Gene 768 (2021) 145312

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short communication

Structure, gene order, and nucleotide composition of mitochondrial genomes in parasitic lice from Amblycera

Andrew D. Sweet [a,b,*], Kevin P. Johnson [c], Yanghui Cao [c], Robert S. de Moya [c,d], Rachel K. Skinner [c,d], Milton Tan [c], Stephany Virrueta Herrera [c,e], Stephen L. Cameron [a]

[a] Department of Entomology, Purdue University, West Lafayette, IN, USA
[b] Department of Biological Sciences, Arkansas State University, State University, AR, USA
[c] Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL, USA
[d] Department of Entomology, University of Illinois, Urbana, IL, USA
[e] Program in Ecology, Evolution, and Conservation Biology, University of Illinois, Urbana, IL, USA

ARTICLE INFO

Keywords:
Evolution
Organelle genomics
AT-bias
Phylogeny

ABSTRACT

Parasitic lice have unique mitochondrial (mt) genomes characterized by rearranged gene orders, variable genome structures, and less AT content compared to most other insects. However, relatively little is known about the mt genomes of Amblycera, the suborder sister to all other parasitic lice. Comparing among nine different genera (including representative of all seven families), we show that Amblycera have variable and highly rearranged mt genomes. Some genera have fragmented genomes that vary considerably in length, whereas others have a single mt chromosome. Notably, these genomes are more AT-biased than most other lice. We also recover genus-level phylogenetic relationships among Amblycera that are consistent with those reported from large nuclear datasets, indicating that mt sequences are reliable for reconstructing evolutionary relationships in Amblycera. However, gene order data cannot reliably recover these same relationships. Overall, our results suggest that the mt genomes of lice, already know to be distinctive, are even more variable than previously thought.

1. Introduction

Most insects have a mitochondrial (mt) genome that consists of a single chromosome comprising 37 genes (13 protein-coding, 22 transfer RNAs (tRNAs), and two ribosomal RNAs (rRNAs)) that are common to other metazoans, as well as a conserved gene order (Boore, 1992; Chen and Butow, 2005). However, there are notable exceptions. Parasitic lice (Phthiraptera) have particularly variable mt genomes. All of the major clades of lice – Philopteridae (feather lice), Anoplura (sucking lice), Trichodectidae (mammal chewing lice), Rhynchophthirina (elephant lice), and Amblycera (bird and mammal chewing lice) – are known to have taxa with rearranged gene orders relative to the inferred ancestral insect mt genome (Covacin et al., 2006; Cameron et al., 2007, 2011). There are also at least 20 species of lice that have fragmented mt genomes, with the 37 genes separated onto multiple circular chromosomes (Shao et al., 2009; Song et al., 2019). All but four of these cases are from the clade composed of mammal lice in Anoplura, Rhynchophthirina, and Trichodectidae (Shao et al., 2001, 2012, 2015; Dong et al., 2014). Four species of lice in the feather louse genus *Columbicola* also have highly fragmented mt genomes, which indicates fragmentation has occurred multiple times within parasitic lice (Sweet et al., 2020). Additionally, Cameron et al. (2011) provided evidence that several other genera of feather lice also have fragmented mt genomes, suggesting fragmentation is more prevalent in lice than assumed previously.

Parasitic lice, along with their closest free-living relatives (booklice), also have less AT-biased mt genomes compared to other insects. Many insect mt genomes have a high AT% (an average of ~76 AT% across

Abbreviations: mt, mitochondrial; tRNA, transfer RNA; rRNA, ribosomal RNA; bp, base pairs; cox1, cytochrome oxidase subunit 1; cox2, cytochrome oxidase subunit 2; cox3, cytochrome oxidase subunit 3; atp6, ATP synthase F0 subunit 6; atp8, ATP synthase F0 subunit 8; nad1, NADH dehyrdrogenase subunit 1; nad2, NADH dehyrdrogenase subunit 2; nad3, NADH dehyrdrogenase subunit 3; nad4, NADH dehyrdrogenase subunit 4; nad4l, NADH dehyrdrogenase subunit 4L; nad5, NADH dehyrdrogenase subunit 5; nad6, NADH dehyrdrogenase subunit 6; cob, cytochrome *b*; BS, bootstrap; PP, posterior probability.

* Corresponding author at: 901 West State Street, West Lafayette, IN, 47907, USA.
E-mail address: sweet19@purdue.edu (A.D. Sweet).

https://doi.org/10.1016/j.gene.2020.145312
Received 30 July 2020; Received in revised form 3 November 2020; Accepted 13 November 2020
Available online 18 November 2020

Pterygota according to Salvato et al. (2008)), likely because of an A + T substitution bias (Tamura 1992, Simon et al., 1994, Jermiin and Crozier 1994). Some groups of insects, such as hymenopterans, have mt genomes with >85 AT% (Silvestre et al., 2008). In contrast, lice tend to have mt genomes with more balanced nucleotide compositions; many species of lice have between 50 and 60 AT% (Yoshizawa and Johnson 2013). However, there is need for a broader comparison across entire mt genomes of lice to confirm this pattern.

Despite a considerable focus on louse mt genomes, the suborder Amblycera (~1,300 species; Price et al., 2003) has received relatively little attention. Complete mt genomes have only been reported from three genera of Amblycera: *Heterodoxus*, *Amyrsidea*, and *Colpocephalum* (Shao et al., 2001, Song et al., 2019). All three taxa have a single mt chromosome, but the three mt genomes are highly rearranged relative to one another and to other lice. However, a broader taxonomic representation of the suborder is necessary for understanding the evolutionary history of the mt genomes in this group. Recent phylogenetic studies indicate that Amblycera is sister to all other parasitic lice (Johnson et al., 2018; de Moya et al., 2019), so focusing on this group will also help to clarify mt genome evolution across Phthiraptera. Mitochondrial sequences could also provide insight into the usefulness of mt data for estimating evolutionary relationships in lice. Many previous phylogenetic studies in lice have relied on mt or a few nuclear sequences (Johnson and Whiting, 2001; Smith et al., 2004; Song et al., 2019), but more recent work has generated highly supported hypotheses based on hundreds of nuclear genes (Johnson et al., 2018). Given the increasing reliance on large-scale genomic data in phylogenetics, comparing phylogenies estimated from nuclear and mt genomes could indicate whether the two types of data are compatible, or whether mt data should be considered reliable for phylogenetic analysis in lice moving forward (Springer et al., 2001; Shaw, 2002; Wiens et al., 2010; Talavera and Vila, 2011).

