Please do not remove this page

Mitochondrial genome evolution and molecular phylogeny of parasitic lice (infraorder: Phthiraptera)

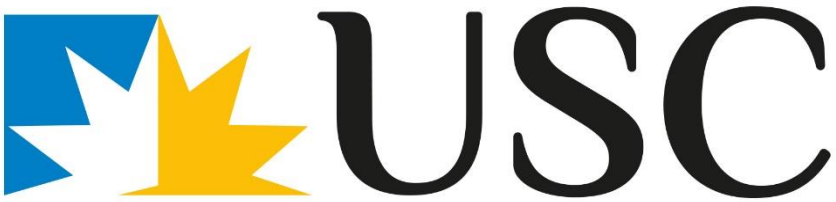
Dong, Yalun

https://research.usc.edu.au/esploro/outputs/doctoral/Mitochondrial-genome-evolution-and-molecular-phylogeny/99684098802621/filesAndLinks?index=0

Dong, Y. (2022). Mitochondrial genome evolution and molecular phylogeny of parasitic lice (infraorder: Phthiraptera) [University of the Sunshine Coast, Queensland]. https://doi.org/10.25907/00712
Document Type: Thesis

UniSC Research Bank: https://research.usc.edu.au
research-repository@usc.edu.au
It's your responsibility to determine if additional rights or permissions are needed for your use.
Downloaded On 2023/11/30 01:35:52 +1000

Please do not remove this page



Mitochondrial genome evolution and molecular phylogeny of parasitic lice (infraorder: Phthiraptera)

Yalun Dong

A thesis submitted for the degree of Doctor of Philosophy at

The University of the Sunshine Coast in 2022

Centre for Bioinnovation, and School of Science, Technology and Engineering

Abstract

Parasitic lice (infraorder Phthiraptera) have five parvorders: Amblycera, Anoplura Ischnocera, Rhynchophthirina and Trichodectera. Sucking lice are in the parvorder Anoplura and chewing lice are in the other four parvorders. Fragmented mitochondrial (mt) genome has been found in lice of all five parvorders. Anoplura, Rhynchophthirina, and Trichodectera species studied so far, all have extensively fragmented mt genomes. For Amblycera and Ischnocera, some species (e.g., pigeon lice) have fragmented mt genomes whereas others have single-chromosome mt genomes. As a shared derived character, mt genome fragmentation supported the close phylogenetic relationship among the three parvorders: Anoplura, Rhynchophthirina and Trichodectera. However, a few studies showed that mt genome fragmentation could occur independently in different lineages of parasitic lice. In this thesis, I investigated the mt genome evolution in a broader range of parasitic lice including both chewing lice and sucking lice from birds and mammals. The potential use of mt genome fragmentation in resolving phylogenetic relationships among parasitic lice was explored and discussed.

This thesis has six chapters. Chapter 1 is a literature review, in which I introduced parasitic lice and their mt genomes. In Chapter 2, I analysed the variation in mt karyotype between the macaque louse and the colobus louse (both in the genus *Pedicinus*) and inferred the mt karyotype of the most recent common ancestor of higher primate lice. I discovered that two of the inferred ancestral mt minichromosomes of 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. The findings of this chapter have been published in Fu & Dong et al. (2020).

In Chapter 3, I reconstructed the phylogeny of sucking lice and inferred the ancestral mt karyotype of *Polyplax* sucking lice. I discovered that: 1) tRNA genes are entirely responsible for mt karyotype variation in three *Polyplax* species studied to date; and 2) tRNA gene translocation is frequent between different types of minichromosomes and towards the boundaries with control regions. A similar tRNA gene translocation pattern can also be seen in other sucking lice with fragmented mt genomes. The findings of this chapter have been published in Dong et al. (2021).

In Chapter 4, I assembled the mt genomes of five seal louse species (family Echinophthiriidae) and the gorilla louse (*Pthirus gorillae*). I then inferred the mt karyotype of the most recent common ancestor of seal lice. My results indicated that: 1) at least three ancestral mt minichromosomes of sucking lice have split in the lineage leading to seal lice; 2) two ancestral minichromosomes of seal lice have merged in the lineage to the northern fur seal louse; and 3) one minichromosome ancestral to primate lice has split in the lineage to the gorilla louse. My analysis of mt karyotypes and gene sequences also indicated the possibility of a host switch of crabeater seal louse to Weddell seals. The findings of this chapter have been published in Dong et al. (2022).

In Chapter 5, I sequenced and assembled the mt genomes of 17 species of Amblycera lice – 14 in the family Menoponidae and three in the family Laemobothriidae. I found that four Menoponidae species in three genera (*Actornithophilus*, *Austromenopon* and *Myrsidea*) have fragmented mt genomes, whereas all the other 13 Amblycera species have the typical single-chromosome mt genomes. My analysis showed that mt genome fragmentation occurred four times independently in each of the four genera: *Actornithophilus*, *Austromenopon*, *Myrsidea* and *Laemobothrion*. The manuscript that reports the findings of this chapter has been externally reviewed by the journal *Cladistics* and a revised manuscript is currently under review with the journal.

In Chapter 6, I summarized my research findings and discussed further study that can be done in the future. This thesis is a significant contribution to our understanding of the evolution of mt genomes of parasitic lice. For the first time, I discovered that: 1) mt karyotype evolved in opposite directions between two closely related congeneric species: macaque louse (*Pedicinus obtusus*) and colobus louse (*Pedicinus badii*); 2) for sucking lice, tRNA gene translocation is frequent between different types of minichromosomes and towards the boundaries with control regions; 3) minichromosomal split was much more frequent than minichromosomal merger in seal lice, and a host switch likely occurred from crabeater seals to Weddell seals; 4) mt genome fragmentation occurred four times independently in each of the four genera of Amblycera: *Actornithophilus*, *Austromenopon*, *Myrsidea* and *Laemobothrion*; and 5) each independent mt genome fragmentation event produced unique shared derived minichromosomal characters that could be informative in resolving the phylogeny of parasitic lice at different taxonomic levels.

Declaration of originality

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of the Sunshine Coast, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of Graduate Research.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications included in this thesis

Fu, Y. T., **Dong, Y.** (equal first author)**,** Wang, W., Nie, Y., Liu, G. H., & Shao, R. (2020). Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse *Pedicinus obtusus* and colobus louse *Pedicinus badii.*
Genomics, 112(6), 4924-4933.
Manuscript: Appendix I

Dong, W. G., **Dong, Y**., Guo, X. G., & Shao, R. (2021).
Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals.
BMC Genomics, 22(1), 1-16.
Manuscript: Appendix II

Dong, Y., Zhao, M., & Shao, R. (2022).
Fragmented mitochondrial genomes of seal lice (family Echinophthiriidae) and gorilla louse (family Pthiridae): frequent minichromosomal splits and a host switch of lice between seals.
BMC Genomics, 23(1), 1-13.
Manuscript: Appendix III

Submitted manuscripts included in this thesis

Dong, Y., Martina, J., Amber, G., Ludovica, V., Gabriel, C., Dominique, P., & Shao, R. (2022). Mitochondrial genome fragmentation occurred multiple times independently in bird lice of the families Menoponidae and Laemobothriidae.
Under review with *Cladistics*

Other publications during candidature

Peer-reviewed papers:

Fu, Y. T., **Dong, Y.** (equal first author)**,** Wang, W., Nie, Y., Liu, G. H., & Shao, R. (2020). Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse *Pedicinus obtusus* and colobus louse *Pedicinus badii.*
Genomics, 112(6), 4924-4933.

Dong, W. G., **Dong, Y**., Guo, X. G., & Shao, R. (2021).
Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals.
BMC Genomics, 22(1), 1-16.

Dong, Y., Zhao, M., & Shao, R. (2022).
Fragmented mitochondrial genomes of seal lice (family Echinophthiriidae) and gorilla louse (family Pthiridae): frequent minichromosomal splits and a host switch of lice between seals.
BMC Genomics, 23(1), 1-13.

Kasimov, V., **Dong, Y**., Shao, R., Brunton, A., Anstey, S. I., Hall, C., ... & Jelocnik, M. (2022). Emerging and well-characterised chlamydial infections detected in a wide range of wild Australian birds.
Transboundary and Emerging Diseases. 2022 Jan 18. doi: 10.1111/tbed.14457

Contributions by others to the thesis

Renfu Shao contributed to the conception and design of this research, as well as the critical revision of the thesis. Contributions by others to individual chapters are acknowledged at the end of each chapter.

Statement of parts of the thesis submitted to qualify for the award of another degree

No works submitted towards another degree have been included in this thesis.

Research involving human or animal subjects

Project number: ANA19154

Project title: Collecting parasitic lice of birds for molecular and evolutionary studies

Start date: 09/11/2020

End date: 31/12/2022

Ethics approval letter: Appendix IV

Acknowledgements

I have learned a lot in phylogenetic analysis and gained skills in assembling Illumina sequence data and reconstructing phylogenetic trees. I am very grateful to all who have contributed to my PhD project.

I would like to thank my principal supervisor Dr Renfu Shao for all the support and encouragement he gave me during my PhD study. His enthusiasm and rigor to science influenced me a lot. He paid a lot of attention to his student and always be helpful when I had difficulty in my project. Many thanks also to my co-supervisor Dr Min Zhao for his guidance and support. Without the assistance of Drs Renfu Shao and Min Zhao, I would not have completed this work.

I gratefully acknowledge the funding received for my PhD from University of the Sunshine Coast and the Australian Biological Resources Study, which provided me the opportunity to work on this great project.

I want to thank Martina Jelocnik (UniSC), Rosemary Booth (AZWH), Amber Gillett (AZWH), Ludovica Valenza (AZWH), Gabriel Conroy (UniSC), Dominique Potvin (UniSC), Danny Brown (AZWH) and other staff in the AZWH, and Vasilli Kasimov (UniSC) for their help in the collection of bird lice used in my research. I also want to thank Wei Wang (UniSC) for his assistance in identification of bird lice.

Finally, I would like to say a heartfelt thank you to my parents and my girlfriend for their constant support, encouragement and patience. Your support and care are also important in helping me accomplish my PhD project. I will always love you all. Thank you!

Financial support

This research was supported by an UniSC Research Scholarship, the Australian Biological Resources Study grant (ABRS, RS217-51), and an ECR and MCR grant from the Centre for Bioinnovation, UniSC.

Keywords

Mitochondrial genome fragmentation, parasitic lice, molecular phylogeny

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060309 Phylogeny and Comparative Analysis
ANZSRC code: 060409 Molecular Evolution

Fields of Research (FoR) Classification

FoR code: 0603 Evolutionary Biology
FoR code: 0604 Genetics

Table of Contents

Thesis abstract……..……2
Declaration of originality……..4
Publications included in this thesis……...5
Submitted manuscripts included in this thesis…….....6
Other publications during candidature…….....6
Contributions by others to the thesis…….7
Statement of parts of the thesis submitted to qualify for the award of another degree….7
Research involving human or animal subjects…….……7
Acknowledgements……...8
Financial support……...9
Keywords……..….9
Australian and New Zealand Standard Research Classifications (ANZSRC)……..9
Fields of Research (FoR) Classification…….....9
List of figures……..16
List of tables…….....21
List of abbreviations used in the thesis……24

Chapter 1: Introduction……..…..….26
1.1. Parasitic lice (infraorder: Phthiraptera)……...27
1.1.1. Origin of parasitic lice …….……29
1.1.2. Morphology of chewing lice ……..….29
1.1.3. Morphology of sucking lice…….30
1.1.4. Effects of lice on their hosts…….30
1.1.5. Host defence……31
1.1.6. Host specificity and co-speciation……...31
1.1.7. Host switch……..32
1.2. Mitochondrial genomes of parasitic lice……33
1.2.1. Typical mitochondrial genomes……..33
1.2.2. Atypical mitochondrial genomes…….34
1.3. Phylogeny of parasitic lice (infraorder Phthiraptera)……..37
1.4. Summary...……...39

1.5. References……39

Chapter 2: Mitochondrial karyotypes evolved in opposite directions between closely related macaque louse *Pedicinus obtusus* and colobus louse *Pedicinus badii*……48

2.1. Abstract…….....49

2.2. Introduction……..49

2.3. Materials and methods……..51

2.3.1. Collection of macaque lice, DNA extraction, sequencing, sequence assembly, minichromosome verification and mitochondrial genome annotation of the macaque louse, *Pedicinus obtusus*……..51

2.3.2. Retrieval and analysis of primate louse sequence data from the Sequence Read Archive……..52

2.3.3. Phylogenetic analysis…….....53

2.3.4. Inference of the ancestral mitochondrial karyotype of higher primate lice…..54

2.4. Results……..56

2.4.1. Mitochondrial genome of the macaque louse, *Pedicinus obtusus*, comprises 12 minichromosomes…….…....56

2.4.2. Mitochondrial genome of the colobus louse, *Pedicinus badii*, comprises 14 minichromosomes……58

2.4.3. *trnK-nad4* mitochondrial minichromosome of the human pubic louse, *Pthirus pubis*, revealed by SRA data…….…….59

2.4.4. Phylogeny of sucking lice reconstructed with mitochondrial genome sequences……..60

2.4.5. Inferred ancestral mitochondrial karyotype of higher primate lice…….……61

2.5. Discussion……....62

2.5.1. Variation in mitochondrial karyotype between *Pedicinus badii* and *Pedicinus obtusus* is the most pronounced among congeneric species of sucking lice observed to date……....62

2.5.2. Mitochondrial karyotypes evolved in opposite directions between *Pedicinus badii* and *Pedicinus obtusus*……..65

2.6. Conclusions……..66

2.7. References……....67
2.8. Supplementary files……..71

Chapter 3: Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals ……....77
3.1. Abstract……..…78
3.2. Introduction…….…..….78
3.3. Materials and methods……..…80
3.3.1. Louse collection, DNA extraction, mitochondrial genome amplification and sequencing……...……..…80
3.3.2. Assembly of sequence reads, gene identification and minichromosome verification…….……81
3.3.3. Phylogenetic analyses……....82
3.3.4. Inference of the ancestral mitochondrial karyotype of *Polyplax* lice……83
3.4. Results and Discussion…….85
3.4.1. Mitochondrial genome of the Asian grey shrew louse, *Polyplax reclinata*……85
3.4.2. Mitochondrial karyotype variation among *Polyplax* species……87
3.4.3. Phylogeny of sucking lice inferred from mt genome sequences……...87
3.4.4. Inferred ancestral mitochondrial karyotype of *Polyplax* lice……88
3.4.5. Mitochondrial tRNA gene translocation in *Polyplax* lice…….90
3.4.6. tRNA gene translocation is frequent and directional towards the control region in parasitic lice with fragmented mitochondrial genomes……91
3.5. Conclusions…….…..94
3.6. References……..…98
3.7. Supplementary files…….....103

Chapter 4: Fragmented mitochondrial genomes of seal lice (family Echinophthiriidae) and gorilla louse (family Pthiridae): frequent minichromosomal splits and a host switch of lice between seals……106
4.1. Abstract……..107

4.2. Introduction……107
4.3. Materials and methods……...…109
4.3.1. Retrieval and assembly of Sequence Read Archive (SRA) data of seal lice and gorilla louse…….109
4.3.2. Phylogenetic analyses…….….110
4.3.3. Inferring the ancestral mitochondrial karyotype of seal lice…….…..111
4.4. Results……112
4.4.1. Mitochondrial minichromosomes of *Lepidophthirus macrorhini* – louse of southern elephant seal (*Mirounga leonine*)……...112
4.4.2. Mitochondrial minichromosomes of *Proechinophthirus fluctus* – louse of northern fur seal (*Callorhinus ursinus*)……...113
4.4.3. Mitochondrial minichromosomes of *Antarctophthirus carlinii* - louse of Weddell seal (*Leptonychotes weddellii*), and *Antarctophthirus lobodontis* - louse of crabeater seal (*Lobodon carcinophagus*)…….…..115
4.4.4. Mitochondrial minichromosomes of *Antarctophthirus microchir* – louse of Australian sea lion (*Neophoca cinerea*)…….115
4.4.5. Mitochondrial minichromosomes of *Pthirus gorillae* – louse of western gorilla (*Gorilla gorilla*)……....116
4.4.6. Phylogeny of sucking lice based on mitochondrial gene sequences…….…..117
4.4.7. Inferred ancestral mitochondrial karyotype of the seal lice…….....119
4.5. Discussion……...120
4.5.1. Mitochondrial minichromosome split occurred more frequently than minichromosome merger in the lineages leading to seal lice and gorilla louse…….....120
4.5.2. The very high mitochondrial gene identity shared between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* does not support them as separate species…….....123
4.6. Conclusions……..……..…124
4.7. References……..125
4.8. Supplementary files…….....129

Chapter 5: Mitochondrial genome fragmentation occurred multiple times independently in bird lice of the families Menoponidae and Laemobothriidae……144
5.1. Abstract……..145
5.2. Introduction……145
5.3. Materials and methods……147
5.3.1. Collection of bird lice……..147
5.3.2. DNA extraction, mitochondrial genome sequencing and assembly……147
5.3.3. PCR verification of mitochondrial minichromosomes……....149
5.3.4. Phylogenetic analysis……...149
5.4. Results……..…150
5.4.1. Bird lice have both fragmented and typical single-chromosome mitochondrial genome organization……..……...150
5.4.2. Mitochondrial karyotype differs between two *Actornithophilus* species……151
5.4.3. The typical single-chromosome mitochondrial genome was retained in *Austromenopon paululum* but broke up into two chromosomes in *Austromenopon atrofulvum*……...152
5.4.4. The mitochondrial genome of *Myrsidea ptilorhynchi* comprises at least four minichromosomes……...153
5.4.5. Three *Laemobothrion* species retained the typical single-chromosome mitochondrial genome organization……154
5.4.6. Phylogeny of Menoponidae lice inferred with mitochondrial genome sequences……155
5.5. Discussion……...158
5.5.1. Mitochondrial genome fragmentation occurred three times independently in the Menoponidae species and once in the Laemobothriidae species investigated to date……..158
5.5.2. Minichromosomal characters derived from mt genome fragmentation provide valuable information for resolving phylogenetic relationships at different taxonomic levels……...159
5.6. Conclusions……....162
5.7. References……..162

5.8. Supplementary files……167

Chapter 6: Summary and future study…….175

Appendix I…….....182
Appendix II……193
Appendix III……..210
Appendix IV……..224
Appendix V……...227

List of figures

Chapter 1

Fig 1.1. Ancestral mitochondrial gene arrangement of insects…….……..…..34

Fig 1.2. The inferred ancestral mitochondrial karyotype of sucking lice…….…36

Fig 1.3. Newly proposed classification of Phthiraptera……....39

Chapter 2

Fig. 2.1. Inferred ancestral mitochondrial karyotype of higher primate lice (Pedicinidae, Pediculidae and Pthiridae).……...55

Fig. 2.2. The mitochondrial genome of macaque louse, *Pedicinus obtusus*..…….….57

Fig. 2.3. Inferred secondary structures of the mitochondrial tRNA-Leu (tag) and tRNA-Leu (taa) of *Pedicinus obtusus*.…….....58

Fig. 2.4. (A) The coding regions of the 14 minichromosomes and conserved AT-rich region and conserved GC-rich region of the colobus louse, *Pedicinus badii*; (B) the coding region of *K-nad4* minichromosomes and conserved AT-rich region and conserved GC-rich region of the human pubic louse, *Pthirus pubis*.……..…….60

Fig. 2.5. Phylogenetic relationships among 14 species of sucking lice (Anoplura) inferred from Bayesian analysis of deduced amino acid sequences of nine mitochondrial genes...…….……..61

Fig. 2.6. Inferred ancestral mitochondrial karyotype of higher primate lice (Pedicinidae, Pediculidae and Pthiridae).……...……62

Fig. 2.7. (A) Two ancestral mitochondrial minichromosomes of higher primate lice 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*.……..66

Supplementary Fig. 2.1. PCR verification of the mitochondrial minichromosomes of the macaque louse, *Pedicinus obtusus*……..71

Supplementary Fig. 2.2. Inference of the positions of tRNA genes in the ancestral mitochondrial minichromosomes of higher primate lice……72

Supplementary Fig. 2.3. Alignment of nucleotide sequences in the NCRs of the mitochondrial minichromosomes of the macaque louse, *Pedicinus obtusus*…….73

Supplementary Fig. 2.4. Alignment of nucleotide sequences of the GC-rich region (A) and the AT-rich region in the mitochondrial minichromosomes of the colobus louse, *Pedicinus badii*……...74

Supplementary Fig. 2.5. Phylogenetic relationships among 14 species of sucking lice (Anoplura) inferred from maximum likelihood (ML) analysis of deduced amino acid sequences of nine mitochondrial genes…….…..75

Chapter 3

Fig. 3.1. Inferring the ancestral mitochondrial karyotype of *Polyplax* lice……..….84

Fig. 3.2. Inferring the location of tRNA genes relative to protein-coding and rRNA genes in the ancestral mitochondrial karyotype of *Polyplax* lice…….…84

Fig. 3.3. The mitochondrial genome of the Asian grey shrew louse, *Polyplax reclinata*……...86

Fig. 3.4. Phylogenetic relationships among 15 species of sucking lice (Anoplura) inferred with combined sequences of eight mitochondrial protein-coding genes (*atp6*, *atp8, cox1*, *cox2*, *cox3, cob*, *nad4L, nad6*) and two rRNA genes (*rrnS* and *rrnL*) with maximum likelihood method (A) and Bayesian method (B)……..……...89

Fig. 3.5. Inferred ancestral mitochondrial karyotype of *Polyplax* lice…….…..…90

Supplementary Fig. 3.1. PCR amplicons generated from the mitochondrial minichromosomes of the Asian grey shrew louse, *Polyplax reclinata*..……..103

Supplementary Fig. 3.2. Alignment of the non-coding region (*NCR*) sequences of 11 mitochondrial minichromosomes of the Asian grey shrew louse, *Polyplax reclinata*.…….….104

Chapter 4

Fig. 4.1. Inference of the ancestral mitochondrial karyotype of seal lice.…….....112

Fig. 4.2. Mitochondrial genomes of five seal louse species: (A) southern elephant seal louse, *Lepidophthirus macrorhini*; (B) northern fur seal louse, *Proechinophthirus fluctus*; (C) Weddell seal louse, *Antarctophthirus carlinii,* and crabeater seal louse, *Antarctophthirus lobodontis*; and (D) Australian sea lion louse, *Antarctophthirus microchir*.……..……114

Fig. 4.3. Mitochondrial genome of the gorilla louse, *Pthirus gorillae*.…….....117

Fig. 4.4. Sequence alignment of degenerate genes with their corresponding full-length genes.…….118

Fig. 4.5. Bayesian phylogenetic tree inferred from the nucleotide sequences of five mitochondrial protein-coding genes (*cob*, *cox1*, *cox2*, *cox3*, *nad4*) of 21 species of sucking lice (Anoplura)…….119

Supplementary Fig. 4.1. Secondary structure inferred with tRNA-Scan from $trnW_1$ and $trnW_2$ gene sequences of the southern elephant seal louse, *Lepidophthirus macrorhini*…….…….129

Supplementary Fig. 4.2. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the southern elephant seal louse, *Lepidophthirus macrorhini*.....129

Supplementary Fig. 4.3. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the northern fur seal louse, *Proechinophthirus fluctus*…….……130

Supplementary Fig. 4.4. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the Weddell seal louse, *Antarctophthirus carlinii*.....130

Supplementary Fig. 4.5. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the crabeater seal louse, *Antarctophthirus lobodontis*……..…..131

Supplementary Fig. 4.6. Conserved non-coding GC-rich motifs among the mitochondrial minichromosomes of the Australian sea lion louse, *Antarctophthirus microchir*……..131

Supplementary Fig. 4.7. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the gorilla louse, *Pthirus gorillae*…….132

Supplementary Fig. 4.8. Phylogenetic relationships among 21 species of sucking lice (Anoplura) inferred by maximum likelihood (ML) analysis of nucleotide sequences of five mitochondrial protein-coding genes…….……132

Supplementary Fig. 4.9. Alignment of partial *18S* rRNA gene, *28S* rRNA gene and *ef1-α* gene sequences between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*……...…133

Chapter 5

Fig. 5.1. The mitochondrial genomes of *Actornithophilus grandiceps* (louse of pied oystercatcher) and *Actornithophilus hoplopteri* (louse of masked lapwing)……...152

Fig. 5.2. The mitochondrial genomes of (A) *Austromenopon paululum* (louse of sooty shearwater) and (B) *Austromenopon atrofulvum* (louse of sooty tern and crested tern)……..…153

Fig. 5.3. The mitochondrial genome of *Myrsidea ptilorhynchi* (louse of satin bowerbird)……...154

Fig. 5.4. Phylogenetic relationships among 23 species of chewing lice (parvorder Amblycera) inferred from Bayesian inference (A) and maximum likelihood (B) analysis of amino acid sequences of all mitochondrial protein-coding genes.…….157

Fig. 5.5. Phylogenetic relationships among 23 species of chewing lice (parvorder Amblycera) inferred from maximum parsimony analysis of amino acid sequences of all mitochondrial protein-coding genes……157

Supplementary Fig. 5.1. The mitochondrial genomes of *Ciconiphilus decimfasciatus* (louse of white-faced heron), *Colpocephalum eucarenum* (louse of Australian pelican), *Colpocephalum spinicollis* (louse of straw-necked ibis), *Eomenopon denticulatum* (louse of scaly breasted lorikeet), *Franciscoloa funerei* (louse of yellow-tailed black cockatoo), *Franciscoloa pallida* (louse of sulfur-crested cockatoo), *Franciscoloa roseicapillae* (louse of little corella, galah and pheasant coucal), *Piagetiella africana* (louse of Australian pelican), *Plegadiphilus threskiornis* (louse of Australian white ibis), *Laemobothrion atrum* (louse of Eurasian coot), *Laemobothrion* sp. (louse of Australian swamphen) and *Laemobothrion maximum* (louse of black kite)……..……167

Supplementary Fig. 5.2. PCR verification of the five mitochondrial minichromosomes (M1-M5) of *Actornithophilus grandiceps* (louse of pied oystercatcher)…….....168

Supplementary Fig. 5.3. PCR verification of the six minichromosomes (M1-M6) of *Actornithophilus hoplopteri* (louse of masked lapwing)…169

Supplementary Fig. 5.4. PCR verification of the two minichromosomes (M1 and M2) of *Austromenopon atrofulvum* (louse of sooty tern louse). Marker band sizes (bp) are indicated on the left side……..….170

Supplementary Fig. 5.5. PCR verification of the four minichromosomes (M1-M4) of *Myrsidea ptilorhynchi* (louse of satin bowerbird)……..171

List of tables

Chapter 2

Table 2.1. Species of parasitic lice included in the phylogenetic analyses in this study……54

Table 2.2. Mitochondrial minichromosomes of the macaque louse, *Pedicinus obtusus*….....56

Table 2.3. Mitochondrial minichromosomes of the colobus louse, *Pedicinus badii*.....59

Supplementary Table 2.1. PCR primers used to verify each mitochondrial minichromosome of the macaque louse, *Pedicinus obtusus*…….76

Chapter 3

Table 3.1. Species of parasitic lice included in the phylogenetic analyses in this study……..…83

Table 3.2. The mitochondrial minichromosomes of the Asian grey shrew louse, *Polyplax reclinata***,** identified by Illumina sequencing……....86

Table 3.3. Comparison of gene content and gene arrangement in the mitochondrial minichromosomes of three *Polyplax* species and their most recent common ancestor (MRCA)……..….87

Table 3.4. Statistical analysis of expected probability and observed probability of tRNA gene translocation in three *Polyplax* species relative to their most recent common ancestor……....91

Table 3.5. Translocated tRNA genes in the most recent common ancestor (MRCA) of *Polyplax* lice relative to the MRCA of sucking lice……..…92

Table 3.6. Statistical analysis of expected probability and observed probability of tRNA gene translocation in the most recent common ancestor (MRCA) of *Polyplax* lice relative to the MRCA of sucking lice……93

Table 3.7. Translocated tRNA genes in non-*Polyplax* sucking lice relative to the most recent common ancestor (MRCA) of sucking lice……....95

Table 3.8. Statistical analysis of expected probability and observed probability of tRNA gene translocation in non-*Polyplax* sucking lice relative to the most recent common ancestor of sucking lice……..…...97

Supplementary Table 3.1. The primers used to amplify the mitochondrial genes, minichromosomes and coding regions of the Asian grey shrew louse, *Polyplax reclinata*……..……..…105

Chapter 4

Supplementary Table 4.1. Species of parasitic lice included in the phylogenetic analyses in this study……..…..……134

Supplementary Table 4.2. Mitochondrial minichromosomes of *Lepidophthirus macrorhini* - louse of southern elephant seal (*Mirounga leonine*)…..…..135

Supplementary Table 4.3. Sequence identities between *trnW*$_1$ and *trnW*$_2$ of the southern elephant seal louse, *Lepidophthirus macrorhini*, between *trnW* genes of *Lepidophthirus macrorhini* and other seal lice, and between *trnW* genes and other tRNA genes of *Lepidophthirus macrorhini*.....136

Supplementary Table 4.4. Mitochondrial minichromosomes of *Proechinophthirus fluctus* - louse of northern fur seal (*Callorhinus ursinus*)…….137

Supplementary Table 4.5. Mitochondrial minichromosomes of *Antarctophthirus carlinii* - louse of Weddell seal (*Leptonychotes weddellii*)…….....138

Supplementary Table 4.6. Mitochondrial minichromosomes of *Antarctophthirus lobodontis* - louse of crabeater seal (*Lobodon carcinophagus*)…….…...…..139

Supplementary Table 4.7. Mitochondrial minichromosomes of *Antarctophthirus microchir* – louse of Australian sea lion (*Neophoca cinerea*)…….……140

Supplementary Table 4.8. Mitochondrial minichromosomes of *Pthirus gorilla* – louse of western gorilla (*Gorilla gorilla*)……...……...141

Supplementary Table 4.9. Sequence identities between *Antarctophthirus carlinii* (louse of Weddell seal, *Leptonychotes weddellii*) and *Antarctophthirus lobodontis* (louse of crabeater seal, *Lobodon carcinophagus*)…….….142

Supplementary Table 4.10. Sequence identities between Weddell seal (*Leptonychotes weddellii*) and crabeater seal (*Lobodon carcinophagus*)……143

Chapter 5

Table 5.1. Species of chewing lice included in the phylogenetic analyses in this study….....148

Supplementary Table 5.1. Specific primers used for PCR verification of minichromosomes……172

Supplementary Table 5.2. Fragmented mitochondrial minichromosomes of four Menoponidae species……...……...172

Supplementary Table 5.3. Derived minichromosomes and gene clusters observed in Menoponidae and Laemobothriidae species……173

List of abbreviations used in the thesis

A: adenine

atp6 and *atp8*: genes for ATP synthase subunits 6 and 8

AZWH: Australia Zoo Wildlife Hospital

bp: base pair

C: cytosine

cob: gene for cytochrome b

cox1, *cox2* and *cox3*: Genes for cytochrome c oxidase subunits 1, 2 and 3

DNA: deoxyribonucleic acid

G: guanine

kb: kilo base pair

min: minute

MRCA: most recent common ancestor

mt: mitochondrial

mtSSB: mitochondrial single-stranded binding protein

Mya: million years ago

nad1, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*: Mitochondrial genes for NADH dehydrogenase subunits 1-6 and 4 L

PCR: polymerase chain reaction

RNA: ribonucleic acid

rRNA ribosomal RNA

rrnS and *rrnL*: genes for small and large subunits of ribosomal RNA

sec: second

SRA: Sequence Read Archive

tRNA: transfer RNA

trnA or *A*: tRNA gene for alanine

trnC or *C*: tRNA gene for cysteine

trnD or *D*: tRNA gene for aspartic acid

trnE or *E*: tRNA gene for glutamic acid

trnF or *F*: tRNA gene for phenylalanine

trnG or *G*: tRNA gene for glycine

trnH or *H*: tRNA gene for histidine

trnI or *I*: tRNA gene for isoleucine

trnK or *K*: tRNA gene for lysine

trnL$_1$ or *L$_1$*: tRNA gene for leucine (anticodon NAG)

trnL$_2$ or *L$_2$*: tRNA gene for leucine (anticodon YAA)

trnM or *M*: tRNA gene for methionine

trnN or *N*: tRNA gene for asparagine

trnP or *P*: tRNA gene for proline

trnQ or *Q*: tRNA gene for glutamine

trnR or *R*: tRNA gene for arginine

trnS$_1$ or *S1*: tRNA gene for serine (anticodon NCU)

trnS$_2$ or *S$_2$*: tRNA gene for serine (anticodon NGA)

trnT or *T*: tRNA gene for threonine

trnV or *V*: tRNA gene for valine

trnW or *W*: tRNA gene for tryptophan

trnY or *Y*: tRNA gene for tyrosine

U: uracil

UniSC: University of the Sunshine Coast

Chapter 1

Introduction

Lice (infraorder Phthiraptera) are obligate ectoparasites of mammals and birds (Kim & Ludwig, 1978; Lyal, 1985; Durden & Musser, 1994; Price et al., 2003; Ramli et al., 2008; Smith, 2008). There are five parvorders in Phthiraptera: Amblycera, Anoplura Ischnocera, Rhynchophthirina and Trichodectera (De Moya et al., 2021). Blood sucking lice are in the parvorder Anoplura while chewing lice are in the other four parvorders. Two types of mitochondrial (mt) genome structures have been found in parasitic lice: typical single-chromosome mt genomes, and fragmented mt genomes (Shao et al., 2009; Shao et al., 2012; Song et al., 2019; Sweet et al., 2020; Sweet et al., 2021). Both typical and fragmented mt genomes have been found in species in Amblycera and Ischnocera while all the species studied in the parvorders Anoplura, Rhynchophthirina and Trichodectera have fragmented mt genomes (Shao et al., 2009; Shao et al 2012; Jiang et al., 2013; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Shao et al., 2017; Song et al., 2019; Sweet et al., 2020; Sweet et al., 2021). Song et al., (2019) showed that mt genome fragmentation as a derived shared character supported the close phylogenetic relationship among the three parvorders: Anoplura, Rhynchophthirina and Trichodectera. A few recent studies, however, showed that mt genome fragmentation could occur independently in different lineages of parasitic lice (Sweet et al., 2020; Sweet et al., 2021). In my PhD project, I studied mt genome evolution in a broader range of parasitic lice and investigated whether mt genome fragmentation and its derived genomic characters could be used to resolve phylogenetic relationships among parasitic lice. As a start, I will review the literature on the biology, mitochondrial genomes, and phylogeny of parasitic lice in this chapter.

1.1. Parasitic lice (infraorder: Phthiraptera)

Traditionally, parasitic lice were grouped as a monophyletic order, Phthiraptera, and divided into four suborders; chewing lice were in three suborders: Amblycera; Ischnocera and Rhynchophthirina, while sucking lice were in one suborder: Anoplura (Kim & Ludwig, 1978; Lyal, 1985; Durden & Musser, 1994; Samuel et al., 2001; Barker et al., 2003; Price et al., 2003). This classification has been challenged in recent studies regarding the status of Ischnocera as a suborder (Johnson et al., 2018; Song et al., 2019; De Moya et al., 2021). Traditionally, Ischnocera has two families: Trichodectidae and Philopteridae (Lyal, 1985, Price et al., 2003). Song et al (2019) showed that the family Trichodectidae was closer to Anoplura and Rhynchophthirina than to the family Philopteridae, thus rejecting the monophyly of Ischnocera. Johnson et al. (2018) also reconstructed the phylogeny of parasitic lice with sequences of 1,107 genes and showed the suborder Ischnocera was paraphyletic. As a result, a new classification was proposed by De Moya

et al. (2021), in which parasitic lice were grouped in the infraorder Phthiraptera of the suborder Troctomorpha, and were divided into five parvorders: Amblycera, Anoplura, Ischnocera, Rhynchophthirina and Trichodectera. The name Ischnocera is retained in the new classification; however, the name Ischnocera has a different meaning between the traditional classification and the new classification. In the traditional classification, Ischnocera contains Philopteridae and Trichodectidae. In the new classification, Ischnocera contains only Philopteridae; Trichodectidae has been moved to the parvorder Trichodectera. In this thesis, the new classification of parasitic lice is adopted.

The parvorder Amblycera has 1,200 described species classified into 102 genera of six families (Price et al., 2003). Three families of Amblycera (Laemobothriidae, Menoponidae and Ricinidae) are bird lice and the other three families (Boopiidae, Gyropidae and Trimenoponidae) are mammal lice (Lyal, 1985; Price et al., 2003). *Therodoxus oweni* is the only species in the Boopiidae parasitizing a bird, cassowary (Price et al., 2003; Smith, 2008). Clay (1971) discussed the occurrence of Boopiidae on cassowaries: either cassowaries acquired lice from marsupials, or ancestral marsupial lice were from birds. Emerson and Price (1976) also described the seventh family of Amblycera, named Abrocomophagidae. However, this family is currently invalid (Price et al., 2003).

The parvorder Ischnocera contains around 3,500 species of chewing lice in one family, Philopteridae (De Moya et al., 2021). Philopteridae has 138 genera and is the most genus rich family of parasitic lice. Species of Philopteridae parasitize only birds except for the lemur louse, *Trichophilopterus babakotophilus*, which is likely switched from birds to lemurs (Price et al., 2003; Cameron et al., 2011).

The parvorder Trichodectera has one family, Trichodectidae (De Moya et al., 2021). Trichodectidae contains less than 500 species in 38 genera (Price et al., 2003; Smith, 2008). All species of Trichodectidae parasitize mammals (Lyal, 1985; Price et al., 2003).

The parvorder Rhynchophthirina contains only one family Haematomyzidae with one genus *Haematomyzus*. There are three described species in the genus *Haematomyzus*: *Haematomyzus*

elephantis (louse of elephants), *Haematomyzus hopkinsi* (louse of desert warthog) and *Haematomyzus porci* (louse of bush pigs) (Richards & Davies, 1977; Lyal, 1985; Price et al., 2003).

The parvorder Anoplura has 532 described species in 49 genera of 15 families; around 840 species of eutherian mammals in 12 of the 29 orders of mammals are reported as hosts of sucking lice (Durden & Musser, 1994; Light & Hafner, 2007; Light et al., 2010). Over 60% of the sucking lice parasitize rodents and most of the rodent lice are grouped in the families Hoplopleuridae and Polyplacidae (Kim, 1988; Durden & Musser, 1994). The three species of human lice are all sucking lice, which inhabit the head, body and pubic areas respectively (Badiaga & Brouqui, 2012). The monophyly of Anoplura has been supported by both morphological data and molecular data (Lyal, 1985; Cruickshank et al., 2001; Barker et al., 2003).

1.1.1. Origin of parasitic lice

Due to a limited number of fossils available, the origin of parasitic lice (Phthiraptera) is still unresolved (Hopkins, 1949; Lyal, 1985). The most supported hypothesis is that parasitic lice appeared between late Cretaceous (60 MYA) and late Carboniferous (280 MYA) (Price et al., 2003). Smith et al. (2011b) suggested that originally, parasitic lice lived on feathered theropod dinosaurs at around 130 MYA. However, Misof et al. (2014) suggested that the origin of Phthiraptera was around 53 MYA.

1.1.2. Morphology of chewing lice

The size of chewing lice varies, ranging from 0.8 mm to 11 mm (Price et al., 2003; Smith, 2008; Waage, 2008). In most species, females are around 20% larger in size than males (Price et al., 2003). They have dorsoventrally flattened bodies with heads horizontally positioned, which helps them move and groom on the feathers or hair of their hosts (Ramli et al., 2008; Smith, 2008). The color of chewing lice can be white, yellowish to brown or even black. In some species, the color of lice is quite the same to that of their hosts (Price et al., 2003; Ramli et al., 2008). Rothschild and Clay (1957) suggested that similarity in color was a strategy for avoiding host detection. However, this hypothesis has not been tested, to my knowledge.

Several morphological features differ among the four parvorders of chewing lice. Firstly, their mouthpart is different. Amblycera lice have palps, which is a primitive feature inherited from their psocopteran ancestors. The mouthpart of Amblycera can move in a vertical plane while Ischnocera

lice move their mouthpart in a horizontal plane parallel to the head (Price et al., 2003). Some Amblycera species, such as *Ricinus* species, have modified chewing mouthpart for blood sucking (Clay, 1949). Rhynchophthirina have weevil-like mouthpart - there is a long proboscis at the end of their mouthpart. Secondly, their antennae vary. The antennae of Amblycera have four segments but this structure is hard to be observed because their antennae are concealed in lateral grooves (Richards & Davies, 1977). Chewing lice in the other three parvorders have obvious antennae with 3-5 segments and some male Ischnocera, such as *Columbicola columbae*, *Neotrichodectes minutus* and *Cebidicola extrarius*, have larger antennae than females (Richards & Davies, 1977; Lyal, 1985; Price et al., 2003; Smith, 2008). With these antennae, males can hold on for hours, even days, clasping females during copulation (Price et al., 2003). Thirdly, the structure of thoracic segments is distinctive between Amblycera and Ischnocera. Amblycera species have three thoracic segments that divide their thorax into prothorax, mesothorax and metathorax while for Ischnocera, only two apparent thoracic segments can be observed due to the fusion of mesothorax and metathorax (Richards & Davies, 1977; Lyal, 1985; Price et al., 2003; Smith, 2008). The abdominal segments are quite unique in different genera of chewing lice. Most chewing lice have 8-10 segments whereas some have 11 abdominal segments (Price et al., 2003; Smith, 2008).

1.1.3. Morphology of sucking lice

Sucking lice (Anoplura) are usually smaller than chewing lice; their body length is between 0.5 mm and 8 mm (Ibarra, 1993). Other than the sucking mouthpart and small body size, sucking lice look similar to chewing lice (Snodgrass, 1944). They have flattened and wingless bodies, tibiotarsal claws for clinging to hosts' hair and elongated mouthpart for penetrating hosts' skin to suck blood (Kim & Ludwig, 1978; Light et al., 2010). Ancestral sucking lice likely share the similar morphology of mouthpart as chewing lice (Lyal, 1985; Light et al., 2010). During their evolution, sucking lice switched to feed directly on host blood because blood was more nutritious than the organic matters on host skin surface; subsequently, their mouthpart evolved and adapted to the blood-sucking lifestyle (Marshall, 1981; Lehane, 1992; Light et al., 2010).

1.1.4. Effects of lice on their hosts

Louse infestation may cause direct damage to host such as anemia, itching, skin sensitization and other allergic reactions; secondary infection and transmission of pathogens may occur as well (Nelson et al., 1975; Nelson et al., 1977). Effects of chewing lice on poultries and livestock can be

severe. Reduction in appetite, body mass and egg or meat production can all be related to chewing louse infestation (Nelson et al., 1977). For example, chicken head lice (*Cuclotogaster heterographus*) can cause severe restlessness and debility or even death of chicks (Price et al., 2003). Some other lice such as *Columbicola columbae* and *Campanulotes compar*, both parasitizing rock dove (*Columba livia*), can cause reduction of plumage density, leading to lower winter survival (Booth et al., 1993). In some wild birds, such as barn swallow (*Hirundo rustica*), the feather damage caused by chewing lice can make birds less competitive in courtship and even decrease their chance of long-term survival (Brown et al., 1995; Kose et al., 1999).

Compared with chewing lice, sucking lice are more likely to be vectors of pathogens: agents of borreliosis, bartonellosis, brucellosis and salmonellosis have been isolated from sucking lice (Samuel et al., 2001). Human body lice can transmit three pathogenic bacteria: *Rickettsia prowazekii*, *Borrelia recurrentis*, and *Bartonella Quintana* (Raoult & Roux, 1999; Brouqui, 2011; Badiaga & Brouqui, 2012). Rodent lice can transmit pathogenic bacteria such as *Haemobartonella muris* (Crystal, 1959) and *Eperythrozoon coccoides* (Berkenkamp & Wescott, 1988). The hog lice (*Haematopinus suis*) can act as vectors of swinepox virus (Bruner, 1963). Infestation of sucking lice can also cause severe diseases (Samuel et al., 2001), e.g. *Linognathus africanus* can cause severe alopecia, depression and anemia in deer (Brunetti et al., 1971; Foreyt et al., 1986).

1.1.5. Host defence

Birds and mammals have various ways to remove lice from their body and avoid louse infestation. Covering their body with dusts, sunning, choosing lice-free individuals as mates are some of their strategies to get rid of lice (Price et al., 2003). The most important methods to clean up lice are oral grooming and scratching. Birds and mammals not only groom themselves but also groom each other. The behavior of grooming between individuals is called “allogrooming”, which plays an important role in controlling louse burden (Bell & Clifford, 1964; Villa et al., 2016). Periodic molting of feathers and hair (Murray, 1957), melanin in feathers (Bonser, 1995; Kose et al., 1999), neurotoxin on feathers (Dumbacher, 1999), and the immune response of hosts (James, 1999) are also effective strategies in controlling parasitic lice.

1.1.6. Host specificity and co-speciation

In each parasite-host system, the phylogeny of parasite and their host can be similar when the parasite-host association meets the following conditions: 1) the parasite is specific to a single host linage; and 2) the parasite speciates with its host and never extinct independent of its host (Page et al., 1998). However, independent speciation, host switch and extinction are also not rare in the evolution of parasites and they may cause difficulty in unravelling the host-parasite association (Page et al., 1998; Reed et al., 2007).

Like other permanent parasites, lice are commonly highly host specific. The divergence time of lice is often the same as their hosts (Page et al., 1998). Thus, lice were considered a model system for analyzing the process of co-speciation (Sweet & P. Johnson, 2016). The entire lifecycle of parasitic lice is completed on their host, which makes the cladogenesis of lice strongly related to that of their hosts (Page et al., 1998). It is also suggested that the phylogeny of lice can be used as a guide in demonstrating the phylogeny of their hosts (Hopkins, 1942). Hopkins (1942) argued that flamingos and waterfowl were closely related because they shared four genera of lice with each other. However, subsequent studies pointed out that there were multiple host switches of parasitic lice from waterfowl to flamingos (Ahlquist, 1990).

Comparison of louse phylogenies with their host phylogenies has been conducted on several species. The co-speciation of gopher lice with their hosts has been well demonstrated (Hafner et al., 1994). Thus, gopher louse (*Geomydoecus:* Trichodectidae) was commonly chosen as a model system in studying co-speciation (Johnson et al., 2001). Leonardi et al. (2019) also concluded the co-speciation of seal lice with their hosts by reconstructing phylogenetic tree of six species of seal lice (using eight mt protein-cding genes) and conpairing the evolution histories of these seal lice with their hosts. However, there are exceptions as well. Although swiftlet lice (*Dennyus:* Menoponidae) exhibit significant degree of co-speciation with their hosts, the host-parasite association of this species is complicated and both host-switch and independent speciation likely occurred (Clayton et al., 1996). Avian lice in the genus *Brueelia* are relatively host-specific but the co-speciation with their hosts was not supported by phylogenetic analysis of these lice and their hosts (Price et al., 2003).

1.1.7. Host switch

Host switch was also observed in some species of parasitic lice (Brunetti et al., 1971; Westrom et al., 1976; Foreyt et al., 1986; Reed et al., 2007), e.g., host switch of gorilla louse to humans (Reed et al., 2007) and host switch of *Heterodoxus spiniger* from wallabies to dingoes, dogs and other hosts (Thompson, 1939). The host switch of gorilla louse to humans was revealed by cophylogenetic analysis of primates and their lice (Reed et al., 2007). Phylogenetic tree of primate lice based on combined *cox1* and *EF-Iα* data supported that human pubic louse was switched from gorillas and this host switch occurred around 3-4 MYA (Reed et al., 2007). *Heterodoxus spiniger* was found on domestic dogs (Thompson, 1939). However, the typical and more common host of this louse was the agile wallabies (*Macropus agilis*) (Thompson, 1939; Kim & Ludwig, 1978; Koungoulos & Contos, 2019). Although Thompson (1939) suggested that *H. spiniger* on dogs may be stragglers (non-permanent parasites), this hypothesis has been rejected because of the wide spread of *H. spiniger* all over the world. Host switch occurred also in another wallaby louse, *H. octoseriantus* (Barker, 1991). *Heterodoxus octoseriantus* initially parasitized the bush-tailed rock-wallaby, *Petrogale penicillata*, but switched to the Hertbert's rock-wallaby, *Petrogale herberti*, when these two species share the same habitat (Barker, 1991).

1.2. Mitochondrial genomes of parasitic lice

Mitochondrion is an organelle that plays critical roles in cell function (John, 2012). The main function of mitochondria is to synthesize adenosine triphosphate (ATP) as energy source consumed in every metabolic process in vivo (Chinnery & Schon, 2003). Studies of mitochondrial DNA (mtDNA) started in early 1960s (Chinnery & Schon, 2003). The genetic system in mitochondria is independent from nuclear genome; the "Serial Endosymbiotic Theory" hypothesized that mitochondria originated from the symbiotic relationship between bacteria and ancestral eukaryotic cells (Ernster & Schatz, 1981). The gene content of mitochondrial (mt) genome is stable in bilaterian animals; there are 37 genes in total: 13 protein coding genes (PCGs); 2 ribosomal RNA (rRNA) genes and 22 transfer RNA (tRNA) genes (Ernster & Schatz, 1981; Chinnery & Schon, 2003). In parasitic lice, different types of mt genome organization have been observed in species of different parvorders (Shao et al., 2001; Shao et al., 2009; Cameron et al., 2011; Shao et al., 2012): the typical mitochondrial genomes were observed in Amblycera and Ischnocera whereas atypical mt genomes were observed in Anoplura, Ischnocera and Rhynchophthirina.

1.2.1. Typical mitochondrial genomes

The typical organization of animal mt genome is a single circular chromosome containing 15,000-20,000 base pairs of DNA (Boore, 1999, Lavrov and Bell, 2014). Before fragmented mt genome in parasitic lice was discovered, the mt genomes of all studied insects were of the typical organization. The arrangement of genes in most insect mt genomes was also similar (Shao et al., 2001). The most common mt gene arrangement of insects was inferred to be the ancestral to insects (Fig 1.1).

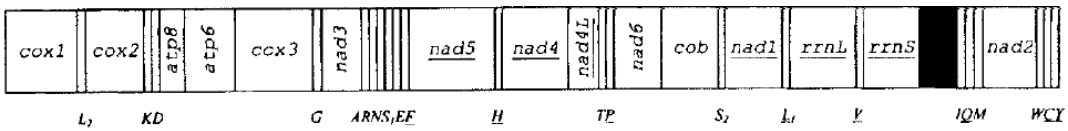


Fig 1.1. Ancestral mitochondrial gene arrangement of insects. Genes underlined are on the same strand while other genes are on the opposite strand (figure from Shao et al., 2001).

The complete mt genome of the wallaby louse (*H. macropus*) was sequenced in 2001 (Shao et al., 2001). It was the first species of parasitic lice that a complete mt genome was sequenced. This mt genome has numerous gene rearrangement relative to the inferred ancestral mt gene arrangement of insects; however, it still retains the typical single-chromosome mt genome organization (Shao et al., 2001). To date, 17 more species of chewing lice have been sequenced for complete mt genomes (Cameron et al., 2011; Song et al., 2019; Sweet et al., 2020; Sweet et al., 2021). Ten of the 17 chewing lice retained the typical single-chromosomal mt genome organization and have gene rearrangement relative to the inferred ancestral mt gene arrangement of insects (Cameron et al., 2011; Song et al., 2019; Sweet et al., 2021). Two of these species (*Campanulotes compar* and *Coloceras* sp.) share similar gene arrangement (only the location of *trnQ* is different); the other eight species (*Colpocephalum griffoneae*, *Amyrsidea minuta*, *Ibidoecus bisignatus*, *Bothriometopus macrocnemis*, *Falcolipeurus quadripustulatus*, *Heterodoxus spiniger*, *Osborniella crotophagae* and *Ricinus* sp.) have different gene arrangement from each other (Cameron et al., 2011; Song et al., 2019; Sweet et al., 2021).

1.2.2. Atypical mitochondrial genomes

Atypical mt genome organization has been observed in several animals such as nematodes (Gibson et al., 2007), sponges (Lavrov et al., 2013), jellyfish (Smith et al., 2011) and insects (Shao et al., 2009; Shao et al 2012; Wei et al., 2012; Jiang et al., 2013; Chen et al., 2014a; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Perlman et al., 2015; Shao et al., 2017; Feng et al., 2018;

Feng et al., 2019; Song et al., 2019; Sweet et al., 2020; Sweet et al., 2021). For example, the mt genome of winged box jellyfish (*Alatina moseri*) was fragmented into eight linear chromosomes ranging from 2.9 to 4.6 kb (Smith et al., 2011).

Among parasitic lice, fragmented mt genome was first discovered in the human body louse (*Pediculus humanus*) (Shao et al., 2009). The mt genome of *P. humanus* has 20 minichromosomes; each minichromosome is 3-4 kb in size and has 1-3 genes (Shao et al., 2009). The human head louse (*Pediculus capitis*) has the same type of fragmented mt genome as the human body louse (Shao et al. 2012). The human pubic louse (*Pthirus pubis*) has 14 mitochondrial minichromosomes and a different gene arrangement from the human head and body louse (Shao et al., 2012). Fragmented mt genomes were also found in five species of rodent and shrew lice (*Polyplax asiatica*, *Polyplax spinulosa*, *Polyplax reclinata*, *Hoplopleura akanezumi* and *Hoplopleura kitii*); these mt genomes have 11 minichromosomes; the gene content of each minichromosome varies among these species (Dong et al., 2014a; Dong et al., 2014b; Dong et al., 2021). The mt genomes of the pig lice (*Haematopinus suis* and *Haematopinus apri*) and the horse louse (*Haematopinus asini*) have nine minichromosomes each; the two pig lice have the same mt gene arrangement to each other whereas the horse louse differs from the pig lice in mt gene arrangement (Jiang et al., 2013; Song et al., 2014). The mt genome of guanaco louse (*Microthoracius praelongiceps*) contains 12 minichromosomes (Shao et al., 2017). The inferred ancestral mt genome organization (i.e. mt karyotype) of sucking lice has 11 minichromosomes (Fig. 1.2) (Shao et al., 2017). During the mt genome evolution of sucking lice, splits and mergers of minichromosomes occurred unevenly in different species: split of minichromosomes occurred several times in primate lice and rodent lice while merger of minichromosome only occurred in pig lice and horse lice (Shao et al., 2017).

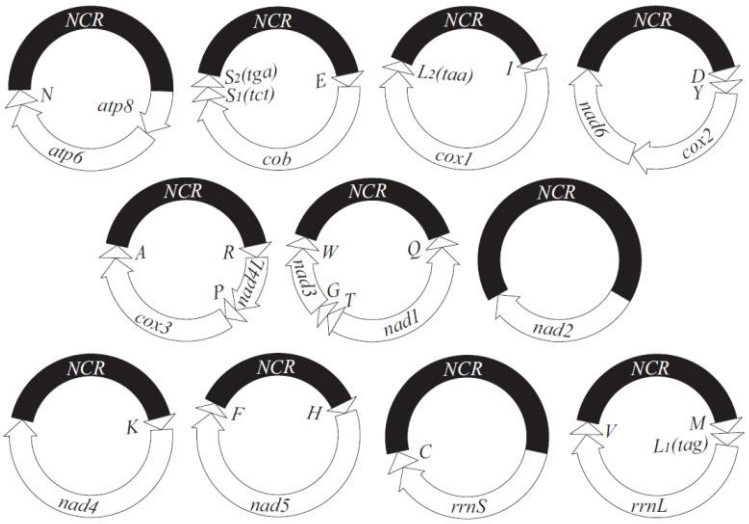


Fig 1.2. The inferred ancestral mitochondrial karyotype of sucking lice (Shao et al., 2017)

Mitochondrial genome fragmentation is also observed in chewing lice. The mt genome of the elephant louse (*Haematomyzus elephantis*) has 10 minichromosomes; nine of these minichromosomes have different gene content and gene arrangement from that of the sucking lice (Shao et al., 2015). Four species in the family Trichodectidae also have fragmented mt genomes (Song et al., 2019). The cattle louse (*Bovicola bovis*) and the sheep louse (*Bovicola ovis*) have 12 mt minichromosomes while the goat louse (*Bovicola caprae*) and the dog louse (*Trichodectes canis*) have 13 mt minichromosomes. The mt karyotype of the cattle louse and sheep louse are the same whereas that of the other two species differ from each other in mt karyotype. Sweet et al. (2020) sequenced the mt genomes of three pigeon lice (family Philopteridae): *Columbicola macrourae*, *Columbicola columbae* and *Columbicola passerinae*; all of them have fragmented mt genomes. *Columbicola macrourae* has 15 minichromosomes; *Columbicola columbae* has 16 minichromosomes. *Columbicola passerinae* have two types of mt genome organization: both types have 17 minichromosomes but only 12 minichromosomes are common between the two types (Sweet et al., 2020). Sweet et al. (2021) assembled the mt genomes of seven Amblycera lice using Illumina sequence data from Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra/) and showed that four of these Amblycera lice have fragmented mt genomes: 1) Australian hobby louse, *Laemobothrion tinnunculi*, has three minichromosomes; 2) citrine warbler louse, *Myrsidea* sp., has three minichromosomes; 3) Incan shrew opossum louse, *Cummingsia maculate*, has three minichromosomes; and 4) Lowland paca louse, *Macrogyropus costalimai*, has seven minichromosomes.

Fragmented mt genome was also observed in booklice. Four species of booklice (*Liposcelis bostrychophila*, *Liposcelis entomophila*, *Liposcelis paeta* and *Liposcelis bostrychophila*) have fragmented mt genomes (Wei et al., 2012; Chen et al., 2014a; Perlman et al., 2015; Feng et al., 2018; Feng et al., 2019). *Liposcelis bostrychophila* contains two circular mt chromosomes that are similar in both size and the number of genes: one of the chromosomes is 8,530 bp with 24 genes while the other one is 7,933 bp with 18 genes (Wei et al., 2012). However, another two species of booklice (*Liposcelis decolor* and *Liposcelis sculptilis*) have the typical mt genome organization (Chen et al., 2014b; Shi et al., 2016).

1.3. Phylogeny of parasitic lice (infraorder Phthiraptera)

The phylogeny of parasitic lice is not fully resolved yet. Both morphological features and molecular data have been used to establish the higher-level phylogeny of parasitic lice (Kim & Ludwig, 1978; Kim & Ludwig, 1982; Lyal, 1985; Barker, 1994; Yoshizawa & Johnson, 2003, 2010; Johnson et al., 2018; Song et al., 2019). The morphology-based phylogeny among the four suborders of parasitic lice (traditional classification) established by Lyal (1985) has been challenged recently by molecular studies (Johnson et al., 2018; Song et al., 2019; De Moya et al., 2021).

Initially, parasitic lice were divided into two orders: Mallophaga (chewing lice) and Anoplura (sucking lice) (Ferris, 1951; Kim & Ludwig, 1982). However, Clay (1970) and Lyal (1985) rejected the order Mallophaga because Mallophaga was a paraphyletic group. Lyal (1985) studied the morphology of parasitic lice in details and classified the parasitic lice into four suborders: Amblycera, Anoplura, Ischnocera and Rhynchophthirina. Lyal (1985) also proposed that: 1) Rhynchophthirina is the sister group of Anoplura; 2) Ischnocera is sister to Rhynchophthirina + Anoplura; and 3) Amblycera is sister to the other three suborders as a group. This suborder-level phylogeny has been widely accepted since Lyal (1985). However, among these four suborders of parasitic lice, only three of them (Amblycera, Anoplura and Rhynchophthirina) were supported to be monophyletic; the monophyly of Ischnocera was not supported (Lyal, 1985).

Cruickshank (2001) sequenced partial *EF1α* gene of 108 species of parasitic lice and showed that the suborder Amblycera was monophyletic but the suborder Ischnocera was paraphyletic. Partial sequences of *cox1* and *18S* genes were also used in reconstructing the phylogeny of parasitic lice (Johnson & Whiting, 2002; Barker et al., 2003; Yoshizawa & Johnson, 2003, 2010; Johnson et al.,

2018). All of these studies support the monophyly of Amblycera, Anoplura and Rhynchophthirina; the monophyly of Ischnocera was supported only in two of these studies (Johnson & Whiting, 2002; Barker et al., 2003) but was rejected in other studies. Single gene analysis provided conflicting results of the monophyly of Ischnocera. *18S* tree supported the monophyly of Ischnocera while the tree based on *EF1α* and *cox1* rejected the monophyly of Ischnocera (Johnson & Whiting, 2002). Combined phylogenetic analysis of *18S + EF1α, 18S + cox1,* and *EF1α + cox1* provided different results on the monophyly of Ischnocera. The monophyly of Ischnocera was supported when *18S* was included in the combined analysis (i.e. *18S + EF1α* and *18S + cox1*) but was rejected when *18S* is not included (*EF1α + cox1*) (Johnson & Whiting, 2002).

Song et al. (2019) reconstructed the phylogeny based on 25 species of parasitic lice: three species from Amblycera; four species from the family Trichodectidae of Ischnocera (traditional classification), five species from the family Philopteridae of Ischnocera, the elephant louse (*Haematomyzus elephantis*) and 12 species from Anoplura. Song et al. (2019) showed that: 1) as a suborder, Ischnocera was paraphyletic; 2) Trichodectidae species were more closely related to Anoplura and Rhynchophthirina than to Philopteridae species (Ischnocera); and 3) all species of eutherian mammal lice (Anoplura, Rhynchophthirina and Trichodectidae) have fragmented mt genomes. Thus, Song et al. (2019) suggested that the family Trichodectidae should be moved out of the suborder Ischnocera and a new suborder named Trichodectera to be created to accommodate the family Trichodectidae.

De Moya et al. (2021) reconstructed phylogenetic tree of over 100 species of Psocodea insects, including 50 speices of parasitic lice, and showed that: 1) the parvorders Anoplura, Rhynchophthirina and Trichodectera formed a monophyletic clade parallel to the parvorder Ischnocera; 2) the parvorder Amblycera is a monophyletic and parallel to the monophyletic clade formed by Anoplura, Ischnocera, Rhynchophthirina and Trichodectera. Based on their phylogenetic analysis, De Moya et al. (2021) proposed a new classification of parasitic lice: infraorder Phthiraptera with five parvorders: Amblycera, Anoplura, Ischnocera, Rhynchophthirina and Trichodectera (Fig. 1.3). This thesis adopts the new classification of De Moya et al. (2021) unless the traditional classification is specifically referred to.

Newly proposed classification:			
Order	Suborder	Infraorder	Parvorder
Psocodea:			
	Trogiomorpha:		
		Prionoglaridetae	
		Psyllipsocetae	
		Atropetae	
	Psocomorpha:		
		Archipsocetae	
		Philotarsetae	
		Epipsocetae	
		Psocetae	
		Caeciliusetae	
		Homilopsocidea	
	Troctomorpha:		
		Amphientometae	
		Sphaeropsocetae	
		Pachytroctetae	
		Liposcelidetae	
		Phthiraptera:	
			Amblycera
			Anoplura
			Rhynchophthirina
			Trichodectera
			Ischnocera

Fig 1.3. Newly proposed classification of Phthiraptera (De Moya et al., 2021)

1.4. Summary

In my PhD project, I intended to study mt genome evolution in a wider range of parasitic lice than that in the previously studies. My aims were to investigate: 1) mt genome evolution in understudied families/genera of sucking lice (Anoplura); 2) how fragmented mt genomes evolved in parasitic lice; 3) how often mt genome fragmentation occurred in bird lice in the parvorder Amblycera; and 4) whether minichromosomal characters derived from mt genome fragmentation are informative in resolving phylogenetic relationships among parasitic lice.

1.5. References

Ahlquist, J. E. (1990). Phylogeny and classification of birds: a study in molecular evolution. ***Yale University Press***.

Badiaga, S., & Brouqui, P. (2012). Human louse-transmitted infectious diseases. ***Clinical Microbiology and Infection***, 18(4): 332-337.

Barker, S. C. (1994). Phylogeny and classification, origins, and evolution of host associations of lice. ***International Journal for Parasitology***, 24(8): 1285-1291.

Barker, S. C., Whiting, M., Johnson, K. P., & Murrell, A. (2003). Phylogeny of the lice (Insecta, Phthiraptera) inferred from small subunit rRNA. ***Zoologica Scripta***, 32(5): 407-414.

Barker, S. J. S. p. (1991). Phylogeny of the *Heterodoxus octoseriatus* group (Phthiraptera: Boopiidae) from rock-wallabies (Marsupialia: Petrogale). ***Systematic Parasitology***, 19(1): 17-24.

Bell, J. F., & Clifford, C. (1964). Effects of limb disability on lousiness in mice. II. Intersex grooming relationships. ***Experimental Parasitology***, 15(4): 340-349.

Berkenkamp, S., & Wescott, R. J. L. a. s. (1988). Arthropod transmission of Eperythrozoon coccoides in mice. ***Laboratory Animal Science***, 38(4): 398-401.

Bonser, R. H. C. (1995). Melanin and the Abrasion Resistance of Feathers. ***The Condor***, 97(2): 590-591.

Booth, D. T., Clayton, D. H., & Block, B. A. (1993). Experimental demonstration of the energetic cost of parasitism in free-ranging hosts. ***Proceedings of the Royal Society of London. Series B: Biological Sciences***, 253(1337): 125-129.

Brouqui, P. (2011). Arthropod-borne diseases associated with political and social disorder. ***Annual Review of Entomology***, 56: 357-374.

Brown, C. R., Brown, M. B., & Rannala, B. (1995). Ectoparasites reduce long-term survival of their avian host. ***Proceedings of the Royal Society of London. Series B: Biological Sciences***, 262(1365): 313-319.

Bruner, D. (1963). The pox diseases of man and animals. ***Diseases transmitted from animals to man***, 386-400.

Brunetti, O., & Cribbs, H. (1971). California deer deaths due to massive infestation by the louse (*Linognathus africanus*). ***California fish and game***.

Cameron, S. L., Yoshizawa, K., Mizukoshi, A., Whiting, M. F., & Johnson, K. P. (2011). Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). ***BMC Genomics***, 12(1): 394.

Chen, S. C., Wei, D. D., Shao, R., Shi, J. X., Dou, W., & Wang, J. J. (2014a). Evolution of multipartite mitochondrial genomes in the booklice of the genus *Liposcelis* (Psocoptera). ***BMC Genomics***, 15(1): 861.

Chen, S. C., Wei, D. D., Shao, R., Dou, W., & Wang, J. J. (2014b). The complete mitochondrial genome of the booklouse, *Liposcelis decolor*: insights into gene arrangement and genome organization within the genus Liposcelis. ***Plos One***, 9(3).

Chinnery, P. F., & Schon, E. A. (2003). Mitochondria. ***Journal of Neurology, Neurosurgery & Psychiatry***, 74(9): 1188-1199.

Clay, T. (1949). Piercing Mouth-Parts in the Biting Lice (Mallophaga). ***Nature***, 164(4171): 617-617.

Clay, T. (1970). Amblycera (Phthiraptera: Insecta). ***Bulletin of the British Museum (Natural History) Entomology***.

Clay, T. (1971). A new genus and two new species of Boopiidae (Phthiraptera: Amblycera). ***Pacific Insects***, 13(3-4): 519-529.

Clayton, D., Price, R., & Page, R. J. S. E. (1996). Revision of Dennyus (Collodennyus) lice (Phthiraptera: Menoponidae) from swiftlets, with descriptions of new taxa and a comparison of host–parasite relationships. ***Systematic Entomology***, 21(3): 179-204.

Cruickshank, R. H., Johnson, K. P., Smith, V. S., Adams, R. J., Clayton, D. H., & Page, R. D. M. (2001). Phylogenetic Analysis of Partial Sequences of Elongation Factor 1α Identifies Major Groups of Lice (Insecta: Phthiraptera). ***Molecular Phylogenetics and Evolution***, 19(2): 202-215.

Crystal, M. M. (1959). The infective index of the spined rat louse, *Polyplax spinulosa* (Burmeister), in the transmission of *Haemobartonella muris* (Mayer) of rats. ***Journal of Economic Entomology***, 52(3): 543-544.

De Moya, R. S., Yoshizawa, K., Walden, K. K., Sweet, A. D., Dietrich, C. H., & Kevin P, J. (2021). Phylogenomics of parasitic and nonparasitic lice (Insecta: Psocodea): combining sequence data and exploring compositional bias solutions in next generation data sets. ***Systematic Biology***, 70(4), 719-738.

Dong, W. G., Song, S., Jin, D. C., Guo, X. G., & Shao, R. (2014a). 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(1): 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 Genomics***, 15: 751.

Dumbacher, J. P. (1999). Evolution of Toxicity in Pitohuis: I. Effects of Homobatrachotoxin on Chewing Lice (Order Phthiraptera). ***The Auk***, 116(4): 957-963.

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.

Emerson, C. K., & Price, D. R. (1976). Abrocomophagidae (Mallophaga: Amblycera), a New Family from Chile. ***Florida Entomologist***, 425-428.

Ernster, L., & Schatz, G. (1981). Mitochondria: a historical review. ***The Journal of Cell Biology***, 91(3 Pt 2): 227s-255s.

Feng, S., Yang, Q., Li, H., Song, F., Stejskal, V., Opit, G. P., ... & Shao, R. (2018). The highly divergent mitochondrial genomes indicate that the booklouse, *Liposcelis bostrychophila* (Psocoptera: Liposcelididae) is a cryptic species. ***G3: Genes, Genomes, Genetics***, 8(3), 1039-1047.

Feng, S., Li, H., Song, F., Wang, Y., Stejskal, V., Cai, W., & Li, Z. (2019). A novel mitochondrial genome fragmentation pattern in *Liposcelis brunnea*, the type species of the genus *Liposcelis* (Psocodea: Liposcelididae). ***International Journal of Biological Macromolecules***, 132, 1296-1303.

Ferris, G. F. (1951). The sucking lice. ***Memoirs of the Pacific Coast Entomological Society***, 1.

Foreyt, W. J., Rice, D. H., & Kim, K. C. (1986). Pediculosis of mule deer and white-tailed deer fawns in captivity. ***Journal of the American Veterinary Medical Association***, 189(9): 1172-1173.

Gibson, T., Blok, V. C., & Dowton, M. (2007). Sequence and characterization of six mitochondrial subgenomes from *Globodera rostochiensis*: Multipartite structure is conserved among close nematode relatives. ***Journal of Molecular Evolution***, 65(3): 308-315.

