

OPEN

# The mitochondrial genomes of palaeopteran insects and insights into the early insect relationships

Nan Song<sup>1\*</sup>, Xinxin Li<sup>1</sup>, Xinming Yin<sup>1</sup>, Xinghao Li<sup>1</sup>, Jian Yin<sup>2</sup> & Pengliang Pan<sup>2</sup>

Phylogenetic relationships of basal insects remain a matter of discussion. In particular, the relationships among Ephemeroptera, Odonata and Neoptera are the focus of debate. In this study, we used a next-generation sequencing approach to reconstruct new mitochondrial genomes (mitogenomes) from 18 species of basal insects, including six representatives of Ephemeroptera and 11 of Odonata, plus one species belonging to Zygentoma. We then compared the structures of the newly sequenced mitogenomes. A tRNA gene cluster of IMQM was found in three ephemeropteran species, which may serve as a potential synapomorphy for the family Heptageniidae. Combined with published insect mitogenome sequences, we constructed a data matrix with all 37 mitochondrial genes of 85 taxa, which had a sampling concentrating on the palaeopteran lineages. Phylogenetic analyses were performed based on various data coding schemes, using maximum likelihood and Bayesian inferences under different models of sequence evolution. Our results generally recovered Zygentoma as a monophyletic group, which formed a sister group to Pterygota. This confirmed the relatively primitive position of Zygentoma to Ephemeroptera, Odonata and Neoptera. Analyses using site-heterogeneous CAT-GTR model strongly supported the Palaeoptera clade, with the monophyletic Ephemeroptera being sister to the monophyletic Odonata. In addition, a sister group relationship between Palaeoptera and Neoptera was supported by the current mitogenomic data.

The acquisition of wings and of ability of flight contribute to the success of insects in the planet. The origin of insect wings and flight remain contentious. Yet, it is generally agreed that wingless species comprising the subclass Apterygota, including bristletails (Archaeognatha) and silverfish (Zygentoma), constitute the most primitive lineages in Insecta. Winged insects, namely the subclass Pterygota occur in the next stage. According to the character of whether insects can fold back the wings over the abdomen, the Pterygota are subdivided into Palaeoptera and Neoptera. The Palaeoptera includes two extant orders: Odonata (dragonflies and damselflies) and Ephemeroptera (mayflies). All other winged insects form the Neoptera. Although the monophyly of the Pterygota is well established, the interrelationships among basal orders are ambiguous. Determining the relationship of Apterygota with Pterygota and the relationship of palaeopteran orders with regard to Neoptera are the keys to understanding evolution of flight and insect diversification.

Members of two wingless insect orders Archaeognatha and Zygentoma are traditionally classified in the order “Thysanura”<sup>1,2</sup>, based on the external morphological similarities. However, the Zygentoma was found to have a closer relationship with winged insects than has the Archaeognatha<sup>3–8</sup>. The mouthpart morphology supports the Dicondylia hypothesis, namely the clade Zygentoma + Pterygota. Probable synapomorphies of the Zygentoma and Pterygota were summarized by Kristensen<sup>1</sup>. Recent molecular studies also supported Zygentoma as a sister group to Pterygota<sup>9,10</sup>. The monophyly of Zygentoma was questioned because of the outside position of the *Tricholepidion*<sup>5,11</sup>. The relic silverfish *Tricholepidion gertschi* is the single extant species of the family Lepidotrichidae. Some morphological<sup>5</sup> and molecular<sup>11</sup> studies placed this species as the sister group to all other Dicondylia, while others supported a monophyletic Zygentoma<sup>9,12–14</sup>.

The issue on the palaeopteran relationships remains one of the major open questions in the field of insect evolution and systematics. Odonata and Ephemeroptera are the earliest-diverging lineages of winged insects. The species belonging to both orders are unable to fold their wings horizontally over their abdomen as most of neopteran insects. This character has been suggested as an evidence for the Palaeoptera monophyly. Hennig (1969)<sup>3</sup> proposed some possible morphological synapomorphies of Ephemeroptera and Odonata. However, most

<sup>1</sup>College of Plant Protection, Henan Agricultural University, Zhengzhou, 450002, China. <sup>2</sup>Xinyang Agriculture and Forestry University, Xinyang, 464001, China. \*email: [songnan@henau.edu.cn](mailto:songnan@henau.edu.cn)

of traits described by Hennig (1969)<sup>3</sup> were considered as plesiomorphic or convergent<sup>1,15</sup>. A recent morphological study based on the head characters supported a monophyletic Palaeoptera<sup>16</sup>. In addition, a molecular study using genome-scale data<sup>10</sup> supported Palaeoptera (Ephemeroptera + Odonata) and a sister group relationship of Palaeoptera to Neoptera, though with limited taxonomic sampling for palaeopteran insects.

Besides the Palaeoptera hypothesis, two other alternative hypotheses has been proposed by various authors. The hypothesis of a sister-group relationship between Odonata and Neoptera (the Metapterygota hypothesis or the basal Ephemeroptera hypothesis) was proposed based on the morphology of the wing veins, the mandibles and the respiratory system<sup>4,17–21</sup>. Several studies using combined molecular and morphological data also supported the basal Ephemeroptera hypothesis<sup>6,22</sup>. In contrast, a sister group relationship between the Odonata and the Ephemeroptera + Neoptera (the Chiasmomyaria hypothesis or the basal Odonata hypothesis) was preferred by Boudreaux<sup>5</sup>, based primarily on the character of direct sperm transfer. Kristensen (1981)<sup>1</sup> criticized Boudreaux's characters as homoplastic. Nevertheless, several analyses based on the single molecular markers supported the basal Odonata hypothesis<sup>12,23–25</sup>. In addition, an analysis using data from expressed sequence tag<sup>26</sup> recovered the relationship of (Odonata + (Ephemeroptera + Neoptera)), though including a single representative from each of Ephemeroptera and Odonata.

The rapid technological advance in molecular sequencing has led to acquisition of large amounts of sequence data in a very cost-effective way. At the same time, the development of assembly algorithms allows for rapidly reconstructing organelle genomes from next-generation sequencing data. Mitochondrial genomes (mitogenomes) as a class of organelle genome data are more easily to be assembled, annotated and to be more suitable for much larger-scale sampling, compared with the whole genome data. Mitochondrial phylogenomic analyses have been used to estimate the phylogenies of basal insects<sup>27–36</sup>. As of January 2019, there are only three mitogenome sequences available for *Zygentoma* in GenBank, 20 for Ephemeroptera and 27 for Odonata hinting at the need for further exploration of the mitogenomic approach in the basal insect groups.

In the present study, we sequenced a nearly complete mitogenome of *Thermobia* (*Zygentoma*: Lepismatidae), six partial or nearly complete mitogenomes of Ephemeroptera and 11 of Odonata, to add evidence to the controversy. Combined with published mitogenome sequences, we investigated the phylogenetic relationships of basal insects, with particular emphasis on the Palaeoptera problem.

