Zoologica Scripta

Mitochondrial genome sequences effectively reveal deep branching events in aphids (Insecta: Hemiptera: Aphididae)

Jing Chen, Yuan Wang, Liyun Jiang & Gexia Qiao

Submitted: 6 October 2016
Accepted: 5 February 2017
doi:10.1111/zsc.12241

Chen, J., Wang, Y., Jiang, L. & Qiao, G. (2017). Mitochondrial genome sequences effectively reveal deep branching events in aphids (Insecta: Hemiptera: Aphididae). —*Zoologica Scripta*, *46*, 706–717.

Over the past dozen years, considerable effort has been devoted to unravelling the higher-level phylogenetic relationships of viviparous aphids (Aphididae). However, to date, there is still no commonly accepted phylogenetic hypothesis. In this study, we establish a robust phylogenetic framework for the Aphididae based on mitochondrial genome sequences of 35 aphid species, 22 of which are newly reported. Phylogenetic inferences are performed using multiple data sets, alternative partitioning schemes and different model-based methods. Our analyses result in well-supported backbone relationships for the major lineages of aphids, suggesting the feasibility of mitogenome data for resolving phylogenetic questions in aphids. Mindarinae is strongly supported as the earliest branching lineage within Aphididae. A monophyletic clade comprising Calaphidinae, Phyllaphidinae and Saltusaphidinae is corroborated to be the sister group to the species-richest subfamily Aphidinae. In addition, the morphologically defined subfamily Eriosomatinae is uncovered to be non-monophyletic.

Corresponding author: *Gexia Qiao*, *Key Laboratory of Zoological Systematics and Evolution*, *Institute of Zoology*, *Chinese Academy of Sciences*, *Beijing 100101*, *China. E-mail: qiaogx@ioz.ac.cn*
Jing Chen, *Key Laboratory of Zoological Systematics and Evolution*, *Institute of Zoology*, *Chinese Academy of Sciences*, *Beijing*, *China. E-mail: chenjing@ioz.ac.cn*
Yuan Wang, *Key Laboratory of Zoological Systematics and Evolution*, *Institute of Zoology*, *Chinese Academy of Sciences*, *Beijing*, *China and Shaanxi International Travel Healthcare Center*, *Shaanxi Entry-Exit Inspection and Quarantine Bureau*, *Xi'an*, *China. E-mail: wangyuan0330@163.com*
Liyun Jiang and Gexia Qiao, *Key Laboratory of Zoological Systematics and Evolution*, *Institute of Zoology*, *Chinese Academy of Sciences*, *Beijing*, *China. E-mails: jiangliyun@ioz.ac.cn*, *qiaogx@ioz.ac.cn*

Introduction

Aphids (Hemiptera: Aphidoidea) are an important group of phloem-feeding insects, many of which are serious pests for agriculture and forestry. This group contains approximately 5000 species within three families, Adelgidae, Phylloxeridae and Aphididae, mainly inhabiting the temperate regions of the Northern Hemisphere (Blackman & Eastop 2000; Favret 2016). Due to their fascinating biological characteristics, aphids are good model organisms for evolutionary and ecological studies. Many aphid species have complex life cycles involving sexual and asexual reproduction and the development of multiple alternative phenotypes (Dixon 1977; Blackman & Eastop 2000). Some species induce gall formation on their host plants (Wool 2004) and even produce specialized soldier castes (Stern & Foster 1996). In addition, aphids harbour different bacterial symbionts, including the obligate nutrient-provisioning symbiont, *Buchnera aphidicola*, and multiple facultative symbionts that provide ecological benefits for aphids (Buchner 1965; Oliver *et al.* 2010; Zytynska & Weisser 2016).

The vast majority of all aphid species (>95%) are included in the family Aphididae (24 subfamilies) (Remaudière & Remaudière 1997; Favret 2016). This group is also called true aphids and differs from its closest relatives Adelgidae and Phylloxeridae by the presence of siphunculi and involving viviparity in life cycles (Zhang & Zhong 1983; Heie 1987). Nevertheless, to date, a well-supported phylogenetic hypothesis is not available for the Aphididae. Earlier studies based on morphological and biological characters yielded conflicting hypotheses on the relationships among major lineages of Aphididae (Heie 1987; Wojciechowski 1992; Zhang *et al.* 1999b; Heie &

Wegierek 2009). Previous molecular phylogenetic studies using very few genes failed to reach an agreement as well. von Dohlen & Moran (2000) firstly reconstructed the Aphididae phylogeny with mitochondrial 12S and 16S rDNA, wherein the monophyly of most morphologically defined subfamilies was not recovered and almost no phylogenetic structure was revealed for deep nodes. They therefore argued that aphids potentially experienced a rapid radiation accompanying the shift from gymnosperms to angiosperms. Utilizing the nuclear LWO gene combined with other three genes, Ortiz-Rivas *et al.* (2004) and Ortiz-Rivas & Martínez-Torres (2010) uncovered three main lineages within the Aphididae, with the subfamily Lachninae being sister to all remaining aphidids. However, in their analyses, most relationships among subfamilies were weakly supported. In Ortiz-Rivas & Martínez-Torres (2010), the LWO phylogeny actually contradicted the results of other single genes. More recently, Nováková *et al.* (2013) built the Aphididae phylogeny using five genes derived from aphids' primary endosymbiont *B. aphidicola*. In contrast, the *Buchnera* trees were inconsistent with previous aphid phylogenetic hypotheses, and their supports for deep relationships were very low. Although aphids and *Buchnera* are cospeciating associated (Moran *et al.* 1993; Clark *et al.* 2000; Liu *et al.* 2013), whether *Buchnera* genes are appropriate to resolve the deep branching events in the evolution of its aphid hosts remains debatable (Liu *et al.* 2014).

Previous molecular phylogenetic analyses of the Aphididae were all conducted using limited data, leading to unstable and incongruent results with weak support or a low degree of deeper resolution. The difficulty of resolving the relationships among major Aphididae taxa may be due either to a real early rapid diversification as proposed by von Dohlen & Moran (2000) or a lack of sufficient phylogenetic signals. Thereby, to avoid methodological artefacts, an extended source of information is greatly needed.

Over the past decade, mitochondrial genomes have been widely utilized in insect phylogenetic studies, covering a broad range of taxonomic levels (Ma *et al.* 2012; Cameron 2014; Li *et al.* 2015; Song *et al.* 2015). Mitogenome sequences could provide a considerable amount of phylogenetically informative signals in insect systematics. In most cases, the mitochondrial phylogenomic studies rarely yield wildly conflicting results with other data sources (e.g. morphology and nuclear genes) and usually produce higher support for deep nodes (Cameron 2014). However, this type of data resource is quite limited for aphids. Up to now, only 13 complete or nearly complete mitogenomes of Aphididae species have been reported, mostly from the subfamily Aphidinae (Thao *et al.* 2004; Wang *et al.* 2013, 2014, 2015, 2016; Zhang *et al.* 2014, 2016a,b; Song *et al.* 2016; Li *et al.* 2017). Accordingly, no comprehensive phylogenetic analysis using mitogenome data has ever been performed on aphids.

In this study, we presented, for the first time, a robust phylogenetic hypothesis for the Aphididae based on a supermatrix comprising mitochondrial genome sequences of broadly sampled Aphididae taxa. To better explore such a largely expanded data source, we employed multiple data sets, different model-based approaches of phylogenetic inference coupled with alternative partitioning schemes and statistical tests for both model fitness and topology selection. Our resulting phylogenetic hypothesis will be helpful in better understanding the evolution and diversification of aphids.

