JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Molecular characterization of Puntius species based on Cytochrome c oxidase subunit I (COI) gene

Sunil kumar.P* and Durga Narayanan

Department of Zoology, Sree Kerala Varma College, Thrissur-680011, India

Abstract

Fishes show an astonishing diversity of shapes, sizes, and colours. The delimitation and recognition of fish species is not only of interest for taxonomy and systematists, but it is also a requirement in studies of natural history and ecology, fishery management, tracking the dispersal patterns of eggs and larvae, estimations of recruitment and spawn areas, and authentication of food products. Fish identification is traditionally based on morphological features. However, due to high diversity and morphological plasticity, in many cases, fish and their diverse developmental stages are difficult to identify by using morphological characteristics alone. For many animal taxa, sequence divergences within the 5′ region of the mitochondrial cytochrome oxidase subunit I (COI) gene are generally much greater between species than within species. This in turn suggests that the approach is extensively applicable among phylogenetically distant animal groups. This investigation successfully identified *Puntius parrah* and *Puntius mahecola* at molecular level using cytochrome oxidase subunit I (COI) gene. The inter- and intra-specific pairwise distances analysis using Neighbor-Joining K2P model have clearly validated prominent range of genetic distance between the *Puntius parrah* and *Puntius mahecola* species.

Keywords: DNA barcoding, mitochondrial genome, phylogenetic analysis, *Puntius parrah*, and *Puntius mahecola*.

INTRODUCTION

DNA barcoding is a method of species identification using a short section of DNA from a species gene or genes. The premise of DNA barcoding is that, by comparison with a reference library of such DNA sections, an individual sequence can be used to uniquely identify an organism to species, in the same way that a super market scanner uses the familiar black stripes of the UPC barcode to identify an item in its stock against its reference database. These barcodes are sometimes used in an effort to identify unknown species, parts of an organism, or simply to catalogue as many taxa as possible or to compare with traditional taxonomy in an effort to determine species boundaries. (Brower.*et.al.,* 1999)

The most commonly used barcode region for animals and some protists is a portion of cytochrome c oxidase1 (COI or COX1) gene, found in mitochondrial DNA. When barcoding is used to identify organisms from a sample containing DNA from more than one organism the term DNA from more than one organism the term DNA meta barcoding is used. DNA barcoding techniques were developed from early DNA sequence work on microbial communities using the 5srRNA gene. In 2003, specific methods and terminology of modern DNA barcoding were proposed as a standardised method for identifying species, as well as potentially allocating unknown sequence to higher taxa such as orders and phyla in a paper by Paul D N. (Hebert.*et. al.,* 2003).

DNA barcoding methods for fish are used to identify groups of fish based on DNA sequences within selected regions of a genome. These methods can be used to study fish as genetic material, in the form of environmental DNA or cells is freely diffused in the water. This allows researchers to identify which species are present in a body of water by collecting a water sample, extracting DNA from the sample and isolating DNA sequence that are specific for the species of interest. In fisheries research, barcoding can be used as an alternative to traditional sampling methods, barcoding methods can often provide information without damage to studied animal. Barcoding and meta barcoding methods are particularly useful in studying endangered or elusive fish as species can be detected without catching or harming animals (Pradhan.*et.al.,* 2015).

Barcoding and meta barcoding approaches yield rigorous and extensive data on recruitment, ecology and geographic ranges of fisheries resource. The methods also improve knowledge of nursery areas and spawning grounds with benefits for fisheries management. Traditional methods for fishery assessment can be highly destructive such as gillnet sampling or trawling. Molecular methods offer an alternative for non-invasive sampling. Barcoding and meta barcoding also power tools in monitoring of fisheries quotes and by catch (Davis and Nixon.*et.al.,* 1992). The study was conducted to identify the fish specimens collected from Trissur District of Kerala using mitochondrial COI barcode region, to construct NJ phylogenetic tree of the identified fish

specimens and to estimate the inter- and intra-specific pairwise distances (K2P model) using the isolated sequence.

