International Journal of Biological Macromolecules 122 (2019) 824–832

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: http://www.elsevier.com/locate/ijbiomac

The mitochondrial genome of *Greenidea psidii* van der Goot (Hemiptera: Aphididae: Greenideinae) and comparisons with other Aphididae aphids

Jing Chen [a], Yuan Wang [b], Man Qin [a,c], Li-Yun Jiang [a], Ge-Xia Qiao [a,c,*]

[a] Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[b] Shaanxi International Travel Healthcare Center, Shaanxi Entry-Exit Inspection and Quarantine Bureau, Xi'an 710068, China
[c] College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 25 September 2018
Received in revised form 29 October 2018
Accepted 29 October 2018
Available online 30 October 2018

Keywords:
Comparative analysis
Mitogenome
Phylogeny

ABSTRACT

The complete mitochondrial genome of *Greenidea psidii* was sequenced and compared with the genomes of other aphid species. The *G. psidii* mitogenome is 16,202 bp long with an A + T content of 85.4%, comprising 37 genes arranged in the same order as the inferred insect ancestral arrangement, a control region and a special repeat region between *trnE* and *trnF*. *G. psidii* is the first representative possessing such a repeat region from the subfamily Greenideinae. These repeat motifs found in Aphidinae, Eriosomatinae and Greenideinae are species-specific, differing in nucleotide composition, length and copy number across different species. All reported complete aphid mitogenomes are A- and C-skewed in the whole genomes and show opposite skewness for C and G between the majority and minority strands. The ratios of nonsynonymous to synonymous substitution rates indicated that the evolution of the aphid mitogenomes has been dominated by purifying selection. Based on whole mitogenome sequences, the phylogenetic relationships among all aphid species with complete mitogenomes were investigated. *G. psidii* was robustly clustered with *Cervaphis quercus*, confirming its affiliation with the subfamily Greenideinae.

1. Introduction

The insect mitochondrial genome (mitogenome) is a closed-circular double-stranded molecule that is 15–18 kb in length [1]. It generally encodes 37 genes, which are oriented in different ways: 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes and two ribosome RNA (rRNA) genes [2]. In addition to these genes, a non-coding A + T-rich region termed the control region is also included in the insect mitogenome. Owing to its small size, maternal inheritance, fast evolutionary rate and limited recombination, insect mitogenomes have been widely used in molecular evolution [3], comparative genomics [4,5], phylogenetics at different taxonomic levels [6–8] and population genetics [9,10].

Aphids are an important insect group of phloem-feeding hemipterans, which comprises approximately 5000 species within three families, Adelgidae, Phylloxeridae and Aphididae (24 subfamilies) [11,12]. Aphids are characterized by complex life cycles, inducing galls on host plants and harboring different endosymbionts [13–16]. In addition, many species are serious pests for agriculture and forestry. Therefore, aphids have been considered good model organisms for ecological and evolutionary studies.

Although the mitogenomes have been widely used in the phylogenetic studies of many insect groups [1,6,7], this is not the case for aphids. Previous aphid molecular phylogenetic studies have been mainly based on only a couple of genes [17–19]. Chen et al. [8] constructed the first mitogenomic phylogeny of viviparous aphids. However, only 13 PCGs, 12S rRNA, tRNA-Val and 16S rRNA genes were used in that study due to the experimental difficulties in sequencing complete aphid mitogenomes. Ren et al. [20] then investigated the phylogenetic relationships of *Rhus* gall aphids using whole mitogenome sequences. To date, only 29 complete mitogenomes of aphids have been reported, covering 26 species under five subfamilies (mostly from Aphidinae and Eriosomatinae) (Table 1) [4,20–32].

In this study, we sequenced the complete mitogenome of *Greenidea psidii* van der Goot using high-throughput sequencing, which is the first species from the aphid tribe Greenideini (Aphididae: Greenideinae). *G. psidii* feeds on the young shoots and undersurfaces of the young leaves of *Psidium guajava* and other Myrtaceae plants. This species is widely distributed in East and Southeast Asia and has been introduced to many other countries and regions, such as Hawaii, California, Brazil, Costa Rica, Mexico, Panama and Venezuela [16]. In Taiwan, alate viviparous females of this species are produced throughout the year [33]. We characterized the complete mitogenome of *G. psidii* and compared it with other aphid mitogenomes by the following aspects: genome size and organization, nucleotide composition, codon usage, evolutionary rate and the special non-coding repeat region. Phylogenetic analyses

* Corresponding author at: Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China.
E-mail address: qiaogx@ioz.ac.cn (G.-X. Qiao).

https://doi.org/10.1016/j.ijbiomac.2018.10.209

Table 1
The complete mitochondrial genomes of aphids.

Subfamily	Tribe	Species	Length (bp)	Accession number	Reference
Aphidinae	Aphidini	*Aphis craccivora* Koch	15,308	NC_031387	[21]
		Aphis craccivora Koch	15,305	KX447142	[21]
		Aphis gossypii Glover	15,869	NC_024581	[22]
		Schizaphis graminum (Rondani)	15,721	NC_006158	[23]
	Macrosiphini	*Acyrthosiphon pisum* (Harris)	16,971	NC_011594	Directly downloaded from GenBank
		Cavariella salicicola (Matsumura)	16,317	NC_022682	[4]
		Diuraphis noxia (Kurdjumov)	15,784	NC_022727	[24]
		Myzus persicae (Sulzer)	17,382	NC_029727	[25]
		Myzus persicae (Sulzer)	15,582	KU877171	Directly downloaded from GenBank
		Sitobion avenae (Fabricius)	15,180	NC_024683	[26]
Eriosomatinae	Eriosomatini	*Eriosoma lanigerum* (Hausmann)	15,640	NC_033352	[27]
	Fordini	*Baizongia pistaciae* (Linnaeu)	15,602	NC_035314	[20]
		Kaburagia ensigallis Tsai & Tang	16,164	MF043984	[20]
		Kaburagia ovatirhusicola Xiang	16,184	MF043985	[20]
		Kaburagia rhusicola Takagi	16,164	MF043986	[20]
		Kaburagia rhusicola Takagi	16,159	MF043987	[20]
		Melaphis rhois (Fitch)	15,436	NC_036065	[28]
		Nurudea choui (Xiang)	15,308	NC_035310	[20]
		Nurudea ibofushi Matsumura	16,054	NC_035311	[20]
		Nurudea meitanensis (Tsai & Tang)	15,301	NC_035316	[20]
		Nurudea shiraii (Matsumura)	15,389	NC_035301	[20]
		Nurudea yanoniella (Matsumura)	15,858	NC_035313	[20]
		Schlechtendalia chinensis (Bell)	16,047	NC_032386	[29]
		Schlechtendalia elongallis (Tsai & Tang)	16,191	NC_035315	[20]
		Schlechtendalia flavogallis (Tang)	16,150	NC_035312	[20]
		Schlechtendalia peitan (Tsai & Tang)	15,609	NC_035302	[20]
Greenideinae	Cervaphidini	*Cervaphis quercus* Takahashi	15,272	NC_024926	[30]
	Greenideini	*Greenidea psidii* van der Goot	16,202	MH844624	This study
Hormaphidinae		*Hormaphis betulae* (Mordvilko)	15,088	NC_029495	[31]
Mindarinae		*Mindarus keteleerifoliae* Zhang	15,199	NC_033410	[32]

based on all complete mitogenome sequences from the Aphididae were carried out using maximum-likelihood and Bayesian methods.

