



Genetic diversity of *Hilsa kelee* collected from the Bay of Bengal and the Arabian Sea

Anirban Sarker^{1,2} · Habibon Naher^{1,2} · Junman Huang¹ · Kishor Kumar Sarker¹ · Mohammad Abdul Baki² · Chenhong Li¹

Received: 6 February 2020 / Revised: 6 August 2020 / Accepted: 4 September 2020
© Senckenberg Gesellschaft für Naturforschung 2020

Abstract

The Kelee shad (*Hilsa kelee*) is currently considered as the only representative species in the genus *Hilsa* (Clupeidae, Clupeiformes) with a wide distribution in almost all coasts of the Indian Ocean. Two marginal marine water bodies of the Indian Ocean, the Bay of Bengal and the Arabian Sea, are important fisheries areas for this fish. In this study, a comparison of genetic diversity of Kelee shad of these two important fisheries areas was conducted by applying a cross-species gene enrichment method. This was also the first challenge to discover the genetic diversity of this fish. Analysis of molecular variance results showed that the variation between groups is 10.85% and among individuals within populations contributed 89% variation. Nucleotide diversity results showed that Kelee shad populations are in almost stable condition and have a moderate population size. Nucleotide diversity of the Bay of Bengal and the Arabian Sea were almost similar; these were 0.012138 and 0.010831, respectively. Network, structure, and PCA results suggested that Kelee shad of the Bay of Bengal and the Arabian Sea have significant genetic differences. These two populations were separated from each other by a F_{ST} value of 0.10955 ($p < 0.05$). It is speculated that Southern India along with Sri Lanka and Ram bridge together with ancient climatic differences acted as means of isolation concerning Kelee shad populations between these two water bodies. Furthermore, from the maximum likelihood phylogenetic tree, it is established that the Arabian Sea can be regarded as the major stock of this species.

Keywords Clupeidae · Kelee Shad · Gene enrichment method · Indian Ocean · Ram Bridge

Introduction

The Kelee shad (*Hilsa kelee*) is currently considered as the single species of the genus *Hilsa* (Family: Clupeidae, Order: Clupeiformes). This fish is distributed in almost all coasts of the Indian Ocean, from Durban, South Africa, and the Madagascar Island to north of Hong Kong and east of Papua New Guinea across the Maputo Bay, the Gulf of Oman, the Arabian Sea, the

Bay of Bengal, the Gulf of Thailand, and the Java Sea (Gjøsaeter and Sousa 1983; Rahman et al. 2009; Dwivedi 2019). Generally, this pelagic fish species prefers coastal waters, but occasionally enters into estuaries down to low salinities of 7 ppt (Whitehead 1985; Huda and Haque 2003).

Other shad species are of high commercial importance, but Kelee shad also has significant economic importance. This fish is eaten fresh, salted, and dried in the areas of its distribution (Rahman et al. 2009). According to FAO landing statistics (1992–1998), total catch from the Bay of Bengal and Indian Ocean was 43,000 metric tons per year and from continental waters of the Indian subcontinent 46,000 metric tons per year. There is no statistics of fisheries from Africa available, but it is a very important fish in the eastern coast of Africa, especially Maputo Bay, Mozambique, with a total estimated catch of 4600 metric tons per year (Charles-Dominique and Albaret 2003). This fish is regarded as an exportable fish species at Oman and Myanmar. In Bangladesh, Kelee shad is used as alternative food fish of the Hilsa Shad because of its cheaper price to poor and middle class people and regarded as a good source of protein.

Communicated by R. Thiel

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12526-020-01114-3>) contains supplementary material, which is available to authorized users.

✉ Chenhong Li
chli@shou.edu.cn

¹ Shanghai Universities Key Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Ocean University, Shanghai 201306, China

² Department of Zoology, Jagannath University, Dhaka 1100, Bangladesh

Some other morphologically similar tropical shads (i.e., *Tenualosa ilisha*, *Tenualosa toli*, and *Tenualosa macrura*) are culturally important in Kelee shad's distributed areas. As per example, *T. ilisha* is treated as an important object in Bengali New Year celebration in Bangladesh and some parts of India. *Tenualosa toli* and *T. macrura* are also closely linked to the life and culture of some parts of Indonesia, Malaysia, and Singapore. Bengkalis of Indonesia is named as "Trubuk city" (Terubuk is the local name of *T. toli* and *T. macrura*). Despite this, *H. keele* attained a very poor research focus except few articles focused on their identification, distribution, biology, and fisheries (Rao 1973; Gjøsæter and Sousa 1983; Panhwar et al. 2012; Cardinale et al. 2014; Dwivedi 2019). On the other hand, all *Tenualosa* species except *T. thibaudeaui* (Mekong Herring) received an immense research focus (Jia 1982; Blaber et al. 1996, 2005; Wang 1996; Brewer et al. 2001; Puvanasundram et al. 2017; Zain et al. 2018). The field of genetic diversity and population status of Kelee shad have not been investigated yet. Discovery of genetic diversity is very important for sustainable ocean fisheries, because it can provide basic information for the fisheries resource assessment and management (Yang et al. 2020). Due to environmental changes, pollution and over exploitation, many fishes are losing genetic diversity currently (Madduppa et al. 2018). However, Kelee shad's genetic status is still unrevealed; thus, it is very difficult to clarify the fisheries management and conservation policy of this fish.

The Arabian Sea and the Bay of Bengal are similar in many aspects, as such as location on the same latitude bands, semi-enclosed nature, and opening on the southern side. However, because of differences in summer monsoon, the precipitation exceeds evaporation in the Bay of Bengal, whereas due to absence of summer monsoon, opposite phenomenon is visualized in the Arabian Sea. The run-off is more, and the annual rainfall is extra 1–2 m in the Bay of Bengal than that of the Arabian Sea. As a result, surface water salinity of the Bay of Bengal (30–33‰) is lesser than that of the Arabian Sea (34 to 37‰). Furthermore, huge run-off leads to strong turbid water and sediment load in the Bay of Bengal especially during monsoon season. Surface temperature fluctuation is also lesser in the Bay of Bengal (27–29 °C) than that of the Arabian Sea (23–29 °C) (Panikkar et al. 1966; Jensen 2001). However, the difference among oceanographic parameters between these two marine water bodies indicated that fish habitats are significantly different from each other. The habitat differences as well as several thousand years of separation may have led to variations between the fish populations of these two water bodies similar to that found by Behera et al. (2015) for the genetic differences of *T. ilisha* populations of both areas when studying mitochondrial DNA Cytochrome b gene sequence.

