- 1. Edwards, P., Tuan, L. A. and Allan, G. L., A survey of marine trash fish and fish meal as aquaculture feed ingredients in Vietnam. Australian Centre for International Agricultural Research Publication, 2004.
- 2. Padmanabhan, P. and Sujana, K. A., Animal products in traditional medicine from Attappady hills of Western Ghats. *Indian J. Tradit. Knowl.*, 2008, **7**, 326–329.
- 3. Charrier, M. and Rouland, C., The digestive oxidases of the snail *Helix aspersa*: localizations and variations according to the nutritional status. *Can. J. Zool*., 1992, **70**, 2234–2241.
- 4. Flari, V. and Charrier, M., Contribution to the study of carbohydrases in the digestive tract of the edible snail *Helix lucorum* L. (Gastropoda: Pulmonata: Stylommatophora) in relation to its age and its physiological state. *Comp. Biochem. Physiol*., 1992, **102**, 363–372.
- 5. Shimizu, E., Ojima, T. and Nishita, K., cDNA cloning of an alginate lyase from abalone, *Haliotis discus hannai*. *Carbohydr. Res*., 2003, **328**, 2841–2852.
- 6. Suzuki, K. I., Ojima, T. and Nishita, K., Purification and cDNA cloning of a cellulase from abalone *Haliotis discus hannai*. *Eur. J. Biochem*., 2003, **270**, 771–778.
- 7. Andreotti, G., Giordano, A., Tramice, A., Mollo, E. and Trincone, A., Hydrolyses and transglycosylations performed by purified -D-glucosidase of the marine mollusk *Aplysia fasciata. J. Biotechnol*., 2006, **122**, 274–284.
- 8. Suzuki, H., Suzuki, K. I., Inoue, A. and Ojima, T., A novel oligo alginate lyase from abalone, *Haliotis discus hannai*, that releases disaccharide from alginate polymer in an exolytic manner. *Carbohydr. Res*., 2006, **34**1, 809–819.
- 9. Ootsuka, S., Saga, N., Suzuki, K. I., Inoue, A. and Ojima, T., Isolation and cloning of an endo-β-1,4-mannanase from pacific abalone *Haliotis discus hannai*. *J. Biotechnol*., 2006, **125**, 269–280.
- 10. Nikapitiya, C., Oh, C., Whang, I., Kim, C. G., Lee, Y. H., Kim, S. J. and Lee, J., Molecular characterization, gene expression analysis and biochemical properties of  $\alpha$ -amylase from the disk abalone, *Haliotis discus discu*s. *Comp. Biochem. Physiol. B*, 2009, **152**, 271–281.
- 11. Walker, A. J., Dmglen, A. and Shewy, P. R., Bacteria associated with the digestive system of the slug *Deroceras recticulum* are not required for protein digestion. *Soil Biol. Biochem*., 1999, **31**, 387– 394.
- 12. Kiran, D. P. *et al.*, Bacterial diversity in different regions of gastrointestinal tract of giant african snail (*Achatina fulica*). *Microbiology*, 2012, **1**, 415–426.
- 13. Miller, G. L., Measurement of carboxymethyl cellulase activity. *Anal. Biochem*., 1960, **1**, 127–132.
- 14. Maloy, S. R., *Experimental Techniques in Bacterial Genetics*, Jones and Bartlett Publishers, 1990, pp. 161–163.
- 15. Suzuki, M., Yoshida, K. and Ashida, K., Purification and characterization of xylanase from the mid-gut gland of the apple snail (*Pomacea canaliculata*). *Agric. Biol. Chem*., 1991, **55**, 693–700.
- 16. Sawabe, T. *et al.*, *Vibrio gallicus* sp. nov., isolated from the gut of French abalone *Haliotis tuberculata*. *Int. J. Syst. Evol. Microbiol*., 2004, **54**, 843–846.
- 17. Sawabe, T., Hayashi, K., Moriwaki, J., Fukui, Y., Thompson, F. L., Swings, J. and Christen, R., *Vibrio neonatus* sp. nov. and *Vibrio ezurae* sp. nov. isolated from the gut of Japanese abalones. *Syst. Appl. Microbiol*., 2004, **27**, 527–534.
- 18. Hayashi, K. J., Moriwaki, T., Sawabe, F., Thompson, J., Swings, N., Gudkovs, R. C. and Ezur, Y., *Vibrio superstes* sp. nov., isolated from the gut of Australian abalones *Haliotis laevigata* and *Haliotis rubra*. *Int. J. Syst. Evol. Microbiol*., 2003, **53**, 1813–1817.
- 19. Pawar, K. D., Dar, M. A., Rajput, B. P. and Kulkarni, G. J., Enrichment and identification of cellulolytic bacteria from the gastrointestinal tract of giant African snail, *Achatina fulica. Appl. Biochem. Biotechnol*., 2015, **175**, 1971–1980.

