Title	Mitochondrial phylogenomics and genome rearrangements in the barklice (Insecta: Psocodea)
Author(s)	Yoshizawa, Kazunori; Johnson, Kevin P.; Sweet, Andrew D.; Yao, Izumi; Ferreira, Rodrigo L.; Cameron, Stephen L.
Citation	Molecular phylogenetics and evolution, 119, 118-127 https://doi.org/10.1016/j.ympev.2017.10.014
Issue Date	2018-02
Doc URL	http://hdl.handle.net/2115/72476
Rights	© 2017 Elsevier Inc. All rights reserved., This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
Rights(URL)	https://creativecommons.org/licenses/by-nc-nd/4.0/
Type	article (author version)
File Information	2018MPE(1).pdf

[Instructions for use](#)

1 *For submission to MPE*

2
3 **Mitochondrial phylogenomics and genome rearrangements in the**
4 **barklice (Insecta: Psocodea)**

5
6 Kazunori Yoshizawa ^{a*}, Kevin P. Johnson ^b, Andrew D. Sweet ^b, Izumi Yao ^a, Rodrigo
7 L. Ferreira ^c, Stephen L. Cameron ^d

8
9 *a Systematic Entomology, School of Agriculture, Hokkaido University, Sapporo 060-*
10 *8589, Japan*

11 *b Illinois Natural History Survey, University of Illinois, Champaign, IL 61820, USA*

12 *c Biology Department, Federal University of Lavras, 37200-000 Lavras, MG, Brazil*

13 *d Department of Entomology, Purdue University, West Lafayette, IN 47907, USA*

14
15 *corresponding author. Tel: +81-11-706-2424. E-mail: psocid@res.agr.hokudai.ac.jp

ABSTRACT

The mitochondrial genome arrangement in the insect order Psocodea (booklice, barklice, and parasitic lice) is extremely variable. Genome organization ranges from the rearrangement of a few tRNAs and protein coding genes, through extensive tRNA and protein coding gene rearrangements, to subdivision into multiple mini-chromosomes. Evolution of the extremely modified mitochondrial genome in parasitic lice (Phthiraptera) has been the subject of several studies, but limited information is available regarding the mitochondrial genome organization of the more plesiomorphic, free-living Psocodea (formerly known as the "Psocoptera"). In particular, the ancestral state of the psocodean mitochondrial genome arrangement and the evolutionary pathway to the rearranged conditions are still unknown. In this study, we addressed mitochondrial evolutionary questions within the Psocodea by using mitochondrial genome sequences obtained from a wide range of Psocoptera, covering all three suborders. We identified seven types of mitochondrial genome arrangements in Psocoptera, including the first example in Psocodea of retention of the ancestral pancrustacean condition in *Prionoglaris* (Prionoglarididae). Two methods (condition-based parsimony reconstruction and common-interval genome distances) were applied to estimate the ancestral mitochondrial arrangement in Psocodea, and both provided concordant results. Specifically, the common ancestor of Psocodea retained the ancestral pancrustacean condition, and most of the gene arrangement types have originated independently from this ancestral condition. We also utilized the genomic data for phylogenetic estimation. The tree estimated from the mitochondrial genomic data was well resolved, strongly supported, and in agreement with previously estimated phylogenies. It also provided the first robust support for the family Prionoglarididae, as its monophyly was uncertain in previous morphological and molecular studies.

Keywords: mitochondrial genome, gene rearrangements, Psocodea, Psocoptera, Prionoglarididae, phylogeny

1. Introduction

During the last couple of decades, sequences of the mitochondrial genome from hundreds of insect species have been obtained. These sequences have been used for phylogenetic analyses at both deep and shallow levels, as well as for analyses of mitochondrial genome organization (Cameron, 2014a, 2014b). As the sequences of more and more insect mitochondrial genomes have been obtained, it has become clear that, although gene arrangement is quite stable throughout many insects (the ancestral Pancrustacean condition is by far the most common mitochondrial genome arrangement observed), rearrangements of a few transfer RNA genes (tRNAs) are also quite common (Cameron, 2014a). In contrast, more extensive gene rearrangements, particularly those involving protein-coding genes (PCGs) or ribosomal RNA genes (rRNAs), are rather rare events and are common in only a few insect orders (Embioptera: Kômoto et al., 2012; Thysanoptera: Shao and Barker, 2003; Psocodea: Shao et al., 2001a; Hymenoptera: Mao et al., 2015). Extensive gene rearrangements, however, also occur in a number of highly derived members of orders in which most other taxa lack major rearrangements (e.g. Cecidomyiidae, Diptera: Beckenbach, 2012; Iberobaeniidae, Coleoptera: Andujar, 2017; Enicocephalidae, Hemiptera: Li et al., 2012; Aleyrodidae, Hemiptera: Thao et al. 2004).

Among insects, the highest variation in mitochondrial gene arrangement occurs in the order Psocodea (booklice, barklice, and parasitic lice, formerly known as two independent orders, "Psocoptera" and Phthiraptera: Yoshizawa and Johnson, 2006). The mitochondrial variation observed in Psocodea ranges from the rearrangement of a few tRNAs and two PCGs in the suborder Psocomorpha (Shao et al., 2001b; Cameron, 2014a; Li et al., 2013), through extensive tRNA and PCG rearrangements in the suborder Trogiomorpha (Shao et al., 2003), the family Liposcelididae (Shi et al., 2016), and Phthiraptera (e.g., Shao et al., 2001a), to extreme subdivision into multiple mini-chromosomes in some Liposcelididae (Chen et al., 2014) and Phthiraptera (Shao et al., 2009, 2015; Cameron et al., 2011). Evolution of the extremely modified mitochondrial genome in Phthiraptera has been the subject of several studies (e.g., Shao et al., 2001a, 2009, 2015; Cameron et al., 2011). However, other than the family Liposcelididae (the sister-group of the parasitic lice, with many reduced traits

similar to parasitic lice; Yoshizawa and Lienhard, 2010), limited information is available for the more plesiomorphic, free-living Psocodea (formerly the "Psocoptera"). Therefore, the ancestral condition of the psocodean mitochondrial genome arrangement is still unknown. In addition, extensive mitochondrial rearrangements are also known from thrips (Thysanoptera) (Shao and Barker, 2003; Yan et al., 2014; Dickey et al., 2015; Liu et al., 2017), an order classified with Psocodea as part of the superorder Paraneoptera (Yoshizawa and Lienhard, 2016). Additional mitochondrial genomic information from free-living Psocodea is thus crucial to inferring both the ancestral mitochondrial genome organization of Psocodea and for understanding supra-ordinal level evolution of the mitochondrial genome.

Additional mitochondrial genome data from Psocodea will also contribute to our phylogenetic understanding of the order. Although the higher-level phylogenetic relationships within Psocodea have been the subject of several studies (e.g., Yoshizawa et al., 2006; Yoshizawa and Johnson, 2010, 2014), unresolved problems still remain. One of these concerns the monophyly of the Prionoglarididae (suborder Trogiomorpha). Because the family is known to retain many plesiomorphic features (Lienhard, 1998; Yoshizawa et al., 2006), its monophyly is highly controversial. Although one previous molecular phylogenetic analysis (Yoshizawa et al., 2006: fig. 2) provided support for the monophyly of Prionoglarididae, analyses with more extensive taxon or gene samplings (Yoshizawa et al., 2006: fig. 3; Yoshizawa and Johnson, 2014) suggested the family may be paraphyletic. As mentioned above, this family retains the most plesiomorphic morphology among the extant Psocodea, and so resolving the status of Prionoglarididae has great impact on how we interpret ancestral states and the evolution of the Psocodea.

In this study, we address both phylogenetic and mitochondrial evolutionary questions within the Psocodea by using the mitochondrial genome sequences obtained from a wide range of free-living Psocodea. The selected taxa cover all three suborders of the "Psocoptera". In particular, three genera representing both subfamilies of the Prionoglarididae were sampled to test the monophyly of this family and also to examine the origin of the extensive gene rearrangements previously recorded in members of the suborder Trogiomorpha. Two methods of the ancestral state estimation, condition-based coding with parsimony reconstruction and common-

interval genome distances (implemented in the TreeREx software: Bernt et al., 2007, 2008), are compared to test the effectiveness of these methods for ancestral state reconstruction.

2. Materials and Methods

2.1. Samples

Ten species (Table 1) were sequenced representing all three of the free-living suborders of Psocodea, including eight different families. In addition, sequences of Lepidopsocidae sp. (Shao et al., 2003), *Stenocaecilius quercus* (= *Caecilius quercus*: Shao et al., 2001b), *Psococerastis albimaculata*, and *Longivalvus hyalospilus* (Li et al., 2013) were obtained from GenBank. Mitochondrial genomes have also been previously sequenced for parasitic lice (Phthiraptera) and booklice (Liposcelididae). However, sequences from both groups were excluded from the present study due to their extremely high rates of mitochondrial genome rearrangement and fragmentation (Cameron et al., 2011; Chen et al., 2014; Shao et al., 2017), which obscure genome evolution events within free-living Psocodea. Two outgroup sequences, *Abidama producta* (Cercopidae, Hemiptera: an order classified to Paraneoptera together with Psocodea) and *Dysmicohermes ingens* (Corydalidae, Megaloptera: an order of Holometabola, the sister taxon of Paraneoptera), were also obtained from GenBank.