In this study, we collected sequence data of 4434 nuclear genes from 20 Kelee samples taken from the Bay of Bengal and the Arabian Sea, applying a cross-species gene

enrichment method (Li et al. 2013), to examine the genetic diversity of the Kelee shad of these two mentioned regions. Our goal is to provide a solid estimation of the genetic diversity of the Kelee shad of this two water bodies using genome-wide data. Our study shall provide a first look into the genetic diversity of this commercially important species in these two important fisheries areas and may be helpful in future conservation and management measures of this species.

Materials and methods

Sample collection and DNA extraction

Fin clips or muscle tissues of the Kelee shad were collected from two distinctly different habitats, the Bay of Bengal and the Arabian Sea (Fig. 1; Table 1). The samples were identified based on morphological features according to the description of Rahman et al. (2009) and Rahman et al. (2018). Samples of the Bay of Bengal were collected from Cox's Bazar Fish landing Station (CB), Adinath Bazar Fish Market, Maheshkhali (MK), and Dhal Char (DC) in Bangladesh. Samples of Arabian Sea (SA) were collected from the Salalah Fish landing Station in Oman. As all of the samples were collected from fish landing stations, which suggests random sampling of fish capture representing the fish population of specific water bodies (i.e., Bay of Bengal & Arabian Sea). The closely related *T. ilisha* was used as outgroup, collected from Meghna River, Chandpur point, Bangladesh. All samples and specimens had been deposited at museum of Shanghai Ocean University. Tissue samples were collected from the fish's dorsal portion or caudal fin. After that, tissue samples were fixed in 100% ethanol then preserved in 95% ethanol for long duration. Finally, samples were stored in 4 °C refrigerator until DNA extraction. DNA was extracted using an Ezup DNA extraction kit following the protocol of the manufacturer (Sangon Biotech, Shanghai, China) from 25 mg of tissue.

Cytochrome c oxidase subunit I (COI) gene analysis

Cytochrome C oxidase subunit 1 (COI) gene was amplified using common COI gene primers designed for the Clupeiformes by the author's lab (clupeiform_COI_F GGAGCTACAATCCGCCGCCTA and clupeiform_COI_R GGTTCGATTCTTCCTTTCTCGTT). Sequences were determined at Sangon Biotech, Shanghai, China. Chromatographs were used to check and edit sequences, and then assembled using Codon Code Aligner software (Codon Code Corporation). The COI gene sequence was compared with sequences lodged in NCBI for species identification using BLAST.

Table 1 Sampling locations with number of samples for each site and sampling dates

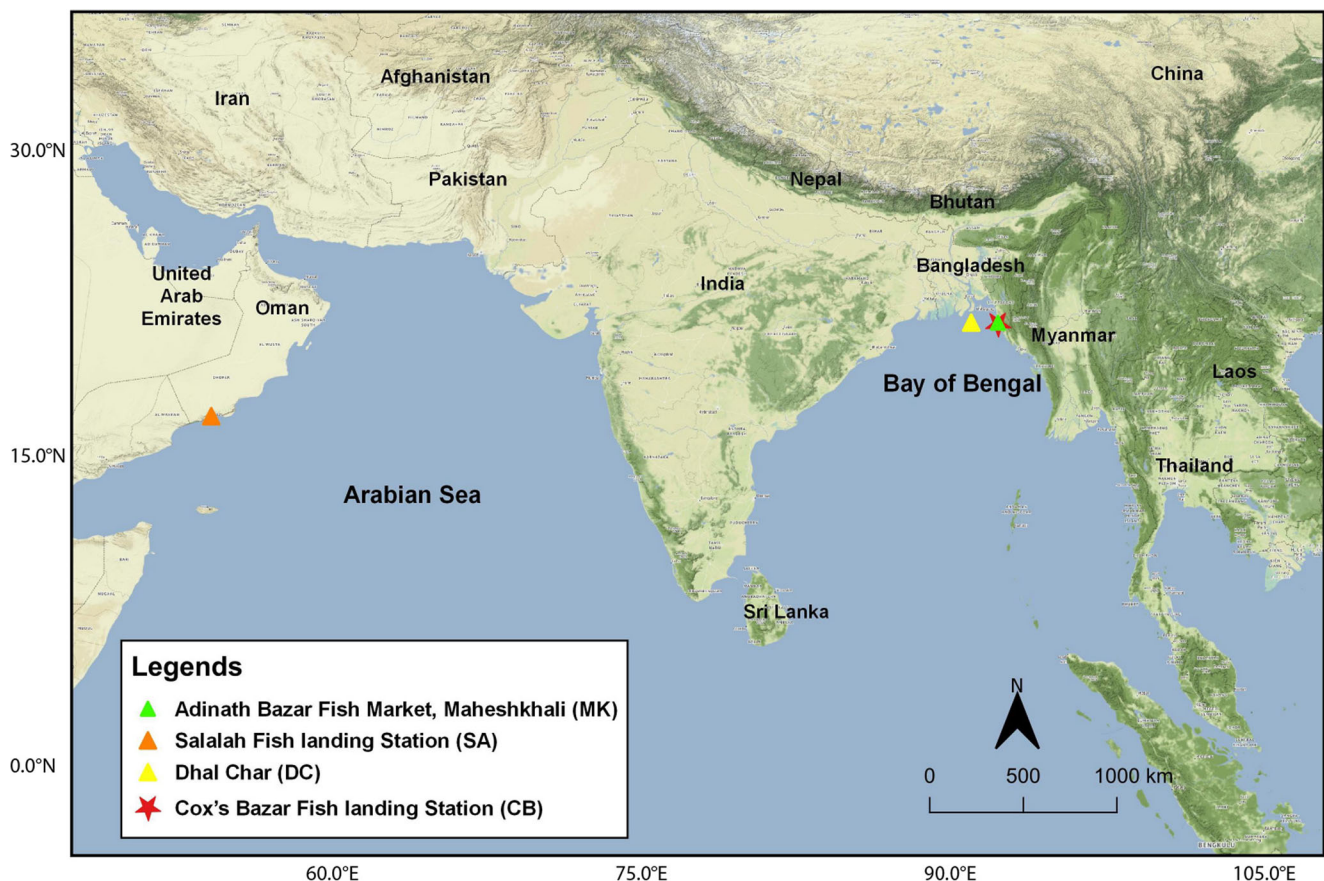
Water body	Voucher number	Sample ID	Locality	No. of samples	Latitude	Longitude	Date of collection
Bay of Bengal	SOU1801029-1 to 7	CL_2050	Cox's Bazar Fish landing Station	07	21° 27' 05.1" N	91° 58' 07.4" E	Sep 6, 2018
Bay of Bengal	SOU1801030-1 to 3	CL_2048	Adinath Bazar Fish Market, Maheshkhali	03	21° 31' 38.4" N	91° 57' 56.2" E	Sep 5, 2018
Bay of Bengal	SOU1801031-1 to 2	DC_CL_1362	Dhal Char	02	21° 31' 50.8" N	90° 40' 41.8" E	Feb 16, 2017
Arabian Sea	SOU1801032-1 to 8	CL_2312	Salalah Fish landing Station	08	17° 01' 07.2" N	54° 05' 58.4" E	May 20, 2019
Outgroup <i>Tenualosa ilisha</i>	SOU1801033-1 to 2	CP_CL_2026	Chandpur	02	23° 11' 42.7" N	90° 37' 55.5" E	Sep 2, 2018

