International Journal of Biology Research
ISSN: 2455-6548
Impact Factor: RJIF 5.22
www.biologyjournal.in
Volume 3; Issue 1; January 2018; Page No. 114-119

DNA divergence and conservational measures of fish fauna in Lake Kolleru based on partial Cytochrome Oxidase I (COI) sequences

Gatreddi Srinu
Department of Zoology & Aquaculture, Acharya Nagarjuna University, Andhra Pradesh, India

Abstract
Lake Kolleru is one of the largest fresh water lakes in India and the only RAMSAR site of Andhra Pradesh. Due to several anthropogenic activities, it is currently experiencing an alarming decline in fish biodiversity and many fishes were categorized as threatened. This emphasizes an immediate need for initiating research based on advanced molecular tools instead of traditional methods. In this study, biodiversity of fish fauna in lake Kolleru has been documented along with their changing patterns based on various studies. To evaluate the taxonomic status of Kolleru fishes and to compare them with their counterparts across India, Cytochrome Oxidase I (COI) based analysis was done for 15 genes available in NCBI GenBank. In total 137 sequences were analyzed. The COI barcodes clearly distinguished all the fish species with high inter-specific genetic distance values than intra-specific values based on K2P (Kimura 2 Parameter) model. In accordance to K2P model, the average genetic distance within species, Genus, Family and order was 0.16%, 2.45%, 5.30%, and 13.71% respectively. In addition to COI divergence estimation, four nucleotide compositions and GC content at three codon positions were calculated and compared with other retrieved COI sequences of different geological regions of the country. In conclusion, emphasis was given to conservational aspects of Lake Kolleru.

Keywords: lake kolleru, cytochrome oxidase I (COI), K2P (kimura 2 parameter)

Introduction

India is well known for its rich natural heritage and harbours a unique biodiversity. India ranks twelfth among 17 megadiverse countries of the world. Even though the exact number of fish species is not known and unexplored because of several taxonomic impediments [1], it was estimated that India housed 2508 fish species [2], of which 856 belonged to fresh water inhabitants [3]. India ranks ninth with a share of 8.9% in the world's fish diversity. Lake Kolleru (81º 5" to 81º 21" E and 16º 32" to 16º 47"N) is the largest natural fresh water lake in India with an extent of 245sq. km. harbouring a rich diversity of fish fauna. Kolleru wetland is a well known breeding ground for many riverine fishes as evident by the presence of robust number juvenile fishes especially the carps. Since Kolleru is connected to Bay of Bengal *via* Upputeru, some marine and migratory fishes are also found in this lake. The natural fishery of the lake is on the decline due to various man-made activities like over exploitation, heavy pollution due to letting in of industrial effluents into the lake, seasonal dry up and agricultural runoff with pesticides and excessive weed growth. In addition, due to shallow nature of the lake (with mean depth 2m.), much of the lake bed is being converted into myriads of fish ponds. At present lake Kolleru is the hub of aquaculture which makes Andhra Pradesh the second largest freshwater fish producing states in India. Pisciculture impacts lake Kolleru with ecological imbalance by introducing invasive species and polluting the lake with aquaculture remnants [4]. The presence of three endemic species, *Rohtee ogilbii*, *Hypselobarbus dobsoni* and *Thynnichthys sandkhol* [5] of the lake which are not encountered in recent times emphasized the need of conservation of Lake Fauna. Previous fish resource estimates emphasized on the presence of 4 endangered, 11 vulnerable and 1 rare species, which need to be conserved to protect them from extinct in near future. Most of the fishes of lake Kolleru are commercially important, of which 16 species are considered as threatened species and suggested to protect them to maintain the biodiversity of fishes as a whole [6, 7]. Evidences emphasising on extinction of several fish species in lake Kolleru [5], Indeed, future scenarios predicted that within next three decades one-third of all fresh water fishes may get vanished [8, 9] which is not exceptional for Kolleru as it contain number of threatened and endemic fish species.

