Molecular Phylogenetics and Evolution 49 (2008) 59–68

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data

J. Daniel Fenn [a,*], Hojun Song [a], Stephen L. Cameron [b], Michael F. Whiting [a]

[a] Department of Biology, Brigham Young University, Provo, UT 84602, USA
[b] Australian National Insect Collection, CSIRO Entomology, PO Box 1700, Canberra, ACT 2601, Australia

ARTICLE INFO

Article history:
Received 6 December 2007
Revised 26 June 2008
Accepted 4 July 2008
Available online 15 July 2008

Keywords:
Mitochondrial genome
Phylogeny
Orthoptera
Data-partitioning
Rearrangements

ABSTRACT

The phylogenetic utility of mitochondrial genomes (mtgenomes) is examined using the framework of a preliminary phylogeny of Orthoptera. This study presents five newly sequenced genomes from four orthopteran families. While all ensiferan and polyneopteran taxa retain the ancestral gene order, all caeliferan lineages including the newly sequenced caeliferan species contain a tRNA rearrangement from the insect ground plan $tRNA^{Lys}$(K)–$tRNA^{Asp}$(D) swapping to $tRNA^{Asp}$ (D)–$tRNA^{Lys}$ (K) confirming that this rearrangement is a possible molecular synapomorphy for this suborder. The phylogenetic signal in mtgenomes is rigorously examined under the analytical regimens of parsimony, maximum likelihood and Bayesian inference, along with how gene inclusion/exclusion, data recoding, gap coding, and different partitioning schemes influence the phylogenetic reconstruction. When all available data are analyzed simultaneously, the monophyly of Orthoptera and its two suborders, Caelifera and Ensifera, are consistently recovered in the context of our taxon sampling, regardless of the optimality criteria. When protein-coding genes are analyzed as a single partition, nearly identical topology to the combined analyses is recovered, suggesting that much of the signals of the mtgenome come from the protein-coding genes. Transfer and ribosomal RNAs perform poorly when analyzed individually, but contribute signal when analyzed in combination with the protein-coding genes. Inclusion of third codon position of the protein-coding genes does not negatively affect the phylogenetic reconstruction when all genes are analyzed together, whereas recoding of the protein-coding genes into amino acid sequences introduces artificial resolution. Over-partitioning in a Bayesian framework appears to have a negative effect in achieving convergence. Our findings suggest that the best phylogenetic inferences are made when all available nucleotide data from the mtgenome are analyzed simultaneously, and that the mtgenome data can resolve over a wide time scale from the Permian (~260 MYA) to the Tertiary (~50 MYA).

1. Introduction

Mitochondrial genomes (mtgenomes) are gaining increasing popularity in higher-level phylogenetic analyses because of their ability to provide better resolution for deep relationships than single or multi-gene analyses and for the feasibility of sequencing the entire genome with relative ease (Cameron et al., 2004; Nardi et al., 2003; Saccone et al., 1999). With more than 15,000 bp of nucleotide data and 37 genes, mtgenomes are one of the most information-rich markers in phylogenetics. In addition to the nucleotide data, other phylogenetically useful information can be extracted from the mtgenomes such as gene rearrangements (Boore and Brown, 1998), gene insertion or deletion events (Rokas and Holland, 2000), and genic or intergenic length variability (Schneider and Ebert, 2004). The phylogenetic utility of mtgenomes has been studied rigorously in the past few years, especially for insects and related groups (Cameron et al., 2007, 2004; Carapelli et al., 2007; Kjer and Honeycutt, 2007). Different methods of analyzing mtgenomes have been proposed and applied, spanning from selecting a few key genes from the entire mtgenome data (Nardi et al., 2003) to inclusion of all available genes excluding control region (Castro and Dowton, 2007), and from analyzing amino acid sequences (Nardi et al., 2001) or purine/pyrimidine recoding of nucleotide sequences (Saitoh et al., 2006) to the traditional approach of analyzing native nucleotide sequences (Cameron et al., 2007). Different optimality criteria and dataset compilation techniques have also been applied to find the best method of analyzing complex mtgenome data (Cameron et al., 2004; Castro and Dowton, 2005; Kim et al., 2005; Stewart and Beckenbach, 2003). The current view, as advanced by Cameron et al. (2007), is that the mtgenome data recover the most phylogenetic information when all available genes are analyzed as nucleotide sequences and the results from different optimality criteria are then

* Corresponding author. Fax: +1 801 422 0090.
E-mail address: danfenn77@gmail.com (J.D. Fenn).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.07.004

compared for the sensitivity of resulting phylogenetic hypotheses to the underlying model assumptions of the tree building methods. In terms of the ability to resolve deeper level relationships in insects, mtgenome data were not able to unambiguously resolve the relationships among major arthropod lineages (Cameron et al., 2004), but were able to resolve intraordinal relationships within Diptera (Cameron et al., 2007) and Hymenoptera (Castro and Dowton, 2007). This suggests that mtgenome data may not have sufficient phylogenetic signal to resolve ancient relationships of the Cambrian to Devonian (600–360 MYA), but may provide strong signals for resolving intraordinal relationships within Diptera and Hymenoptera whose first fossils are found dating to the Upper Triassic (~225 MYA). These findings imply that the maximum resolving capacity of insect mtgenomes may lie somewhere between these two geological times.

Orthoptera is an excellent group to test the phylogenetic utility of mtgenomes. The order is one of the oldest extant insect lineages, with fossils first appearing in the Upper Carboniferous (~290 MYA) (Grimaldi and Engel, 2005; Sharov, 1968), a level of divergence that mtgenomes have not yet been used to resolve. Orthoptera contains many familiar insects including grasshoppers, locusts, katydids, crickets, mole crickets, and wetas, and the order is well known for the ability to jump and their utilization of acoustic communication. It has traditionally been considered a monophyletic group based on several morphological synapomorphies including hind legs adapted for jumping, well-separated hind coxae, a shield-like pronotum with large lateral lobes, and hind tibiae with two rows of dorsal spines (Kevan, 1982; Rentz, 1991). It consists of two suborders, Caelifera (short-horned grasshoppers) and Ensifera (crickets and katydids) that differ in their mtgenome arrangements. Flook et al. (1995b) first discovered that the order of transfer RNAs in the migratory locust *Locusta migratoria* was different from that of most other insect mtgenomes in that it had $tRNA^{Asp}$–$tRNA^{Lys}$ between *cox2* and *atp8* genes, while the typical order is $tRNA^{Lys}$–$tRNA^{Asp}$ in most insects. Flook et al. (1995a) subsequently proposed that this novel rearrangement was synapomorphic for the Caelifera as it occurred in the caeliferan families, Acrididae and Pyrgomorphidae, but not in any of the ensiferan lineages or the other polyneopterous orders studied. Two additional ensiferan mtgenomes have since been published (Kim et al., 2005; Fenn et al., 2007), both of which had the ancestral insect gene arrangement, consistent with the proposal by Flook et al. (1995a).

Here, we present the complete mtgenomes of five new orthopteran species, including representatives of both suborders and four families. Using these new mtgenomes in addition to the previously published mtgenomes of Orthoptera and Polyneoptera, we reconstruct a preliminary phylogeny of Orthoptera as a vehicle to examine the phylogenetic utility of mtgenome data in resolving deep relationships within the order. We do not propose a definitive phylogenetic relationship of Orthoptera due to relatively small taxon sampling, but instead show that mtgenomes are resourceful tools in elucidating these deep relationships. Also, we explore various methods of analyzing mtgenome data in a phylogenetic framework, by testing the effects of different optimality criteria, data partitioning strategies, and data transformation. Finally, we study whether the gene rearrangement provides a useful phylogenetic signal in Orthoptera in light of the five novel mtgenomes.

