GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Complete sequence of the mitochondrial genome of the Japanese buff-tip moth, *Phalera flavescens* (Lepidoptera: Notodontidae)

Q.-Q. Sun[1], X.-Y. Sun[2], X.-C. Wang[1], Y.-H. Gai[2], J. Hu[1], C.-D. Zhu[3] and J.-S. Hao[1]

[1]Laboratory of Molecular Evolution and Biodiversity,
College of Life Sciences, Anhui Normal University, Wuhu, P.R. China
[2]Key Laboratory of Palaeobiology and Stratigraphy,
Nanjing Institute of Geology and Palaeontology,
Chinese Academy of Sciences, Nanjing, P.R. China
[3]Key Laboratory of Zoological Evolution and Systematics,
Institute of Zoology, Chinese Academy of Sciences, Beijing, P.R. China

Corresponding author: J.-S. Hao
E-mail: jshaonigpas@sina.com

Genet. Mol. Res. 11 (4): 4213-4225 (2012)
Received November 23, 2011
Accepted June 30, 2012
Published September 10, 2012
DOI http://dx.doi.org/10.4238/2012.September.10.2

ABSTRACT. We sequenced the complete mitochondrial genome of *Phalera flavescens*. The mitogenome is 15,659 bp in length, including 13 protein-coding genes (*atp6*, *atp8*, *cox1-3*, *nad1-6*, *nad4L*, *cob*), two ribosomal RNAs (*rrnS* and *rrnL*), 22 transfer RNAs and an AT-rich region, a putative control region (*D-loop*). Gene order and orientation were found to be identical to those of other completely sequenced lepidopteran mitogenomes. All 13 protein-coding genes start with the common codon ATN, except for the *cox1* gene, which uses CGA as the initial codon. Nine of the 13 protein-coding genes stop with codon TAA, while the *cox1*, *cox2*, *nad5*, and *nad4* genes stop with the single nucleotide T. All tRNA genes can be folded into canonical cloverleaf secondary structure, except for *trnS1*, which loses the ‘‘DHU’’ arm. Six overlapping sequences totaling 20 bp (1-8 bp for each sequence)

and 16 intergenic spacer sequences, totaling 276 bp (1-58 bp for each sequence) are scattered throughout the genome; the largest intergenic spacer is located between the *trnQ* and *nad2* genes. A microsatellite-like structure $(AT)_6ACC(AT)_6$ and 16-bp poly-T elements preceded by the ATTTA motif are present in the *D-loop* region. Additionally, unexpectedly, an extra 190-bp insertion, with unknown function, was found in the small subunit rRNA gene (*rrnS*); this gene is the longest known (1020 bp) among all of the Lepidoptera.

Key words: Notodontidae; *Phalera flavescens*; Mitochondrial genome; *rrnS*; 190-bp insertion

INTRODUCTION

Mitochondrial genome (mtDNA) is a small molecule of 15-20 kb in length in most metazoans, which plays an important role in the process of metabolism, programmed cell death, illness, aging, etc. It is generally circular with a remarkably conserved set of 37 genes: 13 protein-coding genes (PCGs) (*nad1-6* and *nad4L* for NADH dehydrogenase subunits 1-6, and 4L; *atp6* and *atp8* for ATPase subunits 6 and 8; *cox1-3* for cytochrome oxidase subunits I-III; *cob* for cytochrome b), *rrnS* and *rrnL* for 12S rRNA and 16S rRNA, and 22 tRNAs (Boore, 1999). In addition, it usually contains at least one variable sequence about 1 kb in size, known as the AT-rich region (*D-loop*) (Zhang et al., 1995; Zhang and Hewitt, 1997). During the last few decades, mitochondrial genome sequence data, especially those of PCGs have been employed as a powerful tool for phylogenetics and evolutionary studies.

The order Lepidoptera is one of the largest insect orders, including more than 160,000 described species. Despite this huge species diversity, to date, only about 27 mtDNAs have been fully sequenced, among which only one is of the Notodontidae (Salvato et al., 2008).

The Japanese buff-tip moth *Phalera flavescens* is a member of the family Notodontidae and an important species of insect pests, which is harmful to many economic plants, such as *Malus pumila*, *Pyrus bretschneideri*, *Amygdalus persica*, *Prunus salicina*, and others. This commonly found species is distributed throughout all areas of China and some areas of Japan and Korea. In this study, the complete mitochondrial genome sequence of *P. flavescens* was determined through PCR and primer-walking methods, and described in comparison with that of other sequenced lepidopterans in terms of genome organization, the characteristics of individual tRNAs, rRNAs, PCGs, and the non-coding regions, including the AT-rich region. Thus, the newly sequenced *P. flavescens* mitogenome is very useful to enrich our knowledge about lepidopteran mitogenomes and to supply more information for the studies of relevant areas.

MATERIAL AND METHODS

Specimen collection

The adult specimen of *P. flavescens* (Lepidoptera: Notodontidae) was obtained from Changping District of Beijing city, China, in August 2010. After the specimen collection, the material was placed in 100% ethanol immediately for preservation and fixation, brought to our laboratory and stored at -20°C until use for genomic DNA extraction.

