# POPULATION GENETIC STRUCTURE AND ADAPTIVE VARIATION IN FISHES WITH REFERENCE TO INDIAN OIL SARDINE (*SARDINELLA LONGICEPS*) AND GREEN CHROMIDE (*ETROPLUS SURATENSIS*)

*A Dissertation Submitted to*

**COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY**

*In Partial Fulfillment of the Requirement for the Degree of*

**DOCTOR OF PHILOSOPHY**

*In*

*Marine Science*

*Under The Faculty of Marine Science*

*Department Of Marine Biology, Microbiology and Biochemistry*

*By*

**WILSON SEBASTIAN**
(*Register No - 4638*)

**ICAR–CENTRAL MARINE FISHERIES RESEARCH INSTITUTE**
**Ernakulam North P.O., Cochin-682018, Kerala, INDIA**

AUGUST 2019

# DECLARATION

I do hereby declare that this thesis entitled "POPULATION GENETIC STRUCTURE AND ADAPTIVE VARIATION IN FISHES WITH REFERENCE TO INDIAN OIL SARDINE (*SARDINELLA LONGICEPS*) AND GREEN CHROMIDE (*ETROPLUS SURATENSIS*)" is an authentic record of research work carried out by me under the guidance and supervision of Dr Sandhya Sukumaran, Senior Scientist, Marine Biotechnology Division, ICAR – Central Marine Fisheries Research Institute, Kochi, in partial fulfilment of the requirement for the award of PhD degree under the Faculty of Marine Science in Cochin University of Science and Technology. The thesis or part thereof has not previously been presented the award of any Degree in any University.

Place: Kochi

Date: 21.08.2019

Wilson Sebastian

Wilson Sebastian

भारतीय कृषि अनुसंधान परिषद
Indian Council of Agricultural Research
केन्द्रीय समुद्री मात्स्यिकी अनुसंधान संस्थान
Central Marine Fisheries Research Institute
(कृषि अनुसंधान एवं शिक्षा विभाग, केन्द्र कृषि एवं किसान कल्याण मंत्रालय, भारत सरकार)
(Department of Agricultural Research and Education, Ministry of Agriculture and Farmers' Welfare, Govt. of India)
पोस्ट बोक्स सं. 1603, एरणाकुलम नोर्थ पी.ओ., कोच्ची - 682 018, केरल, भारत
Post Box No. 1603, Ernakulam North P.O., Kochi - 682 018, Kerala, India
Phone: 91 484 2394867/2391407, Fax: 91 484 2394909/2396685, Email: director.cmfri@icar.gov.in, Web: www.cmfri.org.in

CMFRI Platinum Jubilee - Celebrating 70 years of Excellence in Research

# CERTIFICATE

This is to certify that this thesis entitled "POPULATION GENETIC STRUCTURE AND ADAPTIVE VARIATION IN FISHES WITH REFERENCE TO INDIAN OIL SARDINE (*SARDINELLA LONGICEPS*) AND GREEN CHROMIDE (*ETROPLUS SURATENSIS*)" is an authentic record of research work carried out by Mr. Wilson Sebastian (*Register No - 4638*), M.Sc., under my guidance and supervision at ICAR – Central Marine Fisheries Research Institute, Kochi, in partial fulfilment of the requirement for the award of PhD degree under the Faculty of Marine Science in Cochin University of Science and Technology. The thesis or part thereof has not previously been presented for the award of any Degree in any University.

Sandhya

Place: Kochi
Dr. Sandhya Sukumaran

Date: 21.08.2019
(Supervising Guide)

# TABLE OF CONTENTS

Page No

# List Of Tables

## CHAPTER 1

Page No

## CHAPTER 2

Supplementary Tables

## CHAPTER 3

Page No

Supplementary Tables

## CHAPTER 4

Page No

Supplementary Tables

## CHAPTER 5

Page No

Supplementary Tables

## CHAPTER 6

Page No

Supplementary Tables

## CHAPTER 7

Page No

Supplementary Tables

**CHAPTER 8** Page No

Supplementary Tables

**CHAPTER 9** Page No

Supplementary Tables

## List Of Figures

**CHAPTER 1** Page No

**CHAPTER 2** Page No

Supplimentary Figures

**CHAPTER 3** Page No

Supplementary Figures

## CHAPTER 4

Page No

Supplementary Figures

## CHAPTER 5

Page No

Supplimentary Figures

**CHAPTER 6** Page No

Supplementary Figures

**CHAPTER 7** Page No

Supplimentary Figures

**CHAPTER 8** Page No

Supplimentary Figures

**CHAPTER 9** Page No

Page No

Supplementary Figures

# SUMMARY

This thesis includes ten chapters; the first chapter is a general introduction regarding the importance of investigating population structure and adaptive variation in fishes and the molecular methods and markers used for these studies. The next eight chapters presenting the findings from my PhD. The general conclusions and prospects are described in the last chapter.

The second chapter describes the sequencing of the complete mitochondrial genome and phylogeny of Indian oil sardine, *Sardinella longiceps* (Valenciennes, 1847) and Goldstripe sardinella, *Sardinella gibbosa* (Bleeker, 1849) from the Indian Ocean. The entire mitogenome was amplified by polymerase chain reactions (PCR) using primers that amplify overlapping segments of the entire genome, and the products were subsequently used for direct sequencing. The length of assembled mitogenomes of *S. longiceps* and *S. gibbosa* are 16,613 and 16658 bp respectively, contained the 37 mitochondrial structural genes (two ribosomal RNA, 22 transfer RNA, and 13 protein-coding genes) with the gene order identical to that of typical vertebrates. The major non-coding region between the tRNA Pro and tRNA Phe genes considered as the control (D-loop) region has several characteristic conserved sequence blocks (CSB). In the phylogenetic tree, *S. longiceps* and *S. gibbosa* clustered together with species belonging to the family Clupeidae. Clupeidae and its five subfamilies are not monophyletic. Only three of the nine currently recognised family, Engraulidae, Pristigasteridae and Dussumieriidae formed well-supported monophyletic groups, and the relationships among other groups are not well supported. This study is the first report of the complete mitogenome of two commercially important clupeids from Indian waters which form the baseline for further studies on molecular systematics, population genetics, biogeography, historical demography, adaptive variation and conservation of these species.

The third chapter focuses specifically on the selection and population structure in *S. longiceps* population. By whole mitogenome scanning approach, we investigated the adaptive consequences in the DNA of the most important organelle in bioenergetics, “mitochondrion” for getting insights regarding the spatial and temporal distribution of selective signals which provide clues to its potential for survival and resilience. Indian oil sardines were collected from different eco-regions of the Indian Ocean and analysed for

mitogenomic selection patterns by approximate hierarchical Bayesian method (FUBAR, MEME) and TreeSAAP. Non-coding control region was also analysed for selective constraints. Even though, purifying selection was the dominant force influencing mitogenome evolution, signals of diversifying selection were observed in key functional regions involved in OXPHOS (participating in proton translocation, polypeptide binding in inter-chain domain interface and mito-nuclear interactions) indicating OXPHOS gene regulation as the critical factor to meet enhanced energetic demands during uncertain environmental conditions. A characteristic control region with 38-40bp tandem repeat units under strong selective pressure was also observed. These changes were prevalent in the Western Indian Ocean; mainly in fishes from South Eastern Arabian Sea (SEAS) followed by the Northern Arabian Sea (NAS) and rare in the Eastern Indian Ocean or Bay of Bengal (BoB) populations. Significant $\Theta_{ST}$ values were observed in pairwise analyses using whole-genome data set with NAS population as the most genetically differentiated. The selected sites could be used for further investigations by employing them as genetic tags of locally adapted populations for conservation and management as small pelagic fishes contribute to the food security of developing nations. The accelerated substitution rate observed on SEAS has arisen from enhanced mutational rates due to selective pressures contributed by highly variable oceanic environment characterized by seasonal hypoxia, variable SST and food availability. The sites with signals positive selection could be used for further investigations by employing them as genetic tags of locally adapted populations for conservation and management of Indian oil sardine.

In the fourth chapter, we investigated the genetic stock structure of *S. longiceps* using microsatellite markers by collecting a total of 768 individuals from eight locations along the Indian coast and one from the Gulf of Oman over 2 years (2013-2015). Six polymorphic microsatellite markers revealed significant genetic differentiation between populations with the highest $F_{ST}$ value (0.055) between Oman and Indian coastline. Within the Indian coastline, another major subdivision between Mumbai & Mangalore vs. other regions were detected ($F_{ST}$ value 0.047) which was also confirmed in Barrier analysis with the presence of two strong barriers between these eco-regions. There exist pronounced differences in oceanographic and environmental features between Gulf of Oman, Western Indian Ocean and Eastern Indian Ocean (Bay of Bengal) which may act as barriers for effective dispersal and gene flow resulting in genetic differentiation. Even though the samples collected from Calicut, Kollam, Trivandrum, Chennai and Vizag showed the presence of admixed genotypes, the possible presence of distinct populations in some regions was evident in Bayesian analysis which needs to be confirmed further using more widespread sampling design and powerful markers. The present study provided insights into the biocomplexity and intraspecific diversity of Indian oil sardine populations, which needs to be preserved for maintaining the resilience of these important fishes to climate change and habitat alterations in the Indian Ocean.

The fifth chapter examines the population genetic analysis based on ddRAD data of *S. longiceps* from the Indian Ocean, for population genetic structure and adaptive divergence in the backdrop of oceanic environmental heterogeneity. The analysis was performed with 100 samples collected from Oman sea (OMAN), South Eastern Arabian Sea (SEAS), North East Arabian Sea (NEAS) South West Bay of Bengal (SBOB) and North West Bay of Bengal (NBOB) population, sequenced by ddRAD method. The ddRAD libraries were prepared based on the previously published protocol and sequenced. Population genetic statistics (allele frequencies, percentage of polymorphic loci, nucleotide diversity, Wright's F-statistics $F_{IS}$ and $F_{ST}$) were computed using 'population' program in STACKS v 1.40. 48,076.00 polymorphic RAD loci, with 1SNP and 2 alleles were retained from the 100 samples sequenced, after de novo processing (without genome alignment). The average frequency of major alleles (P), ranged from 0.998-0.999 and average observed heterozygosity (Ob Het) ranged from 0.0017 to 0.0020. The overall nucleotide diversity (π) in *S. longiceps* populations ranged from 0.0015 to 0.0028 with samples from the Oman sea recording the lowest level of nucleotide diversity. The allele frequency spectrum of major alleles across the loci varies slightly across the population and was skewed towards 1.00. The pairwise comparison of genetic differentiation ($F_{ST}$ and $R_{ST}$) and STRUCTURE analysis found that the Oman Sea population was highly differentiated from all other populations, with very high significance. The second level of analysis, with PCA and Least-squares estimates of ancestry proportions, identified another level genetic differentiation between NEAS and other Indian ocean group SEAS, SBOB and NBOB. Among the environmental factors analysed the minimum annual sea surface temperature, chlorophyll-*a* concentration and maximum dissolved oxygen concentration was found to be the predominant factor explaining genetic variation across Indian oil sardine population. The analyses also identified a set of candidate loci associated with sea surface temperature, chlorophyll-*a* concentration dissolved oxygen concentration. The loci identified as the candidate can be the representation of genomic regions of local adaptation and isolated genomic regions of divergence with gene flow in *S. longiceps*. Thus, the signals of cryptic structuring/local adaptation can be used as a starting point for more detailed study to identify the genomic region of genetic divergence in *S. longiceps* and Clupeoids. Reanalysis of the RADseq data with a reference genome-based method is necessary for identifying genome-wide distribution/chromosomal regions of genetic divergence.

In the sixth chapter, the study on the adaptive evolution and sequence divergence in the mitogenome of clupeoid fishes was described. The vertebrate mitochondrial genome (mtDNA) evolving towards a reduced size is not only under deamination related constraints but also translational efficiency-related constraints (codon amino acid usage constraints). The observed H and L strand base pair composition differences and codon usage bias in mtDNA is a response to the above constraints. The mitochondrial oxidative phosphorylation (OXPHOS) produces 95% of a eukaryotic cell's energy and the membrane protein involved in this system is under high functional constraints. However, the metabolic requirements and the selection forces vary across species and habitat in different individuals. We evaluated the adaptive evolution of mitochondrial genome of 70 clupeoids species having a wide distribution in marine, brackish and freshwaters of tropical and temperate regions.

By comparative mitogenomic analysis of 70 Clupeoids, we observed that both tRNA anticodon composition and tRNA position along the mtDNA was determined by deamination related constraints. The nucleotide of the tRNA anticodon in Clupeoids was saturated with guanine (G) or Thymine (T), positioned around the $O_L$ according to their GT content and the protein-coding regions evolved towards a codon usage pattern, in which most of them are complementary to the T/G saturated tRNA anticodons in the genome. We also found a codon usage pattern specific to fresh/brackish water adapted (radiated) fishes, in which codons evolved to adapt to anticodons. They have a codon usage pattern highly complementary to the GT saturated anticodons in Clupeoids, contrary to their marine counterparts. The results suggest that the Clupeoids mitogenomes are adapted to deamination mutations in anticodon sites, during replication and transcription. The codon usage pattern in Clupeoids was shaped by deamination mutations related constraints in mtDNA. The observed codon usage pattern in euryhaline and freshwater Clupeoids may be a result of accelerated directional mutation associated with increased energy requirement for adaptation to the euryhaline and freshwater environment.

The presence and persistence of non-coding regions in mtDNA, known as the control region are against its evolutionary trend, evolving towards a reduced size. It is explained by the presence of binding sites in the control region (conserved sequence blocks-CSBs) for nuclear-organized proteins that regulate mtDNA maintenance and expression. We performed a comparative mitogenomic investigation of the noncoding control region in 70 Clupeoids to study its evolutionary trend. We confirmed the ability of sequence flanking conserved sequence elements in the control region to form stable secondary structures similar to the tRNA. This stable secondary structure was maintained through a selective constraint as evidenced by low mutation rate and compensatory base substitutions in the stem forming regions. This is the first report of compensatory base substitutions among species that confirm secondary structure formation. The tandem repeats present in the control region originated from the repeat sequences involved in secondary structures associated with conserved sequence elements. The nucleotide polymorphism observed along the flanking regions can be explained as errors that occur during the enzymatic replication of secondary structure-forming regions and repeat elements. The evidence for selective constraints on secondary structures emphasizes the role of the control region in mitogenome function.

This study provides evidence for positive selection in the OXPHOS protein complex of distantly related clupeoid species distributed from temperate to tropic and marine to the freshwater environment. We performed positive selection test and relate the observed variation with the functional sites of secondary and tertiary protein structure by homology protein modelling. Most of the known key functional regions are highly conserved across species. The signatures of adaptive variation in the complex are generally concentrated to loop regions of transmembrane proteins that function as proton pumps. Variations were observed in the property of amino acids, codon usage and base composition across lineages with specific metabolic requirements such as marine to fresh/brackish water transition. Insights from our study showed the need for future experimental characterisation of specific mutations with the efficiency of oxidative phosphorylation and its physiological impact which will be useful for predicting the response of organisms to future climate change and mitochondrial DNA based genetic improvement.

In the seventh chapter, I described the characterization of complete mitogenome of Green chromide, *Etroplus suratensis* (Bloch, 1790) from Vembanad Lake, Kerala, India. The entire mitogenome was PCR amplified as contiguous, overlapping segments and sequenced. The assembled mitogenome of *E. suratensis* is 16456 bp circle, contained the 37 mitochondrial structural genes; two ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA (tRNA) genes, and thirteen protein-coding genes, one non-coding control region/D-loop with the gene order identical to typical vertebrates. Low G content and high A+T (53.8%) content were observed along with intergenic overlaps at ATP6 & ATP8, ND4 & ND4L and ND5 & ND6 genes. ATG is used as start codon by all coding genes except CO1 (GTG is the start codon), TAA was used as translation terminators for ND1, ND2, CO1, ATP8, ND4L and ND5 and the remaining genes used incomplete stop codon TA-/T--. An anti-G bias in the third codon positions and high pyrimidine presence in the second codon positions along with proteins containing amino acid encoded by A and C were most frequently observed. The major non-coding region (D-loop) has several characteristic conserved sequence blocks (CSB) like CSB 1, CSB2, CSB3 and promoter region. The phylogenetic analysis revealed several bootstraps supported monophyletic groups with *E. suratensis* as Indo-Sri Lankan taxa. Among cichlids, the groups from South America and Africa are monophyletic in origin. The mitogenomic information generated in the present study will be very valuable for further studies on evolution, taxonomy, conservation, environmental adaptation and selective breeding of this species having aquaculture, ornamental and evolutionary importance.

In the eighth chapter, we investigated the intraspecific diversity and adaptation potential of this species by analysing Cytochrome C Oxidase 1 and control region. Besides, partial mitogenomes and low coverage RAD-sequencing of individuals from the selected geographical regions were also sequenced. Significant genetic differentiation was detected between populations from different ecoregions of India indicating restricted gene flow and population structuring. A recent decline in effective population size was evident which can be attributed to the fragmentation of many coastal habitats in addition to anthropogenic impacts like pollution and reclamation. Signals of positive and diversifying selection observed in the mitogenomes were correlated with habitat characteristics. Habitat specific mutational signals observed have adaptive significance as the populations of the study represented humid tropical climatic zones constituting rainforests in the southwest, semi-arid zones in the southeast and humid subtropical zones in the northeast regions of India. Adaptation to these environmentally heterogeneous habitats generates genotypic and phenotypic variants with specific metabolic/bioenergetic requirements. The observed adaptive mitogenome evolution may be the imprints of this geographic variability, genetic drift and selective forces imparted by the distinctive ecoregions which form their habitats. The reduction in genetic diversity observed calls for management measures to protect the natural genetic diversity of this species as successful aquaculture ventures require replenishment of genetic diversity at fixed intervals by way of the introduction of natural broodstocks.

In the ninth chapter, I described the characterisation of some candidate genes involved in stress responses of fishes from mRNA and genomic DNA of *E. suratensis.* This study reports the complete sequences of Aquaporin 1 (AQP1) gene and partial sequences of genes, Sodium/Potassium-Transporting ATPase subunit alpha-1 (Na/K-ATPase α1 subunit), Osmotic Stress Transcription Factor 1 (OSTF1), Transcription Factor II B (TFIIB), Heat Shock Cognate 71 (HSC71) and Heat Shock Protein 90 (HSP90) obtained from mRNA and genomic DNA of *E. suratensis*. AQP1 gene was 2163 bp long. Its mRNA sequence has 55 bp 5' UTR, 783 bp open reading frame (ORF), 119 bp 3' UTR, three intronic regions and 90% identity with AQP1 of *Oreochromis niloticus.* The partial Na/K-ATPase α1subunit gene obtained 5998 bp length with an ORF of 2213 bp and 12 intronic regions. The partial OSTF1, TF IIB, HSC71 and HSP90 mRNA sequences obtained were 1473 bp, 587 bp, 1708 bp and 151 bp in length respectively. All the genes showed high sequence similarity with respective genes reported from fishes. Comparison of AQP1 and Na/K-ATPase α1 genomic DNA sequence of *E. suratensis* collected from different water system showed two types of AQP1 with one synonymous mutation in exon-1 and higher sequence difference in intronic regions (including addition, deletion, transition and transversion mutations) with few synonymous and non-synonymous mutations in the exons of Na/K-ATPase α1. The sequence information of these major candidate genes involved in stress responses will help in further studies on population genetics, adaptive variations and genetic improvement programs of this cichlid species having aquaculture, ornamental and evolutionary importance.

The last chapter, chapter ten is about the general conclusion of this study. It includes our contributions and future direction.

Mitochondrial and nuclear genomic DNA information/resource developed for *S. longiceps and E. suratensis* is an important contribution to its future genetic studies including taxonomy, conservation, adaptation to environmental clines and evolution.

Complete mitogenome based analysis revealed the phylogenetic relationship of *S. longiceps and E. suratensis* with other fishes.

Mitochondrial and nuclear DNA markers revealed population genetic structure in *S. longiceps and E. suratensis* in their habitats. A very strong genetic structure was identified between Oman and Indian Ocean samples of *S. longiceps* and a comparatively low genetic differentiation in the Indian coastal line samples, between North-East Arabian Sea and others (South-East Arabian Sea, South-West Bay of Bengal and North-West Bay of Bengal).

Very high level of sub-structuring was observed between *E. suratensis* samples collected from Indian water. Reduction in the genetic diversity and population size in the contemporary population were also reported.

The association of genetic differentiation with environmental factors and candidate loci/sites identified as under selective constraints from mitochondrial and nuclear DNA of *S. longiceps and E. suratensis* indicate adaptive variations/local adaptation in their habitats. We must extend similar research on other species by characterising the diversity of its natural populations to improve resilience to changing and uncertain environments. Developing genetic tools for monitoring and managing natural diversity and distribution of organisms in the background of changing climate, will help humans to adjust with the climate change stress.

# ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr Sandhya Sukumaran for allowing me to work in her research laboratory and for guidance and support throughout my studies at CMFRI. I would like to thank Dr PU Zacharia (Head DFD, CMFRI and PI of NICRA project) for providing me with the fund and all the support for the study. I also like to thank Dr KK Vijayan (Former head MBTD, CMFRI), Dr Vijayagopal P (Head in-charge MBTD, CMFRI) and Dr A Gopalakrishnan (Director CMFRI) for their support. I am also grateful to Dr PC Thomas (Former in charge of HRD Cell and former in-charge of CFD, CMFRI), Dr MP Paulton (Senior Technical officer MBTD, CMFRI), Dr AA Mohamed Hatha (Professor CUSAT), all external examiners and unknown reviewers who had reviewed my publications, for giving their time and improved my research.

My sincere thanks to Dr Sajesh Kumar (Head, Department of Biotechnology, Mar. Augusthinose college, Kottayam), my teacher for being such a great help, more than he would know it. I would like to thank all my friends and colleagues at CMFRI for all their help and support throughout my long years in Cochin. I also would like to thank Dr Reynold Peter, Shamal, Dr Subin, Dr Anusree, Nikhil, Dr Sandhya, Jijo, Isha Arshad, Dr Jeena, Arun, Shihab, Nevin, Francis, Laksmi, Prince Jose, Clemancy Tomy and all others for their sincere friendship and make me feel happy in Cochin. Thanks a lot for all the great discussions.

I also would like to thank all scientists and staffs of CMFRI and NICRA for their support. They also actively helped to create this thesis and financially supported me as well. In particular, I have to acknowledge the effort of HRD CELL for their support in university registration procedures.

Finally, my family was the invisible hand in the completion of this thesis. Their support and encouragement were there just when I needed it the most. I would like to thank my father, mother and brothers for their extraordinarily support throughout my life. It is good to know to have a home to be welcomed unconditionally.  Thanks for all your help and company.

## DECLARATION OF CO-AUTHORSHIP/ PREVIOUS PUBLICATION

This thesis incorporates the text of [or substantial parts from] one or more papers [jointly authored research] that I have published or submitted for publication. In all cases, the key ideas, primary contributions, experimental designs, data analysis, interpretation, and writing were performed by the author, and the contribution of co-authors was primarily through the supervision, feedback on the refinement of ideas and editing of the manuscript, etc...

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my work.

This thesis includes [11] original papers that have been previously published/submitted for publication in peer-reviewed journals, as follows:

| Thesis Chapter | Publication title/full citation | Publication status |
|---|---|---|
| *Chapter 2* | Wilson Sebastian, Sandhya Sukumaran, P. U. Zacharia, A. Gopalakrishnan.“The complete mitochondrial genome and phylogeny of Indian oil sardine, *Sardinella longiceps* and Goldstripe Sardinella, *Sardinella gibbosa* from the Indian Ocean.” *Conservation genetics resources*, no 10 (2018): 735-739. | *“published”* |
| *Chapter 3* | Wilson Sebastian, Sandhya Sukumaran, P.U. Zacharia, K.R. Muraleedharan, P.K. Dinesh Kumar, A. Gopalakrishnan. “Signals of selection in the mitogenome provide insights into adaptation mechanisms in heterogeneous habitats in a widely distributed pelagic fish.” *Scientific Reports*, no 10 (2020): 1-14. | *“published”* |
| *Chapter 4* | Wilson Sebastian, Sandhya Sukumaran, P. U. Zacharia, and A. Gopalakrishnan. "Genetic population structure of Indian oil sardine, *Sardinella longiceps* assessed using microsatellite markers." *Conservation Genetics* (2017): 1-14. | *“published”* |
| *Chapter 5* | Wilson Sebastian, Sandhya Sukumaran, P U Zacharia, K.R. Muraleedharan, P.K. Dinesh Kumar, A Gopalakrishnan. Low coverage genotyping by double digested restriction site-associated DNA sequencing (ddRAD seq) in Indian oil sardine, *Sardinella longiceps* for population genetic structure analysis. | *To be submitted.* |
| *Chapter 6* | Wilson Sebastian, Sandhya Sukumaran. “Mitochondrial Genome Evolution of Clupeoid Fishes: Evidence of Positive Selection and Convergent Evolution”.<br>Wilson Sebastian, Sandhya Sukumaran, A. Gopalakrishnan. “Selective constraints in the mitochondrial tRNA and control region of Clupeoid fishes: A comparative mitogenomic investigation”.<br>Wilson Sebastian, Sandhya Sukumaran, A. Gopalakrishnan. “tRNA anticodon composition and codon usage in Clupeoid fishes mitochondrial genome; insight into selection and mechanism of adaptation”. | *Submitted.*<br>*Submitted.*<br>*Submitted.* |
| *Chapter 7* | Wilson Sebastian, Sandhya Sukumaran, A. Gopalakrishnan. “The complete mitochondrial genome and phylogeny of Green chromide, *Etroplus suratensis* from Vembanad Lake, Kerala”. *Indian Journal of Fisheries.* | *“published”* |
| *Chapter 8* | Wilson Sebastian, Sandhya Sukumaran, A. Gopalakrishnan. “Population genetic structure of Green chromide, *Etroplus suratensis* from Indian waters”.<br>Wilson Sebastian, Sandhya Sukumaran, A. Gopalakrishnan. “Signals of adaptive mitogenomic evolution in an indigenous Cichlid, *Etroplus suratensis* (Bloch, 1790) from India”. | *Submitted.* |

|  |  |  |
|---|---|---|
| *Chapter 9* | Wilson Sebastian, Sandhya Sukumaran, P U Zacharia, A Gopalakrishnan. "Isolation and characterization of Aquaporin 1 (AQP1), sodium/potassium-transporting ATPase subunit alpha-1 (Na/K-ATPase α1), Heat Shock Protein 90 (HSP90), Heat Shock Cognate 71 (HSC71), Osmotic Stress Transcription Factor 1 (OSTF1) and Transcription Factor II B (TFIIB) genes from a euryhaline fish, *Etroplus suratensis". Molecular Biology Reports* no. 45 (2018): 2783-2789. | *"published"* |

## Ethical Approval

The studies included in this thesis were conducted on fishes collected from commercial fishery and hence no ethical approval was required.

# *Chapter 1*

## GENERAL INTRODUCTION

### 1.INTRODUCTION

The world is facing its 6th mass extinction and global species are experiencing drastic environmental changes like increase in temperature, coupled with ocean acidification, increases in the length and intensity of drought, flood conditions, and changes in the salinity of coastal areas. It is driven mainly by the human disturbance in the ecosystem (Pimm *et al.* 1995; Thomas *et al*. 2004; Pimm *et al*.2006) and will accelerate over the coming decades. The survivability of a species depends on its vulnerability to environmental changes controlled by its genetic constitution (Frankham *et al.* 2002; Allendorf and Luikart 2009). In short term, animals and plants acclimatize to change in the environment via, phenotypic plasticity and expressing some particular traits responding to the local environmental condition whilst, at the next level of response, the organism may shift their habitat to more favourable areas. The third type of response is via; genetic change leading to adaptive evolutionary change, which generates adaptation to continuously changing environment, beyond the limit of phenotypic plasticity (Gienapp *et al.* 2008). The distributions of many species are expected to shift in coming years and it predicts that many areas currently occupied by species will no longer be suitable for them. At the same time, some species are predicted to benefit from the effects of climate change; some invasive species and even some native species are expected to benefit in this way (Hoffmann *et al.* 2015).

Species-specific constraints are the limiting factor determining the distribution of organisms in the ecosystems (Potvin and Tousignant 1996) and that may limit the ability of populations to adapt to environmental changes such as increasing temperature (Somero 2005; Reusch and Wood 2007). Thus, the survival of a population depends on its level of genetic variation, and the traits that limit the distribution and abundance of species (Frankham *et al.* 2002). Large scale climatic changes will affect the range of distribution, genetic diversity and subpopulation structuring within species observed today (Grant and Bowen 1998; Hewitt 2000; Bradshaw and Holzapfel 2001; Umina *et al.* 2005). There is evidence that genetic change occurs rapidly by responding to climatic changes and it is predicted that

# 1.INTRODUCTION

The world is facing its 6th mass extinction and global species are experiencing drastic environmental changes like increase in temperature, coupled with ocean acidification, increases in the length and intensity of drought, flood conditions, and changes in the salinity of coastal areas. It is driven mainly by the human disturbance in the ecosystem (Pimm *et al.* 1995; Thomas *et al*. 2004; Pimm *et al*.2006) and will accelerate over the coming decades. The survivability of a species depends on its vulnerability to environmental changes controlled by its genetic constitution (Frankham *et al.* 2002; Allendorf and Luikart 2009). In short term, animals and plants acclimatize to change in the environment via, phenotypic plasticity and expressing some particular traits responding to the local environmental condition whilst, at the next level of response, the organism may shift their habitat to more favourable areas. The third type of response is via; genetic change leading to adaptive evolutionary change, which generates adaptation to continuously changing environment, beyond the limit of phenotypic plasticity (Gienapp *et al.* 2008). The distributions of many species are expected to shift in coming years and it predicts that many areas currently occupied by species will no longer be suitable for them. At the same time, some species are predicted to benefit from the effects of climate change; some invasive species and even some native species are expected to benefit in this way (Hoffmann *et al.* 2015).

Species-specific constraints are the limiting factor determining the distribution of organisms in the ecosystems (Potvin and Tousignant 1996) and that may limit the ability of populations to adapt to environmental changes such as increasing temperature (Somero 2005; Reusch and Wood 2007). Thus, the survival of a population depends on its level of genetic variation, and the traits that limit the distribution and abundance of species (Frankham *et al.* 2002). Large scale climatic changes will affect the range of distribution, genetic diversity and subpopulation structuring within species observed today (Grant and Bowen 1998; Hewitt 2000; Bradshaw and Holzapfel 2001; Umina *et al.* 2005). There is evidence that genetic change occurs rapidly by responding to climatic changes and it is predicted that environmental changes decrease population diversity leading to inbreeding depression which limits further adaptive response (Hoffmann *et al.* 2003; Kellermann *et al.* 2006).

The terrestrial ecosystem has been receiving increased attention from conservation biologists compared to the marine ecosystem (Avise 1998a; Laikre *et al.* 2010). Marine ecosystem covering 70% of the world surface is also vulnerable to human disturbance. Now, it is known that several marine species are already extinct or at the risk of extinction (Powles *et al.* 2000; Dulvy *et al.* 2003; Reynolds *et al.* 2005) and many marine fish species have shown boundary shifts in response to increased sea temperature (Perry *et al.* 2005). Now, there is an increased awareness about the need to conserve aquatic biodiversity especially species diversity and intraspecific diversity of harvestable resources (Ryman *et al.* 1995; Nielsen and Kenchington 2001; Smedbol and Stephenson 2001; Ruzzante *et al.* 2003; Ruzzante *et al.* 2006).

India is a major producer of marine and freshwater fishes and largest supplier of fish worldwide. The country has 7517 kilometres of marine coastline, 3,827 fishing villages, and 1,914 traditional fish landing centres. Pelagic and mid-water species contributed about 50% of the 65 commercially important marine fish species harvested in India. Fishing and aquaculture in India employ about 14.5 million people and it is an important sector for food security. Indian fisheries and oceans face many challenges including the rapid reduction of major wild fish stocks, increasing market demands, and environmental challenges, like pollution and climate change. For feeding a growing world population with fish, we need to develop new approaches to managing wild fisheries and practising aquaculture.

Population structure is considered an element of conservation biology (Crandall *et al.* 2000). Thus, for making conservation decisions, we need knowledge on how environmental factors structure species into discrete population units (Moritz 2002; van Tienderen *et al.* 2002) or stocks. The stock identification process involves the identification and characterisation of a self-sustaining group in natural populations. It is a central theme in fisheries science as a basic requirement in stock assessment and fishery management programs (Cadrin *et al.* 2013). Most of the applied population models are based on the assumption that individuals have homogeneous growth, maturity, mortality etc. and a closed life cycle in which young fishes are produced by previous generations within the same population. Any study which targets a living resource, either by field sampling or laboratory study should consider its population during sampling and analytical designs (Cadrin *et al.* 2013). Because of this, the genetic basis of population structure and local adaptation has been gaining more attention recently. Conservation of local population needs more attention because locally adapted populations may have a unique portion of the species genetic character specialized for adaptation to that particular environment (Hilborn *et al.* 2003). Along with this, studies on population genetics and environmental adaptation provide us with a window towards understanding evolutionary processes and unique opportunity to study the behaviour of genetic material in a dynamic natural environment.

Recently, the availability of highly variable, neutral markers has enabled us to explore more about the structuring of natural populations. The neutral markers like microsatellites, mitochondrial DNA, single nucleotide polymorphism, amplified fragment length polymorphism etc. are predominated in conservation and management applications in population genetics (Cadrin *et al.* 2013). They are increasingly being used to understand whether environmental changes influence species at the DNA level, the nature of selection by environmental forces and the potential of populations to respond by evolutionary adaptation (Allendorf and Luikart 2009). This information could be used to find out which population needs more attention to conservation decision making (Frankham *et al.* 2002). These markers have the potential to provide information about shifts in population size due to environmental changes and environmental adaptation (van Straalen and Timmermans2002; Rosenblum *et al.* 2007). Even though the information on population genetic structure is increasing, we know little about how selective forces act on fish populations, because the commonly applied, molecular markers are selectively neutral and it varies only when population size decreases or gene flow is interrupted, hence not applicable to study adaptive variations (Allendorf and Luikart 2009).

The neutral markers can be replaced by adaptive genes or candidate loci that are directly involved in an organism’s response to environmental changes (Allendorf and Luikart 2009). Until recently, before the release of the results of studies like the ENCODE project, it was believed that the non-coding DNA or junk DNA has no functional significance (Ecker *et al.* 2012). In this project, they assigned a biochemical function for 80% of the human genome, much of the functional non-coding DNA is involved in the regulation of the expression of coding genes and the expression of each coding gene is controlled by multiple regulatory regions/sequences placed both near and distant from the gene. Genome-wide association studies have determined that more than 90% of single-base-pair differences/SNPs in sequences that are associated with various diseases fall outside of protein-coding regions. Previously it was not clear how these sequence differences/SNPs could influence disease, however, new gene regulatory sites discovered by the ENCODE project provide a better explanation in many cases (Ecker *et al.* 2012). On the other hand, it is now very clear that species are not a homogenous genetic group of individuals but every individual have a unique DNA. The genetic changes at candidate loci have an important influence on populations that helps them to adapt to future environmental challenges (Etterson 2004) and there is evidence that even single gene polymorphisms can change population growth rate (Hanski and Saccheri 2006).

Until the recent technological developments like next-generation sequencing, microarray, cloning, gene transfer etc., had taken place, it was very complicated and time-consuming to demonstrate genetics of selection in non-model organisms like marine fishes (Cadrin *et al.* 2013; Hoffmann *et al.* 2015). The recent revolution in genomics and other omics technologies is providing better methods for insights into evolutionary processes to above mentioned environmental stress and offers an opportunity to improve conservation planning and management decisions (Hoffmann *et al.* 2015). The next-generation sequencing (NGS) has changed genomic and transcriptomic approaches to fish biology. These modern sequencing tools are also valuable for the discovery, validation and assessment of genetic markers in populations. Population-level genotyping and transcription profiles study has provided opportunities to identify the widespread genomic variation within species. The high-quality genome (multi-individual Whole Genome Sequencing - WGS) and transcriptome assemblies will improve the accuracy and power of characterisation of genomic diversity and association of genotypes with desirable traits and environmental resilience (Hoffmann *et al.* 2015). But still, it is a significant financial challenge when multiple populations are under investigation. There are several economical alternatives to multi-individual WGS. Some NGS methods based on Genotyping by Sequencing (GBS) for genome-wide genetic marker development and genotyping methods that use restriction enzyme digestion of target genomes to reduce the complexity of the target region. It includes reduced-representation sequencing (RRS) using reduced-representation sequencing libraries (RRLs), restriction-site-associated DNA sequencing (RAD-seq) and low coverage genotyping by sequencing (Hoffmann *et al.* 2015). These are applicable to both model organisms with reference genome sequences and non-model species with no existing genomic data. This is also applicable to pooled population samples.

In the last decade, many genome resources have been developed from fish species, including DNA markers, expression sequence tags (ESTs), microarrays, next-generation sequence archives (SRA) database, single nucleotide polymorphic (SNP) genotyping platform, databases of the aquaculture genome project and whole-genome sequence assemblies (Saroglia and Zhanjiang 2012; MacKenzie and Jentoft 2016). Now it is quite easy to find genes under selection and link them to phenotypic changes or population responses. The recent increase in information about the genetic organization and structure of the fish genome and technological development in molecular biology has created a renewed interest in monitoring population response to recent climate changes. It provided new opportunities in population genetics and now it is possible to understand the adaptive divergence in the aquatic environment.

## 2. Objectives Of The Study

- **To sequence and analyses the complete mitochondrial DNA of Indian oil sardine (*Sardinella longiceps* Valenciennes, 1847) and Green chromide (*Etroplus suratensis* Bloch, 1790).**
- **To study the phylogenetic relationship of *S. longiceps* and *E. suratensis* with other species of their respective families.**
- **Identify the population genetic structure and adaptive variation in *S. longiceps* and *E. suratensis* using nuclear and mitochondrial DNA markers.**
- **To understand the adaptive genetic variations in *S. longiceps* and *E. suratensis* using a candidate gene approach and mitochondrial genome.**

## 3. Study Organisms

### INDIAN OIL SARDINE (*Sardinella longiceps*)

Scientific classification-

| | |
|---|---|
| Kingdom: | Animalia |
| Phylum: | Chordata |
| Class: | Actinopterygii (Klein, 1885) |
| Order: | Clupeiformes (Goodrich, 1909) |
| Family: | Clupeidae (G. Cuvier, 1817) |
| Genus: | Sardinella (Valenciennes, 1847) |
| Species: | *S. longiceps*(Valenciennes, 1847) |

Binomial name: *Sardinella longiceps*

This fish belongs to the group of small pelagic fishes, feeds on phytoplankton (diatoms) and zooplankton (copepods). Small pelagic fishes like sardines and anchovies form the largest biomass, supported by upwelling regions in the world oceans and thus they are the largest fishery in the globe (Shin *et al*. 1998). They are undergoing depletion by over-exploitation and environmental shift. The small pelagic fishery is one of the major contributors of annual catch for human consumption as an important source of income, protein source and raw material for canning industry (fish meal, fish oil and bait) (Freon *et al.* 2005; Alder *et al.* 2008; Smith *et al.* 2011; Pikitch *et al.* 2014). Because of this global economic importance, a continuous effort has been taken to study their population genetic structure which had revealed patterns from populations with low levels of genetic differentiation (Karaiskou *et al.* 2004, Kasapidis and Magoulas 2008, Ruggeri *et al.* 2013; Sukumaran *et al.* 2016; Sebastian *et al.* 2017) to purely separated stocks (Limborg *et al.* 2009; Vinas *et al.* 2004; Cheng *et al.* 2015). Other studies are paying attention to taxonomic relationships and speciation and it revealed sympatric (Karaiskou *et al.* 2003; Klossa-Kilia *et al.* 2007; Thomas *et al.* 2014) and allopatric speciation in small pelagic fishes (Parrish *et al.* 1989; Catanese *et al.* 2010; Cheng *et al.* 2011; Laakkonen *et al.* 2013).

The **Indian oil sardine** (***Sardinella longiceps***) is a species of ray-finned fish belongs to the genus Sardinella. It is one of the two most important commercial fishes in Indian waters (with the mackerel) which form the largest pelagic fishery of India, with an annual production of 0.34 million tons (CMFRI 2018). It is a cheap source of protein for millions and it contributes to the majority of income from fishing due to its abundance. It also plays a significant role in trophic ecology and food web as a planktivorous, energy-rich small forage fish species which are consumed in large quantities by apex predators along with other sardines, mackerel and anchovy. Large scale feeding on phytoplankton by sardines helps in transferring energy from one location and time to another.

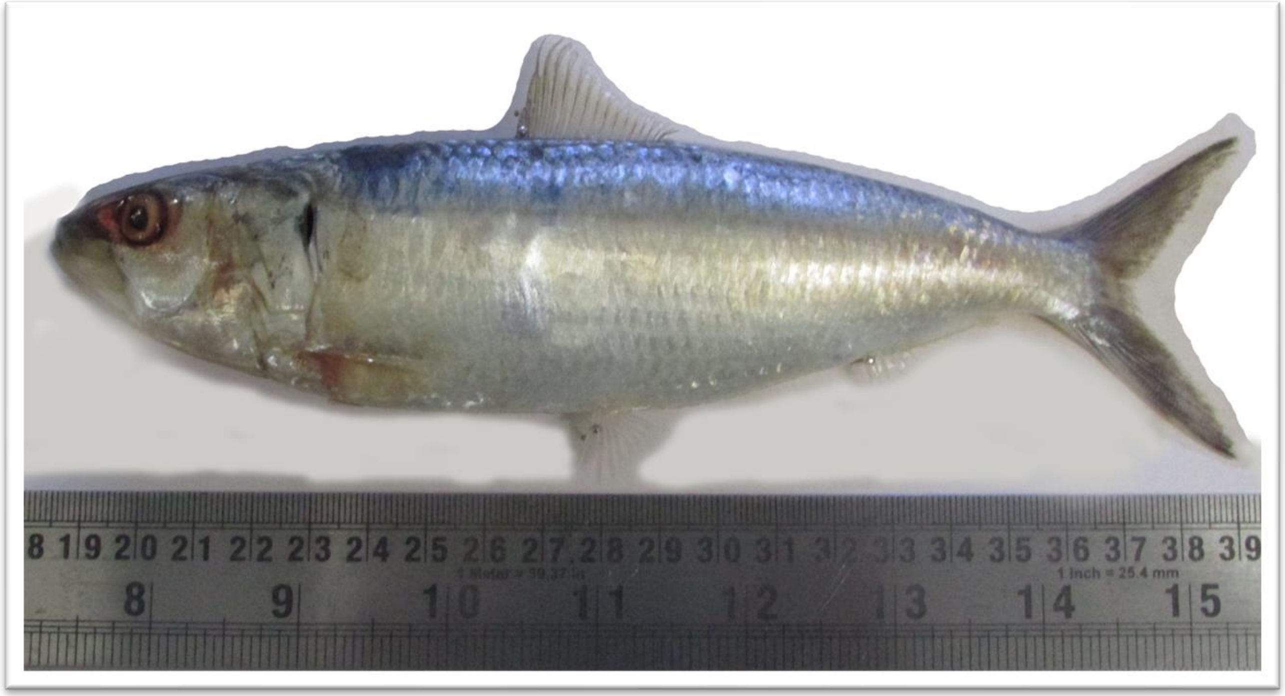


**Fig. 1.1** *Sardinella longiceps*

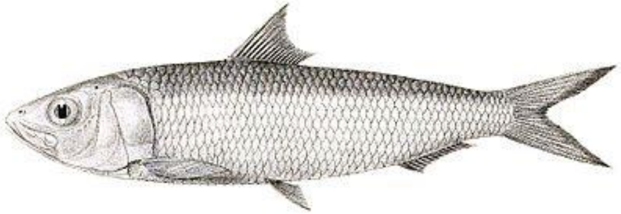


Image Source: FAO Species catalogue Vol. 7. Clupeoid fishes of the world. (Suborder CLUPEOIDEI)
An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, anchovies and wolf-herrings.
Part 1. Chirocentridae, Clupeidae and Pristigasteridae.Whitehead PJP (1985) FAO Fish. Synop., (125)Vol.7 Pt. 1:303 p.

Indian oil sardines inhabit continental shelf waters at a depth range of 20-200m and are distributed along both the east and west coasts of India, Gulf of Oman and Gulf of Aden. They are coastal, pelagic, form schools in coastal waters and undertake localized migrations (Froese and Pauly 2010). It feeds mainly on phytoplankton (especially diatoms) and also on zooplankton. The juveniles are carnivorous but the post-larvae feed mainly on *Fragilaria oceanica* which is considered as a good indicator of oil sardine stock in coastal waters (Nair and Subrahmanyan 1955). Oil sardine reaches a maximum length of 22 cm, and they can weigh up to 200 g (Nair and Chidambaram 1951; Devaraj *et al.* 1997). It breeds once a year at a size of 14-15 cm, during June-July (reported along the south-west coast of India) when temperature and salinity are reduced by the southwest monsoon and spawning peaks in August and September at temperatures from 22 to 28 $^{0}$C (Talwar and Kacker 1984). The exact spawning grounds are not yet located. The pelagic eggs are spherical, range from 1 to 4 mm in diameter and require only 24 hours for development. The pelagic larval development includes minimal movement, but it travels by serpentine swimming and the larval cycle is completed in approximately 40 days (Kuthalingam 1960).

Identifying and characterizing evolutionary significant units of small pelagic fishes for management of its fishery is difficult because these species do not follow the traditional population dynamics models and assumptions (Cadrin *et al.* 2013). They are short-lived, fast-growing, and are characterized by variable levels of natural mortality (Cadrin *et al.* 2013). Their stock size is linked to recruitments, which may be highly variable depending on the presence of an optimal environmental window and hence there exist several hurdles in implementing management measures as compared to longer-lived species (Alheit *et al.* 2009; Alheit *et al.* 2012). Reliability of age-length frequency data and catch-effort analysis is complicated by their size-selective shoaling behaviour (Alheit *et al*. 2009; Alheit *et al.* 2012). There are no species-specific conservation measures in India for Indian oil sardine. But all coastal states have implemented the Marine Fishing Regulation Act by following closed seasons and limiting of fishing zones for different categories of fishing methods. Like other marine pelagic fishes, Indian oil sardine fishery also exhibited fluctuating behaviour, with many population crashes and recoveries during the past century (Devaraj and Martosubroto 1997). Malabar upwelling zones, which is one of the important upwelling zones of the Western Indian Ocean is the largest contributor of the Indian oil sardine fishery and upwelling along these coasts is wind-induced occurring mainly during June–August (Devaraj and Martosubroto 1997; Cailin and Mark SB 2009). Success or failure of sardine recruitment and fishery is highly dependent on the oceanographic features of the Malabar upwelling zone since sardine fishery is dominated by 0 and 1-year class fishes (Devaraj and Martosubroto 1997; Krishnakumar and Bhat 2008). The important factors that determine recruitment and fishery of Indian oil sardines are the intensity of upwelling (Devaraj and Martosubroto 1997), availability of diatoms *F. oceanica* (Nair 1952; Krishnakumar and Bhat 2008) intensity of rainfall (Murty and Edelman 1970), dissolved oxygen, temperature, migratory pattern and survival of the egg and larvae(Devaraj and Martosubroto 1997) and overfishing of immature fishes (Devanesan 1943).

Small pelagic fishes especially sardines of the major oceans like Atlantic and Pacific have been well studied using molecular markers providing improved understanding regarding their biocomplexity and intraspecific diversity (Grant and Bowen 1998; Cadrin *et al*. 2013; Da Silva *et al.* 2015). Indian ocean sardines are less studied using molecular markers compared to their Atlantic and Pacific counterparts except few works using enzyme loci (Venkita Krishnan 1993), cytogenetic, biochemical, and morphometric tools (Mohandas and George 1997) and allozymes (Menezes 1994). All these studies were limited by low sample size and geographical coverage and hence we carried out a comprehensive study using mitochondrial DNA markers (Sukumaran *et al.* 2016) unveiling their historical demography. However, mitochondrial markers were not efficient enough to detect any subpopulation structure in Indian oil sardines and hence we designed a study using whole mitogenome and microsatellite and markers and double digested restriction site-associated DNA sequencing (ddRAD). Microsatellite markers are presumed to be more sensitive markers to detect population subdivision, especially in weakly divergent populations due to their high mutation rates and selective neutrality contributing to high allelic diversity and heterozygosity (DeWoody and Avise 2000; Borrell *et al*. 2012; Putman and Carbone 2014). ddRAD is a powerful tool for genome-wide single nucleotide polymorphism (SNP) markers for non-model organisms like sardines. It has been used for describing fine-scale population structure and detecting the signature of selection. Hence we attempted to understand the population genetic structure and adaptive variation of Indian oil sardines collected from locations along the Indian coast and one location from Gulf of Oman using microsatellite markers developed through a cross-amplification method, whole mitogenome sequencing and ddRAD sequencing.

## GREEN CHROMIDE (*Etroplus suratensis*)

Scientific classification-

| | |
|---|---|
| Kingdom: | Animalia |
| Phylum: | Chordata |
| Class: | Actinopterygii(Klein, 1885) |
| Order: | Cichliformes (R. Betancur *et al*. 2013) |
| Family: | Cichlidae (Bonaparte, 1835) |
| Genus: | Etroplus (G. Cuvier, 1830) |
| Species: | *E. suratensis* (Bloch, 1790) |

Binomial name:*Etroplus suratensis*

The **Green chromide** (***E. suratensis***) is a species of cichlid fish from freshwater and brackish water in southern India and Sri Lanka. Cichlids fishes are candidate species for aquaculture worldwide and as well as for studies on evolutionary diversification and speciation. It is distributed in the fresh and brackish waters of Central and South America, Africa, Madagascar, India and Sri Lanka. The lakes of Africa harbour the richest diversity of Cichlid species, where its massive radiation happened during the past 10 million years. The unique diversity in ecology, morphology and behaviour makes Cichlids good model systems for evolutionary biology, evolutionary genetics and phenotype-genotype relationship studies (Azuma *et al.* 2008). The family Cichlidae comprises more than 700 species, inhabiting fresh and brackish waters of landmasses originated from the Gondwanaland (Africa, South and Central America, India, SriLanka and Madagascar).

Cichlids in India comprise species belonging to the genus *Etroplus,* mainly *Etroplus suratensis*, *Etroplus canarensis* and *Etroplus maculatus*. *E. suratensis* is euryhaline, widely distributed in fresh and brackish water systems of peninsular India and Srilanka whereas *E. maculatus* is mainly confined to brackish waters of Kerala and *E. canarensis* to coastal wetlands of Karnataka. Among these, *E. suratensis* is the most abundant, found in almost all water bodies and river mouths from South Canara in the west coast to the Chilka Lake on the east coast of India (Jayaram 2010; Padmakumar *et al*. 2012) and considered as a very important candidate species for aquaculture.

The family Cichlidae comprises more than 700 species, inhabiting fresh and brackish waters of landmasses originated from the Gondwanaland (Africa, South and Central America, India, Srilanka and Madagascar). The lakes of Africa harbour the richest diversity of Cichlid species, where its massive radiation happened during the past 10 million years. The unique diversity in ecology, morphology and behaviour makes Cichlids good model systems for evolutionary biology, evolutionary genetics and phenotype-genotype relationship studies (Barlow 2000).

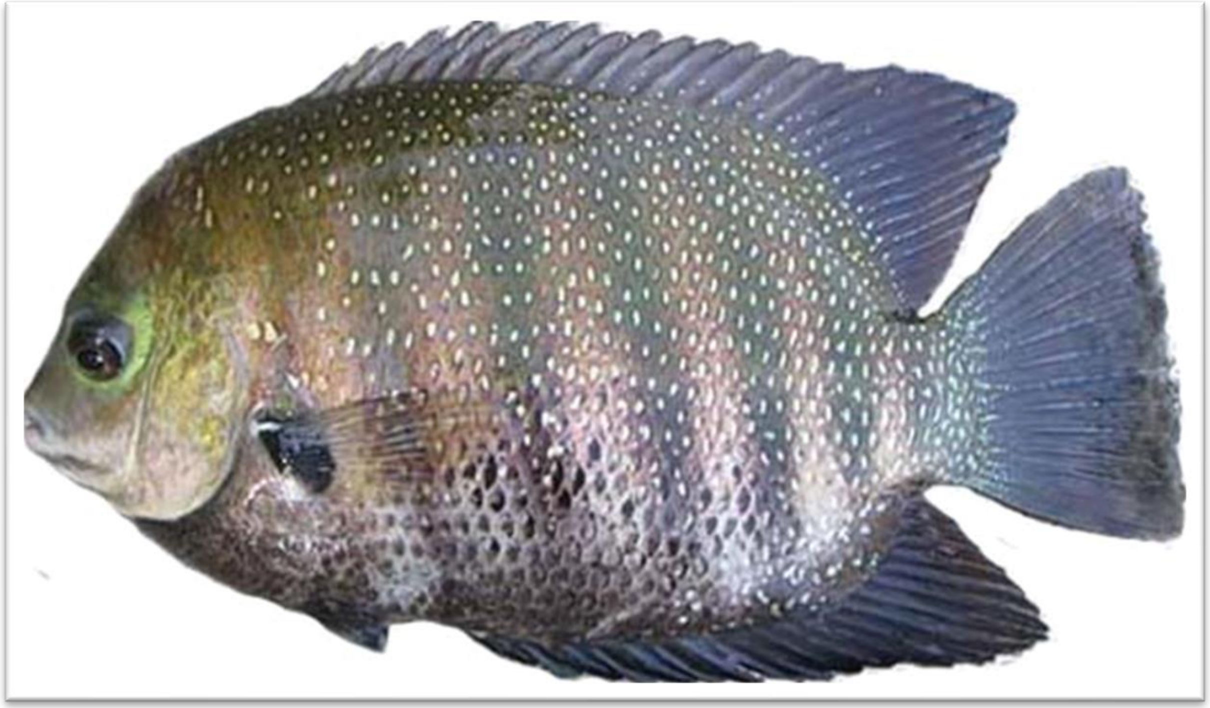


**Fig. 1.2** *Etroplus suratensis*

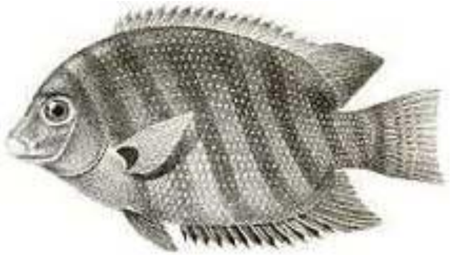


Image source: Francis Day, (1878) The Fishes of India. Volume 2

Allopatric and sympatric speciations have been suggested as mechanisms driving rapid speciation and adaptive radiation of Cichlids in different lakes (Kocher 2004; Watts and Johnson 2004; Genner and Turner 2005). Recent molecular genetic studies in many Cichlids provided evidence for sub-structuring and speciation due to environmental discontinuities over smaller geographical scales (Seehausen 2006; Takeda *et al.* 2013; Brawand *et al.* 2014). Besides, many Cichlids are amenable to culture conditions, making them excellent candidate species for tropical and subtropical aquaculture (Bindu and Padmakumar 2012; Padmakumar *et al.* 2012; Chandrasekar *et al.* 2016).

*E. suratensis*, known as 'Karimeen' in Kerala is characterized by the high adaptive capacity to withstand a wide range of salinity and temperature with highly efficient osmoregulation and cellular stress response mechanisms (Padmakumar *et al.* 2012; Chandrasekar *et al.* 2014) making it popular candidate aquaculture and ornamental species in India (Padmakumar *et al.* 2012). Biology and reproductive characteristics of this species are well known and widely cultured in ponds, tanks, reservoirs and brackish water systems (Jayaprakas *et al.* 1990). The entire life cycle is completed either in fresh or brackish water and it breeds throughout the year with the peak during June to September and February-April (Jayakumar 2002). Even though macrophytes are the predominant food, it also ingests diatoms, molluscs, insects and animal matter (De Silva *et al.* 1984). The backwaters of Kerala are the potential source of *E. suratensis* seed. Wild populations are recorded mainly from Kerala and Tamil Nadu, whereas introduced populations occur in Goa, Andhra Pradesh, Orissa and West Bengal (Jayaram 2010; Abraham 2011). It has also been introduced to other countries like Singapore and Malaysia (Ng and Tan 2010)

Natural populations of *E. suratensis* are facing depletion due to overexploitation (Padmakumar *et al*, 2012) and habitat alterations by the disposal of solid and liquid wastes from increasing urbanisation, increasing number of tourism activities in backwaters/estuaries and a threat from exotic species like *Oreochromis mossambicus* and *Trichogaster trichopterus* (Krishnakumar *et al.* 2009). Despite that, the conservation of natural populations of this species has not attracted sufficient attention from policymakers. Some isolated attempts have been made to create no-fishing zones or aquatic sanctuaries within some of the larger estuaries in addition to captive breeding trials oriented towards conservation (Padmakumar e*t al*. 2012). The major lacunae in conservation efforts are lack of information regarding its present status concerning intra-specific genetic diversity, the potential for adaptation and revival because of the changing climate, habitat and emergence of several diseases in wild and captive populations. Some of the studies have tried to understand phylogenetic relationships and population genetic structure among *E. suratensis* populations using mitochondrial markers indicating the absence of genetic structuring, but all these studies were limited by geographical coverage among sampled populations (Gunawickrama 2012, Dhanya *et al*. 2013, Chandrasekar *et al*. 2016, Alex *et al*. 2016). We did a comprehensive study on understanding the genetic stock structure of *E. suratensis* by collecting samples from all over India. Even though mitogenomes are considered neutral, some of the recent investigations have provided evidence for selection and adaptation in mitochondrial OXPHOS system (Bradbury *et al.* 2008b; Foote *et al.* 2011; Garvin *et al.* 2015a; Teacher *et al.* 2012; Caballero *et al.* 2015) which has been correlated with a wide range of environmental factors like hypoxia (Scott *et al.* 2010), heat stress (Morales *et al.* 2016), cold stress (Cheviron *et al.* 2014; Stier *et al.* 2014), nutrient availability (da Fonseca *et al.* 2008) and expression of genes (Mishmar *et al.* 2003; Garvin *et al.* 2015b; Morales *et al.* 2015). Since *E. suratensis* is widely distributed across geographic gradients, the OXPHOS system may have experienced forces of positive and purifying selection and we investigated this in the present study by characterizing and comparing OXPHOS genes of 37 fish mitogenomes. Also, low coverage genotyping by RAD-seq of fishes collected from different regions of India was employed to understand population connectivity, demographic history and presence of selective forces if any in the nuclear genome.

## 4. Genetic Assessment Of Connectivity Among Fish Population

Natural resource management of fisheries is an important and very critical activity. For that, we need to know the size of a population, its habitat, migratory behaviour, age and size structure, the reproductive pattern of species, natural mortality rate, the rate at which fish are removed by fishing etc. Most of this data can be generated by surveying and analysing the statistical structure of landings. But the most important requirement understanding whether the fish population exist as a single genetic unit or genetically distinct groups (is the species is genetically homogeneous or heterogeneous)? Whether there is any local adaptation in native population?

Fish populations/groups differ in quantitative traits due to the difference in their environment, demographic structure or genetic constitution. But differences in neutral loci are generally considered as the true indicators of stock structure because such difference is generated when gene flow between groups is negligible or reproductively isolated. Studies focusing on genes or gene product involved in selection in the natural population were very much available in 1980 and 1990s (Beaumont 2005). Allozymes and other gene-based markers were replaced by DNA based neutral markers, like microsatellites, mitochondrial DNA, single nucleotide polymorphism, amplified fragment length polymorphism etc., which are predominant in population genetic studies of natural populations (Jarne and Lagoda 1996; van Tienderen *et al.* 2002; Campbell *et al*. 2003; Morin *et al*. 2004). One greater advantage of DNA based markers was that it could be easily generated and applicable to any organism or tissues of varying quality (Cadrin *et al.* 2013). Neutral genetic markers are predominantly used to study population relationship, especially to estimate population parameters like migration rate, genetically effective population size etc., which can’t measure using genetic markers which are under selection (Avise 2004).

The genetic difference can be considered as a sign of population separateness there is an argument that neutral markers are not enough sensitive to identify existing biologically significant structures (Avise 2004). Power of neutral genetic marker depends on population size in question; it is very weak for large populations. Compared to the fresh or brackish water fishes, most of the commercially important marine fishes have a population size which is enough to mask its population genetic difference from neutral markers (Avise 2004). Biologically significant structures can exist even when there is no complete reproductive isolation (Hemmer-Hansen *et al*. 2007). Dispersal rate in populations is different in populations and they have a very high impact on significant genetic differentiation by neutral genetic markers. The slow rate of dispersal allows species to acquire local adaptation to the local ecology. Thus generally genetic differentiation is higher in fresh or brackish water systems. Local adaptation is common in marine fishes even though neutral genetic markers show low-level population genetic differentiation (Teske *et al*. 2019). This indicates the need for use of quantitative traits and genetic marker under selection in fish stock identification.

Now there is an increased interest in identifying molecular genetic markers under selection, for studying adaptive genetic variation in natural populations (Nielsen and Kenchington 2001, McKay and Latta 2002, Luikart *et al*. 2003, Vasemagi and Primmer 2005; Beaumont 2005, Schlotterer and Dieringer 2005, Storz 2005, Joost *et al*. 2007). There are many reasons for focusing on studies to find out molecular genetic changes induced by selection. Distribution of neutral marker variation among populations reveals very little about the adaptive divergence of population. Information on local adaptation not only improves our basic knowledge of evolution but also helps to set management units and priorities for conservation (Fraser and Bernatchez 2001). Second, globally there is an interest in demonstrating the molecular basis of climate change-induced evolution (Gienapp *et al*. 2008; Hoffmann and Willi 2008). Finally, selective harvesting of specific genotype/phenotype is believed to be the main driving force in evolution (Allendorf *et al*. 2008), but we know little about genetic change induced by selection. The 40 years of research effort contribute to revealing huge genetic polymorphism maintained by natural populations (Hedrick 2006; Levasseur *et al*. 2007). But the footprints of selection in the population identified by commonly applied genetic markers are purely by chance (Nielsen *et al.* 2009). So there is a need for a method that accommodates genes responsible for local adaptation and population genetic structure, thus there is a need for genomics.

Population genomics can be defined as a population genetic analysis of a larger number of loci that allow discrimination between locus-specific (selection) and genome-wide effects (drift and migration) (Stinchcombe and Hoekstra 2008). Various types of analyses were used for demonstrating genomic variation distributed in the genome within and between populations. Previously the focus of large sequencing efforts was to develop anonymous markers like microsatellite, AFLP but recently the variations in and around genes are specifically targeted (Bouck and Vision 2007). With the recent advances in sequencing technologies like transcriptome sequencing and whole-genome sequencing approach, large numbers of polymorphic sites are revealed in the coding and non-coding region, which can be exploited using population genomic approaches. Marine fish genomics is still in its infancy, previously it was restricted to model species like Japanese pufferfish and Zebrafish but an array of new large scale projects have been started which are expected to generate knowledge on adaptive variation in the marine environment (Wenne *et al*. 2007).

The employment of recent molecular genetics techniques significantly improves our understanding of the species boundary (Tang *et al.* 2014, Bagley *et al.* 2015, Flot 2015) and population structure (Cowen *et al.* 2007, Henriques *et al*. 2014, Martinez-Takeshita *et al.* 2015). It is providing improved and essential information for developing fisheries management strategies (Reiss *et al.* 2009).

## Types of Genetic Markers_

Differentiating between genotypes with useful characteristic traits is the primary objective in genetics and distinction is not directly based on the traits but the indirect marker-based system. A molecular marker provides polymorphism/allelic variation at a locus of interest. Earlier it was achieved by using phenotypic markers (Begg and Waldman 1999) but now advanced molecular markers developed by molecular biology tools are being used. Molecular markers have been used in many applied biotechnology sectors other than population genetics like genome mapping, phylogenetic reconstruction, forensic application and paternity test. Even though there are many molecular markers, conceptually there are only three basic classes of molecular markers allozymes, DNA sequence polymorphism and, DNA repeats variation.

### *Allozymes*

The first true molecular marker was allozymes and it works on the principle that amino acid variation in protein can be visualized by native gel electrophoresis based on the difference in charge and size caused by amino acid changes. Early studies were using simple starch gel electrophoresis. Bands were visualized by treating gel with a specific staining agent, which contains a substrate of enzyme, co-factor and oxidizing agents. In fishes, biochemical markers like haemoglobin polymorphisms were used initially for characterizing populations of Atlantic cod in the 1960s (Sick 1965a, 1965b). Along with its application in other species, allozyme is used to characterize the genetic population structure of marine fishes with large sample sizes (e.g. Christiansen *et al.* 1976, Winans 1980, Grant and Utter 1980, Kornfield *et al.* 1982). Non-neutral evolution of enzymes (e.g. Hilbish and Koehn 1985; DiMichele *et al.* 1991; Schmidt and Rand 1999) and very less number of useful loci were the limitations of using allozyme markers (e.g. Hulls *et al.* 1996).

*DNA based markers*

Allozymes were replaced with DNA based markers after the arrival of DNA modification methods. DNA based markers survey variation in DNA itself rather than electrophoretic mobility of proteins encoded by DNA. Another important advantage of DNA based markers is that the number of mutations between alleles is countable; it is not possible when allozymes are used.

*RFLPs*- the discovery of restriction endonucleases in the 1960s (by Arber, Smith and Nathans) (Piekarowicz 1979) leads to the generation of a new class of genetic markers called restriction fragment polymorphism (RFLP). It works on the principle that change in the recognition sequence of a restriction enzyme will change the pattern of restriction fragment that it produces. These genetic markers allowed, for the first time, to study noncoding sequences. First DNA based genetic map and first successful association studies were based on RFLP markers (Botstein and White 1980; Kerem *et al*. 1989). RFLP analysis of mitochondrial DNA and ribosomal DNA was very widely used for phylogenetic and population genetic studies of many species including fish populations (Avise 1994). Infinite numbers of RFLP markers are possible but the requirement of suitable hybridization probe prevents its wide application.

*Minisatellites*- Minisatellites contain tandem repeats that show polymorphism in length due to unequal crossing over and other nuclear processes. Similar to RFLP, the first step in minisatellite analysis is digestion of genomic DNA with restriction enzymes followed by electrophoretic separation of DNA fragments. Bands are visualized by hybridization with minisatellites core sequences and it will produce barcode-like band pattern. Because of its high polymorphism, minisatellites have been widely used for forensics and paternity testing (Jeffreys *et al.* 1985). The non-random distribution and complex banding pattern produced by minisatellites prevent its application in population genetics and genome mapping. In addition to that, due to the above reason, a standard population genetic analysis is not possible with minisatellite data. Now single-locus minisatellite is available (Armour *et al*. 1990) but the procedure is still technically complicated and most of the time it needs high-quality molecular DNA.

*PCR –based markers*

One of the important turning points in the history of a molecular marker is the invention of PCR (Saiki *et al.* 1985). It made possible, first in history, to amplify and analyse genomic regions of many individuals without any need of cloning and a large amount of high pure DNA.

*Microsatellites*- Similar to minisatellites, microsatellites are tandemly repeated sequences, but their repeat units are smaller. They are highly polymorphic, abundant and evenly distributed through the genome. Most microsatellite loci are easily amplified by a standard PCR. These advantages popularized microsatellites as a genetic marker for mapping, paternity testing, population genetics etc. (Taulz 1989). This marker also revolutionized marine fish population genetics. These markers and modern statistical techniques (Ryman *et al.* 2006; Waples and Gagiotti 2006) are useful for identifying even low-level structuring found in marine fishes (DeWoody and Avise 2000).

Earlier it was believed that microsatellites mutate by a mechanism called DNA replication slippage which is specific to tandemly repeated sequences (Schlotterer 2000; Ellegren 2000). But the gain and loss of microsatellite repeat unit are more complex which create problems in microsatellite analysis. In addition to that, the PCR stutter bands create difficulties in the automation of microsatellite genotyping.

*RAPDs, ISSRs, IRAPs, and AFLPs*- These markers use PCR primers which can bind with many regions in the genome. Random Amplified Polymorphic DNA-RAPDs use short PCR primers (Williams *et al.* 1990), Inter-Simple Sequence Repeat-ISSRs use primers complement to repeat elements like microsatellites (Zietkiewicz *et al.* 1994), and Inter-

Retrotransposon Amplified Polymorphism- IRAPs use primer complementary to retrotransposons (Kalendar *et al.* 1999). In amplified fragment length polymorphism- AFLPs technique restriction fragments are selectively amplified by adding linkers (Vos *et al.* 1995). In all these techniques, PCR amplification gives multiple bands representing the presence or absence of variations among individuals. The important advantage of these methods is that the previous knowledge about genome or primer sequence is not needed for the study organism (Parsons and Shaw 2002; Castiglioni *et al.* 1998; Cervera *et al*. 2001; Menz *et al*. 2002, Remington *et al*. 1999). But these markers are not reliable because it is very difficult to reproduce the results (Schierwater and Ender 1993).

DNA sequence polymorphism

All the molecular markers that had been discussed up to this point measure DNA variation indirectly but the DNA sequencing and recently popular SNPs detect DNA polymorphism directly.

*SNPs* - SNPs are single nucleotide polymorphism observed widely in the genome. SNPs have been highlighted as the potential marker for future studies in natural population structuring (Morin *et al.* 2004) Various methods are available to identify SNPs, like screening of expression sequence tags (Picoult-Newberg *et al*. 1999; Chen and Sullivan 2003) and generation of whole short gun sequencing using a pool of genomic DNA from many individuals (Weber and Myers 1997; Altshuler *et al*. 2000). High potential for automation is the main advantage of SNPs. SNPs are not only useful for genome mapping but also for characterizing past geographical events such as population expansion and admixture (Brumfield *et al*. 2003). But more loci will have to be screened to get the same accuracy as in microsatellite because information content in few bi-allelic SNPs markers is limited (Kalinowski 2002).

*DNA sequencing* - DNA sequencing of genomic regions of many individuals give most finite genetic information. Recent advances in PCR and sequencing techniques enabled sequence analysis of many individuals (Meyer *et al*. 1999; Schlotterer and Harr 2002; McVean *et al.* 2002). Compared to SNPs, sequencing give full information on analyzed site, it is free from assessment bias and the framework for sequence analysis is well developed (Kreitman 2000).

Other methods for DNA sequence differences

In addition to DNA sequencing and high- throughput SNP genotyping, there are some other methods to detect DNA sequence variation. It is based on the difference in chemical properties of DNA resulting from sequence changes. Most of the common methods detect this change through differences in electrophoretic mobility of DNA.

Some of these methods are denaturing gradient gel electrophoresis, temperature gradient gel electrophoresis, heteroduplex analysis, PCR-RFLP and cleavage amplified polymorphic sequences.

Earlier studies using molecular markers are limited by the availability of methods and types of equipment but the situation has changed. A wide range of genomic resources for a large range of organisms, including non-model organisms is now available. Marker development and analysis can be easily outsourced if experts and instruments are not available.

Forces Acting On Genetic Markers

*Genetic drift*- In a randomly mating population with infinite population size and unaffected by the selection, the frequencies of genes will not change over time. However, in a finite population, allele frequencies will change over generations due to random sampling events/a random sampling of organisms. This evolutionary mechanism is known as *Genetic drift.* The *Genetic drift* occurs in all populations but its intensity varies inversely with the number of breeding individuals in a population., The time required for an allele to be lost from a population by drift is inversely related to the effective population size. Thus its effects are strongest in small populations. As a result, mtDNA (transfer as single copies by females) drift more quickly than nuclear DNA (Birky *et al.* 1989).

Unlike fresh or brackish water fishes, marine species are characterised by high population size and large geographic distribution range which would seem to propose that drift is negligible in the ocean. However, the effects of genetic drift are not determined by the total number of individuals in a population or species. The effective population size, that is the number of individuals participated in the reproduction that contributes genetically to the next generation. Because of the asymmetry in breeding success and larval survival, the effective population size of a species may only be a minute fraction of census population size in many free-spawning aquatic organisms.

*Gene flow*- Theoretically the gene flow is defined as the number of migrants between adjacent populations in each production cycle. This is an important number for ecologists, indicating connectivity between populations. Population genetic models calculate migration rates indirectly as *Nem*, where *m* is the product of the proportion of individuals migrating each generation and *Ne* is the effective population size. Only those individuals that successfully reproduce after migration added to gene flow. There are methods supported with genetic markers to make inferences on gene flow or migration rate between populations (Nielson and Slatkin 2000).

*Mutation rate*- The mutation rate (μ) has an important role in the evaluation of population genetic structure and degree of gene flow. The low mutation rates in the markers may limit the number of variations.in the population and high mutation rates may generate more alleles by mutation before an allele leaving the population where they originated. In general, a marker with high mutation rates (μ approaching m) will overestimate *m,* in frequency-based methods (Slatkin 1995; Neigel 1997). It ultimately leads to the wrong conclusion that the gene flow is higher than that is occurring. This bias is because the estimates of the genetic variation by various models are based on the average difference (allelic or nucleotide difference) between individuals in different populations divided by the average difference between individuals within the populations. In a marker with high mutation rates, the alleles or nucleotides differences begin to saturate and the between-population differences were normalized by within-population differences. Thus it leads to underestimation in the number of differences (allelic or nucleotide difference) between individuals in different populations compared to the differences between individuals in the same population. Higher mutation rates would lead to more (absolute, not relative) differences in markers from populations and make different populations look more similar.

*Selection*- most of the patterns of geographical genetic variation revealed by studies could be explained by some situation involving natural selection acting either directly on the markers or indirectly on markers genetically linked with sites under selection. Selection of different alleles in different populations can increase genetic differentiation. However selective sweeps (where one variant dominate in whole range) or stabilizing selection (where the same selected genotypes are dominated in whole range) could generate homogeneity, which misleads to a conclusion that there are high levels of connectivity.

Several methods have been developed to detect signals of selection in both nuclear and mitochondrial DNA sequences (Skibinski 2000). Simulations studies show that under certain conditions, selection force has an independent action on unlinked loci, so utilizing a broad range of markers is the best way to limit the effect of this force on estimates of gene flow (Slatkin and Barton 1989).

*History*- Similar to the effect of selection, the history of populations also has an important role in shaping the genetic structure of populations. The estimation of many population genetic statistical algorithms like $F_{ST}$, (Neigel 1997) is based on the assumption of equilibrium between processes (such as genetic drift). A recolonized population, after local population extinction, will carry the mark of its genetic history (genetically similar to the source population that provide the individuals for recolonizing population) for a period. In such time interpreting genetic data and inferring genetic connectivity of populations may be misleading. The time needed to go back to equilibrium levels of genetic differentiation after such a demographic event is inversely proportional to the rate of migration between populations. As a result, ignoring historical changes in a population leads to a bias on estimates of gene flow. Species with restricted dispersal will carry the mark of history in its DNA for a longer time than broad dispersers. So, genetic data inferred indirectly from genetic markers of a recolonized population will reflect characters of their source population which leads to an overestimation of gene flow between them. Thus before making conclusions about gene flow between population, we should consider patterns inferred by present-day current patterns (Benzie and Williams, 1997) and direct ecological observations of settlement (Gaines and Bertness 1992; Brown *et al*. 2001) along with genetic markers data.

Patterns of Genetic Differentiation Revealed By Geographic Survey_

The classic marine fishes have large population size, pelagic larvae and wide distribution (Nielsen and Kenchington 2001). Commercially important fishes like fin fishes belong to this group. It includes scombrids (e.g. mackerel, tuna, and bonito), clupeids (e.g. herring, anchovy and sardine), pleuronectids (e.g. plaices, soles and flounders) and gadids (e.g. cod, hake and haddock). In addition to that, some coastal and euryhaline species (like killifish and stickleback) are the primary target of population genetic studies. The study aims to set up fishery management units (Carvalho and Hauser 1994; Hauser and Carvalho 2008). Generally, in marine fishes, intra-population diversity values are high with weak genetic differentiation among populations compared to freshwater fish (DeWoody and Avise 2000). The low level of genetic differentiation may be due to the short population history of marine fishes after post-glacial recolonization and large effective population size (Bradbury *et al*. 2008a) along with the specific character of the marine environment (high ecological homogeneity, lack of dispersal barriers). Even though large genetic differentiation has been identified in marine fishes, it ranges from large geographical level (Bentzen *et al.* 1996; Avise 2000; Heist 2004; OReilly *et al.* 2004; Bremer *et al.* 2005) to small geographical level (Knutsen *et al.* 2003; Pampoulie *et al*. 2004; Hoffman *et al.* 2005; Nielsen *et al.* 2005; Bradbury *et al.* 2008a) with genetic differentiation observed over a few tens of kilometres. An important fact about marine environment is few physical barriers to gene flow, large effective population size and high dispersal capabilities as compared to the freshwater habitats (ponds, lakes and rivers). These features provide a high level of gene flow at the geographical level, which may prevent local adaptation in marine fishes at the geographical level (Kawecki and Ebert 2004). So generally, marine fish populations are less affected by random genetic drift and they respond to even a low level of selective force because locally beneficial alleles have a chance to sweep through populations (Hellberg *et al.* 2002).

In recent investigations, it is clear that the low level of genetic differentiation in marine and some freshwater fish population observed in the studies using neutral markers (neutral region of the genome) is not touching the functionally important genomic region (Leinonen *et al.* 2008). An extensive degree of population differentiation is hidden in the genome of organisms. Even though the evidence for genetic differentiation and adaptive variation is accumulating with high-throughput approaches, it is still scarce in low- throughput approaches.

Several population patterns have been proposed for explaining significant genetic structuring found in the marine environment without any physical barriers to gene flow.

*Closed populations*- Constant genetic differentiation from other populations is the main genetic signature of a *closed population*. The degree of this differentiation depends on the level of spatial and temporal interactions of mutation, drift, migration, selection and population size. When gene flow between populations stopped, the genetic drift and selection will play the key role in differentiation by action on existing genetic variation of each population. The mutations will result in the formation of 'private' alleles because, in the absence of gene flow, genetic variants occur only in the population from which they originated. Over time these private alleles may reach high frequency (via drift or selection) in each isolated population without appearing in other population (Slatkin 1985). The pattern of phylogenetic relationships between haplotypes also indicates a closed population. A pattern in which most similar haplotypes are distributed within populations than between populations indicates a closed population. Because in the absence of gene flow, the new alleles that arise in a population will be from those that already exist within it. In the long run, the continued self-recruitment in closed populations will result in a pattern of reciprocal monophyly because alleles from within-population are more closely related to each other than to those from the distant population, (Cunningham and Collins 1998). Even though the feeble genetic differentiation in haplotype may indicate a closed population, more data is needed to confirm it.

Examples for closed population structure: Highly significant genetic differentiation is common between neighbouring populations of *Tigriopus californicus*, with free-swimming lifestyle (Burton and Feldman 1981; Burton and Lee 1994; Burton 1997). Strong population differentiation among local populations is common among taxa that lack pelagic development (Example, *Excirolana braziliensis* (Lessios *et al.* 1994), gastropods (Hoskin 1997). Organisms showing a strong tendency to stay in or consistently return to a particular area, (natal philopatry) generally have closed populations. Examples: Green turtles (*Chelonia mydas*) (Allard *et al.* 1994) and Atlantic mackerel (*Scomber scombrus*) (Nesbo *et al.* 2000). Closed population is also common in species with pelagic larval stages (Riginos and Nachman 2001). Example, Taiwanese abalone (*Haliotisdiversicolor*) (Conod *et al*. 2002), Haptosquilla pulchella (*Haptosquilla pulchella*) (Barber *et al*. 2000). Analyses of the spatial, temporal samples and temporal stability of allelic frequencies are necessary to detect these type of genetic patterns (Example, *Aequipecten opercularis* (Lewis and Thorpe 1994))

*Abrupt genetic change at a geographical barrier*- This pattern can be defined as a sudden genetic break or an abrupt change in the genetic connection that coincides with a past or present biogeographical barrier formation. Such a pattern is produced when a biogeographical barrier restricted gene flow in a species distribution range for a time which is sufficient for drift to fix new alleles or reciprocal monophyly is produced (Avise *et al.* 1987; Avise 1989). Such abrupt changes can be noticed not only as genealogical history but also as an alteration in gene frequencies or genetic diversity between populations (Ayre *et al.* 1991; Billingham and Ayre 1996). Because of the smaller effective population size, drift to reciprocal monophyly in a divided population appears more rapidly in mtDNA sequences (Birky *et al.* 1989). Thus MtDNA sequences markers are more efficient than nuclear markers to identify this pattern. A set of mtDNA coding or non-coding regions is usually used to investigate reciprocally monophyletic populations. The time taken to arrive at reciprocal monophyly is more when the effective population size is large.

The best example of a marine phylogeographic break is the break that occurred at Cape Canaveral on the eastern coast of Florida. This break is reflected in several marine invertebrates and teleost fishes observed as concordant changes in mtDNA and RFLPs analysis. Example, oyster *Crassostrea virginica* (Reeb and Avise 1990). Another classical biogeographic boundary formation and the resulting phylogeographic break is the Indo-Australian archipelago (Williams and Benzie 1998; Hernawan *et al.* 2017).

If gene flow is possible between populations on either side of the barrier by one or other reasons further intermixing and subsequent recolonization may possible. Such process happens when the conditions act as a barrier in the ecoregions changed, organisms could pass this barrier and establish populations. Sometimes the gene flow may be highly unidirectional as in the anchovy populations (*Engraulis encrasicolus*). The Black Sea haplotype occurs at frequencies near 40% of the Mediterranean, but the Mediterranean sea haplotype is completely absent from the Black Sea (Magoulas *et al.* 1996, 2006). Thus the observed pattern indicated unidirectional gene flow, from the Mediterranean to the Black Sea in the anchovy populations. The biogeographical explanation for the observed pattern is that the Black Sea populations have been repetitively isolated from the Mediterranean

Sea by the Pleistocene climatic changes. mtDNA haplotype distribution of *Acanthinucella spirata* from northern and southern California (Wares and Cunningham 2001; Hellberg *et al.* 2001) is another example. But before such historical interpretations have been made, we should consider some independent verification methods like examination of concordant patterns between species or markers and fossil data. Because sometimes randomly generated lineages may produce the pattern similar to phylogeographic breaks (Avise 1998b). Care should be taken before such historical interpretations have been made.

*Geographic clines*- In this pattern, a constant and progressive change in gene or allele frequencies along a geographic cline can be observed. These patterns are generated by a spatially restricted gene flow between populations and the subsequent generation of a pattern of allele frequency changes. It may also generate when selection gradient acts in geographical clines or by secondary introgression between previously differentiated populations (Endler 1977) (Example, Blue mussel, *Mytilus edulis*, leucine aminopeptidase (*Lap*) locus (Koehn and Siebenaller 1981; Hilbish and Koehn 1985). The frequencies of this allele decline from the mouth of Long Island Sound to its head (Koehn *et al.* 1980; Hilbish and Koehn, 1985). It is concordant with the salinity gradient because, in low salinity environments, this locus is disadvantageous. Another example is geographic clines reported in allozyme and mtDNA marker study of the killifish, *Fundulus heteroclitus* (Ropson *et al*. 1990; Gonzalez- Villasenor and Powers 1990).

*Stepping stone gene flow or isolation by distance*- When gene flow between adjacent populations are restricted to a few migrants (so that the adjacent populations are linked each other via intermediate 'stepping stones'), the genetic identity between the populations decreases with increased geographic distance. Such a pattern in genetic structure is termed as isolation-by-distance (Wright 1943). When populations linked each other via intermediate 'stepping stones, the pairwise genetic distance will be lower for neighbour populations but high for more distant populations. The extent of the connection between genetic distance and geographic distance depends on the interplay of the stepping stones, the genetic drift, the mutation rate and the migration rate among neighbour populations (Slatkin 1993; Hutchison and Templeton 1999). Such patterns have been reported from all oceans, invertebrates and vertebrates. Examples, fishes from the Atlantic Ocean (Pogson *et al*. 2001), Mediterranean Sea (Borsa *et al*. 1997; Naciri *et al.* 1999), Pacific Ocean (Palumbi *et al*. 1997), deep-sea (Vrijenhoek 1997) and solitary coral (*Balanophyllia elegans*)(Hellberg 1995). In red drum (*Sciaenops ocellatus*) and black drum (*Pogonias* *cromis*), two commercially important fish from the Gulf of Mexico (Gold *et al.* 1994; Gold and Richardson 1998), strong isolation by distance has been reported. Isolation by distance was reported in studies of starfish across the entire Pacific Ocean (Benzie and Stoddart 1992a,b).

*Metapopulation*- The recent technological improvements have increased the ability of scientists to collect, analyze and interpret data over spatial and temporal scales. Now we can explain the unique population dynamics, evolution and biogeography of fishes which were not possible earlier. The concept of metapopulation was first established by Richard Levins in 1969. A metapopulation is a spatially structured sub-population or a population of patches connected via dispersal**.** According to Levins' theory (Levins 1969, 1970) the balance between extinction and recolonization rates of local patches that are connected by dispersal. All patches have similar optimum climatic parameters (similar habitat quality), and all habitat areas outside a patch are completely unsuitable. Scientists now used the metapopulation model to explain the dynamics of the population pattern of fishes in the spatial and temporal scale. The book by Kritzer and Sale (2010) provides a detailed review of existing information, understanding and issues in the metapopulation concept.

*Chaotic genetic patchiness*- In some species with pelagic larvae (limpets: Johnson and Black 1984; echinoids: Moberg and Burton 2000; barnacles: Hedgecock 1994; Sotka *et al.* 2014), adult populations show generally lower or no genetic subdivision, but a more spatial and temporal sampling from the same place disclose the existence of genetically differentiated patches. This pattern is known as chaotic or fluctuating genetic patchiness. The spatially and temporal dynamic pattern of pelagic or planktonic larvae recruits explains the reason behind this pattern (Kordos and Burton 1993; Hedgecock 1994; Hedrick 2005). There are three possible explanations for this dynamic population genetic pattern. 1) Source of larvae may differ spatially and temporally, depending on the direction of ocean currents (Kordos and Burton 1993). In such a case, there may be a hidden genetically isolated source population or formerly isolated populations presently making contact (Hare and Avise 1996). 2) Another possibility is that environmental selection on early life stages of species or differential survival rate of genotypes after settlement and before sampling may determine the genetic diversity of larvae. The selection on pelagic larvae of limpet is an example for heterozygotes deficiencies in young bivalves (Green *et al.* 1985; Borsa *et al.* 1991; Burton and Feldman 1982; Hedgecock 1986; Watts *et al*. 1990). 3) Heterogeneous oceanographic factors may affect the reproductive success of marine species. Some species have adaptation for reproducing at localised habitat (Parrish *et al.* 1981; Parrish 1981; Pearse *et al.* 1991; Morgan and Christy 1995; Larry 1995; Christy 2003). It is observed that even though many aquatic species produces millions of gametes, the chances of fertilizing and surviving all of them are not equal. This high variability in reproductive success due to the high failure rates during early life stages suggest that in each season only limited adults may be involved in successful recruitment. It is known as ‘sweepstakes recruitment and it may limit the diversity of recruits (Hedgecock 1986; Hedgecock 1994; Jacobson and MacCall 1995; Caley *et al.* 1996; Flowers *et al.* 2002) and reduce the observed effective population size (Palumbi and Wilson 1990).

*Broad-scale homogeneity*- Some species, especially the marine species with planktotrophic larval phase exhibit high genetic relationship/ low genetic differentiation over a broad geographic range, due to the high gene flow. There are studies using allozyme and mitochondrial markers support the long-distance gene flow in many species. For example; In milkfish, *Chanoschanos* from locations between the Philippine Archipelago to the Hawaiian Islands (Winans 1980), analysis using allozyme loci revealed low levels of genetic differentiation ($F_{ST}$ = 0.039). No significant variation in allelic frequencies over a range spanning more than 12,000 km was also observed. The population studied using Allozyme and mtDNA in *Echinothrix diadema* (Lessios *et al.* 1998; Daniel and Stewart 1998) from Eastern Pacific Barrier (EPB) (a span of over 5400 km) is another example for high levels of gene flow over large geographical distances.

In short, different patterns of geographic genetic differentiation evolve based on the magnitude of the migration rate (m) and the effective population size (Ne). Populations can be completely closed (all recruits from within) when *Nem* is small, or completely open (all recruits from other populations) when *Ne* m is large. Between these two extremes of dispersal, populations may show gradually reduced genetic similarity with increasing geographical isolation owing to restricted dispersal (stepping stone gene flow). More detailed temporal sampling may reveal that open populations consist of mixed cohorts recruited from a relatively small number of breeding adults. In rare cases, the selection on the markers themselves (especially allozymes) may override the forces of ongoing gene flow and drift. Finally, historical effects must always be taken into account, especially when m is small and *Ne* is large.

## 5. Adaptive Genetic Divergence In Fishes

Evolutionary processes in the past have shaped present genetic variation in species and populations to optimize their relative fitness within the environments to which they are exposed through natural selection. Similarly, environmental processes will continue and exert similar selective pressures at local populations to continuously optimize fitness in changing habitats through evolutionary responses based on the present genetic variation. Natural selection may uphold different genotypes in different geographic or ecoregion, thus, population genetic differentiation in some locus may persevere even there is significant gene flow between populations *(*Morgans *et al.* 2014) (Example *Lap* from *Mytilus edulis,* Hsp from *Platicthys flesus* (Hemmer-Hansen *et al.* 2007), *Alticus arnoldorum (*Morgans *et al.* 2014), Thus all the time gene flow is not the most influential power acting on natural population and changing allele frequencies distribution. Sometimes the power of natural selection may be strong enough to overcome the high migration rates/gene flow (Karl and Avise 1992). The appearance of advantageous traits will thus develop over time through selective responses to changing environments. Whereas the nature of future responses of species to environmental change remains difficult to predictions, current patterns of genetic variation within and among species provide us with a window towards understanding evolutionary processes in the past.

To detect local adaptation, various methods have been developed. The process that leads to fitness advantages to local genotypes in the local environment when comparing with non-resident genotypes is called local adaptation (Williams 1966; Kawecki and Ebert 2004). Morphological and genetic differences are strong indications of local adaptation. But linking genetic difference, traits variations and fitness difference between resident and non-resident individuals with local adaptation are very difficult. Such studies are very rare but there is increasing evidence for local adaptation in fishes based on different population genomics approaches.

### *Candidate gene approach*

The applications of candidate genes in population genetics and environmental adaptation have a long history, and it begins from molecular marker analysis (Lewontin 1991). Well studied candidate gene in marine fishes include Pantophysin (Pan I, a membrane protein with unknown function) in stickleback, Haemoglobin (HbI) in Atlantic cod and Walleye Pollock, Lactate dehydrogenase B (Ldh-B involved in glycolysis) in Killifish, Ectodysplasin (EDA, involved in lateral plate armour formation) in Stickleback, heat shock cognate 70 (Hsc70 involved in cellular stress response) in European flounder and major histocompatibility complex (MHC involved in immune response) in Killifish.

Structural or functional genes involved in physiological function and molecular polymorphisms that directly or indirectly affect the phenotypic variation are considered as candidate genes. The evidence of selection can be obtained by DNA sequence comparison or allele frequency-based test. Candidate gene approach can be applied to all marine fishes because the information about the gene can be developed by comparative genome analysis (example, Bargelloni *et al.* 1998; Ford 2001). But the sequence-based approach in marine fishes is very few (example, Ford 2000). This method has the advantage to give information on different environmental forces acting on evolution, but it is time-consuming to locate variation in and around the gene.

The first study by candidate gene approach to detect fish population structure was carried out in the 1960s, and it revealed haemoglobin polymorphism in Atlantic cod populations (Sick 1965b; Moller 1966). Later, studies have demonstrated the difference in haemoglobin affinities between the two haemoglobin variants (Brix *et al.* 1998) and Atlantic cod prefers optimum temperature for their corresponding haemoglobin variant when kept under controlled temperature condition (Petersen and Steffensen 2003). Recently, in the link between haemoglobin genotype and physiological fitness, the two main variants have been demonstrated (Andersen *et al.* 2009). Even though the link between haemoglobin variant and fitness difference in natural population has not yet proved, there is well-supported evidence that haemoglobin is under adaptive evolution in natural populations of Atlantic cod. Membrane protein Pantophysin- Pan I gene is another well-studied candidate gene, which showed signs of adaptive divergence in Walleye Pollock (Canino and Bentzen 2004). In a recent study, comparison between distributions of Pan I allele with allozymes and microsatellites among a natural population of Walleye Pollock revealed a high level of structuring for Pan I locus. These pieces of evidence generate a concept that Pan I gene is under diversifying selection (Canino *et al.* 2005). There are many factors like salinity and temperature suggested being influencing Pan I alleles distribution in the natural population

(Case *et al.* 2005). Even though the function of Pan I gene is unknown it is believed to be affecting fitness and growth behaviour of organisms.

In a recent study, they have adopted a different method by targeting candidate genes for environmental adaptation in the marine environment. Their target was non-coding part of heat shock cognate gene HSC 70 in European flounder (Hemmer-Hansen *et al.* 2007). Heat shock proteins (HSPs) play a major role in eukaryotes in response to stresses like changes in salinity, temperature, pollution etc. (Basu *et al.* 2002). This study leads to the identification of insertion/deletion (indel) polymorphism in HSC locus. Indel genotypes of flounder population were compared with polymorphic microsatellite locus, and they could not find any difference in microsatellite between Baltic Sea and North Sea populations but significant differences were found for HSC locus. Baltic Sea and the North Sea differ in salinity and temperature; this shows a difference in HSC 70 gene in a different environment.

In another study, they proved that there is no difference in the promoter region of the prolactin gene (PrI) in European sea bass, *Dicentrarchus punctatus* (Boutet *et al.* 2008). PrI gene is believed to be playing an important role in salinity tolerance. So the regulatory sequence of PrI gene is believed to be different between two species with two native habitats differing in salinity. Besides, no genetic differences were found between sea bass from marine, brackish and, freshwater environment. So salinity tolerance may be achieved by regulating other genes. Another example is the thermal tolerance and heat shock protein gene expression in common killifish, *Fundulus heteroclitus* (Fangue *et al.*2006).

*Genome scan approach*

This is based on the principle of hitch-hiking (Maynard and Haigh 1974). In this method, the loci identified by genome scan showing significantly higher genetic differentiation than all others (sometimes called outlier loci) between populations due to selection are used as markers. The marker can be AFLP, Microsatellites, or SNPs and EST-linked SNPs or microsatellites from high-throughput transcriptome sequencing. This approach increases the chance to get loci of adaptive variation (Beaumont and Balding 2004; Bonin *et al.* 2006). Statistical tests for outlier are available in many population genetic software packages (Riebler *et al.* 2008).

Compared to freshwater fishes few genome scan studies have been reported in marine fishes (Campbell and Bernatchez 2004; Vasemagi *et al.* 2005). A recent study by genome scan was conducted in three-spined stickleback to know adaptation in marine and freshwater (Makinen *et al.* 2008a). They used 100 EST-based microsatellites and two EDA gene linked indels. In the study, they identified two microsatellites and indels as outliers which show a divergent pattern of genetic variation. Another study was also conducted in the same species (Makinen *et al.* 2008b) by hitch-hiking mapping approach using 24 microsatellites in the flanking region of the candidate gene (Stn 90). They tested for outlier loci in marine and freshwater population and identified many genomic regions showing adaptive evolution.

In a genome scan study to study adaptation in a chemically polluted environment, based on AFLP in Killifish (Williams and Oleksiak 2008) multiple pairwise comparisons were carried out. They identified loci specific to the polluted environment and also loci common to polluted and unpolluted control. This result showed divergence in gene evolution in locally different chemical environments.

In another study in Atlantic cod, (Pogson *et al*. 1995) RFLP loci from cDNA library showed higher differentiation than the allozyme studies. They also found that RFLP loci GM 798, identified as Pantophysin show ten times higher genetic differentiation than the other loci. In a large scale genome scan project, (Moen *et al.* 2008) they have used 318 SNPs from Atlantic cod, for genotyping individuals from Northeast Arctic cod and Norwegian coastal cod. They identified 29 outlier loci (9%), many of them subjected to differential selection. Example for similar studies identified adaptive population divergences is Bradbury *et al*. 2010 and Johansen *et al*. 2011.

All these studies show that adaptive population divergence may be a common phenomenon in high gene flow environment like an ocean ecosystem. Recent studies also have found that the neutral microsatellite loci used in population genetic studies may be under divergent selection/linked to loci under selection (Larsson *et al.* 2007; Skarstein *et al.* 2007; Westgaard and Fevolden 2007)

*Genome mapping_QTL, admixture and association mapping*

Quantitative trait locus can be defined as a stretch of DNA containing linked genes that underlie quantitative traits. It can be used for mapping genomic regions that control the gene involved in a specific quantitative trait variation and they include markers like AFLP, SNPs etc. (Mackay 2001; McKay and Latta 2002). QTL analysis helps to identify the linkage between phenotypic and genotypic variation and provide an explanation for the genetic basis of variation in complex traits. Many QTL studies have been carried out in aquatic species and many studies have been shown that QTL can be shared between species (Somorjai *et al.* 2003).

Association mapping also knew as linkage disequilibrium mapping is based on QTL mapping method that depends on historic linkage disequilibrium to relate phenotype to genotype. It requires a very high number of genetic markers (Hirschhorn and Daly 2005). Instead of genome-wide scanning, targeted approaches using candidate gene or region of genome are also used for association mapping studies (Vasemagi and Primmer 2005).

Admixture mapping uses linkage disequilibrium occurring with a high rate in the natural environment for getting QTL and it requires only a few markers (Smith and O’Brien 2005). Genetic admixture occurs when individuals from geographically separated population begin inbreeding/mixing. It introduces new genetic lineage into a population and slow local adaptation by introducing non-adaptive genes (some time known as gene swamping) which prevent homogenization. Admixture analysis is carried out in inter-specific hybrid zones, which are most commonly found in fishes (Schulte 2001; Nielsen *et al.* 2003).

Very few studies have been conducted in the QTL aspect. Study in Three-spine stickleback is a classic example. In a study using 400 microsatellites (Colosimo *et al.* 2004), they identified Ectodyplasin (EDA) responsible armour plate pattern and evolution of armour plate reduction. In a phylogenetic analysis of the EDA gene sequence, all populations grouped concerning their armour plate pattern in contrast to neutral marker analysis (Colosimo *et al.* 2005). Another contrasting study (Raeymaekers *et al.* 2007) using another EDA-linked microsatellite loci show a high level of structuring in neutral population but

QTL linked to other loci did not show any pattern of divergence. A similar pattern of the result was obtained in other studies (Wang *et al.* 2006).

*Population transcriptomics*

This is based on the hypothesis of King and Washes in 1975 that evolutionary change depends on a change in the mechanism of gene expression regulation than the change within the gene (Khaitovich *et al.* 2006). Microarray targeting thousands of genes, real-time PCR of one or few genes or large scale RNA sequencing were used to demonstrate the genetic basis of population difference in gene expression in response to different environmental stress.

In a study by Whitehead and Crawford (2006), they used different molecular-genetic tools to know the neutral and adaptive genetic variation. They combined data from gene expression and neutral microsatellite of Killifish population kept in a common garden set up. They observed variation in expression of 15% studied genes and a high level of genetic differentiation among the population. In a similar study (Larsen *et al*. 2007, 2008) population from different salinity conditions showed a low level of differentiation in neutral microsatellite loci, but more than 5% of the analysed genes expressed varyingly in the common garden set up. Many studies identified differences in gene expression giving fitness advantage to local populations (Gracey 2007; Schulte 2007; Wittkopp 2007) and they acted as an important component in environmental adaptation in fishes (Cossins and Crawford 2005; Whitehead and Crawford 2006). All these studies demonstrated that gene expression may have an important role in local adaptation (Oleksiak*et al*. 2002; Fisher and Oleksiak2007).

Another approach which is not popular in marine fishes uses large sets of EST associated with candidate gene or QTL linked markers (Rogers and Bernatchez 2005). The combined information from the genome scan and transcription analysis could provide much valuable information.

*Landscape genomics*

Landscape/seascape genetics combine environmental parameters from the geographical position of samples along with genetic markers (Galindo *et al*. 2006; Hansen and Hemmer-Hansen 2007; Selkoe *et al*. 2008). This method can use both neutral markers and markers under selection. The link between genetic divergence with different geographical and environmental patterns can be visualized with this method (Joost *et al*. 2007).

## 6. Statistical Methods For Population Genetic Structure Inference

What qualifies a group of individuals to get differentiated and locally adapted? For measuring and visualizing population structure and adaptive variation within fish species, various methods are used (Crow 1988).

In population genetic studies, the first step is to collect samples of species across the entire study area. Then depending upon the marker used, genotype (for allozymes, microsatellite, SNPs) or haplotype (mt DNA) is assessed for each individual sampled. Data is carefully analysed in various ways to quantify the genetic differentiation. Allele frequency or genotypic frequency is used to investigate the structure of population within species.

Every diploid individual has a pair of chromosomes so that they also have a pair of alleles at each locus in a population. A heterozygote individual has two different alleles at the same locus whereas a homozygote has two of the same alleles at each locus.

Estimation of allele frequency is the starting point of all population genetic analysis. The frequency of an allele *P* is

$$P = 2\,Ho + He/2\,N$$

where *Ho* is the number of the homozygote for that allele, *He* is the number of heterozygotes for the allele and N is the number of individuals scored at the locus.

According to Hardy (1908), once we calculate the frequency of an allele in a population, it is possible to predict the frequency of that allele in offspring from that population. This principle is called the Hardy-Weinberg model. As long as the Hardy-Weinberg model is not affected by environmental processes like selection, lack of random mating, migration etc. allele frequency remains constant from generation to generation (Stem 1943). In a real-life situation, all individuals may not produce the same number of gametes and the gametes may not mix randomly. So even in a population undergoing the Hardy-Weinberg model, there will be a slight deviation from the model in every generation due to a random sampling of organisms. This natural biological variability is called random genetic drift (Crow 2010). This can be variability is calculated as,

Variance of frequency of allele = $P\,(1\text{-}P)/\,2Ne$

Where *P* is the frequency of allele and Ne is effective population size.

Variation in the allele frequency will be small when *Ne* is very large (big population), but in a small population (*Ne* is very low) it will be very high.

Allele frequencies in a population can change over time due to random genetic change and also by selection (one genotype or one allele survive better than others at a locus) or by a pattern of migration or dispersal that may fluctuate in direction or strength over time.

Most applications of genetic data to population questions have used Sewall Wright's island model to relate the geography of gene frequency variation to levels of gene flow (Neigel 1997). In the island model, all populations are linked by equal levels of gene flow, with a proportion of migrants (*m*) every generation. Differences between populations are all assumed to reflect the same parameter and thus are pooled as replicates to provide a single estimate with low variance. Island models may be appropriate in two-population cases, or in describing equally-spaced oceanic islands, but probably do not describe most real population structures (especially those along coastlines) very well. The most commonly considered alternative to the island model is the stepping stone model (Slatkin 1993), in which only adjacent populations exchange migrants. In such circumstances, there is distinct geography, and closer populations are linked by larger amounts of genetic exchange. This seems to reflect the organization of many coastal marine species more accurately, in which dispersal between localities is often related to geographic distance.

Ultimately, conclusions about levels of connectedness between populations are based on the genetic similarity of those populations. Different types of genetic data allow similarity to be assessed and measured in different ways. In frequency-based models, levels of gene flow are estimated as an inverse function of $F_{ST}$, which summarizes departures of heterozygosity from expectations for freely interbreeding populations (Neigel 1997). There can be no degrees of similarity between variants (for example, the similarity of sequences), only degrees of similarity between frequencies of these variants in populations. These models have been applied most often to allozyme data. While the frequencies of different genetic types can also figure into sequence-based models, these data (usually mtDNA sequences) can also provide information about the genealogical relationships of alleles. These models are not so heavily dependent on assumptions of equilibrium as frequency-based models, and as a result, are far better for teasing out the effects of population history from contemporary gene flow (Wakeley 1996; Nielsen and Slatkin 2000).

Formation of the genetically differentiated population in a species is part of its evolution. Because of many physiological or biological reasons, the distribution of species becomes fragmented. For example in the last ice age, aquatic and terrestrial species were fragmented into different habitats. Once fragmented, the allele frequency of most of the loci in that population undergoes random genetic drift. As we have seen earlier, this frequency change will be rapid in small populations when compared to large populations. In this new population, the allele frequency is different from the source population.

In addition to the random genetic drift, local adaptations will occur, involving selection at some loci for particular characteristics and this leads to further differentiation between populations. Such loci under selection can show a higher pattern of differentiation between populations than neutral markers and are sometimes called "outliers".

*F*-statistics ($\mathrm{F}_{\mathrm{ST}}$) and *R*-statistics ($\mathrm{R}_{\mathrm{ST}}$)

There are many types of analysis to quantify genetic variation. One such analysis is F-statistics, developed by Sewall Wright. *F*-statistics or Fixation index was originally developed by Wright (1921) to estimate the effect of inbreeding within samples. According to his definition, this quantity is a correlation coefficient. Later, Wright (1951) redefined this concept to the traditional hierarchical *F*-statistics, $F_{\mathrm{IS}}$, $F_{\mathrm{ST}}$ and $F_{\mathrm{IT}}$ (where *I* stands for individuals, *S* for subpopulations and *T* for the total population) to estimate population subdivisions in a set of populations. He defined $F_{\mathrm{ST}}$, as the correlation between two alleles chosen at random within subpopulations relative to alleles sampled at random from the total population (Wright 1951, 1965). Thus $F_{\mathrm{ST}}$ is a measure of inbreeding due to the correlation among alleles. For example, in two subpopulations with two-allele locus, $F_{\mathrm{ST}}$ will reach a value of one when the two subpopulations are homozygous and a value of zero when the frequencies in the two subpopulations are identical (Wright 1921) (negative values are allowed because correlations vary from –1 to +1).

Therefore, $F_{\mathrm{ST}}$ is a symbol of a measure of the Wahlund principle (Wahlund 1928), (that is, a heterozygote deficiency due to population subdivision) indicated the heterozygote deficit relative to its expectation under Hardy–Weinberg equilibrium (Hartl and Clark 1997).

The principle proposed by Wahlund (1928) can be presented in terms of variance in allele frequency (Wright 1951, 1965; Hartl & Clark 1997):

$F_{\mathrm{ST}} = Vp/[p(1 - p)]$, (When considering a two-allele locus $p$ and $Vp$ are the mean and the variance of the allele frequency among subpopulations)

Thus $F_{\mathrm{ST}}$ quantity is the estimate of the ratio of the observed variance divided by the maximum possible variance (when alleles are fixed in subpopulations).

Later Nei (1977) redefined the fixation indices for multiple alleles as:

$F_{\mathrm{ST}} = (H\mathrm{t} - H\mathrm{s})/H\mathrm{t}$,
($\mathrm{H_S}$ corresponds to the mean heterozygosity averaging over the expected heterozygosity of each subpopulation $\mathrm{H_T}$ corresponds to the expected total heterozygosity of the pooled population) It assumes a diploid locus with alleles in each population, and they are not connected by gene flow and are in Hardy-Weinberg-Equilibrium.

Also, Cockerham and Weir (1987) defined an $F_{\mathrm{ST}}$ related to probabilities of identities:

$F_{\mathrm{ST}} = (f_0 - f_1)/(1 - f_1)$,
where $f_0$ is the probability of identity-in-state for pairs of genes between individuals within subpopulations and $f1$ is the probability of identity-in-state for pairs of genes between individuals within between subpopulations.

Slatkin (1995) devised a statistic, $R_{\mathrm{ST}}$ based on the stepwise mutation model (SMM). According to Slatkin (1995), $R_{\mathrm{ST}}$ can be defined as follows:

$R_{\mathrm{ST}} = (S - S_{\mathrm{w}})/S$,
where $S$ is the average squared difference in allele size between all pairs of alleles, and $S\mathrm{w}$, the average sum of squares of the differences in allele size within each subpopulation.

$R_{\mathrm{ST}}$ is a calculation based on the variances of allele sizes, whereas $F_{\mathrm{ST}}$ is estimated from the variances of allele frequencies. Slatkin (1995) showed that the relationship in equation 4 has the same properties for microsatellites that follow a generalized SMM as does $F_{\mathrm{ST}}$ in the absence of mutation.

Besides, Nei (1973) defined a multiallelic analogue of $F_{\mathrm{ST}}$ among a finite number of subpopulations, called the coefficient of gene differentiation (Nei 1973), $G_{\mathrm{ST}}$.

$G_{\mathrm{ST}} = D_{\mathrm{ST}}/H\mathrm{t} = (H\mathrm{t} – H\mathrm{s})/H\mathrm{t}$,
where $D_{\mathrm{ST}}$ is the average gene diversity between subpopulations, $D_{\mathrm{ST}} = (H\mathrm{t} – H\mathrm{s})$.

A detailed review of *F* statistics and Mutation models used on its algorithms are available in a book by Francois and Nicolas (2002).

Statistical problems associated with population genetic analysis

Along with the availability of modern computational power, alternative ways like contingency table, chi-square test or G-test are used with $\mathrm{F}_{\mathrm{ST}}$ and $\mathrm{G}_{\mathrm{ST}}$ for genetic differentiation, much safe ‘exact’ test can also be employed and these are used in modern genetic analysis software. All these tests tell us whether there is significant heterogeneity in allele frequencies across all populations. Careful removal and rearrangement of populations followed by retesting the data can reveal more details. Care is needed to avoid type one statistical errors.

## 7. Software Packages For Population Genetics

To get accurate results from population genetic studies, it is necessary to correctly analyze and interpret the raw data. Along with the developments in the genotyping technique, new powerful methods have been developed for analysis. In recent years, many computer programs/statistical software packages implemented with these methods have been increasingly available. The statistical software packages are used for aligning both nucleotide and amino acid sequences, analysis of genetic differentiation, analysis of population genetic structure, construction of the phylogenetic tree, identification of demographic history etc. They are successful in hiding the complexity of methodologies used in data interpretation from the user.

Examples of some freely downloadable computer programs/packages commonly used in population genetics data analysis;

Arlequin (http://cmpgunibech/software/arlequin3/),
DnaSP (http://wwwubes/dnasp/),
FSTAT (http://www2unilch/popgen/softwares/fstat.htm)
GDA (http://hydrodictyoneebuconnedu/people/plewis/softwarephp),
Genepop (http://ftpcefecnrsfr/PC/MSDOS/GENEPOP),
GENETIX (http://wwwuniv-montp2fr/~genetix/genetix/genetix.html),
MEGA (http://wwwmegasoftwarenet/),
MSA (http://i122servervu-wienacat/MSA/MSA_download.html),
SPAGeDi (http://wwwulbacbe/sciences/ecoevol/spagedi.html),
BAPS (http://wwwrnihelsinkifi/~jic/bapspage.html),
GeneClass (http://wwwmontpellierinrafr/URLB/index.html),
Geneland (http://wwwinapginrafr/ens_rech/mathinfo/personnel/guillot/Geneland.html),
Structure (http://pritchbsduchicagoedu/software/structure2_1.html),
FDIST2 (http://wwwrubicrdgacuk/~mab/software.html),
LAMARC (http://evolutiongswashingtonedu/lamarc/lamarc_prog.html),
Migrate (http://popgencsitfsuedu/),
Convert (http://wwwagriculturepurdueedu/fnr/html/faculty/Rhodes/Students%20and%20Staff/glaubitz/software.htm),
Formatomatic (http://taylor0biologyuclaedu/~manoukis/Pub_programs/Formatomatic/XML specifications),
Genepop on the web (http://wbiomedcurtineduau/genepop),
MESQUITE (http://mesquiteprojectorg/Mesquite_Folder/docs/mesquite/manual.html),
MR BAYES (http://mrbayescsitfsuedu/),
PHYLIP (http://evolutiongeneticswashingtonedu/phylip.html),
STRUCTURAMA (http://wwwstructuramaorg/)

and some R resources commonly used in population genetics data analysis;

R-project (http://wwwr-projectorg/),
HIERFSTAT (http://www2unilch/popgen/softwares/hierfstat.htm),
Statistical Genetics Resources (http://cranaur-projectorg/src/contrib/Views/Genetics.html).

A detailed description of functionalities and features of major programs used in population genetic analysis available in the review by Excoffier and Heckel 2006.

For handling large volume data like population transcriptomics and reduced representative whole-genome sequencing from NGS based methodologies, highly efficient and fast processing programs are necessary (Lesk 2019). Various tools have been developed to perform different stages of data analysis including quality analysis, filtering, editing, align and assembly. Examples of some commonly used tools for quality analysis of NGS sequences are FastQC, NGSQC, PRINSEQ, FASTX-Toolkit and ContEST (Gollery 2004; Choudhuri 2014).

A detailed review on biocomputing and open-source Bioinformatics tools research is available in books by Haddock and Dunn 2011, Buffalo 2015 respectively. Perl and Python are both perfectly widely used languages for solving a wide variety of biological problems. Many programs written in Perl and Python are available for biological research (Martin 2019). Some of the Programs used for Align/assemble NGS sequence are BFAST, Bowtie, BWA, ELAND, Exonerate, GenomeMapper, GMAP, Gnump, MAQ, MOSAIK, MrFAST and MrsFAST, MUMmer, Novocraft, PASS, RMAP, SeqMap, SHRiMP, Slider, SOAP, SSAHA, SOCS, SWIFT, SX Oligo Search, Vmatch, Zoom etc.

Examples for some program used for RNA-Seq Analysis are; de-novo based- Velvet-Oases, Soapdenovo-Trans (Alternative splicing, differential expression level), Trinity (Reconstruction of transcriptome from RNA-Seq data) and Trans-AByss (Estimate gene expression level, identify potential polyadenylation sites and candidate gene-fusion events); Reference-based Program-Scripture, cufflinks (for details, Korpelainen *et al*. 2014). Some tools used for variant annotation and SNP detection are ANNOVAR, AnnTools, NGS-SNP, Seattleseq, snpEff, SVA, Variant. In addition to this, many genomic data analysis tools Using R programming is also available (Gondro 2015).

## 8. Genomic Resources In Fishes

Huge numbers of genomic resources have been developed from aquatic species. It includes DNA markers, expression sequence tags (EST), microarray, next-generation sequence read archives (SRA) databases, single nucleotide polymorphism (SNPs) genotyping platform, the database for aquaculture genome projects and whole-genome sequence assemblies (Saroglia and Zhanjiang 2012). Allozyme markers have been developed for carp, Atlantic salmon, Atlantic cod, Rainbow trout, Mrigal Karp etc.; mtDNA has been used in Atlantic eels, red drum, Atlantic snapper, carp, red grouper etc.; RAPD markers are available in Atlantic salmon, Asian Arowana etc.; microsatellite markers have been very widely used in Atlantic salmon, Tilapia, Carp, Rainbow trout etc., SNPs have been available in Catfish, Atlantic salmon, Atlantic cod, Japanese flounder, Carp etc..

Earlier the focus of fish genetics programs was to develop Linkage maps, BAC libraries, ESTs and microarray resources in model species but now it is in developing whole-genome and transcriptome assemblies in the model as well as in the nonmodel species (Saroglia and Zhanjiang 2012). The technological advancements in sequencing and computational power have been playing a major role in this development. Many fish species have been fully sequenced including, marine, fresh or brackish water species like *Takifugu rubripes* (puffer fish) (Aparicio *et al*. 2002), *Tetraodon nigroviridis* (puffer fish) (Jaillon *et al.* 2004), *Oryzias latipes* (medaka) (Kasahara *et al*. 2007), *Latimeria chalumnae* ( Lee *et al*. 2013), *Danio rerio* (zebrafish), *Xiphophorus maculatus* (platyfish) (Schartl *et al*. 2013), *Nothobranchius furzeri* (turquoise killifish) (Harel *et al*. 2015) ( Reichwald *et al.* 2015) ( Valenzano *et al*. 2015), *Esox lucius* (northern pike) (Rondeau *et al*. 2014), *Gadus morhua*, Atlantic cod and *Gasterosteus aculeatus*, three-spined stickleback (Jones *et al*. 2012), *Electrophorus electricus* (electric eel) (Gallant *et al*. 2014), *Lepisosteus oculatus* (spotted gar), *Protosalanx hyalocranius* (clearhead icefish) (Liu *et al*. 2017), *Channa argus* (northern snakehead) (Xu *et al*. 2017), *Larimichthys crocea* (large yellow croaker) (Wu *et al*. 2014), *Parachaenichthys charcoti* (Antarctic dragonfish) (Ahn *et al*. 2017), *Sparus aurata* (gilt-head bream) (Pauletto *et al*. 2018), *Atlantic salmon* (Salmo salar) (Lien *et al*. 2016), *Oncorhynchus mykiss* (rainbow trout) (Berthelot *et al*. 2014), *Ictalurus punctatus* (channel catfish) (Liu *et al*. 2016). In addition to this, a large number of ongoing sequencing ventures such as the Genome 10K project which aims to sequence the genome and transcriptome of 10,000 vertebrates, including 4,000 fish genomes is going on (Bernardi *et al*. 2012).

The large genomic information generated has improved the genomic studies and genomic resources have been developed in non-model organisms by using a comparative genomics approach. For example, identification of candidate genes, SNPs, QTL, microsatellite locus etc. in related species, by developing anchor/primer for sequencing region of interest by identifying conserved region across species. It also enabled us to compare and study large genome regions or the entire genome of organisms from different habitats and/ from different genera, which was never possible earlier.

Even though the replacement of traditional markers with DNA-sequences, the new genotyping techniques, computational power and genomic resources has improved population genetic/genomic analysis a lot, we failed to develop and incorporate appropriate realistic mutation model for its better and efficient utilisation. The success of all genetic markers dependent studies is based on the use of the appropriate mutation models, which can trace the underlying mutation processes that generate variations. Still, these processes are poorly understood, because of the complexity of mutation patterns with different markers studied. But of course, it is possible to generate more complex and realistic models for our data with emerging artificial intelligence techniques. We can expect the increased use of artificial intelligence and quantum computing in genetics. Hope that it will improve our understandings about evolution.

*I explore the above-mentioned areas of population genetics and genomics with the help of various bioinformatics/mathematical tools at different spatial and temporal scales concerning Indian oil sardine and Green chromide. Yet the empirical evidence is scarce for some findings, theoretical considerations are adequate to explain it. In the following chapters of this thesis, I explained the materials & methods, results, discussion and conclusions of each experimental study carried out to answer the objectives formulated and the questions that are to be answered. General conclusions and future perspective are described in the last chapter.*

## 9. References

1. Abraham R (2011) *Etroplus suratensis* The IUCN Red List of Threatened Species. The IUCN Red List of Threatened Species 2011:eT172368A6877592
2. Ahn DH, Shin SC, Kim BM, Kang S, Kim JH, Ahn I, Park J, Park H (2017) Draft genome of the Antarctic dragonfish, *Parachaenichthys charcoti*. *GigaScience* 6(8):gix060
3. Alder J, Campbell B, Karpouzi V, Kaschner K, Pauly D (2008) Forage fish: from ecosystems to markets. *Annu Rev Environ Resour* 33:153–166
4. Alex MD, Kumar AB, Kumar US, George S (2016) Analysis of genetic variation in Green Chromide [*Etroplus suratensis* (Bloch)] (Pisces: Cichlidae) using microsatellites and mitochondrial DNA *Indian. J Biotechnol* 15:375-381
5. Alheit J, Pohlmann T, Casini M, Greve W, Hinrichs R, Mathis M, O'Driscoll K, Vorberg R, Wagner C (2012) Climate variability drives anchovies and sardines into the North and Baltic Seas. *Prog Oceanogr* 96(1):128-39
6. Alheit J, Oozeki Y, Roy C (2009) Climate change and small pelagic fish. Cambridge University Press, Cambridge Al-Jufaili SM (2012) Reproductive biology of the Indian oil sardine *Sardinella longiceps* from al-seeb waters off oman. *Fis Aquacult J* 2012:1
7. Allard MW, Miyamoto MM, Bjorndal KA, Bolten AB, Bowen BW (1994) Support for natal homing in green turtles from mitochondrial DNA sequences. *Copeia* 1(1):34-41
8. Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N (2008) Genetic effects of harvest on wild animal populations. *Trends Ecol Evol* 23(6):327–337
9. Allendorf FW, Luikart G (2009) Conservation and the genetics of populations. John Wiley & Sons
10. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407(6803):513-516
11. Andersen O, Wetten OF, De Rosa MC, Andre C, Alinovi CC, Colafranceschi M, Brix O, Colosimo A (2009) Haemoglobinpolymorphisms affect the oxygen-binding properties in Atlantic cod populations. *P Roy Soc B-Biol Sci* 276(1658):833–841
12. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297(5585):1301-1310
13. Armour JA, Povey S, Jeremiah S, Jeffreys AJ (1990) Systematic cloning of human minisatellites from ordered array charomid libraries. *Genomics* 8(3):501-12
14. Avise JC (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution* 43(6):1192-208
15. Avise JC (1994) Molecular Markers, Natural History and Evolution. Chapman and Hall, New York
16. Avise JC (1998a) Conservation genetics in the marine realm. *J Hered* 89(5):377-382
17. Avise JC (1998b) The history and purview of phylogeography: a personal reflection. *Mol Ecol* 7(4):371-9
18. Avise JC (2000) Phylogeography: the history and formation of species. Harvard university press, Cambridge, Massachusetts
19. Avise JC (2004) Molecular Markers, Natural History and Evolution. 2nd edn. Sinauer Associates, Sunderland, Massachusetts
20. Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* 18(1):489-522
21. Ayre DJ, Read J, Wishart J (1991) Genetic subdivision within the eastern Australian population of the sea anemone*Actinia tenebrosa*. *Mar Biol* 109(3):379-90
22. Azuma Y, Kumazawa Y, Miya M, Mabuchi K, Nishida M (2008) Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC Evol Biol* 8(1):215
23. Bagley JC, Alda F, Breitman MF, Bermingham E, van den Berghe EP, Johnson JB (2015) Assessing species boundaries using multilocus species delimitation in a morphologically conserved group of neotropical freshwater fishes, the *Poecilia sphenops* species complex (Poeciliidae). *Plos One* 10(4):e0121139
24. Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2002) Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Mol Ecol* 11(4):659-74

25. Bargelloni L, Marcato S, Patarnello T (1998) Antarctic fish hemoglobins: Evidence for adaptive evolution at subzero temperature. *P Natl Acad Sci USA* 95(15):8670-8675
26. Barlow GW (2000) The Cichlid Fishes: Nature's Grand Experiment in Evolution. Cambridge: Perseus Publishing
27. Basu N, Todgham AE, Ackerman PA, Bibeau MR, Nakano K, Schulte PM, Iwama GK (2002) Heat shock protein genes and their functional significance in fish. *Gene* 295(2):173–183
28. Beaumont MA (2005) Adaptation and speciation: what can $F_{ST}$ tell us? *Trends Ecol Evol* 20(8):435–440
29. Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* 13(4):969–980
30. Begg GA, Waldman JR (1999) An holistic approach to fish stock identification. *Fish Res* 43(1-3):35-44
31. Bentzen P, Taggart CT, Ruzzante DE, Cook D (1996) Microsatellite polymorphism and the population structure of Atlantic cod (*Gadus morhua*) in the northwest Atlantic. *Can J Fish Aquat Sci* 53(12):2706–2721
32. Benzie JA, Stoddart JA (1992a) Genetic structure of outbreaking and non-outbreaking crown-of-thorns starfish (*Acanthaster planci*) populations on the Great Barrier Reef. *Mar Biol* 112(1):119-130
33. Benzie JA, Stoddart JA (1992b) Genetic structure of crown-of-thorns starfish (*Acanthaster planci*) in Australia. *Mar Biol* 112(4):631-639
34. Benzie JA, Williams ST (1997) Genetic structure of giant clam (*Tridacna maxima*) populations in the West Pacific is not consistent with dispersal by present-day ocean currents. *Evolution* 51(3):768–783
35. Bernardi G, Wiley EO, Mansour H, Miller MR, Orti G, Haussler D, O'Brien SJ, Ryder OA, Venkatesh B (2012) The fishes of Genome 10K. *Mar Genom* 7:3–6
36. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C, Labadie K, Alberti A, Aury JM (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun* 22(5):3657
37. Billingham M, Ayre DJ (1996) Genetic subdivision in the subtidal, clonal sea anemone Anthothoe albocincta. *Mar Bio* 125(1):153-63
38. Bindu L, Padmakumar KG (2012) Breeding behaviour and embryonic development in the Orange chromide, *Etroplus maculatus* (Cichlidae, Bloch 1795). *J Mar Biol Assoc India* 54(1):13-19
39. Birky CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121(3):613–627
40. Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol* 23(4): 773–783
41. Borrell YJ, Pinera JA, Sanchez Prado JA, Blanco G (2012) Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES J Mar Sci* 69(8):1357-1371
42. Borsa P, Blanquer A, Berrebi P (1997) Genetic structure of the flounders *Platichthys flesus* and *P. stellatus* at different geographic scales. *Mar Biol* 129(2):233-46
43. Borsa P, Zainuri M, Delay B (1991) Heterozygote deficiency and population structure in the bivalve *Ruditapes decussatus*. *Heredity* 66(1):1
44. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32(3):314
45. Bouck A, Vision T (2007) The molecular ecologist’s guide to expressed sequence tags. *Mol Ecol* 16(5):907–924
46. Boutet I, Quere N, Lecomte F, Agnese JF, Guinand B (2008) Putative transcription factor binding sites and polymorphisms in the proximal promoter of the PRL-A gene in percomorphs and European sea bass (*Dicentrarchus labrax*). *Marine Ecol* 29(3):354-64
47. Bradbury IR, Campana SE, Bentzen P (2008a) Low genetic connectivity in an estuarine fish with pelagic larvae. *Can J Fish Aquat Sci* 65(2):147–158
48. Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, Snelgrove PV, Morris CJ, Gregory RS, Hardie DC, Hutchings JA (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *P Roy Soc Lond B Bio* 277(1701):3725-34
49. Bradbury IR, Laurel B, Snelgrove PVR, Bentzen P, Campana S (2008b) Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life-history. *P Roy Soc B-Biol Sci* 275(1644):1803–1809

50. Bradshaw WE, Holzapfel CM (2001) Genetic shift in photoperiodic response correlated with global warming. *Proc Natl Acad Sci USA* 98(25): 14509–14511
51. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, Turner-Maier J (2014) The genomic substrate for adaptive radiation in African Cichlid fish. *Nature* 513(7518):375-381
52. Bremer JRA, Mejuto J, Gomez-Marquez J, Boán F, Carpintero P, Rodríguez JM, Vinas J, Greig TW, Ely B (2005) Hierarchical analyses of genetic variation of samples from breeding and feeding grounds confirm the genetic partitioning of northwest Atlantic and South Atlantic populations of swordfish (*Xiphias gladius L.*). *J Exp Mar Biol Ecol* 327(2):167–182
53. Brix O, Foras E, Strand I (1998) Genetic variation and functional properties of Atlantic cod haemoglobins: introducing a modified tonometric method for studying fragile haemoglobins. *Comp Biochem Phys A* 119(2):575–583
54. Brown AF, Kann LM, Rand DM (2001) Gene flow versus local adaptation in the acorn barnacle, *Semibalanus balanoides*: Insights from mtDNA polymorphisms. *Evolution* 55(10):1972–1979.
55. Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphism in inferences of population history. *Trends Ecol Evol* 18(5):249-256
56. Buffalo V (2015) Bioinformatics data skills: Reproducible and robust research with open source tools. O Reilly Media, Inc
57. Burton RS (1997) Genetic evidence for long term persistence of marine invertebrate populations in an ephemeral environment. *Evolution* 51(3):993-9
58. Burton RS, Feldman MW (1981) Population genetics of *Tigriopus californicus*. II. Differentiation among neighboring populations. *Evolution* 1(1):1192-1205
59. Burton RS, Feldman MW (1982) Population genetics of coastal and estuarine invertebrates: does larval behavior influence population structure?. In: *Estuarine comparisons*. Academic Press pp. 537-551
60. Burton RS, Lee BN (1994) Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the copepod *Tigriopus californicus*. *P Natl Acad Sci* 91(11):5197-201
61. Caballero S, Duchene S, Garavito MF, Slikas B, Baker CS (2015) Initial evidence for adaptive selection on the NADH subunit Two of freshwater dolphins by analyses of mitochondrial genomes. *Plos One* 10(5):e0123543
62. Cadrin SX, Kerr LA, Mariani S (eds) (2013). Stock identification methods: applications in fishery science. Academic Press
63. Cailin X, Mark SB (2009) Oil sardine (*Sardinella longiceps*) off the Malabar Coast: density dependence and environmental effects. *Fish Oceanogr* 18(5):359–370
64. Caley MJ, Carr MH, Hixon MA, Hughes TP, Jones GP, Menge BA (1996) Recruitment and the local dynamics of open marine populations. *Annu Rev Ecol Syst* 27(1):477-500
65. Campbell D, Bernatchez L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol Biol Evol* 21(5):945–956
66. Campbell D, Duchesne P, Bernatchez L (2003) AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. *Mol Ecol* 12(7):1979–1991
67. Canino MF, Bentzen P (2004) Evidence for positive selection at the pantophysin (Pan I) locus in walleye pollock, *Theragra chalcogramma*. *Mol Biol Evol* 21(7):1391–1400
68. Canino MF, O’Reilly PT, Hauser L, Bentzen P (2005) Genetic differentiation in walleye pollock (*Theragra chalcogramma*) in response to selection at the pantophysin (Pan I) locus. *Can J Fish Aquat Sci* 62(11):2519-2529
69. Carvalho GR, Hauser L (1994) Molecular genetics and the stock concept in fisheries. *Rev Fish Biol Fisher* 4(1):326–350
70. Case RAJ, Hutchinson WF, Hauser L, Van Oosterhout C, Carvalho GR (2005) Macro- and micro-geographic variation in pantophysin (Pan I) allele frequencies in NE Atlantic cod *Gadus morhua*. *Mar Ecol Prog Ser* 301:267–278
71. Castiglioni P, Pozzi C, Heun M, Terzi V, Müller KJ, Rohde W, Salamini F (1998) An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. *Genetics* 149(4):2039-56
72. Catanese G, Manchado M, Infante C (2010) Evolutionary relatedness of mackerels of the genus Scomber based on complete mitochondrial genomes: strong support to the recognition of Atlantic *Scomber colias* and Pacific *Scomber japonicus* as distinct species. *Gene* 452(1):35-43

73. Cervera MT, Storme V, Ivens B, Gusmao J, Liu BH, Hostyn V, Van Slycken J, Van Montagu M, Boerjan W (2001) Dense genetic linkage maps of three Populus species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics* 158(2):787-809
74. Chandrasekar S, Nich T, Tripathi G, Sahu NP, Pal AK, Dasgupta S (2014) Acclimation of brackish water pearl spot (*Etroplus suratensis*) to various salinities: relative changes in abundance of branchial Na+/K+ - ATPase and Na+/K+/2Cl− co-transporter in relation to osmoregulatory parameters. *Fish Physiol Biochem* 40(3):983-996
75. Chandrasekar S, Sivakumar R, Subburaj J, Thangaraj M (2016) Geographical structuring of Indian pearl spot, *Etroplus suratensis* (Bloch, 1790) based on partial segment of the CO1 gene. *Curr Res Microbiol Biotechnol* 45:1536-1539
76. Chen X, Sullivan PF (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J* 3(2):77-96
77. Cheng J, Gao T, Miao Z, Yanagimoto T (2011) Molecular phylogeny and evolution of Scomber (Teleostei: Scombridae) based on mitochondrial and nuclear DNA sequences. *Chinese J Oceanol Limnol* 29(2):297–310
78. Cheng J, Yanagimoto T, Song N, Gao TX (2015) Population genetic structure of chub mackerel *Scomber japonicus* in the Northwestern Pacific inferred from microsatellite analysis. *Mol Biol Rep* 42(2):373–382
79. Cheviron ZA, Connaty AD, McClelland GB, Storz JF (2014) Functional genomics of adaptation to hypoxic cold-stress in high-altitude deer mice: transcriptomic plasticity and thermogenic performance. *Evolution* 68:48-62
80. Choudhuri S (2014) Bioinformatics for beginners: genes, genomes, molecular evolution, databases and analytical tools. Elsevier
81. Christiansen FB, Frydenberg O, Hjorth JP, Simonsen V (1976) Genetics of Zoarces populations. 9. Geographic variation at 3 phosphoglumutase loci. *Hereditas* 83:245- 255.
82. Christy JH (2003) Reproductive timing and larval dispersal of intertidal crabs: the predator avoidance hypothesis. *Revista Chilena de Historia Natural* 76(1):177-185
83. CMFRI 2018 (2018) Annual Report 2017-18. Central Marine Fisheries Research Institute, Kochi
84. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* 307(5717):1928–1933
85. Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D, Kingsley DM (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *Plos Biol* 2(5):635–641
86. Conod N, Bartlett JP, Elliott NG, Evans BS (2002) Comparison of mitochondrial and nuclear DNA analyses of population structure in the blacklip abalone *Haliotis rubra Leach*. *Mar Freshwater Res* 53(3):711-718
87. Cossins AR, Crawford DL (2005) Opinion – Fish as models for environmental genomics. *Nat Rev Genet* 6(4):324–333
88. Cowen RK, Gawarkiewicz G, Pineda J, Thorrold SR, Werner FE (2007) Population connectivity in marine systems an overview. *Oceanography* 20(3):14-21
89. Crandall KA, Bininda-Emonds OR, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends Ecol Evol* 15(7):290-295
90. Crow JF (1988) Eighty years ago. The beginnings of the genetic analysis of population. Genetics 119:473-476
91. Crow JF (2010) Wright and Fisher on inbreeding and random drift. *Genetics* 184(3):609-611
92. Cunningham CW, Collins TM (1998) Beyond area relationships: extinction and recolonization in molecular marine biogeography. In: Molecular approaches to ecology and evolution. 297-321. Birkhauser, Basel
93. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A (2008) The adaptive evolution of the mammalian mitochondrial genome. *BMC genomics* 9(1):119
94. Da Silva C, Booth AJ, Dudley SF, Kerwath SE, Lamberth SJ, Leslie RW, McCord ME, Sauer WH, Zweig T (2015) The current status and management of South Africa's chondrichthyan fisheries. *Afr J Mar Sci* 37(2):233-248
95. Daniel JH, Stewart HB (1998) Oxford University Press, Oxford
96. De Silva SS, Maitipe P, Cumaranatunge RT (1984) Aspects of the biology of the euryhaline Asian Cichlid, *Etroplus suratensis*. *Environ Biol Fish* 10(1-2):77-87
97. Devanesan DW (1943) A brief investigation into the causes of the fluctuations of the annual fishery of the oil sardine of Malabar, *Sardinella longiceps*, determination of its age and an account of the discovery of its eggs and spawning ground. *Madras Fish Bull* 28(1):01-24

98. Devaraj M, Kurup KN, Pillai NGK, Balan K, Vivekanandan E, Sathiadhas R (1997) Status, prospects and management of small pelagic fisheries in India. In: Devaraj M, Martosubroto P (eds) Small pelagic resources and their fisheries in the Asia-Pacific region: proceedings of the APFIC Workshop, pp 91-198
99. Devaraj M, Martosubroto P (eds) (1997) Small pelagic resources and their fisheries in the Asia-Pacific region: proceedings of the APFIC Workshop, pp 91-198
100. DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J Fish Biol* 56(3) 461–473
101. Dhanya AM, Remya M, Biju KA (2013) Morphometric and genetic variations of *Etroplus suratensis* (Bloch) (Actinopterygii: Perciformes: Cichlidae) from two tropical lacustrine ecosystems, Kerala, India. *J Aquat Biol Fisheries* 1(1-2):140-150
102. DiMichele L, Paynter KT, Powers DA (1991) Evidence of lactate dehydrogenase-B allozyme effects in the Teleost, *Fundulusheteroclitus*. *Science* 253(5022): 898-900
103. Dulvy NK, Sadovy Y, Reynolds JD (2003) Extinction vulnerability in marine populations. *Fish Fish* 4(1):25-64
104. Ecker JR, Bickmore WA, Barroso I, Pritchard JK, Gilad Y, Segal E. (2012) Genomics: ENCODE explained. *Nature* 489(7414):52-55
105. Ellegren. H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet* 16(12):551-558
106. Endler JA (1977) Geographic variation, speciation, and clines. Princeton University Press
*107.* Etterson JR (2004) Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. II. Geneticarchitecture of three populations reciprocally plantedalong an environmental gradient in the great plains. *Evolution* 58(7):1459–1471
108. Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nat Rev Genet* 7(10):745-758
109. Fangue NA, Hofmeister M, Schulte PM (2006) Intraspecific variation in thermal tolerance and heat shock protein gene expression in common killifish, *Fundulus heteroclitus*. *J Exp Biol* 209(15):2859–2872
110. Fisher MA, Oleksiak MF (2007) Convergence and divergence in gene expression among natural populations exposed to pollution. *BMC Genomics* 8(1):108
111. Flot JF (2015) Species delimitation’s coming of age. *Syst Biol* 64(6):897–899
112. Flowers JM, Schroeter SC, Burton RS (2002) The recruitment sweepstakes has many winners: genetic evidence from the sea urchin *Strongylocentrotus purpuratus.Evolution* 56(7):1445-53
113. Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, Gilbert MTP, da Fonseca RR (2011) Positive selection on the killer whale mitogenome. *Biol Lett* 7(1):116-118
114. Ford MJ (2000). Effects of natural selection on patterns of DNA sequence variation at the transferrin, somatolactin, and p53 genes within and among chinook salmon (*Oncorhynchus tshawytscha*) populations. *Mol Ecol* 9(7):843-855
115. Ford MJ (2001) Molecular evolution of transferrin: Evidence for positive selection in salmonids. *Mol Biol Evol* 18(4):639-647
116. Francois B, Nicolas L M (2002) The estimation of population differentiation with microsatellite markers.*Molecular Ecology* 11(2):155–165
117. Frankham R, Briscoe DA, Ballou JD (2002) Introduction to conservation genetics. Cambridge university press
118. Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Mol Ecol* 10(12):2741–2752
119. Freon P, Cury P, Shannon L, Roy C (2005) Sustainable exploitation of small pelagic fish stocks challenged by environmental and ecosystem changes: a review. *Bull Mar Sci* 76(2):385–462
120. Froese R, Pauly D (2010) FishBase. http://www.fishbase.org. Accessed 10 November 2013
121. Gaines SD, Bertness MD (1992) Dispersal of juveniles and variable recruitment in sessile marine species. *Nature* 360(6404):579–580
122. Galindo HM, Olson DB, Palumbi SR (2006) Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Curr Biol* 16(16):1622– 1626
123. Gallant JR, Traeger LL, Volkening JD, Moffett H, Chen PH, Novina CD, Phillips GN, Anand R, Wells GB, Pinch M, Güth R (2014) Genomic basis for the convergent evolution of electric organs. *Science* 344(6191):1522-5
124. Garvin MR, Bielawski JP, Sazanov LA, Gharrett AJ (2015a) Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. *J Zool Syst Evol Res* 53(1):1–17

125. Garvin MR, Thorgaard GH, Narum SR (2015b) Differential expression of genes that control respiration contribute to thermal adaptation in redband trout *Oncorhynchus mykiss* gairdneri). *Genome Biol Evol* 7(6):1404–1414
126. Genner MJ, Turner GF (2005) The mbuna Cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish Fish* 6(1):1-34
127. Gienapp P, Teplitsky C, Alho JS, Mills JA, Merila J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Mol Ecol* 17(1):167–178
128. Gold JR, Richardson LR (1998) Mitochondrial DNA diversification and population structure in fishes from the Gulf of Mexico and western Atlantic. *J Hered* 89(5):404-414
129. Gold JR, Richardson LR, Furman C, Sun F (1994) Mitochondrial DNA diversity and population structure in marine fish species from the Gulf of Mexico. *Can J Fish Aquat Sci* 51(S1):205-14
130. Gollery M (2004) Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor, Cold Spring Harbor Laboratory Press,New York
131. Gondro C (2015) Primer to analysis of genomic data using R. Springer,Cham
132. Gonzalez-Villasenor LI, Powers DA (1990) Mitochondrial-DNA restriction-site polymorphisms in the teleost *Fundulus heteroclitus* support secondary intergradation. *Evolution* 44(1):27-37
133. Gracey AY (2007) Interpreting physiological responses to environmental change through gene expression profiling. *J Exp Biol* 210(9):1584–1592
134. Grant WS, Bowen BW (1998) Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. *J Hered* 89(5):415- 426
135. Grant WS, Utter FM (1980) Biochemical genetic variation in walleye pollock, *Theragra chalcogramma* – population structure in the Southwestern Bering Sea and the Gulf of Alaska. *Can J Fish Aquat Sci* 37(7):1093-1100
136. Green RH, Singh SM, Bailey RC (1985) Bivalve molluscs as response systems for modelling spatial and temporal environmental patterns. *Sci Total Environ* 46(1-4):147-169
137. Gunawickrama KS (2012) Morphological heterogeneity and population differentiation in the green chromid *Etroplus suratensis* (Pisces: Cichlidae) in Sri Lanka. *Ruhuna J Sci* 2(1):70-81
138. Haddock SH, Dunn CW (2011) Practical computing for biologists. Sinauer Associates,Sunderland, MA, USA
139. Hansen MM, Hemmer-Hansen J (2007) Landscape genetics goes to sea. *J Biol* 6(3):6
140. Hanski I Saccheri I (2006) Molecular-level variation affects population growth in a butterfly metapopulation. *Plos Biol* 4(5):719–726
141. Hardy HG (1908) Mendelian proportions in a mixed population. *Science* 28;49-50
142. Hare MP, Avise JC (1996) Molecular genetic analysis of a stepped multilocus cline in the American oyster (Crassostrea virginica). *Evolution* 50(6):2305-2315
143. Harel I, Benayoun BA, Machado B, Singh PP, Hu CK, Pech MF, Valenzano DR, Zhang E, Sharp SC, Artandi SE, Brunet A (2015) A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. *Cell* 160(5):1013-1026
144. Hartl DL, Clark AG, Clark AG (1997) Principles of population genetics (Vol. 116). MA: Sinauer associates, Sunderland
145. Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish Fish* 9(4):333–362
146. Hedgecock D (1986) Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *B Mar Sci* 39(2):550-64
147. Hedgecock D (1994) Temporal and spatial genetic structure of marine animal populations in the California Current. California Cooperative Oceanic Fisheries Investigations Reports 35:73-81
148. Hedrick P (2005) Large variance in reproductive success and the Ne/N ratio. *Evolution* 59(7):1596-1599
149. Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: the age of genomics. *Annu Rev Ecol Evol* S 37:67–93.
150. Heist EJ (2004) Genetics of sharks, skates and rays. In: Carrier JC, Musick JA, Heithaus MR (eds) Biology of Sharks and their Relative. CRC Press, New York, pp 471–485
151. Hellberg ME (1995) Stepping-stone gene flow in the solitary coral *Balanophyllia elegans*: equilibrium and nonequilibrium at different spatial scales. *Mar Biol* 123(3):573-81
152. Hellberg ME, Balch DP, Roy K (2001) Climate-driven range expansion and morphological evolution in a marine gastropod. *Science* 292(5522):1707-10
153. Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *B Mar Sci* 70(1):273-90

154. Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V (2007) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platicthys flesus L.*). *Heredity* 99(6):592–600
155. Henriques R, Potts WM, Santos CV, Sauer WH, Shaw PW (2014) Population connectivity and phylogeography of a coastal fish, *Atractoscion aequidens* (Sciaenidae), across the Benguela current region: evidence of an ancient vicariant event. *Plos One* 9:e87907
156. Hernawan UE, van Dijk KJ, Kendrick GA, Feng M, Biffin E, Lavery PS, McMahon K (2017) Historical processes and contemporary ocean currents drive genetic structure in the seagrass T halassia hemprichii in the Indo-Australian Archipelago. *Mol Ecol* 26(4):1008-1021
157. Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405(6789):907-913
158. Hilbish TJ, Koehn RK (1985) The physiological-basis of natural-selection at the LAP Locus. *Evolution* 39(6):1302-1317
159. Hilborn R, Quinn TP, Schindler DE, Rogers DE (2003) Biocomplexity and fisheries sustainability. *P Natl Acad Sci* 100(11):6564-6568
160. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6(2):95-108
161. Hoffman EA, Kolm N, Berglund A, Arguello JR, Jones AG (2005) Genetic structure in the coral-reef-associated Banggai cardinalfish, Pterapogon. *Mol Ecol* 14(5):1367-1375
162. Hoffmann A, Griffin P, Dillon S, Catullo R, Rane R, Byrne M, Jordan R, Oakeshott J, Weeks A, Joseph L, Lockhart P (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. *BMCClim Chang Responses* 2(1):1
163. Hoffmann AA, Hallas RJ, Dean JA, Schiffer M (2003) Low potential for climatic stress adaptation in a rainforest *Drosophila* species. *Science* 301(5629):100–102
164. Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nat Rev Genet* 9(6):421-432
165. Hornell J (1910) Report on the results of a fishery cruise along the Malabar coast and to the Laccadive Islands in 1908. *Madras FishBull* 4:71
166. Hoskin MG (1997) Effects of contrasting modes of larval development on the genetic structures of populations of three species of prosobranch gastropods. *Mar Biol* 127(4):647-56
167. Hulls DM, Moritz C, Mable BK (1996) Molecular Systematics. Sinauer Associates, Sunderland
168. Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53(6):1898-914
169. Jacobson LD, MacCall AD (1995) Stock-recruitment models for Pacific sardine (*Sardinops sagax*). *Can J Fish Aquat Sci* 52(3):566-77
170. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431(7011):946
171. Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends Ecol Evol* 11(10):424-429
172. Jayakumar M (2002) Wetland conservation and Management in Kerala. State Committee on Science Technology and Environment, Thiruvananthapuram, Kerala, INDIA
173. Jayaprakas V, Nair NB, Padmanabhan KG (1990) Sex ratio, fecundity and length-weight relationship of the Indian pearl spot, *Etroplus suratensis* (Bloch). *J Aquacult Trop* 5(2):141-148
174. Jayaram KC (1991) The freshwater fishes of the Indian region. Narendra Publishing House, New Delhi
175. Jayaram KC (2010) The Freshwater Fishes of the Indian Region. Narendra Publishing House, Delhi, INDIA
176. Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable ‘minisatellite’ regions in human DNA. *Nature* 314(6006):67-73
177. Johansen SD, Karlsen BO, Furmanek T, Andreassen M, Jørgensen TE, Bizuayehu TT, Breines R, Emblem A, Kettunen P, Luukko K, Edvardsen RB (2011) RNA deep sequencing of the Atlantic cod transcriptome. Comparative Biochemistry and Physiology Part D: *Genom Proteomics* 6(1):18-22
178. Johnson MS, Black R (1984) Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. *Evolution* 38(6):1371-83
179. Johnson MS, Black R (1998) Increased genetic divergence and reduced genetic variation in populations of the snail *Bembicium vittatum* in isolated tidal ponds. *Heredity* 80(2):163
180. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55

181.Joost S, Bonin A, Bruford MW, Despres L, Conord C, Erhardt G, Taberlet P (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol Ecol* 16(18):3955-3969
182.Kalendar R, Grob T, Regina MI, Suoniemi A, Schulman A (1999) RAP and REMAP: two new retrotransposon- based DNA fingerprinting techniques. *Theor Appl Genet* 98(5):704-711
183.Kalinowski ST (2002) How many alleles per locus should be used to estimate genetic distances? *Heredity* 88(1):62-65
184.Karaiskou N, Apostolidis AP, Triantafyllidis A, Kouvatsi A, Triantaphyllidis C (2003) Genetic identification and phylogeny of three species of the genus Trachurus based on mitochondrial DNA analysis. *Mar Biotechnol* 5(5):493–504
185.Karaiskou N, Triantafyllidis A, Triantaphyllidis C 2004. Shallow genetic structure of three species of the genus Trachurus in European waters. *Mar Ecol Prog Ser* 281:193–205
186.Karl SA, Avise JC (1992) Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* 256(5053):100-2
187.Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447(7145):714
188.Kasapidis P, Magoulas A (2008) Development and application of microsatellite markers to address the population structure of the horse mackerel *Trachurus trachurus*. *Fish Res* 89(2):132-135
189.Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecol Lett* 7(12):1225–1241
190.Kellermann VM, van Heerwaarden B, Hoffmann AA, Sgro CM (2006)Very low additive genetic variance and evolutionary potential in multiple populations of two rainforest Drosophila species. *Evolution* 60(5):1104–1108
191.Kerem BS, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science* 245(4922):1073-80
192.Khaitovich P, Enard W, Lachmann M, Paabo S (2006) Evolution of primate gene expression. *Nat Rev Genet* 7(9):693–702
193.Klossa-Kilia E, Papasotiropoulos V, Tryfonopoulos G, Alahiotis S, Kilias G (2007) Phylogenetic relationships of *Atherina hepsetus* and *Atherina boyeri* (Pisces: Atherinidae) populations from Greece, based on mtDNA sequences. *Biol J Linn Soc* 92(1):151–161
194.Knutsen H, Jorde PE, Andre C, Stenseth NC (2003) Fine-scaled geographical population structuring in highly mobile marine species: the Atlantic cod. *Mol Ecol* 12(2):385–394
195.Kocher TD (2004) Adaptive evolution and explosive speciation: the Cichlid fish model. *Nat Rev Genet* 5(4):288-298
196.Koehn RK, Bayne BL, Moore MN, Siebenaller JF (1980) Salinity related physiological and genetic differences between populations of Mytilus edulis. *Biol J Linn Soc* 14(3-4):319-34
197.Koehn RK, Siebenaller JF (1981) Biochemical studies of aminopeptidase polymorphism in Mytilus edulis. II. Dependence of reaction rate on physical factors and enzyme concentration. *Biochem Genet* 19(11-12):1143-62
198.Kordos LM, Burton RS (1993) Genetic differentiation of Texas Gulf Coast populations of the blue crab *Callinectes sapidus*. *Mari Biol* 117(2):227-33
199.Kornfield I, Sidell BD, Gagnon PS (1982) Stock definition in the Atlantic herring (*Clupea harengus harengus*) – Genetic evidence for discrete fall and spring spawning populations. *Can J Fish Aquat Sci* 39(12):1610-1621
200.Korpelainen E, Tuimala J, Somervuo P, Huss M, Wong G (2014) RNA-seq data analysis: a practical approach. Chapman and Hall/CRC
201.Kreitman M (2000) Methods to detect selection in populations with applications to the human. *Annu Rev Genom Hum Genet* 1(1):539-559
202.Krishnakumar K, Raghavan R, Prasad G, Bijukumar A, Sekharan M, Pereira B, Ali A (2009) When pets become pests–exotic aquarium fishes and biological invasions in Kerala, India. *Curr Sci India* 97(4):474-476
203.Krishnakumar PK, Bhat GS (2008) Seasonal and interannual variations of oceanographic conditions off Mangalore coast (Karnataka, India) in the Malabar upwelling system during 1995–2004 and their influences on the pelagic fishery. *Fish Oceanogr* 17(1):45-60
204.Kritzer J P, Sale PF (2010) Marine Metapopulations. Elsevier
205.Kuthalingam MDK (1960) Observations on the life history and feeding habits of the Indian sardine, *Sardinella longiceps* (Cuv. & Val.). *Treubia* 25(2):207-213
206.Laakkonen HM, Lajus DL, Strelkov P, Vainola R (2013) Phylogeography of amphi-boreal fish: tracing the history of the Pacific herring *Clupea pallasii* in north-east European seas. *BMC Evol Biol* 13(1):67

207. Laikre L, Allendorf FW, Aroner LC, Baker CS, Gregovich DP, Hansen MM, Jackson JA, Kendall KC, Mckelvey KE, Neel MC, Olivieri I (2010) Neglect of genetic diversity in implementation of the convention on biological diversity. *Conserv Biol* 24(1):86–88
208. Larry M (1995) Ecology of marine invertebrate larvae. CRC Press
209. Larsen Pf, Nielsen Ee, Williams Td, Hemmer-Hansen Ja, Chipman Jk, Kruhoffer M, Gronkjaer P, George Sg, Dyrskjot L, Loeschcke V (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Mol Ecol* 16(22):4674–4683
210. Larsen PF, Nielsen EE, Williams TD, Loeschcke V (2008) Intraspecific variation in expression of candidate genes for osmo-regulation, heme-biosynthesis and stress resistance suggests local adaptation in European flounder (*Platichthys flesus*). *Heredity* 101(3):247–259
211. Larsson LC, Laikre L, Palm S, Andre C, Carvalho GR, Ryman N (2007) Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus. *Mol Ecol* 16(6):1135–1147
212. Lee AP, Fan S, Philippe H, MacCallum I, Braasch I, Manousaki T, Schneider I, Rohner N, Organ C, Chalopin D, Smith JJ (2013)The African coelacanth genome provides insights into tetrapod evolution. *Nature* 7445(496):311-316
213. Leinonen T, O'hara RB, Cano JM, Merila J (2008) Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *J Evol Biol* 21(1):1–17
214. Lesk A (2019) Introduction to bioinformatics. Oxford university press
215. Lessios HA, Kessing BD, Robertson DR (1998) Massive gene flow across the world's most potent marine biogeographic barrier. *P Roy Soc Lond B Bio* 265(1396):583-8
216. Lessios HA, Weinberg JR, Starczak VR (1994) Temporal variation in populations of the marine isopod Excirolana: how stable are gene frequencies and morphology? *Evolution* 48(3):549-63
217. Levasseur A, Orlando L, Bailly X, Milinkovitch MC, Danchin EGJ, Pontarotti P (2007) Conceptual bases for quantifying the role of the environment on gene evolution: the participation of positive selection and neutral evolution. *Biol Rev* 82(4):551–572
218. Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull Entomol Soc Am* 15(3):237-240
219. Levins R (1970) Extinction. In: Gesternhaber M (ed), Some Mathematical Problems in Biology. American Mathematical Society, Providence, Rhode Island, pp 77–10
220. Lewis RI, Thorpe JP (1994) Temporal stability of gene frequencies within genetically heterogeneous populations of the queen scallop *Aequipecten (Chlamys) opercularis*. *Mar Biol* 121(1):117-126
221. Lewontin RC (1991) 25 years ago in genetics – electrophoresis in the development of evolutionary genetics – milestone or millstone. *Genetics* 128(4):657–662
222. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature* 533(7602):200
223. Limborg MT, Pedersen JS, Hemmer-Hansen J, Tomkiewicz J, Bekkevold D (2009) Genetic population structure of European sprat *Sprattus sprattus*: differentiation across a steep environmental gradient in a small pelagic fish. *Mar Ecol Prog Ser* 379:213–224
224. Liu K, Xu D, Li J, Bian C, Duan J, Zhou Y, Zhang M, You X, You Y, Chen J, Yu H (2017) Whole genome sequencing of Chinese clearhead icefish, *Protosalanx hyalocranius*. *GigaScience* 6(4):giw012
225. Liu Z, Liu S, Yao J, Bao L, Zhang J, Li Y, Jiang C, Sun L, Wang R, Zhang Y, Zhou T (2016) The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat commun* 2(7):11757
226. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* 4(12):981–994
227. Mackay TFC (2001) The genetic architecture of quantitative traits. *Annu Rev Genet* 35(1):303–339
228. MacKenzie SA, Jentoft S (eds) (2016) Genomics in aquaculture. Academic Press
229. Magoulas A, Castilho R, Caetano S, Marcato S, Patarnello T (2006) Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Mol Phylogenet Evol* 39(3):734-46
230. Magoulas A, Tsimenides N, Zouros E (1996) Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy *(Engraulis encrasicolus*). *Mol Biol Evol* 13(1):178-90
231. Makinen HS, Cano JM, Merila J (2008a) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Mol Ecol* 17(15):3565–3582

232. Makinen HS, Shikano T, Cano JM, Merila J (2008b) Hitchhiking mapping reveals a candidate genomic region for natural selection in three-spined stickleback chromosome VIII. *Genetics* 178(1):453-65
233. Martin J (2019) Python for Biologists: A Complete Programming Course for Beginners. revision number 189, PT Serif and Source Code Pro https://pythonforbiologists.com/index.php/version/
234. Martinez-Takeshita N, Purcell CM, Chabot CL, Craig MT, Paterson CN, Hyde JR, Allen LG (2015) A tale of three tails: cryptic speciation in a globally distributed marine fish of the genus Seriola. *Copeia* 103(2):357–368
235. Maynard Smith J, Haigh J (1974) Hitch-hiking effect of a favorable gene. *Genet Res* 23(1):23–35
236. McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends Ecol Evol* 17(6):285–291
237. McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160(3):12311241
238. Menezes MR (1994) Little genetic variation in the oil sardine, *Sardinella longiceps* Val., from the western coast of India. *Mar Freshw Res* 45(2):257–264
239. Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE (2002). A high-density genetic map of *Sorghum bicolor (L.)* Moench based on 2926 AFLP, RFLP and SSR markers. *Plant Mol Biol* 48(5-6):483-499
240. Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152(3):1103-1110
241. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI (2003) Natural selection shaped regional mtDNA variation in humans. *P Natl Acad Sci USA* 100:171-176
242. Moberg PE, Burton RS (2000) Genetic heterogeneity among adult and recruit red sea urchins, *Strongylocentrotus franciscanus*. *Mar Biol* 136(5):773-84
243. Moen T, Hayes B, Nilsen F, Delghandi M, Fjalestad KT, Fevolden SE, Berg PR, Lien S (2008) Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC Genetics* 9(1):18
244. Mohandas NN, George MK (1997) Population genetic studies on the oil sardine (*Sardinella longiceps*) (Doctoral dissertation) Indian Council of Agricultural Research, Central Marine Fisheries Institute
245. Moller D (1966) Genetic differences between cod groups in the Lofoten area. *Nature* 212(5064):824
246. Morales H E, Pavlova A, Amos N, Major R, Bragg,J, Kilian A et al. (2016) Mitochondrial-nuclear interactions maintain a deep mitochondrial split in the face of nuclear gene flow. *bioRxiv* 095596
247. Morales H E, Pavlova A, Joseph L, Sunnucks P (2015) Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Mol Ecol* 24(11):2820–2837
248. Morgan SG, Christy JH (1995) Adaptive significance of the timing of larval release by crabs. *Am Nat* 145(3):457-79
249. Morgans CL, Cooke GM, Ord TJ (2014) How populations differentiate despite gene flow: sexual and natural selection drive phenotypic divergence within a land fish, the Pacific leaping blenny. *BMC Evol Biol* 14(1):97
250. Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends Ecol Evol* 19(4):208–216
251. Moritz C (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Syst Biol* 51(2):238-254
252. Murty AVS, Edelman MS (1970) On the relation between the intensity of the southwest monsoon and the oil sardine fishery of India. *Indian J Fish* 13: 142-149
253. Naciri M, Lemaire C, Borsa P, Bonhomme F (1999) Genetic study of the Atlantic/Mediterranean transition in sea bass (Dicentrarchus labrax). *J Hered* 90(6):591-6
254. Nair RV (1952) Studies on the revival of the Indian oil sardine fishery. *Proc Indo-Pacif Fish Coun* 2:1-15
255. Nair RV, Chidambaram K (1951) A review of the Indian Oil sardine fishery. *Proc Nat Inst Sci India* 17(1):71-85
256. Nair RV, Subrahmanyan R (1955) The diatom, Fragilaria oceanica Cleve, an indicator of abundance of the Indian oil sardine *Sardinella longiceps* Cuv. & Val. *Curr Sci* 24 (2):41-42
257. Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. *Annals of human genetics* 41(2):225-233
258. Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA*70(12):3321-3323

259. Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. *Ann Rev Ecol Syst* 28(1):105–128
260. Nesbo CL, Rueness EK, Iversen SA, Skagen DW, Jakobsen KS (2000) Phylogeography and population history of Atlantic mackerel (*Scomber scombrus* L.): a genealogical approach reveals genetic structuring among the eastern Atlantic stocks. *Proc R Soc Ser B-Bio* 267(1440):281-292
261. Ng TH, Tan HH (2010) The introduction, origin and life-history attributes of the non-native cichlid *Etroplus suratensis* in the coastal waters of Singapore. *J Fish Biol* 76(9):2238-2260
262. Nielsen EE, Gronkjaer P, Meldrup D, Paulsen H (2005) Retension of juveniles within a hybrid zone between North Sea and Baltic Sea Atlantic cod (*Gadus morhua*). *Can J Fish Aquat Sci* 62(10):2219–2225
263. Nielsen EE, Hansen MM, Ruzzante DE, Meldrup D, Gronkjaer P (2003) Evidence of a hybrid-zone in Atlantic cod (*Gadusmorhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis. *Mol Ecol* 12(6):1497–1508
264. Nielsen EE, Hemmer-hansen JA, Larsen PF, Bekkevold D (2009) population genomics of marine fishes: identifying adaptive variation in space and time. *Mol Ecol* 18(15):3128–3150
265. Nielsen EE, Kenchington E (2001) A new approach to prioritizing marine fish and shellfish populations for conservation. *Fish and Fish* 2(4):328–343
266. Nielson R, Slatkin M (2000) Likelihood analysis of ongoing gene flow and historical association. *Evolution* 54(1):44–50
267. Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nat Genet* 32:261–266
268. Padmakumar KG, Bindu L, Manu PS (2012) *Etroplus suratensis* (Bloch), the State Fish of Kerala. *J Biosci* 37(1):925–931
269. Palumbi SR (1997) Molecular biogeography of the Pacific. *Coral Reefs* 16(1):S47-52
270. Palumbi SR, Wilson AC (1990) Mitochondrial DNA diversity in the sea urchins *Strongylocentrotus purpuratus* and *S. droebachiensis*. *Evolution* 44(2):403-15
271. Pampoulie C, Gysels ES, Maes GE, Hellemans B, Leentjes V, Jones AG, Volckaert FA (2004) Evidence for fine-scale genetic structure and estuarine colonisation in a potential high gene flow marine goby (*Pomatoschistus minutus*). *Heredity* 92(5):434–445
272. Parish J (1981) Reproductive ecology of Naididae (Oligochaeta). *Hydrobiologia* 83(1):115-123
273. Parrish RH, Nelson CS, Bakun A (1981) Transport mechanisms and reproductive success of fishes in the California Current. *Biol Oceanogr* 1(2):175-203
274. Parrish RH, Serra R, Grant WS (1989) The monotypic sardines, Sardina and Sardinops: their taxonomy, distribution, stock structure, and zoogeography. *Can J Fish Aquat Sci* 46(11):2019–2036
275. Parsons YM, Shaw KL (2002) Mapping unexplored genomes: a genetic linkage map of the Hawaiian cricket Laupala. *Genetics* 162(3):1275-1282
276. Pauletto M, Manousaki T, Ferraresso S, Babbucci M, Tsakogiannis A, Louro B, Vitulo N, Quoc VH, Carraro R, Bertotto D, Franch R (2018) Genomic analysis of Sparus aurata reveals the evolutionary dynamics of sex-biased genes in a sequential hermaphrodite fish. *Commun Biol* 1(1):119
277. Pearse JS, McClintock JB, Bosch I (1991) Reproduction of Antarctic benthic marine invertebrates: tempos, modes, and timing. *Am Zool* 31(1):65-80
278. Perry AL, Low PJ, Ellis JR, Reynolds JD (2005) Climate change and distribution shifts in marine fishes. *Science* 308 (5730):1912-1915
279. Petersen MF, Steffensen JF (2003) Preferred temperature of juvenile Atlantic cod *Gadus morhua* with differenthaemoglobin types at normoxia and moderate hypoxia. *J Exp Biol* 206(2):359–364
280. Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. *Genome Res* 9(2):167-174
281. Piekarowicz A (1979) Werner Arber, Daniel Nathans and Hamilton Smith. Nobel prizes for the studies on DNA restriction enzymes. *Postepy biochemii*, 25(2):251-253
282. Pikitch EK, Rountos KJ, Essington TE, Santora C, Pauly D, Watson R, Sumaila UR, Boersma PD, Boyd IL, Conover DO, Cury P (2014) The global contribution of forage fish to marine fisheries and ecosystems. *Fish Fish* 15(1):43–64
283. Pimm S, Raven P, Peterson A, Cagan HS, Sekercioglu, Ehrlich PR (2006) Human impacts on the rates of recent, present, and future bird extinctions. *P Natl Acad Sci USA* 103(29):10941–10946
284. Pimm SL, Russell GJ, Gittelman JL, Brooks TM (1995) The future of biodiversity. *Science* 269:347-350
285. Pogson GH, Mesa KA, Boutilier RG (1995) Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics* 139(1):375–385

286.Pogson GH, Taggart CT, Mesa KA, Boutilier RG (2001) Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution* 55(1):131-46

287.Potvin C, Tousignant D (1996) Evolutionary consequences of simulated global change: genetic adaptation or adaptive phenotypic plasticity? *Oecologia* 108(4):683–693

288.Powles H, Bradford MJ, Bradford RG, Doubleday WG, Innes S, Levings CD. (2000) Assessing and protecting endangered marine species. *ICES J Mar Sci* 57(3):669–676

289.Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol Evol* 4(22):4399-4428

290.Raeymaekers JAM, Joost AM, Van Houdt JKJ, Larmuseau MHD, Geldof S, Volckaert FAM (2007) Divergent selection as revealed by P-ST and QTL-based F-ST in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. *Mol Ecol* 16(4):891–905

291.Reeb CA, Avise JC (1990) A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics* 124(2):397-406

292.Reichwald K, Petzold A, Koch P, Downie BR, Hartmann N, Pietsch S, Baumgart M, Chalopin D, Felder M, Bens M, Sahm A (2015) Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell* 163(6):1527-38

293.Reilly PTO, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between $F_{ST}$ and microsatellite polymorphism in the marine fish, walleye Pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Mol Ecol* 13(7):1799–1814

294.Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ (2009) Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish Fish* 10(4):361–395

295.Remington DL, Whetten RW, Liu BH, O'malley DM (1999) Construction ofan AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. *Theor Appl Genet* 98(8):1279-1292

296.Reusch TBH, Wood TE (2007) Molecular ecology of global change. *Mol Ecol* 19, 3973–3992

297.Reynolds JD, Dulvy NK, Goodwin NB, Hutchings JA (2005) Biology of extinction risk in marine fishes. *P R Soc London B* 272(1579):2337-2344

298.Riebler A, Held L, Stephan W (2008) Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics* 178(3):1817–1829

299.Riginos C, Nachman MW (2001) Population subdivision in marine environments: the contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, *Axoclinus nigricaudus*. *Mol Ecol* 10(6):1439-53

300.Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genomic scans towards the characterization of candidate loci under parallel directional selection in the lake whitefish (*Coregonus clupeaformis*). *Mol Ecol* 14(2):351–361

301.Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR, Lemon C, Bird NH, Koop BF (2014) The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *Plos One* 9(7):e102089

302.Ropson IJ, Brown EC, Powers DA (1990) Biochemical genetics of *Fundulus heteroclitus* (L.). VI. Geographical variation in the gene frequencies of 15 loci. *Evolution* 44(1):16-26

303.Rosenblum EB, Hickerson MJ, Moritz C (2007) Amultilocus perspective on colonization accompanied by selection and gene flow. *Evolution* 61(12):2971–2985

304.Ruggeri P, Splendiani A, Bonanomi S, Arneri E, Cingolani N, Santojanni A, Colella S, Donato F, Giovannotti M, Barucchi VC (2013) Searching for a stock structure in *Sardina pilchardus* from the Adriatic and Ionian seas using a microsatellite DNA-based approach. *Sci Mar* 77(4):1–10

305.Ruzzante DE, Mariani S, Bekkevold D, Andre C, Mosegaard H, Clausen LAW, Dahlgren TG, Hutchinson WF, Hatfield EMC, Torstensen E, Brigham J, Simmonds EJ, Laikre L, Larsson LC, Stet RJM, Ryman N, Carvalho GR. (2006) Biocomplexity in a highly migratory pelagic marine fish, *Atlantic herring*. *P R Soc London B* 273(1593):1459-1464

306.Ruzzante DE, Taggart CT, Cook Hilborn R, Quinn TP, Schindler DE, Rogers DE (2003) Biocomplexity and fisheries sustainability. *P Natl Acad Sci USA* 100(11):6564–6568

307.Ryman N, Palm S, Andre C, Carvalho GR, Dahlgren TG, Jorde PE, Laikre L, Larsson LC, Palme A, Ruzzante DE (2006) Power for detecting genetic divergence: differences between statistical methods and marker loci. *Mol Ecol* 15(8):2031-2045

308.Ryman N, Utter F, Laikre L (1995) Protection of intraspecific biodiversity of exploited fishes. *Rev Fish Biol Fisher* 5(4):417-446

309.Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of (3-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230(4732):1350-1354

310.Saroglia M, Zhanjiang (John) L (eds) (2012) Functional genomics in aquaculture. John Wiley & Sons

311. Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff JN, Lesch KP, Bisazza A (2013) The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat Genet* 45(5):567
312. Schierwater B, Ender A (1993) Different thermostable DNA polymerases may amplify different RAPD products. *Nucleic Acids Res* 21(19):4647-4648
313. Schlotterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109(6):365-371
314. Schlotterer C (2004) The evolution of molecular markers- just a matter of fashion. *Nat Rev Genet* 5(1):63–69
315. Schlotterer C, Dieringer D (2005) A novel test statistics for the identification of local selective sweeps based on microsatellite gene diversity. In: Nurminski D(eds) Selective Sweep. Eurekah.com and Kluwer Academic/Plenum Publishers, Georgetown, TX, USA. pp 55–64
316. Schlotterer C, Harr B (2002) Single nucleotide polymorphisms derived from ancestral populations show no evidence for biased diversity estimates in *Drosophila melanogaster*. *Mol Ecol* 11(5):947-950
317. Schmidt PS, Rand DM (1999) Intertidal microhabitat and selection at MPI: Interlocuscontrasts in the Northern Acorn Barnacle, *Semibalanus balanoides*. *Evolution* 53(1):135-146
318. Schulte PM (2001) Environmental adaptations as windows on molecular evolution. *Comp Biochem Phys B* 128(3):597–611
319. Schulte PM (2007) Responses to environmental stressors in an estuarine fish: interacting stressors and the impacts of local adaptation. *J Therm Biol* 32(3):152–161
320. Scott GR, Schulte PM, Egginton S, Scott AL, Richards JG, Milsom WK (2010) Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol Biol Evol* 28(1):351–363
321. Sebastian W, Sukumaran S, Zacharia PU, Gopalakrishnan A (2017) Genetic population structure of Indian oil sardine, *Sardinella longiceps* assessed using microsatellite markers. *Conserv Genet* 18(4):951-964
322. Seehausen O (2006) African Cichlid fish: a model system in adaptive radiation research. *P Roy Soc Lond B Bio* 273(1597):1987-1998
323. Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish Fish* 9(4):363–377
324. Sick K (1965a) Haemoglobin polymorphism of cod in Baltic and Danish Belt Sea. *Hereditas* 54(1):19-48
325. Sick K. (1965b) Haemoglobin polymorphism of cod in North Sea and North Atlantic Ocean. *Hereditas* 54(1):49-69
326. Skarstein TH, Westgaard JI, Fevolden SE (2007) Comparing microsatellite variation in north-east Atlantic cod (*Gadus morhua L.*) to genetic structuring as revealed by the pantophysin (Pan I) locus. *J Fish Biol* 70 (Suppl. C):271–290
327. Skibinski DOF (2000) DNA tests of neutral theory: applications in marine genetics. *Hydrobiologia* 420(1):137–152
328. Slatkin (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* 47(1):264–279
329. Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution* 39(1):53-65
330. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139(1):457–462
331. Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43(7):1349–1368
332. Smedbol RK, Stephenson R (2001) The importance of managing within-species diversity in cod and herring fisheries of the north-western Atlantic. *J Fish Biol* 59(Supplement A):109-128
333. Smith AD, Brown CJ, Bulman CM, Fulton EA, Johnson P, Kaplan IC, Lozano-Montes H, Mackinson S, Marzloff M, Shannon LJ, Shin YJ (2011) Impacts of fishing low-trophic level species on marine ecosystems. *Science* 333(6046):1147–1150
334. Smith MW, O'Brien SJ (2005) Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat Rev Genet* 6(8): 623–632
335. Somero GN (2005) Linking biogeography to physiology: Evolutionary and acclamatory adjustments of thermal limits. *Front Zool* 2(1):1
336. Somorjai IML, Danzmann RG, Ferguson MM (2003) Distribution of temperature tolerance quantitative trait loci in Arctic charr (*Salvelinus alpinus*) and inferred homologies in rainbow trout (*Oncorhynchus mykiss*). *Genetics* 165(3):1443– 1456

337. Sotka EE, Wares JP, Barth JA, Grosberg RK, Palumbi SR (2004) Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Mol Ecol* 13(8):2143-56
338. Stem C (1943) The Hardy-Weinberg law. *Science* 97(2510):137-138
339. Stier A, Massemin S, Criscuolo F (2014) Chronic mitochondrial uncoupling treatment prevents acute cold-induced oxidative stress in birds. *J Comp Physiol B* 184(8):1021–1029
340. Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100(2):158–170
341. Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 14(3):671–688
342. Sukumaran S, Sebastian W, Gopalakrishnan A (2016) Population genetic structure of Indian oil sardine, *Sardinella longiceps* along Indian coast. *Gene* 576(1):372-378
343. Takeda M, Kusumi J, Mizoiri S, Aibara M, Mzighani SI, Sato T, Terai Y, Okada N, Tachida H (2013) Genetic structure of pelagic and littoral Cichlid fishes from Lake Victoria. *Plos One* 8:e74088
344. Talwar PK, Kacker RK (1984) Commercial Sea fishes of India. Zoological Survey of India, Calcutta, pp 997
345. Tang CQ, Humphreys AM, Fontaneto D, Barraclough TG (2014) Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. Methods *Ecol Evol* 5(10):1086–1094
346. Taulz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17(16):6463-6471
347. Teacher AG, Andre C, Merila J, Wheat CW (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evol Biol* 12(1):248
348. Teske PR, Sandoval-Castillo J, Golla TR, Emami-Khoyi A, Tine M, von der Heyden S, Beheregaray LB (2019) Thermal selection as a driver of marine ecological speciation. *Proc R Soc B* 286(1896):20182023
349. Thomas JA, Telfer MG, Roy DB, Preston CD, Greenwood JJD, Asher J, Fox R, Clarke RT, Lawton JH (2004) Comparative losses of British butterflies, birds, and plants and the global extinction crisis. *Science* 303(5665):1879-1881
350. Thomas Jr RC, Willette DA, Carpenter KE, Santos MD (2014) Hidden diversity in sardines: genetic and morphological evidence for cryptic species in the goldstripe sardinella, *Sardinella gibbosa* (Bleeker, 1849). *Plos One* 9(1):e84719
351. Umina PA, Weeks AR, Kearney MR, McKechnie SW, Hoffmann AA (2005) A rapid shift in a classic clinal pattern in Drosophilareflecting climate change. *Science* 308(5722):691–693
352. Valenzano DR, Benayoun BA, Singh PP, Zhang E, Etter PD, Hu CK, Clement-Ziza M, Willemsen D, Cui R, Harel I, Machado BE (2015) The African turquoise killifish genome provides insights into evolution and genetic architecture of lifespan. *Cell* 163(6):1539-54
353. van Straalen NM, Timmermans M (2002) Genetic variation in toxicant-stressed populations: an evaluation of the ‘genetic erosion’ hypothesis. *Hum Ecol Risk Assess* 8(5):983–1002
354. van Tienderen PH, de Haan AA, van der Linden CG, Vosman B (2002) Biodiversity assessment using markers for ecologically important traits. *Trends Ecol Evol* 17(12):577–582
355. Vasemagi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar L.*). *Mol Biol Evol* 22(4):1067–1076
356. Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol Ecol* 14(12):3623–3642
357. Venkita Krishnan P (1993) Biochemical genetic studies on the oil sardine, *Sardinella longiceps* (cuvier and valenciennes, 1847) from selected centres of the west coast of India (Doctoral dissertation) Indian Council of Agricultural Research, Central Marine Fisheries Institute
358. Vinas J, Bremer JA, Pla C (2004) Inter-oceanic genetic differentiation among albacore (*Thunnus alalunga*) populations. *Mar Biol* 145(2):225–232
359. Vos P, Hogers R, Bleeker M, Reijans M, Lee TV, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23(21):4407-4414
360. Vrijenhoek RC (1997) Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. *J Hered* 88(4):285-93
361. Wakeley J (1996) Distinguishing migration from isolation using the variance of pairwise differences. *Theor Pop Biol* 49(3):369–386
362. Wang CM, Lo LC, Zhu ZY, Yue GH (2006) A genome scan for quantitative trait loci affecting growth-related traits in an F1 family of Asian seabass (*Lates calcarifer*). *BMC Genomics* 7(1):274

363. Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol Ecol* 15(6):1419-1439
364. Wares JP, Cunningham CW (2001) Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* 55(12):2455-2469
365. Watts RJ, Johnson MS (2004) Estuaries, lagoons and enclosed embayments: habitats that enhance population subdivision of inshore fishes. *Mar Freshwater Res* 55(7):641-651
366. Watts RJ, Johnson MS, Black R (1990) Effects of recruitment on genetic patchiness in the urchin*Echinometra mathaei* in Western Australia. *Mar Biol* 105(1):145-151
367. Weber JL, Myers EW (1997) Human whole-genome shotgun sequencing. *Genome Res* 7(5):401-409
368. Wenne R, Boudry P, Hemmer-Hansen J, Lubieniecki KP, Was A, Kause A (2007) What role for genomics in fisheries management and aquaculture? *Aquatic Living Resour* 20(3):241–255
369. Westgaard JI, Fevolden SE (2007) Atlantic cod (*Gadus morhua L.*) in inner and outer coastal zones of northern Norway display divergent genetic signature at non-neutral loci. *Fish Res* 85(3):306–315
370. Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression. *P Natl Acad Sci USA* 103(14):5425–5430
371. Whitehead PJP (1985) FAO species catalogue. Clupeoid fishes of the world (Sub order: Clupeioidei). An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf-herrings. Part 1 – Chirocentridae, Clupeidae and Pristigasteridae, Vol. 7. *FAO Fish Synopsis* 125:1–303
372. WahlundS (1928) Zusammensetzung von Populationen und Korrelationer-scheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* 11:65–106
373. Williams GC (1966) Natural selection, the costs of reproduction, and a refinement of Lack's principle. *The American Naturalist* 100(916):687-690
374. Williams JGK, Kubelik AR, LiVak KJ, Rafaiski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18(22): 6531-6235
375. Williams LM, Oleksiak MF (2008) Signatures of selection in natural populations adapted to chronic pollution. *BMC Evol Biol* 8(1):282
376. Williams ST, Benzie JA (1998) Evidence of a biogeographic break between populations of a high dispersal starfish: congruent regions within the Indo-West Pacific defined by color morphs, mtDNA, and allozyme data. *Evolution* 52(1):87-99
377. Winans GA (1980) Geographic variation in the milkfish *Chanos chanos*. I. Biochemical evidence. *Evolution* 34(3):558-574
378. Wittkopp PJ (2007) Variable gene expression in eukaryotes: a network perspective. *J Exp Biol* 210(9):1567–1575
379. Wright S (1943) Isolation by distance. *Genetics* 28(2):114-138
380. Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19(3):395-420
381. Wright S (1951) The genetical structure of populations. *Ann Eugen (Lond)* 1:323-334
382. Wright S (1921) Systems of mating. *Genetics*6:111-178
383. Wu C, Zhang D, Kan M, Lv Z, Zhu A, Su Y, Zhou D, Zhang J, Zhang Z, Xu M, Jiang L (2014) The draft genome of the large yellow croaker reveals well-developed innate immunity. *Nat Commun* (19)5:5227
384. Xu J, Bian C, Chen K, Liu G, Jiang Y, Luo Q, You X, Peng W, Li J, Huang Y, Yi Y (2017) Draft genome of the Northern snakehead, *Channa argus*. *Gigascience* 6(4):gix011
385. Shin YJ, Roy C, Cury P (1998). Clupeoids reproductive strategies in upwelling areas: a tentative generalization. In : Durand MH, Cury P, Mendelssohn R, Roy C, Bakun A, Pauly D(eds) Global versus local changes in upwelling systems. Paris: ORSTOM, (Colloques et Seminaires). Global Versus Local Changes in Upwelling Systems: Conference, Monterey (USA), pp 409-422
386. Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR) anchored polymerase chain reaction amplification. *Genomics* 20(2):176-183

# Chapter 2

## THE COMPLETE MITOCHONDRIAL GENOME AND PHYLOGENY OF INDIAN OIL SARDINE, *SARDINELLA LONGICEPS* (Valenciennes, 1847) AND GOLDSTRIPE SARDINELLA, *SARDINELLA GIBBOSA* (Bleeker, 1849) FROM THE INDIAN OCEAN

### ABSTRACT

The Indian Oil Sardine, *Sardinella longiceps* (Valenciennes, 1847) and Goldstripe Sardinella, *Sardinella gibbosa* (Bleeker, 1849) are the two commercially important, small pelagic fishes from Indian waters belonging to the family Clupeidae. Accurate identification and characterization of intraspecific diversity of clupeids are very challenging due to cryptic speciation. Characterization of the complete mitogenome is very helpful in resolving taxonomic ambiguities and hence we characterized the complete mitogenome of *S. longiceps* and *S. gibbosa* from Indian waters. The entire mitogenome was amplified by polymerase chain reactions (PCR) using primers that amplify overlapping segments of the entire genome, and the products were subsequently used for direct sequencing. The assembled mitogenomes of *S. longiceps* and *S. gibbosa* are 16,613 and 16658 bp circles respectively, contained the 37 mitochondrial structural genes (two ribosomal RNA, 22 transfer RNA, and 13 protein-coding genes) with the gene order identical to that of typical vertebrates. An anti-G bias in the third codon positions and proteins enriched with amino acids encoded by CA-rich codons was observed in both genomes. The major non-coding region between the tRNA Pro and tRNA Phe genes considered as the control (D-loop) region has several characteristic conserved sequence blocks (CSB). In the phylogenetic tree, *S. longiceps* and *S. gibbosa* clustered together with species belonging to the family Clupeidae. Clupeidae and its five subfamilies are not monophyletic. Only three of the nine currently recognised family, Engraulidae, Pristigasteridae and Dussumieriidae formed well-supported monophyletic groups, and the relationships among other groups are not well supported. This study is the first report of the complete mitogenome of two commercially important clupeids from Indian waters which form the baseline for further studies on molecular systematics, population genetics, biogeography, historical demography, adaptive variation and conservation of these species.

## 1. Introduction

The Indian oil sardine, *Sardinella longiceps* (Valenciennes, 1847) and goldstripe sardinella, *Sardinella gibbosa* (Bleeker, 1849) are the two commercially important species of clupeids available in Indian waters. Indian oil sardine, *S. longiceps* is the most abundant, commercially important species distributed all along the Indian coast with major contributions from southwest and southeast coasts of India. Goldstripe sardinella, *S. gibbosa* contributes to the major share of lesser sardine fishery of the Indian coast with maximum contributions from the southeast coast followed by south-west coast of India (CMFRI 2017). *S. longiceps* is predominantly a phytoplankton feeder whilst *S. gibbosa* is a zooplankton feeder (Devaraj *et al*. 1997). Both the species are pelagic with a depth of occurrence between 10-100m. *S. longiceps* is distributed along Northern and Western Indian Ocean, the Gulf of Oman, and Gulf of Aden whereas *S. gibbosa* is distributed along the Indo-west Pacific and Red Sea (Whitehead *et al.* 1988). The diversity of clupeids is the highest in the Indo-west Pacific region with reports of cryptic speciation and morphological plasticity contributing to taxonomic ambiguity in these groups (Lavoue *et al.* 2007; Lavoue *et al.* 2013; Thomas *et al.* 2014; Stern *et al.* 2016; Sukumaran *et al.* 2016a). These fishes show exemplary bio-complexity and inter- and intra-specific diversity which is very important in providing resilience to environmental fluctuations (Sukumaran *et al.* 2016a; Sukumaran *et al.* 2016b; Sukumaran *et al.* 2017). So accurate identification is the key to characterizing and documenting their diversity for which immediate steps are necessary. Characterizing the complete mitogenome of fishes will act as baseline information for further taxonomic studies which is very important in conservation and further evolutionary studies of these species.

Animal mitochondrial DNA (mtDNA) is a circular molecule, typically 16-20 kb in length, with 37 mitochondrial structural genes encoding two ribosomal RNA (rRNA), 22 transfer RNAs (tRNA) and 13 proteins along with a non-coding control region that regulates replication and transcription (Boore 1999). mtDNA has emerged as a very useful marker for understanding evolutionary relationships, gene flow, hybridisation, introgression and historical demography mainly because of its maternal inheritance, fast evolutionary rate compared to nuclear DNA, lack of recombination and presence of multiple copies in the cell (Meyer 1993; Ballard and Whitlock 2004; Karl *et al.* 2012). mtDNA has been used as a marker to infer genetic population structure of many fishery resources (Curole and

Kocher 1999; Cadrin *et al.* 2013; Miya and Nishida 2015). But often, these inferences were based on a short segment of the mtDNA like D-loop, cytochrome b region or ND2 genes and conclusions on the genetic stock structure using these genes with different evolutionary rates may not reflect the true picture. Complete mitogenomes provide a holistic perspective for comparisons making inferences regarding population structure accurate and effective (Curole and Kocher 1999; Miya and Nishida 2015). The advent of improved techniques like long PCR and next-generation sequencing has made characterisation of complete mitogenomes quicker and easier (Miya and Nishida 1999; Sorenson *et al.* 1999; Morin *et al.* 2010; Jacobsen *et al.* 2012; Miya and Nishida 2015) with more than two thousand mitogenomes available in public databases (http://www.ncbi.nlm.nih.gov/). Recent findings based on mitogenomic data have revolutionized several concepts of molecular phylogeny and evolution across multiple taxonomic levels (Miya and Nishida 2015; Curole and Kocher 1999). Whole mitogenome information has also been recently used to study selection and adaptation in fishes and other organisms in response to environmental and climatic fluctuations (da Fonseca *et al.* 2008; Silva *et al.* 2014; Stager *et al.* 2014; Caballero *et al.* 2015).

To date, several whole mtDNA have been used as molecular markers in the establishment of phylogenetic relationships among clupeidae (Lavoue *et al.* 2007; Lavoue *et al.* 2013). But complete mitogenome studies of fishes from Indian waters have been fragmentary with few freshwater species being characterised. So, this is the first attempt to characterise the complete mitogenomes of marine fishes from Indian waters. Genetic population structure and historical demography of Indian oil sardine and Indian mackerel have been studied recently by the present authors by collecting samples from all over the Indian coast (Sukumaran *et al.* 2016a; Sukumaran *et al.* 2016b; Sukumaran *et al.* 2017). Few studies were also reported in *S. gibbosa* (Thomas *et al.* 2014; Stern *et al.* 2016), but none of the studies has been focussed on resolving taxonomic ambiguities and diversity patterns by complete mitogenome characterisation. Recent investigations have been focussed on selection and adaptation in the mitochondrial oxidative phosphorylation machinery which provided clues to thermal and metabolic adaptations in many fishes (Bradbury *et al.* 2010; Foote *et al.* 2011; Garvin *et al.* 2012; Teacher *et al.* 2012; Caballero *et al.* 2015). Sardines of the Indian Ocean are also important from this viewpoint, as they are widely distributed across environmental clines and are prone to forces of positive and purifying selection. Hence, we investigated the complete mitochondrial genome organization of *S. longiceps* and *S. gibbosa* for the first time followed by phylogenetic resolution of the evolutionary relationships. The present research will provide baseline information for further studies on the taxonomic resolution, conservation, adaptive variation to environmental clines and evolution regarding these commercially and ecologically important species.

## 2. Materials And Methods

### 2.1. Sample collection and preparation

*Sardinella longiceps* was collected from Kochi and *S. gibbosa* from Tuticorin. Skeletal muscle samples were obtained from the tail of each individual and stored in 95% ethanol for DNA extraction. Genomic DNA was isolated by standard phenol/chloroform method after proteinase K digestion (Sambrook and Russell 2001).

### 2.2. PCR Amplification and sequencing Mitochondrial DNA

The entire mitogenome of each species was amplified by polymerase chain reaction (PCR) as contiguous, overlapping segments with novel primer pairs (Table 2.S1.). Primers were designed based on the conserved regions of the mitochondrial genomes of *Sardinella maderensis* (GenBank accession number AP009143), *Sardinella albella* (Gen Bank accession number AP011605) and *Sardinops melanostictus* (Gen Bank accession number AB032554). PCR amplifications were carried out in 25 µl reaction mixture containing 25 ml 10x buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 15 mM $MgCl_2$), 200 µM of each dNTP, 02 µM of each primer, 1 unit of Taq DNA polymerase (Sigma Aldrich), and 50 ng of template DNA. The PCR reaction was carried out in a Biorad T100 thermocycler (Biorad, USA) programmed for an initial denaturation at 94 °C for 4 min followed by 33 cycles of denaturation at 94 °C for 30 sec, annealing at 48 °C - 55 °C for 30 sec, extension at 72 °C for 60 sec and a final extension at 72 °C for 7 min. Purification of the PCR product was carried out using Qiagen PCR purification kit (Qiagen) and sequenced with both primers using the BigDye Terminator Sequencing Ready Reaction v30 kit (Applied Biosystems) following instructions of the manufacturer. Sequencing was carried out on an ABI 3730 automated sequencer (Life Technologies).

### 2.3. Assembly and annotation of the mitochondrial genome

The sequence fragments were assembled into a complete mitochondrial genome using MEGA 6 (Tamura *et al.* 2013) and Geneious R7 (Kearse *et al.* 2012). Annotation and boundary determination of protein-coding genes, rRNA and tRNA were performed using NCBI-BLAST and MitoAnnotator (Iwasaki *et al.* 2013) programs. Nucleotide composition of mitogenome and protein-coding genes were determined using Geneious R7. Codon usage and RSCU values were calculated with MEGA 60. Alignments with previously published closely related bony fishes were carried out to identify the origin of replication and conserved blocks in the non-coding control region. The mtDNA sequences were deposited in NCBI GenBank.

### 2.4. Phylogeny construction

The phylogenetic tree was reconstructed using mitogenome sequences retrieved from NCBI GenBank, aimed to study the relationship of *S. longiceps* and *S. gibbosa* with other clupeids as well as to validate its taxonomic position. The sequences included in the present analysis belonged to the family Denticipitidae, Clupeidae, Engraulidae, Chirocentridae, and Pristigasteridae (Table 2.S2.). The 12 concatenated protein-coding genes were aligned using Geneious R7 (GTR+G+I model was selected as the best model for phylogeny construction) and a maximum likelihood phylogeny was constructed based on 1000 replicates.

## 3. Results And Discussion

### 3.1. Mitogenome organisation

The assembled mitogenome is a 16,613 bp circle for *S. longiceps* and 16658 bp circle for *S. gibbosa* (Fig 2.1.). Both of it contained the 37 mitochondrial structural genes; two ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA (tRNA) genes, and 13 protein-coding genes 1 non-coding control region (D-loop) (Table 2.1.) and with the gene order identical to that in other vertebrates (Boore 1999). The Heavy (H) and Light (L) strand coding pattern previously reported for most vertebrates were also observed in the *S. longiceps* and *S. gibbosa* mitogenome. Except the ND6 and eight tRNA genes (tRNA$^{\text{Gln(TTG)}}$, tRNA$^{\text{Ala(TGC)}}$, tRNAA$^{\text{sn(GTT)}}$, tRNA$^{\text{Cys(GCA)}}$, tRNA$^{\text{Tyr(GTA)}}$, tRNA$^{\text{Ser(TGA)}}$, tRNA$^{\text{Glu(TTC)}}$, and tRNA$^{\text{Pro(TGG)}}$), all other genes were encoded on the H-strand and all genes were similar in length as in other bony fishes (Boore 1999). The overall base composition of the H-strand was as follows: A (26.9%/26.1%), T (25.5%/24.8), C (28.7%/29.5%), G (18.9%/19.7) and G+C (47.6%/49.1%) (SL/SG) (Table 2.2.). The overall sequence similarity is 84% between *S. longiceps* and *S. gibbosa* Similar to other vertebrate low G content and high A+T (52.4/50.9) content were observed in both genome (Broughton *et al*. 2001; Fischer *et al*. 2013). Mitogenome of *S. longiceps* and *S. gibbosa* sequences were deposited in NCBI Gen Bank under accession number KR000002.1 and KU665488.1 respectively.

**Table 2.1** Location and arrangement of genes on the mitogenomes of *S. longiceps* and *S. gibbosa*.

| Gene | *S. longiceps* | | | | | | *S. gibbosa* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Position | | Size (bp) | Strand[a] | Codon[b] | | Position | | Size (bp) | Strand[a] | Codon[b] | |
| | From(bp) | To(bp) | | | Start | Stop | From(bp) | To(bp) | | | Start | Stop |
| tRNA-Phe | 1 | 63 | 63 | H | | | 1 | 68 | 68 | H | | |
| 12S rRNA | 64 | 1017 | 954 | H | | | 69 | 1019 | 951 | H | | |
| tRNA-Val | 1018 | 1089 | 72 | H | | | 1020 | 1091 | 72 | H | | |
| 16S rRNA | 1090 | 2779 | 1690 | H | | | 1092 | 2777 | 1686 | H | | |
| tRNA-Leu | 2780 | 2854 | 75 | H | | | 2778 | 2853 | 76 | H | | |
| ND1 | 2855 | 3829 | 975 | H | ATG | TAA | 2854 | 3828 | 975 | H | ATG | TAA |
| tRNA-Ile | 3837 | 3908 | 72 | H | | | 3837 | 3908 | 72 | H | | |
| tRNA-Gln | 3908 | 3978 | 71 | L | | | 3908 | 3978 | 71 | L | | |
| tRNA-Met | 3978 | 4046 | 69 | H | | | 3978 | 4046 | 69 | H | | |
| ND2 | 4047 | 5093 | 1047 | H | ATG | TAA | 4047 | 5093 | 1047 | H | ATG | TAG |
| tRNA-Trp | 5094 | 5163 | 70 | H | | | 5094 | 5163 | 70 | H | | |
| tRNA-Ala | 5165 | 5233 | 69 | L | | | 5165 | 5233 | 69 | L | | |
| tRNA-Asn | 5235 | 5307 | 73 | L | | | 5235 | 5308 | 74 | L | | |
| tRNA-Cys | 5340 | 5405 | 66 | L | | | 5346 | 5411 | 66 | L | | |
| tRNA-Tyr | 5407 | 5477 | 71 | L | | | 5415 | 5485 | 71 | L | | |
| CO1 | 5479 | 7029 | 1551 | H | GTG | TAA | 5487 | 7037 | 1551 | H | GTG | TAA |
| tRNA-Ser | 7030 | 7097 | 68 | L | | | 7038 | 7108 | 71 | L | | |
| tRNA-Asp | 7102 | 7170 | 69 | H | | | 7113 | 7181 | 69 | H | | |
| CO2 | 7184 | 7874 | 691 | H | ATG | T-- | 7194 | 7884 | 691 | H | ATG | T-- |
| tRNA-Lys | 7875 | 7948 | 74 | H | | | 7885 | 7958 | 74 | H | | |
| ATPase 8 | 7950 | 8117 | 168 | H | ATG | TAA | 7960 | 8127 | 168 | H | ATG | TAA |
| ATPase 6 | 8108 | 8790 | 683 | H | ATG | TA- | 8118 | 8800 | 683 | H | ATG | TA- |
| CO3 | 8791 | 9576 | 786 | H | ATG | TAA | 8801 | 9586 | 786 | H | ATG | TAA |
| tRNA-Gly | 9577 | 9647 | 71 | H | | | 9586 | 9657 | 72 | H | | |
| ND3 | 9648 | 9996 | 349 | H | ATG | T-- | 9658 | 10006 | 349 | H | ATG | T-- |
| tRNA-Arg | 9997 | 10065 | 69 | H | | | 10007 | 10076 | 70 | H | | |
| ND4L | 10066 | 10362 | 297 | H | ATG | TAA | 10077 | 10373 | 297 | H | ATG | TAA |
| ND4 | 10356 | 11736 | 1381 | H | ATG | T-- | 10367 | 11747 | 1381 | H | ATG | T-- |
| tRNA-His | 11737 | 11805 | 69 | H | | | 11748 | 11816 | 69 | H | | |
| tRNA-Ser | 11806 | 11872 | 67 | H | | | 11817 | 11883 | 67 | H | | |
| tRNA-Leu | 11873 | 11944 | 72 | H | | | 11884 | 11955 | 72 | H | | |
| ND5 | 11945 | 13780 | 1836 | H | ATG | TAA | 11956 | 13791 | 1836 | H | ATG | TAG |
| ND6 | 13777 | 14298 | 522 | L | ATG | TAA | 13788 | 14309 | 522 | L | ATG | TAA |
| tRNA-Glu | 14299 | 14367 | 69 | L | | | 14310 | 14378 | 69 | L | | |
| Cyt b | 14374 | 15514 | 1141 | H | ATG | T-- | 14385 | 15525 | 1141 | H | ATG | T-- |
| tRNA-Thr | 15515 | 15586 | 72 | H | | | 15526 | 15597 | 72 | H | | |
| tRNA-Pro | 15586 | 15655 | 70 | L | | | 15597 | 15666 | 70 | L | | |
| D-loop) | 15656 | 16613 | 958 | | | | 15667 | 16658 | 991 | | | |

a H and L, respectively, denote heavy and light strands.

b Codons containing "-"symbols indicate an incomplete stop codon.

3.2. Protein coding gene

In both *S. longiceps and S. gibbosa*, 13 protein-coding genes were of the same size and orientation (Fig 2.1). They are 11427bp in total length and thus represented ~ 69% of the genome. The similarity of the coding sequence is 81% at the nucleotide level and 85% at the amino acid level. All the genes are encoded by heavy strand except ND6 gene which is encoded by light strand. In both the species, ATP6 & ATP8 shared 10 nucleotides, ND4 & ND4L shared 7 and ND5 & ND6 shared 4 nucleotides. ATG is used as start codon by all coding genes except CO1 (GTG is the start codon). Intergenic overlaps of protein-coding regions are common within vertebrate mitogenomes and have been reported for several fish species (Boore 1999; Morin *et al.* 2010; Mu *et al.* 2015). In *S. longiceps*, stop codon TAA was used as translation terminators for ND1, ND2, CO1, ATP8, CO3, ND4L, ND5 and ND6. The remaining genes used incomplete stop codon TA- (ATP6) and T-- (CO2, ND3, ND4 and CYTB). Similarly, in *S. gibbosa*, TAA appears in ND1, CO1, CO3, ND4L, ND6 TAG in ND2, incomplete stop codons TA- in ATP 6, and T-- in CO2, ND3, ND4 and CYTB (Table 2.1). Reading frame overlap and incomplete stop codons are common in mitochondria and post-transcriptional polyadenylation provides the two adenosine nucleotide required for generating the TAA stop codon (Ojala *et al.* 1981). The coding sequences of *S. longiceps* consisted of 24.0% A, 27.6% T, 18.6% G and 29.7% C bases. The corresponding composition for *S. gibbosa* is 23.3% A, 26.6% T, 19.5% G and 30.6% C bases. In both the species, the major coding strand (H-strand) was observed to be relatively AC rich in comparison to the L-strand (53.9% and 53.7% of sites were AC in *S. longiceps* and *S. gibbosa* respectively) (Table 2.2). Variations in the composition of H and L-strand have been reported for vertebrate mitochondrial DNA (Perna and Kocher 1995; Min and Hickey 2007; Fischer *et al.* 2013). As in other vertebrates, an anti-G bias in the codon 3$^{rd}$base position and high pyrimidines presence in the second codon positions were observed in both the genome (Table 2.2). In the second codon position, the anti-G bias was larger (13.95% and 13.85% of sites were G in *S. longiceps* and *S. gibbosa* respectively), similar to reports as in other vertebrate species (Naylor *et al.*, 1995; Boore, 1999). In both the genome, the most frequently used amino acids were Leucine (16.3%/16.4%), followed by Alanine (9.6%/9.6%) and Threonine (8.0%/8.2%) **(**SL/SG) (Table 2.3). Mitogenomes with low GC and high AT content encode proteins highly enriched with amino acids encoded by CA-rich codons (Min and Hickey 2007). Threonine and Proline are amino acids encoded by CA-rich codons and account for ~14% of encoded amino acids. Codon preference for each amino acid in protein-coding gene sequences were identified with the highest estimated RSCU values and were matched to all 22 identified tRNAs in the genome (Table 2.3), except for Alanine, Isoleucine, Leucine, Proline, Serine, Threonine and Valine in *S. longiceps* and Alanine, Proline, Serine, Threonine and Valine in *S. gibbosa.* When considering degenerate third codon positions, codons complementary to the tRNAs ending in A and C were the most frequently observed in both species. G nucleotide was the least frequent in both genomes (Table 2.3). These observations were consistent with the anti-G bias identified in the mitogenome.

**Table 2.2** Nucleotide composition of the mitogenome of *S. longiceps* and *S. gibbosa.*

| *S. longiceps* | | | | *S. gibbosa* | | | |
|---|---|---|---|---|---|---|---|
| % Nucleotide composition | | | | | | | |
| A | C | G | T | A | C | G | T |
| Complete mitogenome (H- Strand) | | | | | | | |
| 23.3 | 30.6 | 19.5 | 26.6 | 24 | 29.7 | 18.6 | 27.6 |
| All protein coding gene concatenated (H- Strand) [a] | | | | | | | |
| 24.7 | 30.4 | 17.9 | 27 | 23.8 | 31.2 | 18.9 | 26.1 |
| ND 6 (L- Strand) [b] | | | | | | | |
| 36.7 | 32.8 | 17.6 | 12.9 | 32.8 | 33.7 | 19.3 | 14.2 |
| 1st codon position[c] | | | | | | | |
| 25.296 | 26.323 | 27.56 | 20.821 | 25.638 | 27.27 | 27.297 | 19.795 |
| 2nd codon position[c] | | | | | | | |
| 17.689 | 28.139 | 13.951 | 40.221 | 17.768 | 28.297 | 13.819 | 40.116 |
| 3rd codon position[c] | | | | | | | |
| 29.166 | 34.667 | 14.293 | 21.874 | 26.533 | 36.141 | 17.478 | 19.847 |

a Based on the 12 protein-coding genes located on the H-strand.
b Base on the ND 6 gene located on the L-strand.
c Based on the 13 protein-coding genes.

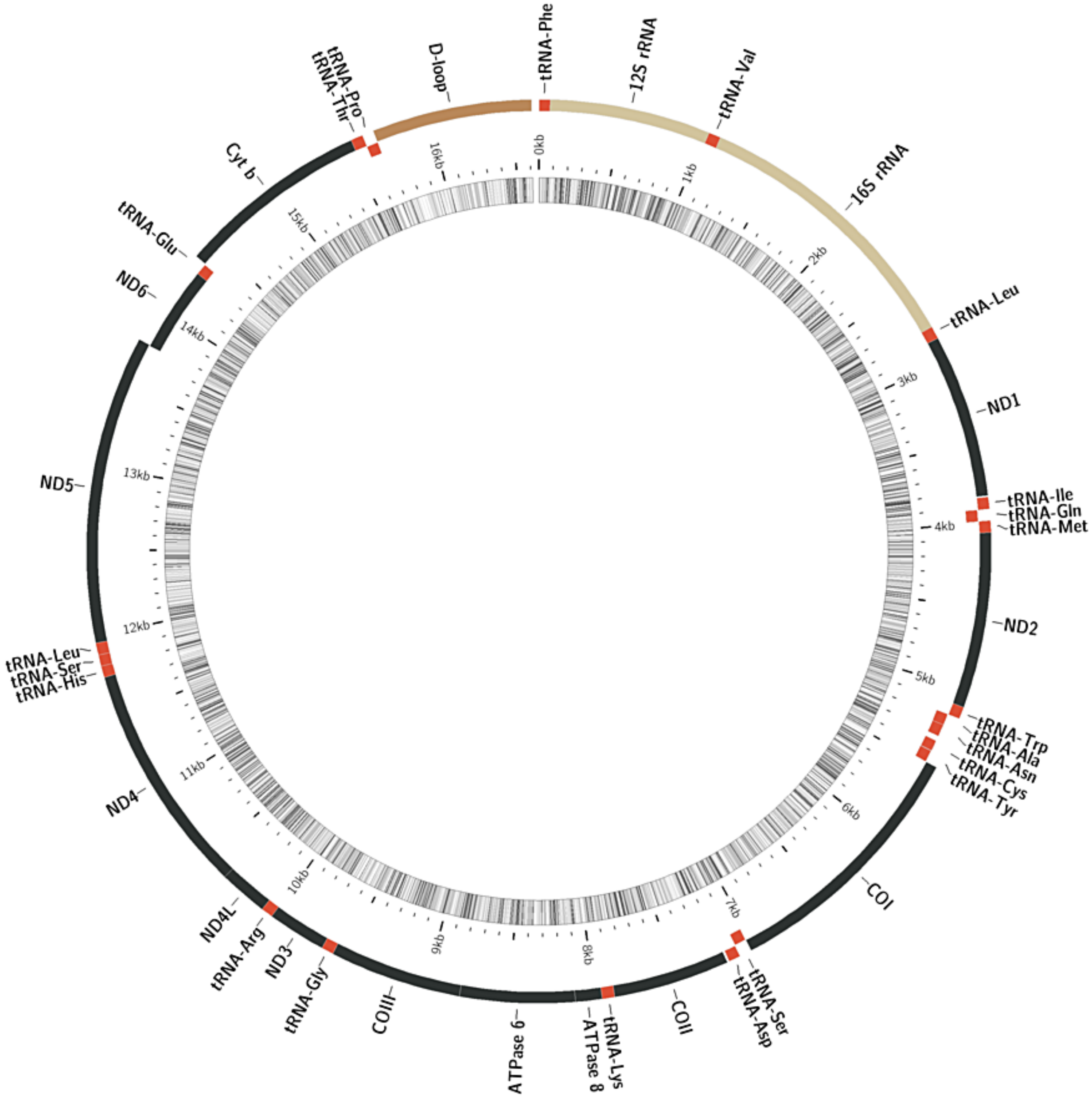


**Fig. 2.1a** Mitogenome map of *S. longiceps* (16,613 bp) (Gen Bank accession no. KR000002.1) generated with MitoAnnotator. Protein-coding genes, tRNAs, rRNAs, and D-loop regions are shown in different colours. Genes located within the outer circle are coded on the H-strand whereas the remaining genes are coded on the L-strand.

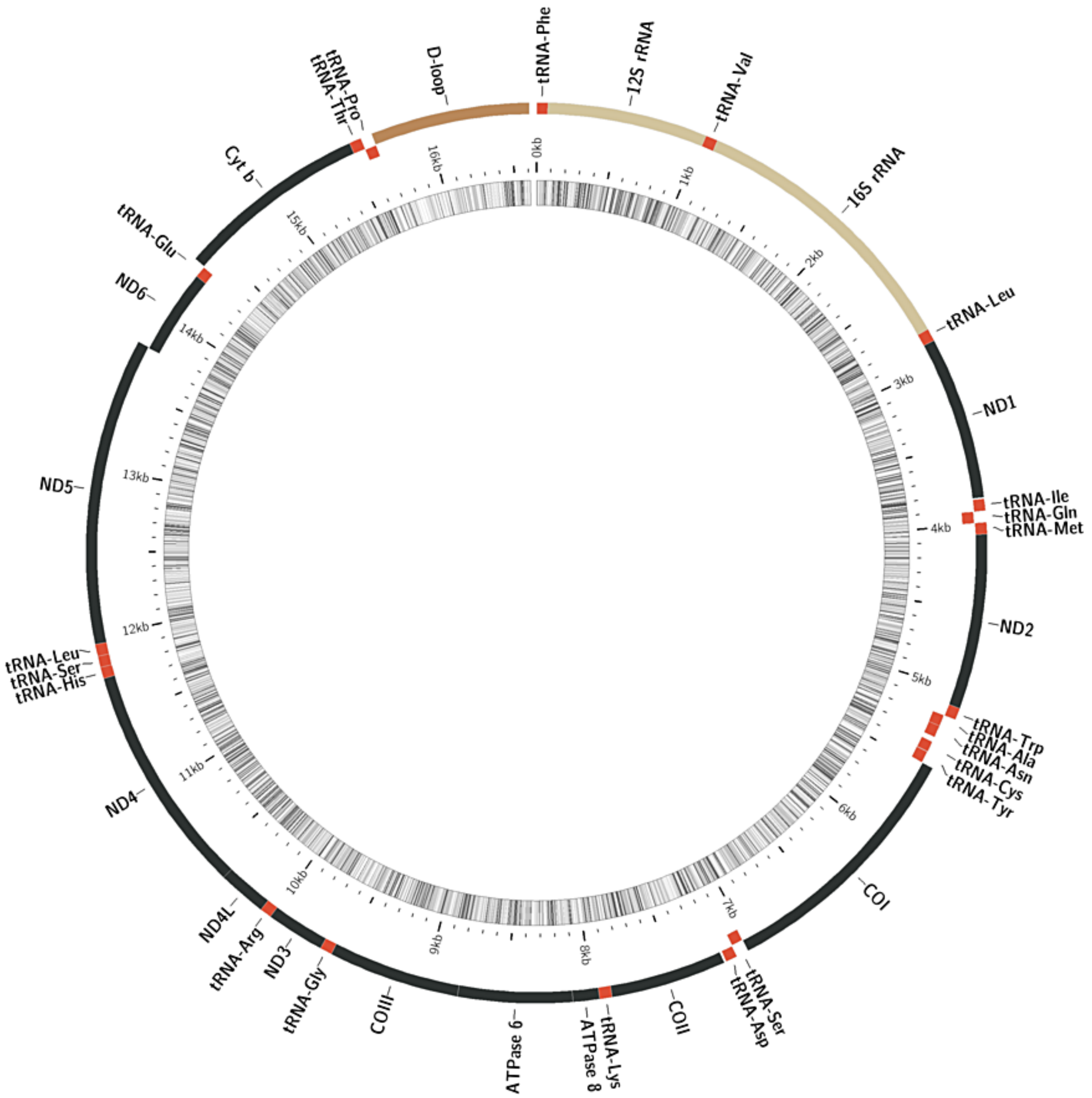


**Fig. 2.1b** Mitogenome map of *S. gibbosa* (16658 bp) (Gen Bank accession no. KU665488.1) generated with MitoAnnotator. Protein-coding genes, tRNAs, rRNAs, and D-loop regions are shown in different colours. Genes located within the outer circle are coded on the H-strand whereas the remaining genes are coded on the L-strand.

**Table 2.3.** Amino acid and codon usage in the mitogenome of *S. longiceps* and *S. gibbosa*.

| Amino acid | *S. longiceps* | | | *S. gibbosa* | | |
|---|---|---|---|---|---|---|
| | %[a] | Codons | RCSUC[b] | % [a] | Codons | RSCU[b] |
| Alanine(Ala/A) | 9.7 | GCU | 0.64 | 9.6 | GCU | 0.69 |
| | | GCC | **1.84** | | GCC | **1.73** |
| | | GCA* | 1.26 | | GCA* | 1.22 |
| | | GCG | 0.26 | | GCG | 0.35 |
| Arginine(Arg/R) | 2 | CGU | 0.31 | 2 | CGU | 0.43 |
| | | CGC | 0.31 | | CGC | 0.43 |
| | | CGA* | **2.63** | | CGA* | **2.22** |
| | | CGG | 0.75 | | CGG | 0.92 |
| Asparagine(Asn/N) | 2.9 | AAU | 0.73 | 2.9 | AAU | 0.63 |
| | | AAC* | **1.27** | | AAC* | **1.37** |
| AsparticAcid(Asp/D) | 2 | GAU | 0.52 | 2 | GAU | 0.38 |
| | | GAC* | **1.48** | | GAC* | **1.62** |
| Cysteine(Cys/C) | 0.8 | UGU | 0.92 | 0.8 | UGU | 0.54 |
| | | UGC* | **1.08** | | UGC* | **1.46** |
| GlutamicAcid(Glu/E) | 2.7 | GAA* | **1.41** | 2.7 | GAA* | **1.22** |
| | | GAG | 0.59 | | GAG | 0.78 |
| Glutamine(Gln/Q) | 2.5 | CAA* | **1.44** | 2.4 | CAA* | **1.4** |
| | | CAG | 0.56 | | CAG | 0.6 |
| Glycine(Gly/G) | 6.5 | GGU | 0.36 | 6.5 | GGU | 0.39 |
| | | GGC | 0.79 | | GGC | 0.95 |
| | | GGA* | **1.88** | | GGA* | **1.44** |
| | | GGG | 0.97 | | GGG | 1.22 |
| Histidine(His/H) | 2.7 | CAU | 0.4 | 2.7 | CAU | 0.42 |
| | | CAC* | **1.6** | | CAC* | 1.58 |
| Isoleucine(Ile/I) | 7.1 | AUU | **1.04** | 6.9 | AUU | 0.94 |
| | | AUC* | 0.96 | | AUC* | 1.06 |
| Leucine(Leu/L) | 16.1 | UUA* | 0.8 | 16.3 | UUA* | 0.41 |
| | | UUG | 0.19 | | UUG | 0.26 |
| | | CUU | **1.51** | | CUU | **1.44** |
| | | CUC | 1.24 | | CUC | 1.14 |
| | | CUA* | **1.5** | | CUA* | **1.68** |
| | | CUG | 0.75 | | CUG | 1.08 |
| Lysine(Lys/K) | 2.1 | AAA* | **1.16** | 2 | AAA* | **1.06** |
| | | AAG | 0.84 | | AAG | 0.94 |
| Methionine(Met/M) | 3.9 | AUA | 0.46 | 4 | AUA | 0.45 |
| | | AUG* | **1.54** | | AUG* | **1.55** |
| Phenylalanine(Phe/F) | 6.2 | UUU | 6.3 | | UUU | 0.52 |
| | | UUC* | **1.37** | | UUC* | **1.48** |
| Proline(Pro/P) | 5.7 | CCU* | 0.84 | 5.7 | CCU* | 1.03 |
| | | CCC | **1.42** | | CCC | **1.35** |
| | | CCA | 1.36 | | CCA | 0.99 |
| | | CCG | 0.38 | | CCG | 0.63 |
| Serine(Ser/S) | 6.3 | UCU | 0.9 | 6.2 | UCU | 0.7 |
| | | UCC | **1.63** | | UCC | **2.21** |
| | | UCA* | **1.45** | | UCA* | 1.21 |
| | | UCG | 0.54 | | UCG | 0.39 |
| | | AGU | 0.21 | | AGU | 0.18 |
| | | AGC* | 1.27 | | AGC* | **1.3** |
| Threonine(Thr/T) | 7.7 | ACU | 0.79 | 8.2 | ACU | 0.89 |
| | | ACC | **1.8** | | ACC | **1.65** |
| | | ACA* | 1.16 | | ACA* | 1.11 |
| | | ACG | 0.26 | | ACG | 0.35 |
| Tryptophan(Trp/W) | 3.3 | UGA* | **1.65** | 3.1 | UGA* | **1.45** |
| | | UGG | 0.35 | | UGG | 0.55 |
| Tyrosine(Tyr/Y) | 3.1 | UAU | 0.76 | 3 | UAU | 0.37 |
| | | UAC | **1.24** | | UAC* | **1.63** |
| Valine(Val/V | 6.7 | GUU | 0.93 | 6.6 | GUU | 0.72 |
| | | GUC | **1.44** | | GUC | 1.3 |
| | | GUA* | 1.18 | | GUA* | **1.38** |
| | | GUG | 0.45 | | GUG | 0.6 |

a % of Amino acid based on the 13 protein-coding genes.

b RSCU relative synonymous codon usage.

* Codons that are complementary to the tRNA genes.

### 3.3. RNA genes

A small (12S rRNA) and large (16S rRNA) ribosomal RNA subunit was identified, where 12S rRNA has 954/951 bp and 16S rRNA has 1689/1687 bp length (SL/SG). The overall similarity of rRNA genes is 90%. In both the species, rRNA genes have high adenine content similar to other vertebrates (Naylor *et al.* 1995; Boore 1999). As in coding genes, 3 of the 22 tRNA genes showed overlaps. In both the species, tRNA Gln shared one nucleotide at both ends, upstream with tRNA Ile and downstream with tRNA Met.

### 3.4. Non-coding region

As in most vertebrates, the origin of light strand replication ($O_L$) in both sardines was located in tRNA Asn and tRNA Cys (WANCY region) and it is from 5308 bp to 5339 bp in *S. longiceps* & 5309 bp to 5345 bp in *S. gibbosa.* This region can fold into a stable stem-loop secondary structure. A major non-coding region between the tRNA-Pro and tRNA-Phe genes were considered as the control region (D-loop) which is 958 bp in *S. longiceps* and 992 bp in *S. gibbosa.* It has several characteristic conserved sequence blocks (CSB) like CSB D, CSB2, CSB3, termination associated sequence (TAS) and Poly T (Fig 2.S1).

### 3.5. Phylogenetic analysis

The mitogenomic phylogenetic tree constructed using Maximum likelihood method showed six moderately supported monophyletic groups within the Clupeidae (Fig. 2.2), as observed in a previous investigation (Lavoue *et al*. 2007, Lavoue *et al*. 2013). The family Clupeidae and its five subfamilies are not monophyletic. Only three of the nine currently recognised family, Engraulidae, Pristigasteridae and Dussumieriidae formed well-supported monophyletic groups and the relationships among other groups were not well supported. Both *S. gobbosa* and *S. longiceps* were grouped with other species in the genus Sardinella, Tenualosa, Gudusia, Potamothrissa, Microthrissa, Pellonula, Odaxothrissa, Ethmalosa, Dorosoma, Harengula, Nematalosa, Clupanodon, Konosirus and Escualosa in the lineage 1.

Even though there are several studies on the major phylogenetic lineages in the clupeoids, the phylogenetic relationships among Clupeids are still under debate (Lavoue *et al*. 2013).

The present study also failed to delimit the formally valid species in the clupeids. Three of the lineages obtained were consistent with the well-defined (by several morphological characters) families Engraulidae, Pristigasteridae and Dussumieriidae. The anchovy family Engraulidae is a well-defined monophyletic group (Grande and Nelson 1985; Lavoue *et al*. 2007, 2009) with 140 species divided into 16 genera found in temperate and tropical regions around the world. Within Engraulinae (subfamily), the New World taxa and Engraulis formed a clade referred to as Engraulini following Lavoue *et al*. (2009). Several morphological characters supported the monophyly of Engraulini, most notably the loss of ventral scutes (Nelson 1970, 1983; Grande 1985; Grande and Nelson 1985), a character present in nearly all other clupeomorph fishes.

Five lineages composed of species formally classified in different clupeid subfamilies (lineage 1-4), make all the traditional clupeid subfamilies monophyletic. None of these five lineages was delimited using morphological characters. Similar to the previous study all these observations lead to a conclusion, that is the phylogenetic signals in the mitochondrial genome are very weak because of the shallow genealogies among formally valid species in clupeid family (Lavoue *et al.* 2013; Thomas *et al.* 2014; Stern *et al.* 2016). In addition, overlapping morphological characters have been reported among species in the family and some of the studies reported reduced number of species in the family (taxonomic over-splitting in clupeid family) (Thomas *et al.* 2014; Stern *et al.* 2016) and the possible existence of different populations or ecotypes of single a species in the clupeid family. (Thomas *et al.* 2014; Stern *et al.* 2016). Advanced investigations using nuclear markers are necessary to resolve this uncertainty.

Even though the uncertainty in the phylogenetic relationship exists, it has been reported that the early diversification of clupeoids occurred in the Tethys sea region (Indo west pacific precursor region) (Lavoue *et al.* 2013). Predicted divergence time showed that they had already diverged significantly in the upper Cretaceous/Early Eocene period. According to the reports of character evolution reconstruction, earlier clupeoids were restricted to marine habitat only, later multiple and independent transitions from marine to freshwater and tropic to temperate habitats occurred. All these transitions occurred at the end of Cretaceous or Early in the Cenozoic Era, at the time of significant global cooling and Cretaceous clupeoids also faced the K-Pg mass extinction period (Dynesius and Jansson 2000; Lavoue *et al.* 2013; Zuloaga *et al.* 2019).

The first report of the complete mitochondrial genome sequence of *S. longiceps* and *S. gibbosa* revealed gene organization, structure, content and order similar to most vertebrates. This will provide baseline information for further studies on the taxonomic resolution, conservation, adaptive variation to environmental clines and evolution regarding these commercially and ecologically important species.

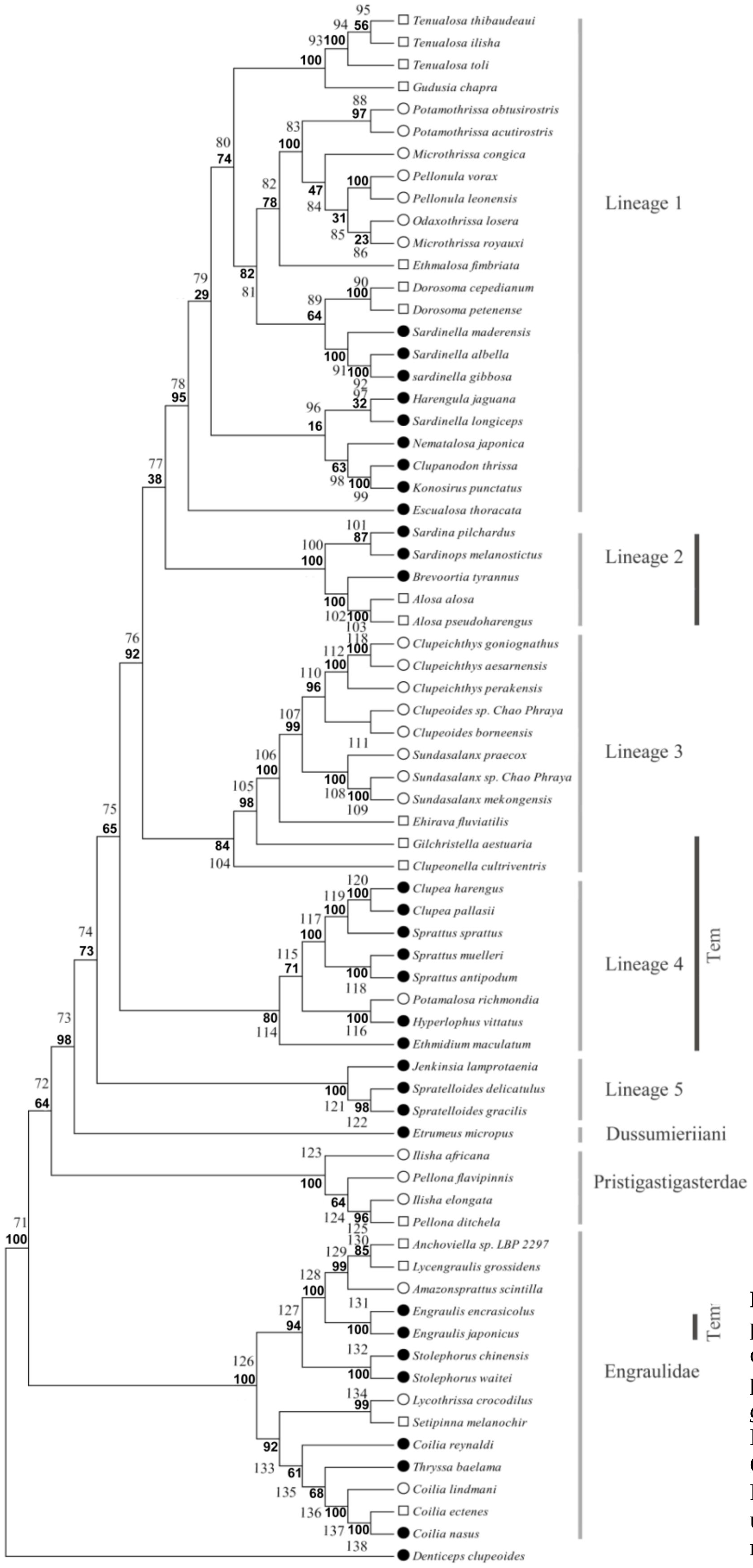


**Fig. 2.2.** Maximum likelihood phylogenetic tree generated by alignment of nucleotide sequences (12 concatenated protein-coding genes) of *S. longiceps, S. gibbosa* and fishes of the family Denticipitidae, Clupeidae, Engraulidae, Chirocentridae, Dussumieriidae and Pristigasteridae. *Denticeps clupeoides* was used as out group. Bootstrap values and node numbers are indicated in bold and grey letters respectively. Black circle, white circles and square in the tree indicates marine, brackish and fresh water species respectively. 'Temp' indicates temperate water species.

# Supplementary Figures and Tables

a) *Sardinella longiceps*

```
           15660     15670     15680     15690     15700     15710     15720     15730     15740     15750
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
tRNA-Phe- AAAACGCGCCAGTTATAGTATTGTCCATTCTGCACACCTAATCACGATAGATTACATTGGCACAGTCAGGGGGTTAAAAATGTCTATGCATAATAGTGCA
                                                           TAS
            15760     15770     15780     15790     15800     15810     15820     15830     15840     15850
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
          TATATTTATGGTGTAGTACATATATATGCATGATTATACATACTTATGGTTTAATACATTAAACTAATGACTACCATATAAAATAGTTAAACCATACAGG

            15860     15870     15880     15890     15900     15910     15920     15930     15940     15950
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
          AAAGACAACACACTAAGGTTGACCCAAACCATCACACAAGATTCAGAAAAAATAGAAAGTCAGCTGATAAATAGAATAATCCCCATAACTGCAATTAAAC

            15960     15970     15980     15990     16000     16010     16020     16030     16040     16050
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
          ACTTTCTATGCGTTATTCATCAAATATAGAAGTCCAGGATAACTGAATGTACTAAGAACCGACCAATGAGATAAATAATTGCATATCATGAATGATAAGA

            16060     16070     16080     16090     16100     16110     16120     16130     16140     16150
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
          TCACGGACAACAATTGTGGGGGTTTCACATGGTGAACTATTCCTGGCATTTGGTTCCTATTTCAGGGCCATTCTTGTAAGTTATTCCTCCCTAGTGAATT
          CSB D
            16160     16170     16180     16190     16200     16210     16220     16230     16240     16250
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
          CTCCGTGCATAAGTTAATGTTAGAGTACTATCGACTCGTTACCCACCAAGCCTAGCATTCACTTATATGCATTTGGTATTTTTTTTTCGGCTCACACTCA

            16260     16270     16280     16290     16300     16310     16320     16330     16340     16350
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
          TCCGCATTTGGCGACTCCTTCCTAATGTTAACTTTCACGGTTGAACATATTCCTTGATTGGAACGTATAAACTGTAAAACTTCATTAGCATTGACAGAAG

            16360     16370     16380     16390     16400     16410     16420     16430     16440     16450
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
          AATTGCATAACTGTTTATCAGGTGCATAAACTATCTATTCTTTCCTCAGAATCTCCATTATCAGTGCCCCCTTTCGCTTCTGCGAGAAGGTTTTCGCGCG

            16460     16470     16480     16490     16500     16510     16520     16530     16540     16550
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
          AACAAACCCCCCTACCCCCCTACGCCGGAAGGAGTCTCTGTTTATTCAATGTCAAACCCCAAAATCCATGGAAGTTCTCGACCAGCGTCTTGCAGCGAGT
                   CSB 2                                      CSB 3
            16560     16570     16580     16590     16600     16610
          ....|....|....|....|....|....|....|....|....|....|....|...
          TCCGTTATTTGCTGTCTTTATATATGCCTCCAAAATTGTATCACTCTGTAAAGCTTGC-tRNA-Pro
```

b) *Sardinella gibbosa*

```
          15670     15680     15690     15700     15710     15720     15730     15740     15750     15760
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
tRNA-Phe- CCGGGCGCGCCCACATAGCACGCCCCACTGTTATAGTACTCTTAGATTAGAGTCCTTAACGAGTACACCAGTATGGTACTAATACATAATATGCATAATT
          TAS
           15770     15780     15790     15800     15810     15820     15830     15840     15850     15860
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
          ATACATACATATATGTACTAGTACATATTATGTATAATTATACATATATATATGTATTAGGTACATATTATGCATGATTATACATATTTATGGTTTAACA

           15870     15880     15890     15900     15910     15920     15930     15940     15950     15960
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
          CATAACATTAATAATCACCCAACTAATAAATAAAACAAACAAGTGAGATAATTAATAAGGTATACCCAAAGCATTCCATTAAGATTCAGAATAATTCTAA

           15970     15980     15990     16000     16010     16020     16030     16040     16050     16060
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
          TATAACCTGATAAACAGATTAACCCCCATAACTGGAGTTAAGCATTTTCCATGCGTTATTCATCAATGATTGAACTTAATAGAAAAGATGTAATAAGAAC

           16070     16080     16090     16100     16110     16120     16130     16140     16150     16160
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
          CGACCAATGAGATAAATAATTGCATATAATGCATGATAGAATCAAGGACAACTATTGTGGGGGTTTCACAGAATGCACTATTCCTGGCATCTGGTTCCTA
          CSB D
           16170     16180     16190     16200     16210     16220     16230     16240     16250     16260
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
          TTTCAGGGCCCATAAACCGGTAGTCCCTCCCTACTTGAATTGTCCTGACATAGGTTAATGGTGGGGTGCTAGAGCTTCTTTACCCCCCATGCCGAGCGCT

           16270     16280     16290     16300     16310     16320     16330     16340     16350     16360
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
          CACTCTAAAGGCCATGGGGTATTTTTTTTTTCGGGTCTCTTTCATCTTGCATCTGGCGACTCCCTCCTAATGTTAACTTACAAGGTGGTCCTATTCTTCT

           16370     16380     16390     16400     16410     16420     16430     16440     16450     16460
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
          TGCTTTATATGTCGTATGTGAAATGATTAGACCACTATTGGAAGATATAACCCCATAATTGATATCAGGTGCATAACAGTACTACTACTTGCTTCACACA

           16470     16480     16490     16500     16510     16520     16530     16540     16550     16560
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
          TATATCTCTCGAGAGCCCCCCTTTCATCCTTTGTAAAAAAATTTTTTACGCCCCCCCCTCCCCCTTACGCCCGAAAAGTCCTGTTACTATTCTTGTTAAA
                                                              CSB 2
           16570     16580     16590     16600     16610     16620     16630     16640     16650
          ...|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|...
          CCCCAAAACCAAATGGAAGTCTCGACCAGCGTCTTCAACGAGTTCTGATGTGTGTTGGTATATATAGTGTTGCAAAAAGGTGCCATTGTGTA-tRNA-Pro
          CSB 3
```

**Fig. 2.S1** Characteristics conserved blocks (CSB), (TAS) and Poly Tin the non-coding region (D-Loop) of *Sardinella longiceps*(a) and *Sardinella gibbosa*(b) mitochondrial DNA.

