

Article

# Comparative Mitogenomics and Phylogenetic Analyses of Pentatomoidea (Hemiptera: Heteroptera)

Shiwen Xu <sup>1</sup> , Yunfei Wu <sup>1</sup>, Yingqi Liu <sup>1</sup> , Ping Zhao <sup>2</sup> , Zhuo Chen <sup>1</sup>, Fan Song <sup>1</sup>, Hu Li <sup>1</sup>  and Wanzhi Cai <sup>1,\*</sup>

<sup>1</sup> Department of Entomology and MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing 100193, China; xusw2019@126.com (S.X.); wuyunfeiakx@126.com (Y.W.); yingqiliu0720@163.com (Y.L.); insectchen625@126.com (Z.C.); fansong@cau.edu.cn (F.S.); tigerleecau@hotmail.com (H.L.)

<sup>2</sup> Key Laboratory of Environment Change and Resources Use in Beibu Gulf (Ministry of Education) and Guangxi Key Laboratory of Earth Surface Processes and Intelligent Simulation, Nanning Normal University, Nanning 530001, China; zpyayjl@126.com

\* Correspondence: caiwz@cau.edu.cn; Tel.: +86-10-62734842

**Abstract:** Pentatomoidea is the largest superfamily of Pentatomomorpha; however, the phylogenetic relationships among pentatomoid families have been debated for a long time. In the present study, we gathered the mitogenomes of 55 species from eight common families (Acanthosomatidae, Cydnidae, Dinidoridae, Scutelleridae, Tessaratomidae, Plataspidae, Urostylididae and Pentatomidae), including 20 newly sequenced mitogenomes, and conducted comparative mitogenomic studies with an emphasis on the structures of non-coding regions. Heterogeneity in the base composition, and contrasting evolutionary rates were encountered among the mitogenomes in Pentatomoidea, especially in Urostylididae, which may lead to unstable phylogenetic topologies. When the family Urostylididae is excluded in taxa sampling or the third codon positions of protein coding genes are removed, phylogenetic analyses under site-homogenous models could provide more stable tree topologies. However, the relationships between families remained the same in all PhyloBayes analyses under the site-heterogeneous mixture model CAT + GTR with different datasets and were recovered as (Cydnidae + (((Tessaratomidae + Dinidoridae) + (Plataspidae + Scutelleridae)) + ((Acanthosomatidae + Urostylididae) + Pentatomidae))). Our study showed that data optimizing strategies after heterogeneity assessments based on denser sampling and the use of site-heterogeneous mixture models are essential for further analysis of the phylogenetic relationships of Pentatomoidea.

**Keywords:** Pentatomoidea; mitochondrial genome; sequence heterogeneity; phylogeny



**Citation:** Xu, S.; Wu, Y.; Liu, Y.; Zhao, P.; Chen, Z.; Song, F.; Li, H.; Cai, W. Comparative Mitogenomics and Phylogenetic Analyses of Pentatomoidea (Hemiptera: Heteroptera). *Genes* **2021**, *12*, 1306. <https://doi.org/10.3390/genes12091306>

Academic Editor: Nico M. Van Straalen

Received: 19 July 2021

Accepted: 24 August 2021

Published: 25 August 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

Pentatomoidea (Hemiptera: Pentatomomorpha), also known as stink bugs and their relatives, is the largest superfamily within Pentatomomorpha, with over 8000 species [1,2]. Most taxa in this widespread terrestrial superfamily are phytophagous, feeding on a variety of fruits, vegetables, grain crops or wild plants, and thus some are economically important pests in agriculture and forestry [2,3].

Although the monophyly of Pentatomoidea has been supported by both morphological and molecular data [4–10], relationships at the family level remain controversial. For instance, there has been a longstanding debate on how many valid families are present in Pentatomoidea (see summaries by Grazia et al. [4] and Rider [2]). The division of Pentatomoidea into 18 families was only generally accepted recently, with a few doubts yet to be addressed [1]. Using morphological characters, Gapud [5] and Xu [11] were the first ones to study the phylogeny of Pentatomoidea. Following their studies, phylogenetic analyses based on molecular data, including nuclear and mitochondrial genes fragments [8,12–14], or whole mitochondrial genomes (mitogenomes) [7,9,10] were subsequently performed. More recently, a phylogenetic study on Pentatomoidea combining morphological and

molecular data was performed by Grazia et al. [4]. The above-mentioned studies, despite having different resulting topologies, reached agreement on relationships among certain lineages. For instance, the sister relationship between Tessaratomidae and Dinidoridae was supported by the majority of these phylogenetic studies. The sister relationship between Urostylididae and Acanthosomatidae was also recovered by different molecular and morphological markers [8,9,11,15]. In addition, mitogenomic analyses performed by different research groups have independently demonstrated that Urostylididae, or the clade involving Urostylididae is the earliest branching lineage of Pentatomoidea [7,10,13,14].

The mitogenome is one of the most widely used molecular markers in insect phylogenetic studies, but the compositional heterogeneity and accelerated evolutionary rates of the mitogenomes in certain groups may result in erroneous grouping of unrelated taxa [16–19]. Recent studies have demonstrated that the site-heterogeneous mixture model implemented in the software PhyloBayes may help in reducing the susceptibility of analyses to heterogeneity and thus provide more stable topologies [16,17,19–21]. Moreover, denser taxa sampling could be another strategy to assess the potential impact of substitutional and compositional biases, and to reduce tree reconstruction artifacts [9,22,23]. However, in regard to Pentatomoidea, some previous phylogenetic analyses had limited taxon sampling with only one species per family, hindering comprehensive data analysis [4,7,12].

So far, phylogenetic studies of Pentatomoidea using mitogenomic data are mainly based on sequences of eight common families (Acanthosomatidae, Cydnidae, Dinidoridae, Scutelleridae, Tessaratomidae, Plataspidae, Urostylididae and Pentatomidae) deposited in GenBank. In the present study, we gathered mitogenomic sequences of 55 species from these eight families, including 20 newly generated mitogenomes to better analyze the impact of sequence heterogeneity. In order to assess better strategies to resolve the phylogeny of this superfamily, we compared the effects of different models for phylogenetic analysis (i.e., site-homogeneous and site-heterogeneous mixture model), as well as some data optimizing methods, on the quality of resulting topology. Based on our analysis, we discuss potential ways to enhance future phylogenetic analysis on this highly diverse insect group.

## 2. Materials and Methods

### 2.1. Sample Collection and DNA Extraction

Voucher specimens of 20 pentatomoid species were collected in the field, preserved in anhydrous ethanol, and stored at  $-20\text{ }^{\circ}\text{C}$  at the Entomological Museum of China Agricultural University. Detailed collection information is provided in Table S1. Genomic DNA of specimens was extracted from thorax muscle tissues using the DNeasy Blood and Tissue Kit (Qiagen, Venlo, The Netherlands) as per the manufacturer's protocol.

