SHORT COMMUNICATION

DNA barcoding of gobiid fishes (Perciformes, Gobioidei)

Divya Viswambharan[1], A. Pavan-Kumar[1], Dhirendra P. Singh[1], A. K. Jaiswar[1], S. K. Chakraborty[1],
J. Rajashekharan Nair[2], and W. S. Lakra[1]

[1]Division of Fish Genetics and Biotechnology, Central Institute of Fisheries Education, Mumbai, Maharashtra, India and [2]Department of Fishery Biology, Kerala University of Fisheries and Ocean Studies, Kochi, Kerala, India

Abstract

Gobiids constitute a major proportion of fish population in both tropical and temperate freshwater as well as marine ecosystem. Due to their small size, cryptic ecology and ambiguous morphological characters, gobiids diversity was not documented completely. In this study, DNA barcodes were generated for 11 species of gobiids, collected from the Ashtamudi Lake, India. The mitochondrial COI gene was amplified using universal primers and the resulted 650 bp amplicon was sequenced. The COI barcodes clearly distinguished all the species with high inter-specific genetic distance values than intra-specific values based on K2P (Kimura 2 Parameter) model. The average genetic distance (K2P model) within species, genus and family was 1.2%, 22.2% and 25.3%, respectively. In addition to barcode-based species identification system, Nucleotide Diagnostic (ND) characters specific for species were identified. The Neighbor-Joining tree revealed distinct clusters shared by the species of same genera.

Introduction

Gobiids are estimated to constitute 35% of the total number of fishes and 20% of the species diversity and they occupy both tropical and temperate environment (Winterbottom et al., 2011). However, due to their small size and often cryptic ecologies, the gobiid diversity was not characterized completely and was often unnoticed. The traditional morphological tools for species identification are constrained by phenotypic plasticity, life stage-specific identification cues and the occurrence of cryptic species. DNA-based approach exploiting sequence diversity among species can be used to identify fishes and to resolve taxonomic ambiguities. Hebert et al. (2004) have demonstrated that the Cytochrome c oxidase subunit I (COI) region is appropriate for discriminating closely related species across diverse animal phyla and this has been used for marine as well as freshwater fishes widely (Hajibabaei et al., 2005; Hubert et al., 2008; Lakra et al., 2011; Steinke et al., 2005; Ward et al., 2008).

This work was carried out with an objective of developing DNA barcodes for Gobiids from Ashtamudi Lake, the second largest wetland ecosystem in India. The aim of the study also includes analyzing Nucleotide diagnostics (NDs) characters for these fishes and supplementing the BOLD database with Indian haplotypes as there is lack of species-specific information.

Materials and methods

A total of 29 Gobiid specimens belonging to 11 species were collected from the Ashtamudi Lake (Lat 8.95 °N and Long 76.6 °E) in Kerala, India, for which the taxonomy and GenBank accession numbers are provided (Table 1). The specimens were identified based on morphological characters using ''Smiths' sea fishes'' (1986) and preserved in absolute alcohol for further molecular study. Total genomic DNA was isolated from the muscle tissue according to the SDS-phenol/chloroform method of Sambrook et al. (2001) with some modifications. The cytochrome c oxidase subunit I (COI) gene was amplified with primers FishF1-5′ TCAACCAACCACAAAGACATTGGCAC 3′ and FishR-1 5′ TAGACTTCTGGG TGGCCAAAGAATCA 3′ (Ward et al., 2005) in a 50-μl volume with 100 ng template DNA, 10 pmol of each specific primer, 200 μM of each dNTPs, 1.0 unit of Taq DNA polymerase and 1x Taq buffer containing 1.5 mM $MgCl_2$. The PCR conditions were initial denaturation at 94 °C for 3 min, followed by 35 cycles of 40 s at 94 °C, 40 s at 54 °C, 60 s at 72 °C and final extension at 72 °C for 10 min. The PCR products were visualized on 1.5% agarose gel and the amplicons were purified by Gel extraction kit (Fermentas, Waltham, MA) following the manufacture's protocol. The purified products were sequenced commercially by Eurofins Pvt Ltd., Bangalore, India.

