



Research Article

The first complete mitochondrial genome of Matsucoccidae (Hemiptera, Coccoidea) and implications for its phylogenetic position

Kai Hu[‡], Sanpeng Yu[‡], Niannian Zhang[‡], Maojuan Tian[‡], Qiming Ban[§], Zhongliang Fan[|], Jiansheng Qiu[‡]

[‡] Guizhou Academy of Forestry, Guiyang, China

[§] Forestry Administration of Wangmo County, Qianxinan, China

[|] Forestry Administration of Huichuan District, Zunyi, China

Corresponding author: Jiansheng Qiu (qiu1803@qq.com)

Academic editor: Jader Oliveira

Received: 15 Sep 2022 | Accepted: 23 Oct 2022 | Published: 09 Nov 2022

Citation: Hu K, Yu S, Zhang N, Tian M, Ban Q, Fan Z, Qiu J (2022) The first complete mitochondrial genome of Matsucoccidae (Hemiptera, Coccoidea) and implications for its phylogenetic position. Biodiversity Data Journal 10: e94915. <https://doi.org/10.3897/BDJ.10.e94915>

Abstract

The mitochondrial genome (mitogenome) has been extensively used to better understand the phylogenetic relationships within the hemipteran suborder Sternorrhyncha, but sequenced mitogenomes remain unavailable for the entire family Matsucoccidae to date. To address this, here we sequenced the complete mitogenome of *Matsucoccus matsumurae*; the first for this family. The mitogenome is 15,360 bp in size and comprises the typical set of 37 mitochondrial genes and a large non-coding region (AT-rich region). Gene order, nucleotide composition and codon usage of protein-coding genes (PCGs) of *M. matsumurae* differ considerably from those of the other two sequenced Coccidae species. All PCGs were initiated by the ATN start codons and ended with the TAA/G or single T-- stop codons. Nine transfer RNA genes could be folded into typical clover-leaf secondary structures. The length and AT content of the ribosomal RNA genes are highly conserved in the Coccoidea mitogenomes. In contrast, the AT-rich control region is highly variable in size and in the number of tandem repeats. The sliding window analysis showed that the *cox1* gene is the most conserved amongst the 13 PCGs, while the ratios of non-synonymous to synonymous substitution rates indicated that the evolution of this

mitogenome has been dominated by positive selection. Phylogenetic analyses, based on nucleotide sequence data of 37 mitochondrial genes and amino acid sequence data of 13 PCGs using Bayesian Inference and Maximum Likelihood methods, showed that Matsucoccidae diverged before the Coccidae.

Keywords

mitogenome, scale insect, phylogenetic analysis, secondary structure, Sternorrhyncha

Introduction

The family Matsucoccidae was erected in 1984 by Koteja (Koteja 1984, Koteja 1990) within the superfamily Coccoidea. It belongs to the Archaeococcoidae clade of the infraorder Cocomorpha, suborder Sternorrhyncha, order Hemiptera. So far, a total of 38 species in two genera of Matsucoccidae have been recorded worldwide (García Morales et al. 2018). Except for six fossil species described from Baltic amber, all 32 extant species exclusively occur on *Pinus* spp. in the Holarctic and Neotropical Regions (García Morales et al. 2018). Amongst them, some species are amongst the most destructive pests of pine trees, causing significant economic losses (McKenzie et al. 1948, Furniss and Carolin 1977, Foldi 2004, Lim et al. 2013, Liu et al. 2014). Based on the fossil evidence and morphological characteristics, Matsucoccidae is considered to be one of the most primitive families amongst the archaeococcoids and may even represent the most ancient of all scale insects (Coccoidea) (Beardsley 1968, Foldi 2004, Koteja and Azar 2008, Hodgson and Hardy 2013, Wang et al. 2016). However, this lacks the support of molecular data.

In insects, the mitochondrial genome (mitogenome) is typically a covalently closed circular double-stranded DNA molecule, usually 15 ~ 18 kb in length, that encodes 37 genes, including 13 protein-coding genes (PCG), two ribosomal RNA genes (rRNA) and 22 transfer RNA genes (tRNA) (Boore 1999, Cameron 2014). In addition, the mitogenome usually includes a non-coding region of variable length that plays a regulatory role in the transcription and replication, namely, the mitochondrial control region (AT-rich region) (Clayton 1982, Boore 1999). The mitogenome, in whole or part, has been widely used as a molecular marker to study population genetics, phylogeny and genetic evolution of insects (Hu et al. 2019).

Currently (23 April 2022), there are no available mitogenomes for Matsucoccidae. In this study, we sequenced, annotated and analysed the mitogenome of *Matsucoccus matsumurae* (the type species of Matsucoccidae) in detail. Furthermore, we sampled all sternorrhynchan superfamilies to investigate the phylogenetic position of Matsucoccidae and also provide insight into the superfamily-level phylogenetic relationships within the suborder Sternorrhyncha.

Materials and methods

Sample collection and DNA extraction

Adult specimens of *M. matsumurae* (Fig. 1) were collected from *Pinus massoniana* in Guizhou Province, China (Suiyang County, 107°2'52.06"E, 27°52'10.02"N, September 2021). All fresh specimens were preserved in 100% ethyl alcohol immediately after collection in the field and deposited at -20°C in the laboratory of Guizhou Academy of Forestry, Guiyang, Guizhou. Identification of the specimens was based on morphological characteristics (Li 2016). The total DNA was extracted from thoracic muscles using the Biospin Insect Genomic DNA Extraction Kit (BioFlux) following the manufacturer's instructions. Voucher specimens are stored in the entomological collection of the Guizhou Academy of Forestry.



Figure 1. [doi](#)

An adult male of *Matsucoccus matsumurae* on *Pinus massoniana* (Guizhou Province, China).

Mitogenome sequencing and assembly

The whole genomic DNA of *M. matsumurae* was sequenced using next-generation sequencing (Illumina HiSeq X10, Biomarker Technologies Corporation, Beijing, China). About 2.13 Gb clean data were assembled into a complete circular mitogenome using NOVOPlasty v.4.3.1 (Dierckxsens et al. 2017) with the *cox1* gene of *Saissetia coffeae* (Lu et al. 2020) as the seed sequence.

Mitogenome annotation and analyses

The annotation of mitogenome was conducted using Geneious Prime v.2022.0.1 (Biomatters, Auckland, New Zealand). The locations and sequences of tRNA genes were determined by the MITOS Web Server (<http://mitos2.bioinf.uni-leipzig.de/index.py>) (Donath et al. 2019) and comparison with homologous mitogenome sequences. Secondary structures of tRNAs were plotted using Adobe Illustrator CC2017 according to MITOS results and manual predictions (following the expected cloverleaf secondary structure of tRNA genes). Two rRNA genes and 13 PCGs were identified, based on their alignments with the other scale insect (Coccoidea) mitogenome sequences. The mitogenomic circular map was depicted with the help of OrganellarGenomeDRAW (OGDRAW) (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) (Greiner et al. 2019).

