Gene 571 (2015) 279–285

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Analysis of the complete mitochondrial genome and characterization of diverse NUMTs of *Macaca leonina*

Qiu-Xiang Hu [a,b], Yu Fan [a,b], Ling Xu [a,b], Wei Pang [a], Shuang Wang [a], Yong-Tang Zheng [a,b,c], Long-Bao Lv [c], Yong-Gang Yao [a,b,c,*]

[a] Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan 650223, China
[b] Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[c] Kunming Primate Research Center, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

ARTICLE INFO

Article history:
Received 7 March 2015
Received in revised form 8 June 2015
Accepted 26 June 2015
Available online 4 July 2015

Keywords:
mtDNA
NUMT
Pig-tailed monkey
Gene conversion
Tandem repeat
Phylogeny

ABSTRACT

As a non-human primate, the pig-tailed macaque has received wide attention because it can be infected by HIV-1. In this study, we determined the complete mtDNA sequence of the northern pig-tailed macaque (*Macaca leonina*). Unexpectedly, during the amplification of the mtDNA control region (D-loop region) we observed several D-loop-like sequences, which were NUMTs (nuclear mitochondrial sequences) and a total of 14 D-loop-like NUMT haplotypes were later identified in five individuals. The neighbor-joining tree and estimated divergence time based on these D-loop-like NUMT sequences of *M. leonina* provide some insights into the understanding of the evolutionary history of NUMTs. D-loop-like haplotypes G and H, which also exist in the nuclear genome of *mulatta*, appear to have been translocated into the nuclear genome before the divergence of *M. mulatta* and *M. leonina*. The other D-loop-like NUMT haplotypes were translocated into the nuclear genome of *M. leonina* after the divergence of the two species. Later sequence conversion was predicted to occur among these 14 D-loop-like NUMT haplotypes. The overall structure of the mtDNA of *M. leonina* was found to be similar to that seen in other mammalian mitochondrial genomes. Phylogenetic analysis based on the maximum likelihood method shows *M. leonina* clustered with *Macaca silenus* among the analyzed mammalian species.

1. Introduction

Pig-tailed macaques characterized by their short tail, carried half-erect and somewhat resembles a pig's tail, have a broad geographic distribution that covers much of the Southeast Asian mainland and islands. Due to its susceptibility to human immunodeficiency virus type 1 (HIV-1) infection (Kuang et al., 2009; Hatziioannou et al., 2014), the pig-tailed macaque has been proposed for use as an experimental primate in biomedical research (Lei et al., 2013). There are many recent studies that have looked at the basic biology for this species (Lei et al., 2014; Zhang et al., 2014; Zheng et al., 2014) and these studies will facilitate its usage in HIV-1 research.

Mitochondrial DNA (mtDNA) has been used as a good tool in phylogenetic, phylogeographic, and population genetic studies because of its rapid rate of evolution, haploidy, and maternal inheritance (Xu et al., 2012; Simon and Hadrys, 2013; Shi et al., 2014). The mtDNA control region (D-loop region) is the most variable portion of the mammalian mtDNA genome, and is commonly variable at the intra-species level, making it useful for the study of population genetic diversity and the reconstruction of the past demographic history. The phylogenetic relationships of the members of the family Cercopithecidae have been widely debated in the past few years (Hayasaka et al., 1996; Harris and Disotell, 1998; Page et al., 1999; Liedigk et al., 2014) and the complete mtDNA sequence of the northern pig-tail macaque (*M. leonina*) is necessary to obtain a clear picture of its phylogenetic position within the Cercopithecidae.

In this study, we aimed to sequence the entire mtDNA genome of *M. leonina*. We also found several different NUMTs of the mtDNA control region of *M. leonina*, with multiple events possibly leading to their

Abbreviations: mtDNA, mitochondrial DNA; NUMT, nuclear mitochondrial sequences; tRNA, transfer ribonucleic acid; rRNA, ribosomal ribonucleic acid; mya, million years ago; DSB, double-strand break; SAP, shrimp alkaline phosphatase; HIV-1, human immunodeficiency virus type 1; *M. mulatta*, *Macaca mulatta*; *M. leonina*, *Macaca leonina*; *M. sylvanus*, *Macaca sylvanus*; *M. thibetana*, *Macaca thibetana*; *M. fascicularis*, *Macaca fascicularis*; MHC, major histocompatibility complex; ND1, NADH-ubiquinone oxidoreductase chain 1; ND2, NADH-ubiquinone oxidoreductase chain 2; COX1, cytochrome c oxidase subunit 1; COX2, cytochrome c oxidase subunit 2; ATP8, ATP synthase F0 subunit 8; ATP6, ATP synthase F0 subunit 6; COX3, cytochrome c oxidase subunit 1; ND3, NADH-ubiquinone oxidoreductase chain 3; ND4L, NADH-ubiquinone oxidoreductase chain 6 like; ND4, NADH-ubiquinone oxidoreductase chain 4; ND5, NADH-ubiquinone oxidoreductase chain 5; ND6, NADH-ubiquinone oxidoreductase chain 6; CYTB, cytochrome b.

* Corresponding author at: Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223 China.
E-mail address: yaoyg@mail.kiz.ac.cn (Y.-G. Yao).

http://dx.doi.org/10.1016/j.gene.2015.06.085
0378-1119/

formation. Phylogenetic analysis based on mtDNA genome information of *M. leonina* and other mammalian species suggested a close relationship between *M. silenus* and *M. leonina*.

2. Materials and methods

2.1. Blood sample and DNA isolation

Platelets normally do not contain nuclear DNA. In order to eliminate the potential contamination of the D-loop-like sequences in the nuclear genome and to obtain the authentic D-loop sequence, we collected platelets from whole blood for DNA isolation. Blood samples (5 mL) of five *M. leonina* monkeys were collected from the Kunming Primate Research Center, Kunming Institute of Zoology, Chinese Academy of Sciences. More details about these five animals are shown in Table 1. The sampling procedures were approved by the Institutional Animal Care and Use Committee of Kunming Institute of Zoology, Chinese Academy of Sciences. Blood platelets were isolated by centrifugation for 15 min at 110 *g* under room temperature. The platelet-rich serum was then transferred to another tube under sterile conditions, with attention being paid so as not to disturb the red blood cell–serum interface. The collected platelet-rich serum was centrifuged for 15 min at 2250 *g*. The resultant pellet from platelet-rich serum was resuspended in washing buffer containing 4.3 mM K_2HPO_4, 4.3 mM Na_2HPO_4, 113 mM NaCl, 4.4 mM NaH_2PO_4, and 5.5 mM glucose, pH 6.5, and centrifuged at 2000 *g* for 15 min. The pellet (platelet fraction) was then resuspended in 100 μL of washing buffer. Around 10 μL of the platelet suspension was digested in a total volume of 40 μL of reaction buffer (10 mM Tris HC1 pH 8.0, 50 mM KCl, and 1% Triton X-100 containing 100 μg/mL of proteinase K) for 20 min at 56 °C and was deactivated for 8 min at 96 °C. Total genomic DNA of blood samples was extracted from the blood by using the AxyPrep™ Blood Genomic DNA Miniprep Kit (Axygen Biosciences, CA, USA) according to the manufacturer'sinstructions.

