Highly Rearranged Mitochondrial Genome in *Falcolipeurus* lice (Phthiraptera: Philopteridae) from Endangered Eagles

Yu Nie  
Hunan Agricultural University

Yi-Tian Fu  
Hunan Agricultural University

Yu Zhang  
Hunan Agricultural University

Yuan-Ping Deng  
Hunan Agricultural University

Ya Tu  
Beijing Wildlife Rescue Center

Guo-Hua Liu (✉ liuguohua5202008@163.com)  
Hunan Agricultural University  [https://orcid.org/0000-0002-8434-899X](https://orcid.org/0000-0002-8434-899X)

Research Article

**Keywords:** Bird lice, Mitochondrial genome, Gene rearrangement, Phylogenetic analyses

**Posted Date:** February 24th, 2021

**DOI:** [https://doi.org/10.21203/rs.3.rs-249932/v1](https://doi.org/10.21203/rs.3.rs-249932/v1)

**License:** ☕️ ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](https://creativecommons.org/licenses/by/4.0/)

**Version of Record:** A version of this preprint was published at Parasites & Vectors on May 20th, 2021. See the published version at [https://doi.org/10.1186/s13071-021-04776-5](https://doi.org/10.1186/s13071-021-04776-5).
Abstract

Background: Fragmented mitochondrial (mt) genomes and extensive mt gene rearrangements have been frequently reported from parasitic lice (Insecta: Phthiraptera). However, relatively little is available about the mt genomes from the family Philopteridae that is the most species-rich family within the suborder Ischnocera.

Methods: Herein, we use next-generation sequencing to decode the mt genome sequences of *Falcolipeurus suturalis* and compared it with the mt genome sequences of *F. quadripustulatus*. Phylogenetic relationship of the concatenated amino acid sequence data for 13 protein-coding genes of the two *Falcolipeurus* lice and selected members of the family Philopteridae was evaluated using Bayesian inference (BI).

Results: The complete mt genome of *F. suturalis* is a circular double-stranded DNA molecule of 16,659 bp, and contains 13 protein-coding genes, 22 transfer RNA genes, two ribosomal RNA genes, as well as three putative non-coding regions. The gene order in *F. suturalis* mt genome was rearranged compared with that of *F. quadripustulatus*, and they were radically different from other louse species and the ancestral insect. Phylogenetic analyses revealed that the clear genetic distinctiveness between *F. suturalis* and *F. quadripustulatus* (Bayesian posterior probabilities=1.0), and the genus *Falcolipeurus* is more closely related to the genus *Ibidoecus* than to other genera (Bayesian posterior probabilities=1.0).

Conclusions: These novel datasets will help to better understand the gene rearrangements and phylogenetic position of *Falcolipeurus* and provide useful genetic markers for systematics and phylogenetic studies of bird lice.

Background

The typical insect mitochondrial (mt) genome is a circular double stranded DNA molecule of about 12–20 kb in length, which commonly contains 37 genes: 13 protein-coding genes (PCGs), 22 transfer RNAs (tRNA) and two ribosomal RNAs (rRNA) [1, 2]. However, in some lineages of parasitic lice (Insecta: Phthiraptera) are notable exceptions. For example, the following groups show extensively fragmented mt genomes, with the 37 genes separated onto multiple circular chromosomes: the families Haematopinidae [3], Hoplopleuridae [4, 5], Menoponidae [6], Pediculidae [7–9], and Polyplacidae [10], Trichodectidae [11]. Parasitic lice are currently divided into chewing lice (including three suborders Ischnocera, Amblycera and Rhynchophthirina) and sucking lice (the suborder Anoplura) by their mouthparts.

Chewing lice are permanent, obligate and host-specific ectoparasites commonly found on birds and mammals. The suborder Ischnocera (about 3,120 species) are currently divided into two families Philopteridae (about 2,600 species) and Trichodectidae. The Philopteridae is a large family within the suborder Ischnocera [12], but the complete mt genomes of only a limited number of species have been sequenced, including *Bothriometopus macrocnemis* [13], *Campanulotes bidentatus compar* [14], *Campanulotes compar* [11], *Coloceras* sp. SLC-2011 [15], *Falcolipeurus quadripustulatus* [11], *Ibidoecus*
**Methods**

**Sample collection and DNA extraction**

Adult samples of *F. suturalis* were taken from *Aquila rapax* from Beijing Wildlife Rescue Center, China. These parasitic lice were washed twice with sterile physiological saline solution (0.85%), and initial identification as *F. suturalis* based on morphological characteristics and host species (Fig. 1) [19], and then stored in 95% (v/v) ethanol at -40°C. The total genomic DNA was extracted from 60 individual bird lice (30 females and 30 males) using DNeasy Tissue Kit (Promega, Madison, USA) following the manufacturer’s instructions. The molecular identity of each bird louse was also confirmed by PCR using previously reported method [20] and then sequenced directly. The both *cox1* and *rrnS* sequences had 86.0% and 84.5% similarity to previously published sequences of *F. quadripustulatus* from China, respectively (GenBank accession nos. NC_039529.1), indicating that these samples of bird lice belong to the genus *Falcolipeurus*.

