International Journal of Biological Macromolecules 121 (2019) 152–160

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: http://www.elsevier.com/locate/ijbiomac

Characterization of the complete mitochondrial genome of *Simulium* (*Byssodon*) *maculatum* (Diptera: Simuliidae) and its phylogenetic implications

Gang Wang, Min Huang *

Key Laboratory of Plant Protection Resources and Pest Management of the Ministry of Education, Entomological Museum, Northwest A&F University, Yangling, Shaanxi 712100, China

ARTICLE INFO

Article history:
Received 4 September 2018
Received in revised form 27 September 2018
Accepted 27 September 2018
Available online 1 October 2018

Keywords:
Simulium (*Byssodon*) *maculatum*
Mitochondrial genome
Phylogenetic analysis

ABSTRACT

The mitochondrial (mt) genome of the black fly *Simulium* (*Byssodon*) *maculatum*, a pest of great importance for both humans and livestock, is sequenced and annotated for the first time. The genome structure, gene order and codon usage are typical among Diptera mt genomes. The mt genome is circular and 15,799 bp in length with 13 protein coding genes (PCGs), 22 transfer RNA genes (tRNAs), two ribosomal RNA genes (rRNAs) and a control region (CR), and with weakly positive AT-skew (0.02) and negative GC-skew (−0.12). Phylogenetic relationships of 16 species representing five families of Culicomorpha and two outgroups, based on mt genome data, were analyzed using both Maximum Likelihood and Bayesian methods. The monophyly of Culicomorpha is well supported, while Chironomoidea is indicated as a paraphyletic group. The well supported monophyletic Simuliidae is the sister group to Culicidae.

1. Introduction

Black flies are significant vectors of disease agents and pests of humans and animals [1,2], ranking the second most important veterinary and medical insect group [3]. >25 species of black flies worldwide are transmitters of the causal agent of human onchocerciasis which is the second most important cause of infectious blindness [4]. *Simulium* (*Byssodon*) *maculatum* (Meign, 1804), a notorious species of black flies (Diptera: Simuliidae), is widely distributed in Asia, Europe and North America [5]. It is one of the most important pests of humans and livestock in floodplains [6], and has even been described as a 'nefarious bloodsucker' in the forest zone of the Palearctic Region [1]. As a widespread species around the Palearctic realm, *S. maculatum* exhibits morphological diversity in different geographic populations and has been described multiple times under different scientific names, resulting in many synonyms (e.g., *Atrctocera pungens* Meigen 1806, *Simulia subfasciatum* Meigen, 1838, *Prosimulium vigintiquaterni* Enderlein 1929, *Echinosimulium echinatum* Baranov 1938, *Simulium* (*Nevermannia*) *maculatum ussurianum* Rubtsov 1940, *Titanopteryx maculatum uralensis* Rubtsov 1956, etc.) as recognized by Crosskey (1988). This suggests that further study is needed to understand cryptic diversity and genetic differentiation in the species [7,8]. The small size and structural homogeneity in morphology of simuliids have hindered attempts to infer phylogenetic relationships of this group based on morphological characteristics [9]. To resolve such problems, partial mitochondrial gene sequences (e.g., COI, COII and Cytb) and nuclear genes (e.g., ECP1, BZF, ITS1 and ITS2) have been extensively utilized for the identification of species complexes and groups and phylogenetic inference of this group [9–15]. However, these genes are usually too

Table 1
Classification and mitochondrial genome information mentioned in this research.

Family	Species	Length(bp)	GenBank accession number
Culicidae	*Aedes albopictus*	17,150	KX383916
	Anopheles gambiae	15,363	NC_002084
	Culex quinquefasciatus	15,579	GU188856
	Sabethes belisarioi	15,911	MF957171
Dixidae	*Dixella aestivalis*	16,465	KT878382
	Dixella sp.	15,574	KM245574
Chironomidae	*Polypedilum vanderplanki*	16,445	NC_028015
	Chironomus tepper	15,652	NC_016167
Simuliidae	*Simulium variegayum*	15,367	NC_033348
	Simulium aureothirtum	15,904	NC_029753
	Simulium maculatum	15,799	MH729190
Ceratopogonidae	*Culicoides arakawae*	18,135	AB361004
	*Culicoides actoni**	9476	AB360971
	*Culicoides cylindratus**	10,428	AB361007
Pachyneuridae	*Cramptonomyia spenceri*	16,274	NC_016203
Keroplatidae	*Arachnocampa flava*	16,923	NC_016204

Note: * the mitochondrial genome of the indicated species is incomplete.

* Corresponding author at: Key Laboratory of Plant Protection Resources and Pest Management of the Ministry of Education, Entomological Museum, Northwest A&F University, Taicheng Road, Yangling, Shaanxi 712100, China.
E-mail address: huangmin@nwsuaf.edu.cn (M. Huang).