## Materials and Methods

**Ethics statement.** No specific permits were required for the insect specimen collection in this study. The specimens were collected in Santan National Forest Park, Guangshui, Hubei province, China (31.86 °N, 113.94 °E). For each species newly sequenced, 1–2 adult individuals were collected. All samples were stored in 95–100% ethanol. Voucher specimens and specimen parts after DNA extraction have been deposited at –20 °C in Entomological Museum of Henan Agricultural University.

The field studies did not involve endangered or protected species. All sequenced insects are common species in China, and are not included in the “List of Protected Animals in China”.

**DNA extraction.** Total genomic DNA was extracted from the thoracic muscle tissue of each individual sample, with the TIANamp Genomic DNA Kit (TIANGEN BIOTECH CO., LTD) following the manufacturer's protocol. Purified DNA was eluted in a single step in 50 µl Buffer TE. The concentration of extracted genomic DNA was measured by Nucleic acid protein analyzer (QUAWELL TECHNOLOGY INC.), and the average values for each species determined were shown in Table S1.

**Library construction and high throughput sequencing.** Genomic DNA for each sample was pooled into twelve different libraries, respectively. Approximately equimolar amounts of genomic DNA for other insect species (*ca.* 20 different species) were mixed into the library. Each pool was designed to include distant taxonomic species in order to reduce the risk of a “contamination” and/or reads assignment errors in the following steps. For library preparation, Illumina TruSeq™ DNA Sample Prep Kit (Illumina, San Diego, CA, USA) was employed, with an average insert size of 350 bp. The indexed libraries were directly sequenced on a HiSeq X Ten platform (Beijing Novogene Bioinformatics Technology Co., Ltd, China), with 150 bp pair-ended reads.

**Reads filtering and de novo assembly.** Raw reads were filtered using NGS QC Toolkit with default settings<sup>37</sup>. The reads containing adapters and poly-N, and low quality reads were removed. At the same time, Q20, Q30, GC-content and sequence duplication level of the cleaned data were calculated. All subsequent genome assembly were based on clean data with high quality (avg. Q20 > 90%, and avg. Q30 > 80%). *De novo* assembly for the high-quality clean reads were performed using IDBA-UD v. 1.1.1<sup>38</sup>. The assemblies were constructed using 200 for the setting of minimum size of contig, and an initial k-mer size of 40, an iteration size of 10, and a maximum k-mer size of 90.

**Mitogenome reconstruction and annotation.** Mitogenome reconstruction method mostly followed a bioinformatics pipeline of Gillett *et al.*<sup>39</sup>. The mitochondrial baiting sequences (i.e., *cox1*, *cob* and *12S*) were amplified and pre-sequenced by the primers designed by Song *et al.*<sup>40</sup>. The mitochondrial scaffolds were identified by blasting the mitochondrial baitings against a local database constructed by BioEdit<sup>41</sup>. The initial mitogenome annotations were conducted using the MITOS<sup>42</sup>, under default settings and the invertebrate genetic code for mitochondria. The gene boundaries were further checked and refined by alignment with homologous sequences of related species (see details in Table S1) in MEGA 7<sup>43</sup>. Mappings to the mitochondrial contigs were performed using BWA v. 0.7.5<sup>44</sup>. Alignments produced in SAM format were converted to sorted BAM format by SAMtools v. 0.1.19<sup>45</sup>. Statistics for nucleotide coverage were generated with Qualimap v.2.2.1<sup>46</sup>.

Representative specimens were identified to species or genus level by checking adult morphological characters, and by Blast matches to *cox1* records from the BOLD database (<http://www.boldsystems.org/>) and NCBI

Gene partitions	NumOTU	Iss	Iss.cSym	Psym	Iss.cAsym	Pasym
PCG1 <sub>NT</sub>	32	0.516	0.808	0.000	0.554	0.000
PCG2 <sub>NT</sub>	32	0.371	0.808	0.000	0.554	0.000
PCG3 <sub>NT</sub>	32	0.897	0.808	0.000	0.554	0.000
rRNA	32	0.703	0.790	0.000	0.520	0.000
tRNA	32	0.689	0.793	0.000	0.524	0.000

**Table 1.** Saturation test based on the datasets of 85taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>RNA. Note: two-tailed tests are used.

Dataset	Reference model	Model used	Cross-validation score	Standard deviation
85taxa_PCG1 <sub>NT</sub> 2 <sub>NT</sub> 3 <sub>RY</sub> RNA	GTR	CAT-GTR	1860.73	±106.882
85taxa_PCG1 <sub>NT</sub> 2 <sub>NT</sub> RNA	GTR	CAT-GTR	1864.43	±163.653

**Table 2.** Cross-validation analyses of the homogeneous and heterogeneous models implemented in PhyloBayes based on nucleotide datasets. Positive scores: better than reference model (GTR).

Genbank (<http://www.ncbi.nlm.nih.gov/genbank/>). The detailed classification information, voucher numbers of species sequenced and accession numbers of the new mitogenome sequences are given in Table S1.

**Multiple sequence alignments.** Each mitochondrial protein-coding gene was aligned separately based on the corresponding amino acid translations using the MUSCLE algorithm<sup>47</sup>, as implemented in TranslatorX<sup>48</sup>. All protein-coding gene alignments were concatenated by using FASconCAT-G<sup>49</sup> to construct the dataset of PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>. The yn00 program of the PAML package<sup>50</sup> was used to calculate the nonsynonymous (*dN*) and synonymous (*dS*) substitution rates of the concatenated 13 protein-coding genes, with the method of Yang and Nielsen<sup>51</sup>. DAMBE 7<sup>52</sup> was used to conduct tests for substitution saturation of each codon position. According to the index of substitution saturation, the third codon positions were significantly saturated (*Iss* > *Iss.cSym* and *Iss* > *Iss.cAsym*) (Table 1). To account for the effect of substitution saturation, two approaches were employed in the further phylogenetic analyses. First, the protein-coding genes were re-concatenated with FASconCAT-G under the parameter option of 3rd sequence position exclusion in order to create a dataset with codon positions 1 and 2 (PCG1<sub>NT</sub>2<sub>NT</sub>). Second, the protein-coding genes were re-concatenated with FASconCAT-G under the parameter option of RY-coding of 3rd sequence positions to compile a dataset with the third codon position nucleotides recoded into two state categories, R (purine) and Y (pyrimidine)<sup>53</sup> (PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>).

Ribosomal and transfer RNA genes were aligned individually with MAFFT<sup>54</sup> under the iterative refinement method incorporating the most accurate local pairwise alignment information (E-INS-i). Gaps of alignments were striped by Gap Strip/Squeeze v2.1.0 with 40% Gap tolerance (<http://www.hiv.lanl.gov/content/sequence/GAPSTREEZE/gap.html>). The dataset PCG1<sub>NT</sub>2<sub>NT</sub> was concatenated with the tRNA and rRNA datasets in order to create a combined dataset of PCG1<sub>NT</sub>2<sub>NT</sub>RNA. The PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub> dataset, tRNA dataset and rRNA dataset were combined together to construct the PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>RNA dataset. Multiple sequence alignments were statistically scored using AliStat<sup>10</sup>. Nucleotide compositions of the mitogenome sequences were estimated using MEGA 7<sup>43</sup>.