2. Materials and methods

2.1. Sample collection and DNA extraction

Specimens of *G. psidii* were collected on *P. guajava* from Hainan, China (Baoting County, 18.68°N, 109.57°E, altitude 260 m). Fresh specimens were preserved in 75% ethanol and liquid nitrogen for slide mounting and DNA extraction, respectively. Both voucher specimens and frozen samples were deposited in the National Zoological Museum of China, Institute of Zoology, Chinese Academy of Sciences, Beijing, China (NZMC no. 26617). The slide-mounted specimens were identified by Ge-Xia Qiao based on the external morphology, then verified using National Center for Biotechnology Information (NCBI) BLAST searches of the *cox1* barcode sequence. Mitochondrial DNA (mtDNA) was extracted from pooled whole-body tissues of approximately 50 aphids using an improved extraction method [34]. The quantity and quality of the isolated mtDNA were determined using a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA).

2.2. Mitogenome sequencing and assembly

The sequencing was performed on an Illumina Hiseq X Ten (Illumina, San Diego, CA, USA) with the strategy of 150 paired-ends and an insert size of 430 bp. After filtering out low-quality and adapter-contaminated reads, high-quality clean reads were assembled using SPAdes v.3.10.1 [35].

2.3. Mitogenome annotation and sequence analyses

A preliminary mitogenome annotation was carried out using MITOS v.2 WebServer [36] with the invertebrate mitochondrial genetic code. The tRNA genes were identified by MITOS and tRNAscan-SE Search Server v.2.0 [37]. Secondary structures for tRNAs were drawn with VARNA [38]. The PCGs and rRNA genes were determined based on their alignments with available aphid mitogenomes. The circular mitogenome was visualized with CGView Server [39].

The nucleotide composition and the relative synonymous codon usage (RSCU) were estimated with MEGA v.7 [40]. The composition skewness was calculated according to the following formulas: AT-skew = [A − T] / [A + T], GC-skew = [G − C] / [G + C] [41]. The nonsynonymous substitution rate (Ka) of PCGs was measured to compare the substitution rate among different aphid species. Using *Pachypsylla venusta* (Osten-Sacken) (GenBank accession no. NC_006157) as a reference, pairwise Ka values were calculated with DnaSP v. 6.12.01 [42], and the genetic distance of each codon position was computed with MEGA. To detect the selective pressure of aphid mitogenomes, the ratio of nonsynonymous to synonymous substitution rates (Ks) (Ka/Ks) and Jukes-Cantor adjusted Ka/Ks (JKa/Jks) were calculated for each PCG with DnaSP. The mitogenome sequence of *Myzus persicae* (Sulzer) (GenBank accession no. KU877171) was excluded from the above analyses as it contained many degenerate bases. The secondary structure of the repeat unit in the repeat region was predicted with the Mfold web server [43].

2.4. Phylogenetic analyses

Based on the whole mitogenome sequences, phylogenetic trees were estimated using maximum-likelihood (ML) approach and Bayesian inference (BI). Grape phylloxera, *Daktulosphaira vitifoliae* (Fitch) (family Phylloxeridae, GenBank accession no. DQ021446), was used as an outgroup. Each PCG was aligned using the TranslatorX online server [44], employing MAFFT to perform the protein alignment. The rRNA and tRNA genes were independently aligned with the MAFFT online service [45,46], and unreliably aligned regions were removed using Gblocks 0.91b [47,48]. The optimal partitioning scheme and substitution models were assessed with PartitionFinder 2.1.1 [49]. All 37 genes totaling 14,324 bp were divided into three partitions and the best-fit model for each partition was GTR + I + G. ML analysis was inferred with 1000 rapid bootstrapping replicates using RAxML v8.2.10 [50]. Bayesian

inference was performed in MrBayes 3.2.6 [51] with four chains, sampling the chains every 1000 generations. Two independent runs of 50,000,000 generations were conducted. Stationarity was assumed when the average standard deviation of split frequencies fell below 0.01 and the effective sample size (ESS) values of all parameters were >200. The first 25% trees were discarded as burn-in.

3. Results and discussion

3.1. General features of the mitochondrial genome of G. psidii

The complete mitogenome sequence of *G. psidii* is a closed-circular molecule of 16,202 bp in length (Fig. 1), and the sequence has been deposited in GenBank under accession number MH844624. The mitogenome contains 13 PCGs, 22 tRNA genes, two rRNA genes, a large non-coding region (control region) and a special non-coding repeat region (Table 2). Twenty-three genes (9 PCGs and 14 tRNA genes) are transcribed on the majority strand (J strand), with the remaining genes being located on the minority strand (N strand). The gene order is identical to the inferred ancestral arrangement of insects [52]. There are a total of 39 overlapping nucleotides between adjacent genes in 10 locations. As most reported aphid mitogenomes [4,23,27,30–32], the longest overlap (20 bp) exists between *atp8* and *atp6*. In addition to the control region and the repeat region, a total of 221 bp intergenic spacers are present at 20 positions, ranging from 1 to 51 bp in length. The longest spacer sequence is located between *nad5* and *trnH*, which is the same as that of the mitogenome of another Greenideinae species, *Cervaphis quercus* Takahashi [30].

The mitogenome of *G. psidii* is larger than most other sequenced aphid species (Table 1). So far, the largest reported mitochondrial genome in aphids is that of *M. persicae* (17,382 bp), while the smallest is that of *Hormaphis betulae* (Mordvilko) (15,088 bp). The nucleotide length variation across different aphid species is relatively low in coding genes. The size variation of aphid mitogenomes occurs mainly due to differences in the number of non-coding nucleotides, especially in the control region and repeat regions [4,27,30,32].

3.2. Nucleotide composition

The overall nucleotide composition of the *G. psidii* mitogenome is 39.2% T, 9.1% C, 46.2% A and 5.5% G, with a strong bias towards A + T (85.4%) (Table 3). Excluding stop codons, the A + T content of the concatenated PCGs is 84.5%, which is slightly lower than that of the mitogenome as a whole. *G. psidii* is slightly A-skewed (AT-skew = 0.082) and is moderately C-skewed (GC-skew = −0.247) in the whole mitogenome (Table 3). The PCGs encoded on the majority strand are slightly T-skewed (AT-skew = −0.074) and are moderately C-skewed (GC-skew = −0.212), whereas the minority strand encoded PCGs are moderately T-skewed (AT-skew = −0.304) and G-skewed (GC-skew = 0.299), suggesting strand heterogeneity in the nucleotide composition.

