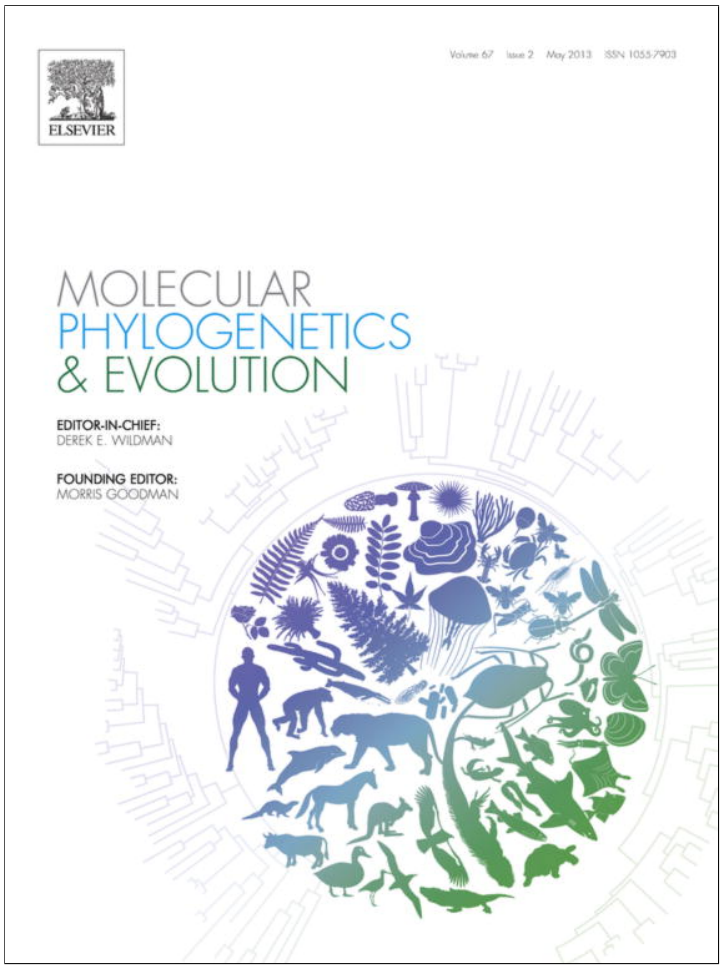

Molecular Phylogenetics and Evolution 67 (2013) 494–508

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Searching for the optimal data partitioning strategy in mitochondrial phylogenomics: A phylogeny of Acridoidea (Insecta: Orthoptera: Caelifera) as a case study

James R. Leavitt [a], Kevin D. Hiatt [a], Michael F. Whiting [a], Hojun Song [a,b,*]

[a] Department of Biology and M.L. Bean Museum, Brigham Young University, Provo, UT 84602, USA
[b] Department of Biology, University of Central Florida, Orlando, FL 32816, USA

ARTICLE INFO

Article history:
Received 1 May 2012
Revised 10 February 2013
Accepted 15 February 2013
Available online 27 February 2013

Keywords:
Phylogeny
Mitochondria
Phylogenomics
Data partition
Orthoptera
Caelifera

ABSTRACT

One of the main challenges in analyzing multi-locus phylogenomic data is to find an optimal data partitioning strategy to account for variable evolutionary histories of different loci for any given dataset. Although a number of studies have addressed the issue of data partitioning in a Bayesian phylogenetic framework, such studies in a maximum likelihood framework are comparatively lacking. Furthermore, a rigorous statistical exploration of possible data partitioning schemes has not been applied to mitochondrial genome (mtgenome) data, which provide a complex, but manageable platform for addressing various challenges in analyzing phylogenomic data. In this study, we investigate the issue of data partitioning in the maximum likelihood framework in the context of the mitochondrial phylogenomics of an orthopteran superfamily Acridoidea (Orthoptera: Caelifera). The present study analyzes 34 terminals representing all 8 superfamilies within Caelifera, which includes newly sequenced partial or complete mtgenomes for 11 families. Using a new partition-selection method implemented in the software PartitionFinder, we compare a large number of data partitioning schemes in an attempt to identify the most effective method of analyzing the mtgenome data. We find that the best-fit partitioning scheme selected by PartitionFinder is superior to any *a priori* schemes commonly utilized in mitochondrial phylogenomics. We also show that over-partitioning is often detrimental to phylogenetic reconstruction. A comparative analysis of mtgenome structures finds that the tRNA gene rearrangement between cytochrome *c* oxidase subunit II and ATP synthase protein 8 does not occur in the most basal caeliferan lineage Tridactyloidea, suggesting that this gene rearrangement must have evolved at least in the common ancestor of Tetrigoidea and Acridomorpha. We find that mtgenome data contain sufficient phylogenetic information to broadly resolve the relationships across Acridomorpha and Acridoidea.

1. Introduction

As sequencing technologies advance at a remarkably rapid rate, the gap between the ability to generate data and the ability to analyze the data in a phylogenetic framework is also increasing rapidly (Delsuc et al., 2005; Philippe et al., 2005, 2004). Mitochondrial phylogenomics, the use of complete mitochondrial genomes (mtgenomes) in phylogenetics, provides a manageable platform for addressing various challenges in analyzing phylogenomic data (Song et al., 2010). A typical insect mtgenome contains 37 genes in a coding region: 13 protein-coding genes (PCGs), two ribosomal RNA genes (16S and 12S) and 22 transfer RNA genes (Wolstenhome, 1992). A number of recent studies have explored various ways of effectively analyzing mtgenome data for inferring phylogenetic relationships among insects (Cameron et al., 2006a, 2006b, 2007, 2009; Castro and Dowton, 2007; Dowton et al., 2009a; Fenn et al., 2008; Pons et al., 2010; Sheffield et al., 2009; Song et al., 2010). Specifically, the issues of gene exclusion (Cameron et al., 2006b; Nardi et al., 2003), data recoding (Cameron et al., 2007; Fenn et al., 2008), data partitioning (Fenn et al., 2008), base compositional bias (Sheffield et al., 2009; Song et al., 2010), and rate heterogeneity (Pons et al., 2010; Song et al., 2010) have been addressed in the context of insect mitochondrial phylogenomics.

Cameron et al. (2004) was the first to suggest that the inclusion of all available data from mtgenomes improved the resolution and support and Cameron et al. (2007) found that biologically realistic data partitioning would produce the best results. Fenn et al. (2008) echoed the same sentiment but warned that over-partitioning of the mtgenome might be more detrimental to the analysis than beneficial. However, there has not been a clear recommendation on how to decide which partitioning method is the best for

* Corresponding author. Address: Department of Biology, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816-2368, USA. Fax: +1 407 823 5769.
E-mail address: song@ucf.edu (H. Song).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.02.019

mitochondrial phylogenomics in general although a number of studies have tackled this very issue in multi-locus analyses of nuclear genes (Brandley et al., 2005; Brown and Lemmon, 2007; Li et al., 2008; McGuire et al., 2007). In deciding the optimal partitioning strategy, one typically compares several datasets differing in partitioning schemes using methods such as Bayes factor (Brandley et al., 2005; Brown and Lemmon, 2007; Nylander et al., 2004), cluster analysis (Li et al., 2008), or the Akaike Information Criterion (AIC) (McGuire et al., 2007). However, a general consensus appears to be that the optimal partitioning strategy depends on both taxon and character sampling and it is difficult to justify a certain partitioning strategy for a particular dataset *a priori*. Recently, Lanfear et al. (2012) developed a method that could overcome this issue by statistically comparing across numerous partitioning schemes for any given sequence data and selecting the best-fit scheme, which was implemented in the program PartitionFinder. They found that typical *a priori* partitioning schemes, such as partitioning by gene or codon position, were often not the most appropriate way of partitioning data across various empirical datasets that they examined (Lanfear et al., 2012). According to Lanfear et al. (2012), the number of possible partitioning schemes follows a relationship known as a Bell number (Bell, 1934). For mitochondrial genome data, this number would be $B_{37} = 5.28 \times 10^{31}$ if 37 genes are treated as individual partitions, and could increase to $B_{63} = 8.25 \times 10^{63}$ if protein-coding genes are further partitioned by codon position. To our knowledge, this new method has not been fully examined in mitochondrial phylogenomics.

Acridoidea (grasshoppers and locusts) is the largest superfamily in the insect order Orthoptera with over 7600 described species and eleven families: Acrididae, Charilaidae, Dericorythidae, Lathiceridae, Lentulidae, Lithidiidae, Ommexechidae, Pamphagidae, Pyrgacrididae, Romaleidae, and Tristiridae. Acridoidea is one of six currently recognized superfamilies that are grasshopper-like in morphology and have been traditionally grouped together and collectively referred to as the Acridomorpha (Dirsh, 1975; Song, 2010). Recent molecular studies (Fenn et al., 2008; Flook et al., 1999, 2000; Flook and Rowell, 1997; Matt et al., 2008) have consistently found this group to be monophyletic although the phylogenetic relationships among the superfamilies remain contentious (Eades, 2000; Flook and Rowell, 1997; Song, 2010). Acridoidea has been considered a monophyletic group based on the morphology of male genitalia and the lack of a basioccipital slit, among other characters (Amédégnato, 1974; Eades, 2000; Kevan, 1982; Roberts, 1941), but the phylogenetic relationships among families within this clade are not well understood (Song, 2010). This is mainly because the higher-level classification of Acridoidea (and Acridomorpha) has been largely influenced by the interpretation of internal male phallic structures, which has not been consistent among different taxonomists (Amédégnato, 1976; Dirsh, 1956; Eades, 2000; Roberts, 1941). Even the latest synthesis on the phylogeny of Acridomorpha (Eades, 2000) depicted the internal relationships of Acridoidea to be largely unresolved. Therefore, this study represents a unique opportunity to reassess the phylogeny of Acridoidea based on a different character system.

The migratory locust, *Locusta migratoria*, was the first hemimetabolous insect to have its mtgenome completely sequenced (Flook et al., 1995a), and the follow-up study (Flook et al., 1995b) suggested that grasshoppers exhibit a unique gene rearrangement within the mtgenome. In grasshoppers, the order of tRNA genes between cytochrome *c* oxidase subunit II (COII) and ATP synthase protein 8 (ATP8) is reversed from the ancestral insect arrangement, inferred from *Drosophila yakuba* (Clary and Wolstenhome, 1985), which exhibits tRNA-Lys preceding tRNA-Asp on the J strand. Flook et al. (1995b) hypothesized that this rearrangement was a unique condition for Caelifera based on an investigation of 5 orthopteran species, and Fenn et al. (2008) supported this hypothesis based on an analysis of 8 orthopteran mtgenomes. Sheffield et al. (2010) recently suggested that this gene rearrangement is a synapomorphy for a subgroup within the orthopteran suborder Caelifera instead of the suborder as a whole. There has been a rapid increase in the number of complete mtgenomes of Acridomorpha available on GenBank (Ding et al., 2007; Erler et al., 2010; Fenn et al., 2008; Liu and Huang, 2008; Ma et al., 2009; Sheffield et al., 2010; Zhang and Huang, 2008), as well as from our ongoing study of orthopteran mtgenomes. Taken all together, grasshoppers represent an excellent system for addressing the various challenges of analyzing mtgenome data in a phylogenetic framework, as well as for understanding the evolution of mtgenomes.

In this study we explore various partitioning schemes using PartitionFinder in the context of the mtgenome phylogeny of the grasshopper superfamily Acridoidea (Orthoptera: Caelifera) in a maximum likelihood framework. We specifically address the following questions: (i) What is the optimal data partitioning strategy in analyzing mtgenome data in a maximum likelihood framework for Acridoidea?; (ii) What are the patterns of mtgenome evolution in Acridoidea in terms of gene rearrangement, anti-codon conservation, and start and stop codon usage?; and (iii) What are the higher-level phylogenetic relationships within Acridoidea and Acridomorpha and how do these compare with previous hypotheses?

2. Materials and methods

2.1. Taxon sampling

We included nine of the 11 extant families within Acridoidea (Table 1). The two omitted families, Dericorythidae and Lathiceridae, which are found in the Middle East and the Namibian desert, respectively, are rare and we were not able to obtain DNA-grade tissues. For Acrididae, which is the largest family within Acridoidea, we included 16 terminals representing 8 subfamilies. For the remaining 8 families, we included one representative per family, except Pamphagidae, for which we included two taxa. Additionally, we included at least one representative for each of the acridomorph superfamilies (Eumastacoidea, Tanaoceroidea, Pneumoroidea, Trigonopterygoidea, and Pyrgomorphoidea) in order to determine the phylogenetic placement of Acridoidea as well as to test the monophyly of Acridoidea within Acridomorpha. For outgroups, we included one member of Tridactyloidea and Tetrigoidea, both of which have been traditionally considered basal lineages within Caelifera. Altogether, our analysis included 34 terminals representing all 8 superfamilies within Caelifera.

For this study, we generated seven complete and four partial mtgenomes representing six superfamilies, 11 families. Tissue samples used for data generation were either collected by the authors or provided by collaborators. The specimens used in this analysis were preserved in 100% ethanol and vouchered to the −80 °C cryofacility in the Insect Genomic Collection at Brigham Young University (BYU-IGC). The voucher information is presented in Supplementary Table S1. The remaining data were obtained from GenBank.

2.2. Character sampling

DNA was extracted from tissues from hind femur or thorax using the Qiagen DNeasy Tissue Extraction Kit (Valencia, California, USA) following the animal tissue protocol. The resulting DNA extracts were vouchered at the BYU-IGC. Mtgenome data were generated using a primer-walking technique. Initial long-range polymerase chain reactions (PCRs) were performed using both universal mtgenome primers (Simon et al., 2006) as well as Orthoptera specific primers, which were the same as those used by

Table 1
Taxonomic information, GenBank accession numbers, and references for the taxa included in this study. Detailed voucher information is found in Supplementary Table S1.

