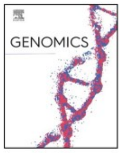
Genomics 112 (2020) 4924–4933

Contents lists available at ScienceDirect

Genomics

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



Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse *Pedicinus obtusus* and colobus louse *Pedicinus badii*

Yi-Tian Fu[a,1], Yalun Dong[b,1], Wei Wang[b], Yu Nie[a], Guo-Hua Liu[a,*], Renfu Shao[b,*]

[a] *Hunan Provincial Key Laboratory of Protein Engineering in Animal Vaccines, College of Veterinary Medicine, Hunan Agricultural University, Changsha 410128, PR China*
[b] *GeneCology Research Centre, School of Science and Engineering, University of the Sunshine Coast, Sippy Downs, Queensland 4556, Australia*

ARTICLE INFO

Keywords:
Sucking lice
Mitochondrial genome fragmentation
Minichromosome
Higher primates

ABSTRACT

We report for the first time the fragmented mitochondrial (mt) genomes of two *Pedicinus* species: *Pedicinus obtusus* and *Pedicinus badii*, and compared them with the lice of humans and chimpanzees. Despite being congeneric, the two monkey lice are distinct from each other in mt karyotype. The variation in mt karyotype between the two *Pedicinus* lice is the most pronounced among the congeneric species of sucking lice observed to date and is attributable to the opposite directions between them in mt karyotype evolution. Two of the inferred ancestral mt minichromosomes of the higher primate lice merged as one in the macaque louse whereas one of the ancestral minichromosomes split into two in the colobus louse after these two species diverged from their most recent common ancestor. Our results showed that mt genome fragmentation was a two-way process in the higher primate lice, and minichromosome merger was more common than previously thought.

1. Introduction

Higher primates (infraorder Semiiformes) comprise Old World monkeys, New World monkeys, apes and humans, with 320 extant species [1]. Higher primates originated ~45 million years ago (MYA) in East Asia [2–4] and expanded all over the world with the modern humans, *Homo sapiens*, the most widely distributed species. Blood-sucking lice (suborder Anoplura) parasitise eutherian mammals exclusively. Of the 540 species of sucking lice described to date, 20 species from three genera (representing three families) are found only on higher primates [5]. Humans are the host of three species of sucking lice: *Pediculus humanus capitis* (head louse), *Pediculus humanus corporis* (body louse) and *Pthirus pubis* (pubic louse). Chimpanzees and New World monkeys share the genus *Pediculus* (family Pediculidae) with humans. Two species of chimpanzees are the hosts of *Pediculus schaeffi*; 11 species of New World monkeys of the family Cebidae are the hosts of *Pediculus mjobergi*. Gorillas share the genus *Pthirus* (Phthiridae) with humans and are the hosts of *Pthirus gorilla*. Forty-one species of Old World monkeys (Cercopithecidae) are the hosts of 14 species of the genus *Pedicinus* (Pedicinidae) [6]. Sucking lice and higher primates have co-evolved for at least 25 million years, during which a mix of events such as parasite duplication, cospeciation, parasite extinction and host switching have occurred [7–9]. Lice have been a favoured model for co-evolutionary studies for several decades [7,10,11]. Host evolution can help understand louse evolution [9]; louse evolution can also help understand host evolution [7].

Mitochondrial (mt) genome information has been used widely in evolutionary studies of animals including parasitic lice [12–15]. Unlike most animals, sucking lice and some chewing lice have fragmented mt genomes with multiple (9 to 20) minichromosomes, in contrast to the typical single-chromosome mt genome of animals [15–26]. Human head and body lice have 20 minichromosomes and are the most fragmented, followed by chimpanzee louse with 18 minichromosomes and human pubic louse with at least 14 minichromosomes [17,18]. At the other end, lice of pigs and horses in the genus *Haematopinus* have nine minichromosomes [19,22], whereas rodent lice, a guanaco louse and four chewing lice are in the middle with 10–13 minichromosomes [15,20,21,24]. Shao et al., (2017) [25] showed that many minichormosome split events occurred in the human lice and the chimpanzee louse whereas minichormosome merger occurred only in the lice of pigs and horses. While 14 of the 20 species of higher primate lice are in the genus *Pedicinus* (family Pedicinidae), which parasitise Old World monkeys [5,6], nothing was known prior to this study about the mt genomes of *Pedicinus* lice except for the sequence of a segment of the

---

[*] Corresponding authors.
E-mail addresses: liuguohua5202008@163.com (G.-H. Liu), rshao@usc.edu.au (R. Shao).
[1] Yi-Tian Fu and Yalun Dong contributed equally.

https://doi.org/10.1016/j.ygeno.2020.09.005
Received 15 July 2020; Received in revised form 8 August 2020; Accepted 3 September 2020
Available online 06 September 2020

rrnS minichromosome of the langur louse, *Pedicinus ancoratus* [17].

To better understand the evolution of fragmented mt genomes in the higher primate lice, we determined the mt karyotypes of the macaque louse, *P. obtusus*, and the colobus louse, *P. badii*, and compared them with the human lice, chimpanzee louse and other sucking lice. Unexpectedly, the two monkey lice are distinct from each other in their mt karyotypes - the variation is the most pronounced among congeneric species of sucking lice observed to date. Their mt karyotypes evolved in opposite directions from one another after they diverged from their most recent common ancestor: ancestral minichromosomes merged in the macaque louse but split in the colobus louse.

2. Materials and methods

2.1. Collection of macaque lice, DNA extraction, sequencing, sequence assembly, minichromosome verification and mitochondrial genome annotation

