

# 1 Using mitochondrial genomes to infer phylogenetic relationships 2 among the oldest extant winged insects (Palaeoptera)

3 Sereina Rutschmann<sup>a,b,c\*</sup>, Ping Chen<sup>d</sup>, Changfa Zhou<sup>d</sup>, Michael T. Monaghan<sup>a,b</sup>

## 4 **Addresses:**

5 *<sup>a</sup>Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587  
6 Berlin, Germany*

7 *<sup>b</sup>Berlin Center for Genomics in Biodiversity Research, Königin-Luise-Straße 6-8, 14195 Berlin,  
8 Germany*

9 *<sup>c</sup>Department of Biochemistry, Genetics and Immunology, University of Vigo, 36310 Vigo, Spain*

10 *<sup>d</sup>The Key Laboratory of Jiangsu Biodiversity and Biotechnology, College of Life Sciences, Nanjing  
11 Normal University, Nanjing 210046, China*

12

13 **\*Correspondence:** Sereina Rutschmann, Department of Biochemistry, Genetics and Immunology,  
14 University of Vigo, 36310 Vigo, E-mail: [sereina.rutschmann@gmail.com](mailto:sereina.rutschmann@gmail.com)

15

16

17

18

## 19 Abstract

20 Phylogenetic relationships among the basal orders of winged insects remain unclear, in particular the  
 21 relationship of the Ephemeroptera (mayflies) and the Odonata (dragonflies and damselflies) with the  
 22 Neoptera. Insect evolution is thought to have followed rapid divergence in the distant past and  
 23 phylogenetic reconstruction may therefore be susceptible to problems of taxon sampling, choice of  
 24 outgroup, marker selection, and tree reconstruction method. Here we newly sequenced three  
 25 mitochondrial genomes representing the two most diverse families of the Ephemeroptera, one of  
 26 which is a basal lineage of the order. We then used an additional 90 insect mitochondrial genomes to  
 27 reconstruct their phylogeny using Bayesian and maximum likelihood approaches. Bayesian analysis  
 28 supported a basal Odonata hypothesis, with Ephemeroptera as sister group to the remaining insects.  
 29 This was only supported when using an optimized data matrix from which rogue taxa and terminals  
 30 affected by long-branch attraction were removed. None of our analyses supported a basal  
 31 Ephemeroptera hypothesis or Ephemeroptera + Odonata as monophyletic clade sister to other insects  
 32 (i.e., the Palaeoptera hypothesis). Our newly sequenced mitochondrial genomes of *Baetis*  
 33 *rutilocylindratus*, *Cloeon dipterum*, and *Habrophlebiodes zijinensis* had a complete set of protein  
 34 coding genes and a conserved orientation except for two inverted tRNAs in *H. zijinensis*. Increased  
 35 mayfly sampling, removal of problematic taxa, and a Bayesian phylogenetic framework were needed  
 36 to infer phylogenetic relationships within the three ancient insect lineages of Odonata,  
 37 Ephemeroptera, and Neoptera. Pruning of rogue taxa improved the number of supported nodes in all  
 38 phylogenetic trees. Our results add to previous evidence for the Odonata hypothesis and indicate that  
 39 the phylogenetic resolution of the basal insects can be resolved with more data and sampling effort.

40 **Keywords:** *Ephemeroptera*; *Long-branch attraction*; *Mayfly*; *Odonata*; *Palaeoptera problem*; *Rogue*  
 41 *taxa*

## 42 1. Introduction

43 Insects are the most diverse branch of metazoan life, yet the relationships among the basal  
44 branches of the winged insects (Pterygota) remains one of the major open questions in the evolution  
45 of life. The unresolved relationship of the basal orders Odonata (dragon- and damselflies) and  
46 Ephemeroptera (mayflies) to the Neoptera (i.e., to the rest of the Pterygota) has been termed the  
47 “Palaeoptera problem” (Hovmöller, 2002; Ogden and Whiting, 2003). The name Paleoptera (“old  
48 wings”) reflects the inability of the Ephemeroptera and Odonata to fold their wings flat over the  
49 abdomen. This has long been considered to be the ancestral condition, in contrast to the more derived  
50 Neoptera (“new wings”) (reviewed by Trautwein et al. 2012). The monophyly of the Neoptera,  
51 including the three major lineages Polyneoptera, Paraneoptera, and Holometabola, is widely  
52 accepted, although relationships among polyneopteran lineages (Whitfield and Kjer, 2008) and the  
53 monophyly of the Paraneoptera (Ishiwata et al., 2011) remain unresolved.

54 There are three competing hypotheses relating to the Palaeoptera problem: the Palaeoptera  
55 hypothesis ((Ephemeroptera + Odonata) + Neoptera), the basal Ephemeroptera hypothesis  
56 (Ephemeroptera + (Odonata + Neoptera)), and the basal Odonata hypothesis (Odonata +  
57 (Ephemeroptera + Neoptera)) (Kjer et al., 2016; Trautwein et al., 2012; Yeates et al., 2012). All  
58 hypotheses are supported by morphological as well as molecular data to varying degrees (reviewed  
59 by Trautwein et al. 2012). Different authors, using the same set of genes and taxa set but different  
60 phylogenetic approaches, namely Bayesian inference (BI) vs. maximum likelihood (ML) (Mallatt  
61 and Giribet, 2006; Meusemann et al., 2010; Ogden and Whiting, 2003), and nucleotide vs. amino  
62 acid (aa) sequences (Song et al., 2016) obtained results supporting distinct hypotheses. The  
63 Palaeoptera hypothesis has received support from several molecular studies based on rRNA genes  
64 and nuclear DNA (Ishiwata et al., 2011; Kjer et al., 2006; Sasaki et al., 2013; Thomas et al., 2013),  
65 and from nDNA phylogenomic analyses (Meusemann et al., 2010; Regier et al., 2010; Reumont et  
66 al., 2012). Thomas et al. (2013) and Kjer et al. (2006) included the most comprehensive taxon

sampling with 35, respectively seven species from each of the three lineages (i.e., Ephemeroptera, Odonata, Neoptera) analyzing seven, respectively eight genes (rRNAs, nuclear DNA (nDNA), mitochondrial DNA (mtDNA)). The basal Ephemeroptera hypothesis was supported by previous studies using mitochondrial genome data (Ma et al., 2014; Song et al., 2016; Zhang et al., 2008), and a combined analysis of rRNA and one nuclear gene (Ogden and Whiting, 2003). Other studies have concluded the relationships are ambiguous (Ogden and Whiting, 2003). Notably, studies supporting the basal Ephemeroptera hypothesis have included either a limited number of mayfly species (e.g., one (Whiting et al., 1997; Zhang et al., 2008), two (Giribet and Ribera, 2000; Ma et al., 2014), four (Ma et al., 2015), or five (Song et al., 2016) species) or a limited number of unlinked loci (e.g., one (Ma et al., 2014; Ma et al., 2015; Song et al., 2016; Zhang et al., 2008), two (Giribet and Ribera, 2000; Wheeler et al., 2001; Whiting et al., 1997), or three (Ogden and Whiting, 2003)). The basal Odonata hypothesis has received support from previous studies based on mitochondrial genome data (Lin et al., 2010; Wan et al., 2012), rRNA genes (Kjer, 2004; Mallatt and Giribet, 2006; Misof et al., 2007; Reumont et al., 2009; Yoshizawa and Johnson, 2005), combining rRNA and nDNA (Blanke et al., 2013), and phylogenomic nDNA (Meusemann et al., 2010; Simon et al., 2009). These studies, with the exceptions of few studies based on *18S* rRNA (Kjer, 2004; Misof et al., 2007; Yoshizawa and Johnson, 2005), included in total between two (Mallatt and Giribet, 2006; Simon et al., 2009) and 13 species (Reumont et al., 2009) from the orders Ephemeroptera and Odonata.

The conflicting phylogenetic signals may result from the ancient radiation of the lineages Ephemeroptera, Odonata, and Neoptera from a common ancestor in the distant past (Whitfield and Kjer, 2008). These three ancient lineages appear to have diverged rapidly, leaving few characteristics to determine their phylogenetic relationships (Whitfield and Kjer, 2008). Evolutionary rate heterogeneity across clades and the representation of old clades by recent extant taxa make ancient relationships such as those of early winged insects the most difficult challenges for phylogenetics (Kjer et al., 2016; Whitfield and Kjer, 2008). Given the weak phylogenetic signal, reconstructions

are more susceptible to systematic errors (Baurain et al., 2007; Rokas and Carroll, 2006; Thomas et al., 2013; Whitfield and Kjer, 2008; Whitfield and Lockhart, 2007). Other sources of conflicting signal in studies of the relationships among these three lineages may result from differences in taxon sampling, sequence data, alignment methods, and phylogenetic methods used including models of evolution (Talavera and Vila, 2011; Thomas et al., 2013; Trautwein et al., 2012). The use of mitochondrial protein-coding genes (PCGs) derived from mitochondrial genomes can circumvent some of these difficulties, since they are easier to align and also appropriate models of molecular evolution are well established (Abascal et al., 2006; 2005; Le et al., 2017). On the other hand, mitochondrial genes include several drawbacks, most importantly the possible presence of pseudogenes (Bensasson, 2001; Rogers and Griffiths-Jones, 2012; Song et al., 2008). However, data derived from high-throughput sequencing, depending on the coverage bears the potential to distinguish between functional genes and pseudogenes based on higher mutation rates of the later (Claes and De Leeneer, 2014).

Overall, mitochondrial markers are well studied and therefore the most widely employed genetic markers in insects, being considered a promising “instrument” for insect systematics (Cameron, 2014). Insect mitochondrial genomes are highly conserved, ranging from 15 to 18 kb in length, containing 37 genes: 13 PCGs, two ribosomal RNAs (rRNAs, *rrnL* and *rrnS*), and 22 transfer RNAs (tRNAs, *trn\**) (Boore, 1999). A non-coding region of variable length, thought to be the origin of initiation of transcription and replication, is typically present (Saito, 2005; Zhang and Hewitt, 1997) and referred to as the AT-rich region. The typical ancestral insect mitochondrial genome differs from the ancestral arthropod mitochondrial genome only by the location of *trnL* (Boore et al., 1995). Significant differences in structure, gene content, and gene arrangement have been found to be the exception for highly derived taxa.

