BCSIR

Available online at www.banglajol.info/index.php/BJSIR

Bangladesh J. Sci. Ind. Res. **59(2)**, 87-104, 2024

BANGLADESH JOURNAL OF SCIENTIFIC AND INDUSTRIAL RESEARCH

ISSN. 0304-9809

Comparative analyses of complete mitochondrial genomes of four sillaginids fish (Perciformes: Sillaginidae) and phylogenetic implications

S. Saha[1,2], C. Jia[3], N. Song[1], T. Suzuki[4], M. A. Baki[2,5], T. Yang[6] and T. Gao[6]*

[1]*Fishery College, Ocean University of China, Qingdao 266003, Shandong, China*
[2]*Department of Zoology, Jagannath University, Dhaka-1100, Bangladesh*
[3]*College of Ecology and Environment, Hainan University, Haikou 570228, Hainan, China*
[4]*Amagasakikita Senior High School, Tsukaguchimachi, Amagasaki, Hyogo 6610002, Japan*
[5]*Noakhali Science and Technology University, Noakhali 3814, Bangladesh*
[6]*Fishery College, Zhejiang Ocean University, Zhoushan 316022, China*

ARTICLE INFO

Received: 07 February 2024

Revised: 04 April 2024

Accepted: 06 May 2024

eISSN 2224-7157/

DOI: https://doi.org/10.3329/bjsir.v59i2.71326

Abstract

Sillago muktijoddhai, *S. mengjialensis*, *S. sihama* and *Sillaginops macrolepis* (Perciformes, Sillaginidae) are common coastal sand borers. The first three species are found in the Northeastern Indian Ocean, and the fourth one is located in the Western Central Pacific Ocean only. Molecular data applying in the prior investigations on systematics and phylogenetic relationships were not only incomplete but also uncertain for these fishes. This study sequenced their complete mitochondrial genomes using Sanger Dideoxy DNA sequencing for the first time. Their complete mitochondrial genome was a circular molecule of 17,022, 16,624, 16,825 and 16,502bp in length for *S. muktijoddhai*, *S. mengjialensis*, *S. sihama* and *Sillaginops macrolepis*, respectively. Most protein-coding genes (PCGs) were initiated with the typical ATG codon and terminated with the TAA or TAG codon and the incomplete termination codon T/TA could be detected in the four species. The majority of AT-skew and GC-skew values of the entire mitogenomes among the four species were negative. The Ka/Ks ratio analyses indicated 13 PCGs were suffering strong purifying selection. In the phylogenetic analysis, *S. muktijoddhai*, *S. mengjialensis*, and *S. sihama* were placed with relative species of the genus *Sillago* supporting morphological phylogeny. However, *S. macrolepis* was situated in the clade of genus *Sillago* that contrasted to morphological phylogeny. Divergence time analysis showed that Sillaginidae species diverged around 61 million years ago.

Keywords: *Sillago muktijoddhai*; *Sillago mengjialensis*; *Sillago sihama*; *Sillaginops macrolepis*; mitogenome; Sillaginidae; phylogeny

Introduction

Sillago sihama and *Sillaginops macrolepis* are well-documented Sillaginidae species among 41 species. Earlier it was supposed that *S. sihama* (Fabricius, 1775), a species with two posterior extensions of the swimbladder, was widely distributed in the Indo-West Pacific region (McKay, 1992). Recent research found that there are at least eight divergent molecular lineages that are currently included under the name *S. sihama* in the Indo-West Pacific region (Cheng *et al*. 2020, Fig. 2). These so-called *S. sihama* are externally similar to *S. sihama* (type locality, southern Red Sea), but are neither identical with this species nor to each other, as these cryptic species belong to structural differences in their swimbladder and well divergent mitochondrial DNA lineages. Two new species such as *Sillago muktijoddhai* and *S. mengjialensis* were recently identified and taxonomically described based on morphology and DNA barcoding approach using partial cytochrome oxidase subunit I (COI) and 16S ribosomal RNA (16SrRNA) gene sequences from the Bay of Bengal,

*Corresponding author's e-mail: gaotianxiang0611@163.com; shilpisahazoo@gmail.com

Bangladesh (Saha *et al.* 2022). These two new species were previously misidentified as *S. sihama* in Bangladesh (Saha *et al.* 2022). Currently it was confirmed that *S. sihama* is only widely distributed in the Indian Ocean, the Red Sea (Eritrea and Saudi Arabia; Golani *et al*. 2013; Bogorodsky *et al*. 2014), Arabian Sea (Western India, Pakistan, Iran; Lakra *et al*. 2011, Cheng *et al*. 2020), South Africa (Mlalazi Estuary, Tugela Bank and Mhlathuze Estuary), and Bay of Bengal, Bangladesh (Saha *et al*. 2024). On the other hand, *Sillaginops macrolepis* is distributed in Indonesian Archipelago, New Britain, Solomon Islands, Philippine Islands (McKay, 1992) and Iriomote Island, Japan (Suzuki *et al*. 2001). The description and phylogeny of *Sillaginops macrolepis* were carried out solely based on morphological characters (McKay, 1992; Suzuki *et al*. 2001; Kaga, 2013). In addition, complete mitochondrial genome sequences of so called *S. sihama* i.e., misidentified *S. sihama* (KR363150, NC_016672) are available (Siyal *et al*. 2015; Liu *et al*. 2012). However, the complete genetic characteristics of *S. muktijoddhai*, *S. mengjialensis*, true *S. sihama* and *Sillaginops macrolepis* are not documented yet.

The mitochondrial (mt) genome offers an easier technique than the nuclear genome; the molecular dynamics and rearrangements that bear differences in the genome can be learned more efficiently. Additionally, for studying evolutionary biology, some characteristics of mtDNA make it an efficient approach, such as small size, cellular abundance, maternal inheritance, compact gene arrangement, and high rate of evolution (Moritz *et al*. 1987; Wolstenholme, 1992). Thus, mtDNA is usually thought to be an excellent molecular marker in fish taxa for phylogenetic analyses (Stepien and Kocher, 1997). Although small mt gene segments showed restriction in determining complex phylogenetic relationships in many fish heredities. However, in longer DNA sequences (e.g., mt genomes), the excessive informatory sites permit the deeper branches and higher-level relationships to be much perfectly determined (Miya and Nishida, 2000). Among 41 Sillaginidae species, 14 mt genome sequences representing nine species, i.e., *S. sihama* (misidentified) (NC016672, KR363150), *S. indica*, *S. parvisquamis*, *S. sinica*, *S. asiatica*, *S. japonica*, *S. aeolus* and *S. panijus* are available in GenBank. The present study sequenced and characterized the complete mt genomes of *Sillaginops macrolepis*, *S. muktijoddhai*, *S. mengjialensis*, and *S. sihama*. The study will give supportive genetic information for further studies in genetics and phylogenetics.

Materials and methods

Sampling and DNA extraction

A single individual of each *S. muktijoddhai*, *S. mengjialensis*, and *S. sihama*, collected from the coastal regions of Bangladesh, and a single individual of *Sillaginops macrolepis* collected from Iriomote Island, Japan, were selected for the present study. Muscle tissue for each species was deposited in the Fishery Ecology Laboratory, Fisheries College, Ocean University of China. Voucher specimen code, collection date, and collection location for each species are as follows: *S. muktijoddhai* - FEL_OUC142275, 2nd October 2018, Cox's Bazar (21.452° N, 91.964° E); *S. mengjialensis* - FEL_OUCCO21972, 16th February 2019, Cox's Bazar; *S. sihama* - FEL_OUCM8192, 6th August 2019, Maheshkhali (21.772° N, 91.888° E); *Sillaginops macrolepis* - FEL_OUCcx1, 28th July 2019, Iriomote Island (24.419° N,123.776° E). White epaxial muscle tissues were collected from fresh specimens and preserved in 95% ethanol at -20ºC. Genomic DNA was extracted by proteinase K digestion followed by a standard phenol-chloroform method (Sambrook *et al*. 1989).

PCR amplification and sequencing

A PCR-based primer walking method was used to ascertain the complete mitochondrial (mt) genome. At first, partial sequences were obtained by the 12S rRNA (MiFish-U-F 12S: GTYGGTAAAWCTCGTGCCAGC and MiFish-U-R 12S: CATAGTGGGGTATCTAATCCYA GTTTG [Miya *et al*. 2015]), 16S rRNA (16S-ar F: 5'-CGCCTGTTTATCAAAAACAT-3' and 16S-br R: 5'-CCGGTCTGAACTCAGATCACGT-3'), COI (FishF1: 5′-TCAACCAACCACAAA GAC ATTGG CAC-3′ and FishR1: 5′-TAGACTTCTGGGTGGCCAAAGAATCA-3′; FishF2: 5'-TCGACTAATCATAAAGATATCGGCAC-3' and FishR2: 5'-ACT TCAGGGTGACCGAAGA ATCAGAA-3' [Ward *et al*., 2005]; L5956-COI: 5'-CACAAAGACATTGGCACCCT-3' and H6558-COI: 5'-CCTCCTGCAGGGTCAAAGAA-3'; F5480-COI: 5'-TTATTACTCAGCC ATCCTACC-3' and R6120-COI: 5'-GGTCCGTAAGTAGCATTGT-3' [Present study]), and Cytb (L14912-Cytb: TTCCTAGCCATACAYTAYAC and H15149-Cytb: TGRGGMCAAATGT CCTTCTGAGGMGCCACC) universal primers.

Then according to the partially amplified fragments and the complete mitochondrial DNA sequences of Sillaginidae fishes reported in GenBank, the stepwise method was used to design the primers gradually in the middle, maintaining more than 50 bp overlap between two contiguous fragments to ensure accuracy of sequencing. Followed the general principles of primer design when designing primers. For example, the length of PCR amplification products was guaranteed to

be 400bp-650bp, the length of primers was generally 19bp-21bp, and the GC content was about 45% -55%. Afterward, Primer Premier 6.0 was used to check the designed primers. According to the principle of primer design, there should be no hairpin structure in the primer, and no dimer should be generated between the primer itself and the primer, trying to avoid a mismatch.

The PCR reaction system was 25 µl, including 17.5 µl ultrapure water, 2.5 µl 10 × PCR buffer, 2 µl dNTPs, 0.15 µl Taq DNA polymerase, 1 µl DNA template, and forward and reverse primers; the reaction conditions were 94°C Pre-denaturation, time 5min; 94°C denaturation, time 45s; 50°C annealing, time 45s; 72°C extension, time 45s; 35 cycles in total; last 72°C extension, time 10min. The purified products were sequenced using a BigDye Terminator cycle sequencing kit v2.0 (Applied Biosystems, Foster City, CA, USA), and sequencing was conducted on an ABI Prism 3730 automatic sequencer (Applied Biosystems) with both forward and reverse primers used for amplification.