DNA extraction, primer design and PCR amplification

Total genomic DNA of *P. flavescens* was extracted from thoracic muscle tissue of the adult specimen using the proteinase K-SOi_2 method of Hao et al. (2005). Some partial short sequences of *cox1*, *cox2*, *nad5*, *cob*, *nad4*, *rrnL*, and *rrnS* were amplified by the PCR protocol: denaturation for 5 min at 95°C, followed by 35 cycles of denaturation for 50 s at 95°C, annealing for 50 s at 55°C, and elongation for 1 min at 72°C, and a final extension step at 72°C for 10 min. The PCR products were then sequenced using the primer set for the complete mitochondrial genomes of all the lepidopterans available or reported in Simon et al. (1994). The amplification of the six long fragments (*rrnS-nad2*, *nad2-cox3*, *cox3-nad5*, *nad5-nad4*, *nad4-cob*, *cob-rrn*S) (Table 1) were performed using TaKaRa LA Taq™ (Takara Bio, Japan) by the long PCR conditions: an initial denaturation for 5 min at 95°C, followed by 15 cycles of denaturation for 50 s at 95°C, annealing for 60 s at 48°-58°C, elongation for 150 s at 68°C; then another 15 cycles of denaturation for 50 s at 95°C, annealing for 60 s at 48°-58°C, and elongation for 150 s per cycle at 68°C, and a final extension at 68°C for 10 min. Two primers 5'-GAAACACTTTCCAGTACCT-3' and 5'-CTAAACCAATTCAACATCC-3' were designed for amplifying the *D-loop* region between the *rrn*S and *nad*2 genes through the comparison of all lepidopteran mitochondrial genomes available.

Table 1. Primers for the long PCR amplication used in this study.

Primers	Upper primer sequence (5'-3')	Lower primer sequence (5'-3')
nad2-cox3	CCCCTCTTTCTTCTAATA	TTGTATGTTTACCTTGGA
cox3-nad5	AAAGGATTACGATGAGGT	TCCAGTTAAGGGTCAAGGACTAT
nad5-nad4	TAAAAAAGGAATCCCACA	GCGTTTACATAGACGAAGA
nad4-cob	CGTCTATGTAAACGCTCA	ATAAGGGTTTTCTACTGGT
cob-rrnS	CGTGTTATTACTTTACTTG	AAACTAGGATTAGATACCC
rrnS-nad2	GAAACACTTTCCAGTACCT	CTAAACCAATTCAACATCC

Sequence analysis

Raw sequence files were proofread and assembled in BioEdit version 7.0 (Hall, 1999). Sequences overlapping fragments were assembled by aligning neighboring fragments using the Clustal X1.83 software (Thompson et al., 1997). The 22 tRNA genes were identified by the tRNAscan-SE v.1.21 software (http://lowelab.ucsc.edu/tRNAscan-SE) (Lowe and Eddy, 1997). Nucleotide composition and codon usage were calculated using the MEGA4.0 software (Kumar et al., 2004). The base composition characteristics showing the relative number of A to T as AT skew (AT skew = [A-T]/[A+T]) and G to C as GC skew (GC skew = [G-C]/[G+C]) were determined as proposed by Perna and Kocher (1995). Protein-coding genes and RNA genes were identified by sequence comparison with the published insect mitogenome sequences using SEQUIN (version 11.0). The mitogenome sequence data have been deposited in the GenBank database under the accession No. JF440342. The secondary structures of *rrnL* and *rrnS* were drawn using the XRNA software (http://rna.ucsc.edu/rnacenter/xrna/xrna.html).

RESULTS AND DISCUSSION

Genome organization

The *P. flavescens* mitochondrial genome was 15,659 bp in length (Figure 1 and Table

2); the size is well within the range of the other lepidopteran mitochondrial genomes from 15,140 bp in *Artogeia melete* (Hong et al., 2009) to 16,094 bp in *Papilio maraho* (Feng et al., 2010). It had the same gene content observed in typical metazoan mitogenomes [13 PCGs, two rRNA (*rrnS* and *rrnL*), 22 tRNA genes and a *D-loop* region]. The mitogenome had the same gene order and orientation as the completely sequenced lepidopteran mitogenomes (Kim et al., 2010), with its major strand coding for nine PCGs and 14 tRNAs, while the minor strand coding for four PCGs, eight tRNAs and two rRNAs. The overall base compositions of the mitogenome L-strand was 40.00% A, 40.82% T, 7.90% G, and 11.28% C, showing a relatively strong AT bias (80.82%). This AT bias value is within the range from 77.9% in *Ochrogaster lunifer* to 82.7% in *Coreana raphaelis* (Kim et al., 2006; Salvato et al., 2008) as well. Additionally, the major strand AT skew was slightly negative (-0.01017), which is also within the range of the other lepidopterans, from -0.04748 in *C. raphaelis* (Kim et al., 2006) to 0.05872 in *Bombyx mori* (Yukuhiro et al., 2002).

