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

De-qiang Pu¹, Hong-ling Liu², Yi-yun Gong¹, Pei-cheng Ji¹, Yue-jian Li³, Fang-sheng Mou¹ & Shu-jun Wei³

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 aging²⁴⁻⁵. 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 region³⁻⁴. 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 rate⁵⁻⁶. 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 groups⁷⁻¹².

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 Schizophora⁸⁻¹⁰⁻¹¹. 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 lineages¹²⁻¹⁶.

¹Industrial Crop Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu 610300, China. ²Institute of Plant Protection, Sichuan Academy of Agricultural Sciences, Chengdu 610066, China. ³Institute of Horticulture Research, Sichuan Academy of Agricultural Sciences, Chengdu 610066, China. ⁴Institute 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)
Table 1. Mitochondrial genomes used in this study.

| Sections | Family | Accession number | References |
|----------|--------|------------------|------------|
| Ingroup  |        |                  |            |
| Schizophora |        |                  |            |
| Aschiza | Episyrphus balteatus | Syrphoidea | Syrphidae | KU351241 | This paper |
|         | Eupeodes corollae | Syrphoidea | Syrphidae | KU379658 | This paper |
|         | Ocytamus sativus | Syrphoidea | Syrphidae | KT272862 | Junqueira, et al.
|         | Simosyrphus grandicornis | Syrphoidea | Syrphidae | DQ866050 | Cameron, et al.
|         | Syrphidae sp. | Syrphoidea | Syrphidae | KM244713 | Tang, et al.
|         | Bactrodera dorsalis | Tephritoidea | Tephritidae | DQ845759 | Zhong, et al.
|         | Drosophila melanogaster | Ephydridae | Drosophilidae | QJ666993 | Chen, et al.
|         | Fergusonia taylori | Oomyzidae | Fungusinaidae | HQ872008 | Nelson, et al.
|         | Nepompa mameae | Scioniomyzidae | Seppidae | KM605250 | Li, et al.
|         | Haematobus irritans | Muscoidea | Muscidae | DQ029097 | Unpublished |
|         | Musca domestica | Muscoidea | Muscidae | KT444442 | Unpublished |
|         | Chrysomya putorius | Oestroidea | Calliphoridae | AF352790 | Junqueira, et al.
|         | Dermatobia hominis | Oestroidea | Oestridae | AY463155 | Unpublished |
|         | Pollenia rudis | Oestroidea | Calliphoridae | JX913761 | Nelson, et al.
|         | Ravinia pernix | Oestroidea | Sarcophagidae | KM676414 | Unpublished |
| Outgroup | Cylidomyia duplonotata | Tabanoidea | Tabanidae | DQ866052 | Cameron, et al.
|         | Trichopentalia punctata | Nemestrinoidea | Nemestrinidae | DQ866051 | Cameron, et al.

Schizophora can be divided into two subsections, the Acalyptratae and Calyptratae, commonly referred to as acalyptate muscoids and calyptrate muscoids, respectively.

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. 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;

As of June 2016, three Syrphidae mitochondrial genome sequences were deposited into GenBank: Ocytamus sativus, Simosyrphus grandicornis and an unknown genus sp. Syrphidae (accession KM244713). 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.

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. Often exhibiting dense migratory swarm behavior, this, and the resemblance to wasps, may panic unaware people. Its feeding habit is rare among adult flies, as it is capable of crushing pollen grains as a food source. 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.

As a result, June 2016, three Syrphidae mitochondrial genome sequences were deposited into GenBank: Ocytamus sativus, Simosyrphus grandicornis and an unknown genus sp. Syrphidae (accession KM244713). 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).
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 rearrangement30.

**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 genomes31,32. 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 78.83% and 78.85%, 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 insects32.

