

# SCIENTIFIC REPORTS



OPEN

## Mitochondrial genomes of the hoverflies *Episyrphus balteatus* and *Eupeodes corollae* (Diptera: Syrphidae), with a phylogenetic analysis of Muscomorpha

Received: 08 September 2016

Accepted: 06 February 2017

Published: 09 March 2017

De-qiang Pu<sup>1</sup>, Hong-ling Liu<sup>2</sup>, Yi-yun Gong<sup>1</sup>, Pei-cheng Ji<sup>1</sup>, Yue-jian Li<sup>3</sup>, Fang-sheng Mou<sup>1</sup> & Shu-jun Wei<sup>4</sup>

The hoverflies *Episyrphus balteatus* and *Eupeodes corollae* (Diptera: Muscomorpha: Syrphidae) are important natural aphid predators. We obtained mitochondrial genome sequences from these two species using methods of PCR amplification and sequencing. The complete *Episyrphus* mitochondrial genome is 16,175 bp long while the incomplete one of *Eupeodes* is 15,326 bp long. All 37 typical mitochondrial genes are present in both species and arranged in ancestral positions and directions. The two mitochondrial genomes showed a biased A/T usage versus G/C. The *cox1*, *cox2*, *cox3*, *cob* and *nad1* showed relatively low level of nucleotide diversity among protein-coding genes, while the *trnM* was the most conserved one without any nucleotide variation in stem regions within Muscomorpha. Phylogenetic relationships among the major lineages of Muscomorpha were reconstructed using a complete set of mitochondrial genes. Bayesian and maximum likelihood analyses generated congruent topologies. Our results supported the monophyly of five species within the Syrphidae (Syrphoidea). The Platypezoidea was sister to all other species of Muscomorpha in our phylogeny. Our study demonstrated the power of the complete mitochondrial gene set for phylogenetic analysis in Muscomorpha.

The mitochondrion is a fundamental eukaryotic organelle. Energy production via oxidative phosphorylation is its most-studied function, though it also functions in apoptosis and cell aging<sup>1,2</sup>. A typical animal mitochondrial genome is a circular DNA molecule, approximately 16 kb long, with a relatively conserved gene content, usually containing 37 genes: 13 protein-coding genes (PCGs), two ribosomal RNA (rRNA) genes, and 22 transfer RNA (tRNA) genes, plus an A + T-rich region<sup>3,4</sup>. Compared with nuclear genomes, animal mitochondrial genomes are characterized by several distinct features, including cellular abundance, small genome size, conserved gene content and organization, lack of extensive recombination, maternal inheritance, and a high nucleotide substitution rate<sup>2,5,6</sup>. In particular, mitochondrial genomes can provide higher phylogenetic resolution than the shorter sequences of individual genes. Consequently, this small molecule has been widely used for phylogenetic analyses in many groups<sup>2,7–12</sup>.

Muscomorpha (Diptera: Brachycera), an infraorder of Brachycera, is a large and diverse group of flies, containing the bulk of the Brachycera. It includes a number of the most common flies, such as the housefly, the fruit fly, and the blowfly. The Muscomorpha can be separated into two sections, the Aschiza and the Schizophora<sup>9,13–15</sup>. Two large superfamilies, Platypezoidea and Syrphoidea were traditionally included in Aschiza. Phoridae and Syrphidae are the two largest families within these superfamilies, respectively. The Schizophora contain the majority of family level diversity among dipterans, and represent a recent rapid radiation of lineages<sup>14,16</sup>.

<sup>1</sup>Industrial Crop Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu 610300, China. <sup>2</sup>Institute of Plant Protection, Sichuan Academy of Agricultural Sciences, Chengdu 610066, China. <sup>3</sup>Institute of Horticulture Research, Sichuan Academy of Agricultural Sciences, Chengdu 610066, China. <sup>4</sup>Institute of Plant and Environmental Protection, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China. Correspondence and requests for materials should be addressed to F.-S.M. (email: fshmu@163.com) or S.-J.W. (email: shujun268@163.com)

	Sections	Species	Superfamily	Family	Accession number	References
Ingroup	Aschiza	<i>Episyrphus balteatus</i>	Syrphoidea	Syrphidae	KU351241	This paper
		<i>Eupeodes corollae</i>	Syrphoidea	Syrphidae	KU379658	This paper
		<i>Ocyrtamus sativus</i>	Syrphoidea	Syrphidae	KT272862	Junqueira, <i>et al.</i> <sup>16</sup>
		<i>Simosyrphus grandicornis</i>	Syrphoidea	Syrphidae	DQ866050	Cameron, <i>et al.</i> <sup>9</sup>
		<i>Syrphidae</i> sp.	Syrphoidea	Syrphidae	KM244713	Tang, <i>et al.</i> <sup>28</sup>
	<i>Megaselia scalaris</i>	Platypoidea	Phoridae	KF974742	Zhong, <i>et al.</i> <sup>51</sup>	
	Schizophora	<i>Bactrocera dorsalis</i>	Tephritoidea	Tephritidae	DQ845759	Zhong, <i>et al.</i> <sup>51</sup>
		<i>Drosophila melanogaster</i>	Ephydroidea	Drosophilidae	JQ686693	Chen, <i>et al.</i> <sup>52</sup>
		<i>Fergusonina taylori</i>	Opomyzoidea	Fergusoninidae	HQ872008	Nelson, <i>et al.</i> <sup>53</sup>
		<i>Nemopoda mamaevi</i>	Sciomyzoidea	Sepsidae	KM605250	Li, <i>et al.</i> <sup>11</sup>
		<i>Haematobia irritans</i>	Muscoidea	Muscidae	DQ029097	Unpublished
		<i>Musca domestica</i>	Muscoidea	Muscidae	KT444442	Unpublished
		<i>Chrysomya putoria</i>	Oestroidea	Calliphoridae	AF352790	Junqueira, <i>et al.</i> <sup>54</sup>
		<i>Dermatobia hominis</i>	Oestroidea	Oestridae	AY463155	Unpublished
		<i>Pollenia rudis</i>	Oestroidea	Calliphoridae	JX913761	Nelson, <i>et al.</i> <sup>55</sup>
<i>Ravinia pernix</i>	Oestroidea	Sarcophagidae	KM676414	Unpublished		
Outgroup	<i>Cydistomyia duplonotata</i>	Tabanoidea	Tabanidae	DQ866052	Cameron, <i>et al.</i> <sup>9</sup>	
	<i>Trichophthalma punctata</i>	Nemestrinoidea	Nemestrinidae	DQ866051	Cameron, <i>et al.</i> <sup>9</sup>	

**Table 1. Mitochondrial genomes used in this study.**

Schizophora can be divided into two subsections, the Acalypratae and Calypratae, commonly referred to as acalyprate muscoids and calyprate muscoids, respectively<sup>15</sup>.