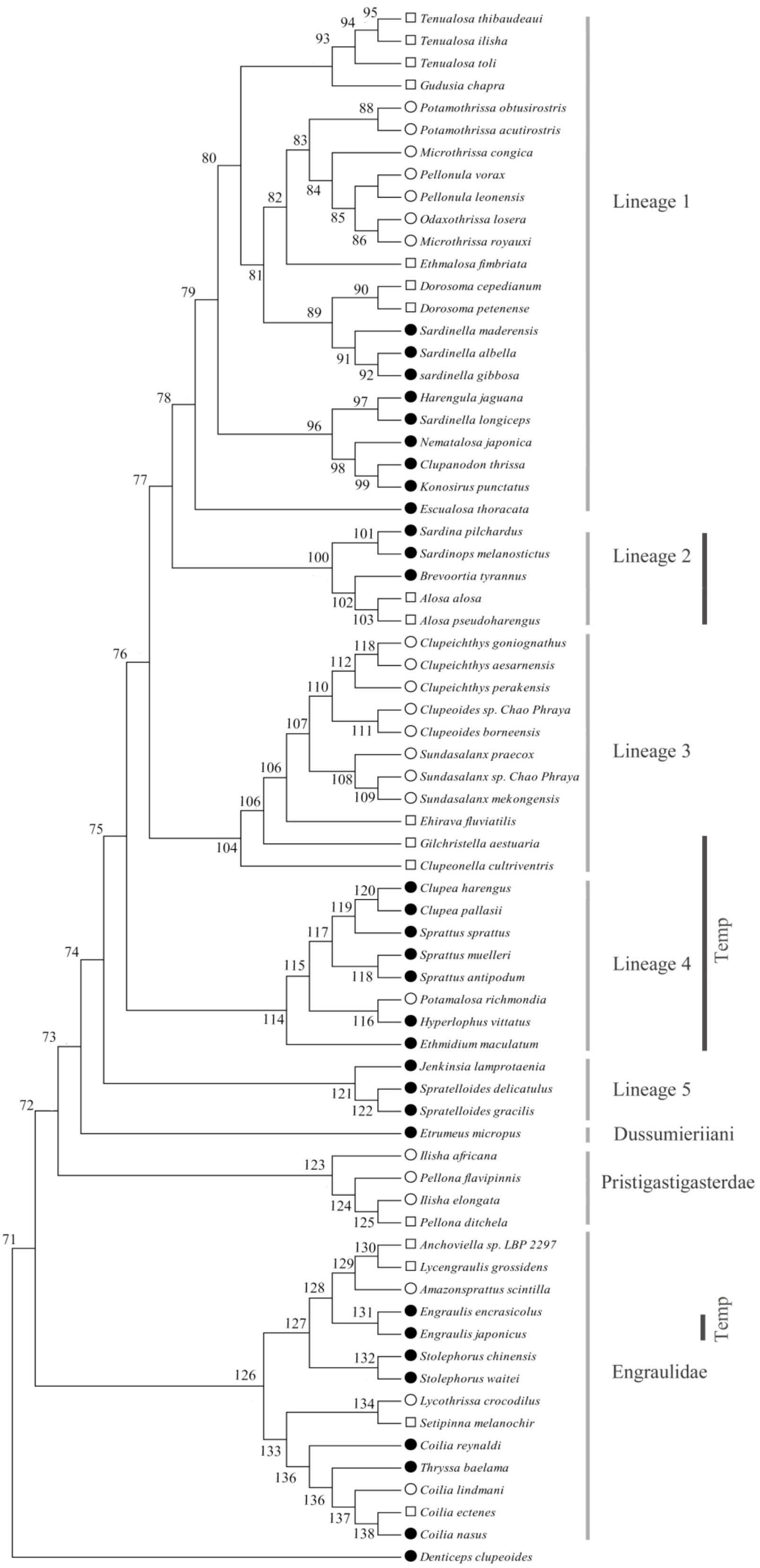


**Fig. 2.S2** Neighbor-joining phylogenetic tree generated by alignment of complete mitogenome nucleotide sequences of *S. longiceps, S. gibbosa* and fishes of the family Denticipitidae, Clupeidae, Engraulidae, Chirocentridae, Dussumieriidae and Pristigasteridae. *Denticeps clupeoides* was used as outgroup. The numbers in the nodes of the phylogenetic tree are node number. Black circle, white circles and square in the tree indicates marine, brackish and freshwater species respectively. 'Temp' indicates temperate water species.

**Table 2.S1** List of Primer pairs used for amplification of *S. longiceps* and *S. gibbosa* mitochondrial DNA.

| Primer Name | | Sequence (5' - 3') | PCR Product length |
|---|---|---|---|
| SPF M 1 | Forward primer | AAGAGGGCCGGTAAAACTCG | 1080 |
| | Reverse primer | GGTTTCGGGGGCTCAAACTA | |
| SPF M 2 | Forward primer | CACAATATTCGCCGCAAGGG | 1140 |
| | Reverse primer | GCGGCCGTTAAACTTTTGGT | |
| SPF M 3 | Forward primer | TCCTGCAGCAAGACATCGTT | 1287 |
| | Reverse primer | AGGCTGGATAGGGCCAAAAC | |
| SPF M 4 | Forward primer | GTTTTGGCCCTATCCAGCCT | 1390 |
| | Reverse primer | TTGGGTCTGGTTAAGACCGC | |
| SPF M 5 | Forward primer | CCACCCCTACCTCCTAACGA | 1267 |
| | Reverse primer | ATGCCATATCAGGTGCTCCG | |
| SPF M 6 | Forward primer | CTCTGTCAGGCAATCTGGCA | 1299 |
| | Reverse primer | ACGCAGGGGTTTAACCTACG | |
| SPF M 7 | Forward primer | CGTAGGTTAAACCCCTGCGT | 1307 |
| | Reverse primer | AATCACCGTAGCAAGCCACA | |
| SPF M 8 | Forward primer | TGTGGCTTGCTACGGTGATT | 1071 |
| | Reverse primer | GCTGCCTCAAACCCAAAGTG | |
| SPF M 9 | Forward primer | ACCACTTTGGGTTTGAGGCA | 1130 |
| | Reverse primer | CATGTGGTTCTGGCTGGCTA | |
| SPF M 10 | Forward primer | GATCATCGCCTCTCTGAGCC | 1131 |
| | Reverse primer | AGAGAGTACCCGGCTGTGAT | |
| SPF M11 | Forward primer | ATCACAGCCGGGTACTCTCT | 1458 |
| | Reverse primer | TTGCTCATCGTTGAGGCTGT | |
| SPF M 12 | Forward primer | ACAGGCACCCCTTTCTTAGC | 1344 |
| | Reverse primer | TCTGGAGCTTGTTGCGTCAT | |
| SPF M 13 | Forward primer | AGAGCTCACCGGGTATTCCT | 1018 |
| | Reverse primer | AAGTGGAACGCGAAAAACCG | |
| SPF M 14 | Forward primer | CGGTTTTTCGCGTTCCACTT | 1287 |
| | Reverse primer | AAGGACTCGCCAGATGCAAA | |

**Table 2.S2** List of species used in the phylogenetic analysis.

| **Species** | **Accession numbers** |
|---|---|
| *Alosa alosa* | AP009131 |
| *Alosa pseudoharengus* | AP009132 |
| *Amazonsprattus scintilla* | AP009617 |
| *Anchoviella sp. LBP 2297* | AP011557 |
| *Brevoortia tyrannus* | AP009618 |
| *Clupanodon thrissa* | JX075099 |
| *Clupea harengus* | AP009133 |
| *Clupea pallasii* | AP009134 |
| *Clupeichthys aesarnensis* | AP011584 |
| *Clupeichthys goniognathus* | AP011589 |
| *Clupeichthys perakensis* | AP011585 |
| *Clupeoides borneensis* | AP011586 |
| *Clupeoides sp. Chao Phraya* | AP011587 |
| *Clupeonella cultriventris* | AP009615 |
| *Coilia ectenes* | JX625133 |
| *Coilia lindmani* | AP011558 |
| *Coilia nasus* | AP009135 |
| *Coilia reynaldi* | AP011559 |
| *Denticeps clupeoides* | AP007276 |
| *Dorosoma cepedianum* | DQ536426 |
| *Dorosoma petenense* | AP009136 |
| *Ehirava fluviatilis* | AP011588 |
| *Engraulis encrasicolus* | AP009137 |
| *Engraulis japonicus* | AB040676 |
| *Escualosa thoracata* | AP011601 |
| *Ethmalosa fimbriata* | AP009138 |
| *Ethmidium maculatum* | AP011602 |
| *Etrumeus micropus* | AP009139 |
| *Gilchristella aestuaria* | AP011606 |
| *Gudusia chapra* | AP011603 |
| *Harengula jaguana* | AP011592 |
| *Hyperlophus vittatus* | AP011593 |
| *Ilisha africana* | AP009140 |
| *Ilisha elongata* | AP009141 |

| | |
|---|---|
| *Jenkinsia lamprotaenia* | AP006230 |
| *Konosirus punctatus* | AP011612 |
| *Lycengraulis grossidens* | AP011563 |
| *Lycothrissa crocodilus* | AP011562 |
| *Microthrissa congica* | AP011598 |
| *Microthrissa royauxi* | AP011596 |
| *Nematalosa japonica* | AP009142 |
| *Odaxothrissa losera* | AP011595 |
| *Pellona ditchela* | AP011609 |
| *Pellona flavipinnis* | AP009619 |
| *Pellonula leonensis* | AP009232 |
| *Pellonula vorax* | AP009231 |
| *Potamalosa richmondia* | AP011594 |
| *Potamothrissa acutirostris* | AP011597 |
| *Potamothrissa obtusirostris* | AP011599 |
| *Sardina pilchardus* | AP009233 |
| *Sardinella albella* | AP011605 |
| *Sardinella maderensis* | AP009143 |
| *Sardinops melanostictus* | AB032554 |
| *Setipinna melanochir* | AP011565 |
| *Spratelloides delicatulus* | AP009144 |
| *Spratelloides gracilis* | AP009145 |
| *Sprattus antipodum* | AP011608 |
| *Sprattus muelleri* | AP011607 |
| *Sprattus sprattus* | AP009234 |
| *Stolephorus chinensis* | AP011566 |
| *Stolephorus waitei* | AP011567 |
| *Sundasalanx mekongensis* | AP006232 |
| *Sundasalanx praecox* | AP011591 |
| *Sundasalanx sp. Chao Phraya* | AP011590 |
| *Tenualosa ilisha* | AP011611 |
| *Tenualosa thibaudeaui* | AP011604 |
| *Tenualosa toli* | AP011600 |
| *Thryssa baelama* | AP009616 |

## 4. References

1. Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Mol Ecol* 13(4):729-744
2. Boore JL, (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27(8):1767-1780
3. Bradbury IR, Hubert S, Higgins B *et al* (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proc R Soc Lond B Biol Sci* 277(1701):3725-3734
4. Broughton RE, Milam JE, Roe BA (2001) The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Res* 11(11):1958-1967
5. Caballero S, Duchene S, Garavito MF, Slikas B, Baker CS (2015) Initial evidence for adaptive selection on the NADH subunit Two of freshwater dolphins by analyses of mitochondrial genomes. *PloS one* 10(5):e0123543
6. Cadrin SX, Kerr LA, Mariani S (2013) Stock identification methods: applications in fishery science. Academic Press
7. CMFRI Kochi (2017) CMFRI Annual Report 2016-2017. Technical Report, CMFRI, Kochi
8. Curole, JP, Kocher TD (1999) Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol Evolut* 14(10):394-398
9. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A (2008) The adaptive evolution of the mammalian mitochondrial genome. *BMC genomics* 9(1):119
10. Devaraj M, Kurup KN, Pillai NGK, Balan K, Vivekanandan E, Sathiadas R (1997) Status, prospects and management of small pelagic fisheries of India. In: Devaraj M, Martosubroto P (eds) Small pelagic resources and their fisheries in the Asia-Pacific Region Proceedings of APFIC working party on Marine Fisheries. RAP Publishers, Thailand
11. Fischer C, Koblmuller S, Gully C, Schlotterer C, Sturmbauer C, Thallinger GG (2013) Complete mitochondrial DNA sequences of the threadfin cichlid (*Petrochromis trewavasae*) and the blunt head cichlid (*Tropheus moorii*) and patterns of mitochondrial genome evolution in cichlid fishes. *Plos One* 8(6):e67048
12. Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, Gilbert MTP, da Fonseca RR (2011) Positive selection on the killer whale mitogenome. *Biol Lett* 7(1):116-118
13. Garvin MR, Bielawski JP, Gharrett AJ (2012) Correction: Positive Darwinian Selection in the Piston That Powers Proton Pumps in Complex I of the Mitochondria of Pacific Salmon. *PloS one* 7(8):e24127
14. Iwasaki W, Fukunaga T, Isagozawa R*et al.* (2013) MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol Biol Evol* 30(11):2531-2540
15. Jacobsen MW, Hansen MM, Orlando L *et al.* (2012) Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp). *Mol Ecol* 21(11):2727-2742
16. Karl SA, Toonen RJ, Grant WS, Bowen BW (2012) Common misconceptions in molecular ecology: echoes of the modern synthesis. *Mol Ecol* 21(17):4171-4189
17. Kearse M, Moir R, Wilson A*et al*. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649
18. Lavoue S, Miya M, Musikasinthorn P, Chen WJ, Nishida M (2013) Mitogenomic evidence for an Indo-west pacific origin of the clupeoidei (Teleostei: Clupeiformes). *Plos One* 8(2):e56485
19. Lavoue S, Miya M, Saitoh K, Ishiguro NB, Nishida M (2007) Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Mol Phylogenet Evol* 43(3):1096-1105
20. Meyer A (1993) Evolution of mitochondrial DNA in fishes. In: Hochachka PW, Mommsen TP (ed) Biochemistry and Molecular Biology of Fishes. Elsevier Science Publishers, New York pp 01-38
21. Min XJ, Hickey DA (2007) DNA asymmetric strand bias affects the amino acid composition of mitochondrial proteins. *DNA Res* 14:201-206

22. Miya M, Nishida M (1999) Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): first example of transfer RNA gene rearrangements in bony fishes. *Mar Biotechnol* 1(5):416-426
23. Miya M, Nishida M (2015) The mitogenomic contributions to molecular phylogenetics and evolution of fishes: a 15-year retrospect. *Ichthyol Res* 62(1):29-71
24. Morin PA, Archer FI, Foote AD *et al.* (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res* 20(7):908-916
25. Mu X, Liu Y, Lai M, Song H, Wang X, Hu Y, Luo J (2015) Characterization of the *Macropodus opercularis* complete mitochondrial genome and family Channidae taxonomy using Illumina based de novo transcriptome sequencing. *Gene* 559:189-195
26. Naylor GJ, Collins TM, Brown WM (1995) Hydrophobicity and phylogeny. *Nature* 373(6515):565-566
27. Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290(5806):470-474
28. Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol* 41:353-358
29. Sambrook J, Russell D (2001) Molecular Cloning: A Laboratory Manual. 3rd edn, Cold Spring Harbor Laboratory Press, New York
30. Silva G, Lima FP, Martel P, Castilho R (2014) Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. *Proc R Soc Lond B Biol Sci* 281(1792):20141093
31. Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol Phylogenet Evol* 12(2):105-114
32. Stager M, Cerasale DJ, Dor R, Winkler DW, Cheviron ZA (2014) Signatures of natural selection in the mitochondrial genomes of Tachycineta swallows and their implications for latitudinal patterns of the pace of life. *Gene* 546(1):104-111
33. Stern N, Rinkevich B, Goren M (2016) Integrative approach revises the frequently misidentified species of Sardinella (Clupeidae) of the Indo-West Pacific Ocean. *J Fish Biol* 89(5):2282-2305
34. Sukumaran S, Gopalakrishnan A, Sebastian W*et al.*(2016a) Morphological divergence in Indian oil sardine, *Sardinella longiceps* Valenciennes, 1847-Does it imply adaptive variation? *J Appl Ichthyol* 32(4):706-711
35. Sukumaran S, Sebastian W, Gopalakrishnan A (2016b) Population genetic structure of Indian oil sardine, *Sardinella longiceps* along Indian coast. *Gene* 576(1):372-378
36. Sukumaran S, Sebastian W, Gopalakrishnan A (2017) Genetic population structure and historic demography of Indian mackerel, *Rastrelliger kanagurta* from Indian peninsular waters. *Fish Res* 191:1-9
37. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725-2729
38. Teacher AG, Andre C, Merila J, Wheat CW (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evol Biol* 12(1):248
39. Thomas Jr RC, Willette DA, Carpenter KE, Santos MD (2014) Hidden diversity in sardines: genetic and morphological evidence for cryptic species in the goldstripe sardinella, *Sardinella gibbosa* (Bleeker, 1849). *PloS one* 9(1):e84719
40. Whitehead PJP, Nelson GJ, Wongratana T (1988) Clupeoid fishes of the world: An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies, and wolf-herrings. *FAO Fish Synop* 125:305-579

# *Chapter 3*

## CHARACTERISING POPULATION STRUCTURE AND ADAPTIVE VARIATION IN THE INDIAN OIL SARDINE *SARDINELLA LONGICEPS* (Valenciennes, 1847) USING MITOCHONDRIAL GENOME

### ABSTRACT

Tropical Indian Ocean has been warming at an accelerated rate compared to all other tropical oceans contributing to an increase in global mean sea surface temperature (SST). Marine organisms especially small pelagic fishes are vulnerable to the changing climate and it is pertinent to understand the molecular changes that ensure resilience. We investigated the adaptive consequences in the DNA of the most important organelle in bioenergetics, “mitochondrion” for getting insights regarding the spatial and temporal distribution of selective signals which provide clues to its potential for survival and resilience. Indian oil sardines were collected from different eco-regions of the Indian Ocean and analysed for mitogenomic selection patterns by approximate hierarchical Bayesian method (FUBAR, MEME) and TreeSAAP. Non-coding control region was also analysed for selective constraints. Even though, purifying selection was the dominant force influencing mitogenome evolution, signals of diversifying selection were observed in key functional regions involved in OXPHOS (participating in proton translocation, polypeptide binding in inter-chain domain interface and mito-nuclear interactions) indicating OXPHOS gene regulation as the critical factor to meet enhanced energetic demands during uncertain environmental conditions. A characteristic control region with 38-40bp tandem repeat units under strong selective pressure was also observed. These changes were prevalent in the Western Indian Ocean; mainly in fishes from South Eastern Arabian Sea (SEAS) followed by the Northern Arabian Sea (NAS) and rare in the Eastern Indian Ocean or Bay of Bengal (BoB) populations. Significant $\Phi_{ST}$ values were observed in pairwise analyses using whole-genome data set with NAS population as the most genetically differentiated. The selected sites could be used for further investigations by employing them as genetic tags of locally adapted populations for conservation and management as small pelagic fishes contribute to the food security of developing nations. The accelerated substitution rate observed on SEAS has arisen from enhanced mutational rates due to selective pressures contributed by highly variable oceanic environment characterized by seasonal hypoxia, variable SST and food availability.

## 1. Introduction

Small pelagic fishes like sardines and anchovies exhibit the remarkable potential to recover from population crash as exemplified by heavy landings after drought regimes (Alheit *et al.* 2009). The capacity to adapt to the uncertain environmental conditions (hypoxia, temperature and productivity) may be imprinted in the mitochondrial and nuclear genome. Indian oil sardine, *Sardinella longiceps* (Valenciennes, 1847) is distributed across wide environmental clines in the Indian Ocean; mainly North East, South East, South West and North West Indian coast, Gulf of Oman and Gulf of Aden (Munroe and Priede 2010). Temperature is the most important factor followed by salinity and dissolved oxygen availability which explains the seasonal fluctuations in distribution and abundance of small pelagic fishes like sardines and anchovies (at species, sub-species and life stage levels) relative to upwelling fronts (Peck *et al.* 2013; Sato *et al.* 2018). The distribution and abundance are also affected by productivity contributed by availability of nitrogen from outside of their habitat by upwelling and other mixing processes especially runoff from rivers (Checkley *et al.* 2017; Reiss *et al.* 2008). So, fluctuations in dissolved oxygen level, temperature and salinity induce physiological stress in planktivorous fishes like sardines and anchovies in dynamic upwelling systems. The metabolic rate of an animal is directly related to the physiological stress and it has a significant impact on survival and persistence. The role of the mitochondrion and mitochondrial DNA cannot be overemphasized in this scenario as it has been proved as one of the vital organelles determining metabolic and energy efficiency and subsequent adaptation.

Temperature and salinity clines are reported in the Indian Ocean especially between the Arabian Sea on the west and Bay of Bengal on the east (Chatterjee *et al.* 2012). The wide distribution of these sardines may provide them with excellent adaptive capacity to environmental gradients.  Indian oil sardines are characterized by localized extinctions and recolonizations, range expansions and contractions in response to environmental forcing, making it one of the important sentinel species for climate change-related investigations (Xu and Boyce 2009). Studies using neutral markers like microsatellites have provided some clues to sub-structuring within the Indian Ocean (Sebastian *et al.* 2017a). Presence of morphologically divergent ecotypes or phenotypic plasticity has also been implicated (Sukumaran *et al.* 2016) at different locations. But, none of the studies has been addressed to finding signals of adaptation in the mitogenome. A wide sampling of mitogenome carries the promise of finding out population structuring and intraspecific selection patterns in response to environmental clines, in addition to finding out marker loci for subsequent investigations on adaptation.

Tropical Indian ocean has been warming for over a century at a rate which is faster than any other region of tropical oceans which influence the sea surface temperature (SST) patterns globally (Roxy *et al.* 2014). The inhabitants of the tropical Indian Ocean will be under a strong selection pressure to cope up with the enhanced energetic demands due to the increased SST as well the changes in salinity, dissolved oxygen, food availability and hydrological factors. Mitochondrial genome adaptations may provide resilience to these climatic factors by changes in the efficiency of OXPHOS complex which could be monitored over time to understand spatial and temporal patterns in the distribution of some sentinel species like Indian oil sardine. Further, conservation and management strategies can be devised to protect or conserve the adapted populations which will ensure food security of the nations as small pelagic contribute substantially to the food security of developing countries.

Empirical evidence for correlations between mitochondrial DNA evolution and mtDNA content with climatic adaptations has been found in recent investigations (Ruiz-Pesini *et al*. 2004; Cheng *et al.* 2013; Lajbner *et al.* 2018). Environmental gradients induce substantial selective pressure on the mitogenomes due to its role in cellular respiration and metabolism in addition to indirect selection due to cytonuclear co-evolution (Ballard and Pichaud, 2014; Morales *et al.* 2016). Absence of recombination also paves the way for selective sweeps (Meiklejohn *et al.* 2007). Thus, mitochondrial OXPHOS complex has been implicated as the vital force in regulating metabolic rate and subsequent adaptation to different thermal and salinity regimes (Garvin *et al.* 2015a).

Mitochondria play important roles in the bioenergetics of tissues by producing 95% of eukaryotic cell energy (ATP) through the process of oxidative phosphorylation. There are five major protein complexes involved in OXPHOS, membrane protein complexes I, II, III, IV and V which are encoded by both nuclear (~88 genes) and mitochondrial (13 genes) genomes whilst complex II is encoded only by the nuclear genome. The respirasome (complex I, III and IV) uses the energy released during electron transfer from NADH to $O_2$ for proton translocation to the intermembrane space and generate a proton gradient across the inner mitochondrial membrane. The ATP synthase (complex V) use the proton motive force generated to synthesize ATP chemi-osmotically (Letts *et al.* 2016). Efficient ATP synthesis is made possible by maintaining the integrity of interactions between mitochondrial and nuclear-encoded subunits of OXPHOS (Lowell and Spiegelman, 2000) as a minor change can influence multiple levels of a biological organization like cellular function, the fitness of organism and ecosystem processes (Latorre-Pellicer *et al.* 2016). Maintaining the optimal mito-nuclear association in OXPHOS system is pivotal as mismatches produce negative effects such as reduced lifespan, fecundity, reduced metabolic rate and diseases (Dowling *et al.* 2008, Gershoni *et al.* 2014, Mossman *et al.* 2016). Small pelagic fishes like sardines and anchovies exhibit the remarkable potential to recover from population crashes as exemplified by heavy landings after drought regimes (Alheit *et al.* 2009). The capacity to adapt to the uncertain life may be imprinted in the mitochondrial and nuclear genome.

Evolutionary studies on mitogenomes of fishes like Pacific salmon revealed that key adaptations in OXPHOS proteins are important in lineage sorting (Garvin *et al.* 2011). Studies on white fish (*Coregonus* spp.) found evidence for relaxed purifying selection in NADH2 gene as a cause of the high rate of non-synonymous mutations (Jacobsen *et al.* 2016). Positively selected sites were detected in cytochrome b region of a widely distributed European anchovy and those sites were correlated with thermal clines (Silva *et al.* 2014). Mutations in OXPHOS genes have been correlated with a wide range of environmental factors like hypoxia (Scott *et al.* 2010; Ekau *et al*. 2010), heat stress (Morales *et al.* 2015), cold stress (Stier *et al.* 2014), nutrient availability (Da Fonseca *et al.* 2008) and the difference in expression of genes (Garvin *et al.* 2015b). Such mutations in human beings have been related to diseases in Human (Gershoni *et al.* 2014), adaptation to different thermal regimes in Drosophila (Doi *et al.* 1999) and differing aerobic capacity in Killifish (Brennan *et al.* 2016). All these studies emphasize the importance of identifying the loci involved in selection as these loci could be used as markers to study environmental adaptation.

The non-coding content present in the mtDNA is known as the control region, which is responsible for the regulation of replication and transcription of mitogenome (Pereira *et al.* 2008). But the exact functions of the control region are not clear. However, the availability of large mitogenome data helped to identify many conserved sequence elements/domains, presence of binding sites for nuclear-encoded factors, replication initiation sites, transcription initiation sites and termination associated sites (Miya and Nishida 2015). But the presence of highly variable sites without any functional elements and heteroplasmy in the control region is still not clearly explained. Intra-strand secondary structures have been identified as recognition sites/binding sites for many regulatory proteins like transcriptional factors (Walberg and Clayton 1981; Katz and Burge 2003; Pereira *et al.* 2008). There is enough evidence that the basic molecular processes like replication, transcription and recombination are controlled/regulated by formation of intra-strand secondary structures by nucleic acids (DNA/RNA) (Pereira *et al.* 2008). There are reports that many control region segments can form stable intra-sequence secondary structures (Katz and Burge 2003; Pereira *et al.* 2008).

In the present study, we characterized 45 complete mitogenomes along with 350 complete mitochondrial control regions of Indian oil sardines from its range of distribution in the Indian ocean mainly; Eastern Indian ocean (Bay of Bengal) and Western Indian ocean (South East Arabian sea and North Arabian Sea). Subsequently, we investigated signals of positive/purifying selection if any correlating with geographical distribution. We also analyse control region sequences and predicted the secondary structure formed by them to understand the most important factor in the evolutionary dynamics of *S. longiceps* mitogenome and its relation to habitat characteristics in the Indian Ocean.

## 2. Materials And Methods

### 2.1. Sample collection, DNA extraction, mitogenome sequencing and assembly

Samples of Indian oil sardines were collected from the three eco-regions mainly, Northern Arabian Sea (NAS), South Eastern Arabian Sea (SEAS) and Bay of Bengal (BoB). A total of 350 individuals were collected (Fig 3.1) during 2015-2017 and DNA extracted. The mitochondrial genome from 45 individuals was amplified as overlapping segments using 16 novel primer pairs (Table 3.T1) designed with *S. longiceps* mitogenome as the template (GenBank Accession No: KR000002.1) and sequenced. Sequences were manually checked, aligned and assembled in MEGA6 (Tamura *et al.* 2013) and Geneious R7 (Kearse *et al.* 2012) against *S. longiceps* mitogenome (Sebastian *et al.* 2017b). Control region of an additional 305 individuals were sequenced and assembled. Three types of sequence datasets were prepared and analysed using MEGA6 and Geneious R7 (Kearse *et al.* 2012); whole mitogenome nucleotide sequence; Nucleotide and amino acid sequences of 13 individual genes, 22 tRNAs and control region as separate data sets; all coding gene concatenated nucleotide and amino acid datasets. Nucleotide sequences of overlapping genes were duplicated and the reverse complement of the ND6 gene sequence was used to get the correct amino acid sequence.

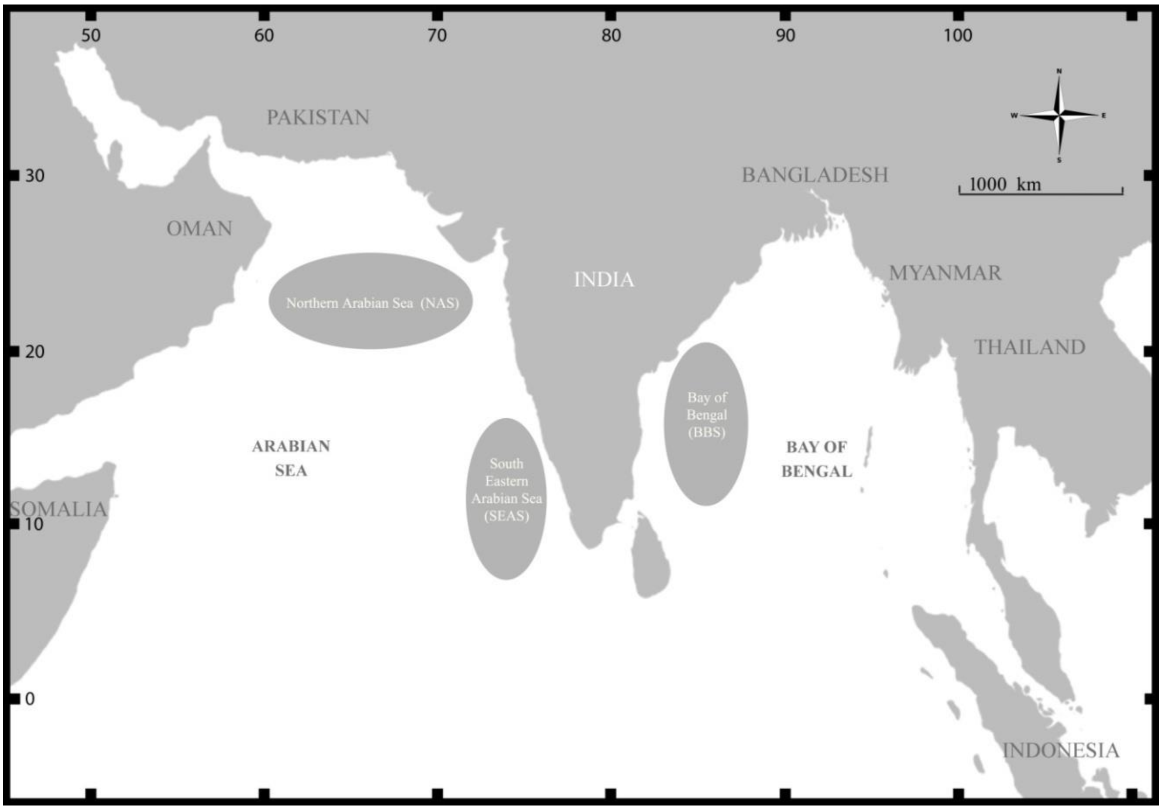


**Fig 3.1** Map showing sampling locations of *S. longiceps* population. The direction of North and an approximate scale are also shown. Sample Site: NAS (Northern Arabian Sea), SEAS (South Eastern Arabian Sea) and BOB (Bay of Bengal)

## 2.2 Population genetic analysis

Descriptive statistics, the number of polymorphic sites (*S*), nucleotide diversity (π) (Nei 1987), haplotype diversity ($H_d$) (Nei 1987), the average number of pairwise nucleotide differences (*K*) (Tajima 1983), the total number of synonymous and non-synonymous mutations, for whole mitogenome and all protein-coding gene concatenated data set were calculated using DnaSP (Librado and Rozas 2009). Harpending raggedness index (Hri) (Harpending 1994), Tajima's *D* (Tajima 1989), Fu's *Fs* (Fu and Li 1993), and *D** were calculated using DnaSP to check deviations from neutrality.

Nucleotide diversity of individual genes, all genes concatenated data sets, control region, whole-genome sequence data sets along with amino acid diversity of individual genes and all gene concatenated amino acid sequence data sets were calculated using MEGA6.

The proportion of variance distributed among population samples was analyzed using the hierarchical analysis of molecular variance procedure (AMOVA) in Arlequin (Excoffier and Lischer 2010). The AMOVA analysis was performed for whole-genome, individual gene and control region nucleotide sequence sets using Arlequin with 10000 permutations. Arlequin was also used to estimate F statistics, pairwise $\Theta_{ST}$ for whole-genome nucleotide sequence.

The number of non-synonymous substitutions per non-synonymous site ($K_a$), number of synonymous substitutions per synonymous site ($K_s$), $K_a/K_s$ and theta ($\theta$) were calculated using MEGA6 and DnaSP. The $\theta$-values were used to calculate the relative mutation rate of individual genes relative to the whole mitogenome using the equation $\mu_{gene}$ = $((\mu_{mitogenome} * \theta_{mitogenome})/\theta_{gene})$ (Jacobsen *et al.* 2016).

Maximum likelihood tree was generated using MEGA6 for whole-genome and all gene concatenated nucleotides sequence data with 1000 bootstrap replicates and GTR substitution model (selected using the J Model Test). *Sardinella maderensis* was used as an outgroup to root both trees. A neighbour-joining tree was also produced for above sequence sets in MEGA6, with mean nucleotide distance. Whole-genome sequences of 45 samples were used to generate haplotype network with the median-joining method in popART (Bandelt *et al.* 1999) software.

## 2.3. Selection analyses

All gene concatenated nucleotide sequence data with Maximum likelihood tree generated from it was used to conduct a whole-genome scan to detect signals of natural selection in mtDNA coding genes. We analysed data with the approximate hierarchical Bayesian method (FUBAR- Fast Unconstrained Bayesian Approximation) and mixed effect method (MEME- Mixed Effect Model of Evolution). These programmes are available in DATAMONKEY (Pond and Frost 2005). MEME model analyses the distribution of synonymous and non-synonymous substitution rates from site to sites and branch to branch at a site. But FUBAR is considered as more dependable when the strength of selection varies across sites because it uses settings which are less sensitive to model specifications. For each method we selected a threshold *P*-value; $P < 0.05$ for MEME and posterior probability > 0.9 for FUBAR. We used TreeSAAP (Woolley *et al.* 2003) to understand changes in physicochemical properties of amino acids caused by replacements, as it compares the amino acid changes inferred from a given tree with a model having 31 predicted physicochemical amino acid property changes, under an assumption of neutrality. Z test was used to analyse the changes in the amino acid properties, which is categorized into eight magnitude groups. The positive and negative Z-scores indicate positive and negative selection respectively. In this analysis, we considered only 6, 7 and 8th category amino acid changes with strong statistical support ($P< 0.001$).

All coding gene amino acid sequence datasets were aligned in MEGA6. 3D homology model of protein subunits with positively selected sites observed was constructed with SWISS-MODEL server (Schwede *et al.* 2003) using the vertebrate protein model of Bovine corresponding to each subunit (available in PDB). Finally, we located positively selected sites identified in the three-dimensional structure of protein subunit and compared it with the functionally important amino acid residues. The number of positively selected sites was compared between eco-regions to correlate it with habitat characteristics and find out signals of local adaptation if any.

## 2.4. Control region sequence analysis and secondary structure prediction

The control region DNA sequence of *S. longiceps* and other clupeoids were assembled in MEGA7. We used ‘mfold’ web server (Zuker 2003) for DNA (with 15 window length and 25 step size) secondary structure prediction by free energy minimization method with nearest neighbour thermodynamic rules. Similarly, the structure and free energy for RNA sequence of control region types and tRNA were calculated using RNA mfold in ‘mfold’ web server. To test the selection effect on control region we calculated Tajima’s D and relative mutation rate in control region data with DnaSP and MEGA7 respectively. We calculated the Tajima’s D statistics for whole mtDNA and region of ~1112bp comprising tRNA pro, control region and tRNA phe with 10bp intervals overlapping at 5bp. To test the functional importance of the formation of secondary structure we compared the conservation status of the sequence forming secondary structures in terms of relative mutation rate and polymorphism. The inter-specific identity was analysed by comparing their sequence with the available control region sequences of fishes belonging to Clupeoidei.

## 2.4. Environmental data

Monthly climatology data of Sea surface Temperature SST (°C) SSS (ppt) and Dissolved Oxygen DO (µmol/kg) was taken from World Ocean Data 2018 available at https://www.nodc.noaa.gov/OC5/woa18/woa18data.html. While monthly average Chlorophyll *a* (mg/m$^3$) data spanning from the year 2002 to 2015 was downloaded from MODIS site (https://modis.gsfc.nasa.gov/data/dataprod/chlor_a.php) and subjected to objective analysis before generating monthly climatology. The data was analysed by using Ferret and visualized in Ocean Data View (ODV 5.1.7, available at https://odv.awi.de/). Seasonal climatology data (Winter - (January, February, March), Spring - (April, May, June), Summer - (July, August, September), Fall - (October, November, December)) for ecoregions were prepared by estimating the mean and standard deviation of annual SST, SSS, DO and Chlorophyll-*a* were estimated as a measure of degree variability in annual climatology. We used generalized linear models in R 3.6.2 with a binomial link to examine variations in the frequencies of amino acid substitutions under selection (as described by Consuegra *et al.* 2015) in NAS, SEAS and BoB with SST (Winter, Spring, Summer and Fall), SSS (Winter, Spring, Summer and Fall), DO (Spring, Summer and Fall), Chlorophyll a (Spring, Summer and Fall), the standard deviation of annual SST (fluctuations in annual SST), the standard deviation of annual SSS (fluctuations in annual SSS), the standard deviation of annual DO (fluctuations in annual DO) and the standard deviation of annual Chlorophyll-*a* (fluctuations in Chlorophyll-*a*).

## 3. Results

### 3.1. Mitogenome sequencing and assembly

MtDNA of 45 individuals were completely sequenced, assembled and annotated. Size of the mitogenomes ranged from 16598 to 16676bp depending on the size variation in the control region. No identical sequences were found. Annotated mitogenomes have been submitted to NCBI, GenBank (Accession numbers MG251937–MG251981). Maximum likelihood and Neighbour-joining trees of whole-genome and all gene concatenated nucleotide sequence data sets revealed different clades even though the bootstrap support was negligible. There were no detectable geographical patterns in clustering in the phylogenetic tree as well as in haplotype network diagram (Fig 3.S2; 3.S3).

### 3.2. Population genetic structure

Descriptive statistics of the entire mitogenome and concatenated protein-coding gene data set are given in Table 3.1. The level of nucleotide diversity was low for the whole mitogenome (ranging from 0.0060 to 0.00132) with only 1131 segregating sites whereas haplotype diversity was high with each genome representing a unique haplotype (45) as evident in haplotype network (Fig 3.S3.). The significant negative Fu's $F_S$ and Tajima's $D$ for whole-genome (-8.642 and -2.319) and concatenated protein-coding gene data set (-11.318 and -2.370) (Table 3.1) indicated an excess of rare nucleotide variants and rare haplotypes respectively compared to that what would be expected under neutrality (Harpending, 1994). The values of Fu's $F_S$, Fu & Li's $F^*$ and Fu's & Li's $D^*$ are -8.642 ($P < 0.0001$), -3.59983 ($P < 0.02$) and -3.75132 ($P < 0.02$) respectively (Table 3.1).

The mismatch analysis using complete genome and concatenated protein-coding gene data sets showed a multimodal pattern of distribution with low and non-significant raggedness index (r = 0.00037 and 0.0019) under the demographic expansion model (Figure 3.S1). A change in the population size, growth or declines will create a distinct pattern in the distribution of pairwise nucleotide differences. A unimodal pattern will be present in populations with a population expansion after a bottleneck and a multimodal pattern in equilibrium populations (Rogers and Harpending 1992). Non-significant raggedness index (p = 0.68) indicates that the data is relatively good fit to a model of population expansion (Harpending 1994). The significant negative Fu's $F_S$ and Tajima's D for the whole genome (-8.642 and -2.319) and concatenated protein-coding gene data set (-11.318and -2.370) (Table 3.1) indicated an excess of rare nucleotide site variants and excess of rare haplotypes respectively compared to that what would be expected under neutrality. Deviations from neutrality in the present study may be due to population expansion or positive selection (Tajima 1989).

Relative mutation rates calculated for different gene regions indicated ND4 gene and ATP8 as high evolving with the highest number of non-synonymous mutations (Ka/Ks (0.25)). No non-synonymous substitutions were observed ND4L and ND6 genes (Table 3.1). Relative mutation rate (π) varied between genes. Control region, ND genes and ATPase evolved faster than other regions. COX evolved slower than ND genes. tRNAs and 12S rRNA were the slowest evolving genes (Table 3.1).

Global $\Theta_{ST}$ values of 0.10359 ($p<0.001$) in whole genome dataset and 0.11387 ($p < 0.001$) in all gene concatenated data sets were obtained in AMOVA analysis (Table 3.S2). Global $\Theta_{ST}$ values ranged from 0.00956 (ND6) to 0.21919 (ND1) when individual gene and control data sets were analysed (Table 3.S2). After sequential Bonferroni correction, values corresponding to ATP6, CO1, CO3, CYTB, ND1, ND2, ND4 and ND5 were significant. Significant $\Theta_{ST}$ values were observed in pairwise analyses using whole genome data set with OMAN population as the most genetically differentiated (Table 3.S3). Similar results were obtained with the whole gene concatenated nucleotide data sets, but with other gene data sets, significant values were very less.

**Table 3.1** Summary of descriptive genetic diversity statistics of entire mitogenome and concatenated protein-coding genes of *S. longiceps* mitochondrial genome.

| | ***S*** | ***π*** | **No of haplotype** | ***H*d** | ***K*** | **Number of Synonymous sites** | **Number of Non-synonymous sites** | ***K*s** | ***K*a** | ***K*a/*K*s** | Θ | **μ relative** | **Tajima's D** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Genome | 1131 | 0.00606 | 45 | 1 (0.005) | 100.628 | - | - | - | - | - | 0.00611 | 1.5 | -2.31923 (*P*< 0.01) |
| Gene concatenated | 859 | 0.00682 | 45 | 1.00 (0.005) | 77,901 | 748 | 136 | 0.023 | 0.001 | 0.043 | 0.00688 | 1.12 | -2.37011 (*P*< 0.01) |
| ATP6 | 47 | 0.00599 | 32 | 0.940 (0.029) | 4.094 | 38 | 13 | 0.015 | 0.003 | 0.2 | 0.00604 | 0.9 | -2.4994 (*P*< 0.01) |
| ATP8 | 4 | 0.00132 | 5 | 0.211 (0.080) | 0.22 | 2 | 2 | 0.004 | 0.001 | 0.25 | 0.00132 | 0.22 | -1.76368 (0.1 >*P* > 0.05) |
| CO1 | 78 | 0.0039 | 43 | 0.998 (0.005) | 6.054 | 66 | 13 | 0.013 | 0.001 | 0.077 | 0.00392 | 0.64 | -2.46843 (*P* < 0.01) |
| CO2 | 54 | 0.00558 | 24 | 0.824 (0.059) | 3.859 | 32 | 23 | 0.014 | 0.003 | 0.214 | 0.00562 | 0.92 | -2.5591 (*P* < 0.001) |
| CO3 | 32 | 0.00455 | 25 | 0.001 (0.032) | 3.569 | 29 | 3 | 0.014 | 0.001 | 0.071 | 0.00457 | 0.75 | -2.1433 (*P*< 0.05) |
| Control region | 107 | 0.01597 | 43 | 0.998 (0.005) | 15.115 | - | - | - | - | - | 0.01631 | 2.66 | -1.5376 (*P*< 0.10) |
| CYTB | 92 | 0.00656 | 39 | 0.984 (0.013) | 7.489 | 79 | 16 | 0.022 | 0.001 | 0.045 | 0.00662 | 1.08 | -2.357 (0.1 >*P* > 0.05) |
| ND1 | 87 | 0.00995 | 34 | 0.987 (0.007) | 9.7 | 84 | 5 | 0.037 | 0.001 | 0.027 | 0.01008 | 1.64 | -1.9555 (*P* < 0.05) |
| ND2 | 100 | 0.00913 | 38 | 0.993 (0.006) | 9.544 | 100 | 4 | 0.033 | 0.001 | 0.03 | 0.00925 | 1.51 | -2.2421 (*P*< 0.01) |
| ND3 | 21 | 0.00512 | 21 | 0.841 (0.045) | 1.781 | 18 | 4 | 0.018 | 0.001 | 0.056 | 0.00515 | 0.84 | -2.1182 (*P*< 0.05) |
| ND4 | 143 | 0.00849 | 42 | 0.997 (0.005) | 11.719 | 117 | 29 | 0.029 | 0.001 | 0.034 | 0.00858 | 1.4 | -2.3601 (*P*< 0.01) |
| ND4L | 7 | 0.00133 | 8 | 0.362 (0.092) | 0.396 | 7 | 0 | 0.005 | 0 | - | 0.00134 | 0.22 | -2.0336 (*P*< 0.05) |
| ND5 | 170 | 0.00887 | 44 | 0.999 (0.005) | 16.293 | 152 | 24 | 0.031 | 0.002 | 0.065 | 0.00898 | 1.46 | -2.3059 (*P*< 0.01) |
| ND6 | 24 | 0.0061 | 28 | 0.904 (0.040) | 3.184 | 24 | 0 | 0.022 | 0 | - | 0.0061 | 0.99 | -1.7519 (0.1 >*P* > 0.05) |
| 12S rRNA | 11 | 0.00265 | 14 | 0.612<br>-0.085 | 1.046 | - | - | - | - | - | 0.0011 | 0.18 | -1.7332<br>(0.1 >*P* > 0.05) |
| 16S rRNA | 74 | 0.00288 | 34 | 0.967<br>-0.019 | 4.857 | - | - | - | - | - | 0.00289 | 0.47 | -2.572<br>(*P* < 0.001) |
| tRNAs | 27 | 0.00117 | 20 | 0.761<br>-0.069 | 1.701 | - | - | - | - | - | 0.00117 | 0.19 | -2.425<br>(*P* < 0.01) |
| | Whole mitogenome nucleotide sequence | | | | | | | Gene concatenated (13 protein-coding genes) | | | | | |
| **Fu's *Fs*** | -8.642 (*P*< 0.0001) | | | | | | | -11.318 (*P*< 0.0001) | | | | | |
| **Fu and Li's *D**** | -3.59983 (*P*< 0.02) | | | | | | | -3.59983 (*P*< 0.02) | | | | | |
| **Fu and Li's *F**** | -3.75132 (*P*< 0.02) | | | | | | | -3.75132 (*P*< 0.02) | | | | | |

S = number of polymorphic sites, ***π* =** nucleotide diversity,*H*d = haplotype diversity, *K* = average number of pairwise nucleotide differences, *K*s =number of synonymous substitutions per synonymous site,*K*a = Number of non-synonymous substitutions per non-synonymous site, Θ = theta from Sandμ = mutation rate.

### 3.3. Evidence for natural selection and adaptive evolution

Relative mutation rates (µ relative) calculated for different gene regions indicated ND4 (1.4) and ND5 (1.46) genes as high evolving with the highest number of non-synonymous mutations (29 and 24 respectively) (Table 3.1). No non-synonymous substitutions were observed in ND4L and ND6 genes. Relative mutation rate varied between genes. Control region, ND genes and ATPase evolved faster than other regions. COX evolved slower than ND genes. tRNAs and 12S rRNAs were the slowest evolving genes (Table 3.1).

Signals of significant selection were evident in many codons among the 3798 codons analysed. FUBAR analysis identified 680 and 10 sites as under pervasive purifying and diversifying selection respectively. Signatures of positive selection were less prevalent than purifying selection and they were concentrated in Complex I (ND1, ND2, ND4 and ND5), Complex III (CYT B), Complex IV (C01, CO2 and CO3) and Complex V (ATP6). MEME analysis showed that there are 26 sites under episodic diversifying selection ($P$< 0.05). The purifying selection has been identified as the dominant force in the mitogenome of *S. longiceps*. TreeSAAP analysis detected many significant amino acid physiochemical property changes in the positively selected regions of *S. longiceps* mitogenome, with conservative amino acid changes dominating over radical changes. Among this, only those sites identified as positively selected, at least by two methods were selected for further discussion (Table 3.2).

**Table 3.2** Codons that are under positive selection in the mitogenome protein-coding genes of *S. longiceps.* The analysis is based on three selection tests: MEME, FUBAR and TreeSAAP method.

| **Gene** | Amino acid position | From Codon To Codon | From Amino acid To Amino acid | **MEME**[a] | **FUBAR**[b] | **TreeSAAP** | Distribution of amino acid replacement across the population |
|---|---|---|---|---|---|---|---|
| | | | | *p*-value | Posterior Probability | Significant properties (category of amino acid changes) | |
| **ND1** | 29 | ATT-TTT | Ile-Phe | 0.0224 | - | - | SEAS |
| **ND1** | 30 | GAG-TTG | Glu-Leu | 0.0006 | - | Average number of surrounding residues (7)<br>Chromatographic index (8)<br>Hydropathy (8)<br>Surrounding hydrophobicity (7) | SEAS |
| **ND2** | 302 | CTT- CAA | Leu-Gln | 0.0061 | - | Polarity (7) | SEAS, NAS |
| **C01** | 25 | CTG- CGA | Leu-Arg | 0.011 | - | Isoelectric point (6)<br>Polarity (7) | SEAS |
| **C01** | 114 | GGC- GCC | Gly-Ala | 0.034 | 0.9101 | - | SEAS, NAS |
| **C01** | 262 | AAT- GAT | Asn-Asp | 0.0435 | 0.9062 | - | SEAS, NAS, BOB |
| **C02** | 50 | CTT- CAA | Leu-Gln | 0.0005 | - | Polarity (7) | NAS |
| **C02** | 63 | GAA-GGA | Glu-Gly | 0.0138 | - | Compressibility (7) | BOB |
| **C02** | 152 | GTT- TCT, TCC | Val-Ser | 0.0006 | - | - | SEAS |
| **ATP6** | 114 | GTA- CTA, CTC, GCA | Val-Leu,Ala | 0.0345 | - | - | SEAS, BOB |
| **ATP6** | 185 | ATT- CAA | Ile-Gln | 0.0413 | - | - | SEAS |
| **C0 3** | 16 | TGA- GGA, TTA, CGA | Trp-Gly, Leu, Arg | 0.0429 | 0.9897 | - | SEAS, BOB |
| **C0 3** | 117 | CCA- TTA, TCT | Pro-Leu,Ser | 0.0374 | 0.9769 | - | SEAS, BOB |
| **ND 4** | 148 | ACC- AAC | Thr-Asn | 0.0208 | - | - | SEAS |
| **ND5** | 9 | TCT- TGA, TAT | Ser-Trp, Tyr | 0.0019 | - | - | NAS |
| **ND5** | 97 | GCC- GGG | Ala-Gly | 0.0015 | - | - | SEAS |
| **ND5** | 98 | CTT- GTT | Leu-Val | 0.0469 | - | - | SEAS |
| **ND5** | 225 | GCC- ACC | Ala-Thr | 0.038 | 0.9055 | - | SEAS, BOB |
| **ND5** | 226 | ACG- ACT | Thr-Asn | 0.0016 | | - | SEAS, NAS |
| **ND5** | 227 | GCC- TGC | Gly-Cys | 0.0423 | 0.9809 | Refractive index (7) | SEAS, BOB |
| **ND5** | 228 | AAA- AAT | Lys-Asn | 0.035 | 0.9745 | Isoelectric point (6) | SEAS, BOB |
| **ND5** | 236 | CCC- TCC, TTT | Pro-Ser,Phe | 0.0061 | | | SEAS, BOB |
| **CYTB** | 70 | TGC- TAC, GTC | Cys-Tyr, Trp | 0.504 | 0.9507 | Chromatographic index (8)<br>Helical contact area (7)<br>Molecular volume (6)<br>Partial specific volume (7) | SEAS, NAS, BOB |
| **CYTB** | 250 | CTA- CAA | Leu-Gln | 0.0493 | | - | NAS |
| **CYTB** | 311 | AAG- CAG | Lys-Gln | 0.0471 | 0.9739 | - | SEAS, NAS |
| **CYTB** | 320 | CTT- ATT | Leu-Ile | 0.0439 | 0.9176 | - | SEAS, NAS, BOB |

MEME - Mixed Effect Model of Evolution, FUBAR - Fast Unconstrained Bayesian Approximation, NAS (Northern Arabian Sea), SEAS (South Eastern Arabian Sea) and BOB (Bay of Bengal).

Sites 29ND1, 30ND1, 302ND2, 148ND4, 9ND5, 97ND5, 98ND5, 225ND5, 226ND5, 227ND5, 228ND5 and 236ND5 were identified as positively selected in Mitochondrial complex I (NADH: ubiquinone oxidoreductase) of *S. longiceps* and all of them were located in transmembrane helices except one which is in the intra-helix loop (228ND5) (Fig 3.2.). Nine of these sites, one in ND2 (#302Leu-Gln) located in C-terminus, one in ND4 (#148Thr-asn) in proton-conducting membrane transporter (Proton_antipo_M) and seven in ND5 (#97Ala-Gly, #98Leu-Val, #225Ala-Thr, #226Thr-Asn, #227Gly-Cys,

#228Lys-Asn & #236Pro-Ser) clustered in Proton_antipo_M & N-terminal (Proton_antipo_N). Position 228 (ND5) showed overlap with amino acid residue that has been reported as one of the key residues in proton translocation (Zhu *et al.* 2016). Asparagine is more polar than Lysine and usually participates in hydrogen bonds as proton donors or acceptors.

Three sites (#25Leu-Arg, #114 Gly-Ala and #262Asn-asp) observed under positive selection in CO1 were located in the transmembrane helix and two of these positions (#25 & #114) showed overlap with amino acid residues that have been reported to participate in polypeptide binding at Subunit I/VIIc interface & Subunit I/VIIa interface respectively. Among three sites observed under positive selection in CO2 gene, amino acid position 50 (Leu-Gln) reside in the intra-helix loop, position 63 (Glu-Gly) in transmembrane helix and 152 (Val-ser) in Beta strand. Among the two sites identified in CO3, position 16 (Trp-Gly) were located in transmembrane helix and position 117 (Pro-Leu, Ser) in the intra-helix loop (Fig 3.3.).

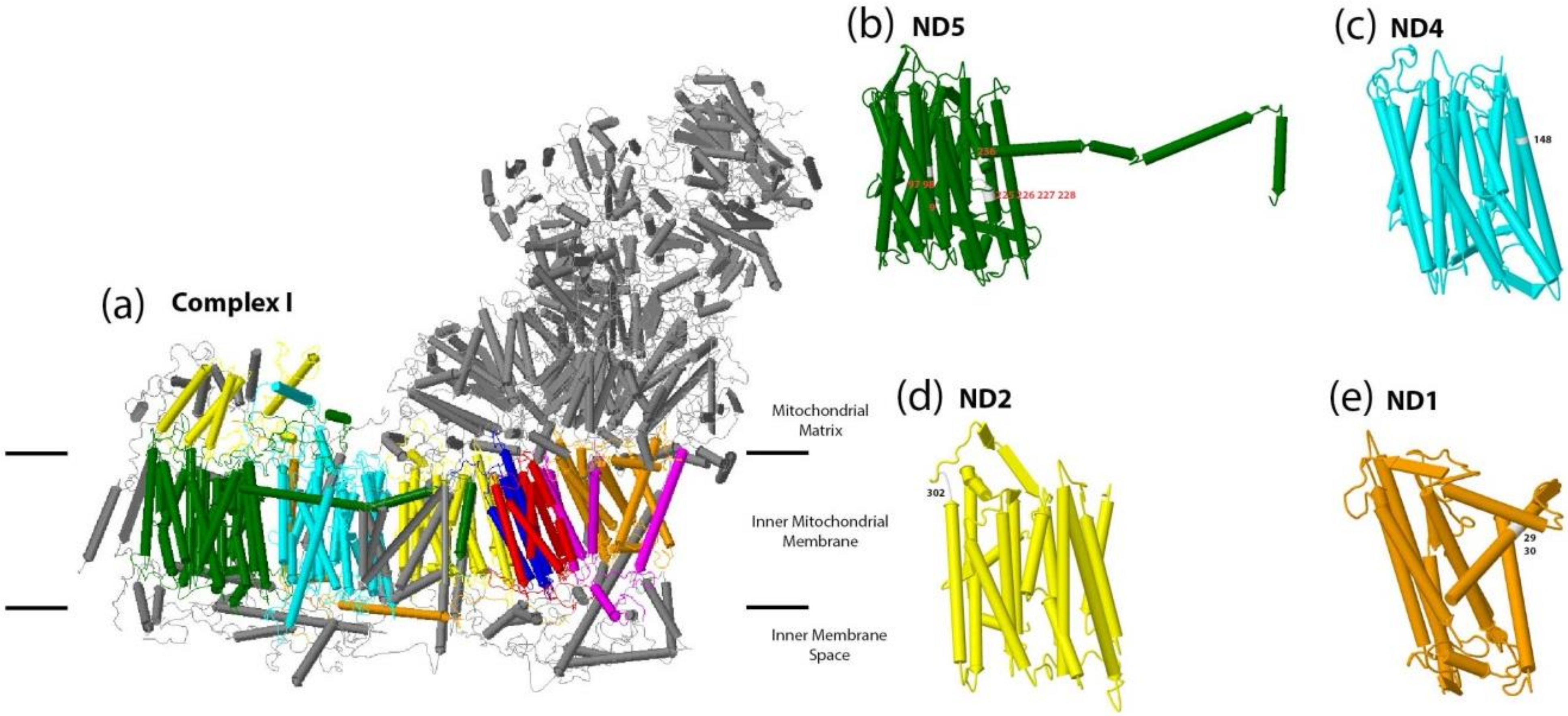


**Fig 3.2 Spatial distribution of positively selected sites identified in NADH dehydrogenase (Complex I) of *S. longiceps.*** Grey structures represent nuclear-encoded subunits. (a) individual OXPHOS Complex I, with mitochondrial-encoded subunits are represented in different coloured as followed: ND2 in yellow; ND4L in blue; ND1 in orange; ND3 in magenta; ND4 in cyan; ND5 in green; ND6 in red. Individual core subunits (b) ND5, (c) ND4, (d) ND2, (e) ND1with amino acid site number on positively selected sites.

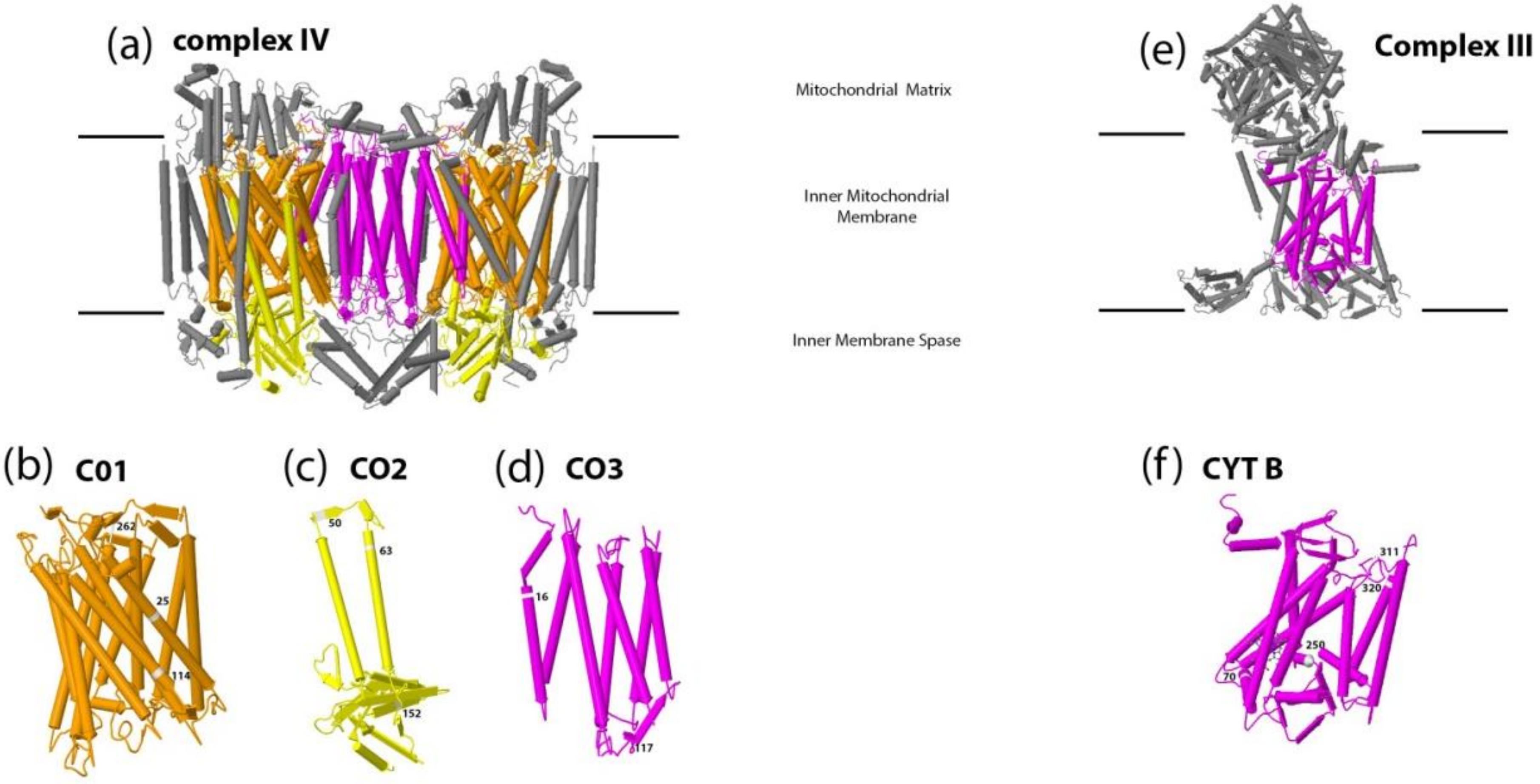


**Fig 3.3 Spatial distribution of positively selected sites identified in Cytochrome C Oxidase (Complex IV) and Cytochrome bc 1 (Complex III) of *S. longiceps.*** Grey structures represent nuclear-encoded subunits. Individual OXPHOS Complex IV (Homodimer) (a) with mitochondrial-encoded subunits is represented in different colours as followed: CO1 in orange; CO2 in yellow; CO3 in magenta. Individual OXPHOS Complex III (e), with mitochondrial-encoded subunit represented in magenta colour. Individual core subunits (b) CO1, (c) CO2, (d) CO3, (f) CYT B with amino acid site number at positively selected sites.

Among four sites (#70Cys-Trp, #250Leu-Gln, #311 Lys-Gln and #320Leu-Ile) that experienced positive selection in CYTB, one (#311) showed overlap with amino acid residue that has been reported to participate in polypeptide binding in inter-chain domain interface and it was located in the transmembrane helix (Fig 3.3.). Among two sites (#114 Val-Cys, ala #185 Ile-Gln) observed under positive selection in ATP6, one (#114) was located in the transmembrane helix-4 and other (#185) in the intra-helix loop connecting helix-5 and 6 (Fig 3.S7.).

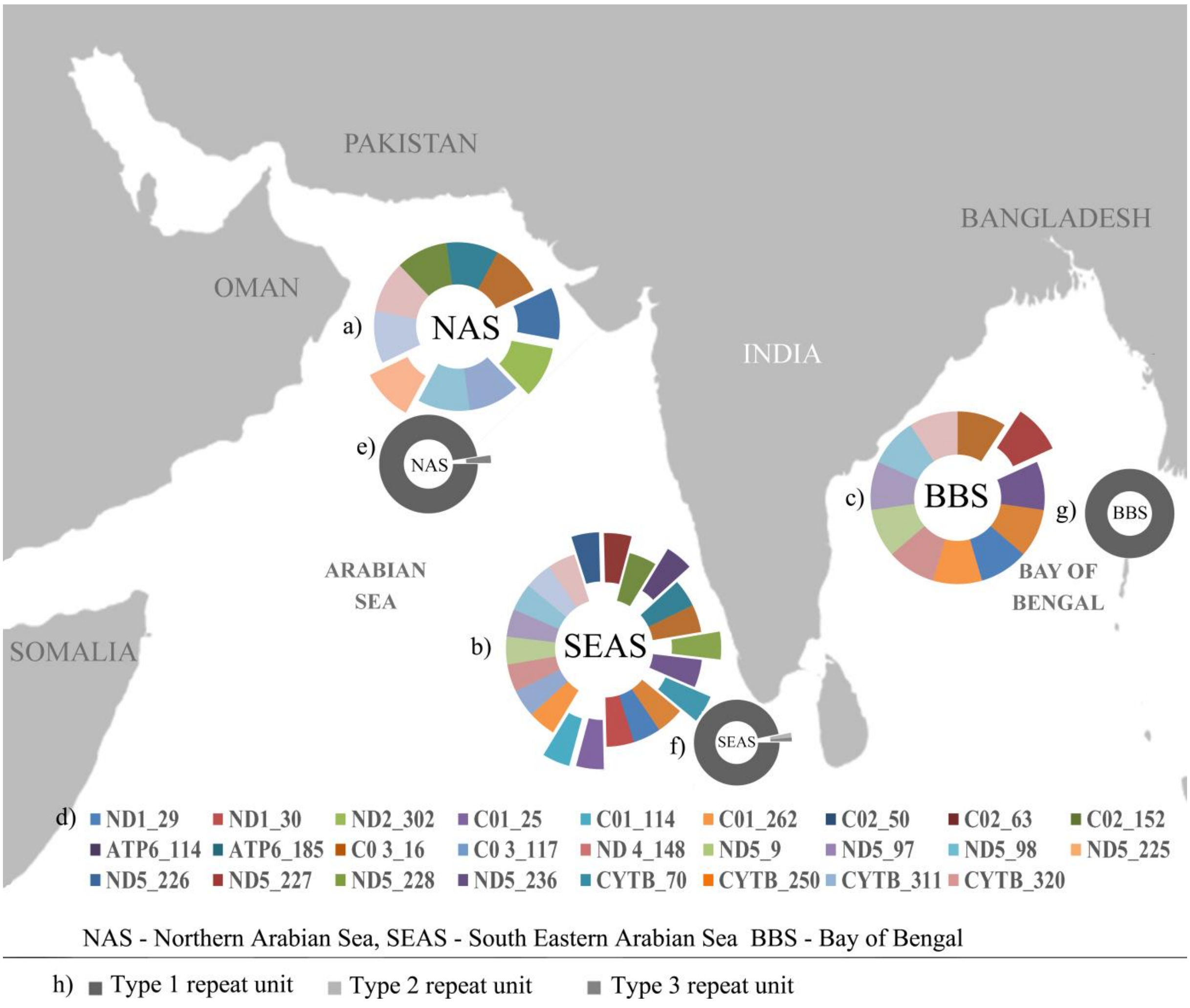


**Fig 3.4 Graphical representation of the geographical distribution of positively selected sites and Control region repeat unit types in the mitogenome of *S. longiceps* in the 3 eco-regions of the Indian Ocean.** a) Frequency of positively selected sites in the northern Arabian sea, b) Frequency of positively selected sites in the south-east Arabian sea, c) Frequency of positively selected sites in Bay of Bengal Ocean, d) Codons that are under positive selection in the mitogenome protein-coding genes, e) Frequency of haplotype with Type 1 repeat unit in the northern Arabian sea, f) Frequency of haplotype with Type 2 repeat unit in the south-east Arabian sea, g) Frequency of haplotype with repeat unit Type 3 in the Bay of Bengal, and h) Haplotype with repeat unit Type 1, 2 and 3.

Individuals with positively selected sites were prevalent in SEAS samples. Among the 26 sites with signals of positive selection, 21, 10 and 9 sites were recorded in individuals from SEAS, NAS and BOB respectively. 8 positively selected sites (Two each in ND1 and ND5 genes and one each in ND4, CO1 and CO2 respectively) were specific to SEAS and one site each in CO2 and CYTB specific to NAS populations (Table 3.2, Fig 3.4) indicating the presence of locally adapted genotypes.

## 3.4. Structure and content of mtDNA control region

The length of the control regions of *S. longiceps* range from 900bp - 980bp (GenBank Accession No: KJ466087–KJ466091; KJ472113–KJ472120; KJ888156–KJ888390; KP000859–KP000897), due to the variation in the number of tandem repeats and a poly-A in different haplotypes (Fig 3.S4.). The control region contains different conserved sequence regions like Termination Associated Sequence (TAS) at 3' end and Conserved Sequence Box (CSB D, CSB1, CSB2 and CSB3). (Sebastian *et al.* 2017a; Fig 3.5.). The tandem repeat sequence of 38-40bp was found in between TAS and poly-A, the repeat units were repeated once (in the majority of haplotypes), twice and three times (in few haplotypes). The repeated unit was 38-40bp in length and at the downstream of the repeat unit, there is a 14bp that have similarity to the 5' end of the repeat unit. But the upstream of the repeat unit is similar to 3' end of the repeat unit. Among the 305 control regions analysed, 259 haplotypes and haplotypes with 3 types of repeats were found. Type 1 with one repeat unit (38bp) were the most abundant, Type 2 (38bp) with two repeat units and Type 3 (40bp) with three repeat units were found only in few individuals, which is from SEAS and North Arabian Sea (Western Indian Ocean). There are sub-types for Type 3 with some variation in the repeat unit (Type 3a, 3b and 3c).

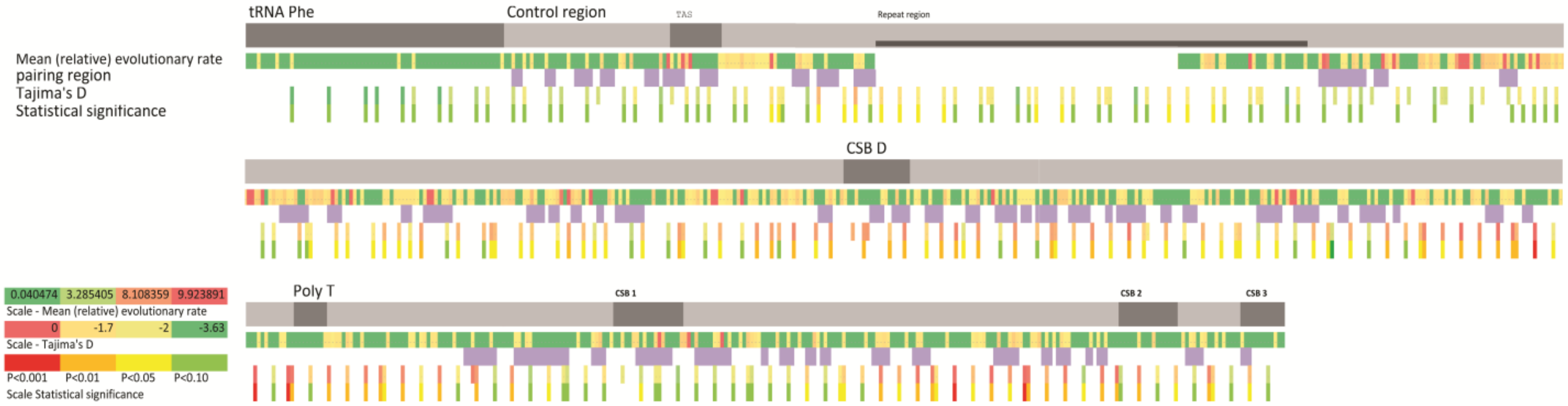
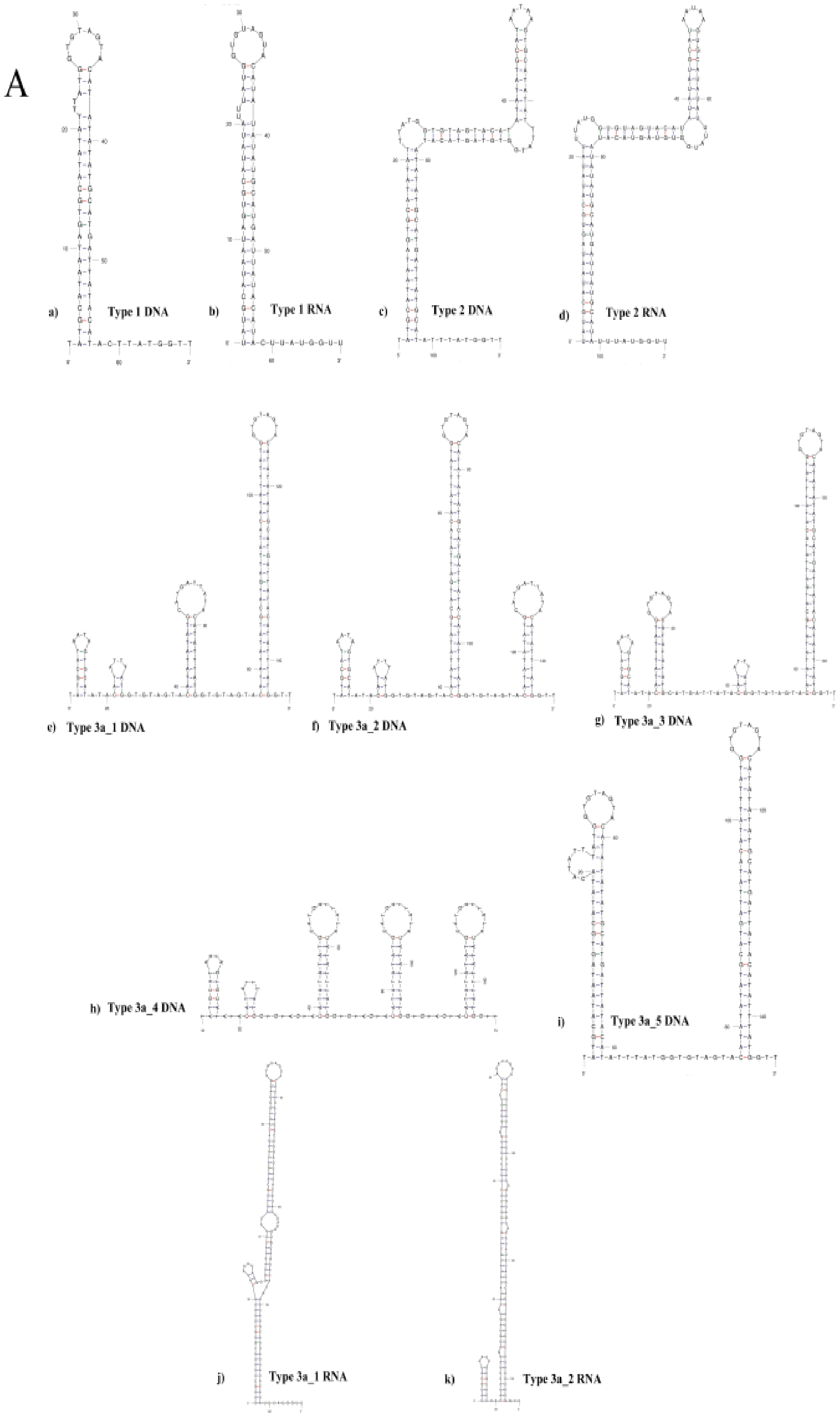


**Fig 3.5 Schematic representation of the *S. longiceps* mtDNA region of ~1112bp comprising tRNA pro, control region and tRNA phe.** The locations of the two tRNA coding flanking regions are indicated in black colour. The characteristic sequence blocks in the control region are indicated as CSBs - conserved blocks, TAS - Termination associated sequence and Poly T. the repeat region between TAS and Poly T is indicated as a black line. Mean (relative) evolutionary rate are shown for each base pair below the site. These rates are scaled such that the average evolutionary rate across all sites is 1. This means that sites showing a rate < 1 are evolving slower than average. Tajima's D and its Statistical significance of ~1112bp comprising tRNA pro, control region and tRNA phe with 10bp intervals overlapping at 5bp are shown.

Several secondary structures with more than 10bp paired bases with varying length were identified in the mtDNA L-strand (Fig 3.S5). The conserved sequences like TAS and CSBs are associated with a secondary structure. All secondary structures predicted for the mtDNA L-strand were also observed in for L-strand mRNA transcript with some minor changes.

The region with length variation and repeat unit is characterized by palindromic sequences within it. Few large and short stem-loop structures with low free energy (-0.101 to - 0.384 ΔG (kcal/mol)) were observed in the repeat region (Fig 3.6, Table 3.3). In haplotypes with Type 3 repeat unit sequences, multiple stem-loop structures have been observed, however, no complex structure observed in Type 1 haplotypes and Type 2 haplotypes. The stem was formed by 3' end of the repeat unit and 5' upstream sequence. In Type 3 repeat unit haplotype sequence, several stem-loop structures with internal bulges were observed. The L strand mRNA transcript of the repeat region is also forming similar structures with greater negative folding energies (ΔG) (Fig 3.6, Table 3.3).



a) Type 1 DNA

b) Type 1 RNA

c) Type 2 DNA

d) Type 2 RNA

e) Type 3a_1 DNA

f) Type 3a_2 DNA

g) Type 3a_3 DNA

h) Type 3a_4 DNA

i) Type 3a_5 DNA

j) Type 3a_1 RNA

k) Type 3a_2 RNA

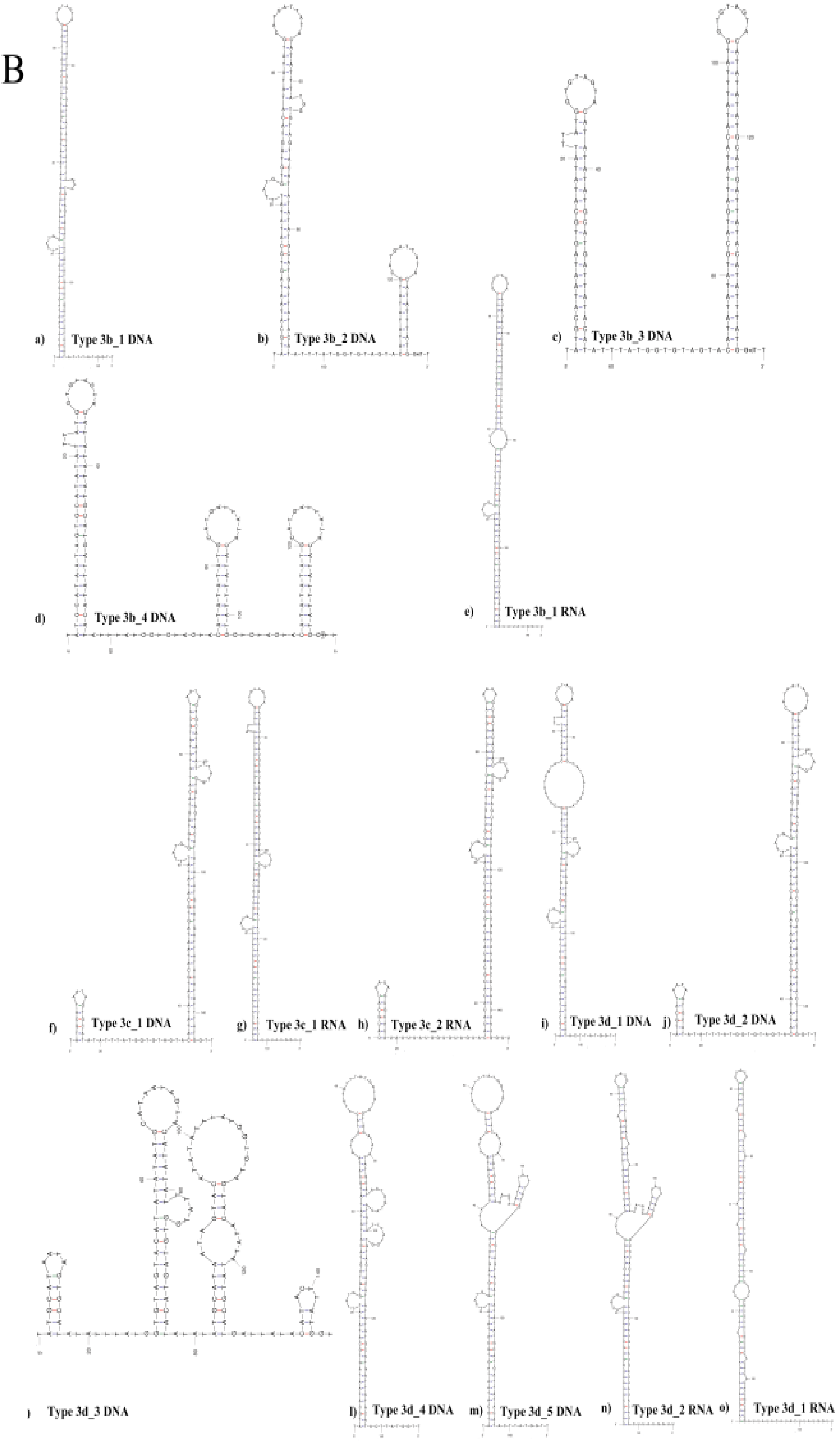


**Fig 3.6 Graphical representation of all predicted secondary structures in repeat unit Type 1, 2 and 3 of mtDNA control region DNA and for the same RNA.** In section A: a) DNA of haplotype with Type 1 repeat unit, b) RNA of haplotype with Type 1 repeat unit, c) DNA of haplotype with Type 2 repeat unit, d) RNA of haplotype with Type 2 repeat unit, e) Structural variant 1 for DNA of haplotype with Type 3 repeat unit variant (Type 3a), f) Structural variant 2 for DNA of

haplotype with Type 3 repeat unit variant (Type 3a), g) Structural variant 3 for DNA of haplotype with Type 3 repeat unit variant (Type 3a), h) Structural variant 4 for DNA of haplotype with Type 3 repeat unit variant (Type 3a), i) Structural variant 5 for DNA of haplotype with Type 3 repeat unit variant (Type 3a), j) Structural variant 1 for RNA of haplotype with Type 3 repeat unit variant (Type 3a), k) Structural variant 2 for RNA of haplotype with Type 3 repeat unit variant (Type 3a). In section B: a) Structural variant 1 for DNA of haplotype with Type 3 repeat unit variant (Type 3b), b) Structural variant 2 for DNA of haplotype with Type 3 repeat unit variant (Type 3b), c) Structural variant 3 for DNA of haplotype with Type 3 repeat unit variant (Type 3b), d) Structural variant 4 for DNA of haplotype with Type 3 repeat unit variant (Type 3b), e) Structural variant 1 for RNA of haplotype with Type 3 repeat unit variant (Type 3b), f) Structural variant 2 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 1 for RNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 2 for RNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 1 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 2 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 3 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 4 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 5 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 1 for RNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 2 for RNA of haplotype with Type 3 repeat unit variant (Type 3c)

**Table 3.3** Folding energy, ΔG for 22 *S. longiceps* mitochondrial tRNA genes and predicted secondary structures of repeat unit types.

| tRNA (RNA) | ΔG(kcal/mol) | Length(bp) | Normalized free energy -ΔG(kcal/mol)/ Length(bp) |
|---|---|---|---|
| tRNA-Ala | -10.77 | 69 | -0.156 |
| tRNA-Arg | -16.3 | 69 | -0.236 |
| tRNA-Asn | -10.12 | 73 | -0.139 |
| tRNA-Asp | -10.37 | 69 | -0.150 |
| tRNA-Cys | -21.7 | 66 | -0.329 |
| tRNA-Gln | -16.21 | 71 | -0.228 |
| tRNA-Glu | -6.1 | 69 | -0.088 |
| tRNA-Gly | -20.3 | 71 | -0.286 |
| tRNA-His | -14.6 | 69 | -0.212 |
| tRNA-Ile | -30.31 | 72 | -0.421 |
| tRNA-Leu | -20.5 | 75 | -0.273 |
| tRNA-Leu | -27.4 | 72 | -0.381 |
| tRNA-Lys | -19.6 | 74 | -0.265 |
| tRNA-Met | -16.24 | 69 | -0.235 |
| tRNA-Phe | -12.34 | 63 | -0.196 |
| tRNA-Pro | -17.1 | 70 | -0.244 |
| tRNA-Ser | -19.2 | 68 | -0.282 |
| tRNA-Ser | -11.31 | 67 | -0.169 |
| tRNA-Thr | -28.2 | 72 | -0.392 |
| tRNA-Trp | -9.07 | 70 | -0.130 |
| tRNA-Tyr | -15.96 | 71 | -0.225 |
| tRNA-Val | -20.3 | 72 | -0.282 |
| Repeat unit predicted structure | ΔG(kcal/mol) | Length(bp) | |
| Type 1 DNA | -8.79 | 67 | -0.131 |
| Type 1 RNA | -17.4 | 67 | -0.260 |
| Type 2 DNA | -24.17 | 108 | -0.224 |
| Type 2 RNA | -41.9 | 108 | -0.388 |
| Type 3a_1 DNA | -15.43 | 147 | -0.105 |
| Type 3a_1 RNA | -39 | 147 | -0.265 |
| Type 3a_2 DNA | -15.43 | 147 | -0.105 |
| Type 3a_2 RNA | -39.5 | 147 | -0.269 |
| Type 3a_3 DNA | -15.22 | 147 | -0.104 |
| Type 3a_4 DNA | -15.21 | 147 | -0.103 |
| Type 3a_5 DNA | -14.97 | 147 | -0.102 |
| Type 3b_1 DNA | -16.64 | 143 | -0.116 |
| Type 3b_1 RNA | -43.9 | 143 | -0.307 |
| Type 3b_2 DNA | -16.42 | 143 | -0.115 |
| Type 3b_3 DNA | -15.87 | 143 | -0.111 |
| Type 3b_4 DNA | -15.65 | 143 | -0.109 |
| Type 3c_1 DNA | -26.24 | 147 | -0.179 |
| Type 3c_1 RNA | -49.4 | 147 | -0.336 |
| Type 3c_2 RNA | -47.5 | 147 | -0.323 |
| Type 3d_1 DNA | -16.05 | 147 | -0.109 |
| Type 3d_5 DNA | -15.23 | 147 | -0.104 |
| Type 3d_1 RNA | -38.65 | 147 | -0.263 |
| Type 3d_2 DNA | -15.73 | 147 | -0.107 |
| Type 3d_2 RNA | -39.5 | 147 | -0.269 |
| Type 3d_3 DNA | -15.58 | 147 | -0.106 |
| Type 3d_4 DNA | -15.35 | 147 | -0.104 |

The number of substitutions/rates of evolution in paired sites was comparatively lower than the unpaired sites and similar sequences were observed in the control region of other Clupeid fishes. Even though the repeat units were present, their sequences slightly differed in different species with the CSBs, TAS and poly-A being highly conserved among clupeid control region. The regions between these conserved regions and repeat units are highly polymorphic among species. Thus, the sequence conservation between species indicates the conformation predicted for the structure in the control region is maintained during evolution or diversification of species.

To assess the robustness of the secondary structure predicted, its folding potential (Free energy, $\Delta G$ = kcal/mol) with the tRNA, which is known to form functional secondary structure were compared. The relative free energy ($\Delta G$/Length) of tRNA range from -0.42 to -0.08 and predicted repeat unit structures range from - 0.384 to - 0.101 (Table 3.3). In the predicted structures, RNA has relatively high relative free energy than its DNA and Type 1 and Type 2 repeat units have lower free energy than Type 3 repeat units (Table 3.3). This indicates a higher folding potential of haplotype with Type 1 and Type 2 over Type 3 repeat units.

For tRNA regions, the Tajima’s D value is zero/negative and significant for most of the coding regions, as expected for functionally constrained regions (Table 3.1). Interestingly the mitochondrial control regions also showed negative and highly significant values (Fig 3.5.). The repeat unit position has a value of -2.14526 ($P < 0.01$) (Fig 3.5.). These results indicate that the coding regions in the mitochondrial DNA are under negative selection force and some regions in the control region (TAS, CSD and repeat unit) are also under negative selection similar to the coding region.

Type 1 haplotypes with one repeat unit were the most abundant in all geographic locations of the Indian Ocean probably because haplotype with one repeat unit has more folding potential (Normalized free energy - $\Delta G$(kcal/mol)/Length(bp) for DNA = 0.131 and RNA = 0.260). Haplotypes with two and three repeat units were less abundant and restricted to the Western Indian Ocean (both eco-regions) (Table 3.3, Fig 3.4).

## 3.5 Environmental data

Wide variations in temperature, dissolved oxygen, salinity, and chlorophyll-*a* were observed between NAS, SEAS and BoB. NAS encompasses the Persian Gulf, Gulf of Oman, Red Sea and the northeast Arabian Sea where a unimodal pattern of sea surface temperature (SST) is observed with the highest temperature (24-27$^0$c) during the northeast monsoon season (October-March) and lowest temperature (20-22$^0$c) during the southwest monsoon season (June–September) (Rao *et al*. 1992). The NAS also is characterized by a very high chlorophyll-*a* concentration (4-10 mg/m$^3$) during May-June, lasting up to October (during the southwest monsoon season). The average sea surface salinity (SSS) is also higher along with NAS throughout the year (36-38ppt) (Chatterjee *et al.* 2012) (Fig. 3.7, Table 3.S7, Fig. 3.S12). SEAS exhibits a typical bimodal pattern of SST with the warm (29-30$^0$C) spring intermonsoon (April–May) and the fall intermonsoon (October–November) and the cool (26-28 $^0$C) southwest monsoon (June–September) and the northeast monsoon seasons (December-March) (Fig. 3.7, Supplementary Table 3.S7, Fig. 3.S12) (Prasanna Kumar *et al.* 2002). High chlorophyll-*a* concentration (Fig. 3.S7, Fig. 3.S12) observed at SEAS (Malabar upwelling zone) is due to the intense coastal upwelling from May to September, and it peaks during July and August (5-10 mg/m$^3$). By October, it recedes to a low (1-2 mg/m$^3$) chlorophyll-*a* concentration and maintains up to May. SEAS is also characterised by a very low dissolved oxygen (1-2 mg/L) during the southwest monsoon season while it is 2-4 mg/L in NAS and 3-5 mg/L in BoB. Coastal upwelling along the Somalia coast and SEAS brings not only subsurface cool, nutrient-rich waters but also less oxygenated waters to surface layer, while coastal upwelling regions in the BoB are well oxygenated (Fig. 3.S7, Fig. 3.S12). During this season, high chlorophyll-*a* concentration along the coast is well corroborated with low temperature and low dissolved oxygen. On the contrary, BoB is characterized by stable dissolved oxygen (3-5 mg/L), reduced salinity (28-33ppt), reduced temperature (28-30$^o$C) and low chlorophyll-*a* (0-3 mg/m$^3$) environment than SEAS and NAS, throughout the year (Qasim 1982). Differences in the standard deviation of annual temperature, dissolved oxygen, salinity, and chlorophyll-*a* of SEAS, NAS and BoB corroborated with the above observations. While the BoB exhibits small deviation (Madhupratap *et al.* 2001; Jouanno *et al.* 2012), the SEAS and NAS exhibits very high and intermediate deviation, indicating very high fluctuations in the environmental parameters of SEAS. Variations in frequencies of the positively selected amino acid substitutions of the dataset were positively and highly correlated to fluctuations in annual Sea surface Temperature (SST) (represented as standard deviation) (parameter estimate = 2.04, SE = 0.04, P = 0.01), fluctuations in annual chlorophyll-*a* (represented as standard deviation) (estimate = 1.24, SE = 0.23, P = 0.03) and negatively to Fall (October, November, December) - Dissolved Oxygen (DO) (estimate = -4.68, SE = 0.27, P = 0.02). Moderate correlation was obtained for the amino acid substitutions under selection with fluctuations in annual DO (represented as standard deviation) (estimate = 0.99, SE = 0.4, P = 0.057), Winter -SST (estimate = 0.47, SE = 0.031, P = 0.057), Spring-DO and Summer Chlorophyll-*a* (estimate = 2.13, SE = 0.13, P = 0.057).

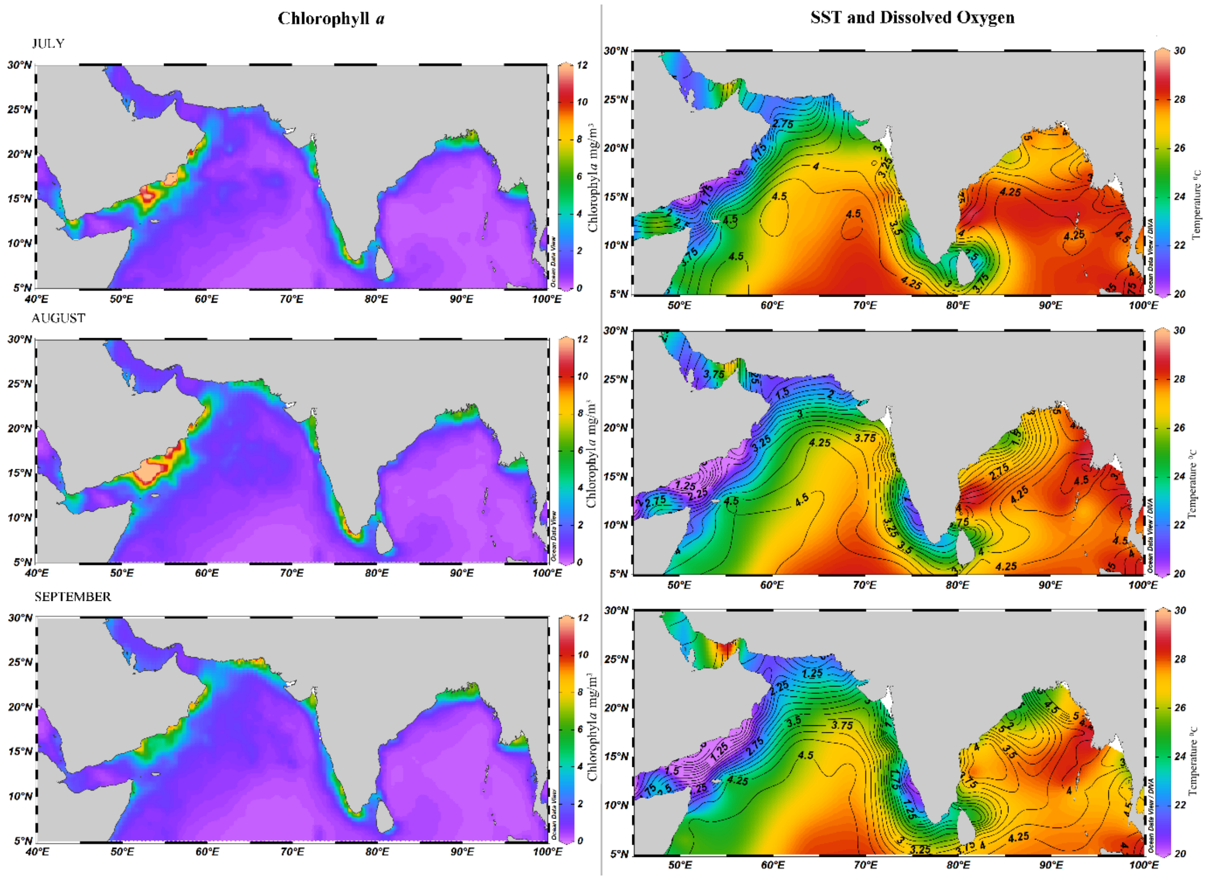


**Figure 3.7** Monthly Chlorophyll-*a* (mg/m$^3$), Sea Surface Temperature- SST ($^0$C) and Dissolved Oxygen (µmol/kg) for the Bay of Bengal and Arabian Ocean during July to September. Chlorophyll *a* and Sea Surface Temperature gradients are represented as coloured shades. Dissolved Oxygen is represented as contour lines. The images were generated in ODV 5.1.7 (https://odv.awi.de/).

## 4. Discussion

The purifying selection has been detected as the dominant force shaping evolution of the mitogenome of *S. longiceps* in the present study. Despite the importance of purifying selection, the positive and diversifying selection was detected for fixed amino acid replacements in key regions involved in oxidative phosphorylation complexes, Complex I: NADH dehydrogenase (ND1, ND2, ND4 and ND5), Complex III: Cytochrome b (CYT B), Complex IV: Cytochrome C Oxidase (C01, CO2 and CO3) and Complex V: ATP Synthase (ATP6 gene). Besides, 8 sites were found specifically in south-east Arabian Sea eco-region and 2 sites in NAS eco-region. *S. longiceps* has a characteristic control region with a 38 – 40 bp tandem repeat units (palindromic sequences within it) and they are under strong selective pressure similar to the coding region. We predicted stable intra-strand secondary structures with low folding energy (- 0.101 to - 0.384 kcal/mol) in the repeat unit. Haplotypes with one repeat unit have lower folding energy and it is the most abundant haplotype in the samples probably due to their enhanced stability (High folding potential). Haplotypes with two and three repeat units are less abundant and they are restricted to the Western Indian Ocean. Correlations of variations in the frequencies of the positively selected amino acid substitutions with variations in the environmental parameters (temperature, dissolved oxygen, and chlorophyll-*a*) indicate that the observed genetic diversity and positive selection in the mtDNA of *S. longiceps* may be driven by the pressures of the heterogeneous environment. High selective pressures were evident in both coding and non-coding regions in samples collected from SEAS followed by NAS. The SEAS is considered as a region with a complex interplay between many oceanographic processes that vary spatially and temporally (Narvekar *et al.* 2017).

*Sardinella longiceps* populations in the Indian Ocean do not show any clear geographical structuring in whole mitogenome analysis. However, our data indicate the low nucleotide diversity in Indian oil sardine with the highest differentiation between samples from NAS and Indian coast (SEAS and BoB). Even though marine fishes have traditionally been included in the low genetically differentiated and weak locally adapted category, the low genetic differentiation found in them is presumed to be biologically meaningful (Knutsen *et al.* 2011). Marine fishes can exist as locally adapted populations with gene flow between them (Hemmer-Hansen *et al.* 2007; Hutchings *et al.* 2007). Migrations homogenize frequency of alleles in the gene pool of populations but diversifying selection will act against it. A previous investigation using two mitochondrial DNA markers indicated a lack of significant genetic differentiation in Indian oil sardine populations along the Indian coast (Sukumaran *et al* 2015) which may be due to the low resolving power of the markers (Hauser and Carvalho 2008). On the contrary, microsatellite markers (*Chapter 4*) indicated significant genetic differentiation between populations of NAS and other regions (Sebastian *et al.* 2017a). The observed sub-structuring in sardines from NAS, SEAS and BoB waters indicating the influence of complex oceanographic factors in determining gene flow (Sebastian *et al.* 2017a). The presence of morphotypes/ecotypes and the possibility of local adaptation in Indian oil sardine has also been implicated (Sukumaran *et al.* 2016). Due to the large effective population size, high dispersal capacity, high fecundity and long planktonic larval phase (40–50 days in the case of *S. longiceps*) of small pelagic fishes like sardines, the contribution of genetic drift in mtDNA evolution will be negligible (Hauser and Carvalho 2008).

In a population with low effective population size ($N_e$), fixation of slightly deleterious mutations by drift may leave a similar signature as positive selection (Pavlova *et al.* 2017). But for a pelagic species like *S. longiceps* with large $N_e$ ($N_e$ mtDNA 1.1 × $10^6$ to 1.31 × $10^9$) (Sukumaran *et al.* 2015), low turnover time and no physical barriers to gene flow, the effect of genetic drift will be minimal in mtDNA evolution. All these evidence points to the possible influence of purifying and the positive selection and the resulting reduction in nucleotide diversity in the mtDNA of Indian oil sardine (Meiklejohn *et al.* 2007). Purifying selection and conservative amino acid changes can result in stabilizing selection that is responsible for the preservation of adaptive characteristics of organisms under constant environmental conditions.

Extensive non-synonymous mutations have been reported in ND genes in many fishes (Garvin *et al.* 2011; Consuegra *et al.* 2015; Jacobsen *et al.* 2016) as in the present study, supporting the relaxed constraints in the ND genes. The high rate of mutations observed in complex I may be due to the position of the ND genes as they are found immediately upstream from the origin of L-strand replication (OriL) and downstream from the origin of H-strand (OriH) replication as these genes stay single-stranded for more time compared to other genes during replication making them prone to high rate of mutation (Marshall *et al.* 2008).

Most of the positively selected sites in complex I were located in the functionally relevant proton-conducting membrane transporter transmembrane domain (ND4 - Proton_antipo_M), (ND5 - Proton_antipo_M & N - terminal) indicating their importance in protein function (Da Fonseca *et al.* 2008). The position 228 (ND5Lys-Asn) under diversifying selection showed overlap with amino acid residue that has been reported as one of the key residues in proton translocation (Zhu *et al.* 2016). Cytochrome C Oxidase (complex IV) catalyses the final step in mitochondrial electron transfer chain and is considered as one of the major regulation sites for OXPHOS (Li *et al.* 2006). Two of the sites in complex IV (#25C01 & #114C01) showed overlap with amino acid residues that have been reported to participate in polypeptide binding at Subunit I/VIIc interface & Subunit I/VIIa interface respectively. Whilst in the intermembrane domain of Cytochrome b of Complex III which is functionally important, more conservation was noticed except one region (#311) that has been reported to participate in polypeptide binding in interchain domain interface which may influence on structure and function of cytochrome b. Control region exhibit a relatively high mutation rate due to reduced functional constraints whereas the low mutation rate in tRNAs and rRNAs may be due to strict purifying selection (Jacobsen *et al.* 2016).

Polymorphisms observed in regions that have been reported to participate in polypeptide binding at mitochondrial and the nuclear-encoded subunits interface (Subunit I/VIIc interface & Subunit I/VIIa interface; complex IV) may change the structure and efficiency of OXPHOS complex possibly playing a role in adaptation. Thus, the co-evolution between mitochondrial and nuclear-encoded subunits due to genome-genome interactions can affect OXPHOS function and regulation in *S. longiceps*. Such co-evolution has been reported in cytochrome c oxidase (complex 4) of primates (Osheroff *et al.* 1983) and NADH dehydrogenase complex of humans (Gershoni *et al.* 2014). Positively selected sites that appear to interact with other COX subunits were also reported from high-performance fish like *Scombroidei* (Dalziel *et al.* 2006). When mito-nuclear interactions are disrupted, it results in reproductive isolation and speciation (Burton *et al.* 2013). The evidence of positive selection in regions involved in mito-nuclear interactions signals possible reproductive isolation, lineage sorting and diversification in sardine which needs to be investigated further.

The amino acid replacements in mitochondrial proteins may have some beneficial and adaptive function in the metabolic performance of *S. longiceps* in a dynamic ocean environment especially in coping up with the challenges from climate change. There are few reports regarding the correlation between genetic diversity of OXPHOS genes, and environmental pressures like hypoxia (Da Fonseca *et al.* 2008), heat stress (Morales *et al.* 2015), cold stress (Stier *et al.* 2014) and nutrient availability (Da Fonseca *et al.* 2008). Adaptive evolution in mitogenome in response to environmental conditions like temperature and salinity has been reported in Atlantic cod (Berg *et al.* 2015), Atlantic salmon, Pacific salmon and Killer whale populations (Foote *et al.* 2011; Garvin *et al.* 2011; Consuegra *et al.* 2015). It is proposed that the coupling efficiency of OXPHOS complex is related to thermal adaptation as it generates heat and ATP during respiration (Lowell and Spiegelman 2000). The heat produced by less coupled mitochondrial membrane protein assembly may be beneficial for endothermic organisms in a cold environment. But heat production may be dangerous in a warm environment because it induces high oxidative stress (high ROS production) (Fangue *et al.* 2009) and affects nutrient uptake (Brand 2000).

The length polymorphism observed in the *S. longiceps* control region is due to the presence of repetitive elements between the TAS and Poly T elements. This repeat region has high folding potential or low folding energy (ΔG) which contributes to the formation of stable stem and loop secondary structures. The folding potential, several substitutions/rates of evolution in paired sites and Tajima’s D statistics analysis showed that they are under strong selection pressure. Haplotypes having repeat unit one with lower folding energy, ΔG (high folding potential) is the most abundant. The haplotypes with two and three repeat units have greater folding energies which are less abundant and restricted to the Western Indian Ocean. The mitochondrial length variation/heteroplasmy due to tandem repeat in the control region is a common phenomenon in animals (Brown *et al.* 1986; Wright 2000). Different models like slipped-strand mispairing (Samuels *et al.* 2004), intermolecular recombination, transposition (Mita *et al.* 1990) and misalignment during replication have been suggested as the mechanism behind observed polymorphisms (Pereira *et al.* 2008). In mammals, deletions in mitogenome are closely linked to mitochondrial diseases and proven to be associated with site-specific breakage hotspots (Samuels *et al.* 2004). The breakage points or deletions are associated with or near the regions with low folding energy, ∆G as well as regions with tandem repeat units. The formation of secondary structures may also promote these deletion mutations.

Our results showed that the distribution of haplotype with repeats units having the highest folding potential and low folding energy is abundant in all oceanic regions where *S. longiceps* is distributed. The distribution of other haplotypes with more than one repeat unit is restricted to the Western Indian Ocean and the repeat units forming secondary structures are under strong negative/purifying selection which indicates that mtDNA is subjected to different selection pressures at different organizational level (at the levels of sequences/amino acids/proteins/structure). The observed differences in substitution rate, negative and highly significant Tajima's D at conserved sequence domains, loops in the predicted secondary structure-forming regions and the presence of tandem repeats emphasize the presence of selective pressures. If the length of the repeat region varies randomly/neutrally, a large difference in repeat region and the number of repeats would have observed (Mignotte *et al.* 1990). But in *S. longiceps* the substitution rate in the loop forming region, several tandem repeats and variation within the repeat unit are negligible. This indicates a selective constraint in this portion which could be explained by a possible function/ presence of protein binding regions in these elements. The secondary structure-forming region might have a role in regulatory function in mtDNA replication and transcription (Pereira *et al.* 2008; Melo-Ferreira *et al.* 2014). Thus, we hypothesize that a selection force is acting at an intra-mitochondrial or inter-cellular level against the inherent tendency of length variation and point mutations which break the secondary structure involved in the efficient regulation of mtDNA functions.

In protein-coding regions, the variability of substitution can be explained by selection force acting on its translated products, but in the non-coding region, it could be explained only by the role of the secondary structure formed by DNA or function of the DNA sequence itself (Wright, 2000). It is clear from the analysis that the TAS, poly-A and secondary structure-forming repeat units were conserved within species and among clupeid fishes. The loop forming regions are protected from a mutation which is more likely to occur during replication as it forms a single-stranded structure (D-loop). Thus, the observed low mutation rate between loop forming regions, CSD, TAS and flanking region strongly indicates the differences in purifying selection pressures acting on it and significant negative Tajima's D strongly support this hypothesis. Thus, a combination of selection pressure and protective effect of intra-strand loop formation acts for some structures of the control region. The position of loop forming region between TAS and poly-A, which includes the D-loop forming region indicates its possible role in replication initiation and/termination of elongation in proposed models of mitochondrial replication (Shadel and Clayton 1997; Yasukawa *et al.* 2005). The occurrence of the secondary structure near the hot spot of polyadenylation sites (Poly A) (Slomovic *et al.* 2005) strengthens its possible role in transcription termination. The secondary structure may act as a punctuation point for correct mRNA processing (Ojala *et al.* 1980). It has been reported that the mitochondrial structural variants/ haplogroups have a clear link with mtDNA copy number variation (by influencing the replication machinery) in humans and contribute to the adaptation of the human population to different climatic zones (Suissa *et al.* 2009; Melo-Ferreira *et al.* 2014; Lajbner *et al.* 2018). Thus, the observed control region variation and its geographical distribution pattern is an indication of locally adapted *S. longiceps* populations to different eco-regions of the Indian Ocean.

Theoretical and empirical evidence have suggested the role of many environmental parameters for species persistence, adaptation, and phenotypic and genotypic diversification (Coyne and Orr 2004; Thompson 2013) Temperature has been proposed as one of the main factors driving evolutionary diversification due to enhanced mutation rates in mitochondrial as well as a nuclear genome in many taxa, which also is the reason for species diversity in tropics as compared to temperate waters (Jablonski 1993; Gillooly *et al.* 2005). It has been demonstrated that fluctuations in the environmental factors promote phenotypic and evolutionary diversification in organisms (Melbinger and Vergassola 2015; Dean et al. 2017; Fuentes and Ferrada 2017; Eddie and Aneil 2019). The Arabian Sea Large Marine Ecosystem is considered as one of the major upwelling systems in the world causing variations in temperature, dissolved oxygen, salinity and chlorophyll-*a*. SEAS exhibit wide fluctuations in temperature values annually as compared to NAS and BoB. Mitochondria play essential roles in aerobic metabolism which is a temperature-sensitive process and consequently, mitochondria are considered as possible sites of processes influencing thermal limits of organisms. Thus, thermal acclimation alters mitochondrial properties to maintain aerobic scope (Iftikar and Hickey 2013). Thermal acclimation in ectotherms may happen by maintaining the stability of OXPHOS proteins for which a few amino acid substitutions may be necessary (Somero 1995; Baris *et al.* 2015). SEAS is also characterized by variations in chlorophyll-*a* values and oxygen minimum zones (hypoxia). Starvation (reduced chlorophyll-*a* in this case) can drive mtDNA evolution by acting as a force to generate energy more efficiently by improving the efficiency of the coupling of energy production in the OXPHOS pathway (Rion and Kawecki 2007; Ballard and Melvin 2010; Ruiz-Pesini *et al.* 2004). Hence, productivity variations of the sardine habitat may lead to the evolution of genotypes with more efficient OXPHOS pathways. Similarly, the occurrence of oxygen minimum zones and consequent hypoxia in SEAS (during southwest monsoon) also demand more efficient coupling of energy production (Solaini *et al.* 2010). Thus, the abundance of selective signatures/higher rate of selected genotypes in SEAS may be a response to the uncertain environmental conditions (hypoxia, temperature and productivity) which warrant ecotypes of high metabolic efficiency for survival and reproduction, compared to the stable environment of NAS and highly stable environment of BoB.

The prevalence of diversifying selection in the SEAS indicates the action of evolutionary forces in mitochondrial OXPHOS complex associated with metabolic adaptation to the dynamic and highly productive environment of the Malabar upwelling zone. The high levels of abundance of sardine populations in this region will also be a factor promoting natural selection and diversification (Harrisson *et al.* 2016; Hughes *et al.* 2017). Positive or diversifying selection may be a factor enhancing metabolic capacity for adapting to their natal habitats and consequently recruits from non-matching natal habitats may be negatively selected due to their competitive disadvantage (Marshall *et al.* 2010). All these observations indicate the presence of locally adapted populations in Indian oil sardine which may have evolved to survive in the uncertain environmental and oceanographic factors to which they are exposed. Two positively selected sites in ND1 (29,30) and ND5 (97,98) genes and one site each in ND4 (148), CO1 (25), CO2 (152) and ATP6 (185) respectively were specific to South-East Arabian sea and one site each in CO2 (50) and CYTB (250) specific to NAS populations. The potential for using these loci as markers for tagging local populations of Indian oil sardine should be explored further as these findings are important for devising management and conservation strategies for Indian oil sardine.

Reports regarding size variation in Indian oil sardine (Sukumaran *et al.* 2016) has also indicated the reduced average size of *S. longiceps* caught from SE Arabian sea in contrast to those from NAS (Oman) waters which also is a well-explained phenomenon for metabolic rates and body size constraints in tropics and temperate regions (Brown *et al.* 2004; Gillooly *et al.* 2005; Allen and Gillooly 2007; Gillooly and Allen 2007). Differences in habitat characteristics between these eco-regions may reduce the success of larval recruitment/ and colonization/survival of specific haplotypes which are locally adapted (Marshall and Morgan 2011) even though they may not act strictly as barriers. The evidence of positive selection in regions participating at mitochondrial and the nuclear-encoded subunits interface interactions (CO1, Subunit I/VIIc interface & Subunit I/VIIa interface in complex IV) from SEAS and NAS signals possible reproductive isolation, sympatric speciation and diversification in sardines. Polymorphisms in regions involved in mito-nuclear interactions may disrupt mito-nuclear interactions resulting in reproductive isolation and speciation (Burton *et al.* 2013). Further investigations using genome-wide markers like SNPs may provide more clarity to these findings.

Even though purifying selection has been detected as the dominant force shaping evolution of sardine mitogenome, the observed diversifying selection may interfere with the conformational coupling of mitochondrial complex, electron translocation and mito-nuclear interactions which may have some evolutionary advantage to provide optimum fitness to heterogeneous ocean habitats. The variation in the geographical distribution frequencies of positively selected sites and control region haplotypes indicates the environmental selection force acting on mtDNA of *S. longiceps*. These functional genes and regulatory elements exhibiting diversifying selection have the potential to act as markers for inferring population genetic structure, plastic responses, adaptation and functional gene evolution in marine fishes and thus could be valuable for management and conservation of this important resource. Further studies could be carried out in Indian oil sardine using whole genome, transcriptome and reduced representation genome scan methods like RAD sequencing to identify genome level adaptations which will provide holistic information for their adaptive capacity. Common-garden experiments (de Villemereuil *et al.* 2016; Gueye *et al.* 2016) can be used investigate the correlation between eco-region characteristics with genotypes and their fitness consequences The present study assumes great relevance from this point of view as Indian oil sardines with enhanced adaptive signatures in mitogenome can be further monitored for their spatial and temporal distribution which may provide clues regarding climatic impacts in the Indian ocean. Further, small pelagic fishes form the mainstay of food security of many coastal states of developing nations and it is imperative to understand their dynamics in space and time.

## Supplementary Tables and Figures

**Table 3.S1** List of Primer pairs used for amplification of *S. longiceps* mitochondrial DNA.

| Primer Name | | Sequence (5' - 3') | PCR Product length |
|---|---|---|---|
| SPF M 1 | Forward primer | AAGAGGGCCGGTAAAACTCG | 1080 |
| | Reverse primer | GGTTTCGGGGGCTCAAACTA | |
| SPF M 2 | Forward primer | CACAATATTCGCCGCAAGGG | 1140 |
| | Reverse primer | GCGGCCGTTAAACTTTTGGT | |
| SPF M 3 | Forward primer | TCCTGCAGCAAGACATCGTT | 1287 |
| | Reverse primer | AGGCTGGATAGGGCCAAAAC | |
| SPF M 4 | Forward primer | GTTTTGGCCCTATCCAGCCT | 1390 |
| | Reverse primer | TTGGGTCTGGTTAAGACCGC | |
| SPF M 5 | Forward primer | CCACCCCTACCTCCTAACGA | 1267 |
| | Reverse primer | ATGCCATATCAGGTGCTCCG | |
| SPF M 6 | Forward primer | CTCTGTCAGGCAATCTGGCA | 1299 |
| | Reverse primer | ACGCAGGGGTTTAACCTACG | |
| SPF M 7 | Forward primer | CGTAGGTTAAACCCCTGCGT | 1307 |
| | Reverse primer | AATCACCGTAGCAAGCCACA | |
| SPF M 8 | Forward primer | TGTGGCTTGCTACGGTGATT | 1071 |
| | Reverse primer | GCTGCCTCAAACCCAAAGTG | |
| SPF M 9 | Forward primer | ACCACTTTGGGTTTGAGGCA | 1130 |
| | Reverse primer | CATGTGGTTCTGGCTGGCTA | |
| SPF M 10 | Forward primer | GATCATCGCCTCTCTGAGCC | 1131 |
| | Reverse primer | AGAGAGTACCCGGCTGTGAT | |
| SPF M11 | Forward primer | ATCACAGCCGGGTACTCTCT | 1458 |
| | Reverse primer | TTGCTCATCGTTGAGGCTGT | |
| SPF M 12 | Forward primer | ACAGGCACCCCTTTCTTAGC | 1344 |
| | Reverse primer | TCTGGAGCTTGTTGCGTCAT | |
| SPF M 13 | Forward primer | AGAGCTCACCGGGTATTCCT | 1018 |
| | Reverse primer | AAGTGGAACGCGAAAAACCG | |
| SPF M 14 | Forward primer | CGGTTTTTCGCGTTCCACTT | 1287 |
| | Reverse primer | AAGGACTCGCCAGATGCAAA | |

**Table 3.S2** AMOVA results for the whole genome, all gene concatenated, individual genes and the control region.

| Source of variation | Variance component | % of variation | $\Theta_{ST}$ | *P*-value |
|---|---|---|---|---|
| WHOLE GENOME | | | | |
| Among population | 5.432 | 10.36 | | |
| Within population | 47.008 | 89.64 | | |
| Total | 52.441 | 100.00 | 0.10359 | < 0.001 |
| ALL GENES CONCATENATED | | | | |
| Among population | 4.627 | 11.39 | | |
| Within population | 36.012 | 88.61 | | |
| Total | 40.641 | 100.00 | 0.11387 | < 0.001 |
| ATP 6 | | | | |
| Among population | 0.4322 | 20.05 | | |
| Within population | 1.7238 | 79.95 | | |
| Total | 2.1661 | 100.00 | 0.20047 | < 0.001 |
| ATP 8 | | | | |
| Among population | 0.0043 | 3.87 | | |
| Within population | 0.1094 | 96.17 | | |
| Total | 0.1137 | 100.00 | 0.03825 | 0.20851 |
| CO 1 | | | | |
| Among population | 0.2672 | 8.57 | | |
| Within population | 2.8503 | 91.43 | | |
| Total | 3.1175 | 100.00 | 0.08572 | 0.00051 |
| CO 2 | | | | |
| Among population | 0.05529 | 2.73 | | |
| Within population | 1.96996 | 97.27 | | |
| Total | 2.02525 | 100.00 | 0.02730 | 0.19158 |
| CO 3 | | | | |
| Among population | 0.38357 | 20.37 | | |
| Within population | 1.49986 | 79.63 | | |
| Total | 1.88343 | 100.00 | 0.20366 | < 0.001 |
| CONTROL REGION | | | | |
| Among population | 0.59306 | 6.66 | | |
| Within population | 0.30531 | 93.34 | | |
| Total | 8.89836 | 100.00 | 0.06665 | 0.00168 |
| CYT B | | | | |
| Among population | 0.28977 | 7.42 | | |
| Within population | 3.61798 | 92.58 | | |
| Total | 3.90775 | 100.00 | 0.07415 | 0.00099 |
| ND 1 | | | | |
| Among population | 1.14923 | 21.92 | | |
| Within population | 4.09380 | 78.08 | | |
| Total | 5.24303 | 100.00 | 0.21919 | < 0.001 |
| ND 2 | | | | |
| Among population | 0.67747 | 13.27 | | |
| Within population | 4.42669 | 86.73 | | |
| Total | 5.10416 | 100.00 | 0.13273 | < 0.001 |
| ND 3 | | | | |
| Among population | 0.10563 | 11.28 | | |
| Within population | 0.83100 | 88.72 | | |
| Total | 0.93663 | 100.00 | 0.11278 | 0.00604 |
| ND 4 | | | | |
| Among population | 0.57774 | 9.36 | | |
| Within population | 5.59671 | 90.64 | | |
| Total | 6.17445 | 100.00 | 0.09357 | < 0.001 |
| ND 4L | | | | |
| Among population | -0.00126 | -0.63 | | |
| Within population | 0.20296 | 100.63 | | |
| Total | 0.20170 | 100.00 | -0.00625 | 0.53475 |
| ND5 | | | | |
| Among population | 0.71526 | 8.36 | | |
| Within population | 7.84554 | 91.64 | | |
| Total | 8.56080 | 100.00 | 0.08355 | < 0.001 |
| ND6 | | | | |
| Among population | -0.01580 | -0.96 | | |
| Within population | 1.66878 | 100.96 | | |
| Total | 1.65298 | 100.00 | -0.00956 | 0.54683 |

The proportion of variance distributed among population samples was analyzed using the hierarchical analysis of molecular variance procedure (AMOVA) in Arlequin.

**Table 3.S3** Pairwise $\Phi_{ST}$ for whole-genome and all gene concatenated sequences**.**

| Whole-genome | | | |
|---|---|---|---|
| | SEAS | NAS | BOB |
| SEAS | | **0.00684** | 0.13965 |
| NAS | 0.02701 | | **0.04398** |
| BOB | 0.00803 | 0.01937 | |
| All gene concatenated | | | |
| | SEAS | NAS | BOB |
| SEAS | | **0.00879** | 0.06055 |
| NAS | 0.03129 | | **0.04391** |
| BOB | 0.01478 | 0.0183 | |

The numbers below the diagonals are $\Phi_{ST}$ and the number above the diagonal are the probability value. NAS (Northern Arabian Sea), SEAS (South Eastern Arabian Sea) and BOB (Bay of Bengal) Arlequin (Excoffier & Lische, 2010) were used to estimate F statistics, pairwise $\Theta_{ST}$ for whole-genome nucleotide sequence.

**Table 3.S4** Sampling locations of *S. longiceps* populations from 3 ecoregions in the Indian Ocean.

| Sampling location | | Latitude | Longitude |
|---|---|---|---|
| NAS n = 117<br>(15 Complete mtDNA, 117 control region) | OMAN | 19.482ºN | 63.744 ºE |
| | VERAVAL | 20.793 ºN | 69.842 ºE |
| | MUMBAI | 19.465 ºN | 72.111 ºE |
| SEAS n = 117<br>(15 Complete mtDNA, 117 control region) | MANGALURU | 13.112 ºN | 74.262 ºE |
| | KOZHIKODE | 11.101 ºN | 75.251 ºE |
| | KOLLAM | 9.344 ºN | 76.112 ºE |
| | THIRUVANANTHAPURAM | 8.182 ºN | 76.933 ºE |
| BoB n = 116<br>(15 Complete mtDNA, 116 control region) | CHENNAI | 13.291 ºN | 18.794 ºE |
| | VISAKHAPATNAM | 17.934 ºN | 84.182 ºE |

NAS (Northern Arabian Sea), SEAS (South Eastern Arabian Sea) and BoB (Bay of Bengal), n number of individuals collected.

**Table 3.S5** Nucleotide diversity of *S. longiceps* populations from 3 ecoregions in the Indian Ocean.

| | NAS | SEAS | BoB |
|---|---|---|---|
| ND1 | 0.0064 | 0.00794 | 0.0065 |
| ND2 | 0.00673 | 0.00878 | 0.0065 |
| CO1 | 0.00275 | 0.00294 | 0.0021 |
| CO2 | 0.00279 | 0.00706 | 0.0025 |
| ATP8 | 0.0008 | 0.00163 | 0 |
| ATP6 | 0.00208 | 0.00606 | 0.0069 |
| CO3 | 0.00161 | 0.00459 | 0.0073 |
| ND3 | 0.00172 | 0.00546 | 0.00507 |
| ND4L | 0.00142 | 0.00089 | 0.00112 |
| ND4 | 0.00639 | 0.00862 | 0.00809 |
| ND5 | 0.00722 | 0.00878 | 0.00842 |
| ND6 | 0.0077 | 0.00457 | 0.00696 |
| CYTB | 0.0021 | 0.00488 | 0.00543 |
| CONTROL | 0.0132 | 0.018 | 0.01292 |
| ALLGENE | 0.0049 | 0.01678 | 0.00284 |
| ALLSEQ | 0.0045 | 0.00605 | 0.00512 |

NAS (Northern Arabian Sea), SEAS (South Eastern Arabian Sea) and BoB (Bay of Bengal).

**Table 3.S6.** Amino acid diversity of *S. longiceps* populations from 3 ecoregions in the Indian Ocean.

| | NAS | SEAS | BoB |
|---|---|---|---|
| ND1 | 0 | 0.000308 | 0.00414 |
| ND2 | 0.000703 | 0.00182 | 0.00048 |
| ND3 | 0.001163 | 0.00204 | 0 |
| ND4L | 0 | 0 | 0 |
| ND4 | 0.00467 | 0.00382 | 0.00217 |
| ND5 | 0.00249 | 0.00505 | 0.00414 |
| ND6 | 0 | 0.000772 | 0 |
| CO1 | 0.00194 | 0.00203 | 0.00064 |
| CO2 | 0.00122 | 0.00849 | 0.00145 |
| CO3 | 0.000423 | 0.0219 | 0.00837 |
| ATP8 | 0 | 0 | 0 |
| ATP6 | 0.00117 | 0.00563 | 0.0066 |
| CYTB | 0.00434 | 0.001706 | 0.00114 |

NAS (Northern Arabian Sea), SEAS (South Eastern Arabian Sea) and BoB (Bay of Bengal).

**Table 3.S7.** Seasonal Climatology of 3 ecoregions, Codons that are under positive selection in the mitogenome protein-coding genes, and the number of repeat units in the control region of *S. longiceps* populations from 3 ecoregions in the Indian Ocean.

| Seasonal Climatology | | | | | | Length polymorphisms in the control region | Codons that are under positive selection in the mitogenome protein-coding genes | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ecoregions | Season | Sea surface temperature (SST) $^0$C | Sea surface salinity (SSS) ppt | Dissolved oxygen (DO) mg/L | Chlorophyll a mg/m$^3$ | Type 1 with one repeat unit, Type 2 with two repeat unit, Type 3 with three repeat unit, | Complex I (12 sites) | | | | | | | Complex IV (8 sites) | | | Complex V (2 sites) | | Complex I I I (4 sites) |
| | | | | | | | ND1 (2 sites) | ND2 (1 sites) | ND 3 | ND4 L | ND4 (1 sites) | ND5 (8 sites) | ND 6 | CO1 (3 sites) | CO2 (3 sites) | CO3 (2 sites) | ATP8 | ATP6 (2 sites) | CYTB (4 sites) |
| NAS | Winter (JFM) | 24.0-27.0 | 36.0-38.0 | - | - | Type 1<br>Type 2 | - | #302 Leu-Gln | - | - | - | #226Thr-Asn | - | #114 Gly-Ala #262 Asn-Asp | #50 Leu-Gln | - | - | - | #70 Cys-Tyr |
| | Spring (AMJ) | 23.0-25.0 | 36.5-38.0 | 3.7-4.0 | 2.0-10.0 | | | | | | | | | | | | | | |
| | Summer (JAS) | 20.0-22.5 | 37-38 | 1.25-2.75 | 4.0-10.0 | | | | | | | | | | | | | | |
| | Fall (OND) | 21.0-22.5 | 36.5-38.0 | 2.5-2.75 | 2.0-5.0 | | | | | | | | | | | | | | |
| SEAS | Winter (JFM) | 28.0-28.5 | 33.0-35.0 | - | - | Type 1<br>Type 2<br>Type 3 | #29 Ile-Phe<br>#30 Glu-Leu | #302 Leu-Gln | - | - | #148 Thr-Asn | #97Ala-Gly<br>#98Leu-Val<br>#225Ala-Thr<br>#226Thr-Asn<br>#227Gly-Cys<br>#228Lys-Asn<br>#236 Pro-Ser | - | #25 Leu-Arg<br>#114 Gly-Ala #262 Asn-Asp | #152 Val-Ser | #16 Trp-Gly, Leu<br>#117 Pro-Leu | - | #114 Val-Leu | #70 Cys-Tyr, Trp |
| | Spring (AMJ) | 29.0-30.0 | 34.5 | 3.25-4.25 | 1.0-5.0 | | | | | | | | | | | | | | |
| | Summer (JAS) | 26.0-28.0 | 34.5 | 1.0-3.25 | 5.0-10.0 | | | | | | | | | | | | | | |
| | Fall (OND) | 28.0-29.0 | 34.5 | 1.0-1.25 | 2.0-3.0 | | | | | | | | | | | | | | |
| BoB | Winter (JFM) | 26.0-27.0 | 31.5-33.0 | - | - | Type 1 | - | - | - | - | - | #225 Ala-Thr<br>#227 Gly-Cys<br>#228 Lys-Asn<br>#236 Pro-Phe | - | #262 Asn-Asp | #63 Glu-Gly | #16 Trp-Arg<br>#117 Pro-Ser | - | #114 Val-Ala | #70 Cys-Tyr, Trp |
| | Spring (AMJ) | 29.0-30.0 | 31.5-33.0 | 4-5 | 0.0-2.0 | | | | | | | | | | | | | | |
| | Summer (JAS) | 28.5-30.0 | 29.5-33.0 | 3.0-4.25 | 1.0-3.0 | | | | | | | | | | | | | | |
| | Fall (OND) | 26.5-27.5 | 32.0-28.5 | 3.25-4.25 | 0.0-2.0 | | | | | | | | | | | | | | |

NAS (Northern Arabian Sea), SEAS (South Eastern Arabian Sea) and BoB (Bay of Bengal).

**Table 3.S8.** Codons that are under purifying selection in the mitogenome protein-coding genes of *S. longiceps.*

| NADH dehydrogenase subunits 1 (ND1)<br>No of sites 79 | | NADH dehydrogenase subunits 2 (ND2)<br>No of sites 94 | | Cytochrome c oxidase subunits 1 (COX1)<br>No of sites 58 | | Cytochrome c oxidase subunits 2 (COX2)<br>No of sites 23 | | ATPase subunits 8 (ATP8)<br>No of sites 3 | | ATPase subunits 6 (APT6)<br>No of sites 32 | | Cytochrome c oxidase subunits 3 (COX3)<br>No of sites 28 | | NADH dehydrogenase subunits 3 (ND3)<br>No of sites 13 | | NADH dehydrogenase subunits 4L (ND4L)<br>No of sites 9 | | NADH dehydrogenase subunits 4 (ND4)<br>No of sites 101 | | NADH dehydrogenase subunits 5 (ND5)<br>No of sites 143 | | NADH dehydrogenase subunits 6 (ND6)<br>No of sites 25 | | Cytochrome b (CYTB),<br>No of sites 73 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α | codon | Posterior Prob β<α |
| 291 | 0.935114 | 27 | 0.935 | 337 | 0.934 | 145 | 0.944 | 6 | 0.96 | 67 | 0.94 | 223 | 0.94 | 90 | 0.94 | 60 | 0.96 | 150 | 0.94 | 376 | 0.94 | 114 | 1 | 180 | 1 |
| 294 | 0.935412 | 295 | 0.935 | 129 | 0.937 | 141 | 0.949 | 11 | 0.96 | 178 | 0.94 | 120 | 0.94 | 103 | 0.95 | 43 | 0.96 | 328 | 0.94 | 214 | 0.94 | 157 | 1 | 100 | 1 |
| 26 | 0.938699 | 242 | 0.935 | 351 | 0.945 | 147 | 0.957 | 44 | 0.97 | 84 | 0.94 | 173 | 0.95 | 82 | 0.95 | 7 | 0.96 | 271 | 0.94 | 496 | 0.94 | 144 | 1 | 246 | 0.999 |
| 228 | 0.943783 | 324 | 0.944 | 224 | 0.946 | 185 | 0.961 | | | 134 | 0.94 | 21 | 0.95 | 39 | 0.95 | 59 | 0.96 | 43 | 0.94 | 29 | 0.94 | 166 | 1 | 105 | 0.999 |
| 87 | 0.944438 | 287 | 0.944 | 25 | 0.947 | 160 | 0.962 | | | 146 | 0.94 | 134 | 0.95 | 113 | 0.96 | 41 | 0.96 | 174 | 0.94 | 24 | 0.94 | 174 | 1 | 282 | 0.999 |
| 209 | 0.946838 | 107 | 0.944 | 328 | 0.947 | 143 | 0.962 | | | 97 | 0.95 | 202 | 0.96 | 32 | 0.96 | 88 | 0.97 | 115 | 0.94 | 135 | 0.94 | 164 | 1 | 251 | 0.999 |
| 308 | 0.947344 | 81 | 0.947 | 232 | 0.948 | 62 | 0.965 | | | 106 | 0.95 | 52 | 0.96 | 38 | 0.97 | 48 | 0.97 | 23 | 0.94 | 270 | 0.94 | 147 | 0.99 | 165 | 0.998 |
| 66 | 0.950452 | 113 | 0.947 | 128 | 0.951 | 133 | 0.965 | | | 65 | 0.95 | 169 | 0.96 | 89 | 0.98 | 45 | 0.99 | 443 | 0.94 | 323 | 0.94 | 142 | 0.99 | 55 | 0.998 |
| 73 | 0.951221 | 162 | 0.948 | 175 | 0.951 | 198 | 0.966 | | | 121 | 0.95 | 185 | 0.96 | 110 | 0.98 | 2 | 1 | 408 | 0.94 | 227 | 0.94 | 118 | 0.99 | 120 | 0.998 |
| 153 | 0.957453 | 114 | 0.948 | 436 | 0.951 | 108 | 0.966 | | | 82 | 0.96 | 67 | 0.96 | 79 | 0.99 | | | 292 | 0.94 | 479 | 0.94 | 126 | 0.99 | 369 | 0.997 |
| 126 | 0.957512 | 320 | 0.949 | 316 | 0.955 | 211 | 0.967 | | | 166 | 0.96 | 140 | 0.96 | 80 | 0.99 | | | 189 | 0.95 | 150 | 0.95 | 156 | 0.99 | 113 | 0.994 |
| 222 | 0.957603 | 173 | 0.949 | 270 | 0.957 | 174 | 0.969 | | | 88 | 0.96 | 239 | 0.96 | 41 | 1 | | | 88 | 0.95 | 167 | 0.95 | 130 | 0.97 | 159 | 0.994 |
| 95 | 0.957955 | 341 | 0.951 | 144 | 0.957 | 119 | 0.969 | | | 138 | 0.96 | 144 | 0.97 | 28 | 1 | | | 190 | 0.95 | 410 | 0.95 | 173 | 0.97 | 361 | 0.994 |
| 137 | 0.958009 | 223 | 0.951 | 413 | 0.957 | 146 | 0.97 | | | 20 | 0.96 | 116 | 0.97 | | | | | 360 | 0.95 | 271 | 0.95 | 143 | 0.97 | 127 | 0.993 |
| 28 | 0.958023 | 120 | 0.957 | 75 | 0.958 | 36 | 0.971 | | | 116 | 0.96 | 174 | 0.97 | | | | | 255 | 0.95 | 306 | 0.95 | 125 | 0.97 | 275 | 0.993 |
| 108 | 0.959467 | 277 | 0.957 | 267 | 0.958 | 57 | 0.971 | | | 64 | 0.96 | 201 | 0.97 | | | | | 347 | 0.95 | 567 | 0.95 | 170 | 0.97 | 115 | 0.993 |
| 183 | 0.960141 | 167 | 0.958 | 192 | 0.958 | 70 | 0.973 | | | 34 | 0.96 | 161 | 0.97 | | | | | 163 | 0.96 | 209 | 0.95 | 152 | 0.97 | 337 | 0.993 |
| 212 | 0.960159 | 256 | 0.958 | 195 | 0.958 | 167 | 0.974 | | | 14 | 0.96 | 117 | 0.97 | | | | | 194 | 0.96 | 535 | 0.95 | 167 | 0.97 | 147 | 0.992 |
| 255 | 0.960179 | 132 | 0.959 | 261 | 0.958 | 59 | 0.988 | | | 111 | 0.97 | 198 | 0.97 | | | | | 171 | 0.96 | 109 | 0.95 | 104 | 0.96 | 299 | 0.992 |
| 314 | 0.960179 | 78 | 0.959 | 379 | 0.959 | 105 | 0.991 | | | 60 | 0.97 | 258 | 0.98 | | | | | 420 | 0.96 | 315 | 0.95 | 172 | 0.96 | 194 | 0.991 |
| 96 | 0.96034 | 84 | 0.96 | 335 | 0.96 | 203 | 0.998 | | | 71 | 0.97 | 23 | 0.99 | | | | | 63 | 0.96 | 91 | 0.95 | 133 | 0.96 | 141 | 0.99 |
| 284 | 0.961381 | 119 | 0.961 | 171 | 0.961 | 137 | 0.998 | | | 110 | 0.97 | 136 | 0.99 | | | | | 349 | 0.96 | 524 | 0.95 | 127 | 0.96 | 81 | 0.989 |
| 272 | 0.963232 | 272 | 0.961 | 482 | 0.961 | 148 | 0.999 | | | 29 | 0.97 | 162 | 0.99 | | | | | 314 | 0.96 | 498 | 0.95 | 100 | 0.96 | 128 | 0.986 |
| 91 | 0.963504 | 181 | 0.961 | 305 | 0.961 | | | | | 171 | 0.97 | 147 | 0.99 | | | | | 202 | 0.96 | 241 | 0.95 | 150 | 0.96 | 213 | 0.984 |
| 239 | 0.96355 | 294 | 0.961 | 213 | 0.963 | | | | | 152 | 0.97 | 123 | 0.99 | | | | | 366 | 0.96 | 565 | 0.96 | 119 | 0.96 | 326 | 0.981 |

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 166 | 0.963688 | 60 | 0.964 | 102 | 0.964 | | | | | 55 | 0.98 | 70 | 1 | | | | | 440 | 0.96 | 46 | 0.96 | | | 266 | 0.979 |
| 99 | 0.96391 | 259 | 0.964 | 392 | 0.964 | | | | | 135 | 0.99 | 93 | 1 | | | | | 425 | 0.96 | 191 | 0.96 | | | 109 | 0.978 |
| 7 | 0.964355 | 229 | 0.964 | 414 | 0.964 | | | | | 164 | 0.99 | 64 | 1 | | | | | 19 | 0.96 | 355 | 0.96 | | | 293 | 0.978 |
| 142 | 0.964745 | 58 | 0.964 | 317 | 0.966 | | | | | 63 | 0.99 | | | | | | | 81 | 0.96 | 612 | 0.96 | | | 258 | 0.974 |
| 22 | 0.966039 | 176 | 0.964 | 209 | 0.966 | | | | | 15 | 1 | | | | | | | 311 | 0.96 | 460 | 0.96 | | | 156 | 0.971 |
| 172 | 0.966459 | 24 | 0.965 | 314 | 0.966 | | | | | 208 | 1 | | | | | | | 164 | 0.96 | 53 | 0.96 | | | 237 | 0.971 |
| 136 | 0.968353 | 280 | 0.965 | 2 | 0.967 | | | | | 204 | 1 | | | | | | | 343 | 0.96 | 399 | 0.96 | | | 256 | 0.97 |
| 193 | 0.968412 | 71 | 0.965 | 349 | 0.967 | | | | | | | | | | | | | 422 | 0.96 | 389 | 0.96 | | | 30 | 0.97 |
| 33 | 0.968926 | 319 | 0.965 | 373 | 0.968 | | | | | | | | | | | | | 184 | 0.96 | 154 | 0.96 | | | 252 | 0.97 |
| 315 | 0.968926 | 225 | 0.965 | 255 | 0.968 | | | | | | | | | | | | | 38 | 0.96 | 324 | 0.96 | | | 347 | 0.969 |
| 229 | 0.969505 | 237 | 0.965 | 474 | 0.968 | | | | | | | | | | | | | 415 | 0.96 | 198 | 0.96 | | | 95 | 0.969 |
| 236 | 0.969822 | 275 | 0.966 | 141 | 0.969 | | | | | | | | | | | | | 26 | 0.96 | 470 | 0.96 | | | 260 | 0.969 |
| 14 | 0.970518 | 257 | 0.966 | 319 | 0.969 | | | | | | | | | | | | | 134 | 0.97 | 245 | 0.96 | | | 228 | 0.969 |
| 24 | 0.970518 | 332 | 0.966 | 333 | 0.97 | | | | | | | | | | | | | 298 | 0.97 | 122 | 0.96 | | | 119 | 0.969 |
| 238 | 0.972163 | 108 | 0.966 | 339 | 0.971 | | | | | | | | | | | | | 132 | 0.97 | 260 | 0.96 | | | 288 | 0.968 |
| 11 | 0.974071 | 95 | 0.966 | 14 | 0.972 | | | | | | | | | | | | | 339 | 0.97 | 332 | 0.96 | | | 306 | 0.968 |
| 200 | 0.974132 | 131 | 0.966 | 97 | 0.974 | | | | | | | | | | | | | 281 | 0.97 | 381 | 0.96 | | | 44 | 0.967 |
| 216 | 0.979368 | 307 | 0.966 | 470 | 0.978 | | | | | | | | | | | | | 283 | 0.97 | 126 | 0.96 | | | 151 | 0.966 |
| 296 | 0.979446 | 172 | 0.966 | 200 | 0.978 | | | | | | | | | | | | | 414 | 0.97 | 452 | 0.96 | | | 178 | 0.966 |
| 109 | 0.980475 | 282 | 0.966 | 456 | 0.981 | | | | | | | | | | | | | 41 | 0.97 | 531 | 0.96 | | | 297 | 0.966 |
| 69 | 0.980659 | 145 | 0.967 | 396 | 0.982 | | | | | | | | | | | | | 309 | 0.97 | 451 | 0.96 | | | 196 | 0.966 |
| 290 | 0.981118 | 179 | 0.967 | 505 | 0.986 | | | | | | | | | | | | | 83 | 0.97 | 433 | 0.96 | | | 212 | 0.966 |
| 285 | 0.982982 | 157 | 0.967 | 211 | 0.989 | | | | | | | | | | | | | 371 | 0.97 | 105 | 0.96 | | | 231 | 0.966 |
| 60 | 0.983604 | 77 | 0.968 | 355 | 0.991 | | | | | | | | | | | | | 140 | 0.97 | 275 | 0.96 | | | 33 | 0.965 |
| 102 | 0.983826 | 85 | 0.968 | 189 | 0.991 | | | | | | | | | | | | | 188 | 0.97 | 106 | 0.96 | | | 160 | 0.965 |
| 268 | 0.985148 | 208 | 0.968 | 480 | 0.993 | | | | | | | | | | | | | 18 | 0.97 | 326 | 0.96 | | | 122 | 0.964 |
| 114 | 0.985734 | 69 | 0.968 | 205 | 0.994 | | | | | | | | | | | | | 405 | 0.97 | 102 | 0.96 | | | 204 | 0.964 |
| 198 | 0.986421 | 92 | 0.968 | 422 | 0.994 | | | | | | | | | | | | | 66 | 0.97 | 311 | 0.97 | | | 248 | 0.963 |
| 164 | 0.987846 | 329 | 0.969 | 350 | 0.994 | | | | | | | | | | | | | 381 | 0.97 | 485 | 0.97 | | | 270 | 0.962 |
| 206 | 0.988031 | 292 | 0.969 | 123 | 0.995 | | | | | | | | | | | | | 441 | 0.97 | 583 | 0.97 | | | 322 | 0.961 |
| 135 | 0.991138 | 268 | 0.969 | 203 | 0.998 | | | | | | | | | | | | | 121 | 0.97 | 284 | 0.97 | | | 314 | 0.961 |
| 143 | 0.991212 | 192 | 0.969 | 401 | 0.999 | | | | | | | | | | | | | 312 | 0.97 | 390 | 0.97 | | | 245 | 0.961 |
| 233 | 0.99246 | 200 | 0.971 | 387 | 0.999 | | | | | | | | | | | | | 91 | 0.97 | 143 | 0.97 | | | 15 | 0.961 |

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | | | | | | | | | | | | | | | | | | | | | | | 5 | |
| 149 | 0.992889 | 300 | 0.971 | | | | | | | | | | | | | | | 240 | 0.97 | 409 | 0.97 | | | 28 | 0.96 |
| 176 | 0.993387 | 121 | 0.973 | | | | | | | | | | | | | | | 141 | 0.97 | 100 | 0.97 | | | 190 | 0.96 |
| 50 | 0.993754 | 128 | 0.973 | | | | | | | | | | | | | | | 430 | 0.97 | 96 | 0.97 | | | 117 | 0.959 |
| 92 | 0.994485 | 343 | 0.978 | | | | | | | | | | | | | | | 389 | 0.97 | 107 | 0.97 | | | 145 | 0.959 |
| 243 | 0.994657 | 233 | 0.98 | | | | | | | | | | | | | | | 388 | 0.97 | 371 | 0.97 | | | 129 | 0.959 |
| 185 | 0.994897 | 146 | 0.986 | | | | | | | | | | | | | | | 320 | 0.97 | 196 | 0.97 | | | 255 | 0.958 |
| 263 | 0.995027 | 199 | 0.986 | | | | | | | | | | | | | | | 55 | 0.98 | 272 | 0.97 | | | 161 | 0.957 |
| 16 | 0.995153 | 168 | 0.988 | | | | | | | | | | | | | | | 456 | 0.98 | 228 | 0.97 | | | 46 | 0.957 |
| 292 | 0.996801 | 283 | 0.988 | | | | | | | | | | | | | | | 274 | 0.98 | 88 | 0.97 | | | 316 | 0.95 |
| 168 | 0.997038 | 337 | 0.99 | | | | | | | | | | | | | | | 331 | 0.98 | 162 | 0.97 | | | 356 | 0.949 |
| 259 | 0.997266 | 303 | 0.99 | | | | | | | | | | | | | | | 304 | 0.98 | 325 | 0.97 | | | 195 | 0.947 |
| 287 | 0.99816 | 227 | 0.991 | | | | | | | | | | | | | | | 120 | 0.98 | 220 | 0.97 | | | 307 | 0.944 |
| 248 | 0.998406 | 211 | 0.991 | | | | | | | | | | | | | | | 178 | 0.98 | 62 | 0.97 | | | 253 | 0.944 |
| 174 | 0.998501 | 297 | 0.991 | | | | | | | | | | | | | | | 162 | 0.98 | 351 | 0.97 | | | 126 | 0.944 |
| 215 | 0.99866 | 64 | 0.991 | | | | | | | | | | | | | | | 323 | 0.99 | 43 | 0.97 | | | 279 | 0.935 |
| 46 | 0.998967 | 74 | 0.992 | | | | | | | | | | | | | | | 183 | 0.99 | 117 | 0.97 | | | | |
| 191 | 0.999381 | 104 | 0.992 | | | | | | | | | | | | | | | 94 | 0.99 | 331 | 0.97 | | | | |
| 163 | 0.999707 | 90 | 0.993 | | | | | | | | | | | | | | | 387 | 0.99 | 553 | 0.97 | | | | |
| 18 | 0.999932 | 253 | 0.993 | | | | | | | | | | | | | | | 85 | 0.99 | 281 | 0.97 | | | | |
| 122 | 0.999986 | 308 | 0.994 | | | | | | | | | | | | | | | 11 | 0.99 | 456 | 0.97 | | | | |
| 165 | 0.999993 | 56 | 0.994 | | | | | | | | | | | | | | | 336 | 0.99 | 571 | 0.97 | | | | |
| | | 125 | 0.994 | | | | | | | | | | | | | | | 181 | 0.99 | 181 | 0.97 | | | | |
| | | 45 | 0.994 | | | | | | | | | | | | | | | 102 | 0.99 | 278 | 0.97 | | | | |
| | | 180 | 0.994 | | | | | | | | | | | | | | | 170 | 0.99 | 305 | 0.97 | | | | |
| | | 139 | 0.994 | | | | | | | | | | | | | | | 231 | 0.99 | 296 | 0.97 | | | | |
| | | 163 | 0.996 | | | | | | | | | | | | | | | 125 | 0.99 | 412 | 0.97 | | | | |
| | | 129 | 0.996 | | | | | | | | | | | | | | | 60 | 1 | 577 | 0.97 | | | | |
| | | 54 | 0.996 | | | | | | | | | | | | | | | 34 | 1 | 476 | 0.97 | | | | |
| | | 210 | 0.997 | | | | | | | | | | | | | | | 383 | 1 | 539 | 0.97 | | | | |
| | | 239 | 0.997 | | | | | | | | | | | | | | | 384 | 1 | 193 | 0.97 | | | | |
| | | 164 | 0.998 | | | | | | | | | | | | | | | 452 | 1 | 165 | 0.97 | | | | |
| | | 143 | 0.999 | | | | | | | | | | | | | | | 380 | 1 | 386 | 0.97 | | | | |
| | | 298 | 0.999 | | | | | | | | | | | | | | | 365 | 1 | 328 | 0.97 | | | | |
| | | 150 | 1 | | | | | | | | | | | | | | | 52 | 1 | 202 | 0.97 | | | | |
| | | 59 | 1 | | | | | | | | | | | | | | | 247 | 1 | 121 | 0.97 | | | | |
| | | 116 | 1 | | | | | | | | | | | | | | | 159 | 1 | 204 | 0.97 | | | | |
| | | | | | | | | | | | | | | | | | | 259 | 1 | 533 | 0.97 | | | | |
| | | | | | | | | | | | | | | | | | | 299 | 1 | 366 | 0.97 | | | | |
| | | | | | | | | | | | | | | | | | | 315 | 1 | 224 | 0.98 | | | | |
| | | | | | | | | | | | | | | | | | | 288 | 1 | 352 | 0.98 | | | | |
| | | | | | | | | | | | | | | | | | | 44 | 1 | 525 | 0.98 | | | | |
| | | | | | | | | | | | | | | | | | | 437 | 1 | 180 | 0.98 | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | 59 | 1 | 382 | 0.98 | | | | |
| | | | | | | | | | | | | | | | | | | | | 112 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 141 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 609 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 276 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 277 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 279 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 36 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 550 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 183 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 443 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 566 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 463 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 142 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 213 | 0.99 | | | | |
| | | | | | | | | | | | | | | | | | | | | 482 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 551 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 584 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 572 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 208 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 422 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 273 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 395 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 166 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 429 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 94 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 203 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 147 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 589 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 327 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 339 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 500 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 595 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 491 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 108 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 185 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 188 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 449 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 557 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 514 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 486 | 1 | | | | |
| | | | | | | | | | | | | | | | | | | | | 189 | 1 | | | | |

The analysis is performed in FUBAR at posterior probability ≥ 0.9.

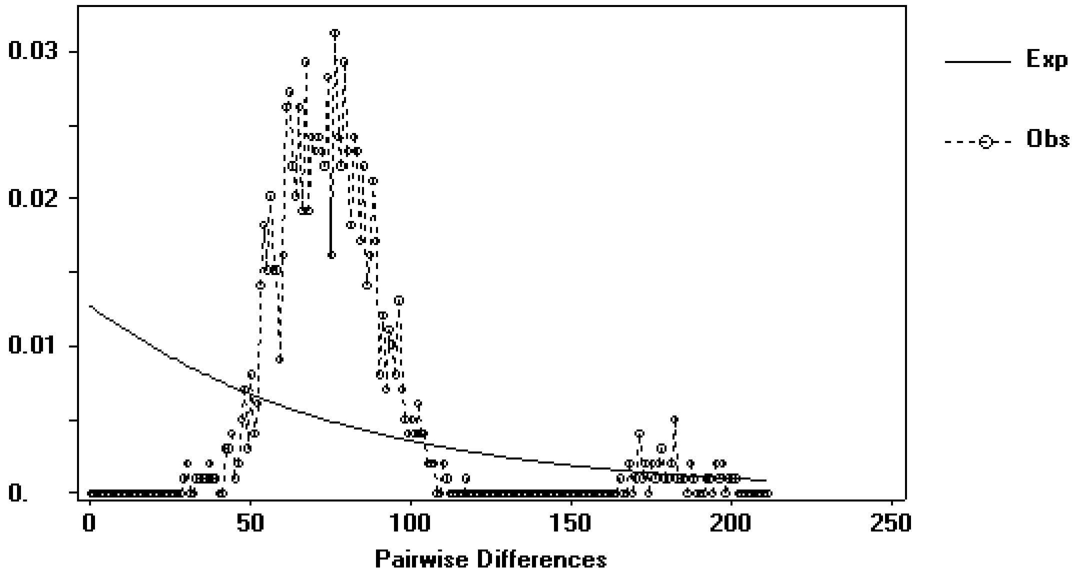


**Figure 3.S1** Mismatch analysis plot for whole-genome nucleotide sequences of *S. longiceps*.

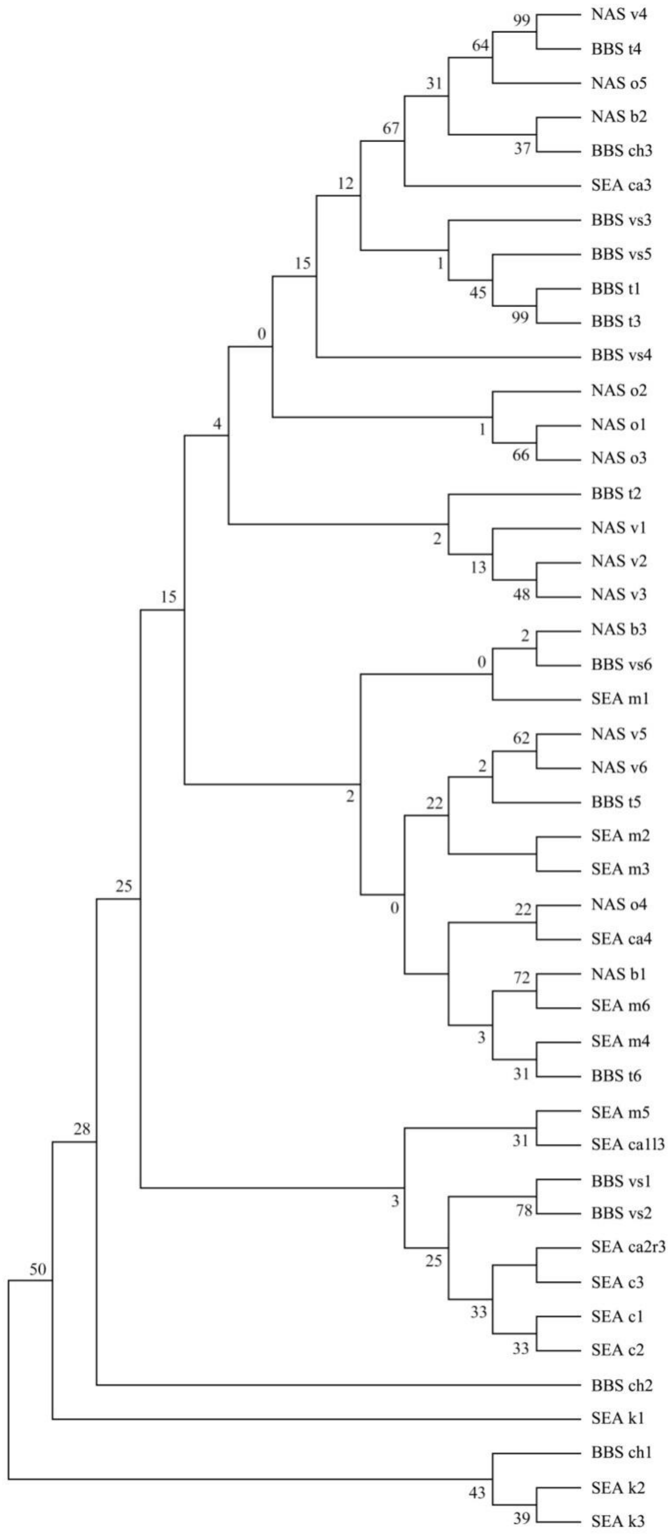


**Fig 3.S2** Neighbour-joining tree for whole mitogenome nucleotide sequences of *S. longiceps*. Bootstrap values for node support are shown.

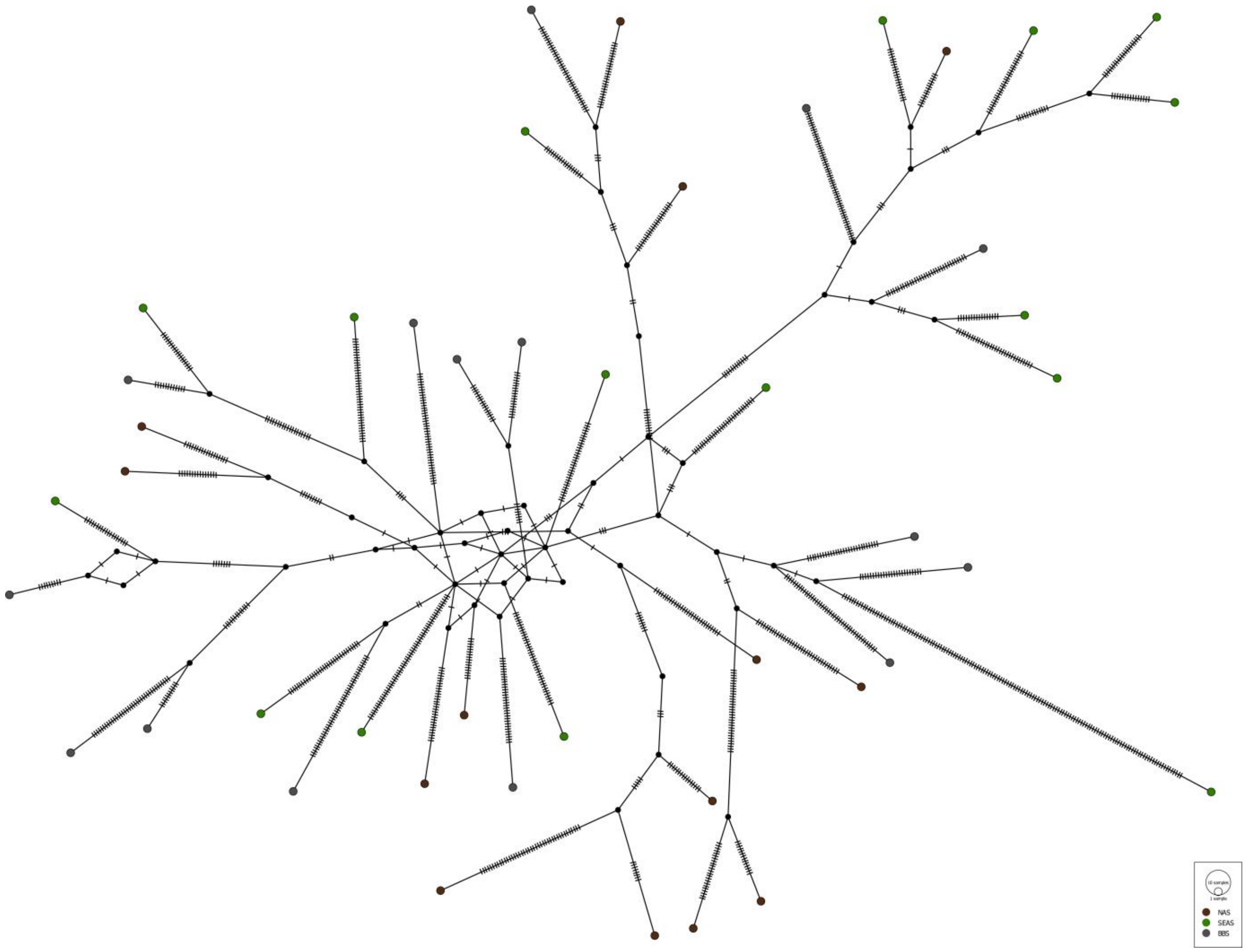


**Fig 3.S3** Median-joining haplotype network tree of whole mitogenome sequences of 45 *S. longiceps*. Haplotypes are represented in circles and mutational steps are indicated as hatch marks.

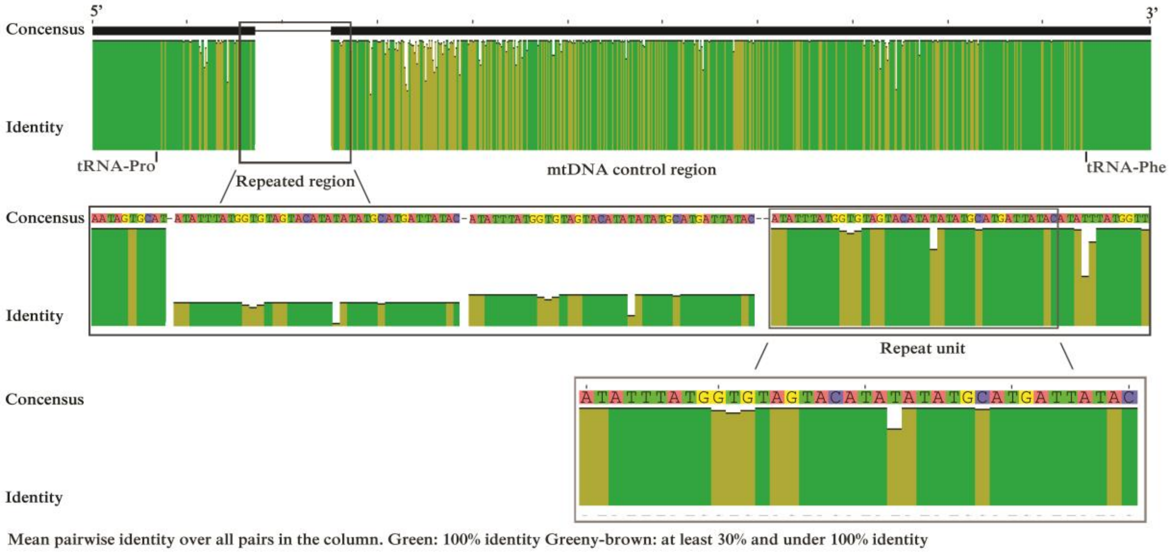


**Fig 3.S4** Schematic representation of the *S. longiceps* mtDNA region of ~1112bp comprising tRNA pro, control region and tRNA phe.

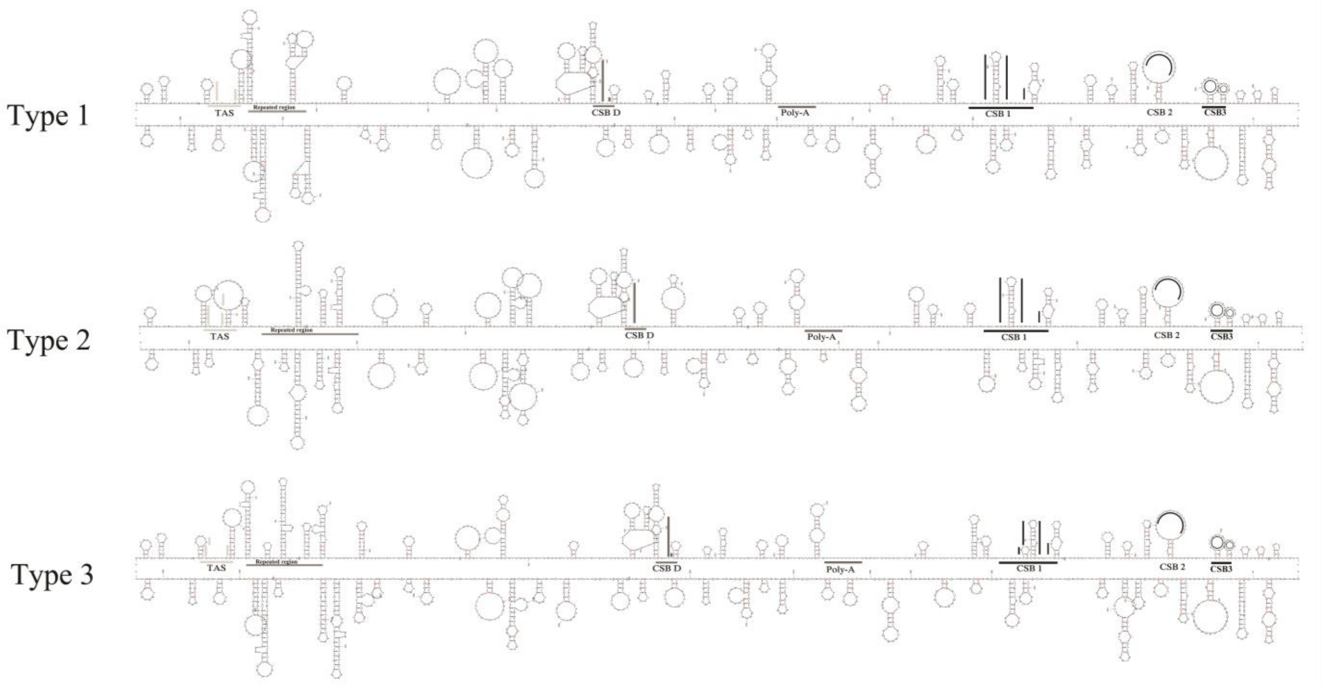


**Fig 3.S5 Secondary structures identified in the mtDNA control region of *S. longiceps.***mtDNA control region haplotype with Type 1, 2 and 3 repeat units are indicated as Type 1, Type 2 and Type 3 respectively.

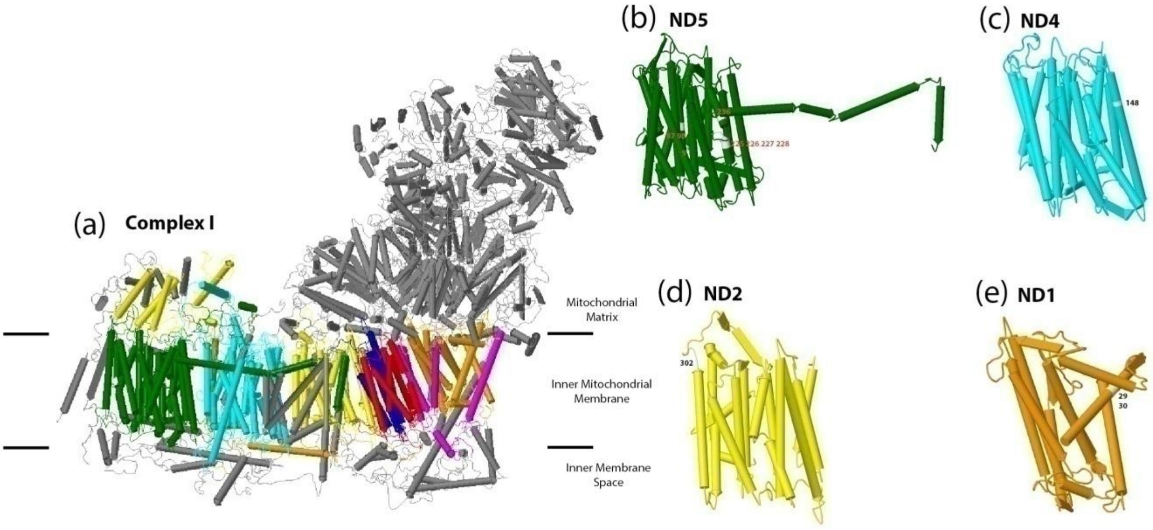


**Fig 3.S6 Spatial distribution of positively selected sites identified in NADH dehydrogenase (Complex I) of *S. longiceps*.** Grey structures represent nuclear-encoded subunits. (a) individual OXPHOS Complex I, with mitochondrial-encoded subunits are represented in different coloured as followed: ND2 in yellow; ND4L in blue; ND1 in orange; ND3 in magenta; ND4 in cyan; ND5 in green; ND6 in red. Individual core subunits (b) ND5, (c) ND4, (d) ND2, (e) ND1with amino acid site number on positively selected sites.

ND1, ND2, ND4 and ND5 subunit proteins showed 75% identity with Chain H (PDB: 5LDX_H), 50% identity with Chain N (PDB: 5LDX_N), 62% identity with Chain M (PDB: 5LDX_M) and 63% identity with Chain L (PDB: 5LDX_L) respectively of *Bos taurus* Respiratory Complex I (Zhu *et al.* 2016). Mitochondrial complex I (NADH: ubiquinone oxidoreductase) contributes to cellular energy production by transferring electrons from NADH to ubiquinone coupled to proton translocation across the membrane. The Key polar amino acid residues which have been reported to participate in proton translocation (ND1 - E198, E149, ND2 - K263, K135, K105, E34, ND4 - E124, K238, E379, K208, ND5 -E149, H253, K397) (Zhu *et al.* 2016) through complex I were conserved in sardine except site 228 in ND5. Site 29ND1, 30ND1, 302ND2, 148ND4, 9ND5, 97ND5, 98ND5, 225ND5, 226ND5, 227ND5, 228ND5 and 236ND5 were identified as positively selected in *S. longiceps* and all of them were located in transmembrane helices except one which is in intra-helix loop (228ND5). Nine of these sites one in ND2 (#302Leu-Gln) were located in C-terminus, one in ND4 (#148Thr-asn) located in proton-conducting membrane transporter (Proton_antipo_M) and seven in ND5 (#97Ala-Gly, #98Leu-Val, #225Ala-Thr, #226Thr-Asn, #227Gly-Cys, #228Lys-Asn & #236Pro-Ser) clustered in Proton_antipo_M & N-terminal (Proton_antipo_N). Position 228 (ND5) showed overlap with amino acid residue that has been reported as one of the key residues in proton translocation.

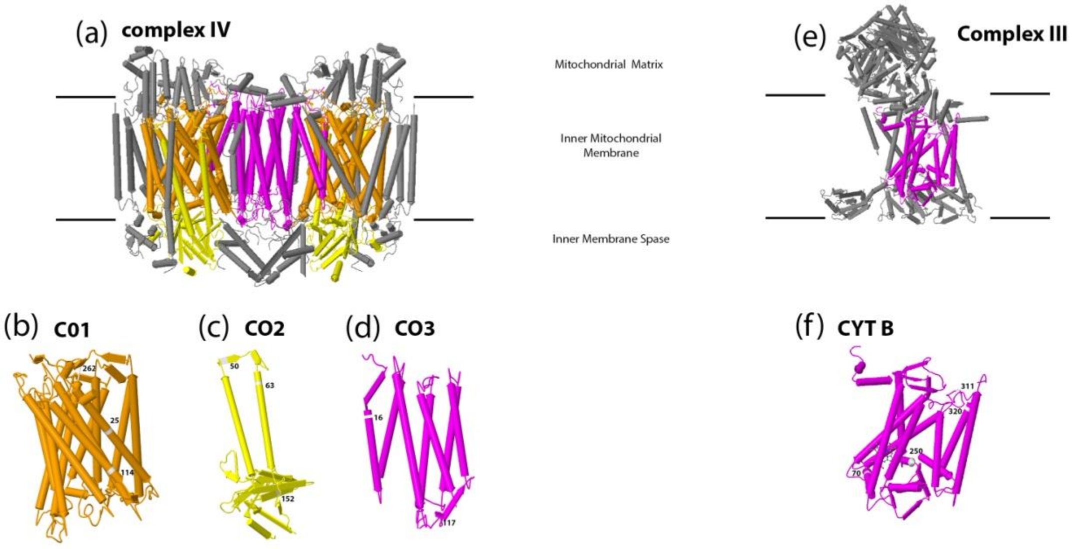


**Fig 3.S7 Spatial distribution of positively selected sites identified in Cytochrome C Oxidase (Complex IV) and Cytochrome bc 1 (Complex III) of *S. longiceps***. Grey structures represent nuclear-encoded subunits. Individual OXPHOS Complex IV (Homodimer) (a) with mitochondrial-encoded subunits is represented in different colours as followed: CO1 in orange; CO2 in yellow; CO3 in magenta. Individual OXPHOS Complex III (e), with mitochondrial-encoded subunit represented in magenta colour. Individual core subunits (b) CO1, (c) CO2, (d) CO3, (f) CYT B with amino acid site number at positively selected sites.

CO1, CO2 and CO3 of *S. longiceps* showed 89% identity with Chain N (PDB: 2OCC_N), 73% identity with Chain B (PDB: 2OCC_B) and 80% identity with Chain C (PDB: 2OCC_C) of *B. taurus* Cytochrome C Oxidase (CcO) respectively (Tsukihara *et al.* 1996). The amino acid residues that have been reported to participate in Electron transfer pathway (F377, R438, R439), D-pathway (Y19, N80, D91, N98, S101, S156, S157, N163, T167), Putative water exit pathway (D227, G232, H233, D364, H368, D369, R438), Ion binding (Binuclear center-heme a3/CuB) (H240, H290, H291, H376), K-pathway (H240, Y244, S255, H290, H291, T316, K319), Putative proton exit pathway (H291, H368, D369, R438, R439), and chemical binding (Low-spin heme a binding site) (H61, H378, S382, T424, S461) in CO1 (Tsukihara *et al.* 1995) were conserved. Three sites (#25Leu-Arg, #114 Gly-Ala and #262Asn-asp) observed under positive selection in CO1 were located in the transmembrane helix and two of these positions (#25 & #114) showed overlap with amino acid residue that have been reported to participate in polypeptide binding at Subunit I/VIIc interface & Subunit I/VIIa interface respectively. Amino acid residues that have been reported to participate in CuA binding site in CO2 and most of the amino acid participated in polypeptide binding & Phospholipid binding in CcO is conserved in *S. longiceps*. Among three sites observed under positive selection in CO2 gene, amino acid position 50 (Leu-gln) reside in the intra-helix loop, position 63 (Glu-gly) in transmembrane helix and 152 (Val-ser) in Beta strand. Among the two sites identified in CO3, position 16 (Trp-Gly) were located in transmembrane helix and position 117 (Pro-Leu,Ser) in the intra-helix loop.

Cytochrome b of *S. longiceps* showed 75% identity with Chain b (PDB: 5LUF_b) of *B. taurus* (Sousa *et al.* 2016). Majority of amino acid sites that have been suggested to participate in Qo binding, Qi binding and chemical binding were conserved. Among four sites (#70Cys-Trp, #250Leu-Gln, #311 Lys-Glnand #320Leu-Ile) that experienced positive selection in CYTB, one (#311) showed overlap with amino acid residue that has been reported to participate in polypeptide binding in inter-chain domain interface and it was located in the transmembrane helix.

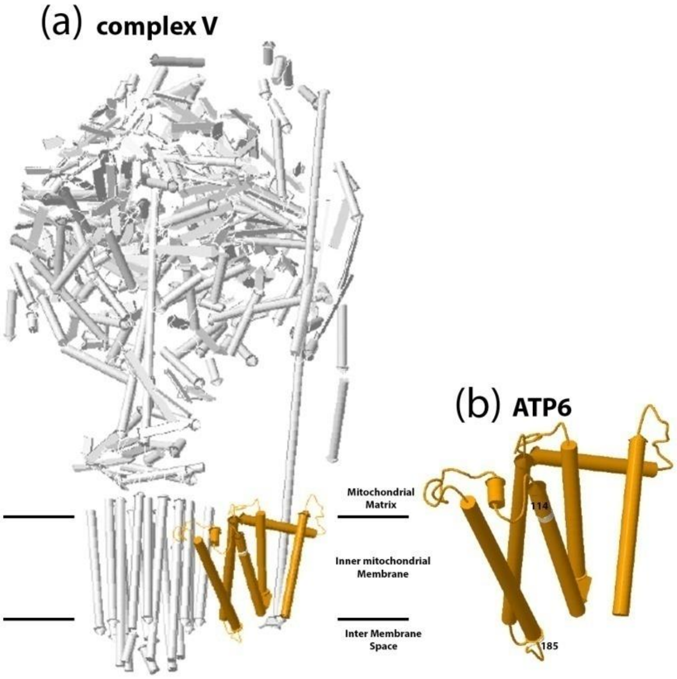


**Fig 3.S8 Spatial distribution of positively selected sites identified in ATP synthase (complex V) of *S. longiceps.***Grey structures represent nuclear-encoded subunits. (a) Individual OXPHOS Complex V, with mitochondrial-encoded subunit in orange colour. (b) Individual core subunits ATP 6 with amino acid site number on positively selected sites.

ATP 6 of *S. longiceps* showed 54% identity with Chain W (PDB: 5ARA_W) of *B. taurus* mitochondrial ATP Synthase (Zhou *et al.* 2015). The highly conserved residue Arg159 in ATP Synthase a subunit showed overlap with Arg at site 160 (middle of helix 5). Among two sites (#114 Val-cys, ala #185 Ile-gln) observed under positive selection, one (#114) was located in the transmembrane helix-4 and other (#185) in intra-helix loop connecting helix-5 and 6.

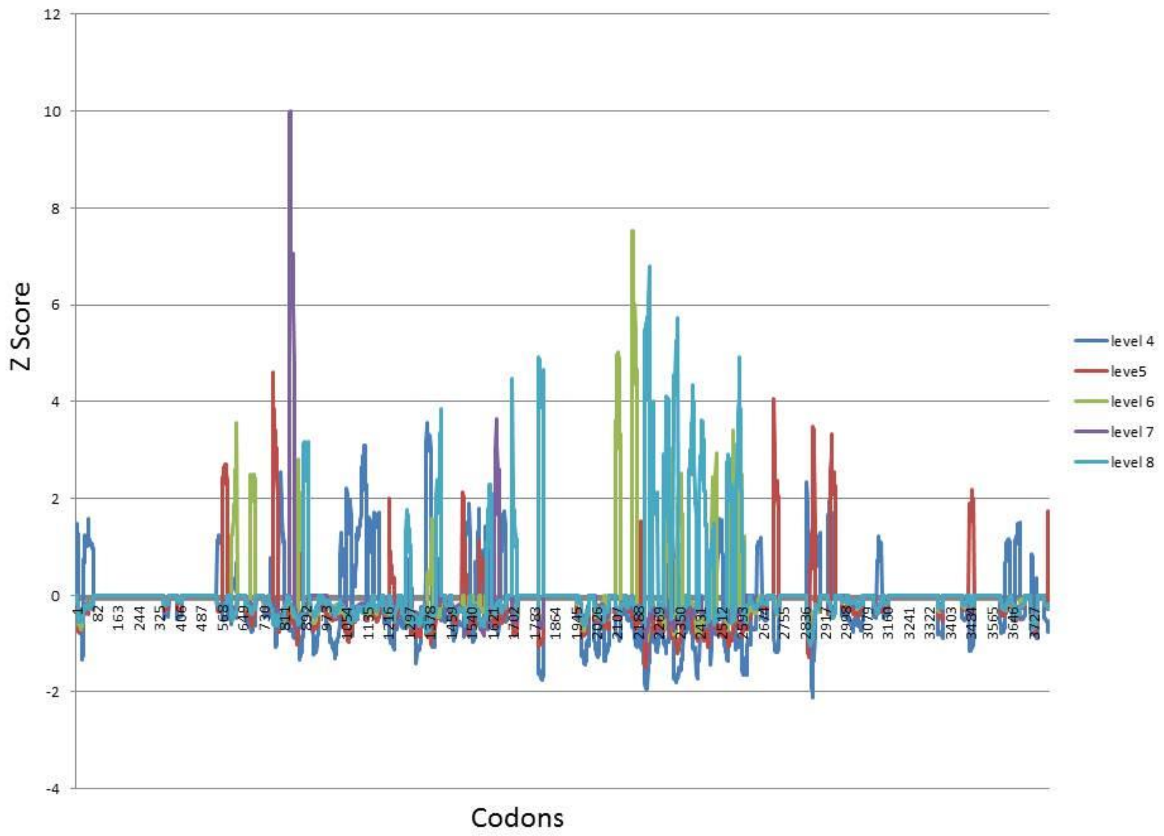
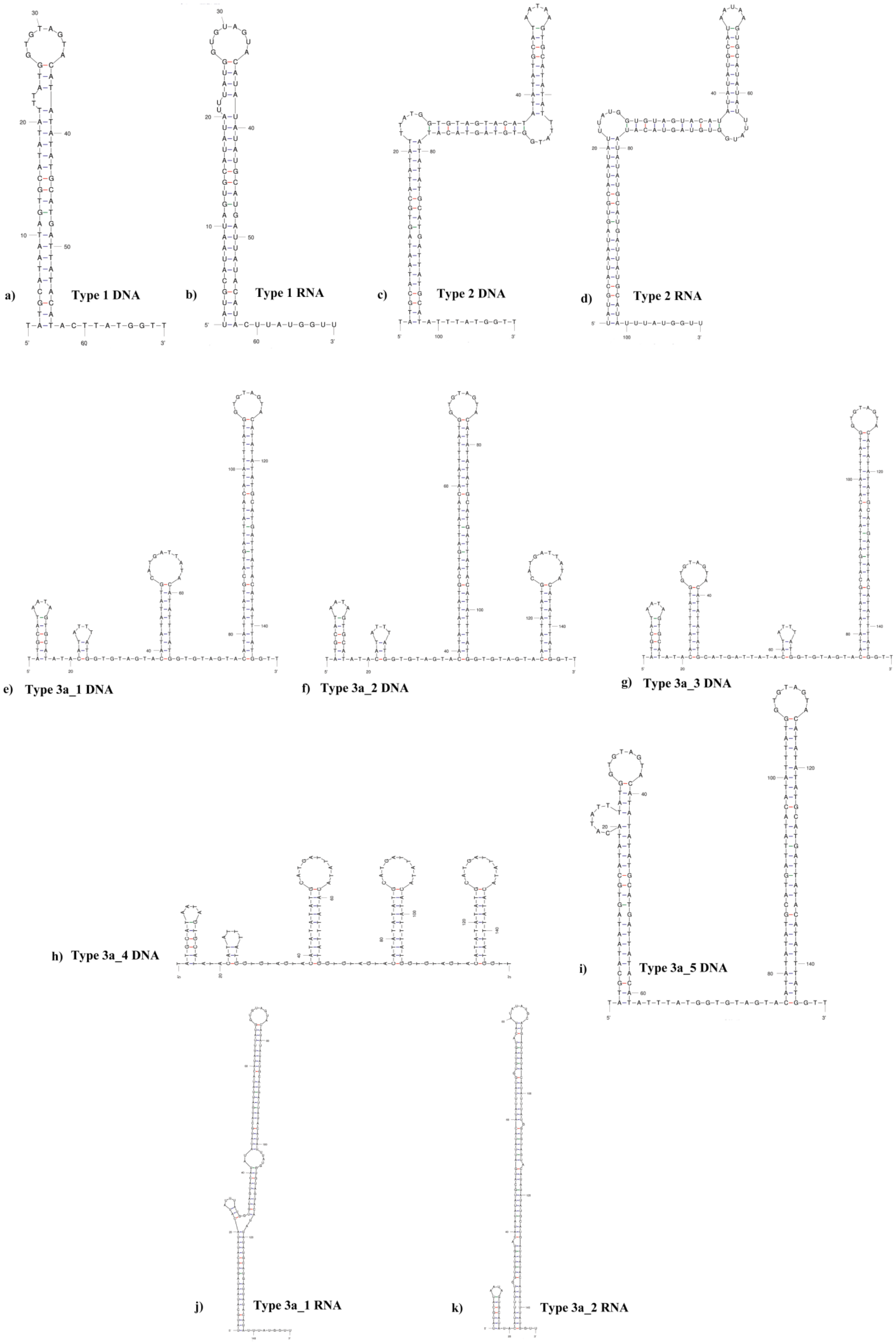


**Fig 3.S9 TreeSAAP results showing the region of the mitochondrial genome under positive disruptive selection.** The Z Score shown with horizontal lines, and vertical numerical number indicate amino acid positions in concatenated gene data set. Category of amino acid physiochemical property changes is represented as level 4 to level 8. Amino acid position of each coding gene in the concatenated gene data set: 1466-1692 ATPase subunits 6 (APT6), 1410-1465 ATPase subunits 8 (ATP8) 675-1179 Cytochrome c oxidase subunits 1 (COX1) 1180-1409 Cytochrome c oxidase subunits 2 (COX2) 1693-1954 Cytochrome c oxidase subunits 3 (COX3) 3416-3795 Cytochrome b (CYTB), 1-325 NADH dehydrogenase subunits 1 (ND1), 326-674 NADH dehydrogenase subunits 2 (ND2), 1955-2070 NADH dehydrogenase subunits 3 (ND3), 2071-2169 NADH dehydrogenase subunits 4L (ND4L), 2170-2629 NADH dehydrogenase subunits 4 (ND4), 2630-3241 NADH dehydrogenase subunits 5 (ND5), 3242-3415 NADH dehydrogenase subunits 6 (ND6).

A



a) Type 1 DNA

b) Type 1 RNA

c) Type 2 DNA

d) Type 2 RNA

e) Type 3a_1 DNA

f) Type 3a_2 DNA

g) Type 3a_3 DNA

h) Type 3a_4 DNA

i) Type 3a_5 DNA

j) Type 3a_1 RNA

k) Type 3a_2 RNA

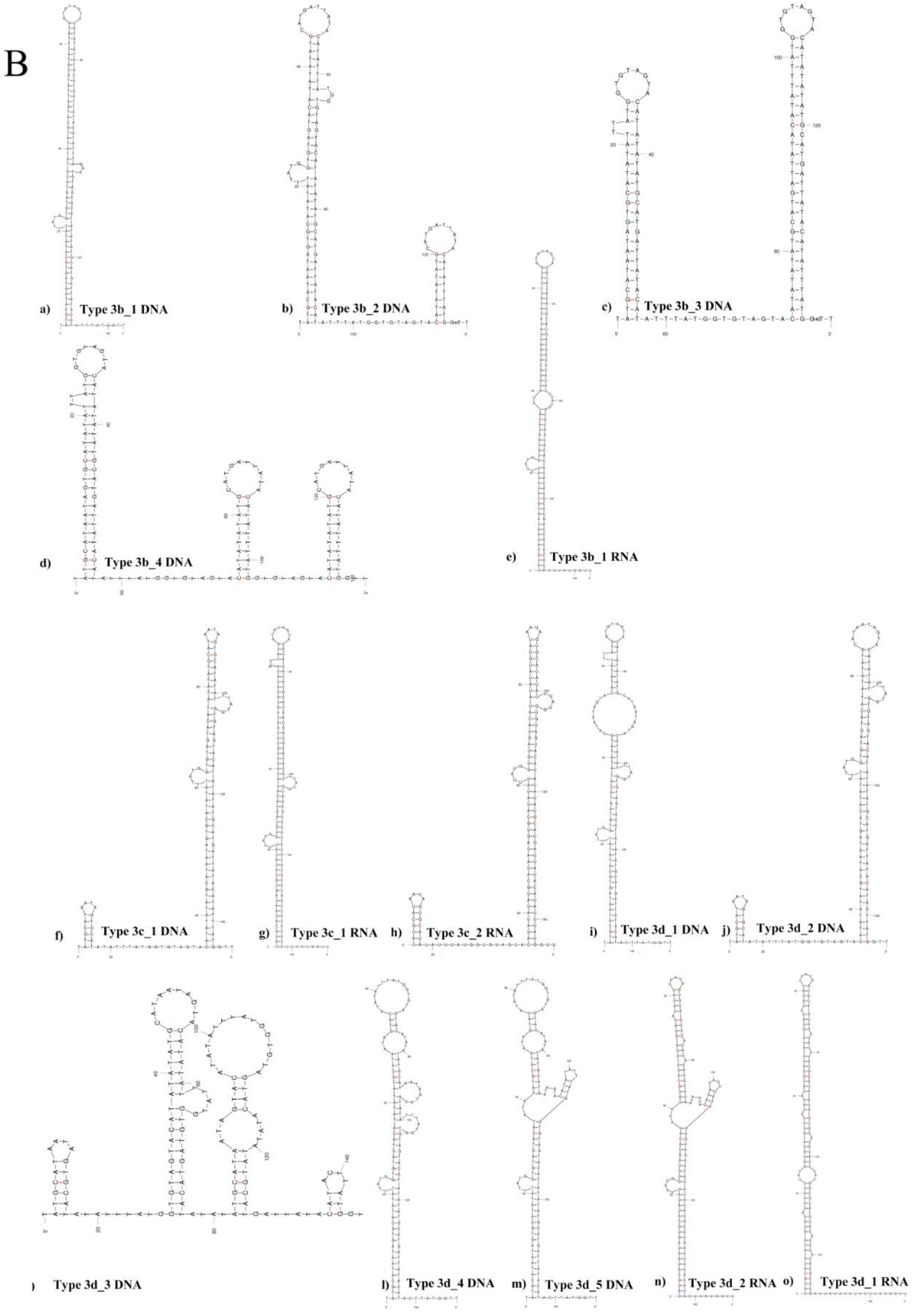


**Figure 3.S10** Graphical representation of all predicted secondary structures in repeat unit Type 1, 2 and 3 of mtDNA controlregion DNA and the same for RNA. In section A: a) DNA of haplotype with Type 1 repeat unit, b) RNA of haplotype with Type 1 repeat unit, c) DNA of haplotype with Type 2 repeat unit, d) RNA of haplotype with Type 2 repeat unit, e) Structural variant 1 for DNA of haplotype with Type 3 repeat unit variant (Type 3a), f) Structural variant 2 for DNA of haplotype with Type 3 repeat unit variant (Type 3a), g) Structural

variant 3 for DNA of haplotype with Type 3 repeat unit variant (Type 3a), h) Structural variant 4 for DNA of haplotype with Type 3 repeat unit variant (Type 3a), i) Structural variant 5 for DNA of haplotype with Type 3 repeat unit variant (Type 3a), j) Structural variant 1 for RNA of haplotype with Type 3 repeat unit variant (Type 3a), k) Structural variant 2 for RNA of haplotype with Type 3 repeat unit variant (Type 3a). In section B: a) Structural variant 1 for DNA of haplotype with Type 3 repeat unit variant (Type 3b), b) Structural variant 2 for DNA of haplotype with Type 3 repeat unit variant (Type 3b), c) Structural variant 3 for DNA of haplotype with Type 3 repeat unit variant (Type 3b), d) Structural variant 4 for DNA of haplotype with Type 3 repeat unit variant (Type 3b), e) Structural variant 1 for RNA of haplotype with Type 3 repeat unit variant (Type 3b), f) Structural variant 2 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 1 for RNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 2 for RNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 1 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 2 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 3 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 4 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 5 for DNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 1 for RNA of haplotype with Type 3 repeat unit variant (Type 3c), Structural variant 2 for RNA of haplotype with Type 3 repeat unit variant (Type 3c).

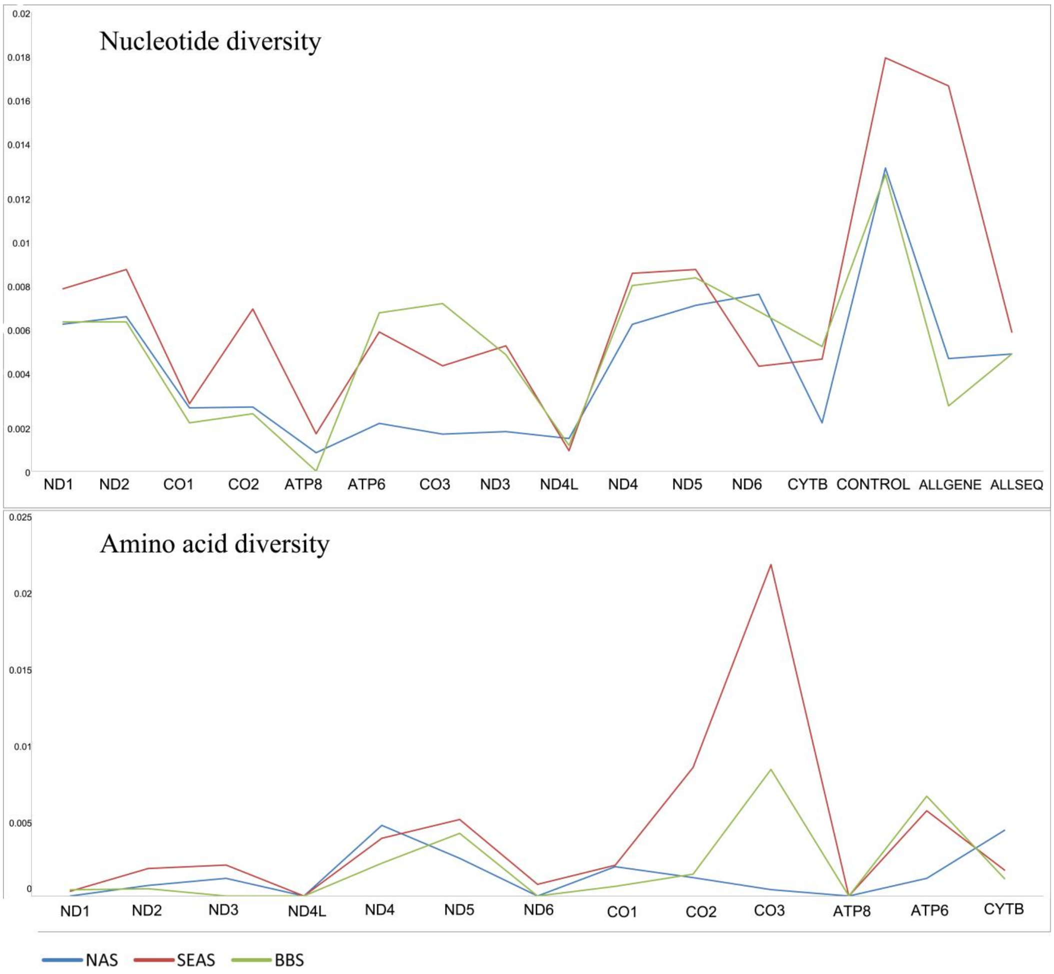


**Figure 3.S11** Nucleotide and amino acid diversity of *S. longiceps* populations from 3 ecoregions in the Indian Ocean. NAS (Northern Arabian Sea), SEAS (South Eastern Arabian Sea) and BoB (Bay of Bengal).

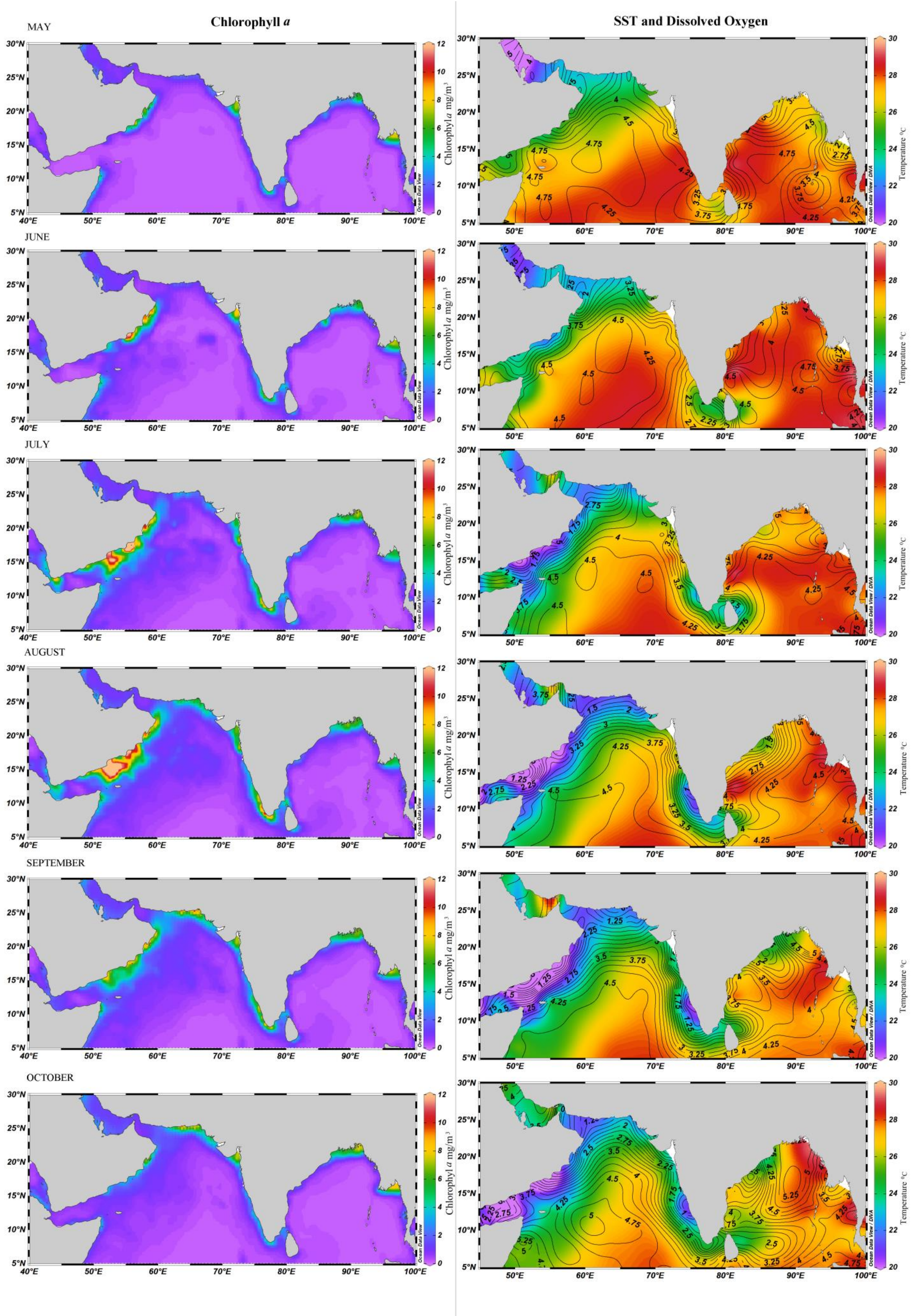


**Figure 3.S12** Monthly Chlorophyll-*a* (mg/m$^3$), Sea Surface Temperature- SST ($^0$C) and Dissolved Oxygen (µmol/kg) for the Bay of Bengal and Arabian Ocean from May to October. Chlorophyll-*a* and Sea Surface Temperature gradients are represented as coloured shades. Dissolved Oxygen is represented as contour lines.

## 5. References

1. Alheit J, Roy C, Kifani S (2009) Decadal-scale variability in populations In: Checkley D Oozeki Y, Roy C, (Eds) Climate change and small pelagic fish Cambridge; Cambridge University Press, United Kingdom
2. Allen AP, Gillooly JF (2007) The mechanistic basis of the metabolic theory of ecology. *Oikos* 116(6):1073-1077
3. Ballard JWO, Pichaud N (2014) Mitochondrial DNA: more than an evolutionary bystander. *Funct Ecol* 28(1):218–231
4. Ballard JWO, Melvin RG (2010) Linking the mitochondrial genotype to the organismal phenotype. *Mol Ecol* 19:1523-1539
5. Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Boil Evol* 16(1):37–48
6. Baris TZ, Crawford DL, Oleksiak MF (2015) Acclimation and acute temperature effects on population differences in oxidative phosphorylation. *Am J Physiol-Reg* I 310:R185-R196
7. Berg PR, Jentoft S, Star B, Ring KH, Knutsen H, Lien S, Jakobsen KS, Andre C (2015) Adaptation to low salinity promotes genomic divergence in Atlantic cod *Gadus morhua* L). *Genome Biol Evol* 7(6):1644–1663
8. Brand MD (2000) Uncoupling to survive? The role of mitochondrial inefficiency in ageing Experimental. *Gerontology* 35(6–7):811–820
9. Brennan RS, Hwang R, Tse M, Fangue NA, Whitehead A (2016) Local adaptation to osmotic environment in killifish, *Fundulus heteroclitus*, is supported by divergence in swimming performance but not by differences in excess post-exercise oxygen consumption or aerobic scope. *Comp Biochem Phy* A 196:11–19
10. Brown GG, Gadaleta G, Pepe G, Saccone C, Sbisa E (1986) Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J Mol Biol* 192(3):503–511
11. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB (2004) Toward a metabolic theory of ecology. *Ecology* 85(7):1771-1789
12. Burton RS, Pereira RJ, Barreto FS (2013) Cytonuclear genomic interactions and hybrid breakdown. *Annu Re Ecol Evol S* 44:281–302
13. Chatterjee A, Shankar D, Shenoi SSC, Reddy GV, Michael GS, Ravichandran M *et al*. (2012) A new atlas of temperature and salinity for the North Indian Ocean Journal of Earth System. *Science* 121(3):559–593
14. Checkley Jr DM, Asch RG, Rykaczewski RR (2017) Climate, anchovy, and sardine. *Annu Rev Mar Sci* 9:469-493
15. Cheng YT, Liu J, Yang LQ, Sun C, Kong QP (2013) Mitochondrial DNA content contributes to climate adaptation using chinese populations as a model. *PloS one* 8(11):pe79536
16. Consuegra S, John E, Verspoor E, De Leaniz CG (2015) Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genet Sel Evol* 47(1):1–10
17. Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland
18. Da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A (2008) The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* 9(1):119
19. Dalziel AC, Moyes CD, Fredriksson E, Lougheed SC (2006) Molecular evolution of cytochrome c oxidase in high-performance fish Teleostei: Scombroidei). *J Mol Evol* 62(3):319–331
20. de Villemereuil P, Gaggiotti OE, Mouterde M, Till-Bottraud I (2016) Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity* 3:249-254
21. Dean AM, Lehman C, Yi X (2017) Fluctuating Selection in the Moran. Genetics 205:1271–1283
22. Doi A, Suzuki H, Matsuura ET (1999) Genetic analysis of temperature-dependent transmission of mitochondrial DNA in *Drosophila*. *Heredity* 82(5):555–560
23. Dowling DK, Friberg U, Lindell J (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecol Evol* 23(10):546–554
24. Eddie KHHo, Aneil F (2019) Agrawal.Mutation accumulation in selfing populations under fluctuating selection. The Society for the Study of Evolution. *Evolution* 72(9):1759–1772
25. Ekau W, Auel H, Portner HO, Gilbert D (2010) Impacts of hypoxia on the structure and processes in pelagic communities (zooplankton, macro-invertebrates and fish). *Biogeosciences* 7(5):1669-1699
26. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10(3):564-567
27. Fangue NA, Richards JG, Schulte PM (2009) Do mitochondrial properties explain intraspecific variation in thermal tolerance? *J Exp Biol* 212(4):514–522

28. Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, Da Fonseca RR (2011) Positive selection on the killer whale mitogenome. *Biol Letters* 7(1):116–118
29. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133(3):693–709
30. Fuentes MA Ferrada E (2017a) Environmental Fluctuations and Their Consequences for the Evolution of Phenotypic Diversity. *Aip Conf Proc* 5:16
31. Garvin MR, Bielawski JP, Gharrett AJ (2011) Positive Darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon. *PloS One* 6(9):e24127
32. Garvin MR, Bielawski JP, Sazanov LA, Gharrett AJ (2015a) Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. *J Zool Syst Evol Res* 53(1):1–17
33. Garvin MR, Thorgaard GH, Narum SR (2015b) Differential expression of genes that control respiration contribute to thermal adaptation in redband trout (*Oncorhynchus mykiss gairdneri*). *Genome Biol Evol* 7(6):1404–1414
34. Gershoni M, Levin L, Ovadia O, Toiw Y, Shani N, Dadon S, Tsur A (2014) Disrupting mitochondrial–nuclear coevolution affects OXPHOS complex I integrity and impacts human health. *Genome Biol Evol* 6(10):2665–2680
35. Gillooly JF, Allen AP, West GB, Brown JH (2005) The rate of DNA evolution: effects of body size and temperature on the molecular clock. *P Natl A Sci Biol* 102(1):140–145
36. Gillooly JF, Allen AP (2007) Linking global patterns in biodiversity to evolutionary dynamics using metabolic theory. *Ecology* 88(8):1890-1894
37. Gueye M, Kantoussan J, Tine M (2016) Common Garden Experiments Confirm the Impact of Salinity on Reproductive Traits that is Observed in Wild Populations of the Back-Chinned Tilapia Sarotherodon melanotheron. *Int J Aquac Fish Sci* 2:031-037
38. Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 66:591–600
39. Harrisson K, Pavlova A, Gan HM, Lee YP, Austin CM, Sunnucks P (2016) Pleistocene divergence across a mountain range and the influence of selection on mitogenome evolution in threatened Australian freshwater cod species. *Heredity* 116(6):506–515
40. Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish Fish* 9(4):333–362
41. Hauser L, Turan C, Carvalho G (2001) Haplotype frequency distribution and discriminatory power of two mtDNA fragments in a marine pelagic teleost (Atlantic herring, *Clupea harengus*). *Heredity* 87:621–630
42. Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V (2007) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platichthys flesus* L.). *Heredity* 99(6):592
43. Hughes LC, Somoza GM, Nguyen BN, Bernot JP, González-Castro M, Díaz de Astarloa JM, Ortí G (2017) Transcriptomic differentiation underlying marine-to-freshwater transitions in the South American silversides *Odontesthes argentinensis* and *O bonariensis* (Atheriniformes). *Ecol Evol* 7(16):5258–5268
44. Hutchings JA, Swain DP, Rowe S, Eddington JD, Puvanendran V, Brown JA (2007) Genetic variation in life-history reaction norms in a marine fish. *P Roy Soc Lond B Bio* 274(1619):1693-1699
45. Iftikar FI, Hickey AJ (2013) Do mitochondria limit hot fish hearts? Understanding the role of mitochondrial function with heat stress in *Notolabrus celidotus*. *Plos One* 8:p.e64120
46. Jablonski D (1993) The tropics as a source of evolutionary novelty through geological time. *Nature* 364(6433):142–144
47. Jablonski D (1993) The tropics as a source of evolutionary novelty through geological time. *Nature* 364:142–144
48. Jacobsen MW, Da Fonseca, RR, Bernatchez L, Hansen MM (2016) Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the NADH2 gene in whitefish Coregonus ssp). *Mol Phylogenet Evol* 95(1):161–170
49. Jouanno J, Sheinbaum J, Barnier B, Molines JM, Candela J (2012) Seasonal and interannual modulation of the eddy kinetic energy in the Caribbean Sea. *J Phys Oceanogr* 42(11):2041–2055
50. Katz L, Burge CB (2003) Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res* 13(9):2042–2051
51. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*.28(12):1647–1649

52. Knutsen H, Olsen EM, Jorde PE, Espeland SH, Andre C, Stenseth NC (2011) Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. Mol Ecol 20(4):768-783
53. Lajbner Z, Pnini R, Camus MF, Miller J, Dowling DK (2018) Experimental evidence that thermal selection shapes mitochondrial genome evolution. *Sci Rep-UK* doi:101038/s41598-018-27805-3
54. Latorre-Pellicer A, Moreno-Loshuertos R, Lechuga-Vieco AV, Sanchez-Cabo F, Torroja C, Acin-Perez R,  Bernad-Miana ML (2016) Mitochondrial and nuclear DNA matching shapes metabolism and healthy ageing. *Nature* 535(7613):561–565
55. Letts JA, Fiedorczuk, K, Sazanov LA (2016) The architecture of respiratory supercomplexes. *Nature* 537(7622):644–648
56. Li Y, Park JS, Deng JH, Bai Y (2006) Cytochrome c oxidase subunit IV is essential for assembly and respiratory function of the enzyme complex. *J Bioenerg Biomembr* 38(5-6):283–291
57. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452
58. Lowell BB, Spiegelman BM (2000) Towards a molecular understanding of adaptive thermogenesis. *Nature* 404(6778):652–660
59. Madhupratap M, Nair KNV, Gopalakrishnan TC, Haridas P, Nair KKC, Venugopal P, Gauns M (2001) Arabian Sea oceanography and fisheries of the west coast of India. *Curr Sci India* 81(4):355–361
60. Marshall DJ, Monro K, Bode M, Keough MJ, Swearer S (2010) Phenotype–environment mismatches reduce connectivity in the sea. *Ecol Lett* 13(1):128–140
61. Marshall DJ, Morgan SG (2011) Ecological and evolutionary consequences of linked life-history stages in the sea. *Curr Biol* 21(18):718–725
62. Marshall HD, Coulson MW, Carr SM (2008) Near neutrality, rate heterogeneity, and linkage govern mitochondrial genome evolution in Atlantic cod (*Gadus morhua*) and other gadine fish. *Mol Biol+* 26(3), 579–589
63. Meiklejohn CD, Montooth KL, Rand DM (2007) Positive and negative selection on the mitochondrial genome. *Trends in Genet* 23(6):259–263
64. Melbinger A, Vergassola M (2015) The Impact of Environmental Fluctuations on Evolutionary Fitness Functions. *Sci Rep* 5:15211
65. Melo-Ferreira J, Vilela J, Fonseca MM, Da Fonseca RR, Boursot P, Alves PC (2014) The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome Biol Evol* 6(4):886–896
66. Mignotte F, Gueride M, Champagne AM, Mounolou JC (1990) Direct repeats in the non-coding region of rabbit mitochondrial DNA: Involvement in the generation of intra-and inter-individual heterogeneity. *Euro J Bio* 194(2):561–571
67. Mita S, Rizzuto R, Moraes CT, Shanske S, Arnaudo E, Fabrizi GM, Koga Y, DiMauro S, Schon EA (1990) Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. *Nucleic Acids Res* 18(3):561–567
68. Miya M, Nishida M, (2015) The mitogenomic contributions to molecular phylogenetics and evolution of fishes: a 15-year retrospect. *Ichthyol Res* 62(1):29–36
69. Morales HE, Pavlova A, Amos N, Major R, Bragg J, Kilian A, Greening C, Sunnucks P (2016) Mitochondrial-nuclear interactions maintain a deep mitochondrial split in the face of nuclear gene flow. *bioRxiv* 095596
70. Morales HE, Pavlova A, Joseph L, Sunnucks P (2015) Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Mol Ecol* 24(11):2820–2837
71. Mossman JA, Biancani LM, Zhu CT, Rand DM (2016) Mitonuclear epistasis for development time and its modification by diet in Drosophila. *Genetics* 203(1):463–484
72. Munroe TA, Priede IG (2010) *Sardinella longiceps* (errata version published in 2017). The IUCN Red List of Threatened Species 2010e:T154989A115258997
73. Narvekar J, D'Mello JR, Prasanna Kumar S, Banerjee P, Sharma V, Shenai-Tirodkar P (2017) Winter-time variability of the eastern Arabian Sea: A comparison between 2003 and 2013. *Geophys Res Lett* 44:6269-6277
74. Nei M (1987) Molecular Evolutionary Genetics. New York, Columbia University Press
75. Ojala D, Merkel C, Gelfand R, Attardi G (1980) The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell* 22(2):393–403
76. Osheroff N, Speck SH, Margoliash E, Veerman EC, Wilms J, Konig BW, Muijsers AO (1983) The reaction of primate cytochromes c with cytochrome c oxidase Analysis of the polarographic assay. *J Biol Chem* 258(9):5731–5738

77. Pavlova A, Gan HM, Lee YP, Austin CM, Gilligan DM, Lintermans M, Sunnucks P (2017) Purifying selection and genetic drift shaped Pleistocene evolution of the mitochondrial genome in an endangered Australian freshwater fish. *Heredity* 118(5):466–476
78. Peck MA, Reglero P, Takahashi M, Catalan IA (2013) Life cycle ecophysiology of small pelagic fish and climate-driven changes in populations. *Prog Oceanogr* 116:220-245
79. Pereira F, Soares P, Carneiro J, Pereira L, Richards MB, Samuels DC, Amorim A (2008) Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region. *Mol Biol Evol* 25(12):2759–2770
80. Pond SLK, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments *Bioinformatics* 21(10):2531–2533
81. Prasanna Kumar S, Muraleedharan PM, Prasad TG, Gauns M, Ramaiah N, De Souza SN, Madhupratap M (2002) Why is the Bay of Bengal less productive during summer monsoon compared to the Arabian Sea? *Geophys Res Lett* 29(24):88-1–88-4
82. Qasim SZ (1982) Oceanography of the northern Arabian Sea. *Deep-Sea Res Part A. Oceanographic Research Papers* 29(9):1041–1068
83. Rao DS, Ramamirtham CP, Murty AVS, Muthusamy S, Kunhikrishnan NP, Khambadkar LR (1992) Oceanography of the Arabian Sea with particular reference to the southwest monsoon. *CMFRI Bulletin* 45:4–8
84. Reiss CS, Checkley Jr DM, Bograd SJ (2008) Remotely sensed spawning habitat of Pacific sardine (*Sardinops sagax*) and Northern anchovy (*Engraulis mordax*) within the California Current Fish Oceanogr 17(2), 126-136 Climate, anchovy, and sardine. *Annu Rev Mar Sci* 9:469-493
85. Rion S, Kawecki TJ (2007) Evolutionary biology of starvation resistance: what we have learned from Drosophila. *J Evolution Biol* 20:1655-1664
86. Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9(3):552-569
87. Roxy MK, Ritika K, Terray P, Masson S (2014) The curious case of Indian ocean warming. *Am J Clim* 27(22), 8501-8509
88. Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303(5655):223–226
89. Samuels DC, Schon EA, Chinnery PF (2004) Two direct repeats cause most human mtDNA deletions. *Trends Genet* 20(9):393–398
90. Sato M, Barth JA, Benoit-Bird KJ, Pierce SD, Cowles TJ, Brodeur RD, Peterson WT (2018) Coastal upwelling fronts as a boundary for planktivorous fish distributions. *Mar Ecol Prog Ser* 595:171-186
91. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31(11):3381–3385
92. Scott GR, Schulte PM, Egginton S, Scott AL, Richards JG, Milsom WK (2010) Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol Biol Evol* 28(1):351–363
93. Sebastian W, Sukumaran S, Zacharia PU, Gopalakrishnan A (2017a) Genetic population structure of Indian oil sardine, *Sardinella longiceps* assessed using microsatellite markers. *Conserv Genet* 18(4):951–964
94. Sebastian W, Sukumaran S, Zacharia PU, Gopalakrishnan A (2017b) The complete mitochondrial genome and phylogeny of Indian oil sardine, *Sardinella longiceps* and Goldstripe *Sardinella, Sardinella gibbosa* from the Indian Ocean. *Conserv Genet Resour*10(4):735–739
95. Shadel GS, Clayton D A (1997) Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem* 66(1):409–435
96. Silva G, Lima F P, Martel P, Castilho R (2014) Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. *P Roy Soc Lond B Bio* 281(1792):20141093
97. Slomovic S, Laufer D, Geiger D, Schuster G (2005) Poly- adenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark. *Mol Cel Biol* 25(15):6427–6435
98. Solaini G, Baracca A, Lenaz G, Sgarbi G (2010) Hypoxia and mitochondrial oxidative metabolism. *BBA-Bioenergetics* 1797:1171-1177
99. Somero GN (1995). Proteins and temperature. *Annu Rev Physiol* 57:43-68
100. Stier A, Massemin S, Criscuolo F (2014) Chronic mitochondrial uncoupling treatment prevents acute cold-induced oxidative stress in birds. *J Comp Physiol B* 184(8):1021–1029
101. Suissa S, Wang Z, Poole J, Wittkopp S, Feder J, Shutt TE *et al*. (2009) Ancient mtDNA genetic variants modulate mtDNA transcription and replication. *Plos Genetics* 5(5):pe1000474

102. Sukumaran S, Gopalakrishnan A, Sebastian W, Vijayagopal P, Nandakumar Rao S *et al*. (2016) Morphological divergence in Indian oil sardine, *Sardinella longiceps* Valenciennes, 1847–Does it imply adaptive variation? *J Appl Ichthyol* 32(4):706–711

103. Sukumaran S, Sebastian W, Gopalakrishnan A (2015) Population genetic structure of Indian oil sardine, *Sardinella longiceps* along Indian coast. *Gene* 576(1):372–378

104. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2):437–460

105. Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics* 123(3):597–601

106. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 60. *Mol Biol Evol* 30(12):2725–2729

107. Thompson JN (2013) Relentless Evolution. University of Chicago Press

108. Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucleic Acids Res* 9(20):5411–5421

109. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19(5):671–672

110. Wright BE (2000) A biochemical mechanism for nonrandom mutations and evolution. *J Bacteriol* 182(11):2993–3001

111. Xu C, Boyce MS (2009) Oil sardine (*Sardinella longiceps*) off the Malabar coast: density dependence and environmental effects. *Fish Oceanogr* 18(5):359–370

112. Yasukawa T, Yang MY, Jacobs HT, Holt IJ (2005) A bidirectional origin of replication maps to the major noncoding region of human mitochondrial DNA. *Mol Cel* 18(6):651–662

113. Zhu J, Vinothkumar KR, Hirst J (2016) Structure of mammalian respiratory complex I. *Nature* 536(7616):354–35

114. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406–3415

# Chapter 4

## POPULATION GENETIC STRUCTURE OF *SARDINELLA LONGICEPS* (Valenciennes, 1847) IN THE INDIAN OCEAN REGION USING MICROSATELLITE DNA MARKERS.

### ABSTRACT

Indian oil sardines, commercially and ecologically important pelagic fishes in Indian waters have not been the focus of major genetic studies as compared to their counterparts in Atlantic and Pacific oceans despite several reports suggesting stock complexity and intraspecific diversity. Hence, we investigated the genetic stock structure of Indian oil sardine, *Sardinella longiceps* using microsatellite markers by collecting a total of 768 individuals from eight locations along the Indian coast and one from the Gulf of Oman over 2 years (2013-2015). Six polymorphic microsatellite markers revealed significant genetic differentiation between populations with the highest $F_{ST}$ value (0.055) between Oman and Indian coastline. Within the Indian coastline, another major subdivision between Mumbai & Mangalore vs. other regions were detected ($F_{ST}$ value 0.047) which was also confirmed in Barrier analysis with the presence of two strong barriers between these eco-regions. There exist pronounced differences in oceanographic and environmental features between Gulf of Oman, Western Indian Ocean and Eastern Indian Ocean (Bay of Bengal) which may act as barriers for effective dispersal and gene flow resulting in genetic differentiation. Even though the samples collected from Calicut, Kollam, Trivandrum, Chennai and Vizag showed the presence of admixed genotypes, the possible presence of distinct populations in some regions was evident in Bayesian analysis which needs to be confirmed further using more widespread sampling design and powerful markers. The present study provided insights into the biocomplexity and intraspecific diversity of Indian oil sardine populations, which needs to be preserved for maintaining the resilience of these important fishes to climate change and habitat alterations in the Indian Ocean.

## 1. INTRODUCTION

Understanding the different mechanisms maintaining intra and interspecific diversity is the primary objective of evolutionary and conservation biology as diversity is the basis of the long-term sustainability of marine fish populations (Hutchings 2000; Santamaria and Mendez 2012). Sustainable harvesting of commercially important species warrants effective management strategy based on scientific studies (Carvalho and Hauser 1995; Frankham *et al*. 2002; Dunlop *et al*. 2009). The on-going climate change, ocean warming and consequent depletion of marine fish populations make it necessary to study population structuring and environmental adaptation in marine fishes (Lecomte *et al*. 2004; Johnson and Welch 2009; Nielsen *et al*. 2009). Marine fishes have traditionally been included in the low genetically differentiated and weak locally adapted category due to the lack of physical barrier to gene flow in the marine environment and large effective population sizes (Smedbol *et al*. 2002; Poulsen *et al*. 2006). However, recent studies detected spatially structured, genetically distinct (Teacher *et al*. 2013; Andre *et al*. 2016) and locally adapted (Larsen *et al*. 2007; Johannesson *et al*. 2011; Wang *et al*. 2013; Brennan *et al*. 2016) populations in marine fishes. The biological characteristics of marine fishes like natal homing (Natoli *et al*. 2005; Svedang *et al*. 2007), larval retention (Cowen *et al*. 2006) along with historic events (Grant and Bowen 1998; Bradbury *et al*. 2008), oceanic current patterns (Cowen *et al*. 2006) and environmental factors (like temperature and salinity gradient) (Larsen *et al*. 2012) contribute to the differentiation process.

Indian oil sardine, *Sardinella longiceps* (Valenciennes, 1847) is one of the most important commercial pelagic fish in Indian waters which form the largest pelagic fishery of India, with an annual production of 0.34 million tons (CMFRI 2015). It is a cheap source of protein for millions and it contributes to the majority of income from fishing due to its abundance. It also plays a significant role in trophic ecology and food web as a planktivorous, energy-rich small forage fish species which are consumed in large quantities by apex predators along with other sardines, mackerel and anchovy. Large scale feeding on phytoplankton by sardines helps in transferring energy from one location and time to another. Indian oil sardines inhabit continental shelf waters at a depth range of 20-200 m and are distributed along both the east and west coasts of India, Gulf of Oman and Gulf of Aden. They are coastal, pelagic, form schools in coastal waters and undertake localized migrations (Froese and Pauly 2009). It breeds once a year, during June-July (reported along south-west coast of India) when temperature and salinity are reduced by the southwest monsoon and spawning peaks in August and September at temperatures from 22 to 28 °C (Talwar and Kacker 1984). The exact spawning grounds are not yet located. The pelagic eggs are spherical, range from 1 to 4 mm in diameter and require only 24 h for development. The pelagic larval development includes minimal movement, but it travels by serpentine swimming and the larval cycle is completed in approximately 40 days (Kuthalingam 1960).

Identifying and characterizing evolutionary significant units of small pelagic fishes for management of its fishery is difficult because these species do not follow the traditional population dynamics models and assumptions (Cadrin *et al*. 2013). They are short-lived, fast-growing, and are characterized by variable levels of natural mortality (Cadrin *et al*. 2013). Their stock size is linked to recruitment, which may be highly variable depending on the presence of an optimal environmental window and hence there exist several hurdles in implementing management measures as compared to longer-lived species (Alheit *et al*. 2009). Reliability of age-length frequency data and catch effort analysis is complicated by their size-selective shoaling behaviour (Alheit *et al*. 2009). There are no species-specific conservation measures in India for Indian oil sardine. But all coastal states have implemented the Marine Fishing Regulation Act by following closed seasons and limiting of fishing zones for different categories of fishing methods. Like other marine pelagic fishes, Indian oil sardine fishery also exhibited fluctuating behaviour, with many population crashes and recoveries during the past century (Devaraj and Martosubroto 1997). Malabar upwelling zones, which is one of the important upwelling zones of the Western Indian Ocean is the largest contributor of the Indian oil sardine fishery and upwelling along these coasts is wind-induced occurring mainly during June-August (Devaraj and Martosubroto 1997). Success or failure of sardine recruitment and fishery is highly dependent on the oceanographic features of the Malabar upwelling zones since sardine fishery is dominated by 0 and 1-year class fishes (Devaraj and Martosubroto 1997; Krishnakumar and Bhat 2008). The important factors that determine recruitment and fishery of Indian oil sardines are intensity of upwelling (Devaraj and Martosubroto 1997), availability of the diatoms *F. oceanica*, (Nair 1952; Krishnakumar and Bhat 2008) intensity of rainfall, dissolved oxygen, temperature, migratory pattern and survival of the egg and larvae (Devaraj and Martosubroto 1997). However, overfishing and capture of juvenile/ immature fishes affect the fishery detrimentally (Devanesan 1943; Mohamed *et al.* 2014).

Small pelagic fishes especially sardines of the major oceans like Atlantic and Pacific have been well studied using molecular markers providing improved understanding regarding their biocomplexity and intra-specific diversity (Grant and Bowen 1998; Cadrin *et al*. 2013; Da Silva *et al*. 2015). Indian ocean sardines are less studied using molecular markers compared to their Atlantic and Pacific counterparts except few works using enzyme loci (Venkita Krishnan 1993), cytogenetic, biochemical, and morphometric tools (Mohandas 1997) and allozymes (Menezes 1994). Venkita Krishnan (1993) used nine enzyme loci to differentiate between subpopulations and inferred that sardines from Cochin, Calicut, Mangalore, Mandapam, and Madras (now Chennai) belong to distinct stocks. Similar conclusions were also made by Mohandas (1997) by using biochemical genetic, morphometric and cytogenetic tools. Contrary to this, Menezes (1994) reported reduced genetic variability in Indian oil sardines from the West Coast of India using allozymes. All these studies were limited by low sample size and geographical coverage and hence the present authors carried out a comprehensive study using mitochondrial DNA markers (Sukumaran *et al*. 2016b) unveiling their historical demography. However, mitochondrial markers were not efficient enough to detect any subpopulation structure in Indian oil sardines and hence this study was designed using microsatellite markers. Microsatellite markers are presumed to be more sensitive markers to detect population subdivision, especially in weakly divergent populations due to their high mutation rates and selective neutrality contributing to high allelic diversity and heterozygosity (DeWoody and Avise 2000; Putman and Carbone 2014; Borrell *et al*. 2012). Hence, we attempted to understand the population genetic structure of Indian oil sardines collected from 8 locations along the Indian coast and one location from Gulf of Oman using microsatellite markers developed through the cross-amplification method. Tests based on classical methods ($F_{ST}$, $R_{ST}$ and hierarchical AMOVA) were combined with Bayesian clustering, principal component analysis, and likelihood estimation of migration rate to derive clues to spatial patterns of structuring. The factors which influence population structuring were studied using mantel tests by using isolation by distance (IBD) and isolation by environment (IBE) algorithms. Our analysis showed a hierarchical population genetic differentiation in the Indian oil sardine.

## 2. Materials And Methods

### 2.1. Sample collection

Indian oil sardine samples were collected from sites across most of the species distribution along the Indian coast as well as from the Gulf of Oman from gillnetters and ring seiners (a mini purse seine) operated near the coast. A total of 800 individuals of *S. longiceps* were collected from seven sites along the Indian coast (Mumbai, Mangalore, Calicut, Kollam, Trivandrum, Chennai and Vizag) and one from the Gulf of Oman, during 2012-2014 (Fig. 4.1). Genomic DNA was isolated from ethanol preserved muscle tissue by standard phenol/chloroform method after proteinase K digestion (Sambrook and Russell 2001). The purified DNA was quantified using Biophotometer (Thermo scientific). All samples were diluted to 50-100 ng/ul using 1X TE (pH 8) before PCR amplification. Each DNA sample was amplified at 12 microsatellite loci using the cross-amplification method; SAR 9 ($GT_{17}$), SAR19B5 ($GT_{48}$) and SARA2F ($GT_{48}$)selected from *Sardina pilchardus* (Gonzalez and Zardoya 2007) and SAR B-D09 ($CA_9$,$GA_8$), SAR B-H04 F ($TG_{18}$), SARB-G09 ($GA_6$,GT,$GA_{36}$), SAR B-H04 ($GT_9$), SAR B-A08 ($CA_{26}$), Sar1-D01 ($CA_{29}$,GG,$CA_3$), Sar1-D06 B ($TG_{18}$), Sar1-H11 B ($TG_{11}$,TA,$TG_6$) and SarB-CO5 ($TC_5$,TT,$TC_4$) from *Sardinops sagax sagax* (Pereyra *et al*. 2004). The forward primers used for amplifying microsatellite loci were labelled with a fluorescent dye (6 FAM).