| Gene | Position (bp) | Strand | Length (bp) | Intergenic nucleotides | Anti or start/stop codon |
|------|---------------|--------|-------------|------------------------|-------------------------|
| *trnI* | 1–66 | + | 66 | 2 | GAT |
| *trnQ* | 69–137 | – | 69 | 14 | TTG |
| *trnM* | 152–220 | + | 69 | 0 | CAT |
| *nad2* | 221–1252 | + | 1032 | 3 | ATT/TAA |
| *trnW* | 1256–1323 | + | 68 | 8 | TCA |
| *trnC* | 1332–1398 | – | 67 | 14 | GCA |
| *trnY* | 1413–1478 | – | 66 | 5 | GTA |
| *cox1* | 1484–3043 | + | 1560 | −5 | ATT/TAA |
| *trnL2* | 3039–3104 | + | 66 | 2 | TAA |
| *cox2* | 3107–3790 | + | 684 | 5 | ATG/TAA |
| *trnK* | 3796–3866 | + | 71 | 60 | CTG |
| *trnD* | 3927–3993 | + | 67 | 0 | GTC |
| *atp8* | 3994–4155 | + | 162 | −7 | ATT/TAA |
| *atp6* | 4149–4826 | + | 678 | 35 | ATG/TAA |
| *cox3* | 4862–5650 | + | 789 | 3 | ATG/TAA |
| *trnG* | 5654–5719 | + | 66 | 0 | TCC |
| *nad3* | 5720–6073 | + | 354 | 4 | AT/TA |
| *trnA* | 6078–6146 | + | 69 | −1 | TGC |
| *trnB* | 6146–6209 | + | 64 | 2 | TCG |
| *trnN* | 6212–6278 | + | 67 | −1 | GTT |
| *trnS1* | 6278–6344 | + | 67 | 1 | GCT |
| *trnE* | 6346–6410 | + | 65 | 23 | TCT |
| *trnF* | 6434–6500 | – | 67 | −1 | GAA |
| *nad5* | 6500–8221 | – | 1722 | 15 | ATT/TAA |
| *trnH* | 8237–8303 | – | 67 | −1 | GTG |
| *nad4* | 8303–9643 | – | 1341 | −7 | ATG/TAA |
| *nad4l* | 9637–9933 | – | 297 | 2 | ATG/TAA |
| *trnT* | 9936–10000 | + | 65 | 0 | TGT |
| *trnP* | 10001–10066 | – | 66 | 2 | TGG |
| *nad6* | 10069–10593 | + | 525 | 3 | AT/TA |
| *cob* | 10597–11733 | + | 1137 | 1 | ATG/TAA |
| *trn32* | 11735–11802 | + | 68 | 16 | TGA |
| *nad1* | 11819–12757 | – | 939 | 10 | ATA/TA |
| *trnl1* | 12768–12832 | – | 65 | 0 | TAG |
| *trnl* | 12833–14170 | – | 1338 | 0 | |
| *trnV* | 14171–14242 | – | 72 | 0 | TAC |
| *trn5* | 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.
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 30,33,34. 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

| Gene    | Position (bp) | Strand | Length (bp) | Intergenic nucleotides | Anti or start/stop codon |
|---------|---------------|--------|-------------|------------------------|--------------------------|
| trn1    | 1–66          | +      | 66          | –3                     | GAT                      |
| trnQ    | 64–133        | –      | 70          | 7                      | TGT                      |
| trnM    | 141–209       | +      | 69          | 0                      | CAT                      |
| nad2    | 210–1238      | +      | 1029        | –2                     | ATTT/TA                 |
| trnW    | 1237–1304     | +      | 68          | 4                      | TCA                      |
| trnC    | 1309–1374     | –      | 66          | 10                     | GCA                      |
| trnY    | 1385–1450     | –      | 66          | 22                     | GTA                      |
| cox1    | 1473–3029     | +      | 1557        | –5                     | ATTT/TA                 |
| trnL2   | 3025–3900     | +      | 66          | 2                      | TAA                      |
| cox2    | 3093–3776     | +      | 684         | 0                      | ATG/TA                  |
| trnK    | 3777–3847     | +      | 71          | 47                     | CTG                      |
| trnD    | 3895–3961     | +      | 67          | 0                      | GTG                      |
| atp8    | 5962–4123     | +      | 162         | –7                     | ATTT/TA                 |
| atp6    | 4117–4794     | +      | 678         | 16                     | ATG/TA                  |
| cox3    | 4811–5599     | +      | 789         | 3                      | ATG/TA                  |
| trnG    | 5603–5669     | +      | 67          | 0                      | TCC                      |
| nad3    | 5670–6023     | +      | 354         | 3                      | ATTT/TA                 |
| trnA    | 6027–6695     | +      | 69          | –1                     | TGC                      |
| trnR    | 6095–6158     | +      | 64          | 13                     | TCG                      |
| trnN    | 6172–6238     | +      | 67          | –1                     | GTT                      |
| trnS1   | 6238–6304     | +      | 67          | 5                      | GCT                      |
| trnE    | 6310–6376     | +      | 67          | 26                     | TTC                      |
| trnF    | 6403–6469     | –      | 67          | –1                     | GAA                      |
| nad5    | 6469–8189     | –      | 1721        | 15                     | ATTT/TA                 |
| trnH    | 8205–8270     | –      | 66          | –1                     | GTG                      |
| nad4    | 8270–9610     | –      | 1341        | –7                     | ATG/TA                  |
| nad4l   | 9604–9900     | –      | 297         | 2                      | ATG/TA                  |
| trnT    | 9903–9967     | +      | 65          | 0                      | TGT                      |
| trnP    | 9968–10033    | –      | 66          | 2                      | TGG                      |
| nad6    | 10036–10560   | +      | 525         | 3                      | ATTT/TA                 |
| cob     | 10564–11700   | +      | 1137        | 7                      | ATGT/TA                 |
| trnS2   | 11708–11775   | +      | 68          | 17                     | TGA                      |
| nad1    | 11792–12730   | –      | 939         | 10                     | ATA/TA                  |
| trnL1   | 12741–12805   | –      | 65          | 0                      | TAG                      |
| trnL    | 12806–14139   | –      | 1334        | 0                      | TAG                      |
| trnV    | 14140–14211   | –      | 72          | 0                      | TAC                      |
| trnS5   | 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.
medial level, whereas three genes of \textit{atp8}, \textit{nad2} and \textit{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\cite{35,36}.

