GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Complete mitochondrial genome sequence of *Marmota himalayana* (Rodentia: Sciuridae) and phylogenetic analysis within Rodentia

Q.J. Chao[1]*, Y.D. Li[2]*, X.X. Geng[1], L. Zhang[1], X. Dai[1], X. Zhang[1], J. Li[1] and H.J. Zhang[1]

[1]Department of Life Sciences, Huai Bei Normal University, Huaibei, China
[2]School of Medicine, Qinghai University, Xining, China

*These two authors contributed equally to this study.
Corresponding author: X.X. Geng
E-mail: xuexiageng@163.com

Genet. Mol. Res. 13 (2): 2739-2751 (2014)
Received December 3, 2012
Accepted April 18, 2013
Published April 14, 2014
DOI http://dx.doi.org/10.4238/2014.April.14.3

ABSTRACT. This is the first report of a complete mitochondrial genome sequence from Himalayan marmot (*Marmota himalayana*, class Marmota). We determined the *M. himalayana* mitochondrial (mt) genome sequence by using long-PCR methods and a primer-walking sequencing strategy with genus-specific primers. The complete mt genome of *M. himalayana* was 16,443 bp in length and comprised 13 protein-coding genes, 2 ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes, and a typical control region (CR). Gene order and orientation were identical to those in mt genomes of most vertebrates. The heavy strand showed an overall A+T content of 63.49%. AT and GC skews for the mt genome of the *M. himalayana* were 0.012 and -0.300, respectively, indicating a nucleotide bias against T and G. The control region was 997 bp in size and displayed some unusual features, including absence of repeated motifs and two conserved sequence blocks (CSB2 and

CSB3), which is consistent with observations from two other rodent species, *Sciurus vulgaris* and *Myoxus glis*. Phylogenetic analysis of complete mt DNA sequences without the control region including 30 taxa of Rodentia was performed with Maximum-Likelihood (ML) and Bayesian Inference (BI) methods and provided strong support for Sciurognathi polyphyly and Hystricognathi monophyly. This analysis also provided evidence that *M. himalayana* mt DNA was closely related to that from *Sciurus vulgaris* (Sciuridae) and was similar to mt DNA from *Myoxus glis*.

Key words: *Marmota himalayana*; Mitochondrial genome; Rodentia; Phylogenetic analysis

INTRODUCTION

Mitochondria are important subcellular organelles present in almost all animal tissue cells. They facilitate the reactions in the tricarboxylic acid cycle and electron transfer in oxidative phosphorylation, required for cellular energy generation. In organisms lacking chloroplasts, mitochondria are the only organelles besides the nucleus that have their own genomic DNA (Giles et al., 1980). The typical vertebrate mitochondrial (mt) genome is a closed circular DNA molecule with two strands (H-strand and L-strand), ranging 15-19 kb in size. The complete mtDNA contains a remarkably conserved set of 37 genes, including 13 protein-coding genes, 2 rRNA genes, and a full set of 22 tRNA genes (Wolstenholme, 1992; Boore, 1999). In addition, noncoding regions are mainly concentrated in a variable control region (CR) or D-loop with important functions in the regulation and initiation of mtDNA replication and transcription (Shadel and Clayton, 1993).

Marmots are large terrestrial rodents widespread across much of northern Eurasia and North America, including the Bering Strait region of western Alaska and eastern Siberia. Today, there are 14 recognized species of marmots throughout the world (Steppan et al., 1999). Among the four marmot species in China, *Marmota himalayana* is one of the species living at the highest altitudes in the region and has the widest distribution, covering parts of the Qinghai-Tibet Plateau and mountains, and meadows and grasslands in India, Nepal, and Pakistan. Marmot is an important animal with high economic value, and marmot products have a long history of development and use in China. To date, studies of *M. himalayana* have mainly focused on its ecology and ethology, and because of their close relationships with ground squirrels and prairie dogs, marmots have featured prominently in theories of social evolution in mammals (Barash, 1989; Arnold, 1990; Blumstein et al., 1997; Armitage, 1975, 1987, 1999). In recent years, helped by the development of long polymerase chain reaction (long-PCR) methods for amplification of mtDNA genes and complete mt genomes, comparative analyses of mt genome sequences and organizations are widely used to study phylogenetic relationships of animal taxa. The unique characters of mt DNA, such as its small size, simple structure, maternal inheritance, fast evolutionary rate, and high informational content, have made mt DNA a very attractive subject of research (Miyata et al., 1982; Brown, 1983; Moritz et al., 1987; Wolstenholme, 1992; Avis, 1994; Castro et al., 1998; Saccone et al., 1999; Yamanoue et

al., 2007; Zhou et al., 2009). More than 1200 complete mt genome sequences are currently available in GenBank, including mt sequences from various vertebrate groups. Although some studies have investigated mitochondrial genes from *M. himalayana*, such as the *Cytb* gene (Steppan et al., 1999), no complete mt genome of the 14 species of marmots has yet been reported, and only one complete mt genome from Sciuridae (*Sciurus vulgaris*, accession No. AJ238588) is available in GenBank (Reyes et al., 2000). Likewise, previous studies of Rodentia phylogenies have mainly focused on rodent positions within mammalian phylogenetic trees (Reyes et al.,1998, 2000, 2004; Kjer and Honeycutt, 2007; David et al., 2007).

In this study, we describe for the first time the complete mt genome of *M. himalayana* and discuss phylogenetic relationships within Rodentia. This provides an important resource for comparative analyses of the evolution of mitochondrial genomes in rodents.