Hoverflies, also called flower flies, compose the insect family Syrphidae. This family contains almost 6,000 described species in 200 genera, and is nearly worldwide in distribution<sup>17,18</sup>. As the common names suggest, these flies are often seen hovering or nectaring at flowers. In contrast to the fairly uniform flower-feeding habits of adult syrphids, the larvae eat a wide range of foods. Traditionally, the Syrphidae has been divided into three subfamilies, the Eristalinae, Microdontinae, and Syrphinae. Larvae of the subfamily Eristalinae are saprophagous in dead wood, eating decaying plant and animal matter in soil, ponds, or streams<sup>19,20</sup>; whereas Microdontinae larvae are inquilines in ants' nests<sup>21</sup>; and Syrphinae larvae are insectivores, preying on aphids, thrips, and other plant-sucking insects<sup>17,22,23</sup>. This three subfamily system (Microdontinae, Eristalinae, and Syrphinae) was adopted for Syrphidae more than 25 years ago<sup>18,23</sup>.

The marmalade hoverfly *Episyrphus balteatus* (Syrphidae) is a relatively small hoverfly (9–12 mm), widespread throughout the Palearctic ecozone, endemic to Europe, North Asia, and North Africa. Its color patterns may appear wasp-like to animals, such as birds, protecting it from predation<sup>24</sup>. Often exhibiting dense migratory swarm behavior, this, and the resemblance to wasps, may panic unaware people<sup>25</sup>. Its feeding habit is rare among adult flies, as it is capable of crushing pollen grains as a food source<sup>26</sup>. *Eupeodes corollae* (Syrphidae) is another common hoverfly species. The adults are often migratory with a worldwide distribution. It has been tested as a biological control agent for aphids and scale insects in greenhouses. However, the larval flies preferred the fruit in the experiment, consuming more fruit than aphids<sup>27</sup>.

As of June 2016, three Syrphidae mitochondrial genome sequences were deposited into GenBank: *Ocyrtamus sativus*, *Simosyrphus grandicornis* and an unknown genus sp. Syrphidae (accession KM244713)<sup>9,16,28</sup>. Here, we sequenced the mitochondrial genomes from *Episyrphus* and *Eupeodes*, representing the first mitochondrial genomes reported from the genera *Episyrphus* and *Eupeodes*. Furthermore, we compared genome features and investigated phylogenetic relationships within Muscomorpha using complete mitochondrial genomes from GenBank along with our two newly sequenced mitochondrial genomes (Table 1).

## Results and Discussion

**General features of the newly sequenced mitochondrial genomes.** The complete *Episyrphus* mitochondrial genome sequence is 16,175 bp long (GenBank accession KU351241) (Table 2). The partial mitochondrial genome of *Eupeodes* mitochondrial genome sequence is 15,326 bp long (GenBank accession KU379658) (Table 3). Particularly A + T-rich region was failed to generate reliable sequence data in both species.

No gene rearrangement was observed in our analyses: (1) as compared with the putative ancestral insect arrangement<sup>29</sup>, (2) as in all sequenced dipteran species<sup>11</sup>, (3) as with the 23 genes encoded on the majority strand (J-strand), and (4) as with the 14 genes encoded on the minority strand (N-strand).

Each of the 37 typical mitochondrial genes is present in both species. The mitochondrial genome of *Episyrphus* has 255 bp of intergenic nucleotides, in 22 different locations, with intergenic spacer lengths ranging from 1 to 60 bp. Seven pairs of genes overlap each other, with overlap lengths ranging from 1 to 7 bp. Eight pairs of genes directly adjacent one another including the pairs of *rrnL-trnV* and *trnV-rrnS*. The mitochondrial genome of *Eupeodes* has 230 bp of intergenic nucleotides, in 19 locations, with intergenic spacer lengths from 2 to 47 bp. Nine pairs of genes overlap each other, with overlap lengths ranging from 1 to 7 bp. Nine pairs of genes directly adjacent one another including the pairs of *rrnL-trnV* and *trnV-rrnS*. In both species, the longest intergenic spacer was located between *trnK* and *trnD*, followed by the one located between *trnE* and *trnF*. The longest overlapping

Gene	Position (bp)	Strand	Length (bp)	Intergenic nucleotides	Anti or start/stop codon
<i>trnI</i>	1–66	+	66	2	GAT
<i>trnQ</i>	69–137	–	69	14	TTG
<i>trnM</i>	152–220	+	69	0	CAT
<i>nad2</i>	221–1252	+	1032	3	ATT/TAA
<i>trnW</i>	1256–1323	+	68	8	TCA
<i>trnC</i>	1332–1398	–	67	14	GCA
<i>trnY</i>	1413–1478	–	66	5	GTA
<i>cox1</i>	1484–3043	+	1560	–5	ATT/TAA
<i>trnL2</i>	3039–3104	+	66	2	TAA
<i>cox2</i>	3107–3790	+	684	5	ATG/TAA
<i>trnK</i>	3796–3866	+	71	60	CTT
<i>trnD</i>	3927–3993	+	67	0	GTC
<i>atp8</i>	3994–4155	+	162	–7	ATT/TAA
<i>atp6</i>	4149–4826	+	678	35	ATG/TAA
<i>cox3</i>	4862–5650	+	789	3	ATG/TAA
<i>trnG</i>	5654–5719	+	66	0	TCC
<i>nad3</i>	5720–6073	+	354	4	ATT/TAA
<i>trnA</i>	6078–6146	+	69	–1	TGC
<i>trnR</i>	6146–6209	+	64	2	TCG
<i>trnN</i>	6212–6278	+	67	–1	GTT
<i>trnS1</i>	6278–6344	+	67	1	GCT
<i>trnE</i>	6346–6410	+	65	23	TTC
<i>trnF</i>	6434–6500	–	67	–1	GAA
<i>nad5</i>	6500–8221	–	1722	15	ATT/TAA
<i>trnH</i>	8237–8303	–	67	–1	GTG
<i>nad4</i>	8303–9643	–	1341	–7	ATG/TAA
<i>nad4l</i>	9637–9933	–	297	2	ATG/TAA
<i>trnT</i>	9936–10000	+	65	0	TGT
<i>trnP</i>	10001–10066	–	66	2	TGG
<i>nad6</i>	10069–10593	+	525	3	ATT/TAA
<i>cob</i>	10597–11733	+	1137	1	ATG/TAA
<i>trnS2</i>	11735–11802	+	68	16	TGA
<i>nad1</i>	11819–12757	–	939	10	ATA/TAA
<i>trnL1</i>	12768–12832	–	65	0	TAG
<i>rrnL</i>	12833–14170	–	1338	0	
<i>trnV</i>	14171–14242	–	72	0	TAC
<i>rrnS</i>	14243–15046	–	804	0	
A+T-rich region	15047–16175	–	1129	0	

