

SCIENTIFIC REPORTS



OPEN

Molecular phylogeny of Polyneoptera (Insecta) inferred from expanded mitogenomic data

Nan Song¹, Hu Li², Fan Song² & Wanzhi Cai²

Received: 22 August 2016

Accepted: 06 October 2016

Published: 26 October 2016

The Polyneoptera represents one of the earliest insect radiations, comprising the majority of hemimetabolous orders, in which many species have great economic importance. Here, we sequenced eleven mitochondrial genomes of the polyneopteran insects by using high throughput pooled sequencing technology, and presented a phylogenetic reconstruction for this group based on expanded mitochondrial genome data. Our analyses included 189 taxa, of which 139 species represent all the major polyneopteran lineages. Multiple results support the monophyly of Polyneoptera, the monophyly of Dictyoptera, and the monophyly of Orthoptera. Sister taxon relationships Plecoptera + Dermaptera, and Zoraptera + Embioptera are also supported by most analyses. Within Dictyoptera, the Blattodea is consistently retrieved as paraphyly due to the sister group relationship of *Cryptocercus* with Isoptera. In addition, the results demonstrate that model selection, data treatment, and outgroup choice can have significant effects on the reconstructed phylogenetic relationships of Polyneoptera.

Polyneoptera includes some early divergent and diverse hemimetabolous insects, which are constituted by 11 higher taxa conventionally ranked as orders: Mantodea (mantids), Phasmatodea (stick insects), Orthoptera (crickets, katydids, grasshoppers, locusts, etc.), Dermaptera (earwigs), Grylloblattodea (rock crawlers), Zoraptera (zorapterans), Plecoptera (stoneflies), Blattodea (cockroaches), Isoptera (termites), Embioptera (webspinners), and the recently discovered insect order Mantophasmatodea (heelwalkers)^{1–4}. The importance as agriculturally dominant insect pests (e.g. locusts), or natural enemies (e.g. mantids), or vectors of disease-causing pathogens (e.g. cockroaches) has resulted in intense and detailed biological study of many species of Polyneoptera. Phylogenetic relationships within Polyneoptera have been widely researched, and various hypotheses have been raised to explain the evolution of these insects^{5–9}. However, few of these studies involved a comprehensive taxon sampling, and evaluated systematically the relationships among polyneopteran orders.

The monophyly of the entire Polyneoptera is generally accepted^{10–12}, yet phylogenetic relationships within this group are still contentious. The superorder Dictyoptera and the order Orthoptera are two well resolved high level relationships of Polyneoptera. Within the Dictyoptera, the hierarchy arrangement of the clade (Mantodea + (Blattodea + Isoptera)) has been well supported by molecular evidence^{12–15}. Termites are actually social cockroaches, with the family Cryptocercidae as their closest relative rendering the Blattodea paraphyletic, while the Mantodea is the sister group to a monophyletic cockroach/termite clade. Within Orthoptera, two suborders are recognized, namely the Ensifera and Caelifera. The monophyly of Orthoptera is well established by mitochondrial genome data^{9,16}.

Despite the sister-group relationship between Embioptera and Phasmatodea is well supported by several morphological character systems^{17–19} and molecular studies^{12,13,20}, some analyses alternatively supported a closer relationship between Embioptera and Zoraptera^{12,21}. Additional hypotheses of embiopteran affinities to other insect lineages had also been suggested². Besides the uncertain placement of Embioptera, a universal consensus on the positions of Zoraptera, Plecoptera, Dermaptera, and Mantophasmatodea remains elusive^{15,22–26}. For example, no less than ten hypotheses on the phylogenetic placement of Zoraptera have been proposed^{21,27}. As addressed above, both morphological characters of wing base structure²¹ and mitochondrial genome data¹¹ supported Zoraptera as sister to Embioptera. However, nuclear genes supported a close relationship between Zoraptera and Dictyoptera^{28,29}. A more recent study based on genome-scale data recovered a sister-group relation between Zoraptera and Dermaptera¹². The latter arrangement was also supported by a combined analysis of morphological and molecular data³⁰.

¹College of Plant Protection, Henan Agricultural University, Zhengzhou, China. ²Department of Entomology, China Agricultural University, Beijing, China. Correspondence and requests for materials should be addressed to N.S. (email: songnan@henau.edu.cn) or W.C. (email: caiwz@cau.edu.cn)

Mitogenome has been extensively used to infer phylogeny of insect^{3,9,11,16,31,32}. Compared to the whole genome, mitogenome is more readily sequenced with an array of relatively conserved primers and a reasonable cost. And it contains more phylogenetic information than single or multi-gene data. However, the lineage-specific elevated rates of nucleotide substitutions or base compositional bias leads to the problem of long-branch attraction/non-stationarity, which lower the phylogenetic resolving power of mitogenome for higher level insect phylogenetics. In model-based phylogenetic analyses, vast majority of evolutionary models assume the stationarity of the replacement rate. That is all taxa in a data set evolving clocklike. The extremely sequence composition heterogeneity may seriously interfere in the stationarity of the data set, and further result in systematic error during inference of phylogeny. Strategies for dealing with these problems have been developed, for example, the advanced evolutionary models accounting for sequence compositional heterogeneity, and data coding methods to avoid the impact of potentially uninformative data partitions.