The nuclear rDNA are organized as multiple copies of a repeated unit that consists of a transcribed zone with coding regions for the 18s, 5.8s and 28 s RNA genes, internal and external transcribed spacers (ITS &ETS) and surrounded by non-transcribed spacers (NTS) sequences. Genetic characterization of fish species is an important step taken towards conservation of fish species that enable precise identification of species and varieties or stocks within species. In view of this, the present study was aimed to characterise and phylogenetic relationship among these Puntius species on the basis of 18s rDNA sequences to facilitate policy makers to make their conservation plans for sustainable utilization (Hillis, D M.*et. al.*, 1992).

Genomic DNA isolation is the total genomic DNA was isolated from the muscle tissue samples following standard phenol:chloroform : isoamyl alcohol method of Sambrook and Russel with minor modifications. The PCR amplifications were performed in 50 µl reaction volume containing 10X taq polymerase buffer, 2m µ Mgcl2 0.2 mM dNTP mix,10 p mol of each primer, 1U taq DNA polymerase and 50 ng genomic DNA. The PCR cycling conditions were initiation denaturation at 94º C for 4 min, followed by 32 cycles of denaturation at 94º C for 30 sec, primer annealing at 55ºC for 30 sec, primer extension at 72ºC for 60 sec, followed final extension at 72ºC for 10 min. The amplified products were resolved in 1.0% agarose gel stained with ethidium bromide and custom sequenced. (Krauskopf *et. al.*, 2003). The raw DNA sequences were edited using bio edit and the pairwise as well as multiple alignment of sequences was performed using clustal W alignment editor. Multiple sequence alignment was also checked manually and the consensus sequences were obtained. Molecular phylogenetic analysis was performed in MEGA (ver.6) software. Sequence data was subsequently analysed for neighbour – joining maximum parsimony and maximum likely hood methods with bootstraps of 1000 replicates. Nucleotide composition analysis was carried out using bio edit. Transitions are transversions were equally weighted. The final consensus sequences were submitted to NCBI database with accession number –JN704069 (*P.conchonius*), JN704072 (*P.chola*), JN704070 (*P.ticto*), JN704071 (*P.sophore*).

18srDNA is a good source of phylogenetic information based on conservation of function, variable mutation rates depending on substructure position and its ubiquity in all taxa. The molecule provides readily obtainable nucleotide sequences because of high rRNA transcript copy number in eukaryotes and ease of PCR primer design. The 18srRNA molecule contains certain conserved regions having low substitution rates and can

provide informative characters for assessing affiliations of evolutionary distant taxa. 18srRNA s also contains variable regions having high substitutions rates. In some organisms these rapidly evolving regions provide phylogenetic signals for discerning relationship between evolutionarily closer clades (Felsenstein.*et.al.,* 1985).

Specimens of snakehead fish were collected from two fresh water bodies of Trissur (Pullu, Manakkodi, Thrissur), Kerala. Caudal fin tissues were taken from each specimen for subsequent molecular studies.

With the tremendous diversity in cypriniformes, understanding of their phylogenetic relationship has not been proceeded at an acceptable rate, despite a plethora of morphological and more recent molecular studies. The phylogenetic relationship based on nuclear rDNA genes was studied in many organisms. The cyprinidae family is monophyletic and this is also in agreement with present study of undertaken Puntius species which showed monophyletic condition. The parsimony, neighbour-joining and maximum likelihood method for 18srDNA resulted in mostly identical tree topologies for undertaken Puntius species (Cavender TM.*et.al.,* 1992).

METHODOLOGY

Specimens of *Puntius parrah* and *P. mahecola* were collected from two fresh water bodies of Trissur (Pullu, Manakkodi, Thrissur), Kerala. Caudal fin tissues were taken from each specimen for subsequent molecular studies. Genomic DNA was isolated from the tissues using NucleoSpin® Tissue Kit (Macherey-Nagel) following manufacturer's instructions. Agarose Gel Electrophoresis was done for DNA Quality check.

The PCR amplification was carried out in a PCR thermal cycler (GeneAmp PCR System 9700, Applied Biosystems). The PCR products were checked using Agarose Gel electrophoresis. It was followed by ExoSAP-IT Treatment and Sequencing using BigDye Terminator v3.1. The cleaned up air dried product was sequenced in ABI 3500 DNA Analyzer (Applied Biosystems). The sequence quality was checked using Sequence Scanner Software v1 (Applied Biosystems). Sequence alignment and required editing of the obtained sequences were carried out using Geneious Pro v5.1.

