



OPEN

## Comparative analysis of four complete mitogenomes from hoverfly genus *Eristalinus* with phylogenetic implications

Hu Li<sup>1,2</sup>✉ & Juan Li<sup>1,2</sup>

The genus *Eristalinus* is widely distributed globally. Four complete mitochondrial genomes (i.e., mitogenomes) of *Eristalinus* were sequenced and analyzed in this study: *Eristalinus viridis* (Coquillett, 1898), *E. quinquestriatus* (Fabricius, 1781), *E. tarsalis* (Macquart, 1855), and *E. sp.* Within these four sequenced mitogenomes, most protein-coding genes (*ND2*, *CO1*, *COX2*, *COX3*, *ND3*, *ND5*, *ND4*, *ND4L*, *ND6*, and *Cytb*) began with a typical ATN (T/C/G/A) start codon and ended with a stop codon TAA or incomplete T, whereas *ND1* began with the start codon TTG. *ND3* ended with TAG. The secondary tRNA structure was that of a typical cloverleaf, and only the *tRNA-Ser1* lacked a DHU arm. Three and five domains appeared in the *12S* and *16S* rRNA secondary structures, respectively. The phylogenetic relationships among the four *Eristalinus* species combined with the published mitogenomes of Syrphidae were reconstructed using the maximum likelihood and Bayesian inference methods, which support the monophyly of the subfamily Syrphinae but do not support that of the subfamily Eristalinae. Of note, Eristalini and Syrphini are monophyletic groups. The mitogenomes of *E. viridis*, *E. quinquestriatus*, *E. sp.*, and *E. tarsalis* are useful for determining the phylogenetic relationships and evolution of Syrphidae.

Syrphidae is a large family with high species diversity within the order Diptera, with more than 6000 species<sup>1,2</sup> distributed worldwide. They are well recognized because most have bright black and yellow patterns on the abdomen and are similar to bumblebees, wasps, or honeybees; this mimicry can help syrphids escape from their natural enemies<sup>3</sup>. Different researchers follow different systems with respect to their higher taxonomy classification<sup>4,5</sup>. The most population is the three subfamilies system, Syrphinae, Eristalinae, and Microdontinae<sup>5</sup>. However, recent studies suggest a four subfamilies system: Syrphinae, Eristalinae, Microdontinae, and Pipizinae<sup>6,7</sup>. Based on mtDNA sequences and morphology, Eristalinae has been deemed as paraphyletic group<sup>8</sup>. However, Eristalinae has been deemed as monophyletic group based on sequence data of nuclear 28S rRNA and mitochondrial cytochrome oxidase c subunit 1 (*COI*) genes in conjunction with larval and adult morphological characteristics of syrphid taxa<sup>9</sup>.

The mitochondrial genome is a circular and double-stranded DNA molecule, with a low molecular (14–36 kb), rapid evolution rate, and stable gene composition. Many insect groups utilize their mitochondrial sequences to solve phylogenetic problems<sup>10–13</sup>. Thus far, 20 complete mitogenomes of Syrphidae have been sequenced and uploaded to GenBank. Sonet et al.<sup>14</sup> published five Afrotropical species of *Eristalinus* (*E. barclayi*, *E. fuscicornis*, *E. vicarians*, *E. aeneus* and *E. tabanooides*) and attempted to resolve the phylogenetic relationships of *Eristalinus* from phylogenetic tree analysis and informativeness of 13 protein-coding genes (PCGs) and 2 rRNAs. More molecular data could help establish Syrphidae in a stable classification system and aid in further understanding its evolutionary history.

In previous studies on mitogenomes, Syrphinae was described as a monophyletic group<sup>14–19</sup> and the tribes Syrphini and Melanostomini as sister groups<sup>19,20</sup>. Eristalinae has not been established as a monophyletic group; moreover, Volucellini, Cheilosini, and Milesiini are strongly related groups, and together clustered a clade as a sister group to the Syrphinae. Eristalini is a cluster, but relation of the genera under Eristalini needs to be further discussed<sup>14,21</sup>.

<sup>1</sup>Shaanxi Key Laboratory of Bio-Resources, School of Biological Science & Engineering, Shaanxi University of Technology, Hanzhong 723000, Shaanxi, China. <sup>2</sup>These authors contributed equally: Hu Li and Juan Li. ✉email: lihu@sntu.edu.cn

*Eristalinus* Rondani, 1845 (Diptera: Syrphidae, Eristalinae) is widely distributed worldwide. This genus contains approximately 75 species, with at least 15 distributed in China. Adults typically visit flowers belonging to the Theaceae, Apiaceae, Liliaceae, and Santales families and feed on pollen and nectar. As observed in other hoverflies, species belonging to *Eristalinus* are often involved in entomophily when pollinating and fertilizing plants such as *Eurya emarginata* (Thunb.), *Santalum album* L., *Eryngium horridum* Malme, and *Allium cepa* L.<sup>22–25</sup>. Saprophygy larvae of *Eristalinus* live in various organic-rich substrates, such as around pools, rotting trees, or other plants, and have even colonized a human corpse as shown in a recent study<sup>26</sup>.

Four complete mitogenomes—those of *Eristalinus viridis* (Coquillett, 1998) (GenBank No. MN494096), *E. quinquestriatus* (Fabricius, 1781) (MT471322), *E. tarsalis* (Macquart, 1855) (MW073114), and *E. sp.* (MT942687)—were sequenced and described in this study. We analyzed the genomic structure and nucleotide composition of these four sequenced species and compared these with other Syrphidae<sup>14–18,20,27,28</sup>; furthermore, we reconstructed phylogenetic relationships combined with current mitochondrial genomes. This study aims to compare and elucidate the phylogenetic relationships among *Eristalinus* and Syrphidae.

## Materials and methods

**Ethics statement.** The specimens studied here were collected in the field by net. The field work permission for specimen collection to *Eristalinus viridis*, *E. quinquestriatus*, *E. tarsalis* was approved by the Changqing National Nature Reserve, Hanzhong, Shaanxi, China, and was performed in accordance with relevant guidelines of the reserve, that for *Eristalinus sp.* is needless due to the location was not privately-owned or protected.

**Sampling, genomic DNA extraction, and polymerase chain reaction (PCR) amplification.** Voucher specimens were deposited in the Museum of Zoology and Botany, Shaanxi University of Technology, Hanzhong, China. Specimens of *E. viridis*, *E. quinquestriatus*, *E. tarsalis*, and *E. sp.* were collected from Shaanxi Province, China (Table S1), identified by Hu Li and Juan Li using the works by Huo et al.<sup>2</sup> and Huang and Cheng<sup>3</sup>. After collection, specimens were transported to the laboratory in absolute ethanol and stored at  $-20^{\circ}\text{C}$ .

Genomic DNA was extracted from adult's thorax and legs using the TIANamp Genomic DNA Kit (TIANGEN, Beijing), and the sample volume was 100  $\mu\text{l}$  for each species. Specific experimental procedures were strictly carried out following the manufacturer's instructions. Genomic DNA concentration reached 20 ng/ $\mu\text{l}$  or more, then at least 50  $\mu\text{l}$  of the sample was sent to Berry Genomics (Beijing, China) for sequencing, the entire mitogenomes of the four species were sequenced using an Illumina NovaSeq6000 platform with 150 bp paired-end reads and insert size of 350 bp, and all four voucher specimens generated 6 GB high-throughput data.

The remaining sample was used for PCR amplification of *COI* by Sangon Biotech (Shanghai, China). The *COI* sequence as a bait sequence was used to obtain the whole mitogenomes sequence. Taq PCR Master Mix (2 $\times$ , blue dye) (BBI Life Sciences, Shanghai) was used in the PCR amplification of *COI*. The primers used for PCR amplification of the *COI* gene were universal for invertebrate phyla (Table S2)<sup>29</sup>. PCR amplification included an initial denaturing step at 94  $^{\circ}\text{C}$  for 4 min, followed by 35 cycles of denaturation at 94  $^{\circ}\text{C}$  for 30 s, annealing at 45  $^{\circ}\text{C}$  for 30 s, elongation at 72  $^{\circ}\text{C}$  for 45 s, and a final elongation step at 72  $^{\circ}\text{C}$  for 10 min. PCR amplification procedures of *COI* were carried out following the manufacturer's manual.

**Mitochondrial genome sequencing, assembly, and annotation.** Complete mitogenomes were assembled using Geneious Prime (v2019 1.3.)<sup>30</sup> combined with PCR amplification of *COI* sequences while using *Eristalinus aenax* (MH321208) and *Eristalinus tabanoides* (MH321207) as references to confirm the accuracy.

The secondary structure and position of 22 tRNAs were predicted by ARWEN version 1.2<sup>31</sup> and tRNAscan-SE version 1.21<sup>32</sup> and were checked manually. Those tRNAs that could not be found were confirmed by alignment with homologous sequences from related species. PCGs were annotated with Geneious Prime (v2019 1.3.)<sup>30</sup> by detecting an open reading frame, which was also confirmed based on BLAST query in GenBank using a published mitogenome from Syrphidae. For rRNA gene identification, the 16S rRNA gene was located between the *tRNA-L2* and *tRNA-V*; the 12S rRNA gene was identified based on comparison with other related species. The secondary structures of 16S rRNA and 12S rRNA were predicted according to data from other species, tobacco hornworm, *Apis mellifera*, *Scopura longa*, and *Andrena chekiangensis*<sup>33–36</sup>. Helical elements were predicted using ClustalX 1.81<sup>37</sup> and RNA Structure 5.2<sup>38</sup>. The control region was identified by the boundaries of *tRNA-I* and 12S rRNA.