Here, we utilize publicly available whole genome sequence data to assemble the mt genomes for seven genera of amblyceran lice. The mt genomes from six of these genera have never been reported before and extend the taxonomic range of our understanding for this group to include representatives from all seven families of Amblycera. We combine these seven mt genomes with previously published data to characterize genome structure, gene order and nucleotide composition in this group of lice, and also use the sequences to estimate genus-level phylogenetic relationships to compare with the most recent nuclear phylogeny.

2. Materials and methods

2.1. Sequence data

We downloaded raw Illumina shotgun sequence data from the NCBI SRA database for representatives of seven genera of lice from the suborder Amblycera: *Cummingsia maculata*, *Heterodoxus spiniger*, *Laemobothrion tinnunculi*, *Macrogyropus costalimai*, *Myrsidea* sp., *Osborniella crotophagae*, and *Ricinus* sp. These data were generated from a single study (Johnson et al., 2018) and include genera from both avian and mammalian hosts (Table 1). All of the genomic data are 100, 150, or 160 bp paired-end reads sequenced on Illumina HiSeq2000 or HiSeq2500 machines. All except the *Osborniella* data were generated from genomic DNA extracted from single individual louse specimens. After obtaining the raw read data, we trimmed each library to remove adapters and low-quality ends with Fastx_clipper v.0.0.14 and Fastx_trimmer v.0.014 (FASTX Toolkit, http://hannonlab.cshl.edu/fastx_toolkit/). We then removed any reads <75 bp.

2.2. Mitochondrial genome assembly and annotation

We assembled the mitochondrial genomes for all seven genera following the approach of Sweet et al. (2020). First, we used aTRAM v.2.1 (Allen et al., 2018) to assemble mitochondrial protein-coding genes for each sample. aTRAM uses BLAST searches to identify genomic reads from a set of target genes, and then uses *de novo* assembly software to assemble the identified reads. We ran aTRAM using amino acid sequences from the published *Heterodoxus macropus* mitochondrial genome as the set of target loci. For each assembly, we ran aTRAM for three iterations with ABySS (Simpson et al., 2009) as the *de novo* assembly software. Following Sweet et al. (2018) and Sweet et al. (2020), we also used a fraction of the the genomic read libraries for assemblies. For these Amblycera libraries, we used fractions of 10%, 20%, or 50%, depending on the library size and ability of aTRAM to successfully assemble genes for a given library. Following the assembly of mitochondrial genes with aTRAM, we used each assembled gene as a starting reference in MITObim v.1.9.1 (Hahn et al., 2013). MITObim uses the MIRA v.4.0.2 assembly program (Chevreux and Suhai 1999) to generate longer mitochondrial contigs using an iterative read mapping approach. We ran MITObim with the –quick option, 100 maximum iterations, and once again using a fraction of the paired-end read libraries. Because by default MIRA calls ambiguities based on only a few alternative mappings, we used miraconvert to call each site based on a majority-rule consensus. For each louse genus, we then aligned the resulting contigs from MITObim using the Geneious *de novo* assembler in Geneious v.11.1.5 (Biomatters, Ltd., Auckland, NZ). We set sensitivity to Medium-Low and otherwise used default parameters.

We used the web version of MITOS2 (Bernt et al., 2013) to initially annotate our assembled mitochondrial contigs. We then manually fine-tuned the automated annotations by identifying open reading frames and comparing to published data (following Cameron, 2014). We identified potential circular contigs by a) identifying genic regions overlapping both the 5′ and 3′ ends of a contig and/or b) identifying long (>50 bp) regions of identical sequence near the 5′ and 3′ ends. We then validated circularity in AWA, which uses paired-end read information to test for continuous coverage across the 5′ and 3′ ends of a potentially circular contig (Machado et al., 2018). If a contig did not meet at least one of the first two conditions or did not pass the AWA test (match > 95%, alignment score > -3, connection coverage > 20X across the 5′ and 3′ junctures), then we did not consider that contig a complete circle. All annotated mt genomes are available on GenBank (Table 1).

In order to characterize the nucleotide composition of amblyceran mt genomes, we calculated nucleotide composition (AT%), AT-skew ((A-T)/(A + T)), and GC-skew ((G-C)/(G + C)) from the available sequences, including the seven mt genomes reported here and the three published

Table 1
Sample and genome information for seven newly assembled mitochondrial genomes of Amblycera lice.

Louse species	Host species	Host class	Architecture	Chromosomes†	AT%	AT-skew	GC-skew	GenBank	SRA
Cummingsia maculata	*Lestoros inca*	Mammalia	Fragmented	3	76.4	−0.01	−0.06	MW199177-MW199179	SRR5308146
Heterodoxus spiniger	*Canis lupus*	Mammalia	Single	1	77.7	−0.03	−0.01	MW199168	SRR5308125
Laemobothrion tinnunculi	*Falco longipennis*	Aves	Fragmented	3	75.9	−0.07	−0.07	MW199169-MW199171	SRR5308127
Macrogyropus costalimai	*Cuniculus paca*	Mammalia	Fragmented	7	75.2	−0.12	0.06	MT644272-MT644278	SRR5308130
Myrsidea sp.	*Myiothylpis luteoviridis*	Aves	Fragmented	3	69.5	−0.05	0.03	MW199172-MW199174	SRR5308132
Osborniella crotophagae	*Crotophaga ani*	Aves	Single	1	76.6	−0.03	0.03	MW199175	SRR5088470
Ricinus sp.	*Myiothylpis luteoviridis*	Aves	Single	1	74.5	−0.03	0.08	MW199176	SRR5308140

† Number of mt fragments assembled from whole genome data.

previously. We also calculated nucleotide composition for coding regions only (i.e., excluding control and intergenic regions), first + second codon positions, and each codon position separately. We then used custom Python scripts to download and calculate AT%, AT-skew, and GC-skew for all full insect mt genomes available in GenBank. The scripts are available on GitHub (https://github.com/adsweet/mitogenomes). We compared nucleotide composition and skew among lice and other insects with Kruskal-Wallis and pairwise Wilcoxon Rank Sum tests to account for the large sample size differences between lice and other insects. We compared different groups of lice with Wilcoxon Rank Sum tests or t-tests.

