



Mitogenomic features of three species of Entomobryoidea (Collembola) from China

NERIVANIA NUNES GODEIRO^{1,2*}, CONG XU², GUOQIANG ZHANG² & FENG ZHANG^{2,3}

¹Natural History Research Center, Shanghai Natural History Museum, Shanghai Science & Technology Museum, Shanghai, 200041, China.

²Department of Entomology, College of Plant Protection, Nanjing Agricultural University, 210095, Nanjing, China.

³<https://orcid.org/0000-0002-1371-266X>

*Corresponding author. [✉ nerivania@gmail.com](mailto:nerivania@gmail.com); [ORCID: https://orcid.org/0000-0002-1669-6124](https://orcid.org/0000-0002-1669-6124)

Abstract

Three novel mitogenomes of Entomobryoidea were assembled and annotated. Sequences final lengths were 14,683 bp for *Alloscopus bannaensis*, 14,884 bp for *Sinhomidia bicolor*, and 14,794 bp for *Coecobrya* sp., including 13 PCGs, 22 tRNA, and 2 rRNA genes. To confirm the morphological identifications, maximum likelihood and Bayesian analyses were performed and the species were placed in the correct phylogenetic position, except *A. bannaensis* which support value was low to confirm its placement. All three mitogenomes are the first ones published for these genera, the genetic data generated here will be useful for further phylogenetic analyses of the group Collembola.

Key words: Entomobryinae, Orchesellidae, de novo assembly, mitochondrial, soil fauna

Introduction

Entomobryoidea *sensu* Zhang *et al.* 2019 is a superfamily of Collembola which currently comprises three families: Orchesellidae, Entomobryidae and Paronellidae (Zhang *et al.*, 2019). *Alloscopus bannaensis* Zhang, 2020 (in: Zhang *et al.* 2020) is an Orchesellidae species, distributed mainly in tropical regions of Asia. *Sinhomidia bicolor* (Yosii, 1965) and *Coecobrya* sp. belong to Entomobryidae family, subfamily Entomobryinae. The species *S. bicolor* was transferred from the genus *Acanthocyrtus* Handschin, 1925 in 2009 to the newly created genus *Sinhomidia* Zhang, 2009 (in: Zhang *et al.* 2009) based only on morphological characters (presence of pointed scales and absence of dental scales). *A. bannaensis* was recently described from the southern region of China and its phylogenetic position was recovered in the same article based on two mitochondrial and two nuclear genes (Zhang *et al.* 2020).

Until January 2022, 158 mitogenomes of Collembola were available on GenBank, but this number is still low when compared to the number of Collembola species described around the world (more than 9,000 species) (Bellinger *et al.* 1996–2022). In view of the lack of publicly available mitogenomes of the Collembola genera *Alloscopus* Börner, 1906, *Sinhomidia* and *Coecobrya* Yosii, 1956, here, we performed low-coverage whole genome sequencing to assemble and characterize the complete mitogenomes of *A. bannaensis*, *S. bicolor* and *Coecobrya* sp. This approach has already been used in other studies of Collembola demonstrating a cost-effective way to recover mitogenomes for evolutionary studies (Godeiro *et al.* 2020, 2021; Sun *et al.* 2020).

Material and methods

A. bannaensis was collected in Yunnan province, *S. bicolor* in Anhui province, and *Coecobrya* sp. in Shaanxi province. All *Coecobrya* sp. specimens analysed were juveniles, which made it impossible to describe, but possibly it is a new species. Sample vouchers of the tree species used here are preserved at –20°C in absolute ethanol at Nanjing Agricultural University, China. One specimen of each sample was used for DNA extraction and purification

