EUROPEAN JOURNAL OF ENTOMOLOGY
ISSN (online): 1802-8829
http://www.eje.cz
Eur. J. Entomol. **114**: 407–415, 2017
doi: 10.14411/eje.2017.052
ORIGINAL ARTICLE

Tandem duplication of two tRNA genes in the mitochondrial genome of *Tagiades vajuna* (Lepidoptera: Hesperiidae)

Fang-Fang LIU[1,*], Yi-Ping LI[1,2,*], Ivan JAKOVLIĆ[3] and Xiang-Qun YUAN[1,]**

[1] Key Laboratory of Plant Protection Resources and Pest Management, Ministry of Education; College of Plant Protection, Northwest A&F University, Yangling, Shaanxi 712100, China; e-mails: hnndlff@163.com (F.-F.L.), lyp2003@126.com (Y.-P.L), yuanxq@nwsuaf.edu.cn (X.-Q.Y.)

[2] Key Laboratory of Applied Entomology, Northwest A&F University, Yangling, Shaanxi 712100, China

[3] Department of Fisheries, Beekeeping, Game Management and Special Zoology, Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25, 10000 Zagreb, Croatia; e-mail: ivanjakovlic@yahoo.com

Key words. Lepidoptera, Hesperiidae, Pyrginae, Eudaminae, *Tagiades vajuna*, mitochondrial phylogenomics, tRNA duplication, butterflies, skippers

Abstract. To explore the debated phylogenetic relationship of two Hesperiidae subfamilies, Pyrginae and Eudaminae, and contribute to the understanding of the evolution of mitogenomic architecture in butterflies, we sequenced the complete mitogenome of *Tagiades vajuna*. The mitogenome is a typical circular duplex molecule of 15,359 bp. Apart from the standard 22 tRNAs, it has a tandem duplication of *trnS*(AGN) and *trnE*, which is unique in lepidopteran insects. Comparison with *Ctenoptilum vasava* indicates that the *trnS1* duplication is not an ancestral state shared with other species of Tagiadini. Independent origin of the *trnS1* duplications was further confirmed by the reconstruction of the ancestral character state based on the topology of the phylogram. Furthermore, comparative analysis of mitogenomes with and without tRNA duplications indicates that tRNA duplication does not alter the codon usage pattern. The mitogenome has negative AT- and GC-skews, and it is highly A+T-biased (79.7%). The AT-rich (or control) region (283 bp) contains "ATAGA" and "ATTTA" motifs. Regarding the phylogenetic analysis, we found that removal of the third codon position (3CP) from datasets used for the mitochondrial phylogenomics of Hesperiidae is likely to produce results that are more consistent: Pyrginae were rendered paraphyletic by Eudaminae in both analyses of the dataset from which the 3CP was removed (13 PCGs + all RNAs), but inclusion of the 3CP resulted in a destabilized topology, resulting in both monophyly and polyphyly. We conclude that even shallow-phylogenies of insects should pay close attention to compositional and mutational biases in mitogenomes.

INTRODUCTION

Skippers (Hesperiidae) are a species-rich family (approximately 3,587 species), which makes up one fifth of the butterfly species in the world. Previous studies have found that neither three mitochondrial genes (Yuan et al., 2015), nor ten nuclear and mitochondrial markers (Sahoo et al., 2016), were capable of providing sufficient phylogenetic resolution to clarify the evolutionary relationships among the Hesperiidae. Particularly difficult to resolve are the relationships among the subfamilies Pyrginae, Eudaminae and Euschemoninae. This led the above authors to propose that future attempts would need much larger datasets.

Due to several advantageous characteristics of mitochondrial genomes, which include small size, abundance in tissues, strict orthology of encoded genes, presence of genes/regions evolving at different rates, uniparental inheritance and general absence of recombination, they are often the tool of choice for resolving phylogenetic relationships among insects (Brown & Wilson, 1979; Ballard & Whitlock, 2004; Gissi et al., 2008; Nelson et al., 2012; Simon & Hadrys, 2013; Wan et al., 2013; Cameron, 2014; Dai et al., 2016). There are some limitations to the application of mitochondrial phylogenomics in insects, such as compositional heterogeneity (Cameron, 2014) and mutation saturation in the third codon position (3CP) of protein-coding genes (PCGs) (Cameron et al., 2012). However, these limitations are relatively well-understood in deep-phylogenies of Insecta lineages (Cameron et al., 2007; Sheffield et al., 2009; Cameron, 2014; Song et al., 2016), the extent of their effects remains incompletely investigated (and occasionally even ignored, see below) in shallow-phylogenies of insects. This might be a consequence of an insufficient

* Contributed equally.

** Corresponding author; e-mail: yuanxq@nwsuaf.edu.cn

number of mitogenomes being available for such phylogenies. The latest two studies of the family Hesperiidae, employing the mitochondrial phylogenomics approach (Zhang et al., 2017b, c), found that the subfamilies Pyrginae and Eudaminae exhibit unresolved polytomies. As we suspect that their results may have been influenced by the aforementioned limitations of this approach, we adopt the most commonly used strategy to alleviate compositional heterogeneity: removal of the 3CP of PCGs. In addition, to improve taxon sampling and phylogenetic resolution, we sequenced the mitochondrial genome of a species belonging to the subfamily Pyrginae, *Tagiades vajuna*. To assess the effects of this strategy, we also conducted analyses on a dataset comprising all three codon positions of all PCGs. Although the removal of RNAs (rRNAs and tRNAs) has been a common approach used in studies utilizing mitogenomic data to investigate the phylogenetic relationships among the major lineages of Lepidoptera (Kim et al., 2011; Yang et al., 2015), there is no evidence of significant incongruence in the phylogenetic signals between RNAs and PCGs in insects (Cameron et al., 2007; Cameron, 2014). RNAs can actually carry a considerable phylogenetic signal and their inclusion has had positive effects on nodal support for some lineages within Lepidoptera (Wan et al., 2013). Thus we have adopted a strategy of maximizing the amount of data by including all 37 mitogenomic genes: 13 PCGs (excluded/included 3CP), two rRNAs and 22 tRNAs.

