LBNL-50963

The complete mitochondrial sequence of the “living fossil” *Tricholepidion gertschi*: structure, phylogenetic implications, and the description of a novel A/T asymmetrical bias.

Nardi, F. (LBNL), Frati, F. (LBNL), Carapelli, A. (LBNL), Dallai, R. (LBNL), Boore, J. (JGI)

Abstract

Introduction

Traditionally, the “Apterygota” has been thought to consist of five orders of wingless hexapods (Protura, Collembola, Diplura, Microcoryphia and Zygentoma) believed to be collectively basal to insects (i.e., the Pterygota). However, some studies have questioned this affinity with insects (Dallai, Abele, Spears, Nardi). Further, within these groups are hotly debated issues, including the monophyly of Entognata (Koch, 1997; Kukalova Peck, 1987), the monophyly of Diplura (Bilinski, 1993; Stys and Bilinski, 1990), the affinity between Collembola and Protura (Dallai, 1994; Kristensen, 1981) and the position of Lepidotrichidae (below). In fact, these relationships constitute one of the most debated issues in hexapod phylogeny.

The family Lepidotrichidae was first described by (Silvestri, 1912) (1912: “Lepidothricinae”) from a Baltic Amber fossil (*Lepidothrix pilifera* Menge). The only living representative of this family is *Tricholepidion gertschi* Wygodzinski. Since this species was first described (Wygodzinsky, 1961) its phylogenetic position has been difficult to establish, due to an “array of unique characters” that are difficult to interpret in a phylogenetic framework. *Tricholepidion* (and therefore the whole family Lepidotrichidae) has been considered either as belonging to the order Zygentoma (Kristensen, 1997; Wygodzinsky, 1961), or basal to the rest of the Zygentoma *plus* the Pterygota (Beutel, 2001; Bitsch and Bitsch, 2000; Staniczek, 2000), although the significance of some of the morphological characters on which these analyses are based have been questioned (Dallai et al., 2001; Kristensen, 1997). If the latter hypothesis proved to be true, the family Lepidotrichidae, would better deserve the ordinal rank.

Since studies based on morphological characters have failed to give a satisfactory answer, a broad scale molecular study is under way ((Nardi et al., 2001), Frati et al, submitted, il Gomphiocephalus) in order to use mitochondrial genome sequences to study the evolution and differentiation of the most basal hexapod groups, including *Tricholepidion*. Mitochondrial genomics, that is analysis of various features of the mitochondrial genome such as gene order and the analysis of the concatenated sequence of its genes, has proved to be a very powerful tool for the study of ancient phylogenetic relationships (Boore, 2000; Boore and Brown, 1995; Boore and Brown, 1998; Garcia-Machado et al., 1999; Hwang et al., 2001; Nardi et al., 2001), and its application seems to be appropriate for the problem under study ((Nardi et al., 2001), this study). In addition, complete mitochondrial sequences, with the advent of automatic sequencing tools, are accumulating rapidly, but there is a strong bias towards the better known or economically important groups, while only two sequences have been produced for the more basal, and evolutionarily more intriguing, hexapod orders. The complete sequence of the mitochondrial genome of *Tricholepidion gertschi* is the second among apterygotans, following the collembolan *T.bielanensis* (Nardi et al., 2001).

Materials and methods

Several specimens of *Tricholepidion gertschi* were collected under the bark of rotting stumps of coast redwood and Douglas fir in the Heath and Marjorie Angelo Coast Range Reserve (Brimascombe, CA, USA). Total DNA was purified from three legs using the DNeasy Tissue Kit (Qiagen), and 1ul of this was used for PCR reactions. Initially, two short fragments were amplified using the universal primer pairs C1-J-1718/C1-N-2191 and LR-J-12887/LR-N-13398 (Simon et al., 1994), that cover 450bp of *cox1* and 500bp of *rrnL*, respectively. Based on these sequences, four primers were designed to amplify the whole mitochondrial genome of *Tricholepidion* in two overlapping pieces: Tg-A (5’-TGCCCCTAGAATTGATGAAACACCG-3’), Tg-B (5’-

