Arthropod Systematics & Phylogeny 27
64 (1) 27–34

Evidence from Mitochondrial Genomics on Interordinal Relationships in Insects

Stephen L. Cameron [1], Andrew T. Beckenbach [2], Mark Dowton [3] & Michael F. Whiting [1]

[1] Department of Integrative Biology, Brigham Young University, Provo, Utah, 84602, United States of America [stephen_cameron@email.byu.edu; Michael_Whiting@byu.edu]
[2] Department of Biological Sciences, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada [abeck@menten.mbb.sfu.ca]
[3] School of Biological Sciences, University of Wollongong, Wollongong, NSW, 2522, Australia [mdowton@uow.edu.au]

Received 26.x.2005, accepted 26.vii.2006.
Available online at www.arthropod-systematics.de

> Abstract

Mitochondrial (mt) genomes are the largest molecular data source for deep level insect phylogenetics that is also obtainable in a reasonable timeframe and for a reasonable cost. Over 100 insect mt genomes have been sequenced, representing 29 of the 30 orders, multiple suborders for a third of the orders, and many representatives of the mega-diverse orders. Genome rearrangements have been found in a third of the insect orders however these rearrangements diagnose groups of ordinal or lower rank. Sequence based phylogenetic hypotheses utilizing mt genomic data are a promising source of data on interordinal relationships however these studies are hampered by base compositional biases, unequal rates of nucleotide substitution across groups and other long-branch effects. Available data from the field of insect mitogenomic phylogenetics is reviewed and future directions in this research outlined.

> Key words

Mitochondria, genomics, phylogenetics, Insecta, interordinal relationships, gene rearrangements.

1. Introduction

The quest to understand insect interordinal relationships is one of the oldest and most important in systematic entomology, without which our conclusions regarding what drove the evolution of the most diverse group of life is necessarily flawed. Because of the difficulties in homologizing structures across the vast diversity which makes up extant insect faunas, considerable hope has been invested in alternative approaches to inferring phylogeny of which molecular biology is probably the most conspicuous. Only a comparatively small number of "standard" genes, including the nuclear ribosomal RNA genes (18S and 28S), some of the histone subunits (principally H3), portions of the mitochondrial (mt) genes (cox1, cox2, cytB, 16S and 12S) and a few developmental genes such as wingless (Wg) and Hox (Hx) are sufficiently conservative to be readily sequenced and compared across all orders. To date these genes have yet to satisfactorily resolve interordinal relationships. While new markers are being pursued to improve phylogenetic resolution between orders, at present none have yet proven to be the magic bullet for which we have all hoped.

One potential source of phylogenetic information which we have extensively investigated is the mitochondrial genome. With 37 genes – 13 protein coding, 2 ribosomal RNA and 22 transfer RNAs – and usually 14–17,000 base pairs in size, the mitochondrial genome of Metazoa is the smallest known genome, making it technically tractable to sequence it in its entirety, but still an order of magnitude larger than most of the single genes used in current insect phylogentic analyses (Saccone 1999; Caterino et al. 2000). Two technical features of the mt genome make its sequencing routine across insects. First, a legacy of the heavy utilization of individual mt genes in insect systematics provides an array of relatively conserved primers which can be used across many insect orders to provide preliminary data on individual genes (e.g. Simon et al. 1994). Secondly, the circular nature of the mt genome means that by amplifying between genes

previously sequenced using standard primers, one can readily sequence the entire genome within a short period of time. Thus, even small amounts of preliminary data can be transformed into whole genome sequences quite readily. This capacity to sequence genomes in their entirety makes the mt genome the largest set of homologous genes which can be compared across animal taxa and thus the largest piece of molecular data which can be readily used in comparisons of gene order or for sequence based phylogenetic work.

