Insect Mitochondrial Genomics: Implications for evolution and phylogeny.

Submitted to *Annual Review of Entomology* as an invited review.

Stephen L. Cameron

Earth, Environment & Biological Sciences School, Science & Engineering Faculty, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, AUSTRALIA.

*Corresponding author:* Stephen Cameron (sl.cameron@qut.edu.au); phone: +61 7 3138 2869; fax: +61 7 3138 2330.
Abstract

The mitochondrial (mt) genome is the most extensively studied genomic system in insects, outnumbering nuclear genomes by tenfold, and representing all orders versus very few. Phylogenomic analysis methods have been extensively tested, identifying compositional bias and rate variation, both within and between lineages, as the principle issues confronting accurate analyses. Major studies at both inter- and intraordinal levels have contributed to our understanding of phylogenetic relationships within many groups. Genome rearrangements are an additional data type data for defining relationships, with rearrangement synapomorphies identified across multiple orders and at many different taxonomic levels. Hymenoptera and Psocodea have greatly elevated rates of rearrangement offering both opportunities and pitfalls for identifying rearrangement synapomorphies in each group. Finally, insects include model systems for studying “aberrant” mt genomes, including truncated tRNAs and multichromosomal genomes. Greater integration of nuclear and mt genomic studies, is the challenge for furthering our understanding of insect genomic evolution.

Keywords: Organelle genomics; Phylogenomics; Insect Phylogenetics; Genome Rearrangements; tRNA editing; Minichromosomes;

Introduction

Due to the extensive use of its individual genes across a wide range of studies, the mitochondrial (mt) genome (= mitogenome or mtgenome) has had an outsized impact on entomological genetics. Initially mt genes were widely used due to the limited number of alternative, nuclear-encoded genes which could be amplified with reliable, near-universal primers across insects (with the exception of the ribosomal RNA genes) (31). Indeed, despite several large scale efforts (e.g. 82, 94, 128), the development of reliable protocols for sequencing nuclear protein coding genes in insects has been so slow that mt genes still predominate in insect molecular systematic datasets. The recent focus of the DNA-barcoding community on the mt encoded cytochrome c oxidase subunit 1 (= cox1, CO1) as a near-exclusive data source for species identification and delimitation (57) has further increased the rate of mt gene sequencing (93). Collectively these forces have resulted in such a flood of data that mt genes are now by orders of magnitude the most common insect sequences on Genbank and understanding the particular dynamics of mt genetic evolution is vital to the appropriate use of this enormous data resource.

Parallel to these general developments, has been a rapid increase in available mt genome data, both for animals in general and insects in particular. At its simplest level the sequencing of whole mt genomes can inform subsequent studies utilising individual mt genes such as by improving primer design or allowing an expanded choice of target genes (e.g. 46). Indeed, such expansion of available genetic resources for a particular species is the most commonly invoked reason for undertaking mt genome sequencing, if the “Introduction” sections of many papers are to be believed. There are however also many studies where mt genome sequencing itself is the “end” goal of the study rather than merely a “means” to some other data set. Whole mt genomes have been used for the same wide array of research goals as individual mt genes including molecular systematics (at both deep and shallow taxonomic scales), population genetics/phylogeography (e.g. 76), diagnostics (e.g. 87), and molecular evolutionary studies (e.g. 30, 100, 104). In addition, whole genome sequencing also allows the study of comparative
and evolutionary genomics questions such as the frequency and type of gene rearrangements (e.g. 24, 45), evolution of genome size etc (e.g. 106). The small size of the mt genome makes it a practical genome study system in insects which nuclear genome sequencing will not equal in the near future.

Since the first insect mt genome was published in 1985, a *Drosophila* appropriately or inevitably depending on your perspective (32), there has been a rapid accumulation of sequenced insect genomes, with representatives of all orders now available (Figure 1). This review will thus focus on the evolving methods for analysing insect mt genomes and empirical findings drawn from this data source. I will not explicitly address studies of single mt genes from insects as this vast field has been covered previously from several perspectives. The insights into mt evolution obtained from whole mt genomes, however, are valuable for the reliable use of single mt genes rather than uniquely applicable to whole genomes.

**Mitochondrial Genomes of Arthropods & Insects**

The mitochondrial genomics of arthropods has been extensively studied; with the complete genomes of over 600 species available, arthropods are second only to the vertebrates as the most studied metazoan phylum. Taxonomic coverage within arthropods is also now quite comprehensive, with multiple representatives sequenced for each of the four extant subphyla, 14 of the 15 classes and 65 of the 103 orders. Of the arthropods, insects are by far the most extensively studied group, representing approximately 80% of the arthropod mt genomes which have been sequenced. Insects are also the most comprehensively sampled at higher taxonomic levels; mt genomes are available from each of the 28 recognised orders (Table 1). Within most orders, mt genomes representing each of the major subordinal lineages are now available and representation at the family level is steadily improving. Data has accumulated rapidly with improved methods of PCR (reviewed in 18) and pyrosequencing (118). There are still a modest number of groups where representation is seriously deficient relative to diversity including dragonflies (Odonata), mt genomes for just 4 of 32 families, stoneflies (Plecoptera) 1 of 17, scale insects (Coccoidea) 0 of 20, and ditrysian moths (Lepidoptera) 17 of 109. However, there is now more than sufficient mt genome data available from insects to reliably draw conclusions about patterns and trends in the genomic evolution of this group.