Materials and methods

Taxon sampling

A total of 35 aphid species were included in this study. The aphid classification system followed Remaudière & Remaudière (1997) and Favret (2016). Thirty-three species belonging to 15 subfamilies of Aphididae were used as ingroups. Two species of Adelgidae and Phylloxeridae were employed as outgroups. Samples for slide mounting and molecular experiments were stored in 75% and 100% ethanol, respectively. Slide-mounted specimens were identified based on the external morphology by following the keys in authoritative monographs and literatures (e.g. Blackman & Eastop 1994, 2000) and by comparison with the original morphological descriptions. All voucher specimens and samples were deposited in the National Zoological Museum of China, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. Voucher information for all samples is listed in Table S1.

DNA extraction, amplification and sequencing

Total DNA was extracted from single aphids using DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. The mitogenomes were amplified by short and long PCRs using universal and specific primers. Detailed procedures generally followed Wang *et al.* (2013, 2014, 2015, 2016). Short PCR products were sequenced directly. All fragments from long PCRs were cloned using pMD19-T Vector System (TaKaRa, Dalian, China), and internal primers were designed to complete the sequencing by primer walking. Sequencing reactions were performed using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) on an ABI PRISM 3730 DNA Analyser (Applied Biosystems).

Sequences were assembled using SEQMAN II (DNAStar, Madison, WI, USA). Complete mitogenome sequences of many sampled species were not obtained due to experimental difficulties. Thus, we used 13 protein-coding genes (PCGs) and 12S, tRNA-Val and 16S rRNA genes (12S/

tRNA-Val/16S) in further analyses, which were equivalent to approximately 80% of the mitogenome. PCGs were identified using the Open Reading Frame Finder (ORFFINDER) at NCBI (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Two rRNAs were identified by sequence similarity with published aphid mitogenomes. The tRNA-Val was predicted by TRNASCAN-SE v.1.21 (Lowe & Eddy 1997). All sequences have been deposited in GenBank. Some published mitogenomes and 12S/tRNA-Val/16S sequences were obtained from GenBank (Table S1).

Sequence alignment and data set concatenation

Protein-coding genes were aligned individually with the TRANSLATORX online server (http://translatorx.co.uk/; Abascal *et al.* 2010), employing MAFFT to perform the protein alignment. The 12S rRNA, 16S rRNA and tRNA-Val genes were independently aligned using the secondary structure-based algorithm Q-INS-i implemented in MAFFT 7.266 (Katoh & Toh 2008; Katoh & Standley 2013), and ambiguously aligned positions were removed with GBLOCKS 0.91b (Castresana 2000; Talavera & Castresana 2007). Alignments of individual genes were concatenated by SEQUENCEMATRIX 1.8 (Meier *et al.* 2006) into five data sets: (i) PCG: 13 PCGs with 10 911 bp; (ii) PCGR: 13 PCGs and 12S/tRNA-Val/16S with 12 969 bp; (iii) PCG12: first and second codon positions of 13 PCGs with 7274 bp; (iv) PCG12R: first and second codon positions of 13 PCGs and 12S/tRNA-Val/16S with 9332 bp; and (v) AA: amino acid sequences of 13 PCGs with 3637 sites.

Phylogenetic analyses

Phylogenetic analyses were firstly performed using Bayesian inference (BI) and maximum-likelihood (ML) methods with homogeneous models. For unpartitioned data sets, the best-fit model of nucleotide substitution and protein evolution was selected using JMODELTEST 2.0.2 (Posada 2008) and PROTTEST 3.4 (Abascal *et al.* 2005) under Bayesian information criterion (BIC) (Schwarz 1978), respectively. Different partitioning schemes were also employed for PCG, PCGR and AA data sets. The optimal partitioning scheme and substitution models for each partition were assessed with PARTITIONFINDER v1.1.1 (Lanfear *et al.* 2012). The BIC and 'greedy' algorithm with branch lengths estimated as 'unlinked' were utilized. The best-fit models for unpartitioned data sets and the optimal partitioning schemes and models for partitioned data sets are listed in Table S2.

Bayesian inferences were performed in MRBAYES 3.2.6 (Ronquist *et al.* 2012). Four Markov chains (three heated and one cold) were run, sampling the chains every 500 generations for amino acid data set and every 1000 generations for nucleotide data set. Two concurrent runs were conducted to verify the results. Stationarity was assumed when the average standard deviation of split frequencies fell below 0.01. The first 25% trees were discarded as burn-in. The number of total Markov chain Monte Carlo (MCMC) generations and burn-in samples for each data set is listed in Table S3. ML analyses were inferred in RAXML v8.2.7 (Stamatakis 2006). A rapid bootstrapping algorithm was applied with 1000 replicates.

To account for the site-specific features of mitochondrial gene evolution, Bayesian analyses were also performed using PHYLOBAYES 4.1c (Lartillot & Philippe 2004) under the site-heterogeneous mixture models CAT and CAT-GTR for all five data sets. Two independent chains were run. Analyses were considered to have converged when the largest discrepancy across bipartitions (maxdiff) dropped below 0.15. The number of total cycles and burn-in samples for each data set is listed in Table S4.

Testing alternative tree topologies and the fitness of homogeneous and heterogeneous models

To evaluate the tree topologies resulting from phylogenetic inferences using different data sets and approaches, the approximately unbiased (AU) test (Shimodaira 2002) was performed. We first calculated the sitewise log likelihoods for each topology by TREE-PUZZLE 5.3 (Schmidt *et al.* 2002) and then submitted the output to CONSEL v.0.1j (Shimodaira & Hasegawa 2001). All competing trees were compared and sorted in the decreasing order of log likelihoods, and the *P*-values of AU test were calculated to assess the level of statistical support for alternative topologies. Additionally, we tested several specific hypotheses concerning three phylogenetic issues: (i) the earliest branching lineage within Aphididae, (ii) the sister group to Aphidinae and (iii) the monophyly of Eriosomatinae (see Table 1). To obtain the phylogenetic trees satisfying given hypotheses, Bayesian inferences were conducted in MRBAYES with the 'constraint' command (constraining monophyly on certain taxa), using the data set and partitioning scheme which yielded the best tree in the first round of tree selection.

The fitness of homogeneous and heterogeneous models was assessed for all five data sets using the cross-validation (CV) test implemented in PHYLOBAYES 4.1c. The GTR and MtREV model was used as the reference model for nucleotide and amino acid data set, respectively. We also compared the fitness of two heterogeneous models CAT and CAT-GTR, employing CAT as the reference model. Each data set was not partitioned and was randomly split into a learning set, including 90% of the original alignment, and a test set, comprising the remaining 10%. The test was run for 1100 cycles with a burn-in of 100. The average CV score over 10 replicates was calculated for each model.