**Phylogenetic inference.** A total of eighty-five species were included to create the full taxon dataset. Of which, three species representing three families in Zygentoma, eighteen representing three families in Ephemeroptera and 31 representing 17 families in Odonata were used in the phylogenetic analysis. In addition, six species of Archaeognatha and 20 species representing 17 orders of Neoptera were also included in the analysis. Two species of Collembola and five of Diplura were selected as outgroups.

Phylogenetic trees were reconstructed using both maximum likelihood (ML) and Bayesian inferences (BI). Partitioned ML analyses were performed with the IQ-TREE<sup>55</sup>, as implemented in the CIPRES Science Gateway<sup>56</sup>. The 13 protein-coding genes were partitioned by gene, whereas the 22 tRNA genes and the two rRNA genes were considered as two separate partitions. The best partitioning schemes (Table S2) for the datasets were selected with PartitionFinder 2<sup>57</sup>. The site-homogeneous GTR model was often chosen as the best-fit model for each partition in every dataset (Table S2). We performed 10,000 ultrafast<sup>58</sup> bootstrap replicates to investigate nodal support across the topology.

Ten-fold Bayesian cross-validation analyses were performed to test the fit of the site-heterogeneous mixture model CAT-GTR and the site-homogeneous model GTR to our full taxon data (85taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>RNA and 85taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>RNA) using PhyloBayes 3.3f<sup>59</sup>. The results showed that the CAT-GTR model was the best fitting model for both datasets (Table 2). BI analyses were conducted with PhyloBayes MPI<sup>60,61</sup> as implemented in the CIPRES Science Gateway<sup>56</sup>, under the CAT-GTR model. For each analysis, two Markov chain Monte Carlo (MCMC) chains were run in parallel, after the removal of constant sites from the alignments. The minimum number of cycles were set to 20,000. Stationarity was considered to be reached when the maxdiff was <0.3 and minimum effective size was >50. The run would be terminated if analysis passed the convergence test. The first 1000 trees of each MCMC were treated as the burn-in, and the majority-rule consensus tree was calculated from the saved trees.

Order	Name	Length	Mapped bases	Mean coverage	Mapped reads
Ephemeroptera	<i>Cloeon dipterum</i>	5,540	50,549	9	337
Ephemeroptera	<i>Epeorus</i> sp.	15,355	1,599,545	104	10,664
Ephemeroptera	<i>Rhithrogena</i> sp.	15,277	2,434,743	159	16,232
Ephemeroptera	<i>Parafronurus</i> sp.	15,214	9,525,907	626	63,510
Ephemeroptera	<i>Isonychia</i> sp.	15,830	35,145,298	2,220	468,740
Ephemeroptera	<i>Ephemera</i> sp.	15,683	42,106,306	2,685	561,487
Odonata	<i>Orthetrum albistylum</i>	10,296	382,641	37	2,551
Odonata	<i>Orthetrum melania</i>	8,609	320,544	37	2,137
Odonata	<i>Ischnura elegans</i>	14,990	1,075,720	72	7,172
Odonata	<i>Paracercion malayanum</i>	15,366	815,970	53	5,446
Odonata	<i>Agriocnemis femina</i>	13,603	1,157,744	85	7,719
Odonata	<i>Platycnemis phyllopoda</i>	15,125	2,444,403	162	16,297
Odonata	<i>Coeliccia cyanomelas</i>	15,100	2,869,718	190	19,133
Odonata	<i>Sympetrum eroticum</i>	14,940	6,600,519	442	44,012
Odonata	<i>Mesopodagrion tibetanum</i>	15,170	10,327,869	681	68,857
Odonata	<i>Anotogaster sieboldii</i>	15,132	10,525,046	696	70,170
Odonata	<i>Mnais tenuis</i>	14,632	10,687,092	730	142,504
Zygentoma	<i>Thermobia</i> sp.	16,586	13,047,410	787	87,090

**Table 3.** The statistics of each mitogenomic contig assembled in this study.

The preliminary full taxon (85 taxa) data ML trees were used for the RogueNaRok<sup>62</sup> analysis, which can identify the taxa being assumed to show uncertain phylogenetic position. The result suggested *Epiophlebia superstes* (Odonata) as a rogue taxon leading to less accurate overall phylogenetic reconstructions. In addition, both ML trees displayed the obviously long branch lengths leading to the outgroup taxa from the Collembola (*Bilobella aurantiaca*, *Cryptopygus antarcticus*) and Diplura (*Lepidocampa weberi*, *Campodea fragilis*, *Campodea lubbocki*). To reduce the potential effect of problematic taxa on the recovered topology, we created the reduced taxon (79 taxa) data which excluded five outgroup taxa mentioned above and the rogue species identified by RogueNaRok. Both ML and Bayesian analyses were repeated based on the reduced taxon datasets (79taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>RNA and 79taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>RNA), with the settings as the analyses from the full taxon data.