2.3. Phylogenetic analysis

To estimate the genus-level phylogenetic relationships among nine Amblycera genera based on their mitochondrial sequences, we aligned protein-coding genes according to their translated codons and rRNA genes using the default parameters in the MAFFT plugin in Geneious (Katoh et al., 2002). We included the previously published sequences from *Amyrsidea minuta* and *Colpocephalum griffoneae* (Song et al., 2019), and used published data from the louse genera *Ibidoecus*, *Campanulotes*, *Pediculus*, and *Haematomyzus* as outgroups (Shao et al., 2009, 2015; Cameron et al., 2011) (Supplementary Table S3). We then created three alignment subsets: 1) a concatenated alignment of all genes and sites, 2)

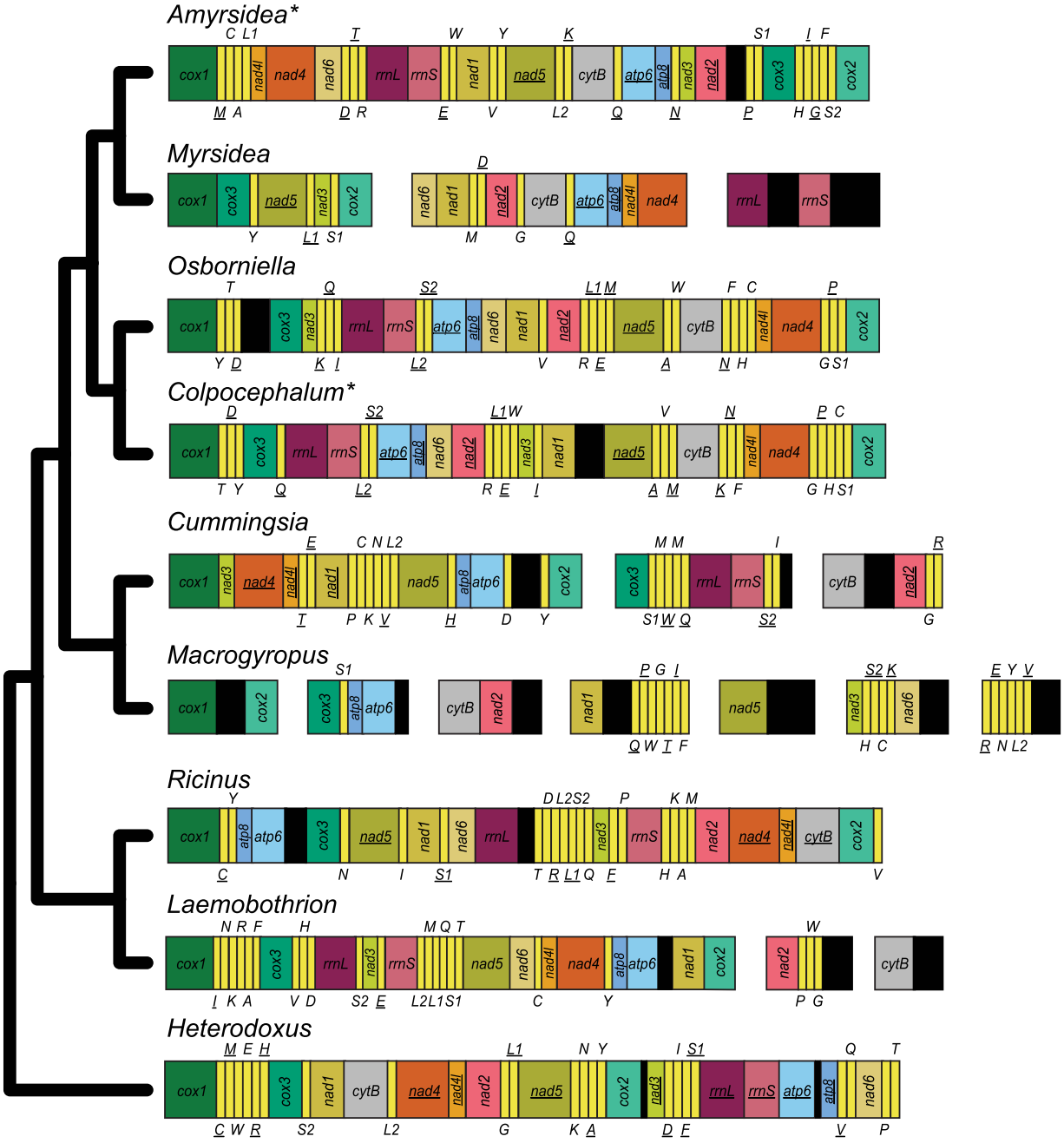

Fig. 1. Mitochondrial (mt) genomes from nine genera of Amblycera lice. Asterisks indicate mt genomes from previous publications. *Cummingsia, Macrogyropus, Myrsidea,* and *Ricinus* are each missing at least one gene. The genomes have been linearized and arbitrarily begin with *cox1* when possible. Genera that have multiple mitochondrial fragments are shown with multiple boxes. tRNAs (indicated by yellow boxes), are given single-letter IACUC abbreviations based on their anticodons. Underlined labels indicate the gene is on the opposite strand than the majority of genes on that chromosome. Black boxes indicate control or non-coding intergenic regions. The cladogram is modified from the ML phylogeny estimated with protein-coding and rRNA genes. See the online version for color.

a concatenated alignment with 3rd codon positions removed from protein-coding genes, and 3) an alignment of *cox1*, a widely used barcoding gene (Hebert et al., 2003). We also tested for saturation by running Xia ISS tests in DAMBE v.7.0.13 (Xia et al., 2003; Xia, 2017) on alignments of the first and second codon positions and separately for the third codon position, and then used the APE package in R (Paradis and Schliep 2018) to plot raw pairwise distances against corrected distances (F84 model) for the two subsets. For each of these alignment types, we estimated optimal substitution models and partitioning strategies using PartitionFinder v.2.1.1 (Lanfear et al., 2016). We then estimated phylogenetic relationships under maximum likelihood (ML) and Bayesian search strategies. For ML, we used IQ-Tree v.1.5.5 with proportionally linked branch lengths and 1,000 ultrafast bootstrap replicates (Nguyen et al., 2015, Hoang et al., 2018). For a Bayesian approach, we used MrBayes v.3.2.7 (Ronquist and Huelsenbeck 2003) on the CIPRES Science Gateway (Miller et al., 2010). We ran three separate MCMC searches, with four chains, 20 million iterations, and sampling every 1,000 iterations in each search. We determined the runs reached stationarity by using Tracer v.1.7.1 (Rambaut et al., 2018) to confirm all parameters had Effective Sample Sizes > 200 and using the R package RWTY (Warren et al., 2017) to confirm average standard deviation of split frequencies to be < 0.01. Finally, we tested whether the mt gene order of Amblycera is phylogenetically informative by running a phylogenetic reconstruction based on a gene order file and 1,000 bootstrap replicates in the Maximum Likelihood for Gene Order (MLGO) web server (Hu et al., 2014). We standardized the gene order files by starting single chromosomes with the *cox1* gene and consistently starting mt genome fragments with the same genes whenever possible.