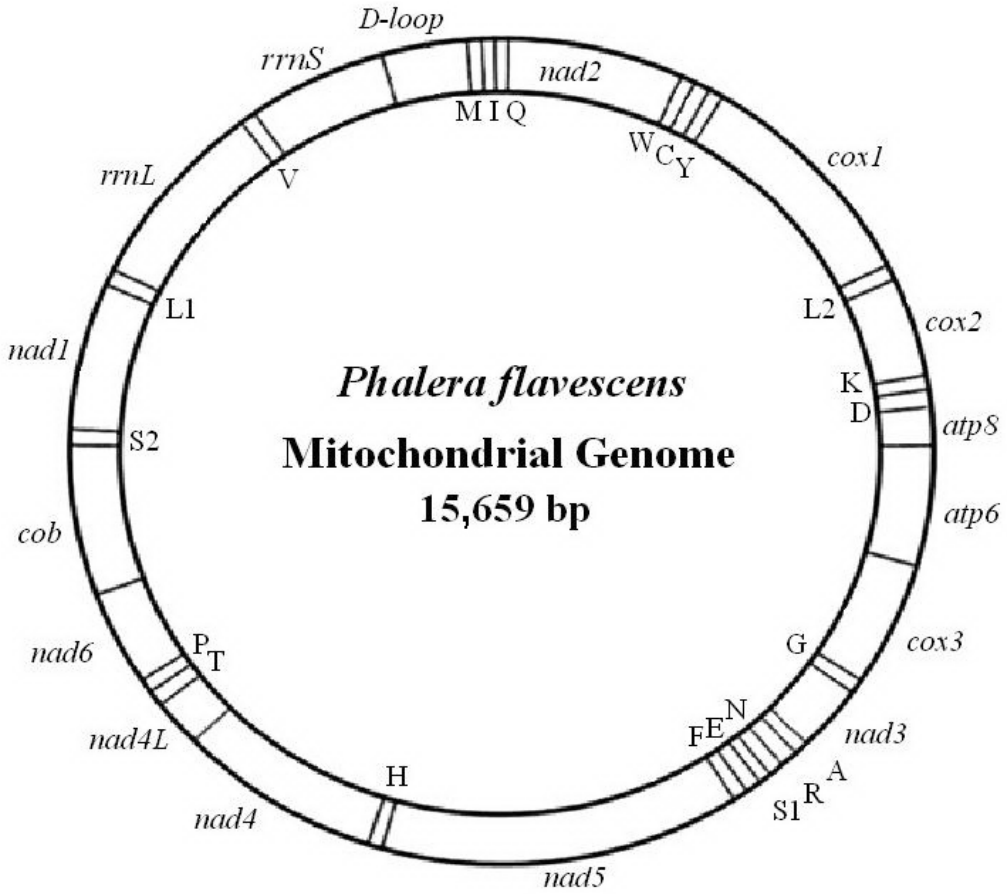

Figure 1. Map of the mitochondrial genome of *Phalera flavescens*. tRNAs are denoted as one-letter symbol according to the IUPAC-IUB single-letter amino acid codes. *cox1-3* = cytochrome oxidase subunits; *cob* = cytochrome b; *nad1-6* = NADH dehydrogenase components.

Table 2. Summary of the *Phalera flavescense* mitogenome.

Gene	Direction	Nucleotide No.	Size (bp)	Intergenic nucleotides	Start codon	Stop codon
trnM	F	1-68	68	0		
trnI	F	69-135	67	0		
trnQ	R	133-201	69	-3		
nad2	F	260-1273	1014	58	ATT	TAA
trnW	F	1289-1360	72	15		
trnC	R	1353-1421	69	-8		
trnY	R	1423-1488	66	1		
cox1	F	1493-3024	1532	5	CGA	T-tRNA
trnL2	F	3025-3091	67	0		
cox2	F	3092-3773	682	0	ATG	T-tRNA
trnK	F	3774-3844	71	0		
trnD	F	3863-3932	70	18		
atp8	F	3933-4091	159	0	ATT	TAA
atp6	F	4088-4762	675	-4	ATG	TAA
cox3	F	4762-5550	789	-1	ATG	TAA
trnG	F	5553-5618	66	2		
nad3	F	5616-5972	357	-3	ATA	TAA
trnA	F	5990-6054	65	17		
trnR	F	6055-6119	65	0		
trnN	F	6126-6191	66	6		
trnS1	F	6199-6267	69	7		
trnE	F	6299-6367	69	31		
trnF	R	6384-6450	67	16		
nad5	R	6451-8188	1738	0	ATT	T-tRNA
trnH	R	8189-8254	66	0		
nad4	R	8255-9632	1378	0	ATA	T-tRNA
nad4L	R	9656-9955	300	23	ATA	TAA
trnT	F	9961-10026	66	5		
trnP	R	10027-10091	65	0		
nad6	F	10097-10585	489	5	ATG	TAA
cob	F	10586-11752	1167	0	ATT	TAA
trnS2	F	11752-11815	64	-1		
nad1	R	11846-12781	936	30	ATG	TAA
trnL1	R	12783-12852	70	1		
rrnL	R	12853-14220	1368	0		
trnV	R	14221-14288	68	0		
rrnS	R	14289-15308	1020	0		
D-loop		15309-15659	351	0		

F = foward; R = reverse.

Protein-coding genes

Like other lepidopterans, all 13 PCGs in the *P. flavescens* mitogenome had ATN as their start codon, except for *cox1*, and they typically showed TAG or TAA as stop codons, except for *cox1*, *cox2*, *nad4*, and *nad5* (T-tRNA) (Table 2).

The start codon of the *cox1* gene has been widely discussed in the insect mtDNA (Yukuhiro et al., 2002). For example, some unusual oligonucleotides, such as the TTA in *Panulirus japonicus* (Yamauchi et al., 2002), TCG in *Bactrocera oleae* (Nardi et al., 2003), ACG in *Vargula hilgendorfii* (Ogoh and Ohmiya, 2004), TATCTA in *Penaeis monodon* (Wilson et al., 2000), TTTTAG in *Bombyx mandarina* (Yukuhiro et al., 2002), and TATTAG in *Ostrinia nubilalis* and *O. furnicalis* (Coates et al., 2005), have been proposed as the *cox1* start codons. As for the start codons for the lepidopteran *cox1* genes, they are also controversial. For example, some unusual start codons are also proposed for three butterfly species (*A. melete*, *Parnassius bremeri* and *Eumenis autonue*) and seven moth species (*Antheraea pemyi*, *Adoxophyes honmai*, *Ochrogaster lunifer*, *Eriogyna pyretorum*, *B. mandarina*, *Lymantria dispar*,

and *Manduca sexta*) (Salvato et al., 2008; Kim et al., 2009; Kim et al., 2010). Owing to the fact that the CGA is present as a conserved region for all lepidopteran insects reported, we tend to consider that CGA is the start codon for *P. flavescens cox1* in this study as suggested by Kim et al. (2009).

The *cox2* gene harbored an incomplete stop codon of a single thymine with no canonical TAN stop codon detected in its terminating region. Such an incomplete stop codon is also detected in 29 other insect mitogenomes available (Cha et al., 2007; Hong et al., 2008; Kim et al., 2010) (Table 3). This single thymine as an incomplete codon detected in PCGs may be caused by posttranscriptional modifications during the mRNA maturation process, such as poly-adenylation. Summarizing, a total of 3727 non-stop codons are used in *P. flavescens* mitogenome PCGs, and this pattern is similar to that in other lepidopteran species as well.