Sucking lice were collected from an adult female rhesus macaque, *Macaca mulatta*, at Changsha Ecological Zoo, Hunan, China, and were identified as *Pedicinus obtusus* based on morphological keys (Fig. 1) [27,28]. *Macaca mulatta* is the host of two *Pedicinus* species: *Pedicinus eurygaster* and *Pedicinus obtusus* [5]. All of the specimens we collected were *Pedicinus obtusus*; no *Pedicinus eurygaster* was found from this macaque. The louse specimens were washed thoroughly in physiological saline solution and kept in ethanol at −20 °C. Genomic DNA was extracted from 50 specimens of *Pedicinus obtusus* (25 females and 25 males) with Wizard® SV Genomic DNA Purification System (Promega). Genomic DNA was checked for quality and concentration by agarose-gel electrophoresis and Bioanalyzer (2100, Agilent). A sequencing library (400 bp inserts) was constructed and paired-end sequencing was carried out on Illumina Hiseq 2500 at Majorbio (Shanghai). Raw sequence reads were cleaned by removing adaptor reads, redundant reads and 'N'-rich reads. 2 Gb clean data (250 bp each, paired-end reads) was obtained. *cox1* fragment (650 bp) of *Pedicinus obtusus* was amplified by PCR (polymerase chain reaction) with a primer pair mtd6 (5′-GGAGGATTTGGAAATTGATTAGTTCC-3′) – mtd11 (3′-ACTGTAAATATATGATGAGCTCA-5′); this fragment was sequenced using Sanger method at the Sangon Company (Shanghai). The *cox1* sequence of *Pedicinus obtusus* and the *rrnS* sequence of *Pedicinus ancoratus* (GenBank accession # EU219983) [17] were used as the initial references to assemble the Illumina sequence reads with Geneious 11.1.5 [29]. The assembly parameters were minimum overlap 150 bp and minimum identity 98% for *cox1* sequence, and minimum overlap 60 bp and minimum identity 60% for *rrnS* sequence. *cox1* and *rrnS* minichromosomes were assembled in full length for the coding regions and, from each end of the coding region, extended ~400 bp into the non-coding regions. Conserved non-coding region sequences between the *cox1* and *rrnS* minichromosomes immediately upstream and downstream the coding region were identified by sequence alignment and were used as references to align with the Illumina sequence dataset. This allowed us to extract the sequence reads derived from the two ends of the coding regions of all other mt minichromosomes. We then assembled these minichromosomes individually in full length for the coding region and extended the contigs into the non-coding regions as we did above for *cox1* and *rrnS* minichromosomes. A pair of specific outward primers was then designed from the coding region of each minichromosome and used in PCR to verify the size and circular organization of each minichromosome of *Pedicinus obtusus* (Table S1, Fig. S1). The forward and reverse primers in each pair were next to each other with a small gap (10–50 bp) in between; PCRs with these primers amplified each circular minichromosome in full length except for the small gap between the two primers. The amplicons from each minichromosome were sequenced individually with Illumina Hiseq 2500 platform as described above, thus the full-length sequences of both coding and non-coding regions of each minichromosome were obtained. Sequences of the mt minichromosomes of the macaque louse were aligned with those of human lice and chimpanzee louse with MAFFT 7.122 [30] to help identify genes and gene boundaries. tRNA genes were identified using ARWEN [31] and tRNAscan-SE [32]. Protein-coding genes and rRNA genes were identified by BLAST searches of GenBank [33].

Fig. 1. The macaque louse, *Pedicinus obtusus* (male), showing three pairs of paratergal plates. Both *Pedicinus obtusus* (Rudow, 1869) and *Pedicinus eurygaster* (Burmeister, 1838) are recorded from *Macaca mulatta* (Lance & Musser, 1994b). These two species can be differentiated by the number of paratergal plates on the abdomen. *Pedicinus obtusus* has three pairs of paratergal plates whereas *Pedicinus eurygaster* has two pairs of paratergal plates (Ferris, 1934; Ferris, 1951).

2.2. Retrieval and analysis of primate louse sequence data from the Sequence Read Archive

We retrieved whole genome sequence data from the publicly available Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra/) for the colobus louse, *Pedicinus badii* (SRR5308136) and the human pubic louse, *Pthirus pubis* (SRR5088475). *Pedicinus badii* is found on three host monkeys: *Colobus polykomos*, *Procolobus badius* and *Procolobus rufomitratus* [5]. The SRA sequence reads (150 bp for *Pedicinus badii* and 160 bp for *Pthirus pubis*, paired-end reads) were assembled with Geneious. For *Pedicinus badii*, partial sequences of mt genes *cob* and *cox1* retrieved from GenBank (accession # FJ267436 and HM171438) were used as initial references for SRA data assembly with the parameters: minimum overlap 60 bp and minimum identity 98%. *cob* and *cox1* minichromosomes were assembled in full length for the coding regions and ~ 250 bp into the non-coding regions from each end of the coding region. Conserved non-coding region sequences between *cob* and *cox1* minichromosomes were identified and were used to extract the SRA reads from the two ends of the coding regions of all other mt minichromosomes. These minichromosomes were then assembled individually to obtain their full-length coding regions flanked by

Table 1

Species of parasitic lice included in the phylogenetic analyses in this study.

Species	Hosts	GenBank accession numbers	References
Pedicinus obtusus	Macaque	MT792495–506	Present study
Pedicinus badii	Colobus	MT721726–39	Present study
Pediculus humanus corporis	Human	FJ499473–90	Shao et al., 2009
Pediculus humanus capitis	Human	JX080388–407	Shao et al., 2012
Pediculus schaeffi	Chimpanzee	KC241882–97, KR706168–69	Herd et al., 2015
Pthirus pubis	Human	JQ976018, EU219987–95, HM241895–8, MT721740	Shao et al., 2012 Present study
Polyplax asiatica	Greater bandicoot rat	KF647751–61	Dong et al., 2014b
Polyplax spinulosa	Asian house rat	KF647762–72	Dong et al., 2014b
Hoplopleura kitti	Bower's white toothed rat	KJ648933–43	Dong et al., 2014a
Hoplopleura akanezumi	Chevrier's field mouse	KJ648932	Dong et al., 2014a
Haematopinus apri	Wild pig	KC814611–19	Jiang et al., 2013
Haematopinus suis	Domestic pig	KC814602–10	Jiang et al., 2013
Haematopinus asini	Horse	KF939318, KF939322, KF939324, KF939326, KJ434034–38	Song et al., 2014
Microthoradus praelongiceps	Guanacos	KX090378-KX090389	Shao et al., 2017
Haematomyzus elephantis	Elephant	KF933032- KF933041	Shao et al., 2015

~250 bp non-coding region sequences. For *Pthirus pubis*, the conserved non-coding region sequences reported in Shao et al. [18] were used to search and assemble the minichromosomes that contained the three genes (*nad4*, *trnK*, *trnN*) that were not identified in Shao et al. [18]. Mitochondrial genes of *Pedicinus badii* and *Pthirus pubis* were identified using the same approach as for *Pedicinus obtusus* (see Section 2.1).

2.3. Phylogenetic analysis

The mt genome sequences of *Pedicinus obtusus*, *Pedicinus badii* and *Pthirus pubis* obtained in the present study were combined with the sequences of 12 other species of sucking lice and the elephant louse, *Haematomyzus elephantis* (used as outgroup), which were available in GenBank from previous studies (Table 1). Deduced amino acid sequences of each mt protein-coding gene except for *nad1*, *nad2*, *nad3* and *nad5* (sequences of these four genes not available to all louse species above) were aligned individually using MAFFT 7.122 and were then concatenated to form a single dataset; ambiguous sites were excluded from further analysis using Gblocks 0.91b [34]. Phylogenetic analyses were conducted using Bayesian inference method (BI) with MrBayes 3.2.6 [35] and Maximum likelihood (ML) with PhyML 3.0 [36]. For BI analysis, four independent Markov chains were run for 1,000,000 metropolis coupled MCMC generations, sampling a tree every 100 generations. The first 2500 trees represented burn-in, and the remaining trees were used to calculate Bayesian posterior probabilities (Bpp). The analysis was run until the potential scale reduction factor approached 1 and the average standard deviation of split frequencies was < 0.01. ML analysis was partitioned by gene and bootstrap was performed using the rapid bootstrapping option with 100 iterations; the MtREV model was used as selected by ProtTest 2.4 [37] based on the Akaike information criterion (AIC). Phylogenetic trees were drawn using FigTree v.1.31 (http://tree.bio.ed.ac.uk/software/figtree).