2. Materials and methods

2.1. Taxon sampling

A total of 14 taxa were analyzed in this study, including eight orthopteran ingroup and six polyneopteran outgroup taxa. We sequenced five new orthopteran species for this study: *Acrida willemsei* (Acrididae), *Calliptamus italicus* (Acrididae), *Gryllotalpa pluvalis* (Gryllotalpidae), *Myrmecophila manni* (Myrmecophilidae), and *Troglophilus neglectus* (Raphidophoridae). Three additional orthopteran species were obtained from previous studies: *Anabrus simplex* (Tettigoniidae, Fenn et al., 2007), *Gryllotalpa orientalis* (Gryllotalpidae, Kim et al., 2005), and *Locusta migratoria* (Acrididae, Flook et al., 1995b). The ingroup sampling therefore represents 2 suborders, 4 superfamilies, and 5 families within Orthoptera. After this manuscript was already well advanced, three new orthopteran genomes have recently become available on GenBank: *Ruspolia dubia* (Zhou et al., 2007a), *Deracantha onos* (Zhou et al., 2007b) and *Oxya chinensis* (Zhang and Huang, 2007). The first two of these genomes are representatives from the family Tettigoniidae (the families, Conocephalidae and Bradyporidae, used by these authors are usually considered subfamilies, Conocephalinae and Bradyporinae, respectively, within Tettigoniidae) whereas the third is a member of the Acrididae. As we already include multiple representatives of these two families and the fact that the genome sequences only became available during the final phase of manuscript preparation we feel justified in not including them at this point. For outgroups we sampled *Grylloblatta sculleni* (Grylloblatodea, Cameron et al., 2006), *Sclerophasma paresisensis* (Mantophasmatodea, Cameron et al., 2006), *Periplaneta fuliginosa* (Blattodea, Yamauchi et al., 2004), *Reticulitermes hageni* (Isoptera, Cameron and Whiting, 2007) *Tamolanica tamolana* (Mantodea, Cameron et al., 2006), and *Timema californicum* (Phasmatodea, Cameron et al., 2006). *Grylloblata* was used to root the trees (Terry and Whiting 2005). This outgroup sampling represents all currently available polyneopteran mtgenomes except for *Pteronarcys princeps* (Plecoptera, Stewart and Beckenbach, 2006). Preliminary analyses suggest that *P. princeps* is a wildcard taxon that introduces instability in phylogenetic reconstruction. Inclusion of this taxon in the analyses resulted in different tree topologies under different optimality criteria and data partitioning, but exclusion of this taxon consistently resulted in a single topology. We suspected a possible effect from base composition heterogeneity that might have been present in *P. princeps*. The mean base composition of all taxa included in our study except *P. princeps* was 32% A, 41% T, 14% G, and 13% C with low variance, while that of *P. princeps* was 37% A, 34% T, 11% G, and 17% C. Similar compositional differences were found when individual codon positions were analyzed separately (see Supplementary Table 1). Therefore, we concluded that the instability introduced by the inclusion of this taxon was possibly due to the base compositional bias. Despite the fact that there is a well-established distance-based method to correct for this type of bias, we decided not to include *P. princeps* in our analysis for the following reasons. First, Plecoptera has never been seriously considered a closely related group to Orthoptera and other outgroup taxa included in our study are sufficient to test the monophyly of Orthoptera in the context of our taxon sampling. Second, we are interested in examining the phylogenetic signal from the mtgenome data, and a simple phenetic method does not contribute much in terms of achieving this objective. Lastly, base compositional heterogeneity is a difficult bias to overcome, especially in the mtgenome data that consist of 37 individual genes, while it might be a relatively simple task in a single-gene study. We are in the process of addressing this very issue in a separate research project. Thus, we feel that our decision to exclude *P. princeps* in this study is justified. Collecting information and GenBank accession numbers for the taxa used in this study are shown in Table 1.

2.2. MtGenome amplification and sequencing

We employed the primer walking technique in generating the complete mtgenomes sequenced for this study. First, short gene

Table 1
Complete taxon sampling used in current study

Order	Family	Genus	Species	Locality	Voucher No.[a]	GenBank [b]	Reference
Grylloblatodea	Grylloblattidae	*Grylloblatta*	*sculleni*	—	—	DQ241796	Cameron et al. (2006)
Isoptera	Rhinotermitidae	*Reticulitermes*	*hageni*	—	—	NC_009501	Cameron and Whiting (2007)
Blattodea	Blattidae	*Periplaneta*	*fuliginosa*	—	—	NC_006076	Yamauchi et al. (2004)
Mantodea	Mantidae	*Tamolanica*	*tamolana*	—	—	NC_007702	Cameron et al. (2006)
Phasmatodea	Timematidae	*Timema*	*californicum*	—	—	DQ241799	Cameron et al. (2006)
Mantophasmatodea	Mantophasmatidae	*Sclerophasma*	*paresisensis*	—	—	NC_007701	Cameron et al. (2006)
Orthoptera	Acrididae	*Locusta*	*migratoria*	—	—	NC_001712	Flook et al. (1995)b
Orthoptera	Acrididae	*Acrida*	*willemsei*	Malaysia, Sabah, Croker Rng. NP, HQ station Rd.	OR059	EU938372	This study.
Orthoptera	Acrididae	*Calliptamus*	*italicus*	Slovenia, Brje pri Komnu	OR073	EU938373	This study.
Orthoptera	Raphidophoridae	*Troglophilus*	*neglectus*	Slovenia, Brje pri Komnu	OR083	EU938374	This study.
Orthoptera	Tettigoniidae	*Anabrus*	*simplex*	—	—	NC_009967	Fenn et al. (2007)
Orthoptera	Myrmecophilidae	*Myrmecophila*	*manni*	USA, WA, Pine Valley Rec. Area	OR022	EU938370	This study.
Orthoptera	Gryllotalpidae	*Gryllotalpa*	*orientalis*	—	—	NC_006678	Kim et al. (2005)
Orthoptera	Gryllotalpidae	*Gryllotalpa*	*pluvalis*	Australia, Queensland, Brisbane, Mitchelton	OR028	EU938371	This study.

Collection localities provided for the new genome sequences provided.
[a] Whiting lab IGC (Insect Genomic Collection) voucher number.
[b] GenBank Accession Numbers.

regions within individual genes (*cox1*, *cox2*, *cytb*, *lrRNA*, and *srRNA*) were amplified and sequenced using conserved insect primers (see Supplementary Table 2 for the primers used). The sequences obtained from these regions were then used to design specific primers to amplify the entire mtgenome in up to 7 fragments ($tRNA^{Met}$ to *cox1*; *cox1* to *cox3*; *cox3* to *nad4*; *nad4* to *cytb*; *cytb* to *lrRNA*; *lrRNA* to *srRNA*; *srRNA* to $tRNA^{Met}$) by long PCR reactions. Specific sequencing primers were designed using sequenced portions of the mtgenome to provide additional sequencing and the cycle repeated to "walk" around the remainder of the mtgenome (Sequencing primers available from JDF upon request). Special care was taken to avoid nuclear coded mitochondrial pseudogenes (numts), which are known to be present in Orthoptera (Bensasson et al., 2001; Spooner and Ritchie, 2006). Previous studies have shown that long PCR amplification of templates greater than 2000 bp in size drastically reduces the chance of amplifying numts (Akanuma et al., 2000). We therefore, performed long PCRs for as much of the mtgenome as possible. By sequencing both majority (J) and minority (N) strands of the mtgenome, we were able to identify possible numts when mismatches arose. In such cases, additional long PCRs were performed to amplify regions outside of the presumptive numts to insure we obtained only genuine mitochondrial sequences.

Long PCRs were performed using the Elongase PCR kit (Invitrogen) with the following cycling conditions: 95 °C for 12 min, 40 cycles of 94 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min and a final elongation of 72 °C for 7 min. Short PCRs were performed where needed using the Elongase enzyme with the following cycling conditions: 92 °C for 2 min, 40 cycles of 92 °C for 30 s, 50 °C for 30 min, 68 °C for 12 min, with a final run-out step of 68 °C for 20 min. Sequencing was performed using ABI BigDye version3 dye terminator chemistry and then fractionated on the ABI 3770 capillary sequencer. Sequencing PCR conditions were as follows: 96 °C for 1 min, 25 cycles of 96 °C for 10 s, 50 °C for 5 s, 60 °C for 75 s. We were unable to successfully sequence the entire control region of some species through primer walking due to highly repetitive A+T-rich portions of this region. These ambiguous regions were resolved using the Topo-TA cloning chemistry (Invitrogen).