Two approaches to phylogenetic utilization of the mt genome have been proposed, uncovering shared genome rearrangements and the use of the whole genome in sequence based phylogenies (BOORE & BROWN 1998; ROKAS & HOLLAND 2000). Some of the first insect genomes to be sequenced, *Apis mellifera* (Hymenoptera) (CROZIER & CROZIER 1993), *Locusta migratoria* (Orthoptera) (FLOOK et al. 1995a), and *Heterodoxus macropus* (Phthiraptera) (SHAO et al. 2001) have moderate to extreme gene order rearrangements relative to the arthropod ground-plan suggesting that the use of "genome morphology" (DOWTON et al. 2002) would be a viable approach to resolving deep nodes in insect evolution. Subsequent sequencing effort has not supported this notion and has lead to an increasing emphasis on using the mt genome sequence in phylogenetic reconstruction. Mt genome phylogenies have ranged from resolving strains within *Drosophila simulans* (BALLARD 2000a) to phylogenies of all arthropods (NARDI et al. 2003a; CAMERON et al. 2004). These phylogenies have produced some remarkable results such as inferring Phthiraptera + Hymenoptera (NARDI et al. 2003a), the polyphyly of Hexapoda (NARDI et al. 2003a; BAE et al. 2004) and Orthoptera + Endopterygota to the exclusion of Paraneoptera (STEWART & BECKENBACH 2003). These results have highlighted the need for rigorously evaluating the phylogenetic behavior of mt genomes to increase confidence that results obtained from this marker are reflective of evolutionary history rather than analytical artifacts or patterns of inheritance that do not reflect the "true" phylogeny.

Whilst the capacity to quickly and cheaply sequence insect mt genomes is now routine, questions remain concerning how best to analyze these data. Over 100 insect mitochondrial genomes have now been sequenced; more than half by the authors of this paper. Phylogenetic coverage is excellent with representatives sequenced for all orders except Zoraptera, representatives of multiple principal subgroups available for 11 orders and wide diversities available within the megadiverse orders Diptera, Lepidoptera, Coleoptera, Hymenoptera and Hemiptera. Finally sufficient data has become available to attempt deep level phylogenies of insects, which are also comprehensive in their coverage, using mt genome data. In the present paper we therefore intend to present the state of the field of mt genomics as it relates to questions of deep level insect relationships, to review some of the approaches which have been taken to this question, and the difficulties which still remain.

2. Genome rearrangements

Gene rearrangements have been used to generate phylogenetic trees since the 1930's (see DOBZHANSKY 1944). It is usually assumed that rearrangements are rare because of the requirement for two or three chromosomal breaks, unique as it seems unlikely that identical chromosome breaks would occur in independent lineages, and irreversible. These features would appear to make rearrangements ideal cladistic markers. Unfortunately, mt genome rearrangements have not lived up to early promise as useful phylogenetic markers for the resolution of interordinal relationships. The majority of insects have the same plesiomorphic gene arrangement that is shared by the Pancrustacea (BOORE et al. 1998) (Fig. 1). To date derived gene orders have been recorded for only 11 of the 29 orders for which data is available. Of these 11 orders the rearrangements found have been diagnostic for groups at or below the ordinal level. The only possible exception is for the Psocodea (Phthiraptera + Psocoptera), however with recent evidence for the polyphyly of the lice (JOHNSON et al. 2004) it appears likely that Psocodea will soon sink from superordinal to ordinal rank and so possible gene rearrangements linking these two groups are correspondingly more likely to be diagnostic at the ordinal level.

Autapomorphic rearrangements have been found in four orders, Embioptera, Thysanoptera, Strepsiptera and Trichoptera (SHAO & BARKER 2003; CARAPELLI et al. 2006; Cameron unpubl. data) for which only a single species has been sequenced, so it is currently impossible to infer the taxonomic levels at which these would be diagnostic. Rearrangements have also been found that are diagnostic for major divisions within orders. Within the Diptera, Culicidae (mosquitoes) share a diagnostic tRNA rearrangement (Arg–Ala rather than Ala–Arg) (BEARD et al. 1993; MITCHELL et al. 1993), which is not found in brachyceran flies. However, the presence or absence of this rearrangement within other "nematoceran" groups has not been established. Caelifera (Orthoptera) share a derived tRNA rearrangement (Asp–Lys), whereas the Ensifera have the plesiomorphic arrangement (Lys–Asp). The derived arrangement in caeliferans was first noted by FLOOK et al. (1995b) in a survey of just five species and we have confirmed it in a wider survey of orthopteran mitochondrial genomes (Acrididae, Pygomorphidae,

