Indian J. Fish., 66(3): 125-130, 2019
DOI: 10.21077/ijf.2019.66.3.86643-16

Note

The complete mitochondrial genome and phylogeny of the green chromide *Etroplus suratensis* (Bloch, 1790) from Vembanad Lake, Kerala, south India

WILSON SEBASTIAN, SANDHYA SUKUMARAN AND A. GOPALAKRISHNAN
ICAR-Central Marine Fisheries Research Institute, Post Box No. 1603, Ernakulam North P. O., Kochi - 682 018 Kerala, India
e-mail: sukumaransandhya@yahoocom

ABSTRACT

The green chromide *Etroplus suratensis* (Bloch, 1790), is a cichlid species which forms an economically valuable food fish and a preferred candidate for brackishwater aquaculture in India. The complete mitogenome of *E. suratensis* collected from Vembanad Lake, Kerala, India has been characterised in the present study. 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, 13 protein-coding genes and 1 non-coding control region/D-loop, with the gene order identical to vertebrates. In the phylogenetic analysis, *E. suratensis* is clustered with other Indo-Sri Lankan taxa. Among cichlids, the groups from South America and Africa are monophyletic in origin. The mitogenomic information generated in this study will be valuable for further studies on evolution, taxonomy, conservation, environmental adaptation and selective breeding of this species having aquaculture, ornamental and evolutionary importance.

Keywords: Cichlids, *Etroplus suratensis,* Complete mitochondrial genome, Phylogenetic status

Etroplus suratensis (Bloch, 1790), is a euryhaline cichlid species distributed mainly in freshwater, brackishwater and river mouths of peninsular India and Sri Lanka and it is the most abundant species among the genus *Etroplus* (Jayaram, 2010). Wild populations of *E. suratensis* have been recorded from the states of Kerala and Tamil Nadu (Jayaram, 2010) and introduced populations from Andhra Pradesh and Odisha. Cichlids have been considered as model organisms for evolutionary biology, evolutionary genetics and phenotype-genotype relationship studies because of their ecological, morphological and behavioural diversity (Azuma *et al.,* 2008). *E. suratensis* is one of the most sought after candidate species for brackishwater aquaculture in Kerala due to its suitability for culture in confinement along with tolerance to a wide range of environmental conditions (Chandrasekar *et al.,* 2014). *E. suratensis* 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). Wild populations of *E. suratensis* are facing habitat deterioration due to increasing urbanisation, tourism activities in backwaters/ estuaries and threats from proliferation of exotic species like *Oreochromis mossambicus* and *Trichogaster trichopterus* (Padmakumar *et al.,* 2002; Kumar *et al.,* 2009). Attempts have been made for captive breeding aimed at conservation along with the creation of aquatic sanctuaries/no fishing zones, within some of the larger estuaries (Padmakumar *et al.,* 2012).

Comparative mitogenomic information has revolutionised several concepts of molecular phylogeny and evolution across multiple taxonomic levels (Miya and Nishida, 2015). Genetic information coupled with biological and behavioural data is crucial for the conservation and management of endangered species. Mitochondrial Oxidative Phosphorylation System (OXPHOS complex) has been indicated as important for selection and adaptation to different environmental regimes in marine fishes (Garvin *et al.,* 2012; Caballero *et al.,* 2015). *E. suratensis* is distributed widely across environmental clines and hence identifying the signals of positive and diversifying selection in OXPHOS machinery of the mitogenome will provide clues regarding their vulnerability to environmental alterations. Considering all these, we characterised the complete mitochondrial genome structure and organisation of *E. suratensis* collected from Vembanad Lake, Kerala followed by phylogenetic analysis using complete mitogenome.

The complete mitogenome of *E. suratensis* collected from Chilka Lake, Odisha, India has already been characterised by Mohanta *et al.* (2016). However, detailed analysis on structure, organisation, 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.