Sequence assembly, mitogenome annotation and sequence analyses

SeqMan (DNAStar, USA) software was used to manually correct, align and assamble the sequences. Finally, the entire mitochondrial DNA sequence was spliced, and the full-length mitochondrial DNA genome was calculated. The start and end positions of each gene were determined by comparing with the complete mitochondrial DNA sequences of the Sillaginidae species published in GenBank (http://www.ncbi.nlm.nih.gov/genbank). For preliminary annotations, MitoAnnotator of Mitofish (https://mitofish.aori.u-tokyo.ac.jp/annotation/input/) was used (Iwasaki *et al*. 2013). Mitochondrial gene structure maps were drawn using CGView Server (Grant and Stothard, 2008) (http://stothard.afns.ualberta.ca/cgview_server/). The online software tRNAscan-SE 2.0 (Lowe and Chan, 2016) was used to identify tRNA genes and predict tRNA II Hierarchical structure diagram. RNA structure Web server (https://rna.urmc.rochester.edu/ RNAstructureWeb/ Servers/ Predict1/Predict1.html) was used to simulate and draw O_L (WANCY) structure. The location of the control area was determined through the position of $tRNA^{Phe}$ and $tRNA^{Pro}$ and analyzed the structure of the control area by comparing with the recognition sites in other sillaginids (Xiao *et al*. 2014; Cai *et al*. 2014; Siyal *et al*. 2015; Saha *et al*., 2022) and some other bony fishes (Cheng *et al*. 2012; Zhang *et al*. 2013). The online software Tandem Repeats Finder (Benson, 1999) was used to search and analyze the tandem repeats in the control region sequence. MEGA 6.0 software (Tamura *et al*. 2013) was used to count nucleotide composition, calculate pairwise sequence identities for entire mt genomes and each gene type, and non-synonymous substitutions rate to synonymous substitutions rate (Ka/Ks) ratios for protein-coding genes (PCGs). AT and GC-skews were calculated using the formulas (A-T)/ (A+T) and (G-C)/ (G+C) (Perna and Kocher, 1995). The MEGA 6.0 software was also used to determine codon usage of PCGs and K2P genetic distances among sillaginids based on 12 PCGs and four mitochondrial genes. One-way analysis of variance (ANOVA) was used to examine for significant distinction in sequence variability of various regions.

Phylogenetic analyses

Phylogenetic analyses were carried out on concatenated sequences of the PCGs except for ND6 due to its heterogeneous base composition and usually bad efficiency (Miya and Nishida, 2000; Miya *et al*. 2003). Fourteen sequences of only nine Sillaginidae species available in the NCBI were used to reveal the phylogenetic relationship. In addition, concatenated sequences of four mitochondrial genes (12S rRNA, 16S rRNA, COI and Cytb) of 24 Sillaginidae species available in the NCBI were also used to get wide species coverage.

Table. S1. Details of mitogenome sequences cited in this study

Species	Sampling location	GenBank accession no.	Reference
Sillaginopsis panijus	Cox's Bazar, Bangladesh	MT460675	Saha *et al.,* 2021
Sillaginopsis panijus	-	AP006802	Miya, Direct submission
Sillago parvisquamis	-	LC387779	-
Sillago indica	-	NC_025298	Xiao *et al.,* 2014a
Sillago sinica	-	NC_030373	Siyal *et al.,* 2015
Sillago asiatica	Sanya, Hainan, China	NC_025337	Xiao *et al.,* 2014b
Sillago aeolus	-	NC_025935	Cai *et al.,* 2014
Sillago sihama	Dongfang, Hainan, China	KR363150	Siyal *et al.,* 2015
	China	NC_016672	Liu *et al.,* 2012
Sillago japonica	-	AP006803	-
	Zhanjiang, China	NC_028228	Niu *et al.,* 2016
	-	MW248468	-
	-	KR363149	Siyal *et al.,* 2015
	-	AP017438	Satoh *et al.,* 2016

Table. S2. Details of sequences cited in phylogenetic analysis based on 4 mitochondrial genes

Species	Individual ID	Sampling location	GenBank accession no.				Reference
			12S rRNA	16S rRNA	COI	Cytb	
Sillaginodes punctata	X1	Queensland Museum, Australia	MF572001	MF572031	KU051719	MF571958	
Sillago aeolus	SB1	Basuo, China	MF572012	MF572042	KU051731	MF571969	
Sillago asiatica	Z1	Sanya, China	MF572013	MF572043	KU051912	MF571971	
Sillago arabica	C188	Karachi, Pakistan	MF572008	MF572038	MF571937	MF571970	
Sillago attenuata	C476	Karachi, Pakistan	MF572014	MF572044	MF571918	MF571972	
Sillago chondropus	C385	Karachi, Pakistan	MF572016	MF572022	MF571940	MF571973	
Sillago flindersi	SF1	Australia	MF572017	MF572023	KU051726	MF571950	
Sillago indica	D1	Karachi, Pakistan	MF572018	MF572024	KM350229	MF571951	
Sillago ingenuua ①	SW1	Dongshan, China	MF572019	MF572025	KU051978	MF571952	
Sillago ingenuua ②	TWW1	Keelung, Taiwan	MF572020	MF572026	KU051989	MF571953	
Sillago japonica	S1	Laizhou, China	MF572021	MF572027	KU051937	MF571954	
Sillago maculata	B1	Australia	MF571999	MF572029	KU051721	MF571956	Cheng *et al.*
Sillago panhwari	C219	Karachi, Pakistan	MF572015	MF572045	MF571947	MF571965	2020
Sillago parvisquamis	XLX1	Japan	MF572000	MF572030	HQ389247	MF571957	
Sillago shaoi	SSP1	Dongshan, China	MF572002	MF572032	KU051886	MF571959	
Sillago cf. *sihama* ①	C358	Karachi, Pakistan	MF572011	MF572041	MF571917	MF571968	
Sillago cf. *sihama* ②	D1	Ranong, Thailand	MF571998	MF572028	KU051822	MF571955	
Sillago cf. *sihama* ③	D2	Dongshan, China	MF572003	MF572033	KU051804	MF571960	
Sillago cf. *sihama* ⑤	C260	Karachi, Pakistan	MF572010	MF572040	MF571912	MF571967	
Sillago cf. *sihama* ⑥	C228	Karachi, Pakistan	MF572009	MF572039	MF571949	MF571966	
Sillago nigrofasciata	FD1	Fuding, China	MF572005	MF572035	KU051808	MF571962	
Sillago sinica	ZG1	Dongying, China	MF572004	MF572034	KU052008	MF571961	
Sillago parasihama	Yang1	Zhanjiang, China	MF572006	MF572036	MF571929	MF571963	
Sillago sp. (Taiwan)	ZP0807085	Taiwan	MF572007	MF572037	MF571941	MF571964	

Accession numbers and sampling details of the downloaded sequences were given in Supplementary Table S1 and S2. *Callanthias japonicus* (AP006796), *Dicentrarchus labrax* (NC_026074) and *Pempheris schwenkii* (NC_026237) were used for these cases as outgroup species according to Near *et al*. (2013) and Betancur-R. *et al*. (2013). Clustal X 2.1 (Larkin *et al*., 2007) was used to align the sequences. The best evolution model of nucleotide substitution was selected by jModelTest v2.1.10 (Darriba *et al*. 2012). The model with the lowest AIC (Akaike Information Criterion) value, such as the GTR+I+G model, was considered the best to describe the substitution pattern. The PAUP* 4.0 (Swofford, 2002) was used to perform maximum likelihood analysis with 1000 bootstrap replications. Partitioned (first, second, and third codon positions of PCGs) Bayesian phylogenetic analysis was done by MrBayes v. 3.2.6 (Ronquist *et al*. 2012) based on ten million generations.

The divergence time was estimated by BEAST.v1.10.4 (Drummond *et al*. 2012) using the lognormal relaxed clock (uncorrelated) model, GTR+I+G substitution model and yule speciation process as a tree prior. The default prior distributions of all parameters were used except ucld.mean (uniform), and calibration point derived from fossil age and putative geological vicariant events (Harrison *et al*. 2006; Mckay, 1992). The root of the family Sillaginidae was calibrated at upper Oligocene (Lower Eocene maximum) and BEAUti input was the same as Cheng *et al*. (2020).

Results and discussion

Genome structure and nucleotide composition

The complete mitogenomes of the four sillaginids were sequenced to be 16,502-17,022 bp in length and were submitted to GenBank with accession numbers MW532121 (*S. muktijoddhai*), MW532122 (*S. mengjialensis*), MW532123 (*S. sihama*) and MZ677462 (*Sillaginops macrolepis*). Each genome consists of 13 typical vertebrate PCGs, 22 tRNA genes, 2 rRNA genes (12S rRNA and 16S rRNA), and 2 non-coding regions (control region and L-strand replication origin; Table I and Figure 1). It has shown the resemblance with the canonical organization of the fish mitochondrial genome in both gene content and order (Boore, 1999). The encoding genes of mitogenome were located on H-strand except for ND6 and eight tRNA (Gln, Ala, Asn, Cys, Tyr, Ser-UCN, Glu and Pro) genes that were transcribed from L-strand as in many teleosts (Zardoya *et al*. 1995; Zardoya and Meyer, 1997; Cheng, 2012).

The GC content of each gene ranged from 29.17% to 60.61%, the highest content was found in $tRNA^{Cys}$ of *S. sihama*. Moreover, the highest AT content was found in $tRNA^{Gly}$ of *S. muktijoddhai* (Table I).

The AT content of the four sillaginids mt genomes was varied from 51.77% to 52.75% (Table II). In addition, the nucleotide base compositions at each codon position covering all 13 mt

Table I. Gene content of the mitochondrial genomes of *S. muktijoddhai* (S1), *S. mengjialensis* (S2), *S. sihama* (S3) and *S. macrolepis* (S4)