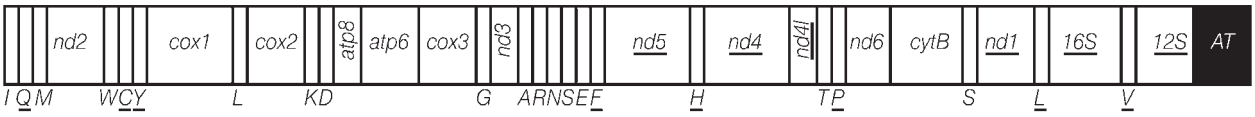

Fig. 1. Ground plan mitochondrial genome arrangement for Pancrustacea including the majority of insect orders (18 of 29 which have been examined). While the mitochondrial genome is circular, it has been linearised for this figure by cutting between the AT-rich region and tRNA-Ile. With the exception of the tRNAs, gene size in the genome is proportional to their size in this figure. Underlined genes are encoded on the minority or N-strand, the remainder are encoded on the majority or J-strand. Genes are labeled with their abbreviated names, nd for nicotinamide dehydrogenase subunits, cox for cytochrome oxidace c subunits, cytB for cytochrome oxidase b, atp for ATP transferase F0 subunits, 16S and 12S for the large and small ribosomal RNA subunits respectively, and AT for the AT-rich or putative control region. Transfer RNA (tRNAs) genes are each labeled with the single letter code for the amino acid corresponding to their anticodon.

Eumastacidae and Catantopidae). It is interesting that this rearrangement is also found in the honeybee (*Apis*) and is one of the few clear examples of a homoplastic mt genome rearrangement. Rearrangements appear to be common in Hymenoptera (Crozier & Crozier 1993; Castro & Dowton 2005; Dowton 1999; Dowton & Austin 1999; Dowton et al. 2001, 2003) and there appears to be considerable variation in the extent of rearrangements between the Symphyta, Parasitica and Aculeata. However, it is currently impossible to generalize about patterns of mt genome rearrangement across the order, as only five complete genomes are available with the majority of information coming from partial genome sequences of the major tRNA blocks. A possible rearrangement synapomorphy for the Ditrysia (Lepidoptera) is the tRNA arrangement Met–Ile but the occurrence of this rearrangement within Ditrysia has yet to be determined.

There is a single rearrangement which is potentially synapomorphic at the ordinal level. In Neuroptera the tRNA arrangement Cys–Trp is found rather than the pancrustacean groundplan condition Trp–Cys. This derived arrangement is not found in either of the other members of the Neuropterida (Megaloptera or Raphidioptera) but within Neuroptera is found in Mantispoidea and Myrmeleontoidea. Recent molecular phylogenies of the Neuroptera (Haring & Aspöck 2004) suggest that these two superfamilies are representative of two of the three major divisions within Neuroptera, Hemerobiiformia and Myrmeleontiformia respectively. As data is not yet available for the third group, Nevrorthiformia, which is currently considered the sister group of the remaining Neuroptera, it is possible that this putative synapomorphy is for a more restricted group than the entire order Neuroptera.

By far the greatest incidence of gene rearrangements in insects occurs within the Paraneoptera (= Acercaria = hemipteroids). Shao & Barker (2003) found that the mt genome of a thysanopteran, *Thrips imaginis*, was highly rearranged. However, as noted above, no other thrips have been sequenced to date and it is unknown for what group this arrangement is diagnostic. Rearrangement rate appears variable within Hemiptera. Sequenced members of Heteroptera and Auchenorhyncha all share the plesiomorphic arrangement (Dotson & Beard 2001; Stewart & Beckenbach 2005). Within Sternorrhyncha, the superfamilies Psylloidea and Aphidoidea have the ground plan condition, whereas the Aleyrodidae possess a rearrangement diagnostic for the family and an additional four derived positions for the block of protein coding genes from cox3 to nadh3. Each is diagnostic for subgroups within Aleyrodidae (Thao et al. 2004). In Psocodea, genome rearrangements appear to be common across both Psocoptera and Phthiraptera s.l., with no representative currently sequenced that possesses the ground plan gene order. The tRNA rearrangement Met–Cys (these genes are separated by approx. 1000 bp in the ground plan) may be diagnostic for Psocodea as it has been found in all Psocomorpha, early branching members of the Trogiomorpha and in Amblycera. It is not found in the derived trogiomorph Lepidopsocidae (Shao et al. 2003) or in the Ischnocera (Covacin et al. 2006; Cameron unpubl. data) probably due to subsequent movement of these tRNAs to new locations. The psocopteran suborder Psocomorpha has a diagnostic rearrangement of the protein coding genes nadh3 and nadh5 and associated tRNAs. Gene order is variable within the suborder Trogiomorpha, with early branching members possessing the putative psocodean ground plan condition whereas derived members have additional rearrangements of tRNA and protein coding genes including some putative reversals (Johnson et al. 2004; Shao et al. 2003). Gene order is highly variable between the two louse suborders, Amblycera and Ischnocera. There are few derived gene boundaries shared with any of the psocopteran suborders, with each other or in the case of Ischnocera between families within this suborder (Shao et al. 2001; Covavin et al. 2006). It is precisely this variability which may be harnessed to improve systematic understanding within the louse suborders, a group which has proven historically very difficult to resolve (Johnson & Clayton 2003).

