bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mitogenome fragmentation evolves multiple times within a major group of parasitic lice

Tomáš Najer[1,2], Jorge Doña[2,3], Aleš Buček[4], Andrew D. Sweet[5], Oldřich Sychra[6], Kevin P. Johnson[2]

[1]Department of Veterinary Sciences, Faculty of Agrobiology, Food, and Natural Resources, Czech University of Life Sciences Prague, Prague 6, Czechia

[2]Illinois Natural History Survey, Prairie Research Institute, University of Illinois Urbana-Champaign, Champaign, IL 61820, USA

[3]Departamento de Biología Animal, Universidad de Granada, Granada, Spain

[4]Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan

[5]Department of Biological Sciences, Arkansas State University, Jonesboro, AR 72401, USA

[6]Department of Biology and Wildlife Diseases, Faculty of Veterinary Hygiene and Ecology, University of Veterinary Sciences Brno, Brno, Czechia

Abstract

Animal mitochondrial genomes (mitogenomes) typically exhibit a highly conserved gene content and organisation, with genes encoded on a single circular chromosome. However, many species of parasitic lice (Insecta: Phthiraptera) are notable exceptions, having mitogenomes fragmented into multiple circular chromosomes. To further understand the process of mitogenome fragmentation, we conducted a large-scale genomic study of a major group of lice, Amblycera. Using genomic information on mitogenome structure and a phylogenomic tree for 90 samples of this group, we found evidence for multiple independent origins of mitogenome fragmentation, some inferred to have occurred less than five million years ago. In addition, the base composition of mitogenomes in Amblycera shows less AT bias than other insects, and fragmentation is related to this reduction of AT bias. By combining phylogenomics and mitochondrial genomics, we provide a detailed portrait of mitogenome evolution across this major group of insects. The evidence of repeated and ongoing fragmentation represents a substantial advance in understanding the repeated nature of this process in lice.

Keywords: mitochondrial genome, Psocodea, Phthiraptera, Amblycera, phylogenomics

Introduction

Mitochondria play a vital role in the metabolism of eukaryotic organisms, supplying cells with most of the energy necessary to function. Likewise, the gene content, structure, and organisation of the mitochondrial genome (mitogenome) remains remarkably stable, especially across animals. In the animal kingdom, mitochondrial genetic information is usually organised on a single circular chromosome which is between 12,000–18,000 base pairs long and contains 37 genes [1]. Any defect in this structure typically results in cell death [2], and is linked to degenerative diseases, ageing, and cancer [3].

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

However, in some animals, mitogenome structure deviates from this conserved state without having a notable effect on mitochondrial functions. Specifically, the mitogenomes of some nematodes (*Globodera*; [4]), cnidarians [5], thrips [6], and lice [7] diverge from this typical structure. In all these groups, the original single mitochondrial chromosome is fragmented into smaller chromosomes of variable size and number, either linear or circular. Similar mitochondrial mutations have detrimental effects in humans [8, 3]. Therefore, understanding mitogenome fragmentation and reorganisation may provide significant insights for research on cell ageing or severe hereditary diseases.

Lice (Insecta: Psocodea) exhibit the most extensive variability in mitogenome structure among all animals. In free-living lice (i.e., bark lice), the mitogenomes are usually single-chromosome, except for book lice, which have two circular mitochondrial chromosomes [9]. In contrast, parasitic lice (Phthiraptera) show a broad range of mitogenome arrangements. Several lineages of parasitic lice maintain mitogenomes on single circular chromosomes [10], while in others, the mitogenomes are highly fragmented and consist of up to 20 small circular chromosomes [11, 12, 13]. Recent evidence of heteroplasmy (multiple gene arrangements in a single individual) in some lice [7] suggests that the initial fragmentation in lice is not a final state but an ongoing process towards increasing fragmentation. Hence, lice offer an exceptional model to document the process of mitogenome fragmentation over time.

To date, studies of mitogenome fragmentation in lice have primarily focused on mammalian lice (Trichodectera; [14]; Anoplura; [13]) and avian feather lice (Ischnocera; [7, 11]). In the case of mammalian lice, Anoplura (sucking lice) and Trichodectera (chewing lice) represent extreme instances of mitochondrial fragmentation. All members of these groups appear to have highly fragmented mitogenomes, with no cases of a single chromosome discovered in these groups to date. Thus, in Anoplura and Trichodectera, mitochondrial fragmentation is highly phylogenetically conserved. However, in avian feather lice (Ischnocera), the opposite appears to be the case [7]. Within this group, mitogenome structure varies dramatically over the tree, with some species possessing a single mitochondrial chromosome while related genera are highly fragmented. This pattern makes it challenging to understand the process of stepwise fragmentation and whether there is phylogenetic conservation of mitogenome structure in Ischnocera. Mitochondrial fragmentation in another major group of lice (Amblycera) is less well studied. A recent study of nine species in this group revealed five cases of single-chromosome mitogenomes and four of fragmented mitogenomes, which seem to have occurred multiple times [10]. Thus, Amblycera provides an excellent opportunity to study the evolutionary dynamics of mitochondrial fragmentation with expanded sampling.

An important feature of louse mitochondrial genomes potentially linked to fragmentation is base composition [15]. Across all animals, mitochondrial genomes typically exhibit a strong AT bias, often over 70% [7]. This phenomenon is most readily explained by biases during replication, where the separated DNA strands are exposed to deamination mutations [16]. Mitochondria lack the extensive repair mechanisms of nuclear genomes [17], making mitogenomes more susceptible to mutations. However, animals with fragmented mitogenomes consistently show less AT biased mitogenomes [7, 10]. When a mitogenome becomes fragmented, the chromosomes are shorter and take less time to replicate. Consequently, less time is spent in a single-stranded state, likely making the genome less

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

vulnerable to mutations. Thus, in fragmented mitogenomes, deamination mutations would be less common and the entire mitogenome less AT biased [7]. Therefore, organisms with fragmented mitogenomes allow the study of these patterns of molecular evolution in more detail.