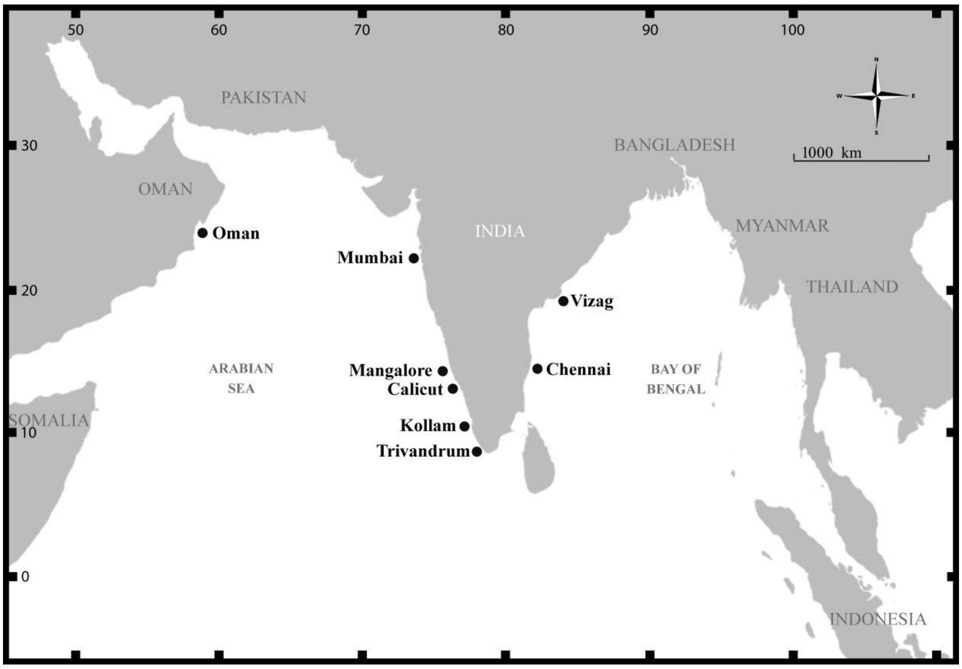


**Fig. 4.1** Map showing Sampling sites (*bold letter*). Scale is approximate

The PCR reactions were performed in 25 µl reaction volume containing 50 ng DNA, 1× reaction buffer (Sigma Aldrich) (10 mM Tris-HCl, pH 8, 500 mM KCl, 1.5 mM MgCl2), 10 mM of each dNTP, 0.5 µM of each primer and 1U Taq DNA polymerase buffer (Sigma Aldrich). The PCR cycles were carried out in a Biorad T100 thermocycler (Biorad, USA) programmed for an initial denaturation at 94 °C for 4 min followed by 33 cycles of; denaturation at 94 °C for 30 s, annealing for 30 s and extension at 72 °C for 30 s and a final extension at 72 °C for 5 min. Annealing temperatures for each locus were: 50 °C (Sar1-D01), 49 °C (Sar1-D06(B)), 51 °C (Sar1-H11(B)), 46 °C (SarB-CO5), 48 °C (SAR B-D09), 50 °C (SAR B-H04 (F)), 51 °C (SAR B-G09), 53 °C (SAR B-H04), 49 °C (SAR B-A08), 50 °C (SAR 9), 48 °C (SAR19B5) and 51 °C (SARA2F). Analysis of fragment length was carried out using an ABI prism genetic analyser (Applied Biosystems, USA) with appropriate size standard.

2.2. Data analyses

Allele calling and sizing of all 12 microsatellite loci from 800 individuals representing eight collection sites were carried out using Gene Mapper v3.7 software (Applied Biosystems, USA) to record their positions. After automated allele calling, each sample was verified by inspecting the peak patterns in electropherogram (exported as pdf file from Gene Mapper v3.7 software) manually, to reduce potential scoring problems caused by PCR artefacts. The quality of the microsatellite data set was again checked by analysing the presence of null alleles, allele size shifts and scoring errors using the software MICROCHECKER (Van Oosterhout *et al.* 2004). Comparative measures of genetic diversity for each sample were calculated like alleles per locus (A), allelic richness (Ar), observed heterozygosity (Ho), expected heterozygosity (He) and coefficient of inbreeding ($F_{IS}$) using ARLEQUIN (Excoffier and Lischer 2010). Deviations from Hardy-Weinberg equilibrium and pairwise linkage disequilibrium were also estimated in ARLEQUIN. Sequential Bonferroni corrections were applied when appropriate (Rice 1989). The loci which exhibited large allele drop out, extreme deviation from Hardy-Weinberg equilibrium and highly significant linkage disequilibrium were eliminated from further analysis. The other six loci; Sar1-D01, Sar1-D06 (B), Sar1-H11 (B), SAR B-D09, SAR B-A08, and SAR 9 were selected for further analysis. The program POWSIM 4.1 (Ryman and Palm 2006) was used to estimate statistical power for detecting genetic differentiation ($F_{ST}$) ranging from 0.00 to 0.05. We used effective population size *Ne* = 7000, the number of generation *t* varied from 0 to 718 and 1000 simulation runs. The proportion of significant outcome was used to estimate statistical power for detecting pairwise genetic differentiation.

To examine genetic differentiation and structuring in sardine populations, multiple approaches were used. We estimated allele identity based (IAM) (Global FST and pairwise $F_{ST}$) and allele size based (SSM) (Global $R_{ST}$ and pairwise $R_{ST}$) statistics in ARLEQUIN and GENEPOP (Raymond and Rousset 1995), respectively. We also calculated the Jost DST (Jost 2008) statistics with the online programme SMOGD (Hedrick 2005; Crawford 2010) separately for each locus and harmonic mean of $D_{ST}$ was used as a measure of heterozygosity based relative differentiation of allele frequencies (actual difference) among samples. We then used the allele size permutation test to compare the relative effect of genetic drift and migration on genetic differentiation of the Indian oil sardine population using the programme SPAGeDi (Hardy and Vekemans 2002). Different allele sizes observed at each locus of the dataset are randomly permutated for 2000 times to test the null hypothesis that stepwise mutation does not contribute to population differentiation. The alternative hypothesis is that genetic differentiation is caused mainly by SSM like mutations.

Mantel tests were used to test for correlation between genetic difference and geographical distance between sampling sites. The geographical distance (shortest sea route) in kilometres were regressed against $F_{ST}/(1-F_{ST})$ with the web-service program isolation by distance (IBDWS) (Jensen *et al.* 2005). Significance of the regression slopes was tested with 1000 permutations. The tests were also calculated with $R_{ST}$ and $D_{ST}$.

Population structure was further characterised by principal component analysis (PCA) based on allele frequency performed using PCAGEN (Goudet 1999) for all eight population samples. The significance of principal component was tested by 15,000 randomisations. A model-based Bayesian MCMC clustering was carried using STRUCTURE v2.3 (Pritchard *et al.* 2000) to determine the number of genetically discrete populations (K) with the highest posterior probability. We simulated K values ranging from 1 to 8 (total sampling sites) under the admixture model with correlated allele frequencies to determine the most likely pattern of population connectivity. Ten independent runs were performed for each K value to verify the results. The program was run for 1,000,000 MCMC steps with a burn-in period of 100,000 steps. The web-based program STRUCTURE HARVESTER (Earl 2012) was used to estimate the most likely value for K with Evanno's delta K method (Evanno *et al*. 2005). The analysis was repeated after excluding the clusters identified in the first run to detect any hierarchical structures. Bayesian clustering analysis was also performed in STRUCTURAMA v.2.1 (Huelsenbeck *et al.* 2011). The run consists of 1,000,000 steps, 10,000 burn-in, numpops = 1-8 and admixture model. BARRIER v2.2 (Manni *et al*. 2004) was used to estimate the area exhibiting largest genetic discontinuities between population pairs. The analysis was conducted using $F_{ST}$ and $R_{ST}$ matrices of genetic distance. The robustness of the barriers was assessed by analyses matrices for each microsatellite locus separately.

To estimate the amount of genetic variability partitioned within and among different sets of populations, an analysis of molecular variance (AMOVA) was performed with ARLEQUIN. Alternative population clustering was used to measure genetic variation within it. Six alternative scenarios were tested based on PCA and Bayesian clustering analysis as in Table 3. We used the mantel test to analyse the correlation between genetic distance ($F_{ST}$) and environmental distance. The environmental distance matrix was constructed using the values of mean temperature and salinity at the sea surface during spawning season (July-September) [estimated from Indian National Centre for Ocean Information Services (INCOIS data)] (Chatterjee *et al*. 2012; Srivastava *et al*. 2015) for location closest to the sampling sites. Mantel tests were performed between the genetic distance matrix and environmental distance matrix. Besides, Partial Mantel tests were performed between genetic distance and environmental distance matrix while controlling the effect for geographical distance. The analysis was performed in the program ZT (Bonnet and Van de Peer 2002) with 300,000 permutations. The computer program TreeFit (Kalinowski 2009) was used to analyse how well a tree fits the genetic data. The tree was constructed from the summarised pairwise $F_{ST}$ matrix based on the neighbour-joining (NJ) method. The R2 value for the NJ tree was calculated. The treeview file generated was visualized in FIGTREE (Andrew 2014), tree-building program.

The heterozygote excess statistics were computed using the software BOTTLENECK (Piry *et al*. 1999), to detect the demographic history of population size variations like recent population decline. Rare alleles would be lost rapidly than common alleles during bottlenecks and hence there will be an excess of heterozygotes when populations experience recent size reduction as compared to populations in equilibrium (Cornuet and Luikart 1996). 95% single-step mutation and 5% multiple-step mutations with 1000 simulation iterations were set under three different mutation models; the infinite allele model (IAM), stepwise mutation model (SMM) and two-phase mutation model (TPM). The programme MIGRATE (Beerli 2006) from CIPRES Science Gateway (Miller *et al*. 2010) was used to estimate population size parameter and migration rate among *S. longiceps* population samples, based on maximum likelihood method (Beerli and Felsenstein 1999). The process was carried out with the SSM model and $F_{ST}$ estimates were used as starting material parameter for estimation. The program was run for 1,000,000 MCMC steps after an initial burn-in of 100,000 interactions.

## 3. Results

The final analysis was carried out using six microsatellite loci for 768 *S. longiceps* samples from the Indian coast and the Gulf of Oman. All loci were polymorphic. The average values of genetic diversity measures like alleles per locus (A), expected heterozygosity ($H_e$) and observed heterozygosity ($H_o$) was similar among samples for the same microsatellite locus (Table 4.1). The average number of alleles per locus (A) ranged from 15.62 (SAR B-A08) to 58.88 (Sar1-D06 (B)) and an average number of alleles per population ranged from 25.00 (MUM) to 39.5 (OMAN). Allelic richness ranged from 23.46 (MUM) to 34. 61 (OMAN), showing high inter-population genetic diversity. The expected ($H_e$) and observed heterozygosities ($H_o$) ranged from 0.856 to 0.931 and 0.814 to 0.860 respectively. In most of the loci within sampling sites, $H_e$ was slightly higher than Ho showing heterozygosity deficit among samples. All *S. longiceps* populations showed positive inbreeding coefficient ($F_{IS}$), except Chennai (−0.004 to 0.090), showing little outcrossing between these sites (Crow 2010). The power analysis of the microsatellite loci revealed that the combination of microsatellite loci and sample sizes used have 90% statistical power to detect a very low ($F_{ST}$ 0.0025) level of genetic differentiation (Table 4.S1). The average number of private alleles (34.7%) for each population is given in Table 4.S6.

**Table 4.1** Summary statistics for all microsatellite loci and samples.

| Location and locus parameters | Abbreviations | SAR 9 | SAR B-A08 | SAR B-D09 | Sar1-D01 | Sar1-D06 (B) | Sar1-H11 (B) | Average |
|---|---|---|---|---|---|---|---|---|
| MUMBAI (*n-96*) | MUM | | | | | | | |
| A | | 16 | 15 | 31 | 19 | 52 | 16 | 25.00 |
| Ar | | 15.792 | 15.000 | 29.463 | 17.507 | 47.696 | 15.331 | - |
| He | | 0.906 | 0.852 | 0.954 | 0.878 | 0.965 | 0.874 | 0.905 |
| Ho | | 0.792 | 0.938 | 0.698 | 0.957 | 0.860 | 0.756 | 0.833 |
| HW | | 0.112 | 0.119 | 0.234 | 0.324 | 0.423 | 0.547 | - |
| $F_{IS}$ | | 0.126 | -0.102 | 0.269 | -0.091 | 0.109 | 0.136 | 0.079 |
| MANGALORE (*n-96*) | MAN | | | | | | | |
| A | | 24 | 20 | 33 | 27 | 62 | 16 | 31.50 |
| Ar | | 21.535 | 18.271 | 30.178 | 22.940 | 53.037 | 14.880 | - |
| He | | 0.901 | 0.868 | 0.957 | 0.892 | 0.978 | 0.851 | 0.908 |
| Ho | | 0.771 | 0.816 | 0.906 | 0.846 | 0.871 | 0.750 | 0.827 |
| HW | | 0.129 | 0.134 | 0.092 | 0.119 | 0.001 | 0.052 | - |
| $F_{IS}$ | | 0.145 | 0.060 | 0.053 | 0.052 | 0.108 | 0.119 | 0.090 |
| CALICUT (*n-96*) | CAL | | | | | | | |
| A | | 26 | 13 | 36 | 23 | 65 | 18 | 31.50 |
| Ar | | 23.014 | 12.746 | 33.963 | 21.107 | 55.750 | 16.643 | - |
| He | | 0.898 | 0.832 | 0.959 | 0.881 | 0.978 | 0.719 | 0.878 |
| Ho | | 0.781 | 0.726 | 0.943 | 0.933 | 0.871 | 0.842 | 0.849 |
| HW | | 0.0003 | 0.0007 | 0.0127 | 0.098 | 0.0041 | 0.5907 | - |
| $F_{IS}$ | | 0.130 | 0.128 | 0.018 | -0.059 | 0.109 | -0.171 | 0.033 |
| KOLLAM (*n-96*) | KLM | | | | | | | |
| A | | 34 | 13 | 29 | 24 | 51 | 24 | 30.33 |
| Ar | | 31.700 | 11.953 | 28.653 | 21.491 | 45.371 | 20.685 | - |
| He | | 0.920 | 0.816 | 0.957 | 0.889 | 0.971 | 0.820 | 0.896 |
| Ho | | 0.842 | 0.948 | 0.847 | 0.863 | 0.883 | 0.708 | 0.848 |
| HW | | 0.026 | 0.097 | 0.0821 | 0.094 | 0.0421 | 0.084 | - |
| $F_{IS}$ | | 0.086 | -0.163 | 0.115 | 0.030 | 0.099 | 0.137 | 0.054 |
| TRIVANDRUM (*n-96*) | TRI | | | | | | | |
| A | | 21 | 19 | 30 | 25 | 53 | 18 | 29.17 |
| Ar | | 19.014 | 17.298 | 27.792 | 23.301 | 47.372 | 15.677 | - |
| He | | 0.9083 | 0.873 | 0.952 | 0.896 | 0.976 | 0.606 | 0.868 |
| Ho | | 0.865 | 0.902 | 0.946 | 0.714 | 0.890 | 0.569 | 0.815 |
| HW | | 0.271 | 0.048 | 0.314 | 0.433 | 0.049 | 0.023 | - |
| $F_{IS}$ | | 0.048 | -0.034 | 0.006 | 0.204 | 0.086 | 0.059 | 0.062 |
| CHNENNAI (*n-96*) | CHN | | | | | | | |
| A | | 20 | 15 | 35 | 21 | 69 | 16 | 29.83 |
| Ar | | 18.307 | 13.376 | 32.472 | 19.379 | 57.845 | 13.882 | - |
| He | | 0.879 | 0.829 | 0.960 | 0.872 | 0.979 | 0.620 | 0.857 |
| Ho | | 0.718 | 0.958 | 0.932 | 0.922 | 0.894 | 0.737 | 0.860 |
| HW | | 0.086 | 0.052 | 0.042 | 0.094 | 0.032 | 0.075 | - |
| $F_{IS}$ | | 0.184 | -0.156 | 0.030 | -0.058 | 0.087 | -0.189 | -0.004 |
| VIZAG (*n-96*) | VKP | | | | | | | |
| A | | 21 | 17 | 33 | 24 | 59 | 26 | 30.50 |
| Ar | | 19.393 | 16.113 | 30.873 | 21.481 | 51.746 | 23.273 | - |
| He | | 0.903 | 0.869 | 0.963 | 0.875 | 0.977 | 0.779 | 0.894 |
| Ho | | 0.854 | 0.945 | 0.770 | 0.954 | 0.872 | 0.723 | 0.853 |
| HW | | 0.134 | 0.184 | 0.243 | 0.154 | 0.154 | 0.144 | - |
| $F_{IS}$ | | 0.054 | -0.088 | 0.201 | -0.090 | 0.096 | 0.072 | 0.044 |
| OMAN (*n-96*) | OMAN | | | | | | | |
| A | | 64 | 13 | 35 | 29 | 60 | 34 | 39.50 |
| Ar | | 55.714 | 12.602 | 31.713 | 24.827 | 52.555 | 30.290 | - |
| He | | 0.978 | 0.857 | 0.957 | 0.884 | 0.978 | 0.932 | 0.931 |
| Ho | | 0.891 | 0.768 | 0.817 | 0.860 | 0.849 | 0.912 | 0.849 |
| HW | | 0.0486 | 0.055 | 0.085 | 0.0921 | 0.0667 | 0.0574 | - |
| $F_{IS}$ | | 0.090 | 0.104 | 0.147 | 0.028 | 0.132 | 0.021 | 0.088 |

*A* number of alleles per locus, *Ar* allelic richness, *Ho* observed heterozygosity, *He* expected heterozygosity, *HW* HW- p-value of Hardye Weinberg equilibrium test as implemented in Genepop, $F_{IS}$ coefficient of inbreeding

The global $F_{ST}$ and $R_{ST}$ values across all 8 Sardine populations were 0.0271 and 0.0778 respectively indicating a high level of genetic differentiation among individuals. In the pairwise analysis, $F_{ST}$ and $R_{ST}$ values were high and significant only when samples from OMAN were compared to those from other locations (0.02750-0.08524 and 0.107-0.248) (Table 4.2). For other comparisons, $F_{ST}$ values were low but significant ($p < 0.001$) in most of them (0.00402-0.0677). However, there was no significant difference between Calicut, Kollam and Trivandrum ($p > 0.05$) in pairwise $F_{ST}$ analysis. $R_{ST}$ values were significant in all comparisons ($p < 0.001$) (0.011-0.248). A similar pattern was observed in $D_{ST}$ values, with relatively higher values than in $F_{ST}$ and $R_{ST}$.

**Table 4.2** Pairwise estimates of $F_{ST}$ (below diagonal), $D_{ST}$ (below diagonal in bracket) and $R_{ST}$ (above diagonal) for microsatellite markers.

| | MUM | MAN | CAL | TRI | CHN | VKP | KLM | OMAN |
|---|---|---|---|---|---|---|---|---|
| MUM | | 0.061* | 0.018* | 0.097* | 0.099* | 0.075* | 0.018* | 0.130* |
| MAN | 0.00402*<br>(0.0575) | | 0.052* | 0.072* | 0.03* | 0.037* | 0.013* | 0.151* |
| CAL | 0.03930*<br>0.0801) | 0.04988*<br>(0.0856) | | 0.051* | 0.041* | 0.036* | 0.010* | 0.203* |
| TRI | 0.05547*<br>(0.0677) | 0.06611*<br>(0.0408) | 0.00254<br>(0.0490) | | 0.03* | 0.011* | 0.014* | 0.150* |
| CHN | 0.05070*<br>(0.0874) | 0.05881*<br>(0.0724) | 0.00351<br>0.0121) | 0.00106*<br>(0.0328) | | 0.012* | 0.009* | 0.121* |
| VKP | 0.02708*<br>(0.0601) | 0.03122*<br>0.0874) | 0.00200<br>0.0315) | 0.00688*<br>(0.0076) | 0.00705*<br>(0.0180) | | 0.013* | 0.248* |
| KLM | 0.01236*<br>(0.0976) | 0.01550*<br>(0.1094) | 0.00222<br>(0.0991) | 0.00698<br>(0.0759) | 0.00629*<br>(0.0685) | 0.00677*<br>(0.0318) | | 0.107* |
| OMAN | 0.02750*<br>0.1942) | 0.02751*<br>(0.2175) | 0.06841*<br>0.3088) | 0.08524*<br>(0.2938) | 0.08371*<br>(0.2575) | 0.04943*<br>(0.252) | 0.03213*<br>(0.2364) | |

*Significant $F_{ST}$ and $R_{ST}$ values after Bonferroni correction for multiple tests. Refer to Table 1 for abbreviations of sampling sites.

Principal component analysis (PCA) indicated a significant proportion of the total genetic variance partitioned in the first two PCs (Fig. 4.S1). PC1 ($p < 0.001$) explained 53.12% of the total genetic variance and grouped samples into two clades corresponding to Arabian Sea and Bay of Bengal with further separation of Oman from all other samples. Samples from Arabian Sea showed south-north trend along PCA 2 (20.79%, $p < 0.01$).The mantel test showed a significant correlation ($r = 0.534$, $R^2$= 0.285, $p < 0.001$) between genetic distance ($F_{ST}/(1 - F_{ST})$) and geographical distance based on all loci (Fig. 4.S2). It was also significant when logarithm of genetic and geographical distance was taken ($r = 0.283$, $R^2 = 0.0807$, $p < 0.001$) (data not shown). It was also significant when both $R_{ST}$ ($r = 0.242$, $R^2 = 0.0588$, $p < 0.01$) and $D_{ST}$ values ($r = 0.551$, $R^2 = 0.305$, $p < 0.02$) were considered. This indicates the existence of strong correlation between genetic and geographical distance among *S. longiceps* populations. In addition to this, a statistically significant correlation, even though weak was evident between genetic distance ($F_{ST}$) and geographical distance (shortest sea route) ($r = 0.286$; $p < 0.02$ for salinity, $r = 0.3932$; $p < 0.04$ for temperature). While, in partial mantel tests carried out by controlling geographical distance, the correlation with salinity ($r = 0.272$; $p < 0.02$) was the most significant.

The cluster analysis on whole data set using Bayesian algorithm implemented in the program STRUCTURE indicates that the most likely number of distinct genetic entities is $K = 2$ or above 2 (Fig. 4.S3) but visual inspection reveals that setting $K > 2$ did not add any meaningful pattern (Fig. 4.2). So, we chose $K = 2$, which probably represents the major structure of the population (Evanno *et al*. 2005) (The clusters represent Oman and Indian Ocean region respectively). Subsequent analysis within Indian Ocean samples (after removing Oman samples) revealed sub-structuring ($K = 2$) with northwest Indian Ocean (Mumbai and Mangalore) and rest of the Indian Ocean region samples (Calicut, Kollam, Trivandrum, Chennai and Vizag) forming two clusters. In further analysis (removing Mumbai and Mangalore) even though populations showed continuous distribution with the neighbouring populations with the presence of admixed individuals, some individuals are strongly assigned to one population and the proportions assigned to each group are asymmetric (Fig. 4.2). This is an indication of population structure and ΔK method suggest $K = 3$. Based on these analyses, the most prominent sub-clustering is between Oman vs. Mumbai & Mangalore vs. other regions (Calicut, Kollam, Trivandrum, Chennai, Vizag). But the signature of sub clustering three groups; Calicut versus Kollam versus (Trivandrum, Chennai, Vizag) within the third cluster (other regions) has been observed which needs to be confirmed further. Kollam samples indicate admixture proportions with Trivandrum, Chennai, Vizag along with a distinct proportion assigned to a separate cluster.

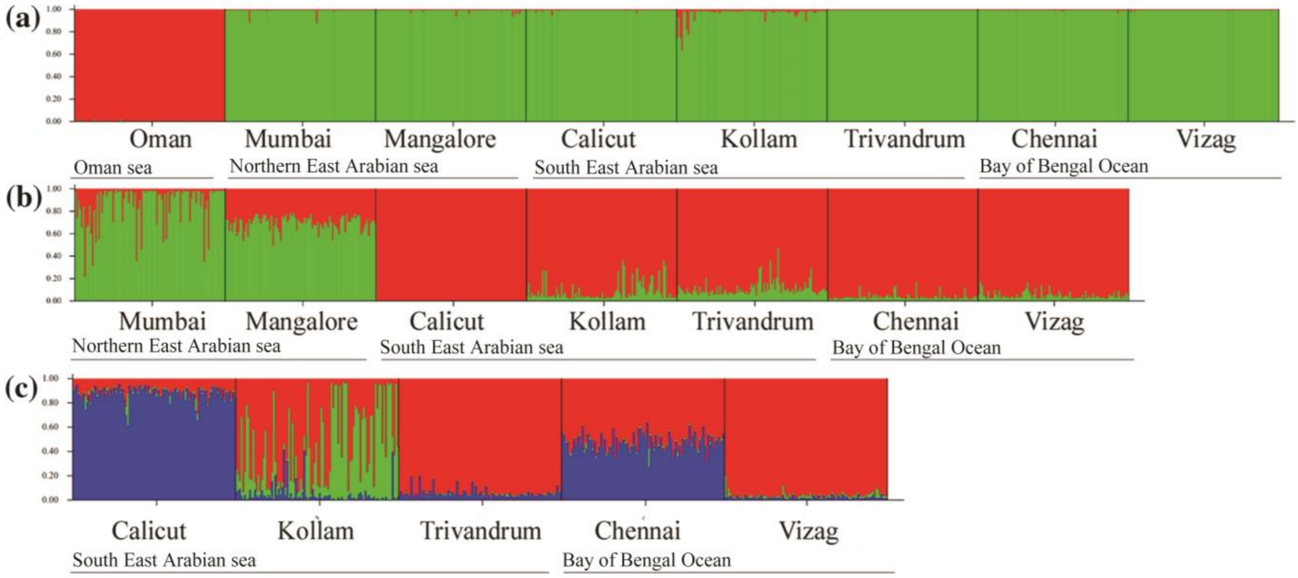


**Fig. 4.2** Graphical results of STRUCTURE analysis of six microsatellite loci in Indian oil Sardine populations. *Vertical lines* represent the probability of individual membership in simulated clusters. **a** Plot for K = 2 (including all the samples), **b** Plot for K = 2 (excluding Oman samples), and **c** K = 3 (Excluding Oman and Mumbai & Mangalore).

**Table 4.3** Results of analysis of molecular variance (AMOVA) for different hierarchical analysis of sardine populations.

| Structure tested | Observed partition | | | |
|---|---|---|---|---|
| | Variance | %total | F Statistics | P |
| 1. One gene pool (Mumbai, Mangalore, Calicut, Trivandrum, Kollam, Chennai, Vishakapatnam, Oman) | | | | |
| Among populations | 0.039 | 3.04 | - | - |
| Within populations | 1.247 | 96.96 | $F_{ST}$ = 0.03040 | <0.001 |
| 2. Two gene pools (Arabian Sea & Bay of Bengal_Mumbai, Mangalore, Calicut, Kollam, Trivandrum, Chennai, Vishakapatnam) vs. (Oman Sea_Oman) | | | | |
| Among groups | 0.04517 | 3.42 | $F_{CT}$ = 0.03422 | 0.12307 |
| Within groups | 0.02780 | 2.11 | $F_{SC}$ = 0.02181 | <0.001 |
| Within populations | 1.24695 | 94.47 | $F_{ST}$ = 0.05528 | <0.001 |
| 3. Three gene pools (Arabian Sea_Mumbai, Mangalore, Calicut, Kollam, Trivandrum) vs. (Bay of Bengal_Chennai, Vishakapatnam) vs. (Oman Sea_Oman) | | | | |
| Among groups | 0.01705 | 1.32 | $F_{CT}$ = 0.01319 | 0.17107 |
| Within groups | 0.02874 | 2.22 | $F_{SC}$ = 0.02253 | <0.001 |
| Within populations | 1.24695 | 96.46 | $F_{ST}$ = 0.03542 | <0.001 |
| 4. Six gene pools (Mumbai) vs. (Mangalore) vs. (Calicut, Kollam, Trivandrum) vs. (Chennai) vs. (Vishakapatnam) vs. (Oman) | | | | |
| Among groups | 0.03793 | 2.94 | $F_{CT}$ = 0.02940 | 0.13783 |
| Within groups | 0.00522 | 0.40 | $F_{SC}$ = 0.00417 | <0.001 |
| Within populations | 1.24695 | 96.65 | $F_{ST}$ = 0.03345 | <0.001 |

The result of AMOVA also revealed significant genetic structuring in our data (Table 4.3). The global AMOVA showed $F_{ST}$ value of 0.03040. The two gene pool comparison was (Oman vs. Indian coast) showing the highest $F_{ST}$ value (0.056, $p < 0.001$) with 95.47 variations within a population. In this case $F_{SC}$ = 0.02181 was significant $p \leq 0.001$and $F_{CT}$ = 0.03422 was not significant ($p = 0.12307$). This indicates data is still structured within the group. In Subsequent analysis with 3, 4 and 5 gene pool groups within-population difference decreased and $F_{CT}$ became significant with the highest level of significance in the 3 gene pool group. But in 6 gene pools groups [(Mumbai, Mangalore) vs. (Calicut) vs. (Kollam) vs. (Trivandrum) vs. (Chennai, Vizag) vs. (Oman)] within-population difference increased and $F_{CT}$ is not significant. The above results were also tested by Bayesian clustering analysis of individuals with Structurama. Structurama results supported five clusters (The highest probability value when K = 5). Barrier analysis on these data sets revealed two major barriers supported by five or more of the six loci. In it one (barrier 1) is between Oman and Mumbai, the other one (barrier 2) is between Mangalore and Calicut (Fig. 4.3). Barrier 1 isolated Oman sample from Indian Ocean and barrier 2 isolated samples from North-West Indian Ocean (Mumbai and Mangalore) from other coasts. The neighbour-joining tree constructed using $F_{ST}$ values showed a pattern concordant with geographic distance between sampling sites with $R^2$ value of 0.988 (Fig. 4.S4).

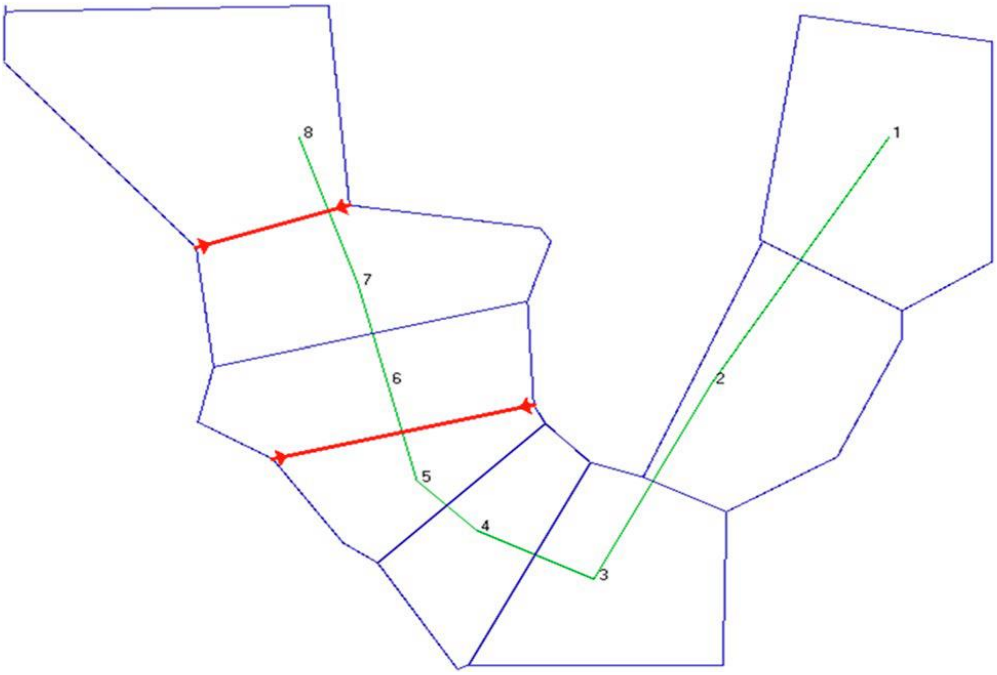


**Fig. 4.3** Genetic barrier to gene flow (*red lines*) among Indian oil sardine calculated using $F_{ST}$ and $R_{ST}$ matrix based on the samples from eight locations. *Blue lines* illustrate the Voronoi tessellation, *numerical letters* 1 to 8 represent population samples Vizag, Chennai, Trivandrum, Kollam, Calicut, Mangalore, Mumbai and Oman respectively.

A null hypothesis of no contributions of stepwise mutation models (SMM) to genetic differentiations $R_{ST} > pR_{ST} = F_{ST}$ was rejected ($p < 0.0001$) based on the data set (Table 4.S2). Overall multilocus $R_{ST}$ values were significantly higher than mean permutated $R_{ST}$ values ($R_{ST} = 0.0733$, $pR_{ST} = 0.0243$ and $p = 0.001$) showing the predominant role of stepwise mutation at the microsatellite loci compared to genetic drift and migration. Even though there are were large differences in $pR_{ST}$ values at each locus, pairwise tests between loci also demonstrated that the shift in average allele size had significantly contributed to population differentiation (Table 4.S3).

The result of the demographic bottleneck was analysed using the Wilcoxon signed-rank test under three microsatellite models (IAM, SMM and TPM). None of the populations showed significant heterozygosity excess under SSM and TPM model (Table 4.T4). However, both MUM and OMAN population showed significant heterozygosity excess under the IAM model.

From the population size parameter θ (i.e. 4 $Ne\mu$, where Ne is effective population size and $\mu$ are mutation rate) effective population size was calculated (assuming a microsatellite mutation rate of $10^{-4}$ per locus per generation) (Whittaker *et al*. 2003). It ranged from 7110 to 8014 respectively. Migration rate ($M$ i.e. $m/\mu$, where $m$ is immigration rate per generation and $\mu$ is mutation rate) analysis showed that there is no significant variation of some immigrants and emigrants between sampling sites observed (Table 4.S5).

## 4. Discussion

Microsatellite markers used in the present study could effectively distinguish three major sub-clusters in the Indian Ocean region with maximum genetic subdivision between Gulf of Oman and Indian coastline followed by another major subdivision within the Indian coastline between Mumbai & Mangalore vs. other parts. Except in these regions of isolation, individual genotypes indicated admixture and geographical connectivity despite the weak genetic structure. Variations in oceanographic and environmental parameters (temperature, salinity, pH and local currents) between geographical locations might have played a decisive role in shaping the genetic structure which needs to be investigated further using adaptive loci.

Even though marine pelagic fishes are characterized by large effective population sizes, high dispersal capacities, high fecundity and long planktonic larval phases, recent studies using microsatellite markers have indicated evidence for weak genetic differentiation in many of them (Gonzalez and Zardoya 2007; Borrell *et al*. 2012; Knutsen *et al*. 2003; Andre *et al*. 2011; Agostini *et al*. 2015; Candy *et al*. 2015) as observed in the present study. The weak genetic structure may be caused by barriers between local communities such as geographic distance, patchiness in the environment, local and global oceanic circulation patterns and environmental gradients which prevents population mixing to some extent (Bailey 1997; Oomen and Hutchings 2015). $F_{ST}$ and $R_{ST}$ were used for comparisons in the present study with $R_{ST}$ showing differentiation between all the samples. $F_{ST}$ is derived from an Infinite Allele Model (IAM) and $R_{ST}$ from a stepwise mutation model (SMM). Allele size variations were taken into account in $R_{ST}$ and hence the large range in allele size variations in abundant pelagic fishes makes them more variable and less meaningful. $F_{ST}$ and $R_{ST}$ ranged between 0.00106 and 0.08524 and 0.009 to 0.248 respectively, similar to those reported for marine pelagic fishes (Knutsen *et al*. 2003; Carlsson 2004; Fauvelot and Borsa 2011). Presence of the high number of private alleles (34.7%) which are uniformly distributed among population samples pointed towards restricted allele sharing between oil sardine samples from different locations. This indicated that comparatively high level of locus polymorphism is the reason for detecting weak genetic structuring (Oreilly *et al*. 2004; Borrell *et al*. 2012) and effect of size homoplasy is limited (Angers *et al*. 2000; Balloux and Lugon-Moulin 2002). The high number of private alleles detected can also be due to the relatively small sample size as the allelic richness was very high in sampled populations. Comparison of $F_{ST}$ and $R_{ST}$ has been used to check the relative contribution of mutation vs. migration rate to population structuring (Hardy *et al*. 2003). The allele size permutation test showed the predominant role of mutation to genetic differentiation of *S. longiceps* populations with relative less contribution of migration rate and drift. Migration rate analysis showed that there is no significant variation of some immigrants and emigrants between sampling sites. The reason for the significant deviations from Hardy Weinberg expectation in some loci, in some population, is the result of deficiencies of heterozygotes (Karlsson and Mork 2005). Possible reasons for the deficiencies were; patchy distribution of population as seen in the STRUCTURE analysis, or the existence of null alleles (Karlsson and Mork 2005).

The Bayesian clustering analysis rejected the null hypothesis of panmixia and inferred three major clusters. Cluster 1 (Oman) and cluster 2 (Mangalore & Mumbai) did not exhibit any pattern of admixture between genotypes, whilst, the presence of admixed genotypes was indicated in Cluster 3. Bayesian clustering analysis results were also supported by Barrier analysis indicating the presence of two strong barriers; first between Oman and Indian Ocean coastline and the second between Mumbai & Mangalore and other parts of the coast. Environmental and oceanographic barriers, larval retention or natal philopatry may be contributing to the restricted mixing of genotypes resulting in significant clustering. The third cluster (Calicut, Kollam, Trivandrum, Chennai, Vizag) in STRUCTURE analysis showed the presence of highly admixed genotypes, with some individuals strongly assigned to one population with the proportions assigned to each group being asymmetric. Use of SNP markers will help in identifying locally adapted populations as they scan functional gene regions which respond to environmental fluctuations by undergoing selection at various levels.

Sardine larvae are pelagic and planktonic with a larval duration of approximately 40 days (Kuthalingam 1960). Sardine shoals are reported to swim at a speed of 5 km/hour (Devaraj and Martosubroto 1997) but very little information is available regarding the migratory potential and pattern. Genetic differentiation is proportional to the number of migrants in each generation and the present study reveals their reduced potential for migration between some sites. It is also not known whether they exhibit any kind of natal philopatry, the fishes returning to spawning grounds. Such patterns of natal homing have been reported in Atlantic herring, (a small pelagic fish belonging to family Clupeidae) with return rates varying between 75-95% (Wheeler and Winters 1984). This pattern of natal homing will contribute substantially to the genetic subdivision among populations. But information regarding spawning grounds and behavioural patterns like natal homing in Indian oil sardine is lacking.

Mantel tests were significant in the present study, indicating the existence of a strong correlation between genetic and geographical distance among *S. longiceps* populations. In addition to this, a statistically significant correlation, even though weak was evident between genetic and environmental distance (temperature and salinity) while controlling geographic distance. The isolation by distance (IBD) and isolation by the environment (IBE) is a common pattern found in many other small marine pelagic fishes (Maes and Volckaert 2002; Bradbury and Bentzen 2007; Cunningham *et al*. 2009; Selkoe and Toonen 2011; Wang *et al*. 2013). But recently, it has been emphasized that hierarchical population structure can easily be mistaken for a pattern of IBD/IBE and the reverse is also possible (Meirmans 2012, 2015). The pattern of autocorrelation deriving from IBD can severely bias commonly used statistical tests like mantel tests and STRUCTURE analysis (Frantz *et al*. 2009). Hence it is advised to select the only biologically interpretable pattern to prevent over-interpretation of results along with alternate approaches of spatial clustering to infer genetic structuring in a data set (Meirmans 2015). Hence in the present study, we considered only the most prominently structured clusters for choosing the number of K in structure analysis. The output from Barrier analysis also supported this finding.

Oceanographic and environmental features of the Gulf of Oman show wide variations as compared to the Arabian Sea Open Ocean. The Arabian Sea Open Ocean exhibits a typical bimodal pattern of sea surface temperature with warming during spring intermonsoon (April-May) and fall intermonsoon (October-November) whereas cooling is observed during southwest monsoon (SWM) (June-September) and North-East monsoon seasons (December-March). Contrary to this, the Persian Gulf, the Gulf of Oman and the Red Sea exhibit a distinct unimodal Sea surface temperature with the lowest temperature during the North-East monsoon season and highest temperature during South-West monsoon season (Rao *et al*. 1992). An intense upwelling has also been reported along the Oman coast during May-June which lasts until October. The average sea surface salinity and chlorophyll levels are also higher along the Oman coast. High salinity in this region is due to the excess of evaporation over precipitation and runoff of high saline water from the Persian Gulf (Qasim 1982). The Upwelling along the Somalia and Oman get intensified during the summer monsoon and enhance primary productivity by bringing a higher amount of nutrients into the upper ocean (Shi *et al*. 2000). Thus, the Gulf of Oman and Indian Ocean coastline can be considered as two distinct marine eco-regions. These factors may act as barriers for the gene flow between the Gulf of Oman and Indian coastline.

The Arabian Sea along the North-West coast of India is comparatively more saline than along the South-West coast of India as it is adjacent to the Persian Gulf and Red sea. During winter, a winter cooling is observed along this coast (October-December) which increases productivity (Prasanna Kumar *et al*. 2002). The northeast trade winds during winter bring in dry continental air into the northern Arabian Sea which enhances evaporation. The combination of this increased evaporation and reduced solar radiation during winter season results in a significant decrease of SST and occurrence of cold surface waters in the Northern Arabian Sea during winter. On the contrary, cooling is observed along the South-West coast of India during the South-West monsoon season (June-August). Also, South-West monsoon causes a reduction in the salinity levels along this region (Rao *et al*. 1992). The Malabar upwelling zone along the South West Indian coast is one of the strongest upwelling zones among world oceans and the upwelling along these coasts is mainly wind-induced occurring during June-August (Bakun *et al*. 1998) resulting in cooling of waters and higher productivity. The fish catch composition along these coasts also shows a clear difference (Madhupratap *et al*. 2001). The observed genetic differentiation between North-West coast and other regions of the Indian Ocean may be related to these environmental and oceanographic features which cause restricted mixing and gene flow. The Bay of Bengal is less productive, cooler and less saline than the Arabian Sea on an average and the reasons for low salinity are increased precipitation along with runoff from three major river systems; the Ganges-Bhramaputra, Irrawadi-Salween and the Krishna-Godavari. There are no major upwelling events along the Bay of Bengal region and thus these waters are less productive and more stratified.

These shifts in habitat characteristics may not be acting strictly as a barrier to migration but may prevent successful larval dispersal and subsequent colonization (Marshall and Morgan 2011). Recruits from non-matching natal environments may be negatively selected due to the competitive disadvantage of genotypes which are adapted to their natal habitat (Marshall *et al*. 2010). Larval retention in natal habitats due to biophysical conditions of the ocean like eddies or gyres will also contribute to restricted mixing and reduced connectivity. The formation of eddies and gyres in the Gulf of Aden, Gulf of Oman and in the Arabian sea as a whole has been reported during South-West and North-East monsoon seasons (Jouanno *et al*. 2012) which coincides with the peak spawning of Indian oil sardine. Similarly, in the Bay of Bengal also, eddies are reported to occur during March-August near the western boundary of Bay of Bengal (Prasanna Kumar *et al*. 2004) which is confined to upper 500 m of the water column with a horizontal dimension of 200-300 km. These bio-physical factors may act as barriers to larval dispersal and mixing.

Surface circulation in the Indian Ocean is undergoing fluctuations in semi-annual scale and that is the reason for the semi-annual reversal of monsoon in the Indian subcontinent (Wyrtki 1973). The major surface currents in the Indian Ocean are summarized in the in Fig. 4.S5. During the winter (northeast monsoon), the surface current systems in the Indian ocean are similar to the general circulation patterns in the Pacific and Atlantic oceans. The Equatorial Counter Current (ECC), the Northeast Monsoon Current (NMC) and the south equatorial current (SEC, not shown in the figure) are the major current observed in this season. Whereas during the summer (southwest monsoon), the surface currents change remarkably from other oceans. The eastward flowing Southwest Monsoon Current (SMC), replaces the westward flowing NMC and the northward flow Somali Current (SC), replaces the southward flowing SC along the Somali coast. A unique surface flow known as Equatorial Jet (EJ) is observed during the transition period of monsoon (April–May and November–December) (Wyrtki 1973). Similar to the above open ocean currents, the boundary currents along the coastal region is also undergoing seasonal reversals (Shetye and Gouveia 1998). The West India Coastal Current (WICC) flows towards north pole during winter and towards south pole during summer along the west coast of India. The East India Coastal Current (EICC) flows towards north pole during winter and towards the south pole during summer (Wyrtki 1973; Shankar *et al.* 2002). The WICC that flows towards south pole during summer along the west coast of India towards Bay of Bengal along NMC may have a significant role in the genetic connectivity observed between Southeast Arabian sea and Bay of Bengal populations. Somali Current (SC), has a significant role in the Northern Arabian Sea along the Oman-Somalia coast. It's northward (summer) and southward (winter) flow influence the productivity (winter cooling) and other coastal ocean characters along Oman-Somalia coast making it a unique ecosystem. This may be the reason for the high genetic differentiation observed with Oman samples. Even though the sea surface circulations promote larval dispersal of *S. longiceps* in the Indian Ocean shifts in habitat characteristics may be acting strictly as a barrier to prevent successful larval dispersal and subsequent colonization (Marshall and Morgan 2011). The WICC possibly disperse larvae along the west coast of India. Despite that, the minor genetic differentiation between Northeast Arabian sea and other Indian coastal samples indicates the predominant role of habitat characteristics in restricted gene flow.

Previous works by Venkita Krishnan (1993) and Mohandas (1997) have pointed to the possibility of sub-structuring in Indian oil sardines. The population genetic structure and historical demography of Indian oil sardine using mitochondrial DNA markers (Sukumaran *et al*. 2016b) indicated the presence of a single evolutionary unit with signals of historic expansion. But microsatellite markers used in the present study could efficiently detect subtle patterns of population genetic structure showing that these markers are more efficient in delineating contemporary patterns of gene flow from historic patterns. Such discordant patterns of genetic structure using mitochondrial and microsatellite markers have been observed in European Sardine, *Sardina pilchardus* (Gonzalez and Zardoya 2007) and other marine fishes (Buonaccorsi *et al*. 2001; Brown *et al*. 2005; Da Silva *et al*. 2015). In all the analysed populations in the present study, high values of genetic diversity (0.856-0.931) were observed within populations without any geographical trend and the values were comparable with other sardine species and anchovies (Pereyra *et al*. 2004; Zarraonaindia *et al*. 2009; Ruggeri *et al*. 2013).

Genetic subdivisions may arise by adaptation of eggs, larvae and adults to local environmental factors in the ocean thus giving resilience to environmental fluctuations (Nielsen *et al*. 2009). But microsatellite markers are assumed to be neutral and hence signatures of selection cannot be detected. Recent studies have found evidence for the detection of signatures of selection in microsatellites which are found in association with functional genes (Larsson *et al*. 2007). In the present study, the microsatellite loci were randomly selected and hence we cannot make any conclusions regarding its role in environmental adaptation. Of late, population genomic studies using genome scans could identify genetic markers which are diverged highly among populations which do not conform to statistical expectations based on a neutral genetic model and these markers are located inside genomic regions where gene loci are under selection (Nielsen *et al*. 2012). Several studies have pointed out the existence of locally adapted fish populations in different environmental clines (Larsen *et al*. 2007; Johannesson *et al*. 2011; Teacher *et al*. 2013; Wang *et al*. 2013; Andre *et al*. 2016; Brennan *et al*. 2016) and gene-associated SNPs have been used as high-resolution tools for population genomic studies and population traceability (Nielsen *et al*. 2012). It has also been reported significant morphological divergence in Indian oil sardines from different locations (Sukumaran *et al*. 2016a) which provide vital clues to the existence of locally adapted populations.

Microsatellite markers used in the present study could effectively delineate the presence of subpopulations in Indian oil sardine. Due to large annual fluctuations in the population size and variability in recruitment of this species, intense fishing pressure is a significant threat to regional sub-populations as it leads to low population size. Such reductions in genetic diversity and stock complexity will affect the resilience of stocks and their ability to recover from low abundance due to extreme climatic events or habitat destruction. Hence management measures are to be devised on a regional scale by assessing the number of spawning components and behavioural groups to preserve stock complexity and prevent overexploitation. Further studies should be carried out by conducting large scale genome scans using SNP markers linked to selection to identify loci under selection and consequent adaptation. Genomic approaches integrated with local ecological sampling will help to get more insight into the exact ecological and genetic mechanism shaping population genetic structure of Indian oil sardine.

## Supplementary Figures and Tables

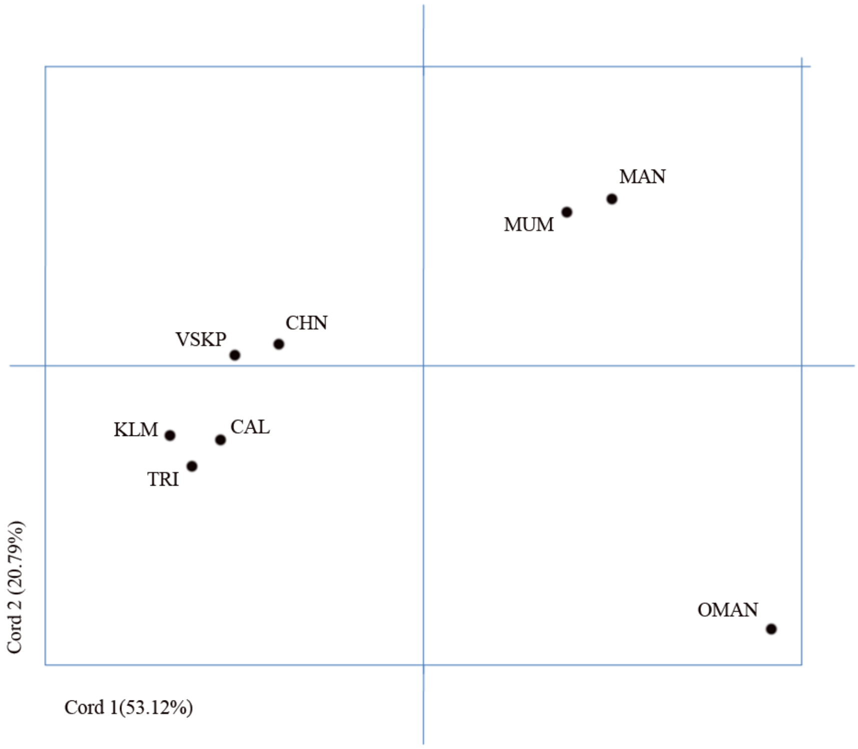


**Fig. 4.S1** Principal component analysis (PCA) based on allele frequency for all the populations.Oman sea_OMAN, North East Arabian sea_MUM & MAN, South East Arabian sea_CAL, KLM & TRI, Bay Of Bengal ocean_ CHN & VSKP.

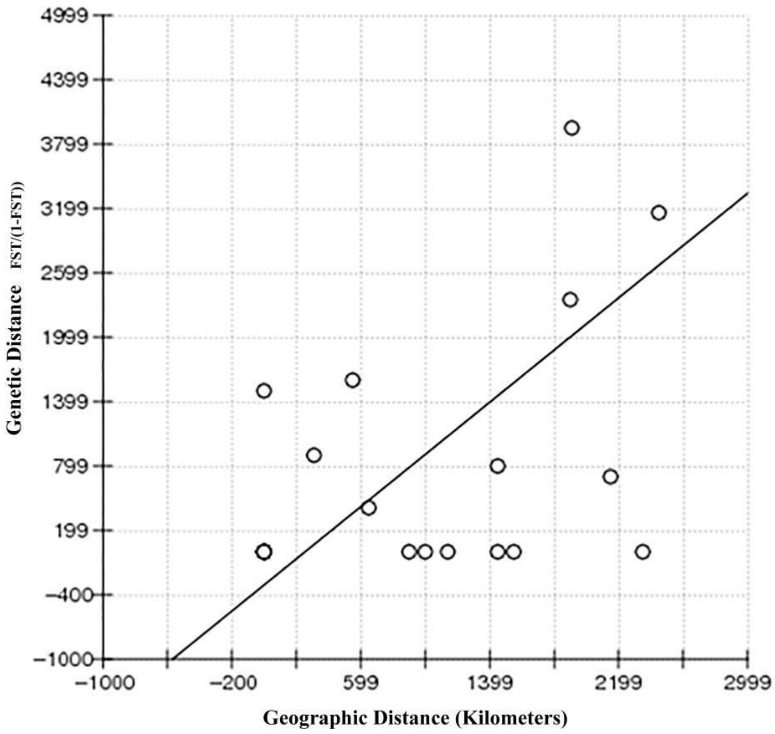
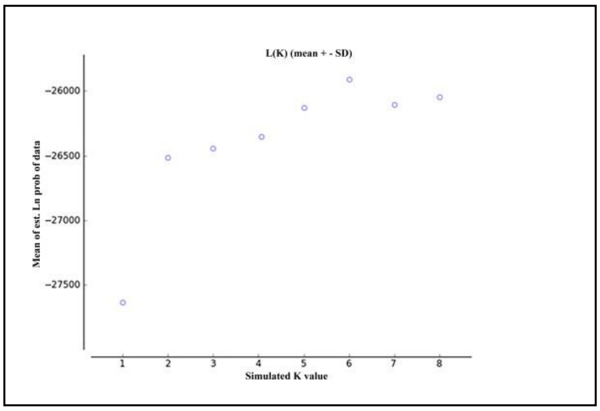
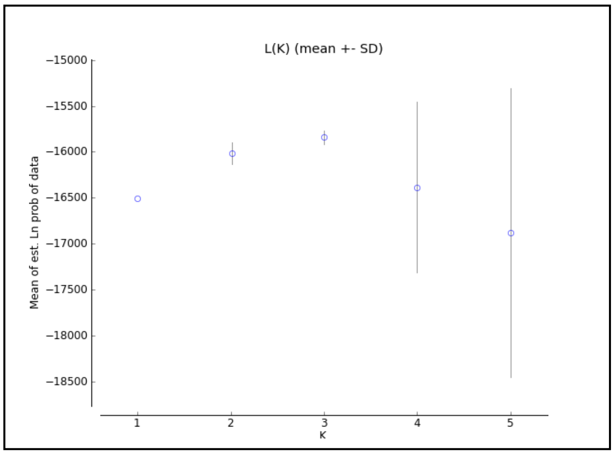


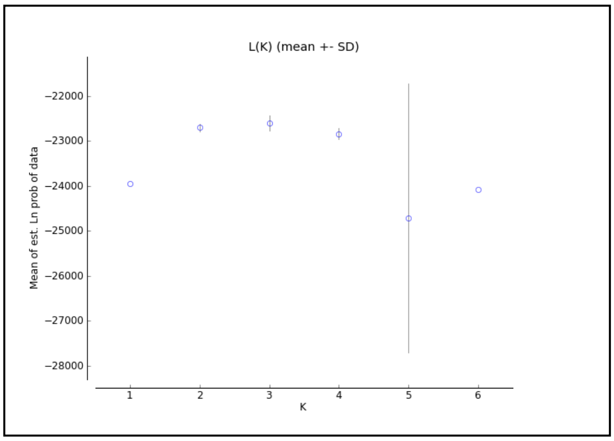
**Fig. 4.S2** Genetic isolation by distance in Indian oil sardine population samples inferred from multilocus estimates of $F_{ST}$ and geographical distance (r = 0.5342, p = 0.003).



a) Populations included in the STRUCTURE analysis - Oman sea_Oman, North East Arabian sea_Mumbai & Mangalore, South East Arabian sea_Calicut, Kollam & Trivandrum,Bay Of Bengal ocean_Chennai & Vizag.



b) Populations included in the STRUCTURE analysis - North East Arabian sea_Mumbai & Mangalore, South East Arabian sea_Calicut, Kollam & Trivandrum,Bay Of Bengal ocean_Chennai & Vizag.
.



c) Populations included in the STRUCTURE analysis -South East Arabian sea_Calicut, Kollam & Trivandrum,Bay Of Bengal ocean_Chennai & Vizag.

**Fig. 4.S3** Probability of each assumed population (K) of Indian oil sardine populations expressed as the mean of likelihood, Ln prob of data [Ln P(D)].

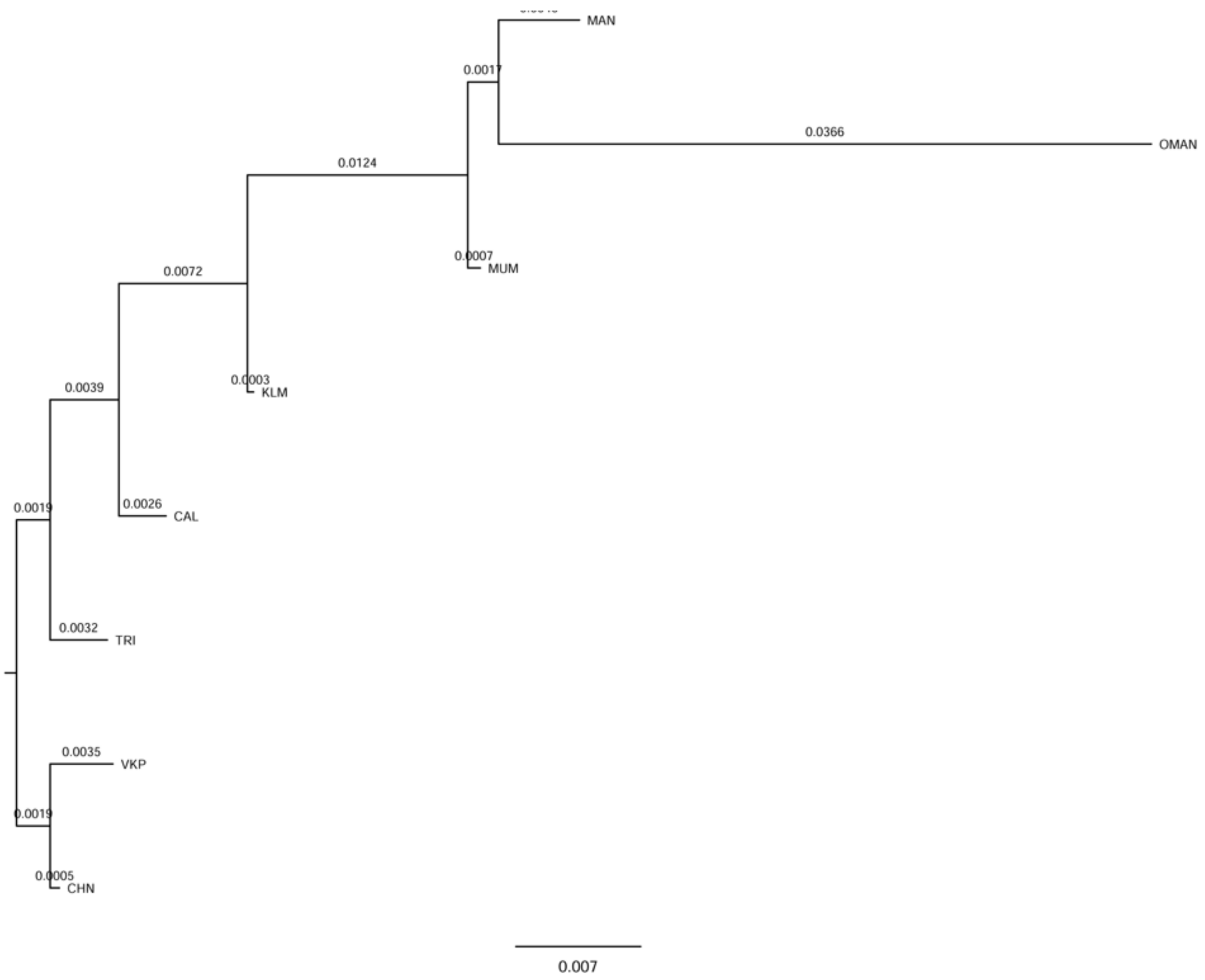


**Fig. 4.S4 Neighbour-joining tree constructed using $F_{ST}$ values of 8 populations of *S. longiceps*.** Branch length is represented in a decimal number, $R^2$ = 0. 988.Oman Sea_Oman (OMAN), North East Arabian Sea_Mumbai (MUM) & Mangalore (MAN), South East Arabian Sea_Calicut (CAL), Kollam (KLM) & Trivandrum (TRI), Bay of Bengal Ocean_Chennai (CHN) & Vizag (VKP).

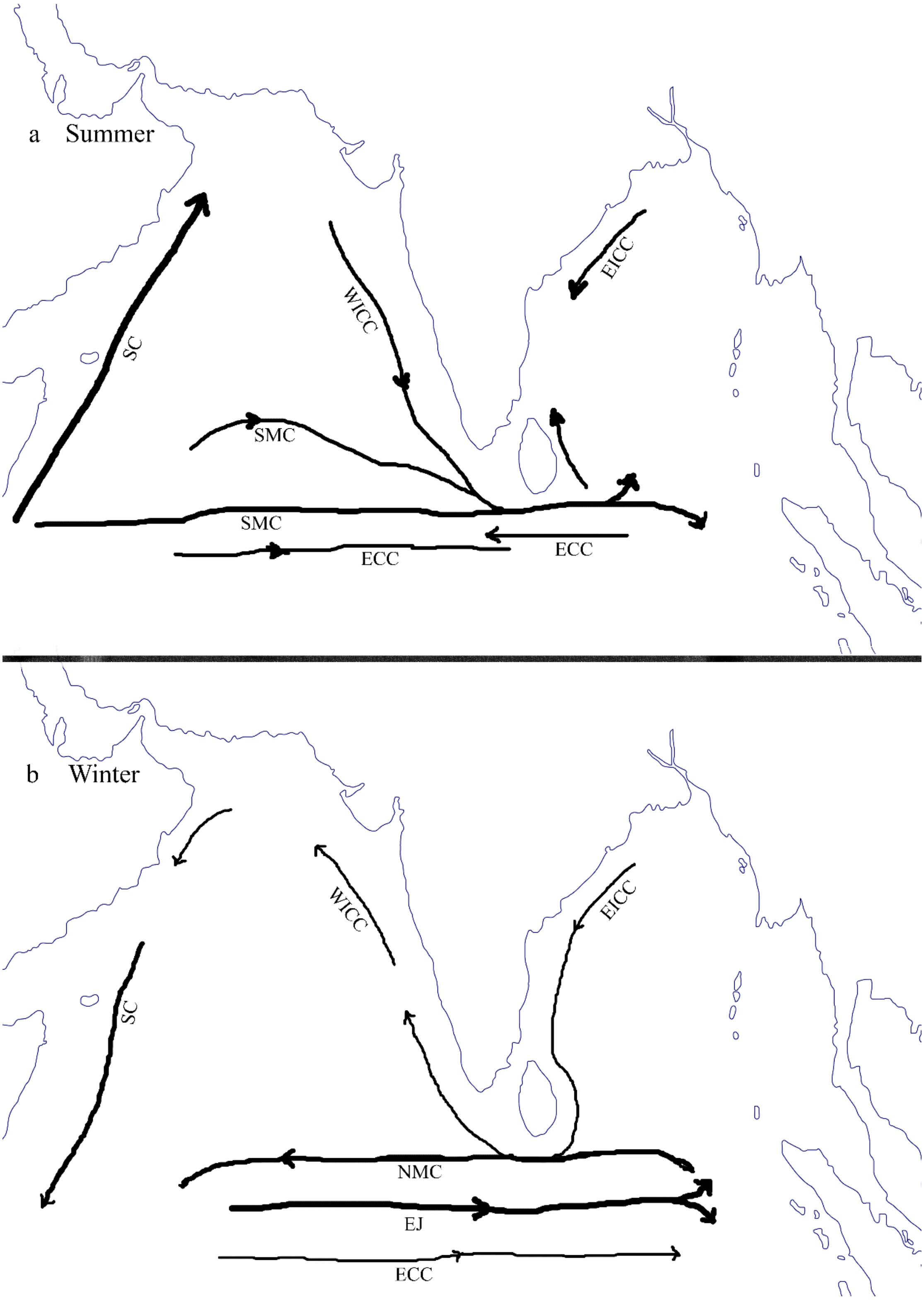


**Fig. 4.S5** Schematic representation of major surface currents in the Indian Ocean during (a) the southwest monsoon (summer) and (b) the northeast monsoon (winter). The major currents are Northeast Monsoon Current (NMC), Equatorial Counter Current (ECC), Equatorial Jet (EJ), Somali Current (SC), Southwest Monsoon Current (SMC), West India Coastal Current (WICC) and East India Coastal Current (EICC). The EJ appears only during the transition period (summer to winter monsoon season) in April-May and November December.

**Table 4.S1** Estimation of statistical power of microsatellite loci using POWSIM.

| Generation time | *Ne* | Expected $F_{ST}$ | Average $F_{ST}$ | $Chi^2$ test | Fisher test |
|---|---|---|---|---|---|
| 00 | 7000 | 0.0000 | 0.0000 | 0.0340 | 0.0500 |
| 14 | 7000 | 0.0010 | 0.0010 | 0.2220 | 0.3020 |
| 35 | 7000 | 0.0025 | 0.0025 | 0.9260 | 0.9600 |
| 70 | 7000 | 0.0050 | 0.0050 | 1.0000 | 1.0000 |
| 140 | 7000 | 0.0100 | 0.0100 | 1.0000 | 1.0000 |
| 282 | 7000 | 0.0200 | 0.0199 | 1.0000 | 1.0000 |
| 354 | 7000 | 0.0250 | 0.0250 | 1.0000 | 1.0000 |
| 718 | 7000 | 0.0500 | 0.0500 | 1.0000 | 1.0000 |

**Table 4.S2** Summary statistics of allele size permutation test for each locus.

| LOCUS NAME | $F_{ST}$ | $pR_{ST}$ (95% C.I) | $R_{ST}$ |
|---|---|---|---|
| SAR 9 | 0.0171 | 0.0153 (-0.0031 to 0.0821) | 0.0778* |
| SAR B-A08 | 0.0560 | 0.0443 (-0.1643 to 0.2256) | 0.2437* |
| SAR B-D09 | 0.0045 | 0.0043 (-0.0034 to 0.0168) | 0.0024* |
| Sar1-D01 | 0.0079 | 0.0071 (-0.0027 to 0.0278) | 0.0066 |
| Sar1-D06 (B) | 0.0028 | 0.0026 (-0.0038 to 0.0132) | 0.0037 |
| Sar1-H11 (B) | 0.0822 | 0.0654 (0.0033 to 0.1621) | 0.2138 |
| MULTI LOCUS | 0.0271 | 0.0243 (0.0049 to 0.0594) | 0.0733* |

* Indicate a significant test after 2000 random permutation.

**Table 4.S3** $R_{ST}$/$pR_{ST}$ results for all loci from SPAGeDI with p values.

| | | All Loci | All loci | All loci | SAR 9 | SAR B-A08 | SAR B-D09 | Sar1-D01 | Sar1-D06 (B) | Sar1-H11 (B) |
|---|---|---|---|---|---|---|---|---|---|---|
| Pairwise Locations | | Observed value | Mean permuted value | p (obs < exp) | | | | | | |
| BOM | MAN | 0.026871 | 0.006371 | 0.967 | 0.7982 | 0.987 | 0.018 | 0.8402 | 0.9441 | 0.8931 |
| BOM | CAL | 0.041107 | 0.015272 | 0.9161 | 0.8392 | 0.2158 | 0.028 | 0.3816 | 0.8851 | 0.997 |
| BOM | TRI | 0.024866 | 0.016469 | 0.7473 | 0.7962 | 0.8661 | 0.2727 | 0.6683 | 0.8062 | 0.9391 |
| BOM | CHN | 0.01067 | 0.019735 | 0.5305 | 0.9411 | 0.0649 | 0.043 | 0.5315 | 0.3886 | 0.8162 |
| BOM | VSKP | -0.00049 | 0.019428 | 0.1099 | 0.7073 | 0.9111 | 0.7642 | 0.7273 | 0.2767 | 0.012 |
| BOM | KLM | 0.004814 | 0.012452 | 0.2957 | 0.978 | 0.2897 | 0.8531 | 0.5275 | 0.3217 | 0.3417 |
| BOM | OMAN | 0.222273 | 0.026955 | 1 | 0.993 | 1 | 0.005 | 0.5005 | 0.004 | 1 |
| MAN | CAL | 0.062378 | 0.022404 | 0.9321 | 0.4326 | 0.952 | 0.019 | 0.8402 | 0.3317 | 0.998 |
| MAN | TRI | 0.03183 | 0.026821 | 0.7283 | 0.2997 | 0.7113 | 0.3896 | 0.6763 | 0.3417 | 0.9341 |
| MAN | CHN | 0.036578 | 0.029026 | 0.7712 | 0.8042 | 0.9281 | 0.1099 | 0.7273 | 0.8462 | 0.8561 |
| MAN | VSKP | 0.028063 | 0.026451 | 0.6314 | 0.2468 | 0.6843 | 0.8182 | 0.9081 | 0.8961 | 0.1888 |
| MAN | KLM | 0.018609 | 0.01833 | 0.5864 | 0.963 | 0.9421 | 0.8631 | 0.5594 | 0.8811 | 0.5744 |
| MAN | OMAN | 0.209812 | 0.027412 | 1 | 0.989 | 1 | 0.01 | 0.6663 | 0.984 | 1 |
| CAL | CAL | 0.000974 | 0.006181 | 0.3007 | 0.6803 | 0.7562 | 0.2478 | 0.4545 | 0.1409 | 0.9391 |
| CAL | CHN | 0.008312 | 0.002218 | 0.8871 | 0.6384 | 0.3916 | 0.0969 | 0.5185 | 0.7203 | 0.999 |
| CAL | VSKP | 0.035773 | 0.004108 | 0.998 | 0.6733 | 0.7123 | 0.7672 | 0.5824 | 0.7243 | 1 |
| CAL | KLM | 0.026059 | 0.011451 | 0.9011 | 0.9221 | 0.1099 | 0.7892 | 0.3167 | 0.6024 | 1 |
| CAL | OMAN | 0.288552 | 0.059289 | 1 | 0.978 | 1 | 0.005 | 0.2667 | 0.9011 | 1 |
| TRI | CHN | 0.002067 | 0.004641 | 0.4446 | 0.958 | 0.7173 | 0.4386 | 0.4396 | 0.6643 | 0.9351 |
| TRI | VSKP | 0.022326 | 0.004438 | 0.981 | 0.5694 | 0.042 | 0.9211 | 0.8162 | 0.7353 | 1 |
| TRI | KLM | 0.013083 | 0.011418 | 0.6374 | 0.992 | 0.8192 | 0.957 | 0.4356 | 0.6953 | 0.9181 |
| TRI | OMAN | 0.25856 | 0.068115 | 1 | 0.986 | 1 | 0.2498 | 0.4805 | 0.8911 | 0.996 |
| CHN | VSKP | 0.008841 | 0.004172 | 0.8062 | 0.965 | 0.7692 | 0.7632 | 0.7453 | 0.009 | 0.9411 |
| CHN | KLM | 0.002922 | 0.00996 | 0.2567 | 0.8761 | 0.3397 | 0.7802 | 0.4286 | 0.032 | 0.6893 |
| CHN | OMAN | 0.253944 | 0.07021 | 1 | 0.961 | 1 | 0.005 | 0.3407 | 0.5195 | 0.997 |
| VSKP | KLM | 0.003946 | 0.004696 | 0.5355 | 0.997 | 0.8811 | 0.4026 | 0.6803 | 0.021 | 0.4096 |
| VSKP | OMAN | 0.20224 | 0.046016 | 1 | 0.986 | 1 | 0.6923 | 0.6474 | 0.2867 | 0.972 |
| KLM | OMAN | 0.209154 | 0.032641 | 1 | 0.976 | 1 | 0.7323 | 0.6474 | 0.3686 | 1 |
| BOM | MAN | 0.026871 | 0.006371 | 0.967 | 0.7982 | 0.987 | 0.018 | 0.8402 | 0.9441 | 0.8931 |
| BOM | CAL | 0.041107 | 0.015272 | 0.9161 | 0.8392 | 0.2158 | 0.028 | 0.3816 | 0.8851 | 0.997 |
| BOM | TRI | 0.024866 | 0.016469 | 0.7473 | 0.7962 | 0.8661 | 0.2727 | 0.6683 | 0.8062 | 0.9391 |
| BOM | CHN | 0.01067 | 0.019735 | 0.5305 | 0.9411 | 0.0649 | 0.043 | 0.5315 | 0.3886 | 0.8162 |
| BOM | VSKP | -0.00049 | 0.019428 | 0.1099 | 0.7073 | 0.9111 | 0.7642 | 0.7273 | 0.2767 | 0.012 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| BOM | KLM | 0.004814 | 0.012452 | 0.2957 | 0.978 | 0.2897 | 0.8531 | 0.5275 | 0.3217 | 0.3417 |
| BOM | OMAN | 0.222273 | 0.026955 | 1 | 0.993 | 1 | 0.005 | 0.5005 | 0.004 | 1 |
| MAN | CAL | 0.062378 | 0.022404 | 0.9321 | 0.4326 | 0.952 | 0.019 | 0.8402 | 0.3317 | 0.998 |
| MAN | TRI | 0.03183 | 0.026821 | 0.7283 | 0.2997 | 0.7113 | 0.3896 | 0.6763 | 0.3417 | 0.9341 |
| MAN | CHN | 0.036578 | 0.029026 | 0.7712 | 0.8042 | 0.9281 | 0.1099 | 0.7273 | 0.8462 | 0.8561 |
| MAN | VSKP | 0.028063 | 0.026451 | 0.6314 | 0.2468 | 0.6843 | 0.8182 | 0.9081 | 0.8961 | 0.1888 |
| MAN | KLM | 0.018609 | 0.01833 | 0.5864 | 0.963 | 0.9421 | 0.8631 | 0.5594 | 0.8811 | 0.5744 |
| MAN | OMAN | 0.209812 | 0.027412 | 1 | 0.989 | 1 | 0.01 | 0.6663 | 0.984 | 1 |
| CAL | Pop4 | 0.000974 | 0.006181 | 0.3007 | 0.6803 | 0.7562 | 0.2478 | 0.4545 | 0.1409 | 0.9391 |
| CAL | CHN | 0.008312 | 0.002218 | 0.8871 | 0.6384 | 0.3916 | 0.0969 | 0.5185 | 0.7203 | 0.999 |
| CAL | VSKP | 0.035773 | 0.004108 | 0.998 | 0.6733 | 0.7123 | 0.7672 | 0.5824 | 0.7243 | 1 |
| CAL | KLM | 0.026059 | 0.011451 | 0.9011 | 0.9221 | 0.1099 | 0.7892 | 0.3167 | 0.6024 | 1 |
| CAL | OMAN | 0.288552 | 0.059289 | 1 | 0.978 | 1 | 0.005 | 0.2667 | 0.9011 | 1 |
| TRI | CHN | 0.002067 | 0.004641 | 0.4446 | 0.958 | 0.7173 | 0.4386 | 0.4396 | 0.6643 | 0.9351 |
| TRI | VSKP | 0.022326 | 0.004438 | 0.981 | 0.5694 | 0.042 | 0.9211 | 0.8162 | 0.7353 | 1 |
| TRI | KLM | 0.013083 | 0.011418 | 0.6374 | 0.992 | 0.8192 | 0.957 | 0.4356 | 0.6953 | 0.9181 |

Oman sea_Oman (OMAN), North East Arabian sea_Mumbai (MUM) & Mangalore (MAN), South East Arabian sea_Calicut (CAL), Kollam (KLM) & Trivandrum (TRI), Bay of Bengal ocean_Chennai (CHN) & Vizag (VSKP).

**Table 4.S4** Wilcoxon signed-rank test under three different mutational models for detecting recent population bottleneck in *S. longiceps.*

| Population | IAM | SMM | TPM |
|---|---|---|---|
| MUM | 0.0030* | 0.4730 | 0.0640 |
| MAN | 0.4650 | 0.1440 | 0.2130 |
| CAL | 0.5810 | 0.5810 | 0.1600 |
| TRI | 0.2750 | 0.1000 | 0.2800 |
| KLM | 0.3220 | 0.4720 | 0.1110 |
| CHN | 0.2970 | 0.4530 | 0.4310 |
| VSKP | 0.3251 | 0.2341 | 0.1768 |
| OMAN | 0.0042* | 0.0891 | 0.0982 |

*Indicate significant values.Oman sea_Oman (OMAN), North East Arabian sea_Mumbai (MUM) & Mangalore (MAN), South East Arabian sea_Calicut (CAL), Kollam (KLM) & Trivandrum (TRI), Bay Of Bengal Ocean_Chennai (CHN) & Vizag (VSKP).

**Table 4.S5** Maximum likelihood estimation of the population size parameter θ (i.e 4 $N_e\mu$, where $N_e$ is effective population size and $\mu$ is mutation rate) and scaled migration rate $M$ (i.e $m/\mu$, where *m*is immigration rate per generation and $\mu$ is mutation rate) for *S. longiceps.*

| Migration rate *M** | | | | | | | | | $N_e$# |
|---|---|---|---|---|---|---|---|---|---|
| Location | BOM | MAN | CAL | TRI | CHN | VSKP | KLM | OMAN | |
| BOM | | 0.98 | 0.88 | 0.61 | 0.84 | 0.96 | 0.90 | 0.74 | 8011 |
| MAN | 0.75 | | 0.96 | 0.85 | 0.70 | 0.77 | 0.62 | 0.96 | 7224 |
| CAL | 0.99 | 0.92 | | 0.75 | 0.93 | 0.78 | 0.69 | 0.76 | 7258 |
| TRI | 0.63 | 0.76 | 0.80 | | 0.78 | 0.80 | 0.76 | 0.85 | 7176 |
| CHN | 0.87 | 0.92 | 0.96 | 0.79 | | 0.81 | 0.95 | 0.89 | 8014 |
| VSKP | 0.91 | 0.90 | 0.59 | 0.63 | 0.68 | | 0.85 | 0.79 | 7576 |
| KLM | 0.83 | 0.64 | 0.59 | 0.66 | 0.69 | 0.84 | | | 7110 |
| OMAN | 0.67 | 0.98 | 0.88 | 0.88 | 0.84 | 0.73 | 0.99 | 0.81 | 7510 |

*Rows and columns are donor and recipient populations. #$N_e$ was calculated assuming a microsatellite mutation rate of $10^{-4}$ per locus per generation.

**Table 4.S6** Average number of private alleles per locus in each population

| Location and locus parameters | Abbreviations | Average number of alleles per locus | Average number of private alleles per locus |
|---|---|---|---|
| BOMBAY | BOMB | 25 | 11.082 |
| MANGLURU | MAN | 31.5 | 10.922 |
| CALICUT | CAL | 31.5 | 9.991 |
| KOLLAM | KLM | 30.33 | 10.83 |
| TRIVANDRUM | TRI | 29.17 | 10.946 |
| CHNENNAI | CHN | 29.83 | 9.795 |
| VISHAKAPATNAM | VKP | 30.5 | 10.088 |
| OMAN | OMAN | 39.5 | 12.073 |

## 5. References

1. Agostini C, Patarnello T, Ashford JR, Torres JJ, Zane L, Papetti C (2015) Genetic differentiation in the ice-dependent fish *Pleuragramma antarctica* along the Antarctic *Peninsula. J Biogeogr* 42(6):1103-1113
2. Alheit J, Oozeki Y, Roy C (2009) Climate change and small pelagic fish. Cambridge University Press, Cambridge Al-Jufaili SM (2012) Reproductive biology of the Indian oil sardine *Sardinella longiceps* from al-seeb waters off oman. *Fis Aquacult J* 2012:1
3. Andre C *et al* (2011) Detecting population structure in a high geneflow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity* 106(2):270-280
4. Andre C *et al* (2016) Population structure in Atlantic cod in the eastern North Sea-Skagerrak-Kattegat: early life stage dispersal and adult migration. *BMC Res Notes* 9(1):1
5. Andrew R (2014) Tree figure drawing tool version 1.4.2 2006-2014, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree.
6. Angers B, Estoup A, Jarne P (2000) Microsatellite size homoplasy, SSCP, and population structure: a case study in the freshwater snail *Bulinus truncatus*. *Mol Biol Evol* 17(12):1926-1932
7. Bailey KM (1997) Structural dynamics and ecology of flatfish populations. *J Sea Res* 37(3):269-280
8. Bakun A, Roy C, Lluch-Cota S (1998) Coastal upwelling and other processes regulating ecosystem productivity and fish production in the Western Indian Ocean. In: Sherman K, Okemwa E, Ntiba M. (eds) Large marine ecosystems of the Indian Ocean: assessment, sustainability and management. Blackwell Science, Cambridge, pp 103-142
9. Balloux F, Lugon - Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Mol Ecol* 11(2):155-165
10. Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22(3):341-345
11. Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152(2):763-773
12. Bonnet E, Van de Peer Y (2002) zt: a software tool for simple and partial Mantel tests. *J Stat Softw* 7(10):1-2
13. Borrell YJ, Pinera JA, Prado JA, Blanco G (2012) Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES J Mar Sci* 69(8):1357-1371
14. Bradbury IR, Bentzen P (2007) Non-linear genetic isolation by distance: implications for dispersal estimation in anadromous and marine fish populations. *Mar Ecol Prog Ser* 340:245-257
15. Bradbury IR, Laurel B, Snelgrove PV, Bentzen P, Campana SE (2008) Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. *Proc R Soc Lond B* 275(1644):1803-1809
16. Brennan RS, Hwang R, Tse M, Fangue NA, Whitehead A (2016) Local adaptation to osmotic environment in killifish, *Fundulus heteroclitus*, is supported by divergence in swimming performance but not by differences in excess post-exercise oxygen consumption or aerobic scope. *Comp Biochem Physiol A* 196:11-19
17. Brown KM, Baltazar GA, Hamilton MB (2005) Reconciling nuclear microsatellite and mitochondrial marker estimates of population structure: breeding population structure of Chesapeake Bay striped bass (*Morone saxatilis*). *Heredity* 94(6):606-615
18. Buonaccorsi VP, McDowell JR, Graves JE (2001) Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*). *Mol Ecol* 10(5):1179-1196
19. Cadrin SX, Kerr LA, Mariani S (2013) Stock identification methods: applications in fishery science. Academic Press, Amsterdam
20. Candy JR, Campbell NR, Grinnell MH, Beacham TD, Larson WA, Narum SR (2015) Population differentiation determined from putative neutral and divergent adaptive genetic markers in

Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Mol Ecol Resour* 15(6):1421-1434

21. Carlsson J, Mcdowell JR, Diaz-Jaimes PI, Carlsson JE, Boles SB, Gold JR, Graves JE (2004) Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. *Mol Ecol* 13(11):3345-3356
22. Carvalho GR, Hauser L (1995) Molecular genetics and the stock concept in fisheries. In: Gary RC, Tony JP (eds) Molecular genetics in fisheries. Springer, Netherlands, pp 55-79
23. Chatterjee A, Shankar D, Shenoi SS, Reddy GV, Michael GS, Ravichandran M, Gopalkrishna VV, Rao ER, Bhaskar TU, Sanjeevan VN (2012) A new atlas of temperature and salinity for the North Indian Ocean. *J Earth Syst Sci* 121(3):559-593
24. CMFRI (2015) Annual report 2014-2015. Central Marine Fisheries Research Institute, Kochi
25. Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144(4):2001-2014
26. Cowen RK, Paris CB, Srinivasan A (2006) Scaling of connectivity in marine populations. *Science* 311(5760):522-527
27. Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Mol Ecol Resour* 10(3):556-557
28. Crow JF (2010) Wright and Fisher on inbreeding and random drift. *Genetics* 184(3):609-611
29. Cunningham KM, Canino MF, Spies IB, Hauser L (2009) Genetic isolation by distance and localized fjord population structure in Pacific cod (Gadus macrocephalus): limited effective dispersal in the northeastern Pacific Ocean. *Can J Fish Aquat Sci* 66(1):153-166
30. Da Silva R, Veneza I, Sampaio I, Araripe J, Schneider H, Gomes G (2015) High levels of genetic connectivity among populations of yellowtail snapper, *ocyurus chrysurus* (Lutjanidae-Perciformes), in the Western South Atlantic revealed through multilocus analysis. *Plos One* 10(3):e0122173
31. Devanesan DW (1943) A brief investigation into the causes of the fluctuations of the annual fishery of the oil sardine of Malabar, *Sardinella longiceps*, determination of its age and an account of the discovery of its eggs and spawning ground. *Madras Fish Bull* No. 28 (Report No. 1)1-24
32. Devaraj M, Martosubroto P (1997) Small pelagic resources and their fisheries in the Asia-Pacific Region. Proceedings of APFIC working party on Marine Fisheries. RAP Publishers, Thailand
33. DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J Fish Biol* 56(3):461-473
34. Dunlop ES, Baskett ML, Heino M, Dieckmann U (2009) Propensity of marine reserves to reduce the evolutionary effects of fishing in a migratory species. *Evol Appl* 2(3):371-393
35. Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359-361
36. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611-2620
37. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564-567
38. Fauvelot C, Borsa P (2011) Patterns of genetic isolation in a widely distributed pelagic fish, the narrow—barred Spanish mackerel (*Scomberomorus commerson*). *Biol J Linn Soc Lond* 104(4):886-902
39. Frankham R, Briscoe DA, Ballou JD (2002) Introduction to conservation genetics. Cambridge University Press, New York
40. Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *J Appl Ecol* 46(2):493-505
41. Froese R. Pauly D (2009) Fish Base. http://www.fishbase.org. Accessed 13 Jan 2013
42. Gonzalez EG, Zardoya R (2007) Isolation and characterization of polymorphic microsatellites for the sardine *Sardina pilchardus* (Clupeiformes: Clupeidae). *Mol Ecol Notes* 7(3):519-921

43. Goudet J (1999) PCAGEN vers 1.2.1. http://www.unil.ch/popgen/ softwares/pcagen.htm. Accessed 18 July 2013
44. Grant WA, Bowen BW (1998) Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J Hered* 89(5):415-426
45. Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2(4):618-620
46. Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* 163(4):1467-1482
47. Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* 59(8):1633-1638
48. Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) Structurama: Bayesian inference of population structure. *Evol Bioinform* 7:55-59
49. Hutchings JA (2000) Collapse and recovery of marine fishes. *Nature* 406(6798):882-885
50. Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genet* 6(1):13
51. Johannesson K, Smolarz K, Grahn M, Andre C (2011) The future of Baltic Sea populations: local extinction or evolutionary rescue? *Ambio* 40(2):179-190
52. Johnson JE, Welch DJ (2009) Marine fisheries management in a changing climate: a review of vulnerability and future options. *Rev Fish Sci* 18(1):106-124
53. Jost LO (2008) $G_{ST}$ and its relatives do not measure differentiation. *Mol Ecol* 17(18):4015-4026
54. Jouanno J, Sheinbaum J, Barnier B, Molines JM, Candela J (2012) Seasonal and interannual modulation of the eddy kinetic energy in the Caribbean Sea. *J Phys Oceanogr* 42(11):2041-2055
55. Kalinowski ST (2009) How well do evolutionary trees describe genetic relationships among populations & quest. *Heredity* 102(5):506-513
56. Karlsson S, Mork J (2005) Deviation from Hardy-Weinberg equilibrium, and temporal instability in allele frequencies at microsatellite loci in a local population of Atlantic cod. *ICES J Mar Sci* 62(8):1588-1596
57. Knutsen H, Jorde PE, Andre C, Stenseth NC (2003) Fine—scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Mol Ecol* 12(2):385-394
58. Krishnakumar PK, Bhat GS (2008) Seasonal and inter annual variations of oceanographic conditions off Mangalore coast (Karnataka, India) in the Malabar upwelling system during 1995-2004 and their influences on the pelagic fishery. *Fish Oceanogr* 17(1):45-60
59. Kuthalingam MDK (1960) Observations on the life history and feeding habits of the Indian sardine, *Sardinella longiceps* Cuv. & Val. *Treubia* 25(2):207-213
60. Larsen PF, Nielsen EE, Williams TD, Hemmer-Hansen J, Chipman JK *et al* (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Mol Ecol* 16(22):4674-4683
61. Larsen PF, Nielsen EE, Meier K, Olsvik PA, Hansen MM, Loeschcke V (2012) Differences in salinity tolerance and gene expression between two populations of Atlantic cod (*Gadus morhua*) in response to salinity stress. *Biochem Genet* 50(5-6):454-466
62. Larsson LC, Laikre L, Palm S, André C, Carvalho GR, Ryman N (2007) Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus. *Mol Ecol* 16(6):1135-1147
63. Lecomte F, Grant WS, Dodson JJ, Rodriguez-Sanchez R, Bowen BW (2004) Living with uncertainty: genetic imprints of climate shifts in East Pacific anchovy (*Engraulis mordax*) and sardine (*Sardinops sagax*). *Mol Ecol* 13(8):2169-2182
64. Madhupratap M, Nair KNV *et al* (2001). Arabian Sea oceanography and fisheries of the west coast of India. *Curr Sci* 81:355-361
65. Maes GE, Volckaert FA (2002) Clinal genetic variation and isolation by distance in the European eel *Anguilla anguilla* (L.). *Biol J Linn Soc Lond* 77(4):509-521
66. Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Hum Biol* 76(2):173-190

67. Marshall DJ, Morgan SG (2011) Ecological and evolutionary consequences of linked life-history stages in the sea. *Curr Biol* 21(18):R718-R725
68. Marshall DJ, Monro K, Bode M, Keough MJ, Swearer S (2010) Phenotype- environment mismatches reduce connectivity in the sea. *Ecol Lett* 13(1):128-140
69. Meirmans PG (2012) The trouble with isolation by distance. *Mol Ecol* 21(12):2839-2846
70. Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. *Mol Ecol* 24(13):3223-3231
71. Menezes MR (1994) Little genetic variation in the oil sardine, *Sardinella longiceps* Val., from the western coast of India. *Mar Freshw Res* 45(2):257-264
72. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES science gateway for inference of large phylogenetic trees. In: Gateway computing environments workshop (GCE 2010). Institute of Electrical and Electronics Engineers, New York, pp 115
73. Mohamed KS, Zacharia PU, Maheswarudu G, Sathianandan TV, Abdussamad EM *et al* (2014) Minimum Legal Size (MLS) of capture to avoid growth overfishing of commercially exploited fish and shellfish species of Kerala. *Mar Fish Inf Serv* 220:3-7
74. Mohandas NN (1997) Population genetic studies on the oil sardine (*Sardinella longiceps*). Dissertation, Cochin University of Science and Technology, Kerala, India
75. Nair RV (1952) Studies on the revival of the Indian oil sardine fishery. *Proc Indo-Pacific Fish Coun* 2:1-5
76. Natoli A, Birkun A, Aguilar A, Lopez A, Hoelzel AR (2005) Habitat structure and the dispersal of male and female bottlenose dolphins (*Tursiops truncatus*). *Proc R Soc Lond B* 272(1569):1217-1226
77. Nielsen EE, Hemmer-hansen JA, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Mol Ecol* 18(15):3128-3150
78. Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *Plos One* 7(7):e37558
79. Oomen RA, Hutchings JA (2015) Variation in spawning time promotes genetic variability in population responses to environmental change in a marine fish. *Conserv Physiol* 3(1):cov027
80. Oreilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F ST and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Mol Ecol* 13(7):1799-1814
81. Pereyra RT, Saillant E, Pruett CL, Rocha-Olivares A, Gold J (2004) Characterization of polymorphic microsatellites in the Pacific sardine *Sardinops sagax sagax* (Clupeidae). *Mol Ecol Notes* 4(4):739-741
82. Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting reductions in the effective size using allele frequencies. J Hered 90:502-503
83. Poulsen N, Nielsen EE, Schierup MH, Loeschcke V, Gronkjaer P (2006) Long—term stability and effective population size in North Sea and Baltic Sea cod (Gadus morhua). *Mol Ecol* 15(2):321-331
84. Prasanna Kumar S, Muraleedharan PM, Prasad TG, Gauns M, Ramaiah N *et al.* (2002) Why is the Bay of Bengal less productive during summer monsoon compared to the Arabian Sea? *Geophys Res Lett* 29(24)
85. Prasanna Kumar S, Nuncio M, Narvekar J, Kumar A, Sardesai S *et al.* (2004) Are eddies nature’s trigger to enhance biological productivity in the Bay of Bengal? *Geophys Res Lett* 31(7)
86. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945-959
87. Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol Evol* 4(22):4399-4428
88. Qasim SZ (1982) Oceanography of the northern Arabian Sea. *Deep Sea Res* 29:1041-1068
89. Rao DS, Ramamirtham CP, Murty AVS *et al* (1992) Oceanography of the Arabian Sea with particular reference to the southwest monsoon. *CMFRI Bull* 45:4-8

90. Raymond M, Rousset F (1995) GENEPOP Version 1.2: population genetics software for exat tests and ecumenicism. J Hered 86(3):248-249
91. Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43(1):223-225
92. Ruggeri P, Splendiani A, Bonanomi S, Arneri E, Cingolani N *et al* (2013) Searching for a stock structure *in Sardina pilchardus* from the Adriatic and Ionian seas using a microsatellite DNAbased approach. *Sci Mar* 77(4):565-574
93. Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Mol Ecol Notes* 6(3):600-602
94. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York
95. Santamaria L, Mendez PF (2012) Evolution in biodiversity policy- current gaps and future needs. *Evol Appl* 5(2):202-218
96. Selkoe KA, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Mar Ecol Prog Ser* 436:291-305
97. Shankar D, Vinayachandran PN, Unnikrishnan AS (2002) The monsoon currents in the north Indian Ocean. *Prog Oceanogr* 52(1):63-120
98. Shi W, Morrison JM, Bohm E, Manghnani V (2000) The Oman upwelling zone during 1993, 1994 and 1995. *Deep-Sea Res* II(47):1227-1247
99. Smedbol RK, McPherson A, Hansen MM, Kenchington E (2002) Myths and moderation in marine metapopulations? *Fish Fish* 3(1):20-35
100. Srivastava A, Dwivedi S, Mishra A (2015) High resolution numerical modeling of the Indian Ocean surface hydrography and circulation. *Discovery* 40(181):34-40
101. Sukumaran S, Gopalakrishnan A, Sebastian W, Vijayagopal P, Nandakumar Rao S *et al* (2016a) Morphological divergence in Indian oil sardine, *Sardinella longiceps* Valenciennes, 1847-Does it imply adaptive variation? *J Appl Ichthyol* 32:706-711
102. Sukumaran S, Sebastian W, Gopalakrishnan A (2016b) Population genetic structure of Indian oil sardine, *Sardinella longiceps* along Indian coast. *Gene* 576(1):372-378
103. Svedang H, Righton D, Jonsson P (2007) Migratory behaviour of Atlantic cod *Gadus morhua*: natal homing is the prime stockseparating mechanism. *Mar Ecol Prog Ser* 345:1-2
104. Talwar PK, Kacker RK (1984) Commercial Sea fishes of India. Zoological Survey of India, Kolkata
105. Teacher AG, Andre C, Jonsson PR, Merila J (2013) Oceanographic connectivity and environmental correlates of genetic structuring in Atlantic herring in the Baltic Sea. *Evol Appl* 6(3):549-567
106. Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4(3):535-538
107. Venkita Krishnan P (1993) Biochemical genetic studies on the oil sardine, *Sardinella longiceps* (Cuvier and Valenciennes, 1847) from selected centers of the west coast of India. Dissertation, Cochin University of Science and Technology, Kerala, India
108. Wang L, Liu S, Zhuang Z, Guo L, Meng Z, Lin H (2013) Population genetic studies revealed local adaptation in a high gene-flow marine fish, the small yellow croaker (*Larimichthys polyactis***)**. *Plos One* 8(12):e83493
109. Wheeler JP, Winters GH (1984) Homing of Atlantic herring (*Clupea harengus* harengus) in Newfoundland waters as indicated by tagging data. *Can J Fish Aquat Sci* 41(1):108-117
110. Whittaker JC, Harbord RM, Boxall N, Mackay I, Dawson G, Sibly RM (2003) Likelihood-based estimation of microsatellite mutation rates. *Genetics* 164(2):781-787
111. Wyrtki K (1973) Physical oceanography of the Indian Ocean. In: The biology of the Indian Ocean. Springer, Berlin, Heidelberg pp 18-36
112. Zarraonaindia I, Pardo MA, Iriondo M, Manzano C, Estonba A (2009) Microsatellite variability in European anchovy (*Engraulis encrasicolus*) calls for further investigation of its genetic structure and biogeography. *ICES J Mar Sci* 66(10):2176-2182

# Chapter 5

## GENOTYPING BY DOUBLE DIGESTED RESTRICTION SITE ASSOCIATED DNA SEQUENCING (ddRAD Seq) IN INDIAN OIL SARDINE, *SARDINELLA LONGICEPS* (Valenciennes, 1847) FOR POPULATION GENETIC STRUCTURE ANALYSIS, DEVELOPING SNPs AND MICROSATELLITE MARKERS

### ABSTRACT

Double digested restriction site-associated DNA sequencing is a powerful tool for generating genome-wide single nucleotide polymorphism (SNPs) markers for non-model organisms. It has been used in non-model organisms for elucidating fine-scale population structure and understanding patterns of selection. In this study, we performed population genomic analysis based on ddRAD data of *Sardinella longiceps* distributed in the Indian Ocean, for identifying population genetic structure and adaptive divergence in the backdrop of oceanic environmental heterogeneity. A total of 100 DNA samples with high quality and quantity were selected for ddRAD sequencing (20 samples each from Oman Sea, North East Arabian Sea (NAS), South-East Arabian Sea (SEAS), South West Bay of Bengal (SBOB) and North West Bay of Bengal (NBOB). The ddRAD libraries were prepared based on the previously published protocol and sequenced. Population genetic statistics (allele frequencies, percentage of polymorphic loci, nucleotide diversity, Wright's F-statistics $F_{IS}$ and $F_{ST}$) were computed using 'population' program in STACKS v 1.40. 48,076.00 polymorphic RAD loci, with 1SNP and 2 alleles were retained from the 100 samples sequenced, after de novo processing (without genome alignment). The average frequency of major alleles (P), ranged from 0.998-0.999 and average observed heterozygosity (Ob Het) ranged from 0.0017 to 0.0020. The overall nucleotide diversity ($\pi$) in *S. longiceps* populations ranged from 0.0015 to 0.0028 with samples from the Oman sea recording the lowest level of nucleotide diversity. The allele frequency spectrum of major alleles across the loci varies slightly across the population and was skewed towards 1.00. The pairwise comparison of genetic differentiation ($F_{ST}$ and $R_{ST}$) and STRUCTURE analysis found that the Oman Sea population was highly differentiated from all other populations, with very high significance. The second level of analysis, with PCA and Least-squares estimates of ancestry proportions, identified another level genetic differentiation between NEAS and other Indian ocean group SEAS, SBOB and NBOB.

Among the environmental factors analysed the minimum annual sea surface temperature, chlorophyll-*a* concentration and maximum dissolved oxygen concentration was found to be the predominant factor explaining genetic variation across Indian oil sardine population. The analyses also identified a set of candidate loci associated with sea surface temperature, chlorophyll-*a* concentration dissolved oxygen concentration. The loci identified as the candidate can be the representation of genomic regions of local adaptation and isolated genomic regions of divergence with gene flow in *S. longiceps*. Thus, the signals of cryptic structuring/local adaptation can be used as a starting point for more detailed study to identify the genomic region of genetic divergence in *S. longiceps* and Clupeoids. Reanalysis of the RADseq data with a reference genome-based method is necessary for identifying genome-wide distribution/chromosomal regions of genetic divergence.

## 1. INTRODUCTION

Indian Oil Sardine, *Sardinella longiceps* (Valenciennes, 1847) is one of the most commercially important fishes in Indian waters forming the largest pelagic fishery, with an annual production of 0.34 million tons (CMFRI Annual report 2018). It forms a cheap source of protein for millions and contributing to the majority of income from fishing due to its abundance (Devaraj and Martosubroto 1997). It also plays a significant role in trophic ecology and food web as a planktivorous, energy-rich small forage fish species which are consumed in large quantities by apex predators along with other sardines, mackerel and anchovy (Ganias *et al.* 2014). Remarkable fluctuations in abundance and distribution have been reported in Indian oil sardine with localized extinctions and recolonizations (Devaraj and Martosubroto 1997). The wide distribution of Indian oil sardines across tropical latitudes (range) makes them excellent models for investigations on adaptive evolution and divergence as wide variations in temperature, salinity, dissolved oxygen and chlorophyll-*a* have been reported across their range of distribution (Sebastian *et al.* 2020). Because of their economic and ecological importance, several investigations have been undertaken to understand their population dynamics and genetic structuring in the Indian Ocean region (Devaraj and Martosubroto 1997; Sukumaran *et al.* 2016a; Sukumaran *et al.* 2016b; Sebastian *et al.* 2017). Mitochondrial markers revealed a lack of genetic differentiation in Indian oil sardines whilst microsatellite markers detected significant genetic differentiation (Sukumaran *et al.* 2016b; Sebastian *et al.* 2017). Comparative mitogenomic investigations provided clues regarding the positive selection and possible locally adapted ecotypes (Sebastian *et al.* 2020). The phenotypic divergence has also been reported in Indian oil sardine pointing towards the possibility of adaptive divergence (Sukumaran *et al.* 2016a). Availability of nitrogen by upwelling and other mixing processes especially runoff from rivers (Checkley Jr *et al*. 2017; Reiss *et al*. 2008) also affect the productivity of oceanic habitats influencing the distribution and abundance of sardines.

Understanding the ecological pressures, its evolutionary impact on the natural population and geographic pattern of genetic variation in marine fishes is vital to conserving them and ensuring resilience to changing climate. Marine fishes are considered to be less diverged than those of freshwater fishes (Smedbol *et al*. 2002). The low degree of genetic differentiation in marine fishes was explained by the low ecological heterogeneity (compared to freshwater), lack of dispersal barriers, large effective population size and short population history after past glacial re-colonisation (Smedbol *et al*. 2002; Poulsen *et al*. 2006). Recent investigations disproved these findings as extensive genomic heterogeneity has been reported in marine fishes by recent investigations employing putative adaptive loci (Yoder *et al*. 2014; Vendrami *et al*. 2019). Population genomic approaches by scanning several parts of the genome of individuals of a population provide information regarding genomic islands of divergence despite gene flow. Loci under divergent selection are relatively protected from homogenizing effects of gene flow and consequently, genome scans provide a better understanding regarding adaptation signals (Narum *et al*. 2013). Partitioning of genetic variation within and among populations has a profound influence on species resilience and several methods and approaches have been employed to understand these patterns (Crandall 2000; van Tienderen *et al.* 2002). Traditionally, these studies were limited to few genetic markers (with insufficient genetic information) creating problems in interpreting the results (especially identifying the recent demographical events) and making conclusions (Cadrin *et al.* 2013; Rosenblum *et al.* 2007).

Next-Generation sequencing technologies like restriction site-associated DNA sequencing (RAD sequencing) have enabled sampling of large parts of the genome (also known as population genomics) even in non-model organisms (Cadrin *et al.* 2013; Hoffmann *et al.* 2015). Restriction site-associated DNA (RAD) sequencing is a method that sequence the DNA flanking the specific restriction enzyme sites in the genome (Davey and Blaxter 2010; Lowry *et al*. 2017). Using a population genomic approach with genotyping-by-sequencing (GBS), RAD sequencing enables sequencing the same genomic region across all the sampled individuals thus generating a reduced representation of the genome for detection of genome-wide nucleotide polymorphisms like single nucleotide polymorphisms (SNPs) (Peterson *et al*. 2012). The RAD seq approach has been applied in many non-model organisms to develop thousands of SNPs (Miller *et al*. 2007; Valencia 2018), linkage maps (Andrews *et al*. 2016) microsatellite markers (Zalapa *et al*. 2012), genome scans (Hohenlohe *et al*. 2010), detection of population differentiation (Emerson *et al*. 2010), and phylogeography (McCormack *et al*. 2012; Zellmer *et al*. 2012) using varied protocols (McCormack *et al*. 2012; Gompert *et al*. 2010; Hyten *et al*. 2010; Williams *et al*. 2010; Peterson *et al*. 2012). Genomic investigations on fishes like the three-spined stickleback *Gasterosteus aculeatus* provided insights regarding diversification of their populations into three life forms (marine, anadromous and freshwater) (Makinen *et al.* 2008a; Makinen *et al.* 2008b), whereas investigations on cichlids of African lakes (Kocher 2004; Genner and Turner 2005), provided information regarding their massive diversification that happened during the past 10 million years (Seehausen 2006; Takeda *et al.* 2013; Brawand *et al.* 2014). All these studies indicated the presence of adaptive diversity patterns not detected by neutral markers providing information regarding additional layers of diversity which needs conservation and management.

Microsatellites/simple sequence repeats (SSR) are widely used genetic markers in population and conservation genetics (Oliveira *et al*. 2006). Traditional methods to develop microsatellite markers include magnetic beads-based enrichment of a DNA library with targeted repeat motifs, followed by cloning and laboratory sequencing (Wang *et al*. 2009) or cross-species amplification method using existing markers from closely related species (Dawson *et al*. 2010; Gu *et al*. 2012). Even though advanced genomic markers like single nucleotide polymorphisms (SNP) have been used widely (Zalapa *et al*. 2012; Meglecz *et al*. 2014) microsatellites still have the potential to resolve fine-scale population structure, demographic events and pedigree patterns (Oliveira *et al*. 2006; Meglecz *et al*. 2014). Next-generation sequencing methods like RAD sequencing is capable of generating hundreds of loci at reduced cost and effort and recently these techniques are being used to develop microsatellite markers in non-model organisms (Zalapa *et al*. 2012).

Understanding genome-wide patterns of genetic diversity are very important in Indian oil sardine to devise conservation and management measures for this important species in Indian waters. Besides, knowledge regarding genomic patterns of divergence is crucial to understand and predict the response of Indian oil sardines to habitat variability and climate change in the Indian Ocean. Even though declines in Indian oil sardine landings have been reported in India, species-specific conservation and management measures are still not implemented except seasonal fishery closures and mesh size regulations applied to the whole fishery (Devaraj *et al*. 1997; Mohamed *et al*. 2014). We explore the spatial pattern of adaptive variation and genetic differentiation among the population with the application of genome-wide genetic markers (produced from ddRAD sequencing) of Indian oil sardine populations collected from Northern Arabian Sea, South-East Arabian

Sea and Bay of Bengal. The reduced representation genomic data was further analysed to detect candidate single nucleotide polymorphisms (SNPs) loci that may be indicative of local adaptation and loci associated with environmental gradient. Microsatellite/Simple sequence repeat (SSR) motifs were also identified in the ddRAD data of *S. longiceps*.

## 2. Materials And Methods

### 2.1. Sample collection, DNA extraction

Matured individuals of *S. longiceps* were collected from four eco-regions mainly, Oman sea (OMAN), North-East Arabian Sea (NEAS), South-East Arabian Sea (SEAS), South-West Bay of Bengal (SBOB) and North-West Bay of Bengal (NBOB) during 2016-2017. The muscle tissue samples were stored in 95% ethanol at -20$^0$c for genomic DNA extraction. Genomic DNA was extracted using DNAeasy blood and tissue kit (Qiagen). The DNA quality was visualized on a 0.8% agarose gel and quantified with NanoDrop™ One (Thermo Fisher Scientific) and Qubit® 3.0 Fluorometer (Thermo Fisher Scientific).

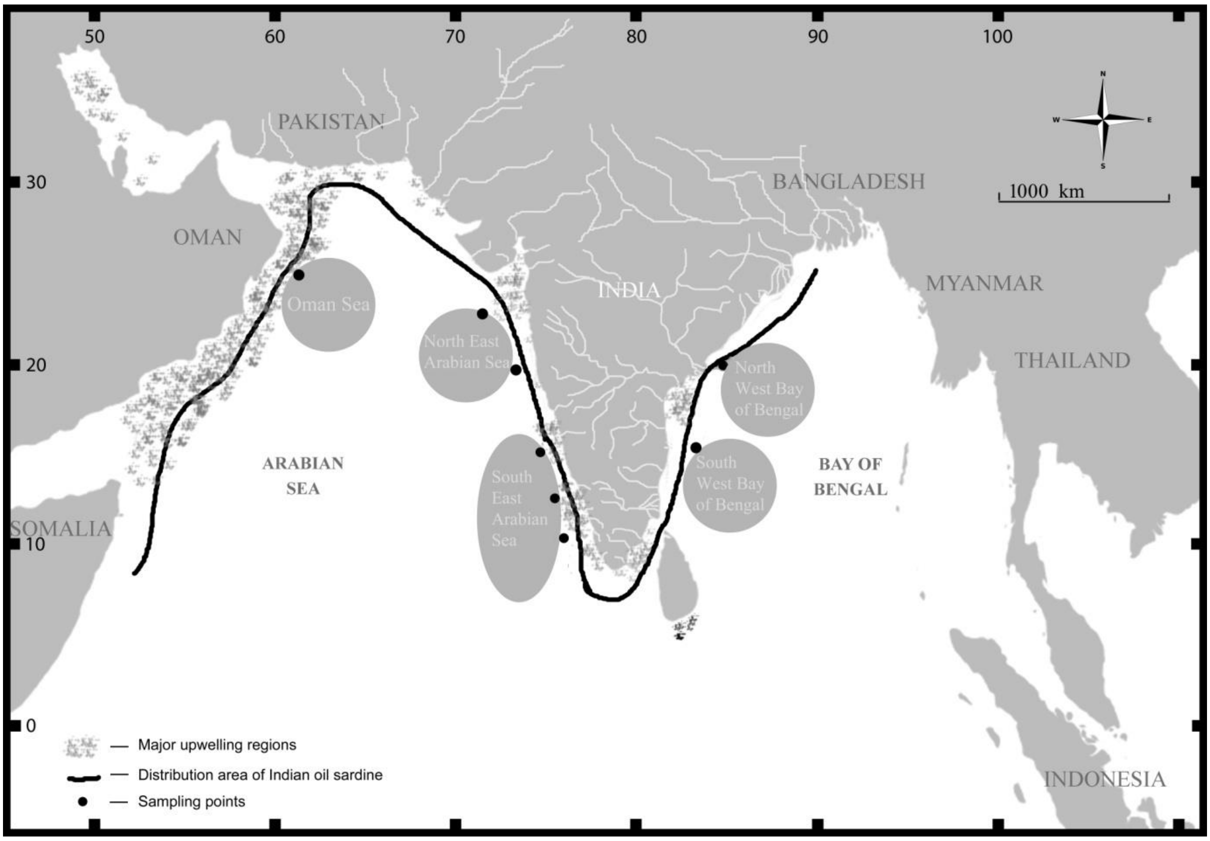


**Figure 5.1** Map showing sampling sites of *S. longiceps*. Oman sea (OMAN), North-East Arabian Sea (NEAS), South-East Arabian Sea (SEAS), South-West Bay of Bengal (SBOB) and North-West Bay of Bengal (NBOB).

## 2.2. ddRAD library construction sequencing and SNP genotyping

A total of 100 DNA samples with high quality and quantity were selected for ddRAD sequencing (20 samples each from OMAN, NEAS, SEAS, SBOB and NBOB). The ddRAD libraries were prepared based on the previously published protocol (Peterson 2012). Briefly, the DNA of each sample was double digested completely with high-fidelity *MspI* and *EcoRI* restriction enzymes (New England Biolabs). The barcode with a unique 5-bp sequence and P1 adapter was ligated to *EcoRI* overhang, separately to individual samples and P2 adapter was ligated to *MspI* overhang. The DNA fragments were selected on an automated size selection technology BluePippin (Sage Science) with a mean size of 300bp on 2% agarose cartridge. The fragments were then PCR amplified and purified with AMPure XP Beads. Libraries were prepared with approximately equal amounts of DNA from each sample. The barcoded ddRAD libraries were sequenced on an Illumina HiSeq 2500 platform with 100bp paired-end sequencing approach.

The raw reads were demultiplexed with a specific barcode index and filtered using 'process_radtags' program in STACKS v1.40 (Catchen *et al*. 2013b; Rochette and Catchen 2017). Reads with low quality (Phred score <20) and uncalled bases were discarded. Lengths of the sequence were trimmed to 85bp. SNPs identification and genotype calling was performed in STACKS using 'denovo_map.pl' program'. ustacks (-m,4) constructed stacks for each sample, cstacks (-M,5; -n,6) used all individual from each population to construct a catalogue of loci and sstacks compared each sample against the catalogue. Number of SNPs was used to determine the values of parameters -m, -M, -N of-of ustacks and -n of csstacks (Paris *et al*. 2017) obtained by evaluating the data with combinations of different values. The parameter -N was set as M+1. The number of SNPs was increased by increasing parameters (-M, -N and -n) until it reached a plateau. The number of SNPs and percentage of polymorphic loci reached a plateau at -m = ~4, -M =~5 and -n =~6.

## 2.3. Estimation of genetic diversity and genetic differentiation.

We used different methods to analyses genetic diversity and pattern of genetic structure within our dataset. Population genetic statistics (allele frequencies, percentage of polymorphic loci, nucleotide diversity, Wright's F-statistics $F_{IS,}$ sites in each population, percentage polymorphic sites and the average frequency of the major allele (P) at the sites.) were computed using the ‘population’ program in STACKS v1.40. (Catchen *et al*. 2012b). We used one random SNP per locus and 90% as the minimum number of populations a locus must be present in to process a locus. Deviations from Hardy-Weinberg equilibrium and the global estimate for genetic differentiation ($R_{ST}$)was assessed using GENEPOP v4.0 (Rousset 2008).

We performed a principal component analysis (PCA) of full data and with R package Adegenet version 2.1.2 (Jombart2008). A neighbour-joining method, clustering of the population as implemented in Neighbor (from Phylip programs) (Felsenstein 1989) was used to generate a phylogenetic tree using average pairwise $F_{ST}$ values as input. The tree was then visualized in FigTree (Andrew 2014).

A simple mantel test using $F_{ST}$ and special distance matric (shortest sea root between the sampling sites) was performed with zt (Bonnet and Van de Peer 2002). We also did a regression analysis with pairwise $F_{ST}/(1-F_{ST})$ and log of the pairwise spatial distance between populations (Rousset 1997).

2.4. Detection of SNP loci associated with environmental variables

We extracted the data for Chlorophyll-*a* concentration (Chl-*a*), Particulate organic carbon concentration (POC), Dissolved oxygen concentration (DO), Sea surface salinity (SSS) and Sea surface temperature (SST) for all the sampling locations. The annual minimum, maximum and mean of each of these environmental factors was then used to test which is the environmental variable best fit with genetic variation among the *S. longiceps* population samples, using the Gradient Forests R package (Ellis *et al*. 2012). In addition to the environmental variable, we also included the latitude and longitude information to measure the importance of geography.

Then we used BAYESCAN v2.1 (Foll and Gaggiotti2008) to identify loci under divergent selection, based on the differences in the allele frequencies between populations. The allele frequencies of each locus in each population/input file ware prepared by converting GENEPOP files to BAYESCAN input file using PGDSpider v2.1.1.5 (Lischer and Excoffier2012). The SNP loci with false discovery rate (FDR)

<0.05 were selected as outlier SNPs and all other optional parameters were set as default. The result was plotted on a graph with the R statistical package.

To assess the local adaptation, we tested the association between SNPs and climate gradients using the latent factor mixed model (LFMM) in LEA (Frichot and François 2015). This method analyses the SNP allele-environment correlation between each SNP and each environmental variable with correcting the background population structure. We calculated the individual admixture coefficients from the genotypic matrix using 'snmf' function, estimated the entropy criterion and chose the number of ancestral populations (K) best explained the genotypic data (Frichot *et al*. 2014). Five independent LFMM runs were conducted using 100000 iterations, burn-in of 10000 and calculated median Z-score (which is the strength of genetic-environmental association) for each locus. Adjusted p-values (q) were calculated using a false discovery rate (FDR) method and inspected the histogram of q as recommended in the LFMM manual (Frichot and Francois 2015). SNP loci with q <0.05 (or FDR < 0.05) were classified as candidate loci.

Each candidate/outlier SNP loci that contained putatively adaptive regions were subject to a BLASTx search of all sequences in the SwissProt, ref seq protein, NCBI non-redundant database e-value = 10. GO Annotator (http://xldb.di.fc.ul.pt/rebil/tools/goa/.) was used for assisting the GO annotation of loci that produced significant blast hits

### 2.5. SSR Identification

Simple sequence repeat (SSR)/microsatellite motifs were identified in the demultiplexed reads by using STR detection software (Fungtammasan *et al*. 2015), targeting di-, tri- and tetra motifs with minimum five perfect repeats. Primer pairs were designed from the flanking sequences of repeat motifs by using PRIMER 3 software (Rozen and Skaletsky 2000).

## 3. Result

### 3.1 Sequence quality and processing

From the 100 samples sequenced, 86 samples passed the minimum number of raw reads of <1000000.00. After de novo (without genome alignment) processing, 49,361.00 polymorphic RAD loci with 1SNP and 2alleles were retained. The number of polymorphic loci genotyped is 48,473.00 and among that 48,076.00 loci were genotyped by at least 50% of individuals (Table 5.1).

**Table 5.1** Summary genetic statistics for restriction-site associated DNA (RAD) sites of *S. longiceps*

| | |
|---|---|
| Nb loci genotyped | 56,358 |
| Nb loci genotyped by at least 50% of individuals | 53,680 |
| Nb polymorphic loci genotyped | 49,361 |
| Nb polymorphic loci genotyped by at least 50% of individuals | 48,473 |
| Nb polymorphic loci with 1 SNP and 2 alleles | 48,076 |

**Table 5.2** Summary of Genetic diversity statistics for restriction-site associated DNA (RAD) sites of *S. longiceps*.

| All positions (variant and fixed) | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pop ID | Private | Sites | Variant Sites | Polymorphic Sites | % Polymorphic Loci | Num Indv | Var | P | Obs Het | Obs Hom | Exp Het | Exp Hom | π | $F_{IS}$ |
| OMAN | 387 | 4393352 | 44876 | 25709 | 0.5852 | 12.6394 | 0.4642 | 0.9984 | 0.002 | 0.998 | 0.0022 | 0.9978 | 0.0020 | 0.0008 |
| NEAS | 88 | 4256641 | 27963 | 20907 | 0.4912 | 11.4331 | 2.1912 | 0.9991 | 0.0012 | 0.9988 | 0.0014 | 0.9986 | 0.0015 | 0.0007 |
| SEAS | 970 | 4691531 | 46269 | 44683 | 0.9524 | 23.1396 | 11.4238 | 0.9983 | 0.0018 | 0.9982 | 0.0025 | 0.9975 | 0.0026 | 0.0025 |
| SBOB | 544 | 4682365 | 45620 | 39288 | 0.8391 | 12.0171 | 3.9287 | 0.9984 | 0.0017 | 0.9983 | 0.0024 | 0.9976 | 0.0025 | 0.0022 |
| NBOB | 563 | 4721272 | 49088 | 42020 | 0.89 | 11.848 | 1.8672 | 0.9982 | 0.002 | 0.998 | 0.0026 | 0.9974 | 0.0028 | 0.0021 |
| Variant positions | | | | | | | | | | | | | | |
| OMAN | 387 | | | | | 12.3226 | 0.6209 | 0.8425 | 0.198 | 0.802 | 0.211 | 0.789 | 0.2398 | 0.0818 |
| NEAS | 88 | | | | | 11.4249 | 1.7164 | 0.8555 | 0.1821 | 0.8179 | 0.2093 | 0.7907 | 0.225 | 0.1139 |
| SEAS | 970 | | | | | 20.0284 | 13.4432 | 0.8255 | 0.1851 | 0.8149 | 0.2534 | 0.7466 | 0.2601 | 0.2583 |
| SBOB | 544 | | | | | 10.2993 | 4.8389 | 0.8308 | 0.1781 | 0.8219 | 0.2429 | 0.7571 | 0.256 | 0.2246 |
| NBOB | 563 | | | | | 10.015 | 2.4759 | 0.8224 | 0.1954 | 0.8046 | 0.2524 | 0.7476 | 0.2678 | 0.1985 |

The average number of individuals genotyped at each locus (Num Indv), the number of variable sites unique to each population (Private), the number of nucleotide sites across the data set (Sites), polymorphic sites across the data set (Polymorphic Sites), percentage of polymorphic loci (% poly), the average frequency of the major allele (P), the average observed heterozygosity per locus (Obs Het), the average nucleotide diversity (π), Average Wright's inbreeding coefficient (FIS), Variance (Var) and Standard Error (StdErr)

### 3.2. Genetic diversity

The average frequency of major alleles (P) ranged from 0.998-0.999 and average observed heterozygosity (Ob Het) ranged from 0.0017 to 0.0020. Whereas in the variable position (at least in one population of the data set) P-value decreased to 0.82 - 0.85 and

Ob Het increased to 0.178 - 0.198 (Table 5.2) for polymorphic positions. The population of Oman Sea and NAS showed a reduced level of genetic diversity when compared to other populations. The overall nucleotide diversity (π) in *S. longiceps* populations ranged from 0.0015 to 0.0028 and samples from the Oman sea population had a low level of nucleotide diversity.

The allele frequency spectrum of major alleles across the loci varied slightly across the population. The spectra of allele frequency were skewed towards 1.00. The allele frequency of the Arabian Sea and Bay of Bengal samples were more skewed towards 1.00 than that of Oman sea samples.

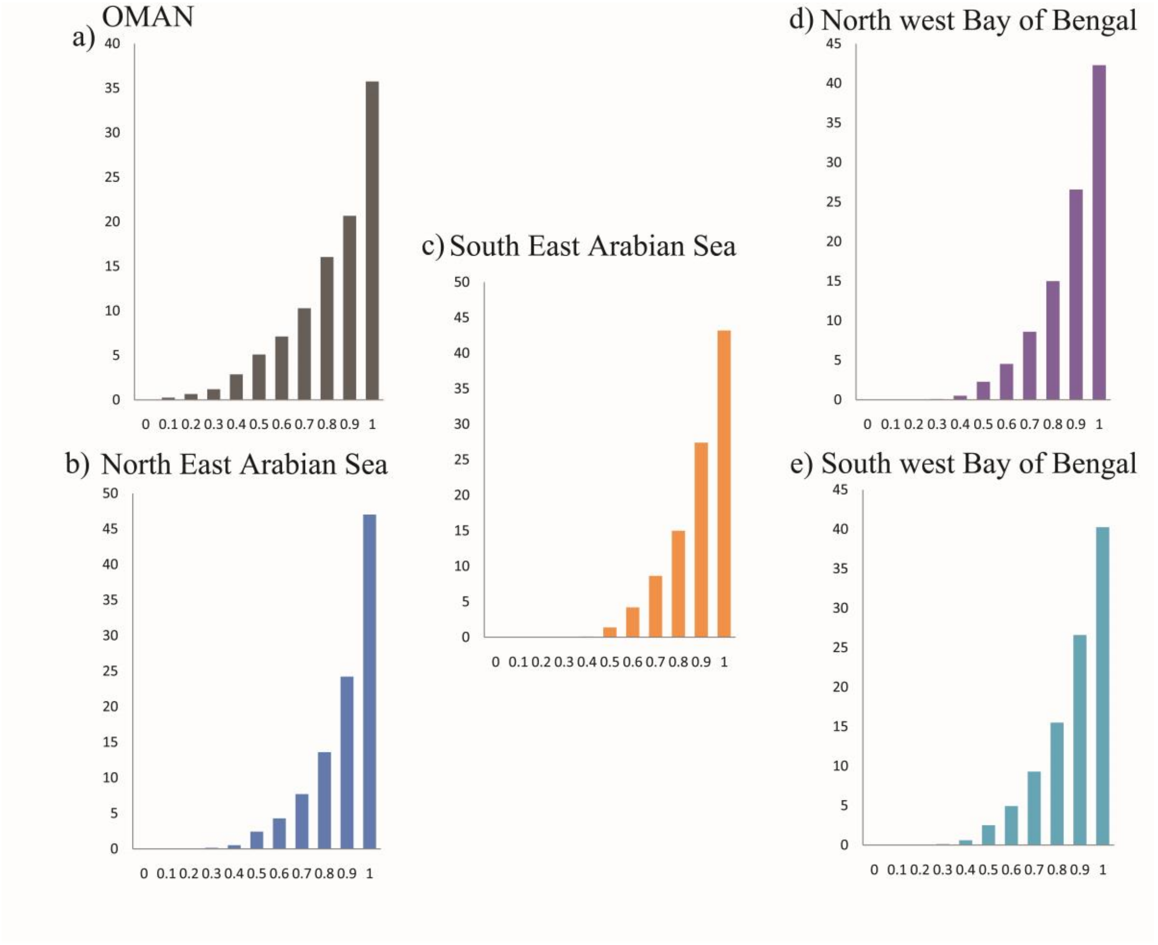


**Fig. 5.2** Allele frequency spectrum distribution for loci among *S. longiceps* populations. The X-axis represents allele frequencies and the Y-axis represents the number of alleles.

The positive average values of $F_{IS}$ did not indicate significant cryptic population structure or assertive mating. Within each population majority of loci had zero or nearly zero $F_{IS}$ value (Fig. 5.3) supporting the absence of cryptic population structure. However, the frequency distribution of $F_{IS}$ value across loci within each population indicated that the South-East Arabian Sea had a fraction of loci with $F_{IS}$ value > 0. When only the polymorphic loci are examined, the $F_{IS}$ value is increased remarkably in Arabian Sea and Bay of Bengal samples especially in Southeast Arabian Sea samples which indicate the possibility of cryptic population structure in the Southeast Arabian Sea. A small fraction of outlier loci with significant $F_{IS}$ is present in all populations, whereas the marked difference observed in the outlier loci of Southeast Arabian Sea also supports the above observations

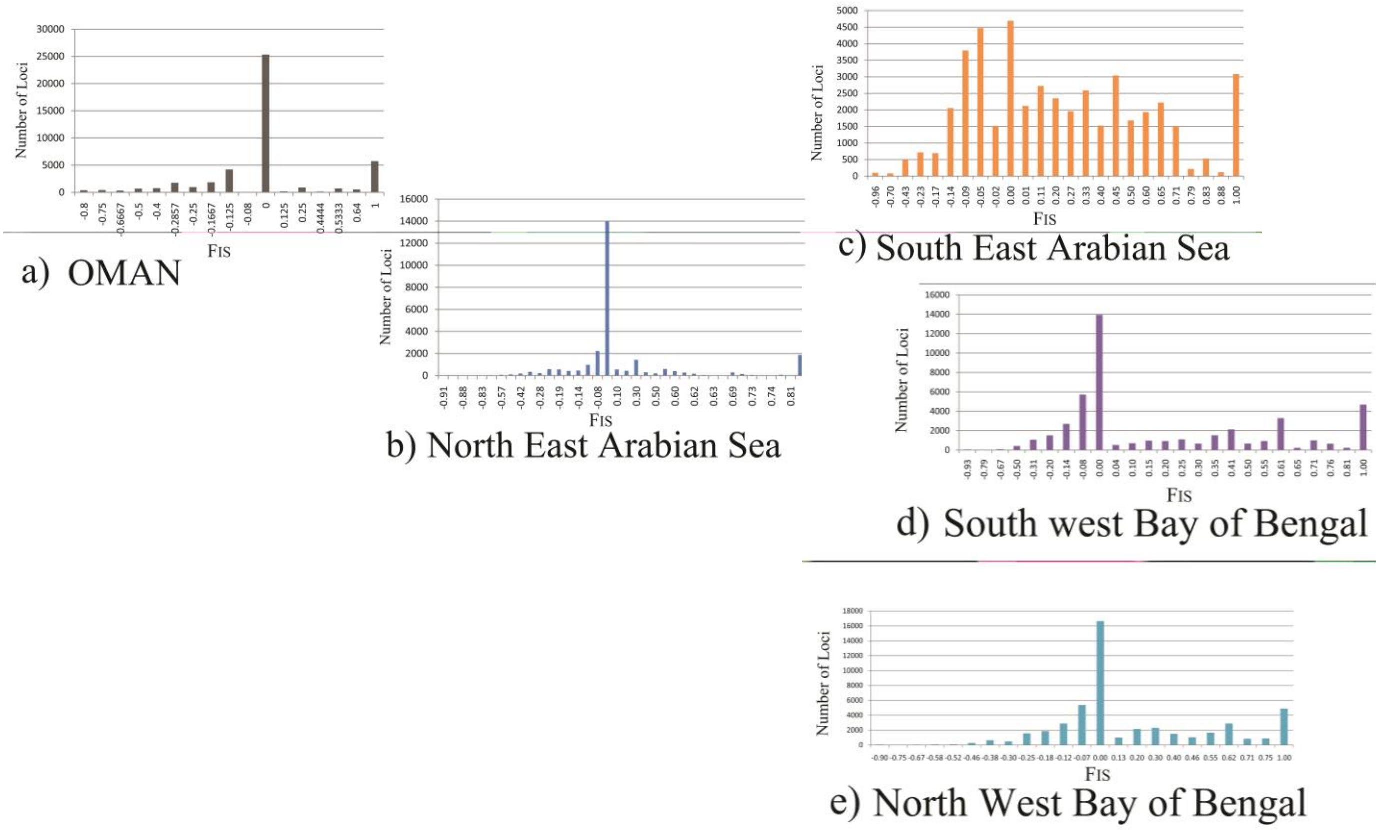


**Fig. 5.3** Frequency distribution of FIS across loci on S. longiceps population. The X-axis represents FIS and the Y-axis represents the number of loci.

3.3. Genetic differentiation

Comparison of pairwise genetic differentiation ($F_{ST}$ and $R_{ST}$) recorded Oman Sea population as highly differentiated from all other populations (Fst/Rst value of 0.0789/0.07632, 0.0657/0.06627, 0.06979/0.06958 and 0.06791/0.06791 for the four pairwise comparisons), with very high significance. In other pairwise comparisons, the highest genetic differentiation was between NEAS and NBOB followed by SEAS and NBOB population, but they are not significant (Table 5.3). While the PCA of full dataset showed that the OMAN samples are separated from other samples along PC1 while the NBOB separated from others along PC2. Individuals from NEAS, SEAS and the majority of individuals from SBOB formed a cluster.

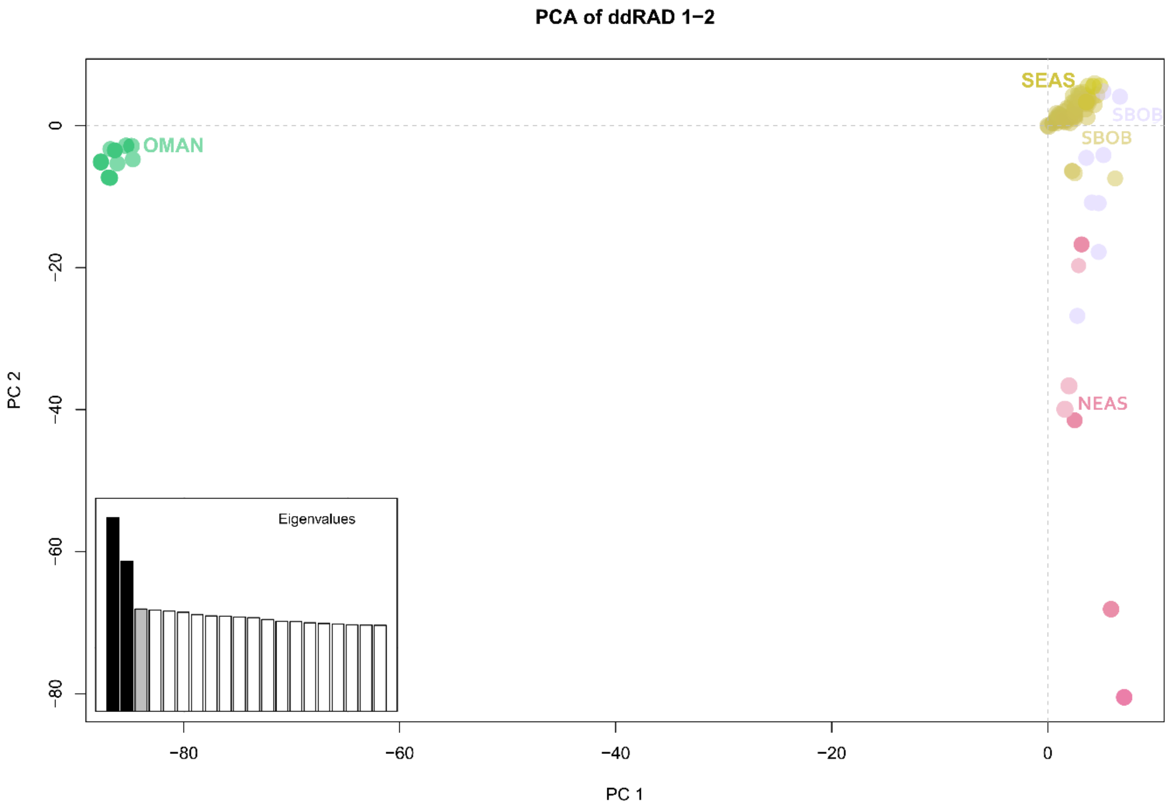


**Fig. 5.4.** Scatter plot showing individual variation in principal component (PC) scores derived from principal component analysis (PCA) of the *S. longiceps* RADseq data. Samples are colour-coded as described in the legend of Fig. 5.1.

**Table 5.3** Pairwise comparison of genetic distance ($F_{ST}$, $R_{ST}$) among *S. longiceps* populations. Below diagonal; genetic divergence among populations as measured by $F_{ST}$, $R_{ST}$. Above diagonal; *P*-value of exact G test for each population pair across all loci by Fisher's method.

| Population | OMAN | NEAS | SEAS | SBOB | NBOB |
|---|---|---|---|---|---|
| OMAN | 0 | Highly sign. | Highly sign. | Highly sign. | Highly sign. |
| NEAS | 0.07589, 0.07632 | 0 | Not sign. | Not sign. | Not sign. |
| SEAS | 0.0657, 0.06627 | 0.0009, 0.001 | 0 | Not sign. | Not sign. |
| SBOB | 0.06979, 0.06958 | 0.00074, 0.00063 | 0.0001, 0.0003 | 0 | Not sign. |
| NBOB | 0.06791, 0.06791 | 0.00159, 0.00132 | 0.00087, 0.00074 | 0.00094, 0.00047 | 0 |

Oman Sea_OMAN, North East Arabian Sea, NEAS, South East Arabian Sea_SEAS, South West Bay of Bengal_SBOB, North West Bay of Bengal_NBOB.

The cluster analysis on whole data set using Bayesian algorithm implemented in the program STRUCTURE indicates that the most likely number of distinct genetic entities is K = 2 (visual inspection reveals that setting K > 2 did not add any meaningful pattern) (Fig. 5.5). So, we chose K = 2, which probably represents the major structure of the population (Evanno *et al*. 2005) (The clusters represent Oman and Indian Ocean region respectively). The Delta K approach in structure analysis showed that K = 2 is the best fit model for the data and the plot of posterior probability clearly showed these two groups (Fig. 5.S2). The second level of analysis, omitting OMAN sea sample showed that the two combinations of alleles are present with varying degrees in each population, supporting the possibility of cryptic population structure by mixing of locally adapted populations.

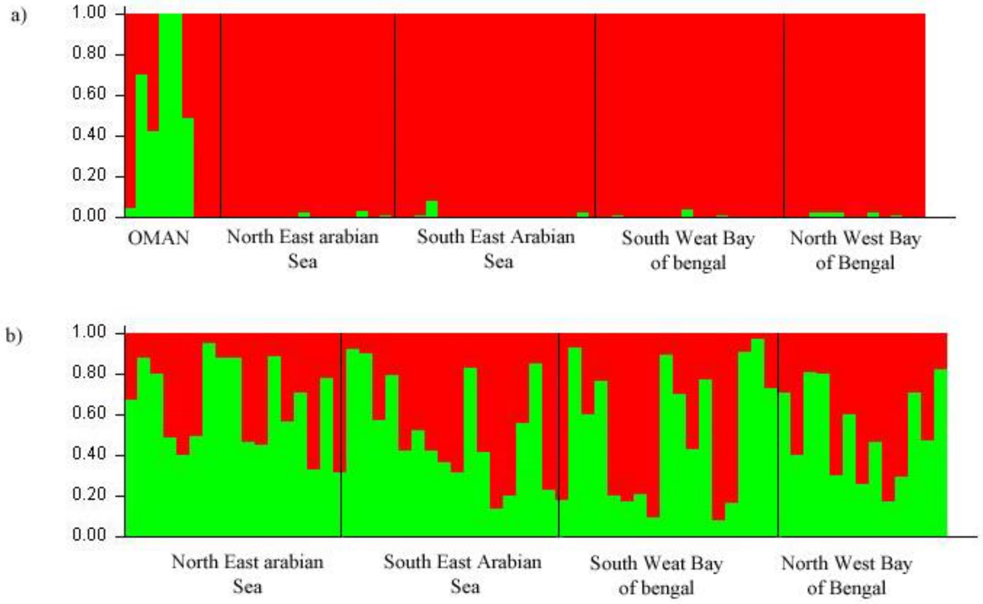


**Fig. 5.5** Graphical results of admixture analysis among all populations derived from 56,358.00 SNPs loci in Structure. *Vertical lines* represent the probability of individual membership in simulated clusters. a) Plot for K = 2 (including all the samples), b) Plot for K = 2 (excluding Oman samples).

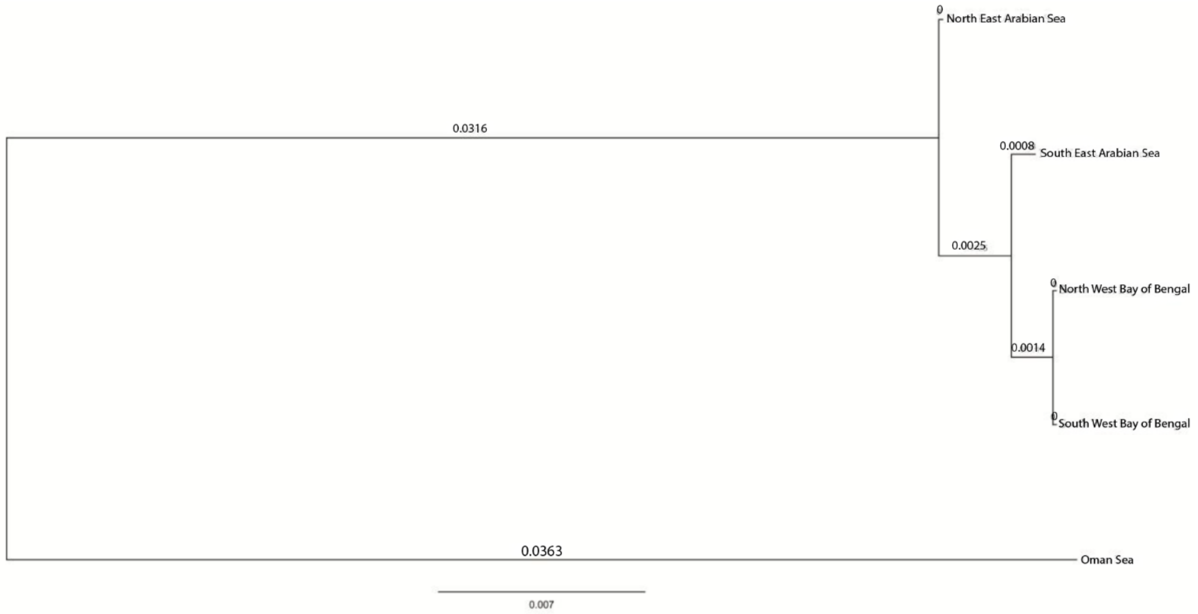


**Fig. 5.6** Nj tree of populations based on average $F_{ST}$ values of 56,358.00 SNPs loci. Populations from different geographical regions are represented as Oman Sea, North East Arabian Sea, South East Arabian Sea, South West Bay of Bengal and North West Bay of Bengal.

### 3.4. Loci associated with environmental variables

The gradient forest analysis showed that the environmental variable showing the greatest importance were related to minimum SST, minimum chlorophyll-*a* and maximum DO (Fig. 5.7a). The 'snmf' function in LEA indicated ancestral population, K = 3 was the best fit for the genotypic matrix used (Fig. 5.8). After combining the results from five independent runs, the histogram of adjusted p-values (q) confirmed correct distribution as recommended in LFMM manual (as expected, flat with a peak close to zero) (Frichot *et al*. 2015; Martins *et al*. 2018) (Fig. 5.S3). Significant association with environmental gradients were detected at 4371 loci (8.8%) by LFMM analysis and among that 3411 SNP loci were unique.38%, 36%,15%, 10% and 1% of them are associated with POC, Chlorophyll-*a*, SSS, DO and SST respectively (Fig. 5.7b). Rest of the 4371 loci 50% SNPs were associated with Chlorophyll-*a* and POC, 10 % with POC and SSS (Fig. 5.7c). Among the LFMM identified loci a total of 36 SNPs was also identified with FDR<0.005. All these loci have a positive alpha coefficient, indicating that these loci are under positive selection (Fig. 5.S6).

Among the 4371 adaptive loci, only 516 loci (11.8%) were showed significant similarity to the known genes in the public database and it has been characterised into 320 groups with molecular function, cellular component and biological process. The adaptive loci encode genes mostly involved in cellular energy metabolism, transcription, cell growth and signalling (Table S3). Most of the candidate/outlier SNP loci (88.2%) were not matched with known genes in the public database and it could be because most of the sequence reads were derived from non-coding regions of the genome.

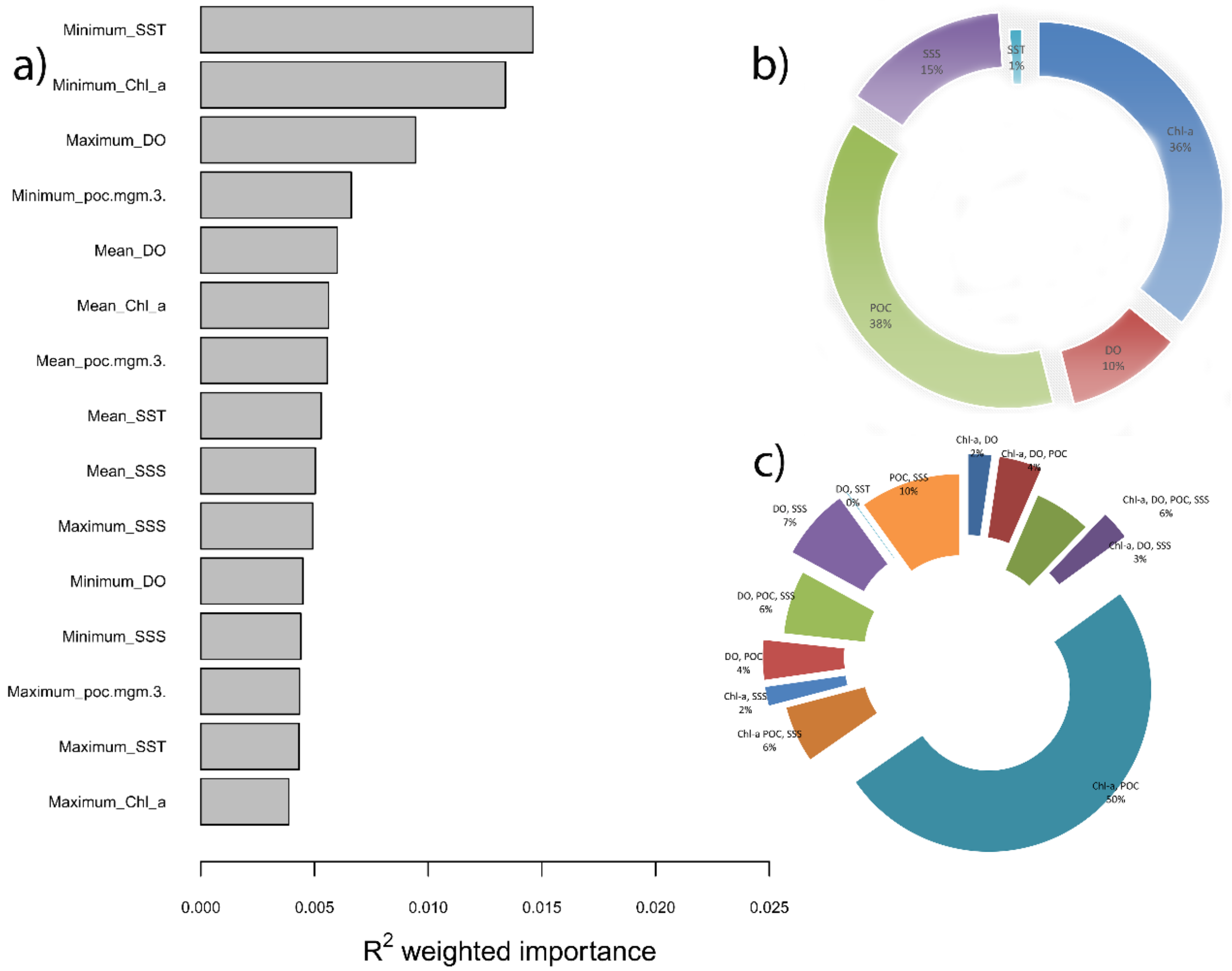


**Fig. 5.7** a) Plot shows the importance of each environmental variable in explaining genetic variation across the population as obtained from gradient forest analysis. b) the plot shows the percentage of loci identified by LFMM analyses of Chlorophyll-*a* concentration (Chl-*a*), Particulate organic carbon concentration (POC), Dissolved oxygen concentration (DO), Sea surface salinity (SSS) and Sea surface temperature (SST) and c) their overlaps.

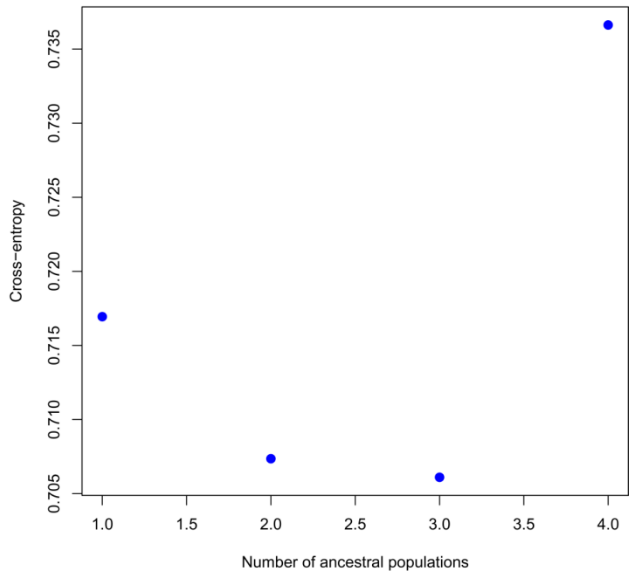


**Fig. 5.8** Least-squares estimates of ancestry proportions. Plot of the value of the cross-entropy criterion as a function of the number of populations in the R function ‘snmf’.

### 3.5. SSR Identification

We obtained ~290000 consensus sequences containing a microsatellite motif with dinucleotide being the most abundant repeat motif, followed by tetra and tri. The dominant dinucleotide motif was AC/CA followed by AG/GA and least common repeat was TC/CT (Fig. 5.S5). Among the trinucleotide most frequent motif were TTC, followed by TTG and AAG. Summary of polymorphic microsatellite loci developed from di, tri and tetra nucleotide motifs of restriction-site associated DNA from *S. longiceps* are given in Table 5.S4.

## 4. Discussion

Genome-wide SNPs generated from ddRAD sequencing data provided vital information regarding the genetic structure of Indian oil sardine across its range of distribution in the Indian ocean. Even though significant genetic differentiation was found between geographically distant populations (Oman sea and other Indian ocean samples) we observed the low overall degree of genetic differentiation among Sardine populations from Indian coastal line ($F_{ST}$ range from 0.07 to 0.0001). The highest difference was obtained between OMAN and other Indian Ocean samples ($F_{ST}$ 0.076 to 0.069), which was similar to our previous study using microsatellite markers (Sebastian *et al*. 2017). Unlike the previous study, we could not find any significant genetic differentiation in samples from Indian Coastline in $F_{ST}$ analysis ($F_{ST}$ range from 0.0015 to 0.0001) but a low genetic differentiation signal between NEAS and other Indian ocean group SEAS, SBOB and NBOB were found in both PCA and Least-squares estimates of ancestry proportions. Similarly, to our results, low but significant genetic differentiation has been reported in marine species like American lobster (Gleason and Burton 2016), Marine snail (Tine *et al*. 2014), European seabass (Xu *et al*. 2016) using ddRAD sequencing approach. High dispersal and genetic admixture have been observed in marine habitats mainly due to the absence of geographic barriers and oceanic currents. Thus, usually only mild genetic difference is observed in marine species over a wide geographical scale (Smedbol *et al*. 2002). The observed low genetic differentiation suggests a restricted amount of gene flow and admixture in Indian Oil sardines along the Indian coast. Among the environmental factors analysed the minimum annual sea surface temperature, chlorophyll-*a* concentration and maximum dissolved oxygen concentration was found to be the predominant factor explaining genetic variation across Indian oil sardine population. The analyses also identified a set of loci associated with sea surface temperature, chlorophyll-*a* concentration dissolved oxygen concentration and so on.

## 4.1 Genetic differentiation

We found significant genetic diversity in each of the sardine populations sampled (π range from 0.0015 to 0.0028). This pattern of genetic diversity is extended across all populations and it is comparable to those reported in other marine fishes (Catchen *et al*. 2013b). In the global distribution of $F_{IS}$, the majority of them were close to zero, whereas a small percentage of above values approaching one. This may indicate regions of the genome shaped by local adaptation and preventing introgressive hybridization (Seehausen *et al*. 2014; Wolf and Ellegren 2017). The $F_{IS}$ value of Southeast Arabian Sea samples varied remarkably from the global pattern, as the majority of them were negative (below zero). This indicates that individuals from this population are less related to each other as expected from random mating. Such a pattern can be generated from a cryptic population structure by the admixture of locally distributed/adapted population in this region (Turner and Hahn2010; Renaut *et al*. 2012; Strasburg *et al*. 2012). But on structure analysis based on Bayesian posterior probability of group assignment of individuals, no such pattern of cryptic diversity was recorded.

The average Expected Heterozygosity observed for the microsatellites was over four times higher for than for SNPs (Table 5.2). The genetic differentiation, quantified as pairwise $F_{ST}$, was similar when measured using microsatellites and SNPs. The $F_{ST}$ values between the Oman and Indian Ocean were highly significant and approximately at the same level in both analyses. STRUCTURE runs converged at $K$ = 2 in the study and previous study using microsatellites (Sebastian *et al*. 2017) and there was low admixture between populations (Fig. 5.4). Microsatellite data analysis, omitting Oman population also revealed sub-structuring with northwest Indian Ocean and rest of the Indian Ocean region samples with the presence of admixed individuals (Sebastian *et al*. 2017). Even though the $F_{ST}$ and STRUCTURE analysis of SNPs did not give any further sub-structuring both PCA and Least-squares estimates of ancestry proportions supported the sub-structuring with northwest Indian Ocean and rest of the Indian Ocean region. Overall, the results indicated that mitochondrial and polymorphic microsatellites and SNPs from RADseq agreed on estimates of population genetic structure of *S. longiceps*. But distinguishing moderately diverged populations from northwest Indian Ocean and rest of the Indian Ocean region samples, microsatellites outperformed mitochondrial DNA and RADseq. Contrary to this the outperformance of RADseq over microsatellite is reported for applications that quantifying relatedness and individual level heterozygosity (Thrasher *et al*. 2018). Comparisons between the results generated from traditional mitochondrial, SSRs and reduced representation libraries/ddRAD-seq methods will be very important for selection and development of the molecular marker in the future conservation genetics studies (Lemopoulos *et al*. 2019).

### 4.2 local adaptation

Local adaptation is expected in most of the species with restricted gene flow (Lessios *et al.* 1994; Hoskin 1997; Smedbol *et al*. 2002; Poulsen *et al*. 2006), but we can also see a growing number of studies reported role of selection in the genetic structuring of species like *S. longiceps* with highly dispersed larvae (Larsen *et al*. 2007; Johannesson *et al*. 2011; Wang *et al*. 2013; Brennan *et al*. 2016). The important factors that determine recruitment and fishery of Indian oil sardines are the intensity of upwelling (Devaraj and Martosubroto 1997), availability of diatoms *F. oceanica* (Nair 1952; Krishnakumar and Bhat 2008) intensity of rainfall (Murty and Edelman 1970), dissolved oxygen, temperature, migratory pattern and survival of the egg and larvae (Devaraj and Martosubroto 1997) and overfishing of immature fishes (Devanesan 1943). The significant number of outlier SNPs loci or adaptive loci (association with environmental gradients) identified from *S. longiceps* using the $F_{ST}$ outlier method and gradient forest analysis can be considered as candidate genes/genomic region playing important roles in local adaptation. Association of these candidate loci with environmental gradients (minimum annual SST and maximum annual DO) confirmed the predominant role of sea surface temperature, dissolved oxygen and chlorophyll-*a* concentration in genetic structuring of *S. longiceps*. The candidate loci associated with SST, DO and Chl-*a* may be a response to the pressure generated by these environmental factors on the growth and survival of *S. longiceps*. Thus, aggregation of spawners/locally adapted *S. longiceps* having candidate loci, at suitable local habitat may be occurring regularly. Fish spawning aggregations are reported in many fish species (Claro and Lindeman 2003; Gruss and Robinson 2015; Cherubin *et al.* 2020) Low sea surface temperature and high oxygen concentration are necessary for the survival of larvae (Dowling and Wiley 1986; Kujawa *et al.* 2015; Yamanaka *et al.* 2017; Nyanti *et al.* 2018; Sswat *et al.* 2018; Roman *et al.* 2019) and Oil sardine prefer the low-temperature season (monsoon period in the Indian ocean) for their spaning (Devaraj and Martosubroto 1997; Murty and Edelman 1970). Most of the candidate loci identified are found to be associated with cellular energy metabolism, transcription, cell growth and signalling. It has also been reported that many of the mitochondrial OXPHOS genes of *S. longiceps* are under positive selection and they are related to the heterogeneous oceanographic pattern in the Indian ocean (Sebastian *et al.* 2020). Positive selection in mitochondrial genes and the oceanographic characters in the Indian Ocean were described in the discussion section of chapter four.

In the present study annotation of many candidate loci did not show any similarity with coding genes of the published genome of fishes. Further re-analysis of the data with reference genome-based methods is necessary (when genome assemblies of *S. longiceps* become available) to identify/annotate the private alleles and outlier loci.

An important benefit of high-density marker loci like SNPs generated by RADseq is the possibility of locating the genomic regions with high population structuring which may restrict gene flow (Fan *et al*. 2012; Nosil and Feder 2012; Feder *et al*. 2012). The loci identified as outliers with $F_{IS}$> zero may be the representation of genomic regions of local adaptation, isolated genomic regions of divergence with gene flow and genomic regions of speciation (Turner and Hahn 2010; Renaut *et al*. 2012; Strasburg *et al*. 2012; Seehausen *et al*. 2014; Wolf and Ellegren 2017) in *S. longiceps*. Thus, the signals of cryptic structuring/assertive matting can be used as a starting point for more detailed study to identify the genomic region of genetic divergence in *S. longiceps* and Clupeoids. Reanalysis of the

RADseq data with a reference genome-based method is necessary for identifying genome-wide distribution/chromosomal regions of genetic divergence.

### 4.2 Microsatellite loci identification from ddRAD data

Microsatellites/simple sequence repeats (SSR) are a widely used genetic markers in population and conservation genetics (Oliveira *et al*. 2006). Traditionally they are isolated and developed using methods such as magnetic beads-based enrichment of a DNA library with targeted repeat motifs, followed by cloning and laboratory sequencing (Wang *et al*. 2009) or cross-species amplification method from closely related species (Dawson *et al*. 2010; Gu *et al*. 2012). Even though the sequence markers and single nucleotide polymorphism (SNP) has become popular (Zalapa *et al*. 2012; Meglecz *et al*. 2014) the microsatellites still have an advantage as a multiallelic marker with potential for resolving fine-scale population structure, demographic events and pedigree analysis (Oliveira *et al*. 2006; Meglecz *et al*. 2014). Now next-generation sequencing is considered as a suitable approach to developing microsatellite compared to traditional methods (Dubreuil *et al*. 2008; Yang 216 *et al*. 2009), because it can generate hundreds of loci at a reduced cost and effort, even in a non-model organism (Zalapa *et al*. 2012) and identification of polymorphic loci in vivo. RADseq is a cost-effective, simple and practical approach to creating reduced representation libraries for microsatellite development strategy in model/nonmodel organism. Here we developed a set of microsatellite markers using a RAD-seq and Illumina sequencing for multiple individuals from *S. longiceps*. The primers were designed for potentially amplifiable loci after polymorphism evaluation. This is the first study to develop microsatellite loci from *S. longiceps,* except the cross-species amplification by Sebastian *et al.* 2017. From the ~290000consensus sequences, we found ~89000 loci containing microsatellites, which included various types of simple sequence repeat motifs that were polymorphic among the ten individuals used in the development process. The results indicate that ddRAD technology is an efficient approach to isolate microsatellite markers from non-model organisms. Further validation and characterisation are needed to standardise the developed microsatellite markers. These novel polymorphic microsatellite loci will be very useful for genetic diversity and population structure studies and these results will provide important information for the conservation and management of this economically and ecologically important species.

The short-read lengths associated with the Illumina platform that we used is limited by the length of the flanking sequence available for optimal primer design, which can be overcome by generating longer sequences. Therefore, the longer paired-end reads of the Illumina Miseq sequencing, Nanopore and Pacbio platform may offer greater efficiency in mining microsatellite loci and primer pairs designing (Wei *et al*. 2014). Availability of reference genomes can provide long sequences upstream and downstream which also improve the efficiency of microsatellite loci identification and primer designing.

## Supplementary Tables and Figures

**Table 5.S1** Summary of zygosity of *S. longiceps* samples used for restriction-site associated DNA (RAD) sites analysis

| **Sample** | **Population id** | **Missing genotype** | **Heterozygote genotype** | **Homozygote genotype** | **Heterozygosity rate (%)** |
|---|---|---|---|---|---|
| OMAN1M | 1 | 8,263 | 8,395 | 39,700 | 17.46 |
| OMAN3M | 1 | 56,339 | 1 | 18 | 5.26 |
| OMAN4M | 1 | 7,769 | 8,399 | 40,190 | 17.29 |
| OMAN5M | 1 | 7,440 | 9,011 | 39,907 | 18.42 |
| OMAN5Ma | 1 | 7,794 | 8,606 | 39,958 | 17.72 |
| OMAN6M | 1 | 10,379 | 5,492 | 40,487 | 11.94 |
| BOM | 2 | 20,097 | 3,538 | 32,723 | 9.76 |
| MALV11 | 2 | 16,737 | 3,040 | 36,581 | 7.67 |
| MALV12 | 2 | 12,583 | 4,554 | 39,221 | 10.4 |
| MALV2 | 2 | 32,781 | 3,579 | 19,998 | 15.18 |
| MALV5 | 2 | 12,018 | 5,075 | 39,265 | 11.45 |
| MALV7 | 2 | 9,023 | 8,531 | 38,804 | 18.02 |
| MALV8 | 2 | 19,320 | 2,379 | 34,659 | 6.42 |
| MALV9 | 2 | 15,482 | 3,398 | 37,478 | 8.31 |
| MANG | 2 | 34,981 | 1,252 | 20,125 | 5.86 |
| MUM1Fa | 2 | 11,754 | 8,319 | 36,285 | 18.65 |
| MUM5Ma | 2 | 10,658 | 8,392 | 37,308 | 18.36 |
| COH1 | 3 | 15,812 | 3,176 | 37,370 | 7.83 |
| COH10 | 3 | 8,772 | 5,696 | 41,890 | 11.97 |
| COH1Ma | 3 | 8,988 | 5,329 | 42,041 | 11.25 |
| COH2 | 3 | 6,338 | 7,900 | 42,120 | 15.79 |
| COH2Fa | 3 | 8,774 | 5,724 | 41,860 | 12.03 |
| COH3 | 3 | 5,021 | 9,447 | 41,890 | 18.4 |
| COH3F | 3 | 36,304 | 339 | 19,715 | 1.69 |
| COH3Fa | 3 | 6,950 | 7,899 | 41,509 | 15.99 |
| COH3Ma | 3 | 6,557 | 7,918 | 41,883 | 15.9 |
| COH4 | 3 | 5,591 | 8,791 | 41,976 | 17.32 |
| COH5 | 3 | 6,279 | 7,599 | 42,480 | 15.17 |
| COH5Ma | 3 | 8,874 | 5,600 | 41,884 | 11.79 |
| COH6 | 3 | 6,314 | 7,696 | 42,348 | 15.38 |
| COH7 | 3 | 6,188 | 7,869 | 42,301 | 15.68 |
| COH8 | 3 | 7,210 | 6,768 | 42,380 | 13.77 |
| COH9 | 3 | 5,960 | 7,977 | 42,421 | 15.83 |
| KNR2 | 3 | 19,135 | 3,785 | 33,438 | 10.17 |
| KNR3 | 3 | 13,159 | 3,809 | 39,390 | 8.82 |
| KNR4 | 3 | 6,045 | 7,807 | 42,506 | 15.52 |
| KNR5 | 3 | 8,938 | 5,527 | 41,893 | 11.66 |
| KNR6 | 3 | 7,088 | 6,729 | 42,541 | 13.66 |
| KNR8 | 3 | 5,898 | 8,354 | 42,106 | 16.56 |
| KNR9 | 3 | 5,766 | 8,391 | 42,201 | 16.59 |
| VIZ1 | 3 | 7,231 | 6,855 | 42,272 | 13.95 |
| VIZ2 | 3 | 6,221 | 7,961 | 42,176 | 15.88 |
| VIZ3 | 3 | 5,361 | 9,567 | 41,430 | 18.76 |
| VIZ4 | 3 | 6,697 | 7,571 | 42,090 | 15.25 |
| MADPM1 | 4 | 8,793 | 5,508 | 42,057 | 11.58 |
| MADPM10 | 4 | 5,468 | 8,559 | 42,331 | 16.82 |
| MADPM12 | 4 | 8,954 | 7,670 | 39,734 | 16.18 |
| MADPM13 | 4 | 12,549 | 3,980 | 39,829 | 9.08 |
| MADPM2 | 4 | 9,016 | 7,790 | 39,552 | 16.45 |
| MADPM3 | 4 | 7,897 | 6,104 | 42,357 | 12.6 |
| MADPM4 | 4 | 15,479 | 2,896 | 37,983 | 7.08 |
| MADPM5 | 4 | 5,853 | 8,214 | 42,291 | 16.26 |
| MADPM6 | 4 | 5,622 | 9,155 | 41,581 | 18.04 |
| MADPM8 | 4 | 11,983 | 3,937 | 40,438 | 8.87 |
| MADPM9 | 4 | 13,243 | 3,526 | 39,589 | 8.18 |
| MAND1M | 4 | 7,645 | 6,298 | 42,415 | 12.93 |
| MAND2F | 4 | 6,568 | 7,857 | 41,933 | 15.78 |
| MAND2M | 4 | 6,236 | 8,498 | 41,624 | 16.95 |
| ODIS1 | 5 | 4,917 | 9,002 | 42,439 | 17.5 |
| ODIS11 | 5 | 7,416 | 6,644 | 42,298 | 13.58 |
| ODIS12 | 5 | 4,852 | 8,967 | 42,539 | 17.41 |
| ODIS13 | 5 | 5,846 | 7,807 | 42,705 | 15.46 |
| ODIS2 | 5 | 5,386 | 9,422 | 41,550 | 18.48 |
| ODIS3 | 5 | 7,259 | 6,489 | 42,610 | 13.22 |
| ODIS4 | 5 | 4,437 | 9,995 | 41,926 | 19.25 |
| ODIS5 | 5 | 5,361 | 8,317 | 42,680 | 16.31 |
| ODIS7 | 5 | 5,250 | 9,412 | 41,696 | 18.42 |
| ODIS8 | 5 | 5,000 | 9,658 | 41,700 | 18.81 |
| VSKP | 5 | 16,513 | 4,182 | 35,663 | 10.5 |

1_Oman Sea_OMAN, 2_North East Arabian Sea, NEAS, 3_South East Arabian Sea_SEAS, 4_South West Bay of Bengal_SBOB, 5_North West Bay of Bengal_NBOB.

**Table 5.S2.** Summary of GO Terms for adaptive loci of *S. longiceps* from the Indian Ocean.

| S NO | GO Terms | Aspect | |
|---|---|---|---|
| 1 | 1,4-alpha-glucan branching enzyme activity | F Molecular Function | enables |
| 2 | acetylcholine-gated cation-selective channel activity | F | enables |
| 3 | acetylgalactosaminyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase activity | F | enables |
| 4 | actin binding | F | enables |
| 5 | actin cytoskeleton organization | P Biological Process | involved_in |
| 6 | actin filament binding | F | enables |
| 7 | amino acid transport | P | involved_in |
| 8 | androgen receptor binding | F | enables |
| 9 | arterial endothelial cell differentiation | P | involved_in |
| 10 | Atg1/ULK1 kinase complex | C Cellular Component | part_of |
| 11 | ATP binding | F | enables |
| 12 | ATP transmembrane transporter activity | F | enables |
| 13 | ATP transport | P | involved_in |
| 14 | atrioventricular valve morphogenesis | P | involved_in |
| 15 | biological_process | P | acts_upstream_of_or_within |
| 16 | blood circulation | P | involved_in |
| 17 | blood vessel morphogenesis | P | involved_in |
| 18 | calcium ion binding | F | enables |
| 19 | carbohydrate metabolic process | P | involved_in |
| 20 | carbohydrate phosphorylation | P | involved_in |
| 21 | carboxylic acid metabolic process | P | involved_in |
| 22 | cardiolipin binding | F | enables |
| 23 | catalytic activity | F | enables |
| 24 | cation binding | F | enables |
| 25 | cation transmembrane transport | P | involved_in |
| 26 | cation transmembrane transporter activity | F | enables |
| 27 | cation transport | P | involved_in |
| 28 | cell adhesion | P | involved_in |
| 29 | cell differentiation | P | involved_in |
| 30 | cell junction | C | part_of |
| 31 | cell maturation | P | acts_upstream_of_or_within |
| 32 | cell projection | C | part_of |
| 33 | cell surface receptor signaling pathway | P | involved_in |
| 34 | cellular component organization | P | involved_in |
| 35 | cellular glucose homeostasis | P | involved_in |
| 36 | cellular iron ion homeostasis | P | involved_in |
| 37 | cellular metabolic process | P | involved_in |
| 38 | cellular protein modification process | P | involved_in |
| 39 | cellular response to DNA damage stimulus | P | involved_in |
| 40 | cellular response to drug | P | involved_in |
| 41 | cellular response to estrogen stimulus | P | involved_in |
| 42 | cellular response to progesterone stimulus | P | involved_in |
| 43 | central nervous system myelination | P | involved_in |
| 44 | centriole replication | P | involved_in |
| 45 | centriole-centriole cohesion | P | involved_in |
| 46 | chaperone-mediated protein complex assembly | P | involved_in |
| 47 | chromatin | C | part_of |
| 48 | chromatin binding | F | enables |
| 49 | cis-regulatory region sequence-specific DNA binding | F | enables |
| 50 | cohesin complex | C | part_of |
| 51 | collagen trimer | C | part_of |
| 52 | cone photoresponse recovery | P | involved_in |
| 53 | copper ion binding | F | enables |
| 54 | cysteine-type peptidase activity | F | enables |
| 55 | cytoplasm | C | part_of |
| 56 | cytoplasmic microtubule | C | part_of |
| 57 | cytoskeleton organization | P | involved_in |
| 58 | cytosol | C | part_of |
| 59 | Derlin-1 retrotranslocation complex | C | part_of |
| 60 | developmental growth | P | acts_upstream_of_or_within |
| 61 | digestive tract development | P | involved_in |
| 62 | digestive tract morphogenesis | P | acts_upstream_of_or_within |
| 63 | DNA binding | F | enables |
| 64 | DNA duplex unwinding | P | involved_in |
| 65 | DNA helicase activity | F | enables |
| 66 | DNA integration | P | involved_in |
| 67 | DNA metabolic process | P | involved_in |

| | | | |
|---|---|---|---|
| 68 | DNA recombination | P | involved_in |
| 69 | DNA repair | P | involved_in |
| 70 | DNA topoisomerase type II (double strand cut, ATP-hydrolyzing) activity | F | enables |
| 71 | DNA topological change | P | involved_in |
| 72 | DNA-binding transcription activator activity | F | enables |
| 73 | DNA-binding transcription activator activity, RNA polymerase II-specific | F | enables |
| 74 | DNA-binding transcription factor activity | F | enables |
| 75 | DNA-binding transcription factor activity, RNA polymerase II-specific | F | enables |
| 76 | dorsal aorta development | P | involved_in |
| 77 | dynein complex | C | part_of |
| 78 | early endosome | C | part_of |
| 79 | embryonic viscerocranium morphogenesis | P | involved_in |
| 80 | endoplasmic reticulum | C | part_of |
| 81 | endoplasmic reticulum membrane | C | part_of |
| 82 | endosome | C | part_of |
| 83 | enteric nervous system development | P | involved_in |
| 84 | ERAD pathway | P | acts_upstream_of_or_within |
| 85 | estrogen receptor binding | F | enables |
| 86 | excitatory postsynaptic potential | P | involved_in |
| 87 | extracellular ligand-gated ion channel activity | F | enables |
| 88 | extracellular matrix | C | part_of |
| 89 | extracellular matrix organization | P | involved_in |
| 90 | extracellular matrix structural constituent | F | enables |
| 91 | extracellular region | C | part_of |
| 92 | extracellular space | C | part_of |
| 93 | extrinsic component of mitochondrial inner membrane | C | part_of |
| 94 | extrinsic component of mitochondrial outer membrane | C | part_of |
| 95 | FAD binding | F | enables |
| 96 | fat cell differentiation | P | involved_in |
| 97 | ferroxidase activity | F | enables |
| 98 | flavin adenine dinucleotide binding | F | enables |
| 99 | G protein-coupled receptor activity | F | enables |
| 100 | G protein-coupled receptor kinase activity | F | enables |
| 101 | G protein-coupled receptor signaling pathway | P | involved_in |
| 102 | G1/S transition of mitotic cell cycle | P | involved_in |
| 103 | glucose binding | F | enables |
| 104 | glycogen biosynthetic process | P | involved_in |
| 105 | glycolytic process | P | involved_in |
| 106 | glycylpeptide N-tetradecanoyltransferase activity | F | enables |
| 107 | Golgi apparatus | C | part_of |
| 108 | GTP binding | F | enables |
| 109 | GTPase activator activity | F | enables |
| 110 | GTPase activity | F | enables |
| 111 | guanyl-nucleotide exchange factor activity | F | enables |
| 112 | guanyl-nucleotide exchange factor complex | C | part_of |
| 113 | heart development | P | involved_in |
| 114 | heart looping | P | involved_in |
| 115 | helicase activity | F | enables |
| 116 | hematopoietic progenitor cell differentiation | P | involved_in |
| 117 | heme transmembrane transporter activity | F | enables |
| 118 | heme transport | P | involved_in |
| 119 | hexokinase activity | F | enables |
| 120 | histone acetylation | P | involved_in |
| 121 | histone acetyltransferase activity | F | enables |
| 122 | histone acetyltransferase complex | C | part_of |
| 123 | histone methylation | P | involved_in |
| 124 | histone methyltransferase activity | F | enables |
| 125 | host cell nucleus | C | part_of |
| 126 | Hrd1p ubiquitin ligase ERAD-L complex | C | part_of |
| 127 | hyaluronic acid binding | F | enables |
| 128 | hydrolase activity | F | enables |
| 129 | hydrolase activity, hydrolyzing O-glycosyl compounds | F | enables |
| 130 | identical protein binding | F | enables |
| 131 | immunoglobulin production in mucosal tissue | P | involved_in |
| 132 | in utero embryonic development | P | acts_upstream_of_or_within |
| 133 | inositol phosphate biosynthetic process | P | involved_in |
| 134 | integral component of membrane | C | part_of |
| 135 | intestinal absorption | P | involved_in |
| 136 | inward rectifier potassium channel activity | F | enables |

| | | | |
|---|---|---|---|
| 137 | ion channel activity | F | enables |
| 138 | ion transmembrane transport | P | involved_in |
| 139 | ion transport | P | involved_in |
| 140 | iron ion transport | P | involved_in |
| 141 | isomerase activity | F | enables |
| 142 | kidney morphogenesis | P | involved_in |
| 143 | kinase activity | F | enables |
| 144 | kinetochore binding | F | enables |
| 145 | kinetochore microtubule | C | part_of |
| 146 | lacrimal gland development | P | acts_upstream_of_or_within |
| 147 | ligase activity | F | enables |
| 148 | liver development | P | involved_in |
| 149 | L-lactate dehydrogenase activity | F | enables |
| 150 | magnesium ion binding | F | enables |
| 151 | maintenance of blood-brain barrier | P | involved_in |
| 152 | melanocyte differentiation | P | acts_upstream_of_or_within |
| 153 | membrane | C | part_of |
| 154 | metal ion binding | F | enables |
| 155 | methylation | P | involved_in |
| 156 | methyltransferase activity | F | enables |
| 157 | microtubule motor activity | F | enables |
| 158 | microtubule plus-end binding | F | enables |
| 159 | microtubule-based movement | P | involved_in |
| 160 | microtubule-based process | P | involved_in |
| 161 | mitochondrial inner membrane | C | part_of |
| 162 | mitochondrial outer membrane | C | part_of |
| 163 | mitochondrion | C | part_of |
| 164 | mitotic sister chromatid cohesion | P | involved_in |
| 165 | monooxygenase activity | F | enables |
| 166 | morphogenesis of a branching epithelium | P | acts_upstream_of_or_within |
| 167 | morphogenesis of an epithelium | P | involved_in |
| 168 | motor activity | F | enables |
| 169 | mRNA binding | F | enables |
| 170 | multicellular organism development | P | involved_in |
| 171 | myosin complex | C | part_of |
| 172 | negative regulation of apoptotic process | P | involved_in |
| 173 | negative regulation of autophagosome assembly | P | involved_in |
| 174 | negative regulation of canonical Wnt signaling pathway | P | involved_in |
| 175 | negative regulation of gene expression | P | involved_in |
| 176 | negative regulation of macroautophagy | P | involved_in |
| 177 | negative regulation of neurogenesis | P | involved_in |
| 178 | negative regulation of NF-kappaB transcription factor activity | P | involved_in |
| 179 | negative regulation of protein kinase activity | P | involved_in |
| 180 | negative regulation of Schwann cell proliferation | P | acts_upstream_of_or_within |
| 181 | negative regulation of transcription by RNA polymerase II | P | involved_in |
| 182 | negative regulation of transcription, DNA-templated | P | involved_in |
| 183 | neural crest cell migration | P | involved_in |
| 184 | neural plate development | P | involved_in |
| 185 | neuronal stem cell population maintenance | P | involved_in |
| 186 | Notch signaling pathway | P | involved_in |
| 187 | nucleic acid binding | F | enables |
| 188 | nucleoplasm | C | part_of |
| 189 | nucleotide binding | F | enables |
| 190 | nucleus | C | part_of |
| 191 | oligodendrocyte development | P | involved_in |
| 192 | oligodendrocyte differentiation | P | involved_in |
| 193 | oxidation-reduction process | P | involved_in |
| 194 | oxidoreductase activity | F | enables |
| 195 | oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen | F | enables |
| 196 | oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen | F | enables |
| 197 | oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor | F | enables |
| 198 | peptidase activity | F | enables |
| 199 | peptidyl-arginine methylation | P | involved_in |
| 200 | peptidyl-arginine N-methylation | P | involved_in |
| 201 | peripheral nervous system development | P | involved_in |
| 202 | peripheral nervous system neuron axonogenesis | P | involved_in |

| | | | |
|---|---|---|---|
| 203 | peroxisomal importomer complex | C | part_of |
| 204 | peroxisomal membrane | C | part_of |
| 205 | peroxisome | C | part_of |
| 206 | peroxisome proliferator activated receptor binding | F | enables |
| 207 | pharyngeal system development | P | involved_in |
| 208 | phosphatidylinositol-3,4,5-trisphosphate binding | F | enables |
| 209 | phosphorylation | P | involved_in |
| 210 | phosphotransferase activity, alcohol group as acceptor | F | enables |
| 211 | photoreceptor activity | F | enables |
| 212 | photoreceptor cell maintenance | P | involved_in |
| 213 | photoreceptor inner segment | C | part_of |
| 214 | phototransduction, visible light | P | involved_in |
| 215 | plasma membrane | C | part_of |
| 216 | plasma membrane organization | P | involved_in |
| 217 | positive regulation of apoptotic process | P | involved_in |
| 218 | positive regulation of autophagosome maturation | P | involved_in |
| 219 | positive regulation of cysteine-type endopeptidase activity involved in apoptotic process | P | involved_in |
| 220 | positive regulation of gene expression | P | acts_upstream_of_or_within |
| 221 | positive regulation of gliogenesis | P | involved_in |
| 222 | positive regulation of GTPase activity | P | involved_in |
| 223 | positive regulation of myelination | P | acts_upstream_of_or_within |
| 224 | positive regulation of neuroblast proliferation | P | acts_upstream_of_or_within |
| 225 | positive regulation of nucleic acid-templated transcription | P | involved_in |
| 226 | positive regulation of TOR signaling | P | involved_in |
| 227 | positive regulation of transcription by RNA polymerase II | P | involved_in |
| 228 | positive regulation of transcription, DNA-templated | P | involved_in |
| 229 | postsynaptic membrane | C | part_of |
| 230 | potassium ion import across plasma membrane | P | involved_in |
| 231 | potassium ion transmembrane transport | P | involved_in |
| 232 | potassium ion transport | P | involved_in |
| 233 | potassium transmembrane transporter activity, phosphorylative mechanism | F | enables |
| 234 | potassium:proton exchanging ATPase activity | F | enables |
| 235 | progesterone receptor binding | F | enables |
| 236 | promoter-specific chromatin binding | F | enables |
| 237 | pronephros development | P | involved_in |
| 238 | protein binding | F | enables |
| 239 | protein deubiquitination | P | involved_in |
| 240 | protein folding | P | involved_in |
| 241 | protein homodimerization activity | F | enables |
| 242 | protein import into peroxisome matrix, docking | P | involved_in |
| 243 | protein kinase activity | F | enables |
| 244 | protein kinase binding | F | enables |
| 245 | protein kinase inhibitor activity | F | enables |
| 246 | protein phosphorylation | P | involved_in |
| 247 | protein secretion | P | involved_in |
| 248 | protein serine/threonine kinase activity | F | enables |
| 249 | protein transport | P | involved_in |
| 250 | protein ubiquitination | P | involved_in |
| 251 | protein-arginine N-methyltransferase activity | F | enables |
| 252 | proteolysis | P | involved_in |
| 253 | proton transmembrane transport | P | involved_in |
| 254 | Rab guanyl-nucleotide exchange factor activity | F | contributes_to |
| 255 | regulation of androgen receptor signaling pathway | P | involved_in |
| 256 | regulation of autophagy | P | involved_in |
| 257 | regulation of blood vessel diameter | P | involved_in |
| 258 | regulation of cell cycle | P | involved_in |
| 259 | regulation of cell morphogenesis | P | involved_in |
| 260 | regulation of developmental process | P | involved_in |
| 261 | regulation of ion transmembrane transport | P | involved_in |
| 262 | regulation of postsynaptic membrane potential | P | involved_in |
| 263 | regulation of receptor-mediated endocytosis | P | involved_in |
| 264 | regulation of TORC1 signaling | P | involved_in |
| 265 | regulation of transcription by RNA polymerase II | P | involved_in |
| 266 | regulation of transcription, DNA-templated | P | involved_in |
| 267 | regulation of vascular permeability | P | involved_in |
| 268 | response to bacterium | P | involved_in |
| 269 | response to endoplasmic reticulum stress | P | acts_upstream_of_or_within |
| 270 | retinoic acid receptor binding | F | enables |
| 271 | retrograde protein transport, ER to cytosol | P | acts_upstream_of_or_within |

| | | | |
|---|---|---|---|
| 272 | Rho GTPase binding | F | enables |
| 273 | rhodopsin kinase activity | F | enables |
| 274 | rhombomere boundary formation | P | involved_in |
| 275 | ribosome | C | part_of |
| 276 | RNA binding | F | enables |
| 277 | RNA polymerase II cis-regulatory region sequence-specific DNA binding | F | enables |
| 278 | RNA-dependent DNA biosynthetic process | P | involved_in |
| 279 | RNA-directed DNA polymerase activity | F | enables |
| 280 | sclerotome development | P | involved_in |
| 281 | semaphorin receptor binding | F | enables |
| 282 | sequence-specific DNA binding | F | enables |
| 283 | signal transduction | P | involved_in |
| 284 | signaling receptor activity | F | enables |
| 285 | signaling receptor binding | F | enables |
| 286 | sister chromatid cohesion | P | involved_in |
| 287 | somatic stem cell population maintenance | P | involved_in |
| 288 | spliceosomal complex assembly | P | involved_in |
| 289 | stem cell differentiation | P | involved_in |
| 290 | structural constituent of ribosome | F | enables |
| 291 | symmetric cell division | P | involved_in |
| 292 | synapse | C | part_of |
| 293 | telomere maintenance | P | involved_in |
| 294 | thiol-dependent ubiquitinyl hydrolase activity | F | enables |
| 295 | thyroid hormone receptor binding | F | enables |
| 296 | tissue morphogenesis | P | involved_in |
| 297 | transcription coactivator activity | F | enables |
| 298 | transcription coregulator activity | F | enables |
| 299 | transcription elongation from RNA polymerase II promoter | P | acts_upstream_of_or_within |
| 300 | transcription factor binding | F | enables |
| 301 | transcription regulatory region sequence-specific DNA binding | F | enables |
| 302 | transferase activity | F | enables |
| 303 | transferase activity, transferring acyl groups | F | enables |
| 304 | transferase activity, transferring glycosyl groups | F | enables |
| 305 | translation | P | involved_in |
| 306 | transmembrane signaling receptor activity | F | enables |
| 307 | transmembrane transport | P | involved_in |
| 308 | transmembrane transporter activity | F | enables |
| 309 | triglyceride metabolic process | P | involved_in |
| 310 | ubiquinone biosynthetic process | P | involved_in |
| 311 | ubiquitin-dependent ERAD pathway | P | involved_in |
| 312 | ubiquitin-dependent protein catabolic process | P | involved_in |
| 313 | ubiquitin-protein transferase activity | F | enables |
| 314 | unfolded protein binding | F | enables |
| 315 | ventral spinal cord interneuron differentiation | P | involved_in |
| 316 | ventriculo bulbo valve morphogenesis | P | involved_in |
| 317 | visual perception | P | involved_in |
| 318 | voltage-gated ion channel activity | F | enables |
| 319 | zinc ion binding | F | enables |

**Table 5.S3.** Summary of the species distribution of Blast hits from BLASTx analysis of adaptive loci of *S. longiceps* from the Indian Ocean.

| S N0 | Species Name |
|---|---|
| 1 | *Acanthochromis polyacanthus* |
| 2 | *Acinetobacter baumannii* |
| 3 | *Acropora digitifera* |
| 4 | *Ailuropoda melanoleuca* |
| 5 | *Amazona aestiva* |
| 6 | *Amphiamblys sp. WSBS2006* |
| 7 | *Amphiprion ocellaris* |
| 8 | *Anas platyrhynchos* |
| 9 | *Anopheles darlingi* |
| 10 | *Aotus nancymaae* |

| | |
|---|---|
| 11 | *Apostichopus japonicus* |
| 12 | *Aptenodytes forsteri* |
| 13 | *Aquila chrysaetos canadensis* |
| 14 | *Astyanax mexicanus* |
| 15 | *Austrofundulus limnaeus* |
| 16 | *Beggiatoa sp. 4572_84* |
| 17 | *Beggiatoa sp. PS* |
| 18 | *Bemisia tabaci* |
| 19 | *Boleophthalmus pectinirostris* |
| 20 | *Bos indicus* |
| 21 | *Bubalus bubalis* |
| 22 | *Buceros rhinoceros silvestris* |
| 23 | *Calidris pugnax* |
| 24 | *Callipepla squamata* |
| 25 | *Callorhinchus milii* |
| 26 | *Calypte anna* |
| 27 | *Camelus ferus* |
| 28 | *Candidatus Entotheonella sp. TSY2* |
| 29 | *Canis lupus familiaris* |
| 30 | *Cathartes aura* |
| 31 | *Cebus capucinus imitator* |
| 32 | *Cervus elaphus hippelaphus* |
| 33 | *Channa striata* |
| 34 | *Chelonia mydas* |
| 35 | *Chenopodium quinoa* |
| 36 | *Chinchilla lanigera* |
| 37 | *Chrysemys picta bellii* |
| 38 | *Ciona intestinalis* |
| 39 | *Clupea harengus* |
| 40 | *Columba livia* |
| 41 | *Corvus brachyrhynchos* |
| 42 | *Crassostrea gigas* |
| 43 | *Crassostrea virginica* |
| 44 | *Cricetulus griseus* |
| 45 | *Crocodylus porosus* |
| 46 | *Cuculus canorus* |
| 47 | *Cynoglossus semilaevis* |
| 48 | *Cyprinodon variegatus* |
| 49 | *Cyprinus carpio* |
| 50 | *Daboia russelii* |
| 51 | *Danio rerio* |
| 52 | *Dasypus novemcinctus* |
| 53 | *Dendroctonus ponderosae* |
| 54 | *Dicentrarchus labrax* |
| 55 | *Echinops telfairi* |
| 56 | *Emys marmorata pallida* |
| 57 | *Enhydra lutris kenyoni* |
| 58 | *Eptesicus fuscus* |
| 59 | *Equus caballus* |
| 60 | *Erinaceus europaeus* |
| 61 | *Esox lucius* |
| 62 | *Exaiptasia pallida* |
| 63 | *Folsomia candida* |
| 64 | *Fukomys damarensis* |
| 65 | *Fundulus heteroclitus* |
| 66 | *Galendromus occidentalis* |
| 67 | *Gallus gallus* |
| 68 | *Gavialis gangeticus* |
| 69 | *Gekko japonicus* |
| 70 | *Halyomorpha halys* |
| 71 | *Haplochromis burtoni* |
| 72 | *Heterocephalus glaber* |
| 73 | *Heteropneustes fossilis* |
| 74 | *Hippocampus comes* |
| 75 | *Homo sapiens* |
| 76 | *Horstia sp. AD1229* |
| 77 | *Hydra vulgaris* |
| 78 | *Hypophthalmichthys nobilis* |
| 79 | *Ictalurus punctatus* |
| 80 | *Ictidomys tridecemlineatus* |
| 81 | *Ixodes scapularis* |

| | |
|---|---|
| 82 | *Kryptolebias marmoratus* |
| 83 | *Labrus bergylta* |
| 84 | *Lachancea mirantina* |
| 85 | *Larimichthys crocea* |
| 86 | *Lasius niger* |
| 87 | *Lates calcarifer* |
| 88 | *Latimeria chalumnae* |
| 89 | *Lepidothrix coronata* |
| 90 | *Lepisosteus oculatus* |
| 91 | *Leptonychotes weddellii* |
| 92 | *Leptosomus discolor* |
| 93 | *Limosa lapponica baueri* |
| 94 | *Limulus polyphemus* |
| 95 | *Lonchura striata domestica* |
| 96 | *Lottia gigantea* |
| 97 | *Loxodonta africana* |
| 98 | *Macaca fascicularis* |
| 99 | *Macaca mulatta* |
| 100 | *Macaca nemestrina* |
| 101 | *Mandrillus leucophaeus* |
| 102 | *Marchantia polymorpha subsp. ruderalis* |
| 103 | *Maylandia zebra* |
| 104 | *Megachile rotundata* |
| 105 | *Meriones unguiculatus* |
| 106 | *Merops nubicus* |
| 107 | *Methylobacterium sp. 174MFSha1.1* |
| 108 | *Microtus ochrogaster* |
| 109 | *Miniopterus natalensis* |
| 110 | *Mizuhopecten yessoensis* |
| 111 | *Monopterus albus* |
| 112 | *Mus caroli* |
| 113 | *Mus musculus* |
| 114 | *Mus pahari* |
| 115 | *Mustela putorius furo* |
| 116 | *Myotis brandtii* |
| 117 | *Myotis davidii* |
| 118 | *Myotis lucifugus* |
| 119 | *Nannospalax galili* |
| 120 | *Nanorana parkeri* |
| 121 | *Natrix tessellata* |
| 122 | *Neolamprologus brichardi* |
| 123 | *Neotoma lepida* |
| 124 | *Nicrophorus vespilloides* |
| 125 | *Nomascus leucogenys* |
| 126 | *Nothobranchius furzeri* |
| 127 | *Notothenia coriiceps* |
| 128 | *Ochotona princeps* |
| 129 | *Octopus bimaculoides* |
| 130 | *Odobenus rosmarus divergens* |
| 131 | *Olea europaea var. sylvestris* |
| 132 | *Oncorhynchus kisutch* |
| 133 | *Oncorhynchus mykiss* |
| 134 | *Ooceraea biroi* |
| 135 | *Opisthocomus hoazin* |
| 136 | *Opisthorchis viverrini* |
| 137 | *Orbicella faveolata* |
| 138 | *Orcinus orca* |
| 139 | *Oreochromis niloticus* |
| 140 | *Ornithorhynchus anatinus* |
| 141 | *Orussus abietinus* |
| 142 | *Orycteropus afer afer* |
| 143 | *Oryctolagus cuniculus* |
| 144 | *Oryzias latipes* |
| 145 | *Pan troglodytes* |
| 146 | *Pantholops hodgsonii* |
| 147 | *Paralichthys olivaceus* |
| 148 | *Parasteatoda tepidariorum* |
| 149 | *Patagioenas fasciata monilis* |
| 150 | *Pelecanus crispus* |
| 151 | *Pelodiscus sinensis* |
| 152 | *Peromyscus maniculatus bairdii* |

| | |
|---|---|
| 153 | *Phalacrocorax carbo* |
| 154 | *Physeter catodon* |
| 155 | *Plecoglossus altivelis* |
| 156 | *Poecilia formosa* |
| 157 | *Poecilia latipinna* |
| 158 | *Poecilia mexicana* |
| 159 | *Poecilia reticulata* |
| 160 | *Pogona vitticeps* |
| 161 | *Priapulus caudatus* |
| 162 | *Protobothrops mucrosquamatus* |
| 163 | *Pseudomyrmex gracilis* |
| 164 | *Pseudopodoces humilis* |
| 165 | *Pundamilia nyererei* |
| 166 | *Pygocentrus nattereri* |
| 167 | *Python bivittatus* |
| 168 | *Rana catesbeiana* |
| 169 | *Rattus norvegicus* |
| 170 | *Rhincodon typus* |
| 171 | *Rhinolophus sinicus* |
| 172 | *Rhinopithecus roxellana* |
| 173 | *Salmo salar* |
| 174 | *Sarcophilus harrisii* |
| 175 | *Scleropages formosus* |
| 176 | *Scomber japonicus* |
| 177 | *Seriola dumerili* |
| 178 | *Seriola lalandi dorsalis* |
| 179 | *Sinocyclocheilus anshuiensis* |
| 180 | *Sinocyclocheilus grahami* |
| 181 | *Sinocyclocheilus rhinocerous* |
| 182 | *Sorex araneus* |
| 183 | *Spinacia oleracea* |
| 184 | *Stegastes partitus* |
| 185 | *Stomoxys calcitrans* |
| 186 | *Strongylocentrotus purpuratus* |
| 187 | *Struthio camelus australis* |
| 188 | *Stylophora pistillata* |
| 189 | *synthetic construct* |
| 190 | *Taeniopygia guttata* |
| 191 | *Takifugu rubripes* |
| 192 | *Tenualosa ilisha* |
| 193 | *Tetraodon nigroviridis* |
| 194 | *Thamnophis sirtalis* |
| 195 | *Thraustotheca clavata* |
| 196 | *Tinamus guttatus* |
| 197 | *Tribolium castaneum* |
| 198 | *Trichinella britovi* |
| 199 | *Trichinella nelsoni* |
| 200 | *Trichinella sp. T9* |
| 201 | *Trichuris suis* |
| 202 | *Tuber melanosporum Mel28* |
| 203 | *Tupaia chinensis* |
| 204 | *Tursiops truncatus* |
| 205 | *Tyto alba* |
| 206 | *ubiquinone* |
| 207 | *Ursus maritimus* |
| 208 | *Vicugna pacos* |
| 209 | *Vollenhovia emeryi* |
| 210 | *Xenopus laevis* |
| 211 | *Xenopus tropicalis* |
| 212 | *Xiphophorus maculatus* |

**Table 5.S4.** Summary of polymorphic microsatellite loci developed from restriction-site associated DNA of *S. longiceps.*

| Repeat motif | Primer name | Primer sequence |
|---|---|---|
| CA | SLSSR:1305:2442:15645 1:N:0:TCTCGCG_1_per2_5F | TGCATGTGTGCACTATTTTCTG |
| | SLSSR:1305:2442:15645 1:N:0:TCTCGCG_1_per2_5R | TGTGTGAGTGGAAGAAGAAGGA |
| CA | SLSSR:1304:18864:2939 1:N:0:TCTCGCG_1_per2_8F | AGAAGGTGCCATTCTCATCTG |
| | SLSSR:1304:18864:2939 1:N:0:TCTCGCG_1_per2_8R | AGTTGCTCACAGTGGGTGTG |
| CA | SLSSR:1307:4717:7219 1:N:0:TCTCGCG_1_per2_6F | GCGCACACGTACCCAGAT |
| | SLSSR:1307:4717:7219 1:N:0:TCTCGCG_1_per2_6R | CTGGCCCCCTGTCCACtat |
| CA | SLSSR:1114:18048:19609 1:N:0:TCTCGCG_1_per2_7F | CTCCGGAACCCCCTATAGAC |

| | | |
|---|---|---|
| | SLSSR:1114:18048:19609 1:N:0:TCTCGCG_1_per2_7R | CCGCTGAATTACTAGGCTACAA |
| CA | SLSSR:1215:2372:88572 1:N:0:TCTCGCG_1_per2_6F | ACCAAAAATGGGGGAGAAAA |
| | SLSSR:1215:2372:88572 1:N:0:TCTCGCG_1_per2_6R | CTGAGGCACCTAGCAACTCC |
| CA | SLSSR:1304:10326:67765 1:N:0:TCTCGCG_1_per2_6F | TATAGCCTCATGCCGAATCA |
| | SLSSR:1304:10326:67765 1:N:0:TCTCGCG_1_per2_6R | AAGTGGCATTTTGCTGGACT |
| CA | SLSSR:1109:20632:63142 1:N:0:TCTCGCG_1_per2_8F | CGCACAAACACTCAGGCATA |
| | SLSSR:1109:20632:63142 1:N:0:TCTCGCG_1_per2_8R | ACCGCTGAATTACTGCTACA |
| CA | SLSSR:1204:4482:45778 1:N:0:TCTCGCG_1_per2_3F | GCCTCATGCTATTCCTTAACTG |
| | SLSSR:1204:4482:45778 1:N:0:TCTCGCG_1_per2_3R | TTCCCACTCCTCACTCAGTC |
| CA | SLSSR:1101:11835:73406 1:N:0:TCTCGCG_1_per2_6F | CAACATAGCAATCAAGACCA |
| | SLSSR:1101:11835:73406 1:N:0:TCTCGCG_1_per2_6R | CCGCTGAATTACAGTGAAAC |
| CA | SLSSR:1108:19625:17638 1:N:0:TCTCGCG_1_PER2_6F | AGCCTCATGCTAATGAGTCAC |
| | SLSSR:1108:19625:17638 1:N:0:TCTCGCG_1_PER2_6R | GCTGAATTCGGTTAGGGTTT |
| CA | SLSSR:1210:10745:68367 1:N:0:TCTCGCG_1_PER2_4F | GGCAAGAGGACAGCAAAGAC |
| | SLSSR:1210:10745:68367 1:N:0:TCTCGCG_1_PER2_4R | GAAAGCGTGGGTATGTGTGA |
| CA | SLSSR:1212:18242:92827 1:N:0:TCTCGCG_1_PER2_8F | CCTCATGCAAACACACATT |
| | SLSSR:1212:18242:92827 1:N:0:TCTCGCG_1_PER2_8R | GTGTAAGGCCTCCCTTGT |
| CA | SLSSR:1214:20083:68669 1:N:0:TCTCGCG_1_PER2_3F | CTCTGGTGACTTTGTTCCAT |
| | SLSSR:1214:20083:68669 1:N:0:TCTCGCG_1_PER2_3R | AATTGGGCATTAGGCTATTT |
| CA | SLSSR:1108:19549:46980 1:N:0:TCTCGCG_1_PER2_8F | ATGCACACACATCGCATAAC |
| | SLSSR:1108:19549:46980 1:N:0:TCTCGCG_1_PER2_8R | TGAGTATGTTTTGGGAAGCAG |
| CA | SLSSR:1205:11452:8398 1:N:0:TCTCGCG_1_PER2_4F | TCATGCACATACACCCACTC |
| | SLSSR:1205:11452:8398 1:N:0:TCTCGCG_1_PER2_4R | CGCTGAATTTATCCCTCTGA |
| CA | SLSSR:1210:16386:31499 1:N:0:TCTCGCG_1_PER2_8F | TTGCATGAATGCAGACACAT |
| | SLSSR:1210:16386:31499 1:N:0:TCTCGCG_1_PER2_8R | CGCTGAATTTCTTAAATAGGC |
| CA | SLSSR:1210:11637:45841 1:N:0:TCTCGCG_1_PER2_6F | GGGGGAACATTCAGGTTTAG |
| | SLSSR:1210:11637:45841 1:N:0:TCTCGCG_1_PER2_6R | CAGATCCATGCCTGCTCTTA |
| CA | SLSSR:1215:21277:98060 1:N:0:TCTCGCG_1_PER2_7F | ATGCACACACATCGCATAAC |
| | SLSSR:1215:21277:98060 1:N:0:TCTCGCG_1_PER2_7R | GGAAGCAGTGCCTACAAGAG |
| CA | SLSSR:1304:20778:29659 1:N:0:TCTCGCG_1_PER2_5F | CCTCATGCAGACATTTCACA |
| | SLSSR:1304:20778:29659 1:N:0:TCTCGCG_1_PER2_5R | TTTTGGTCTAGAGCCTGGTG |
| CA | SLSSR:1207:15305:17450 1:N:0:TCTCGCG_1_PER2_2F | GCCATAGCCTCTCTTCCCTA |
| | SLSSR:1207:15305:17450 1:N:0:TCTCGCG_1_PER2_2R | CCGCTGAATTTGAATGACTAA |
| CT | SLSSR:1305:16321:38518 1:N:0:TCTCGCG_1_PER2_5F | CATATCTGGCAGCTGTGTTA |
| | SLSSR:1305:16321:38518 1:N:0:TCTCGCG_1_PER2_5R | TTCTGTTACGAGCAGCAATA |
| CT | SLSSR:1307:15308:51394 1:N:0:TCTCGCG_1_PER2_9F | CCCAGAGGAAGAGAAGCCTA |
| | SLSSR:1307:15308:51394 1:N:0:TCTCGCG_1_PER2_9R | GCATCTTCTTTTCTGGAGGA |
| GT | SLSSR:1301:9633:92559 1:N:0:TCTCGCG_1_PER2_7F | AGCCTCATGCTAAGTAGTCTGT |
| | SLSSR:1301:9633:92559 1:N:0:TCTCGCG_1_PER2_7R | GCTGAATTACAAAACGTCAA |
| GT | SLSSR:1208:17754:46423 1:N:0:TCTCGCG_1_PER2_9F | AAAAACAGTGGGCAGGAGTG |
| | SLSSR:1208:17754:46423 1:N:0:TCTCGCG_1_PER2_9R | CCCAGACGTAGGGCTTCATA |
| GT | SLSSR:1209:6273:27941 1:N:0:TCTCGCG_1_PER2_8F | TGCATCCAAGTATGAACGTG |
| | SLSSR:1209:6273:27941 1:N:0:TCTCGCG_1_PER2_8R | ATCCAAACTTGGCACTCAGA |
| GT | SLSSR:1110:14725:66314 1:N:0:TCTCGCG_1_PER2_4F | CTCATGCTGAAGCAGATGGA |
| | SLSSR:1110:14725:66314 1:N:0:TCTCGCG_1_PER2_4R | CGCTGAATTCCAGCAATGAT |
| GT | SLSSR:1116:20103:30188 1:N:0:TCTCGCG_1_PER2_3F | GCCTCATGCTAATAAAGCAGAC |
| | SLSSR:1116:20103:30188 1:N:0:TCTCGCG_1_PER2_3R | CTGAATTTCACGCTGCCATA |
| GT | SLSSR:1210:4319:84372 1:N:0:TCTCGCG_1_PER2_7F | CAAACGCACGTTTCTGTATG |
| | SLSSR:1210:4319:84372 1:N:0:TCTCGCG_1_PER2_7R | GACCGGTCACTCCCAAAC |
| GT | SLSSR:1210:14278:84631 1:N:0:TCTCGCG_1_PER2_5F | GAGGCCCCTAGGTAGGTCTT |
| | SLSSR:1210:14278:84631 1:N:0:TCTCGCG_1_PER2_5R | GCTGAATTAAAAAGGCGACA |
| GT | SLSSR:1116:19313:72367 1:N:0:TCTCGCG_1_PER2_5F | CTCCCTTCATCTGTTTCTC |
| | SLSSR:1116:19313:72367 1:N:0:TCTCGCG_1_PER2_5R | TCAGACTGAACAGCCATAG |
| GT | SLSSR:1116:5350:99136 1:N:0:TCTCGCG_1_PER2_6F | CACAAAAGAACACTGTCCA |
| | SLSSR:1116:5350:99136 1:N:0:TCTCGCG_1_PER2_6R | TTACATTCTTGCCACCAC |
| TC | SLSSR:1308:16983:7282 1:N:0:TCTCGCG_1_PER2_1F | CATGCGTACATGCAGATTGT |
| | SLSSR:1308:16983:7282 1:N:0:TCTCGCG_1_PER2_1R | ATTGAGAGAGCGAGGCAAA |
| TAA | SLSSR:1109:10642:70510 1:N:0:TCTCGCG_1_per3_2F | GGCCTATACCAGAGTAATAA |
| | SLSSR:1109:10642:70510 1:N:0:TCTCGCG_1_per3_2R | ACTACAAAAACTAGCGACTG |
| GCA | SLSSR:1104:16062:90989 1:N:0:TCTCGCG_1_per3_6F | GCAGCTGAATGTCCTTGAAA |
| | SLSSR:1104:16062:90989 1:N:0:TCTCGCG_1_per3_6R | AGGGGAGGCTGATAAGAGG |
| GCT | SLSSR:1106:11556:81764 1:N:0:TCTCGCG_1_per3_4F | CAGCTTTGCCACCATAGTCT |
| | SLSSR:1106:11556:81764 1:N:0:TCTCGCG_1_per3_4R | GTGGGACAGAGGAGGTCAG |
| GTT | SLSSR:1108:16597:43447 1:N:0:TCTCGCG_1_per3_3F | GCCCCTTAGTCCTTTAACCA |
| | SLSSR:1108:16597:43447 1:N:0:TCTCGCG_1_per3_3R | ACACTCACTCACCCAAAAGC |
| GCT | SLSSR:1109:19164:37211 1:N:0:TCTCGCG_1_per3_3F | CAGCTTTGCCACCATAGTCT |
| | SLSSR:1109:19164:37211 1:N:0:TCTCGCG_1_per3_3R | GTGGGACAGAGGAGGTCAG |
| CAA | SLSSR:1104:16166:80514 1:N:0:TCTCGCG_1_per3_4F | GCCTCATGCATCAGATAACTT |
| | SLSSR:1104:16166:80514 1:N:0:TCTCGCG_1_per3_4R | CGGCTCCAACAGTCAGATTA |
| CAT | SLSSR:1106:9409:24948 1:N:0:TCTCGCG_1_per3_4F | GCCTCATGCATTTTATGTTG |
| | SLSSR:1106:9409:24948 1:N:0:TCTCGCG_1_per3_4R | TGCTCTGAAGTCGATGACAA |
| GAA | SLSSR:1107:1921:2751 1:N:0:TCTCGCG_1_per3_3F | AGCCTCATGCAATGTTTGAC |
| | SLSSR:1107:1921:2751 1:N:0:TCTCGCG_1_per3_3R | AGTGGTTAAGTGCCTGCAAC |
| TAA | SLSSR:1109:10642:70510 1:N:0:TCTCGCG_1_per3_5F | GACGACGACAACAACAACAA |
| | SLSSR:1109:10642:70510 1:N:0:TCTCGCG_1_per3_5R | GCTGAATTTTAACAGGGACAGA |
| GAT | SLSSR:1104:3046:12994 1:N:0:TCTCGCG_1_per3_6F | TCATGCTCAAGAACAACCAA |
| | SLSSR:1104:3046:12994 1:N:0:TCTCGCG_1_per3_6R | CCGACCATTTAGGTTAACGA |
| TAA | SLSSR:1111:13692:58549 1:N:0:TCTCGCG_1_per3_5F | ATGGTTGTCTTTGGGGAAA |
| | SLSSR:1111:13692:58549 1:N:0:TCTCGCG_1_per3_5R | CCTATTCTCGGACCTCTGGT |
| TAA | SLSSR:1112:16225:89153 1:N:0:TCTCGCG_1_per3_7F | CCTGACGATTCCTTCAATGT |
| | SLSSR:1112:16225:89153 1:N:0:TCTCGCG_1_per3_7R | GCTGAATTTGTCACCTGAGC |

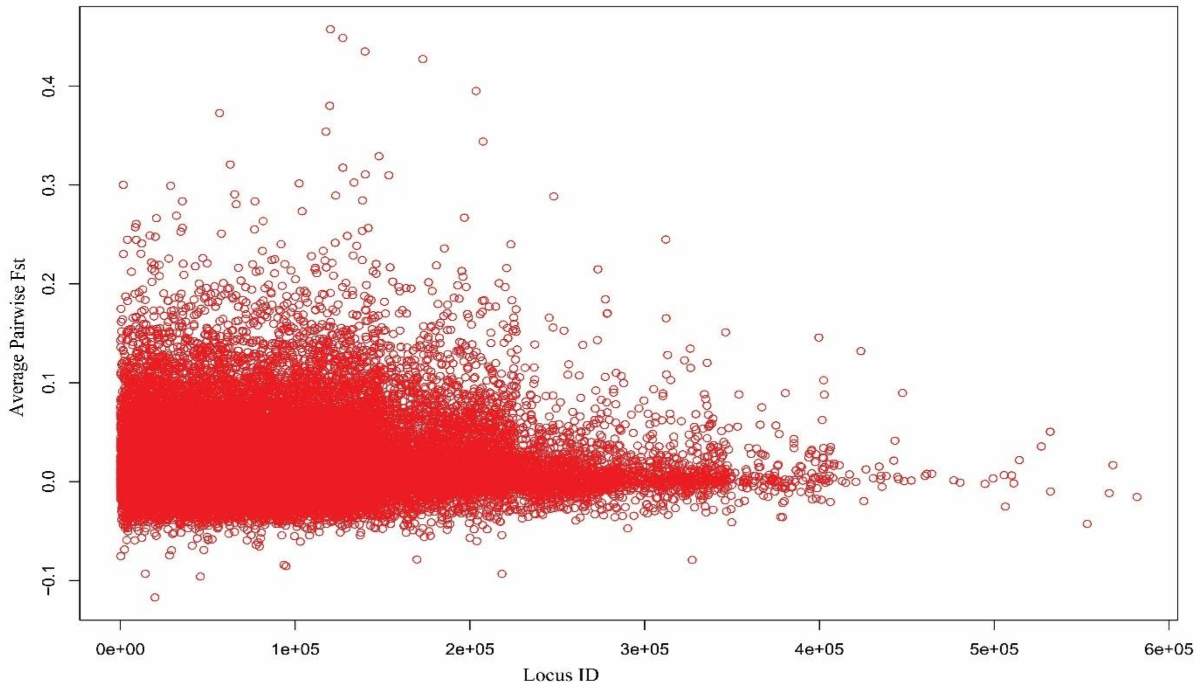


**Fig. 5.S1** Plot of average pairwise $F_{ST}$ of 56,358 SNPs loci between *S. longiceps* population. The x-axis represents the number of ID for each locus and Y-axis indicates the pairwise $F_{ST}$ values.

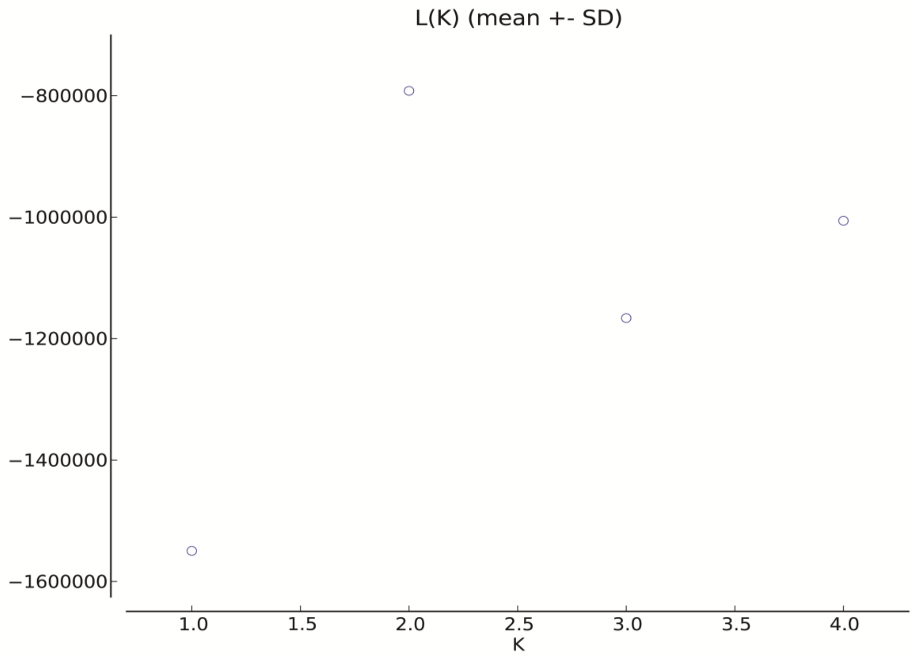


**Fig. 5.S2** Result of structure harvester. Population structuring of *S. longiceps* inferred from STRUCTURE that used the admixture model with correlated allele frequencies when $K = 1–4$. Two genetic clusters were suggested by the maximum value of $\Delta K$ and the order rate of change in posterior likelihood Ln $P$ ($X/K$) per $K$

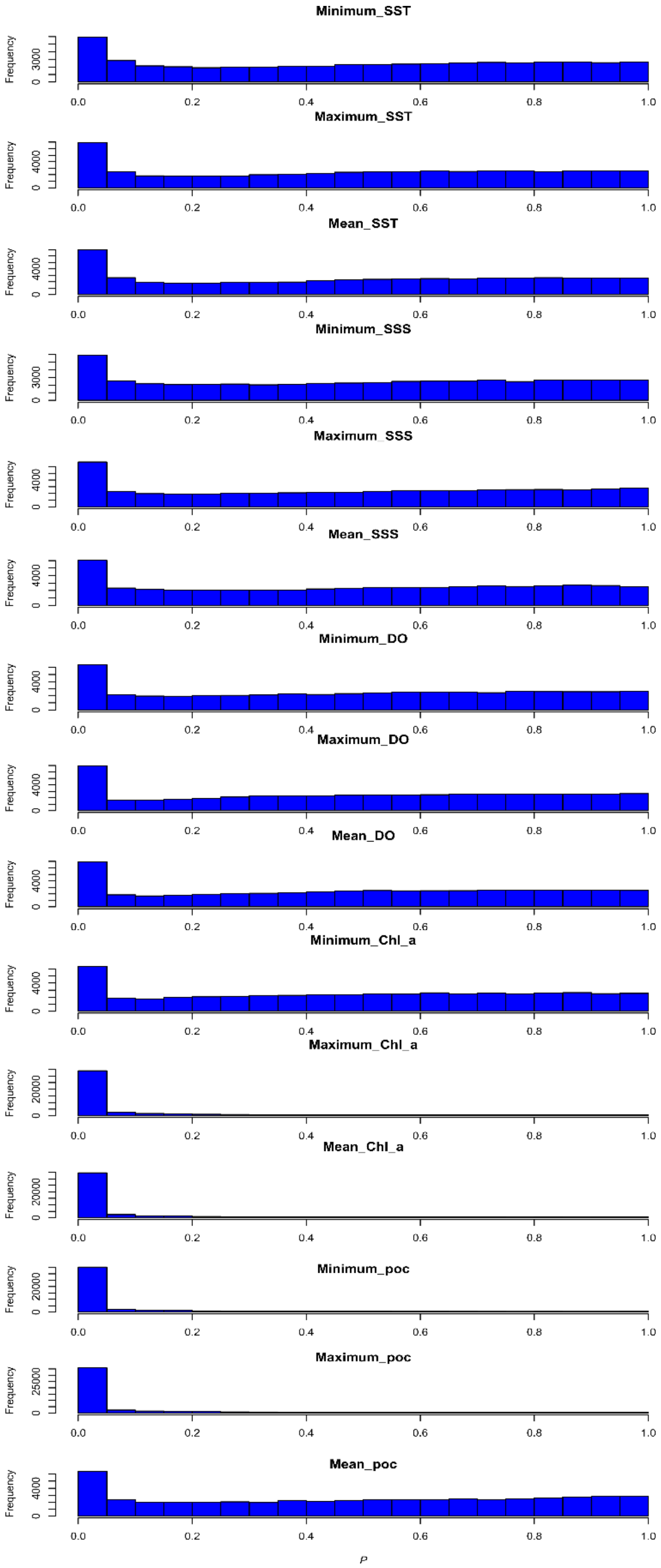


**Fig. 5.S3** The histogram of p-values from LFMM analysis.

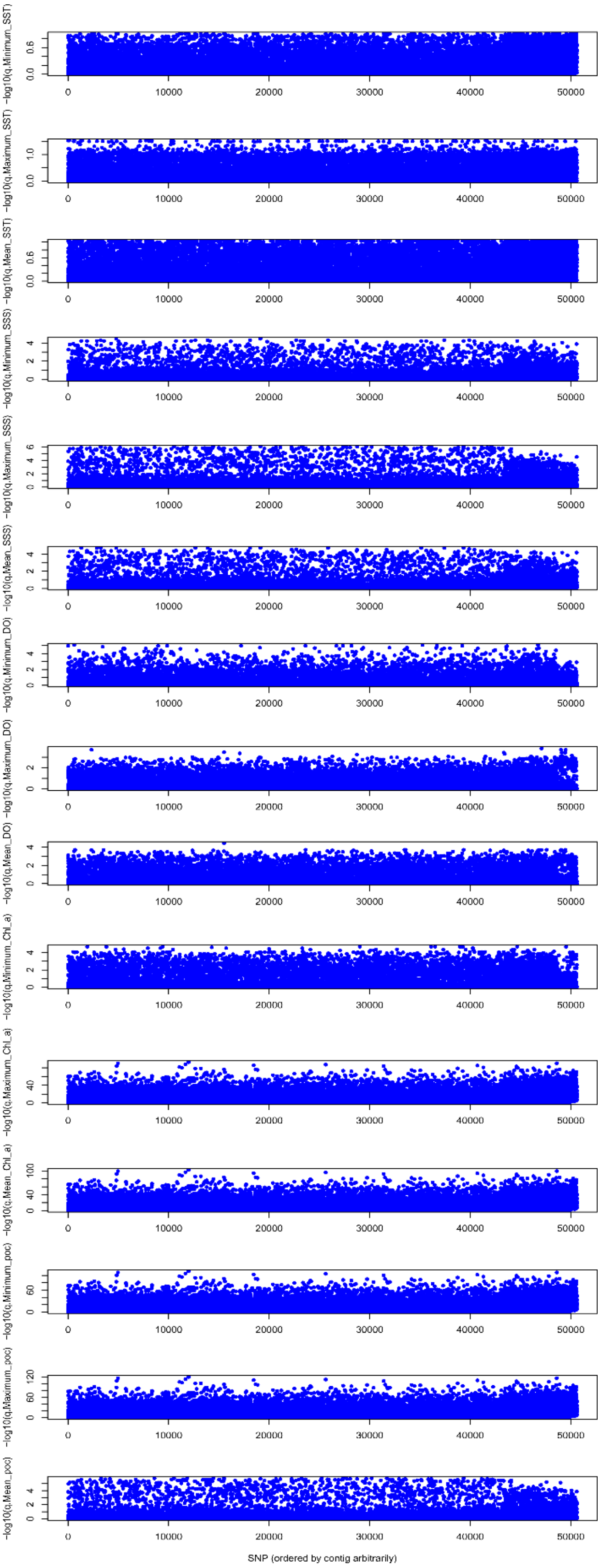


**Fig. 5.S4.** LFMM_Manhattan plot

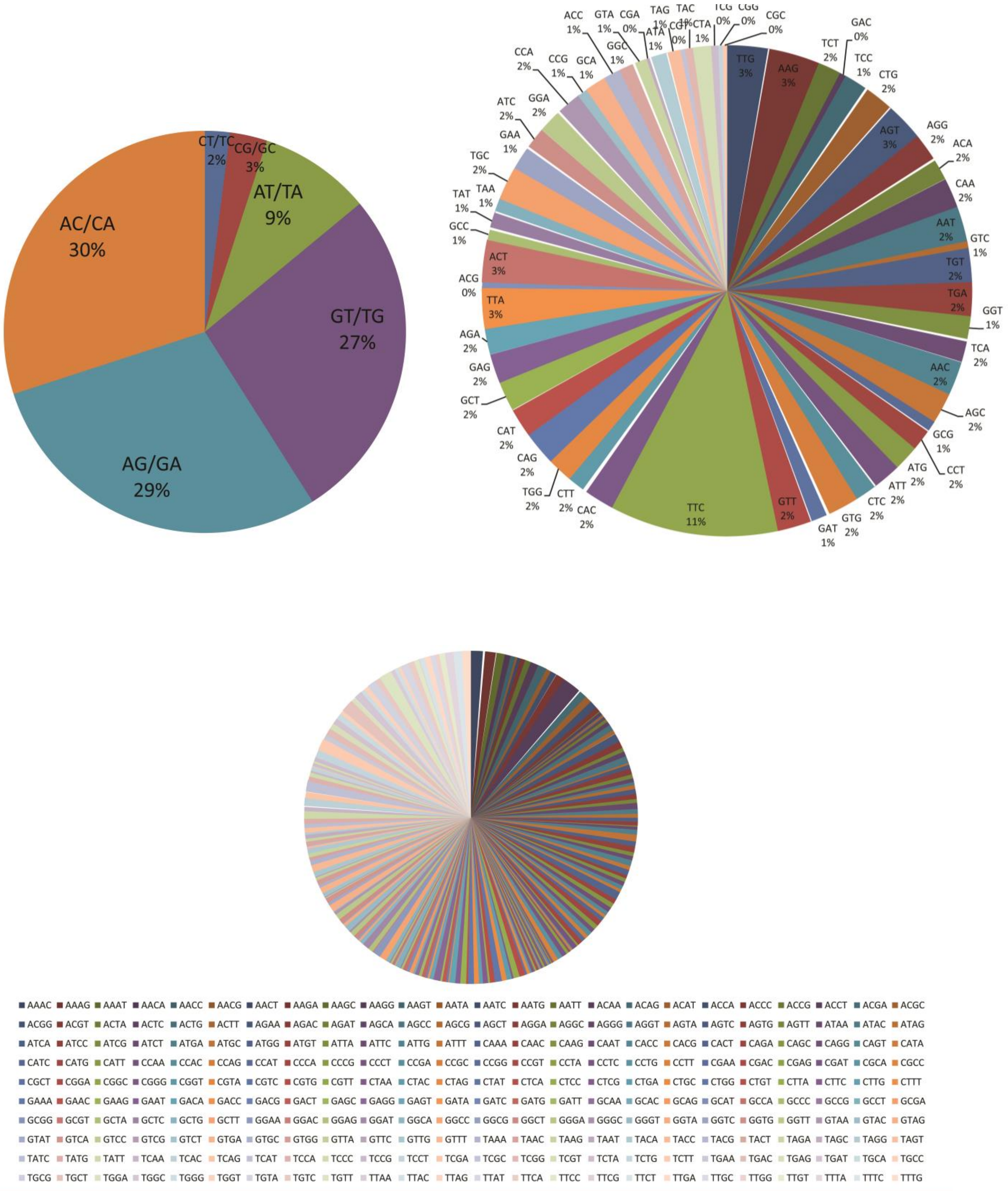


**Fig. 5. S5.** The percentages of di-, tri- and tetra- nucleotide repeats in sequences of SSR motif in *S. longiceps*

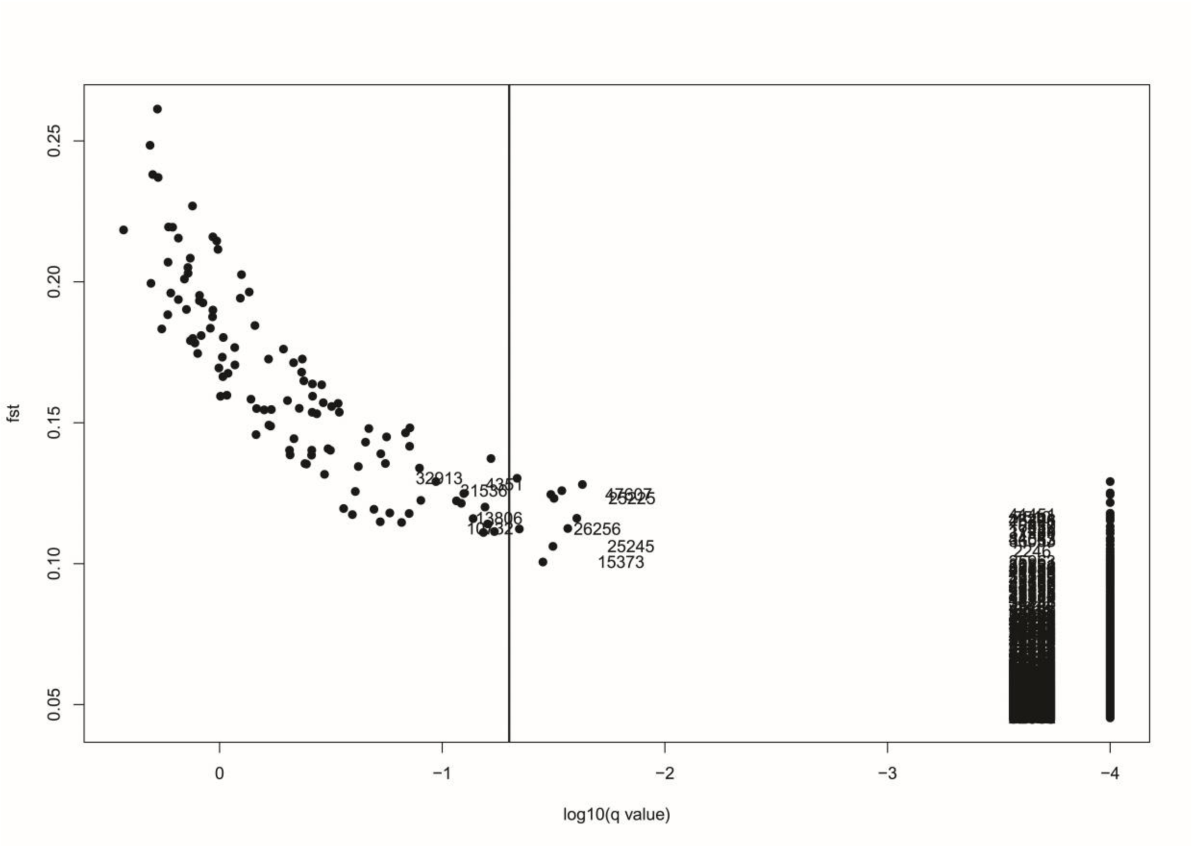


**Fig. 5.S6** F<sub>ST</sub> outlier loci potentially subjected to differential selection among the 56,358.00SNPs loci in *S. longiceps*. The vertical line represents a false discovery threshold of 0.05.

