



DOI: <https://doi.org/10.54692/lgujls.2022.0603219>

Research Article

LGU J. Life. Sci

Vol 6 Issue 3 July - September 2022

ISSN 2519-9404

eISSN 2521-0130

## Molecular identification of museum preserved type specimens of fish species using DNA barcoding

Naila Malkani<sup>1\*</sup>, Saqiba Khalid<sup>1</sup>, Sania Ashraf<sup>1</sup>, Muhammad Kamran<sup>2</sup>, Rizwan Ullah Khan<sup>1</sup>, Atif Yaqub<sup>1</sup>, Azizullah<sup>1</sup>

<sup>1</sup>Department of Zoology, GC University, Lahore Pakistan.

<sup>2</sup>Aquaculture Laboratory, Department of Zoology, University of Sialkot, Pakistan.

**Corresponding Author's Email: [nailamalkani@gcu.edu.pk](mailto:nailamalkani@gcu.edu.pk)**

**ABSTRACT:** *Type specimens of organisms provide the basis for the identification of species. Proper phenotypic data supported by genetic evidence is crucial for every type of specimen keeping in view its significance. Several methods of species identification both morphological and genetic are being used. DNA barcoding using a fragment of cytochrome c oxidase subunit I (COI) mitochondrial gene is gaining popularity because of its accuracy and efficiency. In this study, six type specimens of endemic fishes from Pakistan preserved at Stephenson Natural History Museum, GC University, Lahore, Pakistan were analyzed for their genetic diversity from other members of the genus. COI barcode sequences of *Clupisoma naziri*, *Barilius vagra pakistanicus*, *Nemacheilus naziri*, *Nemacheilus griffithi hazarensis*, *Schizothorax skarduensis*, and *Naziritor zhobensis* were obtained and analyzed. The obtained sequences were approximately 655bp long. The average Kimura-two-parameter(K2P) distances from other members of genera were 0.608%, 0.44%, 0.42%, 0.608%, 0.945%, and 1.364% for *Clupisoma naziri*, *Barilius vagra pakistanicus*, *Nemacheilus naziri*, *Nemacheilus griffithi hazarensis*, *Schizothorax skarduensis*, and *Naziritor zhobensis* respectively. The nodes in K2P distance-based NJ (neighbor-joining) trees were supported by high bootstrap values (100%) in all the species. We conclude that COI sequencing provides an effective way of species identification and barcode generation for fish specimens.*

**Keyword:** DNA barcoding, Kimura-two-parameter, *Schizothorax*, *Clupisoma*, *Barilius*

## INTRODUCTION

Natural history museums are a huge repository of genetic information and

many museums hold type specimens in their collection. Traditionally, these specimens are identified based on the morphological characteristics of the organism and grouped according to the similarities with the previously described species. This identification approach relying on phenotypic traits only can lead to controversies due to issues like cryptic species, differences between larval and adult stages, and sexual dimorphism (Friedheim, 2016).

In the present study, fish holotype specimens are chosen from the collection of Stephenson Natural History Museum, GC University, Lahore, Pakistan. The samples chosen for this study are holotypes of *Clupisoma naziri* (1974), *Barilius vagra pakistanicus* (1965), *Nemacheilus naziri* (1975), *Nemacheilus griffithi hazarensis* (1963), *Schizothorax skarduensis* (1974) and *Naziritor zhobensis* (1979). These specimens were collected from different

S.No.	Specimen	Locality
1	<i>Clupisoma naziri</i>	Indus River at Jinnah Barrage Kalabagh
2	<i>Barilius vagra pakistanicus</i>	Rakhi River near Fort Monroe
3	<i>Nemacheilus naziri</i>	Basin of Kaman Beji River near Harnai
4	<i>Nemacheilus griffithi hazarensis</i>	Swat River near Mingora
5	<i>Schizothorax skarduensis</i>	Indus River at Skardu
6	<i>Naziritor zhobensis</i>	Zhob River

Table 1: Capture site of each sample identified as type specimens on a morphological basis, and preserved in 70% ethanol at Stephenson Natural History Museum, GC University, Lahore. *Barilius vagra pakistanicus*, *Schizothorax skarduensis*, and *Naziritor zhobensis* belong to the order Cypriniformes; family Cyprinidae and are generally distributed in Asia and are endemic to Pakistan (Mirza 1975). *Clupisoma naziri* is from the order Siluriformes; family Schilbeidae while *Nemacheilus* species belong to the order

localities of Pakistan (Table 1). Cypriniformes and the family Nemacheilidae. The precisely defined species are significant for biodiversity maintenance, ecological sustainability, and evolutionary mechanisms. (Bailey, 1970). It is therefore important to explore alternative methods of species identification. DNA barcoding used along with morphological characters can prove as an efficient strategy to deal with the problems of identification faced by taxonomists (Chen et al., 2021). This approach uses a less than 1000 base pair

sequence of a genome which can serve as a standard barcode (Yang et al. 2018). For animal species, DNA barcoding, mitochondrial cytochrome c oxidase subunit I (COI) gene is an efficient marker as it is highly conserved across species. This approach accurately identifies various taxa based on the barcode gap and helps to reveal several animal groups that have not been acknowledged yet at any taxonomic level (Hebert et al., 2003; Hajibabaei et al., 2007; Yaqub et al., 2019; Carugati et al., 2022).

