First complete mitogenomes of Diamesinae, Orthocladiinae, Prodiamesinae, Tanypodinae (Diptera: Chironomidae) and their implication in phylogenetics

Chen-Guang Zheng[1], Xiu-Xiu Zhu[1], Li-Ping Yan[2], Yuan Yao[3], Wen-Jun Bu[1], Xin-Hua Wang[1] and Xiao-Long Lin[1]

[1] College of Life Sciences, Nankai University, Tianjin, China
[2] School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, China
[3] College of Life Sciences, Tianjin Normal University, Tianjin, China

ABSTRACT

Background. The mitochondrial genome (mitogenome) has been extensively used for phylogenetic and evolutionary analysis in Diptera, but the study of mitogenome is still scarce in the family Chironomidae.

Methods. Here, the first complete mitochondrial genomes of four Chironomid species representing Diamesinae, Orthocladiinae, Prodiamesinae and Tanypodinae are presented. Coupled with published mitogenomes of two, a comparative mitochondrial genomic analysis between six subfamilies of Chironomidae was carried out.

Results. Mitogenomes of Chironomidae are conserved in structure, each contains 37 typical genes and a control region, and all genes arrange the same gene order as the ancestral insect mitogenome. Nucleotide composition is highly biased, the control region displayed the highest A + T content. All protein coding genes are under purifying selection, and the ATP8 evolves at the fastest rate. In addition, the phylogenetic analysis covering six subfamilies within Chironomidae was conducted. The monophyly of Chironomidae is strongly supported. However, the topology of six subfamilies based on mitogenomes in this study is inconsistent with previous morphological and molecular studies. This may be due to the high mutation rate of the mitochondrial genetic markers within Chironomidae. Our results indicate that mitogenomes showed poor signals in phylogenetic reconstructions at the subfamily level of Chironomidae.

Subjects Entomology, Evolutionary Studies, Genomics, Molecular Biology, Zoology
Keywords Chironomidae, Diptera, Mitogenome, Phylogeny

Submitted 21 December 2020
Accepted 27 March 2021
Published 6 May 2021

Corresponding author
Xiao-Long Lin, lin880224@gmail.com

Academic editor
Joseph Gillespie

Additional Information and Declarations can be found on page 11

DOI **10.7717/peerj.11294**

INTRODUCTION

The typical mitochondrial genome (mitogenome) of insects is a double-strand circular molecule ranging from 14kb to 20kb in size, which encodes 37 genes (13 protein-coding genes, two ribosomal RNA genes, and 22 transfer RNA genes) and a control region (*Boore, 1999*; *Cameron, 2014*; *Wolstenholme, 1992*). Due to its small genome size, maternal inheritance, low sequence recombination, and fast evolutionary rates (*Brown, George &*

How to cite this article Zheng C-G, Zhu X-X, Yan L-P, Yao Y, Bu W-J, Wang X-H, Lin X-L. 2021. First complete mitogenomes of Diamesinae, Orthocladiinae, Prodiamesinae, Tanypodinae (Diptera: Chironomidae) and their implication in phylogenetics. *PeerJ* **9**:e11294 http://doi.org/10.7717/peerj.11294

Wilson, 1979; *Curole & Kocher, 1999*), the mitogenome is considered as powerful marker for phylogenetic and evolutionary analysis (*Condamine et al., 2018*; *Jacobsen et al., 2012*; *Stokkan et al., 2018*; *Tang et al., 2019b*). Due to high-throughput sequencing technology, an increasing number of complete mitogenomes have been sequenced among the Diptera, covering most families (*Kang, Li & Yang, 2016*; *Li et al., 2020*; *Miao et al., 2020*; *Ramakodi et al., 2015*; *Tang et al., 2019a*). Mitogenomes have been widely used for mitochondrial structure comparison and phylogenetic analysis at different taxonomic level of the Diptera (*Chen et al., 2018*; *De Oliveira Aragão et al., 2019*; *Miao et al., 2020*; *Yan et al., 2019*; *Zhang et al., 2016*; *Zhang et al., 2019b*). However, complete mitogenomes are still scarce for the family Chironomidae, which limits our understanding of their mitochondrial structure and phylogenetic pattern. In addition, it is still unknown whether mitogenomes can effectively resolve phylogenetic relationships at the subfamily level within Chironomidae.

The dipteran family Chironomidae is a diverse aquatic insect group, and are important bioindicators for freshwater ecosystem monitoring. Within Chironomidae, several phylogenetic studies have been conducted based on morphological characters or combining genetic markers to reconstruct the evolutionary history of subfamilies (*Cranston, Hardy & Morse, 2012*; *Sæther, 1977*), but no one has attempted to use mitogenomes. Prior to this study, only five mitogenomes of Chironomidae were available (*Beckenbach, 2012*; *Deviatiiarov, Kikawada & Gusev, 2017*; *Kim, Kim & Shin, 2016*; *Park et al., 2020*; *Zhang et al., 2019a*), representing species from three subfamilies: Chironominae, Podonominae, and Prodiamesinae. However, comparative analysis of mitogenome structure, base composition, substitution and evolutionary rates among subfamilies has not been carried out. In addition, the monophyly of Chironomidae has not been supported by a recent study using mitogenomes of Culicomorpha (*Zhang et al., 2019b*).

In the present study, we provide complete mitogenomes for four species representing the subfamilies Diamesinae, Orthocladiinae, Prodiamesinae, and Tanypodinae. Along with the published mitogenomes of subfamilies Chironominae and Podonominae, the first comparative analysis of the genome structure, base composition, substitution and evolutionary rates among six chironomid subfamilies is presented. In addition, a phylogenomic analysis covering six chironomid subfamilies was carried out.

MATERIALS & METHODS

Taxon sampling

Complete mitogenomes of six chironomid species (Appendix S1), representing six subfamilies, were analyzed in this study, with two ceratopogonid species used as outgroups. The mitogenomes of four non-biting midge species, *Potthastia* sp. (Diamesinae), *Rheocricotopus villiculus* (Orthocladiinae), *Prodiamesa olivacea* (Prodiamesinae) and *Clinotanypus yani* (Tanypodinae) are documented for the first time. The mitogenomes of *Chironomus tepperi* (Chironominae) and *Parochlus steinenii* (Podonominae) were retrieved from GenBank (*Beckenbach, 2012*; *Kim, Kim & Shin, 2016*). The mitogenome of *Propsilocerus akamusi* (MN566452) (*Zhang et al., 2019a*) was excluded from the present study because it is incomplete and lacks annotation. In addition, two species

Table 1 Taxonomic information, sampling metadata, GenBank accession numbers and references of mitochondrial genomes used in the study.
.

