African Journal of Biotechnology Vol. 5 (14), pp. 1308-1318, 16 July 2006
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2006 Academic Journals

Full Length Research Paper

Comparative insect mitochondrial genomes: Differences despite conserved genome synteny

Sathees B. C. Chandra[1]*, Jennifer L. Vlk[1] and Vinayak Kapatral[2]

[1]Department of Biological, Chemical and Physical Sciences, Roosevelt University, Chicago, IL 60605, USA.
[2]Integrated Genomics, 2201 W Campbell Park Dr, Chicago, IL 60612, USA.

Accepted 7 June, 2006

We present a comparative analysis of select insect mitochondrial DNA (mtDNA) representing four insect orders (Diptera, Hymenoptera, Orthoptera and Coleoptera) consisting of 12 different species in an effort to study a common set of genes and to understand the evolution of mitochondrial genome. A functional analysis of mitochondrial genomes was carried out using ERGO bioinformatics suite. To compare the similarity between closely related insect mitochondrial genome sequences, dot-plot comparisons of sequences were performed. LSU and SSU rRNA sequences were used to construct a phylogenetic tree to determine the relationship among four insect orders. LSU rRNA sequences yielded a tree with branching patterns reflecting the expected pattern as insect species belonging to different orders were put into separate clades. Based on the sequence similarity, insect species belonging to four different orders in general appear to be closely related. However, a comparative and functional analysis of insect mitochondria sequences revealed differences in gene organization of mtDNA. Although tRNA species were identical in most species of insects, their position and the transcription orientations were different, reflecting differential transcriptions. Based on this study we conclude that, although the gene types are very similar across these insect orders, significant differences in GC content perhaps suggest multiple mitochondrial ancestors.

Key words: Mitochondria, genome analysis, gene content, gene order, phylogenetic analysis.

INTRODUCTION

Mitochondria are key energy generators in most eukaryotic cells. Research on mitochondria has primarily focused on the process of ATP generation, phylogeny and evolutionary origins. Unique sequence signatures from mitochondrial DNA (mtDNA) have been used not only to categorize species, but also to study animal, bird, and human migration as well as in diagnostics and forensics. With the increase in the whole genome sequencing of eukaryote genomes, mtDNAs are inevitably sequenced and this has facilitated comparative studies. Mitochondria are believed to have evolved in eukaryotes through a process called serial endosymbiosis from an unknown microbial ancestor. An alternate theory proposed by Gray et al. (1999) suggests that mitochondria arose from a common ancestral extinct eukaryote, and evolved concurrently with the nucleus. Although the common mitochondrial ancestor is yet to be identified, several studies have suggested a very close relationship with endosymbionts belonging to α-Proteobacteria such as *Rickettsia* spp., *Anaplasma* spp. and *Ehrlichia* (Gray et al., 1999; Gray et al., 2001). Irrespective of their origins, mtDNA in general appear to have lost genes and have retained identical genes

*Corresponding authors E-mail: schandra@roosevelt.edu; Phone: 1-847-619-7968. Fax: 1-847-619-8555.

(coding and non-coding) by a process commonly referred to as reductive evolution.

The mitochondrial genomes from eukaryotes reveal large size differences between animal, fungal, and plant species. The average size of animal mtDNA is ~16 kb (Boore, 1999), whereas plant mtDNAs range from 200 to 2500 kb (Palmer, 1990). The size of fungal mtDNAs range from 19 kb to 176 kb (Hudspeth, 1995). The mtDNA genome has a limited number of protein coding genes such as NADH ubiquinone oxidoreductase (ND2 to ND5), cytochrome c oxidase (COX), succinate dehydrogenease (SDH), ATP synthase (ATP6), cytochorome (CYTB), etc. Among non-coding genes only rRNA and tRNA genes are found.

Although several mtDNA have been sequenced directly or as part of whole genome sequencing, comparative analysis is limited to fission and budding yeasts (Bullerwell et al., 2003; Langkjaer et al., 2003). Such analyses have identified common and unique signature sequences among the mtDNA from three fission yeasts, *Schizosacchromyces pombe, S. octosporus* and *S. japonicus*. These three species show great variation in size of mtDNA due to the presence of non-coding regions. Although the gene organization between *S. pombe* and *S. octosporus* is similar, the absence of the trnl2 (cau) gene in *S. octosporus* has been reported. Similarly, both rps3 and rnpB genes found in *S. pombe* are absent in *S. japonicus (*Bullerwell 2003). Further the presence of five double hairpin elements (DHEs) found in *S. octosporus* are missing in both *S. pombe* and *S. japonicus*. Similarly, among the budding yeasts, *Saccharomyces castellii* and *S. servazzii* contain fewer introns]and intergenic sequences than *S. cerevisiae* (Langkjaer et al., 2003).

Recently, several mtDNA have been sequenced from insects including *Apis mellifera* (Crozier and Crozier, 1993), *Drosophila melanogaster* (Lewis et al., 1995), *Anopheles gambiae* (Beard et al., 1993), *Anopheles quadrimaculatus* (Mitchell et al., 1993), *Melipona bicolor* (Silverstre and Arias, 2002), *Drosophila simulans* (Ballard, 1999), *Gryllotalpa orientalis* (Kim et al., 2005), *Drosophila mauritiana* (Ballard, 1999), *Locusta migratoria* (Flook et al., 1995), *Pyrocoelia rufa* (Bae et al., 2004), *Crioceris duodecimpunctata* (Stewart and Beckenbach, 2005), *Tribolium castaneum* (Friedrich and Muqim, 2003) and at least 40 more hexapod mitochondrial genome sequences are available from the GenBank non-redundant database to date. We selected these four insect orders involving twelve different species for analysis based on prior behavioral genetic research of honeybees and fruit flies (Chandra et al., 1998; Chandra et al., 2000; Chandra et al., 2001; Chandra and Singh, 2005; Rueppel et al., 2006) which suggest strong similarities these orders. In addition, species belonging to the order Coleoptera and Orthoptera exhibit close evolutionary origins with species belonging to Hymenoptera and Diptera. It is then the purpose of this study to comparatively analyze the protein coding, rRNA, tRNA, and regulatory non-coding sequences of the mtDNA of these twelve species, and to determine if conserved sequences across the four insect orders are suggestive of mtDNA evolutionary origin.

