RESEARCH ARTICLE

The first complete mitochondrial genome from *Bostrychus* genus (*Bostrychus sinensis*) and partitioned Bayesian analysis of Eleotridae fish phylogeny

TAO WEI, XIAO XIAO JIN and TIAN JUN XU*

Laboratory for Marine Living Resources and Molecular Engineering, College of Marine Science, Zhejiang Ocean University, 316000 Zhoushan, People's Republic of China

Abstract

To understand the phylogenetic position of *Bostrychus sinensis* in Eleotridae and the phylogenetic relationships of the family, we determined the nucleotide sequence of the mitochondrial (mt) genome of *Bostrychus sinensis*. It is the first complete mitochondrial genome sequence of *Bostrychus* genus. The entire mtDNA sequence was 16508 bp in length with a standard set of 13 protein-coding genes, 22 transfer RNA genes (tRNAs), two ribosomal RNA genes (rRNAs) and a noncoding control region. The mitochondrial genome of *B. sinensis* had common features with those of other bony fishes with respect to gene arrangement, base composition, and tRNA structures. Phylogenetic hypotheses within Eleotridae fish have been controversial at the genus level. We used the mitochondrial cytochrome b (*cytb*) gene sequence to examine phylogenetic relationships of Eleotridae by using partitioned Bayesian method. When the specific models and parameter estimates were presumed for partitioning the total data, the harmonic mean –lnL was improved. The phylogenetic analysis supported the monophyly of *Hypseleotris* and *Gobiomorphs*. In addition, the *Bostrychus* were most closely related to *Ophiocara*, and the *Philypnodon* is also the sister to *Microphlypnus*, based on the current datasets. Further, extensive taxonomic sampling and more molecular information are needed to confirm the phylogenetic relationships in Eleotridae.

[Wei T., Jin X. X. and Xu T. J. 2013 The first complete mitochondrial genome from *Bostrychus* genus (*Bostrychus sinensis*) and partitioned Bayesian analysis of Eleotridae fish phylogeny. *J. Genet.* **92**, 247–257]

Introduction

Eleotridae is an important and diverse family of fishes and includes 35 genera and about 150 species (FishBase 2000, http://www.fishbase.org/home.htm.). Most Eleotridae fish inhabit the tropical, subtropical coastal and estuarine waters in the world. They are a major component of many fish fauna; however, the species are difficult to sample and the biology of the group is not well understood (Thacker 2003). The Eleotridae fish are generally small in size in comparison with other species, and some species have even losses and simplifications about some aspects of morphology (Akihito Iwata *et al*. 2000). In addition, lack of phylogenetic studies on the different genera has also led to limite understanding of evolutionary relationships of genera within the family.

Currently, molecular phylogenetic studies have succeeded in solving some phylogenetic questions and persistent controversies among teleosts, for example in the Cyprinidae (He *et al*. 2008). The evolutionary relationships of higher teleosts have also explored by using mitogenomic sequence (Miya *et al*. 2003). Despite some studies about morphology and molecular phylogeny in Eleotridae, the phylogenetic relationships have not been well resolved in this family, leading to conflicting phylogenetic hypothesis (Thacker 2003; Thacker and Hardman 2005). The genus *Hypseleotris* was considered to be monophyletic and a sister to the pair of *Eleotris* and *Philypnodon.* The monophyly of the *Gobiomorphs* group was not supported by molecular data, and *Mogurda* was found to be a sister to *Opieleotris. Bostrychus* had also been closely related to *Ophiocara* (Wang *et al*. 2001). However, Thacker (2003) pointed out that *Hypseleotris* and *Philypnodon* were sisters, excluding *Eleotris*. It was different from Wang *et al.*'s (2001) hypothesis to some extent. In addition, the *Mogurda* have been treated as being closely related to *Opieleotris.* Thacker and Hardman (2005) suggested the *Hypseletris* was sister to *Calumia.* Meanwhile, *Bostrychus* was found to have a close relationship with *Ophiocara,* which was consistent with Wang *et al*.'s (2001) conclusion.

*For correspondence. E-mail: tianjunxu@163.com.

Keywords. Eleotridae; mitochondrial genome; cytochrome b; partitioned Bayesian analysis; phylogenetic analysis; *Bostrychus sinensis*.

Previous molecular approaches to Eleotridae phylogenetic reconstruction based on molecular data were simple. When using a single nucleotide substitution model and associated parameters, the ability to explain the data sets which are multiple genes displaying different models of evolution are limited (Brandley *et al*. 2005). In addition, the use of a single model with data may result in mismodelling and significant systematic error in different evolving subsets (Brandley *et al*. 2005; Cheng *et al*. 2012). A solution to these problems would be to apply specific modes and associated parameter estimates to each partition and integrate this into a single maximum likelihood (ML) tree search (Yang 1996). Methods of implementing partitioned analysis using Bayesian/Markov chain Monte Carlo (MCMC) methods are available (Huelsenbeck and Ronquist 2001; Nylander *et al*. 2004). As it accurately models the data, the use of partition-specific modelling can reduce systematic error, and produce more accurate posterior probability estimates and better likelihood values (Cheng *et al*. 2012).

In an attempt to resolve the interrelationships of Eleotridae, we determine the mitogenomic sequence data from *B. sinensis* for the first time in the genus. *B. sinensis* is a small sized camping-cave food fish, belongs to the Eleotridae family (Ip *et al*. 2001). Due to its good taste and high nutritive value, *B. sinensis* has been a promising high quality breeding fish, but few molecular studies have been conducted on the species so far. More studies are needed to determine the phylogenic position of the species within the Eleotridae family. In the present study, we determined the mitochondrial genome sequence of *B. sinensis*. In addition, the *Cytb* gene was used as a genetic marker to infer the phylogenetic relationships within Eleotridae species.

Materials and methods

Sample collection and DNA extraction

The specimens of *B. sinensis* were collected from the Zhoushan fishing grounds, East China Sea (Zhejiang Province, China). Taxonomic status of the fish was identified by morphology. Total genomic DNA was extracted from the muscle tissue by the standard phenol–chloroform method (Sambrook and Russell 2001) and visualized on 1.0% agarose gel, and then it was stored at 4°C.