Whole DNA sequence based diversity assessment, whether directly or indirectly by analysis of proteins was used for species discrimination studies almost 40 years ago [10]. Afterwards, single gene based analysis of ribosomal DNA was used extensively to investigate evolutionary relationships [11]. Recently, Mitochondrial DNA (mtDNA) dependent molecular systematic studies were dominated. Mitochondrial DNA, with the characteristic feature of fast evolution rate than the nuclear DNA, has recently been used to elucidate genetic relationships for many species [12, 13]. Mitochondrial Cytochrome Oxidase I (COI) gene based DNA Barcoding technique also gaining importance as a useful tool for investigating the genetic structure of species apart from its identification application [14]. DNA Barcoding data enabled the researchers to read the genetic information which can be used efficiently in accurate management of species of ecosystem importance. No previous studies were found on genetic diversity, phylogeny studies

based through molecular approaches in Lake Kolleru.

With the aim of filling the gap in the knowledge of current extant of fish diversity in lake Kolleru and its verifying its genetic structure, this study attempts to investigate genetic diversity of available COI sequence data of fishes. After extensive field survey, this study documented the factors that causing the habitat loss of Lake Kolleru. Then, investigation was done for genetic diversity for 15 fish population from Kolleru. Focus was made on examining the genetic distances between interspecific and intraspecific datasets of species and uses this data to explain their genetic characterization.

Materials and methods

Sequence analysis

A total of 16 sequences belonging to 15 fish species of lake Kolleru and 121 sequences of the same fishes from various geographical regions of India, for which the taxonomy and GenBank accession numbers are provided (Table 1). To provide a comprehensive data on fish fauna of Lake Kolleru we compared them with the same species from other territorial parts of Andhra Pradesh and other states. Cytochrome Oxidase I data of the lake Kolleru fishes was complemented by COI sequences from Genbank, as available on 15th October, 2017. Care was taken to avoid bias in divergence assessment by giving to preference to nearby geographical locations and then to distant geographical locations. Disequilibrium in taxa representation may results in skew in divergence distributions. We standardized comparisons of taxa to maximum of ten individuals per each species following the above criteria. Taxa with multiple denominations or taxonomic ranks and suspected sequences that are derived from misidentified were omitted. All over, 137 sequences those includes 15 species, 12 genera, 11 families and 5 orders were included in the analysis part of the total data.

Table 1: List of fish species from Kolleru Lake along with their GenBank accession numbers.

S. No.	Order	Family	Genus	Species	Accession Numbers*
1	Anguilliformes	Anguillidae	*Anguilla*	*bengalensis*	KR021973 JX887591, JX887590, JX260829 JX260828, JX260827, JX260826 JX260825, KP897130, KM875502
2			*Anguilla*	*bicolor bicolor*	KP979655 KM875505, KM875504,KM875503 KY067460, KF182304, AP007236
3	Cypriniformes	Cyprinidae	*Esomus*	*danrica*	KP939356 KX245065, FJ459490, JN673955 FJ459486, KU738848, KJ936709 KX266826, KY290080, KU171302
4			*Laubuca*	*laubuca*	KP939355 KT353103
5	Characiformes	Serrasalmidae	*Piaractus*	*mesopotamicus*	KM519156 JQ667515, KM519156, KM897518, GU701417, KM897143,KM897453 HQ420834, KP856756, KM519156
6	Siluriformes	Bagridae	*Mystus*	*bleekeri*	KP939357 JX983376, JX260918, JX260917 JX260916, KT896741, JN628904 KF824794, KF824797, KX266834
7		Clariidae	*Clarias*	*batrachus*	KM519157 JQ699207, KJ720696, JQ699208 KF742432, KJ959639, KF511567 FJ459459, JN628880, KF214293
8		Heteropneustidae	*Heteropneustes*	*fossilis*	KR021972 JX983313, JX983311, JX260882 JX260879, KX245084, JN628881 JN596578, GQ461897, GQ466395
9		Sisoridae	*Neotropius*	*atherinoides*	KP939358 JX901501, JN628927, KF824819 JN628910, KY290098, KY290041 KF824817, JN628890, JN628911
10	Perciformes	Eleotridae	*Eleotris*	*fusca*	KP979654 JX193751, KU692479, KT960773, KT960771, KT960769, KT960767 KT960768, KT960770, MF611583
11		Gobiidae	*Pseudapocryptes*	*elongatus*	KT124739 & KT124740 KT378133, LC010470, MF594617 LC010471, LC010472, LC010480 LC010481, LC010482
12		Channidae	*Channa*	*punctata*	KP979652 JX983251, JX260843, KY290125