GTAGAAAATGGTGCAGGAACTGGC-3'), Tg-C (5'-TGTAGGCTGGAATGAACGGTTGGAC-3') and Tg-D (5'-GATCTTTATTTCTCACCGTCGCCC-3'). PCR conditions for the primer pair Tg-B/Tg-C were 35 cycles of 1'-94°C, 1'10"-64°C, 10'-68°C, using *rTth* polymerase (Gene Amp XL PCR Kit, Applied Biosystems) according to the manufacturer specifications, with 1.25mM magnesium acetate. The product, of approximately 12 kb, was sequenced at the JGI Joint Genome Institute using a shotgun procedure, with an average of 20 reads of overlap throughout the whole molecule. The rest of the molecule was amplified using the primer pair Tg-A/Tg-D, with PCR conditions: 35 cycles of 1'-94°C, 6'-68°C, using the Expand High Fidelity PCR System (ROCHE) with 2.5 mM magnesium chloride. The product, of approximately 4.5 kb, was cloned (TOPO TA Cloning kit, Invitrogen) and three clones were sequenced via primer walking at the core facility of ENEA, using a dye terminator technology on a ABI 3100 machine (Perkin Elmer). Electropherograms were assembled and manually corrected using the Sequencher software (Gene Codes Corporation), and the resulting sequence was annotated using MacVector (Accelrys), that was also used to calculate basic sequence statistics. Protein coding genes, tRNAs and rRNAs were located by applying functional considerations (Open Reading Frames, secondary structures) and by sequence comparisons with other mitochondrial genomes. All arthropodan complete mitochondrial sequences available at the time of analysis, plus those of three mollusks and two annelids, were retrieved from GenBank. Sequences for all protein coding genes (PCG) were identified, translated applying the appropriate genetic code, and aligned separately using ClustalX (version 1.64b, (Thompson et al., 1997)) software. Phylogenetic analysis was carried out on amino acid sequences using PAUP* (ver. 4.0b10,(Swofford, 1998)) for parsimony analysis and Molphy (version 2.3 (Adachi and Hasegawa, 1996b) running at the UK HGMP Resource Center, Cambridge, UK) for Maximum Likelihood. Details of the analyses are given in the Results and Discussion section.

Results and Discussion

General organization

The entire sequence of the mitochondrial genome of *Tricholepidion gertschi* has been obtained, and deposited in GenBank. The molecule is circular, 15267 bp long, and encodes for the 13 protein encoding genes, 2 rRNAs and 22 tRNAs typical of metazoan mitochondrial genomes, plus a long uninterrupted non-coding region considered homologous to the A+T/Control region of other arthropods. Gene order is the same as in *Drosophila*. The genome is very compact, with only 83 nucleotides in unassigned regions. Genes overlap in 11 cases by 1-8 bp, for a total of 38 nucleotides.

Base composition

Base composition is 39.4% A, 19.4% C, 12.0% G, and 29.2% T, in conformity with the well known A+T bias of hexapod mitochondrial genomes. The content in bases is not uniform along the molecule, and a X^2 test for uniformity in base composition among the 13 protein encoding genes shows that the diversity is highly significant (df=36, P<0.00000). The variation of base composition along the molecule was studied by calculating values in a sliding window of 100 bp. There is significant variation in A and T content, their ratio varying from 2.28 in *nad4* to 0.89 in *atp6*.This suggests a strong correlation between coding direction and base composition. The A/T ratio (TAB1) averages 1,08 for PCGs encoded on the J strand (*sensu* (Simon et al., 1994)) (*nad2*, *cox1*, *cox2*, *atp8*, *atp6*, *cox3*, *nad3*, *nad6*, *cob*) and 2,12 for PCGs on the N strand (*nad5*, *nad4*, *nad4L*, *nad1*). A difference of some kind in base