2.2. Sequencing and assembling

DNA was extracted using a Qiagen QIAamp DNA Micro Kit or DNeasy Blood and Tissue Kit. DNA of *Dorypteryx*, *Prionoglaris*, and *Neotrogla* was sheared using a Covaris M220 instrument to approximately 400 bp and sequence libraries were prepared using a Kapa Library Preparation kit. Libraries were pooled with two other taxa and sequenced together in a single lane with 160 bp paired-end reads on an Illumina HiSeq 2500. Raw reads are deposited in NCBI SRA (SRR5308267, SRR5308282, SRR5308278). To obtain mitochondrial genome sequences from these libraries we generated contigs using a combination of aTRAM (Allen et al., 2015) and MITObim (Hahn et al. 2013). First, aTRAM was used to assemble five protein-coding mitochondrial genes (*cox1*, *cox2*, *cob*, *nad2*, and *nad5*) for each genus using amino acid sequences as targets for these assemblies. In all cases, aTRAM

was run for a single iteration on 10% of the paired-end libraries, and contigs were assembled in aTRAM using ABySS (Simpson et al. 2009). Second, MITObim was used to extend the contigs assembled with aTRAM by using each contig as a starting reference for that species. Additionally, partial previously generated Sanger sequences of *rrnS* were used as starting references for all three genera, and Sanger sequences of *rrnL* as starting references for *Dorypteryx* and *Prionoglaris*. MITObim was then run for each starting reference a maximum of 100 iterations, using either 10% (*Dorypteryx* and *Prionoglaris*) or 20% (*Neotrogla*) fractions of the paired-end libraries. To obtain *trnI* and *trnM* sequences for *Neotrogla*, we used aTRAM with sequences of these tRNAs from the *Prionoglaris* and *Speleketor* genomes in this study. A small region not recovered by aTRAM and MITObim (part of *nad4* and *nad4l* of *Prionoglaris*) was amplified by PCR and sequenced by Beckman CEQ2000 Sanger sequencer (Yoshizawa & Johnson, 2003).

The complete mitochondrial genomes of *Speleketor*, *Stimulopalpus*, *Archipsocus*, *Lachesilla* and *Amphigerontia*, and partial mitochondrial genomes for *Echmepteryx* and *Trogium*, were amplified by long PCR and sequenced by primer walking (Cameron 2014b). Long PCRs were performed with Elongase (Invitrogen), Sanger sequenced with the ABI Big Dye ver3 chemistry and run on an ABI 3770 automated sequencer. Amplification primers are listed in Supplementary Table S1. Long PCR and sequencing conditions match those used in Cameron et al. (2011).

2.3. Annotations

The MITOS server (Bernt et al., 2013) was used for initial annotation. However, the MITOS server often could not correctly identify the start and stop codons, so these were manually annotated by aligning the sequences with the annotated mitochondrial genome data of the other Psocodea downloaded from GenBank (Cameron 2014b).

2.4. Alignment

Protein coding genes (PCGs) were aligned based on translated amino-acids using Muscle (Edgar, 2004) implemented in MEGA 7 (Kumar et al., 2016). Ribosomal RNAs (rRNA) were aligned using MAFFT 6.5 (Katoh and Standley, 2013) with the Q-

INS-i option, in which secondary structure information of RNA is considered. Apparent misalignments were corrected manually. Transfer RNAs (tRNAs) were manually aligned based on secondary structure models estimated in MITOS. Poorly aligned regions (such as hyper variable regions of RNAs near the start and stop codons of PCGs) were excluded from the analyses.

2.5. Data set

We prepared the following six data sets: (1) ALL = all protein coding and RNA genes; (2) ex.3rd = all protein coding genes (third codon position excluded) and RNA genes; (3) PCG = all protein coding genes; (4) PCG12 = all protein coding genes (third codon position excluded); (5) RNA = all RNA genes; (6) AA = amino-acid sequences of the PCG dataset. For each data set, two taxon sets were prepared: (1) all taxa and (2) excluding taxa with missing data (*Stenocaecilius*, *Echmepteryx*, and *Trogium*).

For detecting potential biases affecting the accuracy of phylogenetic estimation using mitochondrial genome data (Sheffield et al., 2009), AT content and P-distances were calculated by using MacClade 4 (Maddison and Maddison, 2000) and PAUP* 4.0a152 (Swofford, 2002), respectively. AT content was calculated for each PCG gene, combined tRNAs, each rRNA, and codon positions (PCG1, 2, 3). A chi-square test of base homogeneity was performed using PAUP*.

2.6. Model selection

The best substitution models and partition schemes for the maximum likelihood (ML) and Bayesian analyses were estimated using PartitionFinder 2.1.1 (Lanfear et al., 2017), with the greedy algorithm. Taxa with missing data were excluded for model estimation to avoid the potential negative effects caused by missing data. The following partitions were predefined for the PartitionFinder analyses: codon positions for each PCGs (13 genes x 3 codons = 39 partitions), tRNAs (22 partitions), and rRNAs (2 partitions)

2.7. Tree Search

We estimated a maximum likelihood tree using IQ-Tree 1.4.3 (Nguyen et al.,

205 2015), with 1000 replicates of ultrafast likelihood bootstrap (Minh et al., 2013) to
206 obtain bootstrap branch support values. Bayesian analyses were performed using
207 MrBayes (Ronquist and Huelsenbeck, 2003). We performed two runs each with four
208 chains for 500 000 generations, and trees were sampled every 100 generations. The
209 first 50% of sampled trees was excluded as burn-in, and a 50% majority consensus
210 tree was computed to estimate posterior probabilities. To evaluate the potential
211 impact of substitution rate and compositional biases on phylogeny estimation, we
212 also performed tree searches using PhyloBayes 4.1 (Lartillot et al., 2009) under a
213 heterogeneous (CAT+GTR) model. We ran two independent tree searches for 10,000
214 cycles. However, for the PCG12 data, the two runs did not converge by 10,000 cycles
215 (maxdiff > 0.3), so we ran 20,000 cycles for this data set. The first 50% of sampled
216 trees were excluded as burn-in, and trees were sampled every 10 cycles. A majority
217 consensus tree was computed from the two combined runs.
218

219 2.8. Character Coding and Ancestral State Estimation

220 Each genome was compared to the inferred ancestral insect mitochondrial
221 genome (present in both outgroup taxa) to examine pairs of adjacent genes or gene-
222 boundaries. Novel gene boundaries, those not observed in the ancestral insect
223 mitochondrial genome, were coded as binary characters (either present or absent).
224 Genome rearrangements result in new gene-pairs from both the insertion of a
225 gene/gene-block at a novel location and from its deletion from the ancestral location.
226 Both of these types of events were coded separately (Fig. 1: insertions labelled with
227 numbers, deletions labelled with letters). For example, in *Speleketor* the translocation
228 of *trnC* results in both *trnC-trnQ*, a novel boundary formed by an insertion, whereas
229 *trnW-trnY* is also a novel boundary but was formed by the deletion of *trnC* from its
230 ancestral location, the *trnW-trnC-trnY* condition. This condition-based data matrix
231 was optimized parsimoniously on the best phylogenomic tree obtained from the ALL
232 dataset (see above) using MacClade.
233

234 We also reconstructed the gene rearrangement history by using TreeREx 1.85
235 (Bradt et al., 2008). TreeREx reconstructs genomic evolution based on common
236 intervals (blocks of genes shared between taxa in a clade) and a defined phylogeny,
237 allowing the inference of tandem-duplicate-random-loss events (TDRL after Boore,

2000), simple transpositions, inversions and inversion-transpositions (the latter three models can also be coded in the condition-based matrix described above). However, there are limitations to TreeREx, particularly that gene-duplications are not allowed, even though these are comparatively common in rearranged mitochondrial genomes and are an inferred mid-point in the TDRL model (after tandem duplication but prior to random loss). Duplicated genes were identified in two of the taxa sequenced in this study, two control regions (CR) are present in *Neotrogla* and *Speleketor*. Therefore, for these, only the CR at a novel position (indicated by asterisk in Fig. 1) was coded.

3. Results

3.1. Sequencing, Annotation, and Data Evaluation

Eight new, complete/nearly complete (missing only a portion of control region) mitochondrial genomes were sequenced representing five additional psocodean families and each of the three free-living suborders: Trogiomorpha (Prionoglarididae: *Prionoglaris stygia* 15,684+ bp at 67x mean coverage, *Neotrogla* sp. 16894+ bp at 81x mean coverage, *Speleketor irwini* 16,849bp at 66x mean coverage, Psyllipsocidae: *Dorypteryx domestica* 18,512+ bp at 320x mean coverage), Troctomorpha (Amphientomidae: *Stimulopalpus japonicus* 14,904bp) , and Psocomorpha (Archipsocidae: *Archipsocus nomas* 15,349bp, Lachesillidae: *Lachesilla anna* 16,236bp Psocidae: *Amphigerontia montivaga* 15,566+ bp). Additionally, partial mitochondrial genomes were sequenced for two additional families of Trogiomorpha, Trogiidae (*Trogium pulsatorium*) and Lepidopsocidae (*Echmepteryx hageni*), to confirm genome rearrangements previously reported in the latter family (Shao et al., 2003) (Supplementary Table S2).