The main goal of this study was to uncover the pattern of mitogenome fragmentation in a major group of lice, Amblycera, with unprecedented taxonomic and temporal resolution. Using genomic sequencing reads, we assembled the mitochondrial genomes of 90 samples from this group and reconstructed a dated phylogenomic tree based on over 2,000 nuclear single-copy ortholog genes. With these data, we traced the phylogenetic pattern of the fragmentation across Amblycera, inferring the number of transitions from non-fragmented to fragmented mitogenomes and shedding light on the dynamics of this process.

Material & Methods

Sequence data

We analysed the mitogenomes of 90 amblyceran lice, a sample including all families, the majority of host groups, and biogeographic regions across which Amblycera occur. Of these, 84 were newly sequenced for this study. We photographed individual specimens as vouchers and extracted total genomic DNA using a Qiagen QIAamp Micro Kit, with a 48-hour initial incubation [18]. We then prepared libraries from these extractions with a Hyper library kit (Kapa Biosystems) and sequenced them on an Illumina NovaSeq 6000, following the 150bp paired-end read protocol described by Johnson et al. [18]. We identified the vouchers to the genus level based on morphology using illustrations and keys [19, 20]. We also included data from six additional amblyceran species analysed by Sweet et al. [10] from NCBI SRA [21]. We conducted a quality check on the raw data from all 90 samples using FastQC v.0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed the reads using BBDuK from the BBMap package (https://sourceforge.net/projects/bbmap/, setup ktrim=r k=23 mink=7 hdist=1 tpe tbo maq=10 qtrim=rl trimq=35 minlength=35). We trimmed adapters automatically and manually trimmed the 5' and 3' ends using the forcetrim= argument.

Mitochondrial genome assembly and annotation

To avoid excessive coverage of the mitochondrial genome and decrease the computational demand, each sequence read library was subsampled for 2 million reads of Read1 and the corresponding Read2 (4 million total reads). The mitochondrial genomes were first assembled using MitoFinder v.1.4.1 [22] with MetaSPAdes as the assembler. From the MetaSPAdes results, contigs similar to the reference (i.e., concatenated nucleotide sequences of previously published mitogenomes; [10, 14]) were selected using TCSF v.2.7.1 [23] with default parameters. From the TCSF results, contigs with high coverage (typically exceeding 100X) and a cumulative length of at least 15 kb were manually selected. These contigs were tested for circularity in Simple-Circularise (https://github.com/Kzra/Simple-Circularise) and AWA [24], searching for *k*-mers up to 40 bp long, mapped on full trimmed reads without subsampling. Manual inspection of gene overlap and sequence similarity, along with the AWA results, further validated circularity (Table S3).

If the assembly failed to provide circular contigs encompassing all mitochondrial genes, the assembly procedure was repeated with subsampling increased to 8 million and 20 million total reads, respectively. Using MITOS2 [25], we annotated the contigs and identified genic regions overlapping the 3' and 5' ends. In the event MITOS2 failed to locate protein-coding genes (PCGs), we manually searched open reading frames (ORFs) identified by ORF

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Finder, part of Sequence Manipulation Suite, [26], and visually inspected ORFs in Jalview 2.11.2.0 [27].

Phylogenomics, dating, and ancestral state reconstruction

For phylogenomic analysis of Amblycera, we used whole genome sequencing data of 90 amblyceran taxa and 29 outgroup taxa from Ischnocera, Trichodectera, Rhynchophthirina, Anoplura, and free-living Psocodea (other taxa from the infraorder Nanopsocetae) [28]. For this analysis, reads were trimmed for adaptors and quality (phred score < 30) using *fastp* v0.20.1 [29] and converted to aTRAM 2.1 [30] databases. We assembled a target set of 2395 single-copy ortholog PCGs from the human head louse (*Pediculus humanus*) for each genomic library using aTRAM, using *tblastn* and the AbySS assembler (iterations=3 and max-target-seqs=3000). The resulting sequences were annotated using Exonerate with the reference protein-coding sequences, exons aligned, and genes trimmed using established tools and parameters [18, 31]. We conducted a phylogenomic analysis on the concatenated alignment under maximum likelihood and using the GTR+G model in IQ-TREE 2 v.2.1.2 [32]. Support was estimated using ultrafast bootstrapping (UFBoot2; [33] with 1000 replicates. We also performed a coalescent analysis in ASTRAL-III v5.7.4 [34] to account for gene-tree/species-tree discordance. Separate gene trees were inferred using IQ-TREE, based on the optimal models. Molecular dating analysis using the concatenated data set was conducted using Least Squares Dating (LSD2) in IQ-TREE 2 with calibration from previously published fossil and codivergence data (split between human and chimpanzee lice 5–7 Mya, split between the lice from Old World primates and Great Apes 20–25 Mya, the minimum age for Menoponidae of 44 Mya based on a fossil; [18, 31]) and root age 127.1 Mya [28]. Ancestral states of the mitogenome (single-chromosome or fragmented) were reconstructed over the dated tree using various models: Equal-Rates (ER), All-Rates-Different (ARD), and a model that does not allow transition from fragmented to single mitogenome organisation (USR). The best model was selected using the corrected Akaike Information Criterion (AICc). The best model (ER; AIC = 109.8825, AICc = 109.928, Table S19) revealed an almost 50% relative likelihood of fragmented ancestral amblyceran mitogenome. Given that more distant outgroups among Psocodea have non-fragmented mitogenomes, a fragmented ancestral state seems unlikely, so we also performed stochastic mapping with 1000 simulations for the USR models. For the reconstructions, we used the *ace* function of the APE v.5.4 R package [35]. The model fit was assessed with the *fitDiscrete* function in the GEIGER v.2.0.7 R package [36], and for the stochastic mapping, we used the PHYTOOLS v.0.7 R package [37].

