Acta Biochim Biophys Sin 2013, 45: 309–319 | © The Author 2013. Published by ABBS Editorial Office in association with Oxford University Press on behalf of the Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. DOI: 10.1093/abbs/gmt009.

Original Article

The complete mitochondrial genome of spittlebug *Paphnutius ruficeps* (Insecta: Hemiptera: Cercopidae) with a fairly short putative control region

Jie Liu[1,2] and Aiping Liang[1*]

[1]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[2]Graduate University of the Chinese Academy of Sciences, Beijing 10049, China
*Correspondence address. Tel: +86-64807223; Fax: +86-64807099; E-mail: LiangAP@ioz.ac.cn

The mitochondrial genome of the spittlebug *Paphnutius ruficeps* is a double-strand DNA circular molecule of 14,841 bp with a total A and T content of 73.8%. It is one of the shortest genomes among published hemipteran mitogenomes and encodes 13 protein-coding genes, 2 ribosome RNA genes and 22 transfer RNA (tRNA) genes. The gene order is consistent with the hypothesized ancestral arthropod genome arrangement. Most of the protein-coding genes use ATG as start and TAA as stop codon. The codons show an evident bias toward the nucleotides T and A at the third codon position and the most commonly used codons contain more A and T than their synonymous ones. The anticodons of the 22 tRNA genes are identical to those of the mitogenome of *Philaenus spumarius*, another studied spittlebug. All the tRNAs could be folded into traditional clover leaf secondary structures. The putative control region (traditionally called A + T-rich region) is the main non-coding part of the mitogenome. The AT content of this region (74.5%) is not significantly higher than that of the total mitogenome (73.8%) and slightly lower than that of the N-chain protein-coding genes (75.3%). The absence of repeat sequences as well as its short length is the most obvious characteristics of the mitochondrial genome of *Paphnutius ruficeps* compared with those of other published hemipteran species.

Keywords mitogenome; putative control region; *Paphnutius ruficeps*

Received: September 27, 2012 Accepted: December 6, 2012

Introduction

Mitochondria are the energy-transducing organelles of eukaryotic cells where adenosine triphosphate (ATP) is produced via the process of oxidative phosphorylation [1]. They are special compared with other organelles as they contain a relatively small genome. The typical animal mitochondrial genome is a covalently closed circular, double-strand DNA molecule. Next to 13 protein-coding genes (PCGs) – three cytochrome oxidase subunits genes (*cox1*, *cox2*, *cox3*), two ATP synthase subunits genes (*atp6*, *atp8*), seven NADH dehydrogenase subunits genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*), and the cytochrome b gene (*cytb*)–it contains two ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes, and a putative control region (also called A + T-rich region) [2]. Compared with nuclear genome, mitogenome has several unique traits such as maternal heredity, lack of homologous recombination, and faster accumulation of mutations, which make it a widely used molecular marker for analyzing phylogenetic, phylogeographic, or population structure problems [3]. Since it features the complete gene expression mechanisms including regulation, transcription, and translation, the mitochondrion can be regarded as a complete genetic system with the mitogenome playing the central role [4]. Although several mitogenomic sequences are available from public databases such as GenBank, our knowledge of the mitogenomes is still limited in comparison to large biological diversity. Up to date, more than 300 complete or nearly complete insect mitogenomes from different species have been sequenced and are available in GenBank. The currently known insect mitogenomes range from 14,670 to 20,456 bp in length and the variations are mainly caused by the sequence differences in putative control region [5].

Hemiptera is one of the most diverse insect orders [6,7] comprising Heteroptera (true bugs), Coleorrhyncha, Sternorrhyncha (including aphids, scale bugs, whiteflies, and psyllids), and Auchenorrhyncha. Auchenorrhyncha includes Cicadomorpha (including leafhoppers, spittlebugs, and cicadas) and Fulgoromorpha (planthoppers). Thereby the phylogenetic relationships among these groups as well as the monophyly of Auchenorrhyncha are still under discussion [8]. Many hemipteran species are known pests on agricultural crops. Spittlebugs (the superfamily of

Cercopoidea) belong to Cicadomorpha and are known for their special habit of spitting out foam to cover their body during the nymph stage. They are currently assigned into four families (Aphrophoridae, Cercopidae, Clastopteridae, and Machaerotidae) and many representatives are economic pest that cause serious damages to crops [9]. Despite the fact that during the last years the mitogenomic data of hemipteran, especially the Heteroptera species increased rapidly [10], only one complete mitogenome of a spittlebug has been published: *Philaenus spumarius* (Aphrophoridae) so far [11].

To increase the knowledge of the hemipteran and Cercopoidea mitogenomes, we sequenced the one of *Paphnutius ruficeps* that belongs to Cercopidae and can be treated as the representative species of the tribe Paphnutini. *P. ruficeps* is a tiny spittlebug with brightly colored wings and the larvae and adults feed on the juices of vascular bundles. The results obtained here will contribute to a more comprehensive understanding of the mitogenomes of Cercopoidea and Hemiptera.

Materials and Methods

Taxon sampling and DNA extraction

The adult specimens of *P. ruficeps* used in this study were collected in Chongqing Municipality, China, during June 2010. They were stored in 100% ethanol at −80°C in the Key Laboratory of Animal Evolution and Systematics, Institute of Zoology, Chinese Academy of Sciences (Beijing, China). The muscle tissue under pronotum was used for the genomic DNA extraction. DNA extraction was performed using TIANamp genomic DNA kit (Tiangen Biotech Co., Ltd., Beijing, China) and the genomic DNA was stored at −20°C.

Polymerase chain reaction amplification and sequencing

Polymerase chain reaction (PCR) amplification was carried out with a Bio-Rad Mycycler (Hercules, USA) using Qiagen *Taq* polymerase (Beijing, China) in a 30-μl reaction volume (following the manual of polymerase). The PCR cycling condition consisted of a 10-min pre-denaturation step at 95°C and 35−40 cycles of denaturation at 95°C for 1 min, annealing at 54°C for 1 min and elongation at 72°C for 90 s, and an additional elongation step at 72°C for 10 min after the cycles. The annealing temperatures and the elongation times in these cycles were adjusted depending on melting temperature (Tm) of different primer pairs and the lengths of the target products.