Importantly there is still an unbalance in the numbers of available molecular markers between the different insect lineages. In particular the early winged insects (i.e., Ephemeroptera and Odonata) are

relatively under-represented. This is (partly) because so far most genomic data were obtained for insect orders that are of economical interest (e.g., as pollinators, model species, agricultural pests, and vectors of human diseases, Robinson et al., 2011). As of January 2017, there were 39 complete or nearly complete palaeopteran mitochondrial genomes (Ephemeroptera: 18, Odonata: 21) from 22 families (eleven of each) available on GenBank; 22 of which were included in publications (Ephemeroptera: ten, Odonata: twelve). Compared to the number of families that have been described (Ephemeroptera: 42 (Barber-James et al., 2008), Odonata: 30 (Dijkstra et al., 2013)) this is still a relatively low number.

The insect order Ephemeroptera comprises over 3,000 species, comprising as the two most species rich families the Baetidae (833 species) and the Leptophlebiidae (608 species) (Barber-James et al., 2008). Their biogeographic origin is probably Pangean, including a greater diversity and higher endemism rate in the Neotropics and Australasia for the Baetidae, and the Neotropical and Afrotropical regions for the Leptophlebiidae, respectively (Barber-James et al., 2008).

We investigated the relationships of the oldest extant winged insects and newly sequenced three mayfly mitochondrial genomes to improve sampling of this under-represented order. New taxa included one representative of the family Leptophlebiidae, for which no mitochondrial genome data were available, and two representatives of the Baetidae, one from each subfamily. We used BI and ML approaches and created matrices that eliminated taxa that reduced overall tree support (rogue taxa) and that were affected by long-branch attraction (LBA). Our analysis included 29 palaeopteran mitochondrial genomes from 20 families and 64 other insect mitochondrial genomes. We expect the increased taxon sampling to improve the phylogenetic signal of the basal nodes and add support for one of the three competing hypotheses.

## 2. Material and methods

### 2.1. Taxon sampling

We newly sequenced three mitochondrial genomes of the two most diverse Ephemeroptera families. *Habrophlebiodes zijingensis* GUI, ZHANG & WU, 1996 is the first representative of the family Leptophlebiidae to have its complete mitochondrial genome sequenced. *Baetis rutilocylindratus* WANG, QIN, CHEN & ZHOU, 2011 and *Cloeon dipterum* L. 1761 are both members of the Baetidae and representative each subfamily: *B. rutilocylindratus* from Baetinae, and *C. dipterum* from Cloeninae. There was a mitochondrial genome from one other member of the Baetinae available on GenBank prior to our study (*Alainites yixiani*, GQ502451, Jia & Zhou). Our analysis also included four species from Archaeognatha, three from Zygentoma (formerly Thysanura), ten from Odonata, 16 additional species from Ephemeroptera, four from Plecoptera, four from Ensifera of Orthoptera, nine from Caelifera of Orthoptera, 13 from Phasmatodea, two from Mantodea, six from Blattodea, 13 from Isoptera, and one each from Collembola, Dermaptera, Grylloblattodea, and Mantophasmatodea (Table 1). Amino acid sequences were obtained from GenBank using a custom Python script (mitogenome\_ncbi.py, [https://github.com/srutschmann/python\\_scripts](https://github.com/srutschmann/python_scripts)).

## 2.2. Sequencing and assembly

The three mayfly species were sequenced with either 454/FLX pyrosequencing (*C. dipterum*) or Sanger sequencing (*B. rutilocylindratus* and *H. zijingensis*). We extracted DNA from *C. dipterum* from twelve to twenty pooled reared subimago specimens, prepared a shotgun library, and sequenced four lanes on a Roche (454) *GS FLX* machine at the Berlin Center for Genomics in Biodiversity Research (BeGenDiv, Berlin, Germany) (see also Rutschmann et al., 2017). The obtained sequence reads were trimmed and *de novo* assembled using the software Newbler v.2.5.3 (454 Life Science Cooperation) under default settings for large data sets. In order to extract the mitochondrial genome of *C. dipterum*, we performed BLASTN searches (Altschul, 1997) using as query all assembled contigs against the NCBI database. We mapped all matching reads back to the mitochondrial genome with BWA (Li and Durbin, 2009), using the BWA-SW algorithm (Li and Durbin, 2010) with settings

suggested for 454 data by CORAL (match score = 2, mismatch penalty = 2, and gap open penalty = 3, Salmela and Schroder, 2011).

Specimens of *B. rutilocylindratus* and *H. zijinensis* were collected in Zijin Hill, Nanjing, China. The DNA was extracted from between two and four larvae using the DNeasy® Blood & Tissue (Qiagen, Leipzig, Germany) kit. Four DNA fragments of each species were amplified with universal primers (*B. rutilocylindratus*: *cob*, *cox1*, *cox3*, *rrnL*; *H. zijinensis*: *cob*, *cox1*, *nad4*, *rrnL*, (Simon et al., 1994)). Subsequently, based on the previously obtained sequence information we designed six (*B. rutilocylindratus*) respective four (*H. zijinensis*) specific primer pairs (Supplementary Table S1; see also Li et al. (2014)). Standard and long polymerase chain reactions (PCRs) were performed on a DNA Engine Peltier Thermal Cycler (Bio-Rad, Shanghai, China). Therefore, we used the rTaq™ DNA polymerase (TaKaRa Bio, Dalian, China) for fragments smaller than two kb and the LA Taq™ polymerase (TaKaRa Bio, Dalian, China) for fragments larger than two kb. All PCR products were purified with the Axygen agarose-out kit. When the PCR amplification signal was too weak to sequence or sequencing resulted in overlap peaks, the products were ligated to pGEM®-T Easy Vector (Promega, Southampton, UK) by *Escherichia coli*, and each resulting clone was sequenced. All purified amplification products were sequenced successively in both directions on an ABI3130xl capillary sequencer. Forward and reverse sequences were assembled and edited using CodonCode Aligner v.3.5.6 (CodonCode Corporation).

### 2.3. Sequence alignments and data matrices

We prepared two data matrices of aa sequences based on the annotated mitochondrial genomes (see 2.5. Annotation and characterization), including a set with all 93 taxa (all\_taxa matrix), and a matrix for which we removed taxa that were either identified as rogue taxa using the RogueNaRok algorithm (Aberer et al., 2012; Wilkinson, 1996) and showing LBA (Felsenstein, 1978; Hedtke et al., 2006) (optimized\_taxa matrix; 88 taxa, Table 1). We used GUIDANCE2 (Landan and Graur, 2008;



189 Sela et al., 2015) to align the sequences, and identify and remove positions detected as ambiguously  
190 aligned regions. To run GUIDANCE we applied the default settings, using as multiple sequences  
191 alignment program MAFFT v.7.050b (L-INS-I algorithm with default settings, Katoh and Standley,  
192 2013).

193 The best-fitting partitioning schemes and corresponding aa substitution models were estimated  
194 with PartitionFinder v.2 (<https://github.com/brettc/partitionfinder>, Lanfear et al., 2012). We used the  
195 Bayesian Information Criterion (BIC) to choose the best model, linked branch lengths, and a greedy  
196 search (Lanfear et al., 2012) with RAxML v.8.2.8 (Stamatakis, 2014). Thereby each gene was  
197 defined as one data block (i.e., possible partition).

198 To identify rogue taxa, we applied the RogueNaRok algorithm (Aberer et al., 2012; Wilkinson,  
199 1996). We used the bootstrap replicates and the best supported maximum likelihood tree based on  
200 the all\_taxa matrix (see 3.1. Data matrices). Identified rogue taxa were removed from the all\_taxa  
201 matrix, resulting in the optimized\_taxa matrix (Table 1). Taxa showing evidence of LBA *sensu*  
202 Bergsten (Bergsten, 2005) were identified using heterogeneities in sequence divergence based on a  
203 model of evolution derived from the entire data matrix. For this, the Relative Composition Frequency  
204 Variability (RCFV) values of each species were calculated with BaCoCa (Kück and Struck, 2014)  
205 using the partitioned concatenated data matrix (i.e., all\_taxa matrix). Taxa with the highest values  
206 were excluded in the optimized\_taxa matrix, apart from the Ephemeroptera, which were retained  
207 because they are the focus of our study and their monophyly is well-established (Hovmöller, 2002;  
208 Ogden and Whiting, 2003).

#### 209 **2.4. Phylogenetic reconstruction**

210 Phylogenetic reconstructions were carried out using MrBayes v.3.2.4 (Ronquist et al., 2012), and  
211 RAxML v.8.1.7 (Stamatakis, 2014). We analyzed both matrices (all\_taxa, optimized\_taxa) using the  
212 best-fitting partitioning scheme and model for each (Table 2). For BI, we unlinked the frequencies,  
213 gamma distributions, substitution rates and the proportion of invariant sites across partitions. For

each matrix, we run two independent analyses of four MCMC chains with  $10^7$  generations, sampling every  $10^3$  generations and discarding a burn-in of 25%. Maximum likelihood inferences were performed with 200 bootstrap replicates. Trees reconstructed from the all\_taxa matrix were rooted with *Tetrodontophora bielanensis* (Collembola). Trees reconstructed from the optimized\_taxa matrix were rooted with *Nesomachilis australica* (Archaeognatha, according to Song et al. 2016) because *Tetrodontophora bielanensis* was found to be a rogue taxon (see 3. Results). All trees were visualized with the ape v.4.1 (Paradis et al., 2004), phangorn v.2.1.1 (Schliep, 2011), and ggtree v.1.4.20 (Yu et al., 2017) packages for R v.3 (R Development Core Team, 2012).