Table 3. Lepidopteran mitogenomes used in this study.

Species	Supperfamily/Family	GenBank accession No.	Reference
Hyphantria cunea	Noctuoidea Arctiidae	NC_014058	Liao et al., 2010
Lymantria dispar	Lymantriidae	NC_012893	Zhu et al., 2010
Helicoverpa armigera	Noctuidae	NC_014668	Yin et al., 2010
Ochrogaster lunifer	Notodontidae	NC_011128	Salvato et al., 2008
Phalera flavescens	Notodontidae	JF440342	This study
Phthonandria atrilineata	Geometroidea Geometridae	NC_010522	Yang et al., 2009
Spilonota lechriaspis	Tortricoidea Tortricidae	NC_014294	Zhao et al., 2011
Adoxophyes honmai	Tortricidae	NC_008141	Lee et al., 2006
Grapholitha molesta	Tortricidae	NC_014806	Gong et al., 2012
Antheraea pernyi	Bombycoidea Saturniidae	NC_004622	Liu et al., 2008
Antheraea yamamai	Saturniidae	NC_012739	Kim SR et al., 2009
Eriogyna pyretorum	Saturniidae	NC_012727	Jiang et al., 2009
Saturnia boisduvalii	Saturniidae	NC_010613	Hong et al., 2008
Chinese Bombyx mandarina	Bombycidae	AY301620	Pan et al., 2008
Bombyx mandarina	Bombycidae	NC_003395	Yukuhiro et al., 2002
Bombyx mori	Bombycidae	NC_002355	Yukuhiro et al., 2002
Manduca sexta	Sphingidae	NC_010266	Cameron and Whiting, 2008
Diatraea saccharalis	Pyraloidea Crambidae	NC_013274	Li et al., 2011
Ostrinia furnacalis	Crambidae	NC_003368	Coates et al., 2005
Ostrinia nubilalis	Crambidae	NC_003367	Coates et al., 2005
Eumenis autonoe	Papilionoidea Nymphalidae	NC_014587	Kim et al., 2010
Acraea issoria	Nymphalidae	GQ376195	Hu et al., 2010
Sasakia charonda	Nymphalidae	NC_014224	Hakozaki et al., unpublished results[a]
Sasakia charonda kuriyamaensis	Nymphalidae	NC_014223	Hakozaki et al., unpublished results[a]
Argyreus hyperbius	Nymphalidae	JF439070	Wang et al., 2011
Papilio maraho	Papilionidae	NC_014055	Feng et al., 2010
Parnassius bremeri	Papilionidae	NC_014053	Kim et al., 2009
Teinopalpus aureus	Papilionidae	NC_014398	Qin et al., 2012
Coreana raphaelis	Lycaenidae	NC_007976	Kim et al., 2006
Artogeia melete	Pieridae	NC_010568	Hong et al., 2009

[a]Hakozaki Y, Ueda J and Sato M.

The codon usage of PCGs of the *P. flavescens* mitochondrial genome is shown in Table 4. The results indicate that the codon usage of all the genes has a strong bias in that the relative synonymous codon usage values of NNU and NNA codons are essentially greater than the others, indicating a higher U+A frequency of the third points compared to the others. The PCG codon usage bias and the third point of A+T bias (92.4%) show a positive correlation. In addition, statistics also showed that UUU (Phe), UUA (Leu), AUU (Ile), AUA (Met), and AAU (Asn) are the most frequently used codons, accounting for 46.85% of all.

Table 4. Codon usage of *Phalera flavescens* mitochondrial genome.

Codon(Aa)		Codon(Aa)		Codon(Aa)		Codon(Aa)	
UUU(F)	338.0 (1.86)	UCU(S)	119.0 (3.00)	UAU(Y)	177.0 (1.90)	UGU(C)	35.0 (1.89)
UUC(F)	26.0 (0.14)	UCC(S)	15.0 (0.38)	UAC(Y)	9.0 (0.10)	UGC(C)	2.0 (0.11)
UUA(L)	472.0 (5.00)	UCA(S)	64.0 (1.62)	UAA(*)	0.0 (0.00)	UGA(W)	89.0 (1.84)
UUG(L)	18.0 (0.19)	UCG(S)	1.0 (0.03)	UAG(*)	0.0 (0.00)	UGG(W)	8.0 (0.16)
CUU(L)	44.0 (0.47)	CCU(P)	61.0 (1.91)	CAU(H)	51.0 (1.59)	CGU(R)	22.0 (1.66)
CUC(L)	5.0 (0.05)	CCC(P)	16.0 (0.50)	CAC(H)	13.0 (0.41)	CGC(R)	1.0 (0.08)
CUA(L)	26.0 (0.28)	CCA(P)	48.0 (1.50)	CAA(Q)	59.0 (1.87)	CGA(R)	30.0 (2.26)
CUG(L)	1.0 (0.01)	CCG(P)	3.0 (0.09)	CAG(Q)	4.0 (0.13)	CGG(R)	0.0 (0.00)
AUU(I)	435.0 (1.92)	ACU(T)	75.0 (2.11)	AAU(N)	245.0 (1.90)	AGU(S)	25.0 (0.63)
AUC(I)	18.0 (0.08)	ACC(T)	12.0 (0.34)	AAC(N)	13.0 (0.10)	AGC(S)	2.0 (0.05)
AUA(M)	256.0 (1.87)	ACA(T)	53.0 (1.49)	AAA(K)	104.0 (1.89)	AGA(S)	87.0 (2.20)
AUG(M)	18.0 (0.13)	ACG(T)	2.0 (0.06)	AAG(K)	6.0 (0.11)	AGG(S)	4.0 (0.10)
GUU(V)	66.0 (1.82)	GCU(A)	66.0 (2.08)	GAU(D)	58.0 (1.73)	GGU(G)	54.0 (1.07)
GUC(V)	5.0 (0.14)	GCC(A)	10.0 (0.31)	GAC(D)	9.0 (0.27)	GGC(G)	2.0 (0.04)
GUA(V)	70.0 (1.93)	GCA(A)	50.0 (1.57)	GAA(E)	67.0 (1.81)	GGA(G)	110.0 (2.18)
GUG(V)	4.0 (0.11)	GCG(A)	1.0 (0.03)	GAG(E)	7.0 (0.19)	GGG(G)	36.0 (0.71)

Data are reported as frequency of codon used with relative synonymous codon usage in parentheses. *Stop codon. Start codons and stop codons were excluded in total codons count of protein coding genes.