These genomes were sequenced by a mix of methods including long-PCR followed by primer walking (Cameron, 2014b), direct NGS sequencing of extracted DNA (also known as genome skimming, Linard et al., 2015), and a combination of both methods. The control region (CR) of *Stimulopalpus* could not be amplified by long-PCR and a combination of PCR and NGS derived sequences allowed the sequences of genes flanking the CR to be determined for this species. The *trnI-trnM* genes of *Neotrogla* were assembled separately from the other mitochondrial genes using NGS approaches, and the two contigs could not be connected into a single

sequence. Therefore, the possibility that they are on separate mini-chromosomes or that they represent pseudogenes cannot be excluded. However, phylogenetic analyses of the *trnI* and *trnM* genes alone placed those from *Neotrogla* in the expected position, consistent with them being the functional copy of these genes in this species. In addition, homologous repeat units (38–42 or 66–73 bp/repeat, Supplementary Fig. S1) were identified at the 3' end of the *trnI-trnM* contig and the 5' end of the *trnQ...rrnS* contig. Therefore, it is very likely that these contigs are connected via this repeat region. No homologous sequence was detected between the 3' end of *rrnS* and the 5' end of *trnI*, except that both are AT rich. Repeat units of this size present known assembly problems for Illumina HiSeq reads, and it seems more likely that these regions failed to assemble rather than the two assembled contigs represent separate mini-chromosomes. Repeat units were also identified in the control regions of several other sequenced species including *Speleketor* (two repeat classes 20 x 30bp, 3 x 44bp respectively), *Stimulopalpus* (3x 108bp), *Lachesilla* (7x 121bp), and *Amphigerontia* (5x 149bp) (Supplementary Table S2), although none of these species failed to assemble into a single contig. Sequence level homology between repeat units in different taxa was not identified.

3.2. Genome Rearrangements

A total of seven genome arrangement types (1–6 and 6') were detected in free-living Psocodea, four of them (1–3, 5) for the first time (Fig. 1). Type 1, identified in *Prionoglaris* (Prionoglarididae) was identical to the ancestral Pancrustacean condition. Both *Neotrogla* (type 2) and *Speleketor* (type 3) possess unique tRNA rearrangements, but they share a novel rearrangement of *trnM* to between duplicated control regions. All species of non-prionoglarid trogiomorphs possess a complicated rearrangement involving 7 tRNAs and *cox2* (type 4), first identified in Lepidopsocidae (Shao et al., 2001ab), but now identified by our study as also occurring in Trogiidae and Psyllipsocidae. The rearranged tRNA block in non-prionoglarid trogiomorphs includes a novel boundary, *trnI-trnM* (Character 1: Fig. 1), that is also observed in *Neotrogla*. *Stimulopalpus* (type 5) also closely resembles the ancestral pancrustacean mitochondrial genome, with one tRNA inverted (*trnI*) and one tRNA transposed (*trnM/trnQ*). All species of Psocomorpha share a complicated

rearrangement of the genes *nad3*, *nad5*, and associated tRNAs (type 6). In addition, *Stenocaecilius* (type 6') likely has a secondary tRNA transposition (*trnE-trnS1*, character 17), but is otherwise the same as other psocomorphans. However, the tRNAs rearrangements between CR and *nad2* identified in other psocomorphans have not been sequenced for *Stenocaecilius*.

In addition to gene rearrangements, a couple of long non-coding regions were identified in *Neotroglia*: 97 bp between CR repeat units and *trnQ*, 96 bp between *trnQ* and *nad2*, and 255 bp between *nad4L* and *trnT*. The former two non-coding regions correspond to the prior positions of *trnI* and *trnM*, respectively, in the ancestral Pancrustacean mitochondrial genome, and may represent 'junk' DNA regions left over from the rearrangement event which resulted in the transposition of these genes. Evidence for this interpretation lies in the identification of a characteristic hair-pin structure similar to the anticodon arm of *trnI* within the 97 bp non-coding region between CR repeats and *trnQ* (Supplementary Fig. S2).

3.3. Mito-phylogenomics (Fig. 2)

The aligned DNA data matrix consisted of 15 360 bp in total length (11 436 bp for PCG and 3 294 for RNA: Supplementary Data S1), of which 1 077 bp of PCG and 684 bp of RNA data were excluded from the analyses because of highly unreliable alignment. Within the PCG data (after excluding unaligned sites), 7 023 sites were variable, of which 1 281 sites were phylogenetically informative. Within the RNA data, 1 545 sites were variable, of which 467 sites were phylogenetically informative. Within the aligned AA data, 2 261 of 3 453 total sites were variable, of which 510 sites were phylogenetically informative.

Plots of P-distance showed that homoplasies caused by multiple substitutions were not problematic for phylogenetic estimation, except for the 3rd codon position where the slope of plots seemed to plateau (Supplementary Fig. S3). Although significant codon heterogeneity was detected by chi-square test in all data sets ($p = 0.000$), comparisons of base composition suggested that there seemed no directional base composition biases causing artificial phylogenetic affinities (Fig. 2; Supplementary Table 3). Comparing datasets including versus excluding third codon positions and RNA genes, and using multiple inference methods, allowed us to

further test if these factors resulted in artefactual relationships or nodal support.

Trees estimated from six data sets, each with two taxon sets (including/excluding taxa with missing data), were all concordant. Only one exception was the placement of *Stenocaecilius* (the taxon with a large amount of missing data: Fig. 1): it was placed as sister to *Lachesilla* with high support values in almost all datasets, but was placed at the base of Psocomorpha by RNA data with very low support values (<50% bootstrap [BS] and posterior probability [PP]). *Stenocaecilius* lacked large amounts of data, including two rRNAs that occupied the largest proportion of the RNA dataset. Although support values for the placement of taxa with missing data were relatively low (66–97% BS and 77–100% PP: *Echmepteryx*, *Trogium*, and *Stenocaecilius* [except for the placement by RNA dataset discussed above]), almost all other branches were supported with very high support values (>99% BP and 100% PP). Therefore, there were almost no detectable differences caused by different data/taxon sets and analytical methods. The only exception concerned the monophyly of Prionoglarididae: the family was consistently recovered as a monophyletic group (Fig. 2), but its support values were significantly lower than other branches (Table 2), although there were no missing data in three prionoglarid taxa. The support values were high in combined PCG+RNA or in separated RNA analyses (over 80% BS and 100% PP) (Table 2). In contrast, when the PCG and amino-acid data were analyzed separately, monophyly of Prionoglarididae generally received lower support values (Table 2). Increasing the size of the data set generally increased support for this clade, as was evident by comparing the results from RNA or PCG to All. Exclusion of the highly homoplasious 3rd codon position did not change the results significantly (ex.3rd and PCG12 datasets: Table 2).

Monophyly of the suborder Trogiomorpha was robustly supported. The trogiomorphs excluding Prionoglarididae formed a clade (*Echmepteryx*–*Dorypteryx* clade), in which *Dorypteryx* placed to the sister of the rest (= infraorder Atropetae). The support values for the relationships among taxa within this clade were relatively low, most probably due to large amount of missing data in *Trogium* and *Echmepteryx*. *Stimulopalpus* was placed sister to Psocomorpha with high support values. *Stimulopalpus* was the only representative sampled here from the suborder Troctomorpha, so the monophyly of this suborder could not be tested. Monophyly of

the suborder Psocomorpha was robustly supported, with *Archipsocus* (Archipsocetae: Archipsocidae) sister to the rest of psocomorphans with high support values. *Stenocaecilius* (Caeciliusetae: Caeciliusidae) and *Lachesilla* (Homilopsocidea: Lachesillidae) formed a clade with high support values. The remaining three samples all belong to the Psocidae, and its monophyly was robust.

3.4. Estimation of the History of Rearrangements

A total of 28 characters (17 insertion and 11 deletion characters) were coded (Supplementary Data S2) from the observed mitochondrial genome arrangements (Fig. 1). Novel tRNA rearrangements observed between the CR and *cox1* in Psocomorpha were treated as missing data for *Stenocaecilius* (Fig. 1).