## 2.5. Annotation and characterization

All mitochondrial genomes were annotated using the MITOS webserver (<http://mitos.bioinf.uni-leipzig.de/index.py>, Bernt et al., 2013), DOGMA (Wyman et al., 2004), and MFannot (<http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>). The predicted PCGs were checked for stop codons and manually adjusted by comparison with the longest predicted open reading frames as implemented in Geneious R7 v.7.1.3 (Biomatters Ltd.), and by comparison with the respective homologous insect sequence alignment. For the tRNA prediction, we additionally used ARWEN v.1.2.3 (Laslett and Canback, 2008) and tRNAscan-SE v.1.21 (<http://lowelab.ucsc.edu/tRNAscan-SE/>, Lowe and Eddy, 1997). The annotated mitochondrial genomes were visualized with OrganellarGenomeDRAW (<http://ogdraw.mpimp-golm.mpg.de/cgi-bin/ogdraw.pl>, Lohse et al., 2013) and manually edited.

Nucleotide contents were retrieved using Geneious. To correct biases in the AT content due to incomplete mitochondrial genomes mostly missing the AT-rich region, we removed all AT-rich regions and the two rRNAs including the five close-by tRNAs (*trnL1-trnM*) and recalculated the nucleotide base pair (bp) compositions. The AT and GC composition skewness were calculated as follows: AT-skew =  $(A - T) / (A + T)$ , and GC-skew =  $(G - C) / (G + C)$  (Perna and Kocher, 1995).

## 238 3. Results

### 239 3.1. Data matrices

240 The final matrices had lengths of 3,675 aa (all\_taxa matrix, including 93 taxa) and 3,673 aa  
 241 (optimized\_taxa matrix, including 88 taxa, Table 1). As best-fitting partitioning scheme for the two  
 242 matrices, we identified either five partitions (all\_taxa matrix, Table 2) or four partitions  
 243 (optimized\_taxa, Table 2). As best-fitting models of aa sequence evolution, we identified MtZoa  
 244 (Rota-Stabelli et al., 2009) and MtArt (Abascal et al., 2006).

245 The RogueNaRok analysis identified the outgroup *T. bielanensis* as well as *Gryllotalpa orientalis*  
 246 (Orthoptera) and *Phraortes* sp. (Phasmatodea) as taxa with uncertain phylogenetic positions, leading  
 247 to less accurate phylogenetic reconstructions. *G. orientalis* (Orthoptera) and *Phraortes* sp.  
 248 (Phasmatodea) clustered within the corresponding order using the all\_taxa matrix, although the  
 249 nodes containing these species were not supported by ML. In the test for LBA, *Aposthonia japonica*  
 250 (Embioptera) and *Challia fletcheri* (Dermaptera) had high RCFV values (0.0237 and 0.0307,  
 251 respectively) suggesting they were affected by LBA. Both species clustered within other orders:  
 252 *Challia fletcheri* (Dermaptera) within the Ephemeroptera as sister taxon to a clade containing all  
 253 representatives of the family Baetidae, and *A. japonica* (Embioptera) within the Phasmatodea as  
 254 sister taxon to a clade containing all species of the order with the exception of *T. californicum* (Table  
 255 3, Supplementary Fig. S1). The mayflies Teloganodidae sp. (0.0305) and the three baetids *A. yixiani*  
 256 (0.0235), *B. rutilocylindratus* (0.0232), and *C. dipterum* (0.0201) showed relatively high RCFV  
 257 values.

### 258 3.2. Phylogenetic reconstruction

259 The Bayesian reconstruction based on the optimized\_taxa matrix produced the tree with most  
 260 support compared to ML or to either analysis of the all\_taxa matrix (Fig. 1, Table 3, Supplementary  
 261 Figs S1-S3). The basal Odonata hypothesis was highly supported in this tree (Bayesian posterior

probability (BPP) = 1.00), whereas neither the basal Ephemeroptera hypothesis nor the Palaeoptera hypothesis received support in any analysis (Table 3). The Odonata were monophyletic in all analyses (BPP = 1.00, BS = 100%) whereas the Ephemeroptera were monophyletic only using the optimized\_taxa matrix (Fig. 1, Table 3, Supplementary Fig. S2).

Most other nodes of interest were consistently supported across different matrices and analyses (Table 3). The Polyneoptera were moderately supported (BS  $\geq$  80%) for the optimized\_taxa matrix using the ML. Within the Polyneoptera, all orders except Orthoptera and Blattodea were monophyletic (BPP  $\geq$  0.95, and BS  $\geq$  90%, Table 3). The two Orthoptera suborders Ensifera and Caelifera were monophyletic (Table 3), whereby the Ensifera clustered together with the Plecoptera for the optimized\_taxa matrix using both ML and BI. The Grylloblattodea, Mantophasmatodea, and Phasmatodea were all well supported using the optimized\_taxa matrix. For the all\_taxa matrix also the Embioptera were recovered in this clade as sister taxa to the Phasmatodea excluding *Timema californicum* (Fig. 1, Supplementary Figs S1, S3). The superorder Dictyoptera was monophyletic clade, consisting of the monophyletic Mantodea and the Blattodea including the Isoptera.

### 3.3. Mitochondrial genomes

The mitochondrial genome sequences have been deposited at GenBank with the accession numbers XXXXXXXXX, XXXXXXXXX, and XXXXXXXXX. The pyrosequencing run of *C. dipterum* using the 454 GS FLX system resulted in 651,306 reads, of which 1.14% mapped to the assembled mitochondrial genome. The depth of coverage for the *C. dipterum* mitochondrial genome was 249.2x ( $\pm$  80.6 SD, Fig. 2a). The three mayfly mitochondrial genomes were 15,407 bp (*C. dipterum*), 14,883 bp (*B. rutilocylindratus*), and 14,355 bp (*H. zijinensis*) long (Table 4). For *B. rutilocylindratus*, the complete mitochondrial genome was sequenced (Fig. 2b). The AT-rich regions were incomplete for genomes of *C. dipterum* and *H. zijinensis*. The three tRNAs between the AT-rich region and *nad2* were also missing in *H. zijinensis* due to the incomplete sequencing (Fig. 2c). All three sequenced mitochondrial genomes contained the entire set of 13 PCGs, two rRNAs, and either 22 tRNAs (*B.*

287 *rutilocylindratus*, *C. dipterum*) or 19 for the incomplete mitochondrial genome of *H. zijinensis*,  
288 including 23 coded at the (+) strand (21 for *H. zijinensis*) and 14 at the (-) strand (13 for *H.*  
289 *zijinensis*) (Fig. 2).

290 The order of the genes was conserved, with the exceptions of the two inverted tRNAs Arginine  
291 and Alanine for *H. zijinensis* (Fig. 2c). All PCGs started with the ATN codons (ATT, ATG, ATA),  
292 and mostly ended with the complete termination codon (TAA or TAG). Exceptions were an  
293 incomplete T termination codon for *nad4* in *B. rutilocylindratus* and *cox1* started with CTC in *C.*  
294 *dipterum*. The AT content of the mitochondrial genomes that were corrected for the missing regions,  
295 was lowest in *B. rutilocylindratus* (59.9%, Table 4, Supplementary Table S2). The mean AT content  
296 for all available mayfly mitochondrial genomes was 65.34%. The average AT-skew was -0.03 ( $\pm$   
297 0.04 SD) and the average GC-skew was -0.13 ( $\pm$  0.14 SD).

## 298 4. Discussion

### 299 4.1. Phylogenetic relationships of Palaeoptera

300 Our Bayesian reconstruction based on the optimized\_taxa matrix provided strong support for the  
301 basal Odonata hypothesis. Previous work based on mitochondrial genome data is (partially)  
302 congruent with our findings (Lin et al., 2010; Song et al., 2016; Wan et al., 2012). In contrast to our  
303 results, a handful of studies based on nucleotide and aa sequences of mitochondrial PCGs and  
304 mitochondrial RNA (mtRNAs) found overall strong support for the basal Ephemeroptera hypothesis  
305 (Ma et al., 2014; Ma et al., 2015; Song et al., 2016; Zhang et al., 2008). Interestingly, Song et al.  
306 (2016), supporting overall the Ephemeroptera hypothesis, found evidence for the Odonata hypothesis  
307 when using aa sequence data. Our results confirm the trend towards a better support for the ancestral  
308 position of the Odonata when using an increased mayfly taxon sampling (more than ten species per  
309 lineage) (Kjer, 2004; Misof et al., 2007; Yoshizawa and Johnson, 2005). The exceptions for this are  
310 the work by Ogden and Whiting (2003) and Thomas et al. (2013), that overall support the

311 Ephemeroptera respective Palaeoptera hypothesis but contain some data sets in favor of the Odonata  
 312 hypothesis. Early inconsistencies based on different sequence alignment strategies (Ogden and  
 313 Whiting, 2003) appear to become less relevant due to the advances of multiple sequence alignment  
 314 programs. Overall, the vast majority of studies were based on BI and ML. Thomas et al. (2013)  
 315 recovered increased resolution using BI in comparison to ML. Different data partitioning (i.e., gene  
 316 trees resulting in different tree topologies) by Kjer et al. (2006) supported different hypotheses for  
 317 different genes. Evidently, using rRNAs (nDNA and mtDNA) and the nDNA EF1- $\alpha$  the Odonata  
 318 hypothesis was recovered whereas the analyses based on the combined and on the mtDNA data  
 319 matrix resulted in a monophyletic palaeopteran clade. It is questionable whether an increase in  
 320 analyzed genomic data (i.e., phylogenomic studies) will be able to resolve the Palaeoptera problem.  
 321 Notably, Misof et al. (2014) and Regier et al. (2010), using 1,478 PCGs and seven palaeopteran  
 322 species, respective 62 PCGs and four palaeoptera species, recovered the Ephemeroptera and Odonata  
 323 as sister clades (i.e., the Palaeoptera hypothesis) but without high support values. Overall, more  
 324 knowledge about the use and limitation of individual markers are needed. For example Simon et al.  
 325 (2012) found that proteins involved in cellular processes and signaling harbor the most phylogenetic  
 326 signal.