3. Results and discussion

3.1. Rearranged and fragmented mt genomes in Amblycera

Our assemblies of complete and partial mt genomes of amblyceran lice show that these genomes have extremely variable architecture (Table 1, Fig. 1). Most notably, we recovered a mix of single and fragmented mt genomes within the suborder. This is the first reported case of fragmented mt genomes in Amblycera. Among the nine genera in our dataset, four have evidence for some level of fragmentation (*Cummingsia*, *Laemobothrion, Macrogyropus*, and *Myrsidea*) and five have the usual single chromosome (*Amyrsidea, Colpocephalum, Heterodoxus, Osborniella*, and *Ricinus*). Nine of the 16 fragments passed our circularity validation test, and all taxa with fragments had at least one fragment pass the test (Supplementary Table S1). The remaining fragments are likely circular as well, but we were unable to confirm complete circularity across the non-coding regions. These non-coding regions often have stretches of highly repetitive sequences, which are difficult to assemble through with both PCR and computation-based approaches. Among the fragmented mt genomes, there is no consistency in the degree of fragmentation. Three of the four genera have three fragments, although the lengths and gene content of these fragments vary (Supplementary Table S1): *Cummingsia* has one longer (8,885 bp) and two medium-length fragments (4,244 and 3,204 bp); *Myrsidea* has three fragments of comparable size (6,194, 6,247, and 5,183 bp); and *Laemobothrion* has one very long fragment (12,244 bp) and two much smaller fragments (1,263 and 1,545 bp). *Macrogyropus* has a noticeably different mt genome architecture, with at least seven fragments between ~ 1,600 – 2,700 bp long. This structure is more similar to the minicircle-type fragments found in mammal lice (Shao et al., 2009, 2015; Song et al., 2019) and *Columbicola* feather lice (Sweet et al., 2020). The variable patterns of the fragmentation, in addition to the phylogenetic position of the genera with fragmentation (Fig. 1), suggest that mt genome fragmentation has occurred multiple times independently throughout the evolutionary history of Amblycera.

We were not able to assemble/annotate all 37 mt genes for each genus (Supplementary Tables S1-S2). *Cummingsia, Macrogyropus, Myrsidea,* and *Ricinus* are each missing at least one gene. Although it is possible some of these genes are legitimately missing from these genera, it is more likely that the genes exist on separate fragments that we were unable to assemble. tRNA genes are particularly difficult to locate if they are on separate fragments, primarily because the genes are relatively short (~55–75 bp) and can differ substantially even between closely related taxa (Helm et al., 2000; Jühling et al., 2012).

The four genera of Amblycera with fragmented mt genomes have fewer fragments compared to other lice with fragmented mt genomes. Mammal lice have between 9 and 20 mt genome fragments (Shao et al., 2012, 2015; Song et al., 2019), and pigeon feather lice in the genus *Columbicola* have 15–17 fragments (Sweet et al., 2020). In comparison, we found that most Amblycera with fragmented mt genomes have relatively few fragments. This level of fragmentation is more similar to what has been found in a few species of thrips and booklice, taxa in which fragmentation is quite rare (Dickey et al., 2015; Shi et al., 2016; Yoshizawa et al., 2018; Feng et al., 2019; Tyagi et al., 2020). However, *Macrogyropus* has at least seven mt genome fragments, so there is certainly variation within Amblycera.

In addition to fragmented mt genomes, we also recovered highly rearranged gene orders among the nine genera. This result is consistent with previously published mt genomes of Amblycera (Shao et al., 2001; Song et al., 2019) and parasitic lice more generally (Covacin et al., 2006; Cameron et al., 2011). Most of the conserved gene boundaries are present in the inferred ancestral gene order in insects: *atp8* and *atp6*, *nad4* and *nad4l*, *cox1* and *cox2*, and *rrnL* and *rrnS* (Boore et al., 1998). However, even these arrangements are not conserved among all Amblycera. For example, *cox1* and *cox2* are separated in *Heterodoxus*. However, despite high levels of rearrangements, there are similarities that reflect phylogenetic relatedness. *Heterodoxus spiniger* has an identical gene order to a previously reported congeneric species, *H. macropus* (Shao et al., 2001). *Osborniella* and *Colpocephalum* share an identical block of gene boundaries between *rrnL* and *nad6*, and a very similar block between *nad5* and *cox3*. *Myrsidea* and *Amyrsidea* also have similarities, including a conserved block between *cob* and *atp6* (Fig. 1). These patterns suggest that denser taxonomic sampling or focusing on particular clades within Amblycera is needed to further understand the evolution of mt gene arrangements in the group.

3.2. Comparative nucleotide composition in the mt genomes of Amblycera

The seven Amblycera mt genomes reported here are AT-biased (mean 75.1%), slightly T-skewed (mean A-skew: −0.05), and slightly G-skewed (mean G-skew: 0.01) (Table 1). This composition is very similar to previously published Amblycera mt genomes, but there are substantial differences with other louse mt genomes (Supplementary Table S3). In particular, feather lice and mammal lice both have significantly lower AT% than Amblycera ($p < 0.001$). However, lice have mt genomes with lower AT content relative to many other insects (Yoshizawa and Johnson 2003, Jermiin and Crozier 1994), and the high AT% in Amblycera are more similar to other insect mt genomes. This is reflected in our comparisons to AT% from over 6,000 insect mt genomes (excluding lice) from GenBank, which showed no significant difference in AT% with Amblycera ($p = 0.18$) (Fig. 2A, Supplementary Table S4). AT- and GC-skew values were more consistent among the different groups of lice, but most lice were significantly more T-skewed ($p < 0.001$) and G-skewed ($p < 0.001$) compared to other insects (Fig. 2C-D, Supplementary Table S4). This pattern is consistent with previous metanalyses of mt genome composition in insects (Salvato et al., 2008, McMahon et al., 2009). There were no significant differences in the AT content of louse mt genomes among coding or different codon sites ($p > 0.21$ in all comparisons) (Supplementary Figure S1, Supplementary Table S3), indicating these comparative patterns are not driven by biases in non-coding regions or third codon positions.