The four species' mitogenome maps in this study were produced using CG View online server using default parameters ([http://stothard.afns.ualberta.ca/cgview\\_server/](http://stothard.afns.ualberta.ca/cgview_server/))<sup>39</sup>. Nucleotide composition was calculated using MEGA 6.0<sup>40</sup>. The AT and GC skew were calculated manually according to formulas: AT skew =  $(A\% - T\%) / (A\% + T\%)$  and GC skew =  $(G\% - C\%) / (G\% + C\%)$ <sup>41</sup>. The codon usage and relative synonymous codon usage (RSCU) of each PCGs were calculated using MEGA 6.0<sup>40</sup>. Homology between control region repeat units in *Eristalinus* species with the control region of other species was determined using a ClustalW<sup>37</sup> sequence alignment implemented in MEGA 6.0<sup>40</sup>. The numbers of nonsynonymous substitutions per nonsynonymous site (Ka) and synonymous substitutions per synonymous site (Ks) were calculated for Syrphidae species using DnaSP v4<sup>42</sup>. The ratio of Ka/Ks was checked manually.

**Phylogenetic analysis.** A total of 24 species mitogenomes of Syrphidae, including four newly sequenced *Eristalinus* species, were used for phylogenetic analyses (Table 1). *Nemopoda mamaevi* (Sepsidae)<sup>43</sup> and *Cestrotus liui* (Lauxaniidae)<sup>44</sup> were used as outgroups.

Each PCG was aligned individually with codon-based multiple alignments using the MAFFT algorithm in the TranslatorX server<sup>45</sup> ([http://pc16141.mnncn.csic.es/index\\_v4.html](http://pc16141.mnncn.csic.es/index_v4.html)). The two rRNA genes were aligned using the MAFFT v7 online server with G-INS-i strategy<sup>46</sup>, and poorly aligned positions were eliminated using Gblocks

	Species	Accession number	Length/bp	Reference
Syrphidae	<i>Episyrphus balteatus</i>	KU351241	16,175	Pu et al. <sup>20</sup>
	<i>Eristalis cerealis</i>	NC050932	15,348	Yan et al. <sup>27</sup>
	<i>Eristalis tenax</i>	MH159199	15,996	Li et al. <sup>21</sup>
	<i>Eristalinus barclayi</i>	MH321205	15,757	Sonet et al. <sup>14</sup>
	<i>Eristalinus fuscicornis</i>	MH321204	15,815	Sonet et al. <sup>14</sup>
	<i>Eristalinus vicarians</i>	MH321206	15,966	Sonet et al. <sup>14</sup>
	<i>Eristalinus aeneus</i>	MH321208	16,245	Sonet et al. <sup>14</sup>
	<i>Eristalinus tabanoides</i>	MH321207	15,792	Sonet et al. <sup>14</sup>
	<i>Eristalinus quinquestriatus</i>	MT471322	15,872	This study
	<i>Eristalinus viridis</i>	MN494096	15,640	This study
	<i>Eristalinus</i> sp.	MT942687	15,883	This study
	<i>Eristalinus tarsalis</i>	NW073114	15,849	This study
	<i>Eristalinus quinquestriatus</i>	MT834869	16,198	Unpublished
	<i>Eupeodes corolla</i>	KU379658	15,326	Pu et al. <sup>20</sup>
	<i>Helophilus virgatus</i>	MN148445	15,742	Li et al. <sup>22</sup>
	<i>Korinchia angustabdomena</i>	MK870078	16,473	Li <sup>23</sup>
	<i>Melanostoma orientale</i>	MN788095	16,229	Chen et al. <sup>24</sup>
	<i>Melanostoma scalare</i>	MN481591	16,126	Unpublished
	<i>Ocyptamus sativus</i>	KT272862	15,214	Junqueira et al. <sup>25</sup>
	<i>Phytomia zonata</i>	MT478107	15,716	Li et al. <sup>26</sup>
	<i>Platycheirus albimanus</i>	MT622646	16,648	Unpublished
	<i>Simosyrphus grandicornis</i>	DQ866050	16,141	Cameron et al. <sup>10</sup>
	<i>Ferdinandea cuprea</i>	MT834868	15,907	Unpublished
<i>Volucella nigricans</i>	MK870079	15,724	Li <sup>23</sup>	
Outgroup	<i>Nemopoda mamaevi</i>	NC026866	15,878	Li et al. <sup>42</sup>
	<i>Cestrotus liui</i>	NC034922	16,171	Li et al. <sup>43</sup>

**Table 1.** Information of complete mitogenomes used for phylogenetic analysis in this study.

0.91b<sup>47</sup> (<https://mafft.cbrc.jp/alignment/server/>). Finally, the aligned sequences of 13 PCGs and two rRNA genes were concatenated manually with MEGA 6.0<sup>40</sup>.

PartitionFinder2<sup>48</sup> on CIPRES Science Gateway<sup>49</sup> was used to select the best-fit partitioning schemes and substitution models for the dataset matrix. The greedy algorithm was employed using unlinked branch lengths, and under the Akaike information criterion to select the optimal partitioning model. Information concerning partition strategies and substitution models used are summarized in Table S3. Phylogenetic trees were constructed using the maximum likelihood method (ML) on the IQ-TREE webserver (<http://www.iqtree.cibiv.univie.ac.at/>)<sup>50</sup> and Bayesian inference (BI) on MrBayes 3.2.6<sup>51</sup> within the CIPRES webserver (<https://www.phylo.org/porta12/login!input.action>)<sup>49</sup> based on the sequences of 13 PCGs and two rRNA genes (PCGRNA, PCG123, PCG12, PCG12RNA, and AA datasets), respectively.

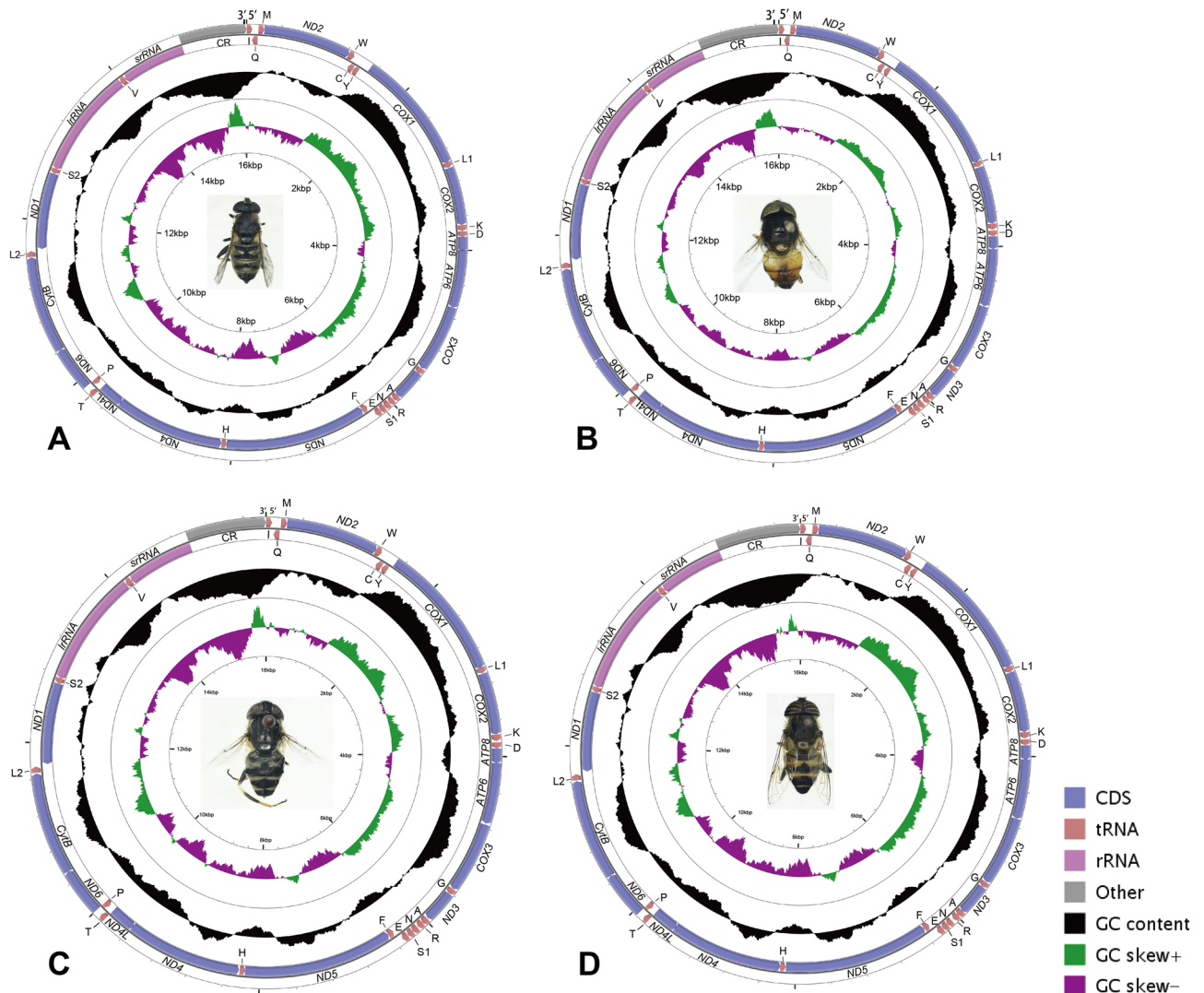
PCGRNA indicates the sequences from PCGs and rRNAs; PCG123 represents all bases PCGs; PCG12 represents all bases excluding the third bases of each protein-coding amino acid genes; PCG12RNA represents the sequences of PCG12 and rRNA. AA represents the amino acids translated from 13 PCGs.