Beyond mitochondrial phylogenomics, mitogenomic gene duplication and the existence of pseudogenes is of interest in studies of the evolutionary history and mechanisms of gene rearrangement and recruitment (Ye et al., 2016). Although tRNAs are considered to be the most "expendable" among the genes encoded by the mitochondrial genomes and their content and order in mitogenomes is rather variable (Gissi et al., 2008), only four tRNA duplications and/or tRNA pseudogenes are recorded in the superfamily Papilionoidea. Among these four, one species belongs to Lycaenidae, *Coreana raphaelis* (*trnS1* duplication) (Kim et al., 2006), one to Nymphalidae, *Acraea issoria* (*trnI* pseudogene) (Hu et al., 2010) and the remaining two are hesperiids: *Ochlodes venata* (*trnL2* pseudogene) and *Ctenoptilum vasava* (*trnS1* duplication and *trnL2* pseudogene) (Hao et al., 2012). That for *C. vasava*, is one of the only two available mitogenomes for the tribe Tagiadini. The other one for *Daimio tethys* is a standard mitogenome. Therefore, it remains unclear whether this tRNA duplication is an autapomorphy for *C. vasava*, or whether it is shared by other, yet unsequenced, mitochondrial genomes in the tribe Tagiadini. To further explore this phenomenon, we have sequenced and characterized the mitogenome of *T. vajuna*, another Tagiadini butterfly, predominantly distributed in South and East Asia. Herein, we record a tandem tRNA duplication event, unique among Lepidoptera. We discuss the evolutionary mechanism for the duplication, and explore the correlation between tRNA duplication and codon usage. We also further compare the tRNA duplication events in the mitogenomes of *T. vajuna* and *C. vasava*, and investigate tRNA gene duplication in the broad phylogenetic context of the family Hesperiidae.

MATERIALS AND METHODS

Sample collection and DNA extraction

A single specimen of an adult *T. vajuna* (Maruyama, 1991) was captured by Shi-Hong Jiang on 14th November 2015 in Yuanshan Park, Shenzhen City, Guangdong Province, China (22°43′–23°15′N; 114°14′–115°16′E). The specimen was taxonomically identified by its morphological characteristics (particularly the genitalia) (Yuan et al., 2015), and *cox1* barcoding (Fig. S1) using 491 of the 493 sequences available for Pyrginae in the BOLD database (Ratnasingham & Hebert, 2013) (two incomplete Pyrginae barcodes were excluded). The complete specimen was immediately preserved in 100% ethanol and stored at –20°C in the Entomological Museum of the Northwest A&F University, Yangling, Shaanxi Province, China. The total DNA was extracted from the thoracic muscles following the manufacturer's instructions (EasyPure^R Genomic DNA Kit, TRAN, TransGen, Beijing, China). As *T. vajuna* is an unprotected invertebrate species, no permits were required for this study.

Sequence analysis

The complete mitochondrial genome of *T. vajuna* was sequenced on an Illumina HiSeq2000 system by the Genesky Biotechnologies Inc. company (Shanghai, China). Illumina sequencing data were quality-trimmed with Trimmomatic v0.35 (Bolger et al., 2014) and used for mitochondrial DNA genome assembly via a two-step MIRA4/Mitobim combined pipeline (Hahn et al., 2013), which implements a hybrid mapping and assembly approach for the targeted assembly of homologous sequences. The annotation of the mitochondrial DNA sequence was carried out in Geneious 8.1.3 (Biomatters, Auckland, New Zealand), using the mitogenome of another species of Hesperiidae, *Parnara guttata* (Hesperiidae: Hesperiinae; GenBank: NC_029136) (Shao et al., 2015), as a reference. Protein-coding genes (PCGs) were determined by finding the ORFs (employing codon table 5), and rRNAs (12S and 16S) were identified using the MITOS Web Server (Bernt et al., 2013). Transfer RNAs, including the two duplicated tRNAs, *trnE* and *trnS*(AGN), were also identified by MITOS and by manually inspecting the potential cloverleaf secondary structures and anticodons. Finally, all genes were visually inspected against the reference mitogenome via alignments in Geneious. Nucleotide composition and codon usage were calculated using MEGA 6.0 (Tamura et al., 2013). Comparative analysis of the codon usage was carried out on 87 complete lepidopteran mitogenomes available from GenBank: all 27 Hesperiids + species randomly chosen from the pool (2–5 species per family), including nine superfamilies and 20 families (Table S2). The mitogenome sequence is deposited in GenBank under the accession number KX865091.

Phylogenetic analysis

Phylogenetic analyses were conducted on the newly sequenced *T. vajuna* mitogenome and 26 complete Hesperiidae mitogenomes retrieved from GenBank. Two species, *Eurema hecabe* (NC_022685, Pieridae) and *Papilio machaon* (NC_018047, Papilionidae), were used as outgroups, adding up to 29 mitogenomes in total (Table S1).

Two GUI-based molecular biology tools, MitoTool (Zhang, 2016b) and BioSuite (Zhang, 2016a), developed by our colleague from the Chinese Academy of Sciences, Dong Zhang, were used to manage sequences and generate statistical tables as described before (Li et al., 2017; Zhang et al., 2017a). Fasta files with

nucleotide sequences for all 37 genes (13 PCGs, 2 rRNAs and 22 tRNAs) were extracted from GenBank files using MitoTool. PCGs were aligned in batches with MAFFT integrated into BioSuite, using codon-alignment mode. All RNAs were aligned with Q-INS-i algorithm (which takes secondary structure information into account) incorporated into MAFFT-with-extensions software (Katoh & Standley, 2013). Phylogenetic analyses were conducted using two different datasets: the complete 13 PCGs + all RNAs (named PCGRT dataset), and 13 PCGs with 3CP removed + all RNAs (named PCG12RT).

Best partitioning strategies and models for the two datasets were selected using PartitionFinder v1.1.1 (Lanfear et al., 2012). We created 16 pre-defined partitions of the two datasets: 13 PCGs + 2 rRNAs + all concatenated tRNAs as a single partition. We utilized the "greedy" algorithm (with branch lengths estimated as "unlinked") and Bayesian information criterion (BIC) to search for the best-fitting scheme (Table S3). Phylogenetic analyses were performed employing the best-fitting partitioning schemes, using maximum likelihood (ML) and Bayesian inference (BI). The ML analyses were performed using RaxML GUI (Silvestro & Michalak, 2012; Stamatakis, 2014), with an ML + rapid bootstrap (BS) algorithm with 1000 replicates. The BI analyses were performed using MrBayes 3.2.6 (Ronquist et al., 2012) with default settings and 6×10^6 MCMC generations (average standard deviation of split frequencies < 0.01, estimated sample size > 200, and potential scale reduction factor ≈ 1). Based on the resultant phylogram, we conducted an ancestral character state reconstruction for the tRNA duplications within the tribe Tagiadini using the MLGO web server (Hu et al., 2014).

RESULTS AND DISCUSSION

Genome features and characteristics

The complete mitogenome of *T. vajuna* (Fig. 1, Table 1) is 15,359 bp-long. It contains the standard 13 PCGs, two ribosomal RNAs (12S and 16S) and the non-coding AT-rich region (also called the control region). Intriguingly, two duplicated tRNA genes (*trnS1* and *trnE*) were found in this mitogenome with the help of the MITOS algorithm (discussed in the "tRNA genes" section), thus adding a further 24 tRNA genes. Fourteen genes are transcribed from the N strand and the remaining 25 genes from the J strand. Apart from the two duplicated *trn* genes, the order of genes in the *T. vajuna* mitogenome is relatively typical of Lepidoptera (Kim et al., 2009).