DNA library preparation, gene capture, and sequencing

Extracted genomic DNA was sheared to about 500 bp using a Covaris M220 Focused-ultrasonicator (Woburn, Massachusetts, USA) according to the manufacturer's instructions. Size of sheared DNA and product of every further step was measured by using agarose gel electrophoresis. According to Li et al. (2013), DNA libraries were constructed, and "with-beads" method was adopted in this protocol to

obtain higher yield. Inline indices were added to the adapter to label the samples in the ligation step of library preparation to ease the possible risk of cross contamination among the samples during subsequent gene capture step, then inline-index coded DNA libraries were pooled together equimolarly.

A cross species gene capture was done by following Li et al. (2013), and genes were captured for two repeated trials that increase the accuracy of enriched gene. The sequences of *Denticeps clupeioides* and *Ilisha elongata* were used for Clupeiform bait design for capturing maximum genes. The

**Fig. 1** Sampling sites in the Arabian Sea and Bay of Bengal

enriched libraries were amplified with IS4 and indexing primers according to Meyer and Kircher (2010). Final products were pooled in equimolar ratios for sequencing on an Illumina HiSeqX 10 l aneat Annoroad (Beijing, China).

Data preparation read assembly and post assembly processing

Data processing, read assembly, and post assembly processing were performed according to the description in Assexon pipeline (Yuan et al. 2019). Raw reads from each sample were parsed according to their 8 bp barcodes on P7 adapter using bcl2fastq v1.8.3 (Illumina). Low-quality bases and sequence adaptors were trimmed by using trim_galore v0.4.1 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Coding frame of each marker sequence is predicted and corrected using a perl script (predict_frames.pl). Coding sequences were extracted and translated into amino acid sequences by using Bio::Seq module in Bioperl (Stajich et al. 2002).

PCR duplicates were excluded by using “-fastx_uniques” command in USEARCH v10.0.240 (Edgar 2010). *Danio rerio* (<https://doi.org/10.5061/dyrad.2j5b4>) was used as a reference species to parse reads to each gene file. Reads are sorted to references with BLAST hit using UBLAST with a relaxed e-value of 1×10^{-4} . Reads of each locus were assembled (De novo assembly) into contigs by a conservative assembler SGA (Simpson and Durbin 2012). Contigs are locally aligned to protein sequences of references using the “protein2dna” model in the package under Exonerate (Slater and Birney 2005). Orthologous genes were picked up by reciprocal blast method.

The output amino acid (AA) sequences were aligned in batch using MAFFT v7.369b (Katoh and Standley 2013). The AA sequences were translated back to DNA after alignment. Poor alignments could interfere in phylogenetic inference, so badly aligned sequences in coding regions were removed by filter.pl according to Yuan et al. (2019). Summary statistics (e.g., number of enriched samples, GC content, percentage of missing data) for coding and flanking region of each locus and sample were extracted by using statistics.pl (Yuan et al. 2019).

SNP calling

A custom Perl script was used to make consensus sequences for each target locus from assembled contigs (Yuan et al. 2016) and then reads with adapter sequences trimmed and low-quality reads were excluded by using BWA v0.7.5. Samtools were used for converting the sequence map format (SAM) files into binary format (BAM) (Li and Durbin 2009). SNP sites were genotyped based on the BAM files using GATK-3.2.2 (McKenna et al. 2010). GATK Best Practices recommendations were followed (DePristo et al. 2011; Van der Auwera et al. 2013). Single SNP per locus with the least

amount of missing data and highest-quality score was kept for most analyses to meet the requirement of linkage equilibrium. Custom Perl script was used again to convert the SNP VCF file into NEXUS file and STRUCTURE input file (Yuan et al. 2016).

Phylogenetic analysis and network based on gene-capture data

The aligned DNA sequences were used for reconstructing phylogeny. Concatenated maximum likelihood (ML) trees were reconstructed under IQtree v1.6.9 with 1000 bootstrap replicates (Nguyen et al. 2015) and displayed by FigTree.v1.4.4. Model selection and data partition were also automatically conducted by IQtree v1.6.9 (Schrempf et al. 2016).

A median-joining network was formed by using Network 5.0.1.1 (Bandelt et al. 1999) to visualize genetic clustering of the different populations. Network only uses polymorphic sites, which are filtered by using Vcftools 0.1.15 (<https://vcftools.github.io/>). Only 523 SNP sites less than 2% missing data were used to build a network among the individuals. The SNP VCF file was converted to .rdf file by VCF2RDF converter to use as an input file in Network.

Genetic variation and population clustering

An analysis of molecular variance (AMOVA) was performed on the SNP data using ARLEQUIN 3.5.2 with 10,000 permutations (Excoffier et al. 2005). PGDspider 2.1.1.2 was used to convert the SNP vcf file into arlequin (.arp) file for using as an input file in ARLEQUIN (Lischer and Excoffier 2012). Nucleotide diversity for each population was projected by using DnaSP v6.12.03 (Rozas et al. 2017). SNP data of variant call format (.vcf) was used as an input file for DnaSP. Pairwise F_{ST} matrix (Weir and Cockerham 1984) supported by ARLEQUIN 3.5.2 was used to calculate the genetic variation among groups, among populations within groups and within populations. Rcmd implemented in ARLEQUIN 3.5.2 was used to generate a simple graphic representation of pair-wise F_{ST} value. Text editor Notepad+ was used to edit the project file at the time of AMOVA and F_{ST} analysis.

Genetic partitioning of the 20 individuals also was assessed using STRUCTURE v2.3.4 (Pritchard et al. 2000) based on the data containing single SNP per locus (976 SNP loci). For the STRUCTURE runs, the initial burn was set in 50,000 replicates, followed by 500,000 replicates for each K (number of genetic clusters). The analysis were run for $K = 1-4$, each replicated three times. STRUCTURE HARVESTER 0.693 (Earl and VonHoldt 2012) was used for best K selector. Finally, the result of STRUCTURE with best K was transferred to plot form in STRUCTURE v2.3.4. Principal component analysis (PCA) was computed in R environment by using the ADE4

Table 2 Results of analysis of molecular variance (AMOVA)

Source of variation	Sum of squares	Variance components	Percentage variation
Among groups	410	14.86	10.85
Among populations within groups	247	0.21	0.15
Within populations	4389	121.93	89
Total	5046	137.00	

R package (Jombart and Ahmed 2011). The SNP VCF file was filtered by using VcfTools 0.1.15 (<https://vcftools.github.io/>) for using as an input file in PCA computation. Only 523 SNP sites less than 2% missing data were used for PCA analysis. Values of pc1 and pc 2 were plotted to draw the genetic clustering of individuals from different populations.