MATERIALS AND METHODS

Insect mitochondrial genomes

The complete mtDNA genome sequences from 12 different insect species representing four orders used in this study were obtained from GenBank (NCBI: http://www.ncbi.nlm.nih.gov). The dataset includes at least two representatives from each of the four insect orders. mtDNA sequences from *Apis mellifera* (NC_001566), *Drosophila melanogaster* (NC_001709), *Anopheles gambiae* (NC_002084), *Anopheles quadrimaculatus* (NC_000875), *Melipona bicolor*, *Drosophila simulans*, *Gryllotalpa orientalis*, *Drosophila mauritiana* (NC_005779), *Locusta migratoria* (NC_001712), *Pyrocoelia rufa* (NC_003970), *Crioceris duodecimpunctata* (NC_003372), *Tribolium castaneum* (NC_003081) were downloaded into the ERGO bioinformatics database for analysis. Nucleotide composition, protein-coding genes, rRNA genes and tRNA genes were analyzed.

Phylogenetic analysis

We initially performed DNA alignment of all LSU rRNA and SSU rRNA sequences using ClastalW (Thompson et al., 1994) for phylogentic analysis followed by neighbor-joining method to construct phylogenies from sequence alignments using “PHYLIP” (Felenstein, 1989) and ClustalX (Thompson et al., 1997). In order to assess the level of significance from the inferred clades, a bootstrap procedure was used following SEQBOOT, DNADIST and NEIGHBOR packages. Finally we extracted the bootstrap tree using the CONSENSE program and the tree was drawn using a TREE VIEW (Page, 1996) and NJPLOT program.

Dot-plot comparison

To compare the similarity between very closely related mitochondrial genome sequences, Dot-plot comparisons of sequences were conducted using either the Microsoft Windows version of DOTTER (Sonnhammer and Durbin, 1995) or DOT-MATRIX (Huang and Zhang, 2004) with a window size of 22 and maximum of one mismatch.

Genome analysis

For comparative and functional analysis, the ERGO bioinformatics suite developed by Integrated Genomics (Overbreek et al., 2003) (http://ergo.integratedgenomics.com/ERGO/) was used. Currently ERGO contains over 850 genomes with annotation and pathway curation, from phylogenetically diverse species.

RESULTS AND DISCUSSION

Genome size

Most mtDNA genome in general are closed circular molecule except for yeast species such as *Candida*

Table 1. General genome features of mitochondria belonging to four different orders. The size of the mitochondrial DNA, their GC content, number of coding genes and non-coding genes among the insect species of different orders are indicated.

Order	Insect species	Size (bp)	GC content (percent)	Coding genes	Non-coding genes	
					tRNA	rRNA
Hymenoptera	*Apis mellifera*	16,343	15.14	13	22	2
	Melipona bicolor	14,422	13.28	13	19	2
Diptera	*Drosophila melanogaster*	19,517	17.84	14	22	2
	Drosophila simulans	14,972	22.07	13	22	2
	Drosophila mauritiana	14,964	22.29	13	22	2
	Anopheles gambiae	15,363	22.44	13	22	2
	Anopheles quadrimaculatus	15,455	22.64	13	22	2
Orthoptera	*Gryllotalpa orientalis*	15,521	29.51	13	22	2
	Locusta migratoria	15,722	24.67	13	22	2
Coleoptera	*Pyrocoelia rufa*	17,739	22.59	13	22	2
	Crioceris duodecimpuntata	15,880	23.11	13	22	2
	Tribolium castaneum	15,881	28.32	13	22	2

parapsilosis (Rycovska et al., 2004), *Pichia pijperi, Williopsis mrakii*, *Pichia jadinii* (Fukuhara et al., 1993) and *Pythium oligandrum* (Martin, 1995) and have a common bacterial triplet gene code. The insect genomes are also closed circular and have the bacterial universal genetic code. The genome size varies from 14,422 bp (in *M. bicolor*) to 19,517 bp (*D. melanogaster*) (Table1). All insect mtDNA are similar in size except for *D. melanogaster* mtDNA which is larger compared to other species of *Drosophila* due to the presence of a non-coding 4553-bp region at the end of genome and it has 14 coding genes compared to 13 in most insects.

The mtDNA from Hymenoptera and Diptera have minor differences in size within each order. For example, *A. mellifera* is 1,921 bp longer than *M. bicolor*, which is attributable to a difference in the number of tRNA genes. *A. mellifera* has 22 tRNA genes compared to 19 in *M. bicolor*. Variation in mitochondrial size is also due in some cases to variation in the repeat length of non-coding regions. However, despite the size differences, the gene order is not altered.

The mtDNA size in the orders Hymenoptera and Diptera are similar, but their GC content varies significantly. The lowest GC content was found among Hymenoptera such as *A. mellifera* and *M. bicolor*, with 15 and 13%, respectively (Table 1). As observed in *Cochliomyia hominivorax* (Diptera), mtDNA has a bias in nucleotide composition, which perhaps has led to an AT-rich genome (Lessinger et al., 2000). This bias is higher at the third position in the codons of protein coding genes due to lower selection and mutation (Jermiin et al., 1994). For all other orders, the A+T composition is very close to the mean observed in insect mtDNA, at 77% (Stewart and Beckenbach, 2005). mtDNA from Orthoptera comprises almost twice the GC percent compared to Hymenoptera with 29% in *G. orientalis*. All Dipterans except *D. melanogaster* mtDNA show a high level of GC content and Coleopterans show 22% (in *P. rufa*) to 29% (in *T. castaneum)* (Table 1).

LSU but not SSU rRNA sequence phylogenetic tree concurs with morphologic classification.

