The complete mitochondrial genome of the taimen, Hucho taimen, and its unusual features in the control region

Article *in* Mitochondrial DNA · August 2011

DOI: 10.3109/19401736.2011.624605 · Source: PubMed

4 authors, including:

Ying Wang
Chinese Academy of Sciences
64 PUBLICATIONS **285** CITATIONS

SEE PROFILE

Xiuyue Zhang
Sichuan University
173 PUBLICATIONS **1,001** CITATIONS

SEE PROFILE

Mitochondrial DNA, August 2011; 22(4): 111–119

informa
healthcare

FULL LENGTH RESEARCH PAPER

The complete mitochondrial genome of the taimen, *Hucho taimen*, and its unusual features in the control region

YING WANG, XIUYUE ZHANG, SHIYONG YANG, & ZHAOBIN SONG

Sichuan Key Laboratory of Conservation Biology on Endangered Wildlife, College of Life Sciences, Sichuan University, Chengdu 610064, People's Republic of China

(Received 16 March 2011; revised 15 July 2011; accepted 14 September 2011)

Abstract
The whole mitochondrial genome of *Hucho taimen* was firstly sequenced and characterized. The genome is 16,833 bp in length and contains 13 protein-coding genes, 2 ribosomal RNAs, 22 transfer RNAs, and a noncoding control region. Twelve protein-coding genes on the heavy strand showed that the content of A + T was higher than that of G + C, whereas the *nd6* protein-coding gene on the light strand displayed an opposite pattern. We described the secondary structure of the origin of light strand (oriL) replication and found that the conserved 5′-GCCGG-3′ sequence motif is variable in *H. taimen* and some other salmonids. We conclude that the control region is variable in length and represents the high A + T content, compared with other mitochondrial control regions available in Salmonidae and other non-salmonids. Additionally, another interesting feature of *H. taimen* mitogenome is that a T-type mononucleotide microsatellite and an 82 bp tandem repeat were identified in the control region.

Keywords: *Mitochondrial genome, Hucho taimen, origin of light strand (oriL) replication, tandem repeats*

Introduction

Taimen, *Hucho taimen* (Pallas), is the world's largest salmonid and a freshwater resident fish (Esteve et al. 2009; Gilroy et al. 2010). At present, overexploitation, environmental pollution, dam constructions, and other reasons have caused a rapid reduction in the wild populations of this species (Tong et al. 2006). Therefore, *H. taimen* has become an increasingly rare and endangered species throughout its native range, which includes Russia, China, Kazakhstan, and Mongolia (Gilroy et al. 2010). *H. taimen* is listed as "vulnerable" on the China Red Data Book of Endangered Animals (Yue and Chen 1998) and "endangered" on the Mongolian red list (Ocock et al. 2006). The studies about *H. taimen* mainly focus on its physiology, ecology, genetics, and behaviors (Matveyev et al. 1998; Tong et al. 2006; Esteve et al. 2009; Gilroy et al. 2010). Apart from a comparative phylogeographic study by using partial DNA sequences (Froufe et al. 2005), and interspecific phylogenetic studies (Osinov and Lebedev 2000; Crespi and Fulton 2004; Phillips et al. 2004), molecular studies about taimen itself are limited. Therefore, molecular and genomic information of *H. taimen* are in demand to obtain.

Mitochondrial DNA (mtDNA) has become one of the most popular genetic markers in animals over the last three decades (Bruford et al. 2003; Waugh 2007), because of its small size and stable organization, maternal inheritance (lack of recombination), high copy numbers, and a relatively rapid mutation rate. Fish mtDNA is a covalently closed-circular molecule usually about 15–18 kb in length, containing two ribosomal RNA (rRNA) genes (12S and 16S), 13 protein-coding genes (*atp6* and *atp8*, *cox1-3*, *cytb*, *nd1-6* and *nd4l*), 22 transfer RNA (tRNA) genes, and

Correspondence: Z. Song, Sichuan Key Laboratory of Conservation Biology on Endangered Wildlife, College of Life Sciences, Sichuan University, Chengdu 610064, People's Republic of China. Tel: + 86 28 85412801. Fax: + 86 28 85414886. E-mail: zbsong@scu.edu.cn

ISSN 1940-1736 print/ISSN 1940-1744 online © 2011 Informa UK, Ltd.
DOI: 10.3109/19401736.2011.624605

Table I. Primers used to amplify the complete mtDNA of *Hucho taimen*.

