

Article

Comparative Mitogenomic Analysis of Heptageniid Mayflies (Insecta: Ephemeroptera): Conserved Intergenic Spacer and tRNA Gene Duplication

Ran Li, Zhiming Lei, Wenjuan Li, Wei Zhang and Changfa Zhou *

The Key Laboratory of Jiangsu Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China; 90716@njnu.edu.cn (R.L.); 201202011@njnu.edu.cn (Z.L.); 201202051@njnu.edu.cn (W.L.); 171201004@njnu.edu.cn (W.Z.)

* Correspondence: zhouchangfa@njnu.edu.cn; Tel.: +86-025-8589-1292

Simple Summary: Heptageniidae is one of the most abundant and widespread families of mayflies, with more than 600 described species and distributed mainly in the Holarctic, Oriental, and Afrotropical regions. Previous phylogenetic analyses in this family mainly focused on morphological characters and were restricted to one or several genes. Ten complete mitochondrial genomes from this family are reported here. A conserved intergenic spacer and tRNA gene duplication in Heptageniidae mitogenomes may be regarded as ancient features and provide evidence for the phylogenetic analysis within Ephemeroptera.

Abstract: Large intergenic spacers and tRNA gene duplications have been reported in several insect groups, although little is known about mitogenomes of mayflies. Here, we determined complete mitogenomes of ten heptageniid species and systemically analyzed their mitogenomic features. Both a conserved intergenic spacer (IGS) and *trnM* duplication were detected in those mitogenomes. The IGS, which was observed in heptageniids, could be further folded into a sTable Stem-loop structure. The tRNA gene duplication was found in almost all analyzed mitogenomes, and a unique gene block *trnI-trnM-trnQ-trnM-ND2* was also discovered. Our analysis demonstrates that the heptageniid gene arrangement pattern can be explained by the tandem duplication-random loss (TDRL) model. Phylogenetic analyses using both Bayesian inference (BI) and maximum likelihood (ML) methods based on the nucleotide and amino acid sequence data recovered the genus *Epeorus* as monophyletic with strong support. Our results provide a better understanding of mitogenomic evolution in Heptageniidae, as well as novel molecular markers for species identification of mayflies.

Keywords: Ephemeroptera; Heptageniidae; mitochondrial genome; intergenic spacer; gene rearrangement



Citation: Li, R.; Lei, Z.; Li, W.; Zhang, W.; Zhou, C. Comparative Mitogenomic Analysis of Heptageniid Mayflies (Insecta: Ephemeroptera): Conserved Intergenic Spacer and tRNA Gene Duplication. *Insects* **2021**, *12*, 170. <https://doi.org/10.3390/insects12020170>

Academic Editor: Fabrice Legeai
Received: 19 January 2021
Accepted: 10 February 2021
Published: 16 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The mitochondrial genome (mitogenome) is a double-stranded DNA molecule with 14–20 kb long and accounts for 1–2% of the total genome found in cells [1,2]. A complete insect mitogenome generally consists of 37 genes, specifying 13 protein-coding genes (PCGs), 22 transfer RNAs (tRNAs), 2 ribosomal RNAs (rRNAs), and a non-coding control region (CR) [3,4]. The sequence of insect mitogenome is a commonly suitable molecular marker for phylogenetic and evolutionary studies, due to its maternal inheritance, high nucleotide substitution rate, simple genetic structure, and lack of recombination [2,5–7]. Dramatic variation in gene organization was found in some insect groups. The gene rearrangements and/or intergenic spacer (IGS) regions appear to be regarded as an additional informative character for phylogenetic reconstruction among some groups [8–10].

Mayflies (Ephemeroptera) are one of the most archaic extant winged insect lineages. The amphibiotic insects with highly reduced terrestrial adult stage have evolved con-

siderably for about 300 Mya [11]. Compared with many other insect orders, the available data of mitogenome from Ephemeroptera are quite scarce. There are currently only 38 complete or partial mitogenomes (among 12 families) deposited to the GenBank (<https://www.ncbi.nlm.nih.gov/>; 20 December 2020). The relative positions of PCGs and rRNAs are usually stable and conserved among different ephemeropteran species, however, tRNAs display different degrees of variation. In the mitogenome of *Siphuriscus chinensis* (Siphuriscidae), for example, *trnK* is duplicated and translocated at the upstream of *trnE*, which generates a new gene cluster (*trnS1-trnK-trnE*) compared with the putative ancestral gene order [12]. The rearrangement of several tRNAs is found in *Alainites yixiani* (Baetidae) with a gene cluster of *trnI-CR-trnC-trnQ-trnY-trnM-ND2-trnW* [13]. Recent studies have also demonstrated that *trnI* of Ephemerellidae is inverted from the downstream of CR in the H-strand to the upstream of CR in the J-strand, forming an identical gene block (*trnI-CR-trnQ-trnM*) [14]. In addition, the duplication of *trnM* gene is always observed in Heptageniidae, except *Paegniodes cupulatus* [13,15,16]. It is apparent that more mitogenomes from diverse groups of mayflies are in demand to well understand the mechanism of these gene rearrangements in the following studies.

Heptageniidae is a family with abundant species-diversity, and more than 600 species belonging to three subfamilies (Ecdyonurinae, Heptageniinae and Rhithrogeninae) have been described [17]. The species are widely distributed in the Palaearctic Region, and their nymphs generally inhabit the surface of rocks, leaves and vegetations in distinct kinds of lotic freshwater habitats (rivers, lakes and streams) [18]. Because of their sensitivity to pollution, many species have been used as excellent models for aquatic biological monitoring and biodiversity researches [19,20]. Over the past decade, the taxonomic studies of this family are mainly based on morphological characters of different life stages [17,21–23]. However, it remains a challenge to identify the heptageniid mayflies because of a lack of dependable characters for both two major stages (nymph and adult) [19]. Consequently, it is generally not possible to obtain definite species names in the research in which immature individuals of the sample account for a large proportion [20,24]. Furthermore, phylogenetic relationships among different groups in this family are still controversial with limited molecular information [25,26]. Only six reliable mitogenome sequences have previously been published for Heptageniidae to date, with four mitogenomes belonging to the genus *Epeorus*, one mitogenome of the genus *Parafonurus*, and one of the genus *Paegniodes*. At the same time, most previous related studies tended to focus on one species and its phylogenetic placement in Ephemeroptera [13,15,16].

Accordingly, we used next-generation sequencing method to obtain mitogenomes for 10 heptageniid mayflies, which represented all three subfamilies: five species from Rhithrogeninae, four from Ecdyonurinae, one from Heptageniinae. We compared the newly generated mitogenomes to the six previously reported sequences, described the structural and compositional features of heptageniid mitogenomes and analyzed the intergenic spacers to access possible evolutionary mechanisms. We also analyzed gene rearrangement patterns and discussed possible rearrangement processes and mechanisms in detail. Finally, the phylogeny of Heptageniidae was reconstructed combining all available mitogenomes.

2. Materials and Methods

2.1. Sample Collection and DNA Extraction

Ten specimens of Heptageniidae were collected between 2019 and 2020 from different sampling sites in China. Detailed information is shown in Supplementary Materials Table S1. Morphological identification of the specimens was conducted by C.F. Zhou based on the key diagnostic features. All samples were deposited in 100% ethanol at -20°C and then cataloged in the voucher collection of Mayfly Museum of Nanjing Normal University, Jiangsu Province, China. Total genomic DNA extraction was performed from the leg tissue using TIANamp Genomic DNA Kit (TIANGEN, China) according to the manufacturer's protocol. DNA quality and concentration were measured on the Nanodrop

2000 spectrophotometer and visualized on 1.0% agarose gel. The qualified DNA was preserved at -20°C and used for sequencing.