RESULT AND DISCUSSION

The present investigation has analysed two fish species at molecular level. The extensive developments in the molecular biology like mitochondrial DNA analysis allow us to complement the conventional taxonomic methods of species identification. A total of 65 COI nucleotide sequences were analysed in this investigation. Among them 63 nucleotide sequences were retrieved from the NCBI database.

All sequences were aligned using CLUSTAL W and all ambiguous position were removed for each sequence pair. This investigation successfully identified *Puntius parrah* and *Puntius mahecola* at molecular level. Figure 1and 2 depicts the phylogenetic analysis of *Puntius parrah* and *Puntius mahecola* respectively. The COI based mDNA analysis of the aforesaid species such as *Puntius parrah* and *Puntius mahecola* revealed distinct clustering of individual fish species within every genus together with strong bootstrap support.

This study has used the following nucleotide sequences with accession numbers to find out the phylogenetic relationship between the identified *Puntius mahecola* with other fishes those are previously reported from the various region over the world. In order to find out the phylogenetic relationship between identified *Puntius parrah* species this study has retrieved the following nucleotide sequences from NCBI , HE681797.1:47-537, JX416126.1:77-566, HE681798.1:34-523, HE681799.1:49-538, HE681801.1:42-531, MW398913.1:71-560, HE681800.1:46-535.

As depicted in figure 2 the nucleotide sequence named *Puntius mahecola*_ CND2 large gene ribosomal RNA is recognized as the isolated sequence of *Puntius mahecola* while the sequence for *Puntius parrah* CND2 large subunit ribosomal RNA gene. As stated in the above section *Puntius parrah*_ CND2 large subunit rRNA gene and *Puntius mahecola* CND2 large subunit rRNA gene have clustered with those of other related species from various regions reported in the NCBI database. Altogether this investigation has identified the two species such as *Puntius parrah* and *Puntius mahecola*. The neighbour joining (N-J) phylogenetic tree constructed using the RNA sequences of *Puntius parrah* and *Puntius mahecola* and was shown in figure 3. The N-J tree for Puntius species formed 27 distinct dades with strongly supported bootstrap values, each representing an individual species.

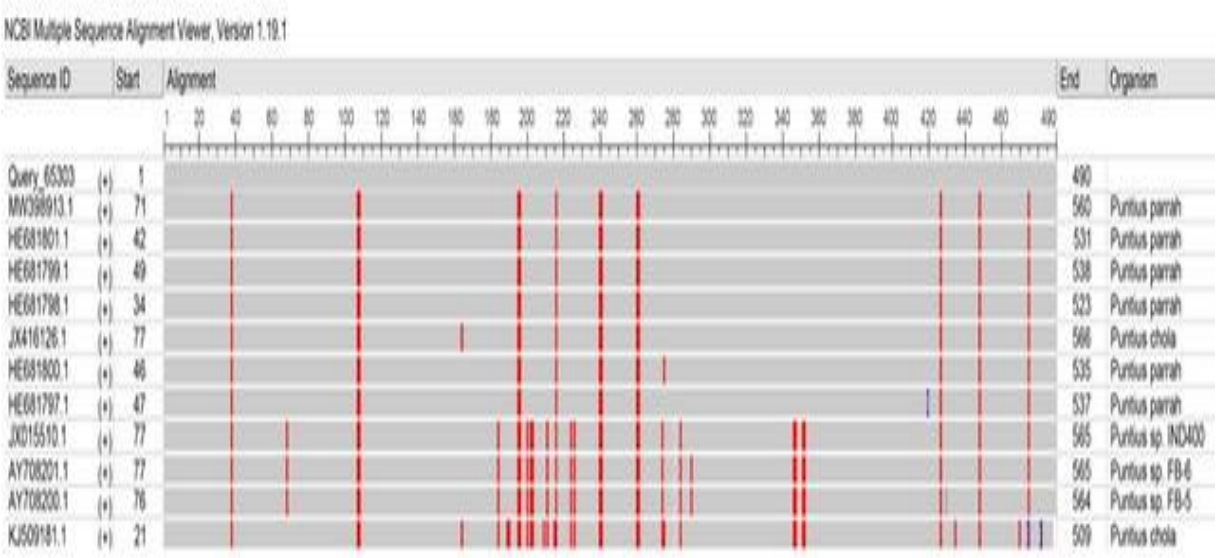

Figure. 1 NCBI Multiple Sequence Alignment For Query_2235 (*Puntius parrah*)