**Sequencing, assembling and verification**

The quality of the extracted genomic DNA was tested using by agarose-gel electrophoresis [21]. The genomic DNA concentration was quantified on the Quibt 2.0 Fluorometer (Thermo Scientific). Genomic DNA library (350 bp inserts) was prepared and sequenced in Novogene Bioinformatics Technology Co. Ltd. (Tianjin, China). The library was sequenced using Illumina HiSeq 2500 (250 bp pair-end reads). Raw reads were filtered and cleaned with Prinseq [22]. The obtained *cox1* and *rrnS* sequences of *F. suturalis* were used as the initial references to assemble the clean reads with Geneious 11.1.5 (minimum overlap identity = 99%, minimum overlap = 150 bp, maximal gap size = 5 bp) [23]. The size and circular
organization of the mt genome assembly were further verified by long PCR using four pairs of designed primers (Table S1; Fig. S1).

**Annotation and sequence analysis**

13 protein-coding genes and two rRNA (rrnL and rrnS) genes were identified by alignment with homologous genes of previously sequenced mt genome of the vulture louse *F. quadripustulatus* [11] using the MAFFT 7.122 software [24]. The 22 tRNA genes were verified using ARWEN [25] and the program tRNAscan-SE [26] with manual adjustment. Correctly annotated mt genomes were illustrated using the visualize module of MitoZ [27]. Nucleotide composition, amino acid sequences of each protein-coding genes and codon usage were mainly analyzed using MEGA 6.0 [28]. Asymmetry of base composition was calculated as the following formula: AT-skew = (A-T)/(A+T), GC-skew = (G-C)/(G+C) [29].

**Phylogenetic analysis**

Total of 10 mt genomes of the family Philopteridae were used for phylogenetic analysis, using one rat louse, *Hoplopleura* sp. (GenBank: MT792483-94) as an outgroup [5]. Each amino acid sequences were aligned individually using MAFFT algorithm. We concatenated the alignments of the individual genes and obtained a single dataset. The ambiguous areas of alignment were removed by Gblocks 0.91b with the options for a less stringent selection [30], and then subjected to phylogenetic analyses under Bayesian inference (BI). BI was conducted with four independent Markov chains run for 1,000,000 metropolis-coupled MCMC generations, sampling a tree every 100 generations in MrBayes 3.1.1 [31]. The initial 25% (2,500) trees of each MCMC were treated as the burn-in and the majority-rule consensus tree were used to calculate Bayesian posterior probabilities (BPP). Phylograms were drawn using FigTree v.1.31.

**Results And Discussion**

**Genome organization**

A total of 3.7 Gb data (about 20-fold coverage) was obtained from Illumina HiSeq 2500 platform which produced 15,178,382×2 raw reads. Extracted reads were cleaned and 9,630,532×2 clean reads were obtained for the assembly of the mt genome. The longest contig is 16,659 bp in size that represented the complete mt genome of *F. suturalis* (GenBank accession: MI456908). We identified and annotated all of the 37 mt genes typical of metazoan mt genomes (Fig. 2; Table 1). This mt genome contains 13 protein-coding genes, 22 tRNA genes, two rRNA genes and three non-coding (AT-rich) regions (Fig. 2; Table 1). The mt gene arrangement and distribution of genes are distinct from those of *F. quadripustulatus* [11] and *I. bisignatus* [15]. The overall nucleotide composition was: A = 27.8%, T = 44.8%, C = 11.1%, G = 16.3%. All mt genes were encoded on the heavy strand, which is similar to the other bird louse species [11, 13]. The three pairs of overlapping regions in the mt genome of *F. suturalis* were observed among *nad4L/nad1*, tRNA-His/tRNA-Asp and tRNA-Asp/tRNA-Arg. The overlapping regions ranged from −4 bp to
-8bp (Table 1). Besides, 22 intergenic regions were observed in this mt genome, ranging from 1 bp to 180 bp in size. The longest space was found between tRNA-S$_2$ and tRNA-G genes (Table 1).
Table 1
The organization of the mt genome of *F. suturalis.*