The nucleotide compositions of all reported complete aphid mitogenomes are summarized in Table S1. All aphid mitogenomes are

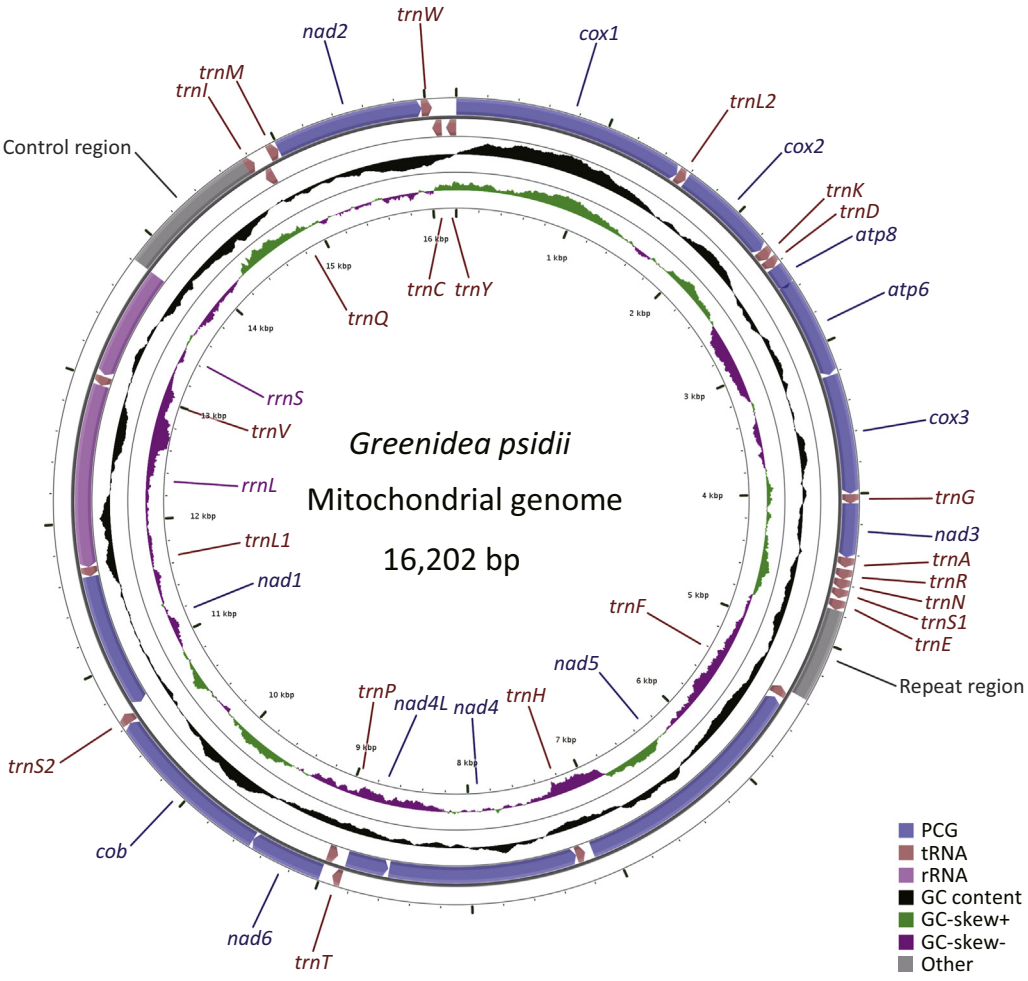

Fig. 1. Circular map of the *Greenidea psidii* mitochondrial genome. Arrows indicate the orientation of gene transcription. The inner circles show GC content and GC-skew, which are plotted as the deviation from the average value of the entire sequence.

Table 2
Organization of the *Greenidea psidii* mitochondrial genome.

Gene	Strand	Position	Length (bp)	Anticodon	Start codon	Stop codon	Intergenic nucleotides (bp)
cox1	J	1–1531	1531		ATA	T	0
trnL2	J	1532–1599	68	TAA			3
cox2	J	1603–2274	672		ATA	TAA	2
trnK	J	2277–2349	73	CTT			0
trnD	J	2350–2416	67	GTC			0
atp8	J	2417–2575	159		ATA	TAA	−20
atp6	J	2556–3209	654		ATT	TAA	7
cox3	J	3217–4002	786		ATG	TAA	2
trnG	J	4005–4067	63	TCC			−3
nad3	J	4065–4421	357		ATA	TAA	−1
trnA	J	4421–4486	66	TGC			11
trnR	J	4498–4564	67	TCG			−1
trnN	J	4564–4630	67	GTT			−1
trnS1	J	4630–4691	62	GCT			11
trnE	J	4703–4770	68	TTC			0
Repeat region		4771–5382	612				0
trnF	N	5383–5448	66	GAA			21
nad5	N	5470–7143	1674		ATT	TAA	51
trnH	N	7195–7259	65	GTG			1
nad4	N	7261–8566	1306		ATA	T	8
nad4L	N	8575–8865	291		ATA	TAA	1
trnT	J	8867–8927	61	TGT			5
trnP	N	8933–9002	70	TGG			1
nad6	J	9004–9498	495		ATT	TAA	−1
cob	J	9498–10,613	1116		ATG	TAA	1
trnS2	J	10,615–10,683	69	TGA			10
nad1	N	10,694–11,629	936		ATT	TAA	0
trnL1	N	11,630–11,695	66	TAG			−1
rrnL	N	11,695–12,962	1268				−1
trnV	N	12,962–13,027	66	TAC			12
rrnS	N	13,040–13,812	773				0
Control region		13,813–14,752	940				0
trnI	J	14,753–14,815	63	GAT			9
trnQ	N	14,825–14,890	66	TTG			34
trnM	J	14,925–14,993	69	CAT			0
nad2	J	14,994–15,974	981		ATA	TAA	−2
trnW	J	15,973–16,043	71	TCA			−8
trnC	N	16,036–16,101	66	GCA			30
trnY	N	16,132–16,201	70	GTA			1

rich in A and T, with the A + T content of whole genomes ranging from 82.2% (*H. betulae*) to 85.4% (*G. psidii*). The base compositions at each codon position of PCGs indicate that the third codon positions harbor the highest A + T content (range: 89.8–96.1%; mean: 93.3%). All species are highly congruent in base skewness, and no subfamily preference for skewness is detected (Fig. 2, Table S1). Like most of the hemipteran mitogenomes, aphids are A- and C-skewed in the whole genomes [53,54]. All species show opposite skewness for C and G nucleotides between the majority and minority strands, which may be due to the spontaneous deamination of C and A in the minority strand during replication [55].

Table 3
Nucleotide composition of the *Greenidea psidii* mitochondrial genome.

	T%	C%	A%	G%	A + T%	AT-skew	GC-skew
Whole mitogenome	39.2	9.1	46.2	5.5	85.4	0.082	−0.247
Protein-coding genes	49.2	8.0	35.3	7.4	84.5	−0.164	−0.039
1st codon positions	40.5	8.2	40.5	10.7	81.0	0.000	0.132
2nd codon positions	53.7	13.3	22.8	10.1	76.5	−0.404	−0.137
3rd codon positions	53.4	2.4	42.7	1.5	96.1	−0.111	−0.231
Protein-coding genes-J	44.9	10.0	38.7	6.5	83.6	−0.074	−0.212
Protein-coding genes-N	56.2	4.8	30.0	8.9	86.2	−0.304	0.299
tRNA genes	40.3	5.8	45.7	8.2	86.0	0.063	0.171
rRNA genes	46.3	4.8	38.8	10.1	85.1	−0.088	0.356
Control region	41.9	6.2	47.9	4.0	89.8	0.067	−0.216
Repeat region	34.2	8.0	53.8	4.1	88.0	0.223	−0.322

Stop codons were excluded in the analyses of protein-coding genes.

3.3. Protein-coding genes

The total length of 13 PCGs is 10,923 bp, which accounts for 67.4% of the entire mitogenome of *G. psidii*. Nine PCGs (*cox1*, *cox2*, *atp8*, *atp6*, *cox3*, *nad3*, *nad6*, *cob* and *nad2*) are encoded on the majority strand, while the remaining four genes (*nad5*, *nad4*, *nad4L* and *nad1*) are encoded on the minority strand (Table 2). All PCGs are initiated by the canonical start codon ATN. Four genes (*atp6*, *nad5*, *nad6* and *nad1*) start with ATT, two genes (*cox3* and *cob*) start with ATG, and the seven remaining genes start with ATA. Eleven PCGs are terminated with the typical stop codon TAA, whereas *cox1* and *nad4* end with a single T. In the mitogenomes of insects, including aphids, the truncated stop codons are common, which could be completed by post-transcriptional polyadenylation [56].