The most parsimonious reconstruction of the condition-based data matrix on the ML phylogenomic tree (Fig. 2) is shown in Fig. 3. The insertions (Character 1–17) contained very little homoplasy (CI = 0.94, RI = 0.98). Translocation of *trnM* was identified as a synapomorphy of *Neotrogla* and *Speleketor*. Both the *Echmepteryx*–*Dorypteryx* clade and the Psocomorpha were characterized by unique gene rearrangements, including a series of non-homoplasious characters (11 and 6 respectively). The pattern seen in *Stenocaecilius* (type 6') could be derived by a single tRNA transposition from the psocomorphan type (type 6). The derived gene boundary, *trnI-trnM*, was identified in both *Neotrogla* and the *Echmepteryx*–*Dorypteryx* clade (Character 1), but they were inferred to have independent origins. In comparison, the deletions (Characters A–K) were more homoplasious (CI = 0.79, RI = 0.92). Furthermore, although deletion of *trnM* from its ancestral position (Character B) was identified in almost all taxa except for *Prionoglaris* and was reconstructed to have occurred in the common ancestor of Psocodea, this interpretation is unlikely (see Discussion).

Reconstructing the pattern of genome rearrangements using the ALL dataset topology (Fig. 2) in TreeREx recovered the following events between the ancestral pancrustacean mitochondrial genome (including arrangement type 1 *Prionoglaris*) and the 6 derived conditions identified above (Figs 1 and 4):

A) a small TDRL involving a 4 gene block (CR to *trnM*) resulting in both the duplication of the CR, and the relative rearrangement of *trnQ* and *trnM* (TDRL I),

in the common ancestor of *Neotroglia* and *Speleketor* (Arrangement types 2 and 3);

B) three rearrangement events including an inversion of *trnI*, transposition of *trnQ*, and a TDRL of an 7 gene block (*trnI* to *trnC*) (TDRL II) in the branch leading to *Speleketor* (Arrangement type 3);

C) an enormous TDRL involving duplication of almost the entire mt genome (33 of 37 genes) and 14 separate block deletions ranging in size from 1 to 9 genes (65 – 5200bp deletions) (TDRL III) in the ancestors of the *Echmepteryx–Dorypteryx* clade (Arrangement type 4) ;

D) a single translocation of *trnM* in the common ancestor of Psocomorpha and Troctomorpha (*Stimulopalpus*) (Arrangement types 5 and 6)

E) a single inversion (*trnI*) in *Stimulopalpus* (Arrangement type 5);

F) two moderate sized TDRLs (TDRL IV 8 genes, TDRL V 4 genes) in the ancestors of the Psocomorpha (Arrangement types 6 and 6');

G) transposition of *trnE* in *Stenocaecilius* (Arrangement type 6').

In addition, there are two possible optimizations for the derived position of *trnM* in the clade *Stimulopalpus*+Psocomorpha (depicted by dotted line in Fig. 4). The transposition of *trnM* could have occurred in the ancestor of this clade or it could have transposed independently in *Stimulopalpus* and as part of the TDRL V event. The number of inferred random losses in TDRL V are the same (4) whether *trnM* was in the insect ancestral genome position or a derived position (*trnI-trnM-trnQ*) prior to this duplication.

4. Discussion

4.1. Mito-phylogenomics

The tree estimated from the mitochondrial genomic data agreed completely with those estimated previously from nuclear and mitochondrial Sanger gene sequencing (Fig. 2: Yoshizawa et al., 2006; Yoshizawa and Johnson, 2010, 2014). Most branches received 100% bootstrap support and posterior probability, except for branches that included taxa with missing data. Tree and support value differences from different data sets were also minimal. In some previous studies, the usefulness of

mitochondrial genomic data for estimating deep insect phylogeny has been questioned (e.g., Cameron et al., 2004 for interordinal relationships). However, for the case of our study of the free-living Psocodea (excluding Liposcelididae), the mitochondrial genome data seems to contain consistent signal for resolving deep phylogenetic relationships between and within suborders.

The only uncertainty and potential conflict with previous studies concerns the monophyly of Prionoglarididae. In a previous analysis, Prionoglaridae was recovered as a monophyletic group (Yoshizawa et al., 2006: fig. 2). However, most of the signal supporting its monophyly was from the nuclear *Histone 3* gene, in which the 3rd codon position shows extremely biased base composition (over 60% AT) for Prionoglarididae species compared to other trogiomorphans (20–34% AT in most cases: Yoshizawa and Johnson, 2010). Therefore, the monophyly of the family recovered in this prior analysis might be an artifact caused by the similarity of base composition. Subsequent analyses with denser taxon and/or gene sampling did not provide support for monophyly of Prionoglarididae (Yoshizawa et al., 2006: fig. 3; Yoshizawa and Johnson, 2014).

In the present analyses, the Prionoglarididae was consistently recovered as a monophyletic group (Fig. 2; Table 2). No obvious directional biases in substitution rate and base composition were identified in the present dataset (Fig. S3, Table S2). Although different datasets provided somewhat variable support values for this clade, they are consistently high. In addition, combining different datasets (e.g., PCG and RNA) provided increased support values (Table 2). Therefore, the mitochondrial data, including the highly variable 3rd codon position, seem to contain consistent signal supporting the monophyly of Prionoglarididae. Alternatively, although monophyly was also supported, support values for Prionoglarididae from the AA data were generally low (Table 2). This pattern of reduced nodal support for the same/highly similar topologies from mitochondrial AA datasets versus nucleotide coding of the same genes has been observed in other insect groups including Polyneoptera (Cameron et al. 2006), Orthoptera (Fenn et al. 2008), and Hymenoptera (Dowton et al. 2009a) and therefore is not surprising at the finer taxonomic scales considered in this study.

4.2. Mitochondrial gene rearrangements

Seven types of mitochondrial genome arrangement were identified in the free-living Psocodea studied here (the extensively rearranged and modified mitochondrial genomes of Liposcelididae and Phthiraptera were excluded) (1–6 and 6' in Fig. 1). Of them, the condition identified in *Stenocaecilius* (type 6') can be simply formed from the condition identified in all other Psocomorpha (type 6) by a single rRNA transposition (*trnE*), and thus is regarded here as its sub-category (Fig. 1). Although mitochondrial gene rearrangements are recognized as rare-genomic change events (Boore et al., 1998; Rokas and Holland, 2000) and widely held to not result in homoplasious convergences, a couple of homoplasies were also evident between closely related members of Psocodea. In the following, we evaluate their gene rearrangement history by comparing the results from two different analytical strategies.

4.2.1. Condition-based coding

The condition-based coding method here proposed can handle transpositions, inversions, and inverse-transpositions but cannot recognize more complicated TDRL events as it breaks them up into multiple observed novel gene-boundaries. Character coding and ancestral state reconstruction can be done without any specific mechanistic assumptions as to how genomes rearrange (e.g. the long-running discussion as to whether mitochondrial recombination occurs in animals or not: Mortiz et al. 1987; Dowton & Campbell 2001; Kraytsberg et al. 2004; Ma & O'Farrell, 2015), which can be an advantage of this method. However, if different assumptions about the cause of rearrangements are applied, two alternative character-coding strategies are possible, potentially allowing a test of those assumptions. If deletion and insertion are recognized as simultaneous or a single event (e.g., as would be the case for recombination within a single mitochondrial genome molecule), then either only insertion or only deletion events should be coded. The consequences of such an approach can be seen in Fig. 3, where the insertion and deletion events are separately coded and the utility of each signal type can be clearly assessed. If insertion and deletion are recognized as different evolutionary events (e.g., recombination between-molecules, which first causes an insertion, then a deletion follows; or as is proposed by the TDRL model), then both insertion and deletion

events may be coded. For example, the present analyses recovered possible remnants of the *trnI* and *trnM* genes in their ancestral position flanking *trnQ* in *Neotrogla* (Supplementary Fig. S2). This strongly suggests that the rearrangements in *Neotrogla* were not caused by within-molecule recombination, but rather that the insertions and deletions occurred as different evolutionary events.

However, the present results showed that inclusion of deletion characters for the ancestral state estimation is highly problematic, even if between-molecules recombination is an assumed mechanism of rearrangement. First, deletion events are more homoplasious, as has been demonstrated in other insect groups (e.g. Hymenoptera: Dowton et al., 2009b). If gene deletion is random with respect to the newly inserted and original copies, then a half of all deletion events should have occurred in the copy located at the original position. Aside from possibly a stretch of non-coding DNA, deletions of newly inserted genes will not leave any evidence of gene transposition, whereas deletions at the original location will always leave evidence of gene transposition in the form of a novel gene boundary between the genes flanking the deleted one. In addition, while there are 36 possible positions for gene insertions, we observe that some genes rearrange considerably more frequently than others, and thus deletions will cluster on these more mobile genes. For instance within the present set of taxa, *trnM* is rearranged in 5 of the 6 genome arrangement types, resulting in 5 instances of the deletion character state B. These heightened rates of transposition by particular tRNAs have been observed in other taxa giving rise to recognition of rearrangement hotspots (e.g. Dowton and Austin, 1999; Dowton et al., 2003) which are also recognized as sites of convergent rearrangements (Dowton et al., 2009b). Therefore, it is obvious that deletion events at the ancestral location are far more frequently observed than convergent insertion events and thus are more homoplasious.