The organisation tables, nucleotide composition and relative synonymous codon usage (RSCU) of the mitogenomes of Coccoidea species were calculated and produced using PhyloSuite v.1.2.2 (Zhang et al. 2020). The nucleotide diversity (π) analyses of 13 PCGs and two rRNA genes of the Coccoidea species and a sliding window analysis (a sliding window of 200 bp and step size of 20 bp), were conducted by DnaSP v.6.0 (Rozas et al. 2017). The non-synonymous substitution rates (K_a), synonymous substitution rates (K_s) and ratios of non-synonymous to synonymous substitution (K_a/K_s) for each of the concatenated 13 PCGs of the Coccoidea mitogenomes were also calculated by DnaSP v. 6.0. Tandem repeats of the control region were identified with the Tandem Repeats Finder server (<https://tandem.bu.edu/trf/trf.html>) (Benson 1999).

Phylogenetic analysis

A total of 34 mitogenomes from five superfamilies of Hemiptera were used for the phylogenetic analyses (Table 1). Of these, 32 species belong to all four superfamilies of Sternorrhyncha (the ingroup), while the remaining two species from the superfamily Fulgoroidea were chosen as outgroups. Nucleotide sequences (without stop codons) for the 13 PCGs were aligned using MAFFT v.7 (Katoch and Standley 2013) with the L-INS-i (accurate) strategy and codon alignment mode (Code table: Invertebrate mitochondrial genetic codon), rRNA and tRNA gene sequences were aligned using MAFFT v.7 (Katoch and Standley 2013) with the G-INS-I algorithm (which takes account of the secondary structure of rRNA and tRNA genes) and normal alignment mode and amino acid sequences of 13 PCGs were aligned using the -auto strategy and normal alignment mode. Ambiguously aligned areas were removed using Gblocks v.0.91b (Talavera and Castresana 2007), respectively. Individual gene alignments were concatenated using PhyloSuite v.1.2.2 (Zhang et al. 2020). For amino acid sequence data, Bayesian Inference (BI) phylogenetic analysis were conducted using PHYLOBAYES MPI v.1.5c (Lartillot et al. 2013) in the CIPRES Science Gateway (Miller et al. 2010) which employs the site-heterogeneous model CAT + GTR. Two independent Markov Chain Monte Carlo (MCMC) chains were run and the analysis was stopped when the two runs had satisfactorily converged (maxdiff. fell below 0.3). A consensus tree was computed from the remaining trees combined from two runs after the initial 25% trees from each MCMC chain run were

discarded as burn-in. For nucleotide sequence data, the best partitioning scheme and nucleotide substitution models for Maximum Likelihood (ML) and BI phylogenetic analyses were selected with PartitionFinder v.2 (Lanfear et al. 2017) using the Bayesian Information Criterion (BIC) (Suppl. materials 1, 2). ML analyses were conducted using IQ-TREE v.1.6.3 (Nguyen et al. 2015) under the ultrafast bootstrap (UFB) approximation approach (Minh et al. 2013) with 5,000 replicates. BI analysis was performed using MrBayes v.3.2.7a (Ronquist et al. 2012) in the CIPRES Science Gateway server (Miller et al. 2010) with four chains (one cold chain and three hot chains). Two independent runs of 5,000,000 generations were carried out with sampling every 1,000 generations. The first 25% of trees were discarded as burn-in. After the average standard deviation of split frequencies fell below 0.01, stationarity was assumed.

Table 1.

Mitogenomes of the 34 Hemiptera insects used in this study.

Superfamily	Family	Species	Accession number	Reference
Aleyrodoidea	Aleyrodidae	<i>Pealius machili</i>	MT015588	Zhang et al. (2020)
		<i>Bemisia tabaci</i>	MK386668	Kunz et al. (2019)
		<i>Aleurocanthus camelliae</i>	KU761949	Chen et al. (2016)
Psylloidea	Calophyidae	<i>Calophya schini</i>	MF431591	Percy et al. (2018)
		<i>Calophya californica</i>	NC_036302	Unpublished
	Carsidaridae	<i>Paracarsidara gigantea</i>	NC_038112	Percy et al. (2018)
		<i>Allocarsidara bakeri</i>	NC_038107	Percy et al. (2018)
	Homotomidae	<i>Mycopsylla gardenensis</i>	MF443235	Percy et al. (2018)
		<i>Homotoma ficus</i>	MG989227	Percy et al. (2018)
	Liviidae	<i>Diaphorina citri</i>	MF614824	Unpublished
		<i>Livia junci</i>	MG989230	Percy et al. (2018)
	Phacopteronidae	<i>Pseudophacopteron</i> sp.	MG989234	Percy et al. (2018)
	Psyllidae	<i>Cyamophila willieti</i>	MN364946	Song et al. (2019)
		<i>Freysuila caesalpiniae</i>	NC_038135	Percy et al. (2018)
	Trioziidae	<i>Trioza urticae</i>	MG989240	Percy et al. (2018)
<i>Pariaconus pele</i>		MG989233	Percy et al. (2018)	
Aphalaridae	<i>Anoeconeossa unicornuta</i>	NC_038108	Percy et al. (2018)	
	<i>Blastopsylla occidentalis</i>	NC_038147	Percy et al. (2018)	
Aphidoidea	Anoeciidae	<i>Anoecia fulviabdominalis</i>	KP722588	Unpublished
	Aphididae	<i>Uroleucon sonchi</i>	MT533446	Unpublished
		<i>Aphis craccivora</i>	MT095075	Voronova et al. (2020)
		<i>Brevicoryne brassicae</i>	MT900510	An et al. (2021)
	Hormaphididae	<i>Hamamelistes spinosus</i>	NC_050942	Lu et al. (2020)
		<i>Pseudoregma bambucicola</i>	MK847518	Zhang et al. (2019)

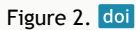
Superfamily	Family	Species	Accession number	Reference
	Mindaridae	<i>Mindarus keteleerifoliae</i>	NC_033410	Unpublished
	Pemphigidae	<i>Schlechtendalia peitan</i>	NC_059063	Unpublished
		<i>Melaphis rhois</i>	KY624581	Ren and Wen (2017)
	Phloeomyzidae	<i>Phloeomyzus passerinii</i>	KP722571	Unpublished
	Thelaxidae	<i>Kurisakia onigurumii</i>	KP722578	Unpublished
Coccoidea	Coccidae	<i>Didesmococcus koreanus</i>	NC_057479	Xu et al. (2021)
		<i>Saissetia coffeae</i>	MN863803	Lu et al. (2020)
	Matsucoccidae	<i>Matsucoccus matsumurae</i>	OM396907	This study
Fulgoroidea	Achilidae	<i>Betatropis formosana</i>	MH324927	Unpublished
	Caliscelidae	<i>Youtuus erythrus</i>	NC_059811	Gong et al. (2021)