The insect mt genome is a compact circular molecule typically 15-18 kb in size. It encodes 37 genes; 13 protein-coding genes (PCGs) that encode subunits from 4 of the 5 mt electron-transport chain complexes plus 2 ribosomal (rRNA) and 22 transfer (tRNA) RNA genes involved in the translation of the PCGs. This set of 37 genes is conserved across bilaterian metazoans, with only a few exceptions (e.g. loss of a small number of genes in some derived groups) (12). In addition to the genes there are a variety of non-coding structural features of the genome of which the largest is termed the “control region” as it contains both an origin of replication and transcription. The arrangement of the genes within the genome is also highly conserved. An ancestral arrangement of the 15 PCG and rRNA genes can be identified for the Bilateria (85), which is only slightly modified in the ancestral ecdysozoan and arthropod (16). The ancestral insect mt genome (Figure 2) differs from the ancestral arthropod only by the location of one tRNA gene (15). While significant departures from the ancestral insect mt genomes in terms of structure, gene content and gene arrangement have been recorded within insects, it is clear from the genomes now available that these “exceptions” are only found in highly derived portions of the insect tree of life.
Insect Mitochondrial Phylogenomics

By far the most widespread use of insect mt genomes is as a source of sequence data for phylogenetic analysis (what I term phylogenomics). To date, over 100 publications report insect phylogenies built using mt genome data and while too many are methodologically naive (e.g. use of a single, seemingly arbitrary inference method, partitioning approach, no discussion of potential biases etc), others extensively test the effectiveness of different analytical approaches. The potential biases in phylogenies based on insect mt genomes have thus been fairly well investigated and while none are unique to mt genomes, several are more extreme than occur in nuclear-gene phylogenetics. Collectively, these studies have addressed a wide range of taxonomic levels from single-species phylogeographic studies (e.g. 76), through intraordinal (4), to interordinal (122) or interclass relationships (19, 86). Representatives of almost all insect orders have been included in mt phylogenomics studies and there is substantial variation between orders in the intensity of analytical biases. Insect mt phylogenomics is not trivial, however, more than enough studies have now been conducted to clearly identify trends for the reliable use of this data source.

Decisions about data inclusion affect phylogenomics studies more acutely than single- or multilocus phylogenetics. The non-coding control region is excluded from almost all studies except phylogeographic studies of a single species (e.g. 46, 76). The remaining 37 genes have been used to varying degrees across previous studies. Some early analyses (e.g. 64, 86) determined the inclusion or exclusion of individual PCGs on the basis of the ratios of gaps and invariant sites. These approaches, however, involve arbitrary cut offs (e.g. 10% gaps good, 10.1% gaps bad) and are highly dependent on taxon selection, as the inclusion of even a single highly divergent species could result in the rejection of all genes (19). In practise almost all studies have included all 13 PCGs, as they comprise \( \approx 75\% \) of the genic sequence, whereas the inclusion of the 2 rRNA (\( \approx 15\% \)) and 22 tRNA (\( \approx 10\% \)) genes has been more variable. The inclusion of these genes has, however, been shown to be beneficial each time it has been tested resulting in at least improved nodal confidence (e.g. in Diptera, 21) to at best stabilising otherwise highly variable backbone relationships (e.g. in Neuroptera, 23). In no instances have they resulted in artefactual relationships being supported when analyses of the PCGs alone were more congruent with independent phylogenetic estimates. There is no justification for the exclusion of rRNA or tRNA genes from insect phylogenomics analyses.

A related issue is the exclusion of portions of genes on either an arbitrary (e.g. third codons of PCGs, loop regions of RNAs) or an algorithmic basis (e.g. by software such as GBlocks, 29). Depending on the taxonomic scale of the analysis the inclusion of third codon positions may result in either serious artefacts (e.g. within Dictyoptera, 25) or are the source of the majority of the phylogenetic signal (e.g. within calliphorid blowflies, 87). Third codon positions are most strongly affected by nucleotide compositional bias and skew (see below) and the removal of third codons may be the only method of effectively dealing with these issues. Given the variable in their phylogenetic performance it should be standard practice to assess the effect on topology and nodal support of inclusion vs. exclusion of third codons by replicate analyses within each study.
Insect mt phylogenomics has largely been spared the tedious debates about alignment methodologies which dogged other areas of deep-level insect phylogenetics over the past two decades (e.g. 64, 115). Alignment of PCGs has been largely a twostep process: initial alignment of amino-acid sequences, followed by translation back to DNA sequences so as to maintain coding frames. Within many orders PCG alignments are almost trivial as there is very limited gene length variability. RNA genes can be aligned with reference to secondary structures or by standard pair-wise alignment methods, however, on the one occasion it has been tested (23) it had no real effect on topology or nodal support. Again for most intraordinal analyses, RNA alignments approach the trivial with indels readily identifiable. Historically inconsistent annotation standards initially created much of the observed variability in gene length, however this is declining as multiple genomes are sequenced for each order. Variable alignment standards are not a major issue within insect mt phylogenomics.

Many early analyses produced clearly artefactual results that were attributed to “long-branch like effects” which in this instance typically meant compositional biases and/or model misspecification. Mt genomes in general, and insect ones in particular, display strong base compositional bias (A+T ≠ G+C) (100). The compositional bias of insect mt genomes varies significantly both across (e.g. A+T% ranges from 64% in termites to 86.7% in bees; 25, 109) and within orders (e.g. beetles, 65 – 78 A+T%; 108). Such compositional heterogeneity is a violation of the stationarity assumption of the widely used, time-homogenous models of nucleotide substitution (GTR, HKY85, F81 etc) (50). Including taxa that have independently evolved significantly biased nucleotide compositions in a single analysis can result in the artefactual groups (48). For example, within beetles both subordinal and superfamily relationships were incorrectly recovered using standard analytical methods which didn’t correct for base compositional heterogeneity (108, 110). In other instances a directional shift in compositional bias appears to be a derived feature of the clade, for instance in termites A+T% is 5.5% lower than their nearest living relative Cryptocercus, and 6.3% lower than other dictyopterans, reinforcing the monophyly of the group (25).