Superfamily	Family	Subfamily	Species	GenBank #	Reference
Tridactyloidea	Tridactylidae	Tridactylinae	*Ellipes minuta*	GU945502	Sheffield et al. (2010)
Tetrigoidea	Tetrigidae	–	*Trachytettix horridus*	JX913766	Present study
Eumastacoidea	Eumastacidae	Paramastacinae	*Paramastax nigra*	JX913772	Present study
Tanaoceroidea	Tanaoceridae	Tanaocerinae	*Tanaocerus kobelei*	JX913774	Present study
Pneumoroidea	Pneumoridae	–	*Physemacris variolosa*	GU945504	Sheffield et al. (2010)
Trigonopterygoidea	Trigonopterygidae	Trigonopteryginae	*Trigonopteryx hopei*	JX913767	Present study
Pyrgomorphoidea	Pyrgomorphidae	Pyrgomorphinae	*Atractomorpha sinensis*	EU263919	Ding et al. (2007)
Pyrgomorphoidea	Pyrgomorphidae	Pyrgomorphinae	*Mekongiella xizangensis*	HM583654	Zhao et al. (2010)
Pyrgomorphoidea	Pyrgomorphidae	Pyrgomorphinae	*Mekongiana xiangchengensis*	HM583653	Zhao et al. (2010)
Acridoidea	Charilaidae	–	*Hemicharilaus monomorphus*	JX913773	Present study
Acridoidea	Pamphagidae	Thrinchinae	*Thrinchus schrenkii*	GU181288	Zhang et al. (2011)
Acridoidea	Pamphagidae	Thrinchinae	*Prionotropis hystrix*	JX913764	Present study
Acridoidea	Pyrgacrididae	Pyrgacridinae	*Pyrgacris descampsi*	JX913771	Present study
Acridoidea	Lithidiidae	Lithidiinae	*Lithidiopsis carinatus*	JX913770	Present study
Acridoidea	Lentulidae	Lentulinae	*Lentula callani*	JX913769	Present study
Acridoidea	Tristiridae	Tristirinae	*Tristira magellanica*	JX913765	Present study
Acridoidea	Ommexechidae	Ommexechinae	*Ommexecha virens*	JX913775	Present study
Acridoidea	Romaleidae	Romaleinae	*Xyleus modestus*	GU945503	Sheffield et al. (2010)
Acridoidea	Acrididae	Calliptaminae	*Calliptamus italicus*	EU938374	Fenn et al. (2008)
Acridoidea	Acrididae	Cyrtacanthacridinae	*Schistocerca gregaria gregaria*	GQ491031	Erler et al. (2010)
Acridoidea	Acrididae	Catantopinae	*Traulia szetschuanensis*	EU914849	Zhang and Huang (2008)
Acridoidea	Acrididae	Oxyinae	*Oxya chinensis*	EF437157	Zhang and Huang (2008)
Acridoidea	Acrididae	Melanoplinae	*Prumna arctica*	GU294758	Sun et al. (2010)
Acridoidea	Acrididae	Melanoplinae	*Ognevia longipennis*	EU914848	Zhang and Huang (2008)
Acridoidea	Acrididae	Acridinae	*Phlaeoba albonema*	EU370925	Shi et al. (2008)
Acridoidea	Acrididae	Acridinae	*Acrida willemsei*	EU938372	Fenn et al. (2008)
Acridoidea	Acrididae	Oedipodinae	*Oedaleus decorus asiaticus*	EU513374	Ma et al. (2009)
Acridoidea	Acrididae	Oedipodinae	*Gastrimargus marmoratus*	EU513373	Ma et al. (2009)
Acridoidea	Acrididae	Oedipodinae	*Locusta migratoria*	X80245	Flook et al. (1995a)
Acridoidea	Acrididae	Gomphocerinae	*Arcyptera coreana*	GU324311	Huang and Liu (unpublished)
Acridoidea	Acrididae	Gomphocerinae	*Euchorthippus fusigeniculatus*	HM583652	Zhao et al. (2010)
Acridoidea	Acrididae	Gomphocerinae	*Chorthippus chinensis*	EU029161	Liu and Huang (2008)
Acridoidea	Acrididae	Gomphocerinae	*Gomphocerippus rufus*	GU294759	Sun et al. (2010)
Acridoidea	Acrididae	Gomphocerinae	*Gomphocerus licenti*	GQ180102	Gao and Huang (unpublished)

Sheffield et al. (2010) to amplify overlapping 8 kb segments (see Supplementary Table S2 for primer details). Each long-range PCR was performed using Elongase (Invitrogen, Carlsbad, CA, USA) and the following amplification profile (2 min/92 °C hot start; 35 cycles of 30 s/92 °C, 30 s/50 °C, 7 min/68 °C). Long-range PCR was used specifically to avoid coamplification of nuclear mitochondrial pseudogenes (numts), which are prevalent in Orthoptera (Bensasson et al., 2000; Moulton et al., 2010; Song et al., 2008). Using the long-range PCR products as templates, nested PCR was performed to amplify ~3 kb segments with the identical PCR profile used for the long-range PCR except for a shorter extension time (4 min) and 20 cycles instead of 35 cycles. When necessary, both the annealing temperature and the extension time were modified in accordance with the melting temperature of the primers being used and the length of the target segment. Species-specific primers were designed to primer-walk each mtgenome to ensure 2× coverage of both the major and minor strands (species-specific primers used available upon request). Successful PCR products were cleaned using PrepEase 96-well Cleaning Kit plates (USB Corporation, Cleveland, OH, USA). Sequencing was performed using ABI Big Dye ver3.0 terminating technology on an ABI 3770 capillary sequencer (BYU Sequencing Center, Provo, UT, USA). Sequences were proofread in Sequencher™ ver. 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA).

A number of recent studies have explored different methods of sequencing mtgenomes using the next-generation sequencing (NGS) technology (Jex et al., 2008; Mikkelsen et al., 2009; Timmermans et al., 2010). One of the mtgenomes included in this study (*Tanaocerus kobelei*) was partly generated by shotgun-sequencing using Roche 454 pyrosequencing technology (GS FLX system). While previous studies used long-range PCR products from mtDNA as templates, we used genomic DNA (gDNA) as templates. Because mtDNA is abundantly present in gDNA, the shotgun sequencing was able to generate mtgenome sequences as by-products. This mtgenome was generated as a part of a single 454 run which included 7 taxa in 3/4 of a plate (Song, unpublished data). A pool of 7 taxa was ligated with Ligation Multiplex Identifiers (MID) supplied by 454, which was later recognized and analyzed by the Newbler software, and contigs were assembled de novo. Average contig size is about 650 bp, with the largest contig size of ~15,000 bp, which corresponds well with the typical size of the mtgenome. For *T. kobelei*, three large contigs were generated which corresponded to three non-overlapping portions of mtgenome: 6582-bp fragment spanning 16S to ATP6 (141× coverage), 2389-bp fragment spanning CYTB to 16S (75× coverage), and 1583-bp fragment spanning ND5 to ND4 (44× coverage). The portions of the mtgenome that the 454 run failed to sequence were complemented with the sequences generated using primer-walking. The resulting sequences were deposited in GenBank with accession numbers: JX913764–JX913775.

2.3. Annotation

All newly generated mtgenomes were uploaded, annotated, and stored in the online program MOSAS (Sheffield et al., 2010), following the same protocols and procedures described by the authors. Previously published mtgenomes available in GenBank were downloaded as raw sequences and re-annotated in MOSAS to ensure consistency in the annotation strategy. tRNAs were positively identified using tRNAscan-SE (Lowe and Eddy, 1997) and Arwen (Laslett and Canbäck, 2008), both implemented in MOSAS. All nucleotides between tRNA-Leu and tRNA-Val were annotated as the large subunit ribosomal RNA (16S). The end of the coding region of each mtgenome was identified using a conserved sequence,

or "end tag", that is located ~800 bp downstream from the last nucleotide of tRNA-Val. All nucleotides between tRNA-Val and the "end tag" were annotated as the small ribosomal subunit (12S). Datasets were organized within MOSAS before being exported for alignment.

2.4. Alignment

Each protein-coding gene (PCG) was aligned individually based on the conservation of reading frames by first translating into amino acids using MacClade 4 (Maddison and Maddison, 2005), by aligning using G-INS-i strategy for global homology using MAFFT (Katoh et al., 2005), and by back-translating to nucleotides. In order to incorporate structural information of ribosomal RNAs during alignment, we first predicted secondary structures of 16S and 12S of *Acrida willemsei* in RNAfold WebServer (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) using minimum free energy (MFE) and partition function. Using these secondary structures as input constraint files, we performed secondary structure-based alignment in RNAsalsa (Stocsits et al., 2009) for 16S and 12S. For transfer RNAs, we predicted secondary structures of individual genes using tRNAscan-SE (Lowe and Eddy, 1997) and Arwen (Laslett and Canbäck, 2008). Individual tRNAs were aligned in MUSCLE (Edgar, 2004) using the default parameters, and the predicted secondary structures were used as a guide to identify stems and loops of each tRNA. After the alignment, both rRNA and tRNA genes were manually dissected into stems and loops guided by the secondary structure model in MEGA 5 (Tamura et al., 2011). The resulting alignments were concatenated in MacClade into a single dataset with 15,404 aligned bp, and then exported in phylip format using Mesquite ver. 2.74 (Maddison and Maddison, 2010).

2.5. Data partitioning schemes

We systematically surveyed a number of different data partitioning schemes using PartitionFinder (Lanfear et al., 2012). Because the software required a user to pre-define partitions and specify in the configuration file, we created an input configuration file that contained a total of 87 partitions, corresponding to individual codon position of each PCG (13 1st codon position, 13 2nd codon position, 13 3rd codon position), stems and loops of each RNA (2 stem of rRNA, 2 loop of rRNA, 22 stem of tRNA, 22 loop of tRNA), which was the most finely partitioned scheme possible in mitochondrial phylogenomics. We used the "greedy" algorithm (heuristic search) with branch lengths estimated as "unlinked" implemented in PartitionFinder to search for the best-fit scheme as well as the worst-fit scheme. Additionally, we also systematically surveyed some of the most frequently used *a priori* partitioning schemes in mitochondrial phylogenomics and compared them against the best-fit scheme selected by PartitionFinder. A total of 21 *a priori* schemes with varying degrees of complexity and the best-fit and the worst-fit schemes were statistically compared (Table 2). We used AIC, a modified version of AIC_c (McGuire et al., 2007), and the Bayesian Information Criterion (BIC) (Miller et al., 2009; Pons et al., 2010), all of which were natively implemented in PartitionFinder, to compare different partitioning schemes. For all decision-making techniques, the analysis with the lowest score was identified as the optimal partitioning scheme.

2.6. Phylogenetic analyses

We performed partitioned maximum likelihood analyses using the best-fit and the worst-fit partitioning schemes recommended by PartitionFinder as well as various conventional *a priori* partitioning schemes (Table 2). Because the goal of this study was to compare the effect of data partitioning schemes, we decided to use the most parameter-rich general time reversible model (GTR + I + Γ) applied to each partition across all analyses. We justified this approach because GTR + I + Γ was selected as the most appropriate model for the vast majority of the partitioned datasets when evaluated using jModelTest (Posada, 2008) and it has been shown that the consequences of over-parameterizing are not as severe as under-parameterizing (McGuire et al., 2007). All datasets were analyzed using RAxML 7.2.8 (Stamatakis et al., 2008) on XSEDE (Extreme Science and Engineering Discovery Environment, https://www.xsede.org) through CIPRES Science Gateway (Miller et al., 2011). Nodal support was evaluated using 1000 replications of rapid bootstrapping implemented in RAxML. In order to compare resulting topologies between the best-fit partitioning scheme and other schemes, we used the program TOPD/FMTS (Puigbò et al., 2007) for calculating nodal distance (nodal method) and identifying taxa with topological disagreement (disagree method). Although there is no comprehensive cladistic analysis of Acridoidea to date, Eades (2000) presented a synthesis evolutionary tree based on all previous studies. We compared all resulting topologies against the Eades' (2000) hypothesis to assess the phylogenetic accuracy.

3. Results

3.1. New mitochondrial genomes

We generated complete mtgenomes for *Lentula callani* (15,944 bp), *Prionotropis hystrix* (15,651 bp), *Tristira magellanica* (16,494 bp), *Pyrgacris descampsi* (15,618 bp), *Lithidiopsis carinatus* (15,652 bp), *Ommexecha virens* (15,536 bp), and *Tanaocerus kobelei* (15,515 bp). We also generated most of the mtgenome for *Hemicharilaus monomorphus* except the tail end of 12S through tRNA-Cys, *Trigonopteryx hopei* except for tRNA-Arg and tRNA-Ser, *Trachytettix horridus* except the control region and ~60 bp of 12S, and *Paramastax nigra* except 16S through tRNA-Met. All newly generated mtgenomes followed the ancestral insect gene order (Clary and Wolstenhome, 1985), with an exception to the inversion of tRNA-Asp and tRNA-Lys between protein-coding genes COII and ATP8. We also found conserved patterns of start and stop codons, functional overlap, and intergenic spacers across Acridomorpha (Table 3). In COI, we found a tetranucleotide start codon (AUGA) for all included caeliferans, which was proposed by Sheffield et al. (2010), except in *T. kobelei* and *Mekongiana xiangchengensis*, which exhibited a substitution of G for the first nucleotide, resulting in the sequence GUGA (Table 4).