Sequence analysis

The COI partial gene sequences obtained for each species were manually assembled using Generunner software (New York, NY). Assembled sequences were end-trimmed to a homologous region to avoid sequencing errors and aligned using Clustal W (Thompson et al., 1997). Only those sequences with more than 550 bp in size were used for the analysis. Sequence divergence values within and between species were calculated using Kimura two Parameter (K2P) distance model implemented in MEGA V.5.0 (Tamura et al., 2011) software (Arizona). The pair-wise deletion option was selected to account for missing sequence information between each compared specimen. Nucleotide Diagnostics (NDs) for each species of interest were identified in the context of the entire pool of available species by MEGA

Correspondence: A. Pavan Kumar, Scientist, Division of Fish Genetics and Biotechnology, Central Institute of Fisheries Education, Versova, Andheri West, Mumbai-61, Maharashtra, India. Tel: 022-26361447, Ext. 452. E-mail: pavankumar@cife.edu.in

V.5.0 (Tamura et al., 2011) software. The COI gene sequence substitution saturation was tested by plotting the number of transitions and transversions against genetic divergence [Kimura 2-parameter (K2P)] using DAMBE version 4.2.13 (Xia & Xie, 2001). The Neighbor-joining (NJ) tree of COI K2P distance was created to give a graphical representation of the divergence pattern between species (Saitou & Nei, 1987) with 1000 bootstrap replications.

Results and discussion

A total of 29 COI sequences were generated from 11 gobiid fish species. Sequence alignment yielded 650 bp per taxon after exclusion of the primer sequence. Simplicity and unambiguity were observed among all the sequences. No insertions, deletions or stop codons were observed in any of the sequence and it proves the absence of NUMTS (Nuclear Mitochondrial DNA, i.e. nuclear DNA sequences originating from mitochondrial DNA). Richly & Leister (2004) reported occurrence of NUMTS (transfers of mtDNA COI sequences into the nuclear genome) after comparing *Fugu rupripes* mtDNA and nuclear DNA sequences. DNA barcoding can overestimate the number of species, if nuclear mitochondrial pseudogenes are co-amplified (Song et al., 2008).

The COI sequence analysis across species in this study revealed average nucleotide frequencies as 24.9% (A), 29.3% (T), 27.1% (C) and 18.7% (G). Min & Hickey (2007) showed strong correlation between the GC composition of COI gene and the entire mitochondrial genome. The overall GC content (45.8%) across the species was slightly lower than the reported GC content for marine teleosts (Lakra et al., 2011; Ward et al., 2005). A higher overall GC content was observed in Eleotridae family (47.0 ± 0.10) compared to Gobiidae family (45.0 ± 0.12) (Table 2). This difference was attributable to the GC content of the codon 3rd base position (43.2% versus 38.3%). Among species, variation in GC content for codon 3rd base position was higher than for position 1, with variation at position 2 being very limited (Figure 1). This showed the absence of synonymous mutations at position 2.

In this study, the average number of transitions (transitional pairs $S_i = 61$) for COI gene is more than the average number of transversions (transversional pairs $S_v = 50$) with an average ratio of 1.22. It indicated that the COI sequence for gobiids is not saturated and can discriminate species. The sequence substitution saturation analysis also showed that the sequences were not saturated as the transitions and transversions were linear when plotted against genetic distance (K80 model) (Figure 2).

Table 1. List of species along with GenBank accession numbers.