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

| Species             | Accession number | PCGs  | rrl   | rrm   |
|---------------------|------------------|-------|-------|-------|
|                     |                  | T     | C     | A     | G     | Length (bp) | AT (%) | AT-skew | GC-skew | Length (bp) | AT (%) | Length (bp) | AT (%) |
| Episyrphus balteatus| KU351241         | 39.90 | 11.66 | 38.95 | 9.49  | 11220      | 78.85  | −0.0121  | −10.24   | 1338        | 84.60  | 804         | 83.96  |
| Euepeodes corallae  | KU379658         | 39.25 | 11.66 | 39.59 | 9.51  | 11211      | 78.83  | 0.0043   | −10.16   | 1334        | 84.78  | 795         | 83.14  |
| Ocyptamus sativus    | KT272862         | 39.71 | 11.68 | 39.26 | 9.35  | 11175      | 78.97  | −0.0058  | −11.106  | 1314        | 84.40  | 778         | 82.78  |
| Simosyrphus grandicornis | DQ866050       | 39.75 | 11.81 | 39.09 | 9.35  | 11208      | 78.84  | −0.0084  | −11.164  | 1339        | 84.99  | 804         | 83.83  |
| Megaselia scalaris   | KP974742         | 37.26 | 14.81 | 37.66 | 10.26 | 13301      | 74.92  | 0.0053   | −18.14   | 1318        | 81.03  | 786         | 79.64  |
| Bactrocera dorsalis   | DQ845759         | 32.94 | 17.85 | 38.16 | 11.05 | 11175      | 71.10  | 0.0734   | −23.53   | 1333        | 79.52  | 790         | 74.94  |
| Drosophila melanogaster | JQ686693       | 38.16 | 12.78 | 39.06 | 10.00 | 11169      | 77.22  | 0.0117   | −12.19   | 1323        | 82.84  | 786         | 80.15  |
| Fergusonina taylori   | HQ872008         | 36.13 | 14.19 | 39.92 | 9.76  | 11160      | 76.05  | 0.0498   | −18.52   | 1306        | 82.54  | 780         | 82.05  |
| Nemapoda mammevi     | KM605250         | 35.84 | 16.28 | 36.88 | 11.00 | 11181      | 72.72  | 0.0144   | −19.34   | 1321        | 80.62  | 783         | 77.65  |
| Haematobia irritans   | DQ029097         | 39.04 | 12.29 | 38.63 | 10.04 | 11178      | 77.67  | −0.0053  | −10.10   | 1321        | 82.66  | 784         | 78.83  |
| Musca domestica      | KT444442         | 37.84 | 13.61 | 38.24 | 10.31 | 11184      | 76.08  | 0.0053   | −13.79   | 1326        | 81.00  | 783         | 78.29  |
| Chrysomyia putoria   | AF352790         | 36.74 | 14.59 | 38.23 | 10.44 | 11187      | 74.97  | 0.0199   | −16.57   | 1329        | 81.79  | 785         | 76.94  |
| Dermatobia hominis   | AY463155         | 35.96 | 15.08 | 39.31 | 9.65  | 11187      | 75.27  | 0.0445   | −21.98   | 1324        | 83.08  | 788         | 78.68  |
| Pollenia rudis       | JX913761         | 37.29 | 14.14 | 38.52 | 10.04 | 11172      | 75.81  | 0.0163   | −16.95   | 1328        | 81.63  | 448         | 74.11  |
| Ravinia pernix       | KM676414         | 36.88 | 14.33 | 38.63 | 10.17 | 11181      | 75.50  | 0.0232   | −16.98   | 1328        | 82.08  | 786         | 77.48  |