Hafner, M. S., Sudman, P. D., Villablanca, F. X., Spradling, T. A., Demastes, J. W., & Nadler, S. A. (1994). Disparate Rates of Molecular Evolution in Cospeciating Hosts and Parasites. ***Science***, 265(5175): 1087-1090.

Hennig, W. (1966). Phylogenetic Systematics. ***University of Illinois Press: Urbana, IL, USA***.

Hopkins, G. H. E. (1942). The Mallophaga as an Aid to the Classification of Birds. ***Ibis***, 84(1): 94-106.

Hopkins, G. H. E. (1949). The Host-Associations of the Lice of Mammals. ***Proceedings of the Zoological Society of London***, 119(2): 387-581.

Ibarra, J. (1993). Lice (Anoplura). ***Medical Insects and Arachnids. Springer, Dordrecht***, 517-528

James, P. J. (1999). Do sheep regulate the size of their mallophagan louse populations? ***International Journal for Parasitology***, 29(6): 869-875.

Jiang, H., Barker, S. C., & Shao, R. (2013). Substantial Variation in the Extent of Mitochondrial Genome Fragmentation among Blood-Sucking Lice of Mammals. ***Genome Biology and Evolution***, 5(7): 1298-1308.

John, J. C. S. (2012). Mitochondrial DNA, mitochondria, disease and stem cells. ***Springer Science & Business Media***.

Johnson, K. P., Drown, D. M., & Clayton, D. H. (2001). A data based parsimony method of cophylogenetic analysis. ***Zoologica Scripta***, 30(2): 79-87.

Johnson, K. P., Nguyen, N.-p., Sweet, A. D., Boyd, B. M., Warnow, T., & Allen, J. M. (2018). Simultaneous radiation of bird and mammal lice following the K-Pg boundary. ***Biology Letters***, 14(5): 20180141.

Johnson, K. P., & Whiting, M. F. (2002). Multiple Genes and the Monophyly of Ischnocera (Insecta: Phthiraptera). ***Molecular Phylogenetics and Evolution***, 22(1): 101-110.

Kim, K. (1988). Evolutionary parallelism in Anoplura and eutherian mammals.

Kim, K. C., & Ludwig, H. W. (1978). The family classification of the Anoplura. ***Systematic Entomology***, 3(3): 249-284.

Kim, K. C., & Ludwig, H. W. (1982). Parallel evolution, cladistics, and classification of parasitic Psocodea. ***Annals of the Entomological Society of America***, 75(5): 537-548.

Kose, M., Mand, R., & Moller, A. P. (1999). Sexual selection for white tail spots in the barn swallow in relation to habitat choice by feather lice. ***Animal Behavior***, 58(6): 1201-1205.

Koungoulos, L., & Contos, P. (2019). Global Expansion of the Australian Biting Louse *Heterodoxus spiniger* Facilitated by Human Transport of Dog (*Canis familiaris*), and Implications for Prehistoric Cultural Interaction in Australasia. ***Environmental Archaeology***, 1-14.

Lavrov, D. V., et al. (2013). Mitochondrial DNA of *Clathrina clathrus* (Calcarea, Calcinea): six linear chromosomes, fragmented rRNAs, tRNA editing, and a novel genetic code. ***Molecular Biology and Evolution***, 30(4): 865-880.

Lavrov, D. V., & Bell, E. (2014). Mitochondrial genomes in invertebrate animals. Molecular life sciences: an encyclopedic reference. New York (NY): Springer New York, 1-8

Lehane, M. (1992). Biology of Blood-sucking insects. ***Parasitology Today***, 8(6): 215.

Leonardi, M. S., Virrueta Herrera, S., Sweet, A., Negrete, J., & Johnson, K. P. (2019). Phylogenomic analysis of seal lice reveals codivergence with their hosts. ***Systematic Entomology***, 44(4): 699-708.

Light, J. E., & Hafner, M. S. (2007). Phylogenetics and host associations of Fahrenholzia sucking lice (Phthiraptera: Anoplura). ***Systematic Entomology***, 32(2): 359-370.

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 Evolutionary Biology***, 10: 292.

Lyal, C. H. C. (1985). Phylogeny and Classification of the Psocodea, with Particular Reference to the Lice (Psocodea, Phthiraptera). ***Systematic Entomology***, 10(2): 145-165.

Marshall, A. G. (1981). ***The ecology of ectoparasitic insects***, Academic Press, London, 459 pp.

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., ... & Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. ***Science***, 346(6210), 763-767.

Murray, M. (1957). The distribution of the eggs of Mammalian Lice on their hosts. 2. Analysis of the Oviposition behaviour of *Damalinia ovis* (L). ***Australian Journal of Zoology***, 5(1), 19-29.

Nelson, B. C., & Murray, M. D. (1971). The distribution of Mallophaga on the domestic pigeon (*Columba livia*). ***International Journal for Parasitology***, 1(1): 21-29.

Nelson, W., Keirans, J. E., Bell, J., & Clifford, C. J. (1975). Host-ectoparasite relationships. ***Journal of Medical Entomology***, 12(2): 143-166.

Nelson, W. A., Bell, J. F., Clifford, C. M., & Keirans, J. E. (1977). Interaction of ectoparasites and their hosts. ***Journal of Medical Entomology***, 13(4-5): 389-428.

Page, R. D., Lee, P. L., Becher, S. A., Griffiths, R., & Clayton, D. H. (1998). A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. ***Molecular Phylogenetics and Evolution***, 9(2): 276-293.

Perlman, S. J., Hodson, C. N., Hamilton, P. T., Opit, G. P., & Gowen, B. E. (2015). Maternal transmission, sex ratio distortion, and mitochondria. ***Proceedings of the National Academy of Sciences***, 112(33), 10162-10168.

Price, R. D., Clayton, D., Hellenthal, R. A., Johnson, K. P., & Palma, R. (2003). The chewing lice: world checklist and biological overview; ***Illinois Natural History Survey Special Publication***, 24: 1-448.

Ramli, R., Cusack, M., Curry, G. B., & Furness, R. W. (2008). Morphological variation of chewing lice (Insecta: Phthiraptera) from different skua taxa. ***Biological Journal of the Linnean Society***, 71(1): 91-101.

Raoult, D., & Roux, V. (1999). The body louse as a vector of reemerging human diseases. ***Clinical Infectious Diseases***, 29(4): 888-911.

Reed, D. L., Light, J. E., Allen, J. M., & Kirchman, J. J. (2007). Pair of lice lost or parasites regained: the evolutionary history of anthropoid primate lice. ***BMC Biology***, 5(1), 7.

Richards, O., & Davies, R. (1977). Mallophaga (biting lice or bird lice). In ***Imms' General Textbook of Entomology*** (pp. 658-6690. Springer. Dordrecht.

Rothschild, M., & Clay, T. (1957). ***Fleas, flukes and cuckoos. A study of bird parasites***.

Samuel, W. M., Pybus, M. J., & Kocan, A. A. (2001). ***Parasitic diseases of wild mammals*** (No. Ed. 2). Iowa state university press.

Shao, R., Barker, S. C., Li, H., Song, S., Poudel, S., & Su, Y. (2015). Fragmented mitochondrial genomes in two suborders of parasitic lice of eutherian mammals (Anoplura and Rhynchophthirina, Insecta). ***Scientific Reports***, 5: 17389.

Shao, R., Campbell, N. J., & Barker, S. C. (2001). Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). ***Molecular Biology and Evolution***, 18(5): 858-865.

Shao, R., Kirkness, E. F., & Barker, S. C. (2009). The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. ***Genome Research***, 19(5): 904-912.

Shao, R., Li, H., Barker, S. C., & Song, S. (2017). The Mitochondrial Genome of the Guanaco Louse, *Microthoracius praelongiceps*: Insights into the Ancestral Mitochondrial Karyotype of Sucking Lice (Anoplura, Insecta). ***Genome Biology and Evolution***, 9(2): 431-445.

Shao, R., Zhu, X.-Q., Barker, S. C., & Herd, K. (2012). Evolution of Extensively Fragmented Mitochondrial Genomes in the Lice of Humans. ***Genome Biology and Evolution***, 4(11): 1088-1101.

Shi, Y., Chu, Q., Wei, D.-D., Qiu, Y. J., Shang, F., Dou, W., & Wang, J. J. (2016). The mitochondrial genome of booklouse, *Liposcelis sculptilis* (Psocoptera: Liposcelididae) and the evolutionary timescale of Liposcelis. ***Scientific Reports***, 6: 30660.

Smith, D. R., Kayal, E., Yanagihara, A. A., Collins, A. G., Pirro, S., & Keeling, P. J. (2011a). First Complete Mitochondrial Genome Sequence from a Box Jellyfish Reveals a Highly Fragmented Linear Architecture and Insights into Telomere Evolution. ***Genome Biology and Evolution***, 4(1): 52-58.

Smith, V. S. (2008). Avian louse phylogeny (Phthiraptera: Ischnocera): a cladistic study based on morphology. ***Zoological Journal of the Linnean Society***, 132(1): 81-144.

Smith, V. S., Ford, T., Johnson, K. P., Johnson, P. C., Yoshizawa, K., & Light, J. E. (2011b). Multiple lineages of lice pass through the K–Pg boundary. ***Biology letters***, 7(5), 782-785.

Snodgrass, R. E. (1944). The feeding apparatus of biting and sucking insects affecting man and animals. ***Smithsonian Miscellaneous Collections***.

Song, F., et al. (2019). Mitochondrial Genome Fragmentation Unites the Parasitic Lice of Eutherian Mammals. ***Systematic Biology***, 68(3): 430-440.

Song, S. D., Barker, S. C., & Shao, R. (2014). Variation in mitochondrial minichromosome composition between blood-sucking lice of the genus *Haematopinus* that infest horses and pigs. ***Parasit Vectors***, 7: 144.

Sweet, A., & P. Johnson, K. (2016). Cophylogenetic analysis of New World ground-doves (Aves: Columbidae) and their parasitic wing lice (Insecta: Phthiraptera: *Columbicola*). ***Molecular Phylogenetics and Evolution***, 103, 122-132.

Sweet, A. D., Johnson, K. P., & Cameron, S. L. J. P. (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., et al. (2021). Structure, gene order, and nucleotide composition of mitochondrial genomes in parasitic lice from Amblycera. ***Gene***, 768: 145312.

Thompson, G. B. (1939). The distribution of *Heterodoxus spiniger* (Enderlein). Papers and Proceedings of the Royal Society of Tasmania, 27-32.

Villa, S. M., Goodman, G. B., Ruff, J. S., & Clayton, D. H. (2016). Does allopreening control avian ectoparasites? ***Biology Letters***, 12(7).

Waage, J. K. (2008). The evolution of insect/vertebrate associations. ***Biological Journal of the Linnean Society***, 12(3): 187-224.

Wei, D. D., Shao, R. F., Yuan, M. L., Dou, W., Barker, S. C., & Wang, J. J. (2012). The Multipartite Mitochondrial Genome of *Liposcelis bostrychophila*: Insights into the Evolution of Mitochondrial Genomes in Bilateral Animals. ***Plos One***, 7(3).

Westrom, D. R., Nelson, B. C., & Connolly, G. E. J. (1976). Transfer of *Bovicola tibialis* (Piaget)(Mallophaga: Trichodectidae) from the introduced fallow deer to the Columbian black-tailed deer in California. ***Journal of Medical Entomology***, 13(2): 169-173.

Yoshizawa, K., & Johnson, K. P. (2003). Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. ***Molecular Phylogenetics and Evolution***, 29(1): 102-114.

Yoshizawa, K., & Johnson, K. P. (2010). How stable is the "Polyphyly of Lice" hypothesis (Insecta: Psocodea): A comparison of phylogenetic signal in multiple genes. ***Molecular Phylogenetics and Evolution***, 55(3): 939-951.

Chapter 2

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

Published as:

Fu, Y. T., **Dong, Y.,** (equal first author) Wang, W., Nie, Y., Liu, G. H., & Shao, R. (2020). Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse *Pedicinus obtusus* and colobus louse *Pedicinus badii. Genomics*, 112(6), 4924-4933.

Manuscript: Appendix I

Contributions to authorship: I retrieved the Illumina sequence data of colobus louse (*Pedicinus badii*) from the Sequence Read Archive (SRA) and assembled its mitochondrial genome. I annotated and submitted the mt genome sequence of colobus louse to NCBI database. I inferred the ancestral mt karyotype of higher primate lice. I also analysed the mt minichromosomal split and merger in macaque louse and colobus louse and prepared 60% of the figures and tables. I wrote the first version of this manuscript. Overall, I contributed at least 50% to this manuscript.

2.1. Abstract

Blood-sucking lice of three genera, *Pedicinus* (family Pedicinidae), *Pediculus* (Pediculidae) and *Pthirus* (Pthiridae), parasitize higher primates (i.e. monkeys, apes and humans) exclusively. Previous studies showed that mitochondrial (mt) genome fragmentation in the *Pediculus* and *Pthirus* lice of humans and chimpanzees was a one-way process with many events of minichromosome split. We report for the first time the fragmented mt genomes of two *Pedicinus* species: the macaque louse, *Pedicinus obtusus*, and the colobus louse, *Pedicinus badii*, and compared them with the lice of humans and chimpanzees. The macaque louse and the colobus louse have 12 and 14 minichromosomes respectively; each minichromosome has a coding region with 1-7 genes and a non-coding region with high sequence similarity across all minichromosomes within a species. Despite being congeneric, the two monkey lice are distinct from each other in mt karyotype (i.e. the number of minichromosomes, the gene content of each mini-chromosome, and the gene arrangement in each minichromosome). 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.

2.2. Introduction

Higher primates (infraorder Semiiformes) comprise Old World monkeys, New World monkeys, apes and humans, with 320 extant species (Wilson & Reeder, 2005). Higher primates originated ~45 million years ago (MYA) in East Asia (Beard et al., 1994; Beard et al., 1996; Gebo et al., 2000) and expanded all over the world with the modern humans, *Homo sapiens*, the most widely distributed species. Blood-sucking lice (parvorder Anoplura) parasitize 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 (Durden & Musser, 1994a). 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* (Pthiridae) with humans and are the hosts of *Pthirus gorilla*. Forty-one species of Old World monkeys (Cercopithecoidea) are the hosts of 14 species of the genus *Pedicinus* (Pedicinidae) (Durden & Musser, 1994b). 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 (Reed et al., 2004; Reed et al., 2007; Light et al., 2010). Lice have been a favored model for co-evolutionary studies for several decades (Hafner & Nadler, 1988; Page, 1996; Reed et al., 2004). Host evolution can help understand louse evolution (Light et al., 2010); louse evolution can also help understand host evolution (Reed et al., 2004).

Mitochondrial (mt) genome information has been used widely in evolutionary studies of animals including parasitic lice (Boore, 1998; Pakendorf & Stoneking, 2005; Cameron, 2014; Song et al., 2019). 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 (Boore, 1999; Shao et al., 2009; Shao et al., 2012; Jiang et al., 2013; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Herd et al., 2015; Shao et al., 2015; Sha et al., 2017; Song et al., 2019; Sweet et al., 2020). 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 (Shao et al., 2009; Shao et al., 2012). At the other end, lice of pigs and horses in the genus *Haematopinus* have nine minichromosomes (Jiang et al., 2013; Song et al., 2014), whereas rodent lice, a guanaco louse and four chewing lice are in the middle with 10-13 minichromosomes (Dong et al., 2014a; Dong et al., 2014b; Shao et al., 2015; Shao et al., 2017; Song et al., 2019) showed that many minichromosome split events occurred in the human lice and the chimpanzee louse whereas minichromosome 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 parasitize Old World monkeys (Durden & Musser, 1994a; Durden & Musser, 1994b), nothing was known prior to this study about mt genomes of *Pedicinus* lice except evidence for a *rrnS*-minichromosome in the langur louse, *Pedicinus ancoratus* (Shao et al., 2009).
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.

2.3. Materials and methods

2.3.1. Collection of macaque lice, DNA extraction, sequencing, sequence assembly, minichromosome verification and mitochondrial genome annotation of the macaque louse, *Pedicinus obtusus*

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 (Ferris, 1934; Ferris, 1951). *Macaca mulatta* is the host of two *Pedicinus* species: *Pedicinus eurygaster* and *Pedicinus obtusus* (Durden & Musser, 1994a). 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) (Shao et al., 2009) were used as the initial references to assemble the Illumina sequence reads with Geneious 11.1.5 (Kearse et al., 2012). 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* (Supplementary Table 2.1, Supplementary Fig. 2.1). 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 (Katoh & Standley, 2016) to help identify genes and gene boundaries. tRNA genes were identified using ARWEN (Laslett & Canbäck, 2008) and tRNAscan-SE (Lowe & Chan, 2016). Protein-coding genes and rRNA genes were identified by BLAST searches of GenBank (Johnson et al., 2008).

2.3.2. 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* (Durden and Musser, 1994a). 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 ~250 bp non-coding region sequences. For *Pthirus pubis*, the conserved non-coding region sequences reported in Shao et al. (2012) were used to search and assemble the minichromosomes that contained the three genes (*nad4*, *trnK*, *trnN*) that were not identified in Shao et al. (2012). Protein-coding genes and rRNA genes were identified with BLAST searches of GenBank (Johnson et al., 2008), and tRNA genes were identified using tRNAscan-SE (Lowe & Chan, 2016) and ARWEN (Laslett & Canbäck, 2008).

2.3.3. 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 2.1). 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 (Talavera & Castresana, 2007). Phylogenetic analyses were conducted using Bayesian inference method (BI) with MrBayes 3.2.6 (Ronquist & Huelsenbeck; 2007) and Maximum likelihood (ML) with PhyML 3.0 (Guindon et al., 2010). 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 2,500 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 (Adascal et al., 2005) based on the Akaike information criterion (AIC). Phylogenetic trees were drawn using Figtree v.1.31 (http://tree.bio.ed.ac.uk/software/figtree).

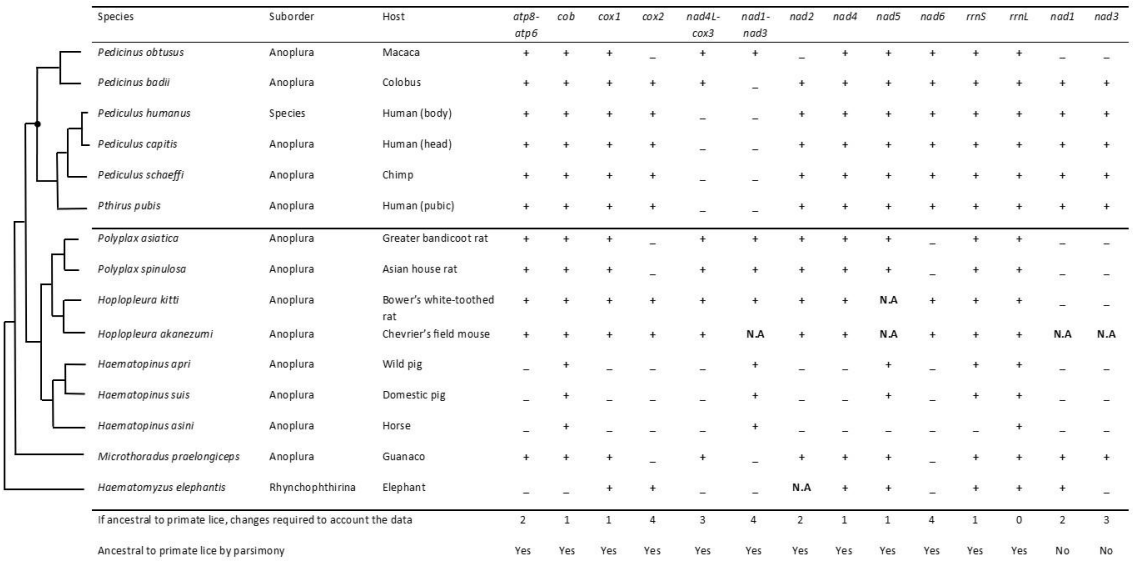
Table 2.1. Species of parasitic lice included in the phylogenetic analyses in this study

Species	Hosts	GenBank accession numbers	References
Pedicinus obtusus	Macaque	MT792495-506	Fu & Dong et al., 2020
Pedicinus badii	Colobus	MT721726-39	Fu & Dong et al., 2020
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; Fu & Dong et al., 2020
Polyplax reclinata	Asian grey shrew	MW291451-MW291461	Dong et al., 2021
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	KJ648923-32	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

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

We used a parsimony method described in Shao et al. (2017) to infer the ancestral mt karyotype of higher primate lice. Three families of sucking lice parasitize 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 (Durden & Musser, 1994a). Each of these families and genera is monophyletic with *Pediculus* and *Pediculus* most closely related and sister to *Pediculus* (Reed et al., 2007; Light et al., 2010). 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. 2.1, Supplementary Fig. 2.2). 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 2.1, Fig. 2.1, Supplementary Fig. 2.2). 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 (Shao et al., 2017), we considered tRNA genes separately from protein-coding and rRNA genes.



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	Species	Human (body)	+	+	+	+	_	_	+	+	+	+	+	+	+	+
Pediculus capitis	Anoplura	Human (head)	+	+	+	+	_	_	+	+	+	+	+	+	+	+
Pediculus schaeffi	Anoplura	Chimp	+	+	+	+	_	_	+	+	+	+	+	+	+	+
Pthirus 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. 2.1. Inferred ancestral mitochondrial karyotype of higher primate lice (Pedicinidae, Pediculidae and Pthiridae). Names and transcription orientation of genes are indicated in the coding region; minichromosomes are in alphabetical order of protein-coding and rRNA gene names.

2.4. Results

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

We obtained a large number of clean sequence reads for *Pedicinus 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 2.2); 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 2,850 to 3,561 bp in size, containing a coding region and a non-coding region (NCR) in a circular organization (Fig. 2.2). The coding regions have 1 to 7 genes each and range from 787 bp to 1,671 bp; the non-coding regions range from 1,551 to 2,177 bp (Table 2.2, Fig. 2.2). 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 *trnL1* (tag) is upstream *rrnL*, in the other type *trnL2* (taa) is upstream *rrnL*. *trnL1* (tag) and *trnL2* (taa) have identical sequences except for the third anti-codon position (Fig. 2.3). In the sequence-read assembly of *trnL1* (tag) and *trnL2* (taa), the third anticodon position was covered 68,396 times in total, of which 41,566 (60.8%) was for *trnL2* (taa) and 26,735 (39.2%) was for *trnL1* (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 (Supplementary Fig. 2.3). The annotated mt genome of Pedicinus obtusus is available in GenBank (Accession numbers MT792495-506).

Table 2.2. Mitochondrial minichromosomes of the macaque louse, *Pedicinus obtusus*

Minichromosome	Size (bp)	Coding region (bp)	Non-coding region (bp)	Number of Illumina sequence-reads[1]
atp8-atp6-C	3,000	907	2,093	690,229
E-cob-S1	3,146	1,182	1,964	1,345,657
1cox1	3,120	1,548	1,572	1,555,266
nad2-Y-cox2-V	3,561	1,799	1,762	1,457,577
R-nad4L-P-cox3-I	3,027	1,261	1,766	1,436,297
T-nad1-Q-N-G-nad3-W	3,524	1,470	2,054	1,304,553
K-nad4	2,919	1,368	1,551	1,792,728
H-nad5	3,406	1,671	1,735	1,084,208
D-A-M-F-nad6	2,864	787	2,077	1,900,118
L1-rrnL	3,009	1,130	1,879	1,684,859[2]

L2-rrnL	3,009	1,130	1,879	1,684,859[2]
rrnS-S2	3,004	827	2,177	1,766,918
Total	37,589	15,080	22,509	17,703,269

[1]Each minichromosome was sequenced individually in full length; [2]*L1-rrnL* minichromosome and *L2-rrnL* minichromosome are from the same sequence-read assembly.

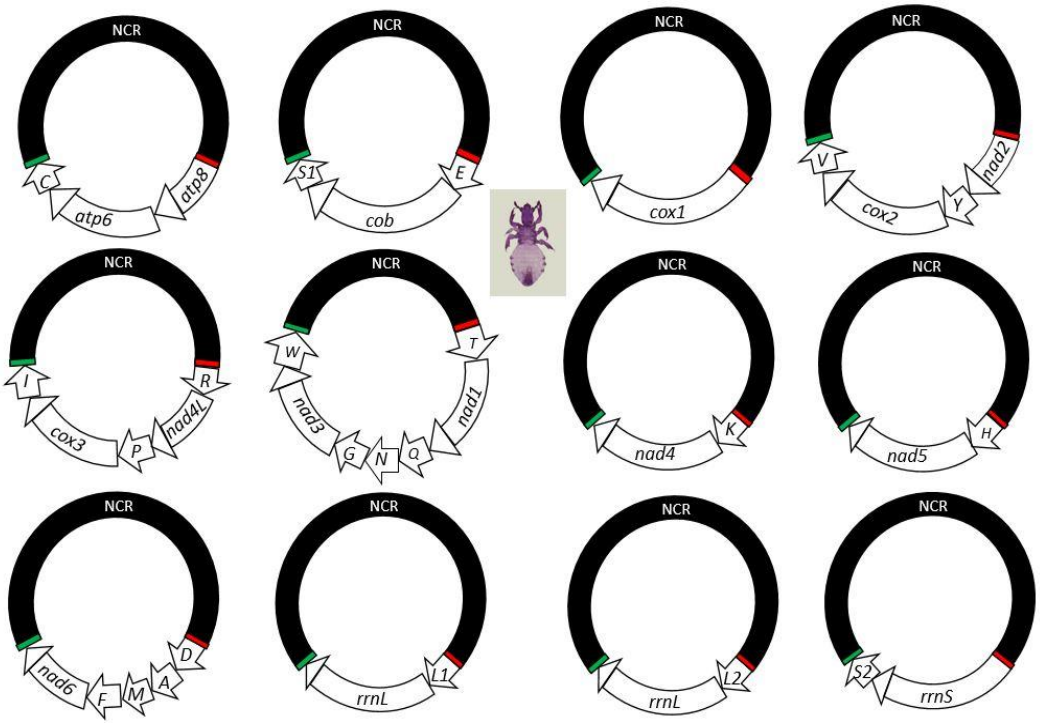


Fig. 2.2. The mitochondrial genome of macaque louse, *Pedicinus obtusus*. Each minichromosome has a coding region and a non-coding region (NCR, in black). Names and transcription orientation of genes are indicated in the coding region; minichromosomes are in alphabetical order of protein-coding and rRNA gene names.

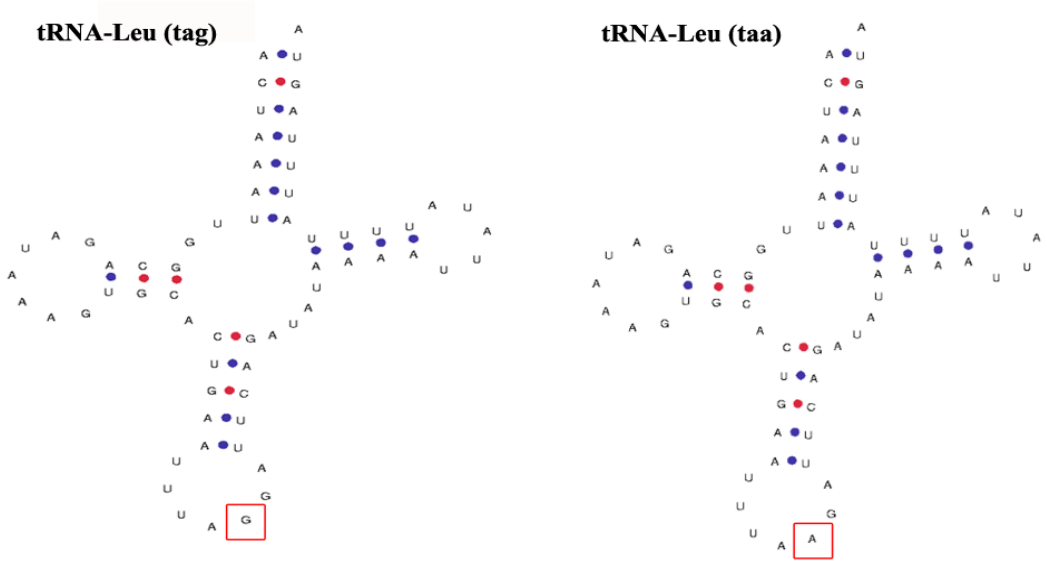


Fig. 2.3. Inferred secondary structures of the mitochondrial tRNA-Leu (tag) and tRNA-Leu (taa) of *Pedicinus obtusus*. The gene for these two tRNAs, *trnL1* (tag) and *trnL2* (taa), have identical sequences except for the third anti-codon position.

2.4.2. 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. 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 (Supplementary Fig. 2.4) – 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. 2.4A). The 14 coding regions range from 74 bp (*trnS2* minichromosome) to 1,664 bp (*trnH-nad5* minichromosome) (Table 2.3). 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, *trnS2*. The other 21 tRNA genes are scattered among 11 coding regions, each coding region with one to five tRNA genes (Fig. 2.4A). 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).

Table 2.3. 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-S1	214	1,226	176	21,695
cox1	214	1,552	177	27,314
Y-cox2	208	734	187	20,672
R-nad4L-P-cox3	212	1,200	190	21,730
T-nad1-Q-N-C	220	1,192	170	20,843
nad2-I	208	1,046	198	22,082
G-nad3-W-D-A-V	215	692	189	20,430
K-nad4	239	1,361	163	23,742
H-nad5	209	1,664	205	22,636
M-F-nad6	209	676	199	19,667
L1- rrnL	207	1,161	174	25,755
L2- rrnS	219	705	212	18,765
S2	206	74	195	12,168
Total	3,001	14,179	2,610	299,810

2.4.3. *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. (2012); 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. 2.4B). 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. (2012). 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 numbers MT721740).

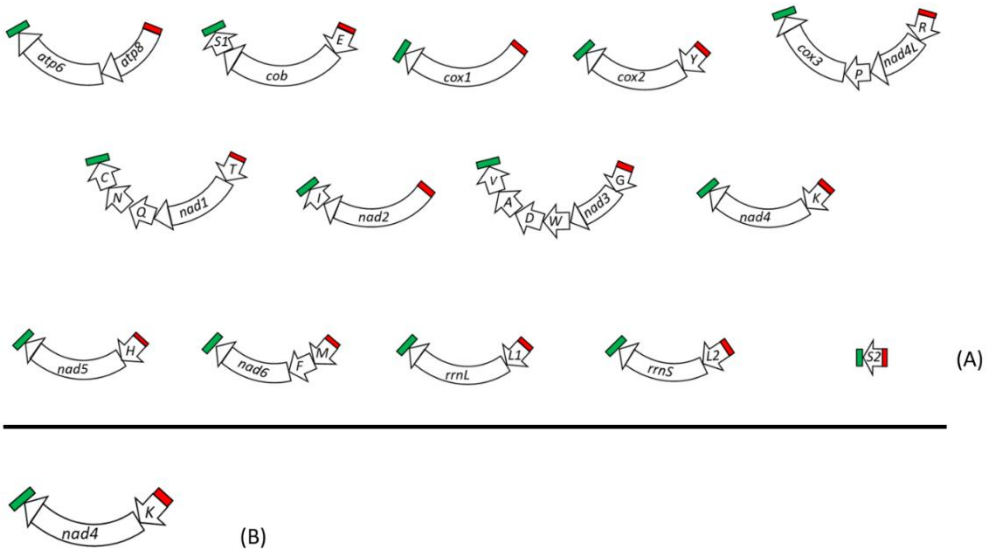


Fig. 2.4. (A) The coding regions of the 14 minichromosomes and conserved AT-rich region (in red) and conserved GC-rich region (in green) of the colobus louse, *Pedicinus badii*; (B) the coding region of *K-nad4* minichromosomes and conserved AT-rich region (in red) and conserved GC-rich region (in green) 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.

2.4.4. 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. 2.5, Supplementary Fig. 2.5). 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. 2.5).

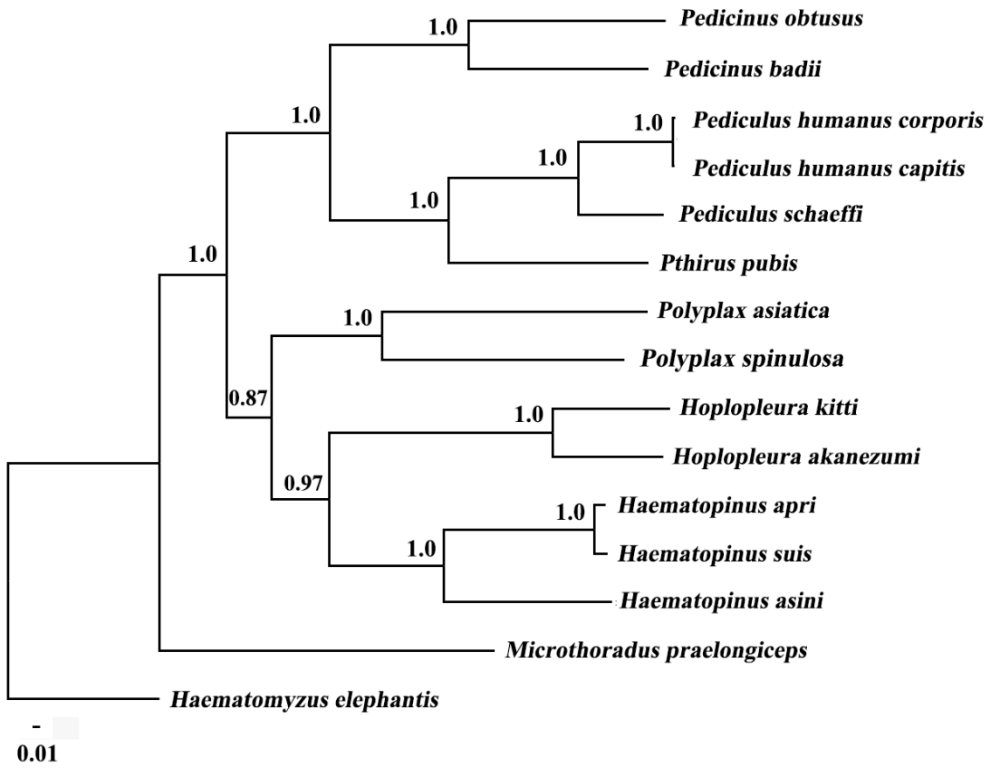


Fig. 2.5. Phylogenetic relationships among 14 species of sucking lice (Anoplura) inferred from Bayesian analysis of deduced amino acid sequences of nine mitochondrial genes. The elephant louse, *Haematomyzus elephantis*, was used as the outgroup. Bayesian posterior probabilities (Bpp) values were indicated at nodes.

2.4.5. Inferred ancestral mitochondrial karyotype of 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 (Herd et al., 2015; Shao et al., 2009; Shao et al., 2012) and *Pthirus pubis* (Shao et al., 2012), in conjunction with the data from other eight species of sucking lice and the elephant louse (Fig. 2.1, Fig. 2.6, Supplementary Fig. 2.2). 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 (Durden & Musser, 1994a). 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. 2.6). 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. 2.6). 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 (Supplementary Fig. 2.2). It is known that tRNA genes are much more mobile than protein-coding and rRNA genes (Shao et al., 2017); 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.

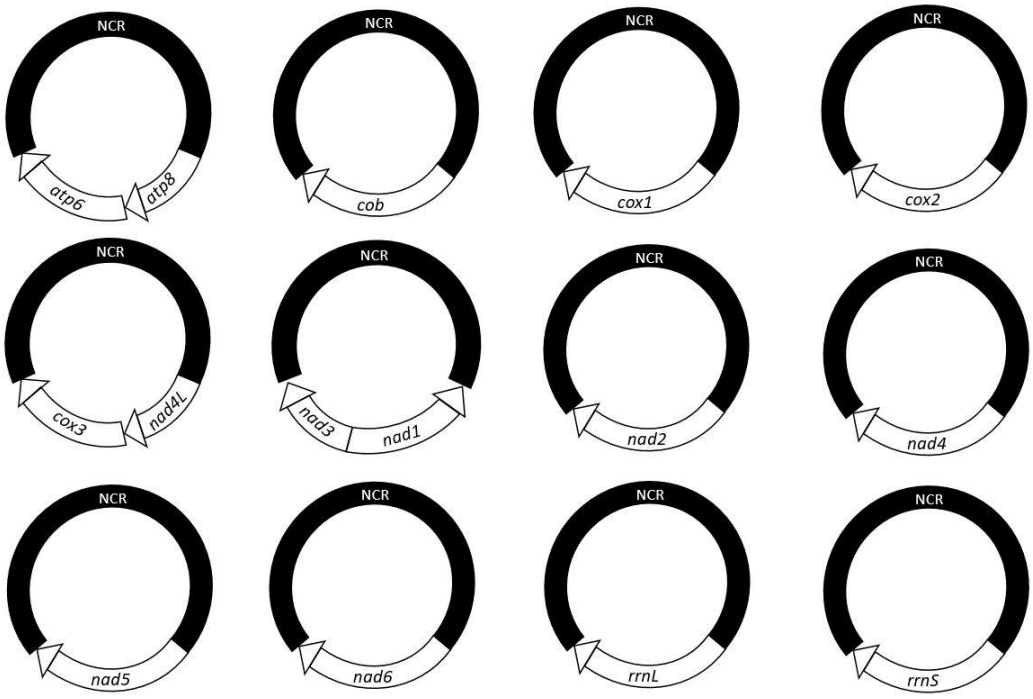


Fig. 2.6. Inferred ancestral mitochondrial karyotype of higher primate lice (Pedicinidae, Pediculidae and Pthiridae). Names and transcription orientation of genes are indicated in the coding region; minichromosomes are in alphabetical order of protein-coding and rRNA gene names.

2.5. Discussion

2.5.1. 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. 2.2) whereas *Pedicinus badii* has 14 minichromosomes (Fig. 2.4A). 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. 2.2, Fig. 2.4A). Likewise, *nad1* and *nad3* are together on the same minichromosome in *Pedicinus obtusus* but are on two separate minichromosomes in Pedicinus badii. *trnS*$_2$ is on the rrnS-minichromosome in Pedicinus obtusus but is on its own minichromosome 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 *trnS*$_2$, six other tRNA genes, *trnA*, *trnC*, *trnD*, *trnI*, *trnL*$_2$ and *trnV*, are also on different minichromosomes between the two *Pedicinus* species (Fig. 2.2, Fig. 2.4A).

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. (2014b) 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 (Dong et al., 2014b). 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: *trnL*$_1$, *trnM*, *trnT* and *trnY* (Dong et al., 2014a). Song et al. (2014) reported that the horse louse, *Haematopinus asini*, has nine minichromosomes, which is the same as the congeneric pig lice, *Haematopinus apri* and *Haematopinus suis* (Jiang et al., 2013). 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* (Song et al., 2014). Herd et al. (2015) reported that the chimpanzee louse, *Pediculus schaeffi*, 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 (Shao et al., 2012).

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 (Shao et al., 2017). 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 (Durden & Musser, 2014b). In general, the mammalian divergence has been studied much more and understood much better than the divergence of their lice (Kumar et al., 2017). 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*) (Kumar et al., 2017) 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 one 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) (Light et al., 2010), 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 (Durden & Musser, 1994a; Durden & Musser, 1994b). 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).

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

Shao et al. (2017) 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* (Shao et al., 2017). 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. 2.7). In the inferred ancestral mt karyotype of higher primate lice, *cox2* and *nad2* are on two separate minichromosomes (Fig. 2.6); this ancestral condition is retained in the colobus louse (Fig. 2.4A). In the macaque louse, *Pedicinus obtusus*, however, these two minichromosomes have merged to form the *nad2-Y-cox2-V* minichromosome (Fig. 2.2, Fig. 2.7A). 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. 2.4A, Fig. 2.7B). The opposite evolutionary directions that these two *Pedicinus* species took explained well the striking variation between them in mt karyotype.

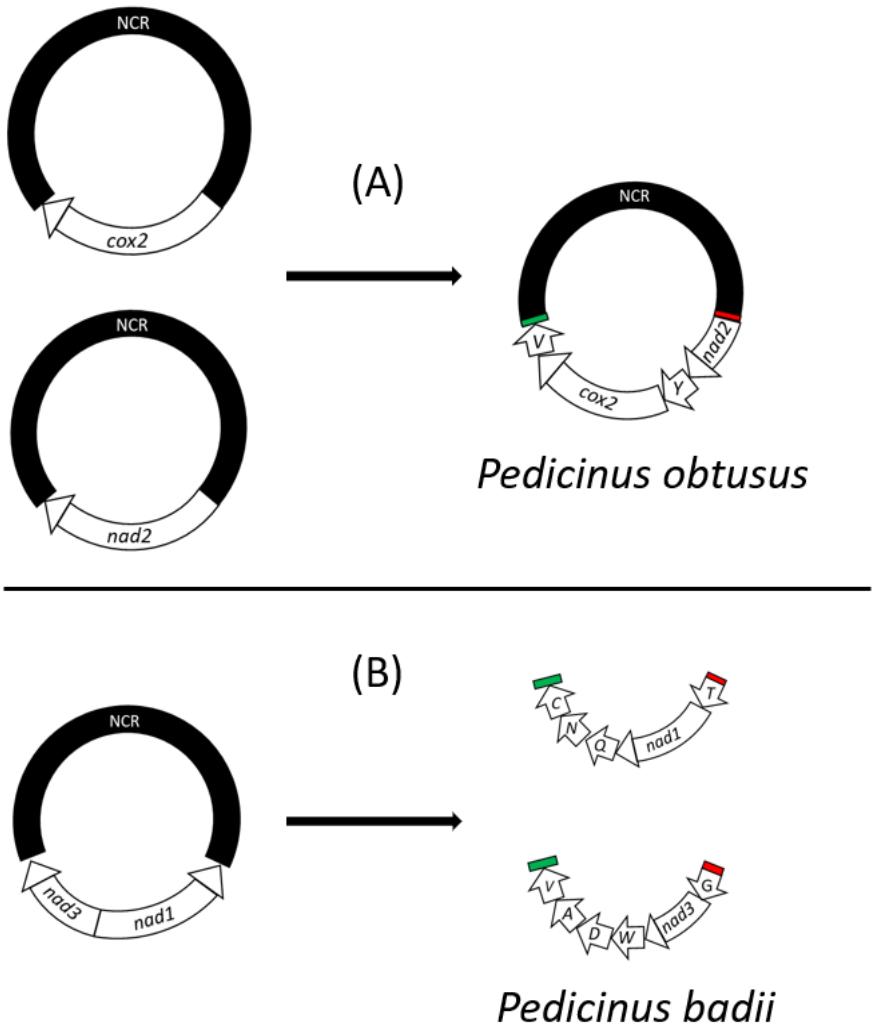


Fig. 2.7. (A) Two ancestral mitochondrial minichromosomes of higher primate lice merged in the macaque louse, *Pedicinus obtusus*; whereas (B) one ancestral mitochondrial minichromosome of higher primate lice split into two in the colobus louse, *Pedicinus badii*. Names and transcription orientation of genes are indicated in the coding region. Non-coding regions are in black.

2.6. Conclusions

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 *Haematopinus*, indicating that minichromosome merger is more common than previously known (Shao et al., 2017). 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.

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.

2.7. References

Abascal, F., Zardoya, R., & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. ***Bioinformatics***, 21(9), 2104-2105.

Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. ***Nucleic Acids Research***, 38(suppl_2), W7-W13.

Beard, K. C., Qi, T., Dawson, M. R., Wang, B., & Li, C. (1994). A diverse new primate fauna from middle Eocene fissure-fillings in southeastern China. ***Nature***, 368(6472), 604-609.

Beard, K. C., Tong, Y., Dawson, M. R., Wang, J., & Huang, X. (1996). Earliest complete dentition of an anthropoid primate from the late middle Eocene of Shanxi Province, China. ***Science***, 272(5258), 82-85.

Boore, J. L., & Brown, W. M. (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. ***Current Opinion in Genetics & Development***, 8(6), 668-674.

Boore, J.L. (1999). Animal mitochondrial genomes. ***Nucleic Acids Research***, 27(8), 1767-1780.

Cameron, S.L. (2014). Insect mitochondrial genomics: implications for evolution and phylogeny. ***Annual Review of Entomology***, 59, 95-117.

Chin T.H. (1999): Taxonomy and fauna of sucking lice (Anoplura) in China. ***Beijing Science Press***.

Durden, L.A., Musser, G.G. (1994a). The mammalian hosts of the sucking lice (Anoplura) of the world: a host-parasite list. ***Bulletin of the Society for Vector Ecology***, 19(2), 130-168.

Durden, L.A., Musser, G.G. (1994b). 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.

Dong, W. G., Song, S., Guo, X. G., Jin, D. C., Yang, Q., Barker, S. C., & Shao, R. (2014a). 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(1), 1-13.

Dong, W. G., Song, S., Jin, D. C., Guo, X. G., & Shao, R. (2014b). 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(1), 1-12.

Ferris, G.B. (1933). Contributions toward a monograph of the sucking lice. Part VII. Stanford University Publications. ***Biological Sciences***, 2(7).

Ferris, G.B. (1951). The sucking lice. ***Memoirs of the Pacific Coast Entomological Society***, 1.

Gebo, D. L., Dagosto, M., Beard, K. C., Qi, T., & Wang, J. (2000). The oldest known anthropoid postcranial fossils and the early evolution of higher primates. ***Nature***, *404*(6775), 276-278.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. ***Systematic Biology***, 59(3), 307-321.

Hafner, M. S., & Nadler, S. A. (1988). Phylogenetic trees support the coevolution of parasites and their hosts. ***Nature***, 332(6161), 258-259.

Herd, K. E., 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 Genomics***, 16(1), 1-8.

Jiang, H., Barker, S. C., & Shao, R. (2013). Substantial variation in the extent of mitochondrial genome fragmentation among blood-sucking lice of mammals. ***Genome Biology and Evolution***, 5(7), 1298-1308.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. ***Nucleic Acids Research***, 36(suppl_2), W5-W9.

Katoh, K., & Standley, D. M. (2016). A simple method to control over-alignment in the MAFFT multiple sequence alignment program. ***Bioinformatics***, 32(13), 1933-1942.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. ***Bioinformatics***, 28(12), 1647-1649.

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. ***Molecular Biology and Evolution***, 34(7), 1812-1819.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., ... & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. ***Bioinformatics***, 23(21), 2947-2948.

Laslett, D., & Canbäck, B. (2008). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. ***Bioinformatics***, 24(2), 172-175.

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 Evolutionary Biology***, 10(1), 1-15.

Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. ***Nucleic Acids Research***, 44(W1), W54-W57.

Page, R. D. (1996). Temporal congruence revisited: comparison of mitochondrial DNA sequence divergence in cospeciating pocket gophers and their chewing lice. ***Systematic Biology***, 45(2), 151-167.

Pakendorf, B., & Stoneking, M. (2005). Mitochondrial DNA and human evolution. ***Annual review of Genomics and Human genetics.*** Genet., 6, 165-183.

Reed, D. L., Smith, V. S., Hammond, S. L., Rogers, A. R., Clayton, D. H., & Barton, N. (2004). Genetic analysis of lice supports direct contact between modern and archaic humans. ***PLoS Biology***, 2(11), e340.

Reed, D. L., Light, J. E., Allen, J. M., & Kirchman, J. J. (2007). Pair of lice lost or parasites regained: the evolutionary history of anthropoid primate lice. ***BMC Biology***, 5(1), 1-11.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. ***Bioinformatics***, 19(12), 1572-1574.

Shao, R., Kirkness, E. F., & Barker, S. C. (2009). The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. ***Genome Research***, 19(5), 904-912..

Shao, R., Zhu, X. Q., Barker, S. C., & Herd, K. (2012). Evolution of extensively fragmented mitochondrial genomes in the lice of humans. ***Genome Biology and Evolution***, 4(11), 1088-1101.

Shao, R., Barker, S. C., Li, H., Song, S., Poudel, S., & Su, Y. (2015). Fragmented mitochondrial genomes in two suborders of parasitic lice of eutherian mammals (Anoplura and Rhynchophthirina, Insecta). ***Scientific Reports***, 5(1), 1-11.

Shao, R., Li, H., Barker, S. C., & Song, S. (2017). The mitochondrial genome of the guanaco louse, *Microthoracius praelongiceps*: insights into the ancestral mitochondrial karyotype of sucking lice (Anoplura, Insecta). ***Genome Biology and Evolution***, 9(2), 431-445.

Song, F., Li, H., Liu, G. H., Wang, W., James, P., Colwell, D. D., ... & Shao, R. (2019). Mitochondrial genome fragmentation unites the parasitic lice of eutherian mammals. ***Systematic Biology***, 68(3), 430-440.

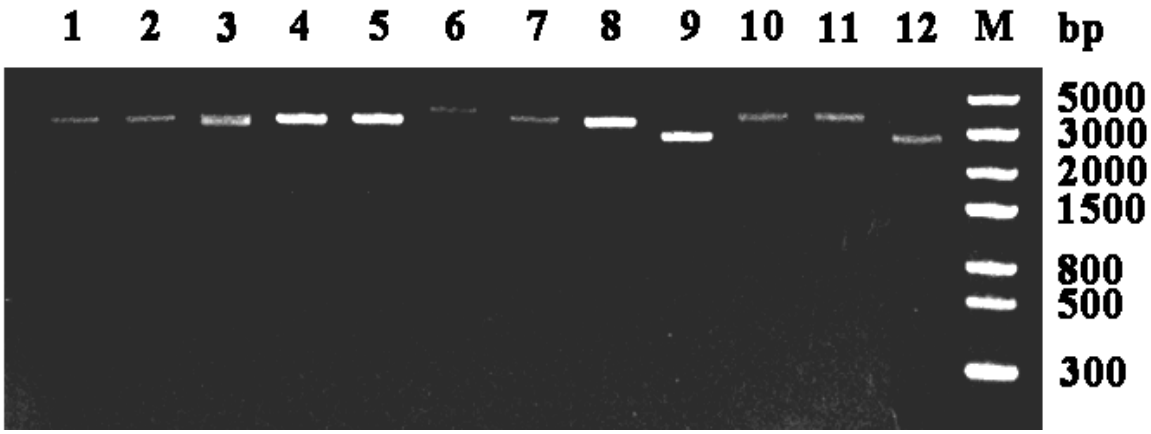
Song, S. D., Barker, S. C., & Shao, R. (2014). Variation in mitochondrial minichromosome composition between blood-sucking lice of the genus *Haematopinus* that infest horses and pigs. ***Parasites & Vectors***, 7(1), 1-8.

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.

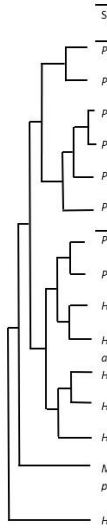
Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. ***Systematic Biology***, 56(4), 564-577.

Wilson, D. E., & Reeder, D. M. (Eds.). (2005). ***Mammal species of the world: a taxonomic and geographic reference*** (Vol. 1). JHU press.

2.8. Supplementary files

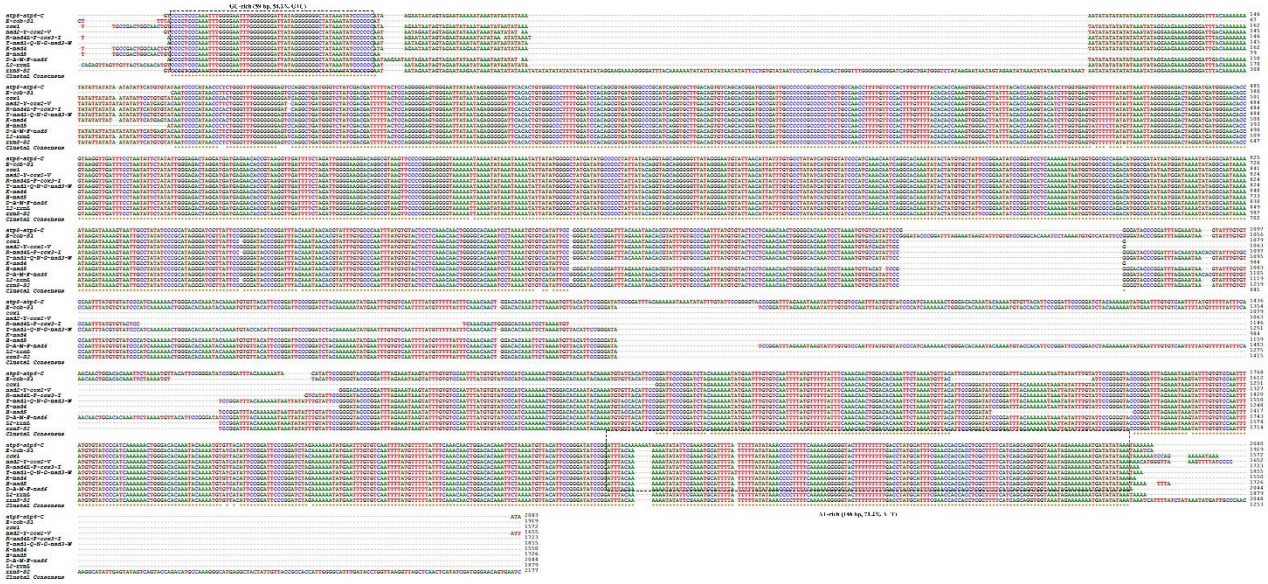


Supplementary Fig. 2.1. PCR verification of the mitochondrial minichromosomes of the macaque louse, *Pedicinus obtusus*. Lane 1–12: PCR amplicons from the 12 minichromosomes of the macaque louse: *cox1*, *L*$_2$*- rrnS*, *atp8-atp6-C*, *L*$_1$*-rrnL*, *L*$_2$*-rrnL*, *nad2-Y-cox2-V*, *R-nad4L-P-cox3-I*, *E-cob-S*$_1$, *K-nad4*, *H-nad5*, *T-nad1-Q-N-G-nad3-W*, *D-A-M-F-nad6*. Genes from which PCR primers were designed are in bold. M: DL2000 DNA marker (band size in bp indicated).



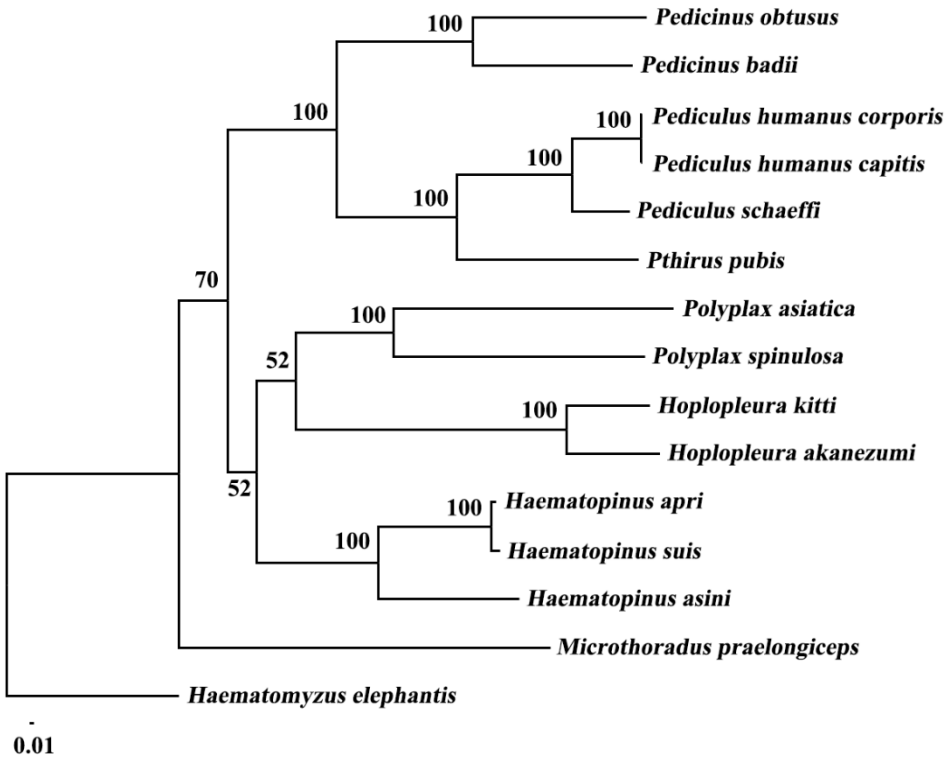
Species	Suborder	Host	*cox3-A*	*rrnS-C*	*E-cob*	*F-nad6*	*G-nad3*	*H-nad5*	*nad2-I*	*K-nad4*	*L1-rrnL*	*L2-rrnL*	*L2-rrnS*	*M*	*P-cox3*	*nad1-Q*	*R-nad4L*	*cob-S1*	*S2*	*T-nad1*	*nad3-W*	*Y-cox2*
Pediculus obtusus	Anoplura	Macaca	_	_	+	+	+	+	_	+	+	+	_	_	+	+	+	+	_	+	+	+
Pediculus badii	Anoplura	Colobus	_	_	+	+	+	+	+	+	+	_	+	_	+	+	+	+	+	+	+	+
Pediculus humanus	Anoplura	Human (body)	+	+	_	+	_	_	+	+	+	+	+	+	_	-	_	_	+	_	_	+
Pediculus capitis	Anoplura	Human (head)	+	+	_	+	_	_	+	+	+	+	+	+	_	+	_	_	+	_	_	+
Pediculus schaeffi	Anoplura	Chimp	+	+	_	+	_	_	+	+	+	+	+	_	_	+	_	+	+	_	_	+
Pthirus pubis	Anoplura	Human (pubic)	+	_	_	+	+	_	+	+	_	+	_	_	_	+	+	+	_	_	_	+
Polyplax asiatica	Anoplura	Greater bandicoot rat	_	+	+	_	+	+	_	+	+	_	_	_	+	_	+	_	_	+	+	+
Polyplax spinulosa	Anoplura	Asian house rat	_	+	+	_	+	+	_	+	_	+	_	_	+[1]	_	+	_	_	+[2]	+	+
Hoplopleura kitti	Anoplura	Bower's white-toothed rat	+	_	+	N.A	+	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	_	_	_	+	+	+	_	+	+	_	+	_	+	+	+	_	_	+	_	+
If ancestral to primate lice, changes required to account the data			3	4	2	3	1	1	4	0	2	6	4	7	1	2	1	4	5	2	3	0
Ancestral to primate lice by parsimony			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Supplementary Fig. 2.2. Inference of the positions of tRNA genes in the ancestral mitochondrial minichromosomes of higher primate lice. Three tRNA genes (*trnD*, *trnN*, *trnV*) are not included as no shared character of these three genes between species were observed. Plus (+) indicates presence; minus (-) indicates absence. [1]pseudo-P upstream from *cox3*; [2]pseudo-T upstream from *cox1*; [3]*trnM* alone on a minichromosome most likely occurred independently in human lice and horse louse (Song et al. 2014; Herd et al. 2015).



Supplementary Fig. 2.3. Alignment of nucleotide sequences in the NCRs of the mitochondrial minichromosomes of the macaque louse, *Pedicinus obtusus*.

Supplementary Fig. 2.4. Alignment of nucleotide sequences of the GC-rich region (A) and the AT-rich region in the mitochondrial minichromosomes of the colobus louse, *Pedicinus badii*. GC-rich region is downstream from the coding region; AT-rich region is upstream from the coding region.



Supplementary Fig. 2.5. Phylogenetic relationships among 14 species of sucking lice (Anoplura) inferred from maximum likelihood (ML) analysis of deduced amino acid sequences of nine mitochondrial genes. The elephant louse, *Haematomyzus elephantis*, was used as the outgroup. Bootstrap values were indicated at nodes.

Supplementary Table 2.1. PCR primers used to verify each mitochondrial minichromosome of the macaque louse, *Pedicinus obtusus*

Primer	Sequence (5' to 3')	Target gene
cox1F	AATGTTTTTGTTACTTCTCACGCTT	*cox1* (cytochrome *c* oxidase subunit 1)
cox1R	TGTGTCTGTCTGATAGAAAGGAACC	*cox1* (cytochrome *c* oxidase subunit 1)
rrnSF	CTAAAAAATAAAGCAAGTCAAGGTA	*rrnS* (small ribosome RNA subunit)
rrnSR	GATTACGCACACCTGCCTCTGATTA	*rrnS* (small ribosome RNA subunit)
atp6F	GAATGAAAGCCCAGAATAAAAAGAT	*atp6* (ATP synthase F0 subunit 6)
atp6R	CATCAAATAACAACCCAAGACCAGA	*atp6* (ATP synthase F0 subunit 6)
rrnLF	CAGAAGGACGAGAAGACCCTGTAGA	*rrnL* (large ribosome RNA subunit)
rrnLR	AAGCCTCTTGTCAACTCTTTCATTC	*rrnL* (large ribosome RNA subunit)
rrnLF	CAGAAGGACGAGAAGACCCTGTAGA	*rrnL* (large ribosome RNA subunit)
rrnLR	AAGCCTCTTGTCAACTCTTTCATTC	*rrnL* (large ribosome RNA subunit)
cox2F	TACTTACACCAGATGGAGCCGATGC	*cox2* (cytochrome *c* oxidase subunit 2)
cox2R	TGCCCTATCACTTTTACCCTCATCC	*cox2* (cytochrome *c* oxidase subunit 2)
cox3F	GGTAGTGAAAATGGCTTGAAATGGA	*cox3* (cytochrome *c* oxidase subunit 3)
cox3R	GAACGAGCCCCCAACAGAATACAAA	*cox3* (cytochrome *c* oxidase subunit 3)
cytbF	GTTAGTTCGTTTCTTTTCTATTCAT	*cob* (cytochrome b)
cytbR	CCTCCTCAAATCCATACTACTAACC	*cob* (cytochrome b)
nad4F	GTGGTGGCTTCCAGCAACCTTTA	*nad4* (NADH dehydrogenase subunit 4)
nad4R	CCAACCGACGAGTAGATTGAAGT	*nad4* (NADH dehydrogenase subunit 4)
nad5F	TGGTCATTTCTGTTCTTCCTCTGCC	*nad5* (NADH dehydrogenase subunit 5)
nad5R	ACAAGAAGGGAAGACCTACCATACA	*nad5* (NADH dehydrogenase subunit 5)
nad3F	CCGTTTGAGTGTGGTGTAATGCCTT	*nad3* (NADH dehydrogenase subunit 3)
nad3R	AAGCGAAAATGCCCACAAACCCTAA	*nad3* (NADH dehydrogenase subunit 3)
nad6F	CCTTCAGCAAAAGTGGATTGGACCT	*nad6* (NADH dehydrogenase subunit 6)
nad6R	ACTACCCCAACCGACAGTATCAACC	*nad6* (NADH dehydrogenase subunit 6)

Chapter 3

Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals

Published as:

Dong, W. G., **Dong, Y**., Guo, X. G., & Shao, R. (2021).

Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals.

BMC Genomics, 22(1), 1-16.

Manuscript: Appendix II

Contributions to authorship: I reconstructed the phylogeny of sucking lice and inferred the ancestral mitochondrial karyotype of *Polyplax* lice. I did statistical analysis of tRNA translocation and wrote the first version of the manuscript. I prepared the figure of phylogenetic trees and over 50% of the tables from statistical analysis. Overall, I contributed at least 50% to this manuscript.

3.1. Abstract

The typical single-chromosome mitochondrial (mt) genome of animals became fragmented in the lineage Mitodivisia, which contains most of the parasitic lice of eutherian mammals. These parasitic lice differ from each other even among closely related species in the same genus in mt karyotype, i.e. the number of minichromosomes, and the gene content and gene order in each minichromosome, which is in stark contrast to the extremely conserved single-chromosome mt genomes across most animal lineages but resembles the nuclear genomes of eukaryotes. How fragmented mt genomes evolved is still poorly understood. We use *Polyplax* sucking lice as a model to investigate how tRNA gene translocation shapes the dynamic mt karyotypes. We sequenced the full mt genome of the Asian grey shrew louse, *Polyplax reclinata*, inferred the ancestral mt karyotype for *Polyplax* lice and compared it with the mt karyotypes of the three *Polyplax* species sequenced to date. We found that tRNA genes are entirely responsible for mt karyotype variation in these *Polyplax* lice. Furthermore, tRNA gene translocation is frequent between different types of minichromosomes and towards the control region. A similar pattern of tRNA gene translocation can also been seen in other sucking lice with fragmented mt genomes. We conclude that inter-minichromosomal tRNA gene translocation orientated towards the control region is a major contributing factor to the highly dynamic mitochondrial genome organization in the parasitic lice of mammals.

3.2. Introduction

Extensive fragmentation of mitochondrial (mt) genome was discovered first in three species of human lice, in which the single mt chromosome typical of animals evolved into 14 and 20 minichromosomes; each minichromosome has 1–5 genes and is 1.8–4 kb in size (Shao et al., 2009; Shao et al., 2012). Since then, 26 more species of parasitic lice have been sequenced for mt genomes (Jiang et al., 2012; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Herd et al., 2015; Shao et al., 2015; Shao et al., 2017; Song et al., 2019; Fu & Dong et al., 2020; Sweet et al., 2021). It appears that mt genome fragmentation occurred at least twice in parasitic lice: once 60-90 million years ago (MYA) in the most recent common ancestor of Mitodivisia - a newly identified clade that contains the vast majority of parasitic lice of eutherian mammals (Light et al., 2010; Smith et al., 2011a; Johnson et al., 2018; Song et al., 2019), and later ~25 MYA in the feather lice of the genus *Columbicola* (Smith et al., 2011; Sweet et al., 2020).

Unlike the extremely conserved single-chromosome mt genomes in most animal lineages (Boore, 1999) but somehow resembling the nuclear genomes of eukaryotes (Wurster & Benirschke, 1970; Qumsiyeh, 1994, Rieseberg, 2001), the mt genome organization is highly dynamic in parasitic lice of the Mitodivisia clade (Song et al., 2019). These lice differ from each other, even among closely related species in the same genus, in mt karyotype, i.e. the number of minichromosomes, and the gene content and gene order in each minichromosome (Shao et al., 2015). The pig lice, *Haematopinus apri* and *Haematopinus suis*, and the horse louse, *Haematopinus asini*, have nine minichromosomes (Jiang et al., 2013; Song et al., 2014), whereas the human head louse, *Pediculus humanus capitis*, and the human body louse, *Pediculus humanus humanus*, have 20 minichromosomes (Shao et al., 2009; Shao et al., 2012). Between these extremes, the elephant louse, *Haematomyzus elephantis*, has 10 minichromosomes (four genes not identified) (Shao et al., 2015). The rodent lice, *Polyplax asiatica*, *Polyplax spinulosa*, *Hoplopleura akanezumi* (nine genes not identified) and *Hoplopleura kitti* (three genes not identified), have 11 minichromosomes (Dong et al., 2014a; Dong et al., 2014b). The guanaco louse, *Microthoracius praelongiceps*, the sheep louse, *Bovicola ovis*, the cattle louse, *Bovicola bovis*, and the macaque louse, *Pedicinus obtusus*, have 12 minichromosomes (Shao et al., 2017; Song et al., 2019; Fu & Dong et al., 2020). The goat louse, *Bovicola caprae*, and the dog louse, *Trichodectes canis*, have 13 minichromosomes (Song et al., 2019). The human pubic louse, *Pthirus pubis* (one gene not identified), and the colobus louse, *Pedicinus badii*, have 14 minichromosomes (Shao et al., 2012; Fu & Dong et al., 2020). Each minichromosome has 1 to 8 genes; the arrangement of genes in a minichromosome varies from species to species even within a genus; the only exceptions are: 1) the human head louse and the human body louse in the genus *Pediculus*, which diverged 83,000-170,000 years ago at the origin of clothing (Kittler et al., 2003; Shao et al., 2009; Shao et al., 2012;); and 2) the domestic pig louse and the wild pig louse in the genus *Haematopinus*, which diverged ~10,000 years ago when pigs were domesticated (Jiang et al., 2013; Ottoni et al., 2013).

Why is the fragmented mt genome organization so variable in these parasitic lice? A few comparative studies to date have suggested four models or mechanisms. First, Shao et al., (2012) showed that point mutation at the third anti-codon position or homologous recombination between tRNA genes accounted for the swap of identity and location between *trnL*$_1$ and *trnL*$_2$ and between *trnR* and *trnG* in the human lice. Dong et al., (2014a) confirmed that this mechanism accounted for the swap of identity and location between *trnL*$_1$ and *trnL*$_2$ in the spiny rat louse, *Polyplax spinulosa*.

Second, Shao et al., (2012) showed that one minichromosome could split into two via gene degeneration followed by deletion, based on analysis of pseudo gene sequences of the human pubic louse. Third, Song et al., (2014) showed that inter-minichromosomal recombination accounted for gene translocation in the horse louse, *Haematopinus asini*, and mt karyotype variation among *Haematopinus* lice. Fourth, Shao et al., (2017) showed that both split and merger of minichromosomes had occurred and contributed to the variation in mt karyotypes in sucking lice.

In the present study, we used *Polyplax* sucking lice, which parasitize rodents and shrews, as a model to investigate the role of tRNA gene translocation in shaping the dynamic mt karyotypes. We show that frequent directional tRNA gene translocation between different types of minichromosomes is a major contributor to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals.

3.3. Materials and methods

3.3.1. Louse collection, DNA extraction, mitochondrial genome amplification and sequencing

Specimens of *Polyplax reclinata* were collected in Yunnan province, China, from the Asian grey shrew, *Crocidura attenuata* (Soricidae) (Sample # 364) (Chin, 1999). Genomic DNA was extracted from individual lice with DNeasy Blood & Tissue kit (QIAGEN). Two pairs of primers, 12SA–12SB (Kambhampati & Smith, 1995) and 16SF–Lx16SR (Shao et al., 2017), were used to amplify fragments of mt genes *rrnS* (375 bp) and *rrnL* (360 bp) (Supplementary Table 3.1). These primers target conserved gene sequence motifs among arthropods. The two gene fragments were sequenced directly using Sanger method at the Thermo Fisher Scientific Genome Sequencing Facility (Guangzhou). Two pairs of specific primers for *P. reclinata*, 12S364F-12S364R and 16S364F-16S364R, were designed from *rrnS* and *rrnL* sequences obtained (Supplementary Table 3.1). The two specific primers in each pair go outwards with 15 and 21 bp in between, respectively. PCRs with these specific primers amplified the near full-length *rrnS* and *rrnL* minichromosomes of *P. reclinata* (Supplementary Fig. 3.1). The amplicons from *rrnS* and *rrnL* minichromosomes, 2.3 and 2.6 kb in size respectively, were sequenced using Sanger method to obtain the sequences of non-coding regions. Another pair of primers specific to *P. reclinata*, 364F–364R (Supplementary Table 3.1), was designed from conserved sequences of the non-coding regions that flank the coding regions of the *rrnS* and *rrnL* minichromosomes. The PCR with the primer pair 364F–364R produced a mixture of amplicons ranging from 1.1 to 1.8 kb in size, expected from the coding

regions of the entire set of mt minichromosomes of *P. reclinata* (Supplementary Fig. 3.1). These amplicons were sequenced from both ends (i.e. paired-end) with Illumina Miseq platform: insert size 400 bp, read length 250 bp. The PCR strategy used in this study was developed from our observations in previous studies that each mt minichromosome has a distinct coding region but a well-conserved non-coding region (Shao et al., 2009; Shao et al., 2012; Jiang et al., 2013; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Shao et al., 2017).