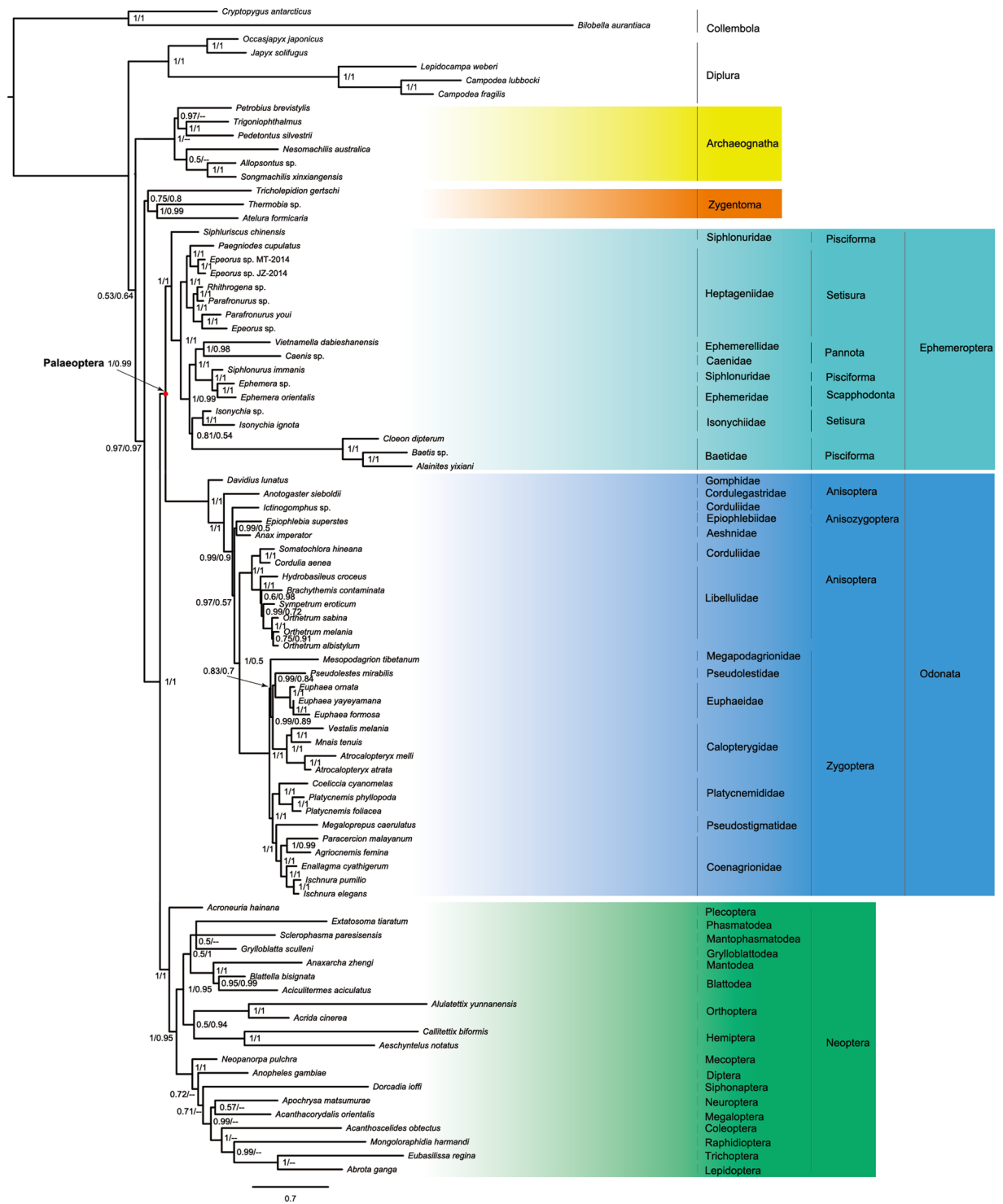
## Results

**Genome sequencing.** Eighteen partial or nearly complete mitogenomes were newly determined for 11 species of Odonata (four dragonflies and seven damselflies), six species of Ephemeroptera and one species of Zygentoma by using a next-generation sequencing method. After filtering, the number of Illumina reads obtained varied from 57,181,127 reads to 83,346,749 reads. In most cases, 0.01% to 0.09% corresponded to mitochondrial reads. Through the BLAST-searches with baiting sequences, the mitogenome identified for each newly sequenced species was assembled on a single contig. The analyzed mitogenomes are around 15,000 nt in size, except for *Agriocnemis femina*, *Orthetrum albistylum*, *Orthetrum melania* and *Cloeon dipterum*. The four species are exhibiting a size ranging 5,540~13,603 nt. The sequencing coverage for each mitochondrial contig varied from 9-fold for *C. dipterum* to 2,685-fold for *Ephemera* sp. (Table 3).

In the fourteen nearly complete mitogenomes (14,632 nt ~ 16,586 nt), we were able to determine their gene arrangement (Fig. 1). The mitochondrial gene order was conserved inside of these newly sequenced species, with the same gene organization as the ancestral insect<sup>63</sup>. A total of 35~37 mitochondrial genes were identified in each nearly complete mitogenome (Fig. 1). The *trnF* was missing in the species of *Coeliccia cyanomelas*, *Parafronurus* sp. and *Rhithrogena* sp. The *trnG* is missing in *Anotogaster sieboldii* and *Sympetrum eroticum*. The *trnP* in *Sympetrum eroticum*, the *trnL(tag)* in *Mnais tenuis* and the *trnS(tga)* in *Platycnemis phyllopoda* were not detected. For the three partial mitogenome sequences with the length of 8,528 nt (*O. melania*), 10,107 nt (*O. albistylum*) and 13,280 nt (*A. femina*), the missing genes were mainly located adjacent to the putative control region. In the species of *C. dipterum*, only nineteen mitochondrial genes were detected (Fig. 1), with a total length of 5,540 nt. Failure to reconstruct the mitogenome of this species may be owing to the lower sequencing depth (Table 3). The mitogenome of the *Thermobia* sp. (Zygentoma) has the AT content of 68% (Table 4). Among the analyzed palaeopteran species, the mean AT content for the new mitogenomes of six mayflies is also 68%, which is lower than the new mitogenomes of four dragonflies (avg. 71%) and of seven damselflies (avg. 72%).

**Phylogenetic analyses.** The full taxon protein-coding gene dataset including all codon positions (PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>NT</sub>) comprises 11,070 nucleotides. Completeness score (Ca) calculated by AliStat for the alignment PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>NT</sub> was 0.9624. The substantial missing data occurred in the species of *C. dipterum*. The alignment PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>RNA comprises approximately 15,218 base positions (Ca = 0.9423). After excluding the third positions, the reduced site matrix PCG1<sub>NT</sub>2<sub>NT</sub>RNA contains 11,528 nucleotide positions (Ca = 0.9359).

The rates of nonsynonymous substitutions (*dN*) ranged from 0.1652 (Mecoptera) to 0.2572 (Collembola), and the rates of synonymous substitutions (*dS*) ranged from 4.5322 (Mantodea) to 4.7947 (Plecoptera) (Table 4). Both Odonata and Ephemeroptera had the relatively slow nonsynonymous substitution rates (Odonata: 0.1668



**Figure 1.** Organization of mitochondrial genomes newly sequenced in this study. The genes above the line indicate the transcriptional direction being from left to right, while those below the line indicate the transcriptional direction being from right to left.

and Ephemeroptera: 0.1791) and the lower ratios of nonsynonymous/synonymous substitution rates (Odonata: 0.0357 and Ephemeroptera: 0.0377).

Partitioned ML analyses (IQ-TRE, GTR model) and Bayesian inference (PhyloBayes, CAT-GTR model) presented the conflict hypotheses. But data treatment methods (excluding or RY-coding the third positions) had no significant effect on tree reconstructions, under the same inference method.

Phylogenetic trees from ML analyses (Fig. S1) recovered Odonata as a sister group to Neoptera, but with low-to-moderate support (bootstrap support, BP  $\leq 77$ ). The Ephemeroptera emerged as monophyletic and as

Order	avg. AT	avg. CG	dN	dS	dN/dS
Collembola	69.47	30.53	0.2572	4.7292	0.0544
Diptera	76.57	23.43	0.1685	4.5708	0.0369
Archaeognatha	69.72	30.28	0.1981	4.7253	0.0419
Zygentoma	65.47	34.53	0.1992	4.7625	0.0418
Ephemeroptera	65.67	34.33	0.1791	4.7496	0.0377
Odonata	70.85	29.15	0.1668	4.6642	0.0357
Plecoptera	61.71	38.29	0.1825	4.7947	0.0381
Phasmatodea	75.46	24.54	0.2358	4.6031	0.0512
Mantophasmatodea	74.34	25.66	0.1914	4.6266	0.0414
Grylloblattodea	70.26	29.74	0.1858	4.7594	0.0390
Mantodea	77.22	22.78	0.2028	4.5322	0.0447
Blattodea	71.01	28.99	0.1822	4.6726	0.0390
Orthoptera	74.94	25.06	0.2175	4.6027	0.0472
Hemiptera	75.69	24.31	0.2337	4.6008	0.0508
Mecoptera	76.00	24.00	0.1652	4.5584	0.0362
Diplura	67.12	32.88	0.2160	4.7141	0.0458
Siphonaptera	78.32	21.68	0.1997	4.5806	0.0436
Neuroptera	76.66	23.34	0.1786	4.5539	0.0392
Megaloptera	75.86	24.14	0.1751	4.6050	0.0380
Coleoptera	75.48	24.52	0.2099	4.6733	0.0449
Raphidioptera	78.83	21.17	0.2123	4.5821	0.0463
Trichoptera	77.80	22.20	0.2253	4.6471	0.0485
Lepidoptera	80.54	19.46	0.1999	4.5499	0.0439