**Table 2.** Annotation of the *Episyrphus balteatus* mitochondrial genome. + majority strand; – minority strand; negative intergenic nucleotides indicate overlapping sequences between adjacent genes.

regions were located between *atp8* and *atp6*, *nad4* and *nad4l*. The intergenic and overlapping region of these two species are similar to most other insect mitochondrial genomes with no gene rearrangement occurred, compared with those with frequent gene rearrangement<sup>30</sup>.

**Base composition.** Three parameters, the AT and GC asymmetries, known as AT-skew and GC-skew, and A + T content, are often used to characterize the nucleotide-compositional behavior of mitochondrial genomes<sup>31,32</sup>. The *Episyrphus* and *Eupeodes* mitochondrial genomes show a very strong bias in nucleotide composition (A + T% > G + C%), which is typical of insect mitochondrial genomes. The 16 Muscomorpha species we analyzed have PCG A + T contents between 71.10% (*Bactrocera dorsalis*) and 78.97% (*Ocyptamus sativus*); The mitochondrial genomes of *Episyrphus* and *Eupeodes* have A + T contents of 78.85% and 78.83%, respectively (Table 4).

The PCG sequence AT-skews of the 16 Muscomorpha species are primarily negative (except in *Episyrphus*, *O. sativus*, *Simosyrphus grandicornis*, and *Haematobia irritans*). All GC-skews are positive, indicating that the PCGs contain a higher percentage of T and C than A and G nucleotides (Table 4), as reported with most other insects<sup>32</sup>.

Gene	Position (bp)	Strand	Length (bp)	Intergenic nucleotides	Anti or start/stop codon
<i>trnI</i>	1–66	+	66	–3	GAT
<i>trnQ</i>	64–133	–	70	7	TTG
<i>trnM</i>	141–209	+	69	0	CAT
<i>nad2</i>	210–1238	+	1029	–2	ATT/TAA
<i>trnW</i>	1237–1304	+	68	4	TCA
<i>trnC</i>	1309–1374	–	66	10	GCA
<i>trnY</i>	1385–1450	–	66	22	GTA
<i>cox1</i>	1473–3029	+	1557	–5	ATT/TAA
<i>trnL2</i>	3025–3090	+	66	2	TAA
<i>cox2</i>	3093–3776	+	684	0	ATG/TAA
<i>trnK</i>	3777–3847	+	71	47	CTT
<i>trnD</i>	3895–3961	+	67	0	GTC
<i>atp8</i>	3962–4123	+	162	–7	ATT/TAA
<i>atp6</i>	4117–4794	+	678	16	ATG/TAA
<i>cox3</i>	4811–5599	+	789	3	ATG/TAA
<i>trnG</i>	5603–5669	+	67	0	TCC
<i>nad3</i>	5670–6023	+	354	3	ATT/TAA
<i>trnA</i>	6027–6095	+	69	–1	TGC
<i>trnR</i>	6095–6158	+	64	13	TCG
<i>trnN</i>	6172–6238	+	67	–1	GTT
<i>trnS1</i>	6238–6304	+	67	5	GCT
<i>trnE</i>	6310–6376	+	67	26	TTC
<i>trnF</i>	6403–6469	–	67	–1	GAA
<i>nad5</i>	6469–8189	–	1721	15	ATT/TA
<i>trnH</i>	8205–8270	–	66	–1	GTG
<i>nad4</i>	8270–9610	–	1341	–7	ATG/TAA
<i>nad4l</i>	9604–9900	–	297	2	ATG/TAA
<i>trnT</i>	9903–9967	+	65	0	TGT
<i>trnP</i>	9968–10033	–	66	2	TGG
<i>nad6</i>	10036–10560	+	525	3	ATT/TAA
<i>cob</i>	10564–11700	+	1137	7	ATG/TAA
<i>trnS2</i>	11708–11775	+	68	17	TGA
<i>nad1</i>	11792–12730	–	939	10	ATA/TAG
<i>trnL1</i>	12741–12805	–	65	0	TAG
<i>rrnL</i>	12806–14139	–	1334	0	
<i>trnV</i>	14140–14211	–	72	0	TAC
<i>rrnS</i>	14212–15006	–	795	0	
A+T-rich region	15007–15326	–			

**Table 3.** Annotation of the *Eupeodes corollae* mitochondrial genome. Symbols are as in Table 2.

**Protein-coding genes, codon usage and nucleotide diversity.** Nine of the 13 mitochondrial PCGs in *Episyrphus* and *Eupeodes* mitochondrial genomes are located on the J-strand; the other four PCGs are located on the N-strand (Tables 1 and 2). Total PCG length in *Episyrphus* is 11,220 bp, while *Eupeodes* has 11,211 bp of PCG.

