International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

DNA Barcoding Application in Study of Icthyo-Biodiversity in Rivers of Pakistan

Asma Karim[1], Rashid Saif[2], Akhtar Ali[3], Beenish Nadeem[4], Hafiza Aqsa Ilyas[5], Waseem Sajjad[6]

[1,3,4,5,6]Department of Zoology, Govt. College of Science, Wahdat Road, Lahore, Pakistan
[2]Decode Genomics, Research Training and Diagnostic Center, Lahore, Pakistan

ABSTRACT:
Fish taxonomy plays a fundamental role in the study of biodiversity. However, traditional methods of fish taxonomy rely on morphological features, which can lead to confusion due to great similarities between closely related species. To overcome this limitation, modern taxonomy employs DNA barcoding as a species identification method. This involves using a short standardized mitochondrial DNA region as a barcode, specifically a 658 base pair fragment near the 5′ end of the mitochondrial cytochrome c oxidase subunit 1 (CO1) gene, to exploit diversity in this region for identification of species. To test the effectiveness and reliability of DNA barcoding, 25 fish specimens from nine different fish species found in various rivers of Pakistan were identified morphologically using a dichotomous key at the start of the study. comprising nine fresh water fish species including *Mystus cavasius, Mystus bleekeri, Osteobrama cotio, Labeo rohita, Labeo culbasu, Labeo gonius, Cyprinus carpio, Catla catla* and *Cirrhinus mrigala* from different rivers of Pakistan were used in present study. DNA was extracted from one of the pectoral fins and partial sequence of CO1 gene was amplified using conventional PCR method. Analysis of the barcodes confirmed that genetically identified fishes were the same as those identified morphologically at the beginning of the study. The sequences were also analyzed for biodiversity and phylogenetic studies. Based on the results of the study, it can be concluded that DNA barcoding is an effective and reliable method for studying biodiversity and conducting phylogenetic analysis of different fish species in Pakistan.

Keywords: DNA barcoding, fresh water fishes, taxonomy, biodiversity, Pakistan

INTRODUCTION:

Taxonomy has remained a challenging job for scientists for various reasons. Many new and promising methodologies for fish identification have been developed, but with few exceptions, these methods in academic research have not yet been made applicable for non-specialists and still needed to mature into globally applicable tools. Different methodologies used so far for fish identification include participation of scientific experts (taxonomists) and local experts, use of taxonomic reference libraries, through image recognition systems, use of field guides based on dichotomous keys, interactive electronic keys, morphometric method including scales and otolith morphology and hydro acoustics. The latest methods for fish identification include genetic methods like (Single nucleotide polymorphisms (SNPs) and DNA barcoding technique in which barcodes are generated for specimen **(Fischer, 2013)**.

DNA barcoding technique is based on the ground that a standardized short region of mitochondrial DNA is used to verify species identity, which typically for fish is portion of cytochrome C oxidase 1 (CO1) gene in mitochondrial DNA **(Hebert *et al,* 2003)**. This region known as barcode is accessible to public and is highly comparable with the barcodes in database. **Hebert *et al*. (2004)** suggested that partial sequence of CO1 gene from mitochondrial DNA has potential for identification of almost all or majority of animal species because of the fact that genetic variations "between" species are always greater than "within' species. He also proposed the use of CO1 gene in mitochondrial DNA as a global taxonomic bio tool for animals.

Clear and correct identification of fish and fish products, has economic importance in many fields. Efficacy of DNA barcoding was assessed in a study where diversity of COI gene was examined within and among 207 Australian commercial fish species during a study to check the efficacy of this gene to identify fishes **(Ward *et al*., 2005)**. This study resulted in the conclusion that CO1 gene is an efficient marker for all 207 fish species identification. DNA barcoding technique involves only small portion of DNA sequence instead of complete genome. The short DNA sequence is taken from particular region of CO1 genome to create DNA barcode. The use of mitochondrial DNA portion instead of any nuclear DNA portion has many advantages **(Hurst & Jiggins, 2005)**. Hybridization among species often creates taxonomic uncertainty, mitochondrial DNA is maternally inherited and any hybrid or subsequent generation would have the maternal species DNA only.