					JN245990, EU342201, KX389275 KJ936637, KJ854469, KU761951
13			*Channa*	*striata*	KP979651 KP842452, KP842443, KJ538701 KJ538675, KY290120, KX389279 EU342204, HM117203, KP842455
14			*Channa*	*orientalis*	KP979653 JX983248, JX983245, KY290045 FJ459480, KJ936643, KJ937374 KJ847127, KF742420, KF742438
15		Cichlidae	*Etroplus*	*suratensis*	KP939359 FJ237544, JX260868, KP316238 KC858286, KF442186, KF442180 KF442165, KF372997, KF442191

*****Accession number in bold indicates sequences from Kolleru whereas remaining were other retrieved sequences

The COI partial gene sequences obtained for each species were assembled and end-trimmed to a homologous region to avoid errors during sequencing and those sequences subjected to aligned using Clustal W analysis tool [15]. Sequences with sufficient length *i.e.* 561 bp only were considered with the view of bringing uniformity in analysis across all species. To ensure homology in heterogeneous sequences, some sequences were trimmed. Sequences with more than 559 bp of length only were used for further analysis. To bring this homogeneity in some sequences, missing sequence parts were adopted from most conserved regions of the sequences available in NCBI GenBank for the same species. Nucleotide composition (A, T, G, C, GC1, GC2 & GC3) calculated for homologous end-trimmed sequences using MEGA V.7.0 [16] software (Arizona). Inter and intra species evolutionary divergences in various hierarchical levels were analysed using Kimura 2 Parameter method [17]. The variation was estimated following the bootstrap method with 1000 bootstrap replicate values. The pair-wise deletion option was selected to treat the gaps or missing data between each compared specimen. Finally, Mutation rates, polymorphic sites and genetic diversity in between the population as well as for the total population was estimated using Dna SP V.5.0 software [18] by defining Kolleru fish COI sequences as Population I and other sequences population dataset II.

Results

A total of 137 sequences were analysed from 15 species out of which 16 sequences were belongs to lake Kolleru covering 12 genera, 11 families and 6 orders. Number of sequences used in this study ranged from two (*Laubuca Laubuca*) for a species to ten (almost remaining all 14 species) based on the availability and uniform distribution geographically. Among listed species included this study 3 were vulnerable species, *viz.,. Mystus bleekeri*, *Clarias batrachus*, *Heteropneustes tossilis* and one is endangered species, *viz., Anguilla bengalensis.* [22]. Four out the listed 15 species *i.e. Channa orientalis, C. punctatus, C. striatus* and *Esomun danrica* were listed in Conservation Assessment and Management Plan (CAMP) workshop report [19]. The report pointed the activities like fishing, loss of habitat, over exploitation and trade are culprits behind the gradual decrease in number of the species in lake Kolleru. These are only available sequences for fish fauna in this lake. The lengths of the available sequences ranged from 590 to 692 with an average of 629 base pairs. Furthermore, no Indels (insertions and deletions) were found in the sequences and also no stop codons were existed in their ExPASy translation stressing that all the available sequences code for functional mitochondrial COI gene and the were no NUMTS (Nuclear Mitochondrial DNA).