An unrooted phylogenetic tree showing the relationship between Diptera, Hymenoptera, Orthoptera and Coleopterans based on mitochondrial (large subunit) LSU rRNA and (small subunit) SSU rRNA sequences were generated using Clustal X (Figure 2a and 2b) respectively. Phylogenetic trees derived from mtDNA large subunit RNA and small subunit RNA sequences placed their host insect species into taxonomic groups similar to that of morphologically derived characteristics. LSU rRNA sequences yielded a tree with branching patterns reflecting the expected pattern as insect species belonging to different orders were put into separate clades (Figure 2a). *A. mellifera* and *M. bicolor* clustered together with a supporting bootstrap value of 97%. Both drosophila and anopheles species were in the same clade with supporting bootstrap value of 100%. All three species from the Coleoptera and two species from Orthoptera grouped into two separate branches with bootstrap values of 100% each suggesting close relatedness among the species.

A phylogenetic tree constructed using SSU rRNA sequences yielded an identical tree profile except for *P. rufa*, which clustered with *M. bicolor, A. mellifera, T. castaneum,* and *C. duodecimpunctat*a. This may be due to length of the DNA, which is 1,859 bp longer than that of the other two Coleoptera species. In addition, there are no recognizable ORFs for a length of 1720 base pairs of *P. rufa* mtDNA (from bp 1,213 to 2,933), as these sequences did not yield any protein coding sequences with ERGO or NCBI Blasts. The size variation may be due to intergenic spacers and the size of the control reg-

ion (Flook et al., 1995). Hence, both insect species from Hymenoptera and Coleoptera orders were grouped in the same clade with supporting bootstrap value of 76% whereas drosophila and anopheles grouped into a distinct cluster with 100% bootstrap value.

Dot-plot comparison reveals highly conserved mtDNA genome synteny.

We initially analyzed whole mtDNA sequences using a dot-matrix software program, which aligns nucleotide *vs* nucleotide between two whole genome sequences. Sequences between and within each order were plotted to visualize sequence breaks and sequence co- linearity. Figure 3 shows the dot-matrix plot generated with a window size of 22 and maximum of one mismatch. Based on the sequence similarity, insect species belonging to four different orders in general appear to be closely related. Not surprisingly, the dot-plot matrix shows a clear picture of gene organization within orders such as *A. mellifera* vs *M. bicolor*, *D. mauritiana* vs *A. gambiae, G. orientalis* vs *L. migratoria*, C*. duodecimpunctata* vs *T. castaneum* (Figure 3*).* Most of the sequence segments of one genome can be found on the other mitochondrial genome and genes are organized in the same order and direction, thereby producing a single straight continuous line with sparse breaks. Between Diptera and Hymenoptera (*D. melanogaster and A. mellifera*) a distinct co-linear sequence alignment was produced. Similarly between Diptera vs Orthoptera (*D. simulans vs G. orientalis*), Diptera *vs* Coleoptera (*D. melanogaster* vs *P. rufa*), Hymenoptera vs Coleoptera (*A. mellifera vs C. duodecimpuntata)*, Hymenoptera vs Orthoptera (*A. mellifera* vs *L. migratoria*), Orthoptera *vs* Coleoptera (*G. orientalis* vs *T. castaneum)* also showed a clear picture of gene organization between species belonging to different orders (not shown in the Figure 3).

Gene content and gene order of mtDNA

With some notable exceptions, the order of mitochondrial genes is highly conserved (Boore, 1999). From our comparative studies, differences in the genes for tRNA have been observed. The typical gene content for an animal mtDNA was observed for all 12 insect species. Almost all species of insect mtDNA encode large and small rRNA, 22 tRNAs, and 13 to 14 polypeptides that participate in the oxidative phosphorylation subunits of NAD dehydrogenase (NADH 1-5 and nad 4L), cytochrome b (Cytb), three cytochrome oxidases (COX 1-3), an hypothetical protein and an ATPase (ATP6) (Figure 1).

Non-coding mtDNA genes

In general, 22 bacterial type tRNA species are found in insect mtDNA. Two ORFs for leucine and two for serine, and one each for the 18 tRNA for different amino acids, were identified. Most insect mtDNA have 22 tRNA ORFs except for *M. bilolor*, which has 19 and lacks both $tRNA^{Cys}$ and $tRNA^{Gln}$ genes. Although tRNA species are identical, their position and the transcription orientations are different, reflecting differential transcriptions. The tRNA gene positions are subject to change more often than the ORFs for rRNAs and protein encoding genes (Thao et al., 2004). Our analysis also revealed a common signature for arrangement of ribosomal RNAs for insects. All 12 species in this study show $tRNA^{val}$ located between ORFs for 23S rRNA and 16S rRNA and are in the same orientation as ND1 except for *M. bicolor* where both 23S and 16S rRNA ORFs are in opposite orientation. However, the $tRNA^{Val}$ ORF is reversed. Interestingly, such differences are not revealed by the dot-plot matrix (Figure 3).

a) Order Hymenoptera: Insect species in this order show considerable differences in the number of tRNA genes, their genomic positions as well as direction of tRNA genes. In *A. mellifera*, all the tRNA genes are encoded on the opposite strand except in the case of $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Arg}$, $tRNA^{Phe}$, $tRNA^{His}$, $tRNA^{Pro}$ and $tRNA^{Val}$. However, in *M. bicolor*, an ORF for $tRNA^{Cys}$ is absent. ORFs for $tRNA^{Tyr}$, $tRNA^{Arg}$, $tRNA^{Lys,}$ $tRNA^{Phe}$, $tRNA^{Pro}$, $tRNA^{His}$, $tRNA^{Thr}$ and $tRNA^{Val}$ are transcribed in the same orientation. However, like other genomes, insects in the Hymenoptera order have maintained 23S rRNA-$tRNA^{val}$-16S rRNA signature sequence. Interestingly, all 12 insect species in our analysis, except *M. bicolor*, have duplications of ORFs for $tRNA^{Ser}$ and $tRNA^{Leu}$ genes.