## 5. REFERENCES

1. Alheit J, Roy C, Kifani S (2009) Decadal-scale variability in populations. In:  Checkley D Oozeki Y, Roy C (eds) Climate change and small pelagic fish. Cambridge; Cambridge University Press, United Kingdom
2. Andrew R (2014) Tree figure drawing tool version 1.4.2 2006–2014, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree.
3. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet* 17(2):81
4. Andrews S (2010) FASTQC. A quality control tool for high throughput sequence data. Cambridge, UK: Babraham Institute.
5. Bonnet E, Van de Peer Y (2002) zt: A sofware tool for simple and partial mantel tests. Journal of Statistical software, 7(10):1
6. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, Turner-Maier J (2014) The genomic substrate for adaptive radiation in African Cichlid fish. *Nature* 513(7518):375-381
7. Brennan R S, Hwang R, Tse M, Fangue N A, Whitehead A (2016) Local adaptation to osmotic environment in killifish, *Fundulus heteroclitus*, is supported by divergence in swimming performance but not by differences in excess post-exercise oxygen consumption or aerobic scope. *Comp Biochem Phy A* 196: 11–19
8. Cadrin SX, Kerr LA, Mariani S (2013) Stock identification methods: applications in fishery science. Academic Press
9. Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, Cresko WA (2013a) The population structure and recent colonization history of O regon threespine stickleback determined using restriction-site associated DNA-sequencing. *Mol Ecol* 22(11):2864-2883
10. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013b) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22(11):3124-3140
11. Checkley Jr, DM, Asch RG, Rykaczewski RR (2017) Climate, anchovy, and sardine. *Annu Rev Mar Sci* 9:469-493
12. Cherubin LM, Dalgleish F, Ibrahim AK, Scharer-Umpierre M, Nemeth RS, Matthews A, Appeldoorn R (2020) Fish Spawning Aggregations Dynamics as Inferred from a Novel, Persistent Presence Robotic Approach. *Front Mar Sci* 6:779
13. Claro R, Lindeman KC (2003) Spawning aggregation sites of snapper and grouper species (Lutjanidae and Serranidae) on the insular shelf of Cuba. *Gulf Caribb Res* 14:91-106
14. CMFRI Kochi (2018) *CMFRI Annual Report* 2017-2018. Technical Report. CMFRI, Kochi
15. Crandall KA, Bininda-Emonds OR, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends Ecol Evol* 15(7):290-295
16. Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. *Brief Funct Genomics* 9(5-6):416-23
17. Dawson DA, Horsburgh GJ, Kupper C, Stewart IR, Ball AD, Durrant KL, Hansson B, Bacon I, Bird S, Klein Á, Krupa AP, Lee J-W, Martín-Galvez D, Simeoni M, Smith G, Spurgin LG, Burke T (2010) New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utilityas demonstrated for birds. *Mol Ecol Res* 10:475-494
18. Devanesan DW (1943) A brief investigation into the causes of the fluctuations of the annual fishery of the oil sardine of Malabar, *Sardinella longiceps*, determination of its age and an account of the discovery of its eggs and spawning ground. *Madras Fish Bull* 28 (Report No. 1):1–24
19. Devaraj M, Martosubroto P (1997) Small pelagic resources and their fisheries in the Asia-Pacific Region. Proceedings of APFIC working party on Marine Fisheries. RAP Publishers, Thailand pp 91-198
20. Dowling DC, Wiley MJ (1986) The effects of dissolved oxygen, temperature, and low stream flow on fishes: a literature review. Illinois Natural History Survey (INHS) Aquatic Biology Section
21. Dubreuil M, Sebastiani F, Mayol M *et al*. (2008) Isolation and characterization of polymorphic nuclear microsatellite loci in Taxus baccata L. *Conserv Genet* 9:1665-1668
22. Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359-361
23. Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests: calculating importance gradients on physical predictors. *Ecology* 93(1):156-168
24. Emerson K, Merz C, Catchen J, Hohenlohe P, Cresko W, Bradshaw W *et al*. Resolving postglacial phylogeography using high- throughput sequencing. *Proc Natl Acad Sci USA* 2010; 107: 16196-16200

25. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611-2620
26. Fan S, Elmer KR, Meyer A (2012) Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and Neotropical lineages. *Philos T R Soc B* 367(1587):385-394
27. Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trend Genet* 28(7):342-350
28. Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5:164-166
29. Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180(2):977-993
30. Frichot E, Schoville S, Bouchard G, François O. (2015) LFMM version 1.0 Reference Manual
31. Frichot, E. and François, O., 2015. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol* 6(8):925-929
32. Fungtammasan A, Ananda G, Hile SE, Su MSW, Sun C, Harris R, Medvedev P, Eckert K, Makova KD (2015) Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res* 25(5):736-749
33. Ganias K (2014*)* Biology and ecology of sardines and anchovies. CRC Press
34. Genner MJ, Turner GF (2005) The mbuna Cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish Fish* 6(1):1-34
35. Gleason LU, Burton RS (2016) Genomic evidence for ecological divergence against a background of population homogeneity in the marine snail *Chlorostoma funebralis*. *Mol Ecol* 25(15):3557-3573
36. Gompert Z, Forister ML, Fordyce JA, Nice CC, Williamson RJ, Buerkle CA (2010) Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of Lycaeides butterflies. *Mol Ecol* 19:2455–2473
37. Gruss A, Robinson J (2015) Fish populations forming transient spawning aggregations: should spawners always be the targets of spatial protection efforts?. *ICES J Mar Sci* 72(2):480-497
38. Gu LY, Liu Y, Wang N, Zhang ZW (2012) A panel of polymorphic microsatellites in the Blue Eared Pheasant (*Crossoptilon auritum*) developed by cross-species amplification. *Chin Birds* 3:103–107
39. Hoffmann A, Griffin P, Dillon S, Catullo R, Rane R, Byrne M, Jordan R, Oakeshott J, Weeks A, Joseph L, Lockhart P (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. *BMC Clim Chang Responses* 2(1):1
40. Hohenlohe P, Bassham S, Etter P, Stiffler N, Johnson E, Cresko W (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genet* 6:e1000862
41. Hoskin MG (1997) Effects of contrasting modes of larval development on the genetic structures of populations of three species of prosobranch gastropods. *Mar Biol* 127(4):647-656
42. Hyten D, Song Q, Fickus E, Quigley C, Lim J, Choi I *et al*. (2010) High-throughput SNP discovery and assay development in common bean. *BMC Genomics* 11:475
43. Johannesson K, Smolarz K, Grahn M, Andre C (2011) The future of Baltic Sea populations: local extinction or evolutionary rescue? *Ambio* 40(2):179–190
44. Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11):1403-1405
45. Kocher TD (2004) Adaptive evolution and explosive speciation: the Cichlid fish model. *Nat Rev Genet* 5(4):288-298
46. Krishnakumar K, Raghavan R, Prasad G, Bijukumar A, Sekharan M, Pereira B, Ali A (2009) When pets become pests - exotic aquarium fishes and biological invasions in Kerala, India. *Curr Sci India* 97(4):474-476
47. Kujawa R, Furgała-Selezniow G, Mamcarz A, Lach M, Kucharczyk D (2015) Influence of temperature on the growth and survivability of sichel larvae *Pelecus cultratus* reared under controlled conditions. *Ichthyol Res* 62(2):163-170
48. Larsen PF, Nielsen EE, Williams TD, Hemmer-Hansen JA, Chipman JK, Kruhoffer M, Gronkjaer P, George SG, Dyrskjot L, Loeschcke V (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Mol Ecol* 16(22):4674–4683
49. Lemopoulos A, Prokkola JM, UusiHeikkilaS, Vasemagi A, Huusko A, Hyvarinen P, Koljonen ML, Koskiniemi J, Vainikka A (2019) Comparing RADseq and microsatellites for estimating genetic diversity and relatedness-Implications for brown trout conservation. *Ecol Evol* 9(4):2106-2120
50. Lessios HA, Weinberg JR, Starczak VR (1994) Temporal variation in populations of the marine isopod Excirolana: how stable are gene frequencies and morphology? *Evolution* 48(3):549-563
51. Lischer HE, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28(2):298-299

52. Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A (2017) Breaking RAD: An evaluation of the utility of restriction site associated DNA sequencing for genome scans of adaptation. *Mol Ecol Resour* 17(2):142-152
53. Makinen HS, Cano JM, Merila J (2008a) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Mol Ecol* 17(15):3565-3582
54. Makinen HS, Shikano T, Cano JM, Merila J (2008b) Hitchhiking mapping reveals a candidate genomic region for natural selection in three-spined stickleback chromosome VIII. *Genetics* 178(1):453-65
55. Martins K, Gugger PF, Llanderal-Mendoza J, González-Rodríguez A, Fitz-Gibbon ST, Zhao JL, Rodríguez-Correa H, Oyama K, Sork VL (2018) Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of Quercus rugosa. *Evol Appl* 11(10):1842-1858
56. McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2012) Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol* 62(13):397-406
57. Meglecz E, Neve G, Biffin E, Gardner MG (2012) Breakdown of phylogenetic signal: a survey of microsatellitedensities in 454 shotgun sequences from 154 non model Eukaryote species. *Plos One* 7:e40861
58. Meglécz E, Pech N, Gilles A, Dubut V, Hingamp P, Trilles A *et al*. (2014) QDD version 3.1: a user-friendly com-puter program for microsatellite selection and primer design revisited: experimental validation of vari-ables determining genotyping success rate. *Mol Ecol* 14:1302-1313
59. Meglecz E, Pech N, Gilles A, Dubut V, Hingamp P, Trilles A, Grenier R, Martin JF (2014) QDD version 3.1: A user-friendly computer program for microsatellite selection and primer design revisited: Experimental validation of variables determining genotyping success rate. *Mol Ecol Res* 14(6):1302-1313
60. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* 17:240–248
61. Munroe TA, Priede IG (2010) *Sardinella longiceps* (errata version published in 2017). *The IUCN Red List of Threatened Species* 2010e:T154989A115258997
62. Murty AVS, Edelman MS (1970) On the relation between the intensity of the southwest monsoon and the oil sardine fishery of India. *Indian J Fish* 13:142-149
63. Nair RV (1952) Studies on the revival of the Indian oil sardine fishery. Proc Indo-Pacific Fish Coun 2:1-5
64. Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol* 22(11):2841-2847
65. Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Phil Trans R Soc B* 367(1):332–342
66. Nyanti L, Soo CL, Ahmad-Tarmizi NN, Ling TY, Sim SF, Grinang J, Ganyai T (2018) Effects of water temperature, dissolved oxygen and total suspended solids on juvenile *Barbonymus schwanenfeldii* (Bleeker, 1854) and *Oreochromis niloticus* (Linnaeus, 1758). *AACL* 11(2):394-406.
67. Oliveira EJ, Gomes Pádua J, Zucchi MI, Vencovsky R, Carneiro Vieira ML (2006) Origin, evolution and genome distribution of microsatellites. *Genet Mol Biol* 29(2):294–307
68. Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: a road map for stacks. *Methods Ecol Evol* 8(10):1360-1373
69. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One* 7(5):e37135.
70. Poulsen N, Nielsen EE, Schierup MH, Loeschcke V, Gronkjaer P (2006) Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*). *Mol Ecol* 15(2):321–331
71. Poulsen N, Nielsen EE, Schierup MH, Loeschcke V, Gronkjaer P (2006) Long-term stability and effective population size in North Sea and Baltic Sea cod (Gadus morhua). *Mol Ecol* 15(2):321–331
72. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959
73. Reiss CS, Checkley Jr, DM, Bograd SJ (2008) Remotely sensed spawning habitat of Pacific sardine (*Sardinops sagax*) and Northern anchovy (*Engraulis mordax*) within the California Current. *Fish Oceanogr* 17(2):126-136
74. Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM, Bernatchez L (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philos T R Soc B* 367(1587):354-363

75. Rochette NC, Catchen JM (2017) Deriving genotypes from RAD-seq short-read data using Stacks. *Nat Protoc* 12(12):2640
76. Roman MR, Brandt SB, Houde ED, Pierson JJ (2019) Interactive effects of hypoxia and temperature on coastal pelagic zooplankton and fish. *Front Mar* Sci 6:139
77. Rosenblum EB, Hickerson MJ, Moritz C (2007) Amultilocus perspective on colonization accompanied by selection and gene flow. *Evolution* 61(12):2971–2985
78. Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics, 145(4):1219-1228
79. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In Bioinformatics methods and protocols, Humana Press, Totowa, NJ. pp 365-386
80. Sebastian W, Sukumaran S, Zacharia PU, Gopalakrishnan A (2017) Genetic population structure of Indian oil sardine, *Sardinella longiceps* assessed using microsatellite markers. *Conser Genet* 18(4):951-964
81. Sebastian W, Sukumaran S, Zacharia PU, Muraleedharan KR, Kumar PD, Gopalakrishnan A (2020) Signals of selection in the mitogenome provide insights into adaptation mechanisms in heterogeneous habitats in a widely distributed pelagic fish. *Sci Rep-UK* 10(1):1-4
82. Seehausen O (2006) African Cichlid fish: a model system in adaptive radiation research. *P Roy Soc Lond B Bio* 273(1597):1987-1998
83. Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre GP, Bank C, Brannstrom A, Brelsford A (2014) Genomics and the origin of species. *Nat Rev Genet* 15(3):176.
84. Smedbol RK, McPherson A, Hansen MM, Kenchington E (2002) Myths and moderation in marine metapopulations?. *Fish Fish* 3(1):20–35
85. Sswat M, Stiasny MH, Jutfelt F, Riebesell U, Clemmesen C (2018) Growth performance and survival of larval Atlantic herring, under the combined effects of elevated temperatures and CO2. *Plos One* 13(1):e0191947
86. Strasburg JL, Sherman NA, Wright KM, Moyle LC, Willis JH, Rieseberg LH (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation?. *Philos T R Soc B* 367(1587):364-373
87. Sukumaran S, Gopalakrishnan A, Sebastian W, Vijayagopal P, Nandakumar RS, Raju N, Ismail S, Abdussamad EM, Asokan PK, Said Koya KP, Rohit P (2016a). Morphological divergence in Indian oil sardine, *Sardinella longiceps* Valenciennes, 1847 - Does it imply adaptive variation?. *J Appl Ichthyol 32*(4):706-711
88. Sukumaran S, Sebastian W, Gopalakrishnan A (2016b) Population genetic structure of Indian oil sardine, *Sardinella longiceps* along Indian coast. *Gene* 576(1):372-378
89. Takeda M, Kusumi J, Mizoiri S, Aibara M, Mzighani SI, Sato T, Terai Y, Okada N, Tachida H (2013) Genetic structure of pelagic and littoral Cichlid fishes from Lake Victoria. *Plos One* 8:e74088
90. Thrasher DJ, Butcher BG, Campagna L, Webster MS, Lovette IJ (2018) Double‐digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Mole Ecol Res* 18(5):953-65
91. Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RST, Hecht J, Knaust F, Belkhir K, Klages S, Dieterich R, Stueber K, Piferrer F, Guinand B, Bierne N, Volckaert FA, Bargelloni L, Power DM, Bonhomme F, Canario AVM, Reinhardt R (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun* 5:5770
92. Turner TL, Hahn MW (2010) Genomic islands of speciation or genomic islands and speciation?. *Mol Ecol* 19(5):848-50
93. Valencia LM, Martins A, Ortiz EM, Di Fiore A (2018) A RAD-sequencing approach to genome-wide marker discovery, genotyping, and phylogenetic inference in a diverse radiation of primates. *Plos One* 13(8):e0201254.
94. van Tienderen PH, de Haan AA, van der Linden CG, Vosman B (2002) Biodiversity assessment using markers for ecologically important traits. *Trends Ecol Evol* 17(12):577–582
95. Vendrami DL, De Noia M, Telesca L, Handal W, Charrier G, Boudry P, Eberhart-Phillips L, Hoffman JI (2019) RAD sequencing sheds new light on the genetic structure and local adaptation of *European scallops* and resolves their demographic histories. *Sci Rep* 9(1):1-13
96. Wang L, Liu S, Zhuang Z, Guo L, Meng Z, Lin H (2013) Population genetic studies revealed local adaptation in a high gene-flow marine fish, the small yellow croaker (*Larimichthys polyactis*). *Plos One* 8(12):e83493
97. Wang N, Liu Y, Zhang ZW (2009) Characterization of nine microsatellite loci for a globally vulnerable species, Reeves's Pheasant (*Syrmaticus reevesii*). *Conserv Genet* 10:1511–1514

98. Wei N, Bemmels JB, Dick CW (2014) The effects of read length, quality and quantity on microsatellite discovery and primer development: from Illumina to PacBio. *Mol Ecol Resour* 14:953-965
99. Williams L, Ma X, Boyko A, Bustamante C, Oleksiak M (2010) SNP identification, verification, and utility forpopulation genetics in a non-model genus. *BMC Genet* 11:32
100. Wolf JB, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet* 18(2):87
101. Xu *J, Li JT*, Jiang Y, Peng W, Yao Z, Chen B, Jiang L, Feng J, Ji P, Liu G, Liu Z, Tai R, Dong C, Sun X, Zhao ZX, Zhang Y, Wang J, Li S, Zhao Y, Yang J, Sun X, Xu P (2016) Genomic basis of adaptive evolution: the survival of Amur ide (*Leuciscus waleckii*) in an extremely alkaline environment. *Mol Biol Evol* 34(1):145-149
102. Yamanaka H, Genkai-Kato M, Kohmatsu Y (2017) Effects of water temperature, dissolved oxygen and body mass on the metabolic scope of larvae and juveniles of the nigorobuna carp, *Carassius auratus grandoculis. Kuroshio Science* 11:97-104
103. Yang JB, Li HT, Li DZ *et al*. (2009) Isolation and characterization of microsatellite markers in the endangered species *Taxus wallichiana* using the FIASCO method. *Hort Science* 44:2043-2045
104. Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P (2014) Genomic signature of adaptation to climate in Medicago truncatula. *Genetics*, 196(4):1263-1275
105. Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E *et al*. (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am J Bot* 99:193–208
106. Zellmer AJ, Hanes MM, Hird SM, Carstens BC (2012) Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. *Syst Biol* 61:763–777

# *Chapter 6*

## ANALYSIS OF MITOCHONDRIAL GENOME EVOLUTION OF CLUPEOID FISHES

### ABSTRACT

The vertebrate mitochondrial genome (mtDNA) evolving towards a reduced size is not only under deamination related constraints but also translational efficiency-related constraints (codon amino acid usage constraints). The observed H and L strand base pair composition differences and codon usage bias in mtDNA is a response to the above constraints. The mitochondrial oxidative phosphorylation (OXPHOS) produces 95% of a eukaryotic cell's energy and the membrane protein involved in this system is under high functional constraints. However, the metabolic requirements and the selection forces vary across species and habitat in different individuals. We evaluated the adaptive evolution of mitochondrial genome of 70 clupeoids species having a wide distribution in marine, brackish and freshwaters of tropical and temperate regions.

By comparative mitogenomic analysis of 70 Clupeoids, we observed that both tRNA anticodon composition and tRNA position along the mtDNA was determined by deamination related constraints. The nucleotide of the tRNA anticodon in Clupeoids was saturated with guanine (G) or Thymine (T), positioned around the $O_L$ according to their GT content and the protein-coding regions evolved towards a codon usage pattern, in which most of them are complementary to the T/G saturated tRNA anticodons in the genome. We also found a codon usage pattern specific to fresh/brackish water adapted (radiated) fishes, in which codons evolved to adapt to anticodons. They have a codon usage pattern highly complementary to the GT saturated anticodons in Clupeoids, contrary to their marine counterparts. The results suggest that the Clupeoids mitogenomes are adapted to deamination mutations in anticodon sites, during replication and transcription. The codon usage pattern in Clupeoids was shaped by deamination mutations related constraints in mtDNA. The observed codon usage pattern in euryhaline and freshwater clupeoids may be a result of accelerated directional mutation associated with increased energy requirement for adaptation to the euryhaline and freshwater environment.

The presence and persistence of a non-coding region in mtDNA, known as the control region are against its evolutionary trend, evolving towards a reduced size. It is explained by

the presence of binding sites in the control region (conserved sequence blocks-CSBs) for nuclear-organized proteins that regulate mtDNA maintenance and expression. We performed a comparative mitogenomic investigation of the noncoding control region in 70 Clupeoids to study its evolutionary trend. We confirmed the ability of sequence flanking conserved sequence elements in the control region to form stable secondary structures similar to the tRNA. This stable secondary structure was maintained through a selective constraint as evidenced by low mutation rate and compensatory base substitutions in the stem forming regions. This is the first report of compensatory base substitutions among species that confirm secondary structure formation. The tandem repeats present in the control region originated from the repeat sequences involved in secondary structures associated with conserved sequence elements. The nucleotide polymorphism observed along the flanking regions can be explained as errors that occur during the enzymatic replication of secondary structure-forming regions and repeat elements. The evidence for selective constraints on secondary structures emphasizes the role of the control region in mitogenome function.

This study provides evidence for positive selection in the OXPHOS protein complex of distantly related clupeoid species distributed from temperate to tropic and marine to the freshwater environment. We performed positive selection test and relate the observed variation with the functional sites of secondary and tertiary protein structure by homology protein modelling. Most of the known key functional regions are highly conserved across species. The signatures of adaptive variation in the complex are generally concentrated to loop regions of transmembrane proteins that function as proton pumps. Variations were observed in the property of amino acids, codon usage and base composition across lineages with specific metabolic requirements such as marine to fresh/brackish water transition. Insights from our study showed the need for future experimental characterisation of specific mutations with the efficiency of oxidative phosphorylation and its physiological impact which will be useful for predicting the response of organisms to future climate change and mitochondrial DNA based genetic improvement.

## 1. Introduction

The order Clupeiformes include sardines, herrings, anchovies and other relatives classified into two major suborders mainly, Denticipitoidei and Clupeoidei consisting of more than 390 species belonging to five families: Clupeidae, Engraulidae, Chirocentridae, Pristigastridae, and Sundasalangidae (Lavoue *et al.* 2014). The most important and abundant forage and food fishes are included in this group with a wide distribution in marine, euryhaline and freshwaters of tropical and temperate regions with the highest diversity in the Indo-West Pacific region and a high degree of endemism (Grant *et al.* 2006). The relationship among order Clupeiformes has been investigated extensively using mitogenomes and the phylogeny reconstructed (Lavoue *et al.* 2013). Molecular and paleontological evidence pointed out that the Eastern Tethys sea region was the Indo-West Pacific precursor region where initial diversification of the Clupeoids occurred (Lavoue *et al.* 2013) during the Cretaceous/Palaeogene period. Subsequently, several independent transitions between marine/freshwater/tropical/temperate regions were the cause of evolutionary diversification of clupeids in the world oceans (Ganias *et al.* 2014). Clupeids occupied the marine environment until mid-Cretaceous when the warm climate of the earth induced uniform thermal conditions from equatorial to polar region (Ganias *et al.* 2014). End of the Cretaceous period was characterized by an increase in sea surface temperature which upset the oceanic and atmospheric circulation patterns leading to a mass extinction of plants and animals on the earth (Cretaceous–Palaeogene (K–Pg) extinction event) (Dynesius and Jansson 2000; Zuloaga *et al.* 2019). During the Palaeogene, continents continued to drift closer to their current positions, prominent reduction in average global temperature occurred along with the intensification of snowfall in high altitudes. Repeated glaciations and melting periods, movements of continents, changes in sea levels, oceanic boundaries, the formation of oceans and atmospheric currents and formation of environmental gradients in the early Cenozoic era (Dynesius and Jansson 2000; Zuloaga *et al.* 2019) functioned as drivers for distribution of ancient Clupeoids to different habitats/regions in the world (populations were trapped in the isolated habitats) and subsequent diversification and colonisation by sympatric and allopatric speciation (example, Teske *et al.* 2019; Jansson and Dynesius 2002; Harrisson 2016). This evolutionary diversification, adaptation and colonisation necessitated positive directional selection in the genome.

A typical animal mitochondrial genome encodes 13 proteins, 2 rRNA genes, and 22 tRNA genes. There is a universal conserved gene arrangement among diverse vertebrates and fishes with some exceptions (Boore 1999; Miya and Nishida 2015). The individual strands of the double-stranded mtDNA molecules are indicated as heavy (H) and light (L) strand based on the difference in buoyant densities in a caesium-chloride gradient. Based on the current replication model, the DNA regions located distant from the $O_L$, in the direction of L-strand replication are exposed as single-stranded for a long time, and hence these regions are more prone to deamination mutations (Shadel and Clayton 1997; Reyes *et al*. 1998). The Heavy (H) strand is replicated first, from the H-strand replication origin (Ori $O_H$) inside the control region, the original H strand is then exposed as single-stranded and acts as lagging strand during the synthesis of Light (L) strand. L-strand is replicated from L-strand replication origin (Ori $O_L$) (in the WANCY region), complementary to the original H strand (Clayton 1991). The genes close to the control region are characterized by a high rate of expression and deamination mutation, due to the presence of transcription initiation and H-strand replication ($O_H$) initiation sites respectively (Xia 2005; Satoh *et al*. 2010). The DNA sequences exposed as single-stranded for a long time (during replication and transcription) are prone to deamination mutations (Lindahl 1993). Thus, the tRNA anticodon sites, tRNA gene order, codon usage and base pair composition in the fish mitogenomes could be under constant mutation pressure and translational selection (Xia 2005; Satoh *et al*. 2010). Some empirical studies explain how mitogenome cope with these pressures (Xia 2005; Satoh *et al*. 2010).

The vertebrate mitochondrial genome (mtDNA) is characterized by an exceptional organization by reducing its content. The vertebrate mtDNA lacks introns and the only non-coding region is the control region (Boore 1999). Coding sequences are found continuous to each other, some genes are overlapping (ATP6 and ATP8) and some termination codons are incomplete (post-transcriptionally completed by polyadenylation) (Boore 1999). Despite the evolutionary trend to reduce the size of the mitogenome, the presence and persistence of a non-coding region, the control region indicates its functional importance. But, the function of the control region in mtDNA replication and transcription is not yet clear. The 13 protein-coding genes in the mitogenome are vital for the proper functioning of the Oxidative phosphorylation machinery as they code for core subunits of electron transport along with nuclear-encoded genes. Along with tRNAs, rRNA genes in the mitogenome are vital for the expression of the 13 protein-coding genes and proper functioning of the Oxidative phosphorylation machinery in the mitochondrion (Boore 1999). In humans, transcription of L-strand is initiated from a single promoter (LSP) and H-strand is initiated from two differentially regulated sites, HSP1 (H1) and HSP2 (H2) (Montoya *et al*. 1982). The polycistronic molecules arising out of transcription, corresponding to L and H strands are further converted as individual tRNA, rRNA, and mRNA molecules through the tRNA processing mechanism known as the tRNA punctuation model (Ojala *et al*. 1981). The presence of binding sites in the control region for nuclear-organized proteins that regulate mtDNA maintenance and expression has been proposed as one important role of the control region (Anderson *et al*. 1981; Murakami *et al*. 2002; Pereira *et al*. 2008). Despite that, a clear explanation is lacking regarding the persistence of a large stretch of a noncoding region with no regulatory elements in the mitogenome which is evolving towards a lower genome size.

The mitochondrial length variation/heteroplasmy due to tandem repeat in the control region is a common phenomenon in animals (Brown *et al*. 1986; Wright 2000) and various studies have reported many conserved and repetitive sequences in the mitochondrial control region of many species (Jamandre *et al*. 2014; Miya and Nishida 2015; Sebastian *et al*. 2017). Subsequently, it leads to a discussion on the possible function of these sequences in mitochondrial metabolic function (Melo-Ferreira *et al*. 2014). Regulated expression of mitochondrial genes is essential for the efficient metabolic process in eukaryotic cells, but still, we know little about the mechanisms of mitochondrial transcription and its regulation (Taanman 1999; Nicholls and Minczuk 2014). It is believed that the major molecular machines in mitochondrial replication and gene expression regulation could be directly influenced by components of the control region (Pereira *et al*. 2008; Nicholls and Minczuk 2014; D'Souza and Minczuk 2018). Many vertebrate mitogenomes exhibit conserved control region organization with binding sites for nuclear-encoded regulatory factors (H-strand origin of replication sites-$O_H$, transcription initiation sites, conserved sequence elements such as termination associated sequences and conserved sequence blocks with possible regulatory function), size variation and presence of variable number tandem repeats (VNTR) (Nicholls and Minczuk 2014; Miya and Nishida 2015). But a clear explanation for the occurrence of control region sequences without any regulatory region has not been proposed yet (Parsons *et al*. 1997; Nicholls and Minczuk 2014).

Several genome sequences provide evidence that synonymous codons are not used in equal frequencies (codon usage bias) and codon usage bias has many important roles in RNA processing, protein translation and protein folding (Perna and Kocher 1995; McLean *et al*. 1998). Two major hypotheses explain codon usage bias. The selection hypothesis is based on the concept that codon usage determines the efficiency and/or the accuracy of protein expression (Xia 2005; Kotlar and Lavner 2006; Satoh *et al*. 2010). Thus, the codon bias is generated and maintained by natural selection. On the contrary, the second is based on the mutational or neutral hypothesis. The codon bias exists because of the nonrandom mutational patterns (Xia 2005; Satoh *et al*. 2010). Advances in technologies helped researchers to test these hypotheses and distinguish between the forces that shaped the codon usage pattern observed across genomes and genes (Xia 2005; Satoh *et al*. 2010). Corroboration of both hypotheses has been reported in many studies (Xia 2005; Satoh *et al*. 2010). The selective and neutral hypotheses for codon usage contradict each other, but both mechanisms have a role in codon usage pattern within and between genomes (Xia 2005; Satoh *et al*. 2010).

The mitochondrion is an organelle important in bioenergetics and mitochondrial genomes play crucial roles in evolutionary diversification and adaptation to different thermal regimes. Metabolic performance of the organisms will be affected by selective mutations in genes involved in OXPHOS (Lajbner *et al.* 2018) and hence purifying selection is a major force driving evolution (Jacobsen *et al.* 2016). Despite that, directional/episodic positive selection in response to shifts in selective pressures like limited oxygen availability or greater energy demand has been observed in several organisms. Evidence for adaptive evolution in mtDNA has been accumulating recently (Mishmar *et al.* 2003; da Fonseca *et al.* 2008; Scott *et al.* 2010; Toews and Brelsford 2012; Cheviron *et al*. 2014; Stier *et al.* 2014; Garvin *et al.* 2015a; Morales *et al.* 2016; Carapelli *et al.* 2019; Baker *et al.* 2019) suggesting its possible role in radiation, successful diversification and adaptation to diverse habitats like marine, euryhaline, cold and warm waters (Garvin *et al.* 2011; Garvin *et al.* 2015a,b; Baker *et al.* 2019). Indirect selection from nuclear genome due to mito-nuclear co-evolution also is a factor influencing evolutionary dynamics of mitogenome (Ballard and Pichaud 2014; Morales *et al.* 2016).

The diversity of habitats colonised by Clupeoid fishes along with a high degree of endemism make them excellent candidates for investigations on adaptive evolution and diversifying selection on the mitogenome. Our first objective was studying the effect of both deamination and translational efficiency-related constraints in the Clupeoids mitogenome evolution.

The presence of conserved sequence elements in the control region and their ability to form secondary structures has previously been predicted in many vertebrate species (Broughton and Dowling 1994; Lee *et al*. 1995; Broughton and Dowling 1997; Freeman *et al*. 2001; Pereira *et al*. 2008; Wang *et al*. 2011). But the structural/functional role of other regions/sequence elements of the large stretch of non-coding control regions has not been deciphered yet. We used a comparative mitogenomic approach to understand the exact function of many unexplained regions of the control region by analyzing the mitogenome of 70 Clupeoid fishes and comparing it with the patterns in tRNA.

We hypothesize that the stress induced by habitat transitions and high energy demand will act as a selective pressure on the nuclear and mitochondrial genome. Thus, we also investigated the selection patterns of protein-coding regions of Clupeoid mitogenomes to gain insights regarding codon usage bias and positive selection at variable habitats. The results of the investigation may provide important clues regarding the dynamics of the mitochondrial genome and the adaptation of clupeoid fishes to diverse habitats of world oceans.

## 2. Materials And Methods

### 2.1. Phylogenetic analysis

The complete mitochondrial genomes of 70 Clupeioid species from all the families were selected for analyses. Mitogenome sequence of *Denticeps clupeoides* (the sister group of clupeoids) was selected as the outgroup. Protein coding gene regions were aligned in MEGA 7 using CLUSTALW and a concatenated data set was produced. Subsequently, a maximum likelihood phylogenetic tree was constructed using the General Time Reversible model (+G+I) of substitution selected using Akaike information criterion with 1000 bootstrap replication. Subsequent analyses were carried out using this tree.

### 2.2. Rate of evolution of genes

The difference in the rate of molecular evolution of genes was compared by analysing genetic distance of all genes against their consensus sequence using Dist mat from the EMBOSS package after removing the out-group. Subsequently, a linear least square regression was conducted with a pairwise distance of 12s rRNA and protein-coding genes as described in Fischer *et al.* 2013. The regression coefficient of related distance value was taken as relative rate. Mean (relative) evolutionary rate of each site in protein-coding genes was estimated in MEGA7 under the General Time Reversible model (+G+I).

### 2.3. Codon usage, amino acid usage, tRNA anticodon composition and tRNA position relative to the control region

Codon and amino acid usage were determined for all protein-coding genes after excluding stop codons in DnaSP v5 (Librado and Rozas 2009), MEGA7 (Kumar *et al.* 2016) and Geneious R7 (Kearse *et al.* 2012), visualised in the form of heat map using Microsoft Excel and mapped onto the tree. The average of GC1 and GC2 (GC12) was used for the analysis of neutrality plot (GC12 vs GC3) (Sueoka 1988, 1999). The nucleotide bias, the skew was calculated as (A-T)/(A+T) or (G-C)/(G+C). The effective number of codons (ENc) was estimated with DAMBE 5 (Xia 2013) and this was used as a measure of codon usage bias in genes (Wright 1990). Relative synonymous codon usage (RSCU) was calculated in MEGA7.

## 2.4. Selective constraints on the secondary structure of the mtDNA control region and transfer RNA (tRNA) genes

The DNA sequence of clupeoids was aligned in MEGA v7 (Kumar *et al*. 2016) using CLUSTAL-W and dataset of individual tRNAs and control region was prepared. The conserved sequence blocks (CSBs) and highly variable regions with repeat units in the control region dataset were identified and annotated by comparing with CSBs reported from fishes (Jamandre *et al*. 2014; Sebastian *et al*. 2017). The nucleotide base composition of tRNAs and control region sequence blocks were calculated using Geneious R7 (Kearse *et al*. 2012).

We used the '_mfold webserver' (Zuker 2003) for DNA secondary structure prediction by free energy minimization method with nearest neighbour thermodynamic rules (with 15 window length and 25 step size) for predicting the secondary structure formed by tRNAs and control region sequences. The structure is then visualized using ViennaRNA web services (Gruber *et al*. 2015). To test the stability of the secondary structures predicted, we compared the free energy, (ΔG = kcal/mol) calculated for the control region elements and tRNAs. To compare this, we selected ΔG calculated for tRNA of *Sardinella longiceps* (as a representative), average ΔG and normalized free energy (ΔG/ Length (bp)) of predicted secondary structures of conserved sequence blocks (CSB 3, CSB 2, CSB 1, CSB D, TAS) in 70 Clupeoids. The stability of secondary structures of highly variable regions with repeat units in the mitochondrial control region of Clupeids was assessed by comparing folding energy (ΔG) and normalized free energy (ΔG/Length(bp)) of 22 *S. longiceps* mitochondrial tRNA genes with highly variable regions of 70 Clupeoids.

The pairing regions involved in the secondary structure formation were identified using the aligned dataset of individual tRNAs and conserved sequence blocks (CSBs) of the control region. Selective constraints on the tRNA and control region secondary structures were analyzed by manually identifying complementary mutations in their pairing regions. The selective constraints on the control region sequence blocks were also analyzed by calculated Tajima's D and relative mutation rate values using DnaSP v5 (Librado and Rozas 2009) and MEGA7 programs respectively. We calculated the Tajima's D statistics for whole mtDNA and region of ~1112bp comprising tRNA-pro, control region and tRNA-phe with 10bp intervals overlapping at 5bp. The inter-specific identity of secondary structure-forming regions in the control region elements was analyzed by comparing the conservation status in terms of relative mutation rate and polymorphism in the sequences using Geneious R7.

### 2.5. Positive selection on protein-coding genes

Positive selection on the 13 protein-coding genes of Clupeoides was analysed using six codon-based selection analysis algorithms; Single Likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR) and Mixed Effects Model of Evolution (MEME). These programmes are available in DATA MONKEY (Pond and Frost 2005). For each method we selected a threshold p-value; p < 0.0.5 for SLAC, FEL, MEME and posterior probability > 0.9 for FUBAR. TreeSAAP (Woolley *et al.* 2003) was used to understand changes in physicochemical properties of amino acids caused by replacements. 3D homology model of the protein subunits with positively selected sites was generated by the SWISS-MODEL server (Schwede *et al.* 2003) using appropriate subunit of the protein structure with *Boss taurus* as a template. The positively selected sites were mapped on to the three-dimensional structure.

## 3. Results

The mitogenomic phylogenetic tree reconstructed using Maximum likelihood method showed seven moderately supported monophyletic groups within the Clupeidae (Fig. 6.1), as observed in a previous investigation (Lavoue *et al*. 2007, Lavoue *et al*. 2013). The family Clupeidae and its five subfamilies are not monophyletic. Engraulidae, Pristigasteridae and Dussumieriidae formed well-supported monophyletic groups, and the relationships among other groups are not well supported. The anchovy family Engraulidae is a well-defined monophyletic group (Grande and Nelson 1985; Lavoue *et al*. 2007) with 140 species divided into 16 genera found in temperate and tropical regions around the world. Most anchovies are highly abundant, marine, planktivorous fishes that form large schools in near-shore habitats. Within Engraulinae (sub-family), the New World taxa and Engraulis formed a clade referred to as Engraulini following Lavoue *et al*. (2014). Several morphological characters supported the monophyly of Engraulini, most notably the loss of ventral scutes (Nelson 1970, 1983; Grande and Nelson 1985), a character present in nearly all other clupeomorph fishes (Nelson 1983, 1984, 1986, 1970, 1971; Grande and Nelson 1985).

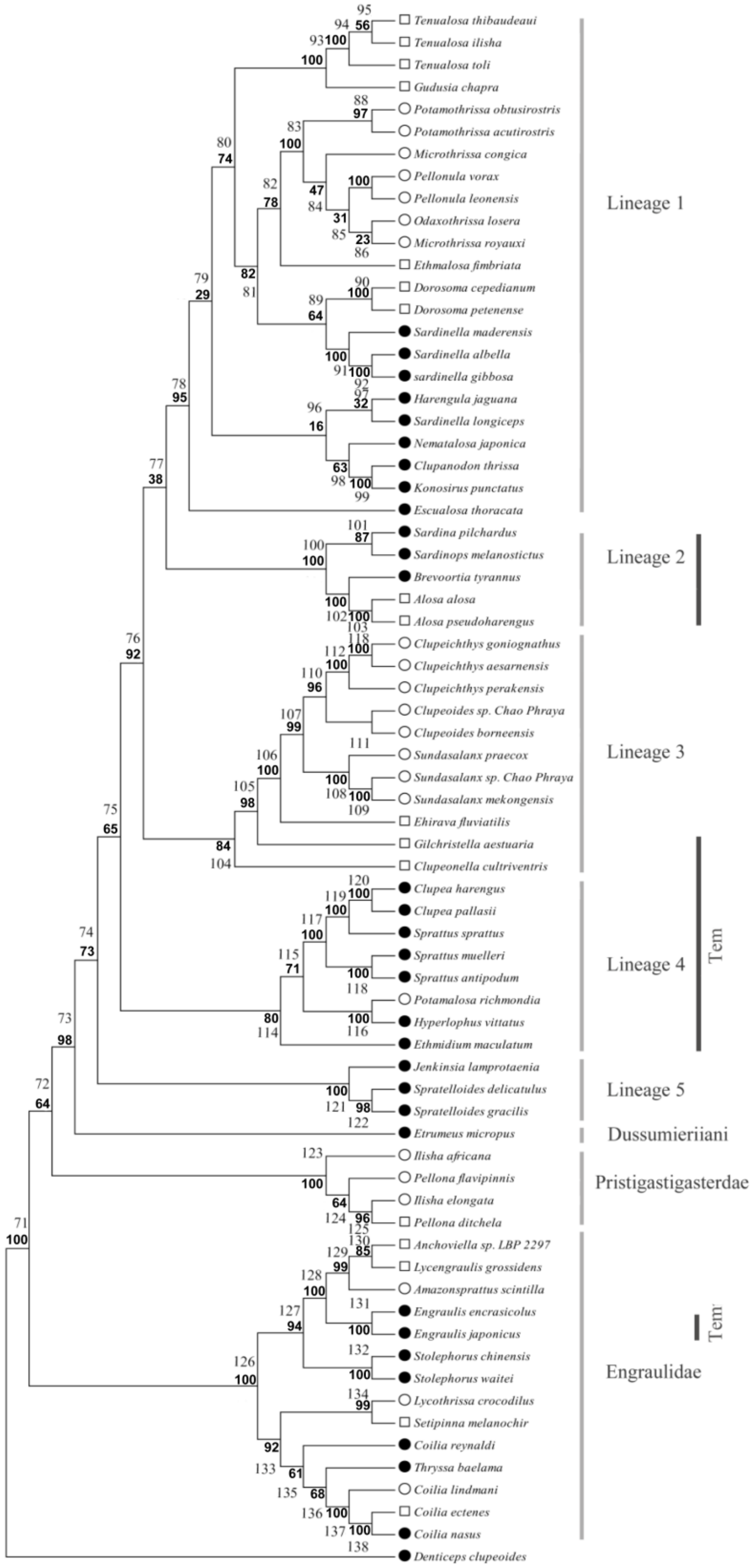
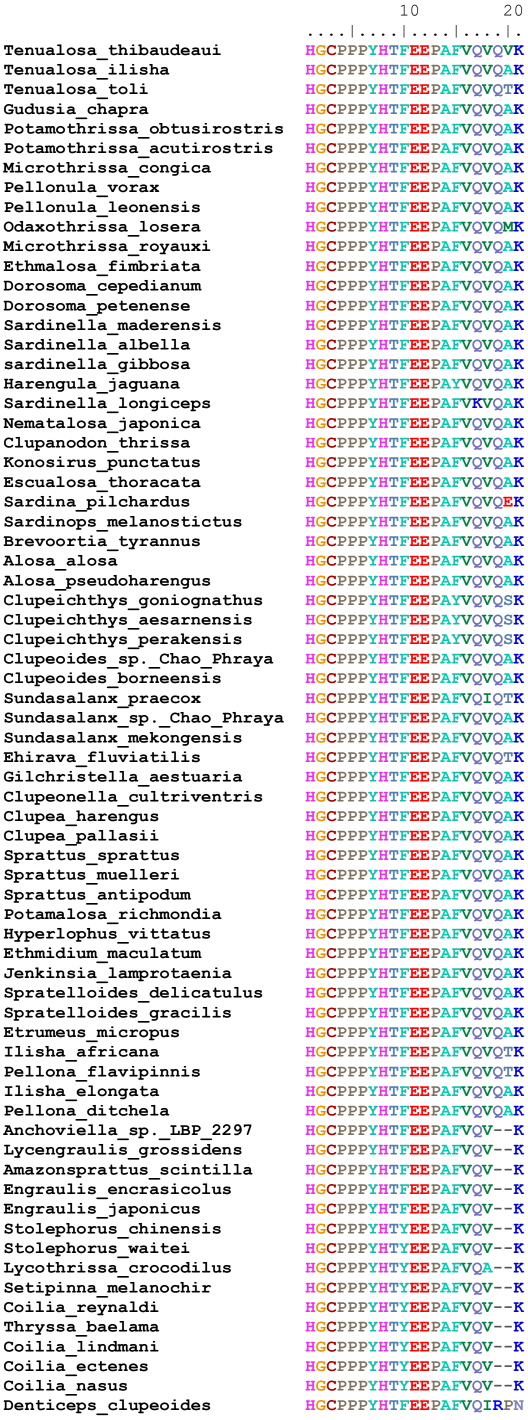


**Fig. 6.1** Maximum likelihood phylogenetic tree generated by alignment of complete mitogenome nucleotide sequences of all considered clupeoid fishes. Bootstrap values and node numbers are indicated in bold and grey letters respectively. Black circle, white circles and square in the tree indicates marine, brackish and freshwater species respectively. 'Temp' indicates temperate water species.

## 3.1. C-terminal end variation in the COI gene

Two variants of the COI gene were observed among Clupeoidei based on the length variation in its 3’ end. This relationship was evident in the phylogenetic tree also as subfamily Engraulidae which lost some residues (loss of 2 amino acids) at its c-terminal formed a separate clade (Fig. 6.2). Even in Denticipitoidei, Cypriniformes and Alocephaliforms COI genes are longer by 2 amino acids.



```
                                          10        20
                                 ....|....|....|....|.
Tenualosa_thibaudeaui            HGCPPPYHTFEEPAFVQVQVK
Tenualosa_ilisha                 HGCPPPYHTFEEPAFVQVQAK
Tenualosa_toli                   HGCPPPYHTFEEPAFVQVQTK
Gudusia_chapra                   HGCPPPYHTFEEPAFVQVQAK
Potamothrissa_obtusirostris      HGCPPPYHTFEEPAFVQVQAK
Potamothrissa_acutirostris       HGCPPPYHTFEEPAFVQVQAK
Microthrissa_congica             HGCPPPYHTFEEPAFVQVQAK
Pellonula_vorax                  HGCPPPYHTFEEPAFVQVQAK
Pellonula_leonensis              HGCPPPYHTFEEPAFVQVQAK
Odaxothrissa_losera              HGCPPPYHTFEEPAFVQVQMK
Microthrissa_royauxi             HGCPPPYHTFEEPAFVQVQAK
Ethmalosa_fimbriata              HGCPPPYHTFEEPAFVQVQAK
Dorosoma_cepedianum              HGCPPPYHTFEEPAFVQVQAK
Dorosoma_petenense               HGCPPPYHTFEEPAFVQVQAK
Sardinella_maderensis            HGCPPPYHTFEEPAFVQVQAK
Sardinella_albella               HGCPPPYHTFEEPAFVQVQAK
sardinella_gibbosa               HGCPPPYHTFEEPAFVQVQAK
Harengula_jaguana                HGCPPPYHTFEEPAYVQVQAK
Sardinella_longiceps             HGCPPPYHTFEEPAFVKVQAK
Nematalosa_japonica              HGCPPPYHTFEEPAFVQVQAK
Clupanodon_thrissa               HGCPPPYHTFEEPAFVQVQAK
Konosirus_punctatus              HGCPPPYHTFEEPAFVQVQAK
Escualosa_thoracata              HGCPPPYHTFEEPAFVQVQAK
Sardina_pilchardus               HGCPPPYHTFEEPAFVQVQEK
Sardinops_melanostictus          HGCPPPYHTFEEPAFVQVQAK
Brevoortia_tyrannus              HGCPPPYHTFEEPAFVQVQAK
Alosa_alosa                      HGCPPPYHTFEEPAFVQVQAK
Alosa_pseudoharengus             HGCPPPYHTFEEPAFVQVQAK
Clupeichthys_goniognathus        HGCPPPYHTFEEPAYVQVQSK
Clupeichthys_aesarnensis         HGCPPPYHTFEEPAYVQVQSK
Clupeichthys_perakensis          HGCPPPYHTFEEPAYVQVQSK
Clupeoides_sp._Chao_Phraya       HGCPPPYHTFEEPAFVQVQAK
Clupeoides_borneensis            HGCPPPYHTFEEPAFVQVQAK
Sundasalanx_praecox              HGCPPPYHTFEEPAFVQIQTK
Sundasalanx_sp._Chao_Phraya      HGCPPPYHTFEEPAFVQVQAK
Sundasalanx_mekongensis          HGCPPPYHTFEEPAFVQVQAK
Ehirava_fluviatilis              HGCPPPYHTFEEPAFVQVQTK
Gilchristella_aestuaria          HGCPPPYHTFEEPAFVQVQAK
Clupeonella_cultriventris        HGCPPPYHTFEEPAFVQVQAK
Clupea_harengus                  HGCPPPYHTFEEPAFVQVQAK
Clupea_pallasii                  HGCPPPYHTFEEPAFVQVQAK
Sprattus_sprattus                HGCPPPYHTFEEPAFVQVQAK
Sprattus_muelleri                HGCPPPYHTFEEPAFVQVQAK
Sprattus_antipodum               HGCPPPYHTFEEPAFVQVQAK
Potamalosa_richmondia            HGCPPPYHTFEEPAFVQVQAK
Hyperlophus_vittatus             HGCPPPYHTFEEPAFVQVQAK
Ethmidium_maculatum              HGCPPPYHTFEEPAFVQVQAK
Jenkinsia_lamprotaenia           HGCPPPYHTFEEPAFVQVQAK
Spratelloides_delicatulus        HGCPPPYHTFEEPAFVQVQAK
Spratelloides_gracilis           HGCPPPYHTFEEPAFVQVQAK
Etrumeus_micropus                HGCPPPYHTFEEPAFVQVQAK
Ilisha_africana                  HGCPPPYHTFEEPAFVQVQTK
Pellona_flavipinnis              HGCPPPYHTFEEPAFVQVQTK
Ilisha_elongata                  HGCPPPYHTFEEPAFVQVQAK
Pellona_ditchela                 HGCPPPYHTFEEPAFVQVQAK
Anchoviella_sp._LBP_2297         HGCPPPYHTFEEPAFVQV--K
Lycengraulis_grossidens          HGCPPPYHTFEEPAFVQV--K
Amazonsprattus_scintilla         HGCPPPYHTFEEPAFVQV--K
Engraulis_encrasicolus           HGCPPPYHTFEEPAFVQV--K
Engraulis_japonicus              HGCPPPYHTFEEPAFVQV--K
Stolephorus_chinensis            HGCPPPYHTYEEPAFVQV--K
Stolephorus_waitei               HGCPPPYHTYEEPAFVQV--K
Lycothrissa_crocodilus           HGCPPPYHTYEEPAFVQA--K
Setipinna_melanochir             HGCPPPYHTYEEPAFVQV--K
Coilia_reynaldi                  HGCPPPYHTYEEPAFVQV--K
Thryssa_baelama                  HGCPPPYHTYEEPAFVQV--K
Coilia_lindmani                  HGCPPPYHTYEEPAFVQV--K
Coilia_ectenes                   HGCPPPYHTYEEPAFVQV--K
Coilia_nasus                     HGCPPPYHTYEEPAFVQV--K
Denticeps_clupeoides             HGCPPPYHTFEEPAFVQIRPN
```

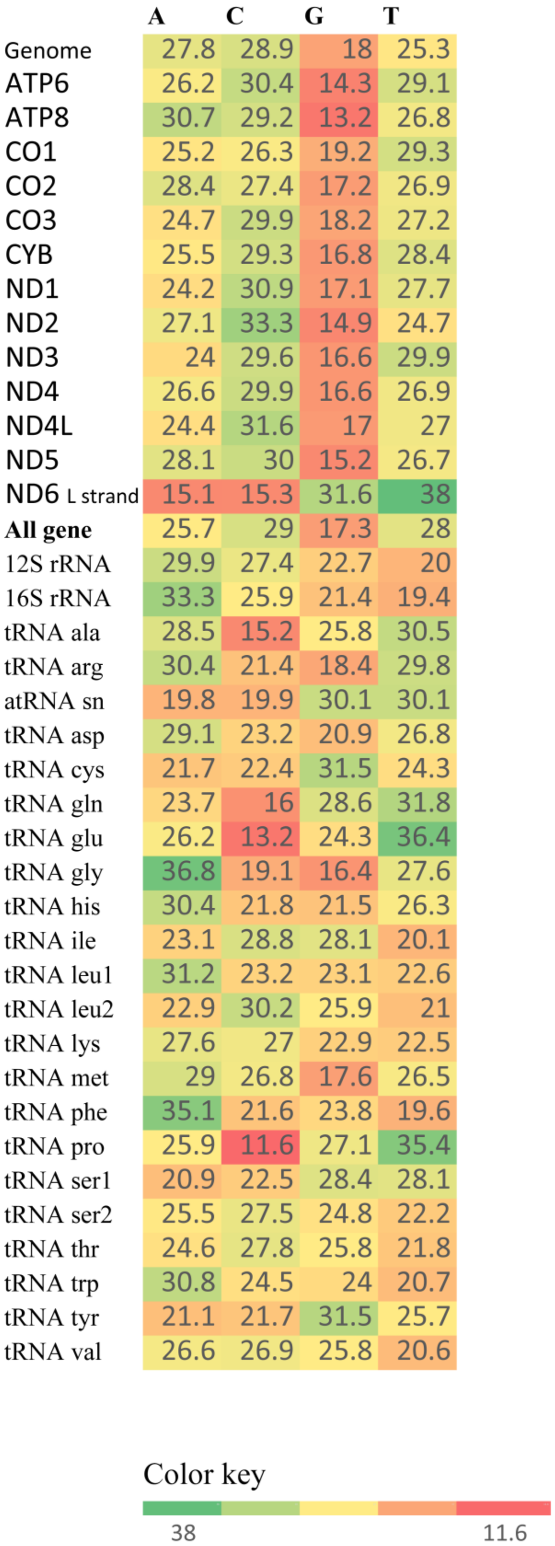
**Fig. 6.2** C-terminal end variation in the CO1 gene of clupeoid fishes.

### 3.2. Relative rate of gene evolution

When the regression-based approach was considered, the highest regression coefficient was observed for the ND3 gene, followed by ND5, CO2 and CO1 genes. Lowest was observed for ND4, ND1, CYTB and ND6 genes (Table 6.S1). The position by position relative rate shows that the second codon position evolves slower than the first and third codon position (Appendix Fig. A3 (a); Fig. A3 (b)). Third codon position is highly evolving and the observation is consistent with the neutral theory of molecular evolution. According to the neutral theory of evolution, the synonymous sites in the protein-coding gene will evolve faster than non-synonymous sites due to the strong selection pressure (Kimura 1983). Most of the changes in the second codon position are non-synonymous, thus they should be under purifying selection. Some mutations at first codon positions and most at third codon positions are synonymous and hence they occur more in population with a chance to get fixed over time. But there is evidence that synonymous sites in vertebrate genes are selectively evolved (Galtier *et al.* 2009; Kunstner *et al.* 2011; Nabholz *et al.* 2011). This indicates that nucleotide and synonymous codon usage bias observed in clupeid mitogenome in the present study is not just because of mutation bias but also due to natural selection.

### 3.3. tRNA anticodon composition and codon usage

The overall base composition of 70 clupeoids mtDNA H-strand dataset consists of A-27.8%, C-28.9%, G-18% and T-25.3% and the coding gene dataset consists of A-25.7%, C-29%, G-17.3% and T-28% (Fig. 6.3). As expected, the base composition of ND6 (L-strand coded) differs from the remaining genes (H-strand coded) with a shift towards T and G (ND6- 15.1% A, 15.3% C, 31.6% G and 38% T; Other genes 24-30.9% A, 26.3-33.3% C, 13.2-19.2% G and 24.7-29.9% T), indicating the difference in the nucleotide composition of H and L strand. The observed low G and high A+T content were similar to the pattern observed in other vertebrates (Boore 1999). GC skews for all the mitochondrial genome and AT skews of most of the genome are negative. This also indicates the richness of T and C in the L strand and asymmetry in nucleotide composition between the two strands. This is a common phenomenon in the mitochondrial genome and generally, the asymmetric pattern in nucleotide composition of DNA strands was explained by the asymmetry in the mutational pressure on DNA strands during replication and transcription (Bulmer 1987, 1991; Necsulea and Lobry 2007).



| | A | C | G | T |
|---|---|---|---|---|
| Genome | 27.8 | 28.9 | 18 | 25.3 |
| ATP6 | 26.2 | 30.4 | 14.3 | 29.1 |
| ATP8 | 30.7 | 29.2 | 13.2 | 26.8 |
| CO1 | 25.2 | 26.3 | 19.2 | 29.3 |
| CO2 | 28.4 | 27.4 | 17.2 | 26.9 |
| CO3 | 24.7 | 29.9 | 18.2 | 27.2 |
| CYB | 25.5 | 29.3 | 16.8 | 28.4 |
| ND1 | 24.2 | 30.9 | 17.1 | 27.7 |
| ND2 | 27.1 | 33.3 | 14.9 | 24.7 |
| ND3 | 24 | 29.6 | 16.6 | 29.9 |
| ND4 | 26.6 | 29.9 | 16.6 | 26.9 |
| ND4L | 24.4 | 31.6 | 17 | 27 |
| ND5 | 28.1 | 30 | 15.2 | 26.7 |
| ND6 L strand | 15.1 | 15.3 | 31.6 | 38 |
| **All gene** | 25.7 | 29 | 17.3 | 28 |
| 12S rRNA | 29.9 | 27.4 | 22.7 | 20 |
| 16S rRNA | 33.3 | 25.9 | 21.4 | 19.4 |
| tRNA ala | 28.5 | 15.2 | 25.8 | 30.5 |
| tRNA arg | 30.4 | 21.4 | 18.4 | 29.8 |
| atRNA sn | 19.8 | 19.9 | 30.1 | 30.1 |
| tRNA asp | 29.1 | 23.2 | 20.9 | 26.8 |
| tRNA cys | 21.7 | 22.4 | 31.5 | 24.3 |
| tRNA gln | 23.7 | 16 | 28.6 | 31.8 |
| tRNA glu | 26.2 | 13.2 | 24.3 | 36.4 |
| tRNA gly | 36.8 | 19.1 | 16.4 | 27.6 |
| tRNA his | 30.4 | 21.8 | 21.5 | 26.3 |
| tRNA ile | 23.1 | 28.8 | 28.1 | 20.1 |
| tRNA leu1 | 31.2 | 23.2 | 23.1 | 22.6 |
| tRNA leu2 | 22.9 | 30.2 | 25.9 | 21 |
| tRNA lys | 27.6 | 27 | 22.9 | 22.5 |
| tRNA met | 29 | 26.8 | 17.6 | 26.5 |
| tRNA phe | 35.1 | 21.6 | 23.8 | 19.6 |
| tRNA pro | 25.9 | 11.6 | 27.1 | 35.4 |
| tRNA ser1 | 20.9 | 22.5 | 28.4 | 28.1 |
| tRNA ser2 | 25.5 | 27.5 | 24.8 | 22.2 |
| tRNA thr | 24.6 | 27.8 | 25.8 | 21.8 |
| tRNA trp | 30.8 | 24.5 | 24 | 20.7 |
| tRNA tyr | 21.1 | 21.7 | 31.5 | 25.7 |
| tRNA val | 26.6 | 26.9 | 25.8 | 20.6 |

Color key

🟩🟩🟨🟧🟥

38 11.6

**Fig. 6.3** Percentage of A, T, G and C of all considered clupeoid fishes mitogenome, protein-coding genes, merged protein-coding gene, 12S rRNA, 16S rRNA and tRNAs.

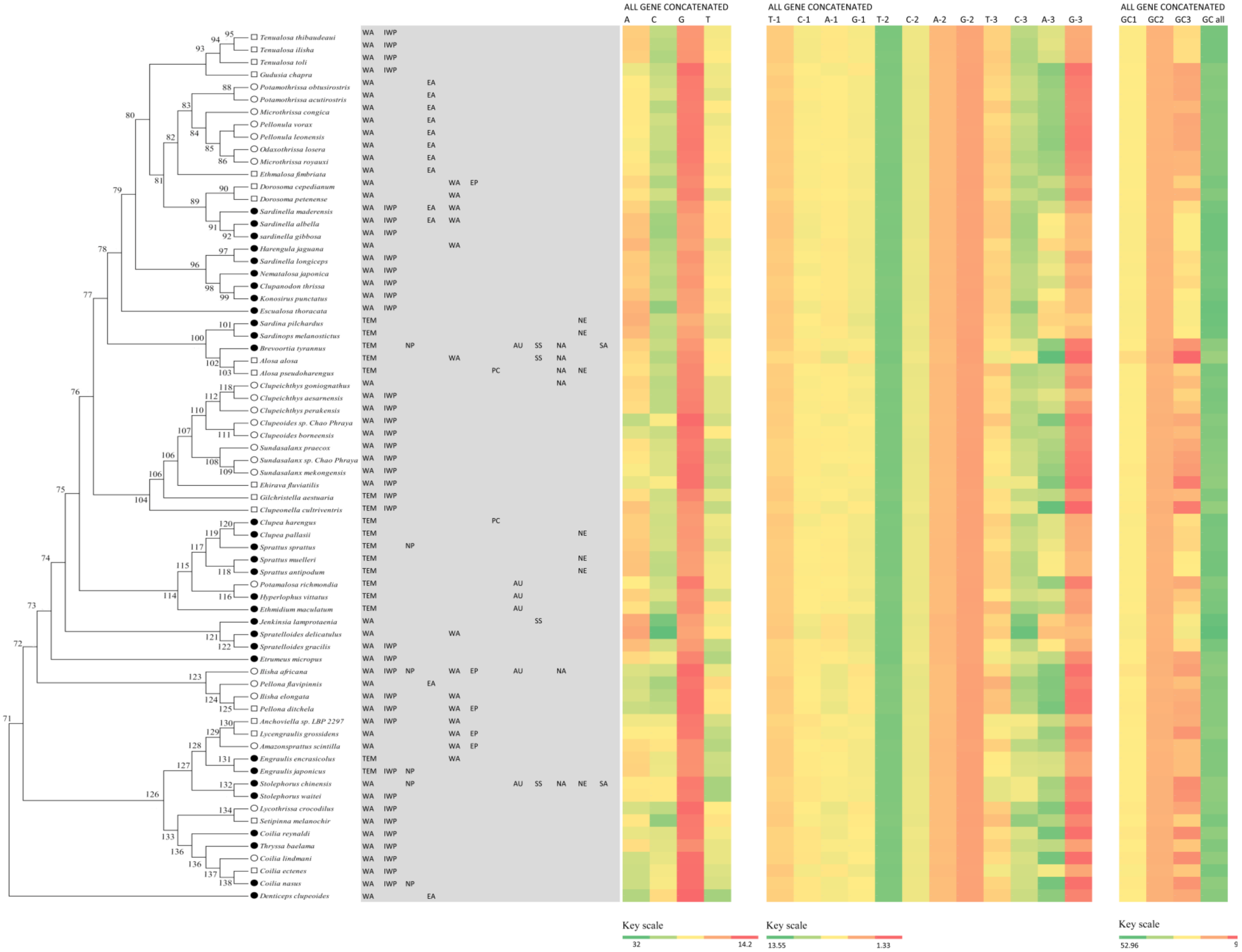


**Fig. 6.4 A, T, G and C contents varying across the clupeoid mitogenomic phylogenetic tree**. A, T, G & C contents at different codon positions and GC content at different codon positions of merged protein coding genes for all considered clupeoid fishes. Number in the node of the phylogenetic tree indicate node-numbers. Black circle, white circles and square in the tree indicates marine, brackish and freshwater species respectively. Biogeographical distribution of the Clupeoidei: IWP Indo-West Pacific, NP North Pacific, EA East Atlantic, WA West Atlantic, EP East Pacific, AU South Australia, SS south South America, NA Northwest Atlantic, NE Northeast Atlantic, SA South Africa, M Marine, F Fresh Water,E Euryhaline, Tem Temperate Water (t<25), Wa Warm Water (t>25)

Distribution of A and G in the marine lineages vs other lineages (fresh and brackish water) showed a remarkable difference. Except for Engraulidae and Tenualosa, all the marine species showed a shift towards high G (18-29%) and low A (20-25%) when compared with euryhaline and freshwater fishes (A 26-29% and G 14-17%) (Fig. 6.4). Even though there is no remarkable difference in the distribution of nucleotides at the 1st and 2nd codon positions between the species, the 3rd codon position showed a clear bias, especially in the Adenine (A3) and Guanine (G3) composition. Both freshwater and euryhaline species preferred A over G, whereas the marine lineages preferred G over A in the third codon position except in Engraulidae. The base composition of both L strand (rich in A+C) and H strand (rich in G+T) genes are consistent with the strand-specific mutational bias observed in the mitogenomes of vertebrate (Boore 1999). The strand-specific base composition was also seen in tRNA (Fig. 6.3). The Clupeoid genome is rich in Leucine (Leu ~ 16%) followed by Alanine (Ala~9%) and Threonine (Thr 8.5%). Asparagine, Arginine, Lysine (2%) and Cysteine (~0.8%) occurred the least. The RSCU results indicated a bias in the codon usage in Clupeoid mitogenome, with a strong anti-G bias in codon usage, codons with A and C at 3rd codon position are abundant than those with T and G (Fig. 6.5). We found a gradient that exists in the arrangement of genes and amino acid composition related to the position of the origin of replication (Ori L and Ori H), control region (CR) and codon usage in Clupeoids mitogenomes (Fig. 6.6). Based on previous reports, the tRNA with anticodons of highly used codons will be positioned near the control region, where transcription efficiency is high (Satoh *et al*. 2010). Among those, the tRNA with anticodons corresponding to the hydrophobic amino acids are high in frequency (Fig. 6.S1; Fig. 6.S8), and consequent encoding of hydrophobic transmembrane protein by mtDNA (Satoh *et al*. 2010). We also found that the nucleotide of the tRNA anticodon in Clupeoids was saturated with guanine (G) or Thymine (T), except tRNA Methionine and Proline (Fig. 6.S1). In addition to this, we found a codon usage pattern specific to fresh/brackish water adapted (radiated) fishes in lineage 1-5 (Fig. 6.4). They have a codon usage pattern highly complementary to the GT saturated anticodons, contrary to their marine counterparts. The results suggest that the tRNA anticodon sites (base composition), tRNA gene order and codon usage in the mitogenome of Clupeoids are adapted to mutational pressure and translational selection (Xia 2005; Satoh *et al*. 2010). The shift in the codon usage pattern of fresh/brackish water radiated Clupeoids may help them to adapt to the new environment.

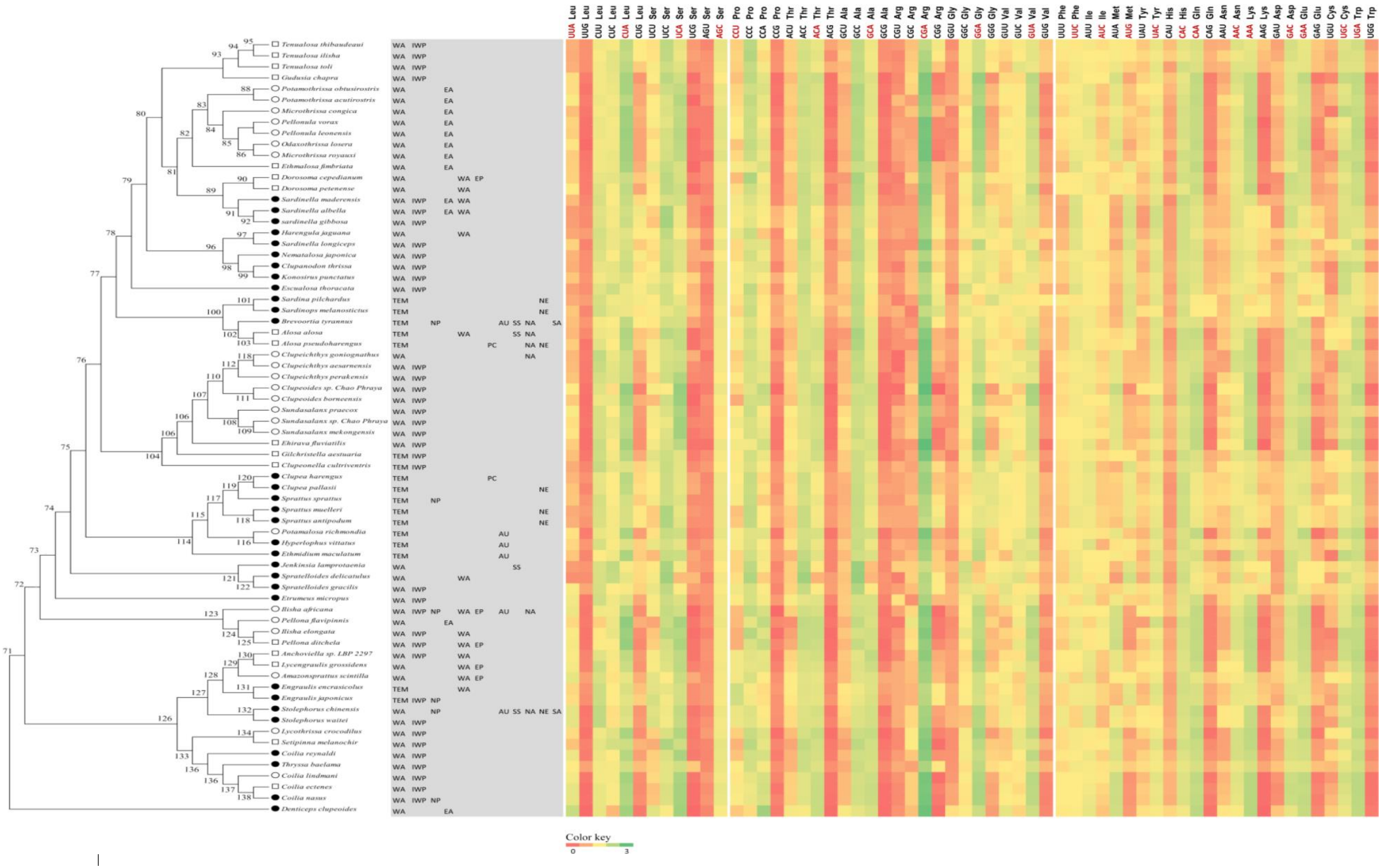


**Fig. 6.5** RSCU values of merged protein-coding genes of Clupeoid fishes. Number in the node of the phylogenetic tree indicate node-numbers. Black circle, white circles and square in the tree indicates marine, brackish and freshwater species respectively.Biogeographical distribution of the Clupeoidei: IWP Indo-West Pacific, NP North Pacific, EA East Atlantic, WA West Atlantic, EP East Pacific, AU South Australia, SS south South America, NA Northwest Atlantic, NE Northeast Atlantic, SA South Africa, M Marine, F Fresh Water, E Euryhaline, Tem Temperate

A

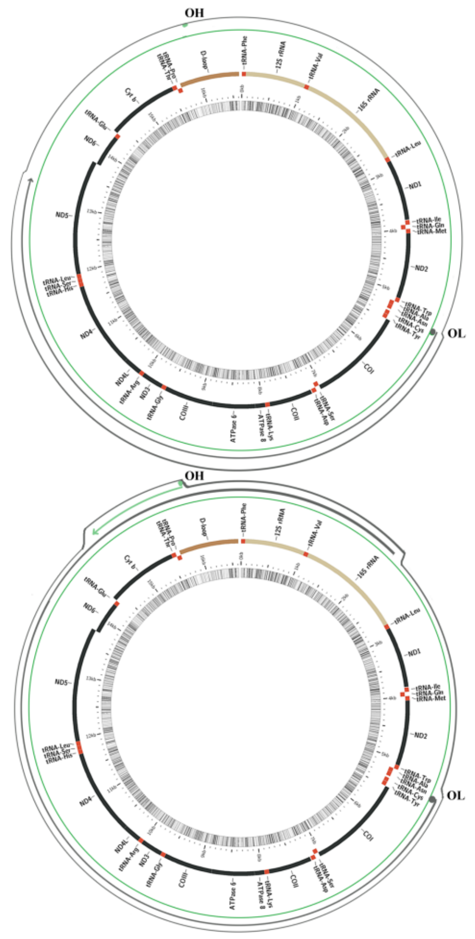
| Gene Order | tRNA Gene | H/L strand coded | Codon | Anticodon in H-strand | Max No of possible GT content | No of GT in H/L-strand tRNA anticodon site | Avilable anticodon in Vrtibrate mtDNA codon table | | | | | | | | Codon Usage | Estimated time duration of single strand exposure (bp |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | No of GT in H-strand anticodon site | | | | | | | | | |
| | | | | | | | 3 | 3 | 2 | 2 | 1 | 1 | 0 | 0 | | |
| CR | | | | | | | | | | | | | | | | |
| 1 | Phe | H | TTC | GAA | 1 | 1 | | | | | | GAA | | AAA | 235 | 1 |
| 5 | Leu1 | H | TTA | TAA | 1 | 1 | | | | | | TAA | | CAA | 368 | 2653 |
| 7 | Ile | H | ATC | GAT | 2 | 2 | | | | GAT | | AAT | | | 271 | 3686 |
| 9 | Met | H | ATG | CAT | 2 | 1 | | | | TAT | | CAT | | | 163 | 3825 |
| 11 | Trp | H | TGA | TCA | 1 | 1 | | | | | | | TCA | CCA | 119 | 4935 |
| OL | | | | | | | | | | | | | | | | |
| 18 | Asp | H | GAC | GTC | 2 | 2 | | | | GTC | | ATC | | | 75.6 | 6941 |
| 20 | Lys | H | AAA | TTT | 3 | 3 | | TTT | | CTT | | | | | 76.5 | 7718 |
| 24 | Gly | H | GGA | TCC | 1 | 1 | | | | | TCC | GCC | ACC | CCC | 243 | 9416 |
| 26 | Arg | H | CGA | TCG | 2 | 2 | | | TCG | GCG | ACG | CCG | | | 75.6 | 9828 |
| 29 | His | H | CAC | GTG | 3 | 3 | | GTG | | ATG | | | | | 104 | 11561 |
| 30 | Ser2 | H | AGC | GCT | 2 | 2 | | | | GCT | | ACT | | | 53.5 | 11630 |
| 31 | Leu2 | H | CTA | TAG | 2 | 2 | | | TAG | GAG | AAG | CAG | | | 247 | 11689 |
| 36 | Thr | H | ACA | TGT | 3 | 3 | TGT | GGT | CGT | AGT | | | | | 419 | 15311 |
| CR | | | | | | | | | | | | | | | | |
| OH | | | | | | | | | | | | | | | | |

| | | | | | | | No of GT in L-strand anticodon site | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 3 | 3 | 2 | 2 | 1 | 1 | 0 | 0 | | |
| CR | | | | | | | | | | | | | | | | |
| 3 | Val | L | GTA | TAC | 1 | 1 | | | | | TAC | GAC | AAC | CAC | 228 | 1026 |
| 8 | Gln | L | CAA | TTG | 3 | 3 | | TTG | | CTG | | | | | 95.7 | 11624 |
| 12 | Ala | L | GCA | TGC | 2 | 2 | | | TGC | GGC | | AGC | CGC | | 358 | 10369 |
| 13 | Asn | L | AAC | GTT | 3 | 3 | | GTT | | ATT | | | | | 116 | 10295 |
| OL | | | | | | | | | | | | | | | | |
| 14 | Cys | L | TGC | GCA | 1 | 1 | | | | | | GCA | | ACA | 30.1 | 10198 |
| 15 | Tyr | L | TAC | GTA | 2 | 2 | | | | GTA | | ATA | | | 113 | 10133 |
| 17 | Ser1 | L | TCA | TGA | 2 | 2 | | | TGA | GGA | CGA | AGA | | | 181 | 8508 |
| 34 | Glu | L | GAA | TTC | 3 | 3 | | TTG | | | | CTC | | | 51.2 | 1282 |
| 37 | Pro | L | CCT | AGG | 3 | 2 | | GGG | AGG | TGG | CGG | | | | 217 | 1 |
| CR | | | | | | | | | | | | | | | | |
| OH | | | | | | | | | | | | | | | | |

Polar
Charged
Non Polar

OL - Origin for L-strand DNA replication
OH - Origin for H-strand DNA replication
CR - control region

B



OL - Origin for L-strand DNA replication
OH - Origin for H-strand DNA replication

C

| | A | C | G | T |
|---|---|---|---|---|
| Genome | 27.8 | 28.9 | 18 | 25.3 |
| ATP6 | 26.2 | 30.4 | 14.3 | 29.1 |
| ATP8 | 30.7 | 29.2 | 13.2 | 26.8 |
| CO1 | 25.2 | 26.3 | 19.2 | 29.3 |
| CO2 | 28.4 | 27.4 | 17.2 | 26.9 |
| CO3 | 24.7 | 29.9 | 18.2 | 27.2 |
| CYB | 25.5 | 29.3 | 16.8 | 28.4 |
| ND1 | 24.2 | 30.9 | 17.1 | 27.7 |
| ND2 | 27.1 | 33.3 | 14.9 | 24.7 |
| ND3 | 24 | 29.6 | 16.6 | 29.9 |
| ND4 | 26.6 | 29.9 | 16.6 | 26.9 |
| ND4L | 24.4 | 31.6 | 17 | 27 |
| ND5 | 28.1 | 30 | 15.2 | 26.7 |
| ND6 L strand | 15.1 | 15.3 | 31.6 | 38 |
| **All gene** | 25.7 | 29 | 17.3 | 28 |
| 12S rRNA | 29.9 | 27.4 | 22.7 | 20 |
| 16S rRNA | 33.3 | 25.9 | 21.4 | 19.4 |
| tRNA ala | 28.5 | 15.2 | 25.8 | 30.5 |
| tRNA arg | 30.4 | 21.4 | 18.4 | 29.8 |
| atRNA sn | 19.8 | 19.9 | 30.1 | 30.1 |
| tRNA asp | 29.1 | 23.2 | 20.9 | 26.8 |
| tRNA cys | 21.7 | 22.4 | 31.5 | 24.3 |
| tRNA gln | 23.7 | 16 | 28.6 | 31.8 |
| tRNA glu | 26.2 | 13.2 | 24.3 | 36.4 |
| tRNA gly | 36.8 | 19.1 | 16.4 | 27.6 |
| tRNA his | 30.4 | 21.8 | 21.5 | 26.3 |
| tRNA ile | 23.1 | 28.8 | 28.1 | 20.1 |
| tRNA leu1 | 31.2 | 23.2 | 23.1 | 22.6 |
| tRNA leu2 | 22.9 | 30.2 | 25.9 | 21 |
| tRNA lys | 27.6 | 27 | 22.9 | 22.5 |
| tRNA met | 29 | 26.8 | 17.6 | 26.5 |
| tRNA phe | 35.1 | 21.6 | 23.8 | 19.6 |
| tRNA pro | 25.9 | 11.6 | 27.1 | 35.4 |
| tRNA ser1 | 20.9 | 22.5 | 28.4 | 28.1 |
| tRNA ser2 | 25.5 | 27.5 | 24.8 | 22.2 |
| tRNA thr | 24.6 | 27.8 | 25.8 | 21.8 |
| tRNA trp | 30.8 | 24.5 | 24 | 20.7 |
| tRNA tyr | 21.1 | 21.7 | 31.5 | 25.7 |
| tRNA val | 26.6 | 26.9 | 25.8 | 20.6 |

Color key

38 11.6

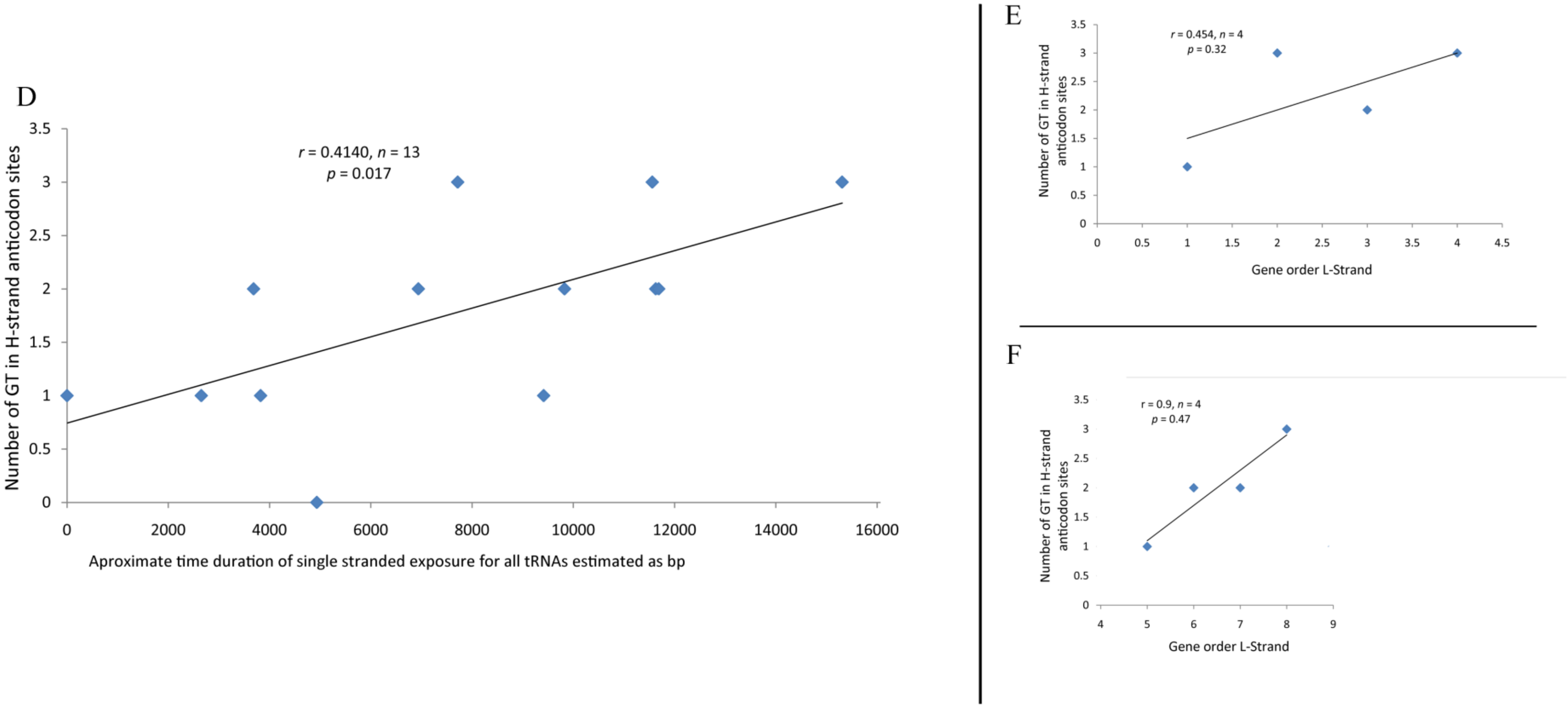


**Fig. 6.6** (A) tRNA genes, its codon, anticodon and order of distribution along H and L strand of Clupeoid mtDNA, (B) Schematic diagram of the mtDNA replication based on the displacement-model of replication, (C) Percentage of A, T, G and C of Clupeoid fishes mitogenome, protein-coding genes, merged protein-coding gene, 12S rRNA, 16S rRNA and tRNAs, (D) Correlation between the number of G and T in the anticodon position of tRNA loci and expected time duration of single-strand exposure during mitochondrial replication, (E) Correlation between the number of G and T in the anticodon position of tRNA loci between $O_H$-$O_L$ and gene order in L-strand, (F) Correlation between the number of G and T in the anticodon position of tRNA loci between $O_L$-$O_H$ and gene order in L-strand.

The neutrality plot (GC12 vs GC3) (R-value is 0.69) indicating GC12 and GC3 is following mutation bias model, there is a moderate correlation between GC12 and GC3 and mutation bias plays a predominant role in shaping the codon usage bias. The effective number of codon (ENc) has been used as a measure of codon usage bias in genes (Wright 1990). Similar to RSCU results, the ENc ranged from 46.4 to 58.1 (which is lower in freshwater lineages, except Engraulide and Tenualosa), indicating a high codon usage bias in the Clupeoides genome (Table 6.S1). ENc plot with concatenated gene data set showed most of the values were above and close to expected ENc plot curve (Fig. 6.7). The standard curve represents the functional relationship between ENc and GC3 under mutation and selection pressure. If the codon usage bias is completely based on mutation bias (GC3 content) all the points will be on the standard curve.

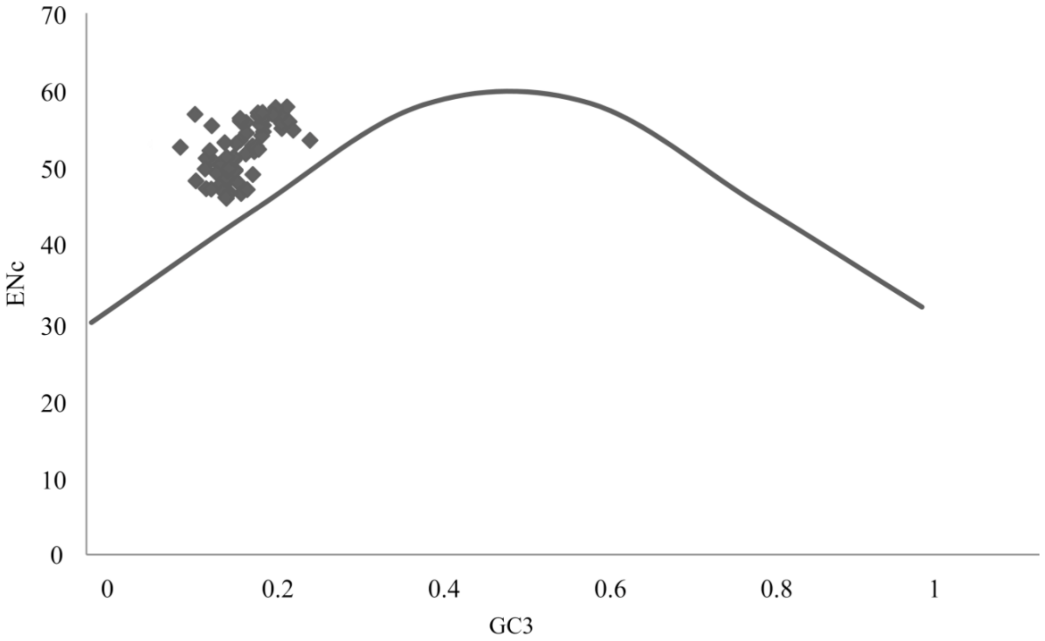


**Fig. 6.7 Relation between ENc and GC3s of Clupeid mitogenomes.** ENc analysis of merged protein-coding genes of all considered clupeoid fishes plotted against GC3s (ENc plot). The expected Enc from GC3 under mutational pressure without selection is shown as a standard curve.

### 3.4. Selective constraints on the secondary structure of the mtDNA control region and transfer RNA (tRNA) genes

#### 3.4.1. Structure and content of mtDNA control region

The control region sequence displayed large length variation, the mean length of all clupeoid fishes analyzed was 953bp. Different conserved sequence regions like the Conserved Sequence Box, CSBs (CSB D, CSB 1, CSB 2 and CSB 3) were identified in all species analyzed (Fig. 6.8) and its relative position was similar to those reported invertebrates and fishes (Sebastian *et al*. 2017). Among the four conserved sequence regions CSB1, CSB 2 and CSB 3 are highly conserved whereas the TAS sequence identified has a high number of polymorphic sites among the clupeoids. The T-homopolymer of more than nine nucleotides was observed between CSB D and CSB 1, T-homopolymer of less than five and a C-homopolymer of less than six were found in all the clupeoid fish species. The base composition of each CSB is unique as follows; CSB D is T rich (41.0%), CSB 3 is AT-rich (A 39.9%, T 27.4%), CSB 2 is C rich (64.3%) and CSB 3 is rich with AC (A 46.7%, C 33.3%) (Table 6.1). The CSBs, TAS and Poly T being highly conserved among clupeoids and the regions between these conserved regions and repeat units are polymorphic among species. The sequences spanning CSBs are characterized by the presence of a high degree of AC and GT/TG repeats. The high A (31.9%) and T (31.4%) content in the control region is also reflecting the presence of high AT repeats. But the interesting thing was that the CSBs in the control region were free from the presence of the above-mentioned repeats.

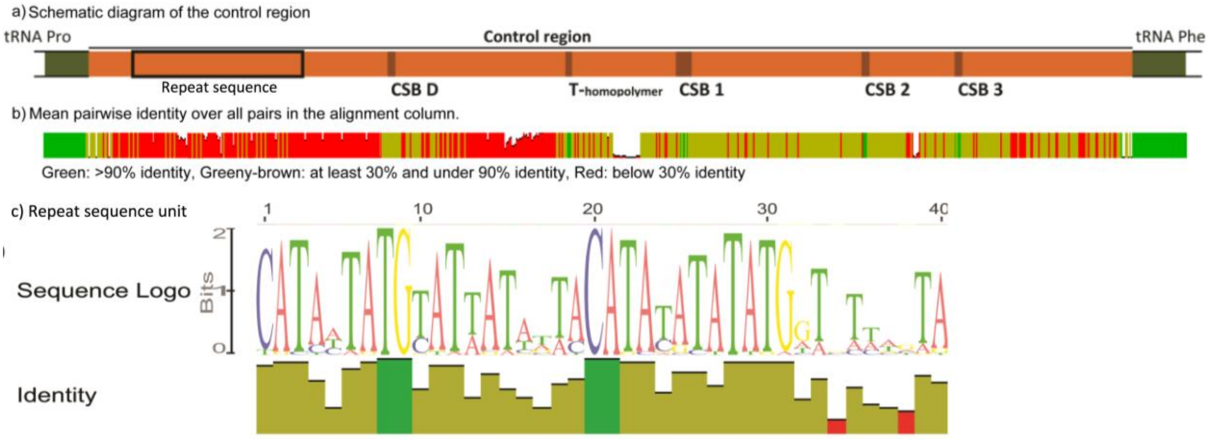

a) Schematic diagram of the control region
tRNA Pro
Control region
tRNA Phe
Repeat sequence
CSB D
T-homopolymer
CSB 1
CSB 2
CSB 3
b) Mean pairwise identity over all pairs in the alignment column.
Green: >90% identity, Greeny-brown: at least 30% and under 90% identity, Red: below 30% identity
c) Repeat sequence unit
1
10
20
30
40
Sequence Logo
Bits
2
1
0
Identity

**Fig. 6.8** Schematic diagram of the control region of the clupeoid fish mitogenome a) Locations of conserved sequence block domains and T-homopolymers of variable regions are mapped b) mean pairwise identity between control region sequences used for analysis c) Sequence log representation of the control region repeat sequence unit/ motif in clupeoids.

**Table 6.1** Features of the four CSBs of the clupeoid fishes

| Base composition | A (%) | C (%) | G (%) | T (%) |
|---|---|---|---|---|
| CSB D | 7.6 | 26.3 | 25.1 | 41.0 |
| CSB 1 | 39.9 | 17.5 | 14.9 | 27.4 |
| CSB 2 | 23.5 | 64.3 | 00.0 | 12 |
| CSB 3 | 46.7 | 33.3 | 6.7 | 13.3 |

The sequences of the clupeoid mtDNA control region were characterized by the presence of an imperfect repeat unit in its highly variable region with palindromic sequences within it (Fig. 6.8c). The repeat unit sequence was ~38-40bp in length and the number of repeat units varied among species. The observed repeat unit sequence was identical among species and they are variants of a common sequence with additions, deletions, and substitutions in some regions. The regions between these conserved regions and repeat units are highly polymorphic among species.

Several secondary structures with more than 10bp paired bases with varying lengths were identified in the mtDNA H and L-strand (Fig. 6.9). The conserved sequences like TAS and CSBs are associated with a secondary structure (Fig. 6.S2, Fig. 6.S3). All secondary structures predicted for the mtDNA L-strand were also observed in the L-strand mRNA transcript with some minor changes. Few large and short stem-loop structures with internal bulges (Fig. 6.S2) and having low free energy (-0.65 to -215.62 ΔG (kcal/mol)) were observed in the repeat region (Table 6.S3). The L strand mRNA transcript of the repeat region is also forming similar structures with greater negative folding energies (ΔG).

**Table 6.2** Folding energy (ΔG) and Normalized free energy -ΔG(kcal/mol)/ Length(bp) for 13 *S. longiceps* mitochondrial tRNA genes and its comparison with predicted secondary structures of highly conserved sequence blocks of the clupeoids mitochondrial control region.

| tRNA (DNA) | ΔG(kcal/mol) | Length(bp) | Normalized free energy - ΔG(kcal/mol)/ Length(bp) |
|---|---|---|---|
| tRNA-Ala | -10.77 | 69 | -0.16 |
| tRNA-Arg | -16.3 | 69 | -0.24 |
| tRNA-Asp | -10.37 | 69 | -0.15 |
| tRNA-Gly | -20.3 | 71 | -0.29 |
| tRNA-His | -14.6 | 69 | -0.21 |
| tRNA-Ile | -30.31 | 72 | -0.42 |
| tRNA-Leu | -27.4 | 72 | -0.38 |
| tRNA-Phe | -12.34 | 63 | -0.20 |
| tRNA-Pro | -17.1 | 70 | -0.24 |
| tRNA-Ser | -11.31 | 67 | -0.17 |
| tRNA-Trp | -9.07 | 70 | -0.13 |
| tRNA-Tyr | -15.96 | 71 | -0.23 |
| tRNA-Val | -20.3 | 72 | -0.28 |
| Species name | Average ΔG(kcal/mol) | Average Length(bp) | Normalized free energy - ΔG(kcal/mol)/ Length(bp) |
| CSB 3 | -11.1 | 41 | -0.27 |
| CSB 2 | -6.29 | 32 | -0.20 |
| CSB 1 | -0.26.32 | 95 | -0.28 |
| CSB D | -0.24 | 101 | -0.24 |
| TAS | -14.72 | 64 | -0.23 |

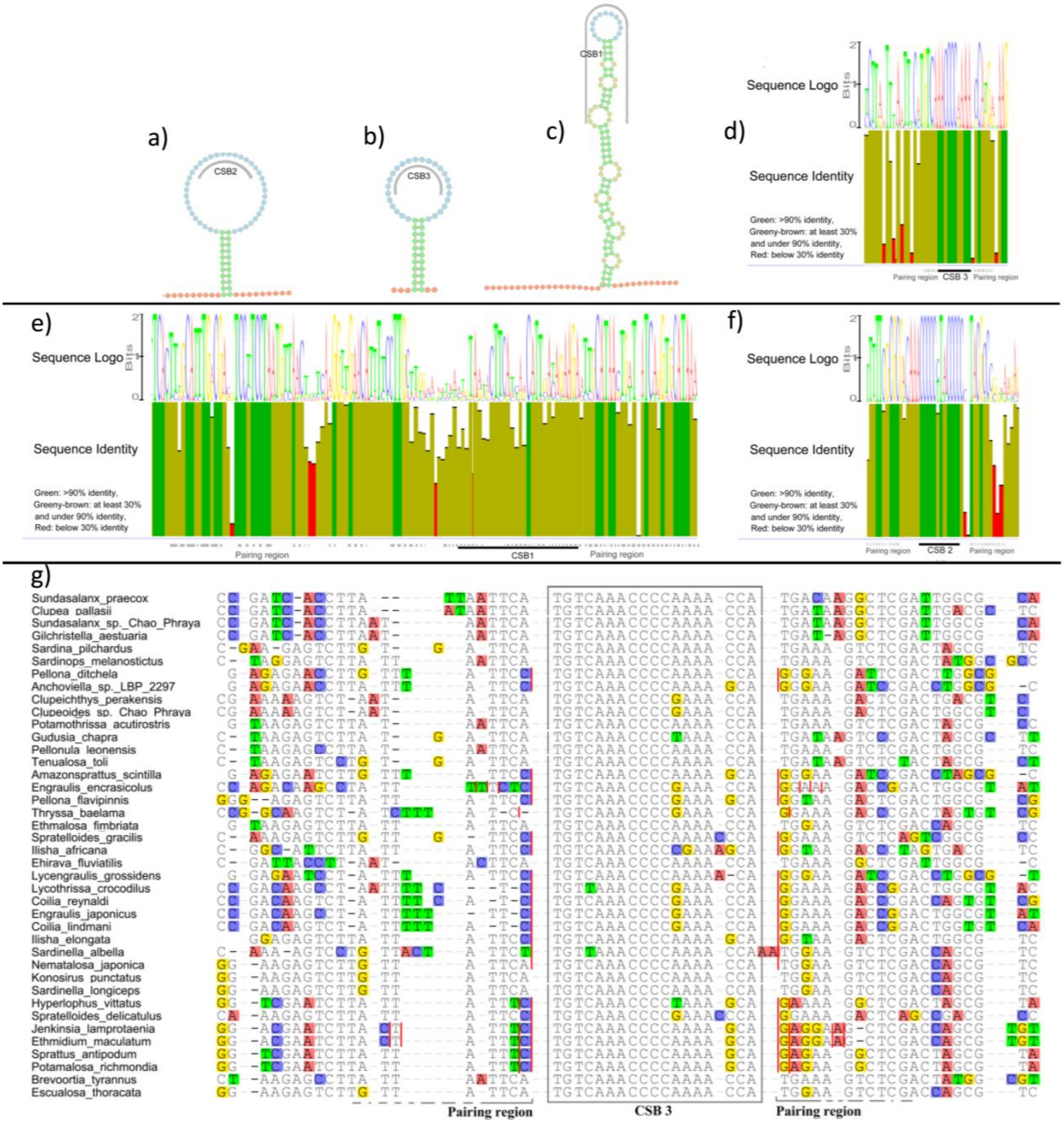


**Fig. 6.9** Potential secondary structure identified a) in CSB2, b) CSB3 and c) CSB1of Clupeoids mtDNA control region. Sequence log representation of d) CSB3, e) CSB1 and f) CSB2 with the pairing flanking sequences. g) Multiple sequence alignment of clupeoid CSB3 and pairing flanking regions. Complementary mutations in its pairing region are marked with vertical red lines.

To assess the robustness of the secondary structure predicted, its folding potential (Free energy, ΔG) was compared with the tRNA, which is known to form a functional secondary structure. The relative free energy of *S. longiceps* (ΔG/Length) tRNA ranged from - 0.42 to -0.08 and the predicted secondary structure of the clupeoids ranged from -0.3 to -0.16 (Table 6.2; Table 6.S2; Table 6.S3). This indicates the higher folding potential of the control regions in the clupeoids mitogenome.

### 6.4.2. Selective constraints on secondary structure

The Tajima's D statistics for Cytochrome b, tRNA-thr, tRNA Pro, control region and tRNA-phe with a length of 3396 bp indicated that the value is negative for all tRNA and most of the regions on cytochrome b. The control region also contained DNA stretches with significant negative Tajima's D, especially the regions locating TAS, repeat sequences and CSBs (Fig. 6.S5).

In both the tRNA and control region sequences, the pairing region (sites paired during secondary structure-formation) was characterized by high inter-specific identity, whereas the regions flanking them are highly polymorphic. A high rate of complementary/compensatory mutation in the stem forming regions in tRNA (Fig. 6.S6) and flanking sequences of CSBs (Fig. 6.9; Fig. 6.S4) was also observed.

### 6.4.3. Structure and content Transfer RNA (tRNA) genes

The length of tRNAs in clupeids ranged from 69-76bp and tRNA sequences folded into a secondary structure similar to the traditional cloverleaf structure. It is composed of four domains, amino acid acceptor (AA) stem, dihydrouridine (D) arm (D-stem + D loop), anticodon (AC) arm (AC stem + AC loop) and thymidine (T) arm (T-stem + T loop). The length of the AC loop, D-stem and T-stem were fixed for each tRNA gene among species, whereas the length of D-loop and T-loop varied in length (mostly in D-loop with 1-3bp) (Fig. 6.S6). Even though the length is fixed among species for each tRNA, high variability was observed between tRNA within and among species. The length of the variable loop is fixed among species for each tRNA and its length ranged from 1-5bp. We observed a very high degree of complementary/complementary mutations in the potential base-pairing region (stem forming region) of tRNA, indicating the action of a force which removes the substitutions that destabilize the functional three-dimensional structure of tRNA (Fig. 6.S6). tRNA-Leu (UUA) has the mitochondrial transcriptional termination factor (mTERF) binding site, in the same region that has been reported for the mammalian mitochondrial genome (5’-TGGCAGAGCCCGG-3’), corresponding to the D-arm (Hyvarinen *et al*. 2007). In most of the clupeid species analyzed, the sequence is 100% identical to the human tridecamer sequence; particularly the 11 out of 13 bases are identical in all the species. In some species, C to T and C to A substitutions were observed in 3’ end of the motif (Fig. 6.S7).

## 6.5 Positive selection

The result from TreeSAAP indicated that several significant physiochemical amino acid changes have occurred with changes among amino acid residues in mitochondrial protein-coding sites. Negative selection dominates in both conservative/moderate (category 1, 2 and 3) and radical changes (category 6, 7 and 8) (total properties 23674(category 1,2,3 +) & 27737(category 1,2,3 -) and 1751 (category 6,7,8 +) & 1964 (category 6,7,8 -)). The proteins ND6, ND2 and ND4 have the highest average number of positive radical amino acid modification (0.92, 0.70 and 0.058 average changes per sites respectively) whereas CO3, CYTB, ATP8 and CO1 have the lowest (0.015, 0.015, 0.013 and 0.008 average changes per site respectively) (Fig. 6.10). There are several positive radical amino acid modifications in the terminal branches/tips than in the interior branches (total number of physiochemical amino acid changes 1034 and 755 respectively at terminal branches/tips and interior branches). In the interior branches, the highest number of radical amino acid changes are in those leading to lineage *Tenualosa* (in lineage 1) (node 80 to 93, 66 (6,7,8 +) and 68 (6,7,8 -) across all proteins), to the lineage 5 (node 74 to 121; 42 (6,7,8 +) and 40 (6,7,8 -) changes across all proteins), to the lineage Pristigasteridae (node 72 to 123; 28 (6,7,8 +) and 68 (6,7,8 -) changes across all proteins), to Engraulidae (node 71 to 126; 33 (6,7,8 +) and 68 (6,7,8 -) changes across all proteins) to lineage 2 (node 77 to 100; 14 (6,7,8 +) and 15 (6,7,8 -) across all proteins) and the branch uniting all clupeoids except the Engraulids (node 71 to 72; 26 (6,7,8 +) and 34 (6,7,8 -) changes across all genes) (Fig. 6.11). The lineage converging temperate water clupeoids in lineage 2 and 4 has a relatively high number of radical amino acid changes (77 to 100, 101, 102; 75 to 114). Similarly, lineage converging at the marine to freshwater transition also showed a high number of amino acid property changes (Fig. 6.11, 6.12). In addition to that, the amino acid residue that has been reported to participate in key functions is not overlapping with sites under radical changes.

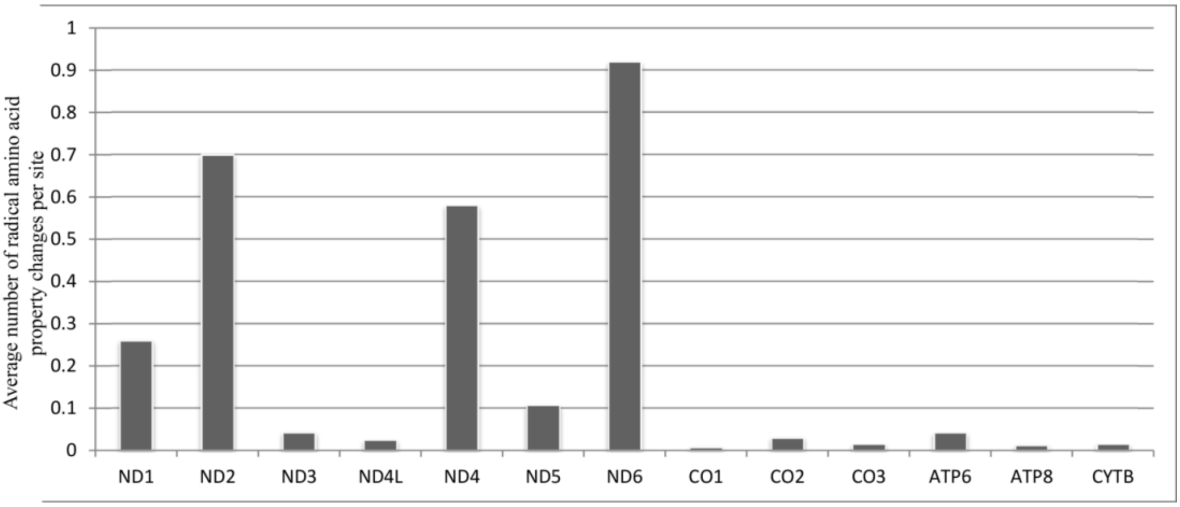


**Fig. 6.10** Radical physicochemical amino acid changes among clupeoid fishes mitochondrial protein-coding genes. An average number of strong positively selected amino acid properties in the 13 mtDNA protein-coding genes for oxidative phosphorylation in the mitogenomes of clupeoid fishes.

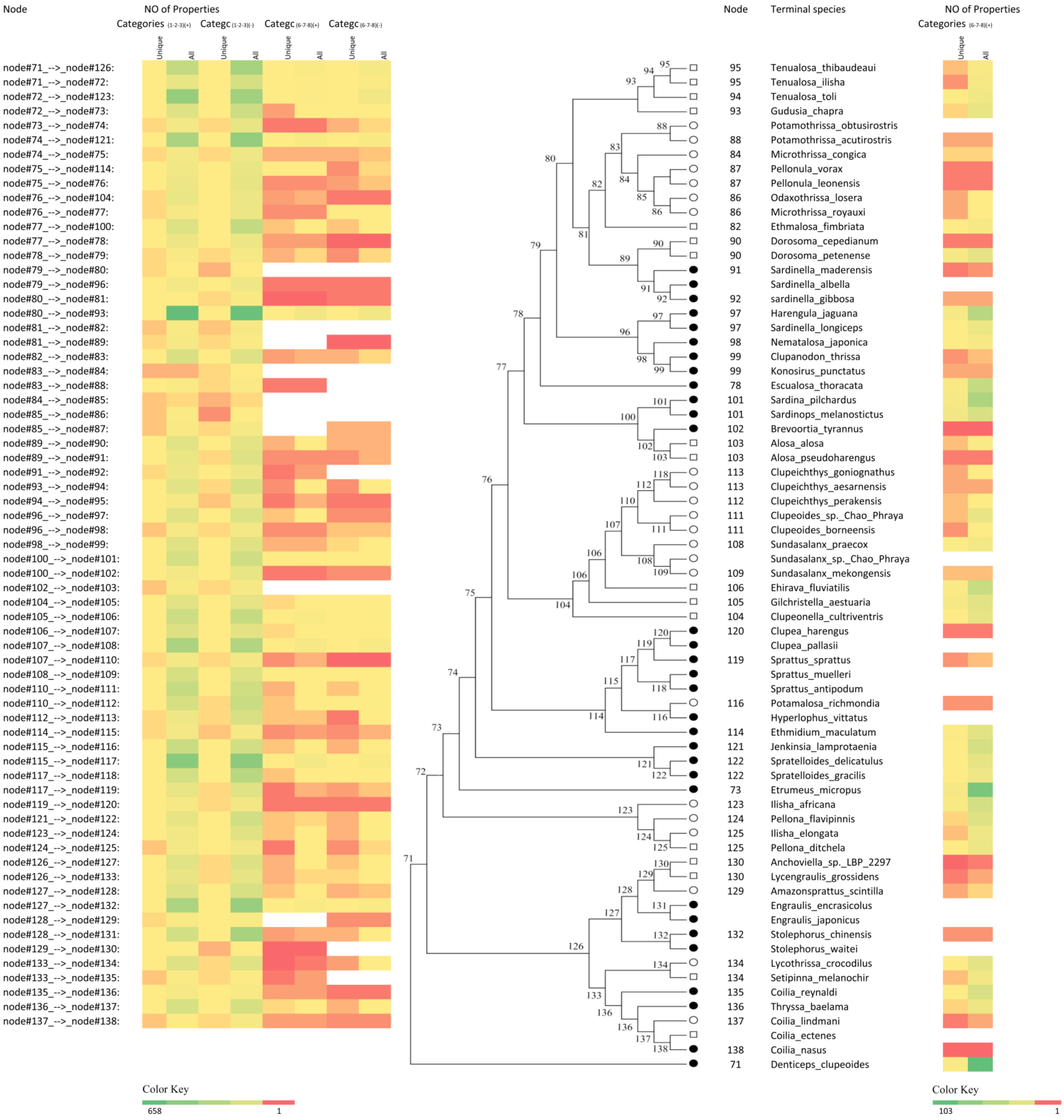


**Fig. 6.11 Physicochemical amino acid changes varying across the clupeoid mitogenomic phylogenetic tree.** Representation of the number of amino acid conservative changes corresponding to conservative categories 1,2 & 3 and radical changes to categories 6, 7 & 8 ($p < 0.001$) across mitochondrial protein-coding genes varying within the branches of the clupeoid mitogenomic phylogenetic tree. Number in the node of the phylogenetic tree indicates node-numbers. Black circle, white circles and square in the tree indicates marine, brackish and freshwater species respectively.

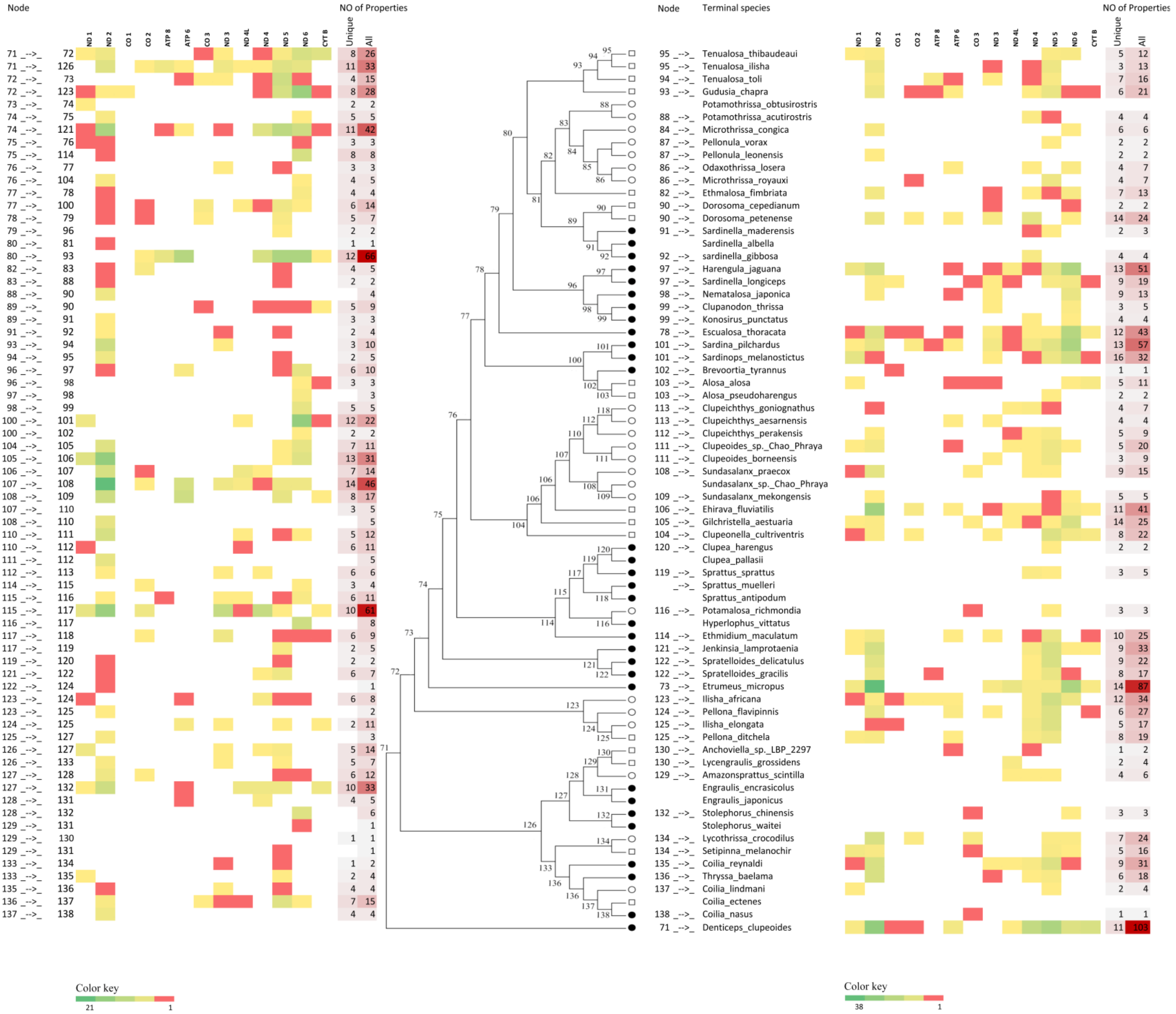


**Fig. 6.12 Radical physicochemical amino acid changes of mitochondrial protein-coding genes varying across the clupeoid mitogenomic phylogenetic tree.** Representation of the number of radical amino acid changes in corresponding categories 6, 7 & 8 ($p < 0.001$) in mitochondrial protein-coding genes varying within the branches of the clupeoid mitogenomic phylogenetic tree. Number in the nodes of the phylogenetic tree indicates node-numbers. Black circle, white circles and square in the tree indicates marine, brackish and freshwater species respectively.

Signatures of positive selection are less prevalent than purifying selection and they were located in complex 1(ND1, ND 2, ND3, ND4, ND4L, ND5 and), complex 2 (CYTB), complex 4 (CO1, CO2 and CO3) and complex 5 (ATP 6). The results of MEME showed 25 positions undergoing episodic diversifying selection (p<0.1), FUBAR analysis found two positions as diversifying selection, FEL and REL identify two positively selected amino acid positions and SLAC method detected one site. In complex I, 6 genes (out of 7 mtDNA encoded sites) were observed to undergo positive selection in MEME analysis mainly; ND1 (site-182), ND2 (site-23, site-86, site- 237, site-325 and site- 343), ND3 (site-6), ND4 (site- 52, site-98 and site-178), ND4L (site-92), and ND5 (site-32, site-189, site- 409, site-538 and site-566). In complex 2 (cytochrome b) MEME detected positive selection at site-379. In complex 4, sites undergoing positive selection were in CO1 (site-21, 133, 187 and 338), CO2 (site-9, 44, 221, 227 and 230) and CO3 (site-47). One of the five sites were identified by all four methods used (site-9 in CO2) and two sites by two methods (MEME and REL) (site-221 and 230 in CO2). In complex 5, 3 amino acid sites (36, 62 and 124) of ATP 6 have been identified by MEME as undergoing positive selection.

Mitochondrial complex I (NADH: ubiquinone oxidoreductase) contributes to cellular energy production by transferring electrons from NADH to ubiquinone coupled to proton translocation across the membrane. The Key polar amino acid residues which have been reported to participate in proton translocation (ND1 - E198, E149, ND2 - K263, K135, K105, E34, ND4 - E124, K238, E379, K208, ND5 - E149, H253, K397, K228) (Zhu *et al*. 2016) through complex I was conserved across all species. Most of the sites that exhibited signatures of positive selection in complex I were restricted to the predicted internal-helix loop region (ND2 site-23, 86, 237; ND4 site-52,178) of their respective proteins (Fig. 6.13). However 6 sites were located in transmembrane helix (ND1 site-182; ND2 site-325, 343; ND3 site-6; ND4L site-92; ND4 site- 98; ND5 site- 189, 538) and one in beta-sheet (ND5 site-32). Sixteen of these sites, three in ND2 (site-23, 86, 237), one in ND4 (site-178) & one in ND5 (site- 189) were located in Proton-conducting membrane transporter (Conserved Protein Domain Family - Proton_antipo_M), one in ND2 (site-325) located in NADH dehydrogenase subunit 2 C-terminus (Conserved Protein Domain Family - NADH_dehy_S2_C), two in ND4 (site- 52, site-98) located in NADH-ubiquinone oxidoreductase chain 4, amino terminus (Conserved Protein Domain Family - Oxidored_q5_N), and one in ND5 (site- 538) clustered in NADH dehydrogenase subunit 5 C-terminus (Conserved Protein Domain Family - NADH5_C) (Fig. 6.13). Majority of amino acid sites that have been suggested to participate in Qo binding, Qi binding and chemical binding were conserved in CYTB (complex 2) (Crofts 2004a, b). In CYTB and ATP6 (complex 5) all sites of positive selection were located in the intra-helix loop.

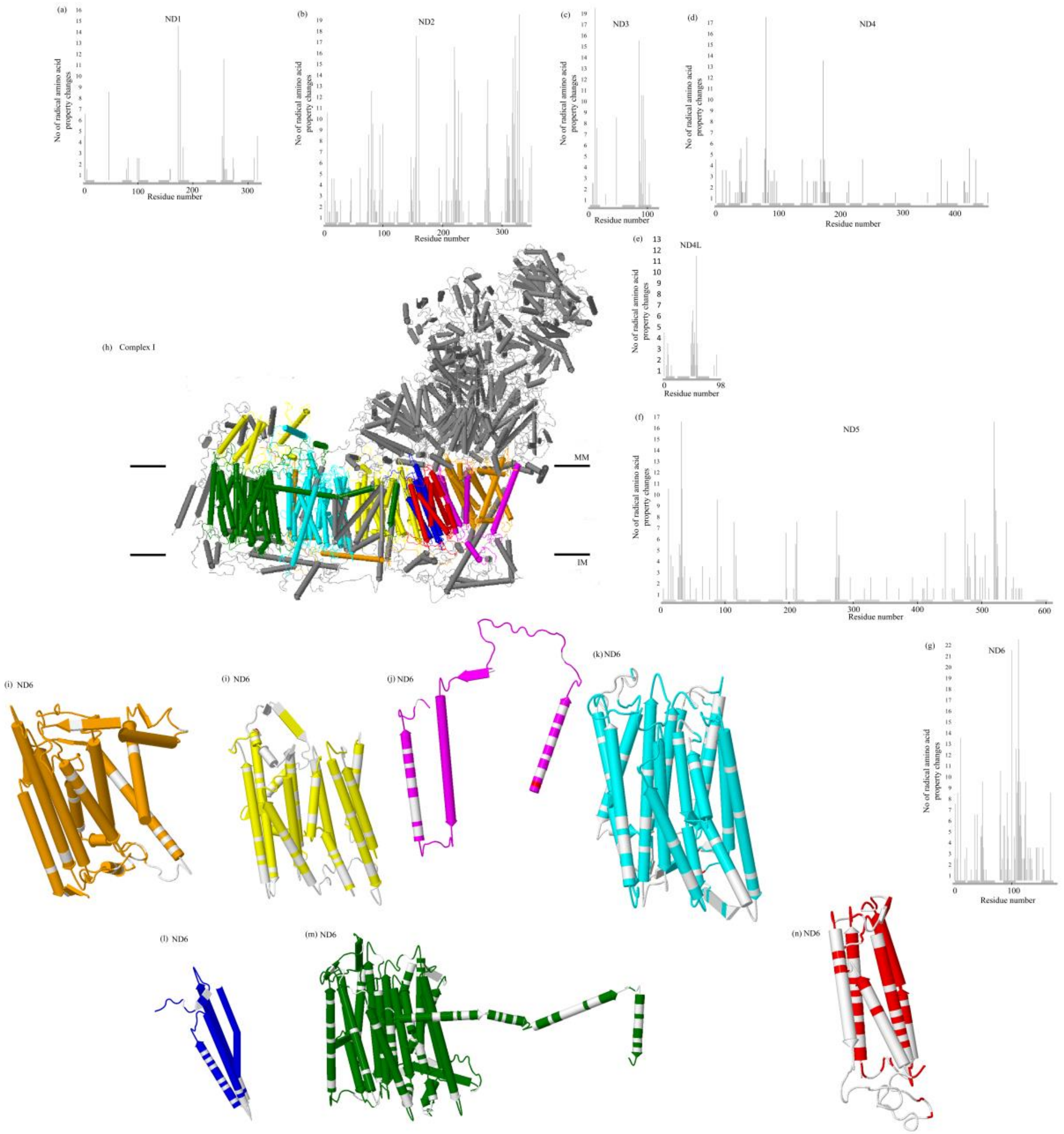


**Fig. 6.13 Amino acid property variation in dehydrogenase (Complex I).** (a) to (g) topological assignment of the sites that has radical amino acid changes under positive destabilising selection in seven subunits of Complex I. Y-axis is the number of radical amino acid changes, X-axis is residue numbers and predicted alpha-helix region is shown in grey. (h) individual OXPHOS Complex I, with mitochondrial-encoded subunits are represented in different coloured as followed: ND2 in yellow; ND4L in blue; ND1 in orange; ND3 in magenta; ND4 in cyan; ND5 in green; ND6 in red. Grey structures represent nuclear-encoded subunits. Individual core subunits (h) ND1, (i) ND2, (g) ND3, (k) ND4, (l) ND4L, (m) ND5 &(n) ND6 with white colour on positively selected amino acid sites.

Cytochrome c oxidase (CcO) (complex IV) is considered as one of the major regulation sites for oxidative phosphorylation and it catalyzes the final step in mitochondrial electron transfer chain by receiving an electron from each of four cytochrome c molecules, transfers them to one oxygen molecule and also translocates four protons across the membrane (Li *et al*. 2006). Sites in complex IV occurred in intra-helix loop (CO1 site-133; CO2 site-227,230), transmembrane helix (CO1 site-21, 187, 338; CO2 site-44, 221; CO3 site-47) and beta-sheet (CO2 site-9) (Fig. 6.14). The amino acid residues that have been reported to participate in Electron transfer pathway (F377, R438, R439), D-pathway (Y19, N80, D91, N98, S101, S156, S157, N163, T167), Putative water exit pathway (D227, G232, H233, D364, H368, D369, R438), Ion binding (Binuclear center-heme a3/CuB) (H240, H290, H291, H376), K-pathway (H240, Y244, S255, H290, H291, T316, K319), Putative proton exit pathway (H291, H368, D369, R438, R439), and chemical binding (Low-spin heme a binding site) (H61, H378, S382, T424, S461) in CO1 (Tsukihara *et al*. 1995; Tsukihara *et al*. 1996) were conserved. In addition to that, amino acid residues that have been reported to participate in CuA binding site in CO2 and the amino acid participated in polypeptide binding (in the subunit interface) and Phospholipid binding in CcO is also conserved across all species in this study.

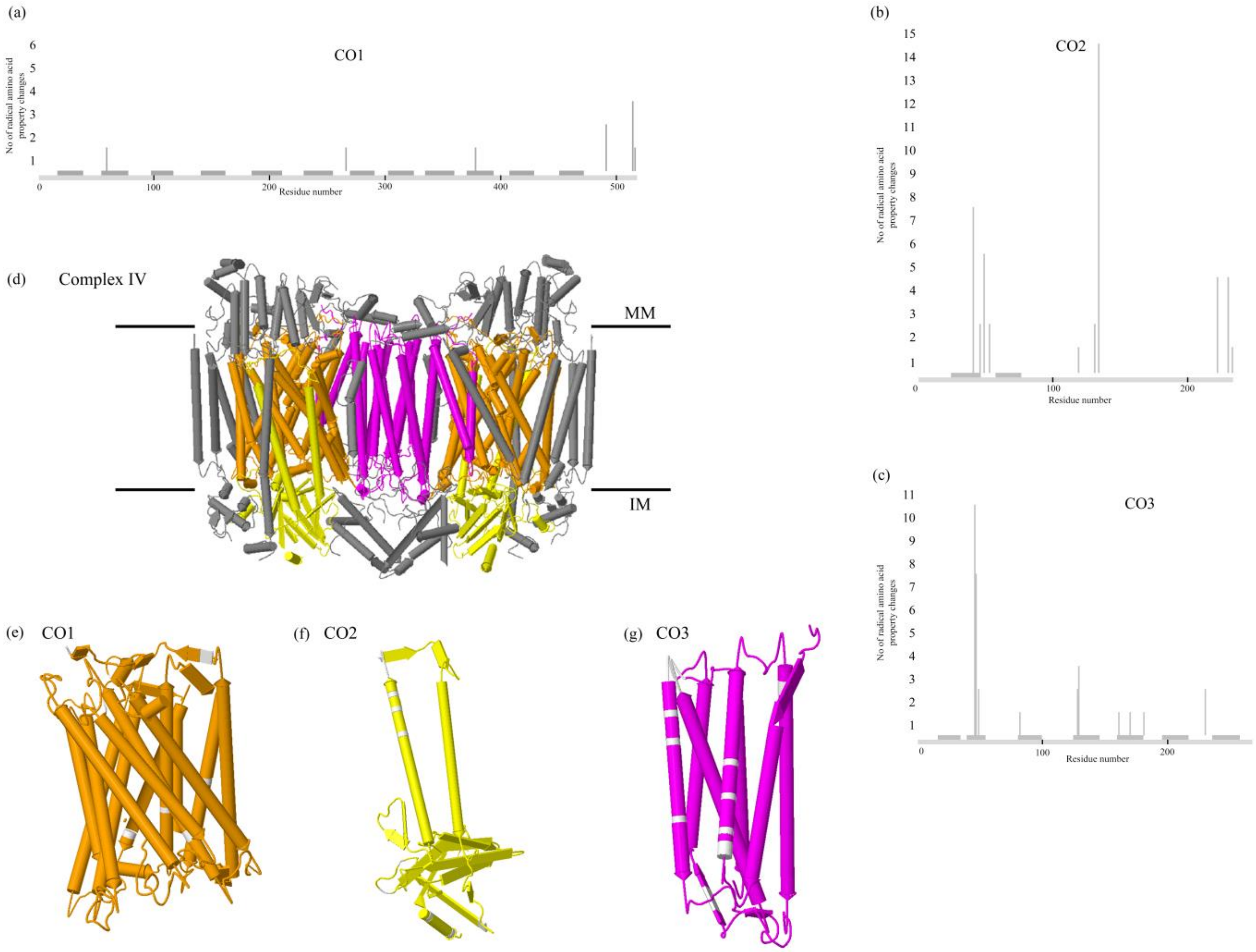


**Fig. 6.14 Amino acid property variation in Cytochrome C Oxidase (Complex IV).** (a) to (c) topological assignment of the sites that has radical amino acid changes under positive destabilising selection in three subunits of Complex IV. Y-axis is the number of radical amino acid changes, X-axis is residue numbers and predicted alpha-helix region is shown in grey. (d) Individual OXPHOS Complex IV (Homodimer) with mitochondrial-encoded subunits is represented in different colours as followed: CO1 in orange; CO2 in yellow; CO3 in magenta. Grey structures represent nuclear-encoded subunits. Individual core subunits (e) CO1, (f) CO2 and (g) CO3 with white colour on positively selected amino acid sites.

The selection analysis indicated that codon sites are under positive selection in 11 genes: CO1, CO2, CO3, ND1, ND2, ND3, ND4, ND5, ND4L, CYTB and ATP6 (except ATP8). But further analysis showed that among this only CO2 (site-44) has a fixed difference between freshwater, marine and euryhaline species among the available mitogenomes (Fig. 6.S9b). Similarly, we identified amino acid changes in ND4, ND5 and ND6 which is specific to temperate water species in lineage 2 and 4 (Appendix Fig. A4).

## 4. Discussion

This study provided evidence for the ability to form stable secondary structures by sequences flanking the conserved sequence elements and evolutionary force (complementary/compensatory mutations in the stem region) maintaining the secondary structures. This indicates the importance of the stable secondary structures in the mitochondrial DNA function and evolution like in the tRNA genes. We also identified sites that are putatively under positive selection in the OXPHOS complex of distantly related clupeoid species distributed from temperate to tropic and marine to the freshwater environment by positive selection test and homology protein modelling using complete mitochondrial genome sequences. Not only amino acids but also the base composition and codon usage bias seems to be shaped by mutational bias and natural selection forces.

### 4.1. The evolution rate of genes

In the analysed clupeoids fishes, mitochondrial genes have evolved at a different rate, ND6 has high substitution rate than other protein-coding genes, along with CYTB and ND1. In clupeoids mitogenome, the second codon position evolved slower than first and third codon position which follows the patterns of the natural theory of molecular evolution (Kimura 1983). Synonymous sites in the protein-coding regions are evolving faster than non-synonymous sites. Some mutations in the first codon position and most in the third codon position are synonymous. Most of the substitutions in second codon position are non-synonymous and thus they should be under purifying selection (Kimura 1983). First and second codon positions vary in populations and chances of them getting fixed in populations are high. In vertebrates, mtDNA genes evolve with the order D-loop > CDS > rRNA > tRNA (Jeffrey 1999; Howell *et al.* 2007). The non-coding region is under least selective pressure similar to third codon position thus evolving faster than other genes. Generally, genes with different functions have different structural and functional constraints so they evolve at a different rate (Wall *et al.* 2005). Hence the complex interactions between dynamic environmental factors and the ability of functional genes to cope with that (examples expression rate, structural stability etc.) will be the systemic determinants of gene evolution (Koonin 2005). The high rate of mutations observed in ND and CYTB may be also associated with the position of the ND genes in the mitogenome. They are found immediately upstream from the origin of L-strand replication (OriL) and immediately downstream from the origin of H-strand (OriH) replication. During replication these genes stay single-stranded for more time compared to other genes, thus they are prone to a high rate of mutation (Marshall *et al.* 2008).

### 4.2. Selective constraints in the mitochondrial tRNA and control region of Clupeoid fishes

We confirmed the ability of conserved sequence elements in the control region to form stable secondary structures similar to the tRNA using comparative mitogenomics of 70 Clupeoids. Similar to tRNA, a selective constraint is acting to maintain this secondary structure as evidenced by low mutation rate and compensatory mutations in the stem forming regions. The presence of discontinuous AT and CG repeats in the flanking regions of conserved sequence motifs promotes the tendency to form secondary structures and stabilize the structure, while the unique sequence composition (absence of these repeats) in conserved sequence motifs helps them to maintain the loops in the stem-loop secondary structure. Thus the control region sequence elements may be functioning through the secondary structure formed by it similar to the tRNAs, and selective pressure is acting on the sequence elements forming secondary or tertiary structure (Chen *et al.* 1999). Even though secondary structures associated with conserved sequences in the mtDNA control region have been predicted (Lee *et al.* 1995), the evidence for its formation has been confirmed from this study by the detection of compensatory/compensatory base substitutions in many species. Accumulation of higher levels of mutations has been proposed during the enzymatic replication of DNA sequences containing repeat units and those sequences having the ability to form secondary structures (Lee *et al.* 1995; Broughton and Dowling 1997). The tandem repeats present in the control region originated from the repeat sequences involved in secondary structures associated with conserved sequence elements during replication. The high polymorphism in the control region may be due to the result of errors that occurred during the enzymatic replication.

Generally, heterogeneity in the protein-coding regions can be explained as due to selective forces, whereas in non-coding regions it can only be explained by the structural or other functional role played by the DNA molecule itself (Wright 2000). The control region exhibits a relatively high mutation rate due to reduced functional constraints whereas the low mutation rate in protein-coding regions, tRNAs and rRNAs may be due to strict purifying selection (Jacobsen *et al*. 2016). The function of conserved sequence sites in the control region has been explained as binding sites for regulatory proteins. (Taanman 1999; Melo-Ferreira *et al*. 2014; Nicholls and Minczuk 2014). DNA secondary structures such as hairpins have been identified as recognition sites for the binding of several proteins involved in the direct interaction with DNA and RNA (Walberg and Clayton 1981; Katz and Burge 2003; Pereira *et al*. 2008). We propose that the conserved sequence in the mitochondrial control region function as a recognition site of proteins/enzymes, by forming secondary structure (stem and loop structure). The loop is occupied by the conserved domains like CSBs and the stem is formed by the flanking sequences, which can pair each other. In addition to this, the characteristic base composition of CSBs avoiding substitution of A and T, protects them from pairing with the flanking repeat sequences consequently forming a loop. There are clear pieces of evidence that the basic molecular processes like replication, transcription, and recombination are controlled/regulated by formation of intra-strand secondary structures by nucleic acids (DNA/RNA) and their interaction with proteins (Pereira *et al*. 2008; Rice and Correll 2008; Spies and Smith 2017) and many control region segments can form stable intra-sequence secondary structures (Lee *et al*. 1995; Katz and Burge 2003; Pereira *et al*. 2008).

Secondary structure formation ability of mtDNA control region sequence elements is evident from the presence of pairing sequences in the regions flanking the conserved sequence motifs (CSBs) similar to the tRNA's stem-loop structure forming regions (Chen *et al*. 1999). The presence of a high percentage of discontinuous AC and TG repeats (present in all the Clupeoids) in the regions flanking CSBs also supports its high folding potential/ability to form potential secondary structures. Folding energy (ΔG) of the predicted secondary structure associated with conserved sequence motifs is comparable with the free energy of the tRNA structures, indicating the stability of stem-loop structures. Thus the primary role of sequences flanking CSBs will be to support the formation of a stem-loop structure so that an enzyme/protein can easily access the conserved sequence in CSBs. The sequence conservation between species indicates its substantial historical stability. Ths the conformation predicted for the structure in the control region is maintained during evolution or diversification of species.

We propose that a selection force is acting at an intra-mitochondrial or inter-cellular level against the mutations which break the secondary structure involved in the efficient regulation of mtDNA functions. The presence of the compensatory mutation in the stem forming flanking sequence of CSBs as in the tRNA indicates that there is a strong pressure acting to maintain the stability of the secondary structure (Lee *et al*. 1995; Chen *et al*. 1999; Pereira *et al*. 2008). Thus the sequences in the secondary structure-forming regions are protected from mutations that break the structure necessary for efficient regulation of mtDNA functions. The Tajima‘s D value is zero/negative and significant for most of the coding and tRNA regions, as expected for functionally constrained regions (Tajima 1989). Similar results of negative Tajima’s D values were also recorded for the mitochondrial control region sequences, CSBs and predicted secondary structure-forming regions flanking them indicating that these regions in the control region are also under negative selection similar to the coding region. Besides, the significant difference in the proportion of substitution/polymorphic positions between the regions forming secondary structures and those flanking them reinforces the presence of strong selective pressure at the structure forming regions.

Large sequence stretches without conserved sequences/binding sites characterized by high mutation rates are also present in the control region. It has been reported that the repeat sequences and secondary structure formation during replication is the major reason for the high rate of mutation observed in some genomic regions (Wright  2000; Samuels *et al*. 2004; Burrow *et al*. 2010). Different models like slipped-strand mispairing (Mita *et al*. 1990; Samuels *et al*. 2004), intermolecular recombination, transposition (Mita *et al*. 1990; Samuels *et al*. 2004) and misalignment during enzymatic replication have been suggested as the mechanism behind the high polymorphisms observed (Pereira *et al*. 2008). Many investigations have reported the evolutionary dynamics of tandem repeats in the mitochondrial DNA control region (Broughton and Dowling 1994; Lee *et al*. 1995; Broughton and Dowling 1997). The presence of repeat sequences which have an inherent tendency of length variation along with secondary structure/single-stranded structure-forming sequences promotes high sequence variability in the control region (Lee *et al*. 1995; Broughton and Dowling 1997; Wright  2000). This is more likely to occur during enzymatic replication of these regions as explained by the different models mentioned earlier. The presence of AT and CG repeats in tandem repeats suggests these regions would have originated from the repeat sequences maintaining the secondary structures associated with conserved sequence elements and subsequently evolved (Broughton and Dowling 1997).

On the contrary, it is clear from the analysis that the secondary structure-forming tandem repeat stretches (between TAS and poly-A) in the control region were conserved among clupeoid fishes which indicated substantial stability along with evolutionary time scales. The position of the conserved tandem repeat sequence region in clupeoids with highly stable intra-strand stem-loop structure, between TAS and poly-A, which includes the D-loop forming region strengthens its possible role in transcription termination (Slomovic *et al*. 2005), replication initiation and/termination of elongation in the proposed models of mitochondrial replication (Shadel and Clayton 1997; Yasukawa *et al*. 2005). It has been reported that the mitochondrial structural variants/ haplogroups have a clear link with mtDNA copy number variation (by influencing the replication machinery) in humans and contribute to the adaptation of the human population to different climatic zones (Suissa *et al*. 2009; Melo-Ferreira *et al*. 2014; Lajbner *et al*. 2018). The secondary structure may also act as a punctuation mark for correct mRNA processing (Ojala *et al*. 1981). The presence of a conserved poly-A after CSB D and CSB 1 may have some functional significance (Slomovic *et al*. 2005). It may promote the formation of secondary structure by initiating displacement during stem-loop structure formation (Cheng *et al*. 1991) as in transcription termination. In mammals, deletions in mitogenome are closely linked to mitochondrial diseases and proven to be associated with site-specific breakage hotspots near the regions with low folding energy (Samuels *et al*. 2004).

### 4.3. tRNA anticodon composition and codon usage in Clupeoid fishes mitochondrial genome; insight into selection and mechanism of adaptation

Clupeoids responded or adapted to deamination related mutation pressure in two ways. The first through fixing anticodon sites saturated with guanine (G) or Thymine (T) (except tRNA Met, Pro) and the second by the positioning of tRNA nearer to the control region. We found that the anticodon of all tRNAs is saturated with the maximum possible G/T substitutions within the constraints of the vertebrate codon table. Along with this, a gradient exists in the position of tRNA between $O_L$ and $O_H$ based on GT content in their anticodons sites. Most of the tRNAs between $O_L$-$O_H$ were placed in increasing order according to the GT content in their anticodon site, along $O_L$-$O_H$ and $O_H$-$O_L$ direction (Fig. 6.6). The frequency of codon usage in mitochondrial proteins is related to the positions of tRNA along mtDNA (codons of tRNA near the control region were highly used) and positions of tRNA were colinear with GT content at its anticodon sites (tRNA with high GT placed near the control region). Thus the Clupeoids mitogenome has adapted to deamination mutation pressure during replication and transcription, by fixing the tRNAs with anticodon saturated with G/T and then positioning them around the $O_H$ according to their degeneracy and GT content. Saturating tRNAs anticodon with G/T, and placing them near control region and OH (the regions of highest deamination pressure) saved them from further deamination pressure. We found a significant correlation between the GT content in the H strand tRNA anticodon sites and the estimated duration of single-stranded exposure/position along the direction of H strand replication (Fig. 2). Similarly, a moderate correlation is found between the GT content in the L strand tRNA anticodon sites and its position between $O_H$-$O_L$ and $O_L$-$O_H$ along the H direction, except tRNA Pro (Fig. 6.6). This suggests that the Clupeoids mitogenomes are adapted to deamination mutations in anticodon sites, during replication and transcription. It also supports the possible role of adaptation to deamination mutations concerning the origin of mt-tRNA position and the evolution of codon usage patterns.

Codon-anticodon adaptation hypothesis for vertebrate mitogenome proposed that highly preferred codon will be matched to the most abundant anticodon (selection hypothesis of anticodon adaptation) (Bulmer 1987, 1991; Xia 2005). Translational selection occurred between synonymous codons translated by a tRNA, when one codon interacts more efficiently than others, with anticodon in its tRNA (Jia and Higgs 2008; Hershberg and Petrov 2008; Charneski *et al.* 2011). In the nuclear gene, translational selection shaped the use of codon corresponding to tRNA gene with high copy number (Hershberg and Petrov 2008). Such association has been observed in genome sequences of Human (Kotlar and Lavner 2006) and *E. coli* (Kanaya *et al*. 1999). However, the translational selection acting on mtDNA may not act at a direction of tRNA gene numbers, because the number of available tRNA is limited to 1 for each amino acids except Leucine and Serine (they have two types of tRNA in mtDNA) (Hershberg and Petrov 2008). Both Nutrality plot and ENc plot indicated that the mutational bias (GC3 content) is the main force shaping the observed codon bias in the clupeoids. Thus the observed codon usage bias is results of adaptation to the tRNA (anticodon saturated with G/T) in the genome. Clupeoids have been adapted to high translational efficiency by using codons complimentary to the tRNA anticodon (saturated with G/T) and codons of the tRNA genes placed close to the control region, where transcription efficiency is high. Besides, they also used hydrophobic amino acids, which are abundantly used in the synthesis of the mitochondrial membrane protein complex. This observation is also consistent with the previous reports based on the vertebrate mitogenome (Satoh *et al*. 2010). Thus we can conclude that translational efficiency-related constraints in mtDNA were shaped by the codon usage pattern in Clupeoids.

Most of the analyses report that the strand-specific mutation bias shaped the anticodon of the tRNAs which drives codon usage bias in the vertebrate mtDNA (mutation hypothesis of anticodon evolution). On the contrary, opposing views assume that selection in codon-anticodon adaptation shaped the anticodon of the tRNAs (selection hypothesis of anticodon adaptation) (Xia 2005; Satoh *et al*. 2010). Consistent with the mutation hypothesis of anticodon evolution, Clupeoids maintained a codon usage bias in the protein-coding region with a strong anti-G bias and abundance of codon with A and C at 3rd codon position than those with T. Whereas, in the fresh/brackish water radiated Clupeoids (in lineage 1-5), a codon usage pattern highly complementary to the GT saturated anticodons, contrary to their marine counterparts have been observed. This may be an adaptation for enhancing osmoregulatory activities (in fresh and brackish waters) by changing their codon usage to a pattern in which the majority of its codons are highly complementary to the fixed GT saturated anticodons. It has been reported that Mitochondrion-rich cells in gills, kidney, and intestine in teleost fishes (Evans *et al*. 2005; Marshall and Grosell 2006) have an important role in osmoregulation/cell homeostasis (ion and water transport across in these tissues) and are primarily involved in adaptation to various osmotic and ionic aquatic habitats (Hwang and Lee 2007; Kaneko *et al*. 2008). Thus the observed codon usage pattern may be a result of accelerated directional mutation associated with increased energy requirement for adaptation to the euryhaline and freshwater environment as observed in some fishes (Kaneko *et al*. 2008; Whitehead *et al*. 2012; Zhang *et al*. 2017). Thus the protein-coding region of Clupeoids mitogenomes evolved towards a codon usage pattern, in which most of them are complementary to the G/T saturated tRNA anticodons in the genome. Hence there is a strong anti-G bias in codon usage and codons with A and C at 3rd codon positions are abundant than those with T in fresh/brackish water radiated Clupeoids contrary to their marine counterparts. Generally, Clupeoids mitogenomes are adapted to deamination mutation pressure during replication and transcription. Efficient mitochondrial gene expression is attained by fixing the tRNAs with anticodon saturated with G/T, then positioning them around the $O_L$ according to their GT content and using codons complementary to tRNA anticodon, which can be translated faster with minimum errors.

The exceptional use of Methionine (has anticodon 5’-CAT-3’instead of TAT, frequent codon ATA) and Proline (has anticodon 5’-AGG-3’instead of GGG, frequent codon CCT) codon/anticodon, deviating from the common codon usage bias discussed above may be associated with the predominant role of selection associated with translational initiation and other function. The use of anticodon TAT for tRNA Met may increase the protein elongation because ATA is the most frequent anticodon. But it may affect initiation rate because the universal start codon that increases the initiation rate is CAT. This suggests that increasing the translation initiation rate is more important than elongation (Xia 2005). The occurrence of outlier tRNA Pro may be associated with other constraints such as the punctuation mark during pre-mRNA processing (Ojala *et al*. 1981) and the availability of rNTPs (Xia 1996). Transcription efficiency can be increased by increasing the use of T in the third codon position due to the high availability of A and low availability of the other three rNTPs in mitochondria (Xia 1996; Hughes *et al.* 2007; Morris *et al.* 2014).

## 4.4. Positive selection in the genes

Selection analysis on sequence alignment confirmed that many codons are under purifying selection and neutrally evolving, but a considerable number of codons were found to be under positive selection and convergent changes have been found among the independent clupeoids lineages. The amino acid identified as positively selected will have some role in speciation and function in the adaptive evolution of clupeoids to different habitats. The destabilising changes/radical amino acid modification tends to concentrate on interior branches on lineages associated with marine to euryhaline or freshwater transitions and tropical to temperate environment transitions. There are several positive radical amino acid modifications in the terminal branches/tips than in the interior branches. In addition to that, the amino acid residue that has been reported to participate in key functions are not overlapping with sites under radical changes and signatures of positive selection are less prevalent than purifying selection and they were located in the complex. These results support the hypothesis that colonisation of clupeoids in different habitat creates a selective regime of positive directional selection in several mitochondrial protein-coding genes and codon usage. However, the key functional amino acid residues have been maintained by a strong purifying selection.

Cytochrome c oxidase (complex IV) was remarkable with length variation in 3' end of CO1 and freshwater specific substitution in the lineage 1 and 3. Cytochrome c oxidase (complex

IV) catalyzes the final step in mitochondrial electron transfer chain and is considered as one of the major regulation sites for oxidative phosphorylation (Li *et al*. 2006). This enzyme is controlled by both nuclear and mitochondrial genomes. Subunits I, II and III are the catalytic core of the enzyme. Subunit I - III and the nuclear subunits are essential for the assembly and catalytic function of complex 4. It receives an electron from each of four cytochrome c molecules which transfer electrons between complex 3 and 4 and transfers them to one oxygen molecule. The Subunits I contains two hemes, cytochrome a and cytochrome a3, and one copper centres, CuB. The Subunits II contains CuA and cytochrome c binding site. The binuclear centre formed by cytochrome a3 and CuB act as a site of oxygen reduction. Cytochrome c, which is reduced by the Subunits III, binds to cytochrome c binding site near the CuA binuclear centre and passes an electron to it. The reduced CuA binuclear centre then passes an electron to cytochrome a, subsequently, it passes an electron to the cytochrome a3-CuB binuclear centre (Scott 1995). During this process, it converts one molecular oxygen to two molecules of water by using four protons from the inner aqueous phase to make water and also translocates four protons across the membrane.

As shown in Fig. 6.S2b amino acid C at 44 in CO2 gene may have some functional importance in speciation and adaptation of clupeoids to freshwater habitat. Because amino acid C (site #44) is common to all freshwater clupeoids in lineage 1 and 3 and it is not C in all other clupeoids except *E. thoracata* and Engraulidae. But it is not specific to freshwater because one marine species *E. thoracata* shared this amino acid C and one freshwater species *P. richmondia* did not possess it. We hypothesize that this substitution could be an ancestral polymorphism rather than a convergent evolution, which could have provided an advantage when freshwater colonisation occurred. Based on the available evidence, the lineage 1, 2 and 3 were formed by one of the three dispersal events crossing the K-Pg extinction boundary and subsequent allopatric cladogenesis. So the adaptation to freshwater occurred in different place and time. We hypothesize that the convergence of this amino acid substitution in CO2 may be associated with increased energy requirement in a freshwater environment. Therefore along with the codon usage bias, these proteins may have some important role in the osmoregulatory process in the freshwater clupeoids. The presence of C in the *E. thoracata* is difficult to explain. It may be an indication of a possible re-invasion of freshwater-adapted *E. thoracata* to coastal marine⁄esturarine habitats along the IWP region. The re-invasion of the marine environment and biome conservatism in the Engraulidae along the northern South American coast has been reported (Bloom and Lovejoy 2012). The conserved nature of most of the key amino acid residues that have been reported to participate in electron transfer pathway, putative water exit pathway, ion/chemical binding and putative proton exit pathway in complex IV indicates that these regions are constrained functionally. Mutations observed outside the key functional residues could be related to relaxed purifying selection (Jacobsen *et al*. 2016).

The observed c-terminal variation in the co1 might have a role in translational regulation of its synthesis. The Carboxyl-terminal end (c-terminal) of CO1 is hydrophilic and exposed to the matrix side of the inner membrane and some functionally interacting residues have been characterized in cichlids CO1 c-terminal (Fischer 2013). It has been reported that the c-terminal domain regulates the assembly and feedback control of co1 synthesis in yeast (Shingu-Vazquez 2010) and co1 is the limiting factor in the assembly of complex IV in fishes (Fischer 2013). Thus the short c-terminal end of co1 that occurred exclusively in Engraulidae may be a factor behind the formation of the clade (Engraulidae) by sympatric speciation/cladogenesis. Based on the available information, Engraulidae and other clupeoids shared a common ancestor around 119 MYA.

The NADH dehydrogenase complex is the first and largest multimeric enzyme of the five complexes constituting the oxidative phosphorylation pathway/respiratory chain (Sazanov 2015). It provides electrons for reduction of quinine to quinol which is available from oxidation of NADH, translocates four protons ($H^+$) across the inner membrane and generates an electrical proton gradient. Complex 1 is L-shaped with all 7 mtDNA encoded hydrophobic protein subunits in the membrane-embedded domain and peripheral domain encoded by the nuclear genome. In addition to 14 basic subunits in bacteria, 32 different subunits have been reported in Bovine heart mitochondrial Complex 1 (Fiedorczuk *et al*. 2016). ND1 and ND2 are located in between peripheral and transmembrane domain whereas ND4 and ND5 occurred at the distal end of the transmembrane domain. Subunit ND2 (homolog of NuoN in *E. coli*), ND4 (NuoM) and ND5 (NuoL) directly act as proton pump for $H^+$ ions. They are homologous to each other and belong to a class of $Na^+$/$H^+$ antiporters. Even though some regions have been assigned with functions, subunits with unknown functions, a mechanism that couple electron transfer and proton pumping are still debated. Amino acid changes in these subunits may have some adaptive value as it interferes with the efficiency of the proton-pumping process. Freshwater clupeoids in lineage 3 is also carrying unique amino acid substitution in ND2 (site#23, 86) and unique substitution at site#566 of ND5 in lineage 3. Similarly, we identified positively selected/radical amino acid changes in ND4 (site#183), ND5 (site#577) and ND6 (site#118) which is specific to temperate water species in lineage 2 and 4. Amino acid substitution C in ND4 (site#183) and A/T/Q in ND5 (site#577) is specific to lineage 3, and D/A in ND6 is specific to lineage 4. We hypothesize that these sites are important for adaptation of clupeoids from tropic water to temperate water habitat.

Even though the Key polar amino acid residues which have been reported to participate in proton translocation (Zhu *et al*. 2016) through complex I were conserved across all species, the highest number of positively selected amino acid sites were found in ND 2, ND4 and ND5 genes. These sites also show the highest average number of positive radical amino acid modification. The predicted transmembrane domain (TM) showed that the sites with higher radical amino acid changes (in ND2, ND4 and ND5) are located mostly in loop regions, suggesting strict functional constraints acting on the TM region, which acts as the proton pumping device. Since none of the substitutions was located directly at known functional regions, they are not likely to be involved in electron translocation and proton pumping. However amino acid changes that hinder/improve the efficiency of proton translocation and conformational coupling of mitochondrial protein domain could affect the performance of complex 1. Comparatively high radical amino acid changes evident in these regions and the observed diversifying selection can be related to relaxed purifying selection (Jacobsen *et al*. 2016)

Many studies have been reported that candidate sites for positive selection are disproportionately concentrated in the complex I in many fishes (Garvin *et al*. 2015a,b; Caballero *et al*. 2015; Consuegra *et al*. 2015; Garvin *et al*. 2012; Jacobsen *et al*. 2016; Teacher *et al*. 2012). OXPHOS complex I produce ~ 40% of the proton-pumping required for ATP synthesis. Polymorphism in this region is also reported in other groups like Hares (Melo-Ferreira *et al*. 2014), Mammals (da Fonseca *et al*. 2008), Tachycineta (Stager *et al*. 2014) and Monkeys (Yu *et al*. 2011).

Cytochrome b is a part of respiratory protein complex III, which is the middle component of the mitochondrial respiratory chain, coupling the transfer of electrons from ubihydroquinone to cytochrome c with the generation of an electrochemical gradient across the mitochondrial membrane. Both Qo and Qi binding site in the cytochrome b subunit plays a key role in the function of complex 3 (Crofts 2004a, b; Kolling *et al*. 2003). The water-binding capacity of the Qi site for the reduction of ubiquinone to ubiquinol is critical in this process (Crofts 2004a, b).

Amino acid sites that have been suggested to participate in Qo binding and Qi binding are conserved and observed amino acid changes in Cytochrome b were away from Qo and Qi binding sites.

ATP synthase (complex V) is composed of a soluble catalytic F1 region and a membrane-inserted $F_O$ region. In mammals, the subunit composition is α3, β3, γ, δ and ε for the $F_1$ region and the $F_O$ region with subunits a, e, f, g, A6L, DAPIT, two membrane-inserted α -helices of subunit b, and the c8-ring (Walker 2013). The rotor subcomplex consists of subunits g, d, e, and the c8-ring. In addition to the rotor, the F1 and FO regions are connected by a peripheral stalk composed of subunits OSCP, d, F6, and the hydrophilic portion of subunit b (Walker 2013; Baker *et al*. 2012). The mechanism by which ATP synthesis and hydrolysis are coupled to rotation of the g subunit is well understood (Walker 2013), but still, it is not clear how the rotation of the central rotor is coupled to proton translocation through the $F_O$ region. According to the most popular model, the proton translocation occurs through two half channels near the a-subunit/c-subunit interface (Junge and Nelson 2005). In this model, one-half channel allows protons to move half-way across the lipid bilayer and protonate the conserved Glu58 residue of one of the c subunits. The other half channel allows the deprotonation of the adjacent c-subunit (Lau and Rubinstein 2012). This will subsequently lead to a net rotation of the entire c-ring by Brownian motion. Electron cryo-microscopy analysis of the bovine mitochondrial ATP synthase suggests that the matrix half channel of the ATP synthase probably formed by the cavity between the c8-ring and the matrix ends of tilted a-helices #5 and #6 of the subunit (Zhou *et al*. 2015). The lumenal half channel in the V-ATPase is formed entirely from the a-helices of a subunit and the corresponding inter-membrane space half channel in the ATP synthase is composed of the intermembrane space ends of a-helices #5 and #6 and one or both of the two trans-membrane a-helices of the b subunit (Zhou *et al*. 2015).

Even though the positive selection in CYTB and ATP6 (complex 5) were located in the intra-helix loop and away from known functional amino acid regions, amino acid replacements can result in regional changes to hydrophobicity and structure within the protein and it has the potential to alter the coupling efficiency of the protein complex. In humans mutations characterized by enhanced binding of water at Qi site have been linked to increased longevity (Beckstead *et al*. 2009) and in yeast, mutation at Qo binding site have been linked to reduced catalysis efficiency and increased oxygen radical production (Wenz *et al*. 2007).

Mitochondrial DNA variation has complex fitness consequences which may get amplified by mito-nuclear interactions and consequently concerted mito-nuclear co-evolution is very essential for the maintenance of metabolic and physiological functions (Ballard and Pichaud 2014; Havird and Sloan 2016; Horan *et al*. 2013; Osada and Akashi 2011; Fox 2012; Wolff *et al*. 2014). When mito-nuclear interactions are disrupted, it results in reproductive isolation and speciation (Burton *et al*. 2013; Dowling *et al*. 2008).

## 4.5. Conclusions

The evidence for the ability to form stable secondary structures by sequences flanking the conserved sequence elements, negative selection (less variability in the stem forming region between clupeoids species, as in the tRNA genes) and positive selection (complementary/compensatory mutations in the stem region, as in the tRNA genes) indicate the importance of the stable secondary structures in the mitochondrial DNA function and evolution. The reason for the persistence of large non-coding regions in the mitochondrial genome is because of the conserved sequence blocks along with the sequences flanking them form secondary structures consequently acting as recognition sites for regulatory proteins. The sequences flanking the secondary structure-forming regions are mutational hot spots with high rates of substitutions, deletions, and insertions. The errors originating during the enzymatic replication of secondary structure-forming sites give rise to high mutation rates in the flanking regions making them mutational hotspots. This can explain the high variability observed in the mitochondrial control region. However, the secondary structure-forming tandem repeat stretches with substantial stability may have greater evolutionary significance in Clupeoid fishes. Further investigations are needed to understand the existence of similar secondary structures in the control region of other animals and the adaptive consequences of control region variations.

Highly conserved tRNA gene arrangement and codon usage in Clupeoids mtDNA are not maintained by the direct action of translational constraints and strand-specific mutation bias respectively. The adaptation to deamination pressure by fixing of tRNAs saturated with G/T at its anticodons and subsequent placing of it around the $O_L$ according to their GT content may be the driving force for codon usage bias. Thus the fixed position G/T saturated tRNA in the mtDNA will be the reason for codon usage bias observed in the Clupeoids. The observed codon usage pattern in euryhaline and freshwater clupeoids may be a result of accelerated directional mutation associated with increased energy requirement for adaptation to the euryhaline and freshwater environment. This is the first empirical evidence for codons evolving to adapt to anticodons in mtDNA.

This study provides evidence for positive selection in the OXPHOS complex of clupeoid species distributed in the wide marine environment. Not only amino acid variation but also the base composition bias and codon usage bias in clupeoids seems to be shaped by mutational bias and natural selection forces. Signatures of positive selection are less prevalent than purifying selection and there are several positive radical amino acid modifications in the terminal branches/tips than in the interior branches. Positively selected/radical amino acid substitutions observed in CO2 and Complex 1 of freshwater and temperate water species respectively, may have some functional importance in speciation and adaptation of clupeoid to freshwater habitat. The short c-terminal end of co1 occurred exclusively in Engraulidae may be a factor behind the formation of the clade (Engraulidae) by sympatric speciation/cladogenesis. The preference of codon corresponding to the abundant tRNA/ affinity to A at 3rd codon position and avoidance of G at 3rd codon position could help in relatively high expression of mitochondrial genes in the clupeoids which are adapted to the euryhaline and freshwater habitat. We believe that convergent evolution occurred by selection at third codon position results in the same codon usage pattern in independently evolved euryhaline and freshwater clupeoid lineages in different oceans. These results support the hypothesis that colonisation of clupeoids in different habitat creates a selective regime of positive directional selection in several mitochondrial protein-coding genes and codon usage. However, the key functional amino acid residues have been maintained by strong purifying selection. Extensive non-synonymous mutations have been reported in NADH dehydrogenase (Complex 1) and Cyt b genes in many fishes as in the present study, supporting the evidence that these genes play an important role on species adaptation and enhanced opportunities for evolutionary radiations in clupeoids.

This study provides molecular evidence that highlights the importance of OXPHOS gene evolution in plasticity, colonization and adaptation to new environments. Insights from our study indicate the need for future experimental characterisation of specific mutations, codon usage pattern and its effect on the efficiency of oxidative phosphorylation and physiological impacts. This will help in predicting the response of organisms to future climate changes and mitochondrial DNA based genetic improvements.

## Supplementary Figures and Tables

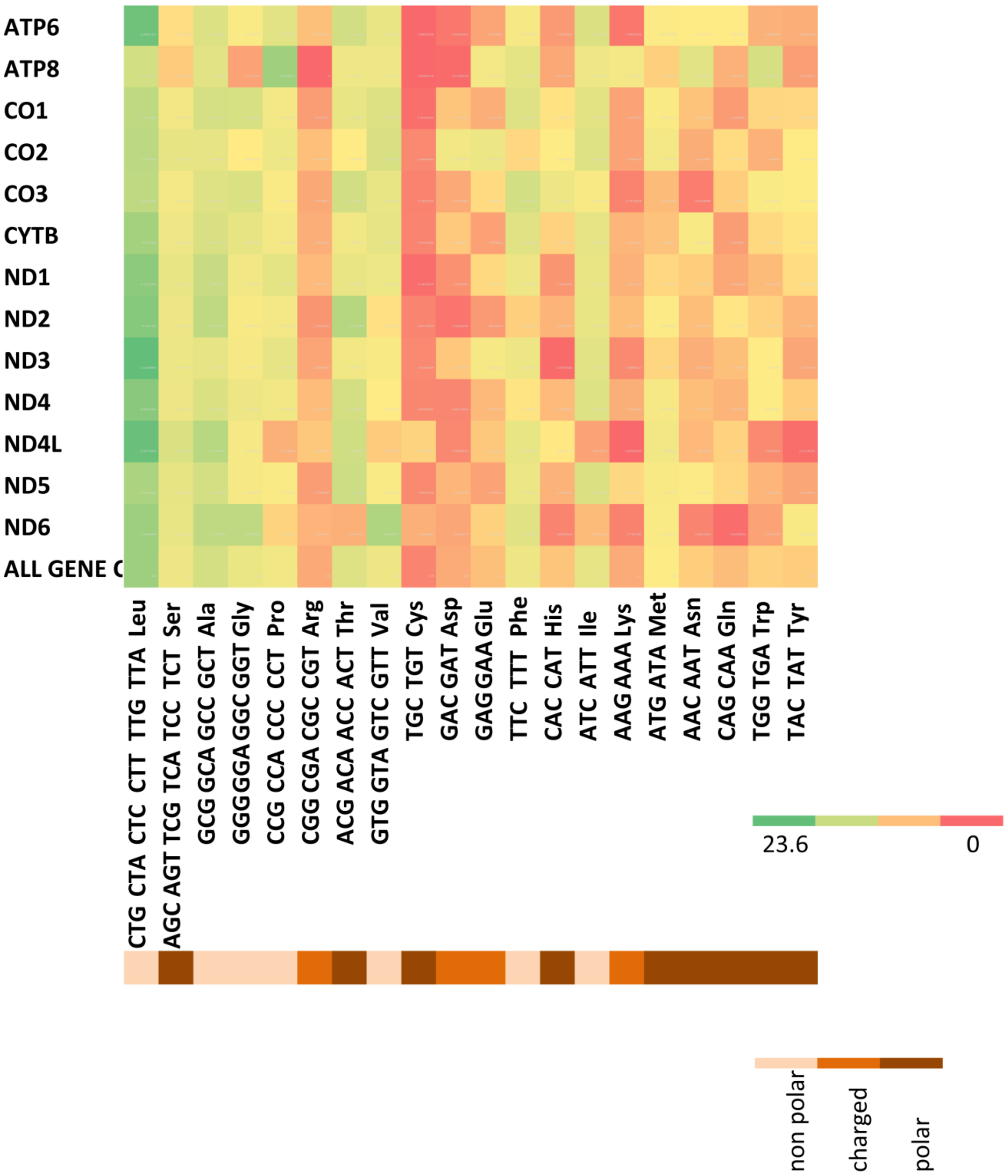


**Fig. 6.S1** Percentage of amino acid contents of merged protein-coding genes of Clupeoid fishes.

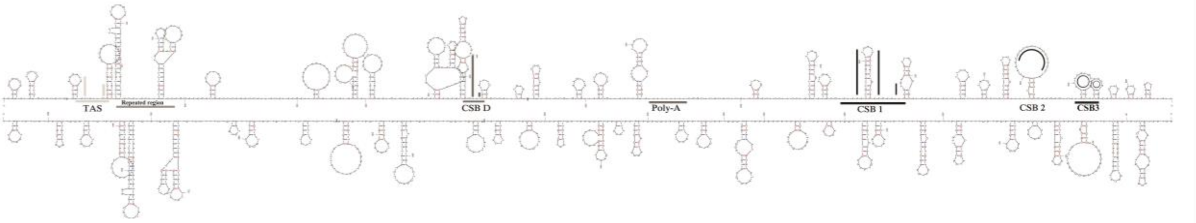


**Fig. 6.S2** Potential secondary structure of repeat sequences identified in the mtDNA control region of *S. longiceps.*

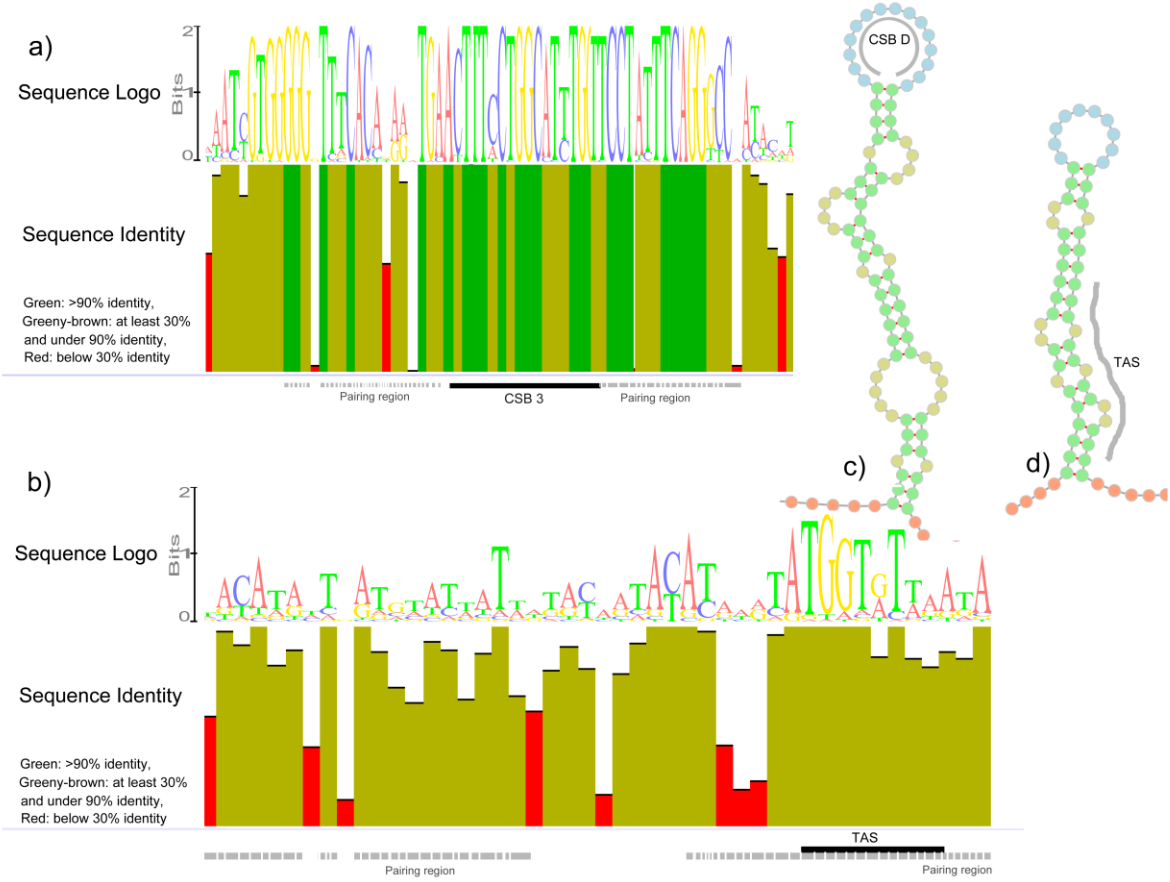
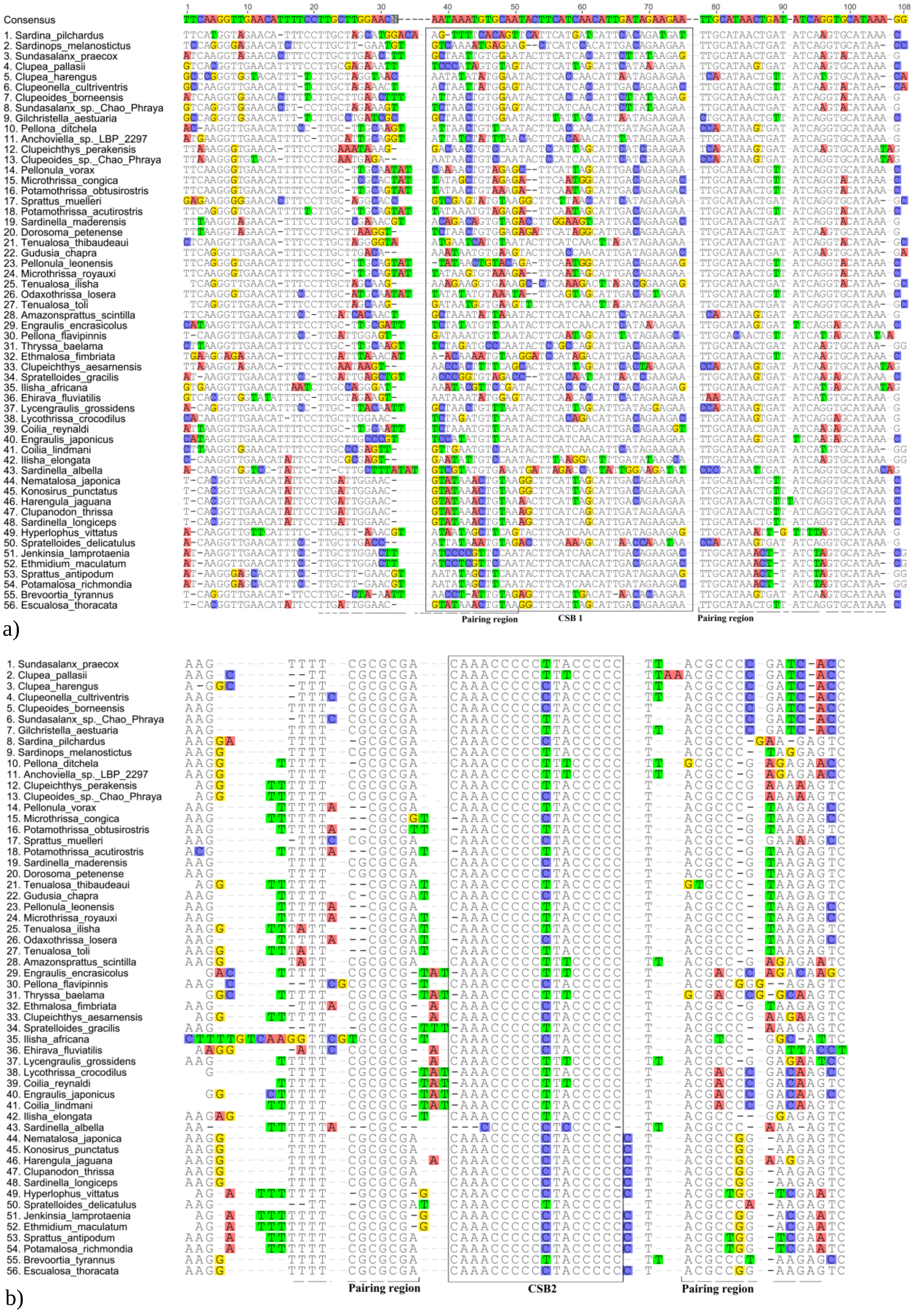


**Fig. 6.S3** Potential secondary structure identified in the CSB D and TAS of Clupeoids mtDNA control region.


Consensus
1. Sardina_pilchardus
2. Sardinops_melanostictus
3. Sundasalanx_praecox
4. Clupea_pallasii
5. Clupea_harengus
6. Clupeonella_cultriventris
7. Clupeoides_borneensis
8. Sundasalanx_sp._Chao_Phraya
9. Gilchristella_aestuaria
10. Pellona_ditchela
11. Anchoviella_sp._LBP_2297
12. Clupeichthys_perakensis
13. Clupeoides_sp._Chao_Phraya
14. Pellonula_vorax
15. Microthrissa_congica
16. Potamothrissa_obtusirostris
17. Sprattus_muelleri
18. Potamothrissa_acutirostris
19. Sardinella_maderensis
20. Dorosoma_petenense
21. Tenualosa_thibaudeaui
22. Gudusia_chapra
23. Pellonula_leonensis
24. Microthrissa_royauxi
25. Tenualosa_ilisha
26. Odaxothrissa_losera
27. Tenualosa_toli
28. Amazonsprattus_scintilla
29. Engraulis_encrasicolus
30. Pellona_flavipinnis
31. Thryssa_baelama
32. Ethmalosa_fimbriata
33. Clupeichthys_aesarnensis
34. Spratelloides_gracilis
35. Ilisha_africana
36. Ehirava_fluviatilis
37. Lycengraulis_grossidens
38. Lycothrissa_crocodilus
39. Coilia_reynaldi
40. Engraulis_japonicus
41. Coilia_lindmani
42. Ilisha_elongata
43. Sardinella_albella
44. Nematalosa_japonica
45. Konosirus_punctatus
46. Harengula_jaguana
47. Clupanodon_thrissa
48. Sardinella_longiceps
49. Hyperlophus_vittatus
50. Spratelloides_delicatulus
51. Jenkinsia_lamprotaenia
52. Ethmidium_maculatum
53. Sprattus_antipodum
54. Potamalosa_richmondia
55. Brevoortia_tyrannus
56. Escualosa_thoracata
Pairing region
CSB 1
Pairing region
a)
1. Sundasalanx_praecox
2. Clupea_pallasii
3. Clupea_harengus
4. Clupeonella_cultriventris
5. Clupeoides_borneensis
6. Sundasalanx_sp._Chao_Phraya
7. Gilchristella_aestuaria
8. Sardina_pilchardus
9. Sardinops_melanostictus
10. Pellona_ditchela
11. Anchoviella_sp._LBP_2297
12. Clupeichthys_perakensis
13. Clupeoides_sp._Chao_Phraya
14. Pellonula_vorax
15. Microthrissa_congica
16. Potamothrissa_obtusirostris
17. Sprattus_muelleri
18. Potamothrissa_acutirostris
19. Sardinella_maderensis
20. Dorosoma_petenense
21. Tenualosa_thibaudeaui
22. Gudusia_chapra
23. Pellonula_leonensis
24. Microthrissa_royauxi
25. Tenualosa_ilisha
26. Odaxothrissa_losera
27. Tenualosa_toli
28. Amazonsprattus_scintilla
29. Engraulis_encrasicolus
30. Pellona_flavipinnis
31. Thryssa_baelama
32. Ethmalosa_fimbriata
33. Clupeichthys_aesarnensis
34. Spratelloides_gracilis
35. Ilisha_africana
36. Ehirava_fluviatilis
37. Lycengraulis_grossidens
38. Lycothrissa_crocodilus
39. Coilia_reynaldi
40. Engraulis_japonicus
41. Coilia_lindmani
42. Ilisha_elongata
43. Sardinella_albella
44. Nematalosa_japonica
45. Konosirus_punctatus
46. Harengula_jaguana
47. Clupanodon_thrissa
48. Sardinella_longiceps
49. Hyperlophus_vittatus
50. Spratelloides_delicatulus
51. Jenkinsia_lamprotaenia
52. Ethmidium_maculatum
53. Sprattus_antipodum
54. Potamalosa_richmondia
55. Brevoortia_tyrannus
56. Escualosa_thoracata
Pairing region
CSB2
Pairing region
b)

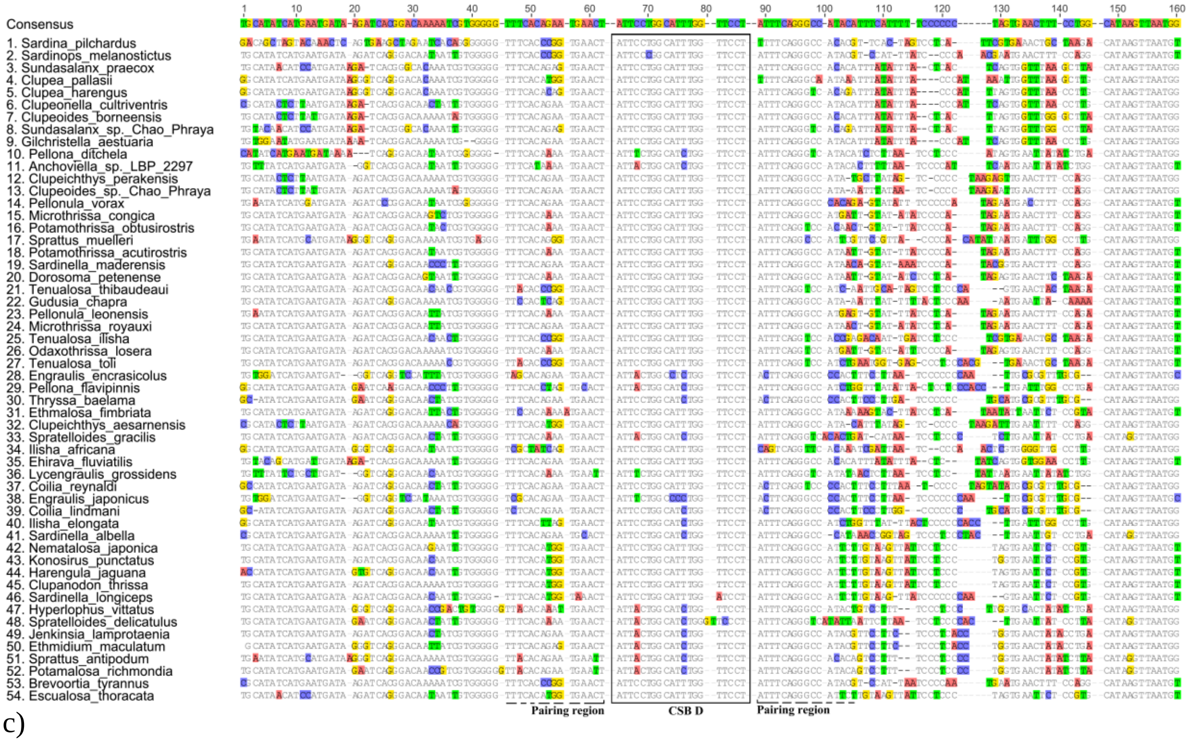


**Fig. 6.S4** Multiple sequence alignment of clupeoids a) CSB1, b) CSB2 and c) CSBD and pairing flanking regions.

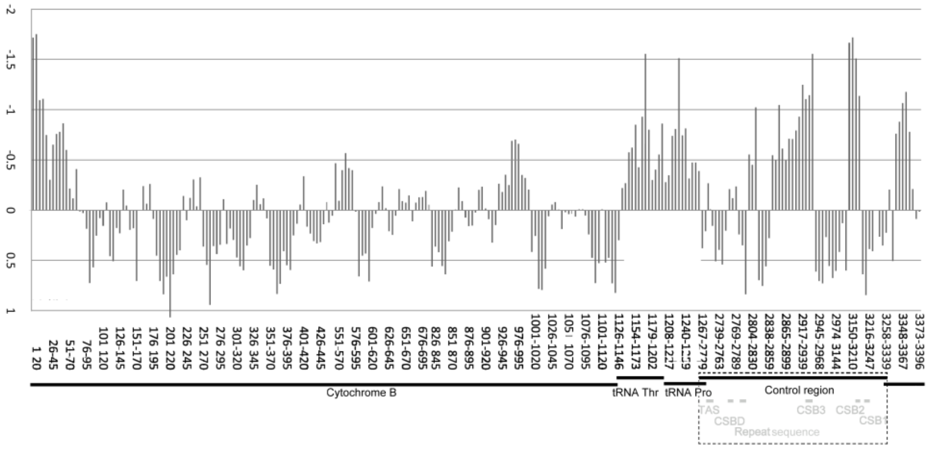


**Fig. 6.S5** Tajima's D value for intervals of 25 bp, overlapping by 5 bp, for DNA sequence alignment which includes Cytochrome b, tRNA Thr, tRNA Pro, control region and tRNA Phe.

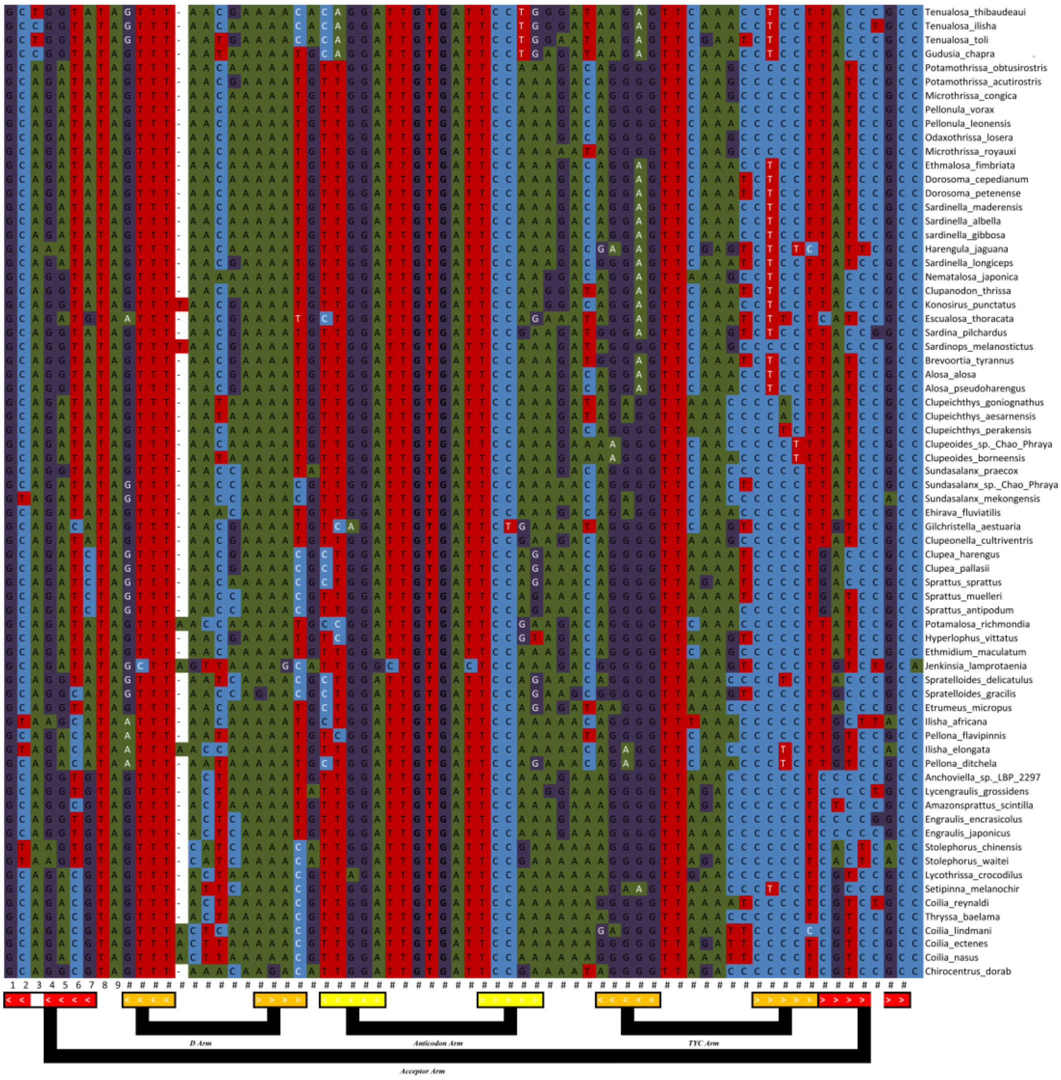


**Fig. 6.S6** Schematic diagram of clupeoid tRNA histidine and complementary mutations in its pairing region

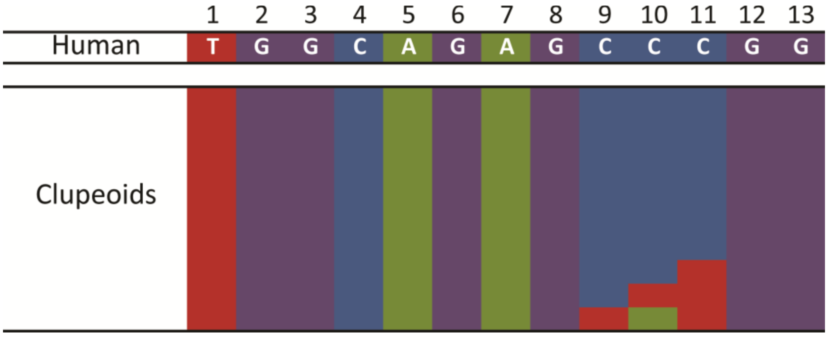


**Fig. 6.S7** Schematic diagram of human transcription termination factor binding site and Base frequencies of the mitochondrial transcription termination factor binding site in the tRNA-Leu (UUR) gene in the mitogenomes of clupeoid fishes.

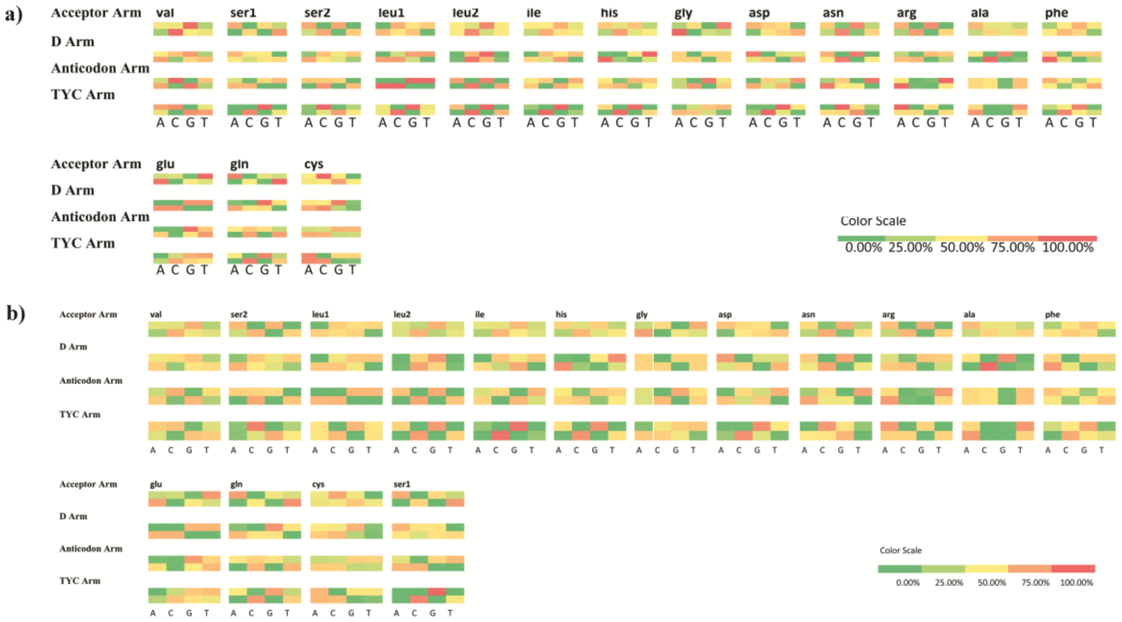
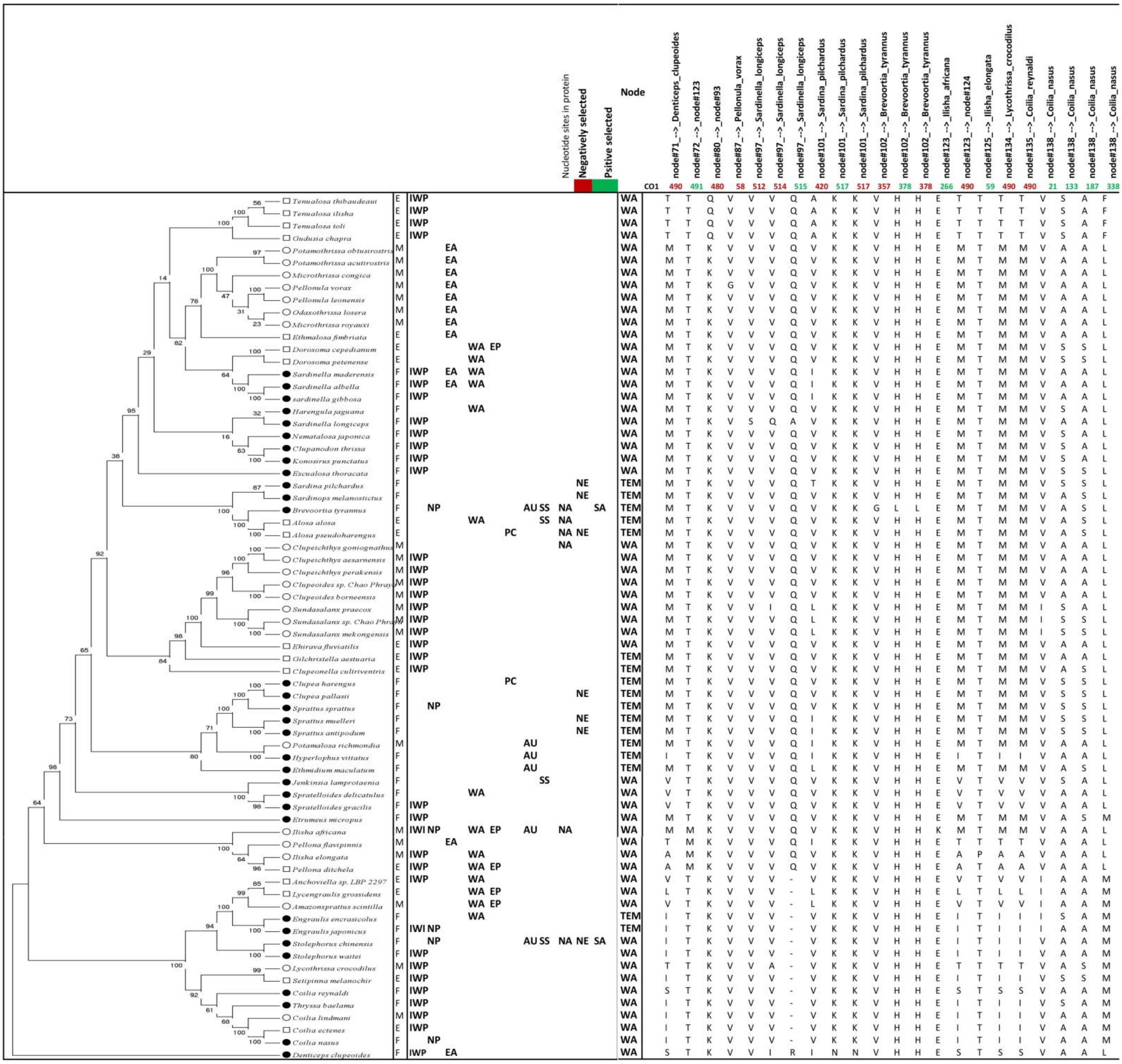


**Fig. 6.S8** Average A,T,G and C content of merged tRNA coding genes of clupeoid fishes. Color scale in a) represents percentage of A, T, G and C in each tRNA gene, b) represents average percentage of A, T, G and C in all tRNA genes.


Nucleotide sites in protein
Negatively selected
Psitive selected
Node
CO1
node#71_-->_Denticeps_clupeoides
node#72_-->_node#123
node#80_-->_node#93
node#87_-->_Pellonula_vorax
node#97_-->_Sardinella_longiceps
node#101_-->_Sardina_pilchardus
node#102_-->_Brevoortia_tyrannus
node#123_-->_Ilisha_africana
node#123_-->_node#124
node#125_-->_Ilisha_elongata
node#134_-->_Lycothrissa_crocodilus
node#135_-->_Coilia_reynaldi
node#138_-->_Coilia_nasus

| | | | | | |
|---|---|---|---|---|---|
| E | EURYHALINE | IWP | Indo-West Pacific | TEM | TEMPERATE WATER (t<25) |
| F | FRESH WATER | NP | North Pacific | WA | WARM WATER (t>25) |
| M | MARINE | EA | East Atlantic | | |
| | | WA | West Atlantic | | |
| | | EP | East Pacific | | |
| | | AU | South Australia | | |
| | | SS | south South America | | |
| | | NA | Northwest Atlantic | | |
| | | NE | Northeast Atlantic | | |
| | | SA | South Africa | | |

**Fig. 6.S9(a)** Amino acid changes under positive and negative/purifying selection in the CO1 subunits of Complex I of clupeoid fishes.

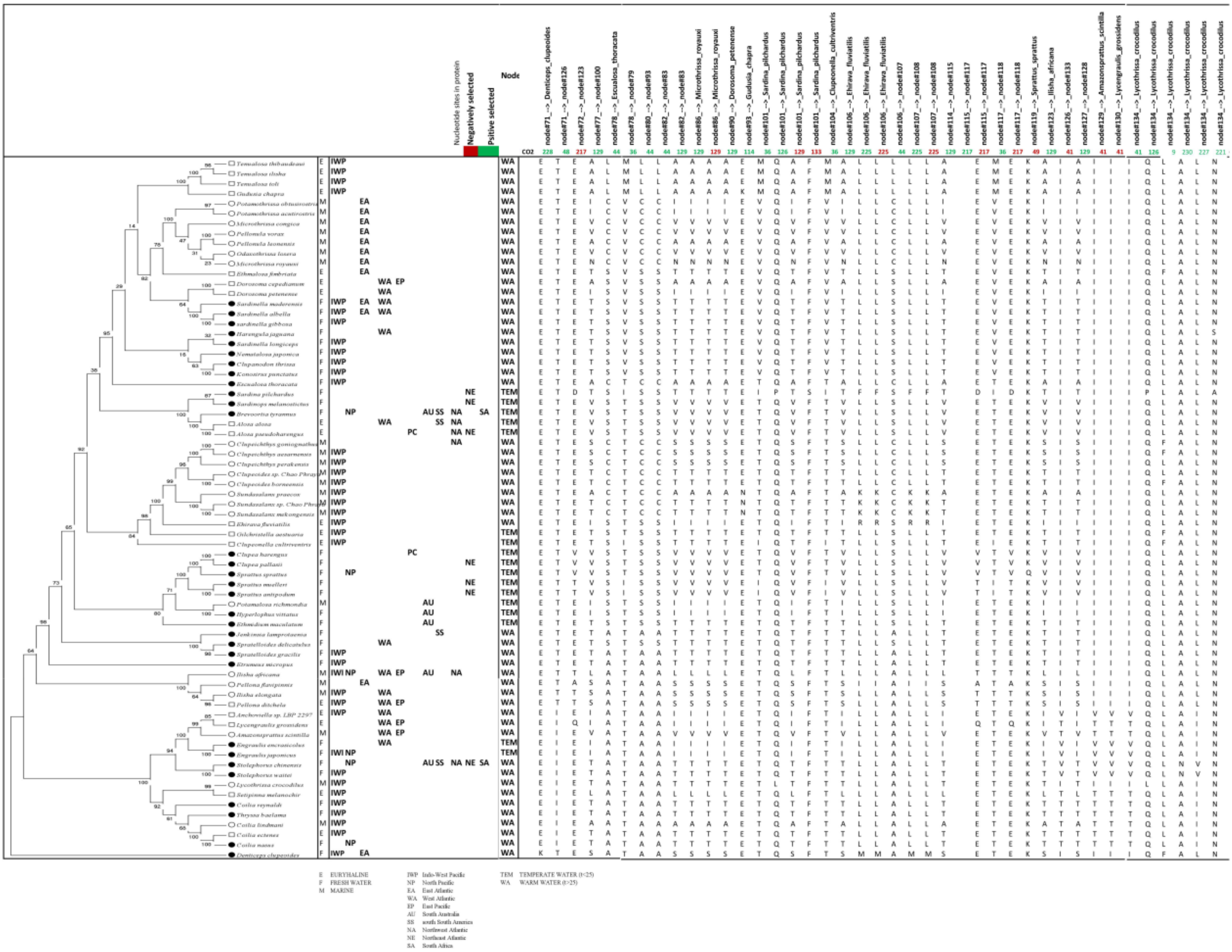


**Fig. 6.S2(b)** Amino acid changes under positive and negative/purifying selection in the CO2 subunits of Complex I of clupeoid fishes.

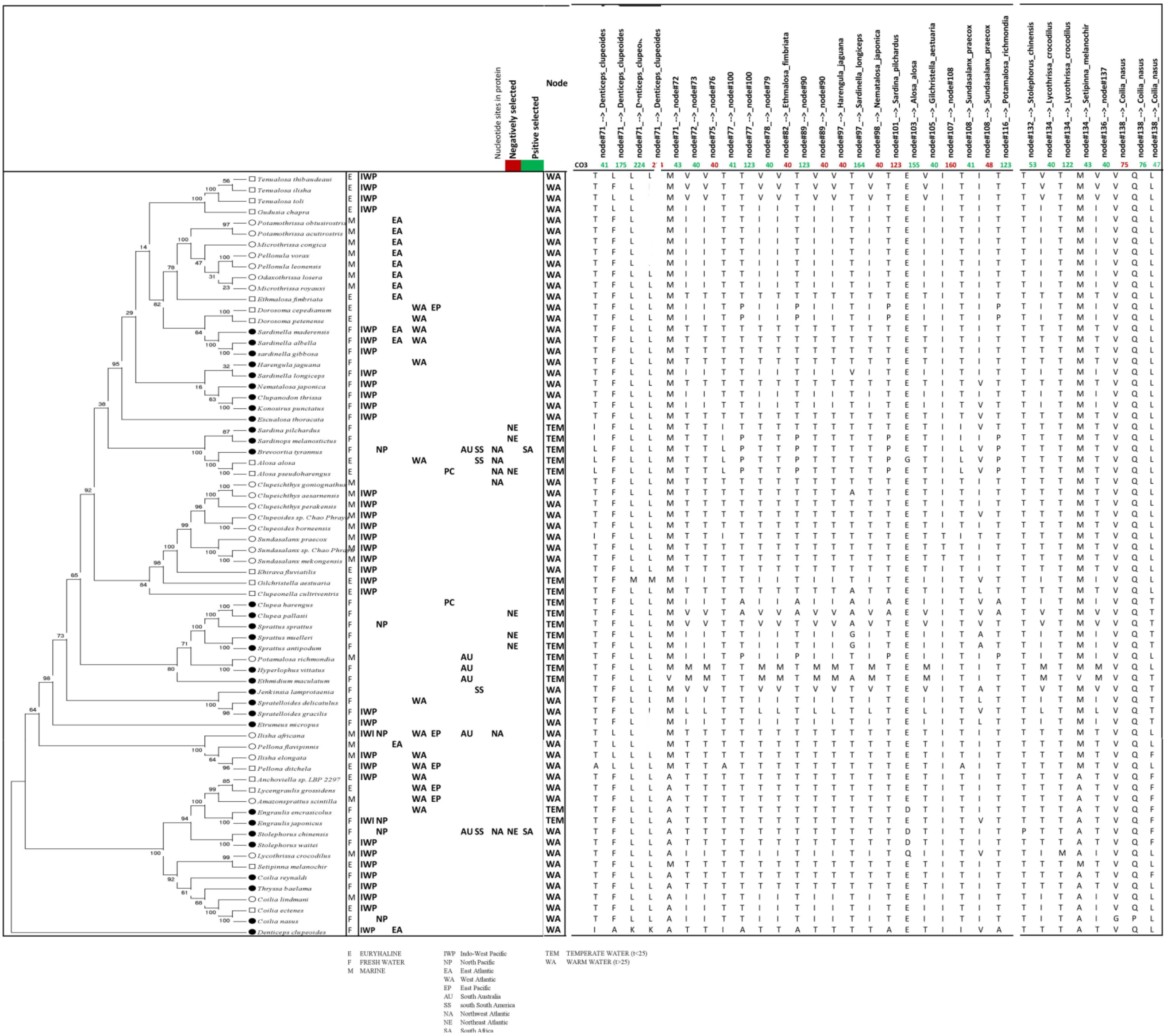


**Fig. 6.S2(c)** Amino acid changes under positive and negative/purifying selection in the CO3 subunits of Complex I of clupeoid fishes.

**Table 6.S1** The genetic distance of protein-coding genes calculated against the consensus sequence of each protein-coding genes of all considered clupeoid fishes. Linear least squares regression with a pairwise distance of 12s rRNA and protein-coding genes of all considered clupeoid fishes

The genetic distance of genes calculated against their consensus sequence

| | ND6 | ND5 | ND4L | ND4 | ND3 | ND2 | ND1 | CYTB | CO3 | CO2 | CO1 | ATP8 | ATP6 | 12S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Alosa_alosa 1* | 12.98 | 9.42 | 8.77 | 12.05 | 9.61 | 10.52 | 10.79 | 10.71 | 5.73 | 6.92 | 7.48 | 3.78 | 9 | 4.44 |
| *Alosa_pseudoharengus 2* | 12.02 | 8.97 | 8.37 | 10.46 | 8.95 | 9.51 | 10.2 | 8.15 | 6.02 | 7.76 | 6.61 | 4.43 | 7.98 | 3.95 |
| *Amazonsprattus_scintilla 3* | 22.31 | 18.01 | 15.21 | 19.64 | 21.91 | 18.57 | 17.8 | 12.41 | 12.18 | 12.6 | 11.77 | 18.67 | 14.15 | 13.58 |
| *Anchoviella_sp._LBP_2297 4* | 14.44 | 14.76 | 13.12 | 15.47 | 15.74 | 15.56 | 11.51 | 10.32 | 10.81 | 10.51 | 9.16 | 20.42 | 9.41 | 13.82 |
| *Brevoortia_tyrannus 5* | 14.9 | 12.2 | 7.2 | 14.58 | 12.38 | 11.54 | 12.46 | 7.83 | 6.58 | 8.1 | 6.81 | 4.43 | 8.61 | 3.96 |
| *Clupanodon_thrissa 6* | 25.43 | 12.01 | 12.78 | 11.73 | 10.62 | 14.62 | 13.14 | 9.69 | 8.3 | 8.23 | 8.17 | 10.65 | 10.35 | 16.06 |
| *Clupea_harengus 7* | 24.81 | 13.06 | 15.81 | 13.62 | 14.79 | 16.88 | 15.49 | 11.88 | 9.47 | 7.86 | 7.26 | 11.28 | 10.16 | 4.92 |
| *Clupea_pallasii 8* | 24.55 | 12.86 | 15.81 | 13.26 | 14.79 | 16.77 | 15.34 | 11.75 | 9.63 | 7.69 | 7.18 | 11.28 | 9.96 | 5.17 |
| *Clupeichthys_aesarnensis 9* | 20.28 | 11.21 | 19.3 | 13.66 | 16 | 21.87 | 11.87 | 9.11 | 9.88 | 9.63 | 9.89 | 12.58 | 9.25 | 4.93 |
| *Clupeichthys_goniognathus 10* | 19.77 | 11.02 | 24.44 | 13.03 | 18.58 | 21.62 | 12.62 | 9.42 | 9.57 | 9.96 | 8.98 | 11.12 | 9.25 | 10.26 |
| *Clupeichthys_perakensis 11* | 18.49 | 11.59 | 22.62 | 13.75 | 14.99 | 22.36 | 12.5 | 9.59 | 9.13 | 9.96 | 9.27 | 10.41 | 9.59 | 10.8 |
| *Clupeoides_borneensis 12* | 21.03 | 11.17 | 17.73 | 11.13 | 15.44 | 21 | 11.59 | 9.3 | 10.19 | 9.13 | 8.99 | 11.79 | 12.34 | 9.76 |
| *Clupeoides_sp._Chao_Phraya 13* | 19.48 | 10.9 | 19.97 | 10.53 | 12.24 | 20.3 | 13.71 | 9.34 | 8.98 | 9.82 | 8.61 | 11.09 | 11.15 | 8.02 |
| *Clupeonella_cultriventris 14* | 17.23 | 12.06 | 11.08 | 15.08 | 14.71 | 16.19 | 14 | 13.67 | 10.64 | 10.81 | 9.67 | 7.77 | 10.09 | 8.39 |
| *Coilia_ectenes 15* | 16.81 | 12.87 | 14.88 | 12.16 | 15.39 | 11.96 | 11.27 | 10.81 | 9.83 | 10.4 | 11.2 | 18.14 | 11.12 | 6.51 |
| *Coilia_lindmani 16* | 17.55 | 12.67 | 15.32 | 13.4 | 16.14 | 12.41 | 10.89 | 11.45 | 8.03 | 10.39 | 10.99 | 17.98 | 11.11 | 10.99 |
| *Coilia_nasus 17* | 17.53 | 12.67 | 14 | 12.96 | 15.77 | 12.08 | 11.27 | 10.62 | 10.25 | 10.23 | 10.66 | 18.14 | 10.26 | 10.99 |
| *Coilia_reynaldi 18* | 19.89 | 14.34 | 11.84 | 13.3 | 15.05 | 15.58 | 11.49 | 10.86 | 10.12 | 10.75 | 11.83 | 16.5 | 11.46 | 11.11 |
| *Denticeps_clupeoides 19* | 21.9 | 19.06 | 15.5 | 20.47 | 20.58 | 23.03 | 17.74 | 16.34 | 14.86 | 11.84 | 14.68 | 18.62 | 16.07 | 11.11 |
| *Dorosoma_cepedianum 20* | 12.45 | 9.58 | 10.77 | 10.41 | 8.65 | 9.9 | 9.56 | 7.25 | 8.16 | 7.45 | 7.56 | 7.81 | 7.96 | 3 |
| *Dorosoma_petenense 21* | 13.09 | 8.67 | 7.94 | 11.12 | 13.38 | 11.38 | 9.82 | 8.35 | 9.52 | 6.96 | 7.11 | 9.19 | 10.71 | 2.89 |
| *Ehirava_fluviatilis 22* | 21.03 | 10.89 | 14.62 | 9.94 | 11.15 | 22.59 | 11.79 | 9.72 | 7.95 | 10.31 | 7.75 | 8.35 | 9.56 | 8.46 |
| *Engraulis_encrasicolus 23* | 23.84 | 17.76 | 18.72 | 18.1 | 18.8 | 19.45 | 12.54 | 11.03 | 11.68 | 13.3 | 9.52 | 20.6 | 13.03 | 13.27 |
| *Engraulis_japonicus 24* | 23.52 | 17.83 | 16.03 | 18.1 | 18.8 | 19.59 | 13.81 | 10.82 | 10.96 | 13.12 | 8.92 | 24.08 | 13.21 | 12.98 |
| *Escualosa_thoracata 25* | 31.78 | 13.98 | 16.45 | 15.46 | 16.22 | 15.15 | 16.05 | 14.55 | 11.04 | 12.36 | 10.84 | 10.5 | 17.71 | 11.61 |
| *Ethmalosa_fimbriata 26* | 12.22 | 8.97 | 9.95 | 8.97 | 10.29 | 10.25 | 10.24 | 6.57 | 7.61 | 8.08 | 6.38 | 7.07 | 9.11 | 3.12 |
| *Ethmidium_maculatum 27* | 15.67 | 9.51 | 8.37 | 10.27 | 10.77 | 10.73 | 11.93 | 9.16 | 7.3 | 7.89 | 9.12 | 5.07 | 7.48 | 3.84 |
| *Etrumeus_micropus 28* | 28.82 | 13.34 | 24.33 | 11.52 | 20.25 | 22.54 | 14.35 | 11.8 | 11.58 | 10.93 | 10.11 | 9.72 | 11.52 | 8.65 |
| *Gilchristella_aestuaria 29* | 19.1 | 12.12 | 11.06 | 11.39 | 13.07 | 14.04 | 13.05 | 9.61 | 9.01 | 10.18 | 6.94 | 9.75 | 9.08 | 5.64 |
| *Gudusia_chapra 30* | 15.68 | 11.84 | 11.06 | 11.22 | 12.41 | 14.74 | 10.73 | 8.63 | 7.76 | 13.41 | 11.76 | 22.26 | 17.15 | 8.08 |
| *Harengula_jaguana 31* | 41.13 | 16.92 | 19.87 | 15.49 | 19.39 | 18.11 | 16.56 | 12.06 | 10.76 | 12.39 | 10.06 | 18.98 | 15.2 | 10.31 |
| *Hyperlophus_vittatus 32* | 24.62 | 11.9 | 10.49 | 15.75 | 11.97 | 14.9 | 13.89 | 11.4 | 8.87 | 8.18 | 8.34 | 5.07 | 10.29 | 4.92 |
| *Ilisha_africana 33* | 21.56 | 12.7 | 16.22 | 13.91 | 13.28 | 13.99 | 12.51 | 11.75 | 10.41 | 12.59 | 12.35 | 24.64 | 11.07 | 8.66 |
| *Ilisha_elongata 34* | 20.31 | 11.36 | 12.25 | 10.99 | 15.85 | 12.78 | 11.89 | 10.25 | 8.76 | 10.83 | 10.84 | 14.33 | 13.01 | 10.25 |
| *Jenkinsia_lamprotaenia 35* | 23.43 | 19.99 | 18.15 | 20.05 | 24.15 | 23.86 | 19.79 | 12.92 | 13.92 | 11.83 | 12.27 | 19.5 | 18.47 | 17.61 |
| *Konosirus_punctatus 36* | 24.74 | 12.93 | 13.17 | 14.22 | 12.71 | 14.49 | 14.21 | 9.88 | 10.24 | 9.89 | 8.81 | 12.85 | 10.83 | 4.19 |
| *Lycengraulis_grossidens 37* | 16.64 | 16.65 | 13.14 | 16.39 | 16.95 | 17.57 | 12.52 | 11.6 | 8.55 | 12.58 | 10.13 | 16.32 | 11.29 | 13.9 |
| *Lycothrissa_crocodilus 38* | 15.4 | 12.28 | 15.65 | 12.71 | 16.71 | 12.66 | 12.52 | 10.01 | 11.18 | 9.36 | 10.52 | 22.26 | 13.42 | 13.01 |
| *Microthrissa_congica 39* | 9.04 | 9.08 | 7.59 | 7.72 | 8.61 | 9.19 | 8.02 | 6.96 | 5.31 | 6.86 | 6.96 | 9.02 | 8.81 | 3.95 |
| *Microthrissa_royauxi 40* | 7.29 | 8.31 | 7.98 | 8.19 | 7.95 | 8.21 | 7.07 | 5.31 | 7.29 | 8.38 | 7.38 | 9.91 | 7.97 | 4.55 |
| *Nematalosa_japonica 41* | 26.57 | 11.47 | 13.64 | 15.25 | 11.72 | 15 | 14.69 | 10.17 | 9.89 | 10.92 | 8.21 | 16.66 | 9.28 | 4.93 |
| *Odaxothrissa_losera 42* | 10.58 | 9.38 | 8.77 | 8.1 | 8.27 | 9.43 | 6.95 | 7.05 | 5.6 | 7.23 | 7.69 | 11.97 | 7.98 | 4.31 |
| *Pellona_ditchela 43* | 24.75 | 11.5 | 13.14 | 12.1 | 16.57 | 13.59 | 13.18 | 9.14 | 9.19 | 13.13 | 11.49 | 17.49 | 12.29 | 10.72 |
| *Pellona_flavipinnis 44* | 20.28 | 12.64 | 13.24 | 13.19 | 14.45 | 12.66 | 10.3 | 9.23 | 7.85 | 10.51 | 12.18 | 20.78 | 12.24 | 11.05 |
| *Pellonula_leonensis 45* | 9.91 | 8.19 | 8.35 | 8.1 | 9.29 | 9.43 | 7.55 | 6.85 | 6.02 | 7.55 | 7.31 | 9.07 | 7.11 | 4.19 |
| *Pellonula_vorax 46* | 10.58 | 8.75 | 8.34 | 8.43 | 9.63 | 8.76 | 7.06 | 7.85 | 4.9 | 8.36 | 7.02 | 9.82 | 5.67 | 3.83 |
| *Potamalosa_richmondia 47* | 16.6 | 9.51 | 9.23 | 10.87 | 9.97 | 9.7 | 9.27 | 8.84 | 6.74 | 7.05 | 8.45 | 7.07 | 8.47 | 4.31 |
| *Potamothrissa_acutirostris 48* | 10.83 | 8.98 | 8.75 | 8.97 | 10.64 | 10.89 | 9.94 | 7.65 | 9.07 | 7.55 | 7.3 | 12.85 | 7.95 | 4.31 |
| *Potamothrissa_obtusirostris 49* | 10.86 | 8.43 | 5.67 | 9.28 | 7.95 | 11.09 | 9.05 | 7.85 | 6.87 | 7.22 | 7.67 | 11.33 | 7.29 | 4.79 |
| *Sardina_pilchardus 50* | 47.02 | 16.89 | 20.18 | 18.2 | 14.85 | 20.09 | 17.97 | 13.18 | 10.51 | 15.06 | 11.49 | 20.7 | 13.22 | 7.27 |
| *Sardinella_albella 51* | 14.39 | 12.36 | 10.27 | 12.86 | 9.97 | 14.57 | 12.59 | 9.22 | 7.72 | 6.94 | 7.34 | 13.62 | 10.54 | 3.37 |
| *sardinella_gibbosa 52* | 14.63 | 12.29 | 10.65 | 13.4 | 10.66 | 13.59 | 12.83 | 9.53 | 8.3 | 6.77 | 7.12 | 14.41 | 9.65 | 3.61 |
| *Sardinella_longiceps 53* | 21.2 | 12.41 | 10.29 | 13.57 | 13.1 | 12.78 | 13.22 | 11.35 | 9.23 | 9.26 | 7.47 | 16.32 | 9.64 | 7.14 |
| *Sardinella_maderensis 54* | 14.09 | 10.94 | 11.91 | 11.69 | 9.31 | 12.07 | 10.77 | 8.84 | 7.72 | 6.95 | 6.46 | 5.78 | 8.82 | 3.24 |
| *Sardinops_melanostictus 55* | 37.59 | 14.04 | 14.4 | 15.37 | 14.42 | 16.3 | 16.17 | 12.5 | 8.8 | 10.45 | 9.63 | 9.82 | 12.1 | 5.54 |
| *Setipinna_melanochir 56* | 17.58 | 13.88 | 16.48 | 13.13 | 17.02 | 13.29 | 13.94 | 10.76 | 9.99 | 11.02 | 11.31 | 15.7 | 10.28 | 13.16 |
| *Spratelloides_delicatulus 57* | 22.27 | 19.95 | 16.81 | 20.33 | 21.73 | 26.27 | 19.54 | 12.01 | 13.95 | 11.7 | 10.85 | 24.66 | 16.49 | 13.86 |
| *Spratelloides_gracilis 58* | 27.3 | 20.4 | 18.72 | 18.48 | 21.46 | 23.83 | 17.7 | 13.6 | 14.28 | 11.65 | 11.26 | 21.05 | 14.65 | 14.19 |
| *Sprattus_antipodum 59* | 27.1 | 14.13 | 15.36 | 15.2 | 16.49 | 19.46 | 15.14 | 14.08 | 10.09 | 11.07 | 9.01 | 13.79 | 11.67 | 5.66 |
| *Sprattus_muelleri 60* | 26.5 | 14.11 | 15.81 | 15.68 | 15.74 | 20.14 | 15.67 | 14.2 | 9.65 | 10.38 | 9.09 | 13.79 | 11.67 | 5.79 |
| *Sprattus_sprattus 61* | 25.55 | 14.13 | 15.36 | 16.35 | 15.82 | 18.02 | 16.17 | 12.28 | 10.2 | 8.53 | 8.13 | 9.86 | 11.04 | 5.42 |
| *Stolephorus_chinensis 62* | 22.14 | 15.28 | 16.79 | 15.9 | 19.76 | 17.25 | 12.5 | 11.33 | 10.68 | 15.28 | 10.97 | 15.05 | 12.39 | 13.4 |
| *Stolephorus_waitei 63* | 22.95 | 15.85 | 19.01 | 16.66 | 19.35 | 17.77 | 13.13 | 11.11 | 10.39 | 15.83 | 11.05 | 14.26 | 11.67 | 13.82 |
| *Sundasalanx_mekongensis 64* | 18.73 | 12.66 | 22.2 | 12.77 | 16.07 | 22.74 | 13.47 | 10.98 | 9.59 | 9.28 | 9.71 | 14.77 | 16 | 12.37 |
| *Sundasalanx_praecox 65* | 22.66 | 12.6 | 19.84 | 13.48 | 13.24 | 24.17 | 14.53 | 10.12 | 10.63 | 11.16 | 10.47 | 14.09 | 16.59 | 10.5 |
| *Sundasalanx_sp._Chao_Phraya 66* | 22.08 | 13.62 | 22.58 | 11.98 | 13.87 | 22.23 | 14.78 | 9.71 | 10.33 | 11.67 | 8.73 | 13.98 | 17.1 | 13.05 |
| *Tenualosa_ilisha 67* | 19.47 | 15.57 | 13.12 | 14 | 13.1 | 20.46 | 13.15 | 10.66 | 10.31 | 13.06 | 12.66 | 22.06 | 20.11 | 9.45 |
| *Tenualosa_thibaudeaui 68* | 20.81 | 13.89 | 12.25 | 15.94 | 11.72 | 19.07 | 12.35 | 12.59 | 10 | 13.41 | 12.06 | 22.16 | 19.87 | 10.08 |
| *Tenualosa_toli 69* | 18.7 | 14.95 | 13.12 | 13.87 | 8.95 | 19.66 | 12.1 | 10.46 | 10.28 | 13.8 | 12.95 | 25.59 | 20.24 | 9.31 |
| *Thryssa_baelama 70* | 20.86 | 15.49 | 17.54 | 17.01 | 18.24 | 19.28 | 13.9 | 11.13 | 10.55 | 12.32 | 11.93 | 26.16 | 14.05 | 10.7 |

Linear least square regression with a pairwise distance of 12s rRNA and protein-coding genes

| | ND6 | ND5 | ND4L | ND4 | ND3 | ND2 | ND1 | CYB | CO3 | CO2 | CO1 | ATP8 | ATP6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Correlation coefficient | 0.323 | 0.7 | 0.585 | 0.567 | 0.724 | 0.57 | 0.466 | 0.465 | 0.649 | 0.676 | 0.665 | 0.608 | 0.597 |
| R square | 0.105 | 0.49 | 0.342 | 0.322 | 0.524 | 0.325 | 0.218 | 0.216 | 0.421 | 0.457 | 0.443 | 0.37 | 0.357 |

**Table 6.S2** Folding energy (ΔG), Normalized free energy -ΔG(kcal/mol)/ Length(bp) for 22 *S. longiceps* mitochondrial tRNA genes and its comparison with predicted secondary structures of conserved sequence blocks of clupeoids mitochondrial control region with repeat units.

| tRNA (DNA) | ΔG(kcal/mol) | Length(bp) | Normalized free energy - ΔG(kcal/mol)/ Length(bp) |
|---|---|---|---|
| tRNA-Ala | -10.77 | 69 | -0.16 |
| tRNA-Arg | -16.3 | 69 | -0.24 |
| tRNA-Asn | -10.12 | 73 | -0.14 |
| tRNA-Asp | -10.37 | 69 | -0.15 |
| tRNA-Cys | -21.7 | 66 | -0.33 |
| tRNA-Gln | -16.21 | 71 | -0.23 |
| tRNA-Glu | -6.1 | 69 | -0.09 |
| tRNA-Gly | -20.3 | 71 | -0.29 |
| tRNA-His | -14.6 | 69 | -0.21 |
| tRNA-Ile | -30.31 | 72 | -0.42 |
| tRNA-Leu | -20.5 | 75 | -0.27 |
| tRNA-Leu | -27.4 | 72 | -0.38 |
| tRNA-Lys | -19.6 | 74 | -0.27 |
| tRNA-Met | -16.24 | 69 | -0.24 |
| tRNA-Phe | -12.34 | 63 | -0.20 |
| tRNA-Pro | -17.1 | 70 | -0.24 |
| tRNA-Ser | -19.2 | 68 | -0.28 |
| tRNA-Ser | -11.31 | 67 | -0.17 |
| tRNA-Thr | -28.2 | 72 | -0.39 |
| tRNA-Trp | -9.07 | 70 | -0.13 |
| tRNA-Tyr | -15.96 | 71 | -0.23 |
| tRNA-Val | -20.3 | 72 | -0.28 |
| Species name | Average ΔG(kcal/mol) | Average Length(bp) | Normalized free energy - ΔG(kcal/mol)/ Length(bp) |
| CSB 3 | -11.1 | 41 | -0.27 |
| CSB 2 | -6.29 | 32 | -0.20 |
| CSB 1 | -0.26.32 | 95 | -0.28 |
| CSB D | -0.24 | 101 | -0.24 |
| TAS | -14.72 | 64 | -0.23 |

**Table 6.S3** Folding energy (ΔG), Normalized free energy -ΔG(kcal/mol)/ Length(bp) for 22 *S. longiceps* mitochondrial tRNA genes and its comparison with predicted secondary structures of highly variable regions of clupeoids mitochondrial control region with repeat units.

| tRNA (DNA) | ΔG(kcal/mol)/Length(bp) | Normalized free energy -ΔG(kcal/mol)/ Length(bp) |
|---|---|---|
| tRNA-Ala | -10.77 /69 | -0.16 |
| tRNA-Arg | -16.3 /69 | -0.24 |
| tRNA-Asn | -10.12 /73 | -0.14 |
| tRNA-Asp | -10.37 /69 | -0.15 |
| tRNA-Cys | -21.7 /66 | -0.33 |
| tRNA-Gln | -16.21 /71 | -0.23 |
| tRNA-Glu | -6.1 /69 | -0.09 |
| tRNA-Gly | -20.3 /71 | -0.29 |
| tRNA-His | -14.6 /69 | -0.21 |
| tRNA-Ile | -30.31 /72 | -0.42 |
| tRNA-Leu | -20.5 /75 | -0.27 |
| tRNA-Leu | -27.4 /72 | -0.38 |
| tRNA-Lys | -19.6 /74 | -0.27 |
| tRNA-Met | -16.24 /69 | -0.24 |
| tRNA-Phe | -12.34 /63 | -0.20 |
| tRNA-Pro | -17.1 /70 | -0.24 |
| tRNA-Ser | -19.2 /68 | -0.28 |
| tRNA-Ser | -11.31 /67 | -0.17 |
| tRNA-Thr | -28.2 /72 | -0.39 |
| tRNA-Trp | -9.07 /70 | -0.13 |
| tRNA-Tyr | -15.96 /71 | -0.23 |
| tRNA-Val | -20.3 /72 | -0.28 |
| Species name | ΔG(kcal/mol) Length(bp) | |
| *Alosa_alosa* | ΔG = -100.63 kcal/mol /532 | -0.19 |
| *Alosa_pseudoharengus* | ΔG = -84.63 kcal/mol /485 | -0.17 |
| *Anchoviella_sp._LBP_2297* | ΔG = -96.25 kcal/mol /493 | -0.2 |
| *Brevoortia_tyrannus* | ΔG = -90.64 kcal/mol /527 | -0.17 |
| *Clupea_harengus* | ΔG = -114.88 kcal/mol /540 | -0.21 |
| *Clupea_pallasii   -0.2* | ΔG = -106.52 kcal/mol /540 | |
| *Clupeichthys_aesarnensis* | ΔG = -97.75 kcal/mol /569 | -0.17 |
| *Clupeichthys_goniognathus* | ΔG = -107.95 kcal/mol /553 | -0.2 |
| *Clupeichthys_perakensis* | ΔG = -90.35 kcal/mol /513 | -0.18 |
| *Clupeoides_borneensis* | ΔG = -92.95 kcal/mol /505 | -0.18 |
| *Clupeoides_sp._Chao_Phraya* | ΔG = -104.66 kcal/mol /479 | -0.22 |
| *Clupeonella_cultriventris* | ΔG = -91.48 kcal/mol /491 | -0.19 |
| *Coilia_lindmani* | ΔG = -117.24 kcal/mol /676 | -0.17 |
| *Coilia_nasus* | ΔG = -135.31 kcal/mol /737 | -0.18 |
| *Dorosoma_petenense* | ΔG = -109.95 kcal/mol /625 | -0.18 |
| *Ehirava_fluviatilis* | ΔG = -100.68 kcal/mol /520 | -0.19 |
| *Engraulis_encrasicolus* | ΔG = -103.89 kcal/mol /506 | -0.21 |
| *Engraulis_japonicus* | ΔG = -99.51 kcal/mol /506 | -0.2 |
| *Escualosa_thoracata* | ΔG = -85.31 kcal/mol /519 | -0.16 |
| *Ethmalosa_fimbriata* | ΔG = -215.62 kcal/mol /721 | -0.3 |
| *Ethmidium_maculatum* | ΔG = -105.64 kcal/mol /577 | -0.18 |
| *Gudusia_chapra* | ΔG = -65.79 kcal/mol /390 | -0.17 |
| *Harengula_jaguana* | ΔG = -126.31 kcal/mol /590 | -0.21 |
| *Ilisha_africana* | ΔG = -78.89 kcal/mol /482 | -0.16 |
| *Ilisha_elongata* | ΔG = -113.55 kcal/mol /631 | -0.18 |
| *Jenkinsia_lamprotaenia* | ΔG = -126.22 kcal/mol /633 | -0.2 |
| *Konosirus_punctatus* | ΔG = -86.66 kcal/mol /524 | -0.17 |
| *Lycengraulis_grossidens* | ΔG = -101.94 kcal/mol /505 | -0.2 |
| *Microthrissa_congica* | ΔG = -135.08 kcal/mol /558 | -0.24 |
| *Microthrissa_royauxi* | ΔG = -126.91 kcal/mol /562 | -0.23 |
| *Nematalosa_japonica* | ΔG = -94.82 kcal/mol /441 | -0.22 |
| *Odaxothrissa_losera* | ΔG = -97.96 kcal/mol /555 | -0.18 |
| *Pellona_ditchela* | ΔG = -108.51 kcal/mol /629 | -0.17 |
| *Pellona_flavipinnis* | ΔG = -90.36 kcal/mol /576 | -0.16 |
| *Pellonula_leonensis* | ΔG = -126.61 kcal/mol /558 | -0.23 |
| *Pellonula_vorax* | ΔG = -150.14 kcal/mol /557 | -0.27 |
| *Potamalosa_richmondia* | ΔG = -90.55 kcal/mol /528 | -0.17 |
| *Potamothrissa_acutirostris* | ΔG = -107.66 kcal/mol /559 | -0.19 |
| *Potamothrissa_obtusirostris* | ΔG = -114.35 kcal/mol /559 | -0.2 |
| *Sardina_pilchardus* | ΔG = -71.67 kcal/mol /351 | -0.2 |
| *Sardinella_albella* | ΔG = -86.42 kcal/mol /473 | -0.18 |
| *sardinella_gibbosa* | ΔG = -92.90 kcal/mol /472 | -0.2 |
| *Sardinella_longiceps1* | ΔG = -84.94 kcal/mol /432 | -0.2 |
| *Sardinella_maderensis* | ΔG = -100.49 kcal/mol /474 | -0.21 |
| *Sardinops_melanostictus* | ΔG = -130.42 kcal/mol /605 | -0.22 |
| *Sardinella_longiceps2* | ΔG = -81.24 kcal/mol /512 | -0.16 |
| *Spratelloides_delicatulus* | ΔG = -83.49 kcal/mol /467 | -0.18 |
| *Spratelloides_gracilis* | ΔG = -89.37 kcal/mol /467 | -0.19 |
| *Sprattus_antipodum* | ΔG = -106.46 kcal/mol /543 | -0.2 |
| *Sprattus_muelleri* | ΔG = -106.63 kcal/mol /563 | -0.19 |
| *Sprattus_sprattus* | ΔG = -97.74 kcal/mol /506 | -0.19 |
| *Stolephorus_chinensis* | ΔG = -91.87 kcal/mol /438 | -0.21 |
| *Sundasalanx_mekongensis* | ΔG = -105.22 kcal/mol /594 | -0.18 |
| *Sundasalanx_praecox* | ΔG = -114.25 kcal/mol /605 | -0.19 |
| *Sundasalanx_sp._Chao_Phraya* | ΔG = -103.15 kcal/mol /606 | -0.17 |
| *Tenualosa_ilisha* | ΔG = -117.71 kcal/mol /621 | -0.19 |
| *Tenualosa_toli* | ΔG = -68.77 kcal/mol /392 | -0.18 |
| *Thryssa_baelama* | ΔG = -65.04 kcal/mol /373 | -0.17 |

**Table 6.S4** The effective number of codon (ENc) and GC3 content of merged protein-coding genes of all considered clupeoid fishes.

| Species Name | Enc | GC3 |
|---|---|---|
| Tenualosa_thibaudeaui | 55.965 | 16.25 |
| Tenualosa_ilisha | 55.7446 | 16.44 |
| Tenualosa_toli | 54.9317 | 16.43 |
| Gudusia_chapra | 48.128 | 13.27 |
| Potamothrissa_obtusirostris | 50.1956 | 13.40 |
| Potamothrissa_acutirostris | 50.6384 | 13.38 |
| Microthrissa_congica | 50.0933 | 14.17 |
| Pellonula_vorax | 50.5181 | 13.31 |
| Pellonula_leonensis | 49.5945 | 13.42 |
| Odaxothrissa_losera | 49.1052 | 13.44 |
| Microthrissa_royauxi | 49.9484 | 14.17 |
| Ethmalosa_fimbriata | 51.314 | 14.08 |
| Dorosoma_cepedianum | 53.1346 | 15.56 |
| Dorosoma_petenense | 51.7172 | 14.04 |
| Sardinella_maderensis | 54.4227 | 16.31 |
| Sardinella_albella | 55.35 | 17.91 |
| sardinella_gibbosa | 56.0372 | 17.88 |
| Harengula_jaguana | 57.3476 | 17.92 |
| Sardinella_longiceps | 55.8302 | 16.32 |
| Nematalosa_japonica | 56.8587 | 16.64 |
| Clupanodon_thrissa | 55.8765 | 16.15 |
| Konosirus_punctatus | 57.3049 | 17.26 |
| Escualosa_thoracata | 56.2222 | 18.50 |
| Sardina_pilchardus | 58.1035 | 18.33 |
| Sardinops_melanostictus | 58.0235 | 17.42 |
| Brevoortia_tyrannus | 55.7022 | 12.29 |
| Alosa_alosa | 52.9578 | 9.77 |
| Alosa_pseudoharengus | 52.3952 | 15.70 |
| Clupeichthys_goniognathus | 53.8746 | 14.58 |
| Clupeichthys_aesarnensis | 53.0379 | 15.88 |
| Clupeichthys_perakensis | 52.0973 | 15.02 |
| Clupeoides_sp._Chao_Phraya | 46.4359 | 13.47 |
| Clupeoides_borneensis | 48.4692 | 14.39 |
| Sundasalanx_praecox | 51.6948 | 13.49 |
| Sundasalanx_sp._Chao_Phraya | 50.7684 | 12.89 |
| Sundasalanx_mekongensis | 49.6757 | 12.66 |
| Ehirava_fluviatilis | 48.6599 | 11.02 |
| Gilchristella_aestuaria | 53.4373 | 14.26 |
| Clupeonella_cultriventris | 57.1558 | 10.95 |
| Clupea_harengus | 56.776 | 16.42 |
| Clupea_pallasii | 56.9426 | 16.02 |
| Sprattus_sprattus | 57.3403 | 15.99 |
| Sprattus_muelleri | 57.1911 | 17.12 |
| Sprattus_antipodum | 57.1544 | 17.21 |
| Potamalosa_richmondia | 49.9592 | 13.55 |
| Hyperlophus_vittatus | 56.138 | 15.05 |
| Ethmidium_maculatum | 52.6757 | 16.05 |
| Jenkinsia_lamprotaenia | 55.1271 | 18.83 |
| Spratelloides_delicatulus | 53.839 | 20.17 |
| Spratelloides_gracilis | 57.3919 | 16.37 |
| Etrumeus_micropus | 53.54 | 13.33 |
| Ilisha_africana | 47.9898 | 12.85 |
| Pellona_flavipinnis | 47.0482 | 14.65 |
| Ilisha_elongata | 47.1438 | 13.74 |
| Pellona_ditchela | 47.1092 | 13.51 |
| Anchoviella_sp._LBP_2297 | 50.2627 | 13.65 |
| Lycengraulis_grossidens | 51.4366 | 12.27 |
| Amazonsprattus_scintilla | 56.2435 | 14.60 |
| Engraulis_encrasicolus | 56.3532 | 14.57 |
| Engraulis_japonicus | 56.6461 | 14.56 |
| Stolephorus_chinensis | 51.5577 | 11.83 |
| Stolephorus_waitei | 52.5362 | 12.13 |
| Lycothrissa_crocodilus | 47.3335 | 13.19 |
| Setipinna_melanochir | 49.4808 | 15.59 |
| Coilia_reynaldi | 50.1849 | 11.76 |
| Thryssa_baelama | 54.7704 | 15.07 |
| Coilia_lindmani | 47.6599 | 11.84 |
| Coilia_ectenes | 47.5468 | 15.14 |
| Coilia_nasus | 47.6204 | 12.25 |
| Denticeps_clupeoides | 46.5509 | 13.41 |

**Table 6.S5** List of species used in this study.

| Classification | | Species | Origin | Accession Nos. |
|---|---|---|---|---|
| Otocephala | | | | |
| Order Clupeiformes | | | | |
| Family Denticipitidae | | *Denticeps clupeoides* Clausen | Bénin, West Africa | AP007276 |
| Family Clupeidae | Subfamily: | *Sardinops melanostictus* (Temminck & Schlegel) | Japan, Northwest Pacific | AB032554 |
| | Clupeinae | *Clupea pallasii* (Valenciennes) | Japan, Northwest Pacific | AP009134 |
| | | *Clupea harengus* (Linnaeus) | North Atlantic | AP009133 |
| | | *Sprattus sprattus* (Linnaeus) | North Atlantic | AP009234 |
| | | *Sprattus muelleri* (Klunzinger) | South Island, New Zealand | AP011607 |
| | | *Sprattus antipodum* (Hector) | South Island, New Zealand | AP011608 |
| | | *Escualosa thoracata* (Valenciennes) | Bangkok, Thailand | AP011601 |
| | | *Clupeonella cultriventris* (Nordmann) | Caspian Sea | AP009615 |
| | | *Harengula jaguana* (Poey) | West Atlantic | AP011592 |
| | | *Sardinella albella* (Valenciennes) | Madagascar | AP011605 |
| | | *Sardinella maderensis* (Lowe) | Near Dakar, Sénégal | AP009143 |
| | | *Sardinella gibbosa* | Indian Ocean | |
| | | *Sardinella longiceps* | Indian Ocean | |
| | | *Sardina pilchardus* (Walbaum) | Europa | AP009233 |
| | Alosinae | *Ethmalosa fimbriata* (Bowdich) | Near Dakar, Sénégal | AP009138 |
| | | *Brevoortia tyrannus* (Latrobe) | North America | AP009618 |
| | | *Ethmidium maculatum* (Valenciennes) | East Pacific, South America | AP011602 |
| | | *Tenualosa ilisha* (Hamilton-Buchanan) | Calcutta, India | AP011610 |
| | | *Tenualosa thibaudeaui* (Durand) | Vientian, northern Laos | AP011604 |
| | | *Tenualosa toli* (Valenciennes) | Calcutta, India | AP011600 |
| | | *Gudusia chapra* (Hamilton-Buchanan) | Calcutta, India | AP011603 |
| | | *Alosa alosa* (Linnaeus) | Vilaine River, France | AP009131 |
| | | *Alosa pseudoharengus* (Wilson) | North America | AP009132 |
| | Dussumeriinae | *Spratelloides delicatulus* (Bennett) | Japan | AP009144 |
| | | *Spratelloides gracilis* (Temminck & Schlegel) | Japan | AP009145 |
| | | *Etrumeus micropus* (Temminck & Schlegel) | Japan | AP009139 |
| | | *Jenkinsia lamprotaenia* (Gosse) | West Africa | AP006230 |
| | Dorosomatinae | *Dorosoma petenense* (Gunther) | North America | AP009136 |
| | | *Dorosoma cepedanium* (LeSueur) | North America | DQ536426 |
| | | *Konosirus punctatus* (Temminck & Schlegel) | Tokyo, Japan, 2007 | AP011612 |
| | | *Clupanodon thryssa* (Linnaeus) | Northwest Pacific | JX075099 |
| | | *Nematalosa japonica* Regan | Okinawa, Japan, 2004 | AP009142 |
| | Pellonulinae | *Pellonula leonensis* (Boulenger) | Ouémé R., Bénin, 2003 | AP009232 |
| | | *Pellonula vorax* (Regan) | Nkomi R., Gabon, 2001 | AP009231 |
| | | *Potamothrissa obtusirostris* (Boulenger) | Lower Congo, 2006 | AP011599 |
| | | *Potamothrissa acutirostris* (Boulenger) | Lower Congo, 2006 | AP011597 |
| | | *Odaxothrissa losera* (Boulenger) | Lower Congo, 2006 | AP011595 |
| | | *Microthrissa royauxi* (Boulenger) | Lower Congo, 2006 | AP011596 |
| | | *Microthrissa congica* (Regan, 1917) | Lower Congo, 2006 | AP011598 |
| | | *Clupeichthys aesarnensis* (Wongratana) | Chao Phraya R., Thailand | AP011584 |
| | | *Clupeichthys perakensis* (Herre) | Thailand | AP011585 |
| | | *Clupeichthys gogniognathus* (Fowler) | Chao Phraya R., Thailand | AP011589 |
| | | *Clupeoides borneensis* (Bleeker) | Chao Phraya R., Thailand | AP011586 |
| | | *Clupeoides sp.*"Chao Phraya" | Chao Phraya R., Thailand | AP011587 |
| | | *Ehirava fluviatilis* (Deraniyagala) | India | AP011588 |
| | | *Gilchristella aestuarius* (Gilchrist) | Kariega estuary? South Africa [catalog number: SAIAB46983] | AP011606 |
| | | *Potamalosa richmondia* (Macleay) | Camden Haven River, Australia [voucher: I.31259-001] | AP011594 |
| | | *Hyperlophus vittatus* (Castelnau) | Western Port, Rhyll, Australia [voucher: NMV A 26036-005] | AP011593 |
| Family Engraulidae | | *Engraulis japonicus* (Temminck & Schlegel) | Japan | AB040676 |
| | | *Engraulis encrasicolus* (Linnaeus) | Northeast Atlantic | AP009137 |
| | | *Coilia nasus* (Temminck & Schlegel) | Japan | AP009135 |
| | | *Coilia ectenes* (Jordan & Seale) | China | JX625133 |
| | | *Coilia lindmani* (Bleeker) | Lake Tonle Sap, Cambodia | AP011558 |
| | | *Coilia reynaldi* (Valenciennes) | Calcutta, India | AP011559 |
| | | *Lycothrissa crocodilus* (Bleeker) | Lake Tonle Sap, Cambodia | AP011562 |
| | | *Setipinna melanochir* (Bleeker) | Lake Tonle Sap, Cambodia | AP011565 |
| | | *Thryssa baelama* (Forsskål) | Indonesia | AP009616 |

|  |  |  |  |  |
|---|---|---|---|---|
|  |  | *Stolephorus cf chinensis* | Bangkok, Thailand | AP011566 |
|  |  | *Stolephorus cf waitei* | Calcutta, India | AP011567 |
|  |  | *Anchiovella sp.* | South America | AP011557 |
|  |  | *Lycengraulis grossidens* (Agassiz) | South America | AP011563 |
|  |  | *Amazonsprattus scintilla* (Roberts) | South America | AP009617 |
| Family Pristigasteridae |  | *Ilisha elongata* (Bennett) | Japan | AP009141 |
|  |  | *Ilisha africana* (Bloch) | East Atlantic | AP009140 |
|  |  | *Pellona flavipinnis* (Valenciennes) | South America | AP009619 |
|  |  | *Pellona ditchela* (Valenciennes) | Bangkok, Thailand, SL | AP011609 |
| Family Sundasalangidae |  | *Sundasalanx mekongensis* (Britz & Kottelat) | Mekong R., Cambodia | AP006232 |
|  |  | *Sundasalanx praecox* (Roberts) | Thailand, | AP011591 |
|  |  | *Sundasalanx sp.* | Bangkok, Thailand | AP011590 |

## 5. References

1. Anderson S, Bankier AT, Barrell BG, Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465
2. Baker EP, Peris D, Moriarty RV, Li XC, Fay JC, Hittinger CT (2019) Mitochondrial DNA and temperature tolerance in lager yeasts. *Sci Adv* 5(1):eaav1869
3. Baker LA, Watt IN, Runswick MJ, Walker JE, Rubinstein JL (2012) Arrangement of subunits in intact mammalian mitochondrial ATP synthase determined by cryo-EM. *P Natl A Sci* 109(29):11675-80
4. Ballard JWO, Pichaud N (2014) Mitochondrial DNA: more than an evolutionary bystander. *Funct Ecol* 28: 218-231
5. Beckstead WA, Ebbert MT, Rowe MJ, McClellan DA (2009) Evolutionary pressure on mitochondrial cytochrome b is consistent with a role of CytbI7T affecting longevity during caloric restriction. *Plos One* 4(6):e5836
6. Bloom DD, Lovejoy NR (2012) Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (family Engraulidae). *J Evol Biol* 25(4): 701-715
7. Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27(8):1767-1780
8. Broughton RE, Dowling TE (1994) Length variation in mitochondrial DNA of the minnow Cyprinella spiloptera. *Genetics* 138:179-190
9. Broughton RE, Dowling TE (1997) Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of the minnow Cyprinella spiloptera. *Mol Biol Evo* 14:1187-1196
10. Brown GG, Gadaleta G, Pepe G, Saccone C, Sbisa E (1986) Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J Mol Biol* 192:503–511
11. Bulmer M (1987) Coevolution of codon usage and transfer RNA abundance. *Nature* 325: 728–730
12. Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129: 897–907
13. Burrow AA, Marullo A, Holder LR, Wang YH (2010) Secondary structure formation and DNA instability at fragile site FRA16B. *Nucleic Acids Res* 38:2865-2877
14. Burton RS, Pereira RJ, Barreto FS (2013) Cytonuclear genomic interactions and hybrid breakdown. *Annu Rev Ecol Evol S*. 44:281-302
15. Caballero S, Duchene S, Garavito MF, Slikas B, Baker CS (2015) Initial evidence for adaptive selection on the NADH subunit two of freshwater dolphins by analyses of mitochondrial genomes. *PloS one* 10(5):e0123543
16. Carapelli A, Fanciulli PP, Frati F, Leo C (2019) Mitogenomic data to study the taxonomy of Antarctic springtail species (Hexapoda: Collembola) and their adaptation to extreme environments. *Polar Biology* 1-8
17. Charneski CA, Honti F, Bryant JM, Hurst LD, Feil EJ (2011) Atypical AT skew in Firmicute genomes results from selection and not from mutation. *Plos Genetics* 7(9): e1002283
18. Chen H, Sun S, Norenburg JL, Sundberg P (2014) Mutation and selection cause codon usage and bias in mitochondrial genomes of ribbon worms (Nemertea). *Plos One* 9(1): e85631
19. Chen Y, Carlini DB, Baines JF, Parsch J, Braverman JM, Tanda S, Stephan W (1999) RNA secondary structure and compensatory evolution. Genes Genet Syst 74: 271-286.
20. Cheng SW, Lynch EC, Leason KR, Shapiro BA, Friedman DI (1991) Functional importance of sequence in the stem-loop of a transcription terminator. Science 254:1205-1207. doi.org/10.1126/science.1835546
21. Cheviron ZA, Connaty AD, McClelland GB, Storz JF (2014) Functional genomics of adaptation to hypoxic cold-stress in high-altitude deer mice: transcriptomic plasticity and thermogenic performance. *Evolution* 68(1):48-62
22. Clayton DA (1991) Replication and transcription of vertebrate mitochondrial DNA. Annu Rev Cell Biol 7, 453–478.
23. Consuegra S, John E, Verspoor E, De Leaniz CG (2015) Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. Genet Sel Evol 47(1):1–10
24. Crofts AR (2004a) The cytochrome bc 1 complex: function in the context of structure. *Annu Rev Physiol* 66: 689-733
25. Crofts AR (2004b) Proton-coupled electron transfer at the Qo-site of the bc1 complex controls the rate of ubihydroquinone oxidation. *BBA-Bioenergetics* 1655:77-92
26. D'Souza AR, Minczuk M (2018) Mitochondrial transcription and translation: overview. Essays Biochem 62:309-320. doi.org/10.1042/EBC20170102
27. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A (2008) The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* 9(1):119

28. Dowling DK, Friberg U, Lindell J (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol Evol* 23(10):546-54
29. Dynesius M, Jansson R (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *P Natl A Sci USA* 97:9115–9120
30. Evans DH, Piermarini PM, Choe KP (2005) The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. *Physiol. Rev.* 85(1), 97-177.
31. Fiedorczuk K, Letts JA, Degliesposti G, Kaszuba K, Skehel M, Sazanov LA (2016) Atomic structure of the entire mammalian mitochondrial complex I. *Nature* 538(7625):406-410
32. Fischer C, Koblmuller S, Gully C, Schlotterer C, Sturmbauer C, Thallinger GG (2013). Complete mitochondrial DNA sequences of the threadfin cichlid (*Petrochromis trewavasae*) and the blunthead cichlid (*Tropheus moorii*) and patterns of mitochondrial genome evolution in cichlid fishes. *Plos One* 8(6):e67048
33. Fox TD (2012) Mitochondrial protein synthesis, import, and assembly. Genetics 192:1203-1234. doi.org/10.1534/genetics.112.141267
34. Freeman AR, Machugh DE, Mckeown S, Walzer C, Mcconnell DJ, Bradley DG (2001) Sequence variation in the mitochondrial DNA control region of wild African cheetahs (*Acinonyx jubatus*). Heredity 86:355-362.
35. Galtier N, Nabholz B, Glemin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol* 18(22):4541-4550
36. Ganias K (2014) Biology and ecology of sardines and anchovies. CRC Press
37. Garvin MR, Bielawski JP, Gharrett AJ (2011) Positive Darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon. *Plos One* 6(9):e24127
38. Garvin MR, Bielawski JP, Gharrett AJ (2012) Correction: Positive Darwinian Selection in the Piston That Powers Proton Pumps in Complex I of the Mitochondria of Pacific Salmon. *Plos One* 7(8):e24127
39. Garvin MR, Bielawski JP, Sazanov LA, Gharrett AJ (2015a) Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. *J Zool Syst Evol Res* 53(1):1-17
40. Garvin MR, Thorgaard GH, Narum SR (2015b) Differential expression of genes that control respiration contribute to thermal adaptation in redband trout (Oncorhynchus mykiss gairdneri). *Genome Biol Evol* 7(6):1404-1414
41. Grant, W. Stewart, and Brian W. Bowen (2006) Living in a tilted world: climate change and geography limit speciation in Old World anchovies (Engraulis; Engraulidae). *Biol J Linn Soc* 88(4): 673-689
42. Grande L, Nelson GJ (1985) Interrelationships of fossil and recent anchovies (Teleostei, Engrauloidea) and description of a new species from the Miocene of Cyprus. American Museum of Natural History
43. Gruber AR, Bernhart SH, Lorenz R (2015) The ViennaRNA web services. In: Picardi E (ed) RNA bioinformatics. Humana Press, New York, USA, pp. 307-326. doi.org/10.1007/978-1-4939-2291-8_19
44. Harrisson K, Pavlova A, Gan HM, Lee YP, Austin CM, Sunnucks P (2016) Pleistocene divergence across a mountain range and the influence of selection on mitogenome evolution in threatened Australian freshwater cod species. *Heredity 116*(6): 506
45. Havird JC, Sloan DB (2016) The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. *Mol Biol Evol* 33(12):3042-53
46. Hershberg R, Petrov DA (2008) Selection on codon bias. *Annu Rev Genet* 42:287–299
47. Horan MP, Gemmell NJ, Wolff JN (2013) From evolutionary bystander to master manipulator: the emerging roles for the mitochondrial genome as a modulator of nuclear gene expression. *Eur J Hum Genet* 21(12):1335
48. Howell N, Elson JL, Howell C, Turnbull DM (2007) Relative Rates of Evolutionin the Coding and Control Regions of African mtDNAs. *Mol Biol Evol* 24:2213–2221
49. Hughes LC, Somoza GM, Nguyen BN, Bernot JP, Gonzalez-Castro M, Díaz de Astarloa JM, Ortí G (2017) Transcriptomic differentiation underlying marine-to-freshwater transitions in the South American silversides *Odontesthes argentinensis* and *O. bonariensis* (Atheriniformes). *Ecol Evol* 7(14):5258-5268
50. Hwang PP, Lee TH (2007) New insights into fish ion regulation and mitochondrion-rich cells. *Comp. Biochem. Phys. A* 148(3), 479-497.
51. Hyvarinen AK, Pohjoismaki JL, Reyes A, Wanrooij S, Yasukawa T, Karhunen PJ, Spelbrink JN, Holt IJ, Jacobs HT (2007) The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA. Nucleic Acids Res 35:6458-6474. doi.org/10.1093/nar/gkm676
52. Jacobsen MW, Da Fonseca RR, Bernatchez L, Hansen MM (2016) Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the NADH2 gene in whitefish (Coregonus ssp.). *Mol Phyl Evol* 95:161-170
53. Jamandre BW, Durand JD, Tzeng WN (2014) High sequence variations in mitochondrial DNA control region among worldwide populations of flathead mullet Mugil cephalus. Int J Zool 2014:564105. doi.org/10.1155/2014/564105

54. Jansson R, Dynesius M (2002) The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annu Rev Ecol Syst* 33(1):741–777.
55. Jeffrey LB (1999) Survey and summary animal mitochondrial genomes. *Nucleic Acids Res* 27(8):1767-1780
56. Jia W, Higgs PG (2008) Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. *Mol Biol Evol* 25: 339-351
57. Junge W, Nelson N (2005) Nature's rotary electromotors. *Science*. 308(5722):642-644
58. Kanaya S, Yamada Y, Kudo Y, Ikemura T (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 238(1), 143-155.
59. Kaneko T, Watanabe S, Lee KM (2008) Functional morphology of mitochondrion-rich cells in euryhaline and stenohaline teleosts. Tokyo: Terrapub.
60. Katz L, Burge CB (2003) Widespread selection for local RNA secondary structure in coding regions of bacterial genes. Genome Res. 13:2042–2051.
61. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12): 1647-1649
62. Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press
63. Kolling DR, Samoilova RI, Holland JT, Berry EA, Dikanov SA, Crofts AR (2003) Exploration of ligands to the Qi site semiquinone in the bc1 complex using high-resolution EPR. *J Biol Chem* 278(41):39747-54
64. Koonin EV (2005) Systemic determinants of gene evolution and function. *Mol Syst Biol* 1: 2005.0021
65. Kotlar D, Lavner Y (2006) The action of selection on codon bias in the human genome is related to frequency, complexity, and chronology of amino acids. *BMC genomics* 7(1), 67.
66. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870-1874
67. Kunstner A, Nabholz B, Ellegren H (2011) Significant selective constraint at 4-fold degenerate sites in the avian genome and its consequence for detection of positive selection. *Genome Biol Eevol* 3: 1381-1389
68. Lajbner Z, Pnini R, Camus MF, Miller J, Dowling DK (2018) Experimental evidence that thermal selection shapes mitochondrial genome evolution. *Sci Rep-UK* 8:1-12
69. Lau WC, Rubinstein JL (2012) Subnanometre-resolution structure of the intact Thermus thermophilus H+-driven ATP synthase. *Nature* 481(7380):214
70. Lavoue S, Konstantinidis P, Chen WJ (2014) Progress in clupeiform systematics. In: Ganias K (eds) Biology and ecology of sardines and anchovies. CRC Press, New Yorks, USA. pp. 3-42
71. Lavoue S, Miya M, Musikasinthorn P, Chen WJ, Nishida M (2013) Mitogenomic evidence for an Indo-west pacific origin of the clupeoidei (Teleostei: Clupeiformes). *Plos One* 8(2):e56485
72. Lavoue S, Miya M, Saitoh K, Ishiguro NB, Nishida M (2007) Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Mol Phylogenet Evol* 43(3):1096-1105
73. Lee WJ, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. J Mol Evol 41:54-66.
74. Li Y, Park JS, Deng JH, Bai Y (2006) Cytochrome c oxidase subunit IV is essential for assembly and respiratory function of the enzyme complex. *J Bioenerg Biomembr* 38(5-6): 283–291
75. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452
76. Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362:709-715.
77. Marshall HD, Coulson MW, Carr SM (2008) Near neutrality, rate heterogeneity, and linkage govern mitochondrial genome evolution in Atlantic cod Gadus morhua) and other gadine fish. *Mol Biol* 26(3), 579–589
78. Marshall WS, Grosell M (2006) Ion transport, osmoregulation, and acid-base balance. *Physiol. Fish.* 3, 177-230.
79. McLean MJ, Wolfe KH, Devine KM (1998) Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. J Mol Evol 47:691–696.
80. Melo-Ferreira J, Vilela J, Fonseca MM, Da Fonseca RR, Boursot P, Alves PC (2014) The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome Biol Evol* 6(4):886–896

81. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI (2003) Natural selection shaped regional mtDNA variation in humans. *P Natl Acad Sci USA* 100:171-176
82. Mita S, Rizzuto R, Moraes CT, Shanske S, Arnaudo E, Fabrizi GM, Koga Y, DiMauro S, Schon EA (1990) Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. Nucleic Acids Res 18:561–567. doi.org/10.1093/nar/18.3.561
83. Miya M, Nishida M (2015) The mitogenomic contributions to molecular phylogenetics and evolution of fishes: a 15-year retrospect. Ichthyol Res 62:29-71.
84. Montoya J, Christianson T, Levens D, Rabinowitz M, Attardi G (1982) Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. P Natl Acad Sci USA 79:7195-7199. doi.org/10.1073/pnas.79.23.7195
85. Morales HE, Pavlova A, Amos N, Major R, Bragg J, Kilian A *et al*. (2016). Mitochondrial-nuclear interactions maintain a deep mitochondrial split in the face of nuclear gene flow. *BioRxiv* 095596
86. Morris MR, Richard R, Leder EH, Barrett RD, Aubin-Horth N, Rogers SM (2014) Gene expression plasticity evolves in response to colonization of freshwater lakes in threespine stickleback. *Mol Ecol* 23(13): 3226-3240
87. Murakami H, Ota A, Simojo H, Okada M, Ajisaka R, Kuno S (2002) Polymorphisms in control region of mtDNA relates to individual differences in endurance capacity or trainability. Jpn J Physiol 52:247–256. https://doi.org/10.2170/jjphysiol.52.247
88. Nabholz B, Kunstner A, Wang R, Jarvis ED, Ellegren H (2011) Dynamic evolution of base composition: causes and consequences in avian phylogenomics. *Mol Biol Evol* 28(8):2197-2210
89. Necsulea A, Lobr JR (2007) A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol Biol Evol* 24;2169-2179
90. Nelson GJ (1971) Paraphyly and polyphyly: redefinitions. *Syst Biol* 20(4):471-472
91. Nelson G (1970) The hyobranchial apparatus of teleostean fishes of the families Engraulidae and Chirocentridae. *Am Mus Novit* 2410:1–30.
92. Nelson G (1983) Anchoa-argentivittata, with notes on other Eastern Pacific anchovies and the Indo-Pacific genus Encrasicholina. *Copeia* 1983: 48–54.
93. Nelson G (1984) Identity of the anchovy Hildebrandichthys-setiger with notes on relationships and biogeography of the genera Engraulis and Cetengraulis. *Copeia* 1984: 422–427.
94. Nelson G (1986) Identity of the anchovy Engraulis-clarki with notes on the species-groups of Anchoa. *Copeia* 1986: 891–902.
95. Nicholls TJ, Minczuk M (2014) In D-loop: 40 years of mitochondrial 7S DNA. Exp Gerontol 56:175-181. doi.org/10.1016/j.exger.2014.03.027
96. Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290(5806), 470-474.
97. Osada N, Akashi H (2011) Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. *Mol Biol Evol* 29(1):337-46
98. Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R., Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363. doi.org/10.1038/ng0497-363
99. Pereira F, Soares P, Carneiro J, Pereira L, Richards MB, Samuels DC, Amorim A (2008) Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region. Mol Biol Evol 25:2759–2770 doi.org/10.1093/molbev/msn225
100. Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358.
101. Pond SLK, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21(10):2531–2533
102. Reyes A, Gissi C, Pesole G, Saccone C (1998) Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol Biol Evol 15:957– 966.
103. Rice PA, Correll CC (2008) Protein-nucleic acid interactions: structural biology. The Royal society of chemistry, Cambridge, UK
104. Samuels DC, Schon EA, Chinnery PF (2004) Two direct repeats cause most human mtDNA deletions. Trends Genet 20:393-398. doi.org/10.1016/j.tig.2004.07.003
105. Satoh TP, Sato Y, Masuyama N, Miya M, Nishida M (2010) Transfer RNA gene arrangement and codon usage in vertebrate mitochondrial genomes: a new insight into gene order conservation. BMC genomics 11: 479.
106. Sazanov LA (2015) A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat Rev Mol Cell Biol* 16(6):375

107. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modelling server. *Nucleic Acids Res* 31(11):3381–3385
108. Scott GR, Schulte PM, Egginton S, Scott AL, Richards JG, Milsom WK (2010) Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol Biol Evol* 28(1):351–363
109. Scott RA (1995) Functional significance of cytochrome c oxidase structure. *Structure* 3(10):981-6
110. Sebastian W, Sukumaran S, Zacharia PU, Gopalakrishnan A (2017) The complete mitochondrial genome and phylogeny of Indian oil sardine, *Sardinella longiceps* and Goldstripe *Sardinella*, *Sardinella gibbosa* from the Indian Ocean. *Conserv Genet Resour* 10(4): 735–739
111. Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem 66, 409-435.
112. Shingu-Vazquez M, Camacho-Villasana Y, Sandoval-Romero L, Butler CA, Fox TD, Perez-Martínez X (2010) The carboxyl-terminal end of Cox1 is required for feedback assembly regulation of Cox1 synthesis in *Saccharomyces cerevisiae* mitochondria. *J Biol Chem* 285(45): 34382-34389
113. Slomovic S, Laufer D, Geiger D, Schuster G (2005) Poly- adenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark. Mol Cel Biol 25:6427–6435. doi.org/10.1128/MCB.25.15.6427-6435.2005
114. Spies M, Smith BO (2017) Protein–nucleic acids interactions: new ways of connecting structure, dynamics and function. Biophys Rev 9:289-291. doi.org/10.1007/s12551-017-0284-4
115. Stager M, Cerasale DJ, Dor R, Winkler DW, Cheviron ZA (2014) Signatures of natural selection in the mitochondrial genomes of Tachycineta swallows and their implications for latitudinal patterns of the pace of life. *Gene* 546(1):104-111
116. Stier A, Bize P, Roussel D, Schull Q, Massemin S, Criscuolo F (2014) Mitochondrial uncoupling as a regulator of life history trajectories in birds: An experimental study in the zebra finch. *J Exp Biol* 217(19):3579-3589
117. Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. *P Natl Acad Sci* 85(8):2653-2657
118. Sueoka N (1999) Two aspects of DNA base composition: G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C. *J Mol Evol* 49(1): 49-62
119. Suissa S, Wang Z, Poole J, Wittkopp S, Feder J, Shutt TE, Wallace DC, Shadel GS, Mishmar D (2009) Ancient mtDNA genetic variants modulate mtDNA transcription and replication. Plos Genetics 5:e1000474. https://doi.org/10.1371/journal.pgen.1000474
120. Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. BBA-Bioenergetics 1410:103-123. doi.org/10.1016/S0005-2728(98)00161-3
121. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.
122. Teacher AG, Andre C, Merila J, Wheat CW (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. BMC Evol Biol 12(1):248
123. Teske PR, Sandoval-Castillo J, Golla TR, Emami-Khoyi A, Tine M, von der Heyden S, Beheregaray LB (2019) Thermal selection as a driver of marine ecological speciation. *Proceedings of the Royal Society B* 286(1896): 20182023
124. Toews DP, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol* 21(16):3907-3930
125. Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S (1995) Structures of metal sites of oxidized bovine heart cytochrome c oxidase at 2.8 A. *Science* 269(5227):1069-1074
126. Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S (1996). The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* 272(5265): 1136-1144
127. Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411–5421
128. Walker JE (2013) The ATP synthase: the understood, the uncertain and the unknown. *Biochem Soc T J* 41 (1) 1-16
129. Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G *et al*. (2005) Functionalgenomic analysis of the rates of protein evolution. *Proc Natl Acad Sci USA* 102:5483–5488
130. Wang L, Zhou X, Nie L (2011) Organization and variation of mitochondrial DNA control region in pleurodiran turtles. Zoologia (Curitiba) 28:495-504.
131. Wenz T, Covian R, Hellwig P, MacMillan F, Meunier B, Trumpower BL, Hunte C (2007) Mutational analysis of cytochrome b at the ubiquinol oxidation site of yeast complex III. *J Biol Chem* 282(6):3977-88

132. Whitehead A, Roach JL, Zhang S, Galvez F (2012) Salinity-and population-dependent genome regulatory response during osmotic acclimation in the killifish (*Fundulus heteroclitus*) gill. *J Exp Biol* 215(8): 1293-1305
133. Wolff JN, Ladoukakis ED, Enríquez JA, Dowling DK (2014) Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Philos T R Soc B* 369(1646):20130443
134. Woolley S, Johnson J, Smith M J, Crandall KA, McClellan DA (2003) TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19(5):671–672
135. Wright BE (2000) A biochemical mechanism for nonrandom mutations and evolution. J Bacteriol 182:2993–3001. doi.org/10.1128/JB.182.11.2993-3001.2000
136. Wright F (1990) The 'effective number of codons' used in a gene. *Gene* 87(1):23-29
137. Xia X (1996) Maximizing transcription efficiency causes codon usage bias. *Genetics* 144(3), 1309-1320.
138. Xia X (2005) Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. *Gene* 345:13–20
139. Xia X (2013) DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol* 30(7):1720-1728
140. Yasukawa T, Yang MY, Jacobs HT, Holt IJ (2005) A bidirectional origin of replication maps to the major noncoding region of human mitochondrial DNA. Mol Cel 18:651–662. doi.org/10.1016/j.molcel.2005.05.002
141. Yu L, Wang X, Ting N, Zhang Y (2011) Mitogenomic analysis of Chinese snub-nosed monkeys: Evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. *Mitochondrion* 11(3):497-503
142. Zhang X, Wen H, Wang H, Ren Y, Zhao J, Li Y (2017) RNA-Seq analysis of salinity stress–responsive transcriptome in the liver of spotted sea bass (Lateolabrax maculatus). *Plos One* 12(3).
143. Zhou A, Rohou A, Schep DG, Bason JV, Montgomery MG, Walker JE *et al* (2015) Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. *Elife* 4:e10180
144. Zhu J, Vinothkumar KR, Hirst J (2016) Structure of mammalian respiratory complex I. *Nature* 536(7616):354–358
145. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415. doi.org/10.1093/nar/gkg595
146. Zuloaga J, Currie DJ, Kerr JT (2019) The origins and maintenance of global species endemism. *Global Ecol Biogeogr* 28(2):170-183

# Chapter 7

## THE COMPLETE MITOCHONDRIAL GENOME AND PHYLOGENY OF GREEN CHROMIDE, *ETROPLUS SURATENSIS* (Bloch, 1790) FROM INDIAN WATERS

### ABSTRACT

The cichlid fish, Green Chromide [*Etroplus suratensis* (Bloch 1790)] is an economically valuable food fish and a preferred candidate for brackishwater aquaculture in India. Genetic composition of complete mitochondrial DNA *E. suratensis* collected from Vembanad Lake of Kerala, India was characterised. The entire mitogenome was PCR amplified as contiguous, overlapping segments and sequenced using the Sanger sequencing method. The assembled mitogenome of *E. Suratensis* is 16456 bp circle, contained the 37 mitochondrial structural genes; two ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA (tRNA) genes, and 13 protein-coding genes, 1 non-coding control region/D-loop with the gene order identical to typical vertebrates. Low G content and high A+T (53.8%) content were observed along with Intergenic overlaps at ATP6 & ATP8, ND4 & ND4L and ND5 & ND6 genes. ATG is used as start codon by all coding genes except CO1 (GTG is the start codon), TAA was used as translation terminators for ND1, ND2, CO1, ATP8, ND4L and ND5 and the remaining genes used incomplete stop codon TA-/T--. An anti-G bias in the third codon positions and high pyrimidines presence in the second codon positions were observed along with proteins with amino acid encoded by A and C were the most frequently observed. The major non-coding region (D-loop) has several characteristic conserved sequence blocks (CSB) like CSB 1, CSB2, CSB3 and Promoter region. The phylogenetic analysis revealed several bootstraps supported monophyletic groups with *E. Suratensis* as Indo-Sri Lankan taxa. The family cichlidae and two continental groups from South America and Africa are monophyletic in origin. This mitogenomic data will provide baseline information for further studies on evolution, taxonomy, conservation, environmental adaptation and selective breeding of this declining species with aquaculture, ornamental and evolutionary importance.