A second violation of model assumptions found in mt genomes is among-site rate heterogeneity (ASRH), variation in substitution rate between different genes or between codon positions within a gene (136). The asymmetrical replication of mt genomes predisposes them to this type of heterogeneity (33). Each strand of the mt genome is replicated separately, with one origin of replication located in the control region and a second approx. 11kb downstream. This means that the second, or lagging, strand spends up to two thirds of the replication cycle in a single-stranded state and is more susceptible to mutations. Asymmetrical replication results in strand-specific nucleotide skew, one A+C rich, the other T+G rich, and the genes encoded on different strands will thus differ in their nucleotide frequencies (96). A compounding factor is that the direction of replication, and thus the direction of nucleotide-skew, has been reversed in several lineages (56). Gene rearrangements, especially inversions can also result in a taxon specific departure from average nucleotide skew in the affected gene. Adequate data partitioning (see below) or the use of a γ-parameter can usually compensate for ASRH, but other methods that have been applied include data transformations such as RY-coding or analysis of translated amino acids. One advantage of analysing amino-acid sequences is that mitochondria specific substitution models have recently been calculated for various taxonomic groups including animals (mtZoa, 98) and arthropods.
Data partitioning, and the ability to apply specific models to different partitions, is ideal for analysing genomes shaped by multiple mutational forces such as the mitochondria. The majority of studies have used relatively intuitive partitioning schemes such as by gene type (PCG, rRNA, tRNA), by gene, by codon position, by codon and gene, or by the strand on which the gene is encoded. Different partitioning strategies can each result in strong nodal support for conflicting topologies at the interordinal level but had no effect at intraordinal levels (21, 44, 47), suggesting that partitioning is most significant at deeper phylogenetic levels. Recently, methods for simultaneously choosing partitioning schemes and substitution models have been developed (69). Their application to insect mt genomes suggested optimal partitions wildly at odds with traditionally used partitions (72). In this dataset, acridomorph grasshoppers, different partitioning schemes resulted in modest topological differences, however, in the absence of independent phylogenetic data it was not possible to determine which partitioning schemes produced the most corroborated topology. The effect of data partitioning needs to be investigated over a wider range of taxonomic scales and in multiple orders before its significance can be fully evaluated but available evidence suggests that it is an underappreciated source of variability between different insect mt phylogenomics studies.

The empirical findings produced over the last decade of insect mt phylogenomics are impressively broad. Interordinal relationships have been explicitly addressed in several analyses. Early studies with very broad ordinal representation have produced clearly incorrect results e.g. non-monophyletic Holometabola, Plecoptera+Diptera, Strepsiptera within Holometabola (28), however subsequent more targeted analyses resulted in much better corroborated results. The monophyly of the Diconyphyla and Pterygota are consistently recovered (27, 91, 122, 138, 139). The “Palaeoptera problem” is as alive with mt genome data as it is with nuclear sequence or transcriptome data (see review in 137), with various studies supporting Metapterygota (91, 138), Chiastomyaria (122) or a likely artefactual grouping of Ephemeroptera+Plecoptera within Neoptera (75, 139). Relationships inferred within the Polyneoptera have been largely congruent with those inferred by nuclear ribosomal (e.g. 116) or protein-coding (61) genes, however all studies are missing one or more orders limiting direct comparisons. Four subclades are consistently recovered within Polyneoptera by both nuclear and mt genomic datasets, Orthoptera (47), Dictyoptera (25, 131), Dermaptera+Plecoptera (122) and Grylloblattodea + Embioptera + Mantophasmatodea + Phasmatodea (GEMP) (20, 67, 122), although relationships between these four groups in not consistent between studies. Zoraptera has yet to be included in an mt phylogenomic study. Relationships within each of these polyneopteran subclades are also relatively uncontroversial with the exception of the last, GEMP. Mt genome data consistently groups Mantophasmatodea as the sister of Phasmatodea, even when, as in (67), Embioptera is grouped within Phasmatodea (as sister to Verophasmatodea). In contrast nuclear genes support Grylloblattodea+Mantophasmatodea (61, 116), however this relationship is sensitive to both alignment (66) and taxon selection (37). Both possible relationships for Mantophasmatodea are supported by different morphological traits, yet no resolution of the conflicting data sets has been made.
In contrast to the the Polyneoptera, there has been comparatively limited mt phylogenomics studies into interordinal relationships within the Holometabola. Many of the intraordinal studies (discussed below) of Holometabola have also commented on interordinal relationships inferred as a by-product of outgroup selection, however, these studies omit too many orders to be of much comparative value. The few genuine interordinal studies within the Holometabola have focused on the Neuroptera, whose relationships are strongly affected by analytical biases (23). A sister-group relationship between Neuroptera and Megaloptera was recovered by both (23) and (123), contrary to recent multigene (126, 129) and morphological (11) phylogenies. Mt phylogenomics also supported the monophyly of Megaloptera (123) which was rendered paraphyletic by Raphidioptera in recent studies (11, 129). The position of Raphidioptera is affected by rate variation and groups as the sister to all Holometabola except Hymenoptera in analyses which do not compensate (23, 123). The near complete resolution of holometabolan relationships by nuclear PCGs has been one of the most significant recent accomplishments in insect phylogenetics (see discussion in 120) but have yet to be really explored with mt genome data.

Intraordinal relationships have also been widely studied using mt phylogenomics, with significant numbers of papers investigating relationships within each of the five megadiverse orders plus Orthoptera (47, 72). Compared to the most extensive nuclear PCG studies (35), mt phylogenomics studies of the Hemiptera are of equivalent scale (36, 111), and are actually larger scale within the Heteroptera (74). There is extensive variation in subordinal relationships both between different mt phylogenomic datasets and between mt genomes and nuclear PCGs; sources of topological variation have yet to be tested. Mt phylogenomics studies of Hymenoptera (e.g. 44, 77) recover broadly similar relationships to nuclear data (58). Extensive gene rearrangements (see below), the most compositional biased mt genomes within insects (77) and highly variable substitution rates between lineages (44), however, make the analysis of hymenopteran relationships with mt phylogenomics challenging. The Coleoptera are a similarly challenging order with substantial base compositional heterogeneity between major clades (108, 110). Rooting trees within the order, using Archostemata as an outgroup, improves phylogenetic resolution (118), however, backbone relationships are not strong supported and in conflict with nuclear studies (60). Due to a parallel sequencing effort, direct comparisons of mt phylogenomic and nuclear PCG analyses can be made for the Diptera (127). Relationships between brachyceran families are highly congruent (21), and mt genomes provided the first evidence for a now-resurrected Orthorrhapha. In contrast relationships within the “nematocera” are quite divergent between nuclear and mt datasets, particularly in the relative positions of the infraorders Culicomorpha and Bibionomorpha (4). Finally, although the range of lepidopteran superfamilies represented is still modest (8 of 43), mt phylogenomics studies are highly congruent with nuclear-gene analyses (84, 95, 135). However, unlike these nuclear datasets, lepidopteran mt genome datasets find significant nodal support for relationships between superfamilies (135). Interestingly, an mt genome phylogeny was the first to show the paraphyly of the Macrolepidoptera, with the butterflies (Papilionoidea) forming part of a basal grade of superfamilies in the Obtectomera (134), a finding not acknowledged by either of the subsequent large-scale phylogenies of Lepidoptera (84, 95).