Order Diptera: Among Diptera, *Drosophila spp* and *Anopheles spp* show very close genomic positions for 22 tRNA genes except for the position between $tRNA^{Arg}$ and $tRNA^{Ala}$, which are switched (Table 2). In Drosophila, all tRNA genes are encoded on the opposite strand except for $tRNA^{Gln}$, $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Phe}$, $tRNA^{His}$,$tRNA^{Pro}$ and $tRNA^{Val}$. In *Anopheles spp* mtDNA, all tRNA genes are encoded as in *Drosophila spp* except $tRNA^{Ser}$. $tRNA^{Ser}$ is transcribed in the opposite orientation in anopheles, however it is encoded in the same orientation in *Drosophila* spp.

Order Orthoptera: Species in this order show very similar arrangement in the direction of coding of tRNA genes and their genomic positions as in Diptera. All tRNA genes are encoded on the opposite strand except in case of $tRNA^{Gln}$, $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Phe}$, $tRNA^{His}$, $tRNA^{Pro}$ and $tRNA^{Val}$. Both *L. migratoria* and *G. orientalis* exhibit close arrangements of tRNA genes except the positions of $tRNA^{Lys}$ and $tRNA^{Asp}$ are switched between them.

Order Coleoptera: Among Coleoptera, all the species have retained identical tRNA positions; however, they differ in their transcription orientation. In *P. rufa* and *C.*

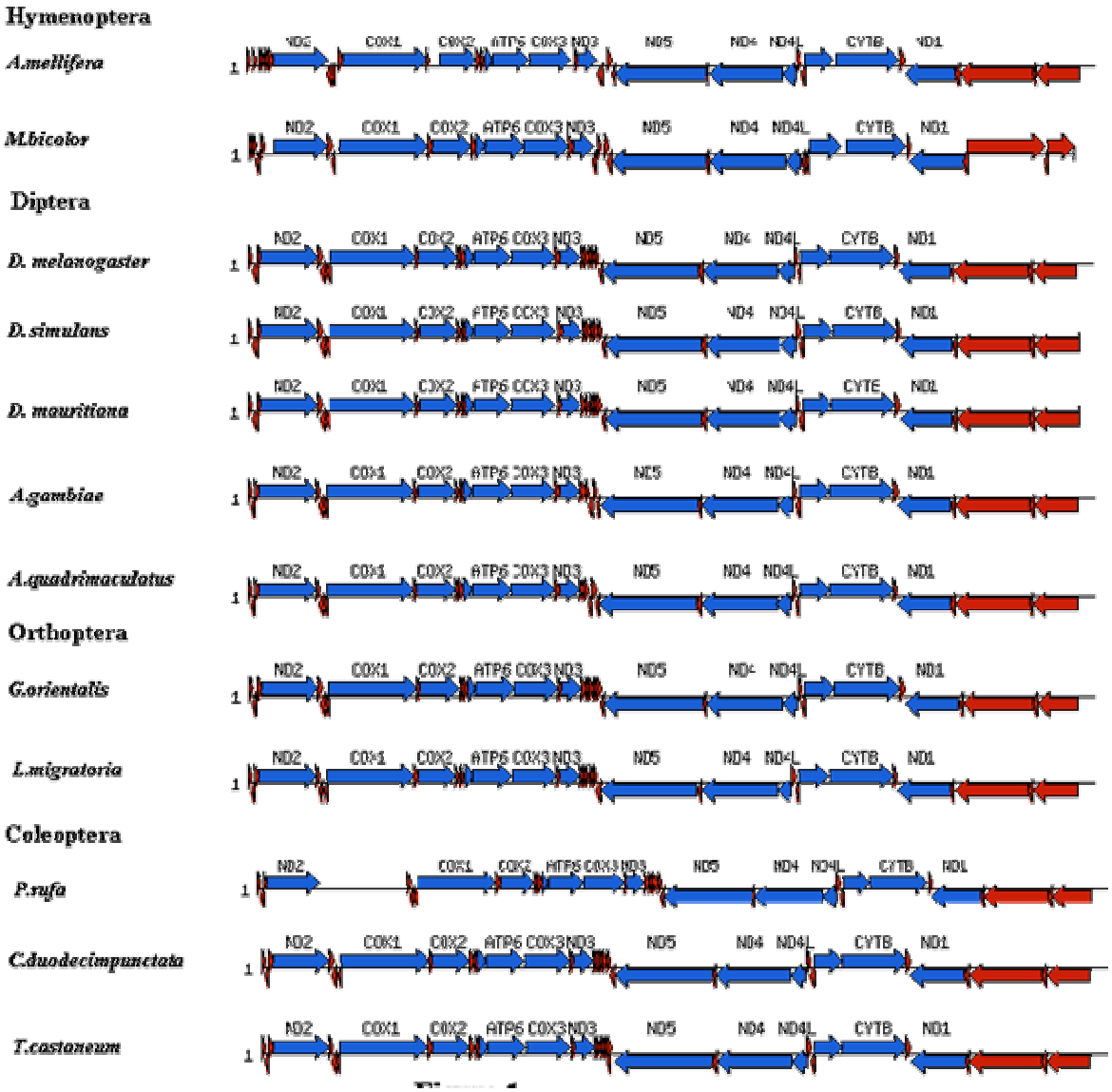

Figure 1. Organization of mitochondrial genes. Gene organization of mitochondrial coding and non-coding gene of 12 insect species belonging to various insect orders are shown.