In the present study, we applied high-throughput, pooled sequencing technology^{33–35} to determine eleven polyneopteran insect mitogenomes (i.e. six from Orthoptera, three from Dermoptera, and each one from Mantodea and Blattodea) to expand currently available mitogenomic data. We conducted a series of molecular phylogenetic analyses to assess the effects of varying analysis parameters including model settings, choice of taxa and data coding methods on phylogenetic inference. Comparisons between different analytical methods were performed, with the aim of providing insights into the suitability of mitogenome for phylogenetic reconstruction of high level relationships within Polyneoptera.

Materials and Methods

Taxon sampling. Besides 11 newly sequenced mitogenomes, the remaining mitogenome sequences of each order in Polyneoptera were retrieved from GenBank (Table S1). Two series of data sets were created. Firstly, 154 taxa data set were created, of which ingroup are comprised of 4 species from Dermoptera, 10 from Plecoptera, 18 from Phasmatodea, 10 from Mantodea, 15 from Blattodea, 17 from Isoptera, 18 from Ensifera of Orthoptera, 43 from Caelifera of Orthoptera, and each one from Grylloblattodea, Mantophasmatodea, Zoraptera, and Embioptera. The Archaeognatha (three species), Thysanura (three species), Ephemeroptera (three species), and Odonata (four species) were selected as outgroup. Secondly, to assess the effect of taxa sampling on the phylogenetic inference of Polyneoptera, additional outgroup taxa were added into the data set of 154 taxa to construct 189 taxa data set. The 189 taxa data set included more outgroup species from Paraneoptera (14 species), and Holometabola (21 species).

DNA extraction and mitogenome assembly. Genomic DNA was extracted from each 95–100% ethanol preserved specimen individually using the TIANamp Micro DNA Kit (TIANGEN BIOTECH CO., LTD) following the manufacturer's protocol. DNA concentration was measured by Nucleic acid protein analyzer (QUAWELL TECHNOLOGY INC.).

Largely identical quantities of genomic DNA from each of the samples were pooled, and the resulting DNA was quantified to be 1.5 ug. The pooled DNA sample was utilized for library construction and sequenced by the Illumina HiSeq2500 platform. For the sequenced sample, 20 Gb paired-end reads of 125 bp length were generated. The NGSQC-Toolkit v2.3.3³⁶ software was used for quality control (avg. Q20 > 85%, avg. Q30 > 80%). Reads containing ambiguous bases (N) or shorter than 88 bp were also removed by NGSQC-Toolkit. Finally, no less than 15 Gb high-quality filtered reads were used in de novo assembly with SOAPdenovo v. 2.0³⁷.

To identify the mitogenome assemblies from the pooled sequencing files, three different fragments of mtDNA (*cox1* 5' region, *cytb* 3' region and *rrnS* 5' region) were amplified as "Bait" sequences by standard PCR reactions using primers designed with reference of Simon *et al.*³⁸ (primers listed in Table S2). Local BLAST searches were conducted with BioEdit version 7.0.5.3³⁹ for each bait sequence reference against all corresponding assemblies. Only hits with 100% pairwise identity were considered a successful identification. The identified mitogenomic sequences were inputted into MITOS web service for initial mitogenome annotation⁴⁰. The resultant gene boundaries were checked and corrected by alignment with published mitogenome sequences of each individual insect order.

Sequence alignment and data concatenation. All thirty-seven genes of insect mitogenome were used for phylogenetic analyses, namely the 13 protein-coding genes, 22 tRNA genes and two rRNA genes. For protein-coding genes, firstly stop codons were excluded. Subsequently, each was aligned based on the invertebrate mitochondrial genetic code with Perl script TransAlign⁴¹. Each tRNA or rRNA gene was aligned using MAFFT (version 7)⁴² under iterative refinement method incorporating the most accurate local (E-INS-i) pairwise alignment information. The resulting alignments were checked in MEGA 6⁴³, and ambiguously aligned positions were manually excluded prior to phylogenetic analyses. Finally, all alignments were concatenated in a single matrix using FASconCAT_v1.0⁴⁴.

Nucleotide homogeneity across taxa was assessed using the Chi-square test⁴⁵ implemented in PAUP*4.0b10⁴⁶. Potential saturation in the combined protein-coding genes was assessed using the index of substitution saturation (*I_{ss}*)⁴⁷ implemented in the DAMBE 5⁴⁸. The genetic distances between major groups were measured using Kimura 2-parameter, Tamura-Nei, and Maximum Composite Likelihood model in the program MEGA with default settings, based on data set of 189taxa_PCGRNA, and using the Archaeognatha as reference. Nucleotide and amino acid composition were also measured with MEGA.

Phylogeny. Data re-coding and partitioning. To alleviate the effect of substitution saturation of codon positions on phylogenetic estimation, protein-coding genes were re-coded by degenerating all sites including synonymous substitutions to IUPAC ambiguity codes through Degen v1.4 (Degen-code)^{49,50}. Thus, two concatenated data sets were used in phylogenetic analyses: PCGRNA and PCGDegenRNA. The deduced amino acid sequences

Gene regions	NumOTU	Iss	Iss.cSym	P	Iss.cAsym	P
1st codons	32	0.420	0.808	0.0000	0.554	0.0000
2nd codons	32	0.284	0.808	0.0000	0.554	0.0000
3rd codons	32	0.765	0.808	0.0000	0.554	0.0000
all codons	32	0.468	0.818	0.0000	0.572	0.0000
RNAs	32	0.477	0.808	0.0000	0.550	0.0000

Table 1. Saturation testing based on data sets with 154 taxa. Note: two-tailed tests are used.

from protein-coding genes were also used for tree reconstruction. Totally, five data sets with 154 taxa (i.e. PCG, PCG_AA, PCGDegen, PCGRNA, and PCGDegenRNA) were analyzed using maximum likelihood, and Bayesian inference methods. In addition, the combined data sets of 189taxa_PCGRNA and 189taxa_PCGDegenRNA were used to investigate the effect of taxon sampling on tree building.

