Comparative analyses of the fragmented mitochondrial genomes of wild pig louse *Haematopinus apri* from China and Japan

Yu Nie, Yi-Tian Fu, Wei Wang, Rong Li, Wan-Qing Tang, Guo-Hua Liu

*Research Center for Parasites & Vectors, College of Veterinary Medicine, Hunan Agricultural University, Changsha, Hunan, 410128, China*

*The Centre for Bioinnovation, School of Science and Engineering, University of the Sunshine Coast, Sippy Downs, QLD, 4556, Australia*

**Abstract**

The wild pig louse *Haematopinus apri* is one of the commonest ectoparasites of wild pigs. In the present study, the entire mitochondrial (mt) genome of wild pig louse *H. apri* from China was sequenced and compared with previously characterized wild pig louse *H. apri* from Japan. We identified all of the 37 mt genes in the wild pig louse *H. apri* from China which are on nine circular minichromosomes. Each mt minichromosome is 2.9 kb size and contains 2-8 genes and one non-coding region (1543 bp-2534 bp). The number of minichromosomes, gene content and gene order in the both mt genomes of wild louse *H. apri* from China and Japan are the same. The identity of the both mt genomes (except for non-coding regions) was 98.3% between wild pig louse *H. apri* from China and Japan. The entire mt genome sequence (except for non-coding regions) of wild pig louse *H. apri* from China is longer (3 bp) than that from Japan. For the 13 protein-coding genes, this comparison showed sequence differences in each gene at both the nucleotide (0.8%-2.4%) and amino acid (0.4%-3.5%) levels. The most conserved of these genes was the nad6, whereas the nad2 was least conserved at the nucleotide levels. This is the first comprehensive comparison of the mt genomes of a louse species from different geographic locations. This useful data provides additional genetic markers to study the phylogeny, systematics and population genetics of wild pig louse *H. apri*.

1. Background

The suborder *Haematopinidae* belongs to the family Haematopinidae of the suborder Anoplura, known as the blood-sucking lice (Barker, 1994; Durden and Musser, 1994; Melenev and Kim, 1974). This genus consists of 21 species, which mainly parasitize even-toed ungulates (Durden and Musser, 1994; Scofield et al., 2012). *Haematopinus* species are vectors of disease-causing microorganisms, such as African swine fever virus (Saegerman et al., 2021), swinepox virus (Thibault et al., 1998), classical swine fever virus (CSFV) (Wall R and Shearer D, 2008) and Ana-plasma spp. (Da Silva et al., 2013). Meanwhile, wild pig louse *H. apri* is one of the most common ectoparasites of wild pigs.

Metazoan mitochondrial (mt) genomes are usually circular DNA organization (13-20 kb) that contain 36 to 37 genes: 12-13 protein-coding genes, two ribosomal RNA genes, and 22 transfer RNA genes (Boore, 1999; Lavrov, 2007; Wolstenholme, 1992). However, fragmented mt genomes have been reported from all sequenced blood-sucking lice (Dong et al., 2014a, b; Herb et al., 2015; Fu et al., 2020a, b; Jiang et al., 2013; Shao et al., 2012, 2015, 2017; Song et al., 2014) and some chewing lice (Song et al., 2019; Sweet et al., 2020, 2021). Although the suborder Anoplura contains 540 species of blood-sucking lice (Durden and Musser, 1994; Kim and Ludwig, 1978), to date, only mt genomes of 15 blood-sucking louse species have been sequenced and deposited in GenBank. In the genus *Haematopinus*, the entire fragmented mt genome was available for only three species (Jiang et al., 2013; Song et al., 2014). A recent study revealed highly divergent mt genomes in the booklouse (*Liposcelis bostrychophila*) from different geographical locations (Feng et al., 2018). This finding raised the possibility that parasitic lice from different geographical origins may have geographic differences in the number of minichromosomes, gene order, and nucleotide composition in their mt genomes. However, there is still no comprehensive comparison of the mt genomes of a louse species from different geographic locations. The mt genome sequences of wild pig louse *H. apri* from wild pig in Japan has been sequenced, but its non-coding regions have not been fully identified (Jiang et al., 2013). In addition, molecular data on wild pig louse *H. apri* from other countries and regions is totally lacking.

