МОЛЕКУЛЯРНАЯ БИОЛОГИЯ, 2011, том 45, № 3, с. 434–445

ГЕНОМИКА. ТРАНСКРИПТОМИКА

УДК 577.21

THE COMPLETE MITOCHONDRIAL GENOME OF THE MARBLED ROCKFISH *Sebastiscus marmoratus* (Scorpaeniformes, Scorpaenidae): GENOME CHARACTERIZATION AND PHYLOGENETIC CONSIDERATIONS

© 2011 Tian-Jun Xu, Yuan-Zhi Cheng, Xue-Zhu Liu, Ge Shi, Ri-Xin Wang*

Key Laboratory for Marine Living Resources and Molecular Engineering, College of Marine Science, Zhejiang Ocean University, China

Received April 02, 2010
Accepted for publication May 12, 2010

The complete mitochondrial genome sequence of the marbled rockfish *Sebastiscus marmoratus* (Scorpaeniformes, Scorpaenidae) was determined and phylogenetic analysis was conducted to elucidate the evolutionary relationship of the marbled rockfish with other Sebastinae species. This mitochondrial genome, consisting of 17 301 bp, is highly similar to that of most other vertebrates, containing the same gene order and an identical number of genes or regions, including 13 protein-coding genes, two ribosomal RNAs, 22 transfer RNAs, and one putative control region. Most of the genes are encoded on the H-strand, while the ND6 and seven tRNA genes (for Gln, Ala, Asn, Tyr, Ser (UCA), Glu, and Pro) are encoded on the L-strand. The reading frame of two pairs of genes overlapped on the same strand (the ATPase 8 and 6 genes overlapped by ten nucleotides; ND4L and ND4 genes overlapped by seven nucleotides). The possibly nonfunctional light-strand replication origin folded into a typical stem-loop secondary structure and a conserved motif (5'-GCCGG-3') was found at the base of the stem within the $tRNA^{Cys}$ gene. An extent termination-associated sequence (ETAS) and conserved sequence blocks (CSB) were identified in the control region, except for CSB-1; unusual long tandem repeats were found at the 3' end of the control region. Phylogenetic analyses supported the view that Sebastinae comprises four genera (*Sebates*, *Hozukius*, *Helicolenus*, and *Sebasticus*).

***Keywords*: *Sebastiscus marmoratus*, complete mitochondrial genome, Sebastinae, phylogenetic analysis, control region.**

The marbled rockfish, *Sebastiscus marmoratus,* belongs to the Scorpaenidae, a diverse clade of Scorpaeniformes. This ovoviviparous fish, inhabiting littoral rocky bottoms from Japan to the East China Sea, is an important commercial species with a worldwide market demand [1]. Although it is a high-value marine food fish species, little is known about the genetic characteristics of marbled rockfish.

Genome-scale approaches have shed light on the evolutionary relationship among diverse organisms [2–4]. Compared with the nuclear genome, the mitochondrial genome has several intrinsic characteristics, such as a single copy nature, relatively short generation time and time for coalescence in lineages, maternal inheritance mode, and lack of genetic recombination. These characteristics make the mitochondrial genomic DNA one of the most suitable tools for phylogenetic studies, such as analysis of introgression, hybridization and gene flow [5]. Liu and Cui [6] sequenced the complete mitochondrial genome of the cutlassfish *Trichiurus japonicus* and utilized it in combination with the data from Scombridae species to resolve a non-monophyletic Trichiuridae, which conflicted with the morphological results. In a phylogeny proposed by Richard et al. [7], basing on previously published mitochondrial genomes and multiple nuclear loci, eight families were recognized within the Cypriniformes, and were considered to form two major clades, the Cyprinoidea and the Cobitoidea, with a high probability these are novel phylogenetic relationships among the earth's most diverse clade of freshwater fishes.

The evolutionary relationships of the live-bearing rockfishes of the subfamily Sebastinae have interested scientists since their original description. However, historicaly studies have mainly focused on the genus *Sebastes*; and relatively little is known about the phylogenetic relationships of the species from *Sebastiscus.* Based on morphological characteristics, *Sebastiscus* was regarded as a subgenus of *Sebastes* [8–11]. Other authors [12, 13] also treated *Sebastiscus* as a subgenus of *Sebastes*. On the other hand, it was shown, that *Helicolenus*, a genus of Sebastinae, was more closely related to *Sebastes* than to *Sebastiscus* using sequence variations in the gene of mitochondrial DNA of cytochrome *c* oxidase 1 [14]. Further

Abbreviations: ATPase 6 and ATPase 8 – ATPase subunit 6 and 8; bp – base pair(s); CO1-3 – Cytochrome *c* Oxidase subunits 1–3; CR – Control Region; CSB – Conserved Sequence Blocks; Cyt *b* – Cytochrome *b*; mitogenome – mitochondrial genome; ND1-6 – NADH Dehydrogenase subunits 1-6; ETAS – Extent Termination Associated Sequence; O_L – light-strand replication origin.

* E-mail: wangrixin1123@126.com

research, in which the variations of existing partial sequences of the cytochrome *b* gene were investigated to infer the phylogenetic position and the radiation of the species of *Sebastes*, suggested that the *Helicolenus* and *Hozukius* are more closely related to *Sebastes* than *Sebastiscus* [15]. Similarly, other authors [16] provided accounts of the position and evolution of the species of the Sebastinae subfamily using sequence data from seven mitochondrial genes and two nuclear genes; in the phylogenetic analyses, *Sebastiscus* was treated as a genus of Sebastinae [16]. The issues surrounding the phylogenetic position of *Sebastiscus* are far from being resolved in systematic biology due to a lack of traditional research using different data types and phylogenetic analyses. In addition, only seven mitogenomes of Scorpaeniformes are currently available in the GenBank, EMBL, and DDBJ databases. No one has yet analyzed the mitochondrial genome of the *S. marmoratus*.

In this study, we present the complete sequence and structure of the mitochondrial genome of *S. marmoratus*. We also report on the gene arrangement and codon usage of *S. marmoratus* mitochondrial genomic DNA and compare it with the mitochondrial genomes of other Scorpaeniformes fishes. Finally, the complete cytochrome *b* gene sequences from 30 previously reported Scorpaenidae species were used to analyze the phylogenetic position of *S. marmoratus* within the Sebastinae subfamily. We determined that *Sebastiscus* is a genus of Sebastinae not a subgenus of *Sebastes*, supporting one of the parties of previous researches. In combination with the homologous data from GenBank, the information reported here might facilitate further investigation of the molecular evolution of the Scorpaeniformes order.

EXPERIMENTAL

Fish sample and DNA extraction. The specimens of *S. marmoratus* were collected in the Zhoushan fishing ground and were identified by morphology. Muscle tissue were removed and kept at –80°C until use. The genomic DNA was extracted using the standard phenol-chloroform method [17] and visualized on a 1.0% aga-rose gel.