2.3. Annotation and alignment

Raw sequence files were proofread and aligned into contigs in Sequencher (ver. 4.2 and 4.6) (GeneCodes Corporation). Contig sequences were checked for ambiguous base calls and only non-ambiguous regions were used for annotation. Transfer RNA analysis and identification was conducted with tRNAscan-SE (Lowe and Eddy, 1997) using mitochondrial predictors and a cove score cut off of one. Reading frames between tRNAs were identified in Sequencher and genes identified using translated BLAST searches (blastn) (Altschul et al., 1997) as implemented in the NCBI website http://www.ncbi.nlm.nih.gov/. Circular mtgenome maps were created based on this annotation using CGView (Stothard and Wishart, 2005), and they were compared against the published orthopteran mtgenomes for *Anabrus, Gryllotalpa*, and *Locusta* (see Supplementary Figs. 1–5).

Annotated mtgenomes were organized into datasets using MEGA 3.0 (Kumar et al., 2004) with each of the 37 genes aligned separately. Protein-coding genes were translated into amino acid sequence using the invertebrate mitochondrial genetic code in MEGA, and aligned based on their amino acid sequence using default settings in ClustalW (as implemented in MEGA). The alignment was back-translated into the corresponding nucleotide sequences. Ribosomal and transfer RNA genes were aligned individually in MUSCLE (Edgar, 2004) using the default settings. Datasets were concatenated in MacClade (Maddison and Maddison, 2003) using the partitioning scheme presented below.

2.4. Phylogenetic analyses and data partitioning

We tested the effect of different inference methods on topology and nodal support in mtgenome phylogenies using parsimony (MP), maximum likelihood (ML) and Bayesian (BA) analyses (See Supplementary Figs. 9–46 for all trees obtained). MP analyses were performed using PAUP* ver. 4.0b10 (Swofford, 2002), with gaps treated both as missing or as a 5th character state (21st character state for amino acid analyses). A total of 1,000,000 random addition searches using TBR were performed for each MP analysis. Bootstrap support was calculated from 1500 bootstrap replicates with 500 random additions per replicate in PAUP*. Partitioned Bremer supports (Baker and DeSalle, 1997) were calculated using a script (Peña et al., 2006) implemented in TNT 1.1 (Goloboff et al., 2003) to assess the relative contribution of each dataset partition to the combined topology. Tree statistics were also calculated in PAUP* (Table 2). Models for ML analyses were chosen for each concatenated dataset using AIC as implemented in ModelTest (Posada and Crandall, 1998). Heuristic ML searches were performed using 1000 random additions in PAUP*. Bootstrap support for ML trees was calculated using 1,500 bootstrap replicates with 10 random additions per replicate. Models for BA analyses were also chosen using AIC, with models assessed independently for each partition

Table 2
Tree statistics for the major analyses under parsimony (PA), maximum likelihood (ML) and Bayesian (BA)

		PA								ML	BA
	No. of characters	Length	CI	RI	RC	Informative characters	Percent informative	CI informative	No. of trees [a]	−Ln*L*	Log likelihood
ALL_DNA123	15377	32781	0.5017	0.3523	0.1767	7719	50.2	0.4586	1	139600.04	136512.5425
ALL_DNA12	11593	18600	0.5387	0.4038	0.2175	4598	39.7	0.4819	1	87704.54	85782.063
ALL_PROT	7809	16558	0.6308	0.4363	0.2752	3547	45.4	0.5892	1	n/a	89174.01
PCG_DNA123	11352	25787	0.4903	0.3463	0.1698	5960	52.5	0.4551	1	108034.77	105698.0705
PCG_DNA12	7568	11604	0.5357	0.4196	0.2248	2839	37.5	0.4874	1	55936.39	54962.8755
PCG_PROT	3784	9567	0.6939	0.4895	0.3397	1788	47.3	0.6686	1	n/a	58322.5495
RIBO	2406	4512	0.5499	0.3745	0.2059	1122	46.6	0.4818	2	19344.70	19362.948
TRAN	1619	2435	0.4657	0.4029	0.2187	637	39.3	0.4657	1	11372.06	54962.92225

[a] Number of most parsimonious trees found. If more than one tree was found a strict consensus was performed.

used in the concatenated datasets. All BA analyses ran four separate runs using unlinked partitions with four chains per run for a total of twenty million generations per run with sampling every 1000 generations in MrBayes versions 3.1.1 (Ronquist and Huelsenbeck, 2003). Tracer v1.4 (Rambaut and Drummond, 2007) was used to analyze the posterior probabilities from each run and assess convergence between chains. An average of 16% of each BA run was treated as burn-in and discarded (see Supplementary Table 3 for specific burn-in values). All phylogenetic analyses were run on the Brigham Young University Life Science's Computational Cluster (http://lsbeast.byu.edu).

We also studied the effect of data partitioning by comparing topology and nodal support values within each of three phylogenetic inference methods we tested (Table 3). Specifically, we created 8 datasets with varying gene content or coding: all genes (protein-coding, ribosomal RNA and transfer RNA genes) coded as nucleotide data (ALL_123), all genes excluding the third codon position of the protein-coding genes (ALL_12), all genes with protein-coding genes translated and analyzed as amino acid sequences (ALL_PROT), protein-coding genes alone coded as nucleotides (PCG_123), protein-coding genes excluding third codon positions (PCG_12), protein-coding genes translated into amino acid sequences (PCG_PROT), transfer RNA genes alone (TRAN), and ribosomal RNA genes alone (RIBO). In this way, we were able to study the phylogenetic information content of different gene types (ALL vs PCG vs RIBO vs TRAN datasets), the effect of substitution saturation at the third codon position of protein-coding genes (123 vs 12 datasets), and the effects of translated amino acids versus nucleotide sequences (123/12 vs PROT datasets). Additionally, the effect of gap coding was explored in MP analyses by coding gaps as missing vs as a new character state (5th state in nucleotide sequences, 21st state in amino acid sequence). In BA analyses for the ALL_123, ALL_12, PCG_123, and PCG_12 datasets, we were able to partition the dataset further into gene-based partitions (GP, 16 partitions in ALL, 13 in PCG), codon partition (CP, 6 partitions in ALL, 3 in PCG), and gene plus codon partition (GCP, 42 partitions in ALL, 39 in PCG), and examine the effect of different partitioning schemes used in BA analyses (see Table 3).

Table 3
A summary of all the analyses completed for the current study

	Parsimony Missing	Parsimony 5th	Maximum likelihood	Bayesian
All_123	X	X	X	–
All_123_CP	–	–	–	X
All_123_GCP	–	–	–	X
All_123_GP	–	–	–	X
All_12	X	X	X	–
All_12_CP	–	–	–	X
All_12_GCP	–	–	–	X
All_12_GP	–	–	–	X
All_PROT	X	X	n/a	X
PCG_123	X	X	X	–
PCG_123_CP	–	–	–	X
PCG_123_GCP	–	–	-	X
PCG_123_GP	–	–	–	X
PCG_12	X	X	X	–
PCG_12_CP	–	–	–	X
PCG_12_GCP	–	–	–	X
PCG_12_GP	–	–	–	X
PCG_PROT	X	X	n/a	X
TRAN	X	X	X	X
RIBO	X	X	X	X

A dash indicates a dataset that does not apply to that particular criterion. n/a refers to the amino acid datasets where ML is unable to analyze amino acid data.

3. Results

3.1. Mtgenome summary

A total of five new mtgenomes were sequenced for this study: *Acrida willemsei* (Acrididae), *Calliptamus italicus* (Acrididae), *Gryllotalpa pluvalis* (Gryllotalpidae), *Myrmecophila manni* (Myrmecophilidae) and *Troglophilus neglectus* (Raphidophoridae). Complete mtgenome sequences were obtained for *A. willemsei* (15,601 bp), *G. pluvalis* (15,525 bp), and *M. manni* (15,324 bp). We were unable to sequence through the entire control region of *C. italicus* and *T. neglectus*, and the partial genomes are 15,675 bp (including a 783 bp portion of the control region) and 15,810 bp (539 bp of the control region), respectively. Our failure to sequence the control regions for these species is due to the high A+T content and multiple stretches of poly-A and poly-T within the control region which interfere with the accuracy of dye terminator sequencing. However, we estimate that the unsequenced control region of *C. italicus* would be as little as 100 bp based on the comparison with other mtgenomes of Orthoptera. Although we were able to sequence only 539 bp of the *T. neglectus* control region, we identified at least two and a half repeat units of 168 bp within the control region. Based on comparison to other orthopterans, there may be at least 200–1000 bp of unsequenced region, but the total length of the control region would depend on the total number of repeat units. Because the length of the control region was approximately 1000 bp based on the size of PCR products, we estimate that there may be up to 12 repeat units if the entire unsequenced portion of the control region is composed of repeat units.