The mitogenome was amplified and sequenced using overlap strategy. The detailed procedure used here was schematically showed in **Fig. 1**. All primers used in this study were listed in **Table 1**. There were two sources for these primers: the first batch of PCR amplifications were performed using the published common primers for insect mitogenomes [12] and amplifications products were sequenced directly; the second one used specific primers that were designed according to yielded sequences.

The PCR products were purified by gel electrophoresis and subsequently sequenced with an ABI Prism 377 Genetic Analyzer (Life Technology, Carlsbad, USA) sequencing machine and ABI BigDye sequencing kit (Life Technology).

Sequence assembly, annotation, and analysis

The sequences of the PCR products were manually proof-checked using the software BioEdit [13] and then the ambiguous sequences were re-sequenced to get clear results. In order to avoid a mix with nuclear mitochondrial sequences in final result, we discriminated the sequences from two aspects. First, the sequences were checked by alignment against the published mitogenomes of other hemipteran species. Additionally, the substitution rates of three different codon positions were calculated and compared for PCGs by aligning the sequences with the ones of the spittlebug *P. spumarius*. Subsequently, the sequences were assembled using the software Condon Code aligner (Li-COR, Inc., Nebraska, USA). Sliding windows analyses (windows size: 20 bp) were performed to visualize the fluctuations of nucleotide distribution within the whole genome. PCGs and rRNA genes were identified via alignment against the mitogenome of *P. spumarius* that is available in GenBank (accession No. AY630340) [11]. Nucleotide composition calculations and genetic code

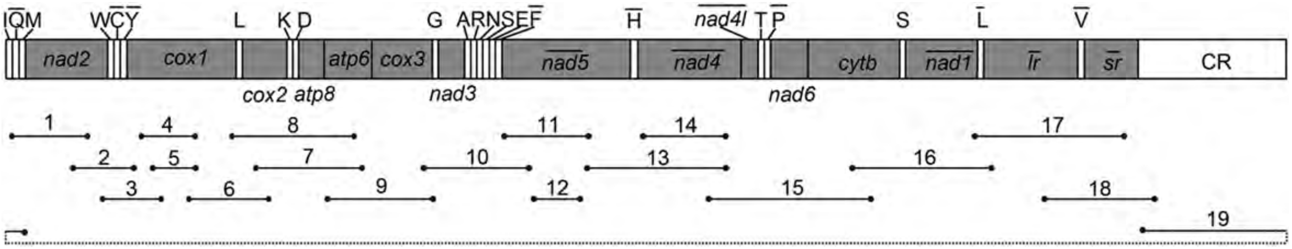

Figure 1 Schematic of the amplification strategy adopted for the mitochondrial genome of *Paphnutius ruficeps* The lines indicate the PCR products and the numbers above the lines are the indexes corresponding to pair numbers shown in **Table 1**.

Table 1 Primers used for the amplification of the mitochondrial genome of *Paphnutius ruficeps*

Pairs	Sense	Sequences (5′ to 3′)	Antisense	Sequences (5′ to 3′)
1	9S	CCTGATTAAAGGATTATTTTGATGT	TWN1284*	ACARCTTTGAAGGYTAWTAGTTT
2	F1-926S	CTCCTTTGCTAGGGTTTTTAC	C1N2353*	GCTCGTGTATCAACGTCTATWCC
3	C1S	AATTGGWGGWTTYGGAAAYTG	F1-22XXA	CCCCAGTCAAACCTCCTATT
4	C1J2195*	TGATTCTTTGGWCACCCWGAAGT	PC12A	CAAATTTCTGAACAYTGACCA
5	PC12S	TGAGCTCATCATATATTTACTGT	PC12A	Listed above
6	3272S	AAAWCWATTGGACATCAATGATA	4440A	ATGWCCWGCAATTATATTWGC
7	4217S	GATCAAGACACCTAGTATTTACACT	N3N5747*	GGRTCAAAYCCACATTCAAATGG
8	3935S	TGAAAATGATAACAAATTTATTTTC	F1-55XXA	TTTGACATTTTCGTTGGGTCC
9	F1-51XXS	TGCGGACTCAATTTATGGTTC	TFN6384*	TATATTTAGAGYATRAYAYTGAAG
10	F1-6029S	ATAGGCGGTTAAATTCCGTTAT	F1-7523A	TTGGGATGGGTTAGGATTGATT
11	7295S	AAAGGGTAATTGAGCTCTCTTAGT	8680A	AAAGCTCATGTWGAAGCTCC
12	N5J7806*	GAMACAARACCTAACCCATCYCA	F1-85XXA	CGGCGTCATTACCTTTACTA
13	8661S	GGAGCTTCWACATGAGCTTT	10715A	CCTCCTCAAATTCATTTTACTA
15	N4LJ9648*	ACCTAAAGCTCCCTCACAWAC	10715A	Listed above
15	F1-105XXS	TTTACACACATATTAGACGAGGT	11753A	GATTTTGCTGAAGGTGAATC
16	11520S	ATCATAACGATAACGAGGTAA	LrN13000*	TTACCTTAGGGATAACAGCGTAA
17	P1612S	CGGTYTGAACTCAGATCATGT	P1612A	TTGYGATAAGTCGTAACAAAGTA
18	13662S	TCAAATTAAATTGAATTGCACAA	SrN14745*	GTGCCAGCAGYYGCGGTTANAC
19	14088S	ACCGCCAAATTCTTTGAAT	169A	AATARGGTATGAACCYATTAGCTT

*Primer was designed based on reference [12], all the remaining primers were designed based on the alignment of the public mitochondrial genome data or derived from the sequencing results.

analyses were conducted with MEGA 4.0 [14]. The tRNA genes were identified using the tRNAScan-SE online server [15] (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/) with the following settings: search mode: tRNA scan only; Source: Mito/Chloroplast; Cove score cutoff: 5. The tRNA-Arg gene cannot be recognized automatically and was therefore identified by alignment with *P. spumarius* and *Drosophila yukuba* (Genbank accession No. X03240) [11,16]. The secondary structure of tRNA-Arg was predicted with the aid of the Mfold web server (http://www.bioinfo.rpi.edu/applications/mfold/) [17] using default settings. Potential secondary structure folds of the putative control regions were also predicted using Mfold web server using default settings. Repeat sequences in the putative control region were identified with the dot-plot function provided by BioEdit and the software of Tandem Repeats Finder version 4.04 [18]. The ends of the two rRNA genes were determined by aligning with published mitogenomes (mitogenomes of *D. yukuba* and *P. spumarius* were used) and the boundary of the adjacent tRNA genes.

