

DNA DIVERGENCE AND PHYLOGENETIC ANALYSIS OF *EPINEPHELUS* SPP. (PERCIFORMES, SERRANIDAE) BASED ON COI SEQUENCES FROM INDIAN WATERS

Gatreddi Srinu, Darwin Chatla* and Suneetha Kola
 Department of Zoology & Aquaculture, Acharya
 Nagarjuna University, Guntur, A.P., India.

*Corresponding author's

ABSTRACT

Despite the ecological and commercial importance of Groupers (fish subfamily Epinephelinae), their taxonomy and systematics are still much debated. Reasons for this are the low level of morphometric variability and the relatively poor phylogenetic information borne by the morpho-anatomical characters used so far in diagnosing species. The present study was carried out to find genetic divergence of six species in the genus *Epinephelus* and to find the phylogenetic signal possessed by using 88 partial mitochondrial Cytochrome Oxidase I (COI) sequences belong to Indian waters. Our results showed that the mean inter-lineage distance was remarkably higher than mean intra-lineage distance based on K2-P (Kimura 2-Parameter) model which is crucial in deciphering the species boundaries. Also, the Neighbour-Joining tree revealed distinct clusters shared by sequences of same species of the genus *Epinephelus*.

KEYWORDS- *Cytochrome Oxidase I (COI) sequence, Epinephelus, Groupers, Phylogenetic signal, Genetic divergence.*

INTRODUCTION

India is one of the mega-diversity nations well known for its biodiversity richness, which is truly reflected by fishes. It was estimated that India housed 3231 valid species belong to fresh and marine waters. Of this total fish diversity, marine fishes constitute 2443 species, which is about 75.6 percent (1). Serranidae is the one of the largest family of marine fishes comprising three subfamilies: Epinephelinae, Serraninae and Anthiaginae (2, 3, 4). The first subfamily recognized as groupers composed of 16 genera with 203 valid species which are distributed world-wide in the tropical and sub-tropical waters (5). Around 69 species of the subfamily Epinephelinae and 31 species of the genus *Epinephelus* have been reported from the Indian waters. It populates a diverse range of habitats mainly coral reefs, rocky areas, sea grass beds and estuaries (6) also exhibits a wide variety of reproductive and growth strategies (7, 8, 9). These are among the most marketable group of tropical marine fishes with high prices at live sea food markets in many parts of the world. Groupers are usually identified based on their color patterns and morpho-meristic characteristics (10). However, identification is not easy as they overlap with many characters of related species. Although some of the recent studies have helped to clear the status of some nominal species, confusion still surrounds the identification of many common species of *Epinephelus*. Studies based on COI sequence divergence of mitochondrial DNA are increasingly used to reinforce our knowledge on the existing taxonomic relationships and species identifications carried out based on morphological characters and biometric data (11). Hence, the present study is an attempt made to evaluate genetic divergence and find out the phylogenetic signal for six species in the genus *Epinephelus* from Indian waters.

MATERIALS AND METHODS

Partial COI gene sequences were obtained from NCBI-GenBank by giving appropriate input terms

