Mitochondrial genomes of two Australian fishflies with an evolutionary timescale of Chauliodinae

Fan Yang, Yunlan Jiang, Ding Yang & Xingyue Liu

Fishflies (Corydalidae: Chauliodinae) with a total of ca. 130 extant species are one of the major groups of the holometabolous insect order Megaloptera. As a group which originated during the Mesozoic, the phylogeny and historical biogeography of fishflies are of high interest. The previous hypothesis on the evolutionary history of fishflies was based primarily on morphological data. To further test the existing phylogenetic relationships and to understand the divergence pattern of fishflies, we conducted a molecule-based study. We determined the complete mitochondrial (mt) genomes of two Australian fishfly species, Archichauliodes deceptor Kimmins, 1954 and Protochauliodes biconicus Kimmins, 1954, both members of a major subgroup of Chauliodinae with high phylogenetic significance. A phylogenomic analysis was carried out based on 13 mt protein coding genes (PCGs) and two rRNAs genes from the megalopteran species with determined mt genomes. Both maximum likelihood and Bayesian inference analyses recovered the Dysmicohermes clade as the sister group of the Archichauliodes clade + the Protochauliodes clade, which is consistent with the previous morphology-based hypothesis. The divergence time estimation suggested that the divergence among the three major subgroups of fishflies occurred during the Late Jurassic and Early Cretaceous when the supercontinent Pangaea was undergoing sequential breakup.

The subfamily Chauliodinae, commonly known as fishflies, is one of the three major groups of the holometabolous order Megaloptera. It belongs to the family Corydalidae, which also includes the subfamily Corydalinae (dorsoflies) and which is regarded as the sister group to the family Sialidae. Compared to Corydalinae, adult fishflies can be distinguished by the absence of postocular plane on head, the reduced cross venation, the callus cerci present on ectoprocts, and the reduced male gonostylus. Fishfly larvae are easily recognized by the absence of ventral tufts and the presence of specialized spiracles on abdominal segment. Currently, over 130 species in 18 genera of extant fishflies are known worldwide. Fishflies live mainly in the subtropical or warm temperate regions, and they occur in all zoogeographical realms. However, they show a remarkably discontinuous distribution due to their absence in the western Palaearctic realm and most parts of the Afrotropical and Neotropical realms. Fishflies are an archaic insect group and many of the extant species qualify as "living fossils", since they originated no later than the Middle Jurassic based on the fossil evidence and display remarkably conservative, unchanged adult and larval morphology between Mesozoic fossil and modern species. Hence, the phylogeny and historical biogeography of fishflies are of high interest and have been recently studied by Liu & Yang (2006), Liu et al. (2012, 2016) and Wang et al. (2012). The current phylogenetic framework of Chauliodinae subdivides the subfamily into three extant groups, i.e., the Dysmicohermes clade, the Protochauliodes clade and the Archichauliodes clade. The Dysmicohermes clade comprises only two western Nearctic endemic genera Dysmicohermes and Orohermes. The Protochauliodes clade is composed of Madachauliodes, Neothermes, Notochauliodes, Protochauliodes and Taeniochauliodes, many of which are distributed in the Southern Hemisphere except Neohermes and some species of Protochauliodes from North America. The Archichauliodes clade includes Platychauliodes from South Africa, Archichauliodes and Apochauliodes from Australia, New Zealand and Chile, and all Asian fishfly genera. The modern fauna of fishflies is thought to be formed by the divergence associated with the sequential breakup and drifting of Gondwana.
However, despite the divergence between Chauliodinae and Corydalinae, which was estimated by a molecular approach to have occurred during the Early Jurassic (~186 MA), all previous hypotheses on phylogeny and historical biogeography of Chauliodinae were proposed without the use of molecular data.

Here we determined and describe the complete mitochondrial (mt) genomes of two fishfly species, namely *Archichauliodes deceptor* Kimmins, 1954 and *Protochauliodes biconicus* Kimmins, 1954. Both are endemic to Australia and are the first fishfly species from the Southern Hemisphere for which the mt genome has been determined. The genome organization, protein-coding genes, transfer RNAs, ribosomal RNAs and the control region were analyzed. A phylogenomic analysis was performed with known mt genome data of Megaloptera to infer the phylogenetic positions of *Archichauliodes* and *Protochauliodes* and to test the previous phylogeny of Chauliodinae based on morphological data. Furthermore, the evolutionary pattern of the three major subgroups of Chauliodinae was reconstructed based on divergence time estimation. The results corroborated the relationships of the three major subgroups of fishflies that was based on morphological data. Additionally, the first molecule-based timescale on the early divergence of fishflies is presented. Lastly, the historical biogeography of fishflies is discussed in light of the new evidence from the molecular data.

