MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 60–71 doi: 10.1111/j.1755-0998.2010.02894.x

DNA BARCODING

# DNA barcoding Indian marine fishes

W. S. LAKRA,* M. S. VERMA,* M. GOSWAMI,* K. K. LAL,* V. MOHINDRA,* P. PUNIA,* A. GOPALAKRISHNAN,* K. V. SINGH,* R. D. WARD† and P. HEBERT‡
*National Bureau of Fish Genetic Resources, Lucknow-226002, India, †CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, Tasmania 7001, Australia, ‡Biodiversity Institute of Ontario, University of Guelph, Ontario, Canada N1G 2WI*

## Abstract

**DNA barcoding has been adopted as a global bio-identification system for animals in recent years. A major national programme on DNA barcoding of fish and marine life was initiated in India by the authors during 2006 and 115 species of marine fish covering Carangids, Clupeids, Scombrids, Groupers, Sciaenids, Silverbellies, Mullids, Polynemids and Silurids representing 79 Genera and 37 Families from the Indian Ocean have been barcoded for the first time using cytochrome c oxidase I gene (COI) of the mtDNA. The species were represented by multiple specimens and a total of 397 sequences were generated. After amplification and sequencing of 707 base pair fragment of COI, primers were trimmed which invariably generated a 655 base pair barcode sequence. The average Kimura two parameter (K2P) distances within species, genera, families, orders were 0.30%, 6.60%, 9.91%, 16.00%, respectively. In addition to barcode-based species identification system, phylogenetic relationships among the species have also been attempted. The neighbour-joining tree revealed distinct clusters in concurrence with the taxonomic status of the species.**

*Keywords*: barcoding, Indian fishes, mtDNA, phylogeny

*Received 2 September 2009; revision accepted 13 May 2010*

## Introduction

Taxonomic ambiguity exists for several fish genera/species, and a proper identification is imperative for management and trade. DNA-based approaches for taxon diagnosis exploiting DNA sequence diversity among species can be used to identify fishes and resolve taxonomic ambiguity including the discovery of new/cryptic species (Hebert *et al.* 2003). India has a rich natural heritage and nurtures a unique bio-diversity, placing it among the 12 most biodiverse countries. Out of 31 100 extant fish species, 2438 are known from Indian subcontinent (Froese & Pauly 2009).

A global DNA-based barcode identification system that is applicable to all animal species will provide a simple, universal tool for the identification of fish species and products. The barcode system is based on sequence diversity in a single gene region (a section of the mitochondrial DNA cytochrome c oxidase I gene, COI). When the reference sequence library is in place, new specimens and products can be identified by comparing their DNA barcode sequences against this barcode reference library.

Correspondence: W. S. Lakra, Tel: +91-522-2442441;
Fax: +91-522-2442403;
E-mail: wslakra@gmail.com

Hebert *et al.* (2004a,b) have demonstrated that the COI region is appropriate for discriminating between closely related species across diverse animal phyla and this has been used for marine and freshwater fishes (Hajibabaei *et al.* 2005; Steinke *et al.* 2005; Ward *et al.* 2005; Hubert *et al.* 2008; Lakra *et al.* 2009). Empirical support for the barcoding concept ranges from studies on invertebrates to birds. Currently, DNA barcoding is being employed to a large variety of organisms ranging from yeasts to humans (Hebert *et al.* 2004a,b; Hogg & Hebert 2004; Moritz & Cicero 2004).These results have prompted international efforts to standardize screening of species diversity and to accelerate the process of cryptic species identification. In recent years, DNA barcodes have been obtained for over 6000 species of fish, including 400 species from the New Zealand, 207 Australian commercial marine fish species, 250 species of marine fish from South African waters and 100 species of fish from Pacific Canada (Ward *et al.* 2009). All the COI sequences have been deposited in the Barcode of Life Data Systems (BOLD, http://www.boldsystems.org), and additional fish COI sequences are available in GenBank (Ward *et al.* 2005; Ratnasingham & Hebert 2007). This study provides the first major barcode records for115 commercially important Indian marine fish species belonging to 37 families.

## Materials and methods

### *Sample collections*

One hundred and fifteen species from 37 families were collected during January, 2006–March, 2010 from the East and West Coast of India. Species identification and nomenclature followed the FAO Fish Identification Sheets. Approximately 100 mg of white muscle tissue and fin-clips from two to five individuals of each species were preserved in 95% ethanol until used. Specimen details and GenBank accession numbers are given in Table 1.

### *DNA isolation*

The DNA was isolated following Ruzzante *et al.* (1996) with minor modifications. The concentration of isolated DNA was estimated using a UV spectrophotometer. The DNA was diluted to a final concentration of 100 ng/μL.

### *Amplification and sequencing*

The COI gene was amplified in a 50 μL volume with 5 μL of 10X Taq polymerase buffer, 2 μL of $MgCl_2$ (50 mM), 0.25 μL of each dNTP (0.05 mM), 0.5 μL of each primer (0.01 mM), 0.6 U of *Taq* polymerase and 5 μl of genomic DNA. The primers used for the amplification of the COI gene were FishF1 – 5′TCAACCAACCACAAAGACATT GGCAC3′ and FishR1-5′TAGACTTCTGGGTGGCCAAA GAATCA3′ (Ward *et al.* 2005). The thermal regime consisted of an initial step of 2 min at 95 °C followed by 35 cycles of 40 s at 94 °C, 40 s at 54 °C and 1 min 10 s at 72 °C followed in turn by final extension of 10 min at 72 °C. The PCR products were visualized on 1.2% agarose gels, and the most intense products were selected for sequencing. Products were labelled using the BigDye Terminator V.3.1 Cycle sequencing kit (Applied Biosystems, Inc) and sequenced bidirectionally using an ABI 3730 capillary sequencer following manufacturer's instructions.

### *Sequence analysis*

Sequences were aligned using CLUSTALW (Thompson *et al.* 1997) and submitted to GenBank (Table 1). The extent of sequence difference between species was calculated by averaging pairwise comparisons of sequence difference across all individuals. The COI sequences of the five individuals of each species were aligned to yield a final sequence of 655 bp. Pairwise evolutionary distance among haplotypes was determined by the Kimura 2-Parameter method (Kimura 1980) using the software program MEGA 3.1 (Molecular Evolutionary Genetics Analysis) (Kumar *et al.* 2004). The neighbour-joining (NJ) tree was constructed using MEGA 3.1 and to verify the robustness of the internal nodes of NJ tree, bootstrap analysis was carried out using 1000 pseudoreplications.

## Results

The results are presented for 115 species representing 79 genera, 37 families and 7 orders. The results inferred from nine subgroups are also given separately.

### *General inference*

A total of 397 sequences were generated from 115 species using multiple specimens for all the species. Sequencing of the COI gene produced 655 nucleotide base pairs per taxon. Simplicity and un-ambiguity were observed among all the sequences, and no insertions, deletions or stop codons were observed in any of the sequences. The sequence analysis revealed average nucleotide frequencies as A = 23.50%, T = 29.40%, G = 18.70% and C = 28.40%. The average K2P distances in percentage within different taxonomic levels are given in Table 2. The average transitional pairs (si = 76) were more frequent than average transversional pairs (sv = 47) with an average ratio of 1.33. The average genetic distance within species, genus, family and order was 0.30%, 6.60%, 9.91% and 16.00%, respectively. The summary form of NJ tree is given in Fig. 1.