Prior to analyses, PartitionFinder⁵¹ was employed to infer the optimal partitioning strategy. Furthermore, we used the Bayesian Information Criterion (BIC) to choose the best model for the final combined nucleotide data set under a greedy search with RAXML⁵². The data blocks were defined by gene types (each of 13 PCGs, each of 22 tRNAs and 2 rRNAs) and by codon position (each of three codon positions for PCGs), totally 63 independent blocks were employed for the full data sets. The partition schemes and best-fit models selected for each data set are provided in Table S3.

Tree searches. Maximum likelihood (ML) analyses were conducted using RAXML⁵² as implemented in the CIPRES Portal⁵³. Branch support was estimated using rapid nonparametric bootstrapping with 1000 replicates. For concatenated data set, the above determined models were applied to the optimal partitions.

Bayesian analyses were performed using a parallel version (pb_mpi1.5a)^{54,55} as implemented on a HP server with twenty-four CPU and 64 G memory. The model CAT-GTR was used for nucleotide analyses, while the model CAT for amino acids. Two chains were run in parallel, and started from a random topology. The Maximum \backslash maxdiff" value to accept was set as 0.1.

The bootstrap supports (BS) of ≥ 75 and posterior probabilities (PP) of ≥ 0.95 were considered to be credible support values for tree nodes. All sequence and tree files constructed in this article are available in the TreeBASE (<http://purl.org/phylo/treebase/phylovs/study/TB2:S19694>).

Tree topology test. In order to determine whether our mitogenome data significantly reject the alternative tree topologies or hypotheses, the topology test were performed on the data set of 154taxa_PCGDegenRNA. The site-log-likelihood values were calculated under the GTR + I + G model using TREE-PUZZLE (Version 5.2)⁵⁶. And then, the obtained values were used as input for the software CONSEL⁵⁷. All constraint likelihood trees were generated by RAXML⁵² as implemented above.

Results

Assembly of mitogenomes. A total of 30 insect species were pooled for high-throughput sequencing. After classification based on morphological and molecular characters, 11 individuals were identified as the polyneopterans and used for this study. The mitogenome sequences for these 11 insects showed a 867–2413 \times coverage that generally resulted in contigs of 8,382–16,522 bp (Fig. S1). Except for two species from Dermaptera (Anisolabididae sp.: 8382 bp, and *Labidura japonica*: 10981 bp), the remaining insects have complete or nearly complete mitogenome assemblies, which range from 15,366 bp (*Tridactylus* sp.) to 16,522 bp (*Ducetia* sp.) in length. The missing fragments were mainly located in the control region and the adjacent regions. All newly determined sequences has been deposited in GenBank (accession numbers: KX673195–KX673205).

Sequence characteristics. Every data set exhibits significant departures from base-compositional homogeneity across taxa ($p = 0.000$, Chi-square test of compositional homogeneity across taxa). Even with Degen-coding treatment, homogeneity could not be retrieved. Thus, non-stationarity was assumed for all data sets in further analyses. The results of the substitution saturation tests showed that the values of substitution saturation index (*Iss*) for the first and second codons in PCGs and all sites of RNAs were significantly smaller than the critical values (*Iss.cSym* or *Iss.cAsym*). However, the *Iss* values of the third codon positions were larger than the *Iss.cAsym* (Table 1). This indicated that the third codon positions might provide poor information for phylogenetics under the assumption of a very asymmetrical true tree.

Pairwise Kimura 2-parameter genetic distance across all main groups ranged from 0.340 (Cryptocercidae) to 0.546 (Strepsiptera). The calculations from both Tamura-Nei and Maximum Composite Likelihood also provided similar results to those based on the Kimura 2-parameter model (Table 2). The results showed that the genetic distances in the mitogenome sequences of Strepsiptera, Zoraptera, Hymenoptera, Embioptera, and Dermaptera were higher than 0.5, while the average genetic distances are lesser than 0.45. The higher genetic distances indicated that the analyzed mitogenome sequences of these insects could have higher evolutionary rate than others.

Next, sequence compositional heterogeneity were investigated through calculating the GC nucleotide content on 1st and 2nd sites and the frequency of GC rich amino acids (G, A, R, P) (Fig. 1). The outgroup taxa and some species of both Plecoptera and Isoptera were characterized by high GC% (low AT%), while most of the Holometabola and Paraneoptera had higher AT% (from the mean 65% of Coleorrhyncha to 76% of Hymenoptera). The ingroup polyneopterans had moderate to high GC % and amino acids content of G, A, R and

Jor lineages	K2P	TN	ML
Cryptocercidae	0.340	0.348	0.347
Blattodea	0.362	0.373	0.372
Mecoptera	0.371	0.383	0.382
Neuroptera	0.371	0.384	0.383
Diptera	0.372	0.384	0.383
Mantophasmatodea	0.394	0.408	0.407
Grylloblattodea	0.398	0.410	0.407
Raphidioptera	0.399	0.414	0.414
Mantodea	0.401	0.415	0.415
Orthoptera	0.404	0.416	0.417
Phasmatodea	0.405	0.420	0.420
Lepidoptera	0.406	0.423	0.420
Psocoptera	0.409	0.425	0.425
Plecoptera	0.412	0.424	0.423
Coleoptera	0.418	0.433	0.434
Odonata	0.419	0.432	0.431
Blattodea	0.423	0.434	0.435
Isoptera	0.432	0.445	0.441
Ephemeroptera	0.434	0.447	0.446
Thysanura	0.462	0.477	0.476
Hemiptera	0.468	0.483	0.487
Dermaptera	0.512	0.534	0.528
Embioptera	0.524	0.547	0.548
Hymenoptera	0.533	0.564	0.563
Zoraptera	0.534	0.555	0.561
Strepsiptera	0.546	0.578	0.577
Avg.	0.429	0.445	0.444

Table 2. Genetic distances for major groups analyzed in this study.