https://doi.org/10.1016/j.ijbiomac.2018.09.205
0141-8130/

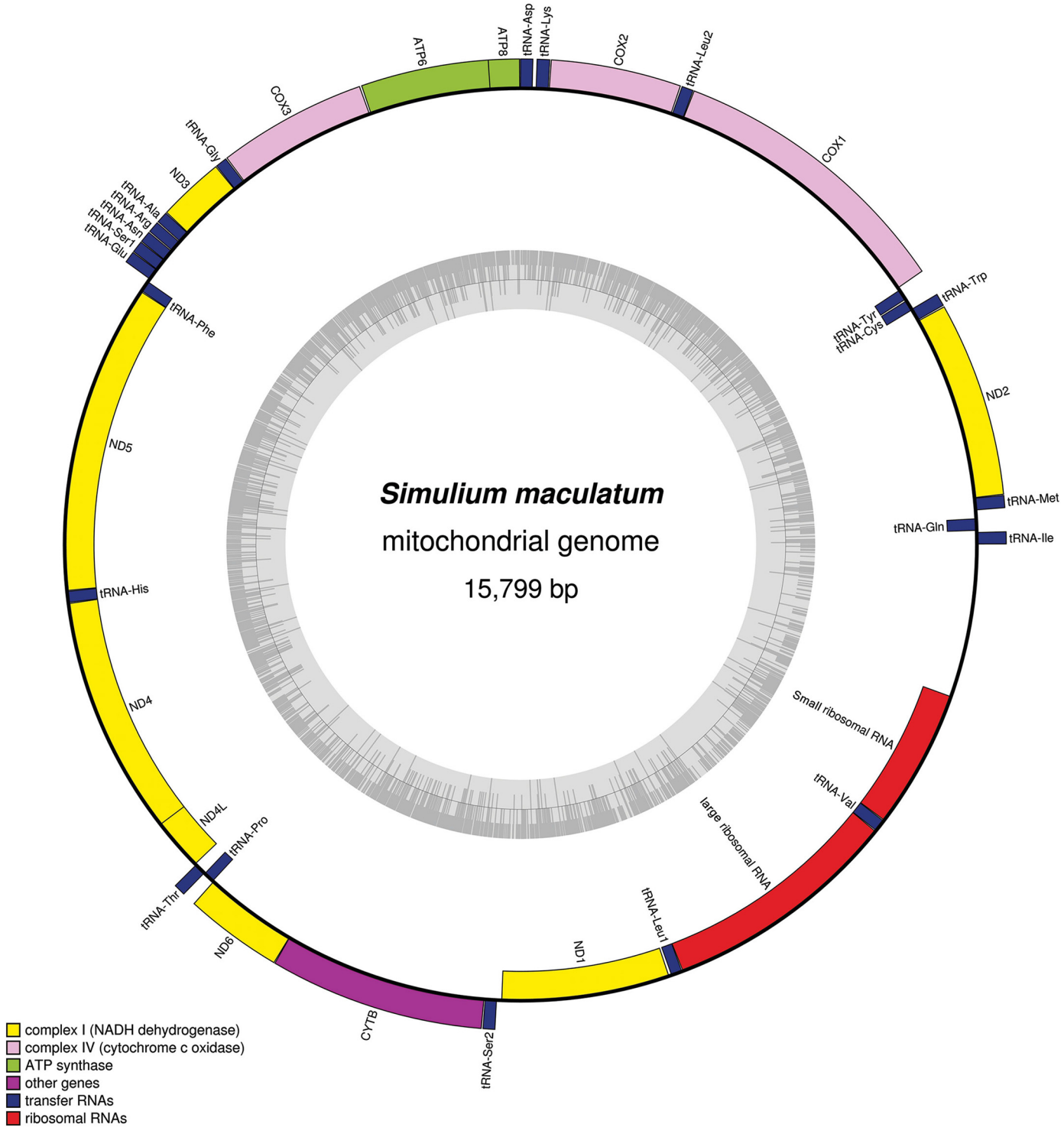

Fig. 1. The mitochondrial genome arrangements of *Simulium maculatum*. Protein-coding and ribosomal genes are shown with standard abbreviations. The J-strand is visualized on the outer circle and the N-strand on the inner circle.

short, with insufficient phylogenetic information to recover higher-level phylogenetic relationships [9]. Complete mt genomes have been shown to be suitable to address such problems, and multi-gene analysis has been widely used for phylogenetic analysis of invertebrates including insects [16]. Typically, the mt genome of insects is a small (about 14–19 kb), closed, circular, and maternally inherited molecule [17]. It comprises 13 protein coding genes (PCGs), two ribosomal RNA genes (the large and small ribosomal subunits, rRNAs) genes, 22 transfer RNA genes (tRNAs) and a non-coding element termed the control region (CR), which contains initiation sites for transcription and replication [18,19]. Moreover, the mt genome is considered a powerful marker which can provide more genome-level information than the shorter sequences of individual genes [20,21] due to its small size, widely abundance (in animal tissues), low level of recombination, faster evolutionary rate, and evolutionary conserved gene products [22–24]. Thus, over the past few decades, many mt genomes have been sequenced and analyzed [25–27], with increasing use in taxonomy [28], population genetics and structure [29–31], phylogenetic inference [32–34], as well

as comparative genome structure and rearrangement [35,36]. But only two mt genomes of black flies have been sequenced so far, without systematic analysis [37,38], and no molecular data of *S. maculatum* have yet been deposited in Genbank hindering further research and management of this and related pests [12].

In this study, we present the complete sequence and functional annotation of *S. maculatum* mitochondrial genome for the first time, and describe its molecular phylogenetic relationship with 14 other species, providing an estimate of the relationships within Culicomorpha and providing a base for further molecular research on related taxa.

2. Materials and methods

2.1. Sample collection and DNA extraction

Larvae and pupae of *S. maculatum* were collected by hand from substrates in the Irtysh River (Xinjiang, China) and were immediately fixed in 95% ethanol. Pupae with pharate adults were selected and isolated individually in 5 ml plastic vials with damp filter paper until adult emergence. Adults were fixed in 95% ethanol after emergence.

All samples were stored in 95% ethanol at −80 °C in the laboratory at the Entomological Museum of Northwest A&F University, Yangling, Shaanxi, China until DNA extraction. The specimens including larvae, pupae, female and male adults were identified morphologically [2,7]. Total genomic DNA was extracted from the thorax of five specimens using DNeasy Blood and Tissue Kit (QIAGEN) following the instruction. The Nanodrop 2000 spectrophotometer was used to quantify the extracted DNA (*S. maculatum*: 8.1 μg) for library construction.

2.2. Sequence analysis

The complete mt genome of *S. maculatum* was sequenced on an Illumina HiSeq™2500 system with paired reads of 2 × 150 bp by the Biomarker Technologies Corporation (Beijing, China). A total of 14,889,248 raw paired reads were retrieved and quality-trimmed with Trimmomatic v0.35 [39]. The resultant 88,458 clean paired reads were used for mt genome reconstruction following the methods of Hahn et al. [40], which employed mt genome of *Simulium* (*Nevermannia*) *aureohirtum* (KP793690) [37] as a bait sequence. The mt genome of *S. maculatum* was annotated with Geneious 8.1.3 (Biomatters, Auckland, New Zealand), using the mt genome of (*Nevermannia*) *aureohirtum* (KP793690) [37] and *Simulium* (*Simulium*) *variegatum* (KU252587) [38] as references. All PCGs were predicted by the ORF Finder in NCBI (https://www.ncbi.nlm.nih.gov/orffinder) and modified by comparison with the previously published mt genomes of Simuliidae. Both rRNAs and tRNAs were identified using the MITOS Web Server [41]; tRNAs genes were also predicted with tRNAscan-SE 1.3.1 [42]. The physical circular map of the mt genome was drawn using Organellar Genome DRAW (OGDRAW) [43]. The nucleotide composition and relative synonymous codon usage (RSCU) values of each PCGs were computed with MEGA 7.0 [44]. Strand asymmetry was calculated using the following formulas: AT skew = [A − T]/[A + T] and GC skew = [G − C]/[G + C] [45]. The mt genome sequence was deposited in GenBank with the accession number MH729190.

2.3. Phylogenetic analysis

A total of 16 species of Nematocera insects were included in our phylogenetic analysis, including 14 Culicpmorpha species and two outgroup species of Keroplatidae and Pachyneuridae, respectively (Table 1). Because there were two incomplete mt genomes obtained, the genes used in the phylogenetic analyses only included seven PCGs (COI, COII, COIII, ND2, ND3 ATP6 and ATP8), 12 tRNAs (Leu(UUA), Ile, Gln, Tyr, Trp, Cys, Asp, Lys, Met, Gly, Ala and Arg) and two rRNAs (lrRNA and srRNA). The third codon positions were removed for phylogeny inference, because they may suffer from saturated sites and rate heterogeneity which can introduce noise to the phylogenetic analysis [46,47]. Two datasets were assembled for phylogenetic analyses: the first dataset consisted of the first and second codon positions of the seven PCGs, two rRNAs and 12 tRNAs; and the second dataset consisted of the first and second codon positions of seven PCGs and two rRNAs.

The L-INS-i strategy and Q-INS-i strategy as implemented in MAFFT version 7 alignment server (http://mafft.cbrc.jp/alignment/server/) [48] were performed with default settings to align PCGs and rRNAs, respectively. Poorly aligned sites of PCGs were removed from the protein alignment using GBlocks v0.91b (http://molevol.cmima.csic.es/castresana/Gblocks_server.html) [49] with default parameters. While, ambiguous positions of rRNAs in the alignment were filtered by hand based on the secondary structures predicted. MXSCARNA was utilized to align tRNA genes based on the predicted secondary structures [50].