All PCGs, including the *COI*, which usually uses non-standard start codons in this group of animals (Ramírez-Ríos et al., 2016), use the standard ATN and TAA (or its

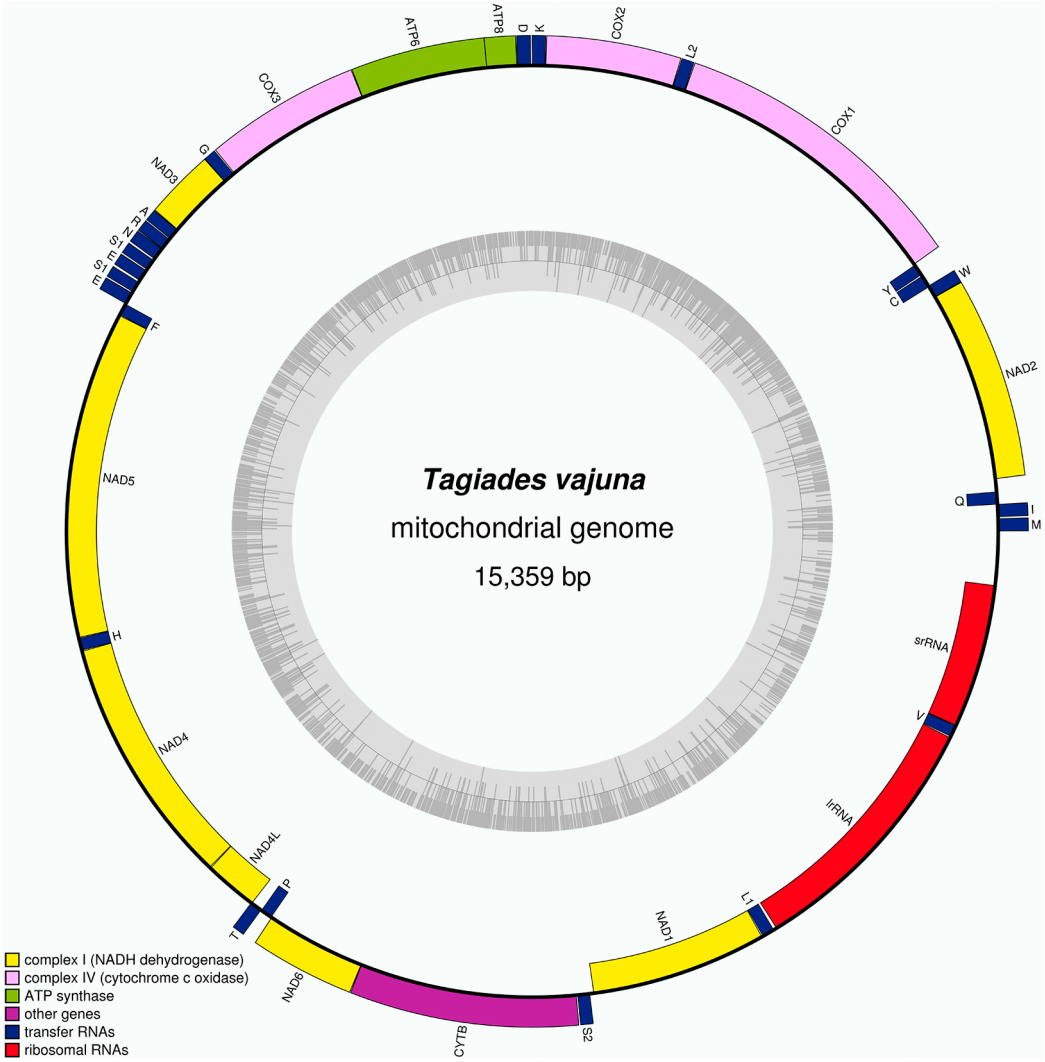

Fig. 1. Circular map of the mitochondrial genome of *T. vajuna*. Protein-coding and ribosomal genes are shown with standard abbreviations. The J-strand is visualized on the outer circle and the N-strand on the inner circle.

Table 1. Organization of the *T. vajuna* mitogenome.

Name	Strand	Location	Size	Anticodon	Start	Stop	IGNc
trnM	J	1–68	68	CAT (32–34)			7
trnI	J	76–143	68	GAT (106–108)			–3
trnQ	N	141–209	69	TTG (177–179)			66
ND2	J	276–1,289	1,014		ATT	TAA	–2
trnW	J	1,288–1,356	69	TCA (1,319–1,321)			–8
trnC	N	1,349–1,419	71	GCA (1,386–1,388)			4
trnY	N	1,424–1,488	65	GTA (1,456–1,458)			3
COI	J	1,492–3,025	1,534		ATG	T--	0
rnL2	J	3,026–3,092	67	TAA (3,056–3,058)			0
COII	J	3,093–3,771	679		ATT	T--	0
trnK	J	3,772–3,842	71	CTT (3,802–3,804)			5
trnD	J	3,848–3,919	72	GTC (3,879–3,881)			0
ATP8	J	3,920–4,081	162		ATT	TAA	–7
ATP6	J	4,075–4,752	678		ATG	TAA	–1
COIII	J	4,752–5,537	786		ATG	TAA	2
trnG	J	5,540–5,604	65	TCC (5,570–5,572)			–3
ND3	J	5,602–5,958	357		ATA	TAA	–1
trnA	J	5,958–6,020	63	TGC (5,987–5,989)			2
trnR	J	6,023–6,086	64	TCG (6,049–6,051)			0
trnN	J	6,087–6,153	67	GTT (6,118–6,120)			7
trnS1	J	6,161–6,221	61	GCT (6,182–6,184)			5
trnE	J	6,227–6,290	64	TTC (6,256–6,258)			9
trnS1	J	6,300–6,360	61	GCT (6,318–6,320)			5
trnE	J	6,366–6,432	67	TTC (6,395–6,397)			0
trnF	N	6,433–6,499	67	GAA (6,464–6,466)			–1
ND5	N	6,499–8,244	1,746		ATT	TAA	0
trnH	N	8,245–8,312	68	GTG (8,728–8,280)			0
ND4	N	8,313–9,651	1,339		ATG	T--	0
ND4L	N	9,652–9,942	291		ATA	TAA	–8
trnT	J	9,935–10,000	66	TGT (9,966–9,968)			0
trnP	N	10,001–10,066	66	TGG (10,034–10,036)			0
ND6	J	10,067–10,600	534		ATA	TAA	–1
CYTB	J	10,600–11,751	1,152		ATG	TAA	4
trnS2	J	11,756–11,822	67	TGA (11,787–11,789)			17
ND1	N	11,840–12,781	942		ATG	TAA	1
trnL1	N	12,783–12,852	70	TAG (12,822–12,824)			12
16S rRNA	N	12,865–14,240	1,376				2
trnV	N	14,243–14,307	65	TAC (14,276–14,278)			–1
12S rRNA	N	14,307–15,076	770				0
AT-rich		15,077–15,359	283				

Note: Sizes are given in bp; IGNc are intergenic nucleotides, where negative numbers indicate overlaps. Start and Stop are codons.

abbreviated version T--) as the start and stop codons, respectively (Table 1). *COI*, *COII* and *ND4* utilize the abbreviated T-- stop codon, which is presumed to be converted into TAA via posttranscriptional polyadenylation (Ojala et al., 1981).