| Gene/Region   | Positions  | Size (bp) | Number of aa<sup>a</sup> | Ini/Ter codons<sup>b</sup> | Anticodon<sup>c</sup> | In  |
|---------------|------------|-----------|--------------------------|---------------------------|-----------------------|-----|
| cox1          | 34-1557    | 1524      | 507                      | ATA/TAA                   | +33                   |     |
| tRNA-Met (M)  | 1574–1637  | 64        |                          | CAT                       | +16                   |     |
| tRNA-Gln (Q)  | 1639–1705  | 67        |                          | TTG                       | +1                    |     |
| tRNA-Glu (E)  | 1706–1770  | 65        |                          | TTC                       | 0                     |     |
| *atp6*        | 1774–2445  | 672       | 223                      | ATA/TAA                   | +3                    |     |
| tRNA-Asn (N)  | 2451–2517  | 67        |                          | GTT                       | +5                    |     |
| *rrnS*        | 2518–3243  | 726       |                          |                           | 0                     |     |
| *rrnL*        | 3244–4318  | 1075      |                          |                           | 0                     |     |
| tRNA-Ala (A)  | 4319–4382  | 64        |                          | TGC                       | 0                     |     |
| *nad6*        | 4385–4858  | 474       | 157                      | ATG/TAA                   | +2                    |     |
| tRNA-Val (V)  | 4861–4922  | 62        |                          | TAC                       | +2                    |     |
| cox3          | 4977–5726  | 750       | 249                      | ATA/TAA                   | +54                   |     |
| tRNA-Lys (K)  | 5746–5808  | 63        |                          | TTT                       | +19                   |     |
| *nad4*        | 5843–7156  | 1314      | 437                      | ATT/TAG                   | +34                   |     |
| AT-loop region| 7157–7985  | 829       |                          |                           |                       |     |
| tRNA-LeuUUR (L<sub>2</sub>) | 7986–8047 | 62   |                         | TAA                       | 0                     |     |
| tRNA-Pro (P)  | 8064–8124  | 61        |                          | TGG                       | +16                   |     |
| *nad2*        | 8130–9101  | 972       | 323                      | ATG/TAA                   | +5                    |     |
| tRNA-Thr (T)  | 9171–9235  | 65        |                          | TGT                       | +69                   |     |

<sup>a</sup>The inferred length of amino acid (aa) sequence of 13 protein-coding genes; <sup>b</sup>Ini/Ter codons: initiation and termination codons; <sup>c</sup>In: Intergenic nucleotides.
| Gene/Region       | Positions       | Size (bp) | Number of aa\(^a\) | Ini/Ter codons\(^b\) | Anticodon\(^c\) | In |
|-------------------|-----------------|-----------|---------------------|----------------------|-----------------|----|
| tRNA-Tyr (Y)      | 9249–9313       | 65        |                     |                      | GTA + 13        |    |
| cox2              | 9314–9991       | 678       | 225                 | ATA/TAA              | 0               |    |
| AT-loop region    | 9992–10713      | 722       |                     |                      |                 |    |
| nad5              | 10714–123889    | 1676      | 558                 | ATG/TA               | 0               |    |
| tRNA-Phe (F)      | 12390–12456     | 67        |                     |                      | GAA 0           |    |
| tRNA-Cys (C)      | 12477–12543     | 67        |                     |                      | GCA + 20        |    |
| atp8              | 12565–12765     | 201       | 66                  | ATG/TAA              | + 21            |    |
| tRNA-SerUCN (S\(_2\)) | 12772–12840    | 69        |                     |                      | TGA + 6         |    |
| tRNA-Gly (G)      | 13021–13091     | 71        |                     |                      | TCC + 180       |    |
| AT-loop region    | 13092–13516     | 425       |                     |                      |                 |    |
| nad3              | 13517–13903     | 387       | 128                 | ATT/TAG              | 0               |    |
| tRNA-LeuCUN (L\(_1\)) | 13905–13966   | 62        |                     |                      | TAG + 1         |    |
| nad4L             | 13992–14264     | 273       | 90                  | ATT/TAA              | + 25            |    |
| nad1              | 14257–15163     | 907       | 302                 | ATG/T                | -8              |    |
| tRNA-SerAGN (S\(_1\)) | 15164–15231    | 68        |                     |                      | TCT 0           |    |
| cytb              | 15232–16323     | 1092      | 363                 | TTG/TAG              | 0               |    |
| tRNA-Trp (W)      | 16330–16396     | 67        |                     |                      | TCA + 6         |    |

\(^a\)The inferred length of amino acid (aa) sequence of 13 protein-coding genes; \(^b\)Ini/Ter codons: initiation and termination codons; \(^c\)In: Intergenic nucleotides.
| Gene/Region   | Positions          | Size (bp) | Number of aa<sup>a</sup> | Ini/Ter codons<sup>b</sup> | Anticodon<sup>c</sup> | ln   |
|--------------|--------------------|-----------|--------------------------|---------------------------|-----------------------|------|
| tRNA-His (H) | 16398–16460        | 63        |                          |                           | GTG                   | +1   |
| tRNA-Asp (D) | 16457–16524        | 68        |                          |                           | GTC                   | -4   |
| tRNA-Arg (R) | 16516–16585        | 70        |                          |                           | ACG                   | -8   |
| tRNA-Ile (I) | 16593–16659        | 67        |                          |                           | GAT                   | +6   |

<sup>a</sup>The inferred length of amino acid (aa) sequence of 13 protein-coding genes;<sup>b</sup>Ini/Ter codons: initiation and termination codons;
<sup>c</sup>In: Intergenic nucleotides.