ML analysis was conducted using IQ-TREE v1.4.1 [51] with the best-fit substitution model automatically selected by ModelFinder [52] in the IQ-TREE package. The node reliability/branch support was assessed by performing bootstrap with 1000 replicates. Prior to BI analysis employed on MrBayes 3.2.6 [53], the GTR + I + G was selected as the best model for two nucleotide datasets under the Akaike Information Criterion (AIC) by ModelTest 3.7 [54]. Bayesian analysis was performed with default settings and four independent runs 2 million generations and sampled every 100 generations (average standard deviation of split frequencies <0.01, estimated sample size (ESS) > 200, and potential scale reduction factor ≈1).

Table 2
Organization of the *Simulium maculatum* mitochondrial genome.

Name	Direction	Location	Length	IGNc	Anticodon	Start	Stop
tRNA-Ile	forward	1–68	68	–	GAT		
tRNA-Gln	reverse	76–144	69	7	TTG		
tRNA-Met	forward	189–257	69	44	CAT		
ND2 CDS	forward	258–1289	1032	0		ATT	TAA
tRNA-Trp	forward	1293–1363	71	3	TCA		
tRNA-Cys	reverse	1356–1419	64	−8	GCA		
tRNA-Tyr	reverse	1425–1489	65	5	GTA		
COX1 CDS	forward	1501–3034	1534	11		TTG	T--
tRNA-Leu	forward	3035–3100	66	0	TAA		
COX2 CDS	forward	3105–3794	690	4		ATG	TAA
tRNA-Lys	forward	3797–3867	71	2	CTT		
tRNA-Asp	forward	3885–3953	69	17	GTC		
ATP8 CDS	forward	3954–4121	168	0		ATC	TAA
ATP6 CDS	forward	4115–4792	678	−7		ATG	TAA
COX3 CDS	forward	4799–5587	789	6		ATG	TAA
tRNA-Gly	forward	5590–5654	65	2	TCC		
ND3 CDS	forward	5655–6008	354	0		ATT	TAA
tRNA-Ala	forward	6009–6073	65	0	TGC		
tRNA-Arg	forward	6076–6139	64	2	TCG		
tRNA-Asn	forward	6142–6206	65	2	GTT		
tRNA-Ser	forward	6207–6273	67	0	GCT		
tRNA-Glu	forward	6274–6339	66	0	TTC		
tRNA-Phe	reverse	6354–6419	66	14	GAA		
ND5 CDS	reverse	6420–8156	1737	0		ATC	TAA
tRNA-His	reverse	8157–8222	66	0	GTG		
ND4 CDS	reverse	8223–9563	1341	0		ATG	TAA
ND4L CDS	reverse	9557–9853	297	−7		ATG	TAA
tRNA-Thr	forward	9856–9922	67	2	TGT		
tRNA-Pro	reverse	9923–9987	65	0	TGG		
ND6 CDS	forward	9990–10,511	522	2		ATT	TAA
CYTB CDS	forward	10,511–11,647	1137	−1		ATG	TAA
tRNA-Ser	forward	11,651–11,717	67	3	TGA		
ND1 CDS	reverse	11,742–12,683	942	24		ATA	TAG
tRNA-Leu	reverse	12,694–12,759	66	10	TAG		
16S rRNA	reverse	12,760–14,092	1333	0			
tRNA-Val	reverse	14,093–14,164	72	0	TAC		
12S rRNA	reverse	14,165–14,947	783	0			
CR	forward	14,948–15,799	852	0			

Note: Sizes are given in bp; IGNc are intergenic nucleotides, where negative numbers indicate overlaps. Start and Stop are codons.CR means control region.

Table 3
Nucleotide composition of the *Simulium maculatum* mitochondrial genome.

Feature	A%	T%	C%	G%	A + T%	C + G%	AT-Skew	GC-Skew
Whole mitgenome	38.3	36.6	14.7	10.4	74.9	25.1	0.02	−0.12
PCGs	36.7	35.7	16.0	11.6	72.4	27.6	0.01	−0.16
1st codon position	38.2	34.1	13.5	14.2	72.3	27.7	0.06	0.03
2st codon position	30.8	34.0	20.0	15.2	64.8	35.2	−0.05	−0.14
3st codon position	41.2	38.8	14.4	5.5	80.0	19.9	0.03	−0.45
tRNA gene	38.6	38.1	12.7	10.5	76.7	23.2	0.01	−0.10
lrRNA	43.9	37.3	12.2	6.7	81.2	18.9	0.08	−0.29
srRNA	40.3	36.7	14.6	8.4	77.0	23.0	0.05	−0.27
Control region	47.4	42.8	6.2	3.5	90.2	9.7	0.05	−0.28

Note: AT-skew = [A − T]/[A + T], GC-skew = [G − C]/[G + C].

3. Results and discussion

3.1. General features of mitochondrial genome organization

The complete mt genome of *S. macultum* is medium-sized (15,799 bp) in comparison with previously reported mt genomes of nematocerous Diptera, which ranged from 15,214 bp in *Ptychoptera* to about 18,600 bp in *Bittacomorphella* [35]. Both gene content and pattern were identical to that of the inferred ancestral insect mt genome, also the same as all previously published simuliid mt genomes, with 23 genes encoded on the majority strand (J-strand) and the remaining 14 genes located on the minority strand (N-strand) (Fig. 1, Table 2).

Within each sequenced *Simuliidae* species, there were three conserved regions in the mt genome, which were also shared with some other dipterans [55,56], in overlapping regions of genes: AAGCCTTA (tRNA Trp-tRNA Cys), ATGATAA (ATP8-ATP6) and TTAACAT (ND4-ND4L). Moreover, we also found two non-coding intergenic regions conserved within dipteran insects, which have been proved the binding sites for a bidirectional transcription termination factor (DmTTF) [35]. One of which is 14 bp in length, located between tRNA-Glu and tRNA-Phe; and the other one ranges from 18 to 24 bp in length, between tRNA-Ser (UCN) and ND1 (Table S1). For Diptera, we found the second non-coding intergenic region in partial Nematocera (except some families such as Culicidae) and in all Brachycera [35,57].

The control region (CR), also called the A + T-rich region or the displacement-loop region (D-loop), is the longest intergenic region in the mt genome. The CR is considered to be involved in regulation of transcription and replication of insect mt genomes [58].

In the Nematocera, the length of CR ranged from 369 bp in *Ptychoptera* to about 3.7 kb in *Bittacomorphella* [35]. The CR of *S. maculatum* is a mid-sized length (spans 848 bp), located between tRNA-Ile and srRNA genes [36], with 90.2% A + T content, and has weakly positive AT-skew (0.05). Large tandem repeats, which are common in the CR of Nematocera, were not detected in *S. maculatum* [36,55]. Some conserved features identified in the CR of other insects were also absent in *S. maculatum*, e.g., (TA)n-like stretch, poly-T stretch, etc. However, we found 'ATTTA' element nearly in the center of the CR, and next to it T/A repeat located.

3.2. Base composition and codon usage

The base composition bias of the *S. maculatum* mt genome was 74.9% A and T (A: 38.3%, C: 14.7, T: 36.6% G: 10.4%), ranking it approximately near the average in nematocerans (73% in Trichoceridae to 83% in Cecidomyiidae) [35]. For PCGs, the AT content of third codon positions (80.0%) was higher than that of the first (72.3%) and second (64.8%) codon positions. In tRNAs, the AT content of the lrRNA genes (81.2%) was higher than that of the srRNA genes (77.0%). The *S. maculatum* mt genome was found with weakly positive AT-skew (0.02) and negative GC-skew (−0.12). Statistics also indicate that the AT-bias was stronger in RNA-encoding genes than in PCGs (Table 3).