2.2. Mitogenome Sequencing and Assembly

Genomic DNA of all samples was sent to Personalbio Inc. (Shanghai, China) for library construction and next-generation sequencing (NGS). One library (Insert size of 400 bp) was prepared for each DNA sample using the TruSeq™ DNA Sample Prep Kit (Illumina, USA). All constructed libraries were then sequenced as 150 bp paired-end on a full run (2×150 PE) using the Illumina NovaSeq platform. After trimming the adapter contamination and removing short and low-quality reads, more than 4 GB (30–41 million reads) clean data for each sample was used in de novo assembly (Table S2). The complete circular mitogenomes were assembled using NovoPlasty 4.0 (k-mer = 33), with the *COI* gene of *S. chinensis* as a seed sequence (Accession number: HQ875717).

2.3. Gene Annotation and Bioinformatic Analysis

All newly determined mitogenomes were firstly annotated using MitoZ 2.4 pipeline [27]. The parameter setting for Genetic Code 5 (invertebrate) was selected, and the Arthropoda database was used to select reference sequences. Secondary structure of tRNAs was predicted by tRNA scan-SE 2.0.2 and ARWEN 1.2 to confirm their accuracy [28,29]. The boundaries of two rRNAs were identified using the ClustalW algorithm in MEGA 7.0 based on alignments of other available heptageniid mitogenomes [30]. PCGs with non-canonical start and stop codons were further adjusted and corrected manually by translating into amino acids.

The nucleotide composition of all components and the relative synonymous codon usage (RSCU) of PCGs were estimated using MEGA 7.0. The base composition values (AT- and GC-skews) were calculated using the following formulas: $\text{AT-skew} = (A - T)/(A + T)$ and $\text{GC-skew} = (G - C)/(G + C)$ [31]. The sliding window analysis (a sliding window of 200 bp and step size of 20 bp), and the nucleotide diversity (P_i) of 13 PCGs among 16 mitogenomes of Heptageniidae were conducted using DnaSP 6.0 [32]. The numbers of synonymous substitutions (Ks) and non-synonymous substitutions (Ka), and the ratios of Ka/Ks for each PCG were also measured in the software DnaSP 6.0. The genetic distances based on Kimura-2-parameter among the 16 mitogenomes were analyzed by MEGA 7.0.

2.4. Phylogenetic Analysis

A total of 17 mitogenomes of Ephemeroptera were used for the phylogenetic analyses, including 16 species of Heptageniidae (10 newly sequenced mitogenomes and six downloaded from the GenBank) as the ingroups and one species of Siphuriscidae (*S. chinensis*) as outgroup (Table 1). Phylogenetic analyses were reconstructed on the concatenated datasets of 13 PCGs at both amino acid and nucleotide levels with Maximum likelihood (ML) and Bayesian analysis (BI) methods. Individual PCG sequences from all the 17 species (excluding the start and stop codons) were aligned individually with codon-based multiple alignments using the software MUSCLE 3.8.31 [33]. The individual aligned sequences were then concatenated by PhyloSuite [34], and conserved regions were identified by the program Gblock 0.91b [35]. The optimal partitioning strategy for two datasets was inferred using PartitionFinder 2 in the PhyloSuite program [36], under a greedy search algorithm with linked branch lengths based on the Bayesian information criterion (BIC) (Tables S3 and S4). Bayesian phylogenetic analyses were conducted using MrBayes 3.2.6 with default settings as implemented in the CIPRES [37,38], and ran for 10^6 generations sampling every 1000 generations. The first 25% of generations were removed as burn-in, and the average standard deviation of split frequencies < 0.01 was considered to reach convergence. ML analyses were carried out by RAxML 8.2.0 with a GTRGAMMAI model. The branch support for each node was evaluated with 1000 bootstrap replicates [39].

Table 1. The species information used in phylogenetic analysis.

Subfamily	Species	Size (bp)	GenBank #
Rhithrogeninae	<i>Epeorus herklotsi</i> 01	15,502	MG870104
	<i>Epeorus herklotsi</i> 02	15,499	MH752075
	<i>Epeorus</i> sp. JZ-2014	15,338	KJ493406
	<i>Epeorus</i> sp. MT-2014	15,456	KM244708
	<i>Epeorus montanus</i>	15,472	This study
	<i>Epeorus melli</i>	15,490	This study
	<i>Epeorus bifurcatus</i>	15,466	This study
	<i>Epeorus pellucidus</i>	15,435	This study
	<i>Paegniodes cupulatus</i> 01	15,715	HM004123
	<i>Paegniodes cupulatus</i> 02	15,721	This study
Ecdyonurinae	<i>Parafronurus youi</i>	15,481	EU349015
	<i>Afronurus furcata</i>	15,334	This study
	<i>Afronurus drepanophyllus</i>	15,242	This study
	<i>Notacanthurus lamellosus</i>	15,693	This study
	<i>Notacanthurus maculosus</i>	15,524	This study
Heptageniinae	<i>Heptagenia ngi</i>	15,495	This study
Outgroup			
Siphuriscidae	<i>Siphuriscus chinensis</i>	16,616	HQ875717

3. Results and Discussion

3.1. Genomic Organization and Composition

We successfully obtained the complete mitogenomes of ten mayflies of the family Heptageniidae. Sequences were deposited in GenBank with the accession numbers: MW381291–MW381300. The total length of all new mitogenomes was well within the range of known complete mitogenomes in Ephemeroptera, with sizes from 14,589 bp in *A. yixiani* to 16,616 bp in *S. chinensis* [12,13]. The lengths of PCGs, tRNAs, rRNAs, and control regions (CRs) for the ten mitogenomes and other six reported mitogenomes of Heptageniidae are compared in Figure 1. Size variation of mitogenomes was mainly due to the difference in the size of the CRs. Nine of the newly determined mitogenomes also contained 38 genes: 13 PCGs, 23 tRNAs, and 2 rRNAs (Figure 2), and an additional *trnM* gene was identified in most Heptageniidae. The mitogenome of *P. cupulatus* encodes the typical 37 genes, which is consistent with a previously reported sequence. The accuracy of the existing mitogenome of *P. cupulatus* sequenced using Sanger technology was verified by our new generated sequence [16]. Most of the genes (9 PCGs and 15 (14) tRNAs) were encoded on the major strand (J-strand), while the remaining genes were on the minor strand (N-strand) (Figure 2).

As is typical for insects, the analyzed mitogenomes had a biased nucleotide composition with A + T content ranging from 63.68% (*Afronurus drepanophyllus*) to 66.77% (*Notacanthurus maculosus*) (Table 2). For most species, the highest A + T content was found in CRs (11 species), and only six species showed the highest A + T content in rRNAs. The relative numbers of A to T and G to C were measured by AT-skew and GC-skew of the base composition in nucleotide sequences. The results of skewness statistics showed that the AT-skews were slightly negative (−0.002 to −0.046) in complete mitogenomes among heptageniids, while GC-skews were obviously negative (−0.176 to −0.245). Meanwhile, PCGs of all mitogenomes also had a negative AT-skew and GC-skew, which indicated that T and C content is significantly more abundant than A and G.