20. Gomare, S., Kim, H. A., Ha, J. H., Lee, M. W. and Park, J. M., Isolation of the polysaccharidase-producing bacteria from the gut of sea snail, *Batillus cornutus. Korean J. Chem. Eng*., 2011, **28**, 1252–1259.

ACKNOWLEDGEMENTS. Bhukya Saida is grateful to the Department of Biotechnology, New Delhi, India for financial support. Md. Imran acknowledges the SRF fellowship from DBT, New Delhi.

Received 7 December 2015; revised accepted 1 February 2016

doi: 10.18520/cs/v110/i11/2170-2172

## **Application of computational methods in fish species identification based on mitochondrial DNA sequences**

## **R. K. Negi1,3,\*, Bheem Dutt Joshi2,3 , J. A. Johnson<sup>2</sup> and S. P. Goyal<sup>2</sup>**

<sup>1</sup>Department of Zoology, University of Delhi, Delhi 110 007, India <sup>2</sup>Wildlife Institute of India, Dehradun 248 001, India <sup>3</sup>Department of Zoology and Environmental Sciences, Gurukula Kangri University, Haridwar 249 404, India

**The great discrepancy in sequence divergence of congeneric (0.4–0.6%) and conspecific (3%) individuals makes it difficult to identify species using DNA. A 650 base pair fragment of the cytochrome** *c* **oxidase subunit I (COI) gene from the fish** *Pethia conchonius*  **was analysed using 30 samples. All the samples were identified as** *P. conchonius* **and two other congeneric species showing <2% sequence divergence with 29 samples out of 30 used, of** *P. conchonius***. Two of the** *P. conchonius* **samples clustered with** *Puntius terio***. Using different computational methods, we identified the sequence that was tagged as** *Puntius chola* **in the NCBI database as the** *P. conchonius* **sequence.**

**Keywords:** Character attribute, *Puntius*, sequence divergence.

SPECIES identification is an important aspect for exploring biodiversity, wildlife forensics, ornamental trade and other fields of biology. Fishes are a highly diverse species constituting more than 50% of all vertebrates globally<sup>1,2</sup>. The segment near 5'-terminus of the mitochondrial cytochrome  $c$  oxidase subunit 1 (COI) is frequently used in taxonomy, and has been selected as the barcode region of the entire animal kingdom<sup>2,3</sup>. The effectiveness of this gene has been validated in different animal groups, and

<sup>\*</sup>For correspondence. (e-mail: negi\_gkv@rediffmail.com)



**Figure 1.** Map showing different sampling locations of *Pethia conchonius* in Uttarakhand.

most species that have been studied show distinct barcode arrays. Identifying closely related species (species complexes or hidden species) and those species recently diverged in the evolutionary time-scale, is difficult. The COI gene may be used to identify species, as this gene exhibits low levels of intra specific sequence divergence and high levels of inter-species sequence divergence in closely related  $taxa^{3,4}$ . However, DNA barcoding has been carried out for a large number of species<sup>5,6</sup> and different threshold levels have been reported for conspecific and congeneric genetic distances (0.02–0.06 and 0.09– 0.110 respectively) of different fish species<sup>7,8</sup>. Some studies have used a genetic distance of  $\geq 2\%$  to define species boundary<sup>9,10</sup>. At the same time, some species exhibit a lower sequence divergence than the aforementioned conspecific and congeneric genetic distances. This makes it difficult to use a single threshold level. Therefore, genetic distances vary from species to species $11,12$ .

