International Journal of Biological Macromolecules 120 (2018) 1048–1054

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

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

Comparative analysis of mitochondrial genomes of the superfamily Grylloidea (Insecta, Orthoptera) reveals phylogenetic distribution of gene rearrangements

Chuan Ma, Jianke Li *

Institute of Apicultural Research/Key Laboratory of Pollinating Insect Biology, Ministry of Agriculture, Chinese Academy of Agricultural Sciences, Beijing 100093, China

ARTICLE INFO

Article history:
Received 15 July 2018
Received in revised form 20 August 2018
Accepted 30 August 2018
Available online 30 August 2018

Keywords:
Start codon
cox1
Gene rearrangement

ABSTRACT

To further characterize mitochondrial genome (mitogenome) features of the superfamily Grylloidea (Insecta, Orthoptera), mitogenomes of *Cacoplistes rogenhoferi* and *Meloimorpha japonica* representing the family Mogoplistidae and three *Ornebius* species of Phalangopsidae were sequenced. A repeat-containing control region (CR) and 37 genes were present in these mitogenomes. Unusual start codons (TCG, CCG, and CTG) of *cox1* and, in *Ornebius*, a partial stop codon (T) of *nad1* followed by a 15–17-bp intergenic spacer were proposed based on transcript information and sequence alignments. The mitogenome-based phylogenetic trees suggest strongly the familial relationships as (((Phalangopsidae + Gryllidae) + Trigonidiidae) + Mogoplistidae). The exclusive occurrence of the *trnE-trnS1-trnN* rearrangement in Phalangopsidae, Gryllidae, and Trigonidiidae is suggestive of its appearance in the common ancestor of these families after the separation of Mogoplistidae. The *trnV* transposition in *O. bimaculatus* and formerly sequenced *Trigonidium sjostedti* (Trigonidiidae) indicates a potential consequence of parallel evolution. This study offers novel insights into mitogenome evolution, especially gene rearrangements, of Grylloidea.

1. Introduction

Grylloidea Laicharting, 1781 represents one of the most species-rich superfamilies of insects in the order Orthoptera. It comprises 5555 extant species including true crickets, scaly crickets, beetle crickets, and sword-tailed crickets according to the Orthoptera Species File (OSF) [1]. Some of them are among the most popular acoustic insects and are kept as pets for entertainment [2]. Grylloidea is currently subdivided into four families (Gryllidae, Mogoplistidae, Phalangopsidae, and Trigonidiidae) and one unclassified subfamily (Pteroplistinae) [1]. Classification schemes of Grylloidea are changing over the years [3–5] and even two families, Gryllotalpidae (mole crickets) and Myrmecophilidae (ant-loving crickets) that are currently classified within Gryllotalpoidea [1], were included in Grylloidea in the OSF before the study by Chintauan-Marquier et al. [5]. The general frame of a phylogenetic classification of Grylloidea has been proposed by Chintauan-Marquier et al. [5] and these hypotheses are popularly adopted as a reference classification, yet unresolved taxa still exist.

The insect mitogenome is a compact circular molecule (15–18 kb in size) encoding a conserved set of 13 protein-coding genes (PCGs), 22 tRNA genes, and two rRNA genes [6,7]. It is featured as maternal inheritance, rare recombination, faster evolution rate, and conserved gene arrangement [8], endowing mitogenome with a strong potential to provide phylogenetic signals to address phylogenetic and evolutionary questions. Indeed, both mitogenome sequence and gene rearrangement have been used for research of molecular evolution, population structure, species/subspecies identification, and phylogenetic inference [6,9,10].

Mitogenomes provide an important source of data to validate taxonomic hypotheses for Grylloidea at the mitogenomic level. However, it is not yet possible due to the low number of available mitogenomes of Grylloidea in GenBank (10 for Gryllidae and one for Trigonidiidae). To make the situation even worse, two of these mitogenomes are found to be contaminated with nuclear mitochondrial pseudogenes (numts) [11,12]. In the sequenced mitogenomes of Grylloidea, two types of gene rearrangements have been reported: a transposition of *trnV* to the site between *rrnS* and the CR, and a local inversion of the ancestral *trnN-trnS1-trnE* to *trnE-trnS1-trnN*. The transposition is found so far exclusively in *T. sjostedti* of Trigonidiidae [13] whereas the inversion occurs in Gryllidae [11,14,15] and Trigonidiidae [13]. Understanding phylogenetic distribution patterns of these rearrangements is fundamental to trace their evolutionary history and to infer phylogenetic relationships within Grylloidea. However, such distribution information of these rearrangements in major lineages of Grylloidea is still lacking. To this end, additional mitogenome sequencing for Grylloidea and in

* Corresponding author at: No. 8 Administration Building, 12 Zhongguancun South Avenue, Beijing, China.
E-mail address: apislijk@126.com (J. Li).

https://doi.org/10.1016/j.ijbiomac.2018.08.181
0141-8130/

particular the unpresented families, Mogoplistidae and Phalangopsidae, are thus of paramount importance.

Here, the mitogenome sequences of two species (*C. rogenhoferi* and *M. japonica*) of Phalangopsidae and three species (*O. fuscicercis*, *O. kanetataki*, and *O. bimaculatus*) of Mogoplistidae were determined and compared. Our study provides reliable mitogenomes of the two families and obtains an accurate annotation of *cox1* and *nad1* positions. Comparative mitogenome analysis reveals phylogenetic distribution of gene rearrangements in Grylloidea, which sheds lights on evolutionary history of these gene rearrangements.