Table 2: Summary of COI genetic divergence (K2P percentage) of Kolleru fishes within various taxonomic levels

Comparison within taxonomic level	Mean± S.E.	Maximum
Species	0.16±0.40	2.30
Genera	2.45±1.32	20.4
Family	5.30± 1.90	31.3
Order	13.71±2.400	31.3

Table 3: Nucleotide composition of COI gene in different orders of Kolleru fishes

Order	A	G	C	T	GC	GC1	GC2	GC3
Anguilliformes	26.5±0.021	17.7±0.026	26.7±0.087	29.1±0.083	44.4±0.084	33.7±0.286	57.2±0.089	42.3±0.000
Cypriniformes	25.1±0.517	17.3±0.295	24.9±0.079	32.7±0.178	42.2±0.252	30.0±0.836	53.4±0.173	43.2±0.144
Characiformes	23.6±0.000	18.7±0.000	28.3±0.000	29.5±0.000	47.0±0.000	41.8±0.000	55.6±0.000	43.6±0.000
Siluriformes	26.7±0.107	16.8±0.123	26.7±0.130	29.8±0.134	43.5±0.175	42.4±0.109	31.7±0.655	56.4±0.183
Perciformes	23.6±0.149	17.6±0.133	29.3±1.226	29.5±1.317	46.9±0.259	43.6±0.135	41.6±0.754	55.5±0.226

Average genetic divergence within species was 0.16% compared with 2.45% for species within genera, while genetic divergence within family and order were 5.30% and 13.71% respectively (Table 2). As an interpretation to these values, increasing genetic divergence was observed with increasing taxonomic levels, indicating noticeable change in genetic divergence at the species boundaries. The average congeneric distance is approximately 15- fold the average conspecific distance. The average nucleotide frequencies for all 15 species are as follows: A= 25.0%, T= 29.6%, G= 17.4%, C= 27.9% (Table 3). The minimum, mean and maximum GC% content for three codon positions of all 15 species given in Figure 1. According to this, unlike other fishes, COI gene data of lake Kolleru contains dominated GC3 content followed by GC1

and GC2 (except for *E. fusca* and *C. orientalis* in which GC2 follows GC3) whereas in other fish fauna, GC2 content occupied majorly followed by GC1 and GC3 differently in different fishes. In overall, the mean GC content values in GC1, GC2 and GC3 codon positions are 44.3, 35.6 and 55.8 respectively, contrast to other fish fauna as 42.7, 48.9, and 42.5 respectively (Table 4).

Table 4: Comparison of Nucleotide frequencies of COI gene between Kolleru fishes and other retrieved sequences from different territorial parts.

	Within Lake Kolleru				Out of lake Kolleru			
	Min.	**Mean**	**Max**	**SE**	**Min.**	**Mean**	**Max**	**SE**
G%	15.9	17.3	18.3	0.164	15.6	17.1	19.8	0.192
C%	25.5	27.9	30.8	0.311	23.5	27.6	30.6	0.325
A%	22.7	25.1	27.8	0.325	22.5	25.3	28.9	0.336
T%	27.4	29.7	32.1	0.248	27.9	30.0	33.2	0.171
GC%	41.4	45.2	49.1	0.237	39.1	44.7	50.4	0.258
GC% Codon Position 1	42.7	44.3	46.1	0.110	34.6	42.7	51.6	0.587
GC% Codon Position 2	25.4	35.6	47.3	0.945	42.3	48.9	56.2	0.359
GC% Codon Position 3	53.2	55.8	57.6	0.201	34.8	42.5	51.1	0.465

Based on this, the difference was attributable to the GC content of the 2nd and 3rd codon base positions. The variation at the GC1 codon position between Kolleru and other fish species is very limited. Order wise nucleotide frequencies along with GC content in various codon positions were given in Table 3.