2.2. PCR amplification, TA cloning and sequencing

The entire mitochondrial genome of *M. leonina* was sequenced by using DNA isolated from platelets. We designed nine pairs of primers to amplify the entire mtDNA genome of *M. leonina* using the long and accurate polymerase chain reaction (LA-PCR) technique according to the manufacturer's instructions (TaKaRa, China). Primers were designed based on the conserved sequences of four species in Cercopithecinae: *M. mulatta* (NC_005943), *Macaca sylvanus* (NC_002764), *Macaca thibetana* (NC_011519), *Macaca fascicularis* (NC_012670). All the primers used in this study are listed in Table S1. PCR amplification was performed in a volume of 50 mL reaction mixture containing 100 ng of DNA, 10 mmol/L Tris-HCl (pH 8.3), 2.5 mmol/L $MgCl_2$, 50 mmol/L KCl, 200 mmol/L of each dNTP, 10 pmol/L of each primer, and one unit of LA Taq DNA polymerase (TaKaRa Biotech Co., Dalian, China). PCR cycles were composed of an initial denaturation cycle for 5 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 30 s at 55–62 °C, and 1.5–4 min at 72 °C, and ended with a final extension at 72 °C for 10 min. Each fragment overlapped with the next one by about 150 bp. PCR products were purified at 37 °C for 40 min using 1.0 U of shrimp alkaline phosphatase (SAP) and 0.5 U of Exonuclease I (TaKaRa Biotechnology Co. Ltd., Dalian, China), followed by incubation at 96 °C for 10 min, then were sequenced directly in the ABI 3730 DNA sequencer. The primer walking method was applied to sequence the LA-PCR products.

To determine whether there were D-loop-like NUMT sequences in the nuclear genome, another primer pair (X_D: F: 5′-CACCCAAAGCTG GCATTCT-3′; R: 5′-GCAAACCCATCTAGGCATT-3′) was designed to amplify the D-loop from the genomic DNA isolated from whole blood. PCR products were ligated into the pMD18-T vector (TaKaRa, China) and transformed to *DH5a* cells (Tiangen, China). For each fragment, five to ten plasmids were selected for sequencing. Plasmid sequencing was performed with universal primers M13F (−47) (5′-CGCCAGGGTT TTCCCAGTCACGAC-3′) and M13R (−48) (5′-AGCGGATAACAATTTCAC ACAGGA-3′). The mtDNA NUMTs and complete mtDNA of *M. leonina* have been deposited in GenBank under accession numbers KP330233–KP330246 (NUMTs) and KP330231 (complete mtDNA sequence).

2.3. Sequence analysis and gene prediction

DNA sequences were analyzed using the DNASTAR program (DNAS Inc., Madison, WI). The locations of protein-coding and rRNA genes were determined by comparison with available sequences of *M. mulatta*. We identified the transfer RNA (tRNA) genes using program tRNA scan-SE 1.21 (Lowe and Eddy, 1997). One tRNA gene, which was not found by tRNA scan-SE, was identified by comparison with a homolog of *M. mulatta*. The secondary structure of tandem repeats in the mtDNA control region was predicted by using GeneQuest in the DNASTAR program (DNAS Inc., Madison, WI).

2.4. Phylogenetic analysis

We retrieved the complete mtDNA sequences of 28 mammalian species from GenBank (Table S2). Among them, 18 species were used in our previous study (Xu et al., 2012) and we added 10 primate species available in GenBank to infer the phylogenetic position of *M. leonina*. The amino acid sequences of 12 protein-coding genes on the mitochondrial H-strand were concatenated for phylogenetic analyses. We excluded the *MT-ND6* gene because it was encoded by the L-strand and its properties deviated from the other 12 genes (Waddell et al., 1999). A maximum likelihood tree was constructed by using MEGA6 (Tamura et al., 2013) with the default settings.

The mtDNA control region sequences of 6 different primate species (*M. mulatta*, *M. sylvanus*, *M. thibetana*, *M. fascicularis*, *Homo sapiens*, *Pan troglodyte*) from GenBank were used for phylogenetic analysis. A neighbor-joining tree was constructed using MEGA6 (Tamura et al., 2013) with the default settings. The divergence time was also estimated using MEGA6 (Tamura et al., 2013) with the divergence time of 2.42–2.48 million years ago (mya) between *M. leonina* and *M. mulatta* (Hayasaka et al., 1996) as the reference.

We aligned the *M. mulatta* mtDNA control region sequence with the *M. mulatta* genome by BLAT at the UCSC server (http://genome.ucsc.edu/) and identified many potential NUMTs. We filtered these sequences with the criteria that more than 80% of the aligned sequence showed an identity below 90%. Finally, we identified 6 high-quality NUMTs (including one NUMT that highly resembles haplotype G in *M. leonina*) and all of them span the region between the start of tRNA-

Table 1
D-loop like NUMT haplotype distribution in five *M. leonina* animals.

Animal number	Blood sample	Gender	Weight (kg)	Age (year)	Haplotype [a]
#2382	#294	Female	5.6	7–9	A, B, C, D, E, G, H, I, N
#445	#295	Male	3	3.5–4.5	B, D, F, G, H, J, K, L
#117	#299	Male	3	3.5–4.5	B, C, D, E, G, I, J, K, M
#872	#300	Male	2.5	3.5–4.5	B, C, E, G, H, I, L, M, N
#1527	#301	Male	2.5	3–3.5	B, C, D, E, F, G, H, I, J

[a] A–N refers to the 14 D-loop-like NUMT haplotypes, which were identified in five *M. leonina* animals.

