Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: Duplication and non-random loss.

Dennis V. Lavrov[1,2], Jeffrey L. Boore[1,3] and Wesley M. Brown[1]

1 Department of Biology, University of Michigan

830 North University Avenue, Ann Arbor, MI 48109-1048 USA

2 Departement de Biochimie, Universite de Montreal, C.P. 6128, Montreal, QC H3C3J7 Canada

3 DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598 USA

Corresponding author: Dennis Lavrov, Département de Biochimie, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Qué., H3C 3J7, Canada; Tel. (514) 343-6111 ext. 5188, Fax (514) 343-2210, email dlavrov@bch.umontreal.ca

Keywords: mitochondrial DNA, gene rearrangement, genome duplication, phylogenetic inference, diplopoda

Running head: Gene rearrangement in millipede mtDNA

ABSTRACT

We determined the complete mtDNA sequences of the millipedes *Narceus annularus* and *Thyropygus* sp. (Arthropoda: Diplopoda) and identified in both genomes all 37 genes typical for metazoan mtDNA. The arrangement of these genes is identical in the two millipedes, but differs from that inferred to be ancestral for arthropods by the location of four genes/gene clusters. This novel gene arrangement is unusual for animal mtDNA, in that genes with opposite transcriptional polarities are clustered in the genome and the two clusters are separated by two non-coding regions. The only exception to this pattern is the gene for cysteine tRNA, which is located in the part of the genome that otherwise contains all genes with the opposite transcriptional polarity. We suggest that a mechanism involving complete mtDNA duplication followed by the loss of genes, predetermined by their transcriptional polarity and location in the genome, could generate this gene arrangement from the one ancestral for arthropods. The proposed mechanism has important implications for phylogenetic inferences that are drawn on the basis of gene arrangement comparisons.

Introduction

Animal mitochondrial DNA is typically a circular molecule of 15-17 kb that encodes 37 genes: 13 for proteins [subunits 6 and 8 of the F_0 ATPase (*atp6* and *atp8*), cytochrome *c* oxidase subunits 1-3 (*cox1-cox3*), apocytochrome *b* (*cob*) and NADH dehydrogenase subunits 1-6 and 4L (*nad1-6* and *nad4L*)]; two for ribosomal RNAs [small and large subunit rRNAs (*rrnS* and *rrnL*)]; and 22 for tRNAs (Boore 1999). In addition, a single large non-coding region is typically present which, for a few animals, is known to contain sequences essential for the initiation of transcription and mtDNA replication (Shadel and Clayton 1997).

Arrangements of metazoan mitochondrial genes are relatively stable, with some having been conserved for hundreds of millions of years [e.g., most vertebrates share an identical arrangement, as do some crustaceans and insects (Boore 1999; unpublished data)]. Only minor rearrangements separate the gene orders of human and the cephalochordate *Branchiostoma floridae*, or *Drosophila yakuba* and the chelicerate *Limulus polyphemus*, whereas numerous, but identifiable, rearrangements separate those of chordates and arthropods (Clary and Wolstenholme 1985; Boore 1999). Because of the low rearrangement rate as well as several other characteristics, mitochondrial gene arrangements promise to be a useful dataset for the study of deep metazoan divergences (Boore and Brown 1998). Indeed, several problematic relationships have already been convincingly resolved using this dataset, including those among classes of echinoderms (Smith et al. 1993) and of arthropods (Boore, Lavrov, and Brown 1998). However, our limited knowledge of gene rearrangement mechanisms hampers our ability to interpret

this dataset and is an impediment to its broader acceptance for phylogenetic studies (Curole and Kocher 1999).

The most detailed and the best supported model explains gene rearrangements by partial duplication of mtDNA due to errors in replication, such as erroneous initiation or termination (Macey et al. 1997) or strand slippage and mispairing (Madsen, Ghivizzani, and Hauswirth 1993), followed by the loss of one copy of each duplicated gene (Moritz, Dowling, and Brown 1987; Boore 2000). This model is supported by the observation of mtDNAs containing duplicated regions (Moritz and Brown 1986; Moritz and Brown 1987), most of which are adjacent to or include the non-coding region; inactivating mutations are frequently found in one copy of each of the genes (Stanton et al. 1994; Arndt and Smith 1998; Kumazawa et al. 1998; Macey et al. 1998). It is commonly assumed that the loss of one of two copies of each duplicated gene happens at random, and the model described above is sometimes referred to as the "duplication-random loss" model (Moritz, Dowling, and Brown 1987; Boore and Brown 1998; Boore 1999; Boore 2000).