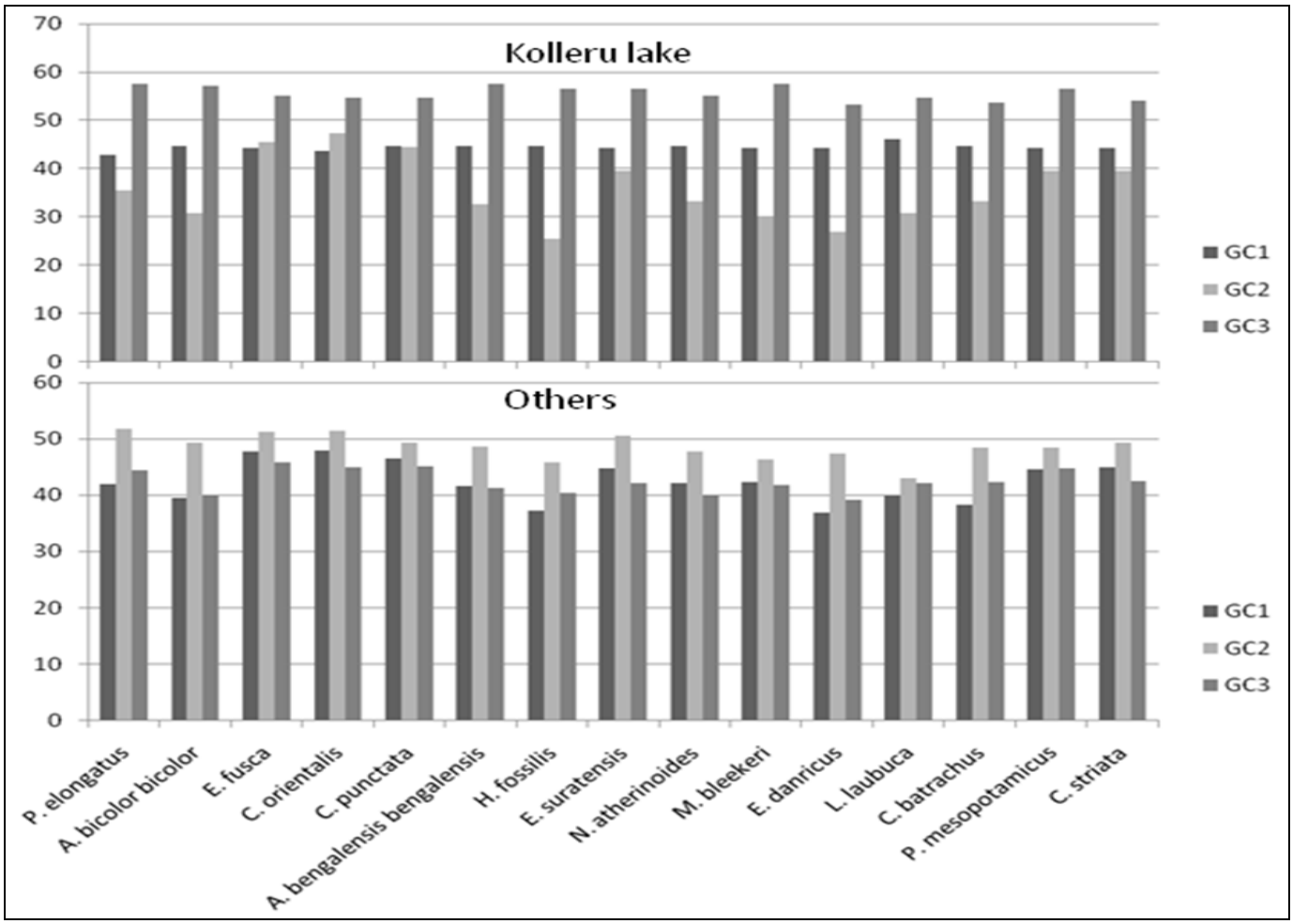

Fig 1: Variation in GC content (%) of COI gene between Kolleru fishes and other retrieved sequences from different teriitorial parts.

Number of mutation and polymorphic sites were showed drastic differences in different species of the total population to understand change in population through the time. Highest mutation rate among all the species was observed in *A. bicolor bicolor* followed by *E. fusca*, C. *batrachus*, C. *orientalis* and *A. Bengalensis* whereas least number of mutations were noticed in *E. Danricus*followed by *E. Suratensis*, *M. Bleekeri*, *H. Fossilis and P. Mesopotamicus*. The same pattern was noticed in nucleotide diversity as well as differences between populations Table 5.