327 The increased mayfly taxon sampling highlighted the phylogenetic diversity (as indicated by the  
 328 branch lengths of the tip taxa, Fig. 1, Supplementary Figs S1-S3) of the Palaeoptera. The overall  
 329 long-branch lengths and high rate-heterogeneity across sites might be one of the explanations for the  
 330 “Palaeoptera problem”. Especially the family of the Baetidae, which was long thought to be the most  
 331 ancestral mayfly family (Ogden and Whiting, 2005), showed high base composition heterogeneity  
 332 possibly evidencing LBA. However, in agreement with other more recent studies we also recovered  
 333 *S. chinensis* as sister taxa to all other mayflies (Li et al., 2014; Ogden et al., 2009). It remains to be  
 334 resolved whether an increased taxon sampling may obviate LBA *sensu* Bergsten (2005).

#### 335 **4.2. Phylogenetic relationships of Polyneoptera**

Besides the Palaeoptera, a universal consensus on the placement of the Plecoptera, Dermaptera, Embioptera, and Xenoptera (i.e., Grylloblattodea and Mantophasmatodea) remains elusive (Song et al., 2016). The monophyly of the Polyneoptera has become more widely accepted (Ishiwata et al., 2011; Letsch and Simon, 2013; Letsch et al., 2012; Misof et al., 2014; Song et al., 2016; Wan et al., 2012), although several studies and data matrices do not support their monophyly (Kjer et al., 2006; Simon et al., 2012; Song et al., 2016). This is also reflected by our results, recovering the Polyneoptera as moderately supported monophyletic clade when using the optimized\_taxa matrix with ML. The placement of the order Plecoptera also remains elusive. The order often clusters among the more ancestral polyneopteran orders (Misof et al., 2014; Song et al., 2016; Wan et al., 2012) and in a close relationship to the Dermaptera, although earlier studies using few species and the aa sequences of the mitochondrial PCGs also recovered a sister relationship to the Ephemeroptera (Lin et al., 2010; Zhang et al., 2009). Notably, Song et al. (2016) was the only study including four dermapteran species. The monophyly of the Orthoptera has been established by previous studies based on mitochondrial genome data and a large number of PCGs (Fenn et al., 2008; Ma et al., 2009; Misof et al., 2014; Song et al., 2016). However, in a recent study, the Ensifera also clustered as more ancestral clade (Song et al., 2016). As for this study, Song et al. (2016) also used BI and ML based on mitochondrial genome data. Notably, Song et al. (2016) explain the more ancestral position of Ensifera by more similar sequence compositional bias in comparison to the outgroups. The relationship between Grylloblattodea and Mantophasmatodea (i.e., Xenoptera superorder) remains elusive. Mostly this is due to the limited amount of available molecular data, including ordinal-level phylogenies based on two (Ishiwata et al., 2011; Song et al., 2016) or three xenopteran species (Misof et al., 2014). Using mitochondrial genome data and one species per order, also Song et al. (2016) did not recover them as a monophyletic clade. Their monophyly was supported by previous studies using one Mantophasmatodea and one Grylloblattodea species (Ishiwata et al., 2011), respective one Mantophasmatodea and two Grylloblattodea species (Misof et al., 2014). The



Phasmatodea have also been found in previous studies as paraphyletic clade due to the outside position of *T. californicum* (Ma et al., 2014; Song et al., 2016). In contrast, Misof et al. (2014), using a large set of nuclear PCGs also found *Timema* as the sister species to all other representatives of the order. The monophyly of the superorder Dictyoptera is generally accepted and the hierarchical clustering (Mantodea + (Blattodea + Isoptera)) has been well supported (Inward et al., 2007; Ishiwata et al., 2011; Kjer et al., 2006; Legendre et al., 2015; Ma et al., 2014; Misof et al., 2014; Song et al., 2016; Wan et al., 2012) using nucleotide and aa sequences of both mtDNA and nDNA.

### 4.3. Optimizing data matrix

The optimization of the data matrix overall resulted in a better-supported tree (as measured by the total number of supported nodes and the fact that all nodes supported in the all\_taxa matrix were also recovered in the optimized\_taxa matrix). In a study on butterfly phylogenetics, the authors also removed rogue taxa and found dramatically increased bootstrap support values (Regier et al., 2013). Comparing the number of supported nodes from the BI and ML approach, the former always resulted in more resolved nodes. This result is consistent with previous work on the “Palaeoptera problem”, recovering more support for analyses using BI (Thomas et al., 2013). Generally, the choice of the outgroup species was reported as being crucial for resolving problematic splits in the tree of life such as insects origin (Thomas et al., 2013). Thus, finding the outgroup as a rogue taxon was perhaps not surprising. The close phylogenetic relationship of the Dermaptera and Ephemeroptera (Supplementary Figs S1, S3) might be misleading due to LBA. Li et al. (2014) also found the order Dermaptera as being closely related to the mayflies. On the other hand studies based on phylogenomic data reported the Dermaptera as sister taxa to the orders Plecoptera (Ishiwata et al., 2011; Ma et al., 2014; Simon et al., 2012; Wan et al., 2012; Wu et al., 2014) Zoraptera (Blanke et al., 2013; Misof et al., 2014). Using mitochondrial genome data and an increased taxon sampling (i.e., four dermapteran species), Song et al. (2016) found the Dermaptera and/or Plecoptera being sister to the remaining taxa within Polyneoptera. Notably, when only using one species of Dermaptera, *C.*



386 *fletcheri* emerged as sister taxa to the Ephemeroptera (Song et al., 2016). This result in agreement  
387 with our findings is presumably due to LBA. Phylogenomic data matrices (e.g., expressed sequence  
388 tags data: Letsch et al., 2012, 1,478 PCGs: Misof et al., 2014, and nuclear genes Ishiwata et al.,  
389 2011) found the Embioptera to be sister clade to the Phasmatodea. Using mitochondrial genomes  
390 data and BI, the Embioptera were also found as sister group to the Phasmatodea (excluding *T.*  
391 *californicum*) (Song et al., 2016). However, in the same study, they also found support for the sister-  
392 relationship of Embioptera-Zoraptera using ML (Song et al., 2016). Overall, more phylogenetic  
393 studies will be needed to clarify this issue; mostly also because the use of few taxa from anomalous  
394 orders (i.e., Dermaptera and Embioptera) tend to evoke LBA (Song et al., 2016).

#### 395 **4.4. Characterization of ephemeropteran mitochondrial genomes**

396 Mayfly mitochondrial genomes are widely conserved in gene content and nucleotide composition  
397 across eleven families, with only a few differences in the content of their tRNAs. The coverage of  
398 the *Cloeon*-mitochondrial genome was in agreement with other studies, using the same sequencing  
399 platform (e.g., 59-281x, Pons et al., 2014). The missing of part of the AT-rich region is due to  
400 reduced sequencing and assembly efficiency of this low complexity region and common among  
401 insect mitochondrial genomes (Li et al., 2014; Tang et al., 2014). The gene order and orientation  
402 were with the exception of two shifted tRNAs for *H. zijinensis* identical to the ancestral insect  
403 mitochondrial genome (Boore, 1999; Cameron, 2014; Simon et al., 1994). Other mayflies are also  
404 reported to miss complete T termination codons in the genes *cox2* and *nad5* (Lee et al., 2009; Li et  
405 al., 2014; Tang et al., 2014). The two rRNAs were located between *trnL* and *trnV* (*rrnL*), and  
406 between *trnV* and the AT-rich region (*rrnS*), respectively. All ephemeropteran mitochondrial  
407 genomes contain an AT-rich region, which is placed between the *rrnS* (- strand) and *trnI* (+ strand).  
408 Li et al. (2014) reported two distinct parts within the AT-rich region in *Siphuriscus chinensis*, which  
409 also seems to be present in *C. dipterum*. Therein, they described the so called CR<sub>1</sub>, which is located  
410 close to the *rrnS* and has a high AT content (71.6%), including six identical 140 bp sequences, and

the CR<sub>2</sub>, which is close to the *trnI* and has a lower AT content (58.1%). *E. orientalis* contains two identical 55 bp long sequences in the AT-rich region (Lee et al., 2009). Few mayfly mitochondrial genomes differ in their gene content from the ancestral insect mitochondrial genome, possessing one additional tRNA. The two heptageniid species *Parafronurus youi* and *Epeorus* sp. encode a second copy of the *trnM* (AUG, *trnM2*) gene located between *trnI* and *trnQ* (Tang et al., 2014; Zhang et al., 2008). For *S. chinensis*, an additional *trnK2* (AAA) gene is described (Li et al., 2014). Overall, Song et al. (2016) found similar genetic distances for the early branching insects lineages. The AT contents for the Ephemeroptera were very similar to the ones of Odonata (62.6%-68.5% for *coxI*, Kim et al., 2014).

## 5. Conclusions

Our analysis based on an increased mayfly taxon sampling and data matrix optimization supported the Odonata hypothesis and unravelled fine structural changes within the palaeopteran mitochondrial genomes. While both BI and ML overall resulted in highly supported trees, only the BI based on an optimized taxa matrix highly supported the sister-relationship of the Ephemeroptera and Neoptera. The optimized taxa matrix, excluding rogue taxa and taxa with LBA, resulted in an overall better supported tree (as measured by the number of supported nodes) for both BI and ML. Our findings highlight the essential need to increase the taxon sampling of under-represented lineages (such as the Palaeoptera) in order to resolve their phylogenetic position. Here we demonstrated the need for an increased taxon sampling in combination with data matrix optimization, and the use of different phylogenetic approaches in order to resolve an ancient radiation. Establishing general recommendation for data matrix optimization requires additional analyses on a broader range of lineages.

## Author contributions

SR and MTM conceived the study; SR sequenced and analyzed the *Cloeon dipterum* genome, and performed all combined analyses; PC and CZ obtained and analyzed the *Baetis rutilocylindratus* and *Habrophlebiodes zijinensis* genomes; SR and MTM wrote the manuscript. All authors made contributions to subsequent revisions and agreed to the final version.

## Acknowledgements

We are grateful to D. H. Funk from the Stround Water Research Center for providing the *Cloeon dipterum* specimens, S. Mbedi, K. Preuß, and L. Wächter for their great help with laboratory work, G. Glöckner, and C. Mazzoni for constructive comments about the analysis of the mitochondrial genomes, and Zedat-HPC at the Freie Universität Berlin, Germany for providing access to high-performance computing clusters. SR thanks the Janggen-Pöhn-Stiftung (<http://www.janggen-poehn.ch/>) for a Postdoctoral stipend. We also want to thank the members of our research groups for constructive working environments. This work was supported by the Leibniz Association PAKT für Forschung und Innovation ('FREDIE' project: SAW-2011-ZFMK-3).