Thus in conclusion it appears clear that gene rearrangement synapomorphies will not contribute much to

resolving insect interordinal relationships. While the possibility exists of finding additional rearrangements within these orders, it is doubtful that such findings would greatly aid our understanding of interordinal relationships. In most cases taxon sampling has focused on capturing the diversity within each order, such as by sequencing representatives of each suborder or major clades, or has already included the earliest branching representatives of that order e.g. *Mastotermes* for Isoptera and *Timema* for Phasmatodea. The possibility that those species which have plesiomorphic genome arrangements actually represent secondary reversions to the insect ground plan is therefore low and it is doubtful that additional interordinal synapomorphies will come to light. In contrast the potential for using genome rearrangements in understanding intraordinal relationships appears bright. Rearrangements have been found in over a third of the insect orders and in those orders where multiple representatives have been examined the phylogenetic signal in rearrangements is often very strong. However, rearrangements are not even remotely clock-like in their evolution across insects and so wide-ranging surveys of genome arrangements are necessary to quantify their potential usefulness within any particular order.

3. Sequence based phylogenetics

The use of the sequence data within mitochondrial genes is an increasingly promising direction in the use of mitochondrial genome data in insect systematics. In general, these studies align some or all of the mitochondrial genes individually, concatenate the resulting alignments and perform phylogenetic analyses on the resulting matrices using normal approaches to molecular phylogenetics such as parsimony, likelihood or Bayesian analyses. In this regard sequence based phylogenies are not substantially different from other molecular phylogenies except that they contain fewer taxa/more data rather than the more taxa/less data approach pursued by most workers. Additionally, unique problems related to the evolution of mitochondrial sequence itself, in particular variation in nucleotide composition and mutational rate, pose special challenges to analyses of this type (CAMERON et al. 2004).

Some of the earliest studies pioneering the use of whole genome data for insect phylogenetics had deliberately limited scope e.g. a phylogeny of the Diptera (LESSINGER et al. 2000), of closely related members of the melanogaster species subgroup of *Drosophila* (BALLARD 2000b) or even between strains of *Drosophila simulans* (BALLARD 2000a). This trend very quickly swung to the other extreme, examining interordinal relationships within the context of analyses of the phylogeny of all arthropods. NARDI et al.'s (2001, 2003a) studies examining the placement of Collembola with wider hexapod and arthropod phylogeny were the leaders in this field. Subsequent studies have tended to follow this approach. For example, STEWART & BECKENBACH (2003), FRIEDRICH & MUQIM (2003) and BAE et al. (2004) examined the relationships of Coleoptera; CASTRO & DOWTON (2005) examined the position of Hymenoptera within Endopterygota; and KIM et al. (2005) examined the position of Orthoptera within the Insecta. The next generation of analyses will directly tackle interordinal phylogenetics as genomic data becomes available for the majority of orders.

There has also been spirited debate on methodological approaches to mitogenomic phylogenetics. Difficulties with aligning genes and concerns about mutational saturation have lead some to argue for the use of amino acid sequences to the exclusion of nucleotide sequences (NARDI et al. 2003a,b). Others have argued for the use of amino acid sequence as an aid for the alignment and analysis of nucleotides (CAMERON et al. 2004), while still others have argued for reductive coding schemes such as purine/pyrimidine (R/Y) coding to address issues of nucleotide compositional bias (DELSUC et al. 2003).