Family	Subfamily	Species	Life stage	Sampling metadata	Accession no	Reference
Ceratopogonidae	Ceratopogoninae	*Culicoides arakawae*			NC_009809	*Matsumoto et al. (2009)*
Ceratopogonidae	Forcipomyiinae	*Forcipomyia makanensis*		Makan, Zunyi, Guizhou, China, 27.630765°N, 106.848949°E	MK000395	*Jiang et al. (2019)*
Chironomidae	Chironominae	*Chironomus tepperi*			JN861749	*Beckenbach (2012)*
Chironomidae	Diamesinae	*Potthastia* sp.	Adult male	Wuying, Yichun, Heilongjiang, China, 48.0869°N, 129.2470°E, 27-Jui-2016, leg. C. Song	MW373523	This study
Chironomidae	Orthocladiinae	*Rheocricotopus villiculus*	Adult male	Tianmu Mountain National Nature Reserve, Hangzhou, Zhejiang, China, 30.3222°N, 119.442°E, 22-Jul-2019, leg. X.-L. Lin	MW373526	This study
Chironomidae	Podonominae	*Parochlus steinenii*		King George Island, West, Antarctica, 62.2333°S, 58.7833°W, summer in 2015	KT003702	*Kim, Kim & Shin (2016)*
Chironomidae	Prodiamesinae	*Prodiamesa olivacea*	Larva	Jiuzhaigou Valley Scenic and Historic Interest Area, Sichuan, China, 33.1928°N, 103.8942°E, 12-Jul-2019, leg. X.-Y. Ge	MW373525	This study
Chironomidae	Tanypodinae	*Clinotanypus yani*	Adult male	Jiulongshan Nature Reserve, Guangyuan, Sichuan, China, 31.976379°N, 106.035644°E, 8-Aug-2017, leg. C. Song	MW373524	This study

of Ceratopogonidae (*Culicoides arakawae* and *Forcipomyia makanensis*) (*Jiang et al., 2019*; *Matsumoto et al., 2009*) were selected as outgroups for phylogenetic analyses since Ceratopogonidae was strongly supported as the sister group of Chironomidae in previous studies (*Kutty et al., 2018*). Detailed taxon sampling information is listed in Table 1. The vouchers of the newly sequenced species are deposited at the college of Life Sciences, Nankai University, Tianjin, China.

DNA extraction, sequencing and assembling

For the newly sequenced species, total genomic DNA was extracted from the body, (except abdomen and genitalia) using a General AllGen Kit (Qiagen, Germany). The entire mitogenome of each species were sequenced using the Illumina NovaSeq 6000 platform

with an insert size of 350-bp and a paired-end 150-bp sequencing strategy by the Allwegene Company and Novogene Co., Ltd. at Beijing, China. About 2 Gb clean data were obtained from each library after trimming using Trimmomatic (*Bolger, Lohse & Usadel, 2014*).

To ensure the accuracy of the mitogenome sequences, three frequently used assembly methods were applied to each sample. A *de novo* assembly was performed using IDBA-UD (*Peng et al., 2012*) with minimum and maximum k values of 40 and 120 bp, respectively. Two reference based assemblies were performed using Geneious 2020.2.1 (*Kearse et al., 2012*) with default setting and MITObim 1.9 (*Hahn, Bachmann & Chevreux, 2013*). The mitogenome sequences obtained by the three methods were aligned, manually compared, and finally compiled into a single sequence in Geneious 2020.2.1 (*Kearse et al., 2012*).

Genome annotation, composition and substitution rate

Genome annotation was conducted as previously described in *Zheng et al. (2020)*. Specifically, the transfer RNA (tRNA) genes and their secondary structures were detected by MITOS2 webserver (http://mitos2.bioinf.uni-leipzig.de/index.py) (*Bernt et al., 2013*) with invertebrate mitochondrial genetic code. The ribosomal RNA (rRNA) genes were predicted by alignment with homologous regions of mitogenome from closely related species. Protein coding genes (PCGs) were initially annotated using the Open Reading Frame Finder (ORF Finder) as implemented at the NCBI website (https://www.ncbi.nlm.nih.gov/orffinder/) and then compared with published mitogenomes of insects using the program BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Newly sequenced mitogenomes were submitted to GenBank (accession numbers: MW373523–MW373526).

CGView Server V 1.0 (*Grant & Stothard, 2008*) was used to draw mitogenome maps. Codon usage of PCGs and nucleotide composition were calculated in MEGA X (*Kumar et al., 2018*). The bias of the nucleotide composition was measured according to the formulas: AT-skew = (A−T)/(A+T) and GC-skew = (G−C)/(G+C). Rates of non-synonymous substitution rate (Ka) and synonymous substitution rate (Ks) were calculated in DnaSP 6.12.03 (*Rozas et al., 2017*).

Phylogenetic analyses

Phylogenetic analyses were conducted using the sequences of 13 PCGs and two rRNAs. The PCGs were aligned based on amino acid sequences using Muscle implemented in MEGA X (*Kumar et al., 2018*). The rRNAs were aligned using MAFFT 7.402 (*Katoh & Standley, 2013*) with algorithm G-INS-i strategy. Alignments of individual genes were then concatenated using SequenceMatrix v1.7.8 (*Vaidya, Lohman & Meier, 2011*) to generate five datasets: PCG123 (all three codon positions of the 13 PCGs), PCG123R (all three codon positions of the 13 PCGs and two rRNAs), PCG12 (the first and second codon positions of the 13 PCGs), PCG12R (the first and second codon positions of the 13 PCGs and two rRNAs), and AA (amino acid sequences of the 13 PCGs). To test substitution saturation, transition and transversion rates were evaluated by DAMBE 5.6.14 (*Xia, 2013*). The program PartitionFinder 2.0 (*Lanfear et al., 2017*) was used to infer the best substitution model (Table S1). The analysis of Bayesian inference (BI) and maximum likelihood (ML) were conducted for each dataset. The BI analyses were performed under

the program MrBayes 3.2.7a (*Ronquist et al., 2012*) with partitioned models (Table S1). Two simultaneous Markov chain Monte Carlo (MCMC) runs of 10,000,000 generations were conducted, trees were sampled every 1000 generations with a burn-in of 25%. The program Tracer 1.7 (*Rambaut et al., 2018*) was used to assess the convergence of runs. The ML analyses were conducted using the program RAxML 8.0.12 (*Stamatakis, 2014*) under the substitution model GTR +GAMMA +I. The nodal support values were calculated with 1,000 bootstrap replicates.