PCR procedures and sequencing. A set of primers (Table 1) for amplifying adjacent, overlapping segments of the complete mitochondrial genome of *S. marmoratus* was designed based upon the mitochondrial genome sequences of the two species closely related to the marbled rockfish, *Sebates schlegelii* (AY491978) [18] and *Helicolenus hilgendorfi* (AP002948) [19]. All PCRs were carried out in a final volume of 50 μl, each containing 5.0 μl of 10 × *Taq* Plus polymerase buffer, 0.2 mM $dNTP_S$, 0.2 μM of the forward and reverse primers, 2 units of *Taq* Plus DNA polymerase with proof-reading characteristic ("TIANGEN", China), and 1 μl of the DNA template. Cycling conditions were 94°C for 4 min, followed by 35 cycles of 94°C for 50 s, 60°C for 60 s and 72°C for 2–3 min, followed by 1 cycle of 72°C for 10 min. PCR was performed on a PTC-200 ("BioRad", USA). The PCR products were electrophoresed on a 1% agarose gel to check the integrity and then visualized by the Molecular Imager Gel Doc XR system ("BioRad"). The PCR products were purified using a QIAEX II Gel Extraction Kit ("Qiagen", China). The purified fragments were ligated into PMD18-T vectors ("Takara", "Otsu", "Shiga", Japan) and transformed TOP10 cells ("TIANGEN") according to the standard protocol. Positive clones were screened via PCR with M13+/– primers and the amplicons were sequenced using the ABI 3730 automated sequencer ("Applied Biosystems", USA) with M13+/– primers.

Table 1. Primers used for PCR amplification of the mtDNA of *S. marmoratus*

Name	Sequence(5'- 3')
Sema-1-F	AACCTCACCCTTCCTTGTTTATCC
Sema-1-R	TCTTATGTTTGTATCCCCGCTTCT
Sema-2-F	GACCTGTATGAATGGCACAACGAG
Sema-2-R	GGTACTGGCGAAGAGTAGTATGGC
Sema-3-F	ATATTCCTCCATCGCACATCTCG
Sema-3-R	GCTCAGGCACCAAATACTAGATAAA
Sema-4-F	ATTACCGCTGTTCTTCTCCTTCT
Sema-4-R	TAGTCTGTTTTCGTTCACCTTCC
Sema-5-F	CAGCATTATGGAAGGTGAACGAA
Sema-5-R	GGATGAGTAAGCGATGAGGGATT
Sema-6-F	TCGCTCAACAACCCTAATAACC
Sema-6-R	ACGGCTCAGAAGAGTGTCAGTAA
Sema-7-F	TATTATCCTTGCCCTCTG
Sema-7-R	GGTTTATGTATGGGTCTGC
Sema-8-F	TTAGCATACCCTATCTTTACGACTC
Sema-8-R	TAGCAGGACTACTCCGATGTTTC
Sema-9-F	CACATCTAAACAACGAAGCCTCAC
Sema-9-R	TTTCTTTGAGCAGTAGGGAGGAC
Sema-10-F	GTCCTCCCTACTGCTCAAAGAAA
Sema-10-R	TGTCCTTCACCTTCAATAACCGT
Sema-11-F	AAAGTCCTCCCTACTGCTCAAA
Sema-11-R	CCCGCTTACTACTAAATCCTCC
Sema-12-F	TATTCCTGGCATTTGGTTC
Sema-12-R	GCTTGTTGGGTTTCGTCTA
Sema-13-F	AACTTCATCCCTGACTTCACTCT
Sema-13-R	TCTAATCCTTCTCCCTACTTTGC

Sequence analysis. DNA sequences were analyzed using the Lasergene software version 5.0 (DNASTAR), contig assembly was performed with the program "Seqman". The location of 13 protein-coding genes and two rRNAs was determined by DOGMA [20] with default settings, and checked by comparing them to the other Scorpaeniformes fishes. Gene predictions were subsequently refined by comparing DNA or amino acid sequences with the known sequences from other rockfishes and the

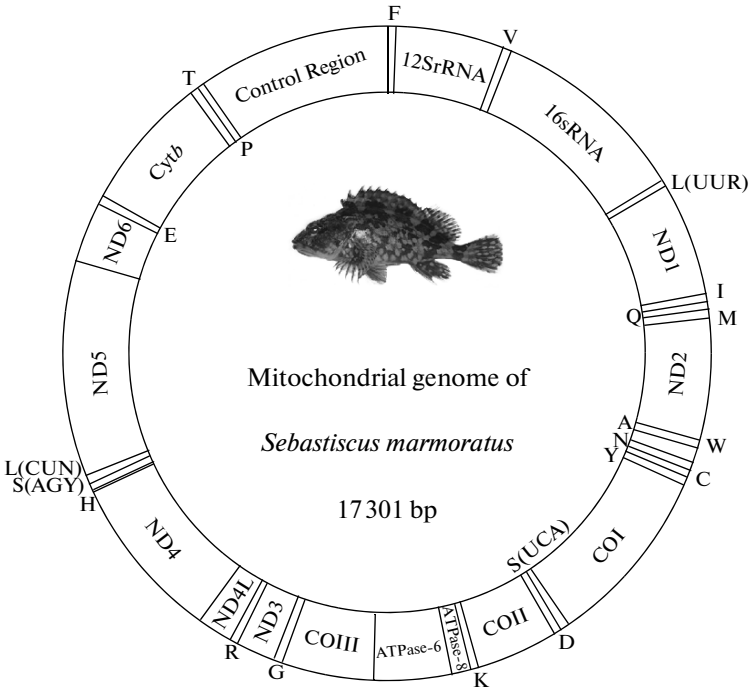

Fig. 1. Organization of the mitogenome of *S. marmoratus* (GenBank Acc. No. GU452728). All protein-coding genes are encoded on the H-strand except for ND6, which is encoded on the L-strand. Both the ribosomal RNA genes are encoded on the H-strand. Transfer RNA genes are designated by single-letter amino acid codes. Genes encoded on the H-strand and L-strands are shown outside and inside the circular gene map, respectively.

codon usage was analyzed with MEGA 4.0 [21]. Most of the 22 tRNA genes were identified with the help of tRNAscan-SE1.21 [22]; the remaining tRNA genes were identified by inspecting the sequence for the tRNA-like secondary structure and anticodons. The putative O_L and CR were identified by sequence homology and proposed secondary structure. The complete mitochondrial genome sequence was deposited in the GenBank (Acc. No GU452728).

Phylogenetic analysis. In this study, a broad range of taxa were chosen for phylogenetic analysis. Nucleotide sequences were analyzed by plotting the number of transitions and transversions on each codon position against the Tamura and Nei genetic distance using the DAMBE program [23]. The transitions and transversions in the mitochondrial Cyt *b* gene were accumulated linearly and showed no saturation patterns at any position; therefore, all of the nucleotide positions were employed in subsequent analysis [24]. Phylogenetic trees were constructed via the Neighbor-Joining method [25].