as on 25th January, 2018. In total, 88 sequences representing six species of genus *Epinephelus* were collected, in which the species with minimum number of sequences was *E. bleekeri* (n = 10) and maximum was *E. coioides* (n = 18) (Table 1). Species with less than 10 sequences were omitted. All sequences were assembled and end-trimmed to get homologous region in order to avoid errors during sequencing and they were aligned using ClustalW analysis tool (12). Only the sequences with sufficient length were considered in the view of bringing uniformity in analysis across all species. To ensure homology in heterogeneous sequences, some bases were trimmed. To bring this homogeneity in some sequences, missing 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 (Molecular Evolutionary Genetic Analysis) (13) software (Arizona). Inter and intra species evolutionary divergences in various hierarchical levels were analysed using Kimura 2- Parameter method (14). The variation was estimated following the bootstrap method with 5000 bootstrap replicate values (15). The pair-wise deletion option was selected to treat the gaps or missing data between each compared specimen. Number of Haplotypes (H) along with Haplotype diversity (Hd) and also the number of Polymorphic sites for six *Epinephelus* species were estimated using Dna SP V.5.0 software (12). Finally, the Neighbour-Joining (N-J) tree among species was created to give evolutionary distance values using bootstrap method in Kimura 2- parameter mode of analysis and the values were represented in terms of number of base substitutions per site (16). Totally, 619 positions in the final dataset were present. To verify the robustness of the nodes of the N-J tree, bootstrap analysis was carried out using 5000 pseudo replicates (15, 17). Both transitions and transversions were cumulated and included as substitutions. Codon positions included were 1st+2nd+3rd+Noncoding. Missing bases or gaps were treated by adopting pair wise deletion method employed in MEGA V 7.0 tool (13).

RESULTS AND DISCUSSION

A total of 88 sequences were retrieved and analysed for six species of the genus *Epinephelus* which are territorially belongs to Indian region. All the retrieved sequences were verified thoroughly and no complexity or ambiguities were observed among them. The lengths of the available sequences ranged from 613 to 649 with an average of 637 bases per sequence. The average length values of the aligned and end-trimmed sequences were presented in Table 1. Furthermore, no Indels (Insertions and Deletions) were found in the sequences and also no stop codons were existed in their ExPASy translation emphasized that all the available sequences code for functional mitochondrial COI gene and there were no NUMTS (Nuclear Mitochondrial DNA).

S.No.	Species(N)	GenBank accession no. & Source	Length
1	<i>Epinephelus coioides</i> (18)	KJ607965, KX090374, KM891742-46, KM226246-48, KJ755858, JX674986-89, JX674982-84	613
2	<i>Epinephelus longispinis</i> (17)	HQ658119, KJ607970, KJ920103, EF609521-22, KM079298, KM226271-76, JX675001-05	643
3	<i>Epinephelus fasciatus</i> (15)	FJ459561-63, EU392207-08, KT835683-84, KM226261-65, JX674994-96	649
4	<i>Epinephelus diacanthus</i> (15)	KJ607971, KJ920101, EF609519-20, EF609516-17, KM079296-97, KM226249-53, HQ589272, EF609518	623
5	<i>Epinephelus chlorostigma</i> (13)	EU392202-04, EF609514-15, KT835685-86, KM226244-45, JX674979-81, KF434771	649
6	<i>Epinephelus bleekeri</i> (10)	MF978166, KM226239-43, JX674970-73	648

Table 1. List of retrieved COI GenBank accession numbers for six *Epinephelus* species (N = No. of Sequences)

The nucleotide frequencies for six *Epinephelus* species were represented in Table 2. The percentage nucleotide values for COI gene were A= 24.1%, T= 29.2%, G= 18.0%, C= 28.3%. Average A/T content (53.5%) is higher than the G/C content (46.5%) for all six *Epinephelus* species. The maximum and minimum values of A/T content observed were 55.0% (*Epinephelus bleekeri*) and 51.4% (*Epinephelus chlorostigma*) respectively. The

calculated AT content in this study is slightly lower than AT content (54.96%) in protein coding genes of mitochondria in hybrid *Epinephelus* species suggested by Gao et al. (18). The maximum and minimum values of G/C content observed were 48.6% (*Epinephelus chlorostigma*) and 45.0% (*Epinephelus bleekeri*) respectively. The comparative analysis for A/T and G/C content values for six *Epinephelus* species revealed that COI gene did not undergo recent evolution (19) and were free from recombination (20). Min & Hickey (2007) (21) showed strong correlation between the GC content of COI gene and the entire mitochondrial genome. The overall GC content across six species (46.5%) was slightly higher than reported GC content of marine teleost fishes (22, 23).