© 2017 Royal Swedish Academy of Sciences, **46**, 6, November 2017, pp 706–717

Table 1 Statistical testing of particular phylogenetic hypotheses

Test	Rank	Hypothesis	Obs	AU *P*-value
I	1	An earliest branching of Mindarinae	Best	0.997
	2	An earliest branching of Lachninae (Ortiz-Rivas *et al.* 2004; Ortiz-Rivas & Martínez-Torres 2010)	525.7	0.003*
	3	An earliest branching of Lachninae + Thelaxinae	706.2	7.0×10^{-6}*
II	1	Aphidinae + (Calaphidinae + Phyllaphidinae + Saltusaphidinae) (Nováková *et al.* 2013)	Best	0.677
	2	Aphidinae + (Calaphidinae + Macropodaphidinae + Phyllaphidinae + Saltusaphidinae)	122.8	0.323
	3	Aphidinae + Lachninae (Mackauer 1965; Heie 1987; Wojciechowski 1992; Heie & Wegierek 2009)	7332.7	1.0×10^{-83}*
	4	Aphidinae + (Lachninae + Thelaxinae)	12275.6	1.0×10^{-42}*
III	1	Non-monophyly of Eriosomatinae	Best	0.934
	2	Monophyly of Eriosomatinae	85.6	0.066

Obs, observed log-likelihood difference to the best topology; AU, approximately unbiased test.

All constrained topologies satisfying particular hypotheses are illustrated in Fig. S5.

*Indicates that the hypothesis received a *P*-value < 0.01 and can be rejected.

Results

Phylogeny

Based on multiple data sets, different model-based phylogenetic inference methods and alternative partitioning schemes, a total of 30 trees were produced (Figs S1–S4). Support values for particular phylogenetic relationships in each tree are summarized in Table 2. All of these trees were compared using AU test (Table S5). The results indicated that the topology resulting from gene partitioned Bayesian analysis of the PCGR data set (Fig. 1) was more likely to present the true tree ($P = 0.996$). The other topologies, however, had *P*-values < 0.01.

In the best tree (BI-PCGR-gene partition, Fig. 1), the Aphididae was retrieved as monophyletic with strong support (posterior probability, PP = 1). Most subfamilies formed well-supported clades, whereas Eriosomatinae was polyphyletic. Within the clade of Aphididae, *Mindarus keteleerifoliae* split off earliest from other taxa (PP = 1). Chaitophorinae was then positioned as a sister group to the remaining Aphididae representatives, which were grouped into a single polytomous clade (PP = 1). Three main clades were recovered within the polytomous backbone. Clade 1 (PP < 0.9; given the overcredibility of Bayesian phylogenetics (Suzuki *et al.* 2002; Cummings *et al.* 2003; Simmons *et al.* 2004), PP above 0.9 is considered strong) comprised *Pemphigus immunis* (Eriosomatinae), Thelaxinae and the monophyletic Lachninae, with the latter two being clustered into a sister group with high support (PP = 1). Clade 2 (PP < 0.9) consisted of Aiceoninae, Anoeciinae, Macropodaphidinae, the monophyletic Hormaphidinae and Greenideinae, and *Eriosoma lanigerum* and *Kaburagia rhusicola* from Eriosomatinae, with inner relationships not highly supported. Within the well-supported Clade 3 (PP = 1), the monophyletic Aphidinae was placed as sister to a robust monophyletic clade, including Phyllaphidinae, Calaphidinae and Saltusaphidinae. Several phylogenetic analyses yielded entirely or largely congruent ingroup topologies with the best tree (i.e. ML-PCGR-gene partition, ML/BI-PCG-gene partition and BI-PCG-CAT/CAT-GTR; see Fig. 1 for support values from these analyses; Figs S1, S2).

In all resulting trees of our phylogenetic analyses, the monophyly of Aphididae was recovered with high support values. All sampled subfamilies represented by multiple taxa were retrieved as well-supported clades in most analyses, with the exception of Eriosomatinae, whose monophyly was only weakly supported by the ML and MrBayes analyses of unpartitioned PCGR (Table 2; Fig. 1, Figs S1–S4). Mindarinae was identified as the sister lineage to all remaining aphidids by all data sets except for PCG12R. Bayesian inferences of PCG, PCGR and AA based on both homogeneous and heterogeneous models provided strong support for this hypothesis (Table 2; Fig. 1, Figs S1, S2, S4). ML and MrBayes analyses of unpartitioned PCG12 and PCG12R, and ML analysis of unpartitioned AA, however, positioned *E. lanigerum* (Eriosomatinae) and Hormaphidinae as the earliest branching lineage within Aphididae, respectively (Figs S3, S4). But their support values were very low in most reconstructions, and these hypotheses were not accepted by statistical testing ($P < 0.001$, Table S5). All phylogenetic inferences strongly supported a sister group relationship between Lachninae and Thelaxinae (Table 2; Fig. 1, Figs S1–S4). In the trees obtained from all analyses except for the ML and MrBayes analyses of unpartitioned PCG12, Calaphidinae, Phyllaphidinae and Saltusaphidinae formed a robust monophyletic clade, which was placed as sister to Aphidinae in most cases (Table 2; Fig. 1, Figs S1–S4). Several inferences (i.e. ML/BI-PCG12-no partition, BI-PCG12-CAT/CAT-GTR, BI-PCG12R-CAT/CAT-GTR and BI-AA-no partition/gene partition) clustered Macropodaphidinae with Calaphidinae + Phyllaphidinae + Saltusaphidinae, but mostly with weak support (Table 2; Figs S3, S4).

Table 2 Sensitivity of particular phylogenetic hypotheses to different data sets and phylogenetic analyses

	PCG				PCGR			
Phylogenetic hypothesis	No partition ML/BI	Gene partition ML/BI	Codon partition ML/BI	BI CAT/CAT-GTR	No partition ML/BI	Gene partition ML/BI	Codon partition ML/BI	BI CAT/CAT-GTR
Aphididae monophyly	100/1	100/1	100/1	1/1	100/1	100/1	100/1	1/1
Aphidinae monophyly	100/1	100/1	100/1	1/1	100/1	100/1	100/1	1/1
Calaphidinae monophyly	100/1	100/1	100/1	1/1	100/1	100/1	100/1	1/1
Chaitophorinae monophyly	100/1	100/1	100/1	1/1	100/1	100/1	100/1	1/1
Eriosomatinae monophyly	NA/NA	NA/NA	NA/NA	NA/NA	+/+	NA/NA	NA/NA	NA/NA
Greenideinae monophyly	100/1	100/1	100/1	1/1	100/1	100/1	100/1	1/1
Hormaphidinae monophyly	+/1	+/1	70/1	+/0.99	+/1	+/1	+/1	+/0.9
Lachninae monophyly	100/1	100/1	100/1	0.98/0.99	100/1	100/1	100/1	1/1
Saltusaphidinae monophyly	100/1	100/1	100/1	1/1	100/1	100/1	100/1	1/1
The earliest branching of Mindarinae	+/0.94	71/1	+/1	0.99/0.96	NA/0.96	+/1	+/NA	0.95/+
Calaphidinae + Phyllaphidinae + Saltusaphidinae	87/1	94/1	96/1	1/1	97/1	99/1	100/1	1/1
Calaphidinae + Macropodaphidinae + Phyllaphidinae + Saltusaphidinae	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA
Aphidinae + (Calaphidinae + Phyllaphidinae + Saltusaphidinae)	+/1	+/0.98	+/+	+/+	+/1	+/1	+/+	NA/NA
Aphidinae + (Calaphidinae + Macropodaphidinae + Phyllaphidinae + Saltusaphidinae)	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA
Lachninae + Thelaxinae	100/1	100/1	100/1	1/1	100/1	100/1	100/1	1/1
Aphidinae + (Lachninae + Thelaxinae)	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/+