The ML analyses were conducted by using 10,000 replications with the “ultrafast” function<sup>52</sup>. For BI analyses, two simultaneous Markov chain Monte Carlo runs running parameters ran 1 million generations with sampling every 1,000 generations. The initial 25% of the sampled data were discarded as burn-ins.

## Results and discussion

**Genome organization and base composition.** The complete mitogenomes of *E. viridis*, *E. quinquestriatus*, *E. tarsalis*, and *E. sp.* were sequenced; the total length of each genome was 15,640 bp, 15,872 bp, 15,849 bp, and 15,883 bp respectively, with each mitogenome including 37 genes (13 PCGs, 22 tRNAs, 2 rRNAs) and non-coding regions (Fig. 1, Tables S4, S5, S6, S7). A total of 23 genes were encoded on the J-strand and another 14 were located on the N-strand, which is consistent with the mitogenome sequences of other Syrphidae species<sup>14–18,20,27,28</sup>. Within Diptera, mitogenomes were found to have gene rearrangements in mosquitoes (Culicidae)<sup>53,54</sup> and gall midges (Cecidomyiidae)<sup>23</sup>. The gene arrangement within the mitogenomes of Syrphidae was the same as that within the mitogenome of the ancestral insect (*Drosophila yakuba*)<sup>55</sup>. The length of Syrphidae genomes ranged from 15,326 to 16,473 bp, with *Korinchia angustabdomena* having the longest genome (16,473 bp) and *Eupeodes corolla* having the shortest (15,326 bp). Overall, the genomic size of the species within these taxa is medium compared with that of other insects.

Within *Eristalinus*, there are 10 complete mitogenomes sequenced in total, including these 4 newly sequenced mitogenomes in the current study and 6 from GenBank. The 10 complete mitogenomes from 9 *Eristalinus*



**Figure 1.** The circle maps of the complete mitochondrial genomes of *Eristalinus*. (A) *E. viridis*, 15,640 bp, (B) *E. quinquestriatus*, 15,872 bp, (C) *E. tarsalis*, 15,849 bp, (D) *E. sp.*, 15,883 bp. PCGs, tRNAs, rRNAs have been given, CR represented control region. The arrow indicates the direction of the gene from start to end. A black pattern shows GC content: four species, pictures taken by Juan Li using CG view ([http://stothard.afns.ualberta.ca/cgview\\_server/](http://stothard.afns.ualberta.ca/cgview_server/)).

species available have a high A + T content: 80.6% in *E. quinquestriatus* (MT834869), 80.2% in *E. quinquestriatus* (MT471322, this study), 80.0% in *E. tabanoides* and *E. viciatians*, 79.9% in *E. fuscicornis* and *E. barclayi*, 79.8% in *E. aeneus* and *E. sp.*, 78.2% in *E. viridis*, and 79.0% in *E. tarsalis*. Furthermore, all species exhibited strong AT bases, and all consisted of positive AT and negative GC skew (Table 2).

**Protein-coding genes and codon usage.** Within the four species sequenced, most PCGs began with a typical ATN (T/C/G/A) start codon and ended with a stop codon TAA or incomplete T; the incomplete T is a common stop codon in insects, which modified into complete TAA via posttranscriptional polyadenylation during mRNA maturation<sup>56</sup> (Tables S4, S5, S6, S7). Within the 10 mitogenome sequences of *Eristalinus* species, *ND1* was relatively conservative and always utilized TTG as the start codon. For stop codons, *ND3* contained TAG, *ND5* ended with an incomplete stop codon T, whereas all others utilized TAA as the stop codon.

PCGs exhibited negative AT (−0.134 to −0.12) and positive GC (0–0.037) skew among the four *Eristalinus* species (Table 2). They all exhibited rich A + T content. The most frequently used amino acids were leucine (Leu), phenylalanine (Phe), and isoleucine (Ile) within the four sequenced mitogenomes of *Eristalinus* (Fig. 2). Summarization of the RSCU showed that each gene codon usage exhibited a strong AT bias and was primarily composed of the nucleotides A and T. The most frequently used codons for all amino acids utilized the nucleotides A or T in the third codon positions (Fig. 2).

Ka/Ks ( $\omega$ ) analysis is a statistical diagnostic method used to detect the form of sequence evolution<sup>57,58</sup>. The 13 PCGs in the Syrphidae mitochondrial genome have values of Ka/Ks < 1, indicating that all these PCGs are under purifying selection (Fig. 3). The gene *ATP8* ( $\omega = 0.553$ ) was predicted to have evolved most rapidly, followed by

Region		Total (bp)	T%	C%	A%	G%	ATskew	GCskew
<i>E. viridis</i>	Whole	15,640	37.3	13.1	40.9	8.6	0.046	-0.207
	PCGs	11,122	43.1	11.9	33.0	11.9	-0.133	0
	tRNAs	1481	39.4	8.6	40.5	11.5	0.0134	0.144
	rRNAs	2126	42.9	5.7	39.7	11.7	-0.039	0.345
	Control region	784	43.1	4.7	49.9	2.3	0.073	-0.343
<i>E. quinquestriatus</i>	Whole	15,872	39.2	11.6	41.0	8.3	0.022	-0.166
	PCGs	11,168	44.3	10.5	33.9	11.3	-0.133	0.037
	tRNAs	1490	39.7	8.7	39.8	11.8	0.001	0.151
	rRNAs	2130	42.9	5.6	40.8	10.7	-0.025	0.313
	Control region	959	46.4	2.8	48.7	2.1	0.024	-0.143
<i>E. sp.</i>	Whole	15,883	38.9	12.0	40.9	8.3	0.025	-0.182
	PCGs	11,170	43.9	10.8	33.8	11.5	-0.12	0.031
	tRNAs	1493	39.4	8.5	40.4	11.7	0.013	0.158
	rRNAs	2139	42.4	5.8	40.6	11.1	-0.022	0.314
	Control region	960	49.6	3.2	45.7	1.5	-0.040	-0.368
<i>E. tarsalis</i>	Whole	15,849	38.2	12.6	40.8	8.5	0.033	-0.194
	PCGs	11,167	43.5	11.4	33.2	11.9	-0.134	0.215
	tRNAs	1488	39.3	8.5	40.5	11.8	0.015	0.163
	rRNAs	2126	42.6	5.7	40.4	11.3	-0.027	0.329
	Control region	883	47.3	5.1	45.6	1.9	-0.018	-0.457

**Table 2.** Nucleotide composition of the four sequenced species complete mitogenomes of *Eristalinus*.

*ND6* ( $\omega = 0.360$ ), *ND5* ( $\omega = 0.221$ ), *ND2* ( $\omega = 0.219$ ), and *ND4* ( $\omega = 0.217$ ); the gene *COI* ( $\omega = 0.0712$ ) was shown to be the most conservative. Concerning gene-specific substitution rates, *Ks* ranged from 0.206 at gene *ATP8* to 0.360 at gene *COI*, while the *Ka* varied from 0.023 at gene *COI* to 0.115 at gene *ND6*. Because the selection pressures upon *ATP8* and *ND6* are relatively weak, and these genes accordingly are relatively unconserved, *COI* and *ND1* are under strong selection pressures and are therefore more conservative, consistent with other Diptera species<sup>59</sup>.

**tRNAs and rRNAs.** Among 22 tRNAs, 8 were encoded on the N-strand and the remaining were encoded on the J-strand (Fig. 1; Tables S4, S5, S6, S7). When comparing the tRNA genes of Syrphidae, the length of all tRNAs was found in the range of 58–72 bp (*tRNA-Lys* in *Ocyptamus sativus* and *tRNA-Val* in 24 sequenced Syrphidae species, respectively). The secondary structure of tRNA genes was a typical cloverleaf structure including a discriminator nucleotide, acceptor stem, T $\psi$ C arm, variable loop, anticodon arm and DHU arm (Fig. 4). In the four sequenced species, the DHU arm was found to be missing in only the *tRNA-SI* gene, whereas the remaining were standard structures, consistent with those of other Syrphidae<sup>14,16,17,27</sup> (Fig. 4). In addition, base mismatches were found in all four species sequenced. 12 G–T and 6 T–T mismatches were found in *E. viridis*, as well as 18 G–T and 5 T–T mismatches found in *E. quinquestriatus*, 18 G–T, 3 T–T, 1 A–G, and 1 C–T mismatches found in *E. sp.*, 17 G–T and 4 T–T mismatches found in *E. tarsalis* (Fig. 4).