RESULTS

Mitogenome organization and composition

The complete mitogenomes of *Chironomus tepperi*, *Potthastia* sp., *Rheocricotopus villiculus*, *Parochlus steinenii*, *Prodiamesa olivacea*, and *Clinotanypus yani* are 15,652, 15,913, 15,985, 16,803, 16,190, and 16,247 bp in size, respectively (Fig. 1; Appendix S2). They are circular molecules, each containing 37 typical mitochondrial genes (13 PCGs, two rRNAs, and 22 tRNAs) and one control region. Among these genes, four PCGs (ND1, ND4, ND4L, and ND5), eight tRNAs (trnC, trnF, trnH, trnL (UAG), trnP, trnQ, trnV, and trnY), and two rRNAs (12S rRNAs and 16S rRNAs) are encoded by the minority strand (N strand), while the other 23 genes are located in the majority strand (J strand). ATP8-ATP6 and ND4L-ND4 overlap by seven nucleotides (ATGATAA and ATGTTAA, respectively) in all six Chironomidae species.

Nucleotide composition (Table 2) of the six Chironomidae species is similar, with a high A+T bias (72.4%–76.8%), the control region has the highest A+T content while the first and the second codon positions of PCGs have the lowest A+T content. All six Chironomidae species exhibited negative AT-skew and GC-skew. All three codon positions of PCGs had negative AT-skew, the GC-skew of the first codon position was positive, while the 2nd and the 3rd codon position were negative. Some gene regions exhibited different nucleotide skew among the six Chironomidae species. For example, in 12S rRNA, the AT-skew of *Chironomus tepperi* and *Clinotanypus yani* are −0.01 and 0.00 respectively, while the AT-skew are positive (0.01–0.05) in the remaining four species.

Protein coding genes

Among Chironomidae species, most PCGs initiate with the standard start codon ATN. The start codon of COI was TTG in *Chironomus tepperi*, *Potthastia* sp., *Rheocricotopus villiculus* and *Prodiamesa olivacea*. The start codon of ND5 in *Chironomus tepperi*, *Potthastia* sp., *Rheocricotopus villiculus*, *Prodiamesa olivacea* and *Clinotanypus yani* was GTG. ND1 started with TTG in *Potthastia* sp., *Rheocricotopus villiculus*, *Parochlus steinenii*, *Prodiamesa olivacea*, and *Clinotanypus yani*. Most PCGs have complete termination codons (TAA or TAG), however, COII in *Parochlus steinenii* and *Clinotanypus yani* has an incomplete termination codon (T-).

Total codon number (except the termination codons) in *Chironomus tepperi*, *Potthastia* sp., *Rheocricotopus villiculus*, *Parochlus steinenii*, *Prodiamesa olivacea*, and *Clinotanypus yani* were 3,730, 3,743, 3,726, 3,729, 3,729, and 3,709, respectively. The most frequently codon families are Ile, Leu2, and Phe (>300), while the least used codon family is Cys (<50)

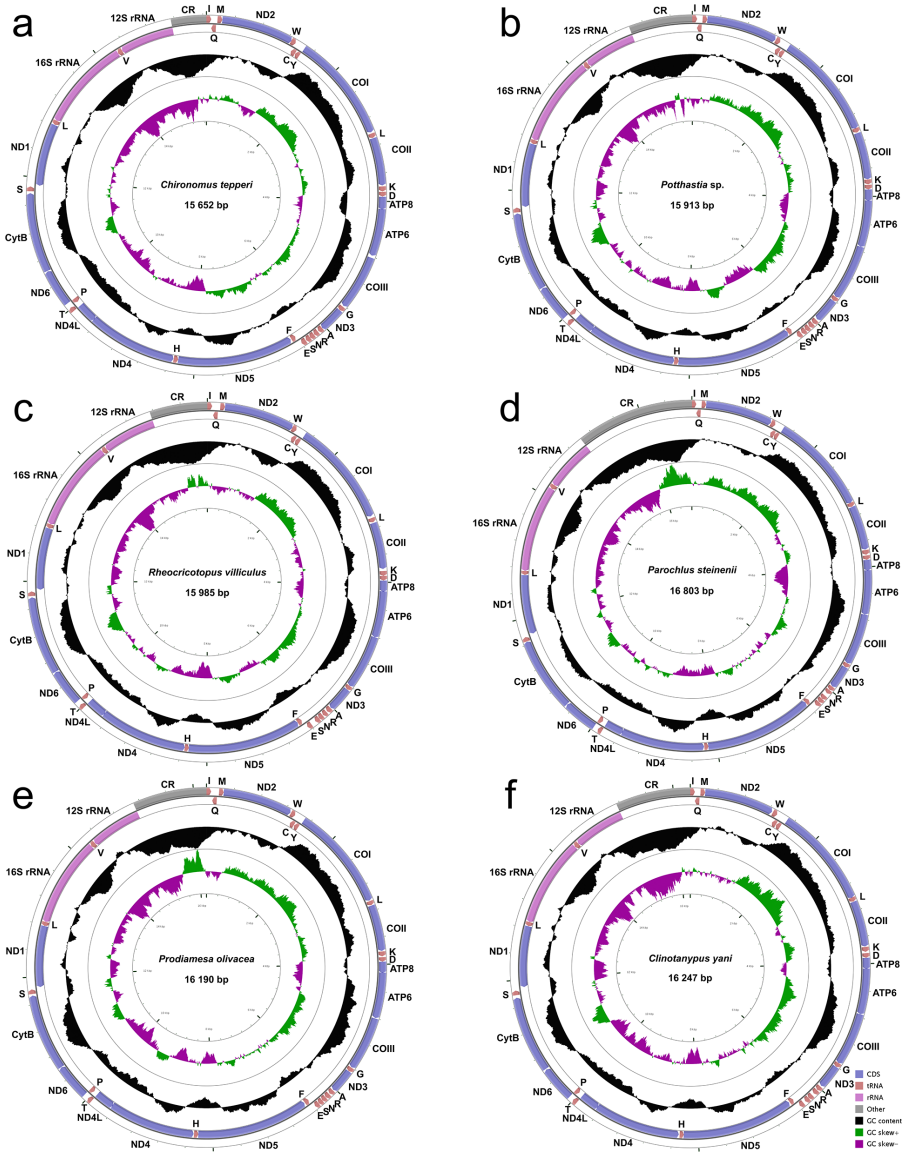

Figure 1 **Mitogenome maps of *Chironomus tepperi* (A), *Potthastia* sp. (B), *Rheocricotopus villiculus* (C), *Rheocricotopus villiculus* (D), *Prodiamesa olivacea* (E), *Clinotanypus yani* (F).** The names of PCGs and rRNAs are indicated by standard abbreviations, while names of tRNAs are represented by a single letter abbreviation. The first circle shows the gene map and arrows indicate the orientation of gene transcription. Blue arrows refer to PCGs, purple arrows refer to rRNAs, red arrows refer to tRNAs and grey arrow refers to control region. The second circle shows the GC content, which is plotted as the deviation from the average GC content of the entire sequence. The third circle shows the GC-skew, which is plotted as the deviation from the average GC-skew of the entire sequence. The innermost circle shows the sequence length.