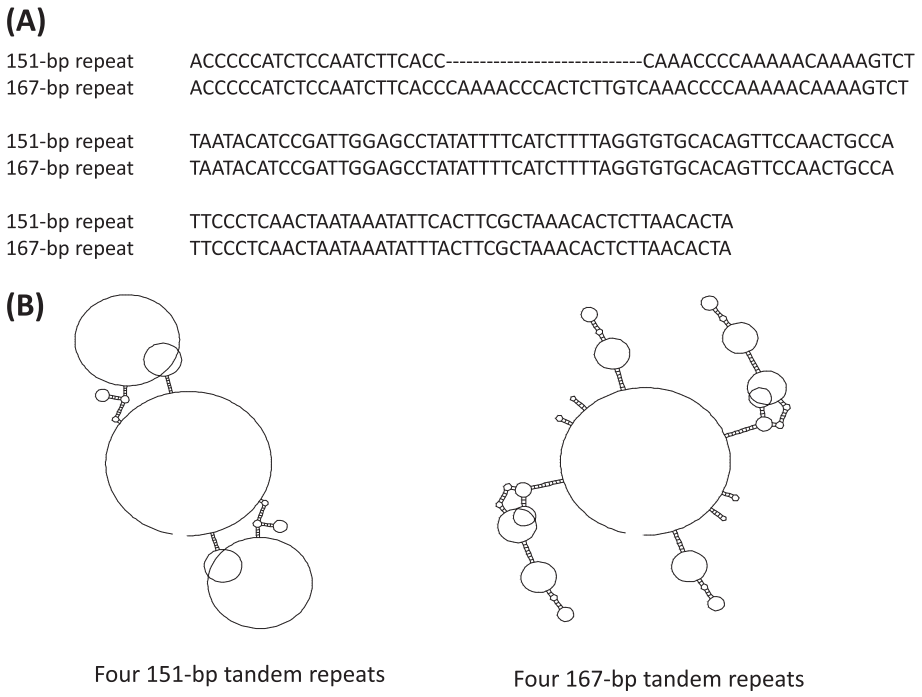

Fig. 1. Structure of the mtDNA control region tandem repeats of *M. leonina*. There are two repeat units with different lengths (151 bp and 167 bp) (A). The secondary structure of the tandem repeats (B) in the mtDNA control region of *M. leonina* was predicted by using GeneQuest in the DNASTAR program.

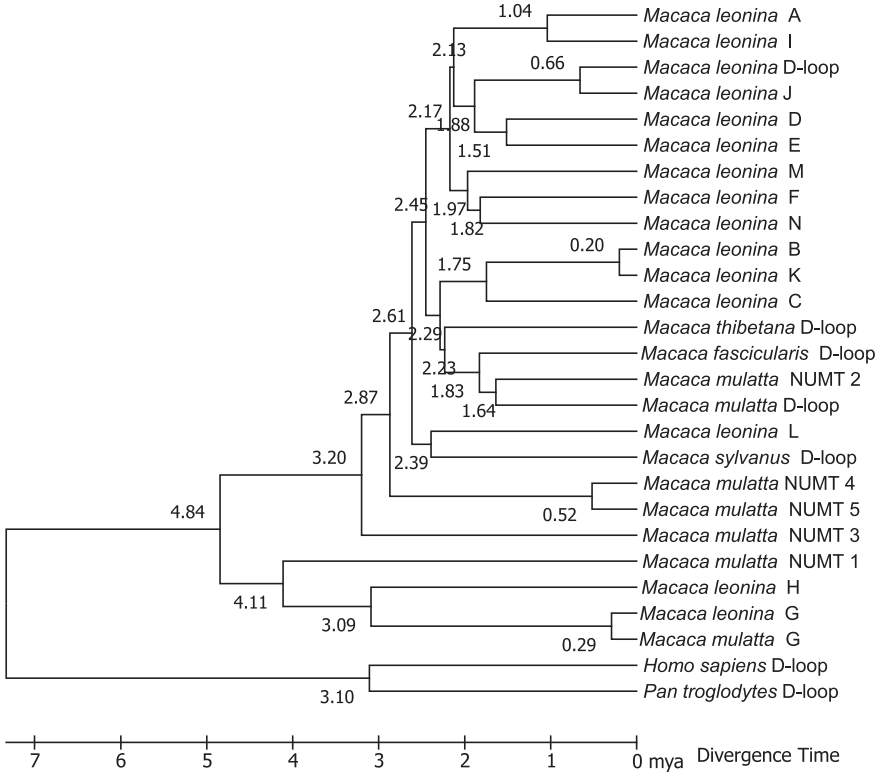

Fig. 2. Neighbor-joining tree of the NUMT and authentic mtDNA control region (marked by a suffix "D-loop" after the species name) sequences of *M. leonina* and related species. The authentic mtDNA control region sequence of *M. thibetana* (GenBank accession number NC_011519), *M. fascicularis* (NC_012670), *M. mulatta* (NC_005943), *M. sylvanus* (NC_002764), *Homo sapiens* (NC_012920), *Pan troglodytes* (NC_001643) were retrieved from GenBank. The D-loop like NUMT haplotypes of *M. leonina* are named from *Macaca leonina* A to *Macaca leonina* N. The NUMTs of *M. mulatta* are named from *Macaca mulatta* NUMT 1 to *Macaca mulatta* NUMT 5. The NUMT of *M. mulatta* that resembles haplotype G of *M. leonina* was marked as "*Macaca mulatta* G". The numbers on the branches refer to divergence time (million years ago, mya).

Pro and the end of tRNA-Phe by manual check. The detailed information of these potential NUMTs of *M. mulatta* was shown in Table S3. Program GENECONV v.1.81 was used to detect sequence conversion events between 14 D-loop-like haplotypes (Sawyer, 1989).

3. Results

3.1. Characteristics of M. leonina D-loop sequence

During our sequencing of the authentic mtDNA control region of *M. leonina* using genomic DNA from whole blood, we observed various D-loop-like sequences in one individual, which suggested that NUMTs of the mtDNA D-loop region existed in the nuclear genome of *M. leonina*. These D-loop-like NUMT sequences were found in five *M. leonina* individuals and we identified a total of 14 D-loop-like NUMT haplotypes (Fig. S1) during our cloning assays. The distribution of these 14 D-loop-like NUMT haplotypes in each animal is shown in Table 1.

Amplification of the mtDNA control region sequence using the mtDNA isolated from platelet showed fragment length polymorphisms. Within the mtDNA control region, we observed 2 types of tandem repeats, composed of a 167-bp or 151-bp basic repeat unit (the 151-bp repeat unit had a 16-base deletion compared with the 167-bp repeat unit) (Fig. 1A) and the number of repeat units ranged from 1 to 4 in different individuals. The symmetrical secondary structure of the four tandem repeats of *M. leonina* mtDNA is shown in Fig. 1B.

Neighbor-joining tree for the 14 D-loop-like NUMT haplotypes (A–N) of *M. leonina*, together with the mtDNA D-loop sequences of other primates and the D-loop-derived NUMTs of *M. mulatta*, showed that all D-loop-like NUMT haplotypes of *M. leonina* were clustered together (Fig. 2), except for haplotypes G and H, suggesting that the translocation of haplotypes G and H into the nuclear genome might be much more ancient than that of the other D-loop-like haplotypes. Intriguingly, a D-loop NUMT of *M. mulatta* resembled haplotype G in *M. leonina*, which would suggest a persistence of the same ancestral translocation after the divergence of the two species. The estimated time for the divergence of the D-loop-like NUMT sequences showed that haplotypes G and H occurred around 3.09 mya, long before the divergence time between *M. leonina* and *M. mulatta* (2.42–2.48 mya (Hayasaka et al., 1996)) (Fig. 2). As shown in Table 2, gene conversion prediction results indicated that the 14 D-loop-like NUMT haplotypes had undergone later sequence conversions.

Table 2
Evidence of gene conversion in 14 D-loop-like haplotypes in the nuclear genome of *M. leonina*.