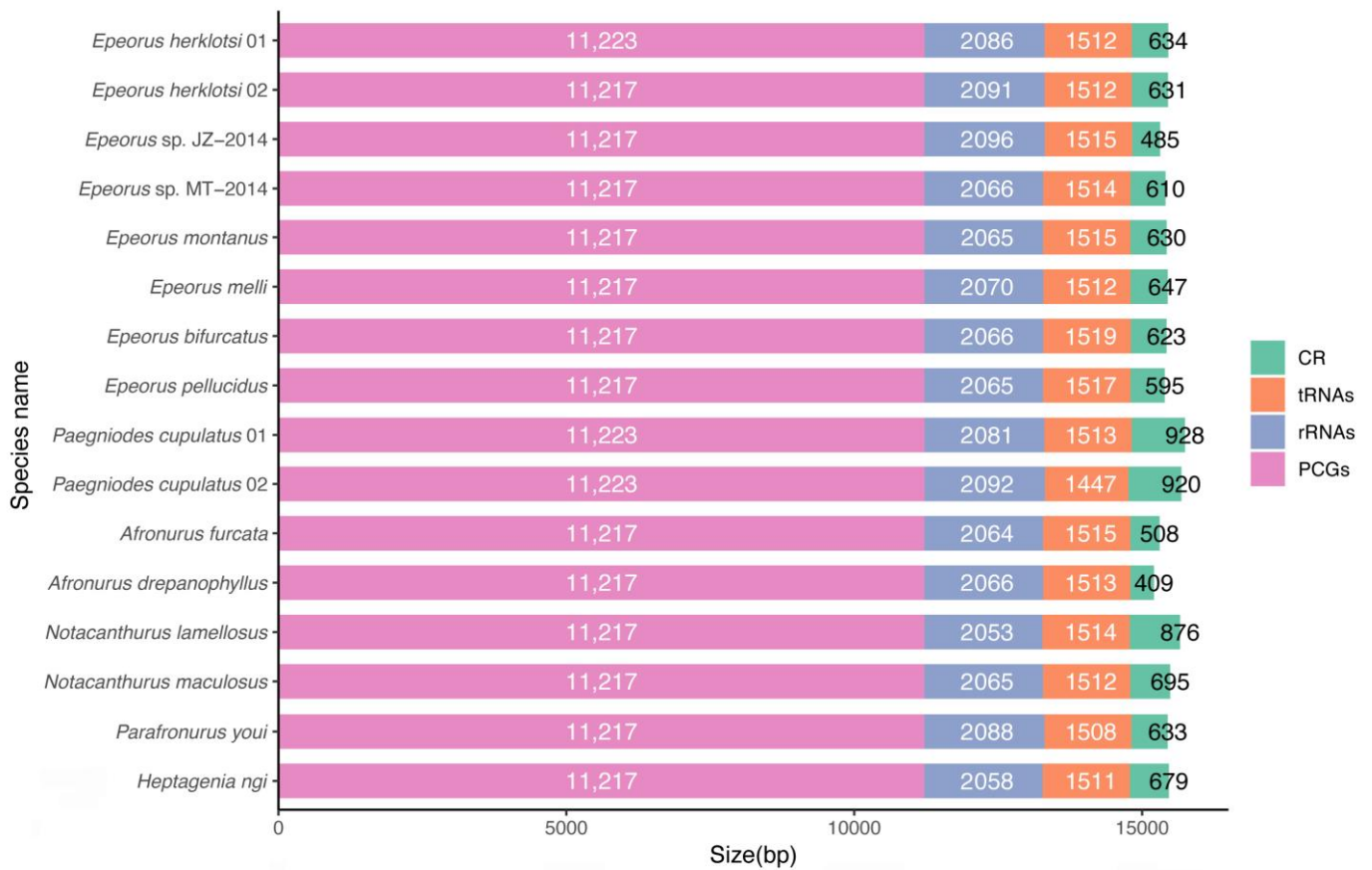


Figure 1. The length of protein-coding genes (PCGs), transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and control regions (CRs) among 16 heptageniid mitogenomes.

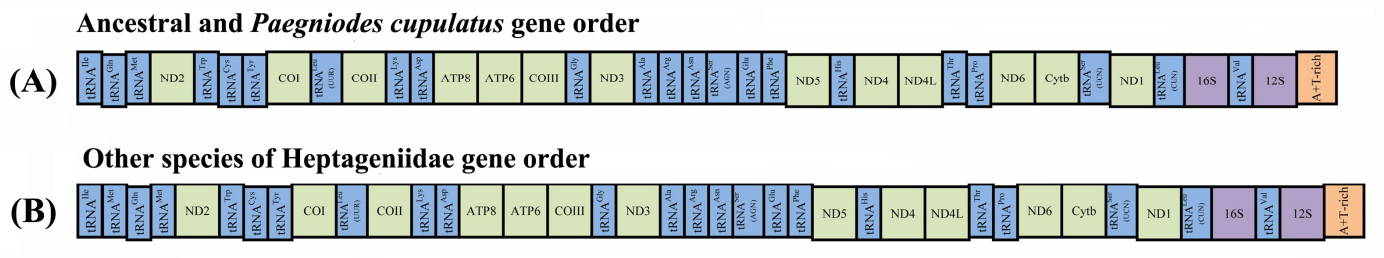


Figure 2. (A) The ancestral and *Paegniodes cupulatus* gene order. (B) The gene order of other heptageniid mitogenomes.

Table 2. Base composition and skewness of heptageniid mitogenomes.

Species	A+T (%)					AT-Skew					GC-Skew				
	All	PCGs	tRNAs	rRNAs	CR	All	PCGs	tRNAs	rRNAs	CR	All	PCGs	tRNAs	rRNAs	CR
<i>Epeorus herklotsi</i> 01	65.66	65.10	65.74	65.77	74.45	−0.002	−0.196	−0.018	0.007	0.017	−0.245	−0.033	0.131	0.311	−0.086
<i>Epeorus herklotsi</i> 02	65.71	65.15	65.87	65.61	74.80	−0.004	−0.197	−0.020	0.007	0.008	−0.243	−0.031	0.140	0.305	−0.082
<i>Epeorus</i> sp. JZ-2014	64.60	64.03	64.62	66.03	72.78	−0.002	−0.201	−0.021	0.012	0.042	−0.209	−0.026	0.123	0.270	0.000
<i>Epeorus</i> sp. MT-2014	64.07	62.53	65.19	67.18	78.03	−0.010	−0.197	−0.023	0.027	0.076	−0.232	−0.023	0.112	0.292	−0.164
<i>Epeorus montanus</i>	64.86	63.52	64.75	67.31	78.89	−0.019	−0.205	−0.019	0.026	0.082	−0.210	−0.021	0.105	0.262	−0.128
<i>Epeorus melli</i>	65.73	65.28	65.01	65.70	74.34	−0.004	−0.199	−0.019	0.004	−0.040	−0.245	−0.022	0.127	0.313	−0.084
<i>Epeorus bifurcatus</i>	65.01	63.72	65.64	67.28	78.01	−0.006	−0.194	−0.021	0.026	0.082	−0.245	−0.023	0.111	0.281	−0.255
<i>Epeorus pellucidus</i>	66.39	65.56	66.78	67.36	76.64	−0.028	−0.194	−0.011	0.041	0.083	−0.215	−0.017	0.115	0.279	−0.036
<i>Paegniodes cupulatus</i> 01	65.59	65.54	65.50	66.41	63.15	−0.009	−0.209	−0.007	0.020	0.154	−0.202	−0.033	0.138	0.313	0.099
<i>Paegniodes cupulatus</i> 02	64.73	64.62	66.14	66.59	58.70	−0.005	−0.213	−0.018	0.025	0.244	−0.203	−0.036	0.135	0.296	−0.011
<i>Parafronurus youi</i>	66.38	67.04	65.38	66.19	57.03	−0.016	−0.199	−0.020	0.017	−0.003	−0.220	−0.018	0.111	0.309	−0.022
<i>Afronurus furcata</i>	64.66	64.85	63.23	65.50	61.22	0.008	−0.184	−0.002	−0.004	0.068	−0.214	−0.021	0.102	0.326	0.096
<i>Afronurus drepanophyllus</i>	63.68	63.26	64.31	64.67	66.26	−0.021	−0.204	−0.026	0.021	0.063	−0.176	−0.025	0.115	0.299	0.087
<i>Notacanthurus lamellosus</i>	66.26	66.51	64.46	66.68	64.95	−0.046	−0.193	−0.023	0.065	−0.072	−0.167	−0.015	0.141	0.249	−0.225
<i>Notacanthurus maculosus</i>	66.77	67.03	65.01	67.17	64.46	−0.014	−0.200	−0.030	0.025	0.045	−0.219	−0.010	0.127	0.298	−0.296
<i>Heptagenia ngi</i>	64.12	64.03	62.81	63.95	68.78	−0.005	−0.196	−0.022	−0.008	−0.058	−0.177	0.000	0.121	0.275	−0.094

3.2. Protein-Coding Genes

All 13 PCGs were detected in the newly generated mitogenomes and their general characters were similar with the total length of 11,217 bp and 11,223 bp (Figure 1). The orientations of the PCGs were identical to other available mitogenomes of mayflies: four genes (*ND1*, *ND4*, *ND4L* and *ND5*) were encoded on the N-strand, and the remaining nine were on the J-strand. The majority of PCGs used the typical start codons ATN (ATA, ATG, ATT and ATC) with the exception of *COI* and *ND5* in most species, which used ACC and GTG (Table S5). Most PCGs terminated with the conventional TAA and TAG as stop codons while three genes (*COII*, *ND4* and *ND5*) terminated as truncated T. Incomplete terminating codons are a common phenomenon, related to post-transcriptional modification during mRNA maturation [40,41]. As shown in Figure 3, the relative synonymous codon usage (RSCU) was generally similar with each other in newly sequenced mitogenomes. Overall codon usage analysis showed that the codons ending with A or T were more preferred. The five most frequently used codons were UUA (Leu2), UCU (Ser2), CCU (Pro), GCU (Ala) and UGU (Cys), and the highest value of RSCU was 3.54 in *N. maculosus* (Figure 3). Our results reveal that the size, orientation and RSCU of PCGs are relatively conserved among the Heptageniidae.