### *Carangids*

Seventeen fish species of 13 genera belonging to the family Carangidae under the order Perciformes were analysed. The average genetic distance within species was 0.32% whereas the average genetic distance between species was 16.1%. The average nucleotide frequencies were 30.20 (T), 27.60 (C), 23.60 (G) and 18.60 (A) %. The average transitional pairs (si = 64) were more frequent than average transversional pairs (sv = 29) with an average ratio of 2.23. The NJ tree revealed distinct clusters shared by the species of same genera (Fig. 2). All assemblages of conspecific individuals had 94–100% bootstrap values and the congeneric species formed the same clade.

### *Clupeids*

Clupeids group consisting of eleven fish species belonging to two families (Clupeidae and Engraulidae) were examined. Seven genera under this group were used for the generation of barcodes. The overall mean distance among the species was very high (20.30%). The average genetic distance within species was 0.41%. The average nucleotide frequencies were 28.20 (T), 28.50 (C), 20.00 (G) and 23.30 (A) %. The average transitional pairs (si = 69) were more frequent than average transversional pairs (sv = 44) with an average ratio of 1.58. The NJ tree clearly

**Table 1** List of species DNA Barcoded along with Genbank accession numbers

| S No. | Order | Family | Genus | Species | No. of individuals | GenBank accession No |
|---|---|---|---|---|---|---|
| 1 | Perciformes | Carangidae | *Decapterus* | *russeli* | 5 | EF609507–EF609511 |
| 2 | | | *Megalaspis* | *cordyla* | 5 | EF609548–EF609552 |
| 3 | | | *Atropus* | *atropus* | 5 | EF609502–EF609506 |
| 4 | | | *Alepes* | *djedaba* | 5 | EF609497–EF609501 |
| 5 | | | | *kleinii* | 3 | FJ347909–FJ347910, FJ237545 |
| 6 | | | *Parastromateus* | *niger* | 5 | EF609567–EF609571 |
| 7 | | | *Selar* | *crumenophthalmus* | 2 | FJ347941–FJ347942 |
| 8 | | | | *boops* | 5 | FJ347888–FJ347892 |
| 9 | | | *Caranx* | *ignobilis* | 3 | EU014220–EU014221, FJ347936 |
| 10 | | | | *hippos* | 2 | FJ347905–FJ347906 |
| 11 | | | *Carangoides* | *malabaricus* | 5 | FJ347878–FJ347881, FJ347935 |
| 12 | | | | *chrysophrys* | 1 | FJ237546 |
| 13 | | | *Alectis* | *indicus* | 3 | FJ347893–FJ347894, FJ347934 |
| 14 | | | *Gnathanodon* | *speciosus* | 3 | EU148561–EU148563 |
| 15 | | | *Trachinotus* | *blochii* | 4 | EU148557–EU148560 |
| 16 | | | *Seriolina* | *nigrofasciata* | 3 | EU014234–EU014236 |
| 17 | | | *Elagatis* | *bipinnulata* | 5 | EU014211–EU014215 |
| 18 | | Scombridae | *Auxis* | *thazard* | 4 | FJ226525–FJ226528 |
| 19 | | | | *rochei* | 5 | FJ226516–FJ226520 |
| 20 | | | *Rastrelliger* | *kanagurta* | 5 | EF60587–EF609589, FJ237547–FJ237548 |
| 21 | | | *Thunnus* | *albacares* | 4 | EF609627–EF609629, EU392206 |
| 22 | | | | *tonggol* | 4 | FJ226521–FJ226524 |
| 23 | | | *Euthynnus* | *affinis* | 5 | EU148527–EU148531 |
| 24 | | | *Katsuwonus* | *pelamis* | 4 | EU014258–EU014261, |
| 25 | | Serranidae | *Epinephelus* | *fasciatus* | 2 | EU392207–EU392208 |
| 26 | | | | *longispinis* | 2 | EF609521–EF609522 |
| 27 | | | | *diacanthus* | 5 | EF609516–EF609520 |
| 28 | | | | *chlorostigma* | 5 | EU392202–EU392204, EF609514–EF609515 |
| 29 | | | | *morrhua* | 2 | EU392188–EU392189 |
| 30 | | | | *tauvina* | 3 | EU148564–EU148566 |
| 31 | | | | *latifasciatus* | 1 | EU014218 |
| 32 | | Scianidae | *Otolithes* | *cuvieri* | 4 | FJ347924–FJ347927 |
| 33 | | | | *ruber* | 3 | FJ237584–FJ237586 |
| 34 | | | *Johnius* | *borneensis* | 5 | FJ347919–FJ347923 |
| 35 | | | | *dussumieri* | 2 | FJ347915–FJ347916 |
| 36 | | | *Dendrophysa* | *russelii* | 2 | EU148580–EU148581 |
| 37 | | | *Nibea* | *maculata* | 4 | EU014247–EU014250 |
| 38 | | Leiognathidae | *Photopectoralis* | *bindus* | 4 | EF609532–EF609535 |
| 39 | | | *Leiognathus* | *daura* | 4 | EU148519–EU148522 |
| 40 | | | | *equlus* | 4 | EU392205, FJ347946, EF609536–EF609537 |
| 41 | | | *Secutor* | *ruconius* | 4 | FJ347950, EF609612–EF609614 |
| 42 | | | *Gazza* | *minuta* | 3 | EF609612–EF609614 |
| 43 | | Mullidae | *Parupeneus* | *forsskali* | 1 | FJ347965 |
| 44 | | | | *barbarinus* | 2 | EU148576–EU148577 |
| 45 | | | | *pleurostigma* | 1 | FJ237573 |
| 46 | | | *Upeneus* | *vittatus* | 3 | FJ347944–FJ347945, FJ237538 |
| 47 | | | | *sulphureus* | 4 | EF609634–EF609637 |
| 48 | | | *Mulloidichthys* | *auriflamma* | 2 | EU014232–EU014233 |
| 49 | | Polynemidae | *Polydactylus* | *sextarius* | 2 | EU392177–EU392178 |
| 50 | | | *Eleutheronema* | *tetradactylum* | 2 | EF609512–EF609513 |
| 51 | | | *Leptomelanoso ma* | *indicum* | 2 | EF609538–EF609539 |
| 52 | | | *Filimanus* | *heptadactyla* | 4 | EF609523–EF609526 |

