

The first complete mitochondrial genome from the subfamily Keroplatinae (Diptera: Keroplatidae): genome structure and comparative analysis

Hanhuiying LU^{1,2}, Yilin YAO^{2,3}, Yanli ZHANG^{2,3}, Zehui KANG^{2①}, Ding YANG³, Xiao ZHANG^{2①}

1. Key Laboratory of National Forestry and Grassland Administration on Grassland Resources and Ecology in the Yellow River Delta, College of Grassland Science, Qingdao Agricultural University, Qingdao, Shandong 266109, China

2. Shandong Engineering Research Center for Environment-Friendly Agricultural Pest Management, College of Plant Health and Medicine, Qingdao Agricultural University, Qingdao, Shandong 266109, China

3. College of Plant Protection, China Agricultural University, Beijing 100193, China

Abstract: Mitochondrial (mt) genomes have been widely used in the phylogenetic study of many insect groups. There is only one complete mt genome of a member of the family Keroplatidae available in GenBank and it belongs to the subfamily Arachnocampinae. Here, we report the first mt genome of a member of the subfamily Keroplatinae in the family Keroplatidae. It is a circular molecule of 15,521 bp with an AT content of 78.9%. This mt genome contains 13 protein coding genes (PCGs), 22 transfer RNA genes (tRNAs), two ribosomal RNA genes (rRNAs) and a control region. The AT content of the whole and each part of this mt genome of Keroplatinae (*Orfelia* sp.) are higher than that in Arachnocampinae (*Arachnocampa flava* Harrison, 1966), except for the control region. The ratio (dN/dS) among Keroplatidae flies shows that *ND5* has the fastest evolutionary rate, while *COI* has the slowest evolutionary rate.

Key words: mitogenome; fungus gnats; *Orfelia*

扁角菌蚊亚科首个完整线粒体基因组——结构和比较分析（双翅目：扁角菌蚊科）

吕韩慧颖^{1,2}, 姚祎琳^{2,3}, 张艳丽^{2,3}, 康泽辉^{2①}, 杨定³, 张晓^{2①}

1. 青岛农业大学草业学院, 黄河三角洲草地资源与生态国家林业和草原局重点实验室, 山东 青岛 266109; 2. 青岛农业大学植物医学学院, 山东省植物病虫害绿色防控工程研究中心, 山东 青岛 266109; 3. 中国农业大学植物保护学院, 北京 100193

摘要: 线粒体基因组已广泛应用于许多昆虫类群的系统发育研究。GenBank 中只有 1 个完整的扁角菌蚊科线粒体基因组, 属于织网菌蚊亚科。在此, 我们公布了扁角菌蚊科另 1 个亚科——扁角菌蚊亚科的第 1 个线粒体基因组, 它是 1 个 15,521 bp 的环状分子, AT 含量为 78.9%。该线粒体基因组包含 13 个蛋白质编码基因、22 个转运 RNA 基因、两个核糖体 RNA 基因和一个控制区。除控制区外, 扁角菌蚊亚科(即 *Orfelia* sp.)线粒体基因组的整体和各部分的 AT 含量均高于织网菌蚊亚科(即 *Arachnocampa flava* Harrison, 1966)。核苷酸替代率 (dN/dS) 分析表明, 扁角菌蚊科线粒体基因组中 *ND5* 基因的进化速度最快, 而 *COI* 基因的进化速度最慢。

关键词: 线粒体基因; 菌蚊; 沃菌蚊属

Accepted 6 June 2023. Published online 18 November 2023. Published 25 December 2023.

① Corresponding authors, E-mails: kangzehui1987@163.com; xzhang_cn@163.com

Introduction

The infraorder Bibionomorpha is a megadiverse group in Nematocera. This is due to the inclusion of the fungus gnats which are comprised of the very large families Keroplatidae, Mycetophilidae and Sciaridae, and the gall midges, the most diverse and species-rich family of Diptera (Ševčík *et al.* 2016). Due to the significance of these groups, the composition of Bibionomorpha families and their relationships to other lower Diptera have been studied many times based on morphological and molecular data, especially mitochondrial (mt) genome data (Beckenbach 2012; Miao *et al.* 2020; Wang *et al.* 2021; Zhang *et al.* 2023).