Genomic DNA was isolated by standard phenol/chloroform method (Sambrook and Russell, 2001). The entire mitogenome was amplified using a long PCR technique with Q5® High-Fidelity DNA polymerase. Primer pairs (Table1) were designed on the basis of known regions of the *E. suratensis* mtDNA and complete mitogenome was amplified as 5 contiguous, overlapping segments and sequenced with both primers using the BigDye Terminator Sequencing Ready Reaction v30 kit (Applied Biosystems). 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. The sequence fragments obtained were assembled using Geneious R7 (Kearse *et al.,* 2012), annotated with NCBI-BLAST (National Centre for Biotechnology Information-The Basic Local Alignment Search Tool) and MitoAnnotator (Iwasaki *et al.,* 2013) and deposited in NCBI GenBank (Accession no. KU665487). The phylogenetic status and nucleotide composition of mitogenome were assessed with MEGA 6 (Tamura *et al.,* 2013).

The mitogenome sequence obtained is a 16456 bp circle with 37 mitochondrial structural genes; two ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA (tRNA) genes, 13 protein-coding genes and 1 non-coding control region/D-loop (Fig. 1**,** Table 2). H-strand was the major coding strand but ND6 and eight tRNA genes were encoded on the L-strand. The gene order, gene length as well as heavy (H) and light (L) strand coding pattern are identical to that in other vertebrates (Boore, 1999). 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%). Similar to other vertebrates, low G content and high A+T (53.8%) content were observed in the genome (Table 3).

Table 1. List of primer pairs used for amplification of *E. suratensis* mitochondrial DNA

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

The 13 protein-coding genes altogether come around 11358 bp. Intergenic overlaps at ATP6 and ATP8 (10 nucleotides), ND4 and ND4L (7 nucleotides) and ND5 and ND6 (4 nucleotides) were observed in their overlapping region which is common within vertebrate mitogenomes and have been reported for several fish species (Boore, 1999; Mu *et al.,* 2015). ATG is used as start codon by all coding genes except CO1 (GTG is the start codon) and TAA was used as stop codon translation terminators for ND1, ND2, CO1, ATP8, ND4L and ND5. The remaining genes used incomplete stop codon TA/T--(Table 1) and post-transcriptional polyadenylation compensate adenosine nucleotide required for generating the stop codon (TAA) (Ojala *et al.,* 1981). The most frequently used amino acids were leucine (17.6%), followed by alanine (8.6%) and isoleucine (7.2%). The highest estimated RSCU were matched to corresponding tRNAs identified in the mitogenome, with the exception of alanine, glycine, leucine, methionine, proline, serine, threonine and valine (Table 4). In the third codon positions, codons complementary to the tRNAs ending in A and C were the most frequently observed and G nucleotide was the least frequent.

Similar to other vertebrates, *E. suratensis* rRNA genes have high adenine content (52.2%) (Boore, 1999) and 3 of the 22 tRNA genes identified showed overlaps.

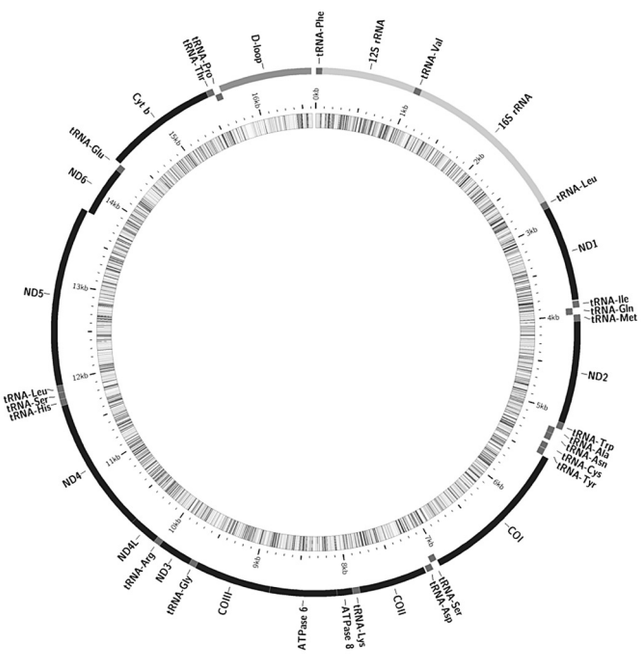

Fig. 1. Mitogenome map of *E. suratensis* (16456 bp) (Gen Bank Accession no. KU665487) generated with MitoAnnotator. Protein-coding genes, tRNAs, rRNAs and D-loop regions are shown in different colours. Genes located within the outer circle are coded on the H-strand whereas the remaining genes are coded on the L-strand.