Among fishes, genus *Puntius*/*Pethia* has more than 130 valid species with many hidden or species complexes $^{13,14}$ . It is difficult to identify species from the DNA database using just one method. Sometimes false results are obtained using online reference repositories<sup>15</sup>. Similaritybased and sequence divergence data may be interpreted incorrectly. Therefore the use of multiple DNA sequence computational methods is suggested in place of single similarity-based methods that yield false results<sup>15</sup>.

In this context, different computational and statistical methods are useful in identifying species from genetic data, such as distance, phylogenetic tree (topology)<sup>2,16</sup>, level of sequence variations<sup>17</sup> and character attribute  $(CA)^{18,19}$  methods. These methods correctly identify recently diverged species. The aim of the present study is to identify *P. conchonius* using COI gene and identify the sequences submitted to NCBI database that have been incorrectly identified (false negative) as well as identify the character attributes of closely related species.

A total 30 samples of *P. conchonius* were used. Ten of these samples were collected from different rivers/ streams in Uttarakhand. The sites include Song river, Diowala (30.18 $\textdegree N$ , 78.13 $\textdegree E$ ), a tributary of the Ganga, Asan river, Vikasnagar  $(30.43\text{°N}, 77.72\text{°E})$ , a tributary of the Yamuna, and Kho river, Kotdwar (29.73°N, 78.52°E), a tributary of the Ramganga (Figure 1). Genomic DNA was obtained from samples (fin clip tissue) using a QIAamp Blood and Tissue Kit according to the manufacturer's protocol (QIAGEN, Germany). A fragment (650 bp) from the extracted DNA was amplified using mitochondrial COI (LCO1490 F-5-GGTCAACAAATCA TAAAGATATTGG3; HCO2198R-5-AAACTTCAGG GTGAC CAA AAAATCA-3') gene<sup>20</sup>. Polymerase chain reaction (PCR) amplification was carried out using  $15 \mu l$ of a PCR master mix with  $1.5 \times$  PCR buffer, 2.5 mM MgCl<sub>2</sub>, 200  $\mu$ M dNTP, 0.4  $\mu$ M of each primer, 0.5 U *Taq* 

## RESEARCH COMMUNICATIONS

**Table 1.** Character attributes observed in different species of genus *Puntius and Pethia.* Nucleotide positions were observed from the complete genome of *Pethia conchonius*

			5	5	5	5	5	5	5	5	5	5	6	6
			6			8	8	8	9	9	9	9	$\overline{0}$	$\theta$
				3	8	$\theta$	7	9		$\overline{c}$	6		6	
Species	Haplotype	Sequence ID	5	5	3		6	5	$\mathbf{0}$	$\overline{2}$	3	4	6	
Pethia conchonius		NC 022856.1	T	G	T	G	A	G	T	А	G	A	T	
P. conchonius	Hap1	$PCD4*(n = 26)$												
P. conchonius	Hap2	PCD5*, 6 $(n = 2)$						С	A	C				
P. conchonius	Hap3	$PCA3*(n = 1)$					G							
P. conchonius	Hap4	$PCA9*(n = 1)$	C											
P. conchonius		AB863607.1 <sup>#</sup>												
P. conchonius		$JO667569.1$ <sup>*</sup>												
P. conchonius		$JX260948.1$ <sup>*</sup>	C											
Puntius terio		$JX260958.1$ <sup>*</sup>	С				G							
P. conchonius		KP712125.1*	С											
Puntius chola		$JX260946.1$ <sup>*</sup>	C											
Pethia khugae		KF511543.1 <sup>#</sup>	C	А		A	G				A			
Pethia khugae		KF511542.1 <sup>#</sup>	C	A	C	A	G				А		С	
P. conchonius		JX260947.1 <sup>#</sup>	C				G							

\*PCD: PC, *Pethia conchonius* (present study); D, Sampling site (Kho river, Kotdwar); A, Sampling site. Doiwala, Song river. #GenBank.