Results and discussion

Mitogenome organisation and nucleotide composition

The complete mitogenome of *M. matsumurae* is a closed circular double-stranded DNA molecule (Fig. 2, Suppl. material 3), 15,360 bp in length. This is medium-sized amongst the available mitogenomes of Coccoidea: from 15,143 bp for *Didesmococcus koreanus* to 15,389 bp for *Saissetia coffeae* (Lu et al. 2020, Xu et al. 2021). The newly-sequenced mitogenome encodes 37 genes (13 PCGs, 22 tRNAs and two rRNAs) and an A+T-rich region (control region), which is a typical architecture for bilaterian animal mitogenomes (Boore 1999). The gene order of *M. matsumurae* (Matsucoccidae) differs considerably from that of the other two Coccidae species (Lu et al. 2020, Xu et al. 2021), but it is comparatively similar to gene orders of most Cimicomorpha, Fulgoroidea, Membracoidea, Psylloidea and Aphidoidea species (only the tRNA cluster of *trnW-trnC-trnY* was rearranged to *trnW-trnY-trnC*) (Fig. 3). Amongst the three scale insect species, twenty-three genes (nine PCGs and 14 tRNAs) are encoded on the majority strand (H strand) and the remaining 14 genes (four PCGs, eight tRNAs and two rRNAs) are encoded on the minority strand (L strand) (Fig. 3). A total of 85 overlapping nucleotides were found in nine pairs of neighbouring genes, with the longest identified overlap between the *trnN* and *trnS1* (22 bp). Furthermore, there are 193 intergenic nucleotides dispersed across 16 gene boundaries. The longest intergenic region (103 bp) is located between *cox3* and *trnG*.

The overall nucleotide composition of the *M. matsumurae* mitogenome is 47.2% A, 6.4% C, 0.6% G and 45.8% T. It therefore exhibits a strong AT bias of 91.1%, which is higher than in other Coccoidea insects (82.5% for *D. koreanus* and 84.7% for *S. coffeae*) (Table 2) (Lu et al. 2020, Xu et al. 2021). The PCGs have the lowest AT content (90.3%) and the control region has the highest (92.9%), which differs from all previously-sequenced scale insects (Table 2) (Lu et al. 2020, Xu et al. 2021). The new mitogenome exhibits negative GC-skews (-0.267) and positive AT-skews (0.056), which is common for scale insects (Table 2) and the Hemiptera in general (Wang et al. 2015, Chen et al. 2019).



Nucleotide diversity and selection pressures

Nucleotide diversity (Pi value) of 13 PCGs and two rRNAs amongst the three scale insect species is shown in Fig. 4. Nucleotide diversity values, calculated for individual genes ranged from 0.249 (*cox1*) to 0.429 (*nad6*). The results indicated that *atp8* (Pi = 0.418) and *nad6* (Pi = 0.429) had comparatively high nucleotide diversity; *atp6*, *cox2*, *cox3*, *cytb*, *nad1*, *nad2*, *nad3*, *nad4l*, *nad4*, *nad5* and *rrnS* (Pi = 0.300-0.384) exhibited an intermediate nucleotide diversity; and *cox1* (Pi = 0.249) and *rrnL* (Pi = 0.286) were highly conserved genes, with low nucleotide diversity (Fig. 4). The lowest nucleotide diversity in *cox1* is in agreement with observations in most sequenced insect mitogenomes (Ma et al. 2019, Ge et al. 2021, Li et al. 2021, Zhang et al. 2022), which presumably accounts for its ability to provide reliable species identification.

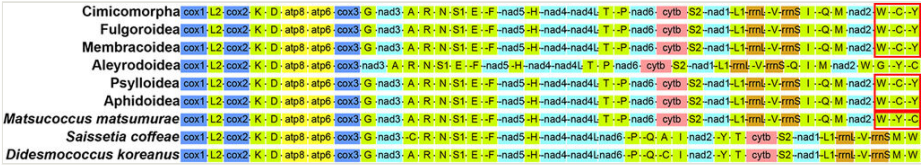


Figure 3. [doi](#)

Gene orders in Hemiptera mitogenomes. With the exception of Coccoidea (*Didesmococcus koreanus*, *Matsucoccus matsumurae* and *Saissetia coffeae*), each superfamily (plus the infraorder Cimicomorpha) is represented by one species. Genes shown with “-” signs are located on the minor strand (L-strand), while others are located on the major strand (H-strand).

Table 2.

Nucleotide composition and skewness of mitogenomes of *Didesmococcus koreanus*, *Matsucoccus matsumurae* and *Saissetia coffeae*.

Feature	Length	A+T%	AT-skew	GC-skew
<i>D. koreanus</i> / <i>M. matsumurae</i> / <i>S. coffeae</i>				
Whole genome	15143/15360/15389	82.5/91.1/84.7	0.142/0.056/0.189	-0.357/-0.267/-0.369
PCGs	10599/10626/10632	81.9/90.3/84.1	-0.084/-0.115/-0.072	-0.108/0.016/-0.107
tRNAs	1300/1330/1373	87.9/92.3/89.3	0.043/0.054/0.054	0.019/-0.059/-0.007
rRNAs	1914/1890/1965	86.2/92.7/87.2	-0.027/-0.062/-0.099	0.426/0.304/0.440
Control region	1350/1404/1536	77.7/92.9/82.1	-0.071/0.015/-0.045	-0.252/-0.818/-0.295

The non-synonymous/synonymous (Ka/Ks) substitution ratio can be used to estimate whether a sequence is undergoing purifying, neutral or positive selection. The rates of non-synonymous (Ka) and synonymous substitutions (Ks) and their ratio (Ka/Ks) were calculated for the 13 PCGs of each of the three scale insect species using *Aphis craccivora* as the reference sequence (Fig. 5). A value of Ka/Ks greater than 1 implies that a gene is evolving predominantly under positive selection. This indicates that non-synonymous mutations are favoured by the Darwinian selection, i.e. that they are retained at a rate greater than synonymous mutations. All Ka/Ks values were above 1, which strongly suggests the presence of positive selection in these species.

Protein-coding genes

The 13 PCGs (length: 10,626 bp) account for 69.2% of the complete mitogenome of *M. matsumurae*. All PCGs were initiated by the typical start codon ATN (ATA/T/G/C) and ended with the TAA/G stop codon or their incomplete form T-. This is almost identical to *D. koreanus* (Xu et al. 2021) and *S. coffeae* (Lu et al. 2020). Such incomplete stop codons are common in insects and believed to be completed by post-transcriptional polyadenylation (Ojala et al. 1981). The AT-skews of PCGs were similar (-0.115 to -0.072) amongst the three available scale insects (Table 2). In *M. matsumurae*, the 13 PCGs encode a total of 3,531 amino acids, amongst which the most frequently used are Leu, Met, Ile and Phe

(accounting for 56.65% of the total amount), whereas Cys is the least frequently used (0.34%) (Fig. 6). These amino acid usage patterns are very consistent in Coccoidea species (Lu et al. 2020, Xu et al. 2021). Relative synonymous codon usage (RSCU) is summarised in Fig. 7. Results indicate that the four most frequently used codons are UUA (Leu), AUU (Ile), UUU (Phe) and AUA (Met). All of them are composed solely of A or U, which is reflected in the high A+T content of PCGs. Excluding TAA and TAG, 61 codons were observed in *D. koreanus* (no CGC), 46 in *M. matsumurae* (no CUG, GUC, GUG, CCC, CCG, ACG, GCC, GCG, CAG, UGC, CGC, CGG, AGC, AGG, GGC and GGG), and 56 in *S. coffeae* (no CUC, ACG, GCG, UGC, CGC and GGC) (Lu et al. 2020, Xu et al. 2021).