Here we describe the mitochondrial DNA of two millipedes. These share a novel gene order, which we propose has been generated from the arrangement that is primitive for arthropods by a novel mechanism: mtDNA duplication and non-random gene loss. According to this model, the destiny of each gene copy in the duplicated region is predetermined by its transcriptional polarity and location in the genome. We discuss the implications of this for phylogenetic inferences based on gene arrangement comparisons.

Materials and methods

A specimen of *Narceus annularus* (order Spirobolida) was collected by Dr. Barry OConner near the Biological Station of the University of Michigan in Northern Michigan; a specimen of *Thyropygus* sp. (order Spirostreptida) was purchased from WARD'S Natural Science Establishment, Inc. Total DNA from body wall muscles or from eggs was prepared as described (Saghai-Maroof et al. 1984). For *N. annularus*, small fragments of *cox3* and *rrnS* were amplified and sequenced and two pairs of primers were designed based on these sequences: Narceus-cox3-F1 (5'-CGTAGCAACAGGCTTTCATGGAC-3'), Narceus-cox3-R1 (5'-AGTCCATGAAAGCCTGTTGCTAC-3'), Narceus-rrnS-F1 (5'-CATAGTCTGAGGGACGTCAAGTC-3') and Narceus-rrnS-R1 (5'-CCTTGACTTGACGTCCCTCAGAC-3'). For *Thyropygus* sp., small fragments of *rrnS* and *cob* were amplified and sequenced and two pairs of primers were designed based on these sequences: Thyropygus-rrnS-F1 (5'-AGGACGTCAAGTCAAGGTGCAGC-3'), Thyropygus-rrnS-R1 (5'-AATCCACCTTCATGATGCACTTC-3'), Thyropygus-cob-F1 (5'-GGATTTGCAGTAGACAATGCCAC-3') and Thyropygus-cob-R1 (5'-GGTGAATAATAACTATGGCTGCGA-3'). The whole mtDNA of each millipede was amplified in two overlapping fragments by using the Perkin Elmer's® XL PCR kit or TaKaRa LA-PCR kit and primer pairs Narceus-rrnS-F1 - Narceus-cox3-F1, Narceus-rrnS-R1 - Narceus-cox3-R1, Thyropygus-rrnS-F1 - Thyropygus-cob-F1 and Thyropygus-rrnS-R1 - Thyropygus-cob-R1. Each PCR reaction yielded a single band when visualized with ethidium bromide staining after electrophoresis in a 1% agarose gel.

PCR reaction products were purified by three serial passages through Ultrafree™ (30,000 NMWL) columns (Millipore) and used as templates in dye-terminator cycle-sequencing reactions according to supplier's (Perkin Elmer®) instructions. Both strands of each amplification product were sequenced by primer walking, using an ABI 377 Automated DNA Sequencer. The sequences of *N. annularus* and *Thyropygus* sp. mtDNAs have been submitted to GenBank under accession numbers AY055727 and AY055728, respectively.

Results and Discussion

Millipede mitochondrial genomes

The complete mtDNA sequences from the millipedes *Narceus annularus* and *Thyropygus* sp. (fig. 1) are 14868 and 15133 bp in size, respectively, and contain all 37 genes typical for animal mtDNA. These genes have an arrangement that differs from any other found in arthropod mtDNA by the positions of at least four genes/gene clusters. In addition, the millipede gene arrangement is unusual because, with the single exception of *trnC*, genes that are transcribed in opposite directions are located in two different parts of the genome and separated by two non-coding regions (fig. 1). By contrast, in most other animal mtDNAs studied, genes are either transcribed from the same strand or arranged in several clusters with alternating transcriptional polarity, and a single large non-coding region is present (Boore 1999). While it is possible that the millipede gene arrangement is the result of four independent rearrangements (fig. 2A) or a large duplication followed by the random loss of genes, the unusual clustering of genes with opposite transcriptional

polarity suggests that a different, non-random mechanism may have generated this gene order.