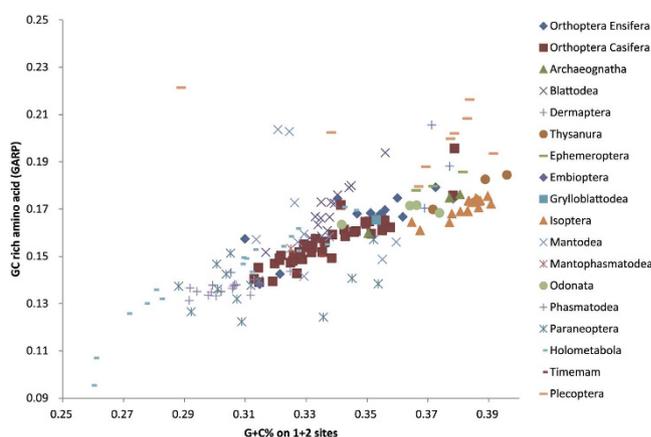


Figure 1. Compositional properties of species analyzed in this study. Both the GC nucleotide content on 1st and 2nd codon positions and the frequency of GC rich amino acids (G, A, R, P) were calculated.

P, and whose values were overlapped with those from outgroup taxa (e.g. the order Archaeognatha and Odonata). These seemed to express a trend on the primitive insect orders with higher GC content in the mtDNA while the relative derived orders with higher AT content. The information shown in the Fig. 2 was consistent with the results of genetic distances presented above, namely that both outgroup Archaeognatha, Thysanura, Odonata, Ephemeroptera and ingroup Polyneoptera) might share similar evolutionary context.

Maximum-likelihood analysis. The different data treatments under maximum likelihood inference with homogeneous GTR model resulted in different tree topology. Only the data set of 154taxa_PCGDegen produced a monophyletic Polyneoptera, while the nested position of Ephemeroptera rendered Polyneoptera to be non-monophyletic in the trees from the remaining four data sets (i.e. 154taxa_PCG, 154taxa_PCG_AA, 154taxa_PCGRNA, and 154taxa_PCGRNA). In addition, the monophyly of Orthoptera was only recovered by

Item	Hypothesis	-Ln likelihood	AU	KH	SH	WKH	WSH
1	(Orthoptera)	509138.17	0.566	0.397	0.934	0.397	0.935
2	(Mantodea, (Blattodea, (<i>Cryptocercus</i> , Isoptera)))	509152.06	0.274	0.324	0.908	0.228	0.916
3	(Isoptera, (Blattodea, Mantodea))	509388.38	0.000	0.000	0.004	0.000	0.000
4	(Phasmatodea)	509186.32	0.009	0.145	0.550	0.006	0.055
5	(Phasmatodea, Embioptera)	509245.48	0.016	0.035	0.156	0.005	0.042
6	((Timematodea, Mantophasmatodea), Phasmatodea)	509241.00	0.011	0.038	0.175	0.007	0.058
7	(Zoraptera, Embioptera)	509150.89	0.532	0.332	0.915	0.332	0.979
8	(Zoraptera, Dermaptera)	509229.22	0.012	0.047	0.211	0.006	0.047
9	(Mantophasmatodea, Grylloblattodea)	509151.56	0.416	0.331	0.904	0.331	0.954
10	(Mantophasmatodea, Mantodea)	509358.74	0.000	0.000	0.005	0.000	0.000
11	(Grylloblattodea, Dermaptera)	509243.56	0.006	0.036	0.177	0.003	0.030
12	(Plecoptera, Embioptera)	509338.30	0.002	0.000	0.006	0.000	0.000
13	(Plecoptera, Dermaptera)	509153.77	0.199	0.252	0.839	0.157	0.658
14	(Plecoptera, all other Polyneoptera)	509141.99	0.420	0.331	0.905	0.331	0.894
15	(Dermaptera, Embioptera)	509271.56	0.006	0.018	0.083	0.001	0.011
16	The monophyletic Polyneoptera	509128.54	0.655	0.603	0.978	0.603	0.981
17	((Mantodea, (Blattodea, (<i>Cryptocercus</i> , Isoptera))), ((Dermaptera, (Orthoptera, Phasmatodea), (Mantophasmatodea, Grylloblattodea))), (Embioptera, (Plecoptera, Zoraptera)))) (Grimaldi and Engel ⁵⁸)	509645.80	0.042	0.000	0.000	0.000	0.000
18	((Orthoptera, (Phasmatodea, Embiidina), ((Mantophasmatodea, Grylloblattodea), (Mantodea, (Blattodea, (<i>Cryptocercus</i> , Isoptera))), (Plecoptera, (Dermaptera, Zoraptera)))) (Terry and Whiting ³⁰)	509413.39	0.000	0.000	0.001	0.000	0.000
19	((Embioptera, Phasmatodea), ((Orthoptera, (Zoraptera, Dictyoptera))), ((Grylloblattodea, Mantophasmatodea), (Dermaptera, Plecoptera)))) (Ishiwata <i>et al.</i> ¹⁵)	509579.43	0.002	0.000	0.000	0.000	0.000
20	((Zoraptera, Embioptera), (Plecoptera, ((Orthoptera, Phasmatodea), ((Mantodea, (Blattodea, (<i>Cryptocercus</i> , Isoptera))), ((Grylloblattodea, Mantophasmatodea), (Dermaptera)))))) (Yoshizawa ¹⁰)	509516.25	0.000	0.000	0.000	0.000	0.000
21	((Plecoptera, Dermaptera), (Orthoptera, ((Mantodea, (Blattodea, (<i>Cryptocercus</i> , Isoptera))), (Grylloblattodea, ((Zoraptera, Embioptera), Phasmatodea), (Timematodea, Mantophasmatodea)))))) (Ma <i>et al.</i> ¹¹)	509146.00	0.357	0.293	0.875	0.293	0.862
22	((Zoraptera, Dermaptera), (Plecoptera, (Orthoptera, ((Grylloblattodea, Mantophasmatodea), (Phasmatodea, Embioptera))), (Mantodea, (Blattodea, (<i>Cryptocercus</i> , Isoptera)))))) (Misof <i>et al.</i> ¹²)	509277.26	0.001	0.000	0.056	0.000	0.000