The codon distribution patterns are highly consistent across all aphid species (Fig. S1). Phe, Ile, Leu (UUR), Met and Asn are the top five abundant codon families, whereas Cys is the least abundant. Among the 51 amino-acid-coding codons in the *G. psidii* mitogenome, the most frequently used codon is UUU (Phe), followed by AUU (Ile), UUA (Leu), AUA (Met) and AAU (Asn), all of which are composed of A or U (Table S2). In addition, the majority of frequently used codons end with A or U (Fig. 3). These two facts seem to lead to the high A + T content of PCGs and the AT bias of the whole mitogenome.

The nonsynonymous substitution rate (Ka) of PCGs was calculated for each aphid species in comparison with *P. venusta* (Fig. 4). There is no significant difference in the nonsynonymous substitution rates among different aphid species (range: 0.446–0.464; mean: 0.450; standard deviation: 0.004). The evolutionary rate of the third codon position

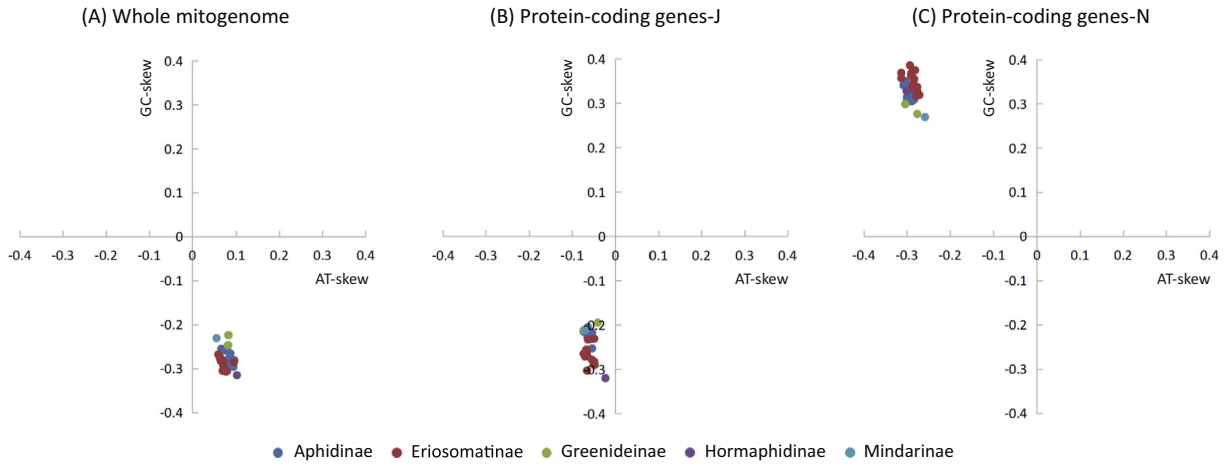

Fig. 2. Nucleotide skewness of aphid mitochondrial genomes. (A) Whole mitogenome; (B) protein-coding genes encoded on the J strand; (C) protein-coding genes encoded on the N strand.

is higher than the first and second codon positions (Fig. 4). The ratios of Ka/Ks and JKa/Jks are below 1 for all PCGs (Fig. 5), suggesting that the major force affecting aphid mitogenome sequences is purifying. Among the 13 PCGs, *cox1* has experienced the strongest purifying pressure (Ka/Ks: 0.067; JKa/Jks: 0.047), while *atp8* (Ka/Ks: 0.670; JKa/Jks: 0.561) and *nad6* (Ka/Ks: 0.529; JKa/Jks: 0.462) have experienced relatively weak purifying pressure.

3.4. Transfer and ribosomal RNA genes

All 22 typical arthropod tRNAs that are encoded in the *G. psidii* mitogenome range from 61 to 73 bp in length and are spread over the entire mitogenome (Table 2). All but one tRNAs display the typical clover-leaf secondary structure (Fig. S2). The truncated secondary structure with the loss of dihydrouridine (DHU) arm occurs in *trnS (AGN)*, which is common in insect mitogenomes. A total of 13 weak-bonded GU base pairs and one mismatched UU base pair are found in the *G. psidii* tRNAs.

Two rRNA genes of *G. psidii* are encoded on the minority strand. *rrnL* is 1268 bp in length and is located between *trnL (CUN)* and *trnV*, while *rrnS* is 773 bp in length and resides between *trnV* and control region (Table 2). The A + T contents of *rrnL* and *rrnS* are 85.7% and 84.2%, respectively.

3.5. Control region and repeat region

The control region of *G. psidii* is located between *rrnS* and *trnI*, with an A + T content of 89.8%. It is 940 bp in length, which falls into the reported range for aphid species (430–2531 bp). The *G. psidii* control region includes a 377 bp lead sequence adjacent to *rrnS*, followed by three conserved structure elements in the Aphididae mitogenomes: an AT-rich zone, a poly-thymidine stretch and a stem-loop region [4,27,30,32]. The lead sequence was also observed in the control region of *Cavariella salicicola* (Matsumura), which may guide the mtDNA replication and transcription [4].

The aphid-specific repeat region between *trnE* and *trnF* is found in the *G. psidii* mitogenome. It is 612 bp in size with an A + T content of 88.0%. This region consists of a 146 bp lead sequence and two 224 bp repeat units separated by an 18 bp sequence, each of which could form a stem-loop structure (Fig. 6).

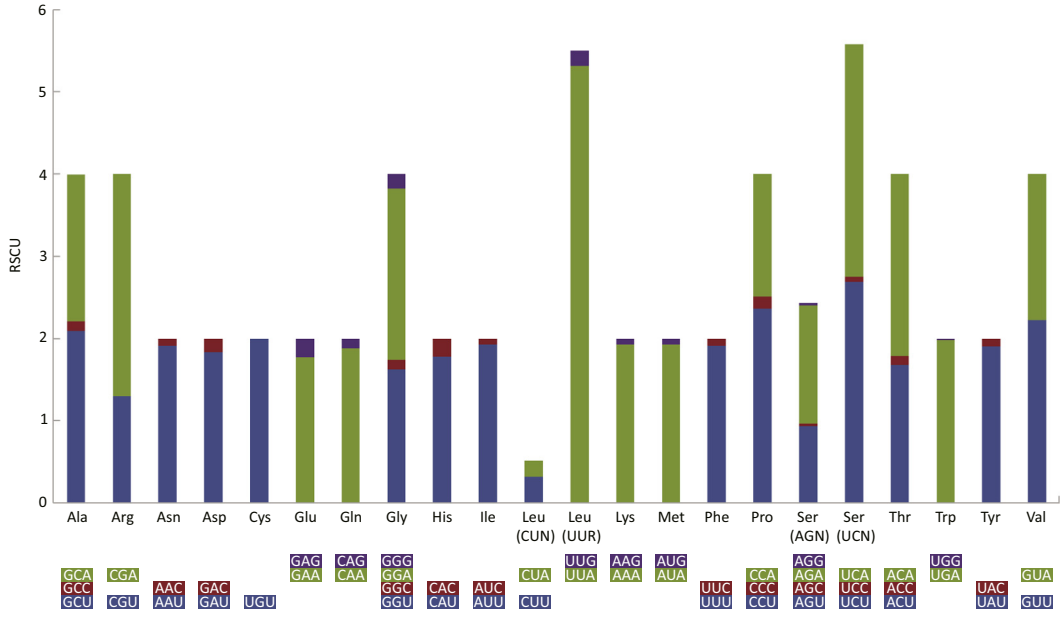

Fig. 3. Relative synonymous codon usage (RSCU) in the *Greenidea psidii* mitochondrial genome.