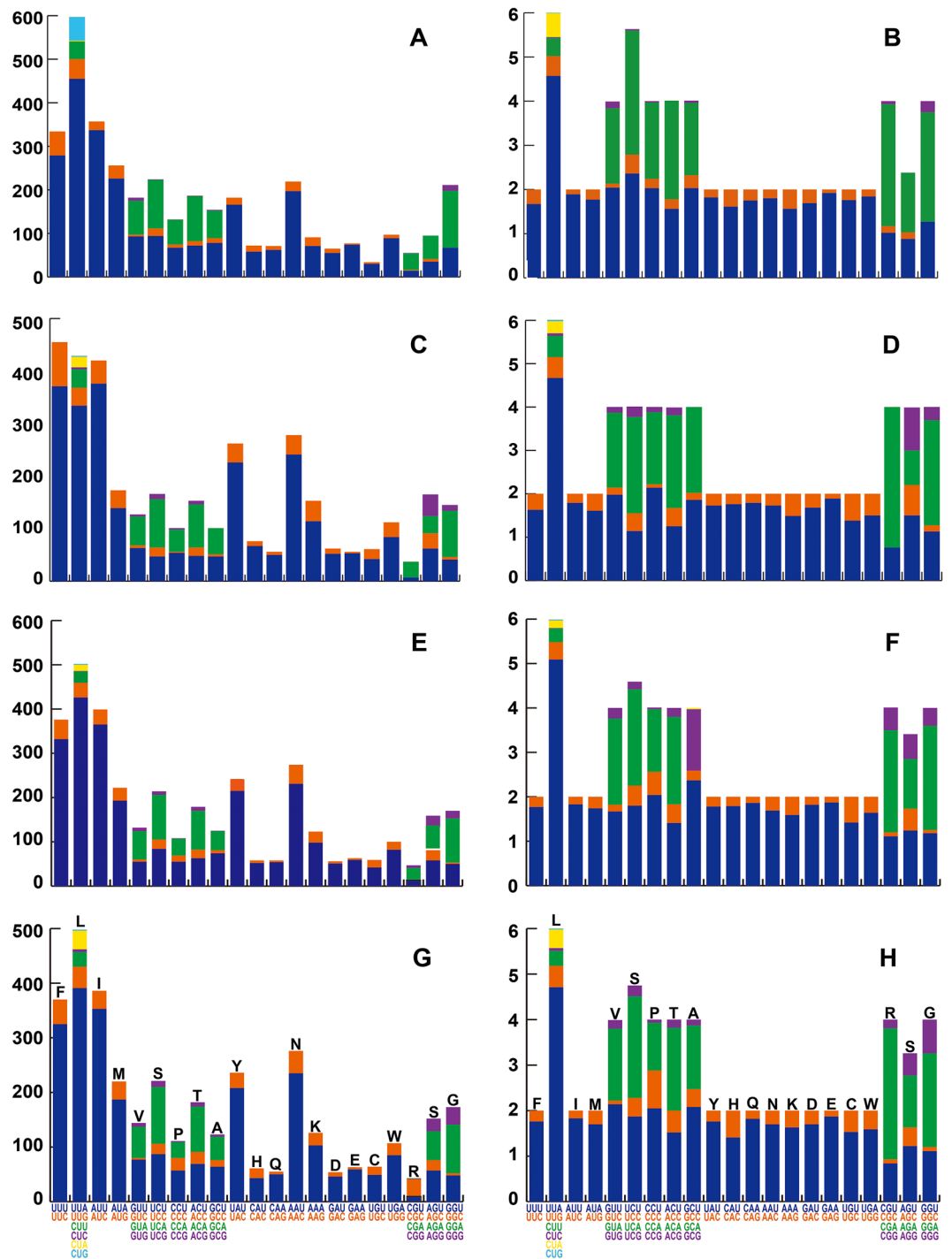
The nucleotide composition of these 22 tRNA genes was significantly biased to A and T nucleotides. The *E. viridis* mitogenome contains 1,481 bp with an A + T content of 79.9%, the *E. quinquestriatus* contains 1,490 bp with an A + T content of 79.5%, the *E. sp.* mitogenome contains 1,493 bp with an A + T content of 79.8%, the *E. tarsalis* mitogenome contains 1,488 bp with an A + T content of 79.8% (Table 2). Four species exhibit positive AT and GC skew.

The two ribosomal RNA genes contain both 16S and 12S rRNA, between either the *tRNA-L2* and *tRNA-V* or between the *tRNA-V* and control region, respectively (Fig. 1). Within the Syrphidae mitogenome, the 16S rRNA length ranged from 1,313 bp (*Ornithopus sativus*) to 1,414 bp (*Melanostoma scalare*), and that of 12S rRNA ranged from 778 bp (*O. sativus* and *M. scalare*) to 824 bp (*K. angustabdomena*).

For *E. viridis*, the length of 12S and 16S rRNA was 793 bp and 1,333 bp, respectively, with an A + T content of 82.6%. Within *E. quinquestriatus*, the 12S and 16S rRNA were 793 bp and 1,337 bp long, with an A + T content of 83.7%; in *E. sp.*, the 12S and 16S rRNA were 793 bp and 1,346 bp in length, with an A + T content of 83.0%; and in *E. tarsalis*, the 12S and 16S rRNA were 790 bp and 1,336 bp long, with an A + T content of 83.0% (Table 2).

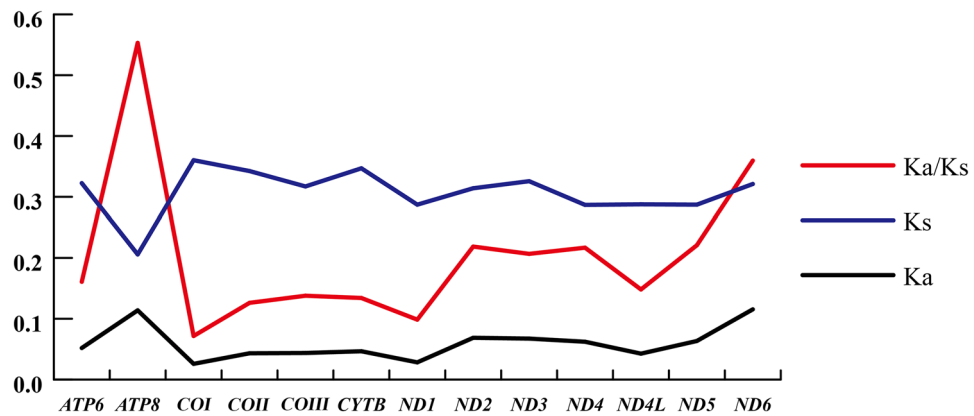
Among the four species of *Eristalinus*, the secondary structure of 16S rRNA includes 5 domains (I, II, IV, V, and VI; domain III was absent in arthropods.) and 43 helices (Fig. 5). Multiple alignments of four species' 16S rRNA gene extended over 1,350 positions and included 1,242 conserved and 108 variable sites. Domain IV was more conserved than other domains structurally.

The secondary structure of 12S rRNA contains 3 domains (I, II, and III) and 24 helices (Fig. 6). Multiple alignments of the four species' 12S rRNA gene extended over 700 positions and comprised 655 conserved and 45 variable sites. Domain III was more conserved than other domains structurally.



**Figure 2.** The codon usage (A, C, E, G) and Relative Synonymous Codon Usage (RSCU) (B, D, F, H) in the mitogenomes of *Eristalinus*. (A, B) *E. viridis*, (C, D) *E. quinquestriatus*, (E, F) *E. sp.*, (G, H) *E. tarsalis*. The code color in the horizontal axis corresponds to the same color in the Figures. The image was computed by MEGA 6.0 (<http://www.megasoftware.net/previousVersions.php>).

**Non-coding region.** The non-coding region contains two parts: gene intervals and a control region (CR; AT-rich region). The *E. viridis* mitogenome contains 10 gene intervals ranging from 1 to 13 bp and has 15 pairs of gene overlaps ranging from 1 to 29 bp (Table S4). The mitogenome of *E. quinquestriatus* contains 12 gene intervals ranging from 1 to 9 bp, with 15 pairs of gene overlaps ranging from 1 to 32 bp (Table S5). The *E. sp.* mitogenome contains 14 gene intervals ranging from 1 to 34 bp and has 11 pairs of gene overlaps ranging from 1 to 9 bp (Table S6). The *E. tarsalis* mitogenome contains 19 gene intervals ranging from 1 to 36 bp and has 7 pairs of gene overlaps ranging from 1 to 7 bp (Table S7).



**Figure 3.** The Ka/Ks analysis of 13 protein-coding genes of the mitochondrial genomes of Syrphidae. The image was computed by DnaSP v4 (<http://www.ub.es/dnasp>).