The very high A+T content of 79.7% is comparable with that recorded for other lepidopteran mitogenomes (Table S1). When broken down by the codon positions of the 13 PCGs, the high A+T content was even somewhat greater at the first codon position (72.8%) than at the second position (69.9%), but by far the greatest at the third codon position – 91.4% (Table 2). This particularly high background mutational pressure towards A/T nucleotides at the third codon position is common in butterfly mitogenomes (92% on average) (Min et al., 2014).

Table 2. Nucleotide composition of the *T. vajuna* mitochondrial genome.

Feature	Size (bp)	Percentage of nucleotides						
		%T	%C	%A	%G	%A+T	AT-skew	GC-skew
Whole genome	15,359	40.7	12.4	39	7.9	79.7	–0.02	–0.22
PCGs	11,184	45.5	11.1	32.3	11.1	77.8	–0.17	0.00
1st codon position	3728	37	11	35.8	16.4	72.8	–0.02	0.20
2nd codon position	3728	48	16.8	21.9	13.3	69.9	–0.37	–0.12
3rd codon position	3728	52	5.4	39.4	3.7	91.4	–0.14	–0.19
tRNAs	1,601	40.5	7.7	41.5	10.2	82	0.01	0.14
rRNAs	2,146	42	4.8	43.2	10	85.2	0.01	0.35
AT-rich region	283	54.4	4.6	37.1	3.9	91.5	–0.19	–0.08

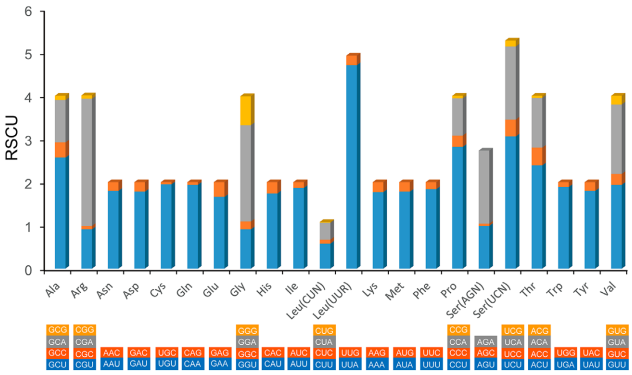

Fig. 2. Relative synonymous codon usage (RSCU) in the mitochondrial genome of *T. vajuna*. Codon families are on the x-axis.

To further explore this high bias towards A and T nucleotides, and demonstrate the frequency of synonymous codon usage, we have calculated the relative synonymous codon usage (RSCU) values (Fig. 2). UUA (*Leu2*), UCU (*Ser2*), CGA (*Arg*) and CCU (*Pro*) were the most frequently used codons and CUG, ACG, UGC and AGC the less frequently used. Three families (*Leu2, Ile, Phe*) account for 34.59% of all codons. Codons ending in A or T were predominant, adding up to 3,390, and accounting for 90.96%. The strong preference for A+T-rich codons over synonymous codons with a lower A+T content in almost all amino acids is observable in Fig. 2. The preference is particularly obvious in *Leu2*(UUR), where the UUA codon was used in 95.44% cases, as opposed to only 4.66% for UUG. This prevalence of NNU and NNA codons, also recorded in other skipper mitogenomes (Hao et al., 2012; Kim et al., 2014), corresponds well with the particularly high AT content at the third codon position.

tRNA genes

The mitogenome of *T. vajuna* harbours 24 tRNA genes interspersed between rRNAs and PCGs and ranging in length from 61 to 72 bp (Fig. 3, Table 1). Among them, 16 are encoded on the J strand and eight on the N strand. With the exception of *trnS*(AGN), which lacks the DHU arm, all of the tRNAs could be folded into cloverleaf secondary structures using MITOS (Fig. 3). The missing DHU stem of *trnS*(AGN) is an ancestral state in butterflies, including skippers, and probably evolved very early in Metazoa (Garey & Wolstenholme, 1989).

Comparative analysis of the selected 87 lepidopteran mitogenomes indicates that the tandem *trnS1-trnE* duplication might be unique among the lepidopteran mitogenomes sequenced. Since it is absent from the phylogenetically closely related *D. tethys* (Fig. 6), it is likely to be specific to the *Tagiades* genus. The *trnS1* duplication, however, was also recorded in two other species: *Ctenoptilum vasava* (Hesperiidae: Pyrginae) and *Coreana raphaelis* (Lycaenidae: Theclinae). Although this indicates that it is possible that the *trnS1* duplication (-$S1_a$-$S1_b$-E-) might be a shared ancestral state for the Tagiadini species, the gene order in *T. vajuna* (-$S1_a$-E_a-$S1_b$-E_b-) indicates that we can reject this hypothesis. This was further reinforced by the results of the ancestral character state reconstruction, which indicate the