2. Materials and methods

2.1. Mitogenome sequencing

Live insects were purchased from a flower and bird market in Shanghai, China and preserved in 100% ethanol at 4 °C. Genomic DNA was extracted individually from each species using a DNeasy Blood & Tissue kit (Qiagen). Library construction and sequencing were performed by Novogene (Beijing, China) on the HiSeq 2500 platform (Illumina Inc.) following the manufacture's protocol for 150-bp paired-end reads. Raw data were initially quality-trimmed by removing reads with adaptor contamination (>15 bp matching to the adaptor sequence), poly-Ns (>5 bp Ns), or >1% error rate (>10 bp with quality score < 20) [16]. Clean reads were assembled into contigs with '-R -F -K 71' implemented in SOAPdenovo-Trans [17] and the resultant contigs were searched against PCGs of a closely-related species *Teleogryllus commodus* (JQ686193) downloaded from GenBank (BLASTX, E-value < 10^{-5}). Clean reads from the hit contigs were used to assemble the full mitogenome in MITObim v1.9 [18] and both ends of the final assembly were manually checked for overlap to establish a circular mitogenome. Finally, these reads were mapped to the obtained mitogenomes using SAMtools [19] to calculate sequencing depth.

To ensure the correctness of the read assembly of the CR usually encompassing large repeats, species-specific PCR primers were designed to amplify the whole CR (Table S1). The PCR was performed using LA Taq™ (TaKaRa Co., Dalian, China) with the following cycling settings: 95 °C for 1 min; 30× (98 °C, 10 s; 52 °C, 10 s; 65 °C, 3.5 min); 65 °C for 5 min. Purified PCR products were ligated to pMD18-T Vector (TaKaRa Co., Dalian, China) followed by Sanger sequencing. The sequencing reads were assembled using the SeqMan software (DNAStar, Inc.).

2.2. Mitogenome annotation and analysis

Mitogenome annotation was initially carried out on the MITOS webserver [20] with the invertebrate mitochondrial code. Gene boundaries were manually corrected based on homologous gene alignments of Grylloidea from GenBank. For PCGs, variable start codons and incomplete stop codons were considered to minimize intergenic spacers and gene overlaps [7,21,22]. The rRNA gene ranges were extended to boundaries of their adjacent tRNA genes. For *O. bimaculatus* that had a tRNA transposition (see below), rRNA gene boundaries were defined by alignment with the two other congeneric mitogenomes.

To confirm *cox1* start codon and *nad1* stop codon positions, the 5′ end of *cox1* transcripts (*M. japonica*, *O. fuscicercis*, and *O. kanetataki*) and 3′ end of *nad1* transcripts (*O. fuscicercis* and *O. kanetataki*) were amplified and sequenced. Specifically, total RNA was extracted with TRIzol (Invitrogen) from the head of single specimens stored at −80 °C. First-strand cDNA synthesis and rapid amplification of cDNA ends (RACE; see gene-specific primers in Table S1) were performed with a SMARTer RACE 5′/3′ Kit (Clontech). PCR products were determined by Sanger sequencing. The assembled and annotated mitogenome sequences were deposited in GenBank (Table 1).

Tandem repeats were identified using the online Tandem Repeats Finder [23] with the maximum period size set to 1000. Sequence identity matrix was calculated with BioEdit [24]. Nucleotide composition was calculated using DAMBE6 [25].

2.3. Phylogenetic analysis

To determine the phylogenetic distribution of gene rearrangements, the phylogeny of the superfamily Grylloidea was reconstructed via the Bayesian inference (BI) and the maximum likelihood (ML) approaches. Nine numt-free mitogenomes accessible in GenBank at the time of data collection (8th May, 2018), combined with our newly obtained mitogenomes, were included for phylogenetic analysis (Table 1). Based on the phylogeny of Ensifera [3,5,14], four mitogenomes from closely related superfamilies (two from Gryllotalpoidea, one from Rhaphidophoroidea, and one from Schizodactyloidea) were selected as outgroup taxa (Table 1). All the 37 gene and 13 protein sequences were extracted from the 18 mitogenomes using a Perl script [26]. The 13 protein sequences, 22 tRNA genes, and two rRNA genes were individually aligned using MUSCLE [27]. The nucleotide sequences of the 13 PCGs were converted into a codon alignment based on the corresponding protein alignment as implemented in PAL2NAL [28]. Poorly aligned positions and divergent regions were eliminated by the Gblocks server [29] with the following settings: minimum number of sequences for a conserved/flanking position (10/15), maximum number of contiguous non-conserved positions (4), minimum length of a block (10), and allowed gap positions (none).

All the 37 gene sequences excluding the third codon positions of the PCGs, which potentially suffered from mutational saturation detected

Table 1
Taxonomic information and GenBank accession numbers for phylogenetic analysis.

Superfamily	Family	Subfamily	Species	GenBank
Grylloidea	Gryllidae	Gryllinae	*Loxoblemmus doenitzi*	NC_033985
Grylloidea	Gryllidae	Gryllinae	*Loxoblemmus equestris*	NC_030763
Grylloidea	Gryllidae	Gryllinae	*Teleogryllus emma*	KU562917
Grylloidea	Gryllidae	Gryllinae	*Teleogryllus commodus*	JQ686193
Grylloidea	Gryllidae	Gryllinae	*Velarifictorus hemelytrus*	NC_030762
Grylloidea	Gryllidae	Oecanthinae	*Oecanthus rufescens*	KX057720
Grylloidea	Gryllidae	Oecanthinae	*Oecanthus sinensis*	NC_034799
Grylloidea	Gryllidae	Podoscirtinae	*Truljalia hibinonis*	NC_034797
Grylloidea	Phalangopsidae	Cachoplistinae	*Cacoplistes rogenhoferi*	MH580272
Grylloidea	Phalangopsidae	Cachoplistinae	*Meloimorpha japonica*	MH580273
Grylloidea	Mogoplistidae	Mogoplistinae	*Ornebius bimaculatus*	MH580274
Grylloidea	Mogoplistidae	Mogoplistinae	*Ornebius fuscicercis*	MH580275
Grylloidea	Mogoplistidae	Mogoplistinae	*Ornebius kanetataki*	MH580276
Grylloidea	Trigonidiidae	Trigonidiinae	*Trigonidium sjostedti*	NC_032077
Gryllotalpoidea	Gryllotalpidae	Gryllotalpinae	*Gryllotalpa orientalis*	NC_006678
Gryllotalpoidea	Myrmecophilidae	Myrmecophilinae	*Myrmecophilus manni*	NC_011301
Rhaphidophoroidea	Rhaphidophoridae	Troglophilinae	*Troglophilus neglectus*	NC_011306
Schizodactyloidea	Schizodactylidae	Comicinae	*Comicus campestris*	NC_028062

by DAMBE6 [25], were concatenated into the final dataset for phylogenetic reconstruction. Best-fit partitioning strategies and models of molecular evolution for phylogenetic inference were identified in PartitionFinder v2.1.1 [30] based on 50 pre-defined partitions (22 for the tRNA genes, two for rRNA genes, and 26 for the first and second positions of the 13 PCGs). The following settings were selected in PartitionFinder v2.1.1: the Bayesian Information Criterion, greedy search algorithms, and unlinked branch lengths. The resulting partition schemes with respective substitution models were used for the BI and ML tree reconstruction.