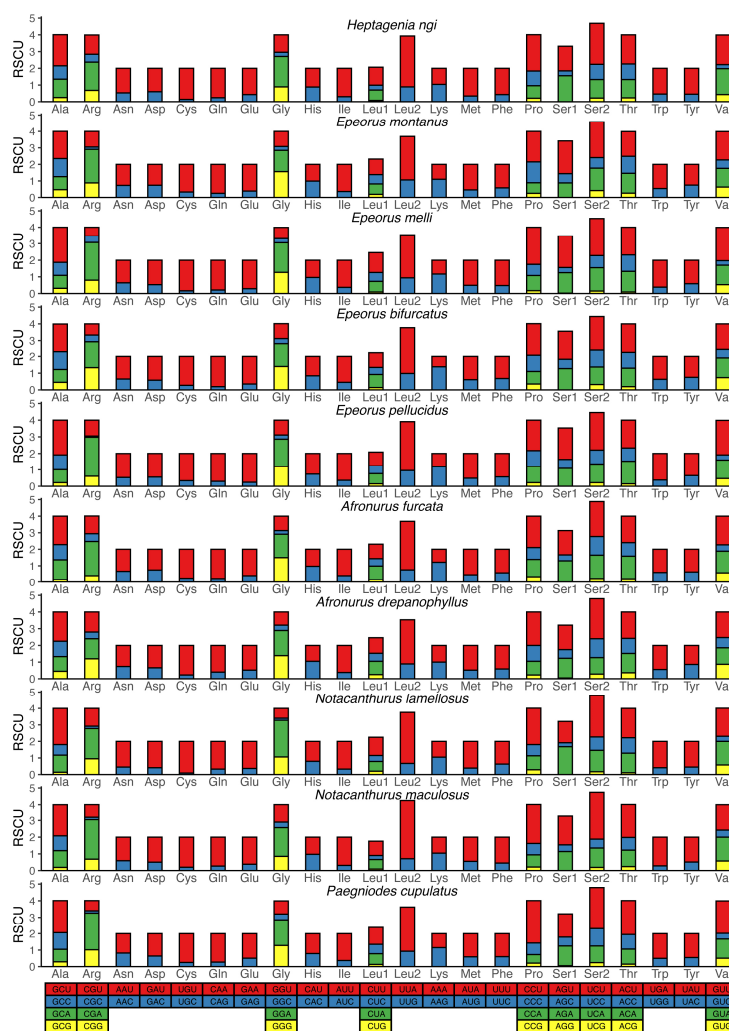


Figure 3. Relative synonymous codon usage (RSCU) of ten heptageniid mitogenomes. Codon families are provided on the X-axis along with the different combinations of synonymous codons that code for that amino acid. RSCU is defined on the Y-axis.

Nucleotide diversity was analyzed with a sliding window of the 13 aligned PCGs (Figure 4A). *ND6* ($P_i = 0.345$) and *ND2* ($P_i = 0.306$) had apparently higher nucleotide diversity than other genes, while *COI* had the lowest P_i of 0.173. Analysis of pairwise genetic distance showed similar results with *ND6* (0.685) and *ND2* (0.555) evolving relatively faster, while *COI* (0.233) and *COII* (0.239) were slower (Figure 4B). Average K_a/K_s ratios were estimated to investigate evolutionary rates of mitogenome PCGs. Ratios ranged from 0.036 for *COI* to 0.324 for *ND6*, which indicated that all PCGs were under purifying selection. Our results showed *ND6* and *ND2* exhibited relaxed purifying selection, while *COI* was under the strongest purifying selection (Figure 4B).

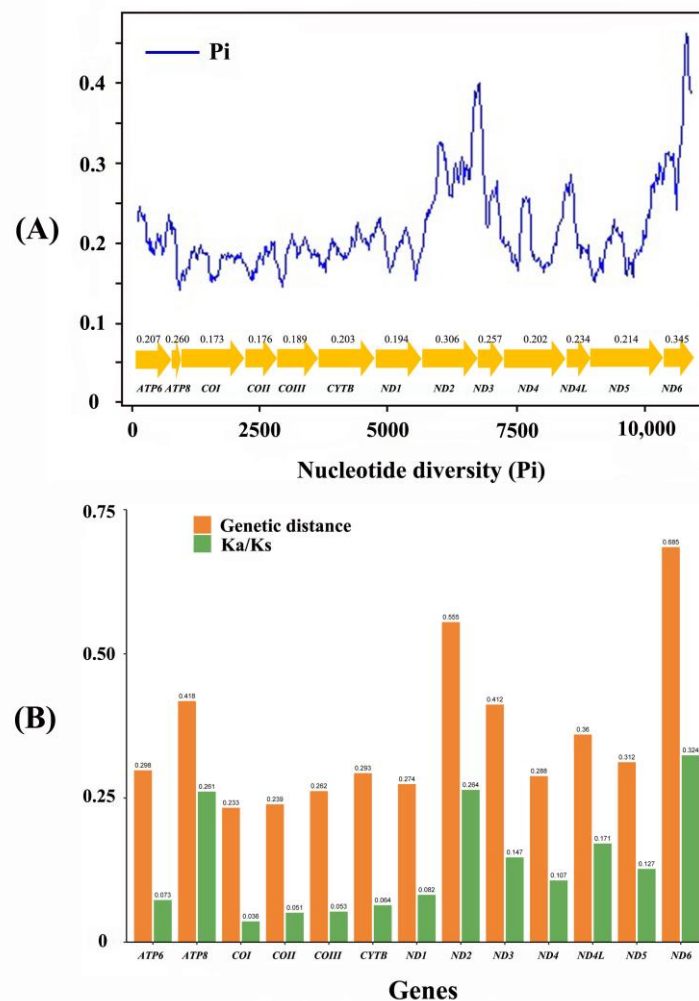


Figure 4. (A) Sliding window analysis based on 13 aligned PCGs. The blue line shows the value of nucleotide diversity P_i . The gene names and P_i values are shown in the graph. (B) Genetic distance (on average) and non-synonymous (K_a) to synonymous (K_s) substitution rates of 13 PCGs among 16 heptageniids.

Analyses of nucleotide diversity, genetic distance and evolutionary rate are useful for designing specific markers among different groups, especially in the taxa with highly variable morphological characters. Our comprehensive analysis showed that *COI* had the lowest evolution rate and evolves under comparative relaxed purifying selection, two genes (*ND6* and *ND2*) exhibited faster evolution rate and diversity than other PCGs, which is inconsistent with previous studies that *COI* was usually considered as one useful marker for species identification and phylogenetic analysis in closely-related taxa [42,43]. Overall, *ND6* and *ND2* might be two potential markers for identifying cryptic species, reconstructing phylogenetic trees and phylogeographic analysis in Heptageniidae.

