Genetics and Molecular Research

Phylogenetic analysis of the Mongolian gerbil (*Meriones unguiculatus*) from China based on mitochondrial genome

C.L. Li[1], X.Y. Du[1], J. Gao[3], C. Wang[1], H.G. Guo[2], F.W. Dai[2], X.Y. Sa[2], W. An[1] and Z.W. Chen[1]

[1]School of Basic Medical Science, Capital Medical University, Beijing, China
[2]Zhejiang Center of Laboratory Animals, Zhejiang Academy of Medical Science, Hangzhou, China
[3]Shanghai Academy of Agriculture, Shanghai, China

Corresponding author: Z.W. Chen
E-mail: czwenteam@163.com

Genet. Mol. Res. 15 (3): gmr.15037703
Received January 4, 2016
Accepted July 25, 2016
Published September 16, 2016
DOI http://dx.doi.org/10.4238/gmr.15037703

ABSTRACT. *Meriones unguiculatus* (Gerbillinae, Rodentia) is widely used as an animal model of human disease. Here, we provide the first report of the complete mitochondrial genome sequence of *M. unguiculatus* (GenBank accession Nos. KF425526 and NC_023263). The sequence contained the conserved vertebrate pattern of 13 protein-coding genes, 2 ribosomal RNAs, 22 transfer RNAs, and 1 major noncoding region. We identified one extended termination-associated sequence and one conserved sequence block in the non-coding region. The putative origin of replication for the light strand (O_L) was 35 bp long. The O_L stem and adjacent sequences were highly conserved, but the loop region differed from those of other rodent species. Base composition and codon usage of the 13 protein-coding genes in *M.*

unguiculatus were compared with those of 23 rodent species with previously sequenced mitochondrial genomes. An A+T content of 63.0% was present in *M. unguiculatus*; this is similar to the Murinae average (62.4 ± 0.8%) and falls between the average for *Mus musculus* (63.1 ± 0.1%) and *Rattus* sp (61.7 ± 0.4%). The AT and GC skew values of *M. unguiculatus* were 0.035 and -0.28, respectively, similar to those of Cricetinae species (0.057 ± 0.05 and -0.31 ± 0.05). The codon families exhibited similar abundance in all 24 species. Analysis of phylogenetic relationships with 23 other rodent species using neighbor-joining and maximum likelihood protocols and the 12 protein-coding regions on the H strand showed that *M. unguiculatus* should be classified as genus *Meriones*, sub-family Gerbillinae, family Muridae.

Key words: Mongolian gerbil; Mitochondrial genome; Phylogeny; Muridae

INTRODUCTION

Meriones unguiculatus (Alston,1876), commonly called the Mongolian gerbil, is primarily distributed in the desert, grasslands, and agricultural fields of Inner Mongolia of China, Mongolia, and Russia (Liu et al., 2007). In the laboratory, *M. unguiculatus* has been widely used as an animal model for human disease. The levels of serum cholesterol of *M. unguiculatus* are susceptible to dietary supplements and do not induce atherosclerosis, a phenomenon that is valuable for studies of cholesterol absorption and metabolism (Hegsted and Gallagher, 1967). On standard diets, approximately 10% of *M. unguiculatus* individuals became obese and showed decreased glucose tolerance, elevated serum immunoreactive insulin, and diabetic changes in the pancreas and other organs (Vincent et al., 1979). Liu et al. (2012) found that *M. unguiculatus* with high body weights displayed the expected increase in serum leptin and insulin concentrations. Due to its lack of significant collateral flow from the vertebral blood supply to the forebrain, a transient bilateral common carotid artery occlusion induces consistent ischemic injury (Du et al., 2010). *M. unguiculatus* infected with *Helicobacter pylori* mimic human *H. pylori* infection, facilitating the analysis of the underlying processes and the study of gastritis and gastric cancer (Wei et al., 2012). *M. unguiculatus* has also been used in studies of nutrition (Ying et al., 2013), aging (Ohlemiller, 2009), epilepsy (Heo and Kang, 2012), parasites (Conchedda et al., 2006), ethology (Moons et al., 2012), and viral infections (Yao et al., 2012). However, although it is an important experimental animal, the available genetic information on gerbils is very limited and its phylogenetic classification is uncertain.

M. unguiculatus is classified under Gerbillinae; however, the taxonomy of Gerbillinae has been debated for a considerable time. Alston (1876) placed Gerbillinae within the family Muridae based principally on mandibular form, but Miller and Gidley (1918) placed Gerbillinae within the family Cricetidae. Pavlinov (2001) found that Cricetidae and gerbils have similar dentition. However, some results from gene molecular analyses of single karyogenes or partially sequenced mitochondrial genes indicated that Gerbillinae is a sister group of Murinae (Chevret and Dobigny, 2005). Nevertheless, Gerbillinae is placed in the family Cricetidae in several textbooks (Han et al., 2005). *M. unguiculatus* (Taxonomy ID: 10047) is classified in the family Muridae on the NCBI taxonomy website and the family Cricetidaeon the Chinese

Biodiversity Information System and Chinese Agriculture Pests Information system (CAPIS) websites. Inrecent reports, two other *Meriones* species, *M. libycus* and *M. meridianus*, were classified into the Cricetidae (Luo and Liao, 2015a,b). Although mitochondrial genome sequences of *M. unguiculatus* have been reported (Kim and Lee, 2016), the taxonomy of *M. unguiculatus* has not yet been analyzed. Sequencing the complete mitochondrial genome should help to resolve the taxonomy of *M. unguiculatus*.

Mitochondria are important subcellular organelles present in almost all animal cells. They facilitate the reactions in the tricarboxylic acid cycle and electron transfer in oxidative phosphorylation that are required for cellular energy generation (Chao et al., 2014). Mitochondrial genomes differ from nuclear genomes in many ways, e.g., they are inherited maternally, experience higher mutation rates, and do not undergo recombination. Consequently, they are useful in phylogenetic, population genetic, and molecular evolution studies (Yoon and Park, 2015). Genome level analyses that include nucleotide composition, codon usage, gene order arrangement, and secondary structures of tRNAs and rRNAs, are powerful tools for inferring higher-level phylogenies and investigating molecular evolution (Boore, 1999). Analyses of complete mitochondrial genome sequences also aid such investigations.

