The mitochondrial genome of the entomophagous endoparasite Xenos vesparum (Insecta: Strepsiptera)

LBNL-59178

The mitochondrial genome of the entomophagous endoparasite *Xenos vesparum* (Insecta: Strepsiptera)

Antonio Carapelli, Laura Vannini, Francesco Nardi, Jeffrey L. Boore, Laura Beani, Romano Dallai, and Francesco Frati

The mitochondrial genome of the entomophagous endoparasite *Xenos vesparum* (Insecta: Strepsiptera)

(Submit to Gene)

Keywords: *mtDNA, gene order, codon usage, tRNA, A+T content, strand nucleotide bias.*

Antonio Carapelli[a], Laura Vannini[a], Francesco Nardi[a], Jeffrey L. Boore[b], Laura Beani[c], Romano Dallai[a], and Francesco Frati[a,*]

[a] *Department of Evolutionary Biology, University of Siena, Siena, Italy*

[b] *U.S. Department of Energy Joint Genome Institute, and Lawrence Berkeley National Laboratory, University of California, Berkeley, U.S.A.*

[c] *Department of Animal Biology and Genetics "Leo Pardi", University of Florence, Florence, Italy*

Abbreviations: *atp6* and *atp8*, genes for ATP synthase subunits 6 and 8; *cox1-3*, genes for subunits I-III of cytochrome *c* oxidase; *cob*, gene for cytochrome *b*; *nad1-6* and *nad4L*, genes for subunits 1-6 and 4L of NADH dehydrogenase; *rrnL* and *rrnS*, genes for the samll and large subunits of ribosomal RNA; *trnX*, genes encoding for transfer RNA molecules with corresponding amino acids denotaed by the one-letter code and anticodon indicated in parentheses (*xxx*) when necessary; tRNA-X, transfer RNA molecules with corresponding amino acids denoted with a one-letter code; bp, base pair; mtDNA, mitochondrial DNA; PCR, Polymerace Chain Reaction;

Address for correspondence:

Francesco Frati
Department of Evolutionary Biology, University of Siena, Via A. Moro 2, 53100, Siena, Italy
Tel.: +39-0577-234417; Fax: +39-0577-234476
email: frati@unisi.it

Abstract

In this study, the nearly complete sequence (14,519 bp) of the mitochondrial DNA (mtDNA) of the entomophagous endoparasite *Xenos vesparum* (Insecta: Strepsiptera) is described. All protein coding genes (PCGs) are in the arrangement known to be ancestral for insects, but three tRNA genes (*trnA*, *trnS(gcu)*, and *trnL(uag)*) have transposed to derived positions and there are three tandem copies of *trnH*, each of which is potentially functional. All of these rearrangements except for that of *trnL(uag)* is within the short span between *nad3* and *nad4* and there are numerous blocks of unassignable sequence in this region, perhaps as remnants of larger scale predisposing rearrangements. *X. vesparum* mtDNA nucleotide composition is strongly biased toward As and Ts, as is typical for insect mtDNAs. There is also significant strand skew in the distribution of these nucleotides, with the J-strand being richer in A than T and in C than G, and the N-strand showing an opposite skew for complementary pairs of nucleotides. The hypothetical secondary structure of the 16S rRNA has also been reconstructed, obtaining a structural model similar to that of other insects.

1. **Introduction**

Animal mitochondrial DNA (mtDNA) is nearly always a covalently closed circular molecule, approximately 16 Kb in size, and extremely conservative in its gene content. Mitochondrial genomes typically contain 37 genes encoding for 13 proteins involved in oxidative phosphorylation (OXPHOS) (*cox1-3*, *cob*, *nad1-6*, *nad4L*, *atp6*, *atp8*), two ribosomal RNA subunits (*rrnS* and *rrnL*), and 22 transfer RNAs (*trnX*, where X refers to the corresponding amino acid) (see Boore, 1999, for a review). Mitochondrial genes may be located on different strands, which are conventionally denoted for arthropods as the J-strand and N-strand (Simon et al., 1994), with the J-strand defined as that with the most Protein Coding Genes (PCGs). In the few cases where it has been studied, genes are transcribed as a polycistronic mRNA, which is post-transcriptionally processed into gene-specific mRNAs (see Boore, 1999).

Except for the "control" region (sometimes called the AT-rich region or D-loop region), which is involved in the regulation and initiation of mitochondrial transcription and replication (Wolstenholme, 1992), the entire genome comprises compactly arranged coding sequences with no introns and with very few intergenic nucleotides (Boore, 1999). Neighboring genes may even slightly overlap in some cases.

Another remarkable feature of mtDNA is the nucleotide bias it exhibits in some taxa. In insects, for example, the mitochondrial genome is usually very rich in As and Ts, with the highest nucleotide bias observed in some Hymenoptera (A+T > 80%) (Crozier & Crozier, 1993; Simon et al., 1994). Further, there is often a bias between the two strands, with one being rich in A and C, and the other being rich in T and G (Perna and Kocher, 1995), perhaps due to asymmetrical mutational bias at replication and/or transcription (Hassanin et al., 2005).

Mitochondrial genes are very commonly used molecular markers for phylogenetic studies at various taxonomic levels (Simon et al., 1994). Its advantages include the (generally) strict orthology of genes, the clonal pattern of maternal inheritance (but see Rawlings et al. 2003), the lack of recombination (although this issue is contended), and its generally rapid rate of evolution (Simon et al., 1994). In addition, comparisons of the relative arrangement of mitochondrial genes has been extremely effective for resolving some phylogenetic relationships (e.g., Boore et al., 1998; Boore and Brown, 1998; Lavrov et al., 2004).

Although the mechanisms responsible for gene translocations are not yet completely understood, several models of mitochondrial genome rearrangements have been proposed (Macey et al., 1997; Boore, 2000; Lavrov et al., 2002), and it has been suggested that some regions of the mitochondrial genome are more subject to such rearrangements than others. In this respect, tRNA-encoding genes are usually more frequently rearranged than either rRNAs and PCGs (Boore, 1999; Curole & Kocher, 1999), with most changes observed in arthropods involving tRNA genes near the control region or in the cluster that is most commonly ordered *trnA*, *trnR*, *trnN*, *trnS*, *trnE*, *trnF* (Boore, 1999).

Within arthropods, the widest sample size of complete mitochondrial genomes so far collected is found in hexapods, with over forty sequences available in GeneBank, and this figure is increasing at a quick pace. However, some orders still remain to be investigated. In this study we report the almost complete mitochondrial genome of the species *Xenos vesparum* (Stylopidae), from the previously uninvestigated order Strepsiptera.

The Strepsiptera are a small, cosmopolitan insect order, whose members exhibit extreme sexual dimorphism, with short-living flying males and neotenic endoparasite females (Kathirithamby, 1989; Beani et al., 2005). Besides the evident morphological oddities, strepsipterans also exhibit other peculiarities at the genetic and molecular level, such as having the smallest nuclear genome recorded among insects (Johnston et al., 2004), unique

insertions in the nuclear 18S rRNA (Gillespie et al., 2005) and accelerated rates of molecular evolution in the 18S rRNA gene (Carmean and Crespi, 1995; Huelsenbeck, 1998). Recent studies using both morphological and molecular data (including mitochondrial gene sequences) have generated four contending hypotheses for the phylogenetic placement of the Strepsiptera: as sister to all the holometabolous insects or, within the Holometabola, as part of the Coleoptera, sister-group to the Coleoptera, or sister-group to the Diptera (Kristensen, 1991; Whiting et al., 1997; Huelsenbeck, 1997; Hwang et al., 1998; Rokas et al., 1999; Grimaldi et al., 2005).

We therefore set out to sequence the complete mitochondrial genome of one member of the Strepsiptera, in order to assess whether the same peculiarities affecting the nuclear genome (Johnston et al., 2004; Gillespie et al., 2005) also influence the mitochondrial genome, and to provide the genetic information necessary for a phylogenetic analysis based on complete mitochondrial genomes (not performed here). Our data have revealed a novel gene order, with the autapomorphic translocation of three tRNA-encoding genes with respect to the putative ancestral hexapod gene arrangement (Lewis et al., 1995), four unusually large (83-173 bp) non-coding sequences, and three copies of *trnH* (*trnH1*, *trnH2*, *trnH3*).