Table 2. Features of the mitogenomes of *E. suratensis*

Gene	Position			Codon[b]	
	From (bp)	To (bp)	Strand[a]	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.

The origin of light strand replication (OL) in *E. suratensis* was located between tRNA Asn and tRNA Cys (WANCY region) and it is from 5303 bp to 5338 bp. WANCY region is a region coding for five mitochondrial tRNAs (tryptophan, alanine, asparagine, cysteine and tyrosine). OL sequence has the potential of forming a stable stem-loop structure in its single-stranded form, which is needed for the initiation of replication (Hixson *et al.,* 1986). A major non-coding region, control region (D-loop) located between the tRNA Pro and tRNA Phe genes (779 bp in size) has several characteristic conserved sequence blocks (CSB) like CSB1, CSB2, CSB3 and promoter region (Fig. 2).

In the phylogenetic tree, *E. suratensis* clustered with cichlids present in Indian and Sri Lankan waters along with one species from Madagascar group (*Paretroplus maculatus*). They formed sister group to all other cichlids (Fig. 3). In the family Cichlidae, species from South America and Africa are monophyletic in origin

Table 3. Nucleotide composition of the mitogenome of *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] Based on the ND 6 gene located on the L-strand; [c] Based on the 13 protein-coding genes.

Table 4. Amino acid and codon usage in mitogenome of *E. suratensis*

Amino acid	%[a]	Codons	RCSUC[b]	Amino acid	%[a]	Codons	RCSUC[b]
Alanine (Ala/A)	8.6	GCU	55	Leucine (Leu/L)	17.6	UUA*	74
		GCC	146			UUG	32
		GCA*	110			CUU	168
		GCG	14			CUC	179
Arginine (Arg/R)	2.0	CGU	11			CUA*	179
		CGC	14			CUG	32
		CGA*	45	Lysine (Lys/K)	1.9	AAA*	71
		CGG	4			AAG	4
Asparagine (Asn/N)	3.0	AAU	29	Methionine (Met/M)	3.9	AUA	104
		AAC*	88			AUG*	42
Aspartic acid (Asp/D)	1.8	GAU	18	Phenylalanine (Phe/F)	6.3	UUU	101
		GAC*	53			UUC*	137
Cysteine (Cys/C)	0.6	UGU	6	Proline (Pro/P)	5.8	CCU*	51
		UGC*	16			CCC	118
Glutamic acid (Glu/E)	2.6	GAA*	84			CCA	49
		GAG	14			CCG	5
Glutamine (Gln/Q)	2.5	CAA*	90	Serine (Ser/S)	6.6	UCU	46
		CAG	7			UCC	97
Glycine (Gly/G)	6.6	GGU	39			UCA*	47
		GGC	94			UCG	5
		GGA*	72			AGU	14
		GGG	38			AGC*	42
Histidine (His/H)	2.8	CAU	37	Threonine (Thr/T)	4.1	ACU	4
		CAC*	74			ACC	153
Isoleucine (Ile/I)	7.2	AUU	137			ACA*	116
		AUC*	148			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] Percentage of amino acid based on the 13 protein-coding genes; [b]RSCU relative synonymous codon usage; *Codons complementary to the tRNA genes.

