
A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders

STEPHEN L. CAMERON, JARON SULLIVAN, HOJUN SONG, KELLY B. MILLER & MICHAEL F. WHITING

Submitted: 7 January 2009

Accepted: 4 March 2009

doi:10.1111/j.1463-6409.2009.00392.x

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. — *Zoologica Scripta*, **, ***-***.

We present a mitochondrial (mt) genome phylogeny inferring relationships within Neuropterida (lacewings, alderflies and camel flies) and between Neuropterida and other holometabolous insect orders. Whole mt genomes were sequenced for *Sialis hamata* (Megaloptera: Sialidae), *Ditaxis latistyla* (Neuroptera: Mantispidae), *Mongoloraphidia barmandi* (Raphidioptera: Raphidiidae), *Macrogyrus oblongus* (Coleoptera: Gyrinidae), *Rhopaea magnicornis* (Coleoptera: Scarabaeidae), and *Mordella atrata* (Coleoptera: Mordellidae) and compared against representatives of other holometabolous orders in phylogenetic analyses. Additionally, we test the sensitivity of phylogenetic inferences to four analytical approaches: inclusion vs. exclusion of RNA genes, manual vs. algorithmic alignments, arbitrary vs. algorithmic approaches to excluding variable gene regions and how each approach interacts with phylogenetic inference methods (parsimony vs. Bayesian inference). Of these factors, phylogenetic inference method had the most influence on interordinal relationships. Bayesian analyses inferred topologies largely congruent with morphologically-based hypotheses of neuropterid relationships, a monophyletic Neuropterida whose sister group is Coleoptera. In contrast, parsimony analyses failed to support a monophyletic Neuropterida as Raphidioptera was the sister group of the entire Holometabola excluding Hymenoptera, and Neuroptera + Megaloptera is the sister group of Diptera, a relationship which has not previously been proposed based on either molecular or morphological data sets. These differences between analytical methods are due to the high among site rate heterogeneity found in insect mt genomes which is properly modelled by Bayesian methods but results in artifactual relationships under parsimony. Properly analysed, the mt genomic data set presented here is among the first molecular data to support traditional, morphology-based interpretations of relationships between the three neuropterid orders and their grouping with Coleoptera.

Corresponding author: *Stephen L. Cameron, Australian National Insect Collection, CSIRO Entomology, Canberra, ACT 2601, Australia. E-mail: stephen.cameron@csiro.au*

Jaron Sullivan, Hojun Song and Michael F. Whiting, Department of Biology, Brigham Young University, Provo, UT 84602, USA

Kelly B. Miller, Department of Biology, and Museum of Southwestern Biology, The University of New Mexico, Albuquerque, NM 87131, USA

Introduction

Comprising over 80% of insect species and more than 50% of animal life, the Holometabola (= Endopterygota) are the most speciose group of organisms ever to have arisen in the history of life (Wilson 1988). Much of this success has been attributed to the development of complete metamorphosis within their life cycles, allowing the partitioning of environmental

resources and life history roles such as feeding and dispersal between morphological specialized immature and adult stages (Norris 1991). Eleven orders are traditionally recognized within Holometabola, and there is strong evidence from both morphological and molecular data sets for the monophyly of nine of these (Whiting 2002a). In contrast to the strong consensus on ordinal monophyly, agreement on the interordinal

relationships within the Holometabola has been far more elusive. The holometabolan orders have been combined into a wide variety of supra-ordinal groupings by different workers (see review by Kristensen 1999), however, two of the most enduring groupings are the superorder Neuropterida and an as yet unnamed grouping of Coleoptera (beetles) + Neuropterida (lacewings and relatives). The neuropterids include three orders, the snake- or camelneck-flies (Raphidioptera) with less than 200 species, the alder- or dobsonflies (Megaloptera) with about 300 described species, and the lacewings (Neuroptera) with over 6500 described species and an estimated global diversity of 10 000 species (Aspöck 2002). The Neuropterida are thus the largest of the holometabolan groups outside of the mega-diverse insect orders (i.e. flies, beetles, wasps, moths).

Neuropteridan monophyly is supported morphologically by the synapomorphic fusion of the third ovipositor valvulae (Achtelig 1975); historically, the presence of a gula has also been considered synapomorphic for the group, although this has been challenged by Hennig (1981). In contrast there is less support from molecular data for neuropterid monophyly as few studies include representatives from all three orders. The only studies with comprehensive taxon sampling rely on the nuclear small subunit (= 18S) rRNA gene and all support neuropterid monophyly (Whiting 2002a; Kjer *et al.* 2006; Misof *et al.* 2007). The Neuropterida are typically considered to be the sister group of the Coleoptera although evidence in support of this relationship is almost entirely lacking. As Kristensen (1991) put it in his review of hexapod phylogeny 'Although the monophyly of this order-group is not supported by conspicuous autapomorphies it has been in vogue for many years'; this judgement, however, did not prevent Kristensen (1991) from depicting Coleoptera and Neuropterida as sister groups in his accompanying tree (fig. 5.5 of that work) or in subsequent reviews of holometabolan relationships (e.g. Figure 9, Kristensen 1999). Other morphological studies have not supported this placement; Hennig (1981) placed Neuropterida as one of five clades in a basal holometabolan polytomy and Boudreaux (1979) suggested that Neuropterida was sister to Mecoptera (Mecoptera, Siphonaptera, Diptera, Trichoptera and Lepidoptera). Molecular phylogenies have thrown up a wide variety of possible sister-group relationships for Neuropterida including Mecoptera + Siphonaptera (Whiting 2002a), a novel clade consisting of (Lepidoptera + Trichoptera) + (Mecoptera + Siphonaptera) (Caterino *et al.* 2002), a novel clade consisting of Strepsiptera + (Coleoptera + Diptera) (Misof *et al.* 2007) and in line with traditional hypotheses, Coleoptera alone (Kjer *et al.* 2006). The nodal support for the sister-group relationship of Neuropterida, however, is low or insignificant in most of these studies.

Relationships within the Neuropterida have also been contentious over time. The traditional view has held a sister-group relationship between Megaloptera and Raphidioptera

(reviewed in Kristensen 1991, 1999). Recent comprehensive cladistic analyses of morphology have supported both Megaloptera + Neuroptera (Aspöck *et al.* 2001) and Megaloptera + Raphidioptera (Wheeler *et al.* 2001). Detailed analysis of ovariole structure suggests Megaloptera + Raphidioptera (Büning 2006). Molecular analyses have variously supported all possible combinations of the neuropterid orders: Raphidioptera + (Megaloptera + Neuroptera) (Haring & Aspöck 2004; Kjer *et al.* 2006); Neuroptera + (Raphidioptera + Megaloptera) (Whiting 2002a; Misof *et al.* 2007); and Megaloptera + (Raphidioptera + Neuroptera) (Caterino *et al.* 2002; Klass 2003).

Thus two critical questions regarding the phylogenetic relationships of the neuropterids remain unanswered: what is the sister group of Neuropterida (Coleoptera or some other order or group of orders); and what is the branching order of the neuropterid orders? We address both these questions using whole mitochondrial (mt) genome data which we have previously found to be particularly useful for resolving deep level relationships such as those between (e.g. Cameron *et al.* 2006) and within insect orders (e.g. Cameron *et al.* 2007; Fenn *et al.* 2008). We present new complete mt genome sequences for representatives of each of the three neuropterid orders as well as three additional beetle species to improve the available taxon sampling within this putative relative of the Neuropterida over what is presently deposited on GenBank. Whole mt genome sequences represent a large increase in data over previous taxonomically comprehensive analyses of holometabolan evolution which have relied almost exclusively on the 18S rRNA gene (e.g. Caterino *et al.* 2002; Whiting 2002a; Kjer *et al.* 2006; Misof *et al.* 2007) and despite using less exemplar taxa is thus the largest data set yet applied to the question of neuropterid relationships. In addition, we test the sensitivity of mt genome data to different analytical approaches. Previously, we have tested the influence of outgroup choice, alignment methodologies, data recoding or transformation strategies, gene exclusion, data partitioning strategies, and tree inference methods on phylogenetic reconstruction using insect mt genomes (Cameron *et al.* 2004, 2006, 2007; Fenn *et al.* 2008). Here we extend our exploration of mt genome phylogenetic utility by testing whether the ribosomal and transfer RNA genes can be usefully incorporated into studies using mt genomic data, whether these structural RNA genes are most effectively aligned using secondary structural models or algorithmic alignment methods and whether ambiguously aligned data are best removed from analyses using algorithmic methods or arbitrarily excluded based on structural regions.

Materials and methods

Mitochondrial genome sequencing

Specimens sequenced for this study include the sialid, *Sialis bamata* collected from Bryant's Fork of the Strawberry

Reservoir (Wasatch Co., UT, USA) by M. Whiting, June 23 1997; the mantispid, *Ditaxis latistyla* from suburban St. Lucia (Brisbane, Queensland, Australia) by J. Nielsen, September 13 2001; the raphidiid, *Mongoloraphidia harmandi* from Yaduki Falls (near Lake Chuzenji, Tochigi Prefecture, Japan) by M. Terry and K. Jarvis, July 12 2002; the gyrinid, *Macrogyrus oblongus* were collected from Neurum Creek (Mt Mee, Queensland, Australia) by S. Cameron, September 23 2001; the scarab, *Rhopaea magnicornis* from suburban Taringa (Brisbane, Queensland, Australia) by S. Cameron, October 29 2001; and the mordellid, *Mordella atrata* from Trout Creek Springs (Wasatch Co.) by S. Cameron July 8 2002. All specimens were snap frozen in liquid nitrogen and stored at -80°C in the insect tissue collection of the Dept. of Biology, Brigham Young University. Australian samples were identified by S. Cameron, American and Japanese samples were identified by K. Miller, H. Song and M. Whiting and voucher specimens have been deposited in the M. L. Bean Museum Insect Genomics Collection (Brigham Young University), accession numbers: IGC-MG69 (*Sialis hamata*), IGC-NE68 (*Ditaxis latistyla*), IGC-RA63 (*Mongoloraphidia harmandi*), IGC-CO132 (*Macrogyrus oblongus*), IGC-CO133 (*Rhopaea magnicornis*), and IGC-CO134 (*Mordella atrata*).