To measure the strength of the phylogenetic signal of fragmentation, we calculated the D-statistic using the *comparative.data* and *phylo.d* functions of the CAPER v.1.0.1 R package [38] over the dated amblyceran tree.

Nucleotide composition

We calculated nucleotide composition of the mitogenome for six sub-datasets for each sample (all sites, coding regions, different codon positions for all three positions, and fourfold degenerate sites of concatenated PCGs) using *Bio* and *Bio.SeqUtils* packages of Biopython 1.80 [39]. We identified the fourfold degenerate sites using MEGA11 [40]. We performed statistical comparisons of AT content using GGPUBR v.0.40 R package [41], the *phylANOVA* function in the PHYTOOLS v.0.7 R package [37], taking into account the concatenated amblyceran tree. Additionally, we calculated AT content for each PCG separately. We used the average AT percentage of different PCGs to test for differences between genes with the Wilcoxon rank sum tests in the rstatix v.0.7.2 R package [42]. We also assessed the correlation between AT content and the length of mitochondrial chromosomes, fitting both linear and

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Phylogenetic Least Squares (PGLS) models to the total AT content and AT content of fourfold degenerate sites, taking into account the concatenated amblyceran tree. For fragmented mitogenomes, we tested both the average length of chromosomal fragments and their actual lengths. For PGLS, we employed the *pgls* function of the CAPER v.1.0.1 R package with both Brownian and Pagel's λ correlations.

Results

Repeated mitochondrial genome fragmentation is widespread among Amblycera

We assembled mitogenomes from 90 samples across 53 genera of Amblycera (Table S1), finding evidence for circularity in most chromosomes (186 out of 197, Table S3). Despite identifying all protein-coding genes in the majority of samples (78, Table S2, Tables S5-S17), there were some fragments that did not meet the circularity criteria. Our phylogenomic analysis based on 2395 nuclear orthologs provided a well-resolved and supported tree for Amblycera and there were only a few differences between the concatenated and coalescent trees, mostly involving rearrangements among some families (Figs. S1, S2, S3). When comparing the mitochondrial genome organisation among closely related taxa, we found a degree of similarity between these species (Fig. 1). Although many species had single circular mitochondrial chromosomes, there were also many instances of mitogenome fragmentation across the concatenated tree (Figs. 2 and S4), and some of this fragmentation occurred over short timescales (< 5 Mya; Figs. 2 and S4).

Our analysis suggests a significant phylogenetic signal for mitogenome fragmentation within Amblycera (D-value = 0.281, P = 0.001, Fig. S9), consistent with a Brownian motion model (P = 0.18). Some clades showed largely consistent levels of mitogenome fragmentation. Among the six currently recognized families of Amblycera (Boopiidae, Gyropidae, Laemobothriidae, Menoponidae, Ricinidae, Trimenoponidae; [19], all except Gyropidae contain samples with single-chromosome mitogenomes (Fig. 1). In particular, all samples from Ricinidae (6 spp.) and Boopiidae (2 spp.) possess single-chromosome mitochondrial genomes with largely consistent gene orders. Notably, Ricinidae, the sister group to all other Amblycera, retains the gene order *cox1-atp8-atp6-cox3*, also seen in free-living book lice [9]. The largest family, Menoponidae, which exclusively inhabits birds, displays a high degree of mitogenome structural variation, ranging from single chromosome to highly fragmented across multiple clades (Fig. 1). The families Gyropidae and Trimenoponidae, both exclusive to mammals, form a clade in our analysis, yet they are not mutually monophyletic. Within these families, mitogenome structure varies from single-chromosome (in *Trimenopon* and *Chinchillophaga*) to highly fragmented with up to nine mitochondrial chromosomes (in *Macrogyropus*).

Our taxon sample included multiple species within several amblyceran genera, allowing us to investigate variation in mitogenome structure among closely related taxa. In particular, we observed transitions from single-chromosome to fragmented mitogenomes within the genera *Menacanthus*, *Amyrsidea*, *Dennyus*, *Austromenopon*, and *Laemobothrion* (Fig. 1). Moreover, we observed changes in the number of fragments between species within *Myrsidea* and *Trinoton* (Fig. 1). Consistent with previous studies on Ischnocera [7], we found that gene order (Fig. 1, Table S2), even for single-chromosome genomes, was massively rearranged between lineages. However, in a few instances, the gene order remained stable among single-chromosome taxa, particularly within Ricinidae and Laemobothriidae (Fig. 1,

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S2). While not definitive, our data suggest a possible trend towards increasing fragmentation within the genus *Myrsidea*. The sister taxon (*Myrsidea sp.* ex *Corcorax melanorhamphos*) to the rest of *Myrsidea* possesses two chromosomes, while the remaining species of *Myrsidea* have three chromosomes, or with some more derived species exhibiting even four or five chromosomes. This pattern suggests a gradual increase in the number of fragments over the course of the evolution of *Myrsidea* (Fig. 1).

It is often assumed that once mitogenome fragmentation occurs, it is irreversible (e.g., [7]). Indeed, we found several highly supported cases where an ancestral state of a single chromosome transitions to a fragmented state (e.g., *Menacanthus*, *Nosopon*, *Piagetiella*, and *Eomenopon*). Despite this, our analysis did not reject the equal rates (ER) model for the evolution of mitochondrial fragmentation across Amblycera, and it was indeed the best-fitting model for this analysis (AIC = 109.8825, AICc = 109.928). When considering only Amblycera, this model suggests that the ancestral states of single versus fragmented have nearly equal likelihoods (Fig. S4). In this scenario, there are only three instances in the phylogenetic tree where a likely transition (>75% relative likelihood) from fragmented to single is inferred (*Hohorstiella* and two cases in *Laemobothrion*). However, for most taxa, the likelihood of the ancestral state does not strongly favour either state. One caveat is that our analysis does not take into account the genome organisation of more distant outgroups among free-living Psocodea, which is predominantly single, as is the case for most other insects. Therefore, it is likely that the ancestral state for Amblycera as a whole would be a single-chromosome in organisation (Fig. 2). The reconstruction that adopts this assumption (non-reversibility of fragmentation; Fig. 2) suggests at least 27 transitions from single-chromosome to fragmented mitochondrial genomes. While it remains to be seen if any definitive case of fragmentation reversal exists, further sampling within *Hohorstiella* and *Laemobothrion* could shed more light on this matter. Overall, fragmentation extensively varies across Amblycera, yet not to such an extent that it obscures overall evolutionary patterns.