Takara Ex Taq was used in the initial short PCRs with the following cycling conditions: 94°C for 1 min; 35 cycles of 98°C for 10 sec, 45°C for 30 sec, 72°C for 1 min; and a final extension of 72°C for 2 min. Takara LA *Taq* was used in the long PCRs with the cycling conditions: 94°C for 1 min; 35 cycles of 98°C for 10 sec, 55–65°C (depending on primers) for 40 sec, 68°C for 4 min; and a final extension of 72°C for 8 min. Negative controls were executed with each PCR experiment. PCR amplicons were checked by agarose gel (1%) electrophoresis; the sizes of amplicons were estimated by comparing with DNA markers. PCR amplicons were purified with Wizard SV Gel/PCR clean-up system (Promega). Sanger sequencing was with the Thermo Fisher Scientific Genome Sequencing Facility (Guangzhou). Illumina sequencing was with the Majorbio Genome Sequencing Facility (Shanghai).

3.3.2. Assembly of sequence reads, gene identification and minichromosome verification

Illumina sequence reads (250 bp each) were assembled into contigs with Geneious 11.1.5 (Kearse et al., 2012). The assembly parameters were minimum overlap 150 bp and minimum identity 98%. tRNA genes were identified using tRNAscan-SE (Lowe & Eddy,1997) and ARWEN (Laslett & Canbäck, 2008). Protein-coding genes and rRNA genes were identified with BLAST searches of GenBank (Gish & States, 1993; Altschul et al., 1997). Sequence alignments were with Clustal X (Larkin et al., 2007). The size and circular organization of each mt minichromosome of *P. reclinata* were verified by PCR using outbound specific primers designed from the coding region of each minichromosome (Supplementary Fig. 3.2). The forward primer and reverse primer in each pair were next to each other with a small gap in between. PCRs with these primers amplified each circular minichromosome in full or near full length (Supplementary Fig. 3.1); these amplicons were also sequenced with Illumina Miseq platform to obtain the full-length sequence of the non-coding region of each minichromosome. PCR set-up, cycling conditions, agarose gel electrophoresis and size measurement were the same as described above. Negative controls were run for each PCR test.

The annotated mt genome sequence of *P. reclinata* was deposited in GenBank (accession numbers MW291451-MW291461).

3.3.3. Phylogenetic analyses

We retrieved the mt genome sequences of 14 species of sucking lice and the elephant louse from GenBank (Table 3.1) and combined these sequences with that of *P. reclinata* generated in the present study. The sequences of eight mt protein-coding genes (*atp6, atp8, cox1, cox2, cox3, cob, nad4L, nad6*) and two rRNA genes (*rrnS* and *rrnL*) of these lice were used in phylogenetic analysis. Other mt protein-coding genes (*nad1, nad2, nad3, nad4*, and *nad5*) were excluded from our analysis because their sequences were not available to all of the species above. Protein-coding gene sequences were aligned based on amino acid sequences using the MAFFT algorithm implemented in TranslatorX online platform (Abascal et al., 2010). rRNA genes were aligned using the MAFFT v7.0 online server with G-INS-i strategy (Katoh & Standley, 2013). Ambiguously aligned sites were omitted using GBlocks v0.91b (Talavera & Castresana, 2007) with default settings. Individual gene alignments were concatenated after removing poorly aligned sites using GBlocks v0.91b. Individual gene alignments were concatenated after removing poorly aligned sites using GBlocks v0.91b. A concatenated alignment, PCGRNA matrix, was used in subsequent analyses; this matrix combines the sequences of the eight protein-coding genes and the two rRNA genes (4,631 bp in total). The matrix was analysed using maximum likelihood (ML) and Bayesian methods with MEGA-X (Kumar et al., 2018) and 2) Bayesian inference method (BI) with MrBayes 3.2.6 (Ronquist & Huelsenbeck, 2003), respectively. For ML analysis, the bootstrap replicates were settled as 500. The substitutional model was *Tamura-Nei* and the ML heuristic method was *Nearest-Neighbor-Interchange (NNI)*. For Bayesian analyses, four independent Markov chains were run for 25,000 MCMC generations, sampling a tree every 10 generations. This analysis was run until the average standard deviation of split frequencies was lower than 0.001. The initial 10,000 trees of each MCMC run were discarded as burn-in. ML trees were directly generated in MEGA-X and the Bayesian trees were drawn with Figtree v1.4.3.

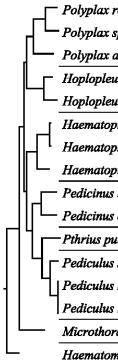
Table 3.1. Species of parasitic lice included in the phylogenetic analyses in this study

Species	Family	Parvorder	Order	GenBank accession number	References
Polyplax reclinata	Polyplacidae	Anoplura	Phthiraptera	MW291451-MW291461	Present study
Hoplopleura akanezumi	Hoplopleuridae	Anoplura	Phthiraptera	KJ648922–32	Dong et al., 2014a
Hoplopleura kitti	Hoplopleuridae	Anoplura	Phthiraptera	KJ648933–43	Dong et al., 2014a
Polyplax spinulosa	Polyplacidae	Anoplura	Phthiraptera	KF647762–72	Dong et al., 2014b
Polyplax asiatica	Polyplacidae	Anoplura	Phthiraptera	KF647751–61	Dong et al., 2014b
Haematopinus asini	Haematopinidae	Anoplura	Phthiraptera	KF939318, KF939322, KF939324, KF939326, KJ434034–38	Song et al., 2014
Haematopinus apri	Haematopinidae	Anoplura	Phthiraptera	KC814611–19	Jiang et al., 2013
Haematopinus suis	Haematopinidae	Anoplura	Phthiraptera	KC814602–10	Jiang et al., 2013
Pthirus pubis	Pthiridae	Anoplura	Phthiraptera	EU219988–95, HM241895–98	Shao et al., 2012
Pediculus schaeffi	Pediculidae	Anoplura	Phthiraptera	KC241882–97, KR706168–69	Herd et al., 2015
Pediculus capitis	Pediculidae	Anoplura	Phthiraptera	JX080388–407	Shao et al., 2012
Pediculus humanus	Pediculidae	Anoplura	Phthiraptera	FJ499473–90	Shao et al., 2009
Pedicinus obtusus	Pedicinidae	Anoplura	Phthiraptera	MT792495-506	Fu et al., 2020
Pedicinus badii	Pedicinidae	Anoplura	Phthiraptera	MT721726-39	Fu et al., 2020
Microthoradus praelongiceps	Microthoraciidae	Anoplura	Phthiraptera	KX090378–89	Shao et al., 2017
Haematomyzus elephantis	Haematomyzidae	Rhyncophth-irina	Phthiraptera	KF933032–41	Shao et al., 2015

3.3.4. Inference of the ancestral mitochondrial karyotype of *Polyplax* lice

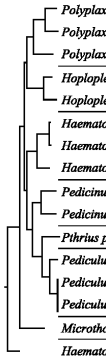
We used a parsimony method described in Shao et al., (2017) to infer the ancestral mt karyotype of *Polyplax* lice based on information from the complete mt genomes of three *Polyplax* species: *P. reclinata* (present study), *P. asiatica* (Dong et al., 2014a) and *P. spinulosa* (Dong et al., 2014a). We mapped characters of mt karyotypes on the phylogenetic tree of sucking lice inferred from mt genome sequences to identify shared characters. We inferred a mt minichromosome character to be ancestral to *Polyplax*: 1) if it was present in at least one *Polyplax* species and also in one or more of the outgroup species; or 2) if it was present in all of the three *Polyplax* species. We considered tRNA genes separately from protein-coding and ribosome RNA genes because protein-coding and

rRNA genes were much more stable than tRNA genes (Shao et al., 2017) (Fig. 3.1; Fig. 3.2). For each inferred ancestral minichromosome character of *Polyplax* lice, we counted the changes needed to explain the observed data of *Polyplax* lice. If two or more minichromosome characters conflicted with one another, the character with the least required changes was inferred to be the ancestral. If two or more characters are equally parsimonious, all characters are inferred to be equally likely to be ancestral (Fig. 3.1, Fig. 3.2).



Species	Family	Host	*atp8-atp6*	*cob*	*cox1*	*cox2-nad6*	*nad4L-cox3*	*nad1*-*nad3*	*nad2*	*nad4*	*nad5*	*rrnS*	*rrnL*
Polyplax reclinata	Polyplacidae	Asian grey shrew	+	+	+	+	+	+	+	+	+	+	+
Polyplax spinulosa	Polyplacidae	Asian house rat	+	+	+	+	+	+	+	+	+	+	+
Polyplax asiatica	Polyplacidae	Greater bandicoot rat	+	+	+	+	+	+	+	+	+	+	+
Hoplopleura kitti	Hoplopleuridae	Bower's white toothed rat	+	+	+	-	+	+	+	+	NA	+	+
Hoplopleura akanezumi	Hoplopleuridae	Chevrier's field mouse	+	+	+	-	+	NA	+	+	NA	+	+
Haematopinus apri	Haematopinidae	Wild pig	-	+	-	-	-	+	-	-	+	+	+
Haematopinus suis	Haematopinidae	Domestic pig	-	+	-	-	-	+	-	-	+	+	+
Haematopinus asini	Haematopinidae	Horse	-	+	-	-	-	+	-	-	-	-	+
Pedicinus badii	Pedicinidae	Colobus	+	+	+	-	+	-	+	+	+	+	+
Pedicinus obtusus	Pedicinidae	Macaque	+	+	+	-	+	+	-	+	+	+	+
Pthrius pubis	Pthiridae	Human (pubic)	+	+	+	-	-	-	+	+	+	+	+
Pediculus schaeffi	Pediculidae	Chimpanzee	+	+	+	-	-	-	+	+	+	+	+
Pediculus humanus corporis	Pediculidae	Human (body)	+	+	+	-	-	-	+	+	+	+	+
Pediculus humanus capitis	Pediculidae	Human (head)	+	+	+	-	-	-	+	+	+	+	+
Microthoradus praelongiceps	Microthoraciidae	Guanaco	+	+	+	+	+	-	+	+	+	+	+
Haematomyzus elephantis	Haematomyzidae	Elephant	-	-	+	-	-	-	NA	+	+	+	+
Ancestral to *Polyplax* species			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
If ancestral to *Polyplax* species, changes required to account the *Polyplax* data			0	0	0	0	0	0	0	0	0	0	0

Fig. 3.1. Inferring the ancestral mitochondrial karyotype of *Polyplax* lice. tRNA genes are excluded. Plus symbol (+) is for presence, minus (-) for absence, and NA for information not available. Underlined *nad1* has an opposite orientation of transcription to other genes. The phylogenetic tree was inferred with mitochondrial genome sequences (see Fig. 3.4 for details).



Species	*cox3-A*	*nad6-A*	*rrnS-C*	*D-Y-cox2*	*E-cob-I*	*nad5-F*	*G-nad3-W*	*H-nad5*	*K-nad4*	*M-L₁-rrnL*	*cox1-L₂*	*Q-nad2-N*	*nad4L-P*	*R-nad4L*	*S₁-S₂-rrnS*	*S₁-S₂-nad1*	*nad1-T*	*T-D-Y-cox2*	*rrnL-V*
Polyplax reclinata	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+
Polyplax spinulosa	-	+	+	+	+	+	+	+	+	-	-	+	+	+	+	-	-	+	+
Polyplax asiatica	+	-	+	+	+	-	+	+	+	+	+	+	-	+	-	+	+	-	+
Hoplopleura kitti	+	-	-	+	-	NA	-	NA	+	+	-	-	+	+	-	-	-	-	+
Hoplopleura akanezumi	+	-	-	+	-	NA	NA	NA	+	-	-	-	+	+	-	-	-	-	+
Haematopinus apri	+	-	+	+	-	+	+	+	+	-	+	-	-	+	-	-	+	-	-
Haematopinus suis	+	-	+	+	-	+	+	+	+	-	+	-	-	+	-	-	+	-	-
Haematopinus asini	+	-	+	+	-	+	+	+	+	-	+	-	-	+	-	-	+	-	-
Pedicinus badii	-	-	-	-	-	-	+	+	+	-	-	-	+	+	-	-	-	-	-
Pedicinus obtusus	-	-	-	-	-	-	+	+	+	-	-	-	+	+	-	-	-	-	-
Pthrius pubis	+	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-
Pediculus schaeffi	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
Pediculus humanus corporis	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
Pediculus humanus capitis	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
Microthoradus praelongiceps	+	-	+	-	-	+	+	+	+	+	+	-	+	+	-	-	+	-	+
Haematomyzus elephantis	-	-	-	-	-	-	-	+	+	-	-	-	-	+	-	-	+	-	+
Ancestral to *Polyplax* species	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Possible	Possible	Yes	No	Yes
If ancestral to *Polyplax* species, changes required to account the *Polyplax* data	1	NA	0	0	0	1	0	0	0	1	1	0	1	0	1	1	1	NA	0

Fig. 3.2. Inferring the location of tRNA genes relative to protein-coding and rRNA genes in the ancestral mitochondrial karyotype of *Polyplax* lice. tRNA genes are indicated with their single-

letter abbreviations. Plus symbol (+) is for presence, minus (-) for absence, and NA for information not available. Underlined genes have opposite orientation of transcription to other genes. The phylogenetic tree was inferred with mitochondrial genome sequences (see Fig. 3.4 for details).

3.4. Results and Discussion

3.4.1. Mitochondrial genome of the Asian grey shrew louse, *Polyplax reclinata*

We obtained the complete mt genome of *Polyplax reclinata* by assembling the 53,915,414 clean Illumina sequence reads (paired-end, 250 bp each) generated from the minichromosome amplicons (Table 3.2). All of the 37 mt genes typical of bilateral animals were identified in *Polyplax reclinata*; these genes were on 11 circular minichromosomes (Fig. 3.3). The mt minichromosomes of *Polyplax reclinata* range from 2,332 to 2,869 bp in size (Table 3.2); each minichromosome has a coding region and a non-coding region (Fig. 3.3; Table 3.2). The coding region of each minichromosome contains 2–6 genes and ranges from 847 bp for *atp8-atp6* minichromosome to 1,809 bp for *H-nad5-F* minichromosome (Note: minichromosomes are named after their genes hereafter). Seven of the 11 minichromosomes of *Polyplax reclinata* have a single protein-coding or rRNA gene each; the other four minichromosomes have two protein-coding genes each. Each minichromosome has 1–4 tRNA genes except *atp8-atp6* minichromosome, which has no tRNA genes (Fig. 3.3). Each of the 37 mt genes in *Polyplax reclinata* is found only in one minichromosome; all genes have the same orientation of transcription relative to the non-coding region except for *nad1*, which has an opposite orientation (Fig. 3.3).

We sequenced the non-coding regions of all of the 11 minichromosomes of *Polyplax reclinata* in full length. The non-coding regions range from 1,060 to 1,578 bp (Table 2.3); the size variation is due to the additional sequences in the middle of the non-coding regions of *M-L$_1$-rrnL-V*, *S$_1$-S$_2$-rrnS-C*, *K-nad4*, *atp8-atp6*, *Q-nad2-N* and *E-cob-I* minichromosomes (Supplementary Fig. 3.2). Excluding the additional sequences, the non-coding regions of all 11 minichromosomes have high sequence similarity to one another. Overall, the non-coding regions have high C and G content in one end but high A and T content in the other end. Indeed, a CG-rich motif (54 bp, 59% C and G) is immediately downstream from the 3'-end of the coding regions and an AT-rich motif (113 bp, 68% A and T) is upstream from the 5'-end of the coding regions in all of the minichromosomes (Supplementary Fig. 3.2)

Table 3.2. The mitochondrial minichromosomes of the Asian grey shrew louse, *Polyplax reclinata*, identified by Illumina sequencing

Minichromosome	Minichromosome size (bp)	Size of coding region (bp)	Size of non-coding region (bp)	Number of Illumina sequence reads	Mean coverage	GenBank accession number
atp8-atp6	2,425	847	1,578	5,836,284	333,785	MW291451
E-cob-I	2,620	1,223	1,397	7,275,348	387,778	MW291453
cox1-L2	2,816	1,580	1,236	4,012,234	204,899	MW291452
T-D-Y-cox2-nad6-A	2,473	1,399	1,074	3,691,415	212,776	MW291461
R-nad4L-P-cox3	2,465	1,191	1,274	6,293,311	365,198	MW291459
nad1-G-nad3-W[1]	2,693	1,450	1,243	9,485,576	351,940	MW291457
Q-nad2-N	2,558	1,124	1,434	5,969,577	330,770	MW291458
K-nad4	2,680	1,340	1,340	4,248,148	218,975	MW291455
H-nad5-F	2,869	1,809	1,060	4,258,208	209,453	MW291454
S1-S2-rrnS-C	2,332	900	1,432	441,782	42,400	MW291460
M-L1-rrnL-V	2,559	1,240	1,319	2,403,531	213,096	MW291456
Total	28,490	14,103	14,387	53,915,414	2,871,070	

Note: [1]underlined *nad1* is opposite to other genes in the orientation of transcription.

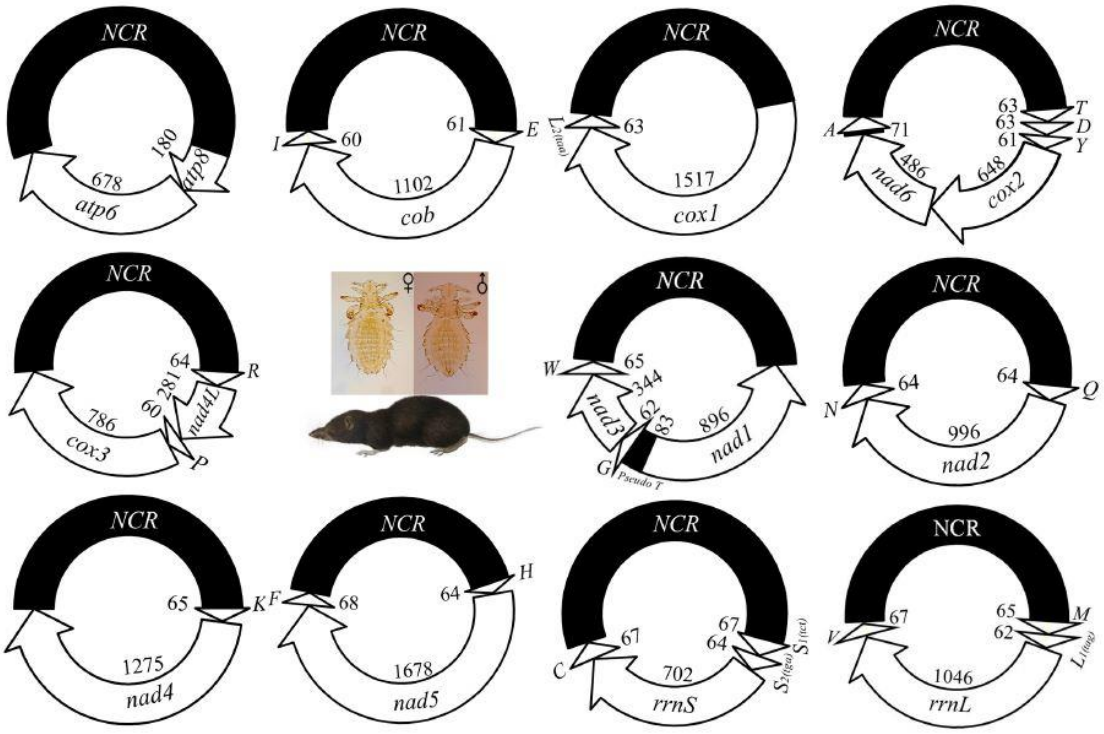


Fig. 3.3. The mitochondrial genome of the Asian grey shrew louse, *Polyplax reclinata*. Each minichromosome has a coding region and a non-coding region (NCR, in black). Names and transcription orientation of genes are indicated in the coding region; minichromosomes are in alphabetical order of protein-coding and rRNA gene names.

3.4.2. Mitochondrial karyotype variation among *Polyplax* species

Polyplax reclinata differs in mt karyotype from the two other *Polyplax* species reported previously (Dong et al., 2014a), although these three species all have 11 minichromosomes. *Polyplax reclinata* shares nine minichromosomes with *P. spinulosa* (louse of the Asian house rat) (Dong et al., 2014a) but differ in the distribution of *trnL1* and *trnL2* between the other two minichromosomes (Table 3.3). In *P. spinulosa*, *trnL1* is downstream from *cox1* and *trnL2* is upstream from *rrnL* (Dong et al., 2014a); in *P. reclinata*, however, it is the opposite (Fig. 3.3, Table 3.3). *Polyplax reclinata* shares only four minichromosomes with *P. asiatica* (louse of the greater bandicoot rat) (Dong et al., 2014a) and differ in the distribution of six tRNA genes among the other seven minichromosomes: *trnA*, *trnF*, *trnP*, *trnS1*, *trnS2* and *trnT* (Table 3.3). If tRNA genes are excluded, the three *Polyplax* species are identical to one another in the distribution of protein-coding genes and rRNA genes among the 11 minichromosomes (Table 3, Fig. 3.3).

Table 3.3. Comparison of gene content and gene arrangement in the mitochondrial minichromosomes of three *Polyplax* species and their most recent common ancestor (MRCA)

Inferred most recent common ancestor of *Polyplax*	*Polyplax reclinata*	*Polyplax spinulosa*	*Polyplax asiatica*
atp8-atp6	*atp8-atp6*	*atp8-atp6*	*atp8-atp6*
E-cob-I	*E-cob-I*	*E-cob-I*	*E-cob-I*
cox1-L2(taa)	*cox1-L2(taa)*	*cox1-**L1(tag)***	*cox1-L2(taa)*
D-Y-cox2-nad6	***T**-D-Y-cox2-nad6-**A***	***T**-D-Y-cox2-nad6-**A***	*D-Y-cox2-nad6*
R-nad4L-P-cox3-A	*R-nad4L-P-cox3*	*R-nad4L-P-cox3*	*R-nad4L-cox3-A*
S1(tct)-S2(tga)-nad1-T-G-nad3-W	*nad1-G-nad3-W*	*nad1-G-nad3-W*	***S1(tct)-S2(tga)**-nad1-T-G-nad3-W*
Q-nad2-N	*Q-nad2-N*	*Q-nad2-N*	*Q-nad2-N-**P***
K-nad4	*K-nad4*	*K-nad4*	*K-nad4-**F***
H-nad5-F	*H-nad5-F*	*H-nad5-F*	*H-nad5*
S1(tct)-S2(tga)-rrnS-C	***S1(tct)-S2(tga)**-rrnS-C*	***S1(tct)-S2(tga)**-rrnS-C*	*rrnS-C*
M-L1(tag)-rrnL-V	*M-L1(tag)-rrnL-V*	*M-**L2(taa)**-rrnL-V*	*M-L1(tag)-rrnL-V*

Notes: The position of *S1* (tct)-*S2* (tga) gene cluster in the ancestral mitochondrial karyotype is either upstream from *rrnS* or, equally parsimoniously, downstream from nad1. The eight tRNA genes that have changed their positions and/or orientation in the three *Polyplax* species since their divergence from their MRCA are in bold.

3.4.3. Phylogeny of sucking lice inferred from mt genome sequences

To assist the inference of the ancestral mt karyotype of *Polyplax* lice, we inferred the phylogeny of sucking lice using mt genome sequences of 15 species of sucking lice of seven genera including *Polyplax reclinata*, which was sequenced in the present study; a chewing louse, the elephant louse,

was used as the outgroup (Table 3.1). We obtained two trees from PCGRNA matrix (4,631 bp of protein-coding and rRNA gene sequences) with ML and Bayesian methods. These two trees have similar topologies to each other and consistently support the monophyly of the genus *Polyplax* (Fig. 3.4): bootstrap support value (BSV) = 100%, posterior probability (PP) =1. Within the *Polyplax*, *P. reclinata* is more closely related to *P. spinulosa* than to *P. asiatica* (BSV=100%, PP=1), despite the fact that the latter two species are found mostly on rodents whereas *P. reclinata* is on shrews (Durden & Musser, 1994a; Durden & Musser, 1994b). The other four genera, *Hoplopleura*, *Pediculus*, *Pedicinus* and *Haematopinus*, also had multiple species from each genus included in our analysis; each of these genera was monophyletic with strong support (BSV=100%, PP=1) (Fig. 3.4). The 15 species of sucking lice were divided into three clades: *Polyplax* and *Hoplopleura* forming a clade (BSV=76.6%: PP=0.998); *Pediculus*, *Pedicinus*, *Pthirus* and *Haematopinus* in another clade (BSV=69.8%; PP=0.983); and *Microthoracius praelongiceps* alone as the earliest branch (BSV=85.2%; PP=1) (Fig. 3.4). Each of the genera included in our analysis also represents a family (Table 3.1). The family level relationships revealed by our analysis are consistent with that proposed by Kim (1988) based on morphological characters and partially consistent with that by Light et al. (2010), Smith et al. (2011a) and Johnson et al. (2018) based on gene sequences.

3.4.4. Inferred ancestral mitochondrial karyotype of *Polyplax* lice

We inferred the ancestral mt karyotype of *Polyplax* lice based on the data available from the three *Polyplax* species and the other 12 species of sucking lice (Table 1, Fig. 3.5). We mapped mt karyotype characters on the phylogeny inferred from the mt genome sequences of these lice to identify shared characters (Fig. 3.1, Fig. 3.2). The inferred ancestral mt karyotype of *Polyplax* lice by parsimony consists of 11 minichromosomes; each minichromosome has a coding region with 2-7 genes and a non-coding region (Fig. 3.5). Seven of the 11 minichromosomes have a single protein-coding or rRNA gene each; the other four minichromosomes have two protein-coding genes each: 1) *atp6* and *atp8*, 2) *cox2* and *nad6*, 3) *cox3* and *nad4L*, and 4) *nad3* and *nad1* (Fig. 3.5). The position of all genes in the ancestral mt karyotype can be inferred without conflict except the position of a cluster of two tRNA genes, S_1-S_2, which is either upstream from *rrnS* or, equally parsimoniously, downstream from *nad1* (Fig. 3.2, Fig. 3.5).

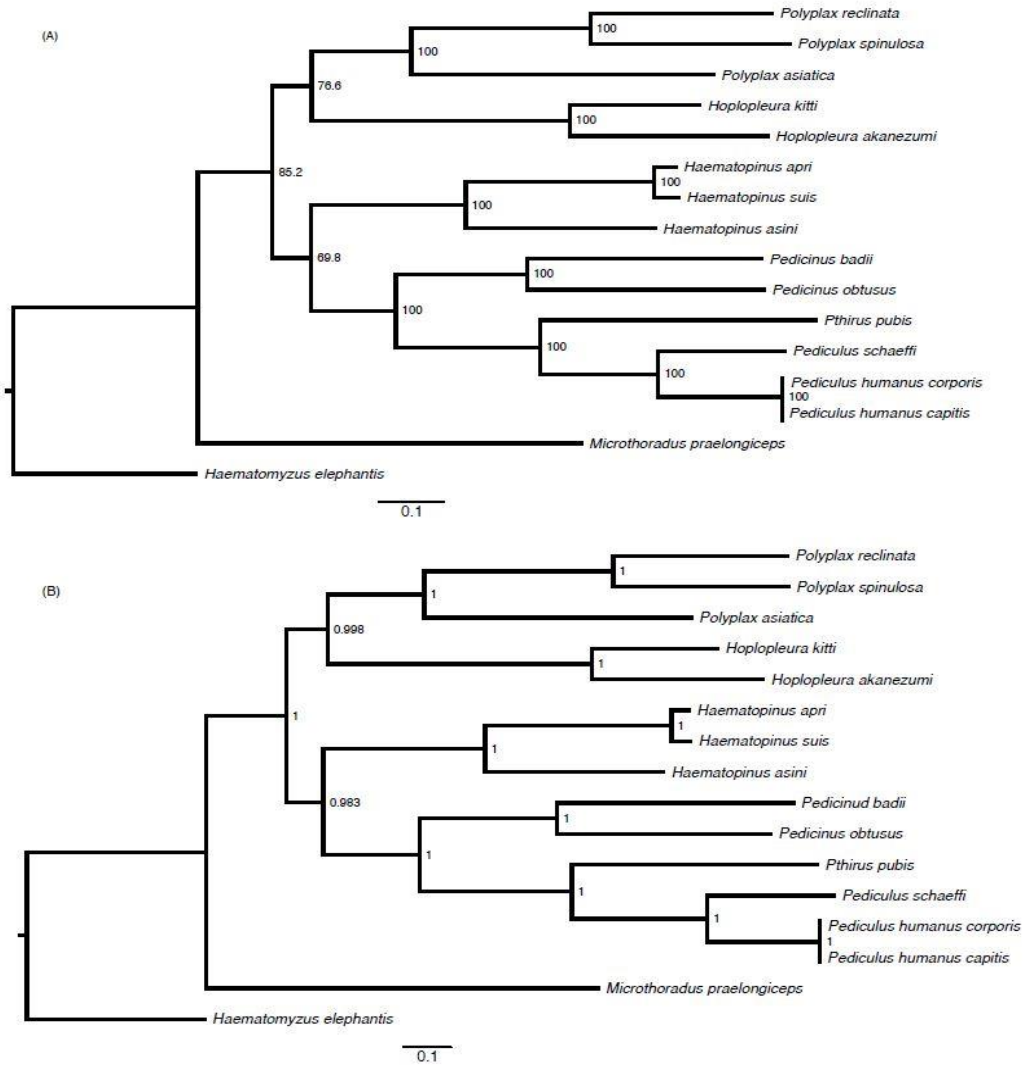


Fig. 3.4. Phylogenetic relationships among 15 species of sucking lice (Anoplura) inferred with combined sequences of eight mitochondrial protein-coding genes (*atp6*, *atp8*, *cox1*, *cox2*, *cox3*, *cob*, *nad4L*, *nad6*) and two rRNA genes (*rrnS* and *rrnL*) with maximum likelihood method (A) and Bayesian method (B). The trees were rooted with the elephant louse, *Haematomyzus elephantis* (Rhynchophthirina). Bootstrap support values (out of 100) and Bayesian posterior probability values (out of 1) were indicated near internal nodes.

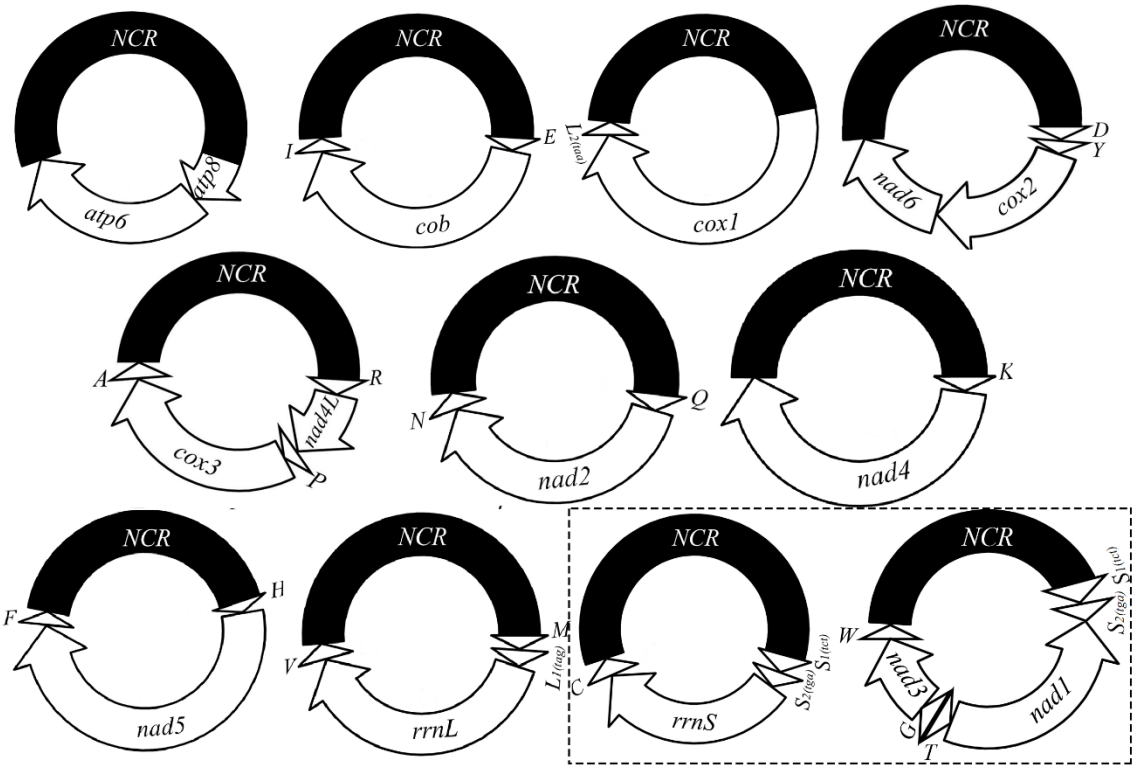


Fig. 3.5. Inferred ancestral mitochondrial karyotype of *Polyplax* lice. Arrow indicates transcription orientation; non-coding regions (*NCR*) are shaded in black. The two minichromosomes containing *S1-S2* gene cluster are boxed in dash line because the location of this cluster is either upstream from *rrnS* or, equally parsimoniously, downstream from *nad1*.

3.4.5. Mitochondrial tRNA gene translocation in *Polyplax* lice

When compared with the inferred ancestral mt karyotype of *Polyplax* lice (Fig. 3.5), all of the three *Polyplax* species sequenced to date retained 11 minichromosomes and the same distribution pattern of protein-coding and rRNA genes among the minichromosomes (Table 3). Therefore, no split nor merger of minichromosomes have occurred in any of these *Polyplax* species since their most recent common ancestor (MRCA) (Fig. 3.4). However, these three *Polyplax* species are different from one another and from their MRCA in the position of tRNA genes. In total, eight tRNA genes have changed their positions in the three *Polyplax* lice since they diverged from their MRCA: *trnA*, *trnF*, *trnL1*, *trnL2*, *trnP*, *trnS1*, *trnS2* and *trnT* (Table 3.3). *trnA* and *trnT* were likely translocated (*trnT* also inverted) in the MRCA of *P. reclinata* and *P. spinulosa* (Fig. 3.4) and thus are shared by these two species (Table 3.3). A pseudo *trnT* can be found in *P. reclinata* (Fig. 3.3) and *P. spinulosa* (Dong et al., 2014a), indicating the translocation of *trnT* occurred relatively recently. *trnL1* and

trnL2 swapped their positions in *P. spinulosa* (Dong et al., 2014a), most likely caused by point mutations at the third anti-codon position (or by homologous recombination between *trnL1* and *trnL2*) – a mechanism suggested by previous studies (Shao et al., 2012; Shao et al., 2017). *trnF* and *trnP* were translocated in *P. asiatica* (Table 3.3); as a cluster, *trnS1-trnS2* was translocated either in *P. asiatica* or equally likely in the MRCA of *P. reclinata* and *P. spinulosa* (Table 3.3, Fig. 3.5).

Table 3.4. Statistical analysis of expected probability and observed probability of tRNA gene translocation in three *Polyplax* species relative to their most recent common ancestor.

tRNA genes	**Expected probability of translocation to be adjacent to NCR if translocation is random**	**Observed probability of translocation to be adjacent to NCR**		**Expected probability of translocation to an intergenic region if translocation is random**	**Observed probability of translocation to an intergenic region**
A	43.75% (21/48)	100% (2/2)		56.25% (27/48)	0%
F	43.75% (21/48)	100% (1/1)		56.25% (27/48)	0%
P	45.83% (22/48)	100% (1/1)		54.12% (26/48)	0%
T	45.83% (22/48)	100% (2/2)		54.12% (26/48)	0%
S1-S2	44.68% (21/47)	100% (3/3)		55.32% (26/47)	0%
Average	44.8%	100%		55.2%	0%
Two tailed *t*-value	-118.61978			115.73698	
p-value	< 0.00001			< 0.00001	
Significant at $p < 0.01$?	Yes			Yes	

Notes: Refer to Table 3 for tRNA gene translocation in *Polyplax* species. "Expected probability of translocation to be adjacent to NCR" = "total sites adjacent to NCR" / "total sites adjacent to NCR + total intergenic sites". "Expected probability of translocation to an intergenic region" = "total intergenic sites" / "total sites adjacent to NCR + total intergenic sites". "Observed probability of translocation to be adjacent to NCR" = "translocation adjacent to NCR" / "translocation adjacent to NCR + translocation to intergenic sites". "Observed probability of translocation to an intergenic region" = "translocation to intergenic sites" / "translocation adjacent to NCR + translocation to intergenic sites". *T*-Test was done at:
https://www.socscistatistics.com/tests/studentttest/default2.aspx

3.4.6. tRNA gene translocation is frequent and directional towards the control region in parasitic lice with fragmented mitochondrial genomes

We noted in the *Polyplax* lice above that mt tRNA gene translocation was closely associated with the control region, i.e. the non-coding region (NCR). With no exception, the destination of all of the translocated tRNA genes (i.e. *trnA*, *trnF*, *trnS1-trnS2*, *trnP* and *trnT* excluding *trnL1* and *trnL2* which swapped positions in *P. spinulosa* as discussed above) were either immediately next to the NCR or

in the case of *trnP* of *P. asiatica*, a short distance (189 bp) inside the NCR (Table 3.3) (Dong et al., 2014a), indicating a possible role of NCR in tRNA gene translocation. The observed probability of tRNA gene translocation to be adjacent to NCR in these *Polyplax* species is significantly higher ($p < 0.00001$) than the expected probability (Table 3.4).

We investigated further if this was a broader pattern for parasitic lice with fragmented mt genomes by comparing the mt karyotypes: (1) between the MRCA of sucking lice (Shao et al., 2017) and the MRCA of *Polyplax* (Fig. 3.5); and (2) between the MRCA of sucking lice and each of the 12 non-*Polyplax* sucking louse species that have been sequenced to date (Shao et al., 2017; Fu & Dong et al., 2020). The MRCA of *Polyplax* species retained the same distribution pattern of protein-coding and rRNA genes and the same number of minichromosomes as the MRCA of sucking lice. However, five tRNA genes have translocated in the MRCA of *Polyplax* relative to the MRCA of sucking lice: all of these tRNA genes were translocated immediately next to the NCR despite alternative intergenic locations available for translocation (Table 3.5). The observed probability of tRNA gene translocation to be adjacent to NCR in the MRCA of *Polyplax* species is significantly higher ($p < 0.00001$) than the expected probability (Table 3.6). When compared with the MRCA of sucking lice, 19 of the 22 tRNA genes in total have translocated in the 12 species of non-*Polyplax* sucking lice: 32 out of the 54 (i.e. 59 %) translocation events resulted in tRNA genes moved to be adjacent to NCR despite alternative intergenic locations available for translocation (Table 3.7). The observed probability of tRNA gene translocation to be adjacent to NCR in these non-*Polyplax* species is 13.6 % higher than the expected probability but this difference is not statistically significant ($p = 0.116709$) (Table 3.8). Most likely tRNA gene translocation to be adjacent to NCR has been under-captured in non-*Polyplax* species (Table 3.7) due to the fact that any newly translocated tRNA gene adjacent to a NCR would inevitably push the previously translocated tRNA gene away from the NCR. After all, there can only be one gene at any time adjacent to each end of the NCR. Among the three comparisons above, the third comparison has the longest evolutionary time frame. The longer the evolutionary time is, the more likely tRNA gene translocation information to be adjacent to NCR will be under-captured.

Table 3.5. Translocated tRNA genes in the most recent common ancestor (MRCA) of *Polyplax* lice relative to the MRCA of sucking lice

tRNA gene	MRCA of sucking lice	MRCA of *Polyplax*	Translocated	Next to NCR
N	*atp8-atp6-N*	*Q-nad2-**N***	Yes	Yes
S1-S2	*E-cob-S1-S2*	***S1-S2**-rrnS-C* *or* ***S1-S2**-nad1-T-G-nad3-W*	Yes	Yes
I	*I-cox1-L2*	*E-cob-**I***	Yes	Yes
Q	*Q-nad1-T-nad1-G-nad3-W*	***Q**-nad2-N*	Yes	Yes

Note: translocated tRNA genes in the MRCA of *Polyplax* lice are in bold.

Table 3.6. Statistical analysis of expected probability and observed probability of tRNA gene translocation in the most recent common ancestor (MRCA) of *Polyplax* lice relative to the MRCA of sucking lice

tRNA genes	Expected probability of translocation to be adjacent to NCR if translocation is random	Observed probability of translocation to be adjacent to NCR		Expected probability of translocation to an intergenic region if translocation is random	Observed probability of translocation to an intergenic region
N	21/46=45.65%	100%		25/46=54.35%	0%
S1-S2	21/45=46.67%	100%		24/45=53.33%	0%
I	21/46=45.65%	100%		25/46=54.35%	0%
Q	21/46=45.65%	100%		25/46=54.35%	0%
Average	45.9%	100%		54.1%	0%
Two tailed *t*-value	-212.13726			212.13726	
p-value	< 0.00001			< 0.00001	
Significant at $p < 0.01$?	Yes			Yes	

Notes: Refer to Table 5 for tRNA gene translocation in the MRCA of *Polyplax* species. Refer to Table 4 notes for probability calculation and *T*-Test.

Clearly, the boundaries between coding regions and NCRs are hot spots for tRNA gene translocation in parasitic lice that have fragmented mt genomes. Of the models and mechanisms proposed in previous studies (introduced above in the Background) (Shao et al., 2012; Dong et al., 2014a; Song et al., 2014; Shao et al., 2017; Song et al., 2019), only the inter-minichromosomal recombination model proposed by Song et al., (2014) based on the mt karyotype variation among *Haematopinus* lice can explain tRNA gene translocation between mt minichromosomes. However, it cannot explain why translocated tRNA genes tend to be in the boundaries between coding regions and NCRs. As DNA double-strand breaks (DSBs) are apparently necessary for gene translocation between mt minichromosomes, the boundary regions may have more frequent DSBs than other regions of the mt minichromosomes, possibly due to the specific AT-rich and GC-rich sequence elements in the boundary regions; these sequence elements are conserved among all mt minichromosomes of each species of parasitic louse (Supplementary Fig. 3.2). Apparently, DSBs do not occur randomly. Tubbs et al. showed that in mouse and human cells, DSBs occur mostly in

AT-rich regions, in particular at homopolymeric (dA/dT) tracts (Tubbs et al., 2018). Sinai et al. showed that integration of AT-dinucleotide rich sequences into human chromosomes created recurrent gaps and breaks at the integration sites, due to a significantly increased tendency to fold into branched secondary structures at these sites (Sinai et al., 2019). Baudat et al. showed that a histone methyl transferase, PRDM9, is a major determinant of meiotic recombination hotspots in humans and mice, and furthermore, the human PRDM9 binding sites contain a 13-mer GC-rich motif and are DSB hotspots (Baudat et al., 2010). Sundararajan et al. showed that DSB hotspots have a strong tendency to occur in genes with high GC content in maize (Sundararajan et al., 2016). He et al. (2017) investigated the DNA sequence of DSB hotspots and identified a 20-bp-long GC-rich degenerate DNA sequence motif (named Maize Hotspot Sequence) present in about 72 % of genic hotspots in maize. In general, noncanonical DNA structures such as hairpins and Gquadruplexes render a genome prone to damage including
DSBs (Puget et al., 2019). We counted the number of translocated tRNA genes in sucking lice at each end of the mt non-coding regions (Tables 3, 5 and 7) – the distribution is rather even (20 vs. 20) between the two ends where the AT-rich and GC-rich motifs were found respectively. Our observed pattern of mt tRNA gene translocation in sucking lice aligns well with the evidence presented in these experimental studies on DSB hotspots. Future studies on DSBs and DNA repair in parasitic lice may help understand further the mechanisms of mt tRNA gene translocation observed in sucking lice in the present study.

3.5. Conclusions

We sequenced the fragmented mt genome of the Asian grey shrew louse, *Polyplax reclinata*, which comprises 11 circular minichromosomes. We also inferred the ancestral mt karyotype of *Polyplax* lice based on the data available to date. We found that tRNA genes are entirely responsible for mt karyotype variation among the three species of *Polyplax* lice sequenced to date. Furthermore, tRNA gene translocation in these *Polyplax* lice is much more frequent than that of protein-coding and rRNA genes; all tRNA gene translocation observed occurs between different types of minichromosomes and is directional towards the boundaries with the control region. A similar pattern of tRNA gene translocation is also conserved in other sucking lice with fragmented mt genomes. Our results show that inter-minichromosomal tRNA gene translocation orientated.

Table 3.7. Translocated tRNA genes in non-*Polyplax* sucking lice relative to the most recent common ancestor (MRCA) of sucking lice

tRNA gene	MRCA of sucking lice	Non-*Polyplax* sucking lice	Species	Translocated	Next to NCR
N	*atp8-atp6-N*	*E-cob-S1-S2-**N***	*Microthoracius praelongiceps*[9]	Yes	Yes
		*cob-S1-**N**-E-M*	*Pediculus schaeffi*[7]	Yes	No
		*S1-**N**-E*	*Pediculus capitis*[2]	Yes	No
			Pediculus humanus[2]		
		*T-nad1-Q-**N**-C*	*Pedicinus badii*[12]	Yes	No
		*T-nad1-Q-**N**-G-nad3-W*	*Pedicinus obtusus*[12]	Yes	No
E	*E-cob-S1-S2*	*cob-S1-N-**E**-M*	*Pediculus schaeffi*[7]	Yes	No
		*F-nad6-**E**-M*	*Pthirus pubis*[2]	Yes	No
		*S1-N-**E***	*Pediculus capitis*[2]	Yes	Yes
			Pediculus humanus[2]		
S1	*E-cob-S1-S2*	*D-Y-cox2-**S1**-S2-P-cox3-A*	*Haematopinus apri*[3]	Yes	No
			Haematopinus suis[3]		
			Haematopinus asini[6]		
		***S1**-N-E*	*Pediculus capitis*[2]	Yes	Yes
			Pediculus humanus[2]		
S2	*E-cob-S1-S2*	*D-Y-cox2-S1-**S2**-P-cox3-A*	*Haematopinus apri*[3]	Yes	No
			Haematopinus suis[3]		
			Haematopinus asini[6]		
		*G-nad3-V-W-**S2***	*Pthirus pubis*[2]	Yes	Yes
I	*I-cox1-L2*	*P-nad2-**I***	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		
			Pthirus pubis[2]		
		*R-nad4L-P-cox3-**I***	*Pedicinus obtusus*[12]	Yes	Yes
L2	*I-cox1-L2*	***L2**-rrnS-C*	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		
D	*D-Y-cox2-nad6*	*T-**D**-H-R-nad4L*	*Pthirus pubis*[2]	Yes	No
		*T-**D**-H*	*Pediculus schaeffi*[7]	Yes	No
			Pediculus capitis[2]		
			Pediculus humanus[2]		
		*G-nad3-W-**D**-A-V*	*Pedicinus badii*[12]	Yes	No
		***D**-A-M-F-nad6*	*Pedicinus obtusus*[12]	Yes	Yes
P	*R-nad4L-P-cox3-A*	***P**-nad2-I*	*Pthirus pubis*[2]	Yes	Yes
			Pediculus schaeffi[7]		
			Pediculus capitis[2]		
			Pediculus humanus[2]		
A	*R-nad4L-P-cox3-A*	*G-nad3-W-D-**A**-V*	*Pedicinus badii*[12]	Yes	No
		*D-**A**-M-F-nad6*	*Pedicinus obtusus*[12]	Yes	No

Table 3.7. Translocated tRNA genes in non-*Polyplax* sucking lice relative to the most recent common ancestor (MRCA) of sucking lice (continued)

tRNA gene	MRCA of sucking lice	Non-*Polyplax* sucking lice	Species	Translocated	Next to NCR
Q	*Q-nad1-T-G-nad3-W*	***Q**-N-E*	*Pediculus humanus*[2]	Yes	Yes
		*T-nad1-**Q**-N-C*	*Pedicinus badii*[12]	Yes	No
		*T-nad1-**Q**-N-G-nad3-W*	*Pedicinus obtusus*[12]	Yes	No
T	*Q-nad1-T-G-nad3-W*	***T**-nad1-Q-N-C*	*Pedicinus badii*[12]	Yes	Yes
		***T**-nad1-Q-N-G-nad3-W*	*Pedicinus obtusus*[12]	Yes	Yes
		*D-Y-cox2-**T***	*Hoplopleura kitti*[5]	Yes	Yes
		*R-nad4L-P-cox3-A-**T***	*Hoplopleura akanezumi*[5]	Yes	Yes
		***T**-D-H-R-nad4L*	*Pthirus pubis*[2]	Yes	Yes
		***T**-D-H*	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		
		***T**-nad1-Q-N-C*	*Pedicinus badii*[12]	Yes	Yes
		***T**-nad1-Q-N-G-nad3-W*	*Pedicinus obtusus*[12]	Yes	Yes
G	*Q-nad1-T-G-nad3-W*	***G**-nad4L-V*	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus capitis[2]		
W	*Q-nad1-T-G-nad3-W*	*C-nad6-**W**-L2*	*Hoplopleura kitti*[5]	Yes	No
			Hoplopleura akanezumi[5]		
		*G-nad3-V-**W**-S2*	*Pthirus pubis*[2]	Yes	No
H	*H-nad5-F*	*T-D-**H**-R-nad4L*	*Pthirus pubis*[2]	Yes	No
		*T-D-**H***	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		
F	*H-nad5-F*	***F**-nad6-E-M*	*Pthirus pubis*[2]	Yes	Yes
		*M-**F**-nad6*	*Pedicinus badii*[12]	Yes	No
		*D-A-M-**F**-nad6*	*Pedicinus obtusus*[12]	Yes	No
C	*rrnS-C*	***C**-nad6-W-L2*	*Hoplopleura kitti*[5]	Yes	Yes
			Hoplopleura akanezumi[5]		
		*T-nad1-Q-N-**C***	*Pedicinus badii*[12]	Yes	Yes
		*atp8-atp6-**C***	*Pedicinus obtusus*[12]	Yes	Yes
M	*M-L1-rrnL-V*	*R-nad4L-nad6-**M***	*Haematopinus apri*[3]	Yes	Yes
			Haematopinus suis[3]		
		*F-nad6-E-**M***	*Pthirus pubis*[2]	Yes	Yes
		*cob-S1-N-E-**M***	*Pediculus schaeffi*[7]	Yes	Yes
		***M**-F-nad6*	*Pedicinus badii*[12]	Yes	Yes
		*D-A-**M**-F-nad6*	*Pedicinus obtusus*[12]	Yes	No
L1	*M-L1-rrnL-V*	***L1**-rrnS-C*	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		

Table 3.7. Translocated tRNA genes in non-*Polyplax* sucking lice relative to the most recent common ancestor (MRCA) of sucking lice (continued)

tRNA gene	MRCA of sucking lice	Non-*Polyplax* sucking lice	Species	Translocated	Next to NCR
			Pediculus humanus[2]		
V	*M-L1-rrnL-V*	*G-nad3-**V**-W-S2*	*Pthirus pubis*[2]	Yes	No
		*G-nad3-W-D-A-**V***	*Pedicinus badii*[12]	Yes	Yes
		*nad2-Y-cox2-**V***	*Pedicinus obtusus*[12]	Yes	Yes
		*G-nad4L-**V***	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		

Notes: Superscript numbers indicate relevant articles listed in the References. Translocation events shared by congeneric species are grouped together and counted as single events. Genes underlined have opposite orientation of transcription to other genes.

Table 3.8. Statistical analysis of expected probability and observed probability of tRNA gene translocation in non-*Polyplax* sucking lice relative to the most recent common ancestor of sucking lice

tRNA genes	Expected probability of translocation to be adjacent to NCR if translocation is random	Observed probability of translocation to be adjacent to NCR		Expected probability of translocation to an intergenic region if translocation is random	Observed probability of translocation to an intergenic region
N	21/45=46.67%	20%		24/45=53.33%	80%
E	21/45=46.67%	33.3%		24/45=53.33%	66.7%
S1	22/46=47.83%	50%		23/46=52.17%	50%
S2	21/45=46.67%	50%		24/45=53.33%	50%
I	21/45=46.67%	100%		24/45=53.33%	0%
L2	21/45=46.67%	100%		24/45=53.33%	0%
D	21/45=46.67%	25%		24/45=53.33%	75%
P	22/46=47.83%	100%		23/46=52.17%	0%
A	21/45=46.67%	0%		24/45=53.33%	100%
Q	21/45=46.67%	33.3%		24/45=53.33%	66.7%
T	22/46=47.83%	100%		23/46=52.17%	0%
G	22/46=47.83%	100%		23/46=52.17%	0%
W	21/45=46.67%	0%		24/45=53.33%	100%
H	21/45=46.67%	50%		24/45=53.33%	50%
F	21/45=46.67%	33.3%		24/45=53.33%	66.7%
C	21/45=46.67%	100%		24/45=53.33%	0%
M	21/45=46.67%	80%		24/45=53.33%	20%
L1	22/46=47.83%	100%		23/46=52.17%	0%
V	21/45=46.67%	75%		24/45=53.33%	25%
Average	46.9%	60.5%		53.1%	39.5%
Two tailed *t*-value	-1.60738			1.61464	
p-value	0.116709			0.11512	
Significant at $p < 0.01$?	No			No	

Notes: Refer to Table 3.7 for tRNA gene translocation in non-*Polyplax* sucking lice. Refer to Table 3.4 notes for probability calculation and *T*-Test.

Acknowledgements

We thank Fan Song for assistance with an initial phylogenetic tree analysis.

3.6. References

Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations**.** ***Nucleic Acids Research***, 38(suppl_2), W7-W13.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. ***Nucleic Acids Research***, 25(17), 3389-3402.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., ... & De Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. ***Science***, 327(5967), 836-840.

Boore, J. L. (1999). Animal mitochondrial genomes. ***Nucleic Acids Research***, 27(8), 1767-1780.

Chin TH. (1999) ***Taxonomy and fauna of sucking lice (Anoplura) in China***. Beijing Science Press.

Dong, W. G., Song, S., Jin, D. C., Guo, X. G., & Shao, R. (2014a). 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(1), 1-12.

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 Genomics***, 15(1), 1-13.

Durden, L. A., & Musser, G. G. (1994a). 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.

Durden, L. A., & Musser, G. G. (1994b). The mammalian hosts of the sucking lice (Anoplura) of the world: a host-parasite list. ***Bulletin of the Society for Vector Ecology***, 19(2), 130-168.

Fu, Y. T., Dong, Y., Wang, W., Nie, Y., Liu, G. H., & Shao, R. (2020). Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse *Pedicinus obtusus* and colobus louse *Pedicinus badii*. ***Genomics***, 112(6), 4924-4933.

Gish, W., & States, D. J. (1993). Identification of protein coding regions by database similarity search. ***Nature Genetics***, 3(3), 266-272.

Herd, K. E., 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 Genomics***, 16(1), 1-8.

He, Y., Wang, M., Dukowic-Schulze, S., Zhou, A., Tiang, C. L., Shilo, S., ... & Pawlowski, W. P. (2017). Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. ***Proceedings of the National Academy of Sciences***, 114(46), 12231-12236.

Irony-Tur Sinai, M., Salamon, A., Stanleigh, N., Goldberg, T., Weiss, A., Wang, Y. H., & Kerem, B. (2019). AT-dinucleotide rich sequences drive fragile site formation. ***Nucleic Acids Research***, 47(18), 9685-9695.

Jiang, H., Barker, S. C., & Shao, R. (2013). Substantial variation in the extent of mitochondrial genome fragmentation among blood-sucking lice of mammals. ***Genome Biology and Evolution***, 5(7), 1298-1308.

Johnson, K. P., Nguyen, N. P., Sweet, A. D., Boyd, B. M., Warnow, T., & Allen, J. M. (2018). Simultaneous radiation of bird and mammal lice following the K-Pg boundary. ***Biology Letters***, 14(5), 20180141.

Kambhampati, S., & Smith, P. T. (1995). PCR primers for the amplification of four insect mitochondrial gene fragments. ***Insect Molecular Biology***, 4(4), 233-236.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. ***Molecular Biology and Evolution***, 30(4), 772-780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. ***Bioinformatics***, 28(12), 1647-1649.

Kim KC. (1988). Evolutionary parallelism in Anoplura and eutherian mammals. In: Service MW, editor. ***Biosystematics of Haematophagous Insects***. Oxford: Claredon Press; Vol. 37. p. 91–114.

Kittler, R., Kayser, M., & Stoneking, M. (2003). Molecular evolution of *Pediculus humanus* and the origin of clothing. ***Current Biology***, 13(16), 1414-1417.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. ***Molecular Biology and Evolution***, 35(6), 1547.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., ... & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. ***Bioinformatics***, 23(21), 2947-2948.

Laslett, D., & Canbäck, B. (2008). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. ***Bioinformatics***, 24(2), 172-175.

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 Evolutionary Biology***, 10(1), 1-15.

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. ***Nucleic Acids Research***, 25(5), 955-964.

Ottoni, C., Girdland Flink, L., Evin, A., Geörg, C., De Cupere, B., Van Neer, W., ... & Larson, G. (2013). Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. ***Molecular Biology and Evolution***, 30(4), 824-832.

Puget, N., Miller, K. M., & Legube, G. (2019). Non-canonical DNA/RNA structures during transcription-coupled double-strand break repair: roadblocks or bona fide repair intermediates? ***DNA Repair***, 81, 102661.

Qumsiyeh, M. B. (1994). Evolution of number and morphology of mammalian chromosomes. ***Journal of Heredity***, 85(6), 455-465.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. ***Trends in Ecology & Evolution***, 16(7), 351-358.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. ***Bioinformatics***, 19(12), 1572-1574.

Shao, R., Kirkness, E. F., & Barker, S. C. (2009). The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. ***Genome Research***, 19(5), 904-912.

Shao, R., Zhu, X. Q., Barker, S. C., & Herd, K. (2012). Evolution of extensively fragmented mitochondrial genomes in the lice of humans. ***Genome Biology and Evolution***, 4(11), 1088-1101.

Shao, R., Barker, S. C., Li, H., Song, S., Poudel, S., & Su, Y. (2015). Fragmented mitochondrial genomes in two suborders of parasitic lice of eutherian mammals (Anoplura and Rhynchophthirina, Insecta). ***Scientific Reports***, 5(1), 1-11.

Shao, R., Li, H., Barker, S. C., & Song, S. (2017). The mitochondrial genome of the guanaco louse, *Microthoracius praelongiceps*: insights into the ancestral mitochondrial karyotype of sucking lice (Anoplura, Insecta). ***Genome Biology and Evolution***, 9(2), 431-445.

Smith, V. S., Ford, T., Johnson, K. P., Johnson, P. C., Yoshizawa, K., & Light, J. E. (2011). Multiple lineages of lice pass through the K–Pg boundary. ***Biology Letters***, 7(5), 782-785.

Song, S. D., Barker, S. C., & Shao, R. (2014). Variation in mitochondrial minichromosome composition between blood-sucking lice of the genus *Haematopinus* that infest horses and pigs. ***Parasites & Vectors***, 7(1), 1-8.

Song, F., Li, H., Liu, G. H., Wang, W., James, P., Colwell, D. D., ... & Shao, R. (2019). Mitochondrial genome fragmentation unites the parasitic lice of eutherian mammals. ***Systematic Biology***, 68(3), 430-440.

Sundararajan, A., Dukowic-Schulze, S., Kwicklis, M., Engstrom, K., Garcia, N., Oviedo, O. J., ... & Mudge, J. (2016). Gene evolutionary trajectories and GC patterns driven by recombination in Zea mays. ***Frontiers in Plant Science***, 7, 1433.

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.

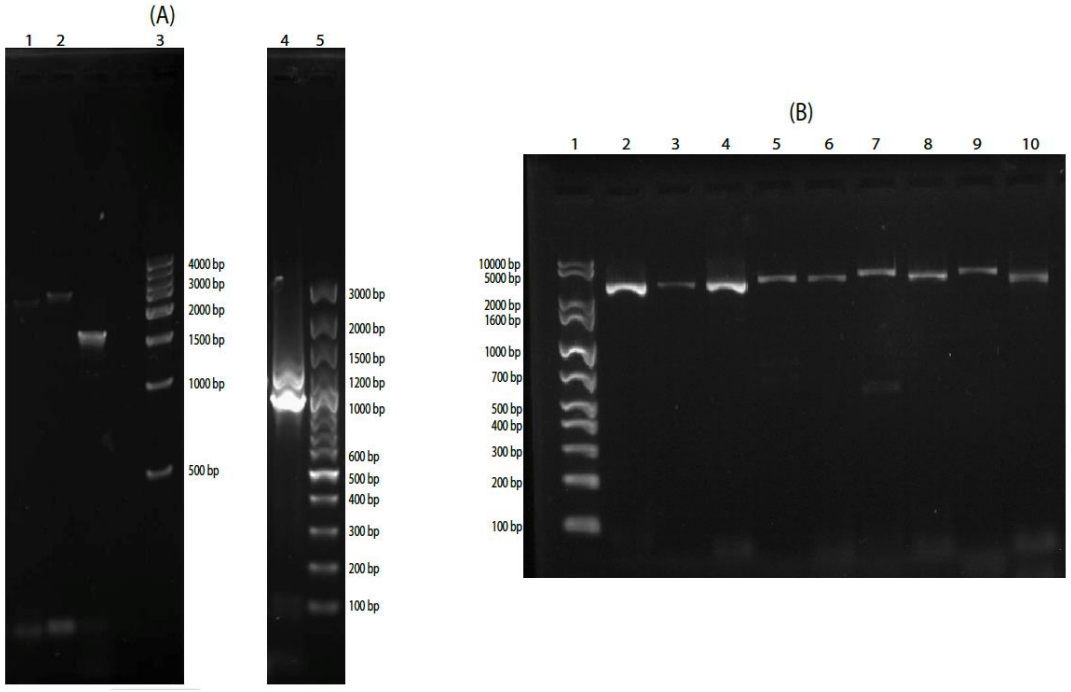
Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. ***Systematic Biology***, 56(4), 564-577.

Toups, M. A., Kitchen, A., Light, J. E., & Reed, D. L. (2011). Origin of clothing lice indicates early clothing use by anatomically modern humans in Africa. ***Molecular Biology and Evolution***, 28(1), 29-32.

Tubbs, A., Sridharan, S., van Wietmarschen, N., Maman, Y., Callen, E., Stanlie, A., ... & Nussenzweig, A. (2018). Dual roles of poly (dA: dT) tracts in replication initiation and fork collapse. ***Cell***, 174(5), 1127-1142.

Wurster, D. H., & Benirschke, K. (1970). Indian muntjac, *Muntiacus muntjak*: a deer with a low diploid chromosome number. ***Science***, 168(3937), 1364-1366.

3.7. Supplementary files



Supplementary Fig. 3.1. PCR amplicons generated from the mitochondrial minichromosomes of the Asian grey shrew louse, *Polyplax reclinata*. (A) Lane 1: amplicon of *S_1-S_2-rrnS-C* minichromosome generated with the primer pair 12S364F-12S364R. Lane 2: amplicon of *M-L_1-rrnL-V* minichromosome generated with the primer pair 16S364F-16S364R. Lane 3: 500 bp Ladder (Tiangen) (band size in bp indicated). Lane 4: amplicons of coding regions of all mt minichromosomes generated with the primer pair 364F-364R. Lane 5: GeneRuler 100 bp DNA Ladder (Thermo Scientific) (band size in bp indicated). The lane to the right of lane 2 is irrelevant to this manuscript. The lane to the left of lane 3 is empty. (B) Lane 1: 1 kb ladder (Tiangen) (band size in bp indicated); Lanes 2-10: amplicons of individual mt minichromosomes, *T-D-Y-cox2-nad6-A*, *R-nad4L-P-cox3*, *Q-nad2-N*, *<u>nad1</u>-G-nad3-W* (gene underlined has opposite transcription orientation to other genes), *K-nad4*, *H-nad5-F*, *E-cob-I*, *cox1-L_2*, *atp8-atp6*.

Supplementary Fig. 3.2. Alignment of the non-coding region (*NCR*) sequences of 11 mitochondrial minichromosomes of the Asian grey shrew louse, *Polyplax reclinata*. The primer pair, 364F and 364R, were used to amplify the coding regions of the 11 minichromosomes (see also Supplementary Fig. 3.1). Asterisk symbol "*" indicates conserved nucleotides; hyphen "-" indicates absent nucleotides.

Supplementary Table 3.1. The primers used to amplify the mitochondrial genes, minichromosomes and coding regions of the Asian grey shrew louse, *Polyplax reclinata*.

Primer	Sequence	Target gene, region or minichromosome
12SA	TACTATGTTACGACTTAT	*rrnS* gene
12SB	AAACTAGGATTAGATACCC	*rrnS* gene
16SF	TTAATTCAACATCGAGGTCGCAA	*rrnL* gene
Lx16SR	GACTGTGCTAAGGTAGCATAAT	*rrnL* gene
12S364F	ATGTCTTGTACATCGCTGTTTAGACACACC	*S_1-S_2-rrnS-C*minichromosome
12S364R	CTCGTGCAGTTTCGTTTATGAACATATGCC	*S_1-S_2-rrnS-C* minichromosome
16S364F	AAGAAGGCTTAGAGTTACCTGGGAGGGGGC	*M-L_1-rrnL-V* minichromosome
16S364R	GGGTCTTCTCGTCCCTCTATGGCATTTGAGC	*M-L_1-rrnL-V* minichromosome
364F	CAGGGTATAGAGGGCGCTCTGGATTGTAAG	Entire coding regions of all minichromosomes
364R	CTTTCCCCCCCCAAAAAGGAGTCAGAGAC	Entire coding regions of all minichromosomes
364atp6F	GCTCCTATTGGACTTTCACCTTTTTTGGTG	*atp8-atp6* minichromosome
364atp6R	GTTGGAAGGAAGTGAGCTAAAGTTCACTC	*atp8-atp6* minichromosome
364cobF	GTTACCATGAGGTCAGATGTCCTTTTGAGG	*E-cob-I* minichromosome
364cobR	GTACCCGAGGAAAGCAGTAGCTATTAAAAC	*E-cob-I* minichromosome
364cox1F	GTGGATAGACGAGCCTACTTTACTAGAGC	*cox1-L_2* minichromosome
364cox1R	CCCAACAGTAAATATATGGTGTGCCCACAC	*cox1-L_2* minichromosome
364cox2F	GAAGAAGTGGTCTTTCCTGAAATCAGGTC	*T-D-Y-cox2-nad6-A* minichromosome
364cox2R	GAGACAATGCAAAGAAGGGAATGCTAAAGC	*T-D-Y-cox2-nad6-A* minichromosome
364cox3F	GCCCTTCGCTTTAGTTGAAGACATTTTTCC	*R-nad4L-P-cox3* minichromosome
364cox3R	GATTGTTCCCACAATCACATGAAGTCCGTG	*R-nad4L-P-cox3* minichromosome
364nad1F	CCCAAACGGTAGTACTGATAAGAACAAAGC	*nad1-G-nad3-W* minichromosome
364nad1R	CAGACTCCTAACTCCTCAGGAGTAATTGC	*nad1-G-nad3-W* minichromosome
364nad2F	CTGTGAGTCATTCTTGGGATGAAAGTCAG	*Q-nad2-N* minichromosome
364nad2R	GCAGAATATACTAAAGCCCGTCGAATAGAC	*Q-nad2-N* minichromosome
364nad4F	CTCATTATGAGGGAGGTTTGTATCTGCTC	*K-nad4* minichromosome
364nad4R	GGCAAACTAAAAAGTCCCAAAATGCCGAC	*K-nad4* minichromosome
364nad5F	GCTTTGTGAAGAGTTAGGGCAGATAGGTC	*H-nad5-F* minichromosome
364nad5R	CTCGATTGACCCTGAGAAGGAGAATAGAAG	*H-nad5-F* minichromosome

Chapter 4

Fragmented mitochondrial genomes of seal lice (family Echinophthiriidae) and gorilla louse (family Pthiridae): frequent minichromosomal splits and a host switch of lice between seals

Published as:

Dong, Y., Zhao, M., & Shao, R. (2022).

Fragmented mitochondrial genomes of seal lice (family Echinophthiriidae) and gorilla louse (family Pthiridae): frequent minichromosomal splits and a host switch of lice between seals**.**

BMC Genomics, 23(1), 1-13.

Manuscript: Appendix III

4.1. Abstract

The mitochondrial (mt) genomes of 15 species of sucking lice from seven families have been studied to date. These louse species have highly dynamic, fragmented mt genomes that differ in the number of minichromosomes, the gene content, and gene order in a minichromosome between families and even between species of the same genus. In the present study, we analysed the publicly available data to understand mt genome fragmentation in seal lice (family Echinophthiriidae) and gorilla louse, *Pthirus gorillae* (family Pthiridae), in particular the role of minichromosome split and minichromosome merger in the evolution of fragmented mt genomes. We show that 1) at least three ancestral mt minichromosomes of sucking lice have split in the lineage leading to seal lice, 2) one minichromosome ancestral to primate lice has split in the lineage to the gorilla louse, and 3) two ancestral minichromosomes of seal lice have merged in the lineage to the northern fur seal louse. Minichromosome split occurred 15-16 times in total in the lineages leading to species in six families of sucking lice investigated. In contrast, minichromosome merger occurred only four times in the lineages leading to species in three families of sucking lice. Further, three ancestral mt minichromosomes of sucking lice have split multiple times independently in different lineages of sucking lice. Our analyses of mt karyotypes and gene sequences also indicate the possibility of a host switch of crabeater seal louse to Weddell seals. We conclude that: 1) minichromosome split contributes more than minichromosome merger in mt genome fragmentation of sucking lice, and 2) mt karyotype comparison helps understand the phylogenetic relationships between sucking louse species.