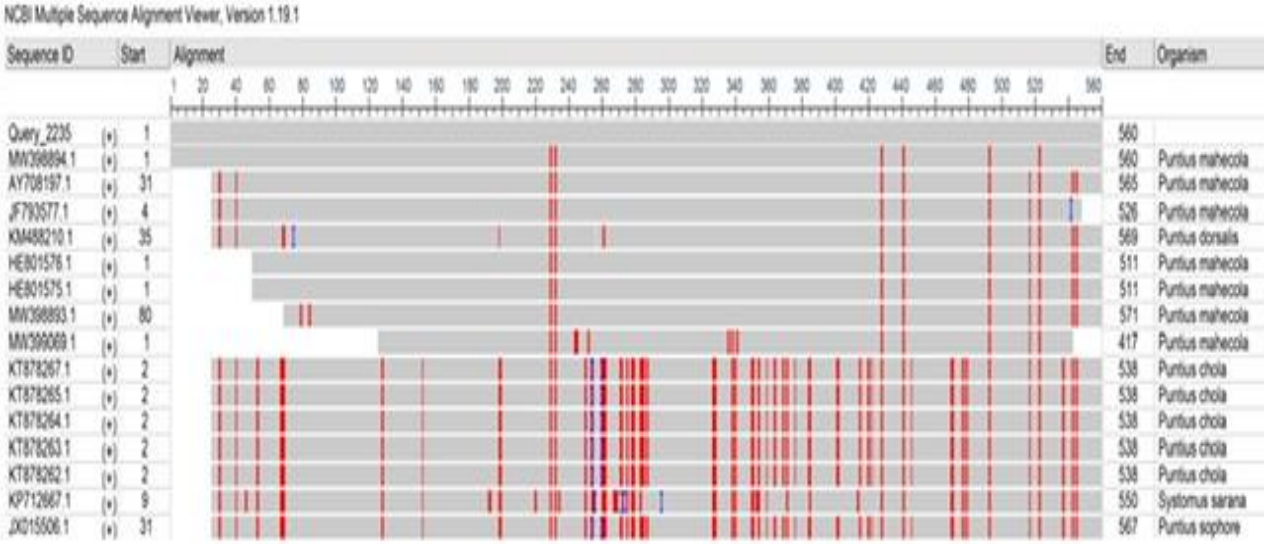

Figure. 2. NCBI Multiple Sequence Alignment For Query_2235 (*Puntius mahecola*)

