Available online at www.sciencedirect.com

ScienceDirect

Gene 399 (2007) 11–19

GENE

www.elsevier.com/locate/gene

The complete mitochondrial genome of the Chinese hook snout carp *Opsariichthys bidens* (Actinopterygii: Cypriniformes) and an alternative pattern of mitogenomic evolution in vertebrate

Xuzhen Wang [a], Jun Wang [b,c,d], Shunping He [a,*], Richard L. Mayden [e]

[a] *Laboratory of Fish Phylogenetics and Biogeography, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China*
[b] *Beijing Institute of Genomics of Chinese Academy of Sciences, Beijing Genomics Institute, Beijing Proteomics Institute, Beijing 101300, China*
[c] *Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK-5230, Odense M, Denmark*
[d] *The Institute of Human Genetics, University of Aarhus, DK-8000 Aarhus C, Denmark*
[e] *Department of Biology, 3507 Laclede Ave., Saint Louis University, St. Louis, MO 63103, USA*

Received 31 January 2007; received in revised form 12 April 2007; accepted 18 April 2007
Available online 27 April 2007

Abstract

The complete mitochondrial genome sequence of the Chinese hook snout carp, *Opsariichthys bidens*, was newly determined using the long and accurate polymerase chain reaction method. The 16,611-nucleotide mitogenome contains 13 protein-coding genes, two rRNA genes (12S, 16S), 22 tRNA genes, and a noncoding control region. We use these data and homologous sequence data from multiple other ostariophysan fishes in a phylogenetic evaluation to test hypothesis pertaining to codon usage pattern of *O. bidens* mitochondrial protein genes as well as to re-examine the ostariophysan phylogeny. The mitochondrial genome of *O. bidens* reveals an alternative pattern of vertebrate mitochondrial evolution. For the mitochondrial protein genes of *O. bidens*, the most frequently used codon generally ends with either A or C, with C preferred over A for most fourfold degenerate codon families; the relative synonymous codon usage of G-ending codons is greatly elevated in all categories. The codon usage pattern of *O. bidens* mitochondrial protein genes is remarkably different from the general pattern found previously in the relatively closely related zebrafish and most other vertebrate mitochondria. Nucleotide bias at third codon positions is the main cause of codon bias in the mitochondrial protein genes of *O. bidens*, as it is biased particularly in favor of C over A. Bayesian analysis of 12 concatenated mitochondrial protein sequences for *O. bidens* and 46 other teleostean taxa supports the monophyly of Cypriniformes and Otophysi and results in a robust estimate of the otophysan phylogeny.

Keywords: Cyprinidae; Ostariophysi; Mitochondrial genome; Evolution

1. Introduction

In most metazoans the extranuclear mitochondrial DNA (mtDNA) is generally a small genome ranging in size from 15 kb to 20 kb. This single circular genome encodes genes for 13 protein subunits of the enzymes of oxidative phosphorylation as well as genes for two rRNAs of the mitochondrial ribosome and the 22 tRNAs necessary for the translation of the proteins encoded by mtDNA (Boore, 1999). Structurally, most animal mitochondrial genomes contain the same 37 genes (Boore, 1999), and the gene order is highly conserved in vertebrates with a few exceptions, e.g. in certain fish (Miya and Nishida, 1999) and amphibian (Liu et al., 2005; Zhang et al., 2005) species. The mitochondrial genome also contains a significant noncoding sequence, the control region, which is involved in regulation of transcription and replication (Shadel and Clayton, 1997).

Complete mitochondrial genomes from numerous vertebrate species have now been determined. Because of their small size and relative autonomy from the nucleus, mitochondrial genomes have

Abbreviations: ATPase 6 and 8, ATPase subunits 6 and 8; bp, base pair(s); COI–III, cytochrome coxidase subunits I–III; cyt *b*, cytochrome *b*; LA PCR, long and accurate polymerase chain reaction; ND1–6, 4L, NADH dehydrogenase subunits 1–6, 4L; rRNA, ribosomal RNA; TAS, termination associated sequence; tRNA, transfer RNA.

* Corresponding author. Institute of Hydrobiology, Chinese Academy of Sciences, 7th Donghu Nanlu, Wuhan 430072, China. Tel.: +86 27 68780430; fax: +86 27 68780071.
E-mail address: clad@ihb.ac.cn (S. He).

0378-1119/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.gene.2007.04.019

proven to be valuable windows on the process of genome evolution (Broughton et al., 2001; Gray et al., 1999) and with respect to the correlation between codon usage and codon–anticodon adaptation (Xia, 2005). Mitochondrial sequences have been the most widely used markers in molecular phylogenetic and phylogeographic analyses due to their mode of maternal inheritance and relative lack of recombination. Furthermore, complete mtDNA sequences have provided valuable insights into phylogenetic problems of many different metazoan groups at various taxonomic levels (Boore and Brown, 1998; Kawaguchi et al., 2001; Lavoue et al., 2005; Mindell et al., 1998; Miya and Nishida, 2000; Miya et al., 2003; Naylor and Brown, 1998; Rasmussen and Arnason, 1999; Zardoya and Meyer, 1997).

In this study, the complete mtDNA sequence of the Chinese hook snout carp, *Opsariichthys bidens*, was newly determined. *Opsariichthys* is a member of the order Cypriniformes (minnows, loaches, and suckers), a large group of primary freshwater fishes distributed throughout North American, Africa and Eurasia. Cypriniformes is in turn placed within the series Otophysi, which also includes the order Characiformes (leporins and piranhas), Siluriformes (catfishes), and Gymnotiformes (South American knifefishes). The series Otophysi and the order Gonorynchiformes (= Anotophysi) further compose Ostariophysi (Fink and Fink, 1981; Nelson, 1994). Within Telesotei, Ostariophysi is one of the most diversified groups and comprises about 93% of primary freshwater fish species (Berra, 2001; Nelson, 1994). The piscivorous minnow, *O. bidens*, is one of the most widespread species of eastern Asia. It occurs almost in all the drainages of main rivers across China. The genus *Opsariichthys* is placed within the family Cyprinidae and morphologically is considered a primitive cyprinid (Howes, 1980; Regan, 1911).