In this study, we report the complete mitochondrial genome of *M. unguiculatus* and investigate gene arrangement, nucleotide composition, and codon usage of this sequence. We also analyzed the phylogenetic relationship of *M. unguiculatus* within the Rodentia using the sequences of 12 protein-coding genes (PCGs) on the H strand.

MATERIAL AND METHODS

Samples, DNA extraction, PCR amplification, and sequencing

The experiments and animal procedures were approved by the Animal Experiments and Experimental Animal Welfare Committee of Capital Medical University (Permit No. 2011-X-009). Genomic DNA was extracted from *M. unguiculatus* liver tissue, which was obtained from the Zhejiang Center of Laboratory Animals. The complete mitochondrial genome was amplified by long distance PCR using a GeneAmp 9700 thermocycler (Applied Biosystems, USA). The primers were synthesized based on reported sequences. The main primer sequences are shown in Table 1.

Table 1. Primers used to amplify the complete mitochondrial genome of *Meriones unguiculatus*.

Primer name	Sense primer (5'-3') sequence	Anti-sense primer (5'-3') sequence
Primer 1	ATCGGACAAGTCGCTTCAAT (*Cytb* 3')	AGCGATGGCTCGTAGTTCTC (*12S* 5')
Primer 2	AATTGAATTGAGCCATGAAGC (*12S* 3')	TAATGGGGGAGGAAGCATCT (*Cox2* 5')
Primer 3	TCCAACCACAGCTTTATACCG (*Cox2* 3')	GGCATGGACTTTTTCTGCAT (*ND3* 5')
Primer 4	GGGGAGTGGGAAGATCAATAA (*ND3* 3')	CGCACAATCTAACAGCATCC (*Cytb* 5')

Four pairs of primers were used to perform long-distance PCR spanning *Cytb* to *12S RNA* (Primer 1), *12S RNA* to *Cox2* (Primer 2), *Cox2* to *ND3* (Primer 3), and *ND3* to *Cytb* (Primer 4). PCR amplification was performed in a final 50-μL reaction mixture containing 5 μL 10 X LA PCR buffer II (TaKaRa, Dalian, China), 8 μL 4 mM dNTP, 200 pmol each primer, 2 U ExTaq polymerase, and 200 ng genomic DNA sample. Amplification was performed using the following protocol: denaturation for 5 min at 94°C, followed by 35 cycles of denaturation

for 30 s at 94°C, annealing for 30 s at 55-60°C, and extension for 2-6 min at 72°C; and a final extension for 10 min at 72°C. PCR products were resolved by electrophoresis on 1.0% agarose gels, and extracted using a DNA Gel Extraction Kit (Qiagen, Valencia, CA, USA). Extracted DNA was sent to Sangon Biotech Co. Ltd. (Sangon, Shanghai, China) for sequencing using a primer-walking strategy.

Four long amplification fragments were obtained; these ranged from approximately 1500 to 6000 bp. The internal sequences were obtained by designing nested sequencing primers for "primer walking" within the amplified products. To confirm the mtDNA sequence, we designed another 12 pairs of primers based on sequences obtained in the first run; these primers amplified sequences from 1000 to 1600 bp. We then performed a second round of PCR and sequenced the PCR products in both directions.

Gene identification and genome analysis

All sequences were aligned and checked by visual inspection using the program BioEdit v7.0.1 (Hall, 1999). The boundaries of protein-coding genes and rRNA genes were predicted using DOGMA (Wyman et al., 2004) with the default settings and refined by alignment with the mitochondrial genomes of genera in the Cricetinae, Arvicolinae, and Murinae. Most tRNA genes were identified by tRNAscan-SE1.2.1 (Lowe and Eddy, 1997) under the default search mode using the vertebrate mitochondrial genetic code and 'Mito/chloroplast' source. RNA structure 5.2 (Reuter and Mathews, 2010) and RNAviz 2.0 (De Rijk et al., 2003) were used to predict and draw the secondary structure of tRNA genes to identify which of the gene products could be folded into the typical cloverleaf secondary structure.

The base compositions of the complete mitochondrial genomes of 24 rodent species were calculated with the EditSeq program from the Lasergene software package. GC-skew = (G-C) / (G+C) and AT-skew = (A-T) / (A+T) were used to compare the base compositions of different strands (Perna and Kocher, 1995). The codon usage of PCGs from 24 rodent species was calculated with the MEGA 5.0 program (Tamura et al., 2011). The mitochondrial genome sequence of *M. unguiculatus* was deposited in GenBank under accession Nos. KF425526 and NC_023263.

Phylogenetic analysis

To investigate the phylogenetic position of *M. unguiculatus*, the complete mitochondrial genomes of 23 rodent species were collected from NCBI (**Table S1**). Twelve PCGs (all but the L-strand coding gene *ND6*) were aligned with MEGA5 (Tamura et al., 2011). Maximum likelihood (ML) and neighbor-joining (NJ) trees were constructed with MEGA5. We identified the most appropriate model of DNA substitution in the ML analysis using MEGA5. Statistical confidence was assessed through the bootstrap analysis with 200 replications of the ML analysis and 1000 replications of the NJ analysis.

RESULTS

Mitochondrial genome structure

The mitochondrial genome of *M. unguiculatus* was a circular molecule of 16,351 bp; it included the 37 genes that are usually present in animal mitochondrial genomes, i.e., 13

PCGs, 22 tRNA genes, and 2 ribosomal genes (Table 2). The PCG *ND6* and 8 of the tRNA genes ($tRNA^{Gln}$, $tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Ser}$, $tRNA^{Glu}$, and $tRNA^{Pro}$) were encoded on the H-strand, and the other genes were encoded on the L-strand. As shown in Table 2, the genes were compactly arranged, with few gaps between them, and some genes overlapped. The longest overlap, 43 nucleotides, involved *ATPase8* and *ATPase6*.