Using DNA eliminates the judgment-based arguments amongst scientists on how to define one feature from the other (Lok et al., 2005). The molecular approaches may be utilized to identify a novel sample as per the existing classification (Rahal et al., 2014). Sequences of mitochondrial cytochrome b (*cyt b*), 16S rRNA, and 648 base pair region of mitochondrial cytochrome c oxidase 1 gene (*COI*) are among the most extensively used genetic markers for fish species identification (Kochzius et al., 2010; Kamran et al., 2020). The closely related species can be identified by using these methods (Folmer and Pennington, 2000; Meier et al., 2022). The applicability of *COI* for species identification in fish created the international advantage for barcoding all

fishes (Ward et al., 2005; Appleyard et al., 2022). The Fish Barcode of Life Initiative (FISH-BOL) is a collaborative international research effort, which seeks to establish a reference library of DNA barcodes for all fish species derived from voucher specimens with authoritative taxonomic identifications (Ivanova et al., 2007; Mir et al., 2021). The primary goal of this study is to compile a reference library for DNA barcode sequences of indigenous type specimens of fishes of Pakistan and to re-evaluate the species delimitation on a molecular basis.

## **MATERIALS AND METHODS**

### **Sample acquisition**

Each specimen was carefully removed from the preservation jar in the safety cabinet to avoid any infection the specimen. With the help of a sharp sterile blade, a very small portion of gills were taken without damaging the fish. The acquired sample from each specimen was transferred to a properly labeled 1.5ml eppendorf tube and proceeded with DNA extraction.

### ***DNA extraction and COI amplification***

DNA was extracted by the salt extraction method (Maurya et al., 2013) however; good quality mtDNA could not be extracted from most of the samples despite repeated efforts, as the

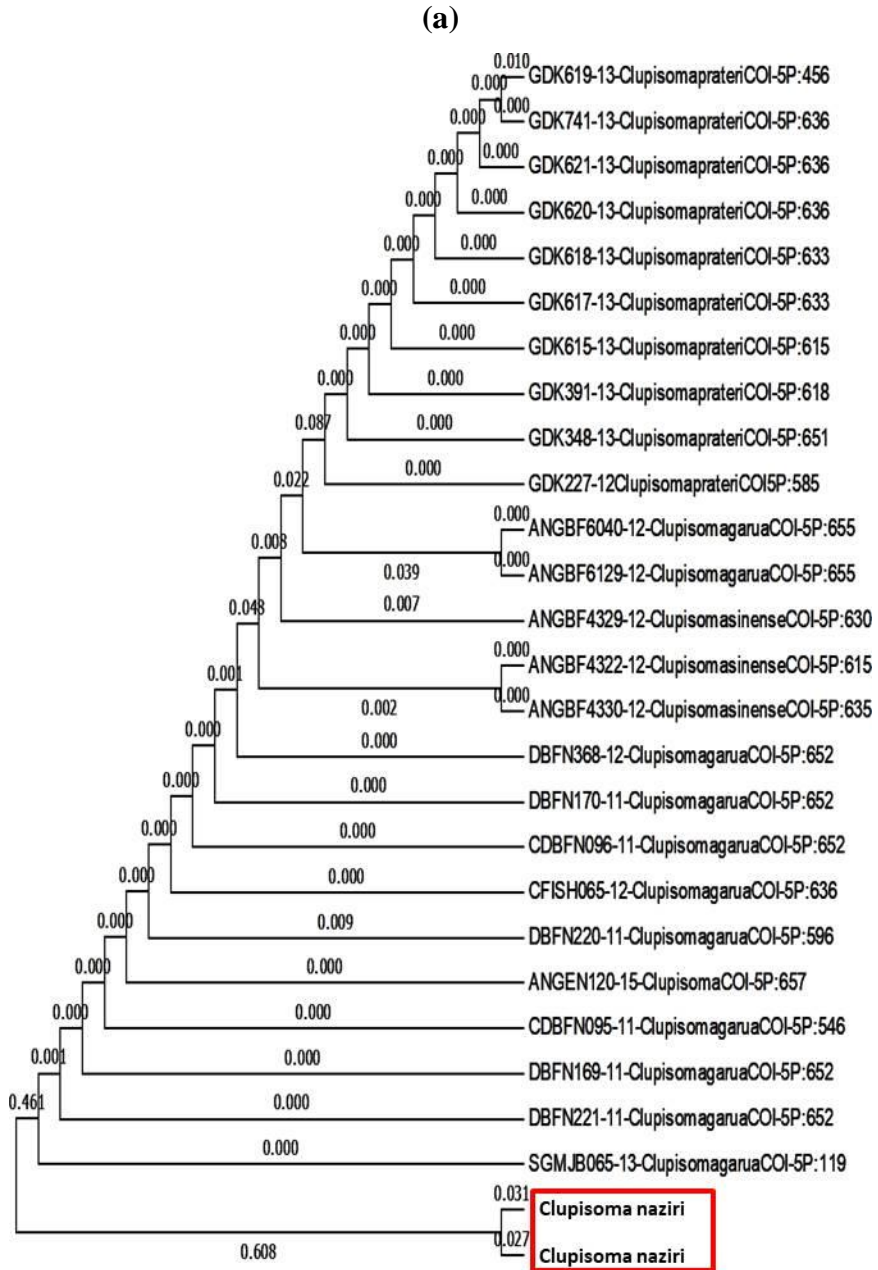
samples were preserved in 10% formalin or 70% ethanol for over four decades. The direct PCR approach from tissue lysate was selected for these specimens. Initially, 50mg of gills were taken from each sample and transferred in 15ml falcon tubes containing 2 ml of lysis buffer (200 mM Tris-HCl pH 8.0; 100 mM EDTA; 250 mM NaCl), 10 $\mu$ l of proteinase K (20 mg/ml) and 60  $\mu$ l of 20% SDS were also added in the mixture. The falcon tubes were kept at 48°C for 2-3 hours in a water bath. After incubation gill tissues were homogenized and the homogenate was subject to standard PCR reaction using primers reported by Ward et al. (2005). (For-TCAACCAACCACAAAGACAT TGGCA; Rev- CTAGACTTCTGGG TGGCAAAGAATCA). PCR reaction was carried out in a thermal cycler (Veriti™ 96-Well Fast Thermal Cycler; Applied Biosystems) with the following conditions: 5 min of initial denaturation at 95 °C; 35 cycles of 30 seconds of denaturation at 95 °C, annealing at 55 °C for 45 seconds and extension at 72 °C for 1 min and final elongation at 72 °C for 10 minutes. The resulting PCR product of 650bp was sent for unidirectional Sanger sequencing to Macrogen Inc. South Korea.