The objectives of the present study were: (i) to determine the number
of minichromosomes, gene order and nucleotide composition of the entire mt genome of wild pig louse *H. apri* from China, (ii) to compare this fragmented mt genome with that of wild pig louse *H. apri* from Japan, (iii) to test the hypothesis that wild pig louse *H. apri* from different geographic locations have geographic differences in the number of minichromosomes, gene order and nucleotide composition in their mt genomes.

2. Methods

2.1. Sample collection and DNA extraction

Adult specimens of wild pig louse *H. apri* were collected from a wild pig in Hunan province, China. These wild pig lice were washed five times in physiological saline solution, identified preliminarily to wild pig louse *H. apri* from China based on morphology and host information (Fig. 1) (Kim and Ludwig, 1978), and stored in 70% (v/v) ethanol at −40 ° C. The total genomic DNA of wild pig louse *H. apri* from China was extracted from individual wild pig louse using a commercially available kit (Promega, Madison, USA) following the manufacturer’s procedure. The identity of each wild pig louse specimen was further confirmed by PCR with primer reported previously (Fu et al., 2020a). The mt cox1 and rrsS gene sequences of the *H. apri* from China had 98.5% and 98.6% similarity with that of wild pig louse *H. apri* from wild pig in Japan (GenBank accession nos. KC814616 and KC814619), respectively.

2.2. Sequencing, assembling and annotation

The DNA concentration was evaluated using the Qubit system (Thermo Fisher Scientific, Waltham, MA, USA). Then, the entire genomic DNA was sequenced on Illumina Miseq Platform. The quality of raw data (300 bp each, paired-end reads) was assessed using FastQC 0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and then were filtered for quality and trimmed using Skewer v0.2.2 (Jiang et al., 2013). Finally, 2 GB clean data was assembled using Geneious 11.1.5 (Kearse et al., 2012) based on obtained mt cox1 and rrsS gene sequences. We assembled all minichromosomes individually in full length as described previously (Fu et al., 2020a). The gene borders were predicted with MITOS web servers (http://mitos.bioinf.uni-leipzig.de/index.py) and manually curated. The location of protein-coding genes was further confirmed with BLAST searches of the NCBI database. tRNA genes were further confirmed using ARWEN (Laslett and Canback, 2008) and the program tRNAscan-SE (Lowe and Chan, 2016). The RNA genes (rrnL and rrnS) were further confirmed by the boundary of the adjacent tRNA genes. The entire fragmented mt genome maps were produced using Microsoft Powerpoint.

2.3. Verification of mt minichromosomes

The size and the circular organization of each mt minichromosome of wild pig louse *H. apri* from China were further verified by PCR using designed primers from the coding regions of each mt minichromosome (Table S1; Fig. S1). The two primers in each pair were next to each other with a small gap in between (10–90 bp). To obtain full-length sequences of each minichromosomes, the amplicons from nine full-length mt minichromosomes of wild pig louse *H. apri* from China were sequenced in a separate batch with high throughput sequencing as described above. The clean data of each minichromosomes was obtained and assembled as described above, respectively.

3. Results and discussion

3.1. Genome organization

By employing the sequencing from 6,099,122 × 2 clean reads, we successfully assembled the entire mt genome of wild pig louse *H. apri* from China. We annotated 37 mt genes, including 13 protein-coding genes, two ribosomal RNA genes, and 22 transfer RNA genes (Fig. 2; Table 1). The nucleotide sequences of the mt minichromosomes of wild pig louse *H. apri* from China were deposited in the GenBank database under the accession numbers ON000914-ON000922. These genes were on nine circular minichromosomes, which was consist with those of previous studies (Jiang et al., 2013; Song et al., 2014). The mt minichromosomes of wild pig louse *H. apri* from China is 2.9 kb–4.2 kb in length, and each minichromosome has a coding region and a non-coding region (NCR) (Table 1). There are 2–8 genes in each coding region, varying in length from 787 bp to 2669 bp (Table 1). Each non-coding region varies in length from 1543 bp (nad2-trnL-cox1-trnL2 mini-chromosome) to 2534 bp (trnR-nad4L-nad6-trnM minichromosome) (Table 1). With the exception of nad1 gene, all mt genes are transcribed in the same direction.