Primer name	Sequences	References
COI barcode region		
VF2_t1	5′-*TGTAAAACGACGGCCAG*TCAACCAACCACAAAGACATTGGCAC-3′*	Ward et al. (2005)
FishF2_t1	5′-*TGTAAAACGACGGCCAG*TCGACTAATCATAAAGATATCGGCAC-3′*	Ward et al. (2005)
FishR2_t1	5′-*CAGGAAACAGCTATGAC*ACTTCAGGGTGACCGAAGAATCAGAA-3′*	Ward et al. (2005)
FR1d_t1	5′-*CAGGAAACAGCTATGA*CACCTCAGGGTGTCCGAARAAYCARAA-3′*	Ivanova et al. (2007)
Cytb(L14724/H15915)	5′-GACTTGAAAAACCACCGTTG-3′ 5′-CTCCGATCTCCGGATTACAAGAC-3′	Xiao et al. (2001)
12S/16S	5′-AATTCAGCAGTGATAAACATT-3′ 5′-AGATAGAAACTGACCTGGATT-3′	Gharrett et al. (2001)
ND1/ND2	5′-ACCTCGATGTTGGATCAGG-3′ 5′-ATTAAAGTGTCTGGGTTGCATTC-3′	Gharrett et al. (2001)
COI/ATP6	5′-AGTTACTGCCGTTCTCTTAC-3′ 5′-TTGAAGGGTGATAAGGC-3′	This study
ATP6 (L8558/L9208)	5′-AGCTTCTTCGACCAATTTATGAG-3′ 5′-TATGCGTGTGCTTGGTGTGCCA-3′	Giuffra et al. (1994)
ND3/ND4	5′-TTACGCGTATAAGTGACTTCCAA-3′ 5′-TTTTGGTTCCTAAGACCAATGGA-3′	Gharrett et al. (2001)
ND4/ND5	5′-CCTAACCTAATGGGAGAACT-3′ 5′-GATCAGGTTACGTACAGGGC-3′	This study
ND5-F/ND5-R	5′-CTCTTGGTGCAAATCCAAGT-3′ 5′-GTGCTGGAGTGTAGTAGGGC-3′	This study
ND5 /Cytb	5′-TTCTTCCCCGCTATCATCCACCG-3′ 5′-GAGCCAAAGTTTCATCA-3′	This study
Cytb/12S	5′-CCAAATCGCCTCCGTAAT-3′ 5′AGGAATGCGGAGACTTGC-3′	This study

*M13 tails were italicised.

a noncoding control region (Guo et al. 2004). Control region is characterized as being the most variable region of mtDNA (Chang and Clayton 1985; Brown et al. 1986), due to nucleotide substitutions, indels, and the number of tandem repeat sequences. Numerous studies have documented the presence of a variable number of tandem repeat arrays and a high level of length heteroplasmy for the fish control region (Lee et al. 1995; Ludwig et al. 2000; Hoarau et al. 2002; Yue et al. 2006; Mjelle et al. 2008).

To date, the complete mitochondrial genomes of 19 salmonids are publicly available (http://www.ncbi.nlm.nih.gov). Among four species within the genus *Hucho*, only the mitogenome of *Hucho bleekeri* was

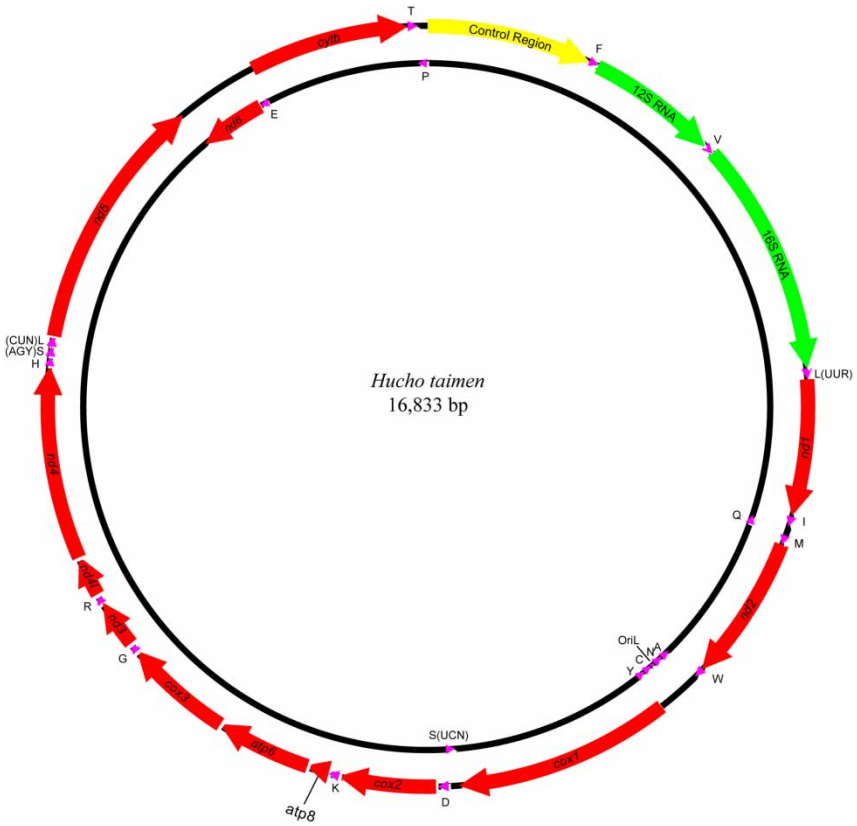

Figure 1. Gene content and organization of mitochondrial genome of *Hucho taimen* (GenBank accession number: HQ897271). The circular gene map representing the mtDNA of *Hucho taimen*. All genes are encoded on the heavy strand except for *nd6* and eight of the tRNA genes (indicated by standard single-letter amino acid codes in the inner loop).

sequenced (Wang et al. 2011). In this study, we reported the complete mtDNA sequence of *H. taimen* for the first time and analyzed its genomic structure. Additionally, we also proposed the secondary structure of the origin of light strand (oriL) replication and described the unusual features of its control region. The information reported in this article will facilitate further investigations of salmonid phylogenetics and a better understanding of *H. taimen* genomics.

Materials and methods

Sample collection and DNA extraction

The wild taimens were captured from the Hutou range of the Ussuri River in Heilongjiang Province in 2002. The fin clips were sampled and then stored in 95% ethanol. Total genomic DNA was extracted from the fin clips according to the standard phenol/chloroform methods (Sambrook and Russell 2001).