**Figure 1. Relative synonymous codon usage (RSCU) of mitochondrial genomes in the two newly sequenced mitochondrial genomes of \textit{Episyrphus balteatus} and \textit{Euepeodes corallae}.** The stop codon is not given.
of Schizophora, which was proved to be a monophyletic group40. We study showed that Schizophora was inter-
ter and the other species of Schizophora. The Opomyzoidea was traditionally considered as a superfamily

C stems, and an A-A pair, located in the A-A stem.

stems; the other is an A-A pair, located in the A-A stem.

respectively. Among the six mismatched base pairs in

Episyrphus

Episyrphus

secondary structures. We found six and five mismatched base pairs in the tRNAs from

trnH

(17 sites), followed by the

trnG

(16 sites) (Fig. 2).

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 length39. The DHU arms in our tRNAs are 2 to 4 bp long, the AC arms

are 4 to 5 bp long, and the T

Ψ

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 trnH 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-U 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 monophy of all superfamilies used in the study. The Aschiza

(lower Cyclorrhapha) was found to be a paraphilic group40. We included two superfamilies of Platypezoidea and

Syrophoidea from Aschiza. Platypezoidea was sister to all other species of Muscomorpha, which is congruent

with previous studies11,41. The five genera of Syrophidae (Syrophoidea) clustered as ((unknown Syrophidae sp.) +

(Ocyptamus + (Eupeodes + (Episyrphus + Simosyrphus)))). Syrophoidea and Opopomizoidea formed a lineage, and

then sister to the other species of Schizophora. The Opopomizoidea was traditionally considered as a superfam-

ily of Schizophora, which was proved to be a monophyletic group40. We study showed that Schizophora was inter-
rupted by Opopomizoidea, which might be caused by the long-branch of Opopomizoidea. In Opopomizoidea, 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 Fergusoniniidae might affect

the evolutionary pattern of their mitochondrial genomes7. The long-branch of Opopomizoidea was also found in

previous study based on mitochondrial genome sequences41. Phylogenetic relationships of other groups included

in our analyses, i.e. ((Sciomyzoidea + Tephritoidae) + (Ephydroidea + (Muscoidea + Oestroidea))), were in accord with previous studies9,13–16.

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 °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 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.
Figure 3. Comparison on the secondary structure of the tRNA genes in Muscomorpha mitochondrial genomes. The secondary structures were draw from tRNA genes of *Episyrphus balteatus*. Variations of each sites 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.
PCR amplification and sequencing. Initially we used a previously designed set of universal primers for insect mitochondrial genomes 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 electrophoresis. 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, New York). The tRNA genes were initially identified using the Mitos WebServer (http://mitos.bioinf.uni-leipzig.de/index.py)37. 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 Syrphidae9,16,28. 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 MEGA643. The AT-skew and GC-skew were calculated according to Hassanin, et al.31: AT-skew = (A% − T%)/(A% + T%) and GC-skew = (G% − C%)/(G% + C%). The interspecifics 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 v544.

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, Platypexoidea 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 Trichophthalma punctata (Tabanomorpha: Tabanoidea: Tabanidae) were used as outgroups because of the close relationship between Tabanomorpha and Muscomorpha11.

MAFFT version 7.205, which implements consistency-based algorithms, was used for the alignment of protein-coding and RNA genes45. We used the G-INS-I and Q-INS-I algorithms in MAFFT46 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 1.147.

Data partitioning, and the ability to apply specific models to different partitions, is ideal for analyzing mitochondrial genomes. We used PartitionFinder version 1.148 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 3.2.549, and the ML method using RAxML version 8.0.050. 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