Ribosomal and transfer RNA genes

In the *P. flavescens* mitogenome, the *rrnL* gene was located between sites 12,853 and 14,220, for a total length of 1368 bp. Its reconstructed secondary structure (Figure 2) broadly conforms to that of *M. sexta* (Cameron and Whiting, 2008). The proposed secondary structure has 49 helices belonging to six domains. The stem regions of H1057/D18, H1087/D19 and H991/D17 in domain II are difficult to fold under the criteria of Watson-Crick pairs, and thus different from *M. sexta* with regard to a large internal loop. The stem H2347/G13 is also highly variable within the order Lepidoptera with regard to sequence variation and the $(AT)_n$ insertion, such as $(AT)_{14}$ in *M. sexta* and $(AT)_{12}$ in *Grapholitha molesta* (Gong et al., 2012). While some bases indentified as highly conserved region, such as the stems H2064/G2, H2077/G3, H2246/G6, H2259/G7, H2282, and their correspondingly large loops are found adjacent to and even within the hypervariable region, these conserved sites may serve as anchor points for the secondary structure of the hypervariable region (Hao et al., 2007).

The *rrnS* gene was located between positions 14,289 and 15,308, for a total 1020 bp in length. Its secondary structure is basically the same as those of the other insect groups with 29 helices belonging to three domains (Niehuis et al., 2006) (Figure 3). However, compared to those of the 29 other lepidopteran species listed in Table 3, this gene was found to be the longest one, for example, a 190-bp extra-insertion was detected between the stem regions H505/20 and H515/21 by comparison with its closely related species *O. lunifer* (Figure 3). These insertion regions are shown to be highly variable among the currently available lepidopteran mtDNAs with their functions unknown. In addition, the stem H1047/38 is significantly variable among different lepidopteran species as well (Niehuis et al., 2006; Cameron and Whiting, 2008).

The *P. flavescens* mtDNA contained 22 tRNA genes, interspersed throughout the whole genome and ranging in size from 65 to 72 bp. All tRNAs except *trnS1* folded into the expected secondary cloverleaf structures (Figure 4). The *trnS1* lacked the DHU stem and this feature is common in the insect mitogenomes (Kim et al., 2006). In the 22 predicted tRNA secondary structures, 24 pairs of base mismatches were detected, and among which 17 were GU, 5 UU, 1 AG, and 1 UC (Figure 4).

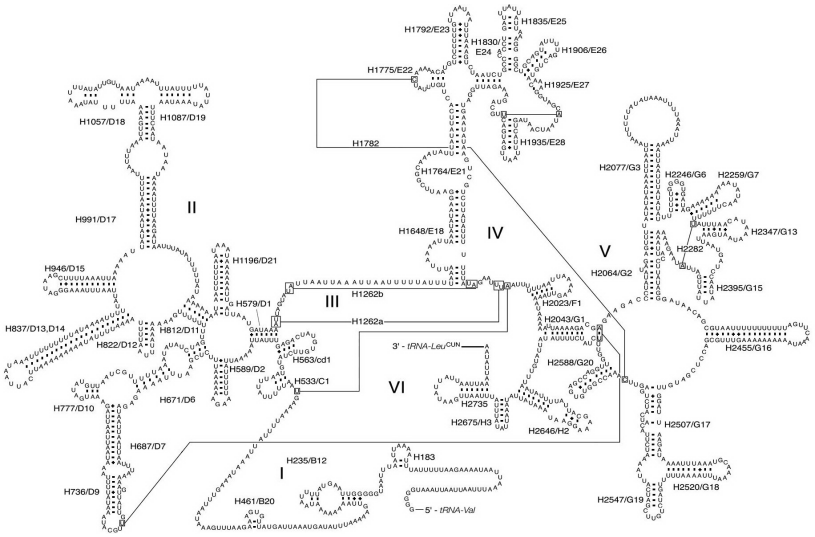

Figure 2. Secondary structure of the *rrnL* of *Phalera flavescens*. Roman numberals denote the conserved domain structure. Helices are numbered according to *Manaduca sexta*. Tertiary structures are denoted by boxed bases joined by solid lines. Watson-Crick pairs are joined by dashes; non-canonical guanine-uracil and other non-canonical interactions are joined by plus signs.

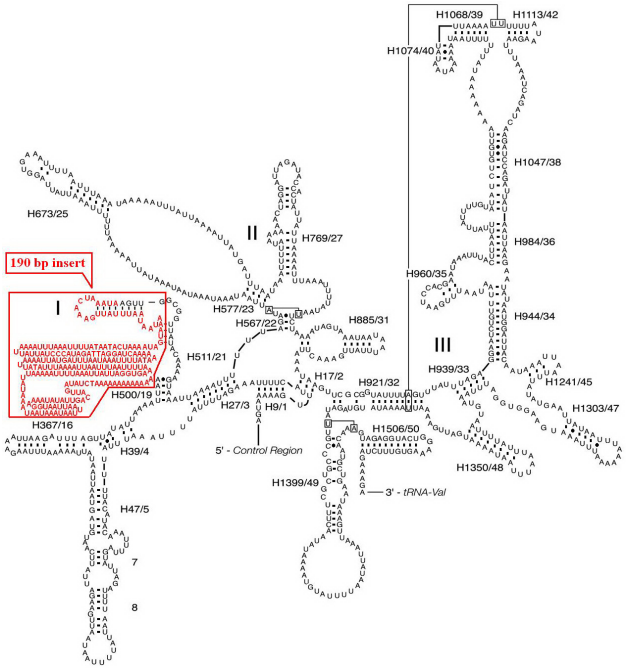

Figure 3. Secondary structure of the *rrnS* of *Phalera flavescens*. Roman numberals denote the conserved domain structure. Helices are numbered according to *Manaduca sexta*. Tertiary structures are denoted by boxed bases joined by solid lines. Watson-Crick pairs and non-canonical guanine-uracil are joined by dashes; other non-canonical interactions are joined by a dot. The 190-bp insertions are marked in red color.