MATERIAL AND METHODS

Sample collection and DNA extraction

Tissue samples were preserved at -20°C until analysis. Total genomic DNA in samples was extracted by a phenol/chloroform/isoamylalcohol method.

PCR amplification and sequencing

To obtain the complete sequence of the mt genome, seven pairs of primers were designed by using NCBI primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) on the basis of alignments and comparisons of the genome sequence of *Sciurus vulgaris* and partial sequences of the *M. himalayana* mt genome (Table 1). The long polymerase chain reaction (long-PCR) technique was carried out to amplify the complete *M. himalayana* mt genome. PCR was performed in 50-µL volume containing 5 µL 10 X LA PCR buffer II (TaKaRa), 8 µL 4 mM dNTP, 2 µL 10 µM of each primer, 2 µL 1 mM $MgCl_2$, 0.5 µL 2.5 U LA Taq polymerase, and 1 µL genomic DNA. PCR cycling conditions were the following: initial denaturation step of 5 min at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 58°C, and 2-5 min at 72°C, and a final extension step at 72°C for 10 min. The resultant PCR products were checked by electrophoresis on a 1.0% agarose gel (Promega, Madison, WI, USA), staining with ethidium bromide, and visualization under UV on a transilluminator, followed by purification with PCR Purification Kit (Qiagen, USA). Two microliter of the purified PCR products were electrophoresed again to determine DNA concentrations by visual comparison of DNA band intensities with intensities of DNA size standards run alongside. The purified products were directly sequenced in both directions on an ABI 3730 XL Sequencer by using the primer walking method with the species-specific primers.

Sequence analysis

Editing and assembly of the mitochondrial sequences were performed with the DNAStar 7.1 software and program CLUSTAL_W 2.1 (Thompson et al., 1997). The inte-

grated mt genome sequence has been deposited in GenBank by the Sequin software (http://www.ncbi.nlm.nih.gov/sequin/index.html) under accession No. JX069958. The 13 protein-coding genes and 2 rRNA genes were identified by DOGMA (Wyman et al., 2004) and their similarity to published gene sequences in NCBI identified by BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST/). The tRNA genes were identified by tRNA Scan-SE version 1.21 (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Eddy, 1997). The remaining $tRNA^{Ser(GCU)}$ gene was not found by tRNAscan-SE and was therefore predicted using its anticodons with the RNAstructure 4.5 and the RNAdraw software (Mathews et al., 1999, 2004). The CR region was identified by sequence homology and proposed secondary structures. Composition skew analysis of AT and the GC bases was conducted to find nucleotide bases that were different among the mt genomes of *M. himalayana* and other species of the *Marmota* genus. Codon usage was calculated by CodonW version 1.3 (http://bioweb.pasteur.fr/seqanal/codonw.html). Tandem repeats of the whole sequence were tested by Tandem Repeats Finder (Benson, 1999).

Table 1. Primers used for the experiment.

Order	Primer name	Sequence (5'-3')	Product size (bp)
1	ND4F	TGCTGGTTCTATAGTTTTAGCCGCC	3330
	CYR	AGAGGGTGGGTTTTGCGGGTG	
2	Mt76	ACGATCCATCCCCAACAAACTAGGA	3124
	16RR	TGGCTGCTTTTAGGCCAACTATGGA	
3	ND3F	CCCATAGGATCTGCTCGCTTACCA	1007
	ND4R	ACCGTTCAGTTTGGTTTCCCCATCG	
4	ND4LF	TTGCCGCATGTGAAGCAGCTGTA	1216
	ND4	GTTAATGGAGGGTGGAAGGGCTAGA	
5	ValF	ACCCGGCTTACACCCGAGAG	8668
	NDCR	GTTTGGGAAGCTCAGGGAAGAGGGA	
6	ND4F	AGCCCACGTCGAAGCTCCCAT	2842
	ND6R	TGAAGATTCAATAGGGGTGGCGGC	
7	ND6	GCTTTGAAGAAACCCCCACAAAGCC	3199
	12RF	CCACCCTATAGGCTACACCTTGACC	

Phylogenetic analyses

The phylogenetic analyses were performed on the complete mt DNA sequences of *M. himalayana* (this study) and from 29 *Glires* species retrieved from GenBank (Table 2). *Lepus europaeus*, *Oryctolagus cuniculus*, and *Ochotona princeps* from the family Lagomorpha were used as outgroups. The nucleotide sequences of 30 complete mt genome - with the control region excluded - were used in the phylogenetic analysis as follows: First, the sequence alignment and a site-specific rate model were constructed according to Kjer et al. (Kjer et al., 2001; Kjer and Honeycutt, 2007). Two different phylogenetic reconstruction approaches including Maximum Likelihood (ML) and Bayesian Inference (BI) were used to reconstruct phylogenetic trees. For ML analysis, 100 bootstraps were performed to estimate the node support with RaxML-7.04 (Stamatakis, 2006). Bayesian analysis was conducted by MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) under the GTR + gamma model, and the Markov chain Monte Carlo (MCMC) chains were run for 10,000,000 generations (sampling every 1000 generations) to allow adequate time for convergence. Discarding the first 1000 trees as burn-in, the remaining trees were used to estimate posterior probabilities.

Table 2. List of mitochondrial genomes analyzed in this study.