Gene rearrangement mechanism

The proposed mechanism is presented in fig. 2B. We assume that the original mitochondrial genome had a gene arrangement identical to that in *Limulus polyphemus*, which appears to be ancestral for arthropods (Lavrov, Boore, and Brown, 2000; unpublished data), and that both of the mtDNA strands were transcribed in their entirety either from the same bi-directional promoter or from two unidirectional promoters, both located in the same non-coding region. Such arrangements have been found in vertebrate mtDNAs, in which most studies on mitochondrial transcription mechanisms have been conducted (Tracy and Stern 1995), and are consistent with the generally observed scarcity of intergenic sequences in arthropod mitochondrial genomes that otherwise could serve as transcriptional promoters (but see Clary, Wahleithner, and Wolstenholme 1983). We further assume that there are unidirectional transcription termination signals located in the same non-coding region of the mtDNA. The presence of such signals has been inferred based on RNase mapping experiments in rats (Sbisa et al. 1990); however, no mechanistic details are known for this or for other animal species.

We envision the first step leading to the millipede gene arrangement to have been a tandem duplication of the entire mtDNA, resulting in a dimeric molecule with the two monomers covalently linked head to tail. Such complex molecules are usually associated with cellular abnormalities and tissue cultures (Clayton and Vinograd 1967;

Wolstenholme et al. 1973), but have also been observed in the normal tissues of rat (Wolstenholme et al. 1973), several species of Drosophila (Shah and Langley 1977) and the isopod *Armadillidium vulgare* (Raimond et al. 1999). In addition to duplicated structural genes, such a dimeric molecule would contain duplicated transcriptional regulators located in the two copies of the non-coding regions. If the functionality of transcriptional promoters in one of the non-coding regions was lost or severely impaired (e.g., by a deletion encompassing the two unidirectional promoters or by a point mutation in a single bi-directional promoter) while the accompanying sets of transcriptional termination signals were retained, then the sets of genes under the control of the disabled promoter(s) would immediately become pseudogenes, and the sequences containing them would be free to mutate and to disappear from the genome (fig. 2B). As a result, the loss of genes would be predetermined by their transcriptional polarity. All genes having one polarity would be lost from one genome copy, and all genes having the opposite polarity would be lost from the other.

The gene order in the two millipede mtDNAs corresponds to the hypothetical gene order that would be derived from the primitive arthropod arrangement by the proposed mechanism, but with two discrepancies present. The first, already mentioned, is the location of *trnC* in the part of the genome that otherwise contains all genes with the opposite transcriptional polarity. The second is the location of *trnT*, which is different from the one expected after the rearrangement proposed. Neither discrepancy necessarily invalidates the proposed mechanism. The product of *trnC* carries the least frequent amino acid in mitochondrial proteins. There are only 31 cysteine residues encoded in *N. annularus* mitochondrial protein genes and 30 in those of *Thyropygus* sp., less than one

percent of the total number of amino acids in both cases. Thus, a weak residual level of transcription could potentially produce a sufficient supply of this gene product. The unexpected location of *trnT* can be explained by an independent translocation event, such as depicted in fig. 2B. Such translocations are relatively common in animal mitochondrial genomes (Boore 1999), and that of *trnT* could have happened either before or after the duplication and loss events postulated in the model. Further sampling of this group of animals may provide data with which to test this hypothesis.

Support for the proposed model

The model proposed here has greater explanatory power for the arrangement found in two millipede mitochondrial genomes than the duplication-random loss model. Unlike the latter model, it predicts the pattern of the gene loss and, therefore, the resulting gene order based on the genes' transcriptional polarities and their positions in the genome. It also explains the presence and predicts the locations of the two non-coding regions that are found in genomes that otherwise have very few non-coding nucleotides (fig. 1).

Indirect support for the proposed model also comes from the study of the mitochondrial genome of the red alga *Chondrus crispus* (Leblanc et al. 1995). This protist has a relatively small and gene-rich mitochondrial genome, with an overall genome organization surprisingly similar to that found in millipedes. As in millipedes, the *C. crispus* mitochondrial genes that are transcribed in opposite directions are located in two different parts of the genome and separated by two non-coding regions. Also as in

millipedes, a single tRNA gene (*trnH*) defies this rule and is located in the "wrong" part of the genome. In addition, the mitochondrial genomes of the millipedes and *C. crispus* share the presence of two large stem-loop structures in the non-coding regions. In *C. crispus* and *N. annularus*, one stem-loop structure is found in each of the two non-coding regions, while in *Thyropygus* sp. both structures are present in the same non-coding region (fig. 3). The study of mitochondrial transcription by Northern hybridization in *C. crispus* revealed the presence of two large transcriptional units, with transcription initiating in one non-coding region (corresponding to the non-coding region adjacent to *trnI* and *rrnS* in millipedes) and terminating in the other non-coding region in one direction and near *trnH* in the other (Richard et al. 1998). Interestingly, while it has been suggested that the stem-loop structures found in *C. crispus* mtDNA may be involved in transcription initiation and termination (Leblanc et al. 1995; Richard et al. 1998), those in millipede mtDNA resemble the stem-loop structure at the origin of light strand replication in vertebrates (Wong and Clayton 1985). If these are functional analogues, then a novel replication mechanism may exist which utilizes both of these structures, perhaps for the separate initiation of replication on each strand of millipede mtDNA.