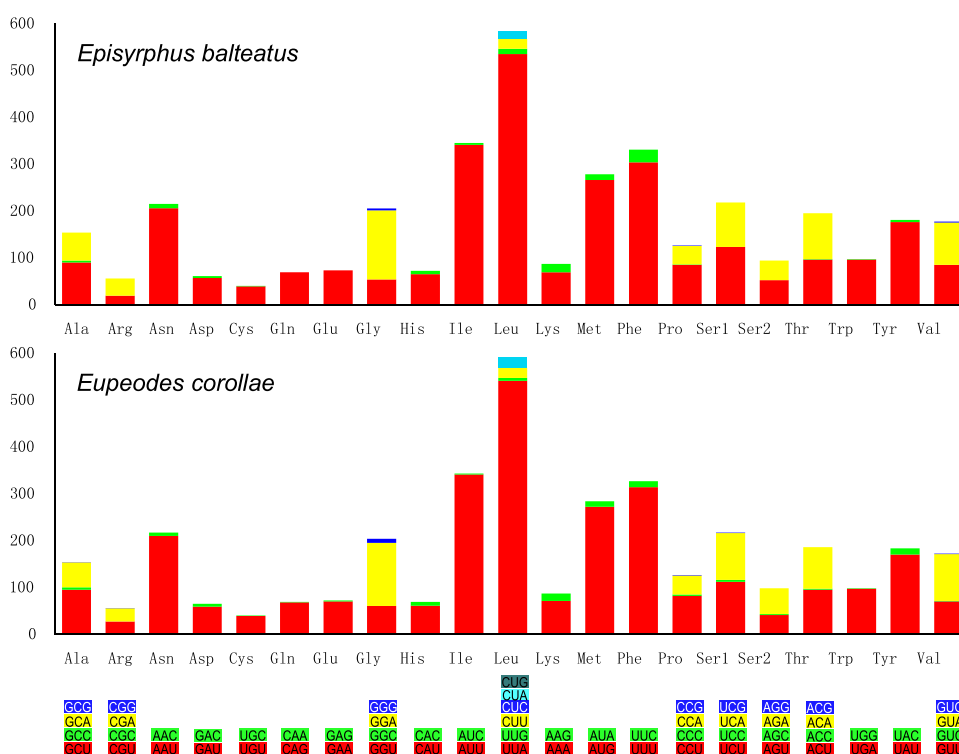
All *Episyrphus* and *Eupeodes* mitochondrial genome PCGs start with ATN codons. One, six, and six of the PCGs start with ATA, ATG, and ATT, respectively. Orthologs from the two species have the same start codons. Most PCG stop codons are the canonical TAA, except for *nad5* in *Eupeodes*, which uses an incomplete TA.

Mitochondrial genome codon usage in *Episyrphus* and *Eupeodes* show a significant bias towards A and T (Fig. 1) as in other species of Muscomorpha (Figure S1). In the *Episyrphus* and *Eupeodes* mitochondrial genomes, Leu, Ile, Phe, and Met are the most frequently encoded amino acids, hence TTA (Leu), ATT (Ile), TTT (Phe), and ATA (Met) are the most frequent codons, as is typical of other insect mitochondrial genomes<sup>30,33,34</sup>. These frequently used codons exclusively consist of A and T, which contribute to the high A + T content seen in most fly mitochondrial genomes (Figure S1). This preferred codon usage is strongly reflected at third positions by high A/T versus G/C frequencies.

Evolutionary rate of protein-coding genes was calculated by using the nucleotide diversity and Jukes and Cantor corrected nucleotide diversity within Muscomorpha. Among the 13 protein-coding genes, five genes of *cox1*, *cox2*, *cox3*, *cob* and *nad1* showed relatively low level, five genes of *atp6*, *nad6*, *nad4*, *nad4l* and *nad5* showed

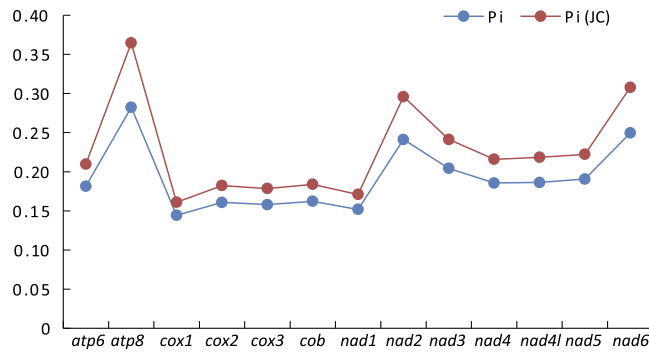
Species	Accession number	PCGs								<i>rrnl</i>		<i>rrns</i>	
		T	C	A	G	Length (bp)	AT (%)	AT-skew	GC-skew	Length (bp)	AT (%)	Length (bp)	AT (%)
<i>Episyrphus balteatus</i>	KU351241	39.90	11.66	38.95	9.49	11220	78.85	-0.0121	-0.1024	1338	84.60	804	83.96
<i>Eupeodes corollae</i>	KU379658	39.25	11.66	39.59	9.51	11211	78.83	0.0043	-0.1016	1334	84.78	795	83.14
<i>Ocyptamus sativus</i>	KT272862	39.71	11.68	39.26	9.35	11175	78.97	-0.0058	-0.1106	1314	84.40	778	82.78
<i>Simosyrphus grandicornis</i>	DQ866050	39.75	11.81	39.09	9.35	11208	78.84	-0.0084	-0.1164	1339	84.99	804	83.83
<i>Megaselia scalaris</i>	KF974742	37.26	14.81	37.66	10.26	11301	74.92	0.0053	-0.1814	1318	81.03	786	79.64
<i>Bactrocera dorsalis</i>	DQ845759	32.94	17.85	38.16	11.05	11175	71.10	0.0734	-0.2353	1333	79.52	790	74.94
<i>Drosophila melanogaster</i>	JQ686693	38.16	12.78	39.06	10.00	11169	77.22	0.0117	-0.1219	1323	82.84	786	80.15
<i>Fergusonina taylori</i>	HQ872008	36.13	14.19	39.92	9.76	11160	76.05	0.0498	-0.1852	1306	82.54	780	82.05
<i>Nemopoda mamaevi</i>	KM605250	35.84	16.28	36.88	11.00	11181	72.72	0.0144	-0.1934	1321	80.62	783	77.65
<i>Haematobia irritans</i>	DQ029097	39.04	12.29	38.63	10.04	11178	77.67	-0.0053	-0.1010	1321	82.66	784	78.83
<i>Musca domestica</i>	KT444442	37.84	13.61	38.24	10.31	11184	76.08	0.0053	-0.1379	1326	81.00	783	78.29
<i>Chrysomya putoria</i>	AF352790	36.74	14.59	38.23	10.44	11187	74.97	0.0199	-0.1657	1329	81.79	785	76.94
<i>Dermatobia hominis</i>	AY463155	35.96	15.08	39.31	9.65	11187	75.27	0.0445	-0.2198	1324	83.08	788	78.68
<i>Pollenia rudis</i>	JX913761	37.29	14.14	38.52	10.04	11172	75.81	0.0163	-0.1695	1328	81.63	448	74.11
<i>Ravinia pernix</i>	KM676414	36.88	14.33	38.63	10.17	11181	75.50	0.0232	-0.1698	1328	82.08	786	77.48

**Table 4.** Nucleotide composition in regions of Muscomorpha mitochondrial genomes.



**Figure 1.** Relative synonymous codon usage (RSCU) of mitochondrial genomes in the two newly sequenced mitochondrial genomes of *Episyrphus balteatus* and *Eupeodes corollae*. The stop codon is not given.

medial level, whereas three genes of *atp8*, *nad2* and *nad6* showed the highest level of nucleotide diversity (Fig. 2). Relative evolutionary rate among the 13 protein-coding genes in Muscomorpha was similar to previous studies of insect mitochondrial genomes<sup>35,36</sup>.