Among louse mt genomes, it is noteworthy that taxa with single mt chromosomes have a significantly higher AT% than those with

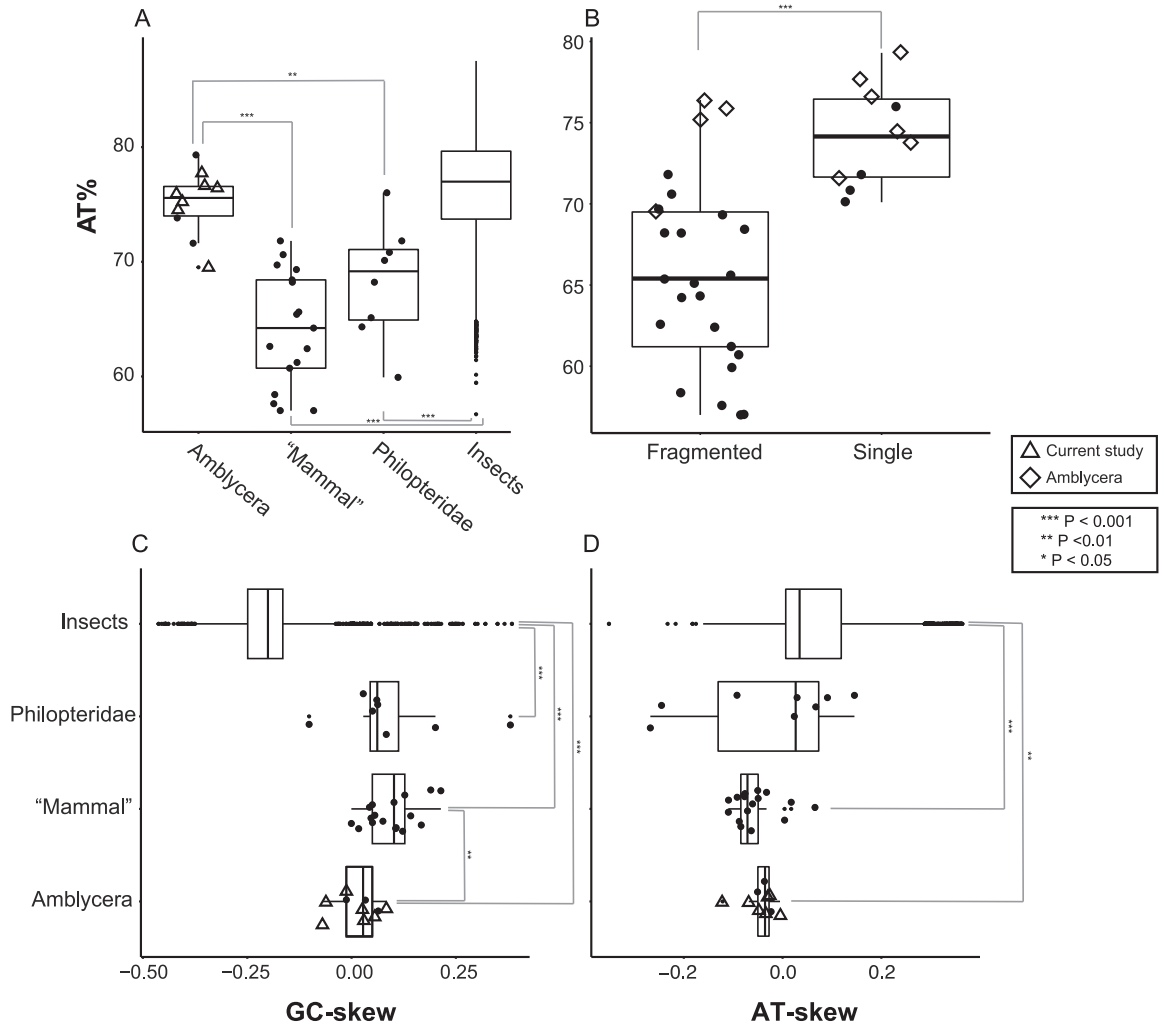

Fig. 2. Box-and-whisker plots showing the distribution of nucleotide composition for the full mitochondrial (mt) genomes of different groups of lice and other insects. "Mammal" lice include Anoplura, Trichodetidae, and Rhynchophthirina. Each box depicts the median (line), first and third quartiles (box edges), minimum/maximum values, and outliers (small dots). Filled and hollow shapes indiciate individual mt genomes, and are shaped according to research study (A, C-D) or suborder of lice (B). Lines and asterisks indicate support values for significant differences between groups. Pairs not linked with lines are not significantly different. Positive GC-skew values indicate a G-skew, and positive AT-skew values indicate an A-skew.

fragmented mt genomes ($t = -5.53$, $p = 5.4e^{-6}$) (Fig. 2B). Yoshizawa and Johnson (2013) postulated that relaxed selection explained the lower AT % in lice compared to other insects (Simon et al., 1994). However, in light of our results, it could be that there is relaxed selection in lice with fragmented mt genomes more specifically, rather than in lice overall. It is also possible that preferential mutational mechanisms (e.g., deamination) are responsible for the differences in AT% between fragmented and non-fragmented mt genomes (Tamura 1992; Reyes et al., 1998), but further work is needed to tease apart the various mechanisms that could explain these patterns. Nevertheless, Amblycera tend to have AT-rich mt genomes regardless of genome structure. If nucleotide composition is a signature of genome fragmentation, Amblycera mt genomes have compositions consistent with single chromosomes, suggesting that fragmentation occurred relatively recently in these taxa.