Full-size DOI: 10.7717/peerj.11294/fig-1

in all six Chironomidae species (Fig. 2). The relative synonymous codon usage (RSCU) patterns among the six Chironomidae species are similar. The RSCU values are showed in Fig. 3. All synonymous codons of 20 amino acids are present. The most frequent used codons are NNU and NNA for each amino acid (Fig. 3).

Table 2 Nucleotide composition of mitochondrial genomes of the six Chironomidae species.

	Species	Whole genome	Protein coding genes	First codon position	Second codon position	Third codon position	tRNA genes	12S rRNA	16S rRNA	Control region
A+T%	*Chironomus tepperi*	76.7	74.3	67.6	67.6	87.6	79.0	82.6	84.3	93.0
	Potthastia sp.	76.8	74.7	69.0	66.1	88.9	76.8	78.1	82.7	93.3
	Rheocricotopus villiculus	77.3	74.4	69.6	67.4	86.0	79.5	84.1	84.4	93.7
	Parochlus steinenii	72.4	69.0	64.6	64.7	77.5	73.2	76.4	80.1	85.5
	Prodiamesa olivacea	75.8	73.4	66.7	65.5	88.2	76.2	78.1	81.9	89.2
	Clinotanypus yani	75.0	72.5	65.4	65.1	87.0	75.7	79.1	81.3	88.7
AT-Skew	*Chironomus tepperi*	−0.14	−0.20	−0.09	−0.41	−0.13	0.03	−0.01	0.00	−0.11
	Potthastia sp.	−0.13	−0.20	−0.10	−0.39	−0.12	0.03	0.01	0.04	−0.05
	Rheocricotopus villiculus	−0.12	−0.18	−0.09	−0.40	−0.08	0.02	0.01	0.05	−0.07
	Parochlus steinenii	−0.11	−0.19	−0.09	−0.40	−0.11	0.04	0.05	0.03	0.06
	Prodiamesa olivacea	−0.12	−0.19	−0.10	−0.40	−0.09	0.03	0.04	0.01	0.02
	Clinotanypus yani	−0.13	−0.19	−0.10	−0.39	−0.10	0.02	0.00	0.03	−0.08
GC-Skew	*Chironomus tepperi*	−0.06	−0.02	0.19	−0.18	−0.12	−0.13	−0.37	−0.36	−0.43
	Potthastia sp.	−0.03	0.02	0.27	−0.16	−0.15	−0.12	−0.24	−0.28	−0.31
	Rheocricotopus villiculus	−0.04	−0.01	0.25	−0.18	−0.17	−0.09	−0.23	−0.34	−0.19
	Parochlus steinenii	−0.04	−0.01	0.21	−0.17	−0.10	−0.06	−0.21	−0.26	−0.18
	Prodiamesa olivacea	−0.04	0.00	0.25	−0.16	−0.21	−0.09	−0.24	−0.29	−0.16
	Clinotanypus yani	−0.06	0.00	0.24	−0.18	−0.18	−0.12	−0.28	−0.34	−0.39

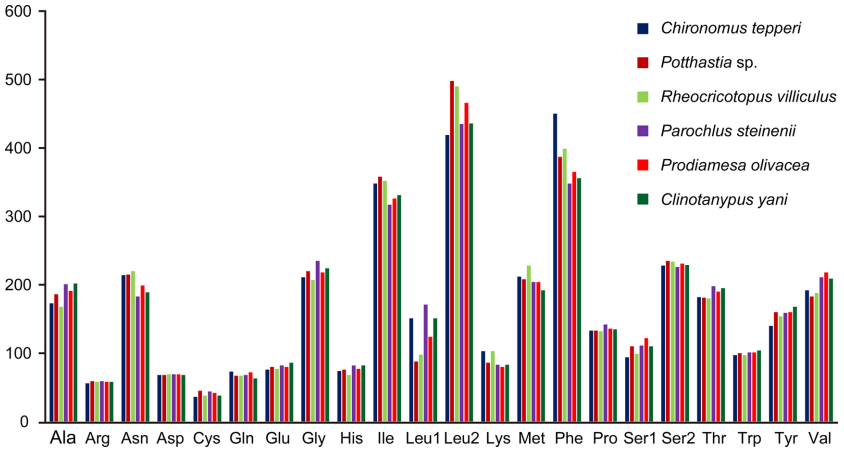

Figure 2 **Patterns of codon usage of the six mitogenomes of six chironomid subfamilies.** The *X*-axis shows the codon families and the *Y*-axis shows the total codons.

Full-size DOI: 10.7717/peerj.11294/fig-2

The Ka/Ks value (ω) was used to test for signatures of natural selection (*Cheng et al., 2018*; *Hu & Banzhaf, 2008*). The ω value of all PCGs are less than 0.6. Among the 13 PCGs, ATP8 has the largest ω value, indicating that ATP8 evolves at the fastest rate. The animal DNA barcoding gene COI has the lowest ω value (Fig. 4).

