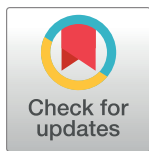


## RESEARCH ARTICLE

## Phylomitogenomic analyses on collembolan higher taxa with enhanced taxon sampling and discussion on method selection

Xin Sun<sup>1,2</sup>, Daoyuan Yu<sup>3\*</sup>, Zhijing Xie<sup>2,4</sup>, Jie Dong<sup>5</sup>, Yinhuan Ding<sup>5</sup>, Haifeng Yao<sup>2,4</sup>, Penelope Greenslade<sup>6,7</sup>

**1** J.F. Blumenbach Institute of Zoology and Anthropology, University of Göttingen, Göttingen, Germany, **2** Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China, **3** Soil Ecology Lab, College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing, China, **4** University of Chinese Academy of Sciences, Beijing, China, **5** Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing, China, **6** Environmental Management, School of Applied and Biomedical Science, Federation University, Ballarat, Victoria, Australia, **7** Division of Biology, Australian National University, Australian Capital Territory, Australia

\* [yudy@njau.edu.cn](mailto:yudy@njau.edu.cn), [dyyu1987@hotmail.com](mailto:dyyu1987@hotmail.com)

## OPEN ACCESS

**Citation:** Sun X, Yu D, Xie Z, Dong J, Ding Y, Yao H, et al. (2020) Phylomitogenomic analyses on collembolan higher taxa with enhanced taxon sampling and discussion on method selection. PLoS ONE 15(4): e0230827. <https://doi.org/10.1371/journal.pone.0230827>

**Editor:** Genlou Sun, Saint Mary's University, CANADA

**Received:** October 31, 2019

**Accepted:** March 9, 2020

**Published:** April 13, 2020

**Copyright:** © 2020 Sun et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All the new sequences will be available from the GenBank database (accession numbers are given in [Table 1](#), but sequences will not be available until acceptance of the manuscript). Furthermore, sequence data of this project has been deposited at the NCBI under the accession numbers MK431895, MK431900, MK431896, MK409685, MK423967, MK431894, MK431898, MK423966, MK409686, MK423968, MK431897, MK431893, MK423965, MK423969, MK423964, and MK431899.

## Abstract

Collembola are a basal group of Hexapoda renowned for both unique morphological characters and significant ecological roles. However, a robust and plausible phylogenetic relationship between its deeply divergent lineages has yet to be achieved. We carried out a mitophylogenomic study based on a so far the most comprehensive mitochondrial genome dataset. Our data matrix contained mitogenomes of 31 species from almost all major families of all four orders, with 16 mitogenomes newly sequenced and annotated. We compared the linear arrangements of genes along mitochondria across species. Then we conducted 13 analyses each under a different combination of character coding, partitioning scheme and heterotachy models, and assessed their performance in phylogenetic inference. Several hypothetical tree topologies were also tested. Mitogenomic structure comparison revealed that most species share the same gene order of putative ancestral pancrustacean pattern, while seven species from Onychiuridae, Poduridae and Symphypleona bear different levels of gene rearrangements, indicating phylogenetic signals. Tomoceroidea was robustly recovered for the first time in the presence of all its families and subfamilies. Monophyly of Onychiuroidea was supported using unpartitioned models alleviating LBA. Paronellidae was revealed polyphyletic with two subfamilies inserted independently into Entomobryidae. Although Entomobryomorpha has not been well supported, more than half of the analyses obtained convincing topologies by placing Tomoceroidea within or near remaining Entomobryomorpha. The relationship between elongate-shaped and spherical-shaped collembolans still remained ambiguous, but Neelipleona tend to occupy the basal position in most trees. This study showed that mitochondrial genomes could provide important information for reconstructing the relationships among Collembola when suitable analytical approaches are implemented. Of all the data refining and model selecting schemes used in this study, the combination of nucleotide sequences, partitioning model and

**Funding:** This work was supported by National Natural Science Foundation of China [grant numbers 41571052, 41971063, 41430857, 41811530086, 31861133006]; Alexander von Humboldt Foundation; and Youth Innovation Promotion Association, CAS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

exclusion of third codon positions performed better in generating more reliable tree topology and higher node supports than others.

## Introduction

Collembola (springtails) are numerically dominant microarthropods in most terrestrial environments, with more than 9000 species reported in the world [1]. They play roles in detrital food webs, and are important to soil ecosystems for modulating litter decomposition processes and forming soil microstructure [2–3]. However, regarding the basal phylogeny of Collembola, it has so far not been resolved. Generally, the basal classification of Collembola has two periods: (1) firstly two groups Arthropleona and Symphypleona *sensu lato* were recognised based on body shape and segmentation, the former contained species with generally elongated body shape and clearly divided body segments, while the later contained species with spherical body shape and fused segments [4–6]; (2) later the two-group system was replaced by a four-group system, in which the Arthropleona was divided into Poduromorpha and Entomobryomorpha mainly based on the status of prothoracic tergum, and Symphypleona *s. l.* was divided into Symphypleona *sensu stricto* and Neelipleona for the morphological differences in body and leg segments [1, 7–11]. To date, the second system has been widely accepted, but was seldom completely recovered by modern phylogenetic analyses, while the first system could also be completely or partially recovered in a few phylogenetic reconstructions.

The phylogenetic relationships within Collembola have been investigated by using either morphological characters or molecular markers in previous studies. D'Haese [9] performed a phylogenetic study based on D1 and D2 regions of 28S rDNA for 55 Collembola species. In his study, Entomobryomorpha and Symphypleona *s. s.*, as well as some families (e.g. Entomobryidae, Hypogastruridae, Neanuridae), were not monophyletic. In the following year, the same author conducted a morphological phylogenetic reconstruction based on 131 characters from 67 taxa [12]. In this study monophyly of four orders (Entomobryomorpha, Neelipleona, Poduromorpha, and Symphypleona) and most families were supported, except for Onychiuridae and Hypogastruridae. Later, several molecular studies based on partial or complete 18S and 28S rDNA also indicated the relationships within Collembola [13–17]. These studies often yielded similar results, but support for some taxa could change significantly depending on taxa sampled. Luan et al. [13] indicated monophyly of Arthropleona, Entomobryomorpha and Poduromorpha, and polyphyly of Symphypleona *s. s.* and Hypogastruridae. Similar results were repeated by Gao et al. [14], except that the later study supported monophyly of Symphypleona *s. s.* von Reumont et al. [16] supported the monophyly of Entomobryomorpha, Poduromorpha and Symphypleona *s. s.*, but rejected Arthropleona. Both Gao et al. [14] and von Reumont et al. [16] did not support Symphypleona *s. l.*, with Neelipleona separated from Symphypleona *s. s.* and occupying the basal branch of Collembola. However, the three aforementioned studies were focused on higher phylogeny of either basal Hexapoda [13–14] or whole Arthropoda [16], thus only limited collembolan species, including 10 in Luan et al. [13], 7 in Gao et al. [14] and 14 in von Reumont et al. [16]. In contrast, Xiong et al. [15] aimed to resolve phylogenetic relationships within Collembola and sampled 30 species belonging to 29 genera and 14 families. This study supported the monophyly of Poduromorpha, Symphypleona *s. s.* and Symphypleona *s. l.*, while Entomobryomorpha was not supported, with Tomoceroidea unexpectedly closer to Poduromorpha. With a further enriched sampling of 54 species, Yu et al. [17] emphasized the monophyly of Tomoceridae, separated Neelipleona from Symphypleona *s. s.*, and doubted again the monophyly of Tomoceroidea and its relationship with other

Entomobryomorpha. However, considering morphological evidence and limitation of taxon sampling in molecular study, the authors suggested keeping the current taxonomic system. With different gene markers (16S rDNA, COX1 and D1–D2 regions of 28S rDNA) and taxon sampling focused on Neelipleona and Symphypleona, Schneider et al. [18] and Schneider and D’Haese [19] revealed a topology different from most other studies, showing that Neelipleona was sister to Arthropleona, and Symphypleona was basal to them, while Entomobryidae and Isotomidae were both polyphyletic. Zhang et al. [20–21] reconstructed the phylogeny of Entomobryoidea with COX1, 16S, 18S and 28S, and found that the family Paronellidae was split into several independent groups included in Entomobryidae. A review of these previous studies shows that different gene markers, taxon samplings and analytical methods can lead to distinct tree topologies reflecting relationships among main groups of Collembola. Overall, the phylogenetic relationships within Collembola are still unresolved to a large extent, with the main controversial points as following: (1) monophylies of some families and superfamilies, e.g. Hypogastruridae, Paronellidae, Onychiuroidea, Tomoceroidea; (2) relationship between Tomoceroidea and other groups; (3) relationship between elongate-shaped and spherical-shaped groups.

The insect mitochondrial genome is usually a compact circular molecule typically 15–18 kb in size. It contains 37 genes: 13 protein-coding genes (PCGs) encoding subunits from four of the five mitochondrial electron-transport chain complexes, and two ribosomal RNA (rRNA) genes and 22 transfer RNA (tRNA) genes involved in the translation of the PCGs [22]. Mitogenomes have become popular multi-loci molecular markers for phylogenetic studies as having advantages of maternal hereditary transmission and accelerated nucleotide substitution rates compared with nuclear markers [22–25], and have recently received extensive use because improved sequencing techniques have made it more efficiency and less costly to obtain the complete sequences [26]. Further, the variation in mitogenomes may be related to the adaption of environmental conditions [27–29]. However, the power of mitogenomes in phylogenetic reconstructions is still controversial. On one hand, genomic data can be organized into different datasets, particularly PCGs can be coded in the forms of nucleotide sequences (all codon positions or first two positions only), codons or amino acids, while the performances of different character coding are not unified [30–31]. On the other hand, heterotachous evolutionary processes of mitogenome sequences are known to mislead phylogenetic inference, and various site-heterogeneous models, such as CAT (classifies sites into categories, [32]) and GHOST (general heterogeneous evolution on a single topology, [33]), have been explored to improve the reliability of results. Moreover, partitioning of alignments is also a common approach to incorporating the heterogeneity, however, its positive effects on tree topology, branch-lengths and bootstrap supports are still debated [34]. As a result, various character coding, partitioning schemes and models have been extensively applied to reconstruct the mitochondrial phylogenetic trees in many insect groups [22], e.g. Hemiptera [35], Heteroptera [36], Psocodea [37], Hymenoptera [38–39], Coleoptera [40], and Lepidoptera [41]. Mitogenomic analyses have also been implemented to solve some phylogenetic and taxonomic problems of Collembola [42–43]. A recent study has reported so far most comprehensive phylogenetic analysis of Collembola based on mitogenomes, with 11 families sampled [44]. In the context of their taxon sampling, they revealed the structural diversity of collembolan mitogenomes, recovered the monophyly of all four orders as well as six families, and contributed to the time frame of collembolan evolution. However, the effectiveness of various alternative analytical approaches has not been sufficiently evaluated, and the position of several significant and most problematic taxa, e.g. Tomoceridae, Oncopoduridae and Paronellidae, remains pending further assessment.