3.2. Annotation

The wild pig louse *H. apri* from China mt genome contains 13 protein-coding genes, which had five initiation codons (ATA, ATG, TTG ATT, and GTG). Among them, four protein-coding genes use ATA (cox3, nad5, nad6, atp6), three genes use ATG (nad1, nad2, nad4L), three genes use TTG (cox1, nad3, atp8), two genes use ATT (cox2, cytB) and one gene use GTG (nad4). This mt genome has three termination codons (TAA,
Table 1
Mitochondrial minichromosomes of the wild pig louse Haematopinus apri from China, identified by Illumina sequencing.

| Minichromosome | Size (bp) | Size of coding region (bp) | Size of non-coding region (bp) | Intergenic region (bp) |
|----------------|-----------|----------------------------|-------------------------------|------------------------|
| nad2-trnI-cox1-trnL_4 | 4212      | 2669                       | 1543                          | 0                      |
| trnD-trnI’-cox2-trnR_5- 
  trnS2-trnR-cox3-trnA | 3832      | 1840                       | 1980                          | 12                     |
| trnK-cyb-trnV | 2780      | 1217                       | 1561                          | 2                      |
| trnQ-nad2-trnL’-trnG- 
  nad3-trnW | 3603      | 1509                       | 2087                          | 7                      |
| trnK-nad4-ap58-ap6-trnN | 3873      | 2297                       | 1568                          | 8                      |
| trnI-nad4-trn6-trnM | 3463      | 872                        | 2534                          | 57                     |
| trn1-nad5-trnF | 3346      | 1777                       | 1563                          | 6                      |
| rnrS-trnC | 3107      | 787                        | 2319                          | 1                      |
| rnrL-trnL | 3378      | 1226                       | 2152                          | 0                      |
| Total | 31594     | 14194                      | 17307                         | 93                     |

TAG and T). Among them, five genes use T (cox1, cox2, cox3, nad2, nad3), five genes use TAA (cyt, nad4, nad5, nad6), three genes use TAG (nad4L, ap56, ap68). The use of the start codon and stop codon of these protein-coding genes in wild pig louse H. apri from China is consistent with that in wild pig louse H. apri from Japan, except for nad4 gene (use ATG for Japan). Incomplete termination codons (TA and T) are also present in other sucking lice, including horse louse H. asiini (Song et al., 2014), rat louse Hoplopleura kiti (Dong et al., 2014b), rat louse Polypilax asiatica (Dong et al., 2014a) and P. spirulosa (Dong et al., 2014a), chimpanzee louse Pediculus schaeffi (Herd et al., 2015), guanaco louse Microthoracius praelongiceps (Shao et al., 2017) and pubic louse Phthirus pubis (Shao et al., 2012). In the mt genome of wild pig louse H. apri from China, the sizes of the rrl and rrnS genes were 718 bp and 1160 bp, respectively. Sequence identity in the rrl and rrnS genes is 98.2% and 99.2% between wild pig louse H. apri from China and Japan, respectively. The 22 tRNA genes identified in the mt genome of wild pig louse H. apri from China ranged from 63 to 75 bp in length. The secondary structure predictions in H. apri from China (not shown) were similar to those of horse louse H. apri from Japan and domestic pig louse H. suis (Jiang et al., 2013).