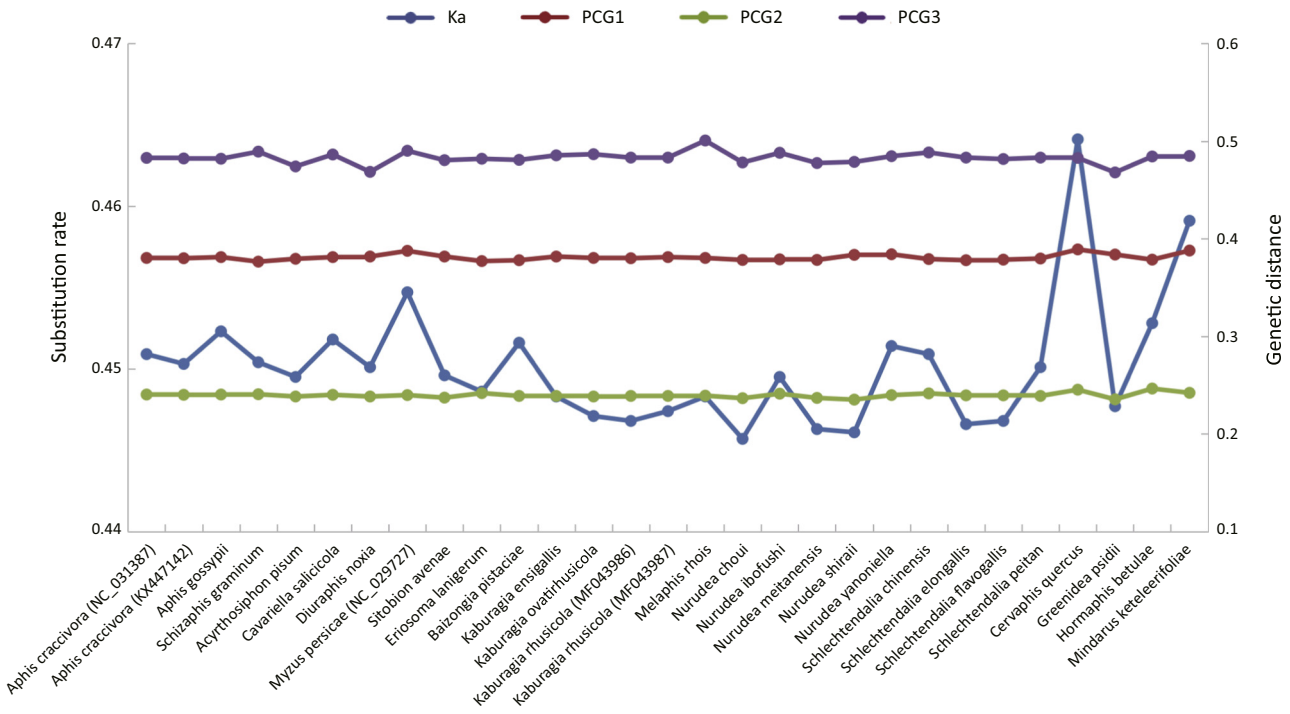

Fig. 4. Nonsynonymous substitution rates (Ka) of aphid mitochondrial protein-coding genes and the genetic distances of each codon positions in comparison with *Pachypsylla venusta*. PCG1, first codon position; PCG2, second codon position; PCG3, third codon position.

Until now, such tandem repeats between *trnE* and *trnF* have been reported in eleven aphid species belonging to three subfamilies: *Acyrthosiphon pisum* (Harris), *Aphis gossypii* Glover, *C. salicicola*, *Diuraphis noxia* (Kurdjumov), *M. persicae*, *Schizaphis graminum* (Rondani), *Sitobion avenae* (Fabricius) (subfamily Aphidinae); *Nurudea ibofushi* Matsumura, *Nurudea yanoniella* (Matsumura), *Schlechtendalia chinensis* (Bell) (subfamily Eriosomatinae); and *G. psidii* (subfamily Greenideinae) (Fig. 6A). This finding refutes the statement that the repeat region is exclusively present in Aphidinae [27,30,32]. The repeat motifs differ in nucleotide composition, length and copy number across different aphid species, leading to varied lengths of the repeat regions (261 bp in *S. avenae* to 1513 bp in *A. pisum*). Considering that the aphid species possessing tandem repeats between *trnE* and *trnF* are scattered within the Aphididae and the repeat regions are species-specific in organization, we assume alternative scenarios that include a single origin in the common ancestor of aphids followed by subsequent losses in some groups, or multiple independent origins in different aphid lineages. The underlying driving forces for its loss or convergent evolution are still unknown and need further study. Similar to the long intergenic spacers in other insect species, the repeat regions in aphids may be an alternative origin of the mtDNA replication [57,58].

3.6. Phylogenetic analyses

To present the relationships among aphid species with complete mitochondrial genomes, we conducted ML and BI phylogenetic analyses

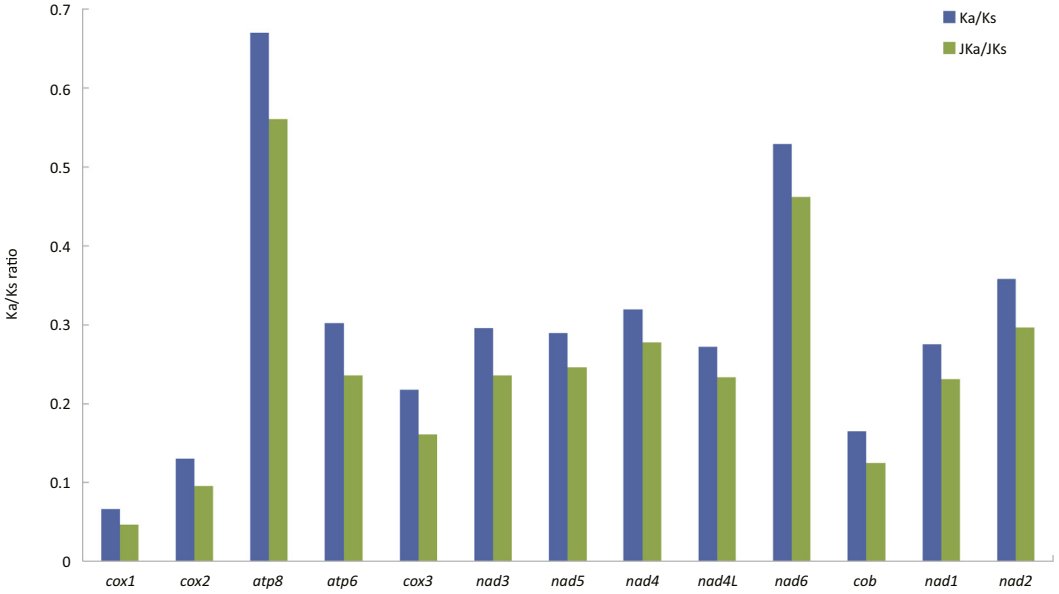

Fig. 5. Ratio of nonsynonymous to synonymous substitution rates (Ka/Ks) and Jukes-Cantor adjusted Ka/Ks (JKa/Jks) for aphid mitochondrial protein-coding genes.

