MITOCHONDRIAL DNA PART B: RESOURCES
2024, VOL. 9, NO. 1, 178–181
https://doi.org/10.1080/23802359.2024.2305407

Taylor & Francis
Taylor & Francis Group

MITOGENOME REPORT

OPEN ACCESS Check for updates

Complete mitochondrial genome of the Korean endemic springtail *Tomocerus caputiviolaceus* Lee 1975 (Collembola: Tomoceridae)

Seon Hwa Jo*, Jachoon Koo* and Kyung-Hwa Park

Division of Science Education and Institute of Science Education, Jeonbuk National University, Jeonju, South Korea

ABSTRACT

The complete mitochondrial genome of *Tomocerus caputiviolaceus* was sequenced and assembled. The complete mitochondrial genome is 15,519 bp in length. The mitogenome contained 37 genes, including 13 protein-coding genes (PCGs), 22 tRNAs, and two rRNAs. In phylogenetic analysis based on the nucleotide sequences of 13 PCGs, *T. caputiviolaceus* is closely related to *Tomocerus qinae* Yu, Yao & Hu, 2016, both of which belong to the genus *Tomocerus* within the family Tomoceridae.

ARTICLE HISTORY
Received 27 September 2023
Accepted 9 January 2024

KEYWORDS
Springtail; Tomoceridae; complete mitochondrial genome; phylogenetic analysis

1. Introduction

The Korean endemic species *Tomocerus caputiviolaceus* Lee 1975 belongs Tomoceridae family. This species was first described from specimens collected from leaf litter from Mt. Deogyusan (Muju), Korea (Collection no. 71-3-3:25) (Lee 1975). This species is characterized by purple head and elongated body. Holotype is deposited in the National Institute of Biological Resources (NIBR), Incheon, Republic of Korea. This is the first mitogenome report for this species. Although approximately 100 species of the genus *Tomocerus* have been described worldwide (https://www.collembola.org/, Bellinger et al. 1996–2023,), one complete mitogenome, *Tomocerus qinae* Yu, Yao, & Hu, 2016, has been recorded (Sun et al. 2020). Therefore, the mitogenome of *T. caputiviolaceus* was sequenced and annotated to provide valuable information for future research of the genetic relationships between *T. caputiviolaceus* and other Collembolan species.

2. Materials

Specimens were collected in July 2021 at Mt. Deogtaesan in Jinan, South Korea (35°40′45″ N, 127°26′42″ E), and stored in 95% ethyl alcohol. The samples in alcohol were imaged using a Nikon SMZ800 stereo microscope equipped with a DS-Fi1 camera (Figure 1). The voucher specimen is deposited and preserved at Insect Collection in the Division of Science Education, Jeonbuk National University, Jeonju, South Korea (Kyung-Hwa Park, pkhsyst@jbnu.ac.kr) under the voucher number '157-Se-TC'. The voucher specimen was identified by comparison with the holotype of *Tomocerus caputiviolaceus* Lee 1975. Species identification was performed based on the body chaetotaxy and dental formula key as the main diagnostic features (Lee 1975; Yu et al. 2018).

Figure 1. Habitus of *T. caputiviolaceus*. Scale bar is 0.5 mm. Photograph was taken by Seon Hwa Jo.

3. Methods

3.1. DNA extraction

Total genomic DNA was isolated from a single specimen using QIAamp DNA Micro Kit (QIAGEN Ltd., Crawley, UK).

CONTACT Kyung-Hwa Park pkhsyst@jbnu.ac.kr Division of Science Education and Institute of Science Education, Jeonbuk National University, Jeonju 54896, South Korea
*These authors contributed equally to this work.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23802359.2024.2305407.