**Results**

**Genome organization and structure.** The complete mt genome of *A. deceptor* is a typical circular, double-strand molecule of 15,797 bp in length (GenBank accession number: KU925864; Fig. 1, Table 1), which is relatively small in size compared to the mt genomes of Megaloptera known thus far, with length ranging from 15,687 bp (*Corydalus cornutus*, Corydalidae, NC_011226) to 16,271 bp (*Dysmicohermes ingens*, Corydalidae, NC_16271). The mt genome contains 37 genes, including 22 tRNAs, 13 PCGs, two rRNAs and a control region. The sequenced part of the *P. biconicus* mt genome is 14,384 bp in length and contains 34 genes with 19 tRNAs, 13 PCGs, 1rRNA and partial 2rRNA (Fig. 2, Table 2). Three tRNAs (i.e., trNA^Ser^, trNA^Met^, trNA^Glu^) and the control region failed to be amplified probably due to high variation and complex secondary structures of this part. In the Megaloptera mt genomes, variations in the length of PCGs, tRNAs, 1rRNA and 2rRNA are inconspicuous except for the length of the control region (see Fig. 3; Table S1).
The gene order is in accordance with the gene order of *Drosophila yakuba*[^6], and no gene rearrangement was found. The published mt genomes of all 12 species of Megaloptera exhibit a highly conserved gene order[^7]. However, the gene order of some reported Neuroptera mt genomes differs slightly from the conserved gene order in the translocation of *tRNACys*, which is located upstream of *tRNATrp* but not at its traditional downstream location of *tRNATrp*. Among all 37 genes in the *A. deceptor* mt genome, 14 genes (4 PCGs, 2 rRNAs and 8 tRNAs) are encoded on the minority strand (N-strand), and 23 genes (9 PCGs and 12 tRNAs) are on the majority strand (J-strand). Gene overlaps were found at 12 and 16 gene junctions in the mt genomes of *A. deceptor* and *P. biconicus*, respectively. Furthermore, *ATP6* and *ATP8* overlap 7 nucleotides (i.e., "ATGATAA"), and this phenomenon is also reported in the mt genome of some related species (e.g., *Neochauliodes bowringi* (McLachlan), *Neochauliodes punctatolosus* Liu & Yang, *Protohermes concolorus* Yang & Yang). Similarly, *ND4L*-*ND4* also had a 7 bp overlap (i.e., "TTAACAT"), but the overlapped sequences between *ND4L*-*ND4* were not always the same in insect mt genomes, such as "TTAACAC" in *N. bowringi* and "ATGTTAA" in *N. punctatolosus*. In addition, there were 13 intergenic regions in the mt genome including 65 nucleotides and ranging from 1 to 16 bp in the *A. deceptor* mt genome. In the *P. biconicus* mt genome, 12 intergenic regions were found, including 96 nucleotides which ranged

| Gene       | Direction | Location | Size (bp) | IGN* | Anticodon | Codon | AT%  |
|------------|-----------|----------|-----------|------|-----------|-------|------|
| tRNA^Ile  | F         | 1–64     | 64        | 0    | GAT       |       | 68.8 |
| tRNA^Gln  | R         | 62–130   | 69        | −3   | TTA       | 79.7  |
| tRNA^Met  | F         | 131–199  | 69        | 0    | CAT       |       | 71   |
| ND2       | F         | 200–1222 | 1023      | 0    | ATG       | TAA   | 79.5 |
| tRNA^Thr  | F         | 1221–1285| 65        | −2   | TCG       |       | 79.4 |
| tRNA^Cys  | R         | 1277–1341| 63        | −9   | GCA       |       | 81   |
| tRNA^Tyr  | R         | 1341–1411| 66        | −1   | GTC       |       | 81.2 |
| COI       | F         | 1398–2939| 1542      | −4   | GAT       | TAA   | 69.2 |
| tRNA^Svu  | F         | 2941–3004| 64        | 1    | TAA       |       | 76.6 |
| COII      | F         | 3007–3691| 685       | 2    | ATG       | T−    | 73.1 |
| tRNA^Asn  | F         | 3692–3762| 71        | 0    | GTT       |       | 74.6 |
| tRNA^Trp  | F         | 3762–3827| 66        | −2   | GTC       |       | 83.3 |
| ATP8      | F         | 3828–3986| 159       | 0    | ATC       | TAA   | 82.4 |
| ATP6      | F         | 3980–4057| 678       | 2    | ATG       | TAA   | 74.5 |
| COIII     | F         | 4657–5445| 789       | −1   | ATG       | TAA   | 71.7 |
| tRNA^Glu  | F         | 5448–5509| 62        | 2    | TCC       |       | 82.2 |
| ND3       | F         | 5510–5863| 354       | 0    | ATT       | TAG   | 78.8 |
| tRNA^Ser  | F         | 5862–5924| 63        | −2   | TGC       |       | 71.4 |
| tRNA^Gly  | F         | 5935–5997| 64        | 10   | TCG       |       | 71.9 |
| tRNA^Ala  | F         | 5997–6062| 66        | −1   | GTT       |       | 80.3 |
| tRNA^Asp  | F         | 6063–6129| 67        | 0    | GCT       |       | 73.2 |
| tRNA^Lys  | F         | 6130–6195| 66        | 0    | TTC       |       | 89.4 |
| tRNA^His  | R         | 6194–6258| 65        | −2   | GAA       |       | 78.5 |
| ND5       | R         | 6259–7975| 1722      | 0    | ATA       | T−    | 77.5 |
| tRNA^Ish  | R         | 7982–8044| 63        | 6    | GTG       |       | 81   |
| ND4       | R         | 8045–9383| 1339      | 0    | ATG       | T−    | 78.9 |
| ND4L      | R         | 9377–9670| 294       | −7   | ATG       | TAA   | 79.6 |
| tRNA^Hip  | F         | 9673–9737| 65        | 2    | TGT       |       | 81.5 |
| tRNA^Giv  | R         | 9738–9803| 66        | 0    | TGG       |       | 81.8 |
| ND6       | R         | 9809–10318| 510      | 5    | ATT       | TAA   | 83   |
| CYTB      | F         | 10335–11471| 1137    | 16   | ATG       | TAA   | 73.5 |
| tRNA^Bas  | F         | 11474–11540| 67      | 2    | TGC       |       | 86.6 |
| ND1       | R         | 11553–12503| 951    | 12   | TTG       | TAA   | 76.8 |
| tRNA^Aau  | R         | 12505–12568| 64      | 1    | TAG       |       | 81.3 |
| rRNA      | R         | 12572–13889| 1318   | 3    | 82
| tRNA^Ish  | R         | 13891–13961| 71      | 1    | TAC       |       | 77.5 |
| sRNA      | R         | 13959–14748| 790     | −3   | 79.4
| Control region | — | 14749–15798 | 1050  | 0    | 86.8 |