PCR amplification and sequencing

A total of 10 sets of primers (table 1) were used for amplifying contiguous, overlapping segments of the complete mitochondrial genome of *B. sinensis*. The PCR reactions were conducted in 50 μL reaction mixtures containing 5.0 μL of a 10× *Taq* Plus polymerase buffer, 0.2 mM of dNTP, 0.2 μM of the forward and reverse primers, 2 U of *Taq* Plus DNA polymerase with proof-reading characteristic (Tiangen, Beijing, China), and 1 μL of DNA template. The PCR reactions consisted of predenaturation at 94°C for 4 min; 35 cycles of denaturation at 94°C for 50 s, annealing at 60°C for 60 s, extension at 72°C for 2 min; and final extension at 72°C for 10 min. The PCR was performed on a Veriti thermocycler (Applied Biosystems, Foster City, USA). The PCR product was electrophoresed on a 1% agarose gel to check integrity, and then it was purified by using the PCR purification kit (Tiangen, Beijing, China). The purified fragments were ligated into PMD-19T vectors (Takara, Dalian, China) and cloned in TOP10 cells (Tiangen, Beijing, China) according to the standard protocol. Positive clones were screened via PCR with M13+/− primers. At least three clones were sequenced from both strands on an ABI3730XL Automated Sequencer with (Applied Biosystems, Foster City, USA) M13 primer.

Table 1. PCR primers in the analysis of *B. sinensis*.

Primer	Sequences (5′–3′)
Bosi-F1	ACTAAAGCATAACACTGAAGAT
Bosi-R1	TTCATTTCTCTTTCAGCTTTCC
Bosi-F2	AAGGGGAGGCAAGTCGTAAC
Bosi-R2	TGTTTAAAGGGCTTAGGTCT
Bosi-F3	GTTTGTTCAACGATTAAAGTCCTACG
Bosi-R3	ATAAAGAAAATTATTACAAAGGCATG
Bosi-F4	TGCCTTTGTAATAATTTTCTT
Bosi-R4	TAGTTGTGAGGGATGTGCCAT
Bosi-F5	CGAGAAAGGGAGGACTTGAACC
Bosi-R5	TATCTTTCCTTGGAATTTAACC
Bosi-F6	ATACCACATAGTAGACCCCAG
Bosi-R6	GACTTTAACCACGAGCTTTTG
Bosi-F7	GGCTCAAAAGCTCGTGGTTAAAG
Bosi-R7	TGCACCAAGAGTTTTTGGTTCCT
Bosi-F8	GGTCTTAGGAACCAAAAACTCTT
Bosi-R8	AATAACAACGGTGGTTTTTCAAG
Bosi-F9	GTG ACTTGA AAA ACCACCGTT
Bosi-R9	CTCCATCTCCGGTTTACA A
Bosi-F10	AGCACCGGTCTTGTAAACCG
Bosi-R10	GGGCTCATCTTAACATCTTCA

Sequence analysis

The SeqMan program (DNAstar, Madison, USA) was used for editing and assembling the contiguous, overlapping sequence. The majority of tRNAs were identified by using tRNAscan-SE 1.21 (http://lowelab.ucsc.edu/tRNAscan-SE/), with default search mode, using vertebrate mitochondrial genetic code as tRNA structure prediction and mitochondrial/chloroplast DNA for the source. The remaining rRNA genes were determined by homology, secondary structure and specific anti-codons (Gutell *et al*. 1993). The DOGMA (http://ogdraw.mpimp-golm.mpg.de/cgi-bin/ogdraw.pl) was used to identify positions of protein-coding genes and rRNAs with default settings (Wyman *et al*. 2004). Subsequently, gene predictions were refined by comparing with other known mitochondrial genome from Eleotridae fish. Repeat sequence patterns in the noncoding region were determined by using Tandem Repeats

Finder software server (http://tandem.bu.edu/trf/trf.basic.submit.html). The codon usage and base composition were calculated by MEGA 4.0 software (Kumar *et al*. 2008). Putative origin of L-strand replication and control region were determined by sequence homology and proposed secondary structure.

Phylogenetic analysis

To understand the phylogenetic relationship within Eleotridae, the data set (*Cytb*) was collected from the sequence available in GenBank. The *Odontobutis obscura* was chosen as an outgroup to root the tree. Multiple alignments of the *Cytb* sequences from 30 Eleotridae species were implemented by using the Clustal W (Thompson *et al*. 1994), and then adjustments were made manually. Phylogenetic trees were performed on MrBayes 3.1 by using partitioned Bayesian method (Huelsenbeck and Ronquist 2001). Here we set three partitions, P1 (all codon positions together), P2 (first and second codon positions; third codon position), and P3 (first codon position; second codon position; third codon position). We determined the best-fit model of molecular evolution for each partition by the akaike information criterion (AIC) using the jModeltest (Posada 2008) program. A smaller AIC value demonstrated the model better fitted to the data. The jModeltest identified GTR+I+G as the best model for *cytb* all codon positions, HKY+I+G for *Cytb* first codon position, HKY+I+G for *cytb* second codon position, GTR+I+G for *Cytb* third codon position, and GTR+I+G for *cytb* first and second codon. The Markov chain Monte Carlo (MCMC) analysis (with random starting trees) were run with one cold and three heated chains. We ran the programme one for 3 million generations and sampled every 100 generations (mcmcp 3000000, nchains = 4, temp = 0.1, samplefreq = 100, burnin = 7500). After the first 7500 as burnin, the remaining 22500 sampled tree were used for estimating the 50% majority rule consensus tree. Bayes factor was used to compare the results for each partitioning strategy, and aimed at accepting the better phylogeny hypothesis for the data matrix.

Results and discussion

Genomic structure, organization and composition

The complete mitochondrial genome sequence of *B. sinensis* was deposited in the GenBank database under accession number NC_017880. The total length of the *B. sinensis*