Table 2. Elements and base composition of *Meriones unguiculatus* mitochondrial genome.

Element	Position	Size	Strand	Strat	Stop	Intergenic nucleotides	T(U)%	C%	A%	G%	A+T%	G+C%
$tRNA^{Phe}$	1-66	66	H				23.4	17.1	40.6	18.7	64.0	35.8
12S rRNA	67-1,023	943	H			0	24.3	23.0	34.8	17.9	59.1	40.9
$tRNA^{Val}$	1,024-1,090	67	H			0	20.9	25.4	38.8	14.9	59.7	40.3
16S rRNA	1,091-2,671	1,581	H			0	27.4	18.7	37.0	16.9	64.4	35.6
$tRNA^{Leu1}$	2,673-2,746	74	H			1	25.6	21.6	36.5	16.2	62.1	37.8
NADH1	2,747-3,703	957	H	GTG	TAG	0	32.0	25.8	29.3	13	61.3	38.8
$tRNA^{Ile}$	3,702-3,769	68	H			-2	33.8	14.7	35.3	16.1	69.1	30.8
$tRNA^{Gln}$	3,767-3,837	71	L			-3	28.1	16.9	21.1	33.8	49.2	50.7
$tRNA^{Met}$	3,848-3,917	70	H			10	20.0	30.0	28.6	21.4	48.6	51.4
NADH2	3,921-4,958	1,038	H	ATG	TAA	4	32.3	24.7	34.0	9.1	66.3	33.8
$tRNA^{Trp}$	4,958-5,022	65	H			-1	28.8	20.0	36.9	18.5	65.7	38.5
$tRNA^{Ala}$	5,025-5,093	69	L			2	31.9	11.6	21.7	34.7	53.6	46.3
$tRNA^{Asn}$	5,099-5,170	72	L			5	29.1	20.8	33.3	16.7	62.4	37.5
O_L	5,171-5,204	34	L			0	22.7	27.3	25.0	25.0	47.7	52.3
$tRNA^{Cys}$	5,202-5,268	67	L			-3	34.3	19.4	31.3	14.9	65.6	34.3
$tRNA^{Tyr}$	5,269-5,334	66	L			0	28.7	18.2	30.3	22.7	59.0	40.9
COX-I	5,336-6,880	1,547	H	ATG	TAA	1	31.8	22.8	29.6	15.7	61.4	38.5
$tRNA^{Ser1}$	6,878-6,946	69	L			-3	29.2	23.5	33.2	14.0	62.4	37.5
$tRNA^{Asp}$	6,950-7,018	69	H			3	32.3	10.2	45.6	11.7	77.9	21.9
COX-II	7,019-7,702	684	H	ATG	TAA	0	29.2	23.5	33.2	14.0	62.4	37.5
$tRNA^{Lys}$	7,705-7,771	67	H			2	35.8	13.4	17.9	32.8	53.7	46.2
ATPase 8	7,773-7,976	204	H	ATG	TAA	1	32.4	25	35.8	6.9	68.2	31.9
ATPase 6	7,934-8,614	681	H	ATG	TAA	-43	31.6	25.3	32.5	10.7	64.1	36.0
COX-III	8,614-9,397	784	H	ATG	T--	-1	30.5	25.6	28.7	15.2	59.2	40.8
$tRNA^{Gly}$	9,398-9,465	68	H			0	29.4	17.6	39.7	13.2	69.1	30.8
NADH3	9,466-9,813	348	H	ATT	TAA	0	36.2	25	27.6	11.2	63.8	36.2
$tRNA^{Arg}$	9,816-9,882	67	H			2	40.3	8.9	40.3	10.4	80.6	19.3
NADH4L	9,884-10,180	297	H	GTG	TAA	1	34.0	23.9	30.0	12.1	64.0	36.0
NADH4	10,174-11,551	1,378	H	ATG	T--	-7	32.1	25.3	32.4	10.1	64.5	35.4
$tRNA^{His}$	11,552-11,616	65	H			0	30.7	18.5	36.9	13.8	67.6	32.3
$tRNA^{Ser2}$	11,617-11,675	59	H			0	31.8	14.5	27.5	26.1	59.3	40.6
$tRNA^{Leu}$	11,676-11,743	68	H			0	30.8	17.6	35.3	16.1	66.1	33.7
NADH5	11,744-13,555	1,812	H	ATT	TAA	0	32.3	24.8	32.4	10.4	64.7	35.2
NADH6	13,552-14,073	522	L	ATG	TAA	-4	40.8	8.6	22.4	28.2	63.2	36.8
$tRNA^{Glu}$	14,074-14,142	69	L			0	42.0	13.0	26.0	18.8	68.0	31.8
Cyt *b*	14,147-15,286	1,040	H	ATG	TAA	3	32.0	26.1	28.2	13.7	60.2	39.8
$tRNA^{Thr}$	15,291-15,357	67	H			4	29.8	17.9	34.3	17.9	64.1	35.8
$tRNA^{Pro}$	15,358-15,427	70	L			0	24.6	20.0	18.5	30.0	43.1	50.0
CR	15,428-16,351	914	H			0	32.7	22.7	32.0	12.6	64.7	35.3

Base composition analysis

The A+T and G+C contents of the H-strand and the A-skew and C-skew values (Perna and Kocher, 1995) for the mitochondrial genomes of *M. unguiculatus* and 23 other rodent species are listed in Figure 1. The base composition of the H-strand of *M. unguiculatus* was A = 5332 (32.6%), T = 4976 (30.4%), G = 2165 (13.2%), and C = 3880 (23.7%). The base composition of the 13 PCGs of *M. unguiculatus* was AT rich (63.0%). A (31.3%) was the most common base, and G (13.6%) was the least common. The base composition of each gene and other sequences are shown in Table 2.