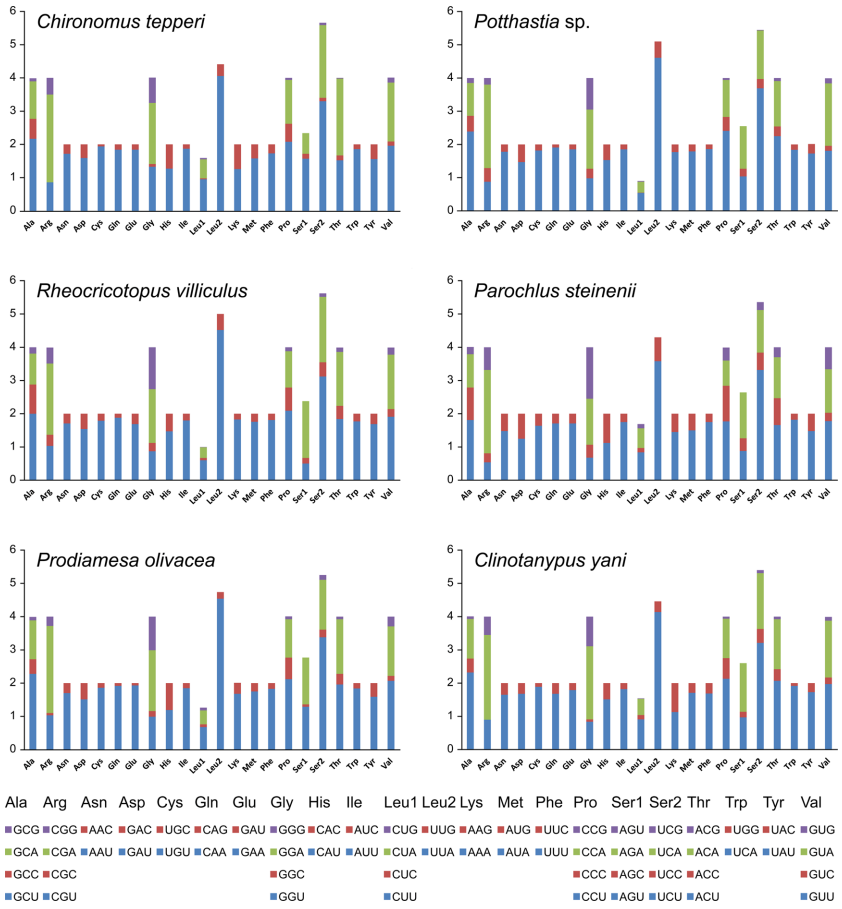

Figure 3 The relative synonymous codon usage (RSCU) in the six mitogenomes of six chironomid subfamilies. The *X*-axis shows the codons and the *Y*-axis shows RSCU values.

Full-size DOI: 10.7717/peerj.11294/fig-3

tRNAs, rRNAs and control region

The typical set of 22 tRNA genes were identified in the mitogenomes of all six Chironomidae species, ranging from 63 to 72 bp in length. The tRNA genes exhibit high A+T content (73.2%–79.5%), positive AT-skew, and negative GC-skew (Table 2). All the predicted tRNAs can be folded into the typical clover-leaf secondary structure except trnS (GCU), which lacks the dihydrouridine (DHU) arm. The non-Watson-crick base pair G-U is common in tRNA genes from all Chironomidae species (Fig. S1–S6).

Both 12S and 16S rRNA genes exhibit similar position and size across the Chironomidae mitogenomes. The A+T content of 12S and 16S rRNA genes ranges from 76.4% to 82.6% and 80.1% to 84.4%, respectively. Both genes exhibit positive AT-skew and negative GC-skew in all Chironomidae species except *Chironomus tepperi*: the AT-skew of 12S rRNA and 16S rRNA in *Chironomus tepperi* is −0.01 and 0.00, respectively (Table 2).

The control regions of *Chironomus tepperi*, *Potthastia* sp., *Rheocricotopus villiculus*, *Parochlus steinenii*, *Prodiamesa olivacea*, and *Clinotanypus yani* are 500, 911, 832, 1,783,

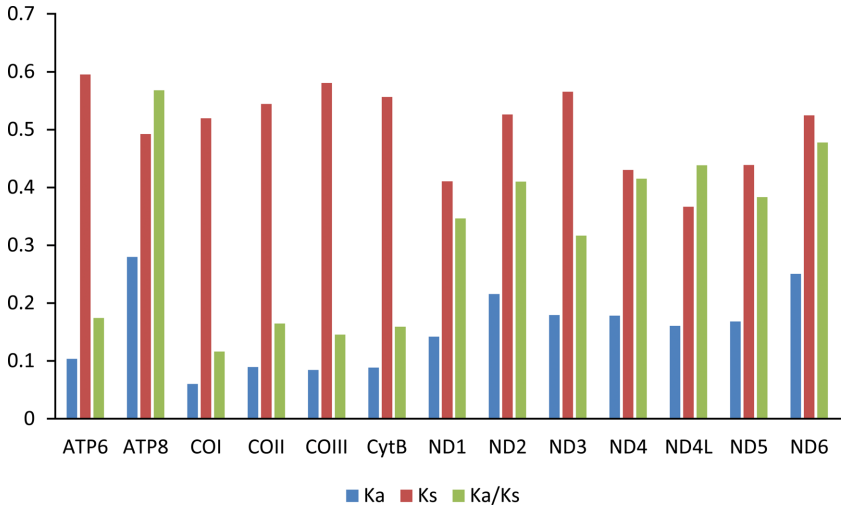

Figure 4 Evolution rate of each PCG of the six mitogenomes of six chironomid subfamilies mitogenomes. Ka refers to non-synonymous substitution rate, Ks refers to synonymous substitution rate, Ka/Ks refers to evolution rate of each PCG.

Full-size DOI: 10.7717/peerj.11294/fig-4

1,079, and 1,095 bp in size, respectively (Appendix S2). All are A + T rich (85.5%–93.7%), much higher than the whole mitogenomes (72.4%–77.3%).

Saturation test and phylogenetic analyses

Saturation tests were performed for the four nucleotide datasets. Each dataset was free of saturation (Fig. S7). In general, phylogenetic trees support the monophyly of the Chironomidae across different datasets in ML and BI analyses (Fig. 5, PP = 1, BS = 100). Within Chironomidae, four topologies were inferred from five datasets: (i) Orthocladiinae + (Chironominae + ((Diamesinae + Prodiamesinae) + (Podonominae + Tanypodinae))) was inferred from the PCG123 and PCGR datasets (Figs. 5A and 5B); (ii) Orthocladiinae + (Chironominae + (Diamesinae+ (Prodiamesinae + (Podonominae + Tanypodinae)))) was inferred from the PCG12 dataset (Fig. 5C); (iii) (Orthocladiinae + Chironominae) + (Diamesinae + (Prodiamesinae + (Podonominae + Tanypodinae))) was inferred from the PCG12R dataset (Fig. 5D); (iv) Chironominae + (Orthocladiinae + (Prodiamesinae + (Diamesinae + (Podonominae + Tanypodinae)))) was inferred from the AA dataset (Fig. 5E). The topology inferred from the AA had the strongest nodal support. Based on five different datasets, Podonominae is sister to Tanypodinae with strong support in both BI (PP $\geq$ 0.98) and, ML (BS = 100) reconstructions, which makes the sister to (Diamesinae + Prodiamesinae) with strong support (PP = 1, BS $\geq$ 91) at the "tip" position. The remaining subfamilies Chironominae and Orthocladiinae are sister to above four subfamilies.