### 2.2. Mitochondrial Genome Sequencing and Assembly

In order to improve sequencing efficiency and avoid resource waste, hybrid libraries were adopted according to Gillett et al. [22]. All Illumina TruSeq libraries were prepared with an average insert size of 350 bp and sequenced using the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA) with 150 bp paired-end reads. Prinseq version 0.20.4 (Edwards Lab, San Diego, CA, USA) [24] was used to remove short and low-quality reads with poly-Ns ( $>15$  bp Ns), or  $>75$  bp bases with a quality score  $\leq 3$ . The remaining reads were de novo assembled using IDBA-UD [25], with minimum and maximum k values of 45 and 145 bp, respectively. The partial sequences of *COI* and *srRNA* of each species were obtained by standard PCR reaction and Sanger sequencing to identify the corresponding mitogenome assemblies. Clean reads were mapped using Geneious version 10.1.3 (<http://www.geneious.com/> (accessed on 22 December 2020)) [26] to evaluate assembly accuracy.

### 2.3. Mitochondrial Genome Annotation and Sequence Analysis

Gene sequences were annotated using MitoZ [27], and further corrected in Geneious. The locations and secondary structures of tRNA genes were determined using tRNAscan-SE version 1.21 [28] and ARWEN version 1.2 [29]. Protein-coding genes (PCGs) and rRNA genes were identified by aligning their sequences with those of homologous genes of other Pentatomoidea species. Nucleotide composition of mitogenomes and the codon usage of PCGs were analyzed with MEGA version 7.0. (IGEM, Philadelphia, PA, USA) [30]. Nucleotide compositional differences were calculated using the following formulae: AT skew =  $(A - T)/(A + T)$  and GC skew =  $(G - C)/(G + C)$  [31]. DnaSP 5.0 [32] was used to calculate the rate of synonymous substitutions (Ks) and the rate of non-synonymous substitutions (Ka) for PCGs.

### 2.4. Phylogenetic Analysis

The 13 PCGs of each species were separately aligned using the L-INS-I strategy in the MAFFT algorithm [33], which was implemented in TranslatorX [34]. Two rRNA genes were aligned individually using the G-INS-I strategy in MAFFT version 7.0 (iFReC, Osaka, Japan) [35]. All alignments were manually checked using MEGA. Gene fragments were imported into Geneious and concatenated into two datasets: (1) the PCGRNA matrix with 12,311 nucleotides, corresponding to 13 PCGs and two rRNA genes; and (2) the PCG12RNA matrix with 8840 nucleotides, corresponding to the first and second codon positions of the 13 PCGs and two rRNA genes. Heterogeneity of sequence divergence within the datasets (i.e., PCGRNA and PCG12RNA) was analyzed using AliGROOVE [36] with the default sliding window size. This metric established pairwise sequence distances between individual terminals or subclades with terminals outside of the focal group. The distances were then compared to distances over the entire data matrix. The metric values can vary between  $-1$ , if distances are very different from the average for the entire data matrix, to  $+1$ , for distances which match the average for the entire matrix.

PartitionFinder2 [37] was used to select the optimal partitioning schemes and substitution models for analyses under site-homogeneous models. We used the “greedy” algorithm with branch lengths estimated as “unlinked” and the Akaike information criterion. The input configuration file containing 15 initial gene partitions (13 PCGs and two rRNA genes) was created and the optimal partitioning schemes for each dataset are shown in Table S2. The site-homogeneous model GTR + I + G was determined to be the most suitable model for all partitions in maximum likelihood (ML) analyses using IQ-TREE web server [38] and Bayesian analyses using MrBayes version 3.2.7 [39]. For IQ-TREE, phylogenetic trees were conducted with 1000 SH-aLRT replicates. For MrBayes, two simultaneous runs of 10 million generations were conducted for the datasets, and trees were sampled every 1000 generations, with the first 25% discarded as burn-in. Stationarity was considered to be reached when the average standard deviation of split frequencies was below 0.01. Bayesian trees were also constructed using PhyloBayes MPI version 1.5a (Robert Cedergren Centre, Montreal, PQ, Canada) [40] under the site-heterogeneous mixture model CAT + GTR [41,42]. The Markov chain Monte Carlo chains were run independently after removing constant sites from the alignment, and they were stopped after the two runs had satisfactorily converged (i.e., maxdiff < 0.3). A consensus tree was produced from the remaining trees after discarding the initial 25% of trees of each run as burn-in.

Phylogenetic analysis was first performed using whole mitogenomes of the 55 species (Table 1). Two species of Pyrrhocoroidea and two species of Aradoidea (Table 1) were selected as outgroups. Because the mitogenomes of Urostylididae showed the highest heterogeneity (see Seciton 3 below) among eight families, we performed another phylogenetic analysis following the same setting but excluding this family to assess the effect of heterogeneity on the resulting topology.

Table 1. Taxa used in the present study.