with the kit QIAamp DNA Micro Kit (QIAGEN GmbH). The DNA quantification was made with Q33230 Qubit™ 1X dsDNA High Sensitive Assay Kit (QIAGEN GmbH). Whole-genome amplification was performed using Qiagen REPLI-g Single Cell kit (QIAGEN GmbH, Shanghai, China), following the manufacturer's protocol. Libraries were constructed and sequenced by Wuhan Frasergen Bioinformatics Co., Ltd., China. MGISEQ2000 platform was used for sequencing paired-end reads with 150 bp length. Approximately 10G of low-coverage data were produced for each sample. NovoPlasty v3.8.3 (Dierckxsens *et al.* 2020) was used to assembly de novo the mitogenome of *A. bannaensis* with kmer value 21 and a previously sequenced COI partial sequence from the same species was used as a reference (accession number MH605084.1). The other two species were assembled de novo using MitoZ v2.4-alpha (Meng *et al.* 2019) with default parameters. Genes were annotated and visualized with MitoZ v2.4-alpha. Genes locations were checked by comparing them to other mitogenomes from the same subfamily. Previously to the phylogenetic analyses, mitogenomes sequences of nine taxa of Entomobryoidea and one outgroup *Desoria trispinata* (Mac Gillivray, 1896) were downloaded from Genbank, taxonomical information and accession numbers are listed in Table 1. The newly assembled mitogenomes were added and the final dataset contained 13 species. Phylogenetic relationships of the three new mitogenomes were inferred from the 13 PCG's. The concatenated matrix had a length of 3,522 sites after trimming. Bayesian phylogenetic inference was performed using PhyloBayes mpi v1.5a (Lartillot *et al.* 2013), with CAT-GTR model, two chains were run until the likelihood had satisfactorily converged (maxdiff < 0.1). Maximum likelihood (ML) inference was performed using IQ-Tree v2.0.7 (Minh *et al.* 2020), ultrafast bootstrap 1,000 replicates (Hoang *et al.* 2018), and SH-aLRT support, model EX-EHO.

TABLE 1. Taxonomical information of the species used in the phylogenetic analyses *sensu* Zhang *et al.* 2019. GenBank numbers marked with (*) correspond to the newly assembled mitogenomes. NA: not applicable.

Species	Superfamily	Family	Subfamily	Voucher code	GenBank number
<i>Desoria trispinata</i> (Mac Gillivray, 1896)	Isotomoidea	Isotomidae	Isotominae	NA	MT611244
<i>Orchesella villosa</i> (Linneus, 1767)	Entomobryoidea	Orchesellidae	Orchesellidae	NA	EU016195
<i>Alloscopus bannaensis</i> Zhang, 2020	Entomobryoidea	Orchesellidae	Heteromurinae	17BN3A	OK037063*
<i>Lepidocyrtus sotoi</i> Bellini & Go-deiro, 2015 (in Bellini <i>et al.</i> 2015)	Entomobryoidea	Entomobryidae	Lepidocyrtinae	NA	MT928545
<i>Pseudosinella tumula</i> Wang, Chen & Christiansen, 2002	Entomobryoidea	Entomobryidae	Lepidocyrtinae	NA	MT611221
<i>Lepidocyrtinus dapeste</i> Santos & Bellini, 2018 (in Bellini <i>et al.</i> 2018)	Entomobryoidea	Entomobryidae	Seirinae	NA	MF716609
<i>Seira dowlingi</i> (Wray, 1953)	Entomobryoidea	Entomobryidae	Seirinae	NA	MW419950
<i>Coecobrya</i> sp.	Entomobryoidea	Entomobryidae	Entomobryinae	COE2	OK037064*
<i>Sinella curviseta</i> Brook, 1882	Entomobryoidea	Entomobryidae	Entomobryinae	NA	MK014212
<i>Sinhomidia bicolor</i> (Yosii, 1965)	Entomobryoidea	Entomobryidae	Entomobryinae	15HWN1	OK037065*
<i>Homidia socia</i> Denis, 1929	Entomobryoidea	Entomobryidae	Entomobryinae	NA	MN480464
<i>Paronellides praefectus</i> Zhang, Ma & Greenslade, 2017	Entomobryoidea	Paronellidae	Paronellinae	NA	MT914178
<i>Zhuqinia jingwanae</i> Zhang, Ma & Greenslade, 2017	Entomobryoidea	Paronellidae	Paronellinae	NA	MT906653

Results and discussion

The total amount of DNA after amplification of each sample was approximately 750 ng/uL, enough to produce high quality libraries and mitogenomes with sufficient reads coverage (Fig. 1). The newly determined circular mitogenomes are 14,683 bp for *A. bannaensis*, 14,884 bp for *S. bicolor*, and 14,794 bp for *Coecobrya* sp. in length,

consisting of 13 protein-coding genes, two ribosomal RNAs (12S rRNA and 16S rRNA), 22 transfer RNA genes, and a noncoding region. All genes did not present any structural rearrangement and the gene order (Fig. 1) is the same as the Pancrustacean model, considered to be the ancestral lineage of springtails (Von Reumont *et al.* 2012). For PCG's, the most frequent pattern for start/end codons was ATT/TAA, followed by ATG/TAA, previous studies found that ATG and TAA are common start and end codons for Collembola (Carapelli *et al.* 2014; Nunes *et al.* 2019). Despite our efforts to manually check the automatic annotation of start/end codons of all genes, some end positions were not identified or were possibly missing (Table 2). The lack of complete termination codons in PCG's, resulting in a T- or TA- 3'-end, is a result of the necessity to compact the genome size to guarantee its faster replication, this is a common feature in many animal mitochondrial genes (Carapelli *et al.* 2014).