RESULTS

Genome composition

The complete mitochondrial genome sequence of *S. marmoratus* is 17301 bp in length, which falls within the range of teleost mitogenomes (Fig. 1). The overall base composition of the H-strand is A – 28.7%; C – 28.1%; T – 26.7%; and G –16.5% (Table 2). This mitogenome included 13 protein-coding genes, 12S rRNA and 16S rRNA genes, 22 tRNA genes, and a control region. Most of the genes are encoded on the H-strand, except for the ND6 gene and seven tRNA ($tRNA^{Gln}$, $tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Tyr}$, $tRNA^{Ser}$ (UCA), tRNA-Glu, and tRNA-Pro) genes. The organization of the marbled rockfish mitogenome conforms to the consensus gene order of other fish mitochondrial genomes [26, 27].

Sequence features of the protein-coding genes

Among 13 protein-coding genes, two overlapping reading-frames were detected on the same strand (Table 3). The ATPase 6 and ATPase 8 overlap by 10 nucleotides, and ND4 and ND4L share seven nucleotides. ND5 and ND6 overlap by four nucleotides on the opposite strand. ATG is the initiation codon of 12 out of the 13 protein-coding genes (ND1, ND2, CO2, ATPase 8, ATPase 6, CO3, ND3, ND4L, ND4, ND5, ND6, and Cyt *b*), while the initiation codon of CO1 is GTG. TAA is the stop codon for nine genes (ND1, ND2, CO2, ATPase 8, ATPase 6, CO3, ND4L, ND4, and ND5), CO1 ends with AGA, and the other genes have incomplete stop codons, either TA or T, which are presumably completed as TAA by post-transcriptional polyadenylation [28].

Codon usage of 13 protein-coding genes was analyzed (Table 4). Of these genes, the L-strand sequence was used only for the ND6 gene. The total number of codons used in the 13 protein-coding genes was 3830 for 20 amino ac-

Table 2. Comparison of the base content in different genes or regions of the marbled rockfish mitochondrial genome

Gene/region	Base composition (%)				
	T	C	A	G	A + T
Protein coding					
1st	21	27.9	26.4	24.8	47.4
2nd	39	28.0	19.4	13.9	58.4
3rd	24	34.3	34.3	7.4	58.3
Total	27.8	30.1	26.7	15.4	54.5
tRNA	23.8	25.4	30.4	20.5	54.2
sRNA	20.4	26.3	31.8	21.5	52.4
Control region	31.6	18.9	37.1	12.4	68.7
Overall	26.7	28.1	28.7	16.5	55.4

ids, the proportion of codons for leucine was 17.2% (658/3830, the largest codon), while cysteine was 0.68% (26/3830, the smallest codon). Leucine and serine are encoded by six different codons. Glycine, proline, threonine, arginine, alanine, and valine are encoded by four different codons. Methionine is encoded by three different codons, including GTG as a start codon, and the others are encoded by two different codons. In addition, a strong anti-G bias of 7.4% was observed in the third codon position (Table 2), which is similar to the findings in other vertebrate mitochondrial genomes [29, 30]. At the second codon position, pyrimidines (T + C = 67.0%) are over-represented in comparison with purines, owing to the hydrophobic character of the proteins [31]. For amino acids with four-fold degenerate third positions, codon families ending in C were the most frequent, followed by codons ending in A and T. Among two-fold degenerate codons, C appeared to be used more than T in the pyrimidine codon family, whereas the purine codon family ended mostly with A. These patterns were in agreement with the frequencies of C, A and T at the third codon position. Consistent with the overall bias against G, G was the least common third position nucleotide in all codon families, except for arginine, where G was similar in frequency to C and T, but still less than A. For glycine, G was more frequent than T (Table 4). All these features are very similar to those observed in other vertebrates [26, 32].

Ribosomal RNA and transfer RNA

The 12S and 16S ribosomal RNA genes of *S. marmoratus* comprise 943 bp and 1692 bp respectively. They are located between $tRNA^{Phe}$ and $tRNA_{Leu}$ (UUR), and are separated by $tRNA^{Val}$, as they are in other vertebrates [33]. The 22 tRNA genes (Table 3) are interspersed in the genome and range in size from 64 to 74 bp and fold into cloverleaf secondary structures with normal base paring. However, there are some base pairs that are G-U wobbles and there are other atypical parings in the stem regions. The average base composition in the tRNAs is 23.8% for T; 25.4% for C; 30.4% for A; 28.5% for G. The tRNA genes contain more A + T bases than the rRNA genes, but less A + T in comparison to the protein-coding genes. All of the postulated tRNA cloverleaf structures generally contain seven bp in the aminoacyl stem, four bp in the TΨC stem, five bp in the anticodon stem, and five bp in the DHU stem. The $tRNA^{Ser}$ (AGY) of the *S. marmoratus* mitogenome does not have seven bp in the aminoacyl stem, as found in most bony fishes [34, 35]. Aberrant tRNA can also fit the ribosome by adjusting its structural conformation and function in a similar way to that of other usual tRNAs in the ribosome [36].

The origin of light strand replication

The putative O_L was confirmed in *S. marmoratus*, and compared with other Scorpaeniformes fishes (*Aptocyclus ventricosus*, *S. schlegelii*, *H. hilgendorfi*, *Satyrichthys amiscus*, *Dactyloptena tiltoni*, *Cottus reinii*, and *Dactyloptena peterseni*); they are almost identical and are located in a cluster of five tRNA genes (the WANCY region) between the $tRNA^{Asn}$ and $tRNA^{Cys}$ genes. The putative O_L serves as the initiation site of the light-strand replication and can fold into a stable stem-loop secondary structure with 13 bp in the stem and 12 bp in the loop. It also possesses a C-rich sequence in the loop where RNA primer synthesis could be initiated. This C-rich sequence has also been found in the O_L loop of other fishes, such as *Oncorynchus mykiss* [37] and *Gadus morhua* [38]. This feature supports the hypothesis that in vertebrates, primer synthesis is most probably initiated by a polypyrimidine tract [39], and not by a stretch of thymines as previously suggested [40]. The conserved sequence motif 5'-GCCGG-3' was found at the base of the stem within the $tRNA^{Cys}$, except for that of *A. ventricosus* and *C. reinii*; this motif seems to be involved in the transition from RNA synthesis to DNA synthesis [41].