**Table 1** Continued

| S No. | Order | Family | Genus | Species | No. of individuals | GenBank accession No |
|---|---|---|---|---|---|---|
| 53 | | Nemipteridae | *Nemipterus* | *japonicaus* | 4 | EF609553–EF609556 |
| 54 | | | | *mesoprion* | 5 | EF609557–EF609561 |
| 55 | | Apogonidae | *Apogon* | *quadrifasciatus* | 5 | EU148585–EU148589 |
| 56 | | | | *norfolcensis* | 5 | FJ237579–FJ237583 |
| 57 | | Chaetodontidae | *Chaetodon* | *trifasciatus* | 2 | FJ237609–FJ237610 |
| 58 | | | | *decussatus* | 5 | FJ237560–FJ237564 |
| 59 | | | | *collare* | 3 | FJ237557–FJ237559 |
| 60 | | | *Heniochus* | *acuminatus* | 3 | EU014237–EU014239 |
| 61 | | Gerreidae | *Pentaprion* | *longimanus* | 4 | EU392179–EU392182 |
| 62 | | | *Thalassoma* | *lunare* | 1 | FJ237565 |
| 63 | | Lethrinidae | *Lethrinus* | *conchyliatus* | 2 | EU148535–EU148536 |
| 64 | | | | *miniatus* | 3 | EU148532–EU148534 |
| 65 | | Lutjanidae | *Lutjanus* | *lutjanus* | 3 | EU148541–EU148543 |
| 66 | | | | *russellii* | 2 | EU148539–EU148540 |
| 67 | | | | *johnii* | 2 | EU148537–EU148538 |
| 68 | | | | *malabaricus* | 5 | EU014227–EU014231 |
| 69 | | Pomacentridae | *Abudefduf* | *vaigiensis* | 3 | FJ237570–FJ237572 |
| 70 | | Sphyraendiae | *Sphyraena* | *jello* | 4 | EF609619–EF609622 |
| 71 | | Terapontidae | *Terapon* | *theraps* | 1 | FJ347958 |
| 72 | | | | *jarbua* | 4 | FJ347885–FJ347887, FJ237549 |
| 73 | | | *Arothron* | *hispidus* | 2 | EU148578–EU148579 |
| 74 | | | | *immaculatus* | 3 | FJ237595–FJ237597 |
| 75 | | Trichiuridae | *Trichiurus* | *lepturus* | 3 | FJ347951–FJ347953 |
| 76 | | | *Lepturacanthus* | *savala* | 4 | EF609540–EF609543 |
| 77 | | Rachycentridae | *Rachycentron* | *canadus* | 5 | EF609582–EF609586 |
| 78 | | Scatophagidae | *Scatophagus* | *argus* | 4 | EF609604–EF609607 |
| 79 | | Priacanthidae | *Priacanthus* | *hamrur* | 4 | EF609574–EF609577 |
| 80 | | Lactariidae | *Lactarius* | *lactarius* | 4 | EF609529–EF609531, FJ347949 |
| 81 | | | | *platypterus* | 2 | EF609527–EF609528 |
| 82 | | Ephippidae | *Ephippus* | *orbis* | 4 | EU014240–EU014243 |
| 83 | | Sparidae | *Accanthopagrus* | *berda* | 3 | EU014244–EU014246 |
| 84 | | | *Argyrops* | *spinifer* | 3 | EU148594–EU148596 |
| 85 | | Ariommatidae | *Ariomma* | *indica* | 5 | EU148514–EU148518 |
| 86 | | Blennidae | *Petroscirtes* | *variabilis* | 5 | EU148523–EU148526, FJ237611 |
| 87 | | Pempheridae | *Pempheris* | *adusta* | 5 | EU148571–EU148575 |
| 88 | | Centrolophidae | *Psenopsis* | *cyanea* | 3 | EU392194–EU392196 |
| 89 | | Menidae | *Mene* | *maculata* | 4 | FJ347937–FJ347940 |
| 90 | Clupeiformes | Clupeidae | *Dussumieria* | *elopsoides* | 5 | FJ347959–FJ347963 |
| 91 | | | | *acuta* | 5 | EU014222–EU014226 |
| 92 | | | *Tenualosa* | *toli* | 4 | EF609623–EF609626 |
| 93 | | | *Hilsa* | *kelee* | 4 | FJ158558–FJ158561 |
| 94 | | | *Sardinella* | *gibbosa* | 2 | FJ237612–FJ237613 |
| 95 | | | | *albella* | 5 | FJ237536–FJ237537, FJ237550–FJ237552 |
| 96 | | | | *longiceps* | 5 | EF609594–EF609598 |
| 97 | | Engraulidae | *Stolephorus* | *indicus* | 2 | FJ347956–FJ347957 |
| 98 | | | *Encrasicholina* | *heteroloba* | 5 | EU392183–EU392187 |
| 99 | | | *Thryssa* | *malabarica* | 4 | FJ347943, FJ347882–FJ347884 |
| 100 | | | | *hamiltonii* | 4 | EU148567–EU148570 |
| 101 | Mugiliformes | Mugilidae | *Liza* | *macrolepis* | 5 | FJ347967, EF609544–EF609547 |
| 102 | Siluriformes | Ariidae | *Osteogeneiosus* | *militaris* | 5 | EF609562–EF609566 |
| 103 | | | *Netuma* | *thalassinus* | 5 | EU014251–EU014255 |
| 104 | | | *Arius* | *subroastratus* | 2 | EU148555–EU148556 |
| 105 | | | | *arius* | 5 | EU148548–EU148552 |
| 106 | Pleuronectiformes | Cynoglsidae | *Cynoglossus* | *macrostomus* | 4 | FJ347954–FJ347955, FJ347911–FJ347912 |

**Table 1** Continued

| S No. | Order | Family | Genus | Species | No. of individuals | GenBank accession No |
|---|---|---|---|---|---|---|
| 107 | | | | *dubius* | 2 | FJ347907–FJ347908 |
| 108 | Beloniformes | Hemiramphidae | *Hemiramphus* | *far* | 2 | EU148546–EU148547 |
| 109 | | | *Hyporhamphus* | *xanthopterus* | 4 | EU148544–EU148545, FJ237601–FJ237602 |
| 110 | | Belonidae | *Strongylura* | *strongylura* | 2 | EU014256–EU014257 |
| 111 | | | | *leiura* | 1 | FJ237566 |
| 112 | Aulopiformes | Synodontidae | *Trachinocephalus* | *myops* | 4 | EF609630–EF609633 |
| 113 | | | *Saurida* | *tumbil* | 5 | EF609599–EF609603 |
| 114 | | | | *undosquamis* | 3 | FJ347930–FJ347932 |
| 115 | | | *Harpadon* | *nehereus* | 3 | EU148582–EU148584 |

**Table 2** Summary of genetic divergences (K2P percentage) within various taxonomic levels

| Comparisons within | Minimum | Maximum | Average | Standard error |
|---|---|---|---|---|
| Species | 0.00 | 00.80 | 00.30 | 0.021 |
| Genera | 0.10 | 12.90 | 06.60 | 0.085 |
| Families | 0.20 | 23.10 | 09.91 | 0.032 |
| Orders | 8.00 | 23.40 | 16.00 | 0.018 |

distinguished all the species. The species belonging to family Clupeidae and Engraulidae were represented by two distinct clades with a boostrap value of 98% (Fig. 3).

### *Scombrids*

The scombrids represented by six genera under the family Scombridae were studied. The average genetic distance within species showed a lower value of 0.3%. The overall mean distance among the species was 9.20%. The average nucleotide composition was T = 29.30, C = 28.60, G = 18.90 and A = 23.20%. The average transitional pairs (si = 38) were more frequent than average transversional pairs (sv = 17) with an average ratio of 2.22. All the species under the six genera were clearly separated by different clusters in the NJ tree with a bootstrap value ranging from 96 to 100% (Fig. 4).