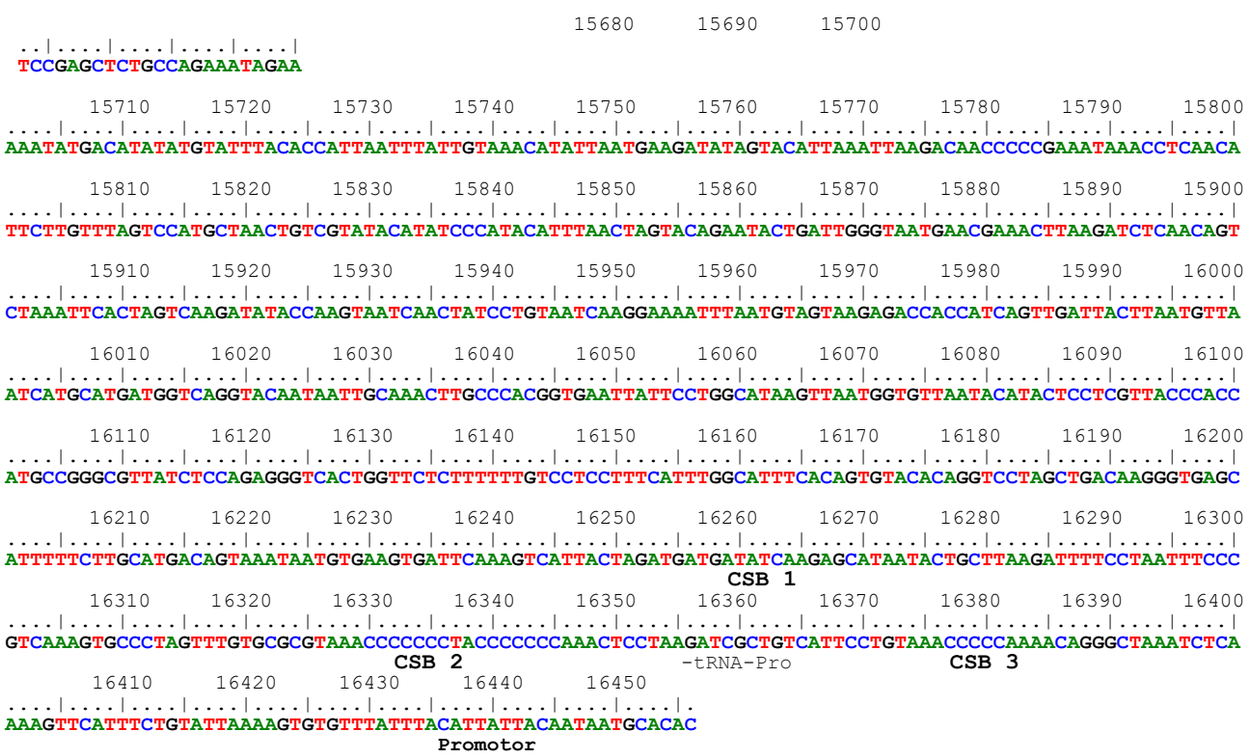

Fig. 2. Characteristic conserved blocks (CSB 1, CSB2, CSB3) and promoter region in the non-conding region (D-Loop) of *E. suratensis* mitochondrial DNA