The second major methodological issue concerns data exclusion strategies, with *a priori* exclusion of genes that are assumed to be noisy or misleading. Whilst some studies have excluded as many as 9 of the 13 protein coding genes (NARDI et al. 2003a,b), the only direct attempts to examine the effects of gene exclusion and to quantify gene quality have found that exclusion reduced phylogenetic signal and that there is relatively little difference between the degree of homoplasy evident in the protein coding genes (CAMERON et al. 2004). The final contentious issue concerns taxon exclusion. Highly variable nucleotide composition (15–60 % GC content), mutational rate heterogeneity and the possibly confounding effect of genome rearrangements can result in extreme divergences of genome sequences between different groups. Inclusion of highly divergent data in a phylogenetic analysis can violate some of the assumptions upon which evolutionary models are constructed and result in grossly incorrect topologies. Exclusion of problematic taxa has been the most popular method used for dealing with this issue, although it cannot be applied when the problematic taxa are the question of interest. The sequencing of alternative, potentially less divergent taxa, as well as developments in analytical approaches are needed to address these concerns.

Now that we have a comprehensive taxon sample which includes 29 of the 30 insect orders, with multiple representatives of many orders, analyses to produce a complete interordinal phylogeny of the insects are in progress (Cameron et al. in prep.). Rather than

preempt that analysis, we will not present a tree here but will instead comment on some of the more strongly supported results and the difficulties which we are encountering in this ongoing analysis.

Dealing first with technical issues, our favoured approach is a total evidence strategy which combines as much data from the mt genome as possible. For that reason, we have concentrated on the use of nucleotide sequences guided by amino acid alignments as a means of producing more reasonable nucleotide alignments. Our analyses have found that alternative alignment strategies, such as multiple alignment of the nucleotides with Clustal or direct optimization approaches (= optimization alignment, Wheeler 1996), have not resulted in reasonable topologies. This may be due to the considerable evolutionary distances between taxa combined with inconsistencies in the definitions of gene location which have resulted in variation of gene length (see below). Further, we favour the analysis of nucleotides over the various reductive coding schemes (such as amino acids or R/Y coding) as we have found that these approaches almost invariably result in a loss of signal for deep nodes (Cameron et al. 2006). As previously reported (Cameron et al. 2004), we have found no evidence to support the *a priori* exclusion of any of the protein coding genes. Further there is a compelling case to be made for the inclusion of the ribosomal and transfer RNA genes in future analyses. The ribosomal RNA genes have only been used in three studies to date (Stewart & Beckenbach 2003; Bae et al. 2004; Kim et al. 2005) and their contribution to the combined analysis assessed in only one (Stewart & Beckenbach 2003). To date the insect ribosomal genes have not been annotated with regard to explicit biological models of the structure of the mature molecule. The large subunit (16S) is defined as the homologous gene segment between flanking tRNA genes; no assessment of indels at either the 5' or 3' ends has been made. The small subunit (12S) is flanked on the 5' end by the control region in most insects, and the boundary between these two regions has not been defined for more than a handful of model organisms (such as *Drosophila*). In both cases, it is clear that alignments of the ribosomal genes potentially include non-functional portions of the genome, and so are not truly homologous comparisons. In contrast, the transfer RNA genes are well defined as their annotation is based on structural criteria of the mature molecules. They also appear to be evolutionarily conservative. Kim et al. (2005) found that the tRNAs had considerably lower pairwise distances between taxa than either the protein coding or ribosomal RNA genes. In our preliminary analyses, we have similarly found that the tRNAs are much less homoplastic than the protein coding genes, and it appears that their inclusion in future phylogenetic analyses is to be highly recommended.

Different analytical approaches yield different toplogies. Simple heuristic searches under parsimony produce a backbone phylogeny for which many nodes are clearly incorrect (e.g. Dermaptera + Hemiptera as the earliest pterygote branch, Ephemeroptera + Plecoptera). Although many clearly incorrect nodes collapse under bootstrap resampling, so do many nodes for which there is considerable corroboration from morphological and other molecular data. Thus Dicondylia, Pterygota and Orthoptera, which are all found by the heuristic search, do not receive significant bootstrap support. Bootstrap support is known to be highly sensitive to overall levels of homoplasy within an analysis and nodal support declines with increasing homoplasy even for datasets with the same level of signal (= informative sites). Mt genome sequences are known to be highly homoplastic, CIs averaging around 0.2 in most analyses, and so alternative ways of measuring support are being sought. A second problem is the existence of clear instances of long-branch attraction/non-stationarity either due to heightened rates of nucleotide substitutions or base compositional bias. Several of the orders with highly divergent genomes, Embioptera, Hymenoptera, Thysanoptera, Psocoptera and Phthiraptera, always group together and usually with high bootstrap support. The grouping of Dermaptera + Hemiptera at the base of the pterygotes is probably due to base compositional effects as genomes from these two orders and the apterygotes have the highest GC contents yet found. Methods of dealing with these long-branch effects such as mild topological constraints may be necessary before more reasonable trees are forthcoming.