2.4. Inference of the ancestral mitochondrial karyotype of higher primate lice

We used a parsimony method described in Shao et al. [25] to infer the ancestral mt karyotype of higher primate lice. Three families of sucking lice parasitise higher primates exclusively: Pedicinidae, Pediculidae and Pthiridae; each family has a single genus – *Pedicinus* (14 species on Old World monkeys), *Pediculus* (2 subspecies on humans, 1 species on chimpanzees, 1 species on New World monkeys) and *Pthirus* (1 species on gorillas, 1 species on humans) respectively [5]. Each of these families and genera is monophyletic with *Pediculus* and *Pediculus* most closely related and sister to *Pediculus* [8,9]. We inferred a mt genome character to be ancestral to higher primate lice if the character was shared: 1) by species of all of the three genera; or 2) by *Pediculus* species and either *Pediculus* or *Pthirus* species; or 3) by species of one or more of the three genera and other parasitic louse species (Fig. 2, Fig. S2). For each inferred ancestral mt genome character to higher primate lice, we counted the changes required to explain the observed character data among the species included in our analysis (Table 1, Fig. 2, Fig. S2). If two or more mt genome characters conflict with each other, the character with the minimum required changes was inferred to be the ancestral whereas others rejected. As tRNA genes were much more mobile than protein-coding and rRNA genes [25], we considered tRNA genes separately from protein-coding and rRNA genes.

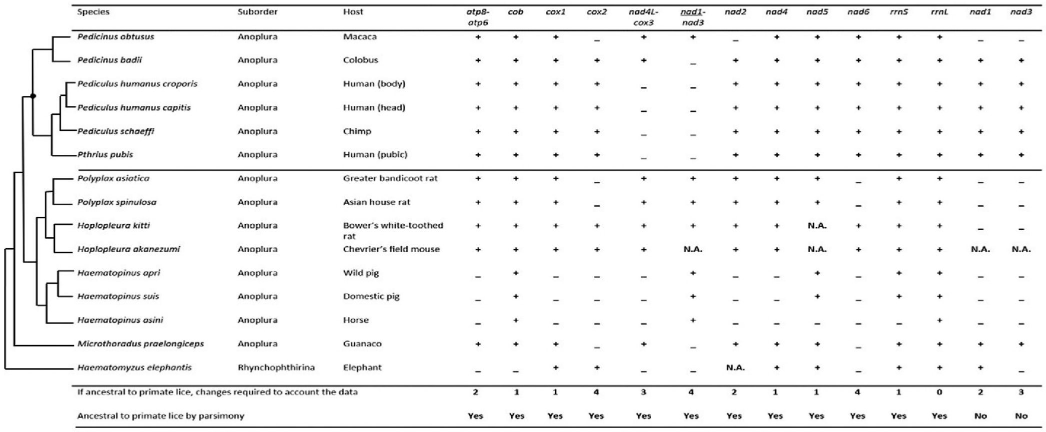
3. Results

3.1. Mitochondrial genome of the macaque louse, Pedicinus obtusus, comprises 12 minichromosomes

We obtained a large number of clean sequence reads for *P. obtusus* using Illumina paired-end sequencing: 1,611,687 pairs from the genomic DNA, and 17,886,211 pairs from the PCR amplicons of individual minichromosomes (Table 2); each sequence read is 250 bp. We assembled these sequence reads into contigs and identified all of the 37 mt genes typical of animals. These genes are on 12 minichromosomes; each minichromosome is 2850 to 3561 bp in size, containing a coding region and a non-coding region (NCR) in a circular organization (Fig. 3). The coding regions have 1 to 7 genes each and range from 787 bp to 1671 bp; the non-coding regions range from 1551 to 2177 bp (Table 2, Fig. 3). Eight minichromosomes have a single protein-coding or rRNA gene each; the other four minichromosomes have two protein-coding genes each. All genes are transcribed in the same orientation relative to the NCR except for *nad1*, which is opposite to all other genes in transcription orientation. *rrnL* gene is found in two types of minichromosome: in one type $trnL_1(tag)$ is upstream *rrnL*, in the other type $trnL_2(taa)$ is upstream *rrnL*. $trnL_1(tag)$ and $trnL_2(taa)$ have identical sequences except for the third anti-codon position (Fig. 4). In the sequence-read assembly of $trnL_1(tag)$ and $trnL_2(taa)$, the third anticodon position was covered 68,396 times in total, of which 41,566 (60.8%) was for $trnL_2(taa)$ and 26,735 (39.2%) was for $trnL_1(tag)$. Each coding region is flanked by a conserved non-coding AT-rich motif (146 bp,71.2%) upstream and a GC-rich motif (59 bp, 54.2%) downstream (Fig. S3). The annotated mt genome of *Pedicinus obtusus* is available in GenBank (Accession numbers MT792495–506).

3.2. Mitochondrial genome of the colobus louse, Pedicinus badii, comprises 14 minichromosomes

The retrieved SRA data of *Pedicinus badii* contains 24,549,545 paired-end sequence reads. Each sequence read is 100 bp in length.



Species	Suborder	Host	atp8-atp6	cob	cox1	cox2	nad4L-cox3	nad1-nad3	nad2	nad4	nad5	nad6	rrnS	rrnL	nad1	nad3
Pedicinus obtusus	Anoplura	Macaca	+	+	+	_	+	+	_	+	+	+	+	+	_	_
Pedicinus badii	Anoplura	Colobus	+	+	+	+	+	_	+	+	+	+	+	+	+	+
Pediculus humanus croporis	Anoplura	Human (body)	+	+	+	+	_	_	+	+	+	+	+	+	+	+
Pediculus humanus capitis	Anoplura	Human (head)	+	+	+	+	_	_	+	+	+	+	+	+	+	+
Pediculus schaeffi	Anoplura	Chimp	+	+	+	+	_	_	+	+	+	+	+	+	+	+
Pthrius pubis	Anoplura	Human (pubic)	+	+	+	+	_	_	+	+	+	+	+	+	+	+
Polyplax asiatica	Anoplura	Greater bandicoot rat	+	+	+	_	+	+	+	+	+	_	+	+	_	_
Polyplax spinulosa	Anoplura	Asian house rat	+	+	+	_	+	+	+	+	+	_	+	+	_	_
Hoplopleura kitti	Anoplura	Bower's white-toothed rat	+	+	+	+	+	+	+	+	N.A.	+	+	+	_	_
Hoplopleura akanezumi	Anoplura	Chevrier's field mouse	+	+	+	+	+	N.A.	+	+	N.A.	+	+	+	N.A.	N.A.
Haematopinus apri	Anoplura	Wild pig	_	+	_	_	_	+	_	_	+	_	+	+	_	_
Haematopinus suis	Anoplura	Domestic pig	_	+	_	_	_	+	_	_	+	_	+	+	_	_
Haematopinus asini	Anoplura	Horse	_	+	_	_	_	+	_	_	_	_	_	+	_	_
Microthoradus praelongiceps	Anoplura	Guanaco	+	+	+	_	+	_	+	+	+	_	+	+	+	+
Haematomyzus elephantis	Rhynchophthirina	Elephant	_	_	+	+	_	_	N.A.	+	+	_	+	+	+	_
If ancestral to primate lice, changes required to account the data			2	1	1	4	3	4	2	1	1	4	1	0	2	3
Ancestral to primate lice by parsimony			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No

Fig. 2. Inference of the ancestral mitochondrial karyotype (tRNA genes excluded) of higher primate lice. Plus (+) indicates presence; minus (−) indicates absence. The phylogeny was from the BI tree generated in this study (Fig. 6). The character of *nad1-nad3* on the same minichromosome conflicts with these two genes on separate minichromosomes. The minimum amount of changes required for *nad1-nad3* on the same minichromosome (4 changes) is less than *nad1* and *nad3* on separate minichromosomes (2 + 3 = 5). Thus, *nad1-nad3* is inferred to be the ancestral.