Keroplatidae is a rather large and cosmopolitan family in the infraorder Bibionomorpha. But there is only one complete mt genome of the family available in GenBank, which belongs to the subfamily Arachnocampinae. In this study, we sequenced and analyzed the mt genome of a member of the genus *Orfelia* Costa, 1857, a fungus gnat. This represents the first complete mt genome of the subfamily Keroplatinae and will be an important addition to explore the phylogeny of Bibionomorpha.

Material and methods

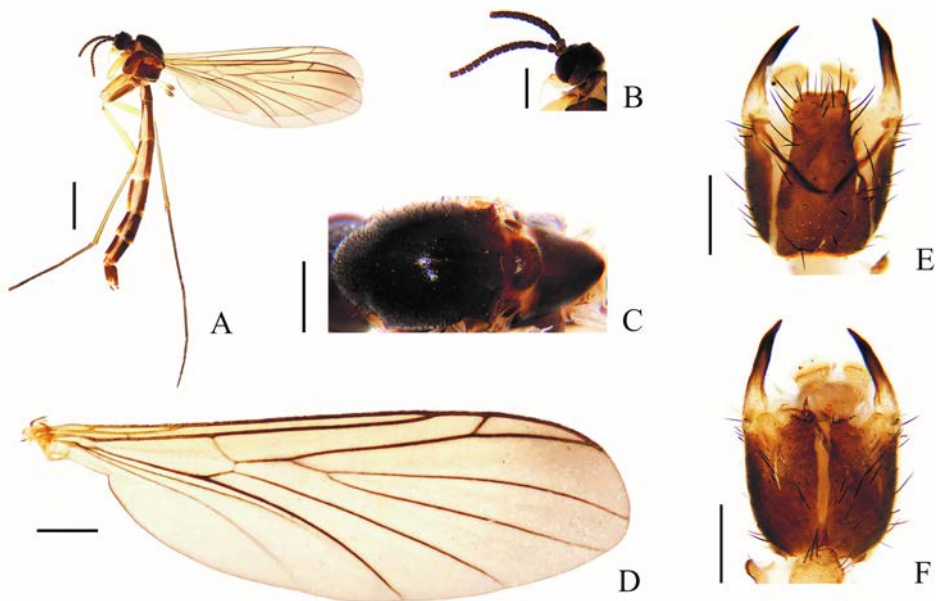


Figure 1. *Orfelia* sp. A. Habitus of male, lateral view; B. Head, lateral view; C. Thorax, dorsal view; D. Wing; E. Male hypopygium, dorsal view; F. Male hypopygium, ventral view. Scale bars = 1.0 mm (A); 0.3 mm (B, C); 0.5 mm (D); 0.2 mm (E, F).

The specimen used in this study was collected from Cirencuo, Derong, Sichuan, China (28°21'12"N 99°22'22"E; 4105 m) and is deposited at the Entomological Museum of Qingdao Agricultural University, China (KER0001). This specimen was identified as belonging to the genus *Orfelia* Costa, 1857 according to Cao (2007) and Cao *et al.* (2008), but it is different from all known species of this genus in China. An illustration of this species (Fig. 1) is provided for further research. Total genomic DNA was extracted from the muscle tissues of

the specimen using the TIANamp Genomic DNA Kit (TIANGEN, Beijing, China). The mt DNA fragments of the specimen were amplified using standard primers conserved across insects (Simon *et al.* 1994) and the missing sequences were amplified using designed primers based on these known nucleotide fragments. Fragments were assembled using MEGA7 (Kumar *et al.* 2016). The complete sequence was annotated by MITOS WebServer (Bernt *et al.* 2013) with checking for predicted transfer RNA genes (tRNAs) and protein-coding genes (PCGs). All the predicted tRNAs were confirmed using the tRNAscan-SE search server (Lowe & Chan 2016). The boundary of the PCGs and the ribosomal RNA genes (rRNAs) were confirmed based on BLASTn [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome] results compared against published sequences of nematoceran species.