3.2. Partitioning strategy assessment

The greedy analysis in PartitionFinder (Lanfear et al., 2012) created 109,737 schemes and selected the best-fit partitioning scheme based on AIC, AIC_c, and BIC. The selected best-fit partitioning scheme had 8 partitions and 599 parameters and the worst-fit scheme had 86 partitions and 6389 parameters (Table 5). Specific schemes for the best-fit and the worst-fit schemes are presented in Table 6. The best-fit scheme grouped different codon positions of different PCGs and stems and loops of different rRNAs and tRNAs. When compared with the *a priori* partitioning schemes, the best-fit scheme had a better information criterion score than any of the typical partitioning schemes (Table 5). The best *a priori* scheme was p5, which divided the dataset into codon positions (i.e. all 1st codon positions of PCGs into a single partition) and one partition of rRNAs and one partition of tRNAs. Partitioning by 13 PCGs, 2 rRNA, and a single partition of tRNAs (p16), which is the most commonly used scheme in mitochondrial phylogenomics, had a mediocre information criterion score. Finely partitioned schemes,

Table 2
A list of *a priori* partitioning schemes tested in this study. (PCG: Protein-coding gene).

p1 = no partition (the entire dataset treated as a single data block)

p2 = partition by strandedness (major and minor strands)

Partition by gene
p3 = PCGs + rRNAs + tRNAs (individual gene categories treated as single data blocks)
p6c = ATPs + COXs + CYTB + NDs + rRNAs + tRNAs (individual PCG categories treated as single data blocks)
p15 = 13 PCG + rRNAs + tRNAs
p16 = 13 PCG + 2 rRNA + tRNAs (most commonly used partition scheme in mitochondrial phylogenomics)
p37 = 13 PCG + 2 rRNA + 22 tRNA

Partition by codon position
p5 = 1st codon PCGs + 2nd codon PCGs + 3rd codon PCGs + rRNAs + tRNAs (individual codon positions of 13 PCGs treated as single data blocks)
p6 = 1st codon PCGs + 2nd codon PCGs + 3rd codon PCGs + 2 rRNA + tRNAs
p7 = 1st codon PCGs + 2nd codon PCGs + 3rd codon PCGs + rRNA stems + rRNA loops + tRNA stems + tRNA loops (stems and loops combined into single partition)
p27 = 1st codon PCGs + 2nd codon PCGs + 3rd codon PCGs + 2 rRNA + 22 tRNA
p51 = 1st codon PCGs + 2nd codon PCGs + 3rd codon PCGs + 2 rRNA stem + 2 rRNA loop + 22 tRNA stem + 22 tRNA loop

Partition by gene and codon position
p28 = 1st & 2nd codons of 13 PCG + 3rd codon of 13 PCG + rRNAs + tRNAs
p29 = 1st & 2nd codons of 13 PCG + 3rd codon of 13 PCG + 2 rRNA + tRNAs
p30 = 1st & 2nd codons of 13 PCG + 3rd codon of 13 PCG + rRNA stems + rRNA loops + tRNA stems + tRNA loops
p50 = 1st & 2nd codons of 13 PCG + 3rd codon of 13 PCG + 2 rRNA + 22 tRNAs

p41 = 1st codon of 13 PCG + 2nd codon of 13 PCG + 3rd codon of 13 PCG + rRNAs + tRNAs
p42 = 1st codon of 13 PCG + 2nd codon of 13 PCG + 3rd codon of 13 PCG + 2 rRNA + tRNAs
p43 = 1st codon of 13 PCG + 2nd codon of 13 PCG + 3rd codon of 13 PCG + rRNA stems + rRNA loops + tRNA stems + tRNA loops
p63 = 1st codon of 13 PCG + 2nd codon of 13 PCG + 3rd codon of 13 PCG + 2 rRNA + 22 tRNA
p87 = 1st codon of 13 PCG + 2nd codon of 13 PCG + 3rd codon of 13 PCG + 2 rRNA stem + 2 rRNA loop + 22 tRNA stem + 22 tRNA loop (most finely partitioned scheme)

in which individual tRNA genes were partitioned or further dissected into stems and loops, performed very poorly in general. In fact, the most finely dissected *a priori* scheme, which partitioned the dataset into individual codon positions and stems and loops of individual RNAs (p87) was the second worst-fit scheme selected by PartitionFinder.

To understand more general trends across different partitioning schemes, we plotted the BIC score against the log likelihood score based on 109,737 schemes from the PartitionFinder search (Fig. 1A) and based on the *a priori* schemes (Fig. 1B). The partitioning scheme that resulted in the best likelihood score did not have the lowest BIC score. There were many partitioning schemes that resulted in highly similar BIC scores (Fig. 1A). When we plotted the BIC score against the number of partitions, there was a strong positive correlation between the two in that the partitioning schemes with more partitions had higher BIC scores (Fig. 1C). However, no data partition at all (p1) or the schemes with a small number of partitions (p2, p3, p4, p6, p6c, and p7) had higher BIC scores than the best-fit scheme (Fig. 1D).

3.3. Topology comparisons

All of the resulting topologies from the different partitioning schemes recovered robust monophyletic relationships for Acridomorpha, Pyrgomorphoidea, and Acridoidea (Supplementary Fig. S1). Specifically, we recovered the following relationship for Acridomorpha at the superfamily level: (Eumastacoidea (Tanaoceroidea ((Pneumoroidea + Trigonopterygoidea) (Pyrgomorphoidea + Acridoidea)))). This relationship was more refined compared to Eades (2000) who did not resolve the phylogenetic position of Trigonopterygoidea. In terms of internal relationships within Acridoidea, all of the partitioning schemes recovered a sister relationship between Charilaidae and Pamphagidae, which was congruent with Eades (2000). A sister relationship between Lentulidae and Lithidiidae was also found across the partitioning schemes, but the phylogenetic placement of Lentulidae was unresolved in Eades (2000). The phylogenetic placements of nine taxa within Acridoidea (*Gastrimargus*, *Locusta*, *Oedaleus*, *Ommexecha*, *Oxya*, *Phlaeoba*, *Pyrgacris*, *Traulia*, *Xyleus*) were affected by different partitioning schemes (Table 7).

The best-fit partitioning scheme produced the following phylogenetic relationships within Acridoidea (Fig. 2): ((Charilaidae + Pamphagidae) (Pyrgacrididae ((Lentulidae + Lithidiidae) (Tristiridae (Acrididae (Ommexechidae + Romaleidae)))))). Because Eades' (2000) hypothesis was largely unresolved at this level, it was difficult to directly compare and contrast our finding with the past findings. Instead, we compared the topology resulting from the analysis of the best-fit partitioning scheme with other *a priori* schemes as well as the worst-fit scheme using nodal distance, a measure developed by Puigbò et al. (2007) which compares the number of nodes separating each taxon from the other taxa in the tree between two trees. This analysis found that the best-fit scheme resulted in an identical topology as the ones from p41 and p42, but the majority of the *a priori* schemes resulted in topologies that were different from the best-fit scheme (Table 7). Using the disagree method implemented in TOPD/FMTS (Puigbò et al., 2007), we were able to identify what taxa were the causes of topological disagreement. Except for the topologies resulting from p41 and p42 schemes, the majority of the *a priori* schemes had between one and six taxa that caused topological disagreements with the best-fit scheme (Table 7).

4. Discussion

4.1. Mitochondrial genome evolution within Orthoptera

The present study adds 11 new species to the growing list of orthopteran mtgenomes, ten of which are new families that have never been sequenced before. Based on this comprehensive taxon sampling, we can make some inferences about the mtgenome evolution within Orthoptera. One of the characteristic features of caeliferan mtgenomes is the tRNA rearrangement between COII and ATP8, which has tRNA-Asp before tRNA-Lys, compared to most other insects that have tRNA-Lys before tRNA-Asp (Fenn et al., 2008; Flook et al., 1995b; Sheffield et al., 2010). In the present study, we included all basal caeliferan lineages to test the new

Table 3
Nucleotide positions and anticodons (for tRNAs) for all genes newly generated for this study.

Gene	Strand	Anticodon	*Trachytettix horridus*[*]	*Paramastax nigra*[*]	*Tanaocerus koebelei*	*Trigonopteryx hopei*[*]	*Hemicharilaus monomorphus*[*]	*Prionotropishystrix*
tRNA-Ile	+	GAU	1–66 (0)	–	1–67 (0)	1–66 (0)	–	1–66 (0)
tRNA-Gln	–	UUG	66–134 (−1)	–	65–133 (−3)	71–139 (4)	–	67–135 (0)
tRNA-Met	+	CAU	134–201 (−1)	–	133–201 (−1)	139–207 (−1)	–	135–203 (−1)
ND2	+		202–1213 (0)	19–1015 (18)	223–1225 (21)	208–1222 (0)	–	204–1226 (0)
tRNA-Trp	+	UCA	1214–1281 (0)	1020–1083 (4)	1226–1291 (0)	1223–1286 (0)	–	1230–1295 (3)
tRNA-Cys	–	GCA	1274–1335 (−8)	1076–1137 (−8)	1284–1348 (−8)	1279–1340 (−8)	4–69 (3)	1288–1352 (−8)
tRNA-Tyr	–	GUA	1336–1399 (0)	1138–1201 (0)	1357–1423 (8)	1353–1421 (12)	79–144 (9)	1362–1427 (9)
COX1	+		1400–2930 (0)	1202–2738 (0)	1425–2955 (1)	1423–2953 (1)	146–1676 (1)	1429–2962 (1)
tRNA-Leu	+	UAA	2931–2997 (0)	2739–2802 (0)	2956–3021 (0)	2954–3018 (0)	1677–1742 (0)	2963–3028 (0)
COX2	+		3010–3676 (12)	2805–3489 (2)	3024–3705 (2)	3021–3702 (2)	1750–2432 (7)	3036–3719 (7)
tRNA-Asp	+	GUC	3677–3739 (0)	3490–3558 (0)	3706–3772 (0)	3703–3767 (0)	2433–2497 (0)	3723–3793 (3)
tRNA-Lys	+	CUU	3741–3807 (1)	3561–3632 (2)	3774–3844 (1)	3769–3838 (1)	2503–2573 (5)	3796–3866 (2)
ATP8	+		3812–3970 (4)	3645–3803 (12)	3858–4016 (13)	3846–4013 (7)	2587–2748 (13)	3884–4042 (17)
ATP6	+		3967–4635 (−4)	3800–4477 (−4)	4013–4684 (−4)	4010–4681 (−4)	2745–3338 (−4)	4039–4713 (−4)
COX3	+		4635–5424 (−1)	4464–5253 (−14)	4684–5474 (−1)	4685–5471 (3)	3414–4214 (75)	4718–5509 (4)
tRNA-Gly	+	UCC	5425–5486 (0)	5254–5316 (0)	5475–5541 (0)	5472–5534 (0)	4215–4280 (0)	5513–5581 (3)
ND3	+		5487–5838 (0)	5317–5668 (0)	5542–5893 (0)	5535–5887 (0)	4281–4634 (0)	5582–5933 (0)
tRNA-Ala	+	UGC	5839–5903 (0)	5669–5732 (0)	5894–5959 (0)	5888–5952 (0)	4636–4700 (1)	5934–5999 (0)
tRNA-Arg	+	UCG	5902–5964 (−2)	5732–5796 (−1)	5960–6026 (0)	–	4708–4772 (7)	6002–6063 (2)
tRNA-Asn	+	GUU	5964–6027 (−1)	5800–5864 (3)	6031–6094 (4)	6021–6085 (68)	4775–4840 (2)	6067–6131 (3)
tRNA-Ser	+	GCU	6028–6094 (0)	5862–5928 (−3)	6095–6161 (0)	6086–6152 (0)	4841–4907 (0)	6132–6198 (0)
tRNA-Glu	+	UUC	6095–6161 (0)	5929–5992 (0)	6162–6227 (0)	6153–6218 (0)	4912–4978 (4)	6199–6264 (0)
tRNA-Phe	–	GAA	6160–6226 (−2)	5991–6055 (−2)	6226–6292 (−2)	6219–6287 (0)	4979–5044 (0)	6267–6332 (2)
ND5	–		6227–7952 (0)	6056–7775 (0)	6293–8028 (0)	6288–8025 (0)	5045–6770 (0)	6333–8064 (0)
tRNA-His	–	GUG	7956–8020 (3)	7779–7842 (3)	8030–8096 (1)	8026–8093 (0)	6777–6843 (6)	8065–8132 (0)
ND4	–		8032–9345 (11)	7844–9166 (1)	8097–9427 (0)	8094–9414 (0)	6844–8177 (0)	8133–9466 (0)
ND4L	–		9351–9641 (5)	9172–9465 (5)	9421–9714 (−7)	9420–9713 (5)	8171–8464 (−7)	9460–9753 (−7)
tRNA-Thr	+	UGU	9644–9707 (2)	9468–9530 (2)	9717–9783 (2)	9716–9779 (2)	8467–8535 (2)	9756–9821 (2)
tRNA-Pro	–	UGG	9708–9773 (0)	9532–9595 (1)	9784–9848 (0)	9781–9844 (1)	8536–8603 (0)	9822–9888 (0)
ND6	+		9775–10,266 (1)	9597–10,100 (1)	9850–10,367 (1)	9847–10,371 (2)	8606–9127 (2)	9894–10,412 (5)
CYTB	+		10,266–11,400 (−1)	10,100–11,237 (−1)	10,368–11,505 (0)	10,371–11,507 (−1)	9131–10,268 (3)	10,416–11,556 (3)
tRNA-Ser	+	UGA	11,401–11,467 (0)	11,238–11,310 (0)	11,506–11,574 (0)	–	10,269–10,338 (0)	11,557–11,626 (0)
ND1	–		11,470–12,417 (2)	11,337–12,284 (26)	11,619–12,542 (44)	11,728–12,517 (0)	10,364–11,311 (25)	11,653–12,600 (26)
tRNA-Leu	–	UAG	12,418–12,484 (0)	12,285–12,350 (0)	12,543–12,608 (0)	12,518–12,582 (0)	11,312–11,377 (0)	12,601–12,666 (0)
16S	–		12,485–13,781 (0)	–	12,609–13,905 (0)	12,583–13,886 (0)	11,378–12,689 (0)	12,667–13,987 (0)
tRNA-Val	–	UAC	13,782–13,849 (0)	–	13,906–13,975 (0)	13,887–13,956 (0)	12,690–12,760 (0)	13,988–14,058 (0)
12S	–		13,850–14,578 (0)	–	13,976–14,739 (0)	13,957–14,735 (0)	12,760–13,291 (−1)	14,059–14,858 (0)
Control	N/A		–	–	14,740–15,515 (0)	14,736–16,732 (0)	–	14,859–15,651 (0)