Species	Classification	Accession No.
Neodon irene	Glires; Rodentia; Sciurognathi; Muroidea; Cricetidae	NC_016055.1
Microtus kikuchii	Glires; Rodentia; Sciurognathi; Muroidea; Cricetidae	AF348082.1
Proedromys liangshanensis	Glires; Rodentia; Sciurognathi; Muroidea; Cricetidae	NC_013563.1
Eothenomys chinensis	Glires; Rodentia; Sciurognathi; Muroidea; Cricetidae	NC_013571.1
Cricetulus griseus	Glires; Rodentia; Sciurognathi; Muroidea; Cricetidae	NC_007936.1
Tscherskia triton	Glires; Rodentia; Sciurognathi; Muroidea; Cricetidae	NC_013068.1
Mesocricetus auratus	Glires; Rodentia; Sciurognathi; Muroidea; Cricetidae	EU660218.1
Mus musculus	Glires; Rodentia; Sciurognathi; Muroidea; Muridae	NC_005089.1
Apodemus chejuensis	Glires; Rodentia; Sciurognathi; Muroidea; Muridae	NC_016662.1
Rattus norvegicus	Glires; Rodentia; Sciurognathi; Muroidea; Muridae	AJ428514.1
Nannospalax ehrenberg	Glires; Rodentia; Sciurognathi; Muroidea; Spalacidae	NC_005315.1
Spalax ehrenbergi	Glires; Rodentia; Sciurognathi; Muroidea; Spalacidae	AJ416891.1
Eospalax baileyi	Glires; Rodentia; Sciurognathi; Muroidea; Spalacidae	NC_018098.1
Sciurus vulgaris	Glires; Rodentia; Sciurognathi; Sciuridae; Sciurinae	NC_002369.1
Marmota himalayana	Glires; Rodentia; Sciurognathi; Sciuridae; Xerinae	NC_018367.1
Jaculus	Glires; Rodentia; Sciurognathi; Dipodidae; Dipodinae	NC_005314.1
Myoxus glis	Glires; Rodentia; Sciurognathi; Gliridae; Glirinae	NC_001892.1
Castor Canadensis	Glires; Rodentia; Sciurognathi; Castoridae; Castor	NC_015108.1
Anomalurus	Glires; Rodentia; Sciurognathi; Anomaluridae; Anomalurus	NC_009056.1
*Octodon degu*s	Glires; Rodentia; Hystricognathi; Octodontidae; Octodon	HM544134.1
Spalacopus cyanus	Glires; Rodentia; Hystricognathi; Octodontidae; Spalacopus	HM544133.1
Tympanoctomys barrerae	Glires; Rodentia; Hystricognathi; Octodontidae; Tympanoctomys	HM544132.1
Ctenomys leucodon	Glires; Rodentia; Hystricognathi; Ctenomyidae; Ctenomys	HM544131.1
Proechimys longicaudatus	Glires; Rodentia; Hystricognathi; Echimyidae; Proechimys	HM544128.1
Cavia porcellus	Glires; Rodentia; Hystricognathi; Caviidae; Cavia	AJ222767
Thryonomys swinderianus	Glires; Rodentia; Hystricognathi; Thryonomyidae; Thryonomys	NC_002658.1
Heterocephalus glaber	Glires; Rodentia; Hystricognathi; Bathyergidae; Heterocephalus	NC_015112.1
Lepus europaeus	Glires; Lagomorpha; Leporidae; Lepus	AJ421471
Oryctolagus cuniculus	Glires; Lagomorpha; Leporidae; Oryctolagus	AJ001588
Ochotona princeps	Glires; Lagomorpha; Leporidae; Oryctolagus	AJ537415

RESULTS

Genome organization and features

The mitochondrial DNA sequence of *Marmota himalayana* was identified as a circular molecule of 16,443 bp consisting of 2 rRNAs (*srRNA* and *lrRNA*), 13 PCGs, 22 tRNAs, and a control region (CR). The three different types of genes and their numbers were the same as previously identified in most vertebrate mt genomes. Among the 37 genes identified, 8 tRNA genes and the *ND6* gene were encoded on the L-strand, and the remaining 28 genes were encoded on the H-strand. The arrangement of the genes in the *M. himalayana* mt genome is summarized in Figure 1 and Table 3. The complete mt genome of *Marmota himalayana* is somewhat shorter than that of the Sciuridae species *Sciurus vulgaris* (16,507 bp). This size variation mainly resulted from differences in the length of the control region among the Rodentia species; therefore, when the CR region was excluded, the mt genome sequences of the Rodentia species were nearly identical in length. In addition, similar to the mt genomes in most vertebrates, overlaps and similar intervals were observed for several genes in the *Marmota himalayana* mt genome, with the most typical interval extending up to 30 nucleotides between the $tRNA^{Asn}$ and $tRNA^{Cys}$ genes.