composition is expected between PCGs encoded in different strands, from functional constraints due to codon base composition and usage. The outcome of that would be a bias that is symmetrical when base composition of genes from each of the two strands are compared. Contrarily, in this case, reverse/complementing genes on the J strand does not restore the base composition typical of genes on the N strand. Additionally, rRNA genes, although their values cannot be directly compared with PCGs due to different functional constrains, show the same tendency of PCGs encoded in the same strand (FIG2). This observation, besides the evident interest for genome-level evolutionary mechanisms, has a major outcome when analysing sequences for phylogenetic reconstructions. All methods of analysis employed so far assume, either implicitly or explicitly, that the probability of a certain substitution remains constant along all the tree. In the likely case of the inversion of one or more PCGs, and possibly rRNA genes, given the bias here described, we must assume a high number of directional mutations in the sequences that are translocated, in order to reequilibrate to the base composition typical of it's new genomic location. Preliminary observations on available complete mitochondrial genomes show that this bias is widespread, and further work is advisable to assess its effect on phylogenetic reconstructions.

Protein Coding Genes

The 13 PCGs typically encoded in Metazoan mitochondrial genomes have been unambiguously located in the molecule: *cox1-3*, *nad1-6* and *4L*, *cob*, *atp6* and *8*. The boundaries of each gene have been identified based on initiation and termination codons, and by sequence comparison with other published genomes, that helped in ambiguous situations. Ten genes begin with canonical initiation codons ATG (6), ATA (1), and ATT (3). The gene for *nad1* starts with TTG, rare among arthropods, but used in *Limulus polyphemus* (a chelicerate) and *Tetrodontophora bielanensis* (a collembolan). Based on aminoacidic alignments, GTG and CTA were assumed to initiate *nad2* and *cox1*, respectively. Neither of the two signalling sequences proposed to initiate *cox1* in other groups, ATAA (*Drosophila* and *Locusta*) and ATTTAA (mosquitoes, *Tetrodontophora*) was present in the area surrounding the initiation of the gene. Eight genes terminate with complete stop codons TAA (6), TAT (1) and TAG (1). Incomplete stop codons, to be completed into UAA by polyadenilation of the transcript, were assumed for the remaining genes, where T (2) or TA (3) at the end of a PCG were directly adjacent to the beginning of the following gene, a tRNA in all cases.

Transfer RNAs

Nineteen transfer RNA were found by the tRNAscan-SE software (version 1.21, (Lowe and Eddy, 1997)), three additional ones (*trnR*, *trnC* and $trnS^{(AGN)}$) were located by eye looking for possible cloverleaf structures in intergenic spacers. Individual tRNAs were assigned by the anticodon sequence, that were found to be identical to those of *Drosophila*. The total set of 22 tRNAs typical of metazoan mitochondrial genomes, one for each amino acid, two for Serine and Leucine, is thus present in the molecule. The structure of each tRNA was reconstructed (FIG3), and no evident abnormalities were found. A total of 9 mismatches were found in acceptor stems, and 6 in anticodon stems, and these are likely to be corrected post-trastcriptionally, given the recent evidences for this kind of modifications (Lavrov et al., 2000b; Yokobori and Paabo, 1995). The discriminator nucleotide is found in regions of gene overlap in 7 cases, and in 2 of these the overlap is limited to this single nucleotide, supporting the hypothesis that this could actually be added post-trascriptionally (Yokobori and Paabo, 1997).

Ribosomal RNAs

Genes were found that encode for the two mitochondrial rRNA subunits. The boundaries of these genes could not be assessed unambiguously by direct comparison with homologous sequences, due to high sequence variability in these areas. The secondary structure of both subunits was therefore reconstructed by hand based on published universal models, and this allowed exact localization of gene boundaries. A description of the evolution and variability of these structures will be given elsewhere (data in preparation).