S.No.	Species	A	T	G	C
1	<i>Epinephelus coioides</i>	24.9±0.015	29.7±0.011	17.0±0.009	28.4±0.016
2	<i>Epinephelus longispinis</i>	24.5±0.031	29.3±0.025	19.1±0.035	27.1±0.036
3	<i>Epinephelus fasciatus</i>	24.2±0.028	29.4±0.023	17.7±0.018	28.0±0.024
4	<i>Epinephelus diacanthus</i>	22.9±0.087	29.4±0.015	18.4±0.123	29.2±0.027
5	<i>Epinephelus chlorostigma</i>	23.9±0.064	27.5±0.159	18.8±0.064	29.8±0.127
6	<i>Epinephelus bleekeri</i>	24.7±0.015	30.3±0.014	17.5±0.032	27.5±0.000

Table 2. Calculated nucleotide frequencies for 4 nucleotides along with their S.E. values

The average GC content in 3 codon positions (GC1, GC2 and GC3) of partial COI sequence for six *Epinephelus* species were 44.0%, 45.8% and 49.3% respectively (Table 3). Overall, GC3 content is dominated followed by GC2 and GC1. This indicates that the synonymous mutations occur mostly at the GC3 followed by GC2 and GC1. This synonymous mutation does not affect resulted protein. The highest percentage of GC3 content for COI gene was noticed in *Epinephelus bleekeri* (56.9%) whereas the lowest value was observed in *Epinephelus longispinis* (39.0%).

S.No.	Species	(G+C)1± S.E.	(G+C)2± S.E.	(G+C)3± S.E.
1	<i>Epinephelus coioides</i>	42.9±0.025	37.3±0.025	56.0±0.025
2	<i>Epinephelus longispinis</i>	56.0±0.061	43.5±0.000	39.0±0.122
3	<i>Epinephelus fasciatus</i>	36.3±0.083	57.8±0.015	43.1±0.000
4	<i>Epinephelus diacanthus</i>	42.2±0.033	42.9±0.095	58.0±0.105
5	<i>Epinephelus chlorostigma</i>	43.9±0.311	58.7±0.050	43.1±0.000
6	<i>Epinephelus bleekeri</i>	43.2±0.036	34.6±0.038	56.9±0.000

Table 3. Calculated GC content in 3 codon positions of COI sequences along with S.E.values

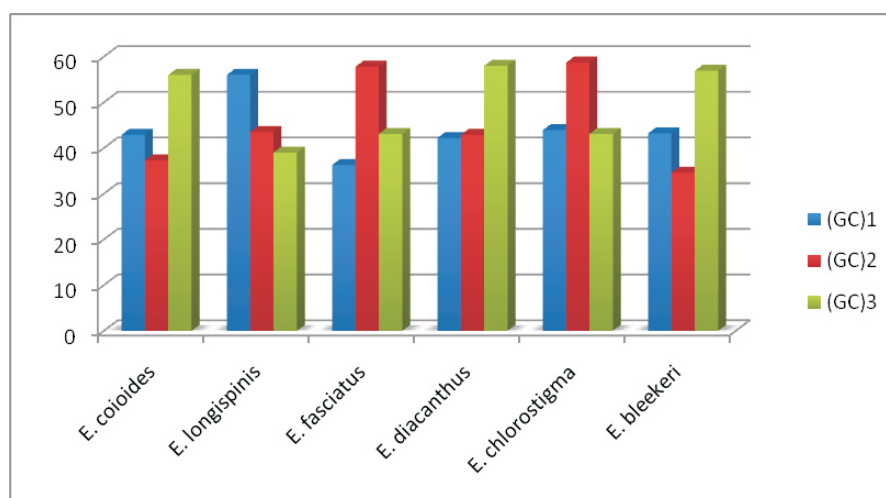


Figure 1. Variation in GC content (%) of COI sequences in 3 codon positions

Altogether, a total of 39 haplotypes recorded in six *Epinephelus* species with 0.793 as an average haplotype diversity. The highest number of haplotypes were observed in *Epinephelus fasciatus* and *Epinephelus chlorostigma* whereas the minimum number were observed in *Epinephelus bleekeri*. Among six *Epinephelus* species, the highest number of polymorphic sites were noticed in *Epinephelus chlorostigma* (78) whereas the least number were observed in *Epinephelus bleekeri* (Table 4).