| Superfamily                | Family           | Species                                    | GenBank Accession Number |
|----------------------------|------------------|--|--------------------------|
| Aradoidea                  | Aradidae         | <i>Aneurus similis</i>                     | NC_030360                |
|                            |                  | <i>Aradus compar</i>                       | NC_030362                |
| Pyrrhocoroidea             | Largidae         | <i>Macrocheraia grandis</i>                | MK809515                 |
|                            | Pyrrhocoridae    | <i>Pyrrhocoris tibialis</i>                | KX355214                 |
|                            |                  | <i>Microdeuterus</i> sp. <sup>a</sup>      | MW847242                 |
|                            |                  | <i>Sastragala esakii</i> <sup>a</sup>      | MW847247                 |
|                            | Acanthosomatidae | <i>Acanthosoma labiduroides</i>            | JQ743670                 |
|                            |                  | <i>Anaxandra taurina</i>                   | NC_042801                |
|                            |                  | <i>Sastragala edessoides</i>               | JQ743676                 |
|                            |                  | <i>Aethus nigritus</i> <sup>a</sup>        | MW847231                 |
|                            |                  | <i>Macroscytus subaeneus</i> <sup>a</sup>  | MW847241                 |
|                            | Cydnidae         | Cydnidae sp.                               | MH643815                 |
|                            |                  | <i>Macroscytus gibbulus</i>                | NC_012457                |
|                            |                  | <i>Scoparipes salvazai</i>                 | NC_042800                |
|                            |                  | <i>Eumenotes</i> sp. <sup>a</sup>          | MW847237                 |
|                            |                  | <i>Megymenum gracilicorne</i>              | NC_042810                |
|                            | Dinidoridae      | <i>Cyclopelta parva</i>                    | NC_037739                |
|                            |                  | <i>Coridius chinensis</i>                  | JQ739179                 |
|                            |                  | <i>Megymenum brevicorne</i>                | JQ739181                 |
|                            |                  | <i>Chrysocoris stolii</i> <sup>a</sup>     | MW847234                 |
|                            |                  | <i>Poecilocoris druracii</i> <sup>a</sup>  | MW847246                 |
|                            |                  | <i>Cantao ocellatus</i>                    | NC_042803                |
|                            | Scutelleridae    | <i>Eurygaster testudinaria</i>             | NC_042808                |
|                            |                  | <i>Eucorysses grandis</i>                  | JQ743671                 |
|                            |                  | <i>Lamprocoris</i> sp.                     | JQ743674                 |
|                            |                  | <i>Poecilocoris nepalensis</i>             | JQ743675                 |
|                            |                  | <i>Dalcantha dilatata</i> <sup>a</sup>     | MW847235                 |
|                            | Tessaratomidae   | <i>Mattiphus splendidus</i>                | MN496304                 |
|                            |                  | <i>Eusthenes cupreus</i>                   | NC_022449                |
|                            |                  | <i>Tessaratoma papillosa</i>               | NC_037742                |
|                            |                  | <i>Brachyplatys subaeneus</i> <sup>a</sup> | MW847232                 |
| Pentatomoidea              | Plataspidae      | <i>Calacta lugubris</i> <sup>a</sup>       | MW847233                 |
|                            |                  | <i>Aponsila</i> sp.                        | MF497710                 |
| <i>Coptosoma bifaria</i>   |                  | NC_012449                                  |                          |
| <i>Megacopta cribraria</i> |                  | NC_015342                                  |                          |
|                            |                  | <i>Urochela caudata</i> <sup>a</sup>       | MW847248                 |
|                            |                  | <i>Urolabida histrionica</i> <sup>a</sup>  | MW847249                 |
|                            | Urostylididae    | <i>Urochela quadrinotata</i>               | NC_020144                |
|                            |                  | <i>Urolabida</i> sp.                       | MF497734                 |
|                            |                  | <i>Urostylis flavoannulata</i>             | NC_037747                |
|                            |                  | <i>Urostylis</i> sp.                       | JQ743679                 |
|                            |                  | <i>Dalpada cinctipes</i> <sup>a</sup>      | MW847236                 |
|                            |                  | <i>Eurydema dominulus</i> <sup>a</sup>     | MW847238                 |
|                            |                  | <i>Glaucias dorsalis</i> <sup>a</sup>      | MW847239                 |
|                            |                  | <i>Hippotiscus dorsalis</i> <sup>a</sup>   | MW847240                 |
|                            |                  | <i>Neojurtina typica</i> <sup>a</sup>      | MW847243                 |
|                            |                  | <i>Pentatoma metallifera</i> <sup>a</sup>  | MW847244                 |
|                            |                  | <i>Plautia lushanica</i> <sup>a</sup>      | MW847245                 |
|                            |                  | <i>Zicrona caerulea</i> <sup>a</sup>       | MW847250                 |
|                            | Pentatomidae     | <i>Arma custos</i>                         | MT535604                 |
|                            |                  | <i>Cazira horvathi</i>                     | NC_042817                |
|                            |                  | <i>Dalsira scabrata</i>                    | NC_037374                |
|                            |                  | <i>Dinorhynchus dybowskyi</i>              | NC_037724                |
|                            |                  | <i>Eocanthecona thomsoni</i>               | NC_042816                |
|                            |                  | <i>Gonopsis affinis</i>                    | NC_036745                |
|                            |                  | <i>Graphosoma rubrolineatum</i>            | NC_033875                |
|                            |                  | <i>Pentatoma semiannulata</i>              | MT985377                 |
|                            |                  | <i>Picromerus griseus</i>                  | NC_036418                |
|                            |                  | <i>Scotinophara lurida</i>                 | NC_042815                |

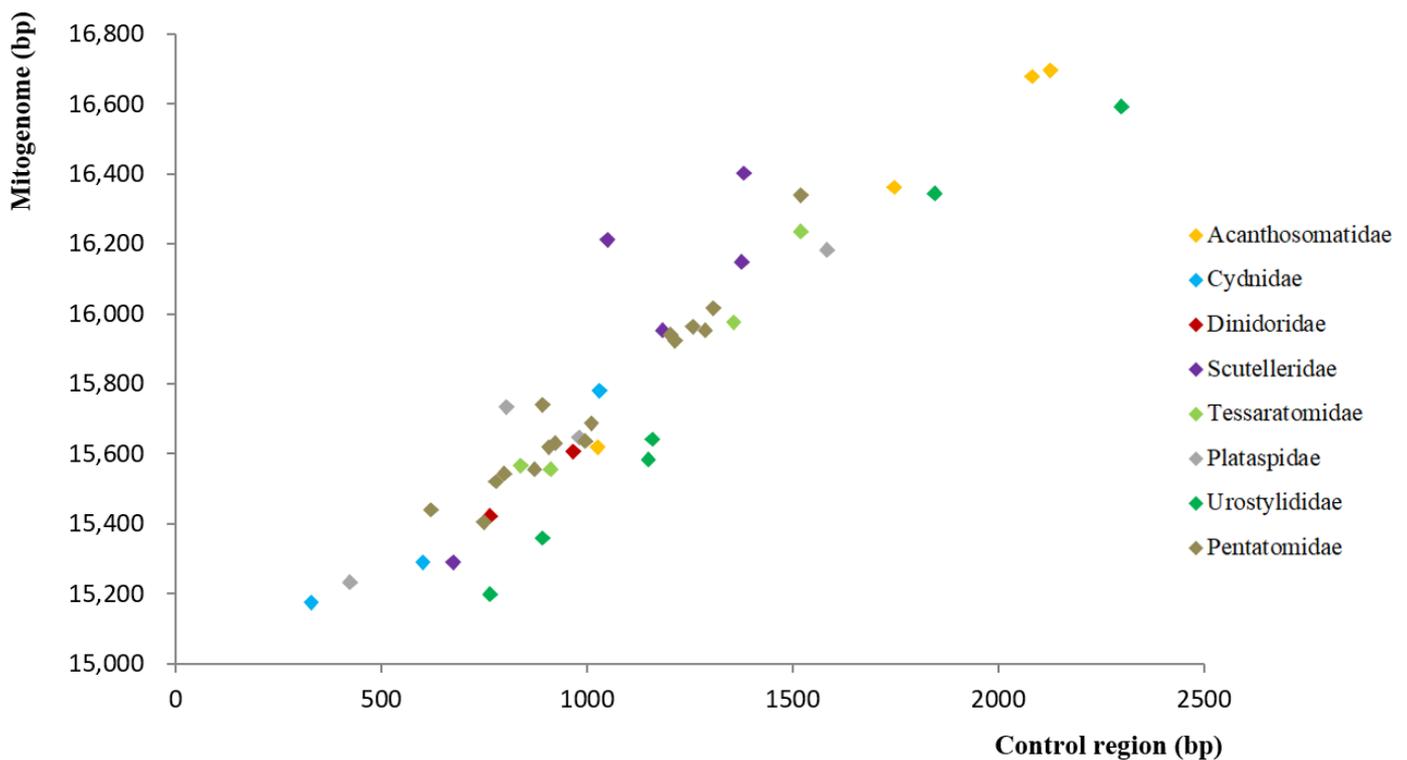
<sup>a</sup> Mitogenomes sequenced in the present study.

### 3. Results and Discussion

#### 3.1. General Features of Pentatomoid Mitogenomes

Among the 55 species involved in the comparative analyses, 44 have complete mitogenomes. A total of 37 typical genes (i.e., 13 PCGs, 2 rRNAs, and 22 tRNAs) and a control region were present in all 55 of the pentatomoid species, as has been reported in most insects [43,44], and no gene rearrangement occurred. There were three conservative gene overlaps within Pentatomioidea, namely an 8 bp overlap between *trnW* and *trnC*, a 7 bp overlap between *ATP8* and *ATP6*, and a 7 bp overlap between *ND4* and *ND4L*. The largest intergenic spacer was located between *trnS2* and *ND1*. Notably, similar overlaps and spacer regions are commonly in other insects of the suborder Heteroptera [43].