Seq name[a]	Sim *P* value[b]	BC KA *P* value[c]	Aligned offsets			Num Poly[d]	Num Dif[e]	Tot Difs[f]	MisM Pen[g]
			Begin	End	Len				
J;L	0.0000	0.0000	636	1045	410	184	0	78	None
A;D	0.0000	0.0000	763	1045	283	137	0	61	None
J;M	0.0000	0.0006	520	727	208	51	0	127	None
E;I	0.0032	0.0096	510	735	226	57	0	97	None
D;I	0.0097	0.0257	856	1001	146	63	0	82	None
I;J	0.0099	0.0278	508	675	168	46	0	110	None
G;L	0.0110	0.0323	373	449	77	18	0	238	None
C;J	0.0154	0.0443	901	977	77	34	0	140	None
B;J	0.0174	0.0483	627	740	114	37	0	129	None
E;M	0.0191	0.0538	520	675	156	39	0	122	None

[a] Seq Name — letters A–N refer to the14 D-loop-like haplotypes in the nuclear genome of *M. leonina*.
[b] Sim*P* value means simulated *P*-values based on 10,000 permutations. Fragments are listed only if Sim*P*-value < = 0.05.
[c] BC KA means Bonferroni-corrected KA (BLAST-like) *P* values. *P* value ≤ 0.05 was regarded as significant for gene conversion.
[d] Num Poly is the number of polymorphic sites in the fragment.
[e] Num Dif is the number of mismatches within the fragment.
[f] Tot Difs is the total number of mismatches between two sequences.
[g] MisM Pen is the penalty per mismatch for the two sequences.

3.2. Structure of the mitochondrial genome and phylogenetic tree

The total length of the complete mtDNA of *M. leonina*, which was determined by using DNA isolated from platelets, ranged from 16,533 bp to 17,050 bp, which was caused by the heteroplasmy of different numbers of tandem repeats in the mtDNA control region. Consistent with other primate species, the complete mtDNA of *M. leonina* is shown to have 13 protein-coding genes, 2 rRNA genes (12S rRNA and 16S rRNA), 22 tRNA genes, and the control region. The overall base composition of the H strand was 32.10% A, 30.40% C, 12.40% G, and 25.10% T. The arrangement of the genes in *M. leonina* mtDNA genome is shown in Table 3.

To obtain a more complete understanding of the evolutionary history of *M. leonina*, a maximum likelihood tree was constructed to infer the phylogenetic position of *M. leonina* based on the concatenated amino acid sequences of 12 protein-coding genes on the mitochondrial H-strand. *M. silenus* showed the closest relationship to *M. leonina* (Fig. 3), both belonging to the genus macaque.

4. Discussion

NUMTs have been found in a wide variety of eukaryotic organisms (Bensasson et al., 2001; Yao et al., 2008; Hazkani-Covo et al., 2010; Tsuji et al., 2012). The suggested mechanism of NUMT formation is

Table 3
Structure of *M. leonina* mitochondrial genome sequence.

Gene	Position	Size(bp)	Strand	Codon	
				Start	Stop
D-loop	16500–17050	1592	H		
	1–1041[a]		H		
tRNA-Phe	1042–1113	72	H		
12s-rRNA	1114–2060	947	H		
tRNA-Val	2061–2129	69	H		
16s-rRNA	2130–3689	1560	H		
tRNA-Leu	3690–3764	75	H		
ND1	3767–4723	957	H	ATG	TAG
tRNA-Ile	4722–4790	69	H		
tRNA-Gln	4788–4859	72	L		
tRNA-Met	4861–4928	68	H		
ND2	4929–5972	1044	H	ATT	TAG
tRNA-Sec	5971–6037	67	H		
tRNA-Ala	6045–6113	69	L		
tRNA-Asn	6115–6187	73	L		
rep_origin	6189–6220	32	H		
tRNA-Cys	6219–6286	68	L		
tRNA-Tyr	6286–6351	66	L		
COX1	6356–7894	1539	H	ATG	CTA
tRNA-Ser	7897–7965	69	L		
tRNA-Asp	7969–8036	68	H		
COX2	8038–8721	684	H	ATG	TAG
tRNA-Lys	8775–8839	65	H		
ATP8	8841–9047	207	H	ATG	TAA
ATP6	9002–9682	681	H	ATG	TAA
COX3	9682–10465	784	H	ATG	T
tRNA-Gly	10466–10533	68	H		
ND3	10534–10879	346	H	ATT	T
tRNA-Arg	10880–10944	65	H		
ND4L	10945–11241	297	H	ATG	TAA
ND4	11235–12612	1378	H	ATG	T
tRNA-His	12613–12681	69	H		
tRNA-Ser	12682–12740	59	H		
tRNA-Leu	12741–12811	71	H		
ND5	12812–14623	1812	H	GTA	TAA
ND6	14624–15151	528	L	TCT	CAT
tRNA-Glu	15152–15220	69	L		
CYTB	15225–16365	1141	H	ATG	T
tRNA-Thr	16366–16430	65	H		
tRNA-Pro	16432–16499	68	L		

[a] There are length polymorphisms in this region. Different number (range from 1 to 4) of the 167 bp or 151 bp tandem repeats was observed in the mtDNA control region of *M. leonina*. The 151 bp repeat has a 16-bp deletion relative to the 167 bp repeat.