To further understand the phylogeny of Collembola, we have newly sequenced and annotated the mitogenomes of 16 collembolan species. Firstly, we identified and compared the

alternative gene arrangements observed along the mtDNA of different taxa. Then, with the enlarged dataset, we performed a range of analyses based on different character coding, partitioning and site-heterogeneous models to address the phylogenetic problems within Collembola mentioned above. Finally, we compared different analytical approaches and assessed the effectiveness of mitogenomic analyses for reconstructing the phylogeny of Collembola.

## Materials and methods

### Taxon sampling

A total of 31 species representing main groups of all four orders of Collembola were selected for the phylogenetic reconstructions (Table 1 and S1 Table). Five families and nine subfamilies were newly introduced to mitophylogenomic analyses in this study, including elongate-shaped Paronellidae (Salininae and Paronellinae), Oncopoduridae, Tomoceridae (Lepidophorellinae and Tomocerinae), Pachyotominae, Heteromurinae and Entomobryinae, and spherical-shaped Katiannidae, Sminthurididae, Sphyrothecinae and Ptenothricinae. Complete and partial mitogenomes of 16 species were originally sequenced in this study, while data of the other 15 Collembola and three outgroups (two Diplura and one Microcoryphia species) were retrieved from the NCBI database. For newly sequenced species, specimens were collected with aspirators or Berlese funnels and preserved in 99% ethanol before morphological examination and DNA extraction. All specimens were morphologically identified to species level before DNA extraction. No permits were required for our collection.

### DNA extraction, amplification and sequencing

DNA was extracted using a QIAamp DNA Micro Kit (QIAGEN GmbH, Shanghai, China). Extractions were performed non-destructively for further morphological examination and identification of the specimens. DNA concentration was measured by Qubit 3.0 using Q33230 Qubit™ 1X dsDNA HS Assay Kit. Mitogenome amplification for less than 50 ng DNA was performed using REPLI-g Single Cell Kit. Each library was sequenced with an insert size of 350 bp on HiSeq X Ten platform (Tianjin Novogene Bioinformatics Technology Co., Ltd, China) generating 150 bp paired-end reads.

The *COX1* was used as a seed sequence for our mitochondrial assembly. Primers for *COX1* were LCO1490/HCO2198 commonly used for Metazoa [45]. Amplification volume and procedure followed Zhang et al. [46]. All PCR products were checked on a 1% agarose gel. Successful products were purified and sequenced in both directions by Tsingke (Beijing, China) on ABI 3730XL DNA Analyser (Applied Biosystems). Sequences were assembled in Sequencher 4.5 (Gene Codes Corporation, Ann Arbor, USA), blasted in GenBank and checked for possible errors, then were preliminarily aligned using MEGA 7.0 [47]. Alignments were checked and corrected manually, with a final 658 bps alignment.

### Mitogenome assembly and annotation

All mitogenomes were assembled with NOVOPlasty v2.7.0 [48] using *COX1* sequence as the initial seed. Mitochondrial gene annotations were performed using MITOS web server [49] and tRNAs gene limits were rechecked with tRNAscan-SE [50]. The mtDNA sequences were deposited in GenBank (Table 1 and S1 Table).

### Phylogenetic analysis

Alignments of 16S rRNA (*rrnL*), 12S rRNA (*rrnS*) and amino-acid sequences of each PCG were conducted by MAFFT v.7.394 [51] with an accurate option L-INS-I, and then an

Table 1. Information of collembolan species used in this study.

GenBank nos.	Order	Family	Subfamily	Species	PCGs	Genes	Size (bp)
MK014212	Entomobryomorpha	Entomobryidae	Entomobryinae	<i>Sinella curviseta</i> Brook, 1882	13	37	14,840
MK431895	Entomobryomorpha	Entomobryidae	Heteromurinae	<i>Dicranocentrus wangi</i> Ma & Chen, 2007	13	37	14,883
MK431900	Entomobryomorpha	Entomobryidae	Lepidocyrtinae	<i>Lepidocyrtus fimetarius</i> Gisin, 1964	13	37	14,698
KT985987	Entomobryomorpha	Entomobryidae	Orchesellinae	<i>Orchesella cincta</i> Linnæus, 1758	13	37	15,728
EU016195	Entomobryomorpha	Entomobryidae	Orchesellinae	<i>Orchesella villosa</i> von Linné, 1767	13	37	14,924
MK431896	Entomobryomorpha	Paronellidae	Paronellinae	<i>Cyphoderus albinus</i> Nicolet, 1842	13	37	14,836
MK409685	Entomobryomorpha	Paronellidae	Salininae	<i>Salina celebensis</i> (Schäffer, 1898)	13	37	14,788
NC_010533	Entomobryomorpha	Isotomidae	Anurophorinae	<i>Cryptopygus antarcticus</i> Willem, 1901	13	37	15,297
KX863671	Entomobryomorpha	Isotomidae	Anurophorinae	<i>Cryptopygus terranovus</i> (Wise, 1967)	13	37	15,352
KU198392	Entomobryomorpha	Isotomidae	Anurophorinae	<i>Folsomia candida</i> Willem, 1902	13	37	15,147
NC_024155	Entomobryomorpha	Isotomidae	Isotominae	<i>Folsomotoma octooculata</i> (Willem, 1901)	13	37	15,338
MK423967	Entomobryomorpha	Isotomidae	Pachytominae	<i>Paranurophorus simplex</i> Denis, 1929	11	28	9,518
MK431894	Entomobryomorpha	Oncopoduridae		<i>Oncopodura yosii</i> Szeptycki, 1977	13	37	14,808
MK431898	Entomobryomorpha	Tomoceridae	Lepidophorellinae	<i>Novacerus tasmanicus</i> (Womersley, 1937)	13	36	15,518
MK423966	Entomobryomorpha	Tomoceridae	Tomocerinae	<i>Tomocerus qinae</i> Yu, Yao & Hu, 2016	13	37	15,045
MK409686	Poduromorpha	Hypogastruridae		<i>Ceratophysella communis</i> (Folsom, 1898)	13	37	15,331
AY191995	Poduromorpha	Hypogastruridae		<i>Gomphiocephalus hodgsoni</i> Carpenter, 1908	13	37	15,075
EU084034	Poduromorpha	Neanuridae	Neanurinae	<i>Bilobella aurantiaca</i> (Caroli, 1912)	13	37	16,312
EU124719	Poduromorpha	Neanuridae	Frieseinae	<i>Friesea grisea</i> (Schäffer, 1891)	13	37	15,442
NC_006074	Poduromorpha	Onychiuridae	Onychiurinae	<i>Thalassaphorura orientalis</i> Stach, 1964	13	34	12,984
MK423968	Poduromorpha	Onychiuridae	Onychiurinae	<i>Thalassaphorura encarpata</i> (Denis, 1931)	13	37	15,213
NC_002735	Poduromorpha	Onychiuridae	Tetrodontophorinae	<i>Tetrodontophora bielensis</i> (Waga, 1842)	13	37	15,455
MK431897	Poduromorpha	Tullbergiidae	Mesaphorurinae	<i>Mesaphorura yosii</i> (Rusek, 1967)	13	37	14,833
NC_006075	Poduromorpha	Poduridae		<i>Podura aquatica</i> Linnæus, 1758	13	34	13,809
MK431893	Neelipleona	Neelidae		<i>Neelides</i> sp.	13	34	13,858
MK423965	Symphyleona	Dicyrtomidae	Ptenothricinae	<i>Ptenothrix huangshanensis</i> Chen & Christiansen, 1996	13	37	15,152
MK423969	Symphyleona	Katiannidae		<i>Sminthurinus signatus</i> (Krausbauer, 1898)	7	20	5,459
MK423964	Symphyleona	Sminthurididae		<i>Sminthurides bifidus</i> Mills, 1934	13	35	14,161
KY618680	Symphyleona	Bourletiellidae		<i>Bourletiella arvalis</i> (Fitch, 1862)	13	37	14,794
NC_010536	Symphyleona	Sminthuridae	Sminthurinae	<i>Sminthurus viridis</i> (Linnæus, 1758)	13	37	14,817
MK431899	Symphyleona	Sminthuridae	Sphyrothecinae	<i>Lipothrix lubbocki</i> (Tullberg, 1872)	13	37	15,141

<https://doi.org/10.1371/journal.pone.0230827.t001>

automated alignment trimming (-automated1) was performed by trimAL v.1.4 [52] for removing gap-only and ambiguous-only positions. After that, we generated codon-based nucleotide sequence alignments of 13 PCGs by trimAL v.1.4 with the option -backtrans based on trimmed amino-acid sequences and unaligned nucleotide sequences of each gene. The final concatenated supermatrices were performed by FASconCAT-G v1.04 [53] as: (A1) nucleotide sequences of 13 PCGs (13fna), (A2) nucleotide sequences of 13 PCGs and two rRNA genes (15fna), (A3) amino acid sequences of 13 PCGs (13faa).