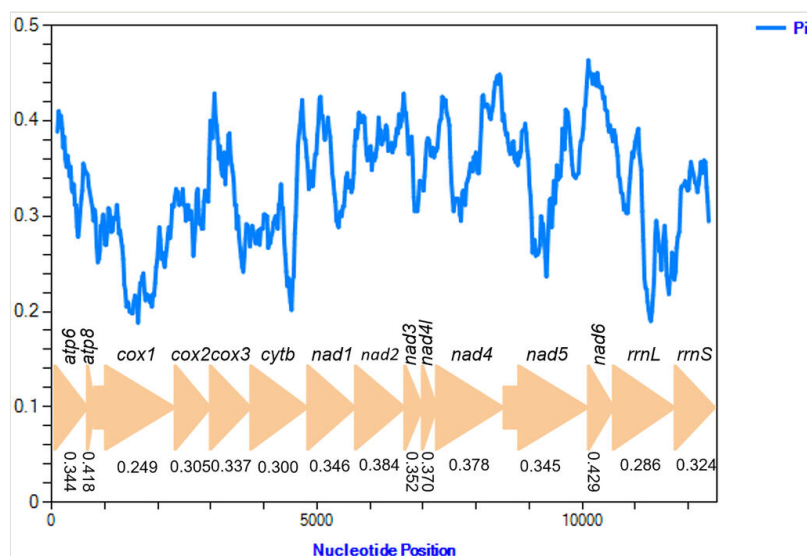


Figure 4. [doi](#)

Sliding window analysis of 13 PCGs and two rRNAs of three Coccoidea species. The blue line shows the value of nucleotide diversity Pi (window size = 200 bp, step size = 20 bp). The Pi value for each gene is shown in the graph.

Transfer RNA genes

All three sequenced scale insect mitogenomes encode 22 tRNA genes (Suppl. material 3) (Lu et al. 2020, Xu et al. 2021). The AT content of tRNA genes is slightly higher than that of the PCGs, ranging from 87.9% (*D. koreanus*) to 92.3% (*M. matsumurae*) (Table 2). Furthermore, the AT-skew values of tRNAs were all greater than zero (0.043 to 0.054). The length of the 22 tRNA genes ranged from 46 bp (*trnS1* of *D. koreanus*) to 78 bp (*trnS1* of *M. matsumurae*) (Suppl. material 3) (Lu et al. 2020, Xu et al. 2021). Most tRNAs lack either the DHU or T ψ C arm and some even lost both DHU and T ψ C arms (Fig. 8) (Lu et al. 2020, Xu et al. 2021). Only ten tRNAs of *D. koreanus*, nine of *M. matsumurae* and 11 of *S. coffeae* could be folded into the common clover-leaf secondary structures (Fig. 8) (Lu et al. 2020, Xu et al. 2021). This suggests that the reduction of DHU or T ψ C arms of tRNA genes

could be a very common phenomenon in the mitogenomes of scale insects. The anticodons of tRNA genes are identical amongst the three scale insects (Fig. 8) (Lu et al. 2020, Xu et al. 2021), except for the *trnE* of *D. koreanus* and *trnK* of *S. coffeae*, which employ UCG and UUU instead of UUC and CUU respectively.

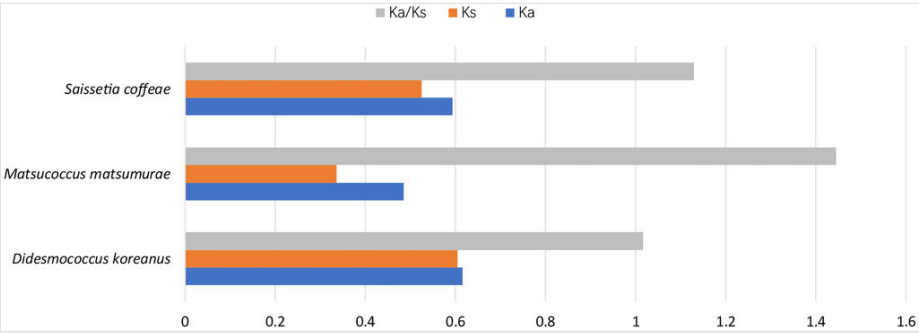


Figure 5. [doi](#)
Evolutionary rates of mitochondrial genomes in the three scale insects. The rate of non-synonymous substitutions (Ka), synonymous substitutions (Ks) and the ratio of Ka/Ks were calculated for the PCGs of each mitogenome, using *Aphis craccivora* as the reference sequence.

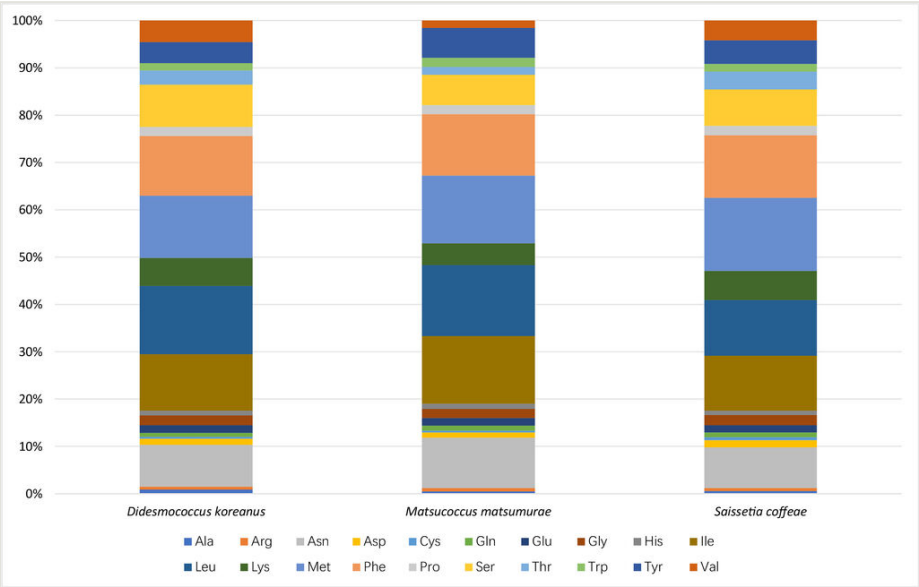


Figure 6. [doi](#)
Percentages of each amino acid within the three scale insects. The stop codon is not included.