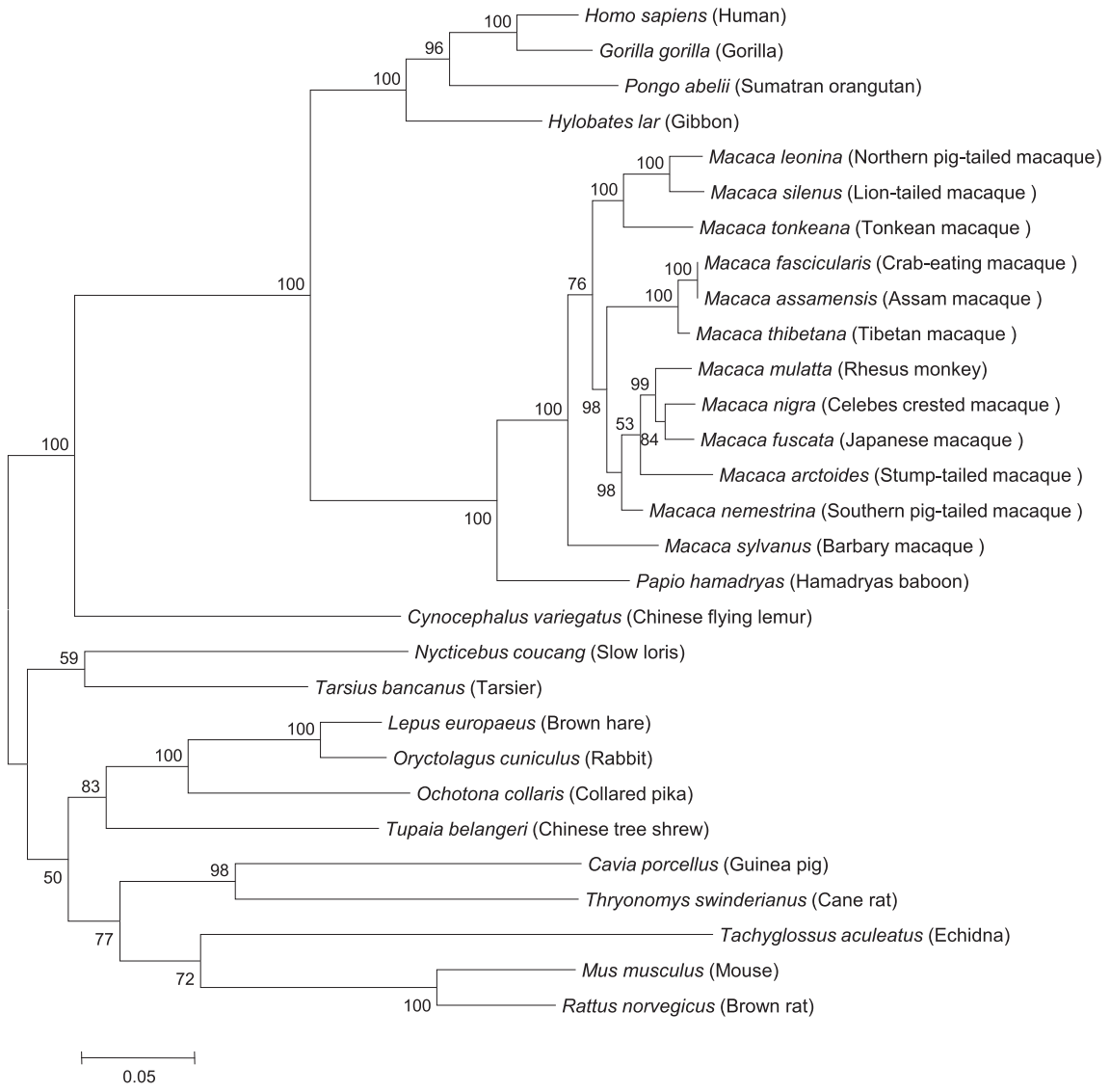

Fig. 3. Maximum likelihood tree of 29 species based on the concatenated amino acid sequences of 12 protein-coding genes on the mitochondrial H-strand. GenBank accession numbers for these species are listed in Table S2. The numbers on the branches refer to bootstrap value.

predominantly associated with DNA-mediated transfer and non-homologous end-joining during nuclear double-strand break (DSB) repair (Blanchard and Schmidt, 1996; Woischnik and Moraes, 2002; Hazkani-Covo and Covo, 2008). After their insertion into the nucleus, NUMTs are not subjected to mitochondrial selective pressure (Perna and Kocher, 1996; Bensasson et al., 2001) and the mutation rate of NUMTs is on average around one order of magnitude slower than the mitochondrial genome (Brown et al., 1982; Lopez et al., 1997). For this reason, NUMTs are commonly believed to be "fossilized" copies of ancient mitochondrial lineages as they are more similar to the ancestral mitochondrial haplotype than the modern mitochondrial counterpart (Perna and Kocher, 1996; Bensasson et al., 2001; Jensen-Seaman et al., 2009). As a result, NUMTs offer good opportunities for the study of mtDNA evolution, and NUMTs can be used as markers for inferring phylogenies (Hazkani-Covo, 2009). A previous study showed that nearly all positions of the mitochondrial genome could be found in the nuclear DNA, but no D-loop-derived NUMTs (Mourier et al., 2001). Within great apes, the frequency of NUMTs has been shown to be increased in gorillas, but these observations were limited to the mtDNA D-loop region (Jensen-Seaman et al., 2004; Thalmann et al., 2004; Anthony et al., 2007). The lack of NUMTs from the D-loop region was probably caused by the significantly higher evolutionary rate of extant mtDNA in this region (Woischnik and Moraes, 2002).

In this study, we determined the entire complete mtDNA sequence of *M. leonina* and we unexpectedly identified various D-loop-like NUMT sequences in this species. Intriguingly, we also found that *M. mulatta* have some D-loop-derived NUMTs (Fig. 2). The reason for the abundance of D-loop-derived NUMTs in *M. leonina* and *M. mulatta* remains unknown. We showed that gene conversions occurred between these similar D-loop-like NUMT haplotypes in *M. leonina*, which would contribute to an increase of the variability (Table 2). Previous studies showed that gene conversion played significant roles in shaping the standing haplotype variations in the *MHC* and *SI* genes, which serve as classic example genes of diversifying selection (Wright, 1939; Takahata, 1990). In addition, there is a report that *Macaca nemestrina* presented the strongest evidence of recombination between subspecies or species based on mtDNA data (Piganeau et al., 2004). Recombination between subspecies or species may thus lead to the diverse D-loop-like NUMT haplotypes in the nuclear genome of *M. leonina*.

Intriguingly, there are tandem repeats in the mtDNA control region of *M. leonina*, which are composed of 167-bp or 151-bp repeat unit. The number of tandem repeats ranged from 1 to 4 among the studied

individuals. Tandemly repeated sequences within the mtDNA control region have been reported in a variety of vertebrate and invertebrate species, with the length of repeat unit from smaller than 10 bp to over 200 bp, and the location of the repeats within the control region was also highly variable (Hoelzel et al., 1993; Lunt et al., 1998; Mundy and Helbig, 2004). The origin of mtDNA tandem repeats was generally thought to be caused by slipped-strand mispairing during mtDNA replication, also known as replication slippage (Buroker et al., 1990; Fumagalli et al., 1996; Broughton and Dowling, 1997; Wilkinson et al., 1997; Hazkani-Covo and Covo, 2008), which was a dominant mechanism responsible for length variation in mammalian mtDNA. Several other mechanisms, including recombination and transposition (Rand and Harrison, 1989) and unequal crossing-over or gene conversion (Hoelzel et al., 1993), might also cause tandem repeat variations. The tandem repeats would fold into stable secondary structures and probably resulted in the insertion or deletion of repeated units during mtDNA replication (Buroker et al., 1990; Hoelzel et al., 1993; Stewart and Baker, 1994). Whether there is a potential relationship between the tandem repeats in the mtDNA control region of *M. leonina* and the occurrence of D-loop-derived NUMTs in this species needs further study.

The structure of the entire mtDNA genome of *M. leonina* was similar to that seen in other mammalian mitochondrial genomes (Anderson et al., 1982; Li et al., 2009; Xu et al., 2012). In the maximum likelihood tree (Fig. 3), the clustering pattern of *M. leonina* and related species was consistent with the known species tree, which suggested the suitability of using mtDNA genome data to reconstruct the phylogenetic relationship among these species.