For PCGs in the supermatrices A1 and A2, we analyzed nucleotide sequences by using either all codon positions or excluding third codon positions (site 1+2) to evaluate the effect of saturation of the third codon positions of nucleotide substitutions. All partition and substitution models on supermatrices A1, A2 and A3 were selected using ModelFinder [54]. In addition, we also inferred trees from supermatrix A1 under the codon model (CODON5). Considering the heterogeneous evolution, GHOST model were applied in all supermatrices with nucleotide- (GHOST\_GTR), codon- (GHOST\_GY), and amino acid- (GHOST\_LG and GHOST\_mtART) models, respectively. Finally, we constructed trees from supermatrix A3

under posterior mean site frequency (PMSF) model [55], a variant of PhyloBayes' CAT model, with 20 amino-acid profile categories under both LG and mtART exchange rate matrices (option: -m LG+C20+F+G and -m mtART+C20+F+G). Therefore, in total 13 analyses were performed as follows: (1) 13 *fna\_GHOST\_GTR*; (2) 13 *fna\_CODON5\_GHOST\_GY*; (3) 13 *fna\_13partition*; (4) 13 *fna\_13partition\_CODON5*; (5) 13 *fna\_13partition\_site1+2*; (6) 15 *fna\_GHOST\_GTR*; (7) 15 *fna\_15partition*; (8) 15 *fna\_13PCGs (site1+2)\_2rrn (2partition)*; (9) 13 *faa\_GHOST\_LG*; (10) 13 *faa\_GHOST\_mtART*; (11) 13 *faa\_PMSF\_LG*; (12) 13 *faa\_PMSF\_mtART*; (13) 13 *faa\_13partition*. The maximum-likelihood (ML) trees were inferred and ultrafast bootstrap [56] with 1,000 replicates were performed in IQ-TREE v1.6.3 [57]. Nodes with a bootstrap value of minimum 95 were considered well-supported in the analyses.

### Tree topology comparison

Tree topology comparison on constraining monophyly was performed by using the RELL approximation method [58]. Seven hypotheses were proposed based on current ambiguities of collembolan phylogeny: (A) best tree without any constrains; (B) Tomoceridae + Oncopoduridae; (C) Neelipleona + Symphypleona s. s.; (D) Tomoceroidea + (Entomobryoidea + Isotomidae); (E) Tomoceroidea + Poduromorpha; (F) Tullbergiidae + Onychiuridae; (G) Symphypleona s. s. + Entomobryomorpha. Approximately unbiased (AU) test [59], bootstrap proportion (BP, [58]), expected likelihood weight (ELW, [60]), Kishino-Hasegawa (KH) test [61], Shimodaira-Hasegawa (SH) test [62], weighted KH (WKH) and weighted SH (WSH) tests were performed in IQ-TREE v.1.6.3 with the options: -au, -zb and -zw. The number of RELL replicates was specified to 10,000.

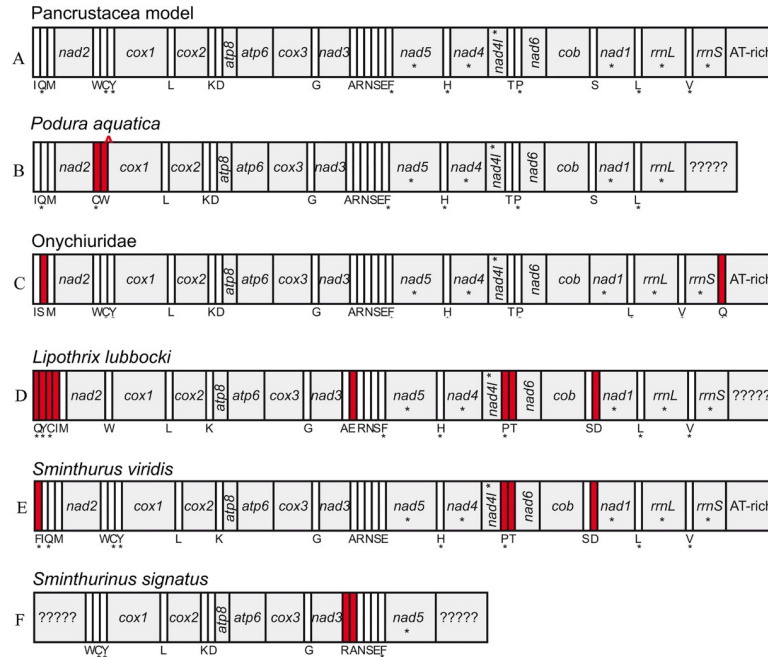
## Results

### Characteristics of collembolan mitochondrial genomes

Complete and partial mitogenomes newly sequenced range from 5,459 bp to 16,312 bp in size (Table 1). The mitogenomes of 24 species possessed the putative ancestral Pancrustacea gene order, while gene rearrangements were observed in seven species with deletion, inversion or translocation of tRNA genes (Fig 1, S2 Table). *Podura aquatica* Linnæus, 1758 (Poduridae) has the deletion of *trnF* between *trnC* and *cox1*, and the inversion between *trnC* and *trnW*. Three species of Onychiuridae have translocation of *trnS (uga)* and *trnQ*. Three species of Symphypleona have different gene rearrangements. *Lipothrix lubbocki* (Tullberg, 1872) (Sminthuridae: Sphyrrothecinae) have the highest level of tRNA rearrangements, including inversions between *trnQ* and *trnI*, between *trnP* and *trnT*, and between *trnC* and *trnY*, and translocations of *trnC*, *trnY*, *trnD*, and *trnE*. *Sminthurus viridis* (Linnæus, 1758) (Sminthuridae: Sminthurinae) have the inversion between *trnP* and *trnT*, and the translocation of *trnD* and *trnF*. *Sminthurinus signatus* (Krausbauer, 1898) (Katiannidae) have the inversion between *trnR* and *trnA*.

### Phylogenetic inference

Among all 13 analyses using different character coding, partitioning scheme and heterotachy models, the combination of 15 *fna\_13PCGs (site1+2)\_2rrn (2partition)*, using nucleotide sequences of 15 genes, partitioning model and exclusion of third codon positions, performed better in generating more plausible tree topology and higher node supports, (Fig 2), which is generally in congruence with external sources of evidence such as morphology and/or other molecular studies (see Discussion Section for details). In this best-resolved tree, the monophyly of most well-defined families and subfamilies is recovered with high support values (support > 80), with the exception of Hypogastruridae, Paronellidae and Entomobryidae.



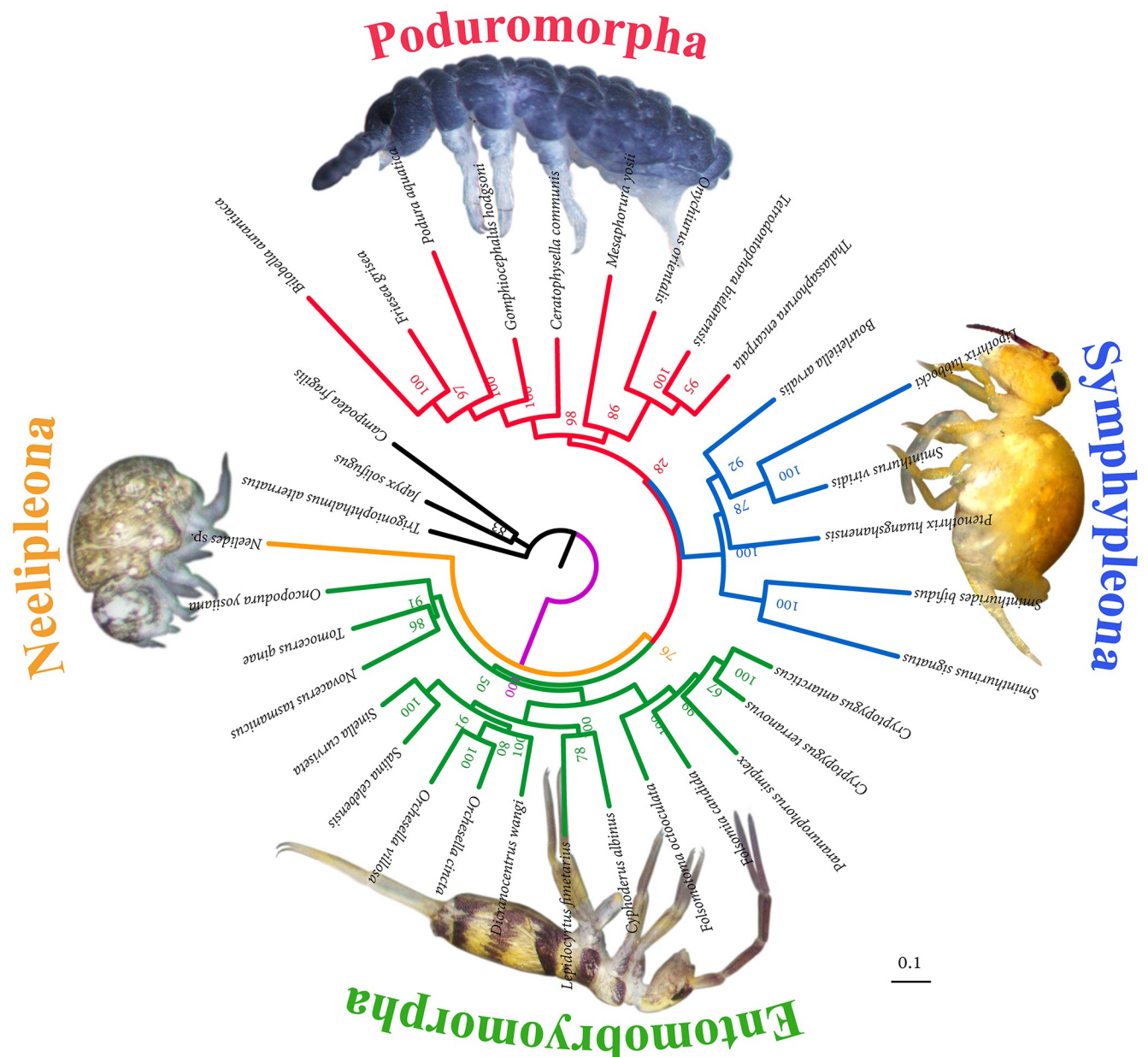
**Fig 1. Comparison of linear arrangements of genes on mitochondrial genomes of collembolan species in this study.** A. Pancrustacea (putative ancestral) model in most species. B. *Podura aquatica*. C. Onychiuridae. D. *Lipothrix lubbocki*. E. *Sminthurus viridis*. F. *Sminthurinus signatus*. The transcriptional direction of mitochondrial genes is from left to right, the asterisk (\*) below the genes indicates the opposite direction, the question marks indicates the unknown genes, the rearrangements are marked in red, the inverted triangle indicates the deletion of gene.