Results

Species identification

The samples were identified based on 11 morphometric and meristic key characters according to the description of Rahman et al. (2009) and Rahman et al. (2018) (Table S1). Moreover, all of the 20 samples showed more than 98% similarities to *H. kelee* COI gene sequence by BLAST to Gene Bank, which is in the range of similarities with the Kelee sequences lodged in GeneBank (Table S2).

Sequencing results (NGS)

Each sample produced 5,329,535 reads on average, after trimming off adapter sequences and reads with low-quality score ($Q < 20$), and this number was 99.56% of raw reads (Table S3). After removing the reads of PCR duplicates, the number of obtained reads was 5,002,704, which is 93.87% of previous cleaned reads. On average from each sample 1002 target loci were obtained with the best one had 1364 loci and the lowest one had 561 captured loci. The average number of captured loci of the out-group sample was 1558. All loci were checked by naked eye. Loci with same sequence for all

samples and weird segments in the alignment were excluded. Loci with samples from only one location or loci with lower than four samples were also excluded. After the exclusion, the total number of loci kept was 2090.

Phylogenetic relationships based on genome-scale nuclear data

Result of network analysis showed that Bay of Bengal's populations (i.e., CB, MK, and DC) clustered together and samples of the SA form a separate group (Fig. 5). The maximum likelihood tree built by IQtree represented that all of the individuals of the Arabian Sea primarily clustered in the base of the clade and polyphyletic to the Bay of Bengal individuals (Fig. 4). All of the individuals of DC (CL_DC_1362), MK (CL_2048), and CB (CL_2050) clustered together and separated from the individuals of SA (CL_2312) with a bootstrap value of 82%. That means all of the samples of the Bay of Bengal form a monophyletic clade and represent a significant isolation from the individuals of the Arabian Sea.

Genetic diversity

Analysis of molecular variance (AMOVA) showed that percentages of variation among groups (MK, CB, DC, and SA) and among population within groups were 10.85% and 0.15%, respectively. The variation between individuals within populations contributed 89% variation (Table 2). Nucleotide diversity (P_i) ranged from 0.014690 in the samples of Cox's Bazar Fish landing Station (CB) to 0.010337 in that of Dhal Char (DC) (Table 3). Pairwise F_{ST} values were low among the populations of CB, MK, and DC, but F_{ST} values between SA

Table 3 Nucleotide diversity (P_i) of each population

Population	Population (full name)	Theta (per site) from P_i
CB	Cox's Bazar Fish landing Station	0.014690
MK	Adinath Bazar Fish Market, Maheshkhali	0.012316
DC	Dhal Char	0.010337
SA (Arabian Sea)	Salalah Fish landing Station	0.010831
Average		0.012044
Bay of Bengal		0.012138

CB Cox's Bazar Fish landing Station, MK Adinath Bazar Fish Market, Maheshkhali, DC Dhal Char, SA Arabian Sea

Table 4 Pairwise differences (F_{ST}) among populations

	MK	CB	DC	SA (Arabian Sea)
MK	0.00000			
CB	−0.00589	0.00000		
DC	0.00900	0.01061	0.00000	
SA	0.09710	0.10979	0.12669	0.00000
Bay of Bengal (MK, CB and DC)				0.10955

CB Cox's Bazar Fish landing Station, MK Adinath Bazar Fish Market, Maheshkhali, DC Dhal Char, SA Arabian Sea

and rest of three populations (Table 4; Fig. 2) were significant. F_{ST} value between Arabian Sea and Bay of Bengal as a whole is 0.10955, which was significant ($p < 0.05$) (Table 4).

Structure analysis

Structure analysis supported two groups ($K = 2$) in the studied population. Samples from three locations of the Bay of Bengal (i.e., CB, MK, and DC) are indistinguishable from each other, but clearly distinct from the population of SA (Fig. 3). Beside this, structure also shows little patch of gene flow between two populations. PCA also revealed almost same pattern like structure result (Fig. S2). Samples of SA and Bay of Bengal

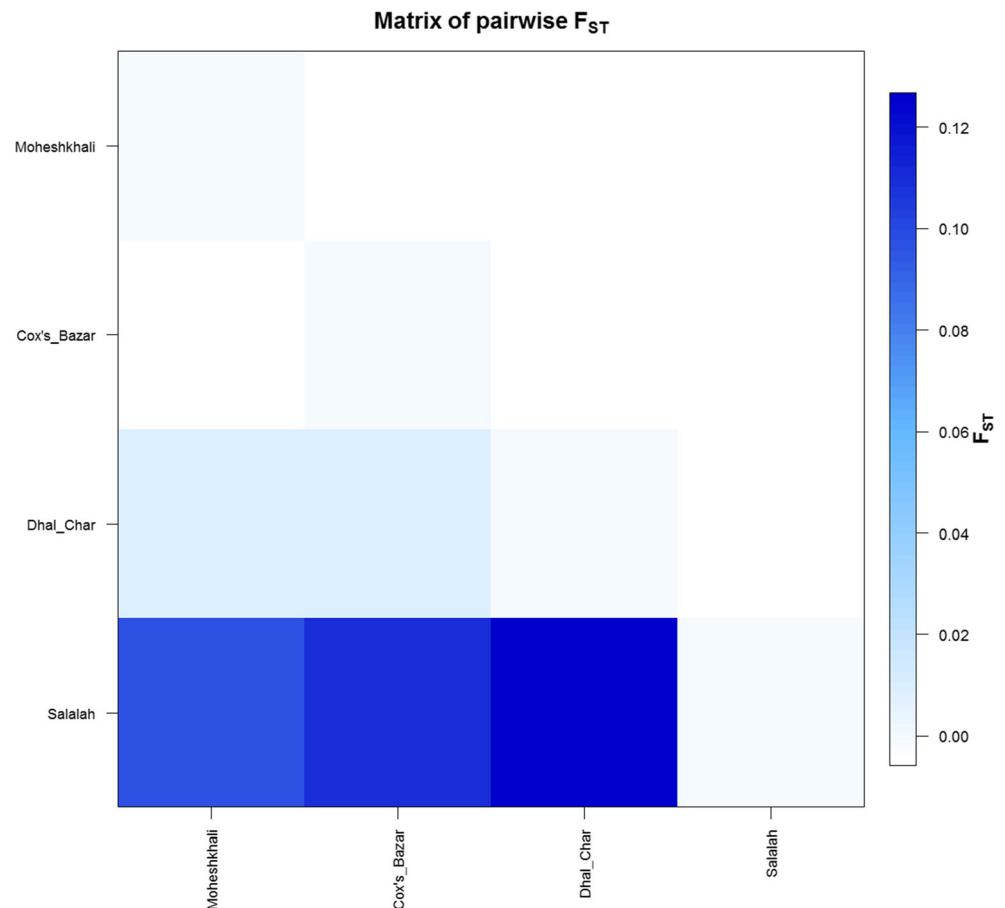
clustered into two separate groups, which showed almost clear separation of two groups (i.e., Arabian Sea and Bay of Bengal).