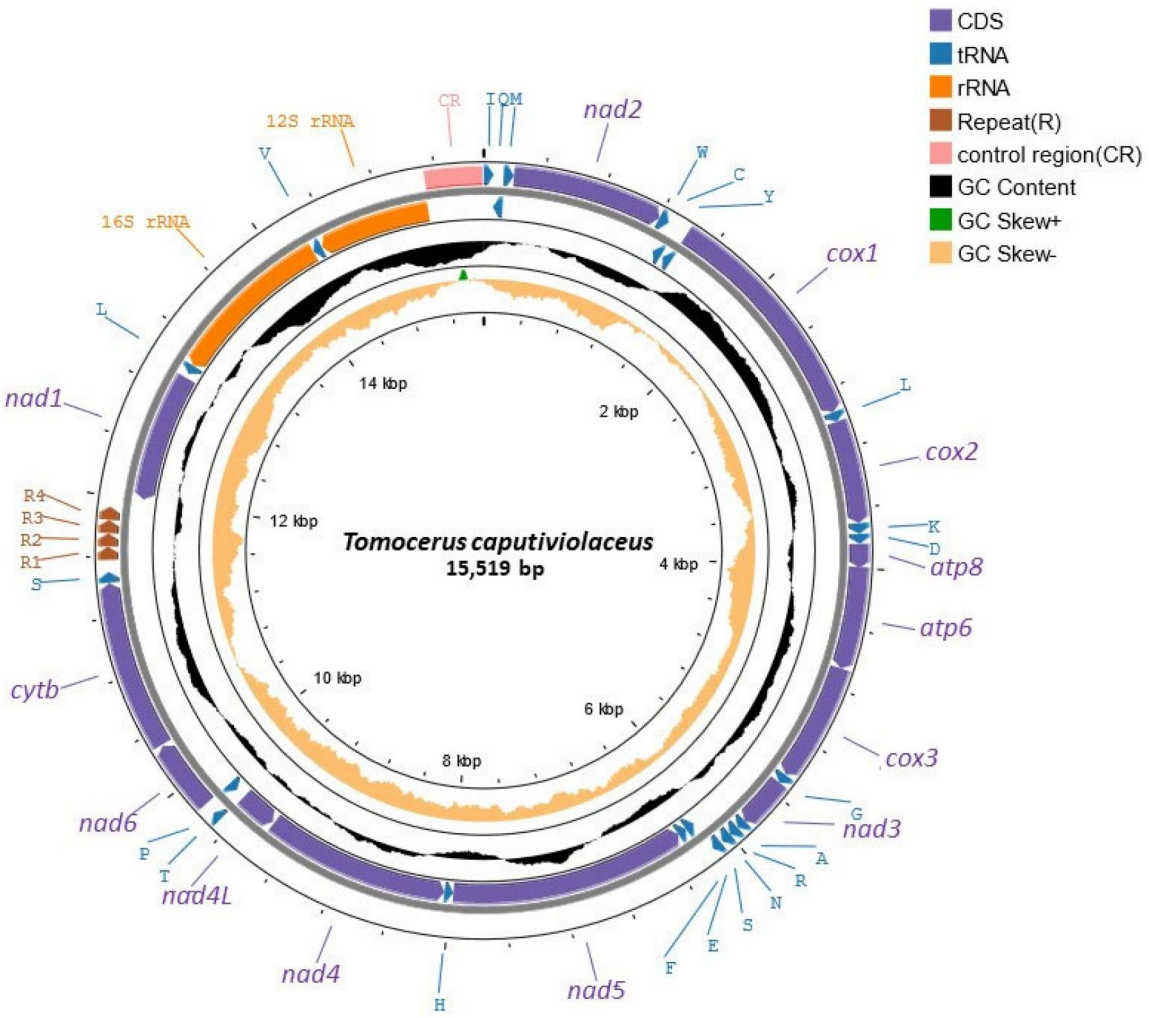

Figure 2. Mitochondrial map of *Tomocerus caputiviolaceus*. Circular maps were generated with the Proksee web server (https://proksee.ca/). protein-coding genes are shown as purple, rRNA genes as orange, tRNA genes as blue, and repeat sequences as brown. The GC content and GC-skew values were calculated using a 500 bp sliding window with a step of 1 bp. The GC content is plotted using a black sliding window. The positive and negative values of GC-skew are plotted using the yellow and green sliding window, respectively. The inner cycle indicates the location of the genes in the mt genome.