Results

Genome content and nucleotide bias

The complete mitogenome of *P. ruficeps* is now available in Genbank under accession No. JF821187.

The mitogenome of *P. ruficeps* is a 14,841-bp covalently closed circular DNA molecular that encodes 13 PCGs, 22 tRNA genes, and 2 rRNA genes. The gene location is the same as in *D. yukuba* [16] and *P. spumarius* [11]. The mitogenome shows a high nucleotide coding efficiency: 14,408 out of a total of 14,841 base pairs are used for coding genes. Next to the putative control region, only very short intragenic spacers were found between neighboring genes. Short gene overlaps are observed between adjacent tRNA genes, PCGs, tRNA and PCGs as well as tRNA and rRNA genes.

The majority-coding strand (J-strand) encodes 23 genes (9 PCGs and 14 tRNA genes) while the remaining 14 genes (4 PCGs, 8 tRNA genes, and 2 rRNA genes) are encoded on the minority-coding strand (N-strand). The detailed positions and other primary information are shown in **Table 2**. The nucleotide composition shows an extreme bias toward A and T: in total the mitogenome contains 73.8% A and T (AT skew is 0.019 and GC-0.015 by J-strand). The results of sliding windows analyses are shown in **Fig. 2(A)**. The frequency distribution of G and C content for 20-bp fragment each is illustrated in **Fig. 2(B)**.

Protein coding genes

The mitogenome of *P. ruficeps* encodes 13 PCGs. They are ATP synthase subunits 6 and 8 (*atp6* and *atp8*), cytochrome c oxidase subunits I, II, and III (*cox1*, *cox2*, and *cox3*); cytochrome b (*cytb*), and NADH dehydrogenase subunits 1

Table 2 Gene position and the primary information in the mitochondrial genome of *Paphnutius ruficeps*

Genes	Begin	End	Size (bp)	Codon	
				Start	Stop
Total genome	1	14,841	14,841		
tRNA-Ile	1	67	67		
tRNA-Gln	66	133	68		
tRNA-Met	133	201	69		
nad2	202	1191	990	TTG	TAG
tRNA-Trp	1195	1261	67		
tRNA-Cys	1254	1317	64		
tRNA-Tyr	1323	1389	67		
cox1	1391	2924	1534	ATG	T - -
tRNA-Leu(TTR)	2925	2991	67		
cox2	2992	3664	673	ATT	T - -
tRNA-Lys	3665	3735	71		
tRNA-Asp	3736	3801	66		
atp8	3802	3954	153	ATT	TAA
atp6	3948	4613	666	ATG	TAA
cox3	4615	5395	781	ATG	T - -
tRNA-Gly	5396	5460	65		
nad3	5461	5814	354	ATG	TAA
tRNA-Ala	5814	5877	64		
tRNA-Arg	5879	5945	67		
tRNA-Asn	5946	6012	67		
tRNA-Ser (AGY)	6012	6079	68		
tRNA-Glu	6079	6145	67		
tRNA-Phe	6144	6208	65		
nad5	6208	7917	1710	TTG	TAA
tRNA-His	7918	7980	63		
nad4	7981	9298	1318	ATG	T - -
nad4l	9292	9579	288	ATG	TAA
tRNA-Thr	9582	9644	63		
tRNA-Pro	9644	9709	66		
nad6	9711	10,223	513	ATA	TAA
cytb	10,223	11,349	1127	ATG	TAG
tRNA-Ser (TCN)	11,355	11,421	67		
nad1	11,442	12,359	918	ATG	TAG
tRNA-Leu (CTN)	12,361	12,428	68		
lrRNA6	12,431	13,692	1262		
tRNA-Val	13,692	13,762	71		
srRNA	13,764	14,531	768		
control region	14,532	14,841	310		

Names of the genes that are located on the N-strand were underlined.

to 6 (*nad1*, *nad2*, *nad4*, *nad4*, *nad4l*, *nad5,* and *nad6*) (**Table 2**).

A total of four different kinds of start codons were identified in the PCGs. ATG is the most widely used one (in *atp6*, *cox1*, *cox3*, *cytb*, *nad1*, *nad3*, *nad4*, and *nad4l*) followed by TTG (*nad2* and *nad5*), ATT (*cox2* and *atp8*), and ATA (*nad6*). Additionally, nine PCGs have complete termination codons. Termination codon TAA is used for six PCGs (*atp8*, *atp6*, *nad3*, *nad5*, *nad4l*, and *nad6*), TAG for other three (*nad2*, *cytb,* and *nad1*), while a single T as incomplete stop codon for the remaining four (*cox1*, *cox2*, *cox3,* and *nad4*).

The relative synonymous codon usage (RSCU) [19] is listed in **Table 3**. It shows an evident bias toward the codons containing U or A at the third codon position. However, no significant difference of RSCU value has been found between J-strand- and N-strand-coded PCGs. The most widely used codons in *P. ruficeps* contain more A and T than their synonymous codons and usually use A or T at the third codon position. **Table 3** shows all used codons.

The A and T biases of different codon positions are shown in **Table 4**. In general it is very similar among the first and second position (average 71.1 and 68.7%, respectively). However, the third position shows an extreme bias toward A and T (average 81.7%). The N-strand coding PCGs use slightly more A and T than J-strand coding genes at each codon position. The amino acid composition in proteins encoded by the mitogenome is summarized and shown in **Fig. 3**. Leucine (Leu), serine (Ser), phenylalanine (Phe), isoleucine (Ile), and methionine (Met) are the most frequently used five amino acids and account for more than 50% of entire genome (**Fig. 3**).

tRNA and rRNA genes

A total of 22 tRNA genes were identified in the mitogenome and their locations are identical to these *P. spumarius* [11]. All tRNAs except tRNA-Arg could be automatically identified by tRNAScane-SE [15] online server. All of tRNAs can be folded into cloverleaf structures. The length of the receptor arms is conservative with seven base pairs and one single base. The exception to this is tRNA-Arg that has only six base pairs. Except for tRNA-Asn and tRNA-Ser (GCU) with four base pairs, all anticodon arms have a stem with five base pairs and a seven base loop that contains anticodons in its center. However, the sizes of DHU and TΨC arms vary considerably among these tRNAs, ranging from 10 to 16 and from 9 to 18 bases, respectively. The predicted cloverleaf fold of $tRNA^{Asp}$ is illustrated as a representative sample in **Fig. 4**.