3.3. Non-coding regions

We obtained the entire non-coding region sequences of all of the nine mt minichromosomes of the wild pig louse H. apri from China, which range from 1543 bp (nad2-trnI-cox1-trnL_4 minichromosome) to 2534 bp (rrnR-nad4L-nad6-trnM minichromosome) (Table 1). The non-coding regions contain tandem repetitive sequences (137 bp) and copy number (2–3 repeats). The non-coding regions of H. apri from China have 52.4%–96.7% pairwise identity to each other. There are sequence motifs in the non-coding regions that are highly conserved in many blood-sucking lice, such as H. suis, H. apri (Jiang et al., 2013), Pedicinus obtusus, P. badii (Fu et al., 2020b) and M. praelongiceps (Shao et al., 2017). As in previous studies (Shao et al., 2012; Jiang et al., 2013), a GC-rich motif (66 bp, 62.1% G and C) is downstream of the 3′-end of the coding region and an AT-rich motif (121 bp, 73.6% A and T) is upstream of the 5′-end. Interestingly, this study found another AT-rich motif (72 bp, 79.2% A and T) upstream of the GC-rich motif.

3.4. Comparative mt genomic analyses of H. apri from China and Japan

A comparison of the nucleotide and the amino acid sequences of mt genes for H. apri from China and Japan are given in Table 2. The nucleotide sequence variation across the entire mt genome (except for non-coding regions) between the wild pig louse H. apri from China and Japan
Japan was 1.7%. Difference across both nucleotide and amino acid sequence differences of the 13 protein-coding genes between the wild pig louse *H. apri* from China and Japan was 1.7% and 1.4%, respectively. The magnitude of nucleotide sequence difference in each gene between the wild pig louse *H. apri* from China and Japan ranged from 0.8% to 2.4%. The greatest difference between the wild pig louse *H. apri* from China and Japan was nad2 gene, whereas the least difference was detected in the *rrn* gene (Table 2). Amino acid sequences inferred from individual mt protein-coding genes of wild pig louse *H. apri* from China and Japan were also compared. The amino acid sequence differences ranged from 0.4% to 3.5%, with cox3 gene being the most conserved protein-coding gene and nad3 gene the least conserved. Previous studies of other lice have also detected low intra-specific sequence variation in mt sequences. For example, the intra-specific sequence variation in *P. pubis* was 0–1.5% for *pcox1* (775 bp) (Amanzougaghene et al., 2020) and 1.0% for *pcbyr* (696 bp) (Light and Reid, 2009; Shao et al., 2012). The intra-specific sequence variation in *P. badii* was less than 0.3% for *pcox1* (346 bp) (Light et al., 2010). Taken together, the molecular evidence presented here indicated the wild pig louse *H. apri* from China and Japan represent the same species.

4. Conclusions

The present study sequenced and compared the entire fragmented mt genomes of *H. apri* from China and Japan, and demonstrated that wild pig louse *H. apri* from different geographic locations is highly conserved in mt genomic composition and structure, refuted our hypothesis. This is the first comprehensive comparison of the mt genomes of a louse species from different geographic locations. The useful data provide useful genetic markers for studying the population genetics, molecular systematics and phylogenetics of blood-sucking lice.

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).

Authors’ contributions

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

Funding

This study was supported in part by the National Natural Science Foundation of China (Grant No. 32172884) and the Training Programme for Excellent Young Innovators of Changsha (Grant No. KQ2106044).

Consent for publication

Not applicable.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijppaw.2022.03.013.