Table 1. Organization of the *Archichauliodes deceptor* mt genome. IGN: Intergenic nucleotide, minus sign indicates overlapping between genes. tRNA^X^: where X is the abbreviation of the corresponding amino acid.

[^6]: The gene order is in accordance with the gene order of *Drosophila yakuba*, and no gene rearrangement was found. The published mt genomes of all 12 species of Megaloptera exhibit a highly conserved gene order. However, the gene order of some reported Neuroptera mt genomes differs slightly from the conserved gene order in the translocation of *tRNACys*, which is located upstream of *tRNATrp* but not at its traditional downstream location of *tRNATrp*. Among all 37 genes in the *A. deceptor* mt genome, 14 genes (4 PCGs, 2 rRNAs and 8 tRNAs) are encoded on the minority strand (N-strand), and 23 genes (9 PCGs and 14 tRNAs) are on the majority strand (J-strand). Among 33 genes in the partial *P. biconicus* mt genome, 12 genes (4 PCGs, 1 rRNA and 7 tRNAs) are encoded on the minority strand (N-strand), and 21 genes (9 PCGs and 12 tRNAs) are on the majority strand (J-strand). Gene overlaps were found at 12 and 16 gene junctions in the mt genomes of *A. deceptor* and *P. biconicus*, respectively. Furthermore, *ATP6* and *ATP8* overlap 7 nucleotides (i.e., "ATGATAA"), and this phenomenon is also reported in the mt genome of some related species (e.g., *Neochauliodes bowringi* (McLachlan), *Neochauliodes punctatolosus* Liu & Yang, *Protohermes concolorus* Yang & Yang). Similarly, *ND4L-ND4* also had a 7 bp overlap (i.e., "TTAACAC"), but the overlapped sequences between *ND4L-ND4* were not always the same in insect mt genomes, such as "TTAACAC" in *N. bowringi* and "ATGTTAA" in *N. punctatolosus*. In addition, there were 13 intergenic regions in the mt genome including 65 nucleotides and ranging from 1 to 16 bp in the *A. deceptor* mt genome. In the *P. biconicus* mt genome, 12 intergenic regions were found, including 96 nucleotides which ranged
from 1 to 19 bp. Fifteen non-coding regions were found in the whole mt genome of *A. deceptr*. The largest non-coding region had 1050 bp and was located between *srRNA* and *tRNA_{Ile}* with the A+T content at 86.8%.

**Protein-coding genes.** The overall A+T content of all 13 PCGs in the genomes of *A. deceptr* and *P. biconicus* was 75.8% and 76%, respectively. Twelve PCGs in the *A. deceptr* mt genome and all 13 PCGs in the *P. biconicus* mt genome use ATN as the start codon, the only exception refers to the *ND1*, which initiate with TTG in the *A. deceptr* mt genome. The stop codons mostly used are TAA, while in the mt genome of two species, *ND3* used TAG as the stop codon, and the other genes used an incomplete stop codon T.

**Transfer RNAs.** According to the secondary structure and corresponding anticodon of tRNAs, we identified 22 tRNA genes in the mt genome of *A. deceptr*, ranging in size from 63 bp (*tRNAGly*), to 71 bp (*tRNALys*, *tRNA_{Val}*). In the partial mt genome of *P. biconicus*, there were 19 tRNA genes ranging in size from 62 bp (*tRNA_{Phe}*), to 71 bp (*tRNA_{Lys}*). Fourteen tRNA genes in *A. deceptr* and 12 genes in *P. biconicus* are encoded on the J-strand, while the remaining tRNAs are encoded on the N-strand. Most tRNAs could be folded as typical clover-leaf structures except for *tRNA_{Ser}(AGN)* due to lack of the DHU arm (Figs 4 and 5), and the AC arm of *tRNA_{Ser}(AGN)* was a stem structure with an independent nucleotide in the middle. This phenomenon is common in sequenced Neuroptera mt genomes. The DHU and T\(^\psi\)C stems are more variable and range in size from 3 bp to 10 bp.

Based on the secondary structure, the amount of mismatched base pairs in tRNAs of *A. deceptr* and *P. biconicus* was 21 and 19, respectively, this includes G-U pairs, A-A mismatches, U-U mismatches and U-C mismatch. These mismatches were mostly found in the 3' end of acceptor stem and DHU arm.