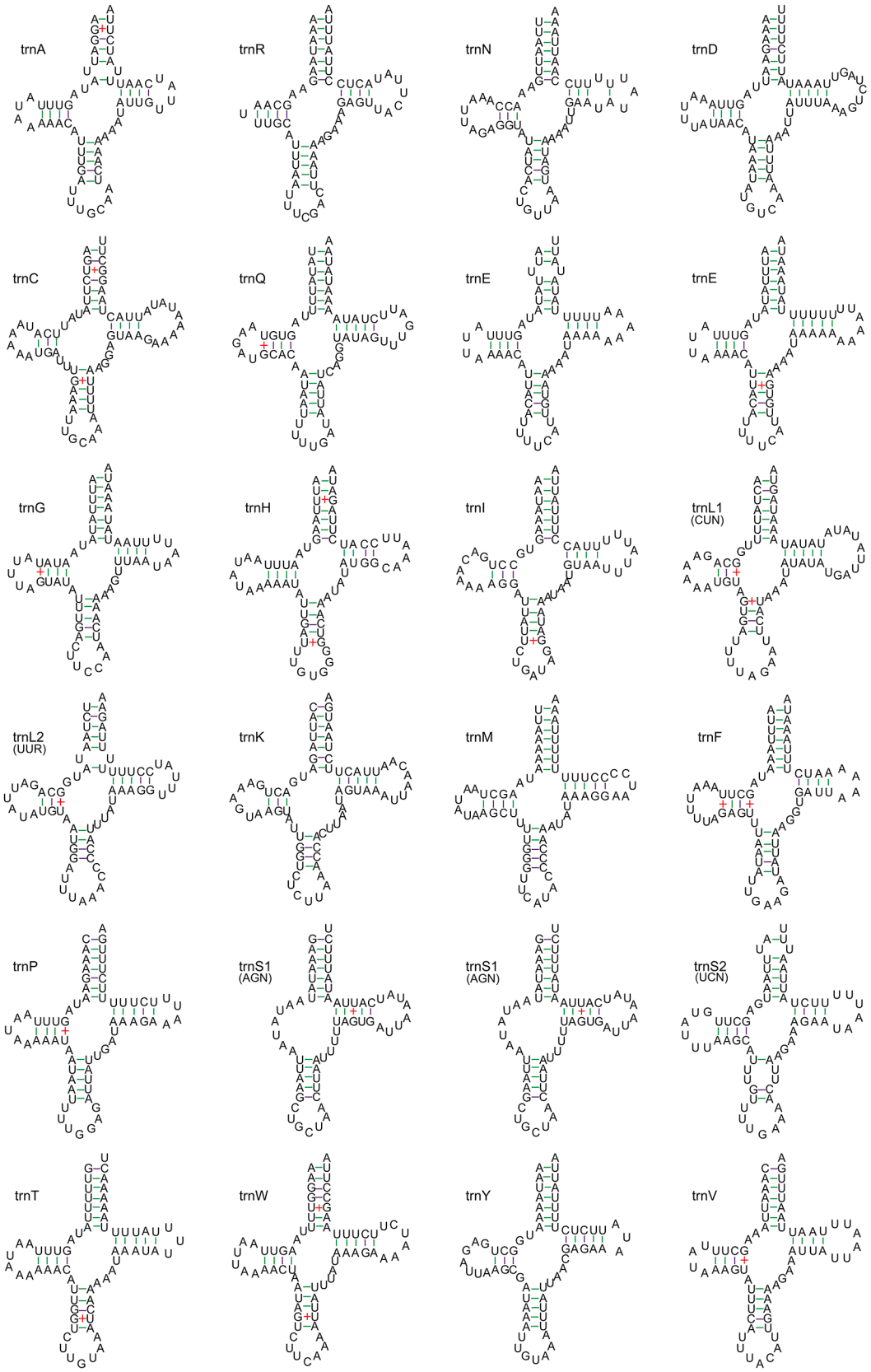

Fig. 3. Predicted secondary cloverleaf structure for the tRNAs of *T. vajuna* mitogenome. Dashes indicate Watson-Crick base pairing; additional sign (+) indicates unmatched base pairing.

typical lepidopteran arrangement -S1-E- as the ancestral state for the two Tagiadini nodes: A1 and A2 (Fig. 6). This confirms the independent origin of the duplicated *trnS1* in these two species: -S1-E- → -$S1_a$-$S1_b$-E- in *C. vasava*, whereas in *T. vajuna* it results from a duplication of the entire segment: -S1-E- → -$S1_a$-E_a-$S1_b$-E_b- (Fig. 5).

Comparison of the copies of the two duplicated tRNAs within and between the species (Fig. 4) lends further sup-

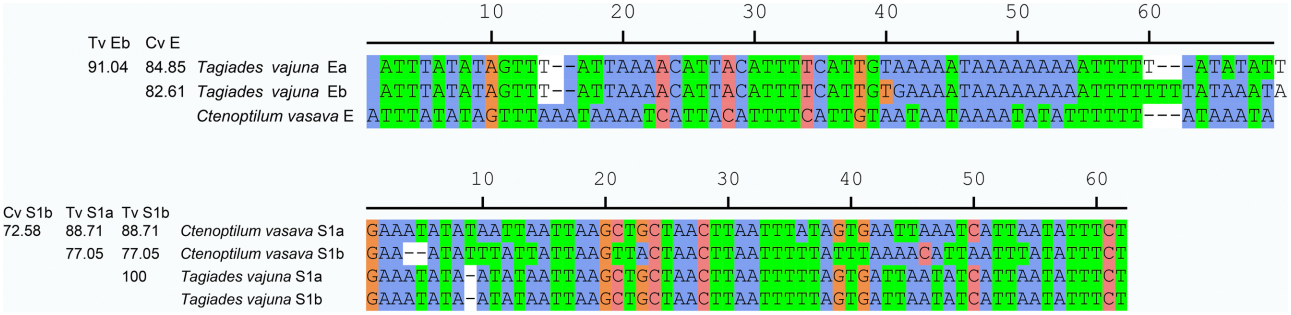

Fig. 4. Comparison of the duplicated *T. vajuna* tRNAs, *trnS1* and *trnE*, with selected orthologs. Similarity (%) between the sequences is indicated in the matrix on the left, where Tv and Cv are abbreviated names of the two species.

port to this scenario: the two *trnS1* copies in *T. vajuna* are identical to each other, but different from the two *C. vasava* orthologs (77.05 and 88.71% similarity, Fig. 4). *Tagiades vajuna trnEa* and *trnEb* copies are not identical, but the differences are relatively minor (91.04% similarity). Comparison with the *C. vasava trnE* sequences indicates that *T. vajuna trnEb* is the faster-evolving copy (*b* = 82.61 and *a* = 84.85% similarity), which is likely to be a consequence of relaxed mutational constraints afforded by the functional redundancy. However, all four copies can be folded into a functional cloverleaf structure (Fig. 3), which indicates that the duplication is relatively recent and that these tRNAs have probably retained their functionality. The codon usage for serine (AGN) (*C. raphaelis*, *C. vasava* and *T. vajuna*) and glutamic acid (only *T. vajuna*) encoded with two tRNA copies is analogous among all lepidopterans (Table S2), which indicates that tRNA duplication did not alter the codon usage pattern.

Intergenic spacers, overlapping sequences and the AT-rich region

Intergenic spacers in the *T. vajuna* mitogenome, distributed in 16 regions, with sizes ranging from 1 to 66 bp, add up to 151 bp. Two of these, Spacer 1 (66 bp, located between *trnQ* and *ND2*) and Spacer 6 (17 bp, *trnS2* and *ND1*; Table 1), are believed to be a constitutive synapomorphic feature of lepidopteran mitogenomes, and likely to have a functional role (Cameron & Whiting, 2008; Kim et al., 2009, 2010; Sheffield et al., 2008; Taanman, 1999).