2. Materials and methods

Parasitized paper wasps *Polistes dominulus* (Hymenoptera, Vespidae) were collected in Tuscany (Italy) on June and July, i.e. when *Xenos vesparum* extrude from their puparia to mate (Hughes et al., 2004). Mature *X. vesparum* male specimens were extracted from their puparia (formed by the overlapped exuviae of their previous larval stages) in order to avoid any risk of contamination with the host tissue. Total DNA was extracted from a *X. vesparum*

male using the Wizard SV Genomic DNA purification system (Promega). Oligonucleotide primers for long-PCR amplifications were designed from short mtDNA fragments initially obtained from *cox1*, *nad5*, *nad4*, and *rrnL*. Owing to difficulties in generating specific long PCR products, we were unable to determine the portion of the genome that in most insects includes part of *rrnS*, the AT-rich region, and *trnI*. A large part of the genome was amplified with a GeneAmp PCR system 2700 (Applied Biosystems) in two long PCR reactions with the species-specific primer pair COI-114-J (5'-TAATAGAGATTGGAGTAGGATCA-3') *vs.* ND4-737-N (5'-ATGGGGTCTATCGTGTTTTAGCA-3'), and the pair ND5-2822-J (5'-GAAGAAACCGGAGTAGGAGCTATAGCTAAAG-3') *vs.* 16S-289-N (5'-CTTAAATGTTTAAGAAAGACGAT-3'). These yielded two fragments including the genome segments *cox1-nad4* (8.3 Kb) and *nad5-rrnL* (5.2 Kb). In both amplifications, made in a reaction volume of 25 μl, 0.25μl of TaKaRa LA Taq (Takara) were added, performing 35 cycles at 94°C for 1 min, 50°C for 1 min and 68°C for 10 min, followed by incubation at 68°C for 15 min. These two overlapping fragments were purified using a Montage PCR Centrifugal Filter Device (Millipore), and processed for DNA sequencing by the production facility of the DOE Joint Genome Institute (Walnut Creek, California). Afterwards, the specific primers COI-355-N (5'-GCGGAGGGTAAACTGTTCATCCTGGACCAGCTCC-3') and ND1-394-J (5'- GATAGAAACCAACCTGGCCTTACGCCG-3') were designed facing "out" from these fragments, and coupled with the primers tRNAI-CD-J (5'-GTGCCTGATWAAAAGGATTAYTTTGATAG-3') and 12S-573-CD-N (5'-GGCAGTATTTTAGGCAAATTAGAGGAATC-3'), respectively. These latter were designed to match conserved regions of a multiple alignment of several *trnH* and *rrnS* genes from available dipteran and coleopteran sequences. PCR conditions for the amplification of the fragment *trnI-cox1* (1.9 Kb) were: 35 cycles at 94°C for 1 min, 50°C for 1 min and 68°C for 2 min and 30 sec. Those for the fragment *nad1-rrnS* (1.7 Kb) were: 35 cycles at 94°C for

1 min; 55°C for 1 min and 68°C for 4 min. Each amplified product was obtained using the TaKaRa LA Taq (Takara) enzyme (as above), and each PCR reaction yielded a single band when visualized with ethidium bromide staining after electrophoresis in a 1% agarose gel. PCR products were directly cloned into a TOPO TA Cloning vector (Invitrogen). Each clone was sequenced on both strands in a CEQ 8000XL automated DNA Analysis System (Beckman Coulter). All sequences were assembled using Sequencher 4.2.2 (Gene Codes) and the chromatograms were examined by eye to verify sequencing accuracy. The number of clones sequenced provided a 35X average coverage for the *cox1-nad4* and the *nad5-rrnL* fragments, and a minimum 3X coverage for the *nad1-rrnS* and the *trnI-cox1* fragments.

Protein-coding and rRNA gene sequences were identified through BLAST searches (Altschul et al., 1990) on GenBank, and then by alignment and comparison with sequences from other insect taxa, particularly *Tribolium castaneum* (Insecta, Coleoptera) and *D. melanogaster* (Insecta, Diptera) (GenBank accession numbers: NC_003081 and NC_001709, respectively). The alignments were performed using the programs ClustalX 1.8 (Thompson et al., 1997) and McClade 4.0 (Maddison and Maddison, 2000). Transfer RNA genes were identified by anticodon sequences and inferred secondary structures. The secondary structures of tRNAs and LSU rRNA were reconstructed by hand using previous models available in the literature and the web site http://www.rna.icmb.utexas.edu/ (Cannone et al., 2002).

The presence of repeated sequences within the non-coding fragments was determined using the mreps software (Kolpakov et al., 2003) available at: http://www.loria.fr/mreps. Secondary structures were then visualized using the program RnaViz 2.0 (De Rijk and De Wachter, 1997). Corresponding analyses of codon usage and RSCU were made with the software CodonW 1.3 (Peden, 1999; at url: http://codonw.sourceforge.net/culong.html#CodonW). Base composition of different sequence fragments were obtained with the program PAUP* 4b8 (Swofford et al., 2001).

Strand asymmetry was calculated using the formulas [A(%)-T(%)]/[A(%)+T(%)] and [C(%)-G(%)]/[C(%)+G(%)] (Lobry, 1995; Perna and Kocher, 1995), for the whole genome, and for two-fold and four-fold third codon position sites of PCGs separately.

The sequence of the amplified mtDNA segment of *X. vesparum* has been deposited to GenBank under the accession number XXXX.

3. Results and discussion

3.1. Genome organization

We have obtained the nearly complete sequence of the mitochondrial genome of *Xenos vesparum*, from *trnQ* to about half of *rrnS* (Fig. 1). In addition to missing half of the sequence for *rrnS* and the sequence of the control region, this omits only one expected gene, *trnI*, which most commonly precedes *trnQ* in arthropod mtDNAs, and so may be in this location for *X. vesparum*. All 13 PCGs are positioned in the arrangement considered to be ancestral for arthropods (Boore, 1999), and found, for instance, in *Limulus polyphemus*, *Daphnia pulex*, and *Drosophila melanogaster*.

On the other hand, three tRNA genes of *X. vesparum* occur in derived positions (Fig. 2). Not surprisingly, two of these changes involve tRNA genes present in the cluster that is commonly *trnA*, *trnR*, *trnN*, *trnS*, *trnE*, *trnF*, previously identified as a rearrangement hot spot of the mitochondrial genome of arthropods (Boore, 1999). The gene *trnA* is transposed to a new position between *trnE* and *trnF* near the other end of this cluster. At its ancestral location, between the 3'-end of *nad3* and the 5'-end of *trnR* is a 42 bp region that is otherwise unassigned and which can be folded into a secondary structure motif (data not shown) similar to that of a partial cloverleaf, suggesting that it may represent the remnants of a duplicated

copy of a tRNA gene (although no significant sequence similarity has been found with any tRNA of *X. vesparum*). The *trnS(gcu)* gene has transposed out of this cluster to a position within an otherwise non-coding region between *nad5* and *nad4*, which also contains three copies of *trnH* (see below). Similarly to what is observed for the translocation of the *trnA*, a short non-coding region (27 bp) is found in the inferred ancestral location of *trnS(gcu)*, i.e. between *trnN* and *trnE*.