3.3. Phylogenetic utility of mt sequences in Amblycera

Our genus-level phylogenetic analyses using protein-coding and rRNA genes from Amblycera recovered trees consistent with the Johnson et al. (2018) parasitic louse phylogeny based on >1,000 nuclear genes (Fig. 3). The trees were well-supported overall, although two branches received low support (<70 bootstrap (BS) and <0.95 posterior probability (PP)). Compared to the all-sites tree, phylogenies based on the alignment without 3rd codon positions and from *cox1* alone had different topologies, lower support for some relationships, and were unstable between ML and Bayesian inference methods (Supplementary Figure S2-S3). These results suggest the amount of information lost from removing 3rd codon positions is more detrimental than any effects of saturation or nucleotide bias (Supplementary Figure S4) (Song et al., 2010; Bourguignon et al., 2016; Yoshizawa et al., 2018), and gives further evidence that the *cox1* barcoding gene is less informative for deeper-level evolutionary relationships (Rubinoff et al., 2006). Gene order information also seems to be unreliable for phylogenetic reconstruction in Amblycera. Our reconstruction using MLGO recovered a very different topology compared to the trees based on nuclear or mitochondrial sequence data (Supplementary Figure S5). The extremely variable mt genome structure among genera in Amblycera likely make it difficult to accurately reconstruct relationships with gene order data, especially with low representation of extant taxa (Tyagi et al., 2020). Overall, our phylogenetic results indicate that nucleotide sequences from mt genes can provide reliable reconstructions of evolutionary relationships in lice, even at relatively deep phylogenetic scales. However, genomic-level nuclear datasets are likely important for phylogenetic estimation among major clades of lice, where nucleotide composition

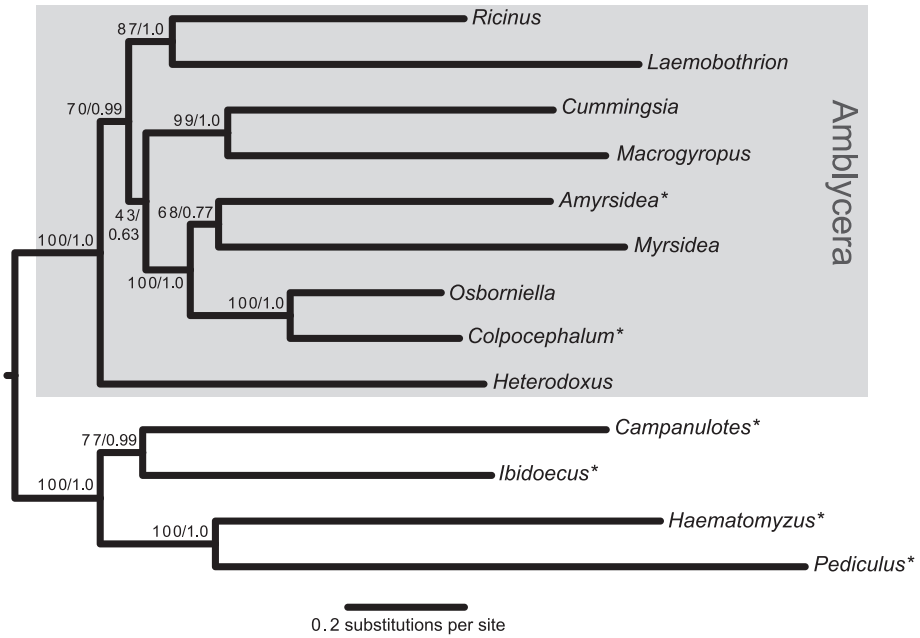

Fig. 3. Genus-level phylogeny of Amblycera lice based on all sites of mitochondrial protein-coding and rRNA genes. Asterisks indicate taxa with sequence data generated from previous studies. Support values are shown above or below branches (bootstrap/posterior probability).

varies drastically between different groups (Galtier and Gouy 1995).

CRediT authorship contribution statement

Andrew D. Sweet: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Kevin P. Johnson:** Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Funding acquisition. **Yanghui Cao:** Methodology, Formal analysis, Investigation, Writing - original draft. **Robert S. de Moya:** Methodology, Formal analysis, Investigation, Writing - original draft. **Rachel K. Skinner:** Methodology, Formal analysis, Investigation, Writing - original draft. **Milton Tan:** Methodology, Formal analysis, Investigation, Writing - original draft. **Stephany Virrueta-Herrera:** Methodology, Formal analysis, Investigation, Writing - original draft. **Stephen L. Cameron:** Conceptualization, Writing - original draft, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Newly reported mitochondrial genome sequences are available on GenBank under accession numbers MW199168-MW199179 and MT644272-MT644278. Relevant data and results files are available on the Dryad Digital Repository (doi:10.5061/dryad.8sf7m0ckn). We are grateful for the comments of Xavier Carette and an anonymous reviewer, which helped improve the manuscript. This work was supported by the National Science Foundation [grant numbers DBI-1906262 to ADS; and DEB-1239788, DEB-1342604, DEB-1855812, and DEB-1925487 to KPJ].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2020.145312.

References

Allen, J.M., LaFrance, R., Folk, R.A., Johnson, K.P., Guralnick, R.P., 2018. aTRAM 2.0: An improved, flexible locus assembler for NGS data, 117693431877454 Evol. Bioinforma. 14. https://doi.org/10.1177/1176934318774546.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenetics Evol. 69 (2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023.

Boore, J.L., Lavrov, D.V., Brown, W.M., 1998. Gene translocation links insects and crustaceans. Nature 392 (6677), 667–668. https://doi.org/10.1038/33577.

Boore, J.L., 1992. Animal Mitochondrial DNA: Structure and Evolution. Int. Rev. Cytol. 141, 173–216. https://doi.org/10.1016/S0074-7696(08)62066-5.

Bourguignon, T., Lo, N., Sobotn, J., Ho, S.Y.W., Iqbal, N., Coissac, E., Lee, M., Jendryka, M.M., Sillam-Dussè, D., et al., 2016. Mitochondrial phylogenomics resolves the global spread of higher termites, ecosystem engineers of the tropics. Mol. Biol. Evol. 34, 589–597. https://doi.org/10.1093/molbev/msw253.

Cameron, S.L., 2014. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research: Sequencing insect mt genomes. Syst. Entomol. 39 (3), 400–411. https://doi.org/10.1111/syen.12071.