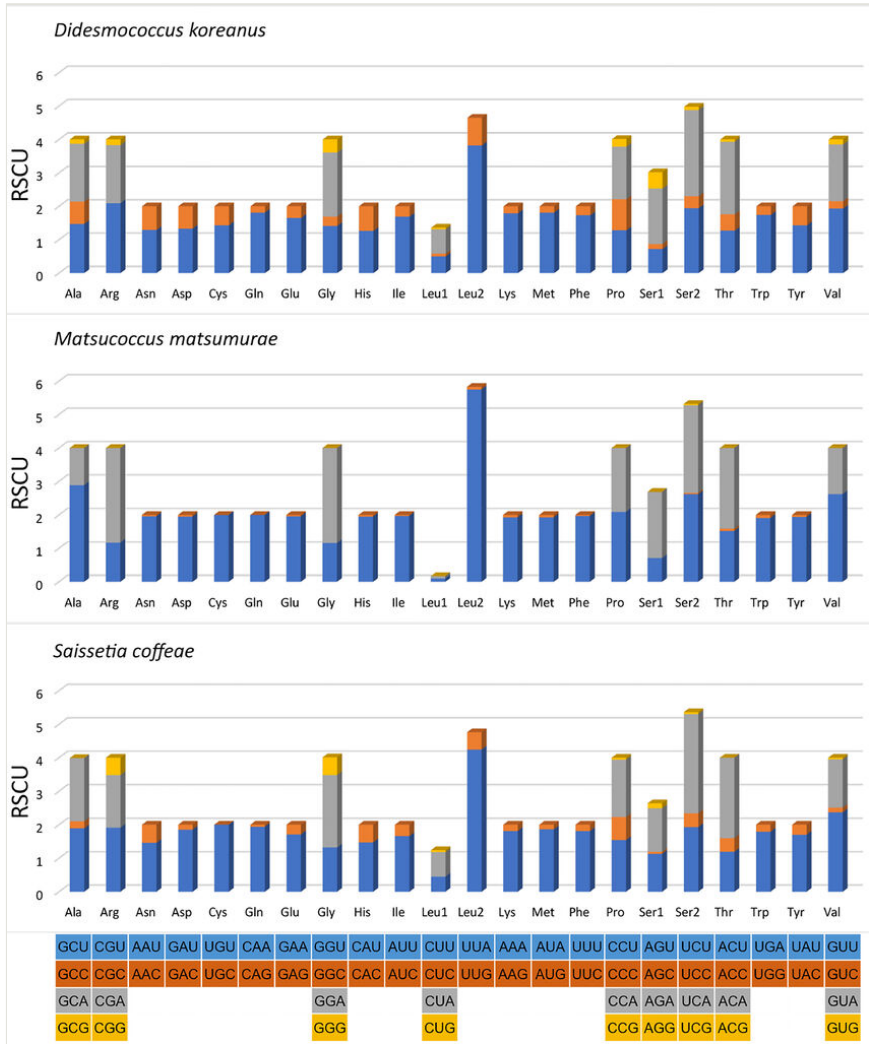


Figure 7. [doi](#)

Relative synonymous codon usage (RSCU) in mitogenomes of the three scale insects. The stop codon is not shown.

Ribosomal RNA genes

The AT nucleotide content of *rrnS* and *rrnL* genes of the three scale insects ranges from 86.2% to 92.7% (Table 2). Two rRNAs in these three mitogenomes show a negative AT-skew (-0.099 to -0.027) (Table 2). The *rrnL* gene, located between *trnL1* and *trnV*, ranged in length from 1,161 bp to 1,177 bp and *rrnS* gene, located between *trnV* and *trnI* (*M. matsumurae*) or *trnM* (*D. koreanus* and *S. coffeae*), ranged from 713 bp to 804 bp in length (Fig. 3, Suppl. material 3) (Lu et al. 2020, Xu et al. 2021). Therefore, the length and AT content of rRNAs are conserved in the Coccoidea.

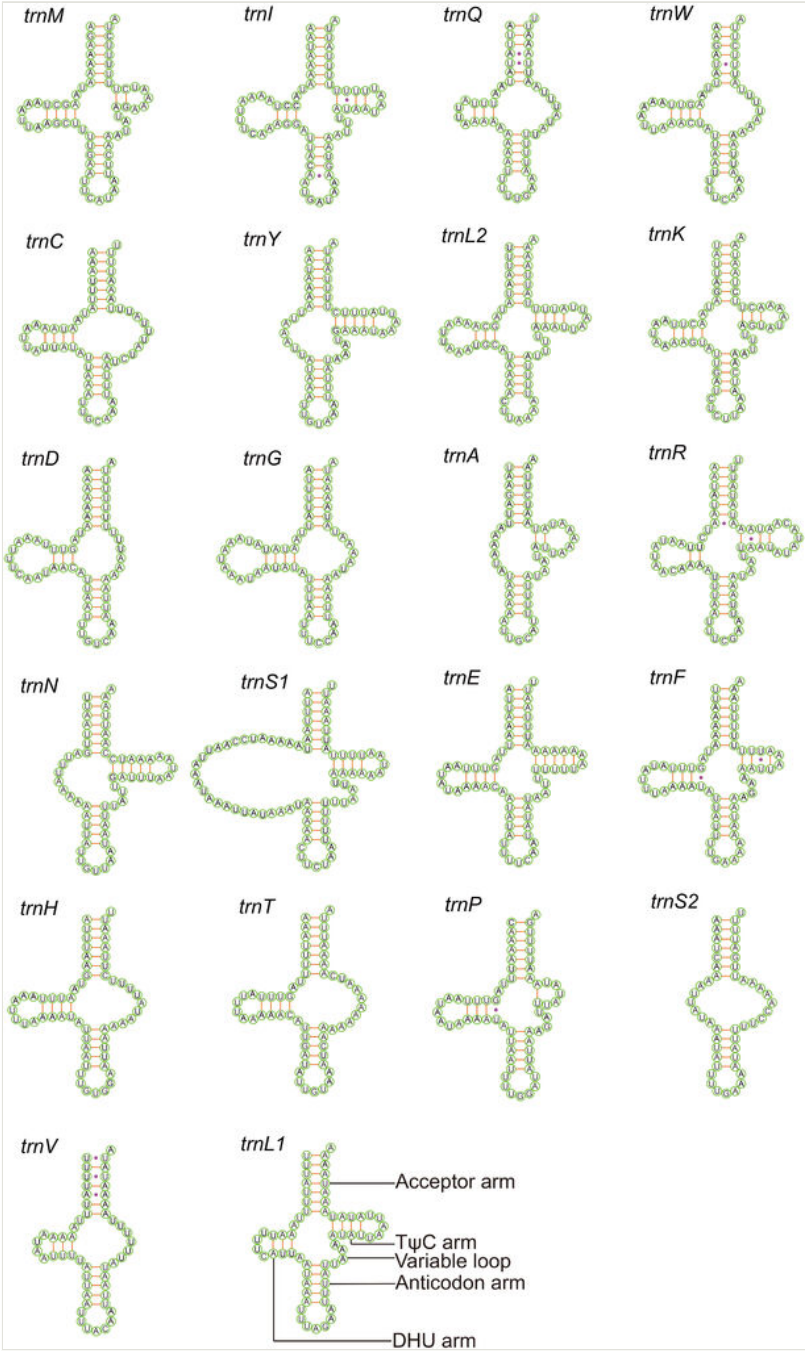


Figure 8. [doi](#)
Secondary structures of 22 tRNAs in the mitogenome of *Matsucoccus matsumurae*. Lines (-) indicate Watson-Crick base pairings, whereas dots (·) indicate unmatched base pairings.