3.2. Mitogenome sequencing, assembly, and annotation

The DNA was fragmented by sonication and used to construct a sequencing library by ligating of 5′ and 3′ adapters using the Illumina TruSeq DNA Nano Library Prep Kit (Illumina Inc., USA). The paired-end (2 × 150 bp) sequence of *T. caputiviolaceus* mitogenome was performed using the Illumina HiSeq-X platform (Illumina Inc., USA) at Macrogen Inc. (Seoul, Korea). 37,496,208 raw reads were processed using Trimmomatic (Bolger et al. 2014) to trim the adaptor sequence and filtered poor-quality reads less than 36 bp in length. The filtered sequences were assembled using SPAdes (Bankevich et al. 2012) and NOVOPlasty (Dierckxsens et al. 2017). Since two assembled sequences showed a little discrepancy at nucleotide position 12,003 (Supplementary Figure S1), the sequence was validated by PCR-based Sanger sequencing. PCR amplifications were performed using the primer set: forward primer, 5′-CTTAATTAGCCCATTGTCTACACC-3′ (position 11,504–11,527) and reverse primer, 5′-TTAACATTGAGTATGGAAG AGGAG-3′ (position 12,325–12,302). PCR program was 95 °C for 3 min, 35 cycles of 95 °C for 40 s/55 °C for 40 s/72 °C for 1 min, and 72 °C for 50 min. The sequence of the PCR product was determined by Macrogen's sequencing service using reverse primer. The sequencing result of the PCR product was identical to the assembled sequence by NOVOplasty (Supplementary Figure S1). In addition, the region of low-depth coverage (a putative control region with TA repeats) was also confirmed by PCR-based Sanger sequencing (Supplementary Figure S2). PCR amplification was performed using the primer set: forward primer, 5′-GGGTATCTAATCCAGGTCTATTGA-3′ (position 14,741–14,764) and reverse primer, 5′-GACTGATTCTGTTGATAGAGAAGC-3′ (position 370–347). The PCR program was the same as above. The depth coverage was measured to validate the final mitogenome sequence of *T. caputiviolaceus* (Supplementary Figure S3). The mitogenome was annotated using MitoZ (Meng et al. 2019), and then annotations were manually curated to adjust gene boundaries based on known annotations of closely related species.

3.3. Phylogenetic analysis

Phylogenetic analysis of *T. caputiviolaceus* was performed based on mitogenomes of nine species in the superfamily Tomoceroidea and two outgroup species. The complete mitogenome sequences were downloaded from NCBI's Genbank.