Table 3. Tree topology testing based on the data set of 154taxa_PCGDegenRNA. Bold indicates the values of $p > 0.05$.

phylogenetic information supporting the sister relationships between Zoraptera and Embioptera. The arrangements of Grylloblattodea, Mantophasmatodea, and Dictyoptera were identical to those from ML analysis.

Effect of additional outgroup taxa. To assess the effect of the addition of outgroup mitogenome sequence data on the phylogenetic estimation, we reran combined analyses including additional 35 outgroup taxa representatives of the Paraneoptera and Holometabola. The results showed that including more distantly related outgroups resulted in a worse resolution of mitogenomic phylogeny of Polyneoptera. The outgroups Paraneoptera and Holometabola always fall within the ingroup, which rendered the Polyneoptera to be a non-monophyletic group. On the whole, Bayesian inferences on the data sets of 189 taxa performed better than corresponding ML analyses. Specially, using CAT-GTR model effectively dragged the outgroup taxa of Paraneoptera and Holometabola into a more basal position of trees.

Hypotheses test. Among alternative hypotheses, the following relationships were supported: the monophyletic Polyneoptera, the monophyletic Orthoptera, the sister relationships between Zoraptera + Embioptera, Mantophasmatodea + Grylloblattodea, and Plecoptera + Dermaptera, the superorder Dictyoptera (Mantodea + “Blattodea” + (*Cryptocercus* + Isoptera))), and the first diverging position of Plecoptera (Table 3). As for the entire Polyneoptera phylogenies suggested by previous studies of Grimaldi and Engel⁵⁸, Terry and Whiting³⁰, Ishiwata *et al.*¹⁵, Yoshizawa¹⁰, Ma *et al.*¹¹, and Misof *et al.*¹², the data analyzed in the present study significantly supported the hypothesis in Ma *et al.*¹¹, and rejected others.

Discussion

Although mitochondrial genome was frequently utilized to phylogenetics of insects, some weakness also limited the phylogenetic resolving power of this kind of DNA maker, such as base compositional heterogeneity and

lineage-specific evolutionary rate^{32,59,60}. As a result, some previous phylogenetic reconstructions from mitogenome were at odds with the traditional views of insect phylogeny. Strategies to ameliorate these problems usually comprise increased taxonomic sampling, employing different data treatment methods, and implementing better-fitting models^{32,61}. As an improvement, in this study we applied the strategies as mentioned above and attempted to provide some new insights into the polyneopteran phylogeny.

Monophyly of Polyneoptera. In previous literatures, some analyses based on nuclear gene sequences found Polyneoptera to be monophyletic^{15,23}. Whereas, other analyses based on the similar gene fragments using different analytical methods implied a paraphyletic Polyneoptera^{13,30}. Thus, the uncertainty from the prior studies required additional researches employing more informative molecular makers to reconstruct the phylogeny of this insect group. In this paper, we provided a comprehensive analysis with sample of 37 genes from the whole mitogenomes of 139 species representing 11 polyneopteran orders.

When only one species of Dermaptera (i.e. *Challia fletcheri*) with remarkably long branch was initially included in the phylogenetic reconstruction, it fell within outgroups and emerged as sister taxon to the Ephemeroptera in ML trees (data not shown). This is presumably due to the effect of long-branch attraction. However, increasing three newly sequenced mitogenomes of Dermaptera into the data set successfully broke down this artificial result. Thus, the sequencing of some important and poorly sampled insect groups proved to be beneficial in tree resolution of Polyneoptera.

We also performed topological tests to examine the monophyly of Polyneoptera. All tests supported the hypothesis of the monophyletic Polyneoptera. In particular, when implementing site-heterogeneous model, the bias correlated with homoplasy in the current mitogenome data can be effectively corrected for. The monophyly of Polyneoptera were consistently retrieved across analyses on the basis of differential data treatments.

We also investigate the effect of outgroup choices on ingroup relationships. When including taxa of Paraneoptera and Holometabola into original 154 taxa data sets, the monophyly of Polyneoptera can't be retrieved by all additional data sets. This may be due to the fact that inclusion of distant outgroups with highly divergent mitogenome sequences led to highly ambiguous alignment, and weakened the resolving power of mitogenome data. This result is in agreement with the limits of mitogenome data in resolving deep phylogeny of insects^{61,62}. Specially, both Paraneoptera and Holometabola comprised some species with rapid evolutionary mitogenome sequences. Their inclusion further confound the phylogenetic reconstruction of Polyneoptera. The sequence character analyses indicate that most polyneopteran insects share close genetic distance and sequence composition with more primitive insect orders. Thus, selection of the most primitive insects as outgroups for polyneopteran phylogeny estimation is preferred to Paraneoptera and Holometabola.