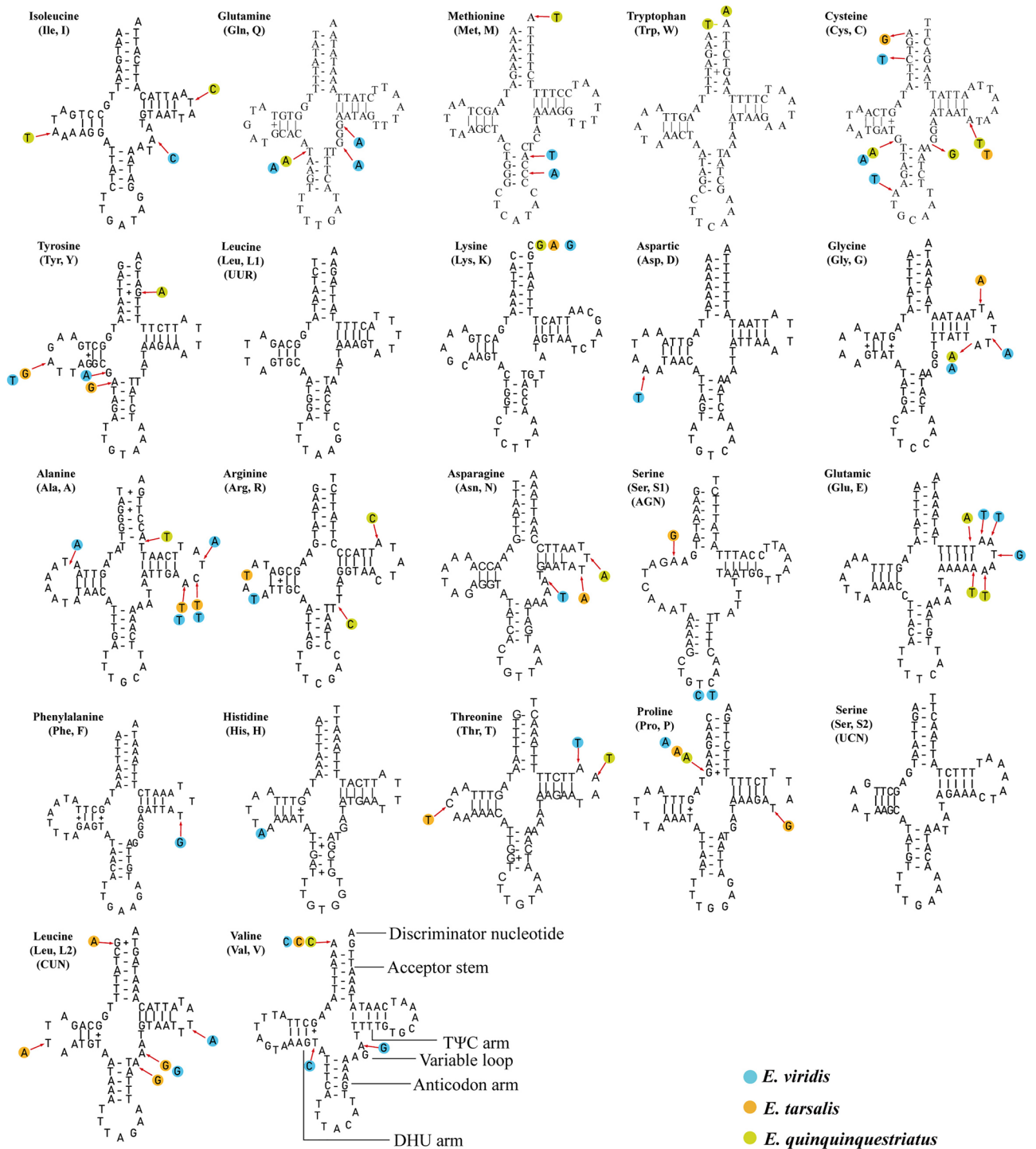
The CR is the largest non-coding region with the largest variation in the entire mitochondrial genome sequence and length. It is primarily involved in mitochondrial DNA replication and transcription<sup>60</sup>. The CRs of *E. viridis*, *E. quinquestriatus*, *E. sp.*, and *E. tarsalis* are 784 bp, 959 bp, 960 bp, and 833 bp in length, respectively (Table 2). This region was shown to be significantly rich in AT with an A + T content of 93.0% in *E. viridis*, 95.1% in *E. quinquestriatus*, 95.3% in *E. sp.*, 92.9% in *E. tarsalis* (Table 2). The CRs of the four species showed slightly positive AT skews and negative GC skews. The CR length in the *Eristalinus* genus is highly variable, ranging in size from 735 bp (*E. tabanoides*) to 1284 bp (*E. quinquestriatus* (MT834869)).

In this study, the CRs of the 10 complete mitogenomes were also observed to contain some structural elements (Fig. 7), such as poly-T/A stretch, stem-loops, and tandem repeats<sup>60</sup>. Nearly all species contained poly-T/A structures in random positions—except for *E. viridis*—with at least one poly-T/A structure near the *tRNA-I*. Stem-loop structures were found close to the boundary of the *tRNA-Ile* (Fig. 7). When comparing CRs of the *Eristalinus* species, we found a conservative sequence of 68 bp containing a stem-loop structure (Fig. S7). Within the CRs of 10 mitogenomes, the “(TA)<sub>n</sub>” (n ≥ 5) motif was common (Fig. 7). There were multiple repetitive units observed, and while only *E. aenax* (Fig. 7G) and *E. quinquestriatus* (GenBank: MT834869) (Fig. 7I) contained one tandem repeat unit, those repeat units within the other species were scattered throughout the CRs (Fig. 7). The number of repeat units varied (*E. tarsalis* has not repeat units), *E. aenax* contained the largest number with four repeat units (Fig. 7G), *E. fuscicornis* (Fig. 7B), *E. quinquestriatus* (Fig. 7D), *E. viridis* (Fig. 7F), and *E. sp.* (Fig. 7H) contained the least with one repeat unit. The repeat unit genomic locations for each species were random. The longest repeat unit sequence was 353 bp long and the shortest was 11 bp long (Fig. 7).

**Phylogenetic analysis.** Phylogenetic analysis employing the PCGRNA (13,187 bp), PCG123 (11,106 bp), PCG12 (7,404 bp), PCG12RNA (9,485 bp), and AA (3701 amino acids) datasets from 24 hoverflies and 2 out-group species showed nearly similar topologies with strong node support under both ML and BI methods (Figs. 8, S8, S9, S10, S11). Because the topology of the PCGRNA datasets significantly conforms to the morphological classification and more previous research, trees from the PCGRNA dataset were chosen in our study<sup>14–19</sup>. In this study, we provided more mitogenome sequences and datasets conducive to understanding the relationship within Syrphidae.

The subfamily Syrphinae has been reconstructed into a monophyletic group supported in both ML and BI analyses [bootstrap support values (BS) = 68, Bayesian posterior probability (pp) = 0.829] (Fig. 8). The tribes Syrphini and Melanostomini are sister taxa with strong support and the six genera relationship of Syrphinae—(*Melanostoma* + (*Platycheirus* + (*Eupeodes* + (*Ocyptamus* + (*Simosyrphus* + *Episyrphus*))))—are strongly supported, and these results are consistent with the previous studies<sup>14,20–23,26</sup>.

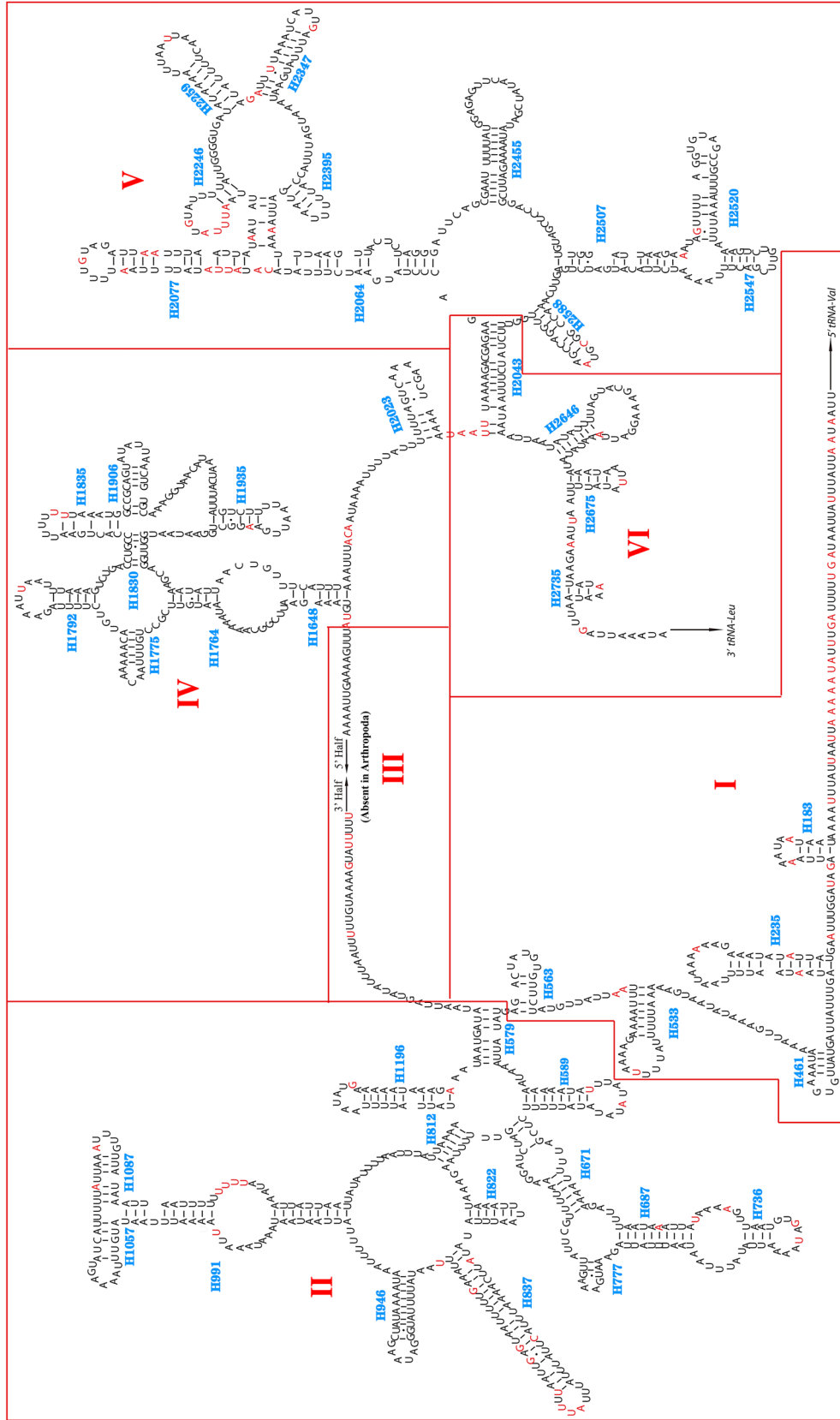
Two main types of clades can be observed in all phylogenetic trees except BI analyses based on the AA dataset (Figs. 8, S8, S9, S10, S11). One included the tribe Volucellini that diverged first, then Milesiini and Cheilosini, and their phylogenetic relationships are as follows: (Syrphinae + (*Volucella* + (*Ferdinandea* + *Korinchia*))) (Figs. 8, S8, S9, S10, S11). Furthermore, the other one was that 14 species of Eristalini were clustered as a single clade with a higher internal node (Figs. 8, S8, S9, S10, S11). *Volucella nigrican* was the sister species to the other Syrphidae only in BI analyses based on the AA dataset (Fig. S11). All Eristalinae species were not gathered into one clade in all 10 trees based on different datasets of mitogenome sequences datasets. According to the topologies obtained from five datasets based on ML and BI methods, Eristalinae tends to be a paraphyletic group. A phylogenetic relationship study of 27 species of Diptera based on *16S*, *12S*, and morphology<sup>8</sup> showed Eristalinae to be paraphyletic. Based on *COI*, *28S*, *18S*, and 111 adults morphological, the phylogenetic relationships indicate that Eristalinae was resolved as non-monophyletic<sup>6</sup>. Eristalinae larvae feeding habits are complex; mitogenomes in the phylogenetic relationship in subfamily Eristalinae contain three tribes (Eristalini, Cheilosini, and Volucellini). Eristalini larvae are aquatic saprophages<sup>2</sup>, Volucellini larvae specialized inquilines in social insect nests, and some feed on dead or dying insects<sup>3,61</sup>. The feeding habits of Volucellini larvae may be close to those of Syrphinae.