<https://doi.org/10.1371/journal.pone.0230827.g001>

Within Hypogastruridae, *Gomphiocephalus hodgsoni* is sister to the clade of Neanuridae + Poduridae, and *Ceratophysella communis* is further sister to them. Two species of Paroneliidae are included in Entomobryidae. Moreover, all three orders represented by more than one species are recovered, despite that Entomobryomorpha does not receive high support (support = 50). Thus this result is generally satisfactory because it conforms well with currently accepted classification system of Collembola based on morphological examination (see Discussion Section for details). Nevertheless, considering that the controversial phylogenetic relationship within Collembola has not been settled to a large extent, the incongruence between analyses may provide alternative valuable hypotheses which are to be tested by further studies, therefore, results of all analyses are thoroughly listed, compared and discussed below.

The phylogenetic trees calculated from all supermatrices (A1, A2 and A3) strongly supported the monophyly of Symphypleona s. s. Symphypleona s. s. clustered with Entomobryoidea + Isotomidae by analyses of 13fna\_GHOST\_GTR, 13fna\_13partition, 15fna\_GHOST\_GTR and 15fna\_15partition, however, with low to medium supports (33.1, 84, 34.5, 79, respectively). The basal position of Neelipleona was strongly supported in most analyses, except 13faa\_GHOST\_mtART and 13faa\_PMSF\_LG clustered Symphypleona s. s. and Neelipleona together (support = 73.9 and 76.1, respectively).

Poduromorpha was supported in most analyses, but Tullbergiidae was clustered with Neelipleona in 13fna\_GHOST\_GTR (support = 78.1), 13fna\_CODON5\_GHOST\_GY (support = 69.5), 13faa\_GHOST\_LG (support = 68.3) and 13faa\_PMSF\_mtART (support = 95.2). In eight other analyses, Tullbergiidae was clustered with Onychiuridae, and six of them were well supported (support > 80). The monophyly of Onychiuridae was well supported in all analyses, but the subfamily Onychiurinae is not monophyletic. Neanuridae was recovered in most



**Fig 2. Maximum likelihood phylogenetic tree inferred from partitioned nucleotide sequences of 15 genes, with third codon position excluded (15fna\_13PCGs (site1+2)\_2rrn (2partition)).** Bootstrap support values are shown in the nodes.

<https://doi.org/10.1371/journal.pone.0230827.g002>

analyses, with Poduridae as sister group, and basal to them were two species of Hypogastruriidae usually forming a paraphyletic group.

The status of Entomobryomorpha was highly incongruent between analyses. Entomobryoidae and Isotomidae were both recovered and clustered together with high supports in all analyses (support > 95). Tomoceroidea was highly supported in most analyses (support > 90 in 10 analyses), and it was clustered with Entomobryoidae + Isotomidae, forming Entomobryomorpha in 13fna\_CODON5\_GHOST\_GY (support = 49.8), 13fna\_13partition\_CODON5



(support = 30.9), 13fna\_13partition\_site1+2 (support = 51), 15fna\_13PCGs(site1+2)\_2rrn (2partition) (support = 50) and 13faa\_13partition(support = 29), but with only low supports. In 13fna\_GHOST\_GTR and 15fna\_GHOST\_GTR, Tomoceroidea was basal to the branch containing Symphypleona + (Entomobryoidea + Isotomidae) (support = 72.9 and 41.2, respectively), while in the other six analyses it was clustered with Poduromorpha. Monophyly of Tomoceridae was recovered in eight analyses (with high support except in 13faa\_GHOST\_LG with support = 18.5), while 13fna\_GHOST\_GTR, 13fna\_13partition, 15fna\_GHOST\_GTR and 15fna\_15partition clustered Tomocerinae with Oncopoduridae (support = 31.8, 83, 27.6 and 86, respectively), and 13fna\_CODON5\_GHOST\_GY clustered Lepidophorellinae with Oncopoduridae (support = 9.6). Paronellidae was polyphyletic in all analyses, with Salininae always clustered with Entomobryinae, and Paronellinae often clustered with Lepidocyrtinae.

Support values for major clades in all analyses were shown in [Table 2](#). Results of two analyses were selected to represent main variations in tree topologies (Figs 3 and 4), while results of other analyses were deposited in [S1 Fig](#).

### Tree topology tests

Most hypotheses were not rejected by tree topology tests in most analyses ([Table 3](#) and [S3 Table](#)). However, in 13fna\_13partition\_CODON5 most hypotheses were significantly rejected except for the best tree without any constraints. In addition, hypothesis (C) Neelipleona + Symphypleona s. s. was significantly rejected in the analyses of 15fna\_GHOST\_GTR and 15fna\_15partition.

Results of the topology tests for two selected trees were shown in [Table 3](#), while the others were deposited in [S3 Table](#).

## Discussion

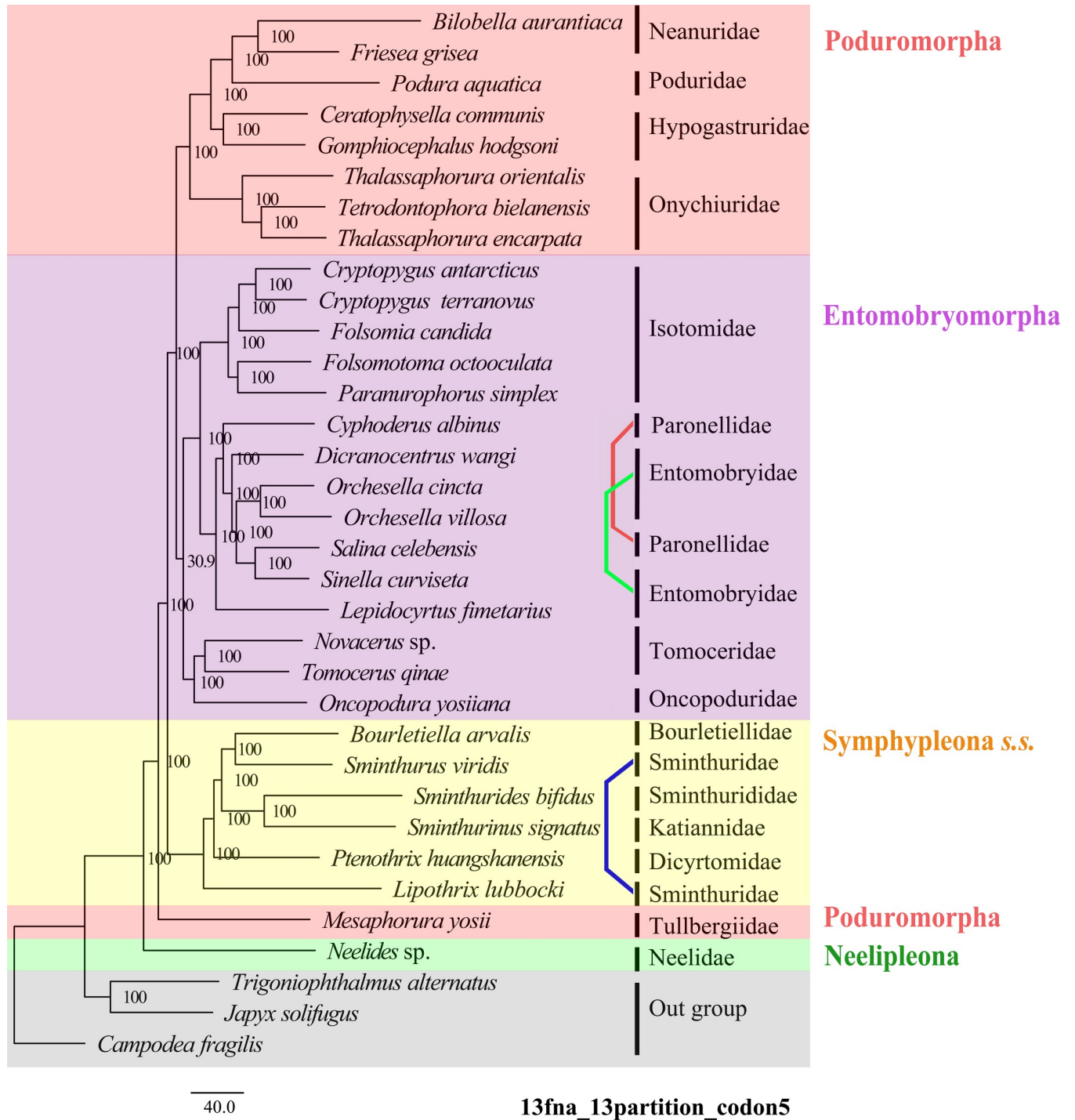
### Mitochondrial gene rearrangements

Similar to nucleotide and amino acid sequences, the arrangements of genes along mitochondrial chromosome may also reflect evolutionary history and phylogenetic relationships [63–64]. Among 31 species in our study, seven species of four families, including four species of Poduromorpha (three Onychiuridae and one Poduridae) and three species of Symphypleona (two Sminthuridae and one Katiannidae), have different mitochondrial gene arrangements. All these gene rearrangements occur in tRNA genes but none in PCGs, as also commonly found in insects [22]. Synapomorphic gene rearrangements were observed on the familial level, which is in line with previous studies in Collembola and many groups of insects [22, 44]. Three species of Onychiuridae have identical translocation of the same genes as a synapomorphy. Besides, the partial sequence of mitogenome of *Tullbergia mixta* Wahlgren, 1906 (Tullbergiidae: Tullbergiinae), also has the same translocation of *trnS* (*uga*) (GenBank accession number KF982833.1), however, no such rearrangement was found in another tullbergiid species, *Mesaphorura yosii* Rusek, 1967 from another subfamily (Mesaphorurinae). Similarly, two species from two subfamilies of Sminthuridae share the translocation of *trnD* and inversion between *trnP* and *trnT* as synapomorphies, but have different rearrangements for other genes. *Podura aquatica* lives almost exclusively on the surface of freshwater bodies, thus is ecologically remote from closely related neanurids and hypogastrurids which are mostly soil dwellers. Our results suggest the rearrangement of mitochondrial genes is taxon-specific and may reflect certain evolutionary events in Collembola. However, most families are represented by only a few species in available mitogenome databases, thus a thorough comparative study is not feasible until more species have been sequenced.