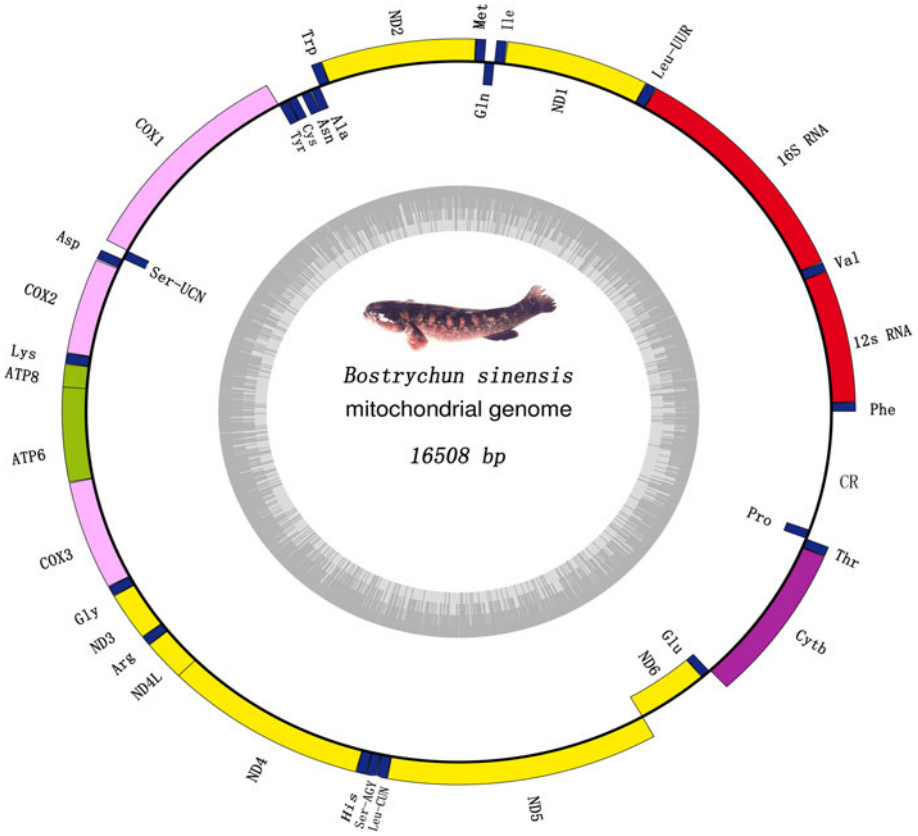

Figure 1. Gene map of the four-eyed sleeper, *B. sinensis* (Perciformes, Eleotridae) mitochondrial genome. Genes encoded on the heavy or light strands are shown outside or inside the circular gene map, respectively. Inner ring displays the GC content.

Table 2. Organization of the complete mitochondrial genome of *B. sinensis*.

Gene	Position		Size (bp)		Codon		Intergenic nucleotide[II]	Strand
	From	To	Nucleotide	Amino acid	Initiation	Stop[I]		
tRNA^{Phe}	1	68	68					H
12S rRNA	69	1021	953					H
tRNA^{Val}	1022	1093	72				2	H
16S rRNA	1096	2774	1679					H
tRNA^{Leu}(UUR)	2775	2849	75					H
ND1	2850	3824	975	324	ATG	TAG	3	H
tRNA^{Ile}	3828	3897	70				−1	H
tRNA^{Gln}	3897	3967	71				−1	L
tRNA^{Met}	3967	4035	69					H
ND2	4036	5082	1047	348	ATG	TA−	−1	H
tRNA^{Trp}	5082	5154	73				3	H
tRNA^{Ala}	5158	5226	69				1	L
tRNA^{Asn}	5228	5300	73				37	L
tRNA^{Cys}	5338	5403	66					L
tRNA^{Tyr}	5404	5474	71				1	L
COI	5476	7029	1554	517	GTG	TAA		H
tRNA^{Ser}(UCN)	7030	7100	71				3	L
tRNA^{Asp}	7104	7175	72				−4	H
COII	7182	7872	691	230	ATG	T–	1	H
tRNA^{Lys}	7873	7948	76					H
ATPase8	7950	8117	168	55	ATG	TAA	−10	H
ATPase6	8108	8791	684	227	ATG	TAA	−1	H
COIII	8791	9575	785	261	ATG	TA−	−1	H
tRNA^{Gly}	9575	9646	72					H
ND3	9647	9997	351	116	ATG	T–	−2	H
tRNA^{Arg}	9996	10064	69					H
ND4L	10065	10361	297	98	ATG	TAG	−2	H
ND4	10355	11735	1381	460	ATG	T–		H
tRNA^{His}	11736	11804	69	ATG				H
tRNA^{Ser}(AGY)	11805	11872	68				4	H
tRNA^{Leu}(CUN)	11877	11949	73					H
ND5	11950	13788	1839	612	ATG	TAA	−4	H
ND6	13785	14306	522	173	ATG	TAA		L
tRNA^{Glu}	14307	14375	69				5	L
Cytb	14381	15521	1141	380	ATG	T–		H
tRNA^{Thr}	15522	15593	72				−1	H
tRNA^{Pro}	15593	15662	70					L
Control region	15663	16508	846					H

[I]TA− and T−, represent incomplete stop codons.
[II]Numbers correspond to the nucleotides separating adjacent genes. Negative numbers indicate overlapping nucleotides.

mitochondrial genome was 16,508 bp, which was similar to other teleost species. The mitogenome contained 13 protein-coding genes, 22 transfer RNA genes, two ribosomal RNAs and a putative noncoding control region (figure 1; table 2). The structure and location of different genes in the genome corresponded to the vertebrate mitogenome model (Hurst *et al*. 1999; Waldbieser *et al*. 2003; Xu *et al*. 2011; Quan *et al*. 2012; Shi *et al*. 2012). Most of the genes were encoded on the H-strand, except that ND6 and eight tRNA genes were encoded on the L-strand, and all genes were similar in size to those in other bony fishes (Tzeng *et al*. 1992; Miya *et al*. 2001). The base composition of *B. sinensis* mitochondrial genome was as follows: 27.53% A, 31.16% C, 16.17% G, 25.14% T (table 3) and G+C content

Table 3. Base composition of *B. sinensis* mitochonondrial genome.

Gene/region	Base composition (%)				
	T	C	A	G	A+T
1st	20.85	27.89	25.70	25.56	46.55
2nd	40.33	28.00	18.20	13.47	58.53
3rd	20.31	41.01	30.65	8.03	50.96
Total	27.17	32.30	24.84	15.69	52.01
tRNA	23.68	25.87	31.00	19.45	54.68
sRNA	20.17	26.67	32.49	20.67	52.66
Control region	29.79	22.22	31.56	16,43	61.35
Overall	25.14	31.16	27.53	16.17	52.67

was 47.33%. The whole mitochondrial genome of *B. sinensis* was biased against GC nucleotides (47.33%), similar to those of *Gillichthys mirabilis* (46.28%), *Rhyacichthys aspro* (48.05%), and *Odontobutis platycephala* (43.79%). The GC and AT skew for *B. sinensis* were determined to be 0.05 and −0.31. These selective nucleotide compositional biases were ascribed to differential mutational pressures or natural selection (Bachtrog 2007). The total base compositions were against G in mitochondrial genome, maybe it was due to a strong bias against the use of G at the third codon position (Peng *et al*. 2006; Wang *et al*. 2007).