Fragmented mitogenomes of Amblycera are less AT biased

Comparison of the base composition (AT percentage) of single-chromosome mitochondrial genomes to fragmented genomes indicates that the AT composition of single-chromosome genomes is significantly higher overall (Fig. S5, Table S4). Although the AT content of fourfold degenerate sites is higher than that of other positions (Fig. S6), the AT content at fourfold degenerate sites from single mitogenomes is notably higher than that of sequences from fragmented mitogenomes (Fig. S7). The decrease in AT bias in fragmented mitogenomes is also observed at all other positions (Fig. S7). These findings align with previous studies [7, 10], reinforcing the hypothesis of differing mutational or selective biases in fragmented versus single mitochondrial genomes. The differences in base composition across different protein-coding genes (Fig. S8, Tables S5–S18) can be primarily attributed to disparities between fourfold degenerate sites and other partitions. In general, fourfold degenerate sites more likely reflect mutational biases rather than direct selection. These differences are especially evident in *cox1-3* and *cob* genes, which also exhibit the most conservation in spatial arrangement (Fig. 1). We did not observe a significant correlation between chromosome length and AT content in fragmented mitogenomes.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Phylogenomics clarifies phylogenetic relationships within Amblycera, uncovering paraphyly of some families

Both concatenated and coalescent analyses strongly support the monophyly of Amblycera (100% support). Within Amblycera (Figs 1, S1, S2), our phylogenomic tree confirms the monophyly of the families Ricinidae, Laemobothriidae, and Boopidae (Fig. 1). However, both the concatenated and coalescent trees suggest that the families Gyropidae and Trimenoponidae are paraphyletic. These two families intertwine to form a single monophyletic clade of lice (Fig. 1, S1, S2) parasitizing Neotropical mammals, primarily rodents and marsupials [19]. The concatenated (Fig. S1) and coalescent (Fig. S2) trees differ in the positions of the genus *Trinoton* and family Boopiidae. In the concatenated tree (Fig. S1), the genus *Trinoton* is sister to Boopiidae, rendering Menoponidae paraphyletic. However, in the coalescent tree (Fig. S2), Boopiidae is sister to all other Amblycera, while *Trinoton* is sister to the remainder of Menoponidae making this latter family monophyletic. In both trees, the remainder of Menoponidae collectively forms a large monophyletic clade (Figs 1, S1, S2). At the generic level, our data also suggest some genera may not be monophyletic, including *Menacanthus, Hohorstiella*, *Colpocephalum*, and *Ricinus* (Figs 1, S1, S2).

Discussion

By leveraging an extensive dataset of 90 samples of lice in the parvorder Amblycera, and integrating mitogenome assembly with nuclear phylogenomics, we gained substantial insights into mitogenome fragmentation within a diverse group of parasitic lice. We observed that mitogenomes undergo fragmentation numerous times (potentially 27 or more), with some fragmentation events occurring as recently as a few million years ago (Figs. 2, S3, S4). In various genera represented by more than one sample, transitions from a single chromosome to a fragmented mitogenome, as well as increasing fragmentation into a larger number of fragments were apparent (Fig. 1). Despite the repeated and ongoing nature of mitogenome fragmentation, the fragmentation process shows some evidence of phylogenetic conservation. Certain louse groups appear to fragment more frequently and rapidly than others, and some broad clades are either entirely fragmented or entirely single-chromosome. Among the six recognised amblyceran families, fragmentation has evolved in at least four of these, although the monophyly of some families is not supported. The reconstruction of the state of the ancestral mitogenome from current data is somewhat uncertain, leading to questions about the direction of fragmentation. However, given the general conservation of mitogenome organisation across insects and in the common ancestor of Amblycera and free-living bark lice, it is highly probable that the common ancestor of Amblycera had a single mitogenome chromosome, with several transitions to fragmented mitogenomes occurring over time. Although transitions from single chromosome to fragmented chromosomes seems the most likely, the merging of chromosomes [43] through homologous and non-homologous mitochondrial recombination [44] has been suggested as a possible mechanism by which mitogenomes might merge back into a single chromosome. However, even under the ER model (Fig. S4), the majority of changes seem to be from single to fragmented chromosomes.

In terms of the process of fragmentation, we observed two primary patterns: 1) a small fragment encompassing a few genes splits from an original singe-chromosome mitogenome (e.g., *Laemobothrion*, Fig. 1, [10]; *Numidicola*, Fig. 1); 2) a single-chromosome mitogenome disintegrates into multiple smaller fragments of approximately equal size (e.g., *Myrsidea*, *Dennyus*, Fig. 1; a similar pattern observed in Ischnocera, [7]). These two patterns are not

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mutually exclusive, and they can co-occur within a single species (e.g., *Quateia,* Fig. 1). In addition, numerous mitogenomes possess a state intermediate to these extremes. Furthermore, the genes involved, fragment sizes, and number of fragments seem to display considerable variation among different fragmentation events. Cameron et al. [12] hypothesized that mitogenomes comprised of numerous small minicircles would evolve incrementally from a few larger fragments rather than directly from a single-chromosome genome. However, the mitogenomes of the genus *Dennyus* (Fig. 1) and the mitogenome of *Piagetiella* (Fig. 1) suggest that a "big bang" fragmentation event might indeed occur in lice, leading to a transition from a single to several mitochondrial chromosomes. Additional sampling within these genera would help clarify whether any intermediate states exist.