S.No.	Species	No. of Haplotypes (Hd)	Polymorphic sites
1	<i>Epinephelus coioides</i>	6 (0.562)	10
2	<i>Epinephelus longispinis</i>	7 (0.860)	15
3	<i>Epinephelus fasciatus</i>	8 (0.867)	10
4	<i>Epinephelus diacanthus</i>	7 (0.829)	18
5	<i>Epinephelus chlorostigma</i>	8 (0.910)	78
6	<i>Epinephelus bleekeri</i>	3 (0.733)	02

Table 4. Calculated no. of Haplotypes and Polymorphic sites for COI sequences (Hd= Haplotype diversity)

Average genetic divergence values of COI partial sequences for *Epinephelus coioides*, *Epinephelus longispinis*, *Epinephelus fasciatus*, *Epinephelus diacanthus*, *Epinephelus chlorostigma* and *Epinephelus bleekeri* were 0.002%, 0.006%, 0.003%, 0.008%, 0.032% and 0.002% respectively. The maximum value of inter species genetic divergence observed in *Epinephelus diacanthus* whereas the minimum value was noticed in *Epinephelus coioides*.

S.No.	Species	A.D. ± S.E.	Min	Max
1	<i>Epinephelus coioides</i>	0.002±0.001	0.000	0.012
2	<i>Epinephelus longispinis</i>	0.006±0.002	0.000	0.019
3	<i>Epinephelus fasciatus</i>	0.003±0.001	0.000	0.009
4	<i>Epinephelus diacanthus</i>	0.008±0.002	0.000	0.018
5	<i>Epinephelus chlorostigma</i>	0.032±0.004	0.000	0.108
6	<i>Epinephelus bleekeri</i>	0.002±0.001	0.000	0.003

Table 5. Kimura 2-parameter distance values within 6 *Epinephelus* species (AD= Average distance, SE= Standard Error)

Pair wise genetic distance matrix of six *Epinephelus* species for COI partial sequences were represented in Table 6. The highest Kimura's 2-parameter (K-2P) genetic distance for COI gene were observed between *Epinephelus coioides* and *Epinephelus chlorostigma* (0.164%) whereas the lowest value found between *Epinephelus longispinis* and *Epinephelus bleekeri* (0.112%). The average value (0.008) of mean inter-lineage distance (Table 5) were remarkably higher than the mean intra-lineage distance (0.127) (Table 6) and even exceeded 15x threshold, which has been suggested by Hebert for new animal species identification (24).

S.No.	Species	EC	EL	EF	ED	ECh	EB
1	<i>Epinephelus coioides</i> (EC)		0.018	0.017	0.018	0.017	0.017
2	<i>Epinephelus longispinis</i> (EL)	0.163		0.018	0.017	0.016	0.015
3	<i>Epinephelus fasciatus</i> (EF)	0.153	0.157		0.018	0.016	0.016
4	<i>Epinephelus diacanthus</i> (ED)	0.159	0.153	0.161		0.016	0.017
5	<i>Epinephelus chlorostigma</i> (ECh)	0.164	0.136	0.149	0.147		0.013
6	<i>Epinephelus bleekeri</i> (EB)	0.152	0.112	0.138	0.153	0.113	

Table 6. Estimates of Inter species pair-wise genetic distances using K-2P model for 6 *Epinephelus* species. Standard error estimates are shown in bold.