Cameron, S.L., Yoshizawa, K., Mizukoshi, A., Whiting, M.F., Johnson, K.P., 2011. Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). BMC Genomics 12 (1). https://doi.org/10.1186/1471-2164-12-394.

Cameron, S.L., Johnson, K.P., Whiting, M.F., 2007. The Mitochondrial Genome of the Screamer Louse *Bothriometopus* (Phthiraptera: Ischnocera): Effects of Extensive Gene Rearrangements on the Evolution of the Genome. J. Mol. Evol. 65, 589–604. https://doi.org/10.1007/s00239-007-9042-8.

Chen, X.J., Butow, R.A., 2005. The organization and inheritance of the mitochondrial genome. Nat. Rev. Genet. 6, 815–825. https://doi.org/10.1038/nrg1708.

Chevreux, B., Suhai, S., 1999. genome sequence assembly using trace signals and additional sequence information. Computer Sci. Biol.: Proceedings of the German Conference on Bioinformaics (GCB) 99, 45–56.

Covacin, C., Shao, R., Cameron, S., Barker, S.C., 2006. Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). Insect Mol. Biol. 15, 63–68. https://doi.org/10.1111/j.1365-2583.2005.00608.x.

de Moya, R.S., Allen, J.M., Sweet, A.D., Walden, K.K.O., Palma, R.L., Smith, V.S., Cameron, S.L., Valim, M.P., Galloway, T.D., Weckstein, J.D., et al., 2019. Extensive host-switching of avian feather lice following the Cretaceous-Paleogene mass extinction event. Commun. Biol. 2, 445. https://doi.org/10.1038/s42003-019-0689-7.

Dickey, A.M., Kumar, V., Morgan, J.K., Jara-Cavieres, A., Shatters, R.G., McKenzie, C.L., Osborne, L.S., 2015. A novel mitochondrial genome architecture in thrips (Insecta: Thysanoptera): extreme size asymmetry among chromosomes and possible recent control region duplication. BMC Genomics 16, 439. https://doi.org/10.1186/s12864-015-1672-4.

Dong, W.-G., Song, S., Jin, D.-C., Guo, X.-G., Shao, R., 2014. Fragmented mitochondrial genomes of the rat lice, *Polyplax asiatica* and *Polyplax spinulosa*: intra-genus variation in fragmentation pattern and a possible link between the extent of fragmentation and

