Sequencing and analysis of the complete mitochondrial genome of the Siberian large-toothed shrew (Sorex daphaenodon) from China

Liu Zhu, Qin Ke Song, Li Jiao Jiao and Dong Ming
College of Life Science and Technology, Mudanjiang Normal University, Mudanjiang, P.R. China

ABSTRACT
The complete mitogenome sequence of the Siberian large-toothed shrew (Sorex daphaenodon) was determined using long PCR. The genome was 17,407 bp in length and contained 13 protein-coding genes, 2 ribosomal RNA genes, 23 transfer RNA genes, 1 origin of L strand replication, and 1 control region. The overall base composition of the heavy strand is A (33.1%), C (24.7%), T (29.0%), and G (13.2%). The base compositions present clearly the A–T skew, which is most obvious in the control region and protein-coding genes. The extended termination-associated sequence domain, the central conserved domain, and the conserved sequence block domain are defined in the mitochondrial genome control region of the Siberian large-toothed shrew. Mitochondrial genome analyses based on MP, ML, NJ, and Bayesian analyses yielded identical phylogenetic trees. The 11 Sorex species formed a monophyletic group with the high bootstrap value (100%) in all examinations.

In this paper, the complete mitochondrial genome of the Siberian large-toothed shrew (Sorex daphaenodon) was sequenced for the first time on ABI 3730XL using a primer walking strategy and the long and accurate Polymerase Chain Reaction (PCR), with 5 pairs of long PCR primers and with 14 pairs of sub-PCR primers. A muscle sample was obtained from a female the Siberian large-toothed shrew captured from Huzhong regions of Daxinganling Mountains in Heilongjiang Province, China (51°44′14″ N, 123°40′44″ E). The muscle tissue was preserved in 95% ethanol and stored at −75°C before use. The specimen and its DNA is stored in Animal and Plant Herbarium of Mudanjiang Normal University. The voucher number is D201620.

The mitochondrial genome is a circular double-stranded DNA sequence that is 17,407 bp long including 13 protein-coding genes, 2 rRNA genes, 23 tRNA genes, 1 origin of L strand replication, and 1 control region. The accurate annotated mitochondrial genome sequence was submitted to GenBank with accession number MK110676. The arrangement of the multiple genes is in line with other Soricidae species (Nikaido et al. 2001; Fontanillas et al. 2005; Kim et al. 2013, 2017; Huang et al. 2014, 2016; Xu et al. 2016; Liu et al. 2016, 2017a, 2017b, 2017x, 2018a, 2018b; Jin et al. 2017) and most mammals (Meganathan et al. 2012; Yoon et al. 2013; Xu et al. 2012, 2013).

The control region of the Siberian large-toothed shrew mitochondrial genome was located between the tRNA-Pro and tRNA-Phe genes, and contains only promoters and regulatory sequences for replication and transcription, but no structural genes. Three domains were defined in the Siberian large-toothed shrew mitochondrial genome control region (Zhang et al. 2009): the extended termination-associated sequence (ETAS) domain, the central conserved domain (CD), and the conserved sequence block (CSB) domain. Three CSBs were found in the CSB domain and they were located in positions 16,608–16,632, 17,111–17,142, and 17,164–17,189. Also, only one repetitive sequence region (RS) was found, which was located between the CSB1 and CSB2, and was rich in A and C. The repetitive pattern of segments in the RS was 5’-TA-(TACACG)n-TA-3’ (n = 55).

The total length of the protein-coding gene sequences was 11,417bp. Most protein-coding genes initiate with ATG except for ND2, ND3, and ND5, which began with ATA or ATT. Six protein-coding genes terminated with TAA whereas the Cyt b gene terminated with AGG. The incomplete stop codons (T– or TA–) were used in ND1, ND2, COX3, ND3, and ND6. A strong bias against A at the third codon position was observed in the protein-coding genes. The frequencies of CTA (Leu), ATT (Ile), TTA (Leu) and ATA (Met) were higher than those of other codons. The length of tRNA genes varied from 57 to 75 bp. Twenty-one of them could be folded into the typical cloverleaf secondary structure except the tRNA-Ser (AGY), whose complete dihydrouridine arm was lacking.

Most the Siberian large-toothed shrew mitochondrial genes were encoded on the H strand, except for the ND6 gene and eight tRNA genes, which were encoded on the L strand. Some reading frame intervals and overlaps were found. One of the most typical was between ATP8 and ATP6. The L-strand replication origin (OL) was located within the WANCY region containing five tRNA genes (tRNA(Trp), tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Tyr). This region was 33 bp long.
and had the potential to fold into a stable stem-loop secondary structure. The total base composition of the Siberian large-toothed shrew mitochondrial genome was A (33.1%), C (24.7%), T (29.0%), and G (13.2%). The base compositions clearly present the A-T skew, which was most obvious in the control region and protein-coding genes.

In order to explore the evolution of Insectivora shrews which include Soricidae and Talpidae, especially the evolution of genus *Sorex* from China, here, we investigate the molecular phylogenetics of Chinese the Siberian large-toothed shrew using complete mitochondrial genome sequence of 32 species. All sequences generated in this study have been deposited in the GenBank (Figure 1).