Control Region

A long unassigned region is present between *rrnS* and *trnI*. This area is nucleotides long, with a base composition of 41,8%A, 12,3%C, 10,9%G, and 35%T. Given it's position, characteristics, and extreme base composition, this can be considered homologous to the A+T/Control region present in most arthropod mitochondrial genomes analysed so far, that contains the origin of replication and transcription. The whole sequence was searched for areas of internal complementarity, capable of forming secondary structures. The software RNAdraw (Matzura and Wennborg, 1996) evidentiated a close array of these structures, the most notable of which are depicted in FIG4. None of these can be homologated, based on consensus sequences and overall similarity, to those described by (Zhang et al., 1995) in many derived hexapods and by (Nardi et al., 2001) in the only basal hexapod examined so far. The evolutionary role of these structures has not been fully understood, even if they seem to play a role in recombination/splicing of these molecules (Stanton et al., 1994) and they may be part of the transcription/replication signalling.

Phylogenetic analysis

Gene order

The gene order of of *Tricholepidion gertschi* mitochondrial genome is identical to that of *Drosophila* and *Daphnia* (and most Hexapods and Crustaceans analysed so far), considered to be the plesiomorphic condition for all Pancrustacea (Crease, 1999). Therefore, gene order does not carry any phylogenetic information.

Sequence analysis

Phylogenetic analysys was carried out by comparing the sequence obtained for *Tricholepidion* with those of all Arthropods for which the complete mitochondrial sequence is available to date: *Limulus polyphemus* (NC_003057, (Lavrov et al., 2000a)), *Ixhodes hexagonus* (NC_002010, (Black and Roehrdanz, 1998)), *Riphicephalus sanguineus* (NC_002074, (Black and Roehrdanz, 1998)), *Lithobius forficatus* (NC_002629, (Lavrov et al., 2000b)), *Narceus annularus* (NC_003343, (Lavrov et al., 2002)), *Thyrophygus* sp. (NC_003344, (Lavrov et al., 2002)), *Pagurus longicarpus* (NC_003058, (Hickerson and Cunningham, 2000)), *Paenaeus monodon* (NC_002184, (Wilson et al., 2000)), *Daphnia pulex* (NC_000844, (Crease, 1999)), *Artemia franciscana* (NC_001620, (Valverde et al., 1994)), *Tetrodontophora bielanensis* (NC_002735, (Nardi et al., 2001)), *Locusta migratoria* (NC_001712; (Flook et al., 1995)), *Tribolium castaneum* (AJ312413, not published), *Crioceris duodecimpunctata* (NC_003372, not published), *Triatoma ditimidata* (AF301594, (Dotson and Beard, 2001)), *Bombyx mori* (NC_002355, not published), *Bombyx mandarina* (NC_003395, not published), *Ostrinia furnacalis* (NC_003368, not published), *Ostrinia nubilialis* (NC_003367, not published), *Anopheles gambiae* (NC_002084, (Beard et al., 1993)), *Anopheles quadrimaculatus* (NC_000875, (Mitchell et al., 1993)), *Drosophila yakuba* (NC_001322, (Clary and Wolstenholme, 1985)), *Drosophila melanogaster* (NC_001709, (Lewis et al., 1995)), *Crysomya chlorophyga* (NC_002097, not published), *Cochliomya hominivorax* (NC_002660, (Lessinger et al., 2000)), *Ceratitis capitata* (NC_000857, (Spanos et al., 2000)), *Apis mellifera* (NC_001566,