**Figure 2. Nucleotide diversity of each protein-coding gene among Muscomorpha.** Pi, nucleotide diversity; Pi(JC), Jukes and Cantor-corrected nucleotide diversity. Species used for calculation were listed in Table 1.

**Transfer RNA and ribosomal RNA genes.** All tRNA anticodons from the two newly sequenced mitochondrial genomes are identical to other Muscomorpha species (Tables 1 and 2). Of the 22 tRNA genes total, 14 are located on the J-strand, with the rest on the N-strand.

The *Episyrphus* mitochondrial genome contains 1,477 bp within tRNA genes, at an A + T content of 80.37%. Individual tRNAs range in length from 64 bp (*trnR*) to 72 bp (*trnV*). The *Eupeodes* mitochondrial genome contains 1,479 bp within tRNA genes, at an A + T content of 80.12%. Individual tRNAs range in length from 64 bp (*trnR*) to 72 bp (*trnV*).

Secondary structure models of the tRNA genes in the two newly sequenced mitochondrial genomes were predicted using the Mitos WebServer<sup>37</sup> (Fig. 3). In *Episyrphus* and *Eupeodes*, all tRNA genes fold into the canonical clover-leaf structure. The dihydrouridine (DHU) arm of all the tRNAs is a large loop, instead of a conserved stem-and-loop structure; however, this is typical of metazoan mitochondrial genomes<sup>38</sup>. The amino acid acceptor (AA) stem and the anticodon (AC) loop are conserved at 7 bp in all of our tRNA genes. The size of the variable- and D-loop often determine overall tRNA length<sup>39</sup>. The DHU arms in our tRNAs are 2 to 4 bp long, the AC arms are 4 to 5 bp long, and the TΨC arms vary in length from 3 to 5 bp. The variable loops are less consistent, ranging from 4 to 8 bp.

We also compared the variation of stem regions of tRNA genes among 15 species of Muscomorpha. Among the 22 tRNA genes, *trnM* was the most conserved one without any nucleotide variation in stem regions, followed by *trnV* and *trnE* with three site mutations. The *trnC* showed the highest number of site mutation on stem regions (17 sites), followed by the *trnH* (16 sites) (Fig. 2).

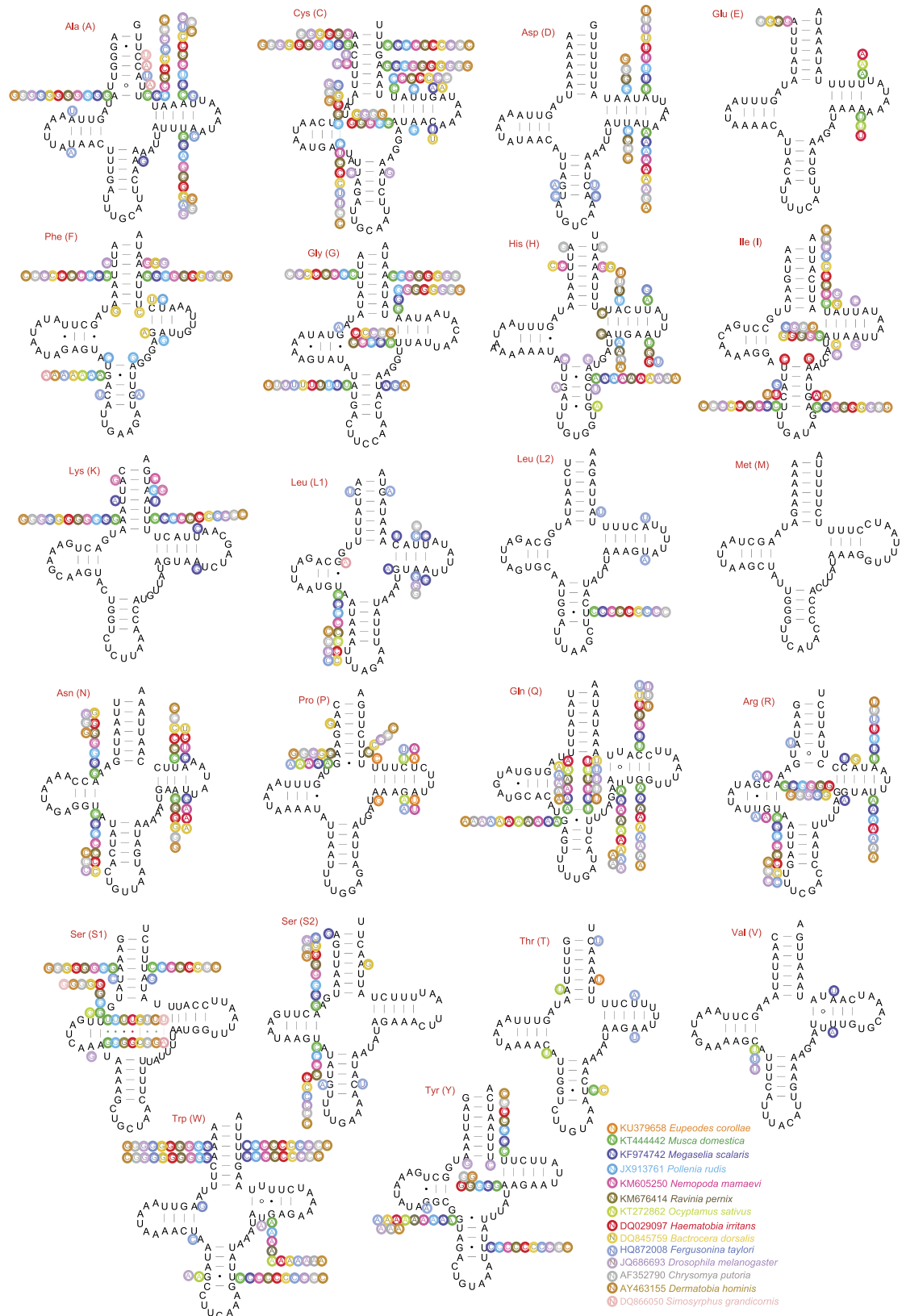
Base pairs other than canonical A-Us and C-Gs are occasionally used in our tRNAs, based on predicted tRNA secondary structures. We found six and five mismatched base pairs in the tRNAs from *Episyrphus* and *Eupeodes*, respectively. Among the six mismatched base pairs in *Episyrphus*, five are U-U pairs, located in the AA and TΨC stems; the other is an A-A pair, located in the A-A stem. *Eupeodes* has four U-U pairs, located in the AA and TΨC stems, and an A-A pair, located in the A-A stem.