## 1. INTRODUCTION

Green chromide(*Etroplus suratensis* Bloch, 1790) is a euryhaline cichlid species that distributed mainly in freshwater, brackish water and river mouths. It is the abundant species among the three indigenous cichlids (The others being *Etroplus maculatus* and *Etroplus canarensis*) native of peninsular India and Sri Lanka. *E. maculatus* occur in most backwaters of Kerala while *E. canarensis* is characterised by its restricted distribution in coastal wetlands of Karnataka (Jayaram 1999). *E suratensis* is with a greyish-green colouration and scales on the sides are with a pearly spot (Costa 2007; Chandrasekar 2014). Even though macrophytes are the predominant food it also consumes diatoms, molluscs, insects and animal matter (de Silva *et al.* 1984). Its wild populations have been recorded from Kerala and Tamil Nadu states of India. There isalso distribution in Goa, Andhra Pradesh, Orissa and West Bengal, probably introduced populations (Jayaram 1999). *E. suratensis* is an economically important food fish, known locally as 'Karimeen' in Kerala (Padmakumar *et al.* 2012) and is designated as the "State fish of Kerala" with backwaters of Kerala being the major source of the wild population and a potential source of its seed (Padmakumar *et al.* 2012). The family cichlidae comprises more than 700 species, inhabit in freshwater and brackish water in landmasses originated from Gondwanaland (Africa, South and Central America, India, SriLanka and Madagascar). The lakes of Africa harbour the largest diversity of cichlid species, where its explosive radiation happened within the past 10 million years (Azuma *et al*. 2008). Because of its rich diversity in terms of ecology, morphology, behaviour cichlids have been used as the best model organism for evolutionary biology, evolutionary genetics and phenotype-genotype relationship studies (Azuma *et al*. 2008). Biology and reproductive aspects of this species were studied as a good aquaculture species, suitability for culture in made habitats like ponds and tanks as it can tolerate a wide range of environmental conditions (Bindu *et al.* 2012; Chandrasekar 2014). Now it is a preferred candidate for brackishwater aquaculture in India and it has been widely introduced to dam reservoirs, lakes and culture ponds in India (Padmakumar *et al.* 2012; Chandrasekar *et al.* 2014) for culture. The backwaters of Kerala are the potential source of *E. suratensis* seed (Chandrasekar *et al.* 2014).

Animal mitochondrial DNA (mtDNA) is a circular molecule, typically 16–20 kb in length, with 37 mitochondrial structural genes encoding two ribosomal RNA (rRNA), 22 transfer

RNAs (tRNA) and 13 proteins along with a non-coding control region that regulates replication and transcription (Boore 1999). Plenty of literature has been published on mtDNA evolution and mtDNA have been using as a very useful marker for understanding evolutionary relationships, gene flow, hybridisation, introgression and historical demography mainly because of its maternal inheritance, fast evolutionary rate compared to nuclear DNA, lack of recombination and presence of multiple copies in the cell (Ballard and Whitlock 2004; Karl *et al*. 2012). mtDNA has been used as a marker to infer genetic population structure of many fishery resources (Curole and Kocher 1999; Cadrin *et al*. 2013; Miya and Nishida 2015). But often, these inferences were based on a short segment of the mtDNA like D-loop, cytochrome b region or ND2 genes and conclusions on the genetic stock structure and phylogenetic relationship using these genes with different evolutionary rates may not reflect the true picture. Complete mitogenomes resources provide a holistic perspective for comparisons making inferences regarding population structure and phylogenetic status accurately and effectively (Curole and Kocher 1999; Miya and Nishida 2015). The advent of new techniques like long PCR and next-generation sequencing techniques has made the characterisation of complete mitogenomes quicker and easier (Miya and Nishida 1999; Sorenson *et al*. 1999; Morin *et al*. 2010; Jacobsen *et al*. 2012; Miya and Nishida 2015). Now more than two thousand fish mitogenomes are available in public databases (http://www.ncbi.nlm.nih.gov/) and recent findings based on mitogenomic data have revolutionized several concepts of molecular phylogeny and evolution across multiple taxonomic levels (Miya and Nishida 2015; Curole and Kocher 1999). Whole mitogenome information has also been recently used to study selection and adaptation in fishes and other organisms in response to environmental and climatic fluctuations (da Fonseca *et al*. 2008; Silva *et al*. 2014; Stager *et al*. 2014; Caballero *et al*. 2015).

Wild populations of *E. suratensis* facing habitat deterioration by the disposal of solid and liquid wastes from increasing urbanisation, an increasing number of tourism activities in backwaters/estuaries, the threat from exotic species like *Oreochromis mossambicus, Trichogaster trichopterus* (Padmakumar *et al*. 2002; Krishnakumar *et al.* 2009) etc. Even though the population is declining and being in high demand, the importance of conservation of its wild populations has not been getting sufficient attention (Padmakumar *et al*. 2002). There have been attempts to create no-fishing zones/aquatic sanctuary within some of the larger estuaries and Captive breeding for the conservation of the wild populations (Bindu and Padmakumar 2014). Genetic information has been widely used for the conservation and management of endangered species. Information on the biological and genetic features of *E. suratensis* is essential for formulating valid programs for its conservation. Few studies have been used mtDNA as molecular markers in the establishment of phylogenetic relationships and population structure among *E. suratensis* population (Suneetha 2007; Dhanya *et al*. 2013; Alex *et al*. 2016). Recent investigations have been focussed on selection and adaptation in the mitochondrial oxidative phosphorylation machinery which provided clues to thermal and metabolic adaptations in many fishes (Bradbury *et al*. 2010; Foote *et al*. 2011; Garvin *et al*. 2012; Teacher *et al*. 2012; Caballero *et al*. 2015). *E suratensis* is important from this viewpoint because they experienced wide climatic fluctuations as they are widely distributed across environmental clines and are prone to forces of positive and purifying selection. Hence, we studied the complete mitochondrial genome structure and organization of *E. suratensis.* So characterizing the complete mitogenome of this commercially and ecologically important species will act as baseline information for further studies on taxonomy, conservation, evolutionary genetics, adaptive variation to the environment, selective breeding etc. This is the first study which accommodates available cichlid mitogenomes (mtDNA) and a mitogenome sequenced from wild *E. suratensis*, which subsequently used for phylogenetic inference, comparative analysis on sequence pattern, and codon usage to get a better view on dynamics on cichlid's mtDNA evolution.

## 2. Materials And Methods

### 2.1 Sample collection and preparation

*E. suratensis* was collected from Vembanad Lake Kerala. Skeletal muscle samples were obtained from the tail of each individual and stored in 95% ethanol for DNA extraction. Genomic DNA was isolated by standard phenol/chloroform method after proteinase K digestion (Sambrook and Russell 2001).

## 2.2 PCR Amplification and sequencing Mitochondrial DNA

The entire mitogenome was amplified by using a long PCR technique with Q5® High-Fidelity DNA Polymerase (NEW ENGLAND BioLabs). Primers pairs 7.T1) were designed based on known regions of the *E. suratensis* mtDNA and complete mitogenome were amplified as 5 contiguous, overlapping segments. PCR amplifications were carried out in 50 µl reaction mixture. Purification of the PCR product was carried out using Qiagen PCR purification kit (Qiagen) and sequenced with both primers using the BigDye Terminator Sequencing Ready Reaction v30 kit (Applied Biosystems) following instructions of the manufacturer. Sequencing was carried out on an ABI 3730 automated sequencer (Life Technologies). The internal region of large fragments was obtained by sequencing of the PCR products with an internal primer designed from the corresponding sequence obtained in the first sequencing process. Short-PCR reactions were carried out in 25µl reaction mixture containing 2.5 µl 10x buffer (10 mM Tris–HCl, 50 mM KCl, 15 mM $MgCl_2$), 200 µM of each dNTP, 0.25 µM of each primer, 1 unit of Taq DNA polymerase (Sigma Aldrich), and 50 ng of template DNA. All the PCR reaction was carried out in a Biorad T100 thermocycler (Biorad, USA).

## 2.3 Assembly and annotation of the mitochondrial genome

The sequence fragments obtained were assembled into the complete mitochondrial genome using MEGA 6 (Tamura *et al*. 2013) and Geneious R7 (Kearse *et al*. 2012). Annotation of protein-coding genes, rRNAs and tRNAs was performed using NCBI-BLAST and MitoAnnotator (Iwasaki *et al*. 2013) programs. Nucleotide composition of mitogenome and protein-coding genes were determined using Geneious R7. Codon usage and RSCU values were calculated with MEGA 6. Alignments with previously published closely related bony fishes were carried out to identify the origin of replication and conserved blocks in the non-coding control region. The mtDNA sequences were deposited in NCBI GenBank.

### 2.4 Phylogeny construction

The phylogenetic tree was reconstructed using mitogenome sequences of cichlids and related persiforms, retrieved from NCBI GenBank, to visualise the relationship of *E. suratensis* with other cichlids as well as to validate its taxonomic position. The sequences included in the phylogenetic analysis belonged to the family to Chondricthyes, Sarcopterygii, Polypteriformes, Acipenseriformes, Amiiformes, Lepisosteiformes, Osteoglossomorpha, Elopomorpha and Otocephala was used as outgroups. The 12 protein-coding genes were aligned and concatenated using Geneious R7 (GTR+G+I model was selected as the best model for phylogeny construction) and a maximum likelihood phylogeny was constructed based on 1000 replicates.

## 3. Results And Discussion

### 3.1 Mitogenome organisation

The assembled mitogenome is a 16456 bp circle (Fig 7.1). It contained the 37 mitochondrial structural genes; two ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA (tRNA) genes, and 13 protein-coding genes, 1 non-coding control region/D-loop (Table 7.1) and the gene order was identical to that in other vertebrates (Boore 1999). As previously reported for most vertebrates *E. suratensis* also followed the Heavy (H) and Light (L) strand coding pattern (Boore 1999). Except the ND6 and eight tRNA genes ($tRNA^{Gln(TTG)}$, $tRNA^{Ala(TGC)}$, $tRNAA^{sn(GTT)}$, $tRNA^{Cys(GCA)}$, $tRNA^{Tyr(GTA)}$, $tRNA^{Ser(TGA)}$, $tRNA^{Glu(TTC)}$, and $tRNA^{Pro(TGG)}$), all other genes were encoded on the H-strand. The overall base composition of the H-strand was as follows: A (28.2%), T (25.6%), C (30.9%), G (15.3%) and G+C (46.2%) (Table 7.2). Similar to other vertebrate low G content and high A+T (53.8%) content was observed in the genome (Broughton *et al*. 2001; Fischer *et al*. 2013). The complete mitogenome of *E. suratensis* sequence was deposited in NCBI Gen Bank under accession number KU665487 respectively.

The complete mitogenome of *E. suratensis* collected from Chilka Lake, Odisha, India has already been characterised by Mohanta *et al.* 2016. But detailed analysis on structure, organization, amino acid content and codon usage have not been reported. In the present investigation, we have conducted an extensive investigation on mitogenome content, structure and phylogenetic position of *E. suratensis*. The phylogenetic analysis included all the available complete mitogenomes of Cichlids to make observations on their divergence.

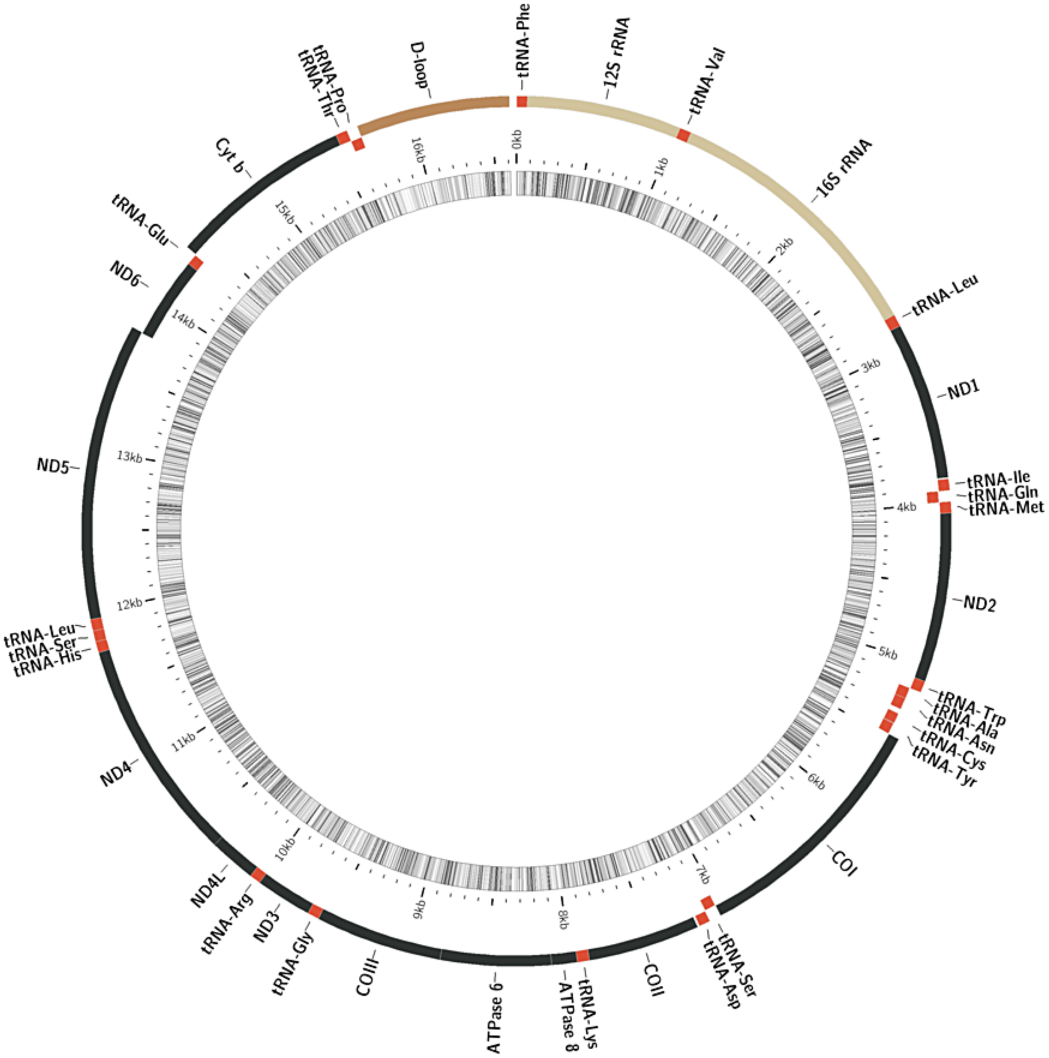


**Fig. 7.1** Mitogenome map of *E. suratensis* (16456 bp) (Gen Bank accession no KU665487) generated with MitoAnnotator. Protein-coding genes, tRNAs, rRNAs, and D-loop regions are shown in different colours. Genes located within the outer circle are coded on the H-strand whereas the remaining genes are coded on the L-strand.

**Table 7.1** Location and arrangement of genes on the mitogenomes of *E. suratensis.*

| Gene | *E. suratensis* | | | | | |
|---|---|---|---|---|---|---|
| | Position | | Size (bp) | Strand[a] | Codon[b] | |
| | From(bp) | To(bp) | | | Start | Stop |
| tRNA-Phe | 1 | 69 | | H | | |
| 12S rRNA | 70 | 1017 | | H | | |
| tRNA-Val | 1018 | 1089 | | H | | |
| 16S rRNA | 1090 | 2780 | | H | | |
| tRNA-Leu | 2781 | 2853 | | H | | |
| ND1 | 2854 | 3828 | | H | ATG | TAA |
| tRNA-Ile | 3832 | 3901 | | H | | |
| tRNA-Gln | 3901 | 3971 | | L | | |
| tRNA-Met | 3971 | 4039 | | H | | |
| ND2 | 4040 | 5086 | | H | ATG | TAA |
| tRNA-Trp | 5087 | 5157 | | H | | |
| tRNA-Ala | 5159 | 5227 | | L | | |
| tRNA-Asn | 5229 | 5301 | | L | | |
| tRNA-Cys | 5339 | 5405 | | L | | |
| tRNA-Tyr | 5406 | 5475 | | L | | |
| CO1 | 5477 | 7033 | | H | GTG | TAA |
| tRNA-Ser | 7050 | 7120 | | L | | |
| tRNA-Asp | 7124 | 7195 | | H | | |
| CO2 | 7201 | 7891 | | H | ATG | T-- |
| tRNA-Lys | 7892 | 7966 | | H | | |
| ATPase 8 | 7968 | 8135 | | H | ATG | TAA |
| ATPase 6 | 8126 | 8808 | | H | ATG | TA- |
| CO3 | 8809 | 9593 | | H | ATG | TA- |
| tRNA-Gly | 9594 | 9663 | | H | | |
| ND3 | 9664 | 10013 | | H | ATG | TA- |
| tRNA-Arg | 10014 | 10081 | | H | | |
| ND4L | 10082 | 10378 | | H | ATG | TAA |
| ND4 | 10372 | 11752 | | H | ATG | T-- |
| tRNA-His | 11753 | 11821 | | H | | |
| tRNA-Ser | 11822 | 11888 | | H | | |
| tRNA-Leu | 11892 | 11964 | | H | | |
| ND5 | 11965 | 13803 | | H | ATG | TAA |
| ND6 | 13801 | 14321 | | L | ATG | TA- |
| tRNA-Glu | 14322 | 14390 | | L | | |
| Cyt b | 14395 | 15488 | | H | ATG | TA- |
| tRNA-Thr | 15535 | 15606 | | H | | |
| tRNA-Pro | 15608 | 15677 | | L | | |
| control region (D-loop) | 15678 | 16456 | | | | |

[a] H and L, respectively, denote heavy and light strands.

[b]Codons containing “- “symbols indicate an incomplete stop codon.

**Table 7.2** Nucleotide composition of the mitogenome of *E. suratensis*

| *E. suratensis* | | | |
|---|---|---|---|
| % Nucleotide composition (GC 46.2) | | | |
| A | C | G | T |
| Complete mitogenome (H- Strand) | | | |
| 28.2 | 30.9 | 15.3 | 25.6 |
| All protein coding gene concatenated (H- Strand)[a] | | | |
| 26.0 | 32.9 | 13.7 | 27.4 |
| ND 6 (L- Strand)[b] | | | |
| 39.7 | 38.3 | 9.6 | 12.3 |
| 1st codon position[c] | | | |
| 26.4 | 28.1 | 24.7 | 20.8 |
| 2nd codon position[c] | | | |
| 17.9 | 28.1 | 13.5 | 40.5 |
| 3rd codon position[c] | | | |
| 31.9 | 39.3 | 6.3 | 22.5 |

[a]Based on the 12 protein-coding genes located on the H-strand.

[b] Base on the ND 6 gene located on the L-strand.

[c] Based on the 13 protein-coding genes.

## 3.2 Protein-coding genes

In *E. suratensis* 13 protein-coding genes were 11358bp in total length and thus represented ~ 69% of the genome. All the genes are encoded by heavy strand except ND6 gene which is encoded by light strand. ATP6 & ATP8 shared 10 nucleotides, ND4 & ND4L shared 7 and ND5 & ND6 shared 4 nucleotides in their overlapping region. ATG is used as start codon by all coding genes except CO1 (GTG is the start codon). Intergenic overlaps of protein-coding regions are common within vertebrate mitogenomes and have been reported for several fish species (Boore 1999; Morin *et al*. 2010; Mu *et al*. 2015). In *E. suratensis*, stop codon TAA was used as translation terminators for ND1, ND2, CO1, ATP8, ND4L and ND5. The remaining genes used incomplete stop codon TA-/T-- (Table 7.1). Reading frame overlap and incomplete stop codons are common in mitochondria and post-transcriptional polyadenylation compensate the two adenosine nucleotide required for generating the TAA stop codon (Ojala *et al*. 1981). The H-strand coding sequences of *E. suratensis* consisted of 26.0% A, 27.4% T, 13.7% G and 32.9% C bases. The corresponding composition for L-strand is 39.7% A, 12.3% T, 9.6% G and 38.3% C bases. The L-strand (AC 78%) was observed to be relatively AC rich in comparison to the major coding strand (AC 58.9% H-strand) (Table 7.2). Variations in the composition of H and L-strand have been reported for vertebrate mitochondrial DNA (Perna and Kocher 1995; Min and Hickey 2007; Fischer *et al*. 2013). Similar to other vertebrates, an anti-G bias in the third codon positions and high pyrimidines presence in the second codon positions were observed in the genome (Table 7.2) (Naylor *et al*. 1995; Boore 1999). The most frequently used amino acids were Leucine (17.6 %), followed by Alanine (8.6 %) and Isoleucine (7.2 %) (Table 7.3). The RSCU values identified were showing codon preference for each amino acid in protein-coding genes. The highest estimated highest RSCU were matched to corresponding tRNAs identified in the mitogenome (Table 7.3), except for Alanine, Glycine, Leucine, Methionine, Proline, Serine, Threonine and Valine. When considering degenerate third codon positions, inconsistent with the anti-G bias identified in the mitogenome of *E. Suratensis,* codons complementary to the tRNAs ending in A and C were the most frequently observed and G nucleotide was the least frequent (Table 7.3).

**Table 7.3** Amino acid and codon usage in mitogenome of *E. suratensis*

| Amino acid | *E. suratensis* | | |
|---|---|---|---|
| | %[a] | Codons | RCSUC[b] |
| Alanine(Ala/A) | 8.6 | GCU | 55 |
| | | GCC | **146** |
| | | GCA* | 110 |
| | | GCG | 14 |
| Arginine(Arg/R) | 2.0 | CGU | 11 |
| | | CGC | 14 |
| | | CGA* | **45** |
| | | CGG | 4 |
| Asparagine(Asn/N) | 3.0 | AAU | 29 |
| | | AAC* | **88** |
| AsparticAcid(Asp/D) | 1.8 | GAU | 18 |
| | | GAC* | **53** |
| Cysteine(Cys/C) | 0.6 | UGU | 6 |
| | | UGC* | **16** |
| GlutamicAcid(Glu/E) | 2.6 | GAA* | **84** |
| | | GAG | 14 |
| Glutamine(Gln/Q) | 2.5 | CAA* | **90** |
| | | CAG | 7 |
| Glycine(Gly/G) | 6.6 | GGU | 39 |
| | | GGC | **94** |
| | | GGA* | 72 |
| | | GGG | 38 |
| Histidine(His/H) | 2.8 | CAU | 37 |
| | | CAC* | **74** |
| Isoleucine(Ile/I) | 7.2 | AUU | 137 |
| | | AUC* | **148** |
| Leucine(Leu/L) | 17.6 | UUA* | 74 |
| | | UUG | 32 |
| | | CUU | 168 |
| | | CUC | **179** |
| | | CUA* | **179** |
| | | CUG | 32 |
| Lysine(Lys/K) | 1.9 | AAA* | **71** |
| | | AAG | 4 |
| Methionine(Met/M) | 3.9 | AUA | **104** |
| | | AUG* | 42 |
| Phenylalanine(Phe/F) | 6.3 | UUU | 101 |
| | | UUC* | **137** |
| Proline(Pro/P) | 5.8 | CCU* | 51 |
| | | CCC | **118** |
| | | CCA | 49 |
| | | CCG | 5 |
| Serine(Ser/S) | 6.6 | UCU | 46 |
| | | UCC | **97** |
| | | UCA* | 47 |
| | | UCG | 5 |
| | | AGU | 14 |
| | | AGC* | 42 |
| Threonine(Thr/T) | 4.1 | ACU | 4 |
| | | ACC | **153** |
| | | ACA* | 116 |
| | | ACG | 8 |
| Tryptophan(Trp/W) | 3.1 | UGA* | **107** |
| | | UGG | 11 |
| Tyrosine(Tyr/Y) | 3.0 | UAU | 34 |
| | | UAC* | **75** |
| Valine(Val/V | 5.8 | GUU | **63** |
| | | GUC | 54 |
| | | GUA* | 61 |
| | | GUG | 21 |

a % of Amino acid based on the 13 protein-coding genes.

b RSCU relative synonymous codon usage.

* Codons that is complementary to the tRNA genes.

### 3.3 RNA genes

A small (12S rRNA) and large (16S rRNA) ribosomal RNA subunit was identified with 948bp and 1691bp in size respectively. Similar to other vertebrate *E. Suratensis* rRNA genes have high adenine content (52.2%) (Naylor *et al*. 1995; Boore 1999). As in coding genes, 3 of the 22 tRNA genes identified showed overlaps, tRNA Gln shared one nucleotide at both ends, upstream with tRNA Ile and downstream with tRNA Met.

### 3.4 Non-coding region

As in most vertebrate's mtDNA, the origin of light strand replication ($O_L$) in *E. Suratensis* was located between tRNA Asn and tRNA Cys (WANCY region) and it is from 5303 bp to 5338 bp. This region can fold into a stable stem-loop secondary structure in its single-stranded form and which is a need for the initiation of replication (Hixson *et al.* 1986). WANCY region is a region coding for five mitochondrial tRNAs (tryptophan, alanine, asparagine, cysteine, and tyrosine). A major non-coding region between the tRNA Pro and tRNA Phe genes were identified (779 bp in size). It is considered as the control region (D-loop) and has several characteristic conserved sequence blocks (CSB) like CSB 1, CSB2, CSB3 and Promoter region (Fig 7.2).

```
                                                                                 15680     15690     15700
                                                                               ..|....|....|....|....|
                                                                   tRNA-Phe-   TCCGAGCTCTGCCAGAAATAGAA

       15710     15720     15730     15740     15750     15760     15770     15780     15790     15800
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AAATATGACATATATGTATTTACACCATTAATTTATTGTAAACATATTAATGAAGATATAGTACATTAAATTAAGACAACCCCCGAAATAAACCTCAACA

       15810     15820     15830     15840     15850     15860     15870     15880     15890     15900
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
TTCTTGTTTAGTCCATGCTAACTGTCGTATACATATCCCATACATTTAACTAGTACAGAATACTGATTGGGTAATGAACGAAACTTAAGATCTCAACAGT

       15910     15920     15930     15940     15950     15960     15970     15980     15990     16000
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CTAAATTCACTAGTCAAGATATACCAAGTAATCAACTATCCTGTAATCAAGGAAAATTTAATGTAGTAAGAGACCACCATCAGTTGATTACTTAATGTTA

       16010     16020     16030     16040     16050     16060     16070     16080     16090     16100
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ATCATGCATGATGGTCAGGTACAATAATTGCAAACTTGCCCACGGTGAATTATTCCTGGCATAAGTTAATGGTGTTAATACATACTCCTCGTTACCCACC

       16110     16120     16130     16140     16150     16160     16170     16180     16190     16200
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ATGCCGGGCGTTATCTCCAGAGGGTCACTGGTTCTCTTTTTTGTCCTCCTTTCATTTGGCATTTCACAGTGTACACAGGTCCTAGCTGACAAGGGTGAGC

       16210     16220     16230     16240     16250     16260     16270     16280     16290     16300
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ATTTTTCTTGCATGACAGTAAATAATGTGAAGTGATTCAAAGTCATTACTAGATGATGATATCAAGAGCATAATACTGCTTAAGATTTTCCTAATTTCCC
   CSB 1
       16310     16320     16330     16340     16350     16360     16370     16380     16390     16400
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
GTCAAAGTGCCCTAGTTTGTGCGCGTAAACCCCCCCTACCCCCCCAAACTCCTAAGATCGCTGTCATTCCTGTAAACCCCCAAAACAGGGCTAAATCTCA
 CSB 2                                  CSB 3
       16410     16420     16430     16440     16450
....|....|....|....|....|....|....|....|....|....|....|.
AAAGTTCATTTCTGTATTAAAAGTGTGTTTATTTACATTATTACAATAATGCACAC-tRNA-Pro
                                   Promotor
```

**Fig. 7.2** Characteristics conserved blocks (CSB 1, CSB2, CSB3) and Promoter region in the non-coding region (D-Loop) of *E. suratensis* mitochondrial DNA.

## 3.5 Phylogenetic analysis

The phylogenetic relationships among 51 cichlid fishes reconstructed using the Maximum likelihood method and Idopacific sergeant (*Abudefduf vaigiensis*) and clownfish (*Amphiprion ocellaris*) as outgroups showed several bootstraps supported monophyletic groups with *E. suratensis* as Indo-Sri Lankan taxa (Fig. 7.3). The phylogenetic relationships obtained for non-cichlids with cichlids were consistent with previous mitogenome studies (Miya *et al.* 2003; Inoue *et al.* 2003; Mabuchi *et al.* 2007) (Fig. 7.S1). Among the 51 cichlids, 29 belonged to Africa, 17 to South America, three to Madagascar and two to Indo-Sri Lanka. Among the lineages, Madagascar and Indo-Sri Lanka taxa are not monophyletic and one Madagascar species (*Paretroplus maculatus*) with Indo-Sri Lanka lineage (with *E. suratensis and E. maculatus*) form a sister group to all other taxa. The other Madagascar taxa (*Paratilapia polleni*, *Ptychochromoides katria*) formed a sister group to the South American lineage. These results are consistent with previous observations using a limited number of mitochondrial DNA, nuclear DNA markers (Zardoya *et al.* 1996; Streelman and Karl 1997; Farias *et al.* 2000; Sparks and Smith 2004) and morphology (Stiassny 1991). The tree topology supported the vicariant divergence scenario (Stiassny 1991; Zardoya *et al.* 1996; Sparks and Smith 2004) than the alternative hypothesis (Vences *et al.* 2001; Briggs 2003). The alternative hypothesis assumes that the origin of cichlids happened in the Cenozoic Africa and dispersed to South America, Madagascar and India by saltwater migration (Vences *et al.* 2001; Briggs 2003). The alternative hypothesis was rejected by the absence of monophyletic Madagascar and Indo-Sri Lankan; African, Madagascar and Indo-Sri Lankan; African and Indo-Sri Lankan cichlids.

The phylogenetic tree supported the proposed Gondwanan origin of Cichlidae and association of divergence pattern of continental cichlid groups with the geological history of continental drift (Gondwanan origin and vicariant divergence of cichlids) (Azuma *et al.* 2008). Similar results were also reported by Sparks and Smith 2004 with mitochondrial and nuclear gene fragments. When estimated divergence times among cichlids (Genner *et al.* 2007; Azuma *et al.* 2008) and the times of continental fragmentation based on geological evidence (Smith *et al.* 1994; Storey 1995; Masters *et al.* 2006) were compared, the divergence time between Madagascar and Indo-Sri Lankan taxa (~87 MYA: 69–106 MYA) is close to the time of separation between Madagascar and India (85–95 MYA) from Gondwanaland. The divergence time estimated between African and South American taxa (~89 MYA: 72–108 MYA) is also matched to the time of separation between African and South American landmasses (~100 MYA). The divergence between African+South American cichlids and Madagascar cichlids (~96 MYA: 78–115 MYA) had happened before the complete separation of the Indo-Madagascar landmass from Gondwanaland (120–130 MYA). The topology of cichlids phylogenetic tree with one Madagascar species (*Paretroplus maculatus*) in Indo-Sri Lanka lineage (*E. suratensis, E. maculatus*) may be an indication of this early divergence.

The complete mitochondrial genome sequence of this economically valuable food fish has gene structure, content and organisation similar to most vertebrates. This mitogenomic data will provide baseline information for further studies on evolution, taxonomy, conservation, environmental adaptation and selective breeding of this declining species with aquaculture, ornamental and evolutionary importance.

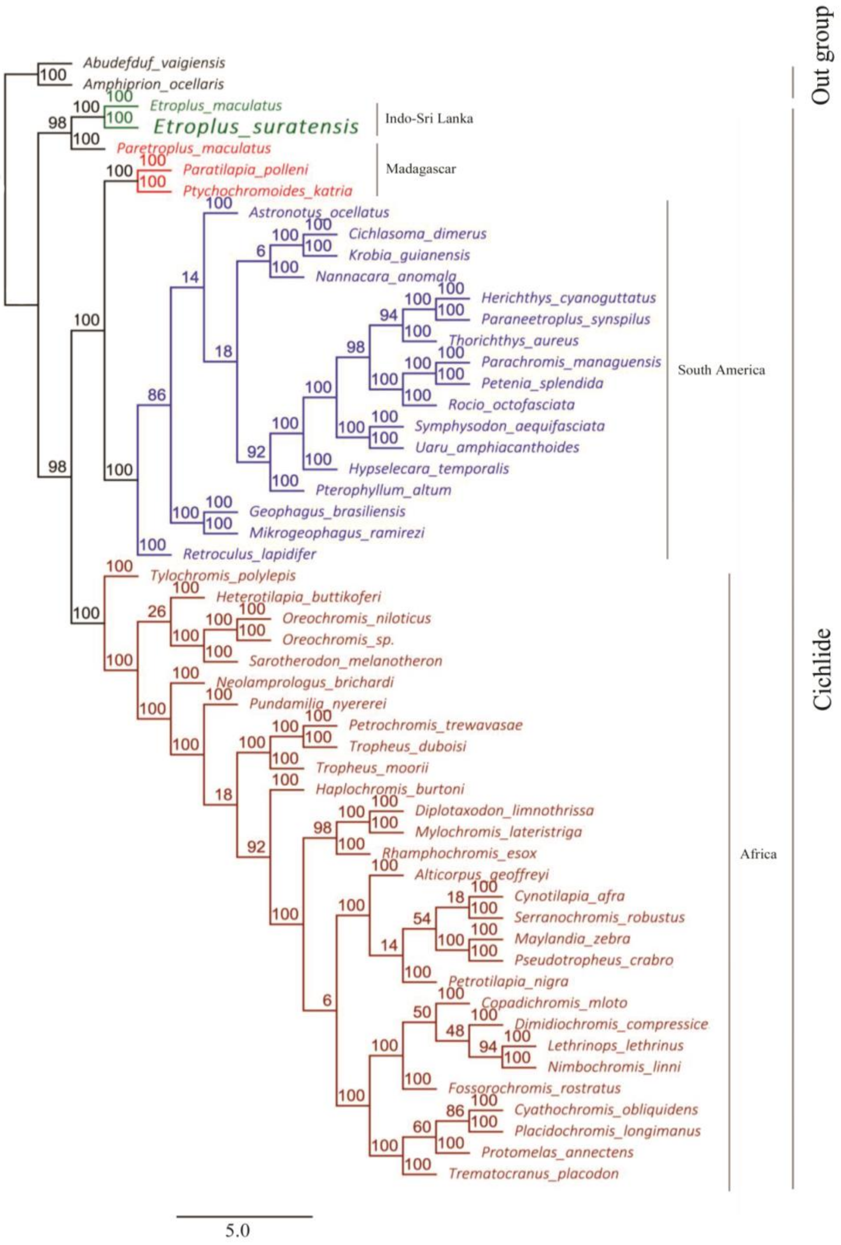


**Fig. 7.3** Maximum likelihood phylogenetic tree generated by alignment of complete mitogenome nucleotide sequences of *E. Suratensis* and other Cichlids. Cichlid species which are represented from Africa, South America, Madagascar and Indo-Sri Lankan regions were used. *Amphiprion ocellaris* and *Abudefduf vaigiensis* were used as outgroup.

## Supplementary Tables and Figures

**Table 7.S1** List of Primer pairs used for amplification of *E. suratensis* mitochondrial DNA.

| Primer Name | | Sequence (5' - 3') |
|---|---|---|
| cichmit 1 | Forward primer | CCTGGCATAAGTTAATGGTG |
| | Reverse primer | AGACAGTTAAGCCCTCGTTA |
| cichmit 2 | Forward primer | ACGGACCGAGTTACCCTAGG |
| | Reverse primer | CCTGCYTCTACWCCAGAGGA |
| cichmit 3 | Forward primer | TTGGTGCCCCYGATATRGCC |
| | Reverse primer | AGGGTGCCGGYGYTRTTTTG |
| cichmit 4 | Forward primer | TRGCCTTYAGYGCAACCGAA |
| | Reverse primer | GGGTTTRAATTGTTTGTTGGTKA |
| cichmit 5 | Forward primer | CCCCGTAATATCYATACCCC |
| | Reverse primer | CTATTGTRGCGGCTGCAATR |
| cichmit 6 | Forward primer | YATTGCAGCCGCYACAATAG |
| | Reverse primer | AGAACCAGTGACCCTCTGGA |

**Table 7.S2** Details of sequences used for phylogenetic analysis**.**

| **Species Name** | **NCBI Accession NO** |
|---|---|
| *Abudefduf_vaigiensis* | (AP006016.) |
| *Acipenser_transmontanus_* | (AB042837.) |
| *Alticorpus_geoffreyi* | (KT277287.) |
| *Amia_calva_mitochondrial* | (AB042952.) |
| *Amphiprion_ocellaris* | (AP006017.) |
| *Anguilla_japonica* | (AB038556.2) |
| *Astronotus_ocellatus* | (AP009127.) |
| *Atractosteus_spatula* | (AP004355.) |
| *Beryx_splendens* | (AP002939.) |
| *Chlorophthalmus_agassizi* | (AP002918.) |
| *Cichlasoma_dimerus* | (KR150876.) |
| *Conger_myriaster* | (AB038381.2) |
| *Copadichromis_mloto* | (KX196155.) |
| *Coregonus_lavaretus* | (AB034824.) |
| *Crossostoma_lacustre* | (M91245.) |
| *Cyathochromis_obliquidens* | (MF033354.) |
| *Cynotilapia_afra* | (JN628861.) |
| *Cyprinus_carpio* | (X61010.) |
| *Danio_rerio* | (AC024175.) |
| *Dimidiochromis_compressiceps* | (JN628856.) |
| *Diplotaxodon_limnothrissa* | (JN628851.) |
| *Engraulis_japonicus* | (AB040676.) |
| *Erpetoichthys_calabaricus* | (AP004350.) |
| *Esox_lucius* | (AP004103.) |
| *Etroplus_maculatus* | (AP009505.) |
| *Etroplus_suratensis* | (NC_029832.) |
| *Fossorochromis_rostratus* | (KT290557.) |
| *Gadus_morhua* | (X99772.) |
| *Gasterosteus_aculeatus* | (AP002944.) |
| *Geophagus_brasiliensis* | (KU531434.) |
| *Gymnothorax_kidako* | (AP002976.) |
| *Halichoeres_melanurus* | (AP006018.) |
| *Haplochromis_burtoni* | (KP641358.) |
| *Helicolenus_hilgendorfi* | (AP002948.) |
| *Herichthys_cyanoguttatus* | (KR150867.) |
| *Heterotilapia_buttikoferi* | (KF866133.) |

| | |
|---|---|
| *Hiodon_alosoides* | (AP004356.2) |
| *Hypselecara_temporalis* | (AP009506.) |
| *Krobia_guianensis* | (KR233978.) |
| *Latimeria_menadoensis* | (AP006858.2) |
| *Lepisosteus_oculatus* | (AB042861.) |
| *Lethrinops_lethrinus* | (KX595334.) |
| *Maylandia_zebra* | (KT166981.) |
| *Mikrogeophagus_ramirezi* | (KR233976.) |
| *Mustelus_manazo* | (AB015962.) |
| *Mylochromis_lateristriga* | (KU056478.) |
| *Nannacara_anomala* | (KU531436.) |
| *Neoceratodus_forsteri* | (AJ584642.) |
| *Neolamprologus_brichardi* | (AP006014.) |
| *Nimbochromis_linni* | (JN628853.) |
| *Notacanthus_chemnitzi* | (AP002975.2) |
| *Oncorhynchus_mykiss* | (L29771.) |
| *Oreochromis_niloticus* | (GU238433.) |
| *Oreochromis_sp.* | (AP009126.) |
| *Oryzias_latipes* | (AP004421.) |
| *Osteoglossum_bicirrhosum* | (AB043025.) |
| *Pantodon_buchholzi* | (AB043068.) |
| *Parachromis_managuensis* | (KP728467.) |
| *Paralichthys_olivaceus* | (AB028664.) |
| *Paraneetroplus_synspilu* | (KF879808.) |
| *Paratilapia_polleni* | (AP009508.) |
| *Paretroplus_maculatus* | (AP009504.) |
| *Petenia_splendida* | (KJ914664.) |
| *Petrochromis_trewavasae* | (HE961974.) |
| *Petrotilapia_nigra* | (JN628852.) |
| *Placidochromis_longimanus* | (KT309044.) |
| *Polymixia_japonica* | (AB034826.) |
| *Polyodon_spathula* | (AP004353.) |
| *Polypterus_ornatipinni* | (AP004351.) |
| *Polypterus_senegalus* | (AP004352.) |
| *Protomelas_annectens* | (KT188786.) |
| *Pseudolabrus_sieboldi* | (AP006019.) |
| *Pseudotropheus_crabro* | (JN628854.) |
| *Pterophyllum_altum* | (KT180164.) |
| *Ptychochromoides_katria* | (AP009507.) |
| *Pundamilia_nyererei* | (KT222896.) |
| *Retroculus_lapidifer* | (KR150871.) |
| *Rhamphochromis_esox* | (JN628860.) |
| *Rocio_octofasciata* | (KR150870.) |
| *Salmo_salar* | (U12143.) |
| *Sardinops_melanostictus* | (AB032554.) |
| *Sargocentron_rubrum* | (AP004432.) |
| *Sarotherodon_melanotheron* | (JF894132.) |
| *Scaphirhynchus_cf._albus* | (AP004354.) |
| *Scyliorhinus_canicula* | (Y16067.) |
| *Serranochromis_robustus* | (KX595333.) |
| *Symphysodon_aequifasciata* | (KT362183.) |
| *Takifugu_rubripes* | (AJ421455.) |
| *Tetraodon_nigroviridis* | (AP006046.) |
| *Thorichthys_aureus* | (KU531435.) |
| *Trematocranus_placodon* | (JN628850.) |
| *Tropheus_duboisi* | (AP006015.) |
| *Tropheus_moorii* | (HE961975.) |
| *Tylochromis_polylepis* | (AP009509.) |
| *Uaru_amphiacanthoides* | (KR150875.) |

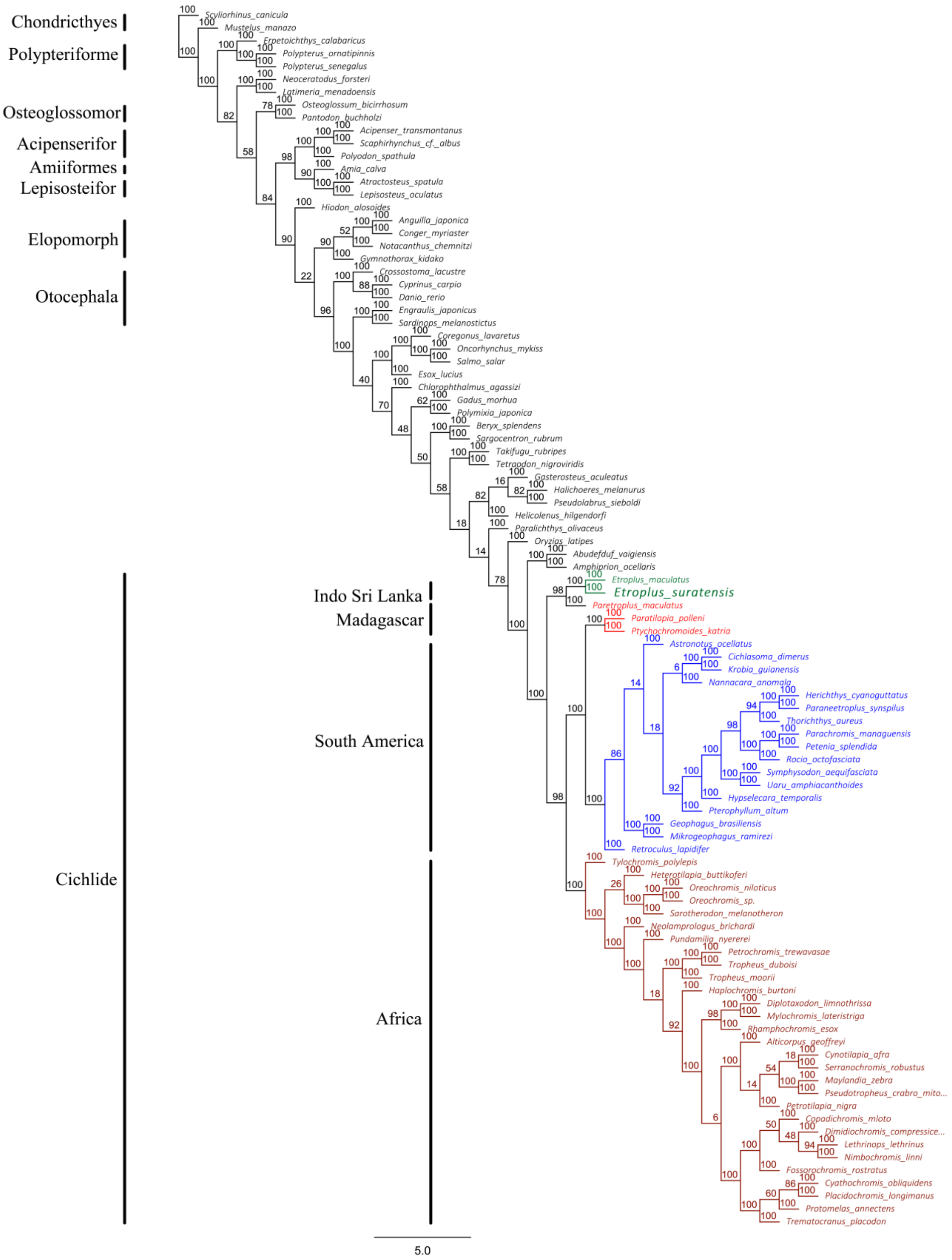


**Fig. 7.S1** Maximum likelihood phylogenetic tree generated by alignment of complete mitogenome nucleotide sequences of *E. Suratensis and* other Cichlids. Fishes belong to Chondrichthyes, Sarcopterygii, Polypteriformes, Acipenseriformes, Amiiformes, Lepisosteiformes, Osteoglossomorpha, Elopomorpha and Otocephala was used as outgroups.

## 4. References

1. Alex MD, Kumar AB, Kumar US, George S (2016) Analysis of genetic variation in Green Chromide [*Etroplus suratensis* (Bloch)] (Pisces: Cichlidae) using microsatellites and mitochondrial DNA. *IndianJ Biotechnol* 15(1):375-381
2. Azuma Y, Kumazawa Y, Miya M, Mabuchi K, Nishida M (2008) Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC Evol Biol* 8(1):215
3. Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Mol Ecol*13(4):729-744
4. Bindu L, Padmakumar KG (2012) Breeding behaviour and embryonic development in the Orange chromide, *Etroplus maculatus* (Cichlidae, Bloch 1795). *J Mar Biol Ass India* 54(1):13-19
5. Bindu L, Padmakumar KG (2014) Reproductive biology of *Etroplus suratensis* (Bloch) from the Vembanad wetland system, Kerala. *Indian J Geo-Mar Sci* 43(4)
6. Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res*27(8):1767-1780
7. Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, Snelgrove PV, Morris CJ, Gregory RS, Hardie DC, Hutchings JA (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature.*Proc R Soc Lond B Biol Sci*277(1701):3725-3734
8. Briggs JC (2003) Fishes and birds: Gondwana life rafts reconsidered. *Syst Biol* 52:548-553
9. Broughton RE, Milam JE, Roe BA (2001) The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Res*11(11):1958-1967
10. Caballero S, Duchene S, Garavito MF, Slikas B, Baker CS (2015) Initial evidence for adaptive selection on the NADH subunit two of freshwater dolphins by analyses of mitochondrial genomes. *PloS one*10(5):e0123543
11. Cadrin SX, Kerr LA, Mariani S (2013) Stock identification methods: applications in fishery science. Academic Press
12. Chandrasekar S, Nich T, Tripathi G, Sahu NP, Pal AK, Dasgupta S (2014) Acclimation of brackish water pearl spot (*Etroplus suratensis*) to various salinities: relative changes in abundance of branchial Na+/K+-ATPase and Na+/K+/2Cl− co-transporter in relation to osmoregulatory parameters. *Fish Physiol Biochem* 40(3):983-96
13. CMFRI Kochi (2017) CMFRI Annual Report 2016-2017. Technical Report CMFRI, Kochi
14. Costa HH (2007) Biological studies of the pearl spot *Etroplus suratensis* (pisces, cichlidae) from three different habitats in Sri Lanka. *Intern Rev Hydrobiol* 68(4):565–580
15. Curole JP, Kocher TD (1999) Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol Evol*14(10):394-398
16. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A (2008) The adaptive evolution of the mammalian mitochondrial genome. *BMC genomics* 9(1):119
17. de Silva SS, Maitipe P, Cumaranatunge RT (1984) Aspects of biology of euryhaline Asian cichlid, *Etroplus suratensis*. *Environ Biol Fishes* 10(1/2):77–87
18. Dhanya AM, Remya M, Biju KA (2013) Morphometric and genetic variations of *Etroplus suratensis* (Bloch) (Actinopterygii: Perciformes: Cichlidae) from two tropical lacustrine ecosystems, Kerala, India. *J Aquat Biol Fisheries* 1(1-2):140-150
19. Farias IP, Orti G, Meyer A (2000) Total evidence: molecules, morphology, and the phylogenetics of cichlid fishes. *J Exp Zool* 288:76-92
20. Fischer C, Koblmuller S, Gully C, Schlotterer C, Sturmbauer C, Thallinger GG (2013) Complete mitochondrial DNA sequences of the threadfin cichlid (*Petrochromis trewavasae*) and the blunthead cichlid (*Tropheus moorii*) and patterns of mitochondrial genome evolution in cichlid fishes. *Plos One* 8(6):e67048
21. Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, Gilbert MTP, da Fonseca RR (2011) Positive selection on the killer whale mitogenome. *Biol Lett*7(1):116-118
22. Garvin MR, Bielawski JP, Gharrett AJ (2012) Correction: Positive Darwinian Selection in the Piston That Powers Proton Pumps in Complex I of the Mitochondria of Pacific Salmon. *PloS one*7(8):e24127
23. Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF (2007) Age of cichlids: new dates for ancient lake fish radiations. *Mol Biol Evol* 24:1269-1282.

24. Hixson JE, Wong TW, Clayton DA (1986) Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. *J Biol Chem* 261(5):2384-2390
25. Inoue JG, Miya M, Tsukamoto K, Nishida M (2003) Basal actinopterygian relationships: a mitogenomic perspective on the phylogenyof the "ancient fish". *Mol Phylogenet Evol* 26:110-120
26. Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, Sado T, Mabuchi K, Takeshima H, Miya M, Nishida M (2013) MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol. Evol.*30(11):2531-2540
27. Jacobsen MW, Hansen MM, Orlando L, Bekkevold D, Bernatchez L, Willerslev E, Gilbert MT (2012) Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp). *Mol Ecol*21(11):2727-2742
28. Jayaram KC (1999) *The Freshwater Fishes of the Indian Region*. Narendra Publishing House, Delhi, India
29. Karl SA, Toonen RJ, Grant WS, Bowen BW (2012) Common misconceptions in molecular ecology: echoes of the modern synthesis. *Mol Ecol* 21(17):4171-4189
30. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649
31. Krishnakumar K, Raghavan R, Prasad G, Bijukumar A, Sekharan M, Pereira B, Ali A (2009) When pets become pests - exotic aquarium fishes and biological invasions in Kerala, India. *Curr Sci India* 97(4): 474– 476
32. Mabuchi K, Miya M, Azuma Y, Nishida M (2007) Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. *BMC Evol Biol* 7:10
33. Masters JC, de Wit MJ, Asher RJ (2006) Reconciling the origins of Africa, India and Madagascar with vertebrate dispersal sce-narios. *Folia Primatol* 77:399-418
34. Min XJ, Hickey DA (2007) DNA asymmetric strand bias affects the amino acid composition of mitochondrial proteins. *DNA Res* 14:201-206
35. Miya M, Nishida M (1999) Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): first example of transfer RNA gene rearrangements in bony fishes. *Mar Biotechnol*1(5):416-426
36. Miya M, Nishida M (2015) The mitogenomic contributions to molecular phylogenetics and evolution of fishes: a 15-year retrospect. *Ichthyol Res* 62(2): 29-71
37. Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, SatohTP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM, Nishida M (2003) Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol Phylogenet Evol* 26:121-138
38. Mohanta SK, Swain SK, Das SP, Bit A, Das G, Pradhan S, Sundaray JK, Jayasankar P, Ninawe AS, Das P (2016) Complete mitochondrial genome sequence of *E. suratensis* revealed by next generation sequencing. *Mitochondr DNA Part B* 1(1):746-747
39. Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res* 20(7):908-916
40. Mu X, Liu Y, Lai M, Song H, Wang X, Hu Y, Luo J (2015) Characterization of the *Macropodus opercularis* complete mitochondrial genome and family Channidae taxonomy using Illumina based de novo transcriptome sequencing. *Gene* 559(2): 189-195
41. Naylor GJ, Collins TM, Brown WM (1995) Hydrophobicity and phylogeny. *Nature* 373(6515):565-566
42. Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290(5806):470-474
43. Padmakumar KG, Bindu L, Manu PS (2012) "*Etroplus suratensis* (Bloch), the State Fish of Kerala." *J Biosci* 37(1): 925-931
44. Padmakumar KG, Krishnan K, Manu PS, Shiny CK, Radhika R (2002) Wetland conservation and Management in Kerala. In: Jayakumar M (eds) Thanneermukkom barrage and fishery decline in Vembanad wetlands, Kerala. State Committee on Science, Technology and Environment, Thiruvananthapuram, Kerala, India, pp 27-36

45. Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol* 41:353-358
46. Sambrook J, Russell D (2001) Molecular Cloning: A Laboratory Manual. 3rd edn. Cold Spring Harbor Laboratory Press, New York
47. Silva G, Lima FP, Martel P, Castilho R (2014) Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. *Proc R Soc Lond B Biol Sci* 281(1792):20141093
48. Smith AG, Smith DG, Funnell BM (1994) Atlas of Mesozoic and Ceno-zoic coastlines. Cambridge University Press, New York
49. Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol Phylogenet Evol1* 2(2):105-114
50. Sparks JS, Smith WL (2004) Phylogeny and biogeography of cichlid fishes (Teleostei: Perciformes: Cichlidae). *Cladistics* 20(6):501-17
51. Sparks JS, Smith WL (2004) Phylogeny and biogeography of cichlid fishes (Teleostei: Perciformes: Cichlidae). *Cladistics* 20:501-517
52. Sparks JS, Smith WL (2005) Freshwater fishes, dispersal ability, and nonevidence: "Gondwana Life Rafts" to the rescue. *Syst Biol* 54:158-165
53. Stager M, Cerasale DJ, Dor R, Winkler DW, Cheviron ZA (2014) Signatures of natural selection in the mitochondrial genomes of *Tachycineta swallows* and their implications for latitudinal patterns of the pace of life. *Gene* 546(1):104-111
54. Stiassny MLJ (1991) Phylogenetic intrarelationships of the family Cichlidae: an overview. In: Keenleyside MHA (eds) Cichlid Fishes: behaviour, ecology and evolution. Chapman & Hall, London, pp 1-35
55. Storey BC (1995) The role of mantle plumes in continental breakup:case histories from Gondwanaland. *Nature* 377:301-308
56. Streelman JT, Karl SA (1997) Reconstructing labroid evolution with single-copy nuclear DNA. *Proc R Soc Lond B* 264:1011-1020
57. Suneetha GKB (2007). Morphological heterogeneity and population differentiation in the green chromide *Etroplus suratensis* (Pisces: Cichlidae) in Sri Lanka. *Ruhuna J Sci* 2:70-81
58. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725-2729
59. Teacher AG, Andre C, Merila J, Wheat CW (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evol Biol* 12(1):248
60. Vences M, Freyhof J, Sonnenberg R, Kosuch J, Veith M (2001) Reconciling fossils and molecules: Cenozoic divergence of cichlid fishesand the biogeography of Madagascar. *J Biogeogr* 28:1091-1099
61. Zardoya R, Vollmer DM, Craddock C, Streelman JT, Karl S, Meyer A (1996) Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proc R Soc Lond B* 263:1589-1598

# Chapter 8

## CHARACTERISING POPULATION STRUCTURE AND ADAPTIVE VARIATION OF *ETROPLUS SURATENSIS* (Bloch, 1790) USING MITOCHONDRIAL GENOME

### ABSTRACT

*Etroplus suratensis* (Bloch, 1790) is one of the most important indigenous Cichlid of the Indian subcontinents with distributed along the estuaries and brackish water lakes of India which make them ideal candidates for investigations on adaptation and selection on mitochondrial OXPHOS genes. Their habitats and populations are facing deterioration due to increasing coastal pollution and climate change. We investigated the intra specific diversity and adaptation potential of this species by analyzing Cytochrome C Oxidase 1 and control region. Besides, partial mitogenomes and low coverage RAD-sequencing of individuals from the selected geographical regions were also sequenced. Significant genetic differentiation was detected between populations from different ecoregions of India indicating restricted gene flow and population structuring. A recent decline in effective population size was evident which can be attributed to the fragmentation of many coastal habitats in addition to anthropogenic impacts like pollution and reclamation. Signals of positive and diversifying selection observed in the mitogenomes were correlated with habitat characteristics. Habitat specific mutational signals observed have adaptive significance as the populations of the study represented humid tropical climatic zones constituting rainforests in the southwest, semi-arid zones in the southeast and humid subtropical zones in the northeast regions of India. Adaptation to these environmentally heterogeneous habitats generates genotypic and phenotypic variants with specific metabolic/bioenergetic requirements. The observed adaptive mitogenome evolution may be the imprints of this geographic variability, genetic drift and selective forces imparted by the distinctive ecoregions which form their habitats. The reduction in genetic diversity observed calls for management measures to protect the natural genetic diversity of this species as successful aquaculture ventures require replenishment of genetic diversity at fixed intervals by way of the introduction of natural broodstocks.

## 1. Introduction

Cichlids are considered as important candidate species for aquaculture worldwide. They are distributed in the fresh and brackish waters of Central and South America, Africa, Madagascar, India and Sri Lanka. They exhibit interesting patterns of adaptive radiation and speciation which makes them excellent models for studying evolutionary diversification and speciation (Azuma *et al*. 2008). Cichlids in India comprise species belonging to the genus *Etroplus*, mainly *Etroplus suratensis*, *Etroplus canarensis* and *Etroplus maculatus*. *E. suratensis* is euryhaline, widely distributed in fresh and brackish water systems of peninsular India and Sri Lanka whereas *E maculatus* is confined to brackish waters of Kerala and *E. canarensis* to coastal wetlands of Karnataka. Among these, *E. suratensis* is the most abundant, found in almost all water bodies and river mouths from South Canara on the west coast to the Chilka Lake on the east coast of India (Jayaram 2010; Padmakumar *et al*. 2012) and considered as a very important candidate species for aquaculture.

The family Cichlidae comprising more than 700 species inhabit fresh and brackish waters of landmasses and hypothesized to originate from the Gondwanaland (Africa, South and Central America, India, Sri Lanka and Madagaskar) (Stiassny 2001). The lakes of Africa harbor the richest diversity of Cichlid species, where massive radiation has happened during the past 10 million years. The unique diversity in ecology, morphology and behaviour makes Cichlids good model systems for evolutionary biology, evolutionary genetics and phenotype-genotype relationship studies (Barlow 2000; Kocher 2004; Genner and Turner 2005; Seehausen 2006; Takeda *et al*. 2013; Brawand *et al*. 2014). In addition, many Cichlids are amenable to culture conditions, making them excellent candidate species for tropical and subtropical aquaculture (Bindu and Padmakumar 2012, Padmakumar *et al*. 2012; Chandrasekar *et al*. 2016).

*Etroplus suratensis*, known as ‘Karimeen’/Pearl spot in Kerala is characterized by high adaptive capacity to withstand wide range of salinity and temperature with highly efficient osmoregulation and cellular stress response mechanisms (Padmakumar *et al*. 2012; Chandrasekar *et al*. 2014) making it a popular candidate aquaculture and ornamental species in India (Padmakumar *et al*. 2012). The biology and reproductive characteristics of this species are well known. It is widely cultured in ponds, tanks, reservoirs and brackish water systems (Jayaprakas *et al*. 1990). The entire life cycle is completed either in fresh or brackish water and it breeds throughout the year with a peak during June-September and February-April (Jayakumar 2002). Even though macrophytes are the predominant food, it also ingests diatoms, molluscs, insects and animal matter (De Silva *et al*. 1984). The backwaters of Kerala are the potential source of *E. suratensis* seeds. Wild populations are recorded mainly from Kerala and Tamil Nadu, but it is also present in Goa, Andhra Pradesh, Orissa and West Bengal (Jayaram 2010; Abraham 2011). It has also been introduced to other countries like Singapore and Malaysia (Ng and Tan 2010).

Natural populations of *E. suratensis* are facing depletion due to overexploitation (Padmakumar *et al*. 2012) and habitat alterations by the disposal of solid and liquid wastes from increasing urbanisation, increasing number of tourism activities in backwaters/estuaries and threats from exotic species like *Oreochromis mossambicus* and *Trichogaster trichopterus* (Krishnakumar *et al*. 2009). In spite of that, the conservation of natural populations of this species has not attracted sufficient attention from policymakers. Some isolated attempts have been made to create no-fishing zones or aquatic sanctuaries within some of the larger estuaries in addition to captive breeding trials oriented towards conservation (Padmakumar *et al*. 2012). The major lacunae in conservation efforts are lack of information regarding its present status with respect to intra specific genetic diversity, the potential for adaptation and revival in view of the changing climate, habitat and emergence of several diseases in wild and captive populations. Some of the studies have tried to understand phylogenetic relationships and population genetic structure among *E. suratensis* populations using mitochondrial markers indicating absence of genetic structuring, but all these studies were limited by geographical coverage among sampled populations (Gunawickrama 2012; Dhanya *et al*. 2013; Chandrasekar *et al*. 2016; Alex *et al*. 2016). This is the first comprehensive study on understanding the genetic stock structure of *E. suratensis* by collecting samples from the representative of widely spaced eco-regions over India. The climate of India hosts a wide range of weather conditions across a vast geographic scale and varied topography with six major climatic subtypes, ranging from tropical wet (Koppen-Geiger climate type_Af) regions in the southwest and the island, semi-arid (Koppen-Geiger climate type_BSh) and arid (Koppen-Geiger climate type_BWh) in the west, tropical wet and dry (Koppen-Geiger climate type_Am) in the east and central and montane (Koppen-Geiger climate type_Cwc) and humid subtropical (Koppen-Geiger climate type_Cwa) climatic zones in the north (Peel *et al*. 2007). We sampled *E. suratensis* from type_Af, type_BSh and type_Cwa as it is distributed only along these climatic zones.

Even though, mitogenomes are considered neutral, some of the recent investigations have provided evidence for selection and adaptation in the mitochondrial Oxidative phosphorylation system (OXPHOS) (Bradbury *et al*. 2010; Foote *et al*. 2011; Garvin *et al*. 2015a; Teacher *et al*. 2012; Caballero *et al*. 2015) which has been correlated with wide range of environmental factors like hypoxia (Scott *et al*. 2010), heat stress (Morales *et al*. 2015), cold stress (Cheviron *et al*. 2014; Stier *et al*. 2014), nutrient availability (da Fonseca *et al*. 2008) and expression of genes (Mishmar *et al*. 2003; Garvin *et al*. 2015b). Since *E. suratensis* is widely distributed across geographic gradients, the OXPHOS system may have experienced forces of positive and purifying selection.

The present study is aimed to investigate the genetic diversity and stock structure of *E. suratensis* by collecting samples from all over India using mitochondrial genes. We also investigated the presence of positive and purifying selection in the OXPHOS system of *E. suratensis* by characterizing and comparing OXPHOS genes of fishes collected from different eco-regions of India.

## 2. Materials And Methods

### 2.1 Sample collection, DNA extraction and mitogenome sequencing

One hundred and forty (141) samples of *E. suratensis* were collected from the estuaries/river mouths of five states of India; Maharashtra (12 nos), Karnataka (12 nos), and Kerala (Kozhikode (Korapuzha) (20 nos), Kochi (Vembanadu lake) (50 nos)) along the West coast and Tamil Nadu (Mandapam) (15 nos), Andhra Pradesh (12 nos) and Odisha (Chilka lake) (20 nos) along the East coast (Fig. 8.1). The muscle tissue samples were stored in 95% ethanol at room temperature for genomic DNA extraction and DNA extracted using DNEASY blood and tissue kit (Qiagen). The quality and quantity of DNA were checked on 0.8% agarose gel and NanoDrop™ One spectrophotometer (Thermo Fisher Scientific, USA) respectively. PCR amplification of the mitochondrial Cytochrome C Oxidase (CO1) and Control region was carried out using specific primers (Table 8.S1) and sequenced in both directions. The PCR reactions for all primer pairs were performed in 25 µL reaction mixture containing 1x Q5 reaction buffer, 200 µM dNTPs, 0.5 µM forward and reverse primers, 100 ng template DNA and 0.005U Q5® High-Fidelity DNA Polymerase (New England Biolabs). PCR products were verified on 1.2% agarose gel, purified using Qiagen PCR purification kit (Qiagen) and sequenced in both directions using Big dye terminator sequencing ready reaction V 3.0 kit (Applied Biosystems, USA) on ABI 3730 automated sequencer. The sequences were assembled and aligned in MEGA 7 (Kumar *et al*. 2016) using *E. suratensis* mitogenome as a reference (GenBank accession number KU665487) (Sebastian *et al*. 2019). Two types of aligned sequence data sets were prepared; (a) dataset of the CO1 gene of 655bp and (b) data set of the control region of 523bp.

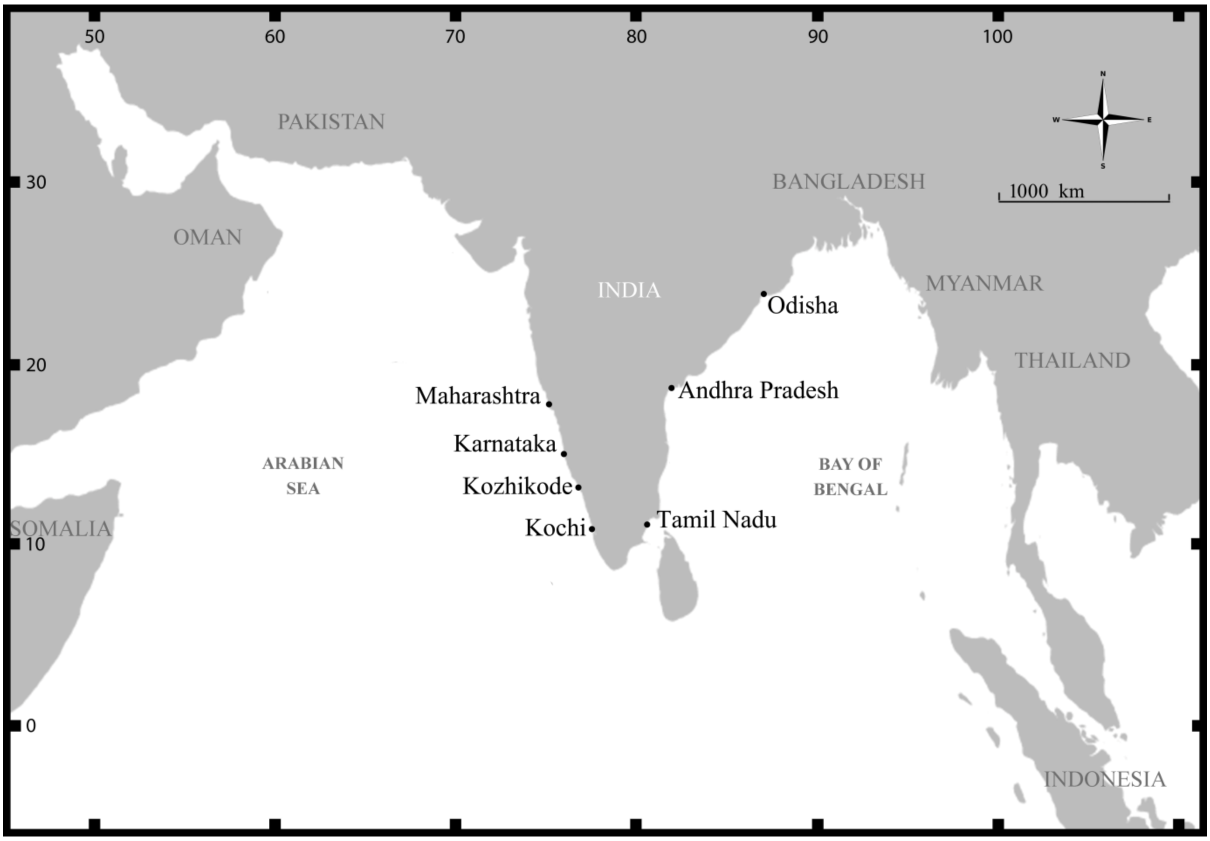


**Fig. 8.1** Map showing sampling locations of *E.suratensis*.

The partial mitogenome of 105 *E. suratensis* from Kerala (Kozhikode, 20 nos; Kochi, 50 nos), Tamil Nadu (15 nos) and Odisha (20 nos) were also amplified as overlapping segments using novel primer pairs (Table 8.S1) based on mitogenome of *E. suratensis* submitted by the previous study (GenBank accession number KU665487) (Chapter 7). The internal regions of each amplified fragment of the mitogenome were sequenced with an internal primer designed from the *E. suratensis* mitogenome (Table 8.S1). The average length of the sequenced mitogenome fragments was approximately 900bp. The targeted region included 7 protein-coding regions (ND4, ND5, ND6, CYTB, ND1, ND2, CO1) flanking control region, which shows the highest sequence variability compared to genes located away from the control region. Sequences were manually checked, aligned and assembled in MEGA 7 by using *E. suratensis* mitogenome as a reference. Two types of aligned sequence data sets were prepared; (a) Partial mitogenome data with seven protein-coding genes, 22 tRNAs, two rRNAs and non-coding control region (11881bp) and (b) concatenated seven protein-coding genes (7200bp). Unique haplotypes from each location were identified and used for further analysis.

## 2.2 Population genetic analysis of mtDNA data

The number of polymorphic sites (S), nucleotide diversity (π) (Nei 1987), haplotype diversity (Hd) (Nei 1987) and the total number of synonymous and non-synonymous mutations were estimated for all four data sets (CO1, control region, partial mitogenome and concatenated seven protein-coding genes data sets) using DnaSP v5 (Librado and Rozas 2009). The differences among samples were estimated using analysis of molecular variance (AMOVA) and estimation of F-statistics ($\Phi_{ST}$) in Arlequin 3.5 (Excoffier and Lischer 2010).

Haplotype networks were constructed for all four data sets (CO1, control region, partial mitogenome and concatenated seven protein-coding genes data sets) using the Median Joining method in popART v1.7 (Bandelt *et al*. 1999). The Akaike information criterion implemented in MEGA7 was used to select the best-fit evolutionary model for the sequences. The Bayesian phylogenetic tree was constructed using the seven concatenated protein-coding gene data (7200bp) under the GTR substitution model in BEAST v1.7.5 (Drummond *et al*. 2012). It was run with four chains for 1,100,000 MCMC generations. For all the analyses, 50000 trees were sampled and 30% of the samples discarded as burn-in. Posterior probabilities at all nodes were estimated for the remaining trees and visualized using tree viewing software FigTree (Rambaut and Drummond 2008).

The comparison using the control region was restricted to Karnataka, Kozhikode, Kochi, Tamil Nadu, Andhra Pradesh, and Odisha because the sequence length of Maharashtra samples was not sufficient for analyses. Analysis using the partial mitogenome of *E. suratensis* was restricted to Kerala (Kozhikode & Kochi), Tamil Nadu and Odisha because the length of sequences from Maharashtra, Karnataka and Andhra Pradesh samples was not sufficient for comparative analysis. All mitogenome sequences have been submitted to NCBI GenBank.

## 2.3 Analysis of historic demography

The demographic history of *E. suratensis* was analyzed using mismatch distribution (Rogers and Harpending 1992; Schneider and Excoffier 1999) in Arlequin 3.5 and DnaSP v5. Mismatch analysis was conducted for the whole population (using all the Control region and concatenated gene sequences) and for 'Kochi' samples (using Control region and concatenated gene sequences) separately. The step-wise expansion model (demographic and spatial) was evaluated using a parametric bootstrapping method comparing the fit with the expected mismatch distribution of the observed and 100 simulated mismatch distributions. The fitness of the model and smoothness of the observed distribution was validated by analyzing the sum of square deviations (SSD) and Harpending's Raggedness index (Hri; Harpending 1994). Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997) statistics were also calculated using DnaSP v5 to test for deviations from neutrality either due to the selection, bottleneck or population expansion. Changes in the effective population size through time were estimated with control region sequence data using Bayesian skyline analysis as implemented in BEAST v1.7.5. Convergence was tested by running the analysis for 10,000,000 chains under the GTR model for a strict clock model and coalescent skyline. All the parameters were automatically optimized and the skyline plot was generated by Tracer v1.6 (Drummond and Rambaut 2007). The mutation rate of 1 x 10-7/site/year as reported for the mitochondrial genome of fish (Jacobsen *et al*. 2012, McMillan and Palumbi 1997) was used for analyses following a strict molecular clock (Ho *et al*. 2011) with a generation time of eight months for *E. suratensis* (Jayakumar 2002). The skyline plots were generated by Tracer 1.5 (Rambaut and Drummond 2008). These analyses were performed only for the control region because of the availability of a standardized value of the mutational rate. Kerala samples from ‘Kochi’ were considered for demographic analyses (Mismatch analysis) as it forms a single phylogenetic lineage and the largest natural habitat (‘Vembanadu Lake’) of *E. suratensis* in India.

### 2.3 Selection analyses

Seven protein-coding genes (7200bp) datasets along with the Bayesian phylogenetic tree was used for detecting any signals of selection on mitochondrial DNA. Several methods have been used to detect positive selection and the statistical performance of each method depends on the assumptions or models. MEME uses a mixed-effects maximum likelihood approach to test the hypothesis that individual sites have been subject to episodic positive or diversifying selection (detect sites evolving under positive selection under a fixed proportion of branches) (Murrell *et al*. 2012). FUBAR uses a Bayesian approach to infer nonsynonymous (dN) and synonymous (dS) substitution rates per-site for a given coding alignment and corresponding phylogeny (Murrell *et al*. 2013). This method assumes that the selection pressure for each site is constant along the entire phylogeny. Both MEME and FUBAR are site-based detection methods (available in DATA MONKEY) (Pond and Frost, 2005), permit synonymous rate variation from site to site, and use likelihood ratio tests (LRTs) at individual sites to assess the significance of positive selection. MEME model analyzes the distribution of synonymous and non-synonymous substitution rates from site to site and branch to branch at a site. FUBAR may have more power than FEL, especially when positive selection is relatively weak and of variable strength across sites because it uses settings that are less sensitive to model specifications (HKY 85 nucleotide substitution model was used for analysis). FEL uses the maximum-likelihood (ML) approach to infer nonsynonymous (dN) and synonymous (dS) substitution rates per site for a given alignment and corresponding phylogeny (Pond and Frost, 2005). This method assumes that the selection pressure for each site is constant along the entire phylogeny. Consequently, different methods may not provide consistent results and selection analysis only determines if there is significant excess or lack of non-synonymous substitutions. For each method, we selected a threshold p-value; $p < 0.05$ for MEME, FEL, SLAC and posterior probability > 0.9 for FUBAR. TreeSAAP results are used to understand changes in the physicochemical property of amino acids caused by replacements (Woolley *et al*. 2003). Z test was used to analyze the changes in the amino acid properties, which is categorized into eight magnitude groups. The most conservative physiochemical changes are represented as category 1 and the most radical changes as category 8. (Scale of 1 to 8, the lowest and highest categories indicate stabilizing and destabilizing selection respectively). We considered only categories 6, 7 and 8 (the most radical changes) with strong statistical support ($p < 0.001$) from TreeSAAP and positively selected sites detected from all methods for further analysis.

3D homology model of the protein subunits with positively selected sites was generated by the SWISS-MODEL server (Schwede *et al.* 2003) using an appropriate subunit of the protein structure with *Boss taurus* as a template. The positively selected sites were mapped on to the three-dimensional structure.

## 2.4 ddRAD library construction sequencing and SNP genotyping

A total of ten samples were selected for ddRAD sequencing (three samples each from Kochi, Tamil Nadu, Odisha and one sample from Kozhikode). The ddRAD libraries were prepared based on the previously published protocol (Peterson *et al*. 2012). Briefly, the DNA of each sample was double digested completely with *MspI* and *EcoRI* restriction enzymes (New England Biolabs). The P1 adapter with a barcode was ligated to *EcoRI* overhang and P2 adapter was ligated to *MspI* overhang. The DNA fragment with 300bp of mean size was selected on a BluePippin (Sage Science, USA) with 2% agarose cartridge. The fragments were then PCR amplified and purified with AMPure XP Beads. The ddRAD libraries were sequenced on an Illumina HiSeq 2500 (Illumina, USA) platform with 100bp paired-end sequencing approach.

The raw reads were demultiplexed with specific barcode index and filtered using process_radtags program in STACKS V 1.40 (Catchen *et al*. 2013). Reads with low quality (Phred score <20) and uncalled bases were discarded. Lengths of the sequence were trimmed to 85bp. SNPs identification and genotype call was performed in STACKS using denovo_map.pl program. Ustacks (-m 4) constructed stacks for each sample, cstacks (-M 3; -n 3) used all individual from each population to construct a catalogue of loci and sstacks compare each sample against the catalog. Population genetic statistics (allele frequencies, percentage of polymorphic loci, nucleotide diversity, Wright's F-statistics $F_{IS}$ and $F_{ST}$) were computed using population program in STACKS.

## 2.5 Development of SNPs and microsatellite markers from ddRAD sequencing data.

Microsatellites/SSR motifs were identified from demultiplexed reads by using STR detection software (Fungtammasan *et al.* 2015), targeting di-, tri- and tetra motifs with minimum five perfect repeats.

## 3. Results

A 655bp region of the CO1 gene and 522bp of Control region from 92 individuals was analyzed for populationgenetic structure (GenBank accession numbers MH923307-MH923344, MT174050-MT174140) CO1 - Maharashtra (12 nos), Karnataka (12 nos), Kerala (Kozhikode (12 nos), Kochi (20 nos)), Tamil Nadu (12 nos), Andhra Pradesh (12 nos) and Odisha (12 nos). Control region - Karnataka (14 nos), Kerala (Kozhikode (14 nos), Kochi (22 nos)), Tamil Nadu (14 nos), Andhra Pradesh (14 nos) and Odisha (14 nos). (Table 8.1). The size of the partial mitogenomes is 11881bp, after multiple alignments. It included seven protein-coding regions (ND1, ND2, CO2 (Partial), ND4, ND5, CYTB, and ND6), 17 tRNAs, 2 rRNAs and a control region (Table 8.S2). The size of the complete mitochondrial genome of *E. suratensis* is 16465bp (NCBI GenBank Accession number KU665487). 105 annotated mitogenomes have been submitted to NCBI, GenBank (Accession numbers MH923307-MH923344).

### 3.1 Population genetic analysis of mtDNA data

The analysis of the CO1 gene of *E. suratensis* revealed 25 haplotypes with the most common haplotype shared among 67 individuals. There were 48 variable sites and 8 parsimony informative sites. Overall haplotype diversity (Hd) and nucleotide diversity (π) were 0.487 and 0.00186 respectively. Analysis of the control region revealed the presence of 35 haplotypes with four major haplotypes representing KOCHI, TAMIL NADU, MAHARASHTRA and ODISHA. The nucleotide diversity (π) and haplotype diversity (Hd) values were estimated as 0.0123 and 0.96 respectively (Table 8.1).

**Table 8.1** Summary of genetic diversity statistics for CO1, Control region, partial mitogenome nucleotide and concatenated protein coding gene sequences of *E. suratensis.*

| | No of sample | **S**[a] | **π** [b] | **No of haplotype** | **Hd**[c] | **K**[d] | **Number of Synonymous sites** | **Number of Non-synonymous sites** | **Θe** | **Tajima'S D** | **Fu's Fs** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CO1(640bp) | 92 | 48 | 0.00186 | 25 | 0.487 | 1.214 | 8 | 40 | 9.424 | -2.78 (P<0.001) | -27.629 (P>0.1) |
| Control(522bp) | 92 | 42 | 0.0123 | 35 | 0.961 | 6.369 | - | - | 8.246 | -0.72 (P>0.1) | -12.58 (P>0.1) |
| Genome (11881bp) | 105 | 174 | 0.0026 | 37 | 0.932 | 23.27 | 1853 | 5346 | 44.806 | -1.84805 (P>0.1) | -14.531 (P>0.1) |
| Concatenated Gene (7200bp) | 105 | 112 | 0.003 | 29 | 0.998 | 15.53 | 1853 | 5346 | 29.28 | -1.80101 (P<0.01) | -15.066 (P < 0.001) |

[a]number of polymorphic sites, [b]nucleotide diversity, [c]haplotype diversity, [d]average number of pairwise nucleotide differences, [e]theta per sequence from the total number of mutations.

The Maximum likelihood phylogenetic tree for CO1 sequences did not resolve any structure, whereas the control region sequence revealed distinct clustering among population samples (Fig. 8.S1). The presence of independent lineages representing Karnataka, Kozhikode, Kochi, Tamil Nadu, Andhra Pradesh, and Odisha was reported with some mixing between lineages. Kozhikode formed sister lineage to all the other lineages. Haplotype network of CO1 sequence was star-shaped with one haplotype common to all the populations and other haplotypes differing by a few mutational steps (Fig. 8.S2). The haplotype network of the control region revealed spatial structuring of samples, consistent with the pattern in the phylogenetic tree (Fig. 8.S3).

Analysis of the 11881bp partial mitogenome data set from 105 samples revealed low nucleotide diversity (π, 0.0026) and high haplotype diversity (Hd, 0.998) with 37 haplotypes. There were 174 variable sites (S) and 90 parsimony informative sites with an average number of nucleotide differences (k) being 23.27. In the concatenated seven protein-coding genes (7200bp) dataset, 29 haplotypes were present with haplotype (Hd) diversity of 0.998 and nucleotide diversity (π) of 0.0003. The total number of variable sites (S) was 112 with the average number of nucleotide differences (k) being 15.53. The basic statistics of the CO1, control region, partial mitogenome sequence and concatenated seven protein-coding genes are presented in Table 8.1.

The Bayesian phylogenetic trees of partial mitogenome indicated the geographical clustering by high posterior probability in the interior branches leading to the major lineages (Fig. 8.3). Tamil Nadu, Odisha, Kochi, and Kozhikode formed lineages 1, 2, 3 and 4 respectively with Kozhikode being the sister lineage. These findings were further corroborated in the haplotype network of partial mitogenomes (Fig. 8.2) and other data sets (Fig. 8.S1; Fig. 8.S2; Fig. 8.S3).

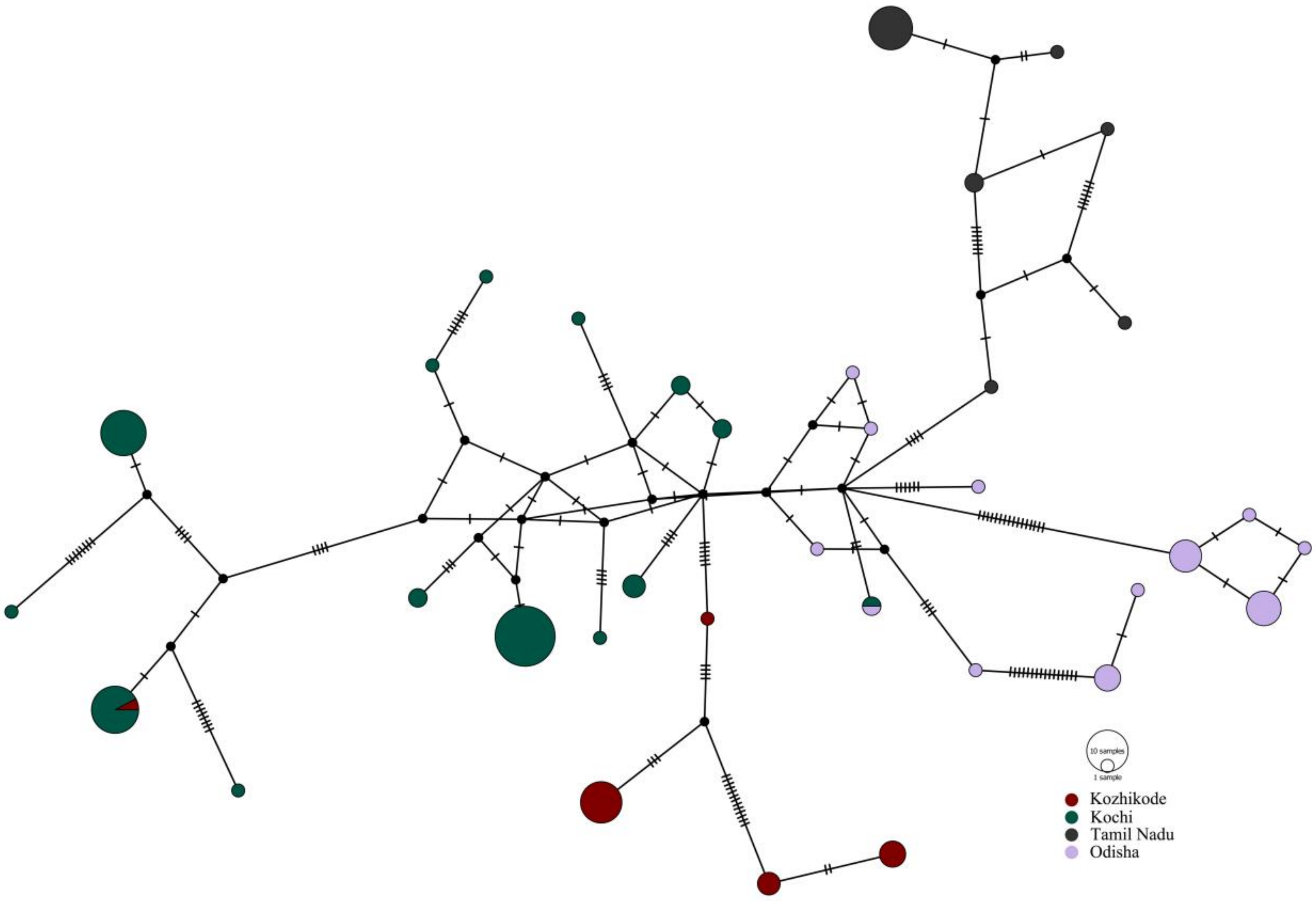


**Fig. 8.2** Haplotype network diagram constructed using partial mitogenome of *E.suratensis* with a median joining method. Haplotypes are represented in circles and colors indicate geographical locations. Mutational steps are indicated as vertical stripes.

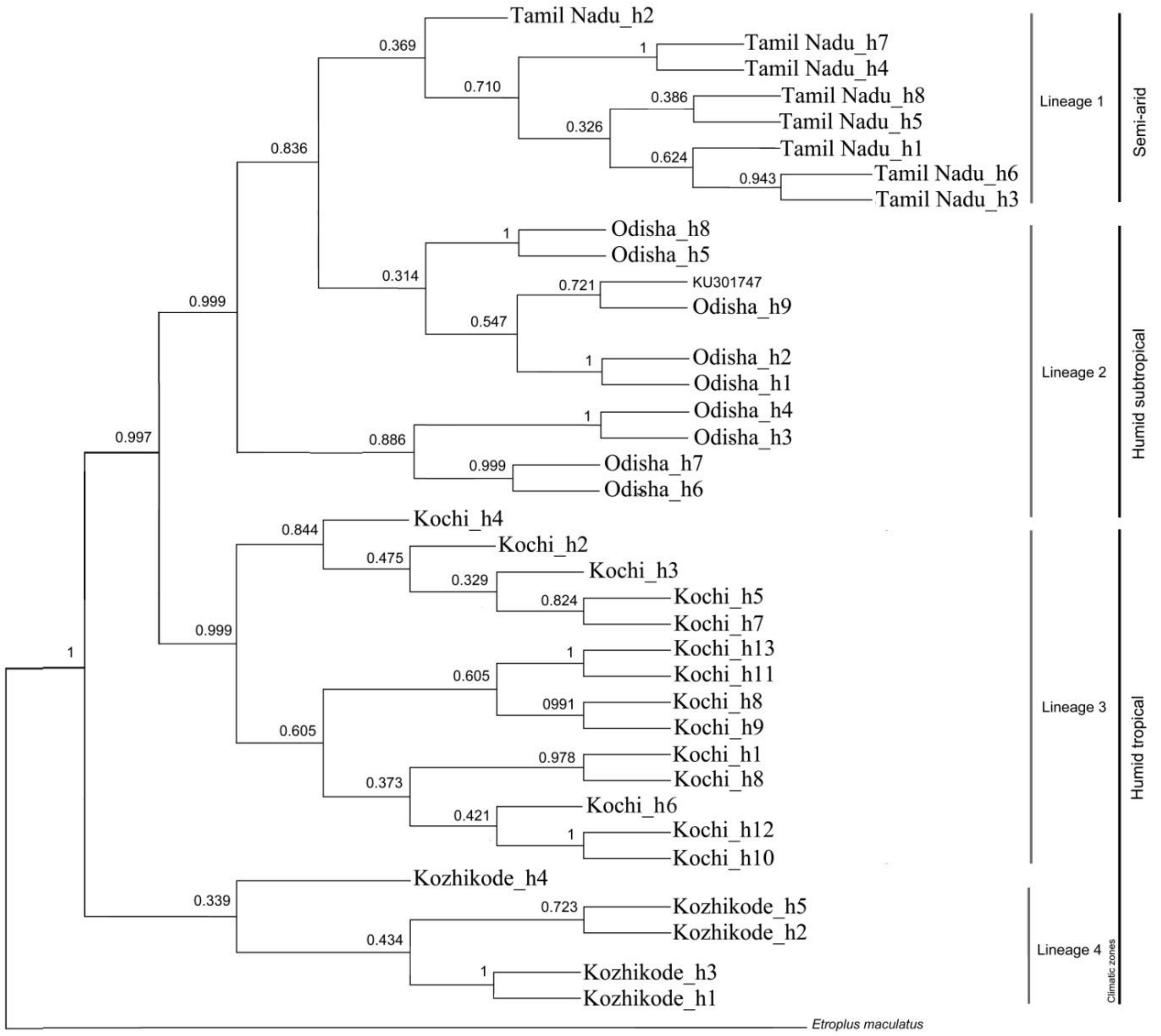


**Fig. 8.3** Bayesian tree for partial mitogenome nucleotide sequences of *E. suratensis. E. maculatus* (GenBank accession number NC_009587) was used as an outgroup to root the tree. Posterior probability values for node support are shown. Refer Fig. 8.1 for Site Name and Sample ID.

Both the Control region and partial mitogenome of *E. suratensis* showed a significant global $\Phi_{ST}$ value of 0.41 and 0.40 respectively (Table 8.S3). Genetic differentiation among samples from different geographical locations was further evident from pairwise $\Phi_{ST}$ comparisons. Significant $\Phi_{ST}$ values were obtained between Karnataka, Kozhikode, Kochi, Tamil Nadu, Andhra Pradesh, and Odisha (Table 8.2) when control region data was analysed. Similar results were obtained in partial mitogenome comparisons with the highest $\Phi_{ST}$ value between Kozhikode and Tamil Nadu samples (Table 8.2).

The mismatch distribution plots of both the control region and concatenated gene sequences showed a bimodal pattern for whole samples (Fig. 8.S4) and Kochi samples indicating historically-stable/declining population size (Rogers and Harpending 1992; Schneider and Excoffier 1999). This was further supported by non-significant values of Fu's Fs and Tajima's D (-12.58 and -0.72 respectively, for control region; -15.066 and -0.658 respectively, for concatenated genes) (Table 8.1) (Tajima 1989; Fu 1997). The bayesian skyline plot of the control region sequence indicated a historically stable population with a recent decline in effective population size (Fig. 8.4).

**Table 8.2** Pairwise $\Phi_{ST}$ for the control region and concatenated genes sequences of *E. suratensis.*

| Control region | | | | | | |
|---|---|---|---|---|---|---|
| | ANDHRA PRADESH | KOCHI | KOZHIKODE | MAHARASHTRA | ODISHA | TAMILNADU |
| ANDHRA PRDESH | 0 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| KOCHI | 0.30899 | 0 | <0.009 | <0.001 | <0.001 | <0.001 |
| KOZHIKIDE | 0.58153 | 0.53703 | 0 | <0.001 | <0.018 | <0.001 |
| MAHARASHTRA | 0.65531 | 0.59585 | 0.54408 | 0 | <0.001 | <0.001 |
| ODISHA | 0.46603 | 0.37636 | 0.67434 | 0.64279 | 0 | <0.001 |
| TAMILNADU | 0.16332 | 0.16725 | 0.42179 | 0.46555 | 0.26919 | 0 |

| Concatenated genes | | | | |
|---|---|---|---|---|
| | KOZHIKIDE | KOCHI | TAMILNADU | ODISHA |
| KOZHIKIDE | 0 | <0.001 | <0.001 | <0.001 |
| KOCHI | 0.63724 | 0 | <0.001 | <0.001 |
| TAMILNADU | 0.78705 | 0.65188 | 0 | <0.001 |
| ODISHA | 0.59807 | 0.48044 | 0.60774 | 0 |

The numbers below the diagonals are $\Phi_{ST}$ and the number above the diagonal are the probability value.

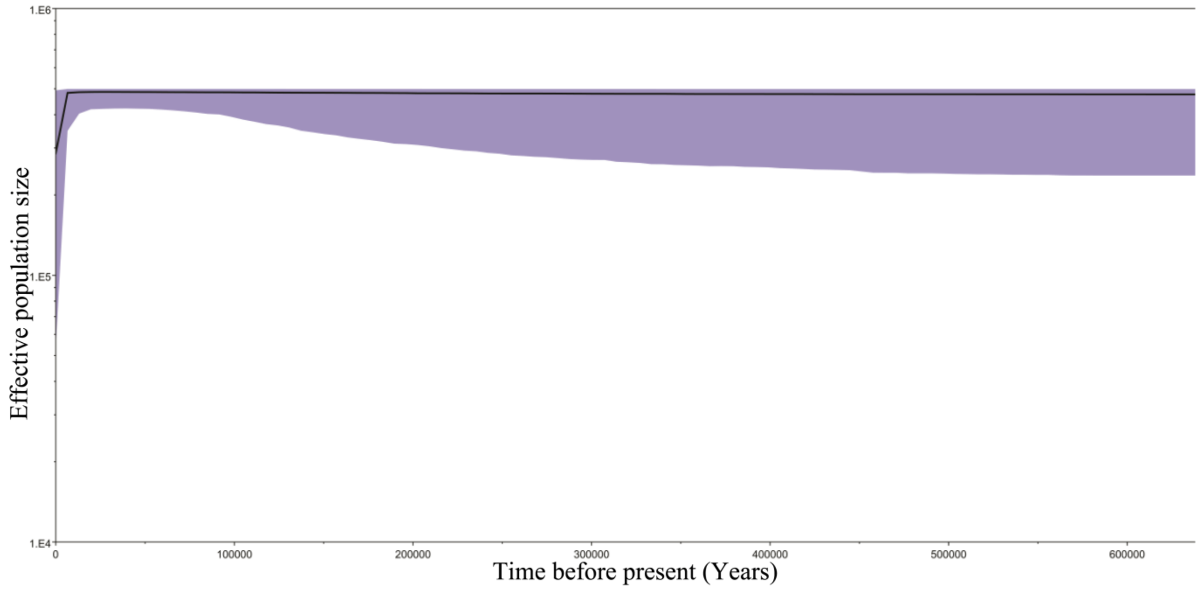


**Fig. 8.4** Bayesian skyline plot constructed using the control region sequence of *E. suratensis.*

### 3.2 Selection analyses

Fu’S Fs and Tajima’s D (-1.84) values (Tajima 1989) were negative (-14.53) indicating excess number of alleles and low-frequency polymorphisms relative to expectation probably from genetic hitchhiking (Fu and Li 1993), and selective sweep and/or purifying selection (Ramos-Onsins and Rozas, 2002) (Table 8.1).

MEME found evidence of episodic positive/diversifying selection at 32 sites (p<0.05). FUBAR inferred 8 sites subject to diversifying positive selection with posterior probability > 0.9. FEL found evidence of pervasive positive/diversifying selection at seven sites and negative/purifying selection at 10 sites whereas SLAC detected pervasive positive/diversifying selection at one site and pervasive negative/purifying selection at nine sites with significant p-values (p<0.05) (Table 8.3, Table 8.S4). Codons associated with significant (p = 0.001) radical (categories 6, 7, 8) changes in physicochemical properties of amino acids were detected in the *E. suratensis* mitochondrial protein-coding genes. Some of the differences in amino acids fixed between the lineages were associated with significant radical physicochemical changes (Table 8.3). Seven of the 28 codons identified by MEME as undergoing episodic positive selection were also associated with significant radical physicochemical properties changes in TreeSAAP analysis (Table 8.3). Among the 50 codons identified by all four methods, 9, 3, 4, 7, 25 and 2 codons were located in ND1, CO1, ND4, ND5, CYTB, and ND6 respectively.

**Table 8.3** Codons that are under positive selection in *E. suratensis* OXPHOS complex, based on three selection tests: FUBAR, MEME, FEL, SLAC and TreeSAAP method.

| Gene | Codon position | From Codon To Codon | From AA To AA | MEME | FUBAR | FEL | SLAC | TreeSAAP | Predicted function of amino acid residue | Distribution of amino acid replacements across lineages |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | p-value (<=0.05) | posterior probability (>= 0.9) | p-value (0.05) | p-value (0.06) | Significant properties, category of amino acid changes 6-7-8(+) | | |
| ND1 | 22 | GCC-TCC | Ala-Ser | 0.03 | NS | NS | NS | NS | - | L2 |
| ND1 | 29 | GTT-ATT | Val-Ile | 0.03 | NS | 0.045 | NS | NS | - | L3,4 |
| ND1 | 39 | CTT-CTC | Leu-Phe | 0.03 | NS | NS | NS | NS | - | L3 |
| ND1 | 47 | GGC-TGC | Gly-Cys | NS | NS | NS | NS | Refractive_index | - | L2 |
| ND1 | 48 | CCC-TTT, TCC | Pro-Phe, Ser | NS | NS | NS | NS | Chromatographic_index/Solvent_accessible_reduction_ratio | - | L2 |
| ND1 | 55 | ATT-GTT | Ile-Val | 0.01 | 0.984 | 0.032 | NS | NS | - | L1,2,3 |
| ND1 | 137 | GGG-GAG, GCG | Gly-Glu, Ala | 0.01 | 0.9887 | 0.002 | 0.05 | NS | - | L1,3 |
| ND1 | 143 | GCC-ACC, CCC | Ala-Thr,Pro | 0.01 | 0.9888 | 0.012 | 0.05 | NS | Proton translocation | L1,2,3,4 |
| ND1 | 169 | CAA-CCA | Gln-Pro | 0.01 | 0.9712 | NS | NS | Compressibility | - | L2,3,4 |
| CO1 | 30 | GGC-GAC | Gly-Asp | NS | NS | NS | NS | Polar_requirement | - | L3 |
| CO1 | 50 | GAC-GAA | Asp-Glu | 0.03 | NS | NS | NS | NS | - | L3 |
| CO1 | 98 | AAC-AAA | Asn-Lys | NS | NS | NS | NS | Isoelectric_point | D-pathway | L2,3 |
| ND4 | 197 | TGA-GGA | Trp-Gly | 0.01 | NS | NS | NS | NS | - | L2,3 |
| ND4 | 259 | AGC-ATC | Ser-Ile | 0.01 | 0.9492 | NS | NS | Bulkiness/Chromatographic_index/Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity | - | L1,3 |
| ND4 | 358 | ACC-CCC | Thr-Pro | NS | NS | NS | NS | Compressibility | (Proton_antipo_M) Proton-conducting membrane transporter | L3 |
| ND4 | 378 | GGC-CGC | Gly-Arg | 0.01 | NS | NS | NS | Helical_contact_area/Isoelectric_point/Molecular_volume/Molecular_weight/Partial_specific_volume/Refractive_index | (Proton_antipo_M) Proton-conducting membrane transporter | L3 |
| ND5 | 302 | AAT-AAA | Asn-Lys | NS | 0.9197 | NS | NS | Isoelectric_point | (Proton_antipo_M) Proton-conducting membrane transporter | L2,3 |
| ND5 | 340 | TTC-TTG | Phe-Leu | 0.04 | NS | NS | NS | NS | (Proton_antipo_M) Proton-conducting membrane transporter | L1 |
| ND5 | 347 | CTA-ATA | Leu-Met | NS | 0.9075 | NS | NS | NS | (Proton_antipo_M) Proton-conducting membrane transporter | L3,4 |
| ND5 | 356 | CTC-ATC | Lue-Ile | NS | 0.9045 | NS | NS | NS | (Proton_antipo_M) Proton-conducting membrane transporter | L3 |
| ND5 | 368 | ATA-TTA | Met-Leu | 0.02 | NS | NS | NS | NS | (Proton_antipo_M) Proton-conducting membrane transporter | L1,3 |
| ND5 | 399 | GAT-GAA | Asp-Glu | 0.03 | NS | NS | NS | NS | (Proton_antipo_M) Proton-conducting membrane transporter | L1 |
| ND5 | 554 | GCA-GAA | Ala-Glu | NS | NS | NS | NS | Polar_requirement | NADH5 C-terminus | L2,4 |
| CYTB | 15 | AAT-AAA | Asn-Lys | NS | NS | NS | NS | Isoelectric_point | - | L2 |
| CYTB | 18 | CTA-ATA | Leu-Met | 0.04 | NS | 0.045 | NS | NS | Intrachain domain | L2 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | interface | |
| CYTB | 21 | CTT-CGT | Leu-Arg | NS | NS | NS | NS | Isoelectric_point/Polarity | - | L2 |
| CYTB | 22 | CCT-CAT | Leu-Pro | 0.03 | NS | 0.041 | NS | NS | Interchain domain interface [polypeptide binding] | L2 |
| CYTB | 33 | TTT-TTG | Phe-Leu | 0.04 | NS | NS | NS | NS | - | L2 |
| CYTB | 51 | CTT-CAA | Leu-Gln | 0.001 | NS | 0.037 | NS | Polarity | Heme bL binding site [chemical binding] | L2 |
| CYTB | 63 | TTC-TTA | Phe-Leu | 0.05 | NS | NS | NS | NS | Interchain domain interface [polypeptide binding] | L2 |
| CYTB | 74 | AAC-AAA | Asn-Lys | NS | NS | NS | NS | Isoelectric_point | Interchain domain interface [polypeptide binding] | L2 |
| CYTB | 78 | TTC-CTC, ATC | Phe-Ile,Leu | 0.02 | NS | NS | NS | NS | interchain domain interface [polypeptide binding] | L2,3 |
| CYTB | 83 | CAT-CAA | His-Qln | 0.03 | NS | NS | NS | NS | heme bL binding site [chemical binding] | L3 |
| CYTB | 84 | GCC-AAC | Ala-Asn | 0.02 | NS | NS | NS | NS | heme bL binding site [chemical binding] | L3 |
| CYTB | 85 | AAT-AAA | Asn-Lys | NS | NS | NS | NS | Isoelectric_point | intrachain domain interface | L2 |
| CYTB | 91 | TTC-ATC, GTC | Phe-Ile, Val | 0.02 | NS | NS | NS | NS | intrachain domain interface | L2 |
| CYTB | 106 | TCT-TAT | Ser-Tyr | NS | NS | NS | NS | Partial_specific_volume | intrachain domain interface | L2 |
| CYTB | 107 | TAC-GAC, TGC | Tyr-Asp,Cys | 0.01 | NS | NS | NS | Polar_requirement/Polarity | intrachain domain interface | L2 |
| CYTB | 109 | TAC-TGC | Tyr-Cys | 0.05 | NS | NS | NS | NS | intrachain domain interface | L2 |
| CYTB | 226 | TAC-GAC | Tyr-Asp | 0.05 | NS | NS | NS | Polar_requirement/Polarity | interchain domain interface [polypeptide binding] | L3 |
| CYTB | 232 | TTT-TGT | Phe-Cys | 0.04 | NS | NS | NS | NS | intrachain domain interface | L3 |
| CYTB | 238 | GCC-ACC | Ala-Thr | 0.03 | NS | NS | NS | NS | - | L3 |
| CYTB | 239 | CTT-CGT | Leu-Arg | NS | NS | NS | NS | Isoelectric_point/Polarity | intrachain domain interface | L3 |
| CYTB | 254 | GAC-GGC | Asp-Gly | 0.05 | NS | NS | NS | NS | intrachain domain interface | L3 |
| CYTB | 258 | CCT-CTT | Pro-Leu | NS | NS | NS | NS | NS | intrachain domain interface | L3 |
| CYTB | 259 | GCC-CCC | Ala-Pro | 0.04 | NS | NS | NS | Solvent_accessible_reduction_ratio/ Surrounding_hydrophobicity | intrachain domain interface | L3 |
| CYTB | 263 | GTA-TGA | Val-Trp | 0.03 | NS | NS | NS | NS | intrachain domain interface | L3 |
| CYTB | 265 | CCC-CTC | Pro-Leu | NS | NS | NS | NS | Solvent_accessible_reduction_ratio/ Surrounding_hydrophobicity | intrachain domain interface | L3 |
| ND6 | 93 | GCA-GTA | Ala-Val | 0.04 | NS | NS | NS | NS | - | L4 |
| ND6 | 105 | TGT-TAT | Cys-Tyr | NS | NS | NS | NS | Chromatographic_index | - | L2,3 |

CO2 - Cytochrome c oxidase subunits 2, CYTB - Cytochrome b, ND1 - NADH dehydrogenase subunits 1, ND4 - NADH dehydrogenase subunits 4, ND5 - NADH dehydrogenase subunits 5, ND6 - NADH dehydrogenase subunits 6

Even though most of these amino acid changes were associated with few individuals, some of them were associated with specific lineages. Amino acid changes at four sites (site #29-ND1, site #347-ND5, site #238-CYTB, site #93-ND6) were restricted to lineage 4 and two sites to (site #340-ND5 and #399-ND5) lineage 1 (Fig. 8.5). Some of the amino acid changes (site #55-ND1 and #143-ND1) were restricted to lineages 1, 2 and 3.

Amino acid substitutions fixed at site 169-ND1 have been shared between Kochi (lineage 3) & Kozhikode (lineage 4) samples even though it is dominantly occurring in Kozhikode (Lineage 4) samples. Similarly, substitution at site 368-ND5 is shared among Tamil Nadu (Lineage 1) & Kochi (lineage 3) samples and dominant in Tamil Nadu. Most of the sites identified in CYTB are restricted to few individuals in Lineage 2, 3 and 4 (Fig. 8.5).

Some of the sites identified as under positive or diversifying selection (Table 8.3) were associated with key functional regions in the mitochondrial OXPHOS complex. In the mitochondrial complex I, (NADH: ubiquinone oxidoreductase) most of the sites (14 sites) that exhibited signatures of positive selection were located in the transmembrane helix (ND1 site- #22, #29, #55, #137, #143, #169; ND4 site- #197, #259, #358, #378; ND5 site- #340, #347, #399; ND6 site- #93, #105), however some sites were restricted to the predicted internal-helix loop region (ND1 site- #39, #47, #48; ND5 site- #302, #356, #368, #554; ND6 site- #105) of their respective proteins (Fig. 8.6). Among these, the site #143-ND1 is involved in proton translocation through ND1 and most of the sites in ND4 (#358-ND4, #378-ND4) and ND5 (#302-ND, #340-ND, #347-ND, #356-ND, #368-ND, #399-ND5) are associated with proton-conducting membrane transporter (Proton_antipo_M) in Complex I (Zhu *et al*. 2016).

Majority of amino acid sites that have been suggested to participate in Qo binding, Qi binding and chemical binding were conserved in CYTB (complex III) (Crofts 2004). However, in CYTB most of the sites under positive selection were located in the transmembrane helix (site- #15, #18, #33, #51, #63, #78, #83-85, #91, #226, #232, #238, #239) and in the beta-sheet (site- #21, #22, #258, #259, #263, #265) (Fig. 4). Among the sites in CYTB, some are associated with intrachain domain interface (site- #18, #85, #91, #106, #107, #109, #232, #239, #254, #258, #259, #263, #265), polypeptide binding (site- #22, #226, #63, #74, #78) and Heme bL binding (site- #83, #84, #51) in complex III.

Sites under positive selection in Cytochrome c oxidase (COX1) (complex IV) occurred in the intrahelix loop (CO1 site- #50) and the transmembrane helix (CO1 site- #30, #98) (Fig. 8.6). The positively selected site 98-CO1 is involved in proton translocation through D-pathway in complex IV (Li *et al*. 2006). Whereas the amino acid residues that have been reported to participate in electron transfer pathway (F#377, R#438, R#439), D-pathway (Y#19, N#80, D#91, N#98, S#101, S#156, S#157, N#163, T#167), putative water exit pathway (D#227, G#232, H#233, D#364, H#368, D#369, R#438), ion binding (Binuclear center-heme a3/CuB) (H#240, H#290, H#291, H#376), K-pathway (H#240, Y#244, S#255, H#290, H#291, T#316, K#319), putative proton exit pathway (H#291, H#368, D#369, R#438, R#439), and chemical binding (Low-spin heme a binding site) (H#61, H#378, S#382, T#424, S#461) in CO1 (Tsukihara *et al*. 1995) were conserved. In addition to that, amino acid residues that have been reported to participate in CuA binding, polypeptide binding (in the subunit interface) and phospholipid binding in COX1 are also conserved across all individuals in this study.

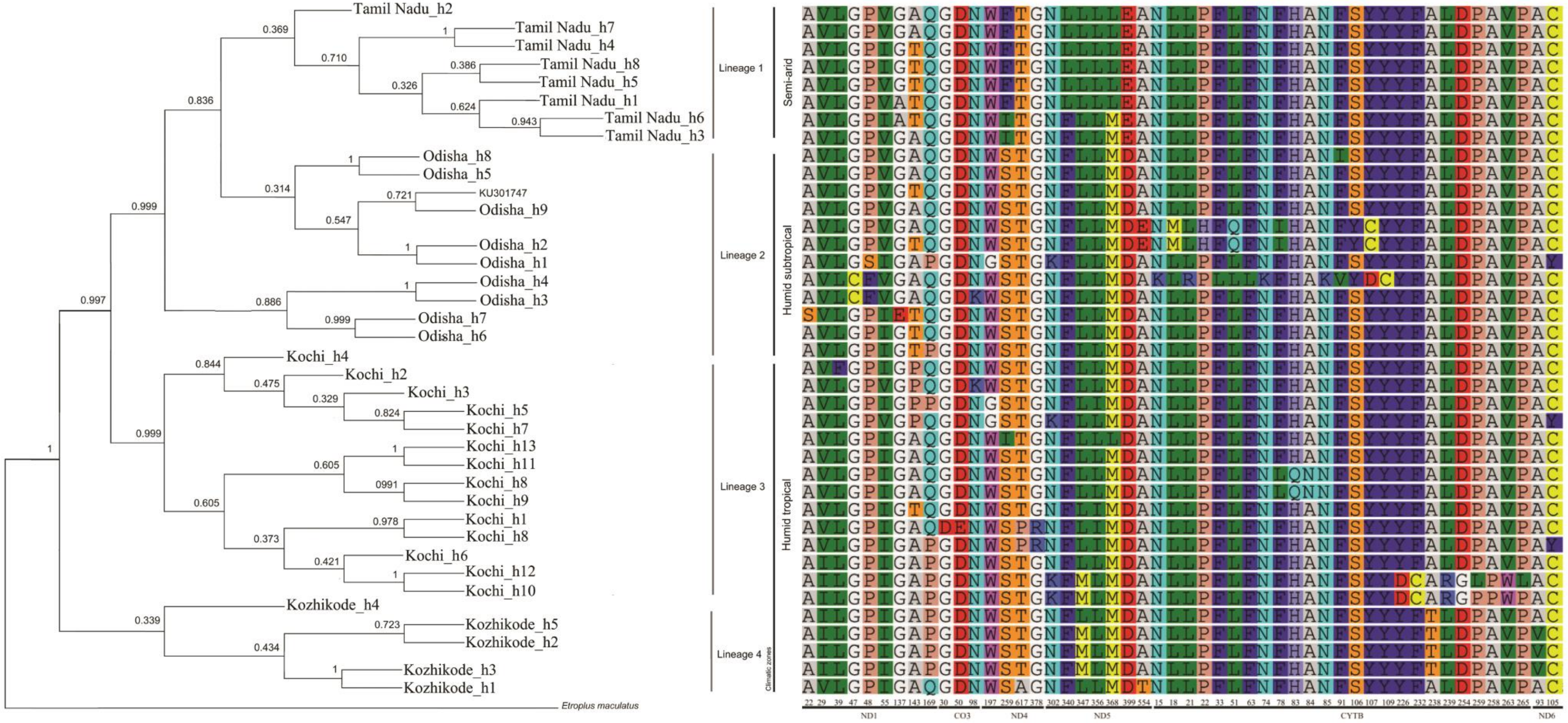


**Fig. 8.5** Positively selected sites in the *E. suratensis* mitogenome OXPHOS complex represented in the phylogenetic tree (Bayesian tree). Refer Fig. 8.1 for Site Name and Sample ID.

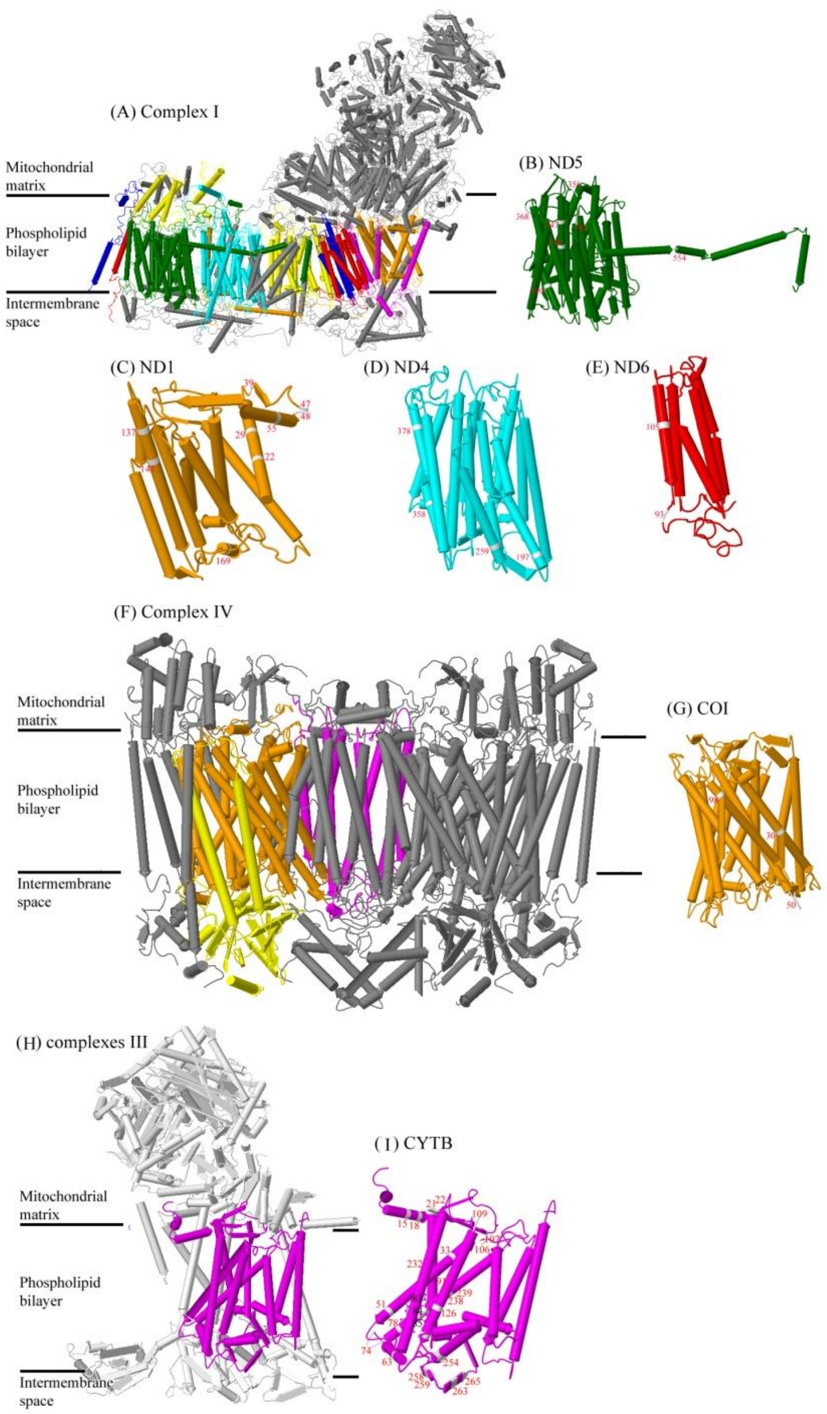


**Fig. 8.6** Amino acid property variation in dehydrogenase (Complex I), Cytochrome C Oxidase (Complex IV) and Cytochrome b (complex III). (A) individual OXPHOS Complex I, with mitochondrial-encoded subunits are represented in different colors as followed: ND2 in yellow; ND4L in blue; ND1 in orange; ND3 in magenta; ND4 in cyan; ND5 in green; ND6 in red. Grey structures represent nuclear-encoded subunits. Individual core subunits (B) ND1, (C) ND4, (D) ND5 and (E) ND6 with white color on positively selected amino acid sites. (F) Individual OXPHOS Complex IV (Homodimer) with mitochondrial-encoded subunits is represented in different colors as followed: CO1 in orange; CO2 in yellow; CO3 in magenta. Grey structures represent nuclear-encoded subunits. (G) Individual core subunit CO1 with white color on positively selected amino acid sites. (H) Individual OXPHOS Complex III with mitochondrial-encoded subunit is represented in magenta. Grey structures represent nuclear-encoded subunits. (I) Individual core subunit CYTB with white color on positively selected amino acid sites**.**

### 3.3 Genomic variation analysis using ddRAD data

Among the ten samples used, only six with more than 100000 raw read counts were selected for further analysis (two samples each from Odisha, Tamil Nadu and Kochi). The overall nucleotide diversity (π) in *E. suratensis* populations was 0.0001 and 0.39 to 0.301 for variant positions. The average major allele frequency (P) was 0.999 and the average observed heterozygosity 0.0001 in the entire data set (Table 8.S5). Whereas P decreased to a range from 0.81 - 0.805 and the observed average heterozygosity increased to 0.3904 - 0.3214 for polymorphic positions. $R_{ST}$ for all three pairwise comparisons of populations in the present study ranged from 0.6 - 0.4 (Table 8.4) but P-value of exact G test was not significant. Large numbers of fixed differences were observed (SNPs with an $F_{ST}$ of 1.0) in the comparison between populations. Since the sample number used in ddRAD data was very low we could not make any further discussions about the results.

**Table 8.4.** Pairwise comparison of genetic distance ($R_{ST}$) among *E.suratensis* populations. Below diagonal; genetic divergence among populations as measured by $R_{ST}$.

| Pop ID | ODISHA | KOCHI | TAMIL NADU |
|---|---|---|---|
| ODISHA | | 0.321 | 0.429 |
| KOCH | 0.4 | | 0.334 |
| TAMIL NADU | 0.5 | 0.6 | |

Above diagonal; P-value of exact G test for each population pair across all loci by Fisher's method

## 4. Discussion

Significant genetic differentiation was evident between *E. suratensis* samples from different water bodies of the Indian subcontinent indicating restricted gene flow. The very low nucleotide diversity and recent reduction ineffective population size as indicated in Bayesian skyline plot call for augmentation of natural genetic diversity by protecting its habitats from pollution, reclamation, and invasive species. Successful aquaculture ventures require enhancement of genetic diversity and prevention of inbreeding by the supply of brood stocks from wild at fixed intervals and hence it is imperative to devise management measures for the protection of its natural diversity.

Even though many of the brackish water estuaries in India are connected, the sampling sites on the West and East Coast of India are not connected to the extent of contributing to substantial gene flow between populations (Saravanan *et al.* 2013). This may be the reason for genetic subdivisions as indicated by significantly high global $\Phi_{ST}$ (0.4-0.6) and pairwise $\Phi_{ST}$ (0.16-0.78) values for *E. suratensis* populations. This was further corroborated by a phylogenetic tree in which distinct lineages were present corresponding to the geographical locations along with discrete clustering in the haplotype network diagram. Previous studies on genetic structuring of *E. suratensis* collected from Kerala waters reported the absence of genetic differentiation which may be due to the limited geographic coverage and lower resolving power of markers (CO1) (Gunawickrama 2012; Dhanya *et al.* 2013; Chandrasekar *et al.* 2016; Alex *et al.* 2016). But none of the previous investigations has addressed genetic structuring of *E. suratensis* from different eco-regions of India.

The limited migration capabilities of *E. suratensis* along with reduced larval dispersal may be a factor contributing to genetic subdivisions (Cadrin *et al.* 2013). The reproductive strategy of *E. suratensis* is characterized by pairing, nest making, pit nursing and parental care (Padmakumar *et al.* 2012) which limits the dispersal capacity of the young ones. Lack of connectivity between inland water bodies along with reduced population size also will exacerbate the effect of genetic drift on the *E. suratensis* population structure (Pavlova *et al.* 2017). In contrast, many marine fishes are characterized by reduced genetic differentiation and high gene flow due to their capacities tomigrate as well as the ability of larvae to disperse to longer distances (Siddall *et al.* 2003).

*E. suratensis* samples used in the present study represented three climatic zones, humid tropical climatic zones supporting rainforests in southwest India, semi-arid places in the south-east and humid subtropical zones in the northeast. A chain of brackish water systems and rivers connects the south-west coast bordering the state of Kerala, India. Kerala has 41 west-flowing rivers reaching the Arabian Sea through estuaries (Saravanan *et al.* 2013). Vembanad Lake (Kochi), the longest lake in India is one among them. The sampling site Kozhikode (Korapuzha) is located in the northern region of brackish water systems bordering the state of Kerala. The brackish water systems and rivers in Tamil Nadu (Mandapam) are connected to the Palk Bay in the Bay of Bengal. Rivers and Lakes situated along the east (Andhra Pradesh and Odisha) coast of India is connected to the Bay of Bengal. The brackish water systems of Maharashtra, Karnataka, and Kerala on the West coast and Tamil Nadu, Andhra Pradesh and Odisha on the East coast of India exhibit wide seasonal variations in important water quality parameters like temperature, salinity, pH and dissolved oxygen which may act as selective forces on the genome of organisms inhabiting these water bodies. Thus, the significant genetic differentiation observed between populations indicates that these fishes may be isolated and adapted to the geographical locations where they inhabit. In a species with low effective population size, fixation of mutations will occur due to the action of genetic drift in addition to the effect of the environment (Hauser and Carvalho 2008; Harrisson *et al*. 2016).

The signals of very low nucleotide diversity (~0.0025), a recent decline in effective population size detected in Bayesian skyline plots and mismatch analysis corroborated the outcome of surveys that documented a decline of *E. suratensis* populations in natural habitats (Kurup and Thomas 2001; Padmakumar *et al*. 2002). The reasons for contemporary low effective population size may be due to the ecological deterioration of Green Chromide habitats due to sea-level fluctuations, land reclamation and pollution (Padmakumar *et al*. 2012; Krishnakumar *et al.* 2009) in addition to overfishing, bottleneck and founder effects (Jayaram 2010; Siddall *et al.* 2003). Thus, the observed low genetic diversity and reduction in effective population size call for augmenting conservation efforts for this species which is very important in aquaculture (Cadrin *et al.* 2013). Even though *E. suratensis* is a candidate species for aquaculture, the intentional introduction to the natural habitat by activities like aquaculture would also have reduced genetic diversity because it homogenized allele frequencies (Husemann *et al.* 2012; Crook *et al.* 2015).Significant genetic differentiation between populations in the present study indicates a lack of connectivity driven byhabitat fragmentation in addition to the fixation of mutations due to genetic drift. The selective forces of the environment also will be important. The reasons for reduction in contemporary effective population size and intra-specific genetic diversity may be due to anthropogenic activities, over-exploitation and competition from invasive species. In addition, the ecological deterioration of Green Chromide habitats due to sea-level fluctuations, land reclamation and pollution (Padmakumar *et al*. 2012; Krishnakumar *et al*. 2009) also accelerate bottleneck and founder effects (Jayaram 2010; Siddall *et al*. 2003). Successful aquaculture ventures require enhancement of genetic diversity and prevention of inbreeding by the supply of brood stocks from wild at fixed intervals and hence it is imperative to devise management measures for its natural diversity by protecting its habitats from pollution, reclamation and invasive species.

In addition to genetic drift and geographical isolation, positive selection in response to environmental effects would also contribute to the observed mitochondrial divergence between *E. suratensis* populations (Morales *et al*. 2015; Harrisson *et al*. 2016) as evident in the present study with many positively selected sites in OXPHOS genes.

Selection analysis indicated signals of positive or diversifying selection in many genes of Oxidative Phosphorylation Complex (Complex I: NADH dehydrogenase (ND1, ND4, ND5, and ND6), Complex III: Cytochrome b (CYT B), and Complex IV: Cytochrome C Oxidase (C01)) coinciding with the habitat characteristics as the populations of the present study represented humid tropical climatic zones constituting rainforests in the southwest (Calicut and Kochi (type_Af)), semi-arid zones (Tamil Nadu (Mandapam) (type_BSh)) in the southeast and humid subtropical zones (Odisha (Chilka) (type_Cwa)) in the northeast regions of India (Peel *et al*. 2007). The unique amino acid and nucleotide substitutions corresponding to each geographic location indicate the presence of positive selection in response to environmental effects (Morales *et al.* 2015; Harrisson *et al*. 2016). Based on this, we classified the lineages as lineages specific to semi-arid regions (lineage 1; Tamil Nadu), humid tropical regions (lineage 2; Odisha) and humid subtropical regions (lineages 3 and 4).

The sampling site Kozhikode (Korapuzha) is located in the northern region of brackish water systems bordering the state of Kerala (average pH =7, average temp = 29$^0$C, Salinity = 0.1-0.2ppt). Kochi belong to the Vembanad Lake is characterized by low salinity and pH due to the high monsoon rainfall in Kerala (average pH =7.2, average temp = 29$^0$C, Salinity = 0.3-20ppt) (Thasneem *et al*. 2018). The coastal ecosystems of Tamil Nadu (Mandapam) are characterized by comparatively high-water temperature and salinity throughout the year (average pH = 8, average temp = 31$^{0}$C, Salinity = 20-30ppt) (Sridhar *et al*. 2006; Srinivasan *et al*. 2018). The brackish water systems and rivers in Mandapam are connected to the Palk Bay in the Bay of Bengal. Chilka Lake (Odisha) situated along the east coast of India is highly alkaline (average pH = 8.5, average temp = 27$^{0}$C, Salinity = 20-30ppt) and is connected to the Bay of Bengal (Sagarika *et al*. 2010). The brackish water systems of Kerala, Tamil Nadu (Mandapam) and Odisha (Chilka lake) exhibit wide seasonal variations in important water quality parameters like temperature, salinity, pH and dissolved oxygen which may act as selective forces on the genome of organisms inhabiting these water bodies.

The differences in climatic conditions and consequent environmental gradients across the range of distribution of a species may demand specific metabolic/bioenergetic adaptations generating genotypic and phenotypic variants (Schoville *et al*. 2012; Franks and Hoffmann 2012). Hence the positively selected sites and consequent amino acid substitutions may be signals of adaptation to suit the specific metabolic requirements in their habitat (da Fonseca *et al*. 2008; Morales *et al*. 2017). The conserved nature of most of the key amino acid residues participating in electron transfer pathway, putative water exit pathway, ion/chemical binding and putative proton exit pathway in the OXPHOS complex indicates that these regions are constrained functionally (under strong purifying selection) (Ballard and Whitlock 2004). However, evidence for positive selection acting on mitochondrial OXPHOS genes is accumulating. Signals of positive selection associated with thermal adaptation, size, diet, salinity, latitude, migratory behavior, and swimming speed have been reported (Ballard *et al*. 2007) in many fishes (Garvin *et al*. 2015a; Yu *et al*. 2011; Teacher *et al*. 2012). In Drosophila, a particular mtDNA (Cytochrome c oxidase) haplotype was reported to be more tolerant of cold which can colonize temperate regions (Ballard *et al*. 2007). Large scale human mitochondrial genome analysis also reported that some haplotypes with different coupling efficiency are good candidates for adaptation to various habitats (Mishmar *et al*. 2003; Zhu *et al*. 2016)

The NADH dehydrogenase complex (complex I) is the first and largest multimeric enzyme of the five complexes constituting the oxidative phosphorylation pathway (Sazanov 2015). It provides electrons for the reduction of quinine to quinol generated out of oxidation of the NADH and translocates four protons ($H^{+}$) across the inner membrane. The subunits ND2, ND4, and ND5 directly act as proton pumps for $H^{+}$ ions and the changes in amino acids may have some adaptive value. Among the eighteen positively selected sites in *E. suratensis*, ten were located in transmembrane helices especially in proton-conducting membrane transporter (Proton_antipo_M) (site- #358-ND4, #378-ND4, #302-ND5, #340-ND5, #347-ND5, #356-ND5, #368-ND5, #399-ND5) and sites involved in proton translocation(site #143-ND1 in lineage 1, 2, 3) associated with Complex I (Zhu *et al*. 2016). The lineage 4 restricted substitutions (#347-ND5, Leu-Met, on Proton_antipo_M) and lineage 1 restricted substitution (#340-ND5, #399-ND5) at key residues may have some direct influence on proton translocation. Even though we are unable to identify the exact functional and evolutionary significance of the observed amino acid changes, the replacements can result in regional changes in hydrophobicity, structure within the protein and the coupling efficiency.

Many studies have reported that candidate sites for positive selection are disproportionately concentrated in the complex I in many fishes (Garvin *et al*. 2015a,b; Caballero *et al*. 2015; Consuegra *et al*. 2015; Jacobsen *et al*. 2016; Teacher *et al*. 2012) which may be related to protein function (da Fonseca *et al*. 2008; Morales *et al*. 2017) as OXPHOS complex I produce ~ 40% of the proton-pumping required for ATP synthesis. Polymorphism in this region is also reported in other groups like Hares (Melo-Ferreira *et al*. 2014), Mammals (da Fonseca *et al*. 2008), birds (Morales *et al*. 2015), Tachycineta (Stager *et al*. 2014) and Monkeys (Yu *et al*. 2011).

Cytochrome b is part of the respiratory protein complex III, which is the middle component of the mitochondrial respiratory chain, coupling the transfer of electrons from ubi hydroquinone to cytochrome c with the generation of an electrochemical gradient across the mitochondrial membrane. Substitutions at amino acids participating in the interaction (site- #18, #85, #91, #106, #107, #109, #232, #239, #254, #258, #259, #263, #265 in CYTB) and polypeptide binding (site- #22, #226, #63, #74, #78 in CYTB) in the inter-chain domain interface can alter the coupling efficiency of complex III, thus influencing the functional structure of cytochrome b (Iwata *et al*. 1998). Polymorphisms in regions reported to participate in polypeptide binding at mitochondrial and the nuclear-encoded subunits interface (site- #22, #226, #63, #74, #78 in CYTB) (Gershoni *et al*. 2014) may be due to co-evolution between mitochondrial and nuclear-encoded subunits in *E. suratensis*. Such co-evolution has been reported in cytochrome c oxidase (complex IV) of primates (Osheroff *et al*. 1983) and NADH dehydrogenase complex of humans (Gershoni *et al*. 2014). Positively selected sites that appear to interact with other COX subunits were also reported from high-performance fishes belonging to *Scombroidei* (Dalziel *et al*. 2006). When mito-nuclear interactions are disrupted, it may result in reproductive isolation and speciation (Burton *et al*. 2013).

Substitutions in the Heme bL binding site (#83, #84, #51 in CYTB) (Iwata *et al*. 1998) may also have some beneficial and adaptive function in the metabolic performance of *E. suratensis* in their local habitats. In humans, mutations characterized by enhanced binding of water at Qi site have been linked to increased longevity (Beckstead *et al*. 2009) whereas in yeast, mutation at Qo binding site linked to reduced catalytic efficiency and increased oxygen radical production (Wenz *et al*. 2007).

Cytochrome c oxidase (complex IV) catalyzes the final step in the mitochondrial electron transfer chain and is considered as one of the major regulation sites for OXPHOS (Li *et al*. 2006). It receives an electron from each of the four cytochrome c molecules which transfers electrons between complex III and IV and transfers them to one oxygen molecule. During this process, it converts one molecular oxygen to two molecules of water by using four protons from the inner aqueous phase to make water and also, translocates four protons across the membrane. The conserved nature of most of the key amino acid residues reported to participate in the electron transfer pathway, putative water exit pathway, ion/chemical binding and putative proton exit pathway in complex IV indicates that these regions are constrained functionally. Mutations observed outside the key functional residues could be related to relaxed purifying selection (Jacobsen *et al*. 2016).

High genetic differentiation between populations observed in the present study indicates a lack of connectivity driven by habitat fragmentation, the influence of historic geographic events and selective forces of the environment to which they are adapted. In addition, reduction in effective population size due to anthropogenic activities, overexploitation, habitat alterations and competition from invasive species contribute to a reduction in intraspecific genetic diversity. The signals of adaptive mitogenome evolution/habitat specific substitutions indicate the influence of habitat on the dynamics of metabolic gene functions in the OXPHOS. Insights from the present investigation are very important for further experiments on genetic improvement of stocks of *E. suratensis* by correlating habitat characteristics with economically important traits like growth and reproduction. In spite of the historic expansion, a recent decline in effective population size was detected in Bayesian skyline plots which call for augmenting conservation efforts for this species which is very important in aquaculture. The reasons for contemporary low effective population size may be due to the ecological deterioration of Green Chromide habitats due to sea-level fluctuations, land reclamation and pollution (Padmakumar *et al*. 2012; Krishnakumar *et al*. 2009) in addition to overfishing, bottleneck and founder effects (Jayaram 2010; Siddall *et al*. 2003).. Insights from the present investigation are very important for further experiments and planning on genetic conservation of natural diversity natural habitats of *E. suratensis.* Augmenting aquaculture of this species important in aquaculture and ornamental fish industry could also be considered as an option for conservation in addition to providing income for farmers.

## Supplementary Tables and Figures

**Table 8.S1** List of Primer pairs used for amplification and sequencing of *E. suratensis* mitogenome

| Primer Name | | Sequence (5' - 3') |
|---|---|---|
| cichmit 1 | Forward primer | CCTGGCATAAGTTAATGGTG |
| | Reverse primer | AGACAGTTAAGCCCTCGTTA |
| cichmit 1IP | Forward primer | CGCCCTGATATGCTCAACAGC |
| | Reverse primer | CGGTAGGTCTGTCACCTCTAC |
| cichmit 1IP2 | Forward primer | CTGAAACTGGCCCTGAAGCGC |
| | Reverse primer | CGATGTACAGGTGTGCGTGGAG |
| cichmit 2 | Forward primer | ACGGACCGAGTTACCCTAGG |
| | Reverse primer | CCTGCYTCTACWCCAGAGGA |
| cichmit 2IP | Forward primer | GTGGCAGAGCCCGGCATTGC |
| | Reverse primer | GAGGGAGGAAGGAGTCAGAAGC |
| cichmit 3 | Forward primer | TTGGTGCCCCYGATATRGCC |
| | Reverse primer | AGGGTGCCGGYGYTRTTTTG |
| cichmit 3IP | Forward primer | CCTTGTCAAGGTGGGATCGTGG |
| | Reverse primer | CGTAGGGAGGGGAGGGCGAT |
| cichmit 4 | Forward primer | TRGCCTTYAGYGCAACCGAA |
| | Reverse primer | GGGTTTRAATTGTTTGTTGGTKA |
| cichmit 4IP | Forward primer | CGTTGAACTCACCACAACAAACG |
| | Reverse primer | TGAGGTCCTGTGTGGGAATTATG |
| cichmit 5 | Forward primer | CCCCGTAATATCYATACCCC |
| | Reverse primer | CTATTGTRGCGGCTGCAATR |
| cichmit 5IP | Forward primer | CCCCTACCCCTGAACTAGGAG |
| | Reverse primer | GAGAGGGGGTCTGTGGCTATG |
| cichmit 6 | Forward primer | YATTGCAGCCGCYACAATAG |
| | Reverse primer | AGAACCAGTGACCCTCTGGA |
| cichmit 6IP | Forward primer | CCCTACCCCTGAACTAGGAG |
| | Reverse primer | GAGAGGGGGTCTGTGGCTATG |
| cichmit 6IP2 | Forward primer | CACCCCCAACTGAGCTCTTACC |
| | Reverse primer | TTGGGCGGATTTTCCGGCTGC |

**Table 8.S2** List of mitogenomic region sequenced and included in the partial mitogenome data (11881bp) of *E. suratensis* mitogenome.

| Protein coding genes | Ribosomal RNA | Transfer RNA | Non-coding region |
|---|---|---|---|
| ND1, ND2, COX1(Partial), ND4, ND5, ND6, CYTB | 12S ribosomal RNA, 16S ribosomal RNA | tRNA-Val, tRNA-Phe,tRNA-Leu, tRNA-Ile,tRNA-Gln, tRNA-Met, tRNA-Trp, tRNA-Ala,tRNA-Asn, tRNA-Cys,tRNA-Tyr, tRNA-His, tRNA-Ser, tRNA-Leu,tRNA-Glu, tRNA-Thr,tRNA-Pro | control region |

**Table 8.S3** AMOVA analysis results for Control Region, Concatenated genes and partial mito-genome nucleotide sequences of *E. suratensis.*

| Structure tested | Variance component | % of variation | Φ Statistics ($\Phi_{ST}$) | *p* value |
|---|---|---|---|---|
| Control Region (One gene pool) | | | | |
| Among population | 1.38202 | 41.14 | 0.41 | <0.0001 |
| Within population | 1.97702 | 58.86 | - | - |
| Concatenated genes (One gene pool) | | | | |
| Among group | 6.67391 | 60.07 | 0.60 | <0.0001 |
| Within group | 4.43 | 39.93 | - | |
| Partial genome (One gene pool) | | | | |
| Among group | 6.305 | 40.08 | 0.40 | <0.0001 |
| Within group | 9.426 | 59.92 | - | |

**Table 8.S4** Sites under negative/purifying selection identified in FEL and SLAC.

| Gene | Codon position | From AA to AA | FEL | SLAC |
|---|---|---|---|---|
| ND1 | 139 | Leu-Phe | 0.05 | NS |
| ND1 | 161 | Ile-Met | 0.05 | NS |
| ND1 | 168 | Leu-Phe | 0.05 | 0.05 |
| ND4 | 171 | Leu-Val | 0.05 | 0.05 |
| ND4 | 203 | Ser-Asn | 0.05 | 0.05 |
| ND4 | 307 | Ser-Pro | NS | 0.05 |
| ND5 | 290 | Thr-Ser | 0.05 | 0.05 |
| ND5 | 391 | Ala-Thr | 0.05 | 0.05 |
| CYTB | 93 | Ile-Thr | 0.05 | 0.05 |
| CYTB | 230 | Leu-Val | 0.05 | 0.05 |
| ND6 | 121 | Val-Ala | 0.05 | 0.05 |

**Table 8.S5** Summary genetic statistics for restriction-site associated DNA (RAD) sites of *E. suratensis*.

|  | All positions (variant and fixed) |  |  |  |  |  |  |  |  |  | # Variant positions |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pop ID | Sites | Variant Sites | Polymorphic Sites | Private | % Poly | N | P | Obs Het | π | $F_{IS}$ | N | P | Obs He | π | $F_{IS}$ |
| CHILKA | 13305577 | 4840 | 2079 | 2977 | 0.016 | 1.735 | 0.9999 | 0.0001 | 0.0001 | 0 | 1.604 | 0.82 | 0.3214 | 0.301 | -0.031 |
| KOCHI | 15060470 | 5189 | 2052 | 95 | 0.014 | 1.343 | 0.9999 | 0.0001 | 0.0001 | 0 | 1.296 | 0.811 | 0.3722 | 0.354 | -0.027 |
| MANDAPAM | 14143882 | 4518 | 1764 | 0 | 0.013 | 1 | 0.9999 | 0.0001 | 0.0001 | 0 | 1 | 0.805 | 0.3904 | 0.39 | 0.0 |

Average number of individuals genotyped at each locus (N), the number of variable sites unique to each population (Private), the number of nucleotide sites across the data set (Sites), polymorphic sites across the data set (Polymorphic Sites) percentage of polymorphic loci (% poly), the average frequency of the major allele (P), the average observed heterozygosity per locus (Obs He), the average nucleotide diversity (π), and the average Wright's inbreeding coefficient ($F_{IS}$)

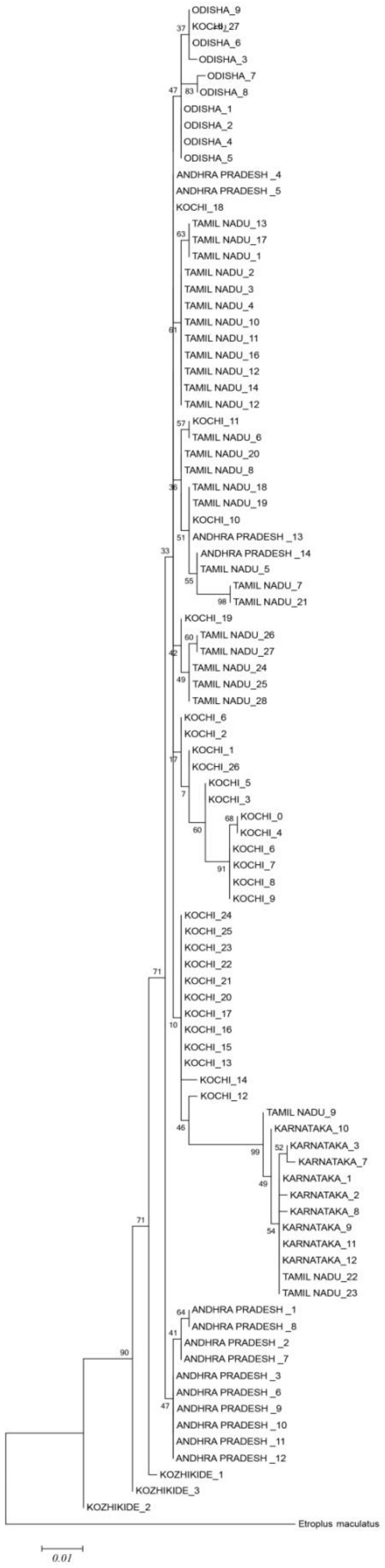


**Fig. 8.S1** Maximum likelihood tree generated by alignment of *E. suratensis* control region sequence.

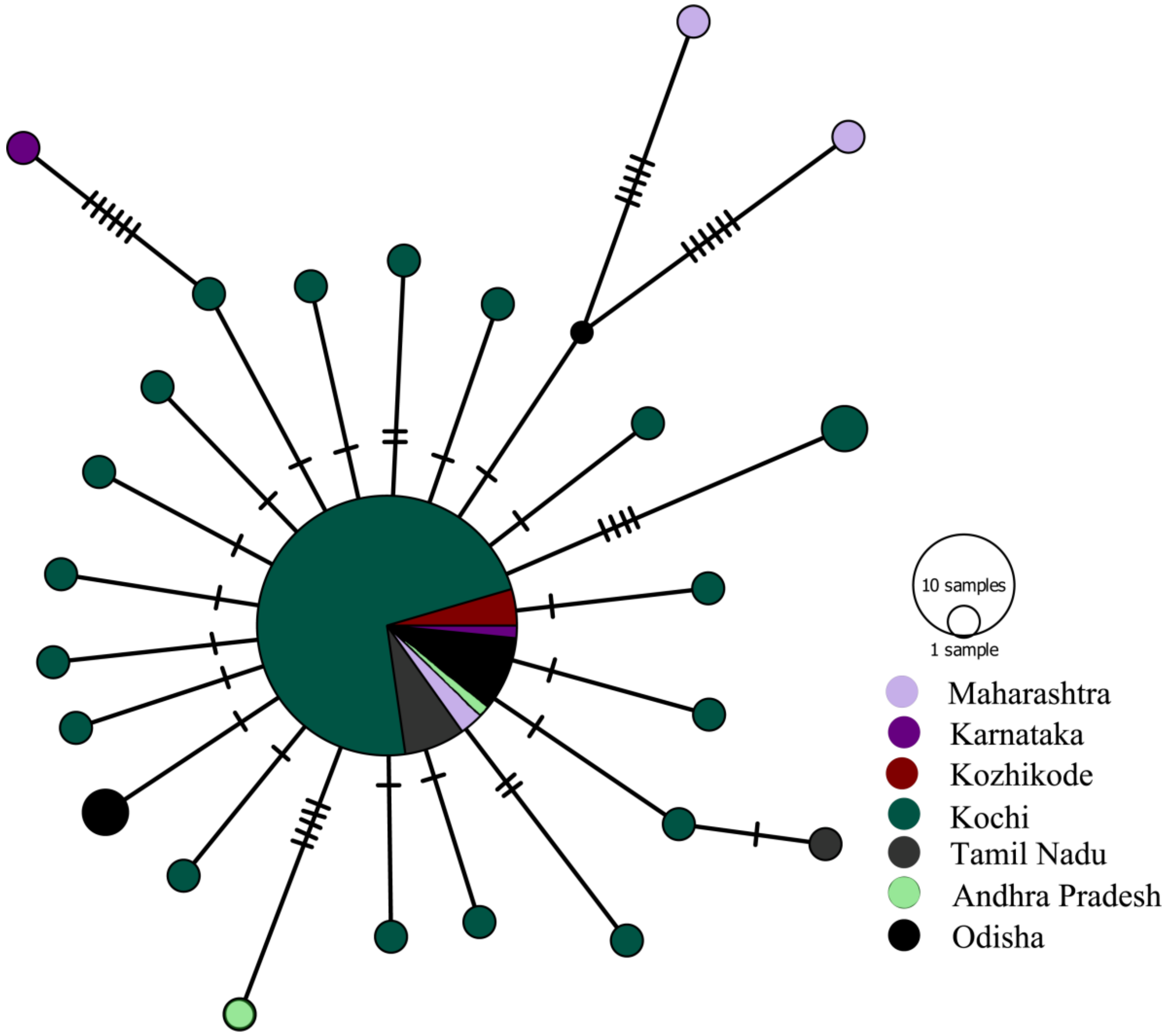


**Fig. 8.S2** Haplotype network diagram constructed with mitochondrial CO1 of *E.suratensis* using a median joining method. Haplotypes are represented in circles and colors indicate geographical locations. Mutational steps are indicated as vertical stripes.

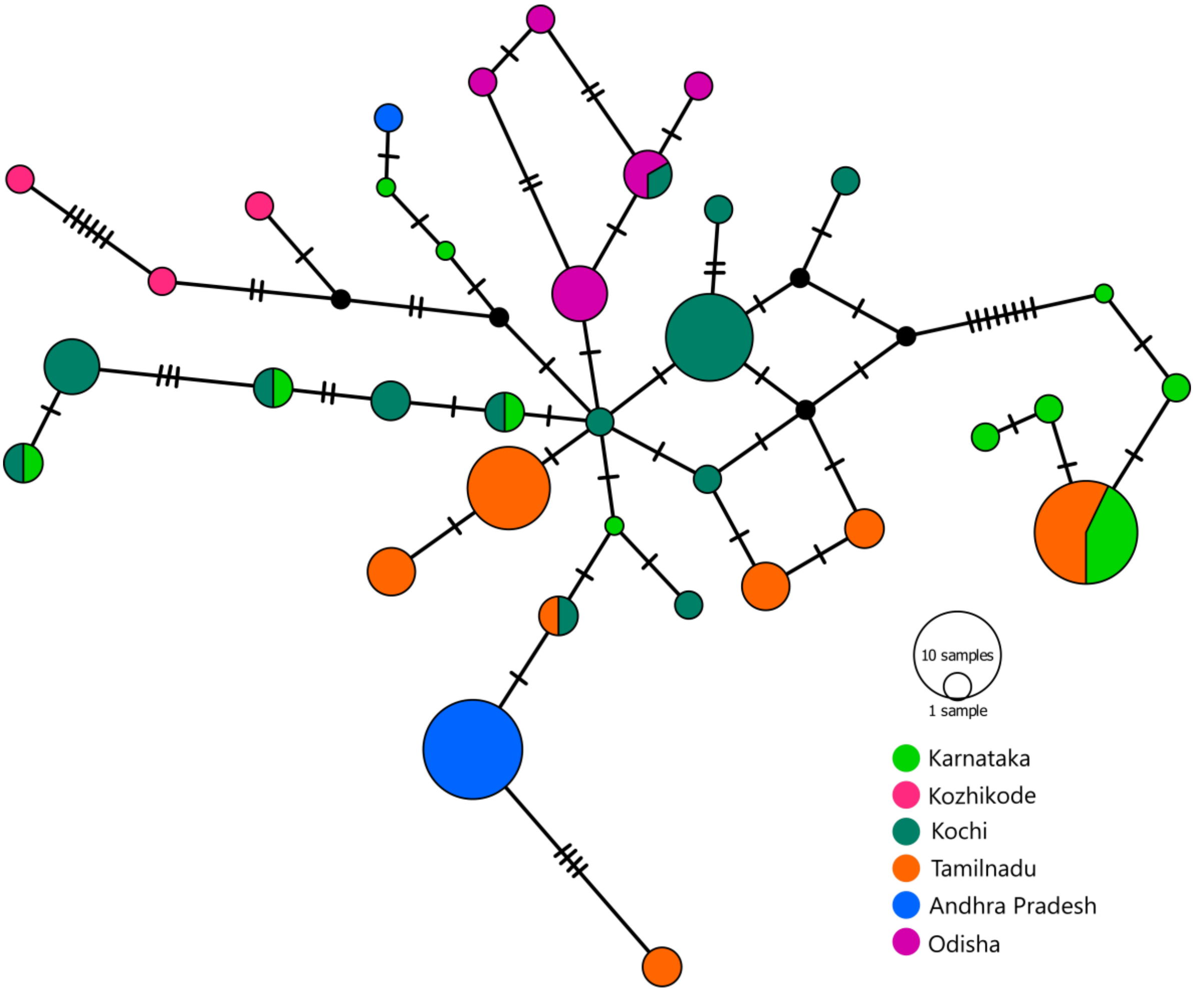


**Fig. 8.S3** Haplotype network constructed with Control region sequences of *E. suratensis* using median-joining method. Haplotypes are represented in circles and colors indicate geographical locations. Mutational steps are indicated as vertical stripes.

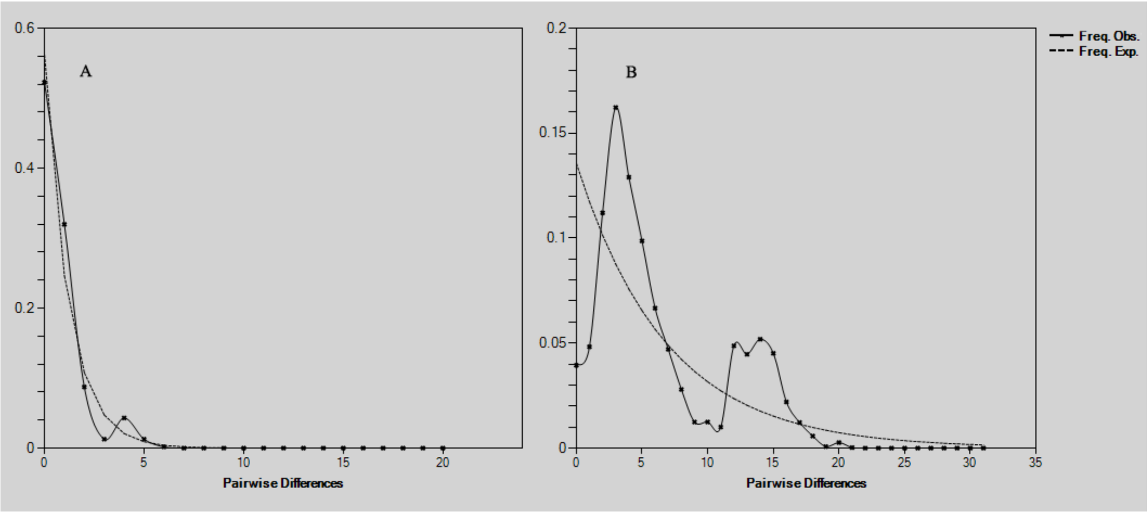


**Fig. 8.S4** Mismatch distribution analysis plots. A) Based on concatenated genes of all samples and B) based on control region of all samples.

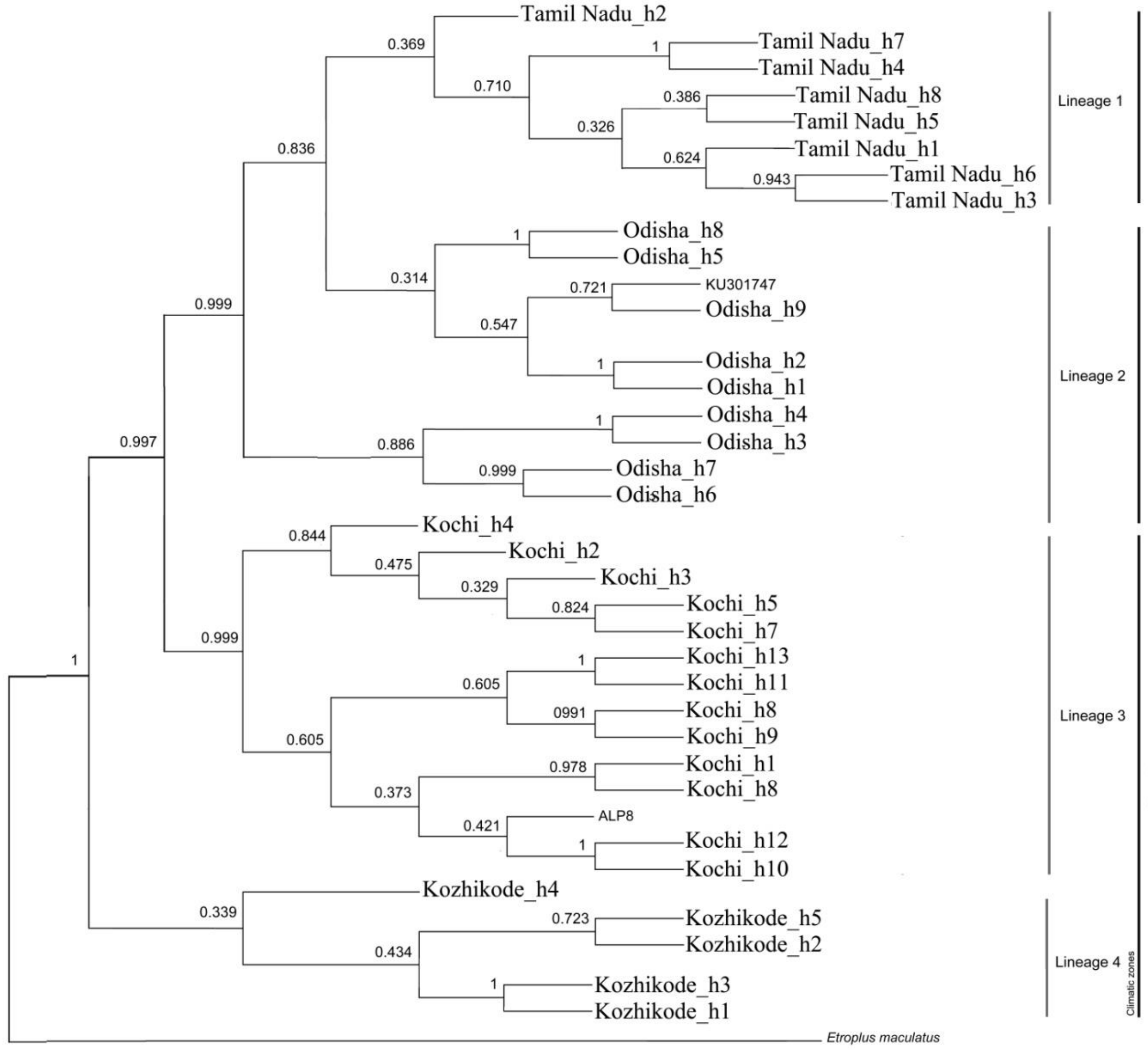


**Fig. 8.S5** The Bayesian tree for 36 haplotypes of *E. suratensis* concatenated mitochondrial protein-coding gene data (7200bp) of *E. suratensis. E. maculatus* (Gen Bank accession number NC_009587) was used as an outgroup to root the tree. Posterior probability values for node support are shown. Refer to Fig 8.1 for Site Name and Sample ID.

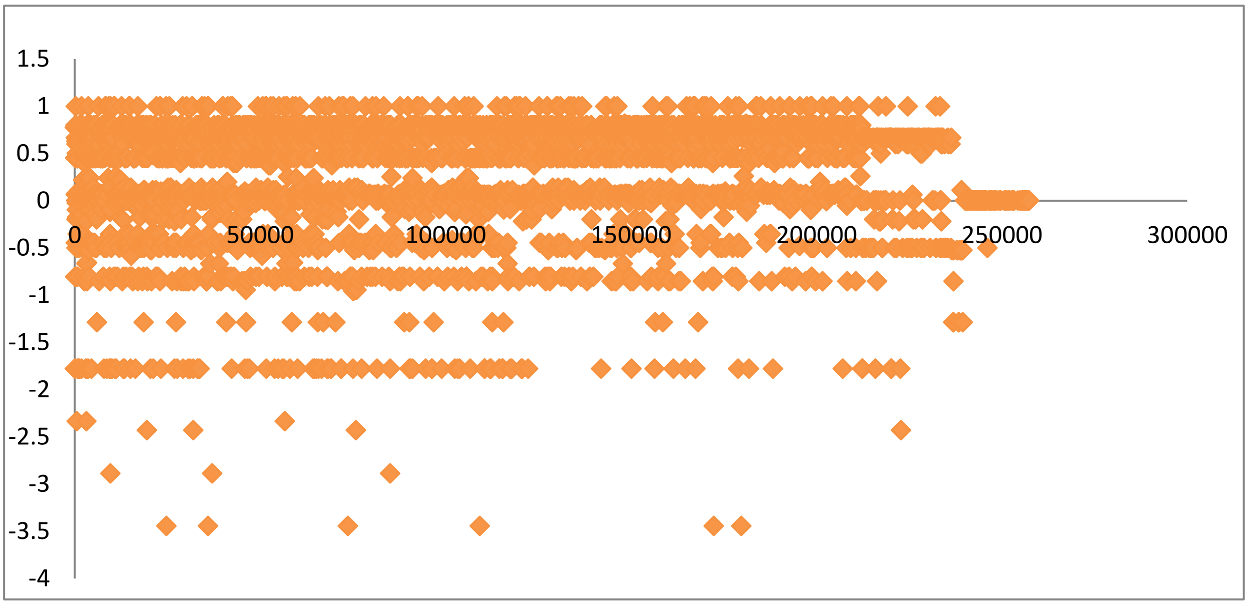


**Fig. 8.S6** Plot of pairwise $F_{ST}$ of 3921 loci between *E. suratensis* population. X-axis represents number of ID for each locus and Y-axis indicates the pairwise $F_{ST}$ values. Large numbers of fixed differences are observed (SNPs with an $F_{ST}$ of 1.0) in comparison between three populations.

## 5. References

1. Abraham R (2011) *Etroplus suratensis* The IUCN Red List of Threatened Species. The IUCN Red List of Threatened Species 2011:eT172368A6877592
2. Alex MD, Kumar AB, Kumar US, George S (2016) Analysis of genetic variation in Green Chromide [*Etroplus suratensis* (Bloch)] (Pisces: Cichlidae) using microsatellites and mitochondrial DNA. *IndianJ Biotechnol* 15(1):375-381
3. Azuma Y, Kumazawa Y, Miya M, Mabuchi K, Nishida M (2008) Mitogenomic evaluation of the historical biogeography of Cichlids toward reliable dating of teleostean divergences. *BMC Evol Biol* 8(1):215
4. Ballard JWO, Melvin RG, Katewa SD, Maas K (2007) Mitochondrial DNA variation is associated with measurable differences in life-history traits and mitochondrial metabolism in *Drosophila simulans*. *Evolution* 61(17):1735-1747
5. Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Mol Ecol* 13(4):729-744
6. Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intras pecific phylogenies. *Mol Biol Evol* 16(1):37-48
7. Barlow G (2000) The Cichlid Fishes: Nature's Grand Experiment in Evolution. Perseus Publishing, Cambridge, Massachusetts, USA
8. Beckstead WA, Ebbert MT, Rowe MJ, McClellan DA (2009) Evolutionary pressure on mitochondrial cytochrome b is consistent with a role of CytbI7T affecting longevity during caloric restriction. *Plos One* 4:e5836.
9. Bindu L, Padmakumar KG (2012) Breeding behaviour and embryonic development in the Orange chromide, *Etroplus maculatus* (Cichlidae, Bloch 1795). *J Mar Biol Assoc India* 54(1):13-19
10. Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, Snelgrove PV, Morris CJ, Gregory RS, Hardie DC, Hutchings JA (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *P Roy Soc Lond B Bio* 277(1701):3725-3734
11. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, Turner-Maier J (2014) The genomic substrate for adaptive radiation in African Cichlid fish. *Nature* 513(7518):375-381
12. Burton RS, Pereira RJ, Barreto FS (2013) Cytonuclear genomic interactions and hybrid breakdown. *Annu Re Ecol Evol* S. 44, 281-302
13. Caballero S, Duchene S, Garavito MF, Slikas B, Baker CS (2015) Initial evidence for adaptive selection on the NADH subunit two of freshwater dolphins by analyses of mitochondrial genomes. *Plos One* 10:e0123543
14. Cadrin SX, Kerr LA, Mariani S (2013) Stock identification methods: applications in fishery science. 2nd edn. Academic Press, UK.
15. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22(11):3124-3140
16. Chandrasekar S, Nich T, Tripathi G, Sahu NP, Pal AK, Dasgupta S (2014) Acclimation of brackish water pearl spot (*Etroplus suratensis*) to various salinities: relative changes in abundance of branchial Na+/K+ - ATPase and Na+/K+/2Cl− co-transporter in relation to osmoregulatory parameters. *Fish Physiol Biochem* 40(3):983-996
17. Chandrasekar S, Sivakumar R, Subburaj J, Thangaraj M (2016) Geographical structuring of Indian pearl spot, *Etroplus suratensis* (Bloch, 1790) based on partial segment of the CO1 gene. *Curr Res Microbiol Biotechnol* 45:1536-1539
18. Cheviron ZA, Connaty AD, McClelland GB, Storz JF (2014) Functional genomics of adaptation to hypoxic cold-stress in high-altitude deer mice: transcriptomic plasticity and thermogenic performance. *Evolution* 68(1):48-62
19. Consuegra S, John E, Verspoor E, De Leaniz CG (2015) Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genet Sel Evol* 47:1-10.
20. Crofts AR (2004). The cytochrome bc 1 complex: function in the context of structure. *Annu Rev Physiol* 66:689-733
21. Crook DA, Lowe WH, Allendorf FW, Eros T, Finn DS, Gillanders BM, Hadwen WL, Harrod C, Hermoso V, Jennings S, Kilada RW (2015) Human effects on ecological connectivity in aquatic ecosystems: integrating scientific approaches to support management and mitigation. *Sci Total Environ* 534:52-64
22. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A (2008) The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* 9(1):119

23. Dalziel AC, Moyes CD, Fredriksson E, Lougheed SC (2006) Molecular evolution of cytochrome c oxidase in high-performance fish Teleostei: Scombroidei). *J Mol Evol* 62:319-331
24. De Silva SS, Maitipe P, Cumaranatunge RT (1984) Aspects of the biology of the euryhaline Asian Cichlid, Etroplus suratensis. *Environ Biol Fish* 10(1-2):77-87
25. Dhanya AM, Remya M, Biju KA (2013) Morphometric and genetic variations of *Etroplus suratensis* (Bloch) (Actinopterygii: Perciformes: Cichlidae) from two tropical lacustrine ecosystems, Kerala, India. *J Aquat Biol Fisheries* 1(1-2):140-150
26. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 17. *Mol Biol Evol* 29(8):1969-1973
27. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
28. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564-567
29. Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, Gilbert MTP, da Fonseca RR (2011) Positive selection on the killer whale mitogenome. *Biol Letters* 7(1):116-118
30. Franks SJ, Hoffmann AA (2012) Genetics of climate change adaptation. *Genetics* 46:185-208
31. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147(2):915-925
32. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133(3):693-709
33. Fungtammasan A, Ananda G, Hile SE, Su MSW, Sun C, Harris R, Medvedev P, Eckert K, Makova KD (2015) Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res* 25(5):736-749
34. Garvin MR, Bielawski JP, Sazanov LA, Gharrett AJ (2015a) Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. *J Zool Syst Evol Res* 53(1):1-17
35. Garvin MR, Thorgaard GH, Narum SR (2015b) Differential expression of genes that control respiration contribute to thermal adaptation in redband trout (*Oncorhynchus mykiss gairdneri*). *Genome Biol Evol* 7(6):1404-1414
36. Genner MJ, Turner GF (2005) The mbuna Cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish Fish* 6(1):1-34
37. Gershoni M, Levin L, Ovadia O, Toiw Y, Shani N, Dadon S, Tsur A (2014) Disrupting mitochondrial-nuclear coevolution affects OXPHOS complex I integrity and impacts human health. *Genome Biol Evol* 6:2665-2680
38. Gunawickrama KS (2012) Morphological heterogeneity and population differentiation in the green chromid *Etroplus suratensis* (Pisces: Cichlidae) in Sri Lanka. *Ruhuna J Sci* 2(1):70-81
39. Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 66:591-600
40. Harrisson K, Pavlova A, Gan HM, Lee YP, Austin CM, Sunnucks P (2016) Pleistocene divergence across a mountain range and the influence of selection on mitogenome evolution in threatened Australian freshwater cod species. *Heredity* 116(6):506-515
41. Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish Fish* 9(4):333-362
42. Ho SY, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigos AG, Cooper A (2011) Time-dependent rates of molecular evolution. *Mol Ecol* 20:3087-3101
43. Husemann M, Ray JW, King RS, Hooser EA, Danley PD (2012) Comparative biogeography reveals differences in population genetic structure of five species of stream fishes. *Biol J Linn Soc* 107(4):867-885
44. Iwata S, Lee JW, Okada K, Lee JK, Iwata M, Rasmussen B, Link TA, Ramaswamy S, Jap BK (1998) Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. *Science* 281(5273):64-71
45. Jacobsen MW, Da Fonseca RR, Bernatchez L, Hansen MM (2016) Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the NADH2 gene in whitefish (Coregonus ssp.). *Mol Phyl Evol* 95:161-170
46. Jacobsen MW, Hansen MM, Orlando L, Bekkevold D, Bernatchez L, Willerslev E, Gilbert MTP (2012) Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (Coregonus spp). *Mol Ecol* 21(11):2727-2742
47. Jayakumar M (2002) Wetland conservation and management in Kerala. State Committee on Science Technology and Environment, Thiruvananthapuram, Kerala, India
48. Jayaprakas V, Nair NB, Padmanabhan KG (1990) Sex ratio, fecundity and length-weight relationship of the Indian pearlspot, *Etroplus suratensis* (Bloch). *J Aquacult Trop* 5(2):141-148

49. Jayaram KC (2010) The Freshwater Fishes of the Indian Region. Narendra Publishing House, Delhi, India
50. Kocher TD (2004) Adaptive evolution and explosive speciation: the Cichlid fish model. *Nat Rev Genet* 5(4):288-298
51. Krishnakumar K, Raghavan R, Prasad G, Bijukumar A, Sekharan M, Pereira B, Ali A (2009) When pets become pests-exotic aquarium fishes and biological invasions in Kerala, India. *Curr Sci India* 97(4):474-476
52. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 70 for bigger datasets. *Mol Bio Evol* 33(7):1870-1874
53. Kurup BM, Thomas KV (2001) Fishery resources of the Ashtamudi estuary. In: Kerry B, Joseph M, Baba M, Kurian N (eds) Developing a Management Plan for Ashtamudi Estuary, Kollam, India, ASR Ltd. Marine and Freshwater Consultants Hamilton, New Zealand and Centre for Earth Science Studies, Thiruvananthapuram, India, pp. 513-546
54. Li Y, Park JS, Deng JH, Bai Y (2006) Cytochrome c oxidase subunit IV is essential for assembly and respiratory function of the enzyme complex. *J Bioenerg Biomembr* 38(5-6):283-291
55. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451-1452
56. McMillan WO, Palumbi SR (1997) Rapid rate of control region evolution in Pacific butterfly fishes (Chaetodontidae). *J Mol Evol* 45:473-484.
57. Melo-Ferreira J, Vilela J, Fonseca MM, Da Fonseca RR, Boursot P, Alves PC (2014) The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome Biol Evol* 6:886-896
58. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI (2003) Natural selection shaped regional mtDNA variation in humans. *P Natl Acad Sci USA* 100(1):171-176
59. Morales HE, Pavlova A, Joseph L, Sunnucks P (2015) Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Mol Ecol* 24(11):2820-2837
60. Morales HE, Sunnucks P, Joseph L, Pavlova A (2017) Perpendicular axes of incipient speciation generated by mitochondrial introgression. *Mol Ecol* 26:3241-3255
61. Murrell B, Joel OW, Sasha M, Thomas W, Konrad S, Pond SLK (2012) Detecting individual sites subject to episodic diversifying selection. *Plos Genetics* 8(7):e1002764
62. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Bio Evol* 30(5)*:*1196-1205
63. Nei M (1987) Molecular evolutionary genetics. Columbia university press, New York, USA
64. Ng TH, Tan HH (2010) The introduction, origin and life-history attributes of the non-native Cichlid *Etroplus suratensis* in the coastal waters of Singapore. *J Fish Biol* 76(9):2238-2260
65. Osheroff N, Speck SH, Margoliash E, Veerman EC, Wilms J, Konig BW, Muijsers AO (1983) The reaction of primate cytochromes c with cytochrome c oxidase Analysis of the polarographic assay. *J Biol Chem* 258:5731-5738.
66. Padmakumar KG, Bindu L, Manu PS (2012) Etroplus suratensis (Bloch), the State Fish of Kerala. *J Biosci* 37(1):925-931
67. Pavlova A, Gan HM, Lee YP, Austin CM, Gilligan DM, Lintermans M, Sunnucks P (2017) Purifying selection and genetic drift shaped Pleistocene evolution of the mitochondrial genome in an endangered Australian freshwater fish. *Heredity* 118(5):466-476
68. Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Koppen-Geiger climate classification. *Hydrol Earth Syst Sc* 4(2):439-473
69. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One* 7(5):e37135
70. Pond SLK, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21(10):2531-2533
71. Rambaut A, Drummond A (2008) "FigTree: Tree figure drawing tool, version 1.2. 2." Institute of Evolutionary Biology, University of Edinburgh. accessed on 13 March 2019
72. Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Bio Evol* 19(12):2092-2100
73. Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *MolBiol Evol* 9(3):552-569
74. Sagarika N, Gayatri N, Gouri CN, Rajani K S (2010) Physicochemical parameters of chilika lake water after opening a new mouth to bay of bangal, orissa, India. *Continental J Env Sci* 4:57- 65

75. Saravanan KR, Sivakumar K, Choudhury BC (2013) Important Coastal and Marine Biodiversity Areas of India. In: Sivakumar K (ed.) Coastal and Marine Biodiversity Areas of India: challenges and way forward, ENVIS Bulletin: Wildlife & protected areas Vol 15, Wildlife institute of India, Dehradun-248001, India, pp 134-188
76. Sazanov, L.A., 2015. A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat Rev Mol Cell Biol* 16:375
77. Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* 152:1079-1089
78. Schoville SD, Bonin A, Francois O, Lobreaux S, Melodelima C, Manel S (2012) Adaptive genetic variation on the landscape: methods and cases. *Annu Rev Ecol Evol* 43:23-43
79. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31:3381-3385.
80. Scott GR, Schulte PM, Egginton S, Scott AL, Richards JG, Milsom WK (2010) Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol Biol Evol* 28(1):351-363
81. Sebastian W, Sukumaran S, Gopalakrishnan A (2019) Complete mitochondrial genome and phylogeny of the green chromide *Etroplus suratensis* (Bloch, 1790) from Vembanad Lake, Kerala, south India. *Indian J Fish* 66:125-130.
82. Seehausen O (2006) African Cichlid fish: a model system in adaptive radiation research. *P Roy Soc Lond B Bio* 273(1597):1987-1998
83. Siddall M, Rohling EJ, Almogi-Labin A, Hemleben C, Meischner D, Schmelzer I, Smeed DA (2003) Sea-level fluctuations during the last glacial cycle. *Nature* 423(6942):853-858
84. Siddall M, Rohling EJ, Almogi-Labin A, Hemleben C, Meischner D, Schmelzer I, Smeed DA (2003) Indian continental margin: An update. *Geol Soc India* 46:157-162.
85. Sridhar R, Thangaradjou T, Senthil Kumar S, Kannan L (2006) Water quality and phytoplankton characteristics in the Palk Bay, southeast coast of India. *J Environ Biol* 27(3):561-566
86. Srinivasan S, Balasubramanian K, Rajaram K, Mukunda K K, Basanta K J (2018) Diurnal Variation and Water Quality Parameters of Three Different Ecosystems in Gulf of Mannar, Southeast Coast of India. *J Marine Sci Res Dev* 8:1-6
87. Stager M, Cerasale DJ, Dor R, Winkler DW, Cheviron ZA (2014) Signatures of natural selection in the mitochondrial genomes of Tachycineta swallows and their implications for latitudinal patterns of the pace of life. *Gene* 546:104-111
88. Stiassny ML (2001) The Cichlid Fishes: Nature's Grand Experiment In Evolution. *Copeia* 3:878-879.
89. Stier A, Bize P, Roussel D, Schull Q, Massemin S, Criscuolo F (2014) Mitochondrial uncoupling as a regulator of life history trajectories in birds: An experimental study in the zebra finch. *J Exp Biol* 217(19):3579-3589
90. Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics* 123(3):597-601
91. Takeda M, Kusumi J, Mizoiri S, Aibara M, Mzighani SI, Sato T, Terai Y, Okada N, Tachida H (2013) Genetic structure of pelagic and littoral Cichlid fishes from Lake Victoria. *Plos One* 8(9):e74088
92. Teacher AG, Andre C, Merila J, Wheat CW (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evol Biol* 12(1):248
93. Thasneem TA, Bijoy NS, Geetha. PN (2018) Water quality status of Cochin estuary, India. *Indian J Geo-Mar Sci* 4:57- 65
94. Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K *et al*. (1995) Structures of Metal Sites of Oxidized Bovine Heart Cytochrome c Oxidase at 2.8 \AA. *Science* 269(5227):1069-1074
95. Wenz, T., Covian, R., Hellwig, P., MacMillan, F., Meunier, B., Trumpower, B.L., Hunte, C., 2007. Mutational analysis of cytochrome b at the ubiquinol oxidation site of yeast complex III. J. Biol. Chem. 282:3977-3988
96. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19(5):671-672
97. Yu L, Wang X, Ting N, Zhang Y (2011) Mitogenomic analysis of Chinese snub-nosed monkeys: Evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. *Mitochondrion*. 11:497-503
98. Zhu J, Vinothkumar KR, Hirst J (2016) Structure of mammalian respiratory complex I. *Nature* 536(7616):354-358

# Chapter 9

## ISOLATION AND CHARACTERIZATION OF STRESS RESPONSE GENES FROM *ETROPLUS SURATENSIS* (Bloch, 1790)

### ABSTRACT

The present study reports the complete sequences of Aquaporin 1 (AQP1) gene and partial sequences of genes, Sodium/Potassium-Transporting ATPase subunit alpha-1 (Na/K-ATPase α1 subunit), Osmotic Stress Transcription Factor 1 (OSTF1), Transcription Factor II B (TFIIB), Heat Shock Cognate 71 (HSC71) and Heat Shock Protein 90 (HSP90) obtained from mRNA and genomic DNA of *Etroplus suratensis* (Bloch, 1790). They are candidate genes involved in stress responses of fishes. AQP1 gene was 2163 bp long. Its mRNA sequence has 55 bp 5' UTR, 783 bp open reading frame (ORF), 119 bp 3' UTR, three intronic regions and 90% identity with AQP1 of *Oreochromis niloticus.* The partial Na/K-ATPase α1subunit gene obtained 5998 bp length with an ORF of 2213 bp and 12 intronic regions. The partial OSTF1, TF IIB, HSC71 and HSP90 mRNA sequences obtained were 1473 bp, 587 bp, 1708 bp and 151 bp in length respectively. All the genes showed high sequence similarity with respective genes reported from fishes. Comparison of AQP1 and Na/K-ATPase α1 genomic DNA sequence of *E. suratensis* collected from different water system showed two type of AQP1 with one synonymous mutation in exon-1 and higher sequence difference in intronic regions (including addition, deletion, transition and transversion mutations) with few synonymous and non-synonymous mutations in the exons of Na/K-ATPase α1. The sequence information of these major candidate genes involved in stress responses will help in further studies on population genetics, adaptive variations and genetic improvement programs of this cichlid species having aquaculture, ornamental and evolutionary importance.

## 1. INTRODUCTION

The Pearl spot (*Etroplus suratensis* Bloch, 1790) is brackish water, euryhaline fish belonging to the family Cichlidae, widely distributed in peninsular India and Sri Lanka inhabiting both fresh and brackish water systems (Jayaram 1999, Padmakumar *et al.* 2012). The species exhibits high levels of adaptive capacity as it can withstand a wide range of salinity and temperature conditions with highly efficient osmoregulation and cellular stress response mechanisms (Padmakumar *et al.* 2012, Chandrasekar *et al.* 2014). It is a popular species for aquaculture, but natural populations are getting depleted due to overexploitation and habitat destruction (Padmakumar *et al.* 2012). The life cycle of this species is completed either in fresh or brackish waters and breeding has been reported throughout the year with a peak from June to September and February-April (Jayakumar 2002).

Most of the fishes undergo feeding or spawning migrations, thus encountering many stressful conditions. Fishes combat these challenges by way of altering their physiological or behavioural patterns, otherwise termed as plasticity. The driver for plasticity lies in the genome as in most of the cases these plastic responses are associated with a change in the expression or mutation in a gene, set of genes or regulatory region (Larsen *et al.* 2007). Selection and adaptive evolution in the functional gene regions or regulatory elements form the key force providing optimum fitness to the organism (Tine *et al.* 2010, Nielsen *et al.* 2009, Tamura *et al.* 2013). *E. suratensis* is a species capable of tolerating a wide range of temperature and salinity conditions and consequently studying the key genes involved in environmental tolerance may be the first step to derive clues regarding their plasticity (Padmakumar *et al.* 2012, Chandrasekar *et al.* 2014). Heat shock proteins and stress-induced transcriptional factors are considered as candidate genes for studying different physiological stresses like temperature and salinity variations. Sodium/Potassium-transporting ATPase, aquaporins and osmotic stress transcriptional factors are important candidate genes for studying osmotic stresses and hence in this study we characterized Aquaporin 1, Na+/K+-ATPase α1, HSP90, HSC71B, OSTF1 and TFIIB genes from *E. suratensis*. The information generated will be useful for further studies on population genomics, adaptive variation and selective breeding of this important aquaculture species.

## 2. Materials And Methods

*E. suratensis* samples (30 numbers, 8-32 cm) were collected from Vembanad estuary, Kerala, India and total RNA extracted from liver and gill tissues using Trizol reagent and quantified using Nanodrop Spectrophotometer (Thermo Fisher Scientific). Total RNA was reverse transcribed to cDNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher). DNA was isolated using a standard phenol-chloroform method. Oligonucleotide primers were designed for each gene based on the information regarding corresponding gene sequences from NCBI GenBank (Table 9.S1).

### 2.1. Genome walking method

Genomic DNA sequence of the target gene was obtained by a genome walking method. Genome walking library from genomic DNA (including EcoRV, DraI, PvuII and StuI restriction enzyme digested) was constructed as described in Genome Walker Universal Kit (Clontech). Each library (4 libraries) was screened by performing primary PCR with P1L & P2L (Table 9.S1) primer of each gene and adapter primer 1 (ADP1, Table 9.S1). The PCR products were subsequently used as a template for the secondary PCR with the respective P2L & P2R and adapter primer 2 (ADP2, Table 9.S1). PCR products were screened on a 1.2% agarose gel and DNA bands in the gel were sliced and purified using MinElute Gel Extraction Kit (Qiagen). Purified PCR products were cloned into PJET cloning vector with CloneJET PCR Cloning Kit (Thermo Fisher Scientific) according to the manufacturer’s instruction. Plasmids were transferred into TOP10 chemically competent *Escherichia coli* and plated on ampicillin LB agar plate. Colonies containing inserts were screened by colony PCR and positive samples were cultured overnight. Plasmids were purified using Gene Jet Plasmid Mini-Prep Kit (Thermo Fisher Scientific) and sequenced both directions using Big Dye Terminator Sequencing Ready Reaction v 3.0 Kit on an ABI 3730 automated sequencer (Applied Biosystems).

Primer for amplification of the first-strand cDNA of each gene was designed from the corresponding sequence generated by genome walking method. PCR was carried out for each gene using corresponding sense and antisense primers to obtain open reading frame/cDNA of the gene (Table 9.S1).

### 2.2. RACE PCR method

In order to get more information about 5’ and 3’ ends of the genes a 5’ RACE and 3’ RACE was performed respectively using the SMARTer® RACE 5’/3’ Kit (Clontech) Primer for amplification of the first-strand cDNA of each gene was designed from corresponding sequence generated by genome walking method. PCR was carried out for each gene using corresponding sense and antisense primers (Table 9.S1) to obtain open reading frame/cDNA of the gene. The 25 µl volume reaction mixture contained 50ng DNA, 1X PCR reaction buffer (10 Mm Tris-HCl, 500Mm KCl, 1.5Mm MgCl2) (Sigma Aldrich), 10mM of each dNTPs, 10mM each primer and 1U Taq DNA Polymerase (Sigma Aldrich). The PCR program consisted of an initial denaturation at $95^0$c for 5min, followed by 30 cycles of $94^0$c for 30 seconds, $60^0$c for 30 seconds, $72^0$c 60 seconds and final extension step of $72^0$c for 10 minutes. The PCR products were eluted after electrophoresis, cloned and then sequenced as described above.

In order to get more information about 5’ and 3’ ends of the genes a 5’ RACE and 3’ RACE was performed respectively using the SMARTer$^{®}$ RACE 5’/3’ Kit (Clontech), according to the manufactures instruction. A universal primer mix with corresponding antisense GSP1L primer and antisense nested GSP2L primer of each gene were used for 5’ RACE. Universal primer mix with corresponding sense primer GSP1R and GSP2R of each gene was used for 3’ RACE. A set of specific primer (Table 9.S1) was designed for Aquaporin 1 and Na/K-ATPase α1 based on their sequence obtained from *E. suratensis.*

A set of specific primers (Table 9.S1) were designed for Aquaporin 1 and Na/K-ATPase α1 based on their sequence obtained from *E. suratensis.* These primers were then used to amplify respective genes from genomic DNA of *E. suratensis* collected from Vembanad Lake, Cochin estuary, Korapuzha-Kozhikode, Mandapam-Tamil Nadu. The PCR products were eluted after electrophoresis, cloned and then sequenced.

### 2.3. Sequence assembly and analysis

Sequence assembly, translations and alignments were prepared using MEGA-6 (Tamura *et al.* 2013). BLAST sequence similarity search tool in the NCBI web site was used for gene identification and sequence similarity analysis. SMART Scan-Prosite program (http://us.expasy.org/tools/scanprosite/) was used to predict the characteristic conserved motifs in the genes. To know the evolutionary relationship between Aquaporin 1, Na/K-ATPase α1, HSP90, HSC71B, OSTF 1 and TFIIB of *E. suratensis* with other cichliformes and bony fishes (available in NCBI, GenBank), phylogenetic trees were constructed using Neighbour-joining method in MEGA-6. Multiple sequence alignment for each gene was prepared with corresponding gene (mRNA) sequences retrieved from NCBI GenBank and the tree topological stability was evaluated by 1000 bootstrap re-sampling. The tertiary structure of the protein was obtained by SWISS-MODEL automated protein modelling server (Schwede *et al.* 2003).

## 3. Results

### 3.1. Aquaporin 1 (AQP1)

The assembled aquaporin gene sequence obtained from *E. suratensis* was 2163 bp long. Its mRNA sequence has 55 bp 5' UTR, 783 bp open reading frame (ORF), 119 bp 3' UTR and three intronic regions. The encoded protein sequence was 261 amino acids long and calculated molecular weight of 27.53 kDa. The sequence has been submitted to GenBank (accession no: MH289467). The channel-forming conserved signature sequence motif NPA (Aspargine-Proline-Alanine) was located between amino acid positions 70-72 and 184-185 (Chrispeels and Agre 1994). Action site of mercurial compounds, which inhibit water permeability, was located before the second NPA motif at site 181 (Cysteine). *E. suratensis* AQP1 nucleotide sequence showed 90%, 84% and 78% identity to *Oreochromis niloticus, Taki fugu obscures* and *Paramormyrops kingsleyae* respectively. Deduced amino acids of the CDS were used to predict the 3D structure of *E. suratensis* AQP1 gene using AQP1 of *Homo sapiens* (PDB ID: 1IH5) as a template with the identity of 60.33% (Fig. 9.1). Multiple sequence alignment and phylogenetic relationship between *E. suratensis* AQP1 nucleotide sequence with other species are shown in Fig. 9.1, Fig. 9.S3 and Fig. S4. Comparison of 1554bp AQP1 genomic DNA sequence of fish collected from Vembanad lake, Cochin estuary, Korapuzha-Kozhikode and Mandapam-Tamil Nadu showed two types of AQP1 with one synonymous mutation in exon one (AAC/AAT, Asn - 120).

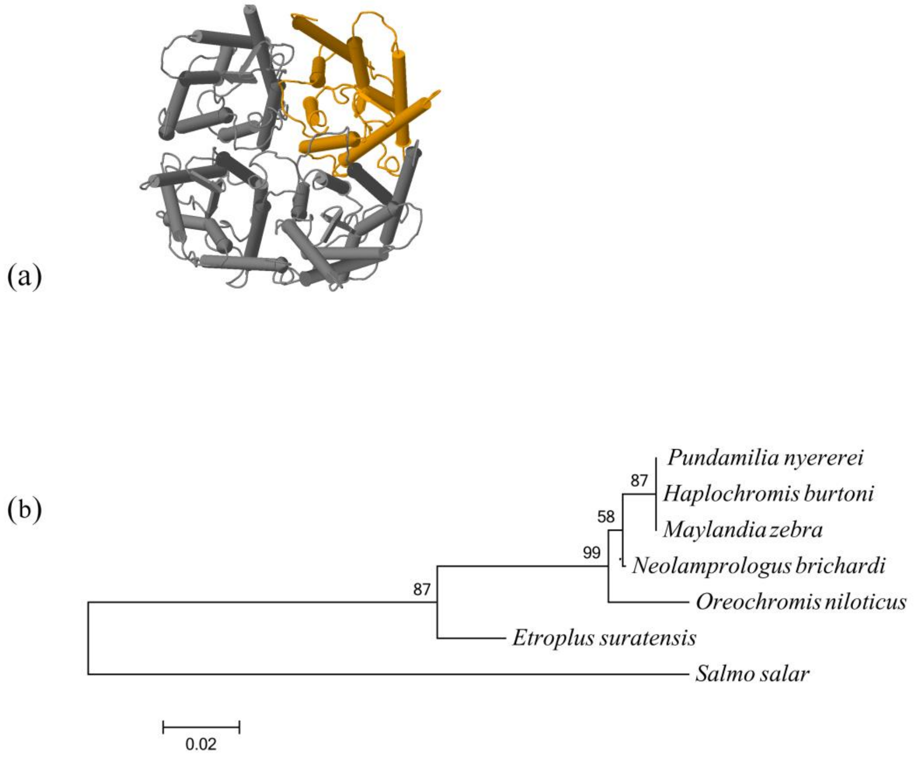


**Fig. 9.1** (a) 3D model of Aquaporin 1 (AQP1) as predicted by Swiss-Model, in the tetramer single unit is represented as yellow. (b) Phylogenetic tree based on AQP1 nucleotide sequences of Cichliformes (available in NCBI, GenBank) by Neighbour Joining method. *Salmo salar* was used as an out group. GenBank accession no: NM 001140000.1 *Salmo salar*, XM 003438085.5 *Oreochromis niloticus*, XM 004542847.2 *Maylandia zebra*, XM 005742590.1 *Pundamilia nyererei*, XM 005912557.2 *Haplochromis burtoni* and XM 006792098.1 *Neolamprologus brichardi*

## 3.2. Sodium/Potassium-Transporting ATPase subunit alpha-1 (Na/K-ATPase α1subunit)

The partial Na/K-ATPase α1subunit gene has 5998 bp length with an ORF of 2213 bp and 12 intronic regions. The partial mRNA sequence encoded 737 amino acids of the respective proteins (GenBank accession no: MH289468). In the amino acids sequence of *E. suratensis* Na/K-ATPase α1subunit gene, a characteristic conserved motif sequence DKTGTLT which is a signature of P-type ATPases was observed which was located between amino acid sites 164-170 (Moller *et al.* 1996). 3D structure of *E. suratensis* Na+/K+-ATPase α1subunit protein was generated using *Sus scrofa* Na/K-ATPase α1 protein (PDB ID 3wgu.1.A) as a template with the identity of 88.87% (Fig. 9.2). Multiple sequence alignment and phylogenetic relationship between *E. suratensis* Na/K-ATPase α1 subunit nucleotide sequence with other species based on the Neighbour-joining method are shown in Fig. 9.2, Fig. 9.S5 and Fig. 9.S6. Comparison of 1798 bp Na+/K+-ATPase α1 subunit sequences of *E. suratensis* collected from different regions indicated higher sequence difference in intronic regions (including addition, deletion, transition and transversion mutations) with few synonymous and non-synonymous mutations in the exons.

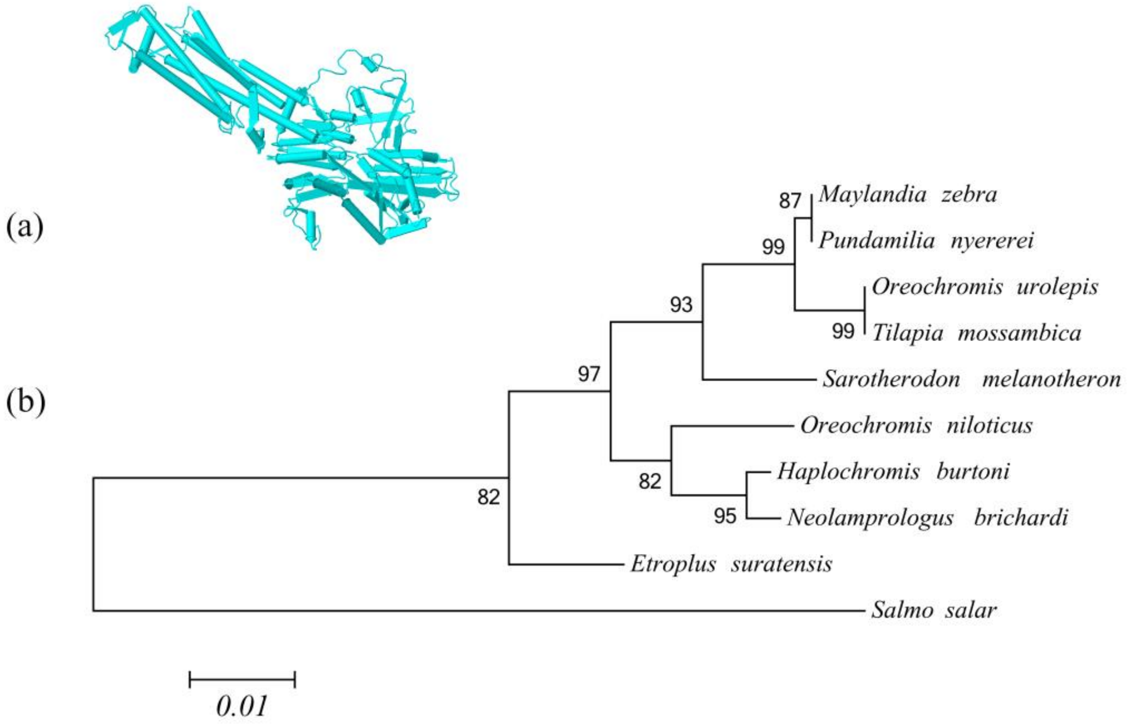


**Fig. 9.2** (a) 3D model of Sodiumpotassium-Transporting ATPase subunit alpha1 (Na/K-ATPase α1) as predicted by Swiss-Model. Phylogenetic tree based on Na/K-ATPase α1nucleotide sequences of Cichliformes (available in NCBI, GenBank) by Neighbour Joining method. *Salmo salar* was used as an out group. GenBank accession no: XM014338426.1 *Haplochromis burtoni*, XM004571251.2 *Maylandia zebra*, XM006792814.1 *Neolamprologus brichardi*, XM005452356.4 *Oreochromis niloticus*, KC702516.1 *Oreochromis urolepis*, XM005749450.2 *Pundamilia nyererei*, GU252208.1 *Sarotherodon melanotheron*, U82549.2 *Tilapia mossambica* and KJ175156.1 *Salmo salar*

The partial OSTF1 mRNA sequence obtained was 1473 bp in length with 378 bp protein-coding region and 1045 bp 3' UTR region (GenBank accession no: MH289469).The partial amino acid sequence of OSTF1 has TSC-22/dip/bun family signature motif MDLVKNHLMYAVREEVE (Ohta *et al*. 1996 ) between amino acid sites36-52. The nucleotide sequence of *E. suratensis* showed 92%, 88% and 92% similarity with OSTF1 of *Oreochromis mossambicus*, *Acanthopagrus schlogelii* and *Amphiprion melanopus* respectively. Phylogenetic relationship between *E. suratensis* OSTF1 nucleotide sequence with other species based on Neighbour joining method is showed in Fig. 9.3.

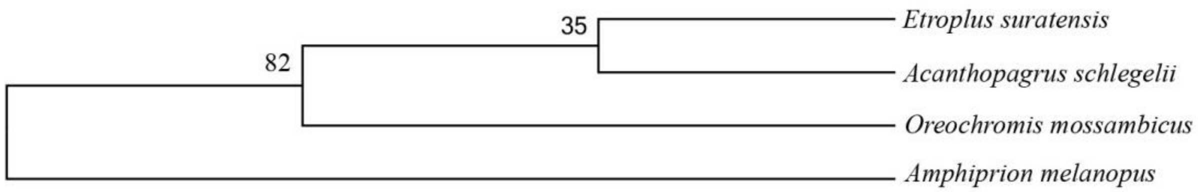


**Fig. 9.3** Phylogenetic tree showing the evolutionary relationship between OSTF1 nucleotide sequences of fishes (available in NCBI, GenBank) and *E. suratensis* by Neighbour Joining method. GenBank accession no: HM037051.1 *Acanthopagrus schlegelii*, JX307115.1 *Amphiprion melanopus*, AY679524.2 *Oreochromis mossambicus*

## 3.4. Transcription Factor IIB (TFIIB)

The TFIIB sequence obtained has 587 bp open reading frame with three intronic regions (GenBank accession no: MH289470). The partial mRNA sequence obtained is encoding 195 amino acid of TFIIB protein with a Transcription factor TFIIB repeat signature motif GRsndAIASACLYIAC (Weinmann 1996) between amino acid site 108 and 123. 3D structure of TFIIB gene was generated using 5iy7.1.M (PDB ID) as a template with the identity of 93.33%. Nucleotide sequence of *E. suratensis* is 94%, 93% and 90% identical to General Transcription Factor IIB of *O. mossambicus, M. zebra and Amphiprion ocellaris* respectively. Multiple sequence alignment and phylogenetic relationship between TFIIB nucleotide sequences are shown in Fig. 9.4, Fig. 9.S1 and Fig. 9.S2.