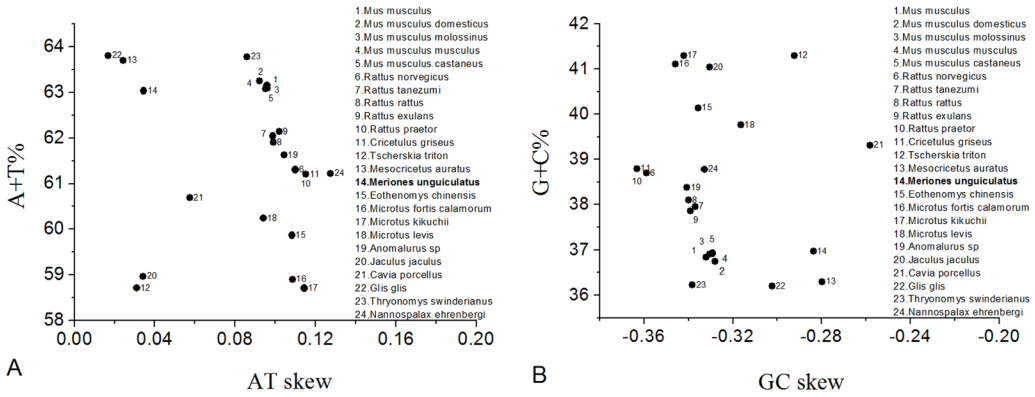

Figure 1. AT skew *vs* A+T% and CG skew *vs* G+C% in the mitochondrial genomes of *Meriones unguiculatus* and 23 other rodent species. Values were calculated using the H-strands of the mitochondrial genomes. The X-axis provides the skew value, while the Y-axis provides the A+T/G+C value. Species names are listed according to their taxonomic placement at the order level (see **Table S1**).

The average genome-wide A+T content for rodent mitochondrial genomes was 61.7 ± 1.7%. It ranged from 58.7% (*Microtus kikuchii*) to 64.2% (*Cricetulus griseus*). The *M. unguiculatus* mitochondrial genome had 63.0% A+T content, similar to the Murinae average (62.4 ± 0.8%) and between that of *Mus* and *Rattus* species. The average A-skew and C-skew values for the analyzed mitochondrial genomes were 0.09 ± 0.03 and -0.33 ± 0.03, respectively. A-skew ranged from 0.02 (*Nannospalax ehrenbergi*) to 0.13 (*Glis glis*), and C-skew ranged from -0.36 (*Rattus praetor*) to -0.26 (*Cavia porcellus*). The A-skew and C-skew values of *M. unguiculatus* were 0.035 and -0.28, respectively, which are close to those of Cricetinae (0.057 ± 0.05 and -0.31 ± 0.05).

PCGs

The total length of the 13 PCGs in *M. unguiculatus* was 10,868 bp, accounting for 66.46% of the complete mitochondrial genome. As shown in Table 2, ATG was the start codon of 9 PCGs, whereas GTG was used by *ND1* and *ND4L*, and ATT was used by *ND3* and *ND5*. The coding regions of *ND2*, *Cox1*, *Cox2*, *ATPase8*, *ATPase6*, *ND3*, *ND4L*, *ND5*, *ND6*, and *Cytb* were terminated by TAA, where as those of *ND1* were terminated by TAG. Two incomplete stop codons T(--) were found in *Cox3* and *ND4*. Several pairs of genes overlapped: *ATPase8* and *ATPase6* (43 nucleotides), *ATPase6* and *Cox2* (1 nucleotide), *ND4L* and *ND4* (7 nucleotides), *ND5* and *ND6* (4 nucleotides) (Table 2).

Codon usage analysis

Codon bias was apparent in the mitochondrial genome of *M. unguiculatus*. AUU (Ile), CUA (Leu), and AUA (Met) were the three most frequently used codons (Table 3). AUU (Ile) was the most frequently used codon; occurring 214 times, twice as a start codon (for *ND3* and *ND5*). The amino acid counts for each of the 24 rodent genomes are presented in **Figure S1**. On average, the 13 PCGs coded 3798.6 ± 8.3 amino acids. The codon count ranged from 3772 in *Thryonomys swinderianus* to 3810 in *Anomalurus* sp and *Rattus rattus*.

The *M. unguiculatus* amino acid count was similar to the Murinae average (3803.0 ± 3.8). The codon families exhibited similar distributions in all 24 species (**Figure S1**). Nine codon families each represented more than 50 of every thousand codons (Ala, Leu1, Ile, Phe, Pro, Met, SeC, Gly, Thr). Together they accounted on average for 65.2 ± 1.3% of the codons in each genome. The Leu1 codon family represented more than 100 of every 1000 codons and on average represented 11.8 ± 0.8% of the codons in each genome (**Figure S1**). The abundance of codon families in PCGs was also investigated in mitochondrial genomes and the results are summarized in **Figure S1** (Salvato et al., 2008). First codons, as well as complete and incomplete stop codons, were excluded from the analysis to avoid biases due to unusual putative start codons and incomplete stop codons.

Table 3. Codon usage of 13 PCGs in the mitochondrial genome of *Meriones unguiculatus*.

Amino acid	Codon	Number	Amino acid	Codon	Number	Amino acid	Codon	Number	Amino acid	Codon	Number
Ala	GCG(A)	6	Cys	UGC(S)	13	Gln	CAG(Q)	14	Ser2	UCG(S)	11
	GCA(A)	79		UGU(S)	12		CAA(Q)	81		UCA(S)	98
	GCC(A)	71	Leu2	UUG(L)	20	His	CAC(H)	42		UCU(S)	70
	GCU(A)	70		UUA(L)	168		CAU(H)	53		UCC(S)	61
Asn	AAC(N)	77	Glu	GAG(E)	31	Ile	AUC(I)	126	Thr	ACG(T)	11
	AAU(N)	83		GAA(E)	73		AUU(I)	214		ACA(T)	138
Arg	CGG(R)	1	Gly	GGG(G)	45	Pro	CCG(P)	7		ACC(T)	75
	CGA(R)	42		GGA(G)	77		CCA(P)	99		ACU(T)	89
	CGC(R)	9		GGC(G)	34		CCC(P)	44	Tyr	UAC(Y)	57
	CGU(R)	7		GGU(G)	41		CCU(P)	53		UAU(Y)	78
Asp	GAC(D)	33	Lys	AAG(K)	10	Ser1	AGG(S)	2	Val	GUG(V)	19
	GAU(D)	38		AAA(K)	87		AGA(S)	1		GUA(V)	72
Leu1	CUG(L)	24	Met	AUG(M)	43		AGC(S)	43		GUC(V)	24
	CUA(L)	206		AUA(M)	184		AGU(S)	21		GUU(V)	43
	CUC(L)	73	Phe	UUC(F)	108	Trp	UGG(W)	13	End	UAA(*)	20
	CUU(L)	116		UUU(F)	145		UGA(W)	85		UAG(*)	15