Here we describe the *O. bidens* mitochondrial genome with respect to gene content and organization, codon usage, nucleotide composition, and putative functional motifs. We aim to test evolutionary hypotheses pertaining to nucleotide and amino acid composition, genome-wide patterns of variability, and evolutionary rates among major ostariophysan lineages. We also compare the pattern of codon usage bias in *O. bidens* with the common pattern in ostariophysans and vertebrates, and we analyze the mechanism of codon usage bias in the course of the *O. bidens* mtDNA evolution. According to a phylogenetic framework based on the mitochondrial genomes of *O. bidens* and 46 other teleostean taxa, we want to characterize *O. bidens* relative to other cypriniforms and we re-examine phylogenetic relationships within the Ostariophysi.

2. Materials and methods

2.1. DNA extraction, LA PCR and sequencing

Total genomic DNA of *O. bidens* was extracted from the muscle tissue using a QIAamp tissue kit following the manufacture's protocol. The mitochondrial genome DNA of *O. bidens* was amplified in its entirety using a LA PCR technique (Miya and Nishida, 1999). The primers designed by Miya and Nishida (2000), and Inoue et al. (2000, 2001a,b) were used to amplify the total mitochondrial genome in two reactions. LA PCR was done in a PTC-100 programmable thermal controller (MJ Research, USA) and reactions were carried out with 30 cycles of a 25 μl reaction volume containing 8.75 μl sterilized distilled water, 2.5 μl 10 × LA PCR buffer II, 4.0 μl deoxy nucleoside triphosphate (dNTP) (5 mM), 2.5 μl $MgCl_2$ (25 mM), 2.5 μl each primer (5 uM), 0.25 μl 2.5 units/μl LA Taq polymerase (Takara), and 5.0 μl template containing approximately 20 ng DNA. The thermal cycle profile was: pre-denaturation at 94 °C for 2 min, and denaturation at 98 °C for 10 s, annealing and extension combined at the same temperature (68 °C) for 16 min, 72 °C for 5 min to denature the Taq polymerase. The long PCR products were electrophoresed on a 0.8% agarose gel and diluted in sterilized distilled water for subsequent use as PCR templates.

We used 24 different primers that amplify contiguous, overlapping segments to get the entire mitochondrial genome of *O. bidens* (Table 2). Some of these primers were versatile, designed from the complete mitochondrial genome of six bony fish species according to Miya and Nishida (2000). The others were specific for *O. bidens*, and designed from the sequence that had been got from the versatile primers. PCR was done in a PTC-100 programmable thermal controller, and reactions were carried out with 30 cycles of 25 μl reaction volume containing 15.5 μl sterilized distilled water, 2.5 μl 10 × PCR buffer, 2.0 μl dNTP (5 mM), 1.8 μl $MgCl_2$ (25 mM), 1.0 μl each primer (5 μM), 0.2 μl 5 units/μl Taq polymerase (Takara), and 1.0 μl long PCR products as template. The thermal cycle profile was: pre-denaturation at 94 °C for 2 min, and denaturation at 94 °C for 15 s, annealing at 52 °C for 15 s, extension at 72 °C for 30 s, and 72 °C for 5 min to denature the Taq polymerase. The PCR products were electrophoresed on a 1.0% agarose gel.

Double-stranded PCR products purified by filtration through Millipore plates were subsequently used for direct cycle sequencing with dye-labeled terminators (ABI). Primers used were the same as those for PCR. All sequencing reactions were performed according to the manufacturer's instructions. Labeled fragments were analyzed on a model MegaBACE 1000 DNA sequencer (GE Healthcare Biosciences, USA).

2.2. Sequence analyses

Annotation of the *O. bidens* mitochondrial genome was based on comparisons to genomes of other cyprinid fishes, including the carp (*Cyprinus carpio*, Chang et al., 1994), the goldfish (*Carassius auratus*, Murakami et al., 1998), and the zebrafish (*Danio rerio*, Broughton et al., 2001). In addition, identification of tRNA genes was verified using the program tRNAscan-SE (Lowe and Eddy, 1997). The potential stem–loop secondary structures within these tRNA gene sequences were calculated using the tRNAscan-SE Search Server available online (http://lowelab.ucsc.edu/tRNAscan-SE/).

The coding genes for the 13 mitochondrial proteins from 39 ostariophysan species and subspecies (Table 1) were concatenated and aligned using Clustal W (Thompson et al., 1994) and corrected by eye to preserve reading frame. The sequence for *ND6* gene is the reverse complement to maintain a consistent reading frame. Codon usage and nucleotide and amino acid frequencies were calculated using the program DAMBE (Xia and Xie, 2001) and relative rate tests were conducted with the program HyPhy (Pond et al., 2005).