The Mitochondrial DNA is considered to be a universal tool that reveals the evolutionary history of the species **(Ballard *et al*., 2004)**. Mitochondrial gene "CO1" is haploid in nature and has highly conserved regions. This gene can be easily amplified without cloning in different species **(Whitworth *et al*., 2007)** and it has also great phylogenetic signals than any other mitochondrial gene. Mitochondrial cytochrome b gene sequence is also used extensively for vertebrate's taxonomy **(Castresana, 2001)**. But there are certain limitations such as it fails to explain relationship among same species as well as phylogenies. DNA barcoding represents a powerful tool for biodiversity assessment and quickly sorting collections into species-like units because of the relative speed at which COI sequence data is generated and analyzed **(Schindel & Miller, 2005)**. DNA barcoding system is suitable, fast and accurate species identification system, which makes ecological system more accessible **(Hebert & Ryan, 2005)**.

Mystus cavasius, Mystus bleekeri and *Osteobram cotio* have very good taste and they are used as food fish in many countries including Bangladesh. These fishes are known to supply calcium and vitamin A to a great extent. These fishes are always marketed in fresh condition. *Osteobrama cotio* is possibly a useful larvicide **(Bhuiyan, 1964** and **Talwar & Jhingran, 1991)**. These fishes have also made their entry in the ornamental fish markets and recently have been documented to be exported as indigenous ornamental fish from India **(Gupta & Banerjee, 2014)**. *Labeo rohita* (Rohu), *Labeo culbasu, Labeo gonius, Cyprinus carpio, Catla catla* and *Cirrhinus mrigala* has high market demand as food fish with high market price **(Talwar & Jhingran, 1991; Rahman *et al*., 2004; Siddiqui *et al*., 2010** and **Hossen *et al*., 2014)** due to good protein content in their flesh. These species are appreciated for good market price and have a great consumer demand **(Roy & Hossain, 2006; Siddiqui *et al*., 2010** and **Ashashree *et al*., 2013)**.

During the present work, molecular taxonomy, biodiversity and phylogenetic analysis of 9 economically important fresh water fish species (*Mystus cavasius , Mystus bleekeri, Osteobrama cotio, Labeo rohita, Labeo culbasu, Labeo gonius, Cyprinus carpio, Catla catla* and *Cirrhinus mrigala*) were carried out.

MATERIALS AND METHODS:

In the present study, fresh water fishes (n=25) from rivers were used. These include *Mystus cavasius* (*n*=6) and *M. bleekeri* (n=3) *Osteobrama cotio* (n=3), *Labeo rohita* (n=4), *L. culbasu* (n=1), *L. gonius* (n=2), *Cyprinus carpio* (n=3), *Catla catla* (n=1) and *Cirrhinus mrigala* (n=2). All these fish belong to Bagridae and Cyprinidae families **(Fig. 1)**.

Fish samples were collected from River Ravi at Head Balloki region in district Kasur; and Rivir Indus at Taunsa Barrage. Fishes were captured with the help of fisherman from fisheries department with casting net. Fishes were brought to the laboratory of Govt. College of Science, Lahore, Pakistan. Fishes were anaesthetized and killed with benzocaine chloride @ 40 mg/l. Fishes were observed closely and when fishes reached the level of anesthesia sufficient to perform surgery, one of the pectoral fin of each fish was removed and transferred immediately into properly labeled autoclaved 1.5 ml Eppendorf tubes. The fishes were then photographed and preserved as voucher specimen in formalin.

DNA Extraction Kit (TIAN ampg) was used to extract DNA from pectoral fins of fish. Extracted DNA was run on the agarose gel to check quality of the extracted nucleic acids DNA bands were seen by passing UV light through the gel in Gel Documentation System (Bio Doc- it TM system). The bands in the gel documentation indicated the quality of DNA. FishF1 and FishR2 primers were used to amplify partial sequences of CO1 gene.

FishF1 5'TCAACCAACCACAAAGACATTGGCAC3'

FishR2 5'ACTTCAGGGTGACCGAAGAATCAGAA3'

The PCR products were run on the gel to see the results. Products with 600 plus bp (gene of interest) were sent for sequencing.

RESULTS:

Good sequence results without stop codon were submitted in Gen Bank. Accession numbers are mentioned in **Table (1)**. Sequences with stop codon were not submitted in Gen Bank but were used in phylogenetic analysis in MEGA 7 and dnaSP5.

The mean read length for all sequences for all these fishes was 621 base pairs. There were no insertions, deletions and stop codons in all sequences submitted to GenBank. The lack of stop codon was consistent in all submitted CO1 gene sequences, this showed that all sequences are functional mitochondrial gene sequences and the average length of all amplified sequences being 621 base pairs suggested that NUMTs (Nuclear DNA sequences originated from mitochondrial DNA sequences) were not sequenced (Vertebrate NUMTs are typically smaller than 600 base pairs **(Zhang & Hewitt, 1996)**.