Table 3. Characteristics of the mitogenome of *S. marmoratus*

Gene	Position		Size (bp)		Codon		Intergenic nucleotide**	Strand
	from	to	nucleotide	amino acid	start	stop*		
$tRNA^{Phe}$	1	68	68				0	H
12S RNA	69	1011	943				3	H
$tRNA^{Val}$	1015	1086	72				0	H
16S RNA	1087	2778	1692				0	H
$tRNA^{Leu(UUR)}$	2779	2852	74				0	H
ND1	2853	3827	975	324	ATG	TAA	4	H
$tRNA^{Ile}$	3832	3901	70				−1	H
$tRNA^{Gln}$	3901	3971	71				−1	L
$tRNA^{Met}$	3971	4041	71				0	H
ND2	4042	5088	1047	348	ATG	TAA	−1	H
$tRNA^{Trp}$	5088	5158	71				1	H
$tRNA^{Ala}$	5160	5228	69				1	L
$tRNA^{Asn}$	5230	5302	73				39	L
$tRNA^{Cys}$	5342	5405	64				0	H
$tRNA^{Tyr}$	5406	5476	71				1	L
CO1	5478	7028	1551	516	GTG	AGA	−3	H
$tRNA^{Ser(UCA)}$	7026	7095	70				3	L
$tRNA^{Asp}$	7099	7171	73				6	H
CO2	7178	7882	705	234	ATG	TAA	−14	H
$tRNA^{Lys}$	7869	7942	74				1	H
ATPase 8	7944	8111	168	55	ATG	TAA	−10	H
ATPase 6	8102	8785	684	227	ATG	TAA	−1	H
CO3	8785	9570	786	261	ATG	TAA	−1	H
$tRNA^{Gly}$	9570	9641	72				0	H
ND3	9642	9992	351	116	ATG	TA–	−2	H
$tRNA^{Arg}$	9991	10059	69				0	H
ND4L	10060	10356	297	98	ATG	TAA	−7	H
ND4	10350	11780	1431	476	ATG	TAA	−53	H
$tRNA^{His}$	11728	11796	69				0	H
$tRNA^{ser(AGY)}$	11797	11864	68				4	H
$tRNA^{Leu(CUN)}$	11869	11941	73				0	H
ND5	11942	13780	1839	612	ATG	TAA	−4	H
ND6	13777	14298	522	173	ATG	TA–	0	L
$tRNA^{Glu}$	14299	14367	69				7	L
Cyt *b*	14375	15515	1141	380	ATG	T–	0	H
$tRNA^{Thr}$	15516	15587	72				−1	H
$tRNA^{Pro}$	15587	15656	70				0	L
Control Region	15657	17301	1645					H

* TA- and T – represent incomplete stop codons.

** Numbers correspond to the nucleotides separating adjacent genes. Negative numbers indicate overlapping nucleotides.

Table 4. Codon usage in mitochondrial protein-coding genes in *S. marmoratus*

Codon	/C (%)	Codon	/C (%)	Codon	/C (%)	Codon	/C (%)
UUU	110(2.8)	UCU	39(1.0)	UAU	62(1.6)	UGU	7(0.18)
Phe		Ser		Tyr		Cys	
UUC	116(3.0)	UCC	67(1.7)	UAC	51(1.3)	UGC	19(0.5)
Phe		Ser		Tyr		Cys	
UUA	103(2.7)	UCA	75(2.0)	UAA	9(0.2)	UGA	104(2.7)
Leu		Ser		Stop		Trp	
UUG	29(0.8)	UCG	4(0.1)	UAG	0(0)	UGG	16(0.4)
Leu		Ser		Stop		Trp	
CUU	154(4)	CCU	47(1.2)	CAU	34(0.9)	CGU	8(0.2)
Leu		Pro		His		Arg	
CUC	137(3.6)	CCC	106(2.8)	CAC	71(1.9)	CGC	18(0.5)
Leu		Pro		His		Arg	
CUA	181(4.7)	CCA	62(1.6)	CAA	80(2.1)	CGA	46(1.2)
Leu		Pro		Gln		Arg	
CUG	59(1.5)	CCG	10(0.3)	CAG	16(0.4)	CGG	11(0.3)
Leu		Pro		Gln		Arg	
AUU	172(4.5)	ACU	55(1.4)	AAU	51(1.3)	AGU	18(0.5)
Ile		Thr		Asn		Ser	
AUC	98(2.6)	ACC	108(2.8)	AAC	59(1.5)	AGC	46(1.2)
Ile		Thr		Asn		Ser	
AUA	91(2.4)	ACA	122(3.2)	AAA	68(1.8)	AGA	1(0.02)
Met		Thr		Lys		Stop	
AUG	56(1.5)	ACG	12(0.3)	AAG	5(0.1)	AGG	0(0)
Met		Thr		Lys		Stop	
GUU	66(1.7)	GCU	73(1.9)	GAU	29(0.7)	GGU	44(1.1)
Val		Ala		Asp		Gly	
GUC	55(1.4)	GCC	163(4.3)	GAC	54(1.4)	GGC	82(2.1)
Val		Ala		Asp		Gly	
GUA	86(2.2)	GCA	110(2.9)	GAA	78(2.0)	GGA	84(2.2)
Val		Ala		Glu		Gly	
GUG	22(0.6)	GCG	16(0.4)	GAG	19(0.5)	GGG	35(0.9)
Val		Ala		Glu		Gly	
GUG	1(0.02)						
Met							
Total codon cout:		3830					

Fig. 2. Alignment of the partial sequences of the mtDNA control regions of *H. hilgendorfi, S. amiscus, S. schlegeli*, and *S. marmoratus* (represented by 1, 2, 3, and 4, respectively). The TAS, CSB-F, CSB-E, CSB-D, CSB-2, and CSB-3 are boxed.

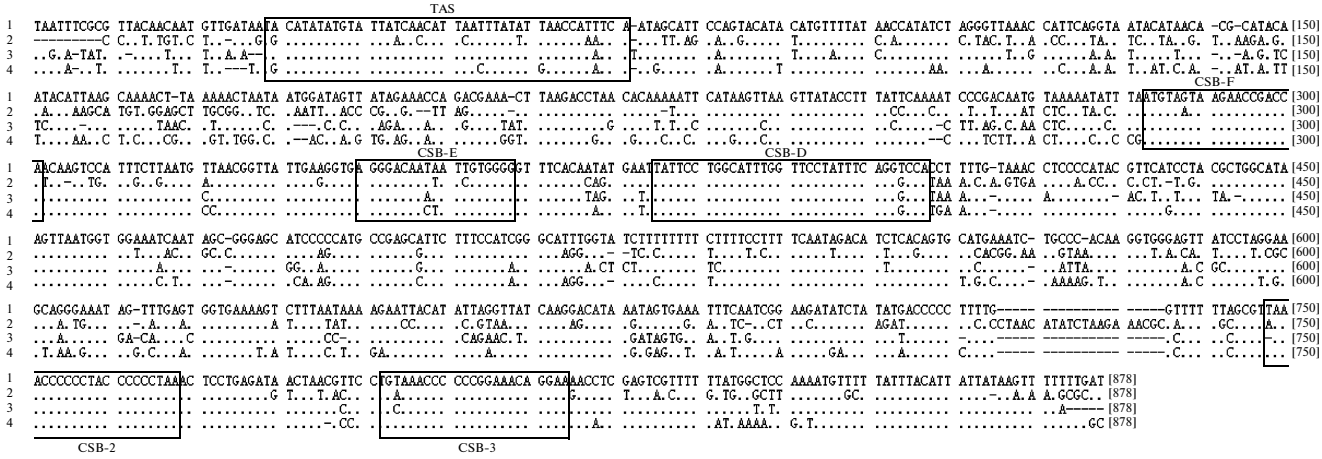

Fig. 3. Complete sequence of the control region of *S. marmoratus*. The sequence is presented as the L-strand sequence from the 5'- to the 3'-end. In the control region, the putative conserved elements (TAS, CSB-F, CSB-E, CSB-D, CSB-2, and CSB-3) are boxed and marked; the key sequence of CSB-1 (GACATA) is also boxed. The tandem repeats and the imperfect repeat are marked. The motif (TAS: TACAT) is underlined.