Table 1
List of fish species analyzed in this study

Classification	Species	Accession no.
Cypriniformes		
Cyprinidae	*Cyprinus carpio*	X61010 (Chang et al., 1994)
	Carassius auratus	AB006953 (Murakami et al., 1998)
	Carassius auratus auratus	AB111951 (Murakami et al., unpublished)
	Carassius carassius	AY714387 (Guo et al., unpublished)
	Coreoleuciscus splendidus	DQ347951 (Lim et al., unpublished)
	Hemibarbus labeo	DQ347953 (Lim et al., unpublished)
	Hemibarbus longirostris	DQ347952 (Lim et al., unpublished)
	Hemibarbus mylodon	DQ345787 (Lim et al., unpublished)
	Sarcocheilichthys variegatus microoculus	AB054124 (Saitoh et al., 2003)
	Phenacobius mirabilis	DQ536431 (Broughton and Reneau, 2006)
	Notropis stramineus	DQ536429 (Broughton and Reneau, 2006)
	Chondrostoma lemmingii	DQ536427 (Broughton and Reneau, 2006)
	Gila robusta	DQ536424 (Broughton and Reneau, 2006)
	Cyprinella spiloptera	DQ536422 (Broughton and Reneau, 2006)
	Campostoma anomalum	DQ536421 (Broughton and Reneau, 2006)
	Rhodeus uyekii	DQ155662 (Kim et al., unpublished)
	Opsariichthys bidens	DQ367044 (this study)
	Danio rerio	AC024175 (Broughton et al., 2001)
Balitoridae	*Crossostoma lacustre*	M91245 (Tzeng et al., 1990)
	Lefua echigonia	AB054126 (Saitoh et al., 2003)
Catostomidae	*Myxocyprinus asiaticus*	AY986503 (Peng et al., 2006)
	Minytrema melanops	DQ536432 (Broughton and Reneau, 2006)
	Carpiodes carpio	AY366087 (Broughton et al., unpublished)
Cobitidae	*Cobitis sinensis*	AY526868 (Chen and Wang, unpublished)
	Cobitis striata	AB054125 (Saitoh et al., 2003)
Characiformes		
Alestiidae	*Phenacogrammus interruptus*	AB054129 (Saitoh et al., 2003)
Characidae	*Chalceus macrolepidotus*	AB054130 (Saitoh et al., 2003)
Gymnotiformes		
Apteronotidae	*Apteronotus albifrons*	AB054132 (Saitoh et al., 2003)
Eigenmanniidae	*Eigenmannia* sp.	AB054131 (Saitoh et al., 2003)
Siluriformes		
Bagridae	*Pseudobagrus tokiensis*	AB054127 (Saitoh et al., 2003)
Callichthyidae	*Corydoras rabauti*	AB054128 (Saitoh et al., 2003)
Ictaluridae	*Ictalurus punctatus*	AF482987 (Waldbieser et al., 2003)
Pangasiidae	*Pangasianodon gigas*	AY762971 (Jondeung and Sangthong, unpublished)
Gonorynchiformes		
Chanidae	*Chanos chanos*	AB054133 (Saitoh et al., 2003)
Gonorynchidae	*Gonorynchus greyi*	AB054134 (Saitoh et al., 2003)
Gonorynchiformes		
Kneriidae	*Cromeria nilotica*	AP007275 (Lavoue et al., 2005)
	Grasseichthys gabonensis	AP007277 (Lavoue et al., 2005)
	Kneria sp.	AP007278 (Lavoue et al., 2005)
	Parakneria cameronensis	AP007279 (Lavoue et al., 2005)
Phractolaemidae	*Phractolaemus ansorgii*	AP007280 (Lavoue et al., 2005)
Clupeiformes		
Chirocentridae	*Chirocentrus dorab*	AP006229 (Ishiguro et al., 2005)
Clupeidae	*Sardinops melanostictus*	AB032554 (Inoue et al., 2000)
	Jenkinsia lamprotaenia	AP006230 (Ishiguro et al., 2005)
Denticipitidae	*Denticeps clupeoides*	AP007276 (Lavoue et al., 2005)
Engraulidae	*Engraulis japonicus*	AB040676 (Inoue et al., 2001a,b)
Sundasalangidae	*Sundasalanx mekongensis*	AP006232 (Ishiguro et al., 2005)
Salmoniformes		
Salmonidae	*Salmo salar*	U12143 (Hurst et al., 1999)

2.3. Phylogenetic analyses

In addition to the newly determined sequence of the *O. bidens* mitochondrial genome, we also included in our phylogenetic analysis the mitochondrial genomes from 39 ostariophysan species and subspecies plus six clupeiform species and one outgroup species, *Salmo salar* (Table 1). The 12 proteins of mtDNA were concatenated and used for phylogenetic analyses. The ND6 protein was encoded by the opposite strand of mtDNA with quite different base and codon biases and was not used. All positions with gaps or ambiguous alignment were excluded. The total number of remaining codons is 3582.

Bayesian phylogenetic analyses were conducted with MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The mtREV24 amino acid substitution model (Adachi and Hasegawa, 1996) with a proportion of invariable sites and the discrete Γ distribution for among-site rate heterogeneity ($I+\Gamma$) was adopted in Bayesian analyses. For the concatenated protein matrix, two independent Bayesian analyses were performed for 500,000 generations by using four independent chains (one cold and three heated) and random starting trees. Parameter values and trees were sampled every 100 generations. The first 1000 samples were discarded as burn-in, and a majority-rule consensus tree calculated from the 4000 remaining trees was used to determine the posterior probabilities of clades.

3. Results

3.1. Mitochondrial genome of *O. bidens*

The total length of the complete *O. bidens* mitochondrial genome nucleotide sequence was 16,611 bp. The genome is composed of 13 protein-coding genes, two rRNA genes (12S and 16S), 22 tRNA genes and a noncoding control region.

All of *O. bidens* mitochondrial protein-coding genes use ATG as the initiation codon with the only exception of *COI*, which uses GTG as the start codon. Stop codons include five TAA and four TAG. The *COII*, *COIII*, *ND4*, and *Cytb* genes possess incomplete stop codons and show a terminal T or TA.

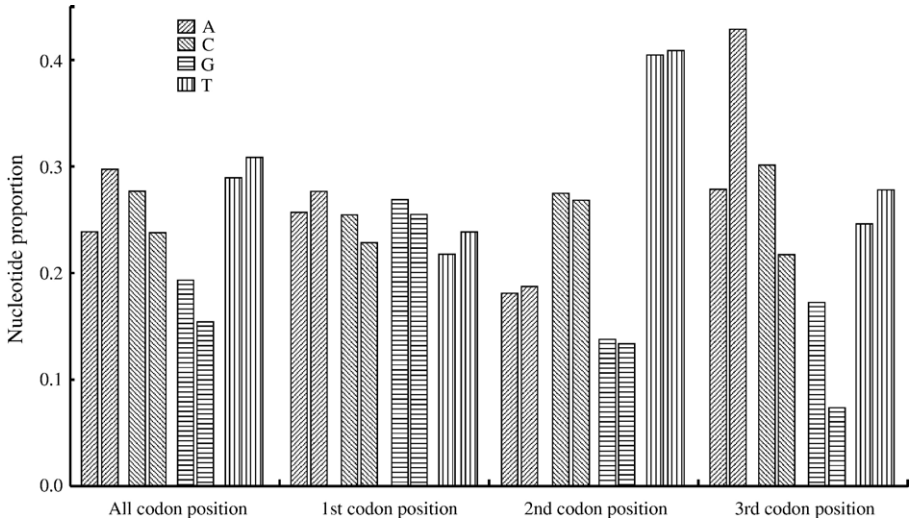

Fig. 1. Nucleotide proportions in the mitochondrial protein-coding genes of *Opsariichthys bidens* (left) and *Danio rerio* (right).

Of the 13 protein-coding genes, gene overlaps can be observed between four pairs of the contiguous genes, *ATP8–ATP6*, *ATP6–COIII*, *ND4L–ND4*, and *ND5–ND6*, and they overlap by seven, seven, one, and four nucleotides respectively.