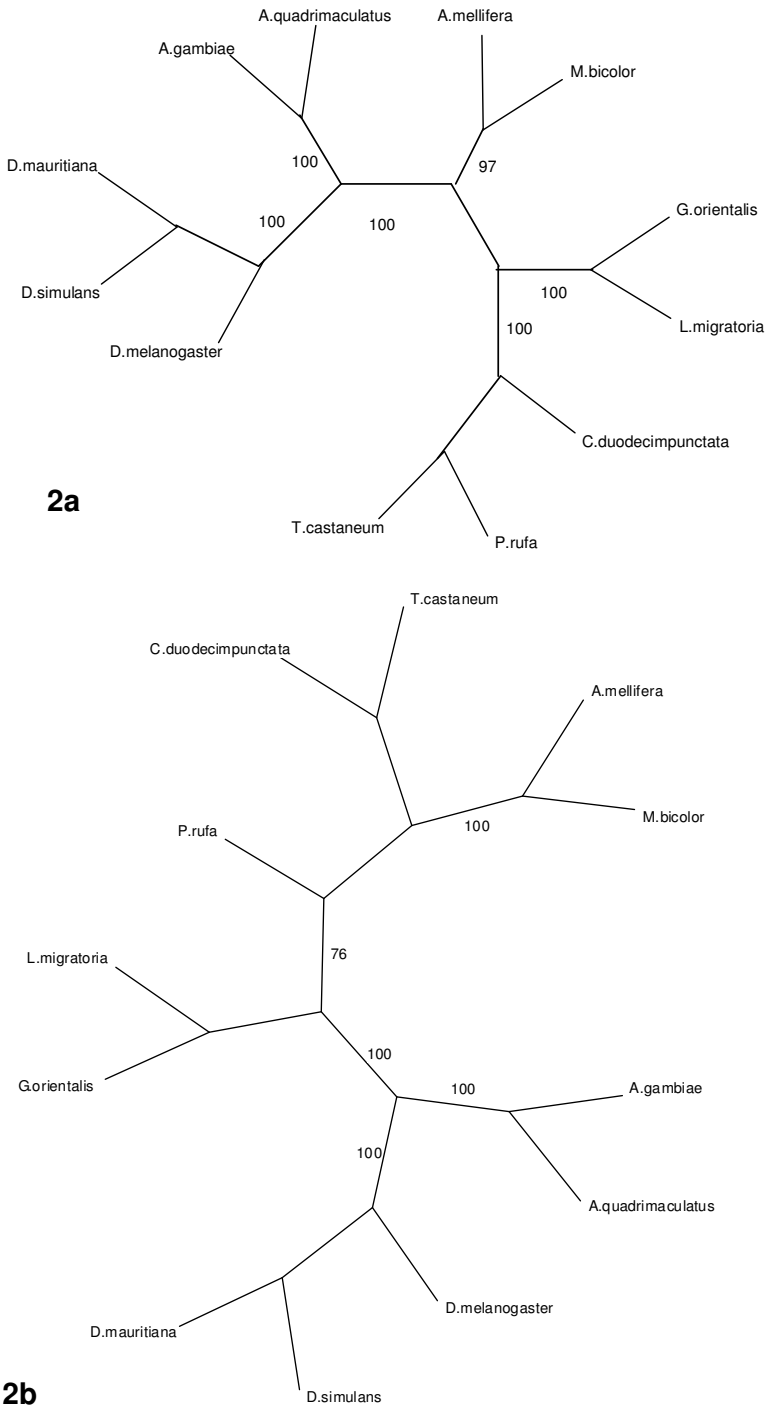

Figure 2 (a and b). An Unrooted phylogenetic tree showing the relationships of members of the four insect orders. The tree is inferred from mtDNA-LSU (2a) and SSU (2b) rDNA sequences using neighbor-joining method. The number above each branch is the percent of 100 bootstrap replications that supported the branch when the data set was analyzed by DNADIST in the PHYLIP version 3.5c (Felsenstein, 1989).

duodecimpunctata, most tRNA genes are encoded in the same orientation as in the case of species belonging to Orthoptera. However, in the case of *T. castaneum*, $tRNA^{Glu}$ is coded in the same orientation unlike in the other two species (*P. rufa and C. duodecimpunctata*).

Table 2. Mitochondrial tRNA genes number and orientation with respect to rest of the genes for insect mitochondrial species. Except *M. bicolor*, all other insect genomes encode 22 tRNA species. Difference between and among insect orders with respect to position and transcription orientation is indicated. The arrows indicate directionality with respect to the rest of the genes.

Order	Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Hymenoptera	**A.mellifera**	Glu	Ser	Met	Gln	Ala	Ile	Cys ←	Tyr ←	Trp	Leu	Asp	Lys	Gly	Arg ←	Asn	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
	M.bicolor	Ile	Ala	Lys ←	Met	Trp	Tyr ←	Leu	Asp	Gly	Arg ←	Asn	Glu				Phe ←	His ←	Thr ←	Pro ←	Ser	Leu ←	Val ←
Diptera	*D.melanogaster*	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Lys	Asp	Gly	Ala	Arg	Asn	Ser	Glu	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
	D.simulans	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Lys	Asp	Gly	Ala	Arg	Asn	Ser	Glu	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
	D.mauritiana	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Lys	Asp	Gly	Ala	Arg	Asn	Ser	Glu	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
	A.gambiae	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Lys	Asp	Gly	Arg	Ala	Asn	Ser ←	Glu	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
	A.quadrimaculatus	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Lys	Asp	Gly	Arg	Ala	Asn	Ser ←	Glu	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
Orthoptera	*G.orientalis*	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Lys	Asp	Gly	Ala	Arg	Asn	Ser	Glu	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
	L.migratoria	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Asp	Lys	Gly	Ala	Arg	Asn	Ser	Glu	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
Coleoptera	*P.rufa*	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Lys	Asp	Gly	Ala	Arg	Asn	Ser	Glu	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
	C.duodecimpunctata	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Lys	Asp	Gly	Ala	Arg	Asn	Ser	Glu	Phe ←	His ←	Thr	Pro ←	Ser	Leu ←	Val ←
	T.castaneum	Ile	Gln ←	Met	Trp	Cys ←	Tyr ←	Leu	Lys	Asp	Gly	Ala	Arg	Asn	Ser	Glu ←	Phe	His ←	Thr	Pro ←	Ser	Leu ←	Val ←