The two ribosomal RNA genes in the mitochondrial genome, *rrnL* and *rrnS*, are 1,338 bp long, with an A + T content of 84.61%; and 804 bp long, with an A + T content of 83.96%, respectively, in *Episyrphus*. In *Eupeodes*, *rrnL* is 1334 bp long, with an A + T content of 84.78%; *rrnS* is 795 bp long, with an A + T content of 83.14% (Table 4).

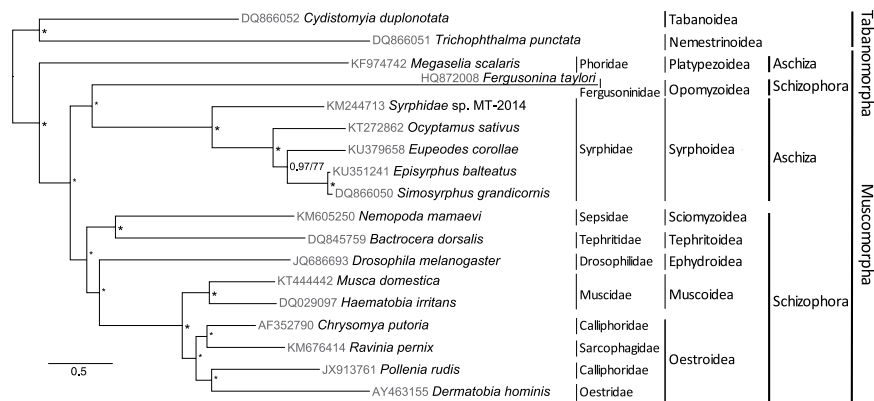
**Phylogenetic relationships.** We reconstructed phylogenies within the Muscomorpha using the nucleotide sequences of the 37 mitochondrial genes. Bayesian and maximum likelihood (ML) methods estimated congruent topologies (Fig. 4). Our analyses supported the monophyly of all superfamilies used in the study. The Aschiza (lower Cyclorrhapha) was found to be a paraphilic group<sup>40</sup>. We included two superfamilies of Platypezoidea and Syrphoidea from Aschiza. Platypezoidea was sister to all other species of Muscomorpha, which is congruent with previous studies<sup>11,41</sup>. The five genera of Syrphidae (Syrphoidea) clustered as ((unknown Syrphidae sp.) + (*Ocyptamus* + (*Eupeodes* + (*Episyrphus* + *Simosyrphus*))))). Syrphoidea and Opomyzoidea formed a lineage, and then sister to the other species of Schizophora. The Opomyzoidea was traditionally considered as a superfamily of Schizophora, which was proved to be a monophyletic group<sup>40</sup>. We study showed that Schizophora was interrupted by Opomyzoidea, which might be caused by the long-branch of Opomyzoidea. In Opomyzoidea, we used one species from family Fergusoninidae, in which, all species are gall-forming flies together with *Fergusobia* (Tylenchida: Neotylenchidae) nematodes. The novel life history of the species from Fergusoninidae might affect the evolutionary pattern of their mitochondrial genomes<sup>2</sup>. The long-branch of Opomyzoidea was also found in previous study based on mitochondrial genome sequences<sup>41</sup>. Phylogenetic relationships of other groups included in our analyses, i.e. ((Sciomyzoidea + Tephritoidea) + (Ephydroidea + (Muscoidea + Oestroidea))), were in accord with previous studies<sup>9,13–16</sup>.

## Methods

**Sampling and DNA extraction.** The specimens were collected from Sichuan Province, China. Specimens were initially preserved in 100% ethanol in the field when collected, and then stored at  $-80^{\circ}\text{C}$  prior to DNA extraction. Whole genomic DNA was extracted from the legs and thorax of the specimens using a DNeasy tissue kit (Qiagen, Hilden, Germany), following the manufacturer's protocols.



**Figure 3. Comparison on the secondary structure of the tRNA genes in Muscomorpha mitochondrial genomes.** The secondary structures were drawn from tRNA genes of *Episyrrhus balteatus*. Variations of each site in other 14 species of Muscomorpha were indicated near corresponding nucleotide. Each species was marked by a unique color as shown on the right bottom of the figure.



**Figure 4. Phylogeny of Muscomorpha inferred from coding nucleotide sequences of the mitochondrial genome (13 PCGs, 22 tRNAs, and 2 rRNA genes), using the Bayesian and maximum likelihood methods.** Numbers separated by “/” indicate the posterior probability and bootstrap values of the corresponding nodes (BI / ML). “\*” indicate that the node was fully supported by both inferences (1/100).

**PCR amplification and sequencing.** Initially we used a previously designed set of universal primers for insect mitochondrial genomes<sup>1,42</sup> to amplify and sequence partial gene segments. Then we designed specific primers based on the sequenced segments to amplify regions that bridged the gaps between different segments (Table S1). PCR cycling consisted of an initial denaturation step at 96 °C for 3 min, followed by 40 cycles of denaturation at 95 °C for 30 s, annealing at 42–53 °C for 30 s, elongation at 60 °C for 1.5 kb/min (depending on the size of target amplicon), and a final elongation step at 60 °C for 10 min. PCR products were evaluated by agarose gel electrophoreses. PCR components were added following the Takara LA Taq protocols. A primer-walking strategy was used for all the amplifications from both strands (Table S1).

**Mitochondrial genome annotation.** Mitochondrion DNA sequences were assembled using Lasergene software (DNASTar, Inc., USA, NewYork). The tRNA genes were initially identified using the Mitos WebServer (<http://mitos.bioinf.uni-leipzig.de/index.py>)<sup>37</sup>. We set the genetic code to “Invertebrate Mito”. Those tRNAs that could not be found using this approach were confirmed by sequence alignment with their homologs from related species. Secondly, protein-coding genes were identified by BLAST searches in GenBank, using other published mitochondrial genomes from Syrphidae<sup>9,16,28</sup>. Finally, the rRNA genes and control regions were identified by the boundary of the tRNA genes, and by comparison with other insect mitochondrial genomes.