Table 5: DNA Divergence values between fish populations of Kolleru and other retrieved sequences.

S. No.	Name of the Sps.	No. of polymorphic sites	No. of mutations	Avg. no. of Nt. Diff (k) in total pop	Avg. no. of Nt. Diff. between pop.	Nt. diversity in total pop
1	*A. bengalensis*	51	55	11.5	8.0	0.021
2	*A. bicolor bicolor*	308	314	105.4	64.8	0.180
3	*E. danricus*	6	6	1.3	0.7	0.002
4	*L. laubuca*	15	15	10	7.5	0.016
5	*P. mesopotamicus*	21	21	5.4	4.0	0.008
6	*M. bleekeri*	17	17	3.8	2.5	0.006
7	*C. batrachus*	110	118	28.0	18.3	0.046
8	*H. fossilis*	20	20	4.7	3.4	0.007
9	*N. atherinoides*	34	34	12.4	9.1	0.020
10	*E. fusca*	122	141	47.8	62.5	0.077
11	*P. elongatus*	25	25	9.1	15.4	0.027
12	*C. punctata*	67	71	16.2	11.1	0.026
13	*C. striata*	28	29	9.6	6.1	0.016
14	*C. orientalis*	84	91	28.5	22.3	0.040
15	*E. suratensis*	11	11	2.2	1.2	0.003

Discussion

The present study represents the first genetic analysis for the ichthyofauna of Kolleru lake, the only RAMSAR site of Andhra Pradesh and the biggest fresh water lake in India. The study shows the effectiveness of COI in the utility of in genetic divergence and phylogenetic studies. The average conspecific, congeneric values observed in the present are lower than previous study reports both fresh water fishes [20] and this is the same scenario with marine fishes [21]. But, even this inter hierarchical divergence values successfully deciphered the fish fauna lake Kolleru. This may contributes in defining threshold levels of interspecies demarcation especially in case of lakes with conservational importance. All the COI sequences used in this study from Kolleru were simple sequences and without any ambiguities. No indels were observed among any of the sequences, which proves the absence of NUMTS (nuclear DNA sequences that are originated from mtDNA sequences and typically less than 600bp in length) as evidenced from some of the previous reports [21, 22] confirms the presence of NUMTS. Results of the current study were similar to Viswambharan *et al*. (2015) [23] with respect of absence of NUMTS. Based on GC content in COI of lake Kolleru, it is evident that substantially more nucleotide change at the 3rd codon position than 1st and 2nd. This reflects the fact that most synonymous mutations occur at the 3rd position with a few synonymous mutations at 1st position and almost negligible change at 2nd position [5]. Mutational rate of the mitochondria proved to be higher than the nuclear genome [24] along with the time course in the evolution [25]. In specific H-strand of the mt genome has more mutational rate [26] where COI gene is located than L-strand. This may be attributes to varying degree of asymmetric skew in the frequencies of bases at codon position 3 [27] with the favour of most occurring type of mutations [28].

Conclusion

The current study is first report for lake Kolleru, with the application of molecular data for estimation of its conservational status in icthtyofauna diversity. Well established pisciculture in lake Kolleru belt areas may result in introduction of Invasive Alien Species (IAS) like *O. nilotics* as these are these are the species extensively being culturing in Kolleru area results in severe disruptions in the ecology of the lake in several aspects. It is further emphasized that by making the molecular information on lake Kolleru readily available to non-taxonomists, it is possible to help accurate biodiversity assessment in future studies.