## Data statement

DNA sequences are available on Genbank under accessions no XXXXXXXXX- XXXXXXXXX. Sequence alignments used for the phylogenetic analyses are available on Mendeley Data <https://doi.org/XXX>.

## References

- Abascal, F., Posada, D., Zardoya, R., 2006. MtArt: A New Model of Amino Acid Replacement for Arthropoda. *Mol. Biol. Evol.* 24, 1–5.
- Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105.
- Aberer, A.J., Krompass, D., Stamatakis, A., 2012. Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Webservice. *Syst. Biol.* 62, 162–166.
- Altschul, S., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Barber-James, H.M., Gattolliat, J.-L., Sartori, M., Hubbard, M.D., 2008. Global diversity of mayflies (Ephemeroptera, Insecta) in freshwater. *Hydrobiologia* 595, 339–350.
- Baurain, D., Brinkmann, H., Philippe, H., 2007. Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors? *Mol. Biol. Evol.* 24, 6–9.
- Bensasson, D., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* 16, 314–321.
- Bergsten, J., 2005. A review of long-branch attraction. *Cladistics* 21, 163–193.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313–319.
- Blanke, A., Greve, C., Wipfler, B., Beutel, R.G., Holland, B.R., Misof, B., 2013. The Identification of Concerted Convergence in Insect Heads Corroborates Palaeoptera. *Syst. Biol.* 62, 250–263.
- Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780.
- Boore, J.L., Collins, T.M., Stanton, D., Daehler, L.L., Brown, W.M., 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* 376, 163–165.
- Cameron, S.L., 2014. Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny.

- 478 Annu. Rev. Entomol. 59, 95–117.
- 479 Claes, K.B.M., De Leeneer, K., 2014. Dealing with pseudogenes in molecular diagnostics in the  
480 next-generation sequencing era. *Methods Mol. Biol.* 1167, 303–315.
- 481 Dijkstra, K.-D.B., Bechly, G., Bybee, S.M., Dow, R.A., Dumont, H.J., Fleck, G., Garrison, R.W.,  
482 Hämäläinen, M., Kalkman, V.J., Karube, H., 2013. The classification and diversity of  
483 dragonflies and damselflies (Odonata), in: Zhang, Z.-Q. (Ed.). pp. 36–45.
- 484 Felsenstein, J., 1978. Cases in which Parsimony or Compatibility Methods Will be Positively  
485 Misleading. *Systematic Zoology* 27, 401.
- 486 Fenn, J.D., Song, H., Cameron, S.L., Whiting, M.F., 2008. A preliminary mitochondrial genome  
487 phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found  
488 within mitochondrial genome data. *Mol. Phylogenet. Evol.* 49, 59–68.
- 489 Giribet, G., Ribera, C., 2000. A Review of Arthropod Phylogeny: New Data Based on Ribosomal  
490 DNA Sequences and Direct Character Optimization. *Cladistics* 16, 204–231.
- 491 Hedtke, S.M., Townsend, T.M., Hillis, D.M., 2006. Resolution of phylogenetic conflict in large data  
492 sets by increased taxon sampling. *Syst. Biol.* 55, 522–529.
- 493 Hovmöller, R., 2002. The Palaeoptera Problem: Basal Pterygote Phylogeny Inferred from 18S and  
494 28S rDNA Sequences. *Cladistics* 18, 313–323.
- 495 Inward, D., Beccaloni, G., Eggleton, P., 2007. Death of an order: a comprehensive molecular  
496 phylogenetic study confirms that termites are eusocial cockroaches. *Biology Letters* 3, 331–335.
- 497 Ishiwata, K., Sasaki, G., Ogawa, J., Miyata, T., Su, Z.-H., 2011. Phylogenetic relationships among  
498 insect orders based on three nuclear protein-coding gene sequences. *Mol. Phylogenet. Evol.* 58,  
499 169–180.
- 500 Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7:  
501 Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780.
- 502 Kim, M.J., Jung, K.S., Park, N.S., Wan, X., Kim, K.-G., Jun, J., Yoon, T.J., Bae, Y.J., Lee, S.M.,  
503 Kim, I., 2014. Molecular phylogeny of the higher taxa of Odonata (Insecta) inferred from COI,  
504 16S rRNA, 28S rRNA, and EF1- $\alpha$  sequences. *Entomological Research* 44, 65–79.
- 505 Kjer, K., 2004. Aligned 18S and Insect Phylogeny. *Syst. Biol.* 53, 506–514.
- 506 Kjer, K.M., Carie, F.L., Litman, J., Ware, J., 2006. A Molecular Phylogeny of Hexapoda. *Arthropod*  
507 *Systematics & Phylogeny* 64, 35–44.
- 508 Kjer, K.M., Simon, C., Yavorskaya, M., Beutel, R.G., 2016. Progress, pitfalls and parallel universes:  
509 a history of insect phylogenetics. *J R Soc Interface* 13, 20160363.
- 510 Kück, P., Struck, T.H., 2014. BaCoCa – A heuristic software tool for the parallel assessment of  
511 sequence biases in hundreds of gene and taxon partitions. *Mol. Phylogenet. Evol.* 70, 94–98.
- 512 Landan, G., Graur, D., 2008. Local reliability measures from sets of co-optimal multiple sequence  
513 alignments. *Pac Symp Biocomput* 15–24.
- 514 Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined Selection of  
515 Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Mol. Biol. Evol.* 29,  
516 1695–1701.
- 517 Laslett, D., Canback, B., 2008. ARWEN: a program to detect tRNA genes in metazoan  
518 mitochondrial nucleotide sequences. *Bioinformatics* 24, 172–175.
- 519 Le, V.S., Dang, C.C., Le, Q.S., 2017. Improved mitochondrial amino acid substitution models for  
520 metazoan evolutionary studies. *BMC Evol. Biol.* 17, 136.
- 521 Lee, E.M., Hong, M.Y., Kim, M.I., Kim, M.J., Park, H.C., Kim, K.Y., Lee, I.H., Bae, C.H., Jin, B.R.,  
522 Kim, I., 2009. The complete mitogenome sequences of the palaeopteran insects *Ephemera*  
523 *orientalis* (Ephemeroptera: Ephemeridae) and *Davidius lunatus* (Odonata: Gomphidae). *Genome*  
524 52, 810–817.
- 525 Legendre, F., Nel, A., Svenson, G.J., Robillard, T., Pellens, R., Grandcolas, P., 2015. Phylogeny of  
526 Dictyoptera: Dating the Origin of Cockroaches, Praying Mantises and Termites with Molecular  
527 Data and Controlled Fossil Evidence. *PLoS One* 10, e0130127.

- Letsch, H., Simon, S., 2013. Insect phylogenomics: new insights on the relationships of lower neopteran orders (Polyneoptera). *Sys. Entomol.* 38, 783–793.
- Letsch, H.O., Meusemann, K., Wipfler, B., Schütte, K., Beutel, R., Misof, B., 2012. Insect phylogenomics: results, problems and the impact of matrix composition. *Proc. R. Soc. B.* 279, 3282–3290.
- Li, D., Qin, J.-C., Zhou, C.-F., 2014. The phylogeny of Ephemeroptera in Pterygota revealed by the mitochondrial genome of *Siphyluriscus chinensis* (Hexapoda: Insecta). *Gene* 545, 132–140.
- Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Lin, C.-P., Chen, M.-Y., Huang, J.-P., 2010. The complete mitochondrial genome and phylogenomics of a damselfly, *Euphaea formosa* support a basal Odonata within the Pterygota. *Gene* 468, 20–29.
- Lohse, M., Drechsel, O., Kahlau, S., Bock, R., 2013. OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 41, W575–W581.
- Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.
- Ma, C., Liu, C., Yang, P., Kang, L., 2009. The complete mitochondrial genomes of two band-winged grasshoppers, *Gastrimargus marmoratus* and *Oedaleus asiaticus*. *BMC Genomics* 10, 156.
- Ma, C., Wang, Y., Wu, C., Kang, L., Liu, C., 2014. The compact mitochondrial genome of *Zorotypus medoensis* provides insights into phylogenetic position of Zoraptera. *BMC Genomics* 15, 1156.
- Ma, Y., He, K., Yu, P., Yu, D., Cheng, X., Zhang, J., 2015. The complete mitochondrial genomes of three bristletails (Insecta: Archaeognatha): the paraphyly of Machilidae and insights into archaeognathan phylogeny. *PLoS One* 10, e0117669.
- Mallatt, J., Giribet, G., 2006. 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.
- Meusemann, K., Reumont, von, B.M., Simon, S., Roeding, F., Strauss, S., Kuck, P., Ebersberger, I., Walz, M., Pass, G., Breuers, S., Achter, V., Haeseler, von, A., Burmester, T., Hadrys, H., Wagele, J.W., Misof, B., 2010. A Phylogenomic Approach to Resolve the Arthropod Tree of Life. *Mol. Biol. Evol.* 27, 2451–2464.
- Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., Reumont, von, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walz, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Yang, H., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767.
- Misof, B., Niehuis, O., Bischoff, I., Rickert, A., Erpenbeck, D., Staniczek, A., 2007. Towards an 18S phylogeny of hexapods: Accounting for group-specific character covariance in optimized mixed nucleotide/doublet models. *Zoology* 110, 409–429.