Protein coding genes and codon usage

In this study, the organization of 13 protein-coding genes in *B. sinensis* was similar to that of other vertebrate species. As shown in table 2, protein coding genes, 12 (*ND1*, *ND2*, *ND3*, *ND4*, *ND4L*, *ND5*, *COI*, *COII*, *COIII*, *ATPase6*, *ATPase8*, and *Cytb*) were encoded on the H-strand. On the contrary, *ND6* gene was coded on the L-strand. The typical initiation codon for ATG was used by 12 of 13 protein coding genes, while *COI* gene employed GTG as the start codon. A different pattern of codon usage was found in stop codons: TAA (*ATPase6*, *ATPase8*, *COI*, *ND5*, and *ND6*), and TAG (*ND1*, *ND4L*). The remaining six genes seemed to end in two incomplete stop codons, TA- (*ND2*, *COIII*) and T-(*COII*, *ND3*, *ND4*, *Cytb*), which were completed via posttranscriptional polyadenylation (Ojala *et al*. 1981). Moreover, as in other vertebrates, there were four overlaps on protein-coding genes of *B. sinensis*. The *ATPase8* and *ATPase6* genes overlapped by 10 nucleotides, the *ND4L* and *ND4* by two nucleotides, and *ATPase6* and *COIII* by one nucleotide. In addition, the *ND5* and *ND6* genes were located on different strands and had overlaps of four nucleotides. The 13 *B. sinensis* protein-coding genes had a strong bias against G (30.65% A; 41.01% C; 8.03% G; 20.31% T) at the third codon position, which was typical in vertebrate mitochondrial genomes (Inoue *et al*. 2000; Ishiguro *et al*. 2001). At the second position, pyrimidines were represented more compared to purines (T+C = 68.33%), as also seen for other vertebrate mitogenomes. But at the first codon position, the ratio of pyrimidines (48.74%) was lower than purines (51.26%). The codon usage pattern of the 13 protein-coding genes in *B. sinensis* is listed in table 4. A total of 3697 codons for 20 amino acids were identified in all protein-coding genes. Codons for leucine were the highest percentage value of

Table 4. Codon usage in *B. sinensis* mitochondrial protein-coding genes.

Amino acid	Codon	Number/frequency (%)	Amino acid	Codon	Number/frequency (%)
Phe	TTT	109/2.9	Tyr	TAT	28/0.8
	TTC	117/3.1		TAC	77/2.1
Leu	TTA	94/2.5	Stop	TAA	10/0.3
	TTG	20/0.5		TAG	2/0.1
	CTT	149/4.0	His	CAT	22/0.6
	CTC	193/5.2		CAC	84/2.3
	CTA	150/4.1	Gln	CAA	85/2.3
	CTG	45/1.2		CAG	12/0.3
Ile	ATT	120/3.3	Asn	AAT	24/0.7
	ATC	132/3.6		AAC	98/2.7
Met	ATA	104/2.8	Lys	AAA	70/1.9
	ATG	53/1.4		AAG	2/0.1
Val	GTT	47/1.3	Asp	GAT	15/0.4
	GTC	72/2.0		GAC	51/1.4
	GTA	62/1.7	Glu	GAA	74/2.0
	GTG	24/0.7		GAG	19/0.5
Ser	TCT	36/1.0	Cys	TGT	6/0.2
	TCC	81/2.2		TGC	16/0.4
	TCA	55/1.5	Trp	TGA	88/2.4
	TCG	5/0.1		TGG	27/0.7
Pro	CCT	59/1.6	Arg	CGT	7/0.2
	CCC	108/2.9		CGC	20/0.5
	CCA	42/1.1		CGA	40/1.1
	CCG	9/0.2		CGG	6/0.2
Thr	ACT	36/1.0	Ser	AGT	13/0.4
	ACC	143/3.9		AGC	43/1.2
	ACA	99/2.7	Stop	AGA	0/0.0
	ACG	13/0.4		AGG	0/0.0
Ala	GCT	63/1.7	Gly	GGT	17/0.5
	GCC	174/4.7		GGC	107/2.9
	GCA	101/2.7		GGA	59/1.6
	GCG	11/0.3		GGG	49/1.3

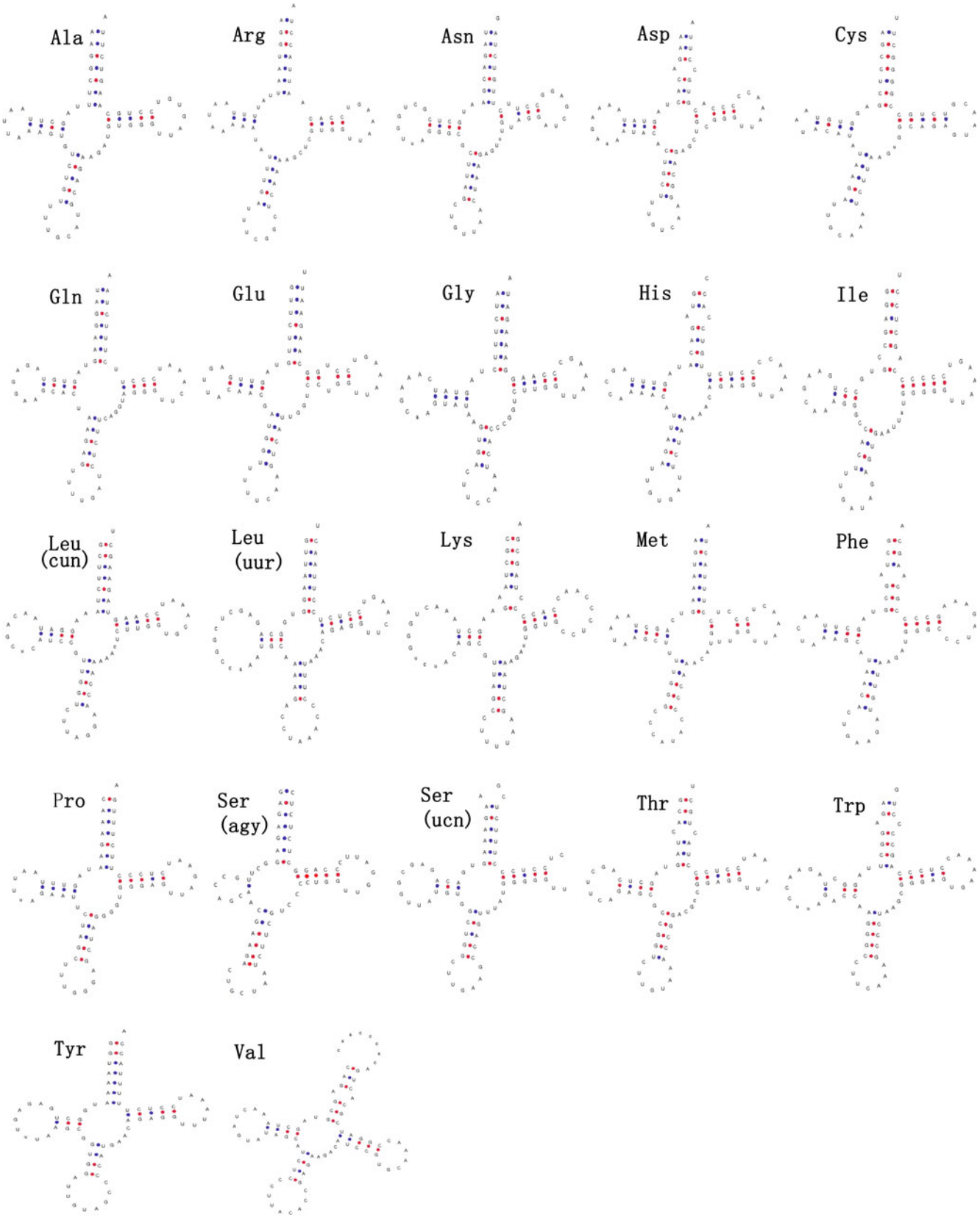

Figure 2. Sequences of *B. sinensis* mitochondrial tRNA genes, represented in the clover-form.