The length of the 44 complete mitogenomes ranged from 15,173 bp (i.e., *Cydnidae* sp.) to 16,694 bp (*Anaxandra taurina*), with an average size of 15.8 kb. The observed length variation among mitogenomes can be mostly attributed to the length variation in control regions (Figure 1). Furthermore, all 55 mitogenomes displayed a strong AT nucleotide bias with the A + T content ranging from 67.2% (*Scoparipes salvazai*) to 78.4% (*Gonopsis affinis*). A positive AT skew (i.e., 0.069 to 0.251) and a negative GC skew (i.e.,  $-0.255$  to  $-0.100$ ) were observed in all 55 Pentatomioidea species, which are common in insect mitogenomes [45–47].



**Figure 1.** Size of the complete mitogenomes and complete control regions of 44 Pentatomioidea species.

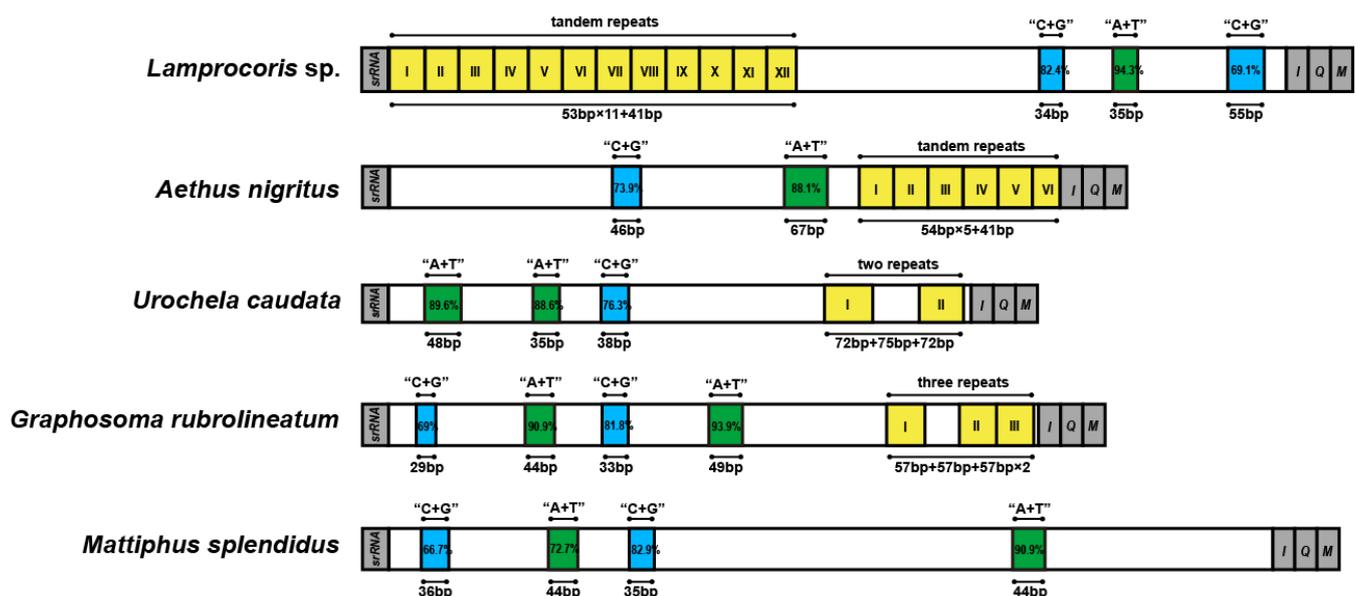
Most PCGs began with a typical ATN codon (i.e., ATA/ATT/ATG/ATC), while the start codon of *COI* was TTG in all pentatomoid species, and a small number of start codons were GTG in *ATP8* and *ATP6*. As for stop codons, they mostly consisted of TAA or TAG, while a single T residue was used as an incomplete stop codon in *COI*, *COII*, *ATP6*, *COIII*, *ND3*, *ND5*, *ND6*, and *ND1*, which has also been detected in other hemipterans [48,49]. In particular, the stop codons of *COII* were all incomplete, except for that in *Mattiphus splendidus*.

The length of tRNA genes ranged from 59 to 76 bp. With the exception of *trnS1* and *trnV*, most tRNA genes could be folded into the typical clover leaf secondary structure.

The dihydrouridine (DHU) arm of *trnS1* was a simple loop in 47 of the 55 examined pentatomoid species, whereas the DHU arm of *trnV* formed the same loop structure in 18 species. Furthermore, the sizes of *lrRNA* and *srRNA* genes ranged from 1238 (*Urolabida histrionica*) to 1387 bp (*S. salvazai*) and from 753 (*A. taurina*) to 878 bp (*Calacta lugubris*), respectively. The average A + T content of *lrRNA* genes was 77.9%, and that of *srRNA* genes was 76.4%.

### 3.2. Non-Coding Regions (NCR) of Pentatomoid Mitogenomes

The control regions (CR) were the NCRs with the greatest length variation, and total lengths of CR ranged from 334 (*S. salvazai*) to 2302 bp (*Urochela quadrinotata*) among the 44 complete pentatomoid mitogenomes. The general structures of the CR of these 44 species are shown in Figure S1. Three kinds of structural elements were summarized: A + T rich sequence blocks, C + G-rich sequence blocks, and repeat sequences. The arrangement and sequences of the A + T-rich and C + G-rich blocks were relatively conserved within the family Pentatomidae—that is, a C + G-rich region of approximately 30 bp and an A + T rich region of approximately 45 bp appeared alternately. Except for four species (i.e., *Eusthenes cupreus*, *M. splendidus* (Figure 2), *S. salvazai*, and *Zicrona caerulea*) that had no repeat sequences in their CR, tandem and non-tandem repeats were abundant among the pentatomoid species that were assessed. In particular, CRs of 30 species contained only one type of repeat unit, while those of the remaining 10 species contained two or three types.



**Figure 2.** Control regions of the representative pentatomoid mitogenomes. Structure elements found in the control region are labeled with different color blocks; green: A + T rich sequence block; blue: C + G rich sequence block; yellow with roman numerals inside: repeat sequences; grey: control regions flanking genes *srRNA*, *trnI*, *trnQ*, or *trnM*.