Overall then the use of mt genomes to infer intraordinal relationships has been widely applied across insects and in most instances is broadly congruent with other phylogenetic data sources. For ancient relationships (e.g. interordinal, major subordinal clades) the nodal support resulting from mt genomes is, for the majority of clades, significantly higher than all but the largest multi-locus phylogenetic studies.
Areas of topological conflict are typically areas of poor nodal support in one or both studies, suggesting that additional data and refined analytical methods could potentially resolve most, if not all, conflicts. A final point that should be made is that these studies have been largely conducted in isolation from nuclear-gene or morphological datasets. Historically, due to technical and financial demands, mt genomes are not simultaneously analysed with nuclear data (see 127 as a major exception). Today, neither reason still has much validity. Greater coordination in the collecting of mt genomes, nuclear gene sequences and morphology for exemplar taxa should be the goal for the future of this field.

Genome Rearrangements & Insect Evolution

The second major use of mt genome data in elucidating insect evolutionary history is the identification of gene rearrangements shared amongst related lineages, or as Mark Dowton (41) termed it “the examination of genome morphology.” More generally, mt genome rearrangements are part of a class of phylogenetic markers termed “rare genomic changes” (97) which are considered to have very low rates of homoplasy. Other examples from the nuclear genome include SINEs/LINEs, genetic code variants and microRNA insertions (97) and in mt genomes, macrorepeats (25), conserved secondary structures (25), tRNA gene conversions (59) and signature base substitutions (83). Conceptually the use of mt genome rearrangements as phylogenetic markers has much to recommend it (14): the gene set is near constant across bilaterian animals, gene homology is usually unambiguous (but see 59), rearrangements appear to be uncommon, gene order is apparently selectively neutral (45) and given the large number of potential gene orders (37! or 1.367 x 10^{43}) the possibility of convergence must be low. Accordingly instances of genome rearrangements within insects have been received considerable attention. Indeed, the possibility that rearrangement data could be used to unravel insect ordinal relationships lay behind the author’s first job in mt genomics as a spotty postdoc over a decade ago.

The different arrangement of two mt genomes can be described by a standard series of gene movements: transposition, inversion and inverse transposition (41). It has been common to refer to “major” and “minor” rearrangements to describe ones which involve, respectively, protein-coding and/or rRNA genes versus tRNA genes alone. Finally the distance which a gene has moved within the genome as part of the rearrangement is often described as short range rearrangements are held to be more frequent. Thus the rearrangement of a pair of genes, e.g. A-B to B-A, is described as “gene shuffling” (45). These terms can be combined, for instance a rearrangement found in Apis, trnD-trnK, can be described as a “minor transposition” from the ancestral insect arrangement trnK-trnD. While it has been historically common to infer the number of rearrangements which separate two genomes by hand, this is less reliable for more extensively rearranged genomes which multiple possible sequences of rearrangement can be inferred. Software such as CREx standardizes such inferences and allows more repeatable interpretation of rearrangement histories (8, 9).

Such descriptive terms, however, are largely independent of the underlying mechanisms which cause mt genome rearrangements. The most widely accepted mechanism for explaining mt genome rearrangements is the “Tandem Duplication, Random Loss” (TDRL) model (13, 81). In the TDRL model, first a portion of the mt genome is duplicated resulting in two copies of a block of genes. Gene duplications could occur via one of several methods, slipped strand mispairing during replication, imprecise termination of replication, dimerization of the genome or recombination (81). Once duplicated,
the accumulation of mutations within the copied genes will eventually render one of them non-functional, at which point the selective pressure for reduction in genome size (92) results in the elimination of the non-functional gene. TDRL readily explains transpositions but cannot explain inversions, consistent with the observed, low frequency of inversions. TDRL also explains why both minor and shuffling rearrangements are most common – a shorter stretches of DNA must be duplicated before a tandem gene block is created. It also explains the increased rate of rearrangements for genes adjacent to the origin of replication, as both strand slippage and imprecisely termination are more likely to include the genes surrounding the origin of replication in the duplicated gene block. Finally several insects have been found which are apparently part-way through the TDRL process, e.g. the scorpionfly Microchorista which a 3.5kb duplication of 3 protein-coding genes and 7 tRNAs each of which display heightened rates of substitutions and indels (5). The existence of intermediates is good proof for the sequence of events outlined in the TDRL model. Across insects, TDRL explains most of the observed mt genome rearrangements, in contrast to vertebrates where it explains almost all observed rearrangements (45).

Two other rearrangement mechanisms have received considerable attention: tandem-duplication, non-random deletion (TDNR) and recombination. TDNR is similar to the TDRL model except that gene-loss is constrained by transcriptional blocks within the genome (71). Genes duplicated across the boundaries of these transcriptional blocks cannot be expressed and so are eliminated; gene loss is thus non-random. TDNR has not often been invoked to explain rearrangements in insects, however the highly rearranged genome of the winter crane fly, Paracladura is consistent with duplication of the entire mt genome followed by the loss of multigene transcription blocks (4). The third mechanism proposed to cause mt genome rearrangements is recombination. Despite a history of contention around the possibility of mitochondrial recombination (81), gene inversions cannot be explained without some form of recombination (40). Inversions are the least common type of rearrangements found in insects and have only been recorded from the orders Dermaptera (122), Hymenoptera (45) and all three of the Paraneopteran orders (101, 103, 117, 125).