Control region

The major non-coding region in *S. marmoratus* is located between $tRNA^{Pro}$ and $tRNA^{Phe}$, and was determined to be 1645 bp in length. It has an overall base composition that is rich in A and T (A + T = 68.7%). This non-coding sequence appeares to correspond to the typical tripartite structure with an extent termination association sequence (ETAS), a central domain, and CSB domains [42]. Several fragments comparable to ETAS, which are thought to act as a signal for the termination of heavy strand elongation, were identified (Fig. 2). Southern first recognized the conserved sequences CSB-B, CSB-C, CSB-D, CSB-E, and CSB-F in the central conserved sequence block domain in mammals [43]; however, only CSB-F, CSB-E, and CSB-D could be identified in fishes [44]. All three of these conserved motifs were identified in the central domain in this study (Fig. 3). In addition, conserved sequence blocks CSB-2 and CSB-3, which are thought to be involved in the positioning of the RNA polymerase both for transcription and for priming replication, were found in the CSB domain. However, no CSB-1 sequence was detected, even though *S. marmoratus* does contain a GACATA-box, which is a typical characteristic of CSB-1 (Fig. 3). These features suggest that the asymmetrical replication mechanism revealed for mammalian mitochondrial DNA also operates in *S. marmoratus*. Interestingly, two unusually long perfect copies (270 bp) and an imperfect copy (151 bp) placed in tandem, were also found at the 3' end of the control region, which is similar to the situation in *S. thompsoni* [45]. Although the presence of tandem re-

peats in the control region has been reported in other teleosts, the size of those sequences tended to be smaller [46, 47]. Slippage and mispairing during replication of the mitogenome could explain the tandem repeats in the control region [48].

Phylogenetic placement of S. marmoratus

Phylogenetic trees were obtained when NJ method was applied to the complete cyt *b* gene sequence of 31 Sebastinae species (Fig. 4). The NJ tree constructed using Tamura and Nei's matrix placed *Sebastiscus* as the first clade among the Sebastinae subfamily. The second cluster, supported by 81% of bootstrap probabilities, contained *Helicolenus*, *Hozukius,* and *Sebates*. The monophyly of species of *Sebastes* was supported with 100% bootstrap probabilities. Thus we propose that *Sebastiscus* is a genus of the Sebastinae subfamily, and not a subgenus of *Sebates* and that the genera *Helicolenus* and *Hozukius* are more closely related to the genus *Sebastes*, with high bootstrap support. But the bootstrap value, which supported the phylogeny of major clusters in the *Sebastes* genus, was weak for the cyt *b* datasets.

DISCUSSION

The complete mitochondrial DNA of *S. marmoratus* is the eighth to be reported for a member of the Scorpaeniformes order. Compared with the complete mitochondrial sequences of other species from the same order, the mitogenome of *S. marmoratus* is 17301 bp in length, and is significantly larger in size than those of *A. ventricosus* (15974 bp), *S. schlegelii* (16525 bp), *S. amiscus* (16526 bp) and *C. reinii* (16561 bp), and similar to those of *H. hilgendorfi* (16728 bp), *D. tiltoni* (16751 bp), and *D. peterseni* (16717 bp). This length variation of the mitogenomes in these species is largely due to the number and size of the repeats inserted in the control region (Fig. 2). Similar phenomena were observed in *S. thompsoni* [45] and *D. tiltoni* [49]; however, it is worth noting that the form of the tandem repeats is different in these species. *S. marmoratus* and *S. thompsoni* both have tandem repeats in the 3' end of the control region, where two perfect copies are followed by an imperfect copy, while the tandem repeats of *D. tiltoni* are at the 5' end of the control region and are comprised of five perfect copies. Additionally, both *D. peterseni* and *H. hilgendorfi* contained one set of repeats that were repeated three and two times, respectively.

Although the high length variation of the control region exists in the Scorpaeniformes fishes, the typical conserved blocks were still recognizable by performing multiple alignments of the sequences of the control regions and comparing them with the key conserved sequences reported in other fishes. CSB-F, CSB-E, CSB-D, CSB-2, and CSB-3 were easily identified in these seven species because they were more conserved than ETAS and CSB-1. Two types of ETAS are recognized in these Scorpaeniformes fishes (except in *D. tiltoni* and *D. peterseni*). The consensus sequence of the first type is "TGCATATATG-TAATTA-C-CA–ATTTAT-T-AA-TCA" (where hyphens indicate nucleotide variations such as transitions, transversions, or deletions) with the conserved motif TG-CAT, found in *C. reinii* and *S. amiscus*. The second type of ETAS is "TACATATATGTAATTA-C-CA–ATT-TAT-T-AA-TCA" with the motif TACAT, found in *S. marmoratus, H. hilgendorfi* and *S. schlegelii*. The conserved ETAS motif in most fishes is TACAT, with one palindromic sequence, ATGTA. In the Scorpaeniformes fishes, there is some variation in ETAS with the conserved motif of TGCAT. The same phenomenon was reported in sinipercine fishes. In the central conserved sequence block domain, we identified the CSB-F, CSB-E, and CSB-D regions. The consensus sequence of CSB-F is "ATGTA-TAAGAA-CGACCA", which serves to differentiate the central conserved sequence block domain from the termination associated sequence domain. CSB-E is located downstream of CSB-F, and its consensus sequence is "AGGGAGA-GTGGGGGT", characterized by the GTGGGG box. CSB-E is followed by CSB-D (consensus sequence "TATTCCTGGCATTTGGTTC-CTA-TTCAGG"). CSB-D is highly conserved in fishes and might function in the regulation of the H-strand replication, in the initiation of the D-loop structure, and might be involved in the mitochondrial metabolism [50]. The consensus sequences of CSB-F, CSB-E, and CSB-D in Scorpaeniformes fishes are highly conserved and consistent with those described in other fishes. CSB-1 is associated with the initiation of mitochondrial DNA duplication; it is highly conserved in mammals but is variable in fishes. The motif GACATA is the most conserved. Although we identified GACATA in Scorpaeniformes fishes, but the consensus sequence of CSB-1 could not be recognized since the segments corresponding to CSB-1 could not be aligned with each other. No conserved sequence was detected in these segments, except for the GACATA, what suggests a more rapid evolution of this domain in comparison to the others in the control region, and might provide information for dissecting the structure/function relationships of the control region. The CSB-2 (TAAACCCCCCTACCCCCCTAA) and CSB-3 (TGAAAACCCCCCGGAAACAGGA) sequences were 100% conserved within the Scorpaeniformes fishes, and is consistent with the CSB-2 and CSB-3 identified in other fishes [6]. The highly conserved natures of the sequences in the central and conserved sequence block domains imply that the structure and function of the control region is mostly determined by these two domains (Fig. 2). The gene order is almost the same in these eight Scorpaeniformes fishes, except for *S. marmoratus* and *S. schlegelii*, who have the $tRNA^{Cys}$ gene encoded by the H-strand, whereas this gene is encoded by the L-strand in most fishes. The overall base composition of the marbled rock fish mitochondrial genome was estimated to be 28.7% for A; 28.1% for C; 26.7% for T; and 16.5% for G respectively, indicating that a composition bias is present in the marbled rock fish. The overall A + T content is 55.4%, what is higher than in the other seven Scorpaeni-