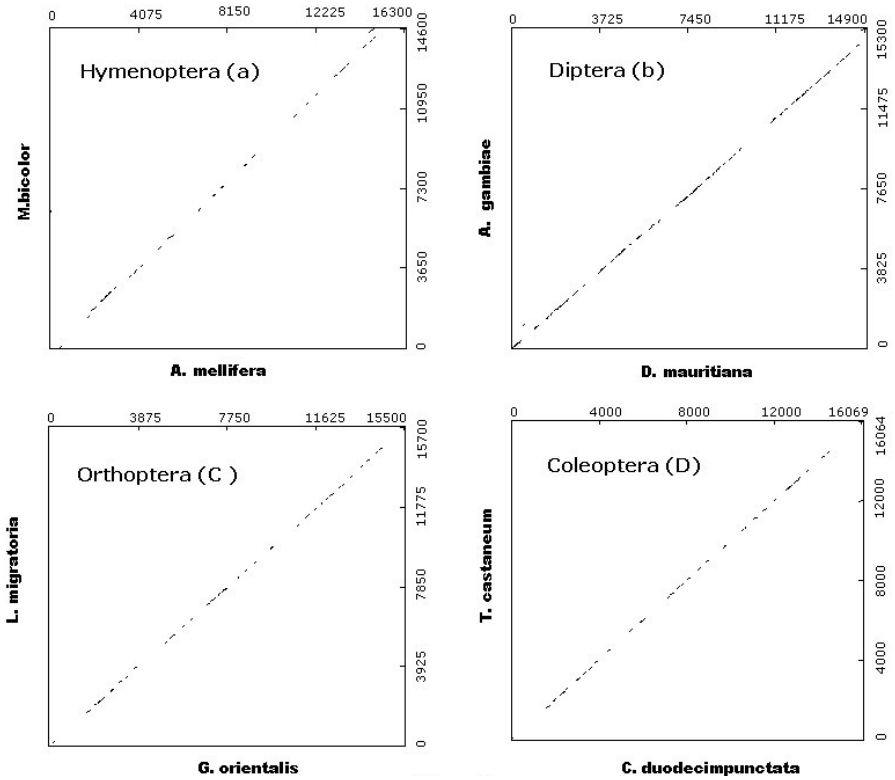

Figure 3. Dot matrix plots comparison of two different mtDNA genomes of insects with in each order. The dot plot was computed using maximum mismatch of one mitochondrial sequence. Similarity search was performed with k-tuple = 5 and a window size of = 22nt.

Mitochondrial protein coding genes

In large part, eukaryotic mtDNA-specified proteins are components of respiratory complexes I (NADH: ubiquinone oxidoreductase encoded by nad genes), II (succinate: ubiquinone oxidoreductase; sdh), III (ubiquinol: cytochrome c oxidoreductase; cob) and IV (cytochrome c oxidase; cox) of the electron transport chain as well as complex V (adenosine triphosphate synthase; atp) (Gray et al. 1999). Interestingly, insects code neither for *SDH* nor *COB* genes. *COB* genes are encoded in both fission and budding yeasts.

The order of protein coding genes in all four insect orders was identical to that of other arthropod species. Interestingly, each species of insect exhibits a variety of initiation and termination codons for diverse protein coding genes. ATP6 and CYTB genes are always initiated by ATG codons in all four orders. Although Hymenoptera show a number of different initiation codons, all protein-coding genes end with termination codon TAA. Only three genes of *A. mellifera* and *M. bicolor* have ATG (ATP6, COX3, CYTB) as initiation codon whereas other genes initiate either with ATC (as in ND2) or ATA (as in COX1, ND3, and ND4) or ATT (as in COX2, ND5, and ND1) codons. Among Dipterans, COX2 and ATP6 genes are always initiated with ATG, and terminated with either CTT, ATT, TTA, TAA or ACT codons. Among Orthoptera, at least 9 genes initiated with start codon ATG (COX1, COX2, COX3, ATP6, ND1, ND4, ND4L, ND5 and CYTB) in the case of *G. orientalis* and at least 8 genes (COX2, COX3, ND1, ND2, ND4, ND4L, ATP6 and CYTB) in the case of *L. migratoria*. Among Coleoptera, COX3, ATP6, ND4 and CYTB genes use ATG as initiation codon and terminate by TAA, ATT, TTT, ACT, or TAG codons. Although mtDNA genes often terminate with stop codons, the transcripts of these genes contain the complete termination signal, UAA, generated by polyadenylation after cleavage of the polycistronic RNA as with metazoan mtDNA (Lavrov et al., 2002).

Order Hymenoptera: In *A. mellifera,* like in most insect mtDNA, all four types of genes are found: oxidoreductase, cytochrome oxidase, cytochrome b and ATP synthase. NADH-ubiquinone oxidoredutase (EC. 1.6.5.3) consists of six subunits 1,2,3,4, 4L and 5. Except for four ORFs ND1, ND4, ND4L, and ND5, all the ORFs are oriented on the opposite strand compared to

the rest of the coding genes. Both LSU rRNA and SSU rRNA are arranged in the same orientation as ND1 and are interrupted by $tRNA^{Val}$. Cytochrome c oxidase (EC 1.9.3.1) polypeptides are arranged and transcribed in the same orientation. COX1 and COX2 are interrupted by $tRNA^{leu}$. COX2 and COX3 are separated by two tRNAs ($tRNA^{Asp}$ and $tRNA^{Lys}$) and an ORF for ATP synthase A chain (EC 3.6.3.14). ND2 and ND3 ORFs are clustered along with seven tRNA, three cytochrome c oxidases (COX1, COX2, COX3) and an ORF for ATP synthase A chain (ATP6). Three tRNA ($tRNA^{Arg}$, $tRNA^{Asn}$, $tRNA^{Phe}$) separate ORFs for ND3 and ND5 and one of them ($tRNA^{Asn}$) is encoded in the opposite direction (Figure 3).

M. bicolor has four similar types of protein coding genes as *A. mellifera.* All the ORFs are encoded in the opposite strand compared to the rest of the coding genes except ORFs ND1, ND4, ND4L and ND5 as in *A. mellifera*. Interestingly, both LSU rRNA and SSU rRNA are arranged in opposite orientation and sandwiched by a gene for $tRNA^{Val}$. Unlike in *A. mellifera*, COX2 and COX3 are separated by an ORF for $tRNA^{Asp}$ and an ORF for ATP synthase A chain. ND2 and ND3 are separated only by 5 tRNA (7 in case of *A. mellifera*), three Cytochrome c oxidases (COX) and an ORF for ATP synthase. Four tRNA interrupt ND3 and ND5 (three in case of *A. mellifera*; $tRNA^{Arg}$, $tRNA^{Asn}$, $tRNA^{Glu}$, $tRNA^{Phe}$). Only four tRNA precede ND2 compared to six in *A. mellifera*.