The mitogenome encodes two rRNA genes, the large ribosomal subunit rRNA (16S) and the small ribosomal subunit rRNA (12S). The former is located between tRNA-Leu and tRNA-Val and the latter between tRNA-Val and the putative control region.

Putative control region

The putative control region is the main and largest noncoding region within the mitogenome of *P. ruficeps*. The A

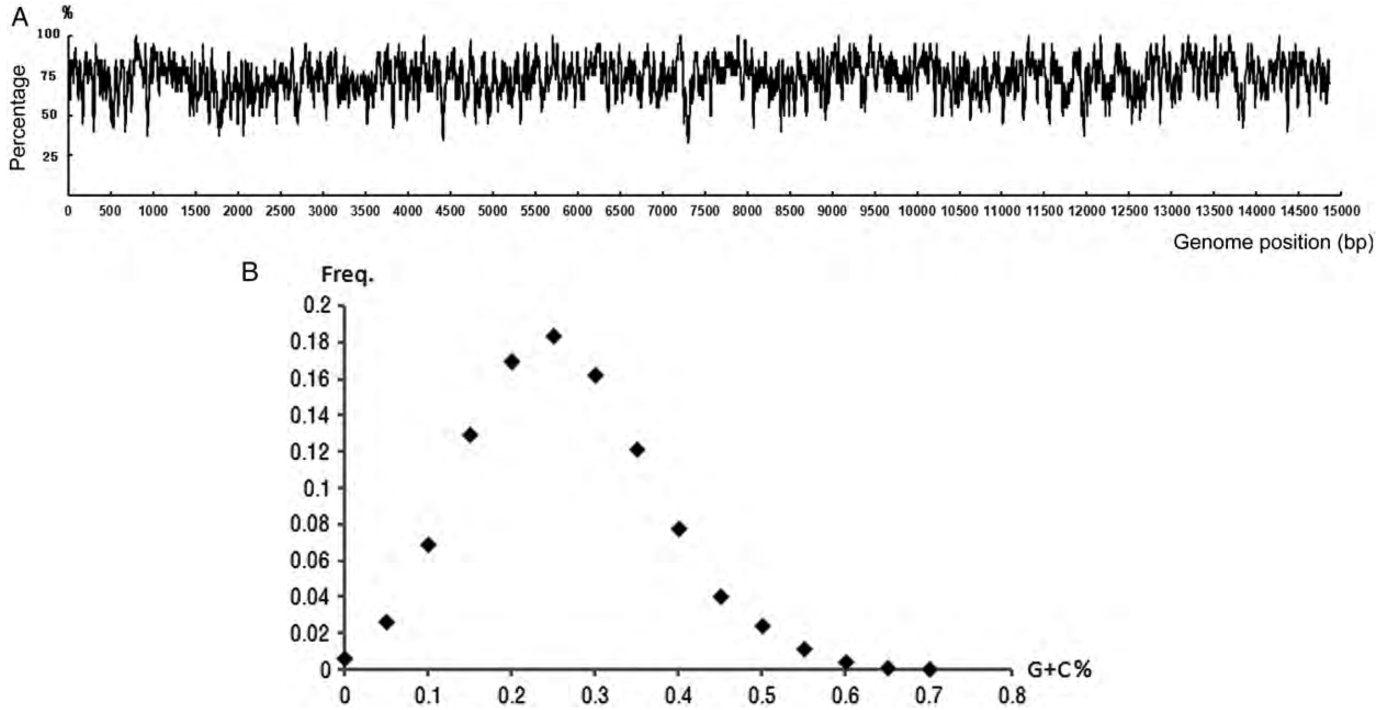

Figure 2 Sliding-window analysis (A) The nucleotide compositional fluctuation along the genome. (B) GC content distribution. The window size is 20 bp.

and T content of the putative control region is 74.5% which is slightly higher than that in total genome (73.8%) and slightly lower than that in N-strand PCGs (75.3%). Thus, we prefer to call this region 'putative control region' rather than 'A + T-rich region' that formally used in many other published insect mitogenomes.

Figure 5 shows the predicted secondary structure of putative control region that contains two adjacent stem-loop structures. The one at 12S rRNA gene side has a considerably long stem (70 bp; marked as '1' in **Fig. 5**) while the one at tRNA-Ile side is considerably shorter (18 bp; marked as '2' in **Fig. 5**).

Discussion

Most insect mitogenomes have a size ranging between 15 and 16 kbps, which is mainly caused by the differences in the non-coding regions [20]. Within them, the lengths of the coding genes is relatively stable while the putative control region that is the largest non-coding region shows extremely significant variations both in length and in patterns. Owing to the short intergenic spaces including the putative control region, the mitogenome of *P. ruficeps* with its 14,841 bp is fairly small compared with other published hemipteran mitogenomes: within this group only the two Sternorrhyncha species, *Neomaskellia andropogonis* and *Pachypsylla venusta* have smaller mitogenomes (14,496 and 14,711 bp, respectively).

No examples of gene rearrangement were found in the mitogenome of *P. ruficeps* or other cicadomorphan species. The gene arrangement, especially the location of the relative larger genes that encode proteins and rRNAs are conservative in most insects [2] and *P. ruficeps* share the putative ancestral state as also found in *Drosophila melanogaster* [21]. This is also shared by most other Hemipteran species except for some Sternorrhyncha [22] and Heteroptera [10].