### **Analysis**

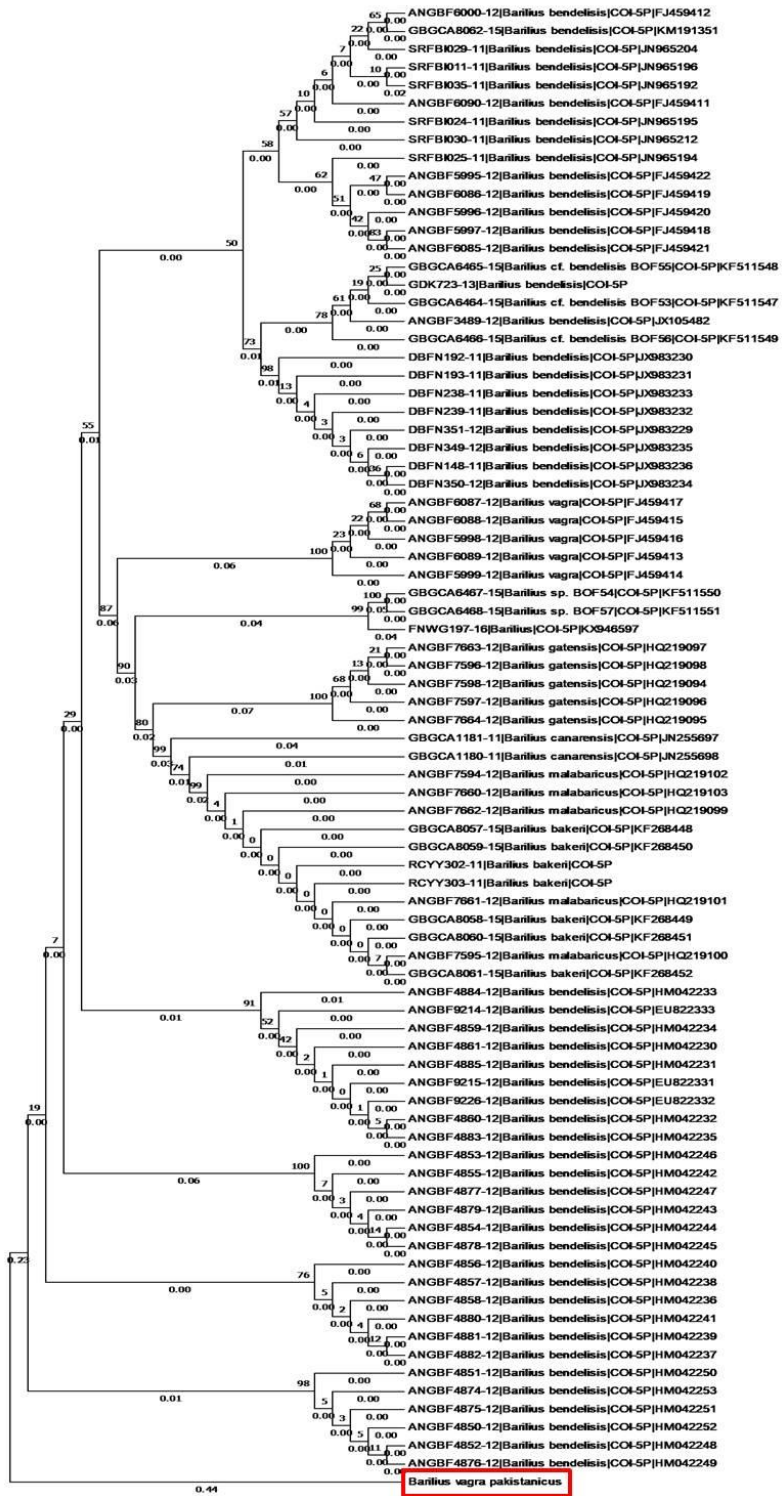
The sequences were aligned with other reported COI genes of the same taxa taken from NCBI database. The evolutionary history was configured using the Neighbour-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch lengths is shown next to the branches. The evolutionary distances were calculated using the Kimura 2-parameter method (Kimura, 1980) and the units represented the number of base substitutions per site. All positions containing gaps and missing data were eliminated. These evolutionary analyses were performed using MEGA7 (Saitou and Nei, 1987). Barcode gap and rank distance was determined using Automatic Barcode Gap Discovery (ABGD) and Barcode Gap Analyses (BGA), respectively.