In summary, we identified various D-loop-derived NUMTs and tandem repeat variations in the mtDNA D-loop region of *M. leonina*. Sequence conversion may explain the occurrence of these D-loop-like NUMT haplotypes. Caution should be taken when we use the mtDNA control region sequence as a genetic marker for estimating the genetic diversity of *M. leonina*.

Role of funding source

This study was supported by the 135 strategic project of Kunming Institute of Zoology, Chinese Academy of Sciences. Y.-T. Z. was supported by the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-EW-R-13).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The authors are grateful to all the participants in this study. We thank Dr. Dong-Dong Wu for helpful comments regarding data analysis, Dr. Wen Zhang and Dr. Rui Bi for experimental assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2015.06.085.

References

Anderson, S., Debruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F., Young, I.G., 1982. Complete sequence of bovine mitochondrial-DNA — conserved features of the mammalian mitochondrial genome. J. Mol. Biol. 156, 683–717.

Anthony, N.M., Clifford, S.L., Bawe-Johnson, M., Abernethy, K.A., Bruford, M.W., Wickings, E.J., 2007. Distinguishing gorilla mitochondrial sequences from nuclear integrations and PCR recombinants: guidelines for their diagnosis in complex sequence databases. Mol. Phylogenet. Evol. 43, 553–566.

Bensasson, D., Zhang, D., Hartl, D.L., Hewitt, G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16, 314–321.

Blanchard, J.L., Schmidt, G.W., 1996. Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. Mol. Biol. Evol. 13, 537–548.

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

Brown, W.M., Prager, E.M., Wang, A., Wilson, A.C., 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. 18, 225–239.

Buroker, N.E., Brown, J.R., Gilbert, T.A., Ohara, P.J., Beckenbach, A.T., Thomas, W.K., Smith, M.J., 1990. Length heteroplasmy of sturgeon mitochondrial-DNA — an illegitimate elongation model. Genetics 124, 157–163.

Fumagalli, L., Taberlet, P., Favre, L., Hausser, J., 1996. Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. Mol. Biol. Evol. 13, 31–46.

Harris, E.E., Disotell, T.R., 1998. Nuclear gene trees and the phylogenetic relationships of the mangabeys (Primates: Papionini). Mol. Biol. Evol. 15, 892–900.

Hatziioannou, T., Del Prete, G.Q., Keele, B.F., Estes, J.D., McNatt, M.W., Bitzegeio, J., Raymond, A., Rodriguez, A., Schmidt, F., Mac Trubey, C., Smedley, J., Piatak Jr., M., KewalRamani, V.N., Lifson, J.D., Bieniasz, P.D., 2014. HIV-1-induced AIDS in monkeys. Science 344, 1401–1405.

Hayasaka, K., Fujii, K., Horai, S., 1996. Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. Mol. Biol. Evol. 13, 1044–1053.

Hazkani-Covo, E., 2009. Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. Mol. Biol. Evol. 26, 2175–2179.

Hazkani-Covo, E., Covo, S., 2008. Numt-mediated double-strand break repair mitigates deletions during primate genome evolution. PLoS Genet. 4, e1000237.

Hazkani-Covo, E., Zeller, R.M., Martin, W., 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 6, e1000834.

Hoelzel, A.R., Hancock, J.M., Dover, G.A., 1993. Generation of VNTRs and heteroplasmy by sequence turnover in the mitochondrial control region of two elephant seal species. J. Mol. Evol. 37, 190–197.

Jensen-Seaman, M.I., Sarmiento, E.E., Deinard, A.S., Kidd, K.K., 2004. Nuclear integrations of mitochondrial DNA in gorillas. Am. J. Primatol. 63, 139–147.

Jensen-Seaman, M.I., Wildschutte, J.H., Soto-Calderon, I.D., Anthony, N.M., 2009. A comparative approach shows differences in patterns of numt insertion during hominoid evolution. J. Mol. Evol. 68, 688–699.

Kuang, Y.Q., Tang, X., Liu, F.L., Jiang, X.L., Zhang, Y.P., Gao, G.X., Zheng, Y.T., 2009. Genotyping of TRIM5 locus in northern pig-tailed macaques (*Macaca leonina*), a primate species susceptible to Human Immunodeficiency Virus type 1 infection. Retrovirology 6, 58.

Lei, A.H., Pang, W., Zhang, G.H., Zheng, Y.T., 2013. Use and research of pigtailed macaques in nonhuman primate HIV/AIDS models. Zool. Res. 34, 77–88.

Lei, A.H., Zhang, G.H., Tian, R.R., Zhu, J.W., Zheng, H.Y., Pang, W., Zheng, Y.T., 2014. Replication potentials of HIV-1/HSIV in PBMCs from northern pig-tailed macaque (*Macaca leonina*). Zool. Res. 35, 186–195.

Li, D.M., Fan, L.Q., Zeng, B., Yin, H.L., Zou, F.D., Wang, H.X., Meng, Y., King, E., Yue, B.S., 2009. The complete mitochondrial genome of *Macaca thibetana* and a novel nuclear mitochondrial pseudogene. Gene 429, 31–36.

Liedigk, R., Roos, C., Brameier, M., Zinner, D., 2014. Mitogenomics of the Old World monkey tribe Papionini. BMC Evol. Biol. 14, 176.

Lopez, J.V., Culver, M., Stephens, J.C., Johnson, W.E., O'Brien, S.J., 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol. Biol. Evol. 14, 277–286.

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

Lunt, D.H., Whipple, L.E., Hyman, B.C., 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. Mol. Ecol. 7, 1441–1455.

Mourier, T., Hansen, A.J., Willerslev, E., Arctander, P., 2001. The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Mol. Biol. Evol. 18, 1833–1837.

Mundy, N.I., Helbig, A.J., 2004. Origin and evolution of tandem repeats in the mitochondrial DNA control region of shrikes (*Lanius* spp.). J. Mol. Evol. 59, 250–257.

Page, S.L., Chiu, C.H., Goodman, M., 1999. Molecular phylogeny of old world monkeys (Cercopithecidae) as inferred from gamma-globin DNA sequences. Mol. Phylogenet. Evol. 13, 348–359.

Perna, N.T., Kocher, T.D., 1996. Mitochondrial DNA: molecular fossils in the nucleus. Curr. Biol. 6, 128–129.

Piganeau, G., Gardner, M., Eyre-Walker, A., 2004. A broad survey of recombination in animal mitochondria. Mol. Biol. Evol. 21, 2319–2325.

Rand, D.M., Harrison, R.G., 1989. Molecular population genetics of mtDNA size variation in crickets. Genetics 121, 551–569.

Sawyer, S., 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. 6, 526–538.

Shi, N.N., Fan, L., Yao, Y.G., Peng, M.S., Zhang, Y.P., 2014. Mitochondrial genomes of domestic animals need scrutiny. Mol. Ecol. 23, 5393–5397.

Simon, S., Hadrys, H., 2013. A comparative analysis of complete mitochondrial genomes among Hexapoda. Mol. Phylogenet. Evol. 69, 393–403.