Mitochondrial genome analyses based on MP, ML, NJ, and Bayesian analyses yielded identical phylogenetic trees, indicating a close phylogenetic affinity of shrews. The phylogram obtained from Maximum Parsimony method is shown in Figure 1. It shows that two major phyletic lineages were present in Insectivora: Soricidae and Talpidae. Soricidae comprised *Crocidura lasiura*, *Crocidura shantungensis*, *Crocidura attenuata*, *Crocidura russula*, *Episoriculus macrurus*, *Episoriculus caudatus*, *Neomys fodiens*, *Nectogale elegans*, *Anourosorex squamipes*, *Blarinella quadricauda*, *Soriculus fumidus*, *Sorex araneus*, *Sorex cylindricauda*, *Sorex unguiculatus*, *Sorex tundrensis*, *Sorex caecutiens*, *Sorex minutissimus* and *Sorex daphaenodon* was supported by bootstrap values of 100%. Talpidae comprised *Talpa europaea*, *Urotrichus talpoides*, *Mogera robusta*, *Galemys pyrenaicus*, *Uropsilus gracilis*, and *Scapanulus oweni* was supported by bootstrap values of 100%. The 11 *Sorex* species formed a monophyletic group with the high bootstrap value (100%) in all examinations.

**Disclosure statement**
The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

**Funding**
This research was supported by the Heilongjiang Provincial Department of Education Filing Project [1352MSY0828], Heilongjiang Provincial Natural Funds [C2017065], and the Doctoral Scientific Research Foundation Project of Mudanjiang Normal University [MNUB201404].
References

Fontanillas P, Depraz A, Giorgi MS, Perrin N. 2005. Nonshivering thermogenesis capacity associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew, Crocidura russula. Mol Ecol. 14:661–670.

Huang T, Yan CC, Tan Z, Tu FY, Yue BS, Zhang XY. 2014. Complete mitochondrial genome sequence of Nectogale elegans. Mitochondrial DNA. 25:253–254.

Huang T, Dang X, An M, Chen L, Zhang J. 2016. The complete mitochondrial genome of the Sorex araneus. Mitochondrial DNA. 27:3655–3656.

Jin ZM, Liu Z, Ma JZ. 2017. Sequencing and analysis of the complete mitochondrial genome of the masked shrew (Sorex caecutiens) from China. Mitochondrial DNA Part B. 2:486–488.

Kim HR, Park JK, Cho JY, Chul Park Y. 2013. Complete mitochondrial genome of an Asian Lesser White-toothed Shrew, Crocidura shantungensis (Soricidae). Mitochondrial DNA. 24:202–204.

Kim TW, Kim YK, Oh DJ, Park JH, Kim D, Adhikari P, Kim G, Park SM, Lee JW, Jung YH, et al. 2017. Complete mitochondrial genome of the Ussuri white-toothed shrew Crocidura lasiura (Insectivora, Soricidae). Mitochondrial DNA Part A. 28:216–217.

Liu Z, Bai W, Wang AN, Tian XM, Li DW. 2018a. Sequencing and analysis of the complete mitochondrial genome of the taiga shrew (Sorex isodon) from China. Mitochondrial DNA Part B. 3:466–468.

Liu Z, Dang YQ, Li JJ. 2018b. Sequencing and analysis of the complete mitochondrial genome of the Eurasian least shrew (Sorex minutissimus) from China. Mitochondrial DNA Part B.

Liu Z, Tian XM, Jin JL, Jin ZM, Li DW, Zhang JS. 2017a. Sequencing and analysis of the complete mitochondrial genome of the slender shrew (Sorex gracillimus) from China. Mitochondrial DNA Part B. 2:642–644.

Liu Z, Tian XM, Jin ZM, Dong M, Zhang JS. 2017b. Sequencing and analysis of the complete mitochondrial genome of the Ussuri shrew (Sorex mirabilis) from China. Mitochondrial DNA Part B. 2:645–647.

Liu Z, Wang AN, Zhang JS, Yang X, Liu H. 2017c. Sequencing and analysis of the complete mitochondrial genome of flat-skulled shrew (Sorex roboratus) from China. Mitochondrial DNA Part B. 2:369–371.

Liu Z, Zhao W, Liu P, Li S, Xu C. 2016. The complete mitochondrial genome of Eurasian water shrew (Neomys fodiens). Mitochondrial DNA Part A. 27:2381–2382.

Meganathan PR, Pagan HJT, McCulloch ES, Stevens RD, Ray DA. 2012. Complete mitochondrial genome sequences of three bats species and whole genome mitochondrial analyses reveal patterns of codon bias and lend support to a basal split in Chiroptera. Gene. 492:121–129.

Nikaido M, Kawai K, Cao Y, Harada M, Tomita S, Okada N, Hasegawa M. 2001. Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. J Mol Evol. 53:508–506.

Xu CZ, Zhang HH, Ma JZ. 2013. The complete mitochondrial genome of sable, Martes flavigula. Mitochondrial DNA. 24:240–242.

Xu CZ, Zhang HH, Ma JZ, Liu ZH. 2012. The complete mitochondrial genome of sable, Martes zibellina. Mitochondrial DNA. 23:167–169.

Xu CZ, Zhao S, Wu HL, Wu SY, Zhang ZW, Wang B, Dou HS. 2016. Sequencing analysis of the complete mitochondrial genome of tundra shrew (Sorex tundrensis) from China. Mitochondrial DNA. 27:2354–2355.

Yoon KB, Kim HR, Kim JY, Jeon SH, Park YC. 2013. The complete mitochondrial genome of the Ussurian tube-nosed bat Murina ussuriensis (Chiroptera: Vespertilionidae) in Korea. Mitochondrial DNA. 24:397–399.

Zhang HH, Xu CZ, Ma JZ. 2009. Structure of the mtDNA control region and phylogeny of the Mustelidae species. Acta Ecol Sin. 29:3585–3592.