There are otherwise two cases of derived gene order in *X. vesparum* mtDNA: (1) the *trnL(uag)* gene has transposed from the ancestral location between *nad1* and *rrnL* to a new location between *nad1* and *trnS(uga)*; (2) within the region between *nad4* and *nad5*, which otherwise contains the transposed *trnS(gcu)*, are three copies of putatively functional *trnH* genes. Although one might judge that the copy adjacent to *nad4* (*trnH1*) is the true ortholog based on positional homology with other arthropod mtDNAs, the two extra copies (*trnH2* and *trnH3*) have the same anticodon (GUG), are remarkably similar in primary sequence, and can be easily folded into a canonical cloverleaf secondary structure (Fig. 3), so there is no reason to doubt that all are functional. A mechanism has been proposed to explain gene rearrangements in the mitochondrial genome whereby a block of two or more genes is duplicated, followed by losses of the supernumerary genes (Boore, 2000; Lavrov et al., 2002). The presence of non-coding nucleotides at the ancestral sites for the transposed *trnA* and *trnS(gcu)* (and especially considering that those of the former can be folded into a tRNA-like structure) supports this as a mechanism. All but one of the gene rearrangements in *X. vesparum* mtDNA are within the short portion of the genome between *nad3* and *nad4* (Fig. 1), supporting speculation that any larger duplication that created the condition for rearrangement might have included all or a large part of this region, and it is possible that the large non-coding block found otherwise in this region may themselves be degenerating pseudogenes that are the vestiges of this process. The arrangement of this peculiar region between *nad5*

and *nad4* has been confirmed by sequencing multiple individuals of *X. vesparum,* although with slight sequence variations.

At least one additional duplication would have had to occur to generate the third copy of *trnH*. It may be that there is some advantage for maintaining these extra copies of *trnH*, or perhaps the event is so recent that none of them has yet become a pseudogene. Interestingly, a duplication of this tRNA gene has been also observed in the remipedian crustacean *Speleonectes tulumensis* (Lavrov et al., 2004), although in a different position.

Long intergenic spacers, also observed in other insects and myriapods (Bae et al., 2004; Lavrov et al., 2002), might be the result of errors during the replication of the mitochondrial genome (Macey et al., 1997; Boore, 2000). Interestingly, the non-coding regions (NC1 to NC4 in Fig. 2 and Table 1) also contain tandemly repeated sequences, TTAAAATTT (repeated twice) in NC1 and TTTTATAAA (repeated twice) in NC4, and the microsatellite $(AT)_9$ in NC4 (Fig. 2). The occurrence of these repeated sequences may be associated with regulatory mechanisms and recombination hot spots, and they might be the result of replication slippage events (Kolpakov et al., 2003). In order to assess the origin of the four non coding regions, we separately run dot plot analyses by plotting each NC region against the whole mitochondrial sequence and found a correspondence between NC1 and *trnH2*. In fact, what seem to be the vestiges of an additional gene for the tRNA-H can be found in NC1 where an impefect copy of tRNA-H, having almost perfect anticodon and acceptor arms, but no canonical DHU and TΨC arms, and with a substitution in the putative anticodon (GUG->UUG), can be reconstructed (Fig. 4). This observation reinforces the hypothesis that major rearrangements/duplications occurred in this region have determined the presence of long intergenic non-coding spacers.

Mitochondrial tRNA genes, with their secondary structure motifs, are thought to play a role in delimiting the positions where polycistronic messenger RNAs are cleaved into gene

specific products. Hence, with a few exceptions aside, they are usually located between protein and rRNA coding genes (Ojala et al., 1981). In this respect, the occurrence of three tRNA genes within a non-coding region (Figs. 1 and 2) is somewhat uncommon. Moreover, several stretches of sequences within the non-coding region can be arranged into complex secondary structure motifs similar to those found in ribosomal rRNA and in the "AT-rich region" (Fig. 4). These sequences may play a role as a signal transcription or post-transcriptional cleavage of polycistronic mRNAs.

Although the sequenced portion of the mtDNA does not contain *trnI*, it seems likely that it is in the same position and orientation as in *D. melanogaster* (and many other arthropods), just before *trnQ* in the unsequenced portion. Strongly supporting this is that one of the primers used to amplify part of the sequenced portion of the mtDNA was designed to match conserved sequences within *trnI*. The first gene identifiable then following that primer annealing site is *trnQ*, as would be expected if the arrangement were *trnI* followed by *trnQ* (Fig. 1). An unusually long fragment of 54 unassigned nucleotides is found before *trnQ*, but no special secondary structure feature is observable. Therefore this fragment is considered a simple intergenic spacer.

3.2 Protein coding genes

The arthropod mitochondrial genetic code was employed to identify ORFs which were then matched to those of *Drosophila* for identifying the 13 PCGs (Table 1). Start codons were chosen to maximize similarity to *Drosophila* PCGs. Ten genes start with codons that ordinarily specify Methionine, ATG (*atp6*, *cox3*, *cob*, and *nad4*) or ATA (*nad1*, *nad2*, *nad3*, *nad4L*, *nad5*, and *nad6*). The remaining three use alternatives, with *atp8* and *cox1* starting with ATT (=Ile), and *cox2* starting with CTT (=Leu); although these would ordinarily encode other amino acids, it is likely that these specify formyl methionine in cases where they are

initiators. Concerning *cox2*, the gene could be considered to start with an ATT codon (=Ile) located 4 codons downstream from the proposed CTT initiation codon. This, however, would imply the inclusion of a total of 12 indels (4 codons) in the nucleotide alignment, a fact which is quite unusual in a region which is otherwise highly conserved in length across a wide number of arthropod species (alignment not shown). A complete stop codon ends the reading frame in 11 genes (TAG for *nad4L* and TAA for 10 others), while *cox2* and *nad5* appear to terminate with a single T (Table 1), presumably completed to a TAA stop codon by polyadenylation of the RNA message after cleavage (Okimoto et al., 1990, Lavrov et al., 2002).

Several cases of sequence overlap between genes are observed (Table 1), which is not uncommon for animal mtDNAs (Wostenholme, 1992). Although the majority of overlaps occur between tRNA genes, in three cases the overlaps involve protein-protein gene junctions (*atp8*/*atp6*, *atp8*/*cox3*, *nad6*/*cob*). All the overlapping genes are encoded on the same strand.

3.3. Genome composition

As expected for an insect mitochondrial genome, there is a considerable compositional bias in the mtDNA of *X. vesparum*, where overall A+T content is 79% (Table 1). This value is consistent with those observed in other holometabolan insects within Coleoptera, Diptera and Lepidoptera. The most AT-rich portion of *X. vesparum* mtDNA is the four non-coding regions (NC1-4) present between *nad5* and *trnH1* (Figs. 2 and 3; Table 1).

Interestingly, the AT-bias is stronger in RNA-encoding genes (tRNAs and rRNAs) than in PCGs. In these latter, however, third codon position sites, on which purifying selection against deleterious mutations is expected to be less severe (Jermiin et al., 1994), have a higher AT content than either first and second codon positions (Table 1). The compositional bias is reflected in codon usage preference of four-fold degenerate codon families, where the codons

ending with A or T greatly outnumber those ending with G or C (Table 2). In addition, six of the nine four-fold degenerate codon families (those encoding for L(CUN), G, R, S(AGN), T, and S(UCN) use codons ending with A more frequently than those ending with T, and their corresponding tRNAs are all encoded in the J-strand. On the other hand, the two four-fold degenerate codon families whose tRNA is encoded in the N-strand use codons ending with T more frequently than those ending with A. The only exception to this correlation is *trnA* (and the corresponding codon family GUN), which is encoded in the J-strand, but uses the GUU codon more frequently than GUA (Table 2). While no reasonable explanation can be given for this observation, it is noteworthy that even tRNAs whose anticodon begins with T prefer codons ending with T. As expected, the compositional bias of mitochondrial protein coding gene sequences influences the selection of a given amino acid as a function of the base content of its codon family. This can be assessed calculating the ratio between the number of GC-rich codons (encoding for A, R, G and P), and that of AT-rich codons (encoding for N, I, K, M, F and Y) (Crozier and Crozier, 1993). In *X. vesparum* this value (0.20) is very similar to the value (0.18) observed in the extremely AT-rich (85%) mtDNA of *Apis mellifera* (Crozier and Crozier, 1993), and is consistently lower than the values (42-43%) observed in Diptera (Lessinger et al., 2000).