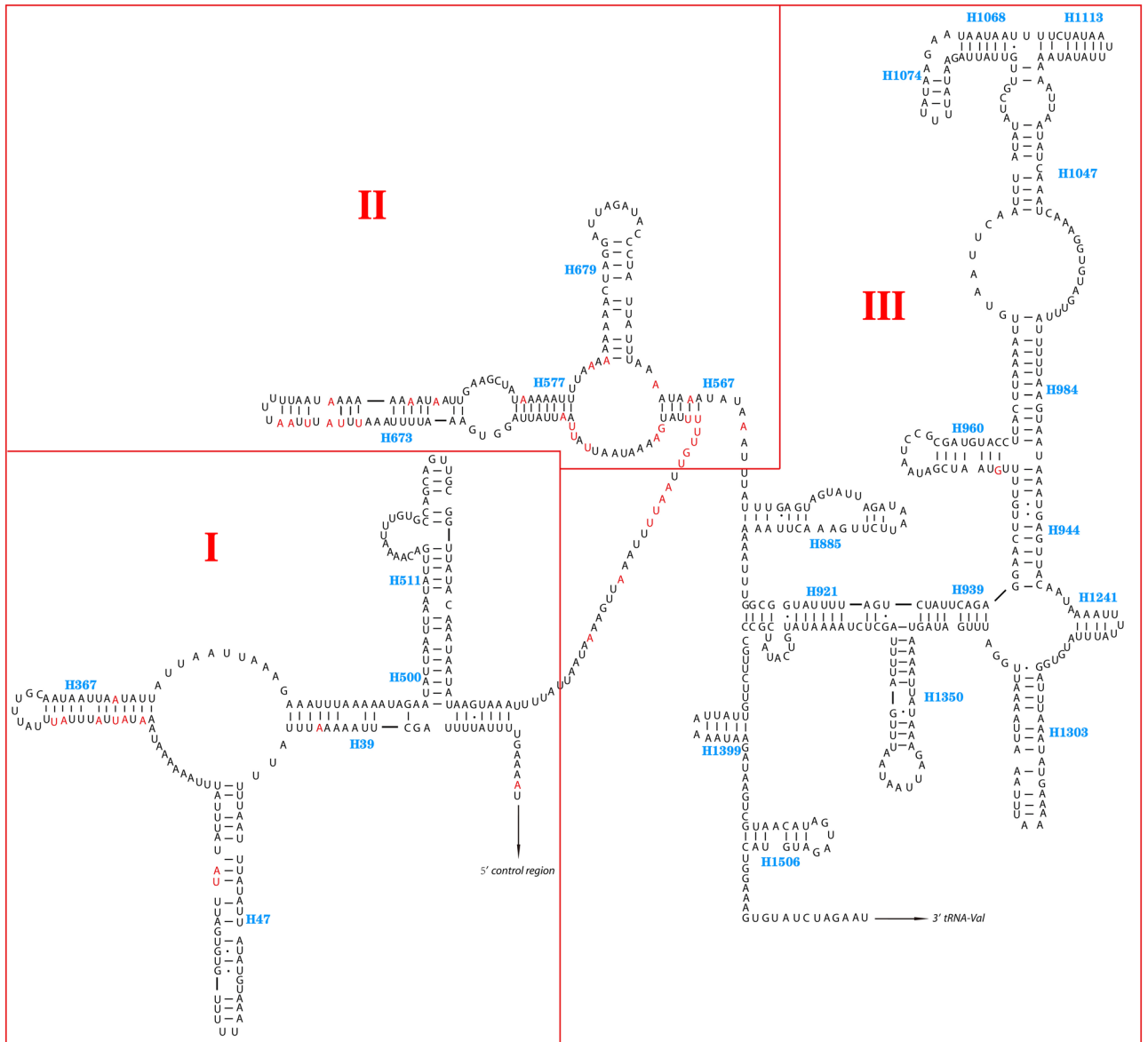
**Figure 4.** Predicted secondary cloverleaf structure for tRNA of four *Eristalinus* species. (A) Isoleucine, (B) Glutamine, (C) Methionine, (D) Tryptophan, (E) Cysteine, (F) Tyrosine, (G, U) Leucine, (H) Lysine, (I) Aspartic, (J) Glycine, (K) Alanine, (L) Arginine, (M) Asparagine, (N, T) Serine, (O) Glutamic, (P) Phenylalanine, (Q) Histidine, (R) Threonine, (S) Proline, (V) Valine. Arrows indicate variations of each site in four species of *Eristalinus*. Each species is marked by unique color (see color legend). The image was predicted by ARWEN version 1.2 (<http://130.235.244.92/bcgi/arwen.cgi>) and tRNAscan-SE version 1.21 (<http://lowelab.ucsc.edu/tRNAscan-SE>), drawing with Adobe Illustrator 2020.

Within Eristalini, phylogenetic trees contained four genera in this study, its major phylogenetic relationships congruently cluster as (*Helophilus* + (*Eristalis* + (*Phytomia* + *Eristalinus*))) (Fig. 8, S8, S9, S10, S11); this result was consistent with the relationships based on *COI* by Sonet et al.<sup>14</sup>, but disagreed with that based on *Cyt b* by





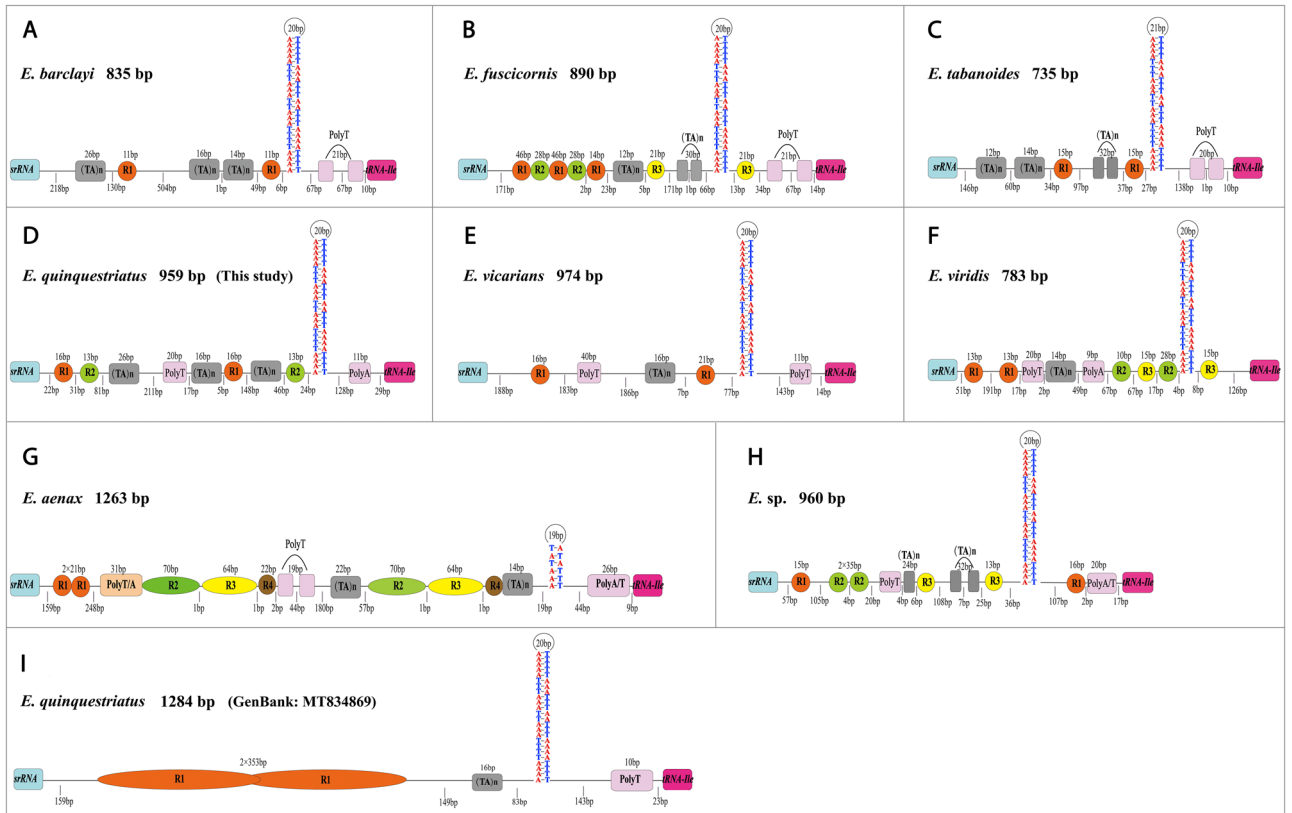
**Figure 5.** Predicted secondary structure for 16S rRNA of *E. sp.* The red color indicates the variation of nucleotide sites in four species. The names of helices are shown in blue “H + numbers.” Areas surrounded by red lines indicate different domains and are respectively numbered I, II, IV, V, and VI in red as in other insects. The image was predicted by ClustalX 1.81 (<http://www.hgmp.mrc.ac.uk/Registered/Option/clustalx.html>) and RNA Structure 5.2 (<http://rna.urmc.rochester.edu/RNAstructure.html>), drawing with Adobe Illustrator 2020.