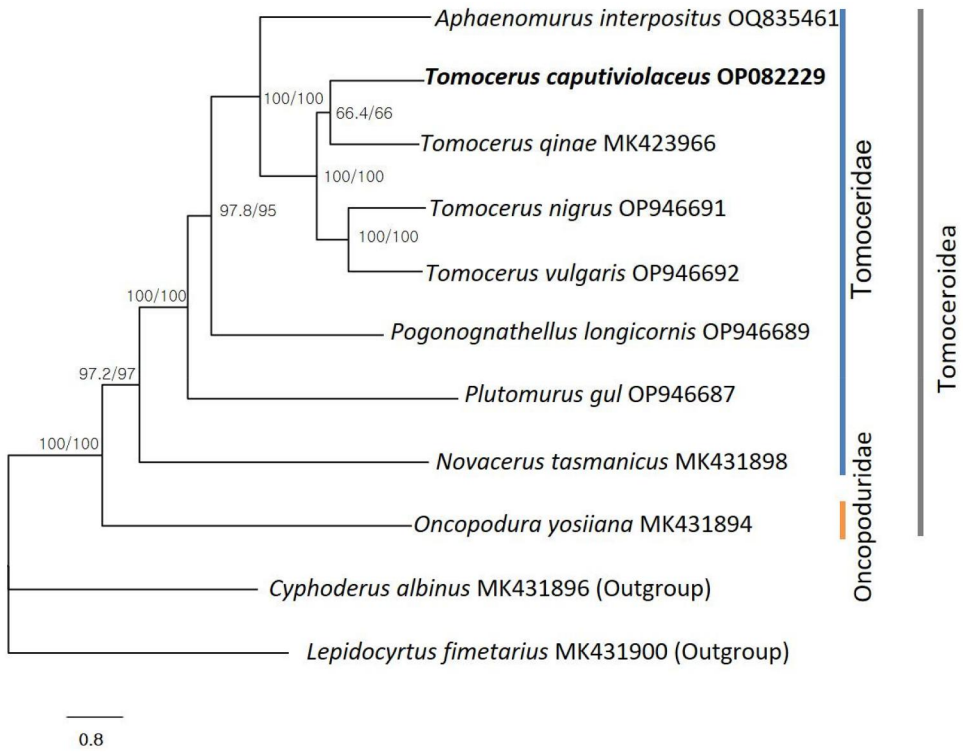

Figure 3. ML tree showing phylogenetic relationships of *Tomocerus caputiviolaceus* (Lee 1975) (marked with bold font) among nine Tomoceroidea species. Two Entomobryoidea species were used as outgroups. Tree was reconstructed using aligned nucleotide sequences of 13 PCGs. The numbers at each node specify the SH-aLRT support (%) and ultrafast bootstrap support (%), respectively. The scale bar indicates the number of substitutions per site. The following sequences were used: *Tomocerus caputiviolaceus* OP082229 (This study), *Aphaenomurus interpositus* OQ835461 (unpublished), *Tomocerus qinae* MK423966 (Sun et al. 2020), *Tomocerus nigrus* OP946691 (unpublished), *Tomocerus vulgaris* OP946692 (unpublished), *Pogonognathellus longicornis* OP946689 (unpublished), *Plutomurus gul* OP946687 (unpublished), *Novacerus tasmanicus* MK431898 (Sun et al. 2020), *Oncopodura yosiiana* MK431894 (Sun et al. 2020), *Cyphoderus albinus* MK431896 (Sun et al. 2020), and *Lepidocyrtus fimetarius* MK431900 (Sun et al. 2020).

Using Geneious (https://www.geneious.com/) software, the nucleotide sequences of 13 protein-coding genes (PCGs) were identified and extracted from the annotated or non-annotated mitogenomes. Multiple alignment was performed separately using MACSE (Ranwez et al. 2018). Gaps were removed with trimAl (Capella-Gutiérrez et al. 2009) and then concatenated into a single dataset at final. Phylogenetic tree was reconstructed using the maximum likelihood (ML) approach in IQ-TREE (Nguyen et al. 2015). Best fit model for three codon sites was automatically in ModelFinder (Kalyaanamoorthy et al. 2017), and the robustness of obtained topology was evaluated with ultrafast bootstrap (Minh et al. 2013) and SH-aLRT (Guindon et al. 2010). The phylogenetic tree was edited using FigTree.

4. Results

4.1. Genome organization and composition

The complete mitogenome of *T. caputiviolaceus* is 15,519 bp in length (Figure 2). It contains 37 genes, including 13 PCGs, 22 transfer RNA (tRNA) genes, and two ribosomal RNA (rRNA) genes. The start codons of 13 PCGs are ATA (*cox1*, *atp8*, *nad3*, *nad4L*), ATG (*atp6*, *cox3*) and ATT (*nad2*, *cox2*, *nad5*, *nad4*, *nad6*, *nad1*, *cytb*), respectively. The stop codons of the 13 PCGs are TAA (*nad2*, *atp8*, *atp6*, *cox3*, *nad5*, *nad4*, *nad4L*, *nad6*, *cytb*, *nad1*), TAG (*nad3*) and the incomplete stop codon T (*cox1* and *cox2*), respectively. The overall nucleotide composition is 40.9% A, 15.1% C, 9.8% G, and 34.2% T, with a highly biased A + T content (75.1%). A 375 bp non-coding sequence is located at the junction between 12S *rRNA* and *trnI*. Unlike other Collembola, 4 repeats of a unique sequence (88 bp in length) are found in position between 11657 and 11941 bp (Figure 2).

4.2. Phylogenetic analysis

Phylogenetic analysis based on ML was performed using the nucleotide sequences of 13 PCGs from nine Tomoceroidea species, of which eight species are from Tomoceridae, and one species is from Oncopoduridae (Figure 3). Two Entomobryoidea species were used as outgroups. The ML phylogenetic tree showed that *T. caputiviolaceus* was first clustered with *Tomocerus qinae*, then together with *Tomocerus nigrus* and *Tomocerus vulgaris*. These *Tomocerus* species form a monophyletic clade with *Novacerus tasmanicus* which belongs to the family Tomoceridae. *Oncopodura yosiiana Szeptycki*, 1977 in the family Oncopoduridae is the sister group of Tomoceridae.