Second and more importantly, homoplasies of deletion characters sometimes can cause very unlikely ancestral state reconstructions. Under both the between-molecules recombination and TDRL scenarios, an insertion event must precede the deletion event. However, for example, as seen in Fig. 3, deletion of *trnM* from between *trnQ* and *nad2* (Character B) is most parsimoniously interpreted to have occurred in the common ancestor of the Psocodea, which was followed by insertions

of *trnM* at multiple different positions in different psocodean lineages: Characters 2, 8, 12 and 15, and reinsertion at its ancestral position in *Prionoglaris* (i.e. reversal of Character B). Therefore, for the most highly supported mitochondrial genome rearrangement models (i.e. between-molecules recombination, TDRL), insertion-only coding provides more accurate ancestral state estimation. If one needs to count the number of actual evolutionary events in the genomic history of a given group, then this can be accomplished by simply doubling the number of insertion events, because deletion events inevitably occurred following the corresponding insertion events.

The mitochondrial genome arrangement of *Prionoglaris* retains the ancestral pancrustacean condition (Fig. 1). Focusing only on the insertion events (i.e., excluding Characters A–K in Fig. 3), four of the five recorded types of novel genome arrangement (2–6 in Fig. 1) were identified as originated independently from the ancestral pancrustacean mitochondrial genome. The majority (10 of 17) of insertion characters are thus autapomorphies. Character 1 (*trnI-trnM*) was homoplasious: it is shared by *Neotrogla* and the *Dorypteryx–Echmepteryx* clade, but their independent origins are quite obvious from the radically different genomic location of the *trnI-trnM* gene pair in these taxa (middle of the CR versus poly-tRNA block between *cox3* and *cox2*, respectively). Only one character (Character 2: *trnM*-CR) was interpreted as a synapomorphic change that groups taxa of different gene arrangement types (2 and 3 in Fig. 1), suggesting multiple rounds of gene rearrangement through time, rather than direct rearrangement from the ancestral pancrustacean mitochondrial genome to the arrangement type seen in these extant genera. Character 2 also supports the close relationship between *Neotrogla* and *Speleketor* (currently grouped in the subfamily Speleketorinae: Lienhard, 2010). Finally, four synapomorphic insertions are identified in the common ancestor of the Psocomorpha (type 6), with only one psocomorphan lineage (*Stenocaecilius*, type 6') having a subsequent rearrangement (Character 17). The type 6' condition was also confirmed recently in a species of Stenopsocidae (*Stenopsocus immaculatus*: Liu et al., 2017), a member of the infraorder Caeciliusetae to which *Stenocaecilius* (Caeciliusidae) is also classified. Therefore, translocation of *trnE* may represent an autapomorphy of the infraorder. The mitochondrial genome of the common ancestor of Psocodea is thus estimated to have retained the pancrustacean ancestral condition. It is also evident from this result

that the extensive rearrangements observed in Psocodea and Thysanoptera have thus occurred independently.

4.2.2. TreeREx analyses

TreeREx software considers tandem-duplication-random-loss (TDRL) as well as transpositions, inversions, and inverted-transpositions (termed ‘reverse-transpositions’ in the software, however this is less precise and can be misinterpreted as transpositions back to an ancestral gene position, i.e. a character reversal in the cladistics sense). Estimation of TDRL events is much harder to recover without the aid of software like CREx or TreeREx (Bernt et al., 2007, 2008). TDRL events cannot be coded using the condition-based coding method. Because of this difference, the rearrangement histories estimated from the condition-based coding and TreeREx analyses are quite different. However, by both estimations, each type of genome arrangement identified in the free-living Psocodea originated via unique history. The mitochondrial genome arrangement of the common ancestor of Psocodea was estimated to retain the ancestral pancrustacean condition also by TreeREx. By using the condition-based matrix, a single transposition event (Character 2: *trnM*) was identified as synapomorphic between *Speleketor* and *Neotrogla*, and TreeREx also recovered a shared TDRL event between them. TreeREx identified that, from the ancestral condition of *Neotrogla* and *Speleketor* (type 2), the condition of *Speleketor* (type 3) was established by one inversion (*trnI*), one transposition (*trnQ*), and one TDRL (TDRL II: Fig. 1) (see Results). However, the arrangement of *Speleketor* can also be achieved by transposition of *trnC* and inverted-transposition of *trnI* only, without any TDRL event. The former less parsimonious output may potentially be caused by incomplete input data: i.e., duplicated CR in *Speleketor* and *Neotrogla* not coded (Supplementary Data S3).

In contrast, while the condition-based analysis did not recover any shared rearrangement event between *Stimulopalpus* and Psocomorpha, TreeREx recovered a transposition of *trnM* as a shared event. However, there is also an equally parsimonious scenario: occurrence of transposition of *trnM* in *Stimulopalpus*, while TDRL V from the ancestral insect genome arrangement (Fig. 1) in Psocomorpha can also explain the final arrangement types with exactly the same numbers of

transposition (1), tandem duplication (1) and independent loss (4) events.

4.2.3. Conclusion

Both methods (character-based coding and TreeREx) provided similar conclusions for the ancestral states of the mitochondrial genome arrangement (Figs 3–4). The effectiveness of these methods cannot be compared directly (e.g. comparing identified number of events by a parsimony criterion) because the different methods use different assumptions for the mechanism of mitochondrial gene rearrangements. Nevertheless, as mentioned above, incorporation of deletion characters into the condition-based matrix involves higher risk of inferring incorrect historical reconstructions and thus should be avoided regardless of the assumed evolutionary mechanisms. The character-based coding method is straightforward, and the constructed matrix can be used directly for ancestral state reconstruction, which provided quite reasonable conclusions in the present case. Each character in the matrix can be considered as an evolutionary event so that the data matrix constructed by the condition-based coding can also be used for phylogenetic estimation. A drawback of the condition-based coding is that it cannot handle TDRL events.

In contrast, TreeREx considers TDRL as well and estimates the rearrangement history directly from the gene order data, without specific character coding. The present analyses, however, recovered some potential flaws of the present TreeREx algorithm. First, TreeREx does not allow the existence of duplicated gene in the input data. Possibly because of this, an apparently less-parsimonious interpretation was obtained for the rearrangement history of *Speleketor*. In addition, TreeREx only outputs a single result, even if there are equally parsimonious possibilities (TreeREx output ACCTAN-type reconstruction, although DELTRAN-type reconstruction is also possible for the transposition of *trnM* in *Stimulopalpus* and Psocomorpha: Fig. 4). Such possibilities must be manually examined based on the phylogenetic relationships and TreeREx output.

Plausibility of different mechanistic assumptions should also be evaluated, not only by parsimony criterion, but also by detailed mitochondrial genome analyses, with dense taxon sampling and strong phylogenetic backbone. Previous evidence has

622 favored the TDRL model (Dowton et al., 2009; Beckenbach, 2011) but, in the present
623 case, the inversion of *trnI* cannot be explained by TDRL. Alternatively, the presence
624 of a potential *trnI* remnant in *Neotroglia* cannot be explained by within-molecule
625 recombination. The between-molecule recombination model can explain both, but
626 this does not overwhelmingly favor that model because each rearrangement event
627 might have been caused by different mechanisms. The present study showed that
628 more highly rearranged mitochondrial genomes can still be quite consistent within
629 higher taxa (i.e., *Echmepteryx*–*Dorypteryx* clade which includes all trogiomorphs
630 except Prionoglarididae, and Psocomorpha from which all major clades were
631 sampled). Therefore, their intermediate genome arrangements cannot be recovered
632 from the extant species. In contrast, variation was identified within the family
633 Prionoglarididae. Only three representatives of Prionoglarididae were included in the
634 present analyses, and there are more genera not analyzed here (e.g., *Sensitibilla* and
635 *Afrotriglia* considered close to *Neotroglia*, and *Siamoglaris* and *Speleopsocus*
636 considered close to *Prionoglaris*) each of which includes multiple species (except for
637 the monotypic *Speleopsocus*). In addition, only a single species (*Stimulopalpus*) was
638 analyzed from the primitive members of the suborder Troctomorpha (i.e., excluding
639 highly derived Liposcelididae), although there are seven more families in this group.
640 Analyses of these taxa may provide further clues to evaluate mitochondrial
641 rearrangement history and mechanisms in the Psocodea.

643 **Acknowledgments**

645 We thank Nico Schneider for supplying valuable specimens. This project was
646 supported by JSPS Grant (15H04409) to KY, the US National Science Foundation
647 (DEB0444972), CSIRO Julius Career Awards, and the Australian Research Council Future
648 Fellowships scheme (FT120100746) to SLC, and NSF DEB-1239788 and DEB-1342604 to
649 KPJ.