The Neighbour-Joining tree for six *Epinephelus* species derived by using all 88 partial sequences using MEGA 6.0 are shown in Figure 2. Branches corresponding to partitions reproduced in less than 50% bootstrap

replicates are collapsed. All six *Epinephelus* species formed monophyletic groups with clear pattern where clades formed by the same species with significant boot strap values. This study revealed that species *Epinephelus diacanthus* and *Epinephelus coioides* were most closely related species followed by *Epinephelus chlorostigma* and *Epinephelus bleekeri*. Among the six species, *Epinephelus fasciatus* proved as more distantly related to remaining groups.

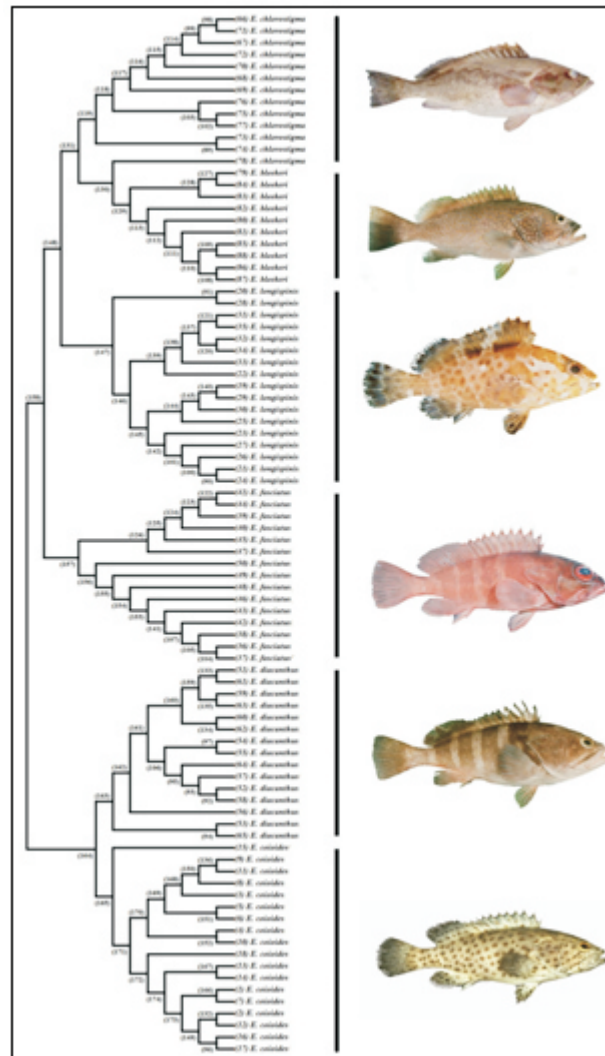


Figure 2. Neighbour-Joining tree with Kimura 2-Parameter model

Groupers are the important resources target by the coastal fisheries in tropical and subtropical areas (10). There are several reports regarding identification of groupers in Indian waters using traditional taxonomic tools (26, 27). However, the *Epinephelus* species are often incorrectly identified in the field because of their closely related morphological features. Given the widespread use of molecular genetic techniques during these decades, phylogenetic relationship of groupers has been well resolved (28, 29, 30), and some studies about DNA barcoding of groupers have also been published (31, 32, 33, 34, 11). Baharum and Nurdalila (35) carried out cytochrome b (cyt b) analysis as molecular marker to investigate the molecular relationship between *Epinephelus hexagonatus* and *Epinephelus fuscoguttatus*. The analysis from the BLAST database of partial sequences of mtDNA cyt b of *Epinephelus fuscoguttatus* showed 99% similarity to the *Epinephelus hexagonatus* and these two species are closely related and their molecular marker result was supported by the level of differentiation on morphology of both species is very low.

CONCLUSION

A total of 88 partial COI sequences belong to six species in the genus *Epinephelus* were analysed in this study. From the results, it is concluded that COI sequences can be used for evolutionary divergence estimation studies for groupers and phylogenetic studies as well. COI sequence based Neighbour-Joining tree revealed remarkably distinct clusters shared by sequences of same species of the genus *Epinephelus*. This study offers important resources for the further study of grouper conservation genetics, speciation, hybridization and other evolutionary traits.