All species possessed the typical metazoan mtgenome composition of 13 protein-coding genes, 2 ribosomal RNAs, 22 transfer RNAs and a single A+T-rich, control region of variable length (Boore, 1999). Gene order within the mtgenome also followed that of the ancestral insect, except two caeliferan species (*A. willemsei* and *C. italicus*), which had a rearrangement of the ancestral tRNA-^{Lys}(K)– tRNA^{Asp}(D) arrangement located between *cox2* and *atp8*, which have swapped to the derived arrangement tRNA^{Asp}–tRNA^{Lys}

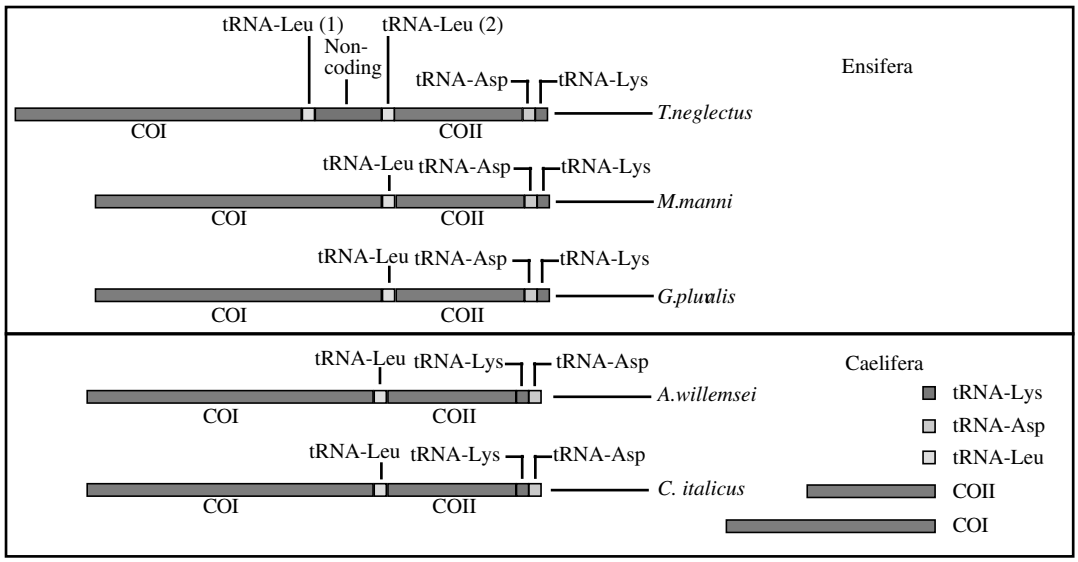

Fig. 1. New Orthopteran Genome Morphology. The five new genomes sequenced for this study showing important gene duplication and rearrangement events. *T. neglectus* shows a tRNA duplication and insertion sequence following *cox1*. The newly sequenced caeliferan species show a rearrangement of tRNA-Lys and tRNA-Asp. See text for explanation.

(Fig. 1). A repeat region was found adjacent to *cox2* in *T. neglectus*, which possessed two copies of $tRNA^{Leu}$ separated by a non-coding region of 357 bp (Fig. 1). We checked this 357 bp region against the GenBank database (Blastn) and found no credible coding region or open reading frame and the sequence has no readily identifiable similarity to other portions of this mtgenome. The additional $tRNA^{Leu}$ had a cove score of 20.1 whereas the original $tRNA^{Leu}$ (located at the 3′ end of *cox1*) cove score was 29.7 as determined by tRNAscan-SE (Supplementary Fig. 6) (Lowe and Eddy, 1997). Although both cove scores are relatively high, the higher cove score for the first gene suggests that this is likely to be the original copy of $tRNA^{Leu}$, and accordingly this copy was used for the phylogenetic analyses. The second $tRNA^{Leu}$ copy differs from the first primarily in the sequence of the DHU stem and loop which is 1 bp shorter and differs at 6 of the 12 remaining sites, the TΨC loop differs at a single base while the anticodon and acceptor stems are identical between the two genes. This suggests that this gene duplication is not evolutionarily recent and is not undergoing concerted evolution unlike most other instances of mitochondrial gene duplication in which sequences are identical between both copies (Caccone et al., 1997; Conrad and Antonarakis, 2007; Tatarenkov and Avise, 2007; Zimmer et al., 1980).

3.2. Phylogenetic analyses

All total evidence analyses (MP_ALL_123, ML_ALL_123, BA_ALL_123_GP) resulted in an identical ingroup topology with high support values within Orthoptera and recovered monophyletic suborders based on the ingroup selection we used in this study (Fig. 2). Within Ensifera, we found Raphidophoroidea (Rhaphidophoridae) to be sister to Tettigonoidea (Tettigoniidae) and that this clade was sister to Grylloidea (Gryllotalpidae + Myrmecophilidae). Within Caelifera, three subfamilies of Acrididae were included and we found a sister relationship between Acridinae (*Acrida willemsei*) and Calliptaminae (*Calliptamus italicus*) to the exclusion of Oedipodinae (*Locusta migratoria*). In terms of the interordinal relationships, we found monophyletic Dictyoptera (Mantodea + (Blattodea + Isoptera)), which was consistently the sister group of Orthoptera. The relationships among the remaining outgroups were not well resolved. Among the total evidence analyses, there was no apparent effect of different optimality criteria in recovering ingroup relationships and the relationships within Dictyoptera (Fig. 1). However, different optimality criteria did produce different relationships among the other outgroups. In both MP and BA analyses, *Timema* was more closely related to *Grylloblatta* than *Sclerophasma*, whereas it was more closely related to *Sclerophasma* in ML analyses. Nodal support values also appeared to be affected by the optimality criteria in that the bootstrap values in MP analyses were generally lower than the bootstrap values in ML and posterior probabilities in BA analyses. For example, the nodal support values for the backbone nodes were relative low in MP (less than 50% for Dictyoptera + Orthoptera, 73% for Dictyoptera and 52% for Orthoptera), but high in ML (94%, 100% and 97%) and BA (1.00, 1.00 and 1.00). However, there was an exception to this general pattern, found in the clade consisting of *A. willemsei* and *C. italicus*, which had a ML bootstrap value of 76%, considerably lower than MP bootstrap value (89%) and BA posterior probability (0.97).

3.3. Effect of data manipulation

Inclusion and exclusion of data had a greater influence in analyzing mtgenome data both in topology and nodal support. We compared the effect of partitioning according to gene type (PCG, TRAN, RIBO) against the total evidence dataset (ALL) and protein-coding genes only (PCG). The analyses based on PCG all produced an identical ingroup topology to the ALL analyses except for relationships within Caelifera in BA analysis (see Supplementary Fig. 7). Nodal support was very high across all nodes, with the exception of the clade consisting of *A. willemsei* and *C. italicus*. When transfer RNAs were analyzed as a single partition (TRAN), a monophyletic Orthoptera was only recovered in one of the four analyses (MP_TRAN_mis), but a monophyletic Caelifera was recovered in all analyses. The analyses based on ribosomal RNAs (RIBO) also performed poorly compared to the combined dataset, resulting in unique and incongruent topologies from each of the different inference methods. Only the ML_RIBO analysis produced a monophyletic Orthoptera and was completely consistent with the ML_ALL topology with the sole exception of *Sclerophasma* moving to be sister to Dictyoptera. The MP_RIBO analysis did not recover a topology congruent with other studies. This dataset scattered members of Dictyoptera throughout the tree and failed to resolve