Discussion

Presence of moderate genetic diversity

The average nucleotide diversity of Kelee Shad was 0.012044 (0.010337–0.014690) (Table 3). This is moderate in shads, higher than Hilsa shad *Tenualosa ilisha* (Mohindra et al. 2019), Elongate ilisha *Ilisha elongata* (Wang et al. 2016), Gold-spotted grenadier anchovy *Coilia dussumieri*

Fig. 2 Matrix of pairwise F_{ST} among populations. The range of color deepness represents the level of genetic distances



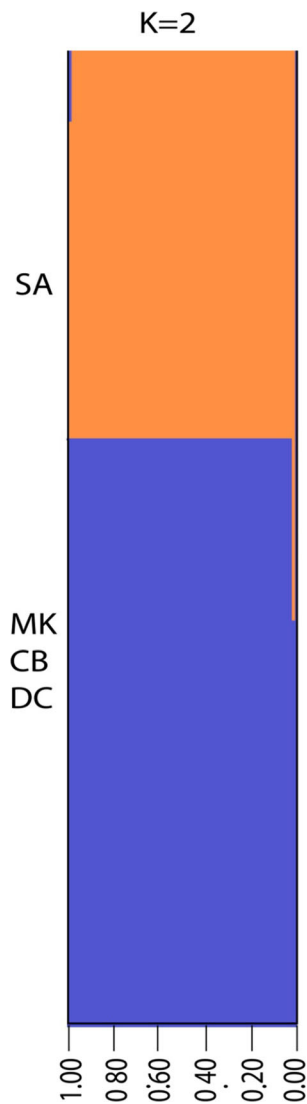


Fig. 3 Structure analysis based on 976 SNP loci. SA Salalah Fish landing Station; MK Adinath Bazar Fish Market, Maheshkhali; CB Cox's Bazar Fish landing Station; DC Dhal Char

(Kathirvelpandian et al. 2014) and Japanese anchovy *Engraulis japonicus* (Zheng et al. 2015) but lower than Pacific herring *Clupea pallasii* (Liu et al. 2011). The nucleotide diversity suggests that the population status of Kelee shad is in almost stable condition and has a moderate population size. However, because of environmental changes and pollution, the species has possibility to lose the stability in the time being, so require placing this species in the conservation and fisheries management plan in the coastal countries of the Indian Ocean.

AMOVA of Kelee shad within population is 89% (Table 2), which is almost similar to Hilsa shad (Asaduzzaman et al. 2019) and Elongate Ilisha (Wang et al. 2016), but Brahmane et al. (2013) found higher value in Hilsa shad in Ganga and Hooghly rivers allied to the Bay of Bengal. Low percentage of variation among populations within groups

(AMOVA) and non-significant F_{ST} value suggests the genetic homogeneity of Kelee shad in Bay of Bengal (i.e., CB, MK, and DC), and significant F_{ST} value with the SA samples represents the genetic heterogeneity between these two water bodies. Mohindra et al. (2019) found similar output in Hilsa shad from three important rivers of India (Hoogli, Brahmaputra, and Padma of Indian part); all of these rivers receive the anadromous migration of Hilsa shad from the Bay of Bengal.

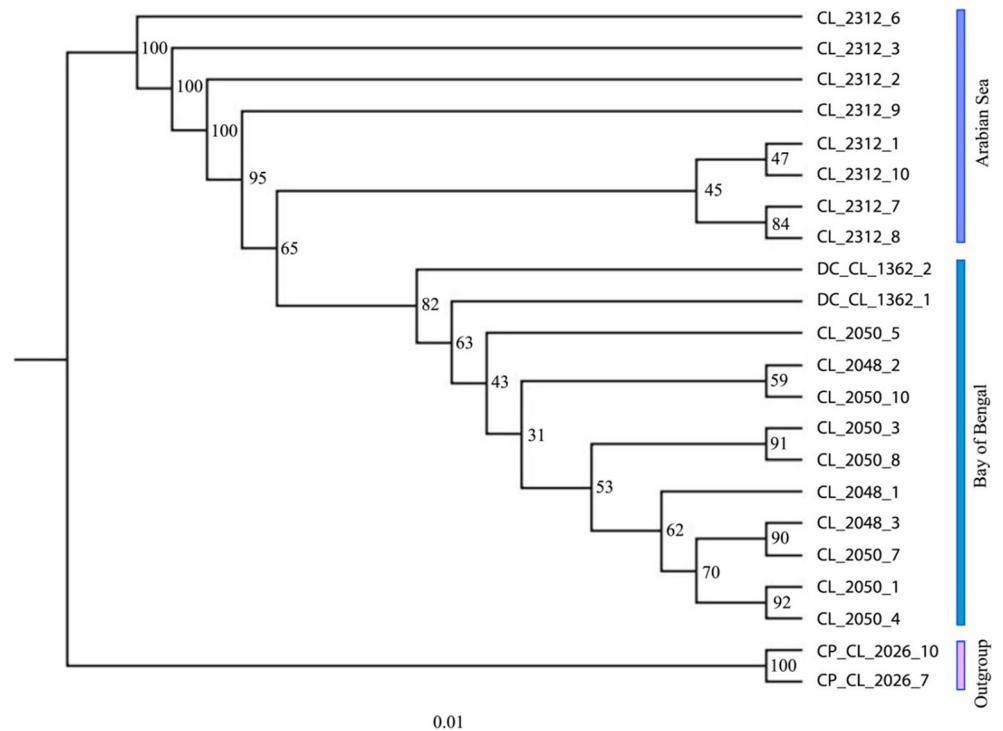
Presence of single population in Bay of Bengal

Kelee shad of all of the three locations of Bay of Bengal (CB, MK, and DC) clustered together in structure, phylogenetic tree, network of SNP loci, and PCA as they belong to the same population. In the maximum likelihood IQtree, samples of all of the three Bay of Bengal locations clustered together and made a monophyletic clade from the samples of Arabian Sea (Fig. 4). F_{ST} values of these populations are also below the significance level ($p > 0.05$). This result is similar to the situation in the closely related *T. ilisha*, which is represented by a single population in the Bay of Bengal (Brahmane et al. 2013). Similar results were also found by Salini et al. (2004) by allozyme and morphological variation analysis in *T. ilisha*. They found no significant differences in allele frequencies detected in Bay of Bengal samples, and that there is a substantial gene flow among the groups of Hilsa samples in Bay of Bengal comprising a single population.