S.No	Order	Family	Subfamily	Genus	Species	No. of specimens	Accession number
1.	Perciformes	Eleotridae	Eleotrinae	*Eleotris*	*fusca*	1	JX193751
2.			Butinae	*Butis*	*butis*	2	JX193740- 41
3.		Gobiidae	Amblyopinae	*Trypauchen*	*vagina*	1	JX193742
4.				*Odontamblyopus*	*rubicundus*	3	JX193743- 45
5.			Gobionellinae	*Oxyurichthys*	*ophthalmonema*	3	JX193748- 50
6.				*Oligolepis*	*acutipennis*	2	JX193738- 39
7.			Gobiinae	*Acentrogobius*	*chlorostigmatoides*	3	JX193727- 29
8.				*Acentrogobius*	*audax*	4	JX193752- 55
9.				*Arcygobius*	*baliurus*	5	JX193733- 37
10.				*Bathygobius*	*fuscus*	2	JX193746- 47
11.				*Psammogobius*	*biocellatus*	3	JX193730- 32

Table 2. Nucleotide composition of COI gene across gobiids.

Sub families	A	G	C	T	GC	GC1	GC2	GC3
Eleotridae	24.3 ± 0.07	18.8 ± 0.15	28.2 ± 0.10	28.8 ± 0.14	47.0 ± 0.12	54.3 ± 0.14	43.3 ± 0.02	43.2 ± 0.12
Gobiidae	24.9 ± 0.05	18.7 ± 0.14	27.1 ± 0.10	29.3 ± 0.12	45.8 ± 0.10	55.6 ± 0.15	43.4 ± 0.01	38.3 ± 0.14

Figure 1. Variation in GC content of COI gene among Eleotridae and Gobiidae family.
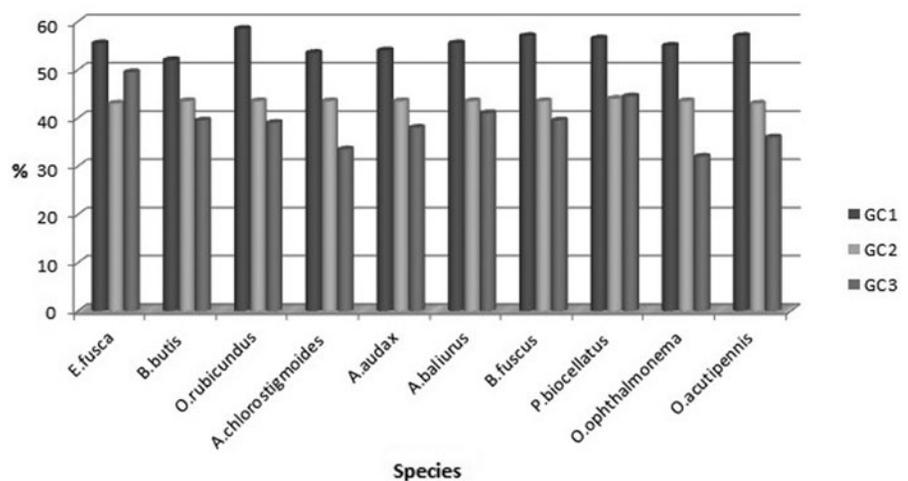
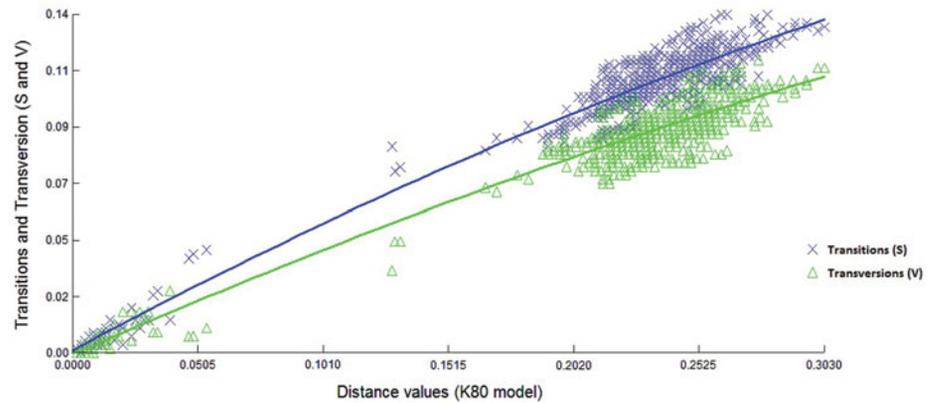Figure 2. Sequence saturation plots for COI gene of gobiids.