Implications for phylogenetic studies

Most studies comparing mitochondrial gene arrangements for phylogenetic inference have not considered the mechanisms for gene rearrangements, but have made the tacit assumption that gene rearrangements are random events in which there is an extremely small probability of convergence. Some studies have also used combinatorics

to estimate the significance of shared gene boundaries, further assuming that each rearrangement has an equal chance to occur (Noguchi et al. 2000). It is clear, however, that if a rearrangement happens by the duplication-loss mechanism, the resulting gene order will not be completely random since *i*) the transcriptional polarity of the genes would not change, *ii*) a gene would only be able to "relocate" within the boundaries of the duplicated region, and *iii*) not all rearrangements that preserve the transcriptional polarity of the genes in this region would be obtainable in a single duplication-loss event (e.g., duplication of ABC to ABCABC cannot lead to CBA by gene loss alone). In addition, the presence of duplication "hot spots", as revealed by comparisons of mtDNAs from different species (Boore and Brown 1998; Boore 1999) and by analysis of different mtDNA molecules from the same organism (Moritz and Brown 1986; Moore, Gudikote, and Van Tuyle 1998), contributes further to non-randomness in mitochondrial gene rearrangements. Some of these considerations have led to our earlier suggestions for down weighting the sharing of certain types of changes (such as exchange of nearest neighbor tRNA genes and movements of genes flanking the control region) in phylogenetic analysis (Boore and Brown 1998).

The present study adds a new dimension to the problem by showing that the destiny of genes in a duplicated region may be determined by biological constraints, rather than by chance. If a duplication includes a non-coding region containing transcriptional control sequences, a condition that appears to be common among metazoan mtDNA duplications, an inactivating mutation in one of those sequences could facilitate the nearly simultaneous and non-random loss of an entire subset of the duplicated genes. If, as proposed here, the entire mtDNA duplicates with the monomers

linked head to tail, it would lead to gene arrangements similar to those found in millipede and red alga mtDNAs. If, instead, duplication resulted in a dimeric mtDNA molecule in which two monomers were linked head to head so that two copies of each gene had the opposite transcriptional polarities, an inactivating mutation in one set of transcriptional control sequences would result in an arrangement in which all genes were transcribed in the same direction. This latter variation of the model may explain the relatively frequent and apparently independent emergence of such arrangements during metazoan mtDNA evolution: mtDNAs having all the genes transcribed from the same strand has been found for some species of nematodes, mollusks, annelids, brachiopods and flatworms (Boore 1999).

Some additional constraints on mitochondrial genome organization could also cause non-random gene loss after a duplication. A transcriptional attenuator is located downstream from the two ribosomal genes in human mtDNA (Montoya, Gaines, and Attardi 1983), and is likely to occur in many other animal mtDNAs (Valverde, Marco, and Garesse 1994). This may constrain the position of the rRNA genes relative to the origin of transcription, to other genes, and to each other. If and when the secondary structures of tRNA sequences serve as the processing signals for mitochondrial primary transcripts, as has been proposed (Ojala, Montoya, and Attardi 1981), the possible rearrangements would be restricted to those that have at least one tRNA gene between each pair of neighboring protein and/or rRNA genes. The conservation of some mitochondrial gene arrangements across diverse phylogenetic groups may also be the result of the physical interaction among their gene products, some of which may require co-translational folding (Dandekar et al. 1998). Possible candidates for this include the

pairs *nad4L-nad4* and *atp8-atp6* which, at least in some animals, are translated from the same messenger RNA. Thus, selection for a certain gene order may operate even in cases in which a complete mtDNA is transcribed as a single operon. An additional constraint on rearrangement may occur if subsets of genes are transcribed from different transcriptional units, as has been reported for the sea urchin *Paracentrotus lividus* (Cantatore et al. 1990).