PCR amplification, cloning, and sequencing

A single individual was used for amplifying the mtDNA genome. We amplified contiguous overlapping segments of the *H. taimen* mitochondrial genome following the universal primers and other primers proposed in the previous studies (Giuffra et al. 1994; Miya and Nishida 1999; Gharrett et al. 2001). Additional primers were designed to close the remaining gaps between the adjacent fragments. The primer sequences used were shown in Table I.

PCR was performed in a MJ PTC-100 thermal cycler, and reactions were carried out in a 25 μl

Table II. Characteristics of the mitochondrial genome of *Hucho taimen*.

Gene	Position		Size	Codon			Intergenic nucleotide†	Strand
	From	To	Nucleotide (bp)	Amino acid	Start	Stop*		
Control region	1	1176	1176				0	H
$tRNA^{Phe}$	1177	1244	68				0	H
12S rRNA	1245	2191	947				0	H
$tRNA^{Val}$	2192	2263	72				0	H
16S rRNA	2264	3943	1680				0	H
$tRNA^{Leu}$(UUR)	3944	4018	75				0	H
nd1	4019	4990	972	323	ATG	TAA	+8	H
$tRNA^{Ile}$	4999	5070	72				−3	H
$tRNA^{Gln}$	5138	5068	71				−1	L
$tRNA^{Met}$	5138	5206	69				0	H
nd2	5207	6256	1050	349	ATG	TAA	−2	H
$tRNA^{Trp}$	6255	6325	71				+2	H
$tRNA^{Ala}$	6396	6328	69				+1	L
$tRNA^{Asn}$	6470	6398	73				0	L
OriL	6505	6471	35				+1	L
$tRNA^{Cys}$	6572	6506	67				0	L
$tRNA^{Tyr}$	6643	6573	71				+1	L
cox1	6645	8195	1551	516	GTG	TAA	0	H
$tRNA^{Ser}$(UCN)	8196	8266	71				+4	L
$tRNA^{Asp}$	8271	8344	74				+14	H
cox2	8359	9049	691	230	ATG	T--	0	H
$tRNA^{Lys}$	9050	9123	74				+1	H
atp8	9125	9292	168	55	ATG	TAA	−10	H
atp6	9283	9966	684	227	ATG	TAA	−1	H
cox3	9966	10,751	786	261	ATG	TAA	−1	H
$tRNA^{Gly}$	10,751	10,820	70				0	H
nd3	10,821	11,169	349	116	ATG	T--	0	H
$tRNA^{Arg}$	11,170	11,239	70				0	H
nd4l	11,240	11,536	297	98	ATG	TAA	−7	H
nd4	11,530	12,910	1381	460	ATG	T--	0	H
$tRNA^{His}$	12,911	12,979	69				0	H
$tRNA^{Ser}$(AGY)	12,980	13,048	69				+1	H
$tRNA^{Leu}$(CUN)	13,050	13,122	73				0	H
nd5	13,123	14,961	1839	612	ATG	TAA	−4	H
nd6	15,479	14,958	522	173	ATG	TAA	0	L
$tRNA^{Glu}$	15,548	15,480	69				+3	L
cytb	15,552	16,692	1141	380	ATG	T--	0	H
$tRNA^{Thr}$	16,693	16,764	72				−1	H
$tRNA^{Pro}$	16,833	16,764	70					L

*T-- represents incomplete stop codons; †numbers correspond to the nucleotides separating adjacent genes. Negative numbers indicate overlapping nucleotides.

reaction volume containing 16.3 μl sterilized distilled water, 2.5 μl 10 × PCR buffer, 2.5 μl $MgCl_2$ (25 mM), 1.0 μl each primer (25 μlM), 1 μl dNTP (2.5 mM), 0.2 μl *Taq* (5 U/μl) (Takara, Dalian, China), and 0.5 μl genomic DNA (appropriate 30 ng). The following PCR program was performed: pre-denaturation at 95°C for 4 min, followed by 35 cycles at 95°C for 30 s, 52–60°C for 45 s, and 72°C for 1–3 min, with a final extension at 72°C for 10 min. All PCR products were electrophoresed on 1.0% agarose gel and sized following the molecular weight marker D2000 (Tiangen, Beijing, China). Most amplified products were directly sequenced by the primer walking methods on a 3730 DNA Analyzer (Applied Biosystems, Inc., Foster City, California, USA). The control region containing a T-type mononucleotide microsatellite was not directly sequenced completely and accurately. Thus, the PCR product of the *H. taimen* control region was first purified using the DNA Agarose Gel Extraction Kit (Omega Bio-Tek, Norcross, Georgia, USA) and cloned into a pMD19-T Vector (Takara, Dalian, China), and then transformed into *E. coli* JM109 competent cells. Positive clones were characterized and sequenced in both directions.