Stewart, D.T., Baker, A.J., 1994. Patterns of sequence variation in the mitochondrial D-loop region of shrews. Mol. Biol. Evol. 11, 9–21.

Takahata, N., 1990. A simple genealogical structure of strongly balanced allelic lines and transspecies evolution of polymorphism. Proc. Natl. Acad. Sci. U.S.A. 87, 2419–2423.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Thalmann, O., Hebler, J., Poinar, H.N., Paabo, S., Vigilant, L., 2004. Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol. Ecol. 13, 321–335.

Tsuji, J., Frith, M.C., Tomii, K., Horton, P., 2012. Mammalian NUMT insertion is non-random. Nucleic Acids Res. 40, 9073–9088.
Waddell, P.J., Cao, Y., Hauf, J., Hasegawa, M., 1999. Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid invariant sites LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. Syst. Biol. 48, 31–53.
Wilkinson, G.S., Mayer, F., Kerth, G., Petri, B., 1997. Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. Genetics 146, 1035–1048.
Woischnik, M., Moraes, C.T., 2002. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res. 12, 885–893.
Wright, S., 1939. The distribution of self-sterility alleles in populations. Genetics 24, 538–552.
Xu, L., Chen, S.Y., Nie, W.H., Jiang, X.L., Yao, Y.G., 2012. Evaluating the phylogenetic position of Chinese tree shrew (*Tupaia belangeri chinensis*) based on complete mitochondrial genome: implication for using tree shrew as an alternative experimental animal to primates in biomedical research. J. Genet. Genomics 39, 131–137.
Yao, Y.G., Kong, Q.P., Salas, A., Bandelt, H.J., 2008. Pseudomitochondrial genome haunts disease studies. J. Med. Genet. 45, 769–772.
Zhang, X.L., Pang, W., Deng, D.Y., Lv, L.B., Feng, Y., Zheng, Y.T., 2014. Analysis of immunoglobulin, complements and CRP levels in serum of captive northern pig-tailed macaques (*Macaca leonina*). Zool. Res. 35, 196–203.
Zheng, H.Y., Zhang, M.X., Zhang, L.T., Zhang, X.L., Pang, W., Lv, L.B., Zheng, Y.T., 2014. Flow cytometric characterizations of leukocyte subpopulations in the peripheral blood of northern pig-tailed macaques (*Macaca leonina*). Zool. Res. 35, 465–473.

Table S1. Primers used for sequencing the complete mtDNA of *Macaca leonina*

Primer name	**Primer sequence(5'-3')**
For PCR amplification	
MT-X1	F: AGGTGCGCTTGGATAAATCA
	R: GCTCCGATTAGGGCGTAGT
MT-X2	F: ATTGCCCTCCTCCTATGAAC
	R: ACTAAGGGCTTTGAAGGCTC
MT-X3	F: ACGCCTAATCTACACCACCTC
	R: CGGGGAGAATAAGGATGTAGACT
MT-X4	F: TTAGGAGCCATCAACTTCATTACC
	R: ATTGCGGGTAAGATAGTTCAGA
MT-X5-2	F: CCCCTAAAACTGGTTTCAAGC
	R: GGATGCGAGTAGTACAGAGGTG
MT-X6	F: AACAACACCTAATGACCCACCA
	R: GGGTAGAGAGCCGGTTAGTGT
MT-X7	F: TAAATGGGGCAACCAAGCA
	R: GCTTGAGGTGGAGAAGGCTAC
MT-X8	F: TGGGGGCTATTACTACCCTATT
	R: ATTTGTAGGGTTAGGCAGGCT
MT-F3	F: CCCTCAAAAATCACCCAATCA
MT-R4	R: GAGCTTGAACGCTTTCTTAATTG
XD	F: CACCCAAAGCTGGCATTCT
	R: GCAAACCCATCTAGGCATT
For sequencing	
L14	TCACCCTATTAACCAGTCACG
L122	AATGCGCCTGCCTTTGAT
L423	AACTGCCATTCCCTCAACTAA
H1140	GGTGTAAAGCACCGTCAAGT
L978	AGTGGCTTTAAAGCTTCTGAAC
L2196	ACTGAACTCCTCACATCACATTG
L4181	CCATCGCAATCTCCAGCAT
L5379	CAAAACCAACACCCCTTATCC
L7839	ATCAAATCAATCGGACACCA
H8407	CGTGGGTGTACCTTGTGGTA
L9788	CTTCATTACCACAGGCTTCCAT
L9938	CTAACCAAATGACCCCTACCTA
L10952	CTCACCATCTGACTCCTACCTCTC
H11251	GCTGTGGGTGCGTTCATA
L11702	GCCCATGGACTCACCTCTT
L12275	CCATCCGTTGACCTTAGGA
L12698	TCCTCACCACCATACTAATTCTAG
L12719	CTTCTCCACCTCAAGCCAAC
H13014	GGTTTTAGTGATACTGGTGATTGG
L14915	ACCTCCTCCGCCTTCTCC
L15555	GATACTTCCTATTTGCATACACAA
H16312	AAAGCGTTTTGAGCTGCGT

Note: We used both the amplification primers and the inner sequencing primers to obtain the entire mtDNA sequence. The “L” and “H” represent the light (L) and heavy (H) strands of the mtDNA, respectively.

Table S2. Mammalian species (re-)analyzed in this study

Species	Common name	Source	GenBank accession number
Eutheria			
Euarchontoglires			
Euarchonta			
Primates			
Nycticebus coucang	Slow loris	(Arnason et al., 2000)	AJ309867
Tarsius bancanus	Tarsier	(Schmitz et al., 2002)	AF348159
Papio hamadryas	Hamadryas baboon	(Arnason et al., 1998)	Y18001
Macaca leonina	Nothern pig-tailed monkey	This study	KP330231
Macaca mulatta	Rhesus monkey	(Gokey et al., 2004)	NC_005943
Macaca arctoides	Stump-tailed macaque	(Liedigk et al., 2014)	NC_025201
Macaca assamensis	Assam macaque	(Jiang et al., 2014)	NC_023795
Macaca fascicularis	Crab-eating macaque		NC_012670
Macaca fuscata	Japanese macaque	(Wang et al., 2014)	NC_025513
Macaca nemestrina	Southern pig-tailed macaque		KP765688
Macaca nigra	Celebes crested macaque	(Du et al., 2014)	NC_026120
Macaca silenus	Lion-tailed macaque	(Liedigk et al., 2014)	NC_025221
Macaca sylvanus	Barbary macaque	(Arnason et al., 2000)	NC_002764
Macaca thibetana	Tibetan macaque	(Li et al., 2009)	NC_011519
Macaca tonkeana	Tonkean macaque	(Liedigk et al., 2014)	NC_025222
Hylobates lar	Gibbon	(Arnason et al., 1996a)	X99256
Pongo abelii	Sumatran orangutan	(Xu and Arnason, 1996b)	X97707
Gorilla gorilla	Gorilla	(Xu and Arnason, 1996a)	X93347
Homo sapiens	Human	(Arnason et al., 1996b)	X93334
Cynocephalus variegatus	Flying lemur	(Xu et al., 2012)	JN800721
Tupaia belangeri	Chinese tree shrew	(Xu et al., 2012)	JN800722
Glires			
Lagomorpha			
Lepus europaeus	Brown hare	(Arnason et al., 2002)	AJ421471
Oryctolagus cuniculus	Rabbit	(Gissi et al., 1998)	AJ001588
Ochotona collaris	Collared pika	(Lin et al., 2002)	AF348080
Rodentia			
Mus musculus	Mouse	(Bibb et al., 1981)	J01420
Rattus norvegicus	Brown rat	.	AJ428514
Cavia porcellus	Guinea pig	(D'Erchia et al., 1996)	AJ222767
Thryonomys swinderianus	Cane rat	(Mouchaty et al., 2001)	AJ301644
Monotremata			
Tachyglossus aculeatus	Echidna	(Janke et al., 2002)	AJ303116