There are eleven gene overlaps in the mitogenome, 1 to 8 bp in size, adding up to 36 bp. The longest two overlaps are between *trnW* / *trnC* and *ND4L* / *trnT* genes (Table 1). Overlapping genes might be a reflection of the selection for a short and economic mitogenome, and they usually involve *trn* genes, because their sequences are constrained by fewer mutations (Doublet et al., 2015).

Similar to some other Lepidoptera (Liao et al., 2010), the AT-rich region (283 bp, A+T = 91.5%) is located between *rrnS* and *trnM* in the *T. vajuna* mitogenome (Table 1, Fig. 1). These regions commonly have an ATTTA motif followed by several runs of microsatellite-like A/T sequences in other Lepidoptera (Cameron & Whiting, 2008). They also possess an ATAGA motif close to the 5'-end of the *12S rRNA* gene, followed by a poly-T stretch of variable length and a poly-A stretch (which can be interrupted or uninterrupted) immediately upstream of *trnM* (Kim et al., 2014). These two motifs also occur in the AT-rich region of the *T. vajuna* mitogenome. The poly-T stretch following ATAGA motif was 19 bp-long, whereas the poly-A stretch was comprised of 14 bp, and interrupted by a T base at position 12. The AT-rich region is believed to be involved in the control of transcription in insects (Zhang et al., 1995).

Phylogenetic analyses

As both methods (BI and ML) used in the phylogenetic analyses produced concordant topologies using the PCG12RT dataset, only the BI tree is shown (Fig. 6, all remaining phylograms are shown in Fig. S2). Using the PCGRT dataset, however, the topologies produced (Fig. S2) were neither concordant with each other, nor with the PCG12RT topology. Hence, we can conclude that removal of the 3CP from datasets used for the mitochondrial phylogenomics of Hesperiidae is likely to produce results that are more consistent. As expected, *T. vajuna* clustered with the other two Tagiadini species, *D. tethys* and *C. vasava*, in all four phylograms (two datasets × two methods) produced. However, phylogenetic relationships among/within the Pyrginae and Eudaminae subfamilies varied among the four phylograms: Pyrginae were rendered paraphyletic by Eudaminae in both analyses of the PCG12RT dataset, but polyphyletic in the ML analysis and monophyletic in the BI analysis of the PCGRT dataset.

Eudaminae recognized as a new hesperiid subfamily on the basis of a combination of molecular and morphological data by Warren et al. (2009) was quickly disputed as there is no morphological evidence for the monophyly of either Eudaminae or Pyrginae in the new sense (Simonsen et al., 2012). Sahoo et al. (2016) report that the position of Eudaminae in relation to Pyrginae is very variable depending on the methodological approach, with particular emphasis

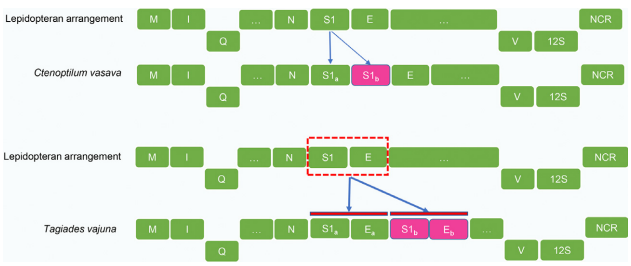

Fig. 5. *trnS1* and *trnE* duplications in lepidopteran mitogenomes. *Coreana raphaelis* is not shown, as its gene arrangement is identical to that of *C. vasava*.

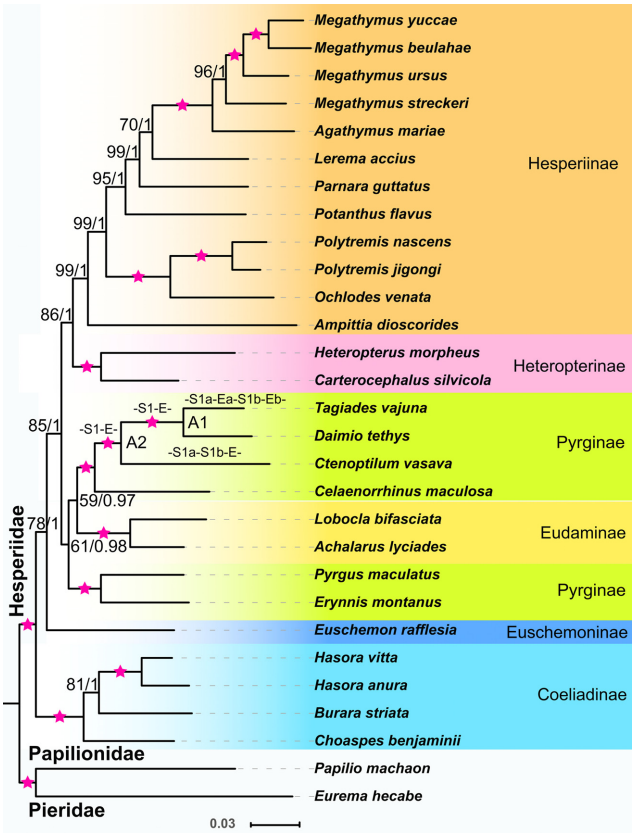

Fig. 6. Phylogenetic relationships of Hesperiidae inferred from mitogenomic sequences. Phylogenetic tree of 29 selected hesperiid butterflies inferred using nucleotide sequences of all 37 genes (with the third codon position of PCGs excluded) produced by the Bayesian inference method. Bootstrap (BS) and posterior probability (BPP) support values lower than the maximum support (BS = 100, BPP = 1.0) produced by the maximum likelihood and Bayesian inference analyses are shown above the branches. Star symbol indicates that both methods resulted in a maximum support value. Subfamilies are in graded colour blocks. Branch lengths correspond to the mutation rate of nucleotide sites. tRNA duplications and results of ancestral state reconstruction are indicated on the tree, where A1 and A2 refer to ancestral nodes.

on the partitioning schemes. They conclude that this is a reflection of the insufficient information in their molecular dataset (ten nuclear and mitochondrial markers) and propose that future attempts will have to rely on phylogenomic approaches. The mitochondrial phylogenomic approach used in this study, however, did not resolve the controversial phylogenetic relationship of the subfamilies Pyrginae and Eudaminae. As the PCG12RT dataset produced more consistent results and higher nodal support (Fig. S2), we hypothesize that this dataset may have produced results more closely reflecting the actual relationships between the Pyrginae and Eudaminae.