Whole genomic DNA was extracted from thoracic muscle tissue with the DNeasy Tissue kit (QIAGEN). Short regions of the *cox1*, *cox2*, *cytB*, *12S*, and *16S* genes were amplified using general insect primers and sequenced (Simon *et al.* 1994; Skerratt *et al.* 2002; Whiting 2002b; Bybee *et al.* 2004). Short sequenced regions were used to design specific primers which in combination with general insect primers allowed us to amplify the whole genome of each species by long PCR. Primer sequence and location for each long PCR is listed in Supplementary Table S1. Within each long PCR product the full, double stranded sequence was determined by primer walking (primers available from S. Cameron upon request). Short PCRs were performed using Elongase (Invitrogen) with the following cycling conditions: 95°C for 12 min; 40 cycles of 94°C for 1 min, 40°C for 1 min, 72°C for 1 min; and a final elongation of 72°C for 7 min. Long PCRs were performed using Elongase with the following cycling conditions: 92°C for 2 min; 40 cycles of 92°C for 30 s, 50°C for 30, 68°C for 12 min; and a final run out step of 68°C for 20 min. Sequencing was performed using ABI BigDye ver3 dye terminator sequencing technology and run on ABI 3770 or ABI 3740 capillary sequencer. Sequencing PCR conditions were 28 cycles of $94^{\circ}\text{C}/10$ s, $50^{\circ}\text{C}/5$ s, $60^{\circ}\text{C}/4$ min. Heteroplasmic regions in several genomes were resolved by cloning using the Topo-TA cloning chemistry (Invitrogen).

Raw sequence files were edited and assembled into contigs in Sequencher ver. 4 (GeneCodes Corporation). Transfer RNA inference was conducted using tRNAscan-SE (Lowe & Eddy 1997) using invertebrate mitochondrial predictors and

a cove score cut off of 1. Reading frames between tRNAs were found in Sequencher and identified using translated BLAST searches (blastx) (Altschul *et al.* 1997) as implemented by the NCBI website (<http://www.ncbi.nlm.nih.gov/>). Annotations of the ribosomal RNA genes were done by eye with reference to previously published insect mt rRNA gene secondary structures (cf. Cameron & Whiting 2008).

Testing regime

Our analyses sought to test the effect of three experimental design approaches on the phylogenetic analysis of mitochondrial genomes. First, we examine the effect of including the rRNA and tRNA genes along with the PCGs vs. analyses of the PCG partitions alone. Second, for those data sets which include the RNA genes we tested whether algorithmic (MUSCLE) or secondary structure guided, by-eye alignment (Secondary) are the most appropriate methods for the alignment of these genes. Finally, we examined the effect of different methods of data exclusion by comparing analyses which included the entire aligned length of each gene (ALL), exclusion by algorithmic assessment of gene regions of high alignment variability (GB) and exclusion by arbitrary removal of sites which have been proposed to have higher levels of homoplasy (ARB). For PCGs the ARB data sets exclude the third codon positions which due to redundancies in the genetic code are under lower selective pressure than other positions (Gojobori 1983). For the RNA genes, length variable stems and loop regions were excluded. This resulted in eight data sets, three of which include the entire aligned gene lengths (ALL-Muscle, ALL-Secondary, ALL-PCG), three with algorithmic removal of variable regions (GB-Muscle, GB-Secondary, GB-PCG) and two with arbitrary removal of variable regions (ARB-Secondary, ARB-PCG). Because the arbitrary removal of variable regions is based on structural position, i.e. stem-loop location, and the algorithmic alignment of RNA genes does not identify the structural homology of individual bases, removal of regions by eye would be based on subjective judgement about alignment variability and does not even have the modest repeatability of structure based exclusion criteria. Finally, each of the eight resulting data sets was analysed by both parsimony (MP) and Bayesian (BA) inference methods to examine the interaction between inference method and the other three analytical factors addressed in this study.

Phylogenetic inference

The majority of mt genomes available on GenBank for Holometabola were used (see Table 1) including representative samplings from within the orders Coleoptera (5 species in 5 families from the suborder Polyphaga), Lepidoptera (5 species in 5 families) and Diptera (11 species in 9 families). In addition we sequenced mt genomes from three additional species of Coleoptera (representing 3 additional families and

Table 1 Taxon sampling and availability.

Species	Order	Family	Accession number	Availability
<i>Tribolium castaneum</i>	Coleoptera	Tenebrionidae	NC003081	Friedrich & Muqim 2003
<i>Adelium</i> sp.	Coleoptera	Tenebrionidae	FJ613422	Sheffield <i>et al.</i> 2009
<i>Mordella atrata</i>	Coleoptera	Mordellidae	FJ859904	Present study
<i>Crioceris duodecimpunctata</i>	Coleoptera	Chrysomelidae	NC003372	Stewart & Beckenbach 2003
<i>Anoplophora glabripennis</i>	Coleoptera	Cerambycidae	NC008221	An <i>et al.</i> unpublished
<i>Rhopaea magnicomis</i>	Coleoptera	Scarabaeidae	FJ859903	Present study
<i>Rhagophthalmus lufengensis</i>	Coleoptera	Phengodidae	NC010969	Present study
<i>Pyrocoelia rufa</i>	Coleoptera	Lampyridae	NC003970	Bae <i>et al.</i> 2004
<i>Macrogyrus oblongus</i>	Coleoptera	Gyrinidae	FJ859901	Present study
<i>Sialis hamata</i>	Megaloptera	Sialidae	FJ859905	Present study
<i>Ditaxis latistyla</i>	Neuroptera	Mantispidae	FJ859906	Present study
<i>Mongoloraphidia harmandi</i>	Raphidioptera	Raphididae	FJ859902	Present study
<i>Antheraea pernyi</i>	Lepidoptera	Saturniidae	NC004622	Liu <i>et al.</i> 2008
<i>Manduca sexta</i>	Lepidoptera	Sphingidae	NC010266	Cameron <i>et al.</i> 2008
<i>Bombyx mori</i>	Lepidoptera	Bombycidae	NC002355	Lee <i>et al.</i> unpublished data
<i>Coreana raphaelis</i>	Lepidoptera	Lycaenidae	NC007976	Kim <i>et al.</i> 2006
<i>Adoxophyes honmai</i>	Lepidoptera	Tortricidae	NC008141	Lee <i>et al.</i> 2006
<i>Anopheles gambiae</i>	Diptera	Culicidae	NC002084	Beard <i>et al.</i> 1993
<i>Cydistomyia duplonotata</i>	Diptera	Tabanidae	NC008756	Cameron <i>et al.</i> 2007
<i>Trichophthalma punctata</i>	Diptera	Nemestrinidae	NC008755	Cameron <i>et al.</i> 2007
<i>Simosyrphus grandicornis</i>	Diptera	Syrphidae	NC008754	Cameron <i>et al.</i> 2007
<i>Drosophila yakuba</i>	Diptera	Drosophilidae	NC001322	Clary & Woolstenholme 1985
<i>Drosophila melanogaster</i>	Diptera	Drosophilidae	NC001709	Lewis <i>et al.</i> 1995
<i>Chrysomya putoria</i>	Diptera	Calliphoridae	NC002697	Junqueira <i>et al.</i> 2004
<i>Cochliomyia hominivorax</i>	Diptera	Calliphoridae	NC002660	Lessinger <i>et al.</i> 2000
<i>Haematobia irritans</i>	Diptera	Muscidae	NC007102	Oliveira <i>et al.</i> 2008
<i>Dermatobia irritans</i>	Diptera	Oestridae	NC006278	Azerado-Espin <i>et al.</i> unpublished data
<i>Ceratitis capitata</i>	Diptera	Tephritidae	NC000857	Spanos <i>et al.</i> 2000
<i>Abispa ephippium</i>	Hymenoptera	Vespidae	NC011520	Cameron <i>et al.</i> (2008)

an additional suborder Adephaga) and three Neuropterida species (representing each of the neuropterid orders). The analyses were rooted using a representative of the Order Hymenoptera. Several recent comprehensive analyses of relationships within Holometabola, using a variety of different marker types (nuclear RNA genes, nuclear protein coding genes, phylogenomic comparisons), have found that Hymenoptera are the sister group of the remaining holometabolous orders (Savard *et al.* 2006; Misof *et al.* 2007; Wiegmann *et al.* in prep.) and so it is an appropriate outgroup for analyses of relationships between the other holometabolous orders. Of the presently available hymenopteran mt genomes, the three bee (Apidae) species have extremely high base compositional bias (Salvato *et al.* 2008) and the genome sequence is incomplete for four of the five remaining species (Castro & Dowton 2005; Castro *et al.* 2006; Cameron *et al.* 2008) so we used the only complete hymenopteran mt genome which was not highly compositionally biased as outgroup, *Abispa ephippium* (Vespidae). Finally, the only sequenced mt genome available from the Order Strepsiptera was omitted from this analysis as we have previously shown that this genome consistently and incorrectly groups within Hymenoptera across a range of

analytical methods (Cameron *et al.* 2008). The phylogenetic position of Strepsiptera within Holometabola cannot be accurately tested using the currently available mt genome data, probably due to the extreme base composition bias found in the only available sequence (Carapelli *et al.* 2006).