When sequences (n=25) were analyzed for polymorphism using DnaSP5, there were 459 sites with alignment gaps or missing data, 117 invariable or monomorphic sites, 199 variable or polymorphic sites, 4 singleton variable sites, 260 mutations and 195 parsimony informative sites found in these barcodes. Sequences were also studied on the of GC content. Percentage concentration of Thymine, Cytosine, Adenine and Guanine is given in **Table (2)**. Analysis using software dnaSP5 showed that there were 11 haplotypes and the haplotype diversity (Hd) was found to be 0.933 in experimental fish species **(Table, 3)**.

25 sequences of fishes were used in phylogenetic analysis. Sequences were aligned using MUSCLE algorithm through MEGA 7 software. Phylogenetic tree was constructed using Neighbor-Joining method **(Saitou & Nei, 1987)** to have insight of their genetic distances which cluster the fish species from Family Cyprinidae and Family Bagridae in separate clads. This phylogenetic tree showed a close relationship

between fishes belonging to same family **(Fig. 1)**. The optimal tree showed the sum of branch length equal to 1.67308992. Bootstrap test (1000 replicates) was conducted and its percentage values are shown next to the branches **(Felsenstein, 1985)**. The evolutionary distances were calculated using Kimmura 2 parameter method **(Kumar *et al*., 2016)**.

DISCUSSION:

enetic identification of species using partial sequence of Cytochrome C Oxidase 1(CO1) gene from mitochondrial DNA has been proved to be a successful method in species identification and at the same time validates the existing taxonomical methods. Present study has strongly confirmed the efficacy of CO1 barcodes for identifying the fish species as well as for having the insight of phylogeny.

According to **Li & Du, (2014)**, the evolutionary history of organisms can be predicted by looking at the %age concentration of GC content in DNA. According to him earlier evolved specimen or ancestral species had more G&C content as compared to more developed/modern species.

Saccone (1999) suggested that there is variation in level of GC content of entire genome, within and among major groups of organisms. The reasons for these variations and their role in evolution are still debatable. In mammalian and avian genomes heterogeneity of GC content is found at 100 kb scale **(Bernardi, 2000** and **Lander, 2001)**. It is an interesting fact that the GC content in genomes of teleost fishes and amphibians are much less heterogeneous than in mammals and birds **(Bernardi Gia & Berardi Gio, 1990** and **Costantini *et al*., 2009)**. Mitochondrial genomes are also relatively uniform and exhibit (taxon-dependent) twists, although in this case, the amount of twist in the coding strand is strongly correlated with overall GC content. Uniform GC content has been reported for the nuclear genomes of fishes and Xenopus. Two main evolutionary processes have been raised to explain why pattern of base composition vary within and among species in the process of mutation such that the rate of change of bases are not constant in time or space and natural selection on overall GC content or on specific pattern of codon usage **(Arne & Adwards, 2000)**.

A haplotype is a group of variants in an organism that are inherited together from a single parent **(By *et al*., 2016)**. Early studies suggested that relatively few (2–6) haplotypes may be sufficient to describe the genetic variations along extended stretches of DNA **(Hobbs *et al*., 2013)**. During present work 11 haplotypes were found among all 25 samples of fresh water fishes indicating genetic variations. The haplotype diversity in our sequences of fishes is 0.933 which shows high haplotype diversity. According to a report haplotype diversity when found in the range 0.75 to 0.92, was considered high, when compared to many other species **(de Jong *et al*., 2011)**.

Phylogenetic analysis was done by using Neighbor- Joining (N-J) tree with a boots trap value of 1000 in MEGA 7 Software package . The cladogram showed inter specific and intra specific evolutionary distances. The distance calculated between fish species belonging to the same family and different family showed remarkable differences.

Two different species of genus *Mystus, i.e., M. bleekeri and M. cavasius* were closely related and these were grouped under the same family Bagridae with the 100% bootstrapping value. All 7 different species of fishes belonging to same family also grouped under the same clad cyprinidae.