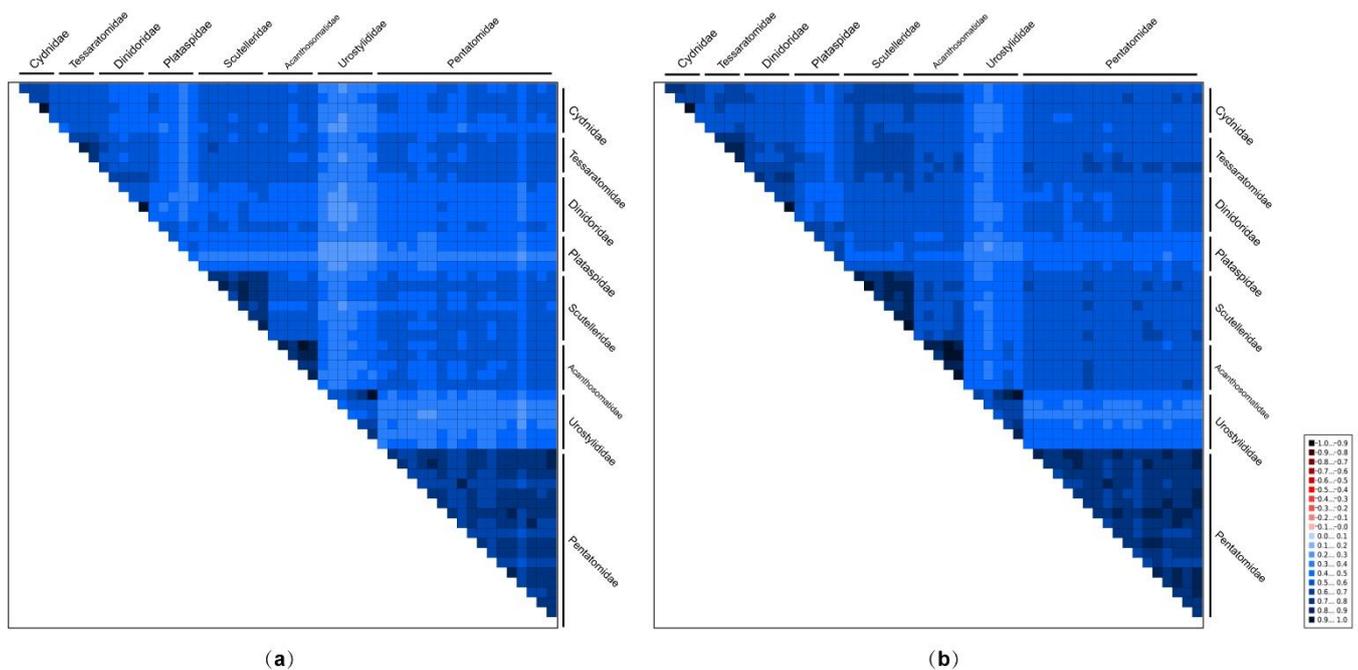
CRs that contained only one type of tandem repeat unit were present in 16 species, and the repeat region was located at either the start (e.g., *Lamprocoris* sp.) or the end (e.g., *Aethus nigratus*) of CR (Figure 2). The tandem repeat units were no more than 150 bp in length, and most of them consisted of small fragments of approximately 50 bp in length. Notably, CRs with more than 10 tandem units occurred in Acanthosomatidae, Urostylididae, Scutelleridae, and Pentatomidae. Apart from Cydnidae, the other seven families showed CRs that contained non-tandem repeats, which could be divided into two types: completely disconnected and with only two units (e.g., *Urochela caudata*; Figure 2), or partially disconnected tandem repeats with at least two units connected to each other (e.g., *Graphosoma rubrolineatum*; Figure 2). Partially disconnected tandem repeats were located in the CR of Acanthosomatidae, Urostylididae, Scutelleridae, Pentatomidae, and Plataspidae.

The intervals between repeats were generally less than 200 bp, although considerably long intervals were observed in *C. lugubris* (531 bp) and *Scotinophara lurida* (467 bp).

In addition to the CR, another long NCR was located between *trnS2* and *ND1* in 50 of the 55 species of Pentatomoidea, and it was 30 bp long on average. Notably, there was a 16 bp region with 84.2% similarity in this NCR.

### 3.3. Sequence Heterogeneity in Pentatomoid Mitogenomes and Phylogenetic Analyses

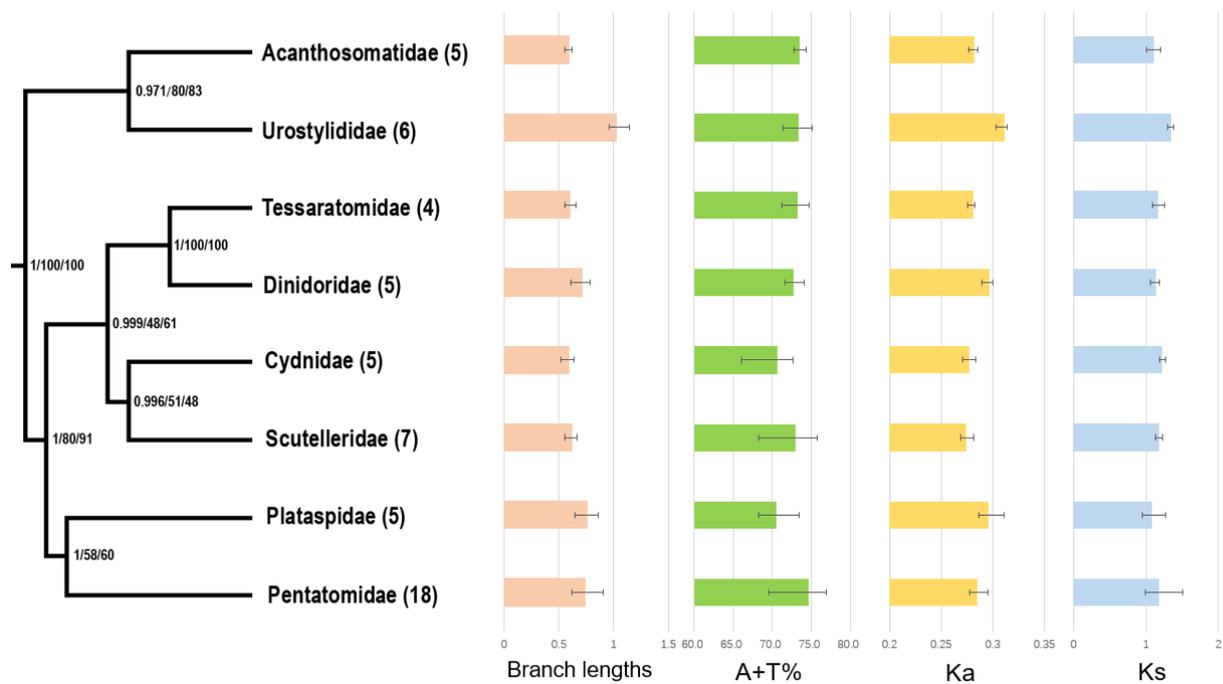
The AliGROOVE analyses indicated that the degree of heterogeneity of the PCG12RNA dataset (mean similarity score, 0.274) was lower than that of the PCGRNA dataset (0.236) (Figure 3). The third codon positions showed higher heterogeneity (lowest similarity score, 0.013) than the first codon positions (0.255) and the second codon positions (0.360) (Figure S2). In addition, the heterogeneity in sequence divergence was stronger for Urostylididae (mean similarity scores: 0.181 for PCGRNA dataset and 0.220 for PCG12RNA dataset) than for other pentatomoid families (0.327 for PCGRNA dataset and 0.376 for PCG12RNA dataset) (Figure 3). Previous studies have demonstrated that the compositional heterogeneity of mitogenomes in certain groups may result in the erroneous grouping of unrelated taxa [16–19], so Urostylididae may be unstably placed or misplaced in phylogenetic trees as a result of the divergences between these taxa.



**Figure 3.** AliGROOVE analysis of 55 pentatomoid species based on PCGRNA (a) and PCG12RNA (b). The mean similarity score between sequences is represented by colored squares, based on AliGROOVE scores ranging from  $-1$ , which indicates a great difference in rates from the remainder of the data set, that is, heterogeneity (red coloring), to  $+1$ , which indicates rates that matched all other comparisons (blue coloring).