Table 3. Summary of COI genetic divergences (K2P percentage) of gobiids within various taxonomic levels.

Comparison within	Minimum	Maximum	Mean ± S.E.
Species	0.00	5.4%	1.2% ± 0.02
Genera	21.2	23.7	22.2% ± 0.03
Family	21.2	30.3	25.3% ± 0.03
Order	21.0	31.2	25.7% ± 0.10

Table 4. NDs for each species, with sample sizes included in parentheses after the species name.

S.No	Species	NDs (nucleotide diagnostics)
1.	*Oligolepis letipennis* (07)	3-G;18-G;**93-A**;195-A;201-T; **255-G;303**-G;**534-A;594-T**
2.	*Oxyurichthys ophthalmonema* (06)	186-G;558-C
3.	*Psammogobius biocellautus* (04)	**252-G;261-T;294-T;315-C; 378-G;390-C**;445-G;446-C; 459-G;513-G; **554-A;**570-T
4.	*Bathygobius fuscus* (04)	**180-C;595-C**
5.	*Arcygobius baliurus* (05)	**57-A;186-A**;291-A;**480-A**
6.	*Acentrogobius audax* (04)	**111-C**;150-T;**249-A;318-T; 321-T;345-G;450-G; 465-G;**504-G
7.	*Acentrogobius chlorostigmoides* (03)	**15-A;63-C;183-A;309-T;579-C**
8.	*Odantamblyopus rubicundus* (04)	**9-A**;43-G;**129-T;135-T; 357-C;525-A;552-T**
9.	*Butis butis* (03)	**12-G;231-G;249-G;285-G; 348-C**;448-A;**516-G**

For species with at least one sND, all sNDs are listed. Nucleotide positions and character states that differ from the species pool by a transversion are in bold.

The average K2P (Kimura 2 Parameter method) distance values for COI increased from lower taxa toward the higher taxonomic rank and also the average congeneric distance value (D = 22.2%) was 18 times more than the average conspecific distance (D = 1.2%) in the present observation which was higher than the previous reports (D = 20.87%) (Table 3). Whereas, the mean divergence values among genera (25.3%) and among families (25.7%) were slightly lower than the previous reports (Hyung-Bae et al., 2012). The gradual increase of genetic variation through increment of taxonomic level clearly indicates

the efficacy of DNA barcodes to discriminate the species. Cluster analysis indicated that the neighbor joining tree of COI for gobiids clearly discriminated all the species and clustered the similar species under same nodes with significant bootstrap values (Figure 3).

The nucleotide diagnostic (ND) approach was applied to strengthen the utility of the DNA barcoding technique to identify species. This approach is similar to the traditional morphology-based methods, where species identification would be based on the presence or absence of a distinct morphological feature. In ND analysis, a distinct DNA character will be assessed which is specific to the species, thereby preventing the ambiguity found in analogue measurement (Wong et al., 2009). All species of each genus were analyzed together to determine the NDs of each species. In this study, total of nine species showed 2 to 10 exclusive ND (simple Nucleotide diagnostic character) in 57 informative sites (Table 4). Most of the NDs were transversions than transitions. Wong et al. (2009) successfully demonstrated the ND approach to discriminate closely related sharks.

DNA barcodes could provide valuable information for conservation biology, like rapid and accurate identification of species, evolutionary histories and phylogenetic diversity. Ardura et al. (2010) have generated DNA barcodes for commercially important fish species from Amazon River and found that seven fish species were commercialized under the same name (Acará). They showed that the DNA barcodes could be used for monitoring species exploitation rate and species-specific trends in trade. Phylogenetic diversity, which measures taxonomic divergence between species, is calculated as the sum of the lengths of all the branches that are members of the corresponding minimum spanning path (Faith, 1992). An index of phylogenetic diversity can inform conservation strategies by using evolutionary lineages (phylogenies) among taxa (Mitchell, 2008). The DNA barcodes developed in this study could be useful for assessing gobiids diversity and monitoring their trade in ornamental fishery. Further, COI sequences reported in this study will supplement the previous reports for estimating phylogenetic diversity values of gobiids.