**Table 2. Support values for the major clades recovered by different datasets and analytical approaches. “N” means this was not supported.**

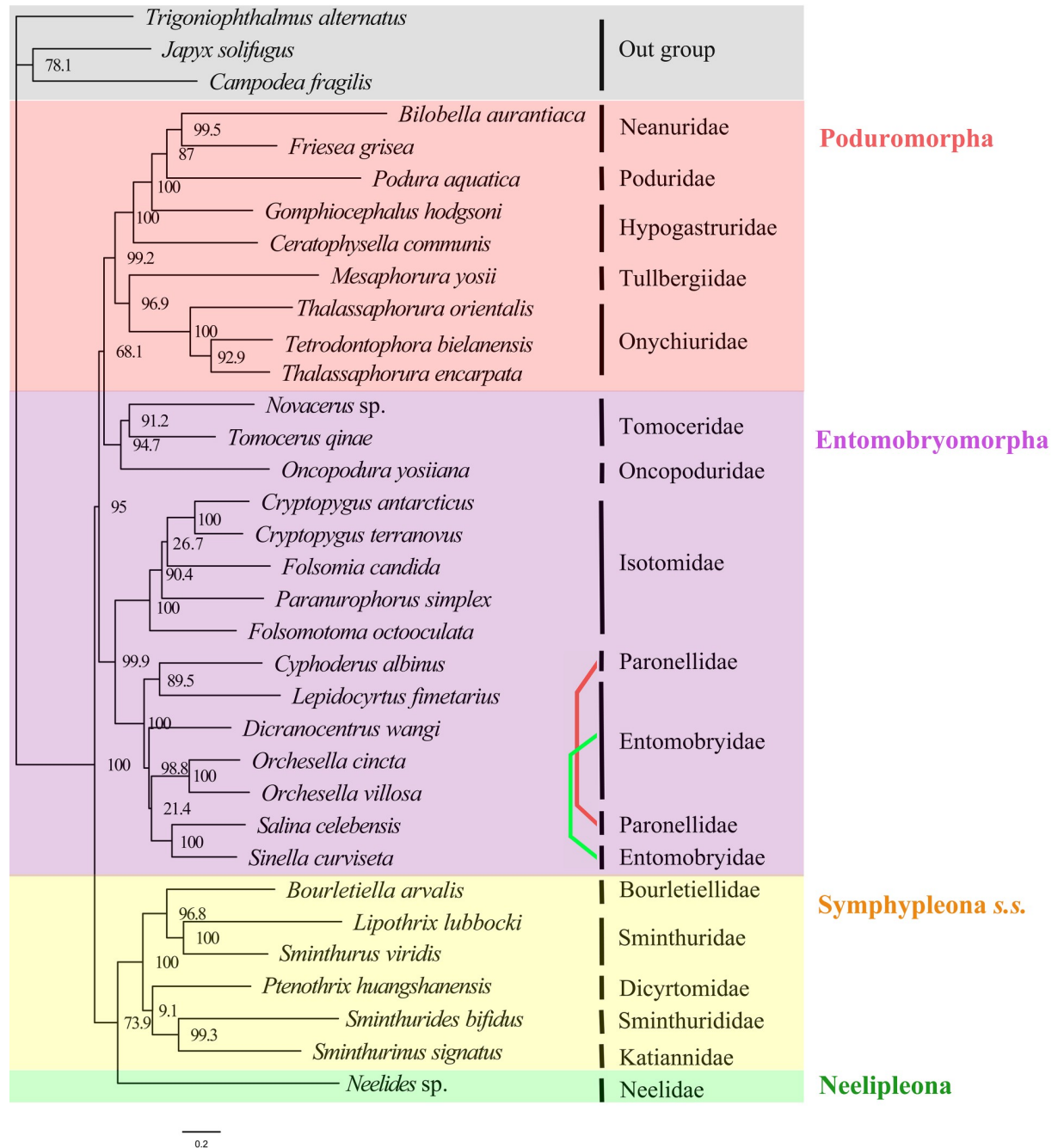
	13fna				15fna				13faa				
	GHOST_GTR	CODON5_GHOST_GY	13partition	13partition_CODON5	13partition_site1+2	GHOST_GTR	15partition	I3PCGs(site1+2)_2rn (2partition)	GHOST_LG	GHOST_mtART	PMSF_LG	PMSF_mtART	I3partition
Poduromorpha	N	N	83.0	N	95.0	99.9	95.0	98.0	N	99.2	98.9	N	51.0
Symphyleona	100.0	100.0	100.0	100.0	100.0	100.0	97.0	100.0	100.0	100.0	100.0	100.0	100.0
Entomobryomorpha	N	49.8	N	30.9	51.0	N	N	50.0	N	N	N	N	29.0
Symphyleona s. l.	N	N	N	N	N	N	N	N	N	73.9	76.1	N	N
Neelipleona at base	100.0	98.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	N	N	24.0	100.0
Tomocerioidea + Poduromorpha	N	N	72.0	N	N	N	68.0	N	21.5	68.1	91.1	78.4	N
Tullbergiidae + Onychiuridae	N	N	82.0	N	96.0	98.0	93.0	98.0	N	96.9	31.4	N	52.0
Entomobryoidae + Isotomidae	100.0	100.0	100.0	100.0	100.0	100.0	97.0	100.0	100.0	99.9	100.0	100.0	100.0
Symphyleona + (Entomobryoidae + Isotomidae)	33.1	N	84.0	N	N	34.5	79.0	N	N	N	N	N	N
Poduridae + Neanuridae	82.0	99.6	87.0	100.0	96.0	72.2	91.0	97.0	98.5	87.0	87.9	74.0	N
Tomoceridae	N	N	N	100.0	81.0	N	N	86.0	18.5	91.2	85.1	88.5	86.0
Hypogastruridae	N	N	N	100.0	N	N	N	N	N	N	N	N	N
Paronellidae	N	N	N	N	N	N	N	N	N	N	N	N	N
Sminthuridae	100.0	100.0	100.0	N	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Onychiurinae	N	N	77.0	N	N	N	N	N	N	N	N	N	N

<https://doi.org/10.1371/journal.pone.0230827.t002>



**Fig 3. Maximum likelihood phylogenetic tree inferred from partitioned nucleotide sequences of 13 PCGs under codon 5 model (13fna\_13partition\_CODON5).** Bootstrap support values are shown in the nodes. Each coloured line linked a same taxon in current taxonomic system but separated in different branches.

<https://doi.org/10.1371/journal.pone.0230827.g003>



**Fig 4. Maximum likelihood phylogenetic tree inferred from amino acid sequences of 13 PCGs under GHOST\_mtART model (13faa\_GHOST\_mtART).** Bootstrap support values are shown in the nodes. Each coloured line linked a same taxon in current taxonomic system separated in different branches.

<https://doi.org/10.1371/journal.pone.0230827.g004>

### Basal phylogeny of Collembola

Of the four orders of Collembola, one, the Symphypleona s. s., was strongly supported by our results, which is in line with most previous studies based on morphological and multi-locus molecular analyses [12, 15, 17]. This order is well characterized by the spherical body shape and can be distinguished from another spherical-shaped order Neelipleona by numerous

**Table 3. Results of tree topology tests of the analyses 13fna\_13partition\_CODON5 and 13faa\_GHOST\_mtART.** Seven hypothetical topologies were (A) best tree without any constrains; (B) Tomoceridae + Oncopoduridae; (C) Neelipleona + Symphypleona s. s.; (D) Tomoceroidea + (Entomobryoidea + Isotomidae); (E) Tomoceroidea + Poduromorpha; (F) Tullbergiidae + Onychiuridae; (G) Symphypleona s. s. + Entomobryomorpha.

Hypotheses	logL	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
<b>13fna_13partition_CODON5</b>									
A	-226076,295	0,000	0.9684 +	0.9710 +	1.0000 +	0.9710 +	0.9997 +	0.9684 +	0.9818 +
B	-226439,654	363,359	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -
C	-226591,485	515,191	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -
D	-226159,904	83,610	0.0291 -	0.029 -	0.2168 +	0.0290 -	0.1332 +	0.0291 -	0.0295 -
E	-226325,936	249,641	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -
F	-226685,661	609,366	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -
G	-226236,015	159,721	0.0025 -	0.0035 -	0.0216 -	0.0035 -	0.0156 -	0.0025 -	0.0029 -
<b>13faa_GHOST_mtART</b>									
A	-116116,632	0,047	0.0000 -	0.4240 +	0.8651 +	0.4240 +	0.9231 +	0.0294 +	0.5346 +
B	-116116,586	0,001	0.0018 -	0.4468 +	0.9604 +	0.4468 +	0.9457 +	0.0307 +	0.7470 +
C	-116116,639	0,054	0.0093 -	0.4160 +	0.8695 +	0.3319 +	0.9636 +	0.0292 +	0.7458 +
D	-116116,585	0,000	0.0196 +	0.5532 +	1.0000 +	0.5772 +	0.9632 +	0.0307 +	0.7241 +
E	-116116,587	0,002	0.0374 +	0.3935 +	0.9447 +	0.3935 +	0.9060 +	0.0307 +	0.7467 +
F	-116116,632	0,047	0.0648 +	0.4228 +	0.8619 +	0.4228 +	0.9494 +	0.0294 +	0.6709 +
G	-116139,142	22,557	0.0144 +	0.1280 +	0.1313 +	0.1280 +	0.3050 +	0.0226 +	0.1397 +

deltaL: logL difference from the maximal logl in the set. bp-RELL: bootstrap proportion using REll method [54]. p-KH: p-value of one sided Kishino-Hasegawa test [57]. p-SH: p-value of Shimodaira-Hasegawa test [58]. p-WKH: p-value of weighted KH test. p-WSH: p-value of weighted SH test. c-ELW: Expected Likelihood Weight [56]. p-AU: p-value of approximately unbiased (AU) test [55]. Plus signs denote the 95% confidence sets. Minus signs denote significant exclusion. All tests performed 10000 resamplings using the REll method.