In short, the hypothesis of polyneopteran monophyly is justified in this study. Furthermore, our analyses demonstrate the applicability of mitogenome data in resolving high level phylogeny of these earliest radiating insect lineages.

Relationships of the entire orders within Polyneoptera. Among the major orders in Polyneoptera, the Orthoptera are the subject of more published papers than other polyneopterans, due in part to their species richness and economic importance. In this study, PhyloBayes analyses under site-heterogeneous model revealed strong support for the monophyly of Orthoptera (posterior probability = 1.0). By contrast, some analyses under homogeneous GTR model could not reconstruct the monophyletic Orthoptera. In some cases, the suborder Ensifera represented by 18 species split earlier in the tree, while the Caelifera was resolved as a more terminal clade. This led to a non-monophyletic Orthoptera. When investigating the causes underlying this arrangement, we found that the Ensifera have more similar sequence compositional bias to outgroups than the Caelifera (Fig. 1). This factor might potentially result in the deeper position of Ensifera. In the hypothesis tests, the constrained ML tree enforcing a monophyletic Orthoptera was significantly supported, and we were thus able to reject results of the non-monophyletic Orthoptera.

Besides the Orthoptera, another sampling density was centered on the Dictyoptera. The dictyopteran monophyly is generally accepted^{22,23,63,64} and the inter-ordinal relationships of (Mantodea, (“Blattodea”, (*Cryptocercus*, Isoptera))) have received strong support^{11,12}. In the present study, these relationships among Dictyoptera orders were consistently recovered by all analyses with strong nodal support. Many biological and morphological features suggested a close affinity of *Cryptocercus* and Isoptera (e.g. the homologous sociality between the subsocial *Cryptocercus* and the eusocial termites, and the shared characters of hindgut flagellates of *Cryptocercus* and lower-grade Isoptera). Therefore, our resolution provides a further support for the currently prevailing hypothesis of Dictyoptera^{31,65,66}. Furthermore, every kind of topological test strongly supported this reconstruction.

The Phasmatodea is rendered paraphyletic in all analyses of this study. Because the *Timema* was consistently placed outside the major clade of the Phasmatodea, and led to the paraphyly of Phasmatodea. Within Polyneoptera, the Zoraptera and/ or Embioptera has a closer relation to Phasmatodea. This is congruent with previous morphological and/or molecular analyses^{1,13}. The study by Ma *et al.*¹¹ also recovered a sister relationship between Zoraptera and Embioptera, but the long branch attraction artifact were suspected for this relation, due to the very long branches exhibited by both orders. The majority of our analyses recovered this sister relationship, and it received significant support in hypothesis testing. However, the PhyloBayes analyses based on data sets of PCG alone recovered Zoraptera as an earlier diverging lineage, and leaved a sister relation of Embioptera to Phasmatodea. This arrangement is in agreement with result from Misof *et al.*¹². Thus, more studies need to elucidate the positions of Zoraptera and Embioptera, and the sequencing of more representatives from both insect orders is urgently required.

The Plecoptera is an enigmatic lineage within Polyneoptera, and its phylogenetic placement always changed across different studies. Some authors agreed on the relative primitive position of the Plecoptera^{2,22–23,67}, while the sister relationship between Plecoptera and Dermaptera has been suggested by recent molecular studies^{15,26}.

In the PhyloBayes tree from the data set of PCGDegenRNA (Fig. 2), the Plecoptera was recovered as the first lineage to diverge, and formed a sister relationship to Dermaptera. In the tree topology tests, the hypothesis of (Plecoptera + Dermaptera) received most wide support, while the hypothesis of Plecoptera as being a sister group to all other Polyneoptera could not be rejected by the current mitogenome data. Therefore, the placement of Plecoptera remains answered.

As to the Dermaptera, traditional morphology-based analyses have suggested affinities with Dictyoptera, Grylloblattaria, or Embioptera^{67–69}, respectively. But none of these arrangements could be strongly upheld². Before this study, only one mitogenome for Dermaptera (i.e. *C. fletcheri*) is available in GenBank. This mitogenome displayed unusual features of strand-bias, and a unique gene arrangement²⁶. With the inclusion of three newly sequenced Dermaptera mitogenome sequences, the long branch of *C. fletcheri* is shortened, and the whole Dermaptera consistently appears as the earliest divergent lineage of Polyneoptera.

Conclusion

In summary, the results obtained in the phylogenetic analyses of the mitogenome data are largely consistent with recently dominant hypotheses of Polyneoptera phylogeny, such as a monophyletic Orthoptera, the Isoptera nested within Blattodea and being as a sister to *Cryptocercus*, and the sister-group relation between Phasmatodea and Embioptera. Conflict in the relationships of Polyneoptera between RAxML and PhyloBayes trees exists mainly on the monophyly of Orthoptera. The result of non-monophyletic Orthoptera generated by RAxML analyses might be ascribed to the sequence compositional bias, which potentially correlates with some systematic errors. Our results also show that the heterogeneous CAT model implemented in the PhyloBayes is a powerful model to account for site-specific nucleotide or amino-acid preferences. By which we can recover a monophyletic Orthoptera and a more reasonable interrelationship of Polyneoptera.