**Ribosomal RNAs.** The *lrRNA* was assumed to fill up the blanks between *tRNA_{Leu}(CUA)* and *tRNA_{Val}* while the *srRNA* was flanked by *tRNA_{Val}* and the control region. In *A. deceptr* the length of *lrRNA* and *srRNA* was determined to be 1319 bp and 786 bp with A+T content as 82% and 79.4%, respectively. The *lrRNA* of *P. biconicus* is 1318 bp in length, while the size of *srRNA* we sequenced was 610 bp.

---

**Figure 2.** Mitochondrial map of *Protochauliodes biconicus*. Circular maps were drawn with CGView\(^25\). The arrows indicate the orientation of gene transcription. The tRNAs are denoted by the color blocks and are labelled according to the IUPAC-IUB single-letter amino acid codes (L1: UUR; L2: CNU; S1: AGN; S2: UCN). The GC content was plotted using a black sliding window, as the deviation from the average GC content of the entire sequence. GC-skew was plotted as the deviation from the average GC-skew of the entire sequence. The inner cycle indicates the location of the genes in the mt genome.
We inferred the secondary structure of lrRNA and srRNA of *A. deceptor* using the published rRNA structure of *Neoneuromus tonkinensis* and *Agriosphodrus dohrni* as models. There were 49 helices in lrRNA in five structural domains (I-II, IV-VI), domain III is absent as in other arthropods (Fig. 6). The multiple alignments of

| Gene        | Direction | Location | Size (bp) | IGN* | Anticodon | Codon | Start | Stop | AT% |
|-------------|-----------|----------|-----------|------|-----------|-------|-------|------|-----|
| ND2         | F         | 5–1027   | 1023      | 0    | ATT       | TAA   | 79.3  |
| tRNA^Trp   | F         | 1026–1092 | 67        | −2   | TCA       |       |       | 76.1 |
| tRNA^Cys   | R         | 1085–1147 | 63        | −8   | GCA       |       |       | 73.0 |
| tRNA^Tyr   | R         | 1148–1213 | 66        | 0    | GTA       |       |       | 66.7 |
| COI         | F         | 1206–2745 | 1540      | −8   | ATT       | T–    | 69.1  |
| tRNA^Gly   | F         | 2751–2815 | 65        | 5    | TAA       |       |       | 73.9  |
| COI         | F         | 2817–3498 | 682       | 5    | ATG       | T–    | 74.5  |
| tRNA^Val   | F         | 3506–3576 | 71        | 7    | CTT       |       | 69.0  |
| tRNA^Glu   | F         | 3576–3641 | 66        | −1   | GTC       |       |       | 80.3  |
| ATP8        | F         | 3642–3808 | 159       | 0    | ATT       | TAA   | 81.8  |
| ATP6        | F         | 3794–4469 | 676       | −7   | ATG       | T–    | 75.4  |
| COII        | F         | 4470–5258 | 789       | 0    | ATG       | TAA   | 70.7  |
| tRNA^Asp   | F         | 5262–5325 | 64        | 3    | TCC       |       | 79.7  |
| ND3         | F         | 5345–5680 | 336       | 19   | ATA       | TAG   | 76.5  |
| tRNA^Asn   | F         | 5679–5743 | 65        | −2   | TGC       |       | 76.9  |
| tRNA^Lys   | F         | 5755–5817 | 63        | 11   | TCG       |       | 74.3  |
| tRNA^Tyr   | F         | 5817–5881 | 65        | −1   | GTT       |       | 78.5  |
| tRNA^Arg   | F         | 5881–5949 | 69        | −1   | GCT       |       | 75.4  |
| tRNA^Pro   | F         | 5949–6012 | 64        | −1   | TTC       |       | 89.1  |
| tRNA^Thr   | R         | 6011–6072 | 62        | −2   | GAA       |       | 77.4  |
| ND5         | R         | 6073–7797 | 1725      | 0    | ATA       | TAA   | 78.1  |
| tRNA^Tyr   | R         | 7804–7866 | 63        | 6    | GTC       |       | 84.2  |
| ND4         | R         | 7865–9206 | 1342      | −2   | ATA       | T–    | 78.8  |
| ND4L        | R         | 9200–9494 | 300       | −7   | ATA       | TAA   | 83.4  |
| tRNA^Val   | F         | 9500–9564 | 65        | 5    | TGT       |       | 83.1  |
| tRNA^Phe   | R         | 9565–9629 | 65        | 0    | TGG       |       | 80.0  |
| ND6         | F         | 9632–10153| 522       | 2    | ATT       | TAA   | 85.2  |
| CYTB        | F         | 10153–11289| 1137     | −1   | ATG       | TAA   | 73.9  |
| tRNA^Glu   | F         | 11289–11355| 67       | −1   | TGA       |       | 82.1  |
| NDJ         | R         | 11371–12325| 955     | 15   | ATT       | T–    | 73.8  |
| tRNA^His   | R         | 12325–12387| 63       | −1   | TAG       |       | 79.4  |
| tRNA^Leu   | R         | 12394–13711| 1318    | 6    |           |       | 81.7  |
| tRNA^Glu   | R         | 13706–13776| 71       | −6   | TAC       |       | 70.4  |
| srRNA       | R         | 13776–14384| 610     | −1   |           |       | 80.1  |

Table 2. Organization of the *Protochauliodes biconicus* mt genome. IGN: Intergenic nucleotide, minus sign indicates overlapping between genes. tRNA^X^: where X is the abbreviation of the corresponding amino acid.