Feature	Strand	Size (bp)				GC_Percent (%)				Amino acids				Codon Start/Stop				Anti-codon	Intergenic Nucleotide * (bp)			
		S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4		S1	S2	S3	S4
$tRNA^{Phe}$	H	68	68	68	68	44.12	47.06	42.65	42.65									GAA	0	0	0	0
12S rRNA	H	954	951	951	951	49.58	50.58	50.47	49.53										0	0	0	0
$tRNA^{Val}$	H	72	73	73	73	50.00	47.95	49.32	46.58									UAC	0	0	0	0
16S rRNA	H	1709	1701	1705	1698	46.58	45.86	48.27	47.29										0	0	0	0
$tRNA^{Leu}$ (UUR)	H	73	73	73	73	46.58	49.32	47.95	50.68									UAA	0	0	0	0
ND1	H	975	975	975	975	51.69	51.18	50.05	48.72	324	324	324	324	ATG/TAG	ATG/TAG	ATG/TAG	ATG/TAG		3	3	3	3
$tRNA^{Ile}$	H	72	71	71	71	50.00	49.30	47.89	45.07									GAU	-1	-1	-1	-1
$tRNA^{Gln}$	L	71	71	71	71	42.25	39.44	45.07	49.30									UUG	-1	-1	-1	-1
$tRNA^{Met}$	H	70	69	69	70	47.14	44.93	43.48	48.57									CAU	0	0	0	0
ND2	H	1046	1046	1046	1047	53.44	50.76	51.43	49.95	348	348	348	348	ATG/TA	ATG/TA	ATG/TA	ATG/TAG		0	0	0	0
$tRNA^{Trp}$	H	72	72	72	72	47.22	48.61	50.00	48.61									UCA	0	0	0	0
$tRNA^{Ala}$	L	69	69	69	69	42.03	43.48	37.68	37.68									UGC	1	1	1	1
$tRNA^{Asn}$	L	73	73	73	73	50.68	49.32	52.05	50.68									GUU	-1	-1	0	0
O_L	H	38	37	38	43	60.53	67.57	73.68	67.44										-3	-3	-3	-3
$tRNA^{Cys}$	L	66	65	66	66	48.48	60.00	60.61	54.55									GCA	0	0	0	0
$tRNA^{Tyr}$	L	70	70	70	70	45.71	52.86	48.57	48.57									GUA	1	1	1	1
CO I	H	1551	1551	1551	1551	46.29	47.78	48.10	46.87	516	516	516	516	GTG/TAG	GTG/TAA	GTG/TAA	GTG/TAA		0	0	0	0
$tRNA^{Ser}$(UCN)	L	71	71	71	71	49.30	47.89	47.89	45.07									UGA	4	4	5	4
$tRNA^{Asp}$	H	69	70	70	72	52.17	45.71	45.71	45.83									GUC	6	6	9	4
CO II	H	691	691	691	691	48.19	45.59	46.74	46.23	230	230	230	230	ATG/T	ATG/T	ATG/T	ATG/T		0	0	0	0
$tRNA^{Lys}$	H	74	74	74	74	44.59	44.59	44.59	44.59									UUU	1	1	1	1
ATPase8	H	165	168	168	168	44.85	48.81	49.40	43.45	54	55	55	55	ATG/TAA	ATG/TAA	ATG/TAA	ATG/TAA		-10	-16	-10	-10
ATPase6	H	683	689	683	683	48.17	47.02	46.41	45.08	227	229	227	227	ATG/TA	GTG/TA	GTG/TA	ATG/TA		0	0	0	0
CO III	H	785	785	786	738	48.15	49.43	48.35	47.97	261	261	261	245	ATG/TA	ATG/TA	ATG/TAA	ATG/AGA		0	0	0	48

Continued

$tRNA^{Gly}$	H	72	72	70	70	29.17	33.33	31.43	37.14									UCC	0	0	0	0
ND3	H	349	349	349	349	50.72	51.29	48.14	50.00	116	116	116	116	ATG/T	ATG/T	ATG/T	ATG/T		0	0	0	0
$tRNA^{Arg}$	H	69	69	69	69	39.13	36.23	37.68	37.68									UCG	0	0	0	0
ND4L	H	297	297	297	297	53.54	54.55	50.84	47.47	98	98	98	98	ATG/TAA	ATG/TAA	ATG/TAA	ATG/TAA		-7	-7	-7	-7
ND4	H	1381	1381	1381	1381	50.69	49.24	47.72	47.25	460	460	460	460	ATG/T	ATG/T	ATG/T	ATG/T		0	0	0	0
$tRNA^{His}$	H	70	69	69	69	30.00	31.88	31.88	34.78									GUG	0	0	0	0
$tRNA^{Ser}$(AGY)	H	67	68	68	68	43.28	55.88	54.41	51.47									GCU	2	2	2	2
$tRNA^{Leu}$(CUN)	H	73	73	73	73	45.21	42.47	46.58	43.84									UAG	0	0	0	0
ND5	H	1866	1839	1839	1839	49.09	49.21	48.40	48.23	621	612	612	612	ATG/TAA	ATG/TAA	ATG/TAA	ATG/TAA		-4	-4	-4	-4
ND6	L	522	522	522	522	52.11	52.68	50.19	47.32	173	173	173	173	ATG/TAA	ATG/TAA	ATG/TAA	ATG/TAA		0	0	1	0
$tRNA^{Glu}$	L	68	68	68	68	47.06	42.65	39.71	44.12									UUC	5	5	5	6
Cyt b	H	1141	1141	1141	1141	51.01	49.69	47.59	46.67	380	380	380	380	ATG/T	ATG/T	ATG/T	ATG/T		0	0	0	0
$tRNA^{Thr}$	H	72	72	72	72	51.39	52.78	58.33	58.33									UGU	2	3	8	1
$tRNA^{Pro}$	L	71	75	72	70	43.66	54.67	50.00	41.43									UGG	0	0	0	0
Control region	H	1319	953	1141	832	38.21	38.09	40.40	41.95										0	0	0	0

* As for the adjacent genes, positive number represented spaced, negative number indicated overlap and zero indicated continuous.

Table II. Nucleotide compositions of different regions in 4 Sillaginids mitochondrial genomes

			S. muktijoddhai					Total number	*S. mengjialensis*					Total number	*S. sihama*					Total number	*S. macrolepis*					Total number
			A	**T**	**G**	**C**	**A+T**		**A**	**T**	**G**	**C**	**A+T**		**A**	**T**	**G**	**C**	**A+T**		**A**	**T**	**G**	**C**	**A+T**	
Complete genome			25.10	26.72	18.90	29.27	51.82	17,022	25.43	26.34	18.41	29.83	51.77	16,624	25.50	26.58	18.68	29.24	52.08	16,825	26.12	26.63	17.99	29.26	52.75	16,502
	Protein coding genes	1st codon	20.7	29	16.6	33.2	49.7	3644	21.6	29	15.4	33.6	50.6	3638	22.0	30	15.4	32.1	52.0	3636	23.1	30	14.4	32.9	53.1	3621
		2nd codon	23.2	28	17.9	31.1	51.2	3642	23.8	28	17.1	31.6	51.8	3636	23.7	28	17.2	30.8	51.7	3635	25.0	28	16.5	30.3	53.0	3619
		3rd codon	22.6	27	20.3	29.9	49.6	3644	23.1	27	19.5	30.5	50.1	3638	23.0	27	19.9	29.9	50.0	3636	23.9	28	19.0	29.5	51.9	3620
		total	22.2	28.2	18.3	31.4	50.4	10930	22.8	27.9	17.3	31.9	50.7	10912	22.9	28.6	17.5	30.9	51.5	10907	24.0	28.4	16.6	30.9	52.4	10860
H Strand	tRNA		29.8	25.9	22.1	22.3	55.7	993	29.6	25.4	22.1	23.0	55.0	993	29.1	25.7	22.3	22.9	54.8	991	29.5	25.1	21.8	23.6	54.6	994
	rRNA		30.6	21.7	22.2	25.5	52.3	2663	30.5	22.0	21.9	25.6	52.5	2652	30.0	20.9	22.6	26.4	50.9	2656	30.4	21.5	22.6	25.5	51.9	2649
	D-loop		30.4	31.4	15.3	22.9	61.8	1319	31.8	30.1	16.2	21.9	61.9	953	31.1	28.5	18.1	22.3	59.6	1141	29.6	28.5	17.3	24.6	58.1	832
	Protein coding gene (ND6)	1st codon	13.2	32	42.0	12.6	45.2	174	13.2	28	42.5	16.7	41.2	174	11.5	26	45.4	16.7	37.5	174	13.8	28	43.7	14.4	41.8	174
		2nd codon	12.6	43	21.3	23.0	55.6	174	12.6	43	22.4	21.8	55.6	174	12.6	42	22.4	23.0	54.6	174	13.8	46	20.1	20.1	59.8	174
L Strand		3rd codon	21.8	21	37.9	19.5	42.8	174	17.8	28	39.7	14.9	45.8	174	20.1	37	32.8	10.3	57.1	174	26.4	30	31.6	12.1	56.4	174
		total	15.9	32.0	33.7	18.4	47.9	522	14.6	32.8	34.9	17.8	47.4	522	14.8	35.1	33.5	16.7	49.9	522	18.0	34.7	31.8	15.5	52.7	522
	tRNA		24.0	29.9	28.3	17.9	53.9	559	23.0	28.3	29.0	19.8	51.3	562	23.2	29.1	28.8	18.9	52.3	560	24.2	29.4	28.0	18.5	53.6	558

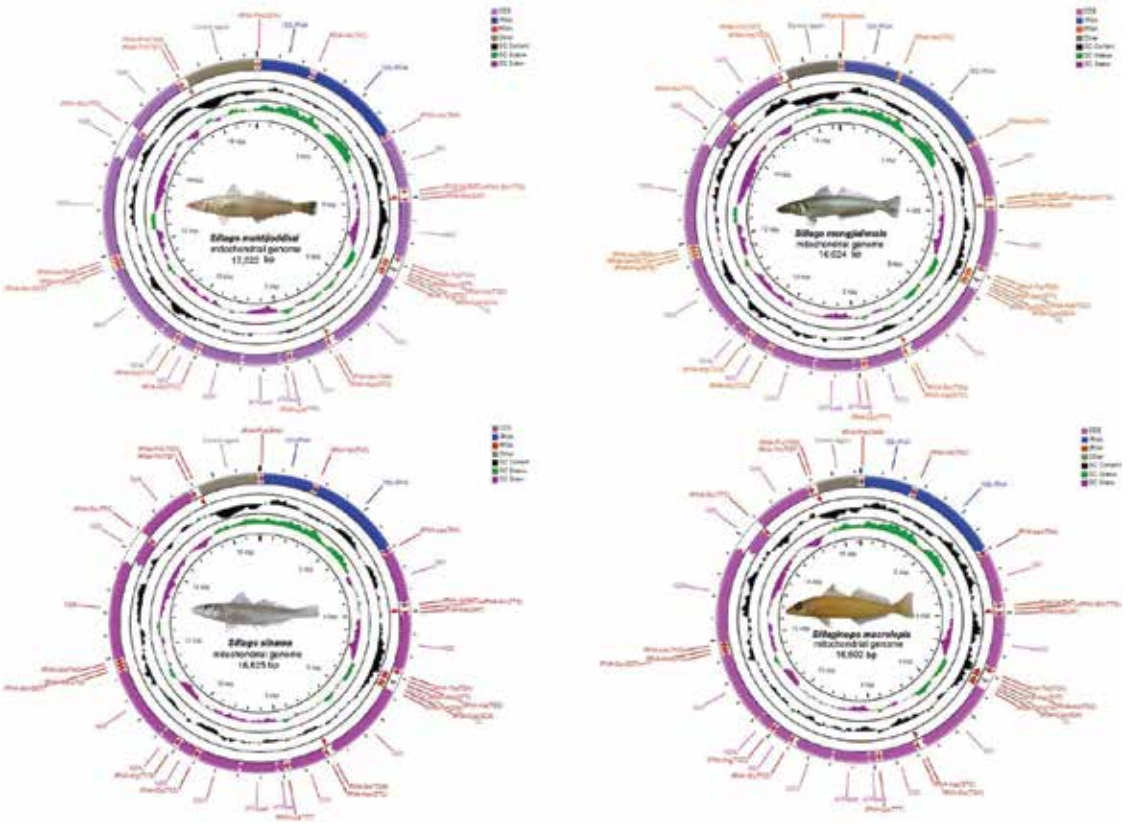

Fig. 1. Circular map of the mitogenome of *S. muktijoddhai*, *S. mengjialensis*, *S. sihama* and *S. macrolepis*. Genes encoded on the heavy or light strands are shown outside or inside the circular gene map, respectively.