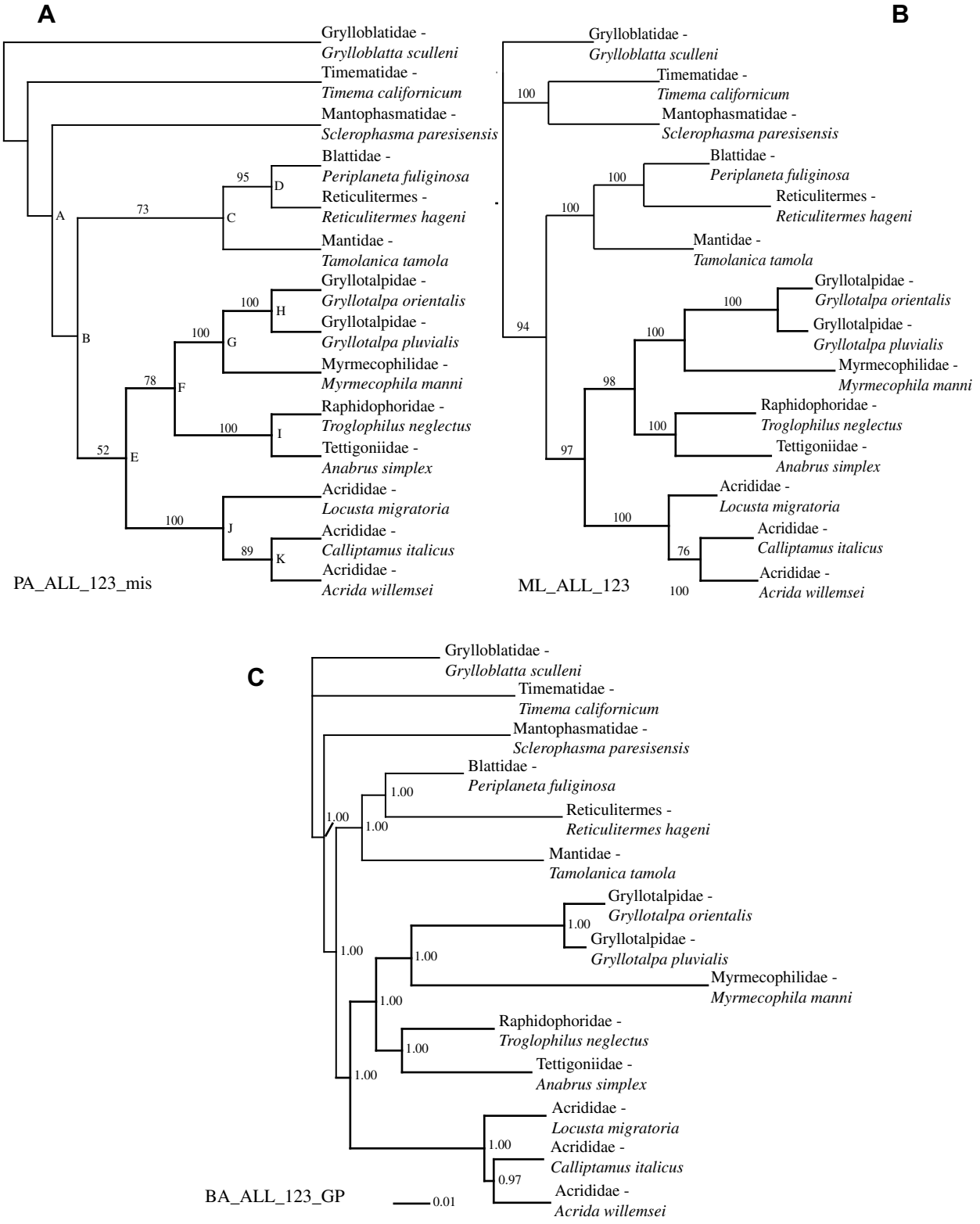

Fig. 2. Phylogenetic reconstruction of Orthoptera using mtgenomes. (A) Parsimony results based on the ALL_123_mis dataset. Applicable bootstrap values are shown. B. Maximum likelihood results based on the ALL_123 dataset. Applicable bootstrap values are shown. (C) Bayesian results based on the ALL_123_GP dataset. Applicable posterior probability values are shown.

most of the clades, resulting in paraphyletic Orthoptera. Nodal support was also very low across all nodes for each of the RIBO analyses.

Exclusion of third codon position from protein-coding genes (ALL_12 and PCG_12) did not appear to have any effect in phylogenetic reconstruction as we recovered the identical ingroup topology as the combined datasets (ALL_123 and PCG_123). This finding indicates that in our study the inclusion of third codon positions, which may be mutationally saturated due to lower selective pressures, do not negatively affect phylogenetic reconstruc-

Table 4
Bremer support values for the MP_ALL_123_mis tree

		atp6[a]	atp8	cox1	cox2	cox3	cytb	nd1	nd2	nd3	nd4	nd4l	nd5	nd6	16s	12s	trans	Total Bremer
A	Mantophasmatodea	0 (0)a	−5	4	4	−1	13	−1	10	3	−5	−5	−10	2	9	1	−8	11
B	Dictyoptera + Orthoptera	−6 (−37.5)	8	21	21	4	−2	8	−5	−5	5	−13	1	−2	−7	−11	−1	16
C	Dictyoptera	−13 (−37.1)	7	22	14	7	−4	4	5	−9	−1	−15	3	−6	11	1	9	35
D	Blattodea + Isoptera	−13 (−20.3)	−7	−6	1	4	30	1	29	2	−7	−4	−6	−1	28	4	9	64
E	Orthoptera	−6 (−37.5)	8	21	21	4	−2	8	−5	−5	5	−13	1	−2	−7	−11	−1	16
J	Caelifera (Acrididae)	19 (3.24)	−2	79	70	29	48	31	56	4	63	13	61	17	40	25	33	586
K	Calliptaminae + Acridinae	−4 (−12.9)	−4	−2	4	7	27	−1	9	0	−4	−4	−17	6	11	4	−1	31
F	Ensifera	−4 (−12.9)	−5	37	31	9	17	8	−7	−4	−1	−11	−11	−9	7	−4	−22	31
I	Rhaphidophoridae + Tettigoniidae	−4 (−4.44)	−1	18	−5	7	11	12	−3	1	−3	4	16	1	13	10	13	90
G	Myrmecophilidae + Gryllotalpidae	0 (0.0)	11	14	8	6	−2	8	25	9	20	4	5	5	8	10	3	134
H	Gryllotalpidae	33 (3.971)	15	43	46	45	57	33	60	19	93	18	125	29	91	52	72	831
	Overall gene contribution (%)[b]	−14.14	3.92	46.76	42.84	12.85	26.94	14.1	11.51	−5.45	1.88	−26.69	−11.14	−2.18	15.07	−8.2	−8.06	=100%

Partitioned Bremers were calculated for each node by 16 gene combinations.
[a] Raw Bremer support values. Percentage support values are shown in parenthesis only for *atp6* as en example to show how total gene contribution percent is calculated.
[b] This value is calculated as a percentage support for the entire gene. For example, under *atp6* the raw nodal support is calculated as a percent of the total support for that node (i.e. −6/16 = −0.375). The entire gene support (the entire column) is then averaged to give the average overall gene contribution percent support.

tion. In contrast, reducing the dataset into the amino acid sequence dataset (ALL_PROT and PCG_PROT) had a considerable effect in phylogenetic reconstruction. In both datasets, MP and BA analyses recovered monophyletic Orthoptera, but a paraphyletic Ensifera, with relative high nodal supports (see supplementary Fig. 8).

Different partitioning strategies of mtgenome data in the Bayesian framework did not appear to affect branch length or support values much, but they did slightly affect topology among the groups which varied between analyses (Caelifera and the outgroups). However, when the mtgenome data were partitioned into both genes and codons (ALL_123_GCP and PCG_123_GCP), the Bayesian analyses failed to converge after twenty million generations using 4 independent runs with 4 chains per run. We reran the analyses with a more thorough search using 25 million generations with 6 independent runs and 10 chains per run, but again failed to converge. This may be the result of over-partitioning such that individual partitions are too small to accurately estimate parameter values and so fail to converge on a result. BA_ALL_123_GCP and BA_PCG_123_GCP topologies presented are thus taken from a single run (Supplementary Table 3).

Different methods of gap coding had varying degrees of influence in phylogenetic reconstruction, especially in relation to the size of dataset. In large datasets (ALL_123, PCG_123, ALL_12, PCG_12), coding gaps as 5th state resulted in the identical overall topology (identical ingroup topology in ALL_12) as when gaps were coded as missing data. However, in small datasets (RIBO, TRAN), we observed a considerable effect of gap coding, resulting in different ingroup relationships, with only well defined groups (such as monophyletic *Gryllotalpa*) remaining consistent between the two. When amino acid sequences were analyzed (PCG_PROT), coding gaps as a 21st state failed to recover ensiferan monophyly.