Many insect species possess the ancestral pancrustacean mt genome arrangement, however independent rearrangements are found in many orders. Unfortunately there is no equivalent of Jeff Boore’s discovery of the Pancrustacea (15): a rearrangement which profoundly altered our understanding of insect evolution. Today, with a representative mt genome sequenced from each order, plus representatives from each of the largest subordinal clades of most orders, it is clear that there are no gene rearrangements shared between orders. Instead the synapomorphic mt genome rearrangements found in insects define clades at a variety of taxonomic scales below the level of order (Table 2); clearly rearrangements are not at all clock-like. In addition there are a large number of taxa in which a unique rearrangement has been noted in a single species but the taxonomic extent of these rearrangements has not been determined. Such examples range from single tRNA rearrangements e.g. Ischalia (Coleoptera: Anthicidae, 118) through to rearrangements of multiple protein-coding genes e.g. Aposthonia (Embioptera: Oligotomidae, 67). It is however apparent that, with the notable exceptions of hemipteroids and hymenopterans (see below), mt genome rearrangements within insects are rare events. In most of the instances noted in Table 2, the synapomorphic rearrangement is the only one found in that order. With the same taxon exceptions, they are also reliable phylogenetic markers. There are few instances where additional rearrangements have resulted in the secondary loss of the synapomorphic gene order and convergence homoplasies are rare. The most notable convergence is the rearrangement trnK-trnD → trnD-trnK which is found in acridomorph grasshoppers (49) and has occurred independently in at least five hymenopteran families:
Apidae (34), Stephanidae (45), Braconidae (124), Formicidae (55), Scelionidae (77). Thus empirically, within insects mt genome rearrangements largely fulfil the criteria of ideal phylogenetic characters which early enthusiasts hypothesized they would possess (14, 97).

There are however two enormous exceptions, Hymenoptera and the Hemipteroids, where extremely high rates of genome rearrangement at best have obscured the phylogenetic reliability of observed rearrangements and at worst resulted in rampant homoplasy. The Hymenoptera have very high rates of tRNA rearrangements; every sequenced hymenopteran species has at least one translocated tRNA. It was established early that tRNA rearrangement were common within Hymenoptera and that the multi-gene tRNA blocks were “hot spots” of rearrangement (39, 42, 43). A review of hymenopteran rearrangements (45), found that of the 67 rearrangements identified to that time, only five were shared between two or more species and of those only two were genuinely synapomorphic. Unique rearrangements were extremely common and convergence between local tRNA rearrangements common. Despite increasing taxonomic coverage and depth of sampling within major groups, the hymenopteran mt genomes sequenced since 2009 have served to confirm pattern. The one exception is the Chalcidoidea which appear to share an inverted block of five protein-coding genes (132). The identification of synapomorphic rearrangements within the Hymenoptera is thus not impossible but is hampered by high rates of noise and significant taxonomic coverage is required to confirm putative synapomorphies.

The Hemipteroids have even greater rates of rearrangement and correspondingly higher difficulties in providing unambiguous interpretations. All three hemipteroid orders include highly rearranged taxa although within the Hemiptera they are confined to the whiteflies (Aleyrodidae) (Thao et al. 2004), the heteropteran Stenopirates (73) and isolated examples of tRNA rearrangements; the majority of families which have been sequenced possessing the ancestral insect arrangement. Both of the available Thysanoptera (thrips) mt genomes are massively rearranged relative the ancestral insect, retaining only 7 of 37 gene boundaries, but quite similar to each other, differing by just 6 tRNA transpositions (101, 133). The lack of mt genomes from thysanopteran families other than Thripidae means that these findings are difficult to contextualise, however a recent phylogeny suggests that the two sequenced genera (Thrips, Frankliniella) represent the most widely divergent clades within this speciose family (17). This suggests that mt genomes within thrips have been stable over long evolutionary timescales after some early massive rearrangement. The final order, the Psocodea includes what was formerly two orders, the barklice (= Pscoptera) and true lice (= Phthiraptera) (62). Available pscopteran mt genomes include independent major rearrangements from the ancestral insect gene order (104, 125, Cameron pers. obs), however louse mt genomes are the most rearranged of all arthropods (103). Louse mt genomes retain between three (Bothriometopus, 22) and none (Ibidoecus, 24) of the derived gene boundaries found in the ancestral insect mt genome. Further, with the exception of two very closely related species from the family Gonioidae, very few novel gene arrangements are shared between lice either. Previously (24), I identified only 11 arrangements shared between any two louse species of which four were convergent when mapped onto an independent phylogenetic tree and five required the postulation of secondary loss to be interpreted as synapomorphies. Confidence in these interpretations was further undermined by the observation of only partial genome sequences for several species due to the presence of genome mini-circles in several louse lineages (see below). Collectively the variation observed within louse mt genomes greatly exceeds our capacity to reliably infer evolutionary patterns from the available species diversity and a huge additional sequencing effort is needed to bring order to this chaos.
The final question is what causes mt genome rearrangements to be elevated in particular groups (as opposed to the mechanisms by which rearrangements occur). This has received considerable attention, with a focus on life-history traits as predictors of rearrangements. Parasitism has been repeatedly invoked as a predisposing factor (42, 103), which is unsurprising given the highly rearranged mt genomes found in several parasitic insect groups. Tests of this hypothesis, however, have largely rejected it. A test of rearrangement rates between parasitic and non-parasitic groups within the Hymenoptera and Diptera, concluding that there was no correlation between parasitism and rearrangements in flies (30). Similarly, more detailed studies of Hymenoptera showed that heightened rates of rearrangement did not correspond with the evolution of parasitism within the order (45). What these studies have shown, is that there is a strong correlation between increased rearrangement and nucleotide substitution rates in parasitic lineages (30, 104), however it is not clear if one causes the other or if they both the common result of a third factor. A second cause of rearrangements that has been proposed is duplication of the control region (CR) which has occurred in thrips and barklice (101, 105, 133). The effect of duplicated CRs on rearrangement rates has not been adequately tested, however, within thrips they actually coincide with quite stable genome arrangements as only six tRNA rearrangements have occurred between the two thripid species with duplicated CRs. Finally there may be a correlation between mt genome rearrangements and the evolution of haplodiploidy. Heightened rearrangement rates are found in four of the eight insect groups which have independently evolved haplodiploidy (sciarid flies, Hymenoptera, Thysanoptera, Aleyrodidae) and mt genome data is not available for the other four clades (Micromalthus beetles, Xyleborini weevils, Hypothenemus weevils and coccoid scales) (88). The one exception is lice which are not haplodiploid but do practise a form of paternal genome elimination which can result in similar genomic inheritance patterns as classical haplodiploidy (80). The causes of mt genome rearrangements in insects are likely multifactorial and much additional research is required.