<https://doi.org/10.1371/journal.pone.0230827.t003>

characters such as the ratio between thoracic and abdominal segments, status of sensory organs on body and length of antennae [65]. Only two analyses clustered *Symphypleona s. s.* and *Neelipleona* together, while other analyses all indicated *Symphypleona s. s.* is more related to non-spherical groups (*Arthropleona*), and *Neelipleona* is basal to them. Additionally, the topology test also rejected *Symphypleona s. l.* hypothesis for the trees based on 15 nucleotide sequences using either partitioning or GHOST model. Four analyses have retrieved the topology of (*Entomobryoidea* + *Isotomidae*) + *Symphypleona s. s.*, which is in line with the result of Leo et al. [44], and is supported by their morphological similarities such as reduction of prothoracic tergum, presence of abdominal bothriotracha and chaetotaxy on legs. But this topology was not strongly supported in our analyses. The ambiguity may be caused by the lack of intermediate forms between spherical and elongated groups in our dataset. Compared to the five families of *Symphypleona s. s.* included in this study, a rarely discovered but widespread family *Mackenziellidae*, has relatively elongated body shape and less fused body segments, which is considered probably more primitive in this order [66] (although also speculated as secondarily derived from globular ancestor by Fjellberg [67]), and is expected to be able to enhance the link between *Symphypleona s. s.* and other *Collembola* in further studies.

*Poduromorpha* was well supported by our analyses, except the unexpected position of *Tullbergiidae* in a few analyses. The unusual grouping of *Tullbergiidae* with *Neelipleona* was most probably caused by long branch attraction (LBA) [68–69]. In morphology, the *Tullbergiidae* is far remote from *Neelipleona* but closest to *Onychiuridae*, which was supported by six of our analyses with high support value (Table 2). However, the sister relationship between *Onychiuridae* and *Tullbergiidae* is still controversial. Similarly, Leo et al. also recovered different

topology for the position of Tullbergiidae by using datasets either including or excluding the third codon positions [44]. Resolving this problem may require further assessment involving other assumed members of the superfamily Onychiuroidea, such as Odontellidae Massoud, 1967 and Pachytullbergiidae Stach, 1954. Moreover, the deep genetic divergence indicated by relatively long branch lengths supported the separation between Tullbergiidae and Onychiuridae, and indicated these two apparently similar groups may have undergone different selective pressures affecting the mitochondrial genomes [43]. From an eco-morphological aspect, although both groups are usually considered as euedaphic lifeform [3, 70], most tullbergiids have a smaller body size and more slender body shape than onychiurids. This indicates they are more adapted to finer pores in deep soil, while in epedaphic habitats (litters, mosses), onychiurids are more frequent than tullbergiids. Additional comparative studies are required to address how habitat preference of Collembola could have influenced their mitogenomes.

The Entomobryomorpha was recognised by only five analyses with low to medium support. Among this order, Isotomidae and Entomobryoidea were clustered together with high support in all analyses, which is in line with most previous studies [14–17, 44]. The problem in the position of the Tomoceroidea still exists, however, compared to other molecular studies focusing on the phylogeny of Collembola or Tomoceroidea [9, 15, 17], the present study for the first time clustered Tomoceroidea with other Entomobryomorpha. Similar results were reported by more comprehensive phylogeny of Arthropoda [16] and Insecta [71]. Among the analyses which did not retrieve Entomobryomorpha, two of them using nucleotide datasets with heterotachy models still indicated the basal position of Tomoceroidea within the branch containing Symphypleona s. s. and Entomobryomorpha. On morphological grounds, the Tomoceroidea is more similar to Entomobryidae and Isotomidae than any other groups [17], especially in the form of the prothoracic tergum and appendages. From an ecological aspect, life form and trophic niche also indicate Tomoceroidea closer to other Entomobryomorpha [3, 72]. Therefore, the previous hypothesis should be kept that the Tomoceroidea is a branch of Entomobryomorpha splitting early from the main trunk [17], resulting in considerable genetic divergences.

### Infra-ordinal phylogeny

Since more species and groups have been added, our study also provided information in infra-ordinal level phylogeny of Collembola. Several current familial and superfamilial groupings were strongly supported in most analyses, including Onychiuridae, Neanuridae, Isotomidae, Entomobryoidea and Sminthuridae, which have also been well defined by morphological characters, such as status of prothorax, pseudocelli, antennal sensory organs, body segment ratios, mouthparts and chaetotaxy [73–74]. Paronellidae was revealed to be polyphyletic, with two subfamilies, Salininae and Paronellinae, related to two other subfamilies Lepidocyrtinae and Entomobryinae of Entomobryidae, respectively. This result coincided with the findings of Zhang et al. [21–22] based on both molecular and morphological evidence, and supported the division between Salininae and Paronellinae [73]. We found the Hypogastruridae was paraphyletic and Tetrodontophorinae was within Onychiurinae. These findings were also in conflict with traditional morphological classification, but similar to the results of previous molecular studies [9, 15, 17, 44, 75]. These problematic families should receive thorough revisions in future studies.

The present study, for the first time, strongly supports the monophyly of Tomoceroidea when all three familial/subfamilial taxa were present. However, the relationships between the three groups was not determined, as seven analyses supported the monophyly of Tomoceridae, four analyses supported Tomocerinae + Oncopoduridae, and two analyses supported

Oncopoduridae + Lepidophorellinae. Tomoceridae was recovered in another study using 18S and 28S genes [17], but the species of Lepidophorellinae included in two studies belong to two morphologically distinct genera and tribes. *Lepidophorella* (tribe Lepidophorellini) used in the previous study has a curved dens and short falcate mucro, while *Novacerus* (tribe Novacerini) used in the present study has a straight dens and elongated multi-dentate mucro. Relationship between the two tribes is so far unclear. Number of eyes and ratio of the length of antennal segments indicate Novacerini is more similar to Lepidophorellini, but the morphology of the jumping organ may group Novacerini with Tomocerinae. However, despite the significant difference in body size, Novacerini and Oncopoduridae are also similar in antennal ratio, presence of postantennal organs (unpublished data) and morphology of mucro. For example, the structure of mucro in *Novacerus* is more similar to that of Oncopoduridae (in particular *Harlommillsia*) than to that of Tomocerinae in arrangement and relative size of dorsal teeth. Therefore, on the basis of current progresses in morphological data and molecular phylogeny, an expected relationship within Tomoceroidea is probably ((Lepidophorellini + Tomocerinae) + Novacerini) + Oncopoduridae, which hypothesis requires further survey.

### Performance of different analytical methods

Both morphological evidence and previous molecular studies have strongly supported Tullbergiidae within Poduromorpha and probably the sister group of Onychiuridae [9, 12, 15, 17]. Therefore, although some analyses showed strong support for Tullbergiidae + Neelipleona, these results can be attributed to long-branch effects caused by model mis-specification [55]. Accordingly, implausible results were yielded by three analyses using PCG sequences with codon and GHOST models and two analyses using amino acids with GHOST\_LG and PMSF\_mtART models. In comparison, all three analyses using both protein-coding and rRNA genes (15fna) supported Tullbergiidae + Onychiuridae (Onychiuroidea) with high support values (support  $\geq 95$ ). This is in line with other mitophylogenomic studies of insects showing that the inclusion of rRNA could improve resolution and nodal support [76–77], although rRNA may have higher levels of homoplasy than PCGs [76]. Two analyses using partitioned PCG sequences without codon model and one analysis using partitioned amino acid sequences also clustered Tullbergiidae with Onychiuridae, showing an advantage of partitioning method. This confirmed that algorithmically optimized partitioning schemes outperform unpartitioned analyses in phylogenetic inference due to better accommodation to the variation in substitution patterns among sites [34]. The other two analyses recovering Tullbergiidae + Onychiuridae were based on amino acids with GHOST\_mtART and PMSF\_LG models, suggesting the importance of selecting suitable exchange rate matrices under different heterotachy models.

Although Tomoceroidea has been previously clustered with Poduromorpha in some phylogenetic analyses based on a few rDNA markers [9, 15, 17], strong evidences from morphology [17, 78], ecology [3, 72], transcriptomes [71] and complete genomes [79] all support its closer relationship with other Entomobryomorpha than with Poduromorpha. Therefore, the topology of Tomoceroidea + Poduromorpha even with apparently robust supports might be considered incorrect. Accordingly, two analyses using partitioned nucleotide sequences excluding third codon positions and one analysis using partitioned amino acid sequences outperformed others among those also successfully recovering Onychiuroidea. Under lower selective pressures, the third codon positions are strongly affected by base compositional bias [22], thus are probably important source of homoplasy [76], which may result in artificial phylogenetic affinities. However, the inclusion of third codon positions do not consistently affect the topology and robustness of phylogenetic trees. For example, previous studies have reported negative effect on Hymenoptera [38] and Dictyoptera [80], neutral effect on Orthoptera [77] and

Psocodea [37], and positive effect on Diptera [76], indicating the necessity of assessing the performance of inclusion or exclusion of third codon positions during each phylogenetic inference.

## Conclusions

Our study based on mitochondrial genomic analyses brought new insight into the high-level phylogenetic relationship within Collembola. The gene rearrangements along mitogenomes were identical within Onychiuridae and similar within Sminthuridae, suggesting that gene orders could provide useful information for inferring relationships among lineages, although their exact phylogenetic signals still need to be assessed. As expected, phylogenetic analyses based on different datasets and models yielded inequivalent tree topologies and nodal supports. In this study, the performance of analyses was increased by sequence partitioning, exclusion of third codon positions and inclusion of two rRNA genes, but not by translating nucleotides into codon or amino acid sequences or application of site- and branch-heterogeneous models. Regarding the phylogeny of Collembola, on infra-ordinal level, most families and superfamilies were strongly supported, except for Paronellidae and Hypogastruridae. Sub-familial-level relationship within Onychiuridae and Tomoceridae also required further clarification. At ordinal level, Symphypleona *s. s.* was most strongly supported, followed by Poduromorpha. Although Entomobryomorpha was not well supported, the recovered topology indicated future denser taxon sampling, data refining and analytical method optimization may enhance the support for this group. The relationship between elongated and spherical groups still remained a major unsolved problem. Although most analyses placed Neelipleona basal to all other orders, the possibility of artifact could not be eliminated given the relatively long branch lengths. In our results Symphypleona *s. s.* could be clustered with any other orders, while the highest supports were found with Entomobryomorpha. It is expected that inclusion of intermediate forms, such as Mackenziellidae and Coenaletidae, will promote the resolution of this phylogenetic problem.