TABLE 2. Gene order and features of the mitochondrial genome of *Sinhomidia bicolor* (SB), *Alloscopus bannaensis* (AB), and *Coecobrya* sp. (COE). (*)End position not identified.

Gene	Size (bp)			Start/stop codons SB: AB : COE
	SB	AB	COE	
ND2	996	978	996	ATG/TAA: ATG/TAA : ATG/TAA
COI	1539	1539	1539	ATT/TAA: ATT/TAA : ATT/TAA
COII	684	684	687	ATT/TAG: ATT/TAA: ATC/TAA
ATP8	171	168	165	ATT/TAA: ATT/TAA : ATT/TAA
ATP6	681	681	681	ATG/TAA: ATG/TAA : ATG/TAA
COIII	804	789	804	ATG/TAG: ATG/TAA : ATG/TAA
ND3	336	336	336	ATT/TAG: ATA/TAA : ATT/TAG
ND5	1698	1318*	1698	ATT/TAA: ATT/* : ATT/TAA
ND4	1345	1345	1345	ATG/*: ATG/TAC : ATG/TAA
ND4L	256	250	258	ATT/TAA: TTA/TAA : TTG/TAA
ND6	475	470	475	ATA/TAA: ATC/TAA : ATT/TAA
CYTB	933*	1134	1134	ATG/*: ATG/TAA : ATG/TAA
ND1	907	899*	925	ATT/TAA: ATT/* : ATA/TAG

Maximum likelihood and Bayesian inferences recovered a similar topology, only the branch containing *A. bannaensis* is uncertain (Fig. 2). *S. bicolor* and *Coecobrya* sp. were recovered in the right position together with the other Entomobryinae (bootstrap/pp values = 100/1), but the phylogenetic relationship of *A. bannaensis* was doubtful, support values in both analyses were low (bootstrap/pp values = 50/0.51). In the ML analysis, *A. bannaensis* was placed as a sister-group of (Lepidocyrtinae + Seirinae), but the Bayesian inference result (Fig. 2) was more congruent to its current morphological classification. Orchesellidae is considered the basal family of Entomobryoidea (Zhang *et al.* 2019), but *A. bannaensis* should be placed in the same branch with *Orchesella villosa* (Linneus, 1767), not in a separate branch. This result was expected for Collembola group because mitogenomes generally are not ideal markers to solve relationships between its families or orders (Cucini *et al.* 2020, Godeiro *et al.* 2021, Sun *et al.* 2020).

Data availability statement

After publication, mitogenomes sequence data that support the findings of this study will be openly available in GenBank of NCBI at (<https://www.ncbi.nlm.nih.gov/>) under the accession numbers: OK037063-5, respectively to *Alloscopus bannaensis*, *Coecobrya* sp. and *Sinhomidia bicolor*. The associated BioProject number is PRJNA656897. SRA and Bio-Sample numbers are SRR15204679/SAMN20255313 for *Sinhomidia bicolor*, SRR15204682/SAMN20255310 for *Alloscopus bannaensis*, and SRR15204681/SAMN20255311 for *Coecobrya* sp.