Model Systems for “Aberrant” Mitochondrial Genomics

The final aspect of study on insect mt genome evolution is as model systems for understanding “aberrant” genomic systems. Given the stability of mt genome structure and function across insects, studying examples which depart from these trends can be highly informative about what factors underlie that stability. Two interesting examples which have been recently discovered from insects are gene truncation in gall midges (6) and genome fragmentation in lice and near relatives (106).

In cecidomyiid gall midges, each tRNA gene has been severely truncated (6). Over 90% of metazoan tRNAs posses a canonical clover-leaf secondary structure composed of four arms of conserved length: the aminoacyl (or acceptor) arm which holds the isotype specific amino acid, the dihydrouridine (or DHU) arm, the anticodon arm which determines tRNA isotype, and the pseudouridine or ТΨС arm (63). The one exception is trnS1 which lacks the DHU arm in most metazoans.

The entire ТΨС arm and the 3’ end of the acceptor stem have been lost from each tRNA gene in cecidomyiids, whereas there is a high sequence-level conservation of the DHU and anticodon stems (6). The only groups where similar truncations have been found are nematodes (130), and some chelicerates (38, 78) in which the ТΨС stem is lost for some tRNA isotypes. Additionally, most tRNAs of spiders lack paired acceptor stems (38). The truncation of cecidomyiid tRNAs is, however, more extreme as it affects all tRNA isotypes and in all instances resulted in physically shorter genes (6). It is unclear what
factors may have lead to these truncated tRNAs, although it may be a result of generalised evolutionary pressures for size reduction in mt genomes (92). Such an explanation however would require the existence of compensatory mechanisms. Polycistronic transcript processing would differ as the tRNA secondary structures no longer form recognition sites for mRNA cleavage (89). Secondly, recognition of tRNA isotype by tRNA-synthetases is based on conserved sequences in both the anticodon and acceptor stems (10). Template-dependent RNA editing, using the 5’ end of the acceptor stem as a template for polymerising the missing bases after transcript cleavage, has been proposed as a means to restore proper functionality to truncated tRNAs (6). This type of RNA editing has been observed in centipedes (70), however the enzymes which perform the editing are unknown.

The second major model “aberrant” system found in insects is the multi-chromosomal mt genomes found in several lineages of lice and bark lice. First observed in the human body louse Pediculus (106), the fragmentation of the mt genome into several chromosomes has occurred multiple times in the Psocodea. Departures from the single chromosome mt genome found in almost all bilaterians are extremely rare, being recorded only in the potato-cyst nematode Globodera (2, 51) the rotifer Brachionus (112). The multi-chromosomal mt genomes found in lice are, however, both more variable in structure and occur in a wider range of taxa than either of these earlier examples. Genome reductions have occurred at least 4 times in lice and one in psocoptera (Liposcelis) and at least three different types of genome structures are found (24, 125). Heteroplasmic genome reductions, where a minicircular 23-gene chromosome coexists with a full sized 37-gene chromosome, are found in goniodid pigeon lice. Stable genomes composed of several, multi-gene chromosomes are found independently in several louse lineages and Liposcelis. Finally, the most extreme genome fragmentation, into a large number of chromosomes each with 1-3 genes and a large conserved non-coding region, appears to be synapomorph for the sucking lice, the Anoplura, and their chewing lice sister-group, the Trichodectidae. Analyses of mt genome variability within human lice has shown that the multi-chromosomal state corresponds to high rates of recombination (107) including the formation of chimeric chromosomes composed of two or more minicircles (102). While mt genome fragmentation was initially linked to blood feeding in derived lice (106), analyses of the nuclear-encoded, mt-targeted maintenance genes in Pediculus suggested that the loss of one of the replisome genes, mt single-stranded binding protein (mtSSB), was likely responsible (24). The replication of full length mt genomes is not possible without this gene however the minicircle sized ones found in Pediculus could be replicated by rest of the polymerase holoenzyme (68). Whether mtSSB is lost or has altered functionality in other psocodean lineages with multi-chromosomal mt genomes is unknown. Deletion- and minicircular mt genomes are found in humans where they cause metabolic, neurodegenerative or age-related diseases (121). Understanding how lice tolerate mt genome structures lethal in other species, presumably purifying selection explains their absence from most lineages, could highly instructive about general mt genome function.

Interestingly both of these examples, gall-midges and lice, are from groups with unusual nuclear genomics in addition to their mt genomic peculiarities. Both have extremely small nuclear genomes (53), and can eliminate chromosomes as part of significant departures from normal Mendelian genetics (80, 88). As discussed above, there is strong correspondence between instances of haplodiploid nuclear genetics and mt genome rearrangements (88). The interactions between the nuclear and mt genome evolution has been largely unexplored to date but are a promising line of investigation in furthering our understanding of the evolutionary dynamics of insect mt genomes.
Summary Points

1. The mt genomes of insects are the best studied of all the invertebrates, with almost 500 species sequenced to data. The representativeness of this dataset is also excellent, including all orders, most suborders and significant sampling of major lineages within the megadiverse orders. While clear gaps in our sampling can be identified, of which the coccoid scales are the most glaring, recent advances in sequencing technology suggest that these can be readily filled in the coming years.

2. Extensive mt phylogenomic analyses of insects has clearly illustrated best practise. All 37 genes should be included as a matter of course, however, potential biases due to the inclusion of third codon positions evaluated in each dataset. Base compositional biases affect many insect mt genomes, however, this effect is not adequately compensated for by standard analytical software. Rate variation is also common, but can be corrected for by partitioning strategies. Partitioning approaches vary significantly between studies, however its effect is rarely tested. Too many recent studies still use a single, arbitrarily chosen analytical design testing and the sensitivity of results should be much more widely tested.