Sequence analysis and genome annotation

The whole mitochondrial sequence was obtained by contiguous assembling with the ContigExpress program. The locations of 13 protein coding and two rRNA (12S and 16S) genes were identified through

Table III. Analyses of mitochondrial genomes for salmonids and other non-salmonids.

			rRNA length (bp)		Control region	
Species (accession numbers)	Size (bp)	A + T (%)	12S	16S	Length (bp)	A + T (%)
Hucho taimen (HQ897271)	16,833	54.7	947	1680	1176	61.9
Hucho taimen (EU760489)	–	–	–	–	1136	61.5
Hucho bleekeri (HM804473)	16,997	54.7	947	1680	1339	62.3
Salmo salar (NC_001960)	16,665	54.7	946	1678	1006	63.1
Salmo trutta trutta (NC_010007)	16,677	54.1	947	1682	1013	63.2
Salmo trutta trutta (AM910409)	–	–	–	–	1013	63.2
Salvelinus alpinus (NC_000861)	16,659	54.6	947	1680	998	62.5
Salvelinus fontinalis (NC_000860)	16,624	54.8	947	1680	964	61.7
Oncorhynchus mykiss (NC_001717)	16,642	54.1	944	1680	1003	61.5
Oncorhynchus tshawytscha (NC_002980)	16,644	54.4	947	1683	986	60.4
Oncorhynchus nerka (NC_008615)	16,658	53.8	947	1678	1004	60.6
Oncorhynchus masou masou (NC_008747)	16,652	55.4	946	1679	998	60.3
Oncorhynchus kisutch (NC_009263)	16,659	53.7	947	1690	1002	61.2
Oncorhynchus keta (NC_009261)	16,659	54.4	946	1681	1005	60.9
Oncorhynchus gorbuscha (NC_010959)	16,785	54.4	947	1678	1131	60.4
Oncorhynchus clarkii henshawi (NC_006897)	16,658	53.9	947	1677	1003	61.4
Oncorhynchus masou (NC_009262)	16,652	55.4	946	1679	998	60.8
Oncorhynchus masou formosanus (NC_008745)	16,652	55.4	946	1679	998	60.4
Oncorhynchus masou ishikawae (NC_008746)	16,652	55.4	946	1679	998	60.7
Coregonus lavaretus (NC_002646)	16,737	53.4	947	1679	1076	60.9
Coregonus lavaretus (AF140602)	–	–	–	–	1356	61.7
Coregonus lavaretus baicalensis (AJ250993)	–	–	–	–	1075	59.8
Thymallus arcticus (NC_012929)	16,644	55.7	946	1680	1000	65.8
Thymallus arcticus (DQ683723)	–	–	–	–	1006	65.7
Thymallus thymallus (NC_012928)	16,657	55.2	946	1678	1001	64.5
Thymallus thymallus (AY841357)	–	–	–	–	938	63.7
Hucho hucho (EU729360)	–	–	–	–	1548	61.6
Brachymystax lenok (EU760490)	–	–	–	–	1214	63.7
Brachymystax lenok (EU760491)	–	–	–	–	1134	63.5
Brachymystax lenok (AF125519)	–	–	–	–	1089	63.9
Acipenser dabryanus (AY510085)	16,438	53.9	960	1701	726	61.9
Acipenser transmontanus(AB042837)	16,692	54.1	960	1701	979	63.3
Heterodontus francisci (AJ310141)	16,708	60.1	952	1676	1068	65.7
Lepidosiren paradoxa (AF302934)	16,403	62.8	930	1578	974	63.2
Carpiodes carpio (AY366087)	16,611	54.2	952	1689	929	62.8
Carassius carassius (NC_006291)	16,580	57.7	954	1681	923	65.0
Cyprinus carpio (NC_001606)	16,575	56.8	951	1681	928	66.2
Esox lucius (NC_004593)	16,695	56.1	942	1660	1072	68.6
Larimichthys crocea (EU339149)	16,466	53.0	947	1693	795	61.2
Myxocyprinus asiaticus (AY986503)	16,623	55.3	950	1687	941	60.3
Pagellus bogaraveo (AB305023)	16,941	53.6	951	1697	1195	62.6

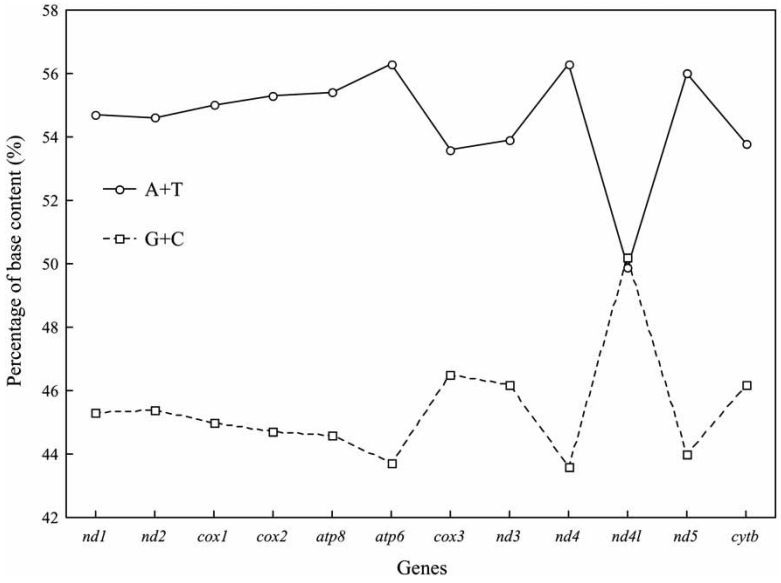

Figure 2. The A + T and G + C percentage of base content for 12 mitochondrial protein-coding genes on the heavy strand in *Hucho taimen* mtDNA. Genes were ordered according to their positions in mitochondrial genome.