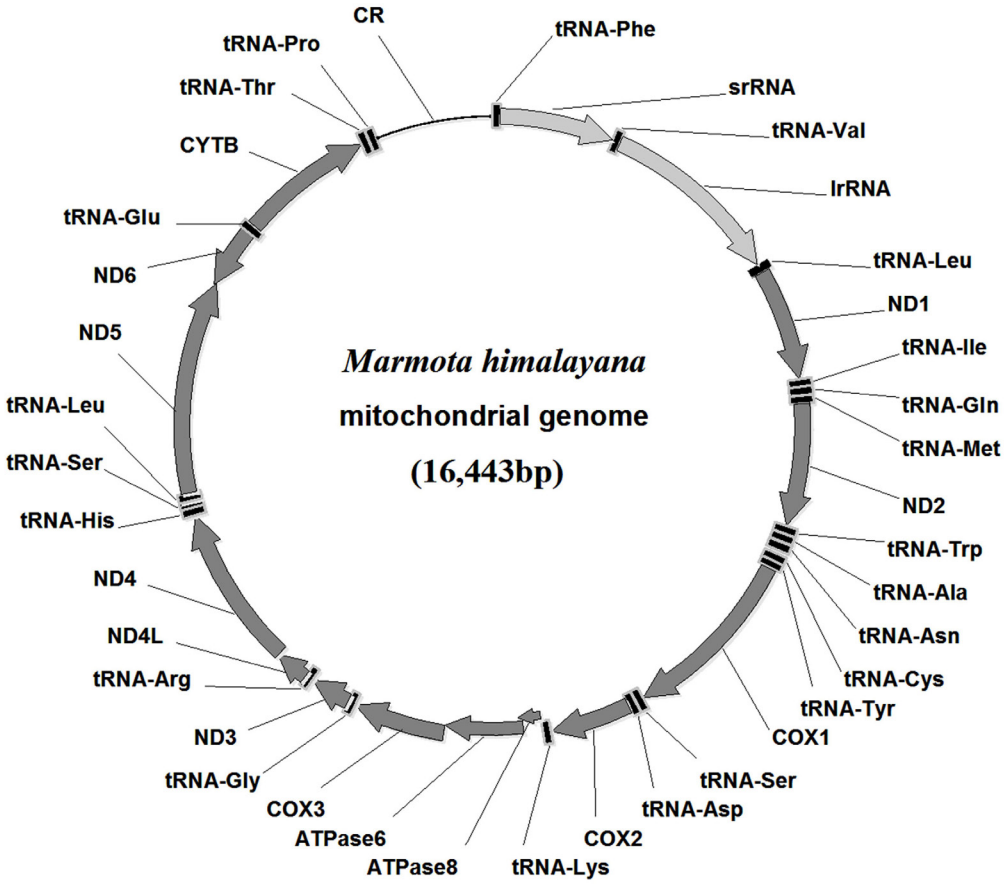

Figure 1. Gene map of the *Marmota himalayana* mtDNA. Protein-coding genes (deep-gray) and rRNA genes (light-gray) are identified by arrows. The *ND6* gene is transcribed in the opposite direction as relative to the rest PCGs. tRNA genes (black) are depicted with their corresponding amino acids.

Nucleotide composition and base bias

The total nucleotide base composition of the *M. himalayana* mt genome was A (32.13%) > T (31.36%) > C (23.73%) > G (12.78%) (Table 4). Its overall A+T content was 63.49%, which is similar to that in the mt genome of *Sciurus vulgaris* (62.97%). To investigate potential base bias, we calculated the AT and GC skews for the *M. himalayana* mt genome, including CR, protein-coding, and ribosomal RNA genes, and identified an AT skew of 0.012 and a GC skew of -0.300. We also noted that all GC skews and most of the AT skews were negative for the protein-coding genes, indicating that C and T residues were more prevalent than G and A residues in these genes, which is consistent with previous observations that vertebrate sequences have a bias against the use of G (Saccone et al., 1999). The nucleotide compositions in the rRNA genes (63.2% AT) and in the CR (62.2% AT) were similar to compositions in the complete mt genome, and, therefore, an A+T bias pattern holds for all functional segments of the *M. himalayana* mt genome.

Table 3. Organization of *Marmota himalayan* mitochondrial genome.

Gene	From	To	Size (bp)	Start codon	Stop codon	Strand	Amino acid
tRNA^{Phe}	1	69	69			H	F
srRNA	70	1040	971			H	
tRNA^{Val}	1041	1109	69			H	V
lrRNA	1110	2677	1568			H	
tRNA$^{Leu(UAA)}$	2678	2751	74			H	L
ND1	2755	3710	956	ATG	TA-	H	+3
tRNA^{Ile}	3711	3779	69			H	I
tRNA^{Gln}	3777	3848	72			L	Q -3
tRNA^{Met}	3852	3920	69			H	M +3
ND2	3921	4962	1042	ATT	T--	H	
tRNA^{Trp}	4963	5030	68			H	W
tRNA^{Ala}	5034	5102	70			L	A +3
tRNA^{Asn}	5108	5180	73			L	N +5
tRNA^{Cys}	5211	5277	67			L	C +30
tRNA^{Tyr}	5278	5343	66			L	Y
CO1	5352	6893	1545	ATG	TAG	H	+8
tRNA$^{Ser(UGA)}$	6896	6964	69			L	S +2
tRNA^{Asp}	6968	7036	69			H	D +3
CO2	7037	7720	684	ATG	TAA	H	
tRNA^{Lys}	7724	7790	67			H	K +3
ATPase8	7792	7995	204	ATG	TAG	H	+1
ATPase6	7953	8632	680	ATG	TT-	H	-3
CO3	8633	9416	784	ATG	T--	H	
tRNA^{Gly}	9417	9486	70			H	G
ND3	9487	9833	347	ATA	TA-	H	
tRNA^{Arg}	9834	9900	67			H	R
ND4L	9902	10198	297	ATG	TAA	H	+1
ND4	10192	11569	1378	ATG	T--	H	+7
tRNA^{His}	11570	11638	69			H	H
tRNA$^{Ser(GCU)}$	11640	11696	57			H	S +1
tRNA$^{Leu(UAG)}$	11698	11767	70			H	L +1
ND5	11768	13585	1818	ATT	TAA	H	
ND6	13569	14093	525	ATG	AGA	L	-17
tRNA^{Glu}	14094	14162	69			L	E
*CYT*B	14167	15306	1140	ATG	AGA	H	+3
tRNA^{Thr}	15307	15373	67			H	T
tRNA^{Pro}	15379	15446	68			L	P +4
CR	15447	16443	997			H	+997

Table 4. Nucleotide composition and skews of *Marmota himalayana* mitochondrial protein-coding and ribosomal RNA genes.