That said there are several findings within our parsimony analyses which are consistent with many previous studies. All orders were found to be monophyletic, except the Hymenoptera. Many of the traditionally supported interordinal clades were also recovered with high support: Dictyoptera, Endopterygota, and Amphiesmenoptera. In contrast other more recently proposed and/or less well supported groupings were not supported: Xenonomia (= Mantophasmatodea + Grylloblattodea) and Eukinolabia (= Embioptera + Phasmatodea) (sensu Terry & Whiting 2005), Orthopteroidea (= Phasmatodea + Orthoptera), Mecopterida (requires the inclusion of the Neuropterida and exclusion of Amphiesmenoptera) and Coleoptera + Neuropterida. Interesting novel hypotheses which do receive significant support include a sister-group relationship between Mantophasmatodea and Phasmatodea and a monophyletic Zygentoma, contrary to the more widely accepted notion of zygentoman paraphyly with *Tricholepidion* as the sister-group to the remaining dicondylian insects.

By contrast, Bayesian analyses produce more resolved trees, more nodes are significantly supported and

there appears to be less influence from long-branch attraction. Nevertheless it is important to keep in mind that these analyses are not wholly immune to long-branch attraction, however the significance with which they are supported is greatly reduced compared to parsimony analyses. There is still a clearly erroneous clade composed of Psocoptera, Embioptera, and Hymenoptera but significant support for relationships between these orders is no longer present (posterior probability < 90 %). Dermaptera still migrates towards the base of the tree (between Apterygota and Pterygota), but again this is not significantly supported. Ephemeroptera appears as the well-supported sister group of Neoptera but Odonata is the weakly supported sister group of Orthoptera as in the parsimony analysis. This may be due to base composition or be indicative of a secondary origin of the palaeopterous condition in odonates necessitated by the move to direct wing musculation. As in the parsimony analyses many interordinal nodes are well supported: Dicondylia, Pterygota, Dictyoptera and Amphiesmenoptera. Intraordinal relationships are the same as those recovered by parsimony and consistent with current notions of the phylogeny of Diptera, Coleoptera and Lepidoptera (those orders for which significant intraordinal sampling is available). Those interordinal hypotheses not receiving support are the same that conflict with by the parsimony analysis: Xenonomia, Eukinolabia, Orthopteroidea, Coleoptera + Neuropterida and Mecopterida. In this instance, however, relationships within Endopterygota differ markedly from the parsimony analysis; no interordinal relationships receive significant support except Neuroptera + Megaloptera (in line with suggestions by HARING & ASPÖCK 2004 and contrary to traditional hypotheses on relationships within the Neuropterida) and Amphiesmenoptera. So while correcting for some of the deficiencies of the parsimony analyses, the Bayesian analyses have fallen prey to others. In particular, the most poorly resolved areas differ in the two – Polyneoptera in parsimony and Endopterygota in Bayesian. Reconciling the strengths and weaknesses of the two approaches will, however, provide tremendous insights into the interordinal relationships of the insects.

4. Conclusions and looking to the future

Mitochondrial genomics has come of age as a data source for the deep level phylogenetics of insects (interordinal and resolving major clades within orders). Genomes have been sequenced for representatives of 29 of the 30 orders and for many orders multiple representatives are available. The existence of a comprehensive data set which surveys the diversity of the Insecta is the first step. The testing of various approaches to phylogenetic reconstruction, understanding the biases or complications inherent in the data, and working towards a comprehensive phylogeny based on these data, is the second.

Genome rearrangements will not serve as useful phylogenetic markers for resolving interordinal relationships due to a lack of variation in gene order. Eighteen of the 29 orders which have been surveyed have no rearrangements and in the remaining 11 orders rearrangements occur sporadically, and are often minor (such as local movements of single tRNAs). Rearrangement synapomorphies are however probably underappreciated as a potential data source at subordinal levels, their rarity enhancing their value as markers due to the reduced likelihood of homoplasy.