**Figure 6.** Predicted secondary structure for 12S rRNA of *E. sp.* The red color indicates the variation nucleotide sites in four species. The names of helices showed in blue “H + numbers.” Areas surrounded by red lines indicate different domains and are respectively numbered I, II, and III in red as in other insects. The image was predicted by ClustalX 1.81 (<http://www.hgmp.mrc.ac.uk/Registered/Option/clustalx.html>) and RNA Structure 5.2 (<http://rna.urmc.rochester.edu/RNAstructure.html>), drawing with Adobe Illustrator 2020.

Zhang & Huo<sup>21</sup>, in which, *Phytomia* diverged first, and *Eristalis* and *Eristalinus* clustered as a sister group, with the cluster formed as (*Phytomia* + (*Helophilus* + (*Eristalis* + *Eristalinus*))). In this study, *Helophilus* is a sister clade to other genera within the Eristalini group and was highly supported [BS = 100, pp = 1] (Fig. 8, S8, S9, S10, S11).

In the genus *Eristalinus*, the seven topologies generated from all five datasets have shown that *E. viridis* is the sister species of other species (Fig. 8, S8, S9, S10, S11). Furthermore, *E. fuscicornis*, *E. barclayi*, *E. tarsalis* and *E. aeneus* are sister species with strong bootstrap support values and posterior probabilities in pair (Fig. 8, S8, S9, S10, S11), it is almost identical with Sonet et al.<sup>14</sup>.



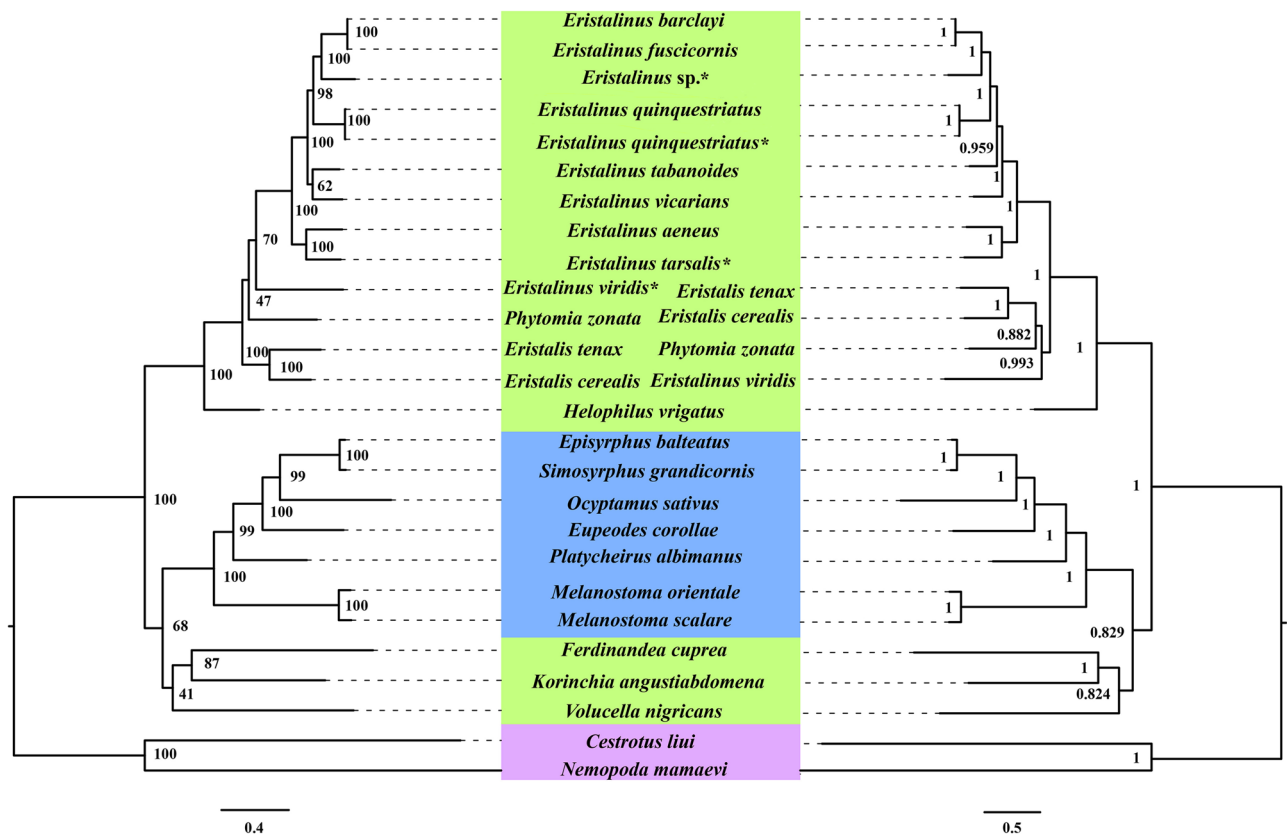
**Figure 7.** Control Regions (CRs) of *Eristalinus*. (A) *E. barclayi*, (B) *E. fuscicornis*, (C) *E. tabanoides*, (D) *E. quinquestriatus*, (E) *E. vicarians*, (F) *E. viridis*, (G) *E. aenax*, (H) *E. sp.*, (I) *E. quinquestriatus*. Different shapes and colors represent the different kinds of sequences. “R” refers to repeat units. The image was predicted by Geneious Prime 2019 ([www.geneious.com](http://www.geneious.com)), drawing with Adobe Illustrator 2020.

The relationships within the genus *Eristalinus* from 3 ML (PCG12RNA, PCG123, and PCG12 datasets) and 2 BI (PCG12 and PCG12RNA datasets) phylogenetic trees are inconsistent. *E. tabanoides* and *E. vicarians* are gathered in a sister relationship, but it is not exposed in other five trees. *E. viridis*' status may need to be further verified and discussed because it gathered into the clade under the genus *Eristalinus* in most seven trees but not in only three trees (AA, PCGRNA and PCG123 datasets from BI inference). These issues may be related to the selections of datasets, the different methods (ML and BI), or limited numbers of complete mitogenomes in the family Syrphidae<sup>62,63</sup>. The morphological characteristic of *E. viridis* is differences with other *Eristalinus*, which its eyes without any spots. This characteristic may suggest its unstable branch in phylogenetic relationship.

### Conclusions

The complete mitogenomes of *E. viridis*, *E. quinquestriatus*, *E. sp.*, and *E. tarsalis* were sequenced and described in the present study. No gene arrangement was found in either of these sequences, and the gene order and direction were similar to the arthropod ancestral mitochondrial genome. Among *Eristalinus*, a conserved stem-loop structure exists near the 3' end of the CR.

Phylogenetic analyses of 24 Syrphidae species support the monophyly of Syrphinae but do not support that of Eristalinae based on ML and BI methods using five datasets. The phylogenetic relationships constructed using the complete mitogenomes effectively interpreted the genus-level relationships within Eristalini, stating that species of *Helophilus* diverged first, followed by *Eristalis*, *Phytomia* and *Eristalinus*. The complete mitochondrial genomes sequenced in this study provided valuable data, which would be useful for determining the phylogenetic relationship of Syrphidae in the future. Thus, additional mitochondrial genome sampling and more molecular data are still required in order to effectively resolve the phylogeny and lineages of Syrphidae species.



**Figure 8.** Phylogenetic trees of Syrphidae based on the PCGRNA dataset using maximum likelihood (ML, left) and Bayesian inference (BI, right) methods. “Bootstrap support” (ML) and “posterior probabilities” (BI) are indicated at their nodes respectively. The species sequenced in this study were marked with \* after the species name. The image was mainly predicted by IQ-TREE (<http://iqtree.cibiv.univie.ac.at/>) and MrBayes in CIPRES webserver (<https://www.phylo.org/portal2/logininput.action>), drawing with Adobe Illustrator 2020.

### Data availability

The complete mitogenomes of *E. viridis*, *E. quinquestriatus*, *E. tarsalis*, and *E. sp.* were deposited in the GenBank database under accession numbers MN494096, MT471322, MW073114, and MT942687 respectively.