Gene	Strand	Anticodon	*Pyrgacris descampsi*	*Lentula callani*	*Lithidiopsis carinatus*	*Tristira magellanica*	*Ommexecha virens*
tRNA-Ile	+	GAU	1–64 (0)	1–65 (0)	1–67 (0)	1–66 (0)	1–64 (0)
tRNA-Gln	–	UUG	81–149 (16)	66–134 (0)	69–137 (1)	67–135 (0)	66–134 (1)
tRNA-Met	+	CAU	149–216 (−1)	135–202 (0)	137–206 (−1)	135–203 (−1)	134–203 (−1)
ND2	+		217–1242 (0)	203–1222 (0)	207–1227 (0)	204–1223 (0)	204–1221 (0)
tRNA-Trp	+	UCA	1249–1317 (6)	1228–1295 (5)	1228–1295 (0)	1890–1955 (666)	1222–1288 (0)
tRNA-Cys	–	GCA	1310–1374 (−8)	1288–1352 (−8)	1288–1352 (−8)	1948–2010 (−8)	1281–1343 (−8)
tRNA-Tyr	–	GUA	1401–1468 (26)	1369–1434 (16)	1358–1424 (5)	2015–2081 (4)	1348–1413 (4)
COX1	+		1470–3000 (1)	1436–2966 (1)	1426–2956 (1)	2083–3613 (1)	1415–2945 (1)
tRNA-Leu	+	UAA	3001–3066 (0)	2967–3033 (0)	2957–3021 (0)	3614–3680 (0)	2946–3010 (0)
COX2	+		3067–3748 (0)	3042–3723 (8)	3027–3709 (5)	3688–4371 (7)	3013–3694 (2)
tRNA-Asp	+	GUC	3749–3811 (0)	3724–3789 (0)	3710–3775 (0)	4372–4436 (0)	3695–3760 (0)
tRNA-Lys	+	CUU	3818–3887 (6)	3794–3864 (4)	3781–3851 (5)	4441–4511 (4)	3763–3833 (2)
ATP8	+		3896–4066 (8)	3868–4041 (3)	3873–4034 (21)	4528–4689 (16)	3849–4010 (15)
ATP6	+		4063–4737 (−4)	4038–4712 (−4)	4031–4705 (−4)	4686–5355 (−4)	4007–4681 (−4)
COX3	+		4742–5532 (4)	4716–5506 (3)	4709–5499 (3)	5356–6156 (0)	4686–5475 (4)
tRNA-Gly	+	UCC	5533–5599 (0)	5507–5573 (0)	5500–5575 (0)	6158–6224 (1)	5476–5543 (0)
ND3	+		5600–5951 (0)	5574–5925 (0)	5576–5928 (0)	6225–6578 (0)	5544–5896 (0)
tRNA-Ala	+	UGC	5952–6016 (0)	5926–5988 (0)	5929–5995 (0)	6579–6642 (0)	5897–5961 (0)
tRNA-Arg	+	UCG	6019–6082 (2)	5989–6056 (0)	5996–6065 (0)	6646–6711 (3)	5964–6028 (2)
tRNA-Asn	+	GUU	6083–6147 (0)	6059–6124 (2)	6067–6131 (1)	6715–6780 (3)	6034–6098 (5)
tRNA-Ser	+	GCU	6148–6213 (0)	6125–6191 (0)	6132–6198 (0)	6781–6847 (0)	6099–6165 (0)
tRNA-Glu	+	UUC	6217–6282 (3)	6193–6256 (1)	6200–6266 (1)	6849–6914 (1)	6166–6230 (0)
tRNA-Phe	–	GAA	6281–6345 (−2)	6257–6319 (0)	6267–6333 (0)	6913–6976 (−2)	6232–6297 (1)
ND5	–		6346–8065 (0)	6320–8047 (0)	6334–8067 (0)	6977–8715 (0)	6298–8023 (0)
tRNA-His	–	GUG	8081–8145 (15)	8049–8115 (1)	8069–8141 (1)	8716–8781 (0)	8030–8094 (6)
ND4	–		8147–9469 (1)	8116–9449 (0)	8143–9477 (1)	8782–10,119 (0)	8095–9432 (0)
ND4L	–		9475–9768 (5)	9443–9736 (−7)	9471–9764 (−7)	10,113–10,406 (−7)	9426–9719 (−7)
tRNA-Thr	+	UGU	9771–9837 (2)	9739–9805 (2)	9767–9834 (2)	10,409–10,474 (2)	9722–9791 (2)
tRNA-Pro	–	UGG	9838–9900 (0)	9806–9868 (0)	9835–9898 (0)	10,475–10,538 (0)	9792–9856 (0)
ND6	+		9903–10,430 (2)	9871–10,392 (2)	9901–10,422 (2)	10,541–11,059 (2)	9859–10,380 (2)
CYTB	+		10,430–11,570 (−1)	10,396–11,544 (3)	10,429–11,568 (6)	11,067–12,204 (7)	10,384–11,520 (3)
tRNA-Ser	+	UGA	11,571–11,640 (0)	11,550–11,618 (5)	11,571–11,640 (2)	12,205–12,274 (0)	11,525–11,594 (4)
ND1	–		11,662–12,609 (21)	11,640–12,587 (21)	11,660–12,607 (19)	12,301–13,248 (26)	11,616–12,563 (21)

(*continued on next page*)

Table 3 (*continued*)

Gene	Strand	Anticodon	*Pyrgacris descampsi*	*Lentula callani*	*Lithidiopsis carinatus*	*Tristira magellanica*	*Ommexecha virens*
tRNA-Leu	–	UAG	12,610–12,675 (0)	12,588–12,651 (0)	12,608–12,673 (0)	13,249–13,314 (0)	12,564–12,628 (0)
16S	–		12,676–13,979 (0)	12,652–13,968 (0)	12,674–13,994 (0)	13,315–14,633 (0)	12,629–13,937 (0)
tRNA-Val	–	UAC	13,980–14,050 (0)	13,969–14,039 (0)	13,995–14,065 (0)	14,634–14,704 (0)	13,938–14,008 (0)
12S	–		14,051–14,822 (0)	14,039–14,834 (−1)	14,066–14,862 (0)	14,705–15,506 (0)	14,009–14,787 (0)
Control	N/A		14,823–15,618 (0)	14,835–15,944 (0)	14,863–15,652 (0)	15,507–16,494 (0)	14,788–15,536 (0)

* Indicates partially sequenced mtgenomes.

Table 4
Start codon of protein-coding genes for all taxa included in the analysis. The start codon for COI is hypothesized to be a conserved tetranucleotide (AUGA), which is found in all caeliferans. Two exceptions are found in *T. kobelei* and *M. xiangchengensis* (indicated by asterisk), both of which have GUGA as the start codon.

Species	*ND2*	*COI*	*COII*	*ATP8*	*ATP6*	*COIII*	*ND3*	*ND5*	*ND4*	*ND4L*	*ND6*	*CYTB*	*ND1*
Ellipes minuta	I	t	M	I	M	M	M	I	M	M	I	M	L
Trachytettix horridus	M	t	M	M	M	M	I	M	I	I	L	M	I
Paramastax nigra	–	t	M	I	M	M	I	M	I	M	I	M	M
Tanaocerus kobelei	I	t*	M	I	M	M	I	M	M	M	I	M	L
Physemacris variolosa	M	t	M	I	M	M	I	M	I	M	I	M	L
Trigonopteryx hopei	I	t	M	I	M	M	I	M	L	M	I	M	–
Atractomorpha sinensis	I	t	M	I	M	M	I	I	M	M	M	M	L
Mekongiella xizangensis	I	t	M	I	M	M	I	I	M	M	M	M	L
Mekongiana xiangchengensis	I	t*	M	I	M	I	M	I	L	M	L	M	L
Hemicharilaus monomorphus	M	t	M	I	M	I	I	L	M	M	M	M	L
Thrinchus schrenkii	M	t	M	I	M	M	I	M	M	M	M	M	L
Prionotropis hystrix	M	t	M	I	M	M	I	M	M	M	M	M	L
Pyrgacris descampsi	M	t	M	I	M	M	M	I	L	M	I	M	L
Lithidiopsis carinatus	M	t	M	M	M	M	I	M	M	M	L	M	L
Lentula callani	M	t	M	I	M	M	I	M	M	M	L	I	L
Tristira magellanica	M	t	M	I	M	I	I	L	M	M	I	M	L
Ommexecha virens	M	t	M	I	M	M	I	M	M	M	M	M	L
Xyleus modestus	M	t	I	I	M	M	I	L	L	M	I	M	L
Calliptamus italicus	M	t	M	I	M	M	I	L	M	M	M	M	L
Schistocerca gregaria gregaria	M	t	M	I	M	M	I	I	M	M	M	M	L
Traulia szetschuanensis	M	t	M	I	M	M	I	M	M	M	M	M	L
Oxya chinensis	M	t	M	I	M	M	I	I	M	M	M	M	L
Prumna arctica	M	t	M	I	M	M	I	M	M	M	M	M	L
Ognevia longipennis	M	t	M	I	M	M	I	L	M	M	M	M	L
Phlaeoba albonema	M	t	M	M	M	M	I	M	M	M	M	M	L
Acrida willemsei	M	t	M	M	M	M	I	M	M	M	M	M	L
Oedaleus decorus asiaticus	M	t	M	I	M	M	I	I	M	M	M	M	L
Gastrimargus marmoratus	M	t	M	I	M	M	I	I	M	M	M	M	L
Locusta migratoria	M	t	M	I	M	M	I	I	M	M	M	M	L
Arcyptera coreana	M	t	M	I	M	M	I	M	I	M	M	M	L
Euchorthippus fusigeniculatus	M	t	M	I	M	M	I	I	L	M	M	I	L
Chorthippus chinensis	M	t	M	I	M	M	I	I	L	M	M	I	L
Gomphocerippus rufus	M	t	M	I	M	M	I	M	M	M	M	I	L
Gomphocerus licenti	M	t	M	I	M	M	I	M	L	M	M	M	L

hypothesis by Sheffield et al. (2010), who suggested that this gene rearrangement is a synapomorphy for a subgroup within the orthopteran suborder Caelifera. We found that Tridactyloidea was the only caeliferan that had the ancestral gene arrangement (Fig. 2). In other words, the tRNA rearrangement probably evolved in the common ancestor of Tetrigoidea and Acridomorpha after it diverged from the ancestral Tridactyloidea and we consider this rearrangement as a definitive molecular synapomorphy for Tetrigoidea + Acridomorpha.

Besides this tRNA rearrangement, gene order, tRNA anticodons, start and stop codons, and gene lengths are highly conserved throughout the mtgenomes of Acridomorpha (Table 3). Based on fossil data, the stem-group Acridomorpha (Locustopseidae) is known from the Early Jurassic (198 MYA) (Zeuner, 1942) and the stem-group Tetrigidae is known from the Early Cretaceous (125–130 MYA) (*Archaeotetrix* and *Prototetrix*) (Sharov, 1968), which suggested that the mtgenome structure has remained unmodified for at least 150–170 million years. This pattern is in contrast with the mtgenome structures in more recently evolved insect orders such as paraneopterans (Covacin et al., 2006; Hua et al., 2008; Shao and Barker, 2003; Shao et al., 2001) and Hymenoptera (Dowton et al., 2009b) that are known to have extensive gene rearrangements.

4.2. Partition strategies for analyzing mtgenome data

The main objective of this study is to address the effect of data partitioning in a maximum likelihood framework, especially when it is applied to complex, multi-locus datasets. With a total of 37 genes, there are numerous possible ways of partitioning mtgenome data, but most mitochondrial phylogenomic studies to date have so far relied on only a subset of partitioning schemes (e.g. Cameron et al., 2004; Fenn et al., 2008), based on decisions made without much justification. The present study represents the first attempt to rigorously explore the vast landscape of mtgenome data partitioning.

An in-depth look at the best-fit partitioning scheme, selected by PartitionFinder, reveals some interesting patterns. The scheme partitioned the mtgenome data into 8 partitions, of which five partitions grouped genes based on some structural characteristics (Table 6). For example, one partition grouped the 2nd codon position of 8 PCGs, and another partition grouped loop regions of 8 RNAs. This finding indicates that there may be some inherent similarities in terms of nucleotide evolution among the same codon positions across different PCGs and among stems and loops across different RNAs. It is worth noting that PartitionFinder was able to find structural similarities among different genes without our

Table 5
Comparison of best-fit and worst-fit scheme as selected by PartitionFinder with commonly used *a priori* partitioning schemes.