The putative control region is the most variable part of the mitogenome and shows complicated structures. No sequence similarity was found when blasting the putative control region sequences with the GenBank nucleotide records. Its considerably small length of only 310 bp is thereby the most obvious characteristic of the *P. ruficeps* mitogenome. In the only other studied spittlebug *P. spumarius*, it is 1835 bp [11] in length while in the sternorrhynchan *Trialeurodes vaporariorum*, it is 3725 bp [22]. The shortest one within Hemiptera was previously observed in *Neomaskellia andropogonis* [22] with only 328 bp. In general, this region shows a high degree of variability among insects, from 4601 bp in *D. melanogaster* [21] to only 70 bp in the Orthopteran *Ruspolia dubia* [23]. Boyce *et al.* [24] reported that the bark weevil has a putative control region longer than 13 kbp, but so far no sequences are available in GenBank. Apparently, these changes occur on every systematic level, so no phylogenetic statement can be made.

Typically, the putative control region has some or all of the following characters: tandem repeated sequences, a long

Table 3 Relative synonymous codon usage of *Paphnutius ruficeps* mitogenome

Amino acids	Codon	Total	Strand	
			J	N
F	**UUU**	1.77	1.75	1.79
	UUC	0.23	0.25	0.21
L	**UUA**	3.52	3.47	3.58
	UUG	1.14	1.12	1.17
	CUU	0.64	0.73	0.52
	CUC	0.02	0.04	0.00
	CUA	0.53	0.50	0.57
	CUG	0.14	0.13	0.16
I	**AUU**	1.83	1.89	1.71
	AUC	0.17	0.11	0.29
M	**AUA**	1.42	1.33	1.56
	AUG	0.58	0.67	0.44
V	**GUU**	1.77	1.96	1.41
	GUC	0.14	0.12	0.17
	GUA	1.65	1.57	1.80
	GUG	0.45	0.36	0.62
S	**UCU**	2.17	2.19	2.15
	UCC	0.48	0.41	0.59
	UCA	2.15	2.19	2.10
	UCG	0.34	0.37	0.29
P	**CCU**	1.97	1.91	2.13
	CCC	0.33	0.36	0.25
	CCA	1.47	1.45	1.50
	CCG	0.23	0.27	0.13
T	**ACU**	1.99	2.06	1.79
	ACC	0.34	0.23	0.63
	ACA	1.48	1.52	1.37
	ACG	0.20	0.19	0.21
A	**GCU**	2.08	2.10	2.05
	GCC	0.20	0.15	0.29
	GCA	1.45	1.54	1.27
	GCG	0.27	0.21	0.39
Y	**UAU**	1.57	1.66	1.47
	UAC	0.43	0.34	0.53
H	**CAU**	1.62	1.54	1.88
	CAC	0.38	0.46	0.13
Q	**CAA**	1.43	1.50	1.29
	CAG	0.57	0.50	0.71
N	**AAU**	1.67	1.63	1.73
	AAC	0.33	0.37	0.27
K	**AAA**	1.39	1.19	1.76
	AAG	0.61	0.81	0.24
D	**GAU**	1.54	1.50	1.61
	GAC	0.46	0.50	0.39
E	**GAA**	1.43	1.35	1.60
	GAG	0.57	0.65	0.40
C	**UGU**	1.57	1.71	1.52
	UGC	0.43	0.29	0.48
W	**UGA**	1.48	1.46	1.54
	UGG	0.52	0.54	0.46
R	CGU	1.25	1.38	1.05
	CGC	0.08	0.13	0.00
	CGA	2.20	2.00	2.53
	CGG	0.47	0.50	0.42
S	AGU	0.88	0.98	0.73
	AGC	0.10	0.10	0.10
	AGA	1.64	1.42	1.95
	AGG	0.24	0.34	0.10
G	GGU	1.38	1.38	1.37
	GGC	0.20	0.21	0.20
	GGA	1.58	1.53	1.66
	GGG	0.84	0.88	0.78
Stop	**UAA**	1.33	1.33	1.33
	UAG	0.67	0.67	0.67

Codons in bold type indicate the most commonly used codons for each amino acid.

sequences of T, a subregion with extremely high A and T content, and stem-loop structures [5,25]. However, unlike the putative control region of the spittlebug *P. spumarius* [11] which has all the above mentioned characteristics, none of them except the stem-loop structures can be found in *P. ruficeps*. Thus, the comparatively simple structure of the putative control region can be considered as another characteristic of *P. ruficeps*. Tandem repeat sequences have been reported in many insect mitogenomes [5]. The other two known Cicadomorpha species, *P. spumarius* [11] and *Homalodisca vitripennis* (unpublished, NC_006899.1) have such repeat segments. Therefore, we conclude that their absence might be a derived character of *P. ruficeps*. However, such sequences are also absent in some species of Orthoptera, Coleoptera, and Diptera [5], so it appears as if there reduction is highly homoplasious.

The biological function of the putative control region is still an unanswered question [5]. It has been shown that it plays a role as the starting point of the mitogenome replication as in the fruit fly [26]. However, based merely on sequence data, we cannot determine whether it has the same function in *P. ruficeps* as in fruit fly.

The A and T content of the putative control region of the *P. ruficeps* mitogenome is only slightly higher than the average level of the whole genome, but lower than the N-strand PCGs. The putative control region is often called A + T-rich region in many published mitogenomes due to