The relative synonymous codon usage (RSCU) in 13 PCGs of the *S. maculatum* mt genome is summarized in Fig. 2. A + T bias was also reflected in the relative codon usage by the PCGs. The AT rich codons ATT (Ile), TTA (Leu), AAT (Asn), TTT (Phe), AAA (Lys) and ATA (Met) were the six most frequently used codons. While, the GC rich codons GAG (Glu) and GCG (Ala) were the least used codons.

Of the 13 PCGs (Table 2), 12 PCGs utilized ATN (ATG, ATT, ATC and ATA) as start condons, whereas CO1 employed TTG as initiation codon. ATG (Met) was the largest number start condon used in six PCGS (ATP6, COI, COII, CYTB, ND4 and ND4L). ATT (Met) followed in second place with 3 PCGs (ND2, ND3 and ND6). Then, ATC was used in 2 PCGs (ATP8 and ND5). However, ATA was only applied in ND1.

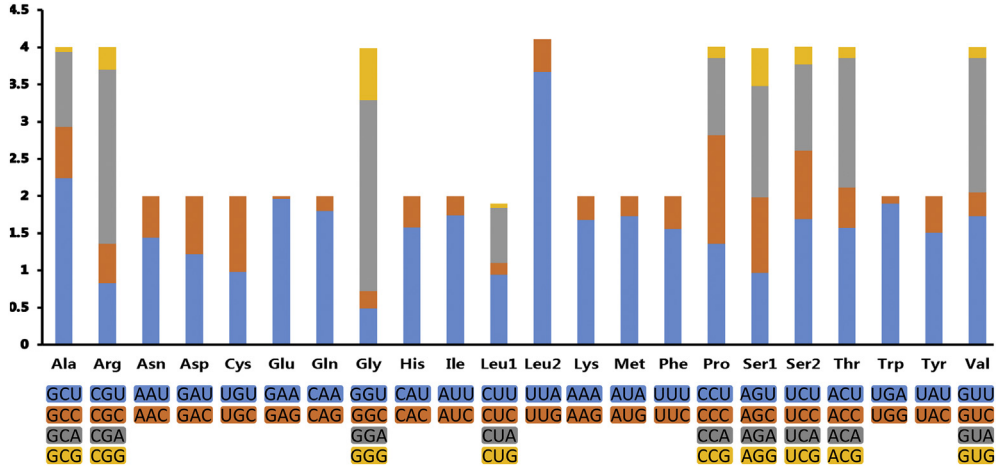

Fig. 2. Relative synonymous codon usage (RSCU) for the mitochondrial genome of *Simulium maculatum*. Codon families were provided on the X-axis.

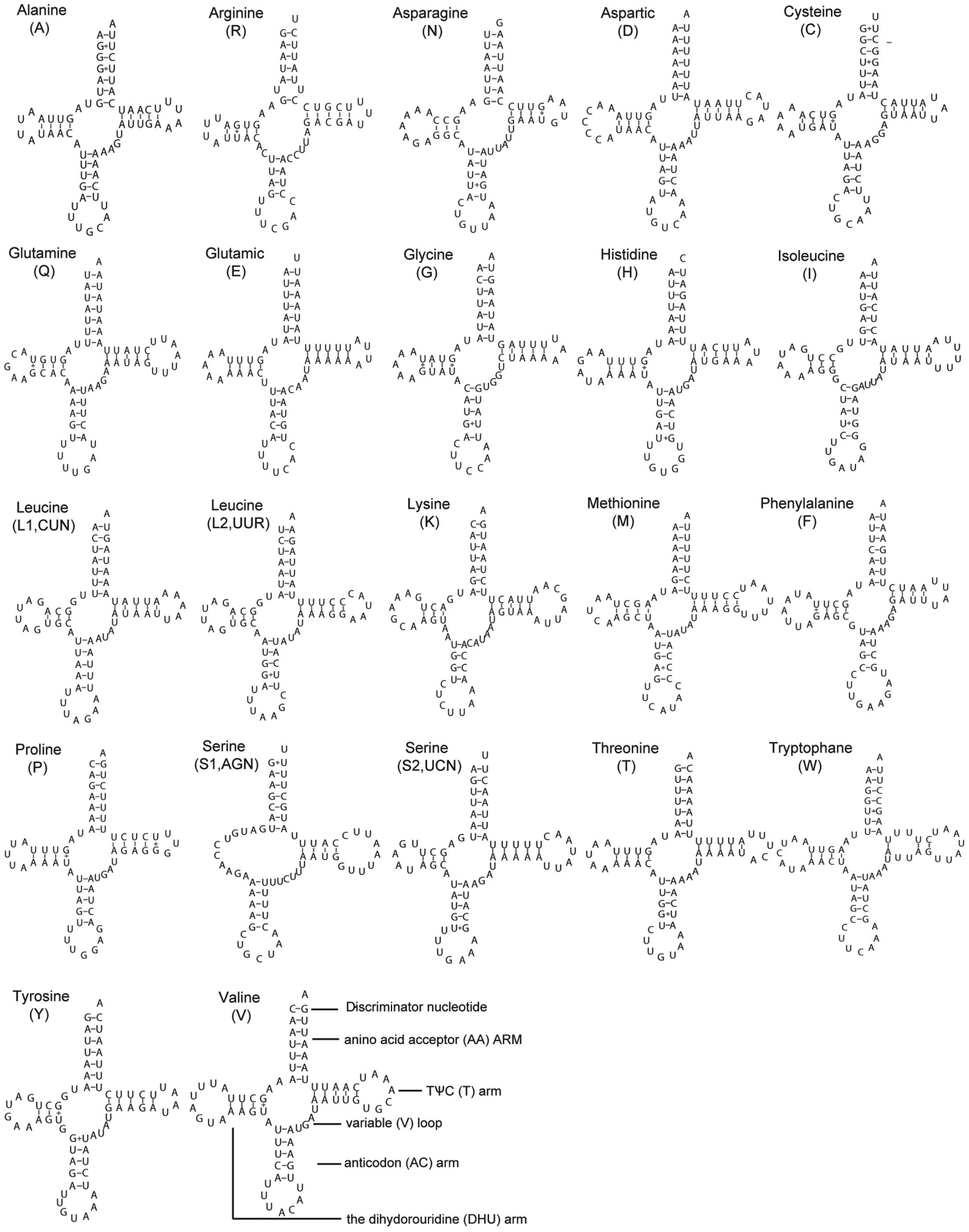

Fig. 3. Inferred abnormal secondary structures of the tRNAs. Dashes indicate Watson-Crick base pairing; additional sign (+) indicates unmatched base pairing.