The observed total A + T and G + C content of the complete mt genome were 73.0% and 27.0%, respectively, which were consistent with those of previous studies [11, 15] (Table 2). A negative AT skew (-23.3) and a positive GC skew (18.9) were calculated in this my genome (Table 2), which are common features of ectoparasites mt genome [11, 15]. All bird lice from Philopteridae reported to date and in the present study show strand asymmetry (GC skew between 6.3% and 38.1%) (Table 2).
Table 2
Nucleotide composition of the mt genomes of Philopteridae species, including that of *Falco*ipeurus *suturalis*.

| Species                      | Nucleotide frequency (%) | Whole genome sequence |
|------------------------------|--------------------------|-----------------------|
|                              | A  | T  | G  | C  | A+T% | AT skew | GC skew |
| *Bothriometopus macrocnemis* | 32.1 | 38.7 | 15.5 | 13.8 | 70.8  | -9.2    | 6.1     |
| *Campanulotes bidentatus*    | 26.5 | 43.7 | 20.67 | 9.77 | 70.1  | -24.5   | 38.1    |
| *Campanulotes compar*        | 26  | 44.5 | 20.4  | 9.1  | 70.5  | -26.3   | 38.1    |
| *Coloceras sp. SLC-2011*     | 27.5 | 42.9 | 19.9  | 9.6  | 70.4  | -21.8   | 35.1    |
| *Ibidoecus bisignatus*       | 35.5 | 40.6 | 13.2  | 10.8 | 76    | -6.7    | 10.2    |
| *Columbicola columbae*       | 39.1 | 29.2 | 16.3  | 15.4 | 68.2  | 14.6    | 2.8     |
| *Columbina picui*            | 33.5 | 31.6 | 18.3  | 16.6 | 65.1  | 2.9     | 5       |
| *Columbina cruziana*         | 32.9 | 31.4 | 19    | 16.7 | 64.3  | 2.4     | 6.3     |
| *Falco*ipeurus *quadripustulatus* | 26.3 | 45.5 | 16.9  | 11.3 | 71.8  | -26.8   | 20.1    |
| *Falco*ipeurus *suturalis*   | 28  | 45  | 16.4  | 11.2 | 73    | -23.3   | 18.9    |

Annotation

As the mt genomes of parasitic lice can contain non-standard initiation codons [1, 5, 13], the identification of initiation codons can sometimes be challenging. In this mt genome, all protein-coding genes had ATA or ATG or ATT or TTG as their initiation codon. 4 genes (*cox1*, *atp6*, *cox3* and *cox2*) start with ATA, 5 genes (*nad6*, *nad2*, *nad5*, *atp8* and *nad1*) start with ATG, 3 genes (*nad3*, *nad4L* and *nad4*) start with ATT and 1 gene (*cyt*bi*b*) use TTG (Table 1). All protein-coding genes had TAA or TAG or TA or T as their termination codon (Table 1). 8 genes (*cox1*, *atp6*, *nad6*, *cox3*, *nad2*, *cox2*, *atp8* and *nad4L*) stop with TAA, 3 genes (*nad4*, *nad3*, and *cyt*bi*b*) stop with TAG, *nad1* gene stop with TA and *nad5* gene use T (Table 1). Incomplete termination codons (TA or T) were identified in *nad1* and *nad5* genes, which is consistent with studies of some other bird lice, including *B. macrocnemis* (*nad1*), *F. quadripustulatus* (*nad5*, *nad6* and *nad1*). In *F. suturalis* mt genome, the *rrnL* genes was located between *rrnS* and tRNA-Ala genes, and *rrnS* genes was between tRNA-Asn and *rrnL* genes (Fig. 2; Table 1). The lengths of the *rrnS* and *rrnL* genes were 726 bp and 1075 bp, respectively. The 22 tRNA genes length varied from 61 to 71 bp (Table 1). All 22 tRNA genes can fold into cloverleaf structure (Fig. 3), which were consistent with those of previous studies [32, 33]. Non-coding region (NC1) (829 bp), located between *nad4* and tRNA-L2, has the highest A + T content of 75.4%. Non-coding region (NC2) (722 bp; A + T = 74.4%) located between *cox2* and *nad5* and Non-coding region (NC3) (425bp; A + T = 75.1%) located between tRNA-G and *nad3* (Table 1).