	PCG12		PCG12R		AA		
Phylogenetic hypothesis	No partition ML/BI	BI CAT/CAT-GTR	No partition ML/BI	BI CAT/CAT-GTR	No partition ML/BI	Gene partition ML/BI	BI CAT/CAT-GTR
Aphididae monophyly	100/1	1/1	100/1	1/1	100/1	100/1	1/1
Aphidinae monophyly	100/1	0.99/1	100/1	1/1	100/1	100/1	1/1
Calaphidinae monophyly	100/1	1/1	100/1	1/1	99/1	99/1	1/1
Chaitophorinae monophyly	100/1	1/1	100/1	1/1	100/1	100/1	1/1
Eriosomatinae monophyly	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA
Greenideinae monophyly	100/1	1/1	100/1	1/1	100/1	100/1	1/1
Hormaphidinae monophyly	+/1	0.98/0.96	+/1	0.99/+	+/NA	+/NA	+/+
Lachninae monophyly	100/1	NA/NA	100/1	0.97/0.99	100/1	98/1	0.98/0.99
Saltusaphidinae monophyly	100/1	1/1	100/1	1/1	100/1	100/1	1/1
The earliest branching of Mindarinae	NA/NA	+/NA	NA/NA	NA/NA	NA/1	+/0.98	0.99/0.96
Calaphidinae + Phyllaphidinae + Saltusaphidinae	NA/NA	0.99/0.99	97/1	1/1	88/1	88/1	0.99/1
Calaphidinae + Macropodaphidinae + Phyllaphidinae + Saltusaphidinae	+/+	+/0.93	NA/NA	+/+	NA/+	NA/0.99	NA/NA
Aphidinae + (Calaphidinae + Phyllaphidinae + Saltusaphidinae)	NA/NA	NA/NA	+/1	NA/NA	NA/NA	+/NA	NA/NA
Aphidinae + (Calaphidinae + Macropodaphidinae + Phyllaphidinae + Saltusaphidinae)	+/+	+/+	NA/NA	0.92/+	NA/+	NA/0.98	NA/NA
Lachninae + Thelaxinae	100/1	1/1	100/1	1/1	100/1	100/1	1/1
Aphidinae + (Lachninae + Thelaxinae)	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	NA/NA	0.99/NA

ML, maximum-likelihood; BI, Bayesian inference; NA, hypothesis not recovered.

Support values are shown for the recovered phylogenetic hypothesis. '+' is shown if ML bootstrap value is below 70% or BI posterior probability is below 0.9.

Model comparison

The results of CV test are listed in Table 3. The positive CV scores indicated that for all unpartitioned data sets, the heterogeneous models (CAT and CAT-GTR) were preferred over homogeneous models and that CAT-GTR was the best-fitting model.

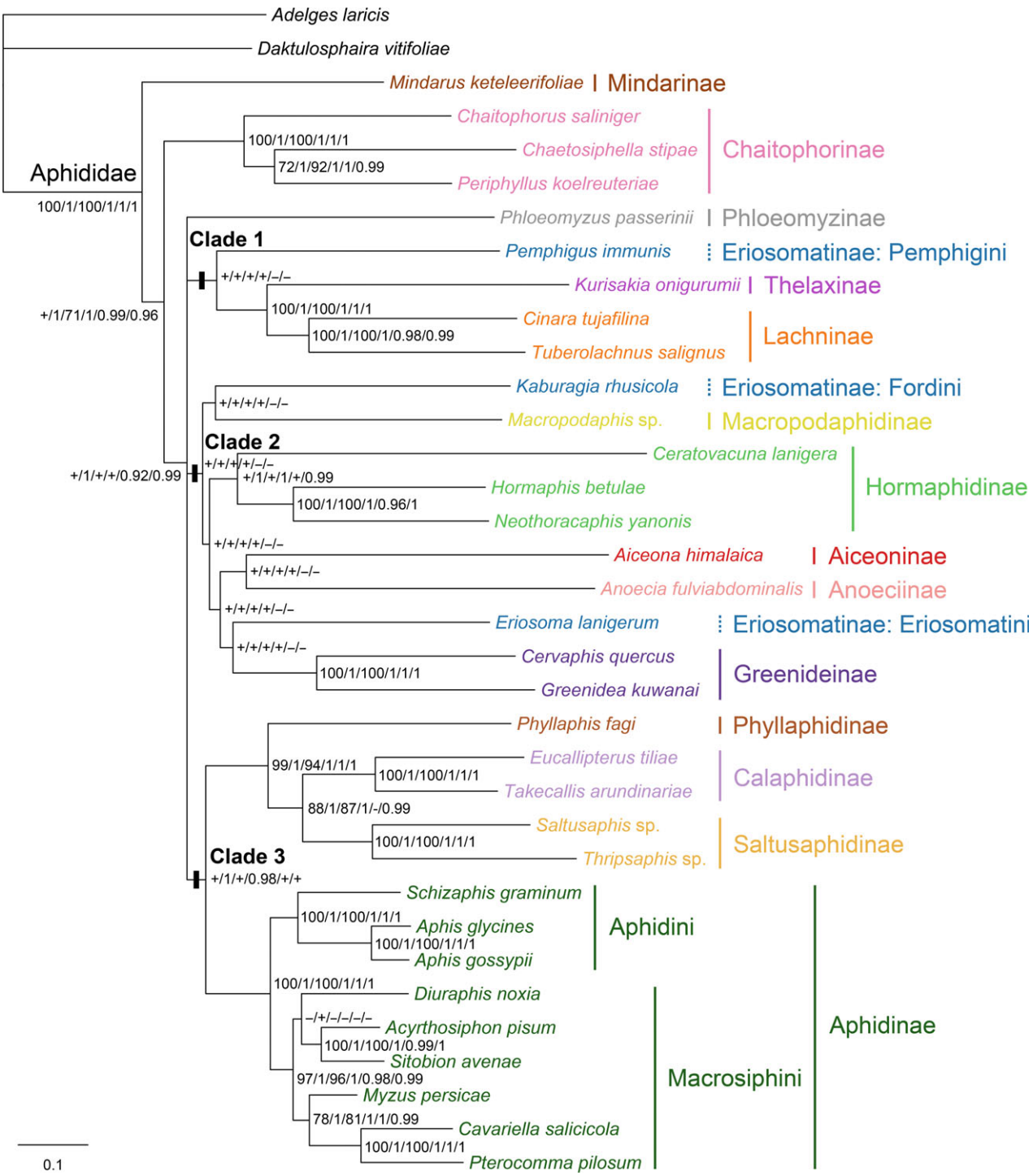

Fig. 1 Aphid tree resulting from the gene partitioned Bayesian inference of PCGR. Values at node indicate maximum-likelihood (ML) bootstraps (>70%) and Bayesian inference (BI) posterior probabilities (PP, >0.9) from the analyses of ML-PCGR-gene partition, BI - PCGR-gene partition, ML-PCG-gene partition, BI-PCG-gene partition, BI-PCG-CAT and BI-PCG-CAT-GTR. '+' is shown if the clade is recovered, but ML bootstrap value is below 70% or Bayesian PP is below 0.9. '–' is shown if the clade is not recovered. Dashed vertical lines correspond to non-monophyletic subfamilies.