4.2. Introduction

Sucking lice (parvorder Anoplura) are obligate ectoparasites of eutherian mammals and feed exclusively on their host blood (De Moya et al., 2021). There are over 500 species of sucking lice, classified into 50 genera and 15 families (Durden & Musser, 1994). Sucking lice have a flattened body to avoid host grooming, elongated mouthparts for blood-sucking, and sharp claws to climb and hold on host hair (Kim & Ludwig, 1978). Unlike most animals, which have single-chromosome mitochondrial (mt) genomes (Boore, 1999), the sucking lice investigated to date have fragmented mt genomes, each with 9-20 minichromosomes (Shao et al., 2009; Shao et al., 2012; Shao et al., 2017; Song et al., 2014). Fifteen species of sucking lice from seven of the 15 families have been sequenced for complete or near complete mt genomes (Dong et al., 2014a; Dong et al., 2014b; Dong et al., 2021; Fu & Dong et al., 2020; Jiang et al., 2013; Shao et al., 2009; Shao et al., 2012;

Shao et al., 2017; Song et al., 2014). Human head louse (*Pediculus humanus capitis*, family Pediculidae) and human body louse (*Pediculus humanus corporis*) have the most fragmented mt genomes with 20 minichromosomes (Shao et al., 2009; Shao et al., 2012), followed by chimpanzee louse (*Pediculus schaeffi*) with 18 minichromosomes (Herd et al., 2015). Human pubic louse (*Pthirus pubis*, family Pthiridae) has 15 minichromosomes (*trnN* gene not identified) (Fu & Dong et al., 2020; Shao et al., 2012). Colobus louse (*Pedicinus badii*, family Pedicinidae) and macaque louse (*Pedicinus obtusus*) have 14 and 12 minichromosomes, respectively (Fu & Dong et al., 2020). Guanaco louse (*Microthoradus praelongiceps*, family Microthoraciidae) has 12 minichromosomes (Shao et al., 2017). Three rodent lice of the family Polyplacidae (*Polyplax asiatica*, *Polyplax spinulosa* and *Polyplax reclinata*) each have 11 minichromosomes (Dong et al., 2014b; Dong et al., 2021). Two rodent lice of the family Hoplopleuridae (*Hoplopleura akanezumi* and *Hoplopleura kitti*) each have at least 11 minichromosomes with five and three genes not identified, respectively (Dong et al., 2014a). Wild pig louse (*Haematopinus apri*, family Haematopinidae), domestic pig louse (*Haematopinus suis*), and horse louse (*Haematopinus asini*) have the least fragmented mt genomes, each with nine minichromosomes (Jiang et al., 2013; Song et al., 2014).

No species in the other eight families of sucking lice have been studied for their mt genome organization. These families are Echinophthiriidae (lice of seals), Enderleinellidae (lice of squirrels), Linognathidae (lice of cattle, sheep and goats), Hamophthiriidae (lice of colugos), Hybophthiridae (lice of aardvarks), Neolinognathidae (lice of elephant shrews), Pecaroecidae (lice of peccaries), and Ratemiidae (lice of horses, donkeys and zebras) (Durden & Musser, 1994; Kim, 1985). In the present study, we analysed the publicly available Sequence Read Archive (SRA) data to understand the mt genome fragmentation in five species of seal lice (*Antarctophthirus carlinii*, *A. lobodontis*, *A. microchir*, *Lepidophthirus macrorhini*, *Proechinophthirus fluctus*) in the family Echinophthiriidae, and the gorilla louse (*Pthirus gorillae*) in the family Pthiridae. We also investigated the role of minichromosome split and merger in the evolution of the highly dynamic mt genome organization observed in sucking lice.

The family Echinophthiriidae has five genera and 13 species (Durden & Musser, 1994). The genus *Antarctophthirus* has seven species, while the other four genera have one to two species. Seal lice are unique among parasitic lice in living in the marine environment (Kim, 1975; Kim et al., 1975). Ancestral seal lice were thought to be terrestrials but evolved unique morphology to adapt to marine

life together with their hosts (Kim, 1985; Kim et al., 1975; Leonardi et al., 2019). The family Pthiridae has only one genus *Pthirus* with two species: *Pthirus pubis* (human pubic louse) and *Pthirus gorillae* (gorilla louse) (Durden & Musser, 1994). These two species diverged 3-4 million years ago (MYA), possibly due to host switch from gorillas to humans (Reed et al., 2007). The mt genome of *Pthirus pubis* has been sequenced previously (Shao et al., 2012); however, the mt genome of *Pthirus gorillae* has not been studied. We find that: 1) seal lice differ substantially from sucking lice in other families in mt genome organization; 2) gorilla louse has a more fragmented mt genome than human pubic louse; 3) minichromosome split occurred in both lineages leading to seal lice and gorilla louse; 4) minichromosome merger occurred in the lineage to the northern fur seal louse; and 5) likely a host switch of louse occurred between crabeater seals and Weddell seals.

4.3. Materials and methods

4.3.1. Retrieval and assembly of Sequence Read Archive (SRA) data of seal lice and gorilla louse

The SRA data of five species of seal lice (*Antarctophthirus carlinii*; *Antarctophthirus lobodontis*, *Antarctophthirus microchir*, *Lepidophthirus macrorhini* and *Proechinophthirus fluctus*) and the gorilla louse (*Pthirus gorillae*) were retrieved from NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra/). These SRA data were produced with Illumina platforms in whole genome sequencing projects and were deposited by researchers in Kevin Johnson's group at the University of Illinois, Urbana-Champaign (Leonardi et al., 2019). The whole dataset of sequence-reads of each species was imported into and assembled in Geneious 11.0.2 (Kearse et al., 2019). A subset of 15,000 sequence reads in each dataset was extracted and assembled *de novo* to obtain seed contigs for mt gene search; the number of contigs to be displayed was set at 100. The consensus sequences of these 100 contigs were searched in a batch in NCBI "Non-redundant protein sequences (NR)" database using BLASTx and default parameters to identify protein sequence matches (McGinnis & Madden, 2004). The consensus sequences that matched significantly (E-value $<10^{-15}$) to the mt protein sequences of sucking lice (parvorder Anoplura) in the NR database were used as reference sequences to assemble the full-length coding region and its adjacent non-coding regions (200-320 bp upstream and downstream) of each mt minichromosome; the entire set of SRA sequence-reads of each species was explored in coding and non-coding region assembly. We did not attempt to assemble the full-length non-coding region of each minichromosome with SRA data because: 1) the non-coding region is highly similar in sequence among different mt minichromosomes of a sucking louse species (Dong et al., 2014a; Dong et al.,

2014b; Dong et al., 2021; Fu & Dong et al., 2020; Jiang et al., 2013; Shao et al., 2009; Shao et al., 2012; Shao et al., 2015; Shao et al., 2017; Song et al., 2014), and 2) the SRA Illumina sequence reads are too short (100-160 bp each) for such attempt. The key assembly parameters were: 1) minimum overlap 60 bp for *Proechinophthirus fluctus* (SRA sequence reads 100 bp each) and 100 bp for other four species of seal lice (SRA sequence reads 150 bp each) and gorilla louse (SRA sequence reads 160 bp each); and 2) minimum identity 95%. Because of the high sequence similarity among the non-coding regions of different minichromosomes within a species, once two or more minichromosomes were assembled, the conserved sequence motifs including the hallmark AT-rich motif and GC-rich motif in the non-coding regions were identified by sequence alignment and used as references to identify and assemble the remaining minichromosomes. An AT-rich motif upstream from coding region and a GC-rich motif downstream from coding region are present in the mt minichromosomes of all sucking lice and the chewing lice of eutherian mammals sequenced to date (Dong et al., 2014a; Dong et al., 2014b; Dong et al., 2021; Fu & Dong et al., 2020; Herd et al., 2015; Jiang et al., 2013; Shao et al., 2009; Shao et al., 2012; Shao et al., 2015; Shao et al., 2017; Song et al., 2014; Song et al., 2019). The steps described above were repeated multiple times for each species until no more additional mt genes could be found. The protein-coding genes and rRNA genes in mt minichromosomes were identified by BLAST search in NCBI database; tRNA genes were identified by tRNA-scan (Lowe & Chan, 2016) and ARWEN (Laslett & Canbäck, 2008).

For *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*, their partial *elongation factor 1-α* gene sequence, *18S* and *28S* rRNA gene sequences were generated by Illumina sequence read assembly, using the available sequences of *Haematopinus eurysternus* (accession numbers HM171457 and HM171381) and *Echinophthirus horr*idus (accession number KX810111) as initial reference sequences. We used Geneious 11.0.2 (Kearse et al., 2012) and started with *medium sensitivity* in reference assembly to find the most conserved regions of these three genes. Then we used the SRA sequence reads that were mapped to the conserved regions as references to initiate and extend the contigs for *elongation factor 1-α*, *18S* rRNA and *28S* rRNA genes of *Antarctophthirus carlinii* and *Antarctophthirus lobodontis.* These genes were then verified by BLAST search in NCBI database.

4.3.2. Phylogenetic analyses

Twenty-two species of parasitic lice (Supplementary Table 4.1) were included in our phylogenetic analysis: 1) the five species of seal lice and the gorilla louse reported in the present study; 2) 15 sucking louse species reported in previous studies; and 3) the elephant louse (*Haematomyzus elephantis*) as the outgroup. The sequences of five mt protein-coding genes (*cob*, *cox1*, *cox2*, *cox3*, *nad4*) were used in our phylogenetic analysis. These five genes were aligned individually using MAFFT 7.471, then concatenated into a single file after removing the poorly aligned sites using Gblocks 0.91b. Two methods were used in our phylogenetic analysis: 1) maximum likelihood (ML) with IQ-Tree (Nguyen et al., 2014), and 2) Bayesian inference method (BI) with MrBayes 3.2.6 (Ronquist & Huelsenbeck, 2003). Model test were done in IQ-TREE (Kalyaanamoorthy et al., 2017) and the best-fit model is TIM+F+R4. For ML analysis, the bootstrap replicates were set at 1,000. For BI analyses, four independent Markov Chains were run for five million MCMC generations, sampling a tree every 100 generations. This analysis was run until the average standard deviation of split frequencies was lower than 0.001. The ML tree and BI tree were drawn with Figtree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

4.3.3. Inferring the ancestral mitochondrial karyotype of seal lice

To infer the ancestral mt karyotype of seal lice, we used a parsimony method described in Shao et al. (2017). The ancestral mt karyotype of seal lice was inferred based on a comparison of mt minichromosomal characters between the five species of seal lice and the inferred most recent common ancestor (MRCA) of sucking lice (Shao et al., 2017). We inferred a minichromosomal character to be ancestral to seal lice if: 1) the character was present in one or more of the five species of seal lice and also in the MRCA of sucking lice; or 2) the character is present in all the five species of seal lice. For example, the character *E-cob* in a minichromosome is present in four of the five seal louse species and also in the MRCA of sucking lice (Fig. 4.1). Thus, *E-cob* in a minichromosome is inferred to be ancestral to seal lice. The MRCA of sucking lice also has S_1-S_2 in the same minichromosome with *E-cob*; these two tRNA genes are most likely translocated to other minichromosome(s) in the MRCA of seal lice as they are not with *E-cob* in any of the five seal louse species (Fig. 4.1). Taking another example, *K-nad4* in a minichromosome is present in all the five seal louse species thus it is inferred to be ancestral to seal lice (Fig. 4.1). In this case, *K-nad4* is also present in the MRCA of sucking lice, which reinforces the inference of *K-nad4* minichromosome to be ancestral to seal lice.

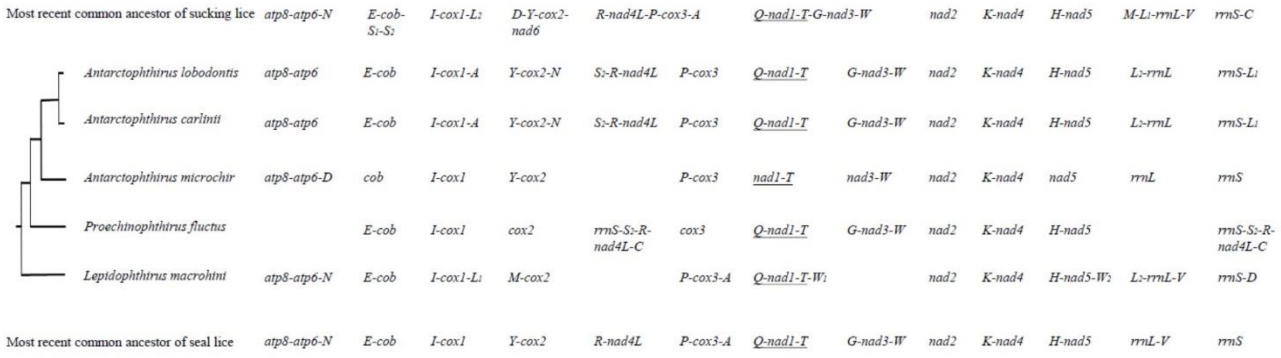


Fig. 4.1. Inference of the ancestral mitochondrial karyotype of seal lice. Minichromosomal characters were inferred to be ancestral to seal lice if a character was present in: 1) one or more of the five species of seal lice and also in the MRCA of sucking lice; or 2) all the five species of seal lice. Genes underlined have opposite transcription orientation to those not underlined.

4.4. Results

4.4.1. Mitochondrial minichromosomes of *Lepidophthirus macrorhini* – louse of southern elephant seal (*Mirounga leonine*)

The Illumina data of *Lepidophthirus macrorhini* (SRR5809351) from SRA database contains 47,978,079 paired-end sequence reads; each sequence read is 150 bp in size. We assembled these sequence reads and identified 27 of the 37 mt genes typical of bilateral animals. These genes are on 11 minichromosomes; each minichromosome contains a single protein-coding or rRNA gene with none or up to three tRNA genes except *atp8-atp6-N* minichromosome, which contains two protein-coding genes (Fig. 4.2A, Supplementary Table 4.2). The genes in all the minichromosomes have the same orientation of transcription relative to the non-coding regions except for *Q-nad1-T-W* minichromosome, in which three genes (underlined) have the opposite transcription orientation to all other genes (Fig. 4.2A, Supplementary Table 4.2). *trnW* gene were identified in two minichromosomes: *Q-nad1-T-W1* minichromosome and *H-nad5-W2* minichromosome. The pairwise identity between these two *trnW* genes is 50% (Supplementary Table 4.3); both have TCA anticodon and can form clove-leafed secondary structure (Supplementary Fig. 4.2). *trnW1* has a higher identity to *trnW* of other seal lice (55.4% to 69.1%) than *trnW2* has (42.6% to 55.1%), indicating *trnW1* is more likely the original copy of *trnW*. However, it is unclear whether *trnW2* is duplicated from *trnW1* or from other tRNA genes (Supplementary Table 4.3). We obtained ~300 bp non-coding sequence both upstream and downstream from the coding region of each

minichromosome (Supplementary Table 4.2). A conserved AT-rich (80.3%) motif (56 bp) and a conserved GC-rich (60.7%) motif (63 bp) were found in the non-coding sequences of all the minichromosomes (Fig. 4.2A, Supplementary Fig. 4.2). Ten mt genes were not identified in our analysis of *Lepidophthirus macrorhini*: *nad3, nad4L, nad6, trnC, trnF, trnG, trnR, trnS$_1$, trnS$_2$* and *trnY*. The annotated mt minichromosomes of *Lepidophthirus macrorhini* were available in GenBank (accession numbers MW803094-104).

4.4.2. Mitochondrial minichromosomes of *Proechinophthirus fluctus* – louse of northern fur seal (*Callorhinus ursinus*)

The Illumina data of *Proechinophthirus fluctus* (SRR5308138) from SRA database contains 5,819,192 paired-end sequence reads; each sequence read is 100 bp in length. We assembled these sequence reads and identified 22 of the 37 typical mt genes. These genes are on 10 minichromosomes; each minichromosome contains a single protein-coding or rRNA gene with none or up to three tRNA genes (Fig. 4.2B, Supplementary Table 4.4). The genes in all the minichromosomes have the same orientation of transcription relative to the non-coding regions except for *Q-nad1-T* minichromosome; these three genes have the opposite transcription orientation to all other genes (Fig. 4.2B). We obtained ~200 bp non-coding sequence both upstream and downstream from the coding region of each minichromosome (Supplementary Table 4.4). A conserved AT-rich (71.8%) motif (85 bp) and a conserved GC-rich (61.7%) motif (81 bp) were found in the non-coding sequences of all the minichromosomes (Fig. 4.2B, Supplementary Fig. 4.3). Fifteen mt genes were not identified in our analysis of the SRA data of *Proechinophthirus fluctus*: *atp6, atp8, nad6, rrnL, trnA, trnD, trnF, trnL$_1$, trnL$_2$, trnM, trnN, trnP, trnS1, trnV* and *trnY*. The annotated mt minichromosomes of *Proechinophthirus fluctus* were available in GenBank (accession numbers MW803105-114).

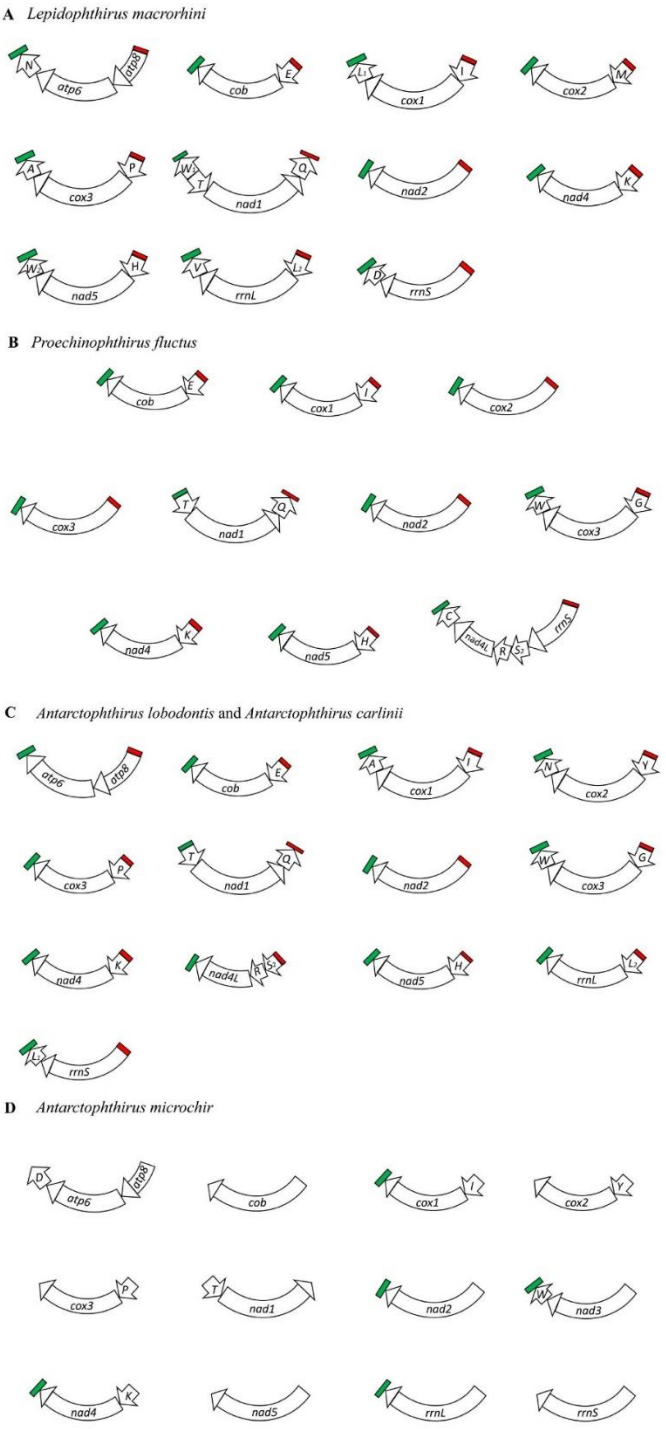


Fig. 4.2. Mitochondrial genomes of five seal louse species: (A) southern elephant seal louse, *Lepidophthirus macrorhini*; (B) northern fur seal louse, *Proechinophthirus fluctus*; (C) Weddell seal louse, *Antarctophthirus carlinii*, and crabeater seal louse, *Antarctophthirus lobodontis*; and (D) Australian sea lion louse, *Antarctophthirus microchir*. Conserved AT-rich motifs are in red and

conserved GC-rich motifs are in green. In *Antarctophthirus microchir*, AT-rich motifs were not identified and GC-rich motifs were identified in only five of the 12 minichromosomes. Names and transcription orientation of genes are indicated in the coding region. tRNA genes are indicated with their single-letter abbreviations of the corresponding amino acids.

4.4.3. Mitochondrial minichromosomes of *Antarctophthirus carlinii* - louse of Weddell seal (*Leptonychotes weddellii*), and *Antarctophthirus lobodontis* - louse of crabeater seal (*Lobodon carcinophagus*)

The Illumina data of *Antarctophthirus carlinii* (SRR5809348) and *Antarctophthirus lobodontis* (SRR5809349) from SRA database contains 39,054,456 and 45,005,741 paired-end sequence reads respectively; each sequence read is 150 bp. We assembled these sequence reads and identified 30 of the 37 typical mt genes on nine minichromosomes in each species (Fig. 4.2C). Each minichromosome contains a single protein-coding gene with none or up to two tRNA genes (Fig. 4.2C, Supplementary Table 4.5 and 4.6). The gene content and gene arrangement in each mt minichromosome are the same between these two species; furthermore, the identity of each homologous gene is 94.3% to 100% between the two species. The genes in all the minichromosomes have the same orientation of transcription relative to the non-coding regions except for *Q-nad1-T* minichromosome, in which the three genes have the opposite transcription orientation to all other genes (Fig. 4.2C). We obtained ~300 bp non-coding sequence both upstream and downstream from the coding region of each minichromosome (Supplementary Table 4.5 and 4.6). In *Antarctophthirus carlinii*, a conserved AT-rich (60%) motif (60 bp) and a conserved GC-rich (67.4%) motif (46 bp) were found in the non-coding sequences of all the minichromosomes (Supplementary Fig. 4.4). Conserved AT-rich (68.7%) motif (65 bp) and GC-rich (64.9%) motif (37 bp) were also found in *Antarctophthirus lobodontis* (Supplementary Fig. 4.5). Seven mt genes were not identified in our analysis of the SRA data of these two *Antarctophthirus* species: *nad6, trnC, trnD, trnF, trnM, trnS1* and *trnV*. The annotated mt minichromosomes of *Antarctophthirus carlinii* (accession numbers MW803073-81) and *Antarctophthirus lobodontis* (accession numbers MW803064-72) were available in GenBank.

4.4.4. Mitochondrial minichromosomes of *Antarctophthirus microchir* – louse of Australian sea lion (*Neophoca cinerea*)

The Illumina data of *Antarctophthirus microchir* (SRR5809347) from SRA database contains 43,650,933 paired-end sequence reads. Each sequence read is 150 bp in size. We assembled these sequence reads and identified 20 of the 37 typical mt genes. These genes are on 12 minichromosomes; each minichromosome contains a single protein-coding or rRNA gene with none or one tRNA gene except *atp8-atp6-D* minichromosome (Fig. 4.2D, Supplementary Table 4.7). We obtained 250 bp non-coding sequence both upstream and downstream from the coding region of each minichromosome (Supplementary Table 4.7). We identified a conserved GC-rich motif in the non-coding sequences of five of the 12 minichromosomes; however, we were unable to identify conserved AT-rich motif in the non-coding sequences of any minichromosomes (Fig. 4.2D, Supplementary Table 4.7, Supplementary Fig. 4.6). Full-length non-coding region sequences are needed to investigate whether the unidentified conserved motifs are in the middle section of the non-coding regions of *Antarctophthirus microchir*. Seventeen mt genes were not identified in our analysis of the SRA data of *Antarctophthirus microchir*: *nad4L, nad6, trnA, trnC, trnE, trnF, trnG, trnH, trnL1, trnL2, trnM, trnN, trnQ, trnR, trnS1, trnS2* and *trnV*. The annotated mt minichromosomes of *Antarctophthirus microchir* were available in GenBank (accession numbers MW803082-93).

4.4.5. Mitochondrial minichromosomes of *Pthirus gorillae* – louse of western gorilla (*Gorilla gorilla*)

The Illumina data of *Pthirus gorillae* (SRR5088474) from SRA database contains 60,425,294 paired-end sequence reads; each sequence read is 160 bp in length. We assembled these sequence reads and identified 36 of the 37 typical mt genes; these genes are on 17 minichromosomes (Fig. 4.3). *trnN* was the only gene not identified in our analysis. Fourteen of the 17 minichromosomes contain a single protein-coding or rRNA gene with none or up to five tRNA genes. Of the other three minichromosomes, *atp8-atp6* minichromosome contains no tRNA gene whereas *trnA* minichromosome and *trnC* minichromosome each have only a tRNA gene (Fig. 4.3, Supplementary Table 4.8). The genes in all the minichromosomes have the same orientation of transcription relative to the non-coding regions (Fig. 4.3). A 65-bp non-coding sequence upstream from *rrnS* has 58.5% identity to *trnL1* and 60% identity to *trnL2* (Fig. 4.4A). As *trnL* gene is upstream from *rrnS* in the human pubic louse (*Pthirus pubis*), human head louse (*Pediculus humanus capitis*) and human body louse (*Pediculus humanus corporis*) (Shao et al., 2009), the 65 bp non-coding sequence upstream from *rrnS* in *Pthirus gorillae* is very likely a degenerate *trnL* gene. Three other

regions are likely degenerate genes too: 1) a 218-bp sequence between *trnF* and *trnT* has 100% identity with a 5' section of *nad6* (Fig. 4.4B); 2) a 320-bp sequence upstream from *trnA* has 100% identity with a middle section of *rrnS* (Fig. 4.4C); and 3) a 320-bp sequence upstream from *trnC* has 53.3% identity to 3' section of *rrnL* (Fig. 4.4D). We obtained ~320 bp non-coding sequence both upstream and downstream from the coding region of each minichromosome (Supplementary Table 4.8). A conserved AT-rich (76.6%) motif (96 bp) was found in the non-coding sequences of all the minichromosomes except *trnA* minichromosome and *trnC* minichromosome (Fig. 4.3; Supplementary Fig. 4.7). A GC-rich (76%) motif (25 bp) was found in the non-coding sequences of all the minichromosomes (Fig. 4.3; Supplementary Fig. 4.7). The annotated mt minichromosomes of *Pthirus gorillae* are available in GenBank (accession numbers MW803115-131).

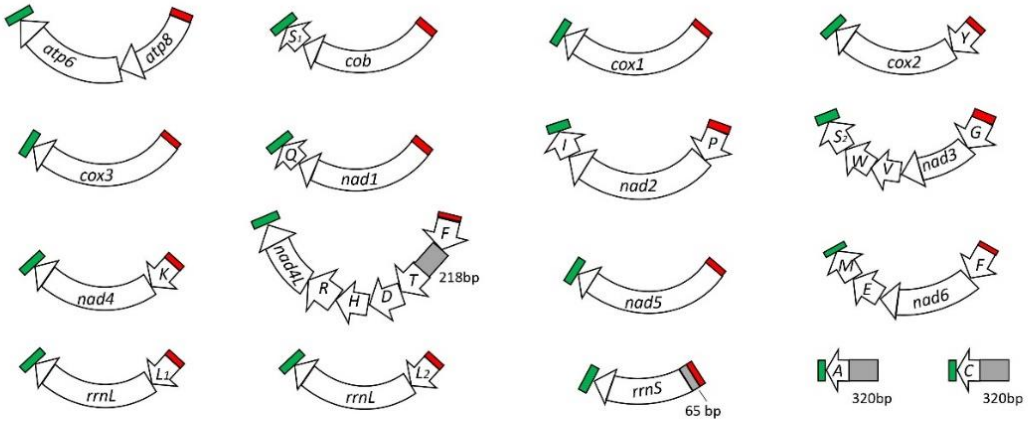


Fig. 4.3. Mitochondrial genome of the gorilla louse, *Pthirus gorillae.* Conserved AT-rich motifs are in red; conserved GC-rich motifs are in green; and degenerate genes are in grey. Names and transcription orientation of genes are indicated in the coding region.

4.4.6. Phylogeny of sucking lice based on mitochondrial gene sequences

We reconstructed the phylogeny of sucking lice (Anoplura) to assist the inference of ancestral mt karyotype of seal lice (family Echinophthiriidae). Both the Bayesian and maximum likelihood (ML) trees strongly support the monophyly of Echinophthiriidae and each of the other seven families of

sucking lice (Fig. 4.5; Supplementary Fig. 4.8). The relationships among the five species of seal lice are resolved with strong support. The three *Antarctophthirus* species are most closely related to each other, among which the Weddell seal louse (*Antarctophthirus carlinii*) and the crabeater seal louse (*Antarctophthirus lobodontis*) are sister to one another, both having very short branch length relative to that of the Australian sea lion louse (*Antarctophthirus microchir*). The *Antarctophthirus* species are more closely related to the northern fur seal louse (*Proechinophthirus fluctus*) than to the southern elephant seal louse (*Lepidophthirus macrorhini*) (Fig. 4.5; Supplementary Fig. 4.8).

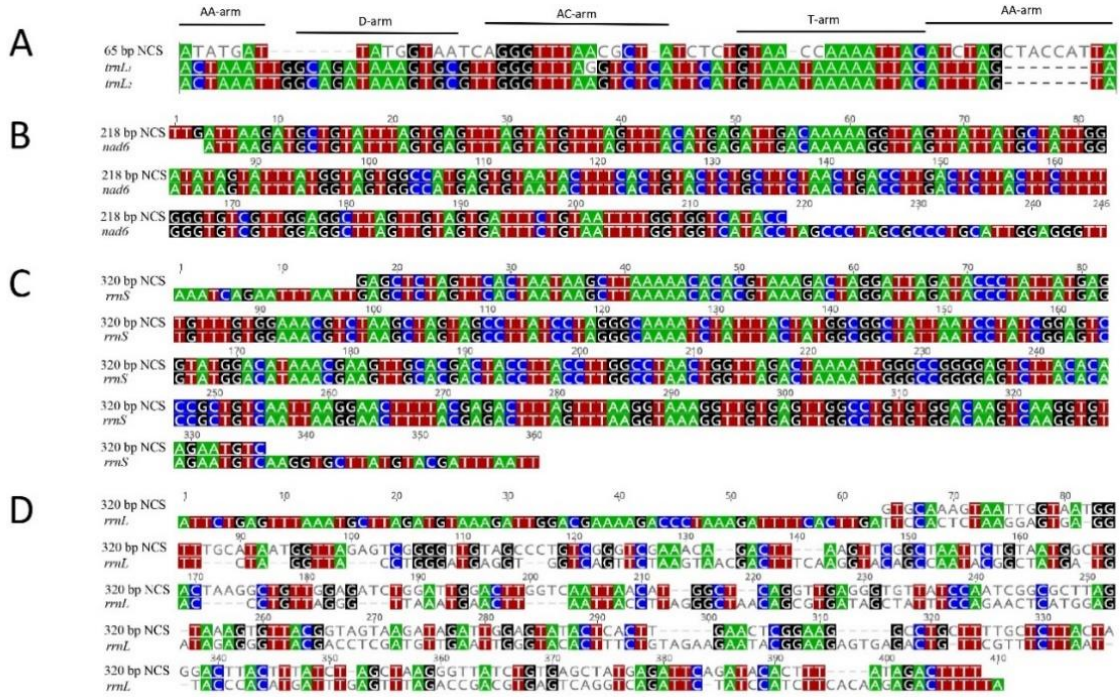


Fig. 4.4. Sequence alignment of degenerate genes with their corresponding full-length genes. (A): the 65-bp non-coding sequence (NCS) in *rrnS* minichromosome has 58.5% identity to *trnL1* gene and 60% identity to *trnL2* gene. (B): The 218-bp intergenic NCS in *F-T-D-H-R-nad4L* minichromosome has 215 bp identical to a 5' section of *nad6* gene (nucleotide 1 to 215). (C) The 320-bp NCS in *A* minichromosome is identical to a middle section of *rrnS* gene (nucleotide 246 to 565). (D) The 320-bp NCS in *C* minichromosome has 53.3% identity to 3' section of *rrnL* gene (nucleotide 838 to 1189).

The monophyly of primate lice (families Pediculidae, Pthiridae and Pedicinidae) is strongly supported and the relationships among the primate lice are resolved with strong support (Fig. 4.5; Supplementary Fig. 4.8). The gorilla louse (*Pthirus gorillae*) is most closely related to the human

pubic louse (*Pthirus pubis*); the *Pthirus* species are more closely related to the *Pediculus* species of human and chimpanzee than to the *Pedicinus* species of monkeys. There is support in both Bayesian and ML trees for: 1) the primate lice to be most closely related to the pig and horse lice (family Haematopinidae); and 2) the rodent lice in the families Polyplacidae and Hoplopleuridae to be closely related. There is support in the Bayesian tree but not in the ML tree for: 1) Pthiridae, Pediculidae, Pedicinidae, Haematopinidae, Hoplopleuridae and Polyplacidae to be more closely related to each other than to Microthoraciidae or Echinophthiriidae; and 2) Echinophthiriidae to be more closely related the group that contains Pthiridae, Pediculidae, Pedicinidae, Haematopinidae, Hoplopleuridae and Polyplacidae than to Microthoraciidae (Fig. 4.5; Supplementary Fig. 4.8).

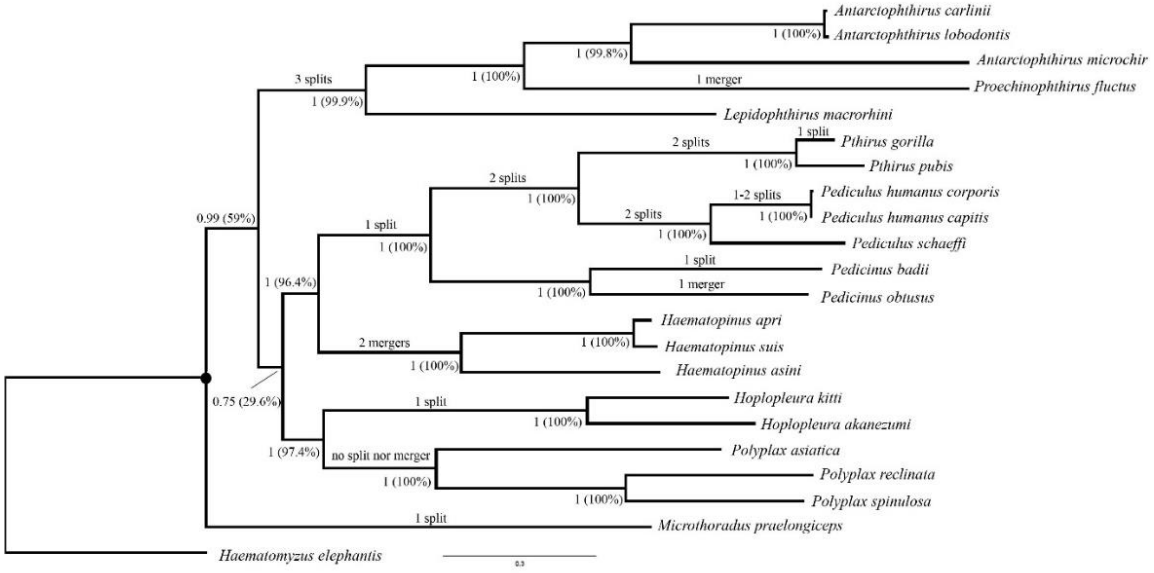


Fig. 4.5. Bayesian phylogenetic tree inferred from the nucleotide sequences of five mitochondrial protein-coding genes (*cob*, *cox1*, *cox2*, *cox3*, *nad4*) of 21 species of sucking lice (Anoplura). The elephant louse, *Haematomyzus elephantis*, was used as the outgroup. Bayesian posterior probability (Bpp) values were indicated near nodes followed by bootstrap support values (in brackets) from maximum likelihood (ML) tree (Supplementary Fig. 4.8).

4.4.7. Inferred ancestral mitochondrial karyotype of the seal lice

The mt minichromosomal information and the sucking louse phylogeny reported above allowed us to infer the partial mt karyotype of the most recent common ancestor of the five species of seal lice.

There were at least 13 minichromosomes in the ancestral mt karyotype of these seal lice (Fig. 4.1). The position and arrangement of 12 of the 13 mt protein-coding genes, two rRNA genes and 14 of the 22 tRNA genes in the ancestral mt karyotype can be inferred based on the minichromosomal information available from the five species of seal lice. The position and arrangement of *nad6* and eight tRNA genes, however, cannot be inferred due to: 1) *nad6* was not identified in any of the five species of seal lice; and 2) the eight tRNA genes are very variable in their arrangement among the five species of seal lice (Fig. 4.1).

4.5. Discussion

4.5.1. Mitochondrial minichromosome split occurred more frequently than minichromosome merger in the lineages leading to seal lice and gorilla louse

Previous studies indicated that both split and merger of mt minichromosomes occurred in sucking lice and were responsible to a large degree to their highly dynamic mt genome organization (Dong et al., 2021; Fu & Dong et al., 2020; Shao et al., 2017). Prior to the current study, split of mt minichromosomes was observed in species from five of the seven families of sucking lice studied: Pediculidae, Pthiridae, Pedicinidae, Microthoraciidae, and Hoplopleuridae, whereas merger of mt minichromosomes was observed in species from two families: Haematopinidae and Pedicinidae (Fu & Dong et al., 2020; Jiang et al., 2013; Shao et al., 2017; Song et al., 2014). Neither split nor merger of mt minichromosome occurred in the family Polyplacidae (Dong et al., 2014b; Dong et al., 2021; Shao et al., 2017). The mt genome organization of other eight families of sucking lice was unknown prior to the current study: Echinophthiriidae (lice of seals), Enderleinellidae (lice of squirrels), Linognathidae (lice of cattle, sheep, and goats), Hamophthiriidae (lice of colugos), Hybophthiridae (lice of aardvarks), Neolinognathidae (lice of elephant shrews), Pecaroecidae (lice of peccaries) and Ratemiidae (lice of horses, donkeys, and zebras).

We showed in the current study that at least three ancestral mt minichromosomes of the sucking lice (Shao et al., 2017) have split further in the lineage leading to seal lice (Echinophthiriidae): 1) *D-Y-cox2-nad6* minichromosome, 2) *R-nad4L-P-cox3-A* minichromosome, and 3) *Q-nad1-T-G-nad3-W* minichromosome (Fig. 4.1). In all the five species of seal lice, *cox2* has its own minichromosome not shared with any other protein-coding gene(s), indicating the split of *D-Y-cox2-nad6* minichromosome occurred in the most recent common ancestor (MRCA) of these seal lice although we could not identify *nad6* in our SRA data analyses (Fig. 4.1). Similarly, *R-nad4L-P-cox3-A*

minichromosome and *Q-nad1-T-G-nad3-W* minichromosome also split in the MRCA of seal lice. *nad4L* has its minichromosome not shared with any other protein-coding gene(s) in *Antarctophthirus lobodontis* and *Antarctophthirus carlinii*, and *cox3* has its minichromosome not shared with any other protein-coding gene(s) in all the five species of seal lice (Fig. 4.1). *nad1* has its minichromosome not shared with any other protein-coding gene(s) in all the five species of seal lice; furthermore, *nad3* has its minichromosomes not shared with any other protein-coding gene(s) in *Antarctophthirus microchir* and *Proechinophthirus fluctus* (Fig. 4.1). We also observed a merger event in the lineage to *Proechinophthirus fluctus* between two ancestral mt minichromosomes inferred for the seal lice: *R-nad4L* and *rrnS* (Fig. 4.1).

The family Pthiridae has a single genus and two species: gorilla louse (*Pthirus gorillae)* and human pubic louse (*Pthirus pubis*); these two species diverged 3-4 MYA (Reed et al., 2007). Previous studies showed that the mt genome of *Pthirus pubis* has 15 minichromosomes with *trnN* gene still not identified (Fu & Dong et al 2020; Shao et al., 2012). In the current study, we showed that the mt genome of *Pthirus gorilla* has 17 minichromosomes (*trnN* not identified either) and is more fragmented than that of *Pthirus pubis* (Fig. 4.3). *cox3-A* minichromosome is ancestral to primate lice and is retained in *Pthirus pubis* (Shao et al., 2012); however, this minichromosome is split into two in *Pthirus gorilla*: one minichromosome contains *cox3* and other has *trnA* (Fig. 4.3).

The frequency of split and merger varies among different families of sucking lice, so does the type of minichromosomes that split and merge

Shao et al. (2017) proposed that split and merger of mt minichromosomes contributed to the complex and dynamic mt genome organization observed in sucking lice (Shao et al., 2017). Shao et al. (2017) reported that: 1) split of mt minichromosomes occurred in sucking louse species from four families after their divergence from the MRCA of sucking lice: Hoplopleuridae, Pthiridae, Pediculidae and Microthoraciidae; 2) merger of mt minichromosomes occurred in species from the family Haematopinidae; and 3) no split nor merger occurred in species from the family Polyplacidae (Shao et al., 2017), which was also confirmed in Dong et al. (2021). Fu & Dong et al. (2020) reported that both split and merger of mt minichromosomes occurred in the family Pedicinidae: split of a minichromosome occurred in the colobus louse *Pedicinus badii* whereas merger of minichromosomes occurred in the macaque louse *Pedicinus obtusus*, relative to the MRCA of higher primate lice. The other eight families of sucking lice were unknown previously for

their mt genome organization: Echinophthiriidae (lice of seals), Enderleinellidae (lice of squirrels), Linognathidae (lice of cattle, sheep and goats), Hamophthiriidae (lice of colugos), Hybophthiridae (lice of aardvarks), Neolinognathidae (lice of elephant shrews), Pecaroecidae (lice of peccaries) and Ratemiidae (lice of horses, donkeys and zebras) (Durden & Musser, 1994; Kim, 1985). In the present study, we showed that split of three minichromosomes occurred in seal lice of the family Echinophthiriidae relative to the MRCA of sucking lice, and merger of two minichromosomes in the lineage leading to the northern fur seal louse, *Proechinophthirus fluctus* (Fig. 4.5). We also showed that split of a minichromosome occurred in the gorilla louse, *Pthirus gorilla*, after its divergence from the human pubic louse, *Pthirus pubis*.

It is apparent that split of minichromosomes occurs much more frequently and in more families than merger of minichromosomes in sucking lice (Fig. 4.5). Furthermore, the frequency of minichromosome split varies from family to family with the two families of great ape lice (Pthiridae, Pediculidae) having the highest number of minichromosome split (8-9 split events) but no split in Haematopinidae and Polyplacidae (Fig. 4.5). It is noteworthy that a few ancestral minichromosomes to sucking lice split multiple times independently in lineages leading to different families. Shao et al. (2017) showed that: 1) *D-Y-cox2-nad6* minichromosome split twice independently, once in Hoplopleuridae, and another time in Pediculidae and Pthiridae; and 2) *Q-nad1-T-G-nad3-W* minichromosome split twice independently, once in Microthoraciidae, and another time in Pediculidae and Pthiridae (Shao et al., 2017). In the current study, we found that these two minichromosomes also split independently in seal lice (Echinophthiriidae). Furthermore, *R-nad4L-P-cox3-A* minichromosome, which split in Pediculidae and Pthiridae (Shao et al., 2017), also split independently in seal lice (Echinophthiriidae).

Much rarer than minichromosome split, minichromosome merger was seen only in species from three families: pig lice and horse louse (Haematopinidae) (Jiang et al., 2013; Shao et al., 2017; Song et al., 2014), macaque louse (Pedicinidae) (Fu & Dong et al., 2020), and northern fur seal louse (Echinophthiriidae). In the pig lice and horse louse, two ancestral minichromosomes to sucking lice, *atp8-atp6-trnN* and *trnK-nad4* merged; two other ancestral minichromosomes, *trnI-cox1-trnL2* and *nad2* also merged (Shao et al., 2017). In the macaque louse, a single merge event occurred between two ancestral minichromosomes to higher primate lice that contain *cox2* and *nad2* genes respectively. In the northern fur seal louse, a single merger event occurred between two

minichromosomes ancestral to seal lice that contain *nad4L* and *rrnS* genes respectively (Fig. 4.1). It is noteworthy that *nad2* minichromosome was involved in two separate merger events whereas the other five minichromosomes were each involved only in a single merger event.

4.5.2. The very high mitochondrial gene identity shared between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* does not support them as separate species

The crabeater seal louse, *Antarctophthirus lobodontis*, was described by Enderlein (1909) and redescribed by Leonardi et al. (2016); the Weddell seal louse, *Antarctophthirus carlinii*, was described by Leonardi et al. (2014). These two species are very similar in morphology (Leonardi et al. 2014; Leonardi et al. 2016). Leonardi et al. (2016) provided two morphological characters to differentiate between these two species: 1) *Antarctophthirus carlinii* has six dorsal posterior long hairs (four marginal and two principal) around the posterior border of the head while *Antarctophthirus lobodontis* has only four marginal long hairs around the posterior border of the head; and 2) *Antarctophthirus lobodontis* has a line of eight spines in the basis of the head and three hairs above the last row of four spines (Leonardi et al. 2016). As reported above, *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* have identical mt karyotypes (Fig. 4.2C), indicating a very close relationship between them. Furthermore, *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* share very high identity of mt genes ranging from 94.3% to 100% (average 99.03%) (Supplementary Table 4.9). We also obtained and compared the partial sequences of three nuclear genes between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*. The sequences of their *elongation factor 1-*α gene (182 bp) are 100% identical; their *18S* sequences (1054 bp) and *28S* rRNA gene sequences (2295 bp) have 99.3% and 99.1% identities respectively (Supplementary Fig. 4.9). Previous studies showed that mt genes of parasitic lice evolved much faster than their hosts. Thus, the identities of mt genes are much lower between parasitic lice than between their hosts (Hafner et al., 1994; Page et al., 1998; Paterson et al., 2000; Johnson et al., 2003). However, this is not the case for *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*. The mt gene identities between these lice are 4.11% higher on average than that between their hosts, the Weddell seal (*Leptonychotes weddellii*) and the crabeater seal (*Lobodon carcinophagus*) - 90.8% to 100% identities with an average of 94.99% (Supplementary Table 4.10). The low genetic divergence between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* does not support them as separate species. Therefore, they are more likely two subspecies than two species. Leonardi et al. (2019) also discussed the possibility that *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*

were the same species based on the low genetic divergence between them. However, based on cophylogenetic analysis, Leonardi et al. (2019) concluded that *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* co-speciated with their hosts thus were different species; this conclusion was not supported by the lower genetic divergence between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* than between their hosts. A relatively recent host switch between the Weddell seal and the crabeater seal is a more plausible explanation for the low genetic divergence observed between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* than co-speciation of *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* with their hosts (Leonardi et al., 2019). Both Weddell seals and crabeater seals have circumpolar distributions in the Antarctic and share the same microhabitats such as land-fast ice and pack ice when they rest and breed (Würsig et al., 2009; Southwell et al., 2012). This could provide ample opportunities for the parasitic lice of one seal species to explore and switch to another seal species. However, it is unclear to us which seal is the original host and which seal is the new host. One possibility is that the host switch occurred from crabeater seal to Weddell seal based on the observation that crabeater seal hosts only one louse species (*Antarctophthirus lobodontis*) whereas Weddell seal hosts two louse species (*Antarctophthirus carlinii* and *Antarctophthirus ogmorhini*) (Durden & Musser, 1994). Both crabeater seal and Weddell seal are in the tribe Lobodontini (subfamily Monachinae), together with Ross seal (*Ommatophoca rossi*) and leopard seal (*Hydrurga leptonyx*) (Durden & Musser, 1994). Weddell seal is the only species in the tribe Lobodontini that hosts two louse species; the other three seal species each host one louse species only ((Durden & Musser, 1994; Leonardi et al., 2014). Further studies on the phylogeny of all louse species of Lobodontini seals should reveal more on the host-parasite relationships in this tribe.

4.6. Conclusions

In this study, we assembled the mt genomes of five species of seal lice and the gorilla louse, and conducted phylogenetic analysis of sucking lice from eight families. We inferred the ancestral mt karyotype of seal lice and analysed the frequency of mt minichromosomal split and merger among sucking lice. Combining mt karyotype comparison and phylogenetic analysis, we concluded that the crabeater seal louse, *Antarctophthirus lobodontis*, and the Weddell seal louse, *Antarctophthirus carlinii*, are likely two subspecies instead of two species. In addition, we also proposed a possibility of a host switch of crabeater seal louse to Weddell seals with gene sequence analysis. We showed that at least three ancestral mt minichromosomes of sucking lice have split in the lineage leading to

seal lice, one minichromosome ancestral to primate lice has split in the lineage leading to gorilla louse, and two ancestral minichromosomes of seal lice have merged in the lineage to the northern fur seal louse. Split of mt minichromosomes occurred 15-16 times in total in the lineages leading to six families of sucking lice studied so far whereas merger of minichromosomes occurred only four times in the lineages leading to three families of sucking lice. Furthermore, three ancestral minichromosomes to sucking lice, *D-Y-cox2-nad6*; *Q-nad1-T-G-nad3-W* and *R-nad4L-P-cox3-A,* have split independently in different lineages of sucking lice. We conclude that: 1) minichromosome split contributes much more than minichromosome merger in mt genome fragmentation of sucking lice, and 2) mt karyotype comparison, in conjunction with gene sequence analysis, helps understand the relationship between sucking louse species.

Acknowledgements

The publicly available SRA data used in this study were produced and deposited by researchers in Kevin Johnson's group at the University of Illinois at Urbana-Champaign. We thank Wei Wang (UniSC) for assistance in checking the morphological characters of *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* in published literature. We thank the associate editor of BMC Genomics and the reviewers for valuable comments that have improved the submitted manuscript greatly.

4.7. References

Boore, J. L. (1999). Animal mitochondrial genomes. ***Nucleic Acids Research***, 27(8): 1767-1780.

De Moya, R. S., Yoshizawa, K., Walden, K. K., Sweet, A. D., Dietrich, C. H., & Kevin P, J. (2021). Phylogenomics of parasitic and nonparasitic lice (Insecta: Psocodea): combining sequence data and exploring compositional bias solutions in next generation data sets. ***Systematic Biology***, 70(4), 719-738.

Dong, W. G., Song, S., Guo, X. G., Jin, D. C., Yang, Q., Barker, S. C., & Shao, R. (2014a). 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: 751.

Dong, W. G., Song, S., Jin, D. C., Guo, X. G., & Shao, R. (2014b). 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(1): 44.

Dong, W. G., Dong, Y., Guo, X. G., & Shao, R. (2021). Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals. ***BMC Genomics***, 22(1): 598.

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***.

Enderlein, G. (1909). Monographie der Robbenläuse, pp 505–516, pls 55–60. In: Enderlein G (ed) ***Die Insekten des Antarktischen Gebietes***. Deutsche Südpolar-Expedition 1901–1903. Zoologie 2:361–528

Fu, Y. T., & Dong, Y., Wang, W., Nie, Y., Liu, G. H., & Shao, R. (2020). Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse *Pedicinus obtusus* and colobus louse *Pedicinus badii*. ***Genomics***, 112(6): 4924-4933.

Hafner, M. S., Sudman, P. D., Villablanca, F. X., Spradling, T. A., Demastes, J. W., & Nadler, S. A. (1994). Disparate rates of molecular evolution in cospeciating hosts and parasites. ***Science***, 265(5175): 1087-1090.

Herd, K. E., 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 Genomics***, 16(1): 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 Biology and Evolution***, 5(7): 1298-1308.

Johnson, K. P., Cruickshank, R. H., Adams, R. J., Smith, V. S., Page, R. D., & Clayton, D. H. (2003). Dramatically elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera). ***Molecular Phylogenetics and Evolution***, 26(2), 231-242.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. ***Nature Methods***, 14(6): 587-589.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. ***Bioinformatics***, 28(12): 1647-1649.

Kim, K. C. (1975). Ecology and morphological adaptation of the sucking lice (Anoplura,

Echinophthiriidae) on the northern fur seal. ***Rapport et Procès verbaux des Réunions du conseil Permanent International pour lExploration de la Mer***, 169: 504-515.

Kim, K. C. (1985). ***Coevolution of parasitic arthropods and mammals***: Wiley.

Kim, K. C., & Ludwig, H. W. (1978). The family classification of the Anoplura. ***Systematic Entomology***, 3(3): 249-284.

Kim, K., Repenning, C., & Morejohn, G. (1975). Specific antiquity of the sucking lice and evolution of otariid seals. ***Rapports et Proces-Verbaux des Reunions***, 169: 544-549.

Laslett, D., & Canbäck, B. (2008). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. ***Bioinformatics***, 24(2): 172-175.

Leonardi, M. S., Poljak, S., Carlini, P., Galliari, J., Bobinac, M., Santos, M., ... & Negrete, J. (2014). *Antarctophthirus carlinii* (Anoplura: Echinophthiriidae), a new species from the Weddell seal *Leptonychotes weddellii*. ***Parasitology Research***, 113(11): 3947-3951.

Leonardi, M. S., Bobinac, M., & Negrete, J. (2016). Redescription of *Antarctophthirus lobodontis* (Anoplura: Echinophthiriidae) from the crabeater seal and identification key for Antarctic lice. ***Polar Biology***, 39(4): 671-676.

Leonardi, M. S., Virrueta Herrera, S., Sweet, A., Negrete, J., & Johnson, K. P. (2019). Phylogenomic analysis of seal lice reveals codivergence with their hosts. ***Systematic Entomology***, 44(4): 699-708.

Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. ***Nucleic Acids Research***, 44(W1), W54-W57.

McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. ***Nucleic Acids Research***, 32(suppl_2): W20-W25.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. ***Molecular Biology and Evolution***, 32(1): 268-274.

Page, R. D., Lee, P. L., Becher, S. A., Griffiths, R., & Clayton, D. H. (1998). A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. ***Molecular Phylogenetics and Evolution***, 9(2), 276-293.

Paterson, A. M., Wallis, G. P., Wallis, L. J., & Gray, R. D. (2000). Seabird and louse coevolution: complex histories revealed by 12S rRNA sequences and reconciliation analyses. ***Systematic Biology***, 49(3): 383-399.

Reed, D. L., Light, J. E., Allen, J. M., & Kirchman, J. J. (2007). Pair of lice lost or parasites

regained: the evolutionary history of anthropoid primate lice. ***BMC Biology***, 5: 7.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. ***Bioinformatics***, 19(12): 1572-1574.

Shao, R., Kirkness, E. F., & Barker, S. C. (2009). The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. ***Genome Research***, 19(5): 904-912.

Shao, R., Zhu, X.-Q., Barker, S. C., & Herd, K. (2012). Evolution of Extensively Fragmented Mitochondrial Genomes in the Lice of Humans. ***Genome Biology and Evolution***, 4(11): 1088-1101.

Shao, R., Barker, S. C., Li, H., Song, S., Poudel, S., & Su, Y. (2015). Fragmented mitochondrial genomes in two suborders of parasitic lice of eutherian mammals (Anoplura and Rhynchophthirina, Insecta). ***Scientific Reports***, 5: 17389.

Shao, R., Li, H., Barker, S. C., & Song, S. (2017). The Mitochondrial Genome of the Guanaco Louse, *Microthoracius praelongiceps*: Insights into the Ancestral Mitochondrial Karyotype of Sucking Lice (Anoplura, Insecta). ***Genome Biology and Evolution***, 9(2): 431-445.

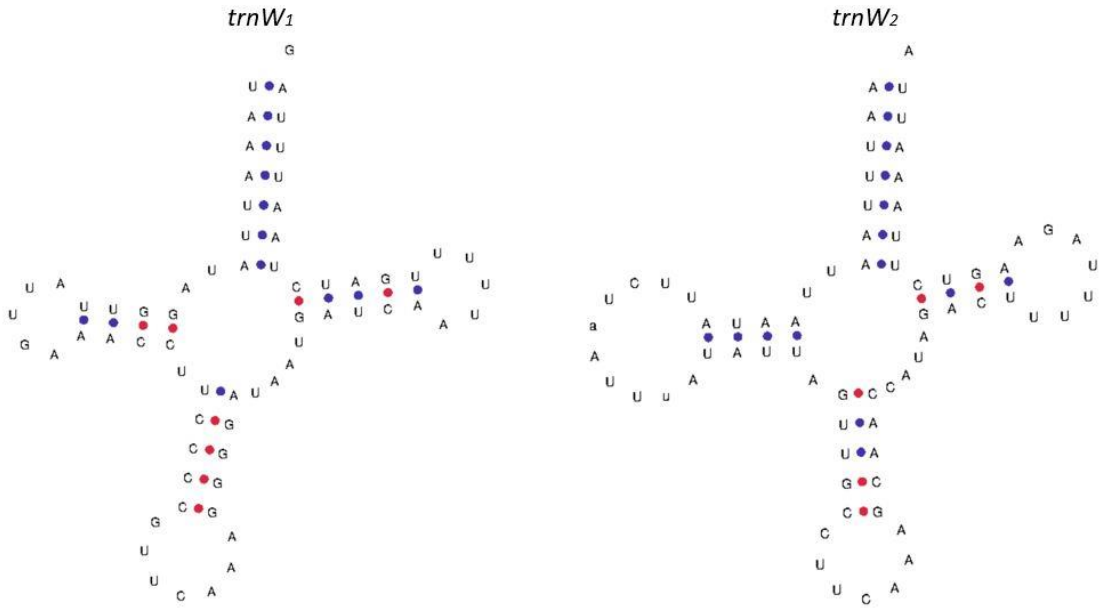
Song, S. D., Barker, S. C., & Shao, R. (2014). Variation in mitochondrial minichromosome composition between blood-sucking lice of the genus *Haematopinus* that infest horses and pigs. ***Parasit Vectors***, 7: 144.

Song, F., Li, H., Liu, G. H., Wang, W., James, P., Colwell, D. D., ... & Shao, R. (2019). Mitochondrial Genome Fragmentation Unites the Parasitic Lice of Eutherian Mammals. ***Systematic Biology***, 68(3): 430-440.

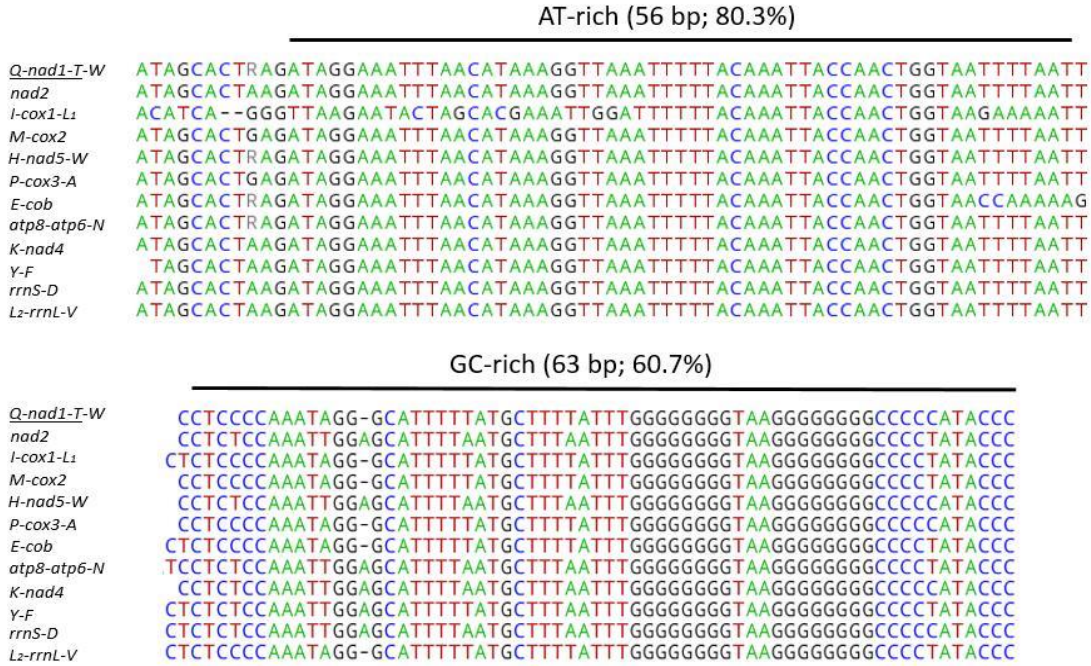
Southwell, C., Bengston, J., Bester, M., Blix, A. S., Bornemann, H., Boveng, P., ... & Trathan, P. (2012). A review of data on abundance, trends in abundance, habitat use and diet of ice-breeding seals in the Southern Ocean. ***Ccamlr Science***, 19: 49-74.

Würsig, B., Perrin, W. F., & Thewissen, J. (2009). ***Encyclopedia of Marine Mammals***: Academic Press.

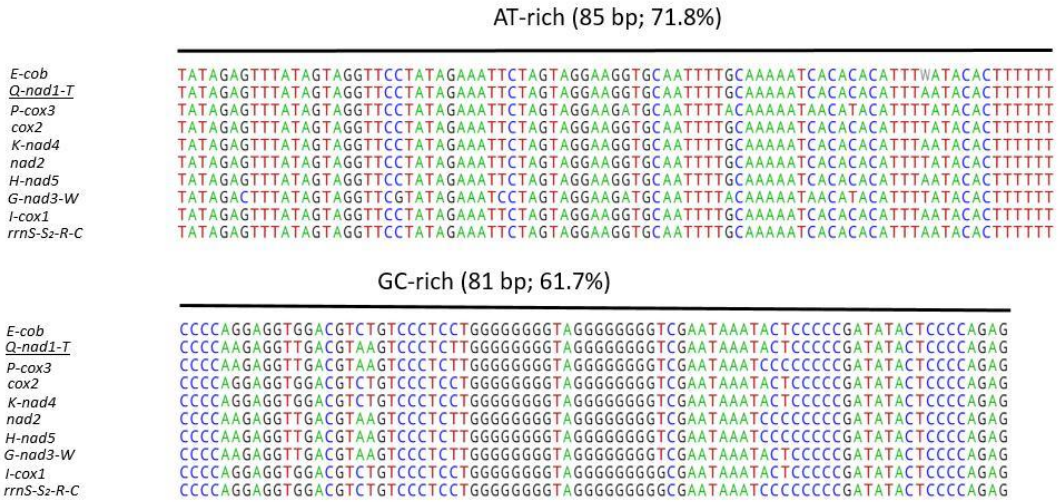
4.8. Supplementary files



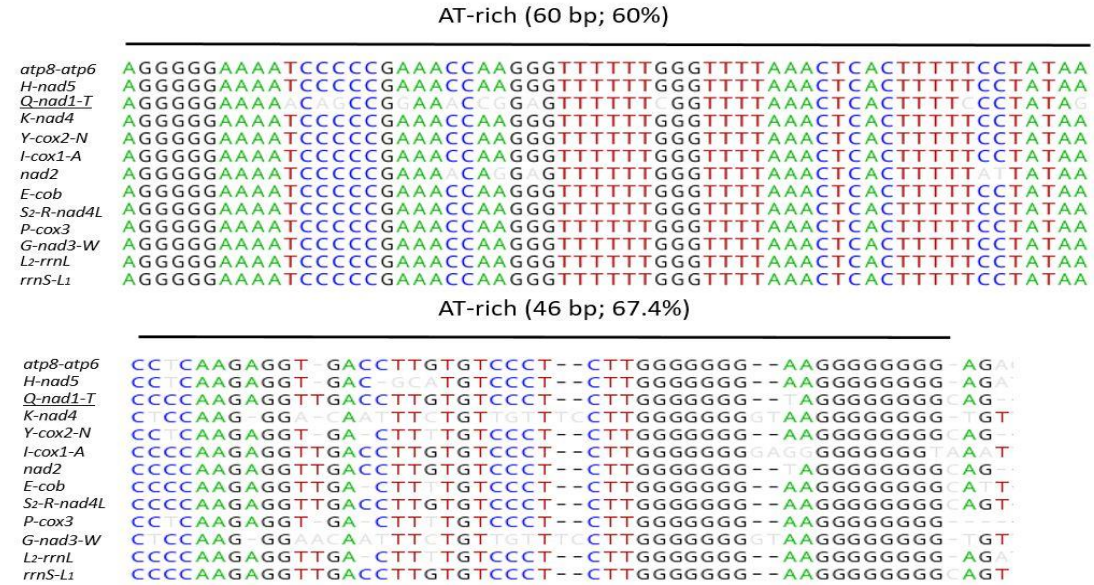
Supplementary Fig. 4.1. Secondary structure inferred with tRNA-Scan from *trnW₁* and *trnW₂* gene sequences of the southern elephant seal louse, *Lepidophthirus macrorhini*.



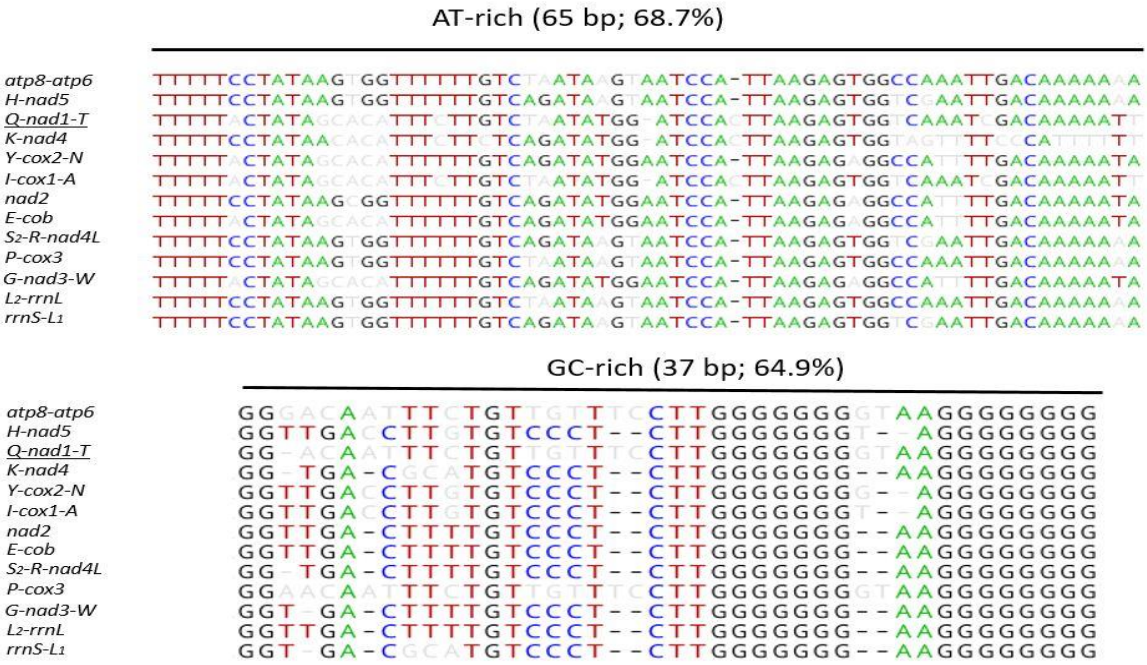
Supplementary Fig. 4.2. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the southern elephant seal louse, *Lepidophthirus macrorhini*.



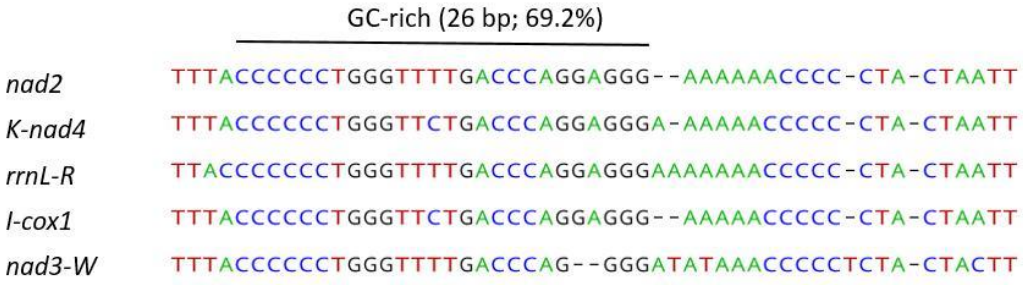
Supplementary Fig. 4.3. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the northern fur seal louse, *Proechinophthirus fluctus*.



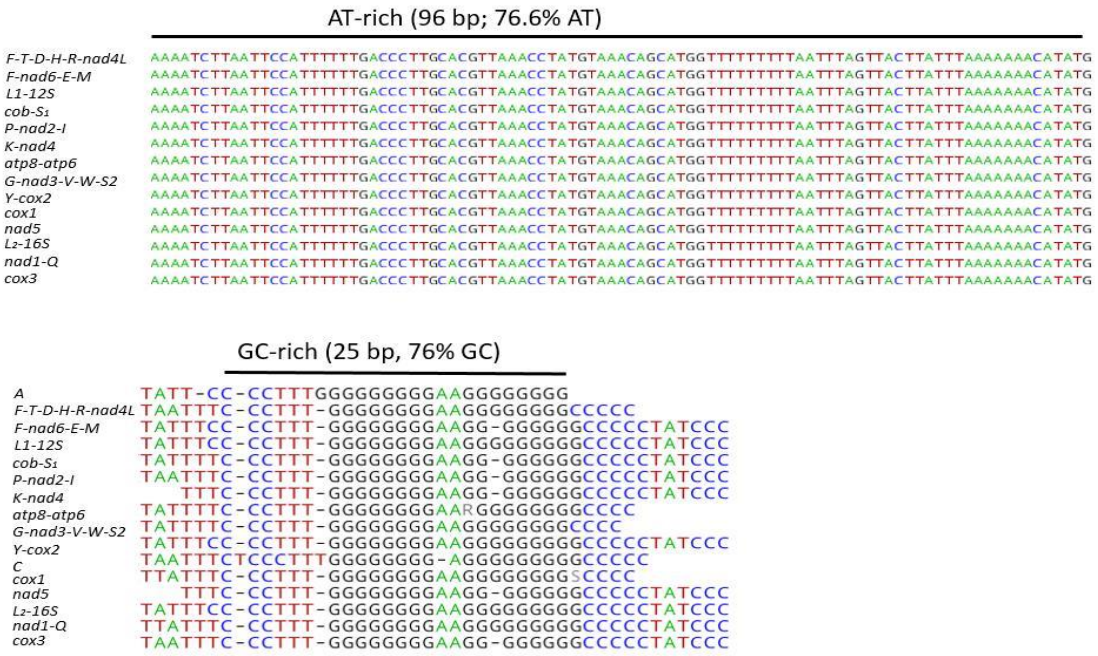
Supplementary Fig. 4.4. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the Weddell seal louse, *Antarctophthirus carlinii*.