References

Amanzougaghene, N., Medianikov, O., Ly, T.D.A., Gautret, P., Davoust, B., Fenollar, F., Iri, A., 2020. Molecular investigation and genetic diversity of Pediculus and Phthirus lice in France. Parasites Vectors 13, 177.
Barker, S.C., 1994. Phylogeny and classification, origins, and evolution of host associations of lice. Int. J. Parasitol. 24, 1285–1291.
Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780.
Da Silva, A.S., Lopes, L.S., Diaz, J.D.S., Tonin, A.A., Stefani, L.M., Araújo, D.N., 2013. Louse outbreak in buffaloes: evidence of Anaplasm marginale transmission by sucking lice Haematopinus equestris. J. Parasitol. 99, 546–547.
Dong, W.G., Song, S., Jin, D.C., Guo, X.G., Shao, R., 2014a. Fragmented mitochondrial genomes of the rat louse, 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. 15, 44.
Dong, W.G., Song, S., Guo, X.G., Jin, D.C., Yang, Q., Barker, S.C., Shao, R., 2014b. 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. 15, 751.
Durden, L.A., Musser, G.G., 1994. The Sucking Lice (Insecta, Anoplura) of the World: A Taxonomic Checklist with Records of Mammalian Hosts and Geographical Distributions. Bulletin of the AMNH no. 218. Sucking lice and hogs.
Feng, S., Yang, Q., Li, H., Song, F., Stejkal, V., Opit, G.P., Cai, W., Li, Z., Shao, R., 2018. The highly divergent mitochondrial genomes indicate that the booklouse, Liposcelis bovaechyphala (Psocoptera: liposcelididae) is a cryptic species. G3 (Bethesda) 8, 1039–1047.
Fu, Y.T., Nie, Y., Duan, D.Y., Liu, G.H., 2020a. Variation of mitochondrial minichromosome composition in *Hoplopleura lice* (*Phthiraptera: hoplopleuridae*) from rats. Parasites Vectors 13, 506.
Fu, Y.T., Dong, Y., Wang, W., Nie, Y., Liu, G.H., Shao, R., 2020b. Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse *Pediculus obtusus* and colobus louse *Pediculus babini*. Genomics 112, 4924–4933.
Herl, K.S., Barker, S.C., Shao, R., 2015. 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. 16, 661.
Jiang, H., Barker, S.C., Shao, R., 2013. Substantial variation in the extent of mitochondrial genome fragmentation among blood-sucking lice of mammals. Genome Biol. Evol. 5, 1298–1308.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinforma. Oxf. Engl. 28, 1647–1649.
Kim, K.C., Ludwig, H.W., 1978. The family classification of the Anoplura. Syst. Entomol. 3, 249–284.
Laslett, D., Canbäck, B., 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinforma. Oxf. Engl. 24, 172–175.
Lavrov, D.V., 2007. Key transitions in animal evolution: a mitochondrial DNA perspective. Integr. Comp. Biol. 47, 734–743.
Light, J.E., Reed, D.L., 2009. Multigene analysis of phylogenetic relationships and divergence times of primate sucking lice (Phthiraptera: Anoplura). Mol. Phylogenet. Evol. 50, 376–395.

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

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

Meloney, W.P., Kim, K.C., 1974. A comparative study of cattle-infesting Haematoptinus, with re description of H. quadripertusus Fahrenholz, 1916 (Anoplura: Haematopinidae). J. Parasitol. 60, 507–522.

Saegerman, C., Bonnet, S., Bouhsira, E., De Regge, N., Fite, J., Etoiré, F., Garigliany, M.-M., Jori, F., Lemperreux, L., Le Potier, M.-F., Quillery, E., Vergne, T., Vial, L., 2021. An expert opinion assessment of blood-feeding arthropods based on their capacity to transmit African swine fever virus in Metropolitan France. Transboundary and emerging diseases 68 (3), 1190–1204.

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

Song, S.D., Barker, S.C., Shao, R., 2014. Variation in mitochondrial minichromosome composition between blood-sucking lice of the genus Haematoptinus that infest horses and pigs. Parasites Vectors 7, 144.

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.

Sweet, A.D., Johnson, K.P., Cao, Y., de Moya, R.S., Skinner, R.K., Tan, M., Virrueta Herrera, S., Cameron, S.L., 2021. Structure, gene order, and nucleotide composition of mitochondrial genomes in parasitic lice from Amblycera. Gene 768, 145312.

Wall, R., Shearer, D., 2008. Lice (phthiraptera). In: Veterinary Ectoparasites: Biology, Pathology and Control. Blackwell Science Ltd, UK, pp. 162–178.

Wolstenholme, D.R., 1992. Genetic novelties in mitochondrial genomes of multicellular animals. Curr. Opin. Genet. Dev. 2, 918–925.