PCGs were calculated (Table II). The 12 PCGs of the H strand showed a strong anti-G bias (20% or lower than 20%, $P < 0.01$) at all codon positions. On the contrary, the PCGs of the L strand (ND6) revealed a strong anti-C bias (17.5% or lower than 17.5%, $P < 0.01$) at the third codon positions. This result proved that mitochondrial anti-G bias is strand-specific instead of a general characteristic for all mt protein genes.

Generally, AT and GC skew describe the strand inconsistency of nucleotide composition. The AT and GC skews were all negative for the entire mitogenomes of four sillaginids (Figure 2). These results revealed that the A content was only slightly lower than T. On the contrary, C was notably more dominant than G. In addition, most of the PCGs encoded by the H-strand displayed negative AT-skews and GC-skews whereas the L-strand displayed negative AT-skews and positive GC-skews (G >C) in four sillaginids (Figure 2). Such skews to a specific nucleotide are assigned to differential mutational pressures attributed on the L- and H-strands (Hassanin *et al*. 2005), which is the result of asymmetric replication of mtDNA (Clayton, 1982; Tanaka and Ozawa, 1994).

Protein coding genes

The accumulative length of the PCGs ranged from 11382 to 11452 bp and 67.3% to 68.97% of the mt genome's total length. There were some overlapping regions between the adjoining PCGs due to the compactness of the animal mtDNA. There were 10 to 16 bp overlaps between ATP8 and ATP6, 7 bp between ND4L and ND4, 4 bp between ND5 and ND6 (Table I). Most of the PCGs started with the typical initial codon ATG, while the COI gene started with the initial codon GTG like many other metazoans (Wolstenholme, 1992). In addition, like most teleost fishes, ATG was the ATPase6 start codon in *S. muktijoddhai* and *Sillaginops macrolepis* but different in *S. mengjialensis* and *S. sihama* that was GTG, also found in *S. sihama* as the ATPase6 start codon (Siyal *et al*. 2015). Most of them had TAA or TAG as the stop codon, while COIII used AGA in *Sillaginops macrolepis*. AGA was not common as COIII stop codon in fish. There was a frameshift mutation in COIII and shortening the length of COIII that resulted in a truncated protein. Incomplete stop codon T or TA was found in COII, COIII, ATP6, ND2, ND3, ND4, and Cytb and were completed with the

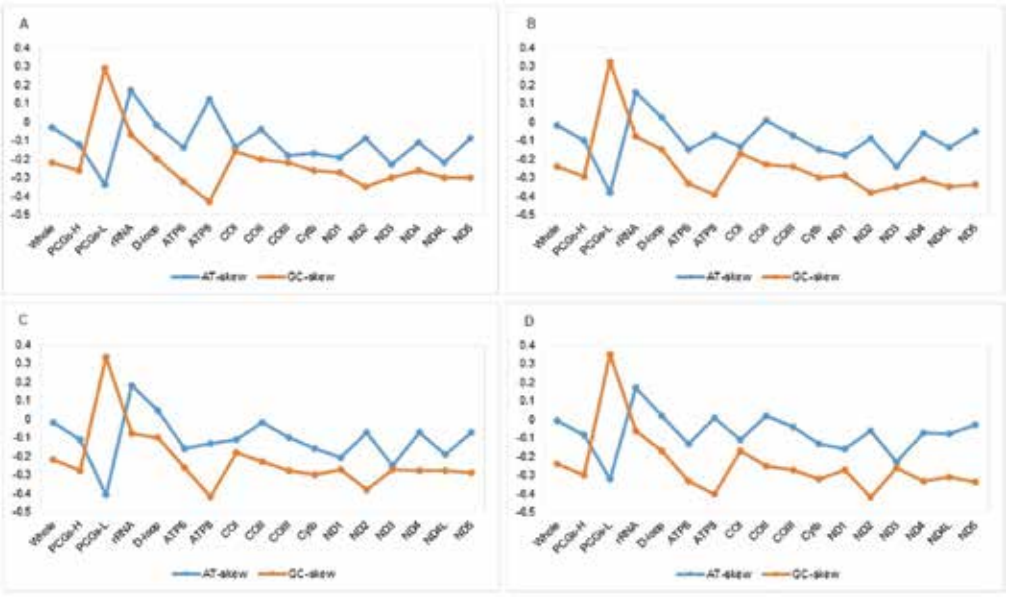

Fig. 2. The nucleotide skewness of four Sillaginids mitochondrial genomes. (A) *S. muktijoddhai*, (B) *S. mengjialensis*, (C) *S. sihama*, (D) *S. macrolepis*. The incomplete T-/TA- of the stop codon is not included.

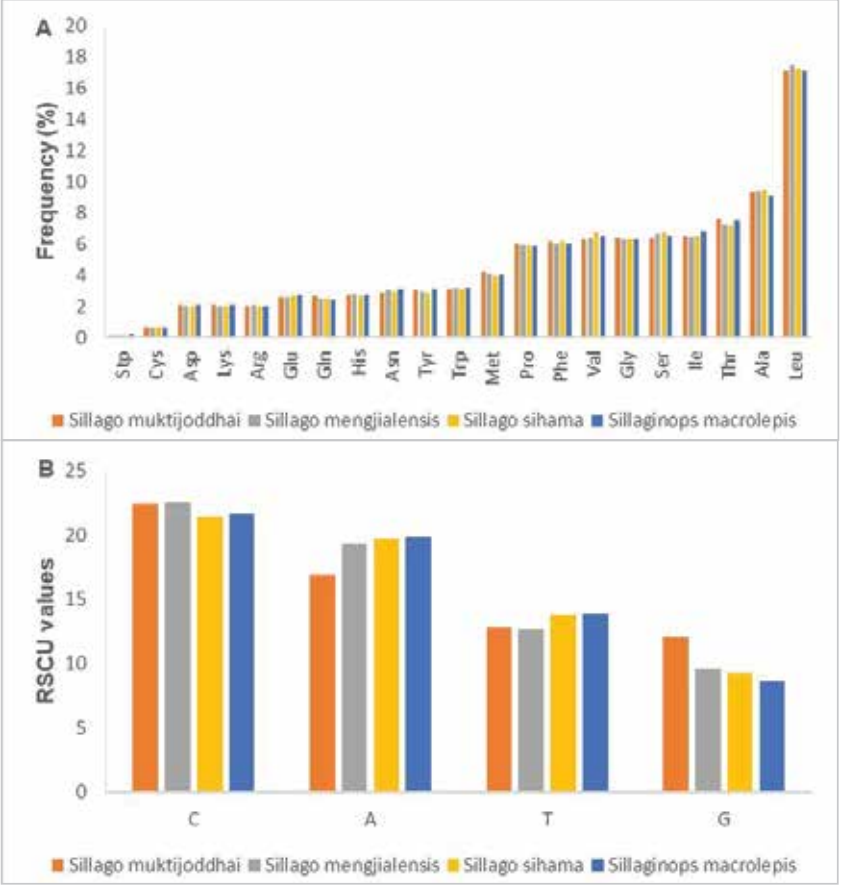

Fig. 3. Codon usage in 13 PCGs of four Sillaginids. (A) Codon frequency and (B) comparison of frequencies of codons ending with the same nucleotide by RSCU (Relative Synonymous Codon Usage) values

addition of 3' adenine residues to the mRNA by post-transcriptional polyadenylation (Coucheron *et al*. 2011; Ojala *et al*. 1981).

In terms of codon usage, the trend of codon usage in the 13 mitochondrial PCGs of four sillaginids was mostly consistent (Figure 3A). The total number of codons varied among four sillaginids from 3792-3814 (excluding incomplete T/TA stop codon) for 20 amino acids. Except for the complete stop codon (Stp), codons encoding cysteine (Cys) had the lowest usage frequency (0.63%-0.81%), and the codons encoding leucine (Leu) had the highest usage frequency (17.1%- 17.5%). Six different codons represented leucine and serine, and other amino acids were represented by only two or four. Comparing the RSCU (Relative synonymous codon usage) values (Figure 3B) showed that the usage trends of codons ending with the same base in the four sillaginids were the same. For example, NNC was more prevalent than NNA. The RSCU values pointed out that codons with C or A in the third position were used in higher content than G or T, so the codons NNC and NNA were more, while the synonymous codons NNG and NNT were less. This was related to the asymmetric replication of mitochondria, and the more unstable G bases were gradually replaced by other bases (Clayton, 1982).

Ribosomal and transfer RNA genes

The mitogenomes of the four Sillaginids contain a small subunit rRNA (12S rRNA) and a large subunit rRNA (16S rRNA). Two rRNAs were situated between $tRNA^{Phe}$ and $tRNA^{Leu}$ (UUA) and differentiated by $tRNA^{Val}$ (Table I). 12S rRNA ranged in size from 951 to 954 bp and 16S rRNA ranged in size from 1691 to 1709 bp. Nucleotide composition for the both two rRNAs encoded by the H strand was as follows A% >C% >G% >T% ($P < 0.01$) (Table II). Here A and C contents were dominant, which was also seen in other bony fishes (Zardoya and Meyer, 1997), different from H strand PCGs and entire mt genomes, which implied that the different regions had a different bias.

Like metazoan mt genomes, four Sillaginids mt genomes had 22 tRNAs, including 14 encoded by the H strand and eight encoded by the L strand, interspersed between rRNA and PCGs and ranged in size from 65 ($tRNA^{Cys}$) to 75 ($tRNA^{Lys}$ and $tRNA^{Pro}$) bp (Table I). Two types of serine (UCN and AGY) and leucine (UUR and CUN) were ascertained, including 22 tRNAs and one particular tRNA gene for the other amino acids. The anticodons of the tRNAs were similar to other vertebrates. All tRNAs can form identical cloverleaf secondary structures like other vertebrates except $tRNA^{ser}$ (AGY), which had no DHU arm and looked like a truncated cloverleaf structure that was also the common feature of vertebrates mt genomes. The presumed tRNA secondary structures had a general character with 7 bp in the amino acid arm, 4-5 bp in the TΨC arm, 4-5 bp in the anticodon arm, and 3-4 bp in the DHU arm. There were some non-complementary base pairs in the arms of some tRNAs. The presence of non-complementary base pairs in the arm was the normal occurrence of mt tRNAs that may be corrected by the post-transcriptional editing process (Lavrov *et al*. 2000)