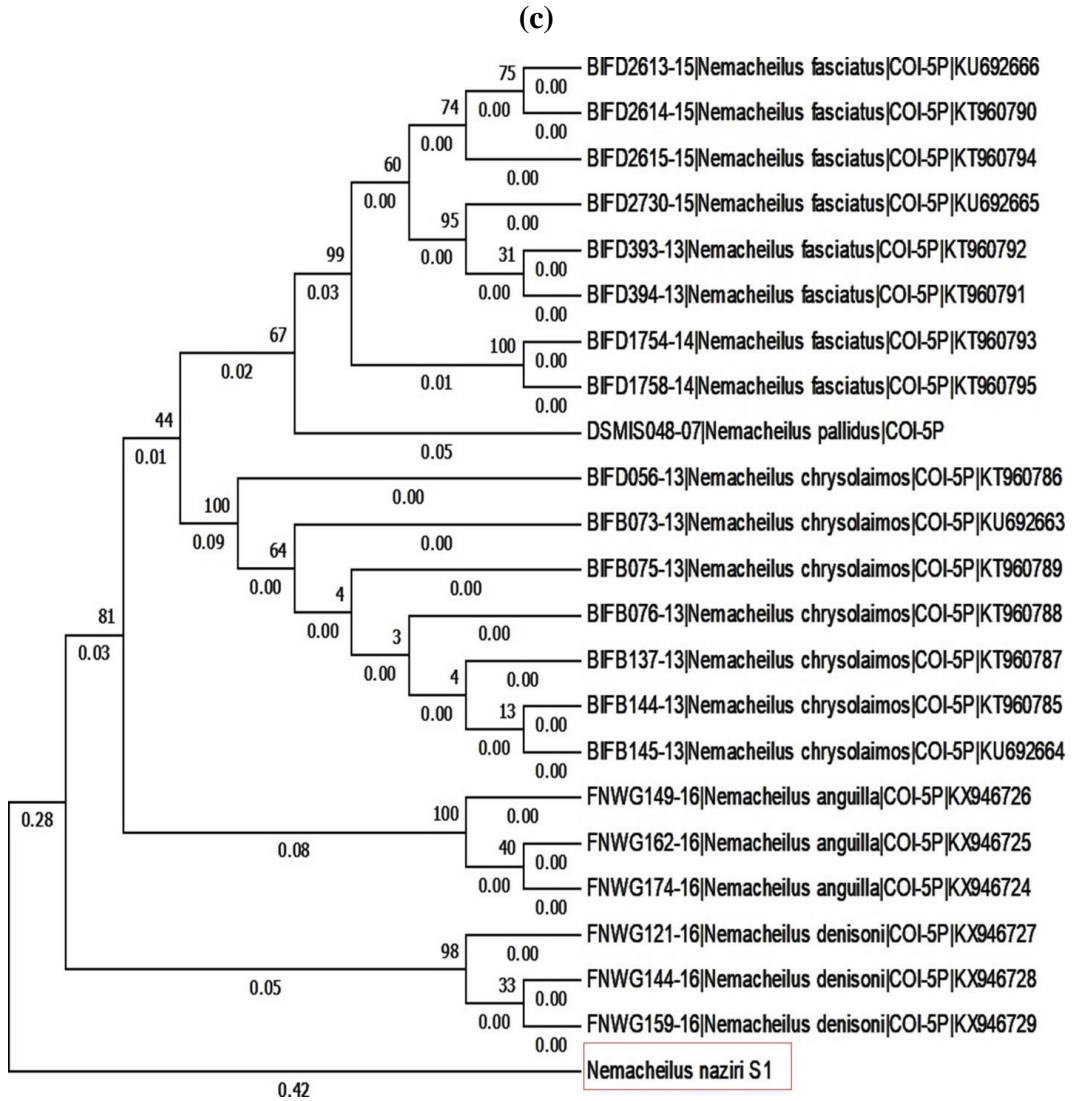
## **RESULTS**

The sequence analysis characterized *Clupisoma naziri*, *Barilius vagra pakistanicus*, *Nemacheilus naziri*, *Nemacheilus griffithi hazarensis*, *Schizothorax skarduensis*, and *Naziritor zhobensis* as six different native fish species, belonging to different genera. The K2P distance of 0.608%, 0.44%, 0.42%, 0.608%, 0.945%, and 1.364% between type specimens and cogenetic species of *Clupisoma naziri*, *Barilius vagra pakistanicus*, *Nemacheilus naziri*, *Nemacheilus griffithi hazarensis*, *Schizothorax skarduensis*, and *Naziritor zhobensis* respectively was found (Fig. 1a, b, c, d, e, and f).