Fourteen coding regions were assembled in full length containing the 37 mt genes typical of animals. Each coding region is flanked by a conserved non-coding AT-rich motif (54 bp 96.3%) upstream and a GC-rich motif (33 bp 56.7%) downstream (Fig. S4) – the same pattern seen in *Pedicinus obtusus* and all other sucking lice sequenced to date, indicating that the mt genome of *Pedicinus badii* comprises 14 minichromosomes (Fig. 5A). The 14 coding regions range from 74 bp (*trnS*$_2$ minichromosome) to 1664 bp (*trnH-nad5* minichromosome) (Table 3). Eleven of the 14 coding regions have a single protein-coding or rRNA gene each; two coding regions have two protein-coding genes each; and one coding region has a single tRNA gene, *trnS*$_2$. The other 21 tRNA genes are scattered among 11 coding regions, each coding region

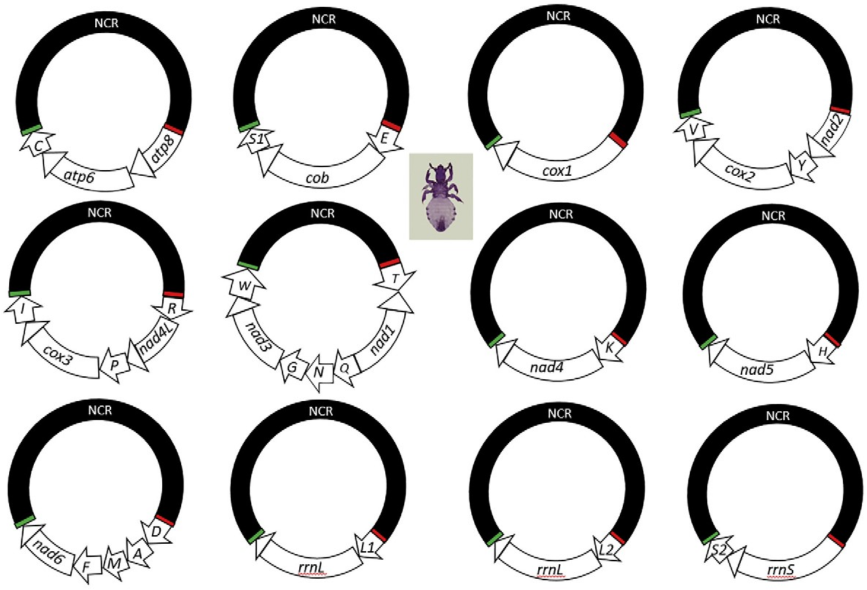


Fig. 3. The mitochondrial genome of macaque louse, *Pedicinus obtusus*. Each minichromosome has a coding region and a non-coding region (NCR, in black). Names and transcription orientation of genes are indicated in the coding region; minichromosomes are in alphabetical order of protein-coding and rRNA gene names. Gene names are: *atp6* and *atp8* for ATP synthase subunits 6 and 8; *cob* for cytochrome *b*; *cox1–3* for cytochrome *c* oxidase subunits 1–3, *nad1–6* and *nad4L* for NADH dehydrogenase subunits 1–6 and 4 L; *rrnS* and *rrnL* for small and large subunits of ribosomal RNA. tRNA genes are indicated with their single-letter abbreviations of the corresponding amino acids.

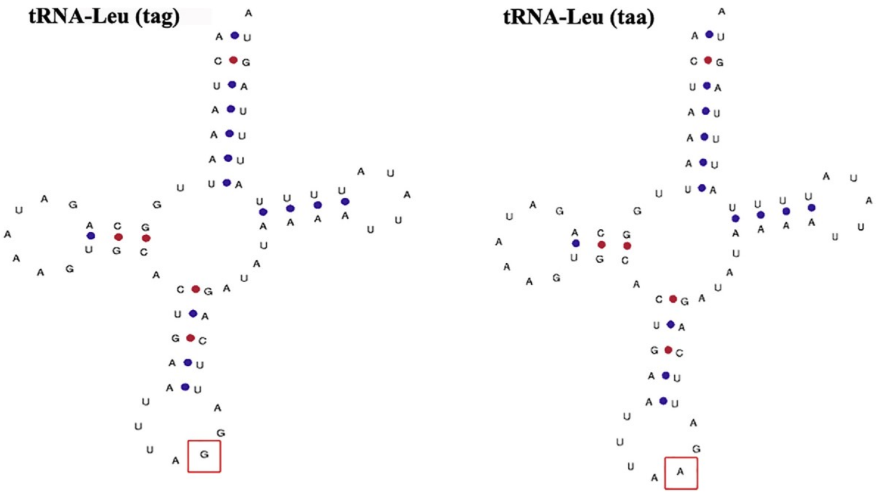


Fig. 4. Inferred secondary structures of the mitochondrial tRNA-Leu (tag) and tRNA-Leu (taa) of *Pedicinus obtusus*. The gene for these two tRNAs, *trnL$_1$(tag)* and *trnL$_2$(taa)*, have identical sequences except for the third anti-codon position.

with one to five tRNA genes (Fig. 5A). Each of the 37 mt genes is present in only one coding region; there is no overlap in gene content between different coding regions. The annotated mt genome of *Pedicinus badii* is available in GenBank (Accession numbers MT721726–39).

3.3. *trnK-nad4* mitochondrial minichromosome of the human pubic louse, *Pthirus pubis*, revealed by SRA data

Fourteen mt minichromosomes containing 34 genes in total have been reported for the human pubic louse, *Pthirus pubis*, in Shao et al. [18]; three mt genes, *nad4*, *trnK* and *trnN*, however, were not found in that study. In the present study, we assembled the SRA data of *Pthirus pubis* and identified *nad4* and *trnK* of this louse; these two genes were together on one minichromosome (Fig. 5B). A conserved AT-rich motif (111 bp 76%) is upstream and a GC-rich motif (77 bp 58%) is downstream the coding region, as seen in the 14 minichromosomes reported previously in Shao et al. [18]. We could not find *trnN* gene of *P. pubis* in our SRA data analysis. The annotated *K-nad4* minichromosome of *Pthirus pubis* is available in GenBank (Accession number MT721740).