CONCLUSIONS

In this study, we report a tandem duplication of two tRNA genes, unique among all the lepidopteran mitogenomes characterized. Comparative analysis and results of an ancestral character state reconstruction indicate the *trnS1* duplication recorded in two species of Tagiadini (*T. vajuna* and *C. vasava*) is not a synapomorphy. Sequencing of further Tagiadini mitogenomes will be needed to determine whether the tandem duplication of *trnS1-trnE* recorded in *T. vajuna* is autapomorphic just for this species or the entire genus (*Tagiades*). Apart from the novel tRNA duplication, the mitogenome of *T. vajuna* has the standard features of Lepidoptera. Although our analyses indicate that the subfamily Pyrginae is most likely paraphyletic, varying topologies produced by different datasets and methods indicate that mitochondrial phylogenomics may not be able to fully resolve the phylogenetic relationships of the subfamilies Eudaminae and Pyrginae. The unstable topologies and weak nodal support recorded in both analyses (BI and ML) of the PCGRT dataset indicate that even shallow-phylogenies of insects should pay close attention to compositional and mutational biases in mitogenomes.

ACKNOWLEDGEMENTS. We thank De-Long Guan and Fan Song for their help with some of the programs used in the analyses. Apart from contributing the programs Dong Zhang also helped us to devise this methodology by contributing valuable ideas. We also thank J.R. Schrock for reviewing the manuscript, as well as three anonymous reviewers, whose constructive comments helped us improve this manuscript. This study was supported by the National Natural Science Foundation of China (Grants Nos31272345, 31071693 and 31772503), Chinese Universities Scientific Fund (Grants No. Z109021710) and Biodiversity Conservation Program of the Ministry of Environmental Protection, China (Grant No. SDZXWJZ01048-2017).

REFERENCES

Ballard J.W. & Whitlock M.C. 2004: The incomplete natural history of mitochondria. — *Mol. Ecol.* **13**: 729–744.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M. & Stadler P.F. 2013: MITOS: Improved de novo metazoan mitochondrial genome annotation. — *Mol. Phylogenet. Evol.* **69**: 313–319.

Bolger A.M., Lohse M. & Usadel B. 2014: Trimmomatic: a flexible trimmer for Illumina sequence data. — *Bioinformatics* **30**: 2114–2120.

Brown W.M. & Wilson A.C. 1979: Rapid evolution of animal mitochondrial DNA. — *Proc. Natl. Acad. Sci. U.S.A.* **76**: 1967–1971.

Cameron S.L. 2014: Insect mitochondrial genomics: implications for evolution and phylogeny. — *Annu. Rev. Entomol.* **59**: 95–117.

Cameron S.L. & Whiting M.F. 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.

Cameron S.L., Lambkin C.L., Barker S.C. & Whiting M.F. 2007: A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. — *Syst. Entomol.* **32**: 40–59.

Cameron S.L., Lo N., Bourguignon T., Svenson G.J. & Evans T.A. 2012: A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): robust support for interfamilial relationships and molecular synapomorphies define major clades. — *Mol. Phylogenet. Evol.* **65**: 163–173.

Dai L.S., Li S., Yu H.M., Wei G.Q., Wang L., Qian C., Zhang C.-F., Li J. & Sun Y. Zhao Y. 2016: Mitochondrial genome of the sweet potato hornworm, *Agrius convolvuli* (Lepidoptera:

Sphingidae), and comparison with other Lepidoptera species. — *Genome* **60**: 128–138.

Doublet V., Ubrig E., Alioua A., Bouchon D., Marcadé I. & Maréchal-Drouard L. 2015: Large gene overlaps and tRNA processing in the compact mitochondrial genome of the crustacean *Armadillidium vulgare*. — *RNA Biol.* **12**: 1159–1168.

Garey J.R. & Wolstenholme D.R. 1989: Platyhelminth mitochondrial DNA: Evidence for early evolutionary origin of a tRNA ser AGN that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. — *J. Mol. Evol.* **28**: 374–387.

Gissi C., Iannelli F. & Pesole G. 2008: Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. — *Heredity* **101**: 301–320.

Hahn C., Bachmann L. & Chevreux B. 2013: Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. — *Nucl. Acids Res.* **41**(13): e129, 9 pp.

Hao J.S., Sun Q., Zhao H., Sun X., Gai Y. & Yang Q. 2012: The complete mitochondrial genome of *Ctenoptilum vasava* (Lepidoptera: Hesperiidae: Pyrginae) and its phylogenetic implication. — *Comp. Funct. Genomics* **2012**(3): 328049, 13 pp.

Hu J., Zhang D., Hao J., Huang D., Cameron S. & Zhu C. 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.

Hu F., Zhou J., Zhou L. & Tang J. 2014: Probabilistic reconstruction of ancestral gene orders with insertions and deletions. — *IEEE/ACM Trans. Comput. Biol. Bioinform.* **11**: 667–672.

Katoh K. & Standley D.M. 2013: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. — *Mol. Biol. Evol.* **30**: 772–780.

Kim I., Lee E.M., Seol K.Y., Yun E.Y., Lee Y.B., Hwang J.S. & Jin B.R. 2006: The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). — *Insect Mol. Biol.* **15**: 217–225.

Kim M.I., Baek J.Y., Kim M.J., Jeong H.C., Kim K.G., Bae C.H., Han Y.S., Jin B.R. & Kim I. 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 M.J., Wan X., Kim K., Hwang J.S. & Kim I. 2010: Complete nucleotide sequence and organization of the mitogenome of endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae). — *Afr. J. Biotechnol.* **9**: 1182–1191.

Kim M.J., Kang A.R., Jeong H.C., Kim K.G. & Kim I. 2011: Reconstructing intraordinal relationships in Lepidoptera using mitochondrial genome data with the description of two newly sequenced lycaenids, *Spindasis takanonis* and *Protantigius superans* (Lepidoptera: Lycaenidae). — *Mol. Phylogenet. Evol.* **61**: 436–445.

Kim M.J., Wang A.R., Park J.S. & Kim I. 2014: Complete mitochondrial genomes of five skippers (Lepidoptera: Hesperiidae) and phylogenetic reconstruction of Lepidoptera. — *Gene* **549**: 97–112.

Lanfear R., Calcott B., Ho S.Y. & Guindon S. 2012: PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. — *Mol. Biol. Evol.* **29**: 1695–1701.

Li W.X., Zhang D., Boyce K., Xi B.W., Zou H., Wu S.G., Li M. & Wang G.T. 2017: The complete mitochondrial DNA of three monozoic tapeworms in the Caryophyllidea: a mitogenomic perspective on the phylogeny of eucestodes. — *Parasit. Vectors* **10**(1): 314, 13 pp.

Liao F., Wang L., Wu S., Li Y.-P., Zhao L., Huang G.-M., Niu C.-J., Liu Y.-Q. & Li M.-G. 2010: The complete mitochondrial genome of the fall webworm, *Hyphantria cunea* (Lepidoptera: Arctiidae). — *Int. J. Biol. Sci.* **6**: 172–186.

Min J.K., Wang A.R., Park J.S. & Kim I. 2014: Complete mitochondrial genomes of five skippers (Lepidoptera: Hesperiidae) and phylogenetic reconstruction of Lepidoptera. — *Gene* **549**: 97–112.