Results and discussion

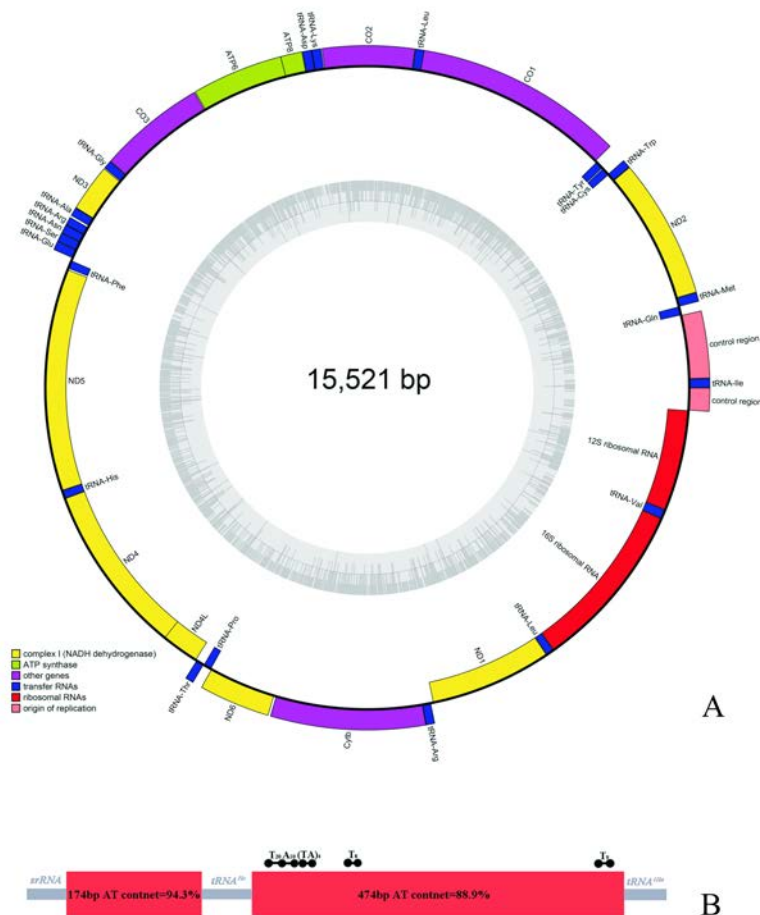


Figure 2. Mitochondrial genome of *Orfelia* sp. A. Gene map; B. Control region. The circular maps were drawn with Geneious v.10.2.2.