The mtDNA of *X. vesparum* also exhibits another type of compositional bias, in the relative frequency of As and Ts in PCGs encoded in different strands (Hassanin et al., 2005). In fact, while all genes encoded in the N-strand show a fairly similar content in As and Ts, in the coding sequence of those encoded in the J-strand, Ts outnumber As (Fig. 5). Strand bias in nucleotide composition can be assessed analyzing the third codon positions of protein coding genes (PCGs), and grouping the codons in two different categories depending on the number of synonymous sites at third codon positions (two-fold and four-fold degenerate sites) (Hassanin et al., 2005). Within the two-fold degenerate codon families, two groups may be

distinguished, those ending with a purine (NNR) and those ending with a pyrimidine (NNY); on the other hand, only one group of four-fold degenerate families exists (NNN). In each of the three groups, compositional bias at third codon positions can be assessed by estimating whether the frequencies of complementary nucleotides (A *vs.* T and G *vs.* C) are different. The rationale for assessing compositional bias at synonymous sites is that substitutions at these sites are unconstrained by the encoded amino acid, and the bias is therefore related to other mutational constraints which drive the asymmetry of base composition between strands. The bias can be quantified with the skew value (Perna and Kocher, 1995). Strand asymmetry is observed in the whole sequenced genome of *X. vesparum* (Table 1), with As outnumbering Ts (AT-skew=0.08) and Cs outnumbering Gs (CG-skew=0.30) in the J-strand. If only PCGs are taken into account (Table 3), the same trend is observed in both two-fold and four-fold degenerate third codon position sites, which are under no or little selective constraints, and it suggests the presence of asymmetric patterns of mutational changes between strands (Lobry, 1995; Sueoka, 1995). This phenomenon has been found widely across arthropods (Hassanin et al., 2005), and it has been related with nucleotide deamination of transiently single-strand DNA during replication (but see Bogenhagen and Clayton, (2003) for the "strand-coupled model") and/or transcription.

3.4. Transfer RNAs and the Large Ribosomal Subunit RNA

The sequences of 23 tRNA genes (including three copies of *trnH*) have been obtained, and evidence suggests that also *trnI* is present and located in its usual position (for arthropods) next to *trnQ*. Therefore, the mtDNA of *X. vesparum* appears to contain the "normal" 22-tRNA complement typical of most metazoan mitochondrial genomes, except for the additional two copies of *trnH*. Each of the 23 tRNA gene sequences can be folded into an almost perfect cloverleaf secondary structure, although mismatches occur in a few of them,

especially in the acceptor arm (Fig. 4). Of the total 451 potential bonds, 393 are canonical A-U and G-C pairings, 23 are non Watson-Crick interactions (G•U and A•C), which are permitted in RNA secondary structures (Hickson et al., 1996), and 34 are considered mismatches. The size of the genes and their anticodons are congruent with those observed in other insect species (Fig. 3 and Table 1). Some tRNAs (most notably tRNA-V, but also tRNA-D, tRNA-M, tRNA-E, tRNA-S(gcu) and the three tRNA-H) have some unusual base-pairings in their arms. Substantial post-transcriptional RNA editing has been shown to occur in some metazoan mitochondrial tRNAs, especially when mismatches fall in regions where the tRNA genes overlap with adjacent down-stream genes (Lonergan and Gray, 1993; Yokobori and Pääbo, 1995; Lavrov et al., 2000; Masta and Boore 2004), and so could potentially occur here to create a more conforming structure. The presence of unusual TΨC and DHU arms for mitochondrial tRNAs has been observed in some other arthropodan taxa (Masta, 2000; Ogoh and Ohmiya, 2004).

Secondary structure models of rRNA molecules are generally established through comparative analysis of multiple sequences (Wuyts et al., 2001). Compensatory substitutions that would maintain secondary structure are considered to be evidence of actual base pairing (Gutell et al., 1994; Buckley et al., 2000). Unfortunately, very few complete secondary structure models have been proposed for insect mitochondrial rRNA molecules (Misof and Fleck, 2003) and a broader comparative analysis in insects has been performed only on a fragment of the mitochondrial LSU rRNA (Buckley et al., 2000). Therefore, we inferred the secondary structure of the LSU rRNA of *X. vesparum* (Fig. 6) by comparison with the model proposed for *D. melanogaster*, available in the Comparative RNA Web Site (Cannone et al., 2002). Homology among structural elements between *X. vesparum* and *D. melanogaster* is evident for each of the four domains of the LSU rRNA. Several helices can be drawn in the same way with negligible structural variability. Apparently, minor structural variation can be

observed only at the 5' end of domain I. As observed in other mitochondrial rRNA molecules, helix stability is frequently ensured by a mixture of canonical Watson-Crick bonds and unconventional base pairings (G•U and A•C) (Hickson et al., 1996; Carapelli et al., 2004).

Acknowledgements

This work was supported by grants from the Italian MIUR (PRIN) and the University of Siena (PAR) to FF and RD. Part of this work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231. We thank J. Kathirithamby for stimulating discussion on strepsipteran biology.

References

Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Bae, J.S., Kim, I., Sohn, H.D., Jin, B.R., 2004. The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. Mol. Phylog. Evol. 32, 978-985.

Beani, L., Giusti, F., Mercati, D., Lupetti, P., Paccagnini, E., Turillazzi, S., Dallai, R., 2005. Mating of *Xenos vesparum* (Rossi) (Strepsiptera, Insecta) revisited. J. Morphol. 265, 291-303.

Bogenhagen, D.F., Clayton, D.A., 2003. The mitochondrial DNA replication bubble has not burst. Trends Biochem. Sci. 28, 357-360.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucl. Acids Res. 27, 1767-1780.

Boore, J.L., 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In: Sankoff, D., Nadeau, J.H. (Eds.), Comparative genomics. Kluwer Academic Publishers, Dordrecht, pp. 133–147.

Boore, J. L., Brown, W.M., 1998. Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. 8, 668-674.

Boore, J.L., Lavrov, D., Brown, W.M., 1998. Gene translocation links insects and crustaceans. Nature 393, 667-668.

Buckley, T.R., Simon, C., Flook, P.K., Misof, B., 2000. Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. Insect Mol. Biol. 9, 565-580.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., Pande, N., Shang, Z., Yu, N., Gutell, R.R., 2002. The Comparative RNA Web (CRW) Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and other RNAs. BioMed Central Bioinformatics 3, 2.

Carapelli, A., Soto-Adames, F.N., Simon, C., Frati, F., Nardi, F., Dallai, R., 2004. Secondary structure, high variability, and conserved motifs for domain III of 12S rRNA in the Arthropleona (Hexapoda; Collembola). Insect Mol. Biol. 13, 659-670.

Carmean, D., Crespi, B., 1995. Do long branches attract flies? Nature 373, 666.

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

Curole, J.P., Kocher, T.D., 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol. Evol. 14, 394-398.

De Rijk, P., De Wachter, R., 1997. RnaViz, a program for the visualisation of RNA secondary structure. Nucl. Acids Res**.** 25, 4679-4684.

Gillespie, J.J., McKenna, C.H., Yoder, M.J., Gutell, R.R., Johnston, J.S., Kathirithamby, J., Cognato, A.I., 2005. Assessing the odd secondary structural proprieties of nuclear small subunit ribosomal RNA sequences (18S) of the twisted-wing parasite (Insecta: Strepsiptera). Insect Mol. Biol., doi: 10.1111/j.1365-2583.2005.00591.x.

Gutell, R.R., Larsen, N. and Woese, C.R., 1994. Lessons from an evolving rRNA mit LSU and 23S structures from a comparative perspective. Microbiol. Rev. 58, 10-26.

Hassanin, A., Léger, N., Deutsch, J., 2005. Evidence for multiple reversals of asymmetrical mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. Syst. Biol. 54, 277-298.

Hickson, R.E., Simon, C., Cooper, A., Spicer, G.S., Sullivan, J., Penny, D., 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Mol. Biol. Evol. 13, 150-169.

Huelsenbeck, J.P., 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46, 69-74.

Huelsenbeck, J.P., 1998 Systematic bias in phylogenetic analysis: Is the Strepsiptera problem solved? Syst. Biol. 47, 519-537.

Hughes, D.P., Kathirithamby, J., Turillazzi, S., Beani, L., 2004. Social wasps desert the colony and aggregate outside if parasitized: parasite manipulation? Behav. Ecol. 15, 1037-1043.

Hwang, U.W., Kim, W., Tautz, D., Friedrich, M., 1998. Molecular phylogenetics at the Felsenstein zone: Approaching the Strepsiptera problem using 5.8S and 28S rDNA

sequences. Mol. Phylog. Evol. 9, 470-480.