Gene	Proportion of nucleotides				%AT	AT skew	GC skew
	A	T	G	C			
ND1	31.0	32.6	11.0	25.4	63.6	-0.025	-0.396
ND2	34.3	30.7	8.6	26.4	65.0	0.055	-0.509
CO1	27.9	34.2	16.3	21.6	62.1	-0.102	-0.140
CO2	32.8	30.4	13.2	23.7	63.2	0.038	-0.285
ATPase8	35.3	32.4	7.8	24.5	67.7	0.043	-0.517
ATPase6	30.2	33.1	10.7	26.0	63.3	-0.046	-0.417
CO3	27.3	32.1	15.1	25.5	59.4	-0.081	-0.256
ND3	27.7	36.0	12.7	23.6	63.7	-0.130	-0.300
ND4L	28.0	37.4	11.4	23.2	65.4	-0.144	-0.341
ND4	31.5	34.0	10.0	24.5	65.5	-0.038	-0.420
ND5	31.3	33.0	10.0	25.7	64.3	-0.026	-0.440
ND6	39.8	23.6	7.3	29.3	63.4	0.256	-0.601
CYTB	28.9	32.9	12.5	25.7	61.8	-0.065	-0.346
srRNA	35.8	25.4	17.1	21.7	61.2	0.170	-0.119
lrRNA	37.0	27.4	16.6	19.0	64.4	0.150	-0.067
CR	30.3	31.9	12.9	24.9	62.2	-0.026	-0.317
Mitochondrial genome	32.1	31.4	12.8	23.7	63.5	0.011	-0.299

Protein-coding genes

The complete mtDNA of *M. himalayana* encodes 13 protein-coding genes (PCGs). The arrangement and directions of all 13 PCGs are the same as observed in other vertebrate mt genomes (Figure 1). Among the 13 PCGs, *ND2* and *ND5* use ATT as start codon, whereas *ND5* appears to use ATT instead of the canonical ATG start codon. As shown in Table 3, the other 10 PCGs had the standard ATG start codon. TAA as a stop codon was observed in *CO2*, *ND4L*, and *ND5*, whereas TAG was the stop codon in *CO1* and *ATPase8*. AGA was observed as the stop codon in *ND6* and *CYTB*, while the remaining 6 PCGs were apparently terminated by an incomplete stop codon, TA_ (*ND1*, *ND3*), TT_ (*ATPase6*), or T_ (*ND2*, *CO3*, *ND4*), thus differing from the typical stop codons found in PCGs (TAA, TAG, TGA). Incomplete stop codons commonly exist in vertebrate mt genomes, and these incomplete codons are typically restored to functional stop codons by posttranscriptional polyadenylation during mRNA maturation (Ojala et al., 1981).

Some of the 13 PCGs displayed sequence overlaps: for instance, the *ATPase8* and *ATPase6* overlapped by a 43-bp long sequence, and *ND4* and *ND4L* shared a seven-nucleotide overlap. Such gene overlaps are found in many vertebrate mtDNAs.

Ribosomal and transfer RNA genes

Small and large ribosomal RNA genes (*srRNA* and *lrRNA*) were 70 and 1110-bp long, respectively, and appeared to be transcribed from the H-strand of the *M. himalayana* mt genome. The rRNA genes were located between $tRNA^{Phe}$ and $tRNA^{Leu(UAA)}$ and separated by the gene for $tRNA^{Val}$, as observed for other vertebrate mt genomes (Delisle and Strobeck, 2002; Hwang et al., 2008).

Twenty-two tRNA genes, ranging 57-74 bp in size and scattered throughout the mt genome, were identified in *M. himalayana*. Of these, 21 were identified by tRNAscan, identified by typical cloverleaf secondary structures with normal base pairing and harboring anticodons that match the vertebrate mitochondrial genetic code (Figure 2). The remaining $tRNA^{Ser(GCU)}$ gene was 57 bp in size determined by visual inspection and comparison with the *Sciurus vulgaris* mt genome sequence. Unlike the other 21 tRNA genes, the $tRNA^{Ser(GCU)}$ gene cannot form a typical cloverleaf structure because it lacks the appropriate sequence to form the dihydrouridine arm and loop. This unusual secondary structure is also found in many mt genomes from other vertebrates, including bear, coyote, dogs, Eurasian otter, and wild camel (Peng et al., 2007; Hou et al., 2007; Cui et al., 2007; Hwang et al., 2008). Mismatches were observed in the amino acid-acceptor arms, anticodon, and TψC stems of some *M. himalayana* tRNA genes [$tRNA^{Phe}$, $tRNA^{Val}$, $tRNA^{Met}$, $tRNA^{Gly}$, $tRNA^{Leu(UAG)}$], which may be restored through RNA-editing mechanisms.