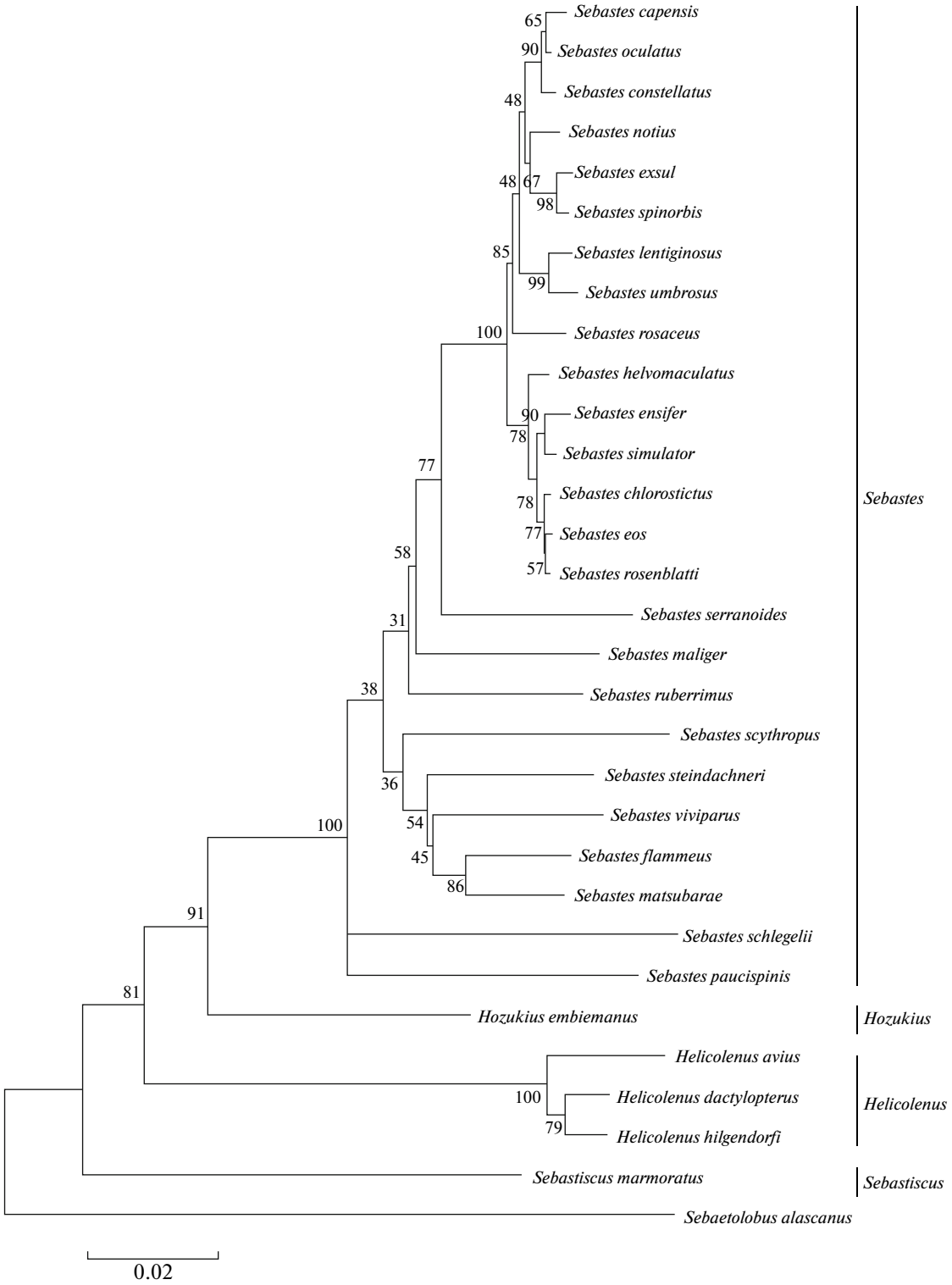

Fig. 4. The phylogenetic tree of the Sebastinae based on the NJ analysis of complete Cyt *b* gene sequences. *Sebastolobus alascanus* was used as an outgroup. The numbers on the branches are the bootstrap values. Reference for the 30 Cyt *b* gene sequences included are: *Helicolenus avius* (DQ678505), *H. dactylopterus* (EU036442), *H. hilgendorfi* (AP002948), *Hozukius emblemarius* (DQ678499), *Sebastes capensis* (AF031503), *S. chlorostictus* (AF031504), *S. constellatus* (AF031505), *S. ensifer* (AF031507), *S. eos* (AF031506), *S. exsul* (AF031514), *S. flammeus* (AB126388), *S. helvomaculatus* (AF031508), *S. lentiginosus* (AF031509), *S. maliger* (AF031500), *S. matsubarae* (AB126390), *S. notius* (AF031510), *S. oculatus* (AF031502), *S. paucispinis* (AF031499), *S. rosaceus* (AF031512), *S. rosenblatti* (AF031511), *S. ruberrimus* (AF031501), *S. schlegelii* (AY491978), *S. scythropus* (AB126389), *S. serranoides* (AF031498), *S. simulator* (AF031513), *S. spinorbis* (AF031515), *S. steindachneri* (AB126387), *S. umbrosus* (AF031516), *S. viviparus* (EU492276), *Sebastolobus alascanus* (AF031497).

Table 5. Mitochondrial genomes of reported Scorpaeniformes

Species (Acc. Ns.)	Size (bp)	A + T (%)	rRNA length (bp)		Control region		Protein
			12S	16S	Length	A + T (%)	A + T (%)
A. ventricosus (NC_008129)	15974	54.4	944	1693	112	88.4	54.1
S. schlegeli (AY491978)	16525	53.8	943	1713	833	64.8	53.5
H. hilgendorfi (AP002948)	16728	54.2	946	1692	1064	65.6	53.7
S. amiscus (AP004441)	16526	54.5	946	1699	856	61.6	54.0
D. tiltoni (AP004440)	16751	53.6	955	1689	959	67.1	52.6
C. reinii (AP004442)	16561	52.1	945	1688	854	60.0	51.5
D. peterseni (AP002947)	16717	53.6	956	1686	932	66.1	52.9
S. marmoratus (GU452728)	17301	55.4	943	1692	1645	68.7	54.5

formes fishes. The control region has a higher A + T content (68.7%) than the overall genome content of *S. marmoratus*, a feature that has been earlier reported in other Scorpaeniformes fishes.