- 578 Ogden, H.T., Whiting, M.F., 2003. The problem with “the Paleoptera Problem:” sense and  
579 sensitivity. *Cladistics* 19, 432–442.
- 580 Ogden, T.H., Gattolliat, J.L., Sartori, M., Staniczek, A.H., Soldán, T., Whiting, M.F., 2009. Towards  
581 a new paradigm in mayfly phylogeny (Ephemeroptera): combined analysis of morphological and  
582 molecular data. *Sys. Entomol.* 34, 616–634.
- 583 Ogden, T.H., Whiting, M.F., 2005. Phylogeny of Ephemeroptera (mayflies) based on molecular  
584 evidence. *Mol. Phylogenet. Evol.* 37, 625–643.
- 585 Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R  
586 language. *Bioinformatics* 20, 289–290.
- 587 Perna, N., Kocher, T., 1995. Patterns of nucleotide composition at fourfold degenerate sites of  
588 animal mitochondrial genomes. *J Mol Evol* 41.
- 589 Pons, J., Bauzá-Ribot, M.M., Jaume, D., Juan, C., 2014. Next-generation sequencing, phylogenetic  
590 signal and comparative mitogenomic analyses in Metacrangonyctidae (Amphipoda: Crustacea).  
591 *BMC Genomics* 15, 566.
- 592 R Development Core Team, 2012. R: A language and environment for statistical computing. R  
593 Foundation for Statistical Computing, Vienna, Austria. *International Statistical Review*.
- 594 Regier, J.C., Mitter, C., Zwick, A., Bazinet, A.L., Cummings, M.P., Kawahara, A.Y., Sohn, J.-C.,  
595 Zwickl, D.J., Cho, S., Davis, D.R., Baixeras, J., Brown, J., Parr, C., Weller, S., Lees, D.C.,  
596 Mitter, K.T., 2013. A large-scale, higher-level, molecular phylogenetic study of the insect order  
597 Lepidoptera (moths and butterflies). *PLoS One* 8, e58568.
- 598 Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham,  
599 C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-  
600 coding sequences. *Nature* 463, 1079–1083.
- 601 Reumont, von, B.M., Jenner, R.A., Wills, M.A., Dell'Ampio, E., Pass, G., Ebersberger, I., Meyer, B.,  
602 Koenemann, S., Iliffe, T.M., Stamatakis, A., Niehuis, O., Meusemann, K., Misof, B., 2012.  
603 Pancrustacean Phylogeny in the Light of New Phylogenomic Data: Support for Remipedia as the  
604 Possible Sister Group of Hexapoda. *Mol. Biol. Evol.* 29, 1031–1045.
- 605 Reumont, von, B.M., Meusemann, K., Szucsich, N.U., Dell'Ampio, E., Gowri-Shankar, V., Bartel,  
606 D., Simon, S., Letsch, H.O., Stocsits, R.R., Luan, Y.-X., Wägele, J., Pass, G., Hadrys, H., Misof,  
607 B., 2009. Can comprehensive background knowledge be incorporated into substitution models to  
608 improve phylogenetic analyses? A case study on major arthropod relationships. *BMC Evol. Biol.*  
609 9, 119.
- 610 Robinson, G.E., Hackett, K.J., Purcell-Miramontes, M., Brown, S.J., Evans, J.D., Goldsmith, M.R.,  
611 Lawson, D., Okamuro, J., Robertson, H.M., Schneider, D.J., 2011. Creating a buzz about insect  
612 genomes. *Science* 331, 1386–1386.
- 613 Rogers, H.H., Griffiths-Jones, S., 2012. Mitochondrial Pseudogenes in the Nuclear Genomes of  
614 *Drosophila*. *PLoS One* 7, e32593.
- 615 Rokas, A., Carroll, S.B., 2006. Bushes in the tree of life. *PLoS Biol.* 4, e352.
- 616 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu,  
617 L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic  
618 Inference and Model Choice Across a Large Model Space. *Syst. Biol.* 61, 539–542.
- 619 Rota-Stabelli, O., Yang, Z., Telford, M.J., 2009. MtZoa: a general mitochondrial amino acid  
620 substitutions model for animal evolutionary studies. *Mol. Phylogenet. Evol.* 52, 268–272.
- 621 Rutschmann, S., Detering, H., Simon, S., Fredslund, J., Monaghan, M.T., 2017. discomark: nuclear  
622 marker discovery from orthologous sequences using draft genome data. *Mol. Ecol. Resour.* 17,  
623 257–266.
- 624 Saito, S., 2005. Replication Origin of Mitochondrial DNA in Insects. *Genetics* 171, 1695–1705.
- 625 Salmela, L., Schroder, J., 2011. Correcting errors in short reads by multiple alignments.  
626 *Bioinformatics* 27, 1455–1461.
- 627 Sasaki, G., Ishiwata, K., Machida, R., Miyata, T., Su, Z.-H., 2013. Molecular phylogenetic analyses

support the monophyly of Hexapoda and suggest the paraphyly of Entognatha. *BMC Evol. Biol.* 13, 236.

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593.

Sela, I., Ashkenazy, H., Katoh, K., Pupko, T., 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* 43, W7–14.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers. *Annals of the Entomological Society of America* 87, 651–701.

Simon, S., Narechania, A., DeSalle, R., Hadrys, H., 2012. Insect Phylogenomics: Exploring the Source of Incongruence Using New Transcriptomic Data. *Genome Biol. Evol.* 4, 1295–1309.

Simon, S., Strauss, S., Haeseler, von, A., Hadrys, H., 2009. A Phylogenomic Approach to Resolve the Basal Pterygote Divergence. *Mol. Biol. Evol.* 26, 2719–2730.

Song, H., Buhay, J.E., Whiting, M.F., Crandall, K.A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13486–13491.

Song, N., Li, H., Song, F., Cai, W., 2016. Molecular phylogeny of Polyneoptera (Insecta) inferred from expanded mitogenomic data. *Nature Publishing Group* 6, 36175.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.

Talavera, G., Vila, R., 2011. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC Evol. Biol.* 11, 315.

Tang, M., Tan, M., Meng, G., Yang, S., Su, X., Liu, S., Song, W., Li, Y., Wu, Q., Zhang, A., Zhou, X., 2014. Multiplex sequencing of pooled mitochondrial genomes--a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Res.* 42, e166–e166.

Thomas, J.A., Trueman, J.W.H., Rambaut, A., Welch, J.J., 2013. Relaxed Phylogenetics and the Palaeoptera Problem: Resolving Deep Ancestral Splits in the Insect Phylogeny. *Syst. Biol.* 62, 285–297.

Trautwein, M.D., Wiegmann, B.M., Beutel, R., Kjer, K.M., Yeates, D.K., 2012. Advances in Insect Phylogeny at the Dawn of the Postgenomic Era. *Annu. Rev. Entomol.* 57, 449–468.

Wan, X., Kim, M.I., Kim, M.J., Kim, I., 2012. Complete Mitochondrial Genome of the Free-Living Earwig, *Challia fletcheri* (Dermaptera: Pygidicranidae) and Phylogeny of Polyneoptera. *PLoS One* 7, e42056.

Wheeler, W.C., Whiting, M., Wheeler, Q.D., Carpenter, J.M., Carpenter, 2001. The Phylogeny of the Extant Hexapod Orders. *Cladistics* 17, 113–169.

Whitfield, J.B., Kjer, K.M., 2008. Ancient Rapid Radiations of Insects: Challenges for Phylogenetic Analysis. *Annu. Rev. Entomol.* 53, 449–472.

Whitfield, J.B., Lockhart, P.J., 2007. Deciphering ancient rapid radiations. *Trends Ecol. Evol.* 22, 258–265.

Whiting, M.F., Carpenter, J.C., Wheeler, Q.D., Wheeler, W.C., 1997. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol.* 46, 1–68.

Wilkinson, M., 1996. Majority-rule reduced consensus trees and their use in bootstrapping. *Mol. Biol. Evol.* 13, 437–444.

Wu, H.-Y., Ji, X.-Y., Yu, W.-W., Du, Y.-Z., 2014. Complete mitochondrial genome of the stonefly *Cryptoperla stilifera* Sivec (Plecoptera: Peltoperlidae) and the phylogeny of Polyneopteran insects. *Gene* 537, 177–183.

Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with

DOGMA. Bioinformatics 20, 3252–3255.

Yeates, D.K., Cameron, S.L., Trautwein, M., 2012. A view from the edge of the forest: recent progress in understanding the relationships of the insect orders. Australian Journal of Entomology 51, 79–87.

Yoshizawa, K., Johnson, K.P., 2005. Aligned 18S for *Zoraptera* (Insecta): phylogenetic position and molecular evolution. Mol. Phylogenet. Evol. 37, 572–580.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol 8, 28–36.

Zhang, D.-X., Hewitt, G.M., 1997. Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochem. Sys. Ecol. 25, 99–120.

Zhang, J., Zhou, C., Gai, Y., Song, D., Zhou, K., 2008. The complete mitochondrial genome of *Parafronurus youi* (Insecta: Ephemeroptera) and phylogenetic position of the Ephemeroptera. Gene 424, 18–24.

Zhang, Y.-Y., Xuan, W.-J., Zhao, J.-L., Zhu, C.-D., Jiang, G.-F., 2009. The complete mitochondrial genome of the cockroach *Eupolyphaga sinensis* (Blattaria: Polyphagidae) and the phylogenetic relationships within the Dictyoptera. Mol. Biol. Rep. 37, 3509–3516.



## 696 Figure legends

697 **Fig. 1.** Phylogenetic relationships of insect major orders using Bayesian inference reconstruction  
 698 based on mitochondrial genome data using the concatenated amino acid sequences of the  
 699 optimized\_taxa matrix (see Table 1). Filled circles indicate supported nodes; whereby black circles  
 700 represent Bayesian posterior probability (BPP) = 1, and grey circles BPP  $\geq$  0.95. Letters indicate  
 701 clades mentioned in the text (see Table 3)

702 **Fig. 2.** Mitochondrial genome maps of three sequenced Ephemeroptera species. The nearly complete  
 703 mitochondrial genome of (a) *Cloeon dipterum*, including coverage depth, the complete  
 704 mitochondrial genome of (b) *Baetis rutilocylindratus*, and the incomplete mitochondrial genome of  
 705 (c) *Habrophlebiodes zijinensis*. Transfer RNA genes are indicated by single-letter IUPAC-IUB  
 706 abbreviations for their corresponding amino acid. Protein coding genes and ribosomal RNA genes  
 707 are listed and colored in the following way: *atp6*, *atp8*, ATP synthase subunits 6 and 8 genes (violet);  
 708 *cob*, cytochrome oxidase *b* gene (red); *cox1-cox3*, cytochrome oxidase *c* subunit 1-3 genes (yellow);  
 709 *nad1-6*, *nad4L*, NADH dehydrogenase subunits 1-6 and 4L (light blue); *rrnS*, *rrnL*, small and large  
 710 ribosomal RNA subunits (orange); AT-rich region (brown). Genes located at the (+) strand appear  
 711 above the central line. Genes located on the (-) strand appear below the central line.