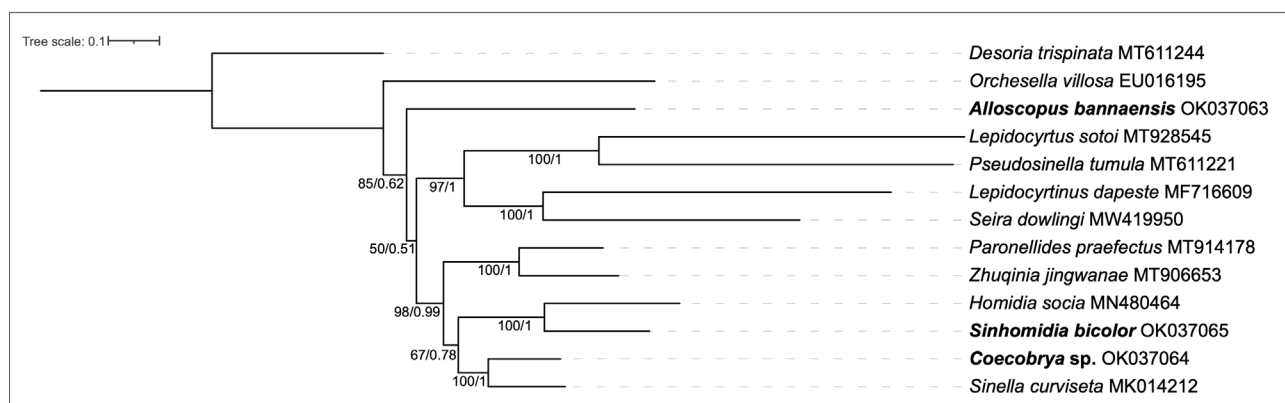


FIGURE 2. Phylogenetic placement of *Alloscopus bannaensis*, *Sinhomidia bicolor*, and *Coecobrya sp.* (in bold) based on Bayesian and maximum likelihood analyses. Numbers in the nodes are the bootstrap value and the posterior probability, respectively. GenBank accession numbers are listed after species names.

Funding

This research was financed by National Natural Science Foundation of China (31970434) and National Science & Technology Fundamental Resources Investigation Program of China (2018FY100300). NNG is currently funded by the Research Foundation of Shanghai Science and Technology Museum and Postdoctoral fund of Haibo Program of Pudong New Area in 2021.

Disclosure statement

The authors report no conflict of interest.

References

- Bellini, B.C., Cipola, N.G. & Godeiro, N.N. (2015) New species of *Lepidocyrtus* Bourlet and *Entomobrya* Rondani (Collembola: Entomobryoidea: Entomobryidae) from Brazil. *Zootaxa* 4027, 227–242.
<https://doi.org/10.11646/zootaxa.4027.2.3>
- Bellini, B.C., Santos, N.M.C., Souza, P.G.C. & Weiner, W.M. (2018) Two new species of Brazilian springtails (Hexapoda: Collembola) with comments on Neotropical *Brachystomella* Ågren and *Seira* (*Lepidocyrtinus*) Börner. *Insect Systematics and Evolution* 1–30.
<https://doi.org/10.1163/1876312X-00002192>
- Bellinger, P.F., Christiansen, K.A. & Janssens, F. (1996–2021) Checklist of the Collembola of the World. Available from: <http://www.collembola.org> (accessed 10 January 2022)
- Börner, C. (1906) Das System der Collembolen nebst Beschreibung neuer Collembolen des Hamburger Naturhistorischen Museums. *Mitteilungen aus den Naturhistorischen Museum in Hamburg* 23, 147–188.
- Brook, G. (1882) On a new genus of Collembola (*Sinella*) allied to *Degeeria* Nicolet. *Journal of the Linnean Society of London (Zoology)* 16, 541–545.
<https://doi.org/10.1111/j.1096-3642.1882.tb02398.x>
- Carapelli, A., Convey, P., Nardi, F. & Frati, F. (2014) The mitochondrial genome of the antarctic springtail *Folsomotoma octooculata* (Hexapoda: Collembola), and an update on the phylogeny of the collembolan lineages based on mitogenomic data. *Entomologia* 2, 46–55.
<https://doi.org/10.4081/entomologia.2014.190>
- Cucini, C., Fanciulli, P.P., Frati, F., Convey, P., Nardi, F. & Carapelli, A. (2020) Re-Evaluating the internal phylogenetic relationships of Collembola by means of mitogenome data. *Genes* 12.
<https://doi.org/10.3390/genes12010044>
- Denis, J.R. (1929) Sur deux Collemboles de Madagascar. *Bulletin du Muséum* 1, 104–106.
- Dierckxsens, N., Mardulyn, P. & Smits, G. (2020) Unraveling heteroplasmy patterns with NOVOPlasty. *NAR Genomics and*