CONCLUSIONS:

It is concluded that DNA barcoding is a useful technique for identification of fresh water fish species caught from River Indus and River Ravi Pakistan. Partial sequence of CO1 gene used as barcode of a

particular species can also be used for the phylogenetic and evolutionary analysis. From our results it is also clear that same species from two different regions show divergence and the same genera with different species have greater divergence and genetic distances as compared to the same species. Presence of haplotypes show the variations in the species from the same region whether from River Indus or River Ravi. The present work can be extended to different species from different water bodies and the divergence among species from these regions can be calculated. The relationship between rate of divergence and environmental factors including biological, physical and chemical factors can also be studied.

ACKNOWLEDGEMENT:
Authors greatly acknowledge the administration of virtual university of Pakistan, for their encouragement, cooperation and support to use the lab facilities in completion of this project.

Conflict of interest:
All authors declare no support from any organization for submitted work and have no financial relationship with any organization that might take interest in the submitted work. All authors are not involved in any other activities or relationships that could appear to have influence the submitted work.

REFERENCES:

1. **Arne, M.** and **Edward, C. (2000).** The evolution of base composition and phylogenetic inference. *Trees.* **15**, 365-369.
2. **Ashashree, H.M.; Venkateshwarlu, M.** and **Sayeswara, H.A. (2013).** Seasonal changes of protein in the tissues of male catfish Mystus cavasius (Ham) in Bhadra reservoir, Karnataka, India. *Int. j appl biol Pharm*. **4**, 264-267.
3. **Ballard, J.; William, O.; Michael, C.** and **Whitlock, M.C. (2004).** The Incomplete Natural History of Mitochondria. *Mol Ecol*. **13**, 729-744.
4. **Bernardi, G**. (2000). Isochores and the evolutionary genomics of vertebrates. Gene, 241(1),
5. 3-17.
6. **Bernardi, G., & Bernardi, G. (1990)**. Compositional transitions in the nuclear genomes of
7. cold-blooded vertebrates. Journal of molecular evolution, 31(4), 282-293.
8. **Bhuiyan, A.L. (1964).** *Fishes of Dacca, Asiat. Soc. Pakistan*, **1**(13): 60-61.
9. **By, C.; Barry, C.; Peter, D.; Moore** and **Richard, L. (2016).** An Ecological and Evolutionary Approach. Biogeography. 118-96858-1 p106
10. **Castresana, J. (2001).** Comparative Genomics and Bioenergetics. Biochimica et Biophysica Acta (BBA) – Bioenergetics, **3**: 147-162.
11. **Costantini, M.; Cammarano, R.** and **Bernardi, G. (2009).** The evolution of isochores pattern in vertebrate genomes. BMC Genomics, **10**:146.
12. **De Jong, M.A.; Wahlberg, N.; van eijk, M.; Breakfeild, P.M.** and **Zwaan, B.J. (2011).** Mitochondrial DNA signature for range-wide populations of *Bicyclus anynana* suggests a rapid expansion from recent refugia. Plos One**, 6:** 6 E21385.
13. **Felsenstein, J. (1985).** Confidence limits on phylogenies: An approach using the bootstrap. Evolution**, 39**: 783-791.