### *Groupers*

Seven species under the genus *Epinephelus* belonging to family Serranidae were investigated in the study. The overall mean distance among the species showed a low value of 12.60%. The average genetic distance within species was very low (0.24%). The sequence analysis revealed nucleotide frequencies as T = 29.40, C = 28.30, G = 18.30 and A = 24.00%. The average transitional pairs (si = 56) were more frequent than average transversional pairs (sv = 18) with an average ratio of 3.10. No individuals were misplaced in the NJ tree and differentiated with a bootstrap value of 94–98% (Fig. 5).

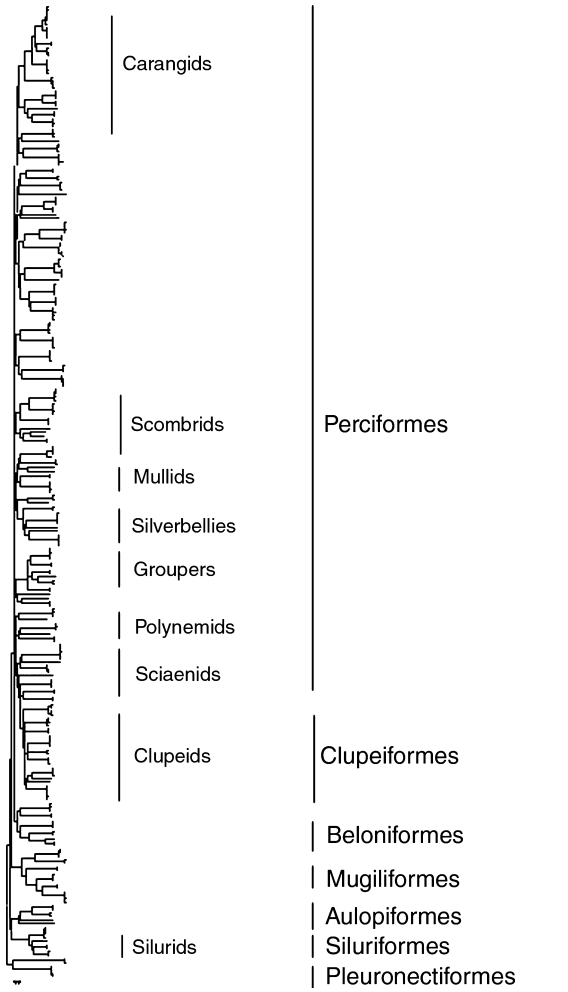



**Fig. 1** Summary form of Neighbour Joining tree of c oxidase I gene sequences derived from 115 fish species using K2P distances.

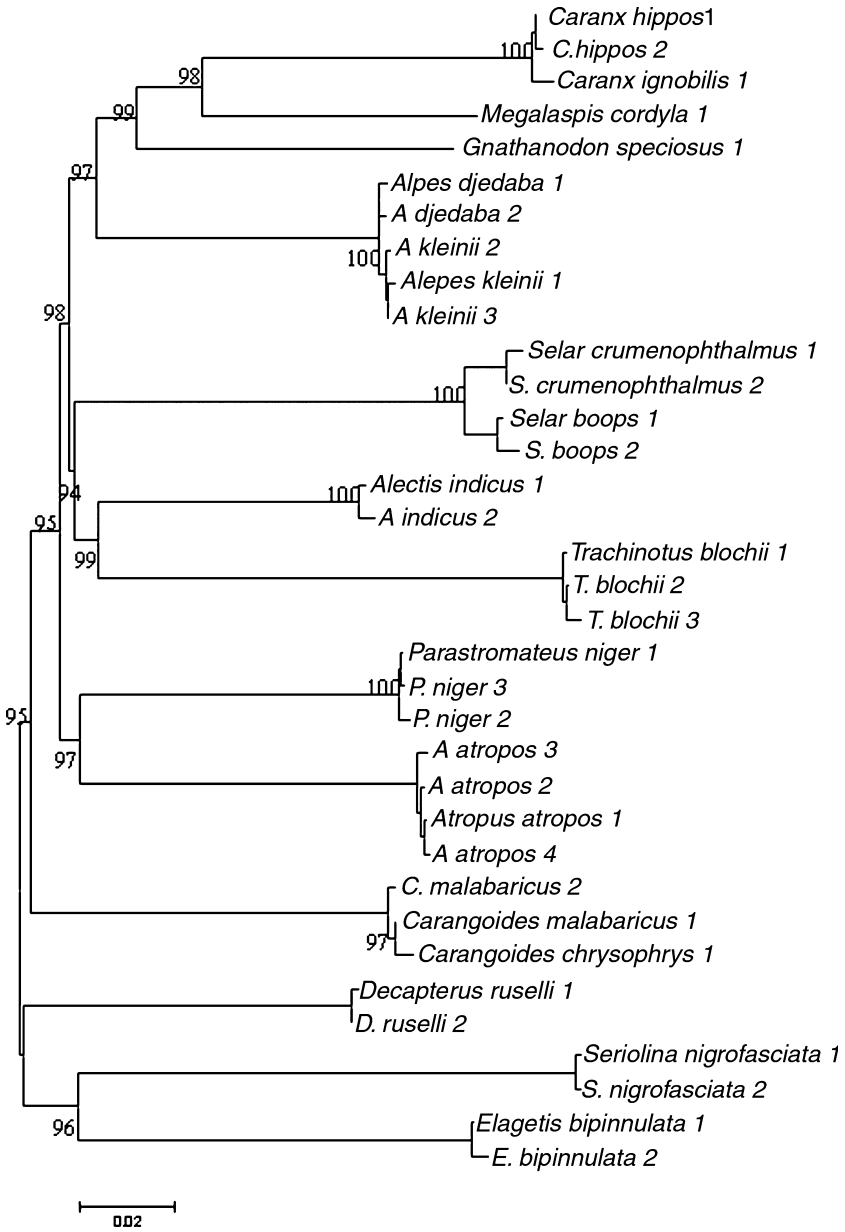



**Fig. 2** Neighbour Joining tree of c oxidase I gene sequences derived from Carangids using K2P distances.

*Sciaenids*

Sciaenids represented by four genera belonging to family Sciaenidae were analysed using six species. The average genetic distance within species was 0.28% whereas the overall mean distance among the species was 18.20%. The sequence analysis revealed nucleotide frequencies as T = 29.90, C = 28.30, G = 18.80 and A = 23.00%. The average transitional pairs (si = 69) were more frequent than average transversional pairs (sv = 32) with an average ratio of 2.12. The NJ tree clearly distinguished the species having same genus under one cluster with a bootstrap value of 96–100% (Fig. 6).

*Silverbellies*

Fifteen DNA barcodes were generated from four species of the genera *Photopectoralis*, *Leiognathus*, *Secutor* and *Gazza*. The average genetic distance within species was 0.20%. The overall mean distance among the species was 16.60%. The sequence analysis revealed nucleotide frequencies as T = 29.50, C = 28.00, G = 17.50 and A = 25.00%. The average transitional pairs (si = 59) were more frequent than average transversional pairs (sv = 34) with an average ratio of 1.74. The NJ tree clearly differentiated the species of the four genera into distinct clusters with a bootstrap value of 97–100% (Fig. 7).