**Comparative analysis of the mitochondrial genomes from Symphyta.** We compared the mitochondrial genomes of 16 species from the Muscomorpha, including our two newly sequenced genomes. Gene arrangement, base composition, and PCG codon usage features were analyzed. Because several tRNA genes were not available for some species, we analyzed base composition using only the PCGs. Furthermore, the unknown *Syrphidae* sp. sequence lacked *nad2* data; therefore, we excluded this species from these analyses.

We calculated base composition using MEGA6<sup>43</sup>. The AT-skew and GC-skew were calculated according to Hassanin, *et al.*<sup>31</sup>:  $AT-skew = (A\% - T\%)/(A\% + T\%)$  and  $GC-skew = (G\% - C\%)/(G\% + C\%)$ . The intergenic spacers and overlapping regions between genes were counted manually. The relative synonymous codon usage (RSCU) of all protein-coding genes was calculated using CodonW (written by John Peden, University of Nottingham, UK). Nucleotide diversity and Jukes and Cantor-corrected nucleotide diversity were calculated for species of Muscomorpha using DnaSP v5<sup>44</sup>.

**Phylogenetic analysis.** We used 14 Muscomorpha species with published mitochondrial genomes, and our two newly sequenced mitochondrial genomes for phylogenetic analyses (Table 1). The 16 species are classified as belonging to two sections, Aschiza and Schizophora. We selected Aschiza sequences belonging to two superfamilies, Platypezoidea and Syrphoidea. Schizophora is classified as two subsections. We selected sequences from both subsections, and selected sequences from six superfamilies within them. *Cydistomyia duplonotata* and *Trichophtalma punctata* (Tabanomorpha: Tabanoidea: Tabanidae) were used as outgroups because of the close relationship between Tabanomorpha and Muscomorpha<sup>11</sup>.

MAFFT version 7.205, which implements consistency-based algorithms, was used for the alignment of protein-coding and RNA genes<sup>45</sup>. We used the G-INS-I and Q-INS-I algorithms in MAFFT<sup>46</sup> for protein-coding and RNA alignment, respectively. The alignment of the nucleotide sequences was guided by the amino acid sequence alignment using the Perl script TranslatorX version 1.1<sup>47</sup>.

Data partitioning, and the ability to apply specific models to different partitions, is ideal for analyzing mitochondrial genomes<sup>2</sup>. We used PartitionFinder version 1.1.1<sup>48</sup> to simultaneously confirm partition schemes and choose substitution models for the matrix. The DNA sequence search model was set to “mrbayes”. The greedy algorithm was used, with estimated, linked branch lengths, to search for the best-fit partitioning model.

We constructed phylogenies among the Muscomorpha with the Bayesian inference method (BI) using MrBayes version 3.2.5<sup>49</sup>, and the ML method using RAxML version 8.0.0<sup>50</sup>. In BI, the GTR + I + G, GTR + G,



HKY + I + G, and HKY + G models were used with corresponding partitions (Table S2). Four simultaneous Markov chains were run for 10 million generations, with tree sampling occurring every 1,000 generations, and a burn-in of 25% of the trees. We used the GTR + G model for each ML analysis. We conducted 200 ML runs to find the highest-likelihood tree, then analyzed 1,000 bootstrap replicates.

## References

- Cameron, S. L. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Syst. Entomol.* **39**, 400–411 (2014).
- Cameron, S. L. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu. Rev. Entomol.* **59**, 95–117 (2014).
- Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
- Simon, C. *et al.* Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**, 651–701 (1994).
- Curole, J. P. & Kocher, T. D. Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends. Ecol. Evol.* **14**, 394–398 (1999).
- Lin, C. P. & Danforth, B. N. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Mol. Phylog. Evol.* **30**, 686–702 (2004).
- Beckenbach, A. T. Mitochondrial genome sequences of Nematocera (lower Diptera): evidence of rearrangement following a complete genome duplication in a winter crane fly. *Genome Biol. Evol.* **4**, 89–101 (2012).
- 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).
- 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).
- Demari-Silva, B. *et al.* Mitochondrial genomes and comparative analyses of *Culex camposi*, *Culex coronator*, *Culex usquatus* and *Culex usquatissimus* (Diptera: Culicidae), members of the coronator group. *BMC Genomics* **16**, 831 (2015).
- 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).
- Zhang, B., Nardi, F., Hull-Sanders, H., Wan, X. & Liu, Y. The complete nucleotide sequence of the mitochondrial genome of *Bactrocera minax* (Diptera: Tephritidae). *PLoS ONE* **9**, e100558 (2014).
- Cook, C. E., Austin, J. J., Henry, R. & Disney, L. A mitochondrial 12S and 16S rRNA phylogeny of critical genera of Phoridae (Diptera) and related families of Aschiza. *Zootaxa* **593**, 1–11 (2004).
- Yeates, D. K. & Wiegmann, B. M. Congruence and controversy: toward a higher-level phylogeny of Diptera. *Annu. Rev. Entomol.* **44**, 397–428 (1999).
- Yeates, D. K. *et al.* Phylogeny and systematics of Diptera: Two decades of progress and prospects. *Zootaxa* **1668**, 565–590 (2007).
- Junqueira, A. C. *et al.* Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. *Sci. Rep.* **6** (2016).
- Mengual, X., Stahls, G. & Rojo, S. Molecular phylogeny of *Allograpta* (Diptera, Syrphidae) reveals diversity of lineages and non-monophyly of phytophagous taxa. *Mol. Phylog. Evol.* **49**, 715–727 (2008).
- Mengual, X., Stahls, 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).
- Grosskopf, G. Biology and life history of *Cheilosia urbana* (Meigen) and *Cheilosia psilophthalma* (Becker), two sympatric hoverflies approved for the biological control of hawkweeds (*Hieracium* spp.) in New Zealand. *Biol. Control* **35**, 142–154 (2005).
- Fischer, O. A. *et al.* Various stages in the life cycle of syrphid flies (*Eristalis tenax*; diptera: Syrphidae) as potential mechanical vectors of pathogens causing mycobacterial infections in pig herds. *Folia Microbiol.* **51**, 147–153 (2006).
- Reemer, M. Review and Phylogenetic Evaluation of Associations between Microdontinae (Diptera: Syrphidae) and Ants (Hymenoptera: Formicidae). *Psyche (Stuttg.)* **2013**, 72–81 (2013).
- Mengual, X., Stahls, G. & Rojo, S. First phylogeny of predatory flower flies (Diptera, Syrphidae, Syrphinae) using mitochondrial COI and nuclear 28S rRNA genes: conflict and congruence with the current tribal classification. *Cladistics* **24**, 543–562 (2008).
- Stahls, G., Hipka, 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).
- Bain, R. S., Rashed, A., Cowper, V. J., Gilbert, F. S. & Sherratt, T. N. The key mimetic features of hoverflies through avian eyes. *Proc. R. Soc. Lond B Biol. Sci.* **274**, 1949–1954 (2007).
- Hondelmann, P. & Borgemeister C. P. Restriction fragment length polymorphisms of different DNA regions as genetic markers in the hoverfly *Episyrphus balteatus* (Diptera: Syrphidae). *Bull. Entomol. Res.* **95**, 349–359 (2005).
- Harmel, N. *et al.* Role of terpenes from aphid-infested potato on searching and oviposition behavior of *Episyrphus balteatus*. *Insect Sci.* **14**, 57–63 (2007).
- Rojo, S., Hopper, K. R. & Marcosgarcia, M. A. Fitness of the hover flies *Episyrphus balteatus* and *Eupeodes corollae* faced with limited larval prey. *Entomol. Exp. Appl.* **81**, 53–59 (1996).
- Tang, M. *et al.* Multiplex sequencing of pooled mitochondrial genomes - a crucial step toward biodiversity analysis using mitogenomics. *Nucleic Acids Res.* **42**, e166 (2014).
- Clary, D. O., Wahleithner, J. A. & Wolstenholme, D. R. Transfer RNA genes in *Drosophila* mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes. *Nucleic Acids Res.* **11**, 2411–2425 (1983).
- Wei, S. J., Shi, M., He, J. H., Sharkey, M. & Chen, X. X. The complete mitochondrial genome of *Diadegma semiclausum* (hymenoptera: ichneumonidae) indicates extensive independent evolutionary events. *Genome* **52**, 308–319 (2009).
- Hassanin, A., Léger, N. & Deutsch, J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Syst. Biol.* **54**, 277–298 (2005).
- Wei, S. J. *et al.* New views on strand asymmetry in insect mitochondrial genomes. *PLoS ONE* **5**, e12708 (2010).
- Mao, M., Valerio, A., Austin, A. D., Dowton, M. & Johnson, N. F. The first mitochondrial genome for the wasp superfamily Platygastroidea: the egg parasitoid *Trissolcus basalus*. *Genome* **55**, 194–204 (2012).
- Wei, S. J., Li, Q., van Achterberg, K. & Chen, X. X. Two mitochondrial genomes from the families Bethyliidae and Mutillidae: independent rearrangement of protein-coding genes and higher-level phylogeny of the Hymenoptera. *Mol. Phylog. Evol.* **77**, 1–10 (2014).
- Wei, S. J., Shi, B. C., Gong, Y. J., Li, Q. & Chen, X. X. Characterization of the Mitochondrial Genome of the Diamondback Moth *Plutella xylostella* (Lepidoptera: Plutellidae) and Phylogenetic Analysis of Advanced Moths and Butterflies. *DNA Cell Biol.* **32**, 173–187 (2013).
- 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).
- Bernt, M. *et al.* MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Mol. Phylog. Evol.* **69**, 313–319 (2013).
- Wolstenholme, D. R. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* **141**, 173–216 (1992).