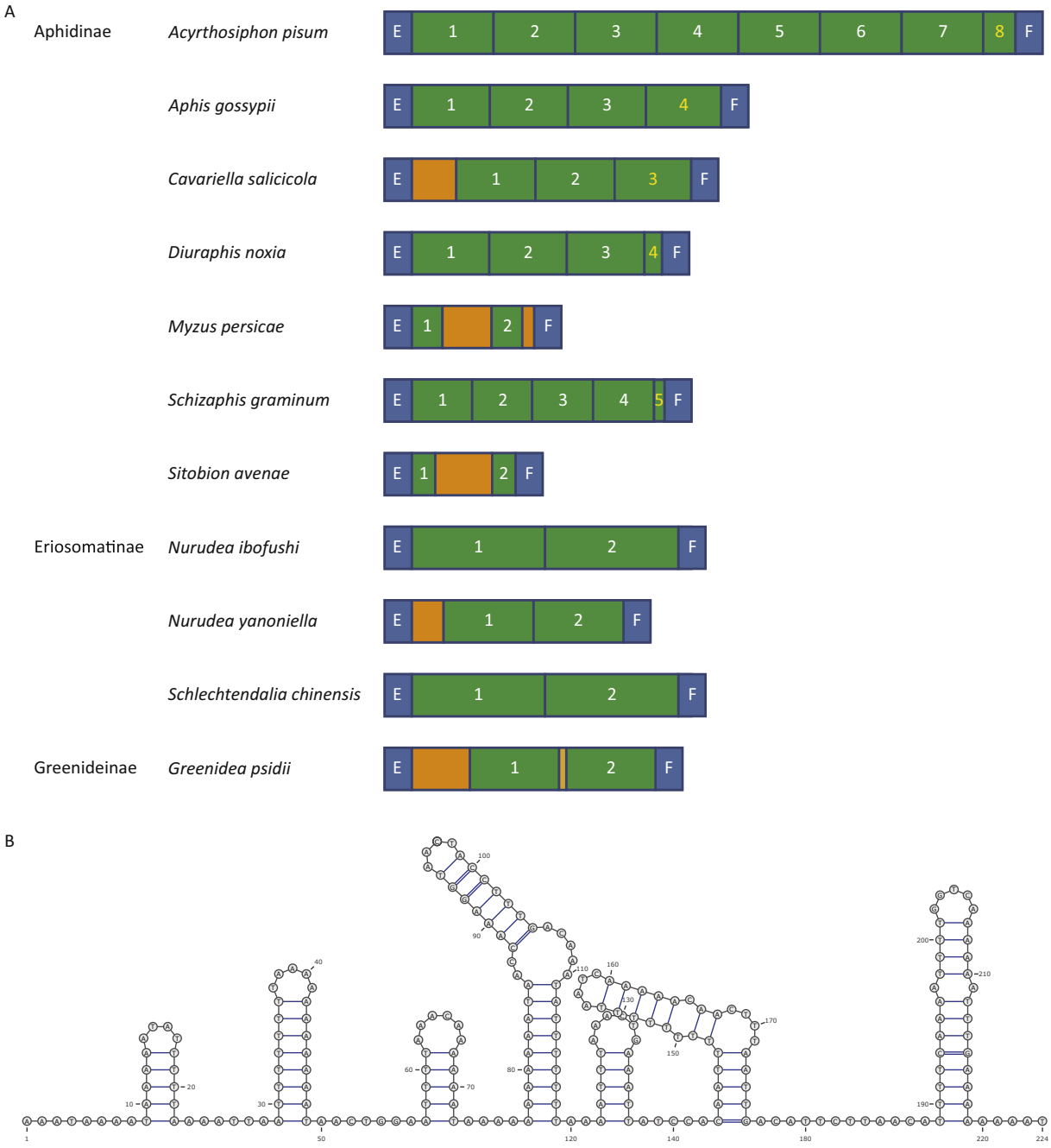

Fig. 6. Structural organizations of the repeat regions in aphid mitochondrial genomes. (A) Organizations of the repeat regions of eleven aphid species. The green rectangle with white and yellow number indicates complete and partial repeat unit, respectively; the orange rectangle indicates non-coding sequence. (B) The putative stem-loop structure of the repeat unit in the *Greenidea psidii* mitochondrial genome.

based on the whole mitogenome sequences. In both analyses (Fig. 7), *G. psidii* was clustered with another greenideine species, *C. quercus*, with strong support (bootstrap, BS = 100%; posterior probability, PP = 1), thus confirming its affiliation with the subfamily Greenideinae. The monophyly of the subfamily Aphidinae and its constituent two tribes was all well supported. Species of the tribe Fordini formed a highly supported monophyletic clade (BS = 100%, PP = 1). However, the subfamily Eriosomatinae was not retrieved as monophyletic, which is consistent with previous studies [8,17,18,59,60]. *Eriosoma lanigerum* (tribe Eriosomatini) was placed as sister to the clades of Greenideinae and Aphidinae in the ML and BI trees, respectively.

Within the clade of Aphididae, *H. betulae* was positioned as the earliest branching taxa in both ML and BI trees, whereas two phylogenetic analyses yielded different relationships for the remaining major clades. In the ML tree (Fig. 7A), the Fordini species formed the sub-basal lineage. The remaining species were split into two clades: one included *E. lanigerum* and the Greenideinae, and the other comprised *Mindarus keteleerifoliae* and all representatives of the Aphidinae. In the BI tree (Fig. 7B), *M. keteleerifoliae* + Greenideinae were placed sub-basally, followed by the Fordini lineage and *E. lanigerum* + Aphidinae. The inner relationships within Aphidinae and Fordini were consistent between different phylogenetic inferences, except for the position of