Non-coding regions

Non-coding mtDNA regions consisted of the control region (CR), the origin of the light strand (O_L), and several concise intergenic spacers (1-48 bp long). CR was the longest, typically situated between $tRNA^{Pro}$ and $tRNA^{Phe}$. It consisted of three conserved domains, namely extended termination associated sequence domain (ETAS), central conserved domain (CCD), and conserved sequence blocks (CSB). The size of the CR was highly variable in the four sillaginids (1319, 953, 1141 and 832 bp in *S. muktijoddhai*, *S. mengjialensis*, *S. sihama* and *Sillaginops macrolepis*, respectively). The length variations were mainly due to tandem repeats in the 5′ region of *S. muktijoddhai* and *S. sihama*. The motif length, copy number and consensus pattern were also different. On the contrary, no tandem repeats were found in the CRs of *S. mengjialensis* and *Sillaginops macrolepis*. In *S. muktijoddhai*, 8.4 copy of the repeat unit was found, which was 37 bp long and in *S. sihama*, 5.9 copy of the repeat unit was found in which consensus size was 36 bp (Figure 4). Each repeat unit contained motif sequences for ETAS, including TAS motif TACAT and reverse complement cTAS motif ATGTA, that can form a stable hairpin structure, important for termination of mtDNA replication (Saccone *et al.* 1991). Repeat in the ETAS region was also found in other sillaginids, e.g., *S. sihama* (Siyal *et al*. 2015). In the ETAS region, a single TAS accompanied with cTAS motif sequence in *Sillaginops macrolepis* and two TAS accompanied with cTAS motif sequence in *S. mengjialensis* were identified. Besides, an incomplete TAS was also identified in the ETAS region of each four species (Figure 4). Six conserved sequence blocks CSB F-A were identified (Figure 4) in the central conserved domain. CSB-E had the distinguishing GTGGG-box of teleost CSB-E (Lee *et al*. 1995). CSB-D was the most conserved (96.2% sequence identity) among the four studied species and *S. panijus*, which was the same as in other teleost fishes (Lee *et al*. 1995) due to its crucial contribution in maintaining the exact regulatory activities of CR. CSB-B and CSB-C had 79.2% and 85.1 % sequence similarity, respectively. CSB-A, E, and F were the least conserved (52.3%, 59.7%, and 67.5% sequence identity). CSB-A was followed by a pyrimidine tract (poly T) and next downstream of it, three conserved sequence blocks CSB-1, CSB-2, and

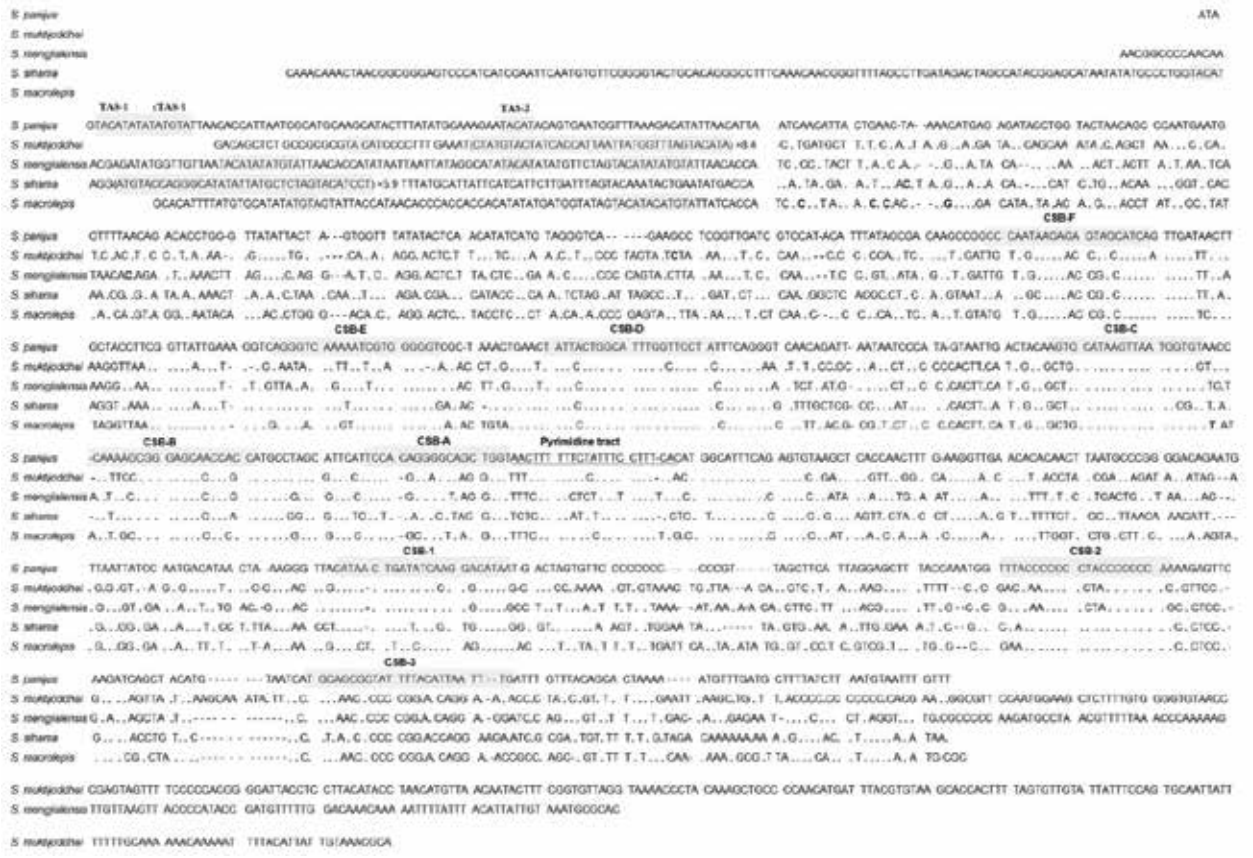

Fig. 4. Alignment of the control regions of *S. muktijoddhai*, *S. mengjialensis*, *S. sihama*, *S. macrolepis* and *S. panijus* mtDNA. The blocks TAS, cTAS; CSB-A, B, C, D, E, F; CSB-1, CSB-2, and CSB-3 are shaded and poly-T underlined. The sequence in parentheses indicates the motif of the tandem repeat, and the Arabic number indicates the copy number. Dots represent similar nucleotides, substitutions are denoted by the corresponding nucleotides, and dashes indicate indels.

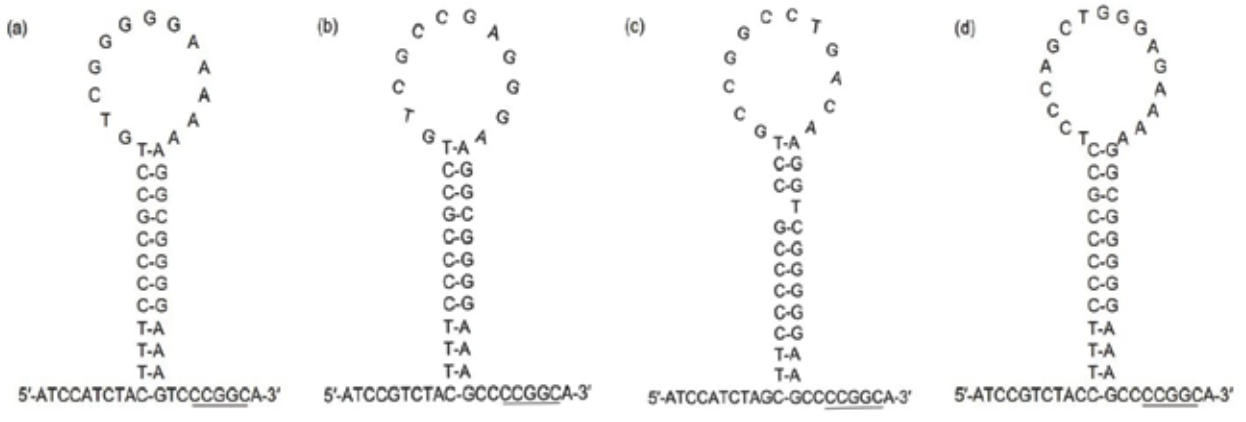

Fig. 5. The OL region stem-loop structures in the mitogenomes of four Sillaginids. The conserved motif sequences were underlined. (a) *S. muktijoddhai* (b) *S. mengjialensis,* (c) *S. sihama* and (d) *S. macrolepis*

CSB-3 were recognized. Like other teleosts, CSB-2 was the most conserved among the three downstream CSBs and distinguished by a poly C stretch separated by TA (Shadel and Clayton, 1997).

O_L was 37-43 bp long, typically situated in the WANCY region between $tRNA^{Asn}$ and $tRNA^{Cys}$. These O_L region sequences can form a stable stem-loop secondary structure, with 24-25 nucleotides in the stem and 12-17 nucleotides in the loop (Figure 5). This stem-loop structure was closely related to the replication of mitochondrial DNA. O_L stem-loop structure also revealed the conserved sequences in the stem than loop region. It had two universal characters of vertebrate O_L like pyrimidine (C, T) rich 5′ side sequence of the stem and conserved motif sequence 5′-GCCGG-3′ at the base of the stem within the $tRNA^{Cys}$ gene that may play a vital role during the transition from RNA to DNA synthesis (Hixson *et al*. 1986).

Sequence variation

On the whole, the highest reserved gene types were rRNA (84.2±0.7%) and tRNA (82.7±0.7%), PCGs showed medium variation (71.6±0.5%), and as desired, the non-coding CR (46.2±2.3%) was the most divergent ($p<0.01$).

Excluding $tRNA^{Pro}$, rRNAs, and the other tRNAs exhibited high levels of sequence identity (sequence similarity > 77%) among the studied species and *S. panijus* (Figure 6A). The conservation of rRNAs and tRNAs was undoubtedly needed to form their secondary structures and for tertiary interactions. This functional constraint would interpret their high sequence similarities despite the high mutation rate in mtDNA. The sequence of the metazoan $tRNA^{Ser}$(AGY) was variable (Garey and Wolstenholme, 1989) may be due to the absence of DHU arm, instead present variable loop, probably reducing structural constraints.

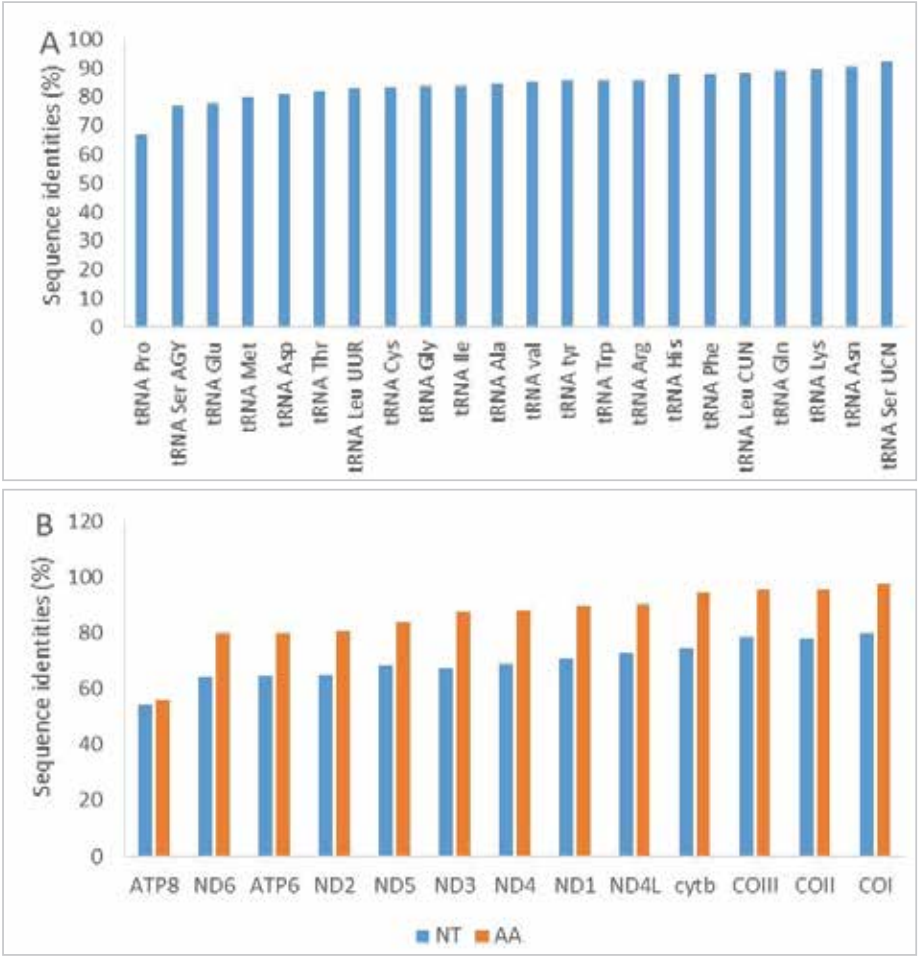

Fig. 6. Sequence differences in mitochondrial genes. Genes were positioned by their sequence identity percentages from small to big (left to right). (A) Sequence identities of 22 tRNA genes. (B) NT% and deduced AA% of 13 PCGs. The AA% positioned genes