Kathirithamby, J., 1989. Review of the order Strepsiptera. Syst. Entomol. 14, 41-92.

Kolpakov, R., Bana, G., Kucherov, G., 2003. mreps: efficient and flexible detection of tandem repeats in DNA. Nucl. Acids Res. 31, 3672-3678.

Kristensen, N.P., 1991. Phylogeny of extant hexapods. In: Naumann, I.D., Carne, P.B., Lawrence, J.F., Nielsen, E.S., Spradberry, J.P., Taylor, R.W., Whitten, M.J., Littlejohn, M.J. (Eds.), Insects of Australia: a textbook for students and research workers. Melbourne Univ. Press, Melbourne, pp. 125-140.

Jermiin, L.S., Graur, D., Lowe, R.M., Crozier, R.H., 1994. Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome *b* genes. J. Mol. Evol. 39, 160-173.

Johnston, J.S., Ross, L.D., Beani, L., Hughes, D.P., Kathirithamby, J., 2004. Tiny genomes and endoreplication in Strepsiptera. Insect Mol. Biol. 13, 581-585.

Lavrov, D.V., Boore J.L., Brown, W.M., 2000. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. Proc. Natl. Acad. Sci. USA 97, 13738–13742.

Lavrov, D.V., Boore, J.L., Brown, W.M., 2002. Complete mt DNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangement: duplication and non-random loss. Mol. Biol. Evol. 19, 163–169.

Lavrov, D.V., Brown, W.M., Boore, J.L., 2004. Phylogenetic position of the Pentastomida and (pan)crustacean relationships. Proc. Biol. Sci. 271, 537-544.

Lessinger, A.C., Martins Junqueira, A.C., Lemos, T.A., Kemper, E.L., da Silva, F.R., Vettore, A.L., Arruda, P., Azeredo-Espin, A.M.L., 2000. The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). Insect Mol. Biol. 9, 521-529.

Lewis, D.L., Farr, C.L., Kaguni, L.S., 1995. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Mol. Biol. 4, 263-278.

Lobry, J.R., 1995. Properties of a general model of DNA evolution under no-strand bias conditions. J. Mol. Evol. 40, 326-330.

Lonergan, K.M., Gray, M.W., 1993. Editing of transfer RNAs in *Acanthamoeba castellanii* mitochondria. Science 259, 812-816.

Macey, J.R., Larson, A., Ananjeva, N.B., Fang, Z., Papenfuss, T.J., Macey, R.J., 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol. Biol. Evol. 14, 91-104.

Maddison, D.R., Maddison, W.P., 2000. McClade. Classroom Version 4.0. Sinauer Associates, Sunderland.

Masta, S.E., 2000. Mitochondrial sequence evolution in spiders: intraspecific variation in tRNAs lacking the TψC arm. *Mol. Biol. Evol.* 17, 1091-1100.

Masta, S. E., Boore, J. L., 2004. The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. Mol. Biol. Evol. 21, 893-902.

Misof, B., Fleck, G., 2003. Comparative analysis of mt LSU rRNA secondary structures of Odonates: structural variability and phylogenetic signal. Insect Mol. Biol. 12, 535-547.

Nardi, F., Carapelli, A., Dallai, R., Frati, F., 2003. The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographic locations. Insect Mol. Biol. 12, 605-611.

Ogoh, K., Ohmiya, Y., 2004. Complete mitochondrial DNA sequence of the sea-firefly, *Vargula hilgendorfii* (Crustacea, Ostracoda) with duplicate control regions. Gene 327, 131-139.

Okimoto, R., Macfarlane, J.L., Wolstenholme, D.R., 1990. Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans.* Nucl. Acids Res. 18, 6113-6118.

Peden, J.F., 1999. Analysis of codon usage, Ph.D. Thesis, Department of Genetics, University of Nottingham, UK.

Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353-358.

Ojala, D., Montoya, J., Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 290, 470-474.

Rawlings, T.A., Collins, T.M., Bieler, T.M., 2003. Changing identities: tRNA duplication and remolding within animal mitochondrial genomes. Proc. Natl. Acad. Sci. USA 100, 15700-15705.

Rokas, A. Kathirithamby, J., Holland, P.W.H., 1999. Intron insertion as a phylogenetic character: the engrailed homeobox of Strepsiptera does not indicate affinity with Diptera. Insect Mol. Biol. 8, 527-530.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651-704.

Sueoka, N., 1995. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J. Mol. Evol. 40, 318-325.

Swofford, D.L., 2001. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer, Associates, Sunderland.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 25, 4876-4882.

Yokobori, S., Pääbo, S., 1995. Transfer RNA editing in land snail mitochondria. Proc. Natl. Acad. Sci. USA 92, 10432–10435.

Whiting, M.F., Carpenter, J.C., Wheeler, Q.D., Wheeler, W.C., 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46, 1-68.

Wolstenholme, D.R., 1992. Animal mitochondrial DNA: structure and evolution, In: Wolstenholme, D.R., Jeon, K.W. (Eds), Mitochondrial Genomes, Academi Press, New York, pp. 173-216.

Wuyts, J., De Rijk, P., Van de Peer, Y., Winkelmans, T., De Wachter, R., 2001. The European Large Subunit Ribosomal RNA Database. Nucl. Acids Res. 29, 175-177.

Yokobori, S. and Pääbo, S. (1995) Transfer RNA editing in land snail mitochondria. Proc. Natl. Acad. Sci. USA 92, 10432-10435.

Table 1. Base composition and gene profile of *X. vesparum* mitochondrial genes.

Region	A	C	G	T	Size[a]	Start codon	Stop codon[b]	Strand	Intergenic nucleotides[c]
trnQ	0.37	0.00	0.17	0.45	64	-	-	N	54
trnM	0.30	0.25	0.10	0.35	69	-	-	J	0
nad2	0.39	0.12	0.10	0.39	921	ATA	TAA	J	+3
trnW	0.49	0.08	0.08	0.34	61	-	-	J	-2
trnC	0.43	0.06	0.06	0.43	62	-	-	N	-1
trnY	0.35	0.08	0.15	0.42	60	-	-	N	-1
cox1	0.33	0.17	0.13	0.37	1515	ATT	TAA	J	-8
trnL(uaa)	0.43	0.06	0.11	0.39	61	-	-	J	-5
cox2	0.40	0.12	0.11	0.37	663	CTT[d]	Taa	J	-12
trnK	0.31	0.16	0.16	0.37	62	-	-	J	0
trnD	0.51	0.08	0.08	0.33	61	-	-	J	-2
atp8	0.48	0.10	0.03	0.40	144	ATT	TAA	J	0
atp6	0.38	0.16	0.07	0.38	612	ATG	TAA	J	-7
cox3	0.39	0.14	0.07	0.40	753	ATG	TAA	J	-1
trnG	0.50	0.03	0.05	0.42	62	-	-	J	0
nad3	0.37	0.14	0.08	0.42	348	ATA	TAA	J	53
trnR	0.41	0.12	0.09	0.37	65	-	-	J	42
trnN	0.55	0.05	0.08	0.32	65	-	-	J	-3
trnE	0.42	0.09	0.05	0.44	64	-	-	J	27
trnA	0.49	0.06	0.06	0.38	63	-	-	J	-1
trnF	0.40	0.02	0.13	0.45	62	-	-	N	-2
nad5	0.29	0.08	0.13	0.50	1620	ATA	Taa	N	0
NC1	0.50	0.06	0.06	0.38	173	-	-	-	-
trnS(gcu)	0.48	0.10	0.05	0.37	73	-	-	J	-
NC2	0.58	0.13	0.06	0.23	86	-	-	-	-
trnH3	0.43	0.01	0.12	0.43	67	-	-	N	-
NC3	0.50	0.07	0.05	0.37	94	-	-	-	-
trnH2	0.48	0.00	0.10	0.42	67	-	-	N	-
NC4	0.47	0.02	0.11	0.39	122	-	-	-	-
trnH1	0.40	0.01	0.12	0.46	65	-	-	N	-
nad4	0.30	0.06	0.13	0.52	1290	ATG	TAA	N	-17
nad4L	0.30	0.02	0.14	0.54	294	ATA	TAG	N	3
trnT	0.57	0.01	0.03	0.39	67	-	-	J	-25
trnP	0.40	0.03	0.13	0.44	63	-	-	N	-1
nad6	0.44	0.14	0.05	0.37	489	ATA	TAA	J	-1
cob	0.35	0.18	0.08	0.40	1095	ATG	TAA	J	-14
trnS(uga)	0.42	0.07	0.11	0.39	66	-	-	J	-1
trnL(uag)	0.38	0.06	0.14	0.41	63	-	-	N	-2
nad1	0.31	0.05	0.15	0.49	954	ATA	TAA	N	33
rrnL	0.45	0.12	0.04	0.38	1208	-	-	N	0
trnV	0.44	0.06	0.06	0.44	66	-	-	N	0
Mean all sites	0.43	0.13	0.07	0.36	14519				
Mean all coding sites	0.34	0.11	0.11	0.43	10698				
Mean PCGs J-strand	0.37	0.15	0.09	0.39	6540				
Mean PCGs N-strand	0.30	0.06	0.13	0.51	4158				
Mean tRNAs	0.43	0.07	0.10	0.40	1478				
First codon pos.	0.39	0.11	0.14	0.36	3566				
Second codon pos.	0.22	0.14	0.13	0.51	3566				
Third codon pos.	0.42	0.09	0.06	0.43	3566				