Populations of Bay of Bengal and Arabian Sea were isolated

Based on structure and PCA result, the populations of Kelee shad of Bay of Bengal and Arabian Sea are separated from each other (Figs. 3 and S2). Median joining network also represented the same trend (Fig. 5). F_{ST} value also supported this separation, because SA samples were separated from the Bay of Bengal samples by F_{ST} value 0.10955 (Table 4), which is under significant level ($p < 0.05$). Maximum Likelihood IQ tree also chains this fact by single separated monophyletic clade of Bay of Bengal samples from Arabian Sea samples (Fig. 4). The same trend was also found in another Indian Shad, *T. ilisha*. Three population-specific haplotypes of *T. ilisha* were observed in Ganga River, which is joined with the Bay of Bengal and two from the Arabian Sea associated River Tapi. Neighbor joining tree based on cytochrome B gene analysis showed that populations of Bay of Bengal and Arabian Sea belong two distinct clusters. F_{ST} value (0.97) also supported the presence of significant genetic difference between the populations of these two-water bodies. Based on allozyme and morphological variation analysis in *T. ilisha*, it was also found that the populations of Bay of Bengal and Persian Gulf (allied to the Arabian Sea) samples have

Fig. 4 A ML tree based on sequences data concatenating 2090 loci reconstructed by using IQtree v1.6.9 with 1000 bootstrap replicates

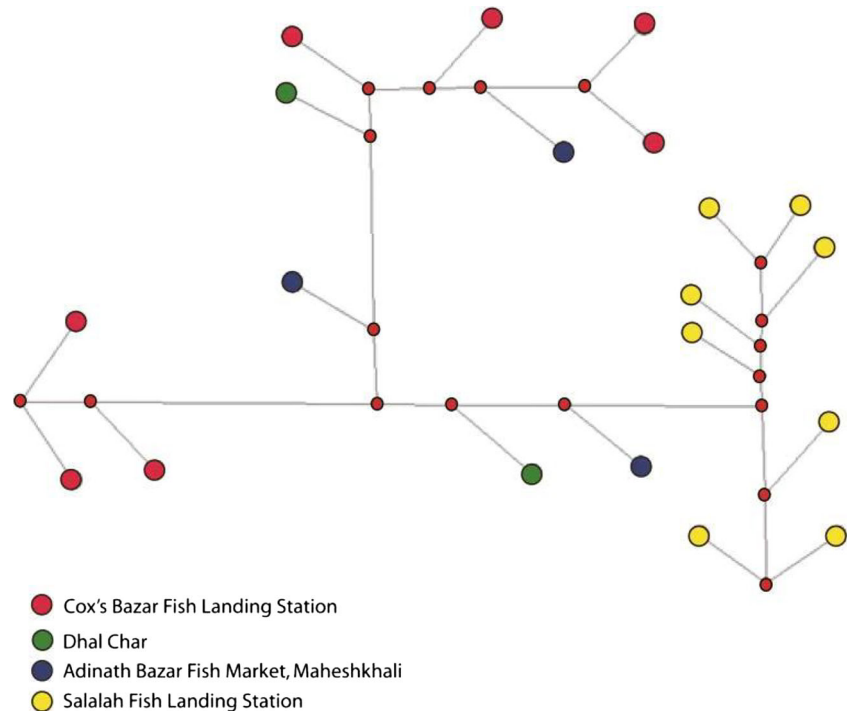


significant differences in allele frequencies (Salini et al. 2004; Behera et al. 2015). In PCA they made separate clusters and also F_{ST} value between these populations was significant.

Once (almost 18,000 years ago) the Arabian Sea had a close connection with Bay of Bengal through Laccadive Sea, Gulf of Mannar, and Palk Bay. So, this likely was an easy way for oceanographic dispersion of Kelee shad between

these two water bodies. After that period, a bridge of limestone shoals, coral reefs, and tombolo called as “Ram Bridge” or “Adam’s Bridge” (about 48 km) originated between Pamban Island off the south-eastern coast of Tamil Nadu, India, and Mannar Island, off the north-western coast of Sri Lanka (Weerakkody 1997; GSI 2003; Karanth 2003). Ram bridge including Pamban Island, Mannar Island, and Sri

Fig. 5 A median joining network of studied individuals based on 523 SNP loci by Network 5.0.1.1



Lanka formed almost 500 km land barrier in that portion of Indian Ocean. Therefore, this land barrier was supposed to halt the dispersion and migration of coastal water dwelling tropical fish species including Kelee shad for several thousand years. Although in the year of 1480 most part of the Ram bridge was broken by a cyclone (Garg 1992) but still the shallow water areas of the former bridge might be inconvenient for dispersion of most of the fish. Because of this, Kelee shad of Arabian Sea and Bay of Bengal might be isolated in its own area and stopped gene flow in the last more than 10,000 years.

Many studies suggested that the oceanographic characters and water quality of these two parts of the Indian Ocean were different from each other due to summer monsoon, river runoff, and weather of adjacent land. Different water parameters like salinity, temperature fluctuation, dissolved oxygen, and nutrient salts made two different water ecosystems in these two water bodies (Panikkar et al. 1966; Jensen 2001). Differentiation of weather of adjacent land had started especially since desert formation in the Arabian Peninsula. Geological record supported that the entire Arabian Peninsula was a lush of grasslands and forest 100,000 years ago (Parton et al. 2015). Before that desert formation, adjacent lands of these two marine water bodies were likely much more similar than they are today. The desert formation makes a huge change in the water quality of Arabian Sea. Kelee shad of these two water bodies adapt in these two different environments more than 100,000 years. Finally, after the Ram bridge formation, Kelee shad population of these water bodies might be separated and gene flow stopped. Because of these two vital reasons, the present genetic differentiation occurred between the Bay of Bengal and the Arabian Sea.

Land barriers likely play an important role in isolation of populations, finally leading to the origin of new species. As for example, the Indo-Pacific barrier resulted in significant differences in the populations of several species, like Peacock Grouper (*Cephalopholis argus*) (Gaither et al. 2011). Another example comes from the Arabian peninsula, where genetic differences were created between the Yellow bar Angel Fish (*Pomacanthus maculosus*) populations of the Western Arabian Peninsula (Red Sea and Gulf of Aden) and the Eastern Arabian Peninsula (Persian Gulf and Sea of Oman) (Torquato et al. 2019). In the same way, it was speculated from the present study that Southern India along with Ram bridge and Sri Lanka acted as a means of isolation inside the Kelee shad population and segregate it into two distinctly different populations in the Bay of Bengal and the Arabian Sea. The oceanographic characters and water quality like salinity, temperature, and nutrients of these two water bodies separated by Southern India are also significantly different from each other.