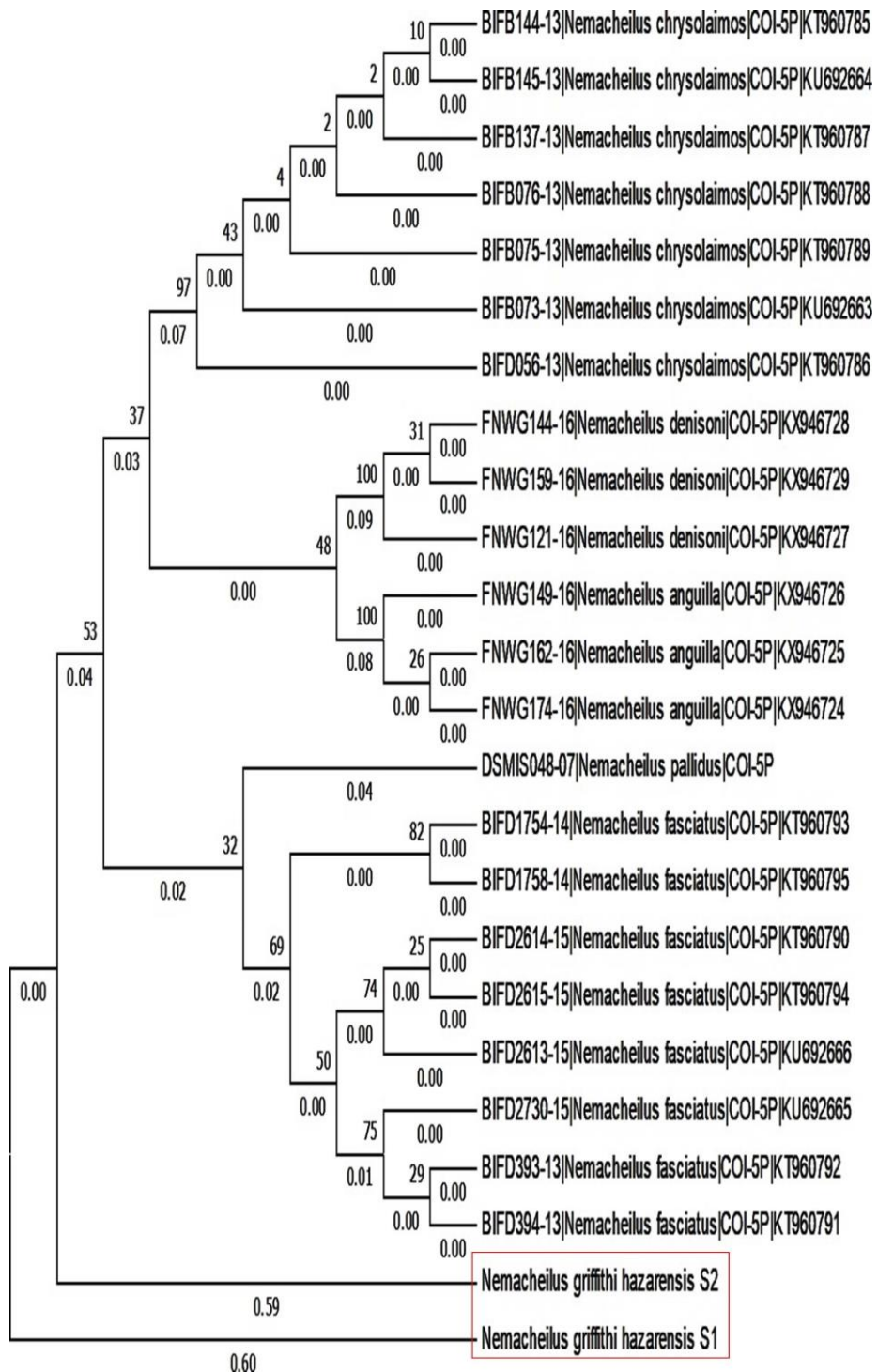


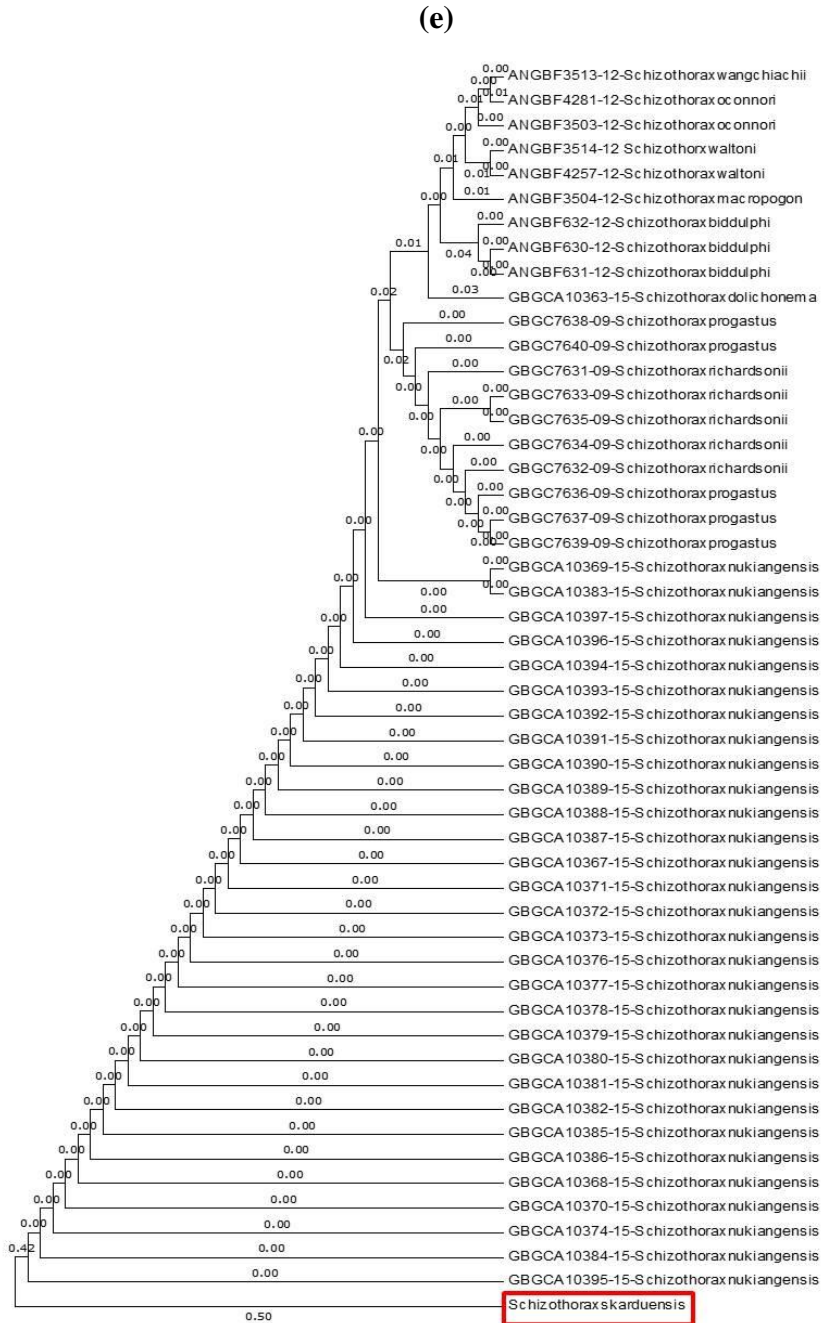
(b)





(d)



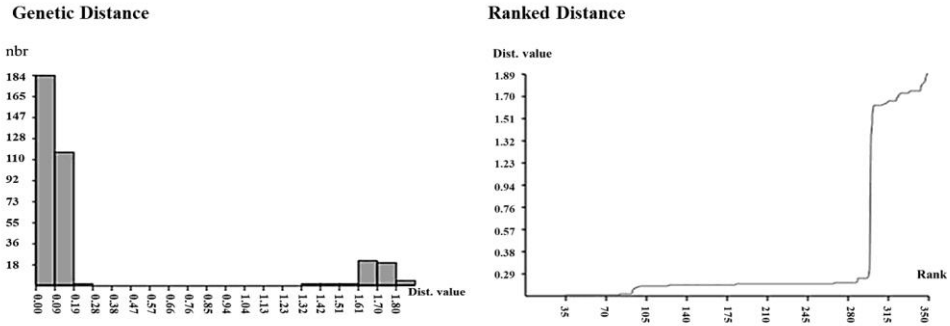


**Fig. 1.** “Evolutionary relationships of taxa of studied fish specimen (a) *Clupisoma naziri*, (b) *Barilius vagra pakistanicus*, (c) *Nemacheilus naziri*, (d) *Nemacheilus griffithi hazarensis*, (e) *Schizothorax skarduensis* and (f) *Naziritor zhobensis* aligned by Neighbor-joining (NJ) method using MEGA 7 software.”