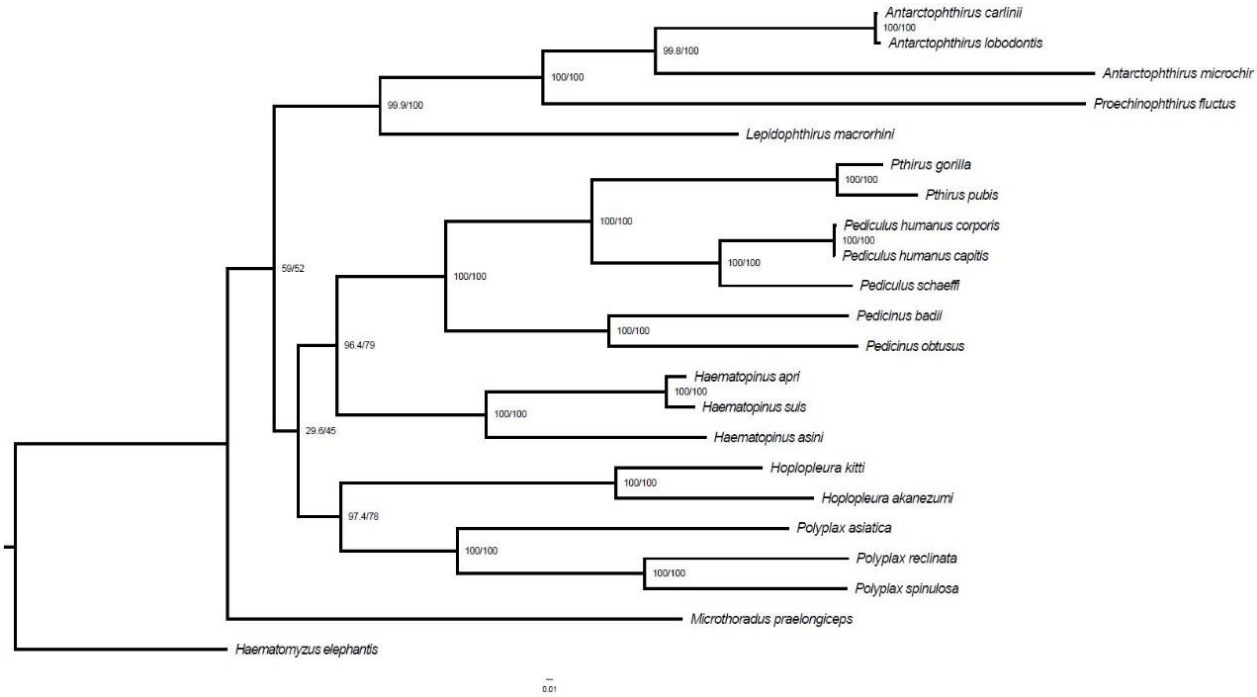
Supplementary Fig. 4.5. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the crabeater seal louse, *Antarctophthirus lobodontis.*



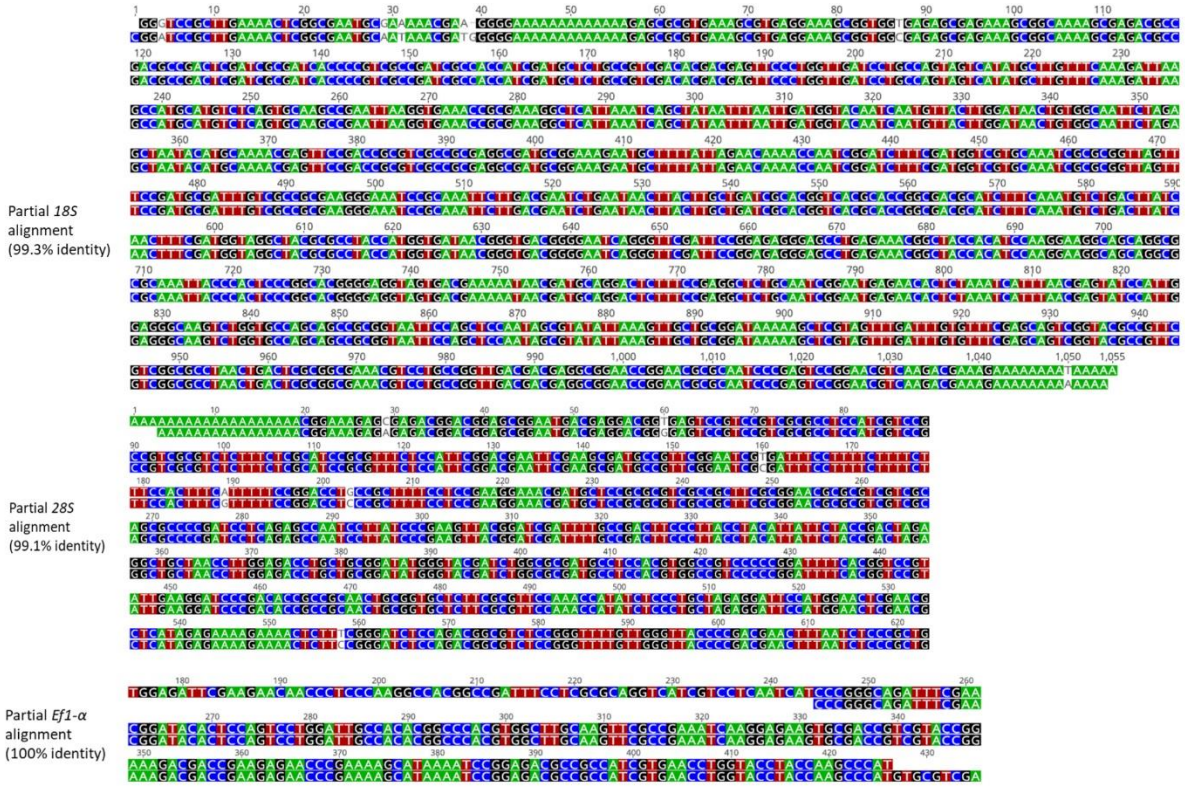
Supplementary Fig. 4.6. Conserved non-coding GC-rich motifs among the mitochondrial minichromosomes of the Australian sea lion louse, *Antarctophthirus microchir*.



Supplementary Fig. 4.7. Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the gorilla louse, *Pthirus gorillae.*



Supplementary Fig. 4.8. Phylogenetic relationships among 21 species of sucking lice (Anoplura) inferred by maximum likelihood (ML) analysis of nucleotide sequences of five mitochondrial protein-coding genes. The elephant louse, *Haematomyzus elephantis*, was used as the outgroup. The ultrafast bootstrap support (%) / SH-aLRT support (%) were indicated near each node.



Supplementary Fig. 4.9. Alignment of partial *18S* rRNA gene, *28S* rRNA gene and *ef1-α* gene sequences between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*.

Supplementary Table 4.1. Species of parasitic lice included in the phylogenetic analyses in this study

Species	Host	GenBank accession number	References
Antarctophthirus microchir	Sea lion	MW803082-93	Present study
Antarctophthirus carlinii	Weddell seal	MW803073-81	Present study
Antarctophthirus lobodontis	Crabeater seal	MW803064-72	Present study
Lepidophthirus macrorhini	Elephant seals	MW803094-104	Present study
Proechinophthirus fluctus	Fur seal	MW803105-114	Present study
Pthirus gorilla	Gorilla	MW803115-131	Present study
Pedicinus obtusus	Macaque	MT792495-506	Fu & Dong et al., 2020
Pedicinus badii	Colobus	MT721726-39	Fu & Dong et al., 2020
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;
Polyplax asiatica	Greater bandicoot rat	KF647751-61	Dong et al., 2014b
Polyplax reclinata	Asian grey shrew	MW291451-61	Dong et al., 2021
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	KJ648923-32	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

Supplementary Table 4.2. Mitochondrial minichromosomes of *Lepidophthirus macrorhini* - louse of southern elephant seal (*Mirounga leonine*)

Minichro-mosome	Coding region size (bp)	Individual gene size (bp)	Assembled upstream non-coding region size (bp)	Assembled downstream non-coding region size (bp)	Number of Illumina sequence-reads	Mean coverage
atp8-atp6-N	937	189-675-70	300	300	27366	1900
E-cob	1152	63-1089	300	300	23907	1720
I-cox1-L1	1668	66-1536-65	300	300	37607	2232
M-cox2	820	67-699	300	300	21714	2023
P-cox3-A	921	67-786-66	300	300	29089	2553
Q-nad1-T-W1	1152	67-64-921-68	300	300	60472	3586
nad2	1114	1047	300	300	71155	5801
K-nad4	1395	69-1326	300	293	53047	3752
H-nad5-W2	1891	66-1539-70	300	170	22603	1434
L2-rrnL-V	1276	71-1139-66	300	273	75116	5679
rrnS-D	848	781-67	300	300	45830	4741

Supplementary Table 4.3. Sequence identities between *trnW*$_1$ and *trnW*$_2$ of the southern elephant seal louse, *Lepidophthirus macrorhini*, between *trnW* genes of *Lepidophthirus macrorhini* and other seal lice, and between *trnW* genes and other tRNA genes of *Lepidophthirus macrorhini*. Identities were generated with ClustalW in Geneious: cost matrix IUB, gap open cost 15, gap extend cost 6.66.

	***trnW*$_1$**	***trnW*$_2$**
***trnW*$_2$**	50%	
trnW (*Antarctophthirus carlinii*)	68.7%	55.1%
trnW (*Antarctophthirus lobodontis*)	68.7%	55.1%
trnW (*Antarctophthirus microchir*)	69.1%	42.6%
trnW (*Proechinophthirus fluctus*)	55.4%	48.5%
trnA	44.8%	44.8%
trnD	50.7%	50%
trnE	48.4%	50.9%
trnH	50%	48.5%
trnI	47.8%	47.8%
trnK	50.7%	53.6%
trnL$_1$	46.9%	52.2%
trnL$_2$	49.2%	43.7%
trnM	51.5%	42.6%
trnN	50.7%	53.5%
trnP	48.5%	38.8%
trnQ	51.5%	49.3%
trnT	50%	50%
trnV	52.2%	47.1%

Supplementary Table 4.4. Mitochondrial minichromosomes of *Proechinophthirus fluctus* - louse of northern fur seal (*Callorhinus ursinus*)

Minichro-mosome	Coding region size (bp)	Individual gene size (bp)	Assembled upstream non-coding region size (bp)	Assembled downstream non-coding region size (bp)	Number of Illumina sequence-reads	Mean coverage
E-cob	1156	64-1092	200	200	23212	1495
I-cox1	1598	73-1503	200	200	26228	1314
cox2	765	765	200	264	24242	1997
P-cox3	849	67-783	200	200	22441	813
Q-nad1-T	1042	66-910-67	200	149	13414	915
nad2	1029	1029	200	200	34297	1370
G-nad3-W	481	67-345-65	200	200	4011	763
K-nad4	1379	70-1308	200	200	23312	1321
H-nad5	1739	60-1680	200	200	12873	631
rrnS-S2-R-nad4L-C	1249	772-68-63-65	200	253	30866	1840

Supplementary Table 4.5. Mitochondrial minichromosomes of *Antarctophthirus carlinii* - louse of Weddell seal (*Leptonychotes weddellii*)

Minichro-mosome	Coding region size (bp)	Individual gene size (bp)	Assembled upstream non-coding region size (bp)	Assembled downstream non-coding region size (bp)	Number of Illumina sequence-reads	Mean coverage
atp8-atp6	829	151-672	300	300	21419	1840
E-cob	1157	65-1092	300	297	11440	1063
I-cox1-A	1670	66-1539-64	300	300	18669	1283
Y-cox2-N	805	65-674-65	300	300	8032	919
P-cox3	848	62-786	300	241	13456	1488
Q-nad1-T	1040	68-903-68	300	300	12074	1173
nad2	1056	1056	300	300	11297	1083
G-nad3-W	529	69-345-72	300	300	17009	1864
K-nad4	1373	67-1305	300	300	13739	1095
H-nad5	1756	70-1681	300	300	16301	1077
S_2-R-nad4L	420	70-69-279	300	300	8957	1098
L_2-rrnL	1421	64-1357	300	300	28669	1779
rrnS-L_1	908	839-69	300	300	19927	1858

Supplementary Table 4.6. Mitochondrial minichromosomes of *Antarctophthirus lobodontis* - louse of crabeater seal (*Lobodon carcinophagus*)

Minichro-mosome	Coding region size (bp)	Individual gene size (bp)	Assembled upstream non-coding region size (bp)	Assembled downstream non-coding region size (bp)	Number of Illumina sequence-reads	Mean coverage
atp8-atp6	829	151-672	300	300	27025	2567
E-cob	1157	65-1092	300	300	39824	3188
I-cox1-A	1670	66-1539-64	300	300	56989	3572
Y-cox2-N	807	65-674-69	300	300	32791	3233
P-cox3	848	62-786	300	241	30577	2977
Q-nad1-T	1040	68-903-68	300	277	29362	2527
nad2	1056	1056	300	300	31499	2645
G-nad3-W	529	69-345-72	300	300	22343	2641
K-nad4	1373	67-1305	300	300	44979	3228
H-nad5	1756	70-1686	300	300	44907	2719
S_2-R-nad4	420	70-69-279	300	300	17670	2322
L_2-rrnL	1429	64-1365	300	266	51799	3296
rrnS-L_1	866	797-69	300	300	24709	2293

Supplementary Table 4.7. Mitochondrial minichromosomes of *Antarctophthirus microchir* – louse of Australian sea lion (*Neophoca cinerea*)

Minichro-mosome	Coding region size (bp)	Individual gene size(bp)	Assembled upstream non-coding region size (bp)	Assembled downstream non-coding region size (bp)	Number of Illumina sequence-reads	Mean coverage
atp8-atp6-D	1083	180-645-67	250	250	79505	7055
cob	1215	1215	250	250	83187	7479
I-cox1	1607	65-1542	250	250	101204	6850
Y-cox2	792	69-723	250	250	47664	5050
P-cox3	908	68-840	250	250	56986	5541
nad1-T	991	82-909	250	250	76311	7167
nad2	996	996	250	250	63260	5917
nad3-W	384	312-68	250	250	30294	4601
K-nad4	1348	69-1278	250	250	71246	5459
nad5	1704	1704	250	250	85929	5561
rrnL-R	1443	1369-74	250	250	110797	9148
rrnS	731	731	250	250	74805	8256

Supplementary Table 4.8. Mitochondrial minichromosomes of *Pthirus gorilla* – louse of western gorilla (*Gorilla gorilla*)

Minichro-mosome	Coding region size (bp)	Individual gene size (bp)	Assembled upstream non-coding region size (bp)	Assembled downstream non-coding region size (bp)	Number of Illumina sequence-reads	Mean coverage
atp8-atp6	833	186-642	320	156	151286	17988
cob-S1	1137	1071-64	320	316	165473	15005
cox1	1551	1551	320	159	169187	13484
Y-cox2	756	69-687	291	320	141626	16356
cox3	807	807	320	309	154255	17083
nad1-Q	970	897-59	320	320	163499	16013
P-nad2-I	1126	66-984-66	320	311	157520	14332
G-nad3-V-W-S2	784	65-363-68-65-65	320	74	144350	18863
K-nad4	1338	68-1269	320	320	166965	13546
F-T-D-H-R-nad4L	859	65-66-70-68-65-276	320	320	137088	14455
nad5	1659	1659	320	320	183897	13054
F-nad6-E-M	749	65-489-64-67	320	199	135817	15392
L1/L2-rrnL	1288	68/68-1220	320	295	193756	16203
rrnS	819	67-752	257	320	135702	15075
A	65	65	320	44	8539	2478
C	64	64	320	34	427	129

Supplementary Table 4.9. Sequence identities between *Antarctophthirus carlinii* (louse of Weddell seal, *Leptonychotes weddellii*) and *Antarctophthirus lobodontis* (louse of crabeater seal, *Lobodon carcinophagus*)

Gene	Identity	Gene	Identity
atp8	100%	*trnG*	100%
atp6	98.1%	*trnH*	100%
cob	98.3%	*trnI*	100%
cox1	99.1%	*trnK*	100%
cox2	98.8%	*trnL1*	100%
cox3	98.5%	*trnL2*	100%
nad1	97.5%	*trnN*	98.5%
nad2	98.3%	*trnO*	100%
nad3	99.1%	*trnP*	100%
nad4	98.7%	*trnR*	98.6%
nad4L	98.9%	*trnS2*	100%
nad5	98.1%	*trnT*	97.1%
rrnL	98.9%	*trnW*	100%
rrnS	94.3%	*trnY*	100%
trnA	100%	Average identity	99.03%
trnE	100%		

Supplementary Table 4.10. Sequence identities between Weddell seal (*Leptonychotes weddellii*) and crabeater seal (*Lobodon carcinophagus*)

Gene	Identity	Gene	Identity
atp8	94.1%	*atp6*	91.2%
cob	92.5%	*cox1*	93%
cox2	93.7%	*cox3*	92.2%
nad1	94%	*nad2*	92.4%
nad3	90.8%	*nad4*	91.5%
nad4L	93.9%	*nad5*	91.6%
nad6	93.4%	*rrnS*	96.1%
rrnL	96.4%	*trnA*	98.6%
trnC	95.5%	*trnD*	94%
trnE	97.1%	*trnF*	97.2%
trnG	98.5%	*trnH*	100%
trnI	95.7%	*trnK*	92.8%
trnL$_1$	97.2%	*trnL*$_2$	97.3%
trnM	100%	*trnN*	93.2%
trnP	95.5%	*trnQ*	91.8%
trnR	95.7%	*trnS*$_1$	98.5%
trnS$_2$	96.6%	*trnT*	92.9%
trnV	94.1%	*trnW*	97%
trnY	98.5%	Average identity	94.99%

Chapter 5

Mitochondrial genome fragmentation occurred multiple times independently in bird lice of the families Menoponidae and Laemobothriidae

Submitted as:

Dong, Y., Martina, J., Amber, G., Ludovica, V., Gabriel, C., Dominique, P., & Shao, R. (2022). Mitochondrial genome fragmentation occurred multiple times independently in bird lice of the families Menoponidae and Laemobothriidae.

Revised manuscript currently under external review with *Cladistics*

Manuscript: Appendix IV

Images of lice and their hosts: Appendix V

5.1. Abstract

Mitochondrial (mt) genome fragmentation has been discovered in all five parvorders (Amblycera, Anoplura, Ischnocera, Rhynchophthirina and Trichodectera) of parasitic lice (infraorder Phthiraptera). To explore whether minichromosomal characters derived from mt genome fragmentation are informative for resolving the phylogeny of parasitic lice at low taxonomic levels, we sequenced the mt genomes of 17 species of bird lice in two families: Menoponidae and Laemobothriidae (Amblycera). Of the 17 species, four species of Menoponidae (*Actornithophilus grandiceps*, *Actornithophilus hoplopteri*, *Austromenopon atrofulvum* and *Myrsidea ptilorhynchi*) have fragmented mt genomes whereas the other 13 species retain the typical single-chromosome mt genomes. The two species of *Actornithophilus* have five and six minichromosomes respectively and differ in the location of six genes in minichromosomes. *Austromenopon atrofulvum* has two mt minichromosomes, in contrast to *Austromenopon paululum*, which has all mt genes on a single chromosome. *Myrsidea ptilorhynchi* has four mt minichromosomes. We mapped mt genome fragmentation events on the phylogeny of Menoponidae and Laemobothriidae species inferred with mt genome sequences. It is clear that mt genome fragmentation occurred four times independently in each of the four genera: *Actornithophilus*, *Myrsidea*, *Austromenopon* and *Laemobothrion*. We found derived minichromosomal characters shared between species of *Myrsidea*, between *Actornithophilus* species, between ischnoceran genera, *Penenirmus* and *Saemundssonia*, and among ischnoceran genera, *Penenirmus*, *Quadraceps* and *Saemundssonia*, respectively. We conclude that while mt genome fragmentation as a general feature does not unite all parasitic lice that have this feature, each independent mt genome fragmentation event does produce shared derived minichromosomal characters that can be informative in resolving the phylogeny of parasitic lice at different taxonomic levels.

5.2. Introduction

Lice (infraorder Phthiraptera) are obligate parasites living on birds and mammals (Price et al., 2003; Durden & Musser 1994; De Moya et al., 2021). Among parasitic lice, chewing lice are in four parvorders: Amblycera, Ischnocera, Rhynchophthirina and Trichodectera, while sucking lice are in the parvorder Anoplura (De Moya et al., 2021). Fragmented mitochondrial (mt) genomes have been found in 21 species of Anoplura (Shao et al., 2009; Shao et al., 2012; Jiang et al., 2013; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Herd et al., 2015; Shao et al., 2017; Fu et al., 2020; Dong et al., 2021; Dong et al., 2022), one species of Rhynchophthirina (Shao et al., 2015), and five

species of Trichodectera (Song et al., 2019; Spradling et al., 2021). Each of these species has 9-20 minichromosomes, which are usually less than 4 kb in size.

Cameron et al. (2011) and three recent studies showed that mt genome fragmentation also occurred in species of Ischnocera (Sweet et al., 2020; Sweet et al., 2022) and Amblycera (Sweet et al., 2021). Cameron et al. (2011) reported five species of Ischnocera with fragmented mt genomes based on PCR test and partial mt genome sequences. Sweet et al. (2020) reported that the mt genomes of three *Columbicola* species of dove lice (Ischnocera) each had 15-17 minichromosomes that were less than 4 kb in size. Sweet et al. (2022) reported that another 14 species of Ischnocera in 14 different genera had fragmented mt genomes; each of these species had 3-10 minichromosomes ranging from around 1 kb to 11 kb in size. Sweet et al. (2021) reported that the mt genomes of four species of Amblycera from different genera (*Cummingsia, Laemobothrion, Macrogyropus* and *Myrsidea*) each had three or seven minichromosomes.

Mitochondrial genome fragmentation occurred at least nine times independently in the parvorder Ischnocera and at least three times independently in the parvorder Amblycera according to Sweet et al. (2022). These fragmentation events are independent from the fragmentation observed in eutherian mammal lice of the parvorders Anoplura, Rhynchophthirina and Trichodectera, which apparently occurred only once in the most common recent ancestor of these three parvorders (Song et al., 2019; Sweet et al., 2020; De Moya et al., 2021; Sweet et al., 2022). The typical single-chromosome mt genome has not been seen in any species from Anoplura, Rhynchophthirina or Trichodectera, but has been seen in 10 species from 10 genera of Ischnocera and seven species of Amblycera from three families (Boopiidae; Menoponidae and Ricinidae) (Shao et al., 2001; Song et al., 2019; Sweet et al., 2021; Gong et al., 2022).

In the current study, we sequenced the mt genomes of 17 species of amblyceran lice: 14 from the family Menoponidae and three from the family Laemobothriidae (Table 5.1). Our aims are to investigate: 1) how widespread mt genome fragmentation occurred in the Amblycera, using Menoponidae as an example; 2) whether other species in Laemobothriidae have fragmented mt genomes as reported for *Laemobothrion tinnunculi* (Sweet et al., 2021); and 3) whether minichromosomal characters derived from mt genome fragmentation are informative for resolving the phylogeny of parasitic lice at low taxonomic levels such as genus and species.

5.3. Materials and methods

5.3.1. Collection of bird lice

We collected 17 species of bird lice; 16 of these species were collected in the Australian Zoo Wildlife Hospital (AZWH) from euthanized birds and one species was from live birds (Australian white ibis) captured at the UniSC Fraser Coast campus and identified by one of the authors (DP) (Table 5.1). Euthanized birds were identified initially by AZWH staff and verified by two of the authors (YD and RS) based on morphology according to Pizzey, Knight and Pizzey (2012) and Menkhorst et al. (2019). We used the post-mortem-ruffling method to collect bird lice (Clayton & Drown, 2001). Ethyl acetate was used to treat euthanized birds if lice were still alive. Ethanol-acetate-soaked cotton balls were put in sealed bags containing euthanized birds, one bird per bag to avoid contamination. After 5-10 minutes, treated birds were placed on a clean A1 size white paper, their feathers ruffled to get lice off bird feathers to the white paper. Lice were kept in 2-mL collection tubes filled with 80% ethanol in -80°C freezer until DNA extraction. Lice were checked under microscope (Nikon SMZ 800N), imaged and identified by morphological features and host records according to the world checklist of chewing lice (Price et al., 2003).

Table 5.1. Species of chewing lice included in the phylogenetic analyses in this study

Louse species	Host	SRA Biosample accession number	Number of Illumina sequence reads obtained	GenBank accession number	References	Collection location in Queensland
Actornithophilus grandiceps	Pied oystercatcher (*Haematopus longirostris*)	SAMN29722645	19,486,216	ON417463-67	Present study	Manly
Actornithophilus hoplopteri	Masked lapwing (*Vanellus miles*)	SAMN29722639	20,263,604	ON380416-21	Present study	North Aramara
Amyrsidea minuta	Green peafowl (*Pavo muticus*)			MH0012227	Song et al., 2019	N.A.
Austromenopon atrofulvum	Sooty tern (*Onychoprion fuscatus*)	SAMN29722646	20,656,938	ON380413-14	Present study	Kings Beach
Austromenopon atrofulvum	Crested tern (*Thalasseus bergii*)	SAMN29722643	15,659,674	ON380422-23	Present study	Caloundra
Austromenopon paululum	Sooty shearwater (*Ardenna grisea*)	SAMN29722648	17,438,954	ON380415	Present study	Minyama
Ciconiphilus decimfasciatus	White-faced heron (*Egretta novaehollandiae*)	SAMN29722644	23,418,092	OM912467	Present study	Rothwell
Colpocephalum eucarenum	Australian pelican (*Pelecanus conspicillatus*)	SAMN29722638	12,309,898	ON204017	Present study	Amity Pt
Colpocephalum griffoneae	Himalayan vulture (*Gyps himalayensis*)			MH001228	Song et al., 2019	N.A.
Colpocephalum spinicollis	Straw-necked ibis (*Threskiornis spinicollis*)	SAMN29722649	15,812,316	ON204018	Present study	Beerburrum
Eomenopon denticulatum	Scaly-breasted lorikeet (*Trichoglossus chlorolepidotus*)	SAMN29722636	12,065,720	ON184000	Present study	Caloundra
Franciscoloa funerei	Yellow-tailed black cockatoo (*Calyptorhynchus funereus*)	SAMN29722642	15,891,768	ON016537	Present study	Curramone
Franciscoloa pallida	Sulfur-crested cockatoo (*Cacatua galerita*)	SAMN29722641	13,324,866	ON016538	Present study	Buderim
Franciscoloa roseicapillae	Little corella (*Cacatua sanguinea*)	SAMN29722635	9,794,058	ON303707	Present study	Moffat Beach
Franciscoloa roseicapillae	Galah (*Eolophus roseicapilla*)	SAMN29722640	14,514,292	ON303708	Present study	Glass House mountains
Franciscoloa roseicapillae	Pheasant coucal (*Centropus phasianinus*)	SAMN29722650	14,626,494	ON303709	Present study	Laceys Creek
Laemobothrion atrum	Eurasian coot (*Fulica atra*)	SAMN29722652	12,096,816	OM935763	Present study	Warana
Laemobothrion maximum	Black kite (*Milvus migrans*)	SAMN29722651	15,152,282	OM935762	Present study	Caloundra
Laemobothrion sp.	Australian swamphen (*Porphyrio porphyrio bellus*)	SAMN29722653	23,631,838	OM912466	Present study	Caloundra
Laemobothrion tinnunculi	Australian hobby (*Falco longipennis*)			MW199169	Sweet et al., 2021	N.A.
Menacanthus cornutus	Red junglefowl (*Gallus gallus*)			NC062859	Gong et al., 2022	N.A.
Myrsidea ptilorhynchi	Satin bowerbird (*Ptilonorhynchus violaceus*)	SAMN29722647	25,679,092	ON417459-60, ON417462, OP271738	Present study	Maleny
Myrsidea sp.	Citrine warbler (*Myiothlypis luteoviridis*)			MW199172-74	Sweet et al., 2021	N.A.
Osborniella crotophagae	Smooth-billed ani (*Crotophaga ani*)			MW199175	Sweet et al., 2021	N.A.
Piagetiella africana	Australian Pelican (*Pelecanus conspicillatus*)	SAMN29722637	19,903,264	ON417468	Present study	Amity Pt
Plegadiphilus threskiornis	Australian white ibis (*Threskiornis moluccus*)	SAMN29722634	9,680,062	ON183999	Present study	Fraser Coast

5.3.2. DNA extraction, mitochondrial genome sequencing and assembly

Genomic DNA of all the 17 bird louse species (20 samples in total; Table 5.1) were extracted with DNeasy Blood and Tissue kit (QIAGEN); 1 to 50 individual lice of each sample (depending on body size of louse) were used in DNA extraction. For the three species of *Laemobothrion*, we used single individual lice for extraction; for the species of Menoponidae, we used 30-50 individual lice collected from the same host individual for DNA extraction. Genomic DNA was checked by Nanodrop and Qubit to meet quality and quantity requirement for sequencing by Novogene (HK) with Illumina Novaseq 6000 platform (PE 150). According to Novogene (HK), the genomic DNA was randomly sheared into short fragments (350 bp); the obtained fragments were end-repaired, A-tailed and ligated with Illumina adapters. The fragments with adapters were PCR amplified, size selected and purified. The library was checked with Qubit and real-time PCR for quantification and with Agilent Bioanalyzer for size distribution. Quantified libraries were sequenced on Illumina platforms. 9.68 to 25.68 million cleaned Illumina sequence reads were obtained from each genomic DNA sample; these reads were deposited in the NCBI SRA database (Table 5.1). Illumina sequence reads were assembled using Geneious 11.0.2 (Kearse et al., 2012). We chose Geneious assembler and medium sensitivity option for de novo assemblies. We chose Geneious mapper and custom sensitivity for reference mapping with the following parameters: 1) 80 bp minimum overlap and 95% minimum identity; 2) maximum 5% gaps per read with 2 bp maximum gap size; and 3) maximum mismatches per read 5% and maximum ambiguity 2. According to Crampton-Platt et al. (2016), the proportion of mt genome reads ranged from 0.5% to 1.4% in whole genome shotgun sequence reads; we thus used 1% of total sequence reads (around 100,000 to 200,000 reads) obtained from each genomic DAN sample as the initial seed dataset for de novo assembly. On average, each initial seed dataset had around 1,000 to 2,000 mt genome reads (150 bp each read), which were sufficient for us to obtain contigs covering multiple mt genes. The consensus sequences of the contigs generated by de novo assembly were searched with BLAST in NCBI databases using the default parameters to identify mt gene sequence matches (Johnson et al., 2008). The consensus sequences that matched significantly (E-value < 10^-18) to mt gene sequences of amblyceran chewing lice were used as initial reference sequences for mapping with the total Illumina sequence reads of each louse sample. For species with the typical single-chromosome mt genomes, extending repeatedly the contigs obtained from the initial references led to the assembly of the full-length mt genomes. For species with fragmented mt genomes, this same approach led to the assembly of multiple full-length mt minichromosomes. We annotated these minichromosomes to identify the mt genes in each minichromosome. We then repeated this process to target unidentified mt genes until

no more mt minichromosomes or genes could be identified. Protein-coding genes and rRNA genes were identified and annotated based on BLAST search results of GenBank (Johnson et al., 2008); tRNA genes were identified and annotated based on secondary structures produced by ARWEN (Laslett et al., 2008) and tRNAscan-SE (Lowe et al., 1997).

5.3.3. PCR verification of mitochondrial minichromosomes

For the four species of Menoponidae with fragmented mt genomes (*Actornithophilus grandiceps, Actornithophilus hoplopteri, Austromenopon atrofulvum* and *Myrsidea ptilorhynchi*), we designed 17 pairs of specific outward primers to verify the circular organization and the size of the 17 minichromosomes obtained from Illumina sequence-read assembly. These primers were designed from the protein-coding or rRNA gene sequences of each minichromosome (Supplementary Table 5.1). The forward and reverse primers in each pair were next to each other with a small gap (200-400 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 Novaseq 6000 platform as described above. PCR setting was: 12.5 ul PrimeSTAR Max Premix (2X), 1 ul forward primer, 1 ul reverse primer, 1 ul DNA template, 9.5 ul ultrapure water (PCR grade). PCR cycling condition was: 94°C for 1m; 35 cycles of 98°C for 10s, 60°C for 15S, 72°C for 1m; 72°C for 10m. Three negative controls were set up for each pair of PCR primers: 1) 1 ul ultrapure water (PCR grade) replacing the DNA template; 2) 1ul ultrapure water replacing the forward primer; and 3) 1 ul ultrapure water replacing the reverse primer. PCR amplicons were checked by agarose gel (1%) electrophoresis and were sequenced with Novaseq 6000 platform (PE 150) at Novogene (HK) with the same library construction methods described above; the Illumina sequence reads were assembled with the same approach described above for genomic DNA sequence assembly (Kearse et al., 2012).

5.3.4. Phylogenetic analysis

We constructed phylogenetic trees of 22 species of bird lice: 18 of these lice are in the family Menoponidae and the other four species are in the family Laemobothriidae (Table 5.1). Sequences of 13 mt protein-coding genes (*atp6*, *atp8*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, nad3, *nad4*, *nad4L*, *nad5* and *nad6*) were used in our phylogenetic analysis. The amino acid sequences of these genes were aligned individually using the MAFFT algorithm implemented in TranslatorX online planform (Abascal et al., 2010). Poorly aligned sites were removed by Gblocks vo.91b (Talavera, G., &

Castresana, J, 2007) with default parameters and gaps not allowed. Then, the alignments of individual genes were concatenated together for phylogenetic analyses. For maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees, we did partitioned analyses with PartitionFinder2 (Lanfear et al., 2017). A MP tree was constructed with PAUP 4.0b10 (Swofford, 2002). We used Heuristic method for MP search with the parameter setting: random additional sequences, 1,000 replicates and TBR branch-swapping algorithm. A ML tree was constructed with IQ-TREE (Nguyen et al., 2014). Ultrafast bootstrap replicates were 1,000. A BI tree was constructed with MrBayes 3.2 6 (Ronquist et al., 2003). MCMC generations were set as 5,000,000 and sample-frequency was set as every 1,000 generations. We set a burn-in of 500,000 generations and the MCMC convergence was checked with Tracer v1.7.2 (Rambaut et al., 2018). The MP tree, the ML tree and the BI tree were edited in Figtree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

5.4. Results

5.4.1. Bird lice have both fragmented and typical single-chromosome mitochondrial genome organization

Among the 17 species of bird lice we sequenced for the first time in this study, 13 of them (10 species of Menoponidae and three species of Laemobothriidae) have the typical single-chromosome mt genomes (Fig. 5.2A, Supplementary Fig. 5.1). The other four species of Menoponidae (*Actornithophilus grandiceps*, *Actornithophilus hoplopteri*, *Austromenopon atrofulvum* and *Myrsidea ptilorhynchi)* have fragmented mt genomes: 1) *Actornithophilus grandiceps* have five minichromosomes and *Actornithophilus hoplopteri* have six minichromosomes (Fig. 5.1); 2) *Austromenopon atrofulvum* has two minichromosomes (Fig. 5.2B); and 3) *Myrsidea ptilorhynchi* has four minichromosomes (Fig. 5.3). All the minichromosomes of these four species except one minichromosome of *Austromenopon paululum* (M1 minichromosome, 9,499 bp in size, more details below) were verified by PCR (Supplementary Fig. 5.2, 5.3, 5.4, 5.5). The failure to amplify M1 minichromosome by PCR was most likely due to its large size. The PCR amplicons from each minichromosome were also sequenced individually and assembled to verify the minichromosomes obtained from genomic DNA sequence assembly. We didn't detect any variation in gene order among individual lice of the same species in the pooled samples sequenced in this study. The mitochondrial genomes (i.e. their sequence, gene content and gene order) of each louse species were identified and confirmed by assembly of Illumina sequence reads obtained from genomic

DNA samples. For the minichromosomes identified, the mean coverage ranges from 422 to 4,396 (Supplementary Table 2). Our sequence comparison revealed a new host record of *Franciscoloa roseicapillae* on pheasant coucal. All the mt protein-coding genes and rRNA genes of *Franciscoloa* lice collected from pheasant coucal (GenBank accession numbers: ON303709) have 99.0-99.9%% sequence identity with their homologous genes of the *Franciscoloa roseicapillae* lice we collected from little corella and galah (*cox1* gene has 99.6% and 99.7% identity) (GenBank accession numbers: ON303707 and ON303708). The mt genomes of the 17 species of bird lice sequenced in the current study were annotated and deposited in GenBank (accession numbers listed in Table 5.1).

5.4.2. Mitochondrial karyotype differs between two *Actornithophilus* species

We assembled the mt genomes of *Actornithophilus grandiceps* (louse of pied oystercatcher) and *Actornithophilus hoplopteri* (louse of masked lapwing), and identified 35 of the 37 typical mt genes in both species (*trnP* and *trnR* not found). These 35 genes are on five minichromosomes in *Actornithophilus grandiceps* but are on six minichromosomes in *Actornithophilus hoplopteri* (Fig. 5.1). The five minichromosomes (M1-M5) of *Actornithophilus grandiceps* range from 2,005 bp to 6,112 bp (Fig. 5.1, Supplementary Table 5.2). The smallest minichromosome, M1, has three genes while the largest minichromosome, M5, has 15 genes. The six minichromosomes (M1-M6) of *Actornithophilus hoplopteri* range from 1,698 bp to 4,361 bp (Fig. 5.1, Supplementary Table 5.2). The smallest minichromosome, M1, has three genes while the largest minichromosome, M6, has 12 genes.

Actornithophilus grandiceps and *Actornithophilus hoplopteri* share obvious similarity in mt karyotype between each other but not with any other parasitic lice (Fig. 5.1): 1) both species have *E-nad4L-nad4* minichromosome (i.e. M1); and 2) both species have five derived gene clusters in different minichromosomes (*V-K-cob-nad1*, *nad6-H-S2*, *nad2-S1-F*, *Y-atp8-atp6-N* and *G-I-cox2-cox1-C-cox3*) (Supplementary Table 5.3). These two species of *Actornithophilus*, however, are distinct from each other in mt karyotype: 1) *Actornithophilus grandiceps* has five minichromosomes while *Actornithophilus hoplopteri* has six minichromosomes; and 2) all their relatable minichromosomes differ in gene content, with the only exception of M1 minichromosome (Fig. 5.1). The difference between *Actornithophilus grandiceps* and *Actornithophilus hoplopteri* in mt karyotype can be accounted for by multiple events that occurred after the divergence of these two species from their most recent common ancestor: 1) a split of one minichromosome (like M2 of

Actornithophilus grandiceps) into two minichromosomes, M2 and M3, in *Actornithophilus hoplopteri*, or equally likely the reverse of this event, i.e. merger of two minichromosomes (like M2 and M3 of *Actornithophilus hoplopteri*) as one in *Actornithophilus grandiceps* to form M2 minichromosome; 2) translocation of six genes (*trnQ*, *trnW*, *trnL1*, *nad5*, *trnD* and *trnL2*) between different minichromosomes in either of the two species; and 3) inversion (i.e. change in transcription orientation) of *trnA*, *trnQ* and *trnW* in either of the two species (Fig. 5.1).

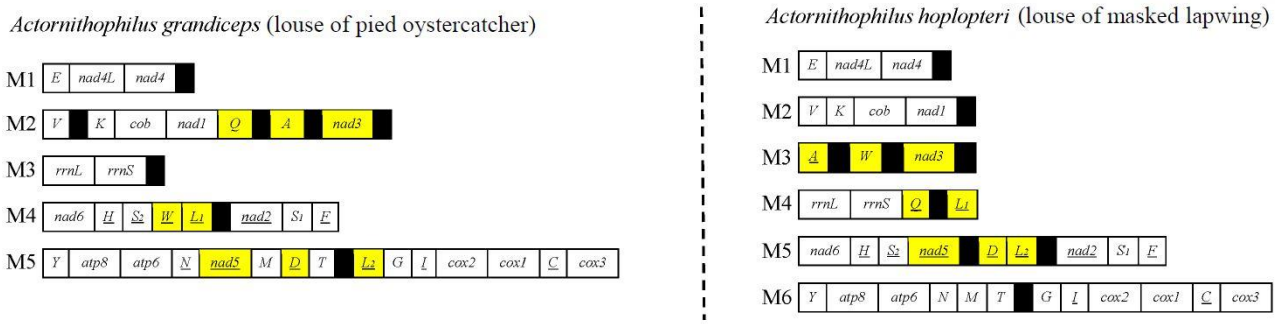


Fig. 5.1. The mitochondrial genomes of *Actornithophilus grandiceps* (louse of pied oystercatcher) and *Actornithophilus hoplopteri* (louse of masked lapwing). tRNA genes are indicated with their single-letter abbreviations of the corresponding amino acids. Genes are transcribed from left to right except underlined, which have an opposite orientation of transcription. Translocated genes between these two species are in yellow. Non-coding regions are in black. Circular minichromosomes are linearised for the sake of illustration.

5.4.3. The typical single-chromosome mitochondrial genome was retained in *Austromenopon paululum* but broke up into two chromosomes in *Austromenopon atrofulvum*

We assembled the mt genomes of *Austromenopon paululum* (louse of sooty shearwater) and *Austromenopon atrofulvum* (two louse samples, one from crested tern and other from sooty tern). All the 37 typical mt genes were identified in both species of *Austromenopon* (Fig. 5.2). The 37 mt genes of *Austromenopon paululum* are on a single chromosome, 14,992 bp in size, which is typical of bilateral animals (Boore, 1999, Lavrov and Bell, 2014). There are two non-coding regions in the mt genome of *Austromenopon paululum*: one is 127 bp in size between *cox1* and *atp6* and other is 327 bp between *trnM* and *trnD* (Fig. 5.2). *Austromenopon atrofulvum* share six derived gene clusters (20 genes in total) with *Austromenopon paululum*: 1) *cox2-cox1-atp6-atp8*, 2) *nad2-nad5-I*, 3) *nad3-Y*, 4) *rrnL-rrnS-L1*, 5) *nad4L-nad4-E*, and 6) *P-K-S1* (Fig. 5.2, Supplementary Table 5.3).

However, the 37 mt genes of *Austromenopon paululum* are on two minichromosomes: M1 is 9,499 bp in size with 16 genes, and M2 is 4,943 bp with 20 genes (Fig. 5.2, Supplementary Table 5.2). There are two non-coding regions in each minichromosome: for M1, one is 50 bp in size between *cox1* and *atp6* and the other is 49 bp between *nad2* and *atp8*; for M2, one is 141 bp in size between *trnS1* and *trnQ* and the other is 80 bp between *trnQ* and *nad4L* (Fig. 5.2).

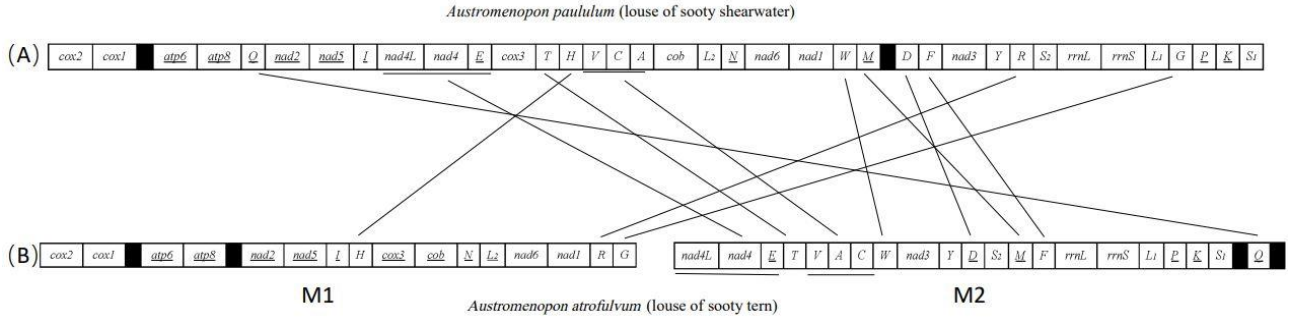


Fig. 5.2. The mitochondrial genomes of (A) *Austromenopon paululum* (louse of sooty shearwater) and (B) *Austromenopon atrofulvum* (louse of sooty tern and crested tern). Translocated genes between *Austromenopon paululum* and *Austromenopon atrofulvum* are linked by lines. Circular minichromosomes are linearised for the sake of illustration.

5.4.4. The mitochondrial genome of *Myrsidea ptilorhynchi* comprises at least four minichromosomes

We assembled the mt genome of *Myrsidea ptilorhynchi* (louse of satin bowerbird) and identified 23 of the 37 typical mt genes (13 mt protein-coding genes, two rRNA genes and eight tRNA genes); the other 14 tRNA genes were not identified (Fig. 5.3). These 23 mt genes are on four minichromosomes (M1-M4) ranging from 1,969 bp to 6,185 bp in size. M1 is the smallest minichromosome with only one gene (*rrnL*) and a 790-bp non-coding region while M4 is the largest minichromosome with 13 genes and a 59-bp non-coding region between *nad2* and *trnM*. M3 is 5,167 bp in size with eight genes and a 27-bp non-coding region between *nad3* and *trnS1*. M2 minichromosome is 3,194 bp in size with only *rrnS* gene and a 2,494 bp non-coding region (Fig. 5.3, Supplementary Table 5.2).

Sweet et al. (2021) reported that the mt genome of an undescribed species of *Myrsidea* (*Myrsidea* sp.) comprised three minichromosomes. All the mt protein-coding genes and rRNA genes were identified in both species of *Myrsidea*; however, we identified only eight mt tRNA genes in *Myrsidea ptilorhynchi* and Sweet et al. (2021) identified only seven mt tRNA genes in *Myrsidea* sp. A major difference between *Myrsidea ptilorhynchi* and *Myrsidea* sp. is on the placement of two rRNA genes. *rrnL* and *rrnS* are on two different minichromosomes in *Myrsidea ptilorhynchi* (Fig. 5.3), while they are on the same minichromosome in *Myrsidea* sp. (Sweet et al. (2021). These two species of *Myrsidea* have one minichromosome (*Y-nad5-L1-nad3-S1-nad2-cox1-cox3*) in common. The placement of other eight mt protein-coding genes are the same between these two species of *Myrsidea*; however, the placement of tRNA genes associated with these protein-coding genes are different between the two species.

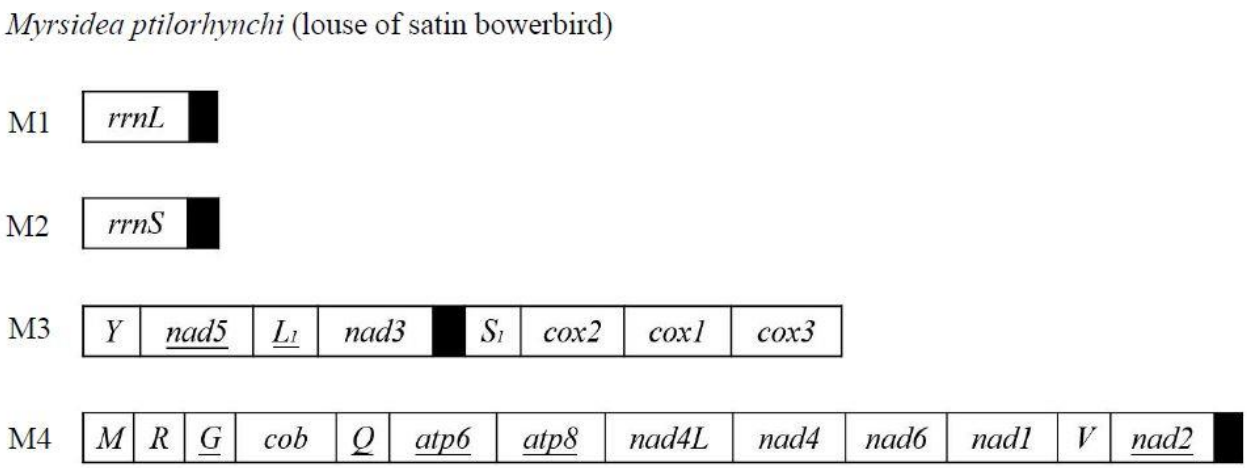


Fig. 5.3. The mitochondrial genome of *Myrsidea ptilorhynchi* (louse of satin bowerbird). Circular minichromosomes are linearised for the sake of illustration.

5.4.5. Three *Laemobothrion* species retained the typical single-chromosome mitochondrial genome organization

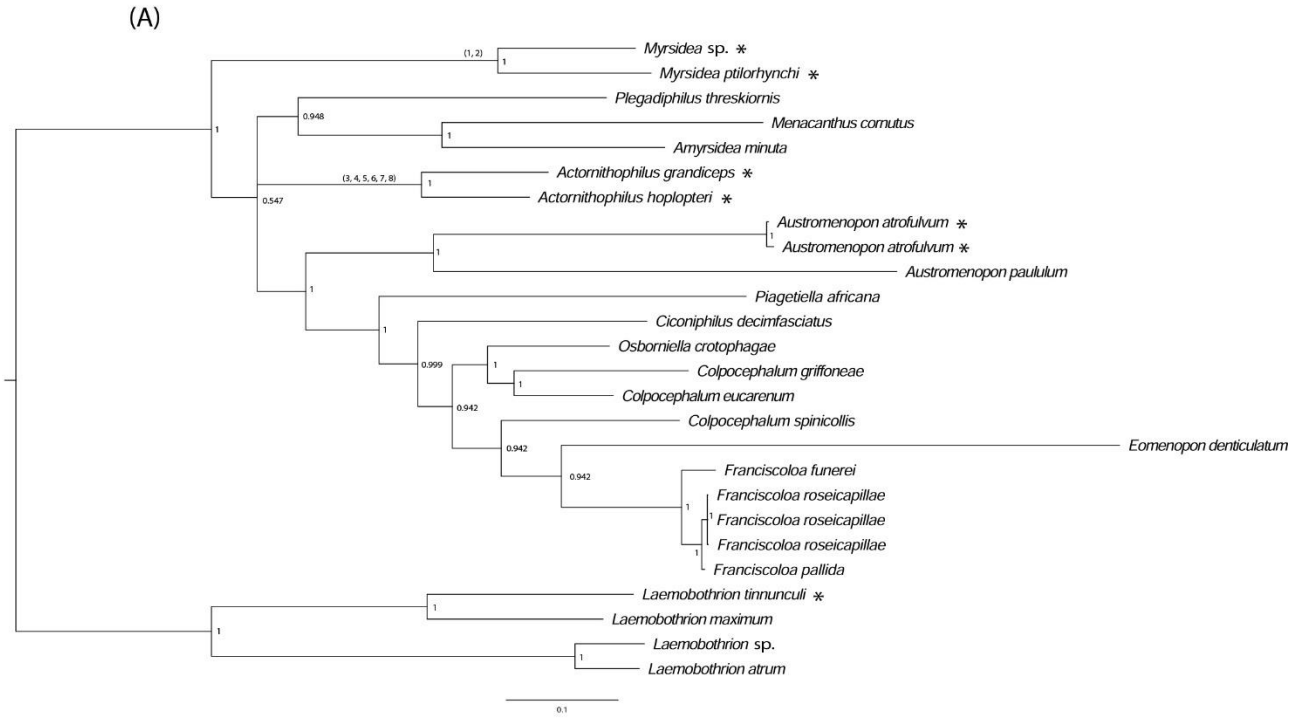
We assembled the mt genomes of three species of *Laemobothrion*: *Laemobothrion atrum* (louse of Eurasian coot), *Laemobothrion maximum* (louse of black kite) and a currently undescribed *Laemobothrion* sp. (louse of Australasian swamphen). All these species of *Laemobothrion* have the typical single-chromosome mt genomes: 14,155 bp (*Laemobothrion atrum*), 14,626 bp (*Laemobothrion maximum*) and 14,308 bp (*Laemobothrion* sp.) in size, respectively (Supplementary Fig. 5.1). All the 37 mt genes typical of bilateral animals were identified in these

three species. *Laemobothrion atrum* and *Laemobothrion* sp. have the same gene arrangement; all mt genes of these two species have the same orientation of transcription except *trnE*, which has an opposite orientation to other genes. The mt gene arrangement of *Laemobothrion maximum* differs from that of the other two *Laemobothrion* species in the location of seven genes: *nad3*, *trnG*, *trnW*, *trnV*, *trnF*, *trnM* and *trnQ* (Supplementary Fig. 5.1). *Laemobothrion atrum* and *Laemobothrion* sp. both have a 111-bp non-coding region between *trnD* and *rrnL*; *Laemobothrion maximum* has a 327-bp non-coding region between *cob* and *nad1* (Supplementary Fig. 5.1). In contrast to the three species of *Laemobothrion* we sequenced in the current study, *Laemobothrion tinnunculi* reported in Sweet et al. (2021) had three minichromosomes: *cob* gene alone was on one minichromosome; *nad2*, *trnP*, *trnG* and *trnW* genes were on another minichromosome; and other mt genes were on the third minichromosome.

5.4.6. Phylogeny of Menoponidae lice inferred with mitochondrial genome sequences

We reconstructed the phylogeny of Menoponidae lice (18 species from 11 genera included) with mt genome sequences to aid the understanding of mt genome fragmentation in this family; four species of Laemobothriidae were used as the outgroup. In both BI tree and ML tree, the 18 species of Menoponidae are divided into five well supported clades: 1) two *Myrsidea* species (both have fragmented mt genomes) form a clade; 2) *Plegadiphilus threskiornis*, *Menacanthus cornutus* and *Amyrsidea minuta* (all have typical single-chromosome mt genomes) form a clade; 3) two species of *Actornithophilus* (both have fragmented mt genomes) form a clade; 4) two species of *Austromenopon* (one has a two-chromosome mt genome whereas the other has a single-chromosome mt genome) form a clade; and 5) other 10 species of Menoponidae from six genera (*Ciconiphilus*; *Colpocephalum*, *Eomenopon*, *Franciscoloa*, *Osborniella* and *Piagetiella*; all have single-chromosome mt genomes) form a clade (Fig. 5.4). Four of these five clades (clade 1-4) were also well supported in the MP tree although the relationships among these clades could not be resolved (Fig. 5.5). The fifth clade was not supported in the MP tree: *Eomenopon denticulatum* alone was on a branch separated from all other species of Menoponidae; the other nine species in the fifth clade were grouped together with weak support (Fig. 5.5). In the outgroup, *Laemobothrion atrum* and the undescribed *Laemobothrion* sp. are most closely related (bootstrap value 100%), both having single-chromosome mt genomes (Fig. 5.4, Fig. 5.5). *Laemobothrion maximum*, which has a single-chromosome mt genome (Supplementary Fig. 1), is most closely related to *Laemobothrion*

tinnunculi (bootstrap value 100%; Fig. 5.4, Fig. 5.5), which has three mt minichromosomes according to Sweet et al. (2021).



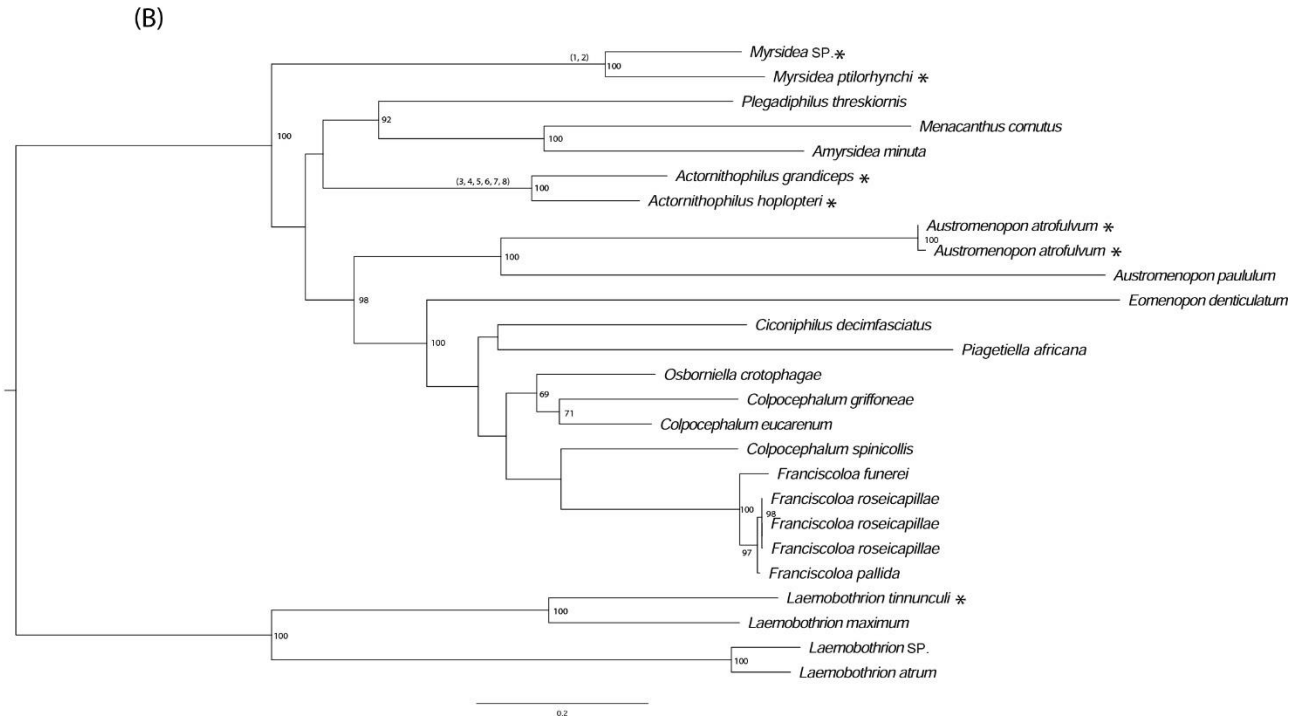


Fig. 5.4. Phylogenetic relationships among 23 species of chewing lice (parvorder Amblycera) inferred from Bayesian inference (A) and maximum likelihood (B) analysis of amino acid sequences of all mitochondrial protein-coding genes. Species of Laemobothriidae were used as the outgroup. Bayesian posterior probabilities were indicated near each node of BI tree. The number of samples of the BI tree is 9,002 and the effective sample size (ESS) is 8,735. The mcmc convergence were checked in Tracer 1.7.2. Ultrafast bootstrap support values over 60 were indicated near each node of ML tree. Species with fragmented mt genomes are marked with * symbol. Numbers in brackets indicate the following shared derived minichromosomal characters: 1) minichromosome *Y-nad5-L1-nad3-S1-cox2-cox1-cox3*; 2) gene cluster *G-cob-Q-atp6-atp8-nad4L-nad4-nad6-nad1*; 3) minichromosome *E-nad4L-nad4*; 4) gene cluster *V-K-cob-nad1*; 5) gene cluster *nad6-H-S2*; 6) gene cluster *nad2-S1-F*; 7) gene cluster *Y-atp8-atp6-N*; and 8) gene cluster *G-I-cox2-cox1-C-cox3*.

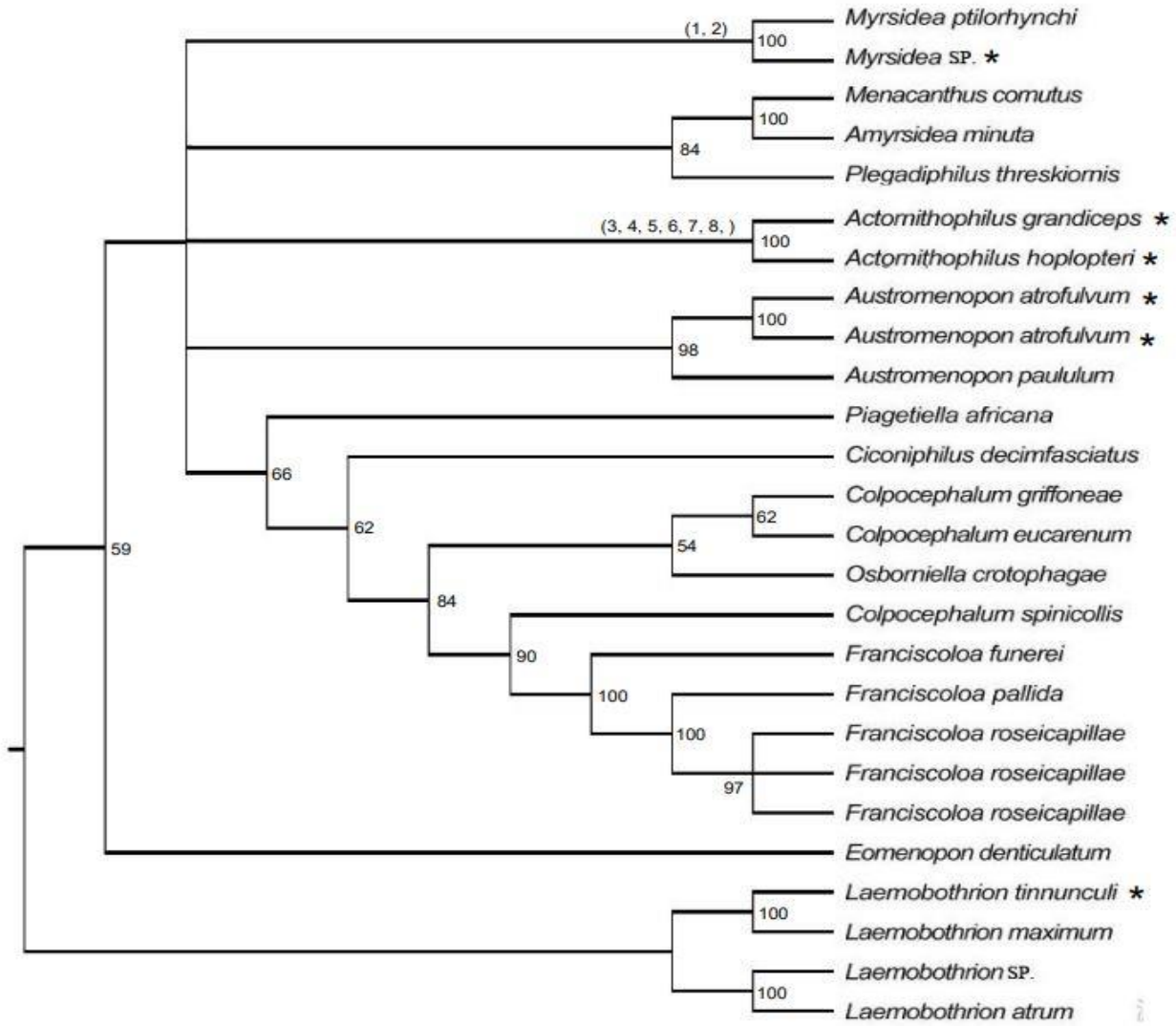


Fig. 5.5. Phylogenetic relationships among 23 species of chewing lice (parvorder Amblycera) inferred from maximum parsimony analysis of amino acid sequences of all mitochondrial protein-coding genes. Species of Laemobothriidae were used as the outgroup. Bootstrap support values were indicated near each node. Species with fragmented mt genomes are marked with * symbol. Numbers in brackets indicated the shared derived minichromosomal characters detailed in Fig. 5.4 legend.

5.5. Discussion

5.5.1. Mitochondrial genome fragmentation occurred three times independently in the Menoponidae species and once in the Laemobothriidae species investigated to date

Our results showed that while most species in the families Menoponidae and Laemobothriidae investigated to date retained the typical single-chromosome mt genomes ancestral to animals, mt genome fragmentation occurred once in the Laemobothriidae and three times independently in the Menoponidae: once in the genus *Myrsidea*, once in the genus *Actornithophilus*, and once in *Austromenopon atrofulvum* (Fig. 5.4, Fig. 5.5). For the genus *Austromenopon*, it is clear that not all species in this genus have fragmented mt genomes as *Austromenopon paululum* has a single-chromosome mt genome. This is also true for the family Laemobothriidae, which has one genus *Laemobothrion* with 20 species (Price et al., 2003): only *Laemobothrion tinnunculi* has a fragmented mt genome whereas other three *Laemobothrion* species have single-chromosome mt genomes. Sweet et al (2021) reported that the mt genome of *Laemobothrion tinnunculi* had three mt minichromosomes: *cob* minichromosome, *nad2-P-G-W* minichromosome and a minichromosome with all other 32 genes. This is in stark contrast to the three *Laemobothrion* species we sequenced in the current study: all 37 mt genes are on the same single chromosome in each species (Supplementary Fig. 5.1). For the genera *Myrsidea* and *Actornithophilus*, it is not clear yet whether all species in these genera have fragmented mt genomes, or not. *Myrsidea* is the most species-rich genus in the Menoponidae with 208 species; *Actornithophilus* is the fifth species-rich genus in the family with 37 species (Price et al. 2003). Only two species from each of these two genera have been studied to date; more species should be investigated in future studies to find out the taxonomic scale of mt genome fragmentation in these genera.

In addition to species of Menoponidae and Laemobothriidae, fragmented mt genomes were found previously in: 1) two other amblyceran species from different families, *Cummingsia maculate*

(Trimenoponidae) and *Macrogyropus costalimai* (Gyropidae) (Sweet et al., 2021); 2) 17 ischnoceran species from 15 genera (Sweet et al., 2020, Sweet et al., 2022); and 3) 22 species of eutherian mammal lice from three parvorders (Anoplura, Rhynchophthirina and Trichodectera) (Song et al., 2019). The evidence available to date indicates that mt genome fragmentation occurred at least 15 times independently among parasitic lice (infraorder Phthiraptera): once in eutherian mammal lice in three parvorders (Anoplura, Rhynchophthirina and Trichodectera) (Song et al., 2019; Sweet et al., 2022), five times in amblyceran lice (Sweet et al., 2022, and the present study), and nine times in ischnoceran lice (Sweet et al., 2022). The independent evolution of fragmented mt genomes is also supported by the observations of the typical single-chromosome mt genomes in 16 amblyceran species from four different families (Boopiidae, Laemobothriidae, Menoponidae and Ricinidae) (Shao et al., 2001; Song et al., 2019; Sweet et al., 2021; Gong et al., 2022 and current study), in 10 ischnoceran species from 10 genera (Covacin et al., 2006, Cameron et al., 2011, Song et al., 2019, Sweet et al., 2021) but not in any species in Anoplura, Rhynchophthirina or Trichodectera (Shao et al., 2009; Shao et al., 2012; Jiang et al., 2013; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Herd et al., 2015; Shao et al., 2015; Shao et al., 2017; Song et al., 2019; Fu & Dong et al., 2020; Dong et al., 2021; Spradling et al., 2021; Dong et al., 2022).

5.5.2. Minichromosomal characters derived from mt genome fragmentation provide valuable information for resolving phylogenetic relationships at different taxonomic levels

Song et al. (2019) showed that mt genome fragmentation and minichromosomal characters united parasitic lice of eutherian mammals in three parvorders: Anoplura, Rhynchophthirina and Trichodectera. All species in these three parvorders studied to date have fragmented mt genomes with varying number of minichromosomes from 9 to 20 in each species (Shao et al., 2009; Shao et al., 2012; Jiang et al., 2013; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Herd et al., 2015; Shao et al., 2015; Shao et al., 2017; Fu et al., 2020; Dong et al., 2021; Spradling et al., 2021; Dong et al., 2022). When only protein-coding and rRNA genes (these genes are more stable than tRNA genes in chromosomal locations) are considered, five minichromosomes, which contain *cox1*, *nad4*, *nad5*, *rrnS* and *rrnL* respectively, are in common among all eutherian mammal lice in these three parvorders whereas other minichromosomes are variable (Song et al., 2019).

In bird lice, extensively fragmented mt genomes were characterized previously in three *Columbicola* species (family Philopteridae, parvorder Ischnocera) (Sweet et al., 2020). Unlike

eutherian mammal lice in Anoplura, Rhynchophthirina and Trichodectera, these *Columbicola* species have very similar number of mt minichromosomes: 1) 15 for *Columbicola macrourae* (two protein-coding genes, *atp8* and *nad6* not identified); 2) 16 for *Columbicola columbae*; and 3) 17 for *Columbicola passerinae*. Furthermore, these *Columbicola* species share the same distribution pattern of all protein-coding and rRNA genes among the minichromosomes: each protein-coding or rRNA gene has its own minichromosome not shared with any other protein-coding or rRNA genes (Sweet et al., 2020). The differences among these *Columbicola* species are entirely in the distribution of tRNA genes among minichromosomes, which varies between species and even between two strains of *Columbicola passerinae* from different host species (Sweet et al., 2020). The genus *Columbicola* has 70 species (Price et al., 2003). It remains to be studied whether other *Columbicola* species have fragmented mt genomes. The minichromosomal characters derived from mt genome fragmentation are potentially a valuable source of information for resolving phylogenies among *Columbicola* species.

Sweet et al. (2022) showed recently that *Columbicola passerinae* (dove louse) was closely related to *Craspedonirmus immer* (louse of common loon), which had a fragmented mt genome with seven minichromosomes. It is possible that these two genera, *Columbicola* and *Craspedonirmus*, may share a common event of mt genome fragmentation. Our comparison of the mt karyotypes between these two genera, however, did not reveal any shared mt minichromosomal characters between them. This is also the case for another two pairs of ischnoceran genera (between *Pessoaiella* and *Osculotes*, between *Cheloptistes* and *Oxylipeurus*) and a pair of amblyceran genera (between *Cummingsia* and *Macrogyropus*) (Sweet et al., 2022); we did not find any shared mt minichromosomal characters between the two genera in each pair. Most of the mt minichromosomes of these genera are partially sequenced so far and most tRNA genes are not identified (Sweet et al., 2022). In addition, only one species from each of these genera has been sequenced except for *Columbicola*. More species from each genus and more complete mt genome data from these genera are needed to better understand the evolution of fragmented mt genomes in these genera. Nevertheless, we did find: 1) a derived *cox1-cob* minichromosome shared between two ischnoceran genera, *Penenirmus* and *Saemundssonia*; and 2) derived *cox2-nad1* and *nad5* minichromosomes shared between three ischnoceran genera, *Penenirmus*, *Quadraceps* and *Saemundssonia*. These three genera and *Alcedoecus* may share a common mt genome fragmentation event according to Sweet et al., (2022).