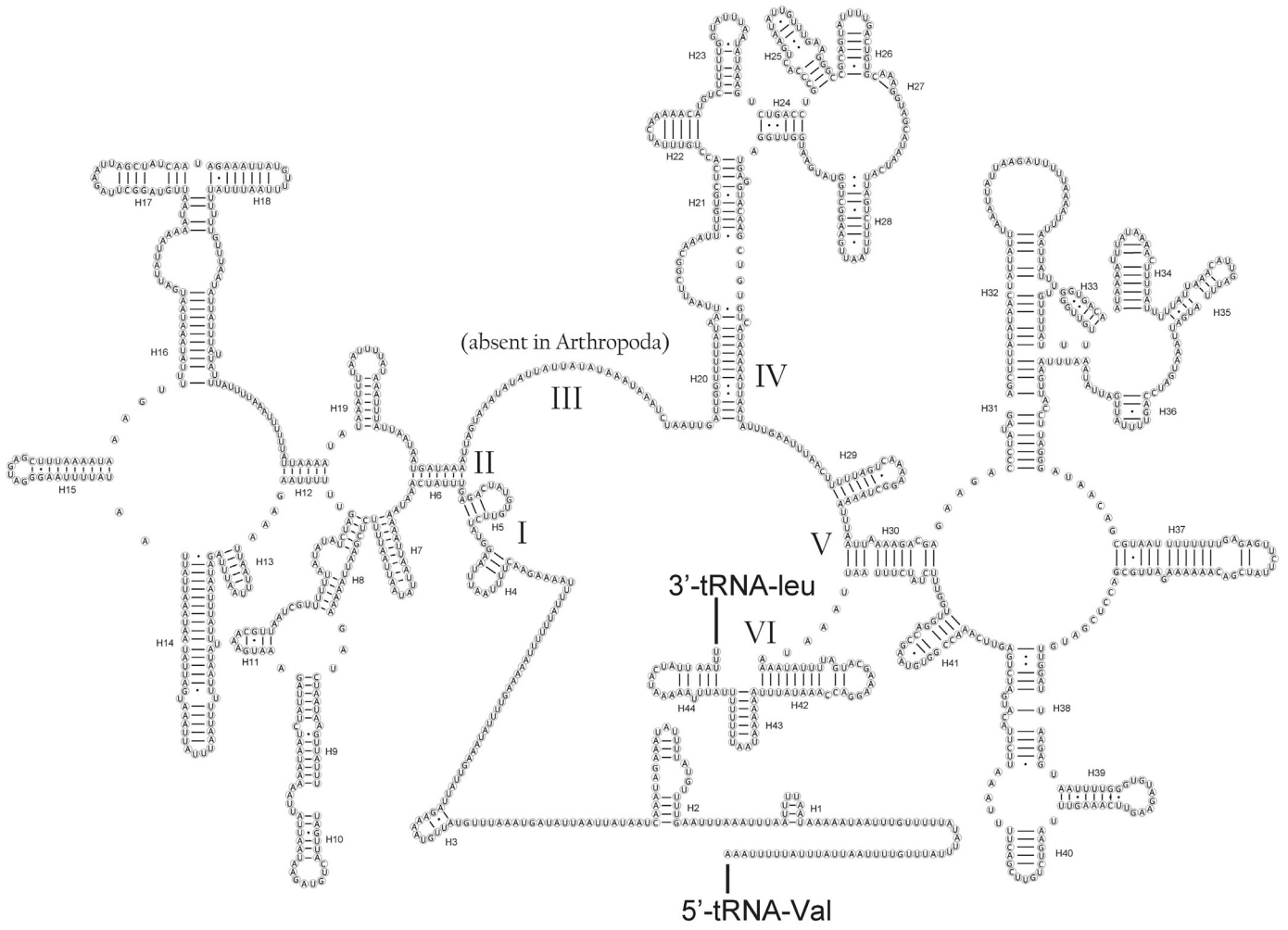

Fig. 4. Predicted secondary structure of the lrRNA gene in *Simulium maculatum*. Inferred Watson–Crick bonds are illustrated by lines, GU bonds by dots.

Similarly, these patterns are also presented in other nematoceran flies [55,56].

As in most other Nematocera, TAA was the most commonly used stop codon in *S. maculatum*, found in all the PCGs except ND1 and COI. The stop codon TAG was used in the ND1 like most other nematoceran flies. The COI gene in *S. maculatum* used the partial stop codon T.

3.3. Transfer RNAs

Twenty-two transfer RNA genes (tRNAs) whose total length was 1473 bp with A + T content of 76.7%, were detected in the mt genome of *S. maculatum*. The length of mt tRNAs ranged from 64 bp (Arg and Cys) to 72 bp (Val). The mt tRNAs had weakly positive AT-skew (0.01) and negative GC-skew (−0.10). As in most previously studied insect mt genomes, most tRNAs could be folded into a typical clover-leaf secondary structure (Fig. 3), but tRNA-Ser (AGN) was an exception in lacking a DHU arm [16]. Two types of mispairings (G-U and A-C) were found in *S. maculatum* mt tRNA secondary structures, including 24 G-U pairs and 1 mismatched base A-C pairs.

3.4. Ribosomal RNAs

It was not easy to determine the boundaries of rRNAs, given the lack of functional annotation features (e.g. start and stop codons). Thus, we identified the boundaries of rRNA genes based on their relative location in the mt genome, following the method of Wang et al. [59]. Among the three sequenced Simuliidae mt genomes, the length of lrRNAs were 1333 bp (*S. maculatum*, present study), 1365 bp (*S. variegatum*), and 1366 bp (*S. aureohirtum*), respectively, and the lengths of srRNAs are 787 bp (*S. maculatum*), 792 bp (*S. aureohirtum*) and 793 bp (*S. variegatum*), respectively [46,47]. Both subunits of rRNAs were encoded on the N-strand as in other insects. Secondary structures of both subunits of rRNAs of *S. maculatum* mt genome were inferred using published rRNAs secondary structures of Tipuloidea (Diptera: Tipulomorpha) [56] and Lauxanioidea (Diptera: Cyclorrhapha) [36] in Fig. 4 and Fig. 5.

The lrRNA had 44 helices in five structural domains (I–II, IV–VI and domain III were absent as in other insects). The multiple alignments of lauxanioid lrRNAs spanned 1333 positions. The srRNA included three domains and 30 helices.

3.5. Phylogenetic analysis

Both ML and Bayesian analyses of two data sets were employed to reconstruct the phylogenetic relationships among 16 species (including two outgroup species) based on the amino acid and nucleotide sequences of the PCGs (with the third codon position removed), tRNAs and rRNAs. Four phylogenetic trees (namely PCG12R-ML, PCG12RT-ML, PCG12R-BI and PCG12RT-BI) reconstructed from both Bayesian and ML analyses had similar topologies across the two datasets (Fig. S1); only PCG12RT-ML had some discordant nodes which will be discussed below. The monophyly of Culicomorpha is strongly