Nucleotide composition of the *O. bidens* mtDNA protein-coding genes reflects a weak bias against G on the light strand, the sense strand for all protein genes except *ND6*, for which the heavy strand serves as the sense strand. In cyprinids including the zebrafish, the carp, and the goldfish, the strong bias against G is particularly marked at third codon positions of mtDNA protein genes (only 6–7% of sites are G). In *O. bidens*, however, G frequency of the mtDNA protein genes is found over 17% at the third codons (Fig. 1). For the protein genes, C is most frequent at third positions in *O. bidens* but A is most frequent at third positions in the zebrafish (Fig. 1). At the second codon positions of *O. bidens* mitochondrial protein-coding genes, pyrimidines are overrepresented compared with purines (C+T=68%).

Codon usage of the *O. bidens* mtDNA protein-coding genes is compared to that of eight other ostariophysan fishes in Table 3. For *O. bidens* mtDNA fourfold degenerate codons (NNN synonymous codon families, the IUB codes for the mixed bases refer to Table 2), the codon families of threonine, proline, serine, valine, and alanine end mostly with C or T (valine), but the glycine codon family ends mostly with G and the codon families of arginine and leucine mostly with A. Among twofold degenerate pyrimidine codons (NNY codon families), the frequency of codons ending in C appears to be somewhat higher than that of codons ending in T. All NNR codon families end mostly with A except methionine (where G is identical in frequency to A). Consistent with the overall bias against G, G is the least common third position nucleotide in all NNN codon families except for arginine (where G is similar in frequency to C and T but still less than A) and glycine (where G is most frequent). However, relative synonymous codon usage of G-ending codons in *O. bidens* are significantly higher than that in the zebrafish (Table 3).

The *O. bidens* mitochondrial genome contains 22 tRNA genes possessing anticodons that match the vertebrate mitochondrial genetic code. These mitochondrial tRNA genes are interspersed between ribosomal RNA and protein-coding genes and range in size from 68 to 77 nucleotides. Each tRNA sequence can be folded into a cloverleaf structure. All potential cloverleaf structures, except that of $tRNA^{Ser(AGY)}$, contain 7 bp in the amino acid stem, 5 bp in the anticodon and TψC stems, and 4 bp

Table 2
PCR and sequencing primers for *Opsariichthys bidens* designed from the complete mitochondrial genome of six bony fish species according to Miya and Nishida (2000)

Forward [a]	Sequence (5′ to 3′) [b]	Reverse [a]	Sequence (5′ to 3′) [b]
L709-12S	TAC ACA TGC AAG TCT CCG CA	H2009-16S	CCT AAG CAA CCA GCT ATA AC
L1969-16S	CGT CTC TGT GGC AAA AGA GTG G	H3058-16S	TCC GGT CTG AAC TCA GAT CAC GTA
L709-12S	TAC ACA TGC AAG TCT CCG CA	H3058-16S	TCC GGT CTG AAC TCA GAT CAC GTA
L2946-16S	GGG ATA ACA GCG CAA TC	H3934-ND1	GCG TAT TCT ACG TTG AAT CC
L3074-16S	CGA TTA AAG TCC TAC GTG ATC TGA GTT CAG	H5937-CO1	TGG GTG CCA ATG TCT TTG TG
L4633-ND2	CAC CGC CCW CGA GCA GTT GA	H8319-Lys	CAC CWG TTT TTG GCT TAA AAG GC
L8329-Lys	AGC GTT GGC CTT TTA AGC	H10035-Gly	CTT TCC TTG GGK TTT AAC CAA G
S3-6-F4	CCC CCC AAC TAA CCC TCC	H10035-Gly	CTT TCC TTG GGK TTT AAC CAA G
L9220-CO3	AAC GTT TAA TGG CCC ACC AAG C	H12293-Leu	TTG CAC CAA GAG TTT TTG GTT CCT AAG ACC
S3-16-F	CAA CTG TTC ATT GGC TGG GAG G	S3-16-R	ATG CTT GTG GTG TTT GCT TAT TCA G
S3-7-F5	GCT AAC TTG TTG ACT CCT TCC C	S3-19-R	GGT TTG CGG GGG TGA AG
L15180-Cyb	CAG ATA TCA TTC TGA GGT GCY ACA GT	H1552-12S	ACT TAC CGT GTT ACG ACT TGC CTC

[a] L and H denote light and heavy strands, respectively.
[b] Positions with mixed bases are labeled with their IUB codes: R=A/G; Y=C/T; K=G/T; M=A/C; S=G/C; W=A/T; N=A/G/C/T.

Table 3
Comparison of codon usage (number of codons) among *Opsariichthys bidens* and selected ostariophysan fishes