(Crozier and Crozier, 1993)), *Heterodoxus macropus* (NC_002351, (Shao et al., 2001)). Sequences of *Lumbricus terrestris* (NC_001673, (Boore and Brown, 1995)), *Platynereis dumerilii* (NC_000931, (Boore, 2001)), *Katarina tunicata* (NC_001636, (Boore and Brown, 1994)), *Lologo bleekeri* (NC_002507, (Sasuga et al., 1999)) and *Albinaria coerulea* (NC_001761, (Hatzoglou et al., 1995)) were used to mark the root of the tree. Thirteen separate alignments were produced, one for each PCG translated into aminoacids. These alignments were checked by eye, and, due to high levels of sequence variability, only those of *cox1*, *cox2*, *cox3* and *cob* were retained, with a single manual correction. The alignments of these 4 genes were concatenated, and positions experiencing gaps or uncertainties in at lest one of the sequences were excluded, mostly to account for small variation in length at the extreme of each sequence. This produced a dataset of 34 sequences of 1305 aminoacidic characters, that was analysed under both Maximum Parsimony and Maximum Likelihood. Unweighted parsimony (heuristic search in PAUP*, random addition, TBR branch swapping) produced two equally parsimonious trees of 6549 steps. Bootstrap support (1000 replications) was very low for most nodes, so that collapsing branches with support lower than 70 (5% significance, (Hillis and Bull, 1993)) yelded a highly unresolved tree, with the only supported nodes being at the ordinal level or below. Diptera, and Lepidoptera were recovered as monoplyletic, while no resolution was present for more ancient relationships. The same data set was analysed under the Maximum Likelihood criterion. Initially a search for the ML tree was carried out using "Quick add OTUs" heuristic search, applying the substitution matrix mtREV24 (Adachi and Hasegawa, 1996a) with frequency of each aminoacid estimated directly from the dataset. The 50 best trees were retained, and each of these was swapped and reoptimized under the same conditions. Most of the searches conveyed to the same tree (lnL=-36675,88) that is shown in FIG5. Local Bootstrap Probabilities (LBPs), were calculated during this last search by the RELL method, to introduce a measure of support at the nodes. The resulting tree is highly resolved, and most of the nodes show medium to high LBP support values. Arthropoda are recovered as a monophyletic group with respect to the five outgroup sequences. A first set of taxa, comprising the tree Myriapods, the tree Chelicerates, *Apis* and *Heterodoxus* form a monophyletic group, but relationships among these species are highly controversial, given that neither of the two classes is recovered as monophyletic, in clear contrast with morphological evidence. The clustering of *Apis* and *Heterodoxus* with the two arachnids, with high levels of support, is a common outcome of mitochondrial-based phylogenies of arthropods, and is probably an artifact due to shared base composition and, secondarily, long branch attraction (unpublished results). The Pancrustacea are recovered as monophyletic, and so are the Crustacea, although with levels of support below 70. The only enthognatous insect, the collembolan *Tetrodontophora bielanensis*, occupies the most basal position within the Pancrustacea, while the four Crustaceans cluster with the rest of the Hexapoda.

This reconstruction has two major implications. One confirms and earlier hypothesis (Nardi et al., 2001) that Hexapoda could be non monophyletic, since *Tetrodontophora* is basal to the node uniting crustaceans to the remaining ectognatan taxa. The second is that crustaceans, at least the species here analysed (4, from 3 orders and 2 classes), are monophyletic with respect to the hexapods. This is in agreement with some recent molecular evidences, but not with other mitochondrial-based analyses, that favour paraphyly of the Crustacea with respect to the Hexapoda (Garcia-Machado et al., 1999; Nardi et al., 2001). Relationships within the Crustacea follow the classical division between Branchiopoda (*Daphnia* and *Artemia*) and Malacostraca (*Pagurus* and *Paenaeus*). Among the Hexapoda, with the exception of *Apis*, *Heterodoxus* and *Tetrodontophora*, phylogenetic relationships broadly follow the accepted taxonomy, supporting the monophyly of Lepidoptera, Diptera, Coleoptera, Eterometabola and Holometabola, but not the supposed sister group relation between Lepidoptera and Diptera. *Tricholepidion* occupies a position basal to all the other ectognatan insects, in agreement with the classical interpretation. Within the Diptera, the main distinction between Nematocera (*Anopheles gambiae* and *A. quadrimaculatus*) and Brachycera (the remaining taxa) is supported. Among Brachycera, the monophyly of the three superfamilies Tephritoidea, Oestroidea and Ephydroidea is recovered, while

the distinction between Acalyptrate and Caliptrate flies is not, in contrast with the results of (Lessinger et al., 2000), although this is based on nodes supported by LBP values lower than 70%.