The two ribosomal genes are 943 bp and 1692 bp in length, respectively. The 12S ribosomal gene is of the same size as the corresponding gene in *S. schlegelii*, and shorter than its counterparts in *A. ventricosus*, *H. hilgendorfi*, *S. amiscus*, *D. tiltoni, D. peterseni,* and *C. reinii*. The 16s is shorter than its counterparts in *A. ventricosus, S. schlegelii,* and *S. amiscus*, but longer than those in *D. tiltoni, C. reinii,* and *D. peterseni* (Table 5). No unusual features were detected in the tRNA genes of *S. marmoratus*, such as the absence of the $tRNA^{Pro}$ gene, which was found in *T. japonicus* and *T. nanhaiensis* [6]. The 22 tRNA genes showed the typical arrangement found in vertebrates. Although numerous non-complementary base pairs were found in the stems, all of the tRNAs could be folded into the typical cloverleaf secondary structure. The majority of these postulated secondary structures generally contain seven bp in the aminoacyl stem, 4 bp in the TΨC stem, five bp in the anticodon stem, and five in the DHU stem. The most differentiated is the $tRNA^{Ser}$ (AGY), which does not contain the typical aminoacyl stem. The usage of initiation codons in all mitochondrial protein-coding genes of *S. marmoratus* is identical to the other fishes investigated. ATG is used for 12 genes (ND1, ND2, CO2, ATPase 8, ATPase 6, CO3, ND3, ND4L, ND4, ND5, ND6, and Cyt *b*), while the CO1 gene uses GTG as the initiation codon. Interestingly, although *H. hilgendorfi* is more closely related to *S. schlegelii* than to *S. marmoratus*, the use of stop codons in *S. marmoratus* is more similar to that of *S. schlegelii*, and among these eight species, only *A. ventricosus* and *C. reinii* have the identical usage of stop codons. The nucleotide distribution at the three codon positions differs significantly in the protein-coding genes, where the nucleotide G shows the most bias; its overall frequency is only 15.4%. The A + T content is also significantly biased at different codon positions; the frequency of A + T varies from 47.4 to 58.4%. At the first position, the percentage of purines (51.2%) was slightly higher than that of pyrimidines (48.9%), whereas at the second position, pyrimidines were over-represented with regard to purines (T + C = 67%).

Barsukov and Chen [8] regarded *Sebastiscus* as a subgenus of *Sebastes*. *Sebastiscus* and *Sebastes* are, however, distantly related, and, according to a previous study, *Hozukius* and *Helicolenus* are more closely related to *Sebastes* than to *Sebastiscus*. Ishii [14] showed that *Helicolenus* was more closely related to *Sebastes* than *Sebastiscus* on the basis of sequence variations in the mitochondrial DNA of the cytochrome *c* oxidase submit 1. Similar studies conducted by Yoshiaki at al. [15] and John and Rusell [16] also showed the same phylogenetic placement of *Sebastiscus* within the *Sebastinae* subfamily. Our results support the view that *Sebastiscus* should be treated as a genus independent from *Sebastes*, which corresponds to the phylogenetic analysis using the sequence data of the mitochondrial genome, but conflicts with the traditional morphology view. However, the low bootstrap values suggest that the phylogenetic relationships of the major clusters in *Sebastes* were poorly resolved by the molecular data. Such uncertainties may have been due to rapid speciation over a relatively short period of time [51], and in the closely related species, cyt *b* sequences may lack the resolving power. To clarify the ambiguous relationships of this genus, a more comprehensive study utilizing more variable molecular markers or nucleotide data with faster rates of evolution is necessary.

ACKNOWLEDGMENTS

This study was supported by the Research Foundation of the Zhejiang province (2008C22026, 20070430 and 2008R40G2110003) and by the National Technology R&D Program of China (2007BAD43B08).

REFERENCES

1. Kita J., Tsuchida S., Setoguma T. 1996. Temperature preference and tolerance, and oxygen consumption of the marbled rockfish, *Sebastiscus marmoratus*. *Mar. Biol.* **123**, 467–471.
2. Hixson J.E., Wong T.W., Clayton D.A. 1986. Both the conserved stem-loop and divergent 5'-flanking se-

quence are required for initiation at the human mitochondrial origin of light-strand DNA replication. *J. Biol. Chem*. **261**, 2384–2390.