Transfer and ribosomal RNA genes

Two rRNA genes (12S rRNA and 16S rRNA) were identified using the default settings of DOGMA and refined through alignment to homologous genes in other mitochondrial genomes. The 12S rRNA gene was 943 bp, and the 16S gene was 1581 bp. They had 59.1 and 64.4% A+T contents, respectively. The lengths of the *M. unguiculatus* tRNA genes ranged from 59 to 72 bp. Twenty-one of the 22 tRNA genes could fold into a cloverleaf secondary structure. The $tRNA^{ser(AGN)}$ was only 59 bp in length and was missing the dihydrouridine stem and loop (Figure 2). Twelve of the *M. unguiculatus* mitochondrial tRNA genes ($tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Glu}$, $tRNA^{Gly}$, $tRNA^{Gln}$, $tRNA^{Met}$, $tRNA^{Phe}$, RNA^{Thr}, $tRNA^{Ser}$, $tRNA^{Pro}$, RNA^{Val}, and $tRNA^{Tyr}$) had mismatches in their stems. The mismatches were located mostly in the acceptor and anticodon stems.

Control region and putative origin of replication for the L-strand (O_L)

The control region (CR) of the *M. unguiculatus* mitochondrial genome was located between $tRNA^{Pro}$ and $tRNA^{Phe}$. The base composition of the CR is reported given in Table 2. We found an extended termination-associated sequence (ETAS-1; 15460-15518) and a conserved sequence block (CSB-1; 16078-16101) in the *M. unguiculatus* mitochondrial genome, consistent with studies of the mitochondrial genomes of other vertebrate species (Sbisà et al., 1997; Jiang et al., 2012). We did not observe CSB-2 or CSB-3 in *M. unguiculatus*.

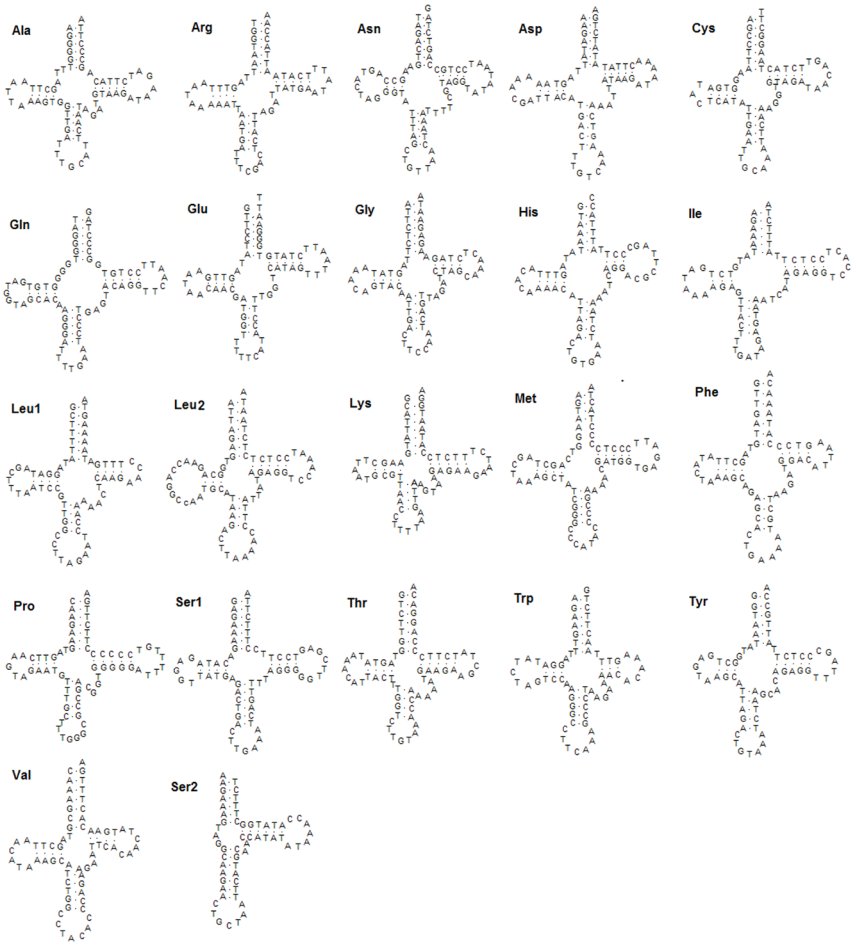

Figure 2. Potential secondary structures of the 22 inferred tRNAs in the *Meriones unguiculatus* mitochondrial genome.

The putative origin of replication for the O_L is located in the WANCY cluster between $tRNA^{Asn}$ and $tRNA^{Cys}$. O_L is reported to be required for L-strand replication (Hixson et al., 1986). We compared the secondary structure of O_L across all genera of Gerbillinae, Murinae, Arvicolinae, and Cricetinae (Figure 3). The *M. unguiculatus* O_L was 34 bp in length, and the O_L motif showed high conservation in the stem and adjacent sequences.

Phylogenetic analysis

To determine the phylogenetic relationships of *M. unguiculatus*, the concatenated nucleotide sequence of the 12 PCGs encoded on the H-strand were used to construct phylogenetic trees using NJ and ML methods. The *ND6* gene, which is encoded on the L-strand, was not used because it deviates markedly in nucleotide and amino acid composition from the other PCGs (Arnason et al., 2002).