[a] without stop codons.

[b] As that are presumably added by polyadenylation to generate complete stop codons are represented with lowercase letters.

[c] numbers correspond to nucleotides separating each gene; negative numbers refer to overlaps between genes.

[d] see section *3.2.* for an alternative hypothesis of initiation codon.

Table 2. Codon usage in the PCGs of the *X. vesparum* mitochondrial genome. RSCU=relative synonymous codon usage.

Amino acid	Codon	n	RSCU	Amino acid	Codon	n	RSCU	Amino acid	Codon	n	RSCU	Amino acid	Codon	n	RSCU
Phe	UUU	341	1.71	Ser	UCU	73	1.62	Tyr	UAU	165	1.74	Cys	UGU	30	1.76
	UUC	59	0.29		UCC	17	0.38		UAC	25	0.26		UGC	4	0.24
Leu	UUA	364	3.89		UCA	91	2.02	TER	UAA	0	0	Trp	UGA	81	1.88
	UUG	35	0.37		UCG	3	0.07		UAG	0	0		UGG	5	0.12
	CUU	56	0.60	Pro	CCU	49	1.83	His	CAU	49	1.61	Arg	CGU	11	1.38
	CUC	11	0.12		CCC	23	0.86		CAC	12	0.39		CGC	4	0.50
	CUA	89	0.95		CCA	33	1.23	Gln	CAA	38	1.65		CGA	13	1.62
	CUG	6	0.06		CCG	2	0.07		CAG	8	0.35		CGG	4	0.50
Ile	AUU	323	1.70	Thr	ACU	44	1.32	Asn	AAU	174	1.71	Ser	AGU	33	0.73
	AUC	56	0.30		ACC	19	0.57		AAC	29	0.29		AGC	11	0.24
Met	AUA	291	1.77		ACA	69	2.08	Lys	AAA	133	1.71		AGA	106	2.36
	AUG	37	0.23		ACG	1	0.03		AAG	23	0.29		AGG	26	0.58
Val	GUU	66	1.83	Ala	GCU	40	2.13	Asp	GAU	47	1.57	Gly	GGU	35	1.12
	GUC	3	0.08		GCC	13	0.69		GAC	13	0.43		GGC	13	0.42
	GUA	59	1.64		GCA	21	1.12	Glu	GAA	60	1.41		GGA	58	1.86
	GUG	16	0.44		GCG	1	0.05		GAG	25	0.59		GGG	19	0.61

Table 3. Nucleotide composition compared for third codon positions for genes encoded on opposite strands. NNR group: two-fold degenerate codon families with G or A (purine) at third position. NNY: two-fold degenerate codon families with T or C (pyrimidine) at third codon position. NNN: four-fold degenerate codon families. AT(2) and CG(2) skews were calculated for two-fold degenerate positions. AT(4) and CG(4) skews were calculated for four-fold degenerate positions.

	Third codon positions												
Strand	NNR group		NNY group		Skew	Skew	NNN group					Skew	Skew
	A%	sites	T%	sites	AT2	CG2	A%	C%	G%	T%	sites	AT4	CG4
J	94.06	640	77.15	792	+0.01	+0.59	55.61	13.64	4.01	26.74	748	+0.35	+0.55
N	79.39	461	96.82	535	-0.10	-0.73	31.54	3.08	12.31	53.08	390	-0.25	-0.60

Figure captions:

Figure 1. Graphical representation of the mtDNA of *X. vesparum* with each gene shown in grey. Those on the outer (J) strand are transcribed counter-clockwise and those on the inner (N) strand are transcribed clockwise. Transfer RNA genes are designated by the one-letter amino acid code for the corresponding amino acid, with the two *trnL* and the two *trnS* differentiated on the basis of their anticodon. Significant non-coding regions are represented in white. The fragment not sequenced in this study is light-shaded.

Figure 2. Linearized representation of the mitochondrial fragment *nad3-rrnL* of *X. vesparum* in comparison with the putative ancestral gene arrangement retained by *D. melanogaster*. Genes are shown in grey. Arrows indicate tRNA gene translocations. Black segments highlight four non-coding regions (NC1-4) between *nad5* and *nad4*. The positions of tandemly repeated sequences within non-coding regions is represented with lowercase letters (a-c).

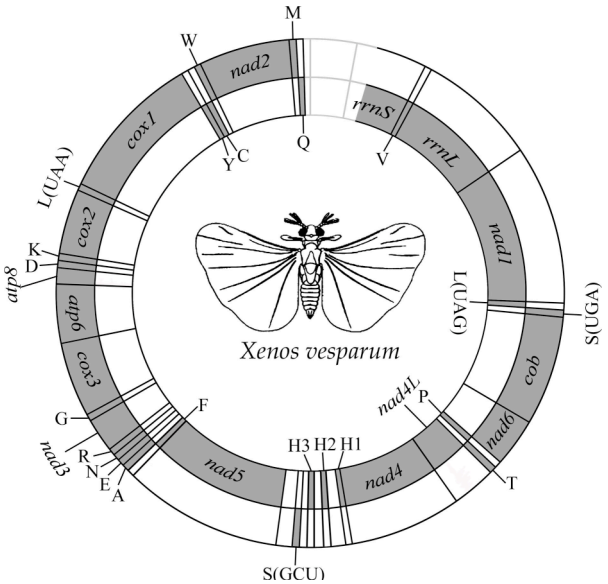
Figure 3. Putative secondary structures for 23 tRNAs of the *X. vesparum* mitochondrial genome. There are three copies of tRNA-H. In the tRNAs for L and S, the codon family to which each tRNA is presumed to correspond is indicated. tRNA-I is not included in this study.

Figure 4. Region between *nad5* and *nad4* on the J-strand of *X. vesparum* mtDNA, with putative secondary structure motifs. Within the fragment are one copy of *trnS(gcu)*, two reverse complement copies of *trnH*. The region highlighted in grey at the 5'-end of NC1

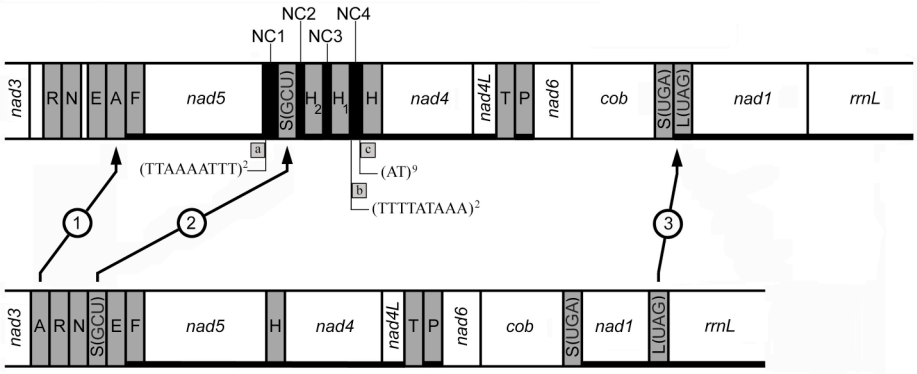
can be folded in a tRNA-like structure, possibly the vestiges of an additional copy of *trnH* (shown in inset). Repeated elements are represented with lowercase letters (a-c).