BLAST searches on NCBI database (http://www.ncbi.nlm.nih.gov) and through comparisons with homologous sequences of other salmonid mtDNA, such as Atlantic salmon (*Salmo salar*), rainbow trout (*Oncorhynchus mykiss*), and Sichuan taimen (*H. bleekeri*). The program tRNAscan-SE1.21 (Lowe and Eddy 1997; http://lowelab.ucsc.edu/tRNAscan-SE) was used to identify tRNA genes within the intergenic regions of the mitochondrial genome. Some tRNA genes that could not be found by tRNAscan-SE were further verified by the inspection of anticondon sequences and proposed secondary structures (Kumazawa and Nishida 1993). The secondary structure of light strand replication origin and the repeat unit from the tandem repeat arrays were constructed with Mfold web server (http://www.bioinfo.rpi.edu/applications/mfold) (Zuker 2003).

Results and discussion

Genome organization

The entire length of the *H. taimen* mitochondrial genome is 16,833 bp (GenBank accession number: HQ897271). The mtDNA of *H. taimen* contains a set of 13 protein-coding genes, 2 rRNA genes (12S and 16S), 22 tRNA genes, and the control region (Figure 1). The compactness of the genome is due to the occurrence of frequent gene overlaps. There are 9 overlapping regions and 10 intergenic spacers among the genes (Table II). The gene content and organization of *H. taimen* are almost identical to other 19 fish species in Salmonidae except for the length variation of their control regions. As with other bony fishes, all mitochondrial genes of the *H. taimen* are encoded on the heavy strand, except for *nd6* and eight tRNA genes. The overall base composition of the heavy strand was A: 27.6%, T: 27.1%, C: 28.7%, and G: 16.6%. With respect to its complete mitochondrial genome, the A + T content (54.7%) is similar to those of other salmonids and non-salmonids (ranging from 53.0% to 62.8%, Table III). Furthermore, these

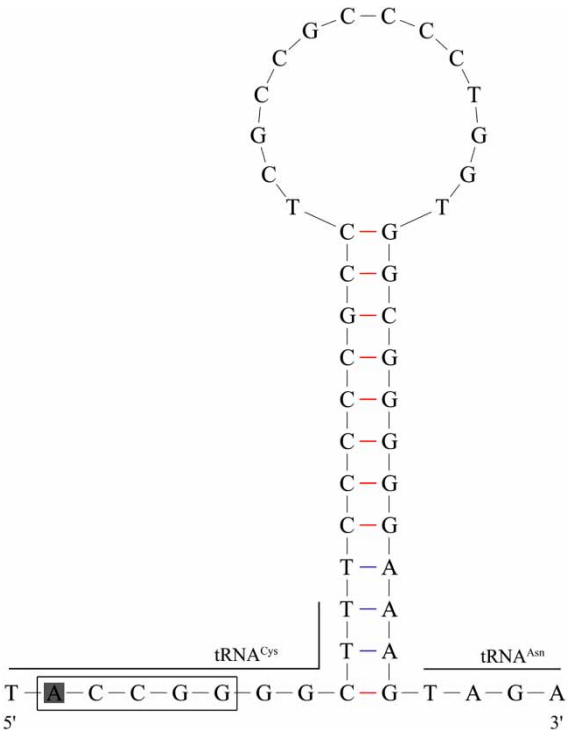

Figure 3. Putative secondary structure of the origin of light strand (oriL) replication. The 5′-ACCGG-3′ motif is indicated by a box, distinguishing from the conserved 5′-GCCGG-3′. The mutational base is indicated in gray color. Lines show the nucleotides partially shared with flanking tRNAs.

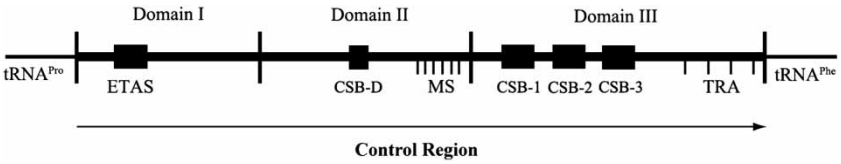

Figure 4. Schematic diagram of *Hucho taimen* mitochondrial control region. ETAS, extended termination-associated sequence; MS, T-type single nucleotide microsatellite; CSB, conserved sequence blocks; CSB-D, central conserved sequence block-D; TRA, tandem repeat array.

values and codon usage are similar to most sequenced bony fish mitochondrial genomes with the lowest frequency for G among the four bases (Meyer 1993).

Protein-coding genes

The protein-coding region in the mitochondrial genome of *H. taimen* is 11,431 bp in length, accounting for 67.9% of the entire genome. All the protein-coding genes begin with the traditional ATG start codon except *cox1* used GTG as the initiation codon (Table II). Accordingly, nine of the 13 protein-coding genes end with the termination codon TAA and four (*cox2*, *nd3*, *nd4*, and *cytb*) do not possess a complete stop codon, but only show a terminal T. This condition is known to be common among fish mitochondrial genomes, and it has been shown that TAA stop codon is created via posttranscriptional polyadenylation (Ojala et al. 1981). Among the 13 protein-coding genes, some overlaps were identified in *atp8-atp6*, *atp6-cox3*, *nd4l-nd4*, and *nd5-nd6* (Table II). The nucleotide compositions of the 11 protein-coding genes on the heavy strand show that the content of A + T is higher than that of G + C, but with respect to *nd4l* gene, the A + T and G + C contents are approximately equal (Figure 2).