In contrast, using the mt genome in sequence based phylogenetic studies has considerable promise at both inter- and intraordinal scales. The ability of mt genome data to consistently recover monophyletic orders, notably including Coleoptera (which has not been found to be monophyletic in any previous molecular analysis), to recover those interordinal relationships for which the weight of evidence for monophyly is very strong (e.g. Pterygota, Endopterygota, Dictyoptera, Amphiesmenoptera), to point to novel relationships for which evidence has been equivocal (e.g. Mantophasmatodea + Phasmatodea) and finally to challenge some poorly supported relationships which continue to receive considerable attention (e.g. Orthopteroidea, Coleoptera + Neuropterida) are all indicative of its utility in deep level insect phylogenetics. Challenges remain to be overcome in the analysis of mt genome data, particularly the issues of long-branch attraction/non-stationarity that are currently so prevalent; the development of novel analytical approaches and software will hopefully achieve this. The sequencing of additional genomes, with a view to increasing diversity within orders, will provide alternatives to some of the highly divergent genomes which are currently available. Indeed total numbers of insect mt genomes looks set to double by the end of the decade. Finally mt genome data needs to be more fully integrated with other data sources. The technical demands of genome sequencing have tended to result in specialized "mitochondrial labs" disconnected from wider efforts within the entomological systematics community. As a result taxonomic consistency between mitochondrial and nuclear sequencing projects and between exemplar morphological datasets has, to date, been poor. Rectifying this will launch ever larger data sets at the problems of insect interordinal phylogenetics and lead us to a final conclusion of the quest for a unified hypothesis regarding the evolution of life's most successful players.

Acknowledgements

This work was supported by NSF grants DEB 0120718 (MFW) and DEB 0444972 (SLC, MFW).

References

Bae, J.S., I. Kim, H.D. Sohn & B.R. Jin 2004. The mitochondrial genome of the firefly *Pyrocoelia rufa*: complete genome sequence, genome organisation and phylogenetic analysis with other insects. – Molecular Phylogenetics and Evolution **32**: 978–985.

Ballard, J.W.O. 2000a. Comparative genomics of mitochondrial DNA in *Drosphila simulans*. – Journal of Molecular Evolution **51**: 64–75.

Ballard, J.W.O. 2000b. Comparative genomics of mitochondrial DNA in members of the *Drosphila melanogaster* subgroup. – Journal of Molecular Evolution **51**: 48–63.

Beard, C.B., D.M. Hamm & F.H. Collins 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. & W.M. Brown 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. – Current Opinion in Genetics and Development **8**: 668–674.

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

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

Cameron, S.L., K.B. Miller, C.A. D'Haese, M.F. Whiting & S.C. Barker 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.

Carapelli, A., L. Vannini, F. Nardi, J.L. Boore, L. Beani, R. Dallai & F. Frati 2006. The mitochondrial genome of the entomophagous endoparasite *Xenos vesparum* (Insecta: Strepsiptera). – Gene **376**: 248–259.

Castro, L.R. & M. Dowton 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.

Caterino, M.S., S. Cho & F.A.H. Sperling 2000. The current state of insect molecular systematics: A thriving tower of babel. – Annual Review of Entomology **45**: 1–54.

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

Crozier, R.H. & Y.C. Crozier 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organisation. – Genetics **133**: 97–117.

Delsuc, F., M.J. Phillips & D. Penny 2003. Comment on "Hexapod origins: Monophyletic or Paraphyletic?" – Science **301**: 1482.

Dobzhansky, T. 1944. Chromosomal races in *Drosophila pseudoobscura* and *D. persimilis*. Pp. 47–145 *in*: T. Dobzhansky & C. Epling (eds.), Contributions to the Genetics, Taxonomy and Ecology of *Drosophila pseudoobscura* and its relatives. – Carnegie Inst. of Wash. Publ. 554.

Dotson, E.M. & C.B. Beard 2001. Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. – Insect Molecular Biology **10**: 205–215.

Dowton, M. 1999. Relationships among the cyclostome braconid (Hymenoptera: Braconidae) subfamilies inferred from a mitochondrial tRNA gene rearrangement. – Molecular Phylogenetics and Evolution **11**: 283–287.

Dowton, M. & A.D. Austin 1999. Evolutionary dynamics of a mitochondrial rearrangement 'hotspot' in the Hymenoptera. – Molecular Biology and Evolution **16**: 298–309.

Dowton, M., R. Belshaw, D.L.J. Quicke & A.D. Austin 2001. Simultaneous molecular and morphological analysis of braconid relationships (Insecta: Hymenoptera: Braconidae) indicates independent mt-tRNA gene inversions within a single wasp family. – Journal of Molecular Evolution **54**: 210–226.