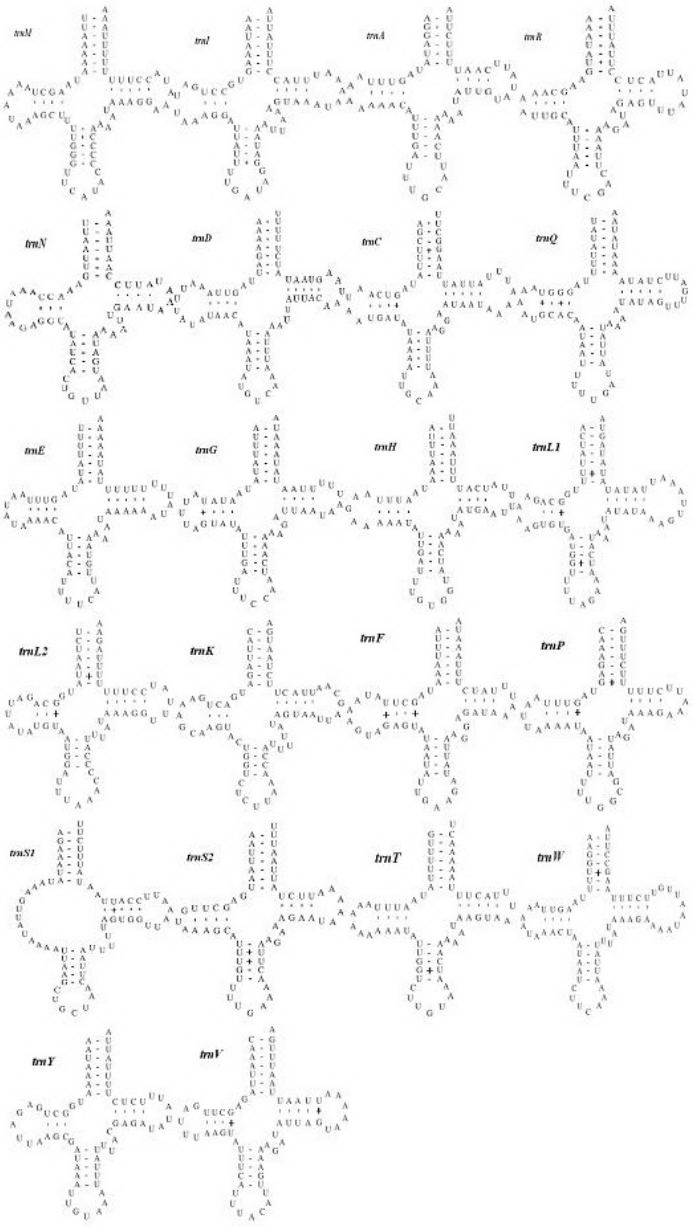

Figure 4. Secondary structures of the 22 tRNAs of the *Phalera flavescens* mitogenomes.

Non-coding regions and overlapping sequences

The non-coding regions included the AT-rich region (*D-loop*) and 16 intergenic spacer sequences. The intergenic spacers ranged from 1 to 58 bp in size, totaling 276 bp in length.

The longest 58-bp spacers were present between the *trnQ* and *nad2* genes. The location of these regions are fixed among the lepidopteran insects, whereas their sequences are highly diverged even in the two congeneric species examined, such as *O. furnicalis* and *O. nubilalis*, *B. mori* and *B. mandarina* (Cameron and Whiting, 2008). Most intergenic spacers include multiple short microsatellite-like repeat regions such as $(TA)_n$, poly-A and poly-T in *P. flavescens*.

The 351-bp AT-rich region is within the size range from 319 bp in *O. lunifer* (Salvato et al., 2008) to 747 bp in *B. mandarina* (Pan et al., 2008) detected in particular lepidopteran species to date (Table 5). This region was located between the *rrnS* and *trnM* genes as we expected, with the A+T content of 93.2%, which is also well within the range of other lepidopterans from 88.0% in *A. melete* (Hong et al., 2009) to 98.5% in *Phthonandria atrilineata* (Yang et al., 2009).

Table 5. Comparison of the total non-coding regions of lepidopteran mitochondrial genomes.

Species	Total size (bp)	A+T-rich (bp)	Non-*D-loop* (bp)
Artogeia melete	15,140	351	14,789
Argyreus hyperbius	15,156	349	14,807
Eumenis autonoe	15,489	678	14,811
Acraea issoria	15,245	430	14,815
Teinopalpus aureus	15,242	419	14,823
Papilio maraho	16,094	1270	14,824
Sasakia charonda kuriyamaensis	15,222	380	14,842
Sasakia charonda	15,224	380	14,844
Parnassius bremeri	15,389	504	14,885
Spilonota lechriaspis	15,368	441	14,927
Eriogyna pyretorum	15,327	357	14,969
Coreana raphaelis	15,360	375	14,985
Antheraea yamamai	15,338	334	15,004
Antheraea pernyi	15,566	552	15,014
Helicoverpa armigera	15,347	328	15,021
Saturnia boisduvalii	15,360	330	15,030
Phthonandria atrilineata	15,499	457	15,042
Hyphantria cunea	15,481	357	15,124
Lymantria dispar	15,569	435	15,134
Bombyx mori	15,643	499	15,144
Diatraea saccharalis	15,490	335	15,165
Bombyx mandarina	15,928	747	15,181
Adoxophyes honmai	15,680	490	15,190
Manduca sexta	15,516	324	15,192
Ochrogaster lunifer	15,593	319	15,274
Phalera flavescens	15,659	351	15,308

Twenty-six lepidopteran species in this study arranged from the shortest to the longest in the size of the non-*D-loop* region.