Partition scheme	*lnL*	#Parameters	#Sites	#Data blocks	AIC	AICc	BIC
best	−211435.02	599	15,404	8	**424068.04**	**424116.59**	**428645.83**
p5	−216500.04	375	15,404	5	433750.08	433768.85	436615.98
p7	−215819.35	523	15,404	7	432684.70	432721.53	436681.67
p6	−216438.69	450	15,404	6	433777.37	433804.52	437216.44
p2	−218696.62	150	15,404	2	437693.24	437696.21	438839.60
p28	−210913.70	2085	15,404	28	425997.39	426650.54	441931.76
p30	−210233.00	2233	15,404	30	424932.01	425689.57	441997.45
p6c	−219083.98	450	15,404	6	439067.96	439095.10	442507.03
p29	−210852.34	2160	15,404	29	426024.68	426729.62	442532.22
p15	−216535.66	1124	15,404	15	435319.32	435496.43	443909.36
p16	−216474.30	1199	15,404	16	435346.61	435549.20	444509.82
p3	−221437.52	225	15,404	3	443325.04	443331.74	445044.57
p1	−222427.65	75	15,404	1	445005.31	445006.05	445578.49
p41	−209214.19	3054	15,404	41	424536.38	426047.43	447876.21
p43	−208533.50	3202	15,404	43	423470.99	425152.17	447941.90
p42	−209152.83	3129	15,404	42	424563.66	426159.52	448476.68
p27	−215157.25	2004	15,404	27	434322.50	434922.25	449637.83
p50	−209570.90	3714	15,404	50	426569.80	428930.57	454953.61
p37	−215192.87	2753	15,404	37	435891.73	437090.43	456931.21
p63	−207871.39	4683	15,404	63	425108.79	429201.17	460898.07
p51	−213594.57	3784	15,404	51	434757.15	437222.50	463675.92
p87	−206308.72	6463	15,404	87	425543.44	434889.49	474936.16
worst	−207469.38	6389	15,404	86	427716.76	436775.05	476543.94

Table 6
The best and the worst partitioning scheme selected by PartitionFinder and the best *a priori* partitioning scheme. p*n* indicates individual codon position of PCGs.

Best: 8 partitions

Partition 1: (ATP6_p1, ATP8_p1, ATP8_p2, Ala_loop, Arg_loop, Asn_loop, Asp_loop, Cys_loop, Glu_loop, Gly_loop, Ile_loop, Leu1_loop, Lys_loop, Met_loop, ND2_p1, ND3_p1, ND6_p1, Ser1_loop, Ser2_loop, Thr_loop, Trp_loop, Tyr_loop)
Partition 2: (ATP6_p2, COX1_p2, COX2_p2, COX3_p2, CYTB_p2, ND2_p2, ND3_p2, ND6_p2)
Partition 3: (ATP6_p3, ATP8_p3, COX1_p3, COX2_p3, COX3_p3, CYTB_p3, ND2_p3, ND3_p3, ND6_p3)
Partition 4: (COX1_p1, COX2_p1, COX3_p1, CYTB_p1)
Partition 5: (12S_stem, 16S_stem, Ala_stem, Arg_stem, Asn_stem, Asp_stem, Cys_stem, Gln_stem, Glu_stem, Gly_stem, His_stem, Ile_stem, Leu1_stem, Leu2_stem, Lys_stem, Met_stem, ND1_p1, ND4_p1, ND5_p1, Phe_stem, Pro_stem, Ser1_stem, Ser2_stem, Thr_stem, Trp_stem, Tyr_stem, Val_stem)
Partition 6: (ND1_p2, ND4L_p1, ND4L_p2, ND4_p2, ND5_p2)
Partition 7: (ND1_p3, ND4L_p3, ND4_p3, ND5_p3)
Partition 8: (12S_loop, 16S_loop, Gln_loop, His_loop, Leu2_loop, Phe_loop, Pro_loop, Val_loop);

p5 (best *a priori*): 5 partitions

Partition 1: (ATP6_p1, ATP8_p1, COX1_p1, COX2_p1, COX3_p1, CYTB_p1, ND1_p1, ND2_p1, ND3_p1, ND4_p1, ND4L_p1, ND5_p1, ND6_p1)
Partition 2: (ATP6_p2, ATP8_p2, COX1_p2, COX2_p2, COX3_p2, CYTB_p2, ND1_p2, ND2_p2, ND3_p2, ND4_p2, ND4L_p2, ND5_p2, ND6_p2)
Partition 3: (ATP6_p3, ATP8_p3, COX1_p3, COX2_p3, COX3_p3, CYTB_p3, ND1_p3, ND2_p3, ND3_p3, ND4_p3, ND4L_p3, ND5_p3, ND6_p3)
Partition 4: (16S, 12S)
Partition 5: (Ala, Arg, Asn, Asp, Cys, Gln, Gly, Glu, His, Ile, Leu1, Leu2, Lys, Met, Phe, Pro, Ser1, Ser2, Thr, Trp, Tyr, Val)

Worst: 86 partitions

(ATP6_p1) (ATP6_p2) (ATP6_p3) (ATP8_p1) (ATP8_p2) (ATP8_p3) (COX1_p1) (COX1_p2) (COX1_p3, ND5_p2) (COX2_p1) (COX2_p2) (COX2_p3) (COX3_p1) (COX3_p2) (COX3_p3) (CYTB_p1) (CYTB_p2) (CYTB_p3) (ND1_p1) (ND1_p2) (ND1_p3) (ND2_p1) (ND2_p2) (ND2_p3) (ND3_p1) (ND3_p2) (ND3_p3) (ND4_p1) (ND4_p2) (ND4_p3) (ND4L_p1) (ND4L_p2) (ND4L_p3) (ND5_p1) (ND5_p3) (ND6_p1) (ND6_p2) (ND6_p3) (16S_stem) (16S_loop) (12S_stem) (12S_loop) (Ala_stem) (Ala_loop) (Arg_stem) (Arg_loop) (Asn_stem) (Asn_loop) (Asp_stem) (Asp_loop) (Cys_stem) (Cys_loop) (Gln_stem) (Gln_loop) (Glu_stem) (Glu_loop) (Gly_stem) (Gly_loop) (His_stem) (His_loop) (Ile_stem) (Ile_loop) (Leu1_stem) (Leu1_loop) (Leu2_stem) (Leu2_loop) (Lys_stem) (Lys_loop) (Met_stem) (Met_loop) (Phe_stem) (Phe_loop) (Pro_stem) (Pro_loop) (Ser1_stem) (Ser1_loop) (Ser2_stem) (Ser2_loop) (Thr_stem) (Thr_loop) (Trp_stem) (Trp_loop) (Tyr_stem) (Tyr_loop) (Val_stem) (Val_loop);

specifying their similarities *a priori*. However, we also found that the remaining three partitions grouped different regions of the genes regardless of their structural similarities. For example, one partition grouped of the 1st codon position of 5 PCGs, the 2nd codon position of 1 PCG, and the loop region of 16 tRNAs. Perhaps this is just noise, but we do not have a good explanation for this kind of pattern at this point. One of the benefits of statistically analyzing partitioning schemes is the possibility of obtaining a robust and justifiable solution through an extensive exploration of potential partitioning schemes rather than trying to speculate the most reasonable partitioning scheme *a priori*.

Indeed, we find that the *a priori* partitioning schemes typically employed in mitochondrial phylogenomics are suboptimal to the best-fit scheme suggested by PartitionFinder in terms of the information criterion scores (Table 5). This is the pattern also observed by Lanfear et al. (2012) who compared partitioning schemes across many types of mitochondrial and nuclear genes. One of the most commonly used partitioning schemes in the mitochondrial phylogenomics includes 16 partitions in which 13 PCGs and 2 rRNAs are individually partitioned and 22 tRNAs are combined into a single partition (Fenn et al., 2008). This scheme (p16) had about twice as many parameters as the best-fit scheme, and 9th of the 20

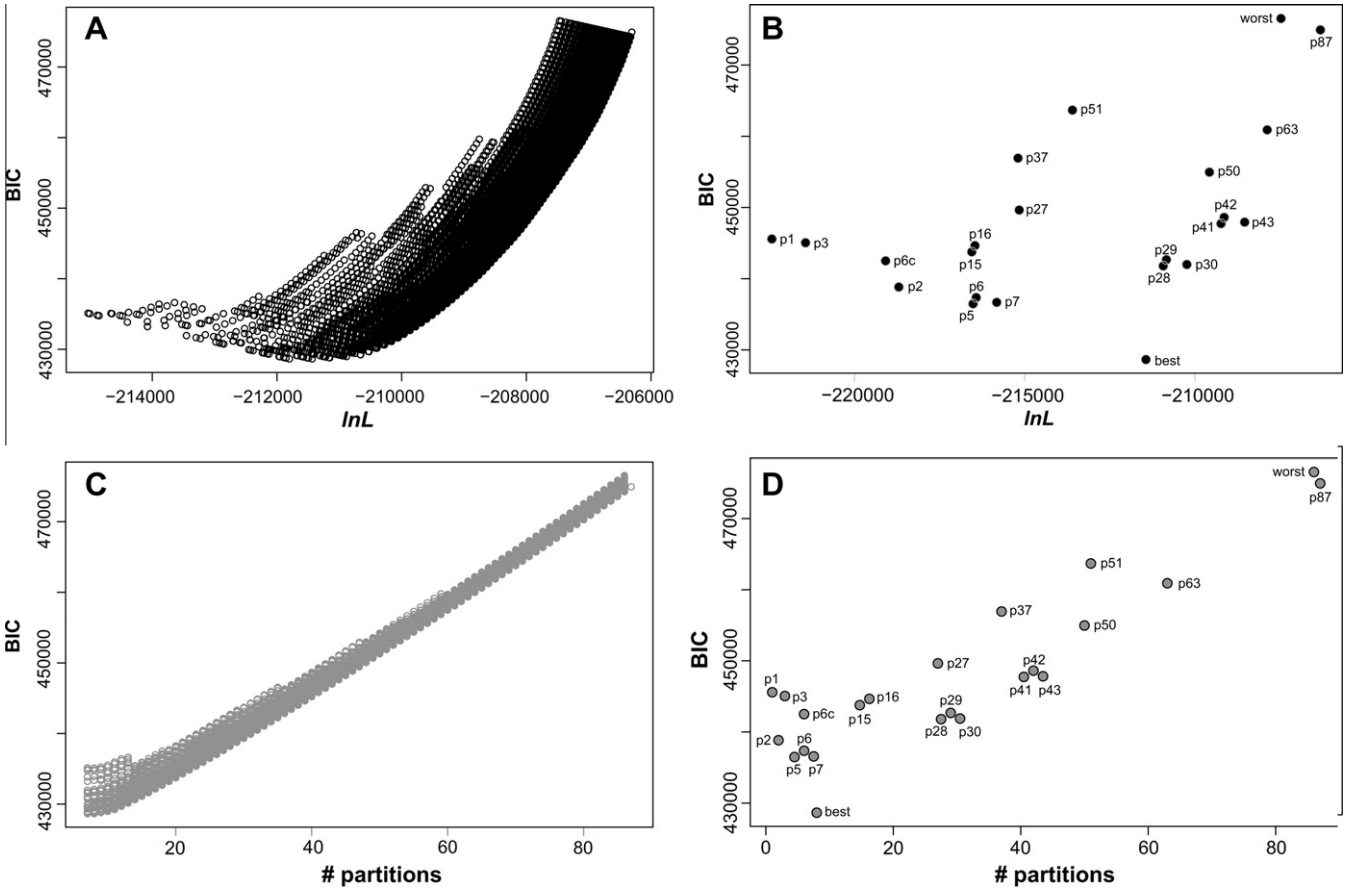

Fig. 1. The relationship between Bayesian Information Criterion (BIC) score and log likelihood (A and B), and between BIC and the number of partitions (C and D). Based on 109,737 schemes explored using PartitionFinder, we find that the best BIC score (the best-fit partitioning scheme) does not necessarily have the best likelihood ratio (A), but there is almost a direct correlation between the BIC score and the number of partitions (C). Among the *a priori* schemes, we find that the schemes that partitioned PCGs in a similar manner tend to result in the similar BIC scores (B and D). For more detailed information about specific partitioning schemes, see Table 2.

Table 7
Comparisons of topologies between best-fit partitioning schemes and other schemes calculated using TOPD/FMTS (Puigbò et al., 2007). Nodal distance is calculated from comparing pairwise distance matrices from the two input tree (the best-fit scheme vs others). The nodal distance score is 0 for identical trees, and increases as the two trees become more dissimilar. Only the genus names are shown for the taxa in disagreement.