ACKNOWLEDGEMENTS

The authors are grateful to the UGC for granting BSR (Basic Scientific Research) fellowship and the authorities of Acharya Nagarjuna University for providing necessary facilities to carry out this work in the Department of Zoology & Aquaculture under the financial support of UGC-SAP- DRS Phase III.

REFERENCES

- Gopi, K.C., & Mishra, S.S. (2015). Diversity of marine fish of India. In Marine Faunal Diversity in India (pp. 171-193).
- Johnson, G.D. (1983). *Nippon spinosus*: a primitive epinepheline serranid, with comments on the monophyly and intrarelationships of the Serranidae. *Copeia*, 777-787.
- Van Der Laan, R., Eschmeyer, W.N., & Fricke, R. (2014). Family-group names of recent fishes. *Zootaxa*, 3882(1): 1-230. <http://dx.doi.org/10.11646/zootaxa.3882.1.1>
- Van der Laan, R., Eschmeyer, W.N. & Fricke, R. (2016). Addenda to Family-group names in Recent Fishes. *Zootaxa*, 3882. (1): 1–230.
- Froese, R. and Pauly, D., FishBase, World Wide Web electronic publication (2017), www.fishbase.org, version (10/2017).
- Rao, D.V. (2009). Checklist of fishes of Andaman and Nicobar Islands, Bay of Bengal. *Environment and Ecology*, 27(1A): 334-353.
- Johannes, R.E. (1978). Reproductive strategies of coastal marine fishes in the tropics. *Environmental biology of fishes*, 3(1): 65-84.
- Shapiro, D.Y. 1987. Reproduction in Groupers. In: *Tropical Snappers and Groupers: Biology and Fisheries Management* (Eds. Polovina, J. J. and Ralston, S.), pp: 295, 327, Westview Press, Boulder, Colorado, USA.
- Morris, A.V., Roberts, C.M., & Hawkins, J.P. (2000). The threatened status of groupers (Epinephelinae). *Biodiversity & Conservation*, 9(7): 919-942.
- Heemstra, P.C., Randall, J.E. (1993). FAO species catalogue vol. 16 Groupers of the world (family Serranidae, subfamily Epinephelinae): An annotated and illustrated catalogue of the grouper, rockcod, hind, coral grouper, and lyretail species known to date. Rome: Food and Agriculture Organization of the United Nations.
- Iswarya, D.V., Kandula, S., & Khedkar, G. D. (2017). DNA barcoding of five species of groupers (Pisces: Serranidae) off Visakhapatnam, central eastern coast of India. *Mitochondrial DNAPartA*: 1-5. <https://doi.org/10.1080/24701394.2017.1339188>
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., & Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research*, 25(24): 4876-4882.
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 33(7): 1870-1874.
- Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16: 111-120.
- Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.
- Saitou, N., and Nei, M. (1987). The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4: 406-425.

17. Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.50 Distributed by the author. Department of Genetics, University of Washington, Seattle, USA.
18. Gao, F., Wei, M., Zhu, Y., Guo, H., Chen, S., and Yang, G. (2016). Characterization of the Complete Mitochondrial Genome of the Hybrid *Epinephelus moara* × *Epinephelus lanceolatus*, and Phylogenetic Analysis in Subfamily Epinephelinae. *J. Ocean Univ. China*, 16 (3): 555-563.
19. D'Onofrio, G., Jabbari, K., Musto, H., Alvarez-Valin, F., Cruveiller, S., Bernardi, G. (1999). Evolutionary genomics of vertebrates and its implications. *Ann NY Acad Sci*. 870:81–94.
20. Lassalle, F., Perian, S., Bataillon, T., Nesme, X., Duret, L., Daubin, V. (2015). GC content evolution in bacterial genomes: the biased gene conversion hypothesis expands. *PLoS Genet*. 11: 1–20.
21. Min, X.J., & Hickey, D.A. (2007). DNA barcodes provide a quick preview of mitochondrial genome composition. *PLOS one*, 2(3): e325.
22. Lakra, W.S., Verma, M.S., Goswami, M., Lal, K.K., Mohindra, V., Punia, Gopalakrishnan, A., Singh, K.V., Ward, R.D. & Hebert, P.D.N. (2011). DNA barcoding Indian marine fishes. *Molecular Ecology Resources*, 11(1): 60-71.
23. Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., & Hebert, P.D.N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1462):1847-1857.
24. Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., & Francis, C.M. (2004). Identification of birds through DNA barcodes. *PLoS biology*, 2(10): e312.
25. Sujatha, K. (2004). Groupers off Visakhapatnam, north east coast of India. *J. Mar. Biol. Ass. India*, 46(1): 87-92.
26. Kirubasankar, R., Dam Roy, S., George, G., Sarma, K., Krishnan, P., Ramkumar, S., Kaliyamoorthy, M. & Goutham Bharathi, M.P. (2013). Fishery and Exploitation of Malabar Grouper, *Epinephelus malabaricus* (Bloch & Schneider 1801) from Andaman Islands. *Asian Fisheries Science*, 26:167-175.
27. Roy, T.S.C., & Gopalakrishnan, A. (2011). Identification of groupers based on pyloric caeca differentiation. *Journal of fish biology*, 79(5): 334-1339.
28. Ding, S., Zhuang, X., Guo, F., Wang, J., Su, Y., Zhang, Q., & Li, Q. (2006). Molecular Phylogenetic relationships of China Seas groupers based on cytochrome b gene fragment sequences. *Science in China Series C*, 49(3): 235-242.
29. Craig, M.T., & Hastings, P.A. (2007). A molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelini. *Ichthyological Research*, 54(1):1-17.
30. Ma, K.Y., Craig, M.T., Choat, J.H., & van Herwerden, L. (2016). The historical biogeography of groupers: Clade diversification patterns and processes. *Molecular phylogenetics and evolution*, 100: 21-30.
31. Govindaraju, G.S., & Jayasankar, P. (2004). Taxonomic relationship among seven species of groupers (genus *Epinephelus*; family Serranidae) as revealed by RAPD fingerprinting. *Marine Biotechnology*, 6(3): 229-237.
32. Lakra, W.S., Goswami, M., & Gopalakrishnan, A. (2009). Molecular identification and phylogenetic relationships of seven Indian Sciaenids (Pisces: Perciformes, Sciaenidae) based on 16S rRNA and cytochrome c oxidase subunit I mitochondrial genes. *Molecular biology reports*, 36(5): 831-839.
33. Ghosh, S., Muktha, M., Behera, P.R., Megharajan, S., Ranjan, R., & Gopalakrishnan, A. (2017). Validation of *Epinephelus coioides* (Hamilton, 1822) occurrence along north-east coast of India. *Indian Journal of Geo-Marine Sciences*, 46(2): 266-271.
34. Basheer, V.S., Vineesh, N., Bineesh, K.K., Kumar, R.G., Mohitha, C., Venu, S., Kathirvelpandian, A., Gopalakrishnan, A. & Jena, J.K. (2017). Mitochondrial signatures for identification of grouper species from Indian waters. *Mitochondrial DNA Part A*, 28(4): 451-457.
35. Baharum, S.N., & Nurdalila, A.A. (2011). Phylogenetic Relationships of *Epinephelus fuscoguttatus* and *Epinephelus hexagonatus* inferred from Mitochondrial Cytochrome b Gene Sequences using Bioinformatic Tools. *International Journal of Bioscience, Biochemistry and Bioinformatics*, 1(1): 47-52.