Identification of tRNA genes

All tRNAs can be detected in the *H. taimen* mitochondrial genome using the program tRNAscan-SE with the exception of $tRNA^{Ser}$ (AGY), which is instead identified by the proposed secondary structures and anticodon (Kumazawa and Nishida 1993). All tRNA genes are located on the heavy strand, except for $tRNA^{Gln}$, $tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Ser}$ (UCN), $tRNA^{Glu}$, and $tRNA^{Pro}$ (Figure 1; Table II). These tRNA genes range in size from 67 to 74 nucleotides and are predicted to fold into the expected cloverleaf secondary structures with normal base pairing.

Ribosomal RNA genes

As the mitochondrial genomes of other fishes, the *H. taimen* mitochondrial genome was found to possess both the small and the large rRNA genes. Both of the ribosomal genes are situated between $tRNA^{Phe}$ and $tRNA^{Leu}$ (UUR) and are separated by $tRNA^{Val}$. The lengths of 12S RNA and 16S RNA are 947 and 1680 bp, respectively, which are within the range observed for other salmonids and non-salmonids (Table III).

Noncoding region

Noncoding region of most fish mitochondrial genome is comprised of the origin of light strand replication and the control region. Similar to that reported for rainbow trout and Atlantic salmon (Zardoya et al. 1995; Hurst et al. 1999), the origin of light strand replication of *H. taimen* is located in the cluster of five tRNA genes (WANCY region) (Figure 1). A stable secondary structure for the origin of light strand replication of Atlantic salmon, rainbow trout, and human has been proposed and supported the suggestion of Zardoya et al. (1995) that initiation of light strand synthesis is not restricted to a stretch of thymines but is initiated in a polypyrimidine tract. For *H. taimen*, a similar stem-loop secondary structure can be also folded, consisting of 12 paired nucleotides in stem region and 14 nucleotides in the loop (Figure 3). The 5′-GCCGG-3′ motif has been demonstrated in the origin of light strand replication of most vertebrates (Zardoya et al. 1995; Hurst et al. 1999; San Mauro et al. 2004), which is involved in transition from RNA synthesis to DNA synthesis in human mtDNA (San Mauro et al. 2004). However, at the base of the stem region of oriL in *H. taimen*, the

Figure 5. Alignment of the repeated 82-bp motif of the related species. The conserved TACAT motif was indicated by box. The variable bases for each other species were showed in gray. Ht, Bl, Hh, and Hb are abbreviations for *Hucho taimen*, *B. lenok*, *Hucho hucho*, and *Hucho bleekeri*, respectively.

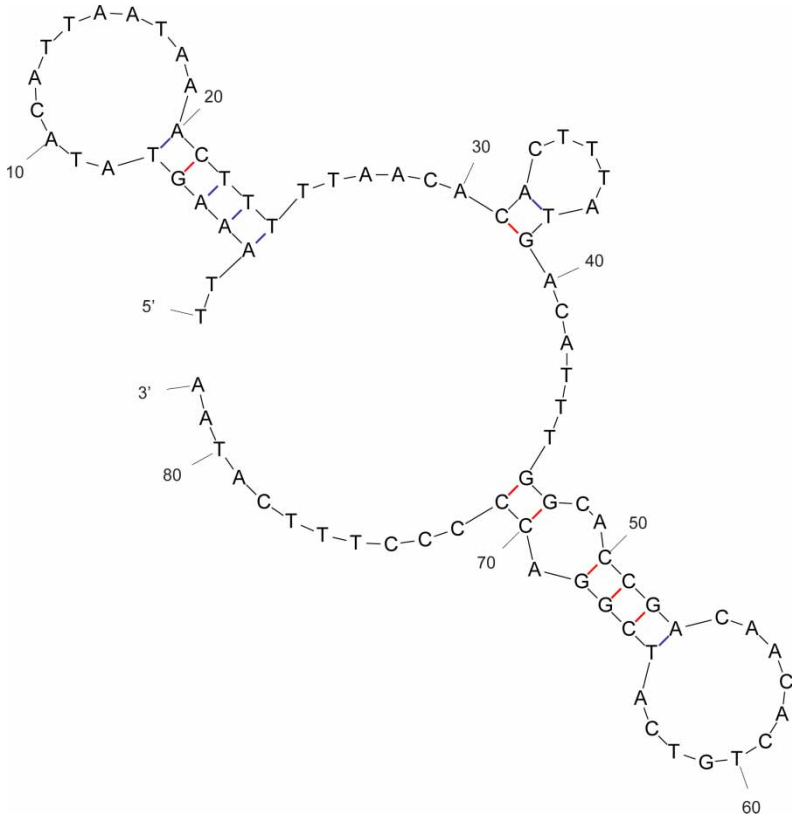

Figure 6. The possible secondary structure of the tandem repeat of 82-bp motif. The estimated free energy was −9.05 kcal at 120 mM Na^+, 2 mM Mg^{2+}, and 20°C.

conserved motif is replaced by 5′-ACCGG-3′ (Figure 3). Interestingly, the unusual motif is also present in *H. bleekeri*, *Coregonus lavaretus*, *Thymallus arcticus*, *Thymallus thymallus*, and *Oncorhynchus gorbuscha*.