Comparative analyses between *F. suturalis* and *F. quadripustulatus*
The entire mt genome of *F. suturalis* is 537 bp longer than that of *F. quadripustulatus* [11]. A comparison of the nucleotide and the amino acid sequences of each protein-coding gene of the two *Falciparum* species is given in Table 3. Nucleotide sequence difference across the entire mt genome was 31.4%. The magnitude of nucleotide sequence variation in each gene between *F. suturalis* and *F. quadripustulatus* ranged from 13.2–27.5%. The greatest variation was observed in the *atp8* gene (27.5%), whereas least differences (13.2%) were found in the *cytb* genes (Table 3). For the *rrnL* and *rrnS* genes, sequence difference was 28.4% and 14.6% between *F. suturalis* and *F. quadripustulatus*, respectively (Table 3). Amino acid sequences inferred from individual mt protein genes of *F. suturalis* were compared with those of *F. quadripustulatus*. The amino acid sequence differences ranged from 4.5%-41.2%, with *cox1* being the most conserved protein, and *atp8* the least conserved (Table 3). This level of amino acid difference is very high. Previous studies of other lice have detected high level difference in protein sequences. For example, difference in amino acid sequences of the 13 protein-coding genes between *C. picui* and *C. cruziana* was 5.5–50% [16], and *C. bidentatus compar* and *C. compar* was 0-37.3% [11, 14]. Taken together, the molecular evidence presented here supports that *F. suturalis* and *F. quadripustulatus* represent distinct louse species.
Table 3
Nucleotide (nt) and/or predicted amino acid (aa) sequence differences in mitochondrial genes
among \textit{Falcolipeurus quadripustulatus} (FQ) and \textit{Falcolipeurus suturalis} (FS) upon pairwise comparison

| Gene/region | Nt sequence length | Nt difference (%) | Number of aa | Number of aa | Number of aa |
|-------------|--------------------|-------------------|--------------|--------------|--------------|
|             | FS     | FQ     | FS/FQ | FS     | FQ     | FS/FQ |
| cox1        | 1524   | 1554   | 15.3  | 507    | 517    | 4.5   |
| atp6        | 672    | 675    | 18.5  | 223    | 224    | 16.1  |
| rrnS        | 726    | 610    | 28.4  |         |         |       |
| rrnL        | 1075   | 1084   | 14.6  |         |         |       |
| nad6        | 474    | 478    | 22    | 157    | 159    | 25.8  |
| cox3        | 750    | 789    | 21.2  | 249    | 265    | 16.2  |
| nad4        | 1314   | 1305   | 24    | 437    | 434    | 27.2  |
| nad2        | 972    | 972    | 27    | 323    | 323    | 32.5  |
| cox2        | 678    | 675    | 14.4  | 225    | 224    | 7.5   |
| nad5        | 1676   | 1711   | 19.6  | 558    | 570    | 21    |
| atp8        | 201    | 204    | 27.5  | 66     | 67     | 41.2  |
| nad3        | 387    | 354    | 27.4  | 128    | 117    | 34.4  |
| nad4L       | 273    | 288    | 20.8  | 90     | 95     | 21.1  |
| nad1        | 907    | 848    | 20    | 302    | 282    | 12.6  |
| cyt b       | 1092   | 1092   | 13.2  | 363    | 363    | 9     |

Gene rearrangement

The mt genome arrangement of two \textit{Falcolipeurus} species substantially differs from those of other bird louse species within the family Philopteridae and from the inferred typical gene arrangement of ancestral insect mt genome (Fig. 4). Only two gene blocks are shared between \textit{B. macrocnemis} and the ancestral insect pattern: G-\textit{nad3} and \textit{atp8-atp6} [13], and one gene block is shared between \textit{Campanulotes} species and the ancestral insect: \textit{atp8-atp6} [11, 14, 34]. However, no derived mt gene arrangements are shared between the two \textit{Falcolipeurus} species. In addition, three gene blocks, V-\textit{cox3}, Y-\textit{cox2} and L_{1}-\textit{nad4L}, are shared by \textit{Falcolipeurus} and \textit{Ibidoecus} [11]. Such a lack of conserved gene arrangement in the mt genome of bird lice precludes the accurate reconstruction and identification of the rearrangement events and model [13].

Usually, the gene arrangement in the mt genome is very conserved within the same genus of ectoparasites [11, 14, 35]. Gene arrangement events between \textit{F. suturalis} and \textit{F. quadripustulatus} were
also analyzed (Fig. 5); at least one translocation could be inferred. The \textit{nad3} gene is located between \textit{cox2} and tRNA-Thr genes in \textit{F. quadripustulatus}, but was found between tRNA-Gly and tRNA-Leu\textsuperscript{CUN} in \textit{F. suturalis} (Fig. 5). The gene arrangement in the mt genomes of two \textit{Falcolipeurus} species indicated that the rate of change in the arrangement of mt genes may vary substantially among closely related groups of lice [36].

One tRNA gene (tRNA-Gly) was lacking and the duplication of three genes (tRNA-Thr, tRNA-Tyr and \textit{cox2}) was detected in the \textit{F. quadripustulatus} mt genome [11]. However, 37 genes have been identified in the \textit{F. suturalis} mt genome. Gene duplication have been also reported in mt genomes of several families of the class Insecta, such as \textit{Brontostoma colossus} [37], \textit{Phalantus geniculatus} [38] and \textit{Reduvius tenebrosus} [39]. In addition, tRNA loss was also found in the mt genome of several families of the class Insecta [11, 40], and this case can be explained by the tandemduplication-random loss (TDRL) model.