Conclusions

Generally speaking, sequence analysis of complete mitochondrial sequences seem to be a powerful tool to unravel phylogenetic relationships among Pancrustacea and Hexapods. Resolution seem to be stronger for more recent relationships, but, with the appropriate statistical and methodological tools, some signal can be recovered also for more ancient relationships. Specifically, the placement of *Tricholepidion* basal to the Insect clade seem to be robust. While answering the question over the correct placement of *Tricholepidion* (with the Zygentoma or basal to all Dicondylia), was not the direct question of this paper, the determination of the complete mitochondrial sequence of *T.gertschi* certainly represent an important step in that direction. Work is in progress to fill this gap. By looking back at the phylogenetic analyses that in recent years have considered complete mitochondrial genomes of arthropods (Flook et al., 1995; Garcia-Machado et al., 1999; Nardi et al., 2001; Wilson et al., 2000), it is evident that resolution at most nodes is greatly improved as more taxa are included. This is certainly true for Hexapods and Crustacea, and may be the case also for Myryapoda and Chelicerata, although the availability of sequences for these latter groups is still extremely scarce and highly biased towards certain taxa.

Parallel to the accumulation of complete mitochondrial sequence data, the knowledge of evolutionary mechanisms and idiosyncrasies of mitochondrial genomes increases, and this is of great help, in turn, for the choice of the right analytical tools to analyse these sequences in a phylogenetic framework. The description of this asymmetrical direction-dependent bias in base frequencies is of great interest, and the effects in terms of evolutionary patterns and phylogenetic reconstructions deserve more in-depth studies.

Acknowledgements

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, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

Table 1 – Base composition of different areas of the molecule calculated for the J strand.

	Strand	A	T	A-T/A+T	G	C	G-C/G+C
Entire molecule	N/A	39.4	29.2	0.15	12.0	19.4	-0.24
nad2	J	36.5	30.8	0.08	10.4	22.3	-0.36
cox1	J	31.5	31.9	-0.01	16.8	19.8	-0.08
cox2	J	34.6	34.6	0.00	14.3	16.5	-0.07
atp8	J	40.7	34.0	0.09	4.9	20.4	-0.61
atp6	J	32.8	36.7	-0.06	11.8	18.7	-0.23
cox3	J	32.3	31.3	0.02	14.8	21.6	-0.19
nad3	J	36.3	33.4	0.04	11.5	18.8	-0.24
nad5	N	45.5	23.8	0.31	10.6	20.1	-0.31
nad4	N	48.1	21.1	0.39	10.8	20.0	-0.30
nad4L	N	48.7	22.0	0.38	7.6	21.7	-0.48
nad6	J	39.9	33.1	0.09	7.5	19.5	-0.44
cob	J	34.4	30.5	0.06	13.3	21.8	-0.24
nad1	N	46.7	22.3	0.35	12.0	19.0	-0.23
rrnL	N	41.9	30.6	0.16	9.4	18.1	-0.32
rrnS	N	42.8	26.3	0.24	11.2	19.7	-0.28
Control Region	N/A	41.8	35.0	0.09	10.9	12.3	-0.06

Adachi, J. and M. Hasegawa. 1996a. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J Mol Evol 42: 459-68.

Adachi, J. and M. Hasegawa. 1996b. MOLPHY version 2.3:programs for molecular phylogenetics based on maximum likelyhood. Computer Science Monographs 28: 1-150.

Beard, C. B., D. M. Hamm and F. H. Collins. 1993. The mitochondrial genome of the mosquito Anopheles gambiae: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Molecular Biology 2: 103-24.

Beutel, R. G. 2001. Ultrastructure of attachment specializations of hexapods (Arthropoda): evolutionary patterns inferred from a revised ordinal phylogeny. J. Zool. Syst. Evol. Research 39: 177-207.