The BI tree reconstruction was performed in MrBayes v3.2 [31] with two runs, each with four Markov Chain Monte Carlo chains. All partitions were allowed to have their own set of parameters and to evolve under different rates. The analysis was run for 10 million generations, sampling trees every 1000 generations. The initial 25% of the trees were discarded as burn-in and the remaining trees were used to generate a 50% majority rule consensus tree with nodal confidence assessed with posterior probabilities (BPP). Bayesian runs achieved sufficient convergence by ascertaining that the average standard deviation of split frequencies between chains was below 0.01 at the end of the runs and that the potential scale reduction factor of each parameter was 1.00.

The ML tree was reconstructed using RAxML v8.2.10 [32] with the GTRGAMMA substitution model. The best-scoring ML tree was obtained by executing 200 ML searches. Bootstrap (BS) values for each node were gained via 1000 replicates.

3. Results

3.1. Mitogenome organization

The five mitogenomes ranged in size from 15,880 bp in *M. japonica* to 16,589 bp in *O. kanetataki* with an average sequencing depth from 301× to 1050×. A strong A + T-bias in nucleotide composition ranging from 72.43% in *M. japonica* to 76.47% in *O. bimaculatus* was found (Table 2). These mitogenomes encoded the typical set of 37 genes for insects consisting of 13 PCGs (*atp6*, *atp8*, *cob*, *cox1–3*, *nad1–6*, and *nad4l*), 22 tRNA genes (two each for Leucine and Serine, one for the other amino acids), and two rRNA genes (*rrnS* and *rrnL*). The ancestral gene arrangement of insects was observed in *O. fuscicercis* and *O. kanetataki* whereas gene rearrangements were found in the three other mitogenomes: a transposition of *trnV* to the site between *rrnS* and the CR in *O. bimaculatus* and a local inversion of the *trnN-trnS1-trnE* to *trnE-trnS1-trnN* in *C. rogenhoferi* and *M. japonica* (Fig. 1A and B, Table S2). Multiple overlaps between adjacent genes were detected, of which 8 bp between *trnW* and *trnC*, 2 bp between *trnY* and *cox1*, 7 bp between *atp8* and *atp6*, and 7 bp between *nad4* and *nad4l* were present in all five mitogenomes.

3.2. Coding genes

For PCGs, an even stronger preference for A + T was detected in third codon positions (82.51%–87.17%) than in the first (66.52%–69.67%) and second (65.83%–68.06%) codon positions (Table 2). All PCGs started with ATN codons with the exception of *cox1* (TCG in *C. rogenhoferi* and *M. japonica*, CTG in *O. bimaculatus*, and CCG in *O. fuscicercis* and *O. kanetataki*), *nad1* (TTG in *C. rogenhoferi* and *M. japonica*), and *nad4* (GTG in *O. fuscicercis*) (Table S2). Most PCGs terminated with a standard stop codon TAA or TAG, whereas the others used an incomplete stop codon T or TA (Table S2). The annotations of *cox1* start codons (*M. japonica*, *O. fuscicercis*, and *O. kanetataki*) and *nad1* stop codons (*O. fuscicercis* and *O. kanetataki*) were confirmed by sequencing the 5′ and 3′ ends of the two gene transcripts, respectively (Fig. 2A and B).

The tRNA genes ranged in size from 58 bp for *trnS1* in both *O. bimaculatus* and *O. kanetataki* to 72 bp for *trnK* in *O. bimaculatus*. Apart from *trnS1* that lacked a stable dihydrouridine arm, all tRNA genes could be folded into typical clover-leaf secondary structures. The 22 tRNAs each shared the same anticodon among the five species. The two rRNA genes, *rrnL* and *rrnS*, were encoded by the minority strand. Each rRNA gene had a similar A + T content among these mitogenomes (Table 2).

3.3. Non-coding regions

The CR was characterized by a high A + T content, varied sizes, and occurrence of long tandem repeats (longer than 100 bp). The highest A + T content was found in *O. bimaculatus* (84.78%) and the lowest in *C. rogenhoferi* (71.99%) (Table 2). The CR ranged in size from 1143 bp in *O. bimaculatus* to 1718 bp in *O. kanetataki*. Their size differences were mainly attributed to tandem repeats identified in the CR: 2.1 copies of a 562-bp repeat in *C. rogenhoferi*, 2.2 copies of a 525-bp repeat in *M. japonica*, 3.1 copies of a 276-bp repeat and two interspersed 165-bp repeats in *O. bimaculatus*, 4.1 copies of a 298-bp repeat in *O. fuscicercis*, and 3.0 copies of a 158-bp repeat and 4.4 copies of a 235-bp repeat in *O. kanetataki* (Fig. 3). Repeat units for each repeat type were highly conserved within each species (sequence identity >99.1%) but divergent between species (<62.6%).

Multiple IGSs were interspersed throughout mitogenomes with 12 IGSs totaling 55 bp in *M. japonica* and 16 IGSs totaling 320 bp in *O. bimaculatus* (Table S2). The two longest IGSs were located between *rrnL* and *rrnS* (115 bp) and between *rrnS* and *trnV* (112 bp) in *O. bimaculatus*. The 115-bp IGS showed no clear similarity to other regions, whereas the 112-bp IGS together with partial sequences of adjoining genes (42 bp of *rrnS* and 11 bp of *trnV*) was completely homologous to a segment (position 15,018 to 15,182 of the mitogenome sequence) located 83-bp downstream in the CR (Fig. 3). Four common IGSs with varied sizes were present between *trnQ* and *trnM* (8–23 bp), *cox1* and *trnL2* (1–34 bp), *nad4l* and *trnP* (2–5 bp), and *trnS2* and *nad1* (15–17 bp) (Table S2).