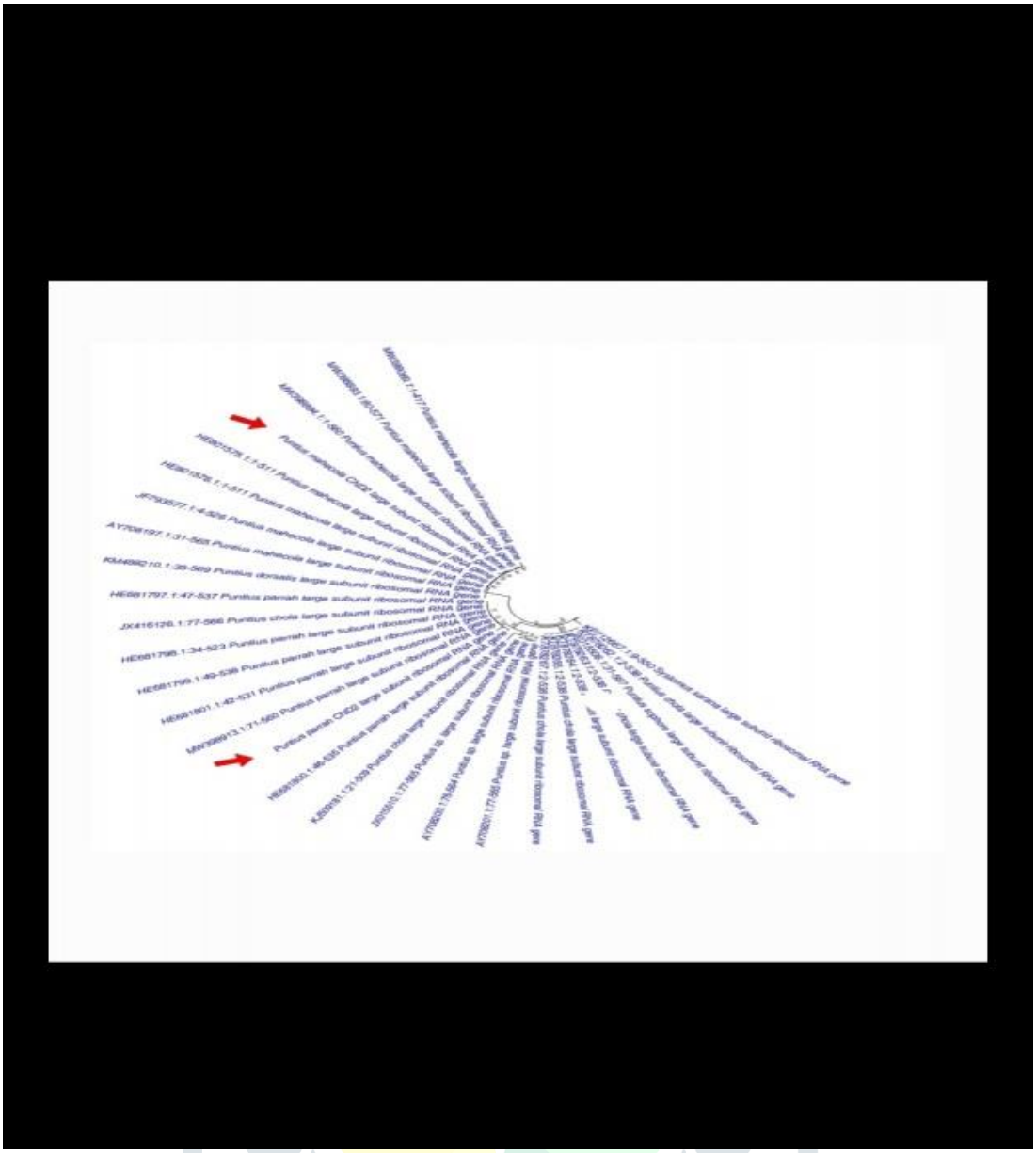

Figure. 3 *Puntius* sp. Evolutionary relationships of taxa. The evolutionary history was inferred using the Neighbor-Joining method (I). The optimal tree with the sum of branch length =0.31456731 is shown. The tree is drawn to scale with branch lengths, in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (3) and are in the units of the number of base differences per site. The analysis involved 46 nucleotide sequences. Codon positions included were $1^{st}+3^{rd}$. All ambiguous positions were removed for each sequence pair. There were a total of 383 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (4).

Inter- and intra-specific pairwise distances

Intra specific pairwise distances between the identified *Puntius parrah* and previously reported *Puntius parrah*

The present study has also find out the genetic distance between the identified two species along with the determination of genetic distance within the same species. The matrix output of intra specific pairwise distance

of *Puntius parrah* revealed that the candidate sequence which is isolated in this study (*Puntius parrah* CND2 large subunit rRNA genes) have formed genetically different clusters with the following nucleotide;

AY708201.1:77-565, AY708200.1:76-564, KJ509181.1:21-509.

The average value for the interspecific genetic divergence was 0.100.

Intra specific pairwise distances between the identified *Puntius mahecola* and previously reported *Puntius mahecola*

This study also determined the genetic distance between the identified *Puntius mahecola* with other reported species worldwide and revealing extensive genetic distance between them. The matrix output of intra specific pair wise distance of *Puntius mahecola* revealed that the candidate sequence which is isolated in this study *Puntius mahecola* CND2 large subunit rRNA gene" have formed genetically different clusters with the following nucleotides such as KT878267.1:2-538,KT878265.1:2-538,KT878264.1:2-538,KT878263.1:2-538,KT878262.1:2-538,KP712667.1:9-550,JX015506.31-567

As mentioned in the former section, the average value for the inter-specific genetic divergence was 0.100. Through DNA barcoding the used material is phylogenetically identified. The materials are identified as *Puntius parrah* and *Puntius mahecola*.