Received: 10 April 2021; Accepted: 3 March 2022

Published online: 09 March 2022

### References

- Ball, S. & Morris, R. *Britain's hoverflies: A field guide, second edition fully revised and updated*, Princeton, pp. 1–312 (2015).
- Huo, K. K., Ren, G. D. & Zheng, Z. M. *Fauna of Syrphidae from MT: Qing-Bashan in China (Insecta: Diptera)*, Beijing, pp. 1–512 (2007).
- Huang, C. M. & Cheng, Y. X. *Fauna Sinica. Insecta Vol. 50. Diptera. Syrphidae*, Beijing, pp. 1–871 (2012).
- Hull, F. M. LIX.—Studies upon Diptera in the British Museum. *Ann. Mag. Nat. Hist.* **2**, 727–746 (1949).
- Thompson, F. C. & Vockeroth, J. R. Family Syrphidae. In: Evenhuis, N. L. (edit): *Catalog of the Diptera of the Australasian and Oceanian Region*. Honolulu (1989).
- Mengual, X., Ståhls, G. & Rojo, S. Phylogenetic relationships and taxonomic ranking of pipizine flower flies (Diptera: Syrphidae) with implications for the evolution of aphidophagy. *Cladistics* **31**, 491–508 (2015).
- Young, A. D. *et al.* Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). *BMC Evol. Biol.* **16**, 143 (2016).
- Skevington, J. H. & Yeates, D. K. Phylogeny of the syrphoidea (Diptera) inferred from mtDNA sequences and morphology with particular reference to classification of the Pipunculidae (Diptera). *Mol. Phylogenet. Evol.* **16**, 212–224 (2000).
- Ståhls, G., Hippa, H., Rotheray, G., Muona, J. & Gilbert, F. Phylogeny of Syrphidae (Diptera) inferred from combined analysis of molecular and morphological characters. *Syst. Entomol.* **28**, 433–450 (2003).
- Cameron, S. L., Lambkin, C. L., Barker, S. C. & Whiting, M. F. A mitochondrial genome phylogeny of Diptera: Whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Syst. Entomol.* **32**, 40–59 (2007).
- Cameron, S. L. Insect mitochondrial genomics: Implications for evolution and phylogeny. *Annu. Rev. Entomol.* **59**, 95–117 (2014).
- Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
- Wang, Y., Chen, J., Jiang, L.-Y. & Qiao, G.-X. Hemipteran mitochondrial genomes: Features, structures and implications for phylogeny. *Int. J. Mol. Sci.* **16**, 12382–12404 (2015).
- Sonet, G. *et al.* First mitochondrial genomes of five hoverfly species of the genus *Eristalinus* (Diptera: Syrphidae). *Genome* **62**, 677–687 (2019).
- Pu, D. *et al.* Mitochondrial genomes of the hoverflies *Episyrphus balteatus* and *Eupeodes corollae* (Diptera: Syrphidae), with a phylogenetic analysis of Muscomorpha. *Sci. Rep.* **7**, 44300 (2017).

16. Li, X. *et al.* The complete mitochondrial genome analysis of *Eristalis tenax* (Diptera, Syrphidae). *Mitochondrial DNA Part B* **2**, 654–655 (2017).
17. Li, H. & Li, J. The complete mitochondrial genome of *Helophilus virgatus* (Diptera: Syrphidae: Eristalinae) with a phylogenetic analysis of Syrphidae. *Mitochondrial DNA Part B* **4**, 3106–3107 (2019).
18. Li, H. Characterization and Phylogenetic Implications of the Complete Mitochondrial Genome of Syrphidae. *Genes* **10**, 563 (2019).
19. Li, J. & Li, H. The first complete mitochondrial genome of genus *Phytomia* (Diptera: Syrphidae). *Mitochondrial DNA Part B* **5**, 2512–2513 (2020).
20. Chen, Q., Niu, X., Fang, Z. & Weng, Q. The complete mitochondrial genome of *Melanostoma orientale* (Diptera: Syrphidae). *Mitochondrial DNA Part B* **5**, 554–555 (2020).
21. Zhang, H.-J. & Huo, K.-K. Mitochondrial *Cty b* Gene Variability and Phylogenetic Relationships among 7 Species of Eristalini (Diptera: Syrphidae). *J. Hanzhong Teach. Coll.* **22**, 54–59 (2004).
22. Saeed, S., Sajjad, A., Kwon, O. & Kwon, Y. J. Fidelity of Hymenoptera and Diptera pollinators in onion (*Allium cepa* L.) pollination. *Entomol. Res.* **38**, 276–280 (2008).
23. Beckenbach, A. T. & Joy, J. B. Evolution of the Mitochondrial Genomes of Gall Midges (Diptera: Cecidomyiidae): Rearrangement and Severe Truncation of tRNA Genes. *Genome Biol. Evol.* **1**, 278–287 (2009).
24. Morales, M. N. & Köhler, A. Espécies de Syrphidae (Diptera) visitantes das flores de *Eryngium horridum* (Apiaceae) no Vale do Rio Pardo, RS, Brasil. *Iheringia. Série Zool.* **96**, 41–45 (2006).
25. Tsuji, K. & Fukami, T. Community-wide consequences of sexual dimorphism: evidence from nectar microbes in dioecious plants. *Ecology* **99**, 2476–2484 (2018).
26. Heo, C. C. *et al.* *Eristalinus arvorum* (Fabricius, 1787) (Diptera: Syrphidae) in human skull: A new fly species of forensic importance. *J. Forensic Sci.* **65**, 276–282 (2020).
27. Junqueira, A. C. M. *et al.* Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. *Sci. Rep.* **6**, 21762 (2016).
28. Yan, J. *et al.* Characterization and phylogenetic analysis of the complete mitochondrial genome of *Eristalia cerealis* (Diptera: Syrphidae). *Mitochondrial DNA Part B* **5**, 1005–1006 (2020).
29. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotech.* **3**, 294–299 (1994).
30. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
31. Laslett, D. & Canback, B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**, 172–175 (2008).
32. Schattner, P., Brooks, A. N. & Lowe, T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucl. Acids Res.* **33**, W686–W689 (2005).
33. Cameron, S. L. & Whiting, M. F. 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**, 112–123 (2008).
34. He, B. *et al.* Characterization of mitochondrial genomes of three *Andrena* bees (Apoidea: Andrenidae) and insights into the phylogenetics. *Int. J. Biol. Macromol.* **127**, 118–125 (2019).
35. Wang, Y., Cao, J.-J., Li, N., Ma, G.-Y. & Li, W.-H. The first mitochondrial genome from Scopuridae (Insecta: Plecoptera) reveals structural features and phylogenetic implications. *Int. J. Biol. Macromol.* **122**, 893–902 (2019).
36. Gillespie, J. J., Johnston, J. S., Cannone, J. J. & Gutell, R. R. Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): structure, organization, and retrotransposable elements. *Insect Mol. Biol.* **15**, 657–686 (2006).
37. Thompson, J. D., Gibson, T. J. & Higgins, D. G. Multiple Sequence Alignment Using ClustalW and ClustalX. *Curr. Protoc. Bioinforma.* **00**, 2.3.1–2.3.22 (2003).
38. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci.* **101**, 7287–7292 (2004).
39. Grant, J. R. & Stothard, P. The CGView Server: a comparative genomics tool for circular genomes. *Nucl. Acids Res.* **36**, W181–W184 (2008).
40. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
41. Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41**, 353–358 (1995).
42. Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497 (2003).
43. Li, X. *et al.* The First Mitochondrial Genome of the Sepsid Fly *Nemopoda mamaevi* Ozerov, 1997 (Diptera: Sciomyzoidea: Sepsidae), with Mitochondrial Genome Phylogeny of Cyclorrhapha. *PLoS One* **10**, e0123594 (2015).
44. Li, X. *et al.* Mitochondrial Genomes Provide Insights into the Phylogeny of Lauxanioidea (Diptera: Cyclorrhapha). *Int. J. Mol. Sci.* **18**, 773 (2017).
45. Abascal, F., Zardoya, R. & Telford, M. J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucl. Acids Res.* **38**, W7–W13 (2010).
46. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).
47. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
48. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701 (2012).
49. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, pp. 1–8 (2010).
50. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).
51. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
52. Minh, B. Q., Nguyen, M. A. T. & Haeseler, A. V. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
53. Beard, C. B., Hamm, D. M. & Collins, F. H. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.* **2**, 103–124 (1993).
54. Behura, S. K. *et al.* Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. *Insect Biochem. Mol. Biol.* **41**, 770–777 (2011).
55. Clary, D. O. & Wolstenholme, D. R. The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* **22**, 252–271 (1985).

56. Ojala, D., Montoya, J. & Attardi, G. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**, 470–474 (1981).
57. Hurst, L. D. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* **18**, 486–487 (2002).
58. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).
59. Zhang, N.-X., Zhang, Y.-J., Yu, G. & Chen, B. Structure characteristics of the mitochondrial genomes of Diptera and design and application of universal primers for their sequencing. *Acta Entomol. Sin.* **56**, 398–407 (2013).
60. Zhang, D.-X., Szymura, J. M. & Hewitt, G. M. Evolution and structural conservation of the control region of insect mitochondrial DNA. *J. Mol. Evol.* **40**, 382–391 (1995).
61. Rotheray, G. E. Descriptions and a key of the larval and puparial stages of north-west European Volucella (Diptera, Syrphidae). *Studia Dipt.* **6**, 103–116 (1999).
62. Sun, Z. *et al.* Mitochondrial genome of *Phalantus geniculatus* (Hemiptera: Reduviidae): trnT duplication and phylogenetic implications. *Int. J. Biol. Macromol.* **129**, 110–115 (2019).
63. Huang, W. & Zhang, Y. Characterization of two complete Mitochondrial Genomes of Ledrinae (Hemiptera: Cicadellidae) and Phylogenetic Analysis. *Insects* **11**, 609 (2020).

## Acknowledgements

This research was supported by a Scientific Research Program Funded by Shaanxi Provincial Education Department (No. 21JK0568), a "City-University Co-construction" Scientific Research Project for State Key Laboratory of Biological Resources and Ecological Environment of Qinling-Bashan (No. SXJ-2102), and an Opening Foundation of Shaanxi University of Technology (No. SLGPT2019KF03-01).

## Author contributions

H. L. and J. L. conceived and designed the experiments; J. L. analyzed the data; H. L. and J. L. drafted the manuscript; H. L. finalized the reviewing and editing. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-08172-6>.

**Correspondence** and requests for materials should be addressed to H.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2022