Connecting the patterns of mitogenome fragmentation with our dating analysis provides a more detailed picture of the temporal dynamics of mitogenome evolution (see Figs. 2, S3, S4). In some cases, mitogenome structure is stable for a long period of time, while in others it appears to change rapidly. In Ricinidae, both genome structure and gene order remained stable for at least around 30 million years (Figs 1, 2, S4). In contrast, over this same timeframe within the family Laemobothriidae, there are multiple transitions in mitogenome structure (Figs 1, 2, S4). In the genus *Myrsidea*, a relatively consistent three-fragment structure remained unchanged for an extended period across much of the genus's diversification (33-10 Mya). However, in more recently evolved species, additional fragmentation took place (Figs 2, S4). Given the notable diversity of *Myrsidea*, containing nearly 400 described species [19], it is likely that novel mitogenome configurations may be discovered in other species within this genus. In the case of *Dennyus*, the closest relative to *Myrsidea* (Figs 1, 2, S1, S2, S3, S4), it seems that its ancestor may have possessed a single-chromosome mitogenome which underwent a recent and rapid fragmentation event (within 2–5 My), resulting in six small minicircles (Figs 1, 2, S3, S4). However, intermediate states of fragmentation cannot be completely ruled out. Even though evidence suggests mitogenome structure typically remains stable within a given species [7], we uncovered instances where considerable variation can occur within a single genus over less than five million years (Figs 2, S4).

Our phylogenetic analysis of fragmentation (Figs. 1) indicates several potential fragmentation hotspots (i.e., lineages in which changes in architecture are more common than elsewhere). However, there is no obvious general pattern in the distribution of these hotspots, such as related to particular taxa, geography, or host associations. The most pronounced examples include the genera *Dennyus*, *Trinoton*, and Amblycera parasitizing pigeons (*Hohorstiella*, *Bonomiella*, and *Quateia*, Fig 1). These hotspots also represent hotspots of massive gene rearrangements, contrary to the cases of isolated fragmentation. In these latter cases, even when a genome fragments, the resulting fragments tend to retain the gene order of the ancestral genome (e.g., *Menacanthus*, *Amyrsidea*, some *Trinoton*, Fig. 1).

Our results also reveal that fragmented mitochondrial genomes exhibit a lower percent AT content compared to non-fragmented (single) ones. Compared to Ischnocera [7], the average AT percentage in Amblycera is marginally higher, possibly due to a higher proportion of single-chromosome mitogenomes. Regarding the AT percentage in different gene partitions, the fourfold degenerate sites could offer valuable insights into base composition biases, as they are not subjected to selection on the amino acid composition. Despite being significant, the difference in AT percentage between fourfold and non-fourfold degenerate

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sites in lice (approximately 5–10%, Figs S6, S7, [7]) is still smaller than the difference in other insects (around 20%, [45]). The disparities among different PCGs (Fig. S8) indicate that the most pronounced effects of base composition biases are evident in the *cox* genes, consistent with the relatively strong purifying selection in these genes [46]. Concerning the initial drivers of fragmentation in lice, although relaxed selection is typically associated with parasitism and reduced morphology, it does not account for why other arthropod parasites (e.g., fleas, ticks) lack fragmented mitogenomes or why fragmented mitogenomes also appear in free-living animals. Cameron et al. [12] posit one hypothesis for the extremely fragmented mitogenomes of Anoplura is the absence of *mtSSB*, a nuclear gene targeted to the mitochondrion that stabilizes mitochondrial DNA during the single-stranded replication stage. The presence of *mtSSB* in lice other than *Pediculus humanus* (Anoplura) is currently unknown; therefore, this hypothesis presents an avenue for further research.

Funding and acknowledgements

TN's work was funded by project No. 22-04386O of the Czech Science Foundation (GACR) "Coevolution of parasitic lice, their hosts and symbionts". This work was also supported by the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ infrastructure (ID:90140), by NSF DEB-1239788, NSF DEB-1925487, and NSF DEB-1926919 to KPJ and the European Commission grant H2020-MSCA-IF-2019 (INTROSYM:886532) to JD. We thank B. Benz, T. Chesser, D. Clayton, K. Dittmar, R. Faucett, T. Galloway, A. Grossi, F. Madeira, J. Malenke, M. Meyer, R. Moyle, B. O'Shea, R. Palma, V. Piacentini, A. Saxena, M. Valim, T. Valqui, J. Weckstein, and R. Wilson for assistance in obtaining samples for this study. We thank S. Virrueta Herrera for assistance with DNA extraction. We thank the A. Hernandez and C. Wright at the University of Illinois Roy J. Carver Biotechnology Center for assistance with Illumina sequencing. We also thank CSIRC personnel (Universidad de Granada, Spain) for assistance and providing computational resources (Albaicin supercomputer).We thank K. K. O. Walden for assistance with the submission of raw read files to NCBI.

Data and code availability

Data associated with this study are available in the Supplementary material and NCBI SRA. Other data are available on reasonable request. The code used to analyse mitogenomes is available on GitHub (https://github.com/tomas-najer/Amblycera-mitogenomes). Additional data, including photos of vouchers are available on figshare (https://figshare.com/account/home#/projects/169538).

Authors contribution.

TN assembled and analysed the mitochondrial genomes, wrote the first draft, edited the manuscript, and obtained funding. JD performed the phylogenomic and dating analyses and edited the manuscript. AB and ADS provided project conceptualization, supervision and edited the manuscript. OS provided a morphological determination of the lice and edited the manuscript. KPJ conceptualised and supervised the work, provided samples and sequences, edited the manuscript, and obtained funding.

Competing interests.