Comparison of sequence similarity of the 13 PCGs by both nucleotides (NT %) and their deduced amino acids (AA %) (Figure 6B) showed different divergences. Cytochrome oxidase subunits, COI, II and III had high sequence identity (NT > 78% and AA > 95%), so most conserved among protein-coding genes. ATP synthase subunit 8 (ATP8) and NADH dehydrogenase subunit 6 (ND6) of complex I had low sequence identity (NT < 64.5% and AA < 80%), so most divergent. This divergence among PCGs can be explained by the fact that ATP synthase subunit and most of the complex I encoded from nuclear genes. On the contrary, COI, COII and COIII in complex IV were mitochondrially encoded (Zardoya *et al*. 1995).

To determine the type of selective pressure on 13 PCGs, we calculated the ratios of non-synonymous substitutions rate (Ka) to synonymous substitutions rate (Ks) (Figure 7). Here Ka/Ks ratios for all PCGs were lower than 1. This result implied strong purifying selection, i.e., natural selection in contrast to deleterious mutations with negative selective coefficients (Yang and Bielawski, 2000) on all PCGs of four sillaginids. However, the average Ka/Ks ratios (0.017 to 0.243) also varied significantly among individual genes revealed that various PCGs have different functional constraints (Lynch *et al*. 2006). For example, the ratio was lowest in COI (0.017) among these genes under the highest purifying selective pressure. On the other hand, the highest in ATP8 (0.243) indicated that under the lowest purifying selective pressure and selection pressures were strand-independent.

Phylogenetic analyses

Two phylogenetic trees were constructed using both BI and ML methods to reveal phylogenetic relationships among Sillaginidae species. The first one including 13 Sillaginidae species was based on concatenated nucleotide sequences (10,955 bp) of the 12 mitochondrial H-strand PCGs (Figure 8A). The second one containing 28 Sillaginidae species was based on concatenated nucleotide sequences (4237 bp) of the four mitochondrial genes (Figure 8B). Thus, both phylogenetic methods (BI and ML) for two phylogenetic trees produced similar topology with high bootstrap support and Bayesian inference value at the node.

In the 12 H-strand PCGs phylogenetic tree, there were two distinct clades for the genus *Sillaginopsis* and *Sillago* within Sillaginidae. Inside the clade of the genus *Sillago*, *Sillaginops macrolepis* appeared as sister taxa with *S. asiatica* and *S. japonica*, which strongly supported its generic status as *Sillago*. Although based on morphology, *Sillaginops macrolepis* was situated outside the clade of the genus *Sillago* and considered including a different genus as *Sillaginops* (Kaga, 2013). *S. sihama* was clustered first with closely related species *S. sihama* (KR363150) and after that *S. muktijoddhai*, *S. mengjialensis* with *S. indica*. This result supported their taxonomic position according to morphology.

In four mt genes phylogenetic tree, within Sillaginidae at first monotypic genera *Sillaginopsis*, *Sillaginopodys* were formed monophyletic clade which agreed with Kaga (2013).

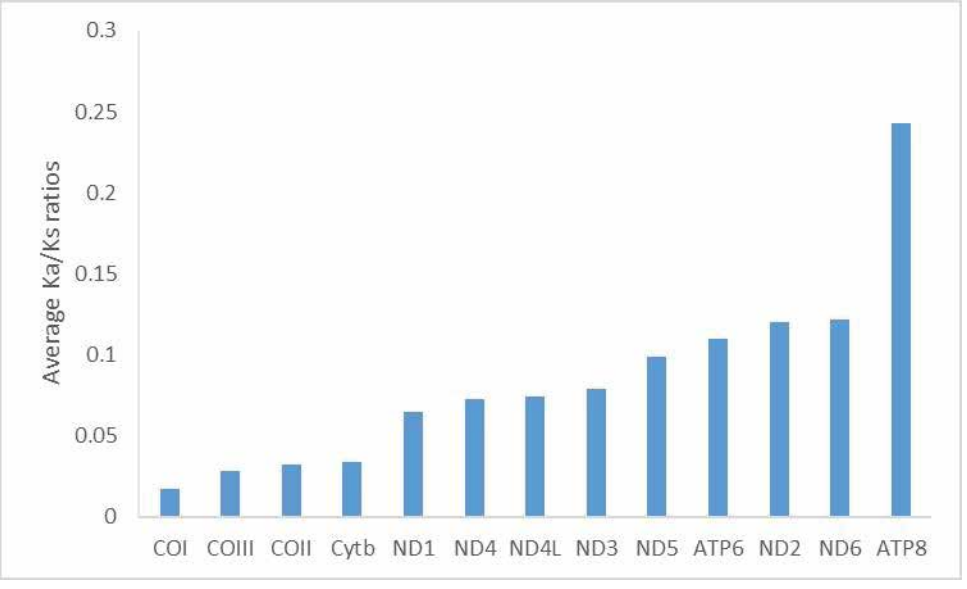

Fig. 7. Average Ka/Ks ratios of the 13 PCGs, where Ka, non-synonymous substitutions rate and Ks, synonymous substitutions rate

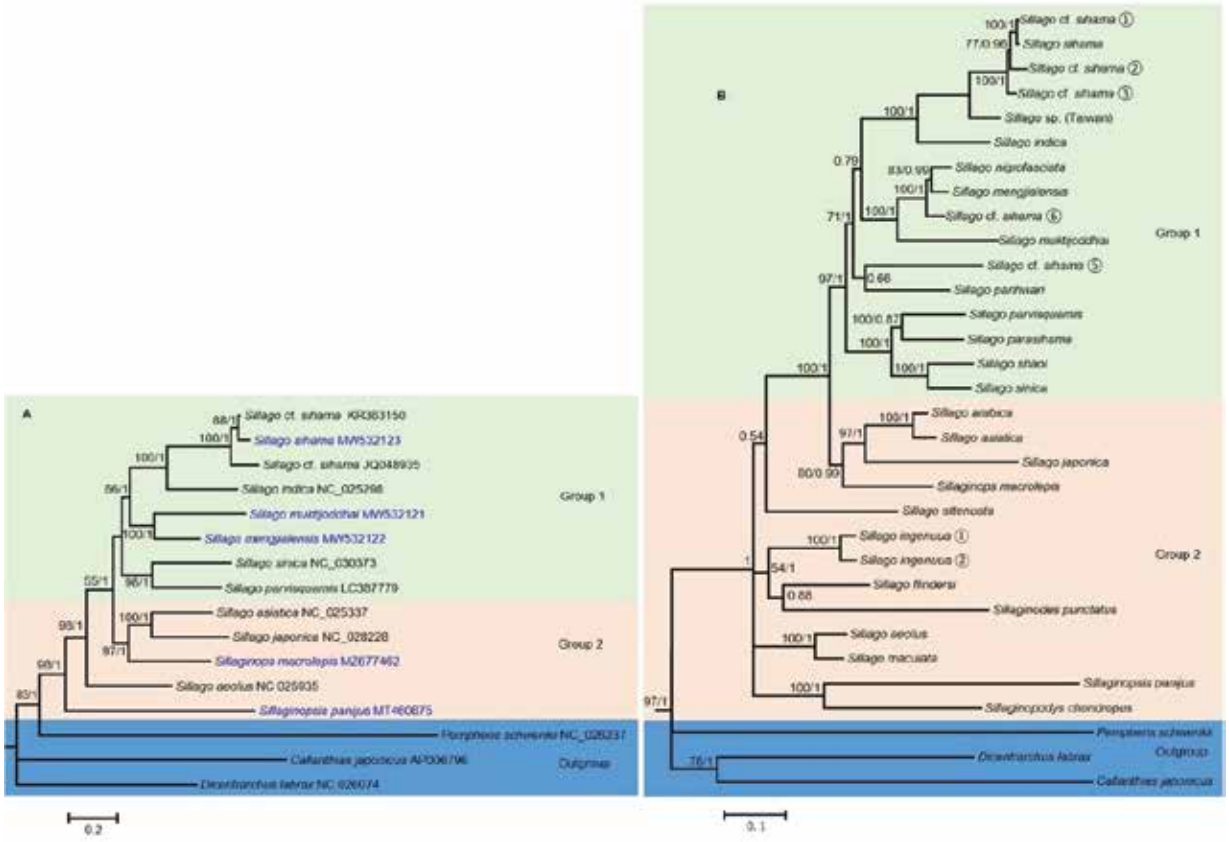

Fig. 8. Phylogenetic tree using BI and ML method. (A) Among 13 Sillaginidae species based on concatenated sequences of the 12 mt PCGs excluding ND6. (B) Among 28 Sillaginidae species based on concatenated sequences of 4 mt genes. The maximum likelihood topology was shown that is similar to the result of Bayesian. Bootstrap support values/ Bayesian posterior probabilities were displayed at branch nodes. Three species of the suborder Percoidei, *Callanthias japonicas, Dicentrarchus labrax* and *Pempheris schwenkii,* were selected as outgroup species.

However, unlike Kaga (2013), monotypic genera *Sillaginodes* (*S. punctata*) and *Sillaginops* (*S. macrolepis*) were inserted within the genus *Sillago*.

At last, the *S. sihama* complex was found as a sister taxon with the dichotomy, including *S. parvisquamis*, *S. parasihama*, *S. shaoi* and *S. sinica*. Species within the *S. sihama* complex were divided into two clades. The *S. panhwari* and Pakistan *S. sihama* ⑤ formed a single clade. Another clade was again divided into two. *S. mengjialensis* was first clustered with *S. nigrofasciata*, then together with Pakistan *S. sihama* ⑥, after that with *S. muktijoddhai* at the basal. In another clade, *S. sihama*, Pakistan *S. sihama* ①, Thai *S. sihama* ② and Chinese *S. sihama* ③ were first associated with *Sillago* sp. (Taiwan), and then with *S. indica* at the basal.