The fragmented mt genomes observed in species of Menoponidae and Laemobothriidae (parvorder Amblycera) in the current study and Sweet et al. (2021) differ from those seen previously in other parvorders. First, the mt genomes of these species of Menoponidae and Laemobothriidae are less fragmented. In contrast to 9 to 20 minichromosomes in eutherian mammal lice (Anoplura, Rhynchophthirina and Trichodectera) (Song et al., 2019) and 15-17 in *Columbicola* species (Ischnocera) (Sweet et al., 2020), *Austromenopon paululum* has only two minichromosomes (Fig. 5.2), *Laemobothrion tinnunculi* has three minichromosomes (Sweet et al., 2021), species of *Myrsidea* have 3-4 minichromosomes (14-15 tRNA genes not identified, Fig. 5.3, Sweet et al., 2021), and species of *Actornithophilus* have 5-6 minichromosomes (Fig. 5.1). Second, the eutherian mammal lice have a single non-coding region, 1,000 to 1,800 bp in size, in most minichromosomes (Shao et al., 2009; Shao et al., 2012; Jiang et al., 2013; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Herd et al., 2015; Shao et al., 2015; Shao et al., 2017; Fu & Dong et al., 2020; Dong et al., 2021). The *Columbicola* species have a single non-coding region, 1,000 to 1,500 bp in size, in most minichromosomes (Sweet et al., 2020). The species of Menoponidae and Laemobothriidae, however, tend to have two or more much shorter non-coding regions in each minichromosome, e.g., the M2 minichromosome of *Actornithophilus grandiceps* has four non-coding regions, 86-832 bp in size (Fig. 5.1). Third, the non-coding regions of eutherian mammal lice and *Columbicola* species are highly conserved in sequence among the minichromosomes within a species (Shao et al., 2009; Shao et al., 2012; Jiang et al., 2013; Dong et al., 2014a; Dong et al., 2014b; Song et al., 2014; Herd et al., 2015; Shao et al., 2015; Shao et al., 2017; Fu & Dong et al., 2020; Sweet et al., 2020). For the four species of Menoponidae with fragmented mt genomes, however, the non-coding region sequences are not conserved among different minichromosomes in any species.

Shared derived minichromosal characters are also observed in the species of Menoponidae. Currently, only two genera of Menoponidae, *Actornithophilus* and *Myrsidea*, each have two species known with fragmented mt genome, whereas *Austromenopon* and *Laemobothrion* each have only a single species known to date with fragmented mt genome (Fig. 5.2, Sweet et al., 2021). The two *Actornithophilus* species share exclusively one minichromosome (*E-nad4L-nad4*) and five gene clusters on other minichromosomes (*V-K-cob-nad1*, *nad6-H-S*$_2$, *nad2-S*$_1$*-F*, *Y-atp8-atp6-N* and *G-I-cox2-cox1-C-cox3*) (Fig. 5.1, Table 5.3). The two species of *Myrsidea* also share exclusively one minichromosome (*Y-nad5-L*$_1$*-nad3-S*$_1$*-cox2-cox1-cox3*) and a gene cluster (*G-cob-Q-atp6-atp8-*

nad4L-nad4-nad6-nad1) on a minichromosome (Supplementary Table 5.3). These unique shared derived characters are examples of possible synapomorphies produced by independent mt genome fragmentation events that can be further studied to help resolve the phylogenetic relationships among species within each of these genera.

5.6. Conclusions

In the current study, we sequenced the mt genomes of 17 species of bird lice in two families: Menoponidae and Laemobothriidae. We found that four species of Menoponidae (*Actornithophilus grandiceps*, *Actornithophilus hoplopteri*, *Austromenopon atrofulvum* and *Myrsidea ptilorhynchi*) have fragmented mt genomes whereas the other 13 species retain the typical single-chromosome mt genomes. We showed that mt genome fragmentation occurred once in the Laemobothriidae and three times independently in the Menoponidae: once in the genus *Myrsidea*, once in the genus *Actornithophilus* and once in *Austromenopon atrofulvum*. Looking more broadly, mt genome fragmentation occurred independently at least 15 times in parasitic lice: 1) once in eutherian mammal lice in the parvorders Anoplura, Rhynchophthirina and Trichodectera (Song et al., 2019); 2) nine times in the parvorder Ischnocera (Sweet et al., 2022); and 3) five times in the parvorder Amblycera (Sweet et al., 2022, and the current study). Each independent mt genome fragmentation event is expected to produce a unique pattern of gene distribution among minichromosomes not seen in other fragmentation events. We found derived minichromosomal characters shared between two species of *Myrsidea*, between two species of *Actornithophilus*, between two ischnoceran genera, *Penenirmus* and *Saemundssonia*, and among three ischnoceran genera, *Penenirmus*, *Quadraceps* and *Saemundssonia*, respectively. As a general feature, mt genome fragmentation does not unite all parasitic lice that have this feature. However, each independent mt genome fragmentation event does produce shared derived minichromosomal characters that can be informative in resolving phylogeny of parasitic lice at different taxonomic levels.

Acknowledgements

We would like to thank Rosemary Booth (AZWH), , Danny Brown (AZWH) and other staff in the AZWH, and Vasilli Kasimov (UniSC), to assist us collect specimens of bird lice. We thank Wei Wang for assistance with identification of bird lice. We thank the Associate Editor and the anonymous reviewers for comments that have greatly improved this manuscript. This study was

partially funded by the Australian Biological Resources Study (ABRS, RF217-51), and partially by an ECR and MCR grant from the Centre for Bioinnovation, University of the Sunshine Coast.

5.7. References

Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. ***Nucleic Acids Research***, 38(Web Server issue): W7-13.

Brown, W. M., George Jr, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. ***Proceedings of the National Academy of Sciences***, 76(4), 1967-1971.

Boore, J. L. (1999). Animal mitochondrial genomes. ***Nucleic Acids Research***, 27(8), 1767-1780.

Cameron, S. L., Yoshizawa, K., Mizukoshi, A., Whiting, M. F., & Johnson, K. P. (2011). Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). ***BMC Genomics***, 12(1), 1-15.

Covacin, C., Shao, R., Cameron, S., & Barker, S. C. (2006). Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). ***Insect Molecular Biology***, 15(1), 63-68.

Crampton-Platt, A., Yu, D. W., Zhou, X., & Vogler, A. P. (2016). Mitochondrial metagenomics: letting the genes out of the bottle. ***GigaScience***, 5(1), s13742-016.

Dong, W. G., Dong, Y., Guo, X. G., & Shao, R. (2021). Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals. ***BMC Genomics***, 22(1): 598.

Dong, W. G., Song,S., Guo,X. G., Jin, D. C., Yang, Q., Barker, S. C., Shao, R. (2014a). 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: 751.

Dong, W. G., Song, S., Jin, D. C., Guo, X. G., & Shao, R. (2014b). 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(1): 44.

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.

Fu, Y. T., & Dong, Y., Wang, W., Nie, Y., Liu, G. H., & Shao, R. (2020). Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse *Pedicinus obtusus* and colobus louse *Pedicinus badii*. ***Genomics***, 112(6): 4924-4933.

Gong, S., Gong, S., Xu, Y., Xu, S., Liang, Y., Tian, L., Cai, W., ... & Song, F. (2022). The complete mitochondrial genome of the chicken body louse, *Menacanthus cornutus*, and evolutionary patterns of extensive gene rearrangements in the mitochondrial genomes of Amblycera (Psocodea: Phthiraptera). ***Genes***, *13*(3), 522.

Graham Pizzey, Frank Knight, Sarah Pizzey (2012). The Field Guide to the Birds of Australia 9th Edition. ***HarperCollins AU***.

Herd, K. E., 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 Genomics***, 16(1): 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 Biology and Evolution***, 5(7): 1298-1308.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. ***Nucleic Acids Research***, 36(suppl_2), W5-W9..

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. ***Nature Methods***, 14(6): 587-589.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., ... & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. ***Bioinformatics***, 28(12): 1647-1649.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. ***Molecular Biology and Evolution***, 34(3), 772-773.

Laslett, D., & Canback, B. (2008). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. ***Bioinformatics***, 24(2): 172-175.

Lavrov, D. V., & Bell, E. (2014). Mitochondrial genomes in invertebrate animals. Molecular life sciences: an encyclopedic reference. ***New York (NY): Springer New York***, 1-8

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. ***Nucleic Acids Research***, 25(5): 955-964.

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., ... & Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. ***Science***, 346(6210), 763-767.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. ***Molecular Biology and Evolution***, 32(1): 268-274.

Peter Menkhorst (Author), Danny Rogers (Author), Rohan Clarke (Author), Jeff Davies (Illustrator), Peter Marsack (Illustrator), Kim Franklin (Illustrator) (2019). The Australian Bird Guide: Revised Edition. ***CSIRO Publishing***.

Price, R. D., Clayton, D., Hellenthal, R. A., Johnson, K. P., & Palma, R. 2003. The Chewing Lice: World Checklist and Biological Overview; ***Illinois Natural History Survey Special Publication***, 24: 1-448.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. ***Systematic Biology***, 67(5), 901-904.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. ***Bioinformatics***, 19(12): 1572-1574.

Shao, R., Barker, S. C., Li, H., Song, S., Poudel, S., & Su, Y. (2015). Fragmented mitochondrial genomes in two suborders of parasitic lice of eutherian mammals (Anoplura and Rhynchophthirina, Insecta). ***Scientific Reports***, 5: 17389.

Shao, R., Kirkness, E. F., & Barker, S. C. (2009). The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. ***Genome Research***, 19(5): 904-912.

Shao, R., Li, H., Barker, S. C., & Song, S. (2017). The mitochondrial genome of the guanaco louse, *Microthoracius praelongiceps*: insights into the ancestral mitochondrial karyotype of sucking lice (Anoplura, Insecta). ***Genome Biology and Evolution***, 9(2): 431-445.

Shao, R., Zhu, X.-Q., Barker, S. C., & Herd, K. (2012). Evolution of extensively fragmented mitochondrial genomes in the lice of humans. ***Genome Biology and Evolution***, 4(11): 1088-1101.

Song, F., Li, H., Liu, G. H., Wang, W., James, P., ... & Shao, R. (2019). Mitochondrial genome fragmentation unites the parasitic lice of eutherian mammals. ***Systematic Biology***, 68(3): 430-440.

Song, S. D., Barker, S. C., & Shao, R. (2014). Variation in mitochondrial minichromosome composition between blood-sucking lice of the genus *Haematopinus* that infest horses and pigs. ***Parasit Vectors***, 7: 144.

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

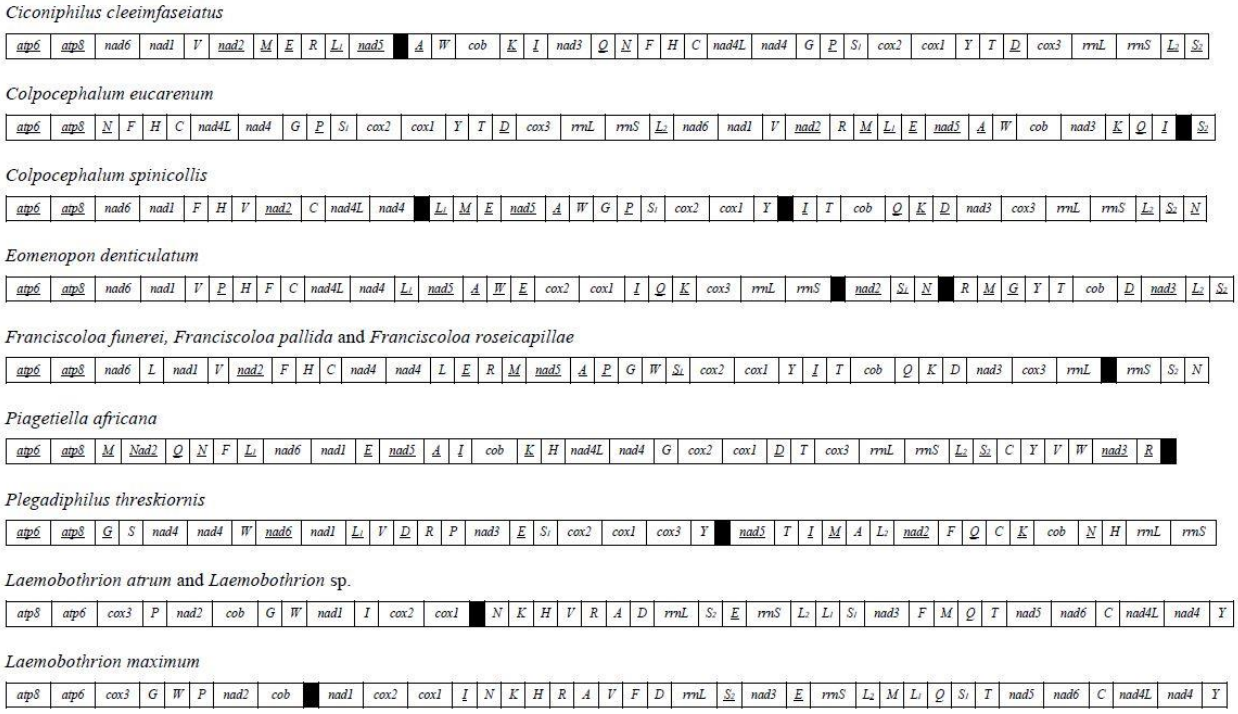
Sweet, A. D., Johnson, K. P., Cao, Y., de Moya, R. S., Skinner, R. K., ... & Cameron, S. L. (2021). Structure, gene order, and nucleotide composition of mitochondrial genomes in parasitic lice from Amblycera. ***Gene***, 768, 145312.

Sweet, A. D., Johnson, K. P., & Cameron, S. L. (2022). Independent evolution of highly variable, fragmented mitogenomes of parasitic lice. ***Communications Biology***, 5(1), 1-10.

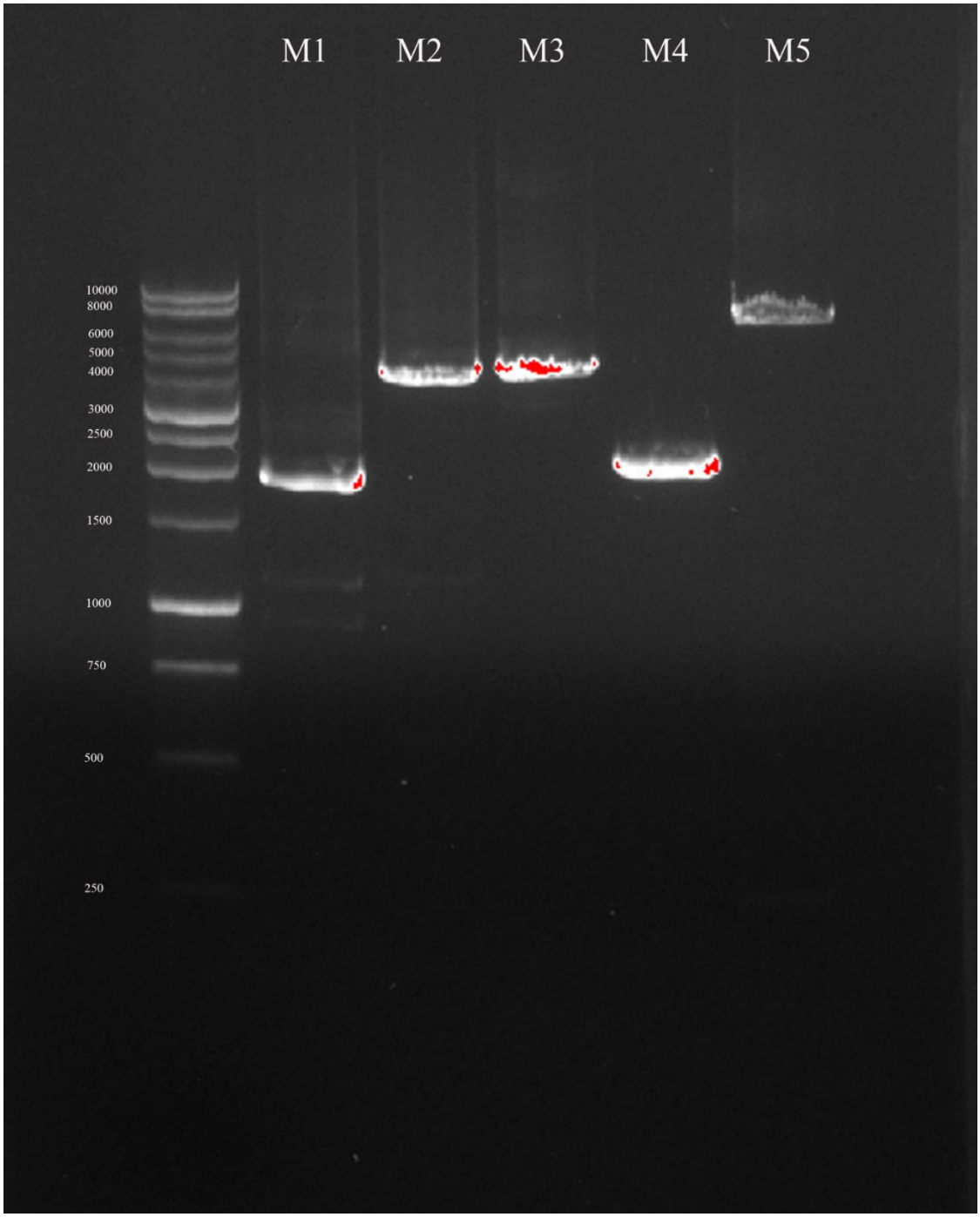
Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. ***Systematic Biology***, 56(4), 564-577.

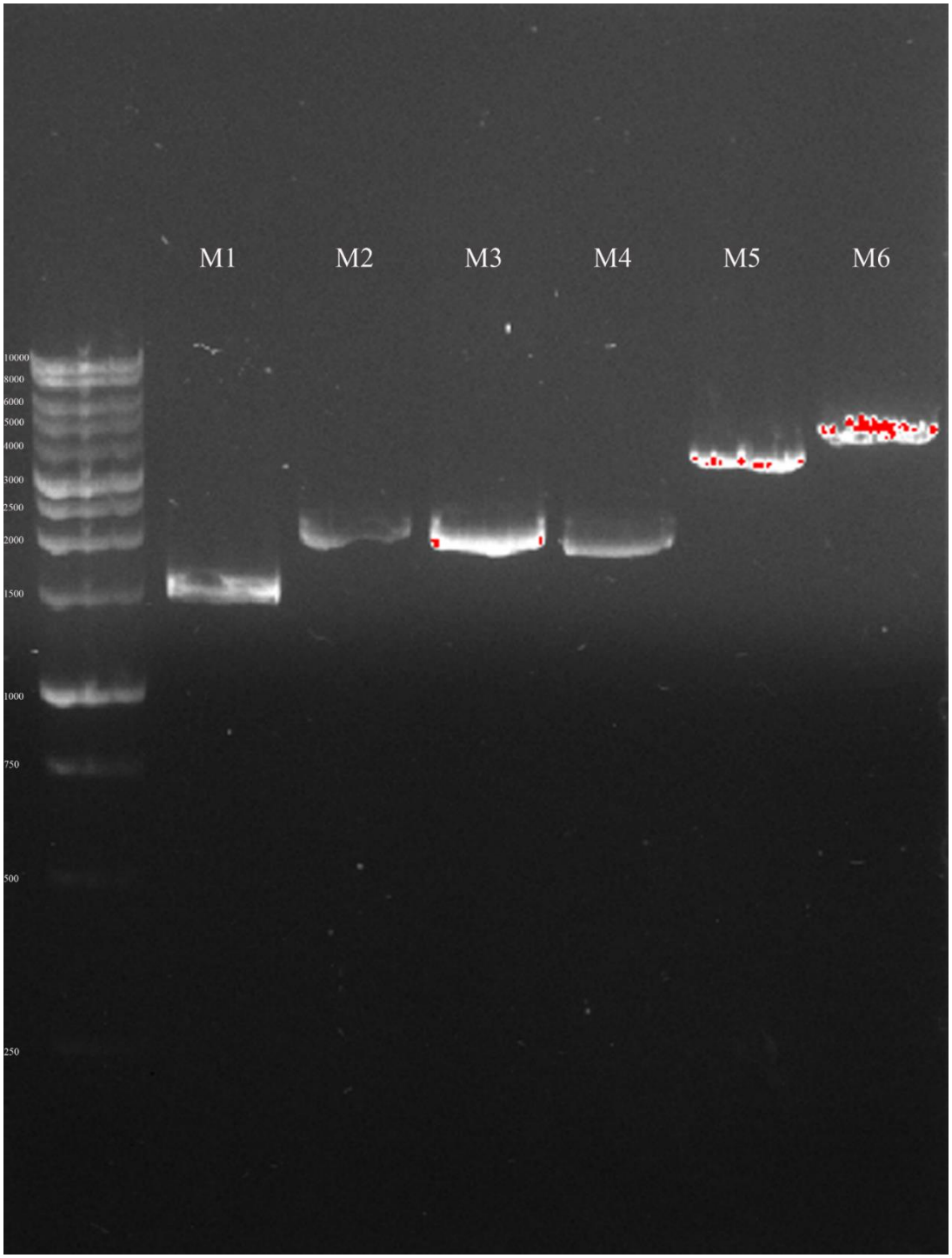
5.8. Supplementary files



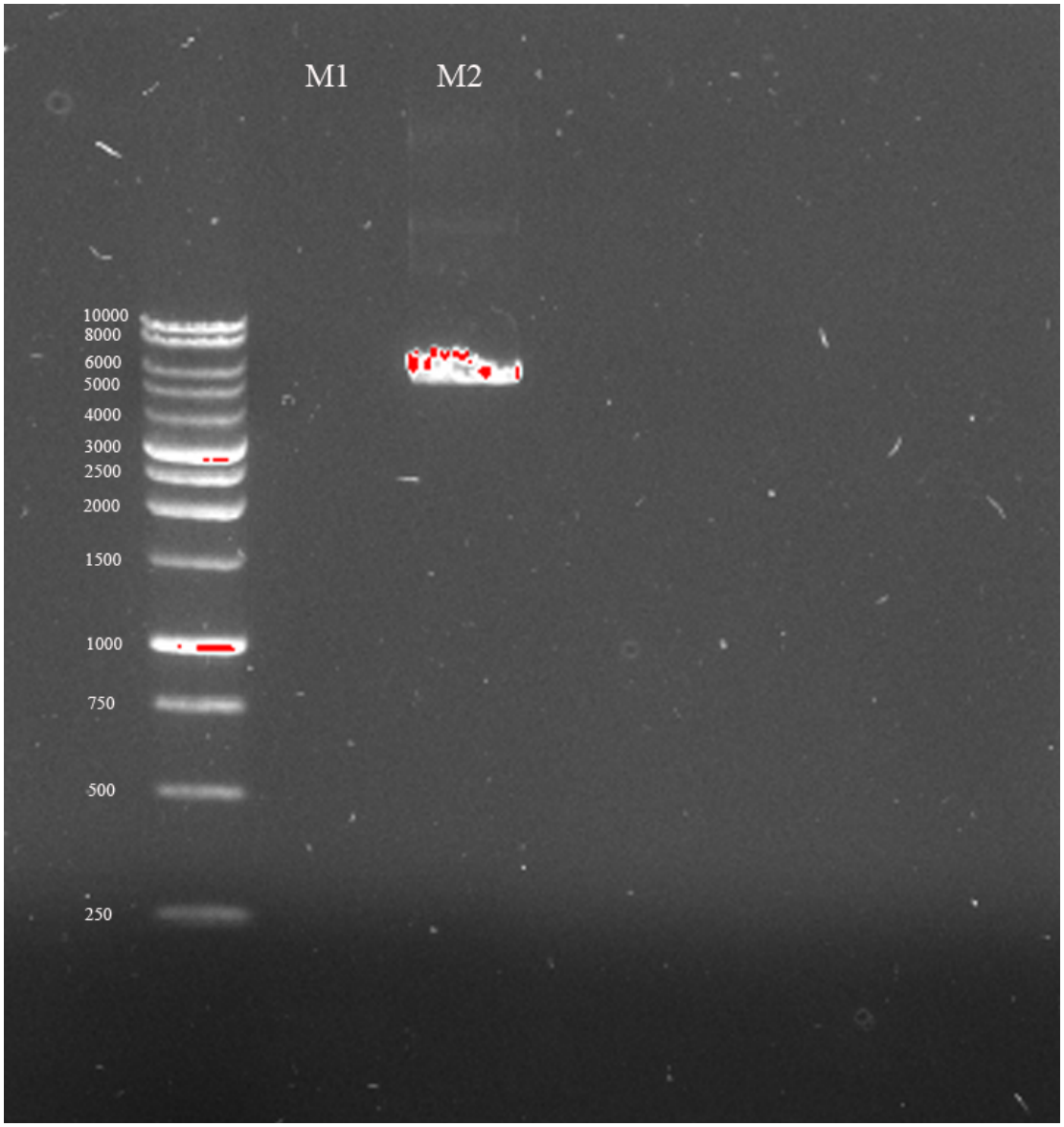
Supplementary Fig. 5.1. The mitochondrial genomes of *Ciconiphilus decimfasciatus* (louse of white-faced heron), *Colpocephalum eucarenum* (louse of Australian pelican), *Colpocephalum spinicollis* (louse of straw-necked ibis), *Eomenopon denticulatum* (louse of scaly breasted lorikeet), *Franciscoloa funerei* (louse of yellow-tailed black cockatoo), *Franciscoloa pallida* (louse of sulfur-crested cockatoo), *Franciscoloa roseicapillae* (louse of little corella, galah and pheasant coucal), *Piagetiella africana* (louse of Australian pelican), *Plegadiphilus threskiornis* (louse of Australian white ibis), *Laemobothrion atrum* (louse of Eurasian coot), *Laemobothrion* sp. (louse of Australian swamphen) and *Laemobothrion maximum* (louse of black kite). tRNA genes are indicated with their single-letter abbreviations of the corresponding amino acids. Genes are transcribed from left to right except those underlined, which have an opposite orientation of transcription. Non-coding regions are in black. Circular mitochondrial genomes are linearised for the sake of illustration.



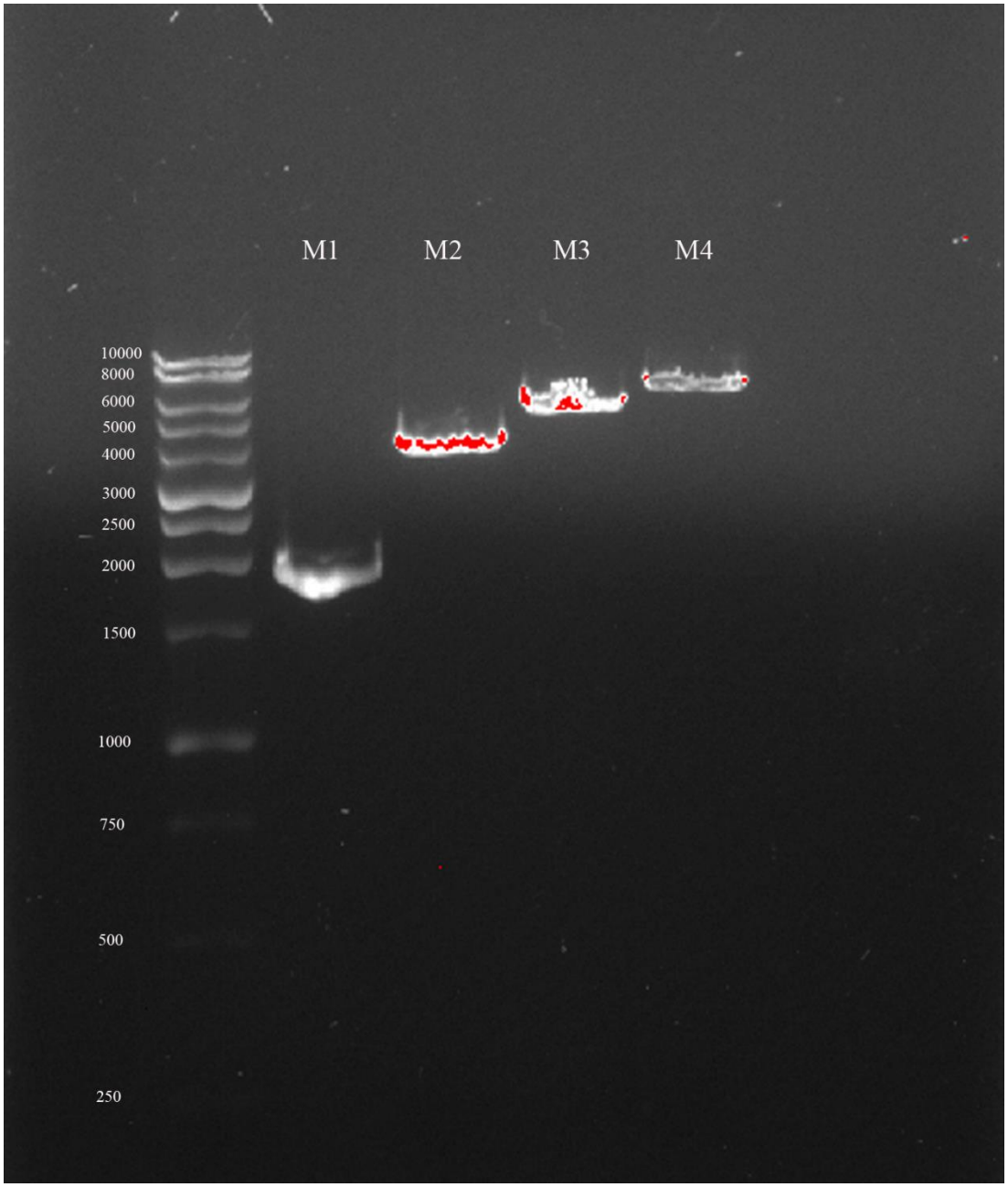
Supplementary Fig. 5.2. PCR verification of the five mitochondrial minichromosomes (M1-M5) of *Actornithophilus grandiceps* (louse of pied oystercatcher). Marker band sizes (bp) are indicated on the left side.



Supplementary Fig. 5.3. PCR verification of the six mitochondrial minichromosomes (M1-M6) of *Actornithophilus hoplopteri* (louse of masked lapwing). Marker band sizes (bp) are indicated on the left side.



Supplementary Fig. 5.4. PCR verification of the two mitochondrial minichromosomes (M1 and M2) of *Austromenopon atrofulvum* (louse of sooty tern louse). Marker band sizes (bp) are indicated on the left side.



Supplementary Fig. 5.5. PCR verification of the four mitochondrial minichromosomes (M1-M4) of *Myrsidea ptilorhynchi* (louse of satin bowerbird). Marker band sizes (bp) are indicated on the left side.

Supplementary Table 5.1. Specific primers used for PCR verification of mitochondrial minichromosomes and the number of Illumina sequence reads obtained from each PCR amplicon

Minichromosome	Species	Forward prime	Reverse Prime	Number of reads
E-nad4L-nad4	*Actornithophilus grandiceps*	AAGGGTTAGGGAGTTTTTACTAGGAATAAG	GAAGTTAACCCATGAGAAATTATGAGGATC	19,380,282
F-S1-nad2-L1-W-S2-H-nad6	*Actornithophilus grandiceps*	CCGAGTGGAGGAAGGGAAGGAACCT	CAGAAAGGGTGGTGGAGCTCGAGAAG	16,833,026
V-K-cob-nad1-Q-A-nad3	*Actornithophilus grandiceps*	GATGATTCAAAACCACATTCAAAAGGAGAA	GATGATTCAAAACCACATTCAAAAGGAGAA	18,067,258
rrnL-rrnS	*Actornithophilus grandiceps*	CTTCCCTTGTTTAACCTTTCATTCCATCAC	CTGATGGTAGACGAGAAGACCCTATAGATC	16,228,942
Y-atp8-atp6-N-nad5-M-D-T-L2-G-I-cox2-cox1-C-cox3	*Actornithophilus grandiceps*	GTAGCAGGTTACATAATCCATTCTTCTTCC	ATCTACTTCTACTAACCCTCTTAGCCTAGC	20,100,600
E-nad4L-nad4	*Actornithophilus hoplopteri*	ATTTTTGATTACCTAAAGCCCACGTTGAGG	GAAGCTTCAATTCGTTCAGGTTGGTAAC	16,105,462
V-K-cob-nad1	*Actornithophilus hoplopteri*	ATCGAGTATGGGGGGATTATGTACACTATG	CAATTGATCGAACCCTTCCTAACATAGC	16,120,914
nad3-A-W	*Actornithophilus hoplopteri*	GGTATCCCTGTGGTGGTTGTATCTTTATAG	CCTAAAATGGTCTTACTCATACCCCTCAG	16,461,668
L1-Q-rrnS-rrnL	*Actornithophilus hoplopteri*	CCTTAGCCCTTCCTCCTTTATTTAAACTTC	GGCAAACGCTTTTCCTGACTGTTTAGC	18,651,904
F-S1-nad2-L2-D-nad5-S2-H-nad6	*Actornithophilus hoplopteri*	ATACAAACCTAACAACAACCATCTCCTATG	GATATTGGTTGCTACTCTATGGGTTAGG	17,282,038
Y-atp8-atp6-N-M-T-G-I-cox2-cox1-C-cox3	*Actornithophilus hoplopteri*	AATTCGAGTTGAGTTAGCTTTTCCAGGGAC	CCTAAAAGCCCAGACCAAATTCCAAATAAC	17,162,222
atp8-atp6-cox1-cox2-G-R-nad1-nad6-L2-N-cob-cox3-H-I-nad5-nad2	*Austromenopon atrofulvum*	GTCCCTCACATTATAAGAGACTGAGTC	GAGCTTTTCTTGGACTCAGGTAAACTTTC	N.A.
K-Q-nad4L-nad4-E-T-V-A-C-W-nad3-Y-D-S1-M-F-rrnL-rrnS-L1-P	*Austromenopon atrofulvum*	TCGTACTTTGAGATATCTGGTTCTTTAGG	ATTAGCCTTTCTTTCCATGGACCATTATGC	18,149,704
rrnL	*Myrsidea ptilorhynchi*	AGTTCCGCGGCTCTTTAAAACTTCAGTGAG	CATGGTAGACGAGAAGACCCTGTAGATC	16,316,178
rrnS	*Myrsidea ptilorhynchi*	GCCCACCGCCGTCTATGTTAACTCAGGAG	GTGTATCTTCGTTTACTGAGCAGTTCCTC	18,548,112
Y-nad5-L1-nad3-S1-cox2-cox1-cox3	*Myrsidea ptilorhynchi*	ATGGGAGCAATTAACTTTATCTCTACCCTC	GTAAAGAGAAAATAGCCATATCCACGGAAG	17,422,022
M-R-G-cob-Q-atp6-atp8-nad4L-nad4-nad6-nad1-V-nad2	*Myrsidea ptilorhynchi*	CTCCTGTGAAGATTAATTTTGTTTTTTTCG	TGGCTGATCCCACCCAAAATATAGCAAGG	16,211,864

Supplementary Table 5.2. The mitochondrial minichromosomes of four bird louse species (family Menoponidae) identified by assembly of Illumina sequence reads (whole genomic DNA sequencing)

Louse species (Bird host)	Minichromosome	Size (bp)	Number of Illumina sequence-reads	Mean coverage
Actornithophilus grandiceps (pied oystercatcher)	*E-nad4L-nad4*	2,005	27,366	1,900
	F-S1-nad2-L1-W-S2-H-nad6	2,068	23,907	1,720
	V-K-cob-nad1-Q-A-nad3	3,950	77,869	2,956
	rrnL-rrnS	3,939	115,421	4,396
	Y-atp8-atp6-N-nad5-M-D-T-L2-G-I-cox2-cox1-C-cox3	6,112	21,714	2,023
Actornithophilus hoplopteri (masked lapwing)	*E-nad4L-nad4*	1,698	7,177	634
	V-K-cob-nad1	2,215	14,085	955
	nad3-A-W	2,102	19,763	1,406
	L1-Q-rrnS-rrnL	2,125	43,077	3,036
	F-S1-nad2-L2-D-nad5-S2-H-nad6	3,645	55,351	2,278
	Y-atp8-atp6-N-M-T-G-I-cox2-cox1-C-cox3	4,361	67,577	2,323
Austromenopon atrofulvum (sooty tern)	*atp8-atp6-cox1-cox2-G-R-nad1-nad6-L2-N-cob-cox3-H-I-nad5-nad2*	9,499	99,206	1,567
	K-Q-nad4L-nad4-E-T-V-A-C-W-nad3-Y-D-S1-M-F-rrnL-rrnS-L1-P	4,943	54,918	1,665
Myrsidea ptilorhynchi (satin bowerbird)	*S1-cox2-cox1-cox3-Y-nad5-L1-nad3*	5,167	58,205	1,690
	Cob-Q-atp8-atp6-nad4L-nad4-nad6-mad1-nad2-M-R-G	6,185	68,633	1,665
	rrnL	1,969	5,551	422
	rrnS	3,914	41,381	1,595

Supplementary Table 5.3. Derived minichromosomes and gene clusters observed in Menoponidae and Laemobothriidae species

Family	Species	Fragmented	*E-nad4L-nad4* (as one minichromosome)	*V-K-cob-nad1*	*nad6-H-S2*	*nad2-S1-F*	*Y-atp8-atp6-N*	*G-I-cox2-cox1-C-cox3*	*cox2-cox1-atp6-atp8*
Menoponidae	*Actornithophilus grandiceps*	+	+	+	+	+	+	+	-
	Actornithophilus hoplopteri	+	+	+	+	+	+	+	-
	Amyrsidea minuta	-	-	-	-	-	-	-	-
	Austromenopon atrofulvum	+	-	-	-	-	-	-	+
	Austromenopon paululum	-	-	-	-	-	-	-	+
	Ciconiphilus Cleeimfaseiatus	-	-	-	-	-	-	-	-
	Colpocephalum eucarenum	-	-	-	-	-	-	-	-
	Colpocephalum griffoneae	-	-	-	-	-	-	-	-
	Colpocephalum spinicollis	-	-	-	-	-	-	-	-
	Eomenopon denticulatum	-	-	-	-	-	-	-	-
	Franciscoloa funerei	-	-	-	-	-	-	-	-
	Franciscoloa pallida	-	-	-	-	-	-	-	-
	Franciscoloa roseicapillae	-	-	-	-	-	-	-	-
	Myrsidea ptilorhynchi	+	-	-	-	-	-	-	-
	Myrsidea sp.	+	-	-	-	-	-	-	-
	Osborniella crotophagae	-	-	-	-	-	-	-	-
	Piagetiella africana	-	-	-	-	-	-	-	-
	Plegadiphilus threskiornis	-	-	-	-	-	-	-	-
Laemobothriidae	*Laemobothrion atrum*	-	-	-	-	-	-	-	-
	Laemobothrion maximum	-	-	-	-	-	-	-	-
	Laemobothrion sp.	-	-	-	-	-	-	-	-
	Laemobothrion tinnunculi	+	-	-	-	-	-	-	-

Supplementary Table 5.3. Derived minichromosomes and gene clusters observed in Menoponidae and Laemobothriidae species (continued)

Family	Species	*nad2-nad5-I*	*nad3-Y*	*rrnL-rrnS-L1*	*nad4L-nad4-E*	*P-K-S1*	*Y-nad5-L1-nad3-S1-cox2-cox1-cox3* (as one minichromosome)	*G-cob-Q-atp6-atp8-nad4L-nad4-nad6-nad1*	*D-rrnL-S2*	*E-rrnS-L2*	*nad5-nad6-C-nad4L-nad4-Y-atp8-atp6*
Menoponidae	*Actornithophilus grandiceps*	-	-	-	-	-	-	-	-	-	-
	Actornithophilus hoplopteri	-	-	-	-	-	-	-	-	-	-
	Amyrsidea minuta	-	-	-	-	-	-	-	-	-	-
	Austromenopon atrofulvum	+	+	+	+	+	-	-	-	-	-
	Austromenopon paululum	+	+	+	+	+	-	-	-	-	-
	Ciconiphilus Cleeimfaseiatus	-	-	-	-	-	-	-	-	-	-
	Colpocephalum eucarenum	-	-	-	-	-	-	-	-	-	-
	Colpocephalum griffoneae	-	-	-	-	-	-	-	-	-	-
	Colpocephalum spinicollis	-	-	-	-	-	-	-	-	-	-
	Eomenopon denticulatum	-	-	-	-	-	-	-	-	-	-
	Franciscoloa funerei	-	-	-	-	-	-	-	-	-	-
	Franciscoloa pallida	-	-	-	-	-	-	-	-	-	-
	Franciscoloa roseicapillae	-	-	-	-	-	-	-	-	-	-
	Myrsidea ptilorhynchi	-	-	-	-	-	+	+	-	-	-
	Myrsidea sp.	-	-	-	-	-	+	+	-	-	-
	Osborniella crotophagae	-	-	-	-	-	-	-	-	-	-
	Piagetiella africana	-	-	-	-	-	-	-	-	-	-
	Plegadiphilus threskiornis	-	-	-	-	-	-	-	-	-	-
Laemobothriidae	*Laemobothrion atrum*	-	-	-	-	-	-	-	+	+	+
	Laemobothrion maximum	-	-	-	-	-	-	-	+	+	+
	Laemobothrion sp.	-	-	-	-	-	-	-	+	+	+
	Laemobothrion tinnunculi	-	-	-	-	-	-	-	+	+	+

Chapter 6

Summary and Future Study

I investigated mt genome evolution and molecular phylogeny of parasitic lice in this thesis. With the assistance of my principal supervisor and collaborators, I collected lice from 121 bird species in the South East Queensland region, mostly from euthanised birds at the Australian Zoo Wildlife Hospital. I sequenced the mt genomes of 83 species of bird lice, among which 17 species were included in this thesis. I analysed the mt genomes of: 1) seven Anoplura species retrieved and assembled from the publicly available Sequence Read Archive data (SRA, http://www.ncbi.nlm.nih.gov/sra/); 2) two Anoplura species sequenced by my collaborators; and 3) 17 Amblycera species that I sequenced in my PhD project. I reconstructed the phylogenies of sucking lice (Anoplura) and chewing lice of the families Menoponidae and Laemobothriidae (Amblycera) using mt genome sequences. I inferred the ancestral mt karyotypes for higher-primate lice, *Polyplax* lice, and seal lice, respectively. I compared the mt karyotypes: 1) between higher primate lice and their MRCA, and 2) between seal lice and their MCRA to understand mt genome fragmentation. I analysed statistically tRNA gene translocation in mammal lice. I did PCR experiments to verify the fragmented mt genomes of bird lice (Menoponidae) that I found initially by Illumina sequence data analysis. I identified shared minichromosomal characters derived from mt genome fragmentation in *Actornithophilus* and *Myrsidea* species. Below, I summarise the major findings in this thesis and discuss future study in this field.

In Chapter 2, I showed that the variation in mt karyotype between two *Pedicinus* species of monkey lice was the most pronounced among congeneric species of sucking lice observed to date. The variation was attributable to the opposite directions between these two species in mt genome evolution. I showed that mt genome fragmentation was a two-way process in these monkey lice: a fragmented mt genome could evolve to become more fragmented or reverse back to be less fragmented.

In Chapter 3, I showed that tRNA gene translocation in mammal lice (Anoplura) occurred much more frequently than that of protein-coding and rRNA genes. tRNA gene translocation is directional towards the boundaries with control regions and is a major contributing factor to the highly dynamic mitochondrial genome organization observed in mammal lice.

In Chapter 4, I showed that: 1) minichromosome split contributes much more than minichromosome merger in mt genome fragmentation of seal lice; and 2) mt karyotype

comparison helps understand the phylogenetic relationships among seal lice. My analyses of mt karyotypes and gene sequences also indicated the possibility of a host switch of crabeater seal louse to Weddell seals.

In chapter 5, I showed that mt genome fragmentation occurred four times independently in each of the four genera: *Actornithophilus*, *Austromenopon*, *Myrsidea* and *Laemobothrion*. I also found unique derived minichromosomal characters shared between *Myrsidea* species, and between *Actornithophilus* species, respectively. My analysis indicated that while mt genome fragmentation as a general feature does not unite all parasitic lice that have this feature, each independent mt genome fragmentation event does produce unique shared derived minichromosomal characters that can be informative in resolving the phylogeny of parasitic lice at different taxonomic levels.

There are several unresolved questions in my PhD project that should be investigated further in future studies. Firstly, how widespread did mt genome fragmentation occur in the parvorders Amblycera and Ischnocera? Amblycera has 1,200 described species in 102 genera of six families and Ischnocera has over 3,500 described species in the family Philopteridae (Price et al., 2003; De Moya et al., 2021). So far, only eight Amblycera species from six genera of four families (Menoponidae, Laemobothriidae, Trimenoponidae, Gyropidae) (Sweet et al., 2021; Chapter 5) and 17 Ischnocera species from 15 genera of the family Philopteridae were reported with fragmented mt genomes (Sweet et al., 2020, Sweet et al., 2022). More broadly sampling of species in the five families mentioned above are necessary to reveal the full scale of mt genome fragmentation in amblyceran lice and ischnoceran lice. Secondly, how minichromosomal characters derived from mt genome fragmentation can be used to resolve the phylogeny of parasitic lice? Derived shared minichromosomal characters are potentially valuable markers for resolving the phylogeny of parasitic lice at different taxonomic levels (Chapter 5). Further sampling and sequencing of parasitic lice in the genera and families in which fragmented mt genomes have been found are needed. Thirdly, why did mt genome fragmentation occur so many times independently in parasitic lice? Among insects (Class: Insecta), fragmented mt genomes have been found in parasitic lice (Chapter 5), booklice (Wei et al., 2012; Chen et al., 2014a; Perlman et al., 2015; Feng et al., 2018; Feng et al., 2019) and thrips (order Thysanoptera) (Dickey et al., 2015). It is still not clear what factors contributed to mt genome fragmentation in parasitic lice. Krikness et al. (2010) hypothesised that the loss of mitochondrial single-

stranded binding protein (mtSSB) might contribute to the mt genome fragmentation in the human body louse. This hypothesis was based on the observation that an orthologue to the *mtSSB* gene of *Drosophila melanogaster* was not found in the human body louse. mtSSB protein plays important roles in mt genome replication, repair and recombination (Kornberg and Barker, 1992). Mutation in *mtSSB* gene can lead to defect in DNA synthesis and cause mt diseases (Farr et al., 2004). Parasitic lice are well suited to the investigation of the role of *mtSSB* (or other genes with similar function) in mt genome fragmentation as both the typical single-chromosome mt genomes and different types of fragmented mt genomes have been found in parasitic lice. With more and more genomes being sequenced, it is possible to compare between louse species with fragmented mt genomes and their closely related species with typical mt genomes to reveal whether *mtSSB* gene or other nuclear genes are indeed associated with mt genome fragmentation.

Genome fragmentation occurred not only in the mitochondria of parasitic lice, but also in: 1) the mitochondria of booklice (Feng et al., 2022); 2) the mitochondria of nematodes (e.g. *Globodera pallida*) (Armstrong et al., 2000); 3) the chloroplasts of dinoflagellates (Barbrook et al., 2019); and 4) the cytomegalovirus (Martin, 1996). The factors contributing to their genome fragmentation and the advantages of fragmented mt genome in their evolution are not resolved yet. Analyzing the shared derived characters among these fragmented genomes can also help understand the phylogeny of these species.

This thesis is a major contribution to our understanding of the evolution of the mt genomes of parasitic lice. For the first time, I showed that: 1) mt karyotype evolved in opposite directions between two closely related congeneric species: macaque louse (*Pedicinus obtusus*) and colobus louse (*Pedicinus badii*); 2) for sucking lice, tRNA gene translocation is frequent between different types of minichromosomes and towards the boundaries with control regions; 3) minichromosomal split was much more frequent than minichromosomal merger in seal lice, and a host switch likely occurred from crabeater seals to Weddell seals; 4) mt genome fragmentation occurred four times independently in each of the four genera of Amblycera: *Actornithophilus*, *Austromenopon*, *Myrsidea* and *Laemobothrion*; and 5) each independent mt genome fragmentation event produced unique shared derived minichromosomal characters that

could be informative in resolving the phylogeny of parasitic lice at different taxonomic levels.

References

Armstrong, M. R., Blok, V. C., & Phillips, M. S. (2000). A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. ***Genetics***, 154(1), 181-192.

Barbrook, A. C., Howe, C. J., & Nisbet, R. E. R. (2019). Breaking up is hard to do: the complexity of the dinoflagellate chloroplast genome. ***Perspectives in Phycology***, 6 (1-2), 31-37. https://doi.org/10.1127/pip/2018/0084

Chen, S. C., Wei, D. D., Shao, R., Shi, J. X., Dou, W., & Wang, J. J. (2014a). Evolution of multipartite mitochondrial genomes in the booklice of the genus *Liposcelis* (Psocoptera). ***BMC Genomics***, 15(1): 861.

De Moya, R. S., Yoshizawa, K., Walden, K. K., Sweet, A. D., Dietrich, C. H., & Kevin P, J. (2021). Phylogenomics of parasitic and nonparasitic lice (Insecta: Psocodea): combining sequence data and exploring compositional bias solutions in next generation data sets. ***Systematic Biology***, 70(4), 719-738.

Dickey, A. M., Kumar, V., Morgan, J. K., Jara-Cavieres, A., Shatters, R. G., McKenzie, C. L., & Osborne, L. S. (2015). A novel mitochondrial genome architecture in thrips (Insecta: Thysanoptera): extreme size asymmetry among chromosomes and possible recent control region duplication. ***BMC Genomics***, 16(1), 1-7.

Farr, C. L., Matsushima, Y., Lagina, A. T., Luo, N., & Kaguni, L. S. (2004). Physiological and biochemical defects in functional interactions of mitochondrial DNA polymerase and DNA-binding mutants of single-stranded DNA-binding protein. ***Journal of Biological Chemistry***, 279(17), 17047-17053.

Feng, S., Yang, Q., Li, H., Song, F., Stejskal, V., Opit, G. P., ... & Shao, R. (2018). The highly divergent mitochondrial genomes indicate that the booklouse, *Liposcelis bostrychophila* (Psocoptera: Liposcelididae) is a cryptic species. ***G3: Genes, Genomes, Genetics***, 8(3), 1039-1047.

Feng, S., Li, H., Song, F., Wang, Y., Stejskal, V., Cai, W., & Li, Z. (2019). A novel mitochondrial genome fragmentation pattern in *Liposcelis brunnea*, the type species of the genus *Liposcelis* (Psocodea: Liposcelididae). ***International Journal of Biological Macromolecules***, 132, 1296-1303.

Feng, S., Pozzi, A., Stejskal, V., Opit, G., Yang, Q., Shao, R., ... & Li, Z. (2022). Fragmentation in mitochondrial genomes in relation to elevated sequence divergence and extreme rearrangements. ***BMC biology***, 20(1), 1-17.

Jahn, C. L., & Klobutcher, L. A. (2002). Genome remodeling in ciliated protozoa. ***Annual review of microbiology***, 56, 489.

Kornberg, A., and Baker, T. A. (1992) ***DNA Replication***, 2nd Ed., W. H. Freeman and Company, New York.

Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, M. A., Clark, J. M., ... & Pittendrigh, B. R. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. ***Proceedings of the National Academy of Sciences***, 107(27), 12168-12173.

Martin, J. (1996). Genetic instability and fragmentation of a stealth viral genome. ***Pathobiology***, 64(1), 9-17

Perlman, S. J., Hodson, C. N., Hamilton, P. T., Opit, G. P., & Gowen, B. E. (2015). Maternal transmission, sex ratio distortion, and mitochondria. ***Proceedings of the National Academy of Sciences***, 112(33), 10162-10168.

Price, R. D., Clayton, D., Hellenthal, R. A., Johnson, K. P., & Palma, R. (2003). The chewing lice: world checklist and biological overview; ***Illinois Natural History Survey Special Publication***, 24: 1-448.

Seegmiller, A., Williams, K. R., & Herrick, G. (1997). Two two‐gene macronuclear chromosomes of the hypotrichous ciliates *Oxytricha fallax* and *O. trifallax* generated by alternative processing of the 81 locus. ***Developmental genetics***, *20*(4), 348-357.

Sweet, A. D., Johnson, K. P., & Cameron, S. L. J. P. (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., ... & Cameron, S. L. (2021). Structure, gene order, and nucleotide composition of mitochondrial genomes in parasitic lice from Amblycera. ***Gene***, 768, 145312.

Sweet, A. D., Johnson, K. P., & Cameron, S. L. (2022). Independent evolution of highly variable, fragmented mitogenomes of parasitic lice. ***Communications Biology***, 5(1), 1-10.

Tree of Life Web Project. 2009a. *Actornithophilus* Ferris, 1916. Version 14 April 2009 (temporary). http://tolweb.org/Actornithophilus/70969/2009.04.14 in The Tree of Life Web Project, http://tolweb.org/

Tree of Life Web Project. 2009b. *Myrsidea* Waterston, 1915. Version 14 April 2009 (temporary). http://tolweb.org/Myrsidea/71733/2009.04.14 in The Tree of Life Web Project, http://tolweb.org/

Wei, D. D., Shao, R. F., Yuan, M. L., Dou, W., Barker, S. C., & Wang, J. J. (2012). The Multipartite Mitochondrial Genome of *Liposcelis bostrychophila*: Insights into the Evolution of Mitochondrial Genomes in Bilateral Animals. ***Plos One***, 7(3).

Appendix I

Manuscript:

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

Published in September 2020

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
0888-7543/ © 2020 Elsevier Inc. All rights reserved.

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′-GGAGG ATTTGGAAATTGATTAGTTCC-3′) – mtd11 (3′-ACTGTAAATATATGAT GAGCTCA-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
Microthoracius 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.

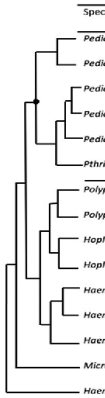
3. Results

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

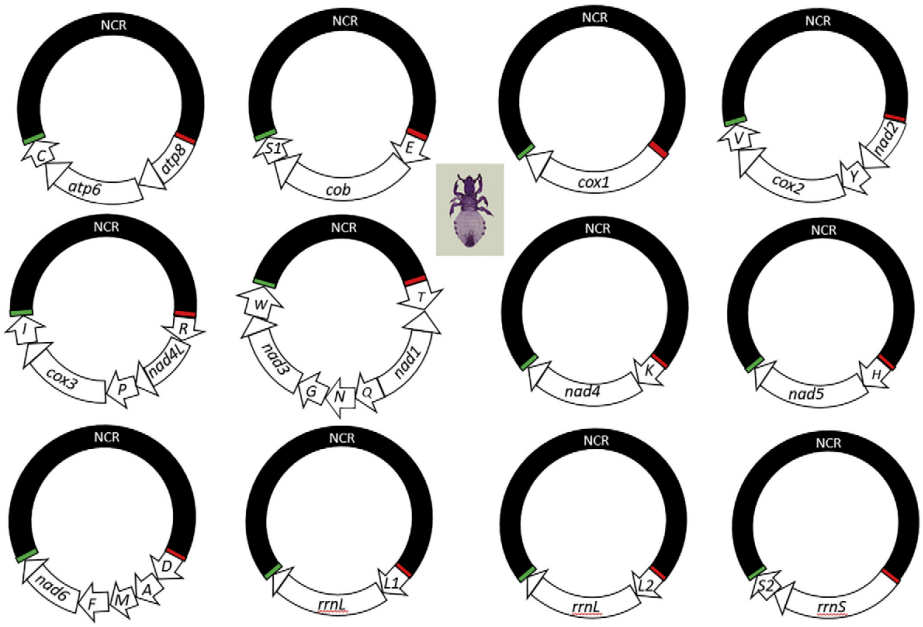


Fig. 3. The mitochondrial genome of macaque louse, *Pedicinus obtusus*. Each minichromosome has a coding region and a non-coding region (NCR, in black). Names and transcription orientation of genes are indicated in the coding region; minichromosomes are in alphabetical order of protein-coding and rRNA 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.

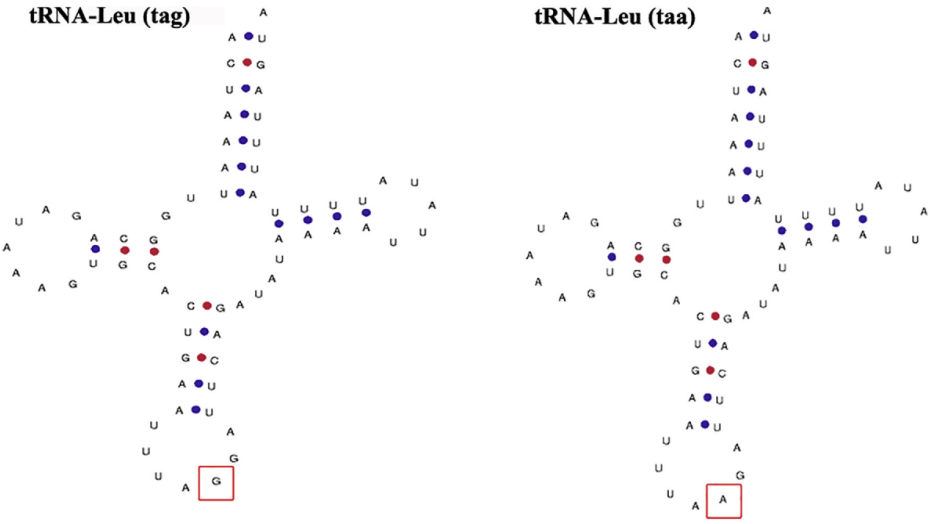


Fig. 4. Inferred secondary structures of the mitochondrial tRNA-Leu (tag) and tRNA-Leu (taa) of *Pedicinus obtusus*. The gene for these two tRNAs, *trnL*$_1$*(tag)* and *trnL*$_2$*(taa)*, have identical sequences except for the third anti-codon position.

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).

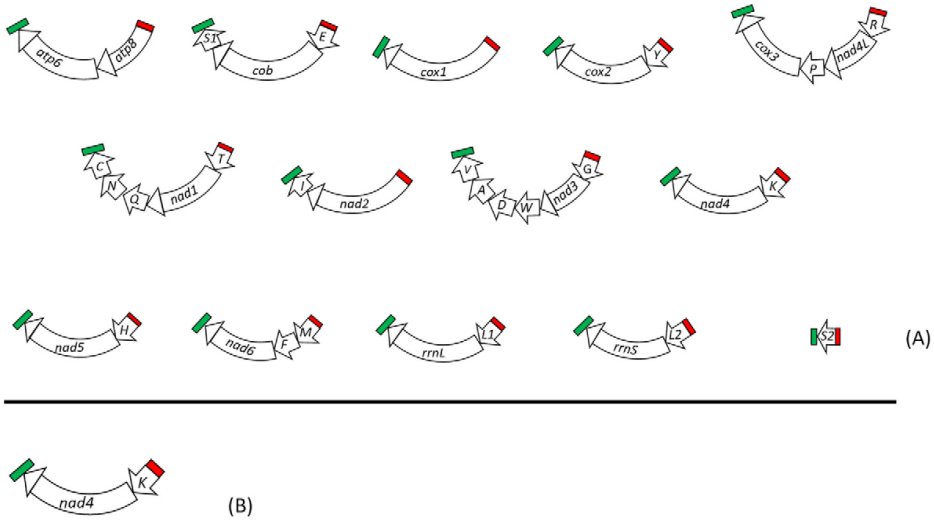


Fig. 5. (A) The coding regions of the 14 minichromosomes and conserved AT-rich region (in red) and conserved GC-rich region (in green) of the colobus louse, *Pedicinus badii*; (B) the coding region of *K-nad4* minichromosomes and conserved AT-rich region (in red) and conserved GC-rich region (in green) 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*. *$trnS_2$* 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 *$trnS_2$*, six other tRNA genes, *trnA*, *trnC*, *trnD*, *trnI*, *$trnL_2$* 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: *$trnL_1$*, *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*,

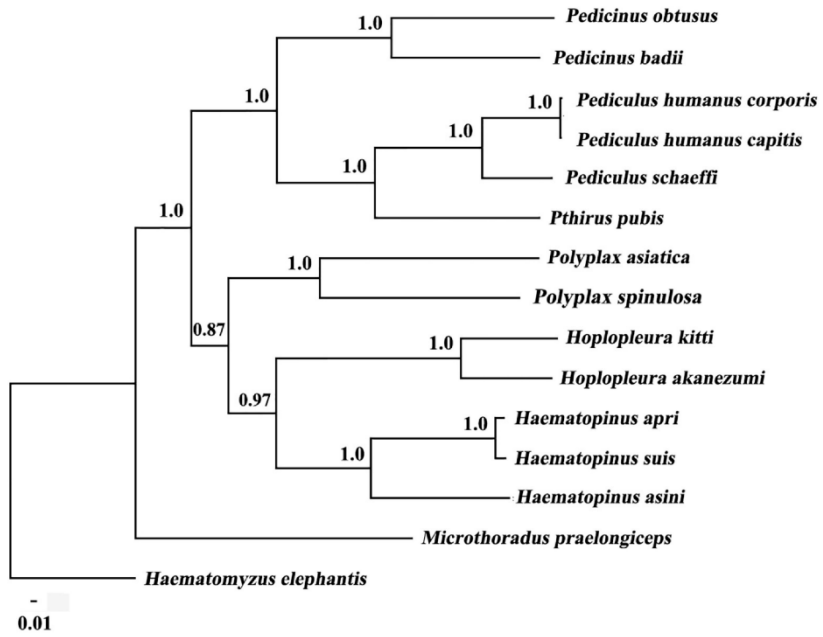


Fig. 6. Phylogenetic relationships among 14 species of sucking lice (Anoplura) inferred from Bayesian analysis of deduced amino acid sequences of nine mi tochondrial proteins. The elephant louse, *Haematomyzus elephantis*, was used as the outgroup. Bayesian posterior probabilities (Bpp) values were indicated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

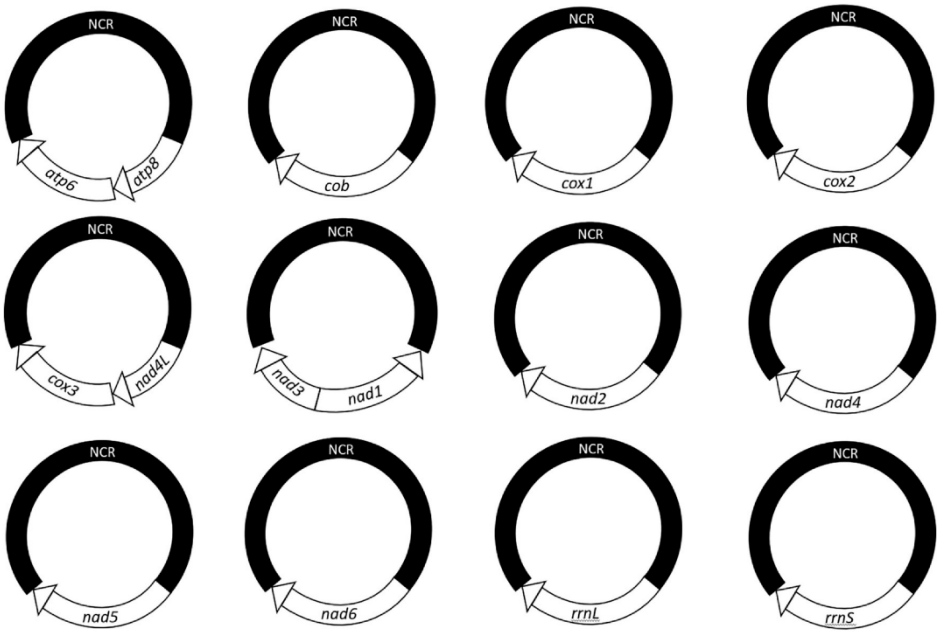


Fig. 7. Inferred ancestral mitochondrial karyotype of higher primate lice (Pedicinidae, Pediculidae and Pthiridae). Names and transcription orientation of genes ar indicated in the coding region; minichromosomes are in alphabetical order of protein-coding and rRNA gene names. See Fig. 3 legend for the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

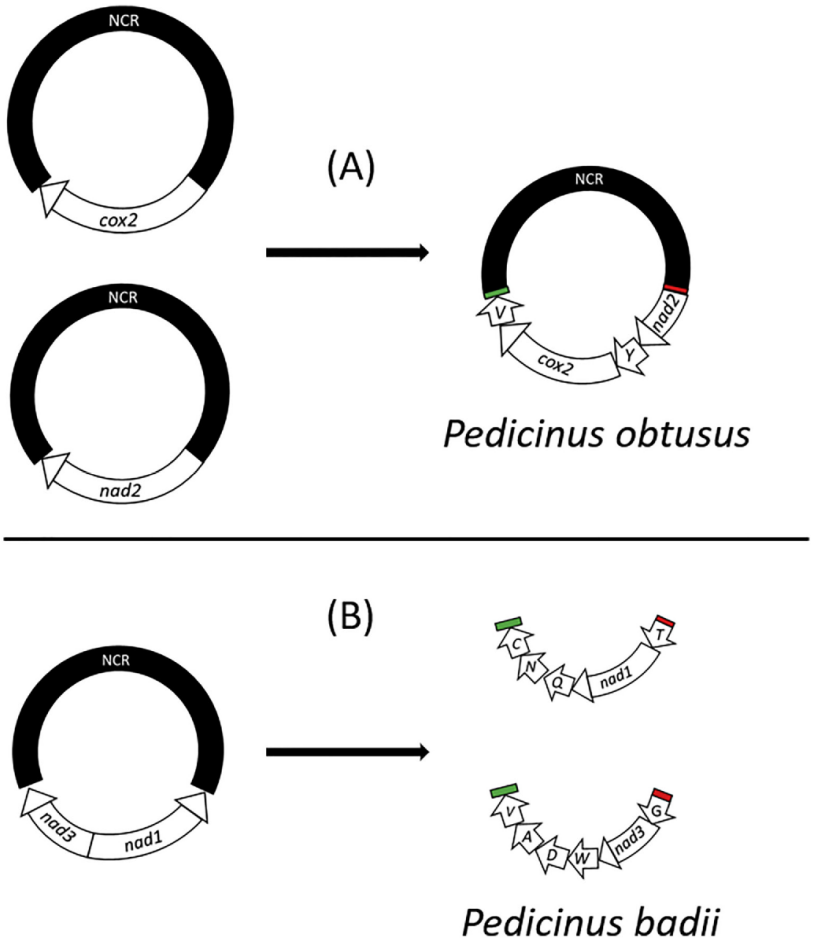


Fig. 8. (A) Two ancestral mitochondrial minichromosomes of higher primate lice (Fig. 7) merged in the macaque louse, *Pedicinus obtusus*; whereas (B) one ancestral mitochondrial minichromosome of higher primate lice split into two in the colobus louse, *Pedicinus badii*. Names and transcription orientation 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 *Haematopinus*, 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.

Appendix II

Manuscript:

Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals

Published in June 2021

Dong et al. BMC Genomics (2021) 22:598
https://doi.org/10.1186/s12864-021-07859-w

BMC Genomics

RESEARCH Open Access

Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals

Wen-Ge Dong[1*], Yalun Dong[2,3], Xian-Guo Guo[1] and Renfu Shao[2,3*]

Abstract

Background: The typical single-chromosome mitochondrial (mt) genome of animals has fragmented into multiple minichromosomes in the lineage Mitodivisia, which contains most of the parasitic lice of eutherian mammals. These parasitic lice differ from each other even among congeneric species in mt karyotype, i.e. the number of minichromosomes, and the gene content and gene order in each minichromosome, which is in stark contrast to the extremely conserved single-chromosome mt genomes across most animal lineages. How fragmented mt genomes evolved is still poorly understood. We use *Polyplax* sucking lice as a model to investigate how tRNA gene translocation shapes the dynamic mt karyotypes.

Results: We sequenced the full mt genome of the Asian grey shrew louse, *Polyplax reclinata*. We then inferred the ancestral mt karyotype for *Polyplax* lice and compared it with the mt karyotypes of the three *Polyplax* species sequenced to date. We found that tRNA genes were entirely responsible for mt karyotype variation among these three species of *Polyplax* lice. Furthermore, tRNA gene translocation observed in *Polyplax* lice was only between different types of minichromosomes and towards the boundaries with the control region. A similar pattern of tRNA gene translocation can also been seen in other sucking lice with fragmented mt genomes.

Conclusions: We conclude that inter-minichromosomal tRNA gene translocation orientated towards the boundaries with the control region is a major contributing factor to the highly dynamic mitochondrial genome organization in the parasitic lice of mammals.

Keywords: tRNA translocation, Genome fragmentation, Mitochondrial karyotype, Parasitic lice

* Correspondence: dongwenge2740@sina.com; rshao@usc.edu.au
[1]Institute of Pathogens and Vectors, Key Laboratory for Preventing and Controlling Plague in Yunnan Province, Dali University, 671000 Dali, China
[2]GeneCology Research Centre, University of the Sunshine Coast, Maroochydore, Queensland, Australia
Full list of author information is available at the end of the article

© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Extensive fragmentation of mitochondrial (mt) genome was discovered first in three species of human lice, in which the single mt chromosome typical of animals evolved into 14 and 20 minichromosomes; each minichromosome has 1–5 genes and is 1.8–4 kb in size [1, 2]. Since then, 26 more species of parasitic lice have been sequenced for mt genomes [3–12]. It appears that mt genome fragmentation occurred at least twice in parasitic lice: once 60–90 million years ago (MYA) in the most recent common ancestor of Mitodivisia - a newly identified clade that contains the vast majority of parasitic lice of eutherian mammals [10, 13–15], and later ~ 25 MYA in the feather lice of the genus *Columbicola* [11, 13].

Unlike the extremely conserved single-chromosome mt genomes in most animal lineages [16] but somehow resembling the nuclear genomes of eukaryotes [17–19], the mt genome organization is highly dynamic in parasitic lice of the Mitodivisia clade [10]. These lice differ from each other, even among closely related species in the same genus, in mt karyotype, i.e. the number of minichromosomes, and the gene content and gene order in each minichromosome [8, 12]. The pig lice, *Haemapinus apri* and *Haemapinus suis*, and the horse louse, *Haemapinus asini*, have nine minichromosomes [3, 6], whereas the human head louse, *Pediculus humanus capitis*, and the human body louse, *Pediculus humanus humanus*, have 20 minichromosomes [1, 2]. Between these extremes, the elephant louse, *Haematomyzus elephantis*, has 10 minichromosomes (four genes not identified) [8]. The rodent lice, *Polyplax asiatica*, *Polyplax spinulosa*, *Hoplopleura akanezumi* (nine genes not identified) and *Hoplopleura kitti* (three genes not identified), have 11 minichromosomes [4, 5]. The guanaco louse, *Microthoracius praelongiceps*, the sheep louse, *Bovicola ovis*, the cattle louse, *Bovicola bovis*, and the macaque louse, *Pedicinus obtusus*, have 12 minichromosomes [9, 10, 12]. The goat louse, *Bovicola caprae*, and the dog louse, *Trichodectes canis*, have 13 minichromosomes [10]. The human pubic louse, *Pthirus pubis* (one gene not identified), and the colobus louse, *Pedicinus badii*, have 14 minichromosomes [2, 12]. Each minichromosome has 1 to 8 genes; the arrangement of genes in a minichromosome varies from species to species even within a genus; the only exceptions are: (1) the human head louse and the human body louse in the genus *Pediculus*, which diverged 83,000-170,000 years ago at the origin of clothing [1, 2, 20–22]; and (2) the domestic pig louse and the wild pig louse in the genus *Haemapinus*, which diverged ~ 10,000 years ago when pigs were domesticated [3, 23].