References

- Rähle, W. Untersuchungen an Kopf und Prothorax von Embia ramburi Rimsky-Korsakow 1906. *Zool. Jahrb. Abt. Anat. Ontog. Tiere*. 248–330 (1970).
- Kristensen, N. P. Phylogeny of insect orders. *Annu. Rev. Entomol.* **26**, 135–157 (1981).
- Cameron, S. L., Barker, S. C. & Whiting, M. F. Mitochondrial genomics and the new insect order Mantophasmatodea. *Mol. Phylogenet. Evol.* **38**, 274–279 (2006).
- Damgaard, J., Klass, K. D., Picker, M. D. & Buder, G. Phylogeny of the heelwalkers (Insecta: Mantophasmatodea) based on mtDNA sequences, with evidence for additional taxa in South Africa. *Mol. Phylogenet. Evol.* **47**, 443–462 (2008).
- Sharov, A. G. Phylogeny of the Orthopteroidea. - Israel Program for Scientific Translations. *Jerusalem*. vi. **251** (1971).
- Kevan, D. K. M. Orthoptera. In: Parker, S. P. (Ed.), *Orthoptera* (McGraw-Hill Book Company, New York, 352–383, 1982).
- Gorochov, A. V. System and evolution of the suborder Ensifera (Orthoptera). *Proc. Zool. Inst., Rus. Acad. Sci.* **260**, 1–207 (1995).
- Gwynne, D. T. Phylogeny of the ensifera (orthoptera): a hypothesis supporting multiple origins of acoustical signalling, complex spermatophores and maternal care in crickets, katydids, and weta. *J. Orthoptera Res.* **4**, 203–218 (1995).
- Fenn, J. D., Song, H., Cameron, S. L. & Whiting, M. F. A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol. Phylogenet. Evol.* **49**, 59–68 (2008).
- Yoshizawa, K. Monophyletic Polyneoptera recovered by wing base structure. *Syst. Entomol.* **36**, 377–394 (2011).
- Ma, C., Wang, Y., Wu, C., Kang, L. & Liu, C. The compact mitochondrial genome of *Zorotypus medoensis* provides insights into phylogenetic position of Zoraptera. *BMC Genomics* **15**, 1156 (2014).
- Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
- Kjer, K. M., Carle, F. L., Litman, J. & Ware, J. A molecular phylogeny of Hexapoda. *Arthropod Syst. Phylo.* **64**, 35–44 (2006).
- Inward, D., Beccaloni, G. & Eggleton, P. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* **3**, 331–335 (2007).
- Ishiwata, K., Sasaki, G., Ogawa, J., Miyata, T. & Su, Z. H. Phylogenetic relationships among insect orders based on three nuclear protein-coding gene sequences. *Mol. Phylogenet. Evol.* **58**, 169–180 (2011).
- Ma, C., Liu, C., Yang, P. & Kang, L. The complete mitochondrial genomes of two band-winged grasshoppers, *Gastrimargus marmoratus* and *Oedaleus asiaticus*. *BMC Genomics* **10**, 156 (2009).
- Blanke, A. *et al.* Revival of Palaeoptera? head characters support a monophyletic origin of Odonata and Ephemeroptera (Insecta). *Cladistics* **28**, 560–581 (2012).
- Friedemann, K., Wipfler, B., Bradler, S. & Beutel, R. G. On the head morphology of Phyllium and the phylogenetic relationships of Phasmatodea (Insecta). *Acta Zool-stockholm.* **93**, 184–199 (2012).
- Mashimo, Y., Beutel, R. G., Dallai, R., Lee, C.-Y. & Machida, R. Embryonic development of Zoraptera with special reference to external morphology, and its phylogenetic implications (Insecta). *J. Morphol.* **275**, 295–312 (2014).
- Letsch, H. & Simon, S. Insect phylogenomics new insights on the relationships of lower neopteran orders (Polyneoptera). *Syst. Entomol.* **38**, 783–793 (2013).
- Yoshizawa, K. The Zoraptera problem evidence for Zoraptera+Embiodea from the wing base. *Syst. Entomol.* **32**, 197–204 (2007).
- Kristensen, N. P. Phylogeny of extant hexapods. *Ins. Aust.* **1**, 125–140 (1991).
- Wheeler, W. C., Whiting, M., Wheeler, Q. D. & Carpenter, J. M. The phylogeny of the extant hexapod orders. *Cladistics* **17**, 113–169 (2001).
- Letsch, H. O., Meusemann, K., Wipfler, B., Schütte, K., Beutel, R. & Misof, B. Insect phylogenomics: results, problems and the impact of matrix composition. *P. Roy. Soc. Lond. B. Bio.* **279**, 3282–3290 (2012).
- Simon, S., Narechania, A., DeSalle, R. & Hadrys, H. Insect phylogenomics: exploring the source of incongruence using new transcriptomic data. *Genome Biol. Evol.* **4**, 1295–1309 (2012).
- Wan, X., Kim, M. I., Kim, M. J. & Kim, I. Complete mitochondrial genome of the free-living earwig, *Challia fletcheri* (Dermaptera: Pygidicranidae) and phylogeny of Polyneoptera. *PLoS ONE* **7**, e42056 (2012).
- Engel, M. S. & Grimaldi, D. The first Mesozoic Zoraptera (Insecta). *Am. Mus. Novit.* **3362**, 1–20 (2002).
- Yoshizawa, K. & Johnson, K. P. Aligned 18S for Zoraptera (Insecta): Phylogenetic position and molecular evolution. *Mol. Phylogenet. Evol.* **37**, 572–580 (2005).
- Wang, Y. *et al.* A unique box in 28S rRNA is shared by the enigmatic insect order Zoraptera and Dictyoptera. *Plos One* **8**, 96–98 (2013).
- Terry, M. D. & Whiting, M. F. Mantophasmatodea and phylogeny of the lower neopteran insects. *Cladistics* **21**, 240–257 (2005).
- Cameron, S. L., Beckenbach, A., Downton, M. & Whiting, M. Evidence from mitochondrial genomics on interordinal relationships in insects. *Arthropod Syst. Phylo.* **64**, 27–34 (2006).
- Cameron, S. L. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu. Rev. Entomol.* **59**, 95–117 (2014).