AT-rich region

The AT-rich region is believed to be involved in regulating the transcription and replication of DNA in insects (Clayton 1982, Cameron 2014). The AT-rich region of *M. matsumurae* is located between *rrnS* and *trnI*, whereas the AT-rich regions of *D. koreanus* and *S. coffeae* are located between *rrnS* and *trnM* and the size of three scale insects ranges from 1,271 bp to 2,261 bp (Fig. 3, Suppl. material 3) (Lu et al. 2020, Xu et al. 2021). Analyses of the AT-rich regions by the Tandem Repeats Finder indicated that *M. matsumurae* and *S. coffeae* have different numbers of absolute tandem repeat units. Two types of absolute tandem repeats were present in *M. matsumurae* (nucleotide positions: 7 to 1,023 and 1,026 to 1,373 of the AT-rich region). The AT-rich region of *S. coffeae* had only one kind of absolute tandem repeat, located between the nucleotide positions 1,096 to 1,533 (Fig. 9). Like in most insect mitogenomes, tandem repeats are common and the size of tandem repeat regions varies depending on the number of copies of the repeating units (Garey and Wolstenholme 1989). Tandem repeats are thought to play an important role in the control of DNA methylation, gene transcription and replication (Zhang et al. 1995, Huang et al. 2013).

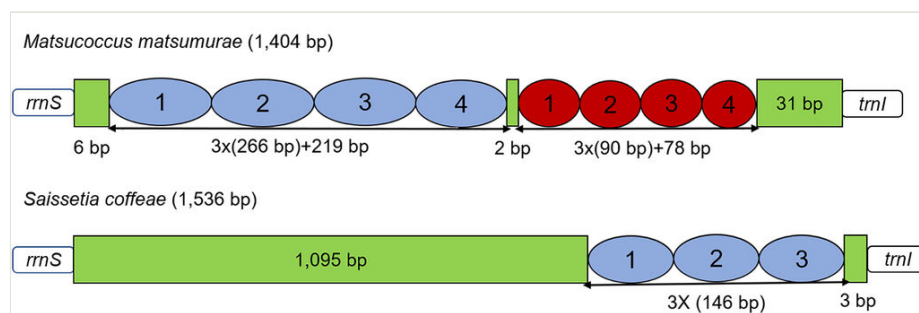


Figure 9. [doi](#)

Structures of AT-rich regions in mitogenomes of the three scale insects. The location and copy number of absolute tandem repeat units are represented by blue and red ovals. Green boxes indicate non-repeat regions.

Phylogenetic relationships

Phylogenetic analyses of 32 species of Sternorrhyncha, including two outgroups, based on ML and BI analyses of nucleotide sequence data of 37 mitochondrial genes, yielded largely congruent topologies, with most branches receiving strong support (Figs 10, 11). As proposed by previous studies (Lu et al. 2020, Xu et al. 2021), the relationships amongst sternorrhynchan superfamilies are inferred as ((Coccoidea + Aphidoidea) + (Aleyrodoidea + Psylloidea)).

All analyses consistently recovered the monophyly of Sternorrhyncha and its four superfamilies (Coccoidea, Aphidoidea, Aleyrodoidea and Psylloidea) with strong support (BS = 86; PP = 1.00) (Figs 10, 11Suppl. material 4). Furthermore, *M. matsumurae*

clustered with the other two coccoid species, *D. koreanus* and *S. coffeae*, with strong support (BS = 100; PP = 1.00) (Figs 10, 11Suppl. material 4), thus confirming its affiliation with the superfamily Coccoidea. Within the clade of Coccoidea, *M. matsumurae* is positioned as the earliest branching taxon in both nucleotide and amino acid trees, which suggests that Matsucoccidae is a more ancient taxon than Coccidae.

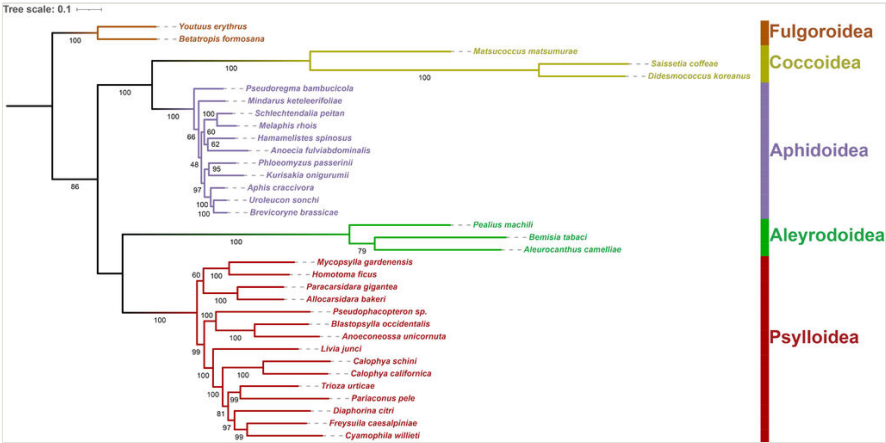


Figure 10. [doi](#)
ML phylogenetic tree for Sternorrhyncha, based on the nucleotide sequence data of 37 mitochondrial genes from *Matsucoccus matsumurae* and other 33 species belonging to five superfamilies of Hemiptera. Bootstrap support values (BS) are indicated on the branches.

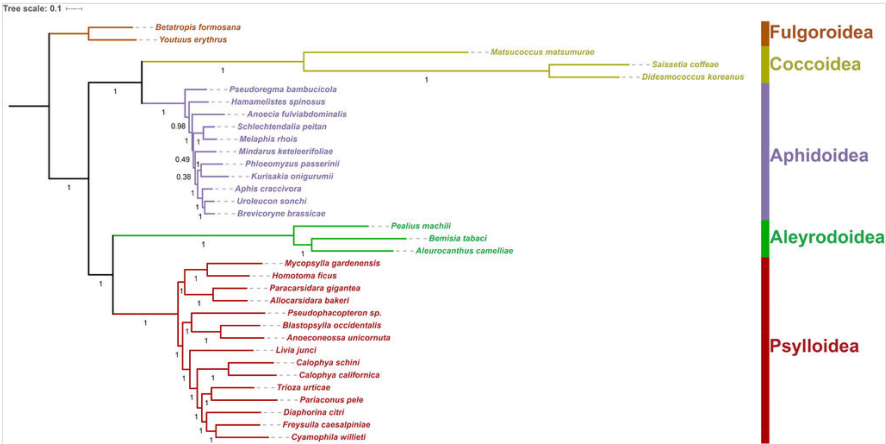


Figure 11. [doi](#)
BI phylogenetic tree for Sternorrhyncha, based on the nucleotide sequence data of 37 mitochondrial genes from *Matsucoccus matsumurae* and other 33 species belonging to five superfamilies of Hemiptera. Bayesian posterior probabilities (PP) are indicated on the branches.

Funding

This study was supported by the Projects of Guizhou Science and Technology Platform and Talent Team (QKHPTRC [2016] 5669) and Survey and Research on Exotic Species in Guizhou Province (AS [2017] 09, KJZXSA [2018] 016, and KJZXSA [2019] 014).

Acknowledgements

We thank Ivan Jakovlić (Bio-Transduction Lab, Wuhan, China) for proofreading of the manuscript and giving us valuable advice.