3. Elmerot C., Janke A., Gojobori T., Arnason U. 2002. The mitochondrial genome of the pufferfish, *Fugu rubripes*, and ordinal teleostean relationships. *Gene*. **295**, 163–172.
4. Fedorova S.A., Stepanov A.D., Adojaan M., Parik J., Argunov V.A., Ozawa T., Khusnutdinova E.K., Villems R. 2008. Phylogenetic analysis of ancient mitochondrial DNA lineages of human remains found in Yakutia**.** *Mol. Biol. (Mosk)*. **42**, 391–398
5. Moore W.S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial gene trees versus nuclear-gene trees. *Evolution*. **49**, 718–726.
6. Liu Y., Cui Z.X. 2009. The complete mitochondrial genome sequence of the cutlassfish *Trichiurus japonicus* (Perciformes: Trichiuridae): Genome characterization and phylogenetic considerations. *Marine Genomics*. **2**, 133–142.
7. Mayden R.L., Chen W.-J., Bart H.L., et al. 2009. Reconstructing the phylogenetic relationships of the earth's most diverse clade of freshwater fishes-order Cypriniformes (Actinopterygii: Ostariophysi): A case study using multiple nuclear loci and the mitochondrial genome. *Mol. Phylogenet. Evol.* **51**, 500–514.
8. Barsukov V.V., Chen L.C. 1978. Review of the subgenus *Sebastiscus* (*Sebastes*, Scorpaenidae) with a description of a new species. *J. Ichthyol*. **18**, 179–193.
9. Barsukov V.V. 1981. A brief review of the subfamily Sebastinae. *J. Ichthyol*. **21**, 1–26.
10. Barsukov V.V. 1988. Rockfishes of the *Sebastes inermis* complex of the subgenus *Sebastodes* (Sebastes, Scorpaenidae). *J. Ichthyol.* **31**, 1–23.
11. Chen L.C. 1985. A study of the *Sebastes inermis* species complex with delimitation of subgenus *Mebarus* (Pisces, Scorpaenidae). *J. Taiwan Mus*. **38**, 23–37.
12. Chen L.C. 1986. Meristic variation in *Sebastes* (Scorpaenidae), with an analysis of character association and bilateral pattern and their significance in species separation. *NOAA Tech. Rep. NMFS*. **45**, 1–17.
13. Kendall J.R., Arthur W. 1991. Systematics and identification of larvae and juveniles of the genus *Sebastes*. *Environ. Biol. Fishes*. **30**, 173–190.
14. *Ishii T.* 1997. The study of speciation from analysis of mitochondrial DNA structures among the fishes in Scorpaenidae (in Japanese with English abstract). *Sophia Life. Sci. Bull*. **16**, 105–111.
15. Yoshiaki K., Nakayama K., Nakabo T. 2003. Molecular phylogenetic perspective on speciation in the genus *Sebastes* (Scorpaenidae) from the Northwest Pacific and the position of *Sebastes* within the subfamily Sebastinae. *Ichthyol. Res*. **50**, 239–244
16. John R.H., Rusell D.V. 2007. The origin, evolution and diversification of rockfishes of the genus *Sebastes* (Cuvier). *Mol. Phylogenet. Evol*. **44**, 790–811.
17. Xu T.J., Chen S.L., Ji X.S., Tian Y.S. 2008. MHC polymorphism and disease resistance to Vibrio anguillarum in 12 selective Japanese flounder (*Paralichthys olivaceus*) families. *Fish. Shellfish. Immun*. **25**, 213–221.
18. Kim I.C., Lee J.S. 2004. The complete mitochondrial genome of the rockfish *Sebastes schlegelii* (Scorpaeniformes, Scorpaenidae). *Mol. Cells.* **17**, 322–328.
19. Miya M., Kawaguchi A., 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*. **18**, 1993–2009.
20. Wyman S.K., Jansen R.K., Boore J.L. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. **20**, 3253–3255.
21. Tamura K., Dudley J., Nei M., Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Bio. Evol*. **24**, 1596–1599.
22. Lowe T.M., Eddy S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genome sequence. *Nucl. Acids Res*. 25, 955–964.
23. Xia X., Xie Z. 2001. DAMBE: Data analysis in molecular biology and evolution. *J. Hered*. 92, 371–373.
24. Xiao H., Chen S.Y. 2005. Molecular phylogeny of *Sinocyclocheilus* (Cypriniformes: Cyprinidae) inferred from mitochondrial DNA sequence. *Mol. Phylogenet. Evol*. **36**, 67–77.
25. Saitou N., Nei M. 1987. The neighbor-joining method, a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. **4**, 406–25.
26. Broughton R.E., Milam J.E., Roe B.A. 2001. The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Res.* **11**, 1958–1967.
27. Delarbre C., Spruyt N., Gallut C., Delmarre C., Barriel V., Janvier P., Laudet V., Gachelin G. 1998. The complete nucleotide sequence of the mitochondrial DNA of the dogfish, *Scyliorhinus canicula*. *Genetics.* **150**, 331–344.
28. Ojala D., Montoya J., Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature*. **290**, 470–474.
29. Janke A., Erpenbeck D., Nilsson M., Arnason U. 2001. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. *Proc. R. Soc. Lond., B Biol. Sci.* **268**, 623–631
30. Zhang P., Chen Y.Q., Liu Y.F., Zhou H., Qu L.H. 2003. The complete mitochondrial genome of the Chinese giant salamander, *Andrias davidianus* (Amphibia: Caudata). *Gene*. **311**, 93–98.
31. Naylor G.J., Collins T.M., Brown W.M. 1995. Hydrophobicity and phylogeny. *Nature*. **373**, 555–556.
32. Peng Z., Wang J., He S. 2006. The complete mitochondrial genome of the helment catfish *Cranoglanis bouderius* (Siluriformes: Cranoglanididae) and the phylogeny of otophysan fishes. *Gene*. **376**, 290–297.
33. Miya M., Nishida M. 2000. Use of mitogenome information in teleostean molecular phylogenetics: tree-based exploration under the maximum-parsimony optimality criterion. *Mol. Phylogenet. Evol*. **17**, 437–455.
34. Lee W.J., Kocher T.D. 1995. Complete sequence of a sea lamprey (*Pettromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. *Genetics*. **139**, 873–887.

35. Noack K., Zardoya R., Meyer A. 1996. The complete mitochondrial DNA sequence of the bichir (*Polypterus ornatipinnis*), a basal ray-finned fish: Ancient establishment of the consensus vertebrate gene order. *Genetics*. **144**, 1165–1180.

36. Ohtsuki T., Kawai G., Watanabe K. 2002. The minimal tRNA: unique structure of *Ascaris suum* mitochondrial tRNAser-UCU having a short T arm and lacking the entire D arm. *FEBS Lett*. **514**, 37–43.

37. Zardoya R., Garrido-Pertierra A., Bautista J.M. 1995. The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*. *J. Mol Evol*. **41**, 942–951.

38. Johansen S., Guddal P.H., Johansen T. 1990. Organization of the mitochondrial genome of atlantic cod, *Gadus morhua*. *Nucl. Acids Res*. 18, 411–419.

39. Taanman J.W. 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochim. Biophys. Acta*. **1410**, 103–123.

40. Wang T.W., Clayton D.A. 1985. *In vitro* replication of human mitochondrial DNA: Accurate initiation at the origin of light-strand synthesis. *Cell*. **42**, 951–958.

41. Hixson J.E., Brown W.M. 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution and phylogenetic implications. *Mol. Biol. Evol*. **3**, 1–18.

42. Sbisa E., Tanzariello F., Reyes A., Pesole G., Saccone C. 1997. Mammalian mitochondrial D-loop region structure analysis: identification of new conserved sequence and their functional and evolutionary implications. *Gene*. **205**, 125–140.

43. Southern S.O., Southern P.J., Dizon A.E. 1998. Molecular characterization of a cloned dolphin mitochondrial genome. *J. Mol. Evol*. **28**, 32–40.

44. Brought R.E., Dowling T.E. 1994. Length variation in mitochondrial DNA of the minnow *Cyprinella spiloptera*. *Genetics*. **138**, 179–190.

45. Higuchi M., Kato K. 2002. Sequence variability in the mitochondrial DNA control region of five *Sebastes* Species. *Fish. Sci*. **68**, 643–650.

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

47. Faber J.E., Stepien C.A. 1998. Tandemly repeated sequence in the mitochondrial DNA control region and phylogeography of the Pike-Perches stizostedion. *Mol. Phylogenet. Evol*. **10**, 310–322.

48. Brought R.E., Dowling T.E. 1997. Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of minnow *Cyprinella spiloptera*. *Mol. Biol. Evol*. **14**, 1187–1196.

49. Miya M., Takeshima H., Endo H., 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.

50. Clayton D.A. 1982. Replication of animal mitochondrial DNA. *Cell*. **28**, 693–705.

51. Rocha-Olivares A., Rosenblatt R.H., Vetter R.D. 1999. Molecular evolution, systematics, and zoogeography of the rockfish subgenus *Sebastomus* (Sebastes, Scorpaenidae) based on mitochondrial cytochrome *b* and control region sequences. *Mol. Phylogenet. Evol*. **11**, 441–458.