Partition scheme	Nodal distance	Taxa in disagreement
Best	–	–
p5	0.9784	*Ommexecha*, *Phlaeoba*, *Pyrgacris*, *Traulia*, *Xyleus*
p7	0.9784	*Ommexecha*, *Phlaeoba*, *Pyrgacris*, *Traulia*, *Xyleus*
p6	0.9784	*Ommexecha*, *Phlaeoba*, *Pyrgacris*, *Traulia Xyleus*
p2	0.8580	*Gastrimargus*, *Locusta*, *Oedaleus*, *Ommexecha*, *Pyrgacris*, *Xyleus*
p28	0.9374	*Gastrimargus*, *Locusta*, *Oedaleus*, *Traulia*
p30	0.9143	*Oxya*, *Gastrimargus*, *Locusta*, *Oedaleus*, *Traulia*
p6c	0.7932	*Gastrimargus*, *Locusta*, *Oedaleus*, *Ommexecha*, *Xyleus*
p29	0.9374	*Gastrimargus*, *Locusta*, *Oedaleus*, *Traulia*
p15	0.3378	*Gastrimargus*, *Locusta*, *Oedaleus*
p16	0.3378	*Gastrimargus*, *Locusta*, *Oedaleus*
p3	0.8580	*Gastrimargus*, *Locusta*, *Oedaleus*, *Ommexecha*, *Pyrgacris*, *Xyleus*
p1	0.8580	*Gastrimargus*, *Locusta*, *Oedaleus*, *Ommexecha*, *Pyrgacris*, *Xyleus*
p41	0.0000	none
p43	0.6716	*Traulia*
p42	0.0000	none
p27	0.9784	*Ommexecha*, *Phlaeoba*, *Pyrgacris*, *Traulia*, *Xyleus*
p50	0.9374	*Gastrimargus*, *Locusta*, *Oedaleus*, *Traulia*
p37	0.7932	*Gastrimargus*, *Locusta*, *Oedaleus*, *Ommexecha*, *Xyleus*
p63	0.7739	*Gastrimargus*, *Locusta*, *Oedaleus*, *Ommexecha*, *Traulia*, *Xyleus*
p51	0.9784	*Ommexecha*, *Phlaeoba*, *Pyrgacris*, *Traulia*, *Xyleus*
p87	0.7517	*Gastrimargus*, *Locusta*, *Oedaleus*, *Traulia*
Worst	0.7739	*Gastrimargus*, *Locusta*, *Oedaleus*, *Ommexecha*, *Traulia*, *Xyleus*

a priori schemes in terms of the BIC score (Table 5). In other words, this commonly used scheme is not the best way to analyze the mtgenome data in our study.

Among the *a priori* schemes, three best performing schemes (p5, p7, and p6) all relied on a partitioning scheme that grouped different codon positions of PCGs into single partitions (i.e. all 1st codon position of 13 PCGs grouping into a single partition). The main difference among these three schemes was how rRNA and tRNA genes were further partitioned. For example, rRNA could be combined into a single partition, divided into two individual genes (16S and 12S), or further divided into stems and loops. The effect of partitioning 16S and 12S into various schemes appeared to have a minimum effect to the performance in terms of BIC scores and resulting topologies. However, when tRNA genes were partitioned into individual genes (22 additional partitions) or divided even further into stems and loops (44 additional partitions), the effect became dramatic. For example, p51 was identical to p6 in terms of PCGs and rRNAs, but differed in how tRNA genes were partitioned, and its performance dropped down to one of the worst *a priori* schemes (Table 5). It is possible that the increasingly poor performance of the higher number of partitions is related to over-parameterization. In general, we find that partitioning schemes that dissect the data finely perform more poorly than those with a small number of partitions (Fig. 1). It has been shown that over-partitioning not only increases computation time, but reduces the amount of information available for estimating parameters, which can lead to increased variance, added weight to divergent lineages, and parameter nonidentifiability (Chang, 1996).

Our comparison among the *a priori* schemes is also revealing in terms of the effect of partitioning by gene or by gene and codon position. The partitioning schemes p5 (best *a priori* scheme), p15, p28, and p41 only differed in how PCGs were divided up. 13 PCGs can be partitioned into three codon-based partitions (p5), 13 gene-based partitions (p15), 26 partitions in which the 1st and 2nd codon positions of each gene are combined and the 3rd codon is treated separately (p28), or 39 partitions in which individual codon positions of each gene are treated separately (p41). We find that as the number of partitions increased, the BIC scores and likelihood scores worsened. Similarly to the effect of over-partitioning of tRNA genes, over-partitioning of PCGs appears to have a detrimental impact in terms of phylogenetic reconstruction. Although partitioning by gene is a more commonly used strategy, we show that partitioning 13 PCGs into three codon-based partitions is a more appropriate method in our dataset. Our findings are also congruent with Fenn et al. (2008) and Pons et al. (2010) who discouraged partitioning schemes that would partition mitochondrial PCGs by both gene and codon position because doing so would result in over-parameterization and stochastic error.

4.3. Effect of data partitioning in terms of phylogenetic reconstruction

We find that our mtgenome data contained strong phylogenetic signal for robustly resolving the higher-level relationships within Acridomorpha because identical superfamily-level relationships were recovered regardless of the partitioning schemes. However, we also show that the relationships within Acridoidea were sensitive to partitioning schemes. The topology resulting from

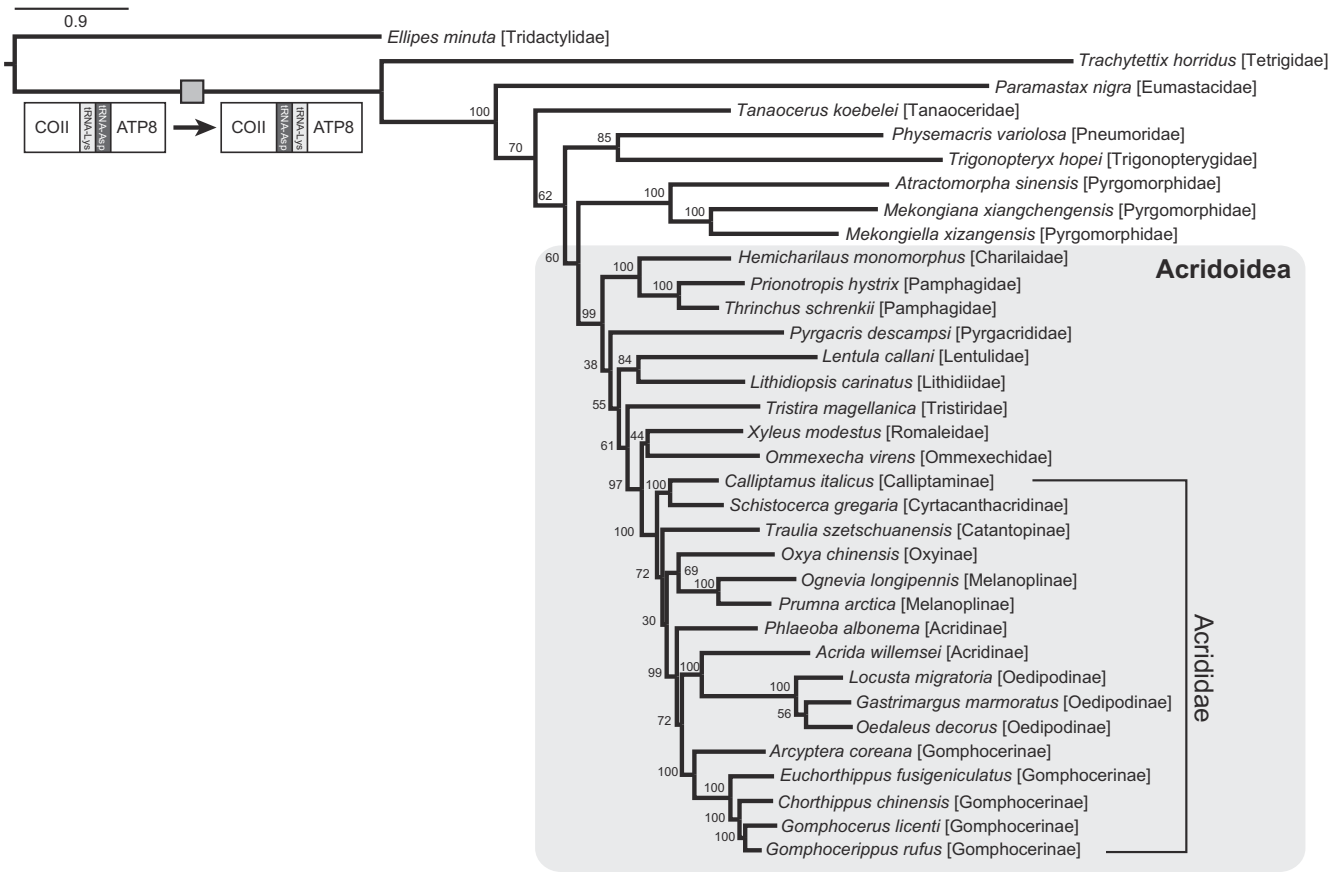

Fig. 2. A maximum likelihood tree of Acridoidea based on the analysis of the best-fit partitioning scheme selected by PartitionFinder. Numbers above nodes are bootstrap values. tRNA rearrangement between COII and ATP8 is found in all caeliferans except Tridactyloidea, suggesting that this rearrangement is a molecular synapomorphy for Tetrigoidea + Acridomorpha.

the best-fit scheme was different from the best *a priori* scheme (p5) in the placement of five taxa. In fact, 18 out of the 20 *a priori* schemes resulted in topologies that were different from the best-fit scheme. The topological conflicts resulted from the alternative placements of nine taxa: *Gastrimargus*, *Locusta*, *Oedaleus*, *Ommexecha*, *Oxya*, *Phlaeoba*, *Pyrgacris*, *Traulia*, and *Xyleus*. Based on the topology resulting from the best-fit scheme, the nodal support values for the placement of these taxa were not robust, often less than the bootstrap value of 50. This pattern may indicate that our mtgenome data did not contain enough phylogenetic information to correctly resolve lower-level relationships or it may be that these taxa are so closely related (especially in the case of *Gastrimargus*, *Locusta*, and *Oedaleus*) that data partitioning itself reduced the number of informative sites to group them correctly. It is notable that these nine taxa do not appear to exhibit long branches, compositional bias, or any other symptoms of possible rogue taxa that are sensitive to phylogenetic reconstruction. Unfortunately, there is no previous phylogenetic hypothesis of Acridoidea to test the phylogenetic accuracy of our study at this level. Given the fact that different partitioning schemes affect the relationships within Acridoidea, it makes sense to select the partitioning scheme that has the best information criterion score as the most likely solution.

Our study represents one of few empirical examples of exploring data partitioning schemes in a maximum likelihood framework. It has been demonstrated that appropriate partitioning schemes within a likelihood framework can greatly improve the phylogenetic reconstruction (Minin et al., 2003; Pagel and Meade, 2004; Roberts et al., 2009) and our study clearly advocates this point. Most of the data partitioning studies have been done in a Bayesian framework (Brandley et al., 2005; Brown and Lemmon, 2007; Miller et al., 2009) because the current implementation of Bayesian phylogenetics readily allows applying different models to different partitions and analyzing them simultaneously (Nylander et al., 2004). The most common way of assessing partitioning strategies is the use of Bayes Factor (Kass and Raftery, 1995), which has been shown to be a reliable and statistically sound way of finding the optimal solution (Brown and Lemmon, 2007). In contrast, a thorough maximum likelihood analysis is computationally expensive (Holder and Lewis, 2003) and partitioning data with mixed models can introduce a prohibitively large burden for computing likelihood. Currently, there are only two phylogenetic programs available (RAxML and GARLI) that can handle partitioned data, but only the latest version of GARLI (ver. 2.0) can handle mixed model partitioning with some limitations. To our knowledge, there is no maximum likelihood software that allows applying different models to different partitions similar to what has been implemented in Bayesian software such as MrBayes. We analyzed our data using RAxML to take advantage of its fast algorithm as well as its capacity to handle partitioned datasets. One drawback, which is a major weakness of our study, is that the current version of RAxML applies the same model of evolution to each partition separately in a partitioned analysis. Given the objective of our study, however, we argue that our application of a single, identical model to datasets that differ in partitioning schemes allows us to better focus the effect of data partitioning, not confounded by other factors.

4.4. Phylogeny of Acridoidea

The present study represents the most character-rich phylogenetic analysis of Acridomorpha and Acridoidea to date. Although taxon sampling is relatively small, it includes all six acridomorph superfamilies and nine of eleven families within Acridoidea to comment on the higher-level relationships. Below we elaborate on the evolution of grasshoppers based on the topology resulting from the best-fit partitioning scheme (Fig. 2). For simplicity, we also present a reduced phylogeny that depicts family-level relationships (Fig. 3).

Among the outgroups, we find that Eumastacoidea, the second largest superfamily including about 1200 species, is the earliest diverging lineage within Acridomorpha. This group is characterized by the lack of tympanum and monoscleritic penis (Dirsh, 1975) and its most basal placement within Acridomorpha has been consistently supported by both morphology (Amédégnato, 1974; Dirsh, 1975; Eades, 2000) and DNA (Flook et al., 1999; Flook and Rowell, 1997). Our study lends additional support for the basal placement of Eumastacoidea within Acridomorpha. Tanaoceroidea, Pneumoroidea, and Trigonopterygoidea are small relict superfamilies whose phylogenetic relationships have not been well understood (Flook et al., 2000). Tanaoceroidea contains a single family (Tanaoceridae) endemic to the southwestern US, characterized by extremely long antennae and a rudimentary male phallic complex (Dirsh, 1955; Grant and Rentz, 1967; Rehn, 1948). Pneumoroidea also contains one family (Pneumoridae) endemic to South Africa, known for its unique femoroabdominal stridulatory mechanism (Dirsh, 1965). Trigonopterygoidea consists of two divergent families, Trigonopterygidae and Xyronotidae. The former is endemic to the Southeast Asia and characterized by reversed male genitalia and foliaceous tegmina (Dirsh, 1952) and the latter contains four species endemic to central Mexico and can be characterized by rudimentary male genitalia and a stridulatory ridge on the third abdominal tergite (Dirsh, 1955). Dirsh (1975) considered Trigonopterygidae to be highly unusual and placed it its own superfamily Trigonopterygoidea and grouped Pneumoridae, Tanaoceridae, and Xyronotidae into a single superfamily Pneumoroidea based on membranous and rudimentary male genitalia. Based on molecular data, Flook et al. (1999) found that Trigonopterygidae and Xyronotidae formed a monophyletic clade and revised the concept of Trigonopterygoidea, which now includes these two families, but they failed to resolve relationships among other basal lineages. Flook et al. (2000) expanded the taxon sampling of basal acridomorphs and rejected Dirsh's (1975) classification based on their findings which found the following relationship: (Eumastacoidea (Tanaoceroidea (Trigonopterygoidea (Pneumoroidea (Pyrgomorphoidea + Acridoidea))))). Our study corroborates with Flook et al. (2000) in the placement of Tanaoceroidea, but disagrees in that we find a strong monophyletic relationship between Trigonopterygoidea and Pneumoroidea. Since our analysis did not include Xyronotidae, we cannot comment on the monophyly of Trigonopterygoidea, but our novel relationship raises an intriguing possibility of these two geographically separated, divergent lineages' sharing an immediate common ancestor.