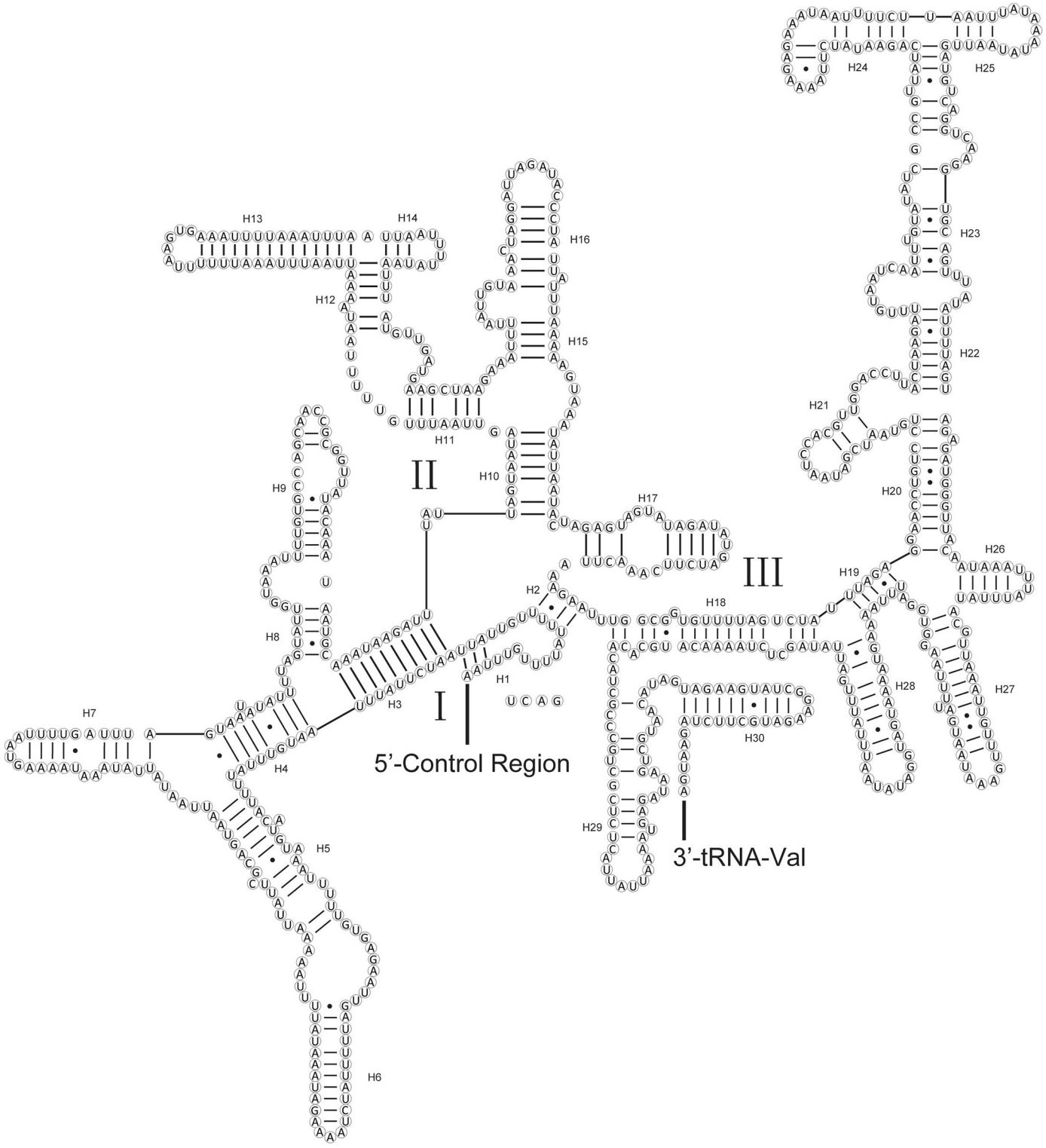

Fig. 5. Predicted secondary structure of the srRNA gene in *Simulium maculatum*. Inferred Watson-Crick bonds are illustrated by lines, GU bonds by dots.

confirmed by BI and ML analyses across both datasets (posterior probability (PP) = 1, ML bootstrap (BS) = 99) (Fig. 6).

Within Culicomorpha, most relationships are robust and congruent across analyses based on the two datasets. Also, the superfamily Culicoidea is monophyletic which is strongly supported in the BI tree but weakly supported in the ML tree (PP = 0.97, BS = 51). Simuliidae formed the sister group to Dixidae + Culicidae (PP = 1, BS = 82). All families represented by two or more taxa are strongly supported as monophyletic including Chironomidae (two taxa), Ceratopogonidae (three taxa), Dixidae (two taxa), Simuliidae (three taxa) and Culicidae (four taxa). Within Culicidae, the monophyletic subfamily Anophelinae is sister to Culicinae (PP = 1, BS = 100). This is consistent with the Bayesian analysis of Reidenbach et al. based on concatenated gene and gene + morphology data sets from 26 mosquito species representing 25 genera of Culicidae [60]. The phylogenetic relationships within the subfamily Culicinae is estimated as (Sabethini (*Sabethes belisarioi*)) +

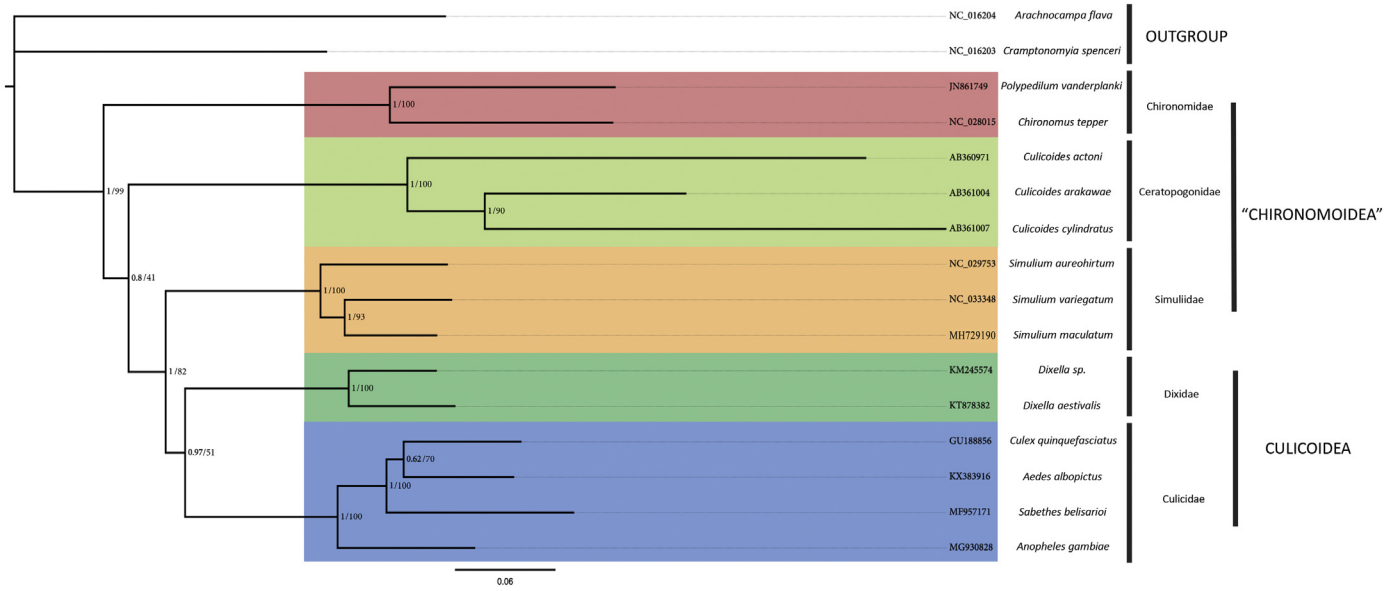

Fig. 6. Phylogenetic relationship of Culicomorpha inferred from two concatenated datasets (PCG12 + rRNA and PCG12 + tRNA + rRNA) and two softwares (IQ-tree and MrBayes). The number on the branches indicates posterior probabilities (BI) and bootstrap (ML).