Our phylogenetic analyses consistently supported the monophyly of Pentatomoidea and each of the sampled families. The results of the MrBayes analysis with the PCGRNA dataset (Figure S3) and the ML analyses with the PCGRNA and PCG12RNA datasets (Figure S4) under the site-homogeneous models show the same topology (Figure 4). The eight major families of Pentatomoidea formed four sister groups: (Acanthosomatidae + Urostylididae), (Tessaratomidae + Dinidoridae), (Cydnidae + Scutelleridae) and (Plataspidae + Pentatomidae), but support values in ML analyses of the latter two were relatively low (i.e.,  $<70$ ). Comparable to many previous phylogenetic studies [7,10,13,14], the clade involving Urostylididae was found to be the earliest branching lineage. The sister groups of (Cydnidae + Scutelleridae) and (Plataspidae + Pentatomidae) were also recovered by

Zhao et al. [10] and Liu et al. [9] using mitochondrial sequences, respectively, but with relatively limited sampling and low support.

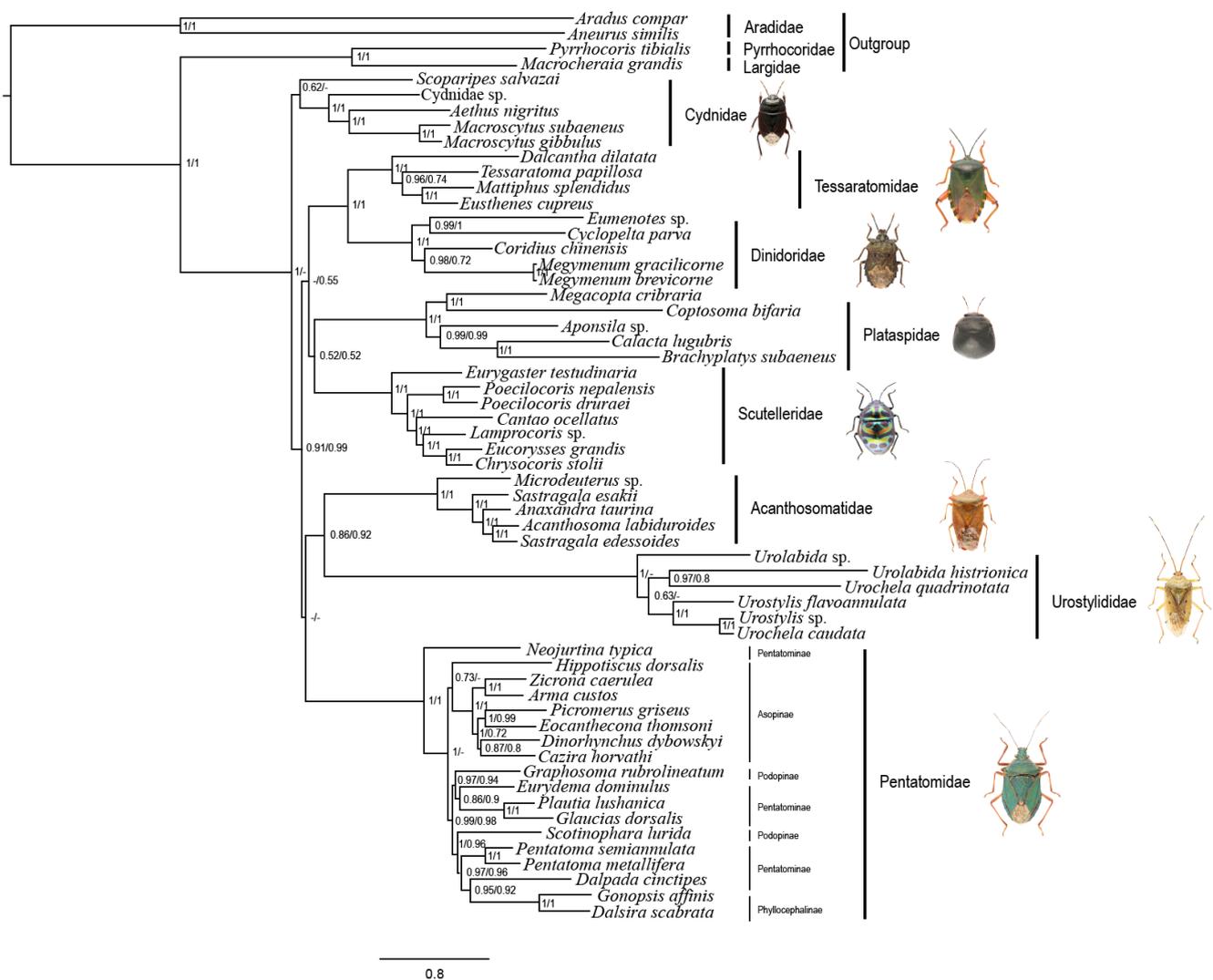


**Figure 4.** Systematic errors in phylogenetic analyses with site-homogeneous models. Supports at nodes (from left to right) obtained via MrBayes based on PCGRNA, IQ-TREE based on PCGRNA and IQ-TREE based on PCG12RNA. Numbers in brackets indicate the number of species in the corresponding family that were used for phylogenetic analysis. The branch lengths were calculated by MrBayes based on PCGRNA datasets. The A + T content (%), rate of non-synonymous substitutions ( $Ka$ ), and rates of synonymous substitutions ( $Ks$ ) were calculated from the protein-coding genes. Error bars represent standard deviations from the data of multiple species.

We calculated the A + T content in the whole mitogenomes as well as  $Ka$  and  $Ks$  of the 13 PCGs of each taxon, and then mapped the values onto the consensus tree (Figure 4). The A + T content in Pentatomidae ( $74.7\% \pm 3.7\%$ ) was higher than that in other families, and the lowest A + T content values were present in Plataspidae ( $70.5\% \pm 2.6\%$ ) and Cydnidae ( $70.6\% \pm 3.3\%$ ). Furthermore, an accelerated evolutionary rate was observed in Urostylididae. The average  $Ka$  (0.311) and  $Ks$  (1.351) in Urostylididae were significantly higher than those of other families, and its branch length (1.021) was the longest. The compositional biases and accelerated evolutionary rates of pentatomoid mitogenomes may be the reason for the unstable phylogenetic relationships and low nodal support values among the corresponding families. After removing Urostylididae from taxa sampling, all MrBayes and ML trees showed a different topology among families (Figures S5 and S6): Acanthosomatidae remained the earliest branching lineage, but the remaining taxa were recovered as (Pentatomidae + (Cydnidae + ((Tessaratomidae + Dinidoridae) + (Plataspidae + Scutelleridae))))). Coincidentally, when removing the third codon positions of PCGs, the MrBayes analysis involving eight families based on PCG12RNA dataset showed the same relationships as above (Figure S7). These results show that sequences with high heterogeneity could affect the stability of tree topologies when using the site-homogeneous model.