**Table 4.** Nucleotide compositions and substitution rates estimated for major lineages.

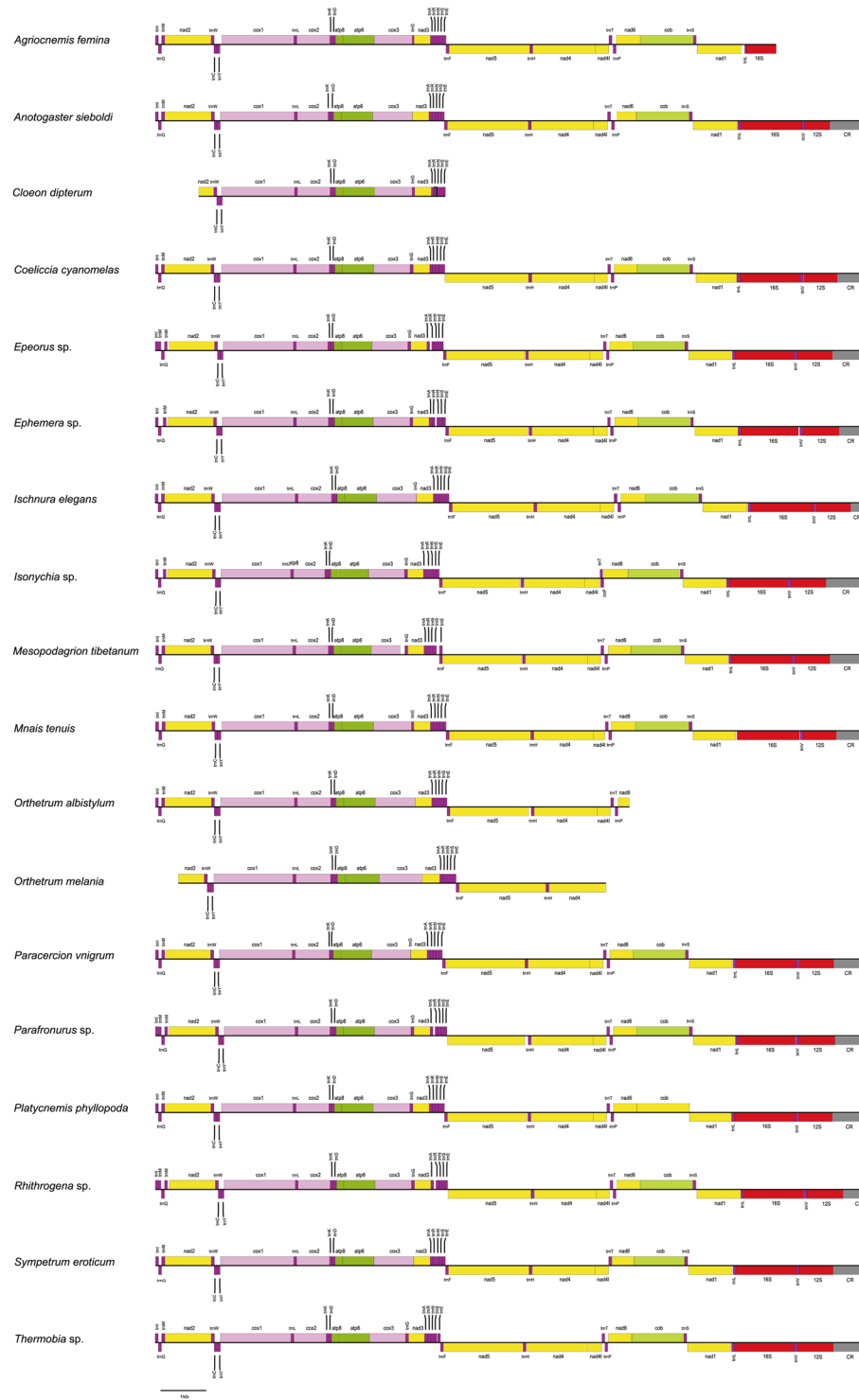
sister group to the single representative of Plecoptera (*Acroneuria hainana*). The Ephemeroptera + Plecoptera clade was the sister group to the remaining pterygote orders. Thus, ML trees substantially supported the basal Ephemeroptera hypothesis. In addition, the following higher-level relationships were consistent across the ML analyses: the monophyly of the Zygentoma was well supported (BP  $\geq$  98) and *T. gertschi* was retrieved as the first clade within this order. The Zygentoma was placed as sister group of all the winged insects, rendering the Apterygota paraphyletic. In the current data matrices, the taxon sampling focused on the Ephemeroptera and Odonata. Thus, the monophyly of some higher-level lineages in both orders can be tested. At the suborder level, the Setisura, Scaphodonta and Pannota were monophyletic within Ephemeroptera. However, the Pisciforma was paraphyletic with respect to *Siphuriscus chinensis*. Within Odonata, the clade Anisoptera including the single representative of Anisozygoptera (*Epiophlebia superstes*) formed a strongly supported sister group relationship to the monophyletic Zygoptera (BP = 100). The reduced taxon datasets (79taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>RNA and 79taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>RNA) resulted in a largely identical ingroup tree topology (Fig. S2) to those inferred from the full taxon datasets.

In contrast, Bayesian analyses under the CAT-GTR model consistently supported the Palaeoptera hypothesis (Fig. 2), with strong nodal support values (posterior probabilities, PP = 1 or PP = 0.99). The monophyly of Zygentoma was recovered in Bayesian analyses, but with lower support values (PP < 0.9). In addition, the subclass Dicondylia was supported because of the sister group relationship between the Zygentoma and pterygote insects (PP = 0.97). The Plecoptera formed a sister group of all other Neoptera. Therefore, the monophyly of Neoptera were supported by the Bayesian inference under the site-heterogeneous CAT-GTR model. Removal of the long-branched outgroup taxa and the rogue species had no significant influence on the ingroup topology, except for the Zygentoma. In the Bayesian analyses with the reduced taxon datasets, the Zygentoma was recovered as paraphyletic with respect to *T. gertschi* (Fig. S3).