**Figure 2.** Clean peaks of each haplotype generated in the present study highlighted in black.

gold polymerase (MBI Fermantas) and 40 ng of genomic DNA. The thermal cycling parameters of PCR, included initial denaturation at  $94^{\circ}$ C for 2 min and at  $94^{\circ}$ C for 45 sec, annealing at  $45^{\circ}$ C for 1 min and  $72^{\circ}$ C for 1 min and one cycle of a final extension for 20 min at  $72^{\circ}$ C. PCR amplification was checked by loading  $4 \mu l$  of reaction mixture on a  $2\%$  (w/v) agarose gel. The amplified PCR products were then processed for cycle sequencing PCR with their respective forward primers, using a master mixture composition suggested by Applied Biosystems. These products were then subjected to DNA sequencing on an ABI 3130 genetic analyser.

The quality of the sequences was checked on Sequencher 4.7 (Gene Codes, USA). Sequences obtained were edited and cleaned, and finally clean sequences (490 bp) were obtained. ClustalW multiple alignments  $(CMA)$  in Bioedit (version 7.0.9.0)<sup>21</sup> were used for further analysis. The generated sequences were compared with published sequences of COI gene in NCBI database and verified. Phylogenetic identification was carried out using a neighbour-joining (NJ) tree, with a bootstrap value of 1000 in the MEGA 6 software package<sup>22</sup>. Character attributes were identified manually to find the specific nucleotide in the data sets in Bioedit and MEGA.

COI gene sequences of *P. conchonius* were generated successfully from 30 individuals. All the generated sequences were of good quality and were identified as *P. conchonius* with 100% similarity, using BLAST (NCBI). Sequences of each haplotype were submitted to GenBank (accession numbers KT957189–KT957196). Only sharp peaks were considered for mutations (Figure 2). Some other sequences (same genus) were found in the NCBI BLAST search, and these sequences showed 98.8% similarity, but they had been tagged as other species. One false negative sequence was also found. This had been tagged as *P. chola*, it showed 100% similarity with *P. conchonius*. A total of 12 parsimonious sites were observed, with six conserved sites that were different from the other sequences of species, viz. *P. terio* and *P. khugae* for *Pethia conchonius*. In contrast, the haplotype PCD5 $&6(n=2)$  has four base pair differences and a genetic distance of 0.008 from the other *P. conchonius*  generated in the present study, and the sequences available in NCBI could be different species, and hence, need to be analysed, taking into consideration more sequences and morphological differences. The sequence of *P. terio* 



PC = *Pethia conchonius*; PCh =*Puntius chola* (false negative); PT = *Puntius terio*; PK = *Pethia khugae*.

![](_page_3_Figure_2.jpeg)

**Figure 3.** Phylogenetic identification using neighbour-joining tree constructed with a bootstrap value of 1000 and *Pethia ticto* as the outgroup using the MEGA 6 software package.

shows only two base pair sequence differences (T into C at position 5675 and A into G at position 5876), and both these differences are transition mutations. But both nucleotides were shared by different *P. conchonius* individuals at the same positions (shared by two individuals). Another individual also shows six character attributes of *P. khugae* and two compound attributes of *P. conchonius*  and *P. terio* (Table 1).

All species used for species identification show sequence divergence values between 0.000 and 0.028. However, *P. khugae* shows a sequence divergence of more than 2% with one *P. conchonius* individual; with the other individuals, it shows a divergence of less than 2% (Table 2). The 2% sequence divergence rule to separate species in the DNA barcode region<sup>23</sup>, is only followed with a single individual in the present study. Other studies have found  $\langle 1\%$  (only 0.6%) sequence divergence between species of the same genus<sup>11</sup>. This suggests that extra care is necessary when defining a species boundary and identifying species.