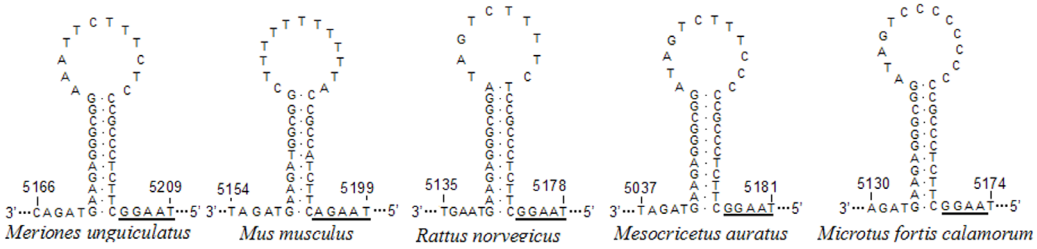

Figure 3. Stem and loop structures of *Meriones unguiculatus*, *Mus musculus*, *Rattus norvegicus*, *Mesocricetus auratus*, and *Microtus fortis calamorum*. The stem and adjacent sequences of O_L showed consistent sequence conservation except in *M. musculus*, but the loop sequences showed some differences among species.

The phylogenetic trees based on NJ and ML analysis of the 12 PCGs are shown in Figure 4. The trees had similar topologies. The model selected for the ML analysis was the general time-reversible model (GTR), with a gamma distributed shape parameter (G) of 0.9427 and a proportion of invariable sites (I) of 0.3527. The model test indicated that the GTR + G + I model (-lnL=116,389) was the most appropriate one for the ML analysis. The two trees provided an overview of the principal divisions and relationships of *M. unguiculatus* within the Rodentia. *M. unguiculatus* was in the Gerbillinae clade and had a high phylogenetic relationship with the Murinae. Gerbillinae, Arvicolinae, Cricetinae, and Murinae were sister groups. Our data confirmed that Gerbillinae, including *M. unguiculatus*, should be classified within the Muridae.

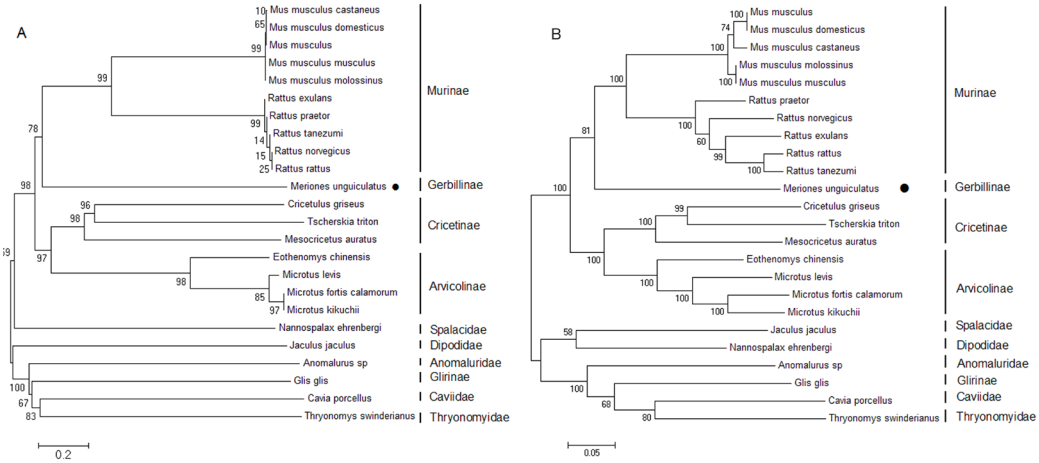

Figure 4. Analysis of phylogenetic relationships using neighbor-joining (NJ; **A**) and maximum likelihood (ML; **B**) methods and the concatenated 12 protein-coding gene sequences. Percentage bootstrap values are shown on interior branches with 1000 replicates for the NJ method and 100 replicates for the ML method.

DISCUSSION

The complete mitochondrial genome of *M. unguiculatus* was obtained by PCR and sequencing and was found to be 16,351 bp; this was slightly longer than the average mitochondrial genome of Murinae (16,301.0 ± 4.08 bp), Cricetinae (16,348.3 ± 121.8 bp), and

Arvicolinae (16,316.8 ± 32.9 bp), but shorter than that of *Anomalurus* sp (NC_009056.1;16,923 bp), *Jaculus jaculus* (NC_005314.1; 16,546 bp), *C. porcellus* (NC_000884.1; 16,801 bp), *G. glis* (NC_001892.1; 16,602 bp), *T. swinderianus* (NC_002658.1; 16,626 bp), and *N. ehrenbergi* (NC_005315.1; 16 408 bp). This suggests that *M. unguiculatus* has a close evolutionary relationship with Murinae, Cricetinae, and Arvicolinae. The complete mitochondrial genome of *M. unguiculatus* included the 37 genes that are usually present in animal mitochondrial genomes and that show a highly conserved gene order in most vertebrates (Boore, 1999; Jiang et al., 2012). Similar to that of Murinae and Cricetinae species, the mitochondrial genome of *M. unguiculatus* was tightly organized with some overlapping of gene sequences. Overlapping genes are common in animal mitochondrial genomes (Boore, 1999; Jiang et al., 2012). The unusual start codons GTG and ATT were observed in 4 PCGs; these codons have also been observed in other mammals (Jiang et al., 2012). The stop codons TAG and T(--) were present in 3 PCGs. Completion of the stop codon TAA can occur post-transcriptionally through polyadenylation of the mRNAs (Ojala et al., 1981). Incomplete stop codons have been observed in many other mammals (Zhong et al., 2010; Jiang et al., 2012).