651 **References**

- Allen, J.M., Huang, D.I., Cronk, Q.C., Johnson, K.P., 2015. aTRAM – automated target restricted assembly method: a fast method for assembling lici across divergent taxa from next-generation sequencing data. *BMC Bioinformatics* 16, 98.
- Andujar, C., Arribas, P., Linard, B., KUndrata, R., Bocak, L., Vogler, A.P., 2017. The mitochondrial genome of *Iberobaenia* (Coleoptera: Iberobaeniidae): first rearrangement of protein-coding genes in the beetles. *Mitochondrial DNA Part A* 28, 156–158.
- Beckenbach, A.T., 2011. Mitochondrial genome sequences of representatives of three families of scorpionflies (Order Mecoptera) and evolution in a major duplication of coding sequence. *Genome* 54, 368–376.
- Beckenbach, A.T., 2012. Mitochondrial genome sequences of Nematocera (Lower Diptera): evidence of rearrangement following a complete genome duplication in a winter crane fly. *Genome Biol. Evol.* 4, 89–101.
- Bern, M., Merkle, D., Ramsch, K., Fritsch, G., Perseke, M., Bernhard, D., Schlegel, M., Stadler, P.F., Middendorf, M., 2007. CREx: inferring genomic rearrangements based on common intervals. *Bioinformatics* 23, 2957–2958.
- Bernt, M., Merkle, D., Middendorf, M., 2008. An algorithm for inferring mitogenome rearrangements in a phylogenetic tree. In: Nelson, C.E., Vialette, S. (Eds.), *Comparative Genomics. RECOMB-CG 2008. Lecture Notes in Computer Science*, vol. 5267. Springer, Berlin, pp. 143–157.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved *de novo* metaxoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313–319.
- Boore, J.L., Lavrov, D.V., Brown, W.M., 1998. Gene translocation links insects and crustaceans. *Nature* 392, 667–668.
- Boore JL. 2000. The duplication random-loss model for gene rearrangement exemplified by mitochondrial genomes of deuterosome animals. In *Comparative Genomics: Empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families*, eds D Sankoff, JH Nadeau, pp. 133–48. Dordrecht: Kluwer Academic Publishers.
- Cameron, S.L., 2014a. Insect mitochondrial genomics: implications for evolution and phylogeny. *Ann. Rev. Entomol.* 59, 95–117.

685 Cameron, S.L., 2014b. How to sequence and annotate insect mitochondrial genomes for
686 systematic and comparative genomics research. *Syst. Entomol.* 39, 400–411.

687 Cameron, S.L., Miller, K.B., D’Haese, C.A., Whiting, M.F., Barker, S.C., 2004.
688 Mitochondrial genome data alone are not enough to unambiguously resolve the
689 relationship of Entognatha, Insecta and Crustacea. *Cladistics* 20, 534–557.

690 Cameron, S.L., Barker, S.C. & Whiting, M.F. 2006. Mitochondrial genomics and the
691 relationships and validity of the new insect order Mantophasmatodea. *Mol.*
692 *Phylogenet. Evol.* 38: 274-279.

693 Cameron, S.L., Yoshizawa, K., Mizukoshi, A., Whiting, M.D., Johnson, K.P., 2011.
694 Mitochondrial genome deletions and mini-circles are common in lice (Insecta:
695 Phthiraptera). *BMC Genomics* 12: 394.

696 Chen, S.C., Wei, D.D., Shao, R., Shi, J.X., DouW., Wang, J.J., 2014. Evolution of
697 multipartite mitochondrial genomes in the booklice of the genus *Liposcelis*
698 (Psocoptera). *BMC Genomics* 15, 861.

699 Dickey, A.M., Kumar, V., Morgan, J.K., Jara-Cavieres, A., Shatters Jr., R.G., Mckenzie,
700 C.L., Osborne, L.S., 2015. A novel mitochondrial genome architecture in thrips
701 (Insecta: Thysanoptera): extreme size asymmetry among chromosomes and possible
702 recent control region duplication. *BMC Genom.* 16, 439.

703 Dowton, M., Austin, A.D., 1999. Evolutionary dynamics of a mitochondrial rearrangement
704 "hot spot" in the Hymenoptera. *Mol. Biol. Evol.* 16, 298–309.

705 Dowton M, Campbell NJH. 2001. Intramitochondrial recombination – is it why some
706 mitochondrial genes sleep around? *Trends Ecol. Evol.* 16: 269-271.

707 Dowton, M., Castro, L.R., Campbell, S.L., Bargon, S.D., Austin, A.D., 2003. Frequent
708 mitochondrial gene rearrangement at the hymenopteran nad3–nad5 junction. *J. Mol.*
709 *Evol.* 56, 517–526.

710 Dowton, M., Cameron, S.L, Austin, A.D. & Whiting, M.F. 2009. Phylogenetic approaches
711 for the analysis of mitochondrial genome sequence data in the Hymenoptera – a
712 lineage with both rapidly and slowly evolving mitochondrial genomes. *Mol.*
713 *Phylogenet. Evol.* 52: 512-519

714 Dowton, M., Cameron, S.L., Dowavic, J.I., Austin, A.D., Whiting, M.F., 2009b.
715 Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera

suggests that mitochondrial tRNA gene position is selectively neutral. *Mol. Biol. Evol.* 26, 1607–1617.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

Fenn, J.D., Song, H., Cameron, S.L. & Whiting, M.F. 2008. A mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol. Phylogenet. Evol.* 49: 59–68.

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nuc. Acid. Res.* 41, e129.

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software version 7: improvement in performance and usability. *Mol. Biol. Evol.* 30, 772–780.

Kômoto, N., Yukuhiro, K., Tomita, S., 2012. Novel gene rearrangements in the mitochondrial genome of a webspinner, *Aposthonia japonica* (Insecta: Embioptera). *Genome* 55, 222–233.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874.

Kraytsberg, Y., Schwartz, M., Brown, T.A., Ebralidse, K., Kunz, W.S., Clayton, D.A., Vissing, J., Khrapko, K. 2004. Recombination of human mitochondrial DNA. *Science* 304: 981.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773.

Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–2288.

Li, H., Liu, H., Shi, A., Stys, P., Zhou, X., Cai, W. 2012. The complete mitochondrial genome and novel gene arrangement of the unique headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). *PLoS ONE* 7, 29419.

Li, H., Shao, R., Song, F., Zhou, X., Yang, Q., Li, Z., Cai, W., 2013. Mitochondrial genomes of two barklice, *Psococerastis albimaculata* and *Longivalvus hyalospilus* (Psocoptera: Psocomorpha): contrasting rates in mitochondrial gene rearrangement between major lineages of Psocodea. *PLoS ONE* 8, e61685.

748 Lienhard, C., 1998. Psocoptères euro-méditerranéens. Faune de France 83, Fédération
749 Française des Sociétés de Sciences naturelles, Paris.

750 Lienhard, C., 2010. A new genus of Sensitibillini from Brazilian caves (Psocodea:
751 ‘Psocoptera’: Prionoglarididae). Rev. Suisse Zool. 117, 611–635.

752 Linard, B., Crampton-Platt, A., Gillett, C.P.D.T., Timmermans, M.J.T.N., Vogler, A.P., 2015.
753 Metagenome skimming of insect specimen pools: potential for comparative genomics.
754 Genome Biol. Evol. 7, 1474–1489.

755 Liu, H., Li, H., Cai, Y., Song, F., Wilson, J.J., Cai, W., 2017. Conserved gene arrangement in
756 the mitochondrial genomes of barklouse families Stenopsocidae and Psocidae. Front.
757 Agr. Sci. Eng. 4, 358–365.

758 Liu, H., Li, H., Song, F., Gu, W., Feng, J., Cai, W., Shao, R., 2017. Novel insights into
759 mitochondrial gene rearrangement in thrips (Insecta: Thysanoptera) from the grass
760 thrips, *Anaphothrips obscurus*. Sci Rep 7, 4284.

761 Ma, H., O’Farrell, P.H. 2015. Selections that isolate recombinant mitochondrial genomes in
762 animals. eLife 4: e07247.

763 Maddison, D.R., Maddison, W.P., 2000. MacClade 4: Analysis of Phylogeny and Character
764 Evolution. Sinauer Assoc., Sunderland, MA.

765 Mao, M., Gibson, T., Dowton, M., 2015. Higher-level phylogeny of the Hymenoptera
766 inferred from mitochondrial genomes. Mol. Phylogenet. Evol. 84, 34–43.

767 Moritz C., Dowling, T.E., Brown, W.M., 1987. Evolution of animal mitochondrial DNA:
768 relevance for population biology and systematics. Ann. Rev. Ecol. Syst. 18: 269-292.

769 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and
770 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol.
771 Biol. Evol. 32, 268–274.

772 Rokas, A., Holland, P.W.H., 2000. Rare genomic changes as a tool for phylogenetics. Trends
773 Ecol. Evol. 15, 454–459.

774 Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under
775 mixed model. Bioinformatics 19, 1572–1574.

776 Shao, R., Barker, S.C., 2003. The highly rearranged mitochondrial genome of the plague
777 thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene
778 boundaries and an extraordinary arrangement of rRNA genes. Mol. Biol. Evol. 20,
779 362–370.

780 Shao, R., Campbell, N.J.H., Barker, S.C., 2001a. Numerous gene rearrangements in the
781 mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera).
782 Mol. Biol. Evol. 18, 858–865.

783 Shao, R., Campbell, N.J.H., Schmidt, E.R., Barker, S.C., 2001b. Increased rate of gene
784 rearrangement in the mitochondrial genomes of three orders of hemipteroid insects.
785 Mol. Biol. Evol. 18, 1828–1832.