Table 4 Nucleotide composition at different codon positions

Genes	Total				Codon position 1st				Codon position 2nd				Codon position 3rd			
	A	T	C	G	A	T	C	G	A	T	C	G	A	T	C	G
nad2	27.5	47.2	9.1	16.2	39.1	38.2	7.3	15.5	17.6	51.2	16.4	14.8	25.8	52.3	3.6	18.2
cox1	26.1	42.0	13.4	18.4	30.9	30.9	13.1	25.2	17.8	43.6	21.5	17.0	29.5	51.7	5.7	13.1
cox2	30.8	39.7	12.2	17.4	32.9	29.3	15.6	22.2	25.4	42.0	16.5	16.1	33.9	47.8	4.5	13.8
atp8	30.1	44.4	11.8	13.7	35.3	43.1	7.8	13.7	27.5	39.2	19.6	13.7	27.5	51.0	7.8	13.7
atp6	28.8	44.7	11.6	14.9	36.0	32.9	12.6	18.5	17.6	52.7	17.6	12.2	32.9	48.6	4.5	14.0
cox3	26.8	42.9	13.1	17.3	27.2	36.0	14.9	21.8	20.8	41.9	18.8	18.5	32.3	50.8	5.4	11.5
nad3	31.6	44.6	8.8	15.0	39.0	33.9	9.3	17.8	19.5	56.8	12.7	11.0	36.4	43.2	4.2	16.1
nad6	40.5	38.0	12.7	8.8	51.5	29.8	7.0	11.7	23.4	49.1	18.7	8.8	46.8	35.1	12.3	5.8
cytb	35.8	36.5	15.2	12.5	37.8	28.6	14.0	19.6	21.2	44.7	20.6	13.5	48.5	36.1	10.9	4.5
J-genes	**30.9**	**42.2**	**12.0**	**14.9**	**36.6**	**33.6**	**11.3**	**18.4**	**21.2**	**46.8**	**18.0**	**14.0**	**34.8**	**46.3**	**6.5**	**12.3**
nad5	32.6	43.0	11.7	12.7	36.3	37.9	8.8	17.0	19.8	49.8	17.5	12.8	41.6	41.4	8.8	8.2
nad4	29.1	44.8	11.5	14.5	30.2	44.1	9.5	16.1	18.5	49.9	14.4	17.3	38.7	40.5	10.7	10.0
nad4l	26.7	52.4	7.3	13.5	24.0	51.0	9.4	15.6	19.8	55.2	8.3	16.7	36.5	51.0	4.2	8.3
nad1	24.7	47.7	9.7	17.9	25.5	42.5	10.1	21.9	19.6	48.0	14.7	17.6	29.1	52.6	4.2	14.1
N-genes	**28.3**	**47.0**	**10.1**	**14.7**	**29.0**	**43.9**	**9.5**	**17.7**	**19.4**	**50.7**	**13.7**	**16.1**	**36.5**	**46.4**	**7.0**	**10.2**
Total	30.1	43.7	11.4	14.8	34.3	36.8	10.7	18.2	20.7	48.0	16.7	14.6	35.3	46.3	6.7	11.6

J-genes and N-genes means the gene located at J-strand and N-strand, respectively. Bolded values are the averages of J-strand and N-strand genes. Names of the genes that are located on the N-strand were underlined.

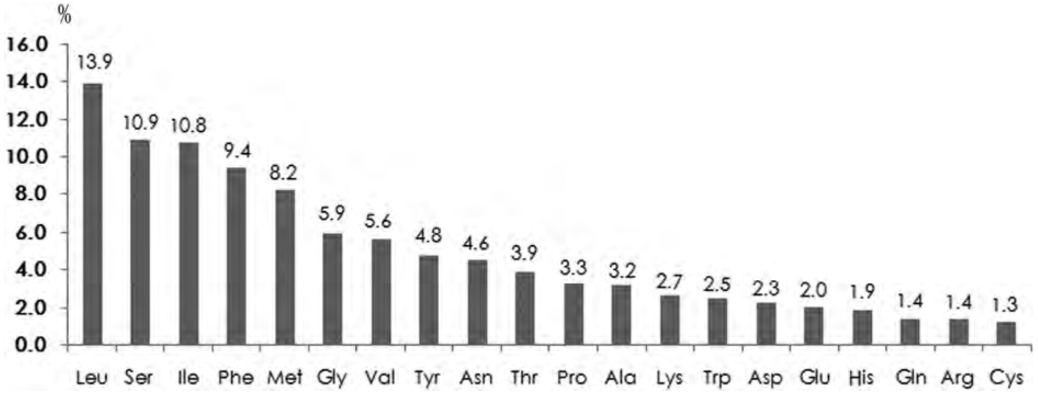

Figure 3 Amino acid content of proteins coded by *Paphnutius ruficeps* mitochondrial genome The amino acids are shown by standard abbreviations.

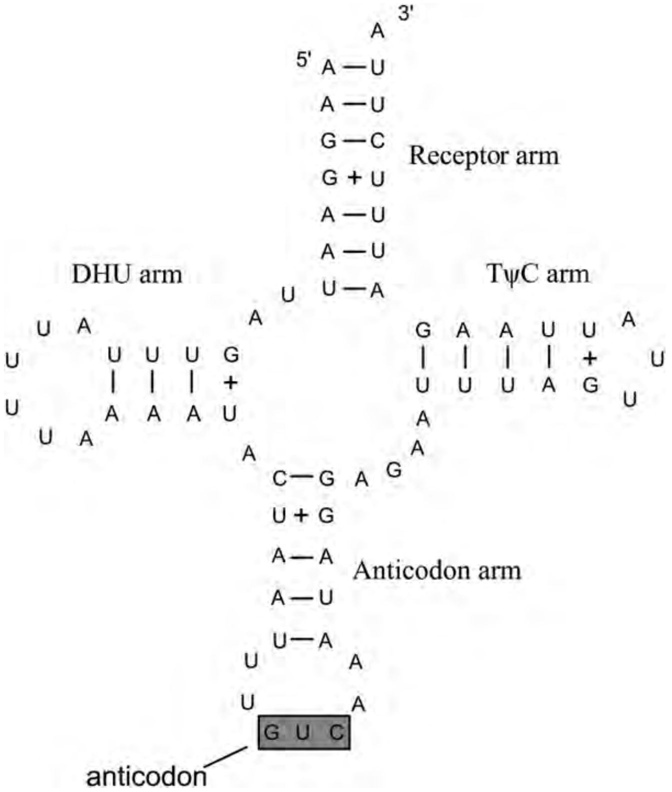

Figure 4 putative clover-leaf secondary structure of $tRNA^{Asp}$ The tRNA is labeled with standard abbreviation of the corresponding amino acid. Waston-Crick pairs are marked with '−' and the non-Waston-Crick pairs (G–U pairs) with '+'.

its comparatively high A and T content [5]. However, some researchers argued that the 'A + T rich' is not a conservative character and the name of A + T-rich region was not suitable for certain species [10]. The mitogenome of *P. ruficeps* is an example that supports this suggestion.

Compared with other hemipteran mitogenomes, the putative control region of *P. ruficeps* is special in the aspects of nucleotide composition, length, and pattern. However, this region exists in all published insect's mitogenomes and its existence itself might be considered as a conservative feature. It is still difficult to understand the important roles played by such variable sequences [27]. This question may be essential to understand the organizational mechanism of the genome, since it represents a complete heredity system despite its small scale. More detailed research including both bioinformatics and experimental works will be needed to answer this question. Experimental research focusing on comparatively 'simple' putative control regions, such as the one of *P. ruficeps*, may be helpful for revealing the exact biological function of this region.