Table 3 Cross-validation (CV) test of the homogeneous and heterogeneous models

Data set	Model 1	Model 2	CV score	Standard deviation
PCG	CAT	GTR	549.14	39.5241
	CAT-GTR	GTR	607.37	36.9872
	CAT-GTR	CAT	58.23	12.4296
PCGR	CAT	GTR	596.99	34.9945
	CAT-GTR	GTR	663.77	36.0922
	CAT-GTR	CAT	66.78	6.8914
PCG12	CAT	GTR	206.33	26.0545
	CAT-GTR	GTR	230.63	23.8939
	CAT-GTR	CAT	24.30	4.5695
PCG12R	CAT	GTR	277.01	47.9735
	CAT-GTR	GTR	298.26	52.0729
	CAT-GTR	CAT	21.25	10.3667
AA	CAT	MtREV	226.79	44.7549
	CAT-GTR	MtREV	276.01	45.6384
	CAT-GTR	CAT	49.22	19.1681

Model 2 is the reference model. A positive CV score indicates better performance than reference model.

Discussion

The feasibility of mitogenome sequences for resolving aphid phylogeny

Although widely used in insect phylogenetic studies, mitogenome data suffer from strong nucleotide compositional bias and substitution rate heterogeneity, thus tending to produce artefactual relationships under inappropriate methods (Yang 1996; Dowton *et al.* 2009; Sheffield *et al.* 2009; Song *et al.* 2010; Cameron 2014). Several solutions have been developed to avoid systematic errors caused by these issues, such as removing third codon positions, analysing translated amino acids, employing adequate data partitioning schemes and applying more sophisticated models (e.g. heterogeneous models) (Fenn *et al.* 2008; Sheffield *et al.* 2009; Cameron 2014; Li *et al.* 2015).

In this study, we performed phylogenetic analyses based on aphid mitogenome sequences using different schemes, trying to better explore the potency of this data. Removal of the third codons yielded odd topologies where *E. lanigerum* (Eriosomatinae) was placed as the earliest branching taxa (ML/BI-PCG12-no partition, ML/BI-PCG12R-no partition, Fig. S3; $P < 0.001$, Table S5), suggesting loss of important phylogenetic signals. Analyses of translated amino acid data set resulted in consistent ingroup topologies with the best tree with respect to major clades (exc. ML-AA-no partition, Fig. S4), but Hormaphidinae and Macrosiphini (Aphidinae) were not always retrieved as monophyletic, and sometimes a decrease in resolution was presented, implying that phylogenetic signals were weakened when nucleotide sequences were translated into corresponding amino acid sequences. Gene partitioned analyses performed well on our mitogenome data. Inferences on PCGR and PCG using gene partitioning schemes produced the best tree (BI-PCGR-gene partition, Fig. 1) and entirely or largely congruent ingroup topologies with the best tree (ML-PCGR-gene partition and ML/BI-PCG-gene partition, Figs S1, S2). Codon partitioned Bayesian analyses on PCGR and PCG, however, resulted in quite different topologies from the best tree (Figs S1, S2), suggesting overpartitioning appear to have a negative effect in accurate phylogenetic reconstruction. CV test results suggested that heterogeneous models showed better performance than homogeneous models for all unpartitioned data sets (Table 3). Bayesian analyses of PCG12 and PCG12R under CAT and CAT-GTR models revealed early branching of Mindarinae and Chaitophorinae rather than *E. lanigerum* (Fig. S3), which also indicated that the employment of heterogeneous models could compensate for the effects of exclusion of third codons to a certain degree.

Utilizing mitochondrial genome sequences, we presented the largest molecular tree ever built for the Aphididae. Our results provided a robust phylogenetic backbone for major lineages of viviparous aphids, in which the deep relationships were highly supported. Mindarinae branched off earliest followed by Chaitophorinae, and all remaining Aphididae representatives formed a well-supported monophyletic clade, which was split into multiple lineages. Several relationships were also well established, such as a sister group relationship between Lachninae and Thelaxinae; a monophyletic clade comprising Calaphidinae, Phyllaphidinae and Saltusaphidinae; and a sister group relationship between this robust monophyletic clade and Aphidinae. The results obtained in our study demonstrated that by applying appropriate methods mitogenome data were quite useful for resolving aphid phylogenetic questions, especially for unravelling the deep branching events.

The earliest branching lineage within Aphididae

In most resulting trees of our analyses, Mindarinae was firmly placed as the earliest branching clade within Aphididae (Table 2; Fig. 1, Figs S1–S4). Its placement was also corroborated by the topology test ($P = 0.997$, Table 1). Whereas in previous molecular phylogenetic studies, Mindarinae was either not included (Ortiz-Rivas *et al.* 2004) or obtained uncertain or unstable systematic positions (von Dohlen & Moran 2000; Ortiz-Rivas & Martínez-Torres 2010; Nováková *et al.* 2013). The subfamily Mindarinae is a small relic group represented by only one fossil genus *Mindarella* Heie dating back to late Oligocene (Heie 1989) and a single extant genus *Mindarus* Koch. *Mindarus* includes nine extant species feeding on Pinaceae (*Abies*, *Keteleeria* and *Picea*) plants (Blackman & Eastop 1994) and

eight fossil species recovered from the Eocene to Oligocene (Heie 2008; Heie & Wegierek 2011; Favret 2016). Besides, Mindarinae is also regarded as one of the most ancient aphid lineages by some taxonomists (Heie 1967, 1987; Hille Ris Lambers 1967; Heie & Pike 1992; Zhang & Qiao 1997; Quednau 2010).

Based on unrooted analyses of a limited number of genes, Ortiz-Rivas *et al.* (2004) and Ortiz-Rivas & Martínez-Torres (2010) recovered three main lineages within Aphididae. Rooted and topologically constrained analyses that kept the three monophyletic clades placed Lachninae as the earliest branching group. However, their results from such phylogenetic methods seem to be not sufficiently convincing, and the position of Lachninae was not highly supported, especially in Ortiz-Rivas *et al.* (2004). In current study based on the greatly expanded mitogenome data and broadly sampled taxa, none of our analyses uncovered Lachninae or Lachninae + Thelaxinae (these two subfamilies were robustly clustered together in all analyses) as the sister to all remaining aphidids, and statistical test rejected these hypotheses with *P*-values < 0.01 (Table 1). Contrary to the conclusions of Ortiz-Rivas *et al.* (2004) and Ortiz-Rivas & Martínez-Torres (2010), some authors agreed on the modern origin of Lachninae (Mackauer 1965; Heie 1987; Wojciechowski 1992; Normark 2000). In addition, considering its high diversity of living species (ca. 400 species), the fossil records of lachine aphids are strikingly rare and very young (Miocene) (Heie & Wegierek 2011).

The sister group to Aphidinae

The subfamily Aphidinae is the species richest and most successful lineage in Aphididae, including greater than half of all described aphid species and occurring throughout the temperate regions of the Northern Hemisphere and subtropical regions (Favret 2016). Much of its species diversity has likely derived from the explosive radiation in the Neogene that coincides with the domination of herbaceous angiosperms (Heie 1990, 1994, 1996). Some authors considered Lachninae to be the sister to Aphidinae (Mackauer 1965; Heie 1987; Wojciechowski 1992; Heie & Wegierek 2009). However, this hypothesis has never been supported by molecular phylogenetic studies (Ortiz-Rivas *et al.* 2004; Ortiz-Rivas & Martínez-Torres 2010; Nováková *et al.* 2013). In the present analyses of mitogenome data, only two inferences clustered Aphidinae with Lachninae + Thelaxinae (BI-PCGR-CAT-GTR, BI-AA-CAT; Table 2; Figs S2, S4). However, this topology as well as the traditional hypothesis of close affinity between Aphidinae and Lachninae was confidently rejected by statistical test with *P*-values <0.001 (Table 1).