Bilinski, S. 1993. Structure of ovaries and oogenesis in Enthognathans (Apterygota). Int. J. Insect Morphol. Embriol. 22: 255-269.

Bitsch, C. and J. Bitsch. 2000. The phylogenetic interrelationships of the higher taxa of apterygote hexapods. Zoologica Scripta 29: 131-156.

Black, W. C. I. and R. L. Roehrdanz. 1998. Mitochondrial gene order is not conserved in arthropods: Prostriate and metastriate tick mitochondrial genomes. Molecular Biology and Evolution 15: 1772-1785.

Boore, J. L. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. Kluwer Academic Publishers, pp. 133-147.

Boore, J. L. 2001. Complete mitochondrial genome sequence of the polychaete annelid Platynereis dumerilii. Mol Biol Evol 18: 1413-6.

Boore, J. L. and W. M. Brown. 1994. Complete DNA sequence of the mitochondrial genome of the black chiton, Katharina tunicata. Genetics 138: 423-43.

Boore, J. L. and W. M. Brown. 1995. Complete sequence of the mitochondrial DNA of the annelid worm Lumbricus terrestris. Genetics 141: 305-319.

Boore, J. L. and W. M. Brown. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opin Genet Dev 8: 668-74.

Clary, D. O. and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of Drosophila yakuba: Nucleotide sequence, gene organization, and genetic code. Journal of Molecular Evolution 22: 252-271.

Crease, T. J. 1999. The complete sequence of the mitochondrial genome of Daphnia pulex (Cladocera: Crustacea). Gene (Amsterdam) 233: 89-99.

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

Dallai, R. 1994. Recent findings on apterygotan sperm structure. Acta Zool. Fennica 195: 23-27.

Dallai, R., P. Lupetti, F. Frati, F. Nardi and B. A. Afzelius. 2001. Binucleate and biflagellate spermatozoa in Tricholepidion gertschi Wygodzynsky (Insecta, Zygentoma). Tissue Cell 33: 606-13.

Dotson, E. M. and C. B. Beard. 2001. Sequence and organization of the mitochondrial genome of the Chagas disease vector, Triatoma dimidiata. Insect Mol Biol 10: 205-15.

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

Garcia-Machado, E., M. Pempera, N. Dennebouy, M. Oliva-Suarez, J. C. Mounolou and M. Monnerot. 1999. Mitochondrial genes collectively suggest the paraphyly of Crustacea with respect to Insecta. J Mol Evol 49: 142-9.

Hatzoglou, E., G. C. Rodakis and R. Lecanidou. 1995. Complete sequence and gene organization of the mitochondrial genome of the land snail Albinaria coerulea. Genetics 140: 1353-66.

Hickerson, M. J. and C. W. Cunningham. 2000. Dramatic mitochondrial gene rearrangements in the hermit crab Pagurus longicarpus (Crustacea, Anomura). Molecular Biology and Evolution 17: 639-644.

Hillis, D. M. and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42: 182-192.

Hwang, U. W., M. Friedrich, D. Tautz, C. J. Park and W. Kim. 2001. Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature 413: 154-7.

Koch, M. 1997. Monophyly and phylogenetic position of the Diplura (Hexapoda). Pedobiologia 41: 9-12.

Kristensen, N. P. 1981. Phylogeny of insect orders. Annual Review of Entomology 26: 135-157.

Kristensen, N. P. 1997. The groundplan and basal diversification of the hexapods. *In* R. A. Fortey and R. H. Thomas (eds), The groundplan and basal diversification of the hexapods. Chapman and Hall, London, pp. 282-293.

Kukalova Peck, J. 1987. New Carboniferous Diplura, Monura and Thysanura, the hexapod ground plan, and the role of toracic lobe in the origin of wings (Insecta). Can. J. Zool. 65: 2327-2345.

Lavrov, D. V., J. L. Boore and W. M. Brown. 2000a. The complete mitochondrial DNA sequence of the horseshoe crab Limulus polyphemus. Molecular Biology and Evolution 17: 813-824.