Mitochondrial gene rearrangements are rare events in animal evolution and, therefore, appear to be well suited for the study of ancient relationships. However, as we accumulate more and more gene arrangement data and as we try to use them for global phylogenetic studies, we must also consider the possible mechanisms that underlie the rearrangements, both to improve our inference of evolutionary relationships and to be able to evaluate the robustness of our conclusions. The model we present provides a model to explain a specific gene rearrangement that has been observed, and is an important step in this direction.

Acknowledgements

This work was supported by the NSF dissertation improvement grant (DEB 9972712) to W.M.B and D.V.L, NSF grant (DEB 9807100) to W.M.B. and J.L.B., and the University of Michigan predoctoral fellowship to D.V.L. Part of this work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098. Thanks to John Moran, B. Franz Lang Kevin Helfenbein and two anonymous reviewers for helpful comments on the manuscript

LITERATURE CITED

Arndt, A., and M. J. Smith. 1998. Mitochondrial gene rearrangement in the sea cucumber genus Cucumaria. Mol. Biol. Evol. **15**:1009–1016.

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. Pp. 133–147 *in* D. Sankoff and J. H. Nadeau, eds. Comparative Genomics. Kluwer Academic Publishers, Dordrecht; Boston; London.

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**:668–674.

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

Cantatore, P., M. Roberti, P. L. Polosa, A. Mustich, and M. N. Gadaleta. 1990. Mapping and characterization of *Paracentrotus lividus* mitochondrial transcripts: multiple and overlapping transcription units. Curr. Genet. **17**:235–245.

Clary, D. O., J. A. Wahleithner, and D. R. Wolstenholme. 1983. Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes. Nucl. Acids Res. **11**:2411–2425.

Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. **22**:252–271.

Clayton, D. A., and J. Vinograd. 1967. Circular dimer and catenate forms of mitochondrial DNA in human leukaemic leucocytes. Nature **216**:652.

Curole, J. P., and T. D. Kocher. 1999. Mitogenomids: digging deeper with complete mitochondrial genomes. Trends Ecol. Evol. **14**:394–398.

Dandekar, T., B. Snel, M. Huynen, and P. Bork. 1998. Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem. Sci. **23**:324–328.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1998. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. Genetics **150**:313–329.

Lavrov, D. V., J. L. Boore, and W. M. Brown. 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Mol. Biol. Evol. **17**:813–824.

Leblanc, C., C. Boyen, O. Richard, G. Bonnard, J. M. Grienenberger, and B. Kloareg. 1995. Complete sequence of the mitochondrial DNA of the rhodophyte *Chondrus crispus* (Gigartinales). Gene content and genome organization. J. Mol. Biol. **250**:484–495.

Macey, J. R., A. Larson, N. B. Ananjeva, Z. Fang, T. J. Papenfuss, and R. J. Macey. 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.

Macey, J. R., J. A. Schulte II, A. Larson, and T. J. Papenfuss. 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol. Biol. Evol. **15**:71–75.

Madsen, C. S., S. C. Ghivizzani, and W. W. Hauswirth. 1993. In vivo and in vitro evidence for slipped mispairing in mammalian mitochondria. Proc. Natl. Acad.

Sci. USA **90**:7671–7675.

Montoya, J., G. L. Gaines, and G. Attardi. 1983. The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell **34**:151–159.

Moore, C. A., J. Gudikote, and G. C. Van Tuyle. 1998. Mitochondrial DNA rearrangements, including partial duplications, occur in young and old rat tissues. Mutat. Res. -Fundam. Mol. Mech. Mutagen. **421**:205–217.

Moritz, C., and W. M. Brown. 1986. Tandem duplications of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science **233**:1425–1427.

Moritz, C., and W. M. Brown. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA **84**:7183–7187.

Moritz, C., T. E. Dowling, and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18**:269–292.

Noguchi, Y., K. Endo, F. Tajima, and R. Ueshima. 2000. The mitochondrial genome of the brachiopod *Laqueus rubellus*. Genetics **155**:245–259.

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

Raimond, R., I. Marcade, D. Bouchon, T. Rigaud, J. P. Bossy, and C. Souty-Grosset. 1999. Organization of the large mitochondrial genome in the isopod *Armadillidium vulgare*. Genetics **151**:203–210.

Richard, O., G. Bonnard, J. M. Grienenberger, B. Kloareg, and C. Boyen. 1998.

Transcription initiation and RNA processing in the mitochondria of the red alga *Chondrus crispus*: convergence in the evolution of transcription mechanisms in mitochondria. J. Mol. Biol. **283**:549–557.

Saghai-Maroof, M. A., K. M. Soliman, R. A. Jorgensen, and R. W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA **81**:8014–8018.