The present four mt genes phylogenetic tree was slightly different from our previous study (Cheng *et al*. 2020). For example, unlike the present study, subgenus *Sillaginopodys* (*S. chondropus*) was inserted into the subgenus *Parasillago*. This difference with Cheng *et al*. (2020) maybe because of including all key groups and although lacking nuclear data during the present study. Key groups included species representing three genera (*Sillaginodes*, *Sillaginopsis*, *Sillago*) and three subgenera of the genus *Sillago* (*Sillaginopodys*, *Sillago* and *Parasillago*) by McKay (1992) and five genera (*Sillaginodes*, *Sillaginopsis*, *Sillaginopodys*, *Sillaginops* and *Sillago*) system of Kaga (2013). Moreover, the present four mt genes phylogenetic tree was congruent with Cheng *et al*. (2020) to divide the species into two groups based on swimbladder development. While group 2 species have less developed swimbladder. On the contrary, group 1 species have well-developed swimbladder, similar to subgenus *Sillago* (McKay, 1992).

Table. S3. K2P distance of 13 Sillaginidae species based on 12 PCGs sequences

	S. panijus	*S. muktijoddhai*	*S. mengjialensis*	*S. sihama*	*S. macrolepis*	*S. aeolus*	*S. asiatica*	*S. japonica*	*S. indica*	*S. sinica*	*S.* cf. *sihama* KR363150	*S.* cf. *sihama* NC016672	*S. parvisquamis*
S. panijus													
S. muktijoddhai	0.335												
S. mengjialensis	0.320	0.225											
S. sihama	0.331	0.275	0.244										
S. macrolepis	0.316	0.264	0.236	0.261									
S. aeolus	0.311	0.274	0.254	0.272	0.263								
S. asiatica	0.327	0.273	0.240	0.268	0.236	0.266							
S. japonica	0.328	0.271	0.253	0.274	0.248	0.272	0.230						
S. indica	0.336	0.266	0.240	0.232	0.251	0.277	0.262	0.265					
S. sinica	0.345	0.281	0.249	0.276	0.270	0.215	0.275	0.281	0.280				
S. cf. *sihama* KR363150	0.330	0.277	0.241	0.041	0.261	0.273	0.266	0.278	0.233	0.280			
S. cf. *sihama* NC016672	0.335	0.294	0.260	0.117	0.273	0.283	0.261	0.283	0.253	0.293	0.088		
S. parvisquamis	0.334	0.262	0.235	0.266	0.256	0.276	0.260	0.265	0.262	0.233	0.265	0.282	

In addition, based on the Kimura-2-parameter model, the net genetic distances of 12 PCGs of 12 Sillaginidae species were calculated (Table S3). The results showed that the genetic distances were between 0.041 and 0.345. The genetic distance between *S. panijus* and *S. sinica* was the largest at 0.345, and the genetic distance between true *S. sihama* and misidentified *S. sihama* (KR363150) was the smallest at 0.041. Besides, the genetic distance was 0.088 between two misidentified *S. sihama* (NC016672, KR363150). This distance revealed that they had no intraspecies relationship instead showed interspecies relationship. This phenomenon may be due to the morphological misidentification of *S. sihama*, as there were at least eight cryptic lineages in the *S. sihama* complex (Cheng *et al*. 2020).

Moreover, based on the Kimura-2-parameter model, the net genetic distances of 4 mt genes of 28 Sillaginidae species were calculated and showed that the genetic distances ranged from 0.025 (between *S. sihama* and *S. sihama* ③) to 0.249 (between *S. punctata* and *S. panijus*). Moreover, *S. sihama* (Bangladesh) has an intraspecies relationship with Pakistan *S. sihama* ① (genetic distance, 0.003).

Divergence time analysis based on 12 mt PCGs showed that Sillaginidae species diverged around 61.30 million years ago (Mya) (Figure 9A). In the family Sillaginidae, the divergence time between *S. panijus* and remaining Sillaginidae species of the genus *Sillago* was the earliest (51.82 Mya). Thus upper Eocene at 51.82 Mya was determined as the divergence period between the genera *Sillaginopsis* and *Sillago*. Then after 13.63 million years' differentiation started within the genus *Sillago* and most of the species of this genus diverged during the lower Eocene to Pliocene (38.19 to 5.36 Mya). However, *S. sihama* (KR363150) and *S. sihama* had the latest differentiation time (1.44 Mya), differentiated during Pleistocene (11,700 yrs to 2.588 Mya) in the Quaternary period.

Divergence time analysis based on 4 mt genes showed that Sillaginidae species diverged around 61.42 Mya (Figure 9B). The divergence within the family Sillaginidae was started 42.93 Mya, and most of the species were differentiated during the late Paleogene (23.03 to 44.27 Mya) and Neogene (2.588 to 23.03 Mya). However, *S. sihama* ② and *S. sihama* had the latest differentiation time (2.12 Mya), diverged during Pleistocene (11,700 yrs to 2.588 Mya) in the Quaternary period.

Same as the present study, Xiao (2015) showed that the origin of the genus *Sillago* probably dated back to the late Eocene based on 12 H-strand mitochondrial PCGs of six *Sillago* species. However, unlike the present study, the root of the family Sillaginidae was in the middle Eocene based on four mitochondrial genes and a single nuclear gene of 24 Sillaginidae species (excluding monotypic genera *Sillaginops* and *Sillaginopsis*) (Cheng *et al*. 2020). Whereas the present study included the most primitive species *Sillaginopsis panijus*

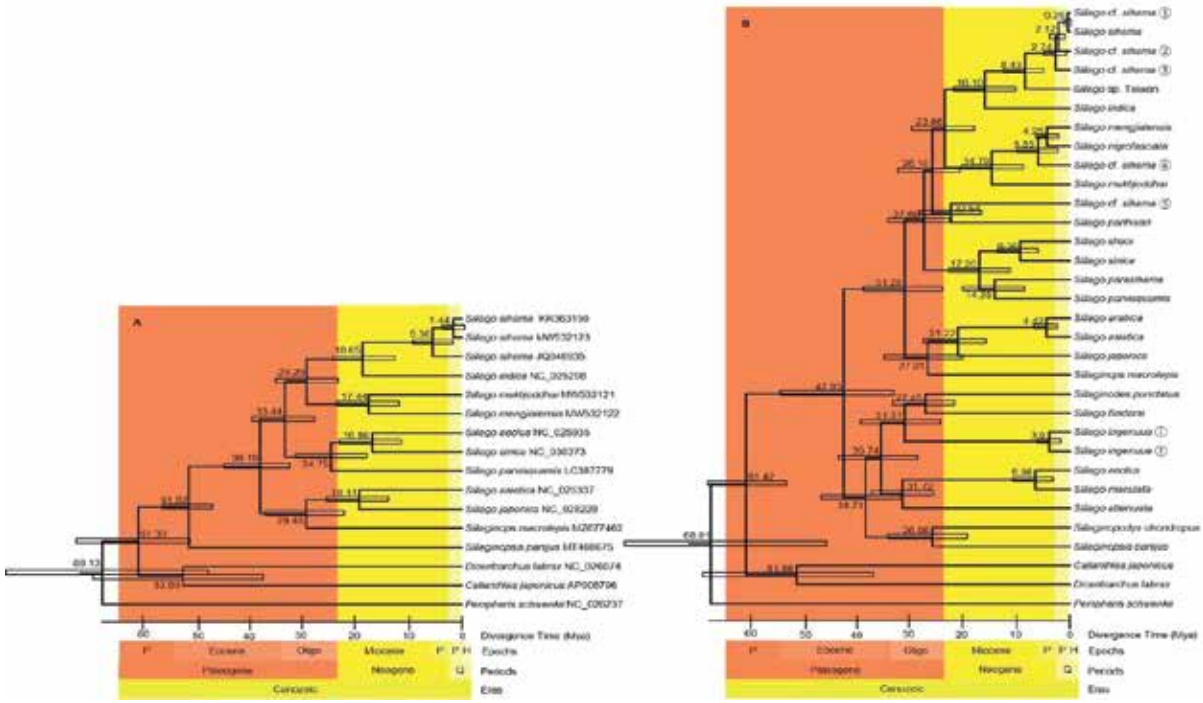

Fig. 9. BEAST phylogeny of the family Sillaginidae. (A) Among 12 species using concatenated sequences of 12 PCGs excluding ND6. (B) Among 28 species using concatenated sequences of 4 mt genes. Numbers near the nodes indicated the estimated divergence time. Bars around each node represent 95% highest posterior density (HPD)

(without swimbladder), therefore, Sillaginidae originated at Paleocene, about 17.9 million years earlier than Cheng *et al*. (2020).

Conclusion

Complete mt genomes of all four sillaginids contained 37 mitochondrial genes (13 protein-coding genes (PCGs), 2 ribosomal RNA (rRNA), 22 transfer RNA (tRNA)), and a control region, where gene arrangement and distribution were canonically identical and consistent with other vertebrates. On the whole, rRNAs and tRNAs were the highest conserved genes, then PCGs and control regions were the lowest conserved. For PCGs, cytochrome oxidase subunit I was the most conserved, while ATP synthase subunit 8 was the most divergent. In the phylogenetic analysis of the 12 H-strand protein-coding genes and 4 mitochondrial genes based on Maximum likelihood and Bayesian inference methods, *S. muktijoddhai*, *S. mengjialensis*, and *S. sihama* were placed in the appropriate position with relative species of the genus *Sillago*. The present phylogeny supported their taxonomic position according to morphology. In comparison, *Sillaginops*. *macrolepis* was situated in the cluster of genus *Sillago* that contrasted to morphological phylogeny. Further study can be done using whole genome survey sequencing to firmly establish the phylogenetic status of the family Sillaginidae.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2023YFD2401903) and the National Natural Science Foundation of China (41976083).

Authors are thankful to Zhi Chen, Chuangeng Cai and Zhengsen Yu for some technical support during the laboratory experiment.

References

Benson G (1999), Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res*. **27**(2): 573-580. https://tandem.bu.edu/trf/trf.html

Betancur RR, Broughton RE, Wiley EO, Carpenter K, López JA, Li C, Holcroft NI, Arcila D, Sanciangco M, Cureton JC, Zhang F, Buser T, Campbell MA, Ballesteros JA, Roa-Varon A, Willis S, Borden WC, Rowley T, Reneau PC and Ortí G (2013), The tree of life and a new classification of bony fishes, *PLoS Curr*. **5**: 1-41. https://doi.org/10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288

Boore JL (1999), Animal mitochondrial genomes, *Nucleic Acids Res*. **27**: 1767-1780. https://doi.org/10.1093/nar/27.8.1767

Cai SS, Song N, Zhang LY and Gao TX (2014), The complete mitochondrial genome of *Sillago aeolus* (Perciformes: Sillaginidae), *Mitochondrial DNA A DNA Mapp Seq Anal.* **27**(2): 1499-1500. https://doi.org/10.3109/19401736.2014.953110.

Cheng J, Ma GQ, Song N and Gao TX (2012), Complete mitochondrial genome sequence of bighead croaker *Collichthys niveatus* (Perciformes: Sciaenidae), A mitogenomic perspective on the phylogenetic relationships of Pseudosciaeniae, *Gene*. **492**: 210-23.

Cheng J, Xiao JG, Song N, Saha S, Qin JG, Nomura H, Panhwar SK, Farooq N, Shao KT and Gao TX (2020), Molecular phylogeny reveals cryptic diversity and swim bladder evolution of Sillaginidae fishes (Perciformes) across the Indo-West Pacific Ocean, *Divers Distrib*. **27**: 82-94. https://doi.org/10.1111/ddi.13171

Clayton DA (1982), Replication of animal mitochondrial DNA, *Cell*. **28**: 693-705. https://doi.org/10.1016/0092-8674(82)90049-6.