## 712 Tables

713 **Table 1.** Set of mitochondrial genomes with according GenBank accession numbers.

Order	Family	Species	Accession
Archaeognatha	Machilidae	<i>Pedetontus silvestrii</i>	NC_011717
Archaeognatha	Machilidae	<i>Petrobius brevistylis</i>	NC_007688
Archaeognatha	Machilidae	<i>Songmachilis xinxiangensis</i>	NC_021384
Archaeognatha	Machilidae	<i>Trigoniophthalmus alternatus</i>	NC_010532
Archaeognatha	Meinertellidae	<i>Nesomachilis australica</i>	NC_006895
Blattodea	Blattidae	<i>Periplaneta americana</i>	NC_016956
Blattodea	Blattidae	<i>Periplaneta fuliginosa</i>	NC_006076
Blattodea	Corydiidae	<i>Eupolyphaga sinensis</i>	NC_014274
Blattodea	Cryptocercidae	<i>Cryptocercus relictus</i>	NC_018132
Blattodea	Ectobiidae	<i>Blattella bisignata</i>	NC_018549
Blattodea	Ectobiidae	<i>Blattella germanica</i>	NC_012901
Collembola	Tetrodontophorinae	<i>Tetrodontophora bielanensis</i> <sup>a</sup>	NC_002735
Dermaptera	Pygidicranidae	<i>Challia fletcheri</i> <sup>a</sup>	NC_018538
Embioptera	Oligotomidae	<i>Aposthonia japonica</i> <sup>a</sup>	AB639034
Ephemeroptera	Ameletidae	<i>Ameletus</i> sp1	KM244682
Ephemeroptera	Baetidae	<i>Alainites yixiani</i>	NC_020034
Ephemeroptera	Baetidae	<i>Baetis rutilocylindratus</i>	XXXXXXXXXX
Ephemeroptera	Baetidae	<i>Cloeon dipterum</i>	XXXXXXXXXX
Ephemeroptera	Caenidae	<i>Caenis pycnacantha</i>	GQ502451
Ephemeroptera	Ephemerellidae	<i>Ephemerella</i> sp.	KM244691
Ephemeroptera	Ephemerellidae	<i>Vietnamella dabieshanensis</i>	NC_020036
Ephemeroptera	Ephemerellidae	<i>Vietnamella</i> sp.	KM244655
Ephemeroptera	Ephemeridae	<i>Ephemerella orientalis</i>	NC_012645
Ephemeroptera	Heptageniidae	<i>Epeorus</i> sp.	KM244708
Ephemeroptera	Heptageniidae	<i>Paegniodes cupulatus</i>	NC_020035
Ephemeroptera	Heptageniidae	<i>Parafronurus youi</i>	NC_011359
Ephemeroptera	Isonychiidae	<i>Isonychia ignota</i>	NC_020037
Ephemeroptera	Leptophlebiidae	<i>Habrophlebiodes zijingensis</i>	XXXXXXXXXX
Ephemeroptera	Potamanthidae	<i>Potamanthus</i> sp.	KM244674
Ephemeroptera	Siphonuridae	<i>Siphonurus immanis</i>	NC_013822
Ephemeroptera	Siphonuridae	<i>Siphonurus</i> sp.	KM244684
Ephemeroptera	Siphuriscidae	<i>Siphuriscus chinensis</i>	HQ875717
Ephemeroptera	Teloganodidae	sp.	KM244670; KM244703 <sup>b</sup>
Grylloblattodea	Grylloblattidae	<i>Grylloblatta sculleni</i>	DQ241796
Isoptera	Hodotermitidae	<i>Microhodotermes viator</i>	NC_018122
Isoptera	Kalotermitidae	<i>Neotermes insularis</i>	NC_018124
Isoptera	Rhinotermitidae	<i>Coptotermes lacteus</i>	NC_018125

Isoptera	Rhinotermitidae	<i>Heterotermes</i> sp.	NC_018127
Isoptera	Rhinotermitidae	<i>Reticulitermes flavipes</i>	NC_009498
Isoptera	Rhinotermitidae	<i>Reticulitermes hageni</i>	NC_009501
Isoptera	Rhinotermitidae	<i>Schedorhinotermes breinli</i>	NC_018126
Isoptera	Termitidae	<i>Drepanotermes</i> sp.	NC_018129
Isoptera	Termitidae	<i>Macrognathotermes errator</i>	NC_018130
Isoptera	Termitidae	<i>Macrotermes subhyalinus</i>	NC_018128
Isoptera	Termitidae	<i>Nasutitermes triodiae</i>	NC_018131
Isoptera	Termopsidae	<i>Porotermes adamsoni</i>	NC_018121
Isoptera	Termopsidae	<i>Zootermopsis nevadensis</i>	NC_024658
Mantodea	Caliridinae	<i>Leptomantella albella</i>	NC_024028
Mantodea	Mantidae	<i>Tamolania tamolana</i>	NC_007702
Mantophasmatodea	Mantophasmatidae	<i>Sclerophasma paresisensis</i>	NC_007701
Odonata	Calopterygidae	<i>Vestalis melania</i>	NC_023233
Odonata	Coenagrionidae	<i>Ischnura pumilio</i>	NC_021617
Odonata	Corduliidae	<i>Cordulia aenea</i>	JX963627
Odonata	Epiophlebiidae	<i>Epiophlebia superstes</i>	NC_023232
Odonata	Euphaeidae	<i>Euphaea formosa</i>	NC_014493
Odonata	Gomphidae	<i>Davidius lunatus</i>	NC_012644
Odonata	Gomphidae	<i>Ictinogomphus</i> sp.	KM244673
Odonata	Libellulidae	<i>Hydrobasileus croceus</i>	KM244659
Odonata	Libellulidae	<i>Orthetrum triangulare melania</i>	AB126005
Odonata	Pseudolestidae	<i>Pseudolestes mirabilis</i>	NC_020636
Orthoptera	Acrididae	<i>Acrida willemsei</i>	NC_011303
Orthoptera	Acrididae	<i>Calliptamus italicus</i>	NC_011305
Orthoptera	Acrididae	<i>Gomphocerus sibiricus</i>	NC_021103
Orthoptera	Acrididae	<i>Oedaleus decorus asiaticus</i>	NC_011115
Orthoptera	Gryllotalpidae	<i>Gryllotalpa orientalis</i> <sup>a</sup>	NC_006678
Orthoptera	Pneumoridae	<i>Tanaocerus koebelei</i>	NC_020777
Orthoptera	Prophalangopsidae	<i>Tarragoilus diuturnus</i>	NC_021397
Orthoptera	Pyrgomorphidae	<i>Atractomorpha sinensis</i>	NC_011824
Orthoptera	Rhaphidophoridae	<i>Troglophilus neglectus</i>	NC_011306
Orthoptera	Romaleidae	<i>Xyleus modestus</i>	NC_014490
Orthoptera	Tetrigidae	<i>Tetrix japonica</i>	NC_018543
Orthoptera	Tettigoniidae	<i>Gampsocleis gratiosa</i>	NC_011200
Orthoptera	Tettigoniidae	<i>Ruspolia dubia</i>	NC_009876
Phasmatodea	Bacillidae	<i>Bacillus rossius</i>	GU001956
Phasmatodea	Diapheromeridae	<i>Micadina phluctainoides</i>	NC_014673
Phasmatodea	Diapheromeridae	<i>Sipyloidea sipyilus</i>	AB477470
Phasmatodea	Heteropterygidae	<i>Heteropteryx dilatata</i>	NC_014680
Phasmatodea	Heteropterygidae	<i>Orestes mouhotii</i>	AB477462
Phasmatodea	Phasmatidae	<i>Entoria okinawaensis</i>	NC_014694
Phasmatodea	Phasmatidae	<i>Extatosoma tiaratum</i>	NC_017748

Phasmatodea	Phasmatidae	<i>Megacrania alpheus adan</i>	NC_014688
Phasmatodea	Phasmatidae	<i>Neohirasea japonica</i>	AB477469
Phasmatodea	Phasmatidae	<i>Phobaeticus serratipes</i>	NC_014678
Phasmatodea	Phasmatidae	<i>Phraortes</i> sp. <sup>a</sup>	NC_014705
Phasmatodea	Phasmatidae	<i>Ramulus irregulariterdentatus</i>	NC_014702
Phasmatodea	Timematidae	<i>Timema californicum</i>	DQ241799
Plecoptera	Perlidae	<i>Dinocras cephalotes</i>	NC_022843
Plecoptera	Perlidae	<i>Kamimuria wangi</i>	NC_024033
Plecoptera	Perlidae	<i>Togoperla</i> sp.	KM409708
Plecoptera	Pteronarcyidae	<i>Pteronarcys princeps</i>	NC_006133
Zygentoma	Lepidotrichidae	<i>Tricholepidion gertschi</i>	NC_005437
Zygentoma	Lepismatidae	<i>Thermobia domestica</i>	NC_006080
Zygentoma	Nicoletiidae	<i>Atelura formicaria</i>	NC_011197

714 <sup>a</sup>, Taxa were excluded in the optimized\_taxa matrix (see 2. Material and methods).

715 <sup>b</sup>, Represents sequence consisting of two genomic fragments (contigs) from the same sample.

**Table 2.** Overview of data matrices, including the best-fitting partitioning and model schemes, and amino acid (aa) sequence length.