3.4. Phylogeny of the superfamily Grylloidea

To test for possible nucleotide substitution saturation, observed substitutions were plotted against GTR-corrected genetic distances (Fig. S1). The observed substitutions increased along genetic distances for the first and second codon positions of PCGs, tRNA genes, *rrnL*, and *rrnS* (slope ≥ 0.70 and regression coefficient $r^2 > 0.99$), indicating little saturation. By contrast, the third codon positions were highly saturated as indicated by the plateau in the scatter plots (slope = 0.25 and $r^2 = 0.74$).

To determine the phylogenetic distribution of gene rearrangements, the phylogeny of the superfamily Grylloidea was reconstructed based on all the 37 genes excluding the saturated third codon positions of PCGs. The BI and ML trees exhibited concordant topologies and strongly recovered the following familial relationships: (((Phalangopsidae + Gryllidae) + Trigonidiidae) + Mogoplistidae) (Fig. 1B).

4. Discussion

We provide here the first complete mitogenome sequences of two families, Mogoplistidae and Phalangopsidae. Comparative mitogenome analysis reveals key structural features, in particular, non-canonical start codons of *cox1* and stop codons of *nad1* as confirmed by transcript evidence, and phylogenetic distribution of gene rearrangements.

4.1. Accurate mitogenome sequencing and assembly

The copy number is at least several hundred-fold higher for mitogenomes than nuclear genomes in animal cells [33], rendering it feasible to assemble mitogenomes from a total DNA library [18]. Even for Orthoptera that has giant nuclear genomes [34–36], the NGS has

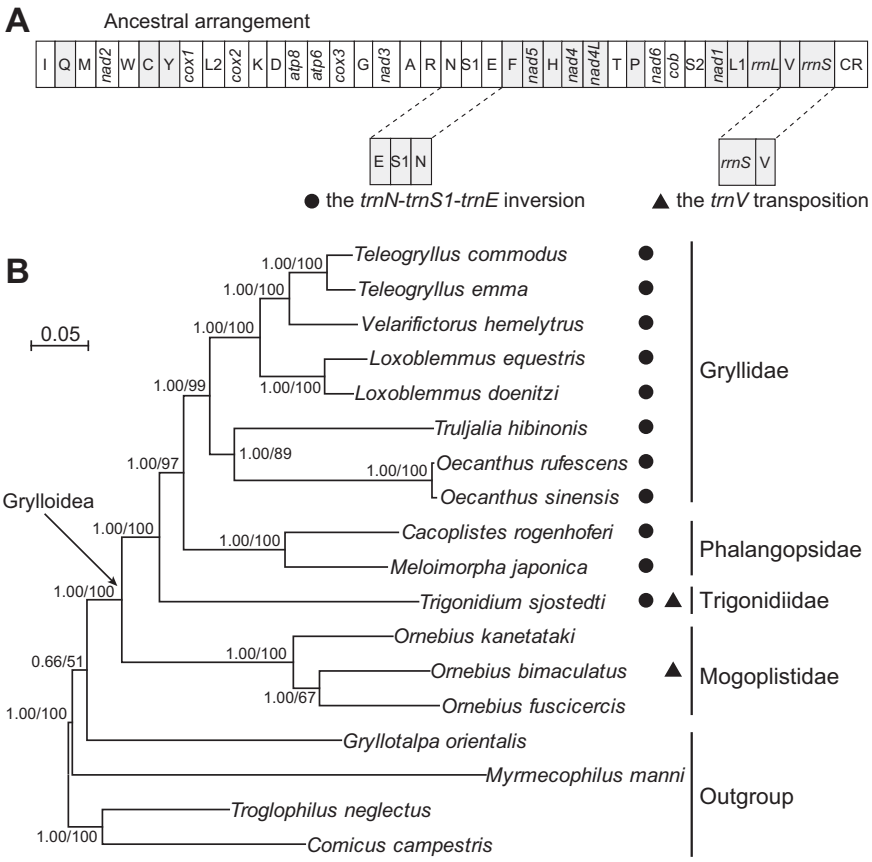

Fig. 1. Linearized representation of mitochondrial gene rearrangements in Grylloidea (A) and their distribution in the Bayesian phylogeny (B). tRNA genes are labeled with the one-letter code for the corresponding amino acid. Genes encoded by the minority strand are shaded. Gene lengths are not to scale. The phylogenetic tree of the superfamily Grylloidea was reconstructed based on all the 37 genes excluding the saturated third codon positions of PCGs. BPP and BS values are shown at nodes.

been successfully applied in mitogenome sequencing [13,14]. In our study, the five mitogenomes were mainly assembled from short-read NGS data (paired-end 2 × 150 bp) of total DNA. The high sequencing depth ranging from 301× to 1050× and further analyses ensure accurate sequencing and assembly of these mitogenomes.

Numt sequences are highly similar to their mitochondrial counterparts and have been reported in most insect orders [37–39]. They cause a substantial threat to accurate interpretation of insect mitogenome data. Examination of the newly assembled five mitogenomes reveals the absence of in-frame stop codons and frame-shift mutations in PCGs, the formation of canonical secondary structures for RNA genes, and similar length for each coding gene. There is hence no evidence for numt contamination in the five mitogenomes.

Repetitive sequences exceeding NGS reads in size usually exert a challenge for de novo assembly of genomes [40–42]. Large repetitive sequences are frequently identified in the CR of insect mitogenomes [7,43]. In such cases, the mitogenomes cannot be determined completely and accurately by short-read NGS alone [44]. To investigate whether the CR contain any repeats in our study, PCR primers have been designed to amplify the CR followed by Sanger sequencing. Indeed, the CR of all the five mitogenomes possesses different sized repeats (Fig. 3), which cannot be spanned by short-reads of NGS. Hence, our study highlights the necessity of combining short-read NGS and Sanger sequencing to obtain a full mitogenome.