Coucheron DH, Nymark M, Breines R, Karlsen BO, Andreassen M, Jørgensen TE and Johansen SD (2011), Characterization of mitochondrial mRNAs in codfish reveals unique features compared to mammals, *Curr Genet*. **57**: 213-222. https://doi.org/10.1007/s00294-011-0338-2

Darriba D, Taboada GL, Doallo R and Posada D (2012), jModelTest 2: more models, new heuristics and parallel computing, *Nature Methods*. **9**: 772. https://doi.org/10.1038/nmeth.2109

Drummond AJ, Suchard MA, Xie D and Rambaut A (2012), Bayesian phylogenetics with BEAUti and the BEAST 1.7, *Mol Biol Evol*. **29**: 1969-1973. https://doi.org/10.1093/molbev/mss075

Garey JR and Wolstenholme DR (1989), Platyhelminth mitochondrial DNA: evidence for early evolutionary origin of a tRNA ser AGN that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons, *J Mol Evol*. **28**: 374-387. https://doi.org/10.1007/BF02603072.

Golani D, Fricke R and Tikochinski Y (2014), *Sillago suezensis*, a new whiting from the northern Red Sea, and status of *Sillago erythraea* Cuvier (Teleostei: Sillaginidae), *J Nat Hist.* **48** (7-8): 413-428. https://doi.org/10.1080/00222933.2013.800609

Grant JR and Stothard P (2008), The CGView Server: A comparative genomics tool for circular genomes, *Nucleic Acids Res*. **36**: 181-184.

Hassanin A, LéGer N and Deutsch J (2005), Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences, *Syst Biol*. **54**: 277-298. https://doi.org/10.1080/10635150590947843.

Harrison T, Krigbaum J and Manser J (2006), Primate biogeography and ecology on the Sunda Shelf Islands: A paleontological and zooarchaeological perspective, *Primate Biogeogr*. **111**: 331-372.

Hixson JE, Wong TW and Clayton DA (1986), Both the conserved stem-loop and divergent 5′-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication, *J Biol Chem*. **261**: 2384-2390.

Iwasaki W, Fukunaga T, Isagozawam R, Yamada K, Maeda Y, Satoh TP, Sado T, Mabuchi K, Takeshima H, Miya M and Nishida M (2013), MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline, *Mol Biol Evol*. **30**: 2531-2540. http://mitofish.aori.u-tokyo.ac.jp/annotation/input.html

Kaga T (2013), Phylogenetic systematics of the family Sillaginidae (Percomorpha: order Perciformes), *Zootaxa*. **3642**(1): 1-105. https://doi.org/10. 11646/zootaxa.3642.1.1

Lakra WS, Verma MS, Goswami M, Lal KK, Mohindra V, Punia P, Gopalakrishnan A, Singh KV, Ward RD and Hebert P (2011), DNA barcoding Indian marine fishes, *Mol Ecol Resour*. **11**(1): 60-71.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007), Clustal W and Clustal X version 2.0, *Bioinform.* **23**: 2947-2948. https://doi.org/10.1093/bioinformatics/btm404

Lavrov DV, Brown WM and Boore JL (2000), A novel type of RNA editing occurs in the mitochondrial trnas of the centipede *Lithobius forficatus*, *Proc Natl Acad Sci. USA*. **97**: 13738-13742.

Lee WJ, Conroy J, Howell WH and Kocher TD (1995), Structure and evolution of teleost mitochondrial control regions, *J Mol Evol*. **41**: 54-66.

Liu D, Guo Y, Wang Z and Liu C (2012), Phylogenetics inferred from mitogenome and control region of silver sillago, *Sillago sihama*, *Mitochondrial DNA* **23**(4): 255-263. https://doi.org/10.3109/ 19401736.2012.674118

Lowe TM and Chan PP (2016), tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes, *Nucleic Acids Res*. **44**(W1): W54-7. https://doi.org/10.1093/nar/gkw413

Lynch M, Koskella B and Schaack S (2006), Mutation pressure and the evolution of organelle genomic architecture, *Science* **311**: 1727-1730.

McKay RJ (1992), FAO Species Catalogue. Vol. 14. Sillaginid fishes of the world (family Sillaginidae). An annotated and illustrated catalogue of the *Sillago*, smelt or Indo-Pacific whiting species known to date. Rome: FAO. FAO Fish. *Synop*. **125**(14): 87p.

Miya M and Nishida M (2000), Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion, *Mol Phylogenet Evol*. **17**: 437-455.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM and Nishida M (2003), Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences, *Mol Phylogenet Evol*. **26**: 121-138.

Miya M, Sato Y, Fukunaga T, Sado T, Poulsen J Y, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M and Iwasaki W (2015), MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species, *R Soc Open Sci*. **2**: 150088. http://doi.org/10.1098/rsos.150088

Moritz C, Dowling T and Brown W (1987), Evolution of animal mitochondrial DNA: relevance for population biology and systematics, *Annu Rev Ecol Syst*. **18**: 269-292. https://doi.org/ 10.1146/ annurev.es-.18.110187. 001413.

Near TJ, Dornburg A, Eytan RI, Keck BP, Smith WL, Kuhn KL, Moore JA, Price SA, Burbrink FT, Friedman M and Wainwright PC (2013), Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes, *Proc Natl Acad Sci U S A*. **110**: 12738-12743. https://doi.org/10.1073/pnas.1304661110

Niu S, Wu R, Liu Y and Wang X (2016), The High throughput sequencing of *Sillago japonica* mitochondrial genome reveals the phylogenetic position within the genus *Sillago*, *Mitochondrial DNA Part A* **27**(5): 3815-3816. https://doi.org/ 10.3109/ 19401736. 2015.1082090

Ojala D, Montoya J and Attardi G (1981), tRNA punctuation model of RNA processing in human mitochondria, *Nature* **290**: 470-474. https://doi.org/10. 1038/ 290470a0. PubMed: 7219536.

Perna NT and Kocher TD (1995), Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes, *J Mol Evol*. **41**: 353-358. https://doi.org/ 10.1007/BF0 1215182. PubMed: 7563121.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA and Huelsenbeck JP (2012), MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space, *Syst Biol*. **61**(3): 539-542.

Saccone C, Pesole G and Sbisá E (1991), The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern, *J Mol Evol*. **33**: 83-91. https://doi.org/10.1007/BF02100199.

Saha S, Bogorodsky SV, Baki MA, Gao TX, McKay RJ, Alpermann TJ and Song N (2024), Assessment of the diversity of the family Sillaginidae in the Indian Ocean, with emphasis on the taxonomic identity of *Sillago sihama*, *Zootaxa* (accepted).

Saha S, Song N, Yu ZS, Baki MA, McKay RJ, Qin JG and Gao TX (2022), Descriptions of two new species, *Sillago muktijoddhai* sp. nov. and *Sillago mengjialensis* sp. nov. (Perciformes: Sillaginidae) from the Bay of Bengal, Bangladesh, *Fishes*. **7**(93): 1-16. https://-doi.org/10.3390/fishes7030093

Saha S, Song N, Baki MA, Yang TY and Gao TX (2021), Characterization and phylogenetic analyses of the complete mitochondrial genome of *Sillaginopsis panijus* (Perciformes: Sillaginidae), *Mitochondrial DNA Part B*. **6**(11): 3202-3203. https://doi.org/ 10.1080/23802359.2021.1989339

Sambrook J, Fritsch EF and Maniatis T (1989), Molecular cloning: a laboratory manual 2nd Ed. New York: Cold Spring Harbor Laboratory Press.

Satoh TP, Miya M, Mabuchi K and Nishida M (2016), Structure and variation of the mitochondrial genome of fishes, *BMC Genomics* **17**: 719. https://doi.org/10. 1186/s12864-016-3054-y

Shadel GS and Clayton DA (1997), Mitochondrial DNA maintenance in vertebrates, *Annu Rev Biochem*. **66**: 409-435. https://doi. org/10.1146/ annurev.biochem-.66.1.409

Siyal FK, Xiao J, Song N and Gao TX (2015), The complete mitochondrial genome of *Sillago sihama* (Perciformes: Sillaginidae), *Mitochondrial DNA A DNA Mapp Seq Anal.* **27**(4): 2933-2934. https://doi. org/10. 3109/ 19401736. 2015.1060447

Stepien CA and Kocher TD (1997), Molecules and morphology in studies of fish evolution, *Mol Sys Fishes*. 1-11.

Suzuki T, Senou H and Hosokawa M (2001), First record of the sillaginid fish, *Sillago macrolepis* from Japan, *I.O.P. Diving News* **12** (2): 2-4.

Swofford DL (2002), PAUP* 4.0: Phylogenetic analysis using parsimony. Sinauer Associates Inc.

Tamura K, Stecher G, Peterson D, Filipski A and Kumar S (2013), MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, *Mol Biol and Evol*. **30**: 2725-2729.

Tanaka M and Ozawa T (1994), Strand asymmetry in human mitochondrial DNA mutations, *Genomics* **22**: 327-335. https://doi.org/10.1006/geno.1994.1391

Ward RD, Zemlak TS, Innes BH, Last PR and Hebert PD (2005), DNA barcoding Australia's fish species, *Philos Trans R Soc B*. **360** (1462)**:** 1847-1857.

Wolstenholme DR (1992), Animal mitochondrial DNA: structure and evolution, *Int Rev Cytol*. **141**: 173-216. https://doi.org/10.1016/S0074-7696(08)62066-5

Xiao JG, Song N, Gao TX and McKay RJ (2014a), The complete mitochondrial genome of *Sillago indica* (Perciformes: Sillaginidae), *Mitochondrial DNA A DNA Mapp Seq Anal.* **27**(2): 1445-1446. http://dx.-doi.org/10.3109/19401736.2014.953085.

Xiao JG, Song N, Gao TX and Zhao Y (2014b), The complete mitochondrial genome of *Sillago asiatica* (Perciformes: Sillaginidae), *Mitochondrial DNA A DNA Mapp Seq Anal.* **27**(3): 1644-1645. http://dx.-doi.org/10.3109/19401736.2014.958709

Xiao JG (2015), The complete mitochondrial genomes and phylogenetic analysis of *Sillago* species, M.Sc., Thesis, Ocean University of China, Qingdao, China.

Yang Z and Bielawski JP (2000), Statistical methods for detecting molecular adaptation, *Trends Ecol Evol*. **15**: 496-503.

Zardoya R and Meyer A (1997), The complete DNA sequence of the mitochondrial genome of a living fossil, the coelacanth (*Latimeria chalumnae*), *Genetics*. **146**: 995-1010.

Zardoya R, Garrido-Pertierra A and Bautista JM (1995), The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*, *J Mol Evol*. **41**: 942-951.

Zhang H, Zhang Y, Zhang X, Song N and Gao TX (2013), Special structure of mitochondrial DNA control region and phylogenetic relationship among individuals of the black rockfish, *Sebastes schlegelii*, *Mitochondrial DNA* **24**(2): 151-157.