Inter specific pairwise distances between the identified *Channa punctata, Channa striata* and previously reported sequences

NJ-K2P methods for calculating inter specific pairwise distances between the identified *Puntius parrah, Puntius mahecola* and previously reported sequences have revealed that extensive genetic distance between them. Divergent clusters with other nucleotide sequences representing respective individual species with an average value 0.100. All the fish specimens analysed in this study can therefore, be identified to species level based on their COI sequences, which offers 100% compatibility between taxonomic and molecular identification and reveals that COI gene based phylogenetic analysis is a convenient approach to complement conventional taxonomy.

The inter- and intra-specific pairwise distances analysis using Neighbor-Joining K2P model have clearly validated prominent range of genetic distance between the *Puntius parrah* and *Puntius mahecola* species.

Conclusion

DNA barcoding has been considered as a global bio identification system for fresh water and marine fishes in recent years. The rapid and accurate identification of closely related fish species such as *Puntius parrah* and *Puntius mahecola* is essential for scientific research and artificial breeding. DNA barcoding can correct field misidentification, reduces ambiguity of species identification, makes species identification more exact, democratises access by creating open access databases and expands technical expertise of taxonomists. The COI gene-based mitochondrial DNA analysis of the aforesaid species such as *Puntius parrah* and *Puntius mahecola* revealed distinct clustering of individual fish species within every genus together with strong bootstrap support. NJ-K2P method for calculating inter specific pairwise distances between the identified *Puntius parrah* and *Puntius mahecola* and previously reported sequences have revealed that extensive genetic distance between them.

Acknowledgements

The authors are thankful to the Principal, Sree Kerala Varma College, Thrissur for permitting to carry out the study in the Zoology Department.

References

Agarwal SS, Saksena DN. A checklist of fishes from Gwalior, Madhya pradesh. J Jiwaji university.1977; 2: 1674-169

Amorim, D. S., Santos, C.M, Krell, F.I, Dubois, A.,Nitri, S. S, Oliveria, O. M., 2016. Timeless standards for species delimitation Zootaxa 4173,121-128. doi:10.11646/zootaxa.4137.1.9.

Brower, A.V.Z (1999). Delimitation of phylogenetic species with DNA sequence:a critique of Davis and Nixon's population aggregation analysis. Syst. Biol. 48,199-213-doi:10.1080/106351599260535.

Cavender T.M, Coburn M M. Phylogenetic relationships of North American cyprinidae. In:Mayden RL ed.systematics , historical ecology and North American freshwater fishes. Palo Alto; Stanford University press 1992, 293-327.

Chakraborty, M., Ghosh, S.K., 2014. An assessment of the DNA barcodes of Indian freshwater fishes. Gene 537, 20-28.

Davis, J.I and Nixon,K.C(1992). Populations genetic variation and the delimitation of phylogenetic species. Syst. biol.41, 421-435.doi:10.1093/sysbio/41.4.421.

Felsenstein J. Confidence limits on phylogenetic :an approach using the bootstrap. Evolution 1985; 39:783-791.

Hebert,P.D.N,cyninska, A.,Ball,S.L & dewaard,J. R. Biological identifications through DNA barcodes. Proc.biol.sci.270, 313-321 (2003).

Hillis,D.M & Huel senbeck , J. P (1992). Signal,noise and reliability in molecular phylogenetic analysis. Journal of Heredity ,83,189-195.10,1093/oxford journals. 11190.

Kullannder,S.O.(2008). Five new species of *Puntius* from Myanmar (Teleostei:cyprinidae) ichthyological exploration of freshwater 19.59-84.

Naylor, R l., Golgburg, R J ., Primavera , J H .,Kautsky, N., Beveridge, M C.,Clay, J., Folke,C., Lubchenco, J., Mooney, H.,Troell, M., 2000. Effect of aquaculture on world fish supplies. Nature 405, 1017-1024.

Nelson JS . Fishes of the world . 4th ed. Hoboken: John wiley and sons , inc, 2006 .

Pradhan V, Kamble Y, Ladniya V, Mogal M (2015). An overview of species identification by DNA barcoding. Int j curr Microbial app. 4: 127-40

Vera M I, Molina A , Pinto R, Reyes M, Alvarez M, Krauskopf E. Genomic organization of the rDNA cistron of the teleost fish *Cyprinus carpio*. Biol res 2003; 36 (2):402-419