Non-coding regions

A total of 1075 bp in the *M. himalayana* mt genome corresponded to non-coding sequences, spread over 15 intergenic segments and ranging 1-997 bp in size. The largest (997 bp) non-coding region, CR, was identified between the $tRNA^{Pro}$ and $tRNA^{Phe}$ genes, with an A+T content of 62.2%. By comparing the complete CR region in *M. himalayana* with the corre-

sponding CRs in other mammalians (Sbisa et al., 1997, Cui et al., 2007; Hwang et al., 2008), we could compartmentalize the CR of *M. himalayana* into three domains: domain (D)-I, -II, and -III (Figure 3), corresponding to the termination-associated sequence domain (TAS), the central conserved domain (CD), and the conserved sequence block domain (CSB) (Sbisa E, 1997). We found a termination-associated sequence (TAS-A) and six conserved sequence blocks (CSB1 and B-F), which contain regulatory sequences controlling replication and transcription. Among Sciuridae animals, the TAS-A, CSB1, and CSBB-F are highly conserved; however, CSB2 and CSB3 are absent from some of them. In addition, we found no repetitive motifs in the CR of the *M. himalayana*, which is identical with that of the *Sciurus vulgaris* and *Myoxus glis*.

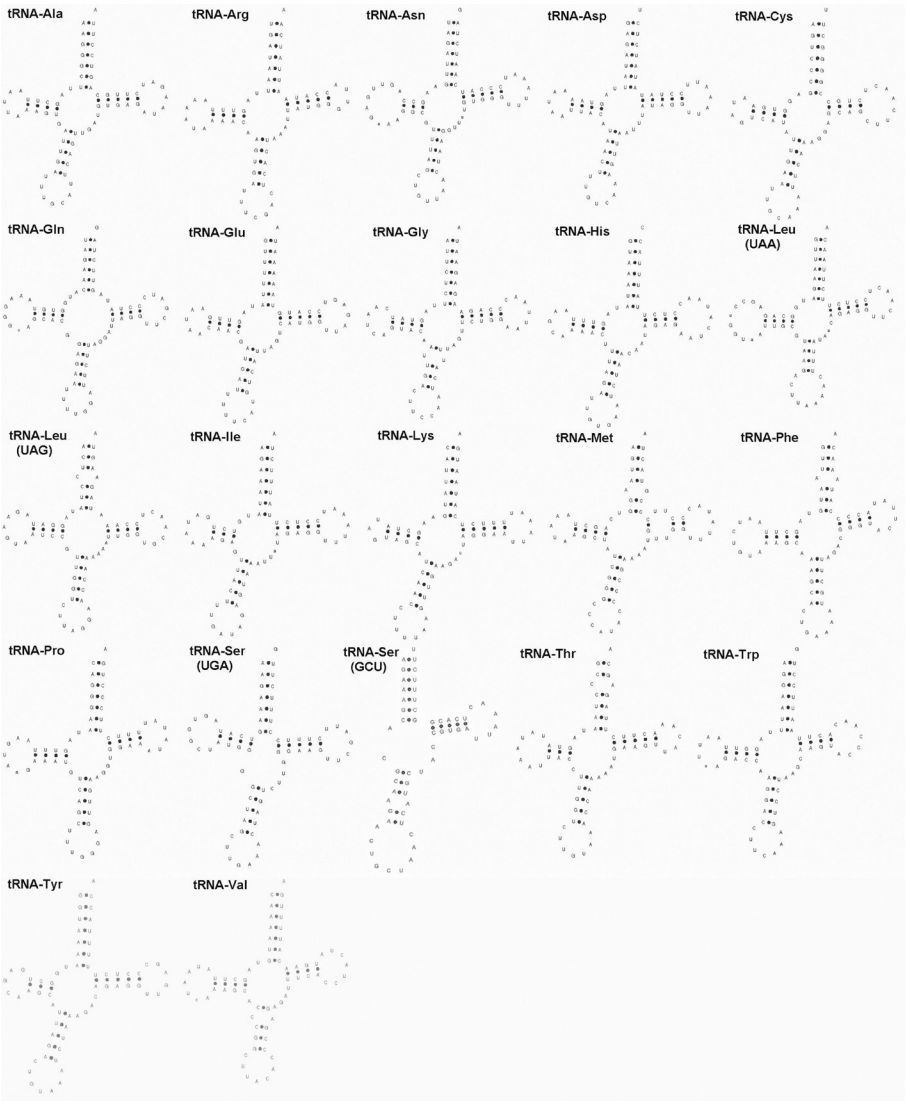

Figure 2. Putative secondary structures of 22 mitochondrial tRNAs from the *Marmota himalayana*.

Figure 3. Organization of the *Marmota himalayana* control region. The control region consists of the extended termination-associated sequence (TAS-A: in DI), and the central conserved (CSB-B, C, D, E, F: in DII) and conserved sequence block (CSB1; in DIII) domains.

Phylogenetic relationships within Rodentia

Tree topologies obtained by ML and BI analyses were almost identical (Figure 4). The phylogenetic tree supports Sciurognathi polyphyly and Hystricognathi monophyly. *M. himalayana* clustered with *S. vulgaris* representing Sciuridae with high statistical support (100% BP under ML and 1.00 PP under Bayesian analysis) and then *M. glis* (BP = 100%, PP = 1.00) formed a basal clade to the Sciurognathi in accordance with morphological and paleontological surveys, providing evidence for the close relationship between Gliridae and Sciuridae. The Hystricognathi taxa formed a highly supported monophyletic clade (100% BP and 1.00 PP), and formed a sister group to the Anomaluridae/Castoridae clade, with lower support (92% BP and 0.69 PP). In addition, both analyses strongly supported the monophyly of the Muroidea/Dipodidae grouping (all with PP = 1.00), and Dipodidae was found at the base of the Muroidea clade with 100% support and assumed to be basal in Muroidea. This observation was identical to a previous finding based on amino acid analysis of the 12 PCGs located on the H-strand (David et al., 2007). Our observations differed from those in some previous studies (Reyes et al., 2004; Kjer and Honeycutt, 2007; David et al., 2007) in that the Hystricognathi with Anomaluridae/Castoridae group was identified as a sister clade to the Muroidea/Dipodidae group with high statistical support (BP = 97%, PP = 0.99), but the Myoxidae/Sciuridae clade was at the base of these two groups, not being a sister clade to the Muroidea/Dipodidae group. The