The microsatellite-like $(AT)_6$, $ACC(AT)_6$ and 16-bp poly-T elements were present in the flanking area of the *P. flavescens D-loop* region, and this case is similar to that found in other lepidopterans, such as $ATTTA(AT)_7$, $ATTTA(TA)_8$, $ATTTA(AT)_9$, $ATTTA(AT)_{10}$, $ATTTA(AT)_7(TA)_3$, and $ATTTA(AT)_{11}$ in *C. raphaelis*, *M. sexta*, *B. mandarina*, *B. mori*, *O. lunifer*, and *A. melete*, respectively (Yukuhiro et al., 2002; Kim et al., 2006; Pan et al., 2008; Cameron and Whiting, 2008; Hu et al., 2010; Liao et al., 2010). The motif poly-T stretch was located upstream of the *rrnS* 5'-end and preceded by the oligonucleotide ATAGA, and this structural feature is remarkably conserved in the majority of lepidopteran insects (Figure 5).

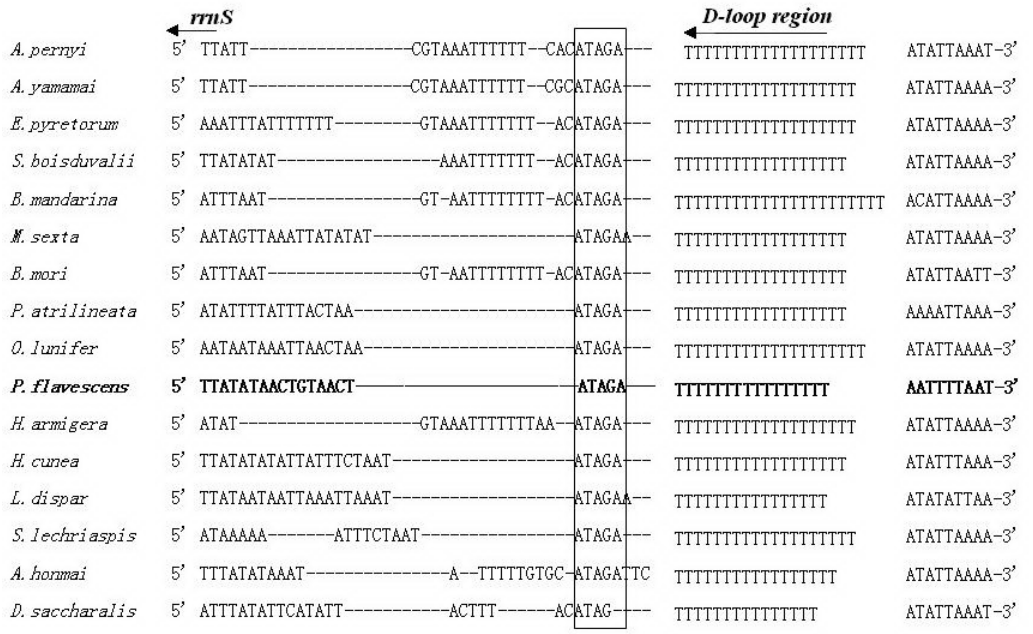

Figure 5. Sequence alignment of the partial *D-loop* region of 16 moth species. The boxed nucleotides indicate the conserved motif ATAGA. The replication direction of ATAGA motif + poly-T is indicated by arrow, the same as rrnS.

ACKNOWLEDGMENTS

Research supported by the Provincial Key Project of the Natural Science Foundation of Anhui Province, China (Grant #KJ2010A142), the CAS/SAFEA International Partnership Program for Creative Research Teams, Chinese Academy of Sciences (Grant #KZCX22Y-W2JC104), and the Opening Funds from the State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences (Grant #104143).

REFERENCES

Boore JL (1999). Animal mitochondrial genomes. *Nucleic Acids Res.* 27: 1767-1780.

Cameron SL and Whiting MF (2008). The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene* 408: 112-123.

Cha SY, Yoon HJ, Lee EM, Yoon MH, et al. (2007). The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). *Gene* 392: 206-220.

Coates BS, Sumerford DV, Hellmich RL and Lewis LC (2005). Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. *Int. J. Biol. Sci.* 1: 13-18.

Feng X, Liu DF, Wang NX, Zhu CD, et al. (2010). The mitochondrial genome of the butterfly *Papilio xuthus* (Lepidoptera: Papilionidae) and related phylogenetic analyses. *Mol. Biol. Rep.* 37: 3877-3888.

Gong YJ, Shi BC, Kang ZJ, Zhang F, et al. (2012). The complete mitochondrial genome of the oriental fruit moth *Grapholita molesta* (Busck) (Lepidoptera: Tortricidae). *Mol. Biol. Rep.* 39: 2893-2900.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid. Symp. Ser.* 41: 95-98.

Hao JS, Li CX, Sun XY and Yang Q (2005). Phylogeny and divergence time estimation of cheilostome bryozoans based on mitochondrial 16S rRNA sequences. *Chin. Sci. Bull.* 12: 1205-1211.

Hao JS, Su CY, Zhu GP, Chen N, et al. (2007). Mitochondrial 16S rDNA molecular morphology of the main butterflies' lineages and its phylogenetic significance. *J. Genet. Mol. Biol.* 18: 109-121.

Hong GY, Jiang ST, Yu M, Yang Y, et al. (2009). The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, *Artogeia melete* (Lepidoptera: Pieridae). *Acta Biochim. Biophys. Sin.* 41: 446-455.