Sbisa, E., M. Nardelli, F. Tanzariello, A. Tullo, and C. Saccone. 1990. The complete and symmetric transcription of the main non coding region of rat mitochondrial genome: in vivo mapping of heavy and light transcripts. Curr. Genet. **17**:247–253.

Shadel, G. S., and D. A. Clayton. 1997. Mitochondrial DNA maintenance in vertebrates. Ann. Rev. Biochem. **66**:409–435.

Shah, D. M., and C. H. Langley. 1977. Complex mitochondrial DNA in Drosophila. Nucl. Acids Res. **4**:2949–2960.

Smith, J. M., A. Arndt, S. Gorski, E. Fajber, and M. J. Smith. 1993. The phylogeny of echinoderm classes based on mitochondrial gene arrangements. J. Mol. Evol. **36**:545–554.

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.

Tracy, R. L., and D. B. Stern. 1995. Mitochondrial transcription initiation: promoter structures and RNA polymerases. Curr. Genet. **28**:205–216.

Valverde, J. R., R. Marco, and R. Garesse. 1994. A conserved heptamer motif for

ribosomal RNA transcription termination in animal mitochondria. Proc. Natl. Acad. Sci. USA **91**:5368–5371.

Wolstenholme, D. R., J. D. McLaren, K. Koike, and E. L. Jacobson. 1973. Catenated oligomeric circular DNA molecules from mitochondria of malignant and normal mouse and rat tissues. J. Cell Biol. **56**:247–255.

Wong, T. W., and D. A. Clayton. 1985. In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. Cell **42**:951–958.

Fig. 1.—The gene maps of *N. annularus* and *Thyropygus* sp. mtDNA. Protein and ribosomal RNA genes are abbreviated as in the text, transfer RNA genes are identified by the one letter code for the corresponding amino acid. Two leucine and two serine tRNA genes are differentiated by their anticodon sequence with *trnL*(*uag*) marked as L1, *trnL*(*uaa*)- as L2, *trnS*(*ucu*)- as S1 and *trnS*(*uga*)- as S2. The direction of transcription for each gene is shown by an arrow. Filled areas depict large non-coding regions, the number attached to each indicates its size. Positive numbers at gene boundaries indicate the number of intergenic nucleotides; negative numbers indicate the number of overlapping nucleotides. Asterisks mark incomplete stop codons (T or TA) that are presumably completed by the addition of 3' A residues to the mRNA.

Fig. 2.—Comparison of gene arrangements in mtDNA of *Limulus polyphemus* and two millipedes (*A*) and the mechanism proposed for the generating of the latter gene arrangement (*B*). The rearrangement of genes or gene blocks in *A* is shown by arrows. The proposed position for transcription initiation and termination in *B* are marked by TI and TT. Genes are not to scale; protein and rRNA genes are indicated by bigger boxes, tRNA genes by smaller boxes, non-coding regions are dotted. Genes are transcribed from left to right except when underlined; underlining indicates the opposite transcriptional polarity. The copies of the genes that became pseudogenes in a hypothetical intermediate arrangement are indicated by filled boxes, the unexpected pattern of retention/loss for two copies of *trnC* (marked by ovals) is discussed in the text. Genes are abbreviated as in Fig. 1.

Fig. 3.—Potential stem loop structures in the non-coding regions of millipede mtDNA. Structural genes adjacent to non-coding regions are shown as arrows; their transcriptional polarities are indicated by the arrows' directions. Runs of thymidines/adenosines in the loop regions are in boldface. Boldface T at the base of the second stem-loop structure in *Thyropygus* sp. mtDNA is inferred to be part of *trnQ*. The 5' end of *N. annularus rrnS* directly adjacent to the stem-loop structure has been inferred based only on sequence comparisons. The gene names are as described in the text.