Amino acid	Codon[a]	Species[b]								
		1	2	3	4	5	6	7	8	9
Lys	AAA*	46	68	62	81	58	67	74	75	63
	AAG	31	7	16	8	18	13	8	7	15
Asn	AAC*	75	67	63	65	82	73	88	70	88
	AAT	38	59	50	56	33	47	34	57	29
Thr	ACA*	98	142	124	140	112	110	135	138	113
	ACG	35	13	19	13	32	9	7	14	20
	ACC	104	86	104	77	103	123	128	110	127
	ACT	60	55	36	63	42	52	54	59	38
Stop	AGA	0	0	0	0	0	0	1	0	0
	AGG	0	0	0	0	0	0	0	0	0
Ser	AGC*	44	30	40	38	42	37	43	46	48
	AGT	15	22	17	14	14	9	8	11	11
Met	ATA	84	129	117	139	101	113	140	143	73
	ATG*	84	54	55	52	70	59	27	59	78
Ile	ATC*	82	60	81	107	115	118	128	93	152
	ATT	185	222	186	201	154	179	166	190	114
Gln	CAA*	60	86	82	84	79	93	87	79	79
	CAG	36	8	16	10	21	7	3	18	19
His	CAC*	66	63	70	68	75	78	73	61	81
	CAT	36	42	34	34	32	31	32	40	21
Pro	CCA*	68	108	82	115	90	68	117	87	113
	CCG	25	4	15	13	22	8	12	17	12
	CCC	77	54	65	40	74	99	64	83	58
	CCT	44	47	49	40	28	34	23	34	32
Arg	CGA*	34	50	37	53	42	43	43	53	52
	CGG	18	6	17	9	12	6	6	8	8
	CGC	14	11	17	4	12	19	19	12	12
	CGT	10	10	4	9	10	6	10	3	4
Leu	CTA*	151	219	197	179	226	254	214	236	223
	CTG	96	38	60	29	98	51	30	43	68
	CTC	108	59	81	38	101	103	130	86	162
	CTT	129	130	135	146	87	92	122	102	108
Glu	GAA*	57	84	67	80	76	81	89	81	81
	GAG	45	17	33	20	29	15	12	16	19
Asp	GAC*	53	48	45	54	57	49	50	54	51
	GAT	25	31	31	30	20	29	22	19	24
Ala	GCA*	87	129	118	132	112	99	96	122	112
	GCG	38	12	23	13	28	15	6	13	22
	GCC	165	138	152	96	164	174	158	141	174
	GCT	56	56	55	87	57	54	73	51	52
Gly	GGA*	70	111	80	116	109	83	90	110	120
	GGG	74	42	58	45	57	49	34	48	48
	GGC	58	53	71	39	56	86	68	58	66
	GGT	44	36	38	36	23	26	44	26	15
Val	GTA*	72	112	97	111	90	85	80	103	86
	GTG	59	17	39	22	46	30	17	28	30
	GTC	51	31	37	21	55	46	38	31	76
	GTT	73	66	72	69	56	61	50	53	55
Stop	TAA	5	7	4	7	7	6	7	7	6
	TAG	4	3	7	3	3	3	1	2	4
Tyr	TAC*	61	48	54	57	78	64	68	48	77
	TAT	52	61	57	55	36	49	50	62	32
Ser	TCA*	61	82	75	108	70	61	75	79	67
	TCG	20	6	14	9	12	7	3	12	13
	TCC	62	51	43	19	70	81	76	46	71
	TCT	48	52	50	57	32	38	43	38	32
Trp	TGA*	79	108	98	107	82	92	113	101	104
	TGG	41	11	20	11	39	30	7	17	18
Cys	TGC*	10	11	15	15	16	17	19	20	22
	TGT	15	15	11	14	10	10	9	9	6
Leu	TTA*	90	132	131	181	76	112	121	125	55
	TTG	52	24	20	26	34	22	22	31	16
Phe	TTC*	119	106	110	88	117	118	141	108	149
	TTT	108	123	114	148	108	108	92	115	86

[a] Codons for which a tRNA with matching anticodon occurs in mitochondria are marked with asterisk.
[b] Species are indicated by numbers. 1, *Opsariichthys bidens*; 2, *Coreoleuciscus splendidus*; 3, *Gila robusta*; 4, *Danio rerio*; 5, *Myxocyprinus asiaticus*; 6, *Icatlurus punctatus*; 7, *Apteronotus albifrons*; 8, *Kneria* sp.; 9, *Chanos chanos*.

in the DHU stem. The *O. bidens* tRNA$^{Ser(AGY)}$ has no recognizable DHU stem and loop. As seen in other vertebrate tRNAs, numerous noncomplementary and T–G base pairs were found in tRNA stem regions. The 12S and 16S ribosomal RNA genes in *O. bidens* are 955 and 1694 nucleotides long, respectively. Both ribosomal RNA genes show great sequence similarity to ostariophysan fishes.

The control region, the largest noncoding region found in the *O. bidens* mtDNA, is 927 nucleotides long. The control region of *O. bidens* was much less similar to non-cyprinid ostariophysan fishes than were the coding sequences, with numerous nucleotide substitutions and insertions and deletions. The control region has an overall nucleotide composition that is rich in A and T (A+T=65.7%). The conserved sequence blocks CSB I–III, which are thought to be involved in positioning RNA polymerase both for transcription and for priming replication (Clayton, 1991; Shadel and Clayton, 1997), are found at the 3′ end of the control region. Another relatively conserved element is the TAS located at the 5′ end of the control region. Two putative TASs are identified in the *O. bidens* control region. Moreover, the palindromic sequence motifs TACAT and ATGTA are repeated three and two times, respectively at the 5′ end of the *O. bidens* control region. These motifs are thought to act as a signal for the termination of heavy strand elongation by forming stable hairpin–loop structure (Saccone et al., 1991) (Fig. 2). The secondary structure of TAS in *O. bidens* shows strong similarity to another East Asian cyprinid fish, the grass carp (*Ctenopharyngodon idellus*, Zhang et al., 1999).

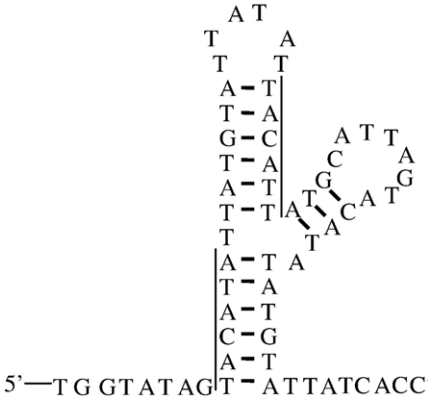

Fig. 2. The potential hairpin–loop structure of TAS at the 5′ end of the *O. bidens* control region.