Partitioned Bremer support values calculated based on the MP_ALL_123_mis dataset (Table 4) indicated that different genes contribute differently to the overall phylogenetic signals of the mtgenome data. Of all the genes, *cox1, cox2* and *cytB* contributed the highest overall signals. Among the protein-coding genes, *atp6*, *nd3*, *nd4l*, *nd5* and *nd6* contributed negatively across the majority of nodes. The two ribosomal genes drastically differed with *16S* providing nearly 15% of the total support whereas *12S* conflicted with the overall topology (negative support of −8%). Collectively, the transfer RNAs are less negative than the *12S* gene but still conflict with the overall topology. In general, total Bremer support for the deeper nodes of the tree (nodes A and B in Fig. 2A) are the least supported (A = 11 and B = 16) corresponding to variation between which interordinal relationships are supported by individual genes. In particular *12S* and the transfer RNAs conflict strongly with these nodes; although, many protein-coding genes provide high support (i.e., *cox1* and *cox2* at node B = 21 and 21, respectively). Surprisingly, node E (Orthoptera) has a support of 16; not nearly as well supported as comparable nodes such as Dictyoptera (node C = 35). Ensifera (node F) is strongly supported by the *cox* genes, but poorly supported by the other protein-coding and RNA genes. Node J (Caelifera/Acrididae) and node H (*Gryllotalpa*) are the highest supported nodes of the tree (J = 586 and H = 831), demonstrating the extremely strong phylogenetic signal within mtgenomes for resolving shallow taxonomic groups, such as families or genera (Table 4). Partitioned Bremer support values for the remaining analyses are presented in Supplementary Table 4.

4. Discussion

4.1. Mitochondrial genomics and gene rearrangement

All of the newly sequenced mtgenomes included in the present study are similar in gene and nucleotide composition when compared to the six available orthopteran mtgenomes as well as to the presumed ancestral hexapod (Boore, 1999; Fenn et al., 2007; Flook et al., 1995b; Kim et al., 2005). However, we found a gene rearrangement within the suborder Caelifera and varying gene content in Rhapidophoridae. Gene rearrangements within the mtgenome have long been considered potentially useful phylogenetic markers and have been shown to be useful in resolving deep level relationships within Arthropoda (Boore et al., 1998). As additional mtgenomes have been sequenced, however, the phylogenetic utility of gene rearrangement in understanding insect evolution now appears less than previously thought because most insect mtgenomes have retained the ancestral gene order and there is not enough phylogenetic signal to resolve interordinal or deeper relationships from the existing gene rearrangements. Exceptions do exist, however, with rearrangements being found in insect orders such as Hymenoptera (Castro et al., 2006; Crozier and Crozier, 1993), Thysanoptera (Shao and Barker, 2003), Phthiraptera (Cameron and Whiting, 2007; Covacin et al., 2006; Shao et al., 2001), and Orthoptera (Flook et al., 1995b). Our study expands previous knowledge of gene rearrangements in Orthoptera and the other polyneopteran orders. We find that all ensiferans including the previously unstudied families, Myrmecophilidae and Rhaphidophoridae, and all the members of the polyneopteran orders included in this study retain the ancestral KD arrangement. Our sampling also includes two new acridid subfamilies that Flook et al. (1995b) did not study, Calliptaminae and Acridinae, which possess the derived gene rearrangement. Incidently, the newly published tettigoniid mtgenomes (Zhou et al., 2007a,b) and an

acridid mtgenome (Zhang and Huang, 2007) have original and derived gene arrangements, respectively, congruent with the present study. Because our taxon sampling of Caelifera is small, only including members of Acrididae, it is difficult to determine whether this gene rearrangement is truly a synapomorphy for Caelifera or not. Flook et al. (1995a) did include one non-acridid caeliferan, a pyrgomorphid, and found that it too has the rearrangement. This suggests that the rearrangement is at least as old as the common ancestor of Acrididae and Pyrgomorphidae. Our study demonstrates that "genome morphology" is a useful source of phylogenetic information at least within Orthoptera. This is a useful marker in that it shows conservation across a large group (possibly an entire suborder) but it is also limited to a single inversion with no detectable addition or deletion of genetic information such as that due to subsequent rearrangement of either gene to other regions within the mtgenome. Additionally, we have identified a second derived rearrangement, the duplicated $tRNA^{Leu}$ in *Troglophilus*, which may be a potential molecular synapomorphy defining some subgroup within Raphidophoroidea. The phylogenetic range over which this rearrangement occurs has yet to be determined.

4.2. Mtgenomes and deep-level phylogenies

The present study represents the most character-heavy analysis of the orthopteran phylogeny to date. The most comprehensive phylogeny of Orthoptera thus far was based on 31 ingroup taxa and three ribosomal loci (Flook et al., 1999). Our dataset is nearly five times larger in terms of character sampling although smaller in taxon sampling. Admittedly, our taxon sampling is too small to make definitive conclusions about all aspects of the evolution of Orthoptera, but the goal of this study is to explore the phylogenetic utility of mtgenome data in resolving deep relationships. Therefore, the phylogenetic inferences set forth in this study are informative only when based on the eight ingroup species specifically tested. We consider this phylogeny to be preliminary until further taxon sampling for mtgenomes exceeds that of studies which utilized more limited character sets (e.g. Flook et al., 1999).

The total evidence datasets, regardless of the optimality criteria used, all recovered monophyletic Orthoptera with two monophyletic suborders, Caelifera and Ensifera under the sampling used. This finding corroborates with the generally accepted classification schemes as well as the study by Flook et al. (1999). Within Ensifera, we found Raphidophoroidea (Rhaphidophoridae) to be sister to Tettigonoidea (Tettigoniidae) and this clade was sister to Grylloidea (Gryllotalpidae + Myrmecophilidae). The relationships within Ensifera are consistent with the findings by Gwynne (1995), Flook et al. (1999), and Desutter-Grandcolas (2003), but in conflict with a recent molecular study by Jost and Shaw (2006) who found that Grylloidea was more closely related to Rhaphidophoroidea than to Tettigonoidea. Because both taxon and character sampling do not overlap much between our study and Jost and Shaw (2006), it is difficult to assess which phylogenetic scheme is more realistic, but our analysis is more in line with the traditional classification of Ensifera. Within Caelifera, three subfamilies of Acrididae were included and the majority of analyses supported a sister relationship between Acridinae (*Acrida willemsei*) and Calliptaminae (*Calliptamus italicus*) to the exclusion of Oedipodinae (*Locusta migratoria*). This relative placement of Oedipodinae within Acrididae has been proposed previously by both molecular studies (Flook and Rowell, 1997, 1998) and a morphological study based on the male phallic complex (Eades, 2000). Although our taxon sampling was small and did not include representatives from all the major orthopteran lineages, our analyses suggests that mtgenome data can resolve relationships dating from the deep Permian (~260 MYA) such as the split between two suborders down to the Tertiary (~50 MYA) such as recent divergence of the three subfamilies of Acrididae (Sharov, 1968). The mtgenome data appear to be very effective in resolving relationships within Orthoptera and our study adds to a growing list of studies showing the phylogenetic versatility of mtgenome data in resolving relationships over very wide time scales.

Our analysis found a relatively robust sister relationship between Orthoptera and Dictyoptera. This is a novel relationship and at odds with many previous studies that found a close relationship between Orthoptera and Phasmatodea (Sharov, 1968; Kamp, 1973; Boudreaux, 1979; Kukalová-Peck, 1991; Flook and Rowell, 1998; Wheeler et al., 2001; Terry and Whiting, 2005). At this point, we are unable to determine whether this is an artifact of small taxon sampling or reflects the information contained within mtgenomes. Previous studies based on mtgenome data found equally incongruent outgroup relationships with respect to Orthoptera. Kim et al. (2005) found Orthoptera to be paraphyletic, with *Locusta migratoria* being sister to a thysanuran *Tricholepidion gertschi*. Cameron et al. (2006) found Orthoptera to be sister to a clade formed by Phasmatodea, Mantophasmatodea, Grylloblattodea, and Dictyoptera. It is possible that the differences in relationships among different studies including ours may be due to the sensitivity of interordinal polyneopteran relationships to both taxon and character sampling. Complete taxon sampling of rich character sets will eventually lead to a more accurate understanding of the phylogenetic relationships within Polyneoptera and the sequencing of complete mtgenomes from additional representatives will aid in this effort.