14. **Fischer, J. (2013).** Fish identification tools for biodiversity and fisheries assessment. Food and Agriculture Organization of United Nations, Rome. Avise J.C. Molecular Markers, Natural History and Evolution. New York: Chapman & Hall.
15. **Gupta, S.** and **Banerjee, S. (2014).** Indigenous ornamental fish trade of West Bengal. Narendra Publishing House, New Delhi, 63.
16. **Hebert, P.D.** and **Ryan, G.T. (2005).** The promise of DNA barcoding for taxonomy. Systematic Biology, **54**: 852-859.
17. **Hebert, P.D.; Cywinska, A.; Ball, S.L.** and **Deewaard, J.R. (2003).** Biological identifications through DNA barcodes. Proc. R. Soc. Lond. Biol. Sci., **270**: 313-321.
18. **Hebert, P.D.; Penton, E.H.; Burns, J.M.; Janzen, D.H.** and **Hallwachs, W. (2004).** Ten species in one: DNA barcoding reveals cryptic species in the neo-tropical skipper butterfly, *Astraptes fulgerator*. Proc. Natl. Acad. Sci. USA., **101**: 14812–14817.
19. **Hobbs, J-P.A.; Herwerden, L-v.; Jerry, D.R.; Jones, G.P.** and **Munday, P.L. (2013).** High genetic diversity in geographically remote populations of Endemic and wide spread coral reef angel fishes (genus: *Centropyge*). *Diversity* **5**, 39-50.
20. **Hossen, M.S.; Reza A.H.; Rakhi, S.F.; Takahashi, K.** and **Hossain, Z. (2014).** Effect of phospholipids in brood stock diets on serum calcium level, gamete quality and spawning of threatened bagrid catfish Gulsha, *Mystus cavasius*. *Int J Res Fisheries and Aquaculture*, **4**: 70-76.
21. **Hurst, G.D.D.** and **Jiggins, F.M. (2005).** Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc. R. Soc. B.*, **272**: 1525-1534.
22. **Kumar, S.; Stecher, G.** and **Tamura, K. (2016). MEGA7**: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution**, 33**(7): 1870-1874.
23. **Lander, E.S. (2001).** Initial sequencing and analysis of the human genome. Nature, **204**: 860-921
24. **Li, X. Q. & Du, D. (2014**). Variation, evolution, and correlation analysis of C+ G content and genome or chromosome size in different kingdoms and phyla. PLoS One, 9(2),
25. e88339
26. **Rahman, M.R.; Rahman, M.A.; Khan, M.N.** and **Hussain, M.G. (2004).** Observation on the embryonic and larval development of silurid catfish, gulsha (*Mystus cavasius* Ham.). *Pak. J.* Biol. Sci., **7:** 1070-1075.
27. **Roy, P.K., Hossain, M.A. (2006).** The fecundity and sex ratio of *Mystus cavasius* (Hamilton) (Cypriniformes: Bagridae). J. Life and Earth Sci., **1**: 65-66.
28. **Saccone, C. (1999).** Evolutionary genomics in Metazoa: the mitochondrial DNA as model system. Gene, **238**: 195-209.
29. **Saitou, N.** and **Nei, M. (1987).** The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution**, 4**: 406-425.
30. **Schindel, D. E., Miller, S. E. (2005).** DNA barcoding: a useful tool for taxonomists. Nature**, 435**: 17.
31. **Siddiqui, M.N.; Biswas, P.K.; Ray, S.; Hasan, M.J.** and **Reza, M. (2010).** Effect of freezing time on the nutritional value of *Mystus gulio* (Nuna Tengra), *Mystus tengara* (Bazari Tengra) and *Mystus cavasius* (Gulsha Tengra). Journal of Science Foundation**, 8**: 119-122
32. **Talwar, P.K.** and **Jhingran, A.G. (1991).** Inland fishes of India and adjacent countries. Vol-1 and Vol-2. First Edition. Oxford and IBH Publishing Co. Pvt. Ltd. New Delhi, Bombay and Calcutta, **1991**, 1063.

33. **Ward, R.D.; Zemlak, T.S.; Ines, B.H.; Last, P.R.** and **Hebert, P.D.N. (2005).** DNA barcoding Australia's fish species, Phil. Trans. R. Soc. Lond. B., **360**: 1847–1857.
34. **Whitworth, T.L.; Dawson, R.D.; Magalon, H.** and **Baudry, E. (2007).** DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). Proceedings of the Royal Society B: Biological Sciences**, 274**: 1731-1739.
35. **Zhang, D.X.** and **Hewitt, G.M. (1996).** Nuclear integration: challenges for mitochondrial DNA markers. Trends Ecol. Evol., **11**(6): 247-51.

Table (1): Fish sequences submitted in Gen Bank with their accession numbers

Family	Species	Common names	River name	Number of nucleotides	Gen Bank Accession numbers
Bagridae	*Mystus cavasius*	Tengra	Ravi	626	**KY465565**
	Mystus cavasius	Tengra	Ravi	622	**KY465566**
	Mystus cavasius	Tengra	Ravi	620	**KY465567**
	Mystus cavasius	Tengra	Ravi	615	**KY465568**
	Mystus cavasius	Tengra	Indus	605	**MT670300**
	Mystus cavasius	Tengra	Indus	610	**MT670301**
Cyprinidae	*Labeo rohita*	Rohu	Indus	651	**MT954413**
	Labeo rohita	Rohu	Indus	653	**MT954411**
	Cyprinus carpio	Common carp	Indus	538	**KP696786**
	Catla catla	Thaila	Indus	632	**MT664962**
	Cirrhinus mrigala	Mori	Indus	659	**KP696784**