The phylogenetic placement of Pyrgomorphidae (the sole member of Pyrgomorphoidea) within Acridomorpha was first hypothesized to be close to Pamphagidae ("Chasmosacci" *sensu* Roberts, 1941) based on the form of the ejaculatory sac (open ejaculatory sac) and associated phallic structures. Both Dirsh (1956) and Amédégnato (1976) considered the family closely related to South African endemic Lentulidae also based on male phallic structures. Using a small molecular dataset, Flook and Rowell (1997) found that Pyrgomorphidae was not related to Pamphagidae and Flook et al. (2000) further confirmed this idea based on a denser taxon sampling of basal Acridomorpha and proposed a new superfamily Pyrgomorphoidea. Synthesizing all known information at the time, paying special attention to male phallic structures, Eades (2000) hypothesized that Pyrgomorphoidea is sister to Acridoidea. Our study corroborates with Eades (2000) in recovering a sister relationship between Pyrgomorphoidea and Acridoidea, despite low nodal support, which also confirms a hypothesis set forth by Eades (1962) that similar male phallic structures must have evolved independently in both Pyrgomorphoidea and Acridoidea.

Our study represents a robust phylogenetic anlaysis of Acridoidea and we report several novel relationships. Here, we discuss about the relationships starting from the base of this monophyletic clade. At the base of Acridoidea, we find a robustly supported clade consisting of Charilaidae and Pamphagidae. Charilaidae, a small family including only five species disjunctly distributed in North and South Africa, has been considered a subfamily of Acrididae (Dirsh, 1953), a subfamily of Pamphagidae (Eades, 1963), and as a separate family (Dirsh, 1956). Dirsh (1975) suggested that Charilaidae might represent a relic lineage that is closely related to Pamphagidae, which has a wider Old World distribution from Africa to Asia. The affinity of Charilaidae to Pamphagidae has been consistently suggested by Dirsh (1956), Amédégnato (1976), and Eades (2000) based on the similarity of phallic structures and our study confirms this hypothesis based on independent molecular data. In our phylogeny, this clade is sister to the remaining Acridoidea and represents the earliest diverging lineage within the superfamily. The phylogenetic position of Pyrgacrididae, which contains two species known only from Mauritius and Réunion Islands in the Western Indian Ocean (Descamps, 1968; Hugel, 2005), has been contentious because of its geographic isolation and the similarity of its male phallic complex to both Pyrgomorphidae and Acrididae (hence the family name) (Eades and Kevan, 1974). Descamps (1968) originally placed it as a subfamily under Acrididae and both Eades and Kevan (1974) and Dirsh (1975) placed it under Pyrgomorphidae. Later Eades (2000) elevated it to its own family and placed it under Acridoidea and suggested that Pyrgacrididae might represent a surviving remnant of a superfamily ancestral to both Pyrgomorphoidea and Acridoidea. Our phylogeny finds that Pyrgacrididae is not closely related to either Pyrgomorphidae or Acrididae, but occupies a rather divergent position near the base of Acridoidea. Mauritius is a volcanic island that emerged in the early Pliocene (6.8–7.8 MYA) and Réunion is even younger (McDougall and Chamalaun, 1969), which suggests that the ancestral Pyrgacrididae that originally colonized the islands cannot be the relict lineage of the common ancestor between Pyrgomorphoidea and Acridoidea.

Another novel relationship in our phylogeny is a sister relationship between two African families, Lithidiidae and Lentulidae. Dirsh (1975) placed Lithidiidae as a subfamily of Acrididae but commented that its phylogenetic position was unclear. Eades (2000) elevated it to family-level because it did not fit in any of the existing families, but suggested a possibility of its affinity to Lentulidae. The sister relationships recovered from our study thus supports Eades' (2000) suggestion. Amédégnato (1976) considered Lentulidae closely related to Pyrgomorphidae based on the similarity of male genitalia, but our analysis does not support her hypothesis, suggesting that the morphological similarity might be due to convergence.

The phylogenetic relationships of the remaining four families included in our study are also novel. We found that Ommexechidae and Romaleidae form a monophyly, which is sister to Acrididae, and these three families are in turn sister to Tristiridae. Of the four families, Ommexechidae, Romaleidae, and Tristiridae are found exclusively in the New World, mainly in South America. Acrididae is cosmopolitan. Considering that all other basal families within Acridoidea are distributed only in the Old World, the placement of South American families within the phylogeny of Acridoidea seems to suggest a possibility of multiple colonization events from the Old World to the New World. Tristiridae comprises 18 genera endemic to the steppe zones of Argentina, Chile and Peru (Cigliano, 1989). Several authors (Amédégnato, 1976; Cigliano, 1989; Eades, 1962; Eades and Kevan, 1974) have expressed that it is a primitive family that has a unique male phallic complex that appears to be intermediate between Robert's (1941) Chasmosacci and Cryptosacci. Because of its morphological affinity to Pyrgomorphidae and Lentulidae, both of which are of the Old World origin, Amédégnato (1993) further hypothesized that Tristiridae might represent remnants from the Antarctic fauna that later colonized South America in the Cenozoic. The phylogenetic position of Tristridae in our study cannot refute Amédégnato's (1993) scenario. Ommexechidae, Romaleidae, and Acrididae are considered modern grasshoppers that fit Robert's (1941) definition of Cryptosacci (Amédégnato, 1976; Eades, 1961). Eades (1961) examined male genitalia of several ommexechids and suggested that the family (which he considered as a subfamily then) had a more affinity to Romaleidae than Acrididae. Amédégnato (1976) expressed the same opinion. Eades (2000), however, later suggested that

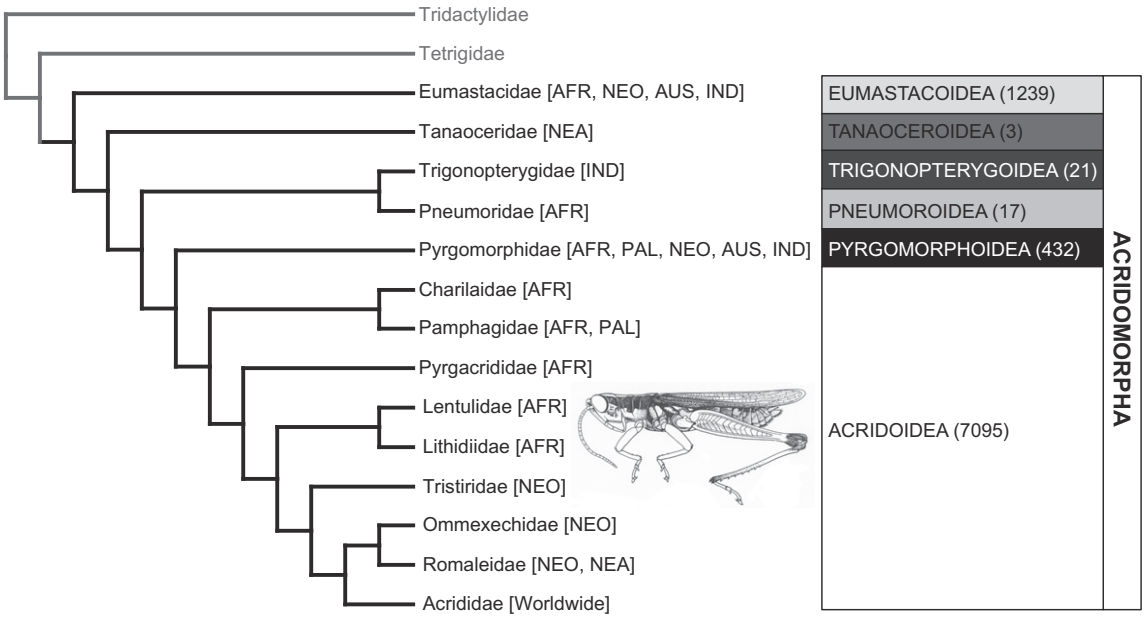

Fig. 3. A family-level phylogeny of Acridomorpha reduced from Fig. 2. The code next to the family name indicates the geographical distribution of each family: AFR, Africa; NEO, Neotropic; AUS, Australia; IND, Indo-Pacific; NEA, Nearctic; PAL, Palearctic. The number next to the superfamily name indicates the number of described species within each superfamily. The figure of the grasshopper was taken from Dirsh (1975).

Ommexechidae was more closely related to Tristiridae. Our study corroborates with the original assessment of Eades (1961) as well as Amédégnato (1976) in finding a sister relationship between Ommexechidae and Romaleidae. Acrididae is the largest family within Acridoidea containing more than 6500 species worldwide. Our taxon sampling within Acrididae does not include all known subfamilies to thoroughly comment on the internal relationships, but we find that the Old World subfamilies (Calliptaminae, Cyrtacantharidinae, Oxyinae, and Catantopinae) are positioned at the base of the clade, which suggests the Old World origin of the family, which has long been suspected (Amédégnato, 1993). In other words, we can speculate that there was an initial divergence between the common ancestor of Ommexechidae and Romaleidae and the ancestral Acrididae, in which the former remained confined mainly to South America and the latter spread from Africa to other parts of the Old World. Following this, there must have been several independent colonization events from the Old World to the New World by ancestral acridids at different times. It is however unclear when the original split might have happened because fossil data for modern grasshoppers are notoriously lacking.

4.5. Final remarks and recommendations

In this study, we have thoroughly demonstrated the effect of data partitioning in a maximum likelihood framework in the context of the mitochondrial phylogenomics of Acridoidea. We have shown that the *a priori* partitioning schemes are suboptimal to the best-fit scheme suggested by PartitionFinder. Based on our study, we recommend the following procedures when analyzing mtgenome data for phylogenetics, but the same idea can be applied to any phylogenetic analyses utilizing multiple loci. The first step is to dissect the dataset into as many partitions as possible when pre-defining the configuration file for PartitionFinder. In our case, we created a total of 87 partitions, which incorporated not only the individual codon positions of PCGs, but also stems and loops of rRNAs and tRNAs. This step is crucial because the finely dissected input data provide PartitionFinder more combinations to search through, therefore increasing the likelihood of finding the most appropriate way of partitioning the data. The "greedy" algorithm implemented in PartitionFinder is an effective heuristic search tool that can search through the vast space of possible partitioning schemes. For our study, this space consisted of $B_{87} = 6.39 \times 10^{96}$ possible schemes and the program compared 109,737 schemes. Admittedly, the schemes that were compared represent only a fraction of the total space, but the best-fit scheme selected by this method produced the mostly likely result, better than any of the *a priori* schemes. We strongly recommend that any future studies of mitochondrial phylogenomics should include a thorough examination of the effect of data partitioning.

Acknowledgments

Funding was provided by National Science Foundation Grants EF-0531665 to MFW and DEB-0816962 (DEB-1064082) to HS and MFW. The following collaborators kindly provided DNA-grade tissue samples: the late Christiane Amédégnato, Maria Marta Cigliano, Michel Lecoq, Paul Lenhart, and Lauren Spearman. Specimens from South Africa were collected and exported under the following permits to HS: ECP CRP 177/08CR and 178/08CR, ECPTA Research Agreement (RA0042). We thank Matt Moulton for help with data generation as well as Rebecca Buckman, Joseph Mugleston, Matthew Bendall, Heather Bracken, Jesse Breinholt, and Seth Bybee for help with the various analyses implemented in this project. We thank Nathan Sheffield for the use of the program MOSAS.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.02.019.