Order Diptera: Insect species belonging to this order have identical gene types as in Hymenoptera. Unlike in *M. bicolor*, both LSU and SSU rRNA are arranged in the same orientation as ND1 and are interrupted by $tRNA^{Val}$ as in Hymenoptera. Cytochrome c oxidase (EC 1.9.3.1) polypeptide is arranged and transcribed in the same orientation as in Hymenoptera and COX1 and COX2 are interrupted by $tRNA^{leu}$. Two tRNAs ($tRNA^{Asp}$ and $tRNA^{Lys}$) and an ORF for ATP synthase A chain flank COX2 and COX3. Unlike in *M. bicolor*, there are three tRNA positioned between ND2 and COX1. There are six tRNA interrupting ND3 and ND5. Interestingly, there are no recognizable ORFs for a length of 4601 base pairs of *D. melanogaster* mtDNA (from bp 14,916 to 19,517), as these sequences did not match with any of the known gene sequences from other known and analyzed genomes. Both *D. simulans* and *D. mauritiana* show great similarity with *D. melanogaster* in terms of types and arrangement of coding genes. One of the conspicuous differences between the Diptera and Hymenoptera in terms of arrangement of coding genes seems to be the number of tRNA sandwiching between ND3 and ND5. Six tRNA interrupt ND3 and ND5 compared to three to four tRNA in case of Hymenoptera. Unlike in Hymenoptera, only three tRNA precede ND2. In both *D. simulans and D. mauritiana*, $tRNA^{Arg}$ is encoded in the opposite direction in all the insect species belonging to the order Diptera, Coleoptera and Orthoptera except Hymenoptera.

A. gambiae and *A. quadrimaculatus* show similar gene organization as in Drosophila, but exhibit clear differences with order Hymenoptera. Unlike in *M. bicolor*, both LSU rRNA and SSU rRNA are arranged in the same orientation as ND1. Unlike in *A. mellifera* or *M. bicolor*, only three tRNA precede ND2. In addition at least 6 tRNA interrupt ND3 and ND5 when compared only 3 to 4 tRNA separate these two ORFs in case of Hymenoptera.

Order Orthoptera: Both *G. oreintalis* and *L. migratoria* exhibit very similar arrangement of all the four type of ORF that code for proteins as in Diptera. All the ORFs are encoded in the opposite strand compared to the rest of the coding genes except in case ORFs ND1, ND4, ND4L and ND5 as in most other insect species. As in Diptera, both LSU rRNA and SSU rRNA are arranged in the same orientation as ND1 and are separated by $tRNA^{Val}$. Both species of Orthopterans show differences with Hymenoptera in terms of arrangement of coding genes. Unlike in Hymenoptera, only three tRNA precede ND2. In case of Hymenoptera four to seven tRNA precede ND2. Only three to four tRNA interrupt between ND3 and ND5. However, in case of Orthopterans like in Diptera, there are six tRNA positioned between ND3 and ND5. Two tRNA and an ORF for ATP synthase A chain separate COX2 and COX3 as in *A. mellifera*, and species of insects in Diptera. Between ND1, ND4L there is an ORF for cytochrome b, a hypothetical protein (RIIE00011: ND6: NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3)) and three tRNA ($tRNA^{Thr}$, $tRNA^{Pro}$, $tRNA^{Ser}$) as in case of both Hymenoptera and Diptera.

Order Coleoptera: All the three species, *P. rufa, C. duodecimpunctata* and *T. castaneum*, have identical protein coding genes similar to other insect species. In all three species, the ORFs are encoded in the opposite strand compared to the rest of the coding genes except for ORFs ND1, ND4, ND4L and ND5 as in other insect mtDNA. Both LSU rRNA and SSU rRNA are arranged in the same orientation as ND1 and separated an ORF for $tRNA^{Val}$. Interestingly, Coleopterans show more similarity with species belonging to Diptera and Orthoptera than with Hymenoptera with respect to gene organization. Like in Diptera and Orthoptera, only three tRNA precede ND2. However, four to seven tRNA precede ND2 in Hymenoptera. Again like in Diptera and Orthoptera, six tRNA positioned between ND3 and ND5. However, only three to four tRNA interrupt ND3 and ND5 in case of Hymenoptera. Also, two tRNA and an ORF for ATP synthase A chain separate COX2 and COX3 in Diptera and Orthoptera. Between ND1, ND4L there is an ORF for cytochrome b, a hypothetical protein (RIIE00011; ND6: NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3)) and three tRNA ($tRNA^{Thr}$, $tRNA^{Pro}$, $tRNA^{Ser}$) present like in rest of the insect orders. Interestingly, there are no recognizable ORFs for a length of 1720 base pairs of *P.*

rufa mtDNA (from bp 1213 to 2933), as these sequences did not match with any of the known gene sequences from other known and analyzed genomes

Conclusion

In our conclusion, Comparative analysis of insect mitochondria has led to identification of differences in gene organization of mtDNA. This study indicates that, although insect mtDNA have different GC content, the gene types are very similar across several insect orders. Significant differences in GC content may suggest multiple mitochondrial ancestors in spite of retaining the similar genome content. Despite overall chromosomal synteny among these species, clearly the gene order and their orientation are different among them. Such orientations may again suggest differences in transcription and gene regulation. Further research involving comparative genome analysis of a large number of insect orders with substantial number of insect species will bring light on the possibility of multiple mitochondrial ancestors.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Robert Haselkorn (University of Chicago) for his valuable comments and suggestions on our manuscript. This work was supported in part from a grant from the Roosevelt University to S.Chandra via the faculty research program, 2005.

REFERENCES

Bae JS, Kim I, Sohn HD, Jin BR (2004). The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. Mol. Phylogenet. Evol. 32: 978-985.