Table (2): GC content in sequences

S. No.	Fish species	T	C	A	G	Total
1	*Mystus cavasius*	30.8	26.8	25.2	17.1	626.0
2	*Mystus cavasius*	30.7	26.7	25.6	17.0	622.0
3	*Mystus cavasius*	24.9	17.9	30.5	26.7	610.0
4	*Mystus cavasius*	25.3	17.6	30.0	27.1	620.0
5	*Mystus cavasius*	25.2	17.7	30.2	26.8	615.0
6	*Mystus cavasius*	24.8	18.0	30.6	26.6	605.0
7	*Mystus bleekeri*	28.1	26.0	26.8	19.1	603.4
8	*Osteobrama cotio*	27.8	27.2	26.9	18.2	644.0
9	*Labeo rohita*	28.5	27.1	26.3	18.1	653.0
10	*Labeo culbasu*	28.2	27.4	27.1	17.4	657.0
11	*Labeo gonius*	28.3	27.0	26.7	18.0	660.0
12	*Cyprinus carpio*	28.1	29.4	25.8	16.6	523.0
13	*Catla catla*	28.3	27.1	26.8	17.7	660.0

14	*Cirrhinus mrigala*	27.9	28.2	24.9	19.0	659.0

Table (3): Haplotypes found in sequences

Haplotype number	Species No.	Species names
1	3	[*Cyprinus carpio*; *C. carpio* (2); *C. carpio* (3)]
2	3	[*Ostiobrama cotio*; *O. cotio* (2); *O. cotio* (3)]
3	3	[*Mystus bleekeri*; *M. bleekeri* (2) *M. bleekeri* (3)]
4	4	[*Labeo rohita*; *L. rohita* (2); *L. rohita* (3); *L. rohita* (4)]
5	1	[*Labeo calbasu*]
6	2	[*Labeo gonius*; *L. gonius* (2)]
7	1	[*Catla catla*]
8	2	[*Cirrhinus mrigala*; *C. mrigala* (2)]
9	2	[*Mystus cavasius*; *M. cavasius* (2)]
10	2	[*Mystus cavasius* (3); *M. cavasius* (6)]
11	2	[*Mystus cavasius* (4); *M. cavasius* (5)]

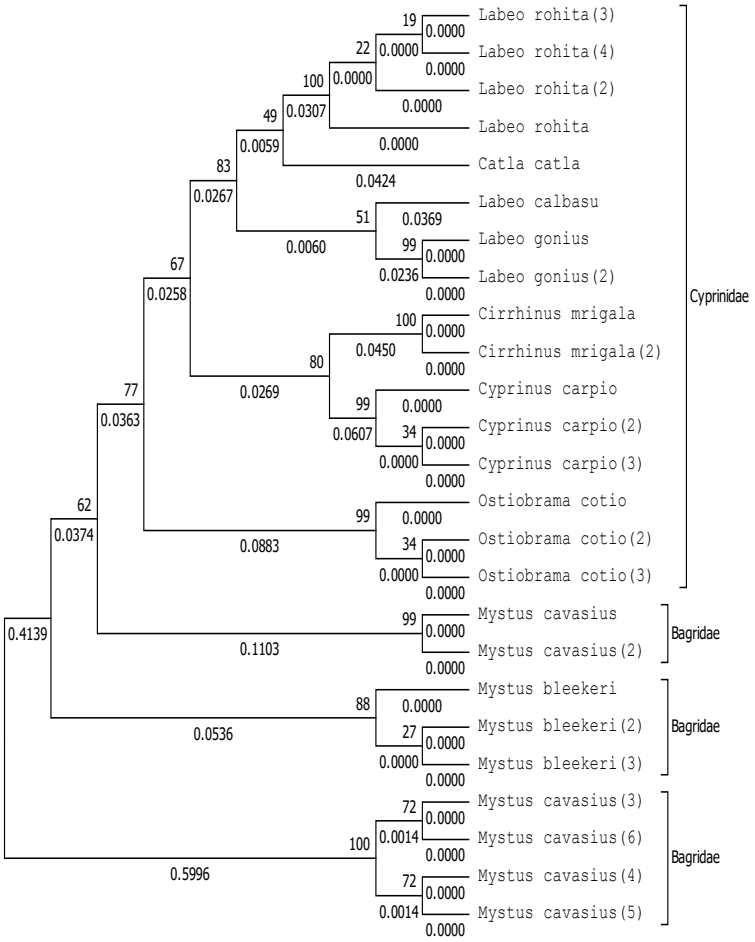

Fig. (1). Phylogenetic tree of some fish species in family Bagridae and family Cyprinidae