References

Amédégnato, C., 1974. Les genres d'acridiens neotropicaux, leur classification par familles, sous-familles et tribus. Acrida 3, 193–203.
Amédégnato, C., 1976. Structure et évolution des genitalia chez les Acrididae et familles apparentées. Acrida 5, 1–15.
Amédégnato, C., 1993. African-American relationships in the Acridians (Insecta, Orthoptera). In: George, W., Lavocat, R. (Eds.), The Africa-South America Connection. Clarendon Press, Oxford, pp. 59–75.
Bell, E.T., 1934. Exponential numbers. Am. Mat. Mon. 41, 411–419.
Bensasson, D., Zhang, D.-X., Hewitt, G.M., 2000. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. Mol. Biol. Evol. 17, 406–415.
Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54, 373–390.
Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56, 643–655.
Cameron, S.L., Barker, S.C., Whiting, M.F., 2006a. Mitochondrial genomics and the new insect order Mantophasmatodea. Mol. Phylogenet. Evol. 38, 274–279.
Cameron, S.L., Beckenbach, A.T., Dowton, M., Whiting, M.F., 2006b. Evidence from mitochondrial genomics on interordinal relationships in insects. Arthropod. Syst. Phylogeny 64, 27–34.
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., Miller, K.B., D'Haese, C.A., Whiting, M.F., Barker, S.C., 2004. Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea *sensu lato* (Athropoda). Cladistics 20, 534–557.
Cameron, S.L., Sullivan, J., Song, H., Miller, K.B., Whiting, M.F., 2009. A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. Zool. Scr. 38, 575–590.
Castro, L.R., Dowton, M., 2007. Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers. Syst. Entomol. 32, 60–69.
Chang, J.T., 1996. Full reconstruction of Markov models on evolutionary trees: identifiability and consistency. Math. Biosci. 137, 51–73.
Cigliano, M.M., 1989. A cladistic analysis of the family Tristiridae (Orthoptera, Acridoidea). Cladistics 5, 379–393.
Clary, D.O., Wolstenhome, D.R., 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22, 252–271.
Covacin, C., Shao, R., Cameron, S., Barker, S.C., 2006. Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). Insect Mol. Biol. 15, 63–68.
Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375.
Descamps, M., 1968. Un Acridoide relique des Mascareignes (Orth. Acridoidea). Bull. Soc. Entomol. Fr. 73, 31–36.
Ding, F.-M., Shi, H.-W., Huang, Y., 2007. Complete mitochondrial genome and secondary structures of lrRNA and srRNA of *Atractomorpha sinensis* (Orthoptera, Pyrgomorphidae). Zool. Res. 29, 580–588.
Dirsh, V.M., 1952. The restroration of the subfamily Trigonopteryginae Walker (Orthoptera, Acrididae). Ann. Mag. Nat. Hist. 5, 82–84.
Dirsh, V.M., 1953. Charilainae, a new subfamily of Acrididae (Orthoptera). Ann. Mag. Nat. Hist. 6, 161–173.
Dirsh, V.M., 1955. Tanaoceridae and Xyronotidae: two new families of Acridoidea (Orthoptera). Ann. Mag. Nat. Hist. 8, 285–288.
Dirsh, V.M., 1956. The phallic complex in Acridoidea (Orthoptera) in relation to taxonomy. Trans. R. Entomol. Soc. Lond. 108, 223–356.
Dirsh, V.M., 1965. Revision of the family Pneumoridae (Orthoptera: Acridoidea). Bull. Br. Mus. (Nat. Hist.) Entomol. 15, 325–396.
Dirsh, V.M., 1975. Classification of the Acridomorphoid Insects. E.W. Classey Ltd., Faringdon, Oxon.
Dowton, M., Cameron, S.L., Austin, A.D., Whiting, M.F., 2009a. Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera – a lineage with both rapidly and slowly evolving mitochondrial genomes. Mol. Phylogenet. Evol. 52, 512–519.
Dowton, M., Cameron, S.L., Dowavic, J.I., Austin, A.D., Whiting, M.F., 2009b. Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. Mol. Biol. Evol. 26, 1607–1617.
Eades, D.C., 1961. The tribes and relationships of the Ommexechinae (Orthoptera, Acrididae). Proc. Acad. Nat. Sci. Phila. 113, 157–172.
Eades, D.C., 1962. Phallic structures, relationships, and components of the Dericorythinae (Orthoptera; Acrididae). Not. Nat. 354, 1–9.

Eades, D.C., 1963. Observations on *Charilaus* and Charilainae (Orthoptera, Pamphagidae). Entomol. News 74, 131–133.

Eades, D.C., 2000. Evolutionary relationships of phallic structures of Acridomorpha (Orthoptera). J. Orthoptera Res. 9, 181–210.

Eades, D.C., Kevan, D.K.M., 1974. The phallic musculature of Pyrgomorphidae, with particular reference to *Atractomorpha sinensis sinensis* Bolivar, and notes on the family Tristiridae and the subfamily Pyrgacridinae, nov. (Orthoptera: Acridoidea). Acrida 3, 247–265.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32, 1792–1797.

Erler, S., Ferenz, H.-J., Moritz, R.F.A., Kaatz, H.-H., 2010. Analysis of the mitochondrial genome of *Schistocerca gregaria gregaria* (Orthoptera: Acrididae). Biol. J. Linn. Soc. 99, 296–305.

Fenn, J.D., Song, H., Cameron, S.L., Whiting, M.F., 2008. A mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. Mol. Phylogenet. Evol. 49, 59–68.

Flook, P.K., Klee, S., Rowell, C.H.F., 1999. Combined molecular phylogenetic analysis of the Orthoptera (Arthropoda, Insecta) and implications for their higher systematics. Syst. Biol. 48, 233–253.

Flook, P.K., Klee, S., Rowell, C.H.F., 2000. Molecular phylogenetic analysis of the Pneumoroidea (Orthoptera, Caelifera): Molecular data resolve morphological character conflicts in the basal Acridomorpha. Mol. Phylogenet. Evol. 15, 345–354.

Flook, P.K., Rowell, C.H.F., 1997. The phylogeny of the Caelifera (Insecta, Orthoptera) as deduced from mtrRNA gene sequences. Mol. Phylogenet. Evol. 8, 89–103.

Flook, P.K., Rowell, C.H.F., Gellissen, G., 1995a. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. 41, 928–941.

Flook, P.K., Rowell, C.H.F., Gellissen, G., 1995b. Homoplastic rearrangements of insect mitochondrial transfer-RNA genes. Naturwissenschaften 82, 336–337.

Grant Jr., H.J., Rentz, D.C.F., 1967. A biosystematic review of the family Tanaoceridae including a comparative study of the proventriculus. Pan-Pac. Entomol. 43, 65–74.

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. 4, 275–284.

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

Hugel, S., 2005. Redécouverte du genre *Pygacris* à l'île de la Réunion: description du mâle de *P. descampi* Kevan, 1975 (Orthoptera, Caelifera). Bull. Soc. Entomol. Fr. 110, 153–159.

Jex, A.R., Hu, M., Littlewood, D.T.J., Waeschenback, A., Gasser, R.B., 2008. Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). BMC Genom. 9, 11.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucl. Acids Res. 33, 511–518.

Kevan, D.K.M., 1982. Orthoptera. In: Parker, S.P. (Ed.), Synopsis and Classification of Living Organisms. McGraw-Hill Book Company, Inc., New York, pp. 352–383.

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

Laslett, D., Canbäck, B., 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24, 172–175.

Li, C., Lu, G., Orti, G., 2008. Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. Syst. Biol. 57, 519–539.

Liu, Y., Huang, Y., 2008. Sequencing and analysis of complete mitochondrial genome of *Chorthippus chinensis* Tarb. Chinese J. Biochem. Mol. Biol. 24, 329–335.

Lowe, T.M., Eddy, S.R., 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl. Acids Res. 25, 955–964.

Ma, C., Liu, C., Yang, P., Kang, L., 2009. The complete mitochondrial genomes of two band-winged grasshoppers, *Gastrimargus marmaratus* and *Oedaleus asiaticus*. BMC Genom. 10, 156.

Maddison, D.R., Maddison, W.P., 2005. MacClade 4. Sinauer Associates, Inc., Sunderland, Massachusetts.

Maddison, W.P., Maddison, D.R., 2010. Mesquite: A Modular System for Evolutionary Analysis. Verson 2.73. <http://mesquiteproject.org>.

Matt, S., Flook, P.K., Rowell, C.H.F., 2008. A partial molecular phylogeny of the Eumastacoidea *s. lat.* (Orthoptera, Caelifera). J. Orthoptera Res. 17, 43–55.

McDougall, I., Chamalaun, F.H., 1969. Isotopic dating and geomagnetic polarity studies on volcanic rocks from Mauritius, Indian Ocean. Geol. Soc. Am. Bull. 80, 1419–1442.

McGuire, J.A., Witt, C.C., Altshuler, D.L., Remsen Jr., J.V., 2007. Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. Syst. Biol. 56, 837–856.

Mikkelsen, M., Rockenbauer, E., Wächter, A., Fendt, L., Zimmermann, B., Parson, W., Nielsen, S.A., Gilbert, T., Willerslev, E., Morling, N., 2009. Application of full mitochondrial genome sequencing using 454 GS FLX pyrosequencing. Forensic. Sci. Int. Genet. Suppl. Ser. 2, 518–519.

Miller, K.B., Bergsten, J., Whiting, M.F., 2009. Phylogeny and classification of the tribe Hydaticini (Coleoptera: Dytiscidae): partition choice for Bayesian analysis with multiple nuclear and mitochondrial protein-coding genes. Zool. Scr. 38, 591–615.

Miller, M.A., Holder, M.T., Vos, R., Midford, P.R., Liebowitz, T., Chan, L., Hoover, P., Warnow, T., 2011. The CIPRES Portals. CIPRES. <http://www.phylo.org/sub_sections/portal>.

Minin, V., Abdo, Z., Joyce, P., Sullivan, J., 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. 52, 674–683.

Moulton, M.J., Song, H., Whiting, M.F., 2010. Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). Mol. Ecol. Resour. 10, 615–627.

Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R., Frati, F., 2003. Hexapod origins: Monophyletic or paraphyletic? Science 299, 1887–1889.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.

Pagel, M., Meade, A., 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Syst. Biol. 53, 571–581.

Philippe, H., Delsuc, F., Brinkmann, H., Lartillot, N., 2005. Phylogenomics. Annu. Rev. Ecol. Evol. Syst. 36, 541–562.

Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21, 1455–1458.

Pons, J., Ribera, I., Bertranpetit, J., Balke, M., 2010. Nucleotide substitution rates for the full set of mitochondrial protein-coding genes in Coleoptera. Mol. Phylogenet. Evol. 56, 796–807.

Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.

Puigbò, P., Garcia-Vallvé, S., McInerney, J.O., 2007. TOPD/FMTS: a new software to compare phylogenetic trees. Bioinformatics 23, 1556–1558.

Rehn, J.A.G., 1948. The locust genus *Tanaocerus* as found in the United States, and the description of a related new genus (Orthoptera: Acridoidea). Proc. Acad. Nat. Sci. Phila. 100, 1–22.

Roberts, H.R., 1941. A comparative study of the subfamilies of the Acrididae (Orthoptera) primarily on the bases of their phallic structures. Proc. Acad. Nat. Sci. Phila. 93, 201–246.

Roberts, T.E., Sargis, E.J., Olson, L.E., 2009. Networks, trees, and treeshrews: assessing support and identifying conflict with multiple loci and a problematic root. Syst. Biol. 58, 257–270.

Shao, R., Barker, S.C., 2003. The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. Mol. Biol. Evol. 20, 362–370.

Shao, R., Campbell, N.J., Barker, S.C., 2001. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol. Biol. Evol. 18, 858–865.

Sharov, A.G., 1968. Phylogeny of the Orthopteroidea. Akad. Nauk SSSR Trudy Paleontol. Inst. 118, 1–216.

Sheffield, N.C., Hiatt, K.D., Valentine, M.C., Song, H., Whiting, M.F., 2010. Mitochondrial genomics in Orthoptera using MOSAS. Mitochondrial DNA 21, 87–104.

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.

Shi, H.-W., Ding, F.-M., Huang, Y., 2008. Complete sequencing and analysis of mtDNA in *Phlaeoba albonema* Zheng. Chinese J. Biochem. Mol. Biol. 24, 604–611.

Simon, C., Buckley, T.R., Frati, F., Stewart, J.B., Beckenbach, A.T., 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compliation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 37, 545–579.

Song, H., 2010. Grasshopper systematics: past, present and future. J. Orthoptera Res. 19, 57–68.

Song, H., Buhay, J.E., Whiting, M.F., Crandall, K.A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc. Natl. Acad. Sci. USA 105, 13486–13491.

Song, H., Sheffield, N.C., Cameron, S.L., Miller, K.B., Whiting, M.F., 2010. When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. Syst. Entomol. 35, 429–448.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A Rapid Bootstrap Algorithm for the RAxML Web-Servers. Syst. Biol. 75, 758–771.

Sun, H., Zheng, Z., Huang, Y., 2010. Sequence and phylogenetic analysis of complete mitochondrial DNA genomes of two grasshopper species *Gomphocerus rufus* (Linnaeus, 1758) and *Primnoa arctica* (Zhang and Jin, 1985) (Orthoptera: Acridoidea). Mitochondrial DNA 21, 115–131.

Stocsits, R.R., Letsch, H., Hertel, J., Misof, B., Stadler, P.F., 2009. Accurate and efficient reconstruction of deep phylogenies from structured RNAs. Nucl. Acids Res. 37, 6184–6193.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Timmermans, M.J.T.N., Dodsworth, S., Culverwill, C.L., Bocak, K., Ahrens, D., Littlewood, D.T.J., Pons, J., Vogler, A.P., 2010. Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. Nucl. Acids Res. 38, e197.

Wolstenhome, D.R., 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141, 173–216.

Zeuner, F.E., 1942. The Locustopsidae and the phylogeny of the Acridodea (Orthoptera). Proc. R. Entomol. Soc. Lond. B. 11, 1–18.

Zhang, C., Huang, Y., 2008. Complete mitochondrial genome of *Oxya chinensis* (Orthoptera, Acridoidea). Acta Biochim. Biophys. Sin. 40, 7–18.

Zhang, D., Zhi, Y., Yin, H., Li, X., Yin, X., 2011. The complete mitochondrial genome of *Thrinchus schrenkii* (Orthoptera: Caelifera, Acridoidea, Pamphagidae). Mol. Biol. Rep. 38, 611–619.

Zhao, L., Zheng, Z.-M., Huang, Y., Sun, H.-M., 2010. A comparative analysis of mitochondrial genomes in Orthoptera (Arthropoda: lnsecta) and genome descriptions of three grasshopper species. Zool. Sci. 27, 662–672.