The topology constructed by PhyloBayes analyses under the site-heterogeneous mixture model CAT + GTR (Figure 5) differed from that of the ML and MrBayes analyses. Cydnidae were the earliest branching lineage of the tree, and the other seven families were divided into two lineages: one was ((Tessaratomidae + Dinidoridae) + (Plataspidae + Scutelleridae)), while the other was ((Acanthosomatidae + Urostylididae) + Pentatomidae)). The remaining topology remained the same when Urostylididae was excluded, whereas the sup-

port values of the relationships among most of the families increased (Figure S8). Notably, the clade ((Tessaratomidae + Dinidoridae) + (Plataspidae + Scutelleridae)) was also supported by the analyses under site-homogeneous models without Urostyloidae in sampling. Consistent with previous phylogenetic analyses on Coleoptera [17], Holometabola [18] and Heteroptera [19], our results show that the site-heterogeneous mixture model can reduce system errors and provide more stable phylogenetic relationships. However, the relatively low support values of some family-level relationships indicated that using mitochondrial data only may not be sufficient to reconstruct the phylogenetic relationships between pentatomoid families, and combining other kinds of data such as nuclear genes and morphological characters is probably needed in future analyses.



**Figure 5.** Phylogenetic tree of Pentatomoidea inferred via Bayesian analyses based on the PCGRNA and PCG12RNA datasets. The nodal support indicates the Bayesian posterior probabilities of PCGRNA/PCG12RNA. “-” indicates node support values below 0.50.

#### 4. Conclusions

Based on a dense taxa sampling, we discovered strong heterogeneity in terms of base composition as well as contrasting evolutionary rates among the pentatomoid mitogenomes. Our study demonstrated that denser sampling, data optimizing (e.g., removing taxa with high heterogeneity and/or the third codon positions of PCGs) and the use of the appropriate evolutionary model (e.g., the site-heterogeneous mixture model CAT + GTR)

are necessary strategies for better resolving the family-level phylogeny of Pentatomoidea in future studies.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/genes12091306/s1>, Figure S1: Control regions of the 44 complete pentatomoid mitogenomes; Figure S2: AliGROOVE analysis of 55 pentatomoid species based on first codon positions of PCGs (a), second codon positions of PCGs (b) and third codon positions of PCGs (c); Figure S3: Phylogenetic tree of Pentatomoidea inferred via MrBayes based on the PCGRNA dataset; Figure S4: Phylogenetic tree of Pentatomoidea inferred via IQ-TREE based on the PCGRNA and PCG12RNA datasets; Figure S5: Phylogenetic tree of Pentatomoidea (excluding Urostylididae) inferred via MrBayes based on the PCGRNA and PCG12RNA datasets; Figure S6: Phylogenetic tree of Pentatomoidea (excluding Urostylididae) inferred via IQ-TREE based on the PCGRNA and PCG12RNA datasets; Figure S7: Phylogenetic tree of Pentatomoidea inferred via MrBayes based on the PCG12RNA dataset; Figure S8: Phylogenetic tree of Pentatomoidea (excluding Urostylididae) inferred via PhyloBayes based on the PCGRNA and PCG12RNA datasets; Table S1: Collection information of taxa used in the present study; Table S2: Partition strategies used in phylogenetic analyses under site-homogeneous model.

**Author Contributions:** Conceptualization, H.L. and W.C.; Methodology, H.L., F.S. and W.C.; Investigation, S.X., Y.W., P.Z. and Z.C.; Formal analysis, S.X., Y.W., Y.L. and F.S.; Writing—original draft preparation, S.X. and Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China [grant numbers 31922012, 31730086].

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All mitochondrial genome sequences used in this study have been deposited into GenBank (accession numbers: MW847231–MW847250).

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

## References

- McPherson, J.E. *Invasive Stink Bugs and Related Species (Pentatomoidea)*; CRC Press: Boca Raton, FL, USA, 2018; p. 819.
- Rider, D.A. Pentatomoidea Home Page. North Dakota State University. Available online: <https://www.ndsu.edu/faculty/rider/Pentatomoidea/> (accessed on 24 November 2015).
- Weirauch, C.; Schuh, R.T. Systematics and evolution of Heteroptera: 25 years of progress. *Annu. Rev. Entomol.* **2011**, *56*, 487–510. [[CrossRef](#)] [[PubMed](#)]
- Grazia, J.; Schuh, R.T.; Wheeler, W.C. Phylogenetic relationships of family groups in Pentatomoidea based on morphology and DNA sequences (Insecta: Heteroptera). *Cladistics* **2008**, *24*, 932–976. [[CrossRef](#)]
- Gapud, V. A generic revision of the subfamily Asopinae with consideration of its phylogenetic position in the family Pentatomidae and superfamily Pentatomoidea (Hemiptera-Heteroptera). *Philipp. Entomol.* **1991**, *8*, 865–961.
- Schuh, R.T.; Weirauch, C. *True Bugs of the World (Hemiptera: Heteroptera) Classification and Natural History*, 2nd ed.; Siri Scientific Press: Manchester, UK, 2020; p. 767.
- Yuan, M.L.; Zhang, Q.L.; Guo, Z.L.; Wang, J.; Shen, Y.Y. Comparative mitogenomic analysis of the superfamily Pentatomoidea (Insecta: Hemiptera: Heteroptera) and phylogenetic implications. *BMC Genom.* **2015**, *16*, 460. [[CrossRef](#)]
- Tian, X.X.; Xie, Q.; Li, M.; Gao, C.Q.; Cui, Y.; Xi, L.; Bu, W.J. Phylogeny of pentatomomorphan bugs (Hemiptera-Heteroptera: Pentatomomorpha) based on six *HOX* gene fragments. *Zootaxa* **2011**, *2888*, 57–68. [[CrossRef](#)]
- Liu, Y.Q.; Li, H.; Song, F.; Zhao, Y.S.; 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](#)]
- Zhao, Q.; Wang, J.; Wang, M.Q.; Cai, B.; Wei, J.F. Complete mitochondrial genome of *Dinorhynchus dybowskyi* (Hemiptera: Pentatomidae: Asopinae) and phylogenetic analysis of Pentatomomorpha species. *J. Insect Sci.* **2018**, *44*, 1–12. [[CrossRef](#)]
- Xu, Z.Q. Phylogeny and cladistic analysis of the Pentatomoidea (Hemiptera: Heteroptera). *Acta Agric. Univ. Pekin.* **1993**, *19*, 49–53.
- Lis, J.A.; Lis, P.; Ziaja, D.J.; Kocorek, A. Systematic position of Dinidoridae within the superfamily Pentatomoidea (Hemiptera: Heteroptera) revealed by the Bayesian phylogenetic analysis of the mitochondrial 12S and 16S rDNA sequences. *Zootaxa* **2012**, *3423*, 61–68. [[CrossRef](#)]
- Li, H.M.; Deng, R.Q.; Wang, J.W.; Chen, Z.Y.; Jia, F.L.; Wang, X.Z. A preliminary phylogeny of the Pentatomomorpha (Hemiptera: Heteroptera) based on nuclear 18S rDNA and mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **2005**, *37*, 313–326. [[CrossRef](#)]