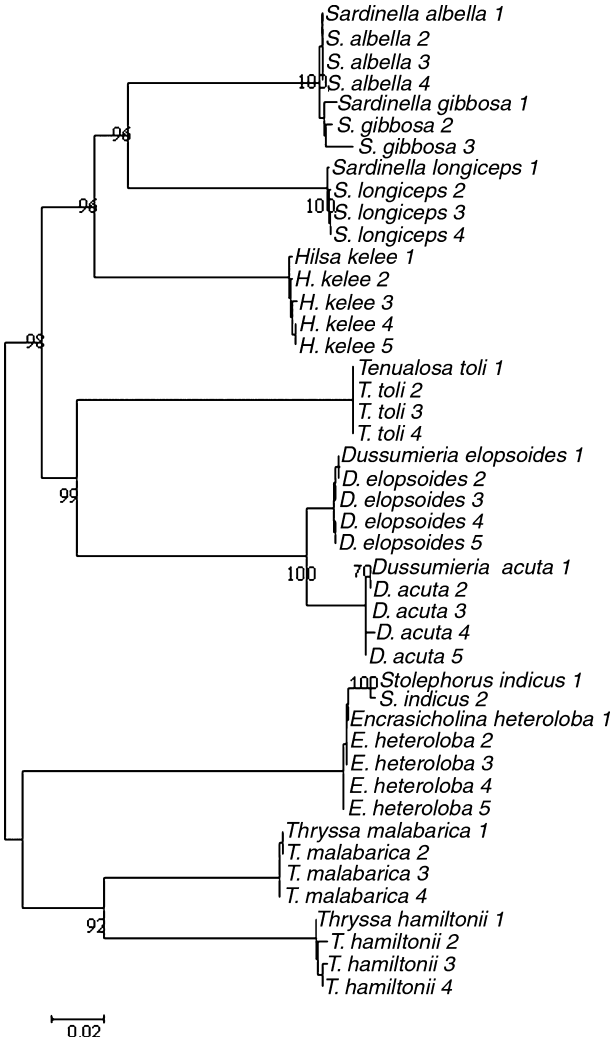



**Fig. 3** Neighbour Joining tree of c oxidase I gene sequences derived from Clupeids using K2P distances.

### *Mullids*

Six fish species commonly called goatfish belonging to Mullidae were characterized in the study. The average genetic distance within species was 0.38% whereas the overall mean distance among the species was 13.90%. The sequence analysis revealed nucleotide frequencies as T = 29.20, C = 29.10, G = 19.10 and A = 22.60%. The average transitional pairs (si = 55) were more frequent than average transversional pairs (sv = 25) with an average ratio of 2.20. The NJ tree revealed that the *Genera Parupeneus, Mulloidichthys* and *Upeneus* formed three separate clusters with a boostrap value of 95–99% (Fig. 8).

### *Polynemids*

Six Polynemids belonging to four genera (*Polydactylus*, *Eleutheronema*, *Leptomelanosoma* and *Filimanus*) were studied. The average K2P distance within species was 0.35%. The mean interspecies distance within the family was 16.30%. The nucleotide composition was estimated as T = 28.90, C = 30.30, G = 18.70 and A = 22.10%. The average transitional pairs (si = 68) were more frequent than average transversional pairs (sv = 23) with an average ratio of 2.90. The NJ tree revealed that three clusters were formed. The first and second cluster were shared by the species of *Genus Polydactylus* and Filimanus, respectively. The third cluster was formed by *Leptomelanosoma* and *Eleutheronema*. The clusters were formed with a bootstrap value ranging from 92–100% (Fig. 9).

### *Silurids*

The catfishes of three genera namely *Osteogeneiosus*, *Netuma* and *Arius* under the family Ariidae were characterized for DNA barcodes. The average K2P distance within species was 0.23%. The mean interspecies distance within the family was very low (8.10%). The sequence analysis revealed nucleotide frequencies as T = 29.20, C = 28.90, G = 17.30 and A = 24.60%. The average ratio (2.15%) of transitional pairs (si = 43) and transversional pairs (sv = 20) was very high in this group. Two clusters were formed in the NJ tree. The first cluster was shared by *Arius subrostratus* and *A. arius*. The second cluster was shared by *Netuma thalassinus* and *Osteogeneiosus militaris*. The clusters were formed with a bootstrap value ranging from 90 to 99% (Fig. 10).

## Discussion

In this study, 115 species representing 7 orders (Perciformes, Clupeiformes, Mugiliformes, Siluriformes, Pleuronectiformes, Beloniformes and Aulopiformes) and 37 families including Carangids, Clupeids, Scombrids, Groupers, Sciaenids, Silverbellies, Mullids, Polynemids and Silurids of Indian marine fishes were characterized for generation of DNA barcodes. The universal primers amplified the target region in all 115 species, generating 397 COI barcodes of 655 bp. No insertions, deletions or stop codons were observed in any of the sequences, supporting the hypothesis that all the amplified sequences derive from a functional mitochondrial COI sequences. The lack of stop codons together with 655 bp length of amplified sequences suggests that NUMTs (Nuclear Mitochondrial DNA: nuclear DNA sequences originating from mitochondrial DNA sequences) were not sequenced, a result in conformity with previous reports (Ward *et al.* 2005). A review of the occurrence of NUMTs in plants and animals did not find any evidence of their existence in Actinopterygii (Bensasson *et al.* 2001). A latter report (Richly & Leister 2004) suggested their presence in *Fugu rupripes*, but this was subsequently