In contrast, in some cases, conspecific $24$  sequence divergence levels close to 3% and congeneric sequence divergence levels >8% have been found. However, these data suggest that the species involved have diverged a long time back. Large differences in the intra species sequence divergence may be a consequence of geographic separation and a disjunct species distribution pattern<sup>25-27</sup>. Considering this, different computational methods are needed for identification of morphologically hidden species. Different methods like NCBI, BLAST (similarity), distance, character attributes and phylogenetic-based methods, that can discriminate species exhibiting low

sequence divergence, have been used in our study. Character-based methods in particular are robust at identifying species. All *P. conchonius* and *P. terio* individuals were grouped into a single clade with a low (70%) bootstrap

value, which may be indicative of separate species that diverged recently from a common ancestor. The clade of *P. conchonius* was split into two major clades and further separated in the internal branch, showing intra-species splits. Only one of the 30 *P. conchonius* samples used clustered in the tree with *P. terio* and one sequence from NCBI (JX260947). The topology of the tree is similar to that of some other fishes $24$ . All the three sequences share the compound character attributes (characters present in other species also), as a result of which they cluster in a single clade.

One sequence identified as being from *P. chola*  (JX260946) was identified in NCBI as being from *P. conchonius* by all the three methods, and one sequence identified as being from *P. conchonius* (JX260947.1) was identified as being from *P. terio* (Figure 3).

In conclusion, all the *P. conchonius* sequences were identified successfully. Two samples (Hap-2), showing a four base pair difference, could be different species, but this needs to be clarified using more sequences. One false negative was also identified, named as *Puntius chola*,

## RESEARCH COMMUNICATIONS

where the sequences of *Puntius terio* show very close proximity with *P. conchonius*. The present study suggests that different computational methods need to be used to identify species from DNA sequence data, as there are several drawbacks in the use of online repositories  $15,28$ .