## Supporting information

**S1 Table. Taxa under study, detail of GenBank accession numbers and sampling location.**  
(DOCX)

**S2 Table. Sequence condition and gene orders of mitogenomes among studied species.**  
(XLSX)

**S3 Table. Results of tree topology tests.**  
(XLSX)

**S1 Fig. Phylogenetic trees obtained from 10 analyses.**  
(PDF)

## Acknowledgments

We are grateful to Dr. Feng Zhang from College of Plant Protection, Nanjing Agricultural University for help in collecting the specimens and analyzing the data, to Dr. Francesco Nardi and Dr. Antonio Carapelli for their valuable comments on this manuscript.

## Author Contributions

**Conceptualization:** Daoyuan Yu, Penelope Greenslade.



**Data curation:** Zhijing Xie, Jie Dong, Yinhuan Ding.

**Formal analysis:** Xin Sun, Zhijing Xie, Jie Dong, Yinhuan Ding, Haifeng Yao.

**Funding acquisition:** Xin Sun, Daoyuan Yu.

**Investigation:** Daoyuan Yu.

**Methodology:** Daoyuan Yu, Jie Dong, Yinhuan Ding.

**Project administration:** Daoyuan Yu.

**Resources:** Daoyuan Yu, Penelope Greenslade.

**Software:** Zhijing Xie, Jie Dong, Yinhuan Ding.

**Supervision:** Daoyuan Yu.

**Validation:** Daoyuan Yu.

**Visualization:** Xin Sun, Zhijing Xie, Jie Dong, Haifeng Yao.

**Writing – original draft:** Xin Sun, Zhijing Xie.

**Writing – review & editing:** Daoyuan Yu.

## References

1. Bellinger P, Christiansen KA, Janssens F. Checklist of the Collembola of the World. 1996–2019 <http://www.collembola.org> [Date of access: 17.07.2019]
2. Petersen H. General aspects of collembolan ecology at the turn of the millennium: proceedings of the Xth international colloquium on Apterygota, České Budějovice 2000: Apterygota at the beginning of the third millennium. *Pedobiologia*. 2002; 46: 246–260.
3. Potapov AA, Semenina EE, Korotkevich AY, Kuznetsova NA, Tiunov AV. Connecting taxonomy and ecology: Trophic niches of collembolans as related to taxonomic identity and life forms. *Soil Biol Biochem*. 2016; 101: 20–31.
4. Börner C. Das system der collembolen, nebst beschreibung neuer collembolen des hamburgener naturhistorischen museums. *Mitt Naturh Mus Hamburg*. 1906; 23: 147–188.
5. Yosii R. Critical check list of the Japanese species of Collembola. *Contri Biol Lab Kyoto Univ*. 1977; 25: 141–170.
6. Christiansen K, Bellinger P. The Collembola of North America North of the Rio Grand. Grinnell Iowa: Grinnell College; 1980–1981.
7. Massoud Z. Contribution à la connaissance morphologique et systématique des Collembolés Neelidae. *Rev Ecol Biol Sol*. 1971; 8: 195–198.
8. Massoud Z. Essai de synthèse sur la phylogénie des Collembolés. *Rev Ecol Biol Sol*. 1976; 13: 241–252.
9. D'Haese CA. Were the first springtails semi-aquatic? A phylogenetic approach by means of 28S rDNA and optimization alignment. *Proc R Soc B*. 2002; 269: 1143. <https://doi.org/10.1098/rspb.2002.1981> PMID: [12061958](https://pubmed.ncbi.nlm.nih.gov/12061958/)
10. Deharveng L. Recent advances in Collembola systematics. *Pedobiologia*. 2004; 48: 415–433.
11. Hopkin SP. A Key to the Collembola (Springtails) of Britain and Ireland. Telford: Field Studies Council (FSC); 2007.
12. D'Haese CA. Morphological appraisal of Collembola phylogeny with special emphasis on Poduromorpha and a test of the aquatic origin hypothesis. *Zool Scr*. 2003; 32: 563–586.
13. Luan YX, Mallatt JM, Xie RD, Yang YM, Yin WY. The phylogenetic positions of the three basal-hexapod groups (Protura, Diplura, and Collembola) based on ribosomal RNA gene sequences. *Mol Biol Evol*. 2005; 22: 1579–1592. <https://doi.org/10.1093/molbev/msi148> PMID: [15845456](https://pubmed.ncbi.nlm.nih.gov/15845456/)
14. Gao Y, Bu Y, Luan YX. Phylogenetic relationships of basal hexapods reconstructed from nearly complete 18S and 28S rRNA gene sequences. *Zool Sci*. 2008; 25: 1139–1145. <https://doi.org/10.2108/zsj.25.1139> PMID: [19267625](https://pubmed.ncbi.nlm.nih.gov/19267625/)

15. Xiong Y, Gao Y, Yin WY, Luan YX. Molecular phylogeny of Collembola inferred from ribosomal RNA genes. *Mol Phylogenet Evol.* 2008; 49: 728–735. <https://doi.org/10.1016/j.ympev.2008.09.007> PMID: [18835455](https://pubmed.ncbi.nlm.nih.gov/18835455/)
16. von Reumont BM, Meusemann K, Szucsich NU, Dell'Ampio E, Gowri-Shankar V, Bartel D, et al. Can comprehensive background knowledge be incorporated into substitution models to improve phylogenetic analyses? A case study on major arthropod relationships. *BMC Evol Biol.* 2009; 9: 119. <https://doi.org/10.1186/1471-2148-9-119> PMID: [19473484](https://pubmed.ncbi.nlm.nih.gov/19473484/)
17. Yu D, Zhang F, Stevens MI, Yan Q, Liu M, Hu F. New insight into the systematics of Tomoceridae (Hexapoda, Collembola) by integrating molecular and morphological evidence. *Zool Scr.* 2016; 45: 286–299.
18. Schneider C, Cruaud C, D'Haese CA. Unexpected diversity in Nelipleona revealed by molecular phylogeny approach (Hexapoda, Collembola). *Soil Org.* 2011; 83(3): 383–398.
19. Schneider C, D'Haese CA. Morphological and molecular insights on *Megalothorax*: the largest Neelipleona genus revisited (Collembola). *Invertebr Syst.* 2013; 27: 317–364.
20. Zhang F, Ma Y, Greenslade P. New Australian Paronellidae (Collembola) reveal anomalies in existing tribal diagnoses. *Invertebr Syst.* 2017; 31: 375–393.
21. Zhang F, Sun D, Yu D, Wang B. Molecular phylogeny supports S-chaetae as a key character better than jumping organs and body scales in classification of Entomobryoidea (Collembola). *Sci Rep*, 2015; 5: 12471. <https://doi.org/10.1038/srep12471> PMID: [26212216](https://pubmed.ncbi.nlm.nih.gov/26212216/)
22. Cameron SL. Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny. *Annu Rev Entomol.* 2014; 59: 95–117. <https://doi.org/10.1146/annurev-ento-011613-162007> PMID: [24160435](https://pubmed.ncbi.nlm.nih.gov/24160435/)
23. Cook CE, Yue Q, Akam M. Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. *Proc R Soc B.* 2005; 272: 1295–1304. <https://doi.org/10.1098/rspb.2004.3042> PMID: [16024395](https://pubmed.ncbi.nlm.nih.gov/16024395/)
24. Cameron SL, Beckenbach AT, Downton M, Whiting MF. Evidence from mitochondrial genomics on interordinal relationships in insects. *Arthropod Syst Phylo.* 2006; 64(1): 27–34.
25. Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annu Rev Ecol Evol Syst.* 2006; 37: 545–579.
26. Smith DR. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Brief Funct Genomics.* 2016; 15(1): 47–54. <https://doi.org/10.1093/bfgp/elv027> PMID: [26117139](https://pubmed.ncbi.nlm.nih.gov/26117139/)
27. Tomasco IH, Lessa EP. The evolution of mitochondrial genomes in subterranean caviomorph rodents: Adaptation against a background of purifying selection. *Mol Phylogenet Evol.* 2011; 61: 64–70. <https://doi.org/10.1016/j.ympev.2011.06.014> PMID: [21723951](https://pubmed.ncbi.nlm.nih.gov/21723951/)
28. Yu L, Wang X, Ting N, Zhang Y. Mitogenomic analysis of Chinese snub-nosed monkeys: Evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. *Mitochondrion.* 2011; 11: 497–503. <https://doi.org/10.1016/j.mito.2011.01.004> PMID: [21292038](https://pubmed.ncbi.nlm.nih.gov/21292038/)
29. Foote AD, Morin PA, Durban JW, Pitman RL, Wade P, Willerslev E, et al. Positive selection on the killer whale mitogenome. *Biol Lett.* 2011; 7: 116–8. <https://doi.org/10.1098/rsbl.2010.0638> PMID: [20810427](https://pubmed.ncbi.nlm.nih.gov/20810427/)
30. Hall BG. Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. *Mol Biol Evol.* 2004; 22(3): 792–802. <https://doi.org/10.1093/molbev/msi066> PMID: [15590907](https://pubmed.ncbi.nlm.nih.gov/15590907/)
31. Simmons MP. Relative benefits of amino-acid, codon, degeneracy, DNA, and purine-pyrimidine character coding for phylogenetic analyses of exons. *J Syst Evol.* 2017; 55(2): 85–109.
32. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 2004; 21(6): 1095–1109. <https://doi.org/10.1093/molbev/msh112> PMID: [15014145](https://pubmed.ncbi.nlm.nih.gov/15014145/)
33. Crotty SM, Minh BQ, Bean NG, Holland BR, Tuke J, Jermiin LS, et al. GHOST: recovering historical signal from heterotachously evolved sequence alignments. *Syst Biol.* 2019; syz051, <https://doi.org/10.1093/sysbio/syz051>.
34. Kainer D, Lanfear R. The effects of partitioning on phylogenetic inference. *Mol Biol Evol.* 2015; 32(6): 1611–1627. <https://doi.org/10.1093/molbev/msv026> PMID: [25660373](https://pubmed.ncbi.nlm.nih.gov/25660373/)
35. Song N, Zhang H, Zhao T. Insights into the phylogeny of Hemiptera from increased mitogenomic taxon sampling. *Mol Phylogenet Evol.* 2019; 137: 236–249. <https://doi.org/10.1016/j.ympev.2019.05.009> PMID: [31121308](https://pubmed.ncbi.nlm.nih.gov/31121308/)
36. Liu Y, Song F, Jiang P, Wilson JJ, Cai W, Li H. Compositional heterogeneity in true bug mitochondrial phylogenomics. *Mol Phylogenet Evol.* 2018; 118: 135–144. <https://doi.org/10.1016/j.ympev.2017.09.025> PMID: [28986237](https://pubmed.ncbi.nlm.nih.gov/28986237/)