3.3. Ribosomal and Transfer RNAs

Both rRNAs (*rrnL* and *rrnS*) were found in the mitogenomes of all ten heptageniid mayflies, encoded on the N-strand and located between the *trnL1* and *trnV*, and between *trnV* and the CR, respectively. Among all analyzed mitogenomes of Heptageniidae, *rrnL* length ranged from 1278 bp in *P. cupulatus* to 1291 bp in *Epeorus* sp. JZ 2014 (Figure 1). The longest *rrnS* was shown in two sequences of *P. cupulatus* and the shortest was in *Notacanthurus lamellosus*. A + T content of the rRNAs ranged from 63.95% to 67.36%, which exhibited a high A + T bias (Table 2). All 22 typical tRNAs found in insect mitogenomes were observed in the newly sequenced mitogenome of *P. cupulatus*, and an additional *trnM* was also found in the remaining nine mitogenomes. The length of all tRNAs ranged from 61 to 71 bp and could be folded into a typical clover-leaf secondary structure with the exception of *trnS1* due to a lack of the DHU arm. Loss of the DHU arm in *trnS1* is a common feature in insect mitogenomes [44].

3.4. Non-Coding Regions

Different numbers and sizes of non-coding regions are observed in metazoan mitogenomes, which results in variable size mitogenome between species [45]. Of the non-coding regions, the largest one is usually thought to be the control region, which contains signals for regulation and initiation of mitochondrial DNA transcription and replication [46,47]. The CRs of heptageniid mitogenomes were located in the conserved position between *rrnS* and *trnI* genes (Figure 2). The length of *P. cupulatus* 02 was longer than that of other species, resulting in having the longest mitogenome (Figure 1). There were large length variations across species, even closely related ones, which supports a previous study [48]. Rapid variation of CRs seemed to provide information for species evolution, but its internal mechanisms need deeper examination.

The most remarkable genomic feature was the presence of one intergenic spacer (IGS) region between *trnA* and *trnR* in all heptageniid mitogenomes. The IGSs of different species had different lengths (32–47 bp) and base compositions, but had a similar secondary structure. As shown in Figure 5, all IGSs could be folded into a stem-loop structure. The features mentioned above (stem-loop hairpin structure, a size of approximately 30 nucleotides, and a stem with complementary bases) were similar to the mitochondrial origins for replication of L-strand (OL) that existed in vertebrate mitogenomes [49,50]. This OL-like region is only observed in heptageniids, but not found in other families of Ephemeroptera. We hence assumed that the conserved region is a molecular synapomorphy of the family Heptageniidae. This region is commonly found in vertebrate mitogenomes, however, it has not been reported in insect mitogenomes so far [51,52].

3.5. Gene Arrangement

Across the order Ephemeroptera, several gene rearrangement events have been observed, such as in *S. chinensis*, *A. yixiani* and *C. fusca* [12–14]. Comparative analysis of the heptageniid mitogenomes indicates that the *trnM* duplication is unique to the family. The ancestral gene composition (22 tRNAs), however, is still recorded in the mitogenome of *P. cupulatus*. Interestingly, only this species shows the ancestral arrangement pattern. However, whether this event was identical with another species of this genus of *Paegniodes* (*P. dao*) remains unclear, due to the lack of sequence information for that group.

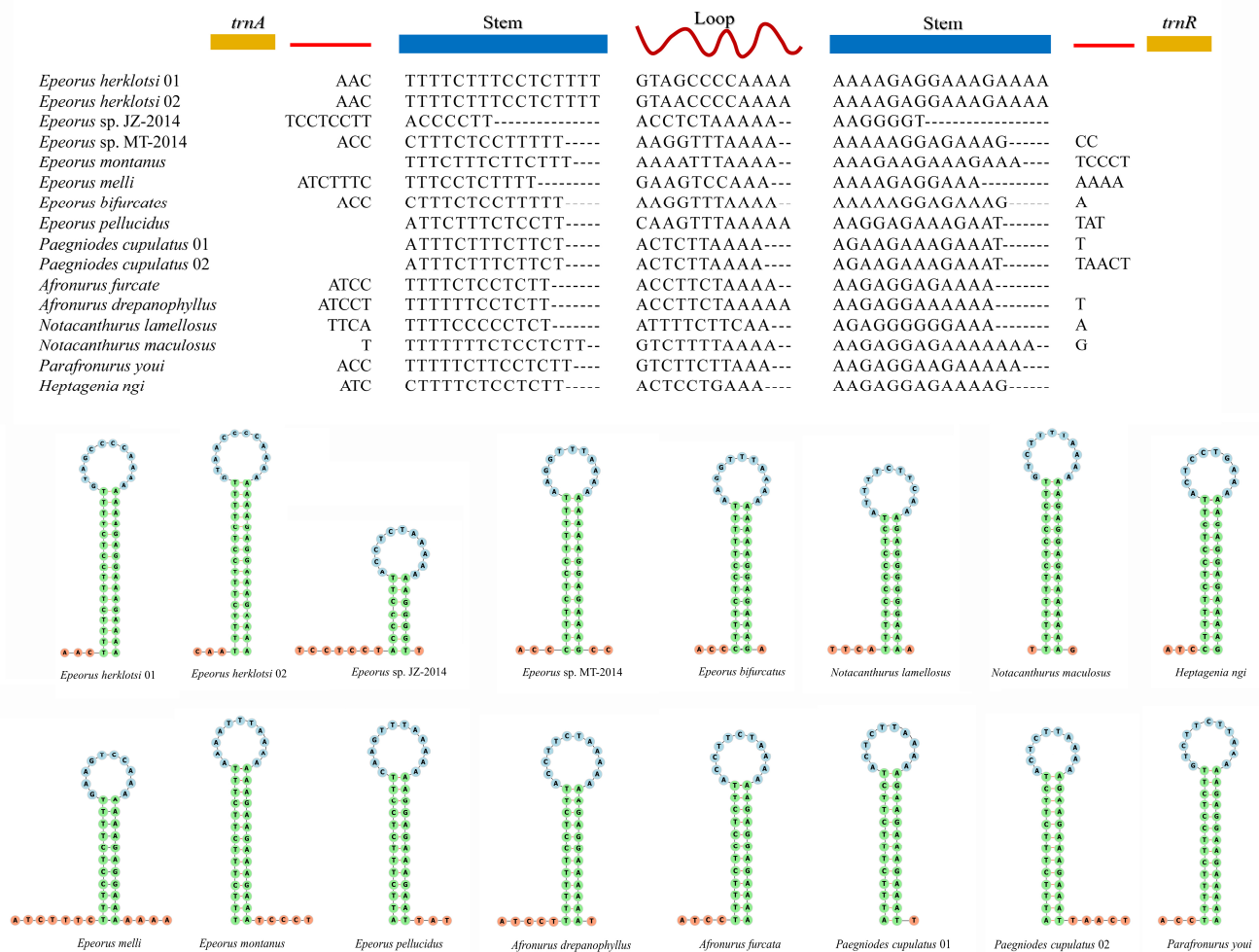


Figure 5. Alignment of the conserved intergenic spacer (IGS) between *trnA* and *trnR* of heptageniid mitogenomes and the putative stem–loop structures.

Compared to the ancestral gene order (Figure 6a), the additional *trnM* was inserted between *trnI* and *trnQ* forming the gene block (*trnI-trnM-trnQ-trnM-ND2*). Two models of tandem duplication-random loss (TDRL) and recombination are generally proposed to explain this rearrangement mechanism [53,54]. The hypothetical process of the gene rearrangement is as follows: The gene block (*trnI-trnQ-trnM*) firstly underwent a tandem duplication, resulting in a new gene cluster with two same sets (*trnI-trnQ-trnM-trnI-trnQ-trnM*) (Figure 6b). Then this consecutive copy was followed by a random loss; *trnQ* of the first set and *trnI* of the second set were subsequently lost (Figure 6c). Finally, the new gene arrangement of heptageniid mitogenomes (*trnI-trnM-trnQ-trnM*) was generated (Figure 6d). It is plausible to hypothesize that the gene arrangement pattern with an extra *trnM* could be explained through the TDRL model [14,55].