- 1. Mittermeier, R. A. and Mitemeir, C. G., Mega diversity earth's biologically wealthiest nation In Global *S V Rankhamb/Rec Res Sci Tech 3 11-13*13 fresh water Biodiversity sea wind Cemex Mexico city (eds mc Allister Lttamiltion, D. E. A. and Harvery, B.), 1997, pp. 1–140.
- 2. Hebert, P. D. N., Cywinska, A., Ball, S. L. and Waard, J. R., Biological identifications through DNA barcodes. *Proc. R. Soc. B*, 2003, **270**, 313–321.
- 3. Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W. and Hebert, P. D. N., DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 968–971.
- 4. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. and Hebert, P. D. N., DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B*, 2005, **360**, 1847–1857.
- 5. Hogg, I. D. and Hebert, P. D. N., Biological identification of springtails (Collembola : Hexapoda) from the Canadian Arctic using mitochondrial DNA barcodes. *Can. J. Zool.*, 2004, **82**, 749– 754.
- 6. Marshall, E., Taxonomy Will DNA bar codes breathe life into classification? *Science*, 2005, **307**, 1037.
- 7. Bhattacharjee, M. J., Laskar, B. A., Dhar, B. and Ghosh. S. K., Identification and re-evaluation of freshwater catfishes through DNA barcoding. *PLoS ONE*, 2012, **7**(11), e49950; doi: 10.1371/ journal.pone.0049950.
- 8. Lara, A., Ponce de León, J. L., Rodríguez, R., Casane, D., Côté, G., Bernatchez, L. and García-Machado, E., DNA barcoding of Cuban freshwater fishes: evidence for cryptic species and taxonomic conflicts. *Mol. Ecol. Res.*, 2010; doi: 10.1111/j.1755- 0998.2009.02785.x.
- 9. Francis, C. M., Borisenko, A. V., Ivanova, N. V., Eger, J. L. and Lim, B. K., The role of DNA barcodes in understanding and conservation of mammal diversity in southeast Asia. *PLoS ONE*, 2010, **5**(9), e12575; doi: 10.1371/journal.pone.0012575.
- 10. Hebert, P. D., Ratnasingham, S. and deWaard, J. R., Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. London Ser. B*, 2003, **270**  (Supplememt 1), S96–S99.
- 11. Mabragana, E., Dıaz de Astarloa, J. M., Hanner, R., Zhang, J. and Gonzalez Castro, M., DNA barcoding identifies argentine fishes from marine and brackish waters. *PLoS ONE*, 2011, **6**(12), e28655; doi: 101371/journalpone0028655.
- 12. Kannan, K., Johnson, J. A., Kumar, A. and Gupta, S. K., Mitochondrial DNA variation in the endangered fish *Dawkinsia tambraparniei* (Actinopterygii : Cypriniformes : Cyprinidae) from southern Western Ghats, India. *Acta Ichthyol. Piscat*., 2014, **44**, 3–8.
- 13. Jayaram, K. C., Revision of the Genus Puntius Hamilton from the Indian Region (Pisces: Cypriniformes Cyprinidae Cyprininae). Occasional Paper No 135 Records of the Zoological Survey of India Kolkata, 1991, p. 178.
- 14. Pethiyagoda, R., Meegaskumbura, M. and Maduwage, K., A synopsis of the South Asian fishes referred to *Puntius*  (Pisces: Cyprinidae). *Ichthyological Exp. Freshwaters*, 2012, **23**, 69–95.
- 15. Koski, L. B. and Golding, G. B., The closest BLAST hit is often not the nearest neighbour. *J. Mol. Evol*., 2001, **52**, 540–542.
- 16. Hajibabei, M., Singer, G. A. C., Clare, E. L. and Hebert, P. D. N., Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biol*., 2007, **5**, 1–15.
- 17. Galan, M., Pages, M. and Cosson, J. F., Next-generation sequencing for rodent barcoding: species identification from fresh degraded and environmental samples. *PLoS ONE*, 2012 **7**(11), e48374.
- 18. Sarkar, I. N., Planet, P. J., Bael, T. E., Stanley, S. E. and Siddall, M., Characteristic attributes in cancer microarrays. *J. Biomed. Inform*., 2002, **35**, 111–122.
- 19. Sarkar, I. N., Planet, P. J. and Desalle, R., CAOS software for use in character based DNA barcoding. *Mol. Ecol. Resour*., 2008, **8**, 1256–1259.
- 20. Folmer, O., Black, M., Hoeh, W., Lutz R. and Vrijenhoek, R., DNA primers for amplification of mitochondrial cytochrome *c*  oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol*., 1994, **3**, 294–299.
- 21. Hall, T. A., Bioedit: a user friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acid Symp. Ser.*, 1999, **41**, 95–98.
- 22. Tamura, K., Stecher, G., Peterson, D. Filipski, A. and Kumar, S., MEGA6: molecular evolutionary genetics analysis Version 6.0. *Mol. Biol. Evol.*, 2013, **30**, 2725–2729.
- 23. Ward, R. D., Hanner, R. and Hebert, P. D. N., The campaign to DNA barcode all fishes. *J. Fish. Biol*., 2009, **74**, 329–356.
- 24. Pereira, L. H. G., Maia, G. M. G., Hanner, R., Foresti, F. and Oliveira, C., DNA barcodes discriminate freshwater fishes from the Paraıba do Sul River Basin Sao Paulo Brazil. *Mitoch. DNA*, 2010, **21**, S2 1–S2 9.
- 25. Cowling, R. M., Richardson, D. M. and Pierce, S. M., Vegetation of Southern Africa. Cambridge: Cambridge University Press, 1997, p. 615.
- 26. Gaston, A. J. and Zacharias, V. J., The recent distribution of endemic and disjunct birds in Kerala State: preliminary results of an ongoing survey. *J. Bom. Nat. Hist. Soc*., 1996, **93**, 389– 400.
- 27. Jayaram, K. C., Ecology and distribution of freshwater fishes, amphibians, and reptiles. In ref. 1, pp. 517–580.
- 28. Collins, R. A. *et al.*, Barcoding and border biosecurity: identifying cyprinid fishes in the aquarium trade. *PLoS ONE*, 2012 **7**(1), e28381; doi: 10.1371/journal.pone.0028381.

ACKNOWLEDGEMENTS. We extend our heartfelt thanks to the Dean, Faculty of Life Sciences and Head, Department of Zoology and Environmental Sciences, Gurukula Kangri University, Haridwar for their consistent support during the work. We also thank the Director and Dean, Wildlife Institute of India, for encouraging this work. We would also like to express our gratitude to our lab mates, who shared their personal experiences with us. We also thank the technical staff of the laboratories of the university and the institute.

Received 23 September 2015; revised accepted 28 January 2016

doi: 10.18520/cs/v110/i11/2172-2176