33. Timmermans, M. J. *et al.* Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Res.* **38**, e197 (2010).
34. Gillett, C. P. *et al.* Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Mol. Biol. Evol.* **31**, 2223–2237 (2014).
35. Cramptonplatt, A. *et al.* Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a bornean rainforest sample. *Mol. Biol. Evol.* **32**, 2302–2316 (2015).
36. Patel, R. K. & Jain, M. NGSQC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* **7**, e30619 (2012).
37. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
38. Simon, C., Buckley, T. R., Frati, F., Stewart, J. B. & Beckenbach, A. T. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved Polymerase Chain Reaction primers for animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* **37**, 545–579 (2006).
39. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).
40. Bernt, M. *et al.* MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).
41. Bininda-Emonds, O. R. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics* **6**, 156 (2005).
42. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
43. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
44. Kuck, P. & Meusemann, K. FASconCAT: Convenient handling of data matrices. *Mol. Phylogenet. Evol.* **56**, 1115–1118 (2010).
45. Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. Testing significance of incongruence. *Cladistics* **10**, 315–319 (1994).
46. Swofford, D. L. PAUP: phylogenetic analysis using parsimony and other methods (Version 4 beta 10. Sinauer Associates, Sunderland, Massachusetts, 2002).
47. Xia, X., Xie, Z., Salemi, M., Chen, L. & Wang, Y. An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* **26**, 1–7 (2003).
48. Xia, X. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* **30**, 1720–1728 (2013).
49. Regier, J. C. *et al.* Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079–1083 (2010).
50. Zwick, A., Regier, J. C. & Zwickl, D. J. Resolving discrepancy between nucleotides and amino acids in deep-level arthropod phylogenomics: differentiating serine codons in 21-amino-acid models. *PLoS ONE* **7**, e47450 (2012).
51. Lanfear, R., Calcott, B., Ho, S. Y. & Guindon, S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701 (2012).
52. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
53. Miller, M. A., Pfeiffer, W. & Schwartz, T. “Creating the CIPRES Science Gateway for inference of large phylogenetic trees” in Proceedings of the Gateway Computing Environments Workshop (GCE) 14 Nov. 2010, New Orleans, LA 1–8 (2010).
54. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).
55. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
56. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
57. Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247 (2001).
58. Grimaldi, D. & Engel, M. S. *Evolution of the Insects* (Cambridge University Press, Cambridge), (2005).
59. Gissi, C., Reyes, A., Pesole, G. & Saccone, C. Lineage-specific evolutionary rate in mammalian mtDNA. *Mol. Biol. Evol.* **17**, 1022–1031 (2000).
60. Boore, J. L., Macey, J. R. & Medina, M. Sequencing and comparing whole mitochondrial genomes of animals. *Method. Enzymol.* **395**, 311–348 (2005).
61. Talavera, G. & Vila, R. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC Evol. Biol.* **11**, 315 (2011).
62. Simon, S. & Hadrys, H. A. comparative analysis of complete mitochondrial genomes among hexapoda. *Mol. Phylogenet. Evol.* **69**, 393–403 (2013).
63. Beutel, R. G. & Gorb, S. N. Ultrastructure of attachment specializations of hexapods (Arthropoda) evolutionary patterns inferred from a revised ordinal phylogeny. *J. Zool. Syst. Evol. Res* **39**, 177–207 (2001).
64. Kjer, K. M. Aligned 18S and insect phylogeny. *Syst. Biol.* **53**, 506–514 (2004).
65. Deitz, L. L., Nalepa, C. A. & Klass, K. D. Phylogeny of the Dictyoptera reexamined. *Ent. Abh.* **61**, 69–91 (2003).
66. Klass, K. D. & Meier, R. A phylogenetic analysis of Dictyoptera (Insecta) based on morphological characters. *Ent. Abh.* **63**, 3–50 (2006).
67. Hennig, W. Die Stammesgeschichte der Insekten. *Frankfurt am Main: Waldemar Kramer* **436** (1969).
68. Giles, E. T. The comparative external morphology and the affinities of the Dermaptera. *Trans. R. Ent. Soc. London* **115**, 95–164 (1963).
69. Kamp, J. W. Numerical classification of orthopteroids, with special reference to the Grylloblattodea. *Can. Ent.* **105**, 1235–1249 (1973).

Acknowledgements

This research is supported by grants from the National Natural Science Foundation of China (No. 31402002), and Key Scientific Research Projects of Henan Province (Grant Nos. 14B210036 and 16A210029).

Author Contributions

N.S. and W.C. designed and performed the research. N.S., H.L. and F.S. analyzed the data. All authors discussed results and implications. All authors have read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Song, N. *et al.* Molecular phylogeny of Polyneoptera (Insecta) inferred from expanded mitogenomic data. *Sci. Rep.* **6**, 36175; doi: 10.1038/srep36175 (2016).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016