Matrix	Model	Length [aa]
<u>all_taxa</u>		
partition_1: <i>atp6, atp8, nad2, nad3, nad6</i>	MTZOA + $\Gamma$ + I + F	840
partition_2: <i>cox1</i>	MTZOA + $\Gamma$ + I	528
partition_3: <i>cob, cox2</i>	MTZOA + $\Gamma$ + I + F	608
partition_4: <i>cox3</i>	MTART + $\Gamma$ + I	268
partition_5: <i>nad1, nad4, nad4L, nad5</i>	MTZOA + $\Gamma$ + I + F	1431
<u>optimized_taxa</u>		
partition_1: <i>atp6, atp8, nad2, nad3, nad6</i>	MTZOA + $\Gamma$ + I + F	843
partition_2: <i>cox1</i>	MTZOA + $\Gamma$ + I	529
partition_3: <i>cob, cox2, cox3</i>	MTZOA + $\Gamma$ + I + F	878
partition_4: <i>nad1, nad4, nad4L, nad5</i>	MTZOA + $\Gamma$ + I + F	1423

**Table 3.** Support for the three hypotheses and node support for clades of interest. Node support is based on Bayesian inference (BI) and maximum likelihood (ML) analyses of two data matrices: optimized\_taxa and all\_taxa, whereby optimized\_taxa had 5 terminal nodes removed based on the occurrence of rogue taxa and long branch attraction. Asterices indicate hypothesis/node support (\*\* = Bayesian posterior probability (BPP)  $\geq 0.95$  and Bootstrap support (BS) = 90%) \* = BPP  $\geq 0.90$  and BS  $\geq 80\%$ ). Superorders are underlined, and suborders are italicised. Letters in brackets refer to monophyletic clades in the Fig. 1.

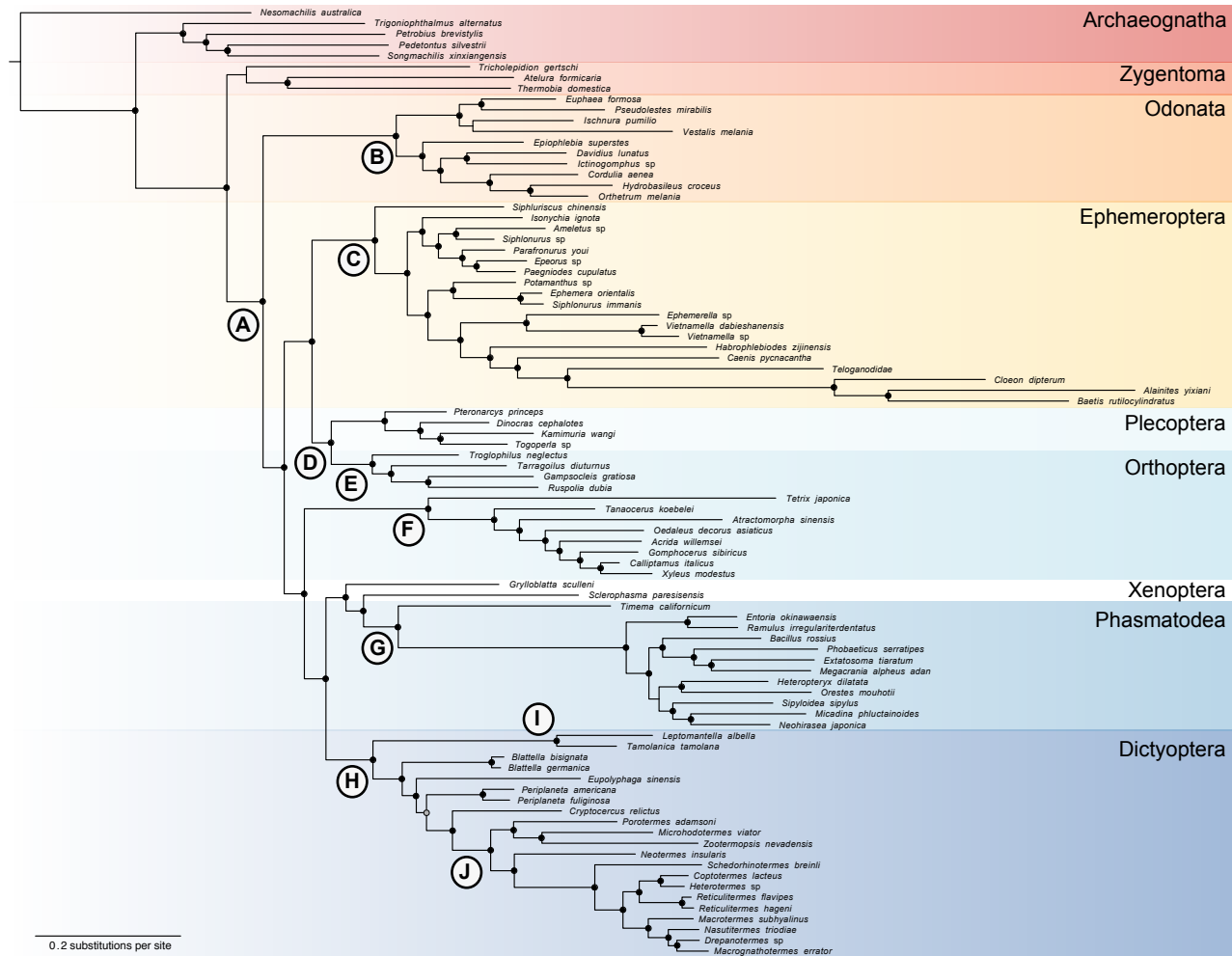
Hypothesis/Clade	BI		ML	
	optimized_taxa	all_taxa	optimized_taxa	all_taxa
Palaeoptera hypothesis				
Ephemeroptera hypothesis				
Odonata hypothesis	**			
Winged insects (A)	**			
<u>Palaeoptera</u>				
Odonata (B)	**	**	**	**
Ephemeroptera (C)	**		**	
<u>Polyneoptera</u>			*	
Plecoptera (D)	**	**	**	**
Orthoptera				
<i>Ensifera</i> (E)	**	**	**	
<i>Caelifera</i> (F)	**	**	**	**
Phasmatodea <sup>a</sup> (G)	**	**	**	**
<u>Dictyoptera</u> (H)	**	**	**	**
Mantodea (I)	**	**	**	**
Isoptera (J)	**	**	**	**

<sup>a</sup>, Includes *Aposthonia japonica* (Embioptera) in reconstructions using the all\_taxa matrix.

**Table 4.** Overview of sequenced ephemeropteran mitochondrial genomes. Total sequence length in base pairs (bp) and individual nucleotide compositions of mayfly mitochondrial genomes calculated based on the whole available sequences and the ones corrected for the incomplete mitochondrial genomes (see 2. Material and methods; for the complete list of all ephemeropteran mitochondrial genomes and Supplementary Table S2).

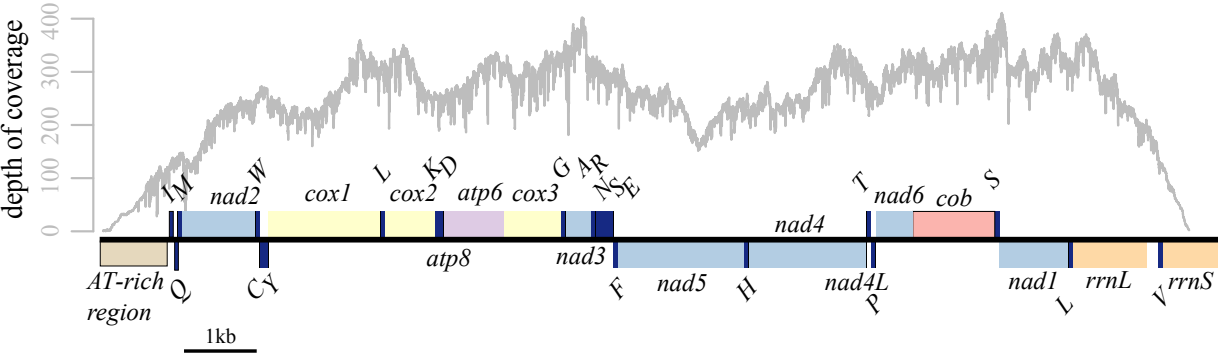
Family	Species	Length [bp]	A%	C%	G%	T%	GC%	AT%
Baetidae	<i>Baetis</i>	14,883 <sup>a</sup>	27.3	19.8	20.1	32.8	39.9	60.1
	<i>rutilocylindratus</i>	12,234	26.7	19.8	20.4	33.2	40.2	59.9
Baetidae	<i>Cloeon dipterum</i>	14,355	30.6	14.7	16.3	38.4	30.9	69.0
		12,197	29.9	14.5	16.9	38.7	31.4	68.6
Leptophlebiidae	<i>Habrophlebiodes</i>	15,407	33.8	20.2	11.0	35.1	31.2	68.9
	<i>zijinensis</i>	12,417	33.6	20.6	11.5	34.3	32.1	67.9

<sup>a</sup>, Indicates complete mitochondrial genomes; rest are incomplete mitochondrial genomes.

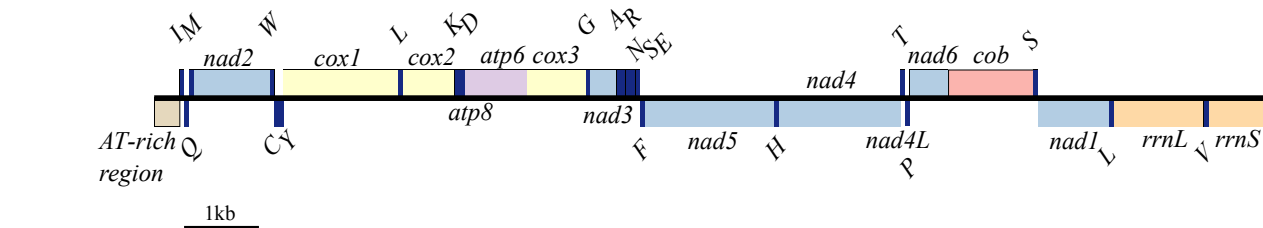




(a) *Cloeon dipterum* L. 1761



(b) *Baetis rutilocylindratus* Wang, Qin, Cheng & Zhou, 2011



(c) *Habrophlebiodes zijinensis* Gui, Zhang & Wu, 1996