4.2. Unusual start codons of cox1 and stop codons of nad1

Determination of start codons of *cox1* in most insects is a long-standing issue since no canonical (ATN) start codons are present in the beginning of the open reading frame downstream *trnY*. As a consequence, unusual start codons of *cox1* have been proposed, including triplets, which encode arginine, asparagine, glutamine, leucine, isoleucine, and serine, and even 4-bp nucleotides [7,22,45–47]. Among these speculative start codons, only CGA encoding arginine in *Maruca vitrata* [48] and TCG encoding serine in *Drosophila melanogaster* [49] have been proved to be the first triplets of their *cox1* transcripts. In our study, the 5′ RACE experiment demonstrates that TCG in *M. japonica* and CCG in *O. fuscicercis* and *O. kanetataki* are the first triplets for the *cox1* transcripts (Fig. 2A). The corresponding triplet in the homologous sequence alignment is TCG in *C. rogenhoferi* and CTG in *O. bimaculatus*. They all have a 2-bp overlap with the upstream *trnY* encoded by the opposite strand (Fig. 2A, Table S2). If *cox1* starts at the very 5′ terminal nucleotide of its transcript as in *D. melanogaster* [49], these triplets should serve as the start codons. Nevertheless, it should be noted that very short (1–3) untranslated nucleotides have been reported in the 5′ end of human mitochondrial mRNAs [50]. The proposed start codons of *cox1* need therefore to be validated by protein sequencing.

Like other insect mitogenomes, most PCGs in our newly determined mitogenomes terminate with either complete stop codons (TAA or TAG) or partial stop codons (T or TA) immediately followed by a tRNA gene (Table S2). These partial stop codons can be converted into a complete stop codon (TAA) by polyadenylation after tRNA cleavage [49,51]. However, stop codons for *nad1* in the three *Ornebius* species are an exception; neither complete stop codons before the downstream *trnS2* nor partial stop codons immediately followed by *trnS2* are present (Table S2). Our 3′ RACE experiment demonstrates that *nad1* transcript ends with U plus polyA (Fig. 2B), which is thought to arise from a partial stop codon (U) followed by polyadenylation during mRNA maturation. The observed intergenic spacer between *nad1* and *trnS2* shows sequence conservation among the five species (Fig. 2B) and even across

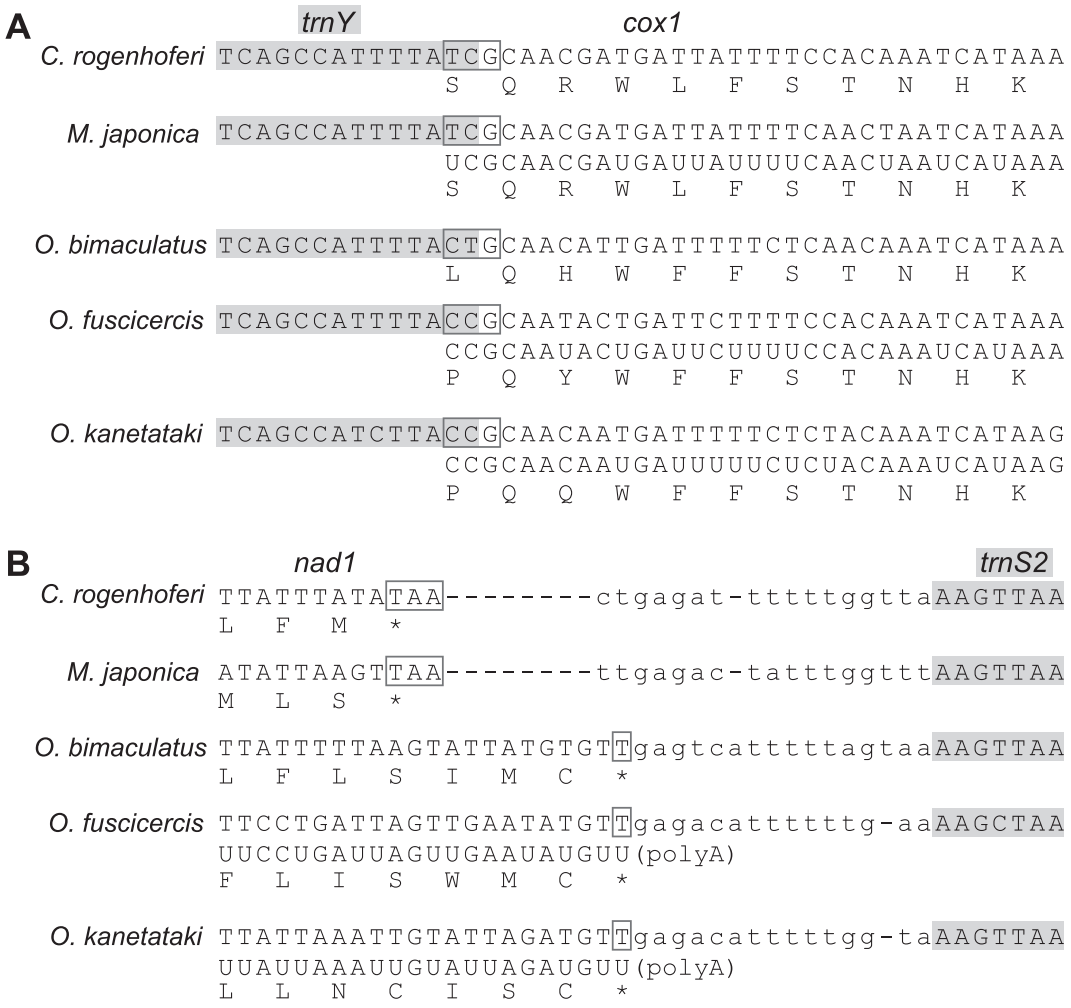

Fig. 2. Sequence alignments for the beginning of *cox1* (A) and the end of *nad1* (B). Adjacent tRNA genes are highlighted in grey. Proposed start codons of *cox1* and stop codons of *nad1* are boxed. The intergenic spacers between *nad1* and *trnS* are shown in lowercase. The mRNA sequences of the 5′ end of *cox1* (*M. japonica*, *O. fuscicercis*, and *O. kanetataki*) and the 3′ end of *nad1* (*O. fuscicercis* and *O. kanetataki*) obtained via RACE are displayed. For *cox1* and *nad1*, a translation is provided based on the invertebrate mitochondrial genetic code.