We declare we have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

[1] Wallace D.C. 1982. Structure and Evolution of Organelle Genomes. Microbiological Reviews, 46(2), 208–240. https://doi.org/10.1128/mr.46.2.208-240.1982

[2] Yoneda M., Katsumata K., Hayakawa M., Tanaka M., Ozawa T. 1995. Oxygen stress induces an apoptotic cell death associated with fragmentation of mitochondrial genome. Biochemical and Biophysical Research Communications, 209(2), 723–729. https://doi.org/10.1006/bbrc.1995.1559

[3] Wallace D.C. 1999. Mitochondrial Diseases in Man and Mouse. Science, 283(5407), 1482–1488. https://doi.org/10.1126/science.283.5407.1482

[4] Gibson R., Blok V.C., Dowton M. 2007. Sequence and Characterization of Six Mitochondrial Subgenomes from Globodera rostochiensis: Multipartite Structure Is Conserved Among Close Nematode Relatives. Journal of Molecular Evolution, 65, 308 – 315. https://doi.org/10.1007/s00239-007-9007-y

[5] Smith D.R., Kayal E., Yanagihara A.A., Collins A.G., Pirro S., Keeling P.J. 2011. First Complete Mitochondrial Genome Sequence from a Box Jellyfish Reveals a Highly Fragmented Linear Architecture and Insights into Telomere Evolution. Genome Biology and Evolution, 4(1), 52–58. https://doi.org/10.1093/gbe/evr127

[6] Dickey A.M., Kumar V., Morgan J.K., Jara-Cavieres A., Shatters R.G. Jr., 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

[7] Sweet A.D., Johnson K.P., Cameron S.L. 2022. Independent evolution of highly variable, fragmented mitogenomes of parasitic lice. Communications Biology, 5, 677. https://doi.org/10.1038/s42003-022-03625-0

[8] Chen X.J., Butow R.A. 2005. The organisation and inheritance of the mitochondrial genome. Nature Reviews Genetics, 6, 815–825. https://doi.org/10.1038/nrg1708

[9] Wei D.-D., Shao R., Yuan M.-L., Dou W., Barker S.C., Wang J.-J. 2012. The Multipartite Mitochondrial Genome of *Liposcelis bostrychophila*: Insights into the Evolution of Mitochondrial Genomes in Bilateral Animals. PloS ONE, 7(3), e33973. https://doi.org/10.1371/journal.pone.0033973

[10] Sweet A.D., Johnson K.P., Cao Y., de Moya R.S., Skinner R.K., Tan M., Herrera S.V., Cameron S.L. 2021. Structure, gene order, and nucleotide composition of mitochondrial genomes in parasitic lice from Amblycera. Gene, 768, 145312. https://doi.org/10.1016/j.gene.2020.145312

[11] 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. http://doi.org/10.7717/peerj.8759

[12] 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, 394. https://doi.org/10.1186/1471-2164-12-394

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[13] 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 Research, 19, 904–912. https://doi.org/10.1101/gr.083188.108

[14] 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. Systematic Biology, 68(3), 430–440. https://doi.org/10.1093/sysbio/syy062

[15]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. Journal of Molecular Evolution, 65, 589 – 604. https://doi.org/10.1007/s00239-007-9042-8

[16] Tamura K. 1992. The rate and pattern of nucleotide substitution in Drosophila mitochondrial DNA. Molecular Biology and Evolution, 9(5), 814–825. https://doi.org/10.1093/oxfordjournals.molbev.a040763

[17] Rong Z., Tu P., Xu P., Sun Y., Yu F., Tu N., Guo L., Yang Y. 2021. The Mitochondrial Response to DNA Damage. Frontiers in Cell and Developmental Biology, 9, 669379. https://doi.org/10.3389/fcell.2021.669379

[18] Johnson K.P., Weckstein J.D., Herrera S.V., Doña J. 2021. The interplay between host biogeography and phylogeny in structuring diversification of the feather louse genus *Penenirmus*. Molecular Phylogenetics and Evolution, 165, 107297. https://doi.org/10.1016/j.ympev.2021.107297

[19] Price R.D., Hellenthal R.A., Palma R.L., Johnson K.P., Clayton D.H. 2003. The chewing lice: world checklist and biological overview. Champaign, IL: Illinois Natural History Survey. https://doi.org/10.5962/bhl.title.154191

[20] Clay T. 1969. A key to the genera of the Menoponidae (Amblycera: Mallophaga: Insecta). Bulletin of the British Museum (Natural History) Entomology, 24, 3–26.

[21] 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. Biology Letters, 14, 20180141. http://dx.doi.org/10.1098/rsbl.2018.0141

[22] Allio R., Schomaker-Bastos A., Romiguier J., Prosdocimi F., Nabholz B., Delsuc F. 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Molecular Ecology Resources, 20(4), 892–905. https://doi.org/10.1111/1755-0998.13160

[23] Kinjo Y., Saitoh S., Tokuda G. 2015. An Efficient Strategy Developed for Next-Generation Sequencing of Endosymbiont Genomes Performed Using Crude DNA Isolated from Host Tissues: A Case Study of *Blattabacterium cuenoti* Inhabiting the Fat Bodies of Cockroaches. Microbes and Environments, 30(3), 208–220. https://doi.org/10.1264/jsme2.ME14153

[24] Machado D.J., Janies D., Brouwer C., Grant T. 2018. A new strategy to infer circularity applied to four new complete frog mitogenomes. Ecology and Evolution, 8(8), 4011–4018. https://doi.org/10.1002/ece3.3918

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[25] Donath A., Jühling F., Al-Arab M., Bernhart S.H., Reinhardt F., Stadler P.F., Middendorf M., Bernt M. 2019. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Research, 47(20), 10543–10552. https://doi.org/10.1093/nar/gkz833

[26] Stothard P. 2020. The Sequence Manipulation Suite: JavaScript Programs for Analysing and Formatting Protein and DNA Sequences. BioTechniques, 28(6), 1064–1225. https://doi.org/10.2144/00286ir01

[27] Waterhouse A.M., Procter J.B., Martin D.M.A., Clamp M., Barton G.J. 2009. Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. Bioinformatics, 25(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