(Aedini (*Anopheles gambiae*) + Culicini (*Culex quinquefasciatus*)), with all nodes having full support (PP = 1, BS = 100), except Aedini + Culicini (PP = 0.62, BS = 70). Within Simuliidae, the subgenus *Simulium* has a closer relationship to the subgenus *Byssodon* than *Nevermannia*, and the monophyletic subgenus *Nevermannia* is sister to *Simulium* + *Byssodon* (PP = 1, BS = 100). This differs from the study of Adler et al. based on morphological, cytological and molecular data, which suggests that the subgenus *Nevermannia* is more closely related to *Byssodon* [2].

Our results suggest that the superfamily Chironomoidea is a paraphyletic group and the Chironomidae is the earliest branch of the Culicomorpha, which is consistent with the previous study of Kutty et al. and Borkent et al. [61,62]. It is incompatible with the results of Kutty et al. based on transcriptome data of all families in Culicomorpha, in which Chironomidae and Ceratopogonidae were not recovered as sister groups in all phylogenetic trees based two datasets for BI and PR for ML [61]. Although PRT for ML has similar topologies with the result of Kutty et al. [61], it still could not fully recover the relationship between Chironomidae and Ceratopogonidae (BS = 47). Moreover, PRT for ML weakly supports Culicidae formed the sister group to (Simuliidae + Dixidae) (BS = 67) and the sister relationship of Simuliidae with Dixidae (BS = 37). As Kutty et al. [61] and Bazinet et al. [63] noted that those poorly supported unstable relationship at the base suggest the need for additional taxonomic sampling, which remains critical for addressing intricate phylogenetic relationships of related groups. Considerable additional data and taxon sampling of Culicomorpha are required to reliably resolve relationships among the eight constituent families. Nevertheless, our results suggest that complete mt genome data are informative of phylogenetic relationships among major lineages of Culicomorpha.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2018.09.205.

Acknowledgments

We are grateful to Professor Chris Dietrich (Illinois Natural History Survey, USA) for reviewing the manuscript and providing valuable comments. This study was supported by Ministry of Science and Technology of China (2013GS650305) and Science and Technology Bureau of Xinjiang Production and Construction Corps (2013BB017, 2012BAH12B03).

References

[1] R.W. Crosskey, The Natural History of Blackflies, John Wiley & Sons, London, 1990.
[2] P.H. Adler, D.C. Currie, D.M. Wood, The Black Flies (Simuliidae) of North America, Cornell University Press, Ithaca, New York, USA, 2004 (DM).
[3] D.C. Currie, P.H. Adler, Global diversity of black flies (Diptera: Simuliidae) in freshwater, Hydrobiologia 595 (2008) 469–475.
[4] P.H. Adler, R.A. Cheke, R.J. Post, Evolution, epidemiology, and population genetics of black flies (Diptera: Simuliidae), Infect. Genet. Evol. 10 (7) (2010) 846–865.
[5] P.H. Adler, R.W. Crosskey, World blackflies (Diptera: Simuliidae): a comprehensive revision of the taxonomic and geographical inventory, http://www.clemson.edu/cafls/departments/esps/biomia/pdfs/blackflyinventory.pdf 2018, Accessed date: 30 May 2018.
[6] A.V. Yankovsky, The ecology and systematics of the striped black flies of the group *Byssodon maculata* Meigen (Diptera, Simuliidae), Entomologicheskoe Obozrenie 57 (1978) 169–179 (In Russian).
[7] H.B. Chen, Chinese Blackflies (Diptera: Simuliidae), Guizhou Science and Technology Publishing House, Guiyang, 2016 (In Chinese).
[8] R.W. Crosskey, An annotated checklist of word blackflies (Diptera: Simuliidae), J. Nat. Hist. 22 (1988) 321–355.
[9] J.K. Moulton, Molecular sequence data resolves basal divergences within Simuliidae (Diptera), Syst. Entomol. 25 (2000) 95–113.
[10] B. Larue, C. Gaudreau, H.O. Bagre, G. Charpentier, Generalized structure and evolution of ITS1 and ITS2 rDNA in black flies (Diptera: Simuliidae), Mol. Phylogenet. Evol. 53 (2009) 749–757.
[11] P. Sriphirom, P.N. Sopaladawan, K. Wongpakam, P. Pramual, Molecular phylogeny of black flies in the *Simulium tuberosum* (Diptera: Simuliidae) species group in Thailand, Genome 57 (2014) 45–55.
[12] S. Anbalagan, M. Kannan, S. Dinakaran, M. Krishnan, Species diversity of black flies (Diptera: Simuliidae) in oriental region and molecular phylogeny of the subgenus *Gomphostilbia* members, J. Vector Dis. 54 (2017) 80–86.
[13] R.G. Harrison, Animal mitochondrial DNA as a genetic marker in population and evolutionary biology, Trends Ecol. Evol. 4 (1989) 6–11.
[14] H. Takaoka, W. Srisuka, V.L. Low, A. Saeung, A new species of the *Simulium* (*Simulium*) *striatum* species-group (Diptera: Simuliidae) from Thailand, and its differentiation from two related species based on a fast-evolving nuclear gene, J. Med. Entomol. 55 (2018) 561–568.
[15] V.L. Low, H. Takaoka, P. Pramual, P.H. Adler, Z. Ya'cob, Y.T. Huang, X.D. Pham, R. Ramli, C.D. Chen, A. Wannaket, M. Sofian-Azirun, Delineating taxonomic boundaries in the largest species complex of black flies (Simuliidae) in the Oriental Region, Sci. Rep. 6 (2016), 20346.
[16] G.L. Senatore, E.A. Alexander, P.H. Adler, J.K. Moulton, Molecular systematics of the *Simulium jenningsi* species group (Diptera: Simuliidae), with three new fast-evolving nuclear genes for phylogenetic inference, Mol. Phylogenet. Evol. 75 (2014) 138–148.
[17] J.L. Boore, Animal mitochondrial genomes, Nucleic Acids Res. 27 (1999) 1767–1780.
[18] C. Gissi, F. Iannelli, G. Pesole, Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species, Heredity 101 (2008) 301–320.
[19] D.R. Wolstenholme, Animal mitochondrial DNA: structure and evolution, Int. Rev. Cytol. 141 (1992) 173–216.