M. himalayana sequence may have been responsible for this discrepancy, since its inclusion tended to cluster the Myoxidae/Sciuridae clade with the outgroup. However, the statistical support for the internal nodes of these clades was strong (92-100 BP and 0.69-1.00 PP) in both ML tree and BI trees.

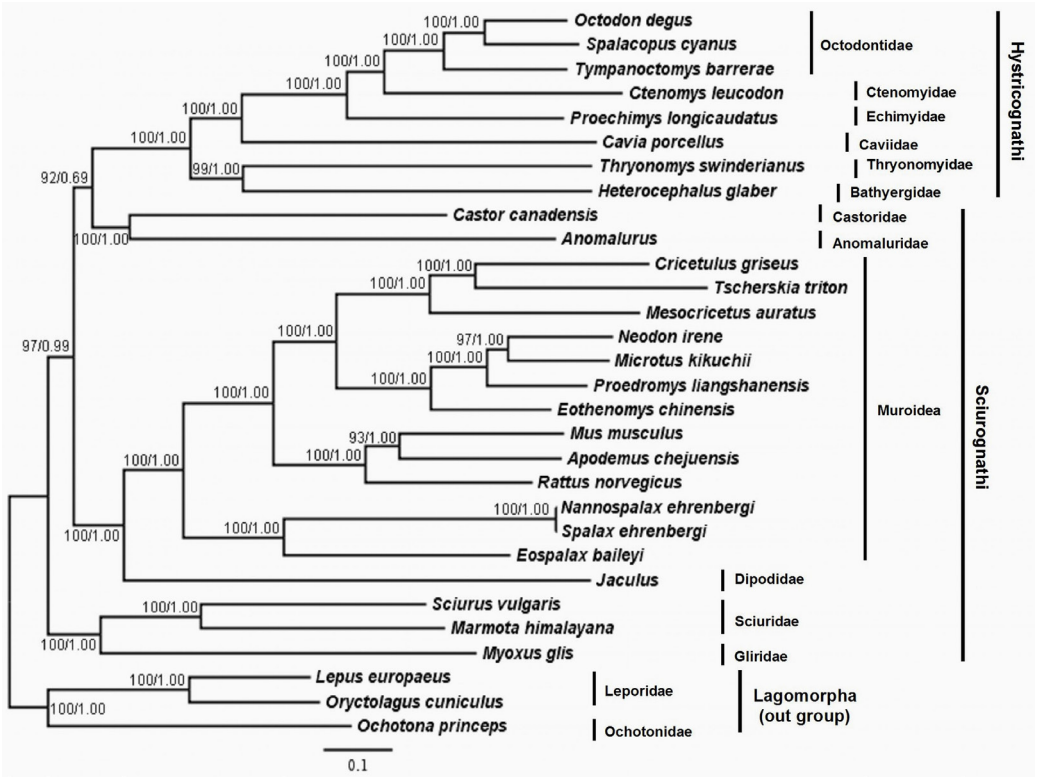

Figure 4. Phylogenetic tree obtained by the maximum-likelihood (ML) method and Bayesian analyses based on the nucleotide sequences of complete mitochondrial genome (without the control region). The first number at each node is the bootstrap probability of ML analyses and the second number is Bayesian posterior probability.

DISCUSSION

Here, we have determined the complete mitochondrial sequence of the marmot *M. himalayana*, and identified and characterized the sequences for the mitochondrial genes and the control region. In particular, we also used the *M. himalayana* mt genome sequence to analyze the phylogenetic relationships within Rodentia, including *M. himalayana* and 26 representative species with the three typical Lagomorpha taxa used as outgroup. Our data supported polyphyly of the Sciurognathi and monophyly of the Hystricognathi. As we expected, *M. himalayana* was found to be most closely related to taxa in the same genus (for example, *S. vulgaris*), followed by Gliridae (*M. glis*), regardless of the method used for phylogenetic reconstruction. The findings obtained in this study offer useful information for additional studies to gain a better understanding of the genetic content of the Rodentia and the commercially important marmota species.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31260259), the Anhui Provincial Natural Science Foundation (#1208085MC45), and the Natural Science Key Foundation (#KJ2010A299) from the Department of Education of Anhui Province.

REFERENCES

Armitage KB (1975). Social behavior and population dynamics of marmots. *Oikos* 26: 341-354.

Armitage KB (1987). Social dynamics of mammals: Reproductive success, kinship and individual fitness. *Trends Ecol. Evol.* 2: 279-284.

Armitage KB (1999). Evolution of sociality in marmots. *J. Mammal.* 80: 1-10.

Arnold W (1990). The evolution of marmot sociality: II. Costs and benefits of joint hibernation. *Behav. Ecol. Sociobiol.* 27: 239-246.

Avis JC (1994). Molecular Markers, Natural History and Evolution. 2nd edn. Kluwer Academic Publishers, New York.