An amino acid alignment was generated in ClustalW (Thompson *et al.* 1994 implemented in MEGA3) for each of the 13 PCGs and a DNA alignment inferred from the amino acid alignment using MEGA3 (Kumar *et al.* 2004) which can translate between DNA and amino acid sequences within alignments. RNA genes were aligned by two different methods—algorithmic alignment using the program MUSCLE (Edgar 2004) as implemented on the EMBL-EBI website (<http://www.ebi.ac.uk/Tools/muscle/index.html>) and a manual alignment based on standard models of the RNA secondary structure (Kjer *et al.* 2009). Secondary structures of the rRNA genes are based on those published in Gillespie *et al.* (2006) and Cameron & Whiting (2008); structure models for tRNA genes were taken from the output of the tRNAScan-SE analyses (Lowe & Eddy 1997). A small number of tRNAs were not identified by tRNAScan-SE in some species, alignments of these genes were based on similarities to the tRNA

structures inferred for other insect species. High variability gene regions were identified for the algorithmic exclusion set (GB) using the program G_blocks ver 0.9b (Castresana 2000) for each gene. For the arbitrary exclusion set (ARB) high variability regions were defined as either loop regions of the alignment which varied in length between taxa (tRNA and rRNA genes), stem-loop regions where stem length varied between taxa (rRNA genes) or third codon positions (PCGs). A list of the excluded regions in the ARB data sets is included in supplementary data (Tables S2 and S3). Alignments of individual genes were then concatenated in MacClade 4.06 (Maddison & Maddison 2003). Each of the PCGs and rRNA genes were included as separate partitions but the tRNAs were joined in a single partition for ModelTest and Bayesian analysis as the number of variable sites within individual tRNAs are too few for accurate parameter calculations.

Phylogenetic analysis was performed with PAUP 4.0b10 (Swofford 2002) for parsimony (MP) and LogDet transformed distance trees, and MrBayes (BA) ver 3.1.1 (Huelsenbeck & Ronquist 2001) for Bayesian analysis. Bootstrap supports were calculated with PAUP 4.0b10 with either 1000 (MP) or 100 (ML) replicates. Tree statistics were calculated in PAUP 4.0b10. All Bayesian analyses were run with unlinked partitions, appropriate models of molecular evolution selected for each partition and each data set analysed using two independent runs, each of four chains (3 hot and 1 cold chain), for 3 million generations with sampling every 1000 generations; convergence was achieved by all analyses within 3 million generations as determined using Tracer ver. 1.4 (Rambaut & Drummond 2007). Completed Bayesian analyses were examined for asymptotic behaviour of each parameter and of total tree likelihood; trees collected prior to this asymptotic point were treated as burn-in and discarded (generally the first 30–60 000 generations). Models for BA were chosen using AIC as implemented in ModelTest (Posada & Crandall 1998). Bayesian run files are available for each analysis from SLC upon request.

Results

Genome sequences

The entire mt genomes of three additional beetle species plus a single representative of each of the neuropterid orders were sequenced and have been submitted to GenBank with the following accession numbers: *Ditaxis* (FJ859906), *Sialis* (FJ859905), *Mongoloraphidia* (FJ859902), *Macrogyrus* (FJ859901), *Rhopaea* (FJ859903), and *Mordella* (FJ859904). These genomes range in size from 15 540 (*Mordella*) to 17 522 bp (*Rhopaea*), i.e. similar in size to other insect mt genomes. As in other mt genomes, the majority of genome size variability is due to variation in the size of the untranslated, putative control regions which ranged in size from 814 (*Sialis*) to 2827 bp (*Rhopaea*). The large repeat regions found

in many insect mt genomes were present only in the Raphidioptera mt genome which consisted of four imperfect, 164 bp repeat units (16, 8 and 7 base mismatches between the units). Additionally, the *Mongoloraphidia* specimen sequenced was heteroplasmic with respect to repeat number, of 8 clones sequenced, 4 clones had 4 repeat units, 3 clones with 3 repeats and 1 clone with 2 repeat units. Small microsatellite regions, consisting of (TA)_n (*Macrogyrus*, *Rhopaea*, *Mordella*, *Ditaxis*, *Sialis*), and (CT)_n (*Mordella*), were a common feature of the control regions sequenced in this study. Each genome has the usual metazoan complement of 37 genes (13 PCGs, 2 rRNAs and 22 tRNAs) and five of the seven studied species had the plesiomorphic pancrustacean genome arrangement (Boore 1999). *Mordella atrata* has a tRNA rearrangement to tRNA-Ala-tRNA-Asn-tRNA-Arg-tRNA-Ser^(AGN), *ANRS* rather than the plesiomorphic *ARNS*. This rearrangement has not been seen in other tenebrionoids and is likely synapomorphic at the family level or lower within Mordellidae. *Ditaxis latistyla* also has a tRNA rearrangement to tRNA-Cys-tRNA-Trp-tRNA-Tyr, *CWY* rather than the plesiomorphic *WCY*. This tRNA arrangement is found in other neuropteran families but not in the other neuropterid orders suggesting that it may be synapomorphic for either the order Neuroptera or some subgroup within this order (Cameron, unpublished data). Additional tRNA genes were inferred by COVE analysis within the control regions of several species (*Mordella*, *Ditaxis*, *Sialis*, *Mongoloraphidia*) however as these regions had limited sequence similarity to their homologous isotypes located elsewhere in the mt genome or in related taxa it is likely that they do not represent functional tRNA genes (cf. Cameron *et al.* 2007). In each species, 12 of the 13 PCGs use the regular mt start (M or I) and stop (TAA, TAG, TA, or T) codons. The exception was *cox1* which, as in other insect groups, appears to use the first in-frame codon after the upstream tRNA-Tyr gene. In three of the four beetle species the *cox1* start codon was N as in other beetles (Sheffield *et al.* 2008); in the adepghan *Macrogyrus* it was F resulting in a translation 4 amino acids longer than the consensus in polyphagan beetles. *Cox1* start codons in Neuropterida were I (*Ditaxis* and *Sialis*) or L (*Mongoloraphidia*) and the translations were respectively 4, 3 or 1 amino acids longer than the beetle and lepidopteran consensus *cox1* alignments and 3, 2, or 0 amino acids longer than in the *cox1* genes of flies.

Mitochondrial genome phylogeny of holometabola

Phylogenetic analyses of the eight data sets yielded five different patterns of the relationships between the holometabolan orders (Figs 1 and 2). Tree statistics for each analysis are listed in Supplementary Table S4 and nodal supports for major clades are listed in Table 2. The largest differences were seen between the two inference methods, parsimony and Bayesian analysis. With parsimony inference, all eight

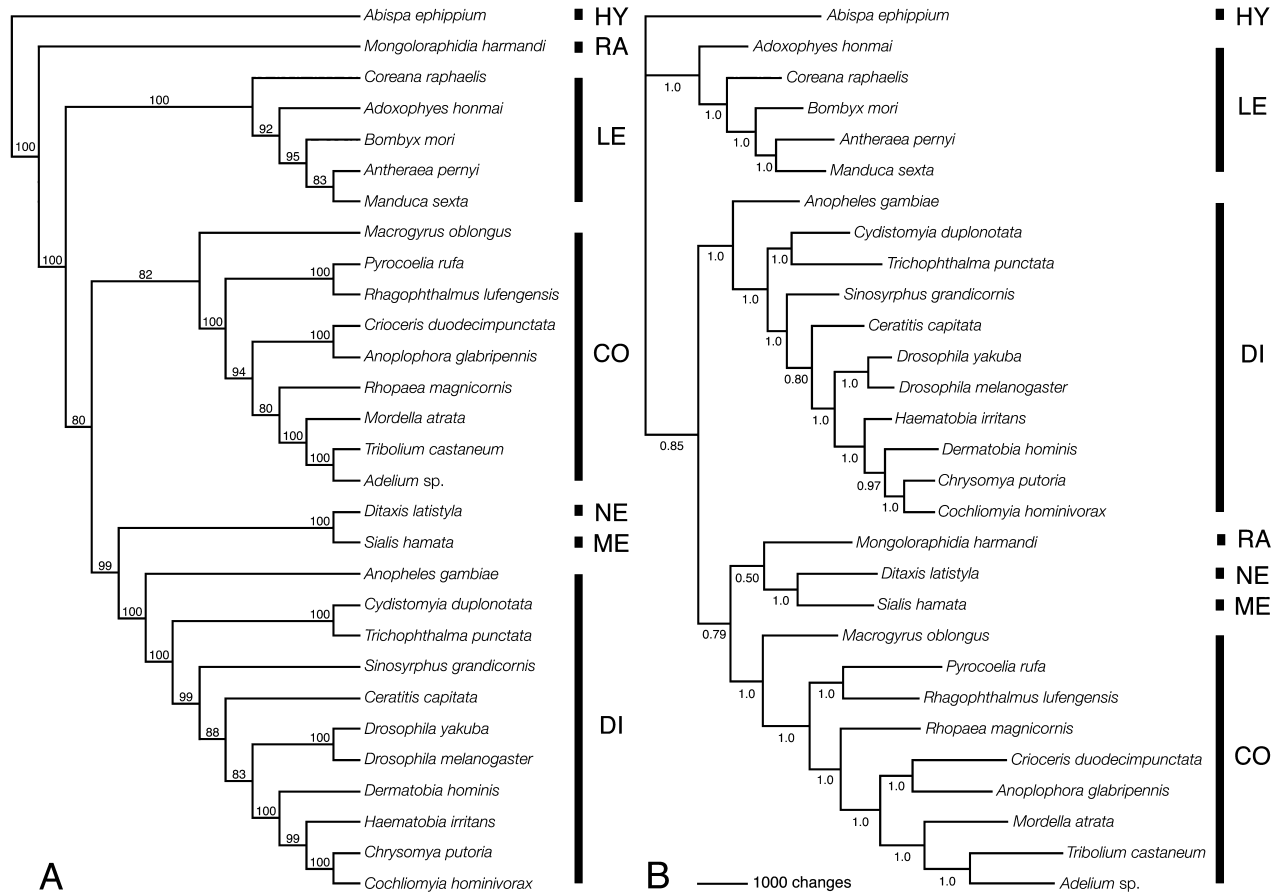


Fig. 1 A, B. Mitochondrial genome phylogeny of the Holometabola inferred from the ALL-Muscle data sets; bars to the right of each tree indicate the insect orders. —A. Parsimony cladogram (bootstrap percentage support above each node). —B. Bayesian phenogram (posterior probability support below each node). CO: Coleoptera; DI: Diptera; HY: Hymenoptera; LE: Lepidoptera; ME: Megaloptera; NE: Neuroptera; RA: Raphidioptera.