Nelson L.A., Lambkin C.L., Batterham P., Wallman J.F., Dowton M., Whiting M.F., Yeates D.K. & Cameron S.L. 2012: Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). — *Gene* **511**: 131–142.

Ojala D., Montoya J. & Attardi G. 1981: tRNA punctuation model of RNA processing in human mitochondria. — *Nature* **290**: 470–474.

Ramírez-Ríos V., Franco-Sierra N.D., Alvarez J.C., Saldamando-Benjumea C.I. & Villanueva-Mejía D.F. 2016: Mitochondrial genome characterization of *Tecia solanivora* (Lepidoptera: Gelechiidae) and its phylogenetic relationship with other lepidopteran insects. — *Gene* **581**: 107–116.

Ratnasingham S. & Hebert P.D. 2013: A DNA-based registry for all animal species: the barcode index number (BIN) system. — *PLoS ONE* **8**(7): e66213, 16 pp.

Ronquist F., Teslenko M., Mark P.V.D., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A. & Huelsenbeck J.P. 2012: MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. — *Syst. Biol.* **61**: 539–542.

Sahoo R.K., Warren A.D., Wahlberg N., Brower A.V., Lukhtanov V.A. & Kodandaramaiah U. 2016: Ten genes and two topologies: an exploration of higher relationships in skipper butterflies (Hesperiidae). — *Peer J.* **4**: e2653, 17 pp.

Shao L., Sun Q. & Hao J. 2015: The complete mitochondrial genome of *Parara guttata* (Lepidoptera: Hesperiidae). — *Mitochondrial DNA* **26**: 724–725.

Sheffield N.C., Song H., Cameron S.L. & Whiting M.F. 2008: A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. — *Mol. Biol. Evol.* **25**: 2499–2509.

Sheffield N.C., Song H., Cameron S.L. & Whiting M.F. 2009: Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. — *Syst. Biol.* **58**: 381–394.

Silvestro D. & Michalak I. 2012: raxmlGUI: a graphical frontend for RAxML. — *Org. Divers. Evol.* **12**: 335–337.

Simon S. & Hadrys H. 2013: A comparative analysis of complete mitochondrial genomes among Hexapoda. — *Mol. Phylogenet. Evol.* **69**: 393–403.

Simonsen T.J., Jong R.D., Heikkilä M. & Kaila L. 2012: Butterfly morphology in a molecular age – Does it still matter in butterfly systematics? — *Arthr. Struct. Dev.* **41**: 307–322.

Song F., Li H., Jiang P., Zhou X., Liu J., Sun C., Vogler A.P. & Cai W. 2016: Capturing the phylogeny of Holometabola with mitochondrial genome data and Bayesian site-heterogeneous mixture models. — *Genome Biol. Evol.* **8**: 1411–1426.

Stamatakis A. 2014: RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. — *Bioinformatics* **30**: 1312–1313.

Taanman J.W. 1999: The mitochondrial genome: structure, transcription, translation and replication. — *Acta Biochim. Biophys.* **1410**: 103–123.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. 2013: MEGA6: molecular evolutionary genetics analysis version 6.0. — *Mol. Biol. Evol.* **30**: 2725–2729.

Wan X., Kim M.J. & Kim I. 2013: Description of new mitochondrial genomes (*Spodoptera litura*, Noctuoidea and *Cnaphalocrocis medinalis*, Pyraloidea) and phylogenetic reconstruction of Lepidoptera with the comment on optimization schemes. — *Mol. Biol. Rep.* **40**: 6333–6349.

Warren A.D., Ogawa J.R. & Brower A.V. 2009: Revised classification of the family Hesperiidae (Lepidoptera: Hesperioidea) based on combined molecular and morphological data. — *Syst. Entomol.* **34**: 467–523.

Yang X., Cameron S.L., Lees D.C., Xue D. & Han H. 2015: A mitochondrial genome phylogeny of owlet moths (Lepidoptera: Noctuoidea), and examination of the utility of mitochondrial genomes for lepidopteran phylogenetics. — *Mol. Phylogenet. Evol.* **85**: 230–237.

Ye F., Lan X.E., Zhu W.B. & You P. 2016: Mitochondrial genomes of praying mantises (Dictyoptera, Mantodea): rearrangement, duplication, and reassignment of tRNA genes. — *Sci. Rep.* **6**: 25634, 9 pp.

Yuan F., Yuan X.Q. & Xue G.X. 2015: *Fauna Sinica: Insecta Vol. 55. Lepidoptera, Hesperiidae.* Science Press, Beijing, 754 pp.

Zhang D. 2016a: *BioSuite Software*. URL: https://github.com/dongzhang0725/BioSuite (last accessed 9 Oct 2016).

Zhang D. 2016b: *MitoTool Software*. URL: https://github.com/dongzhang0725/MitoTool (last accessed 23 Oct 2016).

Zhang D.X., Szymura J.M. & Hewitt G.M. 1995: Evolution and structural conservation of the control region of insect mitochondrial DNA. — *J. Mol. Evol.* **40**: 382–391.

Zhang D., Zou H., Wu S.G., Li M., Jakovlic I., Zhang J., Chen R., Wang G.T. & Li W.X. 2017a: Sequencing, characterization and phylogenomics of the complete mitochondrial genome of *Dactylogyrus lamellatus* (Monogenea: Dactylogyridae). — *J. Helminthol.* [in press].

Zhang J., Cong Q., Fan X.L., Wang R., Wang M. & Grishin N.V. 2017b: Mitogenomes of giant-skipper butterflies reveal an ancient split between deep and shallow root feeders. — *F1000 Res.* **6**: 222, 7 pp.

Zhang J., Cong Q., Shen J., Fan X.-L., Wang M. & Grishin N.V. 2017c: The complete mitogenome of *Euschemon rafflesia* (Lepidoptera: Hesperiidae). — *Mitochondrial DNA (B)* **2**: 136–138.

Received June 21, 2017; revised and accepted September 25, 2017
Published online October 12, 2017

Supplementary files:

Table S1 (http://www.eje.cz/2017/052/S01.xlsx). List of the species included in this study.

Table S2 (http://www.eje.cz/2017/052/S02.xlsx). Codon usage for glutamic acid and serine (AGN) among 87 selected Lepidoptera.

Table S3 (http://www.eje.cz/2017/052/S03.docx). The best partitioning scheme selected by PartitionFinder for the two datasets.

Fig. S1 (http://www.eje.cz/2017/052/S04.pdf). Phylogenetic trees based on 491 *cox1* barcoding sequences.

Fig. S2 (http://www.eje.cz/2017/052/S05.pdf). Phylogenetic trees based on all four analyses (2 datasets × 2 analytical methods).