[28] de Moya R.S., Yoshizawa K., Walden K.K.O., Sweet A.D., Dietrich C.H., Johnson K.P. 2021. Phylogenomics of Parasitic and Nonparasitic Lice (Insecta: Psocodea): Combining Sequence Data and Exploring Compositional Bias Solutions in Next Generation Data Sets. Systematic Biology, 70(4), 719–738. https://doi.org/10.1093/sysbio/syaa075

[29] Chen S., Zhou Y., Chen Y., Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

[30] 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. Evolutionary Bioinformatics, 14, 1–4. https://doi.org/10.1177/1176934318774546

[31] Johnson K.P., Matthee C., Doña J. 2022. Phylogenomics reveals the origin of mammal lice out of Afrotheria. Nature Ecology & Evolution, 6, 1205–1210. https://doi.org/10.1038/s41559-022-01803-1

[32] Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution, 37(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

[33] Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2017. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution, 35(2), 518–522. https://doi.org/10.1093/molbev/msx281

[34] Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics, 19(Suppl 6), 153. https://doi.org/10.1186/s12859-018-2129-y

[35] Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics, 20(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

[36] Pennell M.W., Eastman J.M., Slater G.J., Brown J.W., Uyeda J.C., FitzJohn R.G., Alfaro M.E., Harmon L.J. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics, 30(15), 2216–2218. https://doi.org/10.1093/bioinformatics/btu181

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[37] Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution, 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

[38] Fritz S.A., Purvis A. 2010. Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits. Conservation Biology, 24(4), 1042 – 1051. https://doi.org/10.1111/j.1523-1739.2010.01455.x

[39] Cock P.J.A., Antao T., Chang J.T., Chapman B., Cox C.J., Dalke A., Friedberg I., Hamelryck T., Kauff F., Wilczynski B., de Hoon M.J.L. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

[40] Tamura K., Stecher G., Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution, 38(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

[41] Kassambara A. 2020. ggpubr: “ggplot2” Based Publication Ready Plots.

[42] Kassambara A. 2023. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.2. https://rpkgs.datanovia.com/rstatix/

[43] Dong Y., Zhao M., Shao R. 2022. Fragmented mitochondrial genomes of seal lice (family Echinophthiriidae) and gorilla louse (family Pthiridae): frequent minichromosomal splits and a host switch of lice between seals. BMC Genomics, 23, 283. https://doi.org/10.1186/s12864-022-08530-8

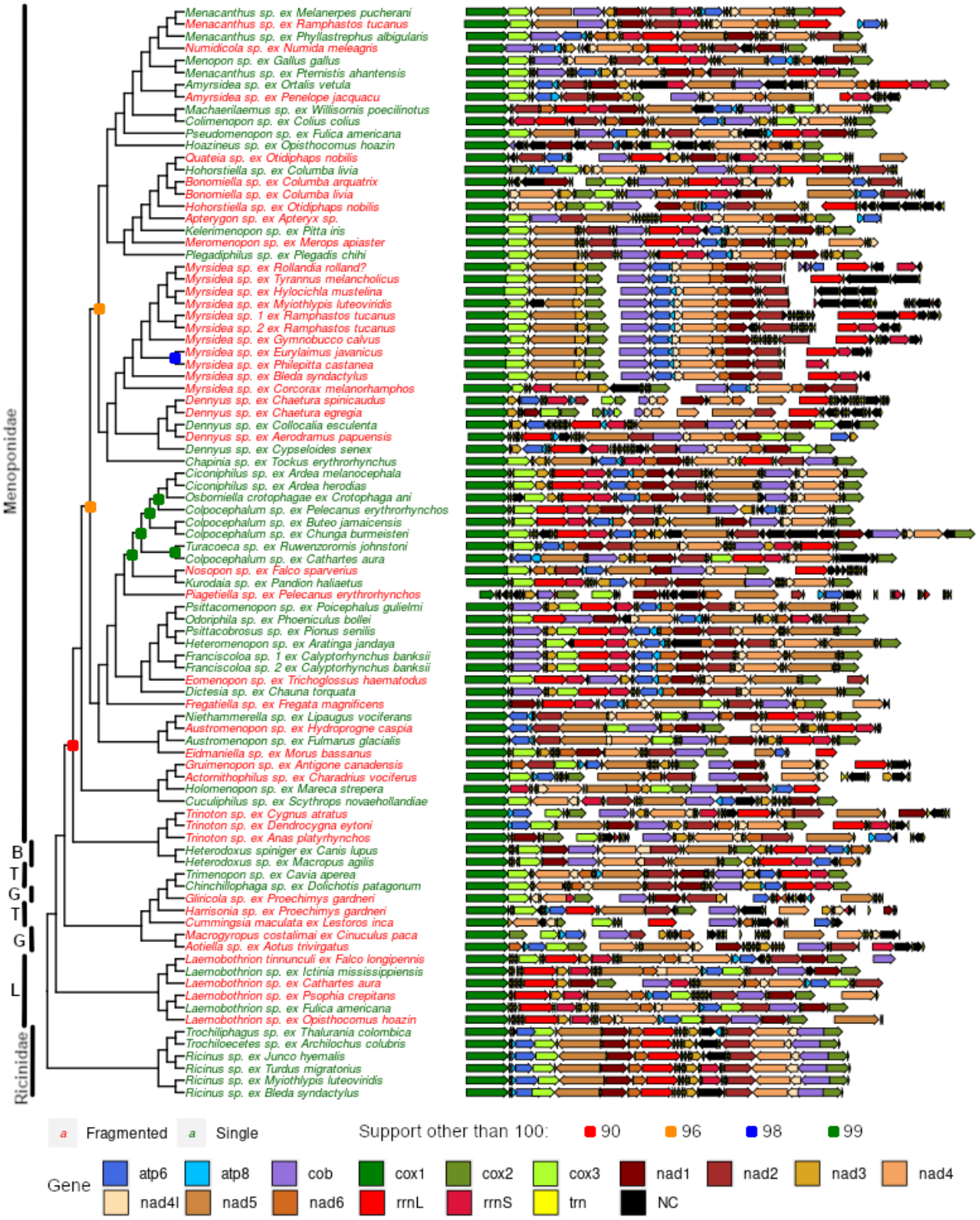
[44] Shao R., Barker S.C. 2011. Chimeric mitochondrial minichromosomes of the human body louse, Pediculus humanus: Evidence for homologous and non-homologous recombination. Gene, 473, 36–43. https://doi.org/10.1016/j.gene.2010.11.002