data sets supported the same topology: Hymenoptera + (Raphidioptera + (Lepidoptera + (Coleoptera + (Diptera + (Megaloptera + Neuroptera))))); bootstrap support for the relationship Diptera + (Megaloptera + Neuroptera) ranging from < 70 to 96% (Figs 1A and 2). In contrast the Bayesian analyses resulting in four different topologies, five data sets (ALL-Muscle, All-Secondary, ARB-Secondary, GB-Muscle, GB-Secondary) supported a very traditional topology: Hymenoptera + (Lepidoptera + (Diptera + (Coleoptera + (Raphidioptera + (Megaloptera + Neuroptera))))); the posterior probability for the relationship Coleoptera + Neuropterida ranged from 0.6 to 0.95 and for the monophyly of Neuropterida from 0.5 to 0.97 (Figs 1b and 2). The three analyses which omitted tRNA and rRNA data partitions (ALL-PCG, ARB-PCG, GB-PCG) resulted in three different topologies (Fig. 2), none of which supported neuropterid monophyly. The topology recovered by all the parsimony analyses was

never recovered by any of the Bayesian analyses. Of the three orders whose monophyly was testable in this study, Coleoptera, Diptera and Lepidoptera, all were monophyletic in seven of the eight analyses using both inference methods. The only exception was Coleoptera which was not monophyletic in either the BA or MP analyses of the ARB-PCG data set due to Adephaga (*Macrogyrus*) failing to group with Polyphaga. Intraordinal relationships recovered for these three orders were largely consistent across the eight data sets. Topology within Diptera varied only with respect to relationships within Muscoidea: Oestridae (*Dermatobia*) + Calliphoridae (*Chrysomya*, *Cochloomyia*) was found in the Muscle and Secondary data sets when analysed by Bayesian methods (Fig. 1B) whereas Muscidae (*Haematobia*) + Calliphoridae was found in PCG data sets when analysed by Bayesian methods and in all data sets analysed by parsimony. Within Lepidoptera, inference method had a modest effect on

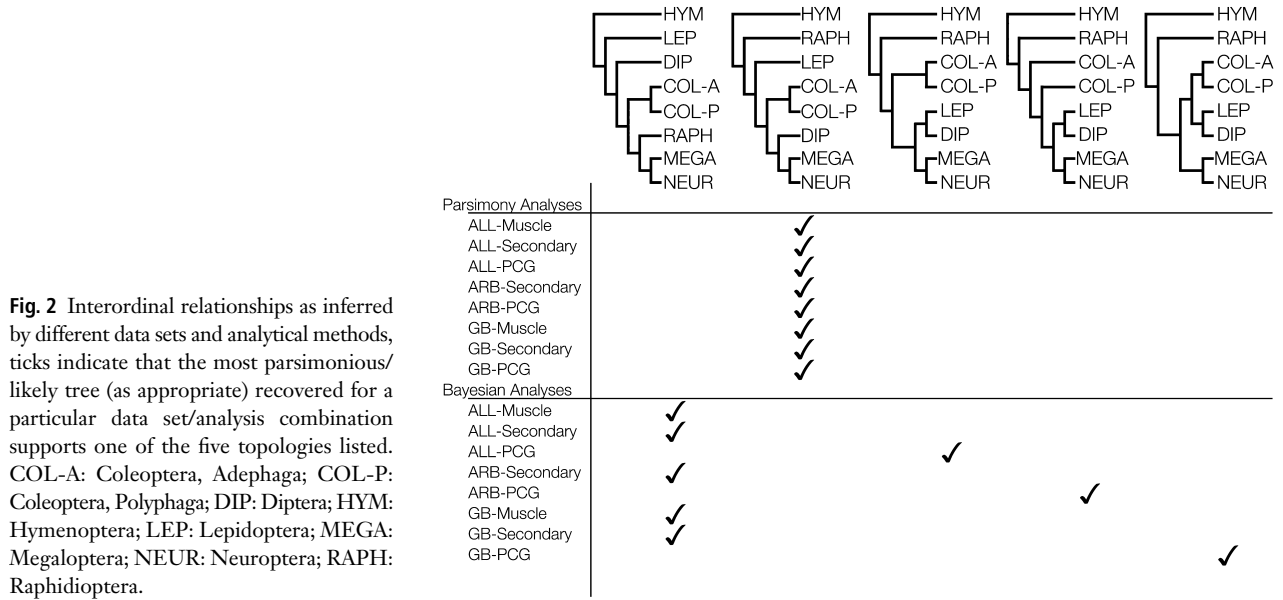


Fig. 2 Interordinal relationships as inferred by different data sets and analytical methods, ticks indicate that the most parsimonious/likely tree (as appropriate) recovered for a particular data set/analysis combination supports one of the five topologies listed. COL-A: Coleoptera, Adephaga; COL-P: Coleoptera, Polyphaga; DIP: Diptera; HYM: Hymenoptera; LEP: Lepidoptera; MEGA: Megaloptera; NEUR: Neuroptera; RAPH: Raphidioptera.

Table 2 Nodal supports for major clades recovered in each analysis. ns, not significant.

Data set	Lepidoptera	Bombycoidea	Diptera	Cyclorapha	Coleoptera	Polyphaga	Neuropterida	Neuroptera + Megaloptera
Parsimony analyses								
ALL-Muscle	100	95	100	99	82	100	ns	100
ALL-PCG	100	87	100	96	78	97	ns	100
ALL-Secondary	100	96	100	99	96	100	ns	100
ARB-PCG	100	98	100	90	Ns	100	ns	81
ARB-Secondary	100	98	100	99	92	100	ns	97
GB-Muscle	100	88	100	97	89	99	ns	100
GB-PCG	100	84	100	93	78	97	ns	99
GB-Secondary	100	96	100	96	96	99	ns	100
Bayesian analyses								
ALL-Muscle	100	100	100	100	100	100	50	100
ALL-PCG	100	100	100	100	99	100	ns	100
ALL-Secondary	100	100	100	100	100	100	60	94
ARB-PCG	100	100	100	100	ns	100	ns	100
ARB-Secondary	100	100	100	100	100	100	97	ns
GB-Muscle	100	100	100	100	100	100	ns	82
GB-PCG	100	100	100	100	100	100	ns	78
GB-Secondary	100	100	100	100	100	100	63	81

topology; Tortricoidea (*Adoxophyes*) was consistently recovered as the sister group of the rest of the order under Bayesian analyses, whereas parsimony analyses consistently recovered Papilionoidea (*Coreana*) as the sister group to the remaining lepidopterans. Variability within coleopteran relationships were the result only of variability in the position of the scarabaeoids (*Rhopaea*) which varied in response to changes in both the data set analysed and the inference method used: sister to Tenebrionoidea in the ALL and GB data sets under MP inference; sister to the clade Chysomeloidea + Tenebrionoidea

in the ALL and GB data sets under BA inference and the ARB data sets under MP inference; or as sister to the clade (Elateroidea + (Chysomeloidea + Tenebrionoidea)) for the ARB data sets under BA inference (Fig. 3).

The four factors which we sought to test within this study varied considerable in their impact on phylogenetic analysis of inter- and intraordinal relationships within the Holometabola. The effect of the inclusion of the rRNA and tRNA genes in a combined analysis of the entire mt genome vs. analysis of the PCGs alone was only significant when the data sets were

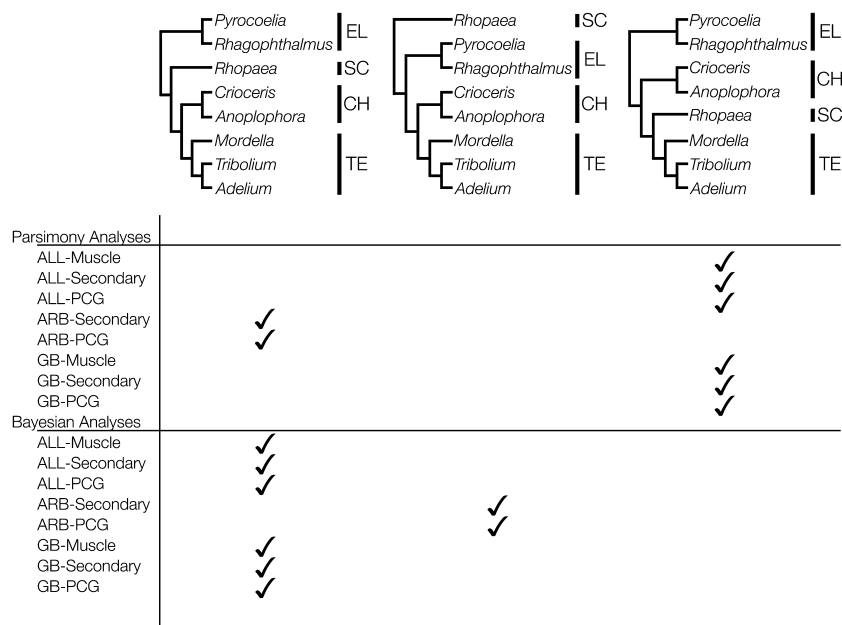


Fig. 3 Relationships within Polyphaga (Coleoptera) as inferred by different data sets and analytical methods, ticks indicate that a particular data set/analysis combination supports one of the three topologies listed; bars to the right of each tree indicate superfamily membership. CH: Chrysomeloidea; EL: Elateroidea; SC: Scarabaeoidea; TE: Tenebrionoidea.

analysed using Bayesian methods. While the topologies for each of the three PCG data sets varied, certain features were in common: Neuropterida is not monophyletic with Raphidioptera placed as the sister to the remaining ingroup taxa (Holometabola excluding Hymenoptera), Megaloptera + Neuroptera is sister to a clade consisting of Lepidoptera + Diptera (ALL-PCG, ARB-PCG) or to Coleoptera + (Lepidoptera + Diptera) (GB-PCG). In contrast the same topology was recovered from the PCG only and combined data sets using parsimony inference, however the bootstrap support recovered for the PCG analyses was generally lower than for the combined data sets. Even though the inclusion of the RNA genes resulted in substantially different topologies and higher levels of nodal support, the method of aligning the RNA partitions seems to be less important than the fact of their inclusion as there was not significant differences in either topology or nodal support between the -Muscle and -Secondary data sets. The three methods of data exclusion only resulted in different topologies within Coleoptera where the placement of *Rhopaea* recovered in the ARB data sets varied from that found in the ALL and GB data sets. Interestingly, it also varied across the two inference methods, but for both MP and BA the position of *Rhopaea* was different from the ARB data sets vs. the ALL and GB data sets suggesting that there is a strong and divergent phylogenetic signal in the third codon position relative to that in the first and second positions.