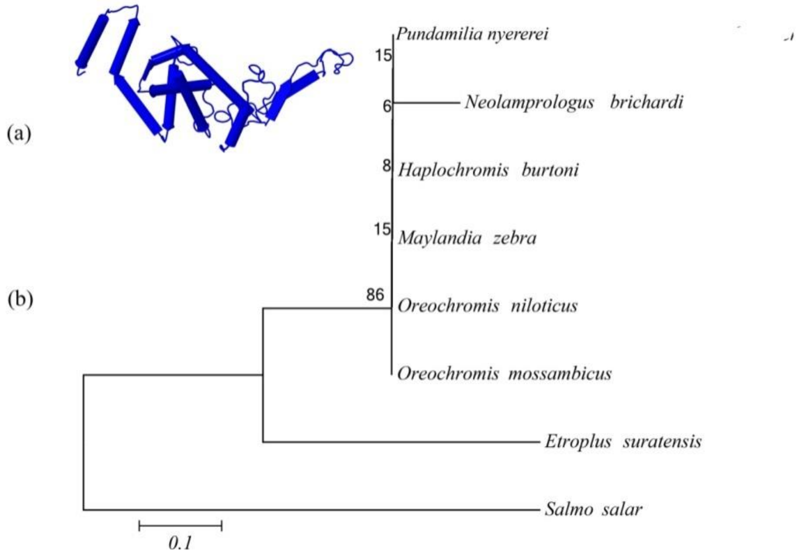


**Fig. 9.4 (**a) 3D model of Transcription factor II B (TF II B) as predicted by Swiss-Model (b) Phylogenetic tree of Cichliformes (available in NCBI, GenBank) and *E. suratensis* based on TF II B nucleotide sequences by Neighbour Joining method. *Salmo salar* was used as an out group. GenBank accession no: BT125312.1 *Salmo salar*, AY679525.1 *Oreochromis mossambicus*, XM_019348123.2 *Oreochromis niloticus*, XM_004570537.4 *Maylandia zebra*, XM_005939374.2 *Haplochromis burtoni*, XM_005720990.2 *Pundamilia nyererei* and XM_006786805.1 *Neolamprologus brichardi*

### 3.5. Heat Shock Cognate 71 (HSC71)

Partial sequence of HSC71 gene obtained was 1708 bp in length with a coding region of 961 bp and 4 intronic regions (GenBank accession no: MH289471). The partial mRNA sequence encodes319 amino acids of the HSC71 protein with two Heat shock Hsp70 proteins family signatures sequences, IDLGTTyS and IFDLGGGTfdvSIL respectively (Lindquist and Craig 1988) at amino acid position 11-18 and 199-212. 3D structure of *E. suratensis* HSC71 gene was obtained using Heat shock cognate protein (PDB ID 1kaz.1.A) of *Bostaurus* as template with the identity of 94.32% (Fig. 9.4). Multiple sequence alignment and Phylogenetic relationship between HSC71 nucleotide sequences of *E. suratensis* with other fishes are shown in Fig. 9.4, Fig. 9.S7 and Fig. 9.S8.

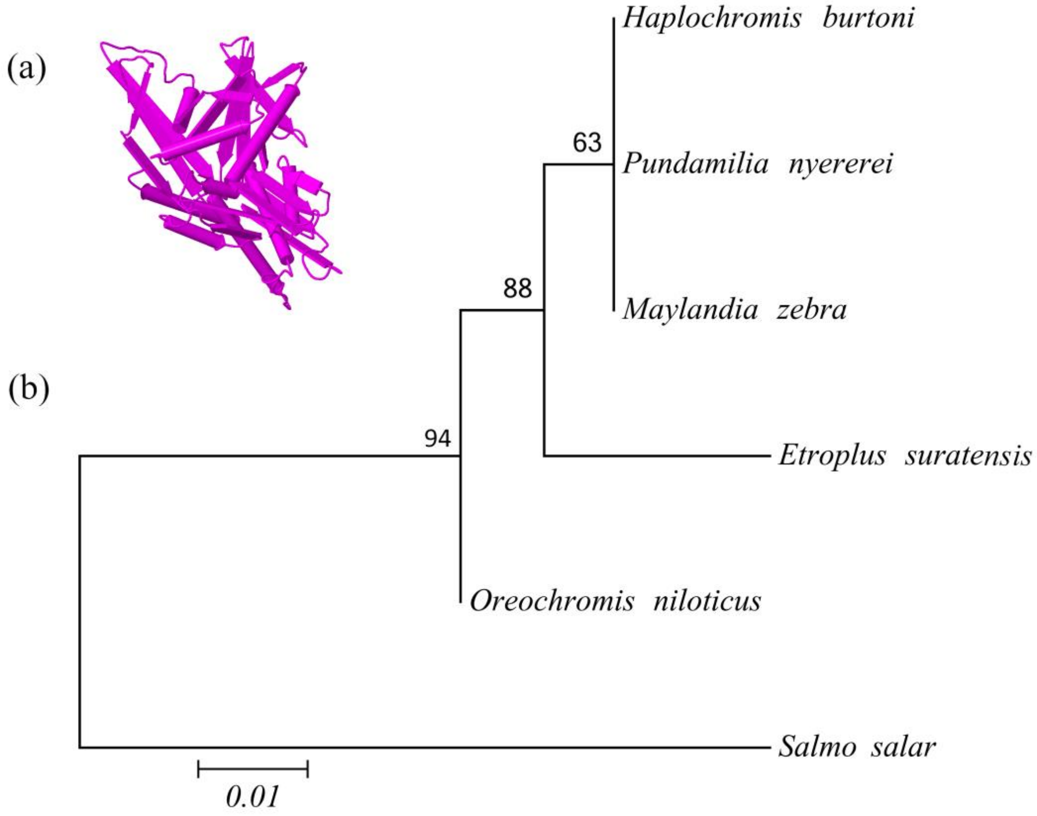


**Fig. 9.5** (a) 3D model of Heat Shock Cognate 71 (HSC71) as predicted by Swiss-Model. (b) Phylogenetic tree based on HSC71 nucleotide sequences of Cichliformes (available in NCBI, GenBank) by Neighbour Joining method. *Salmo salar* was used as an out group. GenBank accession no: XM003455056.5 *Oreochromis niloticus*, XM004558114.4 *Maylandia zebra*, XM005937893.2 *Haplochromis burtoni*, XM005739307.2 *Pundamilia nyererei* and XM014162783.1 *Salmo salar*

### 3.6. Heat Shock Protein 90 (HSP90)

The partial mRNA sequence obtained was 151 bp in length and it encodes 49 nucleotides of the respective protein (GenBank accession no: MH289472). The *E. suratensis* HSP90 gene is closely related to heat shock protein 90 alpha family class B member 1 (Hsp90ab1) of *L. calcarifer* (92%), heat shock protein 90 of *Sinibotia reevesae* (87%) and heat shock protein 90 beta of *Misgurnusan guillicaudatus* (85%). Multiple sequence alignment and phylogenetic relationship between HSP90 nucleotide sequences nucleotide are showed in Fig. 9.6 and Fig. 9.S9.

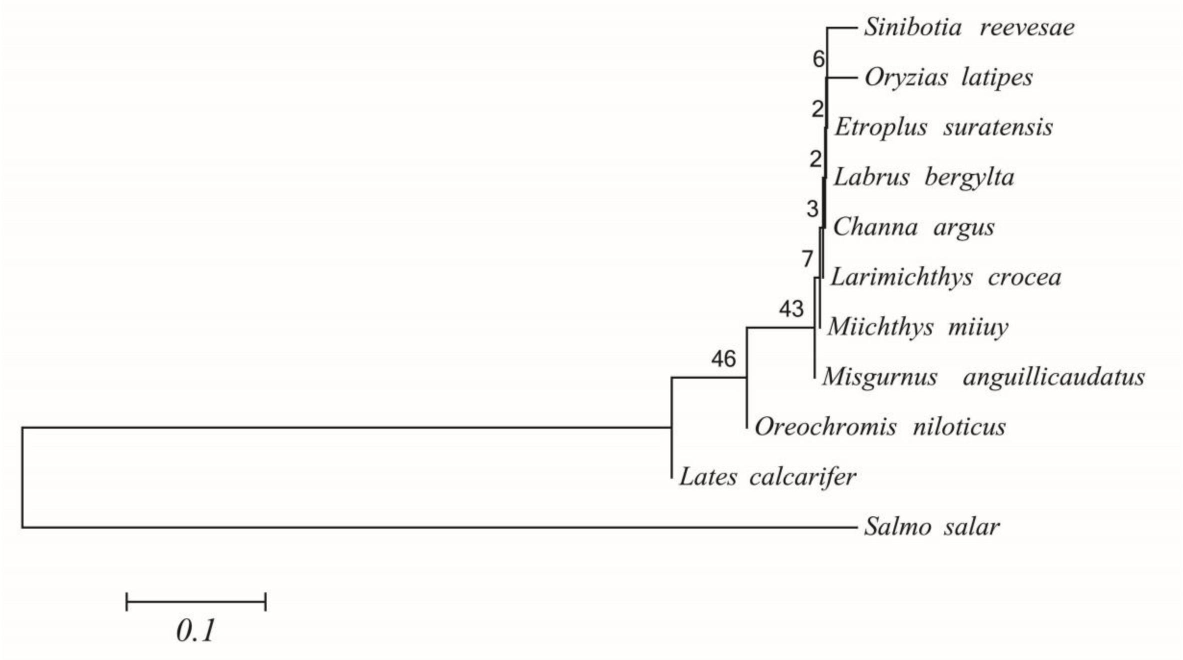


**Fig. 9.6** Phylogenetic tree showing the evolutionary relationship between HSP90 nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratensis* by Neighbour Joining method. *Salmo salar* was used as an out group. GenBank accession no: NM001173702.1 *Salmo salar*, KJ624420.1 *Sinibotia reevesae*, XM020628015.1 *Labrus bergylta*, XM013271984.1 *Oreochromis niloticus*, KU946993.1 *Channa argus*, KT456552.1 *Larimichthys crocea*, JQ929760.1 *Miichthys miiuy*, HQ646106.1 *Lates calcarifer*, CP020802.1 *Oryzias latipes* and KY203858.1 *Misgurnus anguillicaudatus*

## 4. DISCUSSION

Euryhaline fishes like *E. suratensis* have the capacity to acclimatize to a wide range of environmental factors like temperature and salinity (Padmakumar *et al*. 2012, Chandrasekar *et al*. 2014). Gene expression modulations play a major role in this acclimation process involving a network of physiochemical activities known as cellular stress responses which help in regaining cell homeostasis by remodeling of the cell, tissue and organs (Tine *et al*. 2010). Studying the pattern and variation in expression of candidate genes involved in the stress response is important to gain critical knowledge on functional and regulatory mechanisms of local adaptation among natural populations (Charlesworth *et al*. 2017). AQP1, Na/K-ATPase α1subunit, HSC71 and TFIIB genes of *E. suratensis* well diverge from other Cichliform fishes indicating the influence of habitat characteristics. However, *E. suratensis* always clustered in the Cichliformes group when all available bony fish gene sequences were compared.

The aquaporin family proteins are hydrophobic membrane proteins with a major role in water and solute transport which make them ideal candidates for studying osmotic stress (Finn and Cerda 2011). Several aquaporins from animal models including fishes have been studied and its importance in cell homeostasis during osmotic stress conditions like freshwater and seawater acclimatization proved with high levels of expression in gill, kidney, intestine, liver and urinary bladder (Finn and Cerda 2011, Deane *et al*. 2011). AQP1 is a tetramer protein with transmembrane domains which is the most studied aquaporin family in fishes (Finn and Cerda 2011). AQP1 isolated from *E. suratensis* formed a distinct branch with other Cichliformes in the phylogenetic tree with significant bootstrap support. Sodium/Potassium-Transporting ATPase (Na/K ATPase) is another group of potential candidate gene for osmotic stress studies. The fundamental role of Na/K ATPase enzyme is osmoregulation and ion exchange and they are found abundantly in osmoregulatory organs like gill and kidney (Yang *et al*. 2016). It has a heterodimeric structure with a catalytic α subunit and β subunit (Kanai *et al*. 2013). In the osmoregulatory organs of fish, Na/K ATPase actively pump $Na^+$ out and $K^+$ into a cell against concentration and this also provides the driving force for other osmoregulatory activities like transport of other ions (Kanai *et al*. 2013). α and β subunit are encoded by separate genes and they are regulated at transcriptional, translational and assembly level (Kanai *et al*. 2013). Na/K ATPase gene of *E. suratensis* was also well diverged from other Cichliformes. Several studies have examined differential expression of Na/K-

ATPase genes in fishes and observed its elevated expression in hypersaline and freshwater acclimation and consequent adaptation (Deane *et al.* 2011, Tine *et al*. 2010). Isoforms of both α subunit and β subunit have also been reported from fishes (Deane *et al*. 2011).

Stress induced by the alterations in the environment, especially changes in temperature and salinity disrupt cellular homeostasis in fishes and subsequently affect protein synthesis and assembly in cells. Cells overcome this crisis by synthesizing proteins belonging to the heat shock protein family (Hsp) (Tine *et al*. 2010). Among this Hsp70 (60-73kDa) and Hsp90 family (85-90kDa) are the major proteins induced during stress and they function as molecular chaperons by helping refolding, repair and degrading damaged proteins (Basu *et al*. 2002). They are also involved in functions of immune systems, apoptosis and cellular inflammation processes. The Hsp 70 family proteins involved an inducible type Hsp70 and cognate type Heat Shock Cognate 70 (HSC70). HSC 70 is expressed continually, whereas Hsp 70 induced during the stressful condition and it is mediated by binding of heat shock factor 1 (HSF1) on the promoter. HSC proteins are normally present in cytosol, nucleolus and mitochondria where they are involved in various housekeeping homeostasis processes and activities with Hsp proteins (Robert *et al*. 2010). Heat Shock Protein cDNA from *E. suratensis* were identical to heat shock protein 90 alpha family class B member 1 (Hsp90ab1) of *L. calcarifer* (92%), heat shock protein 90 of *S. reevesae* (87%) and HSC71 of *O. niloticus* and *N. brichardi.* Several studies have found evidence of active involvement of Hsp proteins in homeostasis and environmental adaptation in fishes (Tine *et al*. 2010). Intra-specific variation in Hsps and its link to local adaptation, thermal and osmotic tolerance has been demonstrated in fishes (Larsen *et al*. 2007, Tine *et al*. 2010, Nielsen *et al*. 2009, Deane *et al.* 2011) and this gene could be used as a biomarker of stress.

Most of the adaptive/phenotypic plasticity is mediated through transcriptional regulation (Larsen *et al*. 2007) and stress transcriptional factors have major roles in initiating the complex network of cell stress responses. Osmotic stress transcriptional factor 1 (OSTF 1) and transcriptional factor 2 B (TFIIB) are two transcriptional factors induced during hyper osmotic stresses (Fiol *et al*. 2006, Tse 2014). OSTF and TFIIB like sequences from *E. suratensis* were similar to General Transcription Factor IIB of *O. mossambicus* (94%), *M. zebra* (93%) and OSTF1 of *O. mossambicus* (92%), *A. Schlogelii* (88%) respectively.

OSTF 1 was initially identified from *O. mossambicus* and later numerous studies have been carried out to understand its osmoregulatory mechanism (Fiol *et al*. 2006). OSTF 1 is not responding to osmotic stress in zebrafish but its role in embryonic development has been reported. Thus OSTF 1 may not be an osmoregulator in all fishes and it may have some unknown functions (Tse 2014). TFIIB is a general transcriptional factor, binds to beta elements (BRE) in DNA and promotes assembly of poly II complex on promoters (Hampsey 1998). More details about the induction and downstream function of TFIIB induced by stress responses are not well understood. It is suggested that the TFIIB may interact with other transcriptional factors, which are targets of stress genes and facilitate those genes expression (Fiol *et al*. 2006). Thus, SNPs and indels in transcriptional factors like OSTF1 and TFIIB have a potential influence on the rate of expression of genes under their control.

The divergence in sequence characteristics in these genes between other fishes as observed in the phylogenetic tree may be attributed to the environmental conditions of the habitat. The phenotypic plasticity mediated through regulated gene expression can temporarily allow organisms to shift their optimum environmental parameters like temperature tolerance to a new range (Tine *et al*. 2010, Nielsen *et al*. 2009). But evolutionary changes/adaptive modifications in the gene or its regulatory systems can lead to a permanent/hard-wired shift in the optimum tolerance ranges (Tine *et al*. 2010, Nielsen *et al*. 2009). Thus, mutations in transcriptional factors, regulatory regions, 5'/3' UTR may give crucial information about how genes are regulated and expressed within and among species. Understanding the underlying mechanisms of differential acclimatization capabilities of fishes is essential to know susceptibility to environmental alterations and climate change. Detailed sequence data and gene expression analysis based on wide sampling are necessary to achieve this task. The sequence information of major candidate genes involved in stress responses will provide baseline information for further studies on population genetics and adaptive variations in natural populations as well as genetic improvement of cultivated stocks of *E. suratensis*, a species with aquaculture, ornamental and evolutionary importance.

## Supplementary Tables and Figures

**Table 9.S1 PCR primers used in this study.** Oligonucleotide primers were designed for each gene based on the information regarding corresponding gene sequences obtained from NCBI, GenBank. The sequences used for preparing Genome walking primers are; GenBank accession no. 542206014 (*Oreochromis niloticus*, Aquaporin-1), 548404332 (*Pundamilia nyererei*, Heat shock cognate 71), 542219714 (*Oreochromis niloticus*, Heat shock protein HSP 90), 542226640 (*Oreochromis niloticus*, Sodium/potassium-transporting ATPase subunit alpha-1) and 583983155 (*Neolamprologus brichardi*, Transcription initiation factor IIB).

| | | |
|---|---|---|
| Genome walking | | |
| 1 | AQP1 P1L | GTCTCCCAAATTTGATGACTTCCCTGA |
| 2 | AQP1 P2L | TATGCTCCTGTACAAAAGCCCTGGAGT |
| 3 | AQP1 P1R | GCCAGTGACATCACGTCGTCTTTTATC |
| 4 | AQP1 P2R | CATCAAATTTGGGAGACAGCAGGAAAT |
| 5 | HSC.71B P 1L | GTACGAGGGCATCGACTTCTACACCTC |
| 6 | HSC.71B P 2L | CATCGACTTCTACACCTCGATCACCAG |
| 7 | HSC.71B P 1R | CAATCTCCTTCATTTTCAGCAACACCA |
| 8 | HSC.71B P 2R | ATTTTCAGCAACACCATGGAGGAAATC |
| 9 | HSP.90 P 1L | GCTACCCAATCACCCTATTTGTGGAGA |
| 10 | HSP.90 P 2L | GAGGACAAGCCAAAGATAGAGGACGTG |
| 11 | HSP.90 P 2R | TCTCCACAAATAGGGTGATTGGGTAGC |
| 12 | HSP.90 P 1R | ACGTCCTCTATCTTTGGCTTGTCCTCA |
| 13 | NA.K.ATP1 P 1L | GCTCATCAGTATGGCCTACGGACAAAT |
| 14 | NA.K.ATP1 P 2L | GACGTACGAGCGCAAACAAATTGTAGA |
| 15 | NA.K.ATP1 P 1R | GCTTTCTTCAGAGCTGGAGAGTCGTTC |
| 16 | NA.K.ATP1 P 2R | TGTTTGAGCATATCATCCAGCTGCTCT |
| 17 | TF.2B P 1L | CTCAAAGTACCAGAACAGGCGAACCAT |
| 18 | TF.2B P 2L | GGATCGTATCAACTTGCCAAGGAACAT |
| 19 | TF.2B P 1R | CTCCAGTGCCTTCAGTATCAGCTTGAA |
| 20 | TF.2B P 2R | TCTCTTGTCTGCAGGCGATGTAGAGAC |
| 21 | OSTF1 P 1L | GGCCCCGAACAAAAGGCTAAAT |
| 22 | OSTF1 P 2L | ACGCTGCCTTCAAATGCTGACG |
| 23 | OSTF2 P 1R | ACAGGCACTGTTGTCATGCCAT |
| 24 | OSTF2 P 2R | TTGTCRATGGCCACAACGCTAG |
| 25 | ADP1 | GTAATACGACTCACTATAGGGC |
| 26 | ADP2 | ACTATAGGGCACGCGTGGT |
| cDNA amplification | | |
| 27 | TF.2B L | CTCACTTCAGCCTGATTCAAAGAA |
| 28 | TF.2B R | TTCTTTGAATCAGGCTGAAGTGAG |
| 29 | OSTF2 L | TGACTACACCTGCCGTTATCTAAC |
| 30 | OSTF2 R | TTCTTTGAATCAGGCTGAAGTGAG |
| 31 | OSTF1 R | CATATTGGGGCTCTTCTCATTGAG |
| 32 | OSTF1 L | CTCAATGAGAAGAGCCCCAATATG |
| 33 | NA.K.ATP2 L | TATGCTCAAACACCACACTGAAAT |
| 34 | NA.K.ATP2 R | CCTGGCAAATACAATTTCAGTGTG |
| 35 | NA.K.ATP1 L | CATTTCATCCACATCATCACTGGT |
| 36 | NA.K.ATP1 R | CACCAGTGATGATGTGGATGAAAT |
| 37 | HSP.90 L | TTGAGGAGAAGAGGATCAAAGAGA |
| 38 | HSP.90 R | CAGGTACAGGATGATCTTTGTTCC |
| 39 | HSC.71B L | GACTATTTCTGGGCTTAATGTGCT |
| 40 | HSC.71B R | AGCACATTAAGCCCAGAAATAGTC |
| 41 | AQP1 L | CAGTGGTATTATGTATGGAGCACG |
| 42 | AQP1 R | TCGTCTTTTATCAGTGACTGCAAT |
| RACE PCR | | |
| 43 | AQP1 GSP1L | GTCTCCCAAATTTGATGACTTCCCTGA |

| | | |
|---|---|---|
| 44 | AQP1 GSP2L | TATGCTCCTGTACAAAAGCCCTGGAGT |
| 45 | AQP1 GSP1R | GCCAGTGACATCACGTCGTCTTTTATC |
| 46 | AQP1 GSP2R | CATCAAATTTGGGAGACAGCAGGAAAT |
| 47 | HSC.71B GSP 1L | GTACGAGGGCATCGACTTCTACACCTC |
| 48 | HSC.71B GSP 2L | CATCGACTTCTACACCTCGATCACCAG |
| 49 | HSC.71B GSP 1R | CAATCTCCTTCATTTTCAGCAACACCA |
| 50 | HSC.71B GSP 2R | ATTTTCAGCAACACCATGGAGGAAATC |
| 51 | HSP.90 GSP 1L | GCTACCCAATCACCCTATTTGTGGAGA |
| 52 | HSP.90 GSP 2L | GAGGACAAGCCAAAGATAGAGGACGTG |
| 53 | HSP.90 GSP 2R | TCTCCACAAATAGGGTGATTGGGTAGC |
| 54 | HSP.90 GSP 1R | ACGTCCTCTATCTTTGGCTTGTCCTCA |
| 55 | NA.K.ATP1 GSP 1L | GCTCATCAGTATGGCCTACGGACAAAT |
| 56 | NA.K.ATP1 GSP 2L | GACGTACGAGCGCAAACAAATTGTAGA |
| 57 | NA.K.ATP1 GSP 1R | GCTTTCTTCAGAGCTGGAGAGTCGTTC |
| 58 | NA.K.ATP1 GSP 2R | TGTTTGAGCATATCATCCAGCTGCTCT |
| 59 | TF.2B GSP 1L | CTCAAAGTACCAGAACAGGCGAACCAT |
| 60 | TF.2B GSP 2L | GGATCGTATCAACTTGCCAAGGAACAT |
| 61 | TF.2B GSP 1R | CTCCAGTGCCTTCAGTATCAGCTTGAA |
| 62 | TF.2B GSP 2R | TCTCTTGTCTGCAGGCGATGTAGAGAC |
| Population analysis | | |
| 63 | AQP1 pop_gL | TGCCAGATCAGTGTGTTCAAG |
| 64 | AQP1 pop_gR1 | TCATCAAATTTGGGAGACAGC |
| 65 | AQP1 pop_gR2 | TCCACAGTTGTTGCATCGTTA |
| 66 | NA.K.ATP1 pop_gA1L | CATTGCTTTCTTTTCCACCAA |
| 67 | NA.K.ATP1 pop_gA2L | AGGAATCGTCATCAACACTGG |
| 68 | NA.K.ATP1 pop_gAR | TCTCTCATGCCGCTAACAGAT |
| 69 | NA.K.ATP1 pop_gBL | GATCTGTTAGCGGCATGAGAG |
| 70 | NA.K.ATP1 pop_gBR | GGAGGTCCTGGCAAATACAAT |
| 71 | NA.K.ATP1 pop_gCL | GGTGAGCTGAAAGACATGACC |
| 72 | NA.K.ATP1 pop_gCR | GTCCGTAGGCCATACTGATGA |
| 73 | NA.K.ATP1 pop_gDL | GACAAGCTGGTGAATGAGAGG |
| 74 | NA.K.ATP1 pop_gDR | CCATAGCTGTCTTCCAGGTCA |

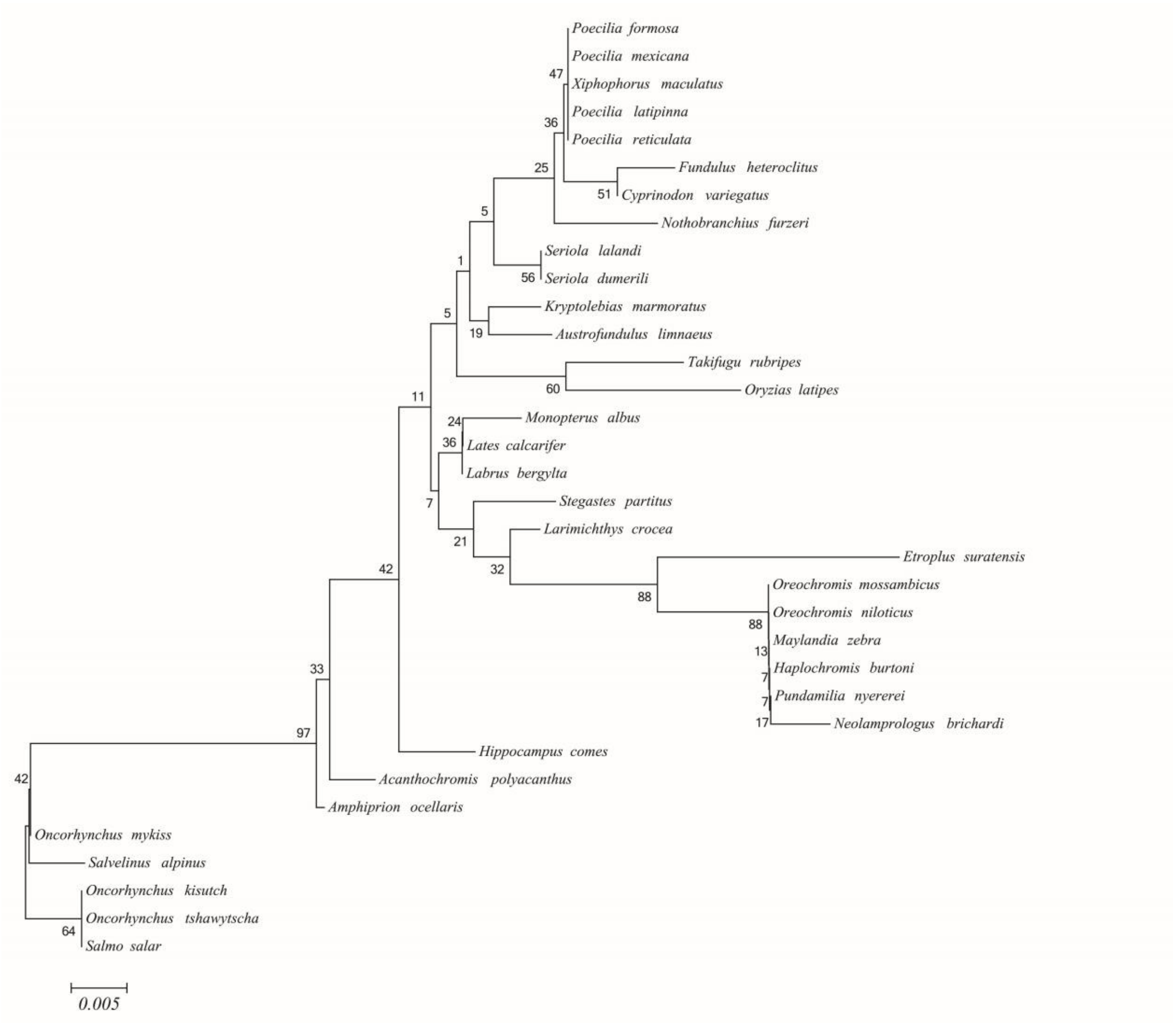
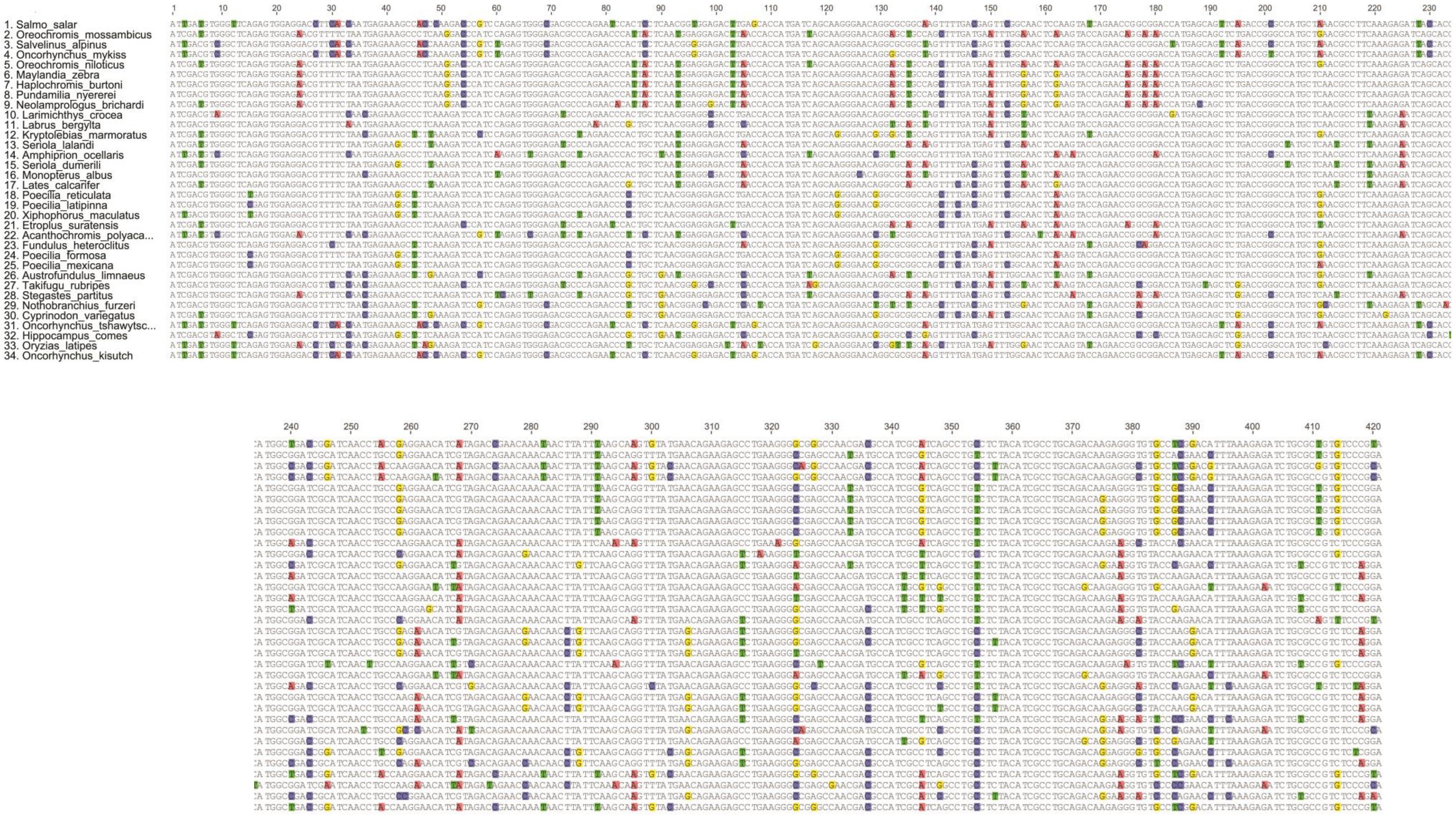


**Fig. 9.S1** Phylogenetic tree based on TF II B nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratensis* by Neighbour Joining method. *Salmo salar* was used as an out group. GenBank accession no: BT125312.1 *Salmo salar*, AY679525.1 *Oreochromis mossambicus*, XM 019348123.2 *Oreochromis niloticus*, XM 004570537.4 *Maylandia zebra*, XM 005939374.2 *Haplochromis burtoni*, XM 005720990.2 *Pundamilia nyererei*, XM 006786805.1 *Neolamprologus brichardi*, XM 010733332.2 *Larimichthys crocea*, XM 020650071.1 *Labrus bergylta*, XM 017430730.2 *Kryptolebias marmoratus*, XM 023414750.1 *Seriola lalandi*, XM 023270576.1 *Amphiprion ocellaris*, XM 022754919.1 *Seriola dumerili*, XM 020615999.1 *Monopterus albus*, XM 018677174.1 *Lates calcarifer*, XM 008433353.2 *Poecilia reticulata*, XM 015050118.1 *Poecilia latipinna*, XM 023335956.1 *Xiphophorus maculatus*, XM 022214317.1 *Acanthochromis polyacanthus*, XM 012860965.2 *Fundulus heteroclitus*, XM 007568728.2 *Poecilia formosa*, XM 015005957.1 *Poecilia mexicana*, XM 014025935.1 *Austrofundulus limnaeus*, XM 003975720.2 *Takifugu rubripes*, XM 008279870.1 *Stegastes partitus*, XM 015940556.1 *Nothobranchius furzeri*, XM 015404400.1 *Cyprinodon variegatus*, XM 024391948.1 *Oncorhynchus tshawytscha*, XM 019877778.1 *Hippocampus comes*, XM 004078643.4 *Oryzias latipes* and XM 020463136.1 *Oncorhynchus kisutch.*


1. Salmo_salar
2. Oreochromis_mossambicus
3. Salvelinus_alpinus
4. Oncorhynchus_mykiss
5. Oreochromis_niloticus
6. Maylandia_zebra
7. Haplochromis_burtoni
8. Pundamilia_nyererei
9. Neolamprologus_brichardi
10. Larimichthys_crocea
11. Labrus_bergylta
12. Kryptolebias_marmoratus
13. Seriola_lalandi
14. Amphiprion_ocellaris
15. Seriola_dumerili
16. Monopterus_albus
17. Lates_calcarifer
18. Poecilia_reticulata
19. Poecilia_latipinna
20. Xiphophorus_maculatus
21. Etroplus_suratensis
22. Acanthochromis_polyaca...
23. Fundulus_heteroclitus
24. Poecilia_formosa
25. Poecilia_mexicana
26. Austrofundulus_limnaeus
27. Takifugu_rubripes
28. Stegastes_partitus
29. Nothobranchius_furzeri
30. Cyprinodon_variegatus
31. Oncorhynchus_tshawytsc...
32. Hippocampus_comes
33. Oryzias_latipes
34. Oncorhynchus_kisutch
1
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420

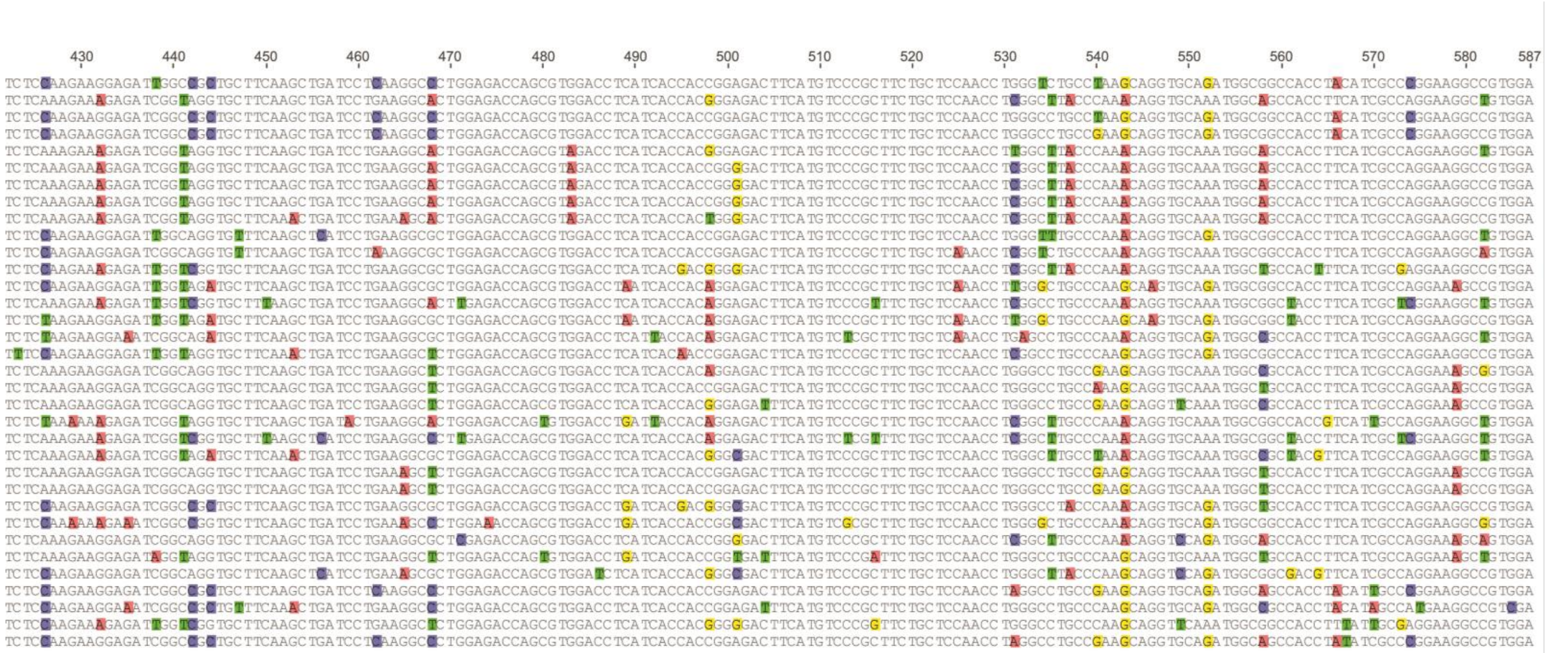


**Fig. 9.S2** Multiple sequence alignment of TF II B nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratensis*. GenBank accession no: BT125312.1 *Salmo salar*, AY679525.1 *Oreochromis mossambicus*, XM 019348123.2 *Oreochromis niloticus*, XM 004570537.4 *Maylandia zebra*, XM 005939374.2 *Haplochromis burtoni*, XM 005720990.2 *Pundamilia nyererei*, XM 006786805.1 *Neolamprologus brichardi*, XM 010733332.2 *Larimichthys crocea*, XM 020650071.1 *Labrus bergylta*, XM 017430730.2 *Kryptolebias marmoratus*, XM 023414750.1 *Seriola lalandi*, XM 023270576.1 *Amphiprion ocellaris*, XM 022754919.1 *Seriola dumerili*, XM 020615999.1 *Monopterus albus*, XM 018677174.1 *Lates calcarifer*, XM 008433353.2 *Poecilia reticulata*, XM 015050118.1 *Poecilia latipinna*, XM 023335956.1 *Xiphophorus maculatus*, XM 022214317.1 *Acanthochromis polyacanthus*, XM 012860965.2 *Fundulus heteroclitus*, XM 007568728.2 *Poecilia formosa*, XM 015005957.1 *Poecilia mexicana*, XM 014025935.1 *Austrofundulus limnaeus*, XM 003975720.2 *Takifugu rubripes*, XM 008279870.1 *Stegastes partitus*, XM 015940556.1 *Nothobranchius furzeri*, XM 015404400.1 *Cyprinodon variegatus*, XM 024391948.1 *Oncorhynchus tshawytscha*, XM 019877778.1 *Hippocampus comes*, XM 004078643.4 *Oryzias latipes* and XM 020463136.1 *Oncorhynchus kisutch.*

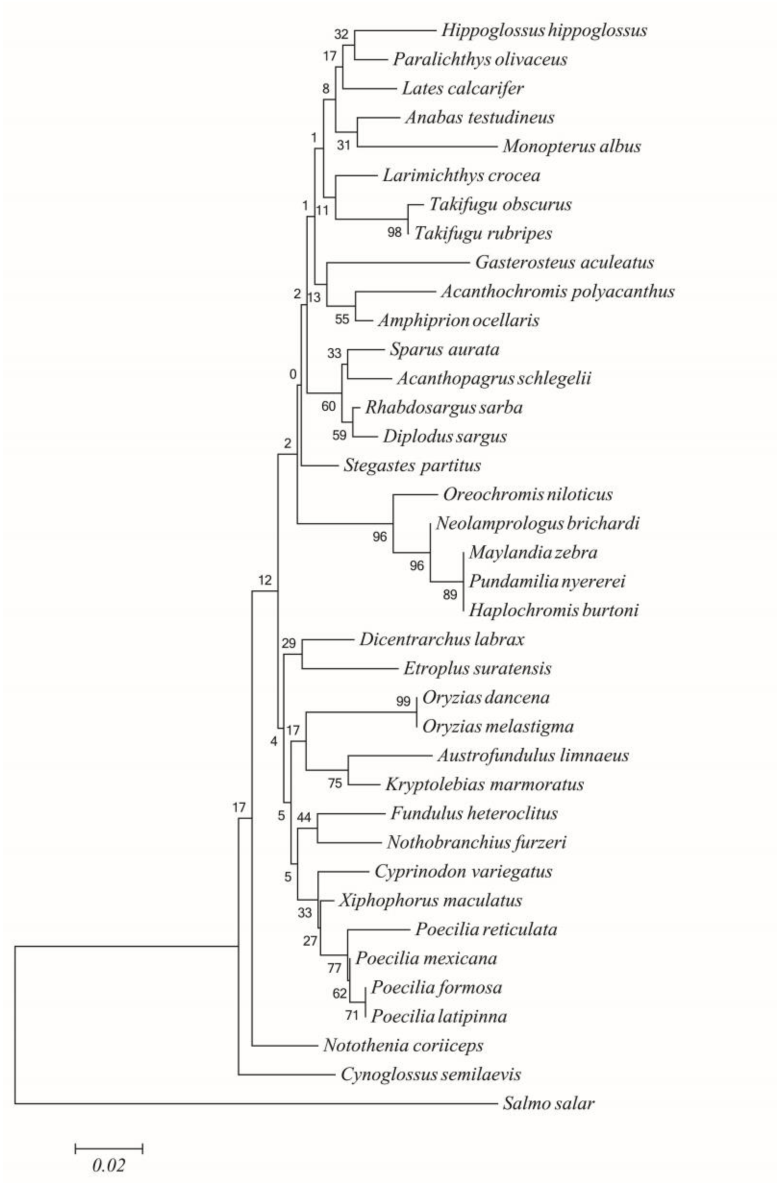
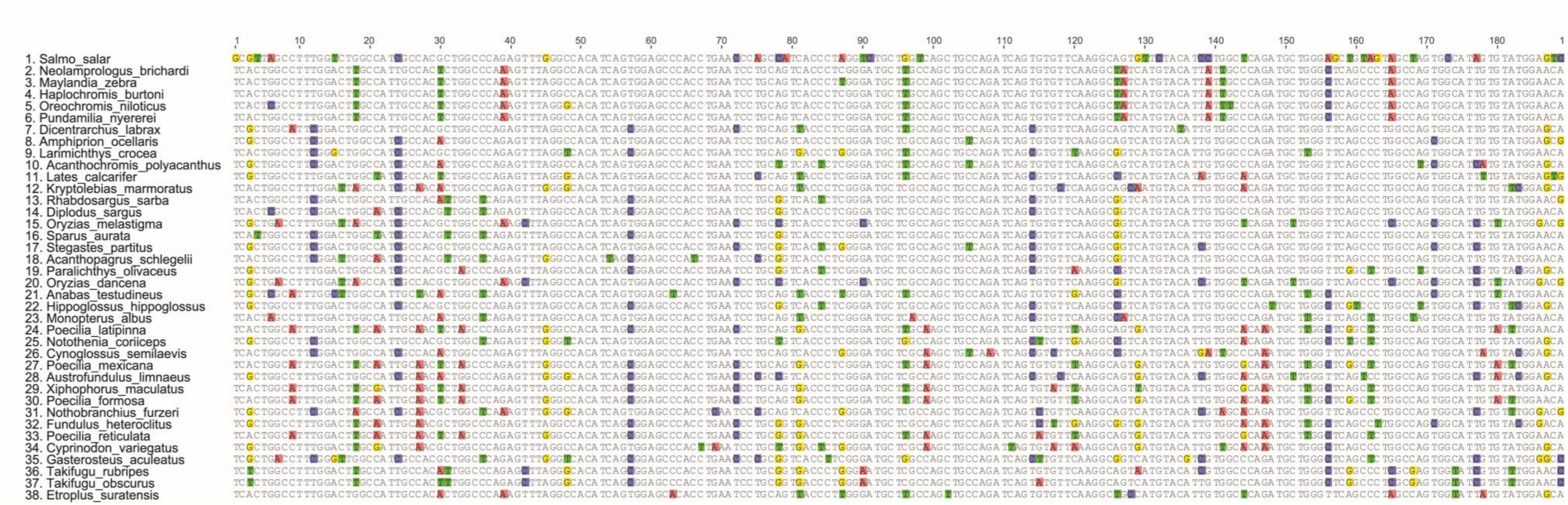
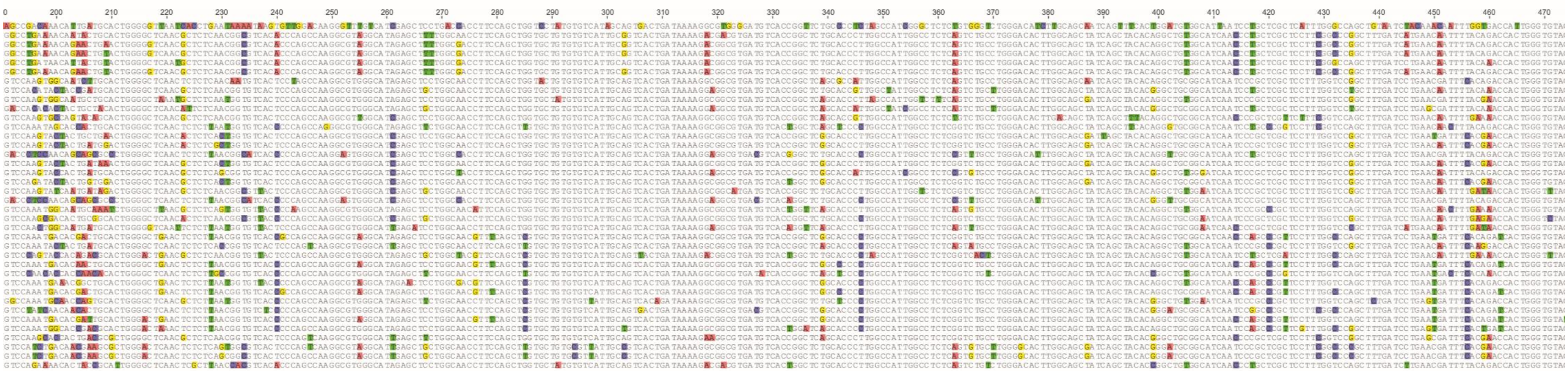


**Fig. 9.S3** Phylogenetic tree showing the evolutionary relationship between AQP1 nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratensis*. *Salmo salar* was used as an out group. GenBank accession no: NM 001140000.1 *Salmo salar*, AB610921.1 *Takifugu obscurus*, AB759556.1 *Oryzias dancena*, AY626939.1 *Sparus aurata*, BT028510.1 *Gasterosteus aculeatus*, DQ924529.3 *Dicentrarchus labrax*, EF451961.1 *Acanthopagrus schlegelii*, HQ185294.1 *Hippoglossus hippoglossus*, JF803845.1 *Rhabdosargus sarba*, JN210582.1 *Diplodus sargus*, JX645188.1 *Anabas testudineus*, NM 001309974.*1 Fundulus heteroclitus*, XM 003438085.5 *Oreochromis niloticus*, XM 003975326.2 *Takifugu rubripes*, XM 004542847.2 *Maylandia zebra*, XM 005742590.*1 Pundamilia nyererei*, XM 005809446.3 *Xiphophorus maculatus*, XM 005912557.2 *Haplochromis burtoni*, XM 006792098.1 *Neolamprologus brichardi*, XM 007548621.2 *Poecilia formosa*, XM 008282895.1 *Stegastes partitus*, XM 008334444.2 *Cynoglossus semilaevis*, XM 008434312.2 *Poecilia reticulata*, XM 010729217.2 *Larimichthys crocea*, XM 010767642.1 *Notothenia coriiceps*, XM 014029414.1 *Austrofundulus limnaeus*, XM 014980920.1 *Poecilia mexicana*, XM 015034612.1 *Poecilia latipinna*, XM 015369857.1 *Cyprinodon variegatus*, XM 015956676.1 *Nothobranchius furzeri*, XM 017425942.2 *Kryptolebias marmoratus*, XM 018676329.1 *Lates calcarifer*, XM 020113387.1 *Paralichthys olivaceus*, XM 020591485.1 *Monopterus albus*, XM 022221094.1 *Acanthochromis polyacanthus*, XM 023270352.1 *Amphiprion ocellaris* and XM 024273629.1 *Oryzias melastigma*.


1
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
1
1. Salmo_salar
2. Neolamprologus_brichardi
3. Maylandia_zebra
4. Haplochromis_burtoni
5. Oreochromis_niloticus
6. Pundamilia_nyererei
7. Dicentrarchus_labrax
8. Amphiprion_ocellaris
9. Larimichthys_crocea
10. Acanthochromis_polyacanthus
11. Lates_calcarifer
12. Kryptolebias_marmoratus
13. Rhabdosargus_sarba
14. Diplodus_sargus
15. Oryzias_melastigma
16. Sparus_aurata
17. Stegastes_partitus
18. Acanthopagrus_schlegelii
19. Paralichthys_olivaceus
20. Oryzias_dancena
21. Anabas_testudineus
22. Hippoglossus_hippoglossus
23. Monopterus_albus
24. Poecilia_latipinna
25. Notothenia_coriiceps
26. Cynoglossus_semilaevis
27. Poecilia_mexicana
28. Austrofundulus_limnaeus
29. Xiphophorus_maculatus
30. Poecilia_formosa
31. Nothobranchius_furzeri
32. Fundulus_heteroclitus
33. Poecilia_reticulata
34. Cyprinodon_variegatus
35. Gasterosteus_aculeatus
36. Takifugu_rubripes
37. Takifugu_obscurus
38. Etroplus_suratensis


0
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470

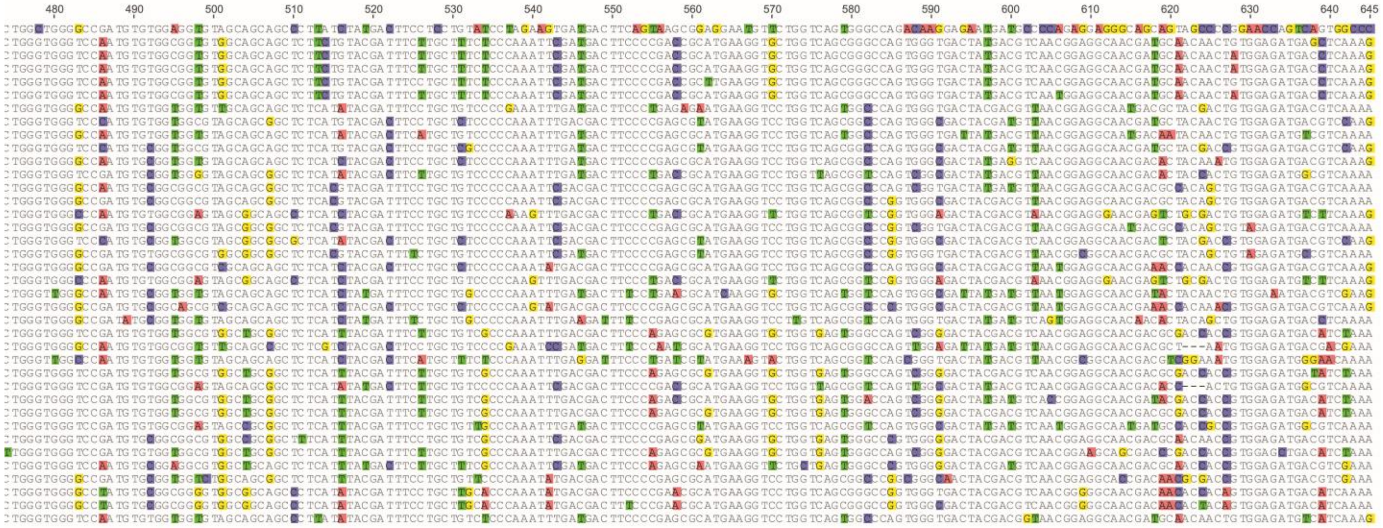


**Fig. 9.S4** Multiple sequence alignment of AQP1 nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratens.* GenBank accession no: NM 001140000.1 *Salmo salar*, AB610921.1 *Takifugu obscurus*, AB759556.1 *Oryzias dancena*, AY626939.1 *Sparus aurata*, BT028510.1 *Gasterosteus aculeatus*, DQ924529.3 *Dicentrarchus labrax*, EF451961.1 *Acanthopagrus schlegelii*, HQ185294.1 *Hippoglossus hippoglossus*, JF803845.1 *Rhabdosargus sarba*, JN210582.1 *Diplodus sargus*, JX645188.1 *Anabas testudineus*, NM 001309974.*1 Fundulus heteroclitus*, XM 003438085.5 *Oreochromis niloticus*, XM 003975326.2 *Takifugu rubripes*, XM 004542847.2 *Maylandia zebra*, XM 005742590.*1 Pundamilia nyererei*, XM 005809446.3 *Xiphophorus maculatus*, XM 005912557.2 *Haplochromis burtoni*, XM 006792098.1 *Neolamprologus brichardi*, XM 007548621.2 *Poecilia formosa*, XM 008282895.1 *Stegastes partitus*, XM 008334444.2 *Cynoglossus semilaevis*, XM 008434312.2 *Poecilia reticulata*, XM 010729217.2 *Larimichthys crocea*, XM 010767642.1 *Notothenia coriiceps*, XM 014029414.1 *Austrofundulus limnaeus*, XM 014980920.1 *Poecilia mexicana*, XM 015034612.1 *Poecilia latipinna*, XM 015369857.1 *Cyprinodon variegatus*, XM 015956676.1 *Nothobranchius furzeri*, XM 017425942.2 *Kryptolebias marmoratus*, XM 018676329.1 *Lates calcarifer*, XM 020113387.1 *Paralichthys olivaceus*, XM 020591485.1 *Monopterus albus*, XM 022221094.1 *Acanthochromis polyacanthus*, XM 023270352.1 *Amphiprion ocellaris* and XM 024273629.1 *Oryzias melastigma.*

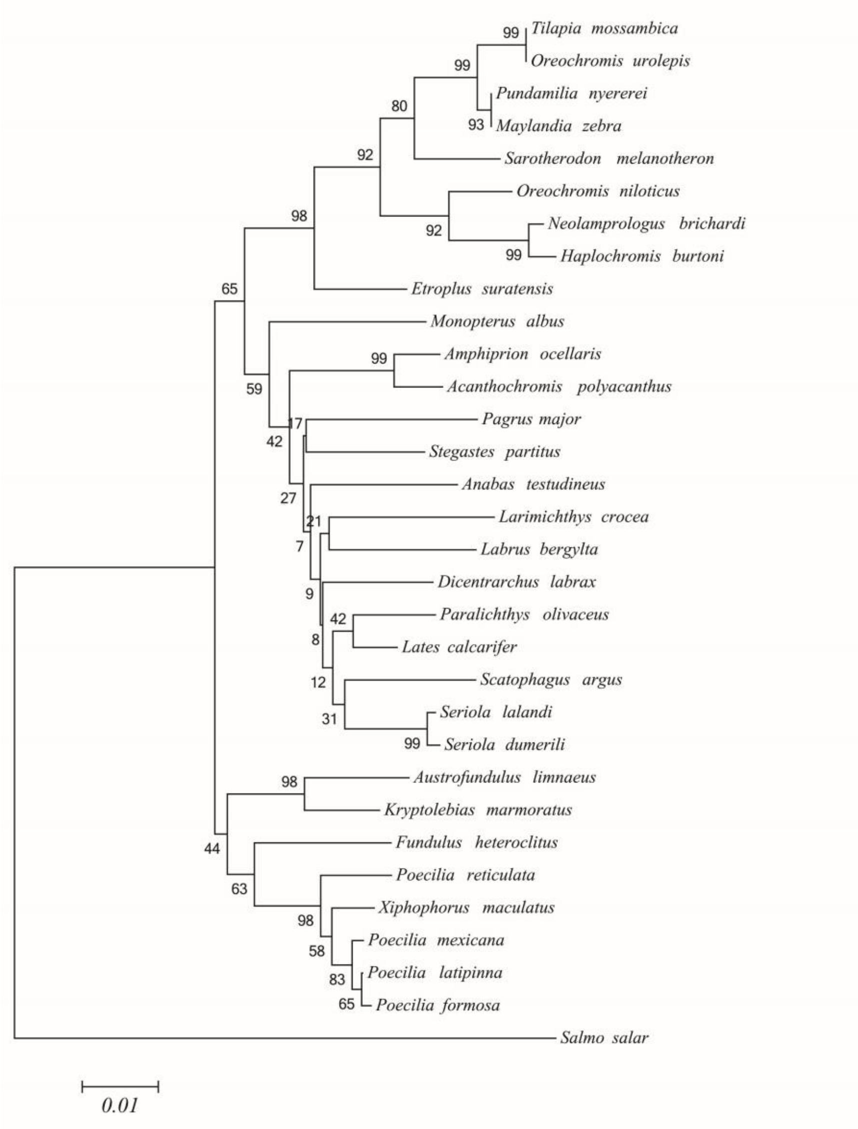
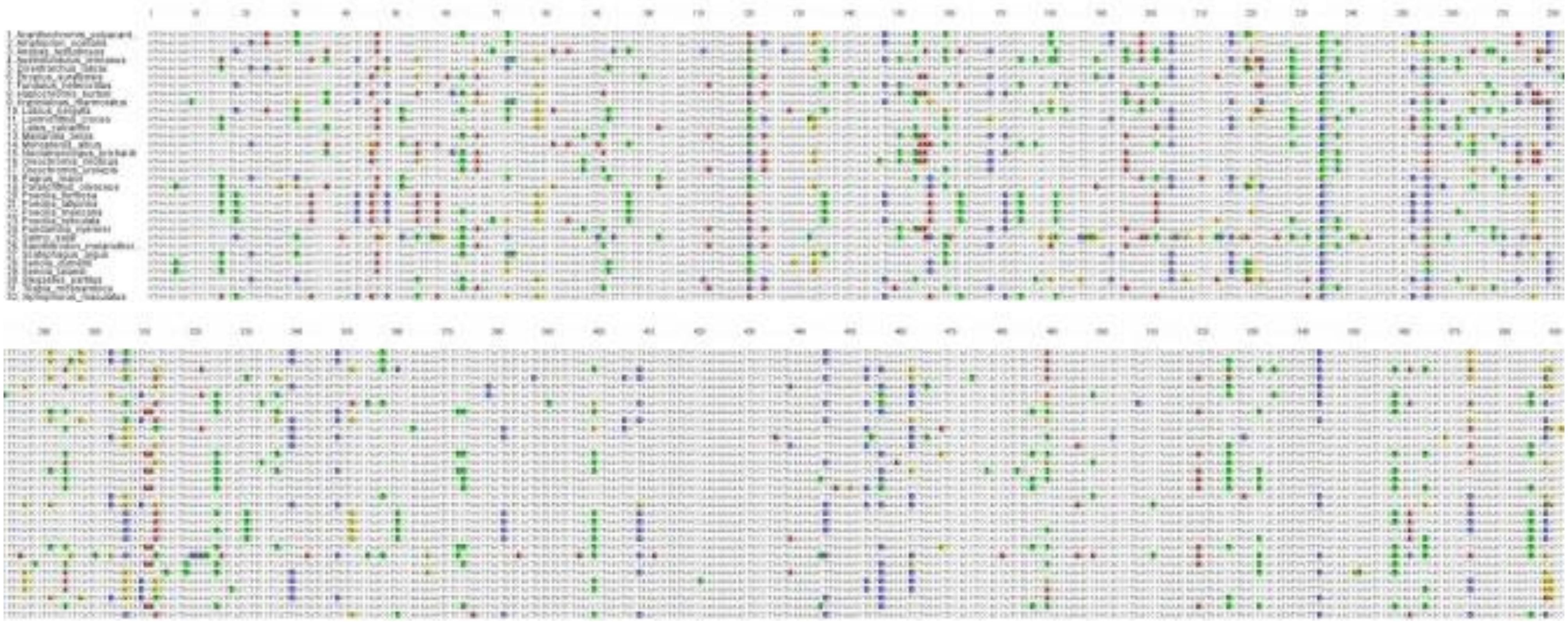
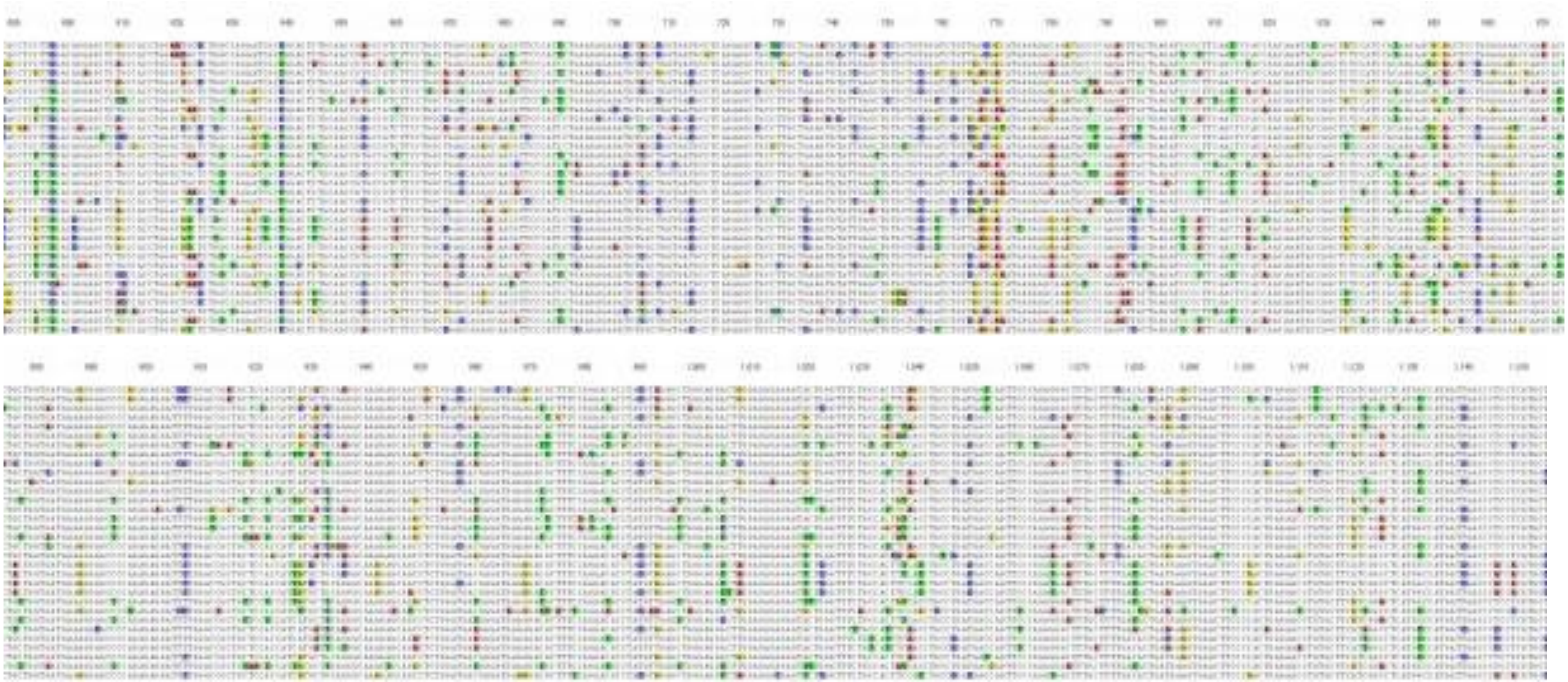
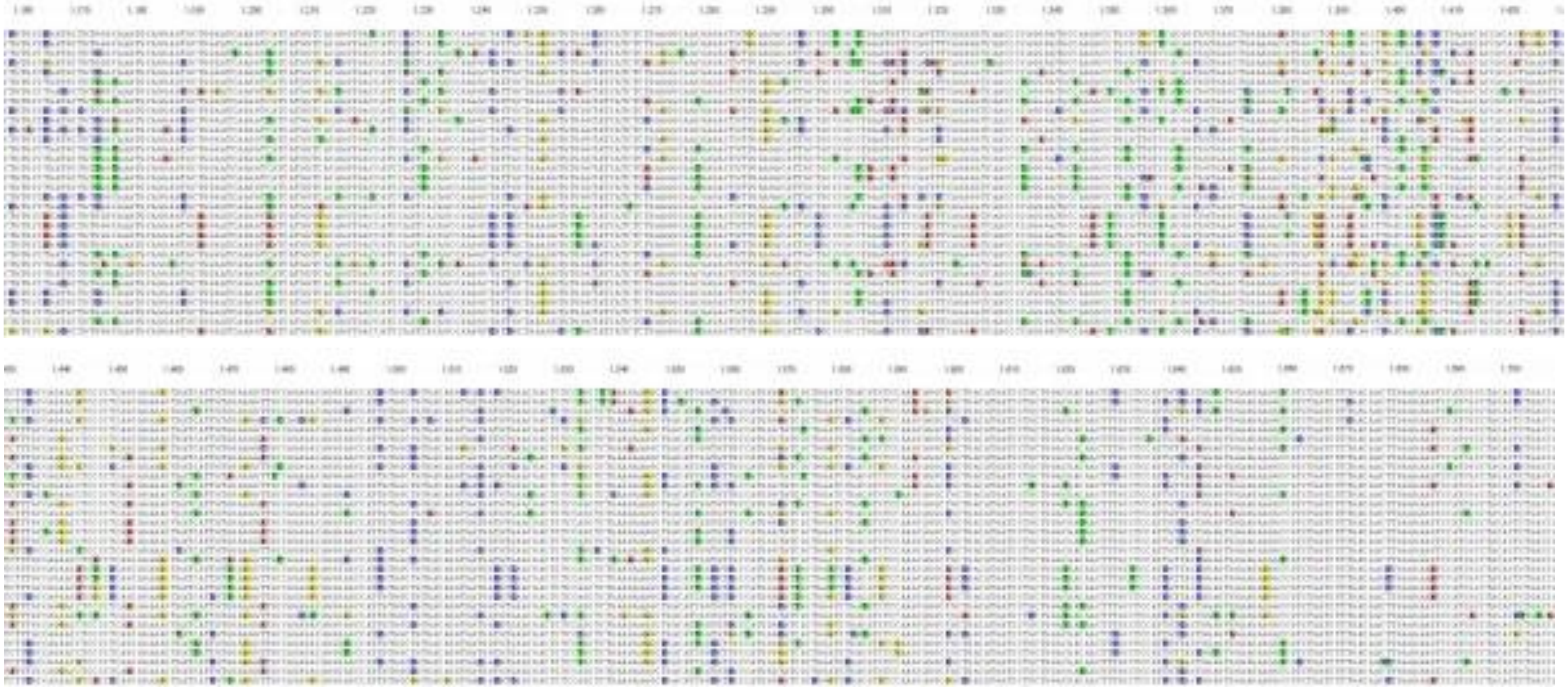


**Fig. 9.S5** Phylogenetic tree showing the evolutionary relationship between Na/K-ATPase α1nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratensis* by Neighbour Joining method. *Salmo salar* was used as an out group. GenBank accession no: XM022222344*.1 Acanthochromis polyacanthus*, XM023265789.1 *Amphiprion ocellaris*, JN180942.1 *Anabas testudineus*, XM014019437.1 *Austrofundulus limnaeus*, KP400258.1 *Dicentrarchus labrax*, NM001310013.1 *Fundulus heteroclitus*, XM014338426.1 *Haplochromis burtoni*, XM017426902.2 *Kryptolebias marmoratus*, XM020642189.1 *Labrus bergylta*, XM019273216.1 *Larimichthys crocea*, XM018661008.1 *Lates calcarifer*, XM004571251.2 *Maylandia zebra*, XM020620527.1 *Monopterus albus*, XM006792814.1 *Neolamprologus brichardi*, XM005452356.4 *Oreochromis niloticus*, KC702516.1 *Oreochromis urolepis*, KT203392.2 *Pagrus major*, XM020104090.1 *Paralichthys olivaceus*, XM007567594.2 *Poecilia formosa*, XM015051596.1 *Poecilia latipinna*, XM014990533.1 *Poecilia mexicana*, XM008423384.2 *Poecilia reticulata*, XM005749450.2 *Pundamilia nyererei*, GU252208.1 *Sarotherodon melanotheron*, KF649217.1 *Scatophagus argus*, XM022761526.1 *Seriola dumerili*, XM023420182.1 *Seriola lalandi*, XM008284129.1 *Stegastes partitus*, U82549.2 *Tilapia mossambica*, XM023329863.1 *Xiphophorus maculatus* and KJ175156.1 *Salmo salar.*







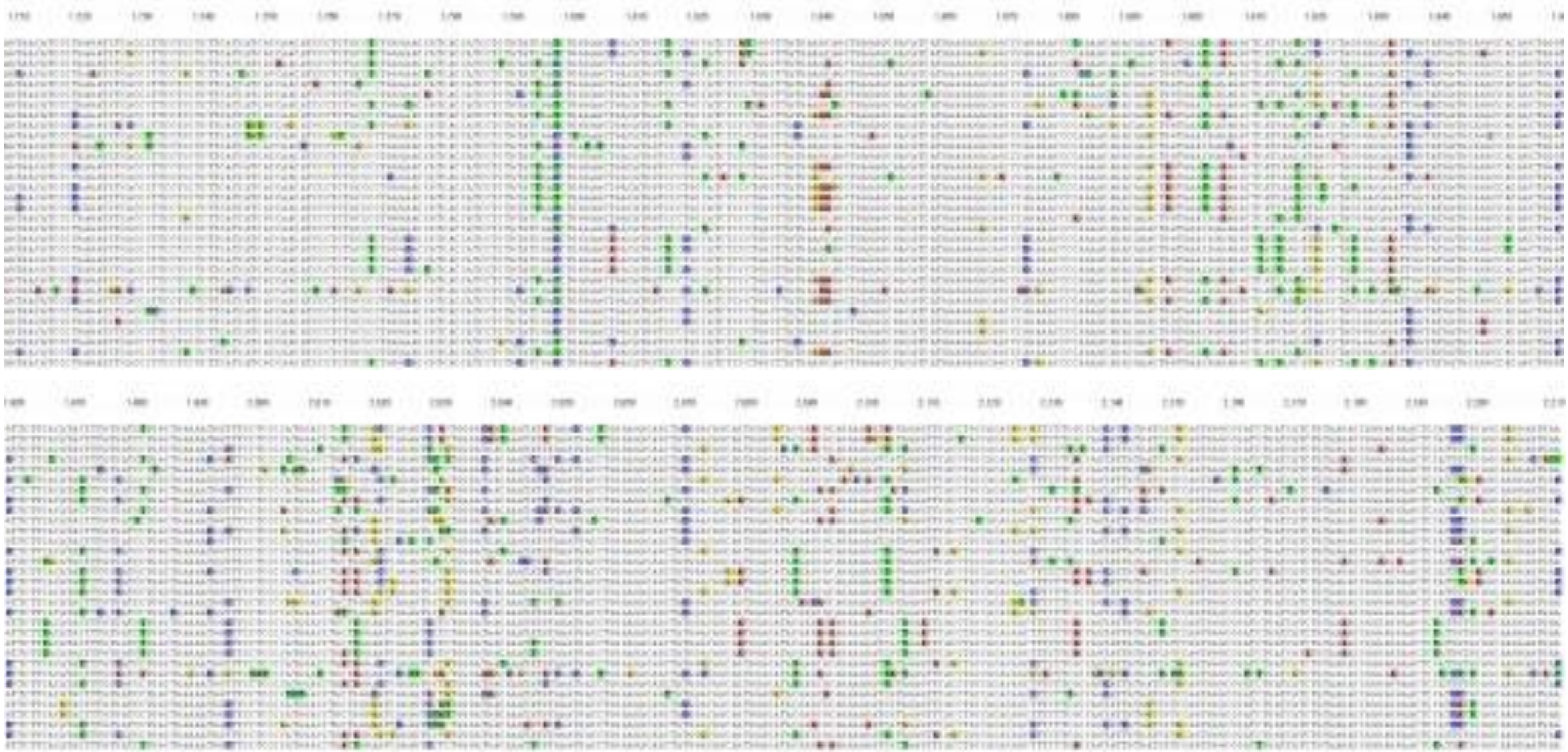


**Fig. 9.S6** Multiple sequence alignment of Na/K-ATPase α1nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratensis*. GenBank accession no: XM022222344.1 *Acanthochromis polyacanthus*, XM023265789.1 *Amphiprion ocellaris*, JN180942.1 *Anabas testudineus*, XM014019437.1 *Austrofundulus limnaeus*, KP400258.1 *Dicentrarchus labrax*, NM001310013.1 *Fundulus heteroclitus*, XM014338426.1 *Haplochromis burtoni*, XM017426902.2 *Kryptolebias marmoratus*, XM020642189.1 *Labrus bergylta*, XM019273216.1 *Larimichthys crocea*, XM018661008.1 *Lates calcarifer*, XM004571251.2 *Maylandia zebra*, XM020620527.1 *Monopterus albus*, XM006792814.1 *Neolamprologus brichardi*, XM005452356.4 *Oreochromis niloticus*, KC702516.1 *Oreochromis urolepis*, KT203392.2 *Pagrus major*, XM020104090.1 *Paralichthys olivaceus*, XM007567594.2 *Poecilia formosa*, XM015051596.1 *Poecilia latipinna*, XM014990533.1 *Poecilia mexicana*, XM008423384.2 *Poecilia reticulata*, XM005749450.2 *Pundamilia nyererei*, GU252208.1 *Sarotherodon melanotheron*, KF649217.1 *Scatophagus argus*, XM022761526.1 *Seriola dumerili*, XM023420182.1 *Seriola lalandi*, XM008284129.1 *tegastes partitus*, U82549.2 *Tilapia mossambica*, XM023329863.1 *Xiphophorus maculatus* and KJ175156.1 *Salmo salar*.

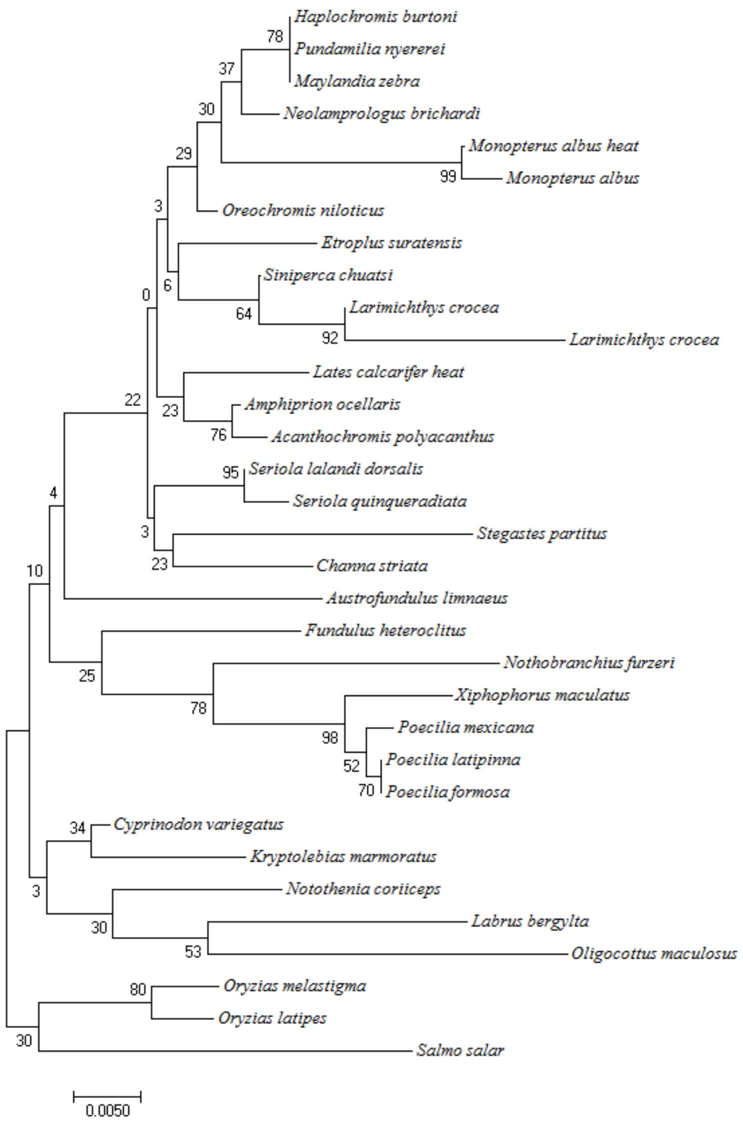


**Fig. 9.S7** Phylogenetic tree showing the evolutionary relationship between HSC 71 nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratensis* by Neighbour Joining method. *Salmo salar* was used as an out group. GenBank accession no: XM003455056.5 *Oreochromis niloticus*, XM004558114.4 *Maylandia zebra*, XM005937893.2 *Haplochromis burtoni*, XM005739307.2 *Pundamilia nyererei*, XM023402313.1 *Seriola lalandi*, KF017616.1 *Siniperca chuatsi*, XM006801582.1 *Neolamprologus brichardi*, AB436469.1 *Seriola quinqueradiata*, XM023265103.1 *Amphiprion ocellaris*, XM022218869.1 *Acanthochromis polyacanthus*, XM020622419.1 *Monopterus albus*, XM018700416.1 *Lates calcarifer*, XM008288792.1 *Stegastes partitus*, XM019270840.1 *Larimichthys crocea*, XM010770172.1 *Notothenia coriiceps*, HF955035.1 *Channa striata*, XM019259354.1 *Larimichthys crocea*, XM015373232.1 *Cyprinodon variegatus*, XM020654713.1 *Labrus bergylta*, XM004075347.4 *Oryzias latipes*, XM015951144.1 *Nothobranchius furzeri*, DQ013308.1 *Oligocottus maculosus*, XM014012411.1 *Austrofundulus limnaeus*, XM012881994.2 *Fundulus heteroclitus*, XM014979569.1 *Poecilia mexicana*, XM015033667.1 *Poecilia latipinna*, XM007554403.2 *Poecilia formosa*, XM024277119.1 *Oryzias melastigma*, XM023265101.1 *Amphiprion ocellaris*, NM001286281.1 *Xiphophorus maculatus*, XM017428353.2 *Kryptolebias marmoratus* and XM014162783.1 *Salmo salar*.

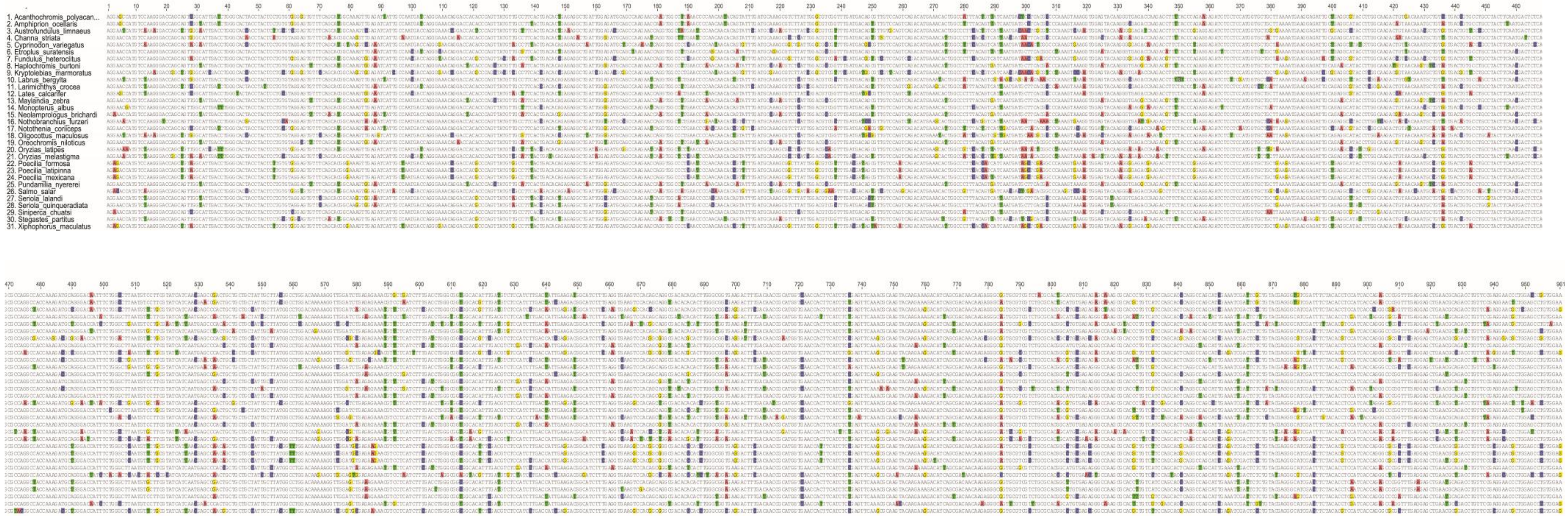


**Fig. 9.S8** Multiple sequence alignment of HSC 71 nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratensis.* GenBank accession no: XM003455056.5 *Oreochromis niloticus*, XM004558114.4 *Maylandia zebra*, XM005937893.2 *Haplochromis burtoni*, XM005739307.2 *Pundamilia nyererei*, XM023402313.1 *Seriola lalandi*, KF017616.1 *Siniperca chuatsi*, XM006801582.1 *Neolamprologus brichardi*, AB436469.1 *Seriola quinqueradiata*, XM023265103.1 *Amphiprion ocellaris*, XM022218869.1 *Acanthochromis polyacanthus*, XM020622419.1 *Monopterus albus*, XM018700416.1 *Lates calcarifer*, XM008288792.1 *Stegastes partitus*, XM019270840.1 *Larimichthys crocea*, XM010770172.1 *Notothenia coriiceps*, HF955035.1 *Channa striata*, XM019259354.1 *Larimichthys crocea*, XM015373232.1 *Cyprinodon variegatus*, XM020654713.1 *Labrus bergylta*, XM004075347.4 *Oryzias latipes*, XM015951144.1 *Nothobranchius furzeri*, DQ013308.1 *Oligocottus maculosus*, XM014012411.1 *Austrofundulus limnaeus*, XM012881994.2 *Fundulus heteroclitus*, XM014979569.1 *Poecilia mexicana*, XM015033667.1 *Poecilia latipinna*, XM007554403.2 *Poecilia formosa*, XM024277119.1 *Oryzias melastigma*, XM023265101.1 *Amphiprion ocellaris*, NM001286281.1 *Xiphophorus maculatus*, XM017428353.2 *Kryptolebias marmoratus* and XM014162783.1 *Salmo salar*.

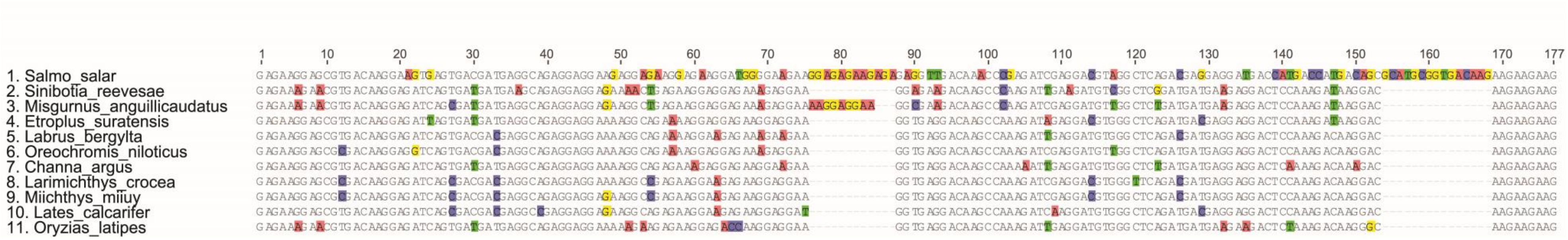


**Fig. 9.S9** Multiple sequence alignment of HSP90 nucleotide sequences of bony fishes (available in NCBI, GenBank) and *E. suratensis.* GenBank accession no: NM001173702.1 *Salmo salar*, KJ624420.1 *Sinibotia reevesae*, XM020628015.1 *Labrus bergylta*, XM013271984.1 *Oreochromis niloticus*, KU946993.1 *Channa argus*, KT456552.1 *Larimichthys crocea*, JQ929760.1 *Miichthys miiuy*, HQ646106.1 *Lates calcarifer*, CP020802.1 *Oryzias latipes* and KY203858.1 *Misgurnus anguillicaudatus.*

## 5. References

1. Jayaram KC (1999) The Freshwater Fishes of the Indian Region. Narendra Publishing House, India
2. Padmakumar KG, Bindu L, Manu PS, (2012) *Etroplus suratensis* (Bloch), the State Fish of Kerala. *J Biosci* 37(1):925–931
3. Chandrasekar S, Nich T, Tripathi G, Sahu NP, Pal AK, Dasgupta S (2014) Acclimation of brackish water pearl spot (*Etroplus suratensis*) to various salinities: relative changes in abundance of branchial Na+/K+-ATPase and Na+/K+/2Cl− co-transporter in relation to osmoregulatory parameters. *Fish Physiol Biochem* 40(3):983–996
4. Jayakumar M (2002) Wetland conservation and Management in Kerala. State Committee on Science, Technology and Environment, Thiruvananthapuram, Kerala, India
5. Larsen PF, Nielsen EE, Williams TD, Hemmer-Hansen JA, Chipman JK, Kruhoffer M, Gronkjaer P, George SG, Dyrskjot L, Loeschcke V (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Mol Ecol* 16(22):4674–4683
6. Tine M, Bonhomme F, McKenzie DJ, Durand J D (2010) Differential expression of the heat shock protein Hsp70 in natural populations of the tilapia, *Sarotherodon melanotheron*, acclimatised to a range of environmental salinities. *BMC Ecol* 10(1):11
7. Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger GL, Ogden R, Carvalho GR (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evol Biol* 9(1):276
8. Charlesworth D, Barton NH, Charlesworth B (2017). The sources of adaptive variation. *Proc R Soc B* 284(1855):20162864
9. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729
10. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31(13):3381–3385
11. Chrispeels M J, Agre P (1994) Aquaporins: water channel proteins of plant and animal cells. *Trends Biochem Sci* 19(10):421–425
12. Moller JV, Juul B, le Maire M (1996) Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochimica Et Biophysica Acta-Reviews on Biomembranes* 1286(1):1–51
13. Ohta S, Shimekake Y, Nagata K (1996) Molecular Cloning and Characterization of a Transcription Factor for the C-Type Natriuretic Peptide Gene Promoter. *FEBS J* 242(3):460–466
14. Weinmann R (1992) The basic RNA polymerase II transcriptional machinery. *Gene expression* 2(2):81–91
15. Lindquist S, Craig EA (1988) The heat-shock proteins. *Annu Rev Genet* 22(1):631–677
16. Finn RN, Cerda J (2011) Aquaporin evolution in fishes. *Front Physiol* 2(1):44
17. Deane EE, Luk JC, Woo N (2011). Aquaporin 1a expression in gill, intestine, and kidney of the euryhaline silver sea bream. *Front Physiol* 2(1):39
18. Yang WK, Kang CK, Hsu AD, Lin CH, Lee TH (2016). Different modulatory mechanisms of renal FXYD12 for Na+-K+-ATPase between two closely related medakas upon salinity challenge. Int *J Biol Sci* 12(6):730–745
19. Kanai R, Ogawa H, Vilsen B, Cornelius F, Toyoshima C (2013) Crystal structure of a Na+-bound Na+, K+-ATPase preceding the E1P state. *Nature* 502(7470):201–206
20. Basu N, Todgham AE, Ackerman PA, Bibeau MR, Nakano K, Schulte PM, Iwama GK (2002). Heat shock protein genes and their functional significance in fish. *Gene* 295(2):173–183
21. Roberts RJ, AgiusC, SalibaC, Bossier P, Sung YY (2010) Heat shock proteins (chaperones) in fish and shellfish and their potential role in relation to fish health: a review. *J Fis Dis* 33(10):789–801
22. Fiol DF, Chan SY, Kultz D (2006) Regulation of osmotic stress transcription factor 1 (Ostf1) in tilapia (*Oreochromis mossambicus*) gill epithelium during salinity stress. *J Exp Biol* 209(16):3257–3265

23. Tse WKF (2014) The role of osmotic stress transcription factor 1 in fishes. *Front Zool* 11(1):86
24. HampseyM (1998) Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol *Mol Biol Rev* 62(2):465–503

# *Chapter 10*

## CONCLUSIONS

The role of nuclear and mitochondrial DNA in energy metabolism, adaptation, diseases and the consequent survival of organisms has been implicated in many scientific investigations. We investigated whether a widely distributed pelagic, planktivorous fish, Indian oil sardine, *Sardinella longiceps* (Valenciennes, 1847) in their dynamic oceanic environment exhibits any population structuring and specific patterns of selection in the mitochondrial and nuclear genome. *S. longiceps* is the most abundant, economic and ecologically important fish resource from the Indian Ocean. We also investigated whether there are population genetic structure and adaptive variations in Green chromide, *Etroplus suratensis* (Bloch, 1790) in their brackish and freshwater habitats, which is fragmented by geographical barriers. *E. suratensis* is an important candidate aquaculture Cichlid species, endemic to India and SriLanka.

Our findings address the three most important aspects relevant to effective management, biodiversity conservation and genetic improvement to sustain the climate change.

- Mechanisms of species and ecosystems resilience
- Biological adaptations and evolutionary processes
- Management in the face of climate change

### 1. Conclusions and Contributions

- The first report of the complete mitochondrial genome sequence of *S. longiceps, S. gibbosa and E. suratensis* revealed gene organization, structure, content and order similar to most vertebrates. This work is a significant contribution to genomic resource development of these fishes which are economically and ecologically important.
- The ddRAD sequences, SNPs and microsatellite markers of *S. longiceps and E. suratensis* are also important as genomic resources.
- Mitochondrial and nuclear genomic DNA information/resource developed for *S. longiceps and E. suratensis* constitute important contributions to future genetic

studies including taxonomy, conservation, adaptation to environmental clines and evolution.

- Complete mitogenome based analysis revealed the phylogenetic relationship of *S. longiceps and E. suratensis* with other fishes.
- Mitochondrial and nuclear DNA markers revealed population genetic structure in *S. longiceps and E. suratensis* in their habitats. A very strong genetic structure was identified between Oman and Indian Ocean samples of *S. longiceps* and a comparatively low genetic differentiation in the Indian coastal line samples, between North East Arabian Sea and others (South East Arabian Sea, South West Bay of Bengal and North West Bay of Bengal).

  Very high level of sub-structuring was observed between *E. suratensis* samples collected from Indian waters. We also reported a reduction in the genetic diversity and effective population size in the contemporary population.

  This information is crucial for developing species-specific management and conservation plans for these two economically and ecologically important species.

- The association of genetic differentiation with environmental factors has been established in our study. The candidate loci/sites identified as under selective constraints from mitochondrial and nuclear DNA of *S. longiceps and E. suratensis* indicate adaptive variations/local adaptation in their habitats.

  These candidate loci can be used further as genetic tags for locally adapted populations and genetic improvement of organisms. We also need to study whether any of these variants will be more successful in future when climatic fluctuations occur and how their distribution will be affected.

- The evolutionary analysis of the mitogenome of Clupeoids and the relation between mitochondrial genomic selection and their distribution in marine, brackish and freshwaters of tropical and temperate regions of the world ocean provide insights regarding the role of vertebrate mitogenome evolution on habitat adaptation.

  Clupeoids mitogenomes are adapted to deamination mutations in anticodon sites, during replication and transcription. Translational efficiency-related constraints in mtDNA were shaped by the codon usage pattern in Clupeoids. Thus, the observed codon usage pattern may be associated with an increased energy requirement for adaptation in the euryhaline and freshwater environment.

We confirmed the ability of sequence flanking conserved sequence elements in the control region (a non-coding region in the mitochondrial genome) to form stable secondary structures similar to the tRNA. The evidence for selective constraints on secondary structures emphasizes the role of the control region in mitogenome function.

We obtained evidence for positive selection in the OXPHOS protein complex of distantly related clupeoid species distributed from temperate to tropic and marine to the freshwater environment.  Variations were observed in the property of amino acids, codon usage and base composition across lineages with specific metabolic requirements such as marine to fresh/brackish water transition.

Insights from our study indicated the need for future experimental characterisation of specific mutations in the oxidative phosphorylation and its physiological impacts which will be useful for predicting the response of organisms to climate change. Further this information can be used for mitochondrial DNA based genetic improvement.

## 2. Future directions

Genetic and genomic approaches could be powerful tools for adaptation and resilience to climate change. In addition, molecular genetic information provides insights regarding the mechanism of evolution, adaptation and diversification of living organisms on earth's diverse habitats. The living organisms in the world are reported to have amazing level of diversity as known in the case of plants and animals resistant to extreme temperature, drought and high level of salt concentration.

- Extreme climatic conditions are being reported worldwide which is negatively affecting the survival and diversification of life on earth by altering endemic regions or shifting ecosystem conditions in which the living organisms evolved and adapted. Changes in the weather patterns like high temperature, drought and sea-level changes have put pressure on our natural food resources like fishery and agricultural systems. Thus, it emphasizes the need for strategies for adaptation to pressures of climate change across many areas.

- Surveying variations (diversity) in the genes/biochemical systems of organisms inhabiting diverse ecosystems is important to identify its genetic background and link them with specific characteristics of the individual/population/species. Such information is not only important in devising conservation strategies (specifically for a species or a group of species in an ecosystem) but also identifying the mechanisms of adaptation or adaptive evolution of living systems in diverse habitat.

- In the natural environment, the diversification and adaptation of organisms happen by repeated natural selection occurring over millions of years on earth. Understanding the exact mechanism (biology and genetics) behind this is the key factor which can serve as biological templates for successful evolutionary pathways and specific adaptations to harsh climates. Thus, we can use such information for identification and generation of animals/plants that can survive/sustain in the changing environments.

- Tools and strategies for the transfer or manipulation of genes and pathways in non-model organisms, in a way that function as in the organism with desired traits (as a biological template) will be a natural solution/adaptation to harsh climates. It will be a natural remedy because what we only do is accelerating the evolution rate or induce a specific mutation (obtained from natural diversity) which provide adaptation to harsh climates. The genome sequencing, population genetics and CRISPR technology will have an important role to exploit and execute this understanding.

- Further studies could be carried out using the whole genome and transcriptome so as to identify genome level adaptations which will provide holistic information with respect to their genomic region of variability and adaptive capacity.

- Common garden experiments are necessary to identify or prove the importance of the identified positive selected regions in the genome. This information will be valid for their effective management, conservation and genetic improvement.

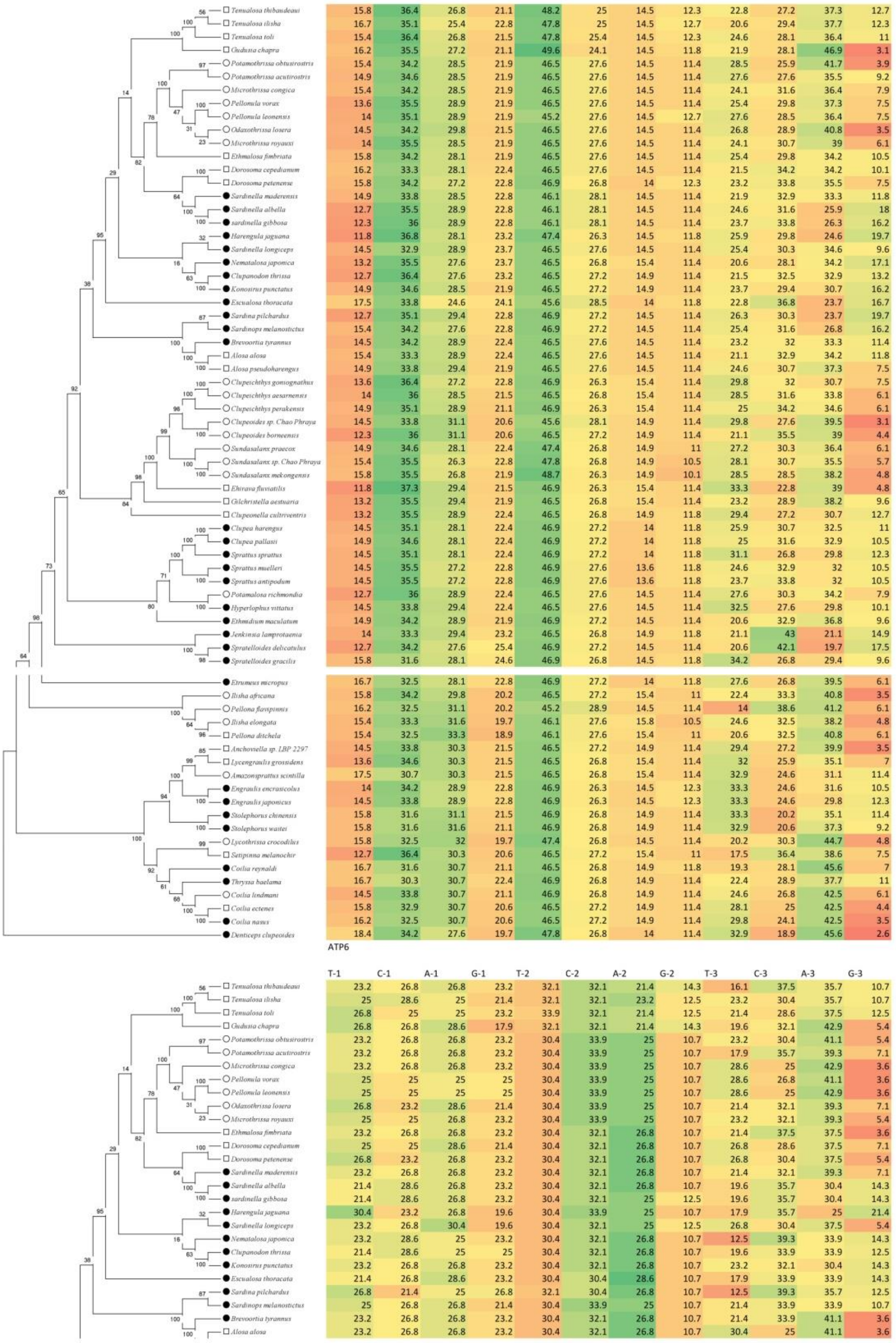
- We also should integrate knowledge about local adaptation, natal homing and spatial processes in fisheries models which is important for the design of habitat reserves in marine and fresh/brackish water regions.

- We must extend similar research on other species by characterising the diversity of its natural populations to improve resilience to changing and uncertain environments. Developing genetic tools for monitoring and managing natural diversity and distribution of organisms in the background of changing climate, will help humans to adjust with the climate change stress.

Genome/DNA in living organisms has always been able to respond and adapt to changing conditions around them. I hope that the applications of genetics and genomics research will help to secure the genetic diversity of living organisms and sustainable future for humanity.

# APPENDIX

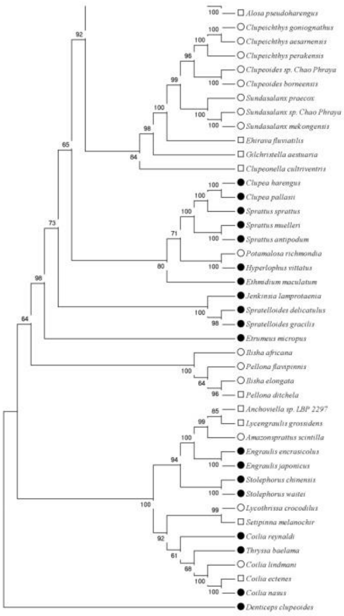
**Fig. A1 Nucleic acid contents varying across the clupeoid mitogenomic phylogenetic tree**. Nucleic acid contents of protein-coding genes of Clupeoid fishes of the present study.



| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 15.8 | 36.4 | 26.8 | 21.1 | 48.2 | 25 | 14.5 | 12.3 | 22.8 | 27.2 | 37.3 | 12.7 |
| □ *Tenualosa ilisha* | 16.7 | 35.1 | 25.4 | 22.8 | 47.8 | 25 | 14.5 | 12.7 | 20.6 | 29.4 | 37.7 | 12.3 |
| □ *Tenualosa toli* | 15.4 | 36.4 | 26.8 | 21.5 | 47.8 | 25.4 | 14.5 | 12.3 | 24.6 | 28.1 | 36.4 | 11 |
| □ *Gudusia chapra* | 16.2 | 35.5 | 27.2 | 21.1 | 49.6 | 24.1 | 14.5 | 11.8 | 21.9 | 28.1 | 46.9 | 3.1 |
| ○ *Potamothrissa obtusirostris* | 15.4 | 34.2 | 28.5 | 21.9 | 46.5 | 27.6 | 14.5 | 11.4 | 28.5 | 25.9 | 41.7 | 3.9 |
| ○ *Potamothrissa acutirostris* | 14.9 | 34.6 | 28.5 | 21.9 | 46.5 | 27.6 | 14.5 | 11.4 | 27.6 | 27.6 | 35.5 | 9.2 |
| ○ *Microthrissa congica* | 15.4 | 34.2 | 28.5 | 21.9 | 46.5 | 27.6 | 14.5 | 11.4 | 24.1 | 31.6 | 36.4 | 7.9 |
| ○ *Pellonula vorax* | 13.6 | 35.5 | 28.9 | 21.9 | 46.5 | 27.6 | 14.5 | 11.4 | 25.4 | 29.8 | 37.3 | 7.5 |
| ○ *Pellonula leonensis* | 14 | 35.1 | 28.9 | 21.9 | 45.2 | 27.6 | 14.5 | 12.7 | 27.6 | 28.5 | 36.4 | 7.5 |
| ○ *Odaxothrissa losera* | 14.5 | 34.2 | 29.8 | 21.5 | 46.5 | 27.6 | 14.5 | 11.4 | 26.8 | 28.9 | 40.8 | 3.5 |
| ○ *Microthrissa royauxi* | 14 | 35.5 | 28.5 | 21.9 | 46.5 | 27.6 | 14.5 | 11.4 | 24.1 | 30.7 | 39 | 6.1 |
| □ *Ethmalosa fimbriata* | 15.8 | 34.2 | 28.1 | 21.9 | 46.5 | 27.6 | 14.5 | 11.4 | 25.4 | 29.8 | 34.2 | 10.5 |
| □ *Dorosoma cepedianum* | 16.2 | 33.3 | 28.1 | 22.4 | 46.5 | 27.6 | 14.5 | 11.4 | 21.5 | 34.2 | 34.2 | 10.1 |
| □ *Dorosoma petenense* | 15.8 | 34.2 | 27.2 | 22.8 | 46.9 | 26.8 | 14 | 12.3 | 23.2 | 33.8 | 35.5 | 7.5 |
| ● *Sardinella maderensis* | 14.9 | 33.8 | 28.5 | 22.8 | 46.1 | 28.1 | 14.5 | 11.4 | 21.9 | 32.9 | 33.3 | 11.8 |
| ● *Sardinella albella* | 12.7 | 35.5 | 28.9 | 22.8 | 46.1 | 28.1 | 14.5 | 11.4 | 24.6 | 31.6 | 25.9 | 18 |
| ● *sardinella gibbosa* | 12.3 | 36 | 28.9 | 22.8 | 46.1 | 28.1 | 14.5 | 11.4 | 23.7 | 33.8 | 26.3 | 16.2 |
| ● *Harengula jaguana* | 11.8 | 36.8 | 28.1 | 23.2 | 47.4 | 26.3 | 14.5 | 11.8 | 25.9 | 29.8 | 24.6 | 19.7 |
| ● *Sardinella longiceps* | 14.5 | 32.9 | 28.9 | 23.7 | 46.5 | 27.6 | 14.5 | 11.4 | 25.4 | 30.3 | 34.6 | 9.6 |
| ● *Nematalosa japonica* | 13.2 | 35.5 | 27.6 | 23.7 | 46.5 | 26.8 | 15.4 | 11.4 | 20.6 | 28.1 | 34.2 | 17.1 |
| ● *Clupanodon thrissa* | 12.7 | 36.4 | 27.6 | 23.2 | 46.5 | 27.2 | 14.9 | 11.4 | 21.5 | 32.5 | 32.9 | 13.2 |
| ● *Konosirus punctatus* | 14.9 | 34.6 | 28.5 | 21.9 | 46.5 | 27.2 | 14.9 | 11.4 | 23.7 | 29.4 | 30.7 | 16.2 |
| ● *Escualosa thoracata* | 17.5 | 33.8 | 24.6 | 24.1 | 45.6 | 28.5 | 14 | 11.8 | 22.8 | 36.8 | 23.7 | 16.7 |
| ● *Sardina pilchardus* | 12.7 | 35.1 | 29.4 | 22.8 | 46.9 | 27.2 | 14.5 | 11.4 | 26.3 | 30.3 | 23.7 | 19.7 |
| ● *Sardinops melanostictus* | 15.4 | 34.2 | 27.6 | 22.8 | 46.9 | 27.2 | 14.5 | 11.4 | 25.4 | 31.6 | 26.8 | 16.2 |
| ● *Brevoortia tyrannus* | 14.5 | 34.2 | 28.9 | 22.4 | 46.5 | 27.6 | 14.5 | 11.4 | 23.2 | 32 | 33.3 | 11.4 |
| □ *Alosa alosa* | 15.4 | 33.3 | 28.9 | 22.4 | 46.5 | 27.6 | 14.5 | 11.4 | 21.1 | 32.9 | 34.2 | 11.8 |
| □ *Alosa pseudoharengus* | 14.9 | 33.8 | 29.4 | 21.9 | 46.5 | 27.6 | 14.5 | 11.4 | 24.6 | 30.7 | 37.3 | 7.5 |
| ○ *Clupeichthys goniognathus* | 13.6 | 36.4 | 27.2 | 22.8 | 46.9 | 26.3 | 15.4 | 11.4 | 29.8 | 32 | 30.7 | 7.5 |
| ○ *Clupeichthys aesarnensis* | 14 | 36 | 28.5 | 21.5 | 46.5 | 26.8 | 15.4 | 11.4 | 28.5 | 31.6 | 33.8 | 6.1 |
| ○ *Clupeichthys perakensis* | 14.9 | 35.1 | 28.9 | 21.1 | 46.9 | 26.3 | 15.4 | 11.4 | 25 | 34.2 | 34.6 | 6.1 |
| ○ *Clupeoides sp. Chao Phraya* | 14.5 | 33.8 | 31.1 | 20.6 | 45.6 | 28.1 | 14.9 | 11.4 | 29.8 | 27.6 | 39.5 | 3.1 |
| ○ *Clupeoides borneensis* | 12.3 | 36 | 31.1 | 20.6 | 46.5 | 27.2 | 14.9 | 11.4 | 21.1 | 35.5 | 39 | 4.4 |
| ○ *Sundasalanx praecox* | 14.9 | 34.6 | 28.1 | 22.4 | 47.4 | 26.8 | 14.9 | 11 | 27.2 | 30.3 | 36.4 | 6.1 |
| ○ *Sundasalanx sp. Chao Phraya* | 15.4 | 35.5 | 26.3 | 22.8 | 47.8 | 26.8 | 14.9 | 10.5 | 28.1 | 30.7 | 35.5 | 5.7 |
| ○ *Sundasalanx mekongensis* | 15.8 | 35.5 | 26.8 | 21.9 | 48.7 | 26.3 | 14.9 | 10.1 | 28.5 | 28.5 | 38.2 | 4.8 |
| □ *Ehirava fluviatilis* | 11.8 | 37.3 | 29.4 | 21.5 | 46.9 | 26.3 | 15.4 | 11.4 | 33.3 | 22.8 | 39 | 4.8 |
| □ *Gilchristella aestuaria* | 13.2 | 35.5 | 29.4 | 21.9 | 46.5 | 26.8 | 15.4 | 11.4 | 23.2 | 28.9 | 38.2 | 9.6 |
| □ *Clupeonella cultriventris* | 13.2 | 35.5 | 28.9 | 22.4 | 46.5 | 26.8 | 14.9 | 11.8 | 29.4 | 27.2 | 30.7 | 12.7 |
| ● *Clupea harengus* | 14.5 | 35.1 | 28.1 | 22.4 | 46.9 | 27.2 | 14 | 11.8 | 25.9 | 30.7 | 32.5 | 11 |
| ● *Clupea pallasii* | 14.9 | 34.6 | 28.1 | 22.4 | 46.9 | 27.2 | 14 | 11.8 | 25 | 31.6 | 32.9 | 10.5 |
| ● *Sprattus sprattus* | 14.5 | 35.1 | 28.1 | 22.4 | 46.9 | 27.2 | 14 | 11.8 | 31.1 | 26.8 | 29.8 | 12.3 |
| ● *Sprattus muelleri* | 14.5 | 35.5 | 27.2 | 22.8 | 46.9 | 27.6 | 13.6 | 11.8 | 24.6 | 32.9 | 32 | 10.5 |
| ● *Sprattus antipodum* | 14.5 | 35.5 | 27.2 | 22.8 | 46.9 | 27.6 | 13.6 | 11.8 | 23.7 | 33.8 | 32 | 10.5 |
| ○ *Potamalosa richmondia* | 12.7 | 36 | 28.9 | 22.4 | 46.5 | 27.6 | 14.5 | 11.4 | 27.6 | 30.3 | 34.2 | 7.9 |
| ● *Hyperlophus vittatus* | 14.5 | 33.8 | 29.4 | 22.4 | 46.5 | 27.6 | 14.5 | 11.4 | 32.5 | 27.6 | 29.8 | 10.1 |
| ● *Ethmidium maculatum* | 14.9 | 34.2 | 28.9 | 21.9 | 46.9 | 27.2 | 14.5 | 11.4 | 20.6 | 32.9 | 36.8 | 9.6 |
| ● *Jenkinsia lamprotaenia* | 14 | 33.3 | 29.4 | 23.2 | 46.5 | 26.8 | 14.9 | 11.8 | 21.1 | 43 | 21.1 | 14.9 |
| ● *Spratelloides delicatulus* | 12.7 | 34.2 | 27.6 | 25.4 | 46.9 | 27.2 | 14.5 | 11.4 | 20.6 | 42.1 | 19.7 | 17.5 |
| ● *Spratelloides gracilis* | 15.8 | 31.6 | 28.1 | 24.6 | 46.9 | 26.8 | 14.5 | 11.8 | 34.2 | 26.8 | 29.4 | 9.6 |
| ● *Etrumeus micropus* | 16.7 | 32.5 | 28.1 | 22.8 | 46.9 | 27.2 | 14 | 11.8 | 27.6 | 26.8 | 39.5 | 6.1 |
| ○ *Ilisha africana* | 15.8 | 34.2 | 29.8 | 20.2 | 46.5 | 27.2 | 15.4 | 11 | 22.4 | 33.3 | 40.8 | 3.5 |
| ○ *Pellona flavipinnis* | 16.2 | 32.5 | 31.1 | 20.2 | 45.2 | 28.9 | 14.5 | 11.4 | 14 | 38.6 | 41.2 | 6.1 |
| ○ *Ilisha elongata* | 15.4 | 33.3 | 31.6 | 19.7 | 46.1 | 27.6 | 15.8 | 10.5 | 24.6 | 32.5 | 38.2 | 4.8 |
| □ *Pellona ditchela* | 15.4 | 32.5 | 33.3 | 18.9 | 46.1 | 27.6 | 15.4 | 11 | 20.6 | 32.5 | 40.8 | 6.1 |
| □ *Anchoviella sp. LBP 2297* | 14.5 | 33.8 | 30.3 | 21.5 | 46.5 | 27.2 | 14.9 | 11.4 | 29.4 | 27.2 | 39.9 | 3.5 |
| □ *Lycengraulis grossidens* | 13.6 | 34.6 | 30.3 | 21.5 | 46.5 | 26.8 | 15.4 | 11.4 | 32 | 25.9 | 35.1 | 7 |
| ○ *Amazonsprattus scintilla* | 17.5 | 30.7 | 30.3 | 21.5 | 46.5 | 26.8 | 15.4 | 11.4 | 32.9 | 24.6 | 31.1 | 11.4 |
| ● *Engraulis encrasicolus* | 14 | 34.2 | 28.9 | 22.8 | 46.9 | 26.3 | 14.5 | 12.3 | 33.3 | 24.6 | 31.6 | 10.5 |
| ● *Engraulis japonicus* | 14.5 | 33.8 | 28.9 | 22.8 | 46.9 | 26.3 | 14.5 | 12.3 | 33.3 | 24.6 | 29.8 | 12.3 |
| ● *Stolephorus chinensis* | 15.8 | 31.6 | 31.1 | 21.5 | 46.9 | 26.8 | 14.9 | 11.4 | 33.3 | 20.2 | 35.1 | 11.4 |
| ● *Stolephorus waitei* | 15.8 | 31.6 | 31.6 | 21.1 | 46.9 | 26.8 | 14.9 | 11.4 | 32.9 | 20.6 | 37.3 | 9.2 |
| ○ *Lycothrissa crocodilus* | 15.8 | 32.5 | 32 | 19.7 | 47.4 | 26.8 | 14.5 | 11.4 | 20.2 | 30.3 | 44.7 | 4.8 |
| □ *Setipinna melanochir* | 12.7 | 36.4 | 30.3 | 20.6 | 46.5 | 27.2 | 15.4 | 11 | 17.5 | 36.4 | 38.6 | 7.5 |
| ● *Coilia reynaldi* | 16.7 | 31.6 | 30.7 | 21.1 | 46.5 | 26.8 | 14.9 | 11.8 | 19.3 | 28.1 | 45.6 | 7 |
| ● *Thryssa baelama* | 16.7 | 30.3 | 30.7 | 22.4 | 46.9 | 26.8 | 14.9 | 11.4 | 22.4 | 28.9 | 37.7 | 11 |
| ○ *Coilia lindmani* | 14.5 | 33.8 | 30.7 | 21.1 | 46.9 | 26.8 | 14.9 | 11.4 | 24.6 | 26.8 | 42.5 | 6.1 |
| □ *Coilia ectenes* | 15.8 | 32.9 | 30.7 | 20.6 | 46.5 | 27.2 | 14.9 | 11.4 | 28.1 | 25 | 42.5 | 4.4 |
| ● *Coilia nasus* | 16.2 | 32.5 | 30.7 | 20.6 | 46.5 | 27.2 | 14.9 | 11.4 | 29.8 | 24.1 | 42.5 | 3.5 |
| ● *Denticeps clupeoides* | 18.4 | 34.2 | 27.6 | 19.7 | 47.8 | 26.8 | 14 | 11.4 | 32.9 | 18.9 | 45.6 | 2.6 |

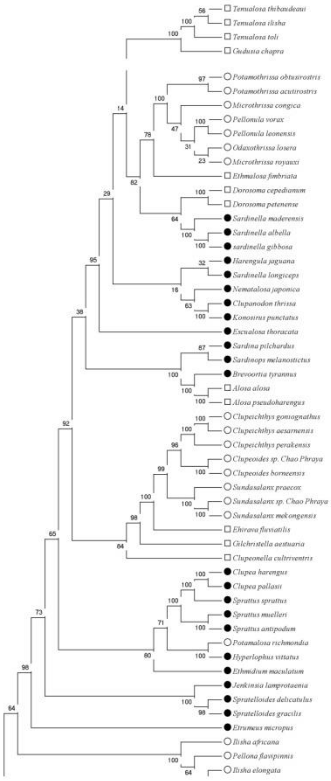
ATP6

| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 23.2 | 26.8 | 26.8 | 23.2 | 32.1 | 32.1 | 21.4 | 14.3 | 16.1 | 37.5 | 35.7 | 10.7 |
| □ *Tenualosa ilisha* | 25 | 28.6 | 25 | 21.4 | 32.1 | 32.1 | 23.2 | 12.5 | 23.2 | 30.4 | 35.7 | 10.7 |
| □ *Tenualosa toli* | 26.8 | 25 | 25 | 23.2 | 33.9 | 32.1 | 21.4 | 12.5 | 21.4 | 28.6 | 37.5 | 12.5 |
| □ *Gudusia chapra* | 26.8 | 26.8 | 28.6 | 17.9 | 32.1 | 32.1 | 21.4 | 14.3 | 19.6 | 32.1 | 42.9 | 5.4 |
| ○ *Potamothrissa obtusirostris* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 33.9 | 25 | 10.7 | 23.2 | 30.4 | 41.1 | 5.4 |
| ○ *Potamothrissa acutirostris* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 33.9 | 25 | 10.7 | 17.9 | 35.7 | 39.3 | 7.1 |
| ○ *Microthrissa congica* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 33.9 | 25 | 10.7 | 28.6 | 25 | 42.9 | 3.6 |
| ○ *Pellonula vorax* | 25 | 25 | 25 | 25 | 30.4 | 33.9 | 25 | 10.7 | 28.6 | 26.8 | 41.1 | 3.6 |
| ○ *Pellonula leonensis* | 25 | 25 | 25 | 25 | 30.4 | 33.9 | 25 | 10.7 | 28.6 | 25 | 42.9 | 3.6 |
| ○ *Odaxothrissa losera* | 26.8 | 23.2 | 28.6 | 21.4 | 30.4 | 33.9 | 25 | 10.7 | 21.4 | 32.1 | 39.3 | 7.1 |
| ○ *Microthrissa royauxi* | 25 | 25 | 26.8 | 23.2 | 30.4 | 33.9 | 25 | 10.7 | 23.2 | 32.1 | 39.3 | 5.4 |
| □ *Ethmalosa fimbriata* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 21.4 | 37.5 | 37.5 | 3.6 |
| □ *Dorosoma cepedianum* | 25 | 25 | 28.6 | 21.4 | 30.4 | 32.1 | 26.8 | 10.7 | 26.8 | 28.6 | 37.5 | 7.1 |
| □ *Dorosoma petenense* | 26.8 | 23.2 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 26.8 | 30.4 | 37.5 | 5.4 |
| ● *Sardinella maderensis* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 21.4 | 32.1 | 39.3 | 7.1 |
| ● *Sardinella albella* | 21.4 | 28.6 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 19.6 | 35.7 | 30.4 | 14.3 |
| ● *sardinella gibbosa* | 21.4 | 28.6 | 26.8 | 23.2 | 30.4 | 32.1 | 25 | 12.5 | 19.6 | 35.7 | 30.4 | 14.3 |
| ● *Harengula jaguana* | 30.4 | 23.2 | 26.8 | 19.6 | 30.4 | 33.9 | 25 | 10.7 | 17.9 | 35.7 | 25 | 21.4 |
| ● *Sardinella longiceps* | 23.2 | 26.8 | 30.4 | 19.6 | 30.4 | 32.1 | 25 | 12.5 | 26.8 | 30.4 | 37.5 | 5.4 |
| ● *Nematalosa japonica* | 23.2 | 28.6 | 25 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 12.5 | 39.3 | 33.9 | 14.3 |
| ● *Clupanodon thrissa* | 21.4 | 28.6 | 25 | 25 | 30.4 | 32.1 | 26.8 | 10.7 | 19.6 | 33.9 | 33.9 | 12.5 |
| ● *Konosirus punctatus* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 23.2 | 32.1 | 30.4 | 14.3 |
| ● *Escualosa thoracata* | 21.4 | 26.8 | 28.6 | 23.2 | 30.4 | 30.4 | 28.6 | 10.7 | 17.9 | 33.9 | 33.9 | 14.3 |
| ● *Sardina pilchardus* | 26.8 | 21.4 | 25 | 26.8 | 32.1 | 30.4 | 26.8 | 10.7 | 12.5 | 39.3 | 35.7 | 12.5 |
| ● *Sardinops melanostictus* | 25 | 26.8 | 26.8 | 21.4 | 30.4 | 33.9 | 25 | 10.7 | 21.4 | 33.9 | 33.9 | 10.7 |
| ● *Brevoortia tyrannus* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 21.4 | 33.9 | 41.1 | 3.6 |
| □ *Alosa alosa* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 30.4 | 25 | 41.1 | 3.6 |

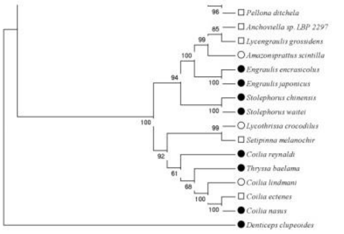


| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Alosa pseudoharengus* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 28.6 | 26.8 | 39.3 | 5.4 |
| ○ *Clupeichthys goniognathus* | 21.4 | 25 | 30.4 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 26.8 | 26.8 | 39.3 | 7.1 |
| ○ *Clupeichthys aesarnensis* | 23.2 | 23.2 | 30.4 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 26.8 | 26.8 | 37.5 | 8.9 |
| ○ *Clupeichthys perakensis* | 23.2 | 23.2 | 30.4 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 23.2 | 33.9 | 35.7 | 7.1 |
| ○ *Clupeoides sp. Chao Phraya* | 28.6 | 25 | 30.4 | 16.1 | 30.4 | 32.1 | 26.8 | 10.7 | 28.6 | 26.8 | 41.1 | 3.6 |
| ○ *Clupeoides borneensis* | 26.8 | 26.8 | 28.6 | 17.9 | 30.4 | 33.9 | 25 | 10.7 | 21.4 | 30.4 | 39.3 | 8.9 |
| ○ *Sundasalanx praecox* | 23.2 | 23.2 | 28.6 | 25 | 30.4 | 35.7 | 23.2 | 10.7 | 23.2 | 30.4 | 35.7 | 10.7 |
| ○ *Sundasalanx sp. Chao Phraya* | 25 | 21.4 | 32.1 | 21.4 | 30.4 | 33.9 | 25 | 10.7 | 32.1 | 23.2 | 39.3 | 5.4 |
| ○ *Sundasalanx mekongensis* | 23.2 | 23.2 | 35.7 | 17.9 | 30.4 | 35.7 | 23.2 | 10.7 | 23.2 | 32.1 | 37.5 | 7.1 |
| □ *Ehirava fluviatilis* | 23.2 | 26.8 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 23.2 | 30.4 | 39.3 | 7.1 |
| □ *Gilchristella aestuaria* | 23.2 | 25 | 28.6 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 28.6 | 33.9 | 28.6 | 8.9 |
| □ *Clupeonella cultriventris* | 26.8 | 23.2 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 25 | 32.1 | 33.9 | 8.9 |
| ● *Clupea harengus* | 25 | 23.2 | 30.4 | 21.4 | 30.4 | 30.4 | 28.6 | 10.7 | 25 | 33.9 | 32.1 | 8.9 |
| ● *Clupea pallasii* | 25 | 23.2 | 30.4 | 21.4 | 30.4 | 30.4 | 28.6 | 10.7 | 26.8 | 32.1 | 32.1 | 8.9 |
| ● *Sprattus sprattus* | 26.8 | 21.4 | 30.4 | 21.4 | 30.4 | 30.4 | 28.6 | 10.7 | 26.8 | 30.4 | 35.7 | 7.1 |
| ● *Sprattus muelleri* | 26.8 | 23.2 | 28.6 | 21.4 | 30.4 | 30.4 | 28.6 | 10.7 | 12.5 | 41.1 | 32.1 | 14.3 |
| ● *Sprattus antipodum* | 26.8 | 23.2 | 28.6 | 21.4 | 30.4 | 30.4 | 28.6 | 10.7 | 16.1 | 37.5 | 32.1 | 14.3 |
| ○ *Potamalosa richmondia* | 25 | 26.8 | 26.8 | 21.4 | 30.4 | 32.1 | 26.8 | 10.7 | 25 | 28.6 | 42.9 | 3.6 |
| ● *Hyperlophus vittatus* | 25 | 25 | 25 | 25 | 30.4 | 32.1 | 26.8 | 10.7 | 19.6 | 32.1 | 37.5 | 10.7 |
| ● *Ethmidium maculatum* | 25 | 25 | 26.8 | 23.2 | 30.4 | 32.1 | 26.8 | 10.7 | 30.4 | 25 | 39.3 | 5.4 |
| ● *Jenkinsia lamprotaenia* | 28.6 | 25 | 25 | 21.4 | 32.1 | 33.9 | 23.2 | 10.7 | 19.6 | 33.9 | 32.1 | 14.3 |
| ● *Spratelloides delicatulus* | 25 | 26.8 | 32.1 | 16.1 | 30.4 | 35.7 | 23.2 | 10.7 | 17.9 | 37.5 | 32.1 | 12.5 |
| ● *Spratelloides gracilis* | 32.1 | 23.2 | 23.2 | 21.4 | 30.4 | 32.1 | 26.8 | 10.7 | 21.4 | 33.9 | 35.7 | 8.9 |
| ● *Etrumeus micropus* | 28.6 | 21.4 | 26.8 | 23.2 | 30.4 | 33.9 | 25 | 10.7 | 21.4 | 28.6 | 41.1 | 8.9 |
| ○ *Ilisha africana* | 26.8 | 28.6 | 37.5 | 7.1 | 33.9 | 28.6 | 28.6 | 8.9 | 28.6 | 25 | 39.3 | 7.1 |
| ○ *Pellona flavipinnis* | 29.1 | 27.3 | 32.7 | 10.9 | 30.9 | 30.9 | 25.5 | 12.7 | 16.4 | 27.3 | 45.5 | 10.9 |
| ○ *Ilisha elongata* | 25 | 28.6 | 32.1 | 14.3 | 30.4 | 32.1 | 28.6 | 8.9 | 19.6 | 26.8 | 50 | 3.6 |
| □ *Pellona ditchela* | 30.4 | 23.2 | 33.9 | 12.5 | 30.4 | 32.1 | 28.6 | 8.9 | 12.5 | 32.1 | 50 | 5.4 |
| □ *Anchoviella sp. LBP 2297* | 26.8 | 28.6 | 26.8 | 17.9 | 33.9 | 30.4 | 25 | 10.7 | 35.7 | 19.6 | 42.9 | 1.8 |
| □ *Lycengraulis grossidens* | 26.8 | 28.6 | 28.6 | 16.1 | 33.9 | 30.4 | 25 | 10.7 | 26.8 | 28.6 | 42.9 | 1.8 |
| ○ *Amazonsprattus scintilla* | 28.6 | 26.8 | 28.6 | 16.1 | 33.9 | 30.4 | 25 | 10.7 | 30.4 | 23.2 | 39.3 | 7.1 |
| ● *Engraulis encrasicolus* | 30.4 | 25 | 26.8 | 17.9 | 33.9 | 30.4 | 25 | 10.7 | 26.8 | 26.8 | 39.3 | 7.1 |
| ● *Engraulis japonicus* | 30.4 | 25 | 26.8 | 17.9 | 33.9 | 30.4 | 25 | 10.7 | 26.8 | 25 | 37.5 | 10.7 |
| ● *Stolephorus chinensis* | 23.2 | 26.8 | 26.8 | 23.2 | 33.9 | 30.4 | 25 | 10.7 | 32.1 | 17.9 | 46.4 | 3.6 |
| ● *Stolephorus waitei* | 23.2 | 26.8 | 26.8 | 23.2 | 33.9 | 30.4 | 25 | 10.7 | 30.4 | 19.6 | 44.6 | 5.4 |
| ○ *Lycothrissa crocodilus* | 25 | 30.4 | 30.4 | 14.3 | 33.9 | 28.6 | 25 | 12.5 | 21.4 | 33.9 | 39.3 | 5.4 |
| □ *Setipinna melanochir* | 25 | 28.6 | 30.4 | 16.1 | 32.1 | 30.4 | 25 | 12.5 | 17.9 | 35.7 | 42.9 | 3.6 |
| ● *Coilia reynaldi* | 25 | 26.8 | 30.4 | 17.9 | 33.9 | 28.6 | 25 | 12.5 | 23.2 | 32.1 | 37.5 | 7.1 |
| ● *Thryssa baelama* | 30.4 | 25 | 28.6 | 16.1 | 35.7 | 26.8 | 25 | 12.5 | 30.4 | 19.6 | 37.5 | 12.5 |
| ○ *Coilia lindmani* | 26.8 | 26.8 | 28.6 | 17.9 | 33.9 | 28.6 | 26.8 | 10.7 | 25 | 28.6 | 42.9 | 3.6 |
| □ *Coilia ectenes* | 26.8 | 26.8 | 26.8 | 19.6 | 33.9 | 26.8 | 26.8 | 12.5 | 28.6 | 26.8 | 41.1 | 3.6 |
| ● *Coilia nasus* | 26.8 | 26.8 | 26.8 | 19.6 | 35.7 | 25 | 26.8 | 12.5 | 26.8 | 28.6 | 41.1 | 3.6 |
| ● *Denticeps clupeoides* | 30.4 | 23.2 | 33.9 | 12.5 | 30.4 | 33.9 | 25 | 10.7 | 32.1 | 17.9 | 46.4 | 3.6 |

ATP8

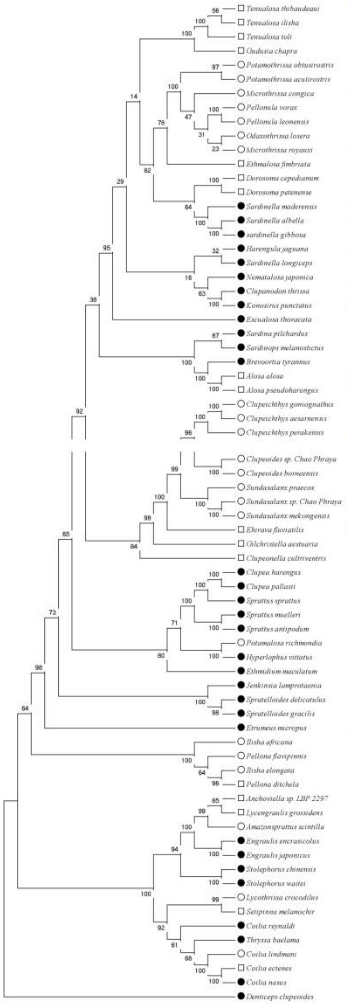


| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 22.7 | 22.7 | 23.8 | 30.8 | 40.5 | 26.4 | 18 | 15.1 | 23.1 | 35.1 | 25.6 | 16.3 |
| □ *Tenualosa ilisha* | 22.9 | 22.5 | 23.8 | 30.8 | 40.3 | 26.6 | 18 | 15.1 | 26.6 | 31.4 | 27.3 | 14.7 |
| □ *Tenualosa toli* | 22.7 | 22.7 | 24 | 30.6 | 40.3 | 26.6 | 18 | 15.1 | 21.1 | 36.6 | 26.7 | 15.5 |
| □ *Gudusia chapra* | 22.7 | 22.7 | 23.8 | 30.8 | 40.3 | 26.6 | 18 | 15.1 | 28.3 | 28.5 | 35.1 | 8.1 |
| ○ *Potamothrissa obtusirostris* | 21.5 | 22.9 | 24.2 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 27.3 | 28.9 | 35.5 | 8.3 |
| ○ *Potamothrissa acutirostris* | 21.5 | 22.9 | 24.2 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 26 | 29.8 | 35.5 | 8.7 |
| ○ *Microthrissa congica* | 20.5 | 23.8 | 24.2 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 24.4 | 31.8 | 33.9 | 9.9 |
| ○ *Pellonula vorax* | 21.7 | 22.9 | 24.2 | 31.2 | 40.5 | 26.2 | 18 | 15.3 | 28.5 | 26.7 | 33.3 | 11.4 |
| ○ *Pellonula leonensis* | 22.1 | 22.5 | 24 | 31.4 | 40.7 | 26.2 | 18 | 15.1 | 27.5 | 28.5 | 33.9 | 10.1 |
| ○ *Odaxothrissa losera* | 21.3 | 23.1 | 24.4 | 31.2 | 40.9 | 26.2 | 18 | 14.9 | 25.8 | 30.2 | 35.1 | 8.9 |
| ○ *Microthrissa royauxi* | 21.3 | 23.1 | 24.2 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 25.8 | 30.4 | 35.3 | 8.5 |
| □ *Ethmalosa fimbriata* | 20.5 | 23.8 | 24.2 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 26.9 | 27.5 | 34.3 | 11.2 |
| □ *Dorosoma cepedianum* | 20.9 | 23.6 | 24 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 22.9 | 33.1 | 29.8 | 14.1 |
| □ *Dorosoma petenense* | 21.5 | 23.1 | 24.2 | 31.2 | 40.7 | 26.4 | 18 | 14.9 | 23.6 | 32.4 | 32.2 | 11.8 |
| ● *Sardinella maderensis* | 20.5 | 23.8 | 24.4 | 31.2 | 40.7 | 26.4 | 18 | 14.9 | 22.9 | 32.2 | 33.1 | 11.8 |
| ● *Sardinella albella* | 20 | 24.2 | 24.4 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 21.9 | 32.9 | 31.2 | 14 |
| ● *sardinella gibbosa* | 20 | 24.2 | 24.6 | 31.2 | 40.7 | 26.4 | 18 | 14.9 | 20.5 | 33.9 | 32 | 13.6 |
| ● *Harengula jaguana* | 20.9 | 23.4 | 24 | 31.6 | 40.5 | 26.4 | 18.2 | 14.9 | 24.8 | 32.9 | 24.6 | 17.6 |
| ● *Sardinella longiceps* | 21.7 | 22.7 | 24.2 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 23.8 | 31.2 | 33.5 | 11.4 |
| ● *Nematalosa japonica* | 21.5 | 23.3 | 24.2 | 31 | 40.7 | 26.4 | 18 | 14.9 | 24.8 | 28.5 | 32 | 14.7 |
| ● *Clupanodon thrissa* | 21.5 | 23.3 | 24.2 | 31 | 40.7 | 26.4 | 18 | 14.9 | 21.7 | 32.8 | 32 | 13.6 |
| ● *Konosirus punctatus* | 21.9 | 22.9 | 24.2 | 31 | 40.7 | 26.4 | 18 | 14.9 | 25 | 30.6 | 27.1 | 17.2 |
| ● *Escualosa thoracata* | 20.3 | 24 | 24.2 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 22.9 | 33.7 | 26.9 | 16.5 |
| ● *Sardina pilchardus* | 21.1 | 23.3 | 24.4 | 31.2 | 40.5 | 26.2 | 18.4 | 14.9 | 26.4 | 29.8 | 22.3 | 21.5 |
| ● *Sardinops melanostictus* | 21.3 | 22.9 | 24.4 | 31.4 | 40.7 | 26.4 | 18 | 14.9 | 26 | 30 | 27.7 | 16.3 |
| ● *Brevoortia tyrannus* | 22.1 | 22.8 | 24.4 | 30.8 | 40.8 | 26.3 | 17.8 | 15.1 | 21.7 | 31.5 | 31.9 | 14.9 |
| □ *Alosa alosa* | 21.5 | 23.1 | 24.4 | 31 | 40.7 | 26.4 | 18 | 14.9 | 21.7 | 32.2 | 30.6 | 15.5 |
| □ *Alosa pseudoharengus* | 21.7 | 22.9 | 24.4 | 31 | 40.7 | 26.4 | 18 | 14.9 | 21.9 | 31.2 | 32.2 | 14.7 |
| ○ *Clupeichthys goniognathus* | 21.1 | 22.7 | 24.8 | 31.4 | 40.5 | 26.4 | 18.2 | 14.9 | 26.9 | 28.3 | 35.1 | 9.7 |
| ○ *Clupeichthys aesarnensis* | 21.5 | 22.3 | 24.8 | 31.4 | 40.5 | 26.4 | 18.2 | 14.9 | 26.6 | 28.3 | 33.5 | 11.6 |
| ○ *Clupeichthys perakensis* | 22.1 | 21.7 | 24.8 | 31.4 | 40.5 | 26.4 | 18.2 | 14.9 | 27.9 | 28.1 | 35.1 | 8.9 |
| ○ *Clupeoides sp. Chao Phraya* | 21.3 | 22.5 | 24.8 | 31.4 | 40.7 | 26.2 | 18 | 15.1 | 28.1 | 25 | 42.1 | 4.8 |
| ○ *Clupeoides borneensis* | 21.1 | 22.5 | 24.6 | 31.8 | 40.7 | 26.4 | 18 | 14.9 | 25.4 | 29.5 | 35.3 | 9.9 |
| ○ *Sundasalanx praecox* | 21.7 | 22.1 | 25.4 | 30.8 | 40.7 | 26.7 | 18 | 14.5 | 26.9 | 29.7 | 33.3 | 10.1 |
| ○ *Sundasalanx sp. Chao Phraya* | 21.1 | 23.1 | 24.8 | 31 | 40.7 | 26.7 | 18 | 14.5 | 26.2 | 31.2 | 33.7 | 8.9 |
| ○ *Sundasalanx mekongensis* | 21.3 | 22.9 | 24.8 | 31 | 40.7 | 26.7 | 18 | 14.5 | 24.2 | 32.8 | 32.9 | 10.1 |
| □ *Ehirava fluviatilis* | 20.5 | 22.9 | 25 | 31.6 | 40.7 | 26.6 | 18 | 14.7 | 30.4 | 24.4 | 38.2 | 7 |
| □ *Gilchristella aestuaria* | 21.1 | 22.3 | 24.8 | 31.8 | 40.7 | 26.6 | 18 | 14.7 | 23.3 | 29.5 | 36.6 | 10.7 |
| □ *Clupeonella cultriventris* | 20.7 | 23.6 | 24.6 | 31 | 40.7 | 26.4 | 18 | 14.9 | 28.1 | 29.3 | 30.4 | 12.2 |
| ● *Clupea harengus* | 20.5 | 23.8 | 24.8 | 30.8 | 40.7 | 26.4 | 18 | 14.9 | 24.6 | 29.8 | 32.8 | 12.8 |
| ● *Clupea pallasii* | 20.5 | 23.8 | 24.8 | 30.8 | 40.7 | 26.4 | 18 | 14.9 | 25.4 | 29.7 | 32 | 13 |
| ● *Sprattus sprattus* | 21.3 | 23.1 | 24.4 | 31.2 | 40.7 | 26.4 | 18 | 14.9 | 27.1 | 28.3 | 32.2 | 12.4 |
| ● *Sprattus muelleri* | 20.7 | 23.8 | 24.8 | 30.6 | 40.7 | 26.4 | 18 | 14.9 | 26.9 | 27.7 | 31.2 | 14.1 |
| ● *Sprattus antipodum* | 20.7 | 23.8 | 24.8 | 30.6 | 40.7 | 26.4 | 18 | 14.9 | 27.1 | 27.3 | 30.8 | 14.7 |
| ○ *Potamalosa richmondia* | 21.9 | 21.9 | 24.6 | 31.6 | 40.7 | 26.4 | 18 | 14.9 | 26.9 | 30.4 | 32.9 | 9.7 |
| ● *Hyperlophus vittatus* | 21.3 | 22.5 | 24.6 | 31.6 | 40.7 | 26.4 | 18 | 14.9 | 27.9 | 28.7 | 33.5 | 9.9 |
| ● *Ethmidium maculatum* | 21.1 | 23.3 | 24.6 | 31 | 40.7 | 26.4 | 18 | 14.9 | 23.4 | 30.8 | 33.9 | 11.8 |
| ● *Jenkinsia lamprotaenia* | 20.9 | 23.1 | 23.8 | 32.2 | 40.7 | 26.4 | 18 | 14.9 | 26 | 32.4 | 26.4 | 15.3 |
| ● *Spratelloides delicatulus* | 20 | 23.8 | 24.2 | 32 | 40.7 | 26.4 | 18 | 14.9 | 21.7 | 38.4 | 22.7 | 17.2 |
| ● *Spratelloides gracilis* | 21.7 | 22.1 | 24 | 32.2 | 40.7 | 26.4 | 18 | 14.9 | 28.9 | 28.3 | 29.1 | 13.8 |
| ● *Etrumeus micropus* | 22.5 | 21.9 | 24 | 31.6 | 40.7 | 26.4 | 18 | 14.9 | 29.7 | 27.9 | 32.2 | 10.3 |
| ○ *Ilisha africana* | 22.9 | 21.1 | 24.8 | 31.2 | 40.9 | 26.2 | 18 | 14.9 | 27.9 | 27.7 | 36.4 | 7.9 |
| ○ *Pellona flavipinnis* | 22.1 | 21.9 | 24.4 | 31.6 | 40.7 | 26.4 | 18 | 14.9 | 20.5 | 30.8 | 36.8 | 11.8 |
| ○ *Ilisha elongata* | 22.1 | 22.1 | 23.6 | 32.2 | 40.7 | 26.4 | 18 | 14.9 | 23.3 | 32.6 | 36.6 | 7.6 |



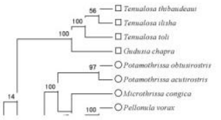
| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Pellona ditchela* | 22.5 | 21.5 | 23.8 | 32.2 | 40.7 | 26.4 | 18 | 14.9 | 24 | 31.2 | 36.4 | 8.3 |
| □ *Anchoviella sp. LBP 2297* | 22 | 21.4 | 24.7 | 31.9 | 40.9 | 26.3 | 17.9 | 15 | 24.5 | 29.2 | 36.4 | 9.9 |
| □ *Lycengraulis grossidens* | 23.2 | 20.6 | 24.7 | 31.5 | 40.9 | 26.3 | 17.9 | 15 | 28.6 | 24.9 | 38.7 | 7.8 |
| ○ *Amazonsprattus scintilla* | 22.8 | 20.8 | 24.5 | 31.9 | 40.9 | 26.3 | 17.9 | 15 | 31.1 | 23.3 | 33.5 | 12.1 |
| ● *Engraulis encrasicolus* | 21 | 22.6 | 24.9 | 31.5 | 40.9 | 26.1 | 17.9 | 15.2 | 28.2 | 27.4 | 34.6 | 9.7 |
| ● *Engraulis japonicus* | 20.6 | 22.8 | 24.9 | 31.7 | 40.9 | 26.1 | 17.9 | 15.2 | 26.7 | 29.4 | 33.3 | 10.7 |
| ● *Stolephorus chinensis* | 22.4 | 21 | 24.7 | 31.9 | 40.7 | 26.3 | 18.1 | 15 | 31.1 | 24.9 | 36.4 | 7.6 |
| ● *Stolephorus waitei* | 22.4 | 21 | 24.7 | 31.9 | 40.7 | 26.3 | 18.1 | 15 | 29.2 | 26.1 | 36.4 | 8.4 |
| ○ *Lycothrissa crocodilus* | 21.4 | 22.4 | 24.9 | 31.3 | 40.5 | 26.5 | 17.9 | 15.2 | 23.9 | 32.5 | 36.6 | 7 |
| □ *Setipinna melanochir* | 22.4 | 21.4 | 23.9 | 32.3 | 40.9 | 26.3 | 17.9 | 15 | 21.2 | 34.8 | 31.5 | 12.5 |
| ● *Coilia reynaldi* | 23.2 | 20.4 | 24.5 | 31.9 | 40.7 | 26.5 | 17.9 | 15 | 25.5 | 29 | 33.9 | 11.7 |
| ● *Thryssa baelama* | 22.4 | 21.2 | 24.7 | 31.7 | 40.9 | 26.1 | 18.1 | 15 | 23.7 | 29 | 35 | 12.3 |
| ○ *Coilia lindmani* | 22.8 | 20.4 | 24.9 | 31.9 | 40.9 | 26.1 | 18.1 | 15 | 23.9 | 28.6 | 42.4 | 5.1 |
| □ *Coilia ectenes* | 23 | 20.2 | 24.9 | 31.9 | 40.9 | 25.9 | 18.1 | 15.2 | 24.7 | 28 | 41.2 | 6 |
| ● *Coilia nasus* | 22.6 | 20.6 | 24.7 | 32.1 | 40.9 | 25.9 | 18.1 | 15.2 | 24.1 | 29 | 40.9 | 6 |
| ● *Denticeps clupeoides* | 23.6 | 20.9 | 25 | 30.4 | 40.7 | 26.2 | 17.8 | 15.3 | 32.8 | 24.6 | 37.4 | 5.2 |

CO1

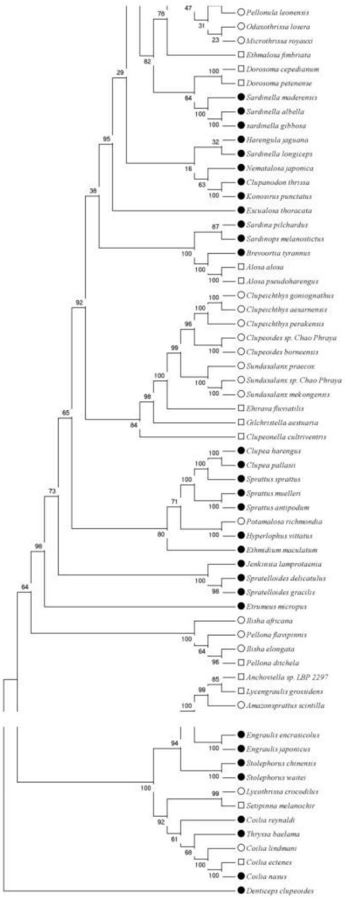


| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 20 | 25.7 | 22.2 | 32.2 | 38.3 | 23 | 27.8 | 10.9 | 19.6 | 33.5 | 33 | 13.9 |
| □ *Tenualosa ilisha* | 17.8 | 27.8 | 22.2 | 32.2 | 38.3 | 23 | 27.8 | 10.9 | 20.9 | 33.9 | 33.5 | 11.7 |
| □ *Tenualosa toli* | 19.1 | 27 | 22.2 | 31.7 | 38.3 | 23 | 27.8 | 10.9 | 20 | 33.9 | 31.7 | 14.3 |
| □ *Gudusia chapra* | 20 | 25.7 | 23 | 31.3 | 38.3 | 23 | 27.8 | 10.9 | 23 | 32.6 | 37.8 | 6.5 |
| ○ *Potamothrissa obtusirostris* | 19.6 | 25.7 | 22.6 | 32.2 | 38.7 | 22.2 | 27.8 | 11.3 | 22.2 | 33 | 39.6 | 5.2 |
| ○ *Potamothrissa acutirostris* | 19.1 | 26.1 | 21.7 | 33 | 38.7 | 22.2 | 27.8 | 11.3 | 21.7 | 32.6 | 40.4 | 5.2 |
| ○ *Microthrissa congica* | 18.3 | 27 | 21.7 | 33 | 38.7 | 22.2 | 27.8 | 11.3 | 22.6 | 34.3 | 38.3 | 4.8 |
| ○ *Pellonula vorax* | 18.7 | 26.5 | 21.3 | 33.5 | 38.7 | 22.2 | 27.8 | 11.3 | 24.3 | 32.6 | 38.7 | 4.3 |
| ○ *Pellonula leonensis* | 18.7 | 26.5 | 21.3 | 33.5 | 38.3 | 22.6 | 27.8 | 11.3 | 26.1 | 30.9 | 39.6 | 3.5 |
| ○ *Odaxothrissa losera* | 19.1 | 26.1 | 21.3 | 33.5 | 38.7 | 22.2 | 27.8 | 11.3 | 23.5 | 33.5 | 39.1 | 3.9 |
| ○ *Microthrissa royauxi* | 17.8 | 27.4 | 22.6 | 32.2 | 38.3 | 22.2 | 28.3 | 11.3 | 21.3 | 34.3 | 36.5 | 7.8 |
| □ *Ethmalosa fimbriata* | 19.1 | 26.1 | 22.2 | 32.6 | 38.3 | 23 | 27.8 | 10.9 | 23.5 | 33.9 | 35.7 | 7 |
| □ *Dorosoma cepedianum* | 18.3 | 27 | 22.2 | 32.6 | 38.3 | 23 | 27.8 | 10.9 | 19.6 | 37.4 | 33 | 10 |
| □ *Dorosoma petenense* | 18.7 | 26.5 | 22.2 | 32.6 | 38.7 | 22.6 | 27.8 | 10.9 | 22.6 | 33.5 | 36.5 | 7.4 |
| ● *Sardinella maderensis* | 18.3 | 27 | 22.2 | 32.6 | 38.3 | 23 | 27.8 | 10.9 | 20.4 | 33.5 | 35.2 | 10.9 |
| ● *Sardinella albella* | 17.4 | 27.8 | 22.2 | 32.6 | 38.3 | 23 | 27.8 | 10.9 | 19.6 | 34.8 | 33.9 | 11.7 |
| ● *sardinella gibbosa* | 17.8 | 27.4 | 22.2 | 32.6 | 38.3 | 23 | 27.8 | 10.9 | 19.6 | 34.8 | 33.9 | 11.7 |
| ● *Harengula jaguana* | 17.4 | 27.8 | 22.6 | 32.2 | 38.3 | 23 | 27.4 | 11.3 | 23.5 | 35.2 | 26.1 | 15.2 |
| ● *Sardinella longiceps* | 18.3 | 27 | 22.2 | 32.6 | 38.3 | 23 | 27.8 | 10.9 | 22.2 | 34.3 | 33.5 | 10 |
| ● *Nematalosa japonica* | 18.7 | 26.5 | 22.6 | 32.2 | 38.3 | 23 | 27.8 | 10.9 | 22.6 | 35.7 | 31.7 | 10 |
| ● *Clupanodon thrissa* | 18.3 | 27 | 22.6 | 32.2 | 38.3 | 23 | 27.8 | 10.9 | 20.9 | 35.7 | 32.6 | 10.9 |
| ● *Konosirus punctatus* | 17.8 | 27.4 | 22.6 | 32.2 | 38.3 | 23 | 27.8 | 10.9 | 20.4 | 34.8 | 31.3 | 13.5 |
| ● *Escualosa thoracata* | 18.7 | 26.5 | 22.2 | 32.6 | 37.8 | 23 | 27.8 | 11.3 | 26.5 | 31.3 | 29.1 | 13 |
| ● *Sardina pilchardus* | 18.7 | 26.1 | 23 | 32.2 | 37.8 | 24.3 | 27 | 10.9 | 26.1 | 33.5 | 27 | 13.5 |
| ● *Sardinops melanostictus* | 18.7 | 27.4 | 21.3 | 32.6 | 38.3 | 23 | 27.8 | 10.9 | 22.6 | 33.5 | 31.3 | 12.6 |
| ● *Brevoortia tyrannus* | 17.8 | 27.8 | 21.3 | 33 | 38.3 | 23 | 27.8 | 10.9 | 23 | 32.2 | 33.5 | 11.3 |
| □ *Alosa alosa* | 16.5 | 29.1 | 21.3 | 33 | 38.3 | 23 | 27.8 | 10.9 | 21.3 | 33 | 37.4 | 8.3 |
| □ *Alosa pseudoharengus* | 17 | 28.7 | 21.3 | 33 | 38.3 | 23 | 27.8 | 10.9 | 25.2 | 28.7 | 36.5 | 9.6 |
| ○ *Clupeichthys goniognathus* | 18.7 | 26.1 | 22.2 | 33 | 37.8 | 23 | 27.8 | 11.3 | 23.9 | 34.8 | 32.6 | 8.7 |
| ○ *Clupeichthys aesarnensis* | 18.7 | 26.1 | 22.2 | 33 | 37.8 | 23 | 27.8 | 11.3 | 24.3 | 33.5 | 33.5 | 8.7 |
| ○ *Clupeichthys perakensis* | 18.7 | 26.1 | 22.2 | 33 | 37.8 | 23 | 27.8 | 11.3 | 27.8 | 30.4 | 34.3 | 7.4 |
| ○ *Clupeoides sp. Chao Phraya* | 18.7 | 25.7 | 23.9 | 31.7 | 37.8 | 23 | 27.8 | 11.3 | 26.1 | 29.6 | 40.9 | 3.5 |
| ○ *Clupeoides borneensis* | 17.4 | 27 | 23 | 32.6 | 37.8 | 23 | 27.8 | 11.3 | 22.2 | 33 | 40.4 | 4.3 |
| ○ *Sundasalanx praecox* | 18.3 | 25.7 | 23.5 | 32.6 | 37.4 | 23 | 28.3 | 11.3 | 27 | 32.2 | 33 | 7.8 |
| ○ *Sundasalanx sp. Chao Phraya* | 18.3 | 25.7 | 23.9 | 32.2 | 37.4 | 23 | 28.3 | 11.3 | 30 | 29.1 | 33.9 | 7 |
| ○ *Sundasalanx mekongensis* | 17.4 | 26.5 | 24.3 | 31.7 | 37.4 | 23 | 28.3 | 11.3 | 23.9 | 33 | 38.7 | 4.3 |
| □ *Ehirava fluviatilis* | 16.6 | 27.5 | 23.6 | 32.3 | 37.8 | 23 | 27.8 | 11.3 | 27.4 | 30 | 40.4 | 2.2 |
| □ *Gilchristella aestuaria* | 18.7 | 25.7 | 23.5 | 32.2 | 37.8 | 23.5 | 27.8 | 10.9 | 26.1 | 32.2 | 36.1 | 5.7 |
| □ *Clupeonella cultriventris* | 18.3 | 26.5 | 22.6 | 32.6 | 38.3 | 23 | 27.8 | 10.9 | 29.1 | 30 | 32.2 | 8.7 |
| ● *Clupea harengus* | 18.7 | 27 | 22.2 | 32.2 | 39.1 | 22.6 | 27.4 | 10.9 | 22.2 | 31.3 | 34.8 | 11.7 |
| ● *Clupea pallasii* | 18.7 | 27 | 22.2 | 32.2 | 39.1 | 22.6 | 27.4 | 10.9 | 22.2 | 31.3 | 35.2 | 11.3 |
| ● *Sprattus sprattus* | 18.3 | 27.8 | 21.7 | 32.2 | 39.1 | 22.6 | 27.4 | 10.9 | 24.8 | 30.9 | 33 | 11.3 |
| ● *Sprattus muelleri* | 18.3 | 27.4 | 22.6 | 31.7 | 39.1 | 22.6 | 27.4 | 10.9 | 23.9 | 33.5 | 27.8 | 14.8 |
| ● *Sprattus antipodum* | 18.3 | 27.4 | 22.6 | 31.7 | 39.1 | 22.6 | 27.4 | 10.9 | 24.3 | 33.5 | 26.5 | 15.7 |
| ○ *Potamalosa richmondia* | 18.3 | 27.4 | 22.6 | 31.7 | 38.3 | 23 | 27.8 | 10.9 | 23 | 32.2 | 38.3 | 6.5 |
| ● *Hyperlophus vittatus* | 18.7 | 27 | 22.6 | 31.7 | 38.3 | 23 | 27.8 | 10.9 | 24.3 | 33 | 33.5 | 9.1 |
| ● *Ethmidium maculatum* | 18.3 | 27.8 | 23.5 | 30.4 | 37.8 | 23.5 | 27.8 | 10.9 | 20.4 | 35.7 | 35.2 | 8.7 |
| ● *Jenkinsia lamprotaenia* | 18.3 | 27.4 | 22.6 | 31.7 | 37.8 | 23.5 | 27.8 | 10.9 | 21.7 | 37.8 | 27.8 | 12.6 |
| ● *Spratelloides delicatulus* | 17.4 | 28.7 | 22.6 | 31.3 | 37.8 | 23.5 | 27.8 | 10.9 | 20.9 | 40.4 | 24.3 | 14.3 |
| ● *Spratelloides gracilis* | 17.8 | 27.4 | 22.6 | 32.2 | 37.8 | 23.5 | 27.8 | 10.9 | 24.3 | 31.7 | 30.9 | 13 |
| ● *Etrumeus micropus* | 18.3 | 27.4 | 22.6 | 31.7 | 37.8 | 23.5 | 27.8 | 10.9 | 28.3 | 28.7 | 37.4 | 5.7 |
| ○ *Ilisha africana* | 20.9 | 25.2 | 23.9 | 30 | 38.3 | 23.5 | 27.4 | 10.9 | 25.7 | 33.9 | 34.8 | 5.7 |
| ○ *Pellona flavipinnis* | 18.3 | 27.4 | 23.9 | 30.4 | 37.8 | 23.9 | 27.4 | 10.9 | 16.1 | 40.4 | 36.1 | 7.4 |
| ○ *Ilisha elongata* | 20 | 26.1 | 23.9 | 30 | 37.8 | 23.9 | 27.4 | 10.9 | 25.7 | 32.2 | 37 | 5.2 |
| □ *Pellona ditchela* | 20.4 | 25.7 | 23.5 | 30.4 | 37.8 | 23.9 | 27.4 | 10.9 | 28.3 | 30.4 | 34.3 | 7 |
| □ *Anchoviella sp. LBP 2297* | 18.7 | 25.2 | 23.5 | 32.6 | 38.7 | 22.6 | 27.8 | 10.9 | 23.5 | 30.9 | 39.1 | 6.5 |
| □ *Lycengraulis grossidens* | 19.1 | 25.2 | 24.3 | 31.3 | 38.3 | 23 | 27.8 | 10.9 | 26.5 | 30.4 | 37.4 | 5.7 |
| ○ *Amazonsprattus scintilla* | 20.4 | 23.5 | 23.5 | 32.6 | 38.3 | 23 | 27.8 | 10.9 | 26.1 | 29.6 | 34.3 | 10 |
| ● *Engraulis encrasicolus* | 18.7 | 24.8 | 23.5 | 33 | 38.7 | 22.6 | 27.8 | 10.9 | 27 | 28.7 | 33.5 | 10.9 |
| ● *Engraulis japonicus* | 18.7 | 24.8 | 23.5 | 33 | 38.7 | 22.6 | 27.8 | 10.9 | 27.4 | 28.7 | 33.5 | 10.4 |
| ● *Stolephorus chinensis* | 20 | 24.3 | 23.5 | 32.2 | 37.8 | 22.6 | 28.7 | 10.9 | 33.9 | 24.3 | 33.9 | 7.8 |
| ● *Stolephorus waitei* | 20 | 24.3 | 23.5 | 32.2 | 37.8 | 22.6 | 28.7 | 10.9 | 33 | 25.7 | 33.5 | 7.8 |
| ○ *Lycothrissa crocodilus* | 19.1 | 24.8 | 24.8 | 31.3 | 38.7 | 23 | 27.4 | 10.9 | 23 | 35.7 | 36.5 | 4.8 |
| □ *Setipinna melanochir* | 19.1 | 25.2 | 24.8 | 30.9 | 38.3 | 23.5 | 27.8 | 10.4 | 18.7 | 40 | 35.2 | 6.1 |
| ● *Coilia reynaldi* | 19.6 | 24.3 | 24.3 | 31.7 | 37.8 | 23.5 | 27.8 | 10.9 | 26.5 | 30.4 | 35.2 | 7.8 |
| ● *Thryssa baelama* | 19.6 | 24.3 | 24.3 | 31.7 | 37.8 | 23.5 | 27.8 | 10.9 | 26.5 | 30 | 33 | 10.4 |
| ○ *Coilia lindmani* | 19.1 | 24.8 | 23.9 | 32.2 | 37.8 | 23.5 | 27.8 | 10.9 | 22.6 | 33.9 | 37.4 | 6.1 |
| □ *Coilia ectenes* | 19.6 | 24.3 | 24.3 | 31.7 | 37.8 | 23.5 | 27.8 | 10.9 | 23.9 | 32.2 | 39.1 | 4.8 |
| ● *Coilia nasus* | 19.6 | 24.3 | 24.3 | 31.7 | 37.8 | 23.5 | 27.8 | 10.9 | 24.8 | 31.3 | 39.1 | 4.8 |
| ● *Denticeps clupeoides* | 19.6 | 25.7 | 23.9 | 30.9 | 37.8 | 23.5 | 27.8 | 10.9 | 27.4 | 30.9 | 38.7 | 3 |

CO2

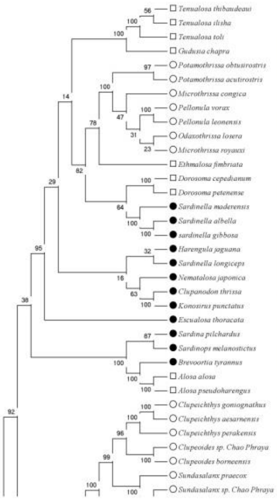


| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 25.3 | 26.1 | 19.9 | 28.7 | 36.4 | 26.4 | 19.9 | 17.2 | 23.8 | 33.3 | 28.7 | 14.2 |
| □ *Tenualosa ilisha* | 25.3 | 26.1 | 19.5 | 29.1 | 36.4 | 26.4 | 19.9 | 17.2 | 20.3 | 37.5 | 30.3 | 11.9 |
| □ *Tenualosa toli* | 25.3 | 26.1 | 20.3 | 28.4 | 36.4 | 26.4 | 19.9 | 17.2 | 23.8 | 34.9 | 31.8 | 9.6 |
| □ *Gudusia chapra* | 25.3 | 26.1 | 20.7 | 28 | 36 | 26.8 | 20.3 | 16.9 | 14.9 | 42.1 | 38.3 | 4.6 |
| ○ *Potamothrissa obtusirostris* | 24.9 | 26.1 | 20.3 | 28.7 | 36 | 26.4 | 20.7 | 16.9 | 23 | 34.5 | 36.8 | 5.7 |
| ○ *Potamothrissa acutirostris* | 25.7 | 25.3 | 20.3 | 28.7 | 36 | 26.4 | 20.7 | 16.9 | 24.1 | 33.3 | 35.6 | 6.9 |
| ○ *Microthrissa congica* | 24.9 | 26.1 | 20.3 | 28.7 | 36 | 26.1 | 20.7 | 17.2 | 18.8 | 38.3 | 38.7 | 4.2 |
| ○ *Pellonula vorax* | 25.7 | 25.7 | 19.9 | 28.7 | 36 | 26.1 | 20.7 | 17.2 | 19.9 | 38.7 | 35.2 | 6.1 |

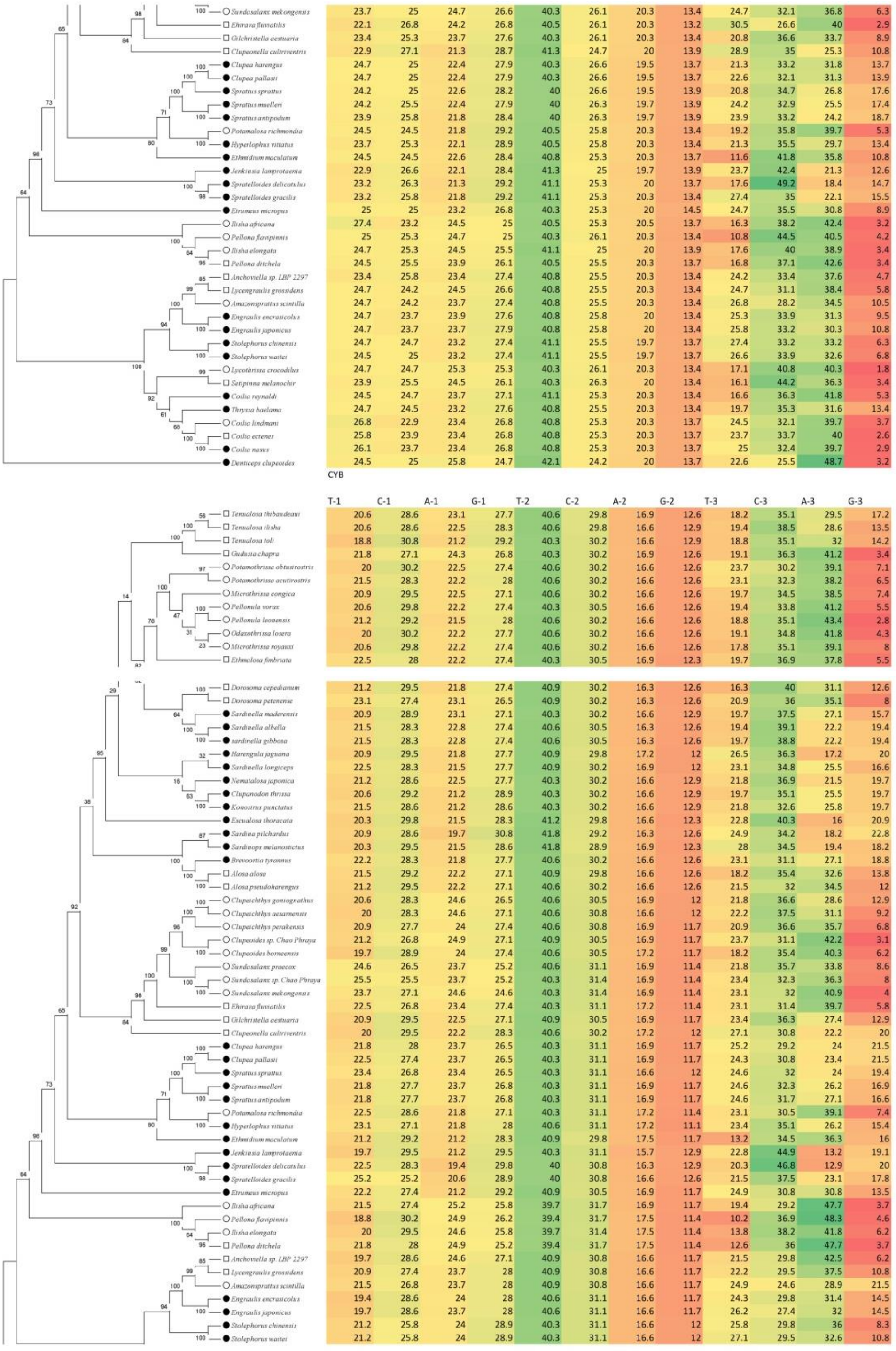


| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ○ *Pellonula leonensis* | 24.5 | 26.4 | 20.3 | 28.7 | 36 | 26.1 | 20.7 | 17.2 | 20.7 | 37.5 | 36.4 | 5.4 |
| ○ *Odaxothrissa losera* | 24.9 | 25.7 | 20.3 | 29.1 | 36 | 26.1 | 20.7 | 17.2 | 21.5 | 36 | 37.9 | 4.6 |
| ○ *Microthrissa royauxi* | 25.3 | 25.7 | 20.3 | 28.7 | 35.6 | 26.8 | 20.7 | 16.9 | 19.9 | 39.8 | 34.5 | 5.7 |
| □ *Ethmalosa fimbriata* | 24.5 | 26.8 | 19.9 | 28.7 | 36 | 26.8 | 20.3 | 16.9 | 24.9 | 35.2 | 32.2 | 7.7 |
| □ *Dorosoma cepedianum* | 24.9 | 27.2 | 18.8 | 29.1 | 36.4 | 26.4 | 20.3 | 16.9 | 18.8 | 41 | 31.8 | 8.4 |
| □ *Dorosoma petenense* | 26.1 | 26.1 | 19.9 | 28 | 36.4 | 26.4 | 20.3 | 16.9 | 19.9 | 39.5 | 31.8 | 8.8 |
| ● *Sardinella maderensis* | 24.5 | 27.2 | 19.5 | 28.7 | 36 | 26.8 | 20.3 | 16.9 | 15.7 | 42.1 | 32.6 | 9.6 |
| ● *Sardinella albella* | 24.1 | 27.6 | 19.2 | 29.1 | 36 | 26.8 | 20.3 | 16.9 | 16.1 | 41.8 | 29.9 | 12.3 |
| ● *sardinella gibbosa* | 24.1 | 27.6 | 19.2 | 29.1 | 36 | 26.8 | 20.3 | 16.9 | 17.2 | 39.8 | 29.9 | 13 |
| ● *Harengula jaguana* | 24.9 | 26.4 | 18.8 | 29.9 | 36 | 26.8 | 19.9 | 17.2 | 18 | 42.9 | 27.2 | 11.9 |
| ● *Sardinella longiceps* | 25.3 | 26.4 | 19.2 | 29.1 | 37.2 | 26.1 | 19.9 | 16.9 | 16.9 | 40.2 | 33 | 10 |
| ● *Nematalosa japonica* | 24.1 | 27.6 | 18.4 | 29.9 | 36 | 26.8 | 20.3 | 16.9 | 17.2 | 42.9 | 24.9 | 14.9 |
| ● *Clupanodon thrissa* | 24.9 | 26.4 | 19.9 | 28.7 | 36.4 | 26.4 | 20.3 | 16.9 | 21.1 | 39.5 | 29.5 | 10 |
| ● *Konosirus punctatus* | 25.3 | 25.3 | 19.9 | 29.5 | 36.4 | 26.4 | 20.3 | 16.9 | 18 | 41 | 29.5 | 11.5 |
| ● *Escualosa thoracata* | 25.3 | 26.4 | 18.8 | 29.5 | 36 | 26.4 | 19.9 | 17.6 | 18 | 42.9 | 24.9 | 14.2 |
| ● *Sardina pilchardus* | 24.5 | 27.2 | 18.4 | 29.9 | 36.4 | 26.4 | 19.9 | 17.2 | 17.6 | 36.8 | 26.1 | 19.5 |
| ● *Sardinops melanostictus* | 24.9 | 27.2 | 19.2 | 28.7 | 36.4 | 26.1 | 20.3 | 17.2 | 18.4 | 38.3 | 28 | 15.3 |
| ● *Brevoortia tyrannus* | 24.1 | 28.4 | 18.4 | 29.1 | 36 | 26.4 | 20.7 | 16.9 | 16.5 | 37.5 | 33.7 | 12.3 |
| □ *Alosa alosa* | 23.4 | 29.1 | 18 | 29.5 | 36.4 | 26.4 | 19.9 | 17.2 | 16.5 | 37.5 | 37.2 | 8.8 |
| □ *Alosa pseudoharengus* | 24.5 | 28 | 18.8 | 28.7 | 36 | 26.4 | 20.7 | 16.9 | 17.6 | 36.8 | 38.3 | 7.3 |
| ○ *Clupeichthys goniognathus* | 25.7 | 26.1 | 19.5 | 28.7 | 36.4 | 26.4 | 19.5 | 17.6 | 19.9 | 36 | 35.6 | 8.4 |
| ○ *Clupeichthys aesarnensis* | 25.3 | 26.4 | 19.2 | 29.1 | 36.4 | 26.4 | 19.5 | 17.6 | 25.3 | 31.4 | 36 | 7.3 |
| ○ *Clupeichthys perakensis* | 25.3 | 26.4 | 19.9 | 28.4 | 36.4 | 26.4 | 19.5 | 17.6 | 16.9 | 39.1 | 36.8 | 7.3 |
| ○ *Clupeoides sp. Chao Phraya* | 25.7 | 25.7 | 19.9 | 28.7 | 36 | 26.4 | 20.3 | 17.2 | 21.1 | 35.6 | 39.1 | 4.2 |
| ○ *Clupeoides borneensis* | 26.4 | 25.3 | 20.3 | 28 | 36 | 26.4 | 20.3 | 17.2 | 16.5 | 39.8 | 39.1 | 4.6 |
| ○ *Sundasalanx praecox* | 26.4 | 24.5 | 20.3 | 28.7 | 35.2 | 26.8 | 20.7 | 17.2 | 20.3 | 36.4 | 37.9 | 5.4 |
| ○ *Sundasalanx sp. Chao Phraya* | 26.4 | 24.5 | 20.3 | 28.7 | 35.2 | 26.8 | 20.3 | 17.6 | 23 | 34.5 | 37.2 | 5.4 |
| ○ *Sundasalanx mekongensis* | 26.1 | 25.3 | 20.3 | 28.4 | 35.2 | 26.8 | 20.7 | 17.2 | 26.4 | 31.4 | 37.9 | 4.2 |
| □ *Ehirava fluviatilis* | 24.1 | 28 | 19.2 | 28.7 | 36 | 26.4 | 19.9 | 17.6 | 28.4 | 34.1 | 34.1 | 3.4 |
| □ *Gilchristella aestuaria* | 23.8 | 27.6 | 19.2 | 29.5 | 36.4 | 26.1 | 19.9 | 17.6 | 23.4 | 36.8 | 33 | 6.9 |
| □ *Clupeonella cultriventris* | 24.5 | 28 | 18.8 | 28.7 | 36 | 26.1 | 20.3 | 17.6 | 22.6 | 37.9 | 28.7 | 10.7 |
| ● *Clupea harengus* | 26.4 | 24.9 | 18 | 30.7 | 35.6 | 27.2 | 20.3 | 16.9 | 18.8 | 37.9 | 33.7 | 9.6 |
| ● *Clupea pallasii* | 26.8 | 24.5 | 17.6 | 31 | 35.6 | 26.8 | 20.3 | 17.2 | 19.2 | 36.4 | 33.7 | 10.7 |
| ● *Sprattus sprattus* | 26.1 | 24.9 | 19.5 | 29.5 | 35.6 | 26.8 | 20.7 | 16.9 | 16.1 | 38.7 | 34.9 | 10.3 |
| ● *Sprattus muelleri* | 24.9 | 25.7 | 19.5 | 29.9 | 35.6 | 26.4 | 20.3 | 17.6 | 13.4 | 42.1 | 29.5 | 14.9 |
| ● *Sprattus antipodum* | 24.9 | 25.7 | 19.5 | 29.9 | 35.6 | 26.4 | 20.3 | 17.6 | 12.6 | 42.5 | 29.1 | 15.7 |
| ○ *Potamalosa richmondia* | 25.7 | 26.1 | 19.9 | 28.4 | 36 | 26.4 | 20.7 | 16.9 | 13.8 | 40.6 | 40.2 | 5.4 |
| ● *Hyperlophus vittatus* | 25.3 | 26.4 | 19.5 | 28.7 | 36 | 26.1 | 20.7 | 17.2 | 19.5 | 37.2 | 34.9 | 8.4 |
| ● *Ethmidium maculatum* | 25.7 | 25.3 | 19.5 | 29.5 | 35.6 | 26.8 | 20.7 | 16.9 | 15.7 | 38.3 | 37.5 | 8.4 |
| ● *Jenkinsia lamprotaenia* | 25.7 | 24.9 | 18.8 | 30.7 | 35.2 | 26.8 | 20.3 | 17.6 | 15.7 | 44.8 | 28 | 11.5 |
| ● *Spratelloides delicatulus* | 24.9 | 26.8 | 18.8 | 29.5 | 36.4 | 26.4 | 20.3 | 16.9 | 20.7 | 42.9 | 22.6 | 13.8 |
| ● *Spratelloides gracilis* | 25.7 | 24.5 | 20.3 | 29.5 | 35.2 | 26.8 | 20.3 | 17.6 | 21.8 | 37.5 | 29.1 | 11.5 |
| ● *Etrumeus micropus* | 27.2 | 23.4 | 20.3 | 29.1 | 35.6 | 26.4 | 20.3 | 17.6 | 24.9 | 32.2 | 34.1 | 8.8 |
| ○ *Ilisha africana* | 26.4 | 24.9 | 19.9 | 28.7 | 35.6 | 26.8 | 20.7 | 16.9 | 20.3 | 34.9 | 39.5 | 5.4 |
| ○ *Pellona flavipinnis* | 23.8 | 27.6 | 19.9 | 28.7 | 36 | 26.8 | 20.3 | 16.9 | 13.8 | 40.2 | 41.8 | 4.2 |
| ○ *Ilisha elongata* | 26.4 | 25.3 | 19.2 | 29.1 | 36 | 26.8 | 20.3 | 16.9 | 18.8 | 34.5 | 42.9 | 3.8 |
| □ *Pellona ditchela* | 27.6 | 23.8 | 19.5 | 29.1 | 35.6 | 26.8 | 20.7 | 16.9 | 16.9 | 37.9 | 39.5 | 5.7 |
| □ *Anchoviella sp. LBP 2297* | 28 | 22.6 | 20.7 | 28.7 | 35.6 | 26.8 | 20.3 | 17.2 | 21.1 | 34.9 | 37.9 | 6.1 |
| □ *Lycengraulis grossidens* | 26.1 | 24.1 | 20.3 | 29.5 | 36 | 26.8 | 19.9 | 17.2 | 21.8 | 35.2 | 37.5 | 5.4 |
| ○ *Amazonsprattus scintilla* | 26.4 | 23.8 | 20.7 | 29.1 | 35.6 | 26.8 | 20.3 | 17.2 | 28.4 | 27.2 | 34.1 | 5.2 |
| ● *Engraulis encrasicolus* | 26.1 | 24.9 | 18.4 | 30.7 | 35.6 | 27.2 | 20.3 | 16.9 | 24.5 | 31.8 | 33.3 | 10.3 |
| ● *Engraulis japonicus* | 26.4 | 24.5 | 19.2 | 29.9 | 35.6 | 27.2 | 20.3 | 16.9 | 24.9 | 32.6 | 31 | 11.5 |
| ● *Stolephorus chinensis* | 26.8 | 24.5 | 20.3 | 28.4 | 36 | 26.8 | 19.9 | 17.2 | 24.9 | 34.1 | 34.1 | 6.9 |
| ● *Stolephorus waitei* | 26.8 | 24.1 | 20.7 | 28.4 | 36 | 26.8 | 19.9 | 17.2 | 26.4 | 32.6 | 34.9 | 6.1 |
| ○ *Lycothrissa crocodilus* | 25.3 | 26.1 | 20.7 | 28 | 36.8 | 26.1 | 19.9 | 17.2 | 24.5 | 32.6 | 38.3 | 4.6 |
| □ *Setipinna melanochir* | 25.7 | 25.3 | 20.3 | 28.7 | 36 | 26.4 | 20.3 | 17.2 | 16.5 | 39.1 | 36 | 8.4 |
| ● *Coilia reynaldi* | 26.1 | 24.9 | 20.7 | 28.4 | 35.6 | 26.8 | 20.3 | 17.2 | 19.2 | 37.5 | 34.9 | 8.4 |
| ● *Thryssa baelama* | 25.7 | 24.9 | 20.7 | 28.7 | 35.6 | 26.8 | 20.3 | 17.2 | 15.7 | 38.3 | 34.9 | 11.1 |
| ○ *Coilia lindmani* | 26.4 | 24.5 | 20.3 | 28.7 | 36 | 26.4 | 19.9 | 17.6 | 19.9 | 36.4 | 41 | 2.7 |
| □ *Coilia ectenes* | 26.8 | 24.1 | 20.3 | 28.7 | 36 | 26.4 | 19.9 | 17.6 | 21.8 | 34.5 | 39.5 | 4.2 |
| ● *Coilia nasus* | 26.8 | 24.1 | 20.3 | 28.7 | 35.6 | 26.8 | 19.5 | 18 | 22.2 | 33.7 | 39.8 | 4.2 |
| ● *Denticeps clupeoides* | 26.1 | 24.5 | 19.5 | 29.9 | 34.9 | 27.2 | 21.1 | 16.9 | 31.4 | 28.4 | 36 | 4.2 |

CO3



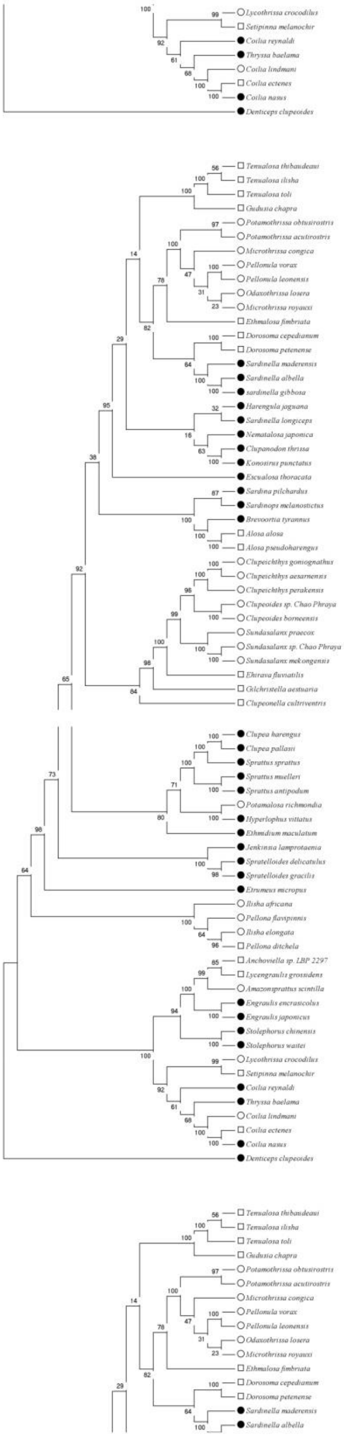
| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 24.2 | 25 | 22.6 | 28.2 | 40.8 | 25.3 | 20.5 | 13.4 | 20 | 38.7 | 26.3 | 15 |
| □ *Tenualosa ilisha* | 24.2 | 25 | 22.9 | 27.9 | 40.3 | 25.5 | 20.5 | 13.7 | 18.9 | 40.5 | 28.4 | 12.1 |
| □ *Tenualosa toli* | 23.9 | 25.3 | 22.9 | 27.9 | 40.3 | 25.5 | 20.5 | 13.7 | 18.4 | 40.8 | 29.2 | 11.6 |
| □ *Gudusia chapra* | 24.5 | 24.7 | 23.9 | 26.8 | 40.3 | 25.5 | 20.5 | 13.7 | 22.6 | 37.1 | 37.9 | 2.4 |
| ○ *Potamothrissa obtusirostris* | 23.9 | 25.5 | 23.4 | 27.1 | 40.3 | 25.8 | 20.3 | 13.7 | 18.7 | 36.3 | 37.6 | 7.4 |
| ○ *Potamothrissa acutirostris* | 25 | 24.7 | 23.2 | 27.1 | 40.3 | 25.8 | 20.3 | 13.7 | 17.4 | 39.2 | 37.4 | 6.1 |
| ○ *Microthrissa congica* | 24.5 | 25.5 | 22.9 | 27.1 | 40.3 | 25.8 | 20.3 | 13.7 | 15 | 38.2 | 41.3 | 5.5 |
| ○ *Pellonula vorax* | 23.2 | 26.8 | 22.6 | 27.4 | 40.3 | 25.8 | 20.3 | 13.7 | 17.1 | 38.2 | 37.6 | 7.1 |
| ○ *Pellonula leonensis* | 25.5 | 24.5 | 22.6 | 27.4 | 40.3 | 25.8 | 20.3 | 13.7 | 14.2 | 40.8 | 40.5 | 4.5 |
| ○ *Odaxothrissa losera* | 23.2 | 26.6 | 23.2 | 27.1 | 40 | 26.1 | 20.3 | 13.7 | 16.6 | 38.7 | 38.9 | 5.8 |
| ○ *Microthrissa royauxi* | 22.9 | 27.1 | 22.9 | 27.1 | 40.3 | 25.8 | 20.3 | 13.7 | 12.9 | 42.6 | 39.5 | 5 |
| □ *Ethmalosa fimbriata* | 23.2 | 26.3 | 22.4 | 28.2 | 40.5 | 25.5 | 20.3 | 13.7 | 16.6 | 38.7 | 39.7 | 5 |
| □ *Dorosoma cepedianum* | 22.4 | 27.4 | 22.9 | 27.4 | 40.3 | 25.3 | 20.5 | 13.9 | 22.6 | 35.5 | 33.9 | 7.9 |
| □ *Dorosoma petenense* | 23.7 | 25.3 | 23.4 | 27.6 | 40.5 | 25.3 | 20.5 | 13.7 | 22.6 | 33.4 | 36.3 | 7.6 |
| ● *Sardinella maderensis* | 22.4 | 27.1 | 21.8 | 28.7 | 40.3 | 25.8 | 20.3 | 13.7 | 19.5 | 40.3 | 28.4 | 11.8 |
| ● *Sardinella albella* | 22.9 | 27.4 | 21.8 | 27.9 | 40 | 26.1 | 20.3 | 13.7 | 21.3 | 38.4 | 26.8 | 13.4 |
| ● *sardinella gibbosa* | 22.9 | 27.4 | 21.8 | 27.9 | 40 | 26.1 | 20.3 | 13.7 | 20.8 | 37.9 | 27.4 | 13.9 |
| ● *Harengula jaguana* | 23.7 | 26.3 | 22.1 | 27.9 | 40 | 26.1 | 20 | 13.9 | 23.4 | 34.2 | 24.2 | 18.2 |
| ● *Sardinella longiceps* | 23.4 | 26.1 | 21.8 | 28.7 | 40.5 | 25.8 | 19.7 | 13.9 | 22.6 | 36.8 | 29.7 | 10.8 |
| ● *Nematalosa japonica* | 23.9 | 25.3 | 22.1 | 28.7 | 40.5 | 25.3 | 20.3 | 13.9 | 19.7 | 40 | 26.8 | 13.4 |
| ● *Clupanodon thrissa* | 23.2 | 26.3 | 22.4 | 28.2 | 40.5 | 25.5 | 20.3 | 13.7 | 19.5 | 38.7 | 27.1 | 14.7 |
| ● *Konosirus punctatus* | 24.7 | 24.7 | 22.1 | 28.4 | 40.5 | 25.5 | 20.3 | 13.7 | 19.2 | 38.4 | 27.1 | 15.3 |
| ● *Escualosa thoracata* | 22.9 | 26.3 | 22.6 | 28.2 | 39.7 | 25.8 | 20.5 | 13.9 | 20.8 | 41.3 | 21.3 | 16.6 |
| ● *Sardina pilchardus* | 24.5 | 23.9 | 22.6 | 28.9 | 40 | 25.8 | 20 | 14.2 | 19.5 | 38.9 | 20.8 | 20.8 |
| ● *Sardinops melanostictus* | 23.4 | 25 | 21.6 | 30 | 40.5 | 25.3 | 20 | 14.2 | 23.2 | 37.6 | 22.4 | 16.8 |
| ● *Brevoortia tyrannus* | 23.7 | 24.5 | 22.6 | 29.2 | 40 | 25.8 | 20 | 14.2 | 18.4 | 39.5 | 31.8 | 10.3 |
| □ *Alosa alosa* | 25.5 | 23.2 | 22.1 | 29.2 | 39.7 | 26.3 | 19.5 | 14.5 | 18.2 | 37.1 | 36.1 | 8.7 |
| □ *Alosa pseudoharengus* | 25.3 | 23.2 | 22.9 | 28.7 | 40 | 25.8 | 20 | 14.2 | 17.1 | 39.5 | 36.8 | 6.6 |
| ○ *Clupeichthys goniognathus* | 22.4 | 26.6 | 23.7 | 27.4 | 40.3 | 26.1 | 20.5 | 13.2 | 22.4 | 38.4 | 30 | 9.2 |
| ○ *Clupeichthys aesarnensis* | 22.4 | 26.6 | 24.2 | 26.8 | 40.5 | 25.8 | 20.5 | 13.2 | 22.9 | 38.4 | 32.1 | 6.6 |
| ○ *Clupeichthys perakensis* | 22.6 | 26.6 | 23.7 | 27.1 | 40.5 | 25.8 | 20.5 | 13.2 | 22.1 | 40.3 | 30.8 | 6.8 |
| ○ *Clupeoides sp. Chao Phraya* | 23.9 | 25 | 24.2 | 26.8 | 40.5 | 25.8 | 20.3 | 13.4 | 19.5 | 35.8 | 42.1 | 2.6 |
| ○ *Clupeoides borneensis* | 24.7 | 24.5 | 23.4 | 27.4 | 40.3 | 26.1 | 20.3 | 13.4 | 14.2 | 38.7 | 43.7 | 3.4 |
| ○ *Sundasalanx praecox* | 23.9 | 24.7 | 23.9 | 27.4 | 40.3 | 26.1 | 20.3 | 13.4 | 23.7 | 36.8 | 32.4 | 7.1 |
| ○ *Sundasalanx sp. Chao Phraya* | 23.7 | 25 | 23.9 | 27.4 | 40.3 | 26.1 | 20.3 | 13.4 | 23.4 | 35.3 | 35.3 | 6.1 |



| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ○ *Sundasalanx mekongensis* | 23.7 | 25 | 24.7 | 26.6 | 40.3 | 26.1 | 20.3 | 13.4 | 24.7 | 32.1 | 36.8 | 6.3 |
| □ *Ehirava fluviatilis* | 22.1 | 26.8 | 24.2 | 26.8 | 40.5 | 26.1 | 20.3 | 13.2 | 30.5 | 26.6 | 40 | 2.9 |
| □ *Gilchristella aestuaria* | 23.4 | 25.3 | 23.7 | 27.6 | 40.3 | 26.1 | 20.3 | 13.4 | 20.8 | 36.6 | 33.7 | 8.9 |
| □ *Clupeonella cultriventris* | 22.9 | 27.1 | 21.3 | 28.7 | 41.3 | 24.7 | 20 | 13.9 | 28.9 | 35 | 25.3 | 10.8 |
| ● *Clupea harengus* | 24.7 | 25 | 22.4 | 27.9 | 40.3 | 26.6 | 19.5 | 13.7 | 21.3 | 33.2 | 31.8 | 13.7 |
| ● *Clupea pallasii* | 24.7 | 25 | 22.4 | 27.9 | 40.3 | 26.6 | 19.5 | 13.7 | 22.6 | 32.1 | 31.3 | 13.9 |
| ● *Sprattus sprattus* | 24.2 | 25 | 22.6 | 28.2 | 40 | 26.6 | 19.5 | 13.9 | 20.8 | 34.7 | 26.8 | 17.6 |
| ● *Sprattus muelleri* | 24.2 | 25.5 | 22.4 | 27.9 | 40 | 26.3 | 19.7 | 13.9 | 24.2 | 32.9 | 25.5 | 17.4 |
| ● *Sprattus antipodum* | 23.9 | 25.8 | 21.8 | 28.4 | 40 | 26.3 | 19.7 | 13.9 | 23.9 | 33.2 | 24.2 | 18.7 |
| ○ *Potamalosa richmondia* | 24.5 | 24.5 | 21.8 | 29.2 | 40.5 | 25.8 | 20.3 | 13.4 | 19.2 | 35.8 | 39.7 | 5.3 |
| ● *Hyperlophus vittatus* | 23.7 | 25.3 | 22.1 | 28.9 | 40.5 | 25.8 | 20.3 | 13.4 | 21.3 | 35.5 | 29.7 | 13.4 |
| ● *Ethmidium maculatum* | 24.5 | 24.5 | 22.6 | 28.4 | 40.8 | 25.3 | 20.3 | 13.7 | 11.6 | 41.8 | 35.8 | 10.8 |
| ● *Jenkinsia lamprotaenia* | 22.9 | 26.6 | 22.1 | 28.4 | 41.3 | 25 | 19.7 | 13.9 | 23.7 | 42.4 | 21.3 | 12.6 |
| ● *Spratelloides delicatulus* | 23.2 | 26.3 | 21.3 | 29.2 | 41.1 | 25.3 | 20 | 13.7 | 17.6 | 49.2 | 18.4 | 14.7 |
| ● *Spratelloides gracilis* | 23.2 | 25.8 | 21.8 | 29.2 | 41.1 | 25.3 | 20.3 | 13.4 | 27.4 | 35 | 22.1 | 15.5 |
| ● *Etrumeus micropus* | 25 | 25 | 23.2 | 26.8 | 40.3 | 25.3 | 20 | 14.5 | 24.7 | 35.5 | 30.8 | 8.9 |
| ○ *Ilisha africana* | 27.4 | 23.2 | 24.5 | 25 | 40.5 | 25.3 | 20.5 | 13.7 | 16.3 | 38.2 | 42.4 | 3.2 |
| ○ *Pellona flavipinnis* | 25 | 25.3 | 24.7 | 25 | 40.3 | 26.1 | 20.3 | 13.4 | 10.8 | 44.5 | 40.5 | 4.2 |
| ○ *Ilisha elongata* | 24.7 | 25.3 | 24.5 | 25.5 | 41.1 | 25 | 20 | 13.9 | 17.6 | 40 | 38.9 | 3.4 |
| □ *Pellona ditchela* | 24.5 | 25.5 | 23.9 | 26.1 | 40.5 | 25.5 | 20.3 | 13.7 | 16.8 | 37.1 | 42.6 | 3.4 |
| □ *Anchoviella sp. LBP 2297* | 23.4 | 25.8 | 23.4 | 27.4 | 40.8 | 25.5 | 20.3 | 13.4 | 24.2 | 33.4 | 37.6 | 4.7 |
| □ *Lycengraulis grossidens* | 24.7 | 24.2 | 24.5 | 26.6 | 40.8 | 25.5 | 20.3 | 13.4 | 24.7 | 31.1 | 38.4 | 5.8 |
| ○ *Amazonsprattus scintilla* | 24.7 | 24.2 | 23.7 | 27.4 | 40.8 | 25.5 | 20.3 | 13.4 | 26.8 | 28.2 | 34.5 | 10.5 |
| ● *Engraulis encrasicolus* | 24.7 | 23.7 | 23.9 | 27.6 | 40.8 | 25.8 | 20 | 13.4 | 25.3 | 33.9 | 31.3 | 9.5 |
| ● *Engraulis japonicus* | 24.7 | 23.7 | 23.7 | 27.9 | 40.8 | 25.8 | 20 | 13.4 | 25.8 | 33.2 | 30.3 | 10.8 |
| ● *Stolephorus chinensis* | 24.7 | 24.7 | 23.2 | 27.4 | 41.1 | 25.5 | 19.7 | 13.7 | 27.4 | 33.2 | 33.2 | 6.3 |
| ● *Stolephorus waitei* | 24.5 | 25 | 23.2 | 27.4 | 41.1 | 25.5 | 19.7 | 13.7 | 26.6 | 33.9 | 32.6 | 6.8 |
| ○ *Lycothrissa crocodilus* | 24.7 | 24.7 | 25.3 | 25.3 | 40.3 | 26.1 | 20.3 | 13.4 | 17.1 | 40.8 | 40.3 | 1.8 |
| □ *Setipinna melanochir* | 23.9 | 25.5 | 24.5 | 26.1 | 40.3 | 26.3 | 20 | 13.4 | 16.1 | 44.2 | 36.3 | 3.4 |
| ● *Coilia reynaldi* | 24.5 | 24.7 | 23.7 | 27.1 | 41.1 | 25.3 | 20.3 | 13.4 | 16.6 | 36.3 | 41.8 | 5.3 |
| ● *Thryssa baelama* | 24.7 | 24.5 | 23.2 | 27.6 | 40.8 | 25.5 | 20.3 | 13.4 | 19.7 | 35.3 | 31.6 | 13.4 |
| ○ *Coilia lindmani* | 26.8 | 22.9 | 23.4 | 26.8 | 40.8 | 25.3 | 20.3 | 13.7 | 24.5 | 32.1 | 39.7 | 3.7 |
| □ *Coilia ectenes* | 25.8 | 23.9 | 23.4 | 26.8 | 40.8 | 25.3 | 20.3 | 13.7 | 23.7 | 33.7 | 40 | 2.6 |
| ● *Coilia nasus* | 26.1 | 23.7 | 23.4 | 26.8 | 40.8 | 25.3 | 20.3 | 13.7 | 25 | 32.4 | 39.7 | 2.9 |
| ● *Denticeps clupeoides* | 24.5 | 25 | 25.8 | 24.7 | 42.1 | 24.2 | 20 | 13.7 | 22.6 | 25.5 | 48.7 | 3.2 |

CYB

| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 20.6 | 28.6 | 23.1 | 27.7 | 40.6 | 29.8 | 16.9 | 12.6 | 18.2 | 35.1 | 29.5 | 17.2 |
| □ *Tenualosa ilisha* | 20.6 | 28.6 | 22.5 | 28.3 | 40.6 | 29.8 | 16.6 | 12.9 | 19.4 | 38.5 | 28.6 | 13.5 |
| □ *Tenualosa toli* | 18.8 | 30.8 | 21.2 | 29.2 | 40.3 | 30.2 | 16.6 | 12.9 | 18.8 | 35.1 | 32 | 14.2 |
| □ *Gudusia chapra* | 21.8 | 27.1 | 24.3 | 26.8 | 40.3 | 30.2 | 16.9 | 12.6 | 19.1 | 36.3 | 41.2 | 3.4 |
| ○ *Potamothrissa obtusirostris* | 20 | 30.2 | 22.5 | 27.4 | 40.6 | 30.2 | 16.6 | 12.6 | 23.7 | 30.2 | 39.1 | 7.1 |
| ○ *Potamothrissa acutirostris* | 21.5 | 28.3 | 22.2 | 28 | 40.6 | 30.2 | 16.6 | 12.6 | 23.1 | 32.3 | 38.2 | 6.5 |
| ○ *Microthrissa congica* | 20.9 | 29.5 | 22.5 | 27.1 | 40.6 | 30.2 | 16.6 | 12.6 | 19.7 | 34.5 | 38.5 | 7.4 |
| ○ *Pellonula vorax* | 20.6 | 29.8 | 22.2 | 27.4 | 40.3 | 30.5 | 16.6 | 12.6 | 19.4 | 33.8 | 41.2 | 5.5 |
| ○ *Pellonula leonensis* | 21.2 | 29.2 | 21.5 | 28 | 40.6 | 30.2 | 16.6 | 12.6 | 18.8 | 35.1 | 43.4 | 2.8 |
| ○ *Odaxothrissa losera* | 20 | 30.2 | 22.2 | 27.7 | 40.6 | 30.2 | 16.6 | 12.6 | 19.1 | 34.8 | 41.8 | 4.3 |
| ○ *Microthrissa royauxi* | 20.6 | 29.8 | 22.2 | 27.4 | 40.6 | 30.2 | 16.6 | 12.6 | 17.8 | 35.1 | 39.1 | 8 |
| □ *Ethmalosa fimbriata* | 22.5 | 28 | 22.2 | 27.4 | 40.3 | 30.5 | 16.9 | 12.3 | 19.7 | 36.9 | 37.8 | 5.5 |
| □ *Dorosoma cepedianum* | 21.2 | 29.5 | 21.8 | 27.4 | 40.9 | 30.2 | 16.3 | 12.6 | 16.3 | 40 | 31.1 | 12.6 |
| □ *Dorosoma petenense* | 23.1 | 27.4 | 23.1 | 26.5 | 40.9 | 30.2 | 16.3 | 12.6 | 20.9 | 36 | 35.1 | 8 |
| ● *Sardinella maderensis* | 20.9 | 28.9 | 23.1 | 27.1 | 40.3 | 30.2 | 16.6 | 12.9 | 19.7 | 37.5 | 27.1 | 15.7 |
| ● *Sardinella albella* | 21.5 | 28.3 | 22.8 | 27.4 | 40.6 | 30.5 | 16.3 | 12.6 | 19.4 | 39.1 | 22.2 | 19.4 |
| ● *sardinella gibbosa* | 21.5 | 28.3 | 22.8 | 27.4 | 40.6 | 30.5 | 16.3 | 12.6 | 19.7 | 38.8 | 22.2 | 19.4 |
| ● *Harengula jaguana* | 20.9 | 29.5 | 21.8 | 27.7 | 40.9 | 29.8 | 17.2 | 12 | 26.5 | 36.3 | 17.2 | 20 |
| ● *Sardinella longiceps* | 22.5 | 28.3 | 21.5 | 27.7 | 40.9 | 30.2 | 16.9 | 12 | 23.1 | 34.8 | 25.5 | 16.6 |
| ● *Nematalosa japonica* | 21.2 | 28.6 | 22.5 | 27.7 | 40.3 | 30.2 | 16.6 | 12.9 | 21.8 | 36.9 | 21.5 | 19.7 |
| ● *Clupanodon thrissa* | 20.6 | 29.2 | 21.2 | 28.9 | 40.3 | 30.2 | 16.6 | 12.9 | 19.7 | 35.1 | 25.5 | 19.7 |
| ● *Konosirus punctatus* | 21.5 | 28.6 | 21.2 | 28.6 | 40.3 | 30.2 | 16.6 | 12.9 | 21.8 | 32.6 | 25.8 | 19.7 |
| ● *Escualosa thoracata* | 20.3 | 29.8 | 21.5 | 28.3 | 41.2 | 29.8 | 16.6 | 12.3 | 22.8 | 40.3 | 16 | 20.9 |
| ● *Sardina pilchardus* | 20.9 | 28.6 | 19.7 | 30.8 | 41.8 | 29.2 | 16.3 | 12.6 | 24.9 | 34.2 | 18.2 | 22.8 |
| ● *Sardinops melanostictus* | 20.3 | 29.5 | 21.5 | 28.6 | 41.8 | 28.9 | 16.9 | 12.3 | 28 | 34.5 | 19.4 | 18.2 |
| ● *Brevoortia tyrannus* | 22.2 | 28.3 | 21.8 | 27.7 | 40.6 | 30.2 | 16.6 | 12.6 | 23.1 | 31.1 | 27.1 | 18.8 |
| □ *Alosa alosa* | 21.5 | 29.2 | 22.2 | 27.1 | 40.9 | 29.8 | 16.6 | 12.6 | 18.2 | 35.4 | 32.6 | 13.8 |
| □ *Alosa pseudoharengus* | 21.2 | 29.5 | 22.2 | 27.1 | 40.6 | 30.2 | 16.6 | 12.6 | 21.5 | 32 | 34.5 | 12 |
| ○ *Clupeichthys goniognathus* | 20.6 | 28.3 | 24.6 | 26.5 | 40.6 | 30.5 | 16.9 | 12 | 21.8 | 36.6 | 28.6 | 12.9 |
| ○ *Clupeichthys aesarnensis* | 20 | 28.3 | 24.6 | 27.1 | 40.6 | 30.8 | 16.6 | 12 | 22.2 | 37.5 | 31.1 | 9.2 |
| ○ *Clupeichthys perakensis* | 20.9 | 27.7 | 24 | 27.4 | 40.6 | 30.8 | 16.9 | 11.7 | 20.9 | 36.6 | 35.7 | 6.8 |
| ○ *Clupeoides sp. Chao Phraya* | 21.2 | 26.8 | 24.9 | 27.1 | 40.9 | 30.5 | 16.9 | 11.7 | 23.7 | 31.1 | 42.2 | 3.1 |
| ○ *Clupeoides borneensis* | 19.7 | 28.9 | 24 | 27.4 | 40.6 | 30.5 | 17.2 | 11.7 | 18.2 | 35.4 | 40.3 | 6.2 |
| ○ *Sundasalanx praecox* | 24.6 | 26.5 | 23.7 | 25.2 | 40.6 | 31.1 | 16.9 | 11.4 | 21.8 | 35.7 | 33.8 | 8.6 |
| ○ *Sundasalanx sp. Chao Phraya* | 25.5 | 25.5 | 23.7 | 25.2 | 40.3 | 31.4 | 16.9 | 11.4 | 23.4 | 32.3 | 36.3 | 8 |
| ○ *Sundasalanx mekongensis* | 23.7 | 27.1 | 24.6 | 24.6 | 40.3 | 31.4 | 16.9 | 11.4 | 23.1 | 32 | 40.9 | 4 |
| □ *Ehirava fluviatilis* | 22.5 | 26.8 | 23.4 | 27.4 | 40.3 | 31.1 | 17.2 | 11.4 | 23.1 | 31.4 | 39.7 | 5.8 |
| □ *Gilchristella aestuaria* | 20.9 | 29.5 | 22.5 | 27.1 | 40.9 | 30.5 | 16.9 | 11.7 | 23.4 | 36.3 | 27.4 | 12.9 |
| □ *Clupeonella cultriventris* | 20 | 29.5 | 22.2 | 28.3 | 40.6 | 30.2 | 17.2 | 12 | 27.1 | 30.8 | 22.2 | 20 |
| ● *Clupea harengus* | 21.8 | 28 | 23.7 | 26.5 | 40.3 | 31.1 | 16.9 | 11.7 | 25.2 | 29.2 | 24 | 21.5 |
| ● *Clupea pallasii* | 22.5 | 27.4 | 23.7 | 26.5 | 40.3 | 31.1 | 16.9 | 11.7 | 24.3 | 30.8 | 23.4 | 21.5 |
| ● *Sprattus sprattus* | 23.4 | 26.8 | 23.4 | 26.5 | 40.3 | 31.1 | 16.6 | 12 | 24.6 | 32 | 24 | 19.4 |
| ● *Sprattus muelleri* | 21.8 | 27.7 | 23.7 | 26.8 | 40.3 | 31.1 | 16.9 | 11.7 | 24.6 | 32.3 | 26.2 | 16.9 |
| ● *Sprattus antipodum* | 21.8 | 27.7 | 23.7 | 26.8 | 40.3 | 31.1 | 16.9 | 11.7 | 24.6 | 31.7 | 27.1 | 16.6 |
| ○ *Potamalosa richmondia* | 22.5 | 28.6 | 21.8 | 27.1 | 40.3 | 31.1 | 17.2 | 11.4 | 23.1 | 30.5 | 39.1 | 7.4 |
| ● *Hyperlophus vittatus* | 23.1 | 27.1 | 21.8 | 28 | 40.6 | 31.1 | 17.2 | 11.1 | 23.4 | 35.1 | 26.2 | 15.4 |
| ● *Ethmidium maculatum* | 21.2 | 29.2 | 21.2 | 28.3 | 40.9 | 29.8 | 17.5 | 11.7 | 13.2 | 34.5 | 36.3 | 16 |
| ● *Jenkinsia lamprotaenia* | 19.7 | 29.5 | 21.2 | 29.5 | 40.3 | 31.1 | 15.7 | 12.9 | 22.8 | 44.9 | 13.2 | 19.1 |
| ● *Spratelloides delicatulus* | 22.5 | 28.3 | 19.4 | 29.8 | 40 | 30.8 | 16.3 | 12.9 | 20.3 | 46.8 | 12.9 | 20 |
| ● *Spratelloides gracilis* | 25.2 | 25.2 | 20.6 | 28.9 | 40 | 30.8 | 16.6 | 12.6 | 21.5 | 37.5 | 23.1 | 17.8 |
| ● *Etrumeus micropus* | 22.2 | 27.4 | 21.2 | 29.2 | 40.9 | 30.5 | 16.9 | 11.7 | 24.9 | 30.8 | 30.8 | 13.5 |
| ○ *Ilisha africana* | 21.5 | 27.4 | 25.2 | 25.8 | 39.7 | 31.7 | 16.9 | 11.7 | 19.4 | 29.2 | 47.7 | 3.7 |
| ○ *Pellona flavipinnis* | 18.8 | 30.2 | 24.9 | 26.2 | 39.4 | 31.7 | 17.5 | 11.4 | 10.2 | 36.9 | 48.3 | 4.6 |
| ○ *Ilisha elongata* | 20 | 29.5 | 24.6 | 25.8 | 39.7 | 31.4 | 17.5 | 11.4 | 13.8 | 38.2 | 41.8 | 6.2 |
| □ *Pellona ditchela* | 21.8 | 28 | 24.9 | 25.2 | 39.4 | 31.7 | 17.5 | 11.4 | 12.6 | 36 | 47.7 | 3.7 |
| □ *Anchoviella sp. LBP 2297* | 19.7 | 28.6 | 24.6 | 27.1 | 40.9 | 30.8 | 16.6 | 11.7 | 21.5 | 29.8 | 42.5 | 6.2 |
| □ *Lycengraulis grossidens* | 20.9 | 27.4 | 23.7 | 28 | 40.9 | 30.8 | 16.6 | 11.7 | 22.2 | 29.5 | 37.5 | 10.8 |
| ○ *Amazonsprattus scintilla* | 21.5 | 26.8 | 23.7 | 28 | 40.9 | 30.8 | 16.6 | 11.7 | 24.9 | 24.6 | 28.9 | 21.5 |
| ● *Engraulis encrasicolus* | 19.4 | 28.6 | 24 | 28 | 40.6 | 31.1 | 16.6 | 11.7 | 24.3 | 29.8 | 31.4 | 14.5 |
| ● *Engraulis japonicus* | 19.7 | 28.6 | 23.7 | 28 | 40.6 | 31.1 | 16.6 | 11.7 | 26.2 | 27.4 | 32 | 14.5 |
| ● *Stolephorus chinensis* | 21.2 | 25.8 | 24 | 28.9 | 40.3 | 31.1 | 16.6 | 12 | 25.8 | 29.8 | 36 | 8.3 |
| ● *Stolephorus waitei* | 21.2 | 25.8 | 24 | 28.9 | 40.3 | 31.1 | 16.6 | 12 | 27.1 | 29.5 | 32.6 | 10.8 |



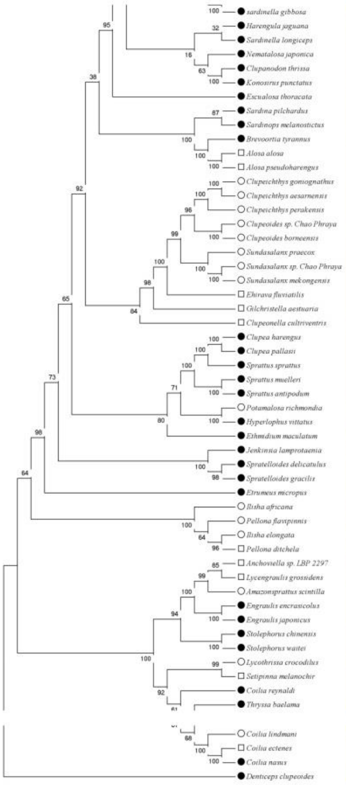
| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ○ Lycothrissa crocodilus | 20.3 | 28 | 24.9 | 26.8 | 40.6 | 31.4 | 16.6 | 11.4 | 16.9 | 36.9 | 42.8 | 3.4 |
| □ Setipinna melanochir | 21.2 | 28 | 24.9 | 25.8 | 40.9 | 30.8 | 16.9 | 11.4 | 16 | 37.8 | 36.3 | 9.8 |
| ● Coilia reynaldi | 22.8 | 25.8 | 24.3 | 27.1 | 40.6 | 31.4 | 17.2 | 10.8 | 18.5 | 32.6 | 43.7 | 5.2 |
| ● Thryssa baelama | 21.2 | 27.4 | 24 | 27.4 | 40.3 | 31.7 | 16.6 | 11.4 | 17.2 | 33.8 | 32 | 16.9 |
| ○ Coilia lindmani | 21.5 | 26.5 | 24.9 | 27.1 | 40.9 | 30.8 | 16.6 | 11.7 | 20.6 | 30.8 | 42.8 | 5.8 |
| □ Coilia ectenes | 21.8 | 26.2 | 25.2 | 26.8 | 40.6 | 31.1 | 16.6 | 11.7 | 20 | 32 | 44 | 4 |
| ● Coilia nasus | 21.8 | 26.2 | 25.2 | 26.8 | 40.6 | 31.1 | 16.6 | 11.7 | 20.3 | 31.7 | 44.3 | 3.7 |
| ● Denticeps clupeoides | 22.8 | 26.5 | 25.8 | 24.9 | 40.6 | 30.2 | 18.2 | 11.1 | 24 | 24.6 | 47.4 | 4 |

ND1

| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ Tenualosa thibaudeaui | 15.2 | 30.7 | 28.7 | 25.5 | 38.7 | 33.2 | 15.2 | 12.9 | 18.1 | 33.2 | 32.4 | 16.3 |
| □ Tenualosa ilisha | 17.2 | 30.1 | 29.8 | 22.9 | 39 | 32.7 | 14.9 | 13.5 | 17.5 | 36.7 | 30.1 | 15.8 |
| □ Tenualosa toli | 15.5 | 30.7 | 28.7 | 25.2 | 38.1 | 33 | 15.8 | 13.2 | 14.9 | 36.7 | 30.9 | 17.5 |
| □ Gudusia chapra | 16.3 | 29.2 | 33.8 | 20.6 | 37 | 36.1 | 15.2 | 11.7 | 15.5 | 37 | 44.7 | 2.9 |
| ○ Potamothrissa obtusirostris | 19 | 27 | 31 | 23 | 35.9 | 36.5 | 15.8 | 11.8 | 17 | 33.3 | 43.4 | 6.3 |
| ○ Potamothrissa acutirostris | 17.5 | 27.9 | 31.6 | 23 | 36.2 | 36.2 | 15.8 | 11.8 | 21.8 | 31.6 | 39.7 | 6.9 |
| ○ Microthrissa congica | 16.1 | 30.2 | 31.3 | 22.4 | 36.5 | 36.5 | 15.5 | 11.5 | 15.5 | 34.5 | 43.7 | 6.3 |
| ○ Pellonula vorax | 16.4 | 29 | 31 | 23.6 | 36.2 | 36.5 | 15.8 | 11.5 | 16.7 | 35.9 | 42.2 | 5.2 |
| ○ Pellonula leonensis | 15.8 | 29.9 | 30.7 | 23.6 | 35.9 | 36.8 | 15.5 | 11.8 | 16.7 | 35.6 | 43.7 | 4 |
| ○ Odaxothrissa losera | 15.2 | 30.5 | 31.9 | 22.4 | 35.6 | 36.8 | 16.1 | 11.5 | 17.8 | 34.5 | 43.1 | 4.6 |
| ○ Microthrissa royauxi | 16.7 | 28.4 | 31.9 | 23 | 35.6 | 37.1 | 15.5 | 11.8 | 15.8 | 36.2 | 43.4 | 4.6 |
| □ Ethmalosa fimbriata | 17 | 29.9 | 29.6 | 23.6 | 37.4 | 35.3 | 15.5 | 11.8 | 18.7 | 33 | 40.8 | 7.5 |
| □ Dorosoma cepedianum | 16.1 | 29.9 | 31 | 23 | 37.1 | 35.6 | 15.5 | 11.8 | 18.1 | 37.9 | 36.5 | 7.5 |
| □ Dorosoma petenense | 16.7 | 29.6 | 31.6 | 22.1 | 37.9 | 34.8 | 15.5 | 11.8 | 21.8 | 33.3 | 37.9 | 6.9 |
| ● Sardinella maderensis | 16.7 | 30.2 | 29 | 24.1 | 36.5 | 35.6 | 15.2 | 12.6 | 19 | 35.3 | 32.5 | 13.2 |
| ● Sardinella albella | 16.1 | 30.7 | 29.3 | 23.9 | 37.1 | 35.1 | 15.5 | 12.4 | 20.7 | 36.5 | 25.3 | 17.5 |
| ● sardinella gibbosa | 16.4 | 30.5 | 29.6 | 23.6 | 37.1 | 35.1 | 15.5 | 12.4 | 21 | 36.8 | 26.1 | 16.1 |
| ● Harengula jaguana | 17 | 28.7 | 28.7 | 25.6 | 35.9 | 35.9 | 15.8 | 12.4 | 20.4 | 37.9 | 21.3 | 20.4 |
| ● Sardinella longiceps | 18.1 | 28.4 | 29 | 24.4 | 36.5 | 36.2 | 15.8 | 11.5 | 19.8 | 35.1 | 30.5 | 14.7 |
| ● Nematalosa japonica | 17 | 29.3 | 29.9 | 23.9 | 37.1 | 35.3 | 15.2 | 12.4 | 20.4 | 31.9 | 30.2 | 17.5 |
| ● Clupanodon thrissa | 16.7 | 29.9 | 31.3 | 22.1 | 36.8 | 36.5 | 15.2 | 11.5 | 21 | 31.3 | 31 | 16.7 |
| ● Konosirus punctatus | 16.4 | 30.2 | 28.2 | 25.3 | 36.2 | 36.8 | 14.9 | 12.1 | 19.5 | 39.1 | 25 | 16.4 |
| ● Escualosa thoracata | 15.5 | 30.5 | 29 | 25 | 37.6 | 35.1 | 14.7 | 12.6 | 17.5 | 42.5 | 23.9 | 16.1 |
| ● Sardina pilchardus | 16.7 | 29.3 | 27.9 | 26.1 | 36.8 | 36.5 | 14.7 | 12.1 | 22.1 | 37.6 | 19.8 | 20.4 |
| ● Sardinops melanostictus | 17.2 | 28.4 | 29 | 25.3 | 35.9 | 37.9 | 15.2 | 10.9 | 22.4 | 37.4 | 20.4 | 19.8 |
| ● Brevoortia tyrannus | 17.8 | 28.7 | 28.7 | 24.7 | 35.6 | 37.6 | 14.9 | 11.8 | 19.3 | 37.9 | 30.2 | 12.6 |
| □ Alosa alosa | 17.5 | 29 | 31 | 22.4 | 35.1 | 38.5 | 15.2 | 11.2 | 16.4 | 40.5 | 32.5 | 10.6 |
| □ Alosa pseudoharengus | 17.2 | 29 | 31 | 22.7 | 35.6 | 37.6 | 15.2 | 11.5 | 17.8 | 39.4 | 33 | 9.8 |
| ○ Clupeichthys goniognathus | 20.1 | 29.9 | 27 | 23 | 37.9 | 35.6 | 14.7 | 11.8 | 17.8 | 36.8 | 34.5 | 10.9 |
| ○ Clupeichthys aesarnensis | 19.8 | 30.2 | 27.3 | 22.7 | 38.8 | 35.1 | 14.7 | 11.5 | 19.3 | 35.1 | 36.2 | 9.5 |
| ○ Clupeichthys perakensis | 19.5 | 30.5 | 26.7 | 23.3 | 38.8 | 34.8 | 14.1 | 12.4 | 18.1 | 37.4 | 37.6 | 6.9 |
| ○ Clupeoides sp. Chao Phraya | 19.8 | 27.9 | 31.6 | 20.7 | 38.5 | 34.8 | 15.5 | 11.2 | 18.7 | 30.7 | 49.4 | 1.1 |
| ○ Clupeoides borneensis | 17.8 | 29 | 31.9 | 21.3 | 37.4 | 35.9 | 15.2 | 11.5 | 15.5 | 35.1 | 44.3 | 5.2 |
| ○ Sundasalanx praecox | 20.7 | 27.9 | 27.6 | 23.9 | 39.4 | 33.9 | 15.8 | 10.9 | 18.7 | 31.6 | 41.1 | 8.6 |
| ○ Sundasalanx sp. Chao Phraya | 19.8 | 27 | 30.2 | 23 | 39.7 | 34.8 | 14.9 | 10.6 | 21.8 | 31.6 | 40.2 | 6.3 |
| ○ Sundasalanx mekongensis | 19.8 | 26.7 | 31.6 | 21.8 | 40.2 | 34.2 | 14.9 | 10.6 | 18.4 | 33.3 | 41.7 | 6.6 |
| □ Ehirava fluviatilis | 17.8 | 27.9 | 31.3 | 23 | 39.7 | 32.8 | 17 | 10.6 | 23 | 29.3 | 42.8 | 4.9 |
| □ Gilchristella aestuaria | 15.2 | 29.3 | 30.5 | 25 | 35.3 | 37.4 | 15.8 | 11.5 | 21.6 | 30.7 | 37.6 | 10.1 |
| □ Clupeonella cultriventris | 17.5 | 28.7 | 29.6 | 24.1 | 36.2 | 37.6 | 15.5 | 10.6 | 25 | 33 | 25 | 17 |
| ● Clupea harengus | 17.8 | 29 | 30.7 | 22.4 | 37.1 | 36.8 | 14.7 | 11.5 | 22.1 | 33 | 26.4 | 18.4 |
| ● Clupea pallasii | 17.5 | 29.3 | 30.7 | 22.4 | 37.4 | 36.5 | 14.7 | 11.5 | 23.6 | 31.9 | 25.9 | 18.7 |
| ● Sprattus sprattus | 18.1 | 28.7 | 29.3 | 23.9 | 37.4 | 36.2 | 14.7 | 11.8 | 25.9 | 32.8 | 23 | 18.4 |
| ● Sprattus muelleri | 17.5 | 29.3 | 29.3 | 23.9 | 37.9 | 34.2 | 14.7 | 13.2 | 20.1 | 38.8 | 21 | 20.1 |
| ● Sprattus antipodum | 17 | 29.6 | 29.3 | 24.1 | 37.9 | 33.9 | 14.7 | 13.5 | 20.4 | 38.5 | 21.8 | 19.3 |
| ○ Potamalosa richmondia | 14.9 | 29.3 | 33.6 | 22.1 | 36.8 | 36.5 | 15.2 | 11.5 | 17.2 | 35.9 | 41.7 | 5.2 |
| ● Hyperlophus vittatus | 15.8 | 28.2 | 32.2 | 23.9 | 37.1 | 36.2 | 15.2 | 11.5 | 24.1 | 31 | 26.7 | 18.1 |
| ● Ethmidium maculatum | 15.5 | 29.3 | 32.2 | 23 | 36.8 | 36.2 | 15.5 | 11.5 | 14.1 | 38.5 | 31.9 | 15.5 |
| ● Jenkinsia lamprotaenia | 18.1 | 30.5 | 26.1 | 25.3 | 37.6 | 36.8 | 13.5 | 12.1 | 23.6 | 44.3 | 19 | 13.2 |
| ● Spratelloides delicatulus | 15.8 | 31.3 | 27 | 25.9 | 37.9 | 34.5 | 14.7 | 12.9 | 21.3 | 41.4 | 16.4 | 21 |
| ● Spratelloides gracilis | 19.8 | 25.9 | 29.6 | 24.7 | 39.1 | 35.1 | 12.9 | 12.9 | 23.6 | 35.3 | 22.7 | 18.4 |
| ● Etrumeus micropus | 20.7 | 27.9 | 27.3 | 24.1 | 39.7 | 33.3 | 15.2 | 11.8 | 23.3 | 33.6 | 35.1 | 8 |
| ○ Ilisha africana | 16.7 | 28.4 | 37.4 | 17.5 | 35.3 | 37.1 | 15.8 | 11.8 | 14.4 | 37.1 | 46.8 | 1.7 |
| ○ Pellona flavipinnis | 15.5 | 29 | 38.2 | 17.2 | 35.3 | 37.4 | 16.4 | 10.9 | 10.3 | 37.1 | 48 | 4.6 |
| ○ Ilisha elongata | 18.1 | 27.9 | 35.3 | 18.7 | 35.3 | 37.4 | 16.4 | 10.9 | 10.9 | 39.1 | 45.7 | 4.3 |
| □ Pellona ditchela | 14.7 | 29.3 | 37.1 | 19 | 35.6 | 37.1 | 16.7 | 10.6 | 9.2 | 38.8 | 49.1 | 2.9 |
| □ Anchoviella sp. LBP 2297 | 16.7 | 28.7 | 34.5 | 20.1 | 37.1 | 36.8 | 15.5 | 10.6 | 22.4 | 32.5 | 39.9 | 5.2 |
| □ Lycengraulis grossidens | 16.7 | 29.6 | 33.3 | 20.4 | 37.4 | 36.5 | 15.2 | 10.9 | 25.9 | 29.9 | 35.1 | 9.2 |
| ○ Amazonsprattus scintilla | 17.5 | 29 | 32.5 | 21 | 37.9 | 35.6 | 14.9 | 11.5 | 22.4 | 30.7 | 32.2 | 14.7 |
| ● Engraulis encrasicolus | 18.4 | 27.6 | 33.3 | 20.7 | 37.6 | 35.6 | 15.2 | 11.5 | 24.7 | 32.2 | 26.7 | 16.4 |
| ● Engraulis japonicus | 18.7 | 27.6 | 33 | 20.7 | 37.6 | 35.9 | 15.2 | 11.2 | 24.7 | 32.2 | 26.1 | 17 |
| ● Stolephorus chinensis | 18.1 | 27.6 | 32.5 | 21.8 | 37.9 | 36.2 | 15.2 | 10.6 | 23 | 31 | 39.7 | 6.3 |
| ● Stolephorus waitei | 18.7 | 27 | 32.8 | 21.6 | 37.9 | 36.2 | 15.2 | 10.6 | 23.9 | 30.5 | 36.8 | 8.9 |
| ○ Lycothrissa crocodilus | 15.5 | 30.2 | 37.4 | 17 | 37.6 | 35.9 | 15.8 | 10.6 | 17.2 | 35.9 | 45.1 | 1.7 |
| □ Setipinna melanochir | 14.1 | 31.7 | 36 | 18.2 | 36.9 | 36.3 | 16.1 | 10.7 | 13.8 | 41.5 | 38 | 6.6 |
| ● Coilia reynaldi | 14.4 | 30.5 | 35.1 | 20.1 | 37.6 | 34.8 | 16.7 | 10.9 | 14.1 | 38.2 | 41.4 | 6.3 |
| ● Thryssa baelama | 19.3 | 28.2 | 31.3 | 21.3 | 37.4 | 35.9 | 15.2 | 11.5 | 18.7 | 35.1 | 30.5 | 15.8 |
| ○ Coilia lindmani | 15.5 | 29.9 | 35.3 | 19.3 | 37.1 | 35.6 | 16.4 | 10.9 | 22.1 | 32.8 | 42.2 | 2.9 |
| □ Coilia ectenes | 15.5 | 29.3 | 35.6 | 19.5 | 37.6 | 35.1 | 16.1 | 11.2 | 22.4 | 33 | 41.4 | 3.2 |
| ● Coilia nasus | 15.2 | 29.6 | 35.6 | 19.5 | 37.6 | 35.1 | 16.1 | 11.2 | 22.4 | 33 | 40.8 | 3.7 |
| ● Denticeps clupeoides | 20.4 | 27.3 | 34.8 | 17.5 | 38.5 | 34.5 | 16.4 | 10.6 | 24.4 | 23 | 50.3 | 2.3 |