insects [52]. The conservation of this intergenic spacer may be maintained by its function as a binding site of a transcription termination factor [53]. It is probable that transcription termination signals contained in the intergenic spacer mediate the precise cleavage at the partial stop codon (U). The novel strategy for *nad1* termination adopted by *Ornebius* species has yet to be elucidated.

4.3. Phylogenetic distribution of gene rearrangements

A robust phylogeny represents an essential prerequisite to decipher phylogenetic distribution and evolutionary history of gene rearrangements. The BI and ML analyses in our study generate concordant tree topologies with strong support values (Fig. 1B) and recover the currently accepted familial relationships within Grylloidea [5]. The tree topologies can be used to map gene rearrangements.

The arrangement of the 37 genes is highly conserved in most insect mitogenomes and the prevalent one is regarded to be the ancestral form for insects [6]. A wide variety of gene rearrangements mainly involving tRNA genes are described in insects including the superfamily Grylloidea, which has been reported to possess two types of gene rearrangements [11,13–15]. The local inversion of the ancestral *trnN-trnS1-trnE* to *trnE-trnS1-trnN* has been previously characterized in Gryllidae [11,14,15] and Trigonidiidae [13] and is now extended to Phalangopsidae represented by *C. rogenhoferi* and *M. japonica* (Fig. 1A and B, Table S2). This inversion does not occur in Mogoplistidae represented by the three *Ornebius* species (Fig. 1A and B, Table S2). This rearrangement seems to be a synapomorphy of Gryllidae, Phalangopsidae, and Trigonidiidae. The gene rearrangement distribution pattern is consistent with the hypothesis that Mogoplistidae is sister to all other families in Grylloidea as revealed by our mitogenome sequence-based phylogenetic analysis (Fig. 1B) and a previous molecular study [5]. These findings suggest that the inversion event happened in the common ancestor of Gryllidae, Phalangopsidae, and Trigonidiidae after Mogoplistidae had diverged from other lineages of Grylloidea. With regard to Pteroplistinae, the unclassified subfamily in Grylloidea, no representative mitogenome has been sequenced to date. However, it

Table 2
Nucleotide compositions (in percentage).

Gene/region	*C. rogenhoferi*	*M. japonica*	*O. bimaculatus*	*O. fuscicercis*	*O. kanetataki*
Whole genome	73.31	72.43	76.47	74.95	74.24
tRNA genes	75.38	75.28	79.07	77.13	77.24
1st codon	68.80	66.52	69.67	67.80	67.83
2nd codon	66.30	65.83	68.06	67.02	66.54
3rd codon	83.89	82.51	87.17	84.98	84.46
rrnS	73.00	72.88	74.28	75.36	72.81
rrnL	75.12	74.82	78.45	77.30	76.93
Control region	71.99	72.97	84.78	82.02	78.23

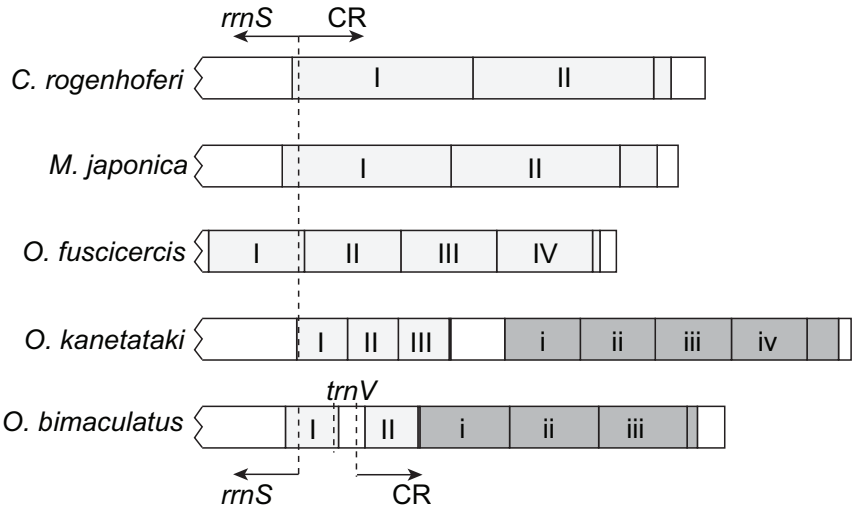

Fig. 3. Repetitive sequences in the CR and adjacent regions. Gene lengths are drawn to scale. Dashed lines represent gene boundaries of *trnV* in *O. bimaculatus* and *rrnS* in all. Two repeats are present in *O. bimaculatus* (165 bp × 2.0 copies and 276 bp × 3.1 copies) and *O. kanetataki* (158 bp × 3.0 copies and 235 bp × 4.4 copies) and one in the others (*C. rogenhoferi*: 562 bp × 2.1 copies; *M. japonica*: 525 bp × 2.2 copies; *O. fuscicercis*: 298 bp × 4.1 copies).

could be expected that this rearrangement occurs in Pteroplistinae, considering its sister-group relationship with Gryllidae + Phalangopsidae [5]. For inversions, the "intramitochondrial recombination" mechanism that involves the breakage and reconnection of DNA double strands in a reverse direction has been proposed [54]. This mechanism has been adopted to explain the *trnN-trnS1-trnE* inversion in Gryllidae [11,15].