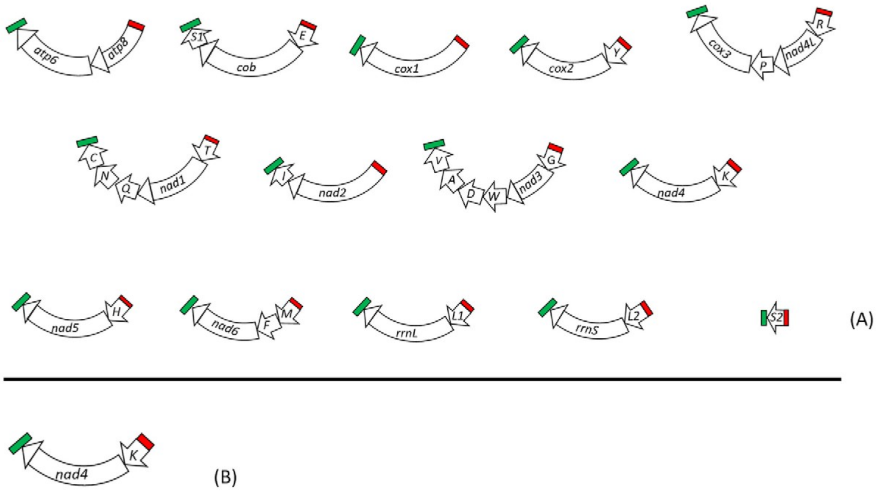


Fig. 5. (A) The coding regions of the 14 minichromosomes and conserved AT-rich region (in red) and conserved GC-rich region (in green) of the colobus louse, *Pedicinus badii*; (B) the coding region of *K-nad4* minichromosomes and conserved AT-rich region (in red) and conserved GC-rich region (in green) of the human pubic louse, *Pthirus pubis*. Names and transcription orientation of genes are indicated in the coding region; minichromosomes are in alphabetical order of protein-coding and rRNA gene names. See Fig. 3 legend for the abbreviation of gene names. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Mitochondrial minichromosomes of the macaque louse, *Pedicinus obtusus.*

Minichromosome	Size (bp)	Coding region (bp)	Non-coding region (bp)	Number of Illumina sequence-reads[a]
atp8-atp6-C	3000	907	2093	690,229
E-cob-S_1	3146	1182	1964	1,345,657
cox1	3120	1548	1572	1,555,266
nad2-Y-cox2-V	3561	1799	1762	1,457,577
R-nad4L-P-cox3-I	3027	1261	1766	1,436,297
T-nad1-Q-N-G-nad3-W	3524	1470	2054	1,304,553
K-nad4	2919	1368	1551	1,792,728
H-nad5	3406	1671	1735	1,084,208
D-A-M-F-nad6	2864	787	2077	1,900,118
L_1*-rrnL*	3009	1130	1879	1,684,859[b]
L_2*-rrnL*	3009	1130	1879	1,684,859[b]
rrnS-S_2	3004	827	2177	1,766,918
Total	37,589	15,080	22,509	17,703,269

[a] Each minichromosome was sequenced individually in full length.

[b] L_1*-rrnL* minichromosome and L_2*-rrnL* minichromosome are from the same sequence-read assembly.

Table 3
Mitochondrial minichromosomes of the colobus louse, *Pedicinus badii.*

Minichromosome	5′ end non-coding region (bp, partial)	Coding region (bp, full length)	3′ end non-coding region (bp, partial)	Number of Illumina sequence-reads
atp8-atp6	212	896	175	22,311
E-cob-S_1	214	1226	176	21,695
cox1	214	1552	177	27,314
Y-cox2	208	734	187	20,672
R-nad4L-P-cox3	212	1200	190	21,730
T-nad1-Q-N-C	220	1192	170	20,843
nad2-I	208	1046	198	22,082
G-nad3-W-D-A-V	215	692	189	20,430
K-nad4	239	1361	163	23,742
H-nad5	209	1664	205	22,636
M-F-nad6	209	676	199	19,667
L_1*-rrnL*	207	1161	174	25,755
L_2*-rrnS*	219	705	212	18,765
S_2	206	74	195	12,168
Total	3001	14,179	2610	299,810

3.4. Phylogeny of sucking lice reconstructed with mitochondrial genome sequences

Bayesian phylogenetic analysis of mt genome sequences divided the 14 species of sucking lice from seven of the 15 Anoplura families into three major clades with strong support: 1) the guanaco louse alone as the earliest branch, 2) the primate lice together as one clade, and 3) the lice of rodents, pigs and horses as one clade, sister to the primate lice (Fig. 6, Fig. S5). Within the clade of primate lice, the human body louse and the human head louse are most closely related to each other and are sister to the chimpanzee louse; all these three species are in the genus *Pediculus*. The human pubic louse is most closely related to the three *Pediculus* species. The macaque louse and the colobus louse, both in the genus *Pedicinus*, are most closely related to each other and are sister to the lice of humans and chimpanzees. In the third clade, the rodent lice in the genus *Hoplopleura* are more closely related to the pig and horse lice in the genus *Haematopinus* than to the rodent lice in the genus *Polyplax*. All groupings are well supported with posterior probability values at 0.87–1.0 (Fig. 6).

3.5. Inferred ancestral mitochondrial karyotype of the higher primate lice

We inferred the ancestral mt karyotype of higher primate lice based on the data available from two *Pedicinus* species (the current study), three *Pediculus* species [17,18,23] and *Pthirus pubis* [18], in conjunction with the data from other eight species of sucking lice and the elephant louse (Fig. 2, Fig. 7, Fig. S2). The *Pedicinus*, *Pediculus* and *Pthirus* species represent all of the three families of sucking lice (Pedicinidae, Pediculidae and Pthiridae) that are found exclusively on higher primates [5]. We were able to establish only the distribution and arrangement of all protein-coding genes and rRNA genes in the ancestral mt karyotype of higher primates (Fig. 7). Excluding tRNA genes, the inferred ancestral mt karyotype of higher primate lice consists of 12 minichromosomes; each minichromosome has a coding region and a non-coding region (Fig. 7). Nine of the 12 minichromosomes have a single protein-coding or rRNA gene each; the other three minichromosomes have two protein-coding genes each: 1) *atp6* and *atp8*, 2) *cox3* and *nad4L*, and 3) *nad3* and *nad1* (*nad1* in opposite orientation of transcription to that of *nad3* and all other genes). The data from the six louse species of higher primates sequenced to date are not sufficient for us to establish the distribution and arrangement of four tRNA genes (*trnM*, *trnD*, *trnN*, *trnV*) in the ancestral mt karyotype (Fig. S2). It is known that tRNA genes are much more mobile than protein-coding and rRNA genes [25]; thus, data from more primate louse species are needed in order to establish the ancestral mt karyotype of higher primate lice with all tRNA genes included.

4. Discussion

4.1. Variation in mitochondrial karyotype between Pedicinus badii and Pedicinus obtusus is the most pronounced among congeneric species of sucking lice observed to date

Despite in the same genus, *Pedicinus badii* and *Pedicinus obtusus* are distinct from each other in their mt karyotypes. *Pedicinus obtusus* has 12 minichromosomes (Fig. 3) whereas *Pedicinus badii* has 14 minichromosomes (Fig. 5). The two *Pedicinus* species have only five minichromosomes in common; the remaining minichromosomes are very different between them. In *Pedicinus obtusus*, *cox2* and *nad2* are together on the same minichromosome; in *Pedicinus badii*, however, these two genes are on two separate minichromosomes (Fig. 3, Fig. 5). Likewise, *nad1* and *nad3* are together on the same minichromosome in *Pedicinus obtusus* but are on two separate minichromosomes in *Pedicinus badii.* *trn*S_2 is on the *rrnS* minichromosome in *Pedicinus obtusus* but is on its own minicromosome in *Pedicinus badii.* Furthermore, *nad1* has the opposite orientation of transcription to all other genes in *Pedicinus obtusus* but has the same orientation as all other genes in *Pedicinus badii.* In addition to *trn*S_2, six other tRNA genes, *trnA*, *trnC*, *trnD*, *trnI*, *trn*L_2 and *trnV*, are also on different minichromosomes between the two *Pedicinus* species (Fig. 3, Fig. 5).