In our study, the sister group relationship between Aphidinae and Calaphidinae + Phyllaphidinae + Saltusaphidinae was recovered in 17 topologies (of 30 topologies), received strong support in some analyses (BI-PCG-no partition/gene partition, BI-PCGR-no partition/gene partition and BI-PCG12R-no partition, Table 2), and was further favoured by the AU test ($P = 0.677$, Table 1). Such sister relationship was also revealed by *Buchnera* genes with high statistical support (Nováková *et al.* 2013). Calaphidinae, Phyllaphidinae and Saltusaphidinae were classified as members of one group by some taxonomists (Börner & Heinze 1957; Quednau 1999; Qiao *et al.* 2005; Heie & Wegierek 2009). These three subfamilies formed a robust monophyletic clade in our analyses. Phyllaphidinae was formerly lumped in Calaphidini (Calaphidinae) and was considered closely related to Saltusaphidinae for sharing the absence of triommatidia in eyes of apterae (Quednau 2010). In addition, both the Saltusaphidinae and Panaphidini (Calaphidinae) aphids have double-filter chambers in midguts (Ponsen 1990; Quednau 2010). Several features are shared by the Aphidinae and species from these three subfamilies: rostrum 4-segmented, eyes with multifacets in apterae and the 1st-instar nymphs, dorsal processes usually present on body, and having accessory glands and filter chambers.

The monophyly of Eriosomatinae

Eriosomatinae is an interesting aphid group characterized by several fascinating morphological and biological characteristics, such as possessing well-developed wax gland plates secreting wax powder or threads (Zhang *et al.* 1999a), obligate alternation between distantly related primary and secondary host plants (Moran 1988), inducing diverse galls on their primary hosts (Wool 2004; Zhang *et al.* 2006) and producing specialized aphid soldiers (Aoki 1977; Stern & Foster 1996). Eriosomatinae includes three tribes strictly associated with different primary hosts: Eriosomatini on *Ulmus* (Ulmaceae), Fordini on *Pistacia* and *Rhus* (Anacardiaceae) (Fordina on *Pistacia*, Melaphidina on *Rhus*) and Pemphigini on *Populus* (Salicaceae).

In the present study, all inferences failed to retrieve a monophyletic Eriosomatinae except for the ML and MrBayes analyses of unpartitioned PCGR, where its monophyly was merely weakly supported (Table 2). Compared with the best tree, although the constrained topology containing monophyletic Eriosomatinae could not be confidently rejected by statistical test, the probability of this hypothesis was extremely low ($P = 0.066$, Table 1). In addition, in all higher-level phylogenetic studies of Aphididae that have been conducted, neither aphid data (von Dohlen & Moran 2000; Ortiz-Rivas *et al.* 2004; Ortiz-Rivas & Martínez-Torres 2010) nor *Buchnera* data (Nováková *et al.* 2013) supported the Eriosomatinae as a

monophyletic clade. Its monophyly was also not corroborated in the morphological cladistic study (Zhang & Chen 1999) and molecular phylogenetic analyses of Eriosomatinae (Zhang & Qiao 2008; Li *et al.* 2014). The subfamily Eriosomatinae has been traditionally recognized as a monophylum based on the synapomorphy of sexuales apterous, dwarfish and lacking rostrum (Heie 1980; Zhang *et al.* 1999a). However, the failure to recover its monophyly in all phylogenetic studies suggests that the dwarfish and non-feeding sexuales do not seem to be apomorphies inherited from a common ancestor (synapomorphies) but result from convergent adaptation to the heteroecious life cycles.

Conclusion

Utilizing mitogenome sequences, our study shed new light on the higher-level phylogeny of Aphididae, with improved support for deep relationships. However, additional data from a broader range of taxa are clearly needed to produce a more detailed phylogeny and to test current hypothesis. Including species of another important conifer-feeding group Neophyllaphidinae and species from more subfamilies within Drepanosiphidae *sensu* Heie & Wegierek (2009) (e.g. Drepanosiphinae, Lizeriinae and Taiwanaphidinae) will be helpful for obtaining a clearer phylogenetic framework and consequently better understanding the evolutionary history of aphids. Finally, although mitogenome sequences provided significant phylogenetic signals for resolving aphid phylogeny, integrating mitochondrial and nuclear genomic data is absolutely necessary in the future to construct a robust phylogeny of Aphididae.

Acknowledgements

We thank all collectors for their assistance in sample collection. Thanks to Fendi Yang for making slides of voucher specimens. The work was supported by the National Natural Sciences Foundation of China (Nos. 31430078, 31402000, 31561163002), the External Cooperation Program of Chinese Academy of Sciences (No. 152111KYSB20130012) and a grant from the Ministry of Science and Technology of the People's Republic of China (No. 2011FY120200).

References

Abascal, F., Zardoya, R. & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, *21*, 2104–2105.

Abascal, F., Zardoya, R. & Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, *2010*, 1–7.

Aoki, S. (1977). *Colophina clematis* (Homoptera, Pemphigidae), an aphid species with "soldiers". *Kontyû*, *45*, 276–282.

Blackman, R. L. & Eastop, V. F. (1994). *Aphids on the World's Trees: An Identification and Information Guide*. Wallingford: CAB International.

Blackman, R. L. & Eastop, V. F. (2000). *Aphids on the Word's Crops: An Identification and Information Guide*, 2nd edn. Chichester: John Wiley & Sons.

Börner, C. & Heinze, K. (1957). Aphidina-Aphidoidea. In P. Sorauer (Ed.) *Handbuch der Pflanzenkrankheiten, 2. Teil, 4. Lief. Homoptera II* (pp. 1–402). Berlin: Paul Parey.

Buchner, P. (1965). *Endosymbiosis of Animals with Plant Microorganisms*. New York: Interscience.

Cameron, S. L. (2014). Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual Review of Entomology*, *59*, 95–117.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*, 540–552.

Clark, M. A., Moran, N. A., Baumann, P. & Wernegreen, J. J. (2000). Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution*, *54*, 517–525.

Cummings, M. P., Handley, S. A., Myers, D. S., Reed, D. L., Rokas, A. & Winka, K. (2003). Comparing bootstrap and posterior probability values in the four-taxon case. *Systematic Biology*, *52*, 477–487.

Dixon, A. F. G. (1977). Aphid ecology: life cycles, polymorphism, and population regulation. *Annual Review of Ecology and Systematics*, *8*, 329–353.

von Dohlen, C. D. & Moran, N. A. (2000). Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biological Journal of the Linnean Society*, *71*, 689–717.

Dowton, M., Cameron, S. L., Austin, A. D. & Whiting, M. F. (2009). Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera – a lineage with both rapidly and slowly evolving mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *52*, 512–519.

Favret, C. (2016). Aphid species file. Version 5.0/5.0. Available via http://Aphid.SpeciesFile.org.

Fenn, J. D., Song, H., Cameron, S. L. & Whiting, M. F. (2008). A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Molecular Phylogenetics and Evolution*, *49*, 59–68.