Here we stated two possible hypotheses of genetic differentiation of Kelee shad between the Bay of Bengal and the

Arabian Sea (i.e., Ram bridge creation and desert formation in the Arabian Peninsula). Nevertheless, lack of sample collection from different coasts of India and Pakistan was a limitation of this present study. If more samples were collected along the coast of the Indian subcontinent, maybe gradual genetic changes of Kelee shad population were revealed. Samples of more locations might help us to make more concrete conclusion about this shad's population structure. This study also makes a space for further studies on this fish species population structure based on more samples from different locations of Indian coasts including the both sides of the mentioned land barrier.

Arabian Sea represents the main stock of Kelee shad

From maximum likelihood IQ tree, it was established that a monophyletic clade of Bay of Bengal's samples was originated from the population of Arabian Sea's samples with a bootstrap value of 82% (Fig. 4) and the Arabian Sea was represented as the main stock. Kelee shad widely distributed in the Arabian Sea and associated Gulf of Oman, Gulf of Aden, coast of Africa, and Maputo Bay and less distributed in the Bay of Bengal and associated coasts and bays (Gjøsaeter and Sousa 1983; Rahman et al. 2009; Panhwar et al. 2013; Cardinale et al. 2014; Dwivedi 2019). This distribution pattern also supported that the Arabian Sea represents the main stock.

Acknowledgments We would like to express our gratitude to the editor and anonymous reviewers for their valuable comments to improve the article. We are grateful to Mr. Sujana Dutta for helping in sample collection and Irin Sultana for some technical support. The Shanghai Oceanus Super-Computing Center (SOSC) provided computational resource for data analyses.

Funding This work was supported by the "Shanghai Universities First-class Disciplines Project of Fisheries and Science and Technology Commission of Shanghai Municipality" (19410740500).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval No animal testing was performed during this study.

Sampling and field studies All necessary permits for sampling and observational field studies have been obtained by the authors from the competent authorities.

Data availability The trimmed fastq data are available in the NCBI repository with accession number, PRJNA490771. Sequence data and other data files are available in the [supplementary materials](#) of this article.

Authors' contributions C.L. and A.S. designed the project and wrote the primary manuscript. A.S. performed lab work, data preparation, and analysis. All authors contributed in editing and revising the manuscript.

References

- Asaduzzaman M, Wahab MA, Rahman MJ, Nahiduzzaman M, Dickson MW, Igarashi Y, Asakawa S, Wong LL (2019) Fine-scale population structure and ecotypes of anadromous Hilsa shad (*Tenualosa ilisha*) across complex aquatic ecosystems revealed by NextRAD genotyping. *Sci Rep* 9(16050):1–14
- Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Behera BK, Singh NS, Paria P, Sahoo AK, Panda D, Meena DK, Das P, Pakrashi S, Biswas DK, Sharma AP (2015) Population genetic structure of Indian shad, *Tenualosa ilisha* inferred from variation in mitochondrial DNA sequences. *J Environ Biol* 36(5):1193–1197
- Blaber SJM, Milton DA, Pang J, Wong P, Ong B, Nyigo L, Lubim D (1996) The biology of the tropical shad *Tenualosa toli* from Sarawak: first evidence of protandry in the Clupeiformes? *Environ Biol Fish* 46:225–242
- Blaber SJM, Fry G, Milton DA, Velde TVD (2005) The life history of *Tenualosa macrura* in Sarawak, further notes on protandry in the genus and management strategies. *Fish Manag Ecol* 12:201–210
- Brahmane MP, Kundu SN, Das MK, Sharma AP (2013) Low genetic diversity and absence of population differentiation of hilsa (*Tenualosa ilisha*) revealed by mitochondrial DNA cytochrome b region in Ganga and Hooghly rivers. *Afr J Biotechnol* 12(22):3383–3389
- Brewer DT, Blaber SJM, Fry G, Merta GS, Efizon D (2001) Sawdust ingestion by the tropical shad (*Tenualosa macrura*, Teleostei: Clupeidae): implications for conservation and fisheries. *Biodivers Conserv* 97:239–249
- Cardinale M, Chacafe OE, Casini M, Chauca I (2014) CPUE trends of *Hilsa kelee* and *Thryssa vitirostris* exploited by the artisanal fish fisheries in Mozambique derived from an on-shore sampling of catches by trip. *Sci Mar* 78(1):1–10
- Charles-Dominique E, Albaret J (2003) African shads with emphasis on the West African shads *Ethmalosa fimbriata*. *Am Fish Soc Symp* 35:27–48
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hertl C et al (2011) A framework for variation discovery and genotyping using next generation DNA sequencing data. *Nat Genet* 43:491–498
- Dwivedi AK (2019) Differentiating three Indian shads by applying shape analysis from digital images. *J Fish Biol*:1–11
- Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinform* 26:2460–2461
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinforma Online* 1:47
- Gaither MR, Bowen BW, Taina-Rae B, Rocha LA, Newman SJ, Gomez JA, Herwerden LV, Craig MT (2011) Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the indo-pacific barrier with contemporary overlap in the coral triangle. *BMC Evol Biol* 11:189
- Garg GR (1992) Adam's Bridge. In: *Encyclopedia of the Hindu World*. South Asia Books, New Delhi 142p
- GIS (2003) Rameswaram. Min Earth Sci, Govt of India, New Delhi
- Gjøsaeter J, Sousa MI (1983) Reproduction, age and growth of the kelee shad, *Hilsa kelee* (Cuvier, 1829) (Pisces: fam: Clupeidae) with information on its fishery in Maputo Bay, Mozambique. *Rev de Invest Pesq* 8:53–80
- Huda MS, Haque ME (2003) Field guide of fin fishes of Sundarban. Aquatic Resources Division, Sundarban Biodiversity Conservation Project, Bangladesh Forest Department, Khulna 197
- Jensen TG (2001) Arabian Sea and bay of Bengal exchange salt and tracers in an ocean model. *Geophys Res Lett* 28(20):3967–3970
- Jia CC (1982) A study on culture of Yangtze River Reeves shad in pond. *Jiangsu Fisher Sci* 2:27–33
- Jombart T, Ahmed I (2011) Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinform* 27:3070–3071
- Karanth KP (2003) Evolution of disjunct distributions among wet-zone species of the Indian subcontinent: testing various hypotheses using a phylogenetic approach. *Curr Sci* 85(9):1276–1283
- Kathirvelpandian A, Gopalakrishnan A, Lakra WS, Krishna G, Sharma R, Divya PR, Kumar R, Jena JK (2014) Mitochondrial ATPase 6/8 genes reveals genetic divergence in the *Coilia dussumieri* (Valenciennes, 1858) populations of north east and northwest coasts of India. *Mol Biol Rep* 41:3723–3731
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
- Li H, Durbin R (2009) Fast and accurate short-read alignment with Burrows-Wheeler transform. *Bioinform* 25:1754–1760
- Li C, Hofreiter M, Straube N, Corrigan S, Naylor GJ (2013) Capturing protein-coding genes across highly divergent species. *Bio Tech* 54: 321–326
- Lischer HE, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinform* 28:298–299
- Liu JX, Tatarenkov A, Beacham TD, Gorbachev V, Wildes S, Avise JC (2011) Effects of Pleistocene climatic fluctuations on the phylogeographic and demographic histories of Pacific herring (*Clupea pallasii*). *Mol Ecol* 20:3879–3893
- Madduppa HH, Timm J, Kochzius M (2018) Reduced genetic diversity in Clown Anemonefish *Amphiprion ocellaris* in exploited reefs of Spermonde Archeipelago, Indonesia. *Front Mar Sci* 5:80
- Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
- Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed Target capture and sequencing. *Cold Spring Harb Protoc* 2010(6):pdb.prot 5448
- Mohindra V, Divya BK, Kumar R, Singh RK, Dwivedi AK, Mandal S, Masih P, Lal KK, Jena JK (2019) Genetic population structure of a highly migratory Hilsa Shad, *Tenualosa ilisha*, in three river systems, inferred from four mitochondrial genes analysis. *Environ Biol Fish* 102:939–954
- Nguyen LT, Schmidt HA, Haeseler VA, Minh BQ (2015) IQ-Tree: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol* 32:268–274
- Panhwar SK, Liu Q, Siddiqui G (2012) Growth, mortality and stock assessment of Kelee shad, *Hilsha kelee* (Fam: Clupeidae) in the coastal waters of Pakistan. *J Ichthyol* 53(5)
- Panikkar NK, Sc FA, Jayaraman R (1966) Biological and oceanographic differences between the Arabian Sea and the Bay of Bengal as observed from the Indian region. *Proc Ind Acad Sci Sec B* 64:231–240
- Parton A, Farrant AR, Leng MJ, Telfer MW, Groucutt HS, Petraglia MD, Parker AG (2015) Alluvial fan records from southeast Arabia reveal multiple windows for human dispersal. *Geology* 43(4):295–298
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multi-locus genotype data. *Genetics* 155:945–959
- Puvanasundram P, Esa YB, Rahim KAA, Amin SMN (2017) Phylogeography and population structure of *Tenualosa toli* inferred from cytochrome b mitochondrial DNA fragment. *J Environ Biol* 26:895–906
- Rahman AKA, Kabir SMH, Ahmad M, Ahmed ATA, Ahmed ZU, Begum ZNT, Hassan MA, Khondker M (2009) *Encyclopedia of Flora and Fauna of Bangladesh*. Asiat Soc Bangladesh 24:78–79