17.5%, followed by alanine (9.4%), and those of cysteine were the least frequent (0.6%). For amino acid with four-fold degenerated codons third position, codons ending with C were the most frequent, followed by codons ending in A and T, but A appeared more frequent than C for arginine. Within two-fold degenerate codons, C appeared to be used more than T for pyrimidine codon family. In addition, the most frequently used codon was CTC (193/3697), and TCG was the least frequently used codon (5/3697); the codon usage bias might be related to the available tRNAs in organism (Lee and Kocher 1995; Xia *et al*. 2007; Cheng *et al*. 2011).

Transfer and ribosomal RNA genes

The 22 tRNA genes and predicted structures for *B. sinensis* were determined and compared with homologues in other vertebrates (figure 2). Of these tRNAs, two forms of tRNA-Ser (UCN-AGY) and tRNA-Leu (UUR-CUN) were identified. Three tRNA clusters (IQM, HSL, and WANCY) were well conserved in *B. sinensis* as in other vertebrate mitochondrial genomes (figure 1). These tRNA genes ranged in size from 66 to 80 nucleotides, and the gene arrangement was typically as in most vertebrates. The most of tRNA secondary structures contained 7 bp in amino acid stem, 5 bp in the TΨC stem, 5 bp in the anticodon stem, and 4 bp in the DHU stem. The tRNA-Ser (AGY) had no discernible DHU stem in the *B. sinensis* mitochondrial genome, which was similar to that in the bony fish (Oh *et al*. 2007, 2008).

The 12S and 16S rRNA genes of *B. sinensis* were 953 bp and 1679 bp in length, respectively (table 2). As in other vertebrate mitochondrial genomes, these genes were located on the H-strand between tRNA Phe and tRNA Leu (UUR) and being separated by tRNA Val. The A: C: G: T base composition of the 12S and 16S rRNAs gene is 32.49%: 26.67%: 20.67%: 20.17% (table 3). The A+T content of *B. sinensis* rRNA genes was 52.66%, which was a slightly higher than protein-coding gene (52.01%), but was less in comparison with tRNA genes (54.68%).

Noncoding regions

As in most vertebrates, there are two noncoding regions in the *B. sinensis* mitochondrial genome: origin of L-strand replication and the control region. The O_L is 36 bp in length and locates in a cluster of five tRNA genes (WANCY region) between tRNA-Asn and tRNA-Cys genes. It is thought to be able to fold into a stable stem-loop secondary structure with 10 bp in the loop and 13 bp in the stem. This stem–loop structure is similar to those of bony fish whose stems are 12 to 13 bp in size and loops about 10–14 bp. Meanwhile, the conserved motif (5′-GCGGG-3′) is identified at the base of the stem in the tRNA-Cys gene, which is involved in the transition from RNA to DNA synthesis (Hixson *et al*. 1986). The conserved stem–loop structures indicate that they play a key role in the origin of mitochondrial DNA replication (Desjardins and Morais 1990).

According to alignment with the mitochondrial genome of other fishes, the control region in *B. sinensis* mitogenome is determined between tRNA Pro and tRNA Phe. The A+T content of the region is 61.35% (A = 31.56%, T = 29.79%, C = 22.22%, G = 16.43%) (table 3), which is higher than the average value of the whole mitogenome (52.67%) of *B. sinensis*. Though the control region is a unique and variable area in mitochondrial genome, termination associated sequence (TAS), and central and conserved sequence (CSB) domains are determined by multiple homologous sequence alignment in *B. sinensis* (figure 3). The TAS might play a key role in terminating the synthesis of the heavy strand (Cheng *et al*. 2010). The conserved sequence blocks (CSB-1, CSB-2 and CSB-3) are thought to be associated with positioning RNA polymerase for both priming replication and transcription (Clayton 1991; Shadel and Clayton 1997). Meanwhile, the CSB-D, CSB-E, CSB-F could also be detected in *B. sinensis* control region. It might provide a little information for examining the structure–function relationships of the control region (Cui *et al*. 2009). Even though the tandem repeats sequence in the control region has been reported in some teleosts (Arnason and Rand 1992; Lee *et al*. 1995), we did not find them in *B. sinensis*.

Effect of partitioning on harmonic mean –lnL, topology, posterior probabilities and Bayes factors

The harmonic mean $-\ln L$ was used for measuring fitness of partitioning data within the entire data set. Partitioning the *Cytb* data set greatly improved harmonic mean $-\ln L$ by codon position partition (table 5), which was similar to previous report (Brown and Lemmon 2007). The partitioning could be thought as a useful method for regulating heterogeneity in the processes of molecular evolution (Cheng *et al*. 2012). The consensus tree topologies were different in different partitioning data analyses, and these differences were inferred to alternative placements of weakly supported nodes (Bayesian posterior probabilities <95%). There were some differences in posterior probabilities among the analysis which relied on whether *Cytb* sequences were partitioned by codon position. All Bayes-factor estimates were much higher than the criterion against a hypothesis. On the basis of Bayes factors, analysis involving the P_3 partition-strategy would provide better explanation of data than the other partition-strategies (tables 5 and 6). Consequently, it is the preferred hypothesis of the Eleotridae phylogeny and next discussion will be confined to this tree about P_3 partition-strategy (figure 4).