Figure 3. The size of PCGs, lrRNA, srRNA and CR, respectively, among sequenced Megaloptera mt genomes.
Chaulioidinae and Megaloptera indicate that conserved nucleotides were distributed unevenly throughout the lrRNA secondary structure. In addition, most invariable positions were found within domain IV, while the lower conserved positions were in domains I-II. The secondary structure of srRNA contains three domains (Fig. 7), and it is less conservative than lrRNA. The H7 region within srRNA is highly variable and difficult to predict among different insects.

**Nucleotide composition and codon usage.** The nucleotide composition of mt genomes of *A. deceptor* and *P. biconicus* is clearly biased towards A/T nucleotides (*A. deceptor*: A = 39.5%, T = 38.0%, C = 13.9%, G = 8.6%; *P. biconicus*: A = 38.0%, T = 39.0%, C = 13.7%, G = 9.3%) (Tables S2–S3). The A + T content is much higher than G + C content in all mt genes of *A. deceptor* and *P. biconicus*, respectively, i.e. PCGs (75.8%, 76.0%),
tRNAs (78.5%, 76.5%), rRNAs (81.0%, 81.2%) and the control region (86.8%). The AT-Skew and GC-Skew in the mt genome of *A. deceptor* are 0.019 and −0.236, while in *P. biconicus* they are −0.013 and −0.191.

The codon usage of all PCGs in the mt genomes of *A. deceptor* and *P. biconicus* is similar to that in other invertebrates (Tables S4–S5). We found that NNU was the most frequently used codon, while NNA was only used in Leu (UUR), Met (AUN), Gln (CAN), Lys (AAN), Glu (GAN) and Gly (GGN). NNG and NNC are less used codons. In addition, A and T bias is reflected in the codon usage, such that the A + T rich codons, i.e. TTT-Phe, TTA-Leu, ATT-Ile, ATA-Met, TAT-Tyr, AAT-Asn and AAA-Lys, are more frequently used than the G + C rich codons.

**Phylogenetic analysis and divergence time estimation.** The ML and BI analyses based on the final dataset of 13,247 nucleotide sites generated the phylogenetic trees with same topologies and high nodal

---

**Figure 5.** Inferred secondary structure of 19 tRNAs in the *Protochauliodes biconicus* mt genome. Most tRNAs are labeled with the abbreviations of their corresponding amino acids. Dash (−) indicates Watson-Crick bonds and dot (⋅) indicates GU bonds.
supports (Fig. 7). Corydalinae and Chaulioidinae were both monophyletic and together formed a monophyletic Corydalidae, which is the sister group of Sialidae. In Chaulioidinae, Dysmicohermes was assigned the sister group of the remaining genera of fishflies. Protochauliodes was recovered as the sister group of Archichauliodes + Neochauliodes.

The estimation of divergence times among the sampled Megaloptera taxa showed a Mesozoic diversification for the extant families and subfamilies of Megaloptera, as well as for the genera of Corydalidae (Fig. 8, Table 3). In Chaulioidinae, the divergence between Dysmicohermes and the other genera of fishflies was dated to be in the Late Jurassic ca. 159 MA (95% HPD 121.01–169.64 MA), which is slightly earlier than the estimate in Wang et al. (ca. 140 MA/95%HPD 67.96–234.89 MA); however, in consideration of the 95% credibility interval, the current estimate falls within the range of the previous study. Protochauliodes was estimated to have diverged from Archichauliodes + Neochauliodes in the Early Cretaceous, ca. 117 MA. Divergence between Archichauliodes and Neochauliodes was also dated in the Early Cretaceous, ca. 102 MA. In Corydalinae, the divergence times among the five dobsonfly genera were estimated to be in the Cretaceous, which corresponds to results in Winterton et al. and Wang et al.

Discussion
Phylogenetic considerations. The species of fishflies investigated here from Archichauliodes and Protochauliodes are significant for inferring the phylogeny of Chaulioidinae, since they represent two major subgroups of fishflies, i.e. the Archichauliodes clade and the Protochauliodes clade, as proposed by a recent morphology-based phylogeny of fishflies. Aside from these two clades, the remaining major subgroup of Chaulioidinae is the Dysmicohermes clade. Our phylogenetic analysis is the first to use molecular data to test the relationships of these three major fishfly subgroups. Our results are generally consistent with the previous morphology-based phylogeny, in which the Dysmicohermes clade was the sister group of the Archichauliodes clade + the Protochauliodes clade, although we could not corroborate the monophyly of each clade due to lack of many genera.

Dysmicohermes, together with its sister genus Orohermes, are basal fishflies and possess a number of plesiomorphic morphological characters, such as the moderately developed male gonocoxites and gonostyli, and the feebly produced spiracles on the larval abdominal segment. Furthermore, the wing venation and larval morphology of Dysmicohermes and Orohermes largely resemble that in Jurochauliodes which is one of the most ancestral fishfly genera currently known from the Middle Jurassic of China. Thus, the Dysmicohermes clade should be considered as the basal most subgroup among extant Chaulioidinae on the basis of morphological and molecular evidence. The divergence time estimation indicated that the evolutionary history of this subgroup is considerably long and dates to the earliest Late Jurassic (ca. 159 MA).