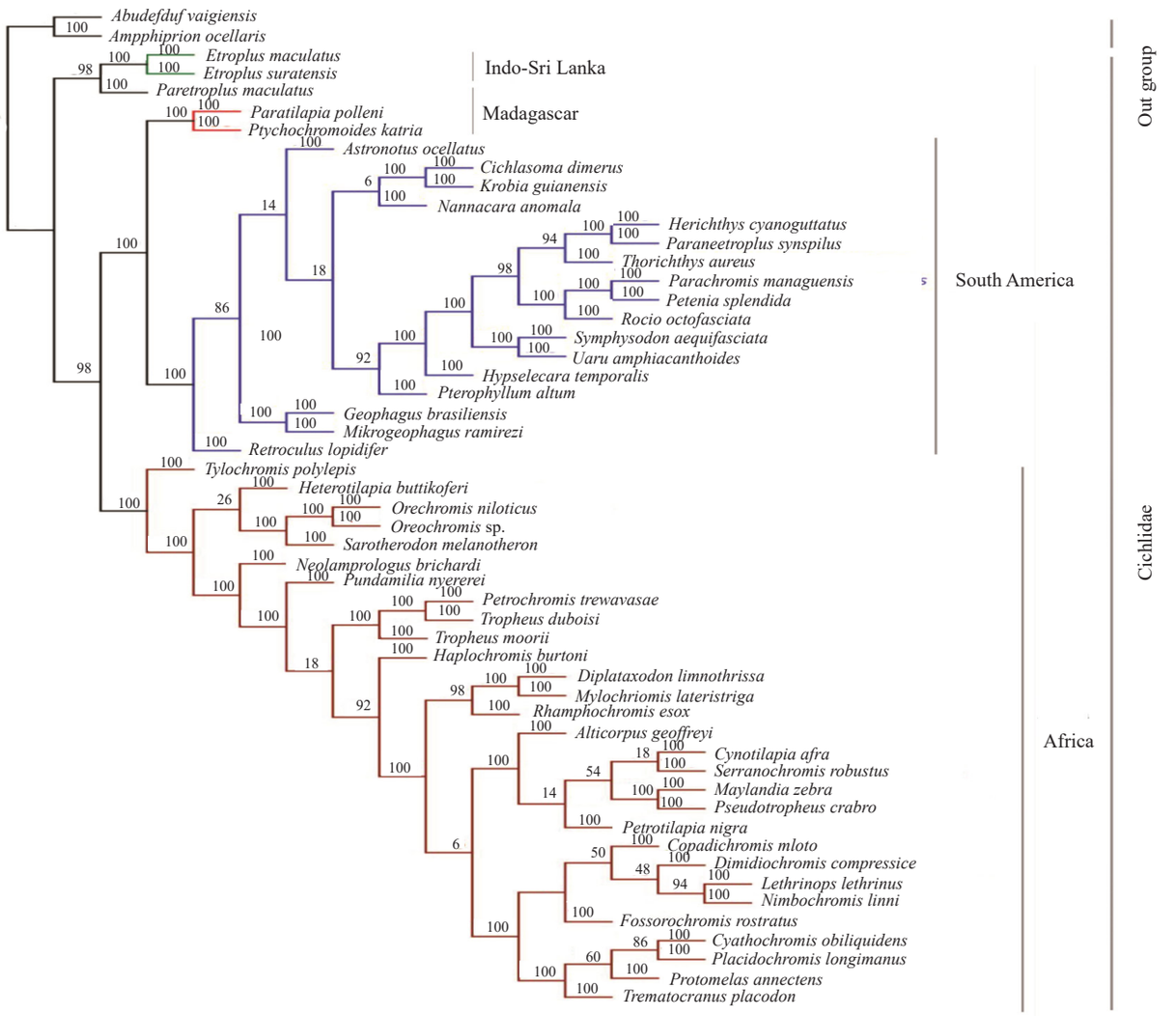

Fig. 3. Neighbor-joining phylogenetic tree generated by alignment of complete mitogenome nucleotide sequences of *E. suratensis a*nd other cichlids. Cichlid species which are represented from Africa, South America, Madagascar and Indo-Sri Lankan regions were used. *Amphiprion ocellaris* and *Abudefduf vaigiensis* were used as outgroups.

whereas Madagascar and Indo-Sri Lankan groups are not monophyletic. The tree also supported the proposed Gondwanan origin of Cichlidae as the divergence pattern of cichlids belonging to each continent was associated with the geological history of continental drift (Azuma *et al.,* 2008). Results of the complete mitogenome phylogeny of this study also strongly supported early diversification events within Cichlidae as well as Gondwanan origin of cichlid lineages as reported by Sparks and Smith (2004) with mitochondrial and nuclear gene fragments.

The mitogenomic information of *E. suratensis* generated in the present study will provide momentum to further studies on evolution, taxonomy, conservation, environmental adaptation and selective breeding of the species.