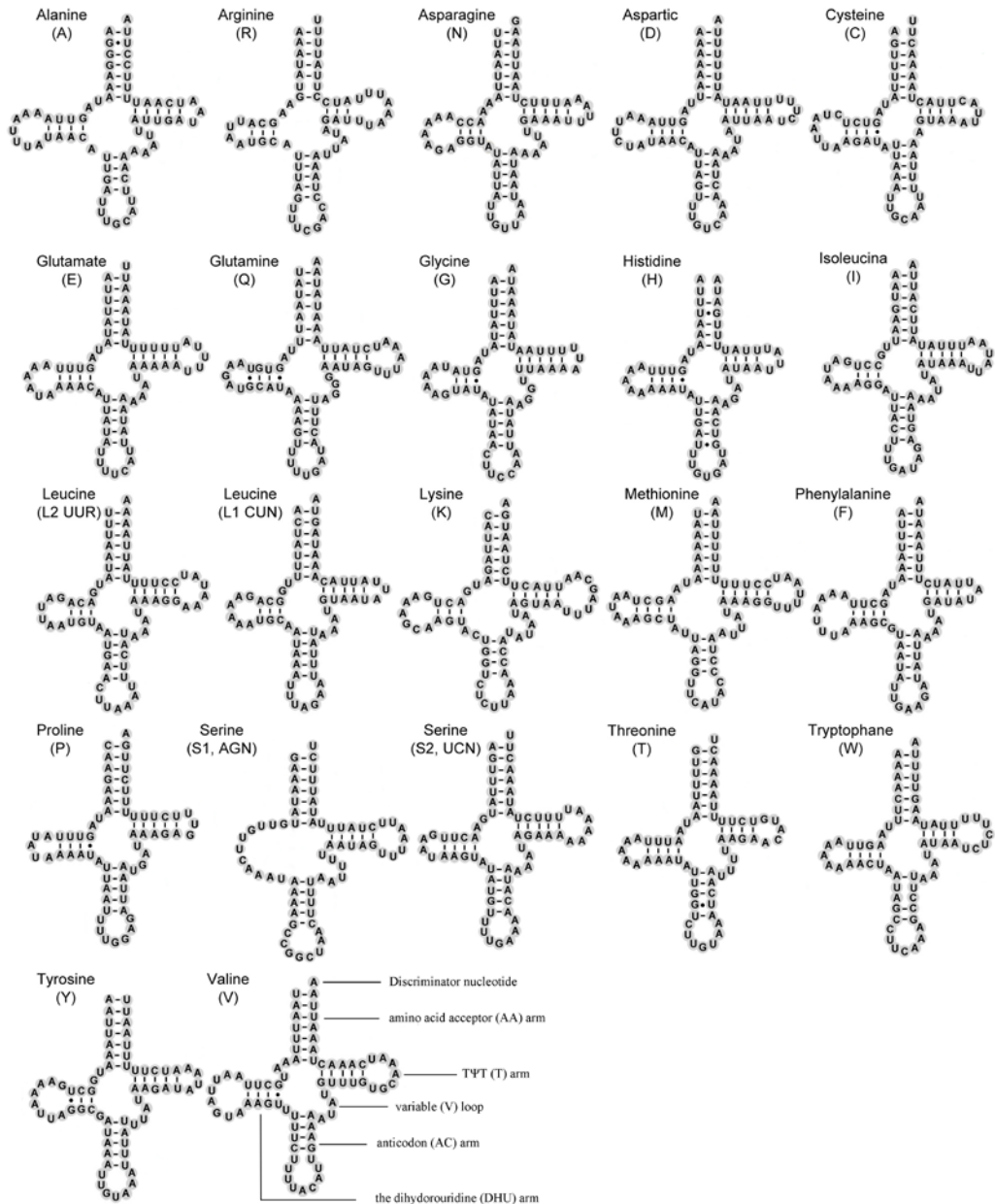


Figure 3. Secondary structures of tRNAs of *Orfelia* sp. Dashes indicate Watson-Crick base pairing and dots indicate G-U base pairing.

The complete mt genome of *Orfelia* sp. (GenBank accession no. MW394227) is 15,521 bp in length. It is a single circular DNA molecule containing 13 PCGs, 22 tRNAs, two rRNAs and two long non-coding regions that are found between *srRNA* and *tRNA^{Ile}*, and between *tRNA^{Ile}* and *tRNA^{Gln}* (Fig. 2). Twenty-three genes are encoded by the majority strand (J-strand), and 14 genes are encoded by the minority strand (N-strand) (Clary & Wolstenholme 1985; Simon *et al.* 2006). Eleven gene overlaps are found in the mt genome, ranging from 1 to 8 bp in length. Twelve small non-coding intergenic spacers are found in the mt genome, ranging

from 2 to 20 bp in length. The largest non-coding intergenic spacer is between *tRNA^{Glu}* and *tRNA^{Phe}*. The canonical mt start codons (ATN) for invertebrate mt genomes (Wolstenholme 1992) are found in 12 PCGs of the mt genome, while *ND5* used uncanonical start codons GTG. The complete stop codons, TAA and TAG, are found in 12 PCGs of the mt genome, except for *ND4* with a single thymine stop codon. The entire twenty-two typical tRNAs are found in the mt genome (Fig. 3), ranging from 64 bp to 72 bp. The *lrRNA* in the mt genome are determined to be 1,322 bp in length and the *srRNA* are 789 bp in length. The nucleotide composition of the mt genomes are biased towards A and T.