DISCUSSION

Mitogenome features

The entire mitogenome length of the six Chironomidae species differs considerably (15,652–16,803 bp), mainly due to the variation in control region size. All Chironomidae

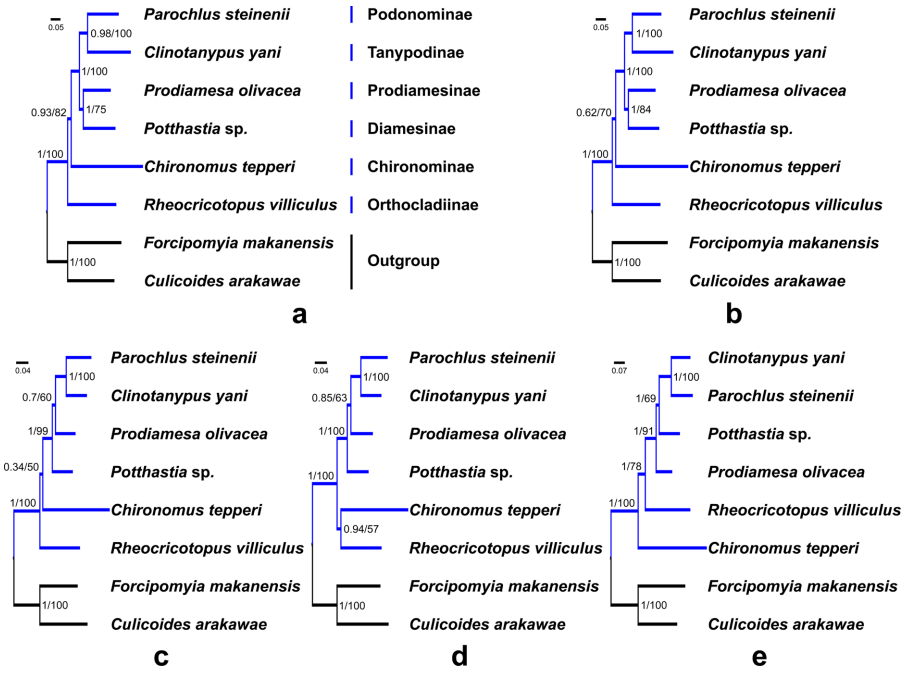

Figure 5 Phylogenetic relationships of six subfamilies within Chironomidae inferred from mitogenomes. (A) Topology obtained based on PCG123; (B) Topology obtained based on PCG123R; (C) Topology obtained based on PCG12; (D) Topology obtained based on PCG12R; (E) Topology obtained based on AA. Numbers at the nodes are BI posterior probabilities (left) and ML bootstrap values (right).
Full-size DOI: 10.7717/peerj.11294/fig-5

mitogenomes contain 37 typical genes and a control region, the order and arrangement of these genes are completely accordant with the ancestral insect gene arrangement (*Clary & Wolstenholme, 1985*). The whole mitogenome of Chironomidae has high A+T content and similar AT/GC-skew, consistent with the similar base composition biases of insect mitochondrial DNA (*Wei et al., 2010*). This type of nucleotide bias may be related to the asymmetric mutation processes during replication (*Hassanin, Leger & Deutsch, 2005*).

Among the Chironomidae mitogenomes, most PCGs have complete termination codons, while the COII gene in *Parochlus steinenii* and *Clinotanypus yani* has an incomplete termination codon (T-) that probably completed by post-transcriptional polyadenylation (*Ojala, Montoya & Attardi, 1981*). The patterns of codon usage among the Chironomidae mitogenomes are nearly the same. The most frequent used codons were NNU and NNA for each amino acid, reflecting the AT bias of nucleotide composition. For most amino acids, the most frequently used codon is not the anti-codon that strictly correspond to tRNA. The low ω value for each PCG indicates that they are all under strong purifying selection. The animal DNA barcoding gene COI has the lowest evolutionary rate, which is consistent with the results observed from other insect groups (*Li et al., 2020*; *Yang, Yu & Du, 2013*; *Zhang & Ye, 2017*).

All six Chironomidae mitogenomes contain the 22 typical tRNA genes, and secondary structure across species is similar. Unlike other tRNA genes, trnS (GCU) lacks the dihydrouridine (DHU) arm. This could be commonly found in published insect

mitogenomes (*Li et al., 2012*; *Lu, Huang & Deng, 2020*; *Zhang et al., 2018*). The A+T contents of 12S rRNA, 16S rRNA, and control region are much higher than that in the whole genome in Chironomidae mitogenomes, indicating a strong A+T bias in these regions.

Phylogenetic analyses

In this study, we applied a variety of strategies to explore the phylogenetic relationships of six subfamilies within the Chironomidae, and confirmed the monophyly of Chironomidae (Fig. 5). However, the topology of subfamilies based on mitogenomes in this study is inconsistent with previous morphological and molecular studies (*Cranston, Hardy & Morse, 2012*; *Sæther, 1977*; *Sæther, 2000*). The present morphological phylogenetics of Chironomidae (*Sæther, 2000*) is composed 11 subfamilies, including (((((Chironominae + Orthocladiinae) + Prodiamesinae) +Diamesinae) + Buchonomyiinae + Chienomyiinae) + ((Usambaromyiinae + Tanypodinae) + Podonominae + Aphroteniinae)) + Telmatogetoninae. The present molecular phylogenetic system of Chironomidae (*Cranston, Hardy & Morse, 2012*) is composed nine subfamilies, including ((((((Chironominae + (Orthocladiinae + Prodiamesinae)) + Diamesinae) + Telmatogetoninae) + Tanypodinae) + Podonominae) + Aphroteniinae) + Buchonomyiinae. Nevertheless, Podonominae and Tanypodinae are ancestral taxa based on both traditional morphological and molecular phylogenies. However, they appear at the "tip" position of mitogenomic phylogenetic tree. Moreover, the "tip" taxa Chironominae and Orthocladiinae appear at the "root" position of the mitogenomic phylogenetic tree. This erroneous phylogenetic reconstruction may be a result of long branch attraction (LBA) (*Siddall & Whiting, 1999*). Due to the high mutation rate of the mitochondrial genetic markers within Chironomidae, some studies (*Ekrem & Willassen, 2004*; *Ekrem, Willassen & Stur, 2010*) have reported that mitochondrial markers (e.g., COI, COII) are not suitable for phylogenetic relationship reconstruction. Here, our mt data reveal different evolutionary history of six subfamilies, which is contradictory with traditional morphology-based systematics. Therefore, we assume that mitogenomes has poor signal for phylogenetic reconstructions at subfamily level in the Chironomidae.

CONCLUSIONS

In this study, we sequenced four complete mitogenomes representing four subfamilies of Chironomidae by whole genome sequencing technologies and did the first comparative analysis of mitogenome base composition and evolutionary history in Chironomidae. The study shows that mitogenomes of Chironomidae are conserved in structure, gene order and nucleotide composition. Our results revealed that mitogenomes have poor phylogenetic signals for subfamily level relationships in Chironomidae.