Author contributions

Investigation, formal analysis and writing—original draft, Kai Hu; resources, Sanpeng Yu, Maojuan Tian, Qiming Ban and Zhongliang Fan; writing—review and editing, Kai Hu, Sanpeng Yu, Niannian Zhang and Jiansheng Qiu; and validation, project administration and funding acquisition, Jiansheng Qiu. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

All authors report no conflicting interests.

References

- An L, Fang Z, Weng QB (2021) The complete mitochondrial genome of *Brevicoryne brassicae* (Hemiptera: Aphididae). Mitochondrial DNA Part B-Resources 6 (3): 974-975. <https://doi.org/10.1080/23802359.2021.1888334>
- Beardsley JW (1968) External morphology of the adult male of *Matsucoccus bisetosus*. Annals of the Entomological Society of America 61 (6): 1449-1459. <https://doi.org/10.1093/aesa/61.6.1449>
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27 (2): 573-580. <https://doi.org/10.1093/nar/27.2.573>
- Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Research 27 (8): 1767-1780. <https://doi.org/10.1093/nar/27.8.1767>
- Cameron SL (2014) Insect mitochondrial genomics: implications for evolution and phylogeny. Annual Review of Entomology 59: 95-117. <https://doi.org/10.1146/annurev-ento-011613-162007>
- Chen J, Wang Y, Qin M, Jiang LY, Qiao GX (2019) The mitochondrial genome of *Greenidea psidii* van der Goot (Hemiptera: Aphididae: Greenideinae) and comparisons with other Aphididae aphids. International Journal of Biological Macromolecules 122: 824-832. <https://doi.org/10.1016/j.ijbiomac.2018.10.209>

- Chen SC, Wang XQ, Li PW, Hu X, Wang JJ, Peng P (2016) The complete mitochondrial genome of *Aleurocanthus camelliae*: Insights into gene arrangement and genome organization within the family Aleyrodidae. *International Journal of Molecular Sciences* 17 (11): 1843. <https://doi.org/10.3390/ijms17111843>
- Clayton DA (1982) Replication of animal mitochondrial DNA. *Cell* 28: 693-705. [https://doi.org/10.1016/0092-8674\(82\)90049-6](https://doi.org/10.1016/0092-8674(82)90049-6)
- Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45 (4): e18.
- Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M (2019) Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Research* 47 (20): 10543-10552. <https://doi.org/10.1093/nar/gkz833>
- Foldi I (2004) The Matsuoccidae in the Mediterranean basin with a world list of species (Hemiptera: Sternorrhyncha: Coccoidea). *Annales de la Société Entomologique de France* 40 (2): 145-168. <https://doi.org/10.1080/00379271.2004.10697412>
- Furniss RL, Carolin VM (1977) *Western Forest Insects*. US Department of Agriculture: Forest Service, Washington DC. <https://doi.org/10.5962/bhl.title.131875>
- García Morales M, Denno BD, Miller DR, Miller GL, Ben-Dov Y, Hardy NB (2018) ScaleNet: A literature-Based Model of Scale Insect Biology and Systematics. <http://scalenet.info>
- Garey JR, Wolstenholme DR (1989) Platyhelminth mitochondrial DNA: Evidence for early evolutionary origin of a tRNA^{ser}AGN that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. *Journal of Molecular Evolution* 28: 374-387. <https://doi.org/10.1007/BF02603072>
- Ge XY, Yuan LL, Kang Y, Liu T, Liu HY, Yang YX (2021) Characterization of the first complete mitochondrial genome of Cyphonocerinae (Coleoptera: Lampyridae) with implications for phylogeny and evolution of fireflies. *Insects* 12 (7): 570. <https://doi.org/10.3390/insects12070570>
- Gong N, Yang L, Chen XS (2021) Structural features and phylogenetic implications of four new mitogenomes of Caliscelidae (Hemiptera: Fulgoromorpha). *International Journal of Molecular Sciences* 22 (3): 1348. <https://doi.org/10.3390/ijms22031348>
- Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47 (W1): W59-W64. <https://doi.org/10.1093/nar/gkz238>
- Hodgson CJ, Hardy NB (2013) The phylogeny of the superfamily Coccoidea (Hemiptera: Sternorrhyncha) based on the morphology of extant and extinct macropterous males. *Systematic Entomology* 38 (4): 794-804.
- Huang WX, Zheng JB, He Y, Luo C (2013) Tandem repeat modification during double-strand break repair induced by an engineered TAL effector nuclease in zebrafish genome. *PLOS One* 8 (12): e84176. <https://doi.org/10.1371/journal.pone.0084176>
- Hu K, Yuan F, Dietrich CH, Yuan XQ (2019) Structural features and phylogenetic implications of four new mitogenomes of Centrotinae (Hemiptera: Membracidae). *International Journal of Biological Macromolecules* 139: 1018-1027. <https://doi.org/10.1016/j.ijbiomac.2019.08.064>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30 (4): 772-780. <https://doi.org/10.1093/molbev/mst010>

- Koteja J (1984) The Baltic amber Matsucoccidae (Homoptera, Coccinea). *Annales Zoologici* 37: 437-49.
- Koteja J (1990) Paleontology, 149-163 In: Rosen, D. (Ed.), *Armored Scale Insects, their Biology, Natural Enemies and Control*, World Crop Pests, Vol. 4A. Elsevier, Amsterdam.
- Koteja J, Azar D (2008) Scale insects from Lower Cretaceous amber of Lebanon (Hemiptera: Sternorrhyncha: Coccinea). *Alavesia* 2: 133-167.
- Kunz D, Tay WT, Court LN, Elfekih S, Gordon KH, Evans GA, De Barro PJ (2019) Draft mitochondrial DNA genome of a 1920 Barbados cryptic *Bemisia tabaci* 'New World' species (Hemiptera: Aleyrodidae). *Mitochondrial DNA Part B-Resources* 4 (1): 1183-1184. <https://doi.org/10.1080/23802359.2019.1591197>
- 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 (3): 772-773.
- Lartillot N, Rodrigue N, Stubbs D, Richer J (2013) PHYLOBAYES MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biology* 62 (4): 611-615. <https://doi.org/10.1093/sysbio/syt022>
- Lim E, Kim DS, Lee SM, Choi KS, Lee DW, Chung YJ, Park CG (2013) Effect of fenitrothion on different life stages of black pine bast scale, *Matsucoccus thunbergianae* *Journal of Asia-Pacific Entomology* 16 (1): 55-59. <https://doi.org/10.1016/j.aspen.2012.10.001>
- Li R, Ma ZX, Zhou CF (2021) The first two complete mitochondrial genomes of Neoephemeridae (Ephemeroptera): Comparative analysis and phylogenetic implication for Furcatergala. *Genes* 12 (12): 1875. <https://doi.org/10.3390/genes12121875>
- Liu WM, Xie YP, Yang Q, Xue JL, Tian F (2014) New Research on *Matsucoccus matsumurae* (Kuwana) (Hemiptera: Matsucoccidae) in China. *Acta Zoologica Bulgarica* S6: 95-102.
- Li Z, et al. (2016) *Zhongguo yuanlin zhiwu jiechong*, 1st ed. Sichuan Publishing House of Science & Technology, Chengd.
- Lu CC, Huang XL, 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. <https://doi.org/10.1016/j.ijbiomac.2020.04.257>
- Lu ZH, Liang YK, Wen J, Ren ZM (2020) Complete mitochondrial genome of the witch-hazel leaf gall aphid *Hamamelistes spinosus* (Hemiptera: Aphididae: Hormaphidinae). *Mitochondrial DNA Part B-Resources* 5 (2): 1388-1389. <https://doi.org/10.1080/23802359.2020.1735965>
- Ma LY, Liu FF, Chiba H, Yuan XQ (2019) The mitochondrial genomes of three skippers: Insights into the evolution of the family Hesperidae (Lepidoptera). *Genomics* 112 (1): 432-441.
- McKenzie HL, Gill LS, Ellis DE (1948) The prescott scale (*Matsucoccus vexillorum*) and associated organisms that cause flagging injury to ponderosa pine in the southwest. *Journal of Agricultural Research* 76 (2): 33-51.
- Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES science gateway for inference of large phylogenetic trees. In proceedings of the gateway computing environments workshop (GCE). Institute of Electrical and Electronics Engineers (IEEE). URL: <https://www.phylo.org/portal2/login!input.action>