The variation in mt karyotype between the two *Pedicinus* species is the most pronounced among congeneric species of sucking lice observed to date. Dong et al. [21] was the first to report variation in mt karyotype between two congeneric species, the rat lice, *Polyplax asiatica* and *Polyplax spinulosa*. These two *Polyplax* species both have 11 mt minichromosomes but differ in the distribution of eight of the 22 tRNA genes among these minichromosomes; the other tRNA genes and all of the protein-coding and rRNA genes are the same in distribution between the two *Polyplax* species [21]. Similarly, two other rodent lice in the genus *Hoplopleura* also share the same pattern for the distribution of all of the protein-coding and rRNA genes but differ in the location of four tRNA genes: *trn*L_1, *trnM*, *trnT* and *trnY* [20]. Song et al. [22] reported that the horse louse, *Haematopinus asini*, has nine minichromosomes, which is the same as the congeneric pig lice, *Haematopinus apri* and *Haematopinus suis* [19]. Six of the nine minichromosomes are identical among these *Haematopinus* species in gene content and gene arrangement. The other three minichromosomes of the horse louse are different from those of the pig lice in the locations of two protein-coding genes and two tRNA genes: *nad4L*, *nad6*, *trnM* and *trnR* [22]. Herd et al. [23] reported that the chimpanzee louse, *Pediculus schaeffi*,

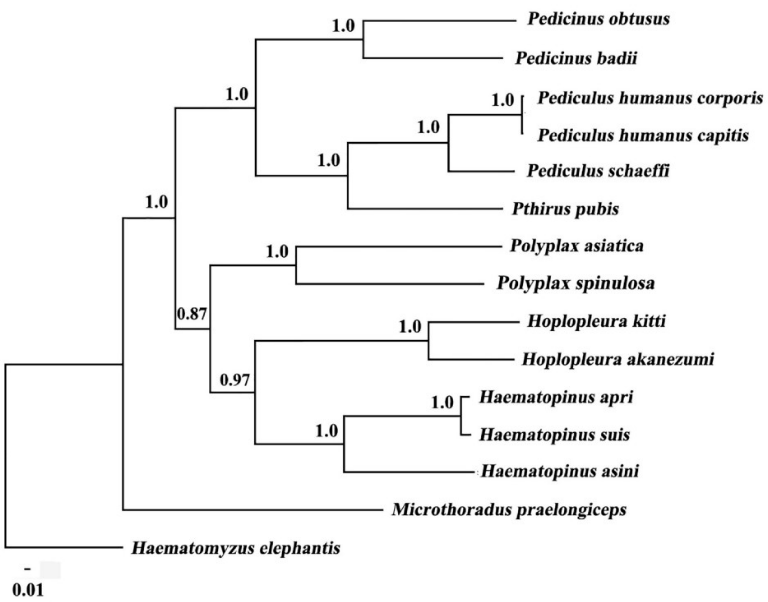


Fig. 6. Phylogenetic relationships among 14 species of sucking lice (Anoplura) inferred from Bayesian analysis of deduced amino acid sequences of nine mitochondrial proteins. The elephant louse, *Haematomyzus elephantis*, was used as the outgroup. Bayesian posterior probabilities (Bpp) values were indicated at nodes

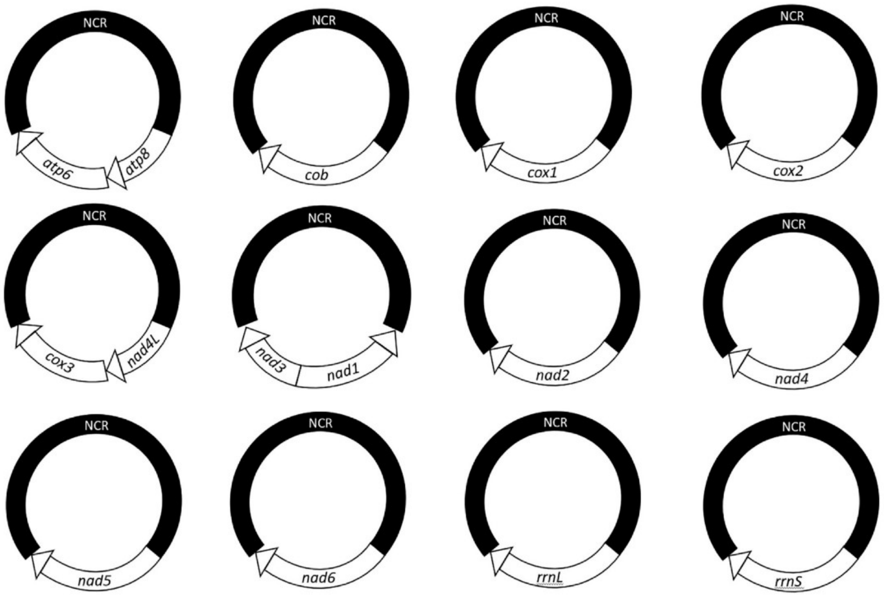


Fig. 7. Inferred ancestral mitochondrial karyotype of higher primate lice (Pedicinidae, Pediculidae and Pthiridae). Names and transcription orientation of genes ar indicated in the coding region; minichromosomes are in alphabetical order of protein-coding and rRNA gene names. See Fig. 3 legend for the abbreviation of gen names.