Heie, O. E. (1967). *Studies on Fossil Aphids (Homoptera: Aphidoidea) Especially in the Copenhagen Collection of Fossils in Baltic Amber*. Copenhagen: Spolia Zoologica Musei Hauniensis.

Heie, O. E. (1980). *The Aphidoidea (Hemiptera) of Fennoscandia and Denmark. I. General Part. The Families Mindaridae, Hormaphididae, Thelaxidae, Anoeciidae, and Pemphigidae. Fauna Entomologica Scandinavica 9*. Klampenborg: Scandinavian Science Press.

Heie, O. E. (1987). Palaeontology and phylogeny. In A. K. Minks & P. Harrewijn (Eds) *Aphids: Their Biology, Natural Enemies and Control*, Vol. 2A (pp. 367–391). Amsterdam: Elsevier.

Heie, O. E. (1989). Fossil aphids (Insecta, Homoptera) from the Tertiary deposits of Bolshaya Svetlovodnaya, the U.S.S.R. *Entomologica Scandinavica*, *19*, 475–488.

Heie, O. E. (1990). Recent advances in palaeoaphidology. *Acta Phytopathologica et Entomologica Hungarica*, *25*, 253–260.

Heie, O. E. (1994). Aphid ecology in the past and a new view on the evolution of Macrosiphini. In S. R. Leather, A. D. Watt, N. J. Mills & K. F. A. Walter (Eds) *Individuals, Populations and Patterns in Ecology* (pp. 409–418). Andover: Intercept.

Heie, O. E. (1996). The evolutionary history of aphids and a hypothesis on the coevolution of aphids and plants. *Bollettino di Zoologia Agraria e di Bachicoltura, Series II*, *28*, 149–155.

Heie, O. E. (2008). Insects in Baltic amber, and a new species of aphid. *Antenna*, *32*, 99–101.

Heie, O. E. & Pike, E. M. (1992). New aphids in Cretaceous amber from Alberta (Insecta, Homoptera). *The Canadian Entomologist*, *124*, 1027–1053.

Heie, O. E. & Wegierek, P. (2009). A classification of the Aphidomorpha (Hemiptera: Sternorrhyncha) under consideration of the fossil taxa. *Redia*, *92*, 69–77.

Heie, O. E. & Wegierek, P. (2011). A list of fossil aphids (Hemiptera, Sternorrhyncha, Aphidomorpha). *Monographs of the Upper Silesian Museum*, *6*, 1–82.

Hille Ris Lambers, D. (1967). On the genus *Neophyllaphis* Takahashi, 1920 (Homoptera, Aphididae) with descriptions of two new species. *Zoologische Mededelingen*, *42*, 55–66.

Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780.

Katoh, K. & Toh, H. (2008). Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics*, *9*, 212.

Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*, 1695–1701.

Lartillot, N. & Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, *21*, 1095–1109.

Li, X. Y., Jiang, L. Y. & Qiao, G. X. (2014). Is the subfamily Eriosomatinae (Hemiptera: Aphididae) monophyletic? *Turkish Journal of Zoology*, *38*, 285–297.

Li, H., Shao, R., Song, N., Song, F., Jiang, P., Li, Z. & Cai, W. (2015). Higher-level phylogeny of paraneopteran insects inferred from mitochondrial genome sequences. *Scientific Reports*, *5*, 8527.

Li, Y. Q., Chen, J. & Qiao, G. X. (2017). Complete mitochondrial genome of the aphid *Hormaphis betulae* (Mordvilko) (Hemiptera: Aphididae: Hormaphidinae). *Mitochondrial DNA Part A*, *28*, 265–266.

Liu, L., Huang, X., Zhang, R., Jiang, L. & Qiao, G. (2013). Phylogenetic congruence between *Mollitrichosiphum* (Aphididae: Greenideinae) and *Buchnera* indicates insect–bacteria parallel evolution. *Systematic Entomology*, *38*, 81–92.

Liu, L., Huang, X., Wang, Y. & Qiao, G. (2014). Revisiting the first case of insect-bacteria cospeciation: phylogenetic incongruence between aphids and their obligate endosymbiont at subfamily level. *PeerJ PrePrints*, *2*, e346v1.

Lowe, T. M. & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*, 955–964.

Ma, C., Yang, P., Jiang, F., Chapuis, M. P., Shali, Y., Sword, G. A. & Kang, L. (2012). Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. *Molecular Ecology*, *21*, 4344–4358.

Mackauer, M. (1965). Parasitological data as an aid in aphid classification. *The Canadian Entomologist*, *97*, 1016–1024.

Meier, R., Kwong, S., Vaidya, G. & Ng, P. K. L. (2006). DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, *55*, 715–728.

Moran, N. A. (1988). The evolution of host-plant alternation in aphids: evidence for specialization as a dead end. *The American Naturalist*, *132*, 681–706.

Moran, N. A., Munson, M. A., Baumann, P. & Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proceedings of the Royal Society of London B: Biological Sciences*, *253*, 167–171.

Normark, B. B. (2000). Molecular systematics and evolution of the aphid family Lachnidae. *Molecular Phylogenetics and Evolution*, *14*, 131–140.

Nováková, E., Hypša, V., Klein, J., Foottit, R. G., von Dohlen, C. D. & Moran, N. A. (2013). Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Molecular Phylogenetics and Evolution*, *68*, 42–54.

Oliver, K. M., Degnan, P. H., Burke, G. R. & Moran, N. A. (2010). Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, *55*, 247–266.

Ortiz-Rivas, B. & Martínez-Torres, D. (2010). Combination of molecular data support the existence of three main lineages in the phylogeny of aphids (Hemiptera: Aphididae) and the basal position of the subfamily Lachninae. *Molecular Phylogenetics and Evolution*, *55*, 305–317.

Ortiz-Rivas, B., Moya, A. & Martínez-Torres, D. (2004). Molecular systematics of aphids (Homoptera: Aphididae): new insights from the long-wavelength opsin gene. *Molecular Phylogenetics and Evolution*, *30*, 24–37.

Ponsen, M. B. (1990). *Phylogenetic implications of the Structure of the Alimentary Tract of the Aphidoidea. I. Greenidea, Israelaphis and Neophyllaphis. II. The Aphis-group*. Wageningen: Wageningen Agricultural University.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, *25*, 1253–1256.

Qiao, G., Zhang, G. & Zhong, T. (2005). *Fauna Sinica. Insecta Vol. 41, Homoptera: Drepanosiphidae*. Beijing: Science Press.

Quednau, F. W. (1999). Atlas of the drepanosiphine aphids of the world. Part I: Panaphidini Oestlund, 1922 - Myzocallidina Börner, 1942 (1930) (Hemiptera: Aphididae: Calaphidinae). *Contributions of the American Entomological Institute*, *31*, 1–281.

Quednau, F. W. (2010). Atlas of the drepanosiphine aphids of the world. Part III: Mindarinae Tullgren, 1909 to Saltusaphidinae Baker, 1920 (Hemiptera: Sternorrhyncha, Aphididae). *Memoirs of the American Entomological Institute*, *83*, 1–361.

Remaudière, G. & Remaudière, M. (1997). *Catalogue des Aphididae du Monde. Homoptera Aphidoidea*. Paris: Institut National de la Recherche Agronomique.

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. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*, 539–542.

Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, *18*, 502–504.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Sheffield, N. C., Song, H., Cameron, S. L. & Whiting, M. F. (2009). Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. *Systematic Biology*, *58*, 381–394.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, *51*, 492–508.