Conclusion

DNA barcodes were developed for 11 species of gobiids from west coast of India (Astamudi Lake). Barcodes discriminated all the analyzed species in congruence with morphological taxonomic analysis. This study also attempted to identify diagnostic nucleotide characters in COI region specific to species. Our study makes an important contribution to BOLD database as some of these species are not represented in the database.

Figure 3. Neighbor joining tree of COI gene
sequences derived from gobiids.
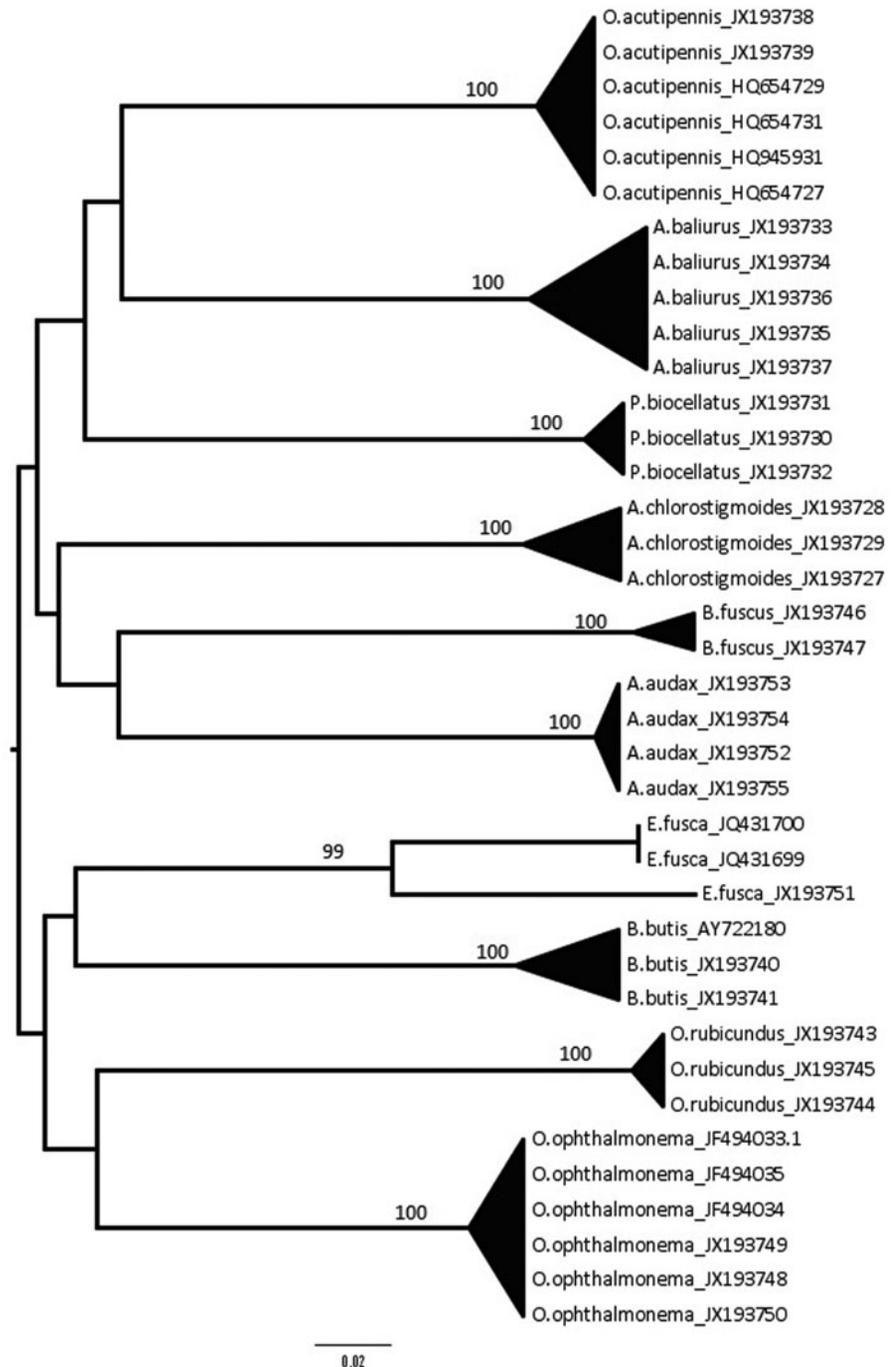
Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Ardura A, Linde AR, Moreira JC, Garcia-Vasquez E. (2010). DNA barcoding for conservation and management of Amazonian commercial fish. Biol Conserv 143:1438–43.