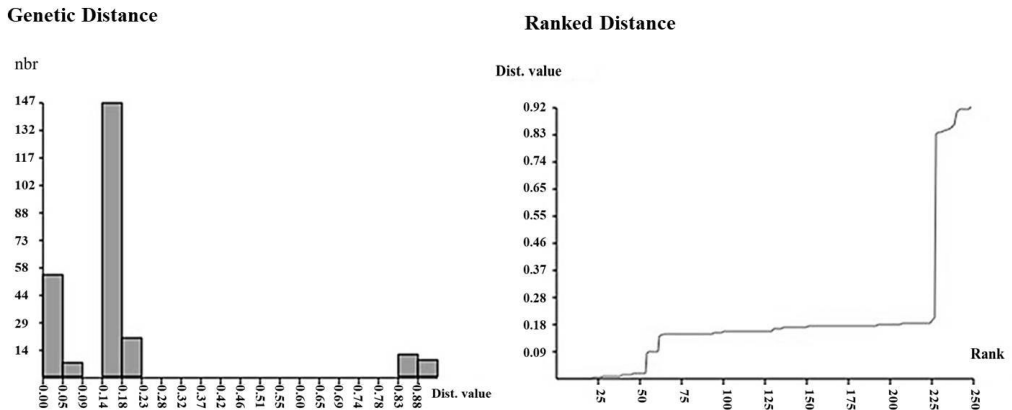
The results of the genetic analysis revealed a clear gap between interspecific and intraspecific distances of the species from other members of

the genus (Fig. 2a, b, c, d, e, and f). The minimum distance to the nearest neighbor (NN) was higher than the maximum intraspecific distance.

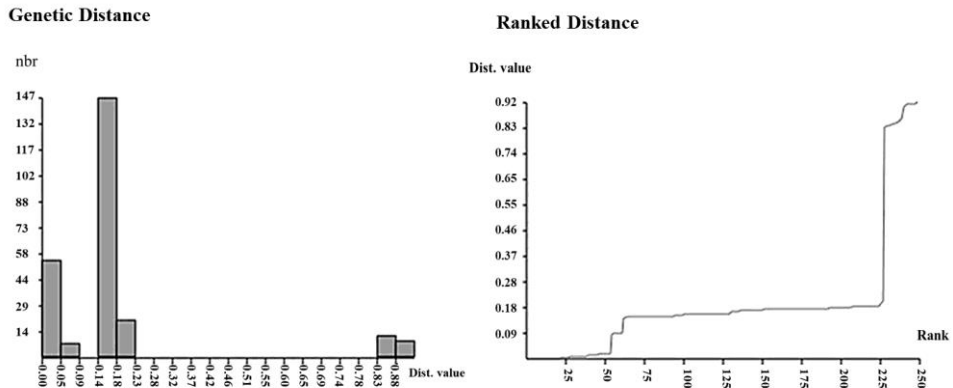
(a)



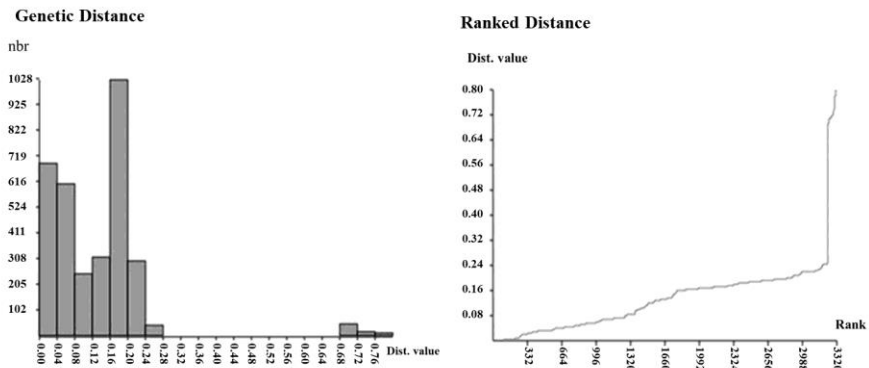
(b)



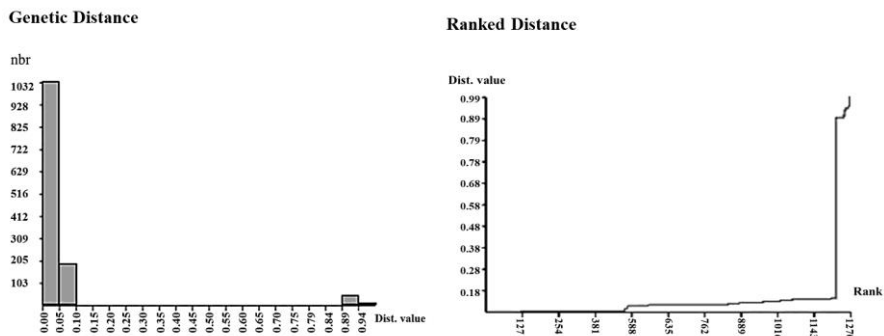
(c)