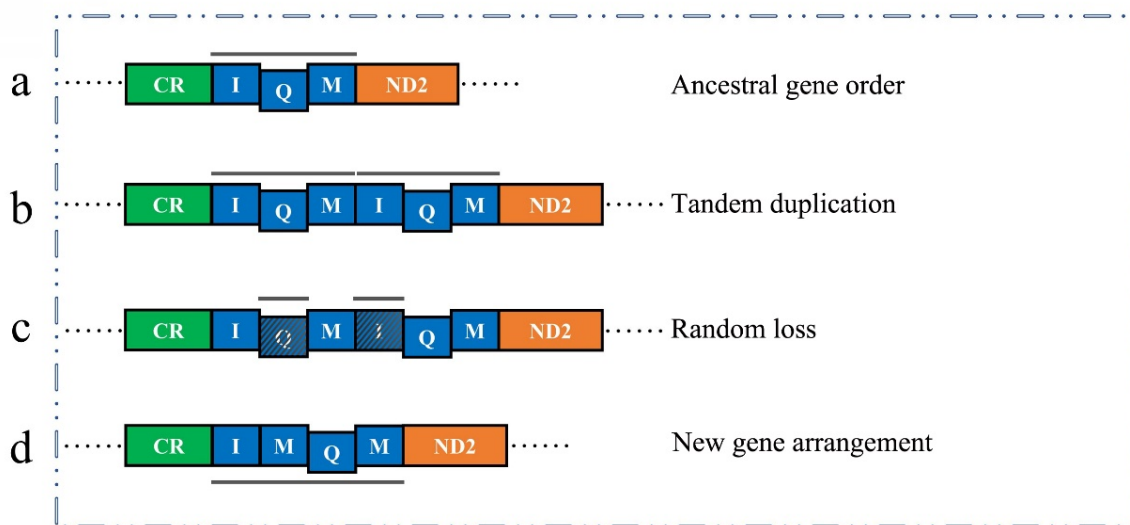


Figure 6. Inferred intermediate processes of the gene rearrangement of Heptageniidae. (a) The ancestral gene order of insects. (b) Duplication of *trnI-trnQ-trnM* region. (c) Random deletion of the copied genes. (d) The final gene order of Heptageniidae.

All gene rearrangement events reported in Ephemeroptera are almost focused on the tRNA genes, characterized as minor rearrangement. This has also been found in other insect orders such as Orthoptera, Hemiptera, Hymenoptera [7,56,57]. This minor rearrangement, however, has rarely been used to reconstruct phylogenetic relationships in insects. In our present study, the rearrangement of *trnM* could be considered a clear molecular synapomorphy for Heptageniidae, and might be an effective molecular marker for the family.

3.6. Phylogenetic Analyses

Phylogenetic analyses based on PCGs of 16 heptageniid species and 1 siphuriscid species (outgroup) were conducted. Two datasets were used in this study: (1) the amino acid matrix (AA) contained 3661 sites including all PCGs' aa sequences; (2) the nucleotide matrix (P123) contained 10,983 sites including all PCGs with three codon positions. The phylogenetic trees generated from two analytical methods (BI and ML) had unique topologies with the same dataset.

The trees based on two datasets exhibited slightly different topology, the incongruence being restricted to the position of *P. cupulatus* (Figure 7). Monophyly of the genus *Epeorus* received strong support in all analyses, which was consistent with the traditional morphological classification [17]. The AA analysis showed that *P. cupulatus* was the sister group to the species of the genus *Epeorus* (Figure 7A). The results supported the previous analysis that the two genera belong to a monophyletic Rhithrogeninae [17]. However, the P123 tree placed *P. cupulatus* at the base of the clade consisting of all other heptageniid species (Figure 7B). The absence of the *trnM* duplication of *P. cupulatus* seems to support this phylogenetic relationship. Additionally, our previous study found that the genus *Paegniodes* possessed more plesiomorphies in Heptageniidae, such as terminal filament and cerci well-developed, setae scattered on the ventral surface of the maxillae, and gills not enlarged or not forming a friction disk [58]. On the basis of the above analyses, *P. cupulatus* might be recognized as the earliest diverging species among the analyzed heptageniids.

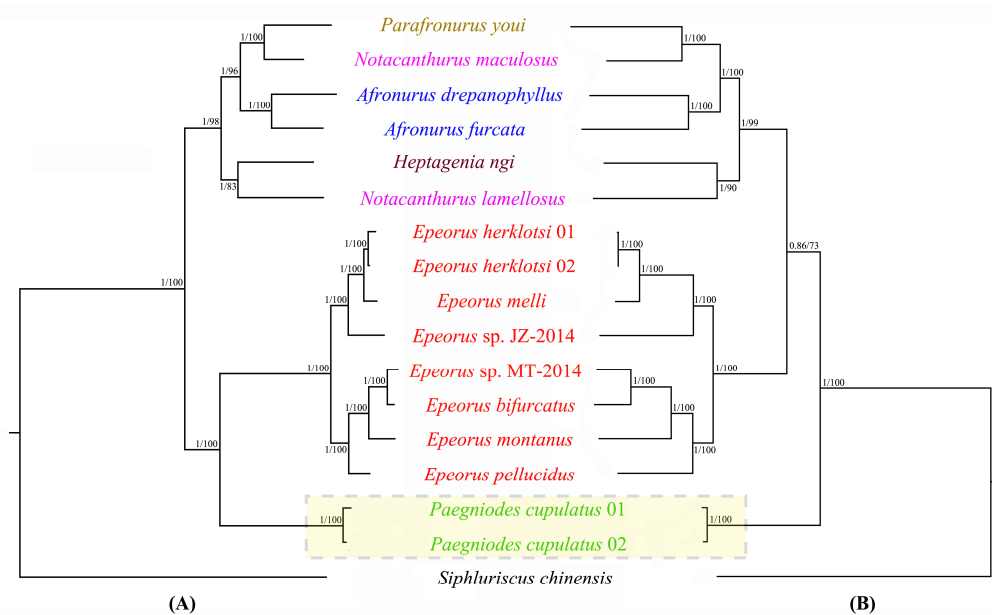


Figure 7. Phylogenetic analyses of Heptageniidae based on amino acid (A) and nucleotide datasets (B). Numbers separated by a slash on the node are posterior probability (PP) and bootstrap value (BV).

For other species, the positions were identical in all topologies (Figure 7). *Heptagenia ngi* (Heptageniinae) was clustered together with other Ecdyonurinae species, and sister to *N. lamellosus*. Because only one species was sampled in Heptageniinae, its current position within Ecdyonurinae needs expanded samplings to reconfirm in the further study. Our analyses revealed that two species of the genus *Notacanthurus* were not placed together. This is consistent with morphological study, which suggests that the genus is probably not monophyletic, but more likely polyphyletic [59].

4. Conclusions

The present study determined ten complete mitogenomes of heptageniid species and is the first detailed comparative genomic and phylogenetic analysis within Heptageniidae. All heptageniid mitogenomes possessed an extra *trnM* except *P. cupulatus*, retaining the typical gene content. The analysis of evolutionary patterns showed that all PCGs were under purifying selection, and that *ND6* and *ND2* exhibited a faster evolution rate and diversity than other genes. A conserved intergenic spacer was observed in all heptageniid mitogenomes, which could be folded into a sTable Stem-loop structure. A tRNA gene rearrangement was found in most mitogenomes of Heptageniidae and the unique gene block *trnI-trnM-trnQ-trnM-ND2* presented. It was plausible to hypothesize that the gene arrangement pattern with an extra *trnM* occurred via the TDRL model. The phylogenetic analyses inferred from mitogenomes strongly supported the monophyly of the genus *Epeorus*. Our results improve the understanding of mitogenomic evolution and phylogenetic relationships in Heptageniidae.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2075-4450/12/2/170/s1>, Table S1. Collection information of Heptageniidae species in this study, Table S2. The information of next generation sequencing for ten samples, Table S3. The partition schemes and best-fitting models selected in amino acid dataset, Table S4. The partition schemes and best-fitting models selected in nucleotide dataset, Table S5. Start and stop codons of protein-coding genes in the mitogenomes of heptageniid mitogenomes.