- Godeiro, N.N., Pacheco, G., Liu, S., Cipola, N.G., Berbel-filho, W.M., Zhang, F., Gilbert, M.T.P. & Bellini, B.C. (2020) Phylogeny of Neotropical Seirinae (Collembola, Entomobryidae) based on mitochondrial genomes. *Zoologica Scripta* 49, 329–339.
<https://doi.org/10.1111/zsc.12408>
- Godeiro, N.N., Bellini, B.C., Ding, N., Xu, C., Ding, Y. & Zhang, F. (2021) A mitogenomic phylogeny of the Entomobryoidea (Collembola): A comparative perspective. *Zoologica scripta* 50, 658–666.
<https://doi.org/10.1111/zsc.12487>
- Handschin, E. (1925) Beiträge zur Collembolenfauna der Sundainseln. *Treubia* 6, 225–270.
<https://doi.org/10.1002/mmnd.48019250305>
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular biology and evolution* 35, 518–522.
<https://doi.org/10.1093/molbev/msx281>
- Lartillot, N., Rodrigues, N., Stubbs, D. & Richer, J. (2013) PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic biology* 62, 611–615.
<https://doi.org/10.1093/sysbio/syt022>
- Linnaeus, C. (1767) *Systema naturae, per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Tom. I. Pars II. Editio Duodecima Reformata. Stockholm, Sweden: Laurenti Salvii, 533–1327.
<https://doi.org/10.5962/bhl.title.156772>
- MacGillivray, A. (1896) The american species of *Isotoma*. *The Canadian Entomologist* 28(2), 47–58.
<https://doi.org/10.4039/Ent2847-2>
- Meng, G., Li, Y., Yang, C. & Liu, S. (2019) MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic acids research*, 1–7.
<https://doi.org/10.1093/nar/gkz173>
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. & Lanfear, R. (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution* 37, 1530–1534.
<https://doi.org/10.1093/molbev/msaa015>
- Nunes, R.C., Godeiro, N.N., Pacheco, G., Liu, S., Gilbert, M.T.P., Valin, F.A., Zhang, F. & Bellini, B.C. (2019) The discovery of Neotropical *Lepidosira* (Collembola, Entomobryidae) and its systematic position. *Zoologica Scripta* 48, 783–800.
<https://doi.org/10.1111/zsc.12377>
- Reumont von, B.M., Jenner, R.A., Wills, M.A., Dell’Ampio, E., Pass, G., Ebersberger, I., Meyer, B., Koenemann, S., Iliffe, T.M., Stamatakis, A., Niehuis, O., Meusemann, K. & Misof, B. (2012) Pancrustacean phylogeny in the light of new phylogenomic data: Support for remipedia as the possible sister group of hexapoda. *Molecular Biology and Evolution* 29, 1031–1045.
<https://doi.org/10.1093/molbev/msr270>
- 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. *PlosOne* 15, 1–21.
<https://doi.org/10.1371/journal.pone.0230827>
- Wang, F., Chen, J. & Christiansen, K. (2002) A new species of the subgenus *Coecobrya* (Collembola: Entomobryidae) from China. *Journal of Entomological Science* 37, 213–218.
<https://doi.org/10.18474/0749-8004-37.3.213>
- Wray, D. L. (1953) Some new species of springtail insects (Collembola). *Nature Notes, Occasional Papers* 1, 1–7.
- Yosii, R. (1956) Hohlencollembolen Japans II. *Japanese Journal of Zoology* 11, 609–627.
- Yosii, R. (1965) On some Collembola of Japan and adjacent countries. *Contributions from the Biological Laboratory Kyoto University* 19, 1–71.
- Zhang, F., Deharveng, L., Greenslade, P. & Chen, J. (2009) Revision of *Acanthocyrtus* (Collembola: Entomobryidae), with description of a new genus from eastern Asia. *Zoological Journal of the Linnean Society* 157, 495–514.
<https://doi.org/10.1111/j.1096-3642.2008.00521.x>
- Zhang, F., Ma, Y. & Greenslade, P. (2017) New Australian Paronellidae (Collembola) reveal anomalies in existing tribal diagnoses. *Invertebrate Systematics* 31, 375–393.
<https://doi.org/10.1071/IS16073>
- Zhang, F., Bellini, B.C. & Soto-Adames, F.N. (2019) New insights into the systematics of Entomobryoidea (Collembola: Entomobryomorpha): first instar chaetotaxy, homology and classification. *Zoological Systematics* 44, 249–278.
<https://doi.org/10.1186/zs.201926>
- Zhang, F., Cipola, N.G., Pan, Z. & Ding, Y. (2020) New insight into the systematics of *Heteromurini* (Collembola: Entomobryidae: Heteromurinae) with special reference to *Alloscopus* and *Sinodicranocentrus* gen. n. *Arthropod Systematics and Phylogeny* 78, 1–16.
<https://doi.org/10.26049/ASP78-1-2020-01>