Barash DP (1989). Marmots: Social Behavior and Ecology. 1st edn. Stanford University Press, Stanford Calif.

Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27: 573-580.

Blumstein DT, Steinmetz J, Armitage KB and Daniel JC (1997). Alarm calling in yellow-bellied marmots: II: The importance of direct fitness. *Anim. Behav.* 53: 173-184.

Boore JL (1999). Animal mitochondrial genomes. *Nucleic Acids Res.* 27: 1767-1780.

Brown WM (1983). Evolution of Genes and Proteins. 3rd edn. Sunderland MA Press, Sinauer Association.

Castro JA, Picornell A and Ramon M (1998). Mitochondrial DNA: a tool for populational genetics studies. *Int. Microbiol.* 1: 327-332.

Cui P, Ji R, Ding F, Qi D, et al. (2007). A complete mitochondrial genome sequence of the wild two-humped camel (*Camelus bactrianus ferus*): an evolutionary history of camelidae. *BMC Genomics* 8: 241.

David SH, Konstantinos L, Reyes A, Carmela G, et al. (2007). Phylogenetic analyses of complete mitochondrial genome sequences suggest a basal divergence of the enigmatic rodent *Anomalurus*. *BMC Evol. Biol.* 7: 16.

Delisle I and Strobeck C (2002). Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. *Mol. Biol. Evol.* 19: 357-361.

Giles RE, Blanc H, Cann HM and Wallace DC (1980). Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* 77: 6715-6719.

Hou WR, Chen Y, Wu X, Hu JC, et al. (2007). A complete mitochondrial genome sequence of Asian black bear Sichuan subspecies (*Ursus thibetanus mupinensis*). *Int. J. Biol. Sci.* 3: 85-90.

Hwang DS, Ki JS, Jeong DH, Kim BH, et al. (2008). A comprehensive analysis of three Asiatic black bear mitochondrial genomes (subspecies ussuricus, formosanus and mupinensis), with emphasis on the complete mtDNA sequence of *Ursus thibetanus ussuricus* (Ursidae). *Mitochondrial DNA* 19: 418-429.

Kjer KM and Honeycutt RL (2007). Site specific rates of mitochondrial genomes and the phylogeny of eutheria. *BMC Evol. Biol.* 7: 8.

Kjer KM, Blahnik RJ and Holzenthal RW (2001). Phylogeny of Trichoptera (caddisflies): characterization of signal and noise within multiple datasets. *Syst. Biol.* 50: 781-816.

Lowe TM and Eddy SR (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955-964.

Mathews DH, Sabina J, Zuker M and Turner DH (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288: 911-940.

Mathews DH, Disney MD, Childs JL, Schroeder SJ, et al. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.* 101: 7287-7292.

Miyata T, Hayashida H, Kikuno R, Hasegawa M, et al. (1982). Molecular clock of silent substitutions: at least a sixfold preponderance of silent changes in mitochondrial genes over those in nuclear genes. *J. Mol. Evol.* 19: 28-35.

Moritz C, Dowling TE and Brown WM (1987). Evolution of animal mtDNA: relevance for population biology and systematics. *Annu Rev. Ecol. Syst.* 18: 269-292.

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

Peng R, Zeng B, Meng X, Yue B, et al. (2007). The complete mitochondrial genome and phylogenetic analysis of the giant

panda (*Ailuropoda melanoleuca*). *Gene* 397: 76-83.
Reyes A, Pesole G and Saccome C (1998). Complete mitochondrial DNA sequence of the fast dormouse, *Glis glis*: further evidence of rodent paraphyly. *Mol. Biol. Evol.* 15: 499-505.
Reyes A, Gissi C, Pesole G, Catzeflis FM, et al. (2000). Where do rodents fit? Evidence from the complete mitochondrial genome of *Sciurus vulgaris*. *Mol. Biol. Evol.* 17: 979-983.
Reyes A, Gissi C, Catzeflis F, Nevo E, et al. (2004). Congruent mammalian trees from mitochondrial and nuclear genes using Bayesian methods. *Mol. Biol. Evol.* 21: 397-403.
Ronquist F and Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.
Saccone C, De Giorgi C, Gissi C, Pesole G, et al. (1999). Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* 238: 195-209.
Sbisa E, Tanzariello F, Reyes A, Pesole G, et al. (1997). Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* 205: 125-140.
Shadel GS and Clayton DA (1993). Mitochondrial transcription initiation. Variation and conservation. *J. Biol. Chem.* 268: 16083-16086.
Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690.
Steppan SJ, Akhverdyan MR, Lyapunova EA, Fraser DG, et al. (1999). Molecular phylogeny of the marmots (Rodentia: Sciuridae): tests of evolutionary and biogeographic hypotheses. *Syst. Biol.* 48: 715-734.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.
Wolstenholme DR (1992). Mitochondrial Genomes. 2nd edn. Academic Press, New York.
Wyman SK, Jansen RK and Boore JL (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252-3255.
Yamanoue Y, Miya M, Matsuura K, Yagishita N, et al. (2007). Phylogenetic position of tetraodontiform fishes within the higher teleosts: Bayesian inferences based on 44 whole mitochondrial genome sequences. *Mol. Phylogenet. Evol.* 45: 89-101.
Zhou Y, Zhang JY, Zheng RQ, Yu BG, et al. (2009). Complete nucleotide sequence and gene organization of the mitochondrial genome of *Paa spinosa* (Anura: Ranoidae). *Gene* 447: 86-96.