The phylogeny of the Eleotridae family

The phylogenetic analyses were used to investigate *B. sinensis* phylogenetic position and the inner relationships

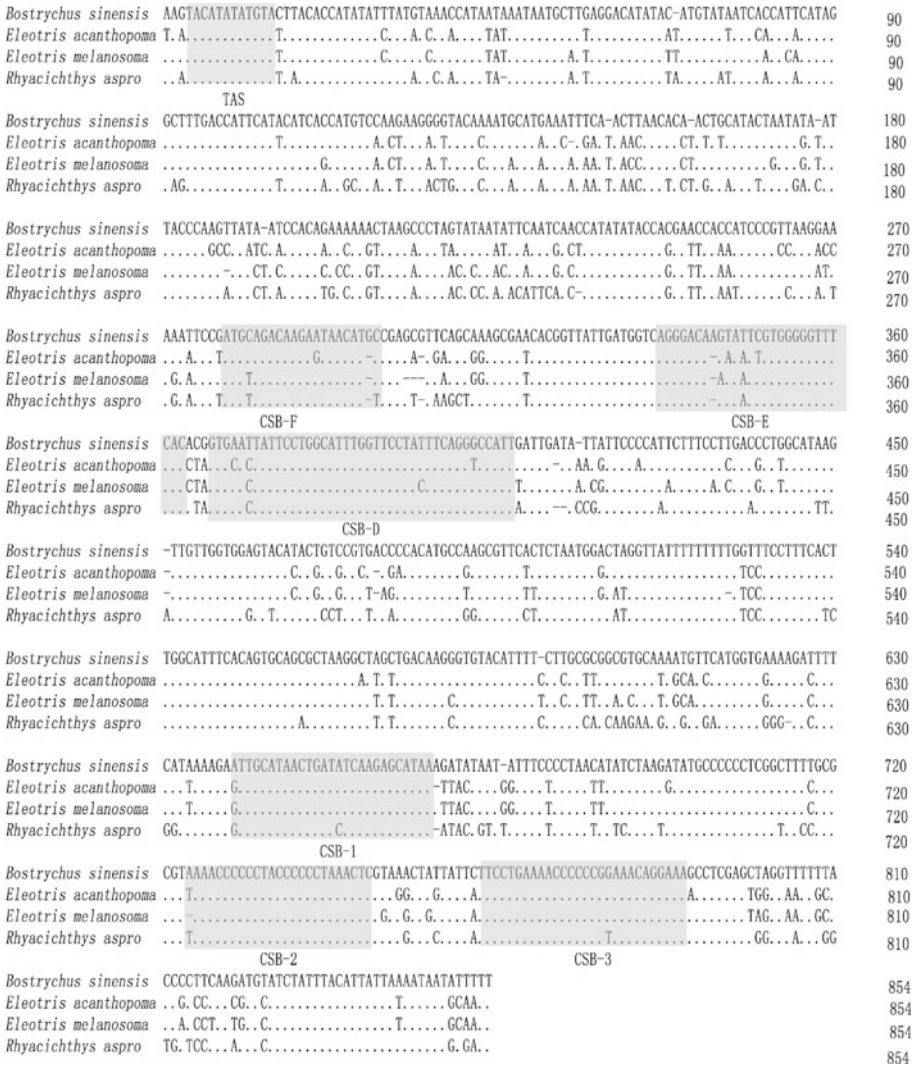

Figure 3. Complete nucleotide sequence of mitochondrial control region of Eleotridae species. Termination associated sequence (TAS), conserved sequence blocks (CSB-1, CSB-2 and CSB-3), and central conserved sequence blocks (CSB-F, CSB-E and CSB-D) are shaded in gray. The species names and GenBank accession numbers are as follows: *Bostrychus sinensis* JQ665462; *Eleotris acanthopoma* AP004455; *Eleotris melanosoma* EU352700; *Rhyacichys aspro* AP004454.

Table 5. Harmonic mean –ln*L* of each partition strategy, and 2ln Bayes factor results from comparisons between all the partitioning strategies. The Bayes factor was calculated from estimated harmonic means of likelihood by the sump command in MrBayes3.1.

	Partition strategy		
	P1	P2	P3
Harmonic mean –ln*L*	12851.73	12841.40	12260.06
P1	–		
P2	20.66	–	
P3	1183.34	1162.68	–

Table 6. Interpretations of the 2ln Bayes factor. Modified from Kass and Raftery (1995).

2ln Bayes factor	Evidence for H_1
< 0	Negative (supports H_0)
0 to 2	Bare worth mentioning
2 to 6	Positive
6 to 10	Strong
> 10	Very strong