39. Navajas, M., Le Conte, Y., Solignac, M., Cros-Arteil, S. & Cornuet, J. M. The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite *Varroa destructor* (Acari: Mesostigmata). *Mol. Biol. Evol.* **19**, 2313–2317 (2002).
40. Yeates, D. K. *et al.* Phylogeny and systematics of Diptera: Two decades of progress and prospects. *Zootaxa*, 565–590 (2007).
41. Ramakodi, M. P., Singh, B., Wells, J. D., Guerrero, F. & Ray, D. A. A 454 sequencing approach to dipteran mitochondrial genome research. *Genomics* **105**, 53–60 (2015).
42. Cameron, S. L., Miller, K. B., D'Haese, C. A., Whiting, M. F. & Barker, S. C. Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). *Cladistics* **20**, 534–557 (2004).
43. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725–2729 (2013).
44. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
45. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
46. Golubchik, T., Wise, M. J., Easteal, S. & Jermini, L. S. Mind the gaps: evidence of bias in estimates of multiple sequence alignments. *Mol. Biol. Evol.* **24**, 2433–2442 (2007).
47. Abascal, F., Zardoya, R. & Telford, M. J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38** (Web Server issue), 7–13 (2010).
48. Lanfear, R., Calcott, B., Ho, S. Y. & Guindon, S. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* **29**, 1695–1701 (2012).
49. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**, 539–542 (2012).
50. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
51. Zhong, M. *et al.* The complete mitochondrial genome of the scuttle fly, *Megaselia scalaris* (Diptera: Phoridae). *Mitochondrial DNA A DNA MappSeq Anal* **27**, 182–184 (2016).
52. Chen, S. *et al.* A cytoplasmic suppressor of a nuclear mutation affecting mitochondrial functions in *Drosophila*. *Genetics* **192**, 483–493 (2012).
53. Nelson, L. A., Cameron, S. L. & Yeates, D. K. The complete mitochondrial genome of the gall-forming fly, *Fergusonina taylori* Nelson and Yeates (Diptera: Fergusoninidae). *Mitochondrial DNA* **22**, 197–199 (2011).
54. Junqueira, A. C. *et al.* The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene* **339**, 7–15 (2004).
55. Nelson, L. A. *et al.* Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). *Gene* **511**, 131–142 (2012).

## Acknowledgements

We thank Yong-Li Yang for her help on the sequencing of the genomes. This research was funded by Sichuan Academy of Agricultural Sciences (2015JSCX-008) and the Tobacco Company of Sichuan (SCYC201402001C, LSYC201603) to D.Q.P., and the National Basic Research Program of China (2013CB127600), Beijing Natural Science Foundation (6162010), and the National Natural Science Foundation of China (31472025) to S.J.W.

## Author Contributions

S.J.W., F.S.M. and Y.J.L. conceived and designed the experiments; Y.Y.G., and P.C.J. collected the specimens; D.Q.P., H.L.L. and S.J.W. performed the experiments and analyzed the data; D.Q.P., and S.J.W. wrote the paper and discussed the results.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Pu, D.-q. *et al.* Mitochondrial genomes of the hoverflies *Episyrphus balteatus* and *Eupeodes corollae* (Diptera: Syrphidae), with a phylogenetic analysis of Muscomorpha. *Sci. Rep.* **7**, 44300; doi: 10.1038/srep44300 (2017).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017