Lavrov, D. V., J. L. Boore and W. M. Brown. 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: Duplication and non-random loss. Mol Biol Evol in press.

Lavrov, D. V., W. M. Brown and J. L. Boore. 2000b. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede Lithobius forficatus. Proc Natl Acad Sci U S A 97: 13738-42.

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

Lewis, D. L., C. L. Farr and L. S. Kaguni. 1995. Drosophila melanogaster mitochondrial DNA: Completion of the nucleotide sequence and evolutionary comparisons. Insect Molecular Biology 4: 263-278.

Lowe, T. M. and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955-64.

Matzura, O. and A. Wennborg. 1996. RNAdraw: An integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. Computer Applications in the Biosciences 12: 247-249.

Mitchell, S. E., A. F. Cockburn and J. A. Seawright. 1993. The mitochondrial genome of Anopheles quadrimaculatus species A: Complete nucleotide sequence and gene organization. Genome 36: 1058-1073.

Nardi, F., A. Carapelli, P. P. Fanciulli, R. Dallai and F. Frati. 2001. The Complete Mitochondrial DNA Sequence of the Basal Hexapod Tetrodontophora bielanensis: Evidence for Heteroplasmy and tRNA Translocations. Mol Biol Evol 18: 1293-304.

Sasuga, J., S. Yokobori, M. Kaifu, T. Ueda, K. Nishikawa and K. Watanabe. 1999. Gene contents and organization of a mitochondrial DNA segment of the squid Loligo bleekeri. J Mol Evol 48: 692-702.

Shao, R., N. J. Campbell and S. C. Barker. 2001. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, Heterodoxus macropus (Phthiraptera). Mol Biol Evol 18: 858-65.

Silvestri, F. 1912. Die Thysanuren des baltischen Bernsteins. Schrift. physik.-okonom. Ges Konigsberg 53: 42-66.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87: 651-701.

Spanos, L., G. Koutroumbas, M. Kotsyfakis and C. Louis. 2000. The mitochondrial genome of the Mediterranean fruit fly, Ceratitis capitata. Insect Molecular Biology 9: 139-144.

Staniczek, A. 2000. The mandible of Silverfish (Insecta: Zygentoma) and Mayflies (Ephemeroptera): it's morphology and phylogenetic significance. Zoologisher Anzeiger 239: 147-178.

Stanton, D. J., L. L. Daehler, C. C. Moritz and W. M. Brown. 1994. Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. Genetics 137: 233-241.

Stys, P. and S. Bilinski. 1990. Ovariole type and the phylogeny of hexapods. Biol. Rev. Cambridge Phyl. Soc.: 401-429.

Swofford, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sinauer, Sunderland, Mass.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins. 1997. The CLUSTAL-X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876-4882.

Valverde, J. R., B. Batuecas, C. Moratilla, R. Marco and R. Garesse. 1994. The complete mitochondrial DNA sequence of the crustacean Artemia franciscana. Journal of Molecular Evolution 39: 400-408.

Wilson, K., V. Cahill, E. Ballment and J. Benzie. 2000. The complete sequence of the mitochondrial genome of the crustacean Penaeus monodon: Are malacostracan crustaceans more closely related to insects than to branchiopods? Molecular Biology and Evolution 17: 863-874.

Wygodzinsky, P. 1961. On a surviving representative of the Lepidotrichidae (Thysanura). Ann. Entomol. Soc. Am. 54: 621-627.

Yokobori, S.-I. and S. Paabo. 1995. Transfer RNA editing in land snail mitochondria. Proceedings of the National Academy of Sciences of the United States of America 92: 10432-10435.

Yokobori, S.-I. and S. Paabo. 1997. Polyadenylation creates the discriminator nucleotide of chicken mitochondrial tRNA-Tyr. Journal of Molecular Biology 265: 95-99.

Zhang, D.-X., J. M. Szymura and G. M. Hewitt. 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. Journal of Molecular Evolution 40: 382-391.