The 21 tRNA genes have the classic cloverleaf structure, but $tRNA^{Ser(AGN)}$ was only 59 bp in length and was missing the dihydrouridine stem and loop (Figure 2). This loss is common among vertebrates (Jiang et al., 2012). We also found that the cloverleaf structure of 12 tRNA genes had mismatches in their stems. Mismatches on tRNA stems have been observed in the following tRNA genes of *Microtus fortis calamorum*: $tRNA^{Ala}$, $tRNA^{Gln}$, RNA^{Val}, $tRNA^{Thr}$, $tRNA^{Ser}$, and $tRNA^{Trp}$ (Jiang et al., 2012). Mismatches in tRNAs may be corrected through well-described mitochondrial RNA-editing mechanisms (Lavrov et al., 2000).

ETAS-1 and CSB-1 were found in the CR of the *M. unguiculatus* mitochondrial genome. The sequence motif "5'-TATAGTACAT" is complementary to the motif "3'-ATATTATGTA" and forms a stable hairpin secondary structure in ETAS-1that serves as a recognition site for the arrest of H-strand synthesis (Jiang et al., 2012). The sequence motif "-ATG-GACATA" was the most highly conserved sequence in CSB-1 in many rodent species (Jiang et al., 2012). CSB-1 has been studied in association with the mitochondrial replication origin (Chang and Clayton, 1986).

DNA synthesis in the L-strand is initiated through priming at the T-rich template sequence in the loop (Hixson et al., 1986). The T-rich sequence is replaced by poly(C) in the O_L loop region in *M. f. calamorum* (Jiang et al., 2012). The T-rich template sequence of *M. unguiculatus* was similar to those in members of Murinae and Cricetinae. *M. musculus* species has uniform O_L sequences and structures. The adjacent sequence 5'-TAAGG-3' at the base of the O_L stem (not 5'-GCCGG-3') was found in all genera of Gerbillinae, Murinae, Arvicolinae, and Cricetinae but not in *M. Musculus* species. The adjacent sequence 5'-TAAGG-3' was not found in the more distantly related *Anomalurus sp.*, *J. jaculus*, *C. porcellus*, *G. glis*, *T. swinderianus*, or *N. ehrenbergi*. The motif 5'-GCCGG-3' has been found in many vertebrates (Fonseca and Harris, 2008) and has been reported to be critical for human mtDNA replication (Hixson et al., 1986).

The 12 PCGs encoded on the H-strand were used to analyze phylogenetic relationships. Consistent with Alston (1876) and Chevret and Dobigny (2005), we placed *M. unguiculatus* within the Rodentia. Our analysis showed *M. unguiculatus* was in the Gerbillinae clade and had a close phylogenetic relationship with the Murinae.

In our study, the complete mitochondrial genome of *M. unguiculatus* was shown to be composed of conserved vertebrate elements. In total, 37 genes were identified: 13 were PCGs, 2 were rRNA genes (12S rRNA and 16S rRNA), and 22 were tRNAs genes. The major non-coding region, theCR, contained ETAS-1 and CSB-1 sequences. Phylogenetic analysis

based on 12 concatenated PCGs confirmed that *M. unguiculatus* should be classified as genus *Meriones*, sub-family Gerbillinae, family Muridae and had a close phylogenetic relationship with the sub-family Murinae.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the National Science Foundation of China (#31000985, #31172168 and #31572348) and Key Projects in the National Science & Technology Pillar Program (#2015BAI09B01).

REFERENCES

Alston ER (1876). On the classification of the order Glires. *Proc. Zool. Soc. Lond.* 44: 61-98. http://dx.doi.org/10.1111/j.1096-3642.1876.tb02543.x

Arnason U, Adegoke JA, Bodin K, Born EW, et al. (2002). Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc. Natl. Acad. Sci. U. S. A.* 99: 8151-8156. http://dx.doi.org/10.1073/pnas.102164299

Boore JL (1999). Animal mitochondrial genomes. *Nucleic Acids Res.* 27: 1767-1780. http://dx.doi.org/10.1093/nar/27.8.1767

Chang DD and Clayton DA (1986). Identification of primary transcriptional start sites of mouse mitochondrial DNA: accurate in vitro initiation of both heavy- and light-strand transcripts. *Mol. Cell. Biol.* 6: 1446-1453. http://dx.doi.org/10.1128/MCB.6.5.1446

Chao QJ, Li YD, Geng XX, Zhang L, et al. (2014). Complete mitochondrial genome sequence of *Marmota himalayana* (Rodentia: Sciuridae) and phylogenetic analysis within Rodentia. *Genet. Mol. Res.* 13: 2739-2751. http://dx.doi.org/10.4238/2014.April.14.3

Chevret P and Dobigny G (2005). Systematics and evolution of the subfamily Gerbillinae (Mammalia, Rodentia, Muridae). *Mol. Phylogenet. Evol.* 35: 674-688. http://dx.doi.org/10.1016/j.ympev.2005.01.001

Conchedda M, Gabriele F and Bortoletti G (2006). Development and sexual maturation of *Echinococcus granulosus* adult worms in the alternative definitive host, Mongolian gerbil (*Meriones unguiculatus*). *Acta Trop.* 97: 119-125. http://dx.doi.org/10.1016/j.actatropica.2005.08.009

De Rijk P, Wuyts J and De Wachter R (2003). RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics* 19: 299-300. http://dx.doi.org/10.1093/bioinformatics/19.2.299

Du X, Chen Z, Li W, Tan Y, et al. (2010). Development of novel microsatellite DNA markers by cross-amplification and analysis of genetic variation in gerbils. *J. Hered.* 101: 710-716. http://dx.doi.org/10.1093/jhered/esq066

Fonseca MM and Harris DJ (2008). Relationship between mitochondrial gene rearrangements and stability of the origin of light strand replication. *Genet. Mol. Biol.* 30: 566-574. http://dx.doi.org/10.1590/S1415-47572008000300027

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.

Han CX, Li JG, Yang LJ, Zhang HL, et al. (2005). China's agriculture and forestry rodents and scientific management. Xi'an: China. Northwest. Univ. Press, P77

Hegsted DM and Gallagher A (1967). Dietary fat and cholesterol and serum cholesterol in the gerbil. *J. Lipid Res.* 8: 210-214.