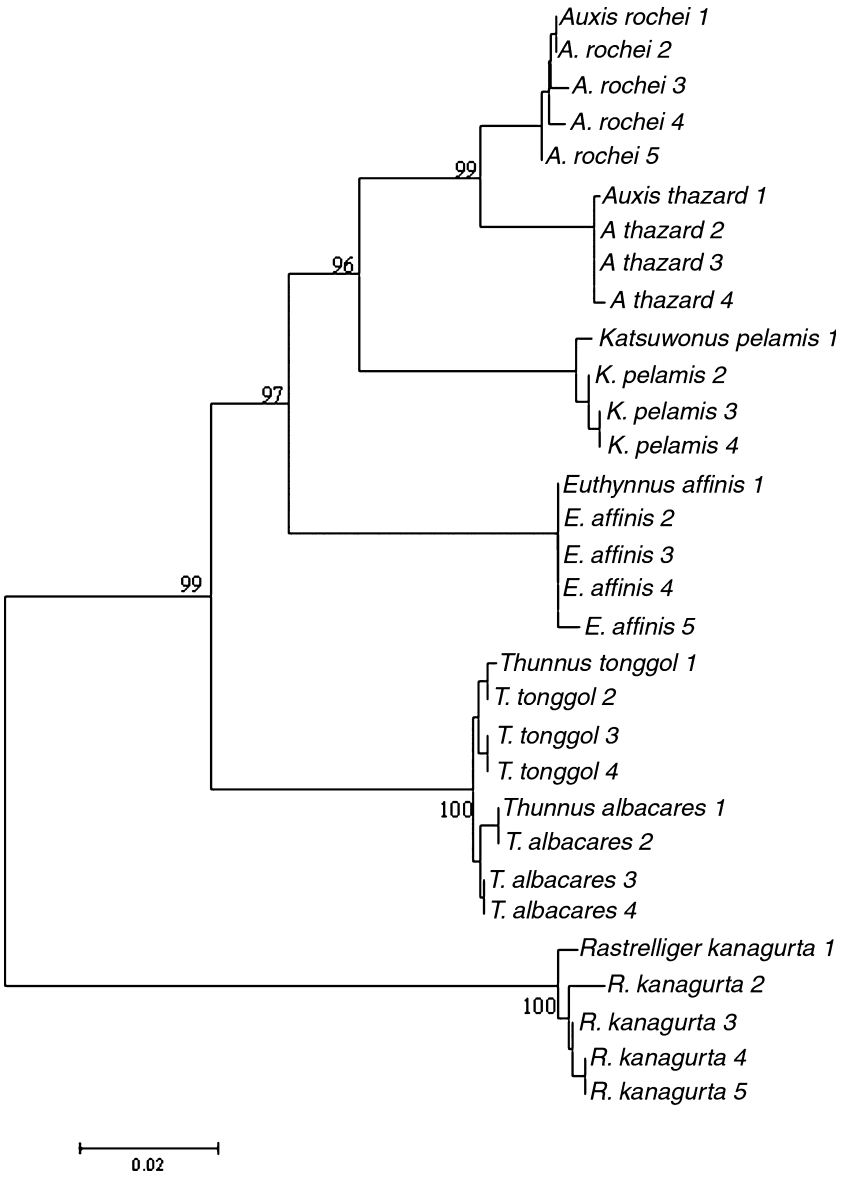



**Fig. 4** Neighbour Joining tree of c oxidase I gene sequences derived from Scombrids using K2P distances.

shown to reflect an error in data interpretation (Ward *et al*. 2009).

The barcode sequences clearly discriminated taxonomic status of all 115 species examined. The mean nucleotide diversity (Pi) among all the species was estimated as 0.2029. It has been shown that lineages diversify more quickly within species than between species (Pons *et al*. 2006). The branch length between species tends to be much deeper than between conspecific individuals leading to a gap in the distribution of the pairwise distance between conspecific individuals and between species that has been referred to the barcoding gap (Meyer & Paulay 2005). The COI locus harbours a high mutational rate even for mtDNA (Saccone *et al*. 1999). This study reveals that the mean genetic distance between conspecific individuals is much smaller than the average distance between individuals of different species. Although barcode analyses primarily seek to delineate species boundaries at the COI locus for the assignment of unknown individuals to known species, unsuspected diversity and overlooked species are often detected through barcodes analyses, sometimes spectacularly (Meyer & Paulay 2005; Kerr *et al*. 2007). In this study, the average K2P distance of individuals within species was estimated as 0.30% whereas it was 6.60% for the species within genera. Hence, there was a 22-fold more sequence difference among congeneric species than conspecific individuals. The variation was more among the congeneric individuals than among the

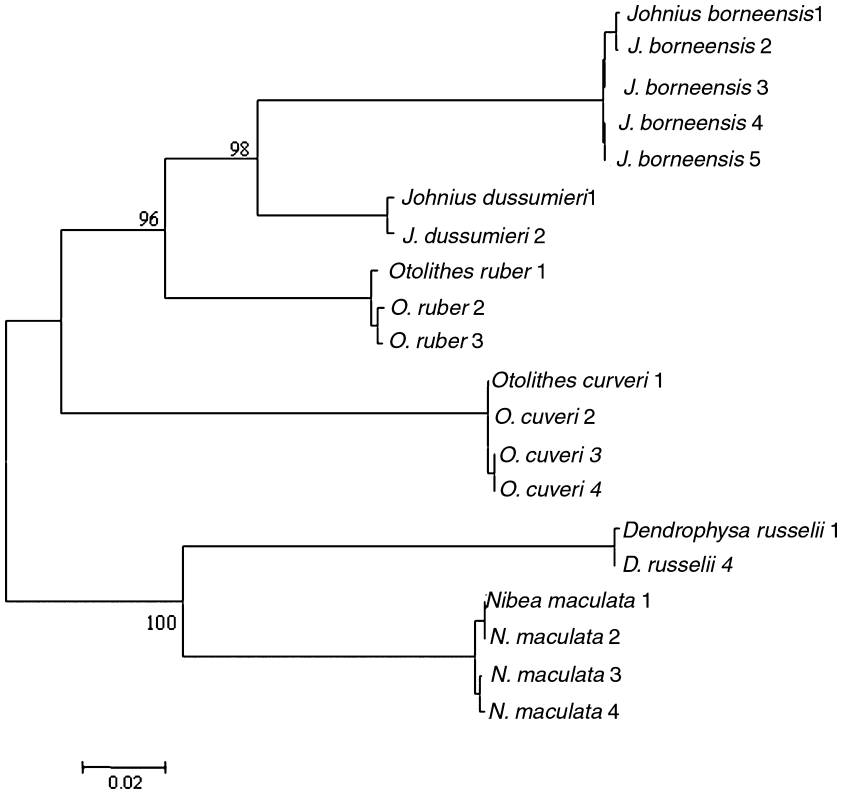



**Fig. 5** Neighbour Joining tree of c oxidase I gene sequences derived from Groupers using K2P distances.

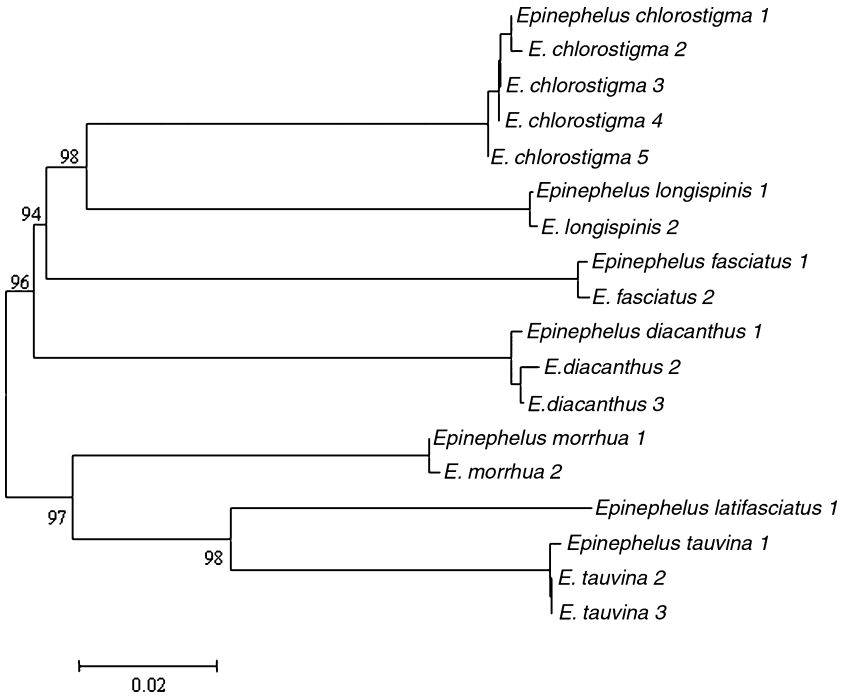
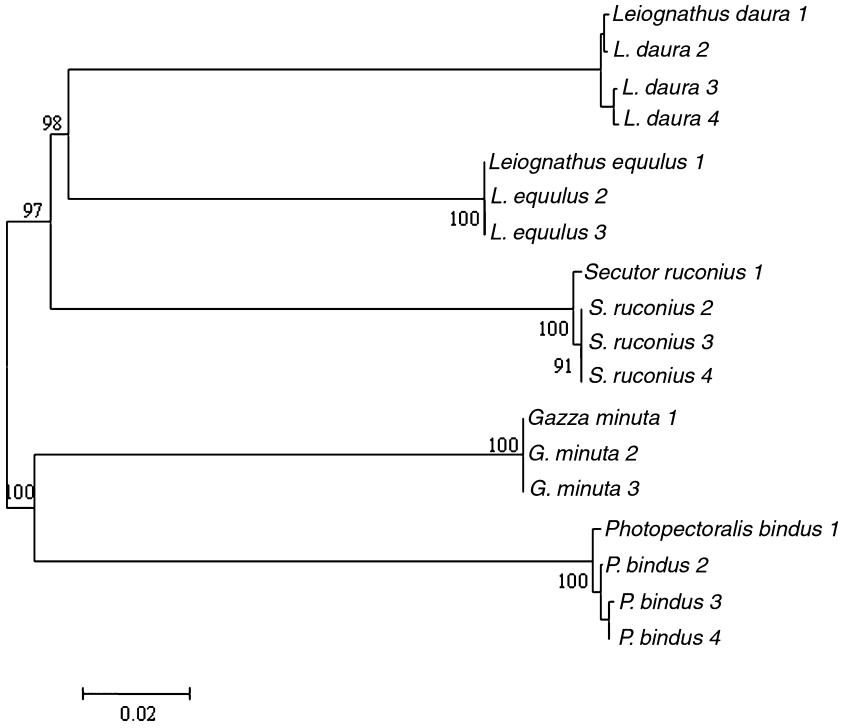