Faith DP. (1992). Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10.

Hajibabaei M, Smith AA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN. (2005). A minimalist barcode can identify a specimen whose DNA is degraded. Mol Ecol Notes 6:959–64.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. (2004). Identification of birds through DNA barcodes. PLoS Biol 2:e312. doi:10.1371/journal.pbio.0020312.

Hubert N, Hanner R, Hom E, Mandrak NE, Taylor E, Burridge M, Watkinson D, et al. (2008). Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE 3:e2490. doi:10.1371/journal.pone.0002490.

Hyung-Bae J, Seung-Ho C, Ho YS. (2012). Exploring the utility of partial cytochrome *c* oxidase subunit 1 for DNA barcoding of gobies. Anim Syst Evol Divers 28:269–78.

Lakra WS, Verma MS, Goswami M, Lal KK, Mohindra V, Punia P, Gopalakrishnan A, et al. (2011). DNA barcoding Indian marine fishes. Mol Ecol Res 11:60–71.

Min XJ, Hickey DA. (2007). DNA barcodes provide a quick preview of mitochondrial genome composition. PLoS ONE 2:e325. doi:10.1371/journal.pone.0000325.

Mitchell A. (2008). DNA barcoding demystified. Aust J Entomol 47:169–73.

Richly E, Leister D. (2004). NUMTs in sequenced eukaryotic genomes. Mol Biol Evol 21:1081–4.

Saitou N, Nei M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–25.

Sambrook SJ, Russel DW, Janssen KA, Irwuin NJ. (2001). Molecular cloning, a laboratory manual (Third Edition). New York: Cold Spring Harbor Laboratory Press.

Smith MM, Heemstra PC. (1986). Smiths' sea fishes (First Edition). Johannesburg, South Africa: Macmillan.

Song H, Buhay JE, Whiting MF, Crandall KA. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. PNAS 36:13486–91.

Steinke D, Vences M, Salzburger W, Meyer A. (2005). Taxi – A software tool for DNA barcoding using distance methods. Philos Trans R Soc Lond B Biol Sci 360:1975–80.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–9.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res 25:4876–82.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. (2005). DNA barcoding Australia's fish species. Philos Trans R Soc Lond B Biol Sci 360:1847–57.

Ward RD, Holmes BH, White WT, Last P. (2008). DNA barcoding Australasian chondrichthyans: Results and potential uses in conservation. Mar Fresh Wat Res 59:57–71.

Winterbottom R, Alofs KM, Marseu A. (2011). Life span, growth and mortality in the western Pacific goby *Trimmabenjamini*, and comparisons with *T. nasa*. Environ Biol Fish 91:295–301.

Wong EHK, Shivji MS, Hanner RH. (2009). Identifying sharks with DNA barcodes: Assessing the utility of a nucleotide diagnostic approach. Mol Ecol Resour 9:243–56.

Xia X, Xie Z. (2001). DAMBE: Data analysis in molecular biology and evolution. J Hered 92:371–3.