**Phylogenetic relationships**

Phylogenetic analysis showed the clear genetic distinctiveness between \textit{F. suturalis} and \textit{F. quadripustulatus} (Bayesian posterior probabilities = 1.0). The branch leading to the two \textit{Falcolipeurus} species is much longer than the branch of two \textit{Columbina} species (looking at branch lengths in tree). The genus \textit{Falcolipeurus} is more closely related to the genus \textit{Ibidoecus} than to other genera (Bayesian posterior probabilities = 1.0) (Fig. 6), which was consistent with that of a previous study [11].

Mt genome sequences are effective molecular markers to study the phylogenetic and systematic relationships at various taxonomic ranks across the phylum Arthropoda, including ectoparasites [41–46]. DNA sequencing provides the opportunity to further evaluate the phylogenetic relationships of the Philopteridae. For examples, Cruickshank et al. analyzed nuclear elongation factor-1 alpha (EF1\textalpha) sequences of 127 species from the four suborders and showed the Philopteridae to be paraphyletic [41]. Yoshizawa and Johnson 2003 analyzed mt 12S and 16S rDNA sequences of 18 species and showed the Philopteridae to be paraphyletic [42]. However, Johnson et al. analyzed 1107 single-copy orthologous genes of 46 species and showed that the Philopteridae to be monophyletic [43]. de Moya et al., analyzed 2,370 orthologous genes and showed that the Philopteridae to be monophyletic [44]. To date, the phylogenetic position of the Philopteridae in deep-level relationships within the order Phthiraptera could not be confidently determined. Although mt genomic data have been proven to be useful genetic marker to explore the phylogenetic relationships among several major lineages of parasitic lice [7, 9, 11], mt genome sequences of many lineages of the family Philopteridae are underrepresented or not represented. Therefore, more complete mt genomes of bird louse species representing this families that have not yet been decoded should be included in future analysis to resolve the phylogenetic position of the family Philopteridae within the order Phthiraptera.

**Conclusions**

The present study presents the entire mt genome sequences of \textit{F. suturalis} and compared it with the mt genome sequences of \textit{F. quadripustulatus}. Gene order is rearranged and represents a new pattern within
the order Phthiraptera. These novel datasets will help to better understand the gene rearrangements and phylogenetic position of *Falcolipeurus* and provide useful genetic markers for systematics and phylogenetic studies of bird lice.

**Abbreviations**

mt: mitochondrial; rDNA: ribosomal DNA; BI: Bayesian inference; *nad2*: NADH dehydrogenase subunit 2; *cox1*: cytochrome c oxidase subunit 1; *cox2*: cytochrome c oxidase subunit 2; *ATP6*: ATP synthase F0 subunit 6; *cox3*: cytochrome c oxidase subunit 3; *nad3*: NADH dehydrogenase subunit 3; *nad5*: NADH dehydrogenase subunit 5; *nad4*: NADH dehydrogenase subunit 4; *nad4L*: NADH dehydrogenase subunit 4L; *nad6*: NADH dehydrogenase subunit 6; *cytb*: cytochrome b; *atp8*: ATP synthase F0 subunit 8; *nad1*: NADH dehydrogenase subunit 1; *tRNA*: transfer RNA; *rrnL*: large subunit of rRNA; *rrnS*: small subunit of rRNA

**Declarations**

**Acknowledgements**

Not applicable.

**Authors’ contributions**

GHL and YT conceived and designed the study, and critically revised the manuscript. YN performed the experiments. YN YTF and GHL analyzed the data. YN and YTF drafted the manuscript. YZ YPD helped in study design, study implementation, and manuscript preparation. All authors read and approved the final manuscript.

**Funding**

This study was supported by the Planned Programme of Hunan Province Science and Technology Innovation (Grant no. 2018RS3085) and the Training Programme for Excellent Young Innovators of Changsha (Grant No. KH2002001).

**Availability of data and materials**

The complete mitochondrial genome sequences of *Falcolipeurus suturalis* have been deposited in the GenBank database under the accession number MI456908.

**Ethics approval and consent to participate**

All procedures involving animals in the present study were approved and this study was approved by the Animal Ethics Committee of Hunan Agricultural University (No. 43321503).

**Consent for publication**
Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

1Hunan Provincial Key Laboratory of Protein Engineering in Animal Vaccines, College of Veterinary Medicine, Hunan Agricultural University, Changsha, Hunan Province 410128, China. 2Beijing Wildlife Rescue Center, Beijing 101300, China.