Author Contributions: Conceptualization, C.Z. and R.L.; methodology, R.L., Z.L., W.L.; investigation, R.L.; data curation, formal analysis and visualization, R.L., Z.L., W.L. and W.Z.; writing—Original draft preparation, R.L.; writing—Review and editing, C.Z.; project administration and funding acquisition, C.Z. and R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 31750002) to Changfa Zhou, and the China Postdoctoral Science Foundation (Grant No. 2019M661877) to Ran Li.

Data Availability Statement: The data that support the findings of this study are openly available in National Center for Biotechnology Information at <https://www.ncbi.nlm.nih.gov/nuccore> (accessed on 10 February 2021), reference numbers MW381291–MW381300.

Acknowledgments: We are grateful to Xiaohan Shu for her help in figure drawing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wolstenholme, D.R. Animal mitochondrial DNA: Structure and evolution. *Int. Rev. Cytol.* **1992**, *141*, 173–216. [[PubMed](#)]
2. Moritz, C.; Dowling, T.E.; Brown, W.M. Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. *Ann. Rev. Ecol. Syst.* **1987**, *18*, 173–216. [[CrossRef](#)]
3. Boore, J.L. Animal mitochondrial genomes. *Nucleic Acids Res.* **1999**, *27*, 1767–1780. [[CrossRef](#)] [[PubMed](#)]
4. Cameron, S.L. Insect mitochondrial genomics: Implications for evolution and phylogeny. *Annu. Rev. Entomol.* **2014**, *59*, 95–117. [[CrossRef](#)] [[PubMed](#)]
5. Avise, J.C.; Arnold, J.; Ball, R.M.; Bermingham, E.; Lamb, T.; Neigel, J.E.; Reeb, C.A.; Saunders, N.C. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.* **1987**, *18*, 489–522. [[CrossRef](#)]
6. Zhou, N.; Dong, Y.; Qiao, P.; Yang, Z. Complete Mitogenomic Structure and Phylogenetic Implications of the Genus *Ostrinia* (Lepidoptera: Crambidae). *Insects* **2020**, *11*, 232. [[CrossRef](#)] [[PubMed](#)]
7. Song, H.; Amédégno, C.; Cigliano, M.M.; Desutter-Grandcolas, L.; Heads, S.W.; Huang, Y.; Otte, D.; Whiting, M.F. 300 million years of diversification: Elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling. *Cladistics* **2015**, *31*, 621–651. [[CrossRef](#)]
8. Du, C.; Zhang, L.; Lu, T.; Ma, J.; Zeng, C.; Yue, B.; Zhang, X. Mitochondrial genomes of blister beetles (Coleoptera, Meloidae) and two large intergenic spacers in *Hycleus* genera. *BMC Genom.* **2017**, *18*, 698. [[CrossRef](#)]
9. Sun, Z.; Liu, Y.; Wilson, J.J.; Chen, Z.; Song, F.; Cai, W.Z.; Li, H. Mitochondrial genome of *Phalantus geniculatus* (Hemiptera: Reduviidae): *trnT* duplication and phylogenetic implications. *Int. J. Biol. Macromol.* **2019**, *129*, 110–115. [[CrossRef](#)]
10. Cao, J.J.; Wang, Y.; Li, W.H. Comparative mitogenomic analysis of species in the subfamily Amphinemurinae (Plecoptera: Nemouridae) reveal conserved mitochondrial genome organization. *Int. J. Biol. Macromol.* **2019**, *138*, 292–301. [[CrossRef](#)]
11. Bauernfeind, E.; Soldán, T. *The Mayflies of Europe (Ephemeroptera)*; Apollo Books: Ollerup, Denmark, 2012; p. 781.
12. 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* **2014**, *545*, 132–140. [[CrossRef](#)] [[PubMed](#)]
13. Gao, X.Y.; Zhang, S.S.; Zhang, L.P.; Yu, D.N.; Zhang, J.Y.; Cheng, H.Y. The complete mitochondrial genome of *Epeorus herklotsi* (Ephemeroptera: Heptageniidae) and its phylogeny. *Mitochondrial DNA B* **2018**, *3*, 303–304. [[CrossRef](#)]
14. Li, R.; Zhang, W.; Ma, Z.X.; Zhou, C.F. Novel gene rearrangement pattern in the mitochondrial genomes of *Torleya mikhaili* and *Cincticostella fusca* (Ephemeroptera: Ephemerellidae). *Int. J. Biol. Macromol.* **2020**, *165*, 3106–3114. [[CrossRef](#)]
15. Zhang, J.Y.; Zhou, C.F.; Gai, Y.H.; Song, D.X.; Zhou, K.Y. The complete mitochondrial genome of *Parafronurus youi* (Insecta: Ephemeroptera) and phylogenetic position of the Ephemeroptera. *Gene* **2008**, *424*, 18–24. [[CrossRef](#)] [[PubMed](#)]
16. Zhou, D.; Wang, Y.Y.; Sun, J.Z.; Han, Y.K.; Zhou, C.F. The complete mitochondrial genome of *Paegniodes cupulatus* (Ephemeroptera: Heptageniidae). *Mitochondrial DNA A* **2016**, *27*, 925–926. [[CrossRef](#)] [[PubMed](#)]
17. Webb, J.M.; McCafferty, W.P. Heptageniidae of the World. Part II: Key to the Genera. *Can. J. Arthrop. Ident.* **2008**, *7*, 1–55.
18. Webb, J.M.; McCafferty, W.P. Contributions to the larvae of North American Nixe (Ephemeroptera: Heptageniidae), with the description of *N. dorotheae* sp. nov. from southern Indiana. *Zootaxa* **2011**, *3065*, 27–37. [[CrossRef](#)]
19. Ball, S.L.; Hebert, P.D.N.; Burian, S.K.; Webb, J.M. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *J. N. Am. Benthol. Soc.* **2005**, *24*, 508–524. [[CrossRef](#)]
20. Barber-James, H.M.; Gattolliat, J.L.; Sartori, M.; Hubbard, M.D. Global diversity of mayflies (Ephemeroptera, Insecta) in freshwater. *Hydrobiologia* **2008**, *595*, 339–350. [[CrossRef](#)]
21. Tshernova, O.A. The generic composition of the mayflies of the family Heptageniidae (Ephemeroptera) in the Holarctic and Oriental region. *Entomol. Obozr.* **1974**, *53*, 801–814.
22. Tomka, I.; Zurwerra, A. Key to the genera of the Heptageniidae (Ephemeroptera) of the Holarctic, Oriental and Ethiopian region. *Entomol. Berich. Luz.* **1985**, *14*, 113–126.
23. Yanai, Z.; Sartori, M.; Dor, R.; Dorchin, N. Molecular phylogeny and morphological analysis resolve a long-standing controversy over generic concepts in Ecdyonurinae mayflies (Ephemeroptera: Heptageniidae). *Syst. Entomol.* **2017**, *42*, 182–193. [[CrossRef](#)]
24. Polášek, M.; Godunko, R.J.; Rutschmann, S.; Svitok, M.; Novikmec, M.; Zahradková, S. Integrative taxonomy of genus *Electrogena* (Ephemeroptera: Heptageniidae): The role of innovative morphological analyses for species delimitation. *Arthropod Syst. Phylogeny* **2018**, *76*, 449–462.