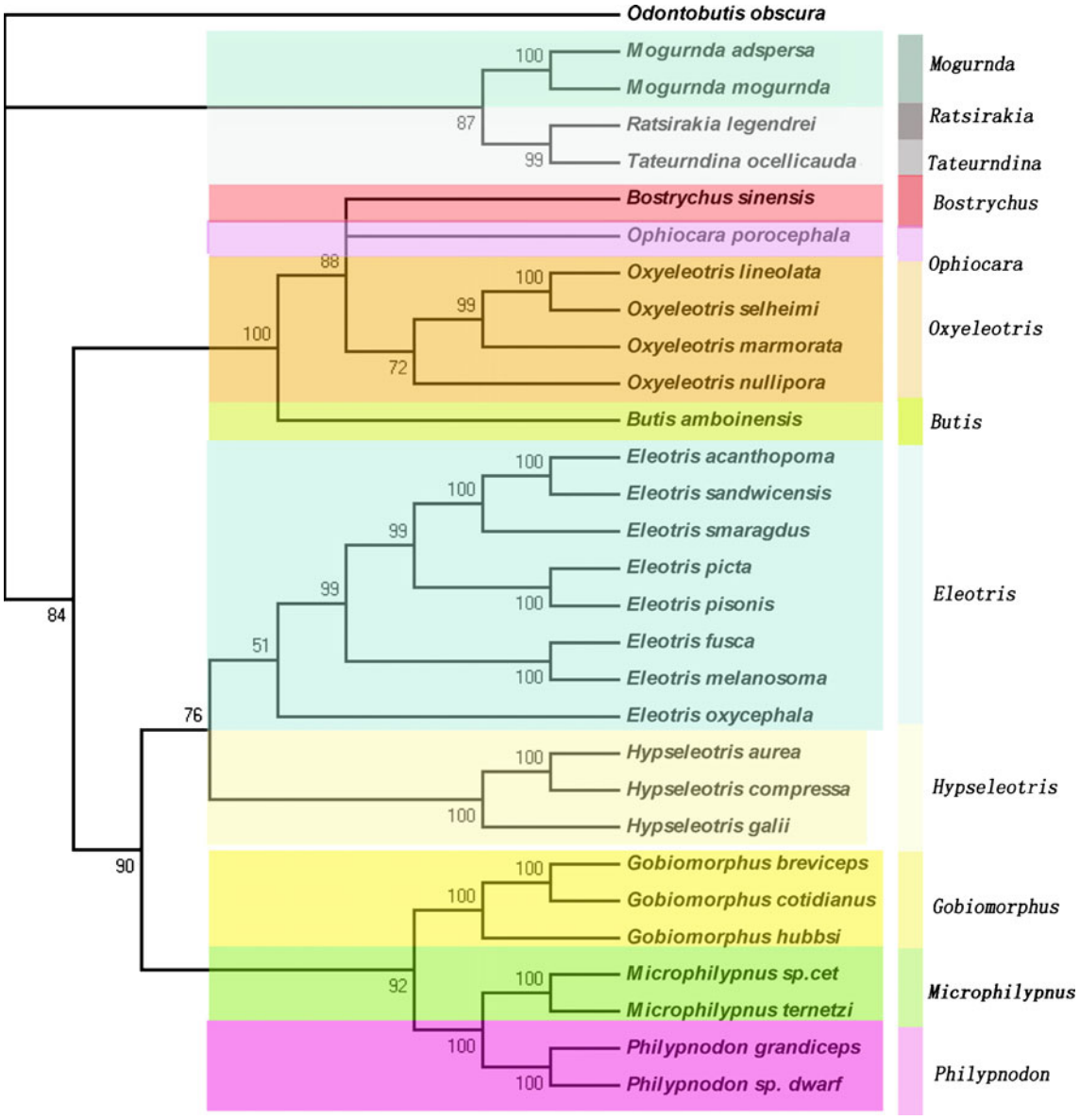

Figure 4. Consensus trees of Eleotridae constructed using Bayesian analysis based on P3 partition-strategies. The species names and the GenBank accession numbers are as follows: *Eleotris acanthopoma* AB021235, *E. fusca* AB021236, *E. melanosoma* AB021237, *E. oxycephala* AB021238, *E. pisonis* AY722229, *E. picta* AY722219, *E. sandwicensis* AY722186, *E. smaragdus* AY722185, *Hypseleotris aurea* AY722187, *H. compressa* AB021240, *H. galii* DQ468295, *Gobiomorphus breviceps* AY722224, *G. cotidianus* AB560890, *G. hubbsi* AB021239, *Microphilypnus* sp. *CET* EU380977, *Microphilypnus ternetzi* AY722253, *Philypnodon* sp. *Dwarf* DQ463936, *Philypnodon grandiceps* DQ463864, *Oxyeleotris lineolata* AY722237, *O. selheimi* AY722238, *O. marmorata* AB021246, *O. nullipora* AY722249, *Mogurnda adspersa* AY722184, *Mogurnda mogurnda* AB021242, *Ratsirakia legendrei* AY722235, *Tateurndina ocellicauda* AB021247, *Butis amboinensis* AB021232, *Ophiocara porocephala* AB021245, *B. sinensis* NC_017880, *Odontobutis obscura* AB021243.

among Eleotridae. The phylogenetic tree based on *Cytb* gene sequence was performed by partitioned Bayesian method, and it included 30 species belonging to 12 genera (figure 4). The results showed that our current analysis disagreed with previous hypothesis in some way. The molecular hypothesis is based on the mitochondrial 12S and tRNA gene, showing a monophyletic *Hypseleotris* and *Hypseleotris* as sister to the pair of *Eleotris* and *Philypnodon* (Wang *et al*. 2001). However, *Hypseleotris* is found to be sister with *Philypnodon*, excluding *Eleotris* based on three mitochondrial genes (*ND1*, *ND2*, and *COI*) (Thacker 2003). Our hypothesis indicates *Hypseleotris* is monophyletic, and sister to *Eleotris*, in addition, the *Philypnodon* are sister to *Microphlypnus*, which is not consistent with previous hypothesis (Wang *et al*. 2001; Thacker 2003). The *Bostrychus* have been most closely related to *Ophiocara*, which is congruent with previous hypotheses (Wang *et al*. 2001; Thacker and Hardman 2005). Wang *et al*. (2001) identified that

the monophyly the *Gobiomorphs* group is not supported by molecular data, and *Mogurda* is found to be sister to *Opieleotris*. However, our molecular phylogeny supports the monophyly of *Gobiomorphs,* and *Mogurda* have been most closely related to *Ratsirakia* and *Tateurndina*. Consequently, the phylogeny of Eleotridae remains to be solved in some respect, so the inclusion of further data from extensive taxonomic sampling and more molecular information are required to illustrate the phylogeny of Eleotridae.

Acknowledgements

This study was supported by National Natural Science Foundation of China (31272661). We thank Prof. Shenglong Zhao for the help of fish sampled and identified by morphology.

References

Akihito Iwata A., Kobayashi T., Ikeo K., Imanishi T., Ono H., Umehara Y. *et al*. 2000 Evolutionary aspects of gobioid fishes based upon a phylogenetic analysis of mitochondrial cytochrome b genes. *Gene* **259**, 5–15.

Arnason E. and Rand D. M. 1992 Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, Gadus morhua. *Genetics* **132**, 211–220.

Bachtrog D. 2007 Reduced selection for codon usage bias in *Drosophila Miranda*. *J. Mol. Evol.* **64**, 586–690.

Brandley M. C., Schmitz A. and Teeder T. W. 2005 Partitioned Bayesian Analyses, Partition Choice and the Phylogenetic Relationships of Scincid Lizards. *Syst. Biol.* **54**, 373–390.

Brown J. M. and Lemmon A. R. 2007 The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst. Biol.* **56**, 643–655.

Cheng Y. Z., Xu T. J., Shi G. and Wang R. X. 2010 Complete mitochondrial genome of the miiuy croaker *Miichthys miiuy* (Perciforms, Scienidae) with phylogenetic consideration. *Mar. Genomic* **3**, 201–209.

Cheng Y. Z., Wang R. X. and Xu T. J. 2011 The mitochondrial genome of the spinyhead croaker *Collchthys lucida*: Genome organization and phylogenetic consideration. *Mar. Genomic* **4**, 17–23.

Cheng Y. Z., Wang R. X., Sun Y. N. and Xu T. J. 2012 The complete mitochondrial genome of the small yellow croaker and partitioned Bayesian analysis of Sciaenidae fish phylogeny. *Genet. Mol. Biol.* **35**, 191–199.

Clayton D. A. 1991 Nuclear gadgets in mitochondrial DNA replication and transcription. *Trends Biochem. Sci.* **16**, 107–111.

Cui Z. X., Liu Y., Li C. P., You F. and Chu K. H. 2009 The complete mitochondrial genome of the large yellow croaker, *Larimichthys cracea* (Perciformes, Sciaenidae): unusual features of its control region and the phylogenetic position of the Sciaenidae. *Gene* **432**, 33–43.

Desjardins P. and Morais R. 1990 Sequence and gene organization of the chicken mitochondrial genome. A novle gene order in higher vertebrates. *J. Mol. Biol.* **212**, 599–634.

Gutell R. R., Gray M. W. and Schnare M. N. 1993 A compilation of large subunit (23S and 23S-like) ribosomal RNA structures. *Nucleic Acids Res.* **21**, 3055–3074.