[20] M. Dowton, L.R. Castro, A.D. Austin, Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome 'morphology', Invertebr. Syst. 16 (2002) 345–356.
[21] S.E. Masta, J.L. Boore, Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes, Mol. Biol. Evol. 25 (2008) 949–959.
[22] C.P. Lin, B.N. Danforth, How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets, Mol. Phylogenet. Evol. 30 (2004) 686–702.
[23] C. Simon, T.R. Buckley, F. Frati, J.B. Stewart, A.T. Beckenbach, Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA, Annu. Rev. Ecol. Evol. Syst. 37 (2006) 545–579.
[24] C. Gissi, F. Iannelli, G. Pesole, Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species, Heredity 101 (4) (2008) 301–320.
[25] A. Gibson, V. Gowri-Shankar, P.G. Higgs, M. Rattray, A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods, Mol. Biol. Evol. 22 (2005) 251–264.
[26] S.L. Cameron, How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research, Syst. Entomol. 39 (2014) 400–411.
[27] S.L. Cameron, Insect Mitochondrial Genomics, Implications for evolution and phylogeny, in: M.R. Berenbaum (Ed.), Annu. Rev. Entomol., Vol. 59, 2014, pp. 95–117.
[28] J. Krzywinski, C. Li, M. Morris, J.E. Conn, J.B. Lima, M.M. Povoa, R.C. Wilkerson, Analysis of the evolutionary forces shaping mitochondrial genomes of a Neotropical malaria vector complex, Mol. Phylogenet. Evol. 58 (2011) 469–477.
[29] L. Mousson, C. Dauga, T. Garrigues, F. Schaffner, M. Vazeille, A.B. Failloux, Phylogeography of *Aedes* (*Stegomyia*) *aegypti* (L.) and *Aedes* (*Stegomyia*) *albopictus* (Skuse) (Diptera: Culicidae) based on mitochondrial DNA variations, Genet. Res. 86 (2005) 1–11.
[30] S.J. Scheffer, M.L. Lewis, Mitochondrial phylogeography of the vegetable pest *Liriomyza trifolii* (Diptera: Agromyzidae): diverged clades and invasive populations, Ann. Entomol. Soc. Am. 99 (2006) 991–998.
[31] A.T. Beckenbach, Mitochondrial genome sequences of Nematocera (lower Diptera): evidence of rearrangement following a complete genome duplication in a winter crane fly, Genome Biol. Evol. 4 (2) (2012) 89–101.
[32] M.T. Oliveira, J.G. Barau, A.C.M. Junqueira, P.C. Feijão, A.C.d. Rosa, C.F. Abreu, A.M.L. Azeredo-Espin, A.C. Lessinger, Structure and evolution of the mitochondrial genomes of *Haematobia irritans* and *Stomoxys calcitrans*: the Muscidae (Diptera: Calyptratae) perspective, Mol. Phylogenet. Evol. 48 (2008) 850–857.
[33] S.L. Cameron, C.L. Lambkin, S.C. Barker, M.F. Whiting, A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision, Syst. Entomol. 32 (2006) 40–59.
[34] M. Moreno, O. Marinotti, J. Krzywinski, W.P. Tadei, A.A. James, N.L. Achee, J.E. Conn, Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time, Malaria J. 9 (2010) 127.
[35] A.T. Beckenbach, J.B. Joy, Evolution of the mitochondrial genomes of gall midges (Diptera: Cecidomyiidae): rearrangement and severe truncation of tRNA genes, Genome Biol. Evol. 1 (2009) 278–287.
[36] X.K. Li, W.L. Li, S.M. Ding, S.L. Cameron, M. Mao, L. Shi, D. Yang, Mitochondrial genomes provide insights into the phylogeny of Lauxanioidea (Diptera: Cyclorrhapha), Int. J. Mol. Sci. 18 (2017) 773.
[37] Q. Wang, L.F. Chen, J.J. Ding, Y.Q. Liu, J.G. Wang, Sequence and analysis of complete mitochondrial genome of *Simuiium aureohirtum*, J. Environ. Entomol. 2 (2016) 361–370 (In Chinese).
[38] J.C. Day, H.S. Gweon, R.J. Post, Sequence and organization of the complete mitochondrial genome of the blackfly *Simulium variegatum* (Diptera: Simuliidae), Mitochondrial DNA B 1 (2016) 799–801.
[39] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.
[40] C. Hahn, L. Bachmann, B. Chevreux, Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-a baiting and iterative mapping approach, Nucleic Acids Res. 41 (2013) e129.
[41] M. Bernt, A. Donath, F. Juhling, F. Externbrink, C. Florentz, G. Fritzsch, J. Putz, M. Middendorf, P.F. Stadler, MITOS: improved de novo metazoan mitochondrial genome annotation, Mol. Phylogenet. Evol. 69 (2013) 313–319.
[42] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964.
[43] M. Lohse, O. Drechsel, S. Kahlau, R. Bock, OrganellarGenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets, Nucleic Acids Res. 41 (2013) w575–w581.
[44] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874.
[45] N.T. Perna, T.D. Kocher, Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes, J. Mol. Evol. 41 (1995) 353–358.
[46] J.W. Breinholt, A.Y. Kawahara, Phylotranscriptomics: saturated third codon positions radically influence the estimation of trees based on next-gen data, Genome Biol. Evol. 5 (2013) 2082–2092.
[47] M.S. Blouin, C.A. Yowell, C.H. Courtney, J.B. Dame, Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics, Mol. Biol. Evol. 15 (1998) 1719–1727.
[48] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780.
[49] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (2000) 540–552.
[50] Y. Tabei, H. Kiryu, T. Kin, K. Asai, A fast structural multiple alignment method for long RNA sequences, BMC Bioinformatics 9 (2008) 33.
[51] L.T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274.
[52] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat. Methods 14 (2017) 587–589.
[53] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, Bioinformatics 19 (2003) 1572–1574.
[54] D. Posada, K.A. Crandall, MODELTEST: testing the model of DNA substitution, Bioinformatics 14 (1998) 817–818.
[55] X.K. Li, S.M. Ding, S.L. Cameron, Z.H. Kang, Y.Y. Wang, D. Yang, The first mitochondrial genome of the Sepsid Fly *Nemopoda mamaevi* Ozerov, 1997 (Diptera: Sciomyzoidea: Sepsidae), with mitochondrial genome phylogeny of Cyclorrhapha, PLoS One 10 (2015), e0123594.
[56] X. Zhang, Z.H. Kang, M. Mao, X.K. Li, S.L. Cameron, H. de Jong, M.Q. Wang, D. Yang, Comparative Mt genomics of the Tipuloidea (Diptera: Nematocera: Tipulomorpha) and its implications for the phylogeny of the Tipulomorpha, PLoS One 11 (2016), e0158167.
[57] S.L. Cameron, M.F. Whiting, The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths, Gene 408 (2008) 112–123.
[58] D.X. Zhang, J.M. Szymura, G.M. Hewitt, Evolution and structural conservation of the control region of insect mitochondrial DNA, J. Mol. Evol. 40 (1995) 382–391.
[59] K. Wang, X.K. Li, S.M. Ding, N. Wang, M. Mao, M.Q. Wang, D. Yang, The complete mitochondrial genome of the *Atylotus miser* (Diptera: Tabanomorpha: Tabanidae), with mitochondrial genome phylogeny of lower Brachycera (Orthorrhapha)s, Gene 586 (2016) 184–196.
[60] K.R. Reidenbach, S. Cook, M.A. Bertone, R.E. Harbach, B.M. Wiegmann, N.J. Besansky, Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology, BMC Evol. Biol. 9 (2009) 298.
[61] S.N. Kutty, W.H. Wong, K. Meusemann, R. Meier, P.S. Cranston, A phylogenomic analysis of Culicomorpha (Diptera) resolves the relationships among the eight constituent families, Syst. Entomol. 43 (2018) 434–446.
[62] A. Borkent, The pupae of the biting midges of the world (Diptera: Ceratopogonidae), with a generic key and analysis of the phylogenetic relationships between genera, Zootaxa 3879 (2014) 1–327.
[63] A.L. Bazinet, K.T. Mitter, D.R. Davis, E.J. Van Nieukerken, M.P. Cummings, C. Mitter, Phylotranscriptomics resolves ancient divergences in the Lepidoptera, Syst. Entomol. 42 (2017) 305–316.