Overlapped genes and none-coding intervals except the putative control region

Like other published animal mitogenomes, the one of *P. ruficeps* is compact and the overlapping genes reflect the high efficiency of nucleotides usages [2]. In insects, two kinds of overlaps between mitochondrial genes can be observed [28]: the first one occurs either between tRNAs or between tRNAs and rRNAs. These overlaps are processed through post-transcription processing [2]. The second one occurs between PCGs. Three such overlapping regions are identified in the mitogenome of *P. ruficeps*: between *atp8* and *atp6*, *nad4* and *nad4l*, *nad6* and *cytb,* respectively. Interestingly, all three of them are between a relative big (*atp6*, *nad4,* and *cytb*) and a small PCG (*atp8*, *nad4l,* and *nad6*). They also appear in the mitogenome of *P. spumarius* [11] and other hemipteran species such as *Neomaskellia andropogonis* (Sternorrhyncha) [22] or *Geisha distinctissima* (Fulgoromorpha) [29]. It is a characteristic of these overlaps that the short open reading frames always appear at the 5′ side of larger ones. Additionally, these overlaps always locate at the 5′ end of the transcription and it is assumed that they might improve translation efficiency [30].

Nucleotide and codon usage of the protein coding genes

As in most published insect mitogenomes [2], the nucleotide usage of the different codon positions shows a bias toward A and T. However, the detailed situation is different

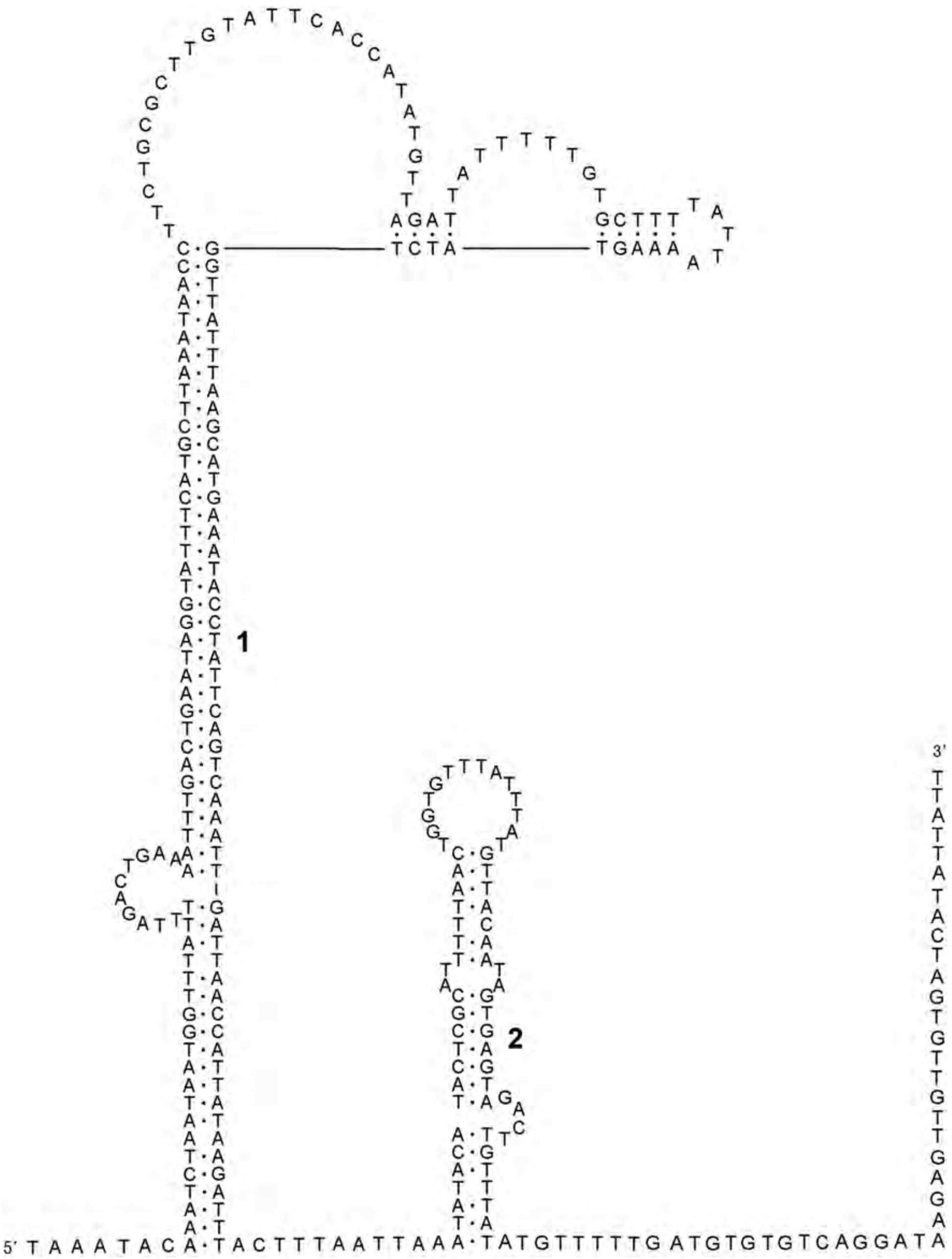

Figure 5 putative secondary structure of the putative control region

among the three codon positions. The second codon position contains the highest G and C content, while the third one has the lowest one. According to the degeneracy of the invertebrate mitochondrial genetic codes, it can be observed that any changes of the nucleotide at the second codon position and most changes of the first codon position will alter the coded amino acid. On the other hand, most changes of the third codon position have no influence on the amino acid. Therefore, the first and second codon positions are under more evolutionary pressures than the third one. In general, it can be assumed that the higher the G and C content in a codon position is, the more conservative it is.

Most PCGs (7 of 13) in the mitogenome of *P. ruficeps* use ATG as start codon. The one with the highest variability is *cox1*, for it has been reported to use irregular start codon such as the four-based codons ATAA, GTAA, and TTAA or the six-based ATTTAA within arthropods [31–34]. However, within Auchenorrhyncha including the two sequenced mitogenomes of spittlebugs [11], *cox1* uses a three based start codon, which can be considered as a potential apomorphy of this clade. Incomplete stop codons are very common among metazoan mitogenomes and it is assumed that they are completed by adding A during the post-transcription processing [35].