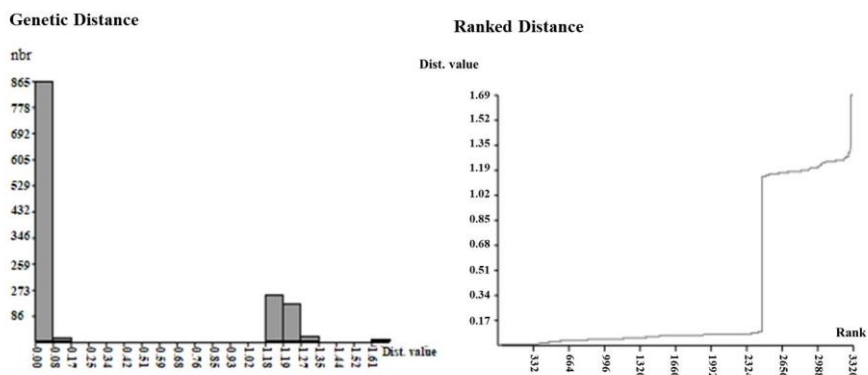
(d)



(e)



(f)



**Fig. 2.** “Histogram of Genetic distance and Ranked distance, created by using ABGD and BGA for (a) *Clupisoma naziri*, (b) *Barilius vagra pakistanicus*, (c) *Nemacheilus naziri*, (d) *Nemacheilus griffithi hazarensis*, (e) *Schizothorax skarduensis* and (f) *Naziritor zhobensis*.”

The FASTA sequences for each species were submitted in the BOLD v3 database under project ID: JANKU and sequence IDs JANKU006-19, JANKU003-19, JANKU001-19, JANKU002-19, JANKU005-19, and JANKU005-19 for *Clupisoma naziri*, *Barilius vagra pakistanicus*, *Nemacheilus naziri*, *Nemacheilus griffithi hazarensis*, *Schizothorax skarduensis*, and *Naziritor zhobensis* respectively were assigned.

## DISCUSSION

Being a permanent reference of a species, type specimens are of critical importance and should be preserved in a way that they are accessible to future scientists. The preservation strategies should be very effective and the specimens should be monitored for their health regularly. However, there is a natural process of decay even at a very slower rate therefore there should be identification tools available other than morphology. Identification of a barcode for a specimen is important as it can serve as a reference even after the physical deterioration of the sample due to an accident. Here we developed a barcode for six important endemic fishes of Pakistan from the type specimens preserved at Stephenson Natural History Museum, GC University, Lahore, Pakistan.

The extraction of DNA from preserved tissues or museum specimens is an elementary component for many scientific researchers (Zimmermann et al., 2008). It has been a basic standard procedure in molecular biology since the 1980s and numerous protocols have been drafted for DNA isolation and subsequent PCR amplification (Jaksch et al., 2016). Methods for DNA isolation and analyzing the sequencing data from specimens not immediately preserved for DNA extraction are improving rapidly (Andersen and Mills, 2012). But still, mitochondrial DNA extraction from several decade-old specimens is very challenging mainly because of the preservation strategies. Several attempts were performed using various protocols and commercially available kits to procure good quality mtDNA from the formalin and ethanol preserved type specimens. But due to the age (more than 40 years) of specimens and preservation solutions, the extraction of DNA was not very successful. Finally, as a despairing attempt, we selected the basic principle of colony PCR. The cell homogenate was directly subjected to PCR using standard fish DNA barcoding primers and amplification was achieved. The sequencing results confirmed the amplification of correct

sequences, which were further used for analysis.

An analysis of the specimen was carried out for generating DNA barcodes. The specific primers for the *COI* gene were employed to produce the barcodes with a mean length of 655 bp for all our samples. Moreover, insertions or deletions, or stop codons were not observed in any of the sequences, justifying the view that all the sequences being amplified are derived from functional *COI* sequences of the mitochondrial genome. The type specimens of *Clupisoma naziri*, *Barilius vagra pakistanicus*, *Nemacheilus naziri*, *Nemacheilus griffithi hazarensis*, *Schizothorax skarduensis*, and *Naziritor zhobensis* were collected from Kalabagh, Fort Monroe, Harnai (basin of Kaman Beji River), River Swat, Skardu and River Zhob (Baluchistan) respectively. Previously, these were characterized morphologically. In our study, a complete K2P model, based on the NJ phenogram, predicts the genetic distance among all specimens that produced a DNA barcode. This DNA barcoding approach sorted these specimen species with 0.608%, 0.44%, 0.42%, 0.608%, 0.945%, and 1.364% distance among individual species of cogeneric species of *Clupisoma naziri*, *Barilius vagra pakistanicus*,

*Nemacheilus naziri*, *Nemacheilus griffithi hazarensis*, *Schizothorax skarduensis*, and *Naziritor zhobensis*. Similar intraspecific variation had been previously observed in the *COI* gene sequences among several teleosts (Lakra et al., 2005). The sequences obtained from the public data portal-BOLD and the consistent analysis of the barcode sequence provided a clear difference in the taxonomic status of all the species under consideration. Moreover, the congeneric and conspecific genetic divergences were analyzed using Kimura's 2 parameter distance followed by the generation of a Neighbor-Joining tree using the MEGA7. These analyses were able to establish phylogenetic relationships among species. From this data, it can be concluded that *Clupisoma naziri*, *Barilius vagra pakistanicus*, *Nemacheilus naziri*, *Nemacheilus griffithi hazarensis*, *Schizothorax skarduensis*, and *Naziritor zhobensis* are not only morphologically different species but also genetically diverse species from other members of the respective genus. The resulted gene sequences may also serve as a milestone for the identification of further species at the molecular level. The DNA barcoding approach separated different species showing average nucleotide divergence among conspecific

individuals. Moreover, it also highlights the support of genetic data along with the morphological evidence for correct species identification.

### **CONCLUSION**

It is concluded from this study that due to complications associated with the morphological identification of species the molecular basis of identification provides an efficient means for classifying the species. In this regard, *COI* provides an excellent choice for the barcode generation of animal species.