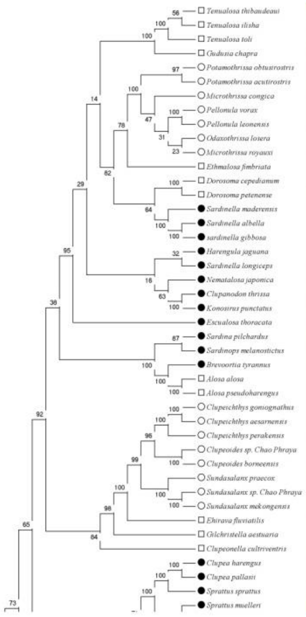
ND2

| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ Tenualosa thibaudeaui | 19.8 | 35.3 | 21.6 | 23.3 | 45.7 | 25.9 | 14.7 | 13.8 | 24.1 | 28.4 | 32.8 | 14.7 |
| □ Tenualosa ilisha | 21.6 | 32.8 | 21.6 | 24.1 | 45.7 | 25 | 16.4 | 12.9 | 19 | 36.2 | 25.9 | 19 |
| □ Tenualosa toli | 20.7 | 33.6 | 20.7 | 25 | 47.4 | 24.1 | 15.5 | 12.9 | 16.4 | 34.5 | 38.8 | 10.3 |
| □ Gudusia chapra | 26.7 | 27.6 | 23.3 | 22.4 | 43.1 | 28.4 | 15.5 | 12.9 | 20.7 | 34.5 | 37.9 | 6.9 |
| ○ Potamothrissa obtusirostris | 25 | 29.3 | 21.6 | 24.1 | 45.7 | 25.9 | 15.5 | 12.9 | 14.7 | 37.1 | 40.5 | 7.8 |
| ○ Potamothrissa acutirostris | 26.7 | 27.6 | 21.6 | 24.1 | 45.7 | 25.9 | 15.5 | 12.9 | 19.8 | 30.2 | 39.7 | 10.3 |
| ○ Microthrissa congica | 24.1 | 30.2 | 21.6 | 24.1 | 44.8 | 26.7 | 15.5 | 12.9 | 12.1 | 44.8 | 32.8 | 10.3 |
| ○ Pellonula vorax | 25 | 28.4 | 21.6 | 25 | 45.7 | 25.9 | 15.5 | 12.9 | 20.7 | 34.5 | 36.2 | 8.6 |
| ○ Pellonula leonensis | 22.4 | 31.9 | 21.6 | 24.1 | 45.7 | 25.9 | 15.5 | 12.9 | 18.1 | 38.8 | 34.5 | 8.6 |
| ○ Odaxothrissa losera | 24.1 | 31 | 20.7 | 24.1 | 45.7 | 25.9 | 15.5 | 12.9 | 20.7 | 33.6 | 38.8 | 6.9 |
| ○ Microthrissa royauxi | 25 | 29.3 | 22.4 | 23.3 | 45.7 | 25.9 | 15.5 | 12.9 | 14.7 | 42.2 | 37.9 | 5.2 |
| □ Ethmalosa fimbriata | 24.1 | 31 | 19.8 | 25 | 46.6 | 25 | 15.5 | 12.9 | 18.1 | 31.9 | 37.9 | 12.1 |
| □ Dorosoma cepedianum | 24.1 | 31 | 19 | 25.9 | 45.7 | 25.9 | 15.5 | 12.9 | 22.4 | 31 | 34.5 | 12.1 |
| □ Dorosoma petenense | 26.7 | 28.4 | 20.7 | 24.1 | 44.8 | 27.6 | 13.8 | 13.8 | 23.3 | 29.3 | 39.7 | 7.8 |
| ● Sardinella maderensis | 22.4 | 31.9 | 20.7 | 25 | 45.7 | 25.9 | 15.5 | 12.9 | 21.6 | 30.2 | 37.1 | 11.2 |
| ● Sardinella albella | 20.7 | 33.6 | 19.8 | 25.9 | 46.6 | 25 | 15.5 | 12.9 | 13.8 | 37.1 | 31.9 | 17.2 |

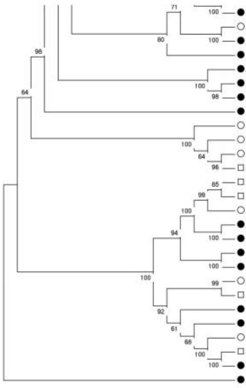


| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ● sardinella gibbosa | 21.6 | 32.8 | 19.8 | 25.9 | 46.6 | 25 | 15.5 | 12.9 | 15.5 | 36.2 | 28.4 | 19.8 |
| ● Harengula jaguana | 25 | 32.8 | 18.1 | 24.1 | 48.3 | 25 | 16.4 | 10.3 | 21.6 | 38.8 | 21.6 | 18.1 |
| ● Sardinella longiceps | 25.9 | 30.2 | 19 | 25 | 46.6 | 25 | 15.5 | 12.9 | 16.4 | 33.6 | 35.3 | 14.7 |
| ● Nematalosa japonica | 20.7 | 32.8 | 21.6 | 25 | 44.8 | 26.7 | 15.5 | 12.9 | 20.7 | 32.8 | 29.3 | 17.2 |
| ● Clupanodon thrissa | 20.7 | 33.6 | 21.6 | 24.1 | 46.6 | 25.9 | 15.5 | 12.1 | 17.2 | 38.8 | 28.4 | 15.5 |
| ● Konosirus punctatus | 21.6 | 32.8 | 21.6 | 24.1 | 45.7 | 25.9 | 15.5 | 12.9 | 22.4 | 31.9 | 22.4 | 23.3 |
| ● Escualosa thoracata | 24.1 | 31 | 20.7 | 24.1 | 46.6 | 25 | 14.7 | 13.8 | 20.7 | 35.3 | 21.6 | 22.4 |
| ● Sardina pilchardus | 23.3 | 28.4 | 19 | 29.3 | 46.6 | 25.9 | 14.7 | 12.9 | 20.7 | 32.8 | 27.6 | 19 |
| ● Sardinops melanostictus | 23.3 | 28.4 | 22.4 | 25.9 | 46.6 | 25.9 | 14.7 | 12.9 | 20.7 | 29.3 | 29.3 | 20.7 |
| ● Brevoortia tyrannus | 25.9 | 27.6 | 21.6 | 25 | 46.6 | 25 | 15.5 | 12.9 | 20.7 | 30.2 | 36.2 | 12.9 |
| □ Alosa alosa | 21.6 | 31.9 | 20.7 | 25.9 | 47.4 | 24.1 | 15.5 | 12.9 | 14.7 | 37.1 | 34.5 | 13.8 |
| □ Alosa pseudoharengus | 23.3 | 30.2 | 20.7 | 25.9 | 46.6 | 25 | 15.5 | 12.9 | 18.1 | 30.2 | 41.4 | 10.3 |
| ○ Clupeichthys goniognathus | 25 | 34.5 | 18.1 | 22.4 | 44 | 26.7 | 16.4 | 12.9 | 27.6 | 27.6 | 33.6 | 11.2 |
| ○ Clupeichthys aesarnensis | 23.3 | 34.5 | 21.6 | 20.7 | 44.8 | 25.9 | 16.4 | 12.9 | 25 | 30.2 | 35.3 | 9.5 |
| ○ Clupeichthys perakensis | 23.3 | 33.6 | 19.8 | 23.3 | 44.8 | 25.9 | 16.4 | 12.9 | 29.3 | 25 | 36.2 | 9.5 |
| ○ Clupeoides sp. Chao Phraya | 21.6 | 34.5 | 20.7 | 23.3 | 45.7 | 25 | 16.4 | 12.9 | 23.3 | 22.4 | 50.9 | 3.4 |
| ○ Clupeoides borneensis | 22.4 | 33.6 | 20.7 | 23.3 | 46.6 | 24.1 | 17.2 | 12.1 | 14.7 | 29.3 | 48.3 | 7.8 |
| ○ Sundasalanx praecox | 23.3 | 32.8 | 17.2 | 26.7 | 45.7 | 25 | 16.4 | 12.9 | 22.4 | 29.3 | 37.9 | 10.3 |
| ○ Sundasalanx sp. Chao Phraya | 23.3 | 33.6 | 19 | 24.1 | 46.6 | 24.1 | 16.4 | 12.9 | 23.3 | 29.3 | 44 | 3.4 |
| ○ Sundasalanx mekongensis | 23.3 | 33.6 | 19 | 24.1 | 46.6 | 24.1 | 17.2 | 12.1 | 30.2 | 23.3 | 42.2 | 4.3 |
| □ Ehirava fluviatilis | 23.3 | 32.8 | 19 | 25 | 43.1 | 27.6 | 17.2 | 12.1 | 25 | 30.2 | 43.1 | 1.7 |
| □ Gilchristella aestuaria | 23.3 | 30.2 | 20.7 | 25.9 | 45.7 | 25.9 | 15.5 | 12.9 | 19.8 | 31.9 | 36.2 | 12.1 |
| □ Clupeonella cultriventris | 20.7 | 33.6 | 18.1 | 27.6 | 47.4 | 24.1 | 15.5 | 12.9 | 24.1 | 29.3 | 30.2 | 16.4 |
| ● Clupea harengus | 23.3 | 28.4 | 17.2 | 31 | 47.4 | 23.3 | 15.5 | 13.8 | 19.8 | 32.8 | 32.8 | 14.7 |
| ● Clupea pallasii | 23.3 | 28.4 | 17.2 | 31 | 47.4 | 23.3 | 15.5 | 13.8 | 20.7 | 31.9 | 35.3 | 12.1 |
| ● Sprattus sprattus | 23.3 | 29.3 | 17.2 | 30.2 | 45.7 | 25 | 15.5 | 13.8 | 20.7 | 31.9 | 34.5 | 12.9 |
| ● Sprattus muelleri | 19.8 | 31.9 | 19 | 29.3 | 46.6 | 25 | 14.7 | 13.8 | 15.5 | 34.5 | 31 | 19 |
| ● Sprattus antipodum | 19.8 | 31.9 | 18.1 | 30.2 | 46.6 | 25 | 14.7 | 13.8 | 16.4 | 33.6 | 29.3 | 20.7 |
| ○ Potamalosa richmondia | 22.4 | 33.6 | 18.1 | 25.9 | 47.4 | 24.1 | 15.5 | 12.9 | 26.7 | 26.7 | 38.8 | 7.8 |
| ● Hyperlophus vittatus | 20.7 | 35.3 | 18.1 | 25.9 | 47.4 | 24.1 | 15.5 | 12.9 | 31 | 29.3 | 30.2 | 9.5 |
| ● Ethmidium maculatum | 24.1 | 30.2 | 19 | 26.7 | 47.4 | 24.1 | 15.5 | 12.9 | 19 | 31.9 | 36.2 | 12.9 |
| ● Jenkinsia lamprotaenia | 22.4 | 29.3 | 19.8 | 28.4 | 45.7 | 27.6 | 12.9 | 13.8 | 23.3 | 36.2 | 23.3 | 17.2 |
| ● Spratelloides delicatulus | 23.3 | 33.6 | 15.5 | 27.6 | 47.4 | 26.7 | 12.9 | 12.9 | 19.8 | 44 | 17.2 | 19 |
| ● Spratelloides gracilis | 25.9 | 28.4 | 18.1 | 27.6 | 47.4 | 26.7 | 13.8 | 12.1 | 24.1 | 30.2 | 31.9 | 13.8 |
| ● Etrumeus micropus | 25 | 29.3 | 18.1 | 27.6 | 47.4 | 26.7 | 14.7 | 11.2 | 24.1 | 29.3 | 35.3 | 11.2 |
| ○ Ilisha africana | 19.8 | 34.5 | 24.1 | 21.6 | 44 | 29.3 | 15.5 | 11.2 | 16.4 | 36.2 | 43.1 | 4.3 |
| ○ Pellona flavipinnis | 22.4 | 29.3 | 25.9 | 22.4 | 44 | 28.4 | 15.5 | 12.1 | 13.8 | 37.1 | 42.2 | 6.9 |
| ○ Ilisha elongata | 24.1 | 29.3 | 23.3 | 23.3 | 44 | 28.4 | 15.5 | 12.1 | 13.8 | 34.5 | 44 | 7.8 |
| □ Pellona ditchela | 23.3 | 30.2 | 25 | 21.6 | 44.8 | 27.6 | 15.5 | 12.1 | 14.7 | 33.6 | 44.8 | 6.9 |
| □ Anchoviella sp. LBP 2297 | 19.8 | 32.8 | 19.8 | 27.6 | 44.8 | 25 | 17.2 | 12.9 | 20.7 | 33.6 | 38.8 | 6.9 |
| □ Lycengraulis grossidens | 21.6 | 31 | 20.7 | 26.7 | 44.8 | 25 | 17.2 | 12.9 | 28.4 | 27.6 | 37.9 | 6 |
| ○ Amazonsprattus scintilla | 25 | 26.7 | 20.7 | 27.6 | 45.7 | 24.1 | 17.2 | 12.9 | 24.1 | 27.6 | 31.9 | 16.4 |
| ● Engraulis encrasicolus | 23.3 | 29.3 | 19.8 | 27.6 | 44.8 | 25.9 | 16.4 | 12.9 | 20.7 | 31.9 | 28.4 | 19 |
| ● Engraulis japonicus | 23.3 | 29.3 | 19 | 28.4 | 44.8 | 25.9 | 16.4 | 12.9 | 21.6 | 31 | 31.9 | 15.5 |
| ● Stolephorus chinensis | 20.7 | 31.9 | 20.7 | 26.7 | 44.8 | 25 | 17.2 | 12.9 | 33.6 | 23.3 | 33.6 | 9.5 |
| ● Stolephorus waitei | 20.7 | 31.9 | 19.8 | 27.6 | 44.8 | 25 | 17.2 | 12.9 | 33.6 | 23.3 | 34.5 | 8.6 |
| ○ Lycothrissa crocodilus | 23.3 | 29.3 | 25 | 22.4 | 44.8 | 25 | 17.2 | 12.9 | 19 | 33.6 | 42.2 | 5.2 |
| □ Setipinna melanochir | 20.7 | 32.8 | 20.7 | 25.9 | 44 | 25.9 | 17.2 | 12.9 | 17.2 | 38.8 | 35.3 | 8.6 |
| ● Coilia reynaldi | 20.7 | 31.9 | 22.4 | 25 | 45.7 | 24.1 | 17.2 | 12.9 | 18.1 | 30.2 | 44 | 7.8 |
| ● Thryssa baelama | 22.4 | 31 | 18.1 | 28.4 | 46.6 | 23.3 | 17.2 | 12.9 | 16.4 | 33.6 | 37.9 | 12.1 |
| ○ Coilia lindmani | 24.1 | 29.3 | 20.7 | 25.9 | 44.8 | 24.1 | 18.1 | 12.9 | 19 | 29.3 | 46.6 | 5.2 |
| □ Coilia ectenes | 21.6 | 31.9 | 21.6 | 25 | 44.8 | 24.1 | 18.1 | 12.9 | 22.4 | 24.1 | 48.3 | 5.2 |
| ● Coilia nasus | 22.4 | 31 | 21.6 | 25 | 44.8 | 24.1 | 18.1 | 12.9 | 23.3 | 23.3 | 48.3 | 5.2 |
| ● Denticeps clupeoides | 26.7 | 25.9 | 25.9 | 21.6 | 44 | 28.4 | 16.4 | 11.2 | 28.4 | 30.2 | 38.8 | 2.6 |

ND3

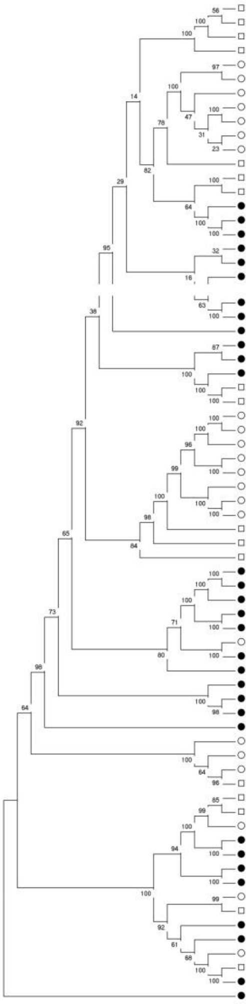


| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ Tenualosa thibaudeaui | 21.5 | 28.5 | 27.8 | 22.2 | 40.7 | 28.5 | 15 | 15.9 | 20.4 | 33 | 32.2 | 14.3 |
| □ Tenualosa ilisha | 19.6 | 30.9 | 26.7 | 22.8 | 40.7 | 28.3 | 14.8 | 16.3 | 18.9 | 35.2 | 31.3 | 14.6 |
| □ Tenualosa toli | 21.7 | 28.5 | 28.3 | 21.5 | 40.2 | 28.5 | 15.4 | 15.9 | 20 | 34.3 | 32.4 | 13.3 |
| □ Gudusia chapra | 20.4 | 29.8 | 29.3 | 20.4 | 41.3 | 27.6 | 15.7 | 15.4 | 18.5 | 34.3 | 42 | 5.2 |
| ○ Potamothrissa obtusirostris | 18.9 | 28.5 | 29.8 | 22.8 | 40.2 | 28.3 | 15.4 | 16.1 | 18.3 | 34.1 | 38.5 | 9.1 |
| ○ Potamothrissa acutirostris | 20.9 | 26.5 | 30.2 | 22.4 | 40.7 | 27.8 | 15.7 | 15.9 | 20.7 | 33 | 40.4 | 5.9 |
| ○ Microthrissa congica | 18.5 | 28.9 | 30 | 22.6 | 40.4 | 27.8 | 15.4 | 16.3 | 16.3 | 35.2 | 39.8 | 8.7 |
| ○ Pellonula vorax | 18.9 | 28.5 | 29.6 | 23 | 40.7 | 27.8 | 15.4 | 16.1 | 19.3 | 33.5 | 40.2 | 7 |
| ○ Pellonula leonensis | 18.3 | 29.1 | 29.1 | 23.5 | 40.7 | 27.8 | 15.4 | 16.1 | 19.6 | 32 | 42 | 6.5 |
| ○ Odaxothrissa losera | 20.2 | 26.7 | 31.3 | 21.7 | 40.7 | 28 | 15.4 | 15.9 | 16.3 | 36.1 | 41.7 | 5.9 |
| ○ Microthrissa royauxi | 18.5 | 28.7 | 30 | 22.8 | 40.4 | 28 | 15.4 | 16.1 | 18.3 | 34.6 | 37.8 | 9.3 |
| □ Ethmalosa fimbriata | 18.9 | 28.3 | 30.4 | 22.4 | 40.9 | 27.8 | 15.2 | 16.1 | 17.6 | 31.7 | 42 | 8.7 |
| □ Dorosoma cepedianum | 18.5 | 28.3 | 30.4 | 22.8 | 40.4 | 27.8 | 15.7 | 16.1 | 19.6 | 33.7 | 32.8 | 13.9 |
| □ Dorosoma petenense | 20.9 | 26.5 | 30.4 | 22.2 | 40.4 | 28 | 15.4 | 16.1 | 21.5 | 30 | 39.3 | 9.1 |
| ● Sardinella maderensis | 19.8 | 28.3 | 28.7 | 23.3 | 40.9 | 27.8 | 15.4 | 15.9 | 17.2 | 32.6 | 35 | 15.2 |
| ● Sardinella albella | 20 | 27.8 | 29.3 | 22.8 | 40.7 | 28 | 15.4 | 15.9 | 15.4 | 36.7 | 27.2 | 20.7 |
| ● sardinella gibbosa | 19.6 | 28.3 | 29.6 | 22.6 | 40.9 | 28 | 15.2 | 15.9 | 16.1 | 36.5 | 26.1 | 21.3 |
| ● Harengula jaguana | 20.4 | 28.5 | 27 | 24.1 | 40.4 | 28.5 | 15.4 | 15.7 | 23 | 34.8 | 23 | 19.1 |
| ● Sardinella longiceps | 18.9 | 29.6 | 28 | 23.5 | 40.7 | 27.8 | 15.4 | 16.1 | 18.5 | 35.7 | 27.4 | 18.5 |
| ● Nematalosa japonica | 19.3 | 28.7 | 29.1 | 22.8 | 40.7 | 28 | 15.2 | 16.1 | 24.3 | 28.7 | 27.8 | 19.1 |
| ● Clupanodon thrissa | 19.6 | 28.7 | 28.7 | 23 | 40.7 | 27.8 | 15.4 | 16.1 | 18.7 | 33.9 | 30.7 | 16.7 |
| ● Konosirus punctatus | 20.4 | 27.4 | 30 | 22.2 | 40.2 | 28 | 15.7 | 16.1 | 20.4 | 31.7 | 25.9 | 22 |
| ● Escualosa thoracata | 18.7 | 30.2 | 28 | 23 | 40.4 | 28.7 | 15.4 | 15.4 | 23 | 36.7 | 20.2 | 20 |
| ● Sardina pilchardus | 21.1 | 26.3 | 27.6 | 25 | 40.9 | 27.2 | 15.4 | 16.5 | 21.1 | 30.9 | 21.7 | 26.3 |
| ● Sardinops melanostictus | 20.4 | 27.8 | 28.9 | 22.8 | 40.7 | 27.6 | 15.7 | 16.1 | 22.4 | 30.9 | 23.5 | 23.3 |
| ● Brevoortia tyrannus | 20.9 | 27.4 | 29.3 | 22.4 | 40.4 | 27.6 | 16.1 | 15.9 | 21.3 | 29.6 | 30.7 | 18.5 |
| □ Alosa alosa | 21.1 | 26.7 | 29.1 | 23 | 40.4 | 27.6 | 16.1 | 15.9 | 17.8 | 34.1 | 34.6 | 13.5 |
| □ Alosa pseudoharengus | 20.7 | 27.2 | 29.3 | 22.8 | 40.4 | 27.6 | 15.9 | 16.1 | 19.6 | 32.6 | 36.5 | 11.3 |
| ○ Clupeichthys goniognathus | 20.9 | 28.7 | 26.5 | 23.9 | 40.4 | 28.5 | 15.7 | 15.4 | 21.5 | 32.6 | 35 | 10.9 |
| ○ Clupeichthys aesarnensis | 20.4 | 29.3 | 27.2 | 23 | 39.8 | 28.7 | 15.7 | 15.9 | 23 | 32.6 | 33.7 | 10.7 |
| ○ Clupeichthys perakensis | 20 | 29.8 | 27.8 | 22.4 | 39.8 | 29.1 | 15.4 | 15.7 | 20.2 | 36.3 | 31.7 | 11.7 |
| ○ Clupeoides sp. Chao Phraya | 21.3 | 27.4 | 30.7 | 20.7 | 40.4 | 28 | 16.1 | 15.4 | 19.6 | 30 | 47.6 | 2.8 |
| ○ Clupeoides borneensis | 19.3 | 29.1 | 30.4 | 21.1 | 40.7 | 27.6 | 16.1 | 15.7 | 15.7 | 34.1 | 42.2 | 8 |
| ○ Sundasalanx praecox | 20.9 | 28 | 28.5 | 22.6 | 39.8 | 28.9 | 15.7 | 15.7 | 23.3 | 32.4 | 37.2 | 7.2 |
| ○ Sundasalanx sp. Chao Phraya | 20.7 | 28.5 | 28.9 | 22 | 39.8 | 28.7 | 15.9 | 15.7 | 23 | 30.9 | 41.5 | 4.6 |
| ○ Sundasalanx mekongensis | 22 | 26.7 | 30.7 | 20.7 | 39.8 | 28.5 | 16.1 | 15.7 | 22.6 | 31.7 | 40 | 5.7 |
| □ Ehirava fluviatilis | 19.8 | 28.7 | 29.1 | 22.4 | 40.2 | 28.3 | 15.7 | 15.9 | 23 | 29.8 | 43.7 | 3.5 |
| □ Gilchristella aestuaria | 19.8 | 28.7 | 28.9 | 22.6 | 40.7 | 28.3 | 15.2 | 15.9 | 23 | 33.5 | 34.3 | 9.1 |
| □ Clupeonella cultriventris | 19.8 | 28.7 | 26.7 | 24.8 | 40.2 | 28.5 | 15.4 | 15.9 | 27.4 | 29.8 | 24.8 | 18 |
| ● Clupea harengus | 18.5 | 29.3 | 29.1 | 23 | 40.2 | 28.3 | 15.7 | 15.9 | 19.1 | 33.9 | 28.5 | 18.5 |
| ● Clupea pallasii | 18.7 | 29.3 | 28.9 | 23 | 40.2 | 28.3 | 15.7 | 15.9 | 19.1 | 33.7 | 28.5 | 18.7 |
| ● Sprattus sprattus | 19.1 | 28.7 | 28.3 | 23.9 | 40.4 | 27.8 | 15.7 | 16.1 | 20.9 | 31.3 | 27 | 20.9 |
| ● Sprattus muelleri | 20.2 | 28.3 | 27.6 | 23.9 | 40.2 | 28 | 15.4 | 16.3 | 18.9 | 33.7 | 27.8 | 19.6 |



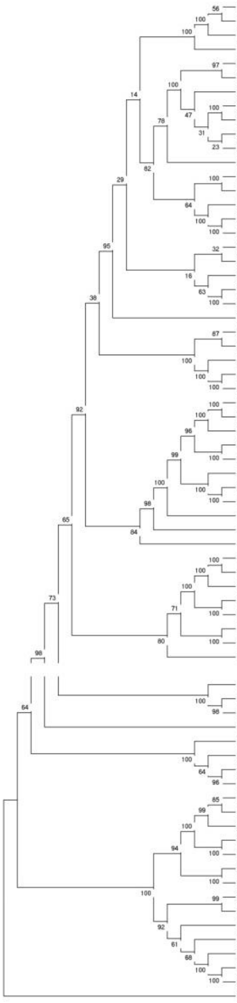
| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ● *Sprattus antipodum* | 20 | 28.5 | 27.4 | 24.1 | 40.4 | 27.8 | 15.4 | 16.3 | 18 | 34.6 | 27.4 | 20 |
| ○ *Potamalosa richmondia* | 20.7 | 27 | 30.2 | 22.2 | 39.8 | 28 | 16.1 | 16.1 | 18.9 | 33.3 | 41.7 | 6.1 |
| ● *Hyperlophus vittatus* | 20.2 | 27.8 | 28.3 | 23.7 | 39.8 | 28 | 16.1 | 16.1 | 24.1 | 30 | 30 | 15.9 |
| ● *Ethmidium maculatum* | 21.5 | 25.4 | 30.9 | 22.2 | 39.6 | 28.7 | 15.4 | 16.3 | 14.6 | 37.2 | 35 | 13.3 |
| ● *Jenkinsia lamprotaenia* | 18.9 | 31.1 | 26.1 | 23.9 | 41.1 | 28.7 | 14.8 | 15.4 | 20.4 | 41.3 | 20.2 | 18 |
| ● *Spratelloides delicatulus* | 20 | 30.7 | 25.4 | 23.9 | 39.8 | 29.1 | 15.2 | 15.9 | 20.4 | 41.7 | 19.3 | 18.5 |
| ● *Spratelloides gracilis* | 22.4 | 27.2 | 27.6 | 22.8 | 40.4 | 28.5 | 15.4 | 15.7 | 22.2 | 33.7 | 25.4 | 18.7 |
| ● *Etrumeus micropus* | 19.8 | 29.3 | 30 | 20.9 | 40.4 | 27.6 | 15.9 | 16.1 | 25.2 | 29.8 | 34.6 | 10.4 |
| ○ *Ilisha africana* | 21.1 | 27.2 | 32.6 | 19.1 | 40 | 28.3 | 16.3 | 15.4 | 22.6 | 32.2 | 39.3 | 5.9 |
| ○ *Pellona flavipinnis* | 18.9 | 30 | 31.7 | 19.3 | 39.3 | 28.3 | 16.7 | 15.7 | 13.5 | 37 | 41.5 | 8 |
| ○ *Ilisha elongata* | 19.6 | 28.5 | 32.2 | 19.8 | 40 | 28 | 16.3 | 15.7 | 14.2 | 37 | 44.9 | 3.9 |
| □ *Pellona ditchela* | 18.7 | 29.3 | 32 | 20 | 39.8 | 28 | 16.5 | 15.7 | 18.3 | 33.3 | 43.9 | 4.6 |
| □ *Anchoviella sp. LBP 2297* | 21.3 | 27.2 | 30.2 | 21.3 | 39.1 | 29.6 | 16.5 | 14.8 | 23.7 | 28.3 | 40.9 | 7.2 |
| □ *Lycengraulis grossidens* | 21.3 | 27.6 | 29.3 | 21.7 | 39.1 | 29.6 | 16.5 | 14.8 | 27 | 27.4 | 36.3 | 9.3 |
| ○ *Amazonsprattus scintilla* | 23.5 | 25 | 29.3 | 22.2 | 38.9 | 29.6 | 16.7 | 14.8 | 23.9 | 29.1 | 29.1 | 17.8 |
| ● *Engraulis encrasicolus* | 20 | 28.9 | 29.3 | 21.7 | 38.9 | 29.8 | 16.3 | 15 | 23.3 | 32 | 29.6 | 15.2 |
| ● *Engraulis japonicus* | 20.2 | 28.7 | 29.3 | 21.7 | 38.9 | 29.8 | 16.3 | 15 | 25 | 30.2 | 29.8 | 15 |
| ● *Stolephorus chinensis* | 21.7 | 27.4 | 29.6 | 21.3 | 38.9 | 30 | 16.1 | 15 | 28.9 | 27.8 | 35.9 | 7.4 |
| ● *Stolephorus waitei* | 21.3 | 27.8 | 29.6 | 21.3 | 38.9 | 30 | 16.1 | 15 | 28.7 | 27.4 | 34.3 | 9.6 |
| ○ *Lycothrissa crocodilus* | 19.3 | 28.9 | 31.1 | 20.7 | 38.5 | 29.8 | 16.5 | 15.2 | 19.6 | 35.4 | 41.7 | 3.3 |
| □ *Setipinna melanochir* | 17.6 | 30.9 | 31.1 | 20.4 | 39.1 | 29.1 | 16.5 | 15.2 | 18.5 | 38.3 | 37.4 | 5.9 |
| ● *Coilia reynaldi* | 20.2 | 28 | 30.4 | 21.3 | 39.1 | 29.1 | 16.7 | 15 | 18 | 35.4 | 38.3 | 8.3 |
| ● *Thryssa baelama* | 21.5 | 27.2 | 30 | 21.3 | 39.8 | 28.9 | 15.9 | 15.4 | 21.1 | 31.5 | 35.2 | 12.2 |
| ○ *Coilia lindmani* | 20.4 | 27.4 | 32.2 | 20 | 38.9 | 29.1 | 16.3 | 15.7 | 18 | 33.5 | 43.9 | 4.6 |
| □ *Coilia ectenes* | 21.3 | 26.7 | 32.2 | 19.8 | 38.9 | 29.1 | 16.3 | 15.7 | 17.6 | 33.5 | 45.9 | 3 |
| ● *Coilia nasus* | 21.7 | 26.3 | 32.2 | 19.8 | 38.9 | 29.1 | 16.3 | 15.7 | 18.5 | 32.6 | 45 | 3.9 |
| ● *Denticeps clupeoides* | 22.8 | 24.3 | 33.9 | 18.9 | 41.7 | 27.4 | 16.3 | 14.6 | 24.1 | 23.5 | 48.3 | 4.1 |

ND4



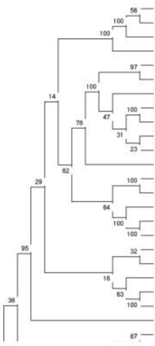
| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 19.2 | 35.4 | 20.2 | 25.3 | 39.4 | 32.3 | 14.1 | 14.1 | 22.2 | 25.3 | 42.4 | 10.1 |
| □ *Tenualosa ilisha* | 21.2 | 33.3 | 20.2 | 25.3 | 38.4 | 33.3 | 14.1 | 14.1 | 18.2 | 31.3 | 36.4 | 14.1 |
| □ *Tenualosa toli* | 20.2 | 34.3 | 20.2 | 25.3 | 39.4 | 32.3 | 13.1 | 15.2 | 13.1 | 34.3 | 32.3 | 20.2 |
| □ *Gudusia chapra* | 22.2 | 32.3 | 23.2 | 22.2 | 38.4 | 33.3 | 14.1 | 14.1 | 19.2 | 31.3 | 46.5 | 3 |
| ○ *Potamothrissa obtusirostris* | 18.2 | 34.3 | 22.2 | 25.3 | 38.4 | 31.3 | 14.1 | 16.2 | 19.2 | 39.4 | 36.4 | 5.1 |
| ○ *Potamothrissa acutirostris* | 18.2 | 34.3 | 22.2 | 25.3 | 38.4 | 31.3 | 14.1 | 16.2 | 28.3 | 30.3 | 35.4 | 6.1 |
| ○ *Microthrissa congica* | 18.2 | 34.3 | 22.2 | 25.3 | 39.4 | 30.3 | 13.1 | 17.2 | 18.2 | 39.4 | 36.4 | 6.1 |
| ○ *Pellonula vorax* | 19.2 | 33.3 | 22.2 | 25.3 | 39.4 | 30.3 | 14.1 | 16.2 | 24.2 | 31.3 | 37.4 | 7.1 |
| ○ *Pellonula leonensis* | 19.2 | 33.3 | 22.2 | 25.3 | 39.4 | 30.3 | 14.1 | 16.2 | 23.2 | 33.3 | 36.4 | 7.1 |
| ○ *Odaxothrissa losera* | 18.2 | 34.3 | 22.2 | 25.3 | 39.4 | 29.3 | 14.1 | 17.2 | 22.2 | 34.3 | 35.4 | 8.1 |
| ○ *Microthrissa royauxi* | 19.2 | 33.3 | 22.2 | 25.3 | 39.4 | 30.3 | 14.1 | 16.2 | 24.2 | 32.3 | 34.3 | 9.1 |
| □ *Ethmalosa fimbriata* | 21.2 | 31.3 | 22.2 | 25.3 | 39.4 | 30.3 | 14.1 | 16.2 | 23.2 | 31.3 | 35.4 | 10.1 |
| □ *Dorosoma cepedianum* | 19.2 | 33.3 | 22.2 | 25.3 | 39.4 | 31.3 | 14.1 | 15.2 | 17.2 | 36.4 | 29.3 | 17.2 |
| □ *Dorosoma petenense* | 18.2 | 34.3 | 21.2 | 26.3 | 39.4 | 31.3 | 14.1 | 15.2 | 17.2 | 32.3 | 39.4 | 11.1 |
| ● *Sardinella maderensis* | 21.2 | 31.3 | 22.2 | 25.3 | 39.4 | 30.3 | 14.1 | 16.2 | 24.2 | 30.3 | 30.3 | 15.2 |
| ● *Sardinella albella* | 20.2 | 32.3 | 22.2 | 25.3 | 39.4 | 30.3 | 14.1 | 16.2 | 15.2 | 35.4 | 29.3 | 20.2 |
| ● *sardinella gibbosa* | 20.2 | 32.3 | 22.2 | 25.3 | 39.4 | 30.3 | 14.1 | 16.2 | 14.1 | 35.4 | 30.3 | 20.2 |
| ● *Harengula jaguana* | 19.2 | 29.3 | 20.2 | 31.3 | 39.4 | 26.3 | 15.2 | 19.2 | 15.2 | 37.4 | 34.3 | 13.1 |
| ● *Sardinella longiceps* | 18.2 | 33.3 | 23.2 | 25.3 | 39.4 | 29.3 | 14.1 | 17.2 | 17.2 | 38.4 | 32.3 | 12.1 |
| ● *Nematalosa japonica* | 19.2 | 34.3 | 21.2 | 25.3 | 40.4 | 29.3 | 14.1 | 16.2 | 22.2 | 32.3 | 28.3 | 17.2 |
| ● *Clupanodon thrissa* | 17.2 | 35.4 | 20.2 | 27.3 | 39.4 | 30.3 | 14.1 | 16.2 | 25.3 | 33.3 | 31.3 | 10.1 |
| ● *Konosirus punctatus* | 17.2 | 33.3 | 23.2 | 26.3 | 39.4 | 30.3 | 14.1 | 16.2 | 20.2 | 34.3 | 33.3 | 12.1 |
| ● *Escualosa thoracata* | 20.2 | 34.3 | 18.2 | 27.3 | 39.4 | 31.3 | 14.1 | 15.2 | 17.2 | 43.4 | 25.3 | 14.1 |
| ● *Sardina pilchardus* | 24.2 | 31.3 | 18.2 | 26.3 | 41.4 | 29.3 | 14.1 | 15.2 | 17.2 | 41.4 | 20.2 | 21.2 |
| ● *Sardinops melanostictus* | 20.2 | 33.3 | 19.2 | 27.3 | 41.4 | 28.3 | 13.1 | 17.2 | 26.3 | 30.3 | 26.3 | 17.2 |
| ● *Brevoortia tyrannus* | 19.2 | 34.3 | 22.2 | 24.2 | 40.4 | 29.3 | 14.1 | 16.2 | 21.2 | 32.3 | 36.4 | 10.1 |
| □ *Alosa alosa* | 20.2 | 33.3 | 22.2 | 24.2 | 40.4 | 29.3 | 15.2 | 15.2 | 21.2 | 32.3 | 34.3 | 12.1 |
| □ *Alosa pseudoharengus* | 20.2 | 33.3 | 22.2 | 24.2 | 40.4 | 29.3 | 14.1 | 16.2 | 22.2 | 31.3 | 37.4 | 9.1 |
| ○ *Clupeichthys goniognathus* | 17.2 | 30.3 | 26.3 | 26.3 | 39.4 | 30.3 | 13.1 | 17.2 | 25.3 | 35.4 | 22.2 | 17.2 |
| ○ *Clupeichthys aesarnensis* | 17.2 | 30.3 | 25.3 | 27.3 | 37.4 | 31.3 | 12.1 | 19.2 | 24.2 | 35.4 | 29.3 | 11.1 |
| ○ *Clupeichthys perakensis* | 20.2 | 29.3 | 23.2 | 27.3 | 37.4 | 32.3 | 13.1 | 17.2 | 22.2 | 36.4 | 30.3 | 11.1 |
| ○ *Clupeoides sp. Chao Phraya* | 21.2 | 27.3 | 26.3 | 25.3 | 40.4 | 29.3 | 14.1 | 16.2 | 19.2 | 31.3 | 46.5 | 3 |
| ○ *Clupeoides borneensis* | 19.2 | 29.3 | 25.3 | 26.3 | 39.4 | 30.3 | 14.1 | 16.2 | 21.2 | 29.3 | 42.4 | 7.1 |
| ○ *Sundasalanx praecox* | 25.3 | 28.3 | 22.2 | 24.2 | 41.4 | 31.3 | 12.1 | 15.2 | 21.2 | 34.3 | 38.4 | 6.1 |
| ○ *Sundasalanx sp. Chao Phraya* | 22.2 | 30.3 | 21.2 | 26.3 | 41.4 | 31.3 | 12.1 | 15.2 | 28.3 | 32.3 | 36.4 | 3 |
| ○ *Sundasalanx mekongensis* | 22.2 | 29.3 | 25.3 | 23.2 | 41.4 | 30.3 | 13.1 | 15.2 | 27.3 | 31.3 | 38.4 | 3 |
| □ *Ehirava fluviatilis* | 20.2 | 30.3 | 21.2 | 28.3 | 39.4 | 31.3 | 12.1 | 17.2 | 17.2 | 34.3 | 41.4 | 7.1 |
| □ *Gilchristella aestuaria* | 20.2 | 32.3 | 21.2 | 26.3 | 40.4 | 29.3 | 14.1 | 16.2 | 23.2 | 31.3 | 36.4 | 9.1 |
| □ *Clupeonella cultriventris* | 20.2 | 34.3 | 21.2 | 24.2 | 39.4 | 30.3 | 14.1 | 16.2 | 23.2 | 29.3 | 32.3 | 15.2 |
| ● *Clupea harengus* | 20.2 | 31.3 | 23.2 | 25.3 | 41.4 | 29.3 | 16.2 | 13.1 | 20.2 | 35.4 | 33.3 | 11.1 |
| ● *Clupea pallasii* | 20.2 | 31.3 | 23.2 | 25.3 | 41.4 | 29.3 | 16.2 | 13.1 | 20.2 | 35.4 | 33.3 | 11.1 |
| ● *Sprattus sprattus* | 19.2 | 32.3 | 23.2 | 25.3 | 41.4 | 29.3 | 16.2 | 13.1 | 15.2 | 40.4 | 30.3 | 14.1 |
| ● *Sprattus muelleri* | 21.2 | 31.3 | 24.2 | 23.2 | 40.4 | 30.3 | 16.2 | 13.1 | 13.1 | 37.4 | 34.3 | 15.2 |
| ● *Sprattus antipodum* | 21.2 | 31.3 | 24.2 | 23.2 | 40.4 | 30.3 | 16.2 | 13.1 | 14.1 | 36.4 | 35.4 | 14.1 |
| ○ *Potamalosa richmondia* | 20.2 | 32.3 | 23.2 | 24.2 | 39.4 | 29.3 | 14.1 | 17.2 | 13.1 | 42.4 | 36.4 | 8.1 |
| ● *Hyperlophus vittatus* | 20.2 | 32.3 | 23.2 | 24.2 | 39.4 | 29.3 | 14.1 | 17.2 | 25.3 | 30.3 | 33.3 | 11.1 |
| ● *Ethmidium maculatum* | 19.2 | 32.3 | 24.2 | 24.2 | 42.4 | 27.3 | 14.1 | 16.2 | 15.2 | 39.4 | 37.4 | 8.1 |
| ● *Jenkinsia lamprotaenia* | 18.2 | 32.3 | 21.2 | 28.3 | 41.4 | 29.3 | 15.2 | 14.1 | 22.2 | 35.4 | 25.3 | 17.2 |
| ● *Spratelloides delicatulus* | 18.2 | 33.3 | 22.2 | 26.3 | 40.4 | 30.3 | 16.2 | 13.1 | 20.2 | 43.4 | 21.2 | 15.2 |
| ● *Spratelloides gracilis* | 19.2 | 30.3 | 24.2 | 26.3 | 40.4 | 29.3 | 14.1 | 16.2 | 28.3 | 30.3 | 26.3 | 15.2 |
| ● *Etrumeus micropus* | 27.3 | 25.3 | 26.3 | 21.2 | 42.4 | 27.3 | 13.1 | 17.2 | 36.4 | 27.3 | 30.3 | 6.1 |
| ○ *Ilisha africana* | 21.2 | 33.3 | 23.2 | 22.2 | 41.4 | 29.3 | 16.2 | 13.1 | 21.2 | 37.4 | 35.4 | 6.1 |
| ○ *Pellona flavipinnis* | 21.2 | 32.3 | 23.2 | 23.2 | 38.4 | 32.3 | 16.2 | 13.1 | 18.2 | 37.4 | 35.4 | 9.1 |
| ○ *Ilisha elongata* | 22.2 | 33.3 | 22.2 | 22.2 | 39.4 | 31.3 | 16.2 | 13.1 | 12.1 | 46.5 | 38.4 | 3 |
| □ *Pellona ditchela* | 22.2 | 32.3 | 22.2 | 23.2 | 39.4 | 31.3 | 16.2 | 13.1 | 14.1 | 42.4 | 39.4 | 4 |
| □ *Anchoviella sp. LBP 2297* | 20.2 | 31.3 | 24.2 | 24.2 | 39.4 | 31.3 | 14.1 | 15.2 | 22.2 | 22.2 | 47.5 | 8.1 |
| □ *Lycengraulis grossidens* | 21.2 | 30.3 | 25.3 | 23.2 | 41.4 | 29.3 | 14.1 | 15.2 | 18.2 | 25.3 | 52.5 | 4 |
| ○ *Amazonsprattus scintilla* | 21.2 | 30.3 | 25.3 | 23.2 | 40.4 | 30.3 | 14.1 | 15.2 | 20.2 | 24.2 | 43.4 | 12.1 |
| ● *Engraulis encrasicolus* | 24.2 | 28.3 | 22.2 | 25.3 | 38.4 | 32.3 | 13.1 | 16.2 | 24.2 | 21.2 | 38.4 | 16.2 |
| ● *Engraulis japonicus* | 23.2 | 29.3 | 23.2 | 24.2 | 38.4 | 32.3 | 14.1 | 15.2 | 23.2 | 21.2 | 40.4 | 15.2 |
| ● *Stolephorus chinensis* | 23.2 | 27.3 | 26.3 | 23.2 | 41.4 | 28.3 | 14.1 | 16.2 | 24.2 | 19.2 | 51.5 | 5.1 |
| ● *Stolephorus waitei* | 24.2 | 26.3 | 26.3 | 23.2 | 41.4 | 28.3 | 14.1 | 16.2 | 24.2 | 18.2 | 49.5 | 8.1 |
| ○ *Lycothrissa crocodilus* | 21.2 | 30.3 | 27.3 | 21.2 | 39.4 | 31.3 | 14.1 | 15.2 | 12.1 | 35.4 | 45.5 | 7.1 |
| □ *Setipinna melanochir* | 23.2 | 28.3 | 25.3 | 23.2 | 39.4 | 31.3 | 14.1 | 15.2 | 18.2 | 35.4 | 36.4 | 10.1 |
| ● *Coilia reynaldi* | 21.2 | 30.3 | 24.2 | 24.2 | 40.4 | 30.3 | 14.1 | 15.2 | 19.2 | 31.3 | 41.4 | 8.1 |
| ● *Thryssa baelama* | 23.2 | 27.3 | 25.3 | 24.2 | 39.4 | 31.3 | 14.1 | 15.2 | 24.2 | 25.3 | 35.4 | 15.2 |
| ○ *Coilia lindmani* | 23.2 | 27.3 | 24.2 | 25.3 | 40.4 | 30.3 | 14.1 | 15.2 | 14.1 | 34.3 | 42.4 | 9.1 |
| □ *Coilia ectenes* | 23.2 | 27.3 | 24.2 | 25.3 | 40.4 | 30.3 | 14.1 | 15.2 | 15.2 | 34.3 | 44.4 | 6.1 |
| ● *Coilia nasus* | 23.2 | 27.3 | 24.2 | 25.3 | 40.4 | 30.3 | 14.1 | 15.2 | 16.2 | 32.3 | 43.4 | 8.1 |
| ● *Denticeps clupeoides* | 22.2 | 31.3 | 24.2 | 22.2 | 37.4 | 31.3 | 17.2 | 14.1 | 24.2 | 31.3 | 42.4 | 2 |

ND4l

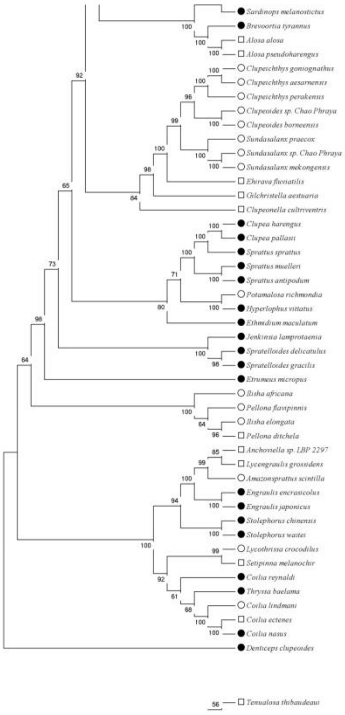


| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 18 | 24.7 | 33.4 | 23.9 | 38.6 | 29.8 | 18.3 | 13.3 | 21.4 | 37.6 | 28 | 12.9 |
| □ *Tenualosa ilisha* | 17.8 | 25.2 | 33.2 | 23.7 | 39.1 | 29.3 | 18.5 | 13.1 | 23.1 | 35.4 | 25.9 | 15.7 |
| □ *Tenualosa toli* | 17.7 | 25.7 | 33.1 | 23.6 | 37.8 | 30.6 | 18.8 | 12.8 | 20.9 | 37.8 | 29.3 | 11.9 |
| □ *Gudusia chapra* | 18.8 | 24.7 | 36.2 | 20.3 | 38.6 | 29.8 | 19 | 12.6 | 19 | 39.6 | 37.2 | 4.3 |
| ○ *Potamothrissa obtusirostris* | 18 | 25 | 32.7 | 24.2 | 38.8 | 30 | 19.3 | 11.9 | 19.3 | 39 | 35.8 | 5.9 |
| ○ *Potamothrissa acutirostris* | 19.5 | 23.9 | 32.7 | 23.9 | 38.8 | 29.6 | 19.5 | 12.1 | 20.5 | 36.7 | 36.5 | 6.4 |
| ○ *Microthrissa congica* | 18.5 | 24.5 | 32.7 | 24.2 | 38.5 | 30.3 | 19.3 | 11.9 | 21.1 | 36.5 | 35.4 | 7 |
| ○ *Pellonula vorax* | 18.5 | 24.2 | 33.2 | 24.1 | 38.6 | 30.1 | 19.3 | 11.9 | 24.5 | 33.2 | 37.2 | 5.1 |
| ○ *Pellonula leonensis* | 18.5 | 24.2 | 33.6 | 23.7 | 38.5 | 30.3 | 19.1 | 12.1 | 20.6 | 37.2 | 36.5 | 5.7 |
| ○ *Odaxothrissa losera* | 19.6 | 23.2 | 34.2 | 22.9 | 38.5 | 30.3 | 19.3 | 11.9 | 19.3 | 37 | 37.8 | 5.9 |
| ○ *Microthrissa royauxi* | 18.7 | 24.1 | 33.6 | 23.7 | 39 | 29.6 | 19.1 | 12.3 | 20.3 | 37.5 | 35.7 | 6.5 |
| □ *Ethmalosa fimbriata* | 19 | 23.9 | 32.9 | 24.2 | 39 | 30.3 | 18.8 | 11.9 | 19.1 | 39.8 | 33.6 | 7.5 |
| □ *Dorosoma cepedianum* | 19.6 | 23.1 | 32.1 | 25.2 | 38.8 | 30.3 | 19 | 11.9 | 21.1 | 38.8 | 31.9 | 8.2 |
| □ *Dorosoma petenense* | 19.3 | 23.4 | 32.9 | 24.4 | 38.8 | 30.3 | 19 | 11.9 | 20.8 | 38 | 35.5 | 5.7 |
| ● *Sardinella maderensis* | 19.3 | 23.7 | 32.4 | 24.5 | 39 | 30 | 18.7 | 12.4 | 19.6 | 39.6 | 26.7 | 14.1 |
| ● *Sardinella albella* | 18.5 | 24.5 | 32.1 | 24.9 | 39 | 30.1 | 19.1 | 11.8 | 20.9 | 39 | 20.8 | 19.3 |
| ● *sardinella gibbosa* | 18.5 | 24.5 | 32.4 | 24.5 | 39 | 30.1 | 19.1 | 11.8 | 20.5 | 39.3 | 22.4 | 17.8 |
| ● *Harengula jaguana* | 19.8 | 23.4 | 32.2 | 24.5 | 38.6 | 29.1 | 19.1 | 13.1 | 23.9 | 38 | 19.8 | 18.3 |
| ● *Sardinella longiceps* | 19.5 | 23.1 | 32.7 | 24.7 | 39 | 30 | 18.8 | 12.3 | 20.5 | 39.9 | 25 | 14.6 |
| ● *Nematalosa japonica* | 19.3 | 23.6 | 32.1 | 25 | 38.8 | 30.1 | 19 | 12.1 | 21.3 | 37.3 | 26.5 | 14.9 |
| ● *Clupanodon thrissa* | 19.3 | 23.6 | 32.6 | 24.5 | 38.3 | 30.8 | 19 | 11.9 | 23.7 | 34.9 | 26.5 | 14.9 |
| ● *Konosirus punctatus* | 19.5 | 23.1 | 32.2 | 25.2 | 38.5 | 30.6 | 19 | 11.9 | 22.4 | 36.7 | 23.2 | 17.7 |
| ● *Escualosa thoracata* | 18.2 | 24.7 | 31.1 | 26 | 38.8 | 30.4 | 18.8 | 11.9 | 23.1 | 39.4 | 20.6 | 16.9 |
| ● *Sardina pilchardus* | 19.6 | 23.7 | 31.1 | 25.5 | 39.6 | 29.5 | 18.3 | 12.6 | 22.9 | 34.4 | 20.9 | 21.8 |
| ● *Sardinops melanostictus* | 19 | 24.2 | 32.4 | 24.4 | 39.3 | 30.1 | 18.3 | 12.3 | 23.1 | 36.8 | 20.8 | 19.3 |
| ● *Brevoortia tyrannus* | 19.1 | 23.6 | 32.2 | 25 | 38.8 | 30.3 | 18.7 | 12.3 | 21.3 | 35.8 | 26.2 | 16.7 |
| □ *Alosa alosa* | 18.7 | 23.7 | 32.6 | 25 | 39 | 30.4 | 18.3 | 12.3 | 18.2 | 39.4 | 32.6 | 9.8 |
| □ *Alosa pseudoharengus* | 18.7 | 24.2 | 32.4 | 24.7 | 38.6 | 30.4 | 18.7 | 12.3 | 20.9 | 37 | 34.2 | 7.9 |
| ○ *Clupeichthys goniognathus* | 19.1 | 24.4 | 33.9 | 22.6 | 39.3 | 29.1 | 19.8 | 11.8 | 20.5 | 40.8 | 29.5 | 9.3 |
| ○ *Clupeichthys aesarnensis* | 20 | 23.7 | 34 | 22.3 | 39 | 29.5 | 19.6 | 11.9 | 18.3 | 41.7 | 32.4 | 7.5 |
| ○ *Clupeichthys perakensis* | 19.3 | 23.9 | 34.5 | 22.3 | 39.4 | 29.1 | 19.6 | 11.8 | 22.7 | 37 | 33.6 | 6.7 |
| ○ *Clupeoides sp. Chao Phraya* | 18.8 | 23.4 | 36.2 | 21.6 | 39.8 | 28.5 | 19.6 | 12.1 | 23.7 | 30.8 | 43 | 2.5 |
| ○ *Clupeoides borneensis* | 19.1 | 23.9 | 35.5 | 21.4 | 39.6 | 28.6 | 19.6 | 12.1 | 17.8 | 36.3 | 40.8 | 5.1 |
| ○ *Sundasalanx praecox* | 20.8 | 22.9 | 33.2 | 23.1 | 39 | 29.8 | 19 | 12.3 | 25.5 | 34 | 33.7 | 6.7 |
| ○ *Sundasalanx sp. Chao Phraya* | 20.3 | 22.9 | 34.7 | 22.1 | 38.5 | 30.1 | 19.5 | 11.9 | 25 | 35 | 34.2 | 5.7 |
| ○ *Sundasalanx mekongensis* | 20 | 23.1 | 35.7 | 21.3 | 38.5 | 30 | 19.6 | 11.9 | 22.9 | 37.5 | 36.3 | 3.3 |
| □ *Ehirava fluviatilis* | 20.1 | 22.7 | 34.9 | 22.3 | 38.8 | 30 | 19.6 | 11.6 | 26.5 | 30.1 | 39.4 | 3.9 |
| □ *Gilchristella aestuaria* | 19.1 | 23.4 | 33.6 | 23.9 | 39.4 | 28.8 | 19.6 | 12.1 | 25 | 34.7 | 31.1 | 9.2 |
| □ *Clupeonella cultriventris* | 18.8 | 24.1 | 32.9 | 24.2 | 39.3 | 29.3 | 19 | 12.4 | 26.7 | 33.7 | 26.4 | 13.3 |
| ● *Clupea harengus* | 18.5 | 23.9 | 33.6 | 24.1 | 39.1 | 29.1 | 19.5 | 12.3 | 23.7 | 33.7 | 25.7 | 16.9 |
| ● *Clupea pallasii* | 18.2 | 24.4 | 33.6 | 23.9 | 39 | 29.5 | 19.5 | 12.1 | 24.1 | 33.7 | 26.2 | 16 |
| ● *Sprattus sprattus* | 19.1 | 23.2 | 33.2 | 24.4 | 38.8 | 29.8 | 19.3 | 12.1 | 24.5 | 34.7 | 24.9 | 15.9 |
| ● *Sprattus muelleri* | 18.7 | 23.4 | 33.9 | 24.1 | 38.5 | 30.6 | 19 | 11.9 | 23.4 | 36.3 | 21.1 | 19.1 |
| ● *Sprattus antipodum* | 18.8 | 23.4 | 33.6 | 24.2 | 38.5 | 30.6 | 19 | 11.9 | 23.2 | 37 | 22.1 | 17.7 |
| ○ *Potamalosa richmondia* | 19.5 | 23.1 | 34.5 | 22.9 | 38.5 | 30 | 19.5 | 12.1 | 19.3 | 36.8 | 37.5 | 6.4 |
| ● *Hyperlophus vittatus* | 20 | 22.9 | 33.2 | 23.9 | 38.5 | 30 | 19.5 | 12.1 | 25 | 33.1 | 28.3 | 13.6 |
| ● *Ethmidium maculatum* | 17.7 | 25.2 | 34.4 | 22.7 | 39.1 | 29.1 | 19.6 | 12.1 | 18.7 | 38.6 | 30.9 | 11.8 |
| ● *Jenkinsia lamprotaenia* | 19.3 | 27 | 28.8 | 24.9 | 39.8 | 29 | 18.8 | 12.4 | 22.1 | 43.9 | 18 | 16 |
| ● *Spratelloides delicatulus* | 18.5 | 27.2 | 28.6 | 25.7 | 40.1 | 28.6 | 19.1 | 12.1 | 21.6 | 45.8 | 15.7 | 16.9 |
| ● *Spratelloides gracilis* | 21.1 | 23.9 | 30.4 | 24.5 | 39.9 | 28.6 | 19.1 | 12.3 | 25.2 | 36.5 | 21.3 | 17 |
| ● *Etrumeus micropus* | 19 | 24.1 | 33.6 | 23.4 | 39.3 | 28.8 | 19.3 | 12.6 | 27.7 | 32.9 | 30 | 9.5 |
| ○ *Ilisha africana* | 18.7 | 25.5 | 35.8 | 20 | 39 | 29.1 | 19.8 | 12.1 | 20 | 39.2 | 37 | 3.8 |
| ○ *Pellona flavipinnis* | 18.3 | 23.9 | 35.8 | 21.9 | 38.1 | 30 | 20 | 11.9 | 15.4 | 39.3 | 40.6 | 4.7 |
| ○ *Ilisha elongata* | 17.3 | 25.2 | 36.8 | 20.6 | 38.9 | 29.5 | 19.8 | 11.8 | 17.2 | 38.6 | 40.8 | 3.4 |
| □ *Pellona ditchela* | 17.3 | 24.4 | 37.2 | 21.1 | 37.8 | 30.3 | 19.6 | 12.3 | 16.2 | 39 | 41.7 | 3.1 |
| □ *Anchoviella sp. LBP 2297* | 19 | 23.4 | 34.2 | 23.4 | 39.4 | 29.3 | 19.6 | 11.6 | 25.5 | 32.9 | 36.3 | 5.2 |
| □ *Lycengraulis grossidens* | 19.5 | 23.1 | 34.9 | 22.6 | 39.1 | 29.6 | 19.3 | 11.9 | 25.2 | 32.7 | 34.4 | 7.7 |
| ○ *Amazonsprattus scintilla* | 20.5 | 22.3 | 33.2 | 24.1 | 39.6 | 29.3 | 19.3 | 11.8 | 26.2 | 31.1 | 31.8 | 11 |
| ● *Engraulis encrasicolus* | 19 | 23.4 | 33.7 | 23.9 | 39.6 | 29.6 | 19 | 11.8 | 25.4 | 32.7 | 26.7 | 15.2 |
| ● *Engraulis japonicus* | 18.8 | 23.6 | 33.4 | 24.2 | 39.6 | 29.8 | 18.8 | 11.8 | 25.5 | 33.2 | 26 | 15.2 |
| ● *Stolephorus chinensis* | 20.5 | 23.2 | 33.2 | 23.1 | 39.8 | 28.5 | 19.8 | 11.9 | 27.7 | 31.3 | 35.7 | 5.4 |
| ● *Stolephorus waitei* | 20.5 | 23.2 | 32.6 | 23.7 | 39.8 | 28.5 | 19.8 | 11.9 | 29 | 30 | 34.2 | 6.9 |
| ○ *Lycothrissa crocodilus* | 18.7 | 25.4 | 35.5 | 20.5 | 38.3 | 30 | 19.8 | 11.9 | 18.8 | 38.3 | 40.4 | 2.5 |
| □ *Setipinna melanochir* | 18.2 | 25.2 | 35.5 | 21.1 | 38 | 30.4 | 19.6 | 11.9 | 17.3 | 42.2 | 34 | 6.4 |
| ● *Coilia reynaldi* | 19.5 | 23.6 | 35.2 | 21.8 | 38.8 | 29.3 | 20.1 | 11.8 | 22.7 | 31.8 | 39.8 | 5.7 |
| ● *Thryssa baelama* | 19.8 | 23.2 | 34.4 | 22.6 | 38.8 | 30 | 19.5 | 11.8 | 21.1 | 35.4 | 32.4 | 11.1 |
| ○ *Coilia lindmani* | 18.8 | 24.2 | 34.9 | 22.1 | 38.6 | 29.5 | 20 | 11.9 | 19 | 35.4 | 41.7 | 3.9 |
| □ *Coilia ectenes* | 19.1 | 23.6 | 35.4 | 21.9 | 38.5 | 29.5 | 20.1 | 11.9 | 22.1 | 32.9 | 41.4 | 3.6 |
| ● *Coilia nasus* | 19.5 | 23.2 | 35.7 | 21.6 | 38.6 | 29.3 | 20.1 | 11.9 | 21.9 | 32.9 | 41.9 | 3.3 |
| ● *Denticeps clupeoides* | 21.3 | 22.8 | 35.2 | 20.7 | 40.7 | 27.5 | 20 | 11.8 | 24.3 | 29 | 43.9 | 2.8 |

ND5



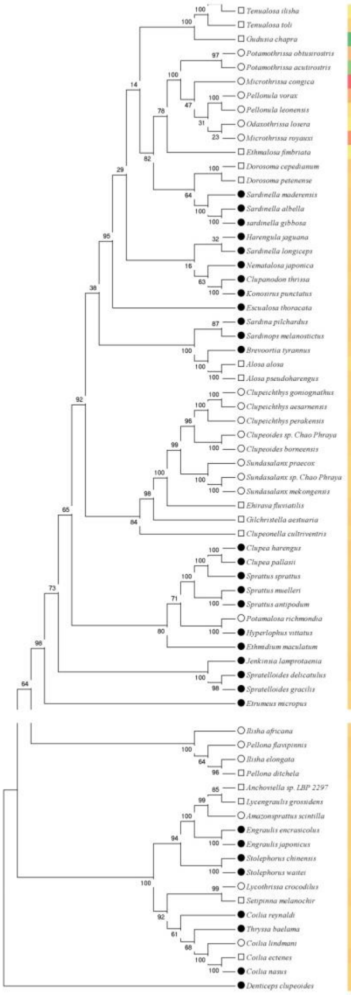
| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 29 | 17 | 14.2 | 39.8 | 45.5 | 21 | 13.1 | 20.5 | 35.2 | 13.6 | 19.9 | 31.3 |
| □ *Tenualosa ilisha* | 29.5 | 15.9 | 15.9 | 38.6 | 42 | 24.4 | 12.5 | 21 | 34.1 | 11.9 | 17.6 | 36.4 |
| □ *Tenualosa toli* | 30.1 | 15.9 | 15.3 | 38.6 | 43.8 | 22.7 | 12.5 | 21 | 35.2 | 11.9 | 13.6 | 39.2 |
| □ *Gudusia chapra* | 34.7 | 11.4 | 14.8 | 39.2 | 45.5 | 21 | 12.5 | 21 | 41.5 | 4.5 | 17.6 | 36.4 |
| ○ *Potamothrissa obtusirostris* | 29 | 13.6 | 14.2 | 43.2 | 46 | 23.9 | 11.9 | 18.2 | 44.3 | 5.7 | 19.3 | 30.7 |
| ○ *Potamothrissa acutirostris* | 31.8 | 10.8 | 15.9 | 41.5 | 45.5 | 23.3 | 12.5 | 18.8 | 42.6 | 8 | 17 | 32.4 |
| ○ *Microthrissa congica* | 30.7 | 11.9 | 13.1 | 44.3 | 46 | 22.2 | 12.5 | 19.3 | 40.9 | 5.7 | 17 | 36.4 |
| ○ *Pellonula vorax* | 33 | 10.2 | 13.6 | 43.2 | 46 | 22.7 | 12.5 | 18.8 | 42.6 | 6.8 | 19.3 | 31.3 |
| ○ *Pellonula leonensis* | 33.5 | 9.7 | 14.2 | 42.6 | 46 | 22.7 | 12.5 | 18.8 | 40.3 | 6.3 | 16.5 | 36.9 |
| ○ *Odaxothrissa losera* | 34.1 | 10.2 | 10.8 | 44.9 | 46.6 | 21 | 13.6 | 18.8 | 43.2 | 4.5 | 18.8 | 33.5 |
| ○ *Microthrissa royauxi* | 33.5 | 9.7 | 13.6 | 43.2 | 46 | 22.7 | 12.5 | 18.8 | 45.5 | 3.4 | 17.6 | 33.5 |
| □ *Ethmalosa fimbriata* | 32.4 | 11.9 | 13.1 | 42.6 | 44.9 | 22.2 | 13.1 | 19.9 | 38.6 | 9.1 | 19.9 | 32.4 |
| □ *Dorosoma cepedianum* | 32.4 | 11.4 | 15.3 | 40.9 | 45.5 | 23.9 | 11.9 | 18.8 | 38.6 | 7.4 | 22.2 | 31.8 |
| □ *Dorosoma petenense* | 34.7 | 10.2 | 14.8 | 40.3 | 46 | 23.3 | 11.4 | 19.3 | 39.8 | 5.7 | 21.6 | 33 |
| ● *Sardinella maderensis* | 28.4 | 13.6 | 15.9 | 42 | 43.8 | 24.4 | 11.9 | 19.9 | 29.5 | 15.3 | 17.6 | 37.5 |
| ● *Sardinella albella* | 27.8 | 14.2 | 14.8 | 43.2 | 44.9 | 23.9 | 11.9 | 19.3 | 27.3 | 18.2 | 15.9 | 38.6 |
| ● *sardinella gibbosa* | 27.3 | 14.8 | 14.8 | 43.2 | 44.9 | 23.9 | 11.9 | 19.3 | 26.1 | 19.3 | 15.9 | 38.6 |
| ● *Harengula jaguana* | 27.8 | 15.3 | 18.2 | 38.6 | 44.9 | 22.7 | 14.2 | 18.2 | 29.5 | 23.3 | 21 | 26.1 |
| ● *Sardinella longiceps* | 28.4 | 15.9 | 13.1 | 42.6 | 46 | 21.6 | 11.4 | 21 | 35.8 | 15.3 | 14.2 | 34.7 |
| ● *Nematalosa japonica* | 25.6 | 19.9 | 14.2 | 40.3 | 44.9 | 24.4 | 13.1 | 17.6 | 34.1 | 17.6 | 18.2 | 30.1 |
| ● *Clupanodon thrissa* | 27.3 | 16.5 | 14.2 | 42 | 45.5 | 25 | 13.1 | 16.5 | 30.1 | 13.6 | 26.1 | 30.1 |
| ● *Konosirus punctatus* | 28.4 | 17.6 | 11.9 | 42 | 47.2 | 22.7 | 12.5 | 17.6 | 31.3 | 14.8 | 18.8 | 35.2 |
| ● *Escualosa thoracata* | 27.8 | 16.5 | 15.9 | 39.8 | 48.3 | 21 | 13.1 | 17.6 | 23.3 | 19.9 | 19.3 | 37.5 |
| ● *Sardina pilchardus* | 22.7 | 19.3 | 15.3 | 42.6 | 43.8 | 21.6 | 14.2 | 20.5 | 31.3 | 27.8 | 21 | 19.9 |



| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ● *Sardinops melanostictus* | 26.1 | 18.2 | 17.6 | 38.1 | 43.8 | 27.3 | 12.5 | 16.5 | 29 | 22.7 | 16.5 | 31.8 |
| ● *Brevoortia tyrannus* | 31.8 | 13.1 | 13.1 | 42 | 44.3 | 21.6 | 12.5 | 21.6 | 27.8 | 18.8 | 18.8 | 34.7 |
| □ *Alosa alosa* | 34.7 | 10.8 | 13.1 | 41.5 | 44.3 | 21.6 | 11.9 | 22.2 | 37.5 | 11.9 | 18.8 | 31.8 |
| □ *Alosa pseudoharengus* | 34.7 | 10.8 | 11.4 | 43.2 | 44.3 | 21.6 | 12.5 | 21.6 | 37.5 | 10.2 | 22.7 | 29.5 |
| ○ *Clupeichthys goniognathus* | 36.4 | 11.4 | 10.8 | 41.5 | 43.2 | 25 | 11.9 | 19.9 | 39.8 | 10.8 | 13.6 | 35.8 |
| ○ *Clupeichthys aesarnensis* | 35.2 | 11.9 | 11.4 | 41.5 | 43.8 | 24.4 | 11.4 | 20.5 | 38.6 | 11.4 | 13.6 | 36.4 |
| ○ *Clupeichthys perakensis* | 36.9 | 11.4 | 11.9 | 39.8 | 43.2 | 25 | 11.9 | 19.9 | 39.2 | 9.1 | 15.9 | 35.8 |
| ○ *Clupeoides sp. Chao Phraya* | 36.9 | 9.7 | 13.6 | 39.8 | 46 | 22.2 | 13.1 | 18.8 | 51.7 | 0.6 | 25 | 22.7 |
| ○ *Clupeoides borneensis* | 35.8 | 10.8 | 13.6 | 39.8 | 44.9 | 23.3 | 13.1 | 18.8 | 48.3 | 3.4 | 18.8 | 29.5 |
| ○ *Sundasalanx praecox* | 34.1 | 14.2 | 11.4 | 40.3 | 43.8 | 24.4 | 12.5 | 19.3 | 43.8 | 8.5 | 23.3 | 24.4 |
| ○ *Sundasalanx sp. Chao Phraya* | 34.1 | 14.2 | 11.4 | 40.3 | 46 | 21.6 | 13.1 | 19.3 | 41.5 | 10.2 | 17.6 | 30.7 |
| ○ *Sundasalanx mekongensis* | 36.9 | 11.9 | 11.4 | 39.8 | 46 | 21.6 | 12.5 | 19.9 | 44.9 | 5.1 | 23.3 | 26.7 |
| □ *Ehirava fluviatilis* | 36.4 | 10.8 | 13.1 | 39.8 | 42.6 | 25 | 13.1 | 19.3 | 45.5 | 2.8 | 26.7 | 25 |
| □ *Gilchristella aestuaria* | 30.1 | 15.3 | 14.2 | 40.3 | 42 | 21.6 | 13.6 | 22.7 | 39.8 | 10.2 | 17 | 33 |
| □ *Clupeonella cultriventris* | 29 | 18.2 | 10.8 | 42 | 44.3 | 22.2 | 12.5 | 21 | 36.4 | 13.1 | 25.6 | 25 |
| ● *Clupea harengus* | 28.4 | 15.3 | 12.5 | 43.8 | 46.6 | 21 | 11.9 | 20.5 | 25.6 | 21 | 21.6 | 31.8 |
| ● *Clupea pallasii* | 27.8 | 15.9 | 13.1 | 43.2 | 46.6 | 20.5 | 12.5 | 20.5 | 29 | 18.2 | 20.5 | 32.4 |
| ● *Sprattus sprattus* | 27.8 | 15.9 | 13.1 | 43.2 | 47.2 | 21 | 11.9 | 19.9 | 31.8 | 18.8 | 21.6 | 27.8 |
| ● *Sprattus muelleri* | 33 | 11.9 | 12.5 | 42.6 | 46 | 21.6 | 12.5 | 19.9 | 26.1 | 21.6 | 21.6 | 30.7 |
| ● *Sprattus antipodum* | 32.4 | 12.5 | 12.5 | 42.6 | 46 | 21.6 | 12.5 | 19.9 | 23.3 | 23.9 | 20.5 | 32.4 |
| ○ *Potamalosa richmondia* | 35.2 | 9.7 | 10.2 | 44.9 | 43.2 | 21.6 | 13.6 | 21.6 | 40.9 | 6.8 | 20.5 | 31.8 |
| ● *Hyperlophus vittatus* | 29 | 15.3 | 10.8 | 44.9 | 43.8 | 21.6 | 13.6 | 21 | 27.8 | 17 | 21.6 | 33.5 |
| ● *Ethmidium maculatum* | 33.5 | 10.8 | 17.6 | 38.1 | 44.3 | 22.2 | 12.5 | 21 | 38.1 | 6.8 | 15.9 | 39.2 |
| ● *Jenkinsia lamprotaenia* | 30.1 | 15.3 | 11.9 | 42.6 | 42.6 | 25 | 10.8 | 21.6 | 23.9 | 16.5 | 13.6 | 46 |
| ● *Spratelloides delicatulus* | 26.7 | 16.5 | 14.8 | 42 | 42 | 22.7 | 11.9 | 23.3 | 22.7 | 19.9 | 15.9 | 41.5 |
| ● *Spratelloides gracilis* | 29 | 17 | 13.6 | 40.3 | 41.5 | 22.7 | 13.1 | 22.7 | 27.8 | 26.1 | 20.5 | 25.6 |
| ● *Etrumeus micropus* | 31.8 | 9.1 | 14.2 | 44.9 | 44.9 | 22.2 | 12.5 | 20.5 | 42 | 9.1 | 24.4 | 24.4 |
| ○ *Ilisha africana* | 33.9 | 9.2 | 12.1 | 44.8 | 45.4 | 21.8 | 13.8 | 19 | 49.4 | 3.4 | 19 | 28.2 |
| ○ *Pellona flavipinnis* | 33.9 | 8 | 10.3 | 47.7 | 48.9 | 17.8 | 12.6 | 20.7 | 54.6 | 1.7 | 8.6 | 35.1 |
| ○ *Ilisha elongata* | 32.2 | 9.8 | 10.3 | 47.7 | 47.7 | 19 | 12.6 | 20.7 | 51.1 | 3.4 | 10.3 | 35.1 |
| □ *Pellona ditchela* | 33.3 | 8.6 | 13.2 | 44.8 | 48.9 | 18.4 | 12.1 | 20.7 | 46.6 | 5.2 | 10.3 | 37.9 |
| □ *Anchoviella sp. LBP 2297* | 35.2 | 8 | 10.8 | 46 | 44.3 | 20.5 | 13.6 | 21.6 | 43.2 | 3.4 | 23.3 | 30.1 |
| □ *Lycengraulis grossidens* | 35.8 | 8 | 11.9 | 44.3 | 43.8 | 21.6 | 13.1 | 21.6 | 36.9 | 8.5 | 28.4 | 26.1 |
| ○ *Amazonsprattus scintilla* | 31.3 | 13.1 | 10.2 | 45.5 | 43.2 | 22.2 | 13.6 | 21 | 33.5 | 13.6 | 26.1 | 26.7 |
| ● *Engraulis encrasicolus* | 31.8 | 11.9 | 11.4 | 44.9 | 42.6 | 23.3 | 14.2 | 19.9 | 26.7 | 18.2 | 23.9 | 31.3 |
| ● *Engraulis japonicus* | 31.3 | 12.5 | 11.4 | 44.9 | 43.2 | 21.6 | 14.2 | 21 | 30.7 | 15.3 | 24.4 | 29.5 |
| ● *Stolephorus chinensis* | 34.1 | 10.8 | 11.4 | 43.8 | 40.9 | 23.9 | 15.9 | 19.3 | 38.6 | 9.7 | 23.3 | 28.4 |
| ● *Stolephorus waitei* | 33 | 11.9 | 10.8 | 44.3 | 40.3 | 24.4 | 15.9 | 19.3 | 38.1 | 8.5 | 25 | 28.4 |
| ○ *Lycothrissa crocodilus* | 35.2 | 6.8 | 11.9 | 46 | 46 | 17.6 | 14.2 | 22.2 | 43.8 | 1.7 | 14.8 | 39.8 |
| □ *Setipinna melanochir* | 35.8 | 8.5 | 8 | 47.7 | 43.8 | 20.5 | 13.1 | 22.7 | 40.9 | 6.8 | 14.2 | 38.1 |
| ● *Coilia reynaldi* | 33 | 11.9 | 7.4 | 47.7 | 43.8 | 19.9 | 12.5 | 23.9 | 44.3 | 4 | 21 | 30.7 |
| ● *Thryssa baelama* | 33.5 | 10.8 | 10.2 | 45.5 | 45.5 | 19.9 | 13.6 | 21 | 38.1 | 11.4 | 19.9 | 30.7 |
| ○ *Coilia lindmani* | 34.7 | 9.7 | 9.7 | 46 | 46 | 19.3 | 14.2 | 20.5 | 43.8 | 5.7 | 22.2 | 28.4 |
| □ *Coilia ectenes* | 34.1 | 9.7 | 10.2 | 46 | 46 | 19.3 | 13.6 | 21 | 46.6 | 3.4 | 19.9 | 30.1 |
| ● *Coilia nasus* | 35.8 | 8.5 | 9.7 | 46 | 46.6 | 18.8 | 13.6 | 21 | 44.9 | 5.1 | 21.6 | 28.4 |
| ● *Denticeps clupeoides* | 33.1 | 7.4 | 17.1 | 42.3 | 44 | 21.7 | 11.4 | 22.9 | 42.3 | 2.9 | 26.9 | 28 |

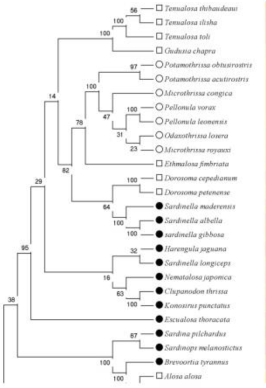
ND6

| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 6.91 | 8.98 | 8.45 | 8.99 | 13.48 | 9.24 | 5.87 | 4.75 | 7.2 | 11.13 | 9.88 | 5.12 |

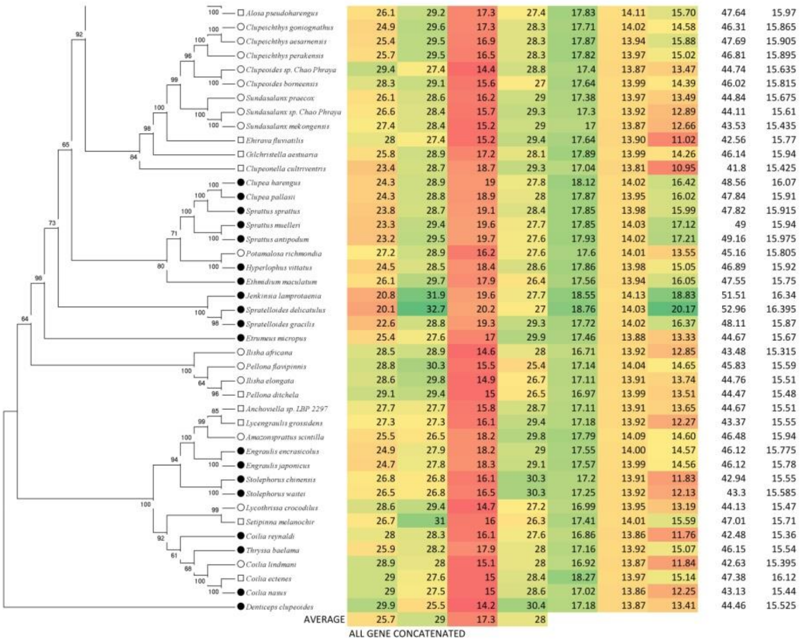


ALL GENE CONCATENATED

| | T-1 | C-1 | A-1 | G-1 | T-2 | C-2 | A-2 | G-2 | T-3 | C-3 | A-3 | G-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa ilisha* | 6.91 | 9.05 | 8.41 | 8.97 | 13.42 | 9.26 | 5.88 | 4.78 | 7.21 | 11.41 | 9.69 | 5.03 |
| □ *Tenualosa toli* | 6.87 | 9.08 | 8.42 | 8.96 | 13.35 | 9.32 | 5.92 | 4.74 | 6.85 | 11.57 | 10.05 | 4.86 |
| □ *Gudusia chapra* | 7.2 | 8.73 | 9 | 8.41 | 13.39 | 9.35 | 5.95 | 4.64 | 7.18 | 11.23 | 12.89 | 2.04 |
| ○ *Potamothrissa obtusirostris* | 6.85 | 8.8 | 8.68 | 9 | 13.33 | 9.45 | 5.97 | 4.59 | 7.48 | 10.79 | 12.44 | 2.61 |
| ○ *Potamothrissa acutirostris* | 7.11 | 8.55 | 8.7 | 8.98 | 13.35 | 9.4 | 5.99 | 4.6 | 7.75 | 10.69 | 12.21 | 2.69 |
| ○ *Microthrissa congica* | 6.77 | 8.92 | 8.64 | 9 | 13.34 | 9.41 | 5.96 | 4.62 | 6.85 | 11.34 | 12.32 | 2.83 |
| ○ *Pellonula vorax* | 6.84 | 8.81 | 8.62 | 9.06 | 13.34 | 9.41 | 5.98 | 4.62 | 7.7 | 10.68 | 12.33 | 2.63 |
| ○ *Pellonula leonensis* | 6.88 | 8.78 | 8.6 | 9.07 | 13.3 | 9.42 | 5.96 | 4.65 | 7.36 | 10.98 | 12.56 | 2.44 |
| ○ *Odaxothrissa losera* | 6.91 | 8.73 | 8.78 | 8.91 | 13.34 | 9.4 | 6 | 4.6 | 7.18 | 11.09 | 12.71 | 2.35 |
| ○ *Microthrissa royauxi* | 6.81 | 8.83 | 8.73 | 8.97 | 13.33 | 9.42 | 5.97 | 4.62 | 6.95 | 11.45 | 12.21 | 2.72 |
| □ *Ethmalosa fimbriata* | 6.91 | 8.82 | 8.56 | 9.05 | 13.4 | 9.39 | 5.95 | 4.6 | 7.28 | 11.02 | 11.97 | 3.06 |
| □ *Dorosoma cepedianum* | 6.84 | 8.85 | 8.57 | 9.06 | 13.38 | 9.41 | 5.95 | 4.59 | 7.09 | 11.69 | 10.69 | 3.87 |
| □ *Dorosoma petenense* | 7.15 | 8.57 | 8.7 | 8.91 | 13.43 | 9.38 | 5.91 | 4.62 | 7.58 | 10.94 | 11.71 | 3.1 |
| ● *Sardinella maderensis* | **6.77** | 8.92 | 8.51 | 9.13 | 13.34 | 9.42 | 5.91 | 4.66 | 6.77 | 11.65 | 10.25 | 4.66 |
| ● *Sardinella albella* | 6.61 | 9.08 | 8.5 | 9.14 | 13.36 | 9.43 | 5.94 | 4.6 | 6.68 | 12 | 8.74 | 5.91 |
| ● *sardinella gibbosa* | 6.6 | 9.09 | 8.55 | 9.1 | 13.37 | 9.43 | 5.92 | 4.61 | 6.62 | 12.05 | 8.84 | 5.83 |
| ● *Harengula jaguana* | 6.85 | 8.91 | 8.36 | 9.21 | 13.35 | 9.32 | 6 | 4.66 | 7.79 | 11.81 | 7.62 | 6.11 |
| ● *Sardinella longiceps* | 6.94 | 8.77 | 8.43 | 9.19 | 13.41 | 9.38 | 5.9 | 4.65 | 7.29 | 11.56 | 9.72 | 4.76 |
| ● *Nematalosa japonica* | 6.75 | 8.98 | 8.46 | 9.14 | 13.36 | 9.4 | 5.96 | 4.62 | 7.41 | 11.13 | 9.28 | 5.51 |
| ● *Clupanodon thrissa* | 7.02 | 8.71 | 8.45 | 9.15 | 13.31 | 9.42 | 5.93 | 4.67 | 7.07 | 11.1 | 10.11 | 5.05 |
| ● *Konosirus punctatus* | 6.91 | 8.79 | 8.46 | 9.17 | 13.35 | 9.45 | 5.95 | 4.59 | 7.32 | 11.28 | 8.77 | 5.98 |
| ● *Escualosa thoracata* | 6.69 | 9.11 | 8.27 | 9.27 | 13.42 | 9.41 | 5.91 | 4.61 | 7.29 | 12.51 | 7.54 | 5.99 |
| ● *Sardina pilchardus* | 6.87 | 8.71 | 8.23 | 9.52 | 13.47 | 9.31 | 5.86 | 4.69 | 7.66 | 11.3 | 7.34 | 7.03 |
| ● *Sardinops melanostictus* | 6.84 | 8.84 | 8.45 | 9.2 | 13.42 | 9.45 | 5.89 | 4.57 | 7.94 | 11.22 | 7.97 | 6.2 |
| ● *Brevoortia tyrannus* | 7.04 | 8.45 | 8.95 | 8.9 | 13.36 | 9.41 | 6.07 | 4.48 | 8.33 | 9.83 | 12.72 | 2.46 |
| □ *Alosa alosa* | 7.67 | 8.1 | 9.42 | 8.14 | 13.55 | 9.18 | 6.12 | 4.49 | 9.17 | 8.23 | 14.39 | 1.54 |
| □ *Alosa pseudoharengus* | 7 | 8.72 | 8.5 | 9.11 | 13.33 | 9.45 | 5.9 | 4.66 | 6.58 | 11.58 | 11.06 | 4.12 |
| ○ *Clupeichthys goniognathus* | 7.02 | 8.84 | 8.6 | 8.87 | 13.35 | 9.41 | 5.97 | 4.61 | 7.92 | 11.2 | 10.83 | 3.38 |
| ○ *Clupeichthys aesarnensis* | 6.87 | 8.79 | 8.59 | 9.08 | 13.43 | 9.36 | 5.96 | 4.58 | 7.72 | 10.6 | 9.73 | 5.28 |
| ○ *Clupeichthys perakensis* | 7.03 | 8.86 | 8.48 | 8.96 | 13.36 | 9.41 | 6 | 4.56 | 7.87 | 11.28 | 10.44 | 3.74 |
| ○ *Clupeoides sp. Chao Phraya* | 6.92 | 8.74 | 9.01 | 8.66 | 13.42 | 9.33 | 6.05 | 4.54 | 6.62 | 11.07 | 13.24 | 2.4 |
| ○ *Clupeoides borneensis* | 7.09 | 8.79 | 8.6 | 8.85 | 13.37 | 9.41 | 5.97 | 4.58 | 7.84 | 11.32 | 11.11 | 3.07 |
| ○ *Sundasalanx praecox* | 7.39 | 8.48 | 8.57 | 8.9 | 13.39 | 9.46 | 5.98 | 4.51 | 8.25 | 10.65 | 11.59 | 2.84 |
| ○ *Sundasalanx sp. Chao Phraya* | 7.31 | 8.53 | 8.72 | 8.77 | 13.42 | 9.44 | 5.99 | 4.48 | 8.53 | 10.44 | 11.92 | 2.45 |
| ○ *Sundasalanx mekongensis* | 7.32 | 8.49 | 9.01 | 8.51 | 13.45 | 9.4 | 6.02 | 4.47 | 8.27 | 10.48 | 12.41 | 2.18 |
| □ *Ehirava fluviatilis* | 6.89 | 8.76 | 8.81 | 8.88 | 13.36 | 9.4 | 6.07 | 4.5 | 9.19 | 9.24 | 13.13 | 1.78 |
| □ *Gilchristella aestuaria* | 6.75 | 8.81 | 8.7 | 9.08 | 13.32 | 9.38 | 6.03 | 4.61 | 8.01 | 10.73 | 11.06 | 3.53 |
| □ *Clupeonella cultriventris* | 7.16 | 8.47 | 9.13 | 8.57 | 13.47 | 9.27 | 6.05 | 4.54 | 8.16 | 9.62 | 14.22 | 1.33 |
| ● *Clupea harengus* | 6.7 | 9.02 | 8.51 | 9.1 | 13.36 | 9.47 | 5.96 | 4.55 | 7.15 | 11.3 | 9.77 | 5.12 |
| ● *Clupea pallasii* | 6.86 | 8.78 | 8.6 | 9.09 | 13.43 | 9.37 | 5.95 | 4.58 | 7.55 | 10.72 | 9.77 | 5.3 |
| ● *Sprattus sprattus* | 6.97 | 8.68 | 8.52 | 9.17 | 13.42 | 9.38 | 5.94 | 4.6 | 8.03 | 10.63 | 9.32 | 5.36 |
| ● *Sprattus muelleri* | 6.93 | 8.75 | 8.56 | 9.1 | 13.39 | 9.38 | 5.91 | 4.65 | 7.37 | 11.28 | 8.85 | 5.84 |
| ● *Sprattus antipodum* | 6.9 | 8.78 | 8.5 | 9.15 | 13.4 | 9.36 | 5.91 | 4.66 | 7.31 | 11.33 | 8.82 | 5.88 |
| ○ *Potamalosa richmondia* | 7.01 | 8.62 | 8.73 | 8.98 | 13.29 | 9.41 | 6.04 | 4.6 | 7.31 | 10.92 | 12.47 | 2.63 |
| ● *Hyperlophus vittatus* | 6.93 | 8.67 | 8.55 | 9.19 | 13.32 | 9.39 | 6.04 | 4.59 | 8.34 | 10.47 | 9.95 | 4.58 |
| ● *Ethmidium maculatum* | 6.93 | 8.68 | 8.84 | 8.88 | 13.38 | 9.33 | 6.02 | 4.61 | 6.08 | 11.67 | 11.2 | 4.38 |
| ● *Jenkinsia lamprotaenia* | 6.8 | 9.13 | 7.98 | 9.42 | 13.44 | 9.46 | 5.76 | 4.67 | 7.44 | 13.28 | 7.06 | 5.55 |
| ● *Spratelloides delicatulus* | 6.66 | 9.31 | 7.91 | 9.45 | 13.42 | 9.35 | 5.88 | 4.68 | 6.91 | 14.07 | 6.26 | 6.1 |
| ● *Spratelloides gracilis* | 7.37 | 8.42 | 8.24 | 9.3 | 13.44 | 9.33 | 5.87 | 4.69 | 8.5 | 11.08 | 8.46 | 5.29 |
| ● *Etrumeus micropus* | 7.33 | 8.43 | 8.55 | 9.03 | 13.51 | 9.22 | 5.94 | 4.66 | 9.1 | 9.99 | 10.91 | 3.34 |
| ○ *Ilisha africana* | 7.25 | 8.53 | 9.38 | 8.18 | 13.29 | 9.42 | 6.12 | 4.5 | 7.45 | 10.92 | 13.04 | 1.93 |
| ○ *Pellona flavipinnis* | 6.84 | 8.74 | 9.35 | 8.4 | 13.18 | 9.53 | 6.12 | 4.51 | 5.4 | 12.02 | 13.28 | 2.63 |
| ○ *Ilisha elongata* | 6.99 | 8.69 | 9.23 | 8.42 | 13.28 | 9.43 | 6.13 | 4.48 | 6.41 | 11.69 | 13.18 | 2.05 |
| □ *Pellona ditchela* | 6.98 | 8.56 | 9.4 | 8.41 | 13.21 | 9.48 | 6.13 | 4.51 | 6.3 | 11.37 | 13.52 | 2.14 |
| □ *Anchoviella sp. LBP 2297* | 7.43 | 8.11 | 8.79 | 9 | 13.37 | 9.39 | 6.05 | 4.52 | 8.98 | 9 | 10.7 | 4.65 |
| □ *Lycengraulis grossidens* | 7.18 | 8.38 | 8.98 | 8.8 | 13.37 | 9.42 | 6.04 | 4.5 | 8.8 | 9.49 | 12.27 | 2.78 |
| ○ *Amazonsprattus scintilla* | 7.01 | 8.72 | 8.53 | 9.07 | 13.3 | 9.43 | 5.94 | 4.66 | 7.13 | 11.05 | 11.6 | 3.55 |
| ● *Engraulis encrasicolus* | 7.01 | 8.5 | 8.77 | 9.05 | 13.35 | 9.47 | 5.99 | 4.53 | 8.59 | 9.97 | 10.18 | 4.6 |
| ● *Engraulis japonicus* | 7.01 | 8.51 | 8.75 | 9.06 | 13.35 | 9.46 | 5.99 | 4.53 | 8.75 | 9.87 | 10.02 | 4.69 |
| ● *Stolephorus chinensis* | 7.31 | 8.26 | 8.83 | 8.94 | 13.35 | 9.4 | 6.08 | 4.51 | 9.63 | 9.14 | 11.87 | 2.69 |
| ● *Stolephorus waitei* | 7.29 | 8.27 | 8.8 | 8.98 | 13.34 | 9.41 | 6.08 | 4.51 | 9.63 | 9.09 | 11.58 | 3.04 |
| ○ *Lycothrissa crocodilus* | 6.95 | 8.65 | 9.4 | 8.34 | 13.32 | 9.41 | 6.07 | 4.54 | 6.97 | 11.37 | 13.18 | 1.82 |
| □ *Setipinna melanochir* | 6.8 | 8.85 | 9.13 | 8.56 | 13.23 | 9.49 | 6.09 | 4.52 | 6.2 | 12.62 | 11.54 | 2.97 |
| ● *Coilia reynaldi* | 7.3 | 8.2 | 9.18 | 8.66 | 13.35 | 9.27 | 6.12 | 4.59 | 7.91 | 10 | 13.67 | 1.76 |
| ● *Thryssa baelama* | 7.33 | 8.27 | 8.84 | 8.89 | 13.38 | 9.38 | 6.04 | 4.54 | 7.23 | 10.57 | 11.03 | 4.5 |
| ○ *Coilia lindmani* | 7.24 | 8.25 | 9.18 | 8.67 | 13.34 | 9.29 | 6.12 | 4.58 | 7.81 | 10.09 | 13.69 | 1.75 |
| □ *Coilia ectenes* | 6.73 | 9.06 | 8.34 | 9.21 | 13.39 | 9.35 | 5.97 | 4.62 | 9.13 | 10.32 | 9.06 | 4.82 |
| ● *Coilia nasus* | 7.2 | 8.32 | 9.12 | 8.7 | 13.35 | 9.31 | 6.13 | 4.55 | 7.41 | 10.36 | 13.67 | 1.89 |
| ● *Denticeps clupeoides* | 7.11 | 8.41 | 9.05 | 8.77 | 13.33 | 9.33 | 6.14 | 4.54 | 7.12 | 10.58 | 12.81 | 2.83 |



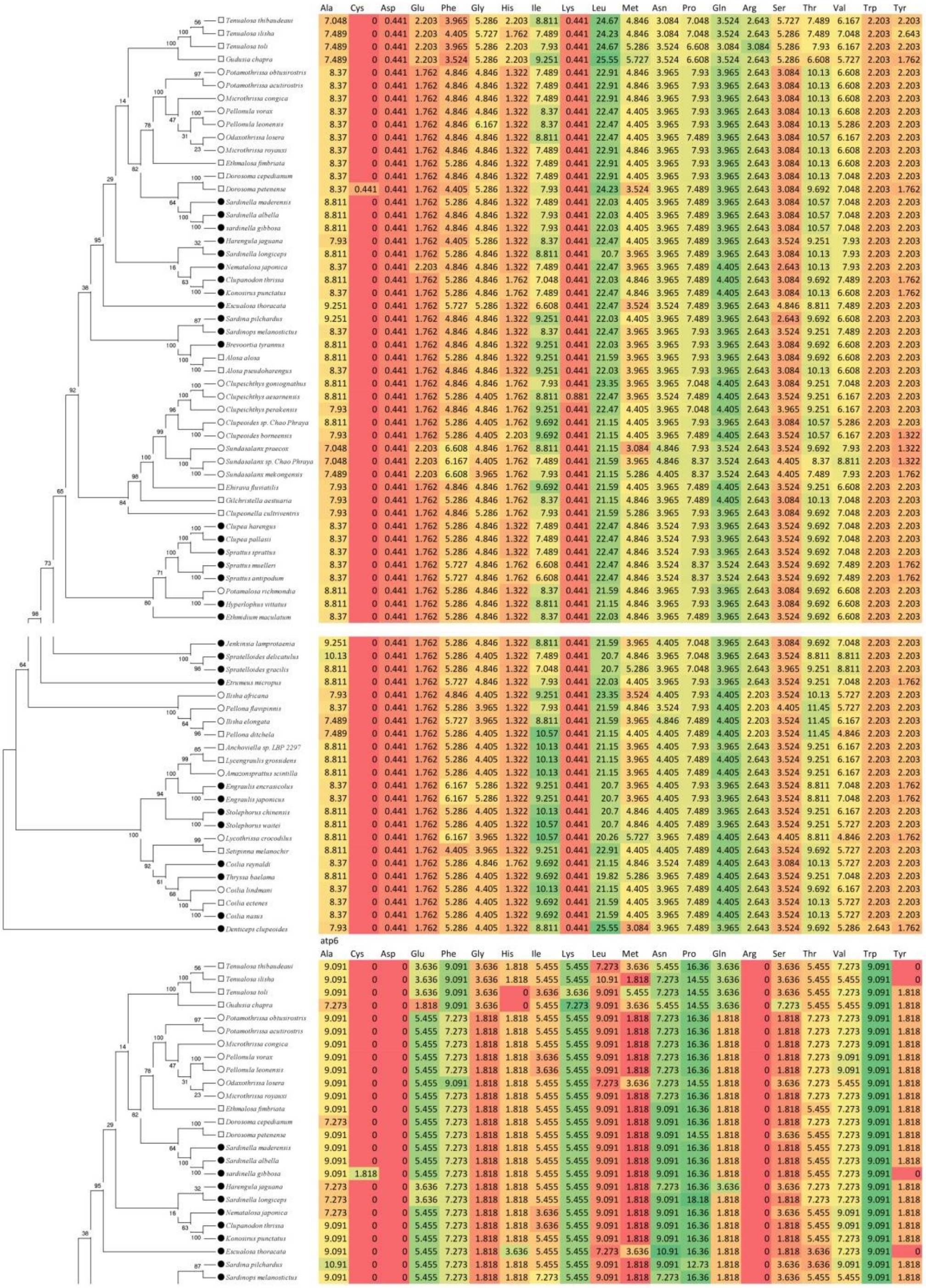
| | A | C | G | T | GC1 | GC2 | GC3 | GC all | GC12 |
|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 24.2 | 29.4 | 18.9 | 27.6 | 17.97 | 13.99 | 16.25 | 48.21 | 15.98 |
| □ *Tenualosa ilisha* | 24 | 29.7 | 18.8 | 27.5 | 18.02 | 14.04 | 16.44 | 48.5 | 16.03 |
| □ *Tenualosa toli* | 24.4 | 30 | 18.6 | 27.1 | 18.04 | 14.06 | 16.43 | 48.53 | 16.05 |
| □ *Gudusia chapra* | 27.9 | 29.3 | 15.1 | 27.8 | 17.14 | 13.99 | 13.27 | 44.4 | 15.565 |
| ○ *Potamothrissa obtusirostris* | 27.1 | 29 | 16.2 | 27.7 | 17.8 | 14.04 | 13.40 | 45.24 | 15.92 |
| ○ *Potamothrissa acutirostris* | 26.9 | 28.6 | 16.3 | 28.2 | 17.53 | 14.00 | 13.38 | 44.91 | 15.765 |
| ○ *Microthrissa congica* | 26.9 | 29.7 | 16.5 | 27 | 17.92 | 14.03 | 14.17 | 46.12 | 15.975 |
| ○ *Pellonula vorax* | 26.9 | 28.9 | 16.3 | 27.9 | 17.87 | 14.03 | 13.31 | 45.21 | 15.95 |
| ○ *Pellonula leonensis* | 27.1 | 29.2 | 16.2 | 27.5 | 17.85 | 14.07 | 13.42 | 45.34 | 15.96 |
| ○ *Odaxothrissa losera* | 27.5 | 29.2 | 15.9 | 27.4 | 17.64 | 14.00 | 13.44 | 45.08 | 15.82 |
| ○ *Microthrissa royauxi* | 26.9 | 29.7 | 16.3 | 27.1 | 17.8 | 14.04 | 14.17 | 46.01 | 15.92 |
| □ *Ethmalosa fimbriata* | 26.5 | 29.2 | 16.7 | 27.6 | 17.87 | 13.99 | 14.08 | 45.94 | 15.93 |
| □ *Dorosoma cepedianum* | 25.2 | 30 | 17.5 | 27.3 | 17.91 | 14.00 | 15.56 | 47.47 | 15.955 |
| □ *Dorosoma petenense* | 26.3 | 28.9 | 16.6 | 28.2 | 17.48 | 14.00 | 14.04 | 45.52 | 15.74 |
| ● *Sardinella maderensis* | 24.7 | 30 | 18.5 | 26.9 | 18.05 | 14.08 | 16.31 | 48.44 | 16.065 |
| ● *Sardinella albella* | 23.2 | 30.5 | 19.7 | 26.6 | 18.22 | 14.03 | 17.91 | 50.16 | 16.125 |
| ● *sardinella gibbosa* | 23.3 | 30.6 | 19.5 | 26.6 | 18.19 | 14.04 | 17.88 | 50.11 | 16.115 |
| ● *Harengula jaguana* | 22 | 30 | 20 | 28 | 18.12 | 13.98 | 17.92 | 50.02 | 16.05 |
| ● *Sardinella longiceps* | 24 | 29.7 | 18.6 | 27.6 | 17.96 | 14.03 | 16.32 | 48.31 | 15.995 |
| ● *Nematalosa japonica* | 23.7 | 29.5 | 19.3 | 27.5 | 18.12 | 14.02 | 16.64 | 48.78 | 16.07 |
| ● *Clupanodon thrissa* | 24.2 | 29.8 | 18.8 | 27.2 | 17.86 | 14.09 | 16.15 | 48.1 | 15.975 |
| ● *Konosirus punctatus* | 23.2 | 29.5 | 19.7 | 27.6 | 17.96 | 14.04 | 17.26 | 49.26 | 16 |
| ● *Escualosa thoracata* | 21.7 | 31 | 19.9 | 27.4 | 18.38 | 14.02 | 18.50 | 50.9 | 16.2 |
| ● *Sardina pilchardus* | 21.4 | 29.3 | 21.2 | 28 | 18.23 | 14.00 | 18.33 | 50.56 | 16.15 |
| ● *Sardinops melanostictus* | 22.3 | 29.5 | 20 | 28.2 | 18.04 | 14.02 | 17.42 | 49.48 | 16.03 |
| ● *Brevoortia tyrannus* | 24.5 | 29.2 | 18.9 | 27.4 | 17.35 | 13.89 | 12.29 | 43.53 | 15.62 |
| □ *Alosa alosa* | 25.5 | 29.8 | 17.9 | 26.9 | 16.24 | 13.67 | 9.77 | 39.68 | 14.955 |



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| □ *Alosa pseudoharengus* | 26.1 | 29.2 | 17.3 | 27.4 | 17.83 | 14.11 | 15.70 | 47.64 | 15.97 |
| ○ *Clupeichthys goniognathus* | 24.9 | 29.6 | 17.3 | 28.3 | 17.71 | 14.02 | 14.58 | 46.31 | 15.865 |
| ○ *Clupeichthys aesarnensis* | 25.4 | 29.5 | 16.9 | 28.3 | 17.87 | 13.94 | 15.88 | 47.69 | 15.905 |
| ○ *Clupeichthys perakensis* | 25.7 | 29.5 | 16.5 | 28.3 | 17.82 | 13.97 | 15.02 | 46.81 | 15.895 |
| ○ *Clupeoides sp. Chao Phraya* | 29.4 | 27.4 | 14.4 | 28.8 | 17.4 | 13.87 | 13.47 | 44.74 | 15.635 |
| ○ *Clupeoides borneensis* | 28.3 | 29.1 | 15.6 | 27 | 17.64 | 13.99 | 14.39 | 46.02 | 15.815 |
| ○ *Sundasalanx praecox* | 26.1 | 28.6 | 16.2 | 29 | 17.38 | 13.97 | 13.49 | 44.84 | 15.675 |
| ○ *Sundasalanx sp. Chao Phraya* | 26.6 | 28.4 | 15.7 | 29.3 | 17.3 | 13.92 | 12.89 | 44.11 | 15.61 |
| ○ *Sundasalanx mekongensis* | 27.4 | 28.4 | 15.2 | 29 | 17 | 13.87 | 12.66 | 43.53 | 15.435 |
| □ *Ehirava fluviatilis* | 28 | 27.4 | 15.2 | 29.4 | 17.64 | 13.90 | 11.02 | 42.56 | 15.77 |
| □ *Gilchristella aestuaria* | 25.8 | 28.9 | 17.2 | 28.1 | 17.89 | 13.99 | 14.26 | 46.14 | 15.94 |
| □ *Clupeonella cultriventris* | 23.4 | 28.7 | 18.7 | 29.3 | 17.04 | 13.81 | 10.95 | 41.8 | 15.425 |
| ● *Clupea harengus* | 24.3 | 28.9 | 19 | 27.8 | 18.12 | 14.02 | 16.42 | 48.56 | 16.07 |
| ● *Clupea pallasii* | 24.3 | 28.8 | 18.9 | 28 | 17.87 | 13.95 | 16.02 | 47.84 | 15.91 |
| ● *Sprattus sprattus* | 23.8 | 28.7 | 19.1 | 28.4 | 17.85 | 13.98 | 15.99 | 47.82 | 15.915 |
| ● *Sprattus muelleri* | 23.3 | 29.4 | 19.6 | 27.7 | 17.85 | 14.03 | 17.12 | 49 | 15.94 |
| ● *Sprattus antipodum* | 23.2 | 29.5 | 19.7 | 27.6 | 17.93 | 14.02 | 17.21 | 49.16 | 15.975 |
| ○ *Potamalosa richmondia* | 27.2 | 28.9 | 16.2 | 27.6 | 17.6 | 14.01 | 13.55 | 45.16 | 15.805 |
| ● *Hyperlophus vittatus* | 24.5 | 28.5 | 18.4 | 28.6 | 17.86 | 13.98 | 15.05 | 46.89 | 15.92 |
| ● *Ethmidium maculatum* | 26.1 | 29.7 | 17.9 | 26.4 | 17.56 | 13.94 | 16.05 | 47.55 | 15.75 |
| ● *Jenkinsia lamprotaenia* | 20.8 | 31.9 | 19.6 | 27.7 | 18.55 | 14.13 | 18.83 | 51.51 | 16.34 |
| ● *Spratelloides delicatulus* | 20.1 | 32.7 | 20.2 | 27 | 18.76 | 14.03 | 20.17 | 52.96 | 16.395 |
| ● *Spratelloides gracilis* | 22.6 | 28.8 | 19.3 | 29.3 | 17.72 | 14.02 | 16.37 | 48.11 | 15.87 |
| ● *Etrumeus micropus* | 25.4 | 27.6 | 17 | 29.9 | 17.46 | 13.88 | 13.33 | 44.67 | 15.67 |
| ○ *Ilisha africana* | 28.5 | 28.9 | 14.6 | 28 | 16.71 | 13.92 | 12.85 | 43.48 | 15.315 |
| ○ *Pellona flavipinnis* | 28.8 | 30.3 | 15.5 | 25.4 | 17.14 | 14.04 | 14.65 | 45.83 | 15.59 |
| ○ *Ilisha elongata* | 28.6 | 29.8 | 14.9 | 26.7 | 17.11 | 13.91 | 13.74 | 44.76 | 15.51 |
| □ *Pellona ditchela* | 29.1 | 29.4 | 15 | 26.5 | 16.97 | 13.99 | 13.51 | 44.47 | 15.48 |
| □ *Anchoviella sp. LBP 2297* | 27.7 | 27.7 | 15.8 | 28.7 | 17.11 | 13.91 | 13.65 | 44.67 | 15.51 |
| □ *Lycengraulis grossidens* | 27.3 | 27.3 | 16.1 | 29.4 | 17.18 | 13.92 | 12.27 | 43.37 | 15.55 |
| ○ *Amazonsprattus scintilla* | 25.5 | 26.5 | 18.2 | 29.8 | 17.79 | 14.09 | 14.60 | 46.48 | 15.94 |
| ● *Engraulis encrasicolus* | 24.9 | 27.9 | 18.2 | 29 | 17.55 | 14.00 | 14.57 | 46.12 | 15.775 |
| ● *Engraulis japonicus* | 24.7 | 27.8 | 18.3 | 29.1 | 17.57 | 13.99 | 14.56 | 46.12 | 15.78 |
| ● *Stolephorus chinensis* | 26.8 | 26.8 | 16.1 | 30.3 | 17.2 | 13.91 | 11.83 | 42.94 | 15.55 |
| ● *Stolephorus waitei* | 26.5 | 26.8 | 16.5 | 30.3 | 17.25 | 13.92 | 12.13 | 43.3 | 15.585 |
| ○ *Lycothrissa crocodilus* | 28.6 | 29.4 | 14.7 | 27.2 | 16.99 | 13.95 | 13.19 | 44.13 | 15.47 |
| □ *Setipinna melanochir* | 26.7 | 31 | 16 | 26.3 | 17.41 | 14.01 | 15.59 | 47.01 | 15.71 |
| ● *Coilia reynaldi* | 28 | 28.3 | 16.1 | 27.6 | 16.86 | 13.86 | 11.76 | 42.48 | 15.36 |
| ● *Thryssa baelama* | 25.9 | 28.2 | 17.9 | 28 | 17.16 | 13.92 | 15.07 | 46.15 | 15.54 |
| ○ *Coilia lindmani* | 28.9 | 28 | 15.1 | 28 | 16.92 | 13.87 | 11.84 | 42.63 | 15.395 |
| □ *Coilia ectenes* | 29 | 27.6 | 15 | 28.4 | 18.27 | 13.97 | 15.14 | 47.38 | 16.12 |
| ● *Coilia nasus* | 29 | 27.5 | 15 | 28.6 | 17.02 | 13.86 | 12.25 | 43.13 | 15.44 |
| ● *Denticeps clupeoides* | 29.9 | 25.5 | 14.2 | 30.4 | 17.18 | 13.87 | 13.41 | 44.46 | 15.525 |
| AVERAGE | 25.7 | 29 | 17.3 | 28 | | | | | |

ALL GENE CONCATENATED

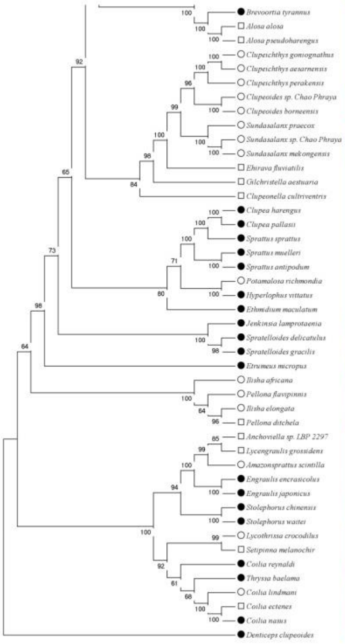
**Fig. A2 Amino acid contents varying across the clupeoid mitogenomic phylogenetic tree**. Amino acid contents of protein coding genes of Clupeoid fishes of the present study.



| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ Tenualosa thibaudeaui | 7.048 | 0 | 0.441 | 2.203 | 3.965 | 5.286 | 2.203 | 8.811 | 0.441 | 24.67 | 4.846 | 3.084 | 7.048 | 3.524 | 2.643 | 5.727 | 7.489 | 6.167 | 2.203 | 2.203 |
| □ Tenualosa ilisha | 7.489 | 0 | 0.441 | 2.203 | 4.405 | 5.727 | 1.762 | 7.489 | 0.441 | 24.23 | 4.846 | 3.084 | 7.048 | 3.524 | 2.643 | 5.286 | 7.489 | 7.048 | 2.203 | 2.643 |
| □ Tenualosa toli | 7.489 | 0 | 0.441 | 2.203 | 3.965 | 5.286 | 2.203 | 7.93 | 0.441 | 24.67 | 5.286 | 3.524 | 6.608 | 3.084 | 3.084 | 5.286 | 7.93 | 6.167 | 2.203 | 2.203 |
| □ Gudusia chapra | 7.489 | 0 | 0.441 | 2.203 | 3.524 | 5.286 | 2.203 | 9.251 | 0.441 | 25.55 | 5.727 | 3.524 | 6.608 | 3.524 | 2.643 | 5.286 | 6.608 | 5.727 | 2.203 | 1.762 |
| ○ Potamothrissa obtusirostris | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 7.489 | 0.441 | 22.91 | 4.846 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.13 | 6.608 | 2.203 | 2.203 |
| ○ Potamothrissa acutirostris | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 7.489 | 0.441 | 22.91 | 4.846 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.13 | 6.608 | 2.203 | 2.203 |
| ○ Microthrissa congica | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 7.489 | 0.441 | 22.91 | 4.846 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.13 | 6.608 | 2.203 | 2.203 |
| ○ Pellonula vorax | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 8.37 | 0.441 | 22.47 | 4.405 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.13 | 6.608 | 2.203 | 2.203 |
| ○ Pellonula leonensis | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 6.167 | 1.322 | 8.37 | 0.441 | 22.47 | 4.405 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.13 | 5.286 | 2.203 | 2.203 |
| ○ Odaxothrissa losera | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 8.811 | 0.441 | 22.47 | 4.405 | 3.965 | 7.489 | 3.965 | 2.643 | 3.084 | 10.57 | 6.167 | 2.203 | 2.203 |
| ○ Microthrissa royauxi | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 7.489 | 0.441 | 22.91 | 4.846 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.13 | 6.608 | 2.203 | 2.203 |
| □ Ethmalosa fimbriata | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 7.489 | 0.441 | 22.91 | 4.405 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.13 | 6.608 | 2.203 | 2.203 |
| □ Dorosoma cepedianum | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 7.489 | 0.441 | 22.91 | 4.405 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.13 | 7.048 | 2.203 | 2.203 |
| □ Dorosoma petenense | 8.37 | 0.441 | 0.441 | 1.762 | 4.405 | 5.286 | 1.322 | 7.93 | 0.441 | 24.23 | 3.524 | 3.965 | 7.489 | 3.965 | 2.643 | 3.084 | 9.692 | 7.048 | 2.203 | 1.762 |
| ● Sardinella maderensis | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 7.489 | 0.441 | 22.03 | 4.405 | 3.965 | 7.489 | 3.965 | 2.643 | 3.084 | 10.57 | 7.048 | 2.203 | 2.203 |
| ● Sardinella albella | 8.811 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 7.93 | 0.441 | 22.03 | 4.405 | 3.965 | 7.489 | 3.965 | 2.643 | 3.084 | 10.57 | 7.048 | 2.203 | 2.203 |
| ● sardinella gibbosa | 8.811 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 7.93 | 0.441 | 22.03 | 4.405 | 3.965 | 7.489 | 3.965 | 2.643 | 3.084 | 10.57 | 7.048 | 2.203 | 2.203 |
| ● Harengula jaguana | 7.93 | 0 | 0.441 | 1.762 | 4.405 | 5.286 | 1.322 | 8.37 | 0.441 | 22.47 | 4.405 | 3.965 | 7.489 | 3.965 | 2.643 | 3.524 | 9.251 | 7.93 | 2.203 | 2.203 |
| ● Sardinella longiceps | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 8.811 | 0.441 | 20.7 | 3.965 | 3.965 | 7.489 | 3.965 | 2.643 | 3.084 | 10.13 | 7.93 | 2.203 | 2.203 |
| ● Nematalosa japonica | 8.37 | 0 | 0.441 | 2.203 | 4.846 | 4.846 | 1.322 | 7.489 | 0.441 | 22.47 | 3.965 | 3.965 | 7.489 | 4.405 | 2.643 | 2.643 | 10.13 | 7.93 | 2.203 | 2.203 |
| ● Clupanodon thrissa | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.762 | 7.048 | 0.441 | 22.03 | 4.846 | 3.965 | 7.489 | 4.405 | 2.643 | 3.084 | 9.692 | 7.489 | 2.203 | 1.762 |
| ● Konosirus punctatus | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.762 | 7.489 | 0.441 | 22.47 | 4.846 | 3.965 | 7.489 | 4.405 | 2.643 | 3.084 | 10.13 | 6.608 | 2.203 | 1.762 |
| ● Escualosa thoracata | 9.251 | 0 | 0.441 | 1.762 | 5.727 | 5.286 | 1.322 | 6.608 | 0.441 | 22.47 | 3.524 | 3.524 | 7.489 | 3.965 | 2.643 | 4.846 | 8.811 | 7.489 | 2.203 | 2.203 |
| ● Sardina pilchardus | 9.251 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 9.251 | 0.441 | 22.03 | 4.405 | 3.965 | 7.489 | 3.965 | 2.643 | 2.643 | 9.692 | 6.608 | 2.203 | 2.203 |
| ● Sardinops melanostictus | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 8.37 | 0.441 | 22.47 | 3.965 | 3.965 | 7.93 | 3.965 | 2.643 | 3.524 | 9.251 | 7.489 | 2.203 | 2.203 |
| ● Brevoortia tyrannus | 8.811 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 9.251 | 0.441 | 22.03 | 3.965 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 9.692 | 6.608 | 2.203 | 2.203 |
| □ Alosa alosa | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 9.251 | 0.441 | 21.59 | 3.965 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 9.692 | 6.608 | 2.203 | 2.203 |
| □ Alosa pseudoharengus | 8.37 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.322 | 9.251 | 0.441 | 22.03 | 3.965 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.13 | 6.608 | 2.203 | 2.203 |
| ○ Clupeichthys goniognathus | 8.811 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.762 | 7.93 | 0.441 | 23.35 | 3.965 | 3.965 | 7.048 | 4.405 | 2.643 | 3.084 | 9.251 | 7.048 | 2.203 | 2.203 |
| ○ Clupeichthys aesarnensis | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.762 | 8.811 | 0.881 | 22.47 | 3.965 | 3.524 | 7.489 | 4.405 | 2.643 | 3.524 | 9.251 | 6.167 | 2.203 | 2.203 |
| ○ Clupeichthys perakensis | 7.93 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.762 | 9.251 | 0.441 | 22.47 | 4.405 | 3.965 | 7.048 | 4.405 | 2.643 | 3.965 | 9.251 | 6.167 | 2.203 | 2.203 |
| ○ Clupeoides sp. Chao Phraya | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.762 | 9.692 | 0.441 | 21.15 | 4.405 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 10.57 | 5.286 | 2.203 | 2.203 |
| ○ Clupeoides borneensis | 7.93 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 2.203 | 9.692 | 0.441 | 21.15 | 4.405 | 3.965 | 7.489 | 4.405 | 2.643 | 3.524 | 10.57 | 6.167 | 2.203 | 1.322 |
| ○ Sundasalanx praecox | 7.048 | 0 | 0.441 | 2.203 | 6.608 | 4.846 | 1.762 | 8.811 | 0.441 | 21.15 | 3.084 | 4.846 | 7.93 | 3.524 | 2.643 | 3.524 | 9.692 | 7.93 | 2.203 | 1.322 |
| ○ Sundasalanx sp. Chao Phraya | 7.048 | 0 | 0.441 | 2.203 | 6.167 | 4.405 | 1.762 | 7.489 | 0.441 | 21.59 | 3.965 | 4.846 | 8.37 | 3.524 | 2.643 | 4.405 | 8.37 | 8.811 | 2.203 | 1.322 |
| ○ Sundasalanx mekongensis | 7.489 | 0 | 0.441 | 2.203 | 6.608 | 3.965 | 1.762 | 7.93 | 0.441 | 21.15 | 5.286 | 4.405 | 8.37 | 3.524 | 2.643 | 4.405 | 7.489 | 7.93 | 2.203 | 1.762 |
| □ Ehirava fluviatilis | 7.93 | 0 | 0.441 | 1.762 | 4.846 | 4.846 | 1.762 | 9.692 | 0.441 | 21.59 | 4.405 | 3.965 | 7.489 | 4.405 | 2.643 | 3.524 | 9.251 | 6.608 | 2.203 | 2.203 |
| □ Gilchristella aestuaria | 7.93 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.762 | 8.37 | 0.441 | 21.15 | 4.846 | 3.965 | 7.489 | 4.405 | 2.643 | 3.084 | 10.13 | 7.048 | 2.203 | 2.203 |
| □ Clupeonella cultriventris | 7.93 | 0 | 0.441 | 1.762 | 4.846 | 5.286 | 1.762 | 7.93 | 0.441 | 21.59 | 5.286 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 9.692 | 7.048 | 2.203 | 2.203 |
| ● Clupea harengus | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 7.489 | 0.441 | 22.47 | 4.846 | 3.524 | 7.93 | 3.965 | 2.643 | 3.524 | 9.692 | 7.048 | 2.203 | 2.203 |
| ● Clupea pallasii | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 7.489 | 0.441 | 22.47 | 4.846 | 3.524 | 7.93 | 3.965 | 2.643 | 3.524 | 9.692 | 7.048 | 2.203 | 2.203 |
| ● Sprattus sprattus | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 7.489 | 0.441 | 22.47 | 4.846 | 3.524 | 7.93 | 3.965 | 2.643 | 3.524 | 9.692 | 7.048 | 2.203 | 2.203 |
| ● Sprattus muelleri | 8.37 | 0 | 0.441 | 1.762 | 5.727 | 4.846 | 1.762 | 6.608 | 0.441 | 22.47 | 4.846 | 3.524 | 8.37 | 3.524 | 2.643 | 3.524 | 9.692 | 7.489 | 2.203 | 1.762 |
| ● Sprattus antipodum | 8.37 | 0 | 0.441 | 1.762 | 5.727 | 4.846 | 1.762 | 6.608 | 0.441 | 22.47 | 4.846 | 3.524 | 8.37 | 3.524 | 2.643 | 3.524 | 9.692 | 7.489 | 2.203 | 1.762 |
| ○ Potamalosa richmondia | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 8.37 | 0.441 | 21.59 | 4.846 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 9.692 | 6.608 | 2.203 | 2.203 |
| ● Hyperlophus vittatus | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 8.811 | 0.441 | 21.15 | 4.846 | 3.965 | 7.93 | 3.965 | 2.643 | 3.084 | 9.692 | 6.608 | 2.203 | 2.203 |
| ● Ethmidium maculatum | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 8.37 | 0.441 | 22.03 | 4.846 | 3.965 | 7.489 | 3.965 | 2.643 | 3.524 | 9.692 | 6.608 | 2.203 | 2.203 |
| ● Jenkinsia lamprotaenia | 9.251 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 8.811 | 0.441 | 21.59 | 3.965 | 4.405 | 7.048 | 3.965 | 2.643 | 3.084 | 9.692 | 7.048 | 2.203 | 2.203 |
| ● Spratelloides delicatulus | 10.13 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 7.489 | 0.441 | 20.7 | 4.846 | 3.965 | 7.048 | 3.965 | 2.643 | 3.524 | 8.811 | 8.811 | 2.203 | 2.203 |
| ● Spratelloides gracilis | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.322 | 7.048 | 0.441 | 20.7 | 5.286 | 3.965 | 7.048 | 3.965 | 2.643 | 3.965 | 9.251 | 8.811 | 2.203 | 2.203 |
| ● Etrumeus micropus | 8.811 | 0 | 0.441 | 1.762 | 5.727 | 4.846 | 1.322 | 7.93 | 0.441 | 22.03 | 4.405 | 3.965 | 7.93 | 3.965 | 2.643 | 3.524 | 9.251 | 7.048 | 2.203 | 1.762 |
| ○ Ilisha africana | 7.93 | 0 | 0.441 | 1.762 | 4.846 | 4.405 | 1.322 | 9.251 | 0.441 | 23.35 | 3.524 | 4.405 | 7.93 | 4.405 | 2.203 | 3.524 | 10.13 | 5.727 | 2.203 | 2.203 |
| ○ Pellona flavipinnis | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 3.965 | 1.322 | 7.93 | 0.441 | 21.59 | 4.846 | 3.524 | 7.93 | 4.405 | 2.203 | 4.405 | 11.45 | 5.727 | 2.203 | 2.203 |
| ○ Ilisha elongata | 7.489 | 0 | 0.441 | 1.762 | 5.727 | 3.965 | 1.322 | 8.811 | 0.441 | 21.59 | 3.965 | 4.846 | 7.489 | 4.405 | 2.203 | 3.524 | 11.45 | 6.167 | 2.203 | 2.203 |
| □ Pellona ditchela | 7.489 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 10.57 | 0.441 | 21.15 | 4.405 | 4.405 | 7.489 | 4.405 | 2.203 | 3.524 | 11.45 | 4.846 | 2.203 | 2.203 |
| □ Anchoviella sp. LBP 2297 | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 10.13 | 0.441 | 21.15 | 3.965 | 4.405 | 7.93 | 3.965 | 2.643 | 3.524 | 9.251 | 6.167 | 2.203 | 2.203 |
| □ Lycengraulis grossidens | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 10.13 | 0.441 | 21.15 | 3.965 | 4.405 | 7.489 | 4.405 | 2.643 | 3.524 | 9.251 | 6.167 | 2.203 | 2.203 |
| ○ Amazonsprattus scintilla | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 10.13 | 0.441 | 21.15 | 3.965 | 4.405 | 7.489 | 4.405 | 2.643 | 3.524 | 9.251 | 6.167 | 2.203 | 2.203 |
| ● Engraulis encrasicolus | 8.37 | 0 | 0.441 | 1.762 | 6.167 | 5.286 | 1.322 | 9.251 | 0.441 | 20.7 | 3.965 | 4.405 | 7.93 | 3.965 | 2.643 | 3.524 | 8.811 | 7.048 | 2.203 | 1.762 |
| ● Engraulis japonicus | 8.37 | 0 | 0.441 | 1.762 | 6.167 | 5.286 | 1.322 | 9.251 | 0.441 | 20.7 | 3.965 | 4.405 | 7.93 | 3.965 | 2.643 | 3.524 | 8.811 | 7.048 | 2.203 | 1.762 |
| ● Stolephorus chinensis | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 10.13 | 0.441 | 20.7 | 4.846 | 4.405 | 7.489 | 3.965 | 2.643 | 3.524 | 9.251 | 6.167 | 2.203 | 2.203 |
| ● Stolephorus waitei | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 10.57 | 0.441 | 20.7 | 4.846 | 4.405 | 7.489 | 3.965 | 2.643 | 3.524 | 9.251 | 5.727 | 2.203 | 2.203 |
| ○ Lycothrissa crocodilus | 8.811 | 0 | 0.441 | 1.762 | 6.167 | 3.965 | 1.322 | 10.57 | 0.441 | 20.26 | 5.727 | 3.965 | 7.489 | 4.405 | 2.643 | 4.405 | 8.811 | 4.846 | 2.203 | 1.762 |
| □ Setipinna melanochir | 8.811 | 0 | 0.441 | 1.762 | 4.405 | 3.965 | 1.322 | 9.251 | 0.441 | 22.91 | 4.405 | 4.405 | 7.489 | 4.405 | 2.643 | 3.524 | 9.692 | 5.727 | 2.203 | 2.203 |
| ● Coilia reynaldi | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.846 | 1.762 | 9.692 | 0.441 | 21.15 | 4.846 | 3.524 | 7.489 | 4.405 | 2.643 | 3.084 | 10.13 | 5.727 | 2.203 | 2.203 |
| ● Thryssa baelama | 8.811 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 9.692 | 0.441 | 19.82 | 5.286 | 3.965 | 7.489 | 4.405 | 2.643 | 3.524 | 9.251 | 7.048 | 2.203 | 2.203 |
| ○ Coilia lindmani | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 10.13 | 0.441 | 21.15 | 4.405 | 3.965 | 7.489 | 4.405 | 2.643 | 3.524 | 9.692 | 6.167 | 2.203 | 2.203 |
| □ Coilia ectenes | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 9.692 | 0.441 | 21.59 | 4.405 | 3.965 | 7.489 | 4.405 | 2.643 | 3.524 | 10.13 | 5.727 | 2.203 | 2.203 |
| ● Coilia nasus | 8.37 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 9.692 | 0.441 | 21.59 | 4.405 | 3.965 | 7.489 | 4.405 | 2.643 | 3.524 | 10.13 | 5.727 | 2.203 | 2.203 |
| ● Denticeps clupeoides | 7.93 | 0 | 0.441 | 1.762 | 5.286 | 4.405 | 1.322 | 8.811 | 0.441 | 25.55 | 3.084 | 3.965 | 7.489 | 3.965 | 2.643 | 3.524 | 9.692 | 5.286 | 2.643 | 1.762 |

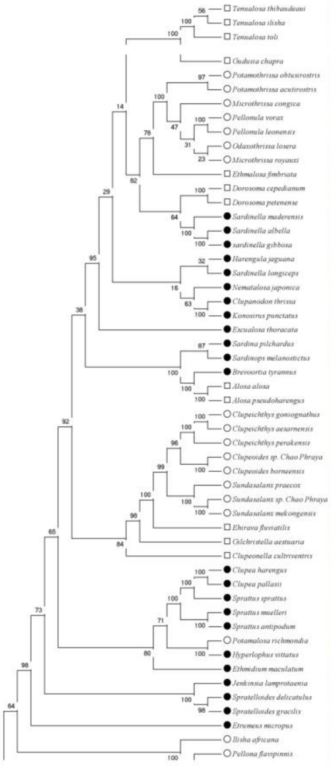
atp6

| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ Tenualosa thibaudeaui | 9.091 | 0 | 0 | 3.636 | 9.091 | 3.636 | 1.818 | 5.455 | 5.455 | 7.273 | 3.636 | 5.455 | 16.36 | 3.636 | 0 | 3.636 | 5.455 | 7.273 | 9.091 | 0 |
| □ Tenualosa ilisha | 9.091 | 0 | 0 | 3.636 | 9.091 | 3.636 | 1.818 | 5.455 | 5.455 | 10.91 | 1.818 | 7.273 | 14.55 | 3.636 | 0 | 3.636 | 5.455 | 5.455 | 9.091 | 0 |
| □ Tenualosa toli | 9.091 | 0 | 0 | 3.636 | 9.091 | 3.636 | 0 | 3.636 | 3.636 | 9.091 | 5.455 | 7.273 | 14.55 | 3.636 | 0 | 3.636 | 5.455 | 7.273 | 9.091 | 1.818 |
| □ Gudusia chapra | 7.273 | 0 | 0 | 1.818 | 9.091 | 3.636 | 0 | 5.455 | 7.273 | 9.091 | 3.636 | 5.455 | 14.55 | 3.636 | 0 | 7.273 | 5.455 | 5.455 | 9.091 | 1.818 |
| ○ Potamothrissa obtusirostris | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 1.818 | 7.273 | 7.273 | 9.091 | 1.818 |
| ○ Potamothrissa acutirostris | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 1.818 | 7.273 | 7.273 | 9.091 | 1.818 |
| ○ Microthrissa congica | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 1.818 | 7.273 | 7.273 | 9.091 | 1.818 |
| ○ Pellonula vorax | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 3.636 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 1.818 | 7.273 | 9.091 | 9.091 | 1.818 |
| ○ Pellonula leonensis | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 3.636 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 1.818 | 7.273 | 9.091 | 9.091 | 1.818 |
| ○ Odaxothrissa losera | 9.091 | 0 | 0 | 5.455 | 9.091 | 1.818 | 1.818 | 5.455 | 5.455 | 7.273 | 3.636 | 7.273 | 14.55 | 1.818 | 0 | 3.636 | 7.273 | 5.455 | 9.091 | 1.818 |
| ○ Microthrissa royauxi | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 1.818 | 7.273 | 7.273 | 9.091 | 1.818 |
| □ Ethmalosa fimbriata | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| □ Dorosoma cepedianum | 7.273 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 7.273 | 7.273 | 9.091 | 1.818 |
| □ Dorosoma petenense | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 14.55 | 1.818 | 0 | 3.636 | 5.455 | 7.273 | 9.091 | 1.818 |
| ● Sardinella maderensis | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| ● Sardinella albella | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| ● sardinella gibbosa | 9.091 | 1.818 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 0 |
| ● Harengula jaguana | 7.273 | 0 | 0 | 3.636 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 3.636 | 0 | 3.636 | 7.273 | 7.273 | 9.091 | 1.818 |
| ● Sardinella longiceps | 7.273 | 0 | 0 | 3.636 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 18.18 | 1.818 | 0 | 1.818 | 7.273 | 7.273 | 9.091 | 1.818 |
| ● Nematalosa japonica | 7.273 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 3.636 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 3.636 | 5.455 | 9.091 | 9.091 | 1.818 |
| ● Clupanodon thrissa | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 3.636 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 9.091 | 9.091 | 1.818 |
| ● Konosirus punctatus | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| ● Escualosa thoracata | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 3.636 | 5.455 | 5.455 | 7.273 | 3.636 | 10.91 | 16.36 | 1.818 | 0 | 1.818 | 3.636 | 7.273 | 9.091 | 0 |
| ● Sardina pilchardus | 10.91 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 12.73 | 1.818 | 0 | 3.636 | 3.636 | 9.091 | 9.091 | 1.818 |
| ● Sardinops melanostictus | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 7.273 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 3.636 | 5.455 | 5.455 | 9.091 | 1.818 |

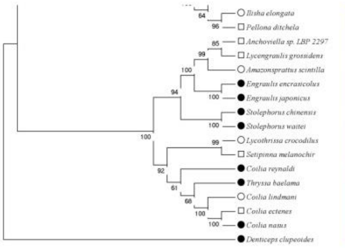


| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ● Brevoortia tyrannus | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| □ Alosa alosa | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| □ Alosa pseudoharengus | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| ○ Clupeichthys goniognathus | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 7.273 | 3.636 | 9.091 | 16.36 | 1.818 | 0 | 0 | 7.273 | 7.273 | 9.091 | 1.818 |
| ○ Clupeichthys aesarnensis | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 7.273 | 3.636 | 9.091 | 16.36 | 1.818 | 0 | 0 | 7.273 | 7.273 | 9.091 | 1.818 |
| ○ Clupeichthys perakensis | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 7.273 | 3.636 | 9.091 | 16.36 | 1.818 | 0 | 0 | 7.273 | 7.273 | 9.091 | 1.818 |
| ○ Clupeoides sp. Chao Phraya | 5.455 | 0 | 0 | 5.455 | 9.091 | 1.818 | 1.818 | 7.273 | 5.455 | 7.273 | 3.636 | 9.091 | 16.36 | 1.818 | 0 | 5.455 | 5.455 | 3.636 | 9.091 | 1.818 |
| ○ Clupeoides borneensis | 5.455 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 3.636 | 7.273 | 18.18 | 1.818 | 0 | 3.636 | 7.273 | 5.455 | 9.091 | 1.818 |
| ○ Sundasalanx praecox | 10.91 | 0 | 0 | 5.455 | 9.091 | 1.818 | 1.818 | 3.636 | 3.636 | 5.455 | 5.455 | 7.273 | 16.36 | 1.818 | 0 | 0 | 9.091 | 7.273 | 9.091 | 1.818 |
| ○ Sundasalanx sp. Chao Phraya | 9.091 | 0 | 0 | 5.455 | 9.091 | 1.818 | 1.818 | 5.455 | 3.636 | 5.455 | 5.455 | 9.091 | 16.36 | 1.818 | 0 | 0 | 9.091 | 5.455 | 9.091 | 1.818 |
| ○ Sundasalanx mekongensis | 7.273 | 0 | 0 | 5.455 | 9.091 | 1.818 | 1.818 | 7.273 | 3.636 | 5.455 | 5.455 | 7.273 | 16.36 | 1.818 | 0 | 0 | 12.73 | 3.636 | 9.091 | 1.818 |
| □ Ehirava fluviatilis | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| □ Gilchristella aestuaria | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 7.273 | 5.455 | 7.273 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| □ Clupeonella cultriventris | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| ● Clupea harengus | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 7.273 | 5.455 | 7.273 | 3.636 | 9.091 | 14.55 | 3.636 | 0 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 |
| ● Clupea pallasii | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 7.273 | 5.455 | 7.273 | 3.636 | 9.091 | 14.55 | 3.636 | 0 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 |
| ● Sprattus sprattus | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 7.273 | 5.455 | 7.273 | 3.636 | 9.091 | 14.55 | 3.636 | 0 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 |
| ● Sprattus muelleri | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 3.636 | 9.091 | 14.55 | 3.636 | 0 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 |
| ● Sprattus antipodum | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 3.636 | 9.091 | 14.55 | 3.636 | 0 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 |
| ○ Potamalosa richmondia | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 7.273 | 10.91 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 |
| ● Hyperlophus vittatus | 9.091 | 0 | 0 | 7.273 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| ● Ethmidium maculatum | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 1.818 | 5.455 | 7.273 | 9.091 | 1.818 |
| ● Jenkinsia lamprotaenia | 7.273 | 0 | 1.818 | 1.818 | 7.273 | 1.818 | 1.818 | 3.636 | 5.455 | 9.091 | 3.636 | 5.455 | 16.36 | 3.636 | 0 | 3.636 | 7.273 | 9.091 | 9.091 | 1.818 |
| ● Spratelloides delicatulus | 9.091 | 0 | 1.818 | 1.818 | 7.273 | 0 | 1.818 | 7.273 | 5.455 | 10.91 | 1.818 | 5.455 | 14.55 | 3.636 | 0 | 3.636 | 10.91 | 3.636 | 9.091 | 1.818 |
| ● Spratelloides gracilis | 10.91 | 0 | 1.818 | 1.818 | 7.273 | 1.818 | 3.636 | 5.455 | 5.455 | 10.91 | 1.818 | 7.273 | 16.36 | 3.636 | 0 | 1.818 | 3.636 | 5.455 | 9.091 | 1.818 |
| ● Etrumeus micropus | 9.091 | 0 | 0 | 5.455 | 7.273 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 1.818 | 7.273 | 16.36 | 1.818 | 0 | 1.818 | 7.273 | 7.273 | 9.091 | 1.818 |
| ○ Ilisha africana | 1.818 | 0 | 0 | 5.455 | 7.273 | 0 | 3.636 | 10.91 | 7.273 | 12.73 | 3.636 | 9.091 | 16.36 | 1.818 | 0 | 3.636 | 7.273 | 0 | 9.091 | 0 |
| ○ Pellona flavipinnis | 3.774 | 0 | 0 | 5.66 | 9.434 | 0 | 1.887 | 7.547 | 7.547 | 11.32 | 1.887 | 5.66 | 16.98 | 1.887 | 0 | 7.547 | 5.66 | 1.887 | 9.434 | 1.887 |
| ○ Ilisha elongata | 3.636 | 0 | 0 | 5.455 | 7.273 | 0 | 3.636 | 5.455 | 7.273 | 10.91 | 1.818 | 9.091 | 16.36 | 1.818 | 0 | 3.636 | 9.091 | 5.455 | 9.091 | 0 |
| □ Pellona ditchela | 3.636 | 0 | 0 | 5.455 | 7.273 | 0 | 1.818 | 7.273 | 7.273 | 10.91 | 1.818 | 9.091 | 18.18 | 1.818 | 0 | 1.818 | 9.091 | 3.636 | 9.091 | 1.818 |
| □ Anchoviella sp. LBP 2297 | 5.455 | 0 | 0 | 7.273 | 7.273 | 1.818 | 1.818 | 7.273 | 5.455 | 12.73 | 3.636 | 5.455 | 16.36 | 1.818 | 0 | 3.636 | 5.455 | 3.636 | 9.091 | 1.818 |
| □ Lycengraulis grossidens | 5.455 | 0 | 0 | 7.273 | 7.273 | 1.818 | 1.818 | 9.091 | 5.455 | 12.73 | 3.636 | 5.455 | 16.36 | 1.818 | 0 | 3.636 | 5.455 | 1.818 | 9.091 | 1.818 |
| ○ Amazonsprattus scintilla | 5.455 | 0 | 0 | 7.273 | 7.273 | 1.818 | 1.818 | 9.091 | 5.455 | 12.73 | 3.636 | 5.455 | 16.36 | 1.818 | 0 | 3.636 | 5.455 | 1.818 | 9.091 | 1.818 |
| ● Engraulis encrasicolus | 5.455 | 0 | 0 | 7.273 | 7.273 | 1.818 | 1.818 | 7.273 | 5.455 | 12.73 | 3.636 | 5.455 | 16.36 | 1.818 | 0 | 3.636 | 5.455 | 3.636 | 9.091 | 1.818 |
| ● Engraulis japonicus | 5.455 | 0 | 0 | 7.273 | 7.273 | 1.818 | 1.818 | 7.273 | 5.455 | 12.73 | 3.636 | 5.455 | 16.36 | 1.818 | 0 | 3.636 | 5.455 | 3.636 | 9.091 | 1.818 |
| ● Stolephorus chinensis | 7.273 | 0 | 0 | 7.273 | 5.455 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 7.273 | 5.455 | 16.36 | 1.818 | 0 | 3.636 | 3.636 | 7.273 | 9.091 | 1.818 |
| ● Stolephorus waitei | 7.273 | 0 | 0 | 7.273 | 5.455 | 1.818 | 1.818 | 5.455 | 5.455 | 9.091 | 7.273 | 5.455 | 16.36 | 1.818 | 0 | 3.636 | 3.636 | 7.273 | 9.091 | 1.818 |
| ○ Lycothrissa crocodilus | 5.455 | 0 | 0 | 7.273 | 7.273 | 1.818 | 1.818 | 10.91 | 5.455 | 12.73 | 3.636 | 5.455 | 16.36 | 1.818 | 0 | 5.455 | 3.636 | 0 | 9.091 | 1.818 |
| □ Setipinna melanochir | 5.455 | 0 | 0 | 7.273 | 5.455 | 1.818 | 1.818 | 7.273 | 5.455 | 12.73 | 5.455 | 5.455 | 18.18 | 1.818 | 0 | 3.636 | 5.455 | 1.818 | 9.091 | 1.818 |
| ● Coilia reynaldi | 5.455 | 0 | 0 | 7.273 | 5.455 | 1.818 | 1.818 | 9.091 | 5.455 | 10.91 | 5.455 | 5.455 | 16.36 | 1.818 | 0 | 5.455 | 3.636 | 3.636 | 9.091 | 1.818 |
| ● Thryssa baelama | 1.818 | 0 | 0 | 7.273 | 5.455 | 1.818 | 1.818 | 7.273 | 5.455 | 12.73 | 5.455 | 5.455 | 16.36 | 1.818 | 0 | 7.273 | 3.636 | 5.455 | 9.091 | 1.818 |
| ○ Coilia lindmani | 5.455 | 0 | 0 | 7.273 | 3.636 | 1.818 | 1.818 | 7.273 | 5.455 | 14.55 | 5.455 | 7.273 | 14.55 | 1.818 | 0 | 5.455 | 3.636 | 3.636 | 9.091 | 1.818 |
| □ Coilia ectenes | 5.455 | 0 | 0 | 7.273 | 5.455 | 1.818 | 1.818 | 5.455 | 5.455 | 12.73 | 5.455 | 7.273 | 14.55 | 1.818 | 0 | 7.273 | 1.818 | 5.455 | 9.091 | 1.818 |
| ● Coilia nasus | 3.636 | 0 | 0 | 7.273 | 5.455 | 1.818 | 1.818 | 5.455 | 5.455 | 12.73 | 5.455 | 7.273 | 14.55 | 1.818 | 0 | 7.273 | 1.818 | 7.273 | 9.091 | 1.818 |
| ● Denticeps clupeoides | 5.455 | 0 | 0 | 3.636 | 7.273 | 0 | 1.818 | 7.273 | 5.455 | 9.091 | 3.636 | 9.091 | 14.55 | 3.636 | 0 | 9.091 | 7.273 | 3.636 | 9.091 | 0 |

atp8

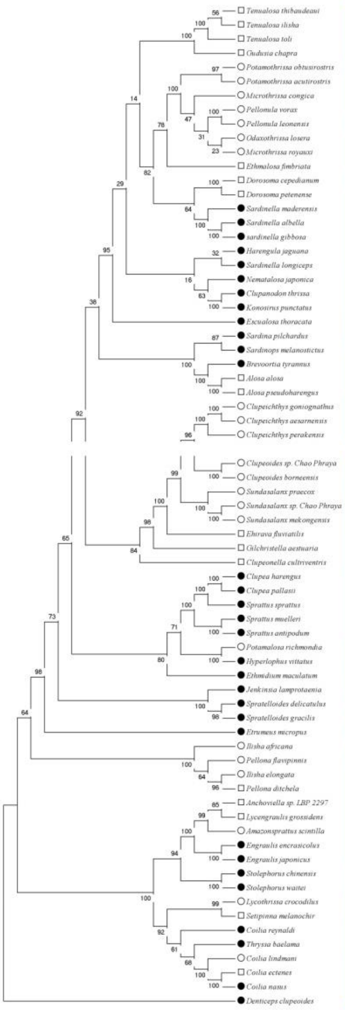


| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ Tenualosa thibaudeaui | 8.35 | 0.194 | 2.718 | 2.136 | 8.35 | 9.126 | 3.689 | 7.184 | 1.553 | 12.04 | 4.466 | 2.718 | 5.437 | 1.748 | 1.553 | 6.602 | 6.99 | 8.544 | 3.301 | 3.301 |
| □ Tenualosa ilisha | 8.544 | 0.194 | 2.718 | 2.136 | 8.35 | 9.126 | 3.689 | 7.184 | 1.553 | 12.04 | 4.466 | 2.718 | 5.437 | 1.748 | 1.553 | 6.602 | 6.99 | 8.35 | 3.301 | 3.301 |
| □ Tenualosa toli | 8.35 | 0.194 | 2.718 | 2.136 | 8.35 | 9.126 | 3.689 | 7.184 | 1.553 | 12.04 | 4.466 | 2.718 | 5.437 | 1.748 | 1.553 | 6.602 | 7.184 | 8.35 | 3.301 | 3.301 |
| □ Gudusia chapra | 8.544 | 0.194 | 2.718 | 2.136 | 8.35 | 9.126 | 3.689 | 7.184 | 1.553 | 12.04 | 4.466 | 2.718 | 5.437 | 1.748 | 1.553 | 6.602 | 6.99 | 8.35 | 3.301 | 3.301 |
| ○ Potamothrissa obtusirostris | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.35 | 3.301 | 3.301 |
| ○ Potamothrissa acutirostris | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.35 | 3.301 | 3.301 |
| ○ Microthrissa congica | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.35 | 3.301 | 3.301 |
| ○ Pellonula vorax | 8.932 | 0.194 | 2.718 | 2.136 | 8.155 | 9.32 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.631 | 1.553 | 1.553 | 5.825 | 6.796 | 8.155 | 3.301 | 3.301 |
| ○ Pellonula leonensis | 9.126 | 0.194 | 2.718 | 2.136 | 8.155 | 9.126 | 3.883 | 7.379 | 1.748 | 12.23 | 4.66 | 2.524 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.35 | 3.301 | 3.301 |
| ○ Odaxothrissa losera | 9.126 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.35 | 3.301 | 3.301 |
| ○ Microthrissa royauxi | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.35 | 3.301 | 3.301 |
| □ Ethmalosa fimbriata | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.35 | 3.301 | 3.301 |
| □ Dorosoma cepedianum | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 6.019 | 6.602 | 8.35 | 3.301 | 3.301 |
| □ Dorosoma petenense | 9.126 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 6.019 | 6.796 | 8.35 | 3.301 | 3.301 |
| ● Sardinella maderensis | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.573 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.155 | 3.301 | 3.301 |
| ● Sardinella albella | 9.515 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.573 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.631 | 6.796 | 8.155 | 3.301 | 3.301 |
| ● sardinella gibbosa | 9.515 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.767 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.631 | 6.796 | 7.961 | 3.301 | 3.301 |
| ● Harengula jaguana | 9.515 | 0.194 | 2.718 | 2.136 | 7.961 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.602 | 8.35 | 3.301 | 3.495 |
| ● Sardinella longiceps | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.35 | 3.301 | 3.301 |
| ● Nematalosa japonica | 8.932 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 6.214 | 6.796 | 8.35 | 3.301 | 3.301 |
| ● Clupanodon thrissa | 8.932 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 6.214 | 6.796 | 8.35 | 3.301 | 3.301 |
| ● Konosirus punctatus | 8.932 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 6.214 | 6.796 | 8.35 | 3.301 | 3.301 |
| ● Escualosa thoracata | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.796 | 8.35 | 3.301 | 3.301 |
| ● Sardina pilchardus | 8.932 | 0.194 | 2.913 | 2.136 | 8.155 | 9.126 | 3.883 | 7.379 | 1.748 | 12.23 | 4.66 | 2.913 | 5.437 | 1.553 | 1.553 | 5.631 | 6.99 | 8.155 | 3.301 | 3.107 |
| ● Sardinops melanostictus | 9.32 | 0.194 | 2.913 | 1.942 | 8.155 | 8.932 | 3.883 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.631 | 6.99 | 8.35 | 3.301 | 3.107 |
| ● Brevoortia tyrannus | 9.091 | 0.193 | 2.901 | 1.934 | 8.317 | 9.091 | 3.482 | 7.35 | 1.741 | 12.57 | 4.836 | 2.708 | 5.416 | 1.547 | 1.547 | 5.996 | 6.77 | 7.737 | 3.288 | 3.482 |
| □ Alosa alosa | 9.126 | 0.194 | 2.913 | 1.942 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 6.019 | 6.796 | 8.155 | 3.301 | 3.301 |
| □ Alosa pseudoharengus | 9.126 | 0.194 | 2.913 | 1.942 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 6.019 | 6.796 | 8.155 | 3.301 | 3.301 |
| ○ Clupeichthys goniognathus | 9.126 | 0.194 | 2.718 | 2.136 | 7.961 | 8.932 | 3.689 | 7.573 | 1.748 | 11.65 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.99 | 8.544 | 3.301 | 3.495 |
| ○ Clupeichthys aesarnensis | 9.126 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.573 | 1.748 | 11.46 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.99 | 8.544 | 3.301 | 3.495 |
| ○ Clupeichthys perakensis | 9.126 | 0.194 | 2.718 | 2.136 | 7.961 | 8.932 | 3.689 | 7.573 | 1.748 | 11.65 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.99 | 8.544 | 3.301 | 3.495 |
| ○ Clupeoides sp. Chao Phraya | 9.126 | 0.194 | 2.913 | 1.942 | 8.155 | 9.126 | 3.689 | 7.573 | 1.748 | 11.84 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.631 | 6.99 | 8.35 | 3.301 | 3.301 |
| ○ Clupeoides borneensis | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 11.65 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.631 | 6.99 | 8.738 | 3.301 | 3.301 |
| ○ Sundasalanx praecox | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.544 | 3.689 | 7.767 | 1.748 | 11.65 | 5.049 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 7.184 | 8.155 | 3.301 | 3.301 |
| ○ Sundasalanx sp. Chao Phraya | 9.32 | 0.194 | 2.718 | 2.136 | 8.155 | 8.544 | 3.689 | 7.573 | 1.748 | 11.84 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 6.019 | 6.99 | 8.35 | 3.301 | 3.301 |
| ○ Sundasalanx mekongensis | 9.32 | 0.194 | 2.718 | 2.136 | 8.35 | 8.544 | 3.689 | 7.573 | 1.748 | 11.65 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 6.019 | 6.99 | 8.35 | 3.301 | 3.301 |
| □ Ehirava fluviatilis | 9.515 | 0.194 | 2.913 | 1.942 | 8.155 | 8.738 | 3.689 | 7.573 | 1.748 | 11.65 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.437 | 7.184 | 8.544 | 3.301 | 3.301 |
| □ Gilchristella aestuaria | 9.709 | 0.194 | 2.718 | 2.136 | 8.155 | 8.738 | 3.689 | 7.573 | 1.748 | 11.65 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.437 | 6.99 | 8.544 | 3.301 | 3.301 |
| □ Clupeonella cultriventris | 9.126 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.573 | 1.748 | 12.04 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 6.019 | 6.796 | 8.155 | 3.301 | 3.301 |
| ● Clupea harengus | 9.126 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.767 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.99 | 7.961 | 3.301 | 3.301 |
| ● Clupea pallasii | 9.126 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.767 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.99 | 7.961 | 3.301 | 3.301 |
| ● Sprattus sprattus | 9.126 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.99 | 8.35 | 3.301 | 3.301 |
| ● Sprattus muelleri | 8.932 | 0.194 | 2.718 | 2.136 | 8.155 | 9.126 | 3.689 | 7.961 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.99 | 7.767 | 3.301 | 3.301 |
| ● Sprattus antipodum | 8.932 | 0.194 | 2.718 | 2.136 | 8.155 | 9.126 | 3.689 | 7.961 | 1.748 | 12.23 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.99 | 7.767 | 3.301 | 3.301 |
| ○ Potamalosa richmondia | 9.709 | 0.194 | 2.718 | 2.136 | 8.35 | 8.932 | 3.689 | 7.379 | 1.748 | 12.04 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.243 | 6.99 | 8.155 | 3.301 | 3.301 |
| ● Hyperlophus vittatus | 9.709 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.573 | 1.748 | 12.04 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.437 | 6.796 | 8.155 | 3.301 | 3.301 |
| ● Ethmidium maculatum | 9.515 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.573 | 1.748 | 12.43 | 4.854 | 2.718 | 5.437 | 1.553 | 1.553 | 5.631 | 6.796 | 7.767 | 3.301 | 3.301 |
| ● Jenkinsia lamprotaenia | 9.515 | 0.194 | 2.718 | 2.136 | 8.155 | 9.126 | 3.689 | 7.184 | 1.748 | 12.23 | 4.466 | 2.718 | 5.437 | 1.553 | 1.553 | 5.243 | 6.99 | 8.738 | 3.301 | 3.301 |
| ● Spratelloides delicatulus | 9.709 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.379 | 1.748 | 12.23 | 4.466 | 2.718 | 5.437 | 1.553 | 1.553 | 5.243 | 6.99 | 8.544 | 3.301 | 3.301 |
| ● Spratelloides gracilis | 9.709 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 7.184 | 1.748 | 12.23 | 4.466 | 2.718 | 5.437 | 1.553 | 1.553 | 5.243 | 6.99 | 8.738 | 3.301 | 3.301 |
| ● Etrumeus micropus | 9.515 | 0.194 | 2.718 | 2.136 | 8.155 | 8.932 | 3.689 | 6.796 | 1.748 | 12.23 | 5.243 | 2.718 | 5.437 | 1.553 | 1.553 | 5.825 | 6.602 | 8.35 | 3.301 | 3.301 |
| ○ Ilisha africana | 9.32 | 0.194 | 2.913 | 1.748 | 8.155 | 8.932 | 3.689 | 7.379 | 1.942 | 12.43 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.243 | 7.184 | 8.35 | 3.301 | 3.301 |
| ○ Pellona flavipinnis | 9.515 | 0.194 | 2.913 | 1.942 | 8.155 | 8.932 | 3.689 | 7.184 | 1.748 | 12.43 | 4.66 | 2.718 | 5.437 | 1.553 | 1.553 | 5.243 | 7.184 | 8.35 | 3.301 | 3.301 |



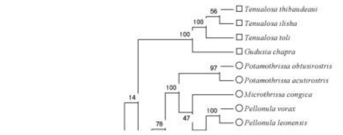
| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ○ *Ilisha elongata* | 9.903 | 0.194 | 2.913 | 1.942 | 8.155 | 8.932 | 3.689 | 7.184 | 1.748 | 12.43 | 4.466 | 2.718 | 5.631 | 1.553 | 1.553 | 5.243 | 6.602 | 8.544 | 3.301 | 3.301 |
| □ *Pellona ditchela* | 9.903 | 0.194 | 2.913 | 1.942 | 8.155 | 8.932 | 3.689 | 7.184 | 1.748 | 12.43 | 4.466 | 2.718 | 5.437 | 1.553 | 1.553 | 5.243 | 6.796 | 8.544 | 3.301 | 3.301 |
| □ *Anchoviella sp. LBP 2297* | 9.747 | 0.195 | 2.729 | 2.144 | 8.187 | 8.967 | 3.704 | 7.992 | 1.754 | 12.09 | 4.288 | 2.729 | 5.458 | 1.365 | 1.559 | 5.068 | 7.018 | 8.382 | 3.314 | 3.314 |
| □ *Lycengraulis grossidens* | 9.747 | 0.195 | 2.729 | 2.144 | 8.187 | 8.967 | 3.704 | 7.992 | 1.754 | 12.48 | 4.288 | 2.729 | 5.458 | 1.365 | 1.559 | 5.068 | 7.018 | 7.992 | 3.314 | 3.314 |
| ○ *Amazonsprattus scintilla* | 9.747 | 0.195 | 2.729 | 2.144 | 8.187 | 8.967 | 3.704 | 7.797 | 1.754 | 12.28 | 4.288 | 2.729 | 5.458 | 1.365 | 1.559 | 5.068 | 7.018 | 8.382 | 3.314 | 3.314 |
| ● *Engraulis encrasicolus* | 9.357 | 0.195 | 2.729 | 2.144 | 8.187 | 9.162 | 3.704 | 7.992 | 1.754 | 12.09 | 4.483 | 2.729 | 5.458 | 1.365 | 1.559 | 5.263 | 7.018 | 8.187 | 3.314 | 3.314 |
| ● *Engraulis japonicus* | 9.552 | 0.195 | 2.729 | 2.144 | 8.187 | 9.162 | 3.704 | 7.992 | 1.754 | 12.09 | 4.483 | 2.729 | 5.458 | 1.365 | 1.559 | 5.068 | 7.018 | 8.187 | 3.314 | 3.314 |
| ● *Stolephorus chinensis* | 9.747 | 0.195 | 2.729 | 2.144 | 7.992 | 8.967 | 3.704 | 7.992 | 1.754 | 12.09 | 4.288 | 2.729 | 5.458 | 1.365 | 1.559 | 5.068 | 7.018 | 8.382 | 3.314 | 3.509 |
| ● *Stolephorus waitei* | 9.747 | 0.195 | 2.729 | 2.144 | 7.992 | 8.967 | 3.704 | 7.992 | 1.754 | 12.09 | 4.288 | 2.729 | 5.458 | 1.365 | 1.559 | 5.068 | 7.018 | 8.382 | 3.314 | 3.509 |
| ○ *Lycothrissa crocodilus* | 9.552 | 0.195 | 2.729 | 2.144 | 8.187 | 9.162 | 3.704 | 7.797 | 1.754 | 12.28 | 4.483 | 2.729 | 5.458 | 1.365 | 1.559 | 5.263 | 7.212 | 7.797 | 3.314 | 3.314 |
| □ *Setipinna melanochir* | 9.357 | 0.195 | 2.729 | 2.144 | 8.187 | 8.967 | 3.704 | 7.212 | 1.754 | 12.09 | 4.288 | 2.729 | 5.458 | 1.365 | 1.559 | 5.458 | 7.018 | 9.162 | 3.314 | 3.314 |
| ● *Coilia reynaldi* | 9.747 | 0.195 | 2.729 | 2.144 | 8.187 | 8.967 | 3.704 | 7.797 | 1.754 | 12.09 | 4.288 | 2.729 | 5.458 | 1.365 | 1.559 | 5.263 | 7.018 | 8.382 | 3.314 | 3.314 |
| ● *Thryssa baelama* | 9.357 | 0.195 | 2.729 | 2.144 | 7.992 | 8.967 | 3.704 | 7.992 | 1.754 | 12.09 | 4.288 | 2.729 | 5.458 | 1.365 | 1.559 | 5.263 | 7.018 | 8.577 | 3.314 | 3.509 |
| ○ *Coilia lindmani* | 9.552 | 0.195 | 2.729 | 2.144 | 7.992 | 8.967 | 3.704 | 7.992 | 1.754 | 11.89 | 4.483 | 2.729 | 5.458 | 1.365 | 1.559 | 5.068 | 7.018 | 8.577 | 3.314 | 3.509 |
| □ *Coilia ectenes* | 9.357 | 0.195 | 2.729 | 2.144 | 7.992 | 9.162 | 3.704 | 7.992 | 1.754 | 11.89 | 4.483 | 2.729 | 5.458 | 1.365 | 1.559 | 5.068 | 7.018 | 8.577 | 3.314 | 3.509 |
| ● *Coilia nasus* | 9.357 | 0.195 | 2.729 | 2.144 | 7.992 | 9.162 | 3.704 | 7.797 | 1.754 | 11.89 | 4.483 | 2.729 | 5.458 | 1.365 | 1.559 | 5.068 | 7.018 | 8.772 | 3.314 | 3.509 |
| ● *Denticeps clupeoides* | 8.932 | 0.194 | 2.913 | 1.942 | 8.155 | 9.126 | 3.689 | 8.155 | 1.553 | 12.23 | 4.66 | 2.913 | 5.631 | 1.359 | 1.748 | 5.825 | 6.796 | 7.573 | 3.301 | 3.301 |

co1

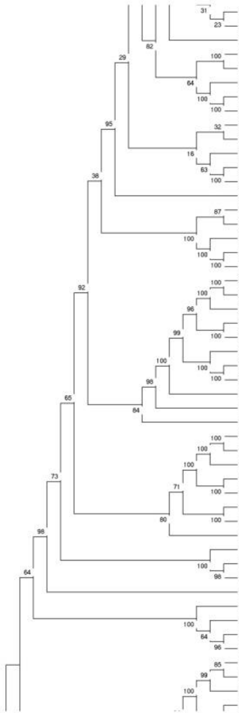


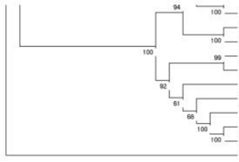
| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 7.391 | 0.87 | 6.087 | 6.522 | 3.478 | 3.913 | 4.348 | 7.391 | 1.739 | 13.04 | 6.087 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.913 | 8.261 | 2.174 | 3.913 |
| □ *Tenualosa ilisha* | 7.391 | 0.87 | 6.087 | 6.522 | 3.478 | 3.913 | 4.348 | 7.391 | 1.739 | 13.04 | 6.087 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.913 | 8.261 | 2.174 | 3.913 |
| □ *Tenualosa toli* | 7.391 | 0.87 | 6.087 | 6.522 | 3.478 | 3.913 | 4.348 | 7.826 | 1.739 | 13.48 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.913 | 7.826 | 2.174 | 3.913 |
| □ *Gudusia chapra* | 7.391 | 0.87 | 6.087 | 6.087 | 3.478 | 3.913 | 4.348 | 7.826 | 2.174 | 13.04 | 6.087 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.913 | 7.826 | 2.174 | 3.913 |
| ○ *Potamothrissa obtusirostris* | 6.522 | 1.304 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 8.261 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.913 | 9.13 | 2.174 | 3.913 |
| ○ *Potamothrissa acutirostris* | 6.957 | 1.304 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.478 | 9.565 | 2.174 | 3.913 |
| ○ *Microthrissa congica* | 6.957 | 1.304 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.478 | 9.565 | 2.174 | 3.913 |
| ○ *Pellonula vorax* | 6.957 | 1.304 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.391 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.478 | 10 | 2.174 | 3.913 |
| ○ *Pellonula leonensis* | 7.391 | 1.304 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.391 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.478 | 9.565 | 2.174 | 3.913 |
| ○ *Odaxothrissa losera* | 6.957 | 1.304 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.391 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.478 | 10 | 2.174 | 3.913 |
| ○ *Microthrissa royauxi* | 6.957 | 1.304 | 5.652 | 6.522 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.652 | 2.609 | 6.087 | 3.478 | 2.609 | 6.957 | 3.478 | 9.13 | 2.174 | 3.913 |
| □ *Ethmalosa fimbriata* | 6.957 | 0.87 | 6.087 | 6.522 | 3.478 | 3.913 | 4.348 | 7.826 | 1.739 | 12.17 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.13 | 2.174 | 3.913 |
| □ *Dorosoma cepedianum* | 7.391 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 8.261 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.478 | 8.696 | 2.174 | 3.913 |
| □ *Dorosoma petenense* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 8.261 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.478 | 9.13 | 2.174 | 3.913 |
| ● *Sardinella maderensis* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.13 | 2.174 | 3.913 |
| ● *Sardinella albella* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.13 | 2.174 | 3.913 |
| ● *sardinella gibbosa* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.13 | 2.174 | 3.913 |
| ● *Harengula jaguana* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 8.261 | 1.739 | 12.61 | 5.652 | 1.304 | 6.087 | 3.478 | 2.609 | 7.826 | 3.913 | 8.696 | 2.174 | 3.913 |
| ● *Sardinella longiceps* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.13 | 2.174 | 3.913 |
| ● *Nematalosa japonica* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 8.261 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 8.696 | 2.174 | 3.913 |
| ● *Clupanodon thrissa* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 8.261 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 8.696 | 2.174 | 3.913 |
| ● *Konosirus punctatus* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 8.261 | 1.739 | 12.61 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 8.696 | 2.174 | 3.913 |
| ● *Escualosa thoracata* | 7.391 | 1.304 | 6.087 | 6.522 | 3.478 | 3.913 | 4.348 | 7.826 | 1.739 | 12.17 | 5.652 | 1.739 | 6.087 | 3.478 | 2.609 | 6.957 | 3.913 | 8.696 | 2.174 | 3.913 |
| ● *Sardina pilchardus* | 7.826 | 0.87 | 6.522 | 6.087 | 3.043 | 3.913 | 3.913 | 9.13 | 1.739 | 13.04 | 4.783 | 1.739 | 6.087 | 3.043 | 2.609 | 7.391 | 4.348 | 7.826 | 2.174 | 3.913 |
| ● *Sardinops melanostictus* | 6.522 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.391 | 1.739 | 13.04 | 5.217 | 1.739 | 6.087 | 3.478 | 2.609 | 7.826 | 3.913 | 9.565 | 2.174 | 3.913 |
| ● *Brevoortia tyrannus* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.391 | 1.739 | 13.04 | 5.217 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.565 | 2.174 | 3.913 |
| □ *Alosa alosa* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.391 | 1.739 | 13.04 | 5.217 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.565 | 2.174 | 3.913 |
| □ *Alosa pseudoharengus* | 6.957 | 0.87 | 6.087 | 6.522 | 3.043 | 3.913 | 4.348 | 7.391 | 1.739 | 13.04 | 5.217 | 1.739 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.565 | 2.174 | 3.913 |
| ○ *Clupeichthys goniognathus* | 7.826 | 1.304 | 5.652 | 6.522 | 3.913 | 3.913 | 4.348 | 7.826 | 1.739 | 11.74 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.522 | 3.913 | 9.13 | 2.174 | 3.913 |
| ○ *Clupeichthys aesarnensis* | 7.826 | 1.304 | 5.652 | 6.522 | 3.913 | 3.913 | 4.348 | 7.826 | 1.739 | 11.74 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.522 | 3.913 | 9.13 | 2.174 | 3.913 |
| ○ *Clupeichthys perakensis* | 7.826 | 1.304 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 7.826 | 1.739 | 12.17 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.522 | 3.913 | 9.13 | 2.174 | 3.913 |
| ○ *Clupeoides sp. Chao Phraya* | 7.826 | 1.304 | 5.652 | 6.522 | 3.913 | 3.913 | 4.348 | 8.696 | 1.739 | 11.74 | 5.652 | 2.174 | 6.087 | 3.478 | 2.609 | 6.087 | 4.348 | 7.826 | 2.174 | 3.913 |
| ○ *Clupeoides borneensis* | 7.826 | 1.304 | 5.652 | 6.522 | 3.913 | 3.913 | 4.348 | 8.261 | 1.739 | 11.74 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.087 | 4.348 | 8.696 | 2.174 | 3.913 |
| ○ *Sundasalanx praecox* | 7.826 | 1.304 | 5.217 | 6.087 | 3.478 | 3.913 | 4.348 | 7.391 | 2.174 | 11.74 | 5.217 | 3.043 | 6.087 | 3.478 | 2.609 | 6.087 | 4.348 | 9.565 | 2.174 | 3.913 |
| ○ *Sundasalanx sp. Chao Phraya* | 7.391 | 1.304 | 5.217 | 6.087 | 3.478 | 3.913 | 4.348 | 7.391 | 2.174 | 11.74 | 5.217 | 3.043 | 6.087 | 3.478 | 2.609 | 6.087 | 4.783 | 9.565 | 2.174 | 3.913 |
| ○ *Sundasalanx mekongensis* | 7.391 | 1.304 | 5.217 | 6.087 | 3.478 | 3.913 | 4.348 | 7.826 | 2.174 | 11.74 | 5.217 | 3.043 | 6.087 | 3.478 | 2.609 | 6.087 | 4.783 | 9.13 | 2.174 | 3.913 |
| □ *Ehirava fluviatilis* | 7.86 | 0.873 | 5.677 | 6.55 | 3.493 | 3.493 | 4.367 | 8.297 | 1.747 | 11.79 | 5.24 | 2.183 | 6.114 | 3.493 | 3.057 | 6.55 | 4.367 | 8.734 | 2.183 | 3.93 |
| □ *Gilchristella aestuaria* | 7.391 | 0.87 | 5.652 | 6.522 | 3.913 | 3.913 | 4.348 | 8.261 | 1.739 | 11.74 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.522 | 4.783 | 8.696 | 2.174 | 3.913 |
| □ *Clupeonella cultriventris* | 7.391 | 0.87 | 5.217 | 6.957 | 3.478 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.522 | 4.348 | 9.13 | 2.174 | 3.913 |
| ● *Clupea harengus* | 6.522 | 0.87 | 5.652 | 6.087 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 13.04 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 10 | 2.174 | 3.913 |
| ● *Clupea pallasii* | 6.522 | 0.87 | 5.652 | 6.087 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 13.04 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 10 | 2.174 | 3.913 |
| ● *Sprattus sprattus* | 6.522 | 0.87 | 5.652 | 6.087 | 3.043 | 3.913 | 4.348 | 7.826 | 1.304 | 13.04 | 5.217 | 2.174 | 6.087 | 3.913 | 2.609 | 7.391 | 3.913 | 10 | 2.174 | 3.913 |
| ● *Sprattus muelleri* | 6.522 | 0.87 | 5.652 | 6.087 | 3.043 | 3.913 | 4.348 | 8.261 | 1.739 | 13.04 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.565 | 2.174 | 3.913 |
| ● *Sprattus antipodum* | 6.522 | 0.87 | 5.652 | 6.087 | 3.043 | 3.913 | 4.348 | 8.261 | 1.739 | 13.04 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 9.565 | 2.174 | 3.913 |
| ○ *Potamalosa richmondia* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.261 | 1.739 | 12.61 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 8.696 | 2.174 | 3.913 |
| ● *Hyperlophus vittatus* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.261 | 1.739 | 12.61 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 3.913 | 8.696 | 2.174 | 3.913 |
| ● *Ethmidium maculatum* | 6.957 | 0.87 | 5.652 | 6.522 | 3.043 | 3.913 | 4.348 | 8.696 | 1.739 | 13.48 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 4.348 | 7.391 | 2.174 | 3.913 |
| ● *Jenkinsia lamprotaenia* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 4.348 | 8.696 | 2.174 | 3.913 |
| ● *Spratelloides delicatulus* | 6.522 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.826 | 4.348 | 8.696 | 2.174 | 3.913 |
| ● *Spratelloides gracilis* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 7.826 | 1.739 | 12.61 | 4.783 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.783 | 9.13 | 2.174 | 3.913 |
| ● *Etrumeus micropus* | 7.391 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 7.826 | 1.739 | 13.04 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.348 | 8.261 | 2.174 | 3.913 |
| ○ *Ilisha africana* | 6.957 | 0.87 | 5.652 | 6.087 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 13.48 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.783 | 7.391 | 2.174 | 3.913 |
| ○ *Pellona flavipinnis* | 7.391 | 0.87 | 5.652 | 6.087 | 3.478 | 3.913 | 4.348 | 9.13 | 1.739 | 12.61 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 4.348 | 7.391 | 2.174 | 3.913 |
| ○ *Ilisha elongata* | 6.957 | 0.87 | 5.652 | 6.087 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 13.04 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 4.783 | 7.391 | 2.174 | 3.913 |
| □ *Pellona ditchela* | 6.957 | 0.87 | 5.652 | 6.087 | 3.478 | 3.913 | 4.348 | 8.261 | 1.739 | 13.04 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 7.391 | 4.783 | 7.826 | 2.174 | 3.913 |
| □ *Anchoviella sp. LBP 2297* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 11.3 | 5.652 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 3.913 | 9.565 | 2.174 | 3.913 |
| □ *Lycengraulis grossidens* | 6.957 | 0.87 | 5.652 | 6.087 | 3.478 | 3.913 | 4.348 | 9.13 | 1.739 | 11.3 | 5.652 | 2.174 | 6.087 | 3.913 | 2.609 | 6.957 | 4.348 | 8.696 | 2.174 | 3.913 |
| ○ *Amazonsprattus scintilla* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.261 | 1.739 | 11.3 | 5.652 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.348 | 9.565 | 2.174 | 3.913 |
| ● *Engraulis encrasicolus* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 11.3 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.522 | 4.348 | 10 | 2.174 | 3.913 |
| ● *Engraulis japonicus* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 11.3 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.522 | 4.348 | 10 | 2.174 | 3.913 |
| ● *Stolephorus chinensis* | 6.522 | 0.87 | 5.652 | 6.522 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 11.74 | 5.652 | 2.609 | 6.087 | 3.478 | 2.609 | 6.957 | 4.348 | 9.565 | 2.174 | 4.348 |
| ● *Stolephorus waitei* | 6.522 | 0.87 | 5.652 | 6.522 | 3.043 | 3.913 | 4.348 | 7.826 | 1.739 | 11.74 | 5.652 | 2.609 | 6.087 | 3.478 | 2.609 | 6.957 | 4.348 | 9.565 | 2.174 | 4.348 |
| ○ *Lycothrissa crocodilus* | 7.391 | 0.87 | 5.652 | 6.522 | 3.478 | 3.478 | 4.348 | 9.565 | 1.739 | 11.74 | 5.652 | 2.174 | 6.087 | 3.043 | 2.609 | 7.391 | 3.913 | 8.261 | 2.174 | 3.913 |
| □ *Setipinna melanochir* | 7.826 | 0.87 | 5.217 | 6.522 | 3.478 | 3.478 | 4.348 | 9.565 | 1.739 | 11.74 | 5.652 | 2.609 | 6.087 | 3.478 | 2.609 | 6.957 | 3.913 | 7.826 | 2.174 | 3.913 |
| ● *Coilia reynaldi* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 11.3 | 5.652 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.783 | 8.696 | 2.174 | 3.913 |
| ● *Thryssa baelama* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 11.3 | 5.652 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.783 | 8.696 | 2.174 | 3.913 |
| ○ *Coilia lindmani* | 7.391 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 11.3 | 5.652 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.348 | 8.696 | 2.174 | 3.913 |
| □ *Coilia ectenes* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 11.3 | 5.652 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.783 | 8.696 | 2.174 | 3.913 |
| ● *Coilia nasus* | 6.957 | 0.87 | 5.652 | 6.522 | 3.478 | 3.913 | 4.348 | 8.696 | 1.739 | 11.3 | 5.652 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.783 | 8.696 | 2.174 | 3.913 |
| ● *Denticeps clupeoides* | 7.391 | 0.87 | 6.087 | 5.652 | 3.913 | 3.913 | 4.348 | 8.696 | 2.174 | 12.17 | 5.217 | 2.174 | 6.087 | 3.478 | 2.609 | 6.957 | 4.348 | 7.826 | 2.174 | 3.913 |

co2



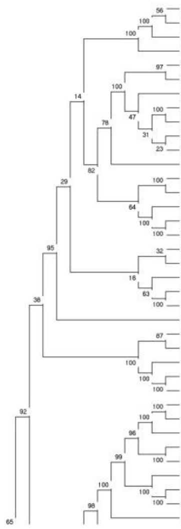
| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 4.981 | 0.766 | 12.64 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 9.962 | 7.28 | 4.598 | 4.215 |
| □ *Tenualosa ilisha* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.981 | 0.766 | 12.26 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 9.579 | 7.28 | 4.598 | 4.215 |
| □ *Tenualosa toli* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 5.364 | 0.766 | 12.64 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 9.962 | 6.897 | 4.598 | 4.215 |
| □ *Gudusia chapra* | 8.429 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 6.13 | 0.766 | 12.64 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 5.747 | 4.598 | 4.215 |
| ○ *Potamothrissa obtusirostris* | 8.046 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 4.981 | 0.766 | 11.88 | 3.065 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 6.897 | 4.598 | 4.598 |
| ○ *Potamothrissa acutirostris* | 8.046 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 4.981 | 0.766 | 11.88 | 3.065 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 6.897 | 4.598 | 4.598 |
| ○ *Microthrissa congica* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.812 | 6.13 | 4.981 | 0.766 | 11.88 | 3.065 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 6.897 | 4.598 | 4.598 |
| ○ *Pellonula vorax* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.812 | 6.13 | 4.981 | 0.766 | 11.88 | 3.065 | 0.766 | 4.598 | 3.065 | 1.916 | 5.747 | 9.195 | 6.897 | 4.598 | 4.598 |
| ○ *Pellonula leonensis* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.812 | 6.13 | 4.981 | 0.766 | 11.88 | 3.065 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 6.897 | 4.598 | 4.598 |



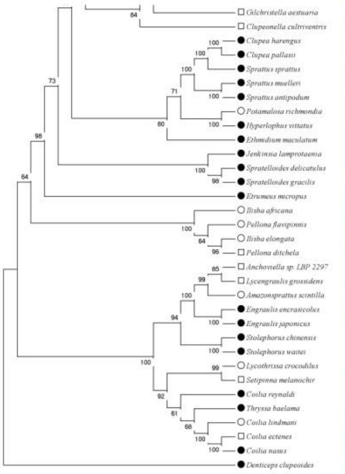


| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ○ *Odaxothrissa losera* | 8.046 | 0.766 | 1.916 | 3.448 | 9.195 | 8.812 | 6.13 | 4.981 | 0.766 | 11.88 | 3.065 | 0.766 | 4.598 | 3.065 | 1.916 | 4.981 | 9.579 | 6.897 | 4.598 | 4.598 |
| ○ *Microthrissa royauxi* | 8.429 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 4.981 | 0.766 | 11.88 | 3.065 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 6.513 | 4.598 | 4.598 |
| □ *Ethmalosa fimbriata* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.598 | 0.766 | 12.26 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.897 | 4.598 | 4.215 |
| □ *Dorosoma cepedianum* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.598 | 0.766 | 12.64 | 2.299 | 0.766 | 4.981 | 3.065 | 1.916 | 5.364 | 9.195 | 7.28 | 4.598 | 4.215 |
| □ *Dorosoma petenense* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 5.747 | 0.766 | 12.64 | 2.299 | 0.766 | 4.981 | 3.065 | 1.916 | 5.364 | 9.195 | 6.13 | 4.598 | 4.215 |
| ● *Sardinella maderensis* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.598 | 0.766 | 12.64 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.897 | 4.598 | 4.215 |
| ● *Sardinella albella* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.215 | 0.766 | 12.64 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 7.28 | 4.598 | 4.215 |
| ● *sardinella gibbosa* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.215 | 0.766 | 12.64 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 7.28 | 4.598 | 4.215 |
| ● *Harengula jaguana* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 4.215 | 0.766 | 12.26 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 7.663 | 4.598 | 4.215 |
| ● *Sardinella longiceps* | 8.046 | 0.766 | 1.916 | 3.448 | 9.962 | 8.429 | 6.13 | 4.981 | 0.766 | 12.64 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.195 | 7.28 | 4.598 | 3.831 |
| ● *Nematalosa japonica* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.513 | 3.831 | 0.766 | 12.26 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 8.046 | 4.598 | 4.215 |
| ● *Clupanodon thrissa* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 5.364 | 0.766 | 12.26 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 6.897 | 4.598 | 4.215 |
| ● *Konosirus punctatus* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.981 | 0.766 | 11.49 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 7.663 | 4.598 | 4.215 |
| ● *Escualosa thoracata* | 7.663 | 0.766 | 1.916 | 3.448 | 9.579 | 9.195 | 6.13 | 4.215 | 0.766 | 12.64 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 7.28 | 4.598 | 4.215 |
| ● *Sardina pilchardus* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 4.215 | 0.766 | 12.64 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 7.663 | 4.598 | 4.215 |
| ● *Sardinops melanostictus* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 5.364 | 0.766 | 12.64 | 2.299 | 0.766 | 4.981 | 3.065 | 1.916 | 5.364 | 8.812 | 6.513 | 4.598 | 4.215 |
| ● *Brevoortia tyrannus* | 8.046 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 4.215 | 0.766 | 13.03 | 2.299 | 0.766 | 4.981 | 3.065 | 1.916 | 5.364 | 9.195 | 7.28 | 4.598 | 4.598 |
| □ *Alosa alosa* | 8.046 | 0.766 | 1.916 | 3.065 | 9.579 | 8.812 | 6.513 | 3.831 | 0.766 | 13.03 | 2.299 | 0.766 | 4.981 | 3.065 | 1.916 | 5.364 | 9.195 | 7.663 | 4.598 | 3.831 |
| □ *Alosa pseudoharengus* | 8.046 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 4.598 | 0.766 | 13.03 | 2.299 | 0.766 | 4.981 | 3.065 | 1.916 | 5.364 | 9.195 | 6.897 | 4.598 | 4.598 |
| ○ *Clupeichthys goniognathus* | 7.663 | 0.766 | 1.916 | 3.448 | 9.962 | 9.195 | 6.13 | 4.981 | 0.766 | 12.64 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.513 | 4.598 | 3.831 |
| ○ *Clupeichthys aesarnensis* | 8.046 | 0.766 | 1.916 | 3.448 | 9.962 | 9.195 | 6.13 | 4.981 | 0.766 | 12.64 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 6.513 | 4.598 | 3.831 |
| ○ *Clupeichthys perakensis* | 7.663 | 0.766 | 1.916 | 3.448 | 9.962 | 9.195 | 6.13 | 5.364 | 0.766 | 12.64 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.13 | 4.598 | 3.831 |
| ○ *Clupeoides sp. Chao Phraya* | 7.663 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 4.598 | 0.766 | 12.26 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.897 | 4.598 | 4.215 |
| ○ *Clupeoides borneensis* | 7.663 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 5.364 | 0.766 | 12.64 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.13 | 4.598 | 4.215 |
| ○ *Sundasalanx praecox* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.812 | 5.747 | 4.215 | 0.766 | 12.26 | 2.682 | 1.149 | 4.598 | 3.065 | 1.916 | 5.364 | 10.34 | 6.897 | 4.598 | 4.598 |
| ○ *Sundasalanx sp. Chao Phraya* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.812 | 6.13 | 4.981 | 0.766 | 11.88 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 10.34 | 6.897 | 4.598 | 4.598 |
| ○ *Sundasalanx mekongensis* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.812 | 6.13 | 4.981 | 0.766 | 12.26 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 10.34 | 6.513 | 4.598 | 4.598 |
| □ *Ehirava fluviatilis* | 7.663 | 0.766 | 1.916 | 3.448 | 9.579 | 9.195 | 6.13 | 4.981 | 0.766 | 12.64 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 9.579 | 6.513 | 4.598 | 4.215 |
| □ *Gilchristella aestuaria* | 7.663 | 0.766 | 1.916 | 3.448 | 9.579 | 9.195 | 6.13 | 4.598 | 0.766 | 12.26 | 2.682 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 7.28 | 4.598 | 4.215 |
| □ *Clupeonella cultriventris* | 7.663 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 4.215 | 0.766 | 13.03 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 6.13 | 9.195 | 6.897 | 4.598 | 4.215 |
| ● *Clupea harengus* | 9.195 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.598 | 0.766 | 11.49 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 8.812 | 7.663 | 4.598 | 4.598 |
| ● *Clupea pallasii* | 8.812 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 4.215 | 0.766 | 11.49 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 8.812 | 8.046 | 4.598 | 4.598 |
| ● *Sprattus sprattus* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.981 | 0.766 | 11.11 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.747 | 9.579 | 7.663 | 4.598 | 4.598 |
| ● *Sprattus muelleri* | 8.046 | 0.766 | 2.299 | 3.065 | 9.962 | 9.195 | 6.13 | 4.981 | 0.766 | 11.11 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 7.28 | 4.598 | 4.215 |
| ● *Sprattus antipodum* | 8.046 | 0.766 | 2.299 | 3.065 | 9.962 | 9.195 | 6.13 | 4.981 | 0.766 | 11.11 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 7.28 | 4.598 | 4.215 |
| ○ *Potamalosa richmondia* | 8.046 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 5.364 | 0.766 | 12.26 | 2.682 | 0.766 | 4.981 | 3.065 | 1.916 | 5.364 | 9.195 | 6.513 | 4.598 | 4.598 |
| ● *Hyperlophus vittatus* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.812 | 6.13 | 4.598 | 0.766 | 12.64 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 6.897 | 4.598 | 4.598 |
| ● *Ethmidium maculatum* | 8.429 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.981 | 0.766 | 11.11 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.747 | 9.195 | 7.28 | 4.598 | 4.598 |
| ● *Jenkinsia lamprotaenia* | 8.429 | 0.766 | 1.916 | 3.448 | 9.195 | 9.195 | 6.13 | 4.215 | 0.766 | 11.49 | 2.682 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 7.663 | 4.598 | 4.598 |
| ● *Spratelloides delicatulus* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.215 | 0.766 | 12.26 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.747 | 9.195 | 7.663 | 4.598 | 4.215 |
| ● *Spratelloides gracilis* | 8.046 | 0.766 | 1.916 | 3.448 | 9.195 | 9.195 | 6.13 | 5.364 | 0.766 | 11.11 | 2.682 | 0.383 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.897 | 4.598 | 4.598 |
| ● *Etrumeus micropus* | 7.663 | 0.766 | 1.916 | 3.448 | 9.195 | 8.812 | 6.13 | 4.981 | 0.766 | 11.49 | 2.682 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 9.962 | 7.28 | 4.598 | 4.598 |
| ○ *Ilisha africana* | 8.046 | 0.766 | 1.916 | 3.448 | 8.812 | 8.429 | 6.13 | 4.598 | 0.766 | 12.64 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.897 | 4.598 | 4.598 |
| ○ *Pellona flavipinnis* | 8.046 | 0.766 | 1.916 | 3.448 | 9.195 | 8.429 | 6.13 | 4.598 | 0.766 | 12.64 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.897 | 4.598 | 4.215 |
| ○ *Ilisha elongata* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.429 | 6.13 | 4.215 | 0.766 | 12.26 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.747 | 9.579 | 7.28 | 4.598 | 4.215 |
| □ *Pellona ditchela* | 8.429 | 0.766 | 1.916 | 3.448 | 8.812 | 8.429 | 6.13 | 4.598 | 0.766 | 12.64 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.579 | 6.897 | 4.598 | 4.598 |
| □ *Anchoviella sp. LBP 2297* | 8.046 | 0.766 | 1.916 | 3.448 | 9.962 | 8.812 | 6.13 | 5.364 | 0.766 | 11.11 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.513 | 4.598 | 4.215 |
| □ *Lycengraulis grossidens* | 8.429 | 0.766 | 1.916 | 3.448 | 10.34 | 8.812 | 6.13 | 4.981 | 0.766 | 11.11 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 4.981 | 9.962 | 6.897 | 4.598 | 3.831 |
| ○ *Amazonsprattus scintilla* | 8.429 | 0.766 | 1.916 | 3.448 | 9.962 | 8.812 | 6.13 | 5.364 | 0.766 | 11.11 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 4.981 | 9.962 | 6.513 | 4.598 | 4.215 |
| ● *Engraulis encrasicolus* | 8.812 | 0.766 | 2.299 | 3.065 | 9.962 | 8.812 | 6.513 | 4.215 | 0.766 | 11.49 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 4.598 | 9.962 | 7.663 | 4.598 | 4.215 |
| ● *Engraulis japonicus* | 8.429 | 0.766 | 1.916 | 3.448 | 9.962 | 8.812 | 6.13 | 4.598 | 0.766 | 11.49 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 4.981 | 9.962 | 7.28 | 4.598 | 4.215 |
| ● *Stolephorus chinensis* | 8.046 | 0.766 | 2.299 | 3.065 | 10.34 | 8.812 | 6.13 | 5.747 | 0.766 | 11.49 | 2.299 | 0.766 | 4.981 | 3.065 | 1.916 | 5.364 | 9.579 | 6.13 | 4.598 | 3.831 |
| ● *Stolephorus waitei* | 8.046 | 0.766 | 2.299 | 3.065 | 10.34 | 8.812 | 6.13 | 5.747 | 0.766 | 11.49 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.13 | 4.598 | 3.831 |
| ○ *Lycothrissa crocodilus* | 8.046 | 0.766 | 1.916 | 3.065 | 9.962 | 8.812 | 6.13 | 6.13 | 0.766 | 11.88 | 2.682 | 0.766 | 4.598 | 3.448 | 1.916 | 5.364 | 9.195 | 6.13 | 4.598 | 3.831 |
| □ *Setipinna melanochir* | 7.663 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 4.981 | 0.766 | 11.88 | 2.682 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.897 | 4.598 | 4.215 |
| ● *Coilia reynaldi* | 8.046 | 0.766 | 1.916 | 3.448 | 9.962 | 8.812 | 6.13 | 5.747 | 0.766 | 11.49 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.13 | 4.598 | 4.215 |
| ● *Thryssa baelama* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 5.747 | 0.766 | 11.49 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.364 | 9.962 | 6.513 | 4.598 | 4.215 |
| ○ *Coilia lindmani* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 5.747 | 0.766 | 11.88 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 9.579 | 6.513 | 4.598 | 4.215 |
| □ *Coilia ectenes* | 8.046 | 0.766 | 1.916 | 3.448 | 9.579 | 8.812 | 6.13 | 5.747 | 0.766 | 11.88 | 2.299 | 0.383 | 4.598 | 3.065 | 1.916 | 5.747 | 9.579 | 6.513 | 4.598 | 4.215 |
| ● *Coilia nasus* | 8.046 | 0.766 | 1.916 | 3.448 | 9.962 | 9.195 | 6.13 | 5.747 | 0.766 | 11.49 | 2.299 | 0.383 | 4.981 | 2.682 | 1.916 | 5.747 | 9.579 | 6.13 | 4.598 | 4.215 |
| ● *Denticeps clupeoides* | 9.195 | 0.766 | 1.916 | 3.448 | 8.429 | 8.429 | 6.13 | 5.364 | 1.149 | 11.88 | 2.299 | 0.766 | 4.598 | 3.065 | 1.916 | 5.747 | 8.812 | 6.897 | 4.598 | 4.598 |

co3

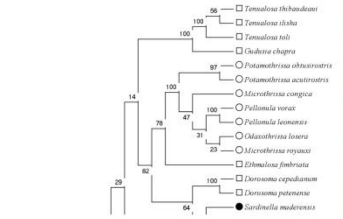


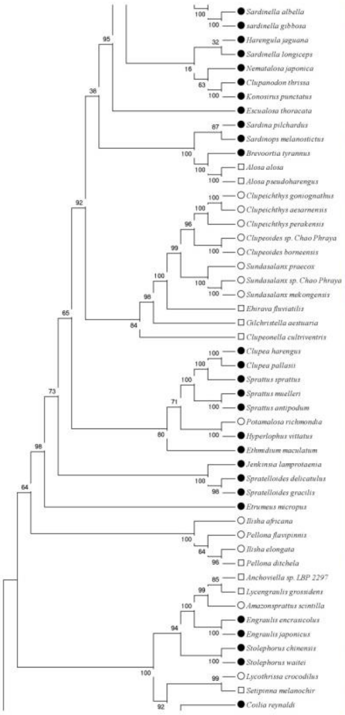
| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 8.421 | 0.789 | 2.632 | 2.105 | 8.158 | 7.105 | 3.158 | 7.105 | 2.368 | 15.26 | 2.368 | 5 | 5.526 | 1.579 | 2.105 | 6.053 | 5.526 | 7.895 | 3.158 | 3.684 |
| □ *Tenualosa ilisha* | 8.684 | 0.789 | 2.632 | 2.105 | 7.895 | 7.105 | 3.158 | 7.368 | 2.368 | 15.26 | 2.368 | 5 | 5.526 | 1.579 | 2.105 | 6.053 | 5.526 | 7.368 | 3.421 | 3.684 |
| □ *Tenualosa toli* | 8.684 | 0.789 | 2.632 | 2.105 | 7.895 | 7.105 | 3.158 | 7.368 | 2.368 | 15.26 | 2.368 | 5 | 5.526 | 1.579 | 2.105 | 6.053 | 5.526 | 7.368 | 3.421 | 3.684 |
| □ *Gudusia chapra* | 8.421 | 0.789 | 2.632 | 2.105 | 7.895 | 7.105 | 3.158 | 7.895 | 2.632 | 15.26 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 5.789 | 6.579 | 3.421 | 3.684 |
| ○ *Potamothrissa obtusirostris* | 8.421 | 0.789 | 2.895 | 1.842 | 7.895 | 7.105 | 3.158 | 7.368 | 2.368 | 15.53 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 6.053 | 6.842 | 3.421 | 3.684 |
| ○ *Potamothrissa acutirostris* | 8.421 | 0.789 | 2.895 | 1.842 | 7.895 | 7.105 | 3.158 | 7.368 | 2.368 | 15.79 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 6.053 | 6.842 | 3.421 | 3.684 |
| ○ *Microthrissa congica* | 8.421 | 0.789 | 2.895 | 1.842 | 8.158 | 7.105 | 3.158 | 7.105 | 2.368 | 15.79 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 6.053 | 6.842 | 3.421 | 3.684 |
| ○ *Pellonula vorax* | 8.684 | 0.789 | 2.632 | 2.105 | 7.895 | 7.105 | 3.158 | 6.842 | 2.368 | 16.32 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 5.789 | 6.053 | 6.842 | 3.421 | 3.684 |
| ○ *Pellonula leonensis* | 8.684 | 0.789 | 2.632 | 2.105 | 8.158 | 7.105 | 3.158 | 6.842 | 2.368 | 16.05 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 5.789 | 6.053 | 6.842 | 3.421 | 3.684 |
| ○ *Odaxothrissa losera* | 8.684 | 0.789 | 2.895 | 1.842 | 7.895 | 7.105 | 3.158 | 7.105 | 2.368 | 15.79 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 6.053 | 6.579 | 3.421 | 3.684 |
| ○ *Microthrissa royauxi* | 8.421 | 0.789 | 2.895 | 1.842 | 7.895 | 7.105 | 3.158 | 7.105 | 2.368 | 16.05 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 6.053 | 6.842 | 3.421 | 3.684 |
| □ *Ethmalosa fimbriata* | 8.684 | 0.789 | 2.895 | 1.842 | 7.895 | 7.105 | 3.158 | 7.105 | 2.368 | 15.53 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 5.526 | 7.632 | 3.421 | 3.684 |
| □ *Dorosoma cepedianum* | 8.421 | 0.789 | 2.632 | 2.105 | 7.895 | 7.368 | 3.158 | 7.368 | 2.368 | 15.79 | 2.368 | 5 | 5.526 | 1.579 | 2.105 | 6.053 | 5.526 | 6.842 | 3.421 | 3.684 |
| □ *Dorosoma petenense* | 8.158 | 0.789 | 2.632 | 2.105 | 7.895 | 7.368 | 3.158 | 7.105 | 2.368 | 15.26 | 2.895 | 5 | 5.526 | 1.579 | 2.105 | 6.053 | 5.789 | 7.368 | 3.158 | 3.684 |
| ● *Sardinella maderensis* | 8.947 | 0.789 | 2.895 | 1.842 | 7.895 | 7.105 | 3.158 | 6.316 | 2.368 | 15.53 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 5.526 | 7.895 | 3.421 | 3.684 |
| ● *Sardinella albella* | 8.947 | 0.789 | 2.895 | 1.842 | 8.158 | 7.105 | 3.158 | 6.579 | 2.368 | 15.79 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.316 | 5.526 | 7.105 | 3.421 | 3.684 |
| ● *sardinella gibbosa* | 8.947 | 0.789 | 2.895 | 1.842 | 8.158 | 7.105 | 3.158 | 6.579 | 2.368 | 15.79 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.316 | 5.526 | 7.105 | 3.421 | 3.684 |
| ● *Harengula jaguana* | 9.474 | 0.789 | 2.895 | 1.579 | 8.158 | 7.105 | 3.158 | 7.105 | 2.368 | 15.53 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.579 | 5 | 6.842 | 3.421 | 3.684 |
| ● *Sardinella longiceps* | 8.947 | 0.789 | 3.158 | 1.579 | 7.632 | 7.105 | 2.895 | 6.842 | 2.368 | 15.53 | 2.632 | 4.474 | 5.526 | 1.579 | 2.105 | 6.842 | 5 | 7.895 | 3.421 | 3.684 |
| ● *Nematalosa japonica* | 8.421 | 0.789 | 2.895 | 1.842 | 7.895 | 7.368 | 3.158 | 7.105 | 2.368 | 15.26 | 2.105 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 5.526 | 8.158 | 3.421 | 3.684 |
| ● *Clupanodon thrissa* | 8.684 | 0.789 | 2.895 | 1.842 | 7.895 | 7.105 | 3.158 | 7.105 | 2.368 | 15.53 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 5.526 | 7.632 | 3.421 | 3.684 |
| ● *Konosirus punctatus* | 8.684 | 0.789 | 2.895 | 1.842 | 7.895 | 7.105 | 3.158 | 6.842 | 2.368 | 15.53 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 5.526 | 7.895 | 3.421 | 3.684 |
| ● *Escualosa thoracata* | 9.211 | 0.789 | 2.895 | 2.105 | 7.632 | 7.368 | 3.158 | 7.105 | 2.105 | 15.53 | 2.895 | 4.737 | 5.526 | 1.579 | 2.105 | 5.789 | 5.526 | 6.579 | 3.421 | 3.947 |
| ● *Sardina pilchardus* | 8.947 | 0.789 | 3.158 | 1.579 | 7.632 | 7.368 | 3.158 | 6.842 | 2.632 | 14.74 | 2.895 | 4.211 | 5.526 | 1.316 | 2.105 | 6.316 | 5.526 | 7.895 | 3.421 | 3.947 |
| ● *Sardinops melanostictus* | 8.947 | 0.789 | 2.895 | 1.842 | 7.632 | 7.368 | 3.158 | 6.316 | 2.632 | 14.74 | 2.895 | 4.211 | 5.526 | 1.316 | 2.105 | 6.316 | 5 | 8.947 | 3.421 | 3.947 |
| ● *Brevoortia tyrannus* | 9.211 | 0.789 | 2.895 | 1.842 | 7.632 | 7.368 | 3.158 | 7.368 | 2.368 | 14.47 | 2.632 | 4.474 | 5.526 | 1.316 | 2.105 | 6.316 | 5.263 | 7.895 | 3.421 | 3.947 |
| □ *Alosa alosa* | 9.737 | 0.789 | 2.895 | 1.842 | 7.632 | 7.368 | 3.421 | 6.842 | 2.368 | 15.26 | 2.632 | 3.947 | 5.526 | 1.053 | 2.105 | 6.316 | 5.526 | 7.368 | 3.421 | 3.947 |
| □ *Alosa pseudoharengus* | 9.211 | 0.789 | 2.895 | 1.842 | 7.632 | 7.368 | 3.158 | 7.368 | 2.368 | 14.74 | 2.895 | 4.474 | 5.526 | 1.316 | 2.105 | 6.316 | 5.263 | 7.368 | 3.421 | 3.947 |
| ○ *Clupeichthys goniognathus* | 8.947 | 0.789 | 3.158 | 1.579 | 7.895 | 6.842 | 3.421 | 7.105 | 2.368 | 15.26 | 3.158 | 4.737 | 5.526 | 1.579 | 2.105 | 5.789 | 6.053 | 6.842 | 3.158 | 3.684 |
| ○ *Clupeichthys aesarnensis* | 8.684 | 0.789 | 3.158 | 1.579 | 7.895 | 6.842 | 3.421 | 7.632 | 2.368 | 15.26 | 3.158 | 4.737 | 5.526 | 1.579 | 2.105 | 5.789 | 6.053 | 6.579 | 3.158 | 3.684 |
| ○ *Clupeichthys perakensis* | 8.684 | 0.789 | 3.158 | 1.579 | 7.895 | 6.842 | 3.421 | 7.105 | 2.368 | 15.53 | 3.158 | 4.737 | 5.526 | 1.579 | 2.105 | 5.789 | 6.053 | 6.842 | 3.158 | 3.684 |
| ○ *Clupeoides sp. Chao Phraya* | 8.684 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.368 | 2.368 | 15.26 | 3.158 | 5 | 5.526 | 1.579 | 2.105 | 5.789 | 6.053 | 6.842 | 3.421 | 3.684 |
| ○ *Clupeoides borneensis* | 8.947 | 0.789 | 3.158 | 1.579 | 7.895 | 6.842 | 3.158 | 7.105 | 2.368 | 15.26 | 3.158 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 5.789 | 6.842 | 3.421 | 3.684 |
| ○ *Sundasalanx praecox* | 9.211 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.368 | 2.368 | 15 | 3.158 | 5 | 5.526 | 1.579 | 2.105 | 5.789 | 5.789 | 6.842 | 3.421 | 3.684 |
| ○ *Sundasalanx sp. Chao Phraya* | 9.211 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.105 | 2.368 | 15 | 3.421 | 5 | 5.526 | 1.579 | 2.105 | 5.789 | 5.789 | 6.842 | 3.421 | 3.684 |
| ○ *Sundasalanx mekongensis* | 9.211 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.632 | 2.368 | 15 | 3.684 | 5 | 5.526 | 1.579 | 2.105 | 5.789 | 5.789 | 6.053 | 3.421 | 3.684 |
| □ *Ehirava fluviatilis* | 9.211 | 0.789 | 2.895 | 1.579 | 7.895 | 6.579 | 3.158 | 7.105 | 2.368 | 15.26 | 3.684 | 5 | 5.526 | 1.579 | 2.105 | 5.789 | 5.789 | 6.579 | 3.421 | 3.684 |



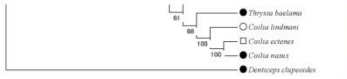
| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Gilchristella aestuaria* | 8.947 | 0.789 | 3.158 | 1.579 | 7.895 | 6.842 | 3.158 | 6.842 | 2.368 | 15 | 3.421 | 4.737 | 5.526 | 1.579 | 2.105 | 5.789 | 6.053 | 7.105 | 3.421 | 3.684 |
| □ *Clupeonella cultriventris* | 8.421 | 1.053 | 2.895 | 1.842 | 7.895 | 7.105 | 3.158 | 7.105 | 2.368 | 16.05 | 1.842 | 4.474 | 5.526 | 1.579 | 2.105 | 5.789 | 5.263 | 8.421 | 3.421 | 3.684 |
| ● *Clupea harengus* | 9.737 | 0.789 | 2.632 | 1.316 | 8.158 | 7.105 | 3.158 | 6.579 | 2.368 | 15.53 | 2.895 | 4.737 | 5.526 | 1.579 | 2.105 | 6.579 | 5.263 | 7.105 | 3.158 | 3.684 |
| ● *Clupea pallasii* | 9.737 | 0.789 | 2.632 | 1.316 | 8.158 | 7.105 | 3.158 | 6.579 | 2.368 | 15.53 | 2.895 | 4.737 | 5.526 | 1.579 | 2.105 | 6.579 | 5.263 | 7.105 | 3.158 | 3.684 |
| ● *Sprattus sprattus* | 9.737 | 0.789 | 2.632 | 1.316 | 8.158 | 7.105 | 3.158 | 6.579 | 2.368 | 15 | 2.895 | 4.737 | 5.526 | 1.579 | 2.105 | 6.316 | 5.526 | 7.368 | 3.421 | 3.684 |
| ● *Sprattus muelleri* | 9.474 | 0.789 | 2.895 | 1.579 | 8.158 | 6.842 | 3.158 | 6.842 | 2.368 | 15.53 | 2.368 | 4.474 | 5.526 | 1.579 | 2.105 | 6.579 | 5.526 | 7.105 | 3.421 | 3.684 |
| ● *Sprattus antipodum* | 9.737 | 0.789 | 2.895 | 1.579 | 8.158 | 6.842 | 3.158 | 6.579 | 2.368 | 15.53 | 2.368 | 4.474 | 5.526 | 1.579 | 2.105 | 6.579 | 5.263 | 7.368 | 3.421 | 3.684 |
| ○ *Potamalosa richmondia* | 9.474 | 0.789 | 2.895 | 1.842 | 7.895 | 6.842 | 3.158 | 6.579 | 2.368 | 15.53 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 5.526 | 5.526 | 8.158 | 3.421 | 3.684 |
| ● *Hyperlophus vittatus* | 9.474 | 0.789 | 2.895 | 1.842 | 7.895 | 6.842 | 3.158 | 6.842 | 2.368 | 15.53 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 5.526 | 5.526 | 7.895 | 3.421 | 3.684 |
| ● *Ethmidium maculatum* | 8.684 | 0.789 | 2.895 | 1.579 | 7.895 | 7.105 | 3.158 | 6.579 | 2.632 | 15.26 | 2.895 | 4.737 | 5.526 | 1.579 | 2.105 | 5.789 | 5.526 | 8.158 | 3.421 | 3.684 |
| ● *Jenkinsia lamprotaenia* | 8.947 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.632 | 2.368 | 15.53 | 2.105 | 4.474 | 5.263 | 1.579 | 2.105 | 6.842 | 4.737 | 8.158 | 3.421 | 3.684 |
| ● *Spratelloides delicatulus* | 8.947 | 0.789 | 3.158 | 1.316 | 8.158 | 6.842 | 2.895 | 6.053 | 2.368 | 15.26 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 6.316 | 5 | 8.947 | 3.421 | 3.947 |
| ● *Spratelloides gracilis* | 8.947 | 0.789 | 3.158 | 1.316 | 7.895 | 6.842 | 3.158 | 6.316 | 2.368 | 15 | 2.895 | 5 | 5.526 | 1.579 | 2.105 | 6.053 | 5 | 8.947 | 3.421 | 3.684 |
| ● *Etrumeus micropus* | 8.421 | 0.789 | 2.895 | 1.579 | 7.895 | 7.632 | 3.158 | 7.368 | 2.368 | 16.05 | 2.632 | 4.474 | 5.526 | 1.579 | 2.105 | 6.053 | 5.789 | 6.316 | 3.421 | 3.947 |
| ○ *Ilisha africana* | 7.632 | 0.789 | 3.158 | 1.579 | 7.632 | 7.105 | 3.158 | 8.158 | 2.368 | 16.58 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 6.053 | 6.316 | 5.526 | 3.421 | 3.947 |
| ○ *Pellona flavipinnis* | 7.895 | 0.789 | 3.158 | 1.579 | 7.632 | 6.842 | 3.158 | 8.158 | 2.368 | 16.32 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 6.316 | 6.579 | 5.526 | 3.421 | 3.684 |
| ○ *Ilisha elongata* | 7.368 | 0.789 | 3.158 | 1.579 | 7.895 | 7.105 | 3.158 | 8.421 | 2.105 | 16.05 | 2.368 | 4.737 | 5.526 | 1.579 | 2.105 | 6.316 | 6.316 | 6.316 | 3.421 | 3.684 |
| □ *Pellona ditchela* | 8.421 | 0.789 | 3.158 | 1.579 | 8.158 | 7.105 | 3.158 | 7.895 | 2.368 | 16.05 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 5.789 | 6.053 | 5.789 | 3.421 | 3.684 |
| □ *Anchoviella sp. LBP 2297* | 9.211 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.632 | 2.105 | 15.79 | 2.632 | 5.263 | 5.526 | 1.579 | 2.105 | 5.526 | 5.526 | 6.842 | 3.421 | 3.684 |
| □ *Lycengraulis grossidens* | 8.684 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.895 | 2.105 | 15.79 | 2.632 | 5.263 | 5.526 | 1.579 | 2.105 | 5.263 | 6.316 | 6.579 | 3.421 | 3.684 |
| ○ *Amazonsprattus scintilla* | 8.684 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.368 | 2.105 | 15.53 | 2.632 | 5.263 | 5.526 | 1.579 | 2.105 | 5.526 | 6.053 | 7.368 | 3.421 | 3.684 |
| ● *Engraulis encrasicolus* | 8.947 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.632 | 2.105 | 14.74 | 3.158 | 5 | 5.526 | 1.579 | 2.105 | 5.789 | 5.789 | 7.368 | 3.421 | 3.684 |
| ● *Engraulis japonicus* | 8.947 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 7.632 | 2.105 | 14.74 | 2.895 | 5 | 5.526 | 1.579 | 2.105 | 5.789 | 5.789 | 7.632 | 3.421 | 3.684 |
| ● *Stolephorus chinensis* | 8.684 | 0.789 | 2.895 | 1.579 | 8.158 | 7.105 | 3.158 | 7.368 | 2.105 | 15.79 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 5.526 | 6.053 | 7.105 | 3.421 | 3.684 |
| ● *Stolephorus waitei* | 8.684 | 0.789 | 2.895 | 1.579 | 8.158 | 7.105 | 3.158 | 7.368 | 2.105 | 15.79 | 2.632 | 4.737 | 5.526 | 1.579 | 2.105 | 5.526 | 6.053 | 7.105 | 3.421 | 3.684 |
| ○ *Lycothrissa crocodilus* | 8.684 | 0.789 | 2.895 | 1.579 | 7.895 | 6.842 | 3.158 | 8.684 | 2.105 | 15.53 | 2.895 | 5 | 5.526 | 1.579 | 2.105 | 5.789 | 6.316 | 5.263 | 3.421 | 3.947 |
| □ *Setipinna melanochir* | 9.211 | 0.789 | 2.895 | 1.579 | 8.158 | 6.842 | 3.158 | 7.895 | 2.105 | 16.05 | 2.632 | 5.263 | 5.526 | 1.579 | 2.105 | 5.526 | 6.316 | 5.526 | 3.421 | 3.421 |
| ● *Coilia reynaldi* | 8.947 | 0.789 | 2.895 | 1.579 | 8.421 | 6.842 | 3.158 | 7.895 | 2.105 | 15.26 | 2.632 | 5.263 | 5.526 | 1.579 | 2.105 | 5.526 | 5.526 | 6.842 | 3.421 | 3.684 |
| ● *Thryssa baelama* | 8.947 | 0.789 | 2.895 | 1.579 | 8.158 | 6.842 | 3.158 | 7.105 | 2.105 | 15.53 | 2.632 | 5.263 | 5.526 | 1.579 | 2.105 | 5.526 | 5.789 | 7.368 | 3.421 | 3.684 |
| ○ *Coilia lindmani* | 8.421 | 0.789 | 2.895 | 1.579 | 8.158 | 6.842 | 3.158 | 7.105 | 2.105 | 15.79 | 2.632 | 5.263 | 5.526 | 1.579 | 2.105 | 6.053 | 5.789 | 7.105 | 3.421 | 3.684 |
| □ *Coilia ectenes* | 8.421 | 0.789 | 2.895 | 1.579 | 8.158 | 6.842 | 3.158 | 7.105 | 2.105 | 15.79 | 2.632 | 5.263 | 5.526 | 1.579 | 2.105 | 6.053 | 5.789 | 7.105 | 3.421 | 3.684 |
| ● *Coilia nasus* | 8.421 | 0.789 | 2.895 | 1.579 | 8.158 | 6.842 | 3.158 | 7.105 | 2.105 | 15.79 | 2.632 | 5.263 | 5.526 | 1.579 | 2.105 | 6.053 | 5.789 | 7.105 | 3.421 | 3.684 |
| ● *Denticeps clupeoides* | 7.105 | 0.789 | 2.632 | 1.579 | 7.895 | 7.105 | 3.158 | 7.105 | 2.368 | 16.32 | 4.474 | 5.263 | 5.789 | 1.579 | 2.105 | 5.263 | 6.316 | 6.316 | 3.421 | 3.421 |

cyb



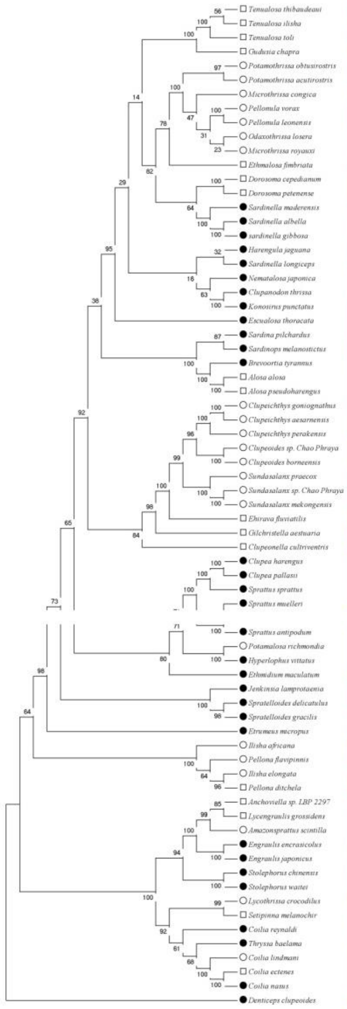


| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 11.11 | 0.309 | 1.235 | 3.395 | 5.556 | 6.173 | 1.235 | 7.099 | 2.16 | 18.52 | 3.704 | 2.778 | 7.099 | 1.852 | 2.469 | 6.79 | 6.173 | 5.864 | 2.469 | 4.012 |
| □ *Tenualosa ilisha* | 10.49 | 0.309 | 1.235 | 3.395 | 6.173 | 5.864 | 1.852 | 6.173 | 2.16 | 17.59 | 3.395 | 2.469 | 7.407 | 1.852 | 2.469 | 7.407 | 6.481 | 7.407 | 2.469 | 3.395 |
| □ *Tenualosa toli* | 10.49 | 0.309 | 1.235 | 3.395 | 5.556 | 6.173 | 1.852 | 4.63 | 2.16 | 18.52 | 3.704 | 2.469 | 7.407 | 1.852 | 2.469 | 7.099 | 6.79 | 8.025 | 2.469 | 3.395 |
| □ *Gudusia chapra* | 10.8 | 0.309 | 1.235 | 3.395 | 6.173 | 5.864 | 1.235 | 7.407 | 2.16 | 16.98 | 4.321 | 2.778 | 7.407 | 2.16 | 2.469 | 7.407 | 6.173 | 5.556 | 2.469 | 3.704 |
| ○ *Potamothrissa obtusirostris* | 11.11 | 0.309 | 1.235 | 3.395 | 5.864 | 5.864 | 1.852 | 7.099 | 2.16 | 18.83 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 7.099 | 6.173 | 5.864 | 2.469 | 3.395 |
| ○ *Potamothrissa acutirostris* | 11.42 | 0.309 | 1.235 | 3.395 | 5.864 | 5.864 | 1.852 | 6.79 | 2.16 | 18.83 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 6.79 | 6.173 | 6.173 | 2.469 | 3.395 |
| ○ *Microthrissa congica* | 10.49 | 0.309 | 1.235 | 3.395 | 5.864 | 5.864 | 1.543 | 6.79 | 2.16 | 18.83 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 7.407 | 6.481 | 6.173 | 2.469 | 3.704 |
| ○ *Pellonula vorax* | 11.11 | 0.309 | 1.235 | 3.395 | 5.864 | 5.864 | 1.543 | 6.79 | 2.16 | 18.83 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 7.407 | 6.173 | 5.864 | 2.469 | 3.704 |
| ○ *Pellonula leonensis* | 11.11 | 0.309 | 1.235 | 3.395 | 5.864 | 5.864 | 1.543 | 6.481 | 2.16 | 18.83 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 7.407 | 5.864 | 6.481 | 2.469 | 3.704 |
| ○ *Odaxothrissa losera* | 10.49 | 0.309 | 1.235 | 3.395 | 5.864 | 6.173 | 1.543 | 7.099 | 2.16 | 18.52 | 2.778 | 2.469 | 7.407 | 1.852 | 2.469 | 7.099 | 6.481 | 6.481 | 2.469 | 3.704 |
| ○ *Microthrissa royauxi* | 10.8 | 0.309 | 1.235 | 3.395 | 5.864 | 5.864 | 1.543 | 7.099 | 2.16 | 18.52 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 7.716 | 5.864 | 6.173 | 2.469 | 3.704 |
| □ *Ethmalosa fimbriata* | 10.8 | 0.309 | 1.235 | 3.395 | 7.099 | 5.864 | 1.852 | 6.481 | 2.16 | 17.59 | 3.086 | 2.778 | 7.407 | 1.852 | 2.469 | 7.099 | 6.481 | 6.173 | 2.469 | 3.395 |
| □ *Dorosoma cepedianum* | 10.8 | 0.309 | 1.235 | 3.395 | 6.173 | 5.864 | 1.543 | 6.79 | 2.16 | 19.14 | 2.778 | 2.469 | 7.407 | 1.852 | 2.469 | 7.407 | 6.173 | 6.173 | 2.469 | 3.395 |
| □ *Dorosoma petenense* | 10.19 | 0.309 | 1.235 | 3.395 | 6.173 | 6.173 | 1.235 | 7.407 | 2.16 | 18.83 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 7.099 | 6.79 | 5.556 | 2.469 | 3.704 |
| ● *Sardinella maderensis* | 10.49 | 0.309 | 1.235 | 3.395 | 5.864 | 6.173 | 1.235 | 6.79 | 2.16 | 18.52 | 3.395 | 2.778 | 7.407 | 1.852 | 2.469 | 7.407 | 6.481 | 5.864 | 2.469 | 3.704 |
| ● *Sardinella albella* | 10.8 | 0.309 | 1.235 | 3.395 | 6.481 | 5.864 | 1.235 | 6.481 | 2.16 | 18.21 | 3.395 | 2.778 | 7.407 | 1.852 | 2.469 | 7.407 | 6.481 | 6.173 | 2.469 | 3.395 |
| ● *sardinella gibbosa* | 10.8 | 0.617 | 1.235 | 3.395 | 6.481 | 5.864 | 1.235 | 6.481 | 2.16 | 18.21 | 3.395 | 2.778 | 7.407 | 1.852 | 2.469 | 7.407 | 6.481 | 6.173 | 2.16 | 3.395 |
| ● *Harengula jaguana* | 11.42 | 0.309 | 1.235 | 3.395 | 6.173 | 5.864 | 1.543 | 7.099 | 2.469 | 19.14 | 2.778 | 2.778 | 7.407 | 1.852 | 2.469 | 6.173 | 5.864 | 5.864 | 2.469 | 3.704 |
| ● *Sardinella longiceps* | 10.8 | 0.309 | 1.235 | 3.395 | 6.173 | 5.864 | 1.543 | 6.79 | 2.16 | 18.52 | 3.086 | 2.778 | 7.407 | 1.852 | 2.469 | 7.099 | 5.864 | 6.481 | 2.469 | 3.704 |
| ● *Nematalosa japonica* | 10.8 | 0.309 | 1.235 | 3.395 | 6.173 | 6.173 | 1.543 | 6.79 | 2.16 | 18.21 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 7.099 | 6.481 | 6.173 | 2.469 | 3.704 |
| ● *Clupanodon thrissa* | 11.11 | 0.309 | 1.235 | 3.395 | 4.938 | 6.481 | 1.543 | 6.481 | 2.16 | 19.14 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 7.099 | 5.864 | 6.79 | 2.469 | 3.704 |
| ● *Konosirus punctatus* | 11.11 | 0.309 | 1.235 | 3.395 | 5.247 | 6.481 | 1.543 | 6.481 | 2.16 | 19.14 | 3.086 | 2.469 | 7.407 | 1.852 | 2.469 | 7.099 | 5.864 | 6.481 | 2.469 | 3.704 |
| ● *Escualosa thoracata* | 10.8 | 0.309 | 1.235 | 3.395 | 6.173 | 5.864 | 1.543 | 6.481 | 2.16 | 18.83 | 2.778 | 2.778 | 7.407 | 1.852 | 2.469 | 6.79 | 6.173 | 7.099 | 2.469 | 3.395 |
| ● *Sardina pilchardus* | 11.11 | 0.309 | 1.235 | 3.086 | 6.79 | 6.481 | 1.235 | 5.556 | 2.16 | 17.59 | 3.086 | 3.086 | 7.407 | 1.852 | 2.469 | 6.79 | 4.938 | 8.951 | 2.469 | 3.395 |
| ● *Sardinops melanostictus* | 10.19 | 0.309 | 1.235 | 3.395 | 6.79 | 5.864 | 1.235 | 5.556 | 2.16 | 17.9 | 3.704 | 3.395 | 7.407 | 1.852 | 2.469 | 6.481 | 5.864 | 8.025 | 2.778 | 3.395 |
| ● *Brevoortia tyrannus* | 10.8 | 0.309 | 1.235 | 3.395 | 7.407 | 6.173 | 1.543 | 6.173 | 2.16 | 17.9 | 3.086 | 2.778 | 7.407 | 1.852 | 2.469 | 6.79 | 6.481 | 6.173 | 2.469 | 3.395 |
| □ *Alosa alosa* | 10.49 | 0.309 | 1.235 | 3.395 | 7.407 | 6.173 | 1.543 | 6.79 | 2.16 | 17.9 | 3.086 | 2.778 | 7.407 | 1.852 | 2.469 | 7.099 | 6.173 | 5.864 | 2.469 | 3.395 |
| □ *Alosa pseudoharengus* | 10.49 | 0.309 | 1.235 | 3.395 | 7.407 | 6.173 | 1.543 | 6.481 | 2.16 | 17.9 | 3.086 | 2.778 | 7.407 | 1.852 | 2.469 | 7.099 | 6.481 | 5.864 | 2.469 | 3.395 |
| ○ *Clupeichthys goniognathus* | 10.49 | 0 | 1.235 | 3.395 | 6.173 | 5.864 | 0.926 | 8.333 | 2.16 | 17.9 | 2.778 | 3.704 | 7.407 | 1.852 | 2.469 | 7.407 | 6.481 | 5.556 | 2.469 | 3.395 |
| ○ *Clupeichthys aesarnensis* | 10.8 | 0 | 1.235 | 3.395 | 6.481 | 5.864 | 0.926 | 8.333 | 2.16 | 17.28 | 2.778 | 3.395 | 7.407 | 1.852 | 2.469 | 7.099 | 6.79 | 5.864 | 2.469 | 3.395 |
| ○ *Clupeichthys perakensis* | 11.11 | 0 | 1.235 | 3.395 | 6.481 | 5.864 | 0.926 | 8.025 | 2.16 | 17.59 | 2.778 | 3.704 | 7.407 | 1.852 | 2.469 | 6.79 | 6.481 | 5.864 | 2.469 | 3.395 |
| ○ *Clupeoides sp. Chao Phraya* | 10.8 | 0 | 1.235 | 3.395 | 6.173 | 5.864 | 0.926 | 8.333 | 2.16 | 17.59 | 3.086 | 3.704 | 7.407 | 1.852 | 2.469 | 6.481 | 6.79 | 5.864 | 2.469 | 3.395 |
| ○ *Clupeoides borneensis* | 10.8 | 0 | 1.235 | 3.395 | 6.173 | 6.173 | 0.926 | 7.407 | 2.16 | 18.21 | 3.086 | 4.012 | 7.407 | 1.852 | 2.469 | 6.173 | 6.79 | 5.864 | 2.469 | 3.395 |
| ○ *Sundasalanx praecox* | 9.877 | 0 | 1.235 | 3.395 | 6.79 | 5.864 | 1.235 | 8.025 | 2.16 | 18.52 | 2.469 | 3.395 | 7.407 | 1.852 | 2.469 | 7.407 | 7.099 | 4.938 | 2.469 | 3.395 |
| ○ *Sundasalanx sp. Chao Phraya* | 9.877 | 0 | 1.235 | 3.395 | 7.099 | 5.864 | 1.235 | 7.716 | 2.16 | 18.21 | 2.469 | 3.395 | 7.407 | 1.852 | 2.469 | 7.407 | 7.407 | 4.938 | 2.469 | 3.395 |
| ○ *Sundasalanx mekongensis* | 9.877 | 0 | 1.235 | 3.395 | 6.79 | 5.864 | 0.926 | 8.333 | 2.16 | 18.52 | 2.469 | 3.704 | 7.407 | 1.852 | 2.469 | 7.407 | 7.407 | 4.321 | 2.469 | 3.395 |
| □ *Ehirava fluviatilis* | 11.73 | 0 | 1.235 | 3.395 | 6.79 | 5.864 | 0.926 | 8.025 | 2.469 | 18.21 | 2.16 | 3.704 | 7.099 | 1.852 | 2.469 | 6.481 | 6.481 | 5.247 | 2.469 | 3.395 |
| □ *Gilchristella aestuaria* | 10.8 | 0.309 | 1.235 | 3.395 | 6.481 | 6.173 | 1.235 | 7.407 | 2.16 | 19.14 | 2.469 | 3.395 | 7.407 | 1.852 | 2.469 | 5.864 | 6.79 | 5.556 | 2.469 | 3.395 |
| □ *Clupeonella cultriventris* | 11.73 | 0.309 | 1.235 | 3.395 | 6.481 | 5.864 | 1.543 | 6.79 | 2.16 | 17.9 | 3.395 | 3.086 | 7.407 | 1.852 | 2.469 | 6.173 | 5.864 | 6.173 | 2.469 | 3.704 |
| ● *Clupea harengus* | 11.42 | 0.309 | 1.235 | 3.395 | 6.173 | 5.247 | 1.543 | 5.864 | 2.16 | 18.52 | 4.63 | 3.086 | 7.099 | 1.852 | 2.469 | 7.099 | 6.79 | 5.247 | 2.469 | 3.395 |
| ● *Clupea pallasii* | 11.42 | 0.309 | 1.235 | 3.395 | 6.173 | 5.247 | 1.543 | 5.864 | 2.16 | 18.52 | 4.63 | 3.086 | 7.099 | 1.852 | 2.469 | 7.099 | 6.79 | 5.247 | 2.469 | 3.395 |
| ● *Sprattus sprattus* | 11.42 | 0.309 | 1.235 | 3.395 | 6.173 | 5.247 | 1.543 | 6.173 | 2.16 | 18.83 | 4.012 | 2.778 | 7.099 | 1.852 | 2.469 | 7.407 | 6.79 | 5.247 | 2.469 | 3.395 |
| ● *Sprattus muelleri* | 11.73 | 0.309 | 1.235 | 3.395 | 6.173 | 5.247 | 1.543 | 6.79 | 2.16 | 18.21 | 4.012 | 3.086 | 7.099 | 1.852 | 2.469 | 7.099 | 6.481 | 5.247 | 2.469 | 3.395 |
| ● *Sprattus antipodum* | 11.73 | 0.309 | 1.235 | 3.395 | 6.173 | 5.247 | 1.543 | 6.79 | 2.16 | 18.21 | 4.012 | 3.086 | 7.099 | 1.852 | 2.469 | 7.099 | 6.481 | 5.247 | 2.469 | 3.395 |
| ○ *Potamalosa richmondia* | 10.8 | 0 | 1.235 | 3.395 | 6.173 | 5.864 | 1.852 | 6.173 | 2.16 | 19.44 | 2.778 | 3.086 | 7.407 | 1.852 | 2.469 | 6.481 | 7.099 | 5.864 | 2.469 | 3.395 |
| ● *Hyperlophus vittatus* | 11.11 | 0 | 1.235 | 3.395 | 6.481 | 5.556 | 1.543 | 5.556 | 2.16 | 19.14 | 2.778 | 3.395 | 7.099 | 1.852 | 2.469 | 6.173 | 7.407 | 6.79 | 2.469 | 3.395 |
| ● *Ethmidium maculatum* | 10.8 | 0 | 0.926 | 3.395 | 6.173 | 6.173 | 1.852 | 5.864 | 2.16 | 19.14 | 2.778 | 3.704 | 7.407 | 1.852 | 2.469 | 6.173 | 6.173 | 7.099 | 2.469 | 3.395 |
| ● *Jenkinsia lamprotaenia* | 12.65 | 0 | 1.235 | 3.086 | 6.173 | 6.173 | 1.235 | 5.864 | 2.16 | 18.21 | 3.704 | 2.469 | 7.716 | 1.852 | 2.469 | 6.79 | 5.556 | 6.481 | 2.778 | 3.395 |
| ● *Spratelloides delicatulus* | 11.73 | 0 | 1.543 | 3.086 | 6.173 | 6.173 | 1.543 | 4.938 | 2.16 | 18.52 | 3.086 | 2.16 | 7.716 | 1.852 | 2.469 | 7.407 | 5.556 | 7.407 | 2.778 | 3.704 |
| ● *Spratelloides gracilis* | 11.73 | 0 | 1.235 | 3.086 | 6.173 | 6.173 | 1.543 | 5.247 | 2.16 | 18.21 | 3.704 | 2.778 | 7.716 | 1.852 | 2.469 | 7.099 | 5.556 | 6.79 | 2.778 | 3.704 |
| ● *Etrumeus micropus* | 10.8 | 0 | 1.235 | 3.395 | 5.864 | 6.173 | 1.543 | 6.173 | 2.16 | 18.21 | 3.086 | 3.086 | 7.407 | 1.852 | 2.469 | 6.79 | 6.173 | 7.716 | 2.469 | 3.395 |
| ○ *Ilisha africana* | 10.8 | 0 | 1.235 | 3.395 | 5.247 | 5.556 | 1.235 | 6.79 | 2.16 | 18.52 | 4.321 | 3.086 | 7.407 | 1.852 | 2.469 | 7.099 | 7.716 | 4.938 | 2.469 | 3.704 |
| ○ *Pellona flavipinnis* | 10.8 | 0 | 1.235 | 3.395 | 5.247 | 5.864 | 1.235 | 6.173 | 2.16 | 18.52 | 4.63 | 3.395 | 7.407 | 1.852 | 2.469 | 6.173 | 8.025 | 4.938 | 2.469 | 4.012 |
| ○ *Ilisha elongata* | 10.49 | 0 | 1.235 | 3.395 | 5.247 | 5.864 | 1.543 | 6.173 | 2.16 | 18.83 | 4.63 | 3.395 | 7.407 | 1.852 | 2.469 | 6.481 | 7.716 | 4.938 | 2.469 | 3.704 |
| □ *Pellona ditchela* | 9.877 | 0 | 1.235 | 3.395 | 5.864 | 6.173 | 1.543 | 5.864 | 2.16 | 18.52 | 4.63 | 3.395 | 7.407 | 2.16 | 2.469 | 6.173 | 8.642 | 4.63 | 2.469 | 3.395 |
| □ *Anchoviella sp. LBP 2297* | 11.11 | 0 | 1.235 | 3.395 | 6.173 | 5.247 | 1.235 | 6.79 | 2.16 | 18.52 | 3.395 | 3.395 | 7.407 | 1.852 | 2.469 | 6.481 | 7.407 | 6.173 | 2.469 | 3.086 |
| □ *Lycengraulis grossidens* | 11.11 | 0 | 1.235 | 3.395 | 5.864 | 5.556 | 1.235 | 6.173 | 2.16 | 18.83 | 3.395 | 3.395 | 7.407 | 1.852 | 2.469 | 6.173 | 7.407 | 6.79 | 2.469 | 3.086 |
| ○ *Amazonsprattus scintilla* | 11.11 | 0 | 1.235 | 3.395 | 6.173 | 5.247 | 1.235 | 5.556 | 2.16 | 18.52 | 3.704 | 3.395 | 7.407 | 1.852 | 2.469 | 6.481 | 7.407 | 7.099 | 2.469 | 3.086 |
| ● *Engraulis encrasicolus* | 11.42 | 0 | 1.235 | 3.395 | 5.864 | 5.247 | 1.235 | 5.864 | 2.16 | 18.83 | 3.395 | 3.395 | 7.407 | 1.852 | 2.469 | 6.173 | 7.716 | 6.79 | 2.469 | 3.086 |
| ● *Engraulis japonicus* | 11.42 | 0 | 1.235 | 3.395 | 5.864 | 5.247 | 1.235 | 5.864 | 2.16 | 19.14 | 3.086 | 3.395 | 7.407 | 1.852 | 2.469 | 6.173 | 7.716 | 6.79 | 2.469 | 3.086 |
| ● *Stolephorus chinensis* | 12.04 | 0 | 1.235 | 3.704 | 5.864 | 5.247 | 0.926 | 6.481 | 2.16 | 17.59 | 3.704 | 3.086 | 7.099 | 1.852 | 2.469 | 7.099 | 6.79 | 6.79 | 2.469 | 3.395 |
| ● *Stolephorus waitei* | 12.04 | 0 | 1.235 | 3.704 | 5.864 | 5.247 | 0.926 | 6.481 | 2.16 | 17.59 | 3.704 | 3.086 | 7.099 | 1.852 | 2.469 | 7.099 | 6.79 | 6.79 | 2.469 | 3.395 |
| ○ *Lycothrissa crocodilus* | 11.73 | 0 | 1.235 | 3.395 | 6.173 | 4.938 | 0.926 | 6.79 | 2.16 | 18.52 | 3.704 | 3.704 | 7.407 | 1.852 | 2.469 | 6.79 | 7.099 | 5.556 | 2.469 | 3.086 |
| □ *Setipinna melanochir* | 10.8 | 0 | 1.235 | 3.395 | 6.481 | 5.247 | 0.926 | 6.79 | 2.16 | 19.14 | 3.395 | 3.395 | 7.407 | 2.16 | 2.16 | 6.481 | 7.716 | 5.247 | 2.469 | 3.395 |
| ● *Coilia reynaldi* | 11.11 | 0 | 1.235 | 3.395 | 6.173 | 5.247 | 0.926 | 6.173 | 2.16 | 18.83 | 3.395 | 4.012 | 7.407 | 1.852 | 2.469 | 5.556 | 8.025 | 6.173 | 2.469 | 3.395 |