References

1. Wolstenholme DR. Animal mitochondrial DNA: structure and evolution. Int Rev Cytol. 1992;141:173–216.
2. Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999;27:1767–80.
3. Song SD, Barker SC, Shao R. Variation in mitochondrial minichromosome composition between blood-sucking lice of the genus Haematopinus that infest horses and pigs. Parasit Vectors. 2014;7:144.
4. Dong WG, Song S, Guo XG, Jin DC, Yang Q, Barker SC, Shao R. 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 Genom. 2014;15:751.
5. Fu YT, Nie Y, Duan DY, Liu GH. Variation of mitochondrial minichromosome composition in Hoplopleura lice (Phthiraptera: Hoplopleuridae) from rats. Parasit Vectors. 2020;13:506.
6. Sweet AD, Johnson KP, Cao Y, de Moya RS, Skinner RK, Tan M, Virrueta Herrera S, Cameron SL. Structure, gene order, and nucleotide composition of mitochondrial genomes in parasitic lice from Amblycera. Gene. 2020;18:145312.
7. Shao R, Kirkness EF, Barker SC. The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, Pediculus humanus. Genome Res. 2009;19:904–12.
8. Herd KE, Barker SC, Shao R. 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 Genom. 2015;16:661.
9. Fu YT, Dong Y, Wang W, Nie Y, Liu GH, Shao R. Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse Pedicinus obtusus and colobus louse Pedicinus badii. Genomics. 2020;112:4924–33.
10. Dong WG, Song S, Jin DC, Guo XG, Shao R. 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 Genom. 2014;15:44.

11. Song F, Li H, Liu GH, Wang W, James P, Colwell DD, Tran A, Gong S, Cai W, Shao R. Mitochondrial genome fragmentation unites the parasitic lice of eutherian mammals. Syst Biol. 2019;68:430–40.

12. Roskov Y, Ower G, Orrell T, Nicolson D, Bailly N, Kirk PM, Bourgoin T, DeWalt RE, Decock W, van Nieukerken EJ, Penev L, editors. Species 2000 & ITIS Catalogue of Life, 2020-12-01.

13. Cameron SL, Johnson KP, Whiting MF. The mt genome of the screamer louse Bothriometopus (phthiraptera: ischnocera): effects of extensive gene rearrangements on the evolution of the genome. J Mol Evol. 2007;65:589–604.

14. Covacin C, Shao R, Cameron S, Barker SC. Extraordinary number of gene rearrangements in the mt genomes of lice (Phthiraptera: Insecta). Insect Mol Biol. 2006;15:63–8.

15. Cameron SL, Yoshizawa K, Mizukoshi A, Whiting MF, Johnson KP. Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). BMC Genom. 2011;12:394.

16. Sweet AD, Johnson KP, Cameron SL. Mitochondrial genomes of Columbicola feather lice are highly fragmented, indicating repeated evolution of minicircle-type genomes in parasitic lice. PeerJ. 2020;8:e8759.

17. Kirchgatter K, de Oliveira Guimarães L, Hugo Yañez Trujillano H, Rafael Arias F, Cáceres AG, de Castro Duarte AMR, Dos Santos Malafronte R, Tubaki RM, Mureb Sallum MA. Phylogeny of Anopheles (Kerteszia) (Diptera: Culicidae) using mitochondrial genes. Insects. 2020;11:324.

18. Kelava S, Mans BJ, Shao R, Moustafa MAM, Matsuno K, Takano A, Kawabata H, Sato K, Fujita H, Ze C, Plantard O, Hornok S, Gao S, Barker D, Barker SC, Nakao R. Phylogenies from mitochondrial genomes of 120 species of ticks: Insights into the evolution of the families of ticks and of the genus Amblyomma. Ticks Tick Borne Dis. 2021;12:101577.

19. Tandan BK. Mallophaga from birds of the Indian subregion. Part VI Falcoleipeurus Bedford. Proceedings of the Royal Entomological Society of London. Series B, Taxonomy, 2009.

20. Shao R, Li H, Barker SC, Song S. The Mitochondrial genome of the guanaco louse, Microthoracius praelongiceps: insights into the ancestral mitochondrial karyotype of sucking Lice (Anoplura, Insecta). Genome Biol Evol. 2017;9:431–45.

21. Almal S, Jeon S, Agarwal M, Patel S, Patel S, Bhak Y, et al. Sequencing and analysis of the whole genome of Indian Gujarati male. Genomics. 2019;111:196–204.

22. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863–4.

23. Kearse M, Moir R, Wilson A, Stones–Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.

24. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
25. Laslett D, Canbäck B. ARWEN: a program to detect tRNA genes in meta-zoan mitochondrial nucleotide sequences. Bioinformatics. 2008;24:172–5.

26. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016;44:W54-7.

27. Meng G, Li Y, Yang C, Liu S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 2019;47:e63.

28. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30:2725–9.

29. Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol. 1995;41:353–8.

30. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56:564–77.

31. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19:1572–4.

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

33. Shao R, Zhu XQ, Barker SC, Herd K. Evolution of extensively fragmented mitochondrial genomes in the lice of humans. Genome Biol Evol. 2012;4:1088–101.

34. Shao R, Campbell NJ, Barker SC. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, Heterodoxus macropus (Phthiraptera). Mol Biol Evol. 2001;18:858–65.