The control region is the major noncoding region found in *H. taimen* of 1176 bp in length with high A + T content (61.9%), and is located between $tRNA^{Pro}$ and $tRNA^{Phe}$ genes (Table III, Figures 1 and 4). As found in other vertebrate mitogenomes, the control region could be divided into three different domains (Figure 4): two variable flanking domains (Domain I and Domain III), and a conservative central domain (Domain II) (Anderson et al. 1981; Brown et al. 1986). The extended termination-associated sequence could be found in the left domain I of the *H. taimen* control region. As with most bony fishes, the critical central conserved sequence blocks (CSBs), CSB-D, CSB-E, and CSB-F with TTCCTGGCATTTGGTTCC, TCAGGGGCAGAAATCGTATTAGG, and ATGTAGTAAGAACCGACC, respectively, could be detected in the central portion of the *H. taimen* control region. According to the previous studies (Sbisà et al. 1997; Pesole et al. 1999; Guo et al. 2003), the functions of the central conserved blocks remain obscure, but CSB-D is highly conserved in fish and may regulate the regulation of heavy strand replication and perhaps also be involved in mitochondrial metabolism (Clayton 1982; Lee et al. 1995). Additionally, the CSB-1, CSB-2, and CSB-3 could be easily identified in *H. taimen* mtDNA (Figure 4).

Several unusual sequence features could be found by aligning the control region of *H. taimen* with that of other salmonids. T-type mononucleotide microsatellite and an 82 bp tandem repeat were identified in this region. T-type mononucleotide microsatellite is an interesting feature of Salmonidae fish, the number of which is variable from 13 to 17 nucleotides. Similarly, AT-repeat microsatellite sequence was identified in some Cyprinidae fishes (Zhang et al. 2009). These different microsatellite sequences might be useful in some interspecies identification.

The 82 bp tandem repeat identified in *H. taimen* also could be found in *H. bleekeri*, *H. hucho*, *Brachymystax lenok*, and was highly similar among these species with only a few base substitutions (Figure 5). The TACAT motif described earlier only from mammals, lungfish (*Protopterus dolloi*) and Asian arowna (*Scleropages formosus*) (Yue et al. 2006) was found in the tandem repeats of several salmonids as well (Figure 5). Four tandem repeat arrays of 82 bp motif are located between the CSB-3 and $tRNA^{Phe}$ gene. The repeat unit could be folded into a

stable secondary structure (Figure 6). The length of control region of salmonids reported is variable among closely related species and within species, due to the presence of a variable number of tandem repeats. The different type of tandem repeats has been reported in most teleost mitochondrial control region, such as Atlantic cod (*Gadus morhua*), yellowtail flounder (*Limanda ferruginea*), winter flounder (*Pseudopleuronectes americanus*), halibuts (*Hippoglossus hippoglossus*, *H. stenolepis*, and *Reinhardtius hippoglossoides*), flatfish (*Platichthy flesus*), sturgeon (*Acipenser sturio*), and Asian arowana (Lee et al. 1995; Ludwig et al. 2000; Hoarau et al. 2002; Yue et al. 2006; Mjelle et al. 2008), and the stable secondary structures of tandem repeats have been proposed. These observations suggest that the copy number of the repeat is highly variable between and within species, which is not fully understood but is best explained by slipped-strand mispairing.

Conclusions

The gene content, arrangement, base composition, and codon usage of *H. taimen* are identical to that observed in most other teleost fishes. Unusual molecular features of taimen are located in the control region of its mitochondrial genome. T-type mononucleotide microsatellite and similar size heteroplasmic tandem repeat arrays of an 82-bp motif were detected in the mitochondrial control regions of several salmonids from the genera of *Hucho* and *Brachymystax*. We attributed the heteroplasmic tandem repeat arrays to the slipped-strand mispairing mechanism.

Acknowledgements

We would like to thank the Heilongjiang River Fishery Research Institute of Chinese Academy of Fishery Sciences for providing the sample of *H. taimen.*

Declaration of interest: This research was funded by the Sichuan Youth Science and Technology Foundation (No. 08ZQ026-019), the Fisheries Research Institute, Sichuan Academy of Agricultural Sciences (No. 10H0024), and the 973 Project (No. 2007CB411605). The authors report no conflict of interest. The authors alone are responsible for the content and writing of the paper.

References

Anderson S, Bankier AT, Barrell BG, de Bruijin MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. Nature 290: 457–465.

Brown GG, Gadaleta G, Pepe G, Saccone C, Sbisà E. 1986. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J Mol Biol 192: 503–511.

Bruford MW, Bradley DG, Luikart G. 2003. DNA makers reveal the complexity of livestock domestication. Nat Rev Genet 4: 900–910.

Chang DD, Clayton DA. 1985. Priming for human mitochondrial DNA replication occurs at the light-strand promoter. Proc Natl Acad Sci USA 82:351–355.

Clayton DA. 1982. Replication of animal mitochondrial DNA. Cell 28:693–705.

Crespi BJ, Fulton MJ. 2004. Molecular systematics of Salmonidae: Combined nuclear data yields a robust phylogeny. Mol Phylogenet Evol 31:658–679.

Esteve M, Gilory D, Mclennan DA. 2009. Spawning behaviour of taimen (*Hucho taimen*) from the Uur River, Northern Monglia. Environ Biol Fish 84:185–189.

Froufe E, Alekseyev S, Knizhin I, Weiss S. 2005. Comparative mtDNA sequence (control region, ATPase6 and NADH-1) divergence in *Hucho taimen* (Pallas) across four Siberian river basins. J Fish Biol 67:1040–1053.