14. Wu, Y.Z.; Rédei, D.; Eger, J., Jr.; Wang, Y.H.; Wu, H.Y.; Carapezza, A.; Kment, P.; Cai, B.; Sun, X.Y.; Guo, P.L.; et al. Phylogeny and the colourful history of jewel bugs (Insecta: Hemiptera: Scutelleridae). *Cladistics* **2018**, *34*, 502–516. [[CrossRef](#)]
15. Singh-Pruthi, H. The morphology of the male genitalia in Rhynchota. *Trans. R. Entomol. Soc. Lond.* **1925**, *73*, 127–267. [[CrossRef](#)]
16. Li, H.; Shao, R.F.; Song, N.; Song, F.; Jiang, P.; Li, Z.H.; Cai, W.Z. Higher-level phylogeny of Paraneopteran insects inferred from mitochondrial genome sequences. *Sci. Rep.* **2015**, *5*, 8527. [[CrossRef](#)]
17. Timmermans, M.J.T.N.; Barton, C.; Haran, J.; Ahrens, D.; Culverwell, C.L.; Ollikainen, A.; Dodsworth, S.; Foster, P.G.; Bocak, L.; Vogler, A.P. Family-level sampling of mitochondrial genomes in Coleoptera: Compositional heterogeneity and phylogenetics. *Genome Biol. Evol.* **2015**, *8*, 161–175. [[CrossRef](#)] [[PubMed](#)]
18. Song, F.; Li, H.; Jiang, P.; Zhou, X.; Liu, J.; Sun, C.; Vogler, A.P.; Cai, W.Z. Capturing the phylogeny of Holometabola with mitochondrial genome data and Bayesian site-heterogeneous mixture models. *Genome Biol. Evol.* **2016**, *8*, 1411–1426. [[CrossRef](#)]
19. Liu, Y.Q.; Song, F.; Jiang, P.; Wilson, J.; Cai, W.Z.; Li, H. Compositional heterogeneity in true bug mitochondrial phylogenomics. *Mol. Phylogenet. Evol.* **2018**, *118*, 135–144. [[CrossRef](#)]
20. Li, H.; Leavengood, J.M.; Chapman, E.G.; Burkhardt, D.; Song, F.; Jiang, P.; Liu, J.P.; Cai, W.Z. Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. *Proc. R. Soc. B* **2017**, *284*, 20171223. [[CrossRef](#)]
21. Nie, R.E.; Breeschoten, T.; Timmermans, M.J.T.N.; Nadein, K.; Xue, H.J.; Bai, M.; Huang, Y.; Yang, X.K.; Vogler, A.P. The phylogeny of Galerucinae (Coleoptera: Chrysomelidae) and the performance of mitochondrial genomes in phylogenetic inference compared to nuclear rRNA genes. *Cladistics* **2017**, *34*, 113–130. [[CrossRef](#)]
22. Gillett, C.P.D.T.; Crampton-Platt, A.; Timmermans, M.J.T.N.; Jordal, B.H.; Emerson, B.C.; Vogler, A.P. Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Mol. Biol. Evol.* **2014**, *31*, 2223–2237. [[CrossRef](#)]
23. Timmermans, M.J.T.N.; Lees, D.C.; Thompson, M.J.; Sáfián, S.; Brattström, O. Mitogenomics of “Old World *Acraea*” butterflies reveals a highly divergent “*Bematistes*”. *Mol. Phylogenet. Evol.* **2016**, *97*, 233–241. [[CrossRef](#)]
24. Schmieder, R.; Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **2011**, *27*, 863–864. [[CrossRef](#)]
25. Peng, Y.; Leung, H.C.M.; Yiu, S.M.; Chin, F.Y.L. IDBA-UD: A de novo assembler for single cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012**, *28*, 1420–1428. [[CrossRef](#)]
26. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Drummond, Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)]
27. Meng, G.L.; Li, Y.Y.; Yang, C.T.; Liu, S.L. MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res.* **2019**, *47*, e63. [[CrossRef](#)]
28. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [[CrossRef](#)]
29. Laslett, D.; Canbäck, B. ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **2008**, *24*, 172–175. [[CrossRef](#)]
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](#)]
32. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [[CrossRef](#)]
33. Katoh, K.; Kuma, K.; Toh, H.; Miyata, T. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **2005**, *33*, 511–518. [[CrossRef](#)]
34. Abascal, F.; Zardoya, R.; Telford, M.J. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **2010**, *38*, W7–W13. [[CrossRef](#)]
35. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)] [[PubMed](#)]
36. Kück, P.; Meid, S.A.; Groß, C.; Wägele, J.W.; Misof, B. AliGROOVE-visualization of heterogeneous sequence divergence within multiple sequence alignments and detection of inflated branch support. *BMC Bioinform.* **2014**, *15*, 294. [[CrossRef](#)] [[PubMed](#)]
37. Lanfear, R.; Calcott, B.; Ho, S.Y.W.; Guindon, S. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Mol. Biol. Evol.* **2012**, *29*, 1695–1701. [[CrossRef](#)]
38. Trifinopoulos, J.; Nguyen, L.T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast-online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [[CrossRef](#)] [[PubMed](#)]
39. 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](#)] [[PubMed](#)]
40. Lartillot, N.; Rodrigue, N.; Stubbs, D.; Richer, J. PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **2013**, *62*, 611–615. [[CrossRef](#)]

41. Lartillot, N.; Philippe, H.A. Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **2004**, *21*, 1095–1109. [[CrossRef](#)]
42. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* **1986**, *17*, 57–86.
43. Sun, Z.Q.; Liu, Y.Q.; 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](#)]
44. Zhao, Y.S.; Jiang, M.J.; Wu, Y.F.; Song, F.; Cai, W.Z.; Li, H. Mitochondrial genomes of three kissing bugs (Reduviidae: Triatominae) and their phylogenetic implications. *Int. J. Biol. Macromol.* **2019**, *134*, 36–42. [[CrossRef](#)] [[PubMed](#)]
45. Wang, J.J.; Wu, Y.F.; Dai, R.H.; Yang, M.F. Comparative mitogenomes of six species in the subfamily Iassinae (Hemiptera: Cicadellidae) and phylogenetic analysis. *Int. J. Biol. Macromol.* **2020**, *149*, 1294–1303. [[CrossRef](#)]
46. Yan, L.P.; Xu, W.T.; Zhang, D.; Li, J.Q. Comparative analysis of the mitochondrial genomes of flesh flies and their evolutionary implication. *Int. J. Biol. Macromol.* **2021**, *174*, 385–391. [[CrossRef](#)]
47. Yua, L.L.; Ge, X.Y.; Xie, G.L.; Liu, H.Y.; Yang, Y.X. First Complete mitochondrial genome of Melyridae (Coleoptera, Cleroidea): Genome description and phylogenetic implications. *Insects* **2021**, *12*, 87.
48. Wang, J.J.; Yang, M.F.; Dai, R.H.; Li, H.; Wang, X.Y. Characterization and phylogenetic implications of the complete mitochondrial genome of Idiocerinae (Hemiptera: Cicadellidae). *Int. J. Biol. Macromol.* **2018**, *120*, 2366–2372. [[CrossRef](#)]
49. Zhao, Q.; Chen, C.; Liu, J.; Wei, J.F. Characterization of the complete mitochondrial genome of *Eysarcoris aeneus* (Heteroptera: Pentatomidae), with its phylogenetic analysis. *Mitochondrial DNA Part B* **2019**, *4*, 2096–2097. [[CrossRef](#)]