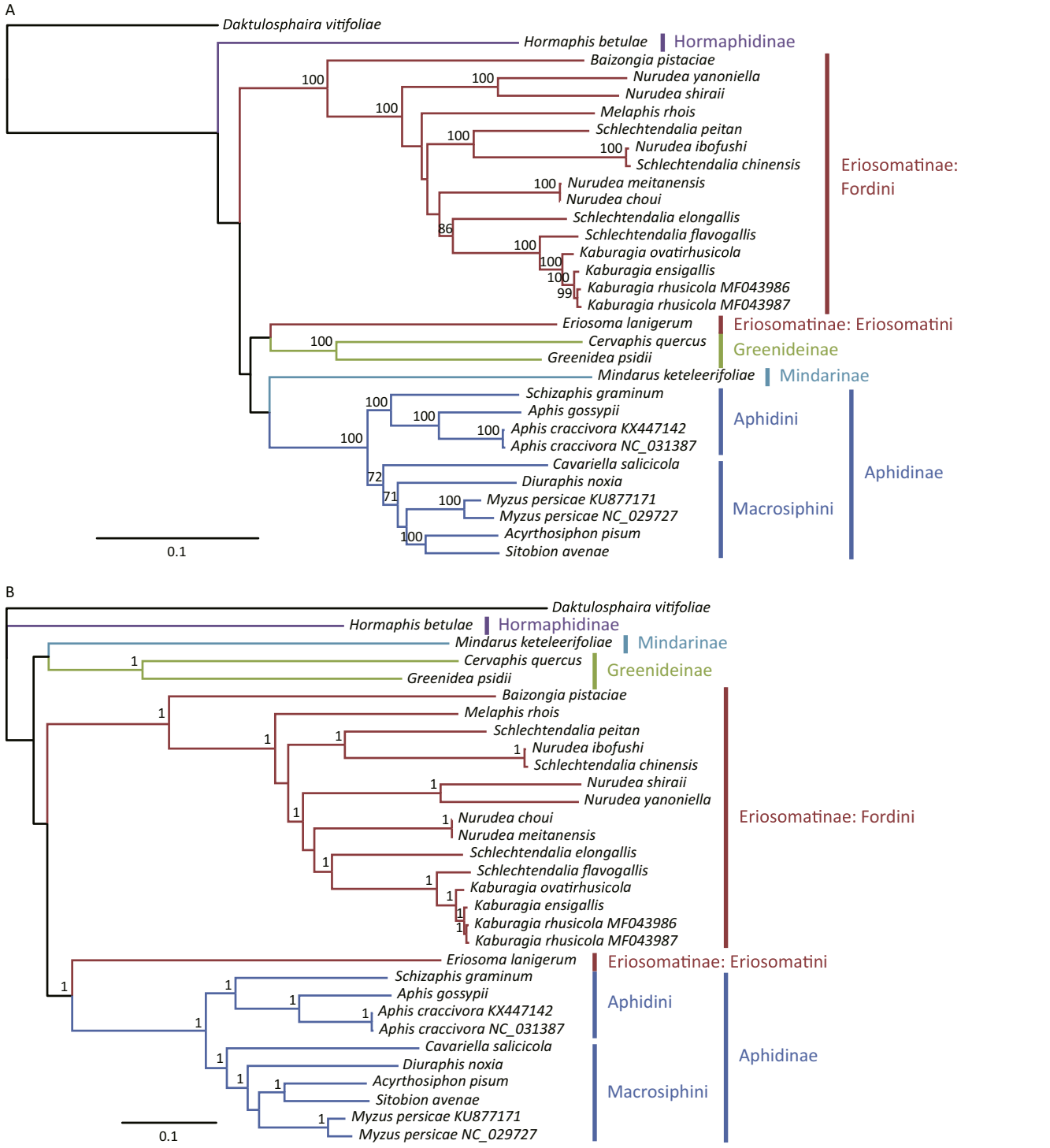

Fig. 7. Aphid phylogenetic trees inferred from whole mitochondrial genomes. (A) Maximum-likelihood (ML) tree. Values at node indicate ML bootstrap (>70%). (B) Bayesian tree. Values at node indicate posterior probabilities (>0.9).

N. yanoniella + *Nurudea shiraii*. The relationships among the Fordini species revealed by ML analysis were congruent with the results of Ren et al. [20]. *Nurudea* and *Schlechtendalia* were not monophyletic, suggesting that these two genera may need taxonomic revisions.

A recent mitochondrial phylogenetic study has proved that mitogenome data was informative and useful for resolving aphid phylogenetic questions [8]. Thereby, more mitogenome sequencing work is absolutely necessary to provide more comprehensive taxon sampling in the future and to, thus, better understand the aphid phylogeny and evolution.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

Thanks to Guang-Chang Ma and Wen-Jing Zhu for their assistance in sample collection. Thanks to Fen-Di Yang for making slides of voucher specimens. The work was supported by the Strategic Priority Research Program A of the Chinese Academy of Sciences (No. XDA19050303)

and the National Natural Science Foundation of China (Nos. 31620103916, 31430078, 31772492).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2018.10.209.