Heo DH and Kang TC (2012). The changes of ERG channel expression after administration of antiepileptic drugs in the hippocampus of epilepsy gerbil model. *Neurosci. Lett.* 507: 27-32. http://dx.doi.org/10.1016/j.neulet.2011.11.043

Hixson JE, Wong TW and Clayton DA (1986). Both the conserved stem-loop and divergent 5′-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. *J. Biol. Chem.* 261: 2384-2390.

Jiang X, Gao J, Ni L, Hu J, et al. (2012). The complete mitochondrial genome of *Microtus fortis calamorum* (Arvicolinae, Rodentia) and its phylogenetic analysis. *Gene* 498: 288-295. http://dx.doi.org/10.1016/j.gene.2012.02.022

Kim EB and Lee SG (2016). The complete mitochondrial genome of the Mongolian gerbil, *Meriones unguiculatus* (Rodentia: Muridae: Gerbillinae). *Mitochondrial DNA A DNA MappSeq Anal* 27: 1457-1458. http://dx.doi.org/10.3109/19401736.2014.953091

Lavrov DV, Brown WM and Boore JL (2000). A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. *Proc. Natl. Acad. Sci. USA* 97: 13738-13742. http://dx.doi.org/10.1073/pnas.250402997

Liu W, Wan X and Zhong W (2007). Population dynamics of the Mongolian gerbils: seasonal patterns and interactions among density, reproduction and climate. *J. Arid Environ.* 68: 383-397. http://dx.doi.org/10.1016/j.jaridenv.2006.07.002

Liu XY, Xu DL and Wang DH (2012). High body weight associated with impaired nonshivering thermogenesis but improved glucose tolerance in Mongolian gerbils (*Meriones unguiculatus*). *J. Therm. Biol.* 37: 524-530. http://dx.doi.org/10.1016/j.jtherbio.2012.06.005

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. http://dx.doi.org/10.1093/nar/25.5.0955

Luo G and Liao J (2015a). The complete mitochondrial genome of *Meriones libycus* (Rodentia: Cricetidae) and its phylogenetic analysis. *Mitochondrial DNA* 27: 2545-2546.

Luo G and Liao J (2015b). The complete mitochondrial genome of *Meriones meridianus* (Rodentia: Cricetidae) and its phylogenetic analysis. *Mitochondrial DNA* 27: 2547-2548.

Miller G and Gidley J (1918). Synopsis of supergeneric groups of rodents. *J. Wash. Acad. Sci.* 8: 431-448. http://dx.doi.org/10.5962/bhl.part.6490

Moons CP, Breugelmans S, Cassiman N, Kalmar ID, et al. (2012). The effect of different working definitions on behavioral research involving stereotypies in Mongolian gerbils (*Meriones unguiculatus*). *J. Am. Assoc. Lab. Anim. Sci.* 51: 170-176.

Ohlemiller KK (2009). Mechanisms and genes in human strial presbycusis from animal models. *Brain Res.* 1277: 70-83. http://dx.doi.org/10.1016/j.brainres.2009.02.079

Ojala D, Montoya J and Attardi G (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290: 470-474. http://dx.doi.org/10.1038/290470a0

Pavlinov IY (2001). Current concepts of gerbillid phylogeny and classification. In: African small mammals (Denys C, Granjon L and Poulet A, eds.). Proceedings of the 8th International Symposium on African Small Mammals. IRD, Paris, 141-149.

Perna NT and Kocher TD (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* 41: 353-358. http://dx.doi.org/10.1007/BF01215182

Reuter JS and Mathews DH (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11: 129. http://dx.doi.org/10.1186/1471-2105-11-129

Salvato P, Simonato M, Battisti A and Negrisolo E (2008). The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics* 9: 331. http://dx.doi.org/10.1186/1471-2164-9-331

Sbisà 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. http://dx.doi.org/10.1016/S0378-1119(97)00404-6

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739. http://dx.doi.org/10.1093/molbev/msr121

Vincent AL, Rodrick GE and Sodeman WA, Jr. (1979). The pathology of the Mongolian Gerbil (*Meriones unguiculatus*): a review. *Lab. Anim. Sci.* 29: 645-651.

Wei J, Noto J, Zaika E, Romero-Gallo J, et al. (2012). Pathogenic bacterium *Helicobacter pylori* alters the expression profile of p53 protein isoforms and p53 response to cellular stresses. *Proc. Natl. Acad. Sci. USA* 109: E2543-E2550. http://dx.doi.org/10.1073/pnas.1205664109

Wyman SK, Jansen RK and Boore JL (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252-3255. http://dx.doi.org/10.1093/bioinformatics/bth352

Yao PP, Qian L, Xia Y, Xu F, et al. (2012). Enterovirus 71-induced neurological disorders in young gerbils, *Meriones unguiculatus*: development and application of a neurological disease model. *PLoS One* 7: e51996. http://dx.doi.org/10.1371/journal.pone.0051996

Ying HZ, Liu YH, Yu B, Wang ZY, et al. (2013). Dietary quercetin ameliorates nonalcoholic steatohepatitis induced by a high-fat diet in gerbils. *Food Chem. Toxicol.* 52: 53-60. http://dx.doi.org/10.1016/j.fct.2012.10.030

Yoon KB and Park YC (2015). Complete mitochondrial genome and codon usage of the Nepalese whiskered bat *Myotis muricola* (Vespertilionidae). *Genet. Mol. Res.* 14: 14637-14645. http://dx.doi.org/10.4238/2015.November.18.27

Zhong HM, Zhang HH, Sha WL, Zhang CD, et al. (2010). Complete mitochondrial genome of the red fox (*Vuples vuples*) and phylogenetic analysis with other canid species. *Zool. Res.* 31: 122-130. http://dx.doi.org/10.3724/SP.J.1141.2010.02122

Supplementary material

Table S1. Sequence data used in the phylogenetic analysis.

Figure S1. Codon families in the mitochondrial genomes of *M. unguiculatus* and 23 other rodent species. The numbers to the left refer to the total number of codons. CDspT, codons per thousand codons. Families are provided on the X axis. The amino acid codes are denoted using single letters.