Our study demonstrates a transposition of *trnV* to the site between *rrnS* and the CR in *O. bimaculatus* and an ancestral gene order in *O. fuscicercis* and *O. kanetataki* (Fig. 1A and B, Table S2). The transposition event could therefore be proposed to occur after the separation of *O. bimaculatus* from its congeneric species. The *trnV* transposition has also been reported in *T. sjostedti* from the family Trigonidiidae [13]. Such a phylogenetic distribution pattern in these taxa reveals that the *trnV* transpositions are not caused by a single rearrangement event; rather, they have occurred independently and are a potential consequence of parallel evolution. This transposition is thus not reliable for phylogenetic inference. Such gene transpositions are widely accepted to be explained by the "tandem duplication/random loss" model, which involves tandem duplication of mitochondrial genes followed by random loss of one of the duplicated copies [6,55].

5. Conclusion

In the present study, we provide five representative mitogenome sequences of two families, Mogoplistidae and Phalangopsidae. Non-canonical start codons of *cox1* and, in *Ornebius*, a partial stop codon (T) of *nad1* that is not immediately followed by a tRNA gene are confirmed by transcript information. The phylogenetic trees based on mitogenome data support the following familial relationships: (((Phalangopsidae + Gryllidae) + Trigonidiidae) + Mogoplistidae). The *trnE-trnS1-trnN* rearrangement occurs in Phalangopsidae, Gryllidae, and Trigonidiidae but not in Mogoplistidae, whereas the *trnV* transposition is present in only one *Ornebius* species (Phalangopsidae) and *T. sjostedti* (Trigonidiidae). The phylogenetic distribution patterns suggest that the *trnE-trnS1-trnN* rearrangement event happened in the common ancestor of the three other families after the separation of Mogoplistidae and that the *trnV* transposition is a potential consequence of parallel evolution. This study uncovers phylogenetic distribution patterns and evolutionary history of gene rearrangements in Grylloidea.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2018.08.181.

Competing interest

The authors declare no conflict of interest.

Acknowledgements

The study was funded by the National Natural Science Foundation of China (31400325), the Natural Science Foundation of Beijing (5182031), the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2015-IAR), and the earmarked fund for Modern Agro-Industry Technology Research System (CARS-44) in China. We thank the two anonymous reviewers for their comments and suggestions on earlier versions of this manuscript.

References

[1] M.M. Cigliano, H. Braun, D.C. Eades, D. Otte, Orthoptera species file. Version 5.0/5.0, http://Orthoptera.SpeciesFile.org 2018, Accessed date: 11 July 2018.
[2] E.M. Costa-Neto, Entertainment with insects: singing and fighting insects around the world. A brief review, Etnobiología 3 (1) (2003) 20–28.
[3] H.J. Song, C. Amedegnato, M.M. Cigliano, L. Desutter-Grandcolas, S.W. Heads, Y. Huang, D. Otte, M.F. Whiting, 300 million years of diversification: elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling, Cladistics 31 (6) (2015) 621–651.
[4] L. Desutter-Grandcolas, Phylogeny and the evolution of acoustic communication in extant Ensifera (Insecta, Orthoptera), Zool. Scr. 32 (6) (2003) 525–561.
[5] I.C. Chintauan-Marquier, F. Legendre, S. Hugel, T. Robillard, P. Grandcolas, A. Nel, D. Zuccon, L. Desutter-Grandcolas, Laying the foundations of evolutionary and systematic studies in crickets (Insecta, Orthoptera): a multilocus phylogenetic analysis, Cladistics 32 (1) (2016) 54–81.
[6] S.L. Cameron, Insect mitochondrial genomics: implications for evolution and phylogeny, Annu. Rev. Entomol. 59 (2014) 95–117.
[7] S.L. Cameron, How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research, Syst. Entomol. 39 (3) (2014) 400–411.
[8] C. Simon, T.R. Buckley, F. Frati, J.B. Stewart, A.T. Beckenbach, 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. 37 (2006) 545–579.
[9] J.L. Boore, Animal mitochondrial genomes, Nucleic Acids Res. 27 (8) (1999) 1767–1780.
[10] S. Simon, H. Hadrys, A comparative analysis of complete mitochondrial genomes among Hexapoda, Mol. Phylogenet. Evol. 69 (2) (2013) 393–403.
[11] J. Yang, Q.L. Ren, Q. Zhang, Y. Huang, Complete mitochondrial genomes of three crickets (Orthoptera: Gryllidae) and comparative analyses within Ensifera mitogenomes, Zootaxa 4092 (4) (2016) 529–547.
[12] J.N. Wolff, D.C.A. Shearman, R.C. Brooks, J.W.O. Ballard, Selective enrichment and sequencing of whole mitochondrial genomes in the presence of nuclear encoded mitochondrial pseudogenes (numts), PLoS ONE 7 (5) (2012).
[13] N. Song, H. Li, F. Song, W.Z. Cai, Molecular phylogeny of Polyneoptera (Insecta) inferred from expanded mitogenomic data, Sci. Rep. 6 (2016).