786 Shao, R., Dowton, M., Murrell, A., Barker, S.C., 2003. Rates of gene rearrangement and
787 nucleotide substitution are correlated in the mitochondrial genomes of insects. Mol.
788 Biol. Evol. 20, 1612–1619.

789 Shao, R., Kirkness, E.F., Barker, S.C., 2009. The single mitochondrial chromosome typical
790 of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus*
791 *humanus*. Genome Res. 19, 904–912.

792 Shao, R., Barker, S.C., Li, H., Song, S., Poudel, S., Su, Y., 2015. Fragmented mitochondrial
793 genomes in two suborders of parasitic lice of eutherian mammals (Anoplura and
794 Rhynchophthirina, Insecta). Sci. Rep. 5, 17389.

795 Shao, R., Li, H., Barker, S.C., Song, S., 2017. The mitochondrial genome of the guanaco
796 louse, *Microthoracius praelongiceps*: insights into the ancestral mitochondrial
797 karyotype of sucking lice (Anoplura, Insecta). Genome Biol. Evol. 9, 431–445.

798 Shi, Y., Chu, Q., Wei, D.D., Qiu, Y.J., Shang, F., Dou, W., Wang, J.J., 2016. The
799 mitochondrial genome of booklouse, *Liposcelis sculptilis* (Psocoptera: Liposcelididae)
800 and the evolutionary timescale of *Liposcelis*. Sci. Rep. 6, 30660.

801 Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, I., 2009. ABySS:
802 A parallel assembler for short read sequence data. Genome Res. 19, 1117–1123.

803 Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other
804 Methods). Version 4. Sinauer Assoc., Sunderland, MA.

805 Yan, D., Tang, Y., Hu, M., Liu, F., Zhang, D., Fan, J., 2012. The mitochondrial genome of
806 *Frankliniella intonsa*: Insights into the evolution of mitochondrial genomes at lower
807 taxonomic levels in Thysanoptera. Genomics 104, 306–312.

808 Yoshizawa, K., Johnson, K.P., 2006. Morphology of male genitalia in lice and their relatives
809 and phylogenetic implications. Syst. Entomol. 31, 350–361.

810 Yoshizawa, K., Johnson, K.P., 2010. How stable is the “Polyphyly of Lice” hypothesis?: a
811 comparison of phylogenetic signal in multiple genes. *Mol. Phylogenet. Evol.* 55, 939–
812 951.

813 Yoshizawa, K., Johnson, K.P., 2014. Phylogeny of the suborder Psocomorpha (Insecta:
814 Psocodea: 'Psocoptera'): congruence and incongruence between morphology and
815 molecules. *Zool. J. Linn. Soc.* 171, 716–731.

816 Yoshizawa, K., Lienhard, C., 2010. In search of the sister group of the true lice: A systematic
817 review of booklice and their relatives, with an updated checklist of Liposcelididae
818 (Insecta: Psocodea). *Arthropod Syst. Phylog.* 68, 181–195.

819 Yoshizawa, K., Lienhard, C., 2016. Bridging the gap between chewing and sucking in the
820 hemipteroid insects: new insights from Cretaceous amber. *Zootaxa* 4079, 229–245.

821 Yoshizawa, K., Lienhard, C., Johnson, K.P., 2006. Molecular systematics of the suborder
822 Trogiomorpha (Insecta: Psocodea: 'Psocoptera'). *Zool. J. Linn. Soc.* 146, 287–299.
823

Captions

Fig. 1. Seven types of the mitochondrial gene arrangements detected from "Psocoptera". Numbers indicate novel gene boundary possibly caused by insertion events, whereas alphabets indicate possible deletion events (condition-based coding: see Fig. 3). Red dotted lines under genome map indicate tandem-duplication-random-loss events (TDRL) identified by TreeREx analysis (see Fig. 4).

Fig. 2. Mitochondrial phylogeny of the "Psocoptera" estimated from ALL dataset. Numbers associated with branches indicate bootstrap/posterior probability values estimated from this data set. Support values for Prionoglarididae estimated from other datasets are provided in Table 2.

Fig. 3. Most parsimonious reconstruction of the condition-based coding data of the mitochondrial gene arrangements. Numbers (gain condition, filled square or triangle) and alphabets (loss condition, open square or triangle) on branches corresponds those scored in Fig.1. Square indicates non-homoplasious condition whereas triangle indicates homoplasious condition. Numbers associated to taxa corresponds the gene arrangement types in Fig. 1.

Fig. 4. Gene rearrangement history as estimated by TreeREx software. See Result section for detailed rearrangement events. Equally parsimonious interpretations are indicated by gray dotted line. A–E correspond to evolutionary events discussed in the text. Abbreviations: Inv.–inversion; TD–tandem duplication; Trans–transposition.

Table 1. List of taxa analyzed in this study, with GenBank accession numbers

Table 2. Support values for Prionoglarididae estimated from different gene and taxon sets with different analytical methods.

Supplements

Data S1. Nexus file of aligned mitochondrial data.

Data S2. Nexus file of the condition-base coding data of gene arrangements.

Data S3. Input data for the TreeREx analysis. Taxa showing the identical genome arrangement were treated as a single terminal taxon.

856 Table S1. Primers used for long PCR.
857 Table S2. Gene annotations.
858 Table S3. AT-content of each gene/taxon.
859 Fig. S1. Repeat units between *trnM* and *-trnQ* of *Neotroglia*.
860 Fig. S2. Hairpin structure between *trnI* anticodon arm and potential *trnI* ruminant
861 detected in *Neotroglia*.
862 Fig. S3. Plots of p-distance calculated from different data sets (taxa with missing data
863 excluded)

(1) *Prionoglaris* = Ancestral Condition of Pancrustacea

(2) *Neotroglia*

(3) *Speleketor*

(4) *Trogiomorpha* ex. *Prionoglarididae*

(5) *Stimulopalpus*

(6) *Psocomorpha* ex. *Stenocaecilius*

(6') *Stenocaecilius*

Table 1. List of taxa analyzed in this study, with GenBank accession numbers

Order	Suborder	Family	Genus	Species	Locality	GenBank #
Psocodea	Trogiomorpha	Prionoglarididae	<i>Prionoglaris</i>	<i>stygia</i>	Luxembourg	given upon acceptance
			<i>Neotrogl</i>	sp.	Brazil: Minas Gerais	given upon acceptance
			<i>Speleketor</i>	<i>irwini</i>	USA: California	given upon acceptance
		Psyllipsocidae	<i>Dorypteryx</i>	<i>domestica</i>	Switzerland: Geneva	given upon acceptance
		Trogiidae	<i>Trogium</i>	<i>pulsatorium</i>	United Kingdom: Sussex	given upon acceptance
		Lepidopsocidae	<i>Genus</i>	sp.	GenBank	NC004816
			<i>Echmepteryx</i>	<i>hageni</i>	USA: Illinois	given upon acceptance
	Troctomorpha	Amphientomidae	<i>Stimulopalpus</i>	<i>japonicus</i>	USA: Illinois	given upon acceptance
	Psocomorpha	Archipsocidae	<i>Archipsocus</i>	<i>nomas</i>	USA: Florida	given upon acceptance
		Caeciliusidae	<i>Stenocaecilius</i>	<i>quercus</i>	GenBank	AH010776.3
		Lachesillidae	<i>Lachesilla</i>	<i>anna</i>	USA: Illinois	given upon acceptance
		Psocidae	<i>Amphigerontia</i>	<i>montivaga</i>	USA: Arizona	given upon acceptance
			<i>Psococerastis</i>	<i>albimaculata</i>	GenBank	JQ910989
			<i>Longivalvus</i>	<i>hyalospilus</i>	GenBank	JQ910986
Hemiptera	Auchenorrhyncha	Cercopidae	<i>Abidama</i>	<i>producta</i>	GenBank	GQ337955
Megaloptera	–	Corydalidae	<i>Dysmicohermes</i>	<i>ingens</i>	GenBank	KJ806318

taxon\data set	All	ex.3rd	RNA	PCG	PCG12	AA
MrBayes_Full	100	100	100	99.4	98.1	98.8
exMissing	100	100	100	99.6	98.7	100
IQtree_Full	90	93	88	72	70	56
exMissing	90	91	81	75	75	70
PhyloBayes_Full	99	99	99	86	80	62

```

-1      CR-trnI-trnM-GRAATDAAGCAGGAATAA-TA---T
-2 AAAGGGGMATADTATTAKGAATGAAGCAGGAMTAA-YA---T
-3 AAAGGGGCATARTATTAGGAATGAAGCAGGAATAA-T-----
-4 -----GGCATAGTATTAGRAATGAAGCAGGAMTAACTAATAT
-5 AAAGGGGAATATTAYTATGAATGAAGCAGGAMTAA-TGRCAT

6 AAAGGGGCATAGTATTAGGAATGAAGCAGGAATAA-TA---T
5 AAAGGGGCATAGTATTAGGAATGAAGCAGGAATAACTAATAT
4 AAAGGGGAATATTATTATGAATGAAGCAGGACTAA-TA-CAT
3 AAAGGGGCATAGTATTAGGAATGAAGCAGGAATAA-TA---T
2 AAAGGGGCATARTATTAGGAATGAAGCAGGAATAA-TA---T
1 AAAGGGGCATAGTATTAGGAATGAAGCAGGACTA-CTA-TAT
- [97 bp of non-coding region
  not homologous to the repeat units] -trnQ.