The Protochauliodes clade currently includes five genera, four of which possess a distinct wing character (i.e., anterior branch of 2 A partially fused with stem of 1 A) which supports the autapomorphic nature of these genera. Furthermore, additional genital characters indicate autapomorphies of the Protochauliodes clade. This
clade was estimated by our analysis to have diverged with the *Archichauliodes* clade during the Early Cretaceous. However, Liu et al. (2012) postulated that these two clades might have diverged during the Early Jurassic. In fact, this hypothesis was proposed based on the inclusion of a fossil fishfly genus in the *Protochauliodes* clade, namely *Eochauliodes* from the Middle Jurassic of China, which accordingly prolonged the interpretation on the evolutionary history of this clade. Nevertheless, the evidence supporting the assignment of *Eochauliodes* in the *Protochauliodes* clade is weak, since it is a homoplasious wing venation character, i.e., bifurcated anterior branch of forewing Rs. Future studies may reveal that this Middle Jurassic fossil genus is distantly related and diverged much earlier than the extant *Protochauliodes* clade and the *Archichauliodes* clade.

**Biogeographic considerations.** Previous study on the phylogeny and historical biogeography of Chaulioidinae suggested a Pangaean origin and a global distribution of the subfamily before the Middle Jurassic, while the earliest diversification of fishflies might have occurred before the initial split of Pangaea. Moreover, Liu et al. (2012) considered that the divergence among the three major fishfly clades might have taken place when Pangaea was not yet separated. However, our estimate for divergence between the *Dysmicohermes* clade and the *Archichauliodes* clade + the *Protochauliodes* clade is slightly after the initial breakup of Pangaea, which led to the formation of Laurasia and Gondwana during an interval of 180–160 MA. Since the *Archichauliodes* and *Protochauliodes* clades were considered to have originated from Gondwana, the divergence of the *Dysmicohermes* clade, which is endemic to western North America, could be correlated to the geographic vicariance formed by the separation of Laurasia and Gondwana.

The *Archichauliodes* and *Protochauliodes* clades include many austral endemic genera, which were thought to have diverged in connection with the sequential breakup of Gondwana. As mentioned above, the molecule-based result of the divergence time between these two clades is much later than that inferred from the morphological data, with the molecular estimate being ca. 117 MA (95% HPD 84.50–138.87 MA) in the Early Cretaceous. The sequential breakup of Gondwana continued into the Early Jurassic and Late Cretaceous. By ca. 120 MA, Gondwana had split into several landmasses, e.g. Africa + northern South America, Madagascar + India, and a landmass including Antarctica, Australia, southern South America, etc. In Chaulioidinae, both *Archichauliodes*
and Protochauliodes clades include extant genera distributed in the areas belonging to at least two of the above main Gondwanan landmasses. If the initial divergence between the Archichauliodes and Protochauliodes clades took place after 120 MA, it would be difficult to infer any correlation of the intergeneric divergence within either of these two major clades to Gondwanan plate drifting. Moreover, the disjunct distribution of the austral endemic genera in the Archichauliodes and Protochauliodes clades seems to be insufficiently explained since fish-flies possess a relatively weak capacity for long-distance dispersal. Therefore, the initial divergence between the Archichauliodes and Protochauliodes clades might have been much earlier than 120 MA. Based on the range of their divergence time presently estimated (95% HPD 84.50–138.87 MA), it is plausible to assume that these two clades diverged...


| Order        | Family/Subfamily           | Species                          | Accession number |
|--------------|---------------------------|----------------------------------|------------------|
| Megaloptera  | Corydalidae/Cauliolidinae | Archichauliodes deceptor         | KU925864         |
|              | Corydalidae/Cauliolidinae | Neochauliodes punctatolousus     | NC_018772        |
|              | Corydalidae/Cauliolidinae | Neochauliodes bowringi           | NC_023444        |
|              | Corydalidae/Cauliolidinae | Dysmicothermes ingens            | NC_024657        |
|              | Corydalidae/Corydalinae   | Protothermes concolorus          | NC_011524        |
|              | Corydalidae/Corydalinae   | Neoneuvromas tonkinensis         | NC_027852        |
|              | Corydalidae/Corydalinae   | Neuvromus exterior               | NC_027851        |
|              | Corydalidae/Corydalinae   | Acanthacorydalis orientalis      | NC_023462        |
|              | Corydalidae/Corydalinae   | Corydalus cornutus               | NC_011276        |
| Sialidae     |                           | Sialis hamata                    | NC_013256        |
| Neuroptera   | Osmylidae                 | Thyridosmylus langii             | NC_021415        |
|              | Ithonidae                 | Rapisma xizangense               | KF626447         |

Table 4. Taxa used in the phylogenetic analysis.

Conclusions

The present study is the first to present a phylogenetic analysis based on mt genomic data to infer relationships among the major subgroups of Chauliodinae. Similar to the previous morphology-based intergeneric phylogeny of fishflies, the present results indicate that the Dysmicothermes clade is the sister group of the Archichauliodes clade + the Protochaulioides clade. However, the timescale we estimated for the divergence among the three major subgroups is much later than that hypothesized from the morphology-based phylogeny, suggesting these major divergences were possibly infected by the sequential breakup of Pangaea during the Late Jurassic and Early Cretaceous. Unfortunately, it is still hard to reveal any clear divergence pattern of the whole subfamily due to lack of many genera, particularly those endemic to certain austral landmasses. Future study should focus on a total-evidence analysis with comprehensive sampling to elucidate the evolutionary history of fishflies.