the length of life cycle. BMC Genomics 15, 44. https://doi.org/10.1186/1471-2164-15-44.
Feng, S., Li, H., Song, F., Wang, Y., Stejskal, V., Cai, W., Li, Z., 2019. A novel mitochondrial genome fragmentation pattern in *Liposcelis brunnea*, the type species of the genus *Liposcelis* (Psocodea: Liposcelididae). Int. J. Biol. Macromol. 132, 1296–1303. https://doi.org/10.1016/j.ijbiomac.2019.04.034.
Galtier, N., Gouy, M., 1995. Inferring phylogenies from DNA sequences of unequal base compositions. PNAS 92, 11317–11321. https://doi.org/10.1073/pnas.92.24.11317.
Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. 41, e129 https://doi.org/10.1093/nar/gkt371.
Hebert, P.D.N., Ratnasingham, S., Dewaard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Biol. Lett. 270, S96–S99. https://doi.org/10.1098/rsbl.2003.0025.
Helm, M., Brulé, H., Friede, D., Giegé, R., Pütz, D., Florentz, C., 2000. Search for characteristic structural features of mammalian mitochondrial tRNAs. RNA 6, 1356–1379. https://doi.org/10.1017/s1355838200001047.
Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.5281/zenodo.854445.
Hu, F., Lin, Y., Tang, J., 2014. MLGO: Phylogeny reconstruction and ancestral inference from gene-order data. BMC Bioinformatics 15, 354. https://doi.org/10.1186/s12859-014-0354-6.
Jermiin, L.S., Crozier, R.H., 1994. The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: Sequence divergence in hymenoptera may be associated with nucleotide content. J. Mol. Evol. 38, 282–294. https://doi.org/10.1007/BF00176090.
Johnson, K.P., Nguyen, N.-P., Sweet, A.D., Boyd, B.M., Warnow, T., Allen, J.M., 2018. Simultaneous radiation of bird and mammal lice following the K-Pg boundary. Biol. Lett. 14, 20180141. https://doi.org/10.1098/rsbl.2018.0141.
Johnson, K.P., Whiting, M.F., 2001. Multiple Genes and the Monophyly of Ischnocera (Insecta: Phthiraptera). Mol. Phy. 22, 101–110. https://doi.org/10.1006/mpev.2001.1028.
Jühling, F., Pütz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., Stadler, P.F., 2012. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Res. 40, 2833–2845. https://doi.org/10.1006/mpev.2001.1028.
Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. https://doi.org/10.1093/nar/gkf436.
Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol. Biol. Evol. 34, 772–773. https://doi.org/10.1093/molbev/msw260.
Machado, D.J., Janies, D., Brouwer, C., Grant, T., 2018. A new strategy to infer circularity applied to four new complete frog mitogenoes. Ecol. Evol. 8, 4011–4018. https://doi.org/10.1002/ece3.3918.
McMahon, D.P., Hayward, A., Kathirithamby, J., 2009. The mitochondrial genome of the "twisted-wing parasite" *Mengenilla australiensis* (Insecta, Strepsiptera): A comparative study. BMC Genomics 10, 603. https://doi.org/10.1186/1471-2164-10-603.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, LA. p. 1–8.
Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.
Paradis, E., Schliep, K., 2018. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633.
Price, R.D., Hellenthal, R.A., Palma, R.L., Johnson, K.P., Clayton, D.H., 2003. The chewing lice: world checklist and biological overview. Illinois Natural History Survey, Champaign, IL.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904. https://doi.org/10.1093/sysbio/syy032.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinforma. Appl. Note 19, 1572–1574. https://doi.org/10.1093/bioinformatics/btg180.
Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Molec. Biol. Evol. 15, 957–966. https://doi.org/10.1093/oxfordjournals.molbev.a026011.
Rubinoff, D., Cameron, S., Will, K., 2006. A genomic perspective on the shortcomings of mitochondrial DNA or "barcoding" identification. J. Heredity 97, 581–594. https://doi.org/10.1093/jhered/esl036.
Salvato, P., Simonato, M., Battisti, A., Negrisolo, E., 2008. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). BMC Genomics 9, 331. https://doi.org/10.1186/1471-2164-9-331.
Shao, R., Barker, S.C., Li, H., Song, S., Poudel, S., Su, Y., 2015. Fragmented mitochondrial genomes in two suborders of parasitic lice of eutherian mammals (Anoplura and Rhynchophthirina, Insecta). Sci. Rep. 5, 17389. https://doi.org/10.1038/srep17389.
Shao, R., Campbell, N.J.H., Barker, S.C., 2001. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol. Biol. Evol. 18, 858–865. https://doi.org/10.1093/oxfordjournals.molbev.a003867.
Shao, R., Kirkness, E.F., Barker, S.C., 2009. The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse. *Pediculus humanus*. Genome Res. 19, 904–912. https://doi.org/10.1101/gr.083188.108.
Shao, R., Zhu, X.-Q., Barker, S.C., Herd, K., 2012. Evolution of Extensively Fragmented Mitochondrial Genomes in the Lice of Humans. Genome Biol. Evol. 4, 1088–1101. https://doi.org/10.1093/gbe/evs088.
Shaw, K.L., 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. PNAS 99, 16122–16127. https://doi.org/10.1073/pnas.242585899.
Shi, Y., Chu, Q., Wei, D.-D., Qiu, Y.-J., Shang, F., Dou, W., Wang, J.-J., 2016. The mitochondrial genome of booklouse, *Liposcelis sculptilis* (Psocoptera: Liposcelididae) and the evolutionary timescale of *Liposcelis*. Sci. Rep. 6, 30660. https://doi.org/10.1038/srep30660.
Silvestre, D., Downton, M., Arias, M.C., 2008. The mitochondrial genome of the stingless bee *Melipona bicolor* (Hymenoptera, Apidae, Meliponini): Sequence, gene organization and a unique tRNA translocation event conserved across the tribe Meliponini. Genet. Mol. Biol. 31, 451–460. https://doi.org/10.1590/S1415-47572008000300010.
Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701. https://doi.org/10.1093/aesa/87.6.651.
Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, İ., 2009. ABySS: A parallel assembler for short read sequence data. Genome Res. 19, 1117–1123. https://doi.org/10.1101/gr.089532.108.
Smith, V.S., Page, R.D.M., Johnson, K.P., 2004. Data incongruence and the problem of avian louse phylogeny. Zool. Scr. 33, 239–259. https://doi.org/10.1111/j.0300-3256.2004.00149.x.
Song, F., Li, H., Liu, G.H., Wang, W., James, P., Colwell, D.D., Tran, A., Gong, S., Cai, W., Shao, R., 2019. Mitochondrial genome fragmentation unites the parasitic lice of Eutherian mammals. Syst. Biol. 68, 430–440. https://doi.org/10.1093/sysbio/syy062.
Song, H., Sheffield, N.C., Cameron, S.L., Miller, K.B., Whiting, M.F., 2010. When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. Syst. Entomol. 35, 429–448. https://doi.org/10.1111/j.1365-3113.2009.00517.x.
Springer, M.S., DeBry, R.W., Douady, C., Amrine, H.M., Madsen, O., de Jong, W.W., Stanhope, M.J., 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol. Biol. Evol. 18, 132–143. https://doi.org/10.1093/oxfordjournals.molbev.a003787.
Sweet, A.D., Johnson, K.P., Cameron, S.L., 2020. Mitochondrial genomes of *Columbicola* feather lice are highly fragmented, indicating repeated evolution of minicircle-type genomes in parasitic lice. PeerJ 8, e8759. https://doi.org/10.7717/peerj.8759.
Sweet, A.D., Boyd, B.M., Allen, J.M., Villa, S.M., Valim, M.P., Rivera-Parra, J.L., Wilson, R.E., Johnson, K.P., 2018. Integrating phylogenomic and population genomic patterns in avian lice provides a more complete picture of parasite evolution. Evolution 72, 95–112. https://doi.org/10.1111/evo.13386.
Talavera, G., Vila, R., 2011. What is the phylogenetic signal limit from mitogenomes? the reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. BMC Evol. Biol. 11, 315. https://doi.org/10.1186/1471-2148-11-315.
Tamura, K., 1992. The rate and pattern of nucleotide substitution in Drosophila mitochondrial DNA. Artic. Mol. Biol. Evol. 9, 814–825. https://doi.org/10.1093/oxfordjournals.molbev.a040763.
Tyagi, K., Chakraborty, R., Cameron, S.L., Sweet, A.D., Chandra, K., Kumar, V., 2020. Rearrangement and evolution of mitochondrial genomes in Thysanoptera (Insecta). Sci. Rep. 10, 1–16. https://doi.org/10.1038/s41598-020-57705-4.
Warren, D.L., Geneva, A.J., Lanfear, R., 2017. RWTY (R We There Yet): An R package for examining convergence of bayesian phylogenetic analyses. Mol. Biol. Evol. 34, 1016–1020. https://doi.org/10.1093/molbev/msw279.
Wiens, J.J., Kuczynski, C.A., Stephens, P.R., 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. Biol. J. Linn. Soc. 99, 445–461. https://doi.org/10.1111/j.1095-8312.2009.01342.x.
Yoshizawa, K., Johnson, K.P., Sweet, A.D., Yao, I., Ferreira, R.L., Cameron, S.L., 2018. Mitochondrial phylogenomics and genome rearrangements in the barklice (Insecta: Psocodea). Mol. Phylogenet. Evol. 119, 118–127. https://doi.org/10.1016/j.ympev.2017.10.014.
Yoshizawa, K., Johnson, K.P., 2013. Changes in base composition bias of nuclear and mitochondrial genes in lice (Insecta: Psocodea). Genetica 141, 491–499. https://doi.org/10.1007/s10709-013-9748-z.
Yoshizawa, K., Johnson, K.P., 2003. Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. Mol. Phylogenet. Evol. 29, 102–114. https://doi.org/10.1016/S1055-7903(03)00073-3.
Xia, X., 2017. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. Molec. Biol. Evol. 35, 1550–1552. https://doi.org/10.1093/molbev/mys073.
Xia, X., Xie, Z., Salemi, M., Chen, L., Wang, Y., 2003. An index of substitutions saturation and its application. Mol. Phylogen. Evol. 26, 1–7. https://doi.org/10.1016/S1055-7903(02)00326-3.