

Sequencing and analysis of the complete mitochondrial genome of the masked shrew (*Sorex caecutiens*) from China

Zhi-Min Jin^{a,b}, Liu Zhu^b and Jian-Zhang Ma^a

^aCollege of Wildlife Resources, Northeast Forestry University, Harbin, P.R. China; ^bCollege of Life Science and Technology, Mudanjiang Normal University, Mudanjiang, P.R. China

ABSTRACT

The complete mitogenome sequence of the masked shrew (*Sorex caecutiens*) was determined using long PCR. The genome was 17,096 bp in length and contained 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, one origin of L strand replication, and one control region. The overall base composition of the heavy strand is A (32.9%), C (24.5%), T (29.3%), and G (13.3%). The base compositions present clearly the A–T skew, which is most obviously 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 masked shrew. Mitochondrial genome analyses based on MP, ML, NJ, and Bayesian analyses yielded identical phylogenetic trees. The five *Sorex* species formed a monophyletic group with the high bootstrap value (100%) in all examinations.

ARTICLE HISTORY

Received 2 July 2017
Accepted 26 July 2017

KEYWORDS

Control region;
mitogenome; phylogenetic
trees; masked shrew

In this paper, the complete mitochondrial genome of the masked shrew (*Sorex caecutiens*) was sequenced for the first time on ABI 3730XL using a primer walking strategy and the long and accurate PCR, with five pairs of long PCR primers and with 14 pairs of sub-PCR primers. A muscle sample was obtained from a female the masked shrew captured from Luobei region of Small Khingan Mountains in Heilongjiang Province, China (45°50′07″ N, 132°84′74″ E). The muscle tissue was preserved in 95% ethanol and stored at –75 °C before use.



The mitochondrial genome is a circular double-stranded DNA sequence that is 17,096 bp long including 13 protein-coding genes, two rRNA genes, 22 tRNA genes, one origin of L strand replication, and one control region. The accurate annotated mitochondrial genome sequence was submitted to GenBank with accession number MF374796. The arrangement of the multiple genes is in line with other Soricidae species (Nikaido et al. 2001; Fontanillas et al. 2005; Huang et al. 2014, 2016; Xu et al. 2016) and most mammals (Meganathan et al. 2012; Xu et al. 2012, 2013; Yoon et al. 2013).

The control region of the masked 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 masked 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 16,426–16,450, 16,813–16,844, and 16,866–16,891. 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-(CACGTACGCCTATA)*n*-CG-3′ (*n* = 16).

The total length of the protein-coding gene sequences was 11,425 bp. Most protein-coding genes initiate with ATG except for ND2, ND3, and ND5, which began with ATA or ATC. Seven 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 ND4. 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 59 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 of the masked 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 (tRNATrp, tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Tyr). This region was 36 bp long and had the potential to fold into a stable stem-loop secondary structure.

CONTACT Jian-Zhang Ma  jianzhangma@163.com  College of Wildlife Resources, Northeast Forestry University, Harbin, P.R. China