Acknowledgements

This work was carried out under the institute project MBTD/GEN/28 receiving funding support from the ICAR, New Delhi. The authors thank Dr P. Vijayagopal, Head, Marine Biotechnology Division, ICAR-CMFRI, and Director, ICAR-CMFRI for providing facilities for carrying out this work. First author received a Senior Research Fellowship from ICAR-NICRA project.

References

Azuma, Y., Kumazawa, Y., Miya, M., Mabuchi, K. and Nishida, M. 2008 Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC Evol. Biol.,* 8: 215. DOI: 10.1186/1471-2148-8-215.

Boore, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.,* 27(8): 1767-1780. DOI: 10.1093/nar/27.8.1767.

Caballero, S., Duchene, S., Garavito, M. F., Slikas, B. and Baker, C. S. 2015. Initial evidence for adaptive selection on the NADH subunit 2 of freshwater dolphins by analyses of mitochondrial genomes. *PloS one* 10(5): e0123543. DOI: 10.1371/journal.pone.0123543.

Chandrasekar, S., Nich, T., Tripathi, G., Sahu, N. P., Pal, A. K. and Dasgupta, S. 2014. Acclimation of brackishwater 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. DOI: 10.1007/s10695-013-9899-y.

Garvin, M. R., Bielawski, J. P. and Gharrett, A. J. 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. DOI: 10.1371/journal.pone.0024127.

Hixson, J. E., Wong, T. W. and Clayton, D. A. 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.

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M. and 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. DOI: doi.org/10.1093/molbev/mst141.

Jayaram, K. C. 2010. *The freshwater fishes of the Indian region*. Narendra Publishing House, New Delhi, India.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. and 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. DOI: 10.1093/bioinformatics/bts199.

Kumar, K., Raghavan, R., Prasad, G., Kumar, B., Sekharan, N. M., Pereira B. and Ali P. H. 2009. When pets become pests-exotic aquarium fishes and biological invasions in Kerala, India. *Curr. Sci.,* 25: 474-476.

Miya, M. and Nishida, M. 2015.The mitogenomic contributions to molecular phylogenetics and evolution of fishes: a 15year retrospect. *Ichthyol. Res,* 62(2): 29-71. DOI: 10.1007/s10228-014-0440-9.

Mohanta, S.K., Swain, S. K., Das, S. P., Bit, A., Das, G., Pradhan, S., Sundaray, J. K., Jayasankar, P., Ninawe, A. S. and Das, P. 2016. Complete mitochondrial genome sequence of *E. suratensis* revealed by next generation sequencing. *Mitochondr. DNA Part B.* 1(1): 746-747. DOI: 10.1080/23802359.2016.1176877.

Mu, X., Liu, Y., Lai, M., Song, H., Wang, X., Hu, Y. and 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. DOI: 10.1016/j.gene.2015.01.056.

Ojala, D., Montoya, J. and Attardi, G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290(5806):470.

Padmakumar, K. G., Bindu, L. and Manu, P. S. 2012. *Etroplus suratensis* (Bloch), the state fish of Kerala. *J. Biosci.,* 37(1): 925-931. DOI: 10.1007/s12038-012-9271-x.

Padmakumar, K. G., Krishnan, K., Manu, P. S., Shiny, C. K. and 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, p. 27-36.

Sambrook, J. and Russell, D. 2001. *Molecular cloning: A laboratory manual*, 3rd edn. Cold Spring Harbour Laboratory Press, New York, USA.

Sparks, J. S. and Smith, W. L. 2004 Phylogeny and biogeography of cichlid fishes (Teleostei: Perciformes: Cichlidae). *Cladistics,* 20(6): 501-517. DOI: org/10.1111/j.1096-0031.2004.00038.x.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.,* 30(12): 2725-2729. DOI: 10.1093/molbev/mst197.

Date of Receipt : 31.01.2019
Date of Acceptance : 11.07.2019