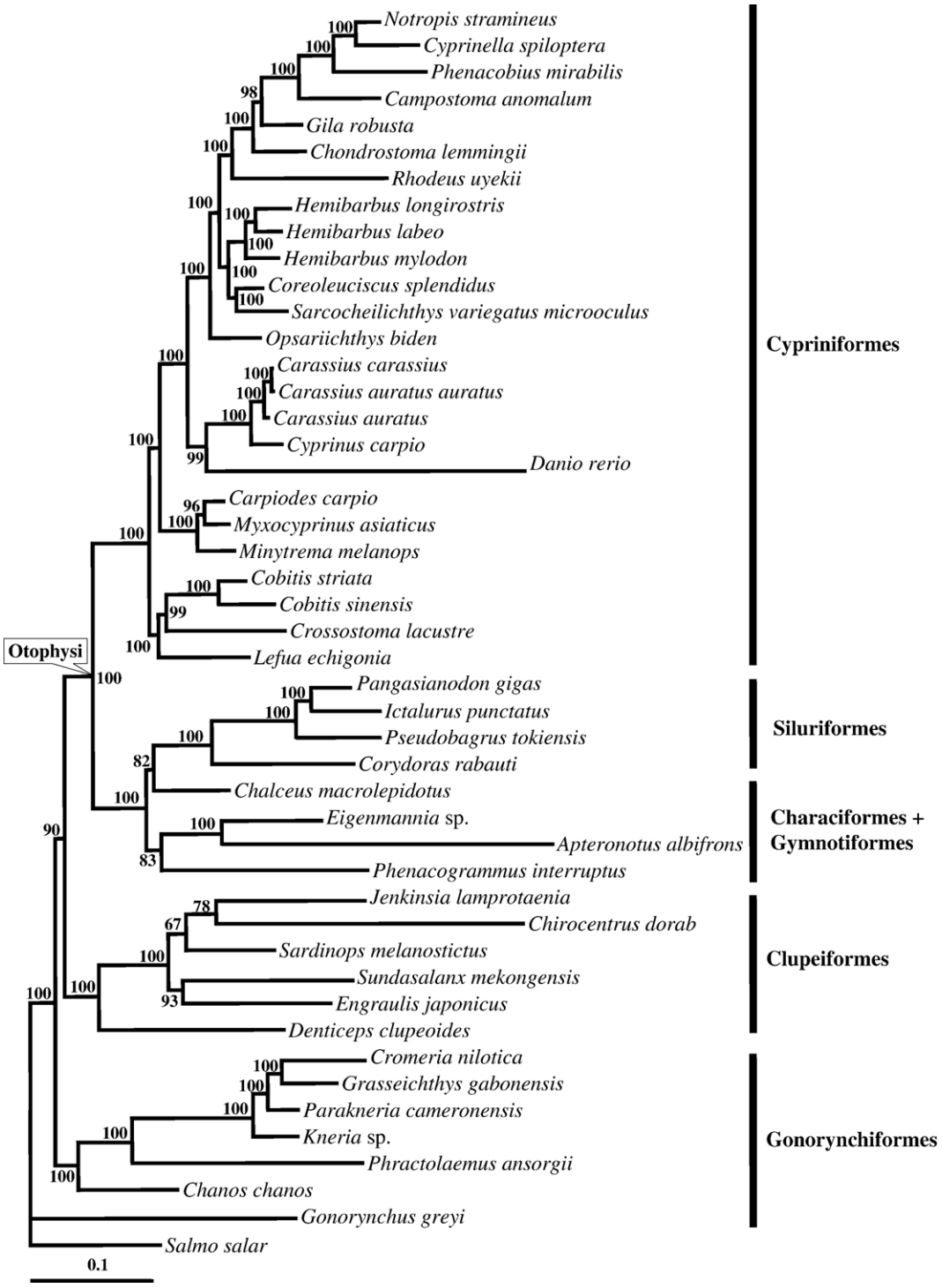

Fig. 3. The 50% majority-rule consensus tree resulting from Bayesian analysis of the concatenated 12 mitochondrial protein sequences dataset. Numbers at nodes represent posterior probabilities for Bayesian analysis.

3.2. Evolutionary patterns in ostariophysan mitochondrial genome

A strong correlation between regional nucleotide variation and amino acid variation for all protein genes is found across mitochondrial genome. Although the spatial pattern of variability is consistent over all positions, the magnitude of variability at first and second positions is consistently lower than at third positions. An extremely high magnitude of variability is found at all positions of *ATP8* gene. Rate of amino acid variation is the lowest for the *COI* gene and somewhat higher for others. A clear pattern of alternating regions of high and low variability is observed in all of the *ND* genes and *ATP6*. Across mitochondrial genomes, peaks of variability of nucleotides and amino acids coincide with extremes of hydropathy, either positive or negative.

Relative evolutionary rates were evaluated by comparing the amount of change in two ostariophysan lineages relative to the outgroup (*Salmo salar*). The test statistic is asymptotically chi-square distributed and can be used to determine whether the number of changes on two lineages is significantly different (Tajima, 1993). All comparisons (data not shown) among ostariophysan orders showed significant differences, and many interfamily comparisons were significantly different as well. However, rate differences within the family Cyprinidae are frequently as great as differences between the cypriniform families. Rates for *Carassius* and *Cobitis* are not significantly different, yet the rate for *Carassius* is significantly different from the rate for *Danio*. Within the family Cyprinidae, the rate of sequence evolution in *Danio* as well as in *Rhodeus* and in *Opsariichthys* significantly differs from the other analyzed cyprinids.

Nucleotide frequencies varied by codon positions and by taxa. At first and second positions, a clear trend of relative frequencies among ostariophysan taxa was T>C>A>G. At third positions, relative frequencies were A>C>T>G (in some cases, e.g. *Danio* and *Coreoleuciscus*, T is somewhat higher in frequency than C) in all cypriniform taxa except for *O. bidens*, where the relative frequencies were C>A>T>G. The *O. bidens* pattern of relative nucleotide frequencies at third positions is not unique within ostariophysans but is also observed in both clupeiform taxa and *Salmo salar* (data not shown). Although nucleotide frequencies among all ostariophysan taxa exhibit significant heterogeneity at first and third codon positions, tests of amino acid frequency indicate that no significant departure from homogeneity occurs within ostariophysan taxa.

3.3. Phylogenetic analysis

Bayesian analysis of the concatenated dataset of 12 mitochondrial proteins, with the mtREV+I+Γ model, produced a well resolved and well-supported phylogenetic tree (Fig. 3). In this tree, the morphology-based groups such as Cyprinidae, Catostomidae, Cypriniformes, Siluriformes, Clupeiformes, and Otophysi were recovered as monophyletic groups with posterior probability (PP) support of 100%, respectively. However, the currently recognized Gonorynchiformes did not form a monophyletic group; *Gonorynchus* was positioned outside other Gonorynchiformes at the basal position of the tree (Fig. 3). The association of the Clupeiformes and the Otophysi was supported by somewhat low PP value (90%).

Within the Otophysi, the Cypriniformes was sister group to the remaining otophysan orders (i.e., Siluriformes, Gymnotiformes, and Characiformes). Within the monophyletic Cypriniformes, the Cyprinidae appeared as sister group to the Catostomidae, and this clade formed sister group to the Balitoridae+Cobitidae.