[14] Z. Zhou, L. Zhao, N. Liu, H. Guo, B. Guan, J. Di, F. Shi, Towards a higher-level Ensifera phylogeny inferred from mitogenome sequences, Mol. Phylogenet. Evol. 108 (2017) 22–33.
[15] W. Ye, J. Dang, L. Xie, Y. Huang, Complete mitochondrial genome of *Teleogryllus emma* (Orthoptera: Gryllidae) with a new gene order in Orthoptera, Zool. Res. 29 (3) (2008) 236–244.
[16] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (2014) 2114–2120.
[17] Y.L. Xie, G.X. Wu, J.B. Tang, et al., SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads, Bioinformatics 30 (12) (2014) 1660–1666.
[18] C. Hahn, L. Bachmann, B. Chevreux, Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-a baiting and iterative mapping approach, Nucleic Acids Res. 41 (13) (2013).
[19] H. Li, B. Handsaker, A. Wysoker, et al., The sequence alignment/map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079.
[20] M. Bernt, A. Donath, F. Juhling, F. Externbrink, C. Florentz, G. Fritzsch, J. Putz, M. Middendorf, P.F. Stadler, MITOS: improved *de novo* metazoan mitochondrial genome annotation, Mol. Phylogenet. Evol. 69 (2) (2013) 313–319.
[21] D.V. Lavrov, J.L. Boore, W.M. Brown, The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*, Mol. Biol. Evol. 17 (5) (2000) 813–824.
[22] N.C. Sheffield, H. Song, L. Cameron, M.F. Whiting, A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles, Mol. Biol. Evol. 25 (11) (2008) 2499–2509.
[23] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, Nucleic Acids Res. 27 (2) (1999) 573–580.
[24] T. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41 (1999) 95–98.
[25] X.H. Xia, DAMBE6: new tools for microbial genomics, phylogenetics and molecular evolution, J. Hered. 108 (4) (2017) 431–437.
[26] W. Minxiao, S. Song, L. Chaolun, S. Xin, Distinctive mitochondrial genome of Calanoid copepod *Calanus sinicus* with multiple large non-coding regions and reshuffled gene order: useful molecular markers for phylogenetic and population studies, BMC Genomics 12 (2011) 73.
[27] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (5) (2004) 1792–1797.
[28] M. Suyama, D. Torrents, P. Bork, PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments, Nucleic Acids Res. 34 (Suppl. 2) (2006) W609–W612.
[29] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (4) (2000) 540–552.
[30] R. Lanfear, P.B. Frandsen, A.M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses, Mol. Biol. Evol. 34 (3) (2017) 772–773.
[31] F. Ronquist, M. Teslenko, P. van der Mark, et al., MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, Syst. Biol. 61 (3) (2012) 539–542.
[32] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (9) (2014) 1312–1313.
[33] R.N. Lightowlers, P.F. Chinnery, D.M. Turnbull, N. Howell, Mammalian mitochondrial genetics: heredity, heteroplasmy and disease, Trends Genet. 13 (11) (1997) 450–455.
[34] S.J. Hanrahan, J.S. Johnston, New genome size estimates of 134 species of arthropods, Chromosom. Res. 19 (6) (2011) 809–823.
[35] F. Dufresne, N. Jeffery, A guided tour of large genome size in animals: what we know and where we are heading, Chromosom. Res. 19 (7) (2011) 925–938.
[36] T.R. Gregory, Animal genome size database, http://www.genomesize.com 2018, Accessed date: 11 July 2018.
[37] E. Richly, D. Leister, NUMTs in sequenced eukaryotic genomes, Mol. Biol. Evol. 21 (6) (2004) 1081–1084.
[38] P. Pamilo, L. Viljakainen, A. Vihavainen, Exceptionally high density of NUMTs in the honeybee genome, Mol. Biol. Evol. 24 (6) (2007) 1340–1346.
[39] L.A.R. Leite, Mitochondrial pseudogenes in insect DNA barcoding: differing points of view on the same issue, Biota Neotrop. 12 (3) (2012) 301–308.
[40] S. Haridas, C. Breuill, J. Bohlmann, T. Hsiang, A biologist's guide to *de novo* genome assembly using next-generation sequence data: a test with fungal genomes, J. Microbiol. Methods 86 (3) (2011) 368–375.
[41] C. Lindqvist, S.C. Schuster, Y. Sun, et al., Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear, Proc. Natl. Acad. Sci. U. S. A. 107 (11) (2010) 5053–5057.
[42] K. Reinert, B. Langmead, D. Weese, D.J. Evers, Alignment of next-generation sequencing reads, Annu. Rev. Genomics Hum. Genet. 16 (2015) 133–151.
[43] D.X. Zhang, G.M. Hewitt, Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies, Biochem. Syst. Ecol. 25 (2) (1997) 99–120.
[44] S. Goodwin, J.D. McPherson, W.R. McCombie, Coming of age: ten years of next-generation sequencing technologies, Nat. Rev. Genet. 17 (6) (2016) 333–351.
[45] N.C. Sheffield, K.D. Hiatt, M.C. Valentine, H.J. Song, M.F. Whiting, Mitochondrial genomics in Orthoptera using MOSAS, Mitochondrial DNA 21 (3–4) (2010) 87–104.
[46] C. Ma, C. Liu, P. Yang, L. Kang, The complete mitochondrial genomes of two band-winged grasshoppers, *Gastrimargus marmoratus* and *Oedaleus asiaticus*, BMC Genomics 10 (2009) 156.
[47] J.W.O. Ballard, Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup, J. Mol. Evol. 51 (1) (2000) 48–63.
[48] V.M. Margam, B.S. Coates, R.L. Hellmich, et al., Mitochondrial genome sequence and expression profiling for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae), PLoS ONE 6 (2) (2011).
[49] J.B. Stewart, A.T. Beckenbach, Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods, Gene 445 (1–2) (2009) 49–57.
[50] R.J. Temperley, M. Wydro, R.N. Lightowlers, Z.M. Chrzanowska-Lightowlers, Human mitochondrial mRNAs-like members of all families, similar but different, Biochim. Biophys. Acta, Bioenerg. 1797 (6–7) (2010) 1081–1085.
[51] D. Ojala, J. Montoya, G. Attardi, tRNA punctuation model of RNA processing in human mitochondria, Nature 290 (5806) (1981) 470–474.
[52] S.L. Cameron, M.F. Whiting, 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 (1–2) (2008) 112–123.
[53] M. Roberti, P.L. Polosa, F. Bruni, C. Musicco, M.N. Gadaleta, P. Cantatore, DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences of *Drosophila melanogaster* mitochondrial DNA, Nucleic Acids Res. 31 (6) (2003) 1597–1604.
[54] A.D. Miller, T.T.T. Nguyen, C.P. Burridge, C.M. Austin, Complete mitochondrial DNA sequence of the Australian freshwater crayfish, *Cherax destructor* (Crustacea: Decapoda: Parastacidae): a novel gene order revealed, Gene 331 (2004) 65–72.
[55] F. Juhling, J. Putz, M. Bernt, A. Donath, M. Middendorf, C. Florentz, P.F. Stadler, Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements, Nucleic Acids Res. 40 (7) (2012) 2833–2845.