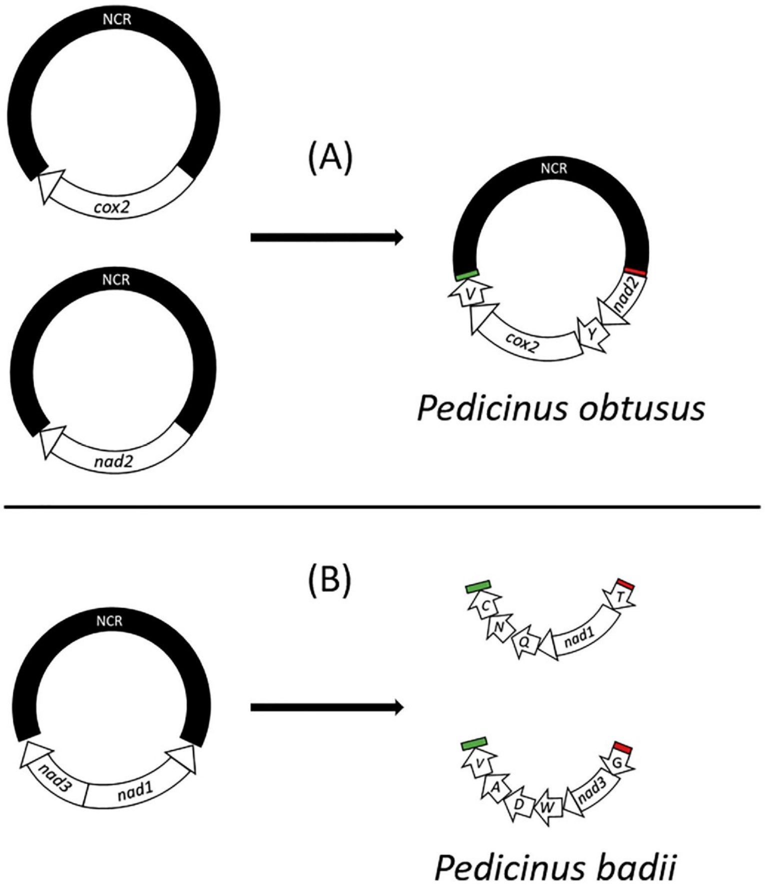


Fig. 8. (A) Two ancestral mitochondrial minichromosomes of higher primate lice (Fig. 7) merged in the macaque louse, *Pedicinus obtusus*; whereas (B) one ancestral mitochondrial minichromosome of higher primate lice split into two in the colobus louse, *Pedicinus badii*. Names and transcription orientation of genes are indicated in the coding region. Non-coding regions are in black. See Fig. 3 legend for the abbreviation of gene names.

has 18 minichromosomes, which is two less than the human lice, *Pediculus humanus corporis* (body louse) and *Pediculus humanus capitis* (head louse). Seventeen of the 18 minichromosomes of the chimpanzee louse match exactly their counterparts of the human lice in gene content and gene arrangement. The other minichromosome of the chimpanzee louse contains a protein-coding gene and four tRNA genes (*cob*, *trnS*$_1$, *trnN*, *trnE* and *trnM*); these five genes are on three minichromosomes in the human lice [18].

Previous studies have showed that fragmented mt genomes of sucking lice are dynamic in genome organization, which is in stark contrast to the highly stable single-chromosome mt genomes of most other animals [25]. The pronounced variation in mt karyotype between the macaque louse and the colobus louse is likely the result of a longer divergence between these two *Pedicinus* species than between the other congeneric species discussed above. Blood-sucking lice are the most host-specific ectoparasites and co-evolved with their mammalian hosts to a varying degree depending on the species [6–9]. In general, the mammalian divergence has been studied much more and understood much better than the divergence of their lice [38]. The host divergence time can serve as an estimate of the louse divergence time in the light of louse-host co-evolution. In this regard, the divergence time 17.6 million years (MY, median) between macaques (genus *Macaca*) and colobuses (genus *Colobus*) [38] can be used as an estimate of the divergence time between *Pedicinus obtusus* and *Pedicinus badii*, and therefore, the time frame for the observed variation between their mt karyotypes to occur. The divergence time and the extent of mt karyotype variation between these two *Pedicinus* species reconcile reasonably well with that: 1) in the genus *Pediculus* between the chimpanzee louse and the human lice (6.4 MY for the variation of a protein-coding gene and four tRNA genes); and 2) between *Polyplax asiatica* and *Polyplax spinulosa* (3.5 MY for *Bandicota-Rattus* divergence, variation of eight tRNA genes); but do not reconcile with that: 1) between *Hoplopleura akanezumi* and *Hoplopleura kitti* (15.9 MY for *Apodemus-Berylmys* divergence, variation of four tRNA genes); and 2) in the genus *Haematopinus* between the horse louse and the pig lice (81 MY for *Equus-Sus* divergence; variation of two protein-coding genes and two tRNA genes). The non-reconciliation with the two latter genera is likely due to host switch thus the louse divergence is likely much more recent than the host divergence. For the genus *Haematopinus*, the 81 MY *Equus-Sus* divergence is even older than the initial divergence event of the sucking lice dated back 77 million years ago

(MYA) [10], and thus cannot indicate at all the divergence time between the horse louse and the pig lice. The genus *Haematopinus* has 21 species, of which 19 species parasitize even-toed ungulates including pigs, cattle and deer, whereas the other two species parasitize odd-toed ungulates including horses and donkeys [6,7]. In the light of the louse-host records and based on the extent of mt karyotype variation, the divergence between the horse louse and the pig lice is likely ~10 MYA due to a host switch of *Haematopinus* from even-toed ungulates to odd-toed ungulates. In this regard, the divergence time between *Hoplopleura akanezumi* and *Hoplopleura kitti* is likely similar to that between *Polyplax asiatica* and *Polyplax spinulosa* (3.5 MY) and is much more recent than their host divergence (15.9 MY for *Apodemus-Berylmys*).

4.2. Mitochondrial karyotypes evolved in opposite directions between Pedicinus badii and Pedicinus obtusus

Shao et al. [26] showed that both splits and mergers of mt minichromosomes occurred in sucking lice and were responsible for their complex and dynamic mt genome organization. While many splits of mt minichromosomes were observed in the lineages leading to the lice of humans, chimpanzees, rodents and guanaco, mergers of minichromosomes were only observed in the lineage leading to the lice of pigs and horses of the genus *Haematopinus* [26]. Intriguingly, mt karyotypes appear to have evolved in opposite directions between the two *Pedicinus* species: two minichromosomes ancestral to higher primate lice merged as one in the lineage leading to the macaque louse, *Pedicinus obtusus*, whereas an ancestral minichromosome split into two in the lineage to the colobus louse, *Pedicinus badii*, after these two lineages diverged from their most recent common ancestor (Fig. 8). In the inferred ancestral mt karyotype of higher primate lice, *cox2* and *nad2* are on two separate minichromosomes (Fig. 7); this ancestral condition is retained in the colobus louse (Fig. 5A). In the macaque louse, *Pedicinus obtusus*, however, these two minichromosomes have merged to form the *nad2-trnY-cox2-trnV* minichromosome (Fig. 3, Fig. 8A). On the other hand, *nad1-nad3* minichromosome (note: *nad1* is opposite to *nad3* in transcription orientation), which is ancestral to higher primate lice, has split into two separate minichromosomes in the colobus louse, *Pedicinus badii*, with *nad1* inverted to have the same transcription orientation as *nad3* (Fig. 5A, Fig. 8B). The opposite evolutionary directions that these two *Pedicinus* species took explained well the striking variation between them in mt karyotype.

In conclusion, we determined the mt karyotypes of the macaque louse and the colobus louse of the genus *Pedicinus* and inferred the ancestral mt karyotype of the higher primate lice. The variation in mt karyotype between the two monkey lice is the most pronounced among the congeneric species of sucking lice observed to date. This is attributable to the opposite directions between the two *Pedicinus* species in their mt karyotype evolution. Two of the ancestral minichromosomes of the higher primate lice merged as one in the macaque louse whereas one of the ancestral minichromosomes split into two in the colobus louse after they diverged from their most recent common ancestor. Minichromosome merger in the macaque louse is the first observed outside the genus *Heamatopinus*, indicating that minichromosome merger is more common than previously known [26]. Our results showed that mt genome fragmentation was a two-way process in the higher primate lice: a fragmented mt genome could evolve to become more fragmented or reverse back to be less fragmented; neither merger nor split of minichromosomes appeared to be strongly selected in the monkey lice.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.09.005.