4. Discussion

4.1. Mitochondrial genome evolution in the Ostariophysi

Despite identical gene order and content among *O. bidens* and the other ostariophysan taxa, the patterns of strand specific nucleotide bias and unequal codon usage observed in *O. bidens* are not conserved among the Ostariophysi. As seen in the zebrafish mitochondrial genome (Broughton et al., 2001), a common feature for the protein-coding genes is that NNY synonymous codon families end mostly in C and NNR and NNN codon families end mostly in A. Such a pattern of codon usage is also universal in vertebrates (Broughton et al., 2001; Xia, 2005). It is noteworthy that the codon bias of *O. bidens* mitochondrial protein-coding genes is quite different from that common codon bias pattern. In *O. bidens*, the most frequently used codons end with C for most NNN codon families, and the relative synonymous codon usage of codons ending in G is greatly elevated in all categories. The *O. bidens* pattern of nucleotide bias at third positions and codon usage bias is also found in other ostariophysan and clupeiform taxa (Table 3). These observations suggest that except for the general codon usage bias pattern found across vertebrates (Broughton et al., 2001; Xia, 2005), there is an alternative pattern of vertebrate mitogenomic evolution.

The reasonable explanations of unequal codon usage in the mitochondrial protein genes are that codon usage is generally biased toward the available tRNA species and codon bias is associated with the strand specific nucleotide bias in mtDNA (Broughton et al., 2001). Because there are only 22 tRNAs in the mitochondria, there is only one specific tRNA species for most amino acids. For each amino acid, only codons ending in A or C will be perfectly matched by a complementary tRNA anticodon. Thus in the zebrafish (Broughton et al., 2001) and most other ostariophysan fishes, the most frequently used codons for most amino acids are those with matching tRNAs. Although there may be some advantage to matched codons and anticodons in protein translation, the phenomenon is not universal, as seen in *O. bidens*. In *O. bidens*, most amino acids with NNN codon families are preferentially specified by codons that do not match the available tRNA (Table 3). This is due to the high number of codons ending in C. It therefore appears that nucleotide bias at third codon positions is the main cause of codon bias in ostariophysan mitochondrial protein-coding genes. Usually, the strand specific nucleotide bias particularly favors A over T at third codon positions of the vertebrate mtDNA protein genes (Broughton et al., 2001), but it dissimilarly favors C over A in *O. bidens* (Fig. 1).

The heavy strand is left as single strand for hours during the asymmetrical replication of vertebrate mtDNA (Clayton, 1982, 2000). Spontaneous hydrolytic deamination of both A and C (Lindahl, 1993; Sancar and Sancar, 1988) occurs frequently in human mitochondrial DNA (Tanaka and Ozawa, 1994). The hydrolytic conversion of C to U and the conversion of A to hypoxanthine generate C→T mutations and A→G mutations, respectively on the heavy strand (consequently G→A mutations and T→C mutations, respectively on the light strand) of mtDNA (Lindahl, 1993; Sancar and Sancar, 1988). Among these two types of spontaneous deamination, the C→T mutations generally occurs more frequently than the A→G mutations (Lindahl, 1993). Therefore, the G→A mutations happening at third positions on the light strand tend to accumulate and would lead to an increased frequency of A-ending codons and substantial bias against G-ending codons, as seen in the zebrafish and most vertebrates (Broughton et al., 2001; Xia, 2005). In *O. bidens* mitochondrion

protein genes, however, the associated T→C mutations happening at third positions on the light strand accumulate and lead to the codon bias toward C-ending codons. The difference in deamination rates of C and A is remarkable (Lindahl, 1993), and would hardly make the A→G mutations happen more frequently than the C→T mutations on the heavy strand of mtDNA. But the effect of mutation fixation can be amplified by different rates of DNA repair. The codon bias toward C-ending codons and the relatively increased frequency of G-ending codons in *O. bidens* mitochondrial protein genes are remarkably different from the common vertebrate codon usage pattern found in the zebrafish mitochondrion, probably because the repair reaction of the deaminated form of C might be significantly more efficient in *O. bidens* mitochondrion than in the zebrafish mitochondrion.

Relative rate tests indicate that evolutionary rates of mitochondrial protein genes vary considerably within and among the ostariophysan orders. Rate differences within the family Cyprinidae and among the cypriniform families show a complicated pattern, which apparently due to rate increases in certain cyprinid lineages. The fact that rates differ among cyprinids indicates that there are still unknown factors that remarkably influence rates of sequence evolution between the closely related taxa.

The other important finding is that nucleotide composition of the mtDNA protein genes varies widely both within and among the ostrariophysan orders, whereas amino acids composition tends to be conserved within the Ostariophysi. Coupled with significant heterogeneity of nucleotide composition, the amino acid frequency homogeneity within the Ostariophysi suggests that a majority of substitutions are occurring at third codon positions, where substitutions are less likely to alter encoded amino acids. The present study also shows that the high number of possible synonymous third position substitutions results in saturation at third positions of the mtDNA protein genes, which could be expected to negatively affect the phylogenetic performance of these gene sequences. For the vertebrate mtDNA protein genes, the strand specific mutational biases in mitochondrion have a major effect on molecular evolution of gene sequences. This result has a strong influence upon the interpretation of mitochondrial phylogenies of ostariophysan fishes based on the mtDNA protein gene sequences. Therefore, for ostariophysan mtDNA proteins, phylogenetic analysis of the amino acid sequences may better reflect patterns of historical descent in ostariophysan fishes than the analysis of nucleotide gene sequences.

4.2. Temporal mitochondrial phylogeny

As expected, the present analysis strongly supports monophyly of the Cypriniformes. The otophysan order Cypriniformes was morphologically recognized as a natural group (Fink and Fink, 1981) and recovered as monophyletic in recent molecular phylogenetic studies (Lavoue et al., 2005; Liu et al., 2002). The mitochondrial protein phylogeny also recovers the well-supported monophyletic Otophysi, comprising the Characiformes, Cypriniformes, Gymnotiformes, and Siluriformes. Our inferred phylogenetic relationships among the otophysan orders are largely consistent with other molecular phylogenies (Dimmick and Larson, 1996; Lavoue et al., 2005; Orti, 1997; Saitoh et al., 2003). However, our ostariophysan mitochondrial protein phylogeny calls into question the monophyly of the Ostariophysi and suggests that the Otophysi is more closely related to the Clupeiformes than to the Gonorynchiformes. Our study strongly supports a clade consisting of the Gonorynchiformes (except *Gonorynchus*), Clupeiformes, and Otophysi.