## Discussion

**Characteristics of the new mitogenomes.** The average AT content of the six newly sequenced mayflies is 68%, which is slightly higher than the average AT content of the whole Ephemeroptera (65.67%). Yet it is still the lowest proportion compared to other insect orders included (Table 4). This result is in agreement with the previous studies<sup>29,36,64,65</sup>. For the newly determined mayflies, there is a tRNA gene cluster of IMQM found in three species, namely, *Parafronurus* sp., *Rhithrogena* sp. and *Epeorus* sp. (Fig. 1). The three species belong to the family Heptageniidae. All previously published mitogenomes from Heptageniidae, except for *Paegniodes cupulatus*, exhibit the arrangement of IMQM<sup>29,36,66,67</sup>. In contrast, the mitogenomes from other families in Ephemeroptera<sup>27,65,67–72</sup> have the ancestral gene order, with the typical IQM tRNA arrangement. The gene rearrangement with IMQM tRNA cluster may serve as a potential synapomorphy for the family Heptageniidae. Further studies needs expanding mitogenomic taxon sampling from Ephemeroptera to affirm this point.

**Relationships among the basal winged insects.** Some previous studies have attempted to resolve the relationships among the basal winged insects and to address the Palaeoptera problem, based on the data from



**Figure 2.** Phylogenetic tree based on Bayesian inference of the nucleotide sequence dataset of 85taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>RNA, under the CAT + GTR model. Values at nodes are Bayesian posterior probability support (Left: the values from dataset of 85taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>3<sub>RY</sub>RNA, right: the values from dataset of 85taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>RNA). “-” indicates the node not being recovered by the dataset of 85taxa\_PCG1<sub>NT</sub>2<sub>NT</sub>RNA. Scale bar represents substitutions/site. The meaning of color is as follows: yellow, Archaeognatha; orange, Archaeognatha; wathet, Ephemeroptera; blue, Odonata; green, Neoptera.

molecular sequences<sup>7,8,10,12,26-28,31,36,68,70,73</sup> and/or morphological characters<sup>9,16</sup>. However, conflicting results were obtained due to limited taxon sampling and various analysis methods (Table S3).

In our analyses, some taxa exist as long branches, for example, the outgroup taxa from Collembola. The tree with long branches may be problematic for accurate estimation of the phylogenetic relationships. The long branch

attraction (LBA) artefact<sup>74–76</sup> is a common phenomenon occurred in the tree reconstructions, where unrelated species can be grouped together artifactually due to the shared long branch lengths. The mitogenome sequences from some insect lineages show the specific evolution rate and the distinct base composition<sup>40,77–79</sup>. Both factors may have a negative impact on the phylogenetic reconstruction from the mitogenomic data. To assess the effect of taxon sampling on our results, we removed the long-branched outgroups and the rogue species (namely that with unstable placement across the trees). However, the results gave rise to the virtually same topologies when analyzing the reduced sequence data under the same inference method.

Two previous study based on the complete mitogenomes<sup>27,36</sup> supported the basal Ephemeroptera hypothesis. Nevertheless, other mitogenomic studies<sup>31,70,73</sup> with various taxon sampling supported the basal position of Odonata. Therefore, the previous studies based on the complete mitogenomes yielded conflict results for the phylogenetic relationships among the most basal extant pterygote lineages. In this study, with the site-homogeneous GTR model, all ML trees recovered a sister group relationship between Odonata and the majority of Neoptera, with weak to moderate support. An abnormal placement of Plecoptera were retained in the ML trees, where the Plecoptera was placed far away from other neopteran insects and appeared a sister group to the monophyletic Ephemeroptera. Both Plecoptera and Ephemeroptera were sister to all the remaining pterygote orders. This arrangement is similar to the basal Ephemeroptera hypothesis, and is in agreement with the studies by Zhang *et al.*<sup>36</sup> and Cai *et al.*<sup>27</sup>.

On the current mitogenomic data, the Bayesian inference analyses with the site-heterogeneous CAT-GTR model yielded topologies consistently supporting the Palaeoptera hypothesis, and suggested that the monophyletic Ephemeroptera is sister group to the monophyletic Odonata. In addition, Bayesian inferences resulted in a monophyletic Neoptera, in which the Plecoptera was sister to all other neopteran lineages (Figs. 2 and S3). Both the prior studies<sup>77,78</sup> and the cross-validation analyses conducted in this study indicated that the site-heterogeneous CAT-GTR model implemented in the PhyloBayes software is more fitting for modeling the evolution of insect mitogenomes than the site-homogeneous GTR model. Moreover, the site-heterogeneous CAT-GTR model has been shown to be least sensitive to long-branch attraction phenomena<sup>40,77–80</sup>. From a point of view of morphology, a prior study by Blanke *et al.*<sup>16</sup> have refuted the possibility of a sister relationship between Plecoptera and Ephemeroptera. All lineages with the sequences that branch near the base of the tree suffer from relatively low apparent substitution rates, which include Plecoptera and Ephemeroptera. The shared sequence similarity might contribute to a kind of convergence that could lead to an artificially deep branching position of the Plecoptera in the ML trees. Therefore, the phylogenetic results from the Bayesian inference analyses using the site-heterogeneous CAT-GTR model (Figs. 2 and S3) should be closer to the species tree.

The Palaeoptera hypothesis was first proposed by Martynow (1924)<sup>81</sup> and Crampton (1924)<sup>82</sup>. Support for a monophyletic origin of Odonata and Ephemeroptera includes various evidence from the morphological characteristics, namely, the shortened antennae<sup>3</sup>, aquatic larvae<sup>82</sup>, the distinct wing joint<sup>83–85</sup>, and a paired penis<sup>86</sup>. Especially, the monophyly of Palaeoptera was supported by head structures in a recent morphological study of Blanke *et al.*<sup>16</sup>. Several molecular studies using nuclear genes supported the Palaeoptera<sup>6,7,9,87–89</sup>. A more recent phylogenomic study also suggested that the Ephemeroptera and Odonata derived from a common ancestor<sup>10</sup>. This study is the first to provide the mitogenomic data supporting a sister group relationship between the monophyletic Ephemeroptera and Odonata. In addition, a sister group relationship of Palaeoptera and Neoptera is strongly supported. The results presented here would be expected to be confirmed by further studies with more extensive taxon sampling.

Received: 6 June 2019; Accepted: 13 November 2019;