Why is the fragmented mt genome organization so variable in these parasitic lice? A few comparative studies to date have suggested four models or mechanisms. First, Shao et al., (2012) [2] showed that point mutation at the third anti-codon position or homologous recombination between tRNA genes accounted for the swap of identity and location between $trnL_1$ and $trnL_2$ and between *trnR* and *trnG* in the human lice. Dong et al., (2014a) [4] confirmed that this mechanism accounted for the swap of identity and location between $trnL_1$ and $trnL_2$ in the spiny rat louse, *Polyplax spinulosa*. Second, Shao et al., (2012) [2] showed that one minichromosome could split into two via gene degeneration followed by deletion, based on analysis of pseudo gene sequences of the human pubic louse. Third, Song et al., (2014) [6] showed that inter-minichromosomal recombination accounted for gene translocation in the horse louse, *Haematopinus asini*, and mt karyotype variation among *Haematopinus* lice. Fourth, Shao et al., (2017) [9] showed that both split and merger of minichromosomes had occurred and contributed to the variation in mt karyotypes in sucking lice.

In the present study, we used *Polyplax* sucking lice, which parasitize rodents and shrews, as a model to investigate the role of tRNA gene translocation in shaping the dynamic mt karyotypes. We show that frequent directional tRNA gene translocation between different types of minichromosomes is a major contributor to the highly dynamic mt genome organization of the parasitic lice of mammals.

Materials and methods

Louse collection, DNA extraction, mitochondrial genome amplification and sequencing

Specimens of *Polyplax reclinata* (family Polyplacidae) were collected in Yunnan Province, China, from the Asian grey shrew, *Crocidura attenuata* (family Soricidae). *Polyplax reclinata* was identified according to Chin (1999) [24]; voucher specimens (# 364) were kept at the Institute of Pathogens and Vectors, Dali University (Additional file 1). Sucking lice collected on the body surface of *Crocidura attenuata* were preserved in 95 % ethanol at − 20 °C prior to DNA extraction. Capture of shrews was approved by health authorities in Yunnan Province, China. Animal capture protocols and procedures were approved by the animal ethics committees at Dali University. Genomic DNA was extracted from individual lice with DNeasy Blood & Tissue kit (QIAGEN). Two pairs of primers, 12SA–12SB [25] and 16SF–Lx16SR [9], were used to amplify fragments of mt genes *rrnS* (375 bp) and *rrnL* (360 bp) (Additional file 2). These primers target conserved gene sequence motifs among arthropods. The two gene fragments were sequenced directly using Sanger method at the Thermo Fisher Scientific Genome Sequencing Facility (Guangzhou). Two pairs of specific primers for *P.*

reclinata, 12S364F-12S364R and 16S364F-16S364R, were designed from *rrnS* and *rrnL* sequences obtained (Additional file 2). The two specific primers in each pair go outwards with 15 and 21 bp in between, respectively. PCRs with these specific primers amplified the near full-length *rrnS* and *rrnL* minichromosomes of *P. reclinata* (Additional file 3A). The amplicons from *rrnS* and *rrnL* minichromosomes, 2.3 and 2.6 kb in size respectively, were sequenced using Sanger method to obtain the sequences of non-coding regions. Another pair of primers specific to *P. reclinata*, 364F–364R (Additional file 2), was designed from conserved sequences of the non-coding regions that flank the coding regions of the *rrnS* and *rrnL* minichromosomes. The PCR with the primer pair 364F–364R produced a mixture of amplicons ranging from 1.1 to 1.8 kb in size, expected from the coding regions of the entire set of mt minichromosomes of *P. reclinata* (Additional file 3A). These amplicons were sequenced from both ends (i.e. paired-end) with Illumina Miseq platform: insert size 400 bp, read length 250 bp. The PCR strategy used in this study was developed from our observations in previous studies that each mt minichromosome has a distinct coding region but also a well-conserved non-coding region [1–6, 9].

Takara Ex Taq was used in the initial short PCRs with the following cycling conditions: 94 °C for 1 min; 35 cycles of 98 °C for 10 s, 45 °C for 30 s, 72 °C for 1 min; and a final extension of 72 °C for 2 min. Takara LA *Taq* was used in the long PCRs with the cycling conditions: 94 °C for 1 min; 35 cycles of 98 °C for 10 s, 55–65 °C (depending on primers) for 40 s, 68 °C for 4 min; and a final extension of 72 °C for 8 min. Negative controls were executed with each PCR experiment. PCR amplicons were checked by agarose gel (1 %) electrophoresis; the sizes of amplicons were estimated by comparing with DNA markers. PCR amplicons were purified with Wizard SV Gel/PCR clean-up system (Promega). Sanger sequencing was done at the Thermo Fisher Scientific Genome Sequencing Facility (Guangzhou). Illumina sequencing was done at the Majorbio Genome Sequencing Facility (Shanghai).

Assembly of sequence reads, gene identification and minichromosome verification

Illumina sequence reads (250 bp each) were assembled into contigs with Geneious 11.1.5 [26]. The assembly parameters were minimum overlap 150 bp and minimum identity 98 %. tRNA genes were identified using tRNAscan-SE [27] and ARWEN [28]. Protein-coding genes and rRNA genes were identified by BLAST searches of GenBank [29, 30]. Sequences were aligned with Clustal X [31]. The size and circular organization of each mt minichromosome of *P. reclinata* were verified by PCR using outbound specific primers designed from the coding region of each minichromosome (Additional file 2). The forward primer and reverse primer in each pair were next to each other with a small gap in between. PCRs with these primers amplified each circular minichromosome in full or near full length (Additional file 3B); these amplicons were also sequenced with Illumina Miseq platform to obtain the full-length sequence of the non-coding region of each minichromosome. PCR set-up, cycling conditions, agarose gel electrophoresis and size measurement were the same as described above. Negative controls were run for each PCR test. The annotated mt genome sequence of *P. reclinata* was deposited in GenBank (accession numbers MW291451-MW291461).

Phylogenetic analyses

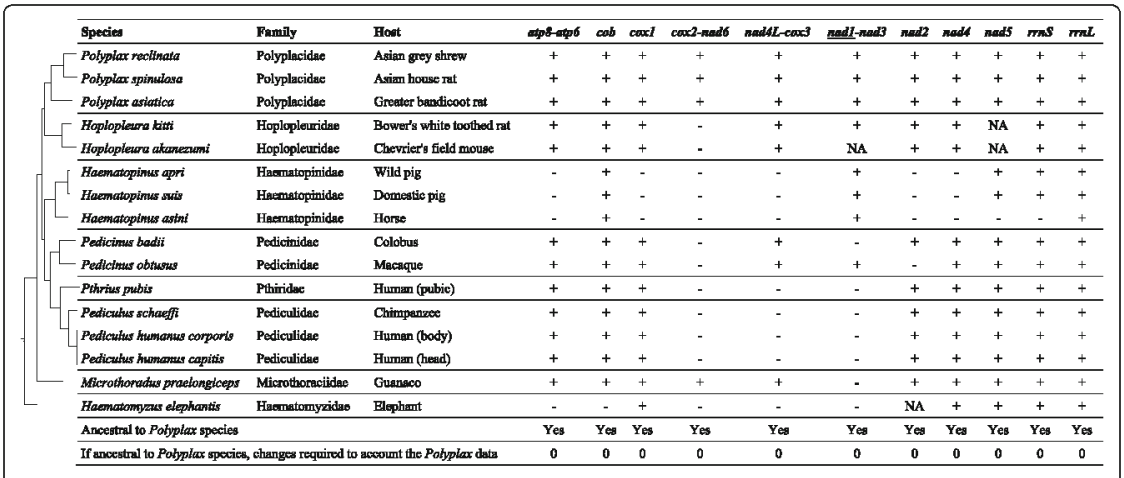
We retrieved the mt genome sequences of all of the 14 species of sucking lice available in GenBank and the elephant louse (Table 1) and combined these sequences with that of *P. reclinata* generated in the present study. The elephant louse was used as the outgroup because it was in the suborder Rhynchophthirina, which was most closely related to sucking lice, the suborder Anoplura [10]. The sequences of eight mt protein-coding genes (*atp6*, *atp8*, *cox1*, *cox2*, *cox3*, *cob*, *nad4L*, *nad6*) and two rRNA genes (*rrnS* and *rrnL*) of these lice were used in phylogenetic analysis. Other mt protein-coding genes (*nad1*, *nad2*, *nad3*, *nad4*, and *nad5*) were excluded from our analysis because their sequences were not available for all of the species above. Protein-coding gene sequences were aligned based on amino acid sequences using the MAFFT algorithm implemented in TranslatorX online platform [32]. rRNA genes were aligned using the MAFFT v7.0 online server with G-INS-i strategy [33]. Individual gene alignments were concatenated after removing poorly aligned sites using GBlocks v0.91b [34]. A concatenated alignment, PCGRNA matrix, was used in subsequent analyses; this matrix combines the sequences of the eight protein-coding genes and the two rRNA genes (4,631 bp in total). The matrix was analyzed using maximum likelihood (ML) and Bayesian methods with MEGA-X [35] and 2) Bayesian inference method (BI) with MrBayes 3.2.6 [36], respectively. For ML analysis, the number of bootstrap replicates was 500; the substitution model selected was *Tamura-Nei* and the heuristic method selected was *Nearest-Neighbor-Interchange (NNI)*. For Bayesian analyses, four independent Markov chains were run for 25,000 MCMC generations, sampling a tree every 10 generations. Bayesian analysis was run until the average standard deviation of split frequencies was below 0.001. The initial 10,000 trees of each MCMC run were discarded as burn-in. ML trees were directly generated in MEGA-X and the Bayesian trees were drawn with Figtree v1.4.3.

Table 1 Species of parasitic lice included in the phylogenetic analyses in this study

Species	Family	Suborder	Order	GenBank accession number	References
Polyplax reclinata	Polyplacidae	Anoplura	Phthiraptera	MW291451-MW291461	Present study
Hoplopleura akanezumi	Hoplopleuridae	Anoplura	Phthiraptera	KJ648922–32	Dong et al., 2014a
Hoplopleura kitti	Hoplopleuridae	Anoplura	Phthiraptera	KJ648933–43	Dong et al., 2014a
Polyplax spinulosa	Polyplacidae	Anoplura	Phthiraptera	KF647762–72	Dong et al., 2014b
Polyplax asiatica	Polyplacidae	Anoplura	Phthiraptera	KF647751–61	Dong et al., 2014b
Haematopinus asini	Haematopinidae	Anoplura	Phthiraptera	KF939318, KF939322, KF939324, KF939326, KJ434034–38	Song et al., 2014
Haematopinus apri	Haematopinidae	Anoplura	Phthiraptera	KC814611–19	Jiang et al., 2013
Haematopinus suis	Haematopinidae	Anoplura	Phthiraptera	KC814602–10	Jiang et al., 2013
Pthirus pubis	Pthiridae	Anoplura	Phthiraptera	EU219988–95, HM241895–98	Shao et al., 2012
Pediculus schaeffi	Pediculidae	Anoplura	Phthiraptera	KC241882–97, KR706168–69	Herd et al., 2015
Pediculus capitis	Pediculidae	Anoplura	Phthiraptera	JX080388–407	Shao et al., 2012
Pediculus humanus	Pediculidae	Anoplura	Phthiraptera	FJ499473–90	Shao et al., 2009
Pedicinus obtusus	Pedicinidae	Anoplura	Phthiraptera	MT792495-506	Fu et al., 2020
Pedicinus badii	Pedicinidae	Anoplura	Phthiraptera	MT721726-39	Fu et al., 2020
Microthoradus praelongiceps	Microthoraciidae	Anoplura	Phthiraptera	KX090378–89	Shao et al., 2017
Haematomyzus elephantis	Haematomyzidae	Rhyncophthirina	Phthiraptera	KF933032–41	Shao et al., 2015

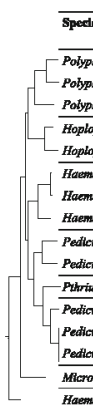
Inference of the ancestral mitochondrial karyotype of *Polyplax* lice

We used a parsimony method described in Shao et al., (2017) [9] to infer the ancestral mt karyotype of *Polyplax* lice based on information from the complete mt genomes of three *Polyplax* species: *P. reclinata* (present study), *P. asiatica* [4] and *P. spinulosa* [4]. We mapped characters of mt karyotypes on the phylogenetic tree of sucking lice inferred from mt genome sequences to identify shared characters. We inferred an mt minichromosome character to be ancestral to *Polyplax*: (1) if it was present in at least one *Polyplax* species and also in



Species	Family	Host	atp8-atp6	cob	cox1	cox2-nad6	nad4L-cox3	nad1-nad3	nad2	nad4	nad5	rrnS	rrnL
Polyplax reclinata	Polyplacidae	Asian grey shrew	+	+	+	+	+	+	+	+	+	+	+
Polyplax spinulosa	Polyplacidae	Asian house rat	+	+	+	+	+	+	+	+	+	+	+
Polyplax asiatica	Polyplacidae	Greater bandicoot rat	+	+	+	+	+	+	+	+	+	+	+
Hoplopleura kitti	Hoplopleuridae	Bower's white toothed rat	+	+	+	-	+	+	+	+	NA	+	+
Hoplopleura akanezumi	Hoplopleuridae	Chevrier's field mouse	+	+	+	-	+	NA	+	+	NA	+	+
Haematopinus apri	Haematopinidae	Wild pig	-	+	-	-	-	+	-	-	+	+	+
Haematopinus suis	Haematopinidae	Domestic pig	-	+	-	-	-	+	-	-	+	+	+
Haematopinus asini	Haematopinidae	Horse	-	+	-	-	-	+	-	-	-	-	+
Pedicinus badii	Pedicinidae	Colobus	+	+	+	-	+	-	+	+	+	+	+
Pedicinus obtusus	Pedicinidae	Macaque	+	+	+	-	+	+	-	+	+	+	+
Pthirus pubis	Pthiridae	Human (pubic)	+	+	+	-	-	-	+	+	+	+	+
Pediculus schaeffi	Pediculidae	Chimpanzee	+	+	+	-	-	-	+	+	+	+	+
Pediculus humanus corporis	Pediculidae	Human (body)	+	+	+	-	-	-	+	+	+	+	+
Pediculus humanus capitis	Pediculidae	Human (head)	+	+	+	-	-	-	+	+	+	+	+
Microthoradus praelongiceps	Microthoraciidae	Guanaco	+	+	+	+	+	-	+	+	+	+	+
Haematomyzus elephantis	Haematomyzidae	Elephant	-	-	+	-	-	-	NA	+	+	+	+
Ancestral to *Polyplax* species			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
If ancestral to *Polyplax* species, changes required to account the *Polyplax* data			0	0	0	0	0	0	0	0	0	0	0

Fig. 1 Inferring the ancestral mitochondrial karyotype of *Polyplax* lice (suborder Anoplura). tRNA genes are excluded. Plus symbol (+) is for presence, minus (-) for absence, and NA for information not available. Underlined *nad1* has an opposite orientation of transcription to other genes. The phylogenetic tree was inferred with mitochondrial genome sequences in the current study



Species	*cox3-A*	*nad6-A*	*rrnS-C*	*D-Y-cox2*	*E-cob-I*	*nad5-F*	*G-nad3-W*	*H-nad5*	*K-nad4*	*M-L1-rrnL*	*cox1-L3*	*Q-nad2-N*	*nad4L-P*	*R-nad4L*	*S1-S2-rrnS*	*S1-S2-nad1*	*nad1-T*	*T-D-Y-cox2*	*rrnL-V*
Polyplax reclinata	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+
Polyplax spinulosa	-	+	+	+	+	+	+	+	+	-	-	+	+	+	+	-	-	+	+
Polyplax asiatica	+	-	+	+	+	-	+	+	+	+	+	+	-	+	-	+	+	-	+
Hoplopleura kitti	+	-	-	+	-	NA	-	NA	+	+	-	-	+	+	-	-	-	-	+
Hoplopleura akanezumi	+	-	-	+	-	NA	NA	NA	+	-	-	-	+	+	-	-	-	-	+
Haematopinus apri	+	-	+	+	-	+	+	+	+	-	+	-	-	+	-	-	+	-	-
Haematopinus suis	+	-	+	+	-	+	+	+	+	-	+	-	-	+	-	-	+	-	-
Haematopinus asini	+	-	+	+	-	+	+	+	+	-	+	-	-	+	-	-	+	-	-
Pediculus badii	-	-	-	-	-	-	+	+	+	-	-	-	+	+	-	-	-	-	-
Pediculus obtusus	-	-	-	-	-	-	+	+	+	-	-	-	+	+	-	-	-	-	-
Pthirus pubis	+	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-
Pediculus schaeffi	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
Pediculus humanus corporis	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
Pediculus humanus capitis	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
Microthoracius praelongiceps	+	-	+	-	-	+	+	+	+	+	+	-	+	+	-	-	+	-	+
Haematomyzus elephantis	-	-	-	-	-	-	-	+	+	-	-	-	-	+	-	-	+	-	+
Ancestral to *Polyplax* species	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Possible	Possible	Yes	No	Yes
If ancestral to *Polyplax* species, changes required to account the *Polyplax* data	1	NA	0	0	0	1	0	0	0	1	1	0	1	0	1	1	1	NA	0

Fig. 2 Inferring the location of tRNA genes relative to protein-coding and rRNA genes in the ancestral mitochondrial karyotype of *Polyplax* lice. tRNA genes are indicated with their single-letter abbreviations. Plus symbol (+) is for presence, minus (-) for absence, and NA for information not available. Underlined genes have opposite orientation of transcription to other genes. The phylogenetic tree was inferred with mitochondrial genome sequences in the current study

one or more non-*Polyplax* species of sucking lice; or (2) if it was present in all of the three *Polyplax* species. We considered tRNA genes separately from protein-coding and ribosome RNA genes because protein-coding and rRNA genes were much more stable than tRNA genes [9] (Figs. 1 and 2). For each inferred ancestral minichromosome character of *Polyplax* lice, we counted the changes needed to explain the observed mt karyotypes of *Polyplax* lice. If two or more minichromosome characters conflicted with one another, the character with the least required changes was inferred to be the ancestral. If two or more characters are equally parsimonious, all characters are inferred to be equally likely to be ancestral (Figs. 1 and 2).

Results and discussion

Mitochondrial genome of the Asian grey shrew louse, *Polyplax reclinata*

We obtained the complete mt genome of *P. reclinata* by assembling the 53,915,414 clean Illumina sequence reads (paired-end, 250 bp each) generated from the minichromosome amplicons (Table 2). All of the 37 mt genes

Table 2 The mitochondrial minichromosomes of the Asian grey shrew louse, *Polyplax reclinata*, identified by Illumina sequencing

Minichromosome	Minichromosome size (bp)	Size of coding region (bp)	Size of non-coding region (bp)	Number of Illumina sequence reads	Mean coverage	GenBank accession number
atp8-atp6	2,425	847	1,578	5,836,284	333,785	MW291451
E-cob-I	2,620	1,223	1,397	7,275,348	387,778	MW291453
cox1-L2	2,816	1,580	1,236	4,012,234	204,899	MW291452
T-D-Y-cox2-nad6-A	2,473	1,399	1,074	3,691,415	212,776	MW291461
R-nad4L-P-cox3	2,465	1,191	1,274	6,293,311	365,198	MW291459
nad1-G-nad3-W [a]	2,693	1,450	1,243	9,485,576	351,940	MW291457
Q-nad2-N	2,558	1,124	1,434	5,969,577	330,770	MW291458
K-nad4	2,680	1,340	1,340	4,248,148	218,975	MW291455
H-nad5-F	2,869	1,809	1,060	4,258,208	209,453	MW291454
S1-S2-rrnS-C	2,332	900	1,432	441,782	42,400	MW291460
M-L1-rrnL-V	2,559	1,240	1,319	2,403,531	213,096	MW291456
Total	28,490	14,103	14,387	53,915,414	2,871,070	

Note: [a]underlined *nad1* is opposite to other genes in the orientation of transcription

typical of bilateral animals were identified in *P. reclinata*; these genes were on 11 circular minichromosomes (Fig. 3). The mt minichromosomes of *P. reclinata* range from 2,332 to 2,869 bp in size (Table 2); each minichromosome has a coding region and a non-coding region (Fig. 3; Table 2). The coding region of each minichromosome has 2–6 genes and ranges from 847 bp for *atp8-atp6* minichromosome to 1,809 bp for *H-nad5-F* minichromosome (minichromosomes named after the genes they contain). Seven of the 11 minichromosomes of *P. reclinata* contain a single protein-coding or rRNA gene each; the other four minichromosomes contain two protein-coding genes each. There are 1–4 tRNA genes in each minichromosome except *atp8-atp6* minichromosome, which has no tRNA genes (Fig. 3). Each of the 37 mt genes in *P. reclinata* is found only in one minichromosome; all genes have the same orientation of transcription relative to the non-coding region except for *nad1*, which has an opposite orientation to other genes (Fig. 3).

We sequenced the non-coding regions of all of the 11 minichromosomes of *P. reclinata* in full length. The non-coding regions range from 1,060 to 1,578 bp (Table 2); the size variation is due to the additional sequences in the middle of the non-coding regions of *M-L_1-rrnL-V*, *S_1-S_2-rrnS-C*, *K-nad4*, *atp8-atp6*, *Q-nad2-N* and *E-cob-I* minichromosomes (Additional file 4). If the additional sequences are excluded, the non-coding regions of all 11 minichromosomes have high sequence similarity to each other. Overall, the non-coding regions have high C and G content in one end but high A and T content in the other end. Indeed, a CG-rich motif (54 bp, 59 % C and G) was found immediately downstream from the 3'-end of the coding regions and an AT-rich motif (113 bp, 68 % A and T) was found upstream from the 5'-end of the coding regions in all of the minichromosomes (Additional file 4).

Mitochondrial karyotype variation among *Polyplax* species

Polyplax reclinata differs in mt karyotype from the two other *Polyplax* species reported previously [4], although these three species all have 11 minichromosomes. *Polyplax reclinata* shares nine minichromosomes with *P. spinulosa* (louse of the Asian house rat) [4] but differs in the distribution of *$trnL_1$* and *$trnL_2$* between the other two minichromosomes (Table 3). In *P. spinulosa*, *$trnL_1$* is downstream from *cox1* and *$trnL_2$* is upstream from *rrnL* [4]; in *P. reclinata*, however, it is the opposite (Fig. 3; Table 3). *Polyplax reclinata* shares only four minichromosomes with *P. asiatica* (louse of the greater bandicoot rat) [4] and differs in the distribution of six tRNA genes among the other seven minichromosomes: *trnA*, *trnF*, *trnP*, *$trnS_1$*, *$trnS_2$* and *trnT* (Table 3). If tRNA genes are excluded, the three *Polyplax* species are identical to one another in the distribution of protein-coding genes and rRNA genes among the 11 minichromosomes (Table 3; Fig. 3).

Phylogeny of sucking lice inferred from mt genome sequences

To assist the inference of the ancestral mt karyotype of *Polyplax* lice, we inferred the phylogeny of sucking lice using mt genome sequences of 15 species of sucking lice of seven genera including *Polyplax reclinata*, which was sequenced in the present study. The elephant louse, which was in the suborder Rhynchophthirina, was used as the outgroup (Table 1). Rhynchophthirina is the sister

Table 3 Comparison of gene content and gene arrangement in the mitochondrial minichromosomes of three *Polyplax* species and their most recent common ancestor (MRCA)

Inferred most recent common ancestor of *Polyplax*	*Polyplax reclinata*	*Polyplax spinulosa*	*Polyplax asiatica*
atp8-atp6	*atp8-atp6*	*atp8-atp6*	*atp8-atp6*
E-cob-I	*E-cob-I*	*E-cob-I*	*E-cob-I*
cox1-$L_{2(taa)}$	*cox1-$L_{2(taa)}$*	*cox1-**$L_{1(tag)}$***	*cox1-$L_{2(taa)}$*
D-Y-cox2-nad6	***T**-D-Y-cox2-nad6-**A***	***T**-D-Y-cox2-nad6-**A***	*D-Y-cox2-nad6*
R-nad4L-P-cox3-A	*R-nad4L-P-cox3*	*R-nad4L-P-cox3*	*R-nad4L-cox3-A*
$S_{1(tct)}$-$S_{2(tga)}$-nad1-T-G-nad3-W	*nad1-G-nad3-W*	*nad1-G-nad3-W*	***$S_{1(tct)}$-$S_{2(tga)}$**-nad1-T-G-nad3-W*
Q-nad2-N	*Q-nad2-N*	*Q-nad2-N*	*Q-nad2-N-**P***
K-nad4	*K-nad4*	*K-nad4*	*K-nad4-**F***
H-nad5-F	*H-nad5-F*	*H-nad5-F*	*H-nad5*
$S_{1(tct)}$-$S_{2(tga)}$-rrnS-C	***$S_{1(tct)}$-$S_{2(tga)}$**-rrnS-C*	***$S_{1(tct)}$-$S_{2(tga)}$**-rrnS-C*	*rrnS-C*
M-$L_{1(tag)}$-rrnL-V	*M-$L_{1(tag)}$-rrnL-V*	*M-**$L_{2(taa)}$**-rrnL-V*	*M-$L_{1(tag)}$-rrnL-V*

Notes: The position of *$S_{1(tct)}$-$S_{2(tga)}$* gene cluster in the ancestral mitochondrial karyotype is either upstream from *rrnS* or, equally parsimoniously, downstream from *nad1*. The eight tRNA genes that have changed their positions and/or orientation in the three *Polyplax* species since their divergence from their MRCA are in bold

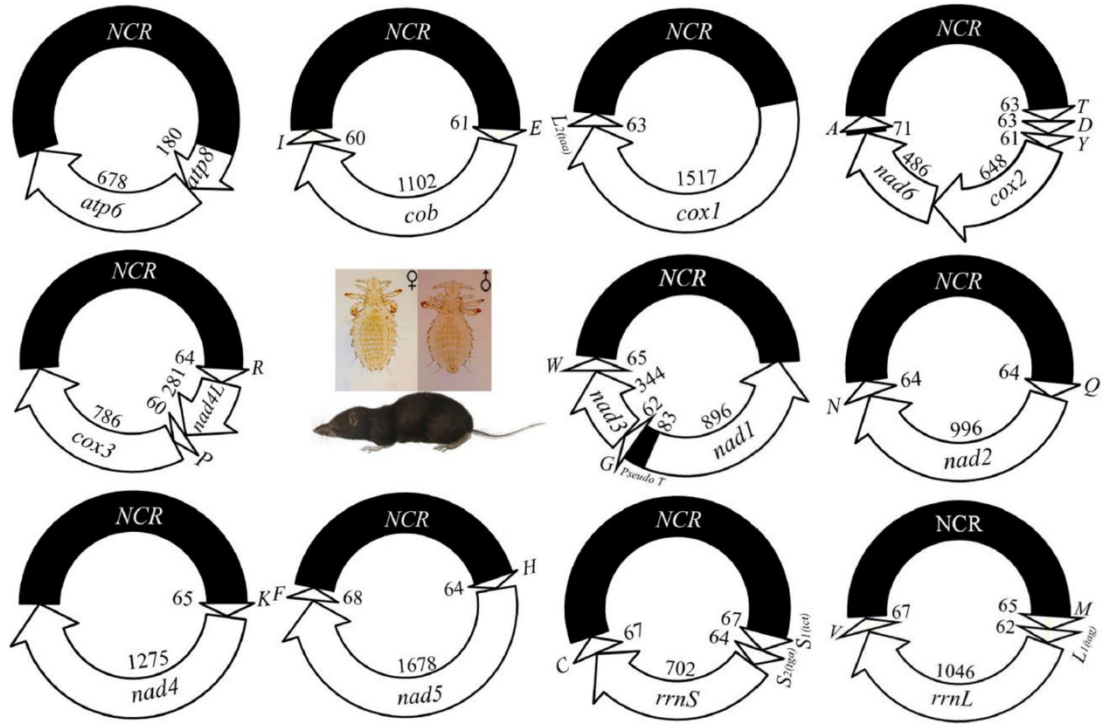


Fig. 3 The mitochondrial genome of the Asian grey shrew louse, *Polyplax reclinata*. Protein-coding genes are: *atp6* and *atp8* (for ATP synthase subunits 6 and 8), *cox1-3* (for cytochrome c oxidase subunits 1–3), *cob* (for cytochrome b), *nad1-6* and *nad4L* (for NADH dehydrogenase subunits 1–6 and 4L). rRNA genes are: *rrnS* and *rrnL* (for small and large subunits of ribosomal RNA). tRNA genes are indicated with their single-letter abbreviations. Arrow indicates transcription orientation; numbers inside inner circles indicate gene length (bp); non-coding regions (*NCR*) are shaded in black

suborder to the sucking lice (suborder Anoplura) [10]. We obtained two trees from PCGRNA matrix (4,631 bp of protein-coding and rRNA gene sequences) with ML and Bayesian methods. These two trees have similar topologies and consistently support the monophyly of the genus *Polyplax* (Fig. 4): bootstrap support value (BSV) = 100 %, posterior probability (PP) = 1. Within the *Polyplax*, *P. reclinata* is more closely related to *P. spinulosa* than to *P. asiatica* (BSV = 100 %, PP = 1), despite the fact that the latter two species are found mostly on rodents whereas *P. reclinata* is found on shrews [37, 38]. The other four genera, *Hoplopleura*, *Pediculus*, *Pedicinus* and *Haematopinus*, also had multiple species from each genus included in our analysis; each of these genera was monophyletic with strong support (BSV = 100 %, PP = 1) (Fig. 4). The 15 species of sucking lice were divided into three clades: *Polyplax* and *Hoplopleura* forming a clade (BSV = 76.6 %: PP = 0.998); *Pediculus*, *Pedicinus*, *Pthirus* and *Haematopinus* in another clade (BSV = 69.8 %; PP = 0.983); and *Microthoracius praelongiceps* alone as the earliest branch (BSV = 85.2 %; PP = 1) (Fig. 4). Each of the genera included in our analysis also represents a family (Table 1). The family level relationships revealed by our analysis are consistent with that proposed by Kim (1988) based on morphological characters [39] and partially consistent with that by Light et al. (2010), Smith et al. (2011) and Johnson et al. (2018) based on gene sequences [13–15].

Inferred ancestral mitochondrial karyotype of *Polyplax* lice

We inferred the ancestral mt karyotype of *Polyplax* lice based on the data available for the three *Polyplax* species and the other 12 species of sucking lice (Table 1; Fig. 5). We mapped mt karyotype characters on the phylogeny inferred from the mt genome sequences of these lice to identify shared characters (Figs. 1 and 2). The inferred ancestral mt karyotype of *Polyplax* lice by parsimony consists of 11 minichromosomes; each minichromosome has a coding region with

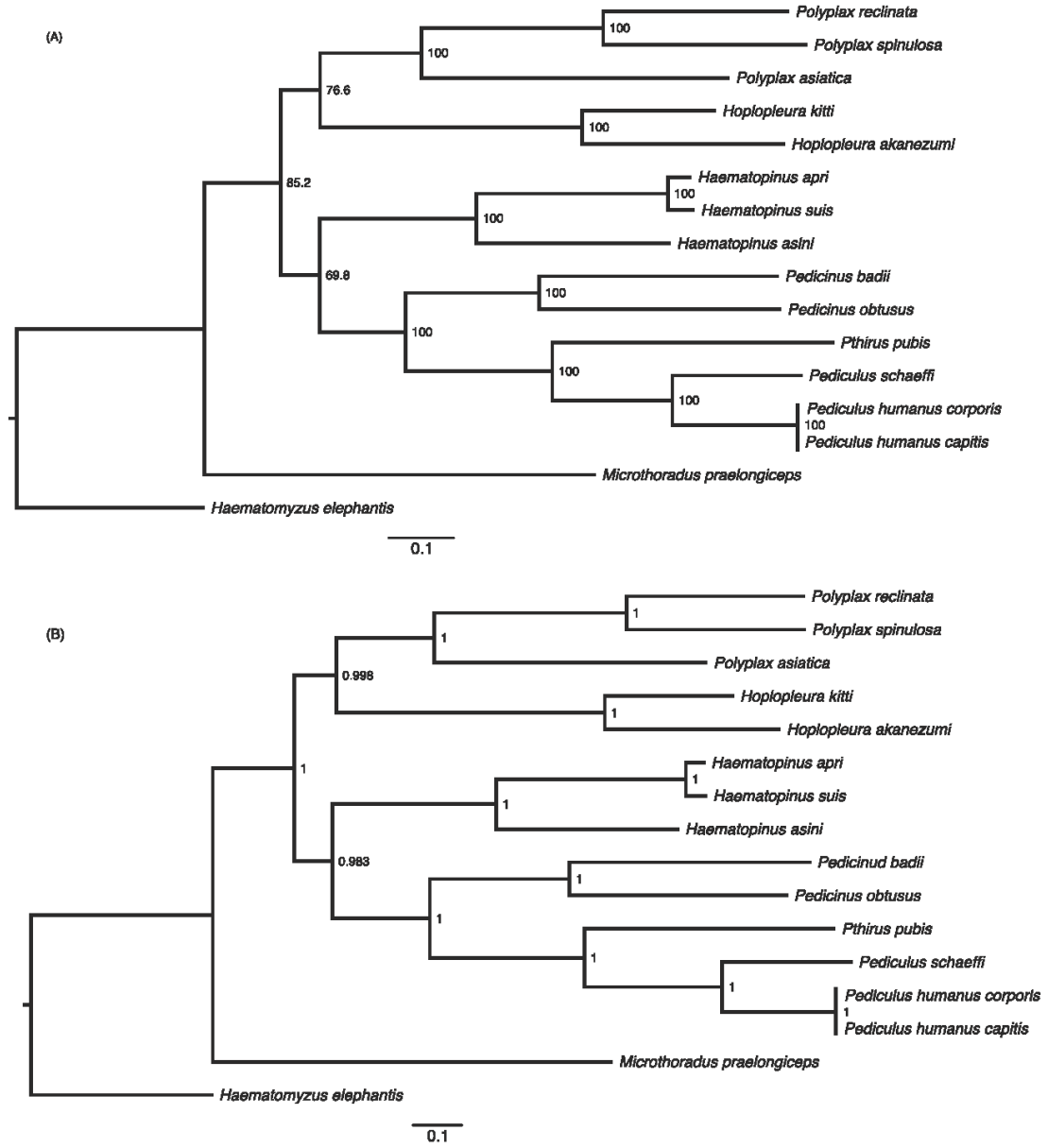


Fig. 4 Phylogenetic relationships among 15 species of sucking lice (suborder Anoplura) inferred with combined sequences of eight mitochondrial protein-coding genes (*atp6*, *atp8*, *cox1*, *cox2*, *cox3*, *cob*, *nad4L*, *nad6*) and two rRNA genes (*rrnS* and *rrnL*) with maximum likelihood method (**A**) and Bayesian method (**B**). The trees were rooted with the elephant louse, *Haematomyzus elephantis* (suborder Rhyncophthirina). Bootstrap support values (out of 100) and Bayesian posterior probability values (out of 1) were indicated near internal nodes

2–7 genes and a non-coding region (Fig. 5). Seven of the 11 minichromosomes have a single protein-coding or rRNA gene each; the other four minichromosomes have two protein-coding genes each: (1) *atp6* and *atp8*, (2) *cox2* and *nad6*, (3) *cox3* and *nad4L*, and (4) *nad3* and *nad1* (Fig. 5). The position of all genes in the ancestral mt karyotype can be inferred without conflict except the position of a cluster of two tRNA genes, S_1-S_2, which is either upstream from *rrnS* or, equally parsimoniously, downstream from *nad1* (Figs. 2 and 5).

Mitochondrial tRNA gene translocation in *Polyplax* lice

When compared with the inferred ancestral mt karyotype of *Polyplax* lice (Fig. 5), all three *Polyplax* species sequenced to date retained 11 minichromosomes and the same distribution pattern of protein-coding and rRNA genes among the minichromosomes (Table 3). Therefore, no split nor merger of minichromosomes have occurred in any of these *Polyplax* species since their most recent common ancestor (MRCA) (Fig. 4). However, these three *Polyplax* species are different from one another and from their MRCA in the position of tRNA genes. In total, eight tRNA genes have changed their positions in the three *Polyplax* lice since they diverged from their MRCA: *trnA*, *trnF*, $trnL_1$, $trnL_2$, *trnP*, $trnS_1$, $trnS_2$ and *trnT* (Table 3). *trnA* and *trnT* were likely translocated (*trnT* also inverted) in the MRCA of *P. reclinata* and *P. spinulosa* (Fig. 4) and thus are shared by these two species (Table 3). A pseudo *trnT* can be found in *P. reclinata* (Fig. 3) and *P. spinulosa* [4], indicating the translocation of *trnT* occurred relatively recently. $trnL_1$ and $trnL_2$ swapped their positions in *P. spinulosa* [4], most likely caused by point mutations at the third anti-codon position (or by homologous recombination between $trnL_1$ and $trnL_2$) – a mechanism suggested by previous studies [2, 9]. *trnF* and *trnP* were translocated in *P. asiatica* (Table 3); as a cluster, $trnS_1$-$trnS_2$ was translocated either in *P. asiatica* or equally likely in the MRCA of *P. reclinata* and *P. spinulosa* (Table 3; Fig. 5).

tRNA gene translocation is frequent and directional towards the boundaries with control region in parasitic lice with fragmented mitochondrial genomes

We noted in the *Polyplax* lice above that mt tRNA gene translocation was closely associated with the control region, i.e. the non-coding region (NCR). With no exception, the destination of all of the translocated tRNA genes (i.e. *trnA*, *trnF*, $trnS_1$-$trnS_2$, *trnP* and *trnT* excluding $trnL_1$ and $trnL_2$ which swapped positions in *P. spinulosa* as discussed above) were either immediately next to the NCR or in the case of *trnP* of *P. asiatica*, a short distance (189 bp) inside the NCR (Table 3) [4], indicating a possible role of NCR in tRNA gene translocation. The observed probability of tRNA gene translocation to be adjacent to NCR in these *Polyplax* species is significantly higher ($p < 0.00001$) than the expected probability (Table 4).

We investigated further if this was a broader pattern for parasitic lice with fragmented mt genomes by comparing the mt karyotypes: (1) between the MRCA of sucking lice [9] and the MRCA of *Polyplax* (Fig. 5); and (2) between the MRCA of sucking lice and each of the 12 non-*Polyplax* sucking louse species that have been sequenced to date [9, 12]. The MRCA of *Polyplax* species retained the same distribution pattern of protein-coding and rRNA genes and the same number of minichromosomes as the MRCA of sucking lice. However, five tRNA genes have translocated in the MRCA of *Polyplax* relative to the MRCA of sucking lice: all of these tRNA genes were translocated immediately next to the NCR despite alternative intergenic locations available for translocation (Table 5). The observed probability of tRNA gene translocation to be adjacent to NCR in the MRCA of *Polyplax* species is significantly higher ($p < 0.00001$) than the expected probability (Table 6). When compared with the MRCA of sucking lice, 19 of the 22 tRNA genes in total have translocated in the 12 species of non-*Polyplax* sucking lice: 32 out of the 54 (i.e. 59 %) translocation events resulted in tRNA genes moved to be adjacent to NCR despite alternative intergenic locations available for translocation (Table 7). The observed probability of tRNA gene translocation to be adjacent to NCR in these non-*Polyplax* species is 13.6 % higher than the expected probability but this difference is not statistically significant ($p = 0.116709$) (Table 8). Most likely tRNA gene translocation to be adjacent to NCR has been under-captured in non-*Polyplax* species (Table 7) due to the fact that any newly translocated tRNA gene adjacent to a NCR would inevitably push the previously translocated tRNA gene away from the NCR. After all, there can only be one gene at any time adjacent to each end of the NCR. Among the three comparisons above, the third comparison has the longest evolutionary time frame. The longer the evolutionary time is, the more likely tRNA gene translocation information to be adjacent to NCR will be under-captured.

Clearly, the boundaries between coding regions and NCRs are hot spots for tRNA gene translocation in parasitic lice that have fragmented mt genomes. Of the models and mechanisms proposed in previous studies (introduced above in the Background) [2, 4, 6, 9, 10], only the inter-minichromosomal recombination model proposed by Song et al., [6] based on the mt karyotype variation among *Haematopinus* lice can explain tRNA gene translocation between mt minichromosomes. However, it cannot explain why translocated tRNA genes tend to be in the boundaries between coding regions and NCRs. As DNA double-strand breaks (DSBs) are apparently necessary for gene translocation between mt minichromosomes, the boundary regions may have more frequent DSBs than other regions of the mt minichromosomes, possibly due to the specific AT-rich and GC-rich sequence elements in the boundary regions; these sequence elements are conserved among all mt minichromosomes of each species of parasitic louse (Additional file 4). Apparently, DSBs do not occur randomly. Tubbs et al. showed that in mouse and human cells, DSBs occur mostly in AT-rich regions, in particular at

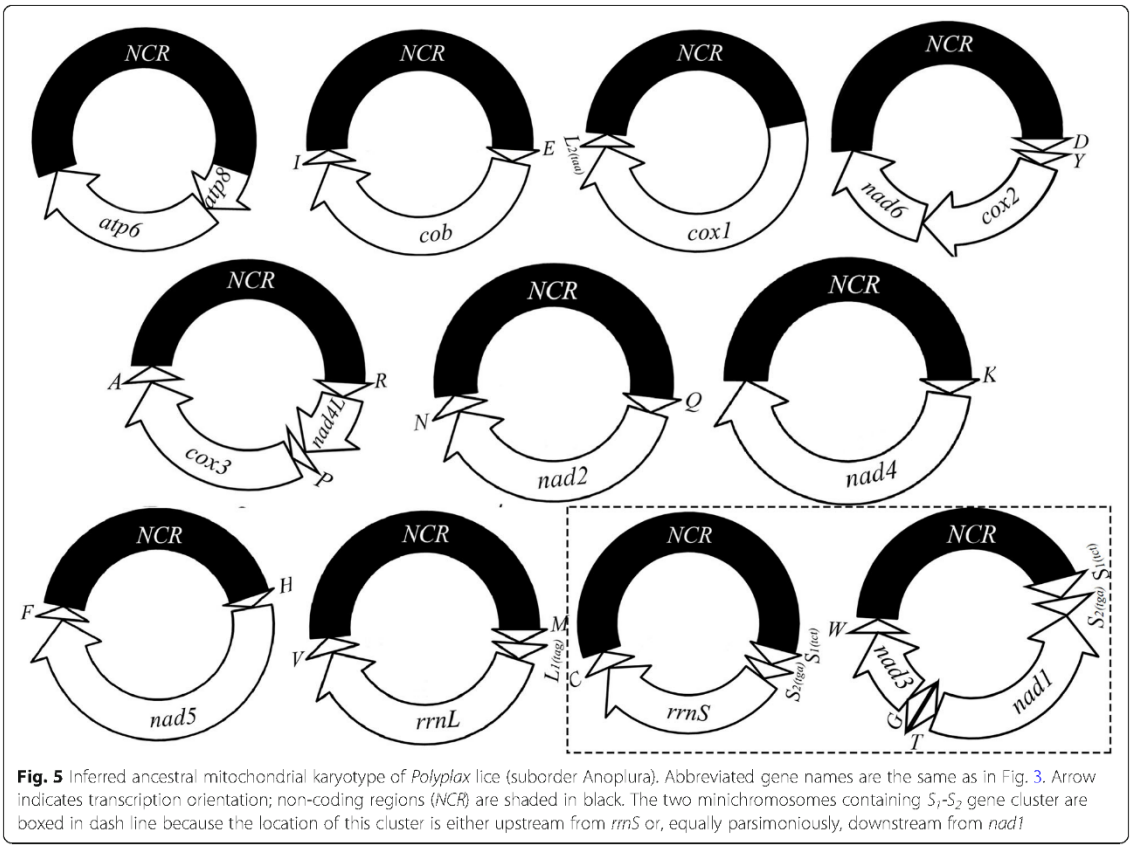


Fig. 5 Inferred ancestral mitochondrial karyotype of *Polyplax* lice (suborder Anoplura). Abbreviated gene names are the same as in Fig. 3. Arrow indicates transcription orientation; non-coding regions (*NCR*) are shaded in black. The two minichromosomes containing S_1-S_2 gene cluster are boxed in dash line because the location of this cluster is either upstream from *rrnS* or, equally parsimoniously, downstream from *nad1*

Table 4 Statistical analysis of expected probability and observed probability of tRNA gene translocation in three *Polyplax* species relative to their most recent common ancestor

tRNA genes	Expected probability of translocation to be adjacent to NCR if translocation is random	Observed probability of translocation to be adjacent to NCR	Expected probability of translocation to an intergenic region if translocation is random	Observed probability of translocation to an intergenic region
A	43.75 % (21/48)	100 % (2/2)	56.25 % (27/48)	0 %
F	43.75 % (21/48)	100 % (1/1)	56.25 % (27/48)	0 %
P	45.83 % (22/48)	100 % (1/1)	54.12 % (26/48)	0 %
T	45.83 % (22/48)	100 % (2/2)	54.12 % (26/48)	0 %
S_1-S_2	44.68 % (21/47)	100 % (3/3)	55.32 % (26/47)	0 %
Average	44.8 %	100 %	55.2 %	0 %
Two tailed *t*-value	-118.61978		115.73698	
p-value	< 0.00001		< 0.00001	
Significant at $p < 0.01$?	Yes		Yes	

Notes: Refer to Table 3 for tRNA gene translocation in *Polyplax* species. "Expected probability of translocation to be adjacent to NCR" = "total sites adjacent to NCR" / "total sites adjacent to NCR + total intergenic sites". "Expected probability of translocation to an intergenic region" = "total intergenic sites" / "total sites adjacent to NCR + total intergenic sites". "Observed probability of translocation to be adjacent to NCR" = "translocation adjacent to NCR" / "translocation adjacent to NCR + translocation to intergenic sites". "Observed probability of translocation to an intergenic region" = "translocation to intergenic sites" / "translocation adjacent to NCR + translocation to intergenic sites". *T*-Test was done at: https://www.socscistatistics.com/tests/studentttest/default2.aspx

Table 5 Translocated tRNA genes in the most recent common ancestor (MRCA) of *Polyplax* lice relative to the MRCA of sucking lice

tRNA gene	MRCA of sucking lice	MRCA of *Polyplax*	Translocated	Next to NCR
N	*atp8-atp6-N*	*Q-nad2-**N***	Yes	Yes
S_1-S_2	*E-cob-*S_1-S_2	***S*$_1$-*S*$_2$**-*rrnS-C* or ***S*$_1$-*S*$_2$**-*nad1-T-G-nad3-W*	Yes	Yes
I	*I-cox1-*L_2	*E-cob-**I***	Yes	Yes
Q	*Q-nad1-T-nad1-G-nad3-W*	***Q**-nad2-N*	Yes	Yes

Note: translocated tRNA genes in the MRCA of *Polyplax* lice are in bold

homopolymeric (dA/dT) tracts [40]. Sinai et al. showed that integration of AT-dinucleotide rich sequences into human chromosomes created recurrent gaps and breaks at the integration sites, due to a significantly increased tendency to fold into branched secondary structures at these sites [41]. Baudat et al. showed that a histone methyl transferase, PRDM9, is a major determinant of meiotic recombination hotspots in humans and mice, and furthermore, the human PRDM9 binding sites contain a 13-mer GC-rich motif and are DSB hotspots [42]. Sundararajan et al. showed that DSB hotspots have a strong tendency to occur in genes with high GC content in maize [43]. He et al. investigated the DNA sequence of DSB hotspots and identified a 20-bp-long GC-rich degenerate DNA sequence motif (named Maize Hotspot Sequence) present in about 72 % of genic hotspots in maize [44]. In general, non-canonical DNA structures such as hairpins and G-quadruplexes render a genome prone to damage including DSBs [45]. We counted the number of translocated tRNA genes in sucking lice at each end of the mt non-coding regions (Tables 3, 5 and 7) – the distribution is rather even (20 vs. 20) between the two ends where the AT-rich and GC-rich motifs were found respectively. Our observed pattern of mt tRNA gene translocation in sucking lice aligns well with the evidence presented in these experimental studies on DSB hotspots. Future studies on DSBs and DNA repair in parasitic lice may help understand further the mechanisms of mt tRNA gene translocation observed in sucking lice in the present study.

In conclusion, we sequenced the fragmented mt genome of the Asian grey shrew louse, *Polyplax reclinata*, which comprises 11 circular minichromosomes. We also inferred the ancestral mt karyotype of *Polyplax* lice based on the data available to date. We found that tRNA genes are entirely responsible for mt karyotype variation among the three species of *Polyplax* lice sequenced to date. Furthermore, tRNA gene translocation in these *Polyplax* lice is much more frequent than that of protein-coding and rRNA genes; all tRNA gene translocation observed occurs between different types of minichromosomes and is directional towards the boundaries with the control region. A similar pattern of tRNA gene translocation is also conserved in other sucking lice with fragmented mt genomes. Our results show that inter-minichromosomal tRNA gene translocation orientated

Table 6 Statistical analysis of expected probability and observed probability of tRNA gene translocation in the most recent common ancestor (MRCA) of *Polyplax* lice relative to the MRCA of sucking lice

tRNA genes	Expected probability of translocation to be adjacent to NCR if translocation is random	Observed probability of translocation to be adjacent to NCR	Expected probability of translocation to an intergenic region if translocation is random	Observed probability of translocation to an intergenic region
N	21/46 = 45.65 %	100 %	25/46 = 54.35 %	0 %
S_1-S_2	21/45 = 46.67 %	100 %	24/45 = 53.33 %	0 %
I	21/46 = 45.65 %	100 %	25/46 = 54.35 %	0 %
Q	21/46 = 45.65 %	100 %	25/46 = 54.35 %	0 %
Average	45.9 %	100 %	54.1 %	0 %
Two tailed *t*-value	-212.13726		212.13726	
p-value	< 0.00001		< 0.00001	
Significant at $p < 0.01$?	Yes		Yes	

Notes: Refer to Table 5 for tRNA gene translocation in the MRCA of *Polyplax* species. Refer to Table 4 notes for probability calculation and *T*-Test

Table 7 Translocated tRNA genes in non-*Polyplax* sucking lice relative to the most recent common ancestor (MRCA) of sucking lice

tRNA gene	MRCA of sucking lice	Non-*Polyplax* sucking lice	Species	Translocated	Next to NCR
N	*atp8-atp6-N*	*E-cob-S_1-S_2-**N***	*Microthoracius praelongiceps*[9]	Yes	Yes
		*cob-S_1-**N**-E-M*	*Pediculus schaeffi*[7]	Yes	No
		*S_1-**N**-E*	*Pediculus capitis*[2]	Yes	No
			Pediculus humanus[2]		
		*T-nad1-Q-**N**-C*	*Pedicinus badii*[12]	Yes	No
		*T-nad1-Q-**N**-G-nad3-W*	*Pedicinus obtusus*[12]	Yes	No
E	*E-cob-S_1-S_2*	*cob-S_1-N-**E**-M*	*Pediculus schaeffi*[7]	Yes	No
		*F-nad6-**E**-M*	*Pthirus pubis*[2]	Yes	No
		*S_1-N-**E***	*Pediculus capitis*[2]	Yes	Yes
			Pediculus humanus[2]		
S_1	*E-cob-S_1-S_2*	*D-Y-cox2-**S_1**-S_2-P-cox3-A*	*Haematopinus apri*[3]	Yes	No
			Haematopinus suis[3]		
			Haematopinus asini[6]		
		***S_1**-N-E*	*Pediculus capitis*[2]	Yes	Yes
			Pediculus humanus[2]		
S_2	*E-cob-S_1-S_2*	*D-Y-cox2-S_1-**S_2**-P-cox3-A*	*Haematopinus apri*[3]	Yes	No
			Haematopinus suis[3]		
			Haematopinus asini[6]		
		*G-nad3-V-W-**S_2***	*Pthirus pubis*[2]	Yes	Yes
I	*I-cox1-L_2*	*P-nad2-**I***	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		
			Pthirus pubis[2]		
		*R-nad4L-P-cox3-**I***	*Pedicinus obtusus*[12]	Yes	Yes
L_2	*I-cox1-L_2*	***L_2**-rrnS-C*	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		
D	*D-Y-cox2-nad6*	*T-**D**-H-R-nad4L*	*Pthirus pubis*[2]	Yes	No
		*T-**D**-H*	*Pediculus schaeffi*[7]	Yes	No
			Pediculus capitis[2]		
			Pediculus humanus[2]		
		*G-nad3-W-**D**-A-V*	*Pedicinus badii*[12]	Yes	No
		***D**-A-M-F-nad6*	*Pedicinus obtusus*[12]	Yes	Yes
P	*R-nad4L-P-cox3-A*	***P**-nad2-I*	*Pthirus pubis*[2]	Yes	Yes
			Pediculus schaeffi[7]		
			Pediculus capitis[2]		
			Pediculus humanus[2]		
A	*R-nad4L-P-cox3-A*	*G-nad3-W-D-**A**-V*	*Pedicinus badii*[12]	Yes	No
		*D-**A**-M-F-nad6*	*Pedicinus obtusus*[12]	Yes	No
Q	*Q-nad1-T-G-nad3-W*	***Q**-N-E*	*Pediculus humanus*[2]	Yes	Yes
		*T-nad1-**Q**-N-C*	*Pedicinus badii*[12]	Yes	No
		*T-nad1-**Q**-N-G-nad3-W*	*Pedicinus obtusus*[12]	Yes	No
T	*Q-nad1-T-G-nad3-W*	***T**-nad1-Q-N-C*	*Pedicinus badii*[12]	Yes	Yes
		***T**-nad1-Q-N-G-nad3-W*	*Pedicinus obtusus*[12]	Yes	Yes

Table 7 Translocated tRNA genes in non-*Polyplax* sucking lice relative to the most recent common ancestor (MRCA) of sucking lice *(Continued)*

tRNA gene	MRCA of sucking lice	Non-*Polyplax* sucking lice	Species	Translocated	Next to NCR
		*D-Y-cox2-**T***	*Hoplopleura kitti*[5]	Yes	Yes
		*R-nad4L-P-cox3-A-**T***	*Hoplopleura akanezumi*[5]	Yes	Yes
		***T**-D-H-R-nad4L*	*Pthirus pubis*[2]	Yes	Yes
		***T**-D-H*	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		
		***T**-nad1-Q-N-C*	*Pedicinus badii*[12]	Yes	Yes
		***T**-nad1-Q-N-G-nad3-W*	*Pedicinus obtusus*[12]	Yes	Yes
G	*Q-nad1-T-G-nad3-W*	***G**-nad4L-V*	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus capitis[2]		
W	*Q-nad1-T-G-nad3-W*	*C-nad6-**W**-L_2*	*Hoplopleura kitti*[5]	Yes	No
			Hoplopleura akanezumi[5]		
		*G-nad3-V-**W**-S_2*	*Pthirus pubis*[2]	Yes	No
H	*H-nad5-F*	*T-D-**H**-R-nad4L*	*Pthirus pubis*[2]	Yes	No
		*T-D-**H***	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		
F	*H-nad5-F*	***F**-nad6-E-M*	*Pthirus pubis*[2]	Yes	Yes
		*M-**F**-nad6*	*Pedicinus badii*[12]	Yes	No
		*D-A-M-**F**-nad6*	*Pedicinus obtusus*[12]	Yes	No
C	*rrnS-C*	***C**-nad6-W-L_2*	*Hoplopleura kitti*[5]	Yes	Yes
			Hoplopleura akanezumi[5]		
		*T-nad1-Q-N-**C***	*Pedicinus badii*[12]	Yes	Yes
		*atp8-atp6-**C***	*Pedicinus obtusus*[12]	Yes	Yes
M	*M-L_1-rrnL-V*	*R-nad4L-nad6-**M***	*Haematopinus apri*[3]	Yes	Yes
			Haematopinus suis[3]		
		*F-nad6-E-**M***	*Pthirus pubis*[2]	Yes	Yes
		*cob-S_1-N-E-**M***	*Pediculus schaeffi*[7]	Yes	Yes
		***M**-F-nad6*	*Pedicinus badii*[12]	Yes	Yes
		*D-A-**M**-F-nad6*	*Pedicinus obtusus*[12]	Yes	No
L_1	*M-L_1-rrnL-V*	***L_1**-rrnS-C*	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		
V	*M-L_1-rrnL-V*	*G-nad3-**V**-W-S_2*	*Pthirus pubis*[2]	Yes	No
		*G-nad3-W-D-A-**V***	*Pedicinus badii*[12]	Yes	Yes
		*nad2-Y-cox2-**V***	*Pedicinus obtusus*[12]	Yes	Yes
		*G-nad4L-**V***	*Pediculus schaeffi*[7]	Yes	Yes
			Pediculus capitis[2]		
			Pediculus humanus[2]		

Notes: Translocation events shared by congeneric species are grouped together and counted as single events. Genes underlined have opposite orientation of transcription to other genes

Table 8 Statistical analysis of expected probability and observed probability of tRNA gene translocation in non-*Polyplax* sucking lice relative to the most recent common ancestor of sucking lice

tRNA genes	Expected probability of translocation to be adjacent to NCR if translocation is random	Observed probability of translocation to be adjacent to NCR	Expected probability of translocation to an intergenic region if translocation is random	Observed probability of translocation to an intergenic region
N	21/45 = 46.67 %	20 %	24/45=53.33%	80 %
E	21/45 = 46.67 %	33.3 %	24/45 = 53.33 %	66.7 %
S_1	22/46 = 47.83 %	50 %	23/46 = 52.17 %	50 %
S_2	21/45 = 46.67 %	50 %	24/45 = 53.33 %	50 %
I	21/45 = 46.67 %	100 %	24/45 = 53.33 %	0 %
L_2	21/45 = 46.67 %	100 %	24/45 = 53.33 %	0 %
D	21/45 = 46.67 %	25 %	24/45 = 53.33 %	75 %
P	22/46 = 47.83 %	100 %	23/46 = 52.17 %	0 %
A	21/45 = 46.67 %	0 %	24/45 = 53.33 %	100 %
Q	21/45 = 46.67 %	33.3 %	24/45 = 53.33 %	66.7 %
T	22/46 = 47.83 %	100 %	23/46 = 52.17 %	0 %
G	22/46 = 47.83 %	100 %	23/46 = 52.17 %	0 %
W	21/45 = 46.67 %	0 %	24/45 = 53.33 %	100 %
H	21/45 = 46.67 %	50 %	24/45 = 53.33 %	50 %
F	21/45 = 46.67 %	33.3 %	24/45 = 53.33 %	66.7 %
C	21/45 = 46.67 %	100 %	24/45 = 53.33 %	0 %
M	21/45 = 46.67 %	80 %	24/45 = 53.33 %	20 %
L_1	22/46 = 47.83 %	100 %	23/46 = 52.17 %	0 %
V	21/45 = 46.67 %	75 %	24/45 = 53.33 %	25 %
Average	46.9 %	60.5 %	53.1 %	39.5 %
Two tailed *t*-value	-1.60738		1.61464	
p-value	0.116709		0.11512	
Significant at $p < 0.01$?	No		No	

Notes: Refer to Table 7 for tRNA gene translocation in non-*Polyplax* sucking lice. Refer to Table 4 notes for probability calculation and *T*-Test

towards the two ends of the control region is a major contributing factor to the highly dynamic mitochondrial genome organization in the parasitic lice of mammals.

Abbreviations

µl: Microliter; *atp6* and *atp8*: Genes for ATP synthase subunits 6 and 8; bp: Base pair; *cob*: Gene for cytochrome b; *cox1*, *cox2* and *cox3*: Genes for cytochrome c oxidase subunits 1, 2 and 3; DNA: Deoxyribonucleic acid; kb: Kilo base pair; min: Minute; MRCA: Most recent common ancestor; mt: Mitochondrial; Mya: Million years ago; *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*: Mitochondrial genes for NADH dehydrogenase subunits 1–6 and 4L; PCR: Polymerase chain reaction; RNA: Ribonucleic acid; rRNA: Ribosomal RNA; *rrnS* and *rrnL*: Genes for small and large subunits of ribosomal RNA; sec: Second; T: Thymine; tRNA: Transfer RNA; *trnA* or *A*: tRNA gene for alanine; *trnC* or *C*: tRNA gene for cysteine; *trnD* or *D*: tRNA gene for aspartic acid; *trnE* or *E*: tRNA gene for glutamic acid; *trnF or F*: tRNA gene for phenylalanine; *trnG or G*: tRNA gene for glycine; *trnH* or *H*: tRNA gene for histidine; *trnI* or *I*: tRNA gene for isoleucine; *trnK* or *K*: tRNA gene for lysine; $trnL_1$ or L_1: tRNA gene for leucine (anticodon NAG); $trnL_2$ or L_2: tRNA gene for leucine (anticodon YAA); *trnM* or *M*: tRNA gene for methionine; *trnN* or *N*: tRNA gene for asparagine; *trnP* or *P*: tRNA gene for proline; *trnQ* or *Q*: tRNA gene for glutamine; *trnR* or *R*: tRNA gene for arginine; $trnS_1$ or S_1: tRNA gene for serine (anticodon NCU); $trnS_2$ or S_2: tRNA gene for serine (anticodon NGA); *trnT* or *T*: tRNA gene for threonine; *trnV* or *V*: tRNA gene for valine; *trnW* or *W*: tRNA gene for tryptophan; *trnY* or *Y*: tRNA gene for tyrosine; U: Uracil

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07859-w.

Additional file 1: Image of the voucher specimens (# 364) of the shrew louse, *Polyplax reclinata*.

Additional file 2: The primers used to amplify the mitochondrial genes, minichromosomes and coding regions of the Asian grey shrew louse, *Polyplax reclinata*.

Additional file 3: PCR amplicons generated from the mitochondrial (mt) minichromosomes of the Asian grey shrew louse, *Polyplax reclinata*. (A) Lane 1: amplicon of S_1-S_2-*rrnS*-*C* minichromosome generated with the

primer pair 12S364F-12S364R. Lane 2: amplicon of *M-L*$_1$*-rrnL-V* minichromosome generated with the primer pair 16S364F-16S364R. Lane 3: 500 bp Ladder (Tiangen) (band size in bp indicated). Lane 4: amplicons of coding regions of all mt minichromosomes generated with the primer pair 364 F-364R. Lane 5: GeneRuler 100 bp DNA Ladder (Thermo Scientific) (band size in bp indicated). The lane to the right of lane 2 is irrelevant to this manuscript. The lane to the left of lane 3 is empty. (B) Lane 1: 1 kb ladder (Tiangen) (band size in bp indicated); Lanes 2–10: amplicons of individual mt minichromosomes, *T-D-Y-cox2-nad6-A*, *R-nad4L-P-cox3*, *Q-nad2-N*, nad1 *-G-nad3-W* (gene underlined has opposite transcription orientation to other genes), *K-nad4*, *H-nad5-F*, *E-cob-I*, *cox1-L*$_2$, *atp8-atp6*. Details of the primers used to amplify individual mt minichromosomes are provided in Additional file 2.

Additional file 4: Alignment of the non-coding region (*NCR*) sequences of 11 mitochondrial minichromosomes of the Asian grey shrew louse, *Polyplax reclinata*. The primer pair, 364 F and 364R, were used to amplify the coding regions of the 11 minichromosomes (see also Additional file 3). Asterisk symbol "*" indicates conserved nucleotides; hyphen "-" indicates absent nucleotides.

Acknowledgements
We thank Dr Fan Song for assistance with an initial phylogenetic tree analysis.

Authors' contributions
WGD, YLD, XGG and RS designed the research. WGD and RS performed the research. WGD and XGG contributed reagents and materials. WGD, YLD and RS analysed the data. WGD and RS wrote the manuscript. All authors have read and approved the final manuscript.

Funding
We acknowledge funding support from the National Natural Science Foundation of China (No. 31660314 and 32060143 to WGD) and the Innovation Team of Vector Biology, Dali University (No. ZKLX2019104 to XGG). The funding agencies had no roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials
The nucleotide sequence of the mitochondrial genome of the Asian grey shrew louse, *Polyplax reclinata*, has been deposited in GenBank (accession number MW291451-MW291461) and will be released publicly once this manuscript is accepted for publication.

Declarations

Ethics approval and consent to participate
Capture of shrews was approved by health authorities in Yunnan Province, China. Animal capture protocols and procedures were approved by the animal ethics committees at Dali University (approval # DLDX-20100421-001). All collection methods were carried out in accordance with the approved guidelines and regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests to declare.

Author details
[1]Institute of Pathogens and Vectors, Key Laboratory for Preventing and Controlling Plague in Yunnan Province, Dali University, 671000 Dali, China. [2]GeneCology Research Centre, University of the Sunshine Coast, Maroochydore, Queensland, Australia. [3]School of Science, Technology and Engineering, University of the Sunshine Coast, Maroochydore, Queensland, Australia.

Received: 2 February 2021 Accepted: 29 June 2021
Published online: 06 August 2021

References

1. 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.
2. 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.
3. Jiang H, Barker SC, Shao R. Substantial variation in the extent of mitochondrial genome fragmentation among blood-sucking lice of mammals. Genome Biol Evol. 2013;5:1298–308.
4. Dong WG, Song S, Jin DC, Guo XG, Shao R. 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 Genomics. 2014;15:44.
5. 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 *Haplopleura* rodent lice (family Hoplopleuridae). BMC Genomics. 2014;15:751.
6. Song SD, Barker SC, Shao R. Variation in mitochondrial minichromosome composition between blood-sucking lice of the genus *Haematopinus* that infest horses and pigs. Parasite Vector. 2014;7:144.
7. Herd KE, Barke 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 Genomics. 2015;16:661.
8. 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.
9. 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.
10. Song F, Li H, Liu GH, Wang W, James P, Colwell DD, Tran A, Gong SY, Cai WZ, Shao R. Mitochondrial genome fragmentation unites the parasitic lice of eutherian mammals. Syst Biol. 2019;68:430–40.
11. 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.
12. 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.
13. Smith VS, Ford T, Johnson KP, Johnson PCD, Yoshizawa K, Light JE. Multiple lineages of lice pass through the K-Pg boundary. Biol Lett. 2011;5:782–5.
14. Light JE, Smith VS, Allen JM, Durden LA, Reed DL. Evolutionary history of mammalian sucking lice (Phthiraptera: Anoplura). BMC Evol Biol. 2010;10:292.
15. Johnson KP, Nguyen N-p, Sweet AD, Boyd BM, Warnow T, Allen JM. Simultaneous radiation of bird and mammal lice following the K-Pg boundary. Biol Lett. 2018;14:141.
16. Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999;27:1767–80.
17. Wurster DH, Benirschke K. Indian muntjac, Muntiacus muntjak: A deer with a low diploid chromosome number. Science. 1970;168:1364–6.
18. Qumsiyeh MB. Evolution of number and morphology of mammalian chromosomes. J Hered. 1994;85:455–65.
19. Rieseberg LH. Chromosomal rearrangements and speciation.Trends Ecol Evol. 2001;16:351–8.
20. Kittler R, Kayser M, Stoneking M. Molecular evolution of Pediculus humanus and the origin of clothing. Curr Biol. 2003;13:1414–7.
21. Kittler R, Kayser M, Stoneking M. Molecular evolution of *Pediculus humanus* and the origin of clothing (Erratum). Curr Biol. 2004;14:2309.
22. Toups MA, Kitchen A, Light JE, Reed DL. Origin of clothing lice indicates early clothing use by anatomically modern humans in Africa. Mol Biol Evol. 2011;28:29–32.
23. Ottoni C, Girdl and Flink L, Evin A, Georg C, Cupere BD, Neer WV, Bartosiewicz L, Linderholm A, Barnett R, Peters J, Decorte R, Waelkens M, Vanderheyden N, Ricaut FX, Cakırlar C, Cevik O, Hoelzel AR, Mashkour M, Fatemeh A, Karimlu M, Seno SS, Daujat J, Brock F, Pinhasi R, Hongo H, Perez-Enciso M, Rasmussen M, Frantz L, Megens HJ, Crooijmans R, Groenen

M, Arbuckle B, Benecke N, Vidarsdottir US, Burger J, Cucchi T, Dobney K, Larson G. Pig domestication and human-mediated dispersal in western eurasia revealed through ancient DNA and geometric morphometrics. Mol Biol Evol. 2013;4:824–32.
24. Chin TH. Taxonomy and fauna of sucking lice (Anoplura) in China. Beijing Science Press; 1999.
25. Kambhampati S, Smith PT. PCR primers for amplification of four insect mitochondrial gene fragments. Insect Mol Biol. 1995;4:233–6.
26. 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. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.
27. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.
28. Laslett D, Canback B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics. 2008;24:172–5.
29. Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genet. 1993;3:266–72.
30. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
31. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947–8.
32. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010;38:W7-W13.
33. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
34. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56:564–77.
35. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolution genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
36. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19:1572–4.
37. Durden LA, Musser GG. The sucking lice (Insecta, Anoplura) of the world: a taxonomic checklist with records of mammalian hosts and geographical distributions. Bull Amer Mus Nat Hist. 1994;218:1–90.
38. Durden LA, Musser GG. The mammalian hosts of the sucking lice (Anoplura) of the world: a host-parasite list. Bull Soc Vector Ecol. 1994;19:130–68.
39. Kim KC. Evolutionary parallelism in Anoplura and eutherian mammals. In: Service MW, editor. Biosystematics of haematophagous insects. Oxford: Claredon Press; 1988. Vol. 37. p. 91–114.
40. Tubbs A, Sridharan S, Wietmarschen N, Maman Y, Callen E, Stanlie A, Wu W, Xia W, Day A, Wong N, Yin M, Canela A, Fu H, Redon C, Pruitt SC, Jaszczyszyn Y, Aladjem M, Aplan PD, Hyrien O, Nussenzweig A. Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. Cell. 2018;174:1127–42.
41. Sinai MIT, Salamon A, Stanleigh N, Goldberg T, Weiss A, Wang YH, Kerem B. AT-dinucleotide rich sequences drive fragile site formation. Nucleic Acids Res. 2019;47:9685–95.
42. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science. 2010;327:836–40.
43. Sundararajan A, Dukowic-Schulze S, Kwicklis M, Engstrom K, Garcia N, Oviedo OJ, Ramaraj T, Gonzales MD, He Y, Wang M, Sun Q, Pillardy J, Kianian SF, Pawlowski WP, Chen C, Mudge J. Gene evolutionary trajectories and GC patterns driven by recombination in *Zea mays*. Front Plant Sci. 2016;7:1433.
44. He Y, Wang M, Dukowic-Schulze S, Zhou A, Tiang CL, Shilo S, Sidhu GK, Eichten S, Bradbury P, Springer NM, Buckler ES, Levy AA, Sun Q, Pillardy J, Kianian PMA, Kianian SF, Chen C, Pawlowski WP. Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. Proc Natl Acad Sci USA. 2017;114:12231–6.
45. Puget N, Miller KM, Legube G. Non-canonical DNA/RNA structures during transcription-coupled double-strand break repair: roadblocks or bona fide repair intermediates? DNA Repair. 2019;81:102661.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- **fast, convenient online submission**
- **thorough peer review by experienced researchers in your field**
- **rapid publication on acceptance**
- **support for research data, including large and complex data types**
- **gold Open Access which fosters wider collaboration and increased citations**
- **maximum visibility for your research: over 100M website views per year**

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

Appendix III

Manuscript:

Fragmented mitochondrial genomes of seal lice (family Echinophthiriidae) and gorilla louse (family Pthiridae): frequent minichromosomal splits and a host switch of lice between seals

Published in March 2022

Dong *et al. BMC Genomics* (2022) 23:283
https://doi.org/10.1186/s12864-022-08530-8

BMC Genomics

RESEARCH Open Access

Fragmented mitochondrial genomes of seal lice (family Echinophthiriidae) and gorilla louse (family Pthiridae): frequent minichromosomal splits and a host switch of lice between seals

Yalun Dong[1,2], Min Zhao[1,2] and Renfu Shao[1,2*]

Abstract

Background: The mitochondrial (mt) genomes of 15 species of sucking lice from seven families have been studied to date. These louse species have highly dynamic, fragmented mt genomes that differ in the number of minichromosomes, the gene content, and gene order in a minichromosome between families and even between species of the same genus.