ACKNOWLEDGEMENTS

A big thank to Dr. Lidong Mo for his help on the manuscript improvement. We also thank Dr. Andrey Krasheninnikov and another two anonymous reviewers for their constructive comments.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was funded by the National Natural Science Foundation of China, grant number 31900344, and China Postdoctoral Science Foundation Grant, grant number 2018M640227. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 31900344.
China Postdoctoral Science Foundation Grant: 2018M640227.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Chen-Guang Zheng analyzed the data, prepared figures and/or tables, and approved the final draft.
- Xiu-Xiu Zhu performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Li-Ping Yan, Wen-Jun Bu and Xin-Hua Wang conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Yuan Yao performed the experiments, prepared figures and/or tables, and approved the final draft.
- Xiao-Long Lin conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The newly sequenced four mitochondrial genomes are available at NCBI SRA (BioProject ID: PRJNA685615), and the assembled sequences are available at GenBank (MW373523–MW373526).

In this study, we also used four published mitochondrial genomes from GenBank: NC_009809, MK000395, JN861749, KT003702

In addition, all mitochondrial genome sequences used in this study are available in the Supplementary File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.11294#supplemental-information.

REFERENCES

Beckenbach AT. 2012. Mitochondrial genome sequences of Nematocera (lower Diptera): evidence of rearrangement following a complete genome duplication in a winter crane fly. *Genome Biology and Evolution* **4**:89–101 DOI 10.1093/gbe/evr131.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**:313–319 DOI 10.1016/j.ympev.2012.08.023.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120 DOI 10.1093/bioinformatics/btu170.

Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Research* **27**:1767–1780 DOI 10.1093/nar/27.8.1767.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **76**:1967–1971 DOI 10.1073/pnas.76.4.1967.

Cameron SL. 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual Review of Entomology* **59**:95–117 DOI 10.1146/annurev-ento-011613-162007.

Chen J-Y, Chang Y-W, Zheng S-Z, Lu M-X, Du Y-Z. 2018. Comparative analysis of the *Liriomyza chinensis* mitochondrial genome with other Agromyzids reveals conserved genome features. *Scientific Reports* **8**:1–10 DOI 10.1038/s41598-018-27213-7.

Cheng YC, Chen MY, Wang JF, Liang AP, Lin CP. 2018. Some mitochondrial genes perform better for damselfly phylogenetics: species-and population-level analyses of four complete mitogenomes of *Euphaea* sibling species. *Systematic Entomology* **43**:702–715 DOI 10.1111/syen.12299.

Clary DO, Wolstenholme DR. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution* **22**:252–271 DOI 10.1007/BF02099755.

Condamine FL, Nabholz B, Clamens A-L, Dupuis JR, Sperling FA. 2018. Mitochondrial phylogenomics the origin, of swallowtail butterflies, and the impact of the number of clocks in Bayesian molecular dating. *Systematic Entomology* **43**:460–480 DOI 10.1111/syen.12284.

Cranston PS, Hardy NB, Morse GE. 2012. A dated molecular phylogeny for the Chironomidae (Diptera). *Systematic Entomology* **37**:172–188 DOI 10.1111/j.1365-3113.2011.00603.x.

Curole JP, Kocher TD. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology & Evolution* **14**:394–398 DOI 10.1016/S0169-5347(99)01660-2.

De Oliveira Aragão A, Neto JPN, Cruz ACR, Casseb SMM, Cardoso JF, Da Silva SP, Ishikawa EAY. 2019. Description and phylogeny of the mitochondrial genome of *Sabethes chloropterus*, *Sabethes glaucodaemon* and *Sabethes belisarioi* (Diptera: Culicidae). *Genomics* **111**:607–611 DOI 10.1016/j.ygeno.2018.03.016.

Deviatiiarov R, Kikawada T, Gusev O. 2017. The complete mitochondrial genome of an anhydrobiotic midge *Polypedilum vanderplanki* (Chironomidae, Diptera). *Mitochondrial DNA Part A* **28**:218–220 DOI 10.3109/19401736.2015.1115849.

Ekrem T, Willassen E. 2004. Exploring Tanytarsini relationships (Diptera: Chironomidae) using mitochondrial COII gene sequences. *Insect Systematics & Evolution* **35**:263–276 DOI 10.1163/187631204788920248.

Ekrem T, Willassen E, Stur E. 2010. Phylogenetic utility of five genes for dipteran phylogeny: a test case in the Chironomidae leads to generic synonymies. *Molecular Phylogenetics and Evolution* **57**:561–571 DOI 10.1016/j.ympev.2010.06.006.

Grant JR, Stothard P. 2008. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Research* **36**:181–184 DOI 10.1093/nar/gkn179.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research* **41**:e129–e129 DOI 10.1093/nar/gkt371.

Hassanin A, Leger N, Deutsch J. 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Systematic Biology* **54**:277–298 DOI 10.1080/10635150590947843.

Hu T, Banzhaf W. 2008. Nonsynonymous to synonymous substitution ratio ka/ks: measurement for rate of evolution in evolutionary computation. In: *International conference on parallel problem solving from nature*. Dortmund: Springer, 448–457.

Jacobsen MW, Hansen MM, Orlando L, Bekkevold D, Bernatchez L, Willerslev E, Gilbert MTP. 2012. Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.). *Molecular Ecology* **21**:2727–2742 DOI 10.1111/j.1365-294X.2012.05561.x.

Jiang X, Han X, Liu Q, Hou X. 2019. The mitochondrial genome of *Forcipomyia makanensis* (Insecta: Diptera: Ceratopogonidae). *Mitochondrial DNA Part B* **4**:344–345 DOI 10.1080/23802359.2018.1544048.

Kang Z, Li X, Yang D. 2016. The complete mitochondrial genome of *Dixella* sp. (Diptera: Nematocera, Dixidae). *Mitochondrial DNA Part A* **27**:1528–1529 DOI 10.3109/19401736.2014.953123.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**:772–780 DOI 10.1093/molbev/mst010.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–1649 DOI 10.1093/bioinformatics/bts199.

Kim S, Kim H, Shin SC. 2016. Complete mitochondrial genome of the Antarctic midge *Parochlus steinenii* (Diptera: Chironomidae). *Mitochondrial DNA Part A* **27**:3475–3476 DOI 10.3109/19401736.2015.1066355.

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**:1547–1549 DOI 10.1093/molbev/msy096.

Kutty SN, Wong WH, Meusemann K, Meier R, Cranston PS. 2018. A phylogenomic analysis of Culicomorpha (Diptera) resolves the relationships among the eight constituent families. *Systematic Entomology* **43**:434–446 DOI 10.1111/syen.12285.

Lanfear R, Frandsen PB, Wright AM, 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**:772–773 DOI 10.1093/molbev/msw260.

Li H, Liu H, Shi A, Štys P, Zhou X, Cai W. 2012. The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). *PLOS ONE* **7**:e29419 DOI 10.1371/journal.pone.0029419.