Material and Methods

Specimens and DNA extraction. The specimen of A. deceptor was collected by S. L. Winterton and J. S. Bartlett at Scrub Rd (27.427°S 152.841°E), Brisbane Forest Pk, SE Queensland, Australia, between December 2007 and January 2008. The specimen of P. biconicus was collected by H. Karube at Brisbane, Australia, on November 12, 2005. After collection, the samples were initially preserved in 95% ethanol in the field, and transferred to −20°C for the long-term storage upon arrival at the China Agricultural University (CAU). All samples were examined and identified by Xingyue Liu. The genomic DNA was extracted and purified from the mesothoracic muscle using TIANamp Genomic DNA Kit (TIANGEN).

PCR amplification and sequencing. The mt genomes of A. deceptor and P. biconicus were generated by amplification of overlapping PCR fragments. PCR primers we used included universal and specifically designed primers (Tables S6–S7). All PCRs were performed using NEB Long Taq DNA polymerase (New England Biolabs, Ipswich, MA) under the following amplification conditions: 95°C for 30 s, 40 cycles of denaturation at 95°C for 10 s, annealing at 43–55°C (depending on the primer pair used) for 50 s, elongation at 65°C for 1 kb/min (depending on the size of amplicon), and the final elongation step at 65°C for 10 min. The quality of PCR products was assessed through electrophoresis in a 1% agarose gel and staining with Gold View.

All PCR products were sequenced in both directions using the BigDye Terminator Sequencing Kit (Applied Bio Systems) and the ABI 3730XL Genetic Analyzer (PE Applied Biosystems, San Francisco, California USA) with two vector-specific primers and internal primers for primer walking.

Bioinformatic analysis. The complete mt genomes of A. deceptor and P. biconicus are deposited in GenBank with accession numbers KU925864 and KY230493, respectively. Sequence assembly was done using ContigExpress. As for the sequence analysis, the tRNAs were identified by tRNAscan-SE Search Server v. 1.2114, while for those tRNAs which could not be detected by this program we compared them with the corresponding tRNAs gene sequence of Neochauliodes punctatolous Liu & Yang15 to determine the position and sequence. The annotations of PCGs and rRNA genes were verified by hand alignment with closely-related species of Chauliodinae. The control region was identified afterwards by the boundary of the tRNA genes and compared with other insect mt genomes. Nucleotide substitution rates, base composition and codon usage were analyzed with MEGA 5.015. The GC and AT skews were measured using the following formulae: AT-skew = (A − T)/(A + T) and GC-skew = (G − C)/(G + C)16.

Phylogenetic analysis. Nine species of Megaloptera with determined mt genomes were included in the ingroup taxa and the outgroup taxa comprised two species of Neuroptera, namely Thyridosmylus langii (McLachlan) (Neuroptera: Osmylidae) and Rapisma xizangense (Neuroptera: Ithonidae) (Table 4). Sequences of 13 PCGs and two rRNAs were used in the present phylogenetic analysis. The PCGs were aligned based on the amino acid alignment using ClustalW in MEGA 5.015. RNA alignment was conducted by G-blocks Server (http://molevol.cmima.csic.es/castresana/Gblocks_server.html). Individual genes were concatenated by
SequenceMatrix v1.7.8. We performed maximum likelihood (ML) and Bayesian inference (BI) using the best-fit partitioning schemes recommended by PartitionFinder. For the ML analysis, we ran 1,000 bootstrap replicates and used the rapid bootstrap feature (random seed value 12345) in RAxML. MrBayes 3.2.3 was used to conduct the BI analysis with the GTR + I + G model as the optimal model selected by PartitionFinder. Two simultaneous runs of 2 million generations were conducted for the dataset, the tree samples were outputted every 1,000 generations with a burnin of 25%.

Divergence time estimation. Estimation of divergence times was conducted with all mt genome data using BEAST version 1.5.3. The taxa and data partitioning we used were consistent with the previous phylogenetic analysis using the GTR + I + G model, estimated base frequencies and Yule process of speciation. Minimum node constraints were assigned a normal prior distribution with standard deviations equal to 12 Ma.

Due to the difficulty of fossilization in habitats associated with fast-flowing water, there are scarce fossil records of Megaloptera. We set two fossil calibrations in our analysis (1) the mean age of Corydalidae + Sialidae was set at 185 Ma with the 95% credibility interval around the mean spanning the period from 204.7 to 165.3 Ma, reflecting the minimum age of these two families, which is based on oldest known fossil of Sialidae (Dobbertinia reticulata) from the Lower Jurassic of Dobbertin, Germany (~185 Ma); (2) the mean age of Chauliodinae + Corydalinae was set at 165 Ma with the 95% credibility interval around the mean spanning the period from 184.7 to 145.3 Ma based on the fossil evidence of an adult fishfly (Jurochauliodes ponomonenkoi) and Zhang from the Middle Jurassic of Inner Mongolia, China (~165 Ma) reported in Liu et al. Two independent MCMC analyses were run for 5 million generations under the uncorrelated lognormal relaxed clock model and sampled every 1000 generations. We combined tree files of both runs using LogCombiner 1.5.3, with the first 25% of the generations from each run discarded as burnin. Finally, we used TreeAnnotator 1.5.3 to calculate divergence time from a combined tree file. The phylogenetic tree was viewed and edited using FigTree 1.3.1.