By far, however, the most significant differences are those between different inference methods. The parsimony analyses consistently recovered the non-monophyly of Neuropterida, with Raphidioptera sister to the rest of the ingroup taxa and

with Megaloptera + Neuroptera as the sister group of Diptera; this set of relationships has not previously been proposed based on morphological or other molecular data sets. Under Bayesian inference, the majority of the data sets supported a traditional set of holometabolan relationships: a monophyletic Neuropterida which was the sister group of Coleoptera, although as noted above this result depends on the inclusion of the RNA data partitions. Inference method also had an effect on intraordinal relationships within Coleoptera. The Bayesian analyses of the ALL and GB data sets supported Scarabaeoidea + (Chrysomeloidea + Tenebrionoidea) whereas parsimony analyses of the same data sets supported Chrysomeloidea + (Scarabaeoidea + Tenebrionoidea); as noted above there is also an effect of data exclusion with the position of the scarabs differing in the ARB data sets from the ALL and GB data sets under both BA and MP methods. Nodal support was also generally lower under parsimony inference than under Bayesian, as is usually found (cf. Cameron *et al.* 2004, 2006, 2007; Fenn *et al.* 2008).

A common explanation for differing topologies recovered between MP and BA is that parsimony based methods fail to account for some aspect of evolution which is successfully modelled in the later method (Felsenstein 2004). Once such factor often cited in phylogenetic analyses of mitochondrial genome data is base compositional heterogeneity (Sheffield *et al.* 2009, plus references therein). The base composition of each taxon and for each of the data partitions are given in Table 3, and the nodes which differ between the MP and BA analyses could be due to taxa with similar base compositions grouping together in the MP analyses. The average AT% of

Table 3 Mitochondrial genome nucleotide composition, A + T%.

Taxon	Protein-coding genes	Third codon positions	Ribosomal RNA genes	Transfer RNA genes	Total
<i>Abispa ephippium</i>	78.7	89.4	81.9	84.4	79.7
<i>Adoxophyes honmai</i>	78.5	92.0	84.2	81.2	79.6
<i>Coreana raphaelis</i>	81.5	96.9	85.5	82.7	82.2
<i>Bombyx mori</i>	79.6	92.8	84.8	81.6	80.5
<i>Antheraea pernyi</i>	78.5	92.3	84.0	81.4	79.6
<i>Manduca sexta</i>	80.3	95.0	85.4	81.8	81.2
Lepidoptera Average	79.7	93.8	84.8	82.3	80.6
<i>Anopheles gambiae</i>	76.1	92.3	81.4	78.0	77.0
<i>Cydistomyia duplonotata</i>	75.6	93.1	81.0	76.9	76.5
<i>Trichophthalma punctata</i>	72.0	83.5	78.2	74.8	73.2
<i>Sinosyrphus grandicornis</i>	78.9	96.6	84.6	80.4	79.9
<i>Ceratitis capitata</i>	75.6	92.3	80.2	76.8	76.4
<i>Drosophila yakuba</i>	76.7	93.8	81.9	76.5	77.4
<i>Drosophila melanogaster</i>	77.2	94.5	81.9	77.1	77.9
<i>Haematobia irritans</i>	77.7	96.3	81.3	76.9	78.2
<i>Dermatobia hominis</i>	75.3	89.8	81.4	77.0	76.4
<i>Chrysomya putoria</i>	75.0	90.3	80.0	76.1	75.8
<i>Cochliomyia hominivorax</i>	74.8	89.5	79.8	76.6	75.7
Diptera Average	75.9	92.0	81.1	77.0	76.8
<i>Macroglyus oblongus</i>	75.2	87.4	82.5	78.4	76.6
<i>Pyrocoelia ruga</i>	76.3	89.4	81.7	78.6	77.3
<i>Rhagoththalmus lufengensis</i>	78.2	94.3	82.3	79.8	78.9
<i>Rhopaea magnicornis</i>	74.5	87.9	77.5	76.2	75.1
<i>Crioceris duodecimpunctata</i>	75.2	87.6	81.3	77.6	76.4
<i>Anoplophora glabripennis</i>	77.0	92.3	80.1	78.3	77.6
<i>Mordella atrata</i>	70.6	80.5	74.4	74.8	71.6
<i>Tribolium castaneum</i>	69.2	76.4	76.1	75.6	70.8
<i>Adelium</i> sp.	70.5	78.0	77.9	76.6	72.1
Coleoptera Average	74.1	86.0	79.3	77.3	75.2
Raphidioptera	78.1	91.3	83.9	81.0	79.2
<i>Ditaxis latistyla</i>	77.3	92.8	82.0	78.6	78.0
<i>Sialis hamata</i>	76.2	91.1	83.3	79.4	77.5
Neuropterida Average	77.2	91.7	83.1	79.7	78.2

the neuropterids (78.28% in the ALL data sets) is closer to that of the dipterans (76.73%) than to that of the coleopterans (74.97%) consistent with the topological differences being the result of nucleotide compositional heterogeneity. To test this hypothesis we analysed the data set using a LogDet transform which is routinely used as a method to account for nucleotide bias. Following removal of invariant sites, the topology recovered by the LogDet transformation (Fig. 4A), is however, almost identical to that recovered by MP in that Neuropterida is not monophyletic and the clade Neuroptera + Megaloptera is the sister group of Diptera, suggesting that base compositional heterogeneity is not the cause of topological differences between the MP and BA data sets. A second factor, much more rarely invoked or investigated, is among-site rate heterogeneity (ASRH) which can be properly accounted for by MrBayes (under the gamma and invgamma parameter options) but is not accounted for by parsimony methods and is only very poorly modelled by LogDet transforms.

Reanalysing the All-Muscle data set without using gamma parameters to model ASRH, i.e. using the GTR + I model, again resulted in a tree with a very similar topology to that recovered by MP and the LogDet methods (Fig. 4B). This suggests the inclusion of a gamma parameter is responsible for the differences between the MP and BA analyses and that the topological differences are driven by significant among site rate heterogeneity present in this data set.

Discussion

Holometabolan phylogeny and mitochondrial genomics

Whole mitochondrial genome sequence data resolve a phylogeny of the Holometabola with high nodal support which is largely consistent with traditional morphology-based interpretations of endopterygote evolution. The relationships depicted in Fig. 1(B) represent our favoured topology because they are the product of more realistic evolutionary models (i.e. Bayesian GTR + I + G), are consistently recovered for

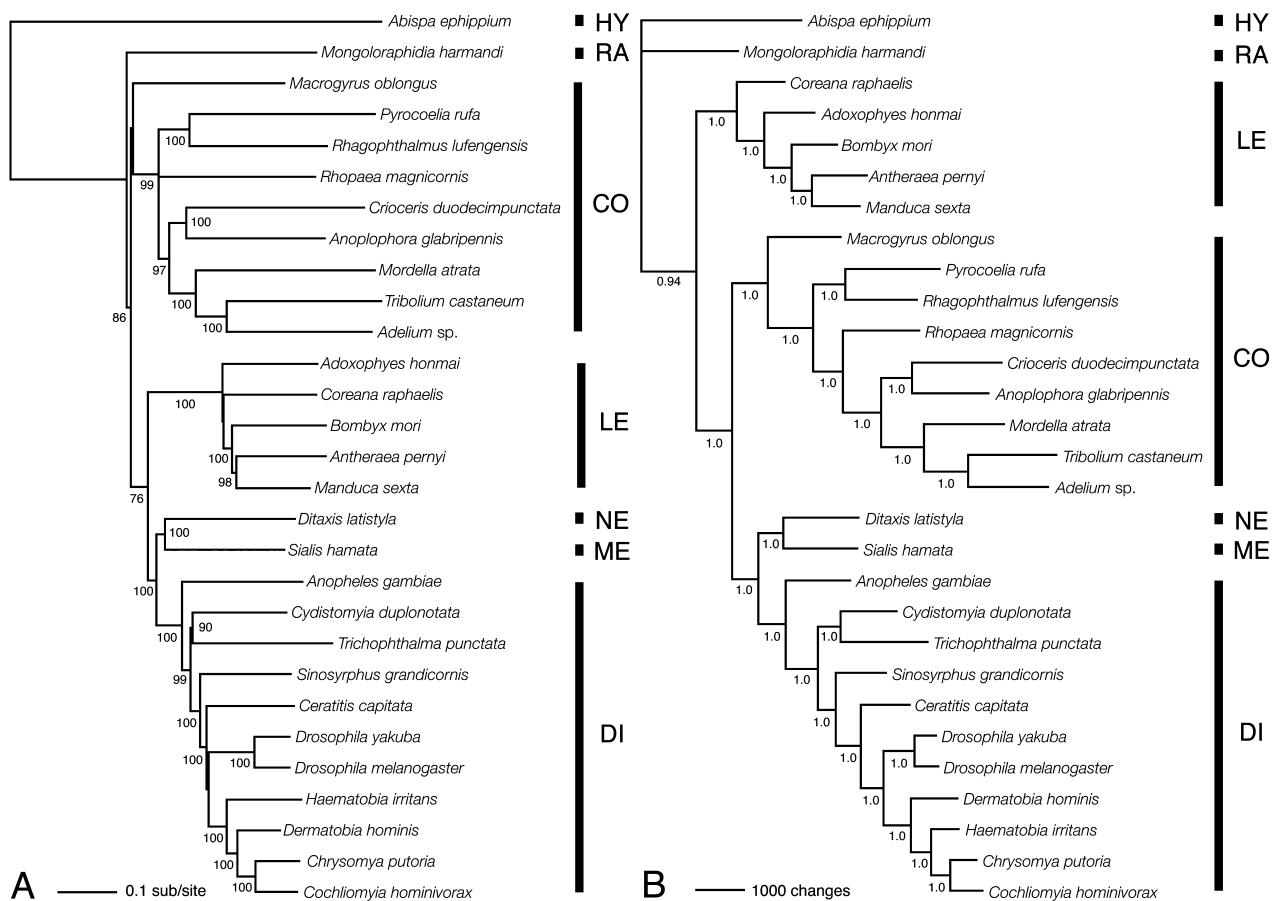


Fig. 4 A, B. Mitochondrial genome phylogeny of the Holometabola inferred from the ALL-Muscle data sets; bars to the right of each tree indicate the insect orders. —A. Phylogram of LogDet transformed data (bootstrap percentage support below each node). —B. Bayesian phenogram using the GTR + I evolutionary model (posterior probability support below each node). CO: Coleoptera; DI: Diptera; HY: Hymenoptera; LE: Lepidoptera; ME: Megaloptera; NE: Neuroptera; RA: Raphidioptera.

data sets which include all mt gene partitions, they better reflect the consensus view of the morphological data and finally because, the main alternative topology (Fig. 1A) would require a far more extensive reinterpretation of holometabolite morphological evolution than the favoured topology. While support is not uniformly high for all nodes presented in this topology (see Table 2), this topology will be used as the starting point for all our discussion of holometabolite evolution and relationships.