3. Empirical results from mt phylogenomics studies are informative at many levels in insect systematics. The results obtained are rarely wildly incongruent with those from morphology or nuclear gene sets and usually have much higher levels of nodal support for deep-level nodes. The wider integration of mt genome data with nuclear phylogenetic studies will serve to strengthen the hypotheses generated by either dataset in isolation.

4. Genome rearrangements are rare in the majority of insect orders. Where they occur they are usually synapomorphic, however, the taxonomic level to which they map varies widely. The Hymenoptera and Psocodea have exceptionally diverse gene orders, with convergent rearrangements found in unrelated taxa. Haplodiploidy, either obligate or via paternal genome elimination, may be a predisposing factor toward mt genome rearrangement.

5. The stability of insect mt genomes across hundreds of millions of years means that groups which differ wildly from norms can be great model systems for understanding how mt genomes function and the selective constraints on them.

6. The future of the field lies in ever greater integration between nuclear and mt genome datasets. Improved phylogenetic resolution will come from large integrated datasets. Understanding the nuclear contribution to mt genome replication and maintenance will improve our understanding of mutational constraints on substitution, gene order and genome structure.

Acknowledgements

I would most like to thank Michael Whiting and Stephen Barker who gave me my star in mitochondrial genomics, my major collaborators on much of this work, Mark Dowton, Hojun Song, Kevin Johnson and Iksoo Kim, and students, Catherine Covacin, Daniel Fenn and Leigh Nelson, without your time, effort, ideas and advice, so much of the primary work on insect mt genomes wouldn’t have been done or wouldn’t have been done nearly as well. This work has been supported by the US National Science Foundation (DEB0444972, EF0531665), CSIRO Julius Career Awards, QUT Vice Chancellor’s Research Fellowship scheme and the Australian Research Council Future Fellowships scheme (FT120100746).
References


Table 1. Available mt genome data (complete/near complete genomes) for Insects. (March 2013, sourced from GenBank, MitoZoa and MetAmiga databases).

<table>
<thead>
<tr>
<th>Order</th>
<th>Suborder</th>
<th>No. mt genomes sequenced</th>
<th>No. Families, recognised</th>
<th>No. Families represented by mt genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaeognatha</td>
<td></td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Zygentoma</td>
<td></td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Ephemeroptera</td>
<td></td>
<td>8</td>
<td>35</td>
<td>6</td>
</tr>
<tr>
<td>Odonata</td>
<td></td>
<td>4</td>
<td>32</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Epiprocta</td>
<td>2</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Zygoptera</td>
<td>2</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>Plecoptera</td>
<td></td>
<td>1</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>Dermaptera</td>
<td></td>
<td>1</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Orthoptera</td>
<td></td>
<td>61</td>
<td>39</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Ensifera</td>
<td>48</td>
<td>27</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Caelifera</td>
<td>13</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Phasmatodea</td>
<td></td>
<td>16</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Embioptera</td>
<td></td>
<td>1</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Grylliobattodea</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mantophasmatodea</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Blattodea</td>
<td></td>
<td>24</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>“cockroaches”</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Isoptera</td>
<td>18</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Mantodea</td>
<td></td>
<td>1</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>Zoraptera</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hemiptera</td>
<td></td>
<td>69</td>
<td>168</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Sternorrhyncha</td>
<td>9</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Auchenorrhyncha</td>
<td>11</td>
<td>36</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Coleorrhyncha</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Heteroptera</td>
<td>47</td>
<td>86</td>
<td>32</td>
</tr>
<tr>
<td>Thysanoptera</td>
<td></td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Psocodea</td>
<td>“Psocodea”</td>
<td>2</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>“Phthiraptera”</td>
<td>8</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td></td>
<td>35</td>
<td>96</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>“Symphyta”</td>
<td>3</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>“Parasitica”</td>
<td>16</td>
<td>50</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Aculeata</td>
<td>16</td>
<td>30</td>
<td>7</td>
</tr>
<tr>
<td>Coleoptera</td>
<td></td>
<td>68</td>
<td>172</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Archostemata</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Myxophaga</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Adephaga</td>
<td>5</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Polyphaga</td>
<td>60</td>
<td>152</td>
<td>31</td>
</tr>
<tr>
<td>Neuroptera</td>
<td></td>
<td>7</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>Megaloptera</td>
<td></td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Raphidioptera</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Trichoptera</td>
<td></td>
<td>1</td>
<td>47</td>
<td>1</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td></td>
<td>74</td>
<td>132</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>“basal” Lepidoptera</td>
<td>2</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Ditrysia</td>
<td>72</td>
<td>109</td>
<td>17</td>
</tr>
<tr>
<td>Siphonaptera</td>
<td></td>
<td>1</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>Mecoptera</td>
<td></td>
<td>4</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Diptera</td>
<td></td>
<td>77</td>
<td>209</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>“Nematocera”</td>
<td>26</td>
<td>35</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Brachycera</td>
<td>51</td>
<td>174</td>
<td>12</td>
</tr>
<tr>
<td>Strepsiptera</td>
<td></td>
<td>2</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 2. Synapomorphic mt genome rearrangements found in insects. References include the first publication to note the rearrangement and, if applicable, the paper which confirmed its taxonomic extent.