Figure 5. Graphical representation of the percentage of As (grey) and Ts (black) calculated in a sliding window of 100 bp. a) Along the J-strand; b) only on PCGs encoded in the N-strand. The position and orientation of each gene is shown below the graphs.

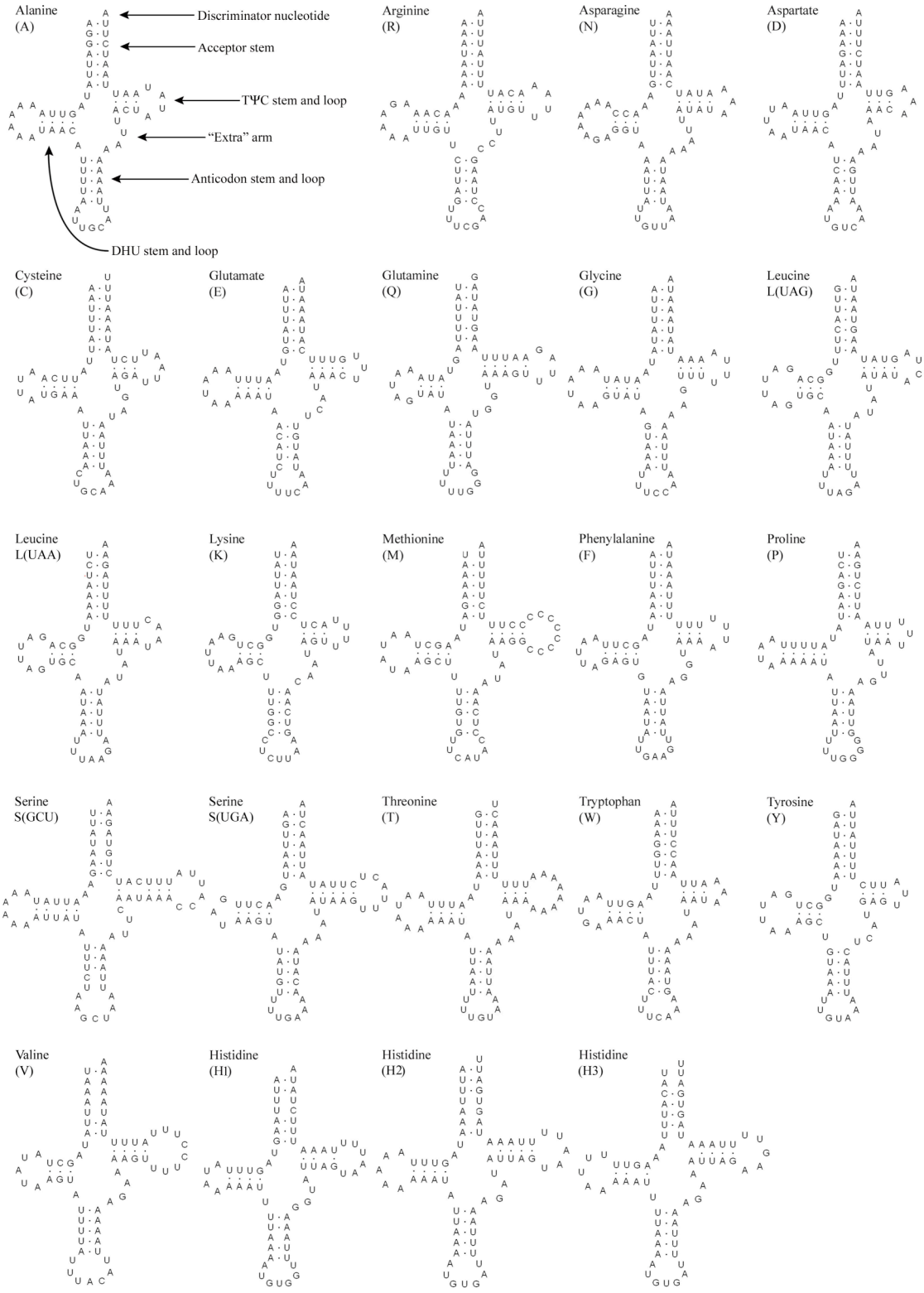
Figure 6. Predicted secondary structure of the LSU rRNA of *X. vesparum* with separation lines indicating the four domains.

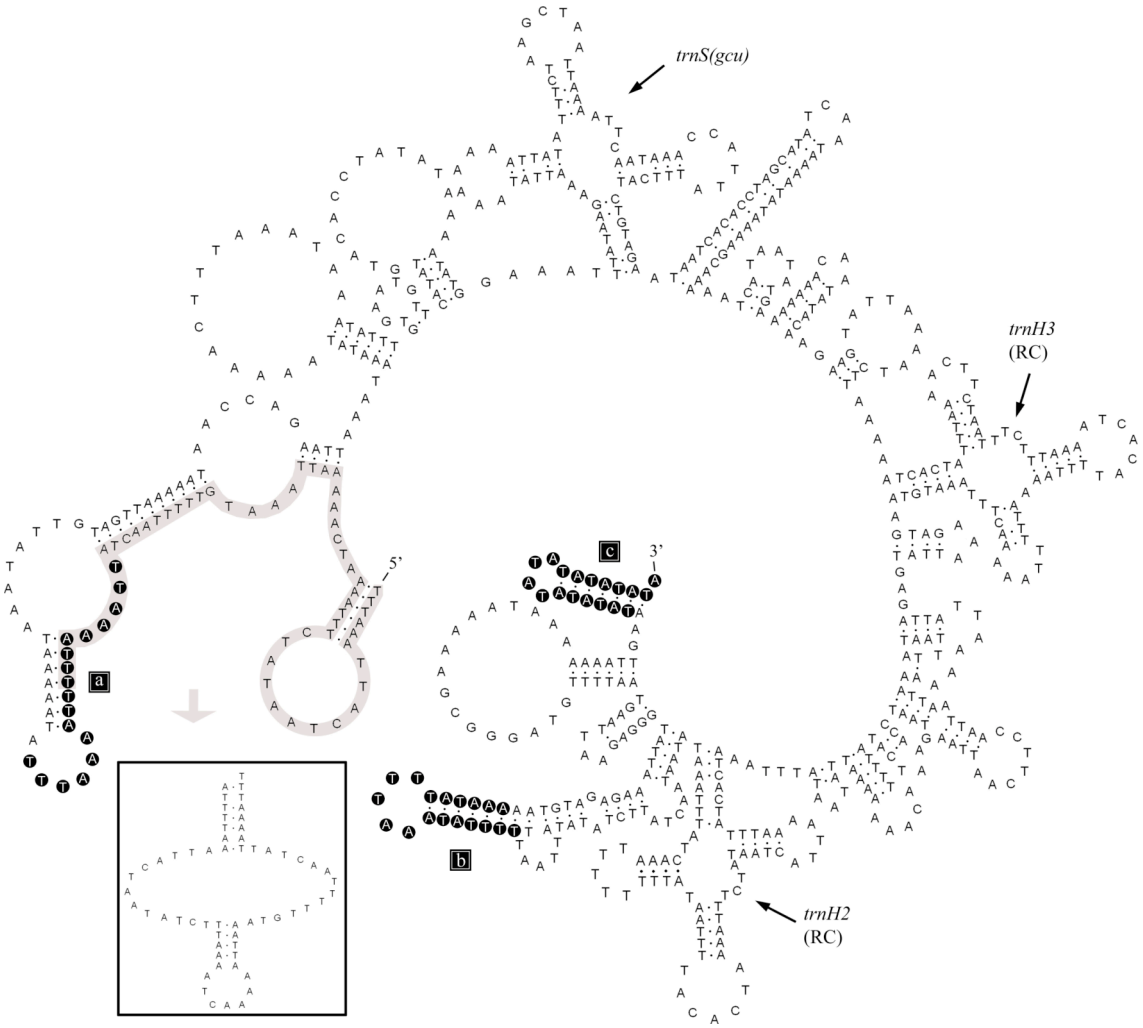

M
W
nad2
rrnS
rrnL
nad1
cox1
Y
C
Q
V
L(UAA)
cox2
K
D
atp8
atp6
cox3
L(UAG)
S(UGA)
cob
Xenos vesparum
nad4L
P
nad6
G
F
H3 H2 H1
nad3
R
N
E
A
nad5
nad4
T
S(GCU)