- Minh BQ, Nguyen MA, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30 (5): 1188-1195. <https://doi.org/10.1093/molbev/mst024>
- Nguyen LT, H.A. S, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32 (1): 268-274. <https://doi.org/10.1093/molbev/msu300>
- Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290: 470-474. <https://doi.org/10.1038/290470a0>
- Percy DM, Crampton-Platt A, Sveinsson S, Lemmon AR, Lemmon EM, Ouvrard D, Burckhardt D (2018) Resolving the psyllid tree of life: Phylogenomic analyses of the superfamily Psylloidea (Hemiptera). *Systematic Entomology* 43 (4): 762-776. <https://doi.org/10.1111/syen.12302>
- Ren ZM, Wen J (2017) Complete mitochondrial genome of the North American Rhus gall aphid *Melaphis rhois* (Hemiptera: Aphididae: Eriosomatinae). *Mitochondrial DNA Part B Resources* 2 (1): 169-170. <https://doi.org/10.1080/23802359.2017.1303345>
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna 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 (3): 539-542. <https://doi.org/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 LargeData Sets. *Molecular Biology and Evolution* 34 (12): 3299-3302. <https://doi.org/10.1093/molbev/msx248>
- Song XM, He YC, Wang XP, Gu X (2019) The complete mitochondrial genome of *Cyamophila willieti* (Wu) (Hemiptera: Psyllidae). *Mitochondrial DNA Part B Resources* 4 (2): 3758-3759. <https://doi.org/10.1080/23802359.2019.1681922>
- 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. <https://doi.org/10.1080/10635150701472164>
- Voronova NV, Levykina S, Warner D, Shulinski R, Bandarenka Y, Zhorov D (2020) Characteristic and variability of five complete aphid mitochondrial genomes: *Aphis fabae mordvilko*, *Aphis craccivora*, *Myzus persicae*, *Therioaphis tenera* and *Appendiseta robiniae* (Hemiptera; Sternorrhyncha; Aphididae). *International Journal of Biological Macromolecules* 149: 187-206. <https://doi.org/10.1016/j.ijbiomac.2019.12.276>
- Wang X, Xie YP, Zhang YF, Liu WM, Wu J (2016) The structure and morphogenic changes of antennae of *Matsucoccus matsumurae* (Hemiptera: Coccoidea: Matsucoccidae) in different instars. *Arthropod Structure & Development* 45 (3): 281-293. <https://doi.org/10.1016/j.asd.2016.01.010>
- Wang Y, Chen J, Jiang LY, Qiao GX (2015) Hemipteran mitochondrial genomes: features, structures and implications for phylogeny. *International Journal of Molecular Sciences* 16 (6): 12382-12404. <https://doi.org/10.3390/ijms160612382>
- Xu H, Liu XC, Li H, Wu SA (2021) The mitogenome of the scale insect *Didesmococcus koreanus* Borchsenius, 1955 (Coccoidea: Coccidae). *Mitochondrial DNA Part B Resources* 6 (4): 1298-1299. <https://doi.org/10.1080/23802359.2021.1906175>
- Zhang D, Gao FL, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data

management and evolutionary phylogenetics studies. *Molecular Ecology Resources* 20 (1): 348-355. <https://doi.org/10.1111/1755-0998.13096>

- Zhang DX, Szymura JM, Hewitt GM (1995) Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution* 40: 382-391. <https://doi.org/10.1007/BF00164024>
- Zhang H, Deng J, Liu Q, Huang XL (2019) The mitochondrial genome of a social aphid, *Pseudoregma bambucicola* (Hemiptera: Aphididae: Hormaphidinae). *Mitochondrial DNA Part B Resources* 4 (2): 2100-2101. <https://doi.org/10.1080/23802359.2019.1622470>
- Zhang H, Fang W, hao XY, Jiang X, Stroiński A, Qin DZ (2022) Comparative analysis of the complete mitochondrial genomes of five species of Ricaniidae (Hemiptera: Fulgoromorpha) and phylogenetic implications. *Biology-Basel* 11 (1): 92. <https://doi.org/10.3390/biology11010092>
- Zhang ZT, Yan X, Yang WJ, Jin DC (2020) Characterization of the complete mitochondrial genome of *Pealius machili* (Hemiptera: Aleyrodidae) with phylogenetic analysis. *Mitochondrial DNA Part B Resources* 5 (2): 1463-1464. <https://doi.org/10.1080/23802359.2020.1741468>

Supplementary materials

Suppl. material 1: Partitioning schemes and substitution models used for ML phylogenetic analyses_ [doi](#)

Authors: Kai Hu and Jiansheng Qiu

Data type: Table

Brief description: The best partitioning schemes and substitution models for PCG123 + tRNA + rRNA dataset comprising 13 PCGs, 22 tRNAs and two rRNAs of 34 species of Hemiptera used for ML phylogenetic analyses

[Download file](#) (18.02 kb)

Suppl. material 2: Partitioning schemes and substitution models used for BI phylogenetic analyses_ [doi](#)

Authors: Kai Hu and Jiansheng Qiu

Data type: Table

Brief description: The best partitioning schemes and substitution models for PCG123 + tRNA + rRNA dataset comprising 13 PCGs, 22 tRNAs and two rRNAs of 34 species of Hemiptera used for BI phylogenetic analyses

[Download file](#) (16.42 kb)

Suppl. material 3: Mitogenomic organisation of *Matsucoccus matsumurae* [doi](#)

Authors: Kai Hu and Jiansheng Qiu

Data type: Table

[Download file](#) (22.46 kb)

Suppl. material 4: BI phylogenetic tree for Sternorrhyncha 

Authors: Kai Hu and Jiansheng Qiu

Data type: Figure

Brief description: BI phylogenetic tree for Sternorrhyncha, based on the amino acid sequence data of 13 PCGs from *Matsucoccus matsumurae* and other 33 species belonging to five superfamilies of Hemiptera. Bayesian posterior probabilities (PP) are indicated on the branches

[Download file](#) (779.95 kb)