[45] Perna N.T., Kocher T.D. 1995. Patterns of Nucleotide Composition at Fourfold Degenerate Sites of Animal Mitochondrial Genomes. Journal of Molecular Evolution, 41, 353–358. https://doi.org/10.1007/BF00186547

[46] Castellana S., Vicario S., Saccone C. 2011. Evolutionary Patterns of the Mitochondrial Genome in Metazoa: Exploring the Role of Mutation and Selection in Mitochondrial Protein-Coding Genes. Genome Biology and Evolution, 3, 1067–1079. https://doi.org/10.1093/gbe/evr040

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1. Phylogeny and pattern of mitogenome organization across Amblycera. Chromosomes have been arbitrarily linearized starting with *cox1* for consistent directionality and to facilitate comparison. Gaps between genes indicate separate mitochondrial fragments. Taxa with fragmented mitogenomes are highlighted in red, and those with single-chromosome mitogenomes in green. All branches supported by 100% ultrafast bootstrap, except those indicated with coloured symbols at node. Abbreviations of amblyceran families: B – Boopidae; T – Trimenoponidae; G – Gyropidae; L – Laemobothriidae.



bioRxiv preprint doi: https://doi.org/10.1101/2023.06.22.546119; this version posted June 26, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

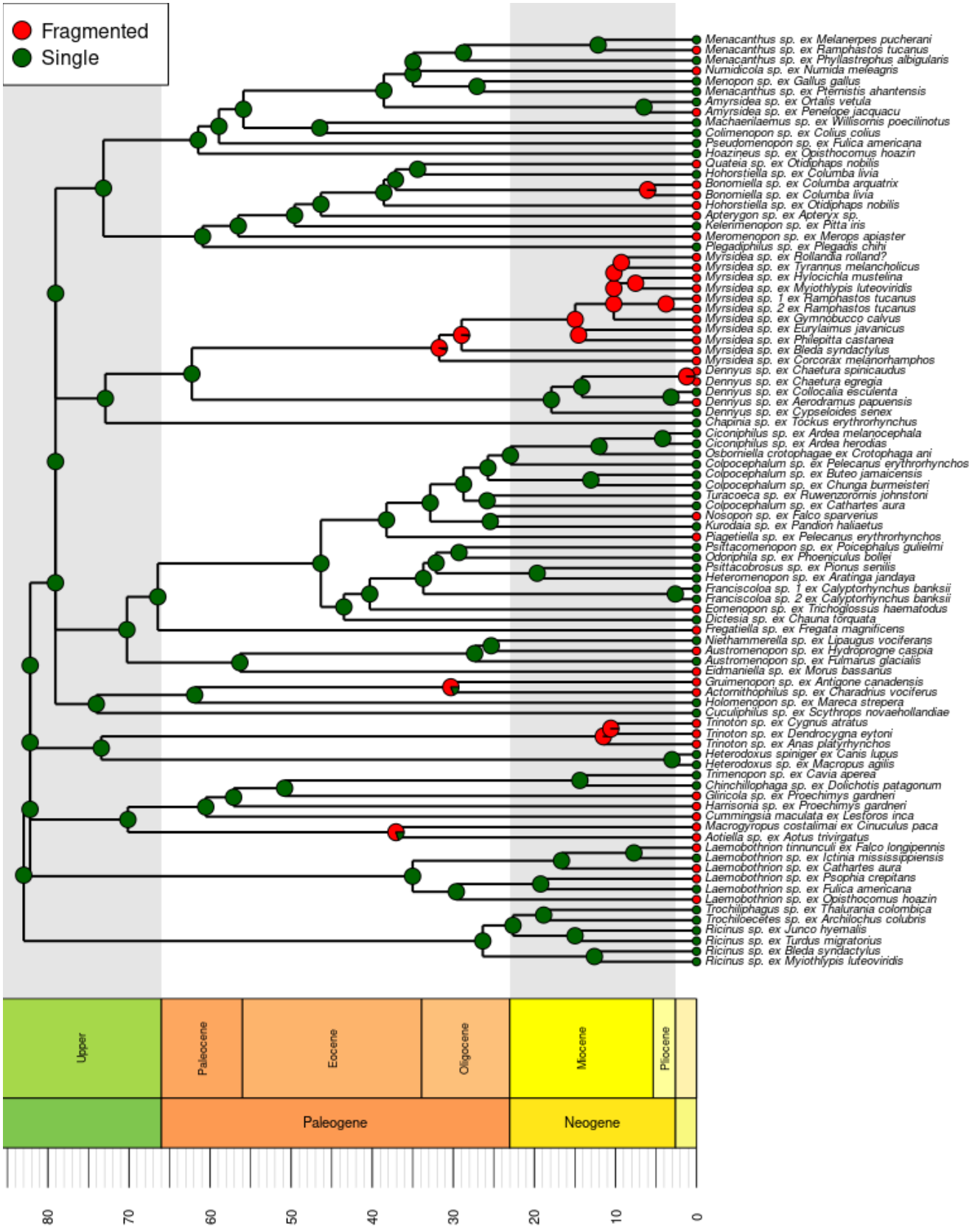



Figure 2. Dated phylogenomic tree with ancestral state reconstruction of mitogenome evolution in Amblycera under irreversible model. Circles at the tips indicate mitogenome structure (single-chromosome vs fragmented). Pie charts at the nodes show the frequency distribution of reconstructed ancestral state after 1000 simulations of stochastic character mapping using an irreversible fragmentation model (USR). Time scale at bottom in million years ago (mya).