Dowton, M., L.R. Castro & A.D. Austin 2002. Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome 'morphology'. – Invertebrate Systematics **16**: 345–356.

Dowton, M., L.R. Castro, S.L. Campbell, S.D. Bargon & A.D. Austin 2003. Frequent mitochondrial gene rearrangements at the hymenopteran nad3nad5 junction. – Journal of Molecular Evolution **56**: 517–526.

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

Flook, P.K., C.H.F. Rowell & G. Gellissen 1995b. Homoplastic rearrangements of insect mitochondrial tRNA genes. – Naturwissenschaften **82**: 336–337.

Friedrich, M. & N. Muqim 2003. Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castanaeum*. – Molecular Phylogenetics and Evolution **26**: 502–512.

Haring, W. & U. Aspöck 2004. Phylogeny of the Neuropterida: A first molecular approach. – Systematic Entomology **29**:415–430.

Johnson, K.P. & D.H. Clayton 2003. The biology, ecology, and evolution of chewing lice. Pp. 449–476 *in*: R.D. Price, R.A. Hellenthal, R.L. Palma, K.P. Johnson & D.H. Clayton (eds.), The Chewing Lice: World Checklist and Biological Overview. – Illinois Natural History Survey Special Publication 24.

Johnson, K.P., K. Yoshizawa & V.S. Smith 2004. Multiple origins of parasitism in lice. – Proceedings of the Royal Society of London B **271**: 1771–1776.

KIM, I., S.Y. CHA, M.H. YOON, J.S. HWANG, S.M. LEE, H.D. SOHN & B.R. JIN 2005. The complete nucleotide sequence and gene organisation of the mitochondrial genome of the oriental mole cricket *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). – Gene **353**: 155–168.

LESSINGER, A.C., A.C. MARTINS JUNQUEIRA, T.A. LEMOS, E.L. KEMPER, F.R. DA SILVA, A.L. VETTORE, P. ARRUDA & A.M.L. AZEREDO-ESPIN 2000. The mitochondrial genome of the primary screwworm fly *Cochliomyia hominovorax* (Diptera: Calliphoridae). – Insect Molecular Biology **9**: 521–529.

MITCHELL, S.E., A.F. COCKBURN & J.A. SEAWRIGHT. 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A.: complete nucleotide sequence and genome organisation. – Genome **36**: 1058–1073.

NARDI, F., A. CARAPELLI, P.P. FANCIULLI, R. DALLAI & F. FRATI 2001. The complete mitochondrial DNA sequence of the haxapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. – Molecular Biology and Evolution **18**: 1293–1304.

NARDI, F., G. SPINSANTI, J.L. BOORE, A. CARAPELLI, R. DALLAI & F. FRATI 2003a. Hexapod origins: monophyletic or paraphyletic? – Science **299**: 1887–1889.

NARDI, F., G. SPINSANTI, J.L. BOORE, A. CARAPELLI, R. DALLAI & F. FRATI 2003b. Response to comment on "Hexapod origins: monophyletic or paraphyletic?" – Science **301**: 1482.

ROKAS, A. & P.W.H. HOLLAND 2000. Rare genomic changes as a tool for phylogenetics. – Trends in Ecology and Evolution **15**: 454–459.

SACCONE, C., C. DE GIORGI, C. GISSI, G. PESOLE & A. REYES 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. – Gene **238**: 195–209.

SHAO, R. & S.C. BARKER 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.

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

SHAO, R., M. DOWTON, A. MURRELL & S.C. BARKER 2003. Rates of gene rearrangements and nucleotide substitution are correlated in the mitochondrial genomes of insects. – Molecular Biology and Evolution **20**: 1612–1619.

SIMON, C.F., A. FRATI, B. BECKENBACH, B. CRESPI, H. LIU & P. FLOOK 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.

STEWART, J.B. & A.T. BECKENBACH 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. & A.T. BECKENBACH 2005. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). – Genome **48**: 46–54.

TERRY, M.D. & M.F. WHITING 2005. Mantophasmatodea and phylogeny of the lower neopterous insects. – Cladistics **21**: 240–257.

THAO, M.L., L. BAUMANN & P. BAUMANN 2004. Organisation of the mitochondrial genomes of whiteflies, aphids and pysllids (Hemiptera, Sternorrhyncha). – BMC Evolutionary Biology **4**: 25.

WHEELER, W.C. 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics? – Cladistics **12**: 1–9.