The nucleotide composition of the *Orfelia* sp. mt genome is biased toward A and T. The AT content of the whole mt genome is 78.9% (A: 40.4%; T: 38.5%; C: 12.5%; G: 8.6%). For PCGs, the AT content of the N strand genes (78.1%) is higher than that of the J strand genes (75.7%). For rRNAs, the AT content of the *lrRNA* (85.0%) is slightly higher than that of the *srRNA* (82.4%). The AT content of the tRNAs is 83.3%. The control region has the highest AT content (90.3%), which is typical of an animal mt genome. The mt genome has a weakly positive AT-skew and a negative GC-skew on the J-strand (Table 1).

Table 1. Comparisons of mitochondrial genomes between two Keroplatidae flies

Regions		<i>Orfelia</i> sp.	<i>Arachnocampa flava</i>
Number of PCGs		13	13
Number of tRNAs		22	22
Number of gene overlaps		11	7
Number of intergenic spacers		11	18
Whole mitogenome	A+T%	78.9	82.0
	G+C%	21.1	18.0
	AT-skew	0.03	-0.03
	GC-skew	-0.19	-0.18
PCGs	A+T%	76.6	78.9
PCGs(J)	A+T%	75.7	77.8
PCGs(N)	A+T%	78.1	80.7
<i>lrRNA</i>	A+T%	85.0	85.2
<i>srRNA</i>	A+T%	82.4	84.2
tRNA genes	A+T%	83.3	84.5
Control region	A+T%	90.3	93.3

Note: The AT-bias and GC-bias of PCGs were calculated by using the formulae: AT-skew = (A-T)/(A+T), GC-skew = (G-C)/(C+G).

Comparisons of mt genomes between the subfamilies Keroplatinae (*Orfelia* sp.) and Arachnocampinae (*Arachnocampa flava* Harrison, 1966) are also listed in Table 1. The AT content of the whole and each part of the mt genome of *Orfelia* sp. are higher than that of *A. flava*, except for the control region. The proportionate number of nonsynonymous differences (dN) is usually divided by the proportionate number of synonymous differences (dS) to quantify the gene evolutionary rates. The resulting ratio (dN/dS) among Keroplatidae flies (Fig. 4) shows that *ND5* has the fastest evolutionary rate, while *COI* has the slowest evolutionary rate. It is worth noting that these results are only based on the current limited

mitochondrial genome data. It is expected that more mitochondrial genome data will be published in the future to improve the accuracy of comparative analysis results and provide more molecular data for phylogenetic study.