References

1. Hoagland KE. The taxonomic impediment and the convention on biodiversity. Association of Systematics Collections Newsletter. 1996; 24(5):61-62.
2. Eschmeyer WN. The catalog of fishes on-line. California Academy of Sciences, San Francisco, California. Link, 2003. http://www.calacademy. org/research/ichthyology/cat alog/fishcatsearch.html.
3. Froese R, Pauly D. Fishbase. World Wide Web Electronic Publications. Available at, 2012. http://www. fishbase.org (accessed June 2012).
4. Woodford DJ, McIntosh AR. Evidence of source–sink metapopulations in a vulnerable native galaxiid fish driven by introduced trout. Ecological Applications. 2010; 20(4):967-977.
5. Barman RP. The Fishes of the Kolleru Lake, Andhra Pradesh, India with Comments on their Conservation. Records of the Zoological Survey of India. 2004; 103(1-2):83-89.
6. Fresh Water Fishes of India. Conservation Assessment and Management Plan (CAMP) workshop report, Edited by Sanjay Molur and Sally walker, published by Zoo outreach organization, CBSG, India, 1998, 156.
7. Menon AGK. Checklist of the Freshwater Fishes of India. Rec. zool. Surv. India, Occ. Paper Zoological Survey of India, Govt. of India. 1999; 175:366.
8. Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, *et al*. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological reviews. 2006; 81(2):163-182.
9. Stiassny ML. The medium is the message: freshwater biodiversity in peril. The living planet in crisis: Biodiversity science and policy, 1999, 53-71.
10. Manwell C, Baker CMA. A sibling species of sea-

cucumber discovered by starch-gel electrophoresis. Comp. Biochem. Physiol. 1963; 10:39-53.
11. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl Acad. Sci. USA. 1977; 97:8392-8396.
12. Arnason U, Gullberg A, Widegren B. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. Journal of Molecular Evolution. 1991; 33(6):556-568.
13. Keskin E, Can A. Phylogenetic relationships among four species and a sub-species of Mullidae (Actinopterygii; Perciformes) based on mitochondrial cytochrome B, 12S rRNA and cytochrome oxidase II genes. Biochemical Systematics and Ecology. 2009; 37(5):653-661.
14. Hajibabaei M, Singer GA, Hebert PDN, Hickey DA. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics. 2007; 23(4):167-172.
15. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic acids research. 1997; 25(24):4876-4882.
16. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution. 2016; 33(7):1870-1874.
17. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution. 1980; 16:111-120.
18. Librado P, Rozas J. Dna SP. v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25(11):1451-1452.
19. Fresh Water Fishes of India Conservation Assessment and Management Plan (CAMP) workshop report, Edited by Sanjay Molur and Sally walker, published by by Zoo outreach organization, CBSG, India 1998, 156.
20. Chakraborty M, Ghosh SK. An assessment of the DNA barcodes of Indian freshwater fishes. Gene. 2014; 537(1): 20-28.
21. Lakra WS, Verma MS, Goswami M, Lal KK, Mohindra V, Punia P, *et al*. DNA barcoding Indian marine fishes. Molecular Ecology Resources. 2011; 11(1):60-71.
22. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2005; 360(1462):1847-1857.
23. Viswambharan D, Pavan-Kumar A, Singh DP, Jaiswar AK, Chakraborty SK, Nair JR, *et al*. DNA barcoding of gobiid fishes Perciformes, Gobioidei. Mitochondrial DNA. 2015; 26(1):15-19.
24. Oleinik AG. On the mutation rates of the mitochondrial and nuclear genomes of salmonid fishes. Russian Journal of Marine Biology. 2000; 26(6):432-438.
25. Lynch M. Evolution of the mutation rate. Trends in Genetics. 2010; 26(8):345-352.
26. Bielawski JP, Gold JR. Mutation patterns of mitochondrial H-and L-strand DNA in closely related cyprinid fishes. Genetics. 2002; 161(4):1589-1597.
27. Faith JJ, Pollock DD. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. Genetics. 2003; 165(2):735-745.
28. Rocha EP, Touchon M, Feil EJ. Similar compositional biases are caused by very different mutational effects. Genome research. 2006; 16(12):1537-1547.