37. Yoshizawa K, Johnson KP, Sweet AD, Yao I, Ferreira RL, Cameron SL. Mitochondrial phylogenomics and genome rearrangements in the barklice (Insecta: Psocodea). *Mol Phylogenet Evol.* 2018; 119: 118–127. <https://doi.org/10.1016/j.ympev.2017.10.014> PMID: 29079378
38. Downton M, Cameron SL, Austin AD, Whiting MF. Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera—A lineage with both rapidly and slowly evolving mitochondrial genomes. *Mol Phylogenet Evol.* 2009; 52: 512–519. <https://doi.org/10.1016/j.ympev.2009.04.001> PMID: 19364540
39. Tang P, Zhu JC, Zheng BY, Wei SJ, Sharkey M, Chen XX, et al. Mitochondrial phylogenomics of the Hymenoptera. *Mol Phylogenet Evol.* 2019; 131: 8–18. <https://doi.org/10.1016/j.ympev.2018.10.040> PMID: 30399430
40. Yuan ML, Zhang QL, Zhang L, Guo ZL, Liu YJ, Shan YY, et al. High-level phylogeny of the Coleoptera inferred with mitochondrial genome sequences. *Mol Phylogenet Evol.* 2016; 104: 99–111. <https://doi.org/10.1016/j.ympev.2016.08.002> PMID: 27497607
41. Yang X, Cameron SL, Lees DC, Xue DY, Han HX. A mitochondrial genome phylogeny of owlet moths (Lepidoptera: Noctuoidea), and examination of the utility of mitochondrial genomes for lepidopteran phylogenetics. *Mol Phylogenet Evol.* 2015; 85: 230–237. <https://doi.org/10.1016/j.ympev.2015.02.005> PMID: 25698356
42. Carapelli A, Convey P, Nardi F, Frati F. The mitochondrial genome of the antarctic springtail *Folsomotoma octooculata* (Hexapoda; Collembola), and an update on the phylogeny of collembolan lineages based on mitogenomic data. *Entomologia.* 2014; 2: 44–55.
43. Carapelli A, Fanciulli PP, Frati F, Leo C. Mitogenomic data to study the taxonomy of Antarctic springtail species (Hexapoda: Collembola) and their adaptation to extreme environments. *Polar Biol.* 2019; 42: 715–732.
44. Leo C, Carapelli A, Cicconardi F, Frati F, Nardi F. Mitochondrial genome diversity in collembola: phylogeny, dating and gene order. *Diversity.* 2019; 11: 169. <https://doi.org/10.3390/d11090169>
45. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome oxidase subunit I from diverse metazoan invertebrates. *Mol Marine Biol Biotechnol.* 1994; 3 (5): 294–299. PMID: 7881515
46. Zhang F, Yu D, Luo Y, Ho SYW, Wang B, Zhu C. Cryptic diversity, diversification and variance in two species complexes of *Tomocerus* (Collembola, Tomoceridae) from China. *Zool Scr.* 2014; 43: 393–404.
47. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016; 33(7): 1870–4. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
48. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 2017; 45(4): e18. <https://doi.org/10.1093/nar/gkw955> PMID: 28204566
49. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritsch G, et al. MITOS: Improved de novo Metazoan Mitochondrial Genome Annotation. *Mol Phylogenet Evol.* 2013; 69(2): 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023> PMID: 22982435
50. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997; 25(5): 955–964. <https://doi.org/10.1093/nar/25.5.955> PMID: 9023104
51. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics.* 2018; 34(14): 2490–2492. <https://doi.org/10.1093/bioinformatics/bty121> PMID: 29506019
52. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009; 25(15): 1972–3. <https://doi.org/10.1093/bioinformatics/btp348> PMID: 19505945
53. Kück P, Longo GC. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool.* 2014; 11: 81. <https://doi.org/10.1186/s12983-014-0081-x> PMID: 25426157
54. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017; 14: 587–589. <https://doi.org/10.1038/nmeth.4285> PMID: 28481363
55. Wang HC, Minh BQ, Susko S, Roger AJ. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol.* 2018; 67: 216–235. <https://doi.org/10.1093/sysbio/syx068> PMID: 28950365

56. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 2018; 35: 518–522. <https://doi.org/10.1093/molbev/msx281> PMID: 29077904
57. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol.* 2015; 32: 268–274. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430
58. Kishino H, Miyata T, Hasegawa M. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J Mol Evol.* 1990; 31(2): 151–160.
59. Shimodaira H. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* 2002; 51: 492–508. <https://doi.org/10.1080/10635150290069913> PMID: 12079646
60. Strimmer K, Rambaut A. Inferring confidence sets of possibly misspecified gene trees. *Proc R Soc B.* 2002; 269: 137–142. <https://doi.org/10.1098/rspb.2001.1862> PMID: 11798428
61. Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol.* 1989; 29(2): 170–179. <https://doi.org/10.1007/bf02100115> PMID: 2509717
62. Shimodaira H, Hasegawa M. Multiple comparison of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol.* 1999; 16: 1114–1116.
63. Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res.* 1999; 27: 1767–1780. <https://doi.org/10.1093/nar/27.8.1767> PMID: 10101183
64. Lavrov DV, Lang BF. Poriferan mtDNA and animal phylogeny based on mitochondrial gene arrangements. *Syst Biol.* 2005; 54(4): 651–659. <https://doi.org/10.1080/10635150500221044> PMID: 16126659
65. Schneider C. Morphological review of the order Neelipleona (Collembola) through the redescription of the type species of *Acanthoneelidus*, *Neelides* and *Neelus*. *Zootaxa.* 2017; 4308(1): 1–94.
66. Yosii R. Phylogenetische Bedeutung der Chaetotaxie bei den Collembolen. *Contri Biol Lab Kyoto Univ.* 1961; 12: 1–37.
67. Fjellberg A. Redescription of *Mackenziella psocoides* Hammer, 1953 and discussion of its systematic position (Collembola, Mackenziellidae) Dallai R. (ed.)—Proceedings Third International Seminar on Apterygota. Siena: University of Siena; 1989. p. 93–105.
68. Felsenstein J. Cases in which parsimony or compatibility methods can be positively misleading. *Syst Zool.* 1978; 27: 401–410.
69. Gribaldo S, Philippe H. Ancient phylogenetic relationships. *Theor Popul Biol.* 2002; 61: 391–408. <https://doi.org/10.1006/tpbi.2002.1593> PMID: 12167360
70. Salmon S, Ponge JF, Gachet S, Deharveng L, Lefebvre N, Delabrosse F. Linking species, traits and habitat characteristics of Collembola at European scale. *Soil Biol Biochem.* 2014; 75: 73–85.
71. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science.* 2014; 346: 763–767. <https://doi.org/10.1126/science.1257570> PMID: 25378627
72. Potapov AA, Korotkevich AY, Tiunov AV. Non-vascular plants as a food source for litter-dwelling Collembola: field evidence. *Pedobiologia.* 2018; 66: 11–17.
73. Deharveng L, Bedos A. Insecta: Collembola. In Yule C.M. and Yong H.S. (Eds). *Freshwater Invertebrates of the Malaysian Region.* Malaysia: Academy of Sciences; 2004. p. 384–393.
74. Cipola NG, da Silva DD, Bellini BC. Class Collembola, In: Thorp and Covich's *Freshwater Invertebrates*, 4th ed. Volume 3: *Keys to Neotropical Hexapoda.* Elsevier; 2018. p. 11–55.
75. Greenslade P, Stevens MI, Torricelli G, D'Haese CA. An ancient Antarctic endemic genus restored: morphological and molecular support for *Gomphiocephalus hodgsoni* (Collembola: Hypogastruridae). *Syst Entomol.* 2011; 36: 223–240.
76. Cameron SL, Lambkin CL, Barker SC, Whiting MF. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Syst Entomol.* 2007; 32: 40–59.
77. Fenn JD, Song H, Cameron SL, Whiting MF. A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol Phylogenet Evol.* 2008; 49: 59–68. <https://doi.org/10.1016/j.ympev.2008.07.004> PMID: 18672078
78. Szeptycki A. Morpho-systematic studies on Collembola. IV. Chaetotaxy of the Entomobryidae and its phylogenetical significance. Kraków: Polska Akademia Nauk; 1979.

79. Zhang F, Ding Y, Zhou QS, Wu J, Luo A, Zhu C. A high-quality draft genome assembly of *Sinella curviseta*: a soil model organism (Collembola). *Genome Biol Evol.* 2019; 11(2): 521–530. <https://doi.org/10.1093/gbe/evz013> PMID: 30668671
80. Cameron SL, Lo N, Bourguignon T, Svenson GJ, Evans TA. A mitochondrial genome phylogeny of termites (Blattodea: Termitoidea): Robust support for interfamilial relationships and molecular synapomorphies define major clades. *Mol Phylogenet Evol.* 2012; 65: 162–173