Table S3. Detailed information of D-loop-derived NUMTs of *Macaca mulatta*

ID	Align %	Identity %	Chr	Strand	Start	End
Macaca mulatta NUMT 1	93.55	91.00	6	+	29662070	29663124
Macaca mulatta NUMT 2	92.72	92.10	5	-	49967443	49968472
Macaca mulatta NUMT 3	89.12	91.50	6	+	34061232	34062206
Macaca mulatta NUMT 4	88.94	90.20	19	+	5865649	5866403
Macaca mulatta NUMT 5	87.83	92.20	7	-	14023875	14024809
Macaca mulatta G	85.53	91.80	4	+	82446882	82447815

We retrieved potential D-loop NUMTs from *M. mulatta* genome by using BLAT with *M. mulatta* D-loop sequence as the search entry. The numbering of start point and end point were scored relative to *M. mulatta* genome.

Align % - the percentage of the aligned sequence; Identity % - the percentage of sequence identity in the aligned sequence; Chr–chromosome; forward and reverse strand was marked as "+" and "-", respectively.

Figure S1. Sequence alignment of the 14 D-loop like NUMT haplotypes (A-N) identified in *Macaca leonina*

Supplementary References

Arnason, U., Adegoke, J.A., Bodin, K., Born, E.W., Esa, Y.B., Gullberg, A., Nilsson, M., Short, R.V., Xu, X. and Janke, A., 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc Natl Acad Sci U S A 99, 8151-6.

Arnason, U., Gullberg, A., Burguete, A.S. and Janke, A., 2000. Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. Hereditas 133, 217-28.

Arnason, U., Gullberg, A. and Janke, A., 1998. Molecular timing of primate divergences as estimated by two nonprimate calibration points. J Mol Evol 47, 718-27.

Arnason, U., Gullberg, A. and Xu, X., 1996a. A complete mitochondrial DNA molecule of the white-handed gibbon, *Hylobates lar*, and comparison among individual mitochondrial genes of all hominoid genera. Hereditas 124, 185-189.

Arnason, U., Xu, X. and Gullberg, A., 1996b. Comparison between the complete mitochondrial DNA sequences of Homo and the common chimpanzee based on nonchimeric sequences. J Mol Evol 42, 145-52.

Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A., 1981. Sequence and gene organization of mouse mitochondrial DNA. Cell 26, 167-80.

D'Erchia, A.M., Gissi, C., Pesole, G., Saccone, C. and Arnason, U., 1996. The guinea-pig is not a rodent. Nature 381, 597-600.

Du, L.N., Shi, F.L., Liu, Z.J. and Zhou, Q.H., 2014. Complete mitochondrial genome of the crested black macaque (Macaca nigra). Mitochondrial DNA, 1-2.

Gissi, C., Gullberg, A. and Arnason, U., 1998. The complete mitochondrial DNA sequence of the rabbit, *Oryctolagus cuniculus*. Genomics 50, 161-9.

Gokey, N.G., Cao, Z., Pak, J.W., Lee, D., McKiernan, S.H., McKenzie, D., Weindruch, R. and Aiken, J.M., 2004. Molecular analyses of mtDNA deletion mutations in microdissected skeletal muscle fibers from aged rhesus monkeys. Aging Cell 3, 319-26.

Janke, A., Magnell, O., Wieczorek, G., Westerman, M. and Arnason, U., 2002. Phylogenetic analysis of 18S rRNA and the mitochondrial genomes of the wombat, *Vombatus ursinus*, and the spiny anteater, *Tachyglossus aculeatus*: increased support for the Marsupionta hypothesis. J Mol Evol 54, 71-80.

Jiang, J., Li, P., Yu, J., Zhao, G., Yi, Y., Yue, B. and Li, J., 2014. The complete mitochondrial genome of Assamese Macaques (Macaca assamensis). Mitochondrial DNA.

Li, D., Fan, L., Zeng, B., Yin, H., Zou, F., Wang, H., Meng, Y., King, E. and Yue, B., 2009. The complete mitochondrial genome of Macaca thibetana and a novel nuclear mitochondrial pseudogene. Gene 429, 31-6.

Liedigk, R., Roos, C., Brameier, M. and Zinner, D., 2014. Mitogenomics of the Old World monkey tribe Papionini. BMC Evol Biol 14, 176.

Lin, Y.H., Waddell, P.J. and Penny, D., 2002. Pika and vole mitochondrial genomes increase support for both rodent monophyly and glires. Gene 294, 119-29.

Mouchaty, S.K., Catzeflis, F., Janke, A. and Arnason, U., 2001. Molecular evidence of an African Phiomorpha-South American Caviomorpha clade and support for Hystricognathi based on the complete mitochondrial genome of the cane rat (Thryonomys swinderianus). Mol Phylogenet Evol 18, 127-35.

Schmitz, J., Ohme, M. and Zischler, H., 2002. The complete mitochondrial sequence of Tarsius bancanus: evidence for an extensive nucleotide compositional plasticity of primate mitochondrial DNA. Mol Biol Evol 19, 544-53.

Wang, J.K., Tang, Y.Q., Li, S.Y., Mai, C. and Gong, Y.F., 2014. The complete mitochondrial genome of Japanese macaque, Macaca fuscata fuscata (Macaca, Cercopithecinae). Mitochondrial DNA, 1-2.

Xu, L., Chen, S.Y., Nie, W.H., Jiang, X.L. and Yao, Y.G., 2012. Evaluating the phylogenetic position of Chinese tree shrew (*Tupaia belangeri chinensis*) based on complete mitochondrial genome: implication for using tree shrew as an alternative experimental animal to primates in biomedical research. J Genet Genomics 39, 131-7.

Xu, X. and Arnason, U., 1996a. A complete sequence of the mitochondrial genome of the western lowland gorilla. Molecular Biology and Evolution 13, 691-8.

Xu, X. and Arnason, U., 1996b. The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. J Mol Evol 43, 431-7.