Of the 11 widely accepted holometabolite orders, including Siphonaptera (contra Whiting 2002b), mitochondrial genome data is now available for eight of these orders and seven are included in the present study; Strepsiptera is excluded due to the known analytical artifacts of the only available strepsipteran mt genome (Cameron *et al.* 2008). This is the most extensive sampling of holometabolite mt genomes assembled to date, is the largest molecular data set, in terms of nucleotides

sequenced per species, to be applied to the question of holometabolite relationships so far and allows testing of several proposed supraordinal groups within the Holometabola. Most notably given the emphasis of this study, we find that Neuropterida is the sister-group of Coleoptera, Neuropterida is monophyletic, and within Neuropterida, Neuroptera is more closely related to Megaloptera than to Raphidioptera. The nodal support for these relationships was variable across the data sets, ranging from very strong in the ARB-Secondary data set (Coleoptera + Neuropterida, 0.95; Neuropterida monophyly, 0.97) to relatively weak in the ALL-Muscle data set (Coleoptera + Neuropterida, 0.79; Neuropterida monophyly, 0.5). While morphological support for the sister grouping of Neuropterida + Coleoptera is limited (Wheeler *et al.* 2001) it has been increasing and now includes characters of the female terminalia (Achtelig 1975; Kristensen 1991), wing base structure (Hörnschemeyer 1998) and ovariole structure

(Büning 2006). Of the molecular analyses of holometabolism relationships, only Kjer *et al.* (2006) supports Coleoptera + Neuroptera. This result, however, is the product of extremely precise analytical conditions and many other studies utilizing the same data source, 18S rRNA genes, arrive at widely different results (cf. Caterino *et al.* 2002; Whiting 2002a; Misof *et al.* 2007). The present study is among the first robust molecular evidence for the long accepted sister group relationship between Coleoptera and Neuroptera.

The other major proposed group within Holometabola, Mecopterida (= Mecoptera, Siphonaptera, Diptera, Trichoptera and Lepidoptera), is not supported in the present study. The only mecopterid orders represented here, Lepidoptera and Diptera, are not sister groups in the majority of analyses (only supported in the PCG data sets which exclude rRNA and tRNA partitions). Mecopterida has not been well supported in other molecular analyses and is not monophyletic in any of the taxonomically comprehensive analyses (Caterino *et al.* 2002; Whiting 2002a; Kjer 2004; Kjer *et al.* 2006; Misof *et al.* 2007) although it is recovered in some data sets which include comparable sampling to that used here (e.g. Hayward *et al.* 2005). This is perhaps not surprising as the morphological features which unite this group are inconspicuous and some are subject to parallelisms with features in other holometabolous orders (Kristensen 1991). Indeed as Hennig (1981) remarked 'Most entomologists and palaeontologists now agree that the Trichoptera, Lepidoptera, Mecoptera and Diptera form a single monophyletic group, but it is not easy to find reliable evidence for this.' Given that mt genome sequences are not yet available for three of the five mecopterid orders the monophyly and relationships of this group cannot in any way be considered testable in the present study. Nuclear phylogenomic data sets (Savard *et al.* 2006) which are almost as limited in taxonomic scope as the present study, do tentatively support Mecopterida albeit with only 4 of the 11 holometabolous orders included (Hymenoptera + (Coleoptera + (Diptera + Lepidoptera))). The utility of mt genomic data in resolving neuropterid relationships gives us confidence that the inclusion of representatives of Trichoptera, Mecoptera and Siphonaptera (work which is in progress by the authors) will also allow meaningful testing and resolution of the Mecopterida.

The relationships within Neuropterida have been contentious for some time. Morphological data has favoured both Raphidioptera + Megaloptera, supported by the structure of the abdominal base and female genitalia, (Achtelig 1975; Kristensen 1981, 1991; Wheeler *et al.* 2001) or ovariolar structure (Büning 2006) as well as Neuroptera + Megaloptera, supported by wing-venation characters (Boudreaux 1979) or the presence of aquatic larvae, shift of larval cardines into the head and elongation of the larval stipites (Aspöck *et al.* 2001). Aspöck *et al.* (2001) are also at pains to point out that while they arrive at the same set of relationships as those proposed

by Boudreaux (1979) they do not accept that wing venation characters support his conclusions. The majority of molecular phylogenetic studies of holometabolism relationships include only a very poor sampling of neuropterid taxa and in general the nodes within Neuropterida are poorly supported (e.g. Whiting 2002a) or unreported (Caterino *et al.* 2002; Misof *et al.* 2006). The two most taxonomically comprehensive analyses each reach different results regarding relationships within Neuropterida. Winterton, (in Klass 2003) using a combination of 18S sequence and morphological characters, found a paraphyletic Megaloptera to be sister group to a clade composed of a monophyletic Raphidioptera and a monophyletic Neuroptera; nodal support for these relationships is, however, poor. Haring & Aspöck (2004) using 18S, elongation factor-1 α , and the mitochondrial *cox3* gene find strong support for Megaloptera + Neuroptera in accordance with an earlier morphological analysis (Aspöck *et al.* 2001). Our data almost universally supports a sister group relationship between Neuroptera and Megaloptera with high nodal support; only the ARB-Secondary data set (when analysed by Bayesian inference) failed to support this relationship recovering instead an unresolved Neuropterida. The concordance between well-sampled molecular data sets (Haring & Aspöck 2004; the present study) and recent morphological analyses (Aspöck *et al.* 2001) suggests that support for the sister grouping of Megaloptera + Neuroptera should now be considered strong. The monophyly of Neuropterida, on the other hand, is less well-supported; in molecular data sets it is dependent on different analytical methods (cf. Fig. 2) and morphological evidence is less obvious (Kristensen 1991; Aspöck 2002).

As in our previous studies, the relationships within each of the three large insect orders included, Coleoptera, Lepidoptera and Diptera, are largely congruent with previous phylogenetic hypotheses for each order. Our results within Diptera mirror those of our previous mt genome phylogeny for this order (Cameron *et al.* 2007) and are very close to the current consensus view of fly relationships (Yeates & Wiegmann 2005). Lepidopteran relationships differ depending on inference method with Tortricioidea sister to the clade Papilionoidea + Bombycoidea in Bayesian analyses vs. Papilionoidea sister to the clade Tortricioidea + Bombycoidea in parsimony analyses. The Bayesian results are consistent with the current consensus view of lepidopteran relationships (Kristensen & Skalski 1999), however, previous mt genome phylogenies have supported the parsimony result (Kim *et al.* 2008) and preliminary nuclear protein coding gene phylogenies of the Lepidoptera suggest similar results (Zwick *pers. comm.*). Coleopteran relationships were more variable with three different topologies supported depending on different analytical methods approaches to excluding variable data (Fig. 3). All the topological variability, however, relates to the placement of Scarabaeoidea (*Rhopaea*) within the phylogeny.

Tenebrionoidea, Chrysomeloidea and Elateroidea are monophyletic in all analyses and the traditionally accepted close relationship between Tenebrionoidea and Chrysomeloidea (both members of the Cucujiformia) is supported in all the Bayesian analyses and some of the parsimony analyses (ARB-PCG, ARB-Secondary data sets). In addition, it should be noted that the present study is the first molecular phylogeny of beetles which includes a wide variety of potential outgroups to resolve a monophyletic Coleoptera (15 of 16 data sets/analysis combinations). Most molecular phylogenies of this group have either used extremely limited outgroups, e.g. Hunt *et al.* (2007) use 1877 ingroup coleopterans vs. 3 outgroup neuropterids, or if a representative selection of other holometabolan insect orders are used as outgroups Coleoptera was not monophyletic. For example, Caterino *et al.* (2002) found Strepsiptera + Diptera (= Halteria) nesting within Coleoptera; Whiting 2002a) found Adephaga (partial) + Amphiesmenoptera and Adephaga (rest) + Halteria; and Misof *et al.* (2007) found Diptera nesting within Coleoptera. The demonstrated capacity of mt genome data to recover a monophyletic Coleoptera as well as to reliably recover widely-accepted monophyletic groups within Coleoptera in present study suggests that this marker will be of considerable utility in resolving relationships within the largest of the insect orders.

Methodological approaches to mitochondrial phylogenomics of insects

In previous studies we have investigated the influence of outgroup, alignment, and data coding, exclusion and partitioning strategies (Cameron *et al.* 2004, 2006, 2007; Fenn *et al.* 2008). In the present instance we examined four analytical approaches: whether to include the rRNA and tRNA genes in total evidence analyses, how best to align these genes, whether it is better to use arbitrary or algorithmic methods of excluding highly variable gene regions and finally how each of these factors interacts with choice of phylogenetic inference methods (parsimony vs. Bayesian inference).