**Fig. 6** Neighbour Joining tree of c oxidase I gene sequences derived from Sciaenids using K2P distances.

conspecific individuals. Mean divergence among species within families increases to 15.5%, and among species within orders and classes it increases to 22.2% and 23.35%, respectively (Ward *et al.* 2005; Spies *et al.* 2006). We found 9.91% average distance among species within families whereas it was 16.00% among species within the



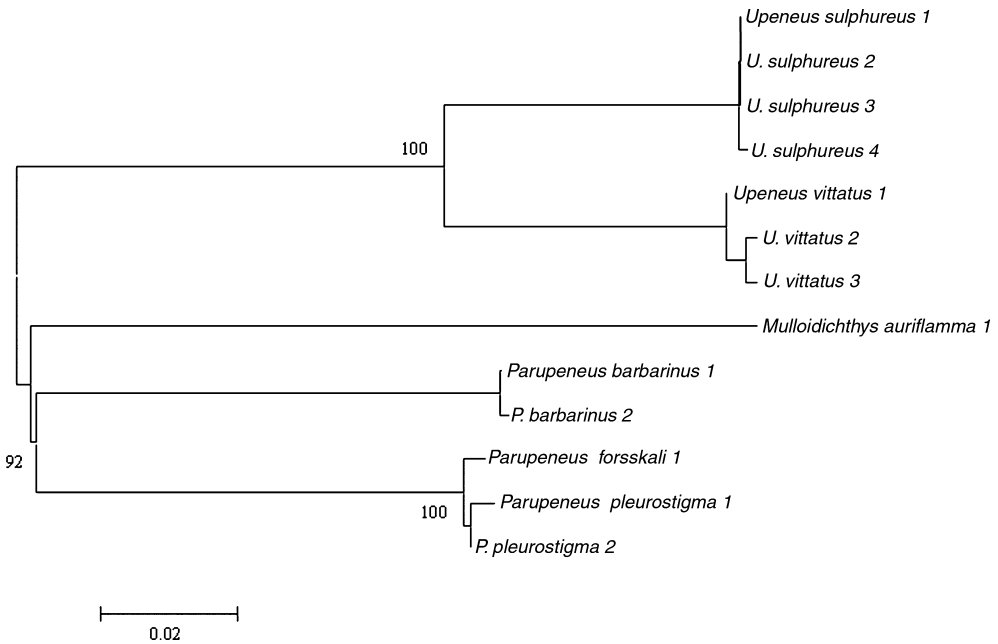




**Fig. 7** Neighbour Joining tree of c oxidase I gene sequences derived from Silverbellies using K2P distances.

**Fig. 8** Neighbour Joining tree of c oxidase I gene sequences derived from Mullids using K2P distances.

order. A steady increase of genetic variation through the increment of taxonomic levels was observed, supporting a marked change of genetic divergence at the species boundaries. This finding supports the previous observations (Hubert *et al*. 2008).

The average transition and transversion ratio was 1.33, while the average GC content was 47.10%, similar to results obtained by Ward *et al*. (2005). The highest GC content (51.20%) was found in the Carangidae while the lowest (44.7%) was observed in the Leognathidae. Saccone *et al*. (1999) reviewed data from the complete mitochondrial genomes of nine Osteichthyes and three Chondrichthyes species, deriving GC contents of 43.2% and 38.4%, respectively. These values