Funding

This work was supported by the Planned Program of Hunan Province Science and Technology Innovation (grant no. 2018RS3085) and the Training Program for Excellent Young Innovators of Changsha (Grant No. KQ1905013). Yalun Dong and Wei Wang are supported by scholarships from the Australian Biological Resources Study (ABRS, RF217–51) and University of the Sunshine Coast.

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 manuscript.

Acknowledgements

We thank Min Zhao for advice in downloading SRA data. We also want to thank Julie Allen, University of Illinois, for sharing the Illumina data of *Pedicinus badii* in NCBI GenBank SRA database.

References

[1] D.E. Wilson, D.M. Reeder, Mammal Species of the World: A Taxonomic and Geographic Reference, JHU Press, 2005.

[2] K.C. Beard, T. Qi, M.R. Dawson, B. Wang, C. Li, A diverse new primate fauna from middle Eocene fissure-fillings in southeastern China, Nature 368 (1994) 604–609.

[3] K.C. Beard, Y. Tong, M.R. Dawson, J. Wang, X. Huang, Earliest complete dentition of an anthropoid primate from the late middle Eocene of Shanxi Province, China, Science 272 (1996) 82–85.

[4] D.L. Gebo, M. Dagosto, K.C. Beard, T. Qi, J. Wang, The oldest known anthropoid postcranial fossils and the early evolution of higher primates, Nature 404 (2000) 276–278.

[5] L.A. Durden, G.G. Musser, The mammalian hosts of the sucking lice (Anoplura) of the world: a host-parasite list, Bull. Soc. Vector Ecol. 19 (1994) 130–168.

[6] L.A. Durden, G.G. Musser, The sucking lice (Insecta, Anoplura) of the world: a taxonomic checklist with records of mammalian hosts and geographical distributions, Bull. AMNH No 218 (1994).

[7] D.L. Reed, V.S. Smith, S.L. Hammond, A.R. Rogers, D.H. Clayton, Genetic analysis of lice supports direct contact between modern and archaic humans, PLoS Biol. 2 (2004).

[8] D.L. Reed, J.E. Light, J.M. Allen, J.J. Kirchman, Pair of lice lost or parasites regained: the evolutionary history of anthropoid primate lice, BMC Biol. 5 (2007) 7.

[9] J.E. Light, V.S. Smith, J.M. Allen, L.A. Durden, D.L. Reed, Evolutionary history of mammalian sucking lice (Phthiraptera: Anoplura), BMC Evol. Biol. 10 (2010) 292.

[10] M.S. Hafner, S.A. Nadler, Phylogenetic trees support the coevolution of parasites and their hosts, Nature 332 (1988) 258–259.

[11] R.D. Page, Temporal congruence revisited: comparison of mitochondrial DNA sequence divergence in cospeciating pocket gophers and their chewing lice, Syst. Biol. 45 (1996) 151–167.

[12] J.L. Boore, W.M. Brown, Big trees from little genomes: mitochondrial gene order as a phylogenetic tool, Curr. Opin. Genet. Dev. 8 (1998) 668–674.

[13] B. Pakendorf, M. Stoneking, Mitochondrial DNA and human evolution, Annu. Rev. Genom. Hum. G 6 (2005) 165–183.

[14] S.L. Cameron, Insect mitochondrial genomics: implications for evolution and phylogeny, Annu. Rev.Entomol. 59 (2014) 95–117.

[15] F. Song, H. Li, G.-H. Liu, W. Wang, P. James, D.D. Colwell, A. Tran, S.Y. Gong, W.Z. Cai, R. Shao, Mitochondrial genome fragmentation unites the parasitic lice of eutherian mammals, Syst. Biol. 68 (2019) 430–440.

[16] J.L. Boore, Animal mitochondrial genomes, Nucleic Acids Res. 27 (1999) 1767–1780.

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

[18] R. Shao, X.-Q. Zhu, S.C. Barker, K. Herd, Evolution of extensively fragmented mitochondrial genomes in the lice of humans, Genome Biol. Evol. 4 (2012) 1088–1101.

[19] H. Jiang, S.C. Barker, R. Shao, Substantial variation in the extent of mitochondrial genome fragmentation among blood-sucking lice of mammals, Genome Biol. Evol. 5 (2013) 1298–1308.

[20] W.-G. Dong, S. Song, X.-G. Guo, D.-C. Jin, Q. Yang, S.C. Barker, R. Shao, Fragmented mitochondrial genomes are present in both major clades of the blood-sucking lice (suborder Anoplura): evidence from two *Hoplopleura* rodent lice (family Hoplopleuridae), BMC Genomics 15 (2014) 751.

[21] W.-G. Dong, S. Song, D.-C. Jin, X.-G. Guo, R. Shao, 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 (2014) 44.

[22] S.D. Song, S.C. Barker, R. Shao, Variation in mitochondrial minichromosome composition between blood-sucking lice of the genus *Haematopinus* that infest horses and pigs, Parasite Vector 7 (2014) 144.

[23] K.E. Herd, S.C. Barker, R. Shao, The mitochondrial genome of the chimpanzee louse, *Pediculus schaeffi*: insights into the process of mitochondrial genome fragmentation in the blood-sucking lice of great apes, BMC Genomics 16 (2015) 661.

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

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

[26] A.D. Sweet, K.P. Johnson, S.L. Cameron, Mitochondrial genomes of *Columbicola* feather lice are highly fragmented, indicating repeated evolution of minicircle-type genomes in parasitic lice, PeerJ 8 (2020) e8759.

[27] G.F. Ferris, Contributions towards a Monograph of the Sucking Lice, Part VII, Stanford University Press, 1934.

[28] G.F. Ferris, The Sucking Lice, Mem. Pacif. Coast Ent. Soc. 1 (1951).

[29] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics 28 (2012) 1647–1649.

[30] K. Katoh, D.M. Standley, A simple method to control over-alignment in the MAFFT multiple sequence alignment program, Bioinformatics 32 (2016) 1933–1942.

[31] D. Laslett, B. Canbäck, ARWEN: a program to detect tRNA genes in metazoan mitochondrial, nucleotide sequences, Bioinformatics 24 (2008) 172–175.

[32] T.M. Lowe, P.P. Chan, tRNAscan-SE On-line: integrating search and context for analysis of transfer, RNA genes, Nucleic Acids Res. 44 (2016) W54–W57.

[33] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, T.L. Madden, NCBI BLAST: a, better web interface, Nucleic Acids Res. 36 (2008) W5–W9.

[34] G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously, aligned blocks from protein sequence alignments, Syst. Biol. 56 (2007) 564–577.

[35] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, Bioinformatics 19 (2003) 1572–1574.

[36] S. Guindon, J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihoodphylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (2010) 307–321.

[37] F. Abascal, R. Zardoya, D. Posada, ProtTest: selection of best-fifit models of protein evolution, Bioinformatics 21 (2005) 2104–2105.

[38] S. Kumar, G. Stecher, M. Suleski, S.B. Hedges, TimeTree: a resource for timelines, timetrees, and divergence times, Mol. Biol. Evol. 34 (2017) 1812–1819.