We find no reason to exclude the rRNA and tRNA genes from mt genome phylogenetic studies. There were no topological differences between data sets including or excluding RNA genes under parsimony analysis although nodal support is generally higher for data sets including these genes (Fig. 2; Table 2). Topologies varied significantly between the three data sets excluding RNA genes under Bayesian analysis and these topologies are less consistent with morphological data than the data sets including RNA genes, i.e. Neuropterida is not monophyletic and Raphidioptera is the sister to the remaining ingroup Holometabola (Fig. 2). In contrast, the data sets which include the RNA genes all give the same interordinal relationships, nodal support is higher and the topology is much more consistent with morphological data (Fig. 2, Table 2). In an earlier analysis (Cameron *et al.* 2007)

we found that including the RNA genes did not significantly alter the recovered topology, however it did improve nodal support as the tRNA genes had significantly lower levels of homoplasy than the protein-coding genes. In the present study we find no effect on topology and a limited effect on nodal support due to the method used to align the RNA genes except for the interordinal nodes in Bayesian analyses where nodal support was generally higher for data sets aligned using secondary structure guides than for algorithmic alignments. Previously, we have proposed that alignment of the rRNA genes might be problematic for mt genome phylogenies as they had much higher levels of homoplasy than either the protein-coding or tRNA genes (Cameron *et al.* 2006). In that analysis, however, we used quite simplistic alignment methods (ClustalW, Thompson *et al.* 1994) compared with those used here (Muscle; Edgar 2004) and it is conceivable that the effect noted in this study is entirely due to the evolutionary timescales over which these alignments were inferred. Including broader ranges of taxa or examining deeper phylogenetic relationships might lead to decreased reliability of algorithmic alignment methods or increased difficulty of applying secondary structure models to more distantly related taxa. It is thus too early to endorse either approach as more suitable for future mt phylogenomic analyses.

Multiple methods of data exclusion have been attempted for insect mt genomic phylogenies ranging from excising whole genes (e.g. Nardi *et al.* 2003; Friedrich & Muqim 2003), to removing regions considered hypervariable and potentially misleading such as third codon positions of the protein-coding genes (e.g. Kim *et al.* 2008). These studies have rarely justified the reasons for data exclusion beyond 'alignment difficulties' or compared topologies between data sets including and excluding the variable regions. There are limited topological or nodal support differences between data sets including all positions and those which remove variable regions based on arbitrary criteria (i.e. gene structures) or algorithmic approaches (i.e. G-blocks; Castresana 2000) except within Coleoptera. Interestingly, the major difference in coleopteran relationships between parsimony and Bayesian analyses are resolved by the arbitrary exclusion of third codon positions. In parsimony analyses of data sets which include third codon positions, ALL and GB data sets, Scarabaeoidea groups with Tenebrionoidea rendering Cucujiformia paraphyletic (Fig. 3). In contrast, all the Bayesian analyses and the parsimony analyses which exclude third codon positions, ARB data sets, recovered Chrysomeloidea + Tenebrionoidea consistent with the generally accepted monophyly of Cucujiformia (Hunt *et al.* 2007). The LogDet transform also results in the same coleopteran relationships as the Bayesian analyses (Fig. 4A). In combination, these all suggest that the unexpected result found by parsimony is a result of base compositional biases at the third codon positions which can

be solved in any of several ways — exclusion of the biased positions (ARB data sets plus parsimony analysis), evolutionary modelling (Bayesian analyses) or data transformation (LogDet transforms). Further, these biases must be less significant at the intraordinal than at the interordinal level as multiple methods adequately correct for them within Coleoptera while only accurate model specification corrects for them at the interordinal level allowing successful recovery of a monophyletic Neuropterida.

By far the most significant factor effecting phylogenetic reconstruction in the present study was the choice of inference method. This stands in some contrast to our previous studies where inference method has either had no effect (e.g. within Diptera, Cameron *et al.* 2007) or the effect was limited compared to other factors tested in the study such as outgroup choice (e.g. between arthropod classes, Cameron *et al.* 2004). In this respect, the present study is most similar to our previous examination of inter-ordinal relationships between Mantophasmatodea and related hemimetabolan insect orders (Cameron *et al.* 2006) where Bayesian analyses converged on a single topology whereas parsimony analyses supported a range of relationships with poor nodal support. In that study however, at least some of the data sets agreed between both parsimony and Bayesian analyses and inference method appeared to be secondary in importance compared to data recoding which was responsible for far more topological variation between data sets. In the present study, the differences between the parsimony and Bayesian analyses are far more absolute; all data sets gave the same result under parsimony and all data sets including the RNA genes gave a single, different result under Bayesian. Inference method also affects intra-ordinal relationships within both Lepidoptera and Coleoptera, in combination with other factors such as data exclusion methods. In all instances however, the Bayesian analysis is more consistent with previous phylogenetic analyses drawing on other molecular markers, morphological data or traditional taxonomic opinion. Fortunately, we can experimentally demonstrate why parsimony methods result in erroneous trees for this data set and which characteristics of these data sets are responsible for these errors (cf. Fig. 4) and so are not placed in the position of choosing preferred trees simply based on prejudices derived from earlier works. These results suggest that accurate inference of inter-ordinal relationships using mt genome data will require sophisticated analytical methods and that data exploration will continue to be vitally important to understanding the resulting phylogenies and exploring unexpected relationships.

Conclusion

Subject to correct use of phylogenetic inference methods, whole mt genome sequence data result in a well supported phylogeny of the Holometabola which corroborate tradi-

tional hypotheses of holometabolan relationships. We find a monophyletic Neuropterida whose sister group is the Coleoptera although nodal support for this relationship varies between different data sets. Each included order whose monophyly could be tested was found to be monophyletic and intra-ordinal relationships are in accordance with both previous morphological and molecular phylogenies of each order. Nodal support was uniformly high across different data sets for the majority of nodes and branch lengths within the ingroup appear to be balanced (i.e. no obvious long branched taxa). The inclusion of the RNA genes, in combination with correct choice of evolutionary model, appears to be vital to inferring the tree which is most consistent with other data sources. The methods used for aligning RNA genes and approaches to data exclusion appear to have limited influence on topology but nodal support of interordinal relationships did vary between approaches. Accordingly, and consistent with our previous studies of the utility of mt genomes in resolving deep-level insect relationships we suggest that all genes be included, all nucleotide positions be included, and the influence of different phylogenetic inference methods be experimentally assessed for each study. Mt genomic data are extremely useful for resolving insect phylogenetic relationships but it needs to be analysed rigorously.

Acknowledgements

The authors would like to thank Brian Wiegmann (North Carolina State University), David Yeates (Australian National Insect Collection) and Nathan Sheffield (Duke University) for their advice and support of this research. This study was supported by NSF grants DEB0120718 & EF0531665 (MFW) and DEB0444972 (MFW, SLC).

References

- Achtelig, M. (1975). Die Abdomenbasis der Neuropteroidea (Insect, Holometabola). Eine vergleichend anatomische Untersuchung des Skeletts und der Muskulatur. *Zoomorphologie*, 82, 201–242.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., 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.
- Aspöck, U. (2002). Phylogeny of the Neuropterida (Insecta: Holometabola). *Zoologica Scripta*, 31, 51–55.
- Aspöck, U., Plant, J. D. & Nemeschkal, H. L. (2001). Cladistic analysis of Neuroptera and their systematic position within Neuropterida (Insecta: Holometabola: Neuropterida: Neuroptera). *Systematic Entomology*, 26, 73–86.
- Bae, J. S., Kim, I., Sohn, H. D. & Jin, B. R. (2004). The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. *Molecular Phylogenetics and Evolution*, 32, 978–985.
- Beard, C. B., Hamm, D. M. & Collins, F. H. (1993). The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence,