The mitochondrial genome of the spittlebug species *P. ruficeps* was sequenced. Its length (14,841 bp) is relatively short compared with other published insect mitogenomes due to a short and structurally simplified putative control region. Tandem repeat sequences that normally exists in many insect mitogenomes are absent in *P. ruficeps*. In summary, the studied mitogenome has 13 PCGs, 2 rRNA genes, and 22 tRNA genes and its gene content and arrangement are identical to the putative ancestral insect state. The genome shows an obvious nucleotide bias toward A and T. All of the 13 PCGs use normal triplet start codons (ATG, TTG, ATT, and ATA). Nine of the PCGs use complete stop codon such as TAA or TAG, while the remaining four use incomplete stop codons. The analyses of the nucleotide bias indicated that the G and C content is positively correlated with the degenerate strictness of different codon positions. The present results expand our knowledge of insect mitochondrial genomes, especially for the hemipteran linage of Cicadomorpha.

Acknowledgement

We would like to thank Dr Ruie Nie (Institute of Zoology, Chinese Academy of Sciences, Beijing, China) for her assistance in sample collection.

Funding

This work was supported by the grants from the National Basic Research Program of China (No. 2011CB302102) and the National Natural Science Foundation of China (Nos. 30970400, 31172128 and 30530110), both awarded to A.P.L.

References

1 Hatefi Y. The mitochondrial electron transport and oxidative phosphorylation system. Annu Rev Biochem 1985, 54: 1015–1069.

2 Boore JL. Animal mitochondrial genomes. Nucleic Acids Res 1999, 27: 1767–1780.

3 Salvato P, Simonato M, Battisti A and Negrisolo E. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). BMC Genomics 2008, 9: 331.

4 Taanman JW. The mitochondrial genome: structure, transcription, translation and replication. Biochim Biophys Acta 1999, 1410: 103–123.

5 Zhang DX, Szymura JM and Hewitt GM. Evolution and structural conservation of the control region of insect mitochondrial DNA. J Mol Evol 1995, 40: 382–391.

6 Kristensen NP. Phylogeny of extant hexapods. In: C.S.I.R. Organization ed. The Insects of Australia, a Textbook for Students and Research Workers, 2nd edn. Victoria: Melbourne University Press, 1991, 125–142.

7 Carpenter FM. Treatise on Invertebrate Paleontology. Vol 3, Superclass Hexapoda. Boulder, Colorado and Lawrence, Kansas: The Geological Society of America and the University of Kansas, 1992.

8 Liang AP. A proposal to stop using the insect order name ‘Homoptera’. Chinese Bull Entomol 2005, 42: 332–337.

9 Holmann F and PeckEconomic DC. Damage caused by spittlebugs (Homoptera: Cercopidae) in Colombia: a first approximation of impact on animal production in *Brachiaria decumbens* Pastures. Neotrop Entomol 2002, 31: 275–284.

10 Hua J, Li M, Dong P, Cui Y, Xie Q and Bu W. Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). BMC Genomics 2008, 9: 610.

11 Stewart JB and Beckenbach AT. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). Genome 2005, 48: 46–54.

12 Simon C, Buckley TR, Frati F, Stewart JB and Beckenbach AT. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu Rev Ecol Evol Syst 2006, 37: 545–579.

13 Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 1999, 41: 95–98.

14 Tamura K, Dudley J, Nei M and Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 2007, 24: 1596–1599.

15 Lowe TM and Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997, 25: 955–964.

16 Clary DO and Wolstenholme DR. The mitochondrial DNA molecule of *Drosophila* yakuba: nucleotide sequence, gene organization, and genetic code. J Mol Evol 1985, 22: 252–271.

17 Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003, 31: 3406–3415.

18 Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 1999, 27: 573–580.

19 Duret L and Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proc Natl Acad Sci USA 1999, 96: 4482–4487.

20 Beard CB, Hamm Mills D and Collins FH. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol Biol 1993, 2: 103–124.

21 Lewis OL, Farr CL and Kaguni LS. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Mol Biol 1995, 4: 263–278.

22 Thao ML, Baumann L and Baumann P. Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). BMC Evol Biol, 2004, 4: 25.

23 Zhou Z, Huang Y and Shi F. The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Conocephalidae) contains a short A + T-rich region of 70 bp in length. Genome 2007, 50: 855–866.

24 Boyce TM, Zwick ME and Aquadro CF. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics 1989, 123: 825–836.

25 Cook CE. The complete mitochondrial genome of the stomatopod crustacean *Squilla mantis*. BMC Genomics 2005, 6: 105.

26 Saito S, Tamura K and Aotsuka T. Replication origin of mitochondrial DNA in insects. Genetics 2005, 171: 1695–1705.

27 Zhang D and Hewitt GM. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol 1997, 25: 99–120.

28 Kim MJ, Wan XL, Kim KG, Hwang JS and Kim I. Complete nucleotide sequence and organization of the mitogenome of endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae). Afr J Biotechnol 2010, 9: 735–754.

29 Song N and Liang A. The complete mitochondrial genome sequence of *Geisha distinctissima* (Hemiptera: Flatidae) and comparison with other hemipteran insects. Acta Biochim Biophys Sin 2009, 41: 206–216.

30 Berthier F, Renaud M, Alziari S and Durand R. RNA mapping on Drosophila mitochondrial DNA: precursors and template strands. Nucleic Acids Res 1986, 14: 4519–4533.

31 Yukuhiro K, Sezutsu H, Itoh M, Shimizu K and Banno Y. Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. Mol Biol Evol 2002, 19: 1385–1389.

32 Ballard JW. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. J Mol Evol 2000, 51: 48–63.

33 Clary DO and Wolstenholme DR. Genes for cytochrome c oxidase subunit I, URF2, and three tRNAs in *Drosophila* mitochondrial DNA. Nucleic Acids Res 1983, 11: 6859–6872.

34 de Bruijn MH. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. Nature 1983, 304: 234–241.

35 Gadaleta G, Pepe G, de Candia G, Quagliariello C, Sbisa E and Saccone C. The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J Mol Evol 1989, 28: 497–516.