35. Kelava S, Mans BJ, Shao R, Moustafa MAM, Matsuno K, Takano A, Kawabata H, Sato K, Fujita H, Ze C, Plantard O, Hornok S, Gao S, Barker D, Barker SC, Nakao R. Phylogenies from mitochondrial genomes of 120 species of ticks: Insights into the evolution of the families of ticks and of the genus Amblyomma.Ticks Tick Borne Dis. 2021;12:101577.

36. Shao R, Campbell NJ, Schmidt ER, Barker SC. Increased rate of gene rearrangement in the mitochondrial genomes of three orders of hemipteroid insects. Mol Biol Evol. 2001;18:1828–32.

37. Kocher A, Kamilari M, Lhuillier E, Coissac E, Péneau J, Chave J, Murienne J. Shotgun assembly of the assassin bug Brontostoma colossus mitochondrial genome (Heteroptera, Reduviidae). Gene. 2014;552:184–94.

38. Sun Z, Liu Y, Wilson JJ, Chen Z, Song F, Cai W, Li H. Mitochondrial genome of Phalantus geniculatus (Hemiptera: Reduviidae): tmT duplication and phylogenetic implications. Int J Biol Macromol. 2019;129:110–15.

39. Jiang P, Li H, Song F, Cai Y, Wang J, Liu J, Cai W. Duplication and remolding of tRNA genes in the mitochondrial genome of Reduvius tenebrosus (Hemiptera: Reduviidae). Int J Mol Sci. 2016;17:951.

40. Tyagi K, Chakraborty R, Cameron SL, Sweet AD, Chandra K, Kumar V. Rearrangement and evolution of mitochondrial genomes in Thysanoptera (Insecta). Sci Rep. 2020;10:695.
41. Cruickshank RH, Johnson KP, Smith VS, Adams RJ, Clayton DH, Page RD. Phylogenetic analysis of partial sequences of elongation factor 1 alpha identifies major groups of lice (Insecta: Phthiraptera). Mol Phylogenet Evol. 2001;19:202–15.

42. Yoshizawa K, Johnson KP. Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. Mol Phylogenet Evol. 2003;29:102–14.

43. Johnson KP, Dietrich CH, Friedrich F, Beutel RG, Wipfler B, Peters RS, Allen JM, Petersen M, Donath A, Walden KKO, Kozlov AM, Podsiadlowski L, Mayer C, Meusemann K, Vasilikopoulos A, Waterhouse RM, Cameron SL, Weirauch C, Swanson DR, Percy DM, Hardy NB, Terry I, Liu S, Zhou X, Misof B, Robertson HM, Yoshizawa K. Phylogenomics and the evolution of hemipteroid insects. Proc Natl Acad Sci U S A. 2018;115:12775–80.

44. de Moya RS, Yoshizawa K, Walden KKO, Sweet AD, Dietrich CH, Johnson KP. Phylogenomics of parasitic and non-parasitic lice (Insecta: Psocodea): combining sequence data and exploring compositional bias solutions in Next Generation Datasets. Syst Biol. 2020: syaa075.

45. Liu GH, Chen F, Chen YZ, Song HQ, Lin RQ, Zhou DH, Zhu XQ. Complete mitochondrial genome sequence data provides genetic evidence that the brown dog tick Rhipicephalus sanguineus (Acari: Ixodidae) represents a species complex. Int J Biol Sci. 2013;9:361–9.

46. Gu XB, Liu GH, Song HQ, Liu TY, Yang GY, Zhu XQ. The complete mitochondrial genome of the scab mite Psoroptes cuniculi (Arthropoda: Arachnida) provides insights into Acari phylogeny. Parasit Vectors. 2014;7:340.

Figures
Figure 1

The back and abdomen of Falcolipeurus suturalis.
Figure 2

The mt genome of Falcolipeurus suturalis. All genes are on the same DNA strand and are transcribed clockwise. Protein-coding and rRNA genes are indicated with the standard nomenclature. tRNA genes are indicated with the one-letter code of their corresponding amino acids. There are two tRNA genes for leucine: L1 for codons CUN and L2 for UUR; and two tRNA genes for serine: S1 for codons AGN and S2 for UCN. “NCR1” refers to the first non-coding region. “NCR2” refers to the second non-coding region. “NCR3” refers to the third non-coding region.
Figure 3

22 tRNA secondary structures from Falcolipeurus suturalis.
Figure 4

Gene rearrangements of mitochondrial genomes of bird lice within the family Philopteridae.
Figure 5
Gene rearrangement in two Falcolipeurus species.

Figure 6
Phylogenetic relationships among 7 species of the family Philopteridae inferred by Bayesian inference from deduced amino acid sequences of 13 protein-coding genes. One rat louse, Hoplopleura sp. as an outgroup. Bootstrap values and Bayesian posterior probabilities (Bpp) were indicated at nodes.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- Fig.S1.tif
- TableS1.docx
- GraphicalAbstract.tif