He S. P., Mayden R. L., Wang X. Z., Wang W., Tang W., Tang K. L. *et al*. 2008 Molecular phylogenetics of the family Cyprindae (Actinopterygii: Cypriniforms) as evidenced by sequence variation in the first intron of S7 ribosomal protein-coding gene: Further evidence from a nuclear gene of the systematic chaos in the family. *Mol. Phylogenet. Evol.* **46**, 818–829.

Hixson J. E., Wong T. W. and Clayton D. A. 1986 Both the conserved and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light strand replication. *J. Biol. Chem.* **261**, 2384–2390.

Huelsenbeck J. P. and Ronquist F. 2001 MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755.

Hurst C. D., Bartlett S. E., Davidson W. S. and Bruce I. J. 1999 The complete mitochondrial DNA sequence of the Atlantic salmon, Salmo salar. *Gene* **239**, 237–242.

Inoue J. G., Miya M., Tsukamoto K. and Nishida M. 2000 Complete mitochondrial DNA sequence of the Japanese sardine Sardinops melanostictus. *Fish Sci.* **66**, 924–932.

Ip Y. K., Chew S F., Leong I. A. W., Jin Y., Lim C. B. and Wu R. S. S. 2001 The sleeper *Bostrichthys sinensis* (Family Eleotridae) stores glutamine and reduces ammonia production during aerial exposure. *J. Comp. Physiol. B* **171**, 357–367.

Ishiguro N., Miya M. and Nishida M. 2001 Complete mitochondrial DNA sequence of ayu Pleoglossus altivelis. *Fish Sci.* **67**, 474–481.

Kass R. E. and Raftery A. E. 1995 Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–735.

Kumar S., Dudley J., Nei M. and Tamura K. 2008 MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* **9**, 299–306.

Lee W. J. and Kocher T. D. 1995 Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. *Genetics* **139**, 873–887.

Lee W. J., Conroy J., Howell W. H. and Kocher T. D. 1995 Structure and Evolution of Teleost Mitochondrial Control Regions. *J. Mol. Evol.* **41**, 54–66.

Miya M., Kawaguchi A. and Nishida M. 2001 Mitogenomic Exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial dna sequences. *Mol. Biol. Evol.* **11**, 1993–2009.

Miya M., Takeshima H., Endo H., Isihiguro N. B., Inoue J. G., Mukai T. *et al*. 2003 Major patterns of higher teleostean phylogenies; a new perspective based on 100 complete mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **26**, 121–138.

Nylander J. A. A., Ronquist F., Huelsenbeck J. P. and Nieves-Aldrey J. L. 2004 Bayesian phylogenetic analysis of combined data. *Syst. Biol.* **53**, 47–67.

Oh D. J., Kim J. Y., Lee J. A., Yoon W. J., Park S. Y. and Jung Y. H. 2007 Complete mitochondrial genome of the rock bream *Oplegnathus fasciatus* (Perciformes, Oplegnathidae) with phylogenetic considerations. *Gene* **392**, 174–180.

Oh D. J., Kim S. J. and Jung Y. H. 2008 Comparison of the mitochondrial genomes of East Asian Pseudolabrus fishes. *Mitochondrial DNA* **19**, 453–460.

Ojala D., Montoya J. and Attardi G. 1981 tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**, 470–474.

Peng Z., Wang J. and He S. 2006 The complete mitochondrial genome of the helmet catfish *Cranoglanis bouderius* (Siluriformes: Cranoglanididae) and the phylogeny of otophysan fishes. *Gene* **376**, 290–297.

Posada D. 2008 jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256.

Quan X. Q., Jin X. X., Wang R. X., Xu T.J. and Shi G. 2012 Complete mitochondrial genome of the Walking goby *Scartelaos histophorus* (Perceformes, Gobiidae). *Mitochondrial DNA* **23**, 298–300.

Sambrook J. and Russell D. W. 2001 *Molecular cloning: a laboratory manual*, 3rd edition. Cold Spring Harbor Laboratory Press, New York, NY.

Shadel G. S. and Clayton D. A. 1997 Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* **66**, 409–435.

Shi G., Jin X. X., Zhao S. L., Xu T. J. and Wang R. X. 2012 Complete mitochondrial genome of *Trypauchen vagina* (Perciformes, Gobioidei). *Mitochondrial DNA* **23**, 151–153.

Thacker C. E. 2003 Molecular phylogeny of gobioid fishes (Teleostei: Perciformes: Gobioidei). *Mol. Phylogenet. Evol.* **26**, 354–368.

Thacker C. E. and Hardman M. A. 2005 Molecular phylogeny of basal gobioid fishes: Rhyacichthyidae, Odontobutidae, Xenisthmidae, Eleotridae. *Mol. Phylogenet. Evol.* **37**, 858–875.

Thompson J. D., Higgins D. G. and Gibson T. J. 1994 CLUSTALW: improving the sensitivity of progresssive multiple sequence alignment through sequence weighting, position-specil gap penalties and weight matrix choice. *Nulcleic Acid Res.* **22**, 4673–4680.

Tzeng C. S., Hui S. F., Shen S. C. and Huang P. C. 1992 The complete nucleotide sequence to the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates. *Nucleic Acids Res.* **20**, 4853–4858.

Waldbieser G. C., Bilodeau A. L. and Nonneman D. J. 2003 Complete sequence and characterization of the channel catfish mitochondrial genome. *Mitochondrial DNA* **14**, 265–277.

Wang H. Y., Tsai M. P., Dean J. and Lee S. C. 2001 Molecular phylogeny of gobioid fishes (Perciformes: Gobioider) based on mitochondrial 12S rRNA sequences. *Mol. Phylogenet. Evol.* **20**, 390–408.

Wang X. Z., Wang J., He S. P. and Mayden R. L. 2007 The complete mitochondrial genome of the Chinese hook snout carp *Opsariichthys bidens* (Actinopterygii: Cypriniformes) and an alternative pattern of mitogenomic evolution in vertebrate. *Gene* **399**, 11–19.

Wyman S. K., Jansen R. K. and Boore J. L. 2004 Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**, 3252–3255.

Xia J., Xia K., Gong J. and Jiang S. 2007 Complete mitochondrial DNA sequence, gene organization and genetic variation of control regions in *Parargyrops edita*. *Fish Sci.* **73**, 1042–1049.

Xu T. J., Cheng Y. Z., Sun Y. N., Shi G. and Wang R. X. 2011 The complete mitochondrial genome of bighead croaker, *Collichthys niveatus* (Perciforms, Sciaenidae): structure of control region and phylogenetic considerations. *Mol. Biol. Rep.* **38**, 4673–4685.

Yang Z. 1996 Maximum likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol.* **42**, 587–596.

Received 1 February 2013, in revised form 20 March 2013, accepted 23 April 2013
Published on the Web: 12 August 2013