Acknowledgements

We are grateful to Dr. Paula M. Mabee whose comments greatly improved the presentation of our manuscript. We also extend our gratitude to the anonymous reviewers for their useful suggestions. Funding support was provided by the grants 30300036, 30225008, and 30530120 from National Natural Science Foundation of China (NSFC) and the Cypriniformes Tree of Life Initiative supported by the USA National Science Foundation (EF-0431326).

References

Adachi, J., Hasegawa, M., 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. 42, 459–468.

Berra, T.M., 2001. Freshwater Fish Distribution. Academic Press, San Diego.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780.

Boore, J.L., Brown, W.M., 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. 8, 668–674.

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.

Broughton, R.E., Reneau, P.C., 2006. Spatial covariation of mutation and nonsynonymous substitution rates in vertebrate mitochondrial genomes. Mol. Biol. Evol. 23, 1516–1524.

Chang, Y.S., Huang, F.L., Lo, T.B., 1994. The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. J. Mol. Evol. 38, 138–155.

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

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

Clayton, D.A., 2000. Transcription and replication of mitochondrial DNA. Hum. Reprod. (Oxford, England) 15 (Suppl 2), 11–17.

Dimmick, W.W., Larson, A., 1996. A molecular and morphological perspective on the phylogenetic relationships of the otophysan fishes. Mol. Phylogenet. Evol. 6, 120–133.

Fink, S.V., Fink, W.L., 1981. Interrelationships of the ostariophysan fishes (Teleostei). Zool. J. Linn. Soc. 72, 297–353.

Gray, M.W., Burger, G., Lang, B.F., 1999. Mitochondrial evolution. Science 283, 1476–1481.

Howes, G., 1980. The anatomy, phylogeny and classification of bariliine cyprinid fishes. Bull. Br. Mus. Nat. Hist., Zool. 37, 129–198.

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

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2001a. Complete mitochondrial DNA sequence of the Japanese anchovy *Engraulis japonicus*. Fish. Sci. 67, 828–835.

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2001b. A mitogenomic perspective on the basal teleostean phylogeny: resolving higher-level relationships with longer DNA sequences. Mol. Phylogenet. Evol. 20, 275–285.

Kawaguchi, A., Miya, M., Nishida, M., 2001. Complete mitochondrial DNA sequence of *Aulopus japonicus* (Teleostei: Aulopiformes), a basal Eurypterygii: longer DNA sequences and higher-level relationships. Ichthyol. Res. 48, 213–223.

Lavoue, S., Miya, M., Inoue, J.G., Saitoh, K., Ishiguro, N.B., Nishida, M., 2005. Molecular systematics of the gonorynchiform fishes (Teleostei) based on whole mitogenome sequences: implications for higher-level relationships within the Otocephala. Mol. Phylogenet. Evol. 37, 165–177.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. Nature 362, 709–715.

Liu, H.Z., Tzeng, C.S., Teng, H.Y., 2002. Sequence variations in the mitochondrial DNA control region and their implications for the phylogeny of the Cypriniformes. Can. J. Zool. 80, 569–581.

Liu, Z.Q., Wang, Y.Q., Su, B., 2005. The mitochondrial genome organization of the rice frog, *Fejervarya limnocharis* (Amphibia: Anura): a new gene order in the vertebrate mtDNA. Gene 346, 145–151.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. U. S. A. 95, 10693–10697.

Miya, M., Nishida, M., 1999. Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): first example of transfer RNA gene rearrangements in bony fishes. Mar. Biotechnol. (NY) 1, 416–0426.

Miya, M., 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., 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.

Murakami, M., Yamashita, Y., Fujitani, H., 1998. The complete sequence of mitochondrial genome from a gynogenetic triploid "ginbuna" (*Carassius auratus langsdorfi*). Zool. Sci. 15, 335–337.

Naylor, G.J., Brown, W.M., 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. 47, 61–76.

Nelson, J.S. (Ed.), 1994. Fishes of the World. John Wiley and Sons Inc., New York.

Orti, G.S., 1997. Radiation of characiform fishes: evidence from mitochondrial and nuclear DNA sequences. In: Kocher, T.D., Stepien, C.A. (Eds.), Molecular Systematics of Fishes. Academic Press, San Diego, pp. 219–243.

Pond, S.L., Frost, S.D., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679.

Rasmussen, A.S., Arnason, U., 1999. Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes. J. Mol. Evol. 48, 118–123.

Regan, C.T., 1911. The classification of the teleostean fishes of the order Ostariophysi-I Cyprinoidea. Ann. Mag. Nat. Hist. 8, 13–32.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Saccone, C., Pesole, G., Sbisa, E., 1991. The main regulatory region of mammalian mitochondrial DNA: structure–function model and evolutionary pattern. J. Mol. Evol. 33, 83–91.

Saitoh, K., Miya, M., Inoue, J.G., Ishiguro, N.B., Nishida, M., 2003. Mitochondrial genomics of ostariophysan fishes: perspectives on phylogeny and biogeography. J. Mol. Evol. 56, 464–472.

Sancar, A., Sancar, G.B., 1988. DNA repair enzymes. Ann. Rev. Biochem. 57, 29–67.

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

Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135, 599–607.

Tanaka, M., Ozawa, T., 1994. Strand asymmetry in human mitochondrial DNA mutations. Genomics 22, 327–335.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Xia, X., 2005. Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. Gene 345, 13–20.

Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92, 371–373.

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

Zhang, F., Mi, Z.Y., Mao, Z.Y., Wu, N.H., 1999. Cloning and structure analysis of the mitochondrial control region and its flanking tRNA genes from grasscarp (*Ctenopharyngodon idellus*). Chinese J. Biochem. Mol. Biol. 15, 417–422.

Zhang, P., Zhou, H., Liang, D., Liu, Y.F., Chen, Y.Q., Qu, L.H., 2005. The complete mitochondrial genome of a tree frog, *Polypedates megacephalus* (Amphibia: Anura: Rhacophoridae), and a novel gene organization in living amphibians. Gene 346, 133–143.