1. Cameron, S. L. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Syst. Entomol.* **39**, 400–411 (2014).
2. Cameron, S. L. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu. Rev. Entomol.* **59**, 95–117 (2014).
3. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
4. 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).
5. Curole, J. P. & Kocher, T. D. Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends. Ecol. Evol.* **14**, 394–398 (1999).
6. 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**, 666–702 (2004).
7. 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).
8. 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).
9. 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).
10. Demari-Silva, B. et al. Mitochondrial genomes and comparative analyses of *Culex campisi*, *Culex coronator*, *Culex usquitans* and *Culex usquitensis* (Diptera:Culicidae), members of the monarch group. *BMC Genomics* **16**, 831 (2015).
11. Li, X. et al. The first mitochondrial genome of the wasp superfamily *Platygastroidea*: the egg parasitoid *Trissolcus basalis* (Hymenoptera: Pteromalidae) indicates extensive independent evolutionary events. *Mol. Phylogen. Evol.* **38**, 201–213 (2008).
12. Zhang, B., Nardi, F., Hull-Sanders, H., Wan, X. & Liu, Y. The complete mitochondrial genome of *Culex pipiens* (Diptera: Culicidae) reveals diversity of lineages and non-monophyly of phyltaphagous taxa. *Mol. Phylog. Evol.* **49**, 715–727 (2008).
13. Junqueira, A. C. et al. Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. *Sci. Rep.* **6** (2016).
14. Mengual, X., Stahls, G. & Rojo, S. Phylogenetic resolution of alloclade (Diptera, Syrphidae) reveals diversity of lineages and non-monophyly of phyltaphagous taxa. *Mol. Phylog. Evol.* **49**, 715–727 (2008).
15. Grosskopf, G. Biology and life history of *Chelisola urbana* (Meigen) and *Chelisola palophilthalma* (Becker), two sympatric hoverflies approved for the biological control of hawkweeds (Hieracium spp.) in New Zealand. *Biol. Control* **35**, 142–154 (2005).
16. 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).
17. Reemer, M. Review and Phylogenetic Evaluation of Associations between Microdontinae (Diptera: Syrphidae) and Ants (Hymenoptera: Formicidae). *Psyche (Stuttg.)* **2013**, 72–81 (2013).
18. Rojo, S., Hopper, K. R. & Marcogarcia, M. A. Fitness of the hover flowers *Episyrphus balteatus* and *Eusephes corollae* faced with limited larval prey. *Entomol. Exp. Appl.* **81**, 53–59 (1996).
19. Tang, M. et al. Mitoupsequencing of pooled mitochondrial genomes - a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Res.* **42**, e166 (2014).
20. 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. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4397–4401 (1978).
21. Harmel, N. et al. First phylogeny of predatory flower flies (Diptera, Syrphidae, Syrphinae) using mitochondrial COI sequences and comparisons to mammalian 28S rRNA genes: conflict and congruence with the current tribal classification. *Syst. Entomol.* **33**, 335–346 (2008).
22. Mengual, X., Stahls, G. & Rojo, S. Phylogenetic relationships and taxonomic ranking of pipistre line wing flies (Diptera: Syrphidae) with implications for the evolution of aphidiophy. *Cladistics* **31**, 491–508 (2005).
23. Cook, C. E., Austin, J. J., Henry, R. & Disney, L. A mitochondrial 12S and 16S rRNA phylogeny of the flies of the family Allograpta (Diptera) and related families of Aschiza. *Zootaxa* **593**, 1–11 (2004).
24. Yeates, D. K., Wiegmann, B. M. Congruence and controversy: toward a higher-level phylogeny of Diptera. *Annu. Rev. Entomol.* **44**, 397–428 (1999).
25. Yeates, D. K. et al. Phylogeny and systematics of Diptera: Two decades of progress and prospects. *Zootaxa* **1668**, 565–590 (2007).
26. Eristalis tenax (Hymenoptera: Formicidae). *Psyche* (2006).
27. Rojo, S., Hopper, K. R. & Marcogarcia, M. A. Fitness of the hover flies *Episyrphus balteatus* and *Eusephes corollae* faced with limited larval prey. *Entomol. Exp. Appl.* **81**, 53–59 (1996).
28. Tang, M. et al. Mitoupsequencing of pooled mitochondrial genomes - a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Res.* **42**, e166 (2014).
29. 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).
30. Wei, S. J., Shi, M., He, J. H., Shawker, T. H. & Chen, X. X. The complete mitochondrial genome of the hoverfly *Eristalis tenax* (Diptera: Syrphidae) as potential mechanical vectors of pathogens causing mycobacterial infections in pig herds. *Folia Microbiol.* **51**, 147–153 (2006).
31. Reemer, M. Review and Phylogenetic Evaluation of Associations between Microdontinae (Diptera: Syrphidae) and Ants (Hymenoptera: Formicidae). *Psyche (Stuttg.)* **2013**, 72–81 (2013).
32. Mengual, X., Stahls, G. & Rojo, S. Phylogeny of predatory flower flies (Diptera, Syrphidae), syrphidae using mitochondrial COI and nuclear 28S rRNA genes: conflict and congruence with the current tribal classification. *Cladistics* **24**, 543–562 (2008).
33. Stahls, 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).
34. 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).
35. Eristalis tenax (Hymenoptera: Formicidae). *Psyche* (2006).
36. 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-Arté, 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., Eastal, S. & Jerumiin, 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. Multiple alignments 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 (2015SCX-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., E.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