Li X-Y, Yan L-P, Pape T, Gao Y-Y, Zhang D. 2020. Evolutionary insights into bot flies (Insecta: Diptera: Oestridae) from comparative analysis of the mitochondrial genomes. *International Journal of Biological Macromolecules* **149**:371–380 DOI 10.1016/j.ijbiomac.2020.01.249.

Lu C, Huang X, Deng J. 2020. The challenge of Coccidae (Hemiptera: Coccoidea) mitochondrial genomes: the case of *Saissetia coffeae* with novel truncated tRNAs and gene rearrangements. *International Journal of Biological Macromolecules* **158**:854–864 DOI 10.1016/j.ijbiomac.2020.04.257.

Matsumoto Y, Yanase T, Tsuda T, Noda H. 2009. Species-specific mitochondrial gene rearrangements in biting midges and vector species identification. *Medical and Veterinary Entomology* **23**:47–55 DOI 10.1111/j.1365-2915.2008.00789.x.

Miao X, Huang J, Menzel F, Wang Q, Wei Q, Lin X-L, Wu H. 2020. Five mitochondrial genomes of black fungus gnats (Sciaridae) and their phylogenetic implications. *International Journal of Biological Macromolecules* **150**:200–205 DOI 10.1016/j.ijbiomac.2020.01.271.

Ojala D, Montoya J, Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**:470–474.

Park K, Jo H, Choi B, Kwak I-S. 2020. Complete mitochondrial genome of *Stictochironomus akizukii* (Tokunaga) (Chironomidae, Diptera) assembled from next-generation sequencing data. *Mitochondrial DNA Part B* **5**:2310–2311 DOI 10.1080/23802359.2020.1750320.

Peng Y, Leung HC, Yiu S-M, Chin FY. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**:1420–1428 DOI 10.1093/bioinformatics/bts174.

Ramakodi MP, Singh B, Wells JD, Guerrero F, Ray DA. 2015. A 454 sequencing approach to dipteran mitochondrial genome research. *Genomics* **105**:53–60 DOI 10.1016/j.ygeno.2014.10.014.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**:901–904 DOI 10.1093/sysbio/syy032.

Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**:539–542 DOI 10.1093/sysbio/sys029.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34**:3299–3302 DOI 10.1093/molbev/msx248.

Sæther OA. 1977. Female genitalia in Chironomidae and other Nematocera: morphology, phylogenies, keys. *Bulletin of the Fisheries Research Board of Canada* **197**:1–209.

Sæther OA. 2000. Phylogeny of the subfamilies of Chironomidae (Diptera). *Systematic Entomology* **25**:393–403 DOI 10.1046/j.1365-3113.2000.00111.x.

Siddall ME, Whiting MF. 1999. Long-branch abstractions. *Cladistics* **15**:9–24.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312– DOI 10.1093/bioinformatics/btu033.

Stokkan M, Jurado-Rivera JA, Oromí P, Juan C, Jaume D, Pons J. 2018. Species delimitation and mitogenome phylogenetics in the subterranean genus *Pseudoniphargus* (Crustacea: Amphipoda). *Molecular Phylogenetics and Evolution* **127**:988–999 DOI 10.1016/j.ympev.2018.07.002.

Tang L, Yan L, Gao Y, Zhang D. 2019a. First report of mitochondrial genome from the subfamily Bengaliinae (Diptera: Calliphoridae). *Mitochondrial DNA Part B* **4**:1560–1561 DOI 10.1080/23802359.2019.1601037.

Tang P, Zhu J-C, Zheng B-Y, Wei S-J, Sharkey M, Chen X-X, Vogler AP. 2019b. Mitochondrial phylogenomics of the Hymenoptera. *Molecular Phylogenetics and Evolution* **131**:8–18 DOI 10.1016/j.ympev.2018.10.040.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* **27**:171–180 DOI 10.1111/j.1096-0031.2010.00329.x.

Wei S-J, Shi M, Chen X-X, Sharkey MJ, Van Achterberg C, Ye G-Y, He J-H. 2010. New views on strand asymmetry in insect mitochondrial genomes. *PLOS ONE* **5**:e12708 DOI 10.1371/journal.pone.0012708.

Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and evolution. *International Review of Cytology* **141**:173–216 DOI 10.1016/S0074-7696(08)62066-5.

Xia XH. 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution* **30**:1720–1728 DOI 10.1093/molbev/mst064.

Yan L, Pape T, Elgar MA, Gao Y, Zhang D. 2019. Evolutionary history of stomach bot flies in the light of mitogenomics. *Systematic Entomology* **44**:797–809 DOI 10.1111/syen.12356.

Yang W, Yu W, Du Y. 2013. The complete mitochondrial genome of the sycamore lace bug *Corythucha ciliata* (Hemiptera: Tingidae). *Gene* **532**:27–40 DOI 10.1016/j.gene.2013.08.087.

Zhang D, Yan L, Zhang M, Chu H, Cao J, Li K, Hu D, Pape T. 2016. Phylogenetic inference of calyptrates, with the first mitogenomes for Gasterophilinae (Diptera: Oestridae) and Paramacronychiinae (Diptera: Sarcophagidae). *International Journal of Biological Sciences* **12**:489–504 DOI 10.7150/ijbs.12148.

Zhang H-L, Ye F. 2017. Comparative mitogenomic analyses of praying mantises (Dictyoptera, Mantodea): origin and evolution of unusual intergenic gaps. *International Journal of Biological Sciences* **13**:367–382 DOI 10.7150/ijbs.17035.

Zhang L-P, Cai Y-Y, Yu D-N, Storey KB, Zhang J-Y. 2018. Gene characteristics of the complete mitochondrial genomes of *Paratoxodera polyacantha* and *Toxodera hauseri* (Mantodea: Toxoderidae). *PeerJ* **6**:e4595 DOI 10.7717/peerj.4595.

Zhang X, Kang Z, Ding S, Wang Y, Borkent C, Saigusa T, Yang D. 2019b. Mitochondrial genomes provide insights into the phylogeny of Culicomorpha (Insecta: Diptera). *International Journal of Molecular Sciences* **20**:747 DOI 10.3390/ijms20030747.

Zhang Q, Xu W, Peng K, Zou L, Li Y, Chen Y, Cai Y, Gong Z. 2019a. The complete mitochondrial genome of *Propsilocerus akamusi* (Diptera, Chironomidae). *Mitochondrial DNA Part B* **4**:3983–3984 DOI 10.1080/23802359.2019.1688703.

Zheng C, Ye Z, Zhu X, Zhang H, Dong X, Chen P, Bu W. 2020. Integrative taxonomy uncovers hidden species diversity in the rheophilic genus *Potamometra* (Hemiptera: Gerridae). *Zoologica Scripta* **49**:174–186 DOI 10.1111/zsc.12401.