<table>
<thead>
<tr>
<th>Order</th>
<th>Clade</th>
<th>Level</th>
<th>Rearrangement</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthoptera</td>
<td>Tettigoniidae +</td>
<td>Infraorder</td>
<td>trnK-trnD → trnD-trnK</td>
<td>49, 72</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>Aleyrodidae</td>
<td>Family</td>
<td>trnC-trnY → trnY-trnC</td>
<td>117</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>multiple aleyrodid genera</td>
<td>Genus/ btw genera</td>
<td>cox3-trnG-nad3 → one of three different locations in mt genome</td>
<td>117</td>
</tr>
<tr>
<td>Thysanoptera</td>
<td>Thripidae</td>
<td>Family</td>
<td>Extensively rearranged mt genome with all but 6 gene positions shared by both thrips species</td>
<td>101, 133</td>
</tr>
<tr>
<td>Psocodea</td>
<td>Psocomorpha</td>
<td>Suborder</td>
<td>nad3-trnA-trnR-trnN-trnS1-trnE-trnF-nad5 → trnA-trnR-trnN-nad5-nad3-trnN-trnS1-trnE</td>
<td>Cameron (pers. obs)</td>
</tr>
<tr>
<td>Psocodea</td>
<td>Ischnocera</td>
<td>Suborder</td>
<td>trnY-cox1 → trnl-cox1</td>
<td>24</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Chalcidoidea</td>
<td>Superfamily</td>
<td>inversion of cox1-trnL2-cox2-trnK-trnD-atp8-atp6-cox3</td>
<td>90, 132</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Formicidae</td>
<td>Family</td>
<td>trnl-trnQ-trnM → trnV-trnM-trnl-trnQ</td>
<td>52, 55</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Apidae</td>
<td>Family</td>
<td>inversion of trnR</td>
<td>45</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Meliponini</td>
<td>Tribe</td>
<td>cox2-trnK-trnD-atp8 → cox2-trnD-atp8</td>
<td>109</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td><em>Apis</em></td>
<td>Genus</td>
<td>trnl-trnQ-trnM → trnM-trnQ-trnA-trnl</td>
<td>34, 113</td>
</tr>
<tr>
<td>Neuroptera</td>
<td>Neuroptera excluding at least Osmylidae</td>
<td>btw Order and Superfamily</td>
<td>trnW-trnC-trnY → trnC-trnW-trnY</td>
<td>23, 140</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Dryopoidea</td>
<td>Superfamily</td>
<td>trnP-nad6 → nad6-trnP</td>
<td>119</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Entiminae-</td>
<td>btw Subfamily and Family</td>
<td>trnA-trnR → trnR-trnA</td>
<td>54, 110</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>Ditrysia</td>
<td>btw Infraorder &amp; Superfamily</td>
<td>trnl-trnQ-trnM → trnM-trnl-trnQ</td>
<td>26, 114</td>
</tr>
<tr>
<td>Strepsiptera</td>
<td>Strepsiptera</td>
<td>Order</td>
<td>trnN-trnS1-trnE → trnN-trnE</td>
<td>27, 79</td>
</tr>
<tr>
<td>Diptera</td>
<td>Culicidae</td>
<td>Family</td>
<td>inversion of trnS1</td>
<td>3, 7</td>
</tr>
<tr>
<td>Diptera</td>
<td>Cecidomyiidae</td>
<td>Family</td>
<td>inversion of trnT and trnP without translocation</td>
<td>6</td>
</tr>
</tbody>
</table>
Legends to Figures.

Figure 1. Accumulation of mitochondrial genome data from insects. Number of species sequenced in each year are blue bars (left axis scale), cumulative total is indicated by red line graph (right axis scale).

Figure 2. Mt genome diagram of ancestral insect. Gene abbreviations are as follows: *atp6, atp8*: ATP synthase subunits 6 and 8 genes; *cob*: cytochrome oxidase b gene; *cox1-cox3*: cytochrome oxidase c subunit 1-3 genes; *nad1-6, nad4L*: NADH dehydrogenase subunits 1-6 and 4L; rRNA: ribosomal RNA; *rns, rnl*: small and large ribosomal RNA subunits; CR: control region; tRNA genes are indicated by the single letter IUPAC-IUB abbreviation for their corresponding amino acid; direct of gene transcription is indicated with an arrow. Gene sizes are roughly proportional to their nucleotide length.
Mini-Glossary

Transpositions: movement of a gene or genes without change of translational direction.

Inversions: movement of a gene or genes with change of translational direction.

Inverse transpositions: movement and inversion of a gene or genes.

Gene boundaries: description of shared adjacent genes e.g. two species which share trnD-trnK share this gene boundary. May be either ancestral, present in ancestral insect, or derived, a novel occurrence in one or more lineages.

Acronyms

mt: mitochondria
PCG protein-coding gene
rRNA ribosomal RNAs
tRNA transfer RNAs
CR control region (also A+T rich region, major non-coding region)

Reference annotations:

[5] Best documented example of a duplicated mt genome still in the process of losing additional gene copies. Strong evidence for the TDRL model of genome rearrangements.

[15] Most significant single tRNA translocation yet discovered, providing the first evidence for the now well accepted Pancrustacea theory of arthropod relationships.

[40] Strong argument that gene inversions are proof of recombination occurring within mitochondria.

[72] Most extensive test to date of the impact of partitioning schemes on inferring phylogenetic relationships using mt genome data.

[76] Example of using mt genomes to infer population history and phylogeography for globally distributed locusts.

[87] Example of using mt genomes to efficiently identify target loci for diagnostics design.

[106] First record of the remarkable phenomena of multichromosomal mt genomes, that have now been found across lice.

[108] Comprehensive examination of the effects of nucleotide compositional heterogeneity on inferring phylogenetic relationships using mt genomes and demonstrates solutions to this widespread problem.

[118] Efficient protocols for applying next-generation sequencing to rapidly collect a large number of insect mt genomes.
What are mitochondria?

The mitochondrion is a fundamental eukaryotic organelle, descended from an alpha-proteobacterium which formed a permanent symbiosis with the ancestral eukaryote roughly 2 billion years ago. Their best studied function is energy production via oxidative phosphorylation – the aerobic breakdown of organic molecules to form ATP. Mitochondria themselves possess two unit membranes, a smooth outer membrane and an inner membrane composed of highly convoluted, transverse folds termed cristae. A series of four protein complexes, the electron transport chain (ETC), located in the mt inner membrane pass electrons from respiration intermediates (NADH or succinate) to oxygen, producing water and pumping protons into the mt intermembrane space. Balancing the electrochemical gradient between the intermembrane space and the mt matrix produced by the ETC is coupled to ATP production by ATP synthetase. The proteins encoded by the mt genome form part of 3 of the 4 ETC complexes and ATP synthetase, however the vast majority of mt active proteins are encoded in the nucleus. Mitochondria also have functions in apoptosis and cell-aging meaning that mt are implicated in many degenerative diseases of aging. Mt are retained even in anaerobic species which cannot undergo oxidative phosphorylation as they are the site of other cellular biochemistry.
Fig. 1

Fig 2.