Gharrett AJ, Gray AK, Brykov V. 2001. Phylogeographic analysis of mitochondrial DNA variation in Alaskan coho salmon, *Oncorhynchus kisutch*. Fish Bull 99:528–544.

Gilroy DJ, Jensen OP, Allen BC, Chandra S, Ganzorig B, Hogan Z, Maxted JT, Vander Zanden MJ. 2010. Home range and seasonal movement of taimen, *Hucho taimen*, in Mongolia. Ecol Freshw Fish 19:545–554.

Giuffra E, Bernatchez L, Guyomard R. 1994. Mitochondrial control region and protein coding genes sequence variation among phenotypic forms of brown trout *Salmo trutta* from northern Italy. Mol Ecol 3:161–171.

Guo XH, Liu SJ, Liu Y. 2003. Comparative analysis of the mitochondrial DNA control region in cyprinids with different ploidy level. Aquaculture 224:25–38.

Guo XH, Liu SJ, Liu Q, Liu Y. 2004. New progresses on mitochondrial DNA in fish. Acta Genet Sin 31:983–1000 (in Chinese with an English abstract).

Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL. 2002. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthy flesus*. Mol Biol Evol 19:2261–2264.

Hurst CD, Bartlett SE, Davidson WS, Bruce IJ. 1999. The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. Gene 239:237–242.

Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN. 2007. Universal primer cocktails for fish DNA barcoding. Mol Ecol Notes 7:544–548.

Kumazawa Y, Nishida M. 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J Mol Evol 37:380–398.

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

Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964.

Ludwig A, May B, Debus L, Jenneckens I. 2000. Heteroplasmy in the mtDNA control region of sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). Genetics 156:1933–1947.

Matveyev AN, Pronin NM, Samusenok VP, Bronte CR. 1998. Ecology of Siberian taimen *Hucho taimen* in the Lake Baikal Basin. J Great Lakes Res 24:905–916.

Meyer A. 1993. Evolution of mitochondrial DNA in fishes. In: Hochachka PW, Mommsen TP, editors. Biochemistry and molecular biology of fishes. Amsterdam: Elsevier Science Publishers. p 1–38.

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 1:416–426.

Mjelle KA, Karlsen BO, Jørgensen TE, Moum T, Johansen SD. 2008. Halibut mitochondrial genomes contain extensive heteroplasmic tandem repeat arrays involved in DNA recombination. BMC Genomics 9:10.

Ocock J, Baasanjav G, Baillie JEM, Erbenebat M, Kottelat M, Mendsaikhan B, Smith K. 2006. Mongolian red list of fishes, Regional Red List series. Vol. 3. London: Zoological Society of London. p 54.

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

Osinov AG, Lebedev VS. 2000. Genetic divergence and phylogeny of Salmoninae based on allozyme data. J Fish Biol 57:354–381.

Pesole G, Gissi C, De Chirico A, Saccone C. 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. J Mol Evol 48:427–434.

Phillips RB, Matsuoka MP, Konko1 NR, Mckay S. 2004. Molecular systematics and evolution of the growth hormone introns in the Salmoninae. Environ Biol Fish 69:433–440.

Sambrook J, Russell DW. 2001. Molecular cloning: A laboratory manual. New York: Cold Spring Harbor Press.

San Mauro D, García-París M, Zardoya R. 2004. Phylogenetic relationships of discoglossid frogs (Amphibia: Anura: Discoglossidae) based on complete mitochondrial genomes and nuclear genes. Gene 343:357–366.

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C. 1997. Mammalian mitochondrial D-loop region structural analysis: Identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125–140.

Tong GX, Kuang YY, Yin JS, Liang LQ, Sun XW. 2006. Isolation of microsatellite DNA and analysis on genetic diversity of endangered fish, *Hucho taimen* (Pallas). Mol Ecol Notes 6: 1099–1101.

Waugh J. 2007. DNA barcoding in animal species: Progress, potential and pitfalls. Bioessays 29:188–197.

Wang Y, Guo R, Li H, Zhang XY, Du J, Song ZB. 2011. The complete mitochondrial genome of the Sichuan taimen (*Hucho bleekeri*): Repetitive sequences in the control region and phylogenetic implications for Salmonidae. Mar Genomics 4: 221–228.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD. 2005. DNA barcoding Australia's fish species. Phil Trans R Soc B 360: 1847–1857.

Xiao WH, Zhang YP, Liu HZ. 2001. Molecular systematics of Xenocyprinae (Teleostei: Cyprinidae): Taxonomy, biogeography, and coevolution of a special group restricted in East Asia. Mol Phylogenet Evol 18:163–173.

Yue PQ, Chen YY. 1998. China red data book of endangered animals, Pisces. Beijing: Science Press. p 29–31 (in Chinese).

Yue GH, Liew WC, Orban L. 2006. The complete mitochondrial genome of a basal teleost, the Asian arowana (*Scleropages formosus*, Osteoglossidae). BMC Genomics 7:242.

Zardoya R, Garrido-Pertierra A, Baustista JM. 1995. The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Onchorynchus mykiss*. J Mol Evol 41:942–951.

Zhang XY, Yue BS, Jiang WX, Song ZB. 2009. The complete mitochondrial genome of rock carp *Procypris rabaudi* (Cypriniformes: Cyprinidae) and phylogenetic implications. Mol Biol Rep 36:981–991.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415.