```

repeat8: L-unit (* and #: repeats within L-unit)

```

*****##### ***** #####

```

```

AAAGGGGCATAGTATGAAGCAGGAATAATGGCATAGTATTAGGAATGAAGCAGGAATAACTAATAT
AAAGG-----GGCATAGTATTAGGAATGAAGCAGGAATAACTAATAT

```

repeat5: S-unit

anticodon arm of *tRNA*

potential remnant of *tRNA*

Table S1a. Long PCR Primers, sequence and location, for *Stimulopalpus japonicus*.

Region	Primer Pair (F & R)	Sequence (5' →3')
Long PCRs		
<i>trnM</i> → <i>cox1</i>	PSOC4 ¹	AAG CTW WTG GGY TCA TAC CYC
	STJA35 ²	TTA ATC CCT GTA GGG ATA GC
<i>cox1</i> → <i>cox3</i>	C1-J-1718 ³	GGA GGA TTT GGA AAT TGA TTA GTT CC
	C3-N-5460 ³	TCA ACA AAG TGT CAG TAT CA
<i>cox3</i> → <i>nad4</i>	STJA2 ⁵	TCA AGG ATT TGA ATA TTG AGA AGC
	STJA3 ⁵	TCA GCC TGA GCG AAT TCA GGC TGG
<i>nad4</i> → <i>cytB</i>	N4-J-8944 ³	GGA GCT TCA ACA TGA GCT TT
	cobR ⁴	GCA TAA GCA AAT AAA AAA TAT CAT TC
<i>cytB</i> → <i>rrnL</i>	STJA6 ²	ATT GAT AAA ATC CCA TTC CAT CC
	STJA7 ²	TTT AAT AAG GGA CGA GAA GAC CC
<i>rrnL</i> → <i>rrnS</i>	16SB ⁵	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594 ⁶	AAA CTA GGA TTA GAT ACC C

¹ Primers designed from consensus sequences, for general amplification of Psocoptera

² Primers specifically designed for sequencing this genome

³ Primers taken from Simon *et al.* (1994)

⁴ Primers taken from Whiting (2002)

⁵ Primers taken from Bybee *et al.* (2004)

⁶ Primer taken from Skerratt *et al.* (2002)

Table S1b. Long PCR Primers, sequence and location, for *Amphigerontia montivaga*.

Region	Primer Pair (F & R)	Sequence (5' →3')
Long PCRs		
<i>cox1</i> → <i>cox3</i>	C1-J-1718 ³	GGA GGA TTT GGA AAT TGA TTA GTT CC
	C3-N-5460 ³	TCA ACA AAG TGT CAG TAT CA
<i>cox3</i> → <i>rrnL</i>	AMMO4 ²	TGC CGA TTC AAT TTA TGG ATC GTC G
	AMMO5 ²	TTA AAA GAC GAG AAG ACC CTA TAG
<i>rrnL</i> → <i>rrnS</i>	16SB ⁵	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594 ⁶	AAA CTA GGA TTA GAT ACC C
<i>rrnS</i> → <i>cox1</i>	AMMO8 ²	TAG AAA GAG AAT GAC GGG CAA TAT G
	AMMO1 ²	ATC AAC TGA TGC TCC TGT ATG TCC

¹ Primers designed from consensus sequences, for general amplification of Psocoptera

² Primers specifically designed for sequencing this genome

³ Primers taken from Simon *et al.* (1994)

⁴ Primers taken from Whiting (2002)

⁵ Primers taken from Bybee *et al.* (2004)

⁶ Primer taken from Skerratt *et al.* (2002)

Table S1c. Long PCR Primers, sequence and location, for *Lachesilla anna*.

Region	Primer Pair (F & R)	Sequence (5' →3')
Long PCRs		
<i>cox2</i> → <i>nad4</i>	FLeu ⁴	TCT AAT ATG GCA GAT TAG TGC
	LAAN1 ⁵	TTG TTT AAA AGA GTA GGT TCC TCC
<i>nad4</i> → <i>cytB</i>	N4-J-8944 ³	GGA GCT TCA ACA TGA GCT TT
	cobR ⁴	GCA TAA GCA AAT AAA AAA TAT CAT TC
<i>cytB</i> → <i>rrnL</i>	LAAN4 ²	TTG ATA AAG CCT CTT TTC ATC CC
	LAAN5 ²	TTA AAA GAC GAG AAG ACC CTA TAG
<i>rrnL</i> → <i>rrnS</i>	16SB ⁵	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594 ⁶	AAA CTA GGA TTA GAT ACC C
<i>rrnS</i> → <i>cox2</i>	LAAN8 ²	AGA GAA TGA CGG GCA ATA TGT GC
	LAAN11 ²	ACA AAA TAC GGA GGG AAG GTA GGG C

¹ Primers designed from consensus sequences, for general amplification of Psocoptera

² Primers specifically designed for sequencing this genome

³ Primers taken from Simon *et al.* (1994)

⁴ Primers taken from Whiting (2002)

⁵ Primers taken from Bybee *et al.* (2004)

⁶ Primer taken from Skerratt *et al.* (2002)

Table S1d. Long PCR Primers, sequence and location, for *Archipsocus nomas*.

Region	Primer Pair (F & R)	Sequence (5' →3')
Long PCRs		
<i>trnM</i> → <i>cox1</i>	ARNO7 ²	ACG TTT TTT TCA ATT TTA CCC CGG
	RLys ⁴	GAG ACC AGT ACT TGC TTT CAG TCA TC
<i>cox2</i> → <i>nad4</i>	ARNO11 ²	TGC CCT TAC TGT CAA AAC TAT TGG TC
	ARNO19 ²	AAC CTA AAG GGT TGG AAG AAC CTG
<i>nad4</i> → <i>rrnL</i>	N4-J-8944 ³	GGA GCT TCA ACA TGA GCT TT
	ARNO3 ⁴	TTT ATG GCG AAT TTA ATT GGG GTG
<i>rrnL</i> → <i>rrnS</i>	16SB ⁵	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594 ⁶	AAA CTA GGA TTA GAT ACC C
<i>rrnS</i> → <i>trnM</i>	ARNO4 ²	ATA TTG CCA GTA AGA TAA TCG TGG
	TM-N-193 ³	TGG GGT ATG AAC CCA GTA GC

¹ Primers designed from consensus sequences, for general amplification of Psocoptera

² Primers specifically designed for sequencing this genome

³ Primers taken from Simon *et al.* (1994)

⁴ Primers taken from Whiting (2002)

⁵ Primers taken from Bybee *et al.* (2004)

⁶ Primer taken from Skerratt *et al.* (2002)

Table S1e. Long PCR Primers, sequence and location, for *Speleketor irwini*.

Region	Primer Pair (F & R)	Sequence (5' →3')
Long PCRs		
<i>trnM</i> → <i>cox1</i>	TM-J-206 ³	TGG GGT ATG AAC CCA GTA GC
	SPIR1 ²	AAG GAG GAT AGA CTG TTC ATC CTG
<i>cox1</i> → <i>cox3</i>	C1-J-1718 ³	GGA GGA TTT GGA AAT TGA TTA GTT CC
	C3-N-5460 ³	TCA ACA AAG TGT CAG TAT CA
<i>cox3</i> → <i>rrnL</i>	SPIR4 ²	ACT ATT ACA TGA GCT CAC CAT GCA C
	SPIR5 ²	TTT ACA TGG AAA GGG TAT TGA AGG
<i>rrnL</i> → <i>rrnS</i>	16SB ⁵	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594 ⁶	AAA CTA GGA TTA GAT ACC C
<i>rrnS</i> → <i>trnM</i>	SPIR6 ²	TAT AGT CTG CAC CTT GAC CTG AC
	TM-N-193 ³	TGG GGT ATG AAC CCA GTA GC

¹ Primers designed from consensus sequences, for general amplification of Psocoptera

² Primers specifically designed for sequencing this genome

³ Primers taken from Simon *et al.* (1994)

⁴ Primers taken from Whiting (2002)

⁵ Primers taken from Bybee *et al.* (2004)

⁶ Primer taken from Skerratt *et al.* (2002)

Table S1f. Long PCR Primers, sequence and location, for *Echmepteryx hageni* and *Trogium pulsatorium*.

Region	Primer Pair (F & R)	Sequence (5' →3')
Long PCRs		
<i>cox3</i> → <i>nad4</i>	PSOC1 ¹	TTG AAG CNG CWG CHT GRT AYT GAC
	PSOC2 ¹	AAR GCT CAT GTK GAR GCW CC

¹ Primers designed from consensus sequences, for general amplification of Psocoptera