Shimodaira, H. & Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, *17*, 1246–1247.

Simmons, M. P., Pickett, K. M. & Miya, M. (2004). How meaningful are Bayesian support values? *Molecular Biology and Evolution*, *21*, 188–199.

Song, H., Sheffield, N. C., Cameron, S. L., Miller, K. B. & Whiting, M. F. (2010). When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. *Systematic Entomology*, *35*, 429–448.

Song, H., Amédégnato, C., Cigliano, M. M., Desutter-Grandcolas, L., Heads, S. W., Huang, Y., Otte, D. & Whiting, M. F. (2015). 300 million years of diversification: elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling. *Cladistics*, *31*, 621–651.

Song, N., Zhang, H., Li, H. & Cai, W. (2016). All 37 mitochondrial genes of aphid *Aphis craccivora* obtained from transcriptome sequencing: implications for the evolution of aphids. *PLoS ONE*, *11*, e0157857.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*, 2688–2690.

Stern, D. L. & Foster, W. A. (1996). The evolution of soldiers in aphids. *Biological Reviews*, *71*, 27–79.

Suzuki, Y., Glazko, G. V. & Nei, M. (2002). Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences*, *99*, 16138–16143.

Talavera, G. & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*, 564–577.

Thao, M. L., Baumann, L. & Baumann, P. (2004). Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evolutionary Biology*, *4*, 25.

Wang, Y., Huang, X. L. & Qiao, G. X. (2013). Comparative analysis of mitochondrial genomes of five aphid species (Hemiptera: Aphididae) and phylogenetic implications. *PLoS ONE*, *8*, e77511.

Wang, Y., Huang, X. L. & Qiao, G. X. (2014). The complete mitochondrial genome of *Cervaphis quercus* (Insecta: Hemiptera: Aphididae: Greenideinae). *Insect Science*, *21*, 278–290.

Wang, Y., Chen, J., Jiang, L. Y. & Qiao, G. X. (2015). The complete mitochondrial genome of *Mindarus keteleerifoliae* (Insecta: Hemiptera: Aphididae) and comparison with other Aphididae insects. *International Journal of Molecular Sciences*, *16*, 30091–30102.

Wang, Y., Jiang, L., Liu, Y., Chen, J. & Qiao, G. (2016). General methods to obtain and analyze the complete mitochondrial genome of aphid species: *Eriosoma lanigerum* (Hemiptera: Aphididae) as an example. *Zoological Systematics*, *41*, 123–132.

Wojciechowski, W. (1992). *Studies on the Systematic System of Aphids (Homoptera, Aphidinea)*. Katowice: Uniwersytet Slaski.

Wool, D. (2004). Galling aphids: specialization, biological complexity, and variation. *Annual Review of Entomology*, *49*, 175–192.

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology & Evolution*, *11*, 367–372.

Zhang, G. & Chen, X. (1999). Study on the phylogeny of Pemphigidae (Homoptera: Aphidinea). *Acta Entomologica Sinica*, *42*, 176–183.

Zhang, G. & Qiao, G. (1997). A study of *Mindarus* Koch in China with descriptions of new species and new subspecies (Homoptera: Mindaridae). *Entomologia Sinica*, *4*, 189–201.

Zhang, H. C. & Qiao, G. X. (2008). Molecular phylogeny of Pemphiginae (Hemiptera: Aphididae) inferred from nuclear gene EF-1α sequences. *Bulletin of Entomological Research*, *98*, 499–507.

Zhang, G. X. & Zhong, T. S. (1983). *Economic Insect Fauna of China. Fasc. 25. Homoptera: Aphidinea, Part I*. Beijing: Science Press.

Zhang, G., Qiao, G., Zhong, T. & Zhang, W. (1999a). *Fauna Sinica, Insecta Vol. 14, Homoptera: Mindaridae and Pemphigidae*. Beijing: Science Press.

Zhang, G. X., Chen, X. L., Qiao, G. X., Zhong, T. S. & Li, J. H. (1999b). *Fauna of Agricultural and Forestry Aphids of Northwest China (Insecta: Homoptera: Aphidinea)*. Beijing: China Environmental Science Press.

Zhang, H. C., Qiao, G. X. & Zhang, G. X. (2006). A study on diversity of aphid's galls in Pemphigidae. *Acta Zootaxonomica Sinica*, *31*, 48–54.

Zhang, B., Ma, C., Edwards, O., Fuller, S. & Kang, L. (2014). The mitochondrial genome of the Russian wheat aphid *Diuraphis noxia*: large repetitive sequences between *trnE* and *trnF* in aphids. *Gene*, *533*, 253–260.

Zhang, B., Zheng, J., Liang, L., Fuller, S. & Ma, C. S. (2016a). The complete mitochondrial genome of *Sitobion avenae* (Hemiptera: Aphididae). *Mitochondrial DNA Part A*, *27*, 945–946.

Zhang, S., Luo, J., Wang, C., Lv, L., Li, C., Jiang, W., Cui, J. & Rajput, L. B. (2016b). Complete mitochondrial genome of *Aphis gossypii* Glover (Hemiptera: Aphididae). *Mitochondrial DNA Part A*, *27*, 854–855.

Zytynska, S. E. & Weisser, W. W. (2016). The natural occurrence of secondary bacterial symbionts in aphids. *Ecological Entomology*, *41*, 13–26.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Aphid trees resulting from the phylogenetic analyses of PCG. Values at node indicate ML bootstraps (>70%) and Bayesian posterior probabilities (>0.9).

Figure S2. Aphid trees resulting from the phylogenetic analyses of PCGR. Values at node indicate ML bootstraps (>70%) and Bayesian posterior probabilities (>0.9).

Figure S3. Aphid trees resulting from the phylogenetic analyses of PCG12 and PCG12R. Values at node indicate

ML bootstraps (>70%) and Bayesian posterior probabilities (>0.9).

Figure S4. Aphid trees resulting from the phylogenetic analyses of AA. Values at node indicate ML bootstraps (>70%) and Bayesian posterior probabilities (>0.9).

Figure S5. Constrained aphid trees obtained from the gene partitioned Bayesian inference of PCGR. Values at node indicate Bayesian posterior probabilities (>0.9). (A) Topology constrained to place Mindarinae as the earliest branching lineage within Aphididae; (B) topology constrained to place Lachninae as the earliest branching lineage within Aphididae; (C) topology constrained to place Lachninae and Thelaxinae as the earliest branching lineages within Aphididae; (D) topology constrained to keep a sister group relationship between Aphidinae and Calaphidinae + Phyllaphidinae + Saltusaphidinae; (E) topology constrained to keep a sister group relationship between Aphidinae and Calaphidinae + Macropodaphidinae + Phyllaphidinae + Saltusaphidinae; (F) topology constrained to keep a sister group relationship between Aphidinae and Lachninae; (G) topology constrained to keep a sister group relationship between Aphidinae and Lachninae + Thelaxinae; (H) topology constrained to keep the monophyly of Eriosomatinae.

Table S1. Voucher information and GenBank accession numbers of aphid species used in this study.

Table S2. Best partitioning schemes and models for partitioned and unpartitioned data sets.

Table S3. Summary of individual MrBayes runs.

Table S4. Summary of individual PhyloBayes runs.

Table S5. Statistical testing of tree topology.