Hong MY, Lee EM, Jo YH, Park HC, et al. (2008). Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects. *Gene* 413: 49-57.

Hu J, Zhang DX, Hao JS, Huang DY, et al. (2010). The complete mitochondrial genome of the yellow coaster, *Acraea issoria* (Lepidoptera: Nymphalidae: Heliconiinae: Acraeini): sequence, gene organization and a unique tRNA translocation event. *Mol. Biol. Rep.* 37: 3431-3438.

Jiang ST, Hong GY, Yu M, Li N, et al. (2009). Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). *Int. J. Biol. Sci.* 5: 351-365.

Kim I, Lee EM, Seol KY, Yun EY, et al. (2006). The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). *Insect Mol. Biol.* 15: 217-225.

Kim MI, Baek JY, Kim MJ, Jeong HC, et al. (2009). Complete nucleotide sequence and organization of the mitogenome of the red-spotted apollo butterfly, *Parnassius bremeri* (Lepidoptera: Papilionidae) and comparison with other lepidopteran insects. *Mol. Cells* 28: 347-363.

Kim SR, Kim MI, Hong MY, Kim KY, et al. (2009). The complete mitogenome sequence of the Japanese oak silkmoth, *Antheraea yamamai* (Lepidoptera: Saturniidae). *Mol. Biol. Rep.* 36: 1871-1880.

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

Kumar S, Tamura K and Nei M (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5: 150-163.

Lee ES, Shin KS, Kim MS, Park H, et al. (2006). The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene* 373: 52-57.

Li WW, Zhang XY, Fan ZX, Yue BS, et al. (2011). Structural characteristics and phylogenetic analysis of the mitochondrial genome of the sugarcane borer, *Diatraea saccharalis* (Lepidoptera: Crambidae). *DNA Cell Biol*. 30: 3-8.

Liao F, Wang L, Wu S, Li YP, et al. (2010). The complete mitochondrial genome of the fall webworm, *Hyphantria cunea* (Lepidoptera: Arctiidae). *Int. J. Biol. Sci.* 6: 172-186.

Liu Y, Li Y, Pan M, Dai F, et al. (2008). The complete mitochondrial genome of the Chinese oak silkmoth, *Antheraea pernyi* (Lepidoptera: Saturniidae). *Acta Biochim. Biophys. Sin.* 40: 693-703.

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

Nardi F, Carapelli A, Dallai R and Frati F (2003). The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations. *Insect Mol. Biol.* 12: 605-611.

Niehuis O, Yen SH, Naumann CM and Misof B (2006). Higher phylogeny of zygaenid moths (Insecta: Lepidoptera) inferred from nuclear and mitochondrial sequence data and the evolution of larval cuticular cavities for chemical defence. *Mol. Phylogenet. Evol.* 39: 812-829.

Ogoh K and Ohmiya Y (2004). Complete mitochondrial DNA sequence of the sea-firefly, *Vargula hilgendorfii* (Crustacea, Ostracoda) with duplicate control regions. *Gene* 327: 131-139.

Pan M, Yu Q, Xia Y, Dai F, et al. (2008). Characterization of mitochondrial genome of Chinese wild mulberry silkworm, *Bomyx mandarina* (Lepidoptera: Bombycidae). *Sci. China C Life Sci.* 51: 693-701.

Perna NT and Kocher TD (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* 41: 353-358.

Qin F, Jiang GF and Zhou SY (2012). Complete mitochondrial genome of the *Teinopalpus aureus* guangxiensis (Lepidoptera: Papilionidae) and related phylogenetic analyses. *Mitochondrial DNA* 2: 123-125.

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

Simon C, Simon C, Bekenbach A, Crespi B, et al. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* 87: 651-701.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The ClustalX windows interface: flexible strategies for

multiple sequences alignment aided by quality analysis tools. *Nucleic Acids Res.* 24: 4876-4882.
Wang XC, Sun XY, Sun QQ, Zhang DX, et al. (2011). Complete mitochondrial genome of the laced fritillary *Argyreus hyperbius* (Lepidoptera: Nymphalidae). *Dongwuxue Yanjiu* 32: 465-475.
Wilson K, Cahill V, Ballment E and Benzie J (2000). The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan crustaceans more closely related to insects than to branchiopods? *Mol. Biol. Evol.* 17: 863-874.
Yamauchi M, Miya M and Nishida M (2002). Complete mitochondrial DNA sequence of the Japanese spiny lobster, *Panulirus japonicus* (Crustacea: Decapoda). *Gene* 295: 89-96.
Yang L, Wei ZJ, Hong GY, Jiang ST, et al. (2009). The complete nucleotide sequence of the mitochondrial genome of *Phthonandria atrilineata* (Lepidoptera: Geometridae). *Mol Biol Rep.* 36: 1441-1449.
Yin J, Hong GY, Wang AM, Cao YZ, et al. (2010). Mitochondrial genome of the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and comparison with other Lepidopterans. *Mitochondrial. DNA* 21: 160-169.
Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, et al. (2002). 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.* 19: 1385-1389.
Zhang DX and Hewitt GM (1997). Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.* 25: 99-120.
Zhang DX, Szymura JM and Hewitt GM (1995). Evolution and structural conservation of the control region of insect mitochondrial DNA. *J. Mol. Evol.* 40: 382-391.
Zhao JL, Zhang YY, Luo AR, Jiang GF, et al. (2011). The complete mitochondrial genome of *Spilonota lechriaspis* Meyrick (Lepidoptera: Tortricidae). *Mol. Biol. Rep*. 38: 3757-3764.
Zhu YJ, Zhou GL, Fang R, Ye J, et al. (2010). The complete sequence determination and analysis of *Lymantria dispar* (Lepidoptera: Lymantriidae) mitochondrial genome. *Plant Quarantine* 4: 6-11.