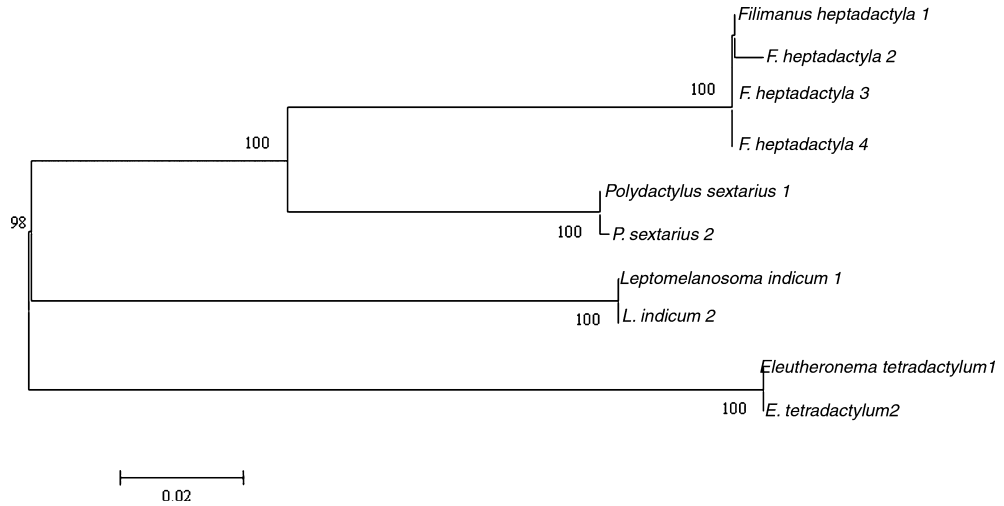



**Fig. 9** Neighbour Joining tree of c oxidase I gene sequences derived from Polynemids species using K2P distances.

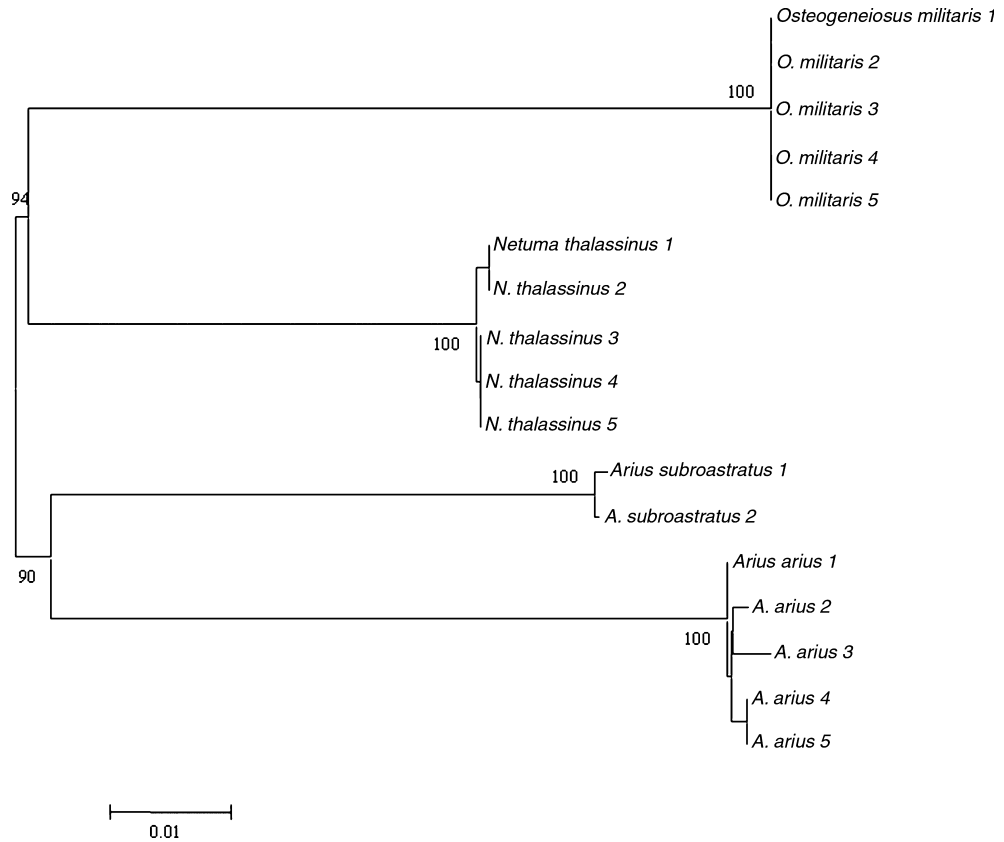



**Fig. 10** Neighbour Joining tree of c oxidase I gene sequences derived from Silurids using K2P distances.

correspond reasonably well to ours especially with respect to the higher GC content of the teleosts. As usual, most nucleotide changes took place at the 3rd codon position than the 1st, and more at the 1st than the 2nd.

The NJ tree revealed identical phylogenetic relationship among the species. The phylogenetic relationship among the species was clearly established, and similar species were clustered under same nodes while dissimilar species were clustered under separate nodes. The

nodes were supported by high bootstrap values (90–100%). Although barcode analysis seeks only to delineate species boundaries, there is clearly some phylogenetic signal in COI sequence data. Congeneric species always clustered together and in most cases so did the confamilial species.

Ward *et al*. (2008) made an interesting revelation in identifying a second species of Asian sea bass (*Lates calcerifer*) based on COI sequence divergences. In addition to the species identification, DNA barcoding has been used for identification of processed fish products (Smith *et al*. 2008). In conclusion, the results from our data are congruent with the taxonomic divisions of the finfish under study, based on morphological characters as reported in FAO identification sheets. This study has strongly authenticated the efficacy of COI in identifying the fish species with designated barcodes. DNA sequences within species need to be similar to one another than to sequences in different species for making DNA barcoding approach successful. Our results suggest that COI barcoding can be taken up as pragmatic approach for resolving unambiguous identification of the fish fauna of Indian Ocean with applications in its management and conservation.

## Acknowledgements

The authors are thankful to Dr S. Ayyappan, Director General, Indian Council of Agricultural Research for the guidance and financial support.

## References

Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology and Evolution*, **16**, 314–321.

FAO (2002) *The state of world fisheries and aquaculture, part 1: world review of fisheries and aquaculture*. Food and Agricultural Organization, Fisheries Department, Rome.

FAO (2004) *Capture production 2002. FAO yearbook of fishery statistics 94/1*. Food and Agricultural Organization, Fisheries Department, Rome.

Froese R, Pauly D (2009) FishBase. World Wide Web electronic publication. http://www.fishbase.org, version (02/2009).

Hajibabaei M, de Waard JR, Ivanova N *et al*. (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B*, **360**, 1959–1967.

Hebert PDN, Cywinska A, Ball SL, Ward JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Series B*, **270**, 313–322.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences USA*, **101**, 14812–14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. *Public Library of Science*, **2**, 1657–1663.

Hogg ID, Hebert PDN (2004) Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology*, **82**, 749–754.

Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E (2008) Identifying Canadian freshwater fishes through DNA barcodes. *Public Library of Science*, **3**, e2490. doi:10.1371/journal.pone.0002490.

Kerr KC, Stoeckle MY, Dove CJ, Weigt LA, Francis CM (2007) Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, **7**, 535–543.

Kimura M (1980) A simple method of estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Lakra WS, Goswami M, Gopalakrishnan A (2009) Molecular identification and phylogenetic relationships of seven Indian Sciaenids (Pisces: perciformes, Sciaenidae) based on 16S rRNA and cytochrome oxidase subunit I mitochondrial genes. *Molecular Biology Reports*, **36**, 831–839.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLOS Biology*, **3**, 2229–2238.

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *Public Library of Science*, **2**, e354.

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–606.

Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Richly E, Leister D (2004) NUMTs in sequenced eukaryotic genomes. *Molecular Biology Evolution*, **21**, 1081–1084.

Ruzzante DE, Taggart CT, Cook C, Goddard S (1996) Genetic differentiation between inshore and offshore Atlantic cod *Gadus morhua*) off Newfoundland: microsatellite DNA variation and antifreeze level. *Canadian Journal of Fisheries and Aquatic Science*, **53**, 634–645.

Saccone C, De GiorgiC, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene*, **238**, 195–209.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopedia of life:an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B*, **360**, 1805–1811.

Smith PJ, McVeagh SM, Steinke D (2008) DNA barcoding for the identification of smoked fish products. *Journal of Fish Biology*, **72**, 464.

Spies IB, Gaichas S, Stevenson DE, Orr JW, Canino MF (2006) DNA based identification of Alaska skates (*Amblyraja, Bathyraja* and *Raja*: Rajidae) using cytochrome c oxidase submit (COI) variation. *Journal of Fish Biology*, **69**(B), 283–292.

Steinke D, Vences M, Salzburger W, Meyer A (2005) TaxI: a software tool for DNA barcoding using distance methods. *Philosophical Transactions of the Royal Society, London, Series B*, **360**, 1847–1857.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Ward RD, Zemlac TC, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B*, **360**, 1847–1857.

Ward RD, Holmes BH, Yearsley GK (2008) DNA barcoding reveals a likely second species of Asian sea bass (barramundi) (*Lates calcarifer*). *Journal of Fish Biology*, **72**, 458–463.

Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-VBOL. *Journal of Fish Biology*, **74**, 329–356.