References

[1] S.L. Cameron, Insect mitochondrial genomics: implications for evolution and phylogeny, Annu. Rev. Entomol. 59 (2014) 95–117.
[2] J.L. Boore, Animal mitochondrial genomes, Nucleic Acids Res. 27 (1999) 1767–1780.
[3] R. Shao, M. Dowton, A. Murrell, S.C. Barker, Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects, Mol. Biol. Evol. 20 (2003) 1612–1619.
[4] Y. Wang, X.L. Huang, G.X. Qiao, Comparative analysis of mitochondrial genomes of five aphid species (Hemiptera: Aphididae) and phylogenetic implications, PLoS One 8 (2013), e77511. .
[5] Y. Wang, H. Li, P. Wang, F. Song, W. Cai, Comparative mitogenomics of plant bugs (Hemiptera: Miridae): identifying the AGG codon reassignments between serine and lysine, PLoS One 9 (2014), e101375. .
[6] H. Li, R. Shao, N. Song, F. Song, P. Jiang, Z. Li, W. Cai, Higher-level phylogeny of paraneopteran insects inferred from mitochondrial genome sequences, Sci. Rep. 5 (2015) 8527.
[7] H. Song, C. Amédégnato, M.M. Cigliano, L. Desutter-Grandcolas, S.W. Heads, Y. Huang, D. Otte, M.F. Whiting, 300 million years of diversification: elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling, Cladistics 31 (2015) 621–651.
[8] J. Chen, Y. Wang, L. Jiang, G. Qiao, Mitochondrial genome sequences effectively reveal deep branching events in aphids (Insecta: Hemiptera: Aphididae), Zool. Scr. 46 (2017) 706–717.
[9] C. Ma, P. Yang, F. Jiang, M.P. Chapuis, Y. Shali, G.A. Sword, L. Kang, Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust, Mol. Ecol. 21 (2012) 4344–4358.
[10] V. Battaglia, P. Gabrieli, S. Brandini, M.R. Capodiferro, P.A. Javier, X.G. Chen, A. Achilli, O. Semino, L.M. Gomulski, A.R. Malacrida, G. Gasperi, A. Torroni, A. Olivieri, The worldwide spread of the tiger mosquito as revealed by mitogenome haplogroup diversity, Front. Genet. 7 (2016) 208.
[11] G. Remaudière, M. Remaudière, Catalogue des Aphididae du Monde, Homoptera Aphidoidea, Institut National de la Recherche Agronomique, Paris, 1997.
[12] C. Favret, Aphid species file, version 5.0/5.0, http://Aphid.SpeciesFile.org 2018 (accessed 11 Aug 2018).
[13] P. Buchner, Endosymbiosis of Animals With Plant Microorganisms, Interscience, New York, 1965.
[14] D. Wool, Galling aphids: specialization, biological complexity, and variation, Annu. Rev. Entomol. 49 (2004) 175–192.
[15] S.E. Zytynska, W.W. Weisser, The natural occurrence of secondary bacterial symbionts in aphids, Ecol. Entomol. 41 (2016) 13–26.
[16] R.L. Blackman, V.F. Eastop, Aphids on the world's plants: an online identification and information guide, http://www.aphidsonworldsplants.info/ 2018 (accessed 11 Aug 2018).
[17] B. Ortiz-Rivas, D. Martínez-Torres, 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, Mol. Phylogenet. Evol. 55 (2010) 305–317.
[18] E. Nováková, V. Hypša, J. Klein, R.G. Foottit, C.D. von Dohlen, N.A. Moran, Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*, Mol. Phylogenet. Evol. 68 (2013) 42–54.
[19] H. Choi, S. Shin, S. Jung, D.J. Clarke, S. Lee, Molecular phylogeny of Macrosiphini (Hemiptera: Aphididae): an evolutionary hypothesis for the *Pterocomma*-group habitat adaptation, Mol. Phylogenet. Evol. 121 (2018) 12–22.
[20] Z. Ren, A.J. Harris, R.B. Dikow, E. Ma, Y. Zhong, J. Wen, Another look at the phylogenetic relationships and intercontinental biogeography of eastern Asian–North American *Rhus* gall aphids (Hemiptera: Aphididae: Eriosomatinae): evidence from mitogenome sequences via genome skimming, Mol. Phylogenet. Evol. 117 (2017) 102–110.
[21] W. Sun, B.L. Huynh, J.A. Ojo, B.S. Coates, F. Kusi, P.A. Roberts, B.R. Pittendrigh, Comparison of complete mitochondrial DNA sequences between old and new world strains of the cowpea aphid, *Aphis craccivora* (Hemiptera: Aphididae), Agri Gene 4 (2017) 23–29.
[22] S. Zhang, J. Luo, C. Wang, L. Lv, C. Li, W. Jiang, J. Cui, L.B. Rajput, Complete mitochondrial genome of *Aphis gossypii* Glover (Hemiptera: Aphididae), Mitochondrial DNA A 27 (2016) 854–855.
[23] M.L. Thao, L. Baumann, P. Baumann, Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha), BMC Evol. Biol. 4 (2004) 25.
[24] B. Zhang, C. Ma, O. Edwards, S. Fuller, L. Kang, The mitochondrial genome of the Russian wheat aphid *Diuraphis noxia*: large repetitive sequences between *trnE* and *trnF* in aphids, Gene 533 (2014) 253–260.
[25] H. Yang, J.J. Wang, R.H. Dai, H.P. Yu, M.F. Yang, Sequence and phylogenetic analysis of the complete mitogenome of *Myzus persicae* (Hemiptera: Aphididae), Acta Entomol. Sin. 60 (2017) 84–94.
[26] B. Zhang, J. Zheng, L. Liang, S. Fuller, C.S. Ma, The complete mitochondrial genome of *Sitobion avenae* (Hemiptera: Aphididae), Mitochondrial DNA A 27 (2016) 945–946.
[27] Y. Wang, L. Jiang, Y. Liu, J. Chen, G. Qiao, General methods to obtain and analyze the complete mitochondrial genome of aphid species: *Eriosoma lanigerum* (Hemiptera: Aphididae) as an example, Zool. Syst. 41 (2016) 123–132.
[28] Z.M. Ren, J. Wen, Complete mitochondrial genome of the North American *Rhus* gall aphid *Melaphis rhois* (Hemiptera: Aphididae: Eriosomatinae), Mitochondrial DNA B Resour. 2 (2017) 169–170.
[29] Z.M. Ren, X. Bai, A.J. Harris, J. Wen, Complete mitochondrial genome of the *Rhus* gall aphid *Schlechtendalia chinensis* (Hemiptera: Aphididae: Eriosomatinae), Mitochondrial DNA B Resour. 1 (2016) 849–850.
[30] Y. Wang, X.L. Huang, G.X. Qiao, The complete mitochondrial genome of *Cervaphis quercus* (Insecta: Hemiptera: Aphididae: Greenideinae), Insect Sci. 21 (2014) 278–290.
[31] Y.Q. Li, J. Chen, G.X. Qiao, Complete mitochondrial genome of the aphid *Hormaphis betulae* (Mordvilko) (Hemiptera: Aphididae: Hormaphidinae), Mitochondrial DNA A 28 (2017) 265–266.
[32] Y. Wang, J. Chen, L.Y. Jiang, G.X. Qiao, The complete mitochondrial genome of *Mindarus keteleerifoliae* (Insecta: Hemiptera: Aphididae) and comparison with other Aphididae insects, Int. J. Mol. Sci. 16 (2015) 30091–30102.
[33] R. Takahashi, Aphididae of Formosa part 2, Rep. Dep. Agric. Govt. Res. Inst. Formosa 4 (1923) 1–173.
[34] J. Chen, R. Guan, S. Chang, T. Du, H. Zhang, H. Xing, Substoichiometrically different mitotypes coexist in mitochondrial genomes of *Brassica napus* L, PLoS One 6 (2011), e17662.
[35] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477.
[36] M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf, P.F. Stadler, MITOS: improved *de novo* metazoan mitochondrial genome annotation, Mol. Phylogenet. Evol. 69 (2013) 313–319.
[37] T.M. Lowe, P.P. Chan, tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes, Nucleic Acids Res. 44 (2016) W54–W57.
[38] K. Darty, A. Denise, Y. Ponty, VARNA: interactive drawing and editing of the RNA secondary structure, Bioinformatics 25 (2009) 1974–1975.
[39] J.R. Grant, P. Stothard, The CGView Server: a comparative genomics tool for circular genomes, Nucleic Acids Res. 36 (2008) W181–W184.
[40] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874.
[41] N.T. Perna, T.D. Kocher, Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes, J. Mol. Evol. 41 (1995) 353–358.
[42] J. Rozas, A. Ferrer-Mata, J.C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado, S.E. Ramos-Onsins, A. Sánchez-Gracia, DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets, Mol. Biol. Evol. 34 (2017) 3299–3302.
[43] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res. 31 (2003) 3406–3415.
[44] F. Abascal, R. Zardoya, M.J. Telford, TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations, Nucleic Acids Res. 38 (2010) W7–W13.
[45] S. Kuraku, C.M. Zmasek, O. Nishimura, K. Katoh, aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity, Nucleic Acids Res. 41 (2013) W22–W28.
[46] K. Katoh, J. Rozewicki, K.D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, Brief. Bioinform. (2017)https://doi.org/10.1093/bib/bbx108.
[47] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (2000) 540–552.
[48] G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, Syst. Biol. 56 (2007) 564–577.
[49] R. Lanfear, P.B. Frandsen, A.M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses, Mol. Biol. Evol. 34 (2016) 772–773.
[50] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (2014) 1312–1313.
[51] F. Ronquist, M. Teslenko, P. van der Mark, D.L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard, J.P. Huelsenbeck, MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, Syst. Biol. 61 (2012) 539–542.
[52] D.O. Clary, D.R. Wolstenholme, The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code, J. Mol. Evol. 22 (1985) 252–271.
[53] N. Song, A. Liang, The complete mitochondrial genome sequence of *Geisha distinctissima* (Hemiptera: Flatidae) and comparison with other hemipteran insects, Acta Biochim. Biophys. Sin. 41 (2009) 206–216.
[54] Y. Wang, J. Chen, L.Y. Jiang, G.X. Qiao, Hemipteran mitochondrial genomes: features, structures and implications for phylogeny, Int. J. Mol. Sci. 16 (2015) 12382–12404.
[55] A. Reyes, C. Gissi, G. Pesole, C. Saccone, Asymmetrical directional mutation pressure in the mitochondrial genome of mammals, Mol. Biol. Evol. 15 (1998) 957–966.
[56] D. Ojala, J. Montoya, G. Attardi, tRNA punctuation model of RNA processing in human mitochondria, Nature 290 (1981) 470.
[57] R.H. Crozier, Y.C. Crozier, The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization, Genetics 133 (1993) 97–117.
[58] E.M. Dotson, C.B. Beard, Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*, Insect Mol. Biol. 10 (2001) 205–215.
[59] C.D. von Dohlen, N.A. Moran, Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation, Biol. J. Linn. Soc. 71 (2000) 689–717.
[60] B. Ortiz-Rivas, A. Moya, D. Martínez-Torres, Molecular systematics of aphids (Homoptera: Aphididae): new insights from the long-wavelength opsin gene, Mol. Phylogenet. Evol. 30 (2004) 24–37.