Published online: 28 November 2019

## References

- Kristensen, N. P. Phylogeny of insect orders. *Annu. Rev. Entomol.* **26**, 135–157 (1981).
- Trautwein, M. D., Wiegmann, B. M., Beutel, R., Kjer, K. M. & Yeates, D. K. Advances in Insect Phylogeny at the Dawn of the Postgenomic Era. *Annu. Rev. Entomol.* **57**, 449–468 (2012).
- Hennig, W. Die Stammesgeschichte der Insekten. *Frankfurt am Main: Waldemar Kramer* **436** (1969).
- Kristensen, N. P. The phylogeny of hexapod “orders”. A critical review of recent accounts. *Z. Zool. Syst. Evolutionsforsch.* **13**, 1–44 (1975).
- Boudreaux, H. B. Arthropod phylogeny, with special reference to insects. *John Wiley & Sons, Inc.* (1979).
- Wheeler, W. C., Whiting, M., Wheeler, Q. D. & Carpenter, J. M. The phylogeny of the extant hexapod orders. *Cladistics* **17**, 113–169 (2001).
- Ishiwata, K., Sasaki, G., Ogawa, J., Miyata, T. & Su, Z. Phylogenetic relationships among insect orders based on three nuclear protein coding gene sequences. *Mol. Phylogenet. Evol.* **58**, 169–180 (2010).
- Regier, J. C. *et al.* Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079–1083 (2010).
- Kjer, K. M., Carle, F. L., Litman, J. E. S. S. E. & Ware, J. E. S. S. I. C. A. A molecular phylogeny of Hexapoda. *Arthropod. Syst. Phylo.* **64**, 35–44 (2006).
- Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
- Giribet, G., Edgecombe, G. D., Carpenter, J. M., D’Haese, C. A. & Wheeler, W. C. Is Ellipura monophyletic? A combined analysis of basal hexapod relationships with emphasis on the origin of insects. *Org. Divers. Evol.* **4**, 319–340 (2004).
- Misof, B. *et al.* Towards an 18S phylogeny of hexapods: accounting for group-specific character covariance in optimized mixed nucleotide/doublet models. *Zoology (Jena)*. **110**, 409–429 (2007).
- Carapelli, A., Liò, P., Nardi, F., van der Wath, E. & Frati, F. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of hexapoda and crustacea. *BMC Evol. Biol.* **7**, S8 (2007).
- Blanke, A., Koch, M., Wipfler, B., Wilde, F. & Misof, B. Head morphology of *Tricholepidion gertschi* indicates monophyletic Zygentoma. *Front. Zool.* **11**, 16 (2014).
- Ogden, T. H. & W., M. F. The problem with “the Paleoptera problem:” sense and sensitivity. *Cladistics* **19**, 432–442 (2003).
- Blanke, A. *et al.* Revival of Palaeoptera-head characters support a monophyletic origin of Odonata and Ephemeroptera (Insecta). *Cladistics* **28**, 560–581 (2012).



17. Beutel, R. G. & Gorb, S. N. A revised interpretation of the evolution of attachment structures in Hexapoda with special emphasis on Mantophasmatodea. *Arthropod Syst. Phylo.* **64**, 3–25 (2006).
18. Kristensen, N. P. Phylogeny of extant hexapods. *The insects of Australia* **2**, 125–140 (1991).
19. Pass, G. *et al.* Phylogenetic relationships of the orders of Hexapoda: contributions from the circulatory organs for a morphological data matrix. *Arthropod Syst. Phylogeny* **64**, 165–203 (2006).
20. Staniczek, A. H. The mandible of silverfish (Insecta: Zygentoma) and mayflies (Ephemeroptera): Its morphology and phylogenetic significance. *Zool. Anz.* **239**, 147–178 (2000).
21. Staniczek, A. H. Der Larvenkopf von *Oniscigaster wakefieldi* McLachlan, 1873:(Insecta: Ephemeroptera: Oniscigastridae); ein Beitrag zur vergleichenden Anatomie und Phylogenie der Eintagsfliegen (Doctoral dissertation, Verlag nicht ermittelbar) (2001).
22. Terry, M. D. & Whiting, M. F. Mantophasmatodea and phylogeny of the lower neopterous insects. *Cladistics* **21**, 240–258 (2005).
23. Kjer, K. Aligned 18S and insect phylogeny. *Syst. Biol.* **53**, 506–514 (2004).
24. Mallatt, J. & Giribet, G. Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. *Mol. Phylogenet. Evol.* **40**, 772–794 (2006).
25. von Reumont, B. M. *et al.* Can comprehensive background knowledge be incorporated into substitution models to improve phylogenetic analyses? A case study on major arthropod relationships. *BMC Evol. Biol.* **9**, 119 (2009).
26. Simon, S., Strauss, S., von Haeseler, A. & Hadrys, H. A phylogenomic approach to resolve the basal pterygote divergence. *Mol. Biol. Evol.* **26**, 2719–2730 (2009).
27. Cai, Y. Y. *et al.* The mitochondrial genome of *Caenis* sp.(Ephemeroptera: Caenidae) and the phylogeny of Ephemeroptera in Pterygota. *Mitochondrial. DNA B* **3**, 577–579 (2018).
28. Comandi, S., Carapelli, A., Podsiadlowski, L., Nardi, F. & Frati, F. The complete mitochondrial genome of *Atelura formicaria* (Hexapoda: Zygentoma) and the phylogenetic relationships of basal insects. *Gene* **439**, 25–34 (2009).
29. Gao, X. Y. *et al.* The complete mitochondrial genome of *Epeorus herklotsi* (Ephemeroptera: Heptageniidae) and its phylogeny. *Mitochondrial. DNA B* **3**, 303–304 (2018).
30. Kim, M. J., Jeong, S. Y., Wang, A. R., An, J. & Kim, I. Complete mitochondrial genome sequence of *Macromia daimoji* Okumura, 1949 (Odonata: Macromiidae). *Mitochondrial. DNA B* **3**, 365–367 (2018).
31. Lin, C. P., Chen, M. Y. & Huang, J. P. The complete mitochondrial genome and phylogenomics of a damselfly, *Euphaea formosa* support a basal Odonata within the Pterygota. *Gene* **468**, 20–29 (2010).
32. Lorenzo-Carballa, M. O., Tsubaki, Y., Plaistow, S. J. & Watts, P. C. The complete mitochondrial genome of the broad-winged damselfly *Mnais costalis* Selys (Odonata: Calopterygidae) obtained by next-generation sequencing. *Inter. J. Odonatol.* **19**, 191–198 (2016).
33. Wang, J. F. *et al.* Complete mitochondrial genome of an enigmatic dragonfly, *Epiophlebia superstes* (Odonata, Epiophlebiidae). *Mitochondrial. DNA* **26**, 718–719 (2015).
34. Yamauchi, M. M., Miya, M. U. & Nishida, M. Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. *Insect molecular biology* **13**, 435–442 (2004).
35. Yu, P., Cheng, X., Ma, Y., Yu, D. & Zhang, J. The complete mitochondrial genome of *Brachythemis contaminata* (Odonata: Libellulidae). *Mitochondrial DNA A* **27**, 2272–2273 (2016).
36. Zhang, J., Zhou, C., Gai, Y., Song, D. & Zhou, K. The complete mitochondrial genome of *Parafronurus youi* (Insecta: Ephemeroptera) and phylogenetic position of the Ephemeroptera. *Gene* **424**, 18–24 (2008).
37. Patel, R. K. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* **7**, e30619 (2012).
38. Peng, Y., Leung, H. C., Yiu, S. M. & Chin, F. Y. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
39. Gillett, C. P. *et al.* Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Mol. Biol. Evol.* **31**, 2223–2237 (2014).
40. Song, N., Li, H., Song, F. & Cai, W. Molecular phylogeny of Polyneoptera (Insecta) inferred from expanded mitogenomic data. *Sci. Rep.-UK.* **6**, 36175 (2016).
41. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series* **41**, 95–98 (1999).
42. Bernt, M. *et al.* MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).
43. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754–1760 (2009).
45. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
46. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multisample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
47. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
48. Abascal, F., Zardoya, R. & Telford, M. J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–W13 (2010).
49. Kuck, P. & Meusemann, K. FASconCAT: Convenient handling of data matrices. *Mol. Phylogenet. Evol.* **56**, 1115–1118 (2010).
50. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
51. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
52. Xia, X. H. DAMBE5: A Comprehensive Software Package for Data Analysis in Mol. Biol. Evol. *Mol. Biol. Evol.* **30**, 1720–1728 (2013).
53. Phillips, M., Delsuc, F. & Penny, D. Genome-scale phylogeny and the detection of systematic biases. *Mol. Biol. Evol.* **21**, 1455–1458 (2004).
54. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
55. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
56. Miller, M., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)* **14**, 1–8 (2010).
57. 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).
58. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2017).
59. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
60. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).

61. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
62. Aberer, A. J., Krompass, D. & Stamatakis, A. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst. Biol.* **62**, 162–166 (2013).
63. Cameron, S. L. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu. Rev. Entomol.* **59**, 95–117 (2014).
64. Lee, E. M. *et al.* The complete mitogenome sequences of the palaeopteran insects *Ephemera orientalis* (Ephemeroptera: Ephemeridae) and *Davidius lunatus* (Odonata: Gomphidae). *Genome* **52**, 810–817 (2009).
65. Zhou, D., Wang, Y. Y., Sun, J. Z. & Zhou, C. F. The complete mitochondrial genome of *Paegniodes cupulatus* (Ephemeroptera: Heptageniidae). *Mitochondrial DNA A.* **27**, 925–926 (2016).
66. Tang, M., Tan, M., Meng, G., Yang, S. & Su, X. Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Res.* **42**, e166 (2014).
67. Ye, Q. M. *et al.* The complete mitochondrial genome of *Isonychia kiangsinensis* (Ephemeroptera: Isonychiidae). *Mitochondrial DNA B* **3**, 541–542 (2018).
68. Rutschmann, S., Chen, P., Zhou, C. & Monaghan, M. T. Using mitochondrial genomes to infer phylogenetic relationships among the oldest extant winged insects (Palaeoptera). *bioRxiv* 164459 (2017).
69. Wu, M. J. & Yu, L. L. The complete mitochondrial genome of *Epeorus herklotsi* (Ephemeroptera: Heptageniidae) from Longquan, Zhejiang, China and its phylogeny. *Mitochondrial DNA B* **3**, 1254–1255 (2018).
70. Li, D., Qin, J. C. & Zhou, C. F. The phylogeny of Ephemeroptera in Pterygota revealed by the mitochondrial genome of *Siphuriscus chinensis* (Hexapoda: Insecta). *Gene* **545**, 132–140 (2014).
71. Lee, E. M. *et al.* The complete mitogenomes of the oriental mayfly, *Ephemera orientalis* (Ephemeroptera: Ephemeridae) and the dragonfly, *Davidius lunatus* (Odonata: Gomphidae). *Korean Society of Applied Entomology*. 179–179 (2008).
72. Lee, E. M. *et al.* The preliminary analysis of the complete mitochondrial genome of the oriental mayfly, *Ephemera orientalis* (Ephemeroptera: Ephemeridae). *Korean Society of Applied Entomology*. 82–82 (2006).
73. Zhang, Y. Y., Xuan, W. J., Zhao, J. L., Zhu, C. D. & Jiang, G. F. The complete mitochondrial genome of the cockroach *Eupolyphaga sinensis* (Blattaria: Polyphagidae) and the phylogenetic relationships within the Dictyoptera. *Mol. Biol. Rep.* **37**, 3509–3516 (2010).
74. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).
75. Hendy, M. D. & Penny, D. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38**, 297–309 (1989).
76. Bergsten, J. A review of long-branch attraction. *Cladistics* **21**, 163–193 (2005).
77. Li, H. *et al.* Higher-level phylogeny of paraneopteran insects inferred from mitochondrial genome sequences. *Sci. Rep.* **5**, 8527 (2015).
78. Li, H. *et al.* Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. *P. Roy. Soc. Lond. B. Bio.* **284** (2017).
79. Song, N., Cai, W. & Li, H. Insufficient power of mitogenomic data in resolving the auchenorrhynchan monophyly. *Zool. J. Linn. Soc. Lond.* **183**, 776–790 (2018).
80. Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC evolutionary biology* **7**, S4 (2007).
81. Martynow, A. V. Über zwei Grundtypen der Flügel bei den Insekten und deren. *Evolution. Z. Morph. Oko. Tiere.* **4**, 465–501 (1924).
82. Crampton, G. C. The phylogeny and classification of insects. *J. Entomol. Zool.* **16**, 33–47 (1924).
83. Haas, F. & Kukalova-Peck, J. Dermapteran hindwing structure and folding: New evidence for familial, ordinal and superordinal relationships within Neoptera (Insecta). *Eur. J. Entomol.* **98**, 445–509 (2001).
84. Kukalova-Peck, J. Arthropod phylogeny and 'basal' morphological structures. *In Arthropod Relat.* 249–268 (1998).
85. Willkommen, J. & Hörnschemeyer, T. The homology of wing base sclerites and flight muscles in Ephemeroptera and Neoptera and the morphology of the pterothorax of *Habroleptoides confusa* (Insecta: Ephemeroptera: Leptophlebiidae). *Arthropod Struct. Dev.* **36**, 253–269 (2007).
86. Bechly, G., Brauckmann, C., Zessin, W. & Gröning, E. New results concerning the morphology of the most ancient dragonflies (Insecta: Odonatoptera) from the Namurian of Hagen-Vorhalle (Germany). *J. Zool. Syst. Evol. Res.* **39**, 209–226 (2001).
87. Hovmöller, R., Pape, T. & Källersjö, M. The Palaeoptera problem: basal pterygote phylogeny inferred from 18S and 28S rDNA sequences. *Cladistics* **18**, 313–323 (2002).
88. Ishiwata, K., Miyata, T. & Su, Z. H. Phylogenetic analysis of winged insects based on several nuclear protein-coding genes. *Genes Genet. Syst.* **85**, 398 (2011).
89. Simon, S., Schierwater, B. & Hadrys, H. On the value of elongation factor-1[alpha] for reconstructing pterygote insect phylogeny. *Mol. Phylogenet. Evol.* **54**, 651–656 (2010).

## Acknowledgements

This research is supported by grants from the Science and Technology Innovation Fund of Henan Agricultural University (KJCX2019A10), the Earmarked Fund for China Agriculture Research System (NO. CARS-27), Key Scientific Research Projects of Henan Province, China (Grant Nos. 16A210029, 18B210006), Study on Technology of Chemical Fertilizer Reduction and Synergism in Tea Garden (2016YFD0200901), Fruit Industry System of Henan Province (S2014-11-G03), Henan Academician Workstation of Pest Green Prevention and Control for Plants in Southern Henan (YZ201601).

## Author contributions

N.S., X.Y. and J.Y. designed this study. N.S. and X.L. performed the molecular experiments. N.S., X.L., X.L. and P.P. analyzed the data. All authors discussed results and implications. N.S. wrote the manuscript. All authors have read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-54391-9>.

**Correspondence** and requests for materials should be addressed to N.S.

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

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons 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) 2019