4.3. Methodological effects of various approaches to phylogenetic reconstruction with mtgenomes

The differences, advantages and disadvantages between the three major phylogenetic optimality criteria (MP, ML and BA) are an ongoing debate within the world of systematics. In this study, we assessed how each criterion would perform under a variety of different datasets derived from the mtgenomes. When all available data were analyzed simultaneously (ALL_123), there was no apparent effect on topology between the parsimony and model-based analyses. Even when a simple model of molecular evolution was uniformly applied to a dataset composed of multiple genes with differing evolutionary rates (ML), topology was the same as the highly precise and partitioned evolutionary models (BA) and high nodal supports were still recovered. These findings indicate that the signals within mtgenomes are so strong that even inadequate models of evolution do not adversely affect the phylogenetic reconstruction. However, it is also possible that the strong congruence among different optimality criteria is a function of small taxon sampling and large character sampling. In fact, when the smaller subsets of data (RIBO or TRAN) or the dataset reduced to amino acid sequences (PROT) are analyzed under different optimality criteria, the effect is more evident in that different topologies were recovered. It will be of great interest to see whether the increase in taxon sampling introduces incongruence between the inference methods.

Because mtgenomes consist of multiple genes of different types, there can be many ways of partitioning and analyzing this marker in a phylogenetic framework. The effect of data partitioning can be highlighted in the caeliferan clade, in which different relationships among three acridid taxa were recovered according to different partitioning schemes (Supplementary Fig. 7). The majority of analyses found the relationships within this clade to be (*Locusta* + (*Calliptamus* + *Acrida*)). However, four analyses (BA_ALL_12_GCP; BA_PCG_12_GCP; BA_TRAN; ML_TRAN) rearranged the relationship to (*Calliptamus* + (*Acrida* + *Locusta*)), and one analysis (BA_PCG_123_GP) found it to be (*Acrida* + (*Locusta* + *Calliptamus*)). Across all analyses, the nodal support for the caeliferan clade (Sup-

plementary Fig. 7A) is maximal, but the terminal relationship has varying degrees of support (Supplementary Fig. 7B). For example, the support values for the clade (*Acrida* + *Calliptamus*) differ considerably across different analyses, never reaching 100% in either MP or ML analyses. At this level, inclusion of third codon position does seem to negatively affect bootstrap values (Supplementary Fig. 7). In a Bayesian framework, further partitioning is possible, but our study suggests that it introduces an unexpected problem during the analyses. Although the topology of BA datasets ALL_123_GP, ALL_123_CP and ALL_123_GCP are all the same, only the first two (GP and CP) resulted in multiple runs converging on the same topology. The ALL_123_GCP dataset with 42 partitions and the PCG_123_GCP dataset with 39 partitions were unable to statistically converge to a single most likelihood score even with 25 million generations. Because the other CP and GP treatments of these datasets converged early on, both within the first million generations, we argue that over-partitioning of the mtgenome may be more detrimental to the analysis than beneficial (Cameron et al., 2007; Nylander et al., 2004). It is possible that allowing more generations and more chains during the tree searching will increase the chance of convergence, but current computational limits seem to preclude accurately using such a highly partitioned scheme for mtgenome data (Hutter, 2007; Mossel and Vigoda, 2006; Toft et al., 2007).

Our study supports the idea that the best way to analyze mtgenome data is a total evidence approach based on nucleotide sequences. We found that reduced coding, such as amino acid recoding of the nucleotide sequences, negatively affect the phylogenetic reconstruction, resulting in paraphyletic Ensifera. Cameron et al. (2006) found that converting nucleotide sequences to amino acid sequences or purine/pyrimidine coding reduced phylogenetic signal and could introduce artificial relationships. Our study concurs with Cameron et al. (2006) that reduced coding introduces more ambiguity than resolution. We promote the inclusion of all data. Frequently, third codon positions and ribosomal genes are excluded *a priori* from many analyses (Brinkmann et al., 2004; Pashley and Ke, 1992). We show that the inclusion of third codon positions does not negatively affect phylogenetic reconstruction when all available data are analyzed together. While we find that ribosomal RNAs and transfer RNAs failed to recover monophyletic Orthoptera or suborders within, they do contribute to the overall signal when analyzed with other available data. Especially, we find that the TRAN dataset had a comparable resolving capacity to the RIBO dataset, despite its smaller size of being approximately 50% of the RIBO dataset. Transfer RNAs are frequently overlooked and excluded from phylogenetic analyses of mtgenomes whereas ribosomal RNAs are often included. We argue that the tRNAs contain as much, if not more, phylogenetic signals as the ribosomal RNAs, and should always be included when mtgenome data are analyzed (Table 4). Partitioned Bremer support values suggest that nearly all the genes of a genome contribute at least some resolving power at various levels in the tree, and that arbitrary gene exclusion may have deleterious affects on portions of the topology or support at individual nodes. The best phylogenetic inference results when all available data are included and analyzed together. Full mtgenomes, partitioned and analyzed appropriately are likely to become commonplace in the world of systematics and be useful over a wide range of evolutionary time scales.

Acknowledgments

We thank Matt Terry, Gavin Svenson, Seth Bybee, April Greenwell and Kelly Miller for their aid in collection and preparation of samples and in manuscript preparation. We also appreciate the valuable comments by anonymous reviewers on the earlier version of this manuscript. This study was supported by NSF Grants DEB0120718 (to M.F.W.) and DEB0444972 (to S.L.C. and M.F.W.).

References

Akanuma, J., Muraki, K., Komaki, H., Nonaka, I., Goto, Y., 2000. Two pathogenic point mutations exist in the authentic mitochondrial genome, not in the nuclear pseudogene. Journal of Human Genetics 45, 337–341.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389–3402.

Baker, R.H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. Systematic Biology 46, 654–673.

Bensasson, D., Zhang, D.X., Hartl, D.L., Hewitt, G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology & Evolution 16, 314–321.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Research 27, 1767–1780.

Boore, J.L., Brown, W.M., 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Current Opinion in Genetics & Development 8, 668–674.

Boore, J.L., Lavrov, D.V., Brown, W.M., 1998. Gene translocation links insects and crustaceans. Nature 392, 667–668.

Boudreaux, H.B., 1979. Arthropod Phylogeny. John Wiley & Sons, New York.

Brinkmann, H., Denk, A., Zitzler, J., Joss, J.J., Meyer, A., 2004. Complete mitochondrial genome sequences of the South American and the Australian lungfish: testing of the phylogenetic performance of mitochondrial data sets for phylogenetic problems in tetrapod relationships. Journal of Molecular Evolution 59, 834–848.

Caccone, A., Milinkovitch, M.C., Sbordoni, V., Powell, J.R., 1997. Mitochondrial DNA rates and biogeography in European newts (Genus *euproctus*). Systematic Biology 46, 126–144.

Cameron, S.L., Barker, S.C., Whiting, M.F., 2006. Mitochondrial genomics and the new insect order Mantophasmatodea. Molecular Phylogenetics and Evolution 38, 274–279.

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. Systematic Entomology 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 (Arthropoda). Cladistics 20, 534–557.

Cameron, S.L., Whiting, M.F., 2007. Mitochondrial genomic comparisons of the subterranean termites from the Genus Reticulitermes (Insecta: Isoptera: Rhinotermitidae). Genome 50, 188–202.

Carapelli, A., Lio, P., Nardi, F., van der Wath, E., Frati, F., 2007. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evolutionary Biology 7, S8.

Castro, L.R., Dowton, M., 2005. The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of *Perga condei* (Hymenoptera: Symphyta: Pergidae). Molecular Phylogenetics and Evolution 34, 469–479.

Castro, L.R., Dowton, M., 2007. Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers. Systematic Entomology 32, 60–69.

Castro, L.R., Ruberu, K., Dowton, M., 2006. Mitochondrial genomes of *Vanhornia eucnemidarum* (Apocrita: Vanhorniidae) and *Primeuchroeus* spp. (Aculeata: Chrysididae): evidence of rearranged mitochondrial genomes within the Apocrita (Insecta: Hymenoptera). Genome 49, 752–766.

Conrad, B., Antonarakis, S.E., 2007. Gene duplication: A drive for phenotypic diversity and cause of human disease. Annual Review of Genomics and Human Genetics 8, 17–35.