Ballard JWO (1999a). Mitochondrian *Drosophila simulans*. Direct submission. Direct submission (NC_005781)

Ballard JWO (1999b). Mitochondrian mitochondrion *Drosophila mauritiana* (fruitfly). Direct submission (NC_005779)

Beard CB, Hamm DM, Collins FH (1993). The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol. Biol. 2: 103-124.

Boore JL (1999). Animal mitochondrial genomes. Nuclei. Acids. Res. 27: 1767-1780.

Bullerwell CE, Leigh J, Forget L, Lang BF (2003). A comparison of three fission yeast mitochondrial genomes. Nucleic Acids Res. 31: 759-68.

Chandra BCS, Geetha L, Abraham VA, Karanth P, Thomas K, Srinivasan MV, Gadagkar R (1998). Uniform Discrimination of Pattern Recognition by Honeybees. Animal Behavior 56, 1391-1398.

Chandra BCS, Hosler JS, Smith BH (2000). Heritable Variation for Latent Inhibition and Its Correlation with Reversal Learning in Honeybees (Apis Mellifera). J. Comparative Psychol. 114: 86-97.

Chandra BCS, Hunt GJ, Cobey S, Smith BH (2001). Quantitative Trait Loci (Qtl) Associated with Reversal Learning and Latent Inhibition in Honey Bees (*Apis Mellifera*). Behavior Genet. 31: 275-285.

Chandra BCS, Singh S (2005). Chemosensory Processing in the Fruit Fly Drosophila Melanogaster: Generalization of a Feeding Response Reveals Overlapping Odor Representations. J. Biosci. 30: 679-688.

Crozier RH, Crozier YC (1993). The mitochondrial genome of the honeybee *Apis mellifera*: Complete sequence and genome organization. Genetics. 133: 97-117.

Felsenstein J (1989). PHYLIP Phylogeny Inference Package. Cladistics. 5: 164-166.

Flook PK, Rowell CH, Gellissen G (1995). The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. 41: 928-941.

Friedrich M, Muqim N (2003). Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castanaeum*. Mol. Phylogenet. Evol. 26: 502-512.

Fukuhara H, Sor F, Drissi R, Dinouel N, Miyakawa I, Rousset S, Viola AM (1993). Linear mitochondrial DNAs of yeasts: frequency of occurrence and general features. Mol. Cell Biol. 13: 2309-2314.

Gray MW, Burger G, Lang BF (1999). Mitochondrial evolution. Science. 283: 1476-1481.

Gray MW, Burger G, Lang BF (2001). The origin and early evolution of mitochondria. Genome Biol. 2: 1-5.

Huang Y, Zhang L (2004). Rapid and sensitive dot-matrix methods for genome analysis. Bioinformatics. 20: 460-466.

Hudspeth MES (1995). The fungal mitochondrial genome--a broader perspective. In: DK Aurora, PD Bridge and D Bhatnagar, eds. Handbook of Fungal Biotechnology. 1st ed. New York: Marcel Dekker, Inc. pp. 213-241.

Jermiin LS, Graur D, Lowe RM, Crozier RH (1994). Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. J. Mol. Evol. 39: 160-73.

Kim I, Cha SY, Yoon MH, Hwang JS, Lee SM Jin BR (2005). The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Gryllotalpidae: Orthoptera). Direct submission. (NC_006678)

Langkjaer RB, Casaregola S, Ussery DW, Gaillardin C, Piskur J (2003). Sequence analysis of three mitochondrial DNA molecules reveals interesting differences among Saccharomyces yeasts. Nucleic Acids Res. 31:3081-3091.

Lavrov DV, Boore JL, Brown WM (2002). Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: Duplication and nonrandom loss. Mol. Biol. Evol. 19: 163–169.

Lessinger AC, Martins JAC, Lemos TA, Kemper EL, da Silva FR, Vettore AL, Arruda P, Azeredo-Espin AM (2000). The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). Insect Mol Biol. 9: 521-9.

Lewis DL, Farr CL, Kaguni LS (1995). *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Mol. Biol. 4: 263-278.

Martin FN (1995). Linear mitochondrial genome organization in vivo in the genus Pythium. Cur. Genet. 28: 225-234.

Mitchell SE, Cockburn AF, Seawright JA (1993). The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. Genome. 36: 1058-1073.

Overbeek R, Larsen N, Walunas T, D'Souza M, Pusch G, Selkov E Jr, Liolios K, Joukov V, Kaznadzey D, Anderson I, Bhattacharyya A, Burd H, Gardner W, Hanke P, Kapatral V, Mikhailova N, Vasieva O, Osterman A, Vonstein V, Fonstein M, Ivanova N, Kyrpides N (2003). The ERGO genome analysis and discovery system. Nucleic Acids Res. 1: 164-71.

Page RDM (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences. 12: 357-358.

Palmer JD (1990). Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet. 6: 115-120.

Rueppell O, Chandra BCS, Pankiw T, Fondrk MK, Beye M, Hunt J, Page RE (2006). The Genetic Architecture of Sucrose Responsiveness in the Honey Bee (Apis Mellifera L.). Genetics 172: 243-251.

Rycovska A, Valach M, Tomaska L, Bolotin-Fukuhara M, Nosek J (2004). Linear versus circular mitochondrial genomes:

intraspecies variability of mitochondrial genome architecture in *Candida parapsilosis* Microbiology. 150: 1571-1580.
Silvestre D, Arias MC (2002). The mitochondrial genome of *Melipona bicolor* (Apidae, Meliponini) Unpublished (NC_004529)
Sonnhammer EL, Durbin R (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene. 167: GC1-10.
Stewart JB,Beckenbach AT (2005). Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). Genome. 48: 46-54.
Stewart JB, Beckenbach AT (2003). Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle *Crioceris duodecimpunctata* Mol. Phylogenet. Evol. 26: 513-526.
Thao ML, Baumann L, Baumann P (2004). Organization of the mitochondrial genomes of whiteflies, aphids and psyllids (Hemiptera, Sternorrhyncha). BMC Evolutionary Biology. 4: 1471-2148.
Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matirx choice. Nucleic Acids Res. 22: 4673-4680.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-82.