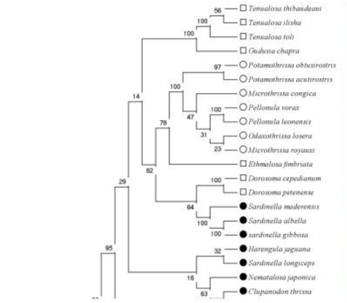
| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ● *Thryssa baelama* | 11.73 | 0 | 1.235 | 3.395 | 5.864 | 5.247 | 1.235 | 6.481 | 2.16 | 19.14 | 3.086 | 3.395 | 7.407 | 1.852 | 2.469 | 6.173 | 7.716 | 5.864 | 2.469 | 3.086 |
| ○ *Coilia lindmani* | 11.11 | 0 | 1.235 | 3.395 | 5.864 | 5.247 | 0.926 | 6.79 | 2.16 | 18.52 | 3.704 | 3.395 | 7.407 | 1.852 | 2.469 | 6.481 | 7.407 | 6.173 | 2.469 | 3.395 |
| □ *Coilia ectenes* | 11.11 | 0 | 1.235 | 3.395 | 5.864 | 5.247 | 0.926 | 6.79 | 2.16 | 18.52 | 3.704 | 3.395 | 7.407 | 1.852 | 2.469 | 6.481 | 7.716 | 5.864 | 2.469 | 3.395 |
| ● *Coilia nasus* | 11.11 | 0 | 1.235 | 3.395 | 5.864 | 5.247 | 0.926 | 6.79 | 2.16 | 18.52 | 3.704 | 3.395 | 7.407 | 1.852 | 2.469 | 6.481 | 7.716 | 5.864 | 2.469 | 3.395 |
| ● *Denticeps clupeoides* | 10.19 | 0 | 1.235 | 3.395 | 5.864 | 5.247 | 1.235 | 8.025 | 2.16 | 18.52 | 3.395 | 4.012 | 7.407 | 2.16 | 2.469 | 6.173 | 7.407 | 4.938 | 2.469 | 3.704 |

nd1

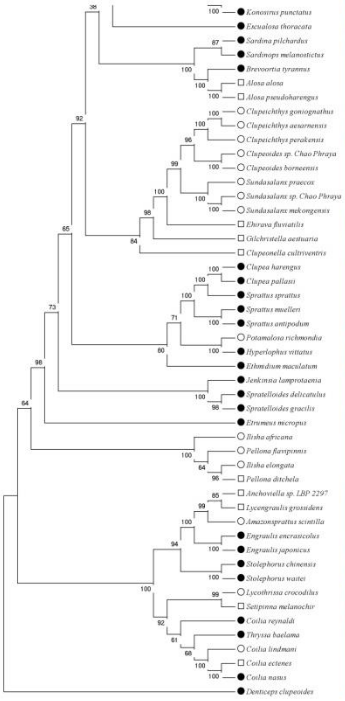


| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 12.61 | 0.86 | 0.573 | 1.433 | 3.152 | 5.158 | 3.152 | 6.877 | 2.579 | 18.05 | 4.871 | 1.719 | 5.444 | 3.725 | 1.719 | 6.59 | 10.6 | 5.731 | 3.152 | 2.006 |
| □ *Tenualosa ilisha* | 12.03 | 0.86 | 0 | 1.433 | 4.011 | 5.444 | 2.865 | 7.45 | 2.579 | 18.34 | 5.158 | 2.006 | 5.731 | 3.725 | 1.719 | 6.877 | 10.32 | 4.011 | 3.152 | 2.292 |
| □ *Tenualosa toli* | 12.89 | 0.86 | 0.287 | 1.719 | 3.438 | 5.158 | 3.725 | 6.304 | 2.579 | 18.05 | 5.158 | 1.719 | 5.158 | 3.725 | 1.719 | 6.59 | 10.6 | 5.158 | 3.152 | 2.006 |
| □ *Gudusia chapra* | 10.89 | 0.86 | 0 | 1.719 | 2.579 | 4.871 | 2.865 | 8.596 | 2.292 | 17.77 | 4.871 | 2.006 | 5.731 | 3.725 | 1.433 | 6.304 | 14.61 | 3.152 | 3.152 | 2.579 |
| ○ *Potamothrissa obtusirostris* | 12.68 | 1.153 | 0 | 2.017 | 2.882 | 4.611 | 2.017 | 6.052 | 2.594 | 20.17 | 3.458 | 2.882 | 5.764 | 3.458 | 1.153 | 5.476 | 14.41 | 3.458 | 3.17 | 2.594 |
| ○ *Potamothrissa acutirostris* | 12.97 | 1.153 | 0 | 2.017 | 2.882 | 4.611 | 2.017 | 6.916 | 2.594 | 19.6 | 3.746 | 2.882 | 5.764 | 3.458 | 1.153 | 5.476 | 13.83 | 3.17 | 3.17 | 2.594 |
| ○ *Microthrissa congica* | 12.39 | 1.153 | 0 | 2.017 | 2.594 | 4.611 | 2.017 | 6.628 | 2.594 | 20.46 | 3.458 | 2.882 | 5.764 | 3.458 | 1.153 | 5.187 | 14.41 | 3.458 | 3.17 | 2.594 |
| ○ *Pellonula vorax* | 13.26 | 1.153 | 0 | 2.017 | 2.882 | 4.611 | 2.017 | 6.628 | 2.594 | 19.31 | 3.746 | 3.17 | 5.764 | 3.458 | 1.153 | 5.187 | 13.54 | 3.746 | 3.17 | 2.594 |
| ○ *Pellonula leonensis* | 13.54 | 1.153 | 0 | 2.017 | 2.882 | 4.611 | 2.305 | 6.916 | 2.594 | 19.31 | 3.458 | 2.594 | 5.764 | 3.458 | 1.153 | 5.476 | 13.54 | 3.458 | 3.17 | 2.594 |
| ○ *Odaxothrissa losera* | 12.39 | 1.153 | 0 | 2.017 | 2.594 | 4.611 | 2.305 | 6.916 | 2.594 | 19.88 | 2.882 | 3.458 | 5.764 | 3.458 | 1.153 | 5.187 | 14.7 | 3.458 | 3.17 | 2.305 |
| ○ *Microthrissa royauxi* | 13.26 | 1.153 | 0 | 2.017 | 2.882 | 4.611 | 2.305 | 6.916 | 2.594 | 19.31 | 3.458 | 2.882 | 5.476 | 3.458 | 1.153 | 5.476 | 14.41 | 3.17 | 3.17 | 2.305 |
| □ *Ethmalosa fimbriata* | 12.39 | 1.153 | 0.865 | 1.441 | 3.17 | 4.611 | 2.017 | 6.916 | 2.594 | 19.31 | 3.458 | 2.305 | 5.476 | 3.746 | 1.441 | 6.052 | 12.97 | 4.323 | 3.17 | 2.594 |
| □ *Dorosoma cepedianum* | 10.95 | 0.865 | 0.865 | 1.729 | 2.882 | 4.611 | 2.305 | 6.052 | 2.594 | 19.6 | 3.458 | 2.305 | 5.476 | 3.458 | 1.441 | 6.34 | 14.7 | 4.899 | 3.17 | 2.305 |
| □ *Dorosoma petenense* | 10.66 | 0.865 | 1.153 | 1.441 | 3.17 | 4.611 | 2.305 | 6.628 | 2.594 | 19.31 | 4.323 | 2.017 | 5.476 | 3.746 | 1.441 | 6.34 | 14.12 | 4.323 | 3.17 | 2.305 |
| ● *Sardinella maderensis* | 12.68 | 1.153 | 0.288 | 1.729 | 2.882 | 4.899 | 2.017 | 5.764 | 2.305 | 19.88 | 3.458 | 2.594 | 5.476 | 3.746 | 1.441 | 6.34 | 12.97 | 4.611 | 3.17 | 2.594 |
| ● *Sardinella albella* | 11.82 | 1.153 | 0.576 | 1.729 | 2.882 | 4.611 | 2.305 | 6.34 | 2.594 | 19.88 | 2.882 | 2.305 | 5.476 | 3.458 | 1.441 | 6.628 | 12.97 | 5.187 | 3.17 | 2.594 |
| ● *sardinella gibbosa* | 11.82 | 1.153 | 0.576 | 1.729 | 2.882 | 4.611 | 2.305 | 6.628 | 2.594 | 19.88 | 2.882 | 2.305 | 5.476 | 3.458 | 1.441 | 6.628 | 12.97 | 4.899 | 3.17 | 2.594 |
| ● *Harengula jaguana* | 13.54 | 1.153 | 0.288 | 1.729 | 3.17 | 5.187 | 2.882 | 6.052 | 2.594 | 17.87 | 4.035 | 2.594 | 5.187 | 3.458 | 1.441 | 6.34 | 12.1 | 4.899 | 3.17 | 2.305 |
| ● *Sardinella longiceps* | 12.97 | 1.153 | 0.288 | 2.017 | 3.458 | 4.611 | 2.017 | 5.187 | 2.882 | 19.31 | 4.035 | 2.594 | 5.476 | 3.17 | 1.153 | 6.052 | 12.97 | 4.611 | 3.17 | 2.882 |
| ● *Nematalosa japonica* | 12.39 | 0.865 | 0.288 | 1.729 | 2.594 | 4.899 | 2.305 | 6.628 | 2.594 | 19.88 | 3.458 | 2.017 | 5.476 | 3.746 | 1.441 | 6.052 | 13.26 | 4.611 | 3.17 | 2.594 |
| ● *Clupanodon thrissa* | 12.1 | 0.576 | 0.865 | 1.729 | 3.17 | 4.899 | 2.305 | 7.493 | 2.594 | 19.88 | 3.746 | 1.729 | 5.476 | 3.458 | 1.441 | 5.764 | 14.41 | 2.594 | 3.17 | 2.594 |
| ● *Konosirus punctatus* | 13.83 | 0.576 | 0.288 | 2.017 | 2.594 | 4.611 | 2.594 | 6.34 | 2.594 | 19.88 | 2.882 | 1.441 | 5.476 | 3.458 | 1.441 | 6.916 | 12.68 | 4.611 | 3.17 | 2.594 |
| ● *Escualosa thoracata* | 13.54 | 1.153 | 0 | 1.441 | 3.458 | 5.476 | 1.441 | 6.052 | 2.594 | 19.88 | 3.746 | 2.305 | 5.476 | 4.611 | 1.153 | 4.899 | 12.68 | 4.611 | 3.17 | 2.305 |
| ● *Sardina pilchardus* | 13.26 | 0.576 | 0.576 | 1.729 | 3.458 | 4.899 | 1.729 | 5.476 | 2.305 | 18.44 | 4.035 | 1.441 | 5.476 | 4.323 | 1.153 | 7.781 | 12.39 | 5.476 | 3.17 | 2.305 |
| ● *Sardinops melanostictus* | 14.7 | 0.576 | 0 | 1.441 | 2.882 | 4.611 | 2.017 | 5.476 | 2.305 | 18.73 | 4.611 | 2.594 | 5.476 | 4.035 | 1.153 | 6.34 | 12.68 | 4.323 | 3.17 | 2.882 |
| ● *Brevoortia tyrannus* | 12.97 | 0.576 | 0.288 | 1.441 | 2.594 | 4.899 | 2.017 | 4.899 | 2.305 | 19.02 | 4.323 | 2.017 | 5.764 | 4.035 | 1.153 | 7.493 | 13.26 | 4.899 | 3.17 | 2.882 |
| □ *Alosa alosa* | 12.68 | 0.288 | 0.288 | 1.441 | 2.594 | 4.899 | 2.017 | 6.34 | 2.305 | 18.73 | 4.323 | 2.305 | 5.764 | 4.035 | 1.153 | 7.781 | 13.83 | 3.17 | 3.17 | 2.882 |
| □ *Alosa pseudoharengus* | 12.39 | 0.576 | 0.288 | 1.441 | 2.594 | 4.899 | 2.017 | 6.34 | 2.305 | 19.02 | 4.323 | 2.305 | 5.764 | 4.035 | 1.153 | 6.916 | 14.12 | 3.458 | 3.17 | 2.882 |
| ○ *Clupeichthys goniognathus* | 11.82 | 0.865 | 0.576 | 1.441 | 4.323 | 5.187 | 2.305 | 6.628 | 2.594 | 18.44 | 4.611 | 2.305 | 6.052 | 2.882 | 1.729 | 8.357 | 10.09 | 4.035 | 3.17 | 2.594 |
| ○ *Clupeichthys aesarnensis* | 12.39 | 0.865 | 0.288 | 1.153 | 4.035 | 5.187 | 2.017 | 7.205 | 2.882 | 19.31 | 4.611 | 2.594 | 5.764 | 3.458 | 1.729 | 7.781 | 9.51 | 3.746 | 3.17 | 2.305 |
| ○ *Clupeichthys perakensis* | 12.39 | 0.865 | 0.288 | 1.441 | 4.611 | 5.764 | 2.017 | 6.916 | 2.882 | 19.88 | 4.035 | 2.017 | 5.187 | 3.17 | 1.729 | 7.781 | 10.09 | 3.458 | 3.17 | 2.305 |
| ○ *Clupeoides sp. Chao Phraya* | 10.95 | 1.153 | 0.576 | 1.153 | 3.746 | 4.899 | 2.305 | 8.069 | 2.594 | 19.31 | 4.323 | 2.882 | 5.764 | 3.746 | 1.153 | 6.052 | 12.68 | 3.17 | 3.17 | 2.305 |
| ○ *Clupeoides borneensis* | 11.24 | 1.153 | 0.576 | 1.153 | 4.035 | 5.187 | 2.017 | 7.493 | 2.882 | 18.73 | 4.035 | 2.594 | 5.764 | 3.746 | 1.153 | 5.764 | 13.83 | 3.17 | 3.17 | 2.305 |
| ○ *Sundasalanx praecox* | 12.1 | 0.865 | 0.865 | 1.153 | 4.899 | 4.899 | 1.729 | 6.052 | 3.458 | 19.88 | 3.746 | 2.594 | 5.187 | 3.746 | 1.441 | 6.052 | 11.24 | 4.899 | 3.17 | 2.017 |
| ○ *Sundasalanx sp. Chao Phraya* | 12.68 | 0.576 | 0.865 | 1.153 | 5.187 | 4.611 | 2.305 | 6.34 | 3.458 | 18.73 | 5.476 | 2.305 | 4.611 | 2.882 | 1.441 | 6.628 | 11.82 | 3.746 | 3.17 | 2.017 |
| ○ *Sundasalanx mekongensis* | 12.1 | 0.865 | 0.865 | 1.153 | 4.611 | 4.611 | 2.017 | 6.34 | 3.458 | 19.02 | 6.916 | 2.305 | 4.611 | 3.17 | 1.441 | 6.052 | 12.1 | 3.17 | 3.17 | 2.017 |
| □ *Ehirava fluviatilis* | 10.95 | 0.865 | 0.288 | 1.729 | 3.17 | 4.899 | 2.594 | 8.069 | 3.746 | 17.87 | 5.476 | 2.305 | 5.764 | 3.17 | 1.153 | 5.476 | 10.95 | 5.187 | 3.17 | 3.17 |
| □ *Gilchristella aestuaria* | 13.54 | 0.865 | 0.576 | 1.729 | 3.17 | 4.611 | 2.017 | 6.628 | 2.594 | 17.58 | 3.458 | 2.594 | 5.764 | 4.035 | 1.441 | 5.476 | 13.83 | 4.611 | 3.17 | 2.305 |
| □ *Clupeonella cultriventris* | 14.41 | 0.576 | 0.288 | 1.729 | 2.594 | 4.611 | 2.305 | 7.205 | 2.882 | 19.6 | 3.746 | 2.017 | 5.187 | 4.035 | 1.153 | 6.628 | 12.39 | 3.17 | 3.17 | 2.305 |
| ● *Clupea harengus* | 12.1 | 0.865 | 0.576 | 1.153 | 3.17 | 4.611 | 1.729 | 5.764 | 2.305 | 19.02 | 5.187 | 2.594 | 6.052 | 3.746 | 1.441 | 6.628 | 13.26 | 4.035 | 3.17 | 2.594 |
| ● *Clupea pallasii* | 11.82 | 0.865 | 0.576 | 1.153 | 3.17 | 4.611 | 1.729 | 5.764 | 2.305 | 19.02 | 5.187 | 2.594 | 6.052 | 3.746 | 1.441 | 6.628 | 13.26 | 4.323 | 3.17 | 2.594 |
| ● *Sprattus sprattus* | 12.39 | 0.865 | 0.576 | 1.153 | 3.17 | 4.611 | 2.017 | 5.187 | 2.305 | 19.02 | 5.187 | 2.305 | 5.764 | 3.746 | 1.441 | 6.916 | 12.68 | 4.899 | 3.17 | 2.594 |
| ● *Sprattus muelleri* | 12.97 | 0.865 | 0.576 | 1.153 | 3.746 | 4.899 | 2.305 | 6.052 | 2.017 | 18.16 | 5.764 | 2.305 | 5.476 | 3.746 | 1.441 | 8.069 | 10.37 | 4.323 | 3.17 | 2.594 |
| ● *Sprattus antipodum* | 12.97 | 0.865 | 0.576 | 1.153 | 3.458 | 5.187 | 2.305 | 6.052 | 2.017 | 18.44 | 5.764 | 2.305 | 5.476 | 3.746 | 1.441 | 7.781 | 10.37 | 4.323 | 3.17 | 2.594 |
| ○ *Potamalosa richmondia* | 12.68 | 0.576 | 0.288 | 1.441 | 2.594 | 4.611 | 2.305 | 6.916 | 2.017 | 19.02 | 5.187 | 2.882 | 5.476 | 4.035 | 1.441 | 4.899 | 14.99 | 3.17 | 3.17 | 2.305 |
| ● *Hyperlophus vittatus* | 12.97 | 0.576 | 0.288 | 1.441 | 2.594 | 4.899 | 2.305 | 6.052 | 2.017 | 19.6 | 4.899 | 2.882 | 5.187 | 4.035 | 1.441 | 4.323 | 14.99 | 4.035 | 3.17 | 2.305 |
| ● *Ethmidium maculatum* | 12.1 | 1.153 | 0.288 | 1.441 | 2.882 | 4.611 | 2.017 | 5.764 | 2.594 | 18.73 | 4.899 | 2.882 | 5.476 | 3.746 | 1.441 | 4.611 | 14.99 | 4.611 | 3.17 | 2.594 |
| ● *Jenkinsia lamprotaenia* | 13.54 | 0.576 | 0.865 | 1.153 | 3.17 | 4.899 | 2.017 | 5.187 | 2.305 | 19.88 | 4.611 | 1.729 | 5.476 | 3.458 | 1.729 | 8.646 | 10.66 | 4.899 | 3.17 | 2.017 |
| ● *Spratelloides delicatulus* | 13.83 | 0.288 | 0.576 | 1.729 | 2.305 | 5.476 | 2.882 | 6.34 | 2.017 | 20.46 | 4.611 | 1.729 | 5.187 | 3.17 | 1.729 | 8.069 | 9.798 | 4.323 | 3.17 | 2.305 |
| ● *Spratelloides gracilis* | 12.39 | 0.576 | 0.288 | 2.017 | 3.458 | 5.187 | 2.017 | 5.187 | 2.017 | 19.31 | 6.34 | 1.441 | 5.476 | 2.882 | 1.729 | 7.493 | 12.1 | 4.899 | 3.17 | 2.017 |
| ● *Etrumeus micropus* | 12.68 | 0.576 | 0 | 1.441 | 4.611 | 4.899 | 2.017 | 6.916 | 2.305 | 18.44 | 4.611 | 2.594 | 5.476 | 4.035 | 1.441 | 7.781 | 9.222 | 5.187 | 3.17 | 2.594 |
| ○ *Ilisha africana* | 10.09 | 1.153 | 0.288 | 1.153 | 2.305 | 4.899 | 2.882 | 8.357 | 2.305 | 18.73 | 4.611 | 3.17 | 5.476 | 4.035 | 1.153 | 5.476 | 17.58 | 1.153 | 3.17 | 2.017 |
| ○ *Pellona flavipinnis* | 9.222 | 0.288 | 0.576 | 1.441 | 2.594 | 4.611 | 2.594 | 7.781 | 2.305 | 17.87 | 5.764 | 2.882 | 6.052 | 4.035 | 1.153 | 6.052 | 17.87 | 1.441 | 3.17 | 2.305 |
| ○ *Ilisha elongata* | 10.37 | 0.288 | 0.288 | 1.729 | 2.594 | 4.899 | 2.594 | 7.781 | 2.594 | 18.73 | 4.899 | 2.882 | 5.476 | 3.746 | 1.153 | 7.205 | 15.85 | 1.441 | 3.17 | 2.305 |
| □ *Pellona ditchela* | 10.95 | 0.288 | 0 | 1.729 | 2.594 | 4.611 | 2.882 | 8.646 | 2.594 | 18.16 | 4.611 | 3.458 | 5.476 | 3.746 | 1.153 | 5.764 | 16.43 | 1.729 | 3.17 | 2.017 |
| □ *Anchoviella sp. LBP 2297* | 12.39 | 0.865 | 0.288 | 1.153 | 2.594 | 4.611 | 2.305 | 8.646 | 2.594 | 20.17 | 4.035 | 2.882 | 5.187 | 4.035 | 1.153 | 4.611 | 15.56 | 1.729 | 3.17 | 2.017 |
| □ *Lycengraulis grossidens* | 12.1 | 0.865 | 0.288 | 1.153 | 2.594 | 4.611 | 2.594 | 8.069 | 2.594 | 20.46 | 4.035 | 2.594 | 5.187 | 4.035 | 1.153 | 5.476 | 14.99 | 2.305 | 3.17 | 1.729 |
| ○ *Amazonsprattus scintilla* | 11.82 | 0.865 | 0.288 | 1.153 | 2.882 | 4.611 | 2.594 | 6.916 | 2.594 | 20.17 | 4.899 | 2.305 | 5.187 | 4.035 | 1.153 | 6.34 | 14.12 | 3.17 | 3.17 | 1.729 |
| ● *Engraulis encrasicolus* | 12.1 | 0.865 | 0.288 | 1.153 | 2.594 | 4.899 | 2.305 | 8.069 | 2.594 | 19.88 | 4.899 | 2.594 | 4.899 | 4.035 | 1.153 | 6.34 | 13.83 | 2.305 | 3.17 | 2.017 |
| ● *Engraulis japonicus* | 12.39 | 0.865 | 0.288 | 1.153 | 2.594 | 4.611 | 2.305 | 7.781 | 2.594 | 20.17 | 4.899 | 2.594 | 4.899 | 4.035 | 1.153 | 6.34 | 13.83 | 2.305 | 3.17 | 2.017 |
| ● *Stolephorus chinensis* | 12.1 | 0.865 | 0.288 | 1.441 | 2.882 | 4.611 | 2.017 | 7.493 | 2.882 | 19.88 | 4.323 | 2.882 | 5.187 | 3.746 | 1.153 | 6.052 | 13.83 | 3.458 | 3.17 | 1.729 |
| ● *Stolephorus waitei* | 12.1 | 0.865 | 0.288 | 1.441 | 2.882 | 4.611 | 2.017 | 7.781 | 2.882 | 19.88 | 4.323 | 2.882 | 5.187 | 3.746 | 1.153 | 6.052 | 13.83 | 3.17 | 3.17 | 1.729 |
| ○ *Lycothrissa crocodilus* | 9.798 | 0.576 | 0.288 | 1.153 | 2.594 | 4.611 | 2.305 | 9.798 | 2.882 | 19.6 | 4.611 | 3.17 | 5.187 | 4.035 | 1.153 | 6.052 | 15.85 | 1.153 | 3.17 | 2.017 |
| □ *Setipinna melanochir* | 10.4 | 0.578 | 0.289 | 1.156 | 2.312 | 4.624 | 2.601 | 9.249 | 2.601 | 19.36 | 4.335 | 3.468 | 5.78 | 4.046 | 1.156 | 5.78 | 15.32 | 1.734 | 3.179 | 2.023 |
| ● *Coilia reynaldi* | 11.82 | 0.576 | 0.288 | 1.153 | 2.017 | 4.611 | 2.882 | 7.205 | 2.594 | 20.46 | 5.764 | 3.746 | 5.476 | 4.035 | 1.153 | 4.899 | 14.12 | 2.305 | 3.17 | 1.729 |
| ● *Thryssa baelama* | 12.39 | 0.865 | 0.288 | 1.153 | 3.17 | 4.611 | 2.017 | 7.493 | 2.305 | 19.6 | 4.323 | 3.458 | 5.476 | 4.035 | 1.441 | 6.916 | 12.39 | 2.882 | 3.17 | 2.017 |
| ○ *Coilia lindmani* | 11.24 | 0.865 | 0.288 | 1.153 | 2.594 | 4.611 | 2.305 | 8.934 | 2.594 | 19.02 | 4.611 | 3.746 | 5.187 | 4.323 | 1.153 | 5.764 | 14.41 | 2.017 | 3.17 | 2.017 |
| □ *Coilia ectenes* | 11.53 | 1.153 | 0.288 | 1.153 | 2.594 | 4.611 | 2.305 | 8.934 | 2.594 | 19.02 | 5.187 | 3.746 | 5.187 | 4.323 | 1.153 | 5.187 | 14.12 | 2.017 | 3.17 | 1.729 |
| ● *Coilia nasus* | 11.53 | 1.153 | 0.288 | 1.153 | 2.594 | 4.611 | 2.305 | 8.934 | 2.594 | 19.02 | 5.187 | 3.746 | 5.187 | 4.323 | 1.153 | 5.187 | 14.12 | 2.017 | 3.17 | 1.729 |
| ● *Denticeps clupeoides* | 9.222 | 0.288 | 0.288 | 1.441 | 2.882 | 4.611 | 2.305 | 7.781 | 2.594 | 20.75 | 5.187 | 3.17 | 5.187 | 4.323 | 1.153 | 7.205 | 14.41 | 2.017 | 3.17 | 2.017 |

nd2

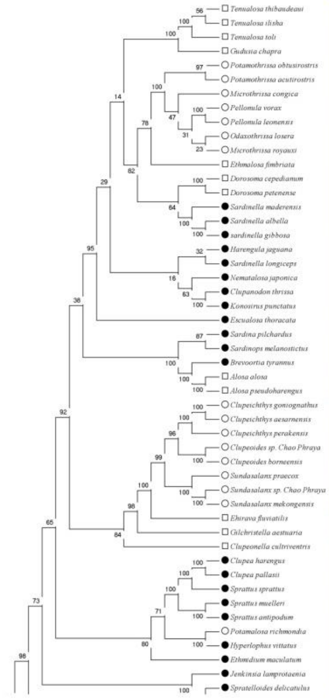


| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 25 | 3.448 | 1.724 | 6.897 | 2.586 | 1.724 | 7.759 | 6.034 | 3.448 | 4.31 | 1.724 |
| □ *Tenualosa ilisha* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 24.14 | 3.448 | 3.448 | 6.897 | 2.586 | 1.724 | 6.897 | 5.172 | 4.31 | 4.31 | 1.724 |
| □ *Tenualosa toli* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 8.621 | 0.862 | 24.14 | 3.448 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 4.31 | 5.172 | 4.31 | 1.724 |
| □ *Gudusia chapra* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 9.483 | 0.862 | 22.41 | 2.586 | 2.586 | 7.759 | 2.586 | 1.724 | 7.759 | 6.897 | 2.586 | 4.31 | 1.724 |
| ○ *Potamothrissa obtusirostris* | 6.034 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 24.14 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 6.897 | 5.172 | 4.31 | 1.724 |
| ○ *Potamothrissa acutirostris* | 6.034 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 8.621 | 0.862 | 23.28 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 7.759 | 6.034 | 5.172 | 4.31 | 1.724 |
| ○ *Microthrissa congica* | 6.034 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 23.28 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 7.759 | 6.897 | 5.172 | 4.31 | 1.724 |
| ○ *Pellonula vorax* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 8.621 | 0.862 | 23.28 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 6.034 | 5.172 | 4.31 | 1.724 |
| ○ *Pellonula leonensis* | 6.034 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 8.621 | 0.862 | 23.28 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 7.759 | 6.034 | 5.172 | 4.31 | 1.724 |
| ○ *Odaxothrissa losera* | 5.172 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 23.28 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 8.621 | 6.034 | 6.034 | 4.31 | 1.724 |
| ○ *Microthrissa royauxi* | 6.034 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 9.483 | 0.862 | 23.28 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 7.759 | 6.034 | 4.31 | 4.31 | 1.724 |
| □ *Ethmalosa fimbriata* | 6.034 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 24.14 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 7.759 | 5.172 | 6.034 | 4.31 | 1.724 |
| □ *Dorosoma cepedianum* | 6.034 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 6.034 | 0.862 | 24.14 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 7.759 | 6.034 | 6.897 | 4.31 | 1.724 |
| □ *Dorosoma petenense* | 9.483 | 0.862 | 2.586 | 5.172 | 6.897 | 3.448 | 0 | 9.483 | 0.862 | 21.55 | 3.448 | 0.862 | 6.897 | 2.586 | 2.586 | 10.34 | 3.448 | 3.448 | 4.31 | 1.724 |
| ● *Sardinella maderensis* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 24.14 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 6.034 | 5.172 | 4.31 | 1.724 |
| ● *Sardinella albella* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 24.14 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 5.172 | 6.034 | 4.31 | 1.724 |
| ● *sardinella gibbosa* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 24.14 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 5.172 | 6.034 | 4.31 | 1.724 |
| ● *Harengula jaguana* | 6.034 | 0.862 | 3.448 | 5.172 | 6.897 | 4.31 | 0 | 7.759 | 0.862 | 25.86 | 2.586 | 2.586 | 7.759 | 2.586 | 1.724 | 6.897 | 4.31 | 5.172 | 3.448 | 1.724 |
| ● *Sardinella longiceps* | 6.034 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 6.897 | 0.862 | 25 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 7.759 | 5.172 | 6.034 | 4.31 | 1.724 |
| ● *Nematalosa japonica* | 7.759 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 8.621 | 0.862 | 23.28 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 6.034 | 4.31 | 4.31 | 1.724 |
| ● *Clupanodon thrissa* | 7.759 | 0.862 | 2.586 | 5.172 | 6.034 | 4.31 | 0 | 7.759 | 0.862 | 24.14 | 4.31 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 5.172 | 4.31 | 4.31 | 1.724 |

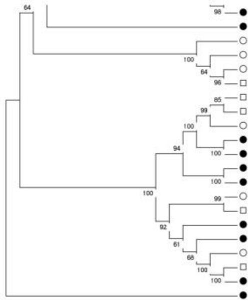


| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ● *Konosirus punctatus* | 6.034 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 24.14 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 6.897 | 5.172 | 4.31 | 1.724 |
| ● *Escualosa thoracata* | 6.897 | 0.862 | 2.586 | 6.034 | 7.759 | 5.172 | 0 | 8.621 | 0.862 | 23.28 | 3.448 | 0.862 | 7.759 | 2.586 | 1.724 | 6.897 | 5.172 | 3.448 | 4.31 | 1.724 |
| ● *Sardina pilchardus* | 9.483 | 0.862 | 2.586 | 5.172 | 6.034 | 4.31 | 0 | 7.759 | 0.862 | 21.55 | 3.448 | 0.862 | 6.897 | 2.586 | 2.586 | 5.172 | 5.172 | 7.759 | 4.31 | 2.586 |
| ● *Sardinops melanostictus* | 7.759 | 0.862 | 2.586 | 6.034 | 6.034 | 5.172 | 0 | 10.34 | 0.862 | 21.55 | 4.31 | 0.862 | 6.897 | 2.586 | 1.724 | 6.897 | 5.172 | 4.31 | 4.31 | 1.724 |
| ● *Brevoortia tyrannus* | 6.897 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 9.483 | 0.862 | 23.28 | 3.448 | 1.724 | 6.897 | 2.586 | 1.724 | 6.897 | 5.172 | 4.31 | 4.31 | 1.724 |
| □ *Alosa alosa* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 8.621 | 0.862 | 23.28 | 3.448 | 2.586 | 6.897 | 2.586 | 1.724 | 6.897 | 4.31 | 6.034 | 4.31 | 1.724 |
| □ *Alosa pseudoharengus* | 6.897 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 8.621 | 0.862 | 23.28 | 3.448 | 1.724 | 6.897 | 2.586 | 1.724 | 6.897 | 5.172 | 5.172 | 4.31 | 1.724 |
| ○ *Clupeichthys goniognathus* | 6.897 | 1.724 | 3.448 | 5.172 | 9.483 | 5.172 | 0.862 | 6.034 | 0.862 | 24.14 | 2.586 | 1.724 | 7.759 | 2.586 | 1.724 | 6.897 | 6.034 | 1.724 | 3.448 | 1.724 |
| ○ *Clupeichthys aesarnensis* | 6.034 | 1.724 | 2.586 | 5.172 | 6.897 | 5.172 | 0 | 7.759 | 0.862 | 25 | 3.448 | 2.586 | 7.759 | 3.448 | 1.724 | 6.897 | 6.034 | 1.724 | 3.448 | 1.724 |
| ○ *Clupeichthys perakensis* | 7.759 | 1.724 | 3.448 | 5.172 | 6.897 | 4.31 | 0.862 | 8.621 | 0.862 | 24.14 | 2.586 | 1.724 | 7.759 | 2.586 | 1.724 | 7.759 | 4.31 | 2.586 | 3.448 | 1.724 |
| ○ *Clupeoides sp. Chao Phraya* | 6.034 | 1.724 | 2.586 | 5.172 | 6.897 | 4.31 | 0.862 | 6.034 | 0.862 | 24.14 | 3.448 | 2.586 | 7.759 | 2.586 | 1.724 | 6.897 | 6.034 | 5.172 | 3.448 | 1.724 |
| ○ *Clupeoides borneensis* | 6.034 | 1.724 | 3.448 | 5.172 | 6.897 | 4.31 | 0.862 | 7.759 | 0.862 | 24.14 | 3.448 | 2.586 | 7.759 | 2.586 | 1.724 | 6.034 | 5.172 | 4.31 | 3.448 | 1.724 |
| ○ *Sundasalanx praecox* | 6.034 | 1.724 | 2.586 | 6.034 | 6.897 | 5.172 | 0 | 6.034 | 0.862 | 24.14 | 1.724 | 1.724 | 6.897 | 3.448 | 1.724 | 6.897 | 6.034 | 6.897 | 3.448 | 1.724 |
| ○ *Sundasalanx sp. Chao Phraya* | 6.034 | 1.724 | 2.586 | 5.172 | 6.897 | 5.172 | 0 | 6.897 | 0.862 | 25 | 2.586 | 2.586 | 6.897 | 3.448 | 1.724 | 6.897 | 5.172 | 5.172 | 3.448 | 1.724 |
| ○ *Sundasalanx mekongensis* | 6.034 | 1.724 | 2.586 | 5.172 | 6.897 | 5.172 | 0 | 6.034 | 0.862 | 25 | 3.448 | 3.448 | 6.897 | 3.448 | 1.724 | 6.034 | 5.172 | 5.172 | 3.448 | 1.724 |
| □ *Ehirava fluviatilis* | 8.621 | 1.724 | 2.586 | 6.034 | 6.034 | 4.31 | 0 | 6.034 | 0.862 | 24.14 | 3.448 | 3.448 | 6.897 | 2.586 | 1.724 | 8.621 | 4.31 | 3.448 | 3.448 | 1.724 |
| □ *Gilchristella aestuaria* | 6.897 | 0.862 | 3.448 | 5.172 | 6.897 | 5.172 | 0 | 7.759 | 0.862 | 22.41 | 3.448 | 1.724 | 7.759 | 2.586 | 1.724 | 6.034 | 6.034 | 5.172 | 4.31 | 1.724 |
| □ *Clupeonella cultriventris* | 7.759 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 24.14 | 3.448 | 1.724 | 6.897 | 2.586 | 1.724 | 6.897 | 3.448 | 6.034 | 4.31 | 1.724 |
| ● *Clupea harengus* | 8.621 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 6.897 | 0.862 | 21.55 | 4.31 | 0.862 | 6.897 | 2.586 | 2.586 | 5.172 | 3.448 | 8.621 | 4.31 | 2.586 |
| ● *Clupea pallasii* | 8.621 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 6.897 | 0.862 | 21.55 | 4.31 | 0.862 | 6.897 | 2.586 | 2.586 | 5.172 | 3.448 | 8.621 | 4.31 | 2.586 |
| ● *Sprattus sprattus* | 8.621 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 5.172 | 0.862 | 22.41 | 4.31 | 0.862 | 6.897 | 2.586 | 2.586 | 5.172 | 5.172 | 7.759 | 4.31 | 2.586 |
| ● *Sprattus muelleri* | 8.621 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 21.55 | 4.31 | 0 | 6.897 | 2.586 | 2.586 | 5.172 | 5.172 | 6.897 | 4.31 | 2.586 |
| ● *Sprattus antipodum* | 9.483 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 21.55 | 4.31 | 0 | 6.897 | 2.586 | 2.586 | 5.172 | 4.31 | 6.897 | 4.31 | 2.586 |
| ○ *Potamalosa richmondia* | 6.897 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 6.897 | 0.862 | 25.86 | 3.448 | 1.724 | 6.897 | 2.586 | 1.724 | 6.897 | 4.31 | 5.172 | 4.31 | 1.724 |
| ● *Hyperlophus vittatus* | 7.759 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 25.86 | 3.448 | 1.724 | 6.897 | 2.586 | 1.724 | 6.897 | 3.448 | 4.31 | 4.31 | 1.724 |
| ● *Ethmidium maculatum* | 7.759 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 9.483 | 0.862 | 23.28 | 3.448 | 0.862 | 7.759 | 2.586 | 1.724 | 6.034 | 3.448 | 5.172 | 4.31 | 2.586 |
| ● *Jenkinsia lamprotaenia* | 9.483 | 0.862 | 3.448 | 5.172 | 6.034 | 5.172 | 0 | 7.759 | 0.862 | 23.28 | 3.448 | 0 | 6.897 | 1.724 | 1.724 | 6.897 | 6.034 | 5.172 | 4.31 | 1.724 |
| ● *Spratelloides delicatulus* | 9.483 | 0.862 | 2.586 | 5.172 | 6.897 | 4.31 | 0 | 6.034 | 0.862 | 25 | 3.448 | 0.862 | 7.759 | 1.724 | 1.724 | 8.621 | 2.586 | 6.034 | 4.31 | 1.724 |
| ● *Spratelloides gracilis* | 8.621 | 0.862 | 3.448 | 5.172 | 6.034 | 4.31 | 0 | 7.759 | 0.862 | 24.14 | 3.448 | 0.862 | 7.759 | 1.724 | 1.724 | 6.897 | 4.31 | 6.034 | 4.31 | 1.724 |
| ● *Etrumeus micropus* | 7.759 | 0.862 | 2.586 | 5.172 | 6.897 | 4.31 | 0 | 6.034 | 0.862 | 24.14 | 2.586 | 1.724 | 7.759 | 2.586 | 1.724 | 6.034 | 6.034 | 7.759 | 3.448 | 1.724 |
| ○ *Ilisha africana* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 4.31 | 0.862 | 7.759 | 0.862 | 24.14 | 3.448 | 1.724 | 7.759 | 2.586 | 1.724 | 4.31 | 10.34 | 2.586 | 4.31 | 1.724 |
| ○ *Pellona flavipinnis* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0.862 | 8.621 | 0.862 | 21.55 | 5.172 | 1.724 | 7.759 | 2.586 | 1.724 | 4.31 | 9.483 | 2.586 | 4.31 | 1.724 |
| ○ *Ilisha elongata* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 5.172 | 0 | 8.621 | 0.862 | 23.28 | 2.586 | 1.724 | 7.759 | 2.586 | 1.724 | 4.31 | 9.483 | 3.448 | 4.31 | 2.586 |
| □ *Pellona ditchela* | 6.897 | 0.862 | 2.586 | 5.172 | 6.034 | 4.31 | 0 | 8.621 | 0.862 | 23.28 | 4.31 | 1.724 | 7.759 | 2.586 | 1.724 | 5.172 | 8.621 | 2.586 | 4.31 | 2.586 |
| □ *Anchoviella sp. LBP 2297* | 8.621 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 6.897 | 1.724 | 24.14 | 3.448 | 2.586 | 6.897 | 2.586 | 1.724 | 4.31 | 5.172 | 4.31 | 4.31 | 1.724 |
| □ *Lycengraulis grossidens* | 7.759 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 6.897 | 1.724 | 24.14 | 3.448 | 2.586 | 6.897 | 2.586 | 1.724 | 4.31 | 6.034 | 4.31 | 4.31 | 1.724 |
| ○ *Amazonsprattus scintilla* | 7.759 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 6.897 | 1.724 | 23.28 | 4.31 | 2.586 | 6.897 | 2.586 | 1.724 | 4.31 | 5.172 | 5.172 | 4.31 | 1.724 |
| ● *Engraulis encrasicolus* | 7.759 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 6.034 | 1.724 | 24.14 | 3.448 | 1.724 | 6.897 | 2.586 | 1.724 | 4.31 | 6.897 | 5.172 | 4.31 | 1.724 |
| ● *Engraulis japonicus* | 8.621 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 6.034 | 1.724 | 24.14 | 3.448 | 1.724 | 6.897 | 2.586 | 1.724 | 4.31 | 6.034 | 5.172 | 4.31 | 1.724 |
| ● *Stolephorus chinensis* | 6.897 | 0.862 | 3.448 | 5.172 | 6.034 | 6.034 | 0 | 6.897 | 1.724 | 24.14 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 4.31 | 6.897 | 5.172 | 4.31 | 1.724 |
| ● *Stolephorus waitei* | 7.759 | 0.862 | 3.448 | 5.172 | 6.034 | 6.034 | 0 | 6.897 | 1.724 | 24.14 | 2.586 | 2.586 | 6.897 | 2.586 | 1.724 | 4.31 | 6.034 | 5.172 | 4.31 | 1.724 |
| ○ *Lycothrissa crocodilus* | 6.034 | 0.862 | 2.586 | 5.172 | 6.897 | 5.172 | 0 | 8.621 | 0.862 | 23.28 | 2.586 | 4.31 | 6.897 | 2.586 | 1.724 | 5.172 | 7.759 | 3.448 | 4.31 | 1.724 |
| □ *Setipinna melanochir* | 5.172 | 0.862 | 3.448 | 5.172 | 6.034 | 6.034 | 0 | 6.034 | 0.862 | 23.28 | 2.586 | 3.448 | 6.897 | 2.586 | 1.724 | 6.034 | 7.759 | 6.034 | 4.31 | 1.724 |
| ● *Coilia reynaldi* | 6.034 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 6.034 | 1.724 | 23.28 | 6.034 | 1.724 | 6.897 | 2.586 | 1.724 | 4.31 | 6.897 | 4.31 | 4.31 | 2.586 |
| ● *Thryssa baelama* | 6.897 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 5.172 | 1.724 | 25 | 3.448 | 2.586 | 6.897 | 2.586 | 1.724 | 4.31 | 5.172 | 6.897 | 4.31 | 1.724 |
| ○ *Coilia lindmani* | 6.897 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 6.897 | 1.724 | 23.28 | 4.31 | 2.586 | 6.897 | 3.448 | 1.724 | 5.172 | 5.172 | 4.31 | 4.31 | 1.724 |
| □ *Coilia ectenes* | 6.034 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 6.897 | 1.724 | 23.28 | 4.31 | 2.586 | 6.897 | 3.448 | 1.724 | 5.172 | 6.034 | 4.31 | 4.31 | 1.724 |
| ● *Coilia nasus* | 6.034 | 0.862 | 2.586 | 6.034 | 6.034 | 6.034 | 0 | 6.897 | 1.724 | 23.28 | 4.31 | 2.586 | 6.897 | 3.448 | 1.724 | 5.172 | 6.034 | 4.31 | 4.31 | 1.724 |
| ● *Denticeps clupeoides* | 6.034 | 0.862 | 3.448 | 4.31 | 6.034 | 4.31 | 0 | 11.21 | 0.862 | 21.55 | 1.724 | 3.448 | 6.897 | 2.586 | 1.724 | 6.897 | 8.621 | 3.448 | 4.31 | 1.724 |

nd3

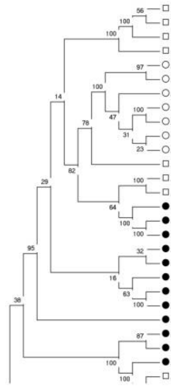


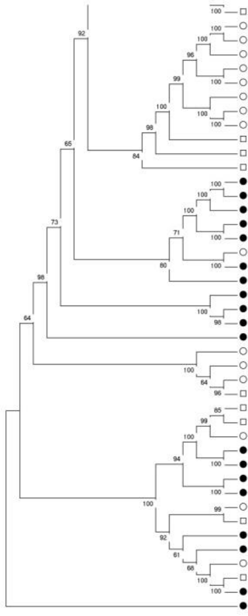
| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 8.261 | 0.87 | 0.87 | 2.391 | 3.913 | 6.304 | 2.391 | 8.696 | 2.174 | 19.13 | 4.565 | 2.174 | 6.304 | 2.174 | 2.609 | 7.174 | 8.478 | 4.348 | 4.348 | 2.826 |
| □ *Tenualosa ilisha* | 8.913 | 0.87 | 0.87 | 2.391 | 3.913 | 6.304 | 2.391 | 8.043 | 2.174 | 19.57 | 4.783 | 1.957 | 6.522 | 2.174 | 2.609 | 7.391 | 7.609 | 4.348 | 4.348 | 2.826 |
| □ *Tenualosa toli* | 8.261 | 0.87 | 0.87 | 2.391 | 3.913 | 6.304 | 2.391 | 8.261 | 2.174 | 19.35 | 5 | 2.391 | 6.304 | 2.174 | 2.609 | 6.957 | 8.696 | 3.696 | 4.348 | 3.043 |
| □ *Gudusia chapra* | 8.043 | 0.87 | 0.87 | 2.391 | 3.913 | 6.087 | 2.391 | 9.13 | 2.174 | 19.78 | 5.435 | 2.826 | 6.087 | 2.174 | 2.609 | 6.739 | 8.261 | 3.043 | 4.348 | 2.826 |
| ○ *Potamothrissa obtusirostris* | 8.913 | 0.87 | 0.87 | 2.391 | 3.696 | 6.522 | 2.391 | 7.826 | 2.174 | 18.7 | 5.87 | 2.391 | 5.87 | 2.174 | 2.609 | 5.435 | 9.783 | 4.13 | 4.348 | 3.043 |
| ○ *Potamothrissa acutirostris* | 8.696 | 0.87 | 0.87 | 2.391 | 3.696 | 6.522 | 2.609 | 8.478 | 2.174 | 18.7 | 5.87 | 2.391 | 6.087 | 2.174 | 2.391 | 5.217 | 9.565 | 3.913 | 4.348 | 3.043 |
| ○ *Microthrissa congica* | 8.696 | 0.87 | 0.87 | 2.391 | 3.696 | 6.739 | 2.391 | 8.696 | 2.174 | 18.7 | 5.435 | 2.391 | 5.87 | 2.174 | 2.609 | 5.435 | 9.565 | 3.913 | 4.348 | 3.043 |
| ○ *Pellonula vorax* | 8.696 | 0.87 | 0.87 | 2.391 | 3.696 | 6.522 | 2.391 | 8.261 | 2.174 | 18.91 | 5.217 | 2.391 | 5.87 | 2.174 | 2.391 | 5.652 | 9.565 | 4.565 | 4.348 | 3.043 |
| ○ *Pellonula leonensis* | 8.913 | 0.87 | 0.87 | 2.391 | 3.696 | 6.522 | 2.391 | 8.043 | 2.174 | 18.91 | 5.217 | 2.391 | 5.87 | 2.174 | 2.391 | 5.652 | 9.348 | 4.783 | 4.348 | 3.043 |
| ○ *Odaxothrissa losera* | 8.261 | 0.87 | 0.87 | 2.391 | 3.696 | 6.304 | 2.391 | 8.478 | 2.174 | 18.48 | 6.087 | 2.391 | 5.87 | 2.174 | 2.391 | 5.652 | 10.22 | 3.913 | 4.348 | 3.043 |
| ○ *Microthrissa royauxi* | 8.913 | 0.87 | 0.87 | 2.391 | 3.696 | 6.522 | 2.391 | 8.478 | 2.174 | 18.48 | 5.652 | 2.391 | 5.87 | 2.174 | 2.609 | 5.435 | 9.565 | 4.13 | 4.348 | 3.043 |
| □ *Ethmalosa fimbriata* | 8.261 | 0.87 | 0.87 | 2.391 | 3.696 | 6.522 | 2.391 | 8.478 | 2.174 | 18.7 | 5.652 | 2.174 | 5.87 | 2.174 | 2.609 | 5.217 | 10.22 | 4.348 | 4.348 | 3.043 |
| □ *Dorosoma cepedianum* | 8.478 | 0.87 | 0.87 | 2.609 | 3.696 | 6.522 | 2.391 | 8.261 | 2.174 | 18.48 | 5.652 | 2.391 | 5.87 | 2.174 | 2.609 | 5 | 10.22 | 4.348 | 4.348 | 3.043 |
| □ *Dorosoma petenense* | 8.261 | 0.87 | 0.87 | 2.609 | 3.913 | 6.304 | 2.391 | 8.478 | 2.174 | 18.48 | 5.435 | 2.391 | 5.87 | 2.174 | 2.609 | 5.435 | 10.22 | 4.13 | 4.565 | 2.826 |
| ● *Sardinella maderensis* | 8.261 | 0.87 | 0.87 | 2.391 | 3.696 | 6.522 | 2.391 | 8.043 | 2.174 | 18.91 | 5 | 2.391 | 5.87 | 2.174 | 2.609 | 5.652 | 9.565 | 5.217 | 4.348 | 3.043 |
| ● *Sardinella albella* | 8.261 | 0.87 | 0.87 | 2.391 | 3.913 | 6.522 | 2.391 | 8.261 | 2.174 | 18.7 | 5 | 2.391 | 5.87 | 2.174 | 2.609 | 5.435 | 10 | 4.783 | 4.348 | 3.043 |
| ● *sardinella gibbosa* | 8.261 | 0.87 | 0.87 | 2.391 | 3.913 | 6.522 | 2.609 | 8.261 | 2.174 | 18.7 | 5.435 | 2.174 | 5.87 | 1.957 | 2.609 | 5.435 | 10 | 4.565 | 4.348 | 3.043 |
| ● *Harengula jaguana* | 9.13 | 0.87 | 0.87 | 2.391 | 3.913 | 6.739 | 2.391 | 7.174 | 2.174 | 19.13 | 5.217 | 2.391 | 6.087 | 2.174 | 2.391 | 5.87 | 8.696 | 5 | 4.348 | 3.043 |
| ● *Sardinella longiceps* | 8.696 | 0.87 | 0.87 | 2.391 | 3.913 | 6.522 | 2.391 | 7.391 | 2.174 | 19.35 | 5 | 2.391 | 5.87 | 2.174 | 2.609 | 5.652 | 9.348 | 5 | 4.348 | 3.043 |
| ● *Nematalosa japonica* | 9.13 | 0.87 | 0.87 | 2.391 | 3.696 | 6.304 | 2.391 | 8.261 | 2.174 | 19.13 | 5.435 | 2.174 | 6.087 | 2.174 | 2.609 | 5.652 | 9.13 | 4.13 | 4.348 | 3.043 |
| ● *Clupanodon thrissa* | 9.13 | 1.087 | 0.87 | 2.391 | 3.696 | 6.304 | 2.391 | 8.261 | 2.174 | 19.13 | 5.217 | 2.391 | 5.87 | 2.174 | 2.609 | 5.652 | 8.913 | 4.348 | 4.348 | 3.043 |
| ● *Konosirus punctatus* | 8.913 | 0.87 | 0.87 | 2.391 | 3.696 | 6.522 | 2.391 | 8.913 | 2.391 | 19.13 | 5 | 2.391 | 5.87 | 2.174 | 2.609 | 5.435 | 9.565 | 3.478 | 4.348 | 3.043 |
| ● *Escualosa thoracata* | 8.696 | 0.87 | 0.87 | 2.391 | 3.696 | 6.304 | 2.391 | 8.043 | 2.174 | 18.91 | 5 | 2.391 | 6.304 | 2.174 | 2.609 | 5.87 | 9.13 | 4.783 | 4.348 | 3.043 |
| ● *Sardina pilchardus* | 8.696 | 1.087 | 0.87 | 2.391 | 3.913 | 6.957 | 2.391 | 7.826 | 2.174 | 17.61 | 5.435 | 2.174 | 5.87 | 2.391 | 2.391 | 6.087 | 8.261 | 6.087 | 4.348 | 3.043 |
| ● *Sardinops melanostictus* | 8.261 | 1.087 | 0.87 | 2.391 | 3.696 | 6.304 | 2.391 | 7.609 | 2.174 | 18.7 | 5.652 | 2.391 | 6.087 | 2.391 | 2.391 | 6.087 | 9.13 | 5 | 4.348 | 3.043 |
| ● *Brevoortia tyrannus* | 8.261 | 1.087 | 0.87 | 2.391 | 3.696 | 6.304 | 2.391 | 8.043 | 2.174 | 18.7 | 5.435 | 2.826 | 5.87 | 2.391 | 2.391 | 6.087 | 9.13 | 4.565 | 4.348 | 3.043 |
| □ *Alosa alosa* | 8.478 | 1.087 | 0.87 | 2.391 | 3.696 | 6.304 | 2.174 | 7.826 | 2.174 | 18.48 | 5.435 | 2.826 | 5.87 | 2.391 | 2.391 | 5.87 | 9.13 | 5 | 4.348 | 3.261 |
| □ *Alosa pseudoharengus* | 8.261 | 1.087 | 0.87 | 2.391 | 3.696 | 6.304 | 2.174 | 7.826 | 2.174 | 18.48 | 5.435 | 2.609 | 5.87 | 2.391 | 2.391 | 6.087 | 9.348 | 5 | 4.348 | 3.261 |
| ○ *Clupeichthys goniognathus* | 9.348 | 0.87 | 0.87 | 2.609 | 4.348 | 6.304 | 2.609 | 7.609 | 1.957 | 19.57 | 4.13 | 2.826 | 5.435 | 1.957 | 2.609 | 6.304 | 8.696 | 4.783 | 4.348 | 2.826 |
| ○ *Clupeichthys aesarnensis* | 8.913 | 0.87 | 1.087 | 2.391 | 4.13 | 6.304 | 2.609 | 7.826 | 1.739 | 19.13 | 4.348 | 2.826 | 5.87 | 1.957 | 2.609 | 6.957 | 8.696 | 4.348 | 4.348 | 3.043 |
| ○ *Clupeichthys perakensis* | 8.478 | 0.87 | 0.87 | 2.391 | 3.696 | 6.304 | 2.826 | 7.609 | 1.957 | 19.13 | 5 | 2.609 | 5.435 | 1.957 | 2.609 | 7.609 | 9.13 | 4.348 | 4.348 | 2.826 |
| ○ *Clupeoides sp. Chao Phraya* | 8.043 | 0.87 | 0.87 | 2.391 | 3.913 | 6.304 | 2.174 | 9.13 | 2.174 | 18.91 | 5.435 | 3.043 | 5.652 | 2.391 | 2.609 | 6.087 | 9.565 | 3.043 | 4.348 | 3.043 |
| ○ *Clupeoides borneensis* | 7.826 | 0.87 | 0.87 | 2.391 | 3.478 | 6.304 | 2.391 | 9.13 | 2.174 | 19.35 | 5 | 2.826 | 5.217 | 2.391 | 2.609 | 6.304 | 9.783 | 3.696 | 4.348 | 3.043 |
| ○ *Sundasalanx praecox* | 9.13 | 0.87 | 0.87 | 2.391 | 3.696 | 6.304 | 2.391 | 8.043 | 1.957 | 19.35 | 4.783 | 2.609 | 5.652 | 2.174 | 2.609 | 6.087 | 9.565 | 3.913 | 4.348 | 3.261 |
| ○ *Sundasalanx sp. Chao Phraya* | 8.913 | 0.87 | 0.87 | 2.609 | 4.348 | 6.304 | 2.609 | 8.913 | 2.174 | 18.91 | 4.348 | 2.391 | 5.652 | 1.957 | 2.609 | 6.087 | 9.565 | 3.261 | 4.348 | 3.261 |
| ○ *Sundasalanx mekongensis* | 8.261 | 0.87 | 0.652 | 2.609 | 4.565 | 6.087 | 2.826 | 9.565 | 2.174 | 18.26 | 4.348 | 2.826 | 5.652 | 1.957 | 2.609 | 6.304 | 10 | 3.043 | 4.348 | 3.043 |
| □ *Ehirava fluviatilis* | 9.348 | 0.87 | 0.87 | 2.391 | 3.696 | 6.304 | 2.391 | 9.13 | 1.957 | 19.13 | 4.783 | 2.609 | 5.435 | 2.391 | 2.609 | 6.304 | 8.913 | 3.478 | 4.348 | 3.043 |
| □ *Gilchristella aestuaria* | 8.913 | 0.87 | 0.87 | 2.391 | 4.13 | 6.304 | 2.391 | 8.913 | 2.174 | 18.48 | 5 | 2.391 | 5.87 | 2.174 | 2.609 | 6.522 | 8.696 | 4.13 | 4.348 | 2.826 |
| □ *Clupeonella cultriventris* | 10 | 0.87 | 0.87 | 2.391 | 3.913 | 6.522 | 2.391 | 7.609 | 2.174 | 18.7 | 5 | 2.391 | 5.652 | 2.174 | 2.609 | 6.304 | 8.043 | 5 | 4.348 | 3.043 |
| ● *Clupea harengus* | 9.348 | 1.087 | 0.87 | 2.609 | 3.696 | 6.087 | 2.391 | 8.261 | 2.174 | 18.91 | 5.217 | 2.826 | 6.087 | 1.957 | 2.609 | 5.652 | 8.913 | 4.13 | 4.348 | 2.826 |
| ● *Clupea pallasii* | 9.348 | 1.087 | 0.87 | 2.609 | 3.696 | 6.087 | 2.391 | 8.043 | 2.174 | 19.13 | 5.217 | 2.826 | 6.087 | 1.957 | 2.609 | 5.652 | 8.913 | 4.13 | 4.348 | 2.826 |
| ● *Sprattus sprattus* | 9.565 | 1.087 | 0.87 | 2.609 | 3.696 | 6.304 | 2.391 | 7.609 | 2.174 | 18.91 | 5.652 | 2.826 | 6.087 | 1.957 | 2.609 | 5.652 | 8.261 | 4.565 | 4.348 | 2.826 |
| ● *Sprattus muelleri* | 8.913 | 1.087 | 0.87 | 2.609 | 3.696 | 6.522 | 2.391 | 7.391 | 2.174 | 19.13 | 5 | 2.609 | 6.087 | 1.957 | 2.609 | 6.087 | 8.696 | 5 | 4.348 | 2.826 |
| ● *Sprattus antipodum* | 8.913 | 1.087 | 0.87 | 2.609 | 3.696 | 6.522 | 2.391 | 7.391 | 2.174 | 19.13 | 5 | 2.609 | 6.087 | 1.957 | 2.609 | 6.087 | 8.478 | 5.217 | 4.348 | 2.826 |
| ○ *Potamalosa richmondia* | 8.696 | 0.87 | 0.652 | 2.391 | 3.478 | 6.522 | 2.391 | 8.043 | 2.174 | 18.7 | 5.652 | 2.826 | 5.87 | 2.391 | 2.391 | 5.87 | 9.565 | 3.913 | 4.348 | 3.261 |
| ● *Hyperlophus vittatus* | 9.565 | 0.87 | 0.87 | 2.391 | 3.478 | 6.522 | 2.391 | 7.609 | 2.174 | 19.35 | 5 | 2.609 | 5.87 | 2.391 | 2.391 | 5.652 | 8.913 | 4.348 | 4.348 | 3.261 |
| ● *Ethmidium maculatum* | 9.13 | 1.087 | 0.87 | 2.391 | 3.478 | 6.304 | 2.391 | 8.478 | 2.174 | 18.04 | 6.087 | 1.957 | 5.87 | 2.391 | 2.391 | 5.87 | 10 | 3.478 | 4.348 | 3.261 |
| ● *Jenkinsia lamprotaenia* | 9.13 | 0.87 | 0.87 | 2.391 | 3.913 | 6.087 | 3.043 | 7.391 | 1.957 | 19.57 | 4.783 | 1.957 | 5.652 | 1.739 | 2.826 | 6.957 | 8.478 | 5.435 | 4.13 | 2.826 |
| ● *Spratelloides delicatulus* | 8.913 | 1.304 | 0.87 | 2.391 | 3.913 | 6.522 | 2.609 | 6.739 | 2.174 | 19.13 | 4.783 | 2.174 | 5.652 | 1.957 | 2.826 | 7.174 | 8.478 | 5.217 | 4.13 | 3.043 |



| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ● *Spratelloides gracilis* | 7.826 | 1.087 | 0.87 | 2.391 | 3.696 | 6.304 | 2.609 | 7.826 | 1.957 | 18.48 | 5 | 2.391 | 5.652 | 2.391 | 2.826 | 7.174 | 9.13 | 5.435 | 4.13 | 2.826 |
| ● *Etrumeus micropus* | 7.826 | 0.87 | 0.87 | 2.391 | 3.261 | 6.087 | 2.391 | 8.478 | 1.957 | 19.78 | 5.217 | 2.826 | 5.652 | 2.174 | 2.609 | 6.957 | 9.348 | 3.696 | 4.348 | 3.261 |
| ○ *Ilisha africana* | 7.391 | 0.87 | 0.87 | 2.391 | 3.696 | 6.304 | 2.391 | 9.13 | 1.957 | 19.13 | 5.87 | 2.826 | 5.652 | 2.609 | 2.609 | 5.87 | 11.09 | 2.174 | 3.913 | 3.261 |
| ○ *Pellona flavipinnis* | 7.391 | 0.87 | 0.652 | 2.391 | 2.609 | 6.304 | 2.609 | 9.565 | 2.174 | 19.78 | 4.783 | 3.261 | 6.304 | 2.391 | 2.391 | 6.087 | 10.22 | 2.609 | 4.348 | 3.261 |
| ○ *Ilisha elongata* | 6.972 | 0.871 | 0.871 | 2.397 | 2.832 | 6.318 | 2.614 | 9.586 | 2.179 | 18.74 | 5.664 | 2.832 | 6.318 | 2.179 | 2.397 | 6.1 | 10.24 | 3.268 | 4.357 | 3.268 |
| □ *Pellona ditchela* | 7.391 | 0.87 | 0.652 | 2.391 | 3.043 | 6.087 | 2.609 | 8.696 | 1.957 | 19.13 | 5.435 | 3.261 | 6.087 | 2.391 | 2.609 | 5.435 | 10.87 | 3.478 | 4.348 | 3.261 |
| □ *Anchoviella sp. LBP 2297* | 9.13 | 0.87 | 0.87 | 2.391 | 3.696 | 5.435 | 2.609 | 8.696 | 1.957 | 17.83 | 5.435 | 2.609 | 6.304 | 3.043 | 2.391 | 6.522 | 9.565 | 3.478 | 4.13 | 3.043 |
| □ *Lycengraulis grossidens* | 9.13 | 0.87 | 0.87 | 2.609 | 3.696 | 5.435 | 2.826 | 7.826 | 1.957 | 18.26 | 5.652 | 2.609 | 5.87 | 2.609 | 2.391 | 7.174 | 9.348 | 3.696 | 4.13 | 3.043 |
| ○ *Amazonsprattus scintilla* | 9.13 | 1.087 | 0.87 | 2.391 | 3.478 | 5.652 | 2.609 | 8.043 | 2.174 | 18.04 | 5.217 | 2.609 | 6.087 | 2.826 | 2.391 | 6.087 | 9.783 | 4.13 | 4.13 | 3.261 |
| ● *Engraulis encrasicolus* | 9.13 | 0.87 | 0.87 | 2.609 | 3.696 | 5.652 | 2.609 | 8.043 | 1.957 | 18.26 | 5.435 | 2.391 | 6.087 | 2.826 | 2.391 | 6.957 | 9.565 | 3.478 | 4.13 | 3.043 |
| ● *Engraulis japonicus* | 8.913 | 0.87 | 0.87 | 2.609 | 3.696 | 5.652 | 2.609 | 7.826 | 1.957 | 18.26 | 5.435 | 2.391 | 6.087 | 2.826 | 2.391 | 6.957 | 9.783 | 3.696 | 4.13 | 3.043 |
| ● *Stolephorus chinensis* | 9.13 | 1.087 | 0.87 | 2.391 | 3.696 | 5 | 2.609 | 7.609 | 2.391 | 18.7 | 5 | 2.391 | 6.087 | 2.391 | 2.391 | 6.957 | 10 | 3.913 | 4.348 | 3.043 |
| ● *Stolephorus waitei* | 9.13 | 1.087 | 0.87 | 2.391 | 3.696 | 5 | 2.609 | 7.609 | 2.391 | 18.7 | 5 | 2.391 | 6.087 | 2.391 | 2.391 | 6.957 | 10 | 3.913 | 4.348 | 3.043 |
| ○ *Lycothrissa crocodilus* | 8.913 | 0.87 | 1.087 | 2.174 | 3.696 | 5.435 | 2.391 | 8.043 | 1.957 | 18.48 | 5.217 | 3.043 | 5.87 | 2.826 | 2.391 | 6.522 | 10.65 | 3.043 | 4.348 | 3.043 |
| □ *Setipinna melanochir* | 8.696 | 0.87 | 1.087 | 2.174 | 3.261 | 5.217 | 2.391 | 8.261 | 1.957 | 19.35 | 5 | 3.043 | 6.087 | 2.826 | 2.391 | 6.304 | 10.43 | 3.261 | 4.348 | 3.043 |
| ● *Coilia reynaldi* | 9.348 | 0.87 | 1.087 | 2.174 | 3.696 | 5.435 | 2.391 | 8.043 | 2.391 | 18.91 | 5.217 | 3.261 | 6.087 | 2.391 | 2.391 | 6.087 | 9.565 | 3.261 | 4.348 | 3.043 |
| ● *Thryssa baelama* | 8.913 | 0.87 | 1.087 | 2.174 | 3.696 | 5.435 | 2.391 | 7.826 | 1.957 | 18.91 | 5.652 | 2.609 | 6.087 | 2.609 | 2.391 | 6.739 | 9.565 | 3.696 | 4.348 | 3.043 |
| ○ *Coilia lindmani* | 8.261 | 0.87 | 1.087 | 2.174 | 3.478 | 5.652 | 2.391 | 8.696 | 2.174 | 18.26 | 5.652 | 2.826 | 5.87 | 2.609 | 2.391 | 6.957 | 10.43 | 2.826 | 4.348 | 3.043 |
| □ *Coilia ectenes* | 8.043 | 0.87 | 1.087 | 2.174 | 3.696 | 5.652 | 2.391 | 8.696 | 2.174 | 18.04 | 5.652 | 2.826 | 5.87 | 2.609 | 2.391 | 7.174 | 10.43 | 2.826 | 4.348 | 3.043 |
| ● *Coilia nasus* | 8.261 | 0.87 | 1.087 | 2.174 | 3.696 | 5.652 | 2.391 | 8.913 | 2.174 | 18.04 | 5.652 | 2.826 | 5.87 | 2.609 | 2.391 | 7.174 | 10.22 | 2.609 | 4.348 | 3.043 |
| ● *Denticeps clupeoides* | 6.087 | 0.87 | 1.087 | 2.174 | 4.565 | 6.522 | 2.609 | 9.565 | 2.174 | 16.96 | 7.609 | 2.609 | 5.87 | 2.391 | 2.174 | 6.087 | 10.65 | 3.043 | 3.696 | 3.261 |

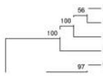
nd4



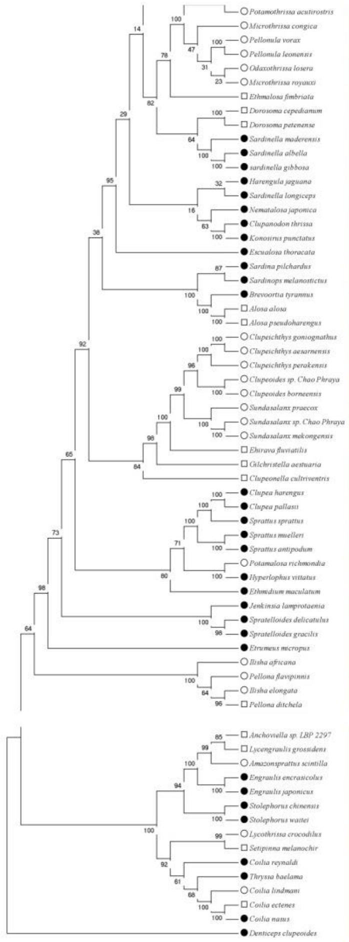


| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 13.27 | 3.061 | 1.02 | 3.061 | 6.122 | 5.102 | 4.082 | 0 | 0 | 24.49 | 6.122 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 3.061 | 1.02 | 0 |
| □ *Tenualosa ilisha* | 14.29 | 3.061 | 1.02 | 3.061 | 6.122 | 5.102 | 4.082 | 0 | 0 | 24.49 | 6.122 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 2.041 | 1.02 | 0 |
| □ *Tenualosa toli* | 13.27 | 3.061 | 1.02 | 3.061 | 6.122 | 5.102 | 4.082 | 0 | 0 | 24.49 | 6.122 | 1.02 | 2.041 | 3.061 | 3.061 | 10.2 | 10.2 | 3.061 | 1.02 | 0 |
| □ *Gudusia chapra* | 12.24 | 3.061 | 1.02 | 3.061 | 6.122 | 5.102 | 4.082 | 0 | 0 | 24.49 | 7.143 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 12.24 | 1.02 | 1.02 | 0 |
| ○ *Potamothrissa obtusirostris* | 14.29 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 2.041 | 1.02 | 0 |
| ○ *Potamothrissa acutirostris* | 14.29 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 2.041 | 1.02 | 0 |
| ○ *Microthrissa congica* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 1.02 | 2.041 | 3.061 | 3.061 | 10.2 | 10.2 | 3.061 | 1.02 | 0 |
| ○ *Pellonula vorax* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 3.061 | 1.02 | 0 |
| ○ *Pellonula leonensis* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 3.061 | 1.02 | 0 |
| ○ *Odaxothrissa losera* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 10.2 | 9.184 | 3.061 | 1.02 | 0 |
| ○ *Microthrissa royauxi* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 3.061 | 1.02 | 0 |
| □ *Ethmalosa fimbriata* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 3.061 | 1.02 | 0 |
| □ *Dorosoma cepedianum* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 8.163 | 11.22 | 3.061 | 1.02 | 0 |
| □ *Dorosoma petenense* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 3.061 | 1.02 | 0 | 23.47 | 4.082 | 2.041 | 2.041 | 4.082 | 3.061 | 8.163 | 11.22 | 4.082 | 1.02 | 0 |
| ● *Sardinella maderensis* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 3.061 | 1.02 | 0 |
| ● *Sardinella albella* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 3.061 | 1.02 | 0 |
| ● *sardinella gibbosa* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 3.061 | 1.02 | 0 |
| ● *Harengula jaguana* | 16.33 | 3.061 | 1.02 | 3.061 | 6.122 | 7.143 | 3.061 | 4.082 | 0 | 21.43 | 4.082 | 2.041 | 1.02 | 4.082 | 3.061 | 9.184 | 5.102 | 4.082 | 1.02 | 1.02 |
| ● *Sardinella longiceps* | 13.27 | 4.082 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 2.041 | 0 | 22.45 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 8.163 | 10.2 | 3.061 | 1.02 | 0 |
| ● *Nematalosa japonica* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 4.082 | 4.082 | 0 | 0 | 24.49 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 10.2 | 9.184 | 4.082 | 1.02 | 0 |
| ● *Clupanodon thrissa* | 15.31 | 3.061 | 1.02 | 3.061 | 6.122 | 6.122 | 4.082 | 1.02 | 0 | 24.49 | 6.122 | 2.041 | 2.041 | 3.061 | 3.061 | 8.163 | 8.163 | 2.041 | 1.02 | 0 |
| ● *Konosirus punctatus* | 15.31 | 3.061 | 1.02 | 3.061 | 6.122 | 5.102 | 4.082 | 1.02 | 0 | 24.49 | 6.122 | 2.041 | 2.041 | 3.061 | 3.061 | 7.143 | 10.2 | 2.041 | 1.02 | 0 |
| ● *Escualosa thoracata* | 16.33 | 4.082 | 1.02 | 3.061 | 7.143 | 4.082 | 5.102 | 1.02 | 0 | 24.49 | 4.082 | 1.02 | 2.041 | 3.061 | 2.041 | 9.184 | 8.163 | 3.061 | 1.02 | 0 |
| ● *Sardina pilchardus* | 13.27 | 4.082 | 1.02 | 3.061 | 10.2 | 4.082 | 5.102 | 1.02 | 0 | 20.41 | 5.102 | 1.02 | 2.041 | 3.061 | 2.041 | 11.22 | 7.143 | 5.102 | 1.02 | 0 |
| ● *Sardinops melanostictus* | 12.24 | 3.061 | 1.02 | 3.061 | 6.122 | 6.122 | 4.082 | 0 | 0 | 24.49 | 6.122 | 1.02 | 2.041 | 3.061 | 3.061 | 10.2 | 8.163 | 5.102 | 1.02 | 0 |
| ● *Brevoortia tyrannus* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 24.49 | 6.122 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 9.184 | 2.041 | 1.02 | 0 |
| □ *Alosa alosa* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 24.49 | 6.122 | 3.061 | 2.041 | 3.061 | 3.061 | 8.163 | 9.184 | 2.041 | 1.02 | 0 |
| □ *Alosa pseudoharengus* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 24.49 | 6.122 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 9.184 | 2.041 | 1.02 | 0 |
| ○ *Clupeichthys goniognathus* | 13.27 | 3.061 | 1.02 | 3.061 | 5.102 | 5.102 | 2.041 | 3.061 | 0 | 21.43 | 6.122 | 2.041 | 1.02 | 4.082 | 3.061 | 11.22 | 10.2 | 4.082 | 1.02 | 0 |
| ○ *Clupeichthys aesarnensis* | 14.29 | 3.061 | 1.02 | 3.061 | 5.102 | 6.122 | 2.041 | 3.061 | 0 | 21.43 | 5.102 | 1.02 | 1.02 | 4.082 | 3.061 | 12.24 | 10.2 | 3.061 | 1.02 | 0 |
| ○ *Clupeichthys perakensis* | 14.29 | 4.082 | 1.02 | 3.061 | 5.102 | 6.122 | 2.041 | 3.061 | 0 | 22.45 | 4.082 | 2.041 | 1.02 | 4.082 | 3.061 | 9.184 | 11.22 | 3.061 | 1.02 | 0 |
| ○ *Clupeoides sp. Chao Phraya* | 11.22 | 3.061 | 1.02 | 3.061 | 5.102 | 6.122 | 2.041 | 3.061 | 0 | 23.47 | 5.102 | 3.061 | 1.02 | 3.061 | 3.061 | 8.163 | 12.24 | 4.082 | 1.02 | 1.02 |
| ○ *Clupeoides borneensis* | 11.22 | 3.061 | 1.02 | 3.061 | 5.102 | 6.122 | 2.041 | 2.041 | 0 | 23.47 | 4.082 | 3.061 | 1.02 | 3.061 | 3.061 | 8.163 | 13.27 | 5.102 | 1.02 | 1.02 |
| ○ *Sundasalanx praecox* | 12.24 | 4.082 | 1.02 | 3.061 | 6.122 | 5.102 | 2.041 | 4.082 | 0 | 22.45 | 6.122 | 2.041 | 1.02 | 3.061 | 3.061 | 12.24 | 8.163 | 3.061 | 1.02 | 0 |
| ○ *Sundasalanx sp. Chao Phraya* | 14.29 | 4.082 | 1.02 | 3.061 | 6.122 | 5.102 | 2.041 | 4.082 | 0 | 22.45 | 6.122 | 2.041 | 1.02 | 3.061 | 3.061 | 11.22 | 7.143 | 3.061 | 1.02 | 0 |
| ○ *Sundasalanx mekongensis* | 11.22 | 4.082 | 1.02 | 3.061 | 6.122 | 5.102 | 2.041 | 4.082 | 0 | 22.45 | 6.122 | 3.061 | 1.02 | 3.061 | 3.061 | 10.2 | 10.2 | 3.061 | 1.02 | 0 |
| □ *Ehirava fluviatilis* | 13.27 | 3.061 | 1.02 | 4.082 | 8.163 | 6.122 | 2.041 | 1.02 | 0 | 22.45 | 4.082 | 2.041 | 1.02 | 2.041 | 4.082 | 9.184 | 11.22 | 4.082 | 1.02 | 0 |
| □ *Gilchristella aestuaria* | 14.29 | 3.061 | 1.02 | 3.061 | 8.163 | 5.102 | 3.061 | 1.02 | 0 | 23.47 | 5.102 | 2.041 | 1.02 | 3.061 | 3.061 | 9.184 | 9.184 | 3.061 | 1.02 | 1.02 |
| □ *Clupeonella cultriventris* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 25.51 | 4.082 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 2.041 | 1.02 | 0 |
| ● *Clupea harengus* | 12.24 | 3.061 | 0 | 3.061 | 7.143 | 5.102 | 3.061 | 2.041 | 0 | 22.45 | 5.102 | 4.082 | 3.061 | 5.102 | 2.041 | 6.122 | 10.2 | 5.102 | 1.02 | 0 |
| ● *Clupea pallasii* | 12.24 | 3.061 | 0 | 3.061 | 7.143 | 5.102 | 3.061 | 2.041 | 0 | 22.45 | 5.102 | 4.082 | 3.061 | 5.102 | 2.041 | 6.122 | 10.2 | 5.102 | 1.02 | 0 |
| ● *Sprattus sprattus* | 12.24 | 3.061 | 0 | 3.061 | 7.143 | 5.102 | 3.061 | 2.041 | 0 | 22.45 | 5.102 | 4.082 | 3.061 | 5.102 | 2.041 | 6.122 | 10.2 | 5.102 | 1.02 | 0 |
| ● *Sprattus muelleri* | 12.24 | 3.061 | 0 | 3.061 | 7.143 | 5.102 | 3.061 | 3.061 | 0 | 22.45 | 5.102 | 4.082 | 3.061 | 5.102 | 2.041 | 7.143 | 10.2 | 3.061 | 1.02 | 0 |
| ● *Sprattus antipodum* | 12.24 | 3.061 | 0 | 3.061 | 7.143 | 5.102 | 3.061 | 3.061 | 0 | 22.45 | 5.102 | 4.082 | 3.061 | 5.102 | 2.041 | 7.143 | 10.2 | 3.061 | 1.02 | 0 |
| ○ *Potamalosa richmondia* | 14.29 | 3.061 | 1.02 | 3.061 | 6.122 | 5.102 | 4.082 | 2.041 | 0 | 23.47 | 7.143 | 1.02 | 2.041 | 3.061 | 3.061 | 10.2 | 8.163 | 1.02 | 1.02 | 1.02 |
| ● *Hyperlophus vittatus* | 14.29 | 3.061 | 1.02 | 3.061 | 6.122 | 5.102 | 4.082 | 2.041 | 0 | 23.47 | 7.143 | 1.02 | 2.041 | 3.061 | 3.061 | 10.2 | 8.163 | 1.02 | 1.02 | 1.02 |
| ● *Ethmidium maculatum* | 10.2 | 3.061 | 1.02 | 3.061 | 8.163 | 5.102 | 4.082 | 3.061 | 0 | 21.43 | 5.102 | 2.041 | 2.041 | 3.061 | 3.061 | 9.184 | 10.2 | 5.102 | 1.02 | 0 |
| ● *Jenkinsia lamprotaenia* | 15.31 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 22.45 | 7.143 | 4.082 | 2.041 | 2.041 | 3.061 | 7.143 | 7.143 | 4.082 | 1.02 | 0 |
| ● *Spratelloides delicatulus* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 4.082 | 1.02 | 0 | 22.45 | 6.122 | 3.061 | 2.041 | 4.082 | 2.041 | 7.143 | 10.2 | 4.082 | 1.02 | 0 |
| ● *Spratelloides gracilis* | 13.27 | 3.061 | 1.02 | 4.082 | 7.143 | 5.102 | 4.082 | 3.061 | 0 | 21.43 | 6.122 | 2.041 | 2.041 | 2.041 | 3.061 | 9.184 | 9.184 | 3.061 | 1.02 | 0 |
| ● *Etrumeus micropus* | 11.22 | 5.102 | 1.02 | 2.041 | 8.163 | 5.102 | 4.082 | 4.082 | 1.02 | 21.43 | 7.143 | 2.041 | 2.041 | 1.02 | 3.061 | 8.163 | 9.184 | 2.041 | 1.02 | 1.02 |
| ○ *Ilisha africana* | 11.22 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 5.102 | 1.02 | 0 | 25.51 | 6.122 | 3.061 | 2.041 | 3.061 | 2.041 | 7.143 | 11.22 | 2.041 | 1.02 | 0 |
| ○ *Pellona flavipinnis* | 13.27 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 5.102 | 1.02 | 0 | 24.49 | 5.102 | 3.061 | 2.041 | 3.061 | 2.041 | 7.143 | 12.24 | 1.02 | 1.02 | 0 |
| ○ *Ilisha elongata* | 11.22 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 5.102 | 1.02 | 0 | 26.53 | 3.061 | 3.061 | 2.041 | 3.061 | 2.041 | 7.143 | 13.27 | 2.041 | 1.02 | 0 |
| □ *Pellona ditchela* | 12.24 | 3.061 | 1.02 | 3.061 | 7.143 | 5.102 | 5.102 | 1.02 | 0 | 25.51 | 4.082 | 3.061 | 2.041 | 3.061 | 2.041 | 7.143 | 12.24 | 2.041 | 1.02 | 0 |
| □ *Anchoviella sp. LBP 2297* | 13.27 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 2.041 | 0 | 21.43 | 7.143 | 3.061 | 3.061 | 3.061 | 3.061 | 7.143 | 10.2 | 2.041 | 1.02 | 0 |
| □ *Lycengraulis grossidens* | 12.24 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 2.041 | 0 | 21.43 | 9.184 | 3.061 | 3.061 | 3.061 | 3.061 | 7.143 | 9.184 | 2.041 | 1.02 | 0 |
| ○ *Amazonsprattus scintilla* | 12.24 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 2.041 | 0 | 21.43 | 8.163 | 3.061 | 3.061 | 3.061 | 3.061 | 7.143 | 10.2 | 2.041 | 1.02 | 0 |
| ● *Engraulis encrasicolus* | 13.27 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 2.041 | 0 | 21.43 | 5.102 | 2.041 | 3.061 | 3.061 | 3.061 | 9.184 | 10.2 | 3.061 | 1.02 | 0 |
| ● *Engraulis japonicus* | 13.27 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 2.041 | 0 | 21.43 | 6.122 | 3.061 | 3.061 | 3.061 | 3.061 | 8.163 | 10.2 | 2.041 | 1.02 | 0 |
| ● *Stolephorus chinensis* | 12.24 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 3.061 | 0 | 20.41 | 9.184 | 3.061 | 3.061 | 3.061 | 3.061 | 8.163 | 8.163 | 2.041 | 1.02 | 0 |
| ● *Stolephorus waitei* | 12.24 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 3.061 | 0 | 20.41 | 9.184 | 3.061 | 3.061 | 3.061 | 3.061 | 8.163 | 8.163 | 2.041 | 1.02 | 0 |
| ○ *Lycothrissa crocodilus* | 10.2 | 3.061 | 1.02 | 2.041 | 6.122 | 6.122 | 5.102 | 2.041 | 0 | 21.43 | 8.163 | 3.061 | 3.061 | 2.041 | 3.061 | 8.163 | 12.24 | 2.041 | 1.02 | 0 |
| □ *Setipinna melanochir* | 12.24 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 5.102 | 3.061 | 0 | 20.41 | 7.143 | 3.061 | 3.061 | 2.041 | 3.061 | 8.163 | 10.2 | 2.041 | 1.02 | 0 |
| ● *Coilia reynaldi* | 12.24 | 3.061 | 2.041 | 2.041 | 8.163 | 6.122 | 4.082 | 2.041 | 0 | 20.41 | 8.163 | 2.041 | 3.061 | 3.061 | 3.061 | 7.143 | 10.2 | 2.041 | 1.02 | 0 |
| ● *Thryssa baelama* | 13.27 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 2.041 | 0 | 20.41 | 8.163 | 3.061 | 3.061 | 3.061 | 3.061 | 7.143 | 10.2 | 2.041 | 1.02 | 0 |
| ○ *Coilia lindmani* | 12.24 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 2.041 | 0 | 20.41 | 7.143 | 3.061 | 3.061 | 3.061 | 3.061 | 7.143 | 10.2 | 4.082 | 1.02 | 0 |
| □ *Coilia ectenes* | 12.24 | 3.061 | 1.02 | 2.041 | 8.163 | 6.122 | 4.082 | 2.041 | 0 | 19.39 | 7.143 | 3.061 | 3.061 | 3.061 | 3.061 | 7.143 | 10.2 | 4.082 | 1.02 | 0 |
| ● *Coilia nasus* | 12.24 | 3.061 | 1.02 | 2.041 | 7.143 | 6.122 | 4.082 | 2.041 | 0 | 20.41 | 7.143 | 3.061 | 3.061 | 3.061 | 3.061 | 7.143 | 10.2 | 4.082 | 1.02 | 0 |
| ● *Denticeps clupeoides* | 12.24 | 4.082 | 1.02 | 3.061 | 5.102 | 5.102 | 5.102 | 2.041 | 1.02 | 24.49 | 5.102 | 2.041 | 2.041 | 2.041 | 2.041 | 7.143 | 12.24 | 1.02 | 1.02 | 2.041 |

nd4l



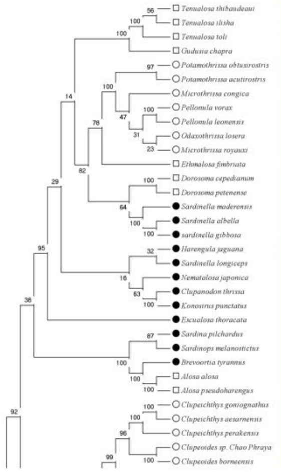
| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 9.002 | 1.146 | 2.619 | 1.473 | 6.056 | 5.728 | 2.291 | 8.347 | 3.273 | 14.08 | 5.074 | 3.601 | 4.91 | 3.437 | 1.637 | 7.692 | 10.64 | 5.074 | 2.291 | 1.637 |
| □ *Tenualosa ilisha* | 9.329 | 1.146 | 2.455 | 1.473 | 6.383 | 5.728 | 2.455 | 8.674 | 3.273 | 14.08 | 5.237 | 4.092 | 5.565 | 3.273 | 1.637 | 7.038 | 9.656 | 4.746 | 2.291 | 1.473 |
| □ *Tenualosa toli* | 9.493 | 1.146 | 2.619 | 1.473 | 5.892 | 5.237 | 2.455 | 7.365 | 3.273 | 15.06 | 4.746 | 4.092 | 5.074 | 3.273 | 1.637 | 7.365 | 11.13 | 4.746 | 2.291 | 1.637 |
| □ *Gudusia chapra* | 8.347 | 1.146 | 2.291 | 1.473 | 6.056 | 5.074 | 2.291 | 9.165 | 3.273 | 14.89 | 5.401 | 4.91 | 4.91 | 3.11 | 1.637 | 8.02 | 10.97 | 3.11 | 2.291 | 1.637 |
| ○ *Potamothrissa obtusirostris* | 10.31 | 0.982 | 2.619 | 1.637 | 6.383 | 5.237 | 2.128 | 8.511 | 3.273 | 14.57 | 4.91 | 4.419 | 4.746 | 3.273 | 1.637 | 6.874 | 9.82 | 4.419 | 2.291 | 1.964 |



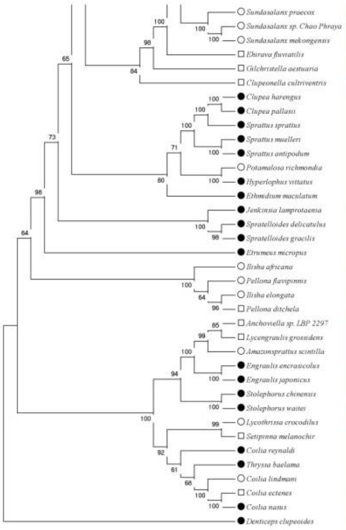
| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ○ *Potamothrissa acutirostris* | 9.656 | 0.982 | 2.455 | 1.637 | 6.383 | 5.401 | 2.291 | 8.511 | 3.273 | 14.57 | 4.583 | 4.583 | 4.91 | 3.273 | 1.637 | 6.874 | 9.984 | 4.746 | 2.291 | 1.964 |
| ○ *Microthrissa congica* | 9.82 | 0.982 | 2.946 | 1.637 | 6.383 | 5.237 | 2.128 | 8.511 | 3.273 | 14.4 | 4.583 | 4.092 | 4.91 | 3.273 | 1.637 | 6.874 | 10.47 | 4.583 | 2.291 | 1.964 |
| ○ *Pellonula vorax* | 9.984 | 0.982 | 2.619 | 1.637 | 6.547 | 5.237 | 2.128 | 8.674 | 3.273 | 14.24 | 4.583 | 4.419 | 4.746 | 3.273 | 1.637 | 6.71 | 10.47 | 4.583 | 2.291 | 1.964 |
| ○ *Pellonula leonensis* | 9.82 | 0.982 | 2.455 | 1.637 | 6.383 | 5.237 | 2.128 | 8.511 | 3.273 | 14.4 | 4.583 | 4.419 | 4.91 | 3.273 | 1.637 | 6.71 | 10.8 | 4.583 | 2.291 | 1.964 |
| ○ *Odaxothrissa losera* | 9.493 | 0.982 | 2.619 | 1.637 | 6.383 | 5.237 | 2.128 | 8.674 | 3.273 | 14.4 | 5.074 | 4.419 | 4.91 | 3.273 | 1.637 | 6.71 | 10.97 | 3.928 | 2.291 | 1.964 |
| ○ *Microthrissa royauxi* | 9.656 | 0.982 | 2.782 | 1.637 | 6.383 | 5.565 | 2.128 | 9.165 | 3.437 | 14.4 | 4.91 | 3.928 | 4.91 | 3.273 | 1.637 | 6.874 | 10.15 | 4.092 | 2.128 | 1.964 |
| □ *Ethmalosa fimbriata* | 9.984 | 0.982 | 2.455 | 1.8 | 6.383 | 5.401 | 2.291 | 8.674 | 3.11 | 14.57 | 4.746 | 4.419 | 4.91 | 2.946 | 1.637 | 6.71 | 10.31 | 4.583 | 2.291 | 1.8 |
| □ *Dorosoma cepedianum* | 10.47 | 0.982 | 2.619 | 1.637 | 6.383 | 5.401 | 2.291 | 8.511 | 3.273 | 14.24 | 4.583 | 4.092 | 4.91 | 3.11 | 1.637 | 6.547 | 9.984 | 5.074 | 2.291 | 1.964 |
| □ *Dorosoma petenense* | 9.82 | 0.982 | 2.782 | 1.637 | 6.383 | 5.401 | 2.291 | 8.674 | 3.273 | 14.24 | 4.746 | 3.928 | 4.91 | 3.11 | 1.637 | 6.547 | 10.64 | 4.746 | 2.291 | 1.964 |
| ● *Sardinella maderensis* | 9.984 | 0.982 | 2.455 | 1.637 | 6.383 | 5.565 | 2.291 | 8.511 | 3.273 | 14.4 | 4.746 | 3.928 | 4.91 | 3.11 | 1.637 | 7.038 | 9.984 | 4.91 | 2.291 | 1.964 |
| ● *Sardinella albella* | 9.984 | 0.982 | 2.455 | 1.8 | 6.383 | 5.237 | 2.291 | 8.02 | 3.273 | 14.4 | 4.746 | 4.255 | 4.91 | 3.11 | 1.637 | 6.71 | 10.15 | 5.401 | 2.291 | 1.964 |
| ● *sardinella gibbosa* | 9.984 | 0.982 | 2.455 | 1.8 | 6.383 | 5.237 | 2.291 | 8.347 | 3.273 | 14.4 | 4.746 | 4.255 | 4.91 | 3.11 | 1.637 | 6.71 | 10.15 | 5.074 | 2.291 | 1.964 |
| ● *Harengula jaguana* | 9.493 | 1.309 | 2.455 | 1.637 | 6.383 | 5.565 | 2.291 | 8.674 | 3.437 | 13.75 | 4.419 | 4.255 | 4.91 | 2.946 | 1.8 | 7.201 | 9.493 | 5.401 | 2.455 | 2.128 |
| ● *Sardinella longiceps* | 10.15 | 1.146 | 2.291 | 1.964 | 6.547 | 5.401 | 2.291 | 8.838 | 3.273 | 14.08 | 4.583 | 4.092 | 4.746 | 3.11 | 1.637 | 6.71 | 10.15 | 4.91 | 2.291 | 1.8 |
| ● *Nematalosa japonica* | 10.15 | 0.982 | 2.455 | 1.8 | 6.383 | 5.401 | 2.291 | 8.183 | 3.11 | 14.4 | 4.583 | 4.419 | 4.746 | 2.946 | 1.637 | 7.038 | 9.984 | 5.237 | 2.291 | 1.964 |
| ● *Clupanodon thrissa* | 10.47 | 0.982 | 2.455 | 1.637 | 6.874 | 5.237 | 2.291 | 8.183 | 3.273 | 14.08 | 4.419 | 4.419 | 4.91 | 3.11 | 1.637 | 6.71 | 10.47 | 4.746 | 2.291 | 1.8 |
| ● *Konosirus punctatus* | 10.97 | 0.982 | 2.619 | 1.637 | 6.547 | 5.237 | 2.291 | 8.511 | 3.273 | 14.08 | 4.583 | 4.255 | 4.583 | 3.11 | 1.637 | 7.038 | 9.82 | 4.746 | 2.291 | 1.8 |
| ● *Escualosa thoracata* | 11.13 | 0.982 | 2.455 | 1.637 | 6.383 | 5.401 | 2.291 | 8.02 | 3.11 | 14.4 | 4.583 | 4.255 | 4.746 | 3.11 | 1.637 | 6.71 | 9.493 | 5.401 | 2.291 | 1.964 |
| ● *Sardina pilchardus* | 9.82 | 0.982 | 2.128 | 1.8 | 6.219 | 5.565 | 2.128 | 8.347 | 3.273 | 14.4 | 4.419 | 3.764 | 4.746 | 3.437 | 1.637 | 7.856 | 9.165 | 6.219 | 2.291 | 1.8 |
| ● *Sardinops melanostictus* | 9.82 | 0.982 | 2.291 | 1.637 | 6.547 | 5.401 | 2.128 | 8.347 | 3.273 | 14.4 | 4.746 | 4.092 | 5.074 | 3.11 | 1.637 | 7.201 | 9.984 | 5.237 | 2.291 | 1.8 |
| ● *Brevoortia tyrannus* | 10.64 | 0.982 | 2.291 | 1.637 | 6.383 | 5.401 | 2.128 | 8.674 | 3.273 | 13.91 | 4.746 | 4.092 | 5.074 | 3.273 | 1.637 | 7.038 | 9.493 | 5.074 | 2.291 | 1.964 |
| □ *Alosa alosa* | 10.64 | 0.982 | 2.291 | 1.637 | 6.547 | 5.401 | 2.128 | 8.511 | 3.273 | 13.75 | 5.074 | 4.092 | 5.237 | 3.11 | 1.637 | 6.874 | 9.656 | 5.074 | 2.291 | 1.8 |
| □ *Alosa pseudoharengus* | 10.64 | 0.982 | 2.291 | 1.637 | 6.383 | 5.401 | 1.964 | 8.347 | 3.273 | 14.08 | 5.074 | 4.255 | 5.237 | 3.273 | 1.637 | 7.038 | 9.493 | 4.746 | 2.291 | 1.964 |
| ○ *Clupeichthys goniognathus* | 8.674 | 0.982 | 2.291 | 1.8 | 6.874 | 5.237 | 2.291 | 8.511 | 3.273 | 14.24 | 5.074 | 4.91 | 4.746 | 3.437 | 1.473 | 7.201 | 10.31 | 4.583 | 2.291 | 1.8 |
| ○ *Clupeichthys aesarnensis* | 8.674 | 0.982 | 2.291 | 1.8 | 7.038 | 5.237 | 2.128 | 8.674 | 3.273 | 14.08 | 4.91 | 4.91 | 4.583 | 3.437 | 1.473 | 7.856 | 10.31 | 4.255 | 2.291 | 1.8 |
| ○ *Clupeichthys perakensis* | 9.165 | 0.982 | 2.128 | 1.8 | 6.874 | 5.074 | 2.128 | 8.674 | 3.273 | 14.57 | 5.237 | 5.074 | 4.419 | 3.437 | 1.473 | 7.201 | 10.31 | 4.092 | 2.291 | 1.8 |
| ○ *Clupeoides sp. Chao Phraya* | 8.347 | 0.982 | 2.128 | 1.8 | 6.71 | 5.237 | 1.964 | 9.493 | 3.437 | 13.75 | 5.728 | 4.746 | 4.419 | 3.601 | 1.473 | 7.201 | 10.64 | 4.092 | 2.291 | 1.964 |
| ○ *Clupeoides borneensis* | 8.183 | 0.982 | 2.128 | 1.8 | 6.547 | 5.237 | 1.964 | 9.329 | 3.437 | 14.4 | 5.237 | 4.746 | 4.583 | 3.601 | 1.473 | 7.365 | 10.64 | 4.092 | 2.291 | 1.964 |
| ○ *Sundasalanx praecox* | 9.165 | 0.982 | 2.291 | 1.964 | 7.038 | 5.401 | 2.128 | 9.002 | 3.437 | 13.75 | 4.91 | 3.601 | 4.583 | 3.437 | 1.473 | 8.02 | 10.15 | 4.255 | 2.291 | 2.128 |
| ○ *Sundasalanx sp. Chao Phraya* | 9.165 | 0.982 | 2.455 | 1.8 | 6.874 | 5.074 | 1.8 | 8.838 | 3.273 | 13.91 | 5.237 | 4.255 | 4.255 | 3.601 | 1.473 | 7.856 | 10.97 | 3.601 | 2.291 | 2.291 |
| ○ *Sundasalanx mekongensis* | 8.511 | 0.982 | 2.291 | 1.8 | 6.874 | 5.401 | 2.128 | 9.002 | 3.273 | 13.75 | 5.565 | 4.583 | 4.419 | 3.437 | 1.473 | 7.365 | 11.46 | 3.273 | 2.291 | 2.128 |
| □ *Ehirava fluviatilis* | 9.165 | 0.982 | 2.455 | 1.8 | 6.547 | 5.074 | 1.964 | 9.165 | 3.273 | 14.73 | 4.583 | 4.583 | 4.419 | 3.601 | 1.473 | 6.71 | 11.46 | 3.764 | 2.291 | 1.964 |
| □ *Gilchristella aestuaria* | 9.656 | 0.982 | 2.291 | 1.8 | 6.547 | 5.401 | 2.128 | 8.674 | 3.273 | 13.75 | 5.728 | 4.746 | 4.583 | 3.437 | 1.637 | 7.038 | 9.329 | 4.746 | 2.291 | 1.964 |
| □ *Clupeonella cultriventris* | 9.493 | 0.982 | 2.455 | 1.637 | 6.547 | 5.401 | 2.128 | 8.511 | 3.273 | 14.24 | 4.746 | 4.255 | 4.746 | 3.273 | 1.637 | 7.201 | 9.984 | 5.237 | 2.291 | 1.964 |
| ● *Clupea harengus* | 9.656 | 0.982 | 2.128 | 1.637 | 6.383 | 5.237 | 2.291 | 8.347 | 3.273 | 13.91 | 5.074 | 4.91 | 4.746 | 3.273 | 1.637 | 7.038 | 9.82 | 5.401 | 2.291 | 1.964 |
| ● *Clupea pallasii* | 9.82 | 0.982 | 2.128 | 1.637 | 6.383 | 5.074 | 2.291 | 8.183 | 3.273 | 14.08 | 5.074 | 4.91 | 4.746 | 3.273 | 1.637 | 7.038 | 9.984 | 5.237 | 2.291 | 1.964 |
| ● *Sprattus sprattus* | 9.82 | 0.982 | 2.128 | 1.8 | 6.383 | 5.074 | 2.291 | 7.856 | 3.273 | 14.08 | 4.91 | 4.746 | 4.583 | 3.11 | 1.637 | 7.201 | 10.31 | 5.565 | 2.291 | 1.964 |
| ● *Sprattus muelleri* | 10.15 | 0.982 | 2.128 | 1.637 | 6.219 | 5.237 | 2.128 | 8.347 | 3.273 | 14.08 | 4.91 | 4.583 | 4.746 | 3.273 | 1.637 | 6.547 | 10.97 | 4.91 | 2.291 | 1.964 |
| ● *Sprattus antipodum* | 10.31 | 0.982 | 2.128 | 1.637 | 6.219 | 5.237 | 2.128 | 8.347 | 3.273 | 14.24 | 4.746 | 4.583 | 4.746 | 3.273 | 1.637 | 6.547 | 10.8 | 4.91 | 2.291 | 1.964 |
| ○ *Potamalosa richmondia* | 9.329 | 0.982 | 2.291 | 1.8 | 6.383 | 5.237 | 1.964 | 8.674 | 3.437 | 13.75 | 5.401 | 4.583 | 4.91 | 3.273 | 1.637 | 7.201 | 10.47 | 4.255 | 2.291 | 2.128 |
| ● *Hyperlophus vittatus* | 9.656 | 0.982 | 2.291 | 1.8 | 6.219 | 5.237 | 1.964 | 8.183 | 3.437 | 14.4 | 4.746 | 4.583 | 4.91 | 3.273 | 1.637 | 7.038 | 10.31 | 4.91 | 2.291 | 2.128 |
| ● *Ethmidium maculatum* | 9.002 | 0.982 | 1.964 | 1.964 | 6.219 | 5.401 | 2.291 | 8.674 | 3.273 | 14.08 | 5.728 | 4.583 | 4.91 | 3.437 | 1.473 | 7.038 | 10.15 | 4.419 | 2.291 | 2.128 |
| ● *Jenkinsia lamprotaenia* | 9.493 | 1.146 | 2.619 | 1.637 | 6.71 | 5.401 | 2.455 | 6.71 | 3.601 | 16.37 | 4.255 | 3.764 | 5.074 | 3.273 | 1.637 | 7.856 | 8.511 | 5.728 | 2.291 | 1.473 |
| ● *Spratelloides delicatulus* | 9.656 | 0.982 | 2.455 | 1.637 | 6.547 | 5.728 | 2.291 | 6.547 | 3.437 | 16.53 | 4.255 | 4.092 | 5.074 | 3.437 | 1.637 | 6.874 | 8.674 | 6.219 | 2.128 | 1.8 |
| ● *Spratelloides gracilis* | 9.493 | 0.982 | 2.291 | 1.8 | 6.383 | 5.728 | 2.619 | 7.856 | 3.273 | 16.04 | 4.419 | 4.092 | 4.746 | 3.273 | 1.637 | 6.874 | 9.165 | 5.237 | 2.291 | 1.8 |
| ● *Etrumeus micropus* | 9.329 | 0.982 | 2.128 | 1.8 | 6.219 | 5.401 | 2.128 | 8.838 | 3.273 | 14.57 | 4.91 | 4.255 | 4.746 | 3.601 | 1.637 | 7.038 | 9.984 | 4.746 | 2.291 | 2.128 |
| ○ *Ilisha africana* | 7.705 | 0.984 | 1.967 | 1.639 | 5.902 | 5.41 | 2.787 | 9.344 | 3.279 | 15.25 | 5.082 | 4.918 | 4.918 | 3.279 | 1.639 | 7.049 | 11.31 | 3.279 | 2.295 | 1.967 |
| ○ *Pellona flavipinnis* | 8.674 | 0.982 | 2.128 | 1.964 | 6.219 | 5.237 | 2.291 | 8.02 | 3.437 | 14.24 | 5.728 | 5.074 | 4.583 | 3.601 | 1.473 | 7.038 | 11.62 | 3.928 | 2.291 | 1.473 |
| ○ *Ilisha elongata* | 7.869 | 0.984 | 2.131 | 1.639 | 5.902 | 5.246 | 2.787 | 8.852 | 3.443 | 14.75 | 5.738 | 5.082 | 4.754 | 3.607 | 1.475 | 6.721 | 11.97 | 3.607 | 2.295 | 1.148 |
| □ *Pellona ditchela* | 8.838 | 0.982 | 2.128 | 1.637 | 5.728 | 5.401 | 2.619 | 8.511 | 3.437 | 14.08 | 6.383 | 4.91 | 4.746 | 3.437 | 1.473 | 7.038 | 11.78 | 3.11 | 2.291 | 1.473 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Anchoviella sp. LBP 2297* | 9.329 | 0.982 | 2.128 | 1.964 | 6.383 | 5.237 | 2.291 | 8.838 | 3.273 | 14.57 | 4.91 | 4.91 | 4.583 | 3.437 | 1.473 | 6.383 | 10.64 | 4.746 | 2.291 | 1.637 |
| □ *Lycengraulis grossidens* | 9.002 | 0.982 | 2.291 | 2.128 | 6.219 | 5.237 | 2.291 | 9.002 | 3.273 | 14.73 | 5.237 | 4.419 | 4.583 | 3.273 | 1.473 | 7.038 | 10.97 | 3.928 | 2.291 | 1.637 |
| ○ *Amazonsprattus scintilla* | 9.165 | 0.982 | 2.455 | 1.8 | 6.547 | 5.401 | 2.128 | 8.674 | 3.273 | 14.24 | 4.91 | 4.419 | 4.583 | 3.601 | 1.473 | 6.874 | 10.31 | 5.237 | 2.291 | 1.637 |
| ● *Engraulis encrasicolus* | 9.002 | 0.982 | 2.128 | 1.964 | 6.547 | 5.237 | 2.291 | 8.674 | 3.437 | 14.08 | 4.746 | 4.255 | 4.583 | 3.437 | 1.473 | 7.038 | 10.8 | 5.565 | 2.291 | 1.473 |
| ● *Engraulis japonicus* | 9.165 | 0.982 | 2.128 | 1.964 | 6.547 | 5.237 | 2.291 | 8.511 | 3.437 | 14.08 | 4.746 | 4.092 | 4.583 | 3.437 | 1.473 | 7.038 | 10.8 | 5.728 | 2.291 | 1.473 |
| ● *Stolephorus chinensis* | 8.674 | 1.146 | 2.455 | 1.8 | 6.219 | 5.237 | 1.964 | 9.493 | 3.437 | 15.22 | 3.928 | 4.746 | 4.583 | 3.601 | 1.473 | 7.201 | 9.82 | 4.91 | 2.291 | 1.8 |
| ● *Stolephorus waitei* | 8.838 | 1.146 | 2.455 | 1.8 | 6.056 | 5.237 | 1.964 | 9.002 | 3.437 | 15.38 | 3.928 | 4.746 | 4.583 | 3.601 | 1.473 | 7.201 | 9.656 | 5.401 | 2.291 | 1.8 |
| ○ *Lycothrissa crocodilus* | 8.02 | 0.982 | 2.291 | 1.964 | 6.547 | 5.237 | 2.128 | 8.511 | 3.437 | 15.38 | 4.91 | 4.91 | 4.91 | 3.437 | 1.473 | 7.201 | 11.78 | 2.946 | 2.291 | 1.637 |
| □ *Setipinna melanochir* | 8.183 | 0.982 | 2.128 | 1.964 | 6.71 | 5.237 | 2.291 | 8.347 | 3.601 | 14.57 | 4.746 | 4.746 | 5.074 | 3.11 | 1.473 | 7.038 | 12.11 | 3.601 | 2.291 | 1.8 |
| ● *Coilia reynaldi* | 8.674 | 0.982 | 2.128 | 1.964 | 5.728 | 5.237 | 2.455 | 9.82 | 3.601 | 14.89 | 4.583 | 4.746 | 4.583 | 3.273 | 1.473 | 7.201 | 10.64 | 3.764 | 2.291 | 1.964 |
| ● *Thryssa baelama* | 8.674 | 0.982 | 2.128 | 1.8 | 6.383 | 5.074 | 2.128 | 8.347 | 3.601 | 14.08 | 5.074 | 4.583 | 4.91 | 3.437 | 1.473 | 7.529 | 10.8 | 4.91 | 2.291 | 1.8 |
| ○ *Coilia lindmani* | 8.674 | 0.982 | 2.128 | 2.128 | 7.038 | 5.237 | 2.291 | 8.511 | 3.764 | 14.24 | 4.91 | 4.746 | 4.583 | 3.11 | 1.473 | 7.201 | 10.97 | 3.928 | 2.291 | 1.8 |
| □ *Coilia ectenes* | 8.674 | 0.982 | 2.128 | 2.128 | 6.71 | 5.237 | 2.291 | 8.674 | 3.764 | 14.24 | 5.074 | 4.746 | 4.583 | 3.11 | 1.473 | 7.038 | 11.13 | 3.764 | 2.291 | 1.964 |
| ● *Coilia nasus* | 8.347 | 0.982 | 2.128 | 2.128 | 6.71 | 5.237 | 2.291 | 8.838 | 3.764 | 14.24 | 5.074 | 4.746 | 4.583 | 3.11 | 1.473 | 7.038 | 11.29 | 3.764 | 2.291 | 1.964 |
| ● *Denticeps clupeoides* | 8.21 | 0.985 | 1.97 | 1.806 | 6.897 | 5.09 | 2.627 | 9.195 | 3.284 | 14.78 | 6.076 | 4.926 | 4.433 | 3.941 | 1.478 | 7.225 | 9.688 | 3.612 | 2.299 | 1.478 |

nd5

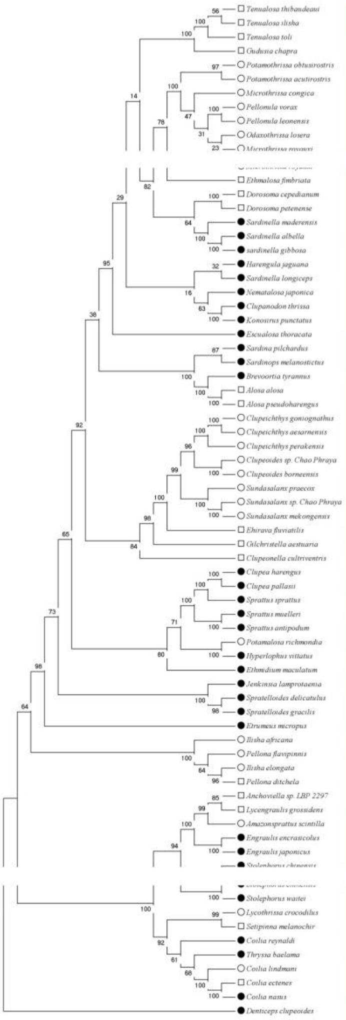


| | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| □ *Tenualosa thibaudeaui* | 11.43 | 1.714 | 2.286 | 2.286 | 9.143 | 12.57 | 1.714 | 2.857 | 1.143 | 17.71 | 4.571 | 0.571 | 3.429 | 0 | 2.286 | 6.857 | 2.286 | 11.43 | 1.143 | 4.571 |
| □ *Tenualosa ilisha* | 12.64 | 1.724 | 1.724 | 2.874 | 7.471 | 12.64 | 1.724 | 3.448 | 0.575 | 17.24 | 5.172 | 0.575 | 4.598 | 0 | 2.299 | 7.471 | 2.874 | 9.195 | 1.149 | 4.598 |
| □ *Tenualosa toli* | 11.49 | 1.724 | 1.724 | 2.874 | 8.046 | 12.64 | 1.149 | 2.874 | 0.575 | 17.82 | 5.172 | 0.575 | 3.448 | 0 | 2.299 | 8.046 | 2.874 | 10.34 | 1.149 | 5.172 |
| □ *Gudusia chapra* | 10.92 | 1.724 | 1.724 | 2.299 | 9.195 | 13.22 | 1.724 | 2.874 | 0.575 | 17.24 | 5.172 | 1.149 | 3.448 | 0 | 2.299 | 6.897 | 2.299 | 11.49 | 1.149 | 4.598 |
| ○ *Potamothrissa obtusirostris* | 14.86 | 1.143 | 2.286 | 2.857 | 9.714 | 11.43 | 1.143 | 2.857 | 0.571 | 15.43 | 6.286 | 0.571 | 2.857 | 0 | 2.286 | 6.857 | 1.714 | 12 | 1.143 | 4 |
| ○ *Potamothrissa acutirostris* | 13.71 | 1.143 | 2.286 | 2.857 | 10.29 | 12 | 1.143 | 2.286 | 0.571 | 14.29 | 8 | 0.571 | 2.857 | 0 | 2.286 | 6.857 | 2.286 | 10.86 | 1.143 | 4.571 |
| ○ *Microthrissa congica* | 12 | 1.143 | 2.286 | 2.857 | 9.714 | 12.57 | 1.143 | 1.714 | 0.571 | 13.71 | 6.286 | 0.571 | 2.857 | 0 | 2.857 | 6.857 | 2.286 | 14.86 | 1.143 | 4.571 |
| ○ *Pellonula vorax* | 13.14 | 1.143 | 2.286 | 2.857 | 10.29 | 12 | 1.143 | 1.714 | 0.571 | 14.86 | 6.286 | 0.571 | 2.857 | 0 | 2.286 | 6.857 | 2.286 | 13.14 | 1.143 | 4.571 |
| ○ *Pellonula leonensis* | 13.14 | 1.143 | 2.286 | 2.857 | 10.29 | 12 | 1.143 | 2.286 | 0.571 | 14.86 | 6.286 | 0.571 | 2.857 | 0 | 2.286 | 6.857 | 2.286 | 12.57 | 1.143 | 4.571 |
| ○ *Odaxothrissa losera* | 12.57 | 1.714 | 1.714 | 4 | 10.29 | 12.57 | 1.143 | 1.714 | 0.571 | 15.43 | 5.143 | 1.143 | 2.857 | 0 | 2.286 | 5.714 | 1.143 | 14.29 | 1.143 | 4.571 |
| ○ *Microthrissa royauxi* | 13.14 | 1.143 | 2.857 | 2.286 | 10.29 | 12 | 1.143 | 1.714 | 0.571 | 14.86 | 6.286 | 0.571 | 2.857 | 0 | 2.286 | 6.857 | 2.286 | 13.14 | 1.143 | 4.571 |
| □ *Ethmalosa fimbriata* | 11.43 | 0.571 | 2.286 | 2.857 | 9.714 | 13.14 | 1.143 | 1.714 | 0.571 | 14.86 | 5.714 | 0.571 | 3.429 | 0 | 2.286 | 7.429 | 2.286 | 13.14 | 1.714 | 5.143 |
| □ *Dorosoma cepedianum* | 13.14 | 1.143 | 1.714 | 2.857 | 9.143 | 12 | 1.143 | 2.857 | 0.571 | 16 | 6.286 | 0.571 | 2.857 | 0 | 2.286 | 7.429 | 2.857 | 11.43 | 1.143 | 4.571 |
| □ *Dorosoma petenense* | 10.29 | 1.714 | 1.714 | 3.429 | 10.29 | 12 | 0.571 | 2.286 | 0.571 | 14.86 | 5.714 | 0.571 | 2.857 | 0 | 2.286 | 9.143 | 3.429 | 13.14 | 1.143 | 4 |
| ● *Sardinella maderensis* | 14.29 | 1.143 | 1.714 | 2.857 | 8.571 | 13.14 | 1.143 | 4.571 | 0.571 | 15.43 | 5.143 | 0.571 | 2.857 | 0 | 2.286 | 6.857 | 2.857 | 10.29 | 1.143 | 4.571 |
| ● *Sardinella albella* | 14.29 | 1.143 | 1.714 | 2.857 | 9.143 | 12.57 | 1.143 | 4 | 0.571 | 14.86 | 5.143 | 0.571 | 2.857 | 0 | 2.286 | 6.857 | 2.286 | 12 | 1.143 | 4.571 |
| ● *sardinella gibbosa* | 14.29 | 1.143 | 1.714 | 2.857 | 9.143 | 12.57 | 1.143 | 4 | 0.571 | 14.86 | 5.143 | 0.571 | 2.857 | 0 | 2.286 | 6.857 | 2.286 | 12 | 1.143 | 4.571 |
| ● *Harengula jaguana* | 10.29 | 1.714 | 1.143 | 5.143 | 8 | 9.143 | 0 | 5.143 | 0.571 | 15.43 | 3.429 | 1.143 | 3.429 | 1.143 | 1.714 | 6.857 | 5.143 | 13.14 | 2.857 | 4.571 |
| ● *Sardinella longiceps* | 12 | 2.857 | 1.143 | 3.429 | 7.429 | 13.14 | 0.571 | 2.286 | 0.571 | 18.29 | 5.143 | 1.143 | 3.429 | 0 | 2.286 | 5.714 | 2.286 | 13.14 | 1.143 | 4 |
| ● *Nematalosa japonica* | 11.43 | 1.143 | 2.286 | 2.857 | 6.286 | 11.43 | 0.571 | 2.857 | 1.143 | 19.43 | 4 | 1.143 | 4.571 | 0 | 2.286 | 6.857 | 3.429 | 12.57 | 1.143 | 4.571 |
| ● *Clupanodon thrissa* | 12 | 1.143 | 1.714 | 3.429 | 7.429 | 10.86 | 0 | 1.143 | 1.143 | 16 | 6.857 | 0.571 | 3.429 | 0 | 2.857 | 6.286 | 4 | 14.29 | 1.143 | 5.714 |
| ● *Konosirus punctatus* | 10.86 | 1.143 | 2.286 | 4 | 8 | 11.43 | 0 | 1.714 | 0.571 | 17.71 | 6.286 | 0 | 3.429 | 0 | 3.429 | 6.286 | 2.857 | 13.71 | 1.143 | 5.143 |
| ● *Escualosa thoracata* | 9.714 | 2.286 | 1.714 | 2.857 | 9.143 | 12 | 0.571 | 4.571 | 1.143 | 14.29 | 6.857 | 1.714 | 4.571 | 0 | 2.857 | 5.143 | 1.714 | 13.71 | 0.571 | 4.571 |
| ● *Sardina pilchardus* | 11.36 | 3.977 | 2.273 | 5.114 | 6.818 | 10.8 | 1.705 | 4.545 | 1.705 | 14.77 | 4.545 | 1.136 | 5.114 | 0 | 2.841 | 6.25 | 1.136 | 13.07 | 0.568 | 2.273 |
| ● *Sardinops melanostictus* | 12.57 | 2.857 | 1.143 | 3.429 | 8.571 | 9.714 | 0 | 4.571 | 0.571 | 15.43 | 4 | 2.286 | 4 | 1.143 | 2.286 | 5.714 | 5.714 | 11.43 | 1.143 | 3.429 |
| ● *Brevoortia tyrannus* | 11.43 | 2.286 | 2.286 | 2.857 | 8 | 12 | 1.143 | 1.143 | 0.571 | 16 | 5.714 | 0.571 | 2.857 | 0 | 2.286 | 8 | 2.286 | 13.71 | 2.286 | 4.571 |
| □ *Alosa alosa* | 11.43 | 2.286 | 1.714 | 2.857 | 8 | 12.57 | 1.143 | 1.143 | 0.571 | 16.57 | 5.714 | 0.571 | 2.857 | 0 | 2.286 | 8 | 2.286 | 13.14 | 2.286 | 4.571 |
| □ *Alosa pseudoharengus* | 12 | 2.286 | 2.286 | 2.857 | 8 | 12 | 1.143 | 0.571 | 0.571 | 16 | 5.714 | 0.571 | 2.857 | 0 | 2.286 | 8.571 | 1.143 | 14.29 | 2.286 | 4.571 |
| ○ *Clupeichthys goniognathus* | 13.22 | 2.299 | 2.299 | 2.874 | 10.34 | 12.07 | 1.149 | 2.299 | 0.575 | 16.09 | 3.448 | 0.575 | 4.598 | 0 | 2.299 | 6.322 | 2.299 | 11.49 | 1.724 | 4.023 |
| ○ *Clupeichthys aesarnensis* | 12.64 | 2.299 | 2.299 | 2.874 | 9.77 | 12.07 | 1.149 | 2.299 | 0.575 | 16.09 | 4.023 | 0 | 4.598 | 0 | 2.299 | 6.897 | 2.299 | 12.07 | 1.724 | 4.023 |
| ○ *Clupeichthys perakensis* | 12.64 | 2.299 | 1.724 | 3.448 | 10.92 | 12.07 | 0.575 | 2.299 | 0.575 | 16.09 | 4.023 | 0.575 | 4.598 | 0 | 2.299 | 6.322 | 2.874 | 10.34 | 1.724 | 4.598 |
| ○ *Clupeoides sp. Chao Phraya* | 9.714 | 2.286 | 1.714 | 3.429 | 9.714 | 12 | 0.571 | 2.857 | 0.571 | 16.57 | 4 | 1.143 | 3.429 | 1.143 | 1.714 | 7.429 | 3.429 | 13.14 | 1.143 | 4 |
| ○ *Clupeoides borneensis* | 9.714 | 2.286 | 1.714 | 3.429 | 8.571 | 11.43 | 0.571 | 2.857 | 0.571 | 16.57 | 3.429 | 1.143 | 3.429 | 1.143 | 1.714 | 9.143 | 3.429 | 13.71 | 1.143 | 4 |



| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12.57 | 3.429 | 1.143 | 3.429 | 8 | 10.86 | 1.143 | 1.143 | 0.571 | 18.86 | 3.429 | 1.143 | 3.429 | 0.571 | 2.286 | 5.714 | 4 | 12.57 | 1.714 | 4 |
| 10.29 | 3.429 | 1.714 | 2.286 | 6.857 | 11.43 | 1.143 | 1.143 | 0.571 | 18.86 | 4.571 | 1.143 | 4 | 1.143 | 2.286 | 4.571 | 3.429 | 14.86 | 1.714 | 4.571 |
| 10.86 | 3.429 | 1.714 | 2.286 | 7.429 | 11.43 | 1.143 | 2.286 | 0.571 | 18.29 | 4.571 | 0.571 | 4 | 1.143 | 2.286 | 5.714 | 2.286 | 13.71 | 1.714 | 4.571 |
| 10.29 | 1.714 | 1.714 | 2.857 | 8.571 | 11.43 | 1.143 | 2.286 | 1.714 | 15.43 | 2.857 | 1.143 | 3.429 | 0 | 2.857 | 10.86 | 2.857 | 13.71 | 1.143 | 4 |
| 9.714 | 2.286 | 2.286 | 1.714 | 7.429 | 14.29 | 2.286 | 1.714 | 0.571 | 15.43 | 5.143 | 1.143 | 2.857 | 0.571 | 2.286 | 6.857 | 4 | 12.57 | 2.286 | 4.571 |
| 12 | 1.714 | 2.857 | 2.286 | 8 | 13.14 | 1.143 | 0.571 | 0.571 | 19.43 | 4.571 | 1.143 | 3.429 | 0 | 2.286 | 6.286 | 2.286 | 12 | 2.286 | 4 |
| 13.14 | 2.286 | 1.714 | 2.286 | 9.714 | 12 | 0.571 | 1.143 | 1.143 | 16.57 | 4.571 | 0.571 | 3.429 | 0.571 | 2.286 | 4 | 2.857 | 14.86 | 1.714 | 4.571 |
| 12.57 | 2.286 | 1.714 | 2.286 | 9.714 | 12 | 0.571 | 1.143 | 1.143 | 16.57 | 4.571 | 1.143 | 3.429 | 0.571 | 2.286 | 4 | 2.857 | 14.86 | 1.714 | 4.571 |
| 12.57 | 2.286 | 1.714 | 2.286 | 9.714 | 11.43 | 0.571 | 1.143 | 1.143 | 16 | 5.143 | 0.571 | 3.429 | 0.571 | 2.286 | 4.571 | 2.857 | 15.43 | 1.714 | 4.571 |
| 12.57 | 2.286 | 1.714 | 2.286 | 8.571 | 12 | 0 | 1.714 | 1.143 | 17.14 | 4.571 | 0.571 | 3.429 | 0 | 2.286 | 4.571 | 2.857 | 14.29 | 1.714 | 6.286 |
| 12.57 | 2.286 | 1.714 | 2.286 | 8.571 | 12 | 0 | 1.714 | 1.143 | 17.14 | 4.571 | 0.571 | 3.429 | 0 | 2.286 | 4.571 | 2.857 | 14.29 | 1.714 | 6.286 |
| 12 | 1.714 | 1.714 | 4 | 5.714 | 13.14 | 0.571 | 1.143 | 0.571 | 18.29 | 4 | 0.571 | 2.857 | 0 | 2.286 | 7.429 | 1.714 | 14.29 | 2.286 | 5.714 |
| 12.57 | 1.143 | 1.714 | 4 | 6.286 | 12.57 | 0.571 | 0.571 | 0.571 | 17.71 | 5.143 | 0.571 | 2.857 | 0 | 2.286 | 8 | 1.143 | 14.29 | 2.286 | 5.714 |
| 11.43 | 2.286 | 1.714 | 2.857 | 6.286 | 11.43 | 1.143 | 4 | 1.143 | 17.14 | 6.286 | 1.143 | 2.857 | 0.571 | 2.286 | 7.429 | 2.857 | 10.86 | 2.857 | 3.429 |
| 13.14 | 2.857 | 2.286 | 2.286 | 5.714 | 11.43 | 0.571 | 2.857 | 0.571 | 17.14 | 3.429 | 0.571 | 2.286 | 0.571 | 2.286 | 10.86 | 1.714 | 13.71 | 2.286 | 3.429 |
| 12 | 2.857 | 2.286 | 2.286 | 5.143 | 12.57 | 0.571 | 3.429 | 0.571 | 15.43 | 5.143 | 1.143 | 2.286 | 0 | 2.286 | 9.714 | 1.714 | 13.14 | 2.857 | 4.571 |
| 12.57 | 2.286 | 2.286 | 2.286 | 6.286 | 11.43 | 0.571 | 2.286 | 1.143 | 16.57 | 4.571 | 0.571 | 2.857 | 0.571 | 2.286 | 10.29 | 1.143 | 12 | 2.857 | 5.143 |
| 10.98 | 0.578 | 3.468 | 3.468 | 9.827 | 13.29 | 0 | 4.046 | 1.156 | 14.45 | 2.89 | 0.578 | 2.89 | 0 | 2.312 | 6.936 | 3.468 | 14.45 | 2.312 | 2.89 |
| 10.98 | 2.89 | 2.312 | 4.046 | 8.671 | 10.98 | 0.578 | 2.312 | 1.156 | 12.14 | 5.78 | 0.578 | 3.468 | 0 | 1.734 | 7.514 | 1.156 | 16.76 | 2.312 | 4.624 |
| 9.884 | 2.907 | 1.744 | 3.488 | 7.558 | 12.21 | 0 | 2.907 | 0.581 | 12.79 | 5.233 | 0.581 | 2.907 | 0 | 2.907 | 5.814 | 0 | 20.93 | 1.744 | 5.814 |
| 11.05 | 2.907 | 2.326 | 2.907 | 8.14 | 12.21 | 0 | 2.907 | 0.581 | 12.21 | 5.233 | 0.581 | 2.907 | 0 | 2.907 | 5.814 | 0 | 19.77 | 1.744 | 5.814 |
| 9.827 | 3.468 | 2.312 | 2.89 | 7.514 | 11.56 | 0 | 4.046 | 0.578 | 12.72 | 6.358 | 0.578 | 2.89 | 0 | 2.89 | 6.358 | 0.578 | 18.5 | 1.734 | 5.202 |
| 13.22 | 2.299 | 1.724 | 3.448 | 6.897 | 11.49 | 0.575 | 1.149 | 0.575 | 16.67 | 3.448 | 1.149 | 2.874 | 0 | 1.724 | 7.471 | 0.575 | 16.67 | 2.299 | 5.747 |
| 13.29 | 2.89 | 1.734 | 3.468 | 6.936 | 10.98 | 0.578 | 1.734 | 0.578 | 16.76 | 3.468 | 0.578 | 2.89 | 0 | 1.734 | 8.092 | 1.156 | 15.61 | 1.734 | 5.78 |
| 13.14 | 2.286 | 1.714 | 3.429 | 5.714 | 12.57 | 0.571 | 2.286 | 0.571 | 17.71 | 2.857 | 1.143 | 2.857 | 0 | 1.714 | 7.429 | 1.143 | 14.86 | 2.286 | 5.714 |
| 13.79 | 2.874 | 1.724 | 4.023 | 5.747 | 11.49 | 0.575 | 3.448 | 0.575 | 17.24 | 2.299 | 1.149 | 2.874 | 0 | 1.724 | 5.747 | 2.299 | 14.37 | 2.299 | 5.747 |
| 12.64 | 2.874 | 1.724 | 4.023 | 5.747 | 12.07 | 0.575 | 3.448 | 0.575 | 17.24 | 2.299 | 1.149 | 2.874 | 0 | 1.724 | 6.322 | 1.724 | 14.94 | 2.299 | 5.747 |
| 14.29 | 2.286 | 1.714 | 3.429 | 4.571 | 10.29 | 1.143 | 3.429 | 1.143 | 16 | 2.857 | 0.571 | 2.857 | 0 | 2.286 | 8 | 1.143 | 14.29 | 2.286 | 7.429 |
| 14.86 | 2.286 | 1.143 | 4 | 4.571 | 10.29 | 1.143 | 3.429 | 1.143 | 16 | 2.286 | 0.571 | 2.857 | 0 | 2.286 | 8 | 1.143 | 14.29 | 2.286 | 7.429 |
| 10.34 | 2.874 | 1.149 | 3.448 | 6.897 | 13.79 | 0.575 | 2.299 | 1.149 | 14.37 | 5.172 | 1.724 | 2.299 | 0 | 1.724 | 6.322 | 0 | 17.82 | 2.299 | 5.747 |
| 13.22 | 3.448 | 1.149 | 4.023 | 5.172 | 12.64 | 0.575 | 1.724 | 0.575 | 18.39 | 1.724 | 0.575 | 2.874 | 0 | 1.724 | 6.322 | 0.575 | 17.24 | 2.299 | 5.747 |
| 12.64 | 3.448 | 1.724 | 3.448 | 6.322 | 13.79 | 1.149 | 1.149 | 0.575 | 17.82 | 2.299 | 0.575 | 2.874 | 0 | 1.724 | 6.897 | 0 | 16.67 | 2.299 | 4.598 |
| 11.49 | 2.874 | 1.724 | 4.023 | 5.747 | 12.07 | 0.575 | 3.448 | 0.575 | 17.24 | 2.874 | 0.575 | 2.874 | 0 | 1.724 | 6.897 | 0.575 | 16.67 | 2.299 | 5.747 |
| 12.07 | 2.874 | 1.149 | 4.023 | 5.747 | 10.92 | 0.575 | 2.299 | 0.575 | 17.82 | 2.299 | 1.724 | 2.874 | 0 | 1.724 | 6.897 | 0 | 18.39 | 2.299 | 5.747 |
| 12.64 | 2.874 | 1.149 | 4.023 | 6.322 | 10.92 | 0.575 | 2.874 | 0.575 | 17.24 | 2.299 | 1.149 | 2.874 | 0 | 1.724 | 6.897 | 0 | 17.82 | 2.299 | 5.747 |
| 11.49 | 2.874 | 1.149 | 4.023 | 6.322 | 11.49 | 0.575 | 2.874 | 0.575 | 17.24 | 2.299 | 1.149 | 2.874 | 0 | 1.724 | 6.897 | 0 | 18.39 | 2.299 | 5.747 |
| 10.98 | 1.156 | 0.578 | 3.468 | 6.358 | 14.45 | 0.578 | 4.046 | 0.578 | 13.87 | 6.936 | 0.578 | 2.312 | 0 | 2.312 | 6.936 | 3.468 | 12.72 | 3.468 | 5.202 |

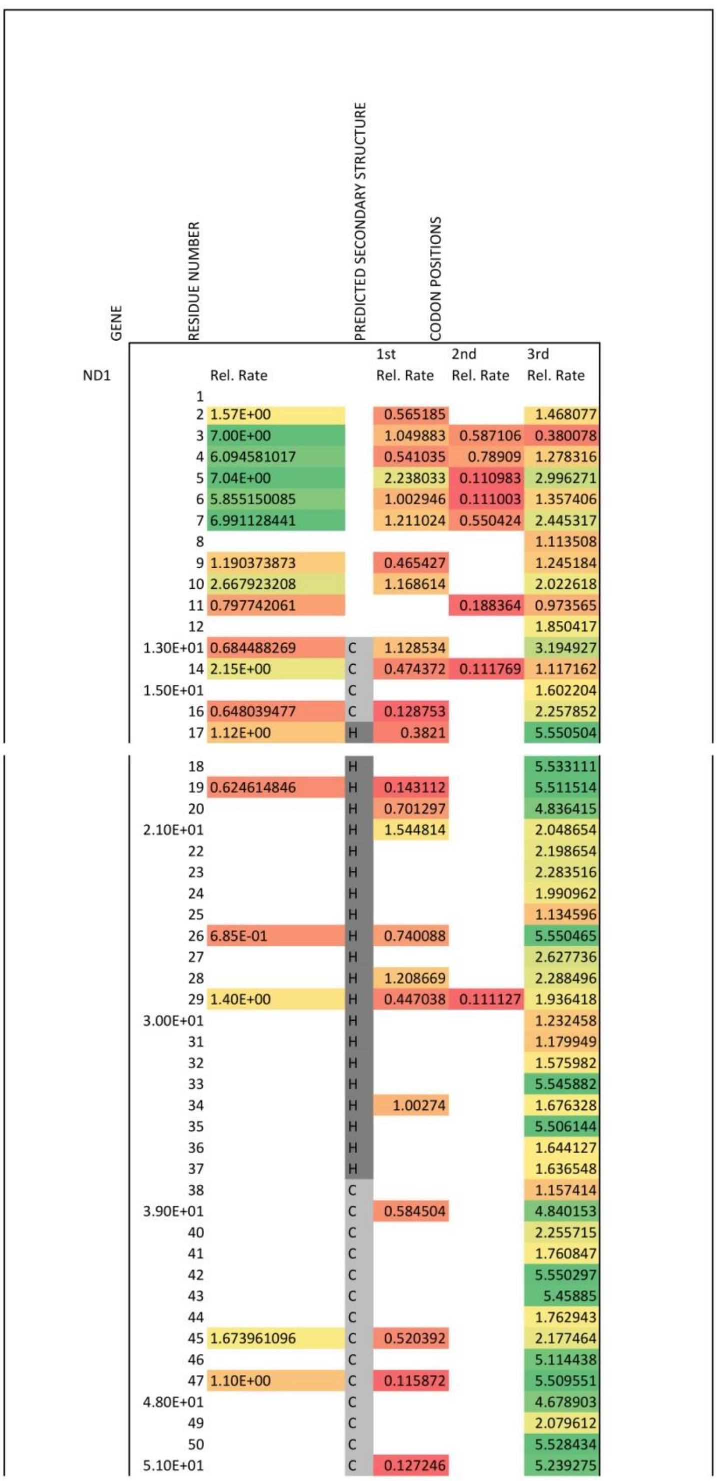
nd6



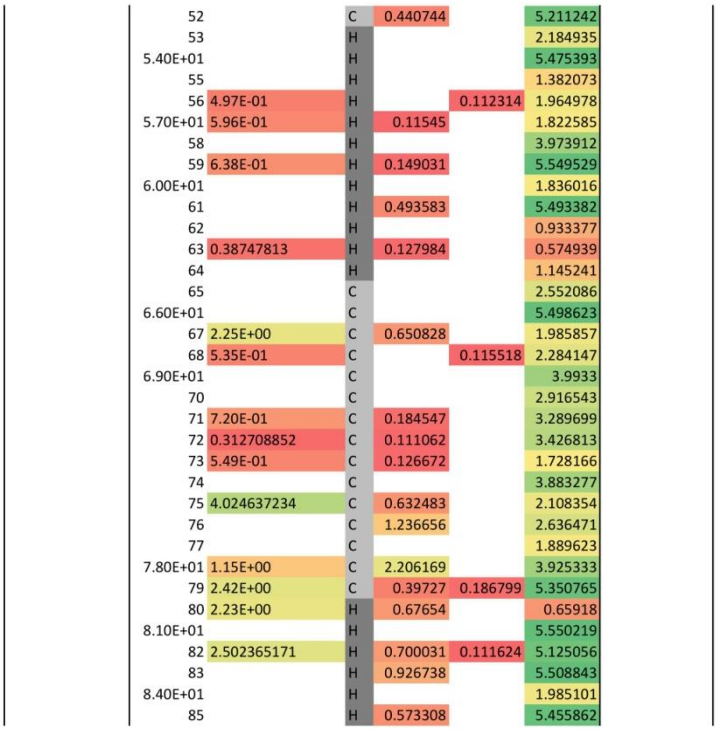
| Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9.1 | 0.8 | 2.1 | 2.6 | 6.1 | 6.7 | 2.9 | 7.1 | 2 | 16.4 | 4.3 | 2.6 | 5.7 | 2.6 | 2.1 | 6.8 | 7.7 | 6.5 | 3.1 | 2.9 |
| 9.3 | 0.8 | 1.9 | 2.6 | 6.3 | 6.7 | 2.9 | 7 | 2 | 16.4 | 4.4 | 2.8 | 5.9 | 2.5 | 2.1 | 6.8 | 7.4 | 6.3 | 3.1 | 2.9 |
| 9.2 | 0.8 | 2 | 2.7 | 6.1 | 6.6 | 3 | 6.6 | 1.9 | 16.6 | 4.4 | 2.8 | 5.7 | 2.5 | 2.1 | 6.8 | 7.9 | 6.4 | 3.1 | 2.9 |
| 8.8 | 0.8 | 1.9 | 2.6 | 6.1 | 6.5 | 2.8 | 7.7 | 2 | 16.4 | 4.6 | 3 | 5.7 | 2.5 | 2 | 6.8 | 8.2 | 5.3 | 3.1 | 2.9 |
| 9.8 | 0.8 | 2 | 2.7 | 6.2 | 6.4 | 2.7 | 7 | 2 | 16.4 | 4.3 | 3 | 5.7 | 2.5 | 2 | 6 | 8.3 | 6.1 | 3.1 | 3 |
| 9.7 | 0.8 | 2 | 2.7 | 6.2 | 6.5 | 2.8 | 7.1 | 2 | 16.3 | 4.4 | 3 | 5.8 | 2.5 | 2 | 6 | 8.2 | 6.1 | 3.1 | 3 |
| 9.4 | 0.8 | 2.1 | 2.7 | 6.2 | 6.5 | 2.7 | 7 | 2 | 16.3 | 4.2 | 2.9 | 5.7 | 2.5 | 2 | 6.1 | 8.4 | 6.3 | 3.1 | 3.1 |
| 9.6 | 0.8 | 2 | 2.7 | 6.2 | 6.5 | 2.7 | 7 | 2 | 16.3 | 4.2 | 3 | 5.7 | 2.5 | 2 | 6.1 | 8.2 | 6.3 | 3.1 | 3.1 |
| 9.7 | 0.8 | 2 | 2.7 | 6.2 | 6.6 | 2.8 | 6.9 | 2 | 16.3 | 4.2 | 2.9 | 5.7 | 2.5 | 2 | 6.1 | 8.3 | 6.3 | 3.1 | 3.1 |
| 9.3 | 0.8 | 2 | 2.7 | 6.2 | 6.5 | 2.7 | 7.1 | 2 | 16.2 | 4.3 | 3.1 | 5.7 | 2.5 | 2 | 6.1 | 8.5 | 6.2 | 3.1 | 3 |
| 9.6 | 0.8 | 2 | 2.7 | 6.2 | 6.5 | 2.7 | 7.2 | 2 | 16.2 | 4.3 | 2.9 | 5.7 | 2.5 | 2 | 6.1 | 8.3 | 6.1 | 3.1 | 3 |
| 9.4 | 0.7 | 2.1 | 2.6 | 6.4 | 6.5 | 2.8 | 7 | 2 | 16.2 | 4.2 | 2.9 | 5.7 | 2.4 | 2 | 6.1 | 8.2 | 6.4 | 3.1 | 3 |
| 9.5 | 0.7 | 2.1 | 2.7 | 6.2 | 6.5 | 2.8 | 6.9 | 2 | 16.4 | 4.2 | 2.9 | 5.7 | 2.4 | 2 | 6.1 | 8.3 | 6.5 | 3.1 | 3 |
| 9.2 | 0.8 | 2.1 | 2.7 | 6.3 | 6.5 | 2.7 | 7.2 | 2 | 16.2 | 4.2 | 2.8 | 5.7 | 2.5 | 2.1 | 6.3 | 8.4 | 6.3 | 3.1 | 2.9 |
| 9.7 | 0.8 | 2 | 2.6 | 6.2 | 6.6 | 2.7 | 6.9 | 2 | 16.3 | 4.1 | 2.9 | 5.7 | 2.5 | 2 | 6.2 | 8.2 | 6.5 | 3.1 | 3 |
| 9.6 | 0.8 | 2 | 2.7 | 6.3 | 6.5 | 2.7 | 6.9 | 2 | 16.3 | 4 | 2.9 | 5.7 | 2.4 | 2 | 6.2 | 8.2 | 6.6 | 3.1 | 3 |
| 9.6 | 0.8 | 2 | 2.7 | 6.3 | 6.5 | 2.8 | 7 | 2 | 16.3 | 4.1 | 2.9 | 5.7 | 2.4 | 2 | 6.2 | 8.2 | 6.5 | 3.1 | 3 |
| 9.7 | 0.8 | 2 | 2.7 | 6.2 | 6.5 | 2.7 | 7.1 | 2.1 | 16.1 | 3.9 | 2.9 | 5.7 | 2.5 | 2 | 6.3 | 7.7 | 6.7 | 3.2 | 3.1 |
| 9.6 | 0.9 | 2 | 2.7 | 6.2 | 6.5 | 2.7 | 6.8 | 2 | 16.4 | 4.1 | 2.9 | 5.7 | 2.4 | 2 | 6.2 | 8 | 6.7 | 3.1 | 3 |
| 9.5 | 0.7 | 2 | 2.7 | 6 | 6.5 | 2.8 | 6.9 | 2 | 16.5 | 4 | 2.9 | 5.8 | 2.5 | 2 | 6.2 | 8.1 | 6.7 | 3.1 | 3 |
| 9.7 | 0.7 | 2.1 | 2.7 | 6.1 | 6.4 | 2.7 | 6.9 | 2 | 16.4 | 4.2 | 2.9 | 5.7 | 2.5 | 2.1 | 6.1 | 8.2 | 6.4 | 3.1 | 3 |
| 9.8 | 0.7 | 2.1 | 2.7 | 6.1 | 6.5 | 2.8 | 7 | 2 | 16.5 | 4.1 | 2.8 | 5.7 | 2.5 | 2.1 | 6.2 | 8.1 | 6.4 | 3.1 | 3 |
| 9.9 | 0.9 | 2 | 2.6 | 6.3 | 6.6 | 2.7 | 6.8 | 2 | 16.3 | 4.1 | 2.9 | 5.8 | 2.6 | 2 | 6 | 7.7 | 6.7 | 3.1 | 3 |
| 9.8 | 0.9 | 2.1 | 2.7 | 6.2 | 6.6 | 2.7 | 6.9 | 2 | 15.8 | 4.2 | 2.7 | 5.7 | 2.5 | 2 | 6.4 | 7.5 | 7.4 | 3.1 | 2.9 |
| 9.7 | 0.8 | 1.9 | 2.6 | 6.2 | 6.4 | 2.7 | 6.8 | 2 | 16.1 | 4.3 | 2.9 | 5.8 | 2.6 | 2 | 6.3 | 7.9 | 6.9 | 3.1 | 2.9 |
| 9.7 | 0.8 | 2.1 | 2.6 | 6.2 | 6.6 | 2.7 | 6.8 | 2 | 16.1 | 4.3 | 2.9 | 5.8 | 2.5 | 2 | 6.4 | 7.9 | 6.6 | 3.2 | 3.1 |
| 9.7 | 0.8 | 2 | 2.6 | 6.2 | 6.6 | 2.7 | 6.8 | 2 | 16.1 | 4.3 | 2.9 | 5.8 | 2.5 | 2 | 6.4 | 7.9 | 6.5 | 3.2 | 3 |
| 9.6 | 0.8 | 2.1 | 2.6 | 6.2 | 6.5 | 2.7 | 6.9 | 2 | 16.1 | 4.3 | 2.9 | 5.8 | 2.5 | 2 | 6.4 | 7.9 | 6.4 | 3.2 | 3.1 |
| 9.4 | 0.8 | 2.1 | 2.7 | 6.6 | 6.5 | 2.8 | 7.1 | 2 | 16.1 | 4.1 | 3.2 | 5.7 | 2.5 | 2 | 6.4 | 7.9 | 6.2 | 3.1 | 2.9 |
| 9.4 | 0.8 | 2 | 2.6 | 6.6 | 6.5 | 2.7 | 7.3 | 2 | 16 | 4.1 | 3.1 | 5.7 | 2.6 | 2 | 6.6 | 7.8 | 6 | 3.1 | 2.9 |
| 9.5 | 0.8 | 2 | 2.7 | 6.5 | 6.5 | 2.7 | 7.3 | 2 | 16.2 | 4.2 | 3.1 | 5.6 | 2.5 | 2 | 6.5 | 7.9 | 5.9 | 3.1 | 2.9 |
| 8.9 | 0.8 | 2 | 2.6 | 6.4 | 6.4 | 2.6 | 7.8 | 2 | 15.9 | 4.5 | 3.3 | 5.7 | 2.6 | 1.9 | 6.1 | 8.5 | 5.8 | 3.1 | 3 |
| 8.8 | 0.8 | 2 | 2.6 | 6.3 | 6.4 | 2.6 | 7.6 | 2.1 | 16.1 | 4.2 | 3.3 | 5.6 | 2.7 | 1.9 | 6.2 | 8.6 | 6.1 | 3.1 | 2.9 |
| 9.4 | 0.9 | 1.9 | 2.7 | 6.6 | 6.4 | 2.6 | 7.1 | 2.1 | 16.1 | 4.1 | 3.1 | 5.6 | 2.6 | 2 | 6.3 | 8.2 | 6.3 | 3.1 | 2.9 |
| 9.3 | 0.9 | 2 | 2.6 | 6.7 | 6.3 | 2.7 | 7.1 | 2.1 | 16 | 4.3 | 3.2 | 5.6 | 2.5 | 2 | 6.4 | 8.3 | 6.1 | 3.1 | 3 |
| 9 | 0.9 | 1.9 | 2.6 | 6.7 | 6.3 | 2.7 | 7.4 | 2.1 | 15.9 | 4.6 | 3.3 | 5.6 | 2.5 | 2 | 6.2 | 8.5 | 5.7 | 3.1 | 2.9 |
| 9.5 | 0.8 | 2 | 2.7 | 6.3 | 6.3 | 2.7 | 7.6 | 2.1 | 16 | 4.1 | 3.2 | 5.6 | 2.5 | 2.1 | 6.2 | 8.2 | 6.1 | 3.1 | 3 |
| 9.6 | 0.8 | 2.1 | 2.6 | 6.3 | 6.6 | 2.7 | 7.2 | 2 | 15.7 | 4.3 | 3.1 | 5.7 | 2.6 | 2 | 5.9 | 8.2 | 6.4 | 3.2 | 3 |
| 9.8 | 0.8 | 2 | 2.6 | 6.2 | 6.6 | 2.8 | 6.9 | 2 | 16.5 | 4.1 | 2.9 | 5.7 | 2.5 | 2 | 6.2 | 7.7 | 6.6 | 3.2 | 3 |
| 9.8 | 0.8 | 1.9 | 2.5 | 6.3 | 6.3 | 2.7 | 6.7 | 2 | 16.1 | 4.5 | 3.1 | 5.8 | 2.6 | 2 | 6.1 | 7.9 | 6.7 | 3.1 | 3 |
| 9.7 | 0.8 | 1.9 | 2.5 | 6.3 | 6.3 | 2.7 | 6.7 | 2 | 16.1 | 4.5 | 3.2 | 5.8 | 2.6 | 2 | 6.1 | 8 | 6.7 | 3.1 | 3 |
| 9.8 | 0.8 | 1.9 | 2.6 | 6.3 | 6.3 | 2.7 | 6.5 | 2 | 16 | 4.5 | 3.1 | 5.7 | 2.6 | 2 | 6.1 | 8 | 6.9 | 3.1 | 3 |
| 9.8 | 0.8 | 2 | 2.5 | 6.3 | 6.5 | 2.7 | 6.8 | 2 | 16 | 4.4 | 3 | 5.8 | 2.5 | 2 | 6.2 | 7.9 | 6.6 | 3.1 | 3.1 |
| 9.8 | 0.8 | 2 | 2.5 | 6.3 | 6.5 | 2.7 | 6.8 | 2 | 16.1 | 4.4 | 3 | 5.8 | 2.5 | 2 | 6.2 | 7.8 | 6.6 | 3.1 | 3.1 |
| 9.7 | 0.7 | 1.9 | 2.7 | 6 | 6.5 | 2.7 | 6.9 | 2 | 16.3 | 4.4 | 3.1 | 5.7 | 2.6 | 2 | 6 | 8.3 | 6.2 | 3.2 | 3.1 |
| 9.9 | 0.7 | 2 | 2.7 | 6 | 6.5 | 2.7 | 6.7 | 2 | 16.5 | 4.3 | 3.1 | 5.7 | 2.6 | 2 | 5.9 | 8.1 | 6.5 | 3.2 | 3.1 |
| 9.4 | 0.8 | 1.9 | 2.6 | 6.1 | 6.4 | 2.8 | 7.1 | 2.1 | 16.1 | 4.6 | 3.1 | 5.7 | 2.6 | 2 | 6 | 8.3 | 6.3 | 3.2 | 3.1 |
| 10.1 | 0.8 | 2.1 | 2.5 | 6.1 | 6.5 | 2.8 | 6.4 | 2 | 16.6 | 4.1 | 2.7 | 5.7 | 2.4 | 2.1 | 6.8 | 7.3 | 7.1 | 3.2 | 2.8 |
| 9.9 | 0.8 | 2.1 | 2.5 | 6.1 | 6.6 | 2.8 | 6.2 | 2 | 16.7 | 4.1 | 2.8 | 5.6 | 2.5 | 2.1 | 6.7 | 7.2 | 7.2 | 3.2 | 3 |
| 9.7 | 0.8 | 2.1 | 2.6 | 6.1 | 6.6 | 2.8 | 6.5 | 1.9 | 16.3 | 4.4 | 2.8 | 5.7 | 2.4 | 2.1 | 6.6 | 7.6 | 7.1 | 3.2 | 3 |
| 9.3 | 0.7 | 2 | 2.6 | 6.4 | 6.6 | 2.7 | 7.1 | 2 | 16.2 | 4.3 | 3 | 5.7 | 2.6 | 2 | 6.4 | 7.8 | 6.6 | 3.1 | 3 |
| 8.5 | 0.8 | 2 | 2.6 | 5.9 | 6.3 | 2.9 | 7.7 | 2 | 16.6 | 4.5 | 3.3 | 5.7 | 2.6 | 1.9 | 6 | 9.4 | 5.2 | 3.1 | 3 |
| 8.7 | 0.7 | 2 | 2.6 | 5.8 | 6.3 | 2.8 | 7.3 | 2 | 16.2 | 4.7 | 3.2 | 5.8 | 2.7 | 1.9 | 6.1 | 9.4 | 5.5 | 3.1 | 3 |
| 8.5 | 0.7 | 2 | 2.6 | 5.9 | 6.4 | 2.9 | 7.5 | 2 | 16.3 | 4.5 | 3.3 | 5.8 | 2.6 | 1.9 | 6.1 | 9.2 | 5.8 | 3.1 | 3 |
| 8.8 | 0.8 | 1.9 | 2.6 | 5.9 | 6.3 | 2.9 | 7.5 | 2 | 16.1 | 4.7 | 3.4 | 5.7 | 2.6 | 1.9 | 5.8 | 9.3 | 5.5 | 3.1 | 3 |
| 9.7 | 0.8 | 1.9 | 2.7 | 6.1 | 6.3 | 2.7 | 7.6 | 2 | 16.1 | 4.2 | 3.3 | 5.7 | 2.7 | 1.9 | 5.7 | 8.4 | 6.1 | 3.1 | 2.9 |
| 9.6 | 0.8 | 1.9 | 2.7 | 6.1 | 6.3 | 2.8 | 7.5 | 2 | 16.3 | 4.3 | 3.2 | 5.6 | 2.6 | 1.9 | 6 | 8.5 | 5.9 | 3.1 | 2.9 |
| 9.6 | 0.8 | 2 | 2.7 | 6.1 | 6.4 | 2.7 | 7.2 | 2 | 16.1 | 4.3 | 3.2 | 5.7 | 2.7 | 1.9 | 5.9 | 8.4 | 6.4 | 3.1 | 2.9 |
| 9.6 | 0.8 | 1.9 | 2.7 | 6.1 | 6.4 | 2.8 | 7.3 | 2 | 16 | 4.2 | 3 | 5.7 | 2.6 | 1.9 | 6 | 8.5 | 6.5 | 3.1 | 2.9 |
| 9.6 | 0.8 | 1.9 | 2.7 | 6.1 | 6.4 | 2.7 | 7.2 | 2 | 16 | 4.2 | 3.1 | 5.7 | 2.6 | 1.9 | 6 | 8.5 | 6.5 | 3.1 | 2.9 |
| 9.6 | 0.8 | 2 | 2.7 | 6 | 6.2 | 2.7 | 7.5 | 2.1 | 16.1 | 4.1 | 3.2 | 5.7 | 2.6 | 1.9 | 6.2 | 8.2 | 6.3 | 3.2 | 3.1 |
| 9.7 | 0.8 | 2 | 2.7 | 6 | 6.2 | 2.7 | 7.5 | 2.1 | 16.1 | 4.1 | 3.2 | 5.6 | 2.6 | 1.9 | 6.2 | 8.2 | 6.4 | 3.2 | 3.1 |
| 9 | 0.8 | 1.9 | 2.6 | 6.2 | 6.3 | 2.7 | 8 | 2 | 16.1 | 4.5 | 3.4 | 5.7 | 2.6 | 1.9 | 6.2 | 8.8 | 5.2 | 3.2 | 2.9 |
| 9.1 | 0.8 | 1.9 | 2.7 | 6 | 6.2 | 2.7 | 7.5 | 2 | 16.5 | 4.1 | 3.4 | 5.8 | 2.6 | 1.9 | 6 | 9 | 5.7 | 3.2 | 3 |
| 9.4 | 0.8 | 2 | 2.7 | 6 | 6.4 | 2.8 | 7.5 | 2.1 | 16.2 | 4.4 | 3.4 | 5.7 | 2.6 | 1.9 | 5.8 | 8.5 | 5.8 | 3.2 | 2.9 |
| 9.4 | 0.8 | 1.9 | 2.7 | 6.1 | 6.3 | 2.7 | 7.3 | 2 | 16.1 | 4.3 | 3.2 | 5.7 | 2.6 | 1.9 | 6.2 | 8.2 | 6.4 | 3.2 | 3 |
| 9.2 | 0.8 | 1.9 | 2.7 | 6.1 | 6.3 | 2.7 | 7.6 | 2.1 | 16 | 4.3 | 3.4 | 5.6 | 2.6 | 1.9 | 6.1 | 8.5 | 6 | 3.2 | 3 |
| 9.1 | 0.8 | 1.9 | 2.7 | 6.1 | 6.3 | 2.7 | 7.6 | 2.1 | 16 | 4.4 | 3.3 | 5.6 | 2.6 | 1.9 | 6.1 | 8.6 | 6 | 3.2 | 3 |
| 9 | 0.8 | 1.9 | 2.7 | 6.1 | 6.3 | 2.7 | 7.6 | 2.1 | 16 | 4.4 | 3.3 | 5.7 | 2.6 | 1.9 | 6.1 | 8.6 | 6 | 3.2 | 3 |
| 8.3 | 0.7 | 1.9 | 2.5 | 6.2 | 6.4 | 2.8 | 8 | 2.1 | 16.3 | 5 | 3.4 | 5.6 | 2.7 | 1.9 | 6.3 | 8.6 | 5.2 | 3.2 | 2.9 |

All gen concatenated

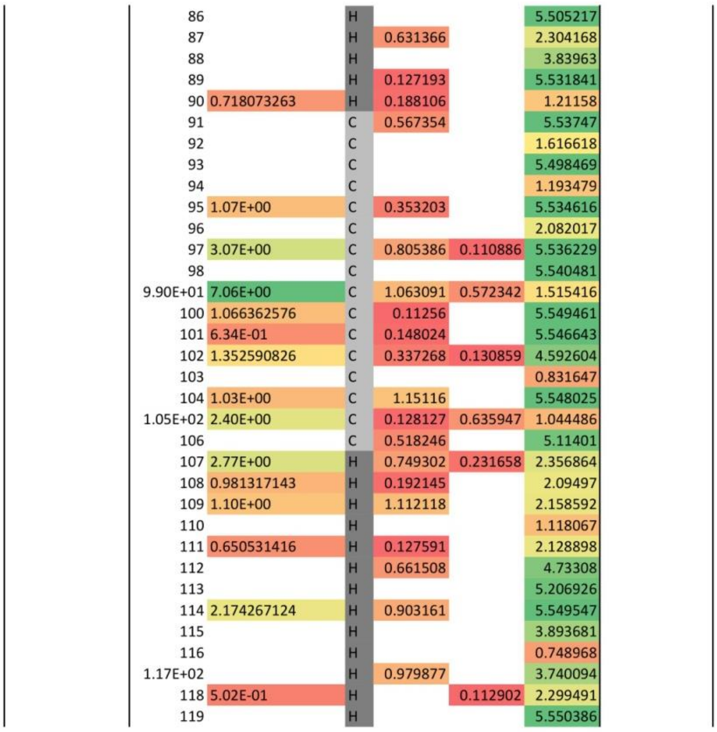
**Fig. A3(a)** Relative rate of evolution on codon position 1, 2 and 3 of Clupeoids ND1 gene.



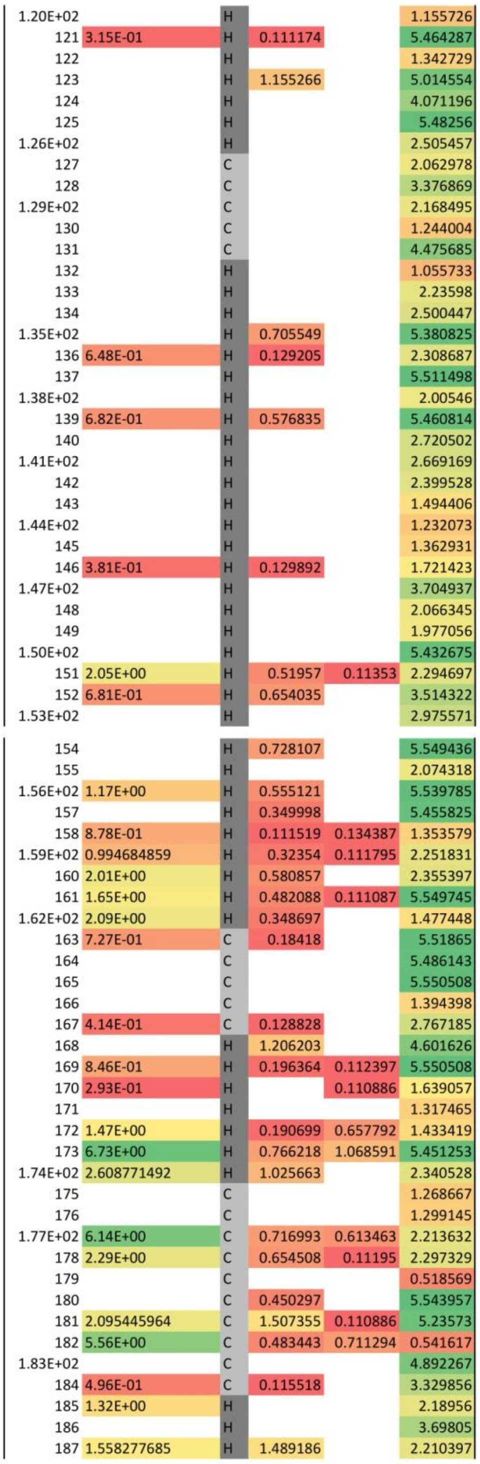
| GENE | RESIDUE NUMBER | | PREDICTED SECONDARY STRUCTURE | CODON POSITIONS | | |
|---|---|---|---|---|---|---|
| | | | | 1st | 2nd | 3rd |
| ND1 | | Rel. Rate | | Rel. Rate | Rel. Rate | Rel. Rate |
| | 1 | | | | | |
| | 2 | 1.57E+00 | | 0.565185 | | 1.468077 |
| | 3 | 7.00E+00 | | 1.049883 | 0.587106 | 0.380078 |
| | 4 | 6.094581017 | | 0.541035 | 0.78909 | 1.278316 |
| | 5 | 7.04E+00 | | 2.238033 | 0.110983 | 2.996271 |
| | 6 | 5.855150085 | | 1.002946 | 0.111003 | 1.357406 |
| | 7 | 6.991128441 | | 1.211024 | 0.550424 | 2.445317 |
| | 8 | | | | | 1.113508 |
| | 9 | 1.190373873 | | 0.465427 | | 1.245184 |
| | 10 | 2.667923208 | | 1.168614 | | 2.022618 |
| | 11 | 0.797742061 | | | 0.188364 | 0.973565 |
| | 12 | | | | | 1.850417 |
| | 1.30E+01 | 0.684488269 | C | 1.128534 | | 3.194927 |
| | 14 | 2.15E+00 | C | 0.474372 | 0.111769 | 1.117162 |
| | 1.50E+01 | | C | | | 1.602204 |
| | 16 | 0.648039477 | C | 0.128753 | | 2.257852 |
| | 17 | 1.12E+00 | H | 0.3821 | | 5.550504 |
| | 18 | | H | | | 5.533111 |
| | 19 | 0.624614846 | H | 0.143112 | | 5.511514 |
| | 20 | | H | 0.701297 | | 4.836415 |
| | 2.10E+01 | | H | 1.544814 | | 2.048654 |
| | 22 | | H | | | 2.198654 |
| | 23 | | H | | | 2.283516 |
| | 24 | | H | | | 1.990962 |
| | 25 | | H | | | 1.134596 |
| | 26 | 6.85E-01 | H | 0.740088 | | 5.550465 |
| | 27 | | H | | | 2.627736 |
| | 28 | | H | 1.208669 | | 2.288496 |
| | 29 | 1.40E+00 | H | 0.447038 | 0.111127 | 1.936418 |
| | 3.00E+01 | | H | | | 1.232458 |
| | 31 | | H | | | 1.179949 |
| | 32 | | H | | | 1.575982 |
| | 33 | | H | | | 5.545882 |
| | 34 | | H | 1.00274 | | 1.676328 |
| | 35 | | H | | | 5.506144 |
| | 36 | | H | | | 1.644127 |
| | 37 | | H | | | 1.636548 |
| | 38 | | C | | | 1.157414 |
| | 3.90E+01 | | C | 0.584504 | | 4.840153 |
| | 40 | | C | | | 2.255715 |
| | 41 | | C | | | 1.760847 |
| | 42 | | C | | | 5.550297 |
| | 43 | | C | | | 5.45885 |
| | 44 | | C | | | 1.762943 |
| | 45 | 1.673961096 | C | 0.520392 | | 2.177464 |
| | 46 | | C | | | 5.114438 |
| | 47 | 1.10E+00 | C | 0.115872 | | 5.509551 |
| | 4.80E+01 | | C | | | 4.678903 |
| | 49 | | C | | | 2.079612 |
| | 50 | | C | | | 5.528434 |
| | 5.10E+01 | | C | 0.127246 | | 5.239275 |



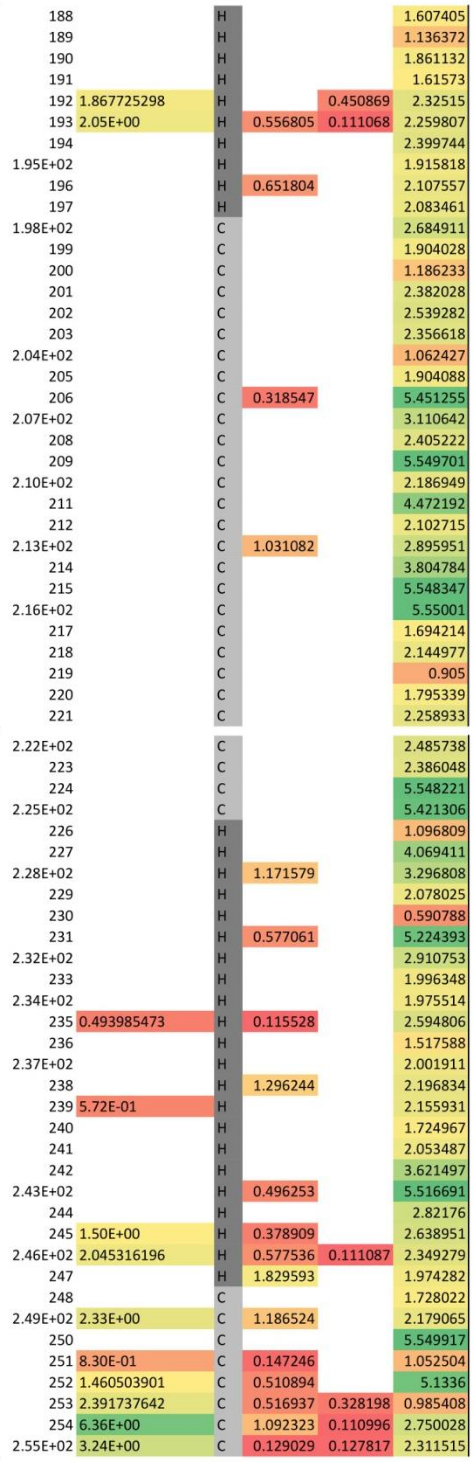
| | | | | | |
|---|---|---|---|---|---|
| 52 | | C | 0.440744 | | 5.211242 |
| 53 | | H | | | 2.184935 |
| 5.40E+01 | | H | | | 5.475393 |
| 55 | | H | | | 1.382073 |
| 56 | 4.97E-01 | H | | 0.112314 | 1.964978 |
| 5.70E+01 | 5.96E-01 | H | 0.11545 | | 1.822585 |
| 58 | | H | | | 3.973912 |
| 59 | 6.38E-01 | H | 0.149031 | | 5.549529 |
| 6.00E+01 | | H | | | 1.836016 |
| 61 | | H | 0.493583 | | 5.493382 |
| 62 | | H | | | 0.933377 |
| 63 | 0.38747813 | H | 0.127984 | | 0.574939 |
| 64 | | H | | | 1.145241 |
| 65 | | C | | | 2.552086 |
| 6.60E+01 | | C | | | 5.498623 |
| 67 | 2.25E+00 | C | 0.650828 | | 1.985857 |
| 68 | 5.35E-01 | C | | 0.115518 | 2.284147 |
| 6.90E+01 | | C | | | 3.9933 |
| 70 | | C | | | 2.916543 |
| 71 | 7.20E-01 | C | 0.184547 | | 3.289699 |
| 72 | 0.312708852 | C | 0.111062 | | 3.426813 |
| 73 | 5.49E-01 | C | 0.126672 | | 1.728166 |
| 74 | | C | | | 3.883277 |
| 75 | 4.024637234 | C | 0.632483 | | 2.108354 |
| 76 | | C | 1.236656 | | 2.636471 |
| 77 | | C | | | 1.889623 |
| 7.80E+01 | 1.15E+00 | C | 2.206169 | | 3.925333 |
| 79 | 2.42E+00 | C | 0.39727 | 0.186799 | 5.350765 |
| 80 | 2.23E+00 | H | 0.67654 | | 0.65918 |
| 8.10E+01 | | H | | | 5.550219 |
| 82 | 2.502365171 | H | 0.700031 | 0.111624 | 5.125056 |
| 83 | | H | 0.926738 | | 5.508843 |
| 8.40E+01 | | H | | | 1.985101 |
| 85 | | H | 0.573308 | | 5.455862 |



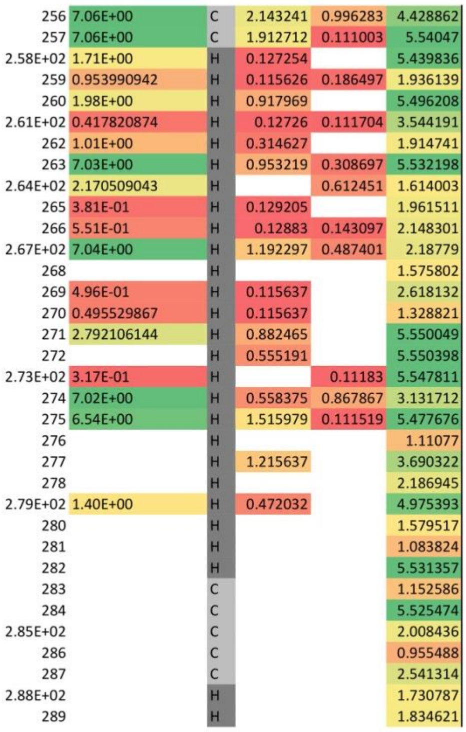
| | | | | | |
|---|---|---|---|---|---|
| 86 | | H | | | 5.505217 |
| 87 | | H | 0.631366 | | 2.304168 |
| 88 | | H | | | 3.83963 |
| 89 | | H | 0.127193 | | 5.531841 |
| 90 | 0.718073263 | H | 0.188106 | | 1.21158 |
| 91 | | C | 0.567354 | | 5.53747 |
| 92 | | C | | | 1.616618 |
| 93 | | C | | | 5.498469 |
| 94 | | C | | | 1.193479 |
| 95 | 1.07E+00 | C | 0.353203 | | 5.534616 |
| 96 | | C | | | 2.082017 |
| 97 | 3.07E+00 | C | 0.805386 | 0.110886 | 5.536229 |
| 98 | | C | | | 5.540481 |
| 9.90E+01 | 7.06E+00 | C | 1.063091 | 0.572342 | 1.515416 |
| 100 | 1.066362576 | C | 0.11256 | | 5.549461 |
| 101 | 6.34E-01 | C | 0.148024 | | 5.546643 |
| 102 | 1.352590826 | C | 0.337268 | 0.130859 | 4.592604 |
| 103 | | C | | | 0.831647 |
| 104 | 1.03E+00 | C | 1.15116 | | 5.548025 |
| 1.05E+02 | 2.40E+00 | C | 0.128127 | 0.635947 | 1.044486 |
| 106 | | C | 0.518246 | | 5.11401 |
| 107 | 2.77E+00 | H | 0.749302 | 0.231658 | 2.356864 |
| 108 | 0.981317143 | H | 0.192145 | | 2.09497 |
| 109 | 1.10E+00 | H | 1.112118 | | 2.158592 |
| 110 | | H | | | 1.118067 |
| 111 | 0.650531416 | H | 0.127591 | | 2.128898 |
| 112 | | H | 0.661508 | | 4.73308 |
| 113 | | H | | | 5.206926 |
| 114 | 2.174267124 | H | 0.903161 | | 5.549547 |
| 115 | | H | | | 3.893681 |
| 116 | | H | | | 0.748968 |
| 1.17E+02 | | H | 0.979877 | | 3.740094 |
| 118 | 5.02E-01 | H | | 0.112902 | 2.299491 |
| 119 | | H | | | 5.550386 |



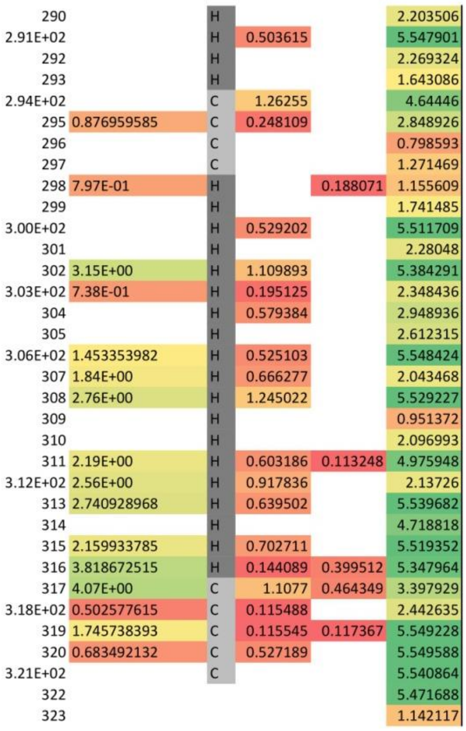
| | | | | | |
|---|---|---|---|---|---|
| 1.20E+02 | | H | | | 1.155726 |
| 121 | 3.15E-01 | H | 0.111174 | | 5.464287 |
| 122 | | H | | | 1.342729 |
| 123 | | H | 1.155266 | | 5.014554 |
| 124 | | H | | | 4.071196 |
| 125 | | H | | | 5.48256 |
| 1.26E+02 | | H | | | 2.505457 |
| 127 | | C | | | 2.062978 |
| 128 | | C | | | 3.376869 |
| 1.29E+02 | | C | | | 2.168495 |
| 130 | | C | | | 1.244004 |
| 131 | | C | | | 4.475685 |
| 132 | | H | | | 1.055733 |
| 133 | | H | | | 2.23598 |
| 134 | | H | | | 2.500447 |
| 1.35E+02 | | H | 0.705549 | | 5.380825 |
| 136 | 6.48E-01 | H | 0.129205 | | 2.308687 |
| 137 | | H | | | 5.511498 |
| 1.38E+02 | | H | | | 2.00546 |
| 139 | 6.82E-01 | H | 0.576835 | | 5.460814 |
| 140 | | H | | | 2.720502 |
| 1.41E+02 | | H | | | 2.669169 |
| 142 | | H | | | 2.399528 |
| 143 | | H | | | 1.494406 |
| 1.44E+02 | | H | | | 1.232073 |
| 145 | | H | | | 1.362931 |
| 146 | 3.81E-01 | H | 0.129892 | | 1.721423 |
| 1.47E+02 | | H | | | 3.704937 |
| 148 | | H | | | 2.066345 |
| 149 | | H | | | 1.977056 |
| 1.50E+02 | | H | | | 5.432675 |
| 151 | 2.05E+00 | H | 0.51957 | 0.11353 | 2.294697 |
| 152 | 6.81E-01 | H | 0.654035 | | 3.514322 |
| 1.53E+02 | | H | | | 2.975571 |
| 154 | | H | 0.728107 | | 5.549436 |
| 155 | | H | | | 2.074318 |
| 1.56E+02 | 1.17E+00 | H | 0.555121 | | 5.539785 |
| 157 | | H | 0.349998 | | 5.455825 |
| 158 | 8.78E-01 | H | 0.111519 | 0.134387 | 1.353579 |
| 1.59E+02 | 0.994684859 | H | 0.32354 | 0.111795 | 2.251831 |
| 160 | 2.01E+00 | H | 0.580857 | | 2.355397 |
| 161 | 1.65E+00 | H | 0.482088 | 0.111087 | 5.549745 |
| 1.62E+02 | 2.09E+00 | H | 0.348697 | | 1.477448 |
| 163 | 7.27E-01 | C | 0.18418 | | 5.51865 |
| 164 | | C | | | 5.486143 |
| 165 | | C | | | 5.550508 |
| 166 | | C | | | 1.394398 |
| 167 | 4.14E-01 | C | 0.128828 | | 2.767185 |
| 168 | | H | 1.206203 | | 4.601626 |
| 169 | 8.46E-01 | H | 0.196364 | 0.112397 | 5.550508 |
| 170 | 2.93E-01 | H | | 0.110886 | 1.639057 |
| 171 | | H | | | 1.317465 |
| 172 | 1.47E+00 | H | 0.190699 | 0.657792 | 1.433419 |
| 173 | 6.73E+00 | H | 0.766218 | 1.068591 | 5.451253 |
| 1.74E+02 | 2.608771492 | H | 1.025663 | | 2.340528 |
| 175 | | C | | | 1.268667 |
| 176 | | C | | | 1.299145 |
| 1.77E+02 | 6.14E+00 | C | 0.716993 | 0.613463 | 2.213632 |
| 178 | 2.29E+00 | C | 0.654508 | 0.11195 | 2.297329 |
| 179 | | C | | | 0.518569 |
| 180 | | C | 0.450297 | | 5.543957 |
| 181 | 2.095445964 | C | 1.507355 | 0.110886 | 5.23573 |
| 182 | 5.56E+00 | C | 0.483443 | 0.711294 | 0.541617 |
| 1.83E+02 | | C | | | 4.892267 |
| 184 | 4.96E-01 | C | 0.115518 | | 3.329856 |
| 185 | 1.32E+00 | H | | | 2.18956 |
| 186 | | H | | | 3.69805 |
| 187 | 1.558277685 | H | 1.489186 | | 2.210397 |



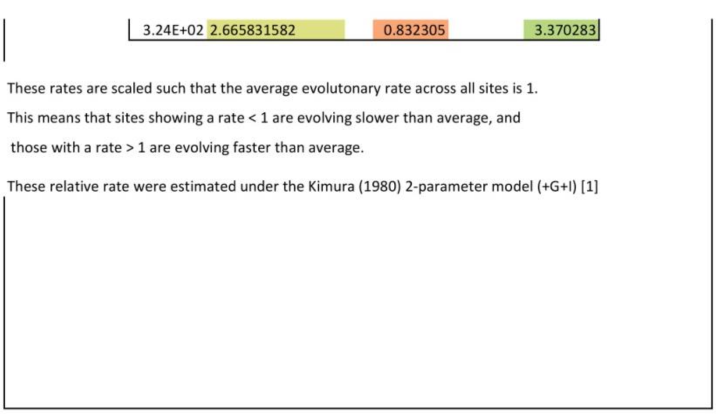
| | | | | | |
|---|---|---|---|---|---|
| 188 | | H | | | 1.607405 |
| 189 | | H | | | 1.136372 |
| 190 | | H | | | 1.861132 |
| 191 | | H | | | 1.61573 |
| 192 | 1.867725298 | H | | 0.450869 | 2.32515 |
| 193 | 2.05E+00 | H | 0.556805 | 0.111068 | 2.259807 |
| 194 | | H | | | 2.399744 |
| 1.95E+02 | | H | | | 1.915818 |
| 196 | | H | 0.651804 | | 2.107557 |
| 197 | | H | | | 2.083461 |
| 1.98E+02 | | C | | | 2.684911 |
| 199 | | C | | | 1.904028 |
| 200 | | C | | | 1.186233 |
| 201 | | C | | | 2.382028 |
| 202 | | C | | | 2.539282 |
| 203 | | C | | | 2.356618 |
| 2.04E+02 | | C | | | 1.062427 |
| 205 | | C | | | 1.904088 |
| 206 | | C | 0.318547 | | 5.451255 |
| 2.07E+02 | | C | | | 3.110642 |
| 208 | | C | | | 2.405222 |
| 209 | | C | | | 5.549701 |
| 2.10E+02 | | C | | | 2.186949 |
| 211 | | C | | | 4.472192 |
| 212 | | C | | | 2.102715 |
| 2.13E+02 | | C | 1.031082 | | 2.895951 |
| 214 | | C | | | 3.804784 |
| 215 | | C | | | 5.548347 |
| 2.16E+02 | | C | | | 5.55001 |
| 217 | | C | | | 1.694214 |
| 218 | | C | | | 2.144977 |
| 219 | | C | | | 0.905 |
| 220 | | C | | | 1.795339 |
| 221 | | C | | | 2.258933 |
| 2.22E+02 | | C | | | 2.485738 |
| 223 | | C | | | 2.386048 |
| 224 | | C | | | 5.548221 |
| 2.25E+02 | | C | | | 5.421306 |
| 226 | | H | | | 1.096809 |
| 227 | | H | | | 4.069411 |
| 2.28E+02 | | H | 1.171579 | | 3.296808 |
| 229 | | H | | | 2.078025 |
| 230 | | H | | | 0.590788 |
| 231 | | H | 0.577061 | | 5.224393 |
| 2.32E+02 | | H | | | 2.910753 |
| 233 | | H | | | 1.996348 |
| 2.34E+02 | | H | | | 1.975514 |
| 235 | 0.493985473 | H | 0.115528 | | 2.594806 |
| 236 | | H | | | 1.517588 |
| 2.37E+02 | | H | | | 2.001911 |
| 238 | | H | 1.296244 | | 2.196834 |
| 239 | 5.72E-01 | H | | | 2.155931 |
| 240 | | H | | | 1.724967 |
| 241 | | H | | | 2.053487 |
| 242 | | H | | | 3.621497 |
| 2.43E+02 | | H | 0.496253 | | 5.516691 |
| 244 | | H | | | 2.82176 |
| 245 | 1.50E+00 | H | 0.378909 | | 2.638951 |
| 2.46E+02 | 2.045316196 | H | 0.577536 | 0.111087 | 2.349279 |
| 247 | | H | 1.829593 | | 1.974282 |
| 248 | | C | | | 1.728022 |
| 2.49E+02 | 2.33E+00 | C | 1.186524 | | 2.179065 |
| 250 | | C | | | 5.549917 |
| 251 | 8.30E-01 | C | 0.147246 | | 1.052504 |
| 252 | 1.460503901 | C | 0.510894 | | 5.1336 |
| 253 | 2.391737642 | C | 0.516937 | 0.328198 | 0.985408 |
| 254 | 6.36E+00 | C | 1.092323 | 0.110996 | 2.750028 |
| 2.55E+02 | 3.24E+00 | C | 0.129029 | 0.127817 | 2.311515 |



| | | | | | |
|---|---|---|---|---|---|
| 256 | 7.06E+00 | C | 2.143241 | 0.996283 | 4.428862 |
| 257 | 7.06E+00 | C | 1.912712 | 0.111003 | 5.54047 |
| 2.58E+02 | 1.71E+00 | H | 0.127254 | | 5.439836 |
| 259 | 0.953990942 | H | 0.115626 | 0.186497 | 1.936139 |
| 260 | 1.98E+00 | H | 0.917969 | | 5.496208 |
| 2.61E+02 | 0.417820874 | H | 0.12726 | 0.111704 | 3.544191 |
| 262 | 1.01E+00 | H | 0.314627 | | 1.914741 |
| 263 | 7.03E+00 | H | 0.953219 | 0.308697 | 5.532198 |
| 2.64E+02 | 2.170509043 | H | | 0.612451 | 1.614003 |
| 265 | 3.81E-01 | H | 0.129205 | | 1.961511 |
| 266 | 5.51E-01 | H | 0.12883 | 0.143097 | 2.148301 |
| 2.67E+02 | 7.04E+00 | H | 1.192297 | 0.487401 | 2.18779 |
| 268 | | H | | | 1.575802 |
| 269 | 4.96E-01 | H | 0.115637 | | 2.618132 |
| 270 | 0.495529867 | H | 0.115637 | | 1.328821 |
| 271 | 2.792106144 | H | 0.882465 | | 5.550049 |
| 272 | | H | 0.555191 | | 5.550398 |
| 2.73E+02 | 3.17E-01 | H | | 0.11183 | 5.547811 |
| 274 | 7.02E+00 | H | 0.558375 | 0.867867 | 3.131712 |
| 275 | 6.54E+00 | H | 1.515979 | 0.111519 | 5.477676 |
| 276 | | H | | | 1.11077 |
| 277 | | H | 1.215637 | | 3.690322 |
| 278 | | H | | | 2.186945 |
| 2.79E+02 | 1.40E+00 | H | 0.472032 | | 4.975393 |
| 280 | | H | | | 1.579517 |
| 281 | | H | | | 1.083824 |
| 282 | | H | | | 5.531357 |
| 283 | | C | | | 1.152586 |
| 284 | | C | | | 5.525474 |
| 2.85E+02 | | C | | | 2.008436 |
| 286 | | C | | | 0.955488 |
| 287 | | C | | | 2.541314 |
| 2.88E+02 | | H | | | 1.730787 |
| 289 | | H | | | 1.834621 |



| | | | | | |
|---|---|---|---|---|---|
| 290 | | H | | | 2.203506 |
| 2.91E+02 | | H | 0.503615 | | 5.547901 |
| 292 | | H | | | 2.269324 |
| 293 | | H | | | 1.643086 |
| 2.94E+02 | | C | 1.26255 | | 4.64446 |
| 295 | 0.876959585 | C | 0.248109 | | 2.848926 |
| 296 | | C | | | 0.798593 |
| 297 | | C | | | 1.271469 |
| 298 | 7.97E-01 | H | | 0.188071 | 1.155609 |
| 299 | | H | | | 1.741485 |
| 3.00E+02 | | H | 0.529202 | | 5.511709 |
| 301 | | H | | | 2.28048 |
| 302 | 3.15E+00 | H | 1.109893 | | 5.384291 |
| 3.03E+02 | 7.38E-01 | H | 0.195125 | | 2.348436 |
| 304 | | H | 0.579384 | | 2.948936 |
| 305 | | H | | | 2.612315 |
| 3.06E+02 | 1.453353982 | H | 0.525103 | | 5.548424 |
| 307 | 1.84E+00 | H | 0.666277 | | 2.043468 |
| 308 | 2.76E+00 | H | 1.245022 | | 5.529227 |
| 309 | | H | | | 0.951372 |
| 310 | | H | | | 2.096993 |
| 311 | 2.19E+00 | H | 0.603186 | 0.113248 | 4.975948 |
| 3.12E+02 | 2.56E+00 | H | 0.917836 | | 2.13726 |
| 313 | 2.740928968 | H | 0.639502 | | 5.539682 |
| 314 | | H | | | 4.718818 |
| 315 | 2.159933785 | H | 0.702711 | | 5.519352 |
| 316 | 3.818672515 | H | 0.144089 | 0.399512 | 5.347964 |
| 317 | 4.07E+00 | C | 1.1077 | 0.464349 | 3.397929 |
| 3.18E+02 | 0.502577615 | C | 0.115488 | | 2.442635 |
| 319 | 1.745738393 | C | 0.115545 | 0.117367 | 5.549228 |
| 320 | 0.683492132 | C | 0.527189 | | 5.549588 |
| 3.21E+02 | | C | | | 5.540864 |
| 322 | | | | | 5.471688 |
| 323 | | | | | 1.142117 |



| | | | |
|---|---|---|---|
| 3.24E+02 | 2.665831582 | 0.832305 | 3.370283 |

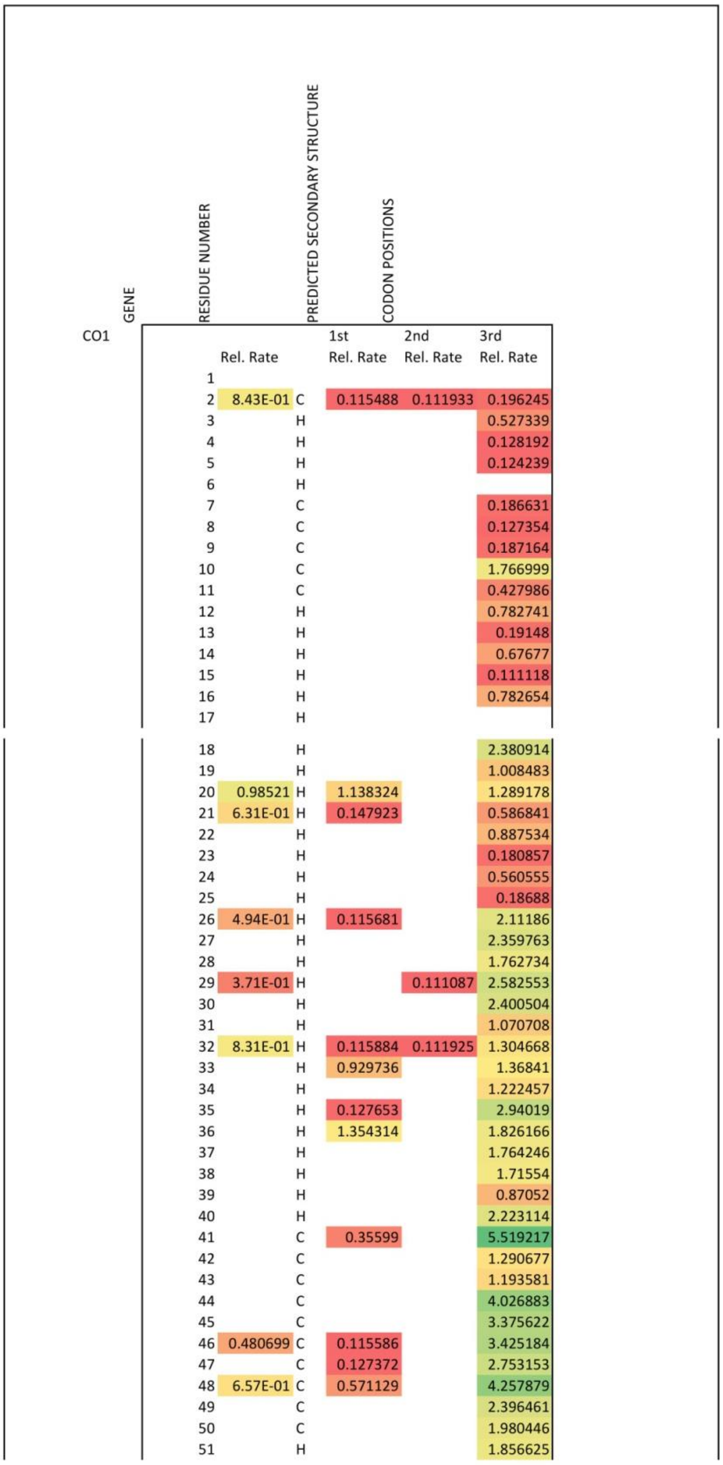
These rates are scaled such that the average evolutonary rate across all sites is 1.

This means that sites showing a rate < 1 are evolving slower than average, and

those with a rate > 1 are evolving faster than average.

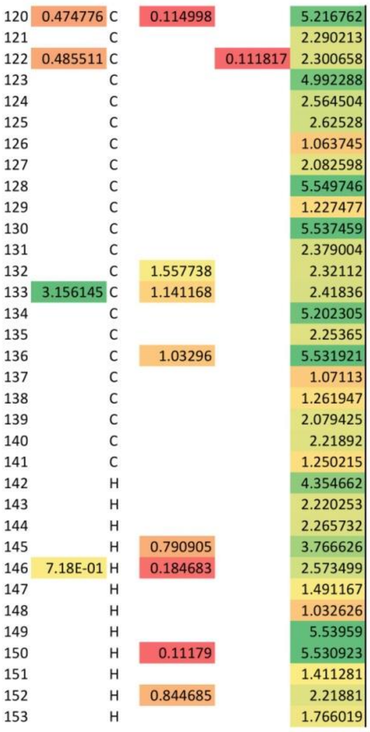
These relative rate were estimated under the Kimura (1980) 2-parameter model (+G+I) [1]

**Fig. A3(b)** Relative rate of evolution on codon position 1, 2 and 3 of Clupeoids CO1 gene.

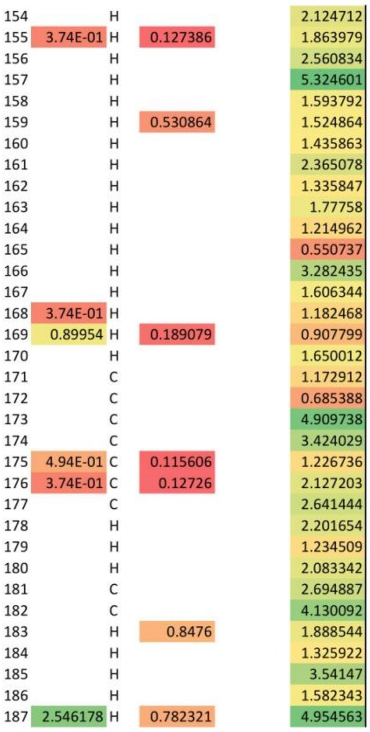


| GENE | RESIDUE NUMBER | | PREDICTED SECONDARY STRUCTURE | CODON POSITIONS | | |
|---|---|---|---|---|---|---|
| CO1 | | | | 1st | 2nd | 3rd |
| | | Rel. Rate | | Rel. Rate | Rel. Rate | Rel. Rate |
| | 1 | | | | | |
| | 2 | 8.43E-01 | C | 0.115488 | 0.111933 | 0.196245 |
| | 3 | | H | | | 0.527339 |
| | 4 | | H | | | 0.128192 |
| | 5 | | H | | | 0.124239 |
| | 6 | | H | | | |
| | 7 | | C | | | 0.186631 |
| | 8 | | C | | | 0.127354 |
| | 9 | | C | | | 0.187164 |
| | 10 | | C | | | 1.766999 |
| | 11 | | C | | | 0.427986 |
| | 12 | | H | | | 0.782741 |
| | 13 | | H | | | 0.19148 |
| | 14 | | H | | | 0.67677 |
| | 15 | | H | | | 0.111118 |
| | 16 | | H | | | 0.782654 |
| | 17 | | H | | | |
| | 18 | | H | | | 2.380914 |
| | 19 | | H | | | 1.008483 |
| | 20 | 0.98521 | H | 1.138324 | | 1.289178 |
| | 21 | 6.31E-01 | H | 0.147923 | | 0.586841 |
| | 22 | | H | | | 0.887534 |
| | 23 | | H | | | 0.180857 |
| | 24 | | H | | | 0.560555 |
| | 25 | | H | | | 0.18688 |
| | 26 | 4.94E-01 | H | 0.115681 | | 2.11186 |
| | 27 | | H | | | 2.359763 |
| | 28 | | H | | | 1.762734 |
| | 29 | 3.71E-01 | H | | 0.111087 | 2.582553 |
| | 30 | | H | | | 2.400504 |
| | 31 | | H | | | 1.070708 |
| | 32 | 8.31E-01 | H | 0.115884 | 0.111925 | 1.304668 |
| | 33 | | H | 0.929736 | | 1.36841 |
| | 34 | | H | | | 1.222457 |
| | 35 | | H | 0.127653 | | 2.94019 |
| | 36 | | H | 1.354314 | | 1.826166 |
| | 37 | | H | | | 1.764246 |
| | 38 | | H | | | 1.71554 |
| | 39 | | H | | | 0.87052 |
| | 40 | | H | | | 2.223114 |
| | 41 | | C | 0.35599 | | 5.519217 |
| | 42 | | C | | | 1.290677 |
| | 43 | | C | | | 1.193581 |
| | 44 | | C | | | 4.026883 |
| | 45 | | C | | | 3.375622 |
| | 46 | 0.480699 | C | 0.115586 | | 3.425184 |
| | 47 | | C | 0.127372 | | 2.753153 |
| | 48 | 6.57E-01 | C | 0.571129 | | 4.257879 |
| | 49 | | C | | | 2.396461 |
| | 50 | | C | | | 1.980446 |
| | 51 | | H | | | 1.856625 |

| | | | | | |
|---|---|---|---|---|---|
| 52 | | H | | | 1.562104 |
| 53 | 3.74E-01 | H | 0.127489 | | 1.428102 |
| 54 | | H | | | 1.33468 |
| 55 | 0.473844 | H | 0.127763 | | 0.95825 |
| 56 | | H | | | 2.826132 |
| 57 | 3.77E-01 | H | | | 1.219926 |
| 58 | 3.65E-01 | H | | 0.111037 | 1.203256 |
| 59 | 4.17E-01 | H | 0.127906 | | 1.998049 |
| 60 | | H | | | 1.809098 |
| 61 | | H | | | 1.788021 |
| 62 | | H | | | 1.224269 |
| 63 | | H | | | 1.551723 |
| 64 | | H | | | 1.760368 |
| 65 | | H | | | 1.465092 |
| 66 | | H | | | 1.184403 |
| 67 | | H | | | 0.202296 |
| 68 | | H | | | 0.612565 |
| 69 | | H | | | 0.774654 |
| 70 | | H | | | 1.511458 |
| 71 | | H | | | 1.629581 |
| 72 | | H | | | 3.000371 |
| 73 | 0.635861 | H | 0.127736 | | 1.652906 |
| 74 | 1.098907 | H | 1.409358 | | 2.079193 |
| 75 | | H | | | 1.988499 |
| 76 | | H | | | 2.565561 |
| 77 | | H | | | 5.545548 |
| 78 | | H | | | 0.98308 |
| 79 | | H | | | 1.782146 |
| 80 | | H | | | 0.942516 |
| 81 | | H | | | 0.94279 |
| 82 | | H | 1.119305 | | 4.340083 |
| 83 | 1.367301 | H | 0.470668 | | 5.282506 |
| 84 | | H | | | 4.457404 |
| 85 | | H | 0.500052 | | 2.556 |
| 86 | | H | | | 0.781103 |
| 87 | 1.091218 | C | 0.51603 | | 1.435331 |
| 88 | | C | | | 2.424943 |
| 89 | | C | | | 2.396109 |
| 90 | | C | | | 2.382182 |
| 91 | | C | | | 1.824294 |
| 92 | | C | | | 1.01037 |
| 93 | | C | | | 0.861772 |
| 94 | | C | | | 0.498286 |
| 95 | | H | | | 4.520012 |
| 96 | | H | | | 0.860459 |
| 97 | | H | | | 1.729852 |
| 98 | | H | | | 1.873157 |
| 99 | | H | | | 1.523935 |
| 100 | | H | | | 1.138166 |
| 101 | | H | | | 0.314513 |
| 102 | | H | | | 0.524293 |
| 103 | | H | | | 1.24061 |
| 104 | | H | 0.127406 | | 2.652757 |
| 105 | | H | 0.111773 | | 5.549412 |
| 106 | | H | | | 4.988252 |
| 107 | | H | | | 2.206942 |
| 108 | | H | | | 2.730435 |
| 109 | | H | | | 1.145061 |
| 110 | | H | 0.127608 | | 2.700817 |
| 111 | 6.84E-01 | H | 0.127781 | | 3.482735 |
| 112 | | H | 0.463769 | | 5.548359 |
| 113 | | H | 1.040548 | | 4.032137 |
| 114 | 1.12E+00 | H | 0.241626 | | 1.789125 |
| 115 | | H | | | 3.600382 |
| 116 | | H | | | 5.533836 |
| 117 | 1.70436 | H | | 0.149467 | 4.130415 |
| 118 | | C | | | 2.953847 |
| 119 | | C | | | 1.373214 |

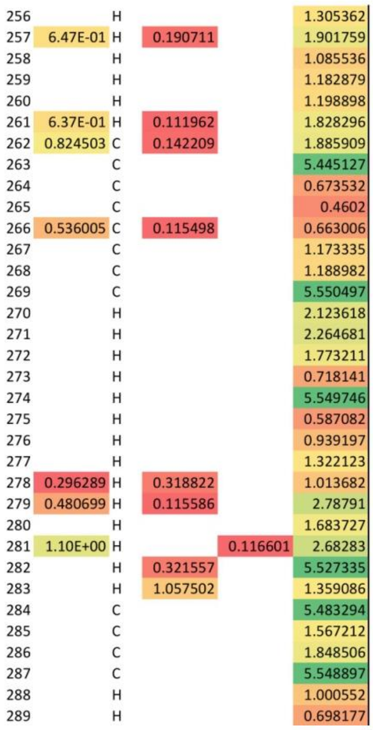


| | | | | | |
|---|---|---|---|---|---|
| 120 | 0.474776 | C | 0.114998 | | 5.216762 |
| 121 | | C | | | 2.290213 |
| 122 | 0.485511 | C | | 0.111817 | 2.300658 |
| 123 | | C | | | 4.992288 |
| 124 | | C | | | 2.564504 |
| 125 | | C | | | 2.62528 |
| 126 | | C | | | 1.063745 |
| 127 | | C | | | 2.082598 |
| 128 | | C | | | 5.549746 |
| 129 | | C | | | 1.227477 |
| 130 | | C | | | 5.537459 |
| 131 | | C | | | 2.379004 |
| 132 | | C | 1.557738 | | 2.32112 |
| 133 | 3.156145 | C | 1.141168 | | 2.41836 |
| 134 | | C | | | 5.202305 |
| 135 | | C | | | 2.25365 |
| 136 | | C | 1.03296 | | 5.531921 |
| 137 | | C | | | 1.07113 |
| 138 | | C | | | 1.261947 |
| 139 | | C | | | 2.079425 |
| 140 | | C | | | 2.21892 |
| 141 | | C | | | 1.250215 |
| 142 | | H | | | 4.354662 |
| 143 | | H | | | 2.220253 |
| 144 | | H | | | 2.265732 |
| 145 | | H | 0.790905 | | 3.766626 |
| 146 | 7.18E-01 | H | 0.184683 | | 2.573499 |
| 147 | | H | | | 1.491167 |
| 148 | | H | | | 1.032626 |
| 149 | | H | | | 5.53959 |
| 150 | | H | 0.11179 | | 5.530923 |
| 151 | | H | | | 1.411281 |
| 152 | | H | 0.844685 | | 2.21881 |
| 153 | | H | | | 1.766019 |

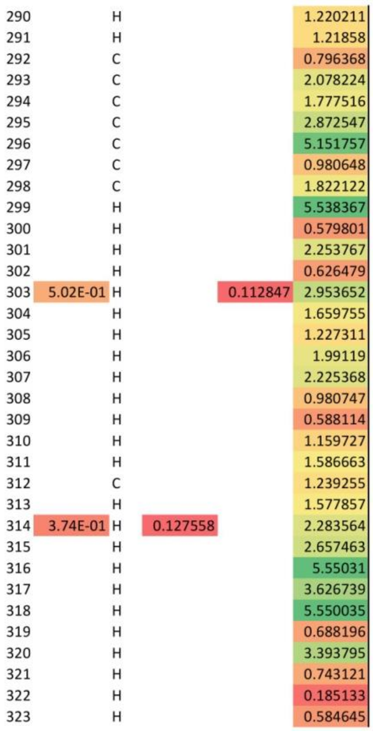


| | | | | | |
|---|---|---|---|---|---|
| 154 | | H | | | 2.124712 |
| 155 | 3.74E-01 | H | 0.127386 | | 1.863979 |
| 156 | | H | | | 2.560834 |
| 157 | | H | | | 5.324601 |
| 158 | | H | | | 1.593792 |
| 159 | | H | 0.530864 | | 1.524864 |
| 160 | | H | | | 1.435863 |
| 161 | | H | | | 2.365078 |
| 162 | | H | | | 1.335847 |
| 163 | | H | | | 1.77758 |
| 164 | | H | | | 1.214962 |
| 165 | | H | | | 0.550737 |
| 166 | | H | | | 3.282435 |
| 167 | | H | | | 1.606344 |
| 168 | 3.74E-01 | H | | | 1.182468 |
| 169 | 0.89954 | H | 0.189079 | | 0.907799 |
| 170 | | H | | | 1.650012 |
| 171 | | C | | | 1.172912 |
| 172 | | C | | | 0.685388 |
| 173 | | C | | | 4.909738 |
| 174 | | C | | | 3.424029 |
| 175 | 4.94E-01 | C | 0.115606 | | 1.226736 |
| 176 | 3.74E-01 | C | 0.12726 | | 2.127203 |
| 177 | | C | | | 2.641444 |
| 178 | | H | | | 2.201654 |
| 179 | | H | | | 1.234509 |
| 180 | | H | | | 2.083342 |
| 181 | | C | | | 2.694887 |
| 182 | | C | | | 4.130092 |
| 183 | | H | 0.8476 | | 1.888544 |
| 184 | | H | | | 1.325922 |
| 185 | | H | | | 3.54147 |
| 186 | | H | | | 1.582343 |
| 187 | 2.546178 | H | 0.782321 | | 4.954563 |

| | | | | |
|---|---|---|---|---|
| 188 | | H | | 5.51826 |
| 189 | 6.86E-01 | H | 0.664224 | 1.313684 |
| 190 | 1.313999 | H | 0.483116 | 3.47949 |
| 191 | | H | | 2.170936 |
| 192 | | H | | 2.159853 |
| 193 | | H | | 5.382967 |
| 194 | | H | 0.313483 | 3.100901 |
| 195 | | H | 0.128635 | 1.652974 |
| 196 | | H | 0.111773 | 4.491497 |
| 197 | | H | 0.494216 | 5.537518 |
| 198 | | H | | 5.158033 |
| 199 | | H | | 5.499061 |
| 200 | | H | | 4.880603 |
| 201 | | H | | 5.550496 |
| 202 | | H | 0.882506 | 2.189266 |
| 203 | | H | | 2.115096 |
| 204 | | H | | 4.132228 |
| 205 | | H | | 2.251646 |
| 206 | | H | | 0.90394 |
| 207 | | H | | 2.237209 |
| 208 | | H | | 1.30458 |
| 209 | | H | 0.111762 | 5.49492 |
| 210 | | H | | 4.501164 |
| 211 | | H | | 1.497747 |
| 212 | | H | | 2.004491 |
| 213 | | H | | 0.835566 |
| 214 | | H | | 2.073821 |
| 215 | | C | 0.498202 | 2.89019 |
| 216 | | C | | 2.145341 |
| 217 | | C | | 3.978738 |
| 218 | | C | | 2.436284 |
| 219 | | C | | 0.563962 |
| 220 | | C | | 1.298408 |
| 221 | | C | | 0.817813 |
| 222 | | H | | 5.448473 |
| 223 | | H | | 1.779682 |
| 224 | | H | | 1.990314 |
| 225 | | C | | 2.15526 |
| 226 | | C | | 1.859294 |
| 227 | | C | | 1.134606 |
| 228 | | H | | 5.15338 |
| 229 | | H | | 1.238981 |
| 230 | | H | 0.491446 | 5.545907 |
| 231 | | H | | 2.004115 |
| 232 | | H | | 0.673798 |
| 233 | | H | | 1.184484 |
| 234 | | H | 0.754782 | 2.249362 |
| 235 | | H | | 0.657413 |
| 236 | | H | | 0.969886 |
| 237 | | H | | 0.318377 |
| 238 | | H | | 2.224389 |
| 239 | | H | | 5.545285 |
| 240 | | H | | 0.442299 |
| 241 | | H | | 5.470702 |
| 242 | | H | | 0.458547 |
| 243 | | H | | 5.542285 |
| 244 | | H | | 2.253944 |
| 245 | | H | | 0.518817 |
| 246 | | H | 0.546289 | 5.459237 |
| 247 | | H | | 0.545239 |
| 248 | | H | 1.007092 | 5.420415 |
| 249 | | H | | 5.528162 |
| 250 | | H | | 2.600467 |
| 251 | | H | | 2.036717 |
| 252 | | H | | 2.041211 |
| 253 | 7.43E-01 | H | 0.184504 | 1.119959 |
| 254 | | H | | 2.124661 |
| 255 | | H | | 4.0643 |



|  |  |  |  |  |  |
|---|---|---|---|---|---|
| 256 |  | H |  |  | 1.305362 |
| 257 | 6.47E-01 | H | 0.190711 |  | 1.901759 |
| 258 |  | H |  |  | 1.085536 |
| 259 |  | H |  |  | 1.182879 |
| 260 |  | H |  |  | 1.198898 |
| 261 | 6.37E-01 | H | 0.111962 |  | 1.828296 |
| 262 | 0.824503 | C | 0.142209 |  | 1.885909 |
| 263 |  | C |  |  | 5.445127 |
| 264 |  | C |  |  | 0.673532 |
| 265 |  | C |  |  | 0.4602 |
| 266 | 0.536005 | C | 0.115498 |  | 0.663006 |
| 267 |  | C |  |  | 1.173335 |
| 268 |  | C |  |  | 1.188982 |
| 269 |  | C |  |  | 5.550497 |
| 270 |  | H |  |  | 2.123618 |
| 271 |  | H |  |  | 2.264681 |
| 272 |  | H |  |  | 1.773211 |
| 273 |  | H |  |  | 0.718141 |
| 274 |  | H |  |  | 5.549746 |
| 275 |  | H |  |  | 0.587082 |
| 276 |  | H |  |  | 0.939197 |
| 277 |  | H |  |  | 1.322123 |
| 278 | 0.296289 | H | 0.318822 |  | 1.013682 |
| 279 | 0.480699 | H | 0.115586 |  | 2.78791 |
| 280 |  | H |  |  | 1.683727 |
| 281 | 1.10E+00 | H |  | 0.116601 | 2.68283 |
| 282 |  | H | 0.321557 |  | 5.527335 |
| 283 |  | H | 1.057502 |  | 1.359086 |
| 284 |  | C |  |  | 5.483294 |
| 285 |  | C |  |  | 1.567212 |
| 286 |  | C |  |  | 1.848506 |
| 287 |  | C |  |  | 5.548897 |
| 288 |  | H |  |  | 1.000552 |
| 289 |  | H |  |  | 0.698177 |



|  |  |  |  |  |  |
|---|---|---|---|---|---|
| 290 |  | H |  |  | 1.220211 |
| 291 |  | H |  |  | 1.21858 |
| 292 |  | C |  |  | 0.796368 |
| 293 |  | C |  |  | 2.078224 |
| 294 |  | C |  |  | 1.777516 |
| 295 |  | C |  |  | 2.872547 |
| 296 |  | C |  |  | 5.151757 |
| 297 |  | C |  |  | 0.980648 |
| 298 |  | C |  |  | 1.822122 |
| 299 |  | H |  |  | 5.538367 |
| 300 |  | H |  |  | 0.579801 |
| 301 |  | H |  |  | 2.253767 |
| 302 |  | H |  |  | 0.626479 |
| 303 | 5.02E-01 | H |  | 0.112847 | 2.953652 |
| 304 |  | H |  |  | 1.659755 |
| 305 |  | H |  |  | 1.227311 |
| 306 |  | H |  |  | 1.99119 |
| 307 |  | H |  |  | 2.225368 |
| 308 |  | H |  |  | 0.980747 |
| 309 |  | H |  |  | 0.588114 |
| 310 |  | H |  |  | 1.159727 |
| 311 |  | H |  |  | 1.586663 |
| 312 |  | C |  |  | 1.239255 |
| 313 |  | H |  |  | 1.577857 |
| 314 | 3.74E-01 | H | 0.127558 |  | 2.283564 |
| 315 |  | H |  |  | 2.657463 |
| 316 |  | H |  |  | 5.55031 |
| 317 |  | H |  |  | 3.626739 |
| 318 |  | H |  |  | 5.550035 |
| 319 |  | H |  |  | 0.688196 |
| 320 |  | H |  |  | 3.393795 |
| 321 |  | H |  |  | 0.743121 |
| 322 |  | H |  |  | 0.185133 |
| 323 |  | H |  |  | 0.584645 |

| | | | | | |
|---|---|---|---|---|---|
| 324 | | H | 0.111624 | | 1.477614 |
| 325 | | H | | | 2.502006 |
| 326 | | H | | | 5.438978 |
| 327 | | H | 0.497387 | | 5.499083 |
| 328 | | C | | | 1.114907 |
| 329 | | C | | | 5.460573 |
| 330 | | C | | | 4.162975 |
| 331 | 0.574679 | C | 0.111696 | 0.113738 | 2.241586 |
| 332 | | C | | | 1.684592 |
| 333 | | C | | | 0.696121 |
| 334 | | C | | | 1.112948 |
| 335 | 1.99E+00 | C | | | 2.213089 |
| 336 | 0.423731 | H | 0.126885 | | 4.984228 |
| 337 | | H | | | 5.447813 |
| 338 | 1.343706 | H | 0.21143 | | 2.823265 |
| 339 | | H | 0.311836 | | 5.547056 |
| 340 | | H | | | 1.129858 |
| 341 | | H | | | 1.90867 |
| 342 | | H | 0.574931 | | 3.211726 |
| 343 | | H | | | 5.423527 |
| 344 | | H | | | 2.356928 |
| 345 | | H | | | 1.581831 |
| 346 | | H | | | 0.322543 |
| 347 | | H | 0.583959 | | 3.160984 |
| 348 | | H | | | 2.194618 |
| 349 | | H | | | 1.637706 |
| 350 | | H | | | 5.550033 |
| 351 | | H | | | 5.549439 |
| 352 | | H | | | 5.262784 |
| 353 | | H | 0.563075 | | 3.725948 |
| 354 | | H | | | 1.471623 |
| 355 | | H | | | 2.290531 |
| 356 | | H | | | 1.096286 |
| 357 | 3.66E-01 | H | | 0.111096 | 5.523542 |
| 358 | | H | 1.341691 | | 2.134443 |
| 359 | 0.937357 | H | 0.215012 | | 2.322073 |
| 360 | | C | | | 1.180796 |
| 361 | | H | | | 2.316997 |
| 362 | | H | | | 2.258031 |
| 363 | | H | 1.047725 | | 2.220771 |
| 364 | | H | | | 1.87186 |
| 365 | | H | | | 0.512357 |
| 366 | | H | | | 5.134412 |
| 367 | | H | 0.184368 | | 5.523856 |
| 368 | | C | | | 1.95572 |
| 369 | | C | | | 0.548679 |
| 370 | | C | | | 2.257652 |
| 371 | | H | | | 1.213663 |
| 372 | | H | | | 1.8281 |
| 373 | | H | | | 0.464444 |
| 374 | | H | | | 0.953652 |
| 375 | | H | | | 1.332259 |
| 376 | | H | | | 0.519211 |
| 377 | | H | | | 0.209478 |
| 378 | 4.79E-01 | H | | 0.12815 | 0.566497 |
| 379 | | H | | | 1.572595 |
| 380 | | H | | | 5.449965 |
| 381 | | H | 1.213329 | | 3.521404 |
| 382 | | H | | | 2.311487 |
| 383 | | H | | | 1.498925 |
| 384 | | H | | | 4.478955 |
| 385 | | H | | | 2.535546 |
| 386 | | H | | | 2.051233 |
| 387 | | H | | | 1.291914 |
| 388 | | H | | | 2.244156 |
| 389 | | H | | | 2.208883 |
| 390 | 2.97E-01 | H | 0.126277 | | 1.557696 |
| 391 | 0.50795 | H | | 0.11353 | 2.065021 |

|  |  |  |  |  |  |
|---|---|---|---|---|---|
| 392 | 1.21E+00 | H |  | 0.2052 | 1.100555 |
| 393 |  | H |  |  | 2.143573 |
| 394 | 1.432627 | H | 0.476019 |  | 5.548283 |
| 395 |  | H |  |  | 0.567497 |
| 396 |  | H |  |  | 0.581473 |
| 397 |  | H |  |  | 0.449921 |
| 398 |  | H |  |  | 5.550495 |
| 399 |  | H | 0.50775 |  | 1.618039 |
| 400 |  | H |  |  | 2.176003 |
| 401 | 2.03E+00 | H | 0.671707 |  | 2.121107 |
| 402 |  | C |  |  | 2.638745 |
| 403 | 1.074993 | C |  | 0.189937 | 2.439006 |
| 404 |  | C |  |  | 1.79426 |
| 405 |  | C | 0.111815 |  | 5.488675 |
| 406 |  | C |  |  | 0.941392 |
| 407 | 7.70E-01 | H | 0.187892 | 0.11556 | 0.500447 |
| 408 | 4.21E-01 | H | 0.127591 |  | 1.967633 |
| 409 |  | H |  |  |  |
| 410 | 9.78E-01 | H | 0.326417 |  | 0.528378 |
| 411 |  | H |  |  | 0.673524 |
| 412 | 0.37658 | H | 0.127992 |  | 0.786853 |
| 413 |  | H |  |  | 0.744431 |
| 414 |  | H |  |  | 1.909384 |
| 415 | 1.06E+00 | H |  | 0.116137 | 2.322381 |
| 416 | 1.37E+00 | H | 0.472069 |  | 5.550244 |
| 417 |  | H |  |  | 1.154194 |
| 418 |  | H |  |  | 1.755538 |
| 419 | 2.781438 | H | 0.952623 | 0.124054 | 5.54886 |
| 420 |  | H |  |  | 2.349669 |
| 421 |  | H |  |  | 5.533066 |
| 422 |  | H |  |  | 1.166254 |
| 423 | 1.74E+00 | H | 1.232088 |  | 2.801456 |
| 424 |  | H |  |  | 2.11689 |
| 425 |  | H |  |  | 0.327726 |
| 426 |  | H |  |  | 0.185358 |
| 427 |  | H |  |  | 2.757565 |
| 428 |  | H |  |  | 1.781481 |
| 429 |  | H |  |  | 1.057828 |
| 430 |  | H |  |  | 0.447676 |
| 431 |  | H | 0.199899 |  | 2.407748 |
| 432 |  | H |  |  | 3.723224 |
| 433 |  | H | 0.63794 |  | 1.311415 |
| 434 |  | H |  |  | 1.834803 |
| 435 |  | C |  |  | 1.637302 |
| 436 |  | C |  |  | 0.750368 |
| 437 |  | C |  |  | 1.083216 |
| 438 |  | C |  |  | 1.193668 |
| 439 |  | C |  |  | 2.929495 |
| 440 |  | C |  |  | 1.512284 |
| 441 |  | C |  |  | 4.948152 |
| 442 |  | C |  |  | 0.443267 |
| 443 |  | C |  |  | 0.554662 |
| 444 |  | C |  |  | 2.198532 |
| 445 |  | H |  |  | 1.060573 |
| 446 |  | H |  |  | 0.864073 |
| 447 |  | H |  |  | 1.271311 |
| 448 |  | H |  |  | 1.623364 |
| 449 | 6.84E-01 | H | 0.111952 |  | 3.946204 |
| 450 |  | H |  |  | 0.189298 |
| 451 |  | H |  |  | 1.751675 |
| 452 |  | H |  |  | 2.572587 |
| 453 | 3.66E-01 | H | 0.115545 |  | 5.153672 |
| 454 |  | H |  |  | 2.585041 |
| 455 |  | H |  |  | 0.530723 |
| 456 |  | H |  |  | 2.171176 |
| 457 |  | H |  |  | 5.099459 |
| 458 |  | H |  |  | 2.291875 |
| 459 |  | H | 0.921998 |  | 1.114899 |

| | | | | | |
|---|---|---|---|---|---|
| 460 | 1.147015 | H | 0.446279 | | 2.331985 |
| 461 | | H | | | 3.729682 |
| 462 | | H | 0.522533 | | 3.880017 |
| 463 | 0.363622 | H | 0.115488 | | 1.458314 |
| 464 | 8.37E-01 | H | | 0.130103 | 1.80327 |
| 465 | | H | | | 4.806001 |
| 466 | | H | | | 1.102595 |
| 467 | 0.511292 | H | 0.185782 | | 1.160651 |
| 468 | | H | | | 0.338818 |
| 469 | | H | 2.151567 | | 1.748085 |
| 470 | | H | | | 1.184468 |
| 471 | | H | | | 0.965894 |
| 472 | 1.637777 | H | 0.568355 | | 1.932425 |
| 473 | | H | | | 0.828803 |
| 474 | | H | | | 0.524334 |
| 475 | | H | | | 0.862266 |
| 476 | | H | | | 2.015916 |
| 477 | 8.32E-01 | H | 0.144358 | | 2.732604 |
| 478 | 0.812244 | H | 0.143124 | | 2.104364 |
| 479 | 5.27E-01 | C | 0.126979 | | 0.617096 |
| 480 | | C | | | 0.328625 |
| 481 | | C | | | 0.799498 |
| 482 | | C | | | 3.54946 |
| 483 | 2.76E+00 | C | 0.804278 | 0.111624 | 0.624899 |
| 484 | 3.15E-01 | C | 0.111432 | | 1.053313 |
| 485 | | C | | | 2.046484 |
| 486 | | C | | | 1.382648 |
| 487 | | C | 0.552813 | | 4.160637 |
| 488 | 7.15E-01 | C | 0.186671 | | 1.550194 |
| 489 | 2.66E+00 | C | 0.674592 | 0.442747 | 2.307978 |
| 490 | 0.407862 | C | | 0.111817 | 1.953379 |
| 491 | | C | | | 1.138522 |
| 492 | | H | | | 1.231529 |
| 493 | | H | | | 2.155081 |
| 494 | | H | | | 0.672322 |
| 495 | | C | 0.320739 | | 3.923934 |
| 496 | | C | | | 0.793945 |
| 497 | | C | | | 1.042359 |
| 498 | | C | | | 0.372976 |
| 499 | | C | | | 1.162944 |
| 500 | | C | | | 2.937397 |
| 501 | | C | | | 5.500895 |
| 502 | | C | | | 0.904343 |
| 503 | | C | | | 0.490623 |
| 504 | | C | | | 1.455525 |
| 505 | 0.957483 | C | | 0.128363 | 1.082887 |
| 506 | | C | | | 2.087508 |
| 507 | | C | | | 2.182612 |
| 508 | | C | | | 1.890686 |
| 509 | | C | | | 1.341924 |
| 510 | 1.32E+00 | C | | 0.127477 | 1.476231 |
| 511 | 0.370098 | C | 0.115897 | 0.11109 | 5.052555 |
| 512 | 0.391917 | | 0.112043 | 0.127998 | 0.329565 |
| 513 | 1.14E+00 | | 0.243862 | 0.125442 | 2.42358 |
| 514 | 0.545863 | | | | 0.701052 |

These rates are scaled such that the average evolutonary rate across all sites is 1.
This means that sites showing a rate < 1 are evolving slower than average, and
those with a rate > 1 are evolving faster than average.
These relative rate were estimated under the Kimura (1980) 2-parameter model (+G+I) [1]

**Table A1** Example of Summery of TreeSAAP analysis on CO2 gene. Result indicated significant (p<0.001) amino acid physiochemical property changes (Categories (6-7-8)) in the positively selected codons or amino acid regions of clupeoids mitogenome,

| Gene | Codon/ Amino acid position | Branches | Total no of amino acid properties | Physiochemical property changes |
|---|---|---|---|---|
| CO 2 | 1226 | node#101_-->_Sardina_pilchardus | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1226 | node#104_-->_Clupeonella_cultriventris | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1226 | node#117_-->_node#118 | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1226 | node#78_-->_node#79 | 1 | Solvent_accessible_reduction_ratio |
| CO 2 | 1231 | node#134_-->_Lycothrissa_crocodilus | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1234 | node#106_-->_node#107 | 1 | Refractive_index |
| CO 2 | 1234 | node#78_-->_Escualosa_thoracata | 1 | Refractive_index |
| CO 2 | 1234 | node#80_-->_node#93 | 2 | Bulkiness/Chromatographic_index |
| CO 2 | 1234 | node#82_-->_node#83 | 1 | Refractive_index |
| CO 2 | 1238 | node#71_-->_node#126 | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1304 | node#93_-->_Gudusia_chapra | 1 | Isoelectric_point |
| CO 2 | 1316 | node#101_-->_Sardina_pilchardus | 1 | Compressibility |
| CO 2 | 1316 | node#134_-->_Lycothrissa_crocodilus | 1 | Hydropathy |
| CO 2 | 1319 | node#106_-->_Ehirava_fluviatilis | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1319 | node#114_-->_node#115 | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1319 | node#123_-->_Ilisha_africana | 2 | Bulkiness/Chromatographic_index |
| CO 2 | 1319 | node#127_-->_node#128 | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1319 | node#77_-->_node#100 | 1 | Solvent_accessible_reduction_ratio |
| CO 2 | 1319 | node#82_-->_node#83 | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1319 | node#86_-->_Microthrissa_royauxi | 1 | Polarity |
| CO 2 | 1319 | node#90_-->_Dorosoma_petenense | 2 | Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1407 | node#115_-->_node#117 | 4 | Chromatographic_index/Hydropathy/Solvent_accessible_reduction_ratio/Surrounding_hydrophobicity |
| CO 2 | 1415 | node#106_-->_Ehirava_fluviatilis | 2 | Isoelectric_point/Polarity |
| CO 2 | 1415 | node#107_-->_node#108 | 2 | Polar_requirement/Polarity |
| CO 2 | 1418 | node#71_-->_Denticeps_clupeoides | 1 | Isoelectric_point |