Covacin, C., Shao, R., Cameron, S., Barker, S.C., 2006. Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). Insect Molecular Biology 15, 63–68.

Crozier, R.H., Crozier, Y.C., 1993. The mitochondrial genome of the Honeybee *Apis–Mellifera*—complete sequence and genome organization. Genetics 133, 97–117.

Desutter-Grandcolas, L., 2003. Phylogeny and the evolution of acoustic communication in extant Ensifera (Insecta, Orthoptera). Zoologica Scripta 32, 525–561.

Eades, D.C., 2000. Evolutionary relationships of phallic structures of Acridomorpha (Orthoptera). Journal of Orthopteran Research 9, 181–210.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 1–19.

Fenn, J.D., Cameron, S.L., Whiting, M.F., 2007. The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Molecular Biology 16, 239–252.

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

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. Systematic Biology 48, 233–253.

Flook, P.K., Rowell, C.H.F., 1997. The phylogeny of the Caelifera (insecta, orthoptera) as deduced from mtrRNA gene sequences. Molecular Phylogenetics and Evolution 8, 89–103.

Flook, P.K., Rowell, C.H.F., 1998. Inferences about orthopteroid phylogeny and molecular evolution from small subunit nuclear ribosomal DNA sequences. Insect Molecular Biology 7, 163–178.

Flook, P.K., Rowell, C.H.F., Gellissen, G., 1995b. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. Journal of Molecular Evolution 41, 928–941.

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2003. T.N.T.: tree analysis using new technology. Program and documentation, available from the authors, and at: <www.zmuc.dk/public/phylogeny>.

Grimaldi, D., Engel, M.S., 2005. Evolution of the Insects. Cambridge University Press, New York.

Gwynne, D.T., 1995. Phylogeny of the Ensifera (Orthoptera): a hypothesis supporting multiple origins of acoustical signalling, complex spermatophores and maternal care in crickets, katydids, and weta. Journal of Orthoptera Research 4, 203–218.

Hutter, M., 2007. On universal prediction and Bayesian confirmation. Theoretical Computer Science 384, 33–48.

Jost, M.C., Shaw, K.L., 2006. Phylogeny of Ensifera (Hexapoda: Orthoptera) using three ribosomal loci, with implications for the evolution of acoustic communication. Molecular Phylogenetics and Evolution 38, 510–530.

Kamp, J.W., 1973. Numerical classification of the orthopteroids, with special reference to the Grylloblattodea. Canadian Entomologist 105, 1235–1249.

Kevan, D.K.M., 1982. Orthoptera. In: Parker, S.P. (Ed.), Orthoptera. McGraw-Hill Book Company, New York, pp. 352–383.

Kim, I., Cha, S.Y., Yoon, M.H., Hwang, J.S., Lee, S.M., Sohn, H.D., Jin, B.R., 2005. The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). Gene 353, 155–168.

Kjer, K.M., Honeycutt, R.L., 2007. Site specific rates of mitochondrial genome and phylogeny of eutheria. BMC Evolutionary Biology 7, 8.

Kukalová-Peck, J., 1991. Fossil history and the evolution of hexapod structures. The Insects of Australia, A Textbook for Students and Research Workers, 2nd ed., Melbourne University Press, Paris, pp. 141–179.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5, 150–163.

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

Maddison, W.P., Maddison, D.R., 2003. MacClade ver 4.06. Sinauer Associates, Sunderland.

Mossel, E., Vigoda, E., 2006. Limitations of Markov chain Monte Carlo algorithms for Bayesian inference of phylogeny. Annals of Applied Probability 16, 2215–2234.

Nardi, F., Carapelli, A., Fanciulli, P.P., Dallai, R., Frati, F., 2001. The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. Molecular Biology and Evolution 18, 1293–1304.

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. Systematic Biology 53, 47–67.

Pashley, D.P., Ke, L.D., 1992. Sequence evolution in mitochondrial ribosomal and Nd-1 genes in Lepidoptera—implications for phylogenetic analyses. Molecular Biology and Evolution 9, 1061–1075.

Peña, C., Wahlberg, N., Weingartner, E., Kodandaramaiah, U., Nylin, S., Freitas, A.V.L., Brower, A.V.Z., 2006. Higher level phylogeny of Satyrinae butterflies (Lepidoptera: Nymphalidae) based on DNA sequence data. Molecular Phylogenetics and Evolution 40, 29–49.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.4. Available from: <http://beast.bio.ed.ac.uk/Tracer>.

Rentz, D.C.F., 1991. Orthoptera (Grasshoppers, locusts, katydids, crickets). In: Naumann, I.D. (Ed.), The Insects of Australia: A Textbook for Students and Research Workers. Melbourne University Press, Carlton, Vic., pp. 369–393.

Rokas, A., Holland, P.W., 2000. Rare genomic changes as a tool for phylogenetics. Trends in Ecology & Evolution 15, 454–459.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Saccone, C., De Giorgi, C., Gissi, C., Pesole, G., Reyes, A., 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene 238, 195–209.

Saitoh, K., Sado, T., Mayden, R.L., Hanzawa, N., Nakamura, K., Nishida, M., Miya, M., 2006. Mitogenomic evolution and interrelationships of the *Cypriniformes* (Actinopterygii: Ostariophysi): the first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. Journal of Molecular Evolution 63, 826–841.

Schneider, A., Ebert, D., 2004. Covariation of mitochondrial genome size with gene lengths: evidence for gene length reduction during mitochondrial evolution. Journal of Molecular Evolution 59, 90–96.

Shao, R., Campbell, N.J.H., Barker, S.C., 2001. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). Molecular Biology and Evolution 18, 858–865.

Shao, R.F., 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. Molecular Biology and Evolution 20, 362–370.

Sharov, A.G., 1968. Phylogeny of the Orthopteroidea. Akademiya Nauk SSSR Trudy Paleontologicheskogo Instituta 118, 1–216.

Spooner, L.J., Ritchie, M.G., 2006. An unusual phylogeography in the bushcricket Ephippiger ephippiger from Southern France. Heredity 97, 398–408.

Stewart, J.B., Beckenbach, A.T., 2003. Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle *Crioceris duodecimpunctata*. Molecular Phylogenetics and Evolution 26, 513–526.

Stewart, J.B., Beckenbach, A.T., 2006. Insect mitochondrial genomics 2: the complete mitochondrial genome sequence of a giant stonefly, *Pteronarcys princeps*, asymmetric directional mutation bias, and conserved plecopteran A+T-region elements. Genome 49, 815–824.

Stothard, P., Wishart, D.S., 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21, 537–539.

Swofford, D.L., 2002. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Tatarenkov, A., Avise, J.C., 2007. Rapid concerted evolution in animal mitochondrial DNA. Proceedings of the Royal Society B—Biological Sciences 274, 1795–1798.

Terry, M.D., Whiting, M.F., 2005. Mantophasmatodea and phylogeny of the lower neopterous insects. Cladistics 21, 240–257.

Toft, N., Innocent, G.T., Gettinby, G., Reid, S.W.J., 2007. Assessing the convergence of Markov Chain Monte Carlo methods: an example from evaluation of diagnostic tests in absence of a gold standard. Preventive Veterinary Medicine 79, 244–256.

Wheeler, W.C., Whiting, M.F., Wheeler, Q.D., Carpenter, J.M., 2001. The phylogeny of the extant hexapod orders. Cladistics 17, 113–169.

Yamauchi, M.M., Miya, M.U., Nishida, M., 2004. Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. Insect Molecular Biology 13, 435–442.

Zhang, C., Huang, Y., 2007. Complete mitochondrial genome of Oxya chinensis (Orthoptera: Acridoidea). Acta Biochimica et Biophysica Sinica 40, 7–18.

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

Zhou, Z.J., Huang, Y., Shi, F.M., Ye, H.Y., 2007b. The complete mitochondrial genome of Deracantha onos (Orthoptera: Bradyporidae). Molecular Biology Reports, in press. Available from: <http://www.springerlink.com/content/d1171050u6176223/?p=c7e7c9fbf20346e9bf17af2cd932b84f&pi=0>.

Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W., Wilson, A.C., 1980. Rapid duplication and loss of genes-coding for the alpha-chains of hemoglobin. Proceedings of the National Academy of Sciences of the United States of America—Biological Sciences 77, 2158–2162.