- genome organisation and comparisons with mitochondrial sequences of other insects. *Insect Molecular Biology*, 2, 103–124.
- Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, 27, 1767–1780.
- Boudreaux, H. B. (1979). *Arthropod Phylogeny, with Special Reference to Insects*. New York: John Wiley & Sons.
- Büning, J. (2006). Ovariole structure supports sistergroup relationship of Neuropterida and Coleoptera. *Arthropod Systematics and Phylogeny*, 64, 115–126.
- Bybee, S. M., Taylor, S. D., Nelson, C. R. & Whiting, M. F. (2004). A phylogeny of robber flies (Diptera: Asilidae) at the subfamilial level: molecular evidence. *Molecular Phylogenetics and Evolution*, 30, 789–797.
- Cameron, S. L., D’Hearse, C. A., Miller, K. B., 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., Barker, S. C. & Whiting, M. F. (2006). Mitochondrial genomics and the relationships and validity of 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). Utility of mitochondrial genomes as phylogenetic markers for insect intraordinal relationships — A case study from flies (Diptera). *Systematic Entomology*, 32, 40–59.
- Cameron, S. L., Dowton, M., Castro, L. R., Ruberu, K., Whiting, M. F., Austin, A. D., Diement, K. & Stevens, J. (2008). The sequence of the mitochondrial genomes of two vespid wasps reveals a number of derived tRNA gene rearrangements. *Genome*, 51, 800–808.
- Cameron, S. L. & Whiting, M. F. (2008). The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene*, 408, 112–123.
- Carapelli, A., Vannini, L., Nardi, F., Boore, J. L., Beani, L., Dallai, R. & Frati, F. (2006). The mitochondrial genome of the entomophagous endoparasite *Xenos vesparum* (Insecta: Strepsiptera). *Gene*, 376, 248–259.
- Castresana, J. (2000). Selection of conservation blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17, 540–552.
- 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., Ruberu, K. & Dowton, M. (2006). Mitochondrial genomes of *Vanbornia eucnemidarum* (Apocrita: Vanhorniidae) and *Primeuchbroeus* spp. (Aculeata: Chrysididae): Evidence of rearranged genomes with Apocrita (Insecta: Hymenoptera). *Genome*, 49, 752–766.
- Caterino, M. S., Shull, V. L., Hammond, P. M. & Vogler, A. P. (2002). Basal relationships of Coleoptera inferred from 18S rDNA sequences. *Zoologica Scripta*, 31, 41–49.
- Clary, D. O. & Wolstenholme, D. R. (1985). The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution*, 22, 252–271.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.
- Felsenstein, J. (2004). *Inferring Phylogenies*. Sunderland: Sinauer Associates.
- 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. *Molecular Phylogenetics and Evolution*, 49, 59–68.
- Friedrich, M. & Muqim, N. (2003). Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castaneum*. *Molecular Phylogenetics and Evolution*, 26, 502–512.
- Gillespie, J. J., Johnston, J. S., Cannone, J. J. & Gutell, R. R. (2006). Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) Rna genes of *Apis mellifera* (Insecta: Hymenoptera): Structure, organization and retrotransposable elements. *Insect Molecular Biology*, 15, 657–686.
- Gojobori, T. (1983). Codon substitution in evolution and the ‘saturation’ of synonymous changes. *Genetics*, 105, 1011–1027.
- Haring, E. & Aspöck, U. (2004). Phylogeny of the Neuropterida: a first molecular approach. *Systematic Entomology*, 29, 415–430.
- Hayward, D. C., Trueman, J. W., Bastiani, M. J. & Ball, E. E. (2005). The structure of the USP/RXR of *Xenos pecki* indicates that Strepsiptera are not closely related to Diptera. *Development Genes and Evolution*, 215, 213–219.
- Hennig, W. (1981). *Insect Phylogeny*. Chichester: John Wiley & Sons.
- Hörnchemeyer, T. (1998). Morphologie und evolution des flügelgehlnks der coleoptera und neuropterida. *Bonner Zoologische Monographien*, 43, 1–126.
- Huelsensbeck, J. P. & Ronquist, F. R. (2001). MrBayes: Bayesian inference of phylogeny. *Biometrics*, 17, 754–755.
- Hunt, T., Bergsten, J., Levakanicova, Z., Papanicolaou, A., St John, O., Wild, R., Hammond, P. M., Ahrens, D., Balke, M., Caterino, M. S., Gomez-Zurita, J., Ribera, I., Barraclough, T. G., Bocakova, M., Bocak, L. & Vogler, A. P. (2007). A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science*, 318, 1913–1916.
- Junqueira, A. C. M., Lessinger, A. C., Torres, T. T., Rodrigues da Silva, F., Vettore, A. L., Arruda, P. & Azeredo Espin, A. M. L. (2004). The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene*, 339, 7–15.
- Kim, I., Lee, E. M., Seol, K. Y., Yun, E. Y., Lee, Y. B., Hwang, J. S. & Jin, B. R. (2006). The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). *Insect Molecular Biology*, 15, 217–225.
- Kim, M. Y., Lee, E. M., Jo, Y. H., Park, H. C., Kim, S. R., Hwang, J. S., Jin, B. R., Kang, P. D., Kim, K.-G., Han, Y. S. & Kim, I. (2008). Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects. *Gene*, 413, 49–57.
- Kjer, K. M. (2004). Aligned 18S and insect phylogeny. *Systematic Biology*, 53, 506–514.
- Kjer, K. M., Carle, F. L., Litman, J. & Ware, J. (2006). A molecular phylogeny of Hexapoda. *Arthropod Systematics and Evolution*, 64, 35–44.
- Kjer, K. M., Roshan, U. & Gillespie, J. G. (2009). Structural and Evolutionary Considerations for Multiple Sequence alignment of RNA, and the challenges for Algorithms that ignore them. In M. Rosenberg (Ed.) *Perspectives on Biological Sequence Alignment*:

- Where, How, and Why It Matters*. Berkeley: University of California Press.
- Klass, K.-D. (2003). Proceedings of the 1st Dresden meeting on insect phylogeny: 'Phylogenetic relationships within the insect orders' (Dresden, September 19–21, 2003). *Entomologische Abhandlungen*, 61, 119–172.
- Kristensen, N. P. (1981). Phylogeny of insect orders. *Annual Review of Entomology*, 26, 135–157.
- Kristensen, N. P. (1991). Phylogeny of extant hexapods. In I. D. Naumann, P. B. Carne, J. F. Lawrence, E. S. Nielsen, J. P. Spradbery, R. W. Taylor & M. J. Whitten & M. J. Littlejohn (Eds) *The Insects of Australia* (pp. 125–140). Melbourne: Melbourne University Press.
- Kristensen, N. P. (1999). Phylogeny of endopterygote insects, the most successful lineage of living organisms. *European Journal of Entomology*, 96, 237–253.
- Kristensen, N. P. & Skalski, A. W. (1999). Phylogeny and palaeontology. In N. P. Kristensen (Ed.) *Handbuch der Zoologie Band IV Arthropoda: Insecta Part 35 Lepidoptera, Moths and Butterflies* (pp. 7–25). Berlin: Walter de Gruyter Publishing.
- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, 5, 150–163.
- Lee, E.-S., Shin, K.-S., Kim, M.-S., Park, H., Cho, S. & Kim, C.-B. (2006). The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene*, 373, 52–57.
- Lessinger, A. C., Junqueira, A. C. M., Lemos, T. A., Kemper, E. L., Rodrigues da Silva, F., Vettore, A. L., Arruda, P. & Azeredo Espin, A. M. L. (2000). The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Molecular Biology*, 9, 521–529.
- Lewis, D. L., Farr, C. L. & Kaguni, L. S. (1995). *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. *Insect Molecular Biology*, 4, 263–278.
- Liu, Y., Li, Y., Pan, M., Dai, F., Zhu, X., Lu, C. & Xiang, Z. (2008). The complete mitochondrial genome of the Chinese oak silkworm, *Antheraea pernyi* (Lepidoptera: Saturniidae). *Acta Biochimica et Biophysica Sinica*, 40, 693–703.
- 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. & Maddison, D. (2003). *Macclade* Ver 4.06. Sunderland: Sinauer Associates.
- Misof, B., Niehuis, O., Bischoff, I., Rickert, A., Erpenbeck, D. & Stankiczek, A. (2007). Towards an 18S phylogeny of hexapods: Accounting for group-specific character covariance in optimized mixed nucleotide/doublet models. *Zoology*, 110, 409–429.
- Nardi, F., Spinsanti, G., Boore, J. L., Carapelli, A., Dallai, R. & Frati, F. (2003). Hexapod origins: monophyletic or paraphyletic? *Science*, 299, 1887–1889.
- Norris, K. R. (1991). General biology. In I. D. Naumann, P. B. Carne, J. F. Lawrence, E. S. Nielsen, J. P. Spradbery, R. W. Taylor & M. J. Whitten & M. J. Littlejohn (Eds) *The Insects of Australia* (pp. 68–108). Melbourne: Melbourne University Press.
- Oliveira, M. T., Grande Barau, J., Junqueira, A. C. M., Feijao, P. C., da Rosa, A. C., Abreu, C. F., Azeredo Espin, A. M. L. & Lessinger, A. C. (2008). Structure and evolution of the mitochondrial genomes of *Haematobia irritans* and *Stomoxys calcitrans*: The Muscidae (Diptera: Calyptratae) perspective. *Molecular Phylogenetics and Evolution*, 48, 850–857.
- Posada, D. & Crandall, K. A. (1998). ModelTest: Testing the best-fit model of nucleotide substitution. *Bioinformatics*, 14, 817–818.
- Rambaut, A. & Drummond, A. J. (2007). *Tracer V1.4*. Available via <http://beast.bio.ed.ac.uk/Tracer>.
- Salvato, P., Simonato, M., Battisti, A. & Negrisola, E. (2008). The complete mitochondrial genome of the bag-shelter moth *Ocbrogaster lunifer* (Lepidoptera: Notodontidae). *BMC Genomics*, 9, 331.
- Savard, J., Tautz, D., Richards, S., Weinstock, G. M., Gibbs, R. A., Werren, J. H., Tettelin, H. & Lercher, M. J. (2006). Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome Research*, 16, 1334–1338.
- Sheffield, N. C., Song, H., Cameron, S. L. & Whiting, M. F. (2008). A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. *Molecular Biology and Evolution*, 25, 2499–2509.
- Sheffield, N. C., Song, H., Cameron, S. L. & Whiting, M. F. (2009). Resolving base compositional heterogeneity in beetle mitochondrial genomes (Insecta: Coleoptera): a survey of phylogenetic methods. *Systematic Biology*, in press.
- Simon, C., Frati, F., Backenbach, A. T., Crespi, B., Liu, B. & Flook, P. (1994). Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, 87, 651–701.
- Skerratt, L. F., Campbell, N. J. H., Murrell, A., Walton, S., Kemp, D. & Barker, S. C. (2002). The mitochondrial 12S gene is a suitable marker of populations of *Sarcoptes scabiei* from wombats, dogs and humans in Australia. *Parasitology Research*, 88, 376–379.
- Spanos, L., Koutroumbas, G., Kotsyfakis, M. & Louis, C. (2000). The mitochondrial genome of the mediterranean fruit fly, *Ceratitis capitata*. *Insect Molecular Biology*, 9, 139–144.
- 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.
- Swofford, D. L. (2002). *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sunderland: Sinauer Associates.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673–4680.
- Wheeler, W. C., Whiting, M. F., Wheeler, Q. D. & Carpenter, J. M. (2001). The phylogeny of the extant hexapod orders. *Cladistics*, 17, 113–169.
- Whiting, M. F. (2002a). Phylogeny of the holometabolous insect orders: molecular evidence. *Zoologica Scripta*, 31, 3–15.
- Whiting, M. F. (2002b). Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zoologica Scripta*, 31, 93–104.
- Wilson, E. O. (1988). The current state of biological diversity. In E. O. Wilson (Ed.) *Biodiversity* (pp. 3–18). Washington DC: National Academy Press.
- Yeates, D. K. & Wiegmann, B. M. (2005). Phylogeny and evolution of Diptera: Recent insights and new perspectives. In D. K. Yeates & B. M. Wiegmann (Eds) *The Evolutionary Biology of Flies* (pp. 14–44). New York City: Columbia University Press.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Primers, sequence and location for long PCRs.

Table S2 Regions deleted from 16S alignment.

Table S3 Regions deleted from the 12S alignment.

Table S4 Tree statistics. CI: consistency index; RI: retention index; RC: rescaled consistency index.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.