5. Discussion and conclusion

The overall mitogenomic content of *T. caputiviolaceus* is strongly biased toward A and T (A + T = 75.1%). In the tree, the family Tomoceridae forms a monophyletic group with high support value. The monophyly of Tomoceridae is

consistent with previous studies based on mitochondrial PCG sequences of Collembola (Sun et al. 2020; Bellini et al. 2022). The monophyly of the superfamily Tomoceroidea was also presented in the previous study (Yu et al. 2022). Further analysis of a larger number of mitogenomes is required to accurately support a monophyletic of the Tomoceroidea superfamily. The mitogenome of *T. caputiviolaceus* will serve as a key reference for these further studies.

Acknowledgements

We sincerely thank Inae Lee, Hakseo Kim, and Gyudong Jang for their assistance in collecting the samples. We would like to thank Editage (www.editage.co.kr) for English language editing.

Author contributions

SHJ and JK performed the experiments, analyzed the data, drafted the manuscript, and prepared figures. KHP devised the main conceptual ideas, designed the study, and approved the final version of this manuscript to be published. All authors discussed and critically revised the results, and agreed to be accountable for all aspects of the work.

Ethics statement

This research does not involve ethical research. Meanwhile, no other permissions are necessary to be provided because the species for this study are not included in the list of protected animals.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

No funding was received.

ORCID

Seon Hwa Jo http://orcid.org/0000-0003-2283-8982
Jachoon Koo http://orcid.org/0000-0002-3559-326X
Kyung-Hwa Park http://orcid.org/0000-0001-6697-8819

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at (https://www.ncbi.nlm.nih.gov/) under the accession number OP082229 the associated BioProject, BioSample, and SRA numbers are PRJNA756712, SAMN29932228, and SRR20662604, respectively.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477. doi:10.1089/cmb.2012.0021.

Bellinger PF, Christiansen KA, Janssens F. 1996–2023. Checklist of the Collembola of the world [accessed 2023 March 31]. http://www.collembola.org.

Bellini BC, Zhang F, de Souza PGC, dos Santos-Costa RC, Medeiros G, Godeiro NN. 2022. The evolution of Collembola higher taxa (Arthropoda, Hexapoda) based on mitogenome data. Diversity. 15(1):7. doi:10.3390/d15010007.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25(15):1972–1973. doi:10.1093/bioinformatics/btp348.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45(4):e18. doi:10.1093/nar/gkw955.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321. doi:10.1093/sysbio/syq010.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 14(6):587–589. doi:10.1038/nmeth.4285.

Lee BH. 1975. Ètude de la faune Corèenne des insects Collemboles. VI. Sur la famille des Tomoceridae, èdaphiques, avec la description de quatre nouvelles espèces et d'une nouvelle sous- espèce. Bulletin du Muséum National D'Histoire Naturelle. 317(224):945–961.

Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 47(11):e63. doi:10.1093/nar/gkz173.

Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 30(5):1188–1195. doi:10.1093/molbev/mst024.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274. doi:10.1093/molbev/msu300.

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol Biol Evol. 35(10):2582–2584. doi:10.1093/molbev/msy159.

Sun X, Yu D, Xie Z, Dong J, Ding Y, Yao H, Greenslade P. 2020. Phylomitogenomic analyses on collembolan higher taxa with enhanced taxon sampling and discussion on method selection. PLoS One. 15(4):e0230827. doi:10.1371/journal.pone.0230827.

Yu D, Ding Y, Tihelka E, Cai C, Hu F, Liu M, Zhang F. 2022. Phylogenomics of elongate-bodied springtails reveals independent transitions from aboveground to belowground habitats in deep time. Syst Biol. 71(5):1023–1031. doi:10.1093/sysbio/syac024.

Yu D, Qin C, Ding Y, Hu F, Zhang F, Yu D. 2018. Revealing species diversity of *Tomocerus ocreatus* complex (Collembola: Tomoceridae): integrative species delimitation and evaluation of taxonomic characters. Arthropod Syst Phylogeny. 76(1):147–172. doi:10.3897/asp.76.e31949.