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

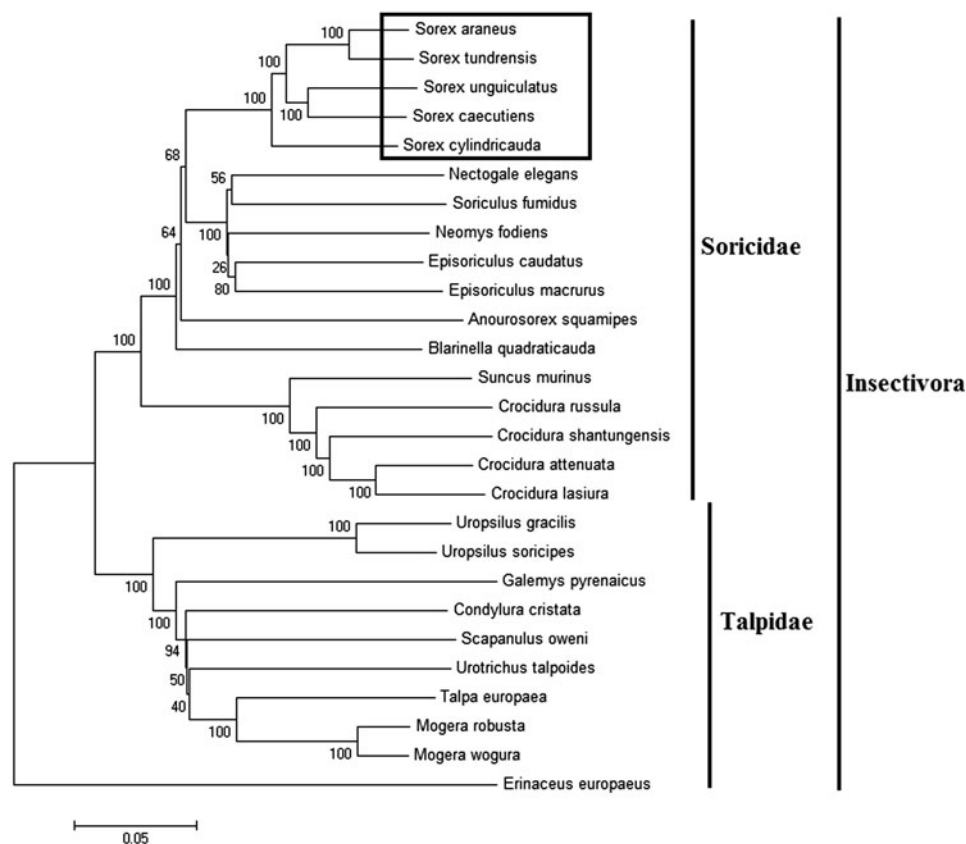


Figure 1. Phylogenetic tree generated using the Maximum Parsimony method based on complete mitochondrial genomes. *Crocidura lasiura* (KR007669), *Crocidura shantungensis* (JX968507), *Crocidura attenuata* (KP120863), *Crocidura russula* (AY769264), *Episoriculus macrurus* (KU246040), *Episoriculus caudatus* (KM503097), *Neomys fodiens* (KM092492), *Nectogale elegans* (KC503902), *Anourosorex squamipes* (KJ545899), *Blarinella quadraticauda* (KJ131179), *Suncus murinus* (KJ920198), *Soriculus fumidus* (AF348081), *Sorex araneus* (KT210896), *Sorex cylindricauda* (KF696672), *Sorex unguiculatus* (AB061527), *Sorex tundrensis* (KM067275), *Sorex caecutiens* (MF374796), *Talpa europaea* (Y19192), *Urotrichus talpoides* (AB099483), *Uropsilus soricipes* (JQ658979), *Uropsilus gracilis* (KM379136), *Mogera wogura* (AB099482), *Mogera robusta* (KT934322), *Condylura cristata* (KU144678), *Galemys pyrenaicus* (AY833419), *Scapanulus oweni* (KM506754), *Erinaceus europaeus* (NC002080).

The total base composition of the masked shrew mitochondrial genome was A (32.9%), C (24.5%), T (29.3%), and G (13.3%). The base compositions clearly present the A-T skew, which was most obviously 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 masked shrew using complete mitochondrial genome sequence of 26 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 quadraticauda*, *Soriculus fumidus*, *Suncus murinus*, *Sorex araneus*, *Sorex tundrensis*, *Sorex caecutiens*, *Sorex cylindricauda*, and *Sorex unguiculatus* and was supported by bootstrap values of 100%. Talpidae comprised *Talpa europaea*, *Urotrichus talpoides*, *Mogera wogura*,

Condylura cristata, *Uropsilus soricipes*, *Mogera robusta*, *Galemys pyrenaicus*, *Uropsilus gracilis*, and *Scapanulus oweni* and was supported by bootstrap values of 100%. The three *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 Department of Education, Heilongjiang Province filing project [1351MSYYB009], the Doctoral Scientific Research Foundation Project of Mudanjiang Normal University [MNUB201404], and Key Innovative Advanced Research Project of Mudanjiang Normal University [GY2014001].

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, Dang X, An M, Chen L, Zhang J. 2016. The complete mitochondrial genome of the *Sorex araneus*. *Mitochondrial DNA A DNA Mapp Seq Anal*. 27:3655–3656.

- 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.
- 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 and analysis of the complete mitochondrial genome of tundra shrew (*Sorex tundrensis*) from China. *Mitochondrial DNA*. 27:2354–2355.
- Yoon KB, Rikim H, 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.