References
1. Yang, D. & Liu, X. Y. Fauna Sinica Insecta Vol.51 Megaloptera (Science Press, 2010).
2. Liu, X. Y., Liu, Y. N., Aspöck, H., Yang, D. & Aspöck, U. Homology of the genital sclerites of Megaloptera (Insecta: Neoptera) and their phylogenetic relevance. Syst. Entomol. 41, 256–286 (2016).
3. New, T. R. & Theischinger, G. Megaloptera (Alderflies, Dobsonflies). Handbuch der Zoologie 4, 1–97 (1993).
4. Liu, X. Y., Wang, Z. H., Wang, D. & Yang, D. Early evolution and historical biogeography of fishflies (Megaloptera: Chauliodinae): implications from a phylogeny combining fossil and extant taxa. PloS ONE 7(5), e34035 (2012).
5. Wang, Y. Y., Liu, X. Y., Winterton, S. L. & Yang, D. The first mitochondrial genome for the fishfly subfamily Chauliodinae and implications for the higher phylogeny of Megaloptera. PloS ONE 7(10), e47302 (2012).
6. Clary, D. & Wolstenholme, D. R. The mitochondrial DNA molecule of Drosophila yakuba: Nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22, 252–271 (1985).
7. Beckenbach, A. T. & Stewart, J. B. Insect mitochondrial genomics 3: the complete mitochondrial genome sequences of representatives from two neopteroid orders: a dobsonfly (order Megaloptera) and a giant lacewing and an owlfly (order Neoptera). Genome 52, 31–38 (2009).
8. Jiang, Y. L. et al. Complete mitochondrial genomes of two Oriental dobsonflies, Nonneuronus tokinenensis (van der Weele) and Neurromus medius (Narits) (Megaloptera: Corydaliae), and phylogenetic implications of Corydalia. Zootaxa 3961(1), 044–062 (2015).
9. Li, H., Gao, J. Y., Liu, H. Y., Liang, A. P. & Cai, W. Z. The architecture and complete sequence of mitochondrial genome of an assassin bug Agriopodus dohrni (Hemiptera: Reduviidae). Int. J. Biol. Sci. 7(6), 792–804 (2011).
10. Dowton, M., Castro, L. R. & Austin, A. D. Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome ‘morphology’. Invertebr. Syst. 16, 345–356 (2002).
11. Wang, Y. Y. et al. Mitochondrial phylogenomics illuminates the evolutionary history of Neoptera. Cladistics. 33, 1–20 (2017).
12. Winterton, S. L., Hardy, N. B. & Wiegemann, B. M. On wings of face: phylogeny and Bayesian divergence time estimates of Neopterida (Insecta) based on morphological and molecular data. Syst. Entomol. 35, 349–378 (2010).
13. Sammartin, L. & Ronquist, F. Southern hemisphere biogeography inferred by event-based models: Plant versus animal patterns. Syst. Biol. 53, 216–243 (2004).
14. Lowe, T. M. & Eddy, S. R. RNAxscan–SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964 (1997).
15. Tamura, K. et al. MEGAS: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and parsimony methods. Mol. Biol. Evol. 28, 2731–2739 (2011).
16. 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).
17. Vaidya, G., Lohman, D. J. & Meier, R. SequenceMatrix: concatenation software for the fast assembly of multi-genome datasets with character set and codon information. Cladistics. 27, 171–180 (2010).
18. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analysis. Mol. Biol. Evol. 29, 1695–1701 (2012).
19. Stamatakis, A., Handley, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57, 758–771 (2008).
20. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574 (2003).
21. Drummond, A. & Ramult, A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214 (2007).
22. Ponomarenko, A. Corydalidae (Megaloptera) from Cretaceous deposits of northern Asia. Entomol. Obozr. 55, 425–433 (1976).
23. Ansorge, J. Dobbertinia reticulata Handlirsch 1920 from the Lower Jurassic of Dobbertin (Mecklenburg/Germany) – the oldest representative of Sialidae (Megaloptera). N. Jb. Geol. Paläont. Mh. 55, 3–56 (2001).
24. Rambaut, A. FigTree version 1.3.1. 1. Computer program distributed by the author Available: http://tree.bio.ed.ac.uk/software/figtree/ (Accessed 2011 Jan 4).
25. Grant, J. R. & Stothard, P. The Cytogenetic Database: a comparative genomics tool for circular genomes. Nucleic Acids Res. 36, W181–W184 (2008).

Acknowledgements
We express our sincere thanks to S.L. Winterton, J.S. Bartlett and H. Karube for collecting the studied specimens, and to J.D. Plant for linguistic improvement of the manuscript. This research was supported by the National Natural Science Foundation of China (Nos 41271063, 31672322, 31322051), the Beijing Natural Science Foundation (No. 5162016) and Chinese Universities Scientific Fund (No. 2017TC031).
Author Contributions
F.Y., Y.L.J., D.Y. and X.Y.L. participated in molecular studies, X.Y.L. designed the program, F.Y. and X.Y.L. prepared the manuscript.

Additional Information
Supplementary information accompanies this paper at doi:10.1038/s41598-017-04799-y

Competing Interests: The authors declare that they have no competing interests.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017