25. Gattolliat, J.L.; Monaghan, M.T.; Sartori, M.; Elouard, J.M.; Vogler, A.P. A molecular analysis of the Afrotropical Baetidae. In *International Advances in the Ecology, Zoogeography and Systematics of Mayflies and Stoneflies*; University of California Press: Berkeley, CA, USA, 2008; pp. 219–232.
26. Ogden, T.H.; Osborne, J.T.; Jacobus, L.M.; Whiting, M.F. Combined molecular and morphological phylogeny of Ephemerellinae (Ephemerellidae: Ephemeroptera), with remarks about classification. *Zootaxa* **2009**, *1991*, 28–42. [[CrossRef](#)]
27. Meng, G.; Li, Y.; Yang, C.; Liu, S. MitoZ: A toolkit for mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res.* **2019**, *47*, e63. [[CrossRef](#)] [[PubMed](#)]
28. Lowe, T.M.; Chan, P.P. tRNAscan-SE Online: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57. [[CrossRef](#)]
29. Laslett, D.; Canbäck, B. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **2008**, *24*, 172–175. [[CrossRef](#)] [[PubMed](#)]
30. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
31. Perna, N.T.; Kocher, T.D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **1995**, *41*, 353–358. [[CrossRef](#)] [[PubMed](#)]
32. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Alejandro, S.G. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [[CrossRef](#)]
33. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
34. Zhang, D.; Gao, F.; Jakovlic, I.; Zhou, H.; Zhang, J.; Li, W.X.; Wang, G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Res.* **2020**, *20*, 348–355. [[CrossRef](#)]
35. Talavera, G.; Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **2007**, *56*, 564–577. [[CrossRef](#)]
36. Lanfear, R.; Calcott, B.; Ho, S.Y.; Guindon, S. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **2012**, *29*, 1695–1701. [[CrossRef](#)] [[PubMed](#)]
37. Ronquist, F.; Teslenko, M.; Van Der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)]
38. Miller, M.A.; Wayne, P.; Terri, S. The CIPRES science gateway: A community resource for phylogenetic analyses. In *Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery*, Salt Lake City, UT, USA, 18–21 July 2011; pp. 1–8.
39. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
40. Ojala, D.; Montoya, J.; Attardi, G. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **1981**, *290*, 470–474. [[CrossRef](#)]
41. Zhang, R.Y.; Li, J.; Geng, S.; Yang, J.; Zhang, X.; An, Y.X.; Li, C.; Cui, H.R.; Li, X.Y.; Wang, Y.Y. The first mitochondrial genome for Phaudidae (Lepidoptera) with phylogenetic analyses of Zygaenoidea. *Int. J. Biol. Macromol.* **2020**, *149*, 951–961. [[CrossRef](#)] [[PubMed](#)]
42. Ma, L.Y.; Liu, F.F.; Chiba, H.; Yuan, X.Q. The mitochondrial genomes of three skippers: Insights into the evolution of the family Hesperidae (Lepidoptera). *Genomics* **2020**, *112*, 432–441. [[CrossRef](#)] [[PubMed](#)]
43. Jia, W.Z.; Yan, H.B.; Guo, A.J.; Zhu, X.Q.; Wang, Y.C.; Shi, W.G.; Chen, H.T.; Fang, Z.; Zhang, S.H.; Fu, B.Q. Complete mitochondrial genomes of *Taenia multiceps*, *T. hydatigena* and *T. pisiformis*: Additional molecular markers for a tapeworm genus of human and animal health significance. *BMC Genom.* **2010**, *11*, 447. [[CrossRef](#)] [[PubMed](#)]
44. Su, T.; He, B.; Li, K.; Liang, A. Comparative analysis of the mitochondrial genomes of oriental spittlebug tribe Cosmoscirtini: Insights into the relationships among closely related taxa. *BMC Genom.* **2018**, *19*, 1–13. [[CrossRef](#)] [[PubMed](#)]
45. Hu, K.; Yuan, F.; Dietrich, C.H.; Yuan, X.Q. Structural features and phylogenetic implications of four new mitogenomes of Centrotinae (Hemiptera: Membracidae). *Int. J. Biol. Macromol.* **2019**, *139*, 1018–1027. [[CrossRef](#)] [[PubMed](#)]
46. Clayton, D.A. Replication and transcription of vertebrate mitochondrial DNA. *Annu. Rev. Cell Biol.* **1991**, *7*, 453–478. [[CrossRef](#)]
47. Fernandez-Silva, P.; Enriquez, J.A.; Montoya, J. Replication and transcription of mammalian mitochondrial DNA. *Exp. Physiol.* **2003**, *88*, 41–56. [[CrossRef](#)] [[PubMed](#)]
48. Ruan, H.T.; Li, M.; Li, Z.H.; Huang, J.J.; Chen, W.Y.; Sun, J.J.; Liu, L.; Zou, K.S. Comparative analysis of complete mitochondrial genomes of three *Gerres* fishes (Perciformes: Gerreidae) and primary exploration of their evolution history. *Int. J. Mol. Sci.* **2020**, *21*, 1874. [[CrossRef](#)]
49. Yu, X.L.; Tan, W.; Zhang, H.Y.; Jiang, W.L.; Gao, H.; Wang, W.X.; Gao, H.; Wang, W.X.; Liu, Y.X.; Wang, Y.; et al. Characterization of the complete mitochondrial genome of *Harpalus sinicus* and its implications for phylogenetic analyses. *Genes* **2019**, *10*, 724. [[CrossRef](#)]
50. Gong, L.; Liu, L.Q.; Guo, B.Y.; Ye, Y.Y.; Lü, Z.M. The complete mitochondrial genome characterization of *Thunnus obesus* (Scombriformes: Scombridae) and phylogenetic analyses of *Thunnus*. *Conserv. Genet. Res.* **2017**, *9*, 379–383. [[CrossRef](#)]

51. Seligmann, H.; Krishnan, N.M.; Rao, B.J. Possible multiple origins of replication in primate mitochondria: Alternative role of tRNA sequences. *J. Theor. Biol.* **2006**, *241*, 321–332. [[CrossRef](#)]
52. Seligmann, H.; Labra, A. The relation between hairpin formation by mitochondrial WANCY tRNAs and the occurrence of the light strand replication origin in Lepidosauria. *Gene* **2014**, *542*, 248–257. [[CrossRef](#)]
53. Downton, M.; Campbell, N.J. Intramitochondrial recombination—is it why some mitochondrial genes sleep around? *Trends Ecol. Evol.* **2001**, *16*, 269–271. [[CrossRef](#)]
54. Mao, M.; Gibson, T.; Downton, M. Evolutionary dynamics of the mitochondrial genome in the Evaniomorpha (Hymenoptera)—A group with an intermediate rate of gene rearrangement. *Genome Biol. Evol.* **2014**, *6*, 1862–1874. [[CrossRef](#)] [[PubMed](#)]
55. Gong, L.; Lu, X.T.; Luo, H.R.; Zhang, Y.; Shi, W.; Liu, L.Q.; Lü, Z.M.; Liu, B.J.; Jiang, L.H. Novel gene rearrangement pattern in *Cynoglossus melampetalus* mitochondrial genome: New gene order in genus *Cynoglossus* (Pleuronectiformes: Cynoglossidae). *Int. J. Biol. Macromol.* **2020**, *149*, 1232–1240. [[CrossRef](#)]
56. Liu, Y.; Li, H.; Song, F.; Zhao, Y.; Wilson, J.J.; Cai, W.Z. Higher-level phylogeny and evolutionary history of Pentatomomorpha (Hemiptera: Heteroptera) inferred from mitochondrial genome sequences. *Syst. Entomol.* **2019**, *44*, 810–819. [[CrossRef](#)]
57. Feng, Z.; Wu, Y.; Yang, C.; Gu, X.; Wilson, J.J.; Li, H.; Cai, W.Z.; Yang, H.L.; Song, F. Evolution of tRNA gene rearrangement in the mitochondrial genome of ichneumonoid wasps (Hymenoptera: Ichneumonoidea). *Int. J. Biol. Macromol.* **2020**, *164*, 540–547. [[CrossRef](#)] [[PubMed](#)]
58. Ma, Z.X.; Han, N.; Zhang, W.; Zhou, C.F. Position and definition of the genus *Paegniodes* Eaton, 1881 based on redescription on the type species *Paegniodes cupulatus* (Eaton, 1871) (Ephemeroptera: Heptageniidae). *Aquat. Insect.* **2018**, *39*, 362–374. [[CrossRef](#)]
59. Zhang, W.; Zhang, M.; Han, N.; Zhou, C.F. Two new species of the genus *Notacanthurus* from China (Ephemeroptera: Heptageniidae, Ecdyonurinae). *Zootaxa* **2020**, *4802*, 335–348. [[CrossRef](#)]