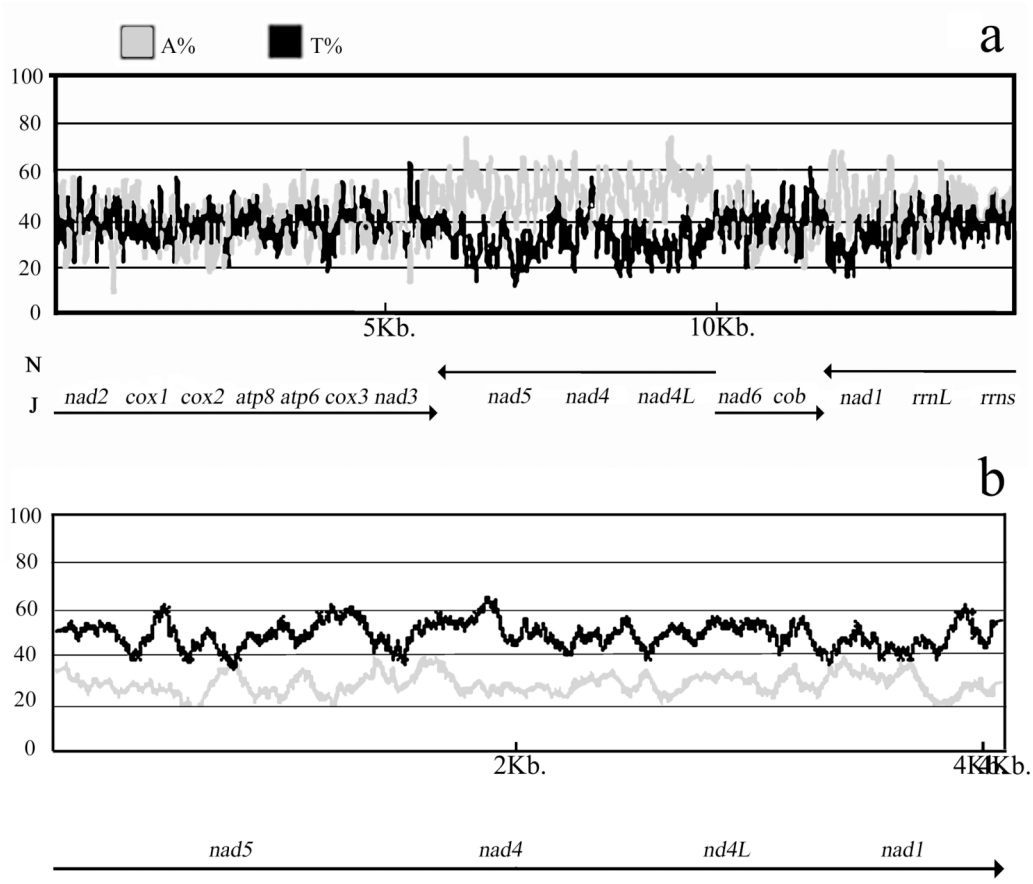
Gene arrangement in the mitochondrial genome of *Xenos vesparum*

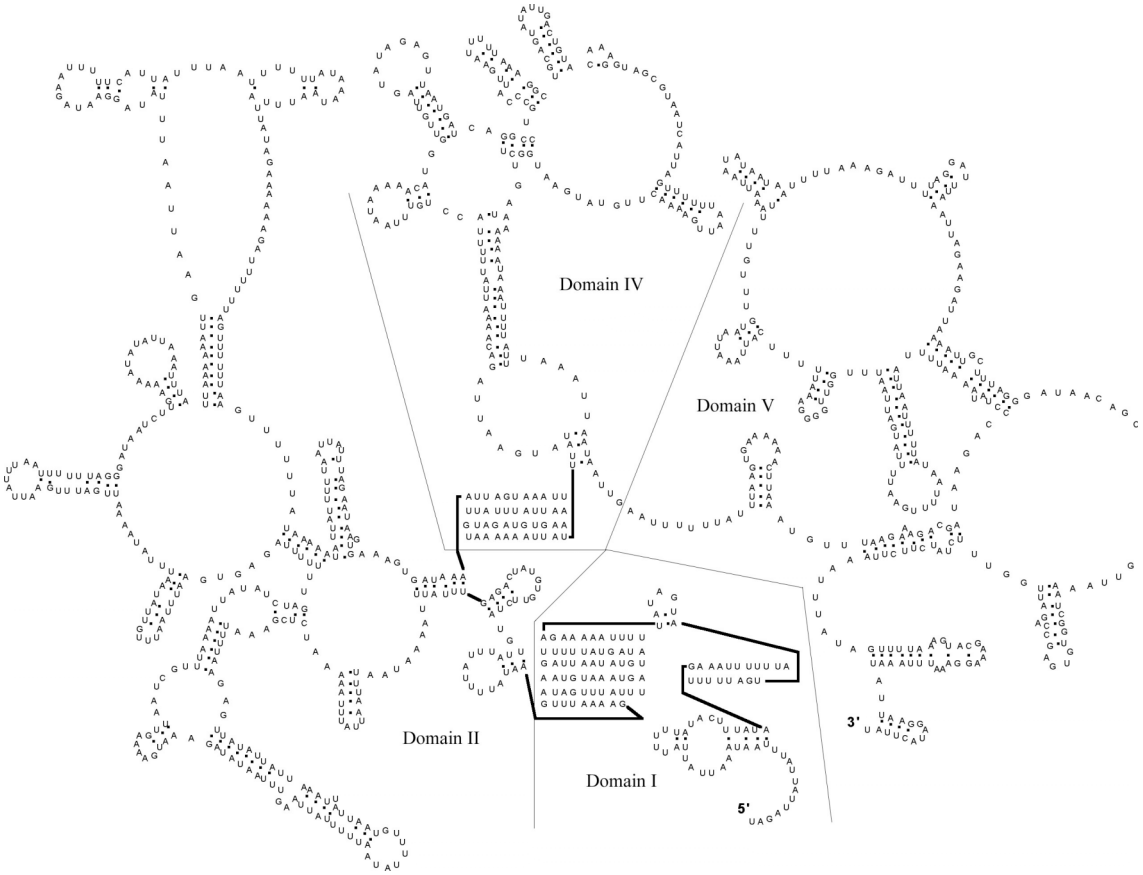


Drosophila melanogaster gene arrangement


Alanine
(A)
Discriminator nucleotide
Acceptor stem
TΨC stem and loop
"Extra" arm
Anticodon stem and loop
DHU stem and loop
Arginine
(R)
Asparagine
(N)
Aspartate
(D)
Cysteine
(C)
Glutamate
(E)
Glutamine
(Q)
Glycine
(G)
Leucine
L(UAG)
Leucine
L(UAA)
Lysine
(K)
Methionine
(M)
Phenylalanine
(F)
Proline
(P)
Serine
S(GCU)
Serine
S(UGA)
Threonine
(T)
Tryptophan
(W)
Tyrosine
(Y)
Valine
(V)
Histidine
(Hl)
Histidine
(H2)
Histidine
(H3)


trnS(gcu)
trnH3
(RC)
trnH2
(RC)
5'
3'
a
b
c


A%
T%
a
100
80
60
40
20
0
5Kb.
10Kb.
N
J
nad2 cox1 cox2 atp8 atp6 cox3 nad3
nad5 nad4 nad4L nad6 cob
nad1 rrnL rrns
b
100
80
60
40
20
0
2Kb.
4Kb.
nad5
nad4
nd4L
nad1


Domain IV
Domain V
Domain II
Domain I
3'
5'