Results: In the present study, we analyzed the publicly available data to understand mt genome fragmentation in seal lice (family Echinophthiriidae) and gorilla louse, *Pthirus gorillae* (family Pthiridae), in particular the role of minichromosome split and minichromosome merger in the evolution of fragmented mt genomes. We show that 1) at least three ancestral mt minichromosomes of sucking lice have split in the lineage leading to seal lice, 2) one minichromosome ancestral to primate lice has split in the lineage to the gorilla louse, and 3) two ancestral minichromosomes of seal lice have merged in the lineage to the northern fur seal louse. Minichromosome split occurred 15-16 times in total in the lineages leading to species in six families of sucking lice investigated. In contrast, minichromosome merger occurred only four times in the lineages leading to species in three families of sucking lice. Further, three ancestral mt minichromosomes of sucking lice have split multiple times independently in different lineages of sucking lice. Our analyses of mt karyotypes and gene sequences also indicate the possibility of a host switch of crabeater seal louse to Weddell seals.

Conclusions: We conclude that: 1) minichromosome split contributes more than minichromosome merger in mt genome fragmentation of sucking lice, and 2) mt karyotype comparison helps understand the phylogenetic relationships between sucking louse species.

Keywords: Seal lice, Gorilla louse, Mitochondrial minichromosomes, Split and merge, Host switch

Background

Sucking lice (parvorder Anoplura) are obligate ectoparasites of eutherian mammals and feed exclusively on their host blood [1]. There are over 500 species of sucking lice, classified into 50 genera and 15 families [2]. Sucking lice have a flattened body to avoid host grooming, elongated mouthparts for blood-sucking, and sharp claws to climb and hold on host hair [3]. Unlike most animals, which have single-chromosome mitochondrial

*Correspondence: rshao@usc.edu.au
[1] Centre for Bioinnovation, University of the Sunshine Coast, 90 Sippy Downs Drive, Sippy Downs, Queensland 4556, Australia
Full list of author information is available at the end of the article

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(mt) genomes [4], the sucking lice investigated to date have fragmented mt genomes, each with 9-20 minichromosomes [5–8]. Fifteen species of sucking lice from seven of the 15 families have been sequenced for complete or near complete mt genomes [5–13]. Human head louse (*Pediculus humanus capitis*, family Pediculidae) and human body louse (*Pediculus humanus corporis*) have the most fragmented mt genomes with 20 minichromosomes [5, 6], followed by chimpanzee louse (*Pediculus schaeffi*) with 18 minichromosomes [14]. Human pubic louse (*Pthirus pubis*, family Pthiridae) has 15 minichromosomes (*trnN* gene not identified) [6, 12]. Colobus louse (*Pedicinus badii*, family Pedicinidae) and macaque louse (*Pedicinus obtusus*) have 14 and 12 minichromosomes, respectively [12]. Guanaco louse (*Microthoradus praelongiceps*, family Microthoraciidae) has 12 minichromosomes [8]. Three rodent lice of the family Polyplacidae (*Polyplax asiatica*, *Polyplax spinulosa* and *Polyplax reclinata*) each have 11 minichromosomes [11, 13]. Two rodent lice of the family Hoplopleuridae (*Hoplopleura akanezumi* and *Hoplopleura kitti*) each have at least 11 minichromosomes with five and three genes not identified, respectively [10]. Wild pig louse (*Haematopinus apri*, family Haematopinidae), domestic pig louse (*Haematopinus suis*), and horse louse (*Haematopinus asini*) have the least fragmented mt genomes, each with nine minichromosomes [7, 9].

No species in the other eight families of sucking lice have been studied for their mt genome organization. These families are Echinophthiriidae (lice of seals), Enderleinellidae (lice of squirrels), Linognathidae (lice of cattle, sheep and goats), Hamophthiriidae (lice of colugos), Hybophthiridae (lice of aardvarks), Neolinognathidae (lice of elephant shrews), Pecaroecidae (lice of peccaries), and Ratemiidae (lice of horses, donkeys and zebras) [2, 15]. In the present study, we analyzed the publicly available Sequence Read Archive (SRA) data to understand the mt genome fragmentation in five species of seal lice (*Antarctophthirus carlinii*, *A. lobodontis*, *A. microchir*, *Lepidophthirus macrorhini*, *Proechinophthirus fluctus*) in the family Echinophthiriidae, and the gorilla louse (*Pthirus gorillae*) in the family Pthiridae. We also investigated the role of minichromosome split and merger in the evolution of the highly dynamic mt genome organization observed in sucking lice.

The family Echinophthiriidae has five genera and 13 species [2]. The genus *Antarctophthirus* has seven species, while the other four genera have one to two species. Seal lice are unique among parasitic lice in living in the marine environment [16, 17]. Ancestral seal lice were thought to be terrestrials but evolved unique morphology to adapt to marine life together with their hosts [15, 16, 18]. The family Pthiridae has only one genus *Pthirus* with two species: *Pthirus pubis* (human pubic louse) and *Pthirus gorillae* (gorilla louse) [2]. These two species diverged 3-4 million years ago (MYA), possibly due to host switch from gorillas to humans [19]. The mt genome of *Pthirus pubis* has been sequenced previously [6, 12]; however, the mt genome of *Pthirus gorillae* has not been studied. We find that: 1) seal lice differ substantially from sucking lice in other families in mt genome organization; 2) gorilla louse has a more fragmented mt genome than human pubic louse; 3) minichromosome split occurred in both lineages leading to seal lice and gorilla louse; 4) minichromosome merger occurred in the lineage to the northern fur seal louse; and 5) likely a host switch of louse occurred between crabeater seals and Weddell seals.

Results

Mitochondrial minichromosomes of *Lepidophthirus macrorhini* – louse of southern elephant seal (*Mirounga leonine*)

The Illumina data of *Lepidophthirus macrorhini* (SRR5809351) from SRA database contains 47,978,079 paired-end sequence reads; each sequence read is 150 bp in size. We assembled these sequence reads and identified 27 of the 37 mt genes typical of bilateral animals. These genes are on 11 minichromosomes; each minichromosome contains a single protein-coding or rRNA gene with none or up to three tRNA genes except *atp8-atp6-N* minichromosome, which contains two protein-coding genes (Fig. 1A, Supplementary Table 1). The genes in all the minichromosomes have the same orientation of transcription relative to the non-coding regions except for *Q-nad1-T-W* minichromosome, in which three genes (underlined) have the opposite transcription orientation to all other genes (Fig. 1A, Supplementary Table 1). *trnW* gene were

(See figure on next page.)

Fig. 1 Mitochondrial genomes of five seal louse species: **A** southern elephant seal louse, *Lepidophthirus macrorhini*; **B** northern fur seal louse, *Proechinophthirus fluctus*; **C** Weddell seal louse, *Antarctophthirus carlinii*, and crabeater seal louse, *Antarctophthirus lobodontis*; and **D** Australian sea lion louse, *Antarctophthirus microchir*. Conserved AT-rich motifs are in red and conserved GC-rich motifs are in green. In *Antarctophthirus microchir*, AT-rich motifs were not identified and GC-rich motifs were identified in only five of the 12 minichromosomes. Names and transcription orientation of genes are indicated in the coding region. 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, *nad1-5* and *nad4L* for NADH dehydrogenase subunits 1-5 and 4 L; *rnS* and *rnL* for small and large subunits of ribosomal RNA. tRNA genes are indicated with their single-letter abbreviations of the corresponding amino acids

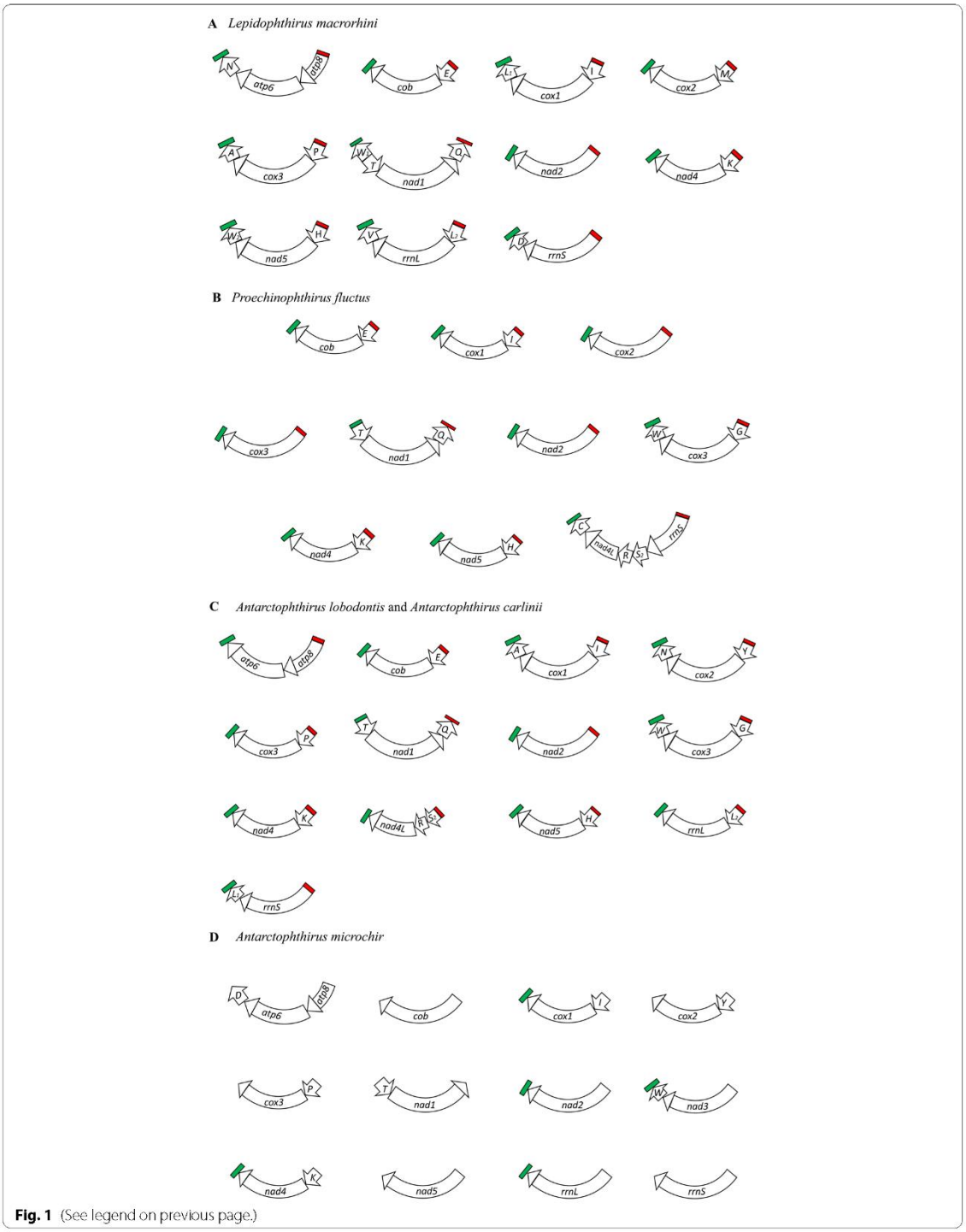


Fig. 1 (See legend on previous page.)

identified in two minichromosomes: *Q-nad1-T-*W_1 minichromosome and *H-nad5-*W_2 minichromosome. The pairwise identity between these two *trnW* genes is 50% (Supplementary Table 2); both have TCA anticodon and can form clove-leafed secondary structure (Supplementary Fig. 1). *trn*W_1 has a higher identity to *trnW* of other seal lice (55.4 to 69.1%) than *trn*W_2 has (42.6 to 55.1%), indicating *trn*W_1 is more likely the original copy of *trnW*. However, it is unclear whether *trn*W_2 is duplicated from *trn*W_1 or from other tRNA genes (Supplementary Table 2). We obtained ~300 bp non-coding sequence both upstream and downstream from the coding region of each minichromosome (Supplementary Table 1). A conserved AT-rich (80.3%) motif (56 bp) and a conserved GC-rich (60.7%) motif (63 bp) were found in the non-coding sequences of all the minichromosomes (Fig. 1A, Supplementary Fig. 2). Ten mt genes were not identified in our analysis of *Lepidophthirus macrorhini*: *nad3, nad4L, nad6, trnC, trnF, trnG, trnR, trn*S_1, *trn*S_2 and *trnY*. The annotated mt minichromosomes of *Lepidophthirus macrorhini* were available in GenBank (accession numbers MW803094-104).

Mitochondrial minichromosomes of *Proechinophthirus fluctus* – louse of northern fur seal (*Callorhinus ursinus*)

The Illumina data of *Proechinophthirus fluctus* (SRR5308138) from SRA database contains 5,819,192 paired-end sequence reads; each sequence read is 100 bp in length. We assembled these sequence reads and identified 22 of the 37 typical mt genes. These genes are on 10 minichromosomes; each minichromosome contains a single protein-coding or rRNA gene with none or up to three tRNA genes (Fig. 1B, Supplementary Table 3). The genes in all the minichromosomes have the same orientation of transcription relative to the non-coding regions except for *Q-nad1-T* minichromosome; these three genes have the opposite transcription orientation to all other genes (Fig. 1B). We obtained ~200 bp non-coding sequence both upstream and downstream from the coding region of each minichromosome (Supplementary Table 3). A conserved AT-rich (71.8%) motif (85 bp) and a conserved GC-rich (61.7%) motif (81 bp) were found in the non-coding sequences of all the minichromosomes (Fig. 1B, Supplementary Fig. 3). Fifteen mt genes were not identified in our analysis of the SRA data of *Proechinophthirus fluctus*: *atp6, atp8, nad6, rrnL, trnA, trnD, trnF, trn*L_1, *trn*L_2, *trnM, trnN, trnP, trn*S_1, *trnV* and *trnY*. The annotated mt minichromosomes of *Proechinophthirus fluctus* were available in GenBank (accession numbers MW803105-114).

Mitochondrial minichromosomes of *Antarctophthirus carlinii* - louse of Weddell seal (*Leptonychotes weddelli*), and *Antarctophthirus lobodontis* - louse of crabeater seal (*Lobodon carcinophagus*)

The Illumina data of *Antarctophthirus carlinii* (SRR5809348) and *Antarctophthirus lobodontis* (SRR5809349) from SRA database contains 39,054,456 and 45,005,741 paired-end sequence reads respectively; each sequence read is 150 bp. We assembled these sequence reads and identified 30 of the 37 typical mt genes on nine minichromosomes in each species (Fig. 1C). Each minichromosome contains a single protein-coding gene with none or up to two tRNA genes (Fig. 1C, Supplementary Tables 4 and 5). The gene content and gene arrangement in each mt minichromosome are the same between these two species; furthermore, the identity of each homologous gene is 94.3 to 100% between the two species. The genes in all the minichromosomes have the same orientation of transcription relative to the non-coding regions except for *Q-nad1-T* minichromosome, in which the three genes have the opposite transcription orientation to all other genes (Fig. 1C). We obtained ~300 bp non-coding sequence both upstream and downstream from the coding region of each minichromosome (Supplementary Tables 4 and 5). In *Antarctophthirus carlinii,* a conserved AT-rich (60%) motif (60 bp) and a conserved GC-rich (67.4%) motif (46 bp) were found in the non-coding sequences of all the minichromosomes (Supplementary Fig. 4). Conserved AT-rich (68.7%) motif (65 bp) and GC-rich (64.9%) motif (37 bp) were also found in *Antarctophthirus lobodontis* (Supplementary Fig. 5). Seven mt genes were not identified in our analysis of the SRA data of these two *Antarctophthirus* species: *nad6, trnC, trnD, trnF, trnM, trn*S_1 and *trnV*. The annotated mt minichromosomes of *Antarctophthirus carlinii* (accession numbers MW803073-81) and *Antarctophthirus lobodontis* (accession numbers MW803064-72) were available in GenBank.

Mitochondrial minichromosomes of *Antarctophthirus microchir* – louse of Australian sea lion (*Neophoca cinerea*)

The Illumina data of *Antarctophthirus microchir* (SRR5809347) from SRA database contains 43,650,933 paired-end sequence reads. Each sequence read is 150 bp in size. We assembled these sequence reads and identified 20 of the 37 typical mt genes. These genes are on 12 minichromosomes; each minichromosome contains a single protein-coding or rRNA gene with none or one tRNA gene except *atp8-atp6-D* minichromosome (Fig. 1D, Supplementary Table 6). We obtained 250 bp non-coding sequence both upstream and downstream from the

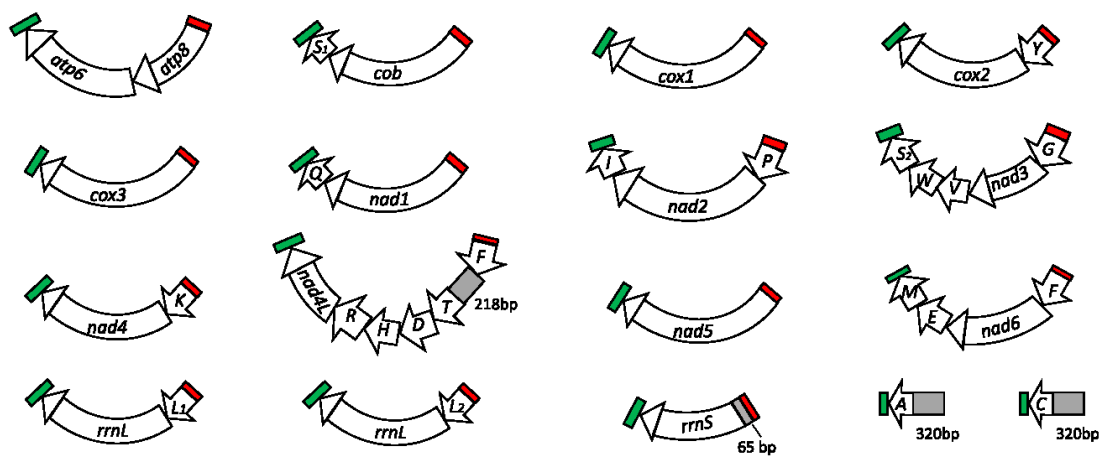


Fig. 2 Mitochondrial genome of the gorilla louse, *Pthirus gorillae*. Conserved AT-rich motifs are in red; conserved GC-rich motifs are in green; and degenerate genes are in grey. Names and transcription orientation of genes are indicated in the coding region. *nad6* is the gene for NADH dehydrogenase subunit 6. Names of other genes are described in Fig. 1 legend

coding region of each minichromosome (Supplementary Table 6). We identified a conserved GC-rich motif in the non-coding sequences of five of the 12 minichromosomes; however, we were unable to identify conserved AT-rich motif in the non-coding sequences of any minichromosomes (Fig. 1D, Supplementary Table 6, Supplementary Fig. 6). Full-length non-coding region sequences are needed to investigate whether the unidentified conserved motifs are in the middle section of the non-coding regions of *Antarctophthirus microchir*. Seventeen mt genes were not identified in our analysis of the SRA data of *Antarctophthirus microchir*: *nad4L, nad6, trnA, trnC, trnE, trnF, trnG, trnH, trnL*$_1$, *trnL*$_2$, *trnM, trnN, trnQ, trnR, trnS*$_1$, *trnS*$_2$ and *trnV*. The annotated mt minichromosomes of *Antarctophthirus microchir* were available in GenBank (accession numbers MW803082-93).

Mitochondrial minichromosomes of *Pthirus gorillae* – louse of Western gorilla (*Gorilla gorilla*)

The Illumina data of *Pthirus gorillae* (SRR5088474) from SRA database contains 60,425,294 paired-end sequence reads; each sequence read is 160 bp in length. We assembled these sequence reads and identified 36 of the 37 typical mt genes; these genes are on 17 minichromosomes (Fig. 2). *trnN* was the only gene not identified in our analysis. Fourteen of the 17 minichromosomes contain a single protein-coding or rRNA gene with none or up to five tRNA genes. Of the other three minichromosomes, *atp8-atp6* minichromosome contains no tRNA gene whereas *trnA* minichromosome and *trnC* minichromosome each have only a tRNA gene (Fig. 2, Supplementary Table 7). The genes in all the minichromosomes have the same orientation of transcription relative to the non-coding regions (Fig. 2). A 65-bp non-coding sequence upstream from *rrnS* has 58.5% identity to *trnL*$_1$ and 60% identity to *trnL*$_2$ (Fig. 3A). As *trnL* gene is upstream from *rrnS* in the human pubic louse (*Pthirus pubis*), human head louse (*Pediculus humanus capitis*) and human body louse (*Pediculus humanus corporis*) [5], the 65 bp non-coding sequence upstream from *rrnS* in *Pthirus gorillae* is very likely a degenerate *trnL* gene. Three other regions are likely degenerate genes too: 1) a 218-bp sequence between *trnF* and *trnT* has 100% identity with a 5′ section of *nad6* (Fig. 3B); 2) a 320-bp sequence upstream from *trnA* has 100% identity with a middle section of *rrnS* (Fig. 3C); and 3) a 320-bp sequence upstream from *trnC* has 53.3% identity to 3′ section of *rrnL* (Fig. 3D). We obtained ~320 bp non-coding sequence both upstream and downstream from the coding region of each minichromosome (Supplementary Table 7). A conserved AT-rich (76.6%) motif (96 bp) was found in the non-coding sequences of all the minichromosomes except *trnA* minichromosome and *trnC* minichromosome (Fig. 2; Supplementary Fig. 7). A GC-rich (76%) motif (25 bp) was found in the non-coding sequences of all the minichromosomes (Fig. 2; Supplementary Fig. 7). The annotated mt minichromosomes of *Pthirus gorillae* are available in GenBank (accession numbers MW803115-131).

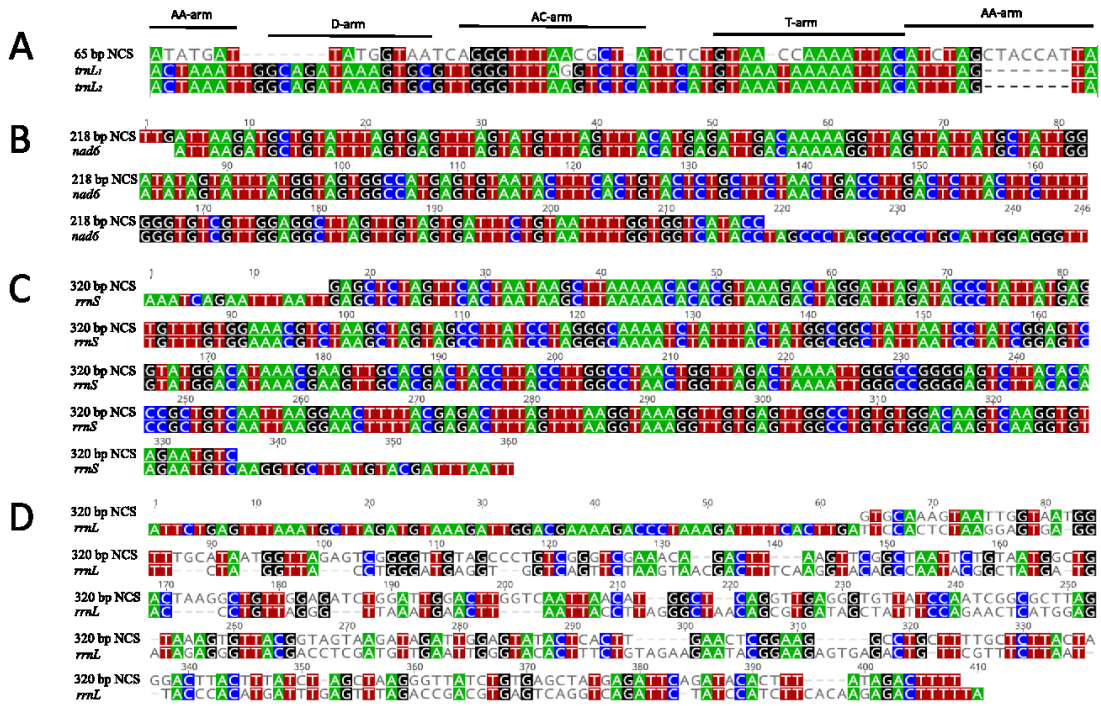


Fig. 3 Sequence alignment of degenerate genes with their corresponding full-length genes. **A**: the 65-bp non-coding sequence (NCS) in *rrnS* minichromosome has 58.5% identity to *trnL1* gene and 60% identity to *trnL2* gene. **B**: The 218-bp intergenic NCS in *F-T-D-H-R-nad4L* minichromosome has 215 bp identical to a 5′ section of *nad6* gene (nucleotide 1 to 215). **C** The 320-bp NCS in *A* minichromosome is identical to a middle section of *rrnS* gene (nucleotide 246 to 565). **D** The 320-bp NCS in *C* minichromosome has 53.3% identity to 3′ section of *rrnL* gene (nucleotide 838 to 1189)

Phylogeny of sucking lice based on mitochondrial gene sequences

We reconstructed the phylogeny of sucking lice (Anoplura) to assist the inference of ancestral mt karyotype of seal lice (family Echinophthiriidae). Both the Bayesian and maximum likelihood (ML) trees strongly support the monophyly of Echinophthiriidae and each of the other seven families of sucking lice (Fig. 4; Supplementary Fig. 8). The relationships among the five species of seal lice are resolved with strong support. The three *Antarctophthirus* species are most closely related to each other, among which the Weddell seal louse (*Antarctophthirus carlinii*) and the crabeater seal louse (*Antarctophthirus lobodontis*) are sister to one another, both having very short branch length relative to that of the Australian sea lion louse (*Antarctophthirus microchir*). The *Antarctophthirus* species are more closely related to the northern fur seal louse (*Proechinophthirus fluctus*) than to the southern elephant seal louse (*Lepidophthirus macrorhini*) (Fig. 4; Supplementary Fig. 8).

The monophyly of primate lice (families Pediculidae, Pthiridae and Pedicinidae) is strongly supported and the relationships among the primate lice are resolved with strong support (Fig. 4; Supplementary Fig. 8). The gorilla louse (*Pthirus gorillae*) is most closely related to the human pubic louse (*Pthirus pubis*); the *Pthirus* species are more closely related to the *Pediculus* species of human and chimpanzee than to the *Pedicinus* species of monkeys. There is support in both Bayesian and ML trees for: 1) the primate lice to be most closely related to the pig and horse lice (family Haematopinidae); and 2) the rodent lice in the families Polyplacidae and Hoplopleuridae to be closely related. There is support in the Bayesian tree but not in the ML tree for: 1) Pthiridae, Pediculidae, Pedicinidae, Haematopinidae, Hoplopleuridae and Polyplacidae to be more closely related to each other than to Microthoraciidae or Echinophthiriidae; and 2) Echinophthiriidae to be more closely related the group that contains Pthiridae, Pediculidae, Pedicinidae, Haematopinidae, Hoplopleuridae and Polyplacidae than to Microthoraciidae (Fig. 4; Supplementary Fig. 8).

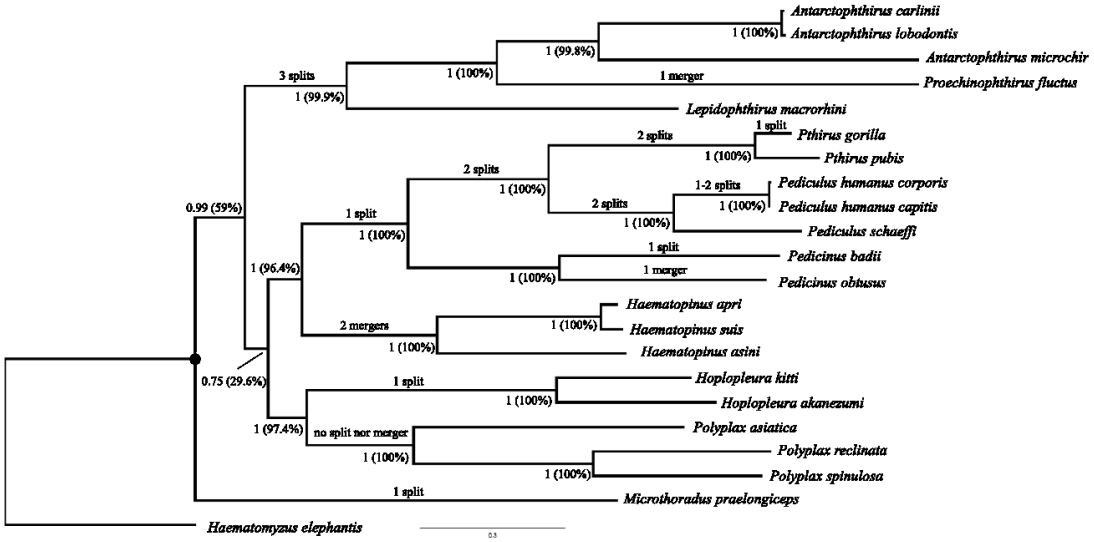


Fig. 4 Bayesian phylogenetic tree inferred from the nucleotide sequences of five mitochondrial protein-coding genes (*cob*, *cox1*, *cox2*, *cox3*, *nad4*) of 21 species of sucking lice (Anoplura). The elephant louse, *Haematomyzus elephantis*, was used as the outgroup. Bayesian posterior probability (Bpp) values were indicated near nodes followed by bootstrap support values (in brackets) from maximum likelihood (ML) tree (Supplementary Fig. 8)

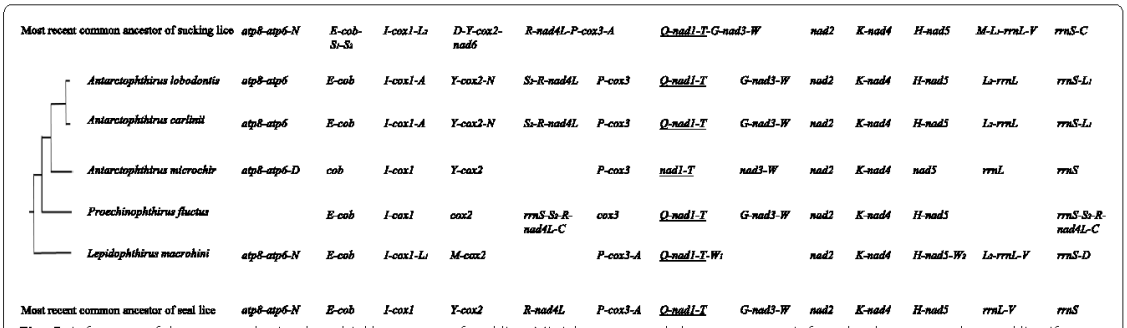


Fig. 5 Inference of the ancestral mitochondrial karyotype of seal lice. Minichromosomal characters were inferred to be ancestral to seal lice if a character was present in: 1) one or more of the five species of seal lice and also in the MRCA of sucking lice; or 2) all the five species of seal lice. Gene names were described in Fig. 1 legend. Genes underlined have opposite transcription orientation to those not underlined

Inferred ancestral mitochondrial karyotype of the seal lice

The mt minichromosomal information and the sucking louse phylogeny reported above allowed us to infer the partial mt karyotype of the most recent common ancestor of the five species of seal lice. There were at least 13 minichromosomes in the ancestral mt karyotype of these seal lice (Fig. 5). The position and arrangement of 12 of the 13 mt protein-coding genes, two rRNA genes and 14 of the 22 tRNA genes in the ancestral mt karyotype can be inferred based on the minichromosomal information available from the five species of seal lice. The position and arrangement of *nad6* and eight tRNA genes, however, cannot be inferred due to: 1) *nad6* was not identified in any of the five species of seal lice; and 2) the eight tRNA genes are very variable in their arrangement among the five species of seal lice (Fig. 5).

Discussion

Mitochondrial minichromosome split occurred more frequently than minichromosome merger in the lineages leading to seal lice and gorilla louse

Previous studies indicated that both split and merger of mt minichromosomes occurred in sucking lice and were responsible to a large degree to their highly dynamic mt genome organization [8, 12, 13]. Prior to the current study, split of mt minichromosomes was observed in species from five of the seven families of sucking lice studied: Pediculidae, Pthiridae, Pedicinidae, Microthoraciidae, and Hoplopleuridae, whereas merger of mt minichromosomes was observed in species from two families: Haematopinidae and Pedicinidae [7–9, 12]. Neither split nor merger of mt minichromosome occurred in the family Polyplacidae [8, 11, 13]. The mt genome organization of other eight families of sucking lice was unknown prior to the current study: Echinophthiriidae (lice of seals), Enderleinellidae (lice of squirrels), Linognathidae (lice of cattle, sheep, and goats), Hamophthiriidae (lice of colugos), Hybophthiridae (lice of aardvarks), Neolinognathidae (lice of elephant shrews), Pecaroecidae (lice of peccaries) and Ratemiidae (lice of horses, donkeys, and zebras).

We showed in the current study that at least three ancestral mt minichromosomes of the sucking lice [8] have split further in the lineage leading to seal lice (Echinophthiriidae): 1) *D-Y-cox2-nad6* minichromosome, 2) *R-nad4L-P-cox3-A* minichromosome, and 3) *Q-nad1-T-G-nad3-W* minichromosome (Fig. 5). In all the five species of seal lice, *cox2* has its own minichromosome not shared with any other protein-coding gene(s), indicating the split of *D-Y-cox2-nad6* minichromosome occurred in the most recent common ancestor (MRCA) of these seal lice although we could not identify *nad6* in our SRA data analyses (Fig. 5). Similarly, *R-nad4L-P-cox3-A* minichromosome and *Q-nad1-T-G-nad3-W* minichromosome also split in the MRCA of seal lice. *nad4L* has its minichromosome not shared with any other protein-coding gene(s) in *Antarctophthirus lobodontis* and *Antarctophthirus carlinii*, and *cox3* has its minichromosome not shared with any other protein-coding gene(s) in all the five species of seal lice (Fig. 5). *nad1* has its minichromosome not shared with any other protein-coding gene(s) in all the five species of seal lice; furthermore, *nad3* has its minichromosomes not shared with any other protein-coding gene(s) in *Antarctophthirus microchir* and *Proechinophthirus fluctus* (Fig. 5). We also observed a merger event in the lineage to *Proechinophthirus fluctus* between two ancestral mt minichromosomes inferred for the seal lice: *R-nad4L* and *rrnS* (Fig. 5).

The family Pthiridae has a single genus and two species: gorilla louse (*Pthirus gorillae)* and human pubic louse (*Pthirus pubis*); these two species diverged 3-4 MYA [19]. Previous studies showed that the mt genome of *Pthirus pubis* has 15 minichromosomes with *trnN* gene still not identified [6, 12]. In the current study, we showed that the mt genome of *Pthirus gorilla* has 17 minichromosomes (*trnN* not identified either) and is more fragmented than that of *Pthirus pubis* (Fig. 2). *cox3-A* minichromosome is ancestral to primate lice and is retained in *Pthirus pubis* [6]; however, this minichromosome is split into two in *Pthirus gorilla*: one minichromosome contains *cox3* and other has *trnA* (Fig. 2).

The frequency of split and merger varies among different families of sucking lice, so does the type of minichromosomes that split and merge

Shao et al. [8] proposed that split and merger of mt minichromosomes contributed to the complex and dynamic mt genome organization observed in sucking lice. Shao et al. [8] reported that: 1) split of mt minichromosomes occurred in sucking louse species from four families after their divergence from the MRCA of sucking lice: Hoplopleuridae, Pthiridae, Pediculidae and Microthoraciidae; 2) merger of mt minichromosomes occurred in species from the family Haematopinidae; and 3) no split nor merger occurred in species from the family Polyplacidae, which was also confirmed in Dong et al. [13]. Fu et al. [12] reported that both split and merger of mt minichromosomes occurred in the family Pedicinidae: split of a minichromosome occurred in the colobus louse *Pedicinus badii* whereas merger of minichromosomes occurred in the macaque louse *Pedicinus obtusus*, relative to the MRCA of higher primate lice. The other eight families of sucking lice were unknown previously for their mt genome organization: Echinophthiriidae (lice of seals), Enderleinellidae (lice of squirrels), Linognathidae (lice of cattle, sheep and goats), Hamophthiriidae (lice of colugos), Hybophthiridae (lice of aardvarks), Neolinognathidae (lice of elephant shrews), Pecaroecidae (lice of peccaries) and Ratemiidae (lice of horses, donkeys and zebras) [2, 15]. In the present study, we showed that split of three minichromosomes occurred in seal lice of the family Echinophthiriidae relative to the MRCA of sucking lice, and merger of two minichromosomes in the lineage leading to the northern fur seal louse, *Proechinophthirus fluctus* (Fig. 4). We also showed that split of a minichromosome occurred in the gorilla louse, *Pthirus gorilla*, after its divergence from the human pubic louse, *Pthirus pubis*.

It is apparent that split of minichromosomes occurs much more frequently and in more families than merger of minichromosomes in sucking lice (Fig. 4). Furthermore, the frequency of minichromosome split varies from family to family with the two families of great ape

lice (Pthiridae, Pediculidae) having the highest number of minichromosome split (8-9 split events) but no split in Haematopinidae and Polyplacidae (Fig. 4). It is noteworthy that a few ancestral minichromosomes to sucking lice split multiple times independently in lineages leading to different families. Shao et al. [8] showed that: 1) *D-Y-cox2-nad6* minichromosome split twice independently, once in Hoplopleuridae, and another time in Pediculidae and Pthiridae; and 2) *Q-nad1-T-G-nad3-W* minichromosome split twice independently, once in Microthoraciidae, and another time in Pediculidae and Pthiridae. In the current study, we found that these two minichromosomes also split independently in seal lice (Echinophthiriidae). Furthermore, *R-nad4L-P-cox3-A* minichromosome, which split in Pediculidae and Pthiridae [8], also split independently in seal lice (Echinophthiriidae).

Much rarer than minichromosome split, minichromosome merger was seen only in species from three families: pig lice and horse louse (Haematopinidae) [8], macaque louse (Pedicinidae) [12], and northern fur seal louse (Echinophthiriidae). In the pig lice and horse louse, two ancestral minichromosomes to sucking lice, *atp8-atp6-trnN* and *trnK-nad4* merged; two other ancestral minichromosomes, *trnI-cox1-trnL*$_2$ and *nad2* also merged [8]. In the macaque louse, a single merge event occurred between two ancestral minichromosomes to higher primate lice that contain *cox2* and *nad2* genes respectively. In the northern fur seal louse, a single merger event occurred between two minichromosomes ancestral to seal lice that contain *nad4L* and *rrnS* genes respectively (Fig. 5). It is noteworthy that *nad2* minichromosome was involved in two separate merger events whereas the other five minichromosomes were each involved only in a single merger event.

The very high mitochondrial gene identity shared between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* does not support them as separate species

The crabeater seal louse, *Antarctophthirus lobodontis*, was described by Enderlein [20] and redescribed by Leonardi et al. [21]; the Weddell seal louse, *Antarctophthirus carlinii*, was described by Leonardi et al. [22]. These two species are very similar in morphology [21, 22]. Leonardi et al. [21] provided two morphological characters to differentiate between these two species: 1) *Antarctophthirus carlinii* has six dorsal posterior long hairs (four marginal and two principal) around the posterior border of the head while *Antarctophthirus lobodontis* has only four marginal long hairs around the posterior border of the head; and 2) *Antarctophthirus lobodontis* has a line of eight spines in the basis of the head and three hairs above the last row of four spines. As reported above, *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* have identical mt karyotypes (Fig. 1C), indicating a very close relationship between them. Furthermore, *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* share very high identity of mt genes ranging from 94.3 to 100% (average 99.03%) (Supplementary Table 8). We also obtained and compared the partial sequences of three nuclear genes between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*. The sequences of their *elongation factor 1-α* gene (182 bp) are 100% identical; their *18S* sequences (1054 bp) and *28S* rRNA gene sequences (2295 bp) have 99.3 and 99.1% identities respectively (Supplementary Fig. 9). Previous studies showed that mt genes of parasitic lice evolved much faster than their hosts. Thus, the identities of mt genes are much lower between parasitic lice than between their hosts [23–26]. This is, however, not the case for *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*. The mt gene identities between these lice are 4.11% higher on average than that between their hosts, the Weddell seal (*Leptonychotes weddellii*) and the crabeater seal (*Lobodon carcinophagus*) - 90.8 to 100% identities with an average of 94.99% (Supplementary Table 9). The low genetic divergence between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* does not support them as separate species; more likely they are two subspecies instead of two species. Leonardi et al. [18] also discussed the possibility that *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* were the same species based on the low genetic divergence between them. However, based on cophylogenetic analysis, Leonardi et al. [18] concluded that *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* co-speciated with their hosts thus were different species; this conclusion was not supported by the lower genetic divergence between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* than between their hosts. A relatively recent host switch between the Weddell seal and the crabeater seal is a more plausible explanation for the low genetic divergence observed between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* than co-speciation of *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* with their hosts [18]. Both Weddell seals and crabeater seals have circumpolar distributions in the Antarctic and share the same microhabitats such as land-fast ice and pack ice when they rest and breed [27, 28]. This could provide ample opportunities for the parasitic lice of one seal species to explore and switch to another seal species. It is unclear to us, however, which seal is the original host and which seal is the new host. One possibility is that the host switch occurred from crabeater seal to Weddell seal based on the observation that crabeater seal hosts only one louse species (*Antarctophthirus lobodontis*) whereas Weddell seal hosts two louse species (*Antarctophthirus carlinii* and *Antarctophthirus*

ogmorhini) [2]. Both crabeater seal and Weddell seal are in the tribe Lobodontini (subfamily Monachinae), together with Ross seal (*Ommatophoca rossii*) and leopard seal (*Hydrurga leptonyx*) [2]. Weddell seal is the only species in the tribe Lobodontini that hosts two louse species; the other three seal species each host one louse species only [2]. Further studies on the phylogeny of all louse species of Lobodontini seals should reveal more on the host-parasite relationships in this tribe.

Conclusion

In this study, we assembled the mt genomes of five species of seal lice and the gorilla louse, and conducted phylogenetic analysis of sucking lice from eight families. We inferred the ancestral mt karyotype of seal lice and analyzed the frequency of mt minichromosomal split and merger among sucking lice. Mt karyotype comparison, gene sequence analysis and phylogenetic analysis all indicated that the crabeater seal louse, *Antarctophthirus lobodontis*, and the Weddell seal louse, *Antarctophthirus carlinii*, are likely two subspecies instead of two species, and the possibility of a host switch of crabeater seal louse to Weddell seals. We showed that at least three ancestral mt minichromosomes of sucking lice have split in the lineage leading to seal lice, one minichromosome ancestral to primate lice has split in the lineage leading to gorilla louse, and two ancestral minichromosomes of seal lice have merged in the lineage to the northern fur seal louse. Split of mt minichromosomes occurred 15-16 times in total in the lineages leading to six families of sucking lice studied so far whereas merger of minichromosomes occurred only four times in the lineages leading to three families of sucking lice. Furthermore, three ancestral minichromosomes to sucking lice, *D-Y-cox2-nad6*, *Q-nad1-T-G-nad3-W* and *R-nad4L-P-cox3-A*, have split independently in different lineages of sucking lice. We conclude that: 1) minichromosome split contributes much more than minichromosome merger in mt genome fragmentation of sucking lice, and 2) mt karyotype comparison, in conjunction with gene sequence analysis, helps understand the relationship between sucking louse species.

Materials and methods

Retrieval and assembly of Sequence Read Archive (SRA) data of seal lice and gorilla louse

The SRA data of five species of seal lice (*Antarctophthirus carlinii*; *Antarctophthirus lobodontis*, *Antarctophthirus microchir*, *Lepidophthirus macrorhini* and *Proechinophthirus fluctus*) and the gorilla louse (*Pthirus gorillae*) were retrieved from NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra/). These SRA data were produced with Illumina platforms in whole genome sequencing projects and were deposited by researchers in Kevin Johnson's group at the University of Illinois, Urbana-Champaign [18]. The whole dataset of sequence-reads of each species was imported into and assembled in Geneious 11.0.2 [29]. A subset of 15,000 sequence reads in each dataset was extracted and assembled de novo to obtain seed contigs for mt gene search; the number of contigs to be displayed was set at 100. The consensus sequences of these 100 contigs were searched in a batch in NCBI "Non-redundant protein sequences (NR)" database using BLASTx and default parameters to identify protein gene sequence matches [30]. The consensus sequences that matched significantly (E-value $< 10^{-15}$) to the mt protein sequences of sucking lice (parvorder Anoplura) in the NR database were used as reference sequences to assemble the full-length coding region and its adjacent non-coding regions (200-320 bp upstream and downstream) of each mt minichromosome; the entire set of SRA sequence-reads of each species was explored in coding and non-coding region assembly. We did not attempt to assemble the full-length non-coding region of each minichromosome with SRA data because: 1) the non-coding region is highly similar in sequence among different mt minichromosomes of a sucking louse species [5–13], and 2) the SRA Illumina sequence reads are too short (100-160 bp each) for such attempt. The key assembly parameters were: 1) minimum overlap 60 bp for *Proechinophthirus fluctus* (SRA sequence reads 100 bp each) and 100 bp for other four species of seal lice (SRA sequence reads 150 bp each) and gorilla louse (SRA sequence reads 160 bp each); and 2) minimum identity 95%. Because of the high sequence similarity among the non-coding regions of different minichromosomes within a species, once two or more minichromosomes were assembled, the conserved sequence motifs including the hallmark AT-rich motif and GC-rich motif in the non-coding regions were identified by sequence alignment and used as references to identify and assemble the remaining minichromosomes. An AT-rich motif upstream from coding region and a GC-rich motif downstream from coding region are present in the mt minichromosomes of all sucking lice and the chewing lice of eutherian mammals sequenced to date [5–14, 31, 32]. The steps described above were repeated multiple times for each species until no more additional mt genes could be found. The protein-coding genes and rRNA genes in mt minichromosomes were identified by BLAST search in NCBI database; tRNA genes were identified by tRNA-scan [33] and ARWEN [34].

For *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*, their partial *elongation factor 1-α* gene sequence, *18S* and *28S* rRNA gene sequences were generated by Illumina sequence read assembly, using the available sequences of *Haematopinus eurysternus* (accession

numbers HM171457 and HM171381) and *Echinophthirus horridus* (accession number KX810111) as initial reference sequences. We used Geneious 11.0.2 [29] and started with *medium sensitivity* in reference assembly to find the most conserved regions of these three genes. Then we used the SRA sequence reads that were mapped to the conserved regions as references to initiate and extend the contigs for *elongation factor 1-α*, *18S* rRNA and *28S* rRNA genes of *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*. These genes were then verified by BLAST search in NCBI database.

Phylogenetic analyses

Twenty-two species of parasitic lice (Supplementary Table 10) were included in our phylogenetic analysis: 1) the five species of seal lice and the gorilla louse reported in the present study; 2) 15 sucking louse species reported in previous studies; and 3) the elephant louse (*Haematomyzus elephantis*) as the outgroup. The sequences of five mt protein-coding genes (*cob*, *cox1*, *cox2*, *cox3*, *nad4*) were used in our phylogenetic analysis. These five genes were aligned individually using MAFFT 7.471, then concatenated into a single file after removing the poorly aligned sites using Gblocks 0.91b. Two methods were used in our phylogenetic analysis: 1) maximum likelihood (ML) with IQ-Tree [35], and 2) Bayesian inference method (BI) with MrBayes 3.2.6 [36]. Model test were done in IQ-TREE [37] and the best-fit model is TIM + F + R4. For ML analysis, the bootstrap replicates were set at 1000. For BI analyses, four independent Markov Chains were run for 5 million MCMC generations, sampling a tree every 100 generations. This analysis was run until the average standard deviation of split frequencies was lower than 0.001. The ML tree and BI tree were drawn with Figtree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

Inferring the ancestral mitochondrial karyotype of seal lice

To infer the ancestral mt karyotype of seal lice, we used a parsimony method described in Shao et al. [8]. The ancestral mt karyotype of seal lice was inferred based on a comparison of mt minichromosomal characters between the five species of seal lice and the inferred most recent common ancestor (MRCA) of sucking lice. We inferred a minichromosomal character to be ancestral to seal lice if: 1) the character was present in one or more of the five species of seal lice and also in the MRCA of sucking lice; or 2) the character is present in all the five species of seal lice. For example, the character *E-cob* in a minichromosome is present in four of the five seal louse species and also in the MRCA of sucking lice (Fig. 5). Thus, *E-cob* in a minichromosome is inferred to be ancestral to seal lice. The MRCA of sucking lice also has S_1-S_2 in the same minichromosome with *E-cob*; these two tRNA genes are most likely translocated to other minichromosome(s) in the MRCA of seal lice as they are not with *E-cob* in any of the five seal louse species (Fig. 5). Taking another example, *K-nad4* in a minichromosome is present in all the five seal louse species thus it is inferred to be ancestral to seal lice (Fig. 5). In this case, *K-nad4* is also present in the MRCA of sucking lice, which reinforces the inference of *K-nad4* minichromosome to be ancestral to seal lice.

Abbreviations

mt: Mitodhindrial; *atp6* and *atp8*: Genes for ATP synthase subunits 6 and 8; bp: Base pair; *cob*: Gene for cytochrome b; *cox1*, *cox2* and *cox3*: Genes for cytochrome c oxidase subunits 1, 2 and 3; DNA: Deoxyribonucleic acid; kb: Kilo base pair; min: Minute; MRCA: Most recent common ancestor; mt: Mitochondrial; Mya: Million years ago; *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*: Mitochondrial genes for NADH dehydrogenase subunits 1-6 and 4 L; PCR: Polymerase chain reaction; RNA: Ribonucleic acid; rRNA: Ribosomal RNA; *rrnS* and *rrnL*: Genes for small and large subunits of ribosomal RNA; sec: Second; T: Thymine; tRNA: Transfer RNA; *trnA* or *A*: tRNA gene for alanine; *trnC* or *C*: tRNA gene for cysteine; *trnD* or *D*: tRNA gene for aspartic acid; *trnE* or *E*: tRNA gene for glutamic acid; *trnF* or *F*: tRNA gene for phenylalanine; *trnG* or *G*: tRNA gene for glycine; *trnH* or *H*: tRNA gene for histidine; *trnI* or *I*: tRNA gene for isoleucine; *trnK* or *K*: tRNA gene for lysine; $trnL_1$ or L_1: tRNA gene for leucine (anticodon NAG); $trnL_2$ or L_2: tRNA gene for leucine (anticodon YAA); *trnM* or *M*: tRNA gene for methionine; *trnN* or *N*: tRNA gene for asparagine; *trnP* or *P*: tRNA gene for proline; *trnQ* or *Q*: tRNA gene for glutamine; *trnR* or *R*: tRNA gene for arginine; $trnS_1$ or S_1: tRNA gene for serine (anticodon NCU); $trnS_2$ or S_2: tRNA gene for serine (anticodon NGA); *trnT* or *T*: tRNA gene for threonine; *trnV* or *V*: tRNA gene for valine; *trnW* or *W*: tRNA gene for tryptophan; *trnY* or *Y*: tRNA gene for tyrosine; U: Uracil.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08530-8.

Additional file 1: Supplementary Fig. 1. Secondary structure inferred with tRNA-Scan [33] from *trnW1* and *trnW2* gene sequences of the southern elephant seal louse, *Lepidophthirus macrorhini*. **Supplementary Fig. 2.** Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the southern elephant seal louse, *Lepidophthirus macrorhini*. **Supplementary Fig. 3.** Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the northern fur seal louse, *Proechinophthirus fluctus*. **Supplementary Fig. 4.** Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the Weddell seal louse, *Antarctophthirus carlinii*. **Supplementary Fig. 5.** Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the crabeater seal louse, *Antarctophthirus lobodontis*. **Supplementary Fig. 6.** Conserved non-coding GC-rich motifs among the mitochondrial minichromosomes of the Australian sea lion louse, *Antarctophthirus microchir*. **Supplementary Fig. 7.** Conserved non-coding AT-rich motifs and GC-rich motifs among the mitochondrial minichromosomes of the gorilla louse, *Pthirus gorillae*. **Supplementary Fig. 8.** Phylogenetic relationships among 21 species of sucking lice (Anoplura) inferred by maximum likelihood (ML) analysis of nucleotide sequences of five mitochondrial protein-coding genes. The elephant louse, *Haematomyzus elephantis*, was used as the outgroup. The ultrafast bootstrap support (%) / SH-aLRT support (%) were indicated near each node. **Supplementary Fig. 9.** Alignment of partial *18S* rRNA gene, *28S* rRNA gene and *ef1-α* gene sequences between *Antarctophthirus carlinii* and *Antarctophthirus lobodontis*. **Supplementary Table 1.** Mitochondrial minichromosomes of *Lepidophthirus macrorhini* - louse of southern elephant seal (*Mirounga leonine*).

Supplementary Table 2. Sequence identities between *trnW1* and *trnW2* of the southern elephant seal louse, *Lepidophthirus macrorhini*, between *trnW* genes of *Lepidophthirus macrorhini* and other seal lice, and between *trnW* genes and other tRNA genes of *Lepidophthirus macrorhini*. Identities were generated with ClustalW in Geneious [29]: cost matrix IUB, gap open cost 15, gap extend cost 6.66. **Supplementary Table 3.** Mitochondrial minichromosomes of *Proechinophthirus fluctus* - louse of northern fur seal (*Callorhinus ursinus*). **Supplementary Table 4.** Mitochondrial minichromosomes of *Antarctophthirus carlinii* - louse of Weddell seal (*Leptonychotes weddelli*). **Supplementary Table 5.** Mitochondrial minichromosomes of *Antarctophthirus lobodontis* - louse of crabeater seal (*Lobodon carcinophagus*). **Supplementary Table 6.** Mitochondrial minichromosomes of *Antarctophthirus microchir* – louse of Australian sea lion (*Neophoca cinerea*). **Supplementary Table 7.** Mitochondrial minichromosomes of *Pthirus gorilla* – louse of western gorilla (*Gorilla gorilla*). **Supplementary Table 8.** Sequence identities between *Antarctophthirus carlinii* (louse of Weddell seal, *Leptonychotes weddelli*) and *Antarctophthirus lobodontis* (louse of crabeater seal, *Lobodon carcinophagus*). **Supplementary Table 9.** Sequence identities between Weddell seal (*Leptonychotes weddelli*) and crabeater seal (*Lobodon carcinophagus*). **Supplementary Table 10.** Species of parasitic lice included in the phylogenetic analyses in this study.

Acknowledgements
The publicly available SRA data used in this study were produced and deposited by researchers in Kevin Johnson's group at the University of Illinois at Urbana-Champaign. We thank Wei Wang for assistance in checking the morphological characters of *Antarctophthirus carlinii* and *Antarctophthirus lobodontis* in published literature. We thank the associate editor and the reviewers for valuable comments that have improved this manuscript greatly.

Authors' contributions
YD and RS designed the research, analyzed the data and wrote the manuscript. MZ assisted YD with SRA data retrieval and edited the manuscript. All authors have read and approved the final manuscript.

Funding
This study was funded by the Australian Biological Resources Study (ABRS, RF217-51).

Availability of data and materials
Sequence data generated in this study are available in NCBI (accession numbers MW803064-803131).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests to declare.

Author details
[1]Centre for Bioinnovation, University of the Sunshine Coast, 90 Sippy Downs Drive, Sippy Downs, Queensland 4556, Australia. [2]School of Science, Technology and Engineering, University of the Sunshine Coast, 90 Sippy Downs Drive, Sippy Downs, Queensland 4556, Australia.

Received: 30 November 2021 Accepted: 28 March 2022
Published online: 08 April 2022

References

1. De Moya RS, Yoshizawa K, Walden KK, Sweet AD, Dietrich CH, Kevin P, et al. Phylogenomics of parasitic and nonparasitic lice (Insecta: Psocodea): combining sequence data and exploring compositional bias solutions in next generation data sets. Syst Biol. 2021;70(4):719–38.
2. Durden LA, Musser GG. 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; 1994.
3. Kim KC, Ludwig HW. The family classification of the Anoplura. Syst Entomol. 1978;3(3):249–84.
4. Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999;27(8):1767–80.
5. 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(5):904–12.
6. Shao R, Zhu X-Q, Barker SC, Herd K. Evolution of extensively fragmented mitochondrial genomes in the lice of humans. Genome Biol Evol. 2012;4(11):1088–101.
7. 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.
8. 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(2):431–45.
9. Jiang H, Barker SC, Shao R. Substantial variation in the extent of mitochondrial genome fragmentation among blood-sucking lice of mammals. Genome Biol Evol. 2013;5(7):1298–308.
10. Dong WG, et al. 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. 2014a;15:751.
11. Dong WG, et al. 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. 2014b;15(1):44.
12. Fu YT, Dong W, et al. Fragmented mitochondrial genomes evolved in opposite directions between closely related macaque louse Pedicinus obtusus and colobus louse Pedicinus badii. Genomics. 2020;112(6):4924–33.
13. Dong WG, Dong Y, Guo XG, Shao R. Frequent tRNA gene translocation towards the boundaries with control regions contributes to the highly dynamic mitochondrial genome organization of the parasitic lice of mammals. BMC Genomics. 2021;22(1):598.
14. 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 Genomics. 2015;16(1):661.
15. Kim KC. Coevolution of parasitic arthropods and mammals. New York: Wiley; 1985.
16. Kim K, Repenning C, Morejohn G. Specific antiquity of the sucking lice and evolution of otariid seals. Rapports Proces Verbaux Reun. 1975;169:544–9.
17. Kim KC. Ecology and morphological adaptation of the sucking lice (Anoplura, Echinophthiriidae) on the northern fur seal, vol. 169: Rapport et Procès verbaux des Réunions du conseil Permanent International pour lExploration de la Mer. Copenhague: Le Conseil, 1903-1989; 1975. p. 504–15.
18. Leonardi MS, et al. Phylogenomic analysis of seal lice reveals codivergence with their hosts. Syst Entomol. 2019;44(4):699–708.
19. Reed DL, Light JE, Allen JM, Kirchman JJ. Pair of lice lost or parasites regained: the evolutionary history of anthropoid primate lice. BMC Biol. 2007;5:7.
20. Enderlein G, J. D. I. d. A. G. D. S. E. Monographie der Robbenläuse 2; 1909. p. 505–16.
21. Leonardi MS, Bobinac M, Negrete J. Redescription of *Antarctophthirus lobodontis* (Anoplura: Echinophthiriidae) from the crabeater seal and identification key for Antarctic lice. Polar Biol. 2016;39(4):671–6.
22. Leonardi MS, et al. *Antarctophthirus carlinii* (Anoplura: Echinophthiriidae), a new species from the Weddell seal *Leptonychotes weddelli*. Parasitol Res. 2014;113(11):3947–51.
23. Hafner MS, et al. Disparate rates of molecular evolution in cospeciating hosts and parasites. Science. 1994;265(5175):1087–90.

24. Page RD, et al. A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. Mol Phylogenet Evol. 1998;9(2):276–93.
25. Paterson AM, Wallis GP, Wallis LJ, Gray RD. Seabird and louse coevolution: complex histories revealed by 12S rRNA sequences and reconciliation analyses. Syst Biol. 2000;49(3):383–99.
26. Johnson KP, et al. Dramatically elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera). Mol Phylogenet Evol. 2003;26(2):231–42.
27. Würsig B, Perrin WF, Thewissen J. Encyclopedia of marine mammals: Academic Press; 2009.
28. Southwell C, et al. A review of data on abundance, trends in abundance, habitat use and diet of ice-breeding seals in the Southern Ocean. Ccamlr Sci. 2012;19:49–74.
29. Kearse M, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
30. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004;32(suppl_2):W20–5.
31. Shao R, et al. Fragmented mitochondrial genomes in two suborders of parasitic lice of eutherian mammals (Anoplura and Rhynchophthirina, Insecta). Sci Rep. 2015;5:17389.
32. Song F, et al. Mitochondrial genome fragmentation unites the parasitic lice of Eutherian mammals. Syst Biol. 2019;68(3):430–40.
33. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64.
34. Laslett D, Canback B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics. 2008;24(2):172–5.
35. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2014;32(1):268–74.
36. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572–4.
37. Kalyaanamoorthy S, et al. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

Appendix IV

Ethics approval letter

9 November 2020

A/Prof Andrew Crowden
Chair, Animal Ethics Committee
Tel: +61 7 5430 2823
Email: animalethics@usc.edu.au

Dr Renfu Shao
Dr Dominique Potvin
Mr Wei Wang
Mr Yalun Dong

Dear Investigators

Animal ethics approval for project: Collecting parasitic lice of birds for molecular and evolutionary studies (ANA19154)

This letter is to confirm that the USC Animal Ethics Committee has reviewed and granted ethics approval of this project subject to the standard conditions listed below.

The period of ethics approval is from 9 November 2020 to 31 December 2022. The ethics approval number for the project is ANA19154.

Ethics approval indicates that this project meets the requirements of the National Health and Medical Research Council's (NHMRC) *Australian code for the care and use of animals for scientific purposes, 8th edition (2013)*. This does not negate the need for other approvals where relevant. It is the investigators' responsibility to ensure that all approvals relevant to this project are obtained.

If you have any queries in relation to this or if you require further information please contact us using the details above.

Yours sincerely

A/Prof Andrew Crowden
Chair, Animal Ethics Committee

STANDARD CONDITIONS OF ETHICS APPROVAL

The standard conditions of approval for all animal teaching or research projects are the following:

1) Comply with the *Australian code for the care and use of animals for scientific purposes, 8th edition (2013)*, and any legislation relevant to the juristiction that this project is conducted.

2) Conduct the research project strictly in accordance with the application which was granted ethics approval, including any conditions nominated in the ethics approval correspondence.

3) Make no change to an ethics approved project, including any approved documentation wording, without prior written approval from the Animal Ethics Committee.

4) Inform the Animal Ethics Committee immediately of anything which may warrant further review of the ethical acceptability of the project, including serious or unexpected adverse events or impacts on animals. An unexpected adverse event report must be submitted no later than one working day after an unexpected adverse event occours.

5) Provide the Animal Ethics Committee with a copy of all licences, permits and approvals related to the conduct of this project.

6) Comply with the monitoring conditions of the project, including the maintinance of animal monitoring records and the conduct of inspections by the Animal Ethics Committee (or representative) throughout the project.

7) Provide the Animal Ethics Committee with a written progress report for each calendar year that ethics approval is active and on completion of the project. Monitoring records will be requested as part of this reporting process.

8) Advise the Animal Ethics Committee in writing if the project is discontinued.

Compliance with these conditions is a requirement of USC's Animal Ethics – Governing Policy and the *Australian code for the care and use of animals for scientific purposes, 8th edition (2013).*

USC Animal ethics policies, procedures, guidelines, and forms are available from Blackboard via Organisations > Research > Research Ethics & Integrity > Animal Ethics; and MyUSC via Research > Research Ethics, Integrity and Compliance > Animal Ethics.

Appendix V

Images of lice and their hosts

Actornithophilus grandiceps (host: pied oystercatcher, *Haematopus longirostris*)

Actornithophilus hoplopteri (host: masked lapwing, *Vanellus miles*)

Austromenopon atrofulvum (host: sooty tern, *Onychoprion fuscatus*)

Austromenopon atrofulvum (host: crested tern, *Thalasseus bergii*)

Austromenopon paululum (host: sooty shearwater, *Ardenna grisea*)

Ciconiphilus decimfasciatus (host: white-faced heron, *Egretta novaehollandiae*)

Colpocephalum eucarenum (host: Australian Pelican, *Pelecanus conspicillatus*)

Colpocephalum spinicollis (host: straw-necked ibis, *Threskiornis spinicollis*)

Eomenopon denticulatum (host: scaly-breasted lorikeet, *Trichoglossus chlorolepidotus*)

Franciscoloa funerei (host: yellow-tailed black cockatoo, *Calyptorhynchus funereus*)

Franciscoloa pallida (host: sulfur-crested cockatoo, *Cacatua galerita*)

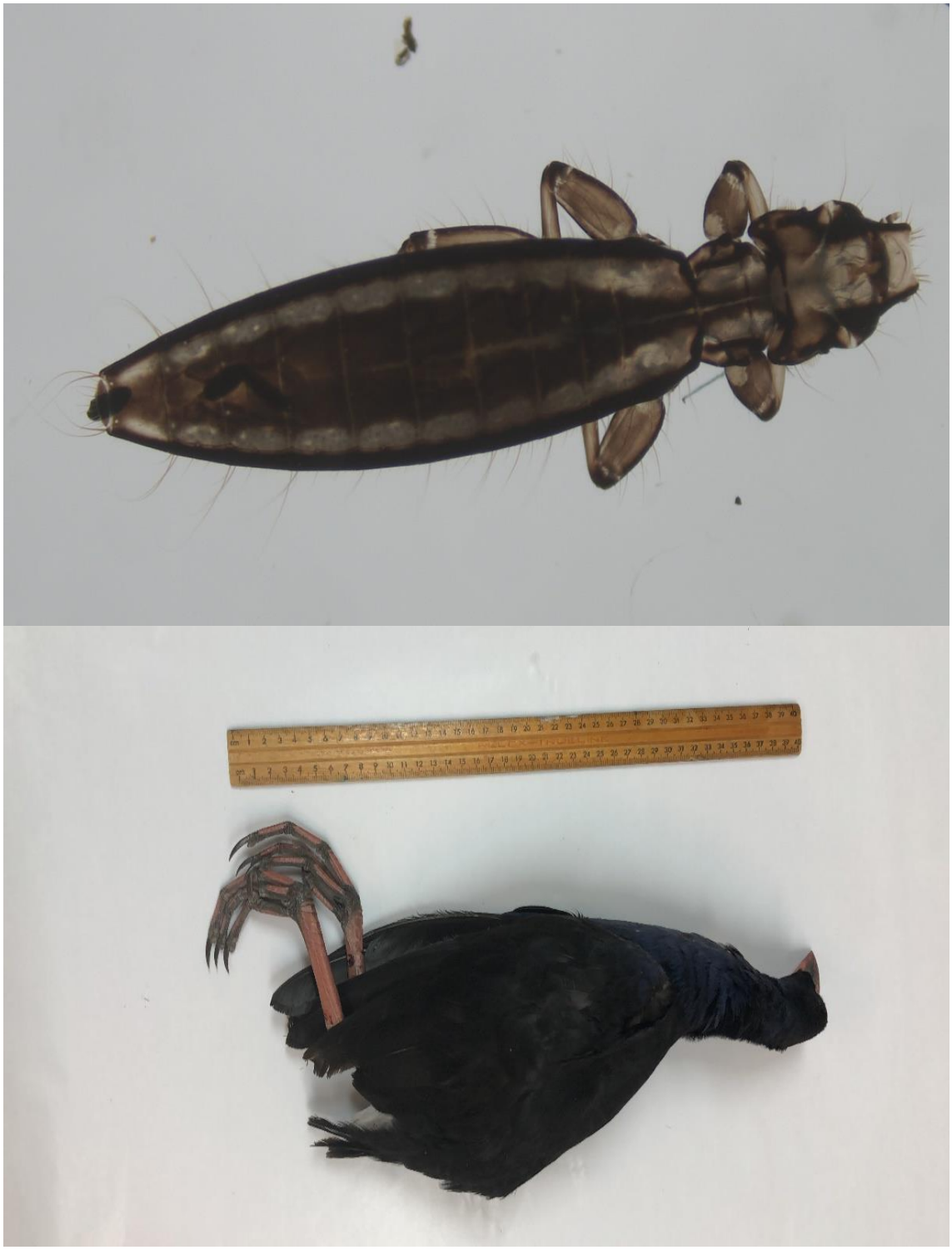
Franciscoloa roseicapillae (host: little corella, *Cacatua sanguinea*)

Franciscoloa roseicapillae (host: galah, *Eolophus roseicapilla*)

Franciscoloa roseicapillae (host: pheasant coucal, *Centropus phasianinus*)

Laemobothrion atrum (host: Eurasian coot, *Fulica atra*)

Laemobothrion maximum (host: black kite, *Milvus migrans*)



Laemobothrion sp. (host: Australian swamphen, *Porphyrio porphyrio bellus*)

Myrsidea ptilorhynchi (host: satin bowerbird, *Ptilonorhynchus violaceus*)

Piagetiella africana (host: Australian pelican, *Pelecanus conspicillatus*)