The complete mitochondrial genome of the scorpion Centruroides vittatus (Arachnida: Scorpionidae)

Tsunemi Yamashita, Douglas Rhoads and Jeff Pummill

Department of Biological Sciences, Arkansas Tech University, Russellville, AR, USA; Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA; Arkansas High Performance Computing Center, University of Arkansas, Fayetteville, AR, USA

ABSTRACT
The complete mitochondrial genome (mitogenome) of the Striped scorpion (Centruroides vittatus) was assembled from Illumina-based whole genome sequencing. The circular genome is 14,602 bp in length with 13 protein coding genes, 21 tRNA, two rRNAs, a translocation-inversion of tRNA^Leu compared to the horse shoe crab mitogenome, and the absence of tRNA^Asp. The A+T content of the mitogenome is 68.1%. Our Bayesian and maximum likelihood phylogenetic analyses placed the C. vittatus mitogenome as a sister group of C. limpidus and nestled within the new world Buthids.

Introduction
Arachnids are a diverse and ancient arthropod taxa, yet they are under represented in genomic studies, including mitochondrial genomes (165 arthropod species/2477 arthropod species). Within the arachnids, the bulk of mitogenomes submitted to NCBI include spiders (36 species), scorpions (eight species), harvestmen (three species), sun spiders (two species), whip spiders (two species), pseudoscorpions (two species), and whip scorpions (two species). The Buthid scorpion family compose the majority of the published scorpion mitogenomes; however, within the genus Centruroides, only C. limpidus is published. Here, we present the complete mitogenome for the striped scorpion, C. vittatus.

The striped scorpion, C. vittatus, is a common scorpion in the Midwestern United States and the northeastern states of Mexico. Although, it is medically important to humans, due to its neurotoxic venom, its venom appears to show reduced adverse effects compared to the western C. sculpturatus.

Total genomic DNA was extracted from a male scorpion collected in Pope County, AR (35° 31’36.96"N, 93° 08’29.21"W) with the Fast ID genomic DNA extraction kit (Genetic IDNA, Inc.). The genomic DNA quality was analysed through 0.9% agarose gel electrophoresis and by UV spectrophotometry. The genomic DNA was sent to the DNA sequencing core facility at UAMS (Univ. AR Med. Sci.) for library generation and 2 × 300 paired end sequencing in an Illumina MiSeq small genome sequencer. The de novo assembly was conducted at the High Performance Computing Center at the University of AR-Fayetteville. The raw sequence data were processed (adapter removal, quality trimming, and merging of data) with BBmap (Bushnell 2016). The processed data were assembled with SPAdes v3.7.1 (Bankevich et al. 2012) and Ray v2.3.1 (Boisvert et al. 2010) with quality assessment with the program Quast v4.0 (Gurevich et al. 2013). The mitogenome was extracted from the assembly, confirmed by template reassembly using NGen (DNASTar 12.3) and annotated using BASys v1.0 (Van Domselaar et al. 2005) and with reference alignment to the published C. limpidus mitogenome (Dávila et al. 2005). The finished genome was deposited in NCBI with the accession number Bankit2047317 Centruroides MF975702.

The C. vittatus mitogenome matches the published C. limpidus mitogenome (81.6%) with a similar genome size (14,602–14,519 bp in C. limpidus) and order of the 13 protein coding genes, 21 tRNA, and two rRNAs. Both Centruroides mitogenomes differ from the horseshoe crab mitogenome with a translocation-inversion of tRNA^Leu and the absence of tRNA^Asp. Both genomes are similar in size to other scorpion genomes, but smaller than for Mesobuthus martensii (15,034 bp) (Choi et al. 2007). The A+T content of the C. vittatus mtDNA assembly is also similar to C. limpidus (68.1–64.46%, respectively). Most of the nucleotide variation between the C. vittatus and C. limpidus mtDNA genome occurs in the putative non-coding control region with C. vittatus showing a 97 bp insertion.

We created a phylogeny showing the placement of the C. vittatus mitogenome with respect to eight scorpion, four spider, and one mite mitogenomes. The consensus tree was created through Bayesian methods with MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003). The mite Tetranychus truncatus was selected as the outgroup (Figure 1). A phylogenetic second tree was created with maximum likelihood methods.

CONTACT Tsunami Yamashita tyamashita@atu.edu Department of Biological Sciences, Arkansas Tech University, Russellville, AR, USA

This work was authored as part of the Contributor’s official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law. 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.
in RAxML v8.0.0 (Stamatakis 2014). As this tree was identical to the Bayesian tree, we only present the Bayesian tree.

**Disclosure statement**

The authors report no conflicts of interest. The authors share responsibility for the content and writing of the paper.

**Funding**

This project was supported by the Arkansas INBRE Program, with a grant from the National Institute of General Medical Sciences (NIGMS), P20 GM103429 from the National Institutes of Health.

**References**

Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin V, Nikolenko S, Pham S, Prjibelski A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19:455–477.

Boisvert S, Laviolette F, Corbeil J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol. 17:1519–1533.

Bushnell B. 2016. BBMap short read aligner. Berkeley (CA): University of California. http://sourceforge.net/projects/bbmap/

Choi EH, Park SJ, Jang KH, Hwang W. 2007. Complete mitochondrial genome of a Chinese scorpion Mesobuthus martensi (Chelicerata, Scorpiones, Buthidae). DNA Seq. 18:459–471.

Dávila S, Piñero D, Bustos P, Cevallos MA, Dávila G. 2005. The mitochondrial genome sequence of the scorpion Centruroides limpidus (Karsch 1879) (Chelicerata; Arachnida). Gene. 360:92–102.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29:1072–1075.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.

Van Domelaar GH, Stothard P, Shivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS. 2005. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. 33:W455–W459.