### **REFERENCES**

1. Andersen JC, Mills NJ (2012). DNA extraction from museum specimens of parasitic Hymenoptera. *PloS One* 7(10).
2. Appleyard SA, Maher S, Miskiewicz AG, Lara-Lopez A, Matis P, Fielder DS, Suthers IM (2022). Genetic and morphological identification of formalin fixed, preserved larval fishes; can we have the best of both worlds?. *J. Exp. Marine Biol. Ecol.* 553(1): 151763.
3. Bailey RM (1970). A list of common and scientific names of fishes from the United States and Canada. 3<sup>rd</sup> edn. American Fisheries Society Special publication No. 6: 150.
4. Carugati L, Melis R, Cariani A, Cau A, Crobe V, Ferrari A, Follesa MC, Geraci ML, Iglésias SP, Pesci P, Tinti F, Cannas, R (2022). Combined COI barcode-based methods to avoid mislabelling of threatened species of deep-sea skates. *Anim. Conserv.* 25(1): 38-52.
5. Chen Y, Zhu X, Loukopoulos P, Weston LA, Albrecht DE, Quinn JC (2021). Genotypic identification of *Panicum* spp. in New South Wales, Australia using DNA barcoding. *Scientific Rep.* 11(1): 1-10.
6. Folmer O, Pennington M (2000). A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. *Fish Res.* 49(2): 165-178.
7. Friedheim S (2016). Comparison of Species Identification Methods: DNA Barcoding versus Morphological Taxonomy. *Horizons* 1 (1).
8. Hajibabaei M, Singer GA, Hebert PD, Hickey DA (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Gen.* 23(4): 167-172.
9. Hebert PD, Cywinska A, Ball SL (2003). Biological identifications through DNA barcodes. *Proceed. Royal Soc. B: Biologi. Sci.* 270(1512): 313-321.

10. Ivanova NV, Zemlak TS, Hanner RH, Hebert PD (2007). Universal primer cocktails for fish DNA barcoding. *Mol. Ecol. Notes* 7(4): 544-548.
11. Jaksch K, Eschner A, Rintelen TV, Haring E (2016). DNA analysis of molluscs from a museum wet collection: a comparison of different extraction methods. *BMC Reserch Notes* 9(1): 348.
12. Kamran M, Yaqub A, Malkani N, Anjum KM, Awan MN, Paknejad H (2020). Identification and Phylogenetic Analysis of *Channa* Species from Riverine System of Pakistan Using COI Gene as a DNA Barcoding Marker. *J. Biores. Manage.* 7(2): 10.
13. Kimura M (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evo.* 16(2): 111-120.
14. Kochzius M, Nölte M, Weber H, Silkenbeumer N, Hjørleifsdottir S, Hreggvidsson G, Marteinson V, Kappel K, Planes S, Tinti F (2008). DNA microarrays for identifying fishes. *Marine Biotechnol.* 10(2): 207-217.
15. Lakra WS, Singh M, Goswami M, Gopalakrishnan A, Lal K, Mohindra V, Sarkar U, Punia P, Singh K, Bhatt J (2016). DNA barcoding Indian freshwater fishes. *Mitochondrial DNA Part A* 27(6): 4510-4517.
16. Lok AS, Ghany MG, Goodman ZD, Wright EC, Everson GT, Sterling RK, Everhart JE, Lindsay KL, Bonkovsky HL, Di Bisceglie AM (2005). Predicting cirrhosis in patients with hepatitis C based on standard laboratory tests: results of the HALT-C cohort. *Hepatology*. 42(2): 282-292.
17. Maurya R, Kumar B, Sundar S (2013). Evaluation of salt-out method for the isolation of DNA from whole blood: a pathological approach of DNA based diagnosis. *Int. J. Life Sci. Biotechnol. Pharma Res.* 2(2):53-57.
18. Meier R, Blaimer BB, Buenaventura E, Hartop E, von Rintelen T, Srivathsan A, Yeo D (2022). A re-analysis of the data in Sharkey et al.'s (2021) minimalist revision reveals that BINs do not deserve names, but BOLD Systems needs a stronger commitment to open science. *Cladistics*. 38(2): 264-275.
19. Mir RA, Bhat KA, Rashid G, Ebinezer LB, Masi A, Rakwal R, Shah AA, Zargar SM (2021). DNA barcoding: a way forward to obtain

- deep insights about the realistic diversity of living organisms. *The Nucleus*. 64(2): 157-165.
20. Mirza MR (1975). Freshwater fishes and zoogeography of Pakistan. *Bijdr Dierk Amsterdam*. 45(20): 143-180.
21. Rahal A, Kumar A, Singh V, Yadav B, Tiwari R, Chakraborty S, Dhama K (2014). Oxidative stress, prooxidants, and antioxidants: the interplay. *BioMed Res. Int*.
22. Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4(4): 406-425.
23. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD (2005). DNA barcoding Australia's fish species. *Phil. Trans. Royal Soc. B: Biol. Sci.* 360(1462): 1847-1857.
24. Yang F, Ding F, Chen H, He M, Zhu S, Ma X, Jiang L, Li H (2018). DNA barcoding for the identification and authentication of animal species in traditional medicine. *Evidence Based Complementary and Alternative Medicine*.
25. Zimmermann J, Hajibabaei M, Blackburn DC, Hanken J, Cantin E, Posfai J, Evans TC (2008). DNA damage in preserved specimens and tissue samples: a molecular assessment. *Front. Zool.* 5(1): 18.
26. Yaqub A, Kamran M, Malkani N, Anjum KM, Faheem M, Iqbal M, Khan RU (2019). Mitochondrial *coi* gene based molecular identification and phylogenetic analysis in exotic fish (*Oreochromis mossambicus*) of Pakistan. *J. Anim. Plant Sci.* 29(5).