- Rahman ASMS, Begum MH, Reza HA, Ahsan MN, Ahmed MS (2018) Molecular characterization of Hilsa shads in Bangladesh using Cytochrome C oxidase subunit 1 (CO1) gene. *I J Fauna Bio Stud* 5(3):1–5
- Rao KS (1973) Eggs and early developmental stages of *Hilsa kelee* (Cuvier). *Ind J Fish* 20(1):250–255
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302
- Salini JP, Milton DA, Rahman MJ, Hussain MG (2004) Allozyme and morphological variation throughout the geographic range of the tropical shad, hilsa *Tenualosa ilisha*. *Fish Res* 66:53–69
- Schrempf D, Minh BQ, Maio ND, Haeseler AV, Kosiol C (2016) Reversible polymorphism-aware phylogenetic models and their application to tree inference. *J Theor Biol* 407:362–370
- Simpson JT, Durbin R (2012) Efficient de novo assembly of large genomes using compressed data structures. *Genet Res* 22:549–556
- Slater GS, Bimney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinform* 6:31
- Stajich JE, Block D, Boulez K et al (2002) Thebioperl toolkit: perl modules for the life sciences. *Genet Res* 12:1611–1618
- Torquato F, Range P, Ben-Hamadou R, Sigsgaard EE, Thomsen PF et al (2019) Consequences of marine barriers for genetic diversity of the coral specialist yellow bar angel fish from the northwestern Indian Ocean. *Eco Evol* 9:11215–11226
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A et al (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43:11.10.11–11.10.33
- Wang HP (1996) Status and conservation of Reeves shad resources in China. *World Fish Cent* 19(1):20–22
- Wang Q, Zhang J, Matsumoto H, Kim J, Li C (2016) Population structure of elongate ilisha *Ilisha elongata* along the Northwestern Pacific Coast revealed by mitochondrial control region sequences. *Fish Sci* 82:771–785
- Weerakkody U (1997) Potential impact of accelerated sea level rise on beaches of Sri Lanka. *J Coast Res* 24:225–242
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evol* 38:1358–1370
- Whitehead PJP (1985) Clupeoid fishes of the world (suborder Clupeoidei). An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf-herrings. Part 1. Chirocentridae, Clupeidae and Pristigasteridae. *FAO Fish Synop* 125(7/1):1–303
- Yang TY, Gao TX, Meng W, Jiang YL (2020) Genome-wide population structure and genetic diversity of Japanese whiting *Sillago japonica* inferred from genotyping-by-sequencing (GBS): implications for fisheries management. *Fish Res* 225:105501
- Yuan H, Jiang J, Jimenez FA, Hoberg EP, Cook JA, Galbreath KE, Li C (2016) Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms. *Mol Ecol Resour* 16:1095–1106
- Yuan H, Atta C, Tornabene L, Li C (2019) Assexon: assembling exon using gene capture data. *Evol Bioinforma* 15:1–13
- Zain NAM, Amin SMN, Arshad A, Yusoff FM, Kamarudin MS, Bokhari F (2018) Length-weight relationship models of two tropical shads, *Tenualosa macrura* (Bleeker, 1852) and *Tenualosa toli* (Valenciennes, 1847) in the coastal waters of Sarawak, Malaysia. *J Appl Ichthyol* 34:992–994
- Zheng W, Zou L, Han Z (2015) Genetic analysis of the populations of Japanese anchovy *Engraulis japonicus* from the Yellow Sea and East China Sea based on mitochondrial cytochrome b sequence. *Biochem Syst Ecol* 58:169–177

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.