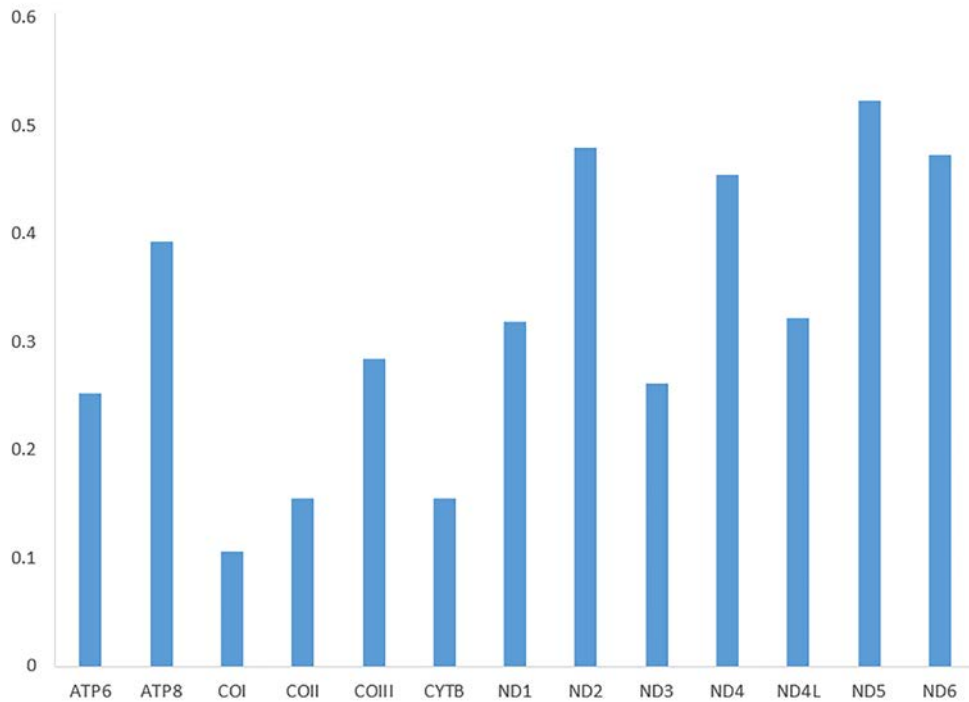


Figure 4. Ratio (dN/dS) of 13 protein-coding genes (PCGs) of mitochondrial genomes among Keroplastidae flies.

Acknowledgements

The authors express their sincere thanks to Jiaqi SHAO, Gang GAO, Yiqun LI, Wenxiao CUI, Qingyun LI and Muhammad Yasir Ali for their great help during this study. This work was supported by the National Natural Science Foundation of China (41901061; 32100356) and the High-level Talents Funds of Qingdao Agricultural University, China (663-1119008; 663-1118015).

References

- Beckenbach A. 2012. Mitochondrial genome sequences of Nematocera (Lower Diptera): evidence of rearrangement following a complete genome duplication in a winter crane fly genome. *Evolutionary Biology*, 4(2): 89–101.
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M & Stadler P. 2013. MITOS: improved de novo metazoan mitochondrial genome 3 annotation. *Molecular Phylogenetics and Evolution*, 69(2): 313–319.
- Cao J. 2007. *Study on Taxonomy of the Family Keroplastidae from China (Diptera: Keroplastidae)*. Master

- Dissertation, Sichuan Agricultural University, Ya'an, Sichuan.
- Cao J, Xu H, Zhou Z, Wu H & Gu Y. 2008. First record of the genus *Orfelia* from China, with descriptions of three new species (Diptera: Keroplatidae). *Entomological News*, 119(3): 271–277.
- Clary D & Wolstenholme D. 1985. The ribosomal RNA genes of *Drosophila* mitochondrial DNA. *Nucleic Acids Research*, 13(11): 4029–4045.
- Costa A. 1857. Contribuzione alla fauna ditterologica Italiana. II. *Giambattista Vico Giornale Scientifico Napoli*, 2: 438–460.
- Harrison R. 1966. Australian glow-worms of the genus *Arachnocampa* Edwards (Diptera: Mycetophilidae). *Pacific Insects*, 8: 877–883.
- Kumar S, Stecher G & Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7): 1870–1874.
- Lowe T & Chan P. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, 44: W54–W57.
- Miao X, Huang J, Menzel F, Wang Q, Wei C, Lin X & Wu H. 2020. Five mitochondrial genomes of black fungus gnats (Sciariidae) and their phylogenetic implications. *International Journal of Biological Macromolecules*, 150: 200–205.
- Ševčík J, Kasprák D, Mantič M, Fitzgerald S, Ševčíková T, Tóthová A & Jaschhof M. 2016. Molecular phylogeny of the megadiverse insect infraorder Bibionomorpha sensu lato (Diptera). *PeerJ*, 4: e2563.
- Simon C, Frati F, Beckenbach A, Crespi B, Liu H & Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, 87(6): 651–701.
- Simon C, Buckley T, Frati F, Stewart J & Beckenbach A. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics*, 37: 55–579.
- Wang Q, Huang J & Wu H. 2021. Mitogenomes provide insights into the phylogeny of Mycetophilidae (Diptera: Sciaroidea). *Gene*, 783: 145564.
- Wolstenholme D. 1992. Animal mitochondrial DNA: structure and evolution. *International Review of Cytology*, 141: 173–216.
- Zhang X, Kang Z & Yang D. 2023. New data on the mitochondrial genome of Nematocera (lower Diptera): features, structures and phylogenetic implications. *Zoological Journal of the Linnean Society*, 197(1): 229–245.