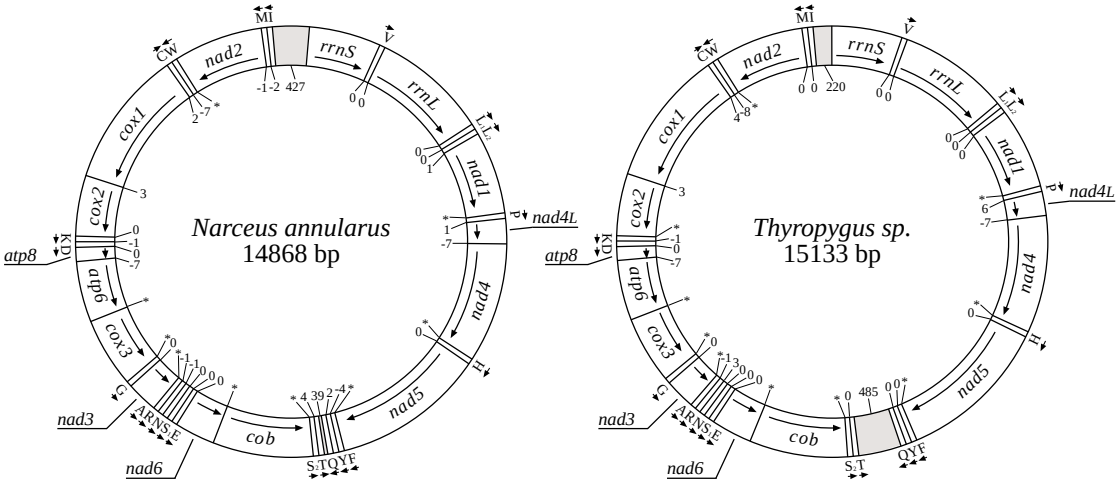
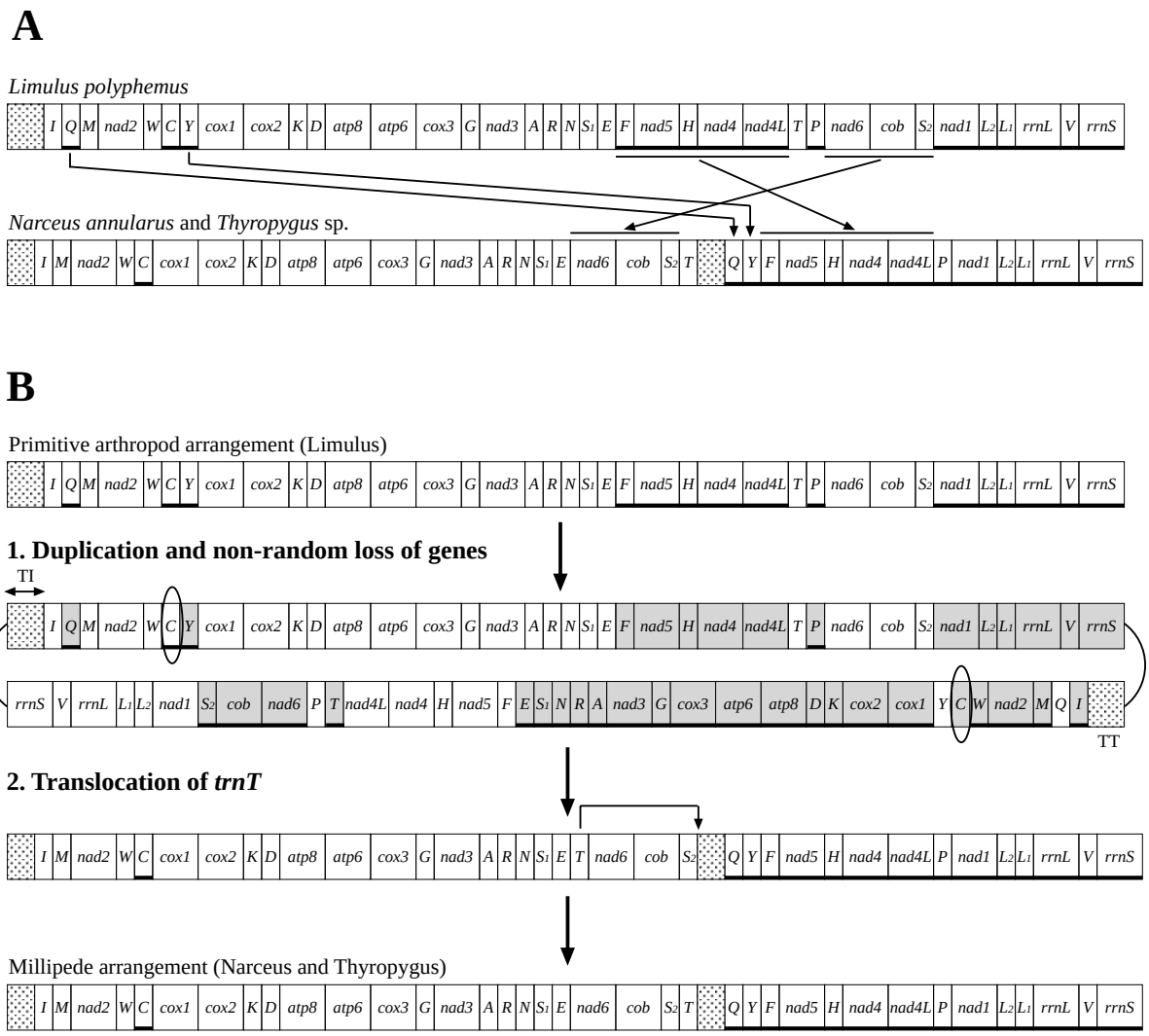
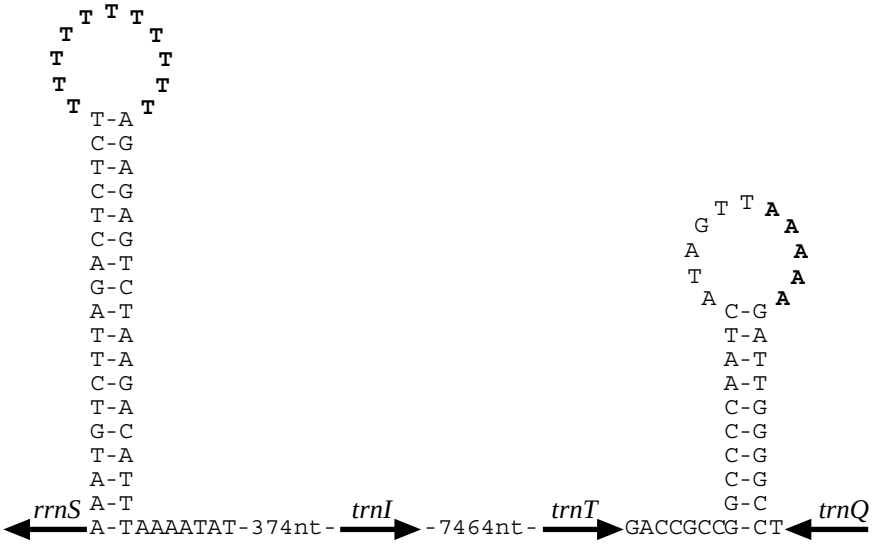
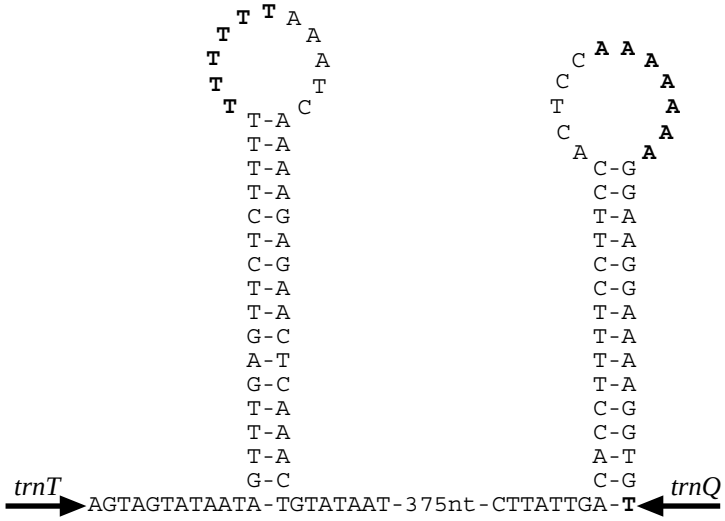

Narceus annularus
14868 bp
Thyropygus sp.
15133 bp
nad2
rrnS
rrnL
nad1
nad4
nad5
cob
cox3
atp6
cox2
cox1
nad3
nad6
atp8
nad4L
M
V
L1
P
H
S2
Q
K
G
A
R
N
S1
E
F
C
W
Y

Figure 1


A
Limulus polyphemus
I Q M nad2 W C Y cox1 cox2 K D atp8 atp6 cox3 G nad3 A R N S1 E F nad5 H nad4 nad4L T P nad6 cob S2 nad1 L2 L1 rrnL V rrnS
Narceus annularus and Thyropygus sp.
I M nad2 W C cox1 cox2 K D atp8 atp6 cox3 G nad3 A R N S1 E nad6 cob S2 T Q Y F nad5 H nad4 nad4L P nad1 L2 L1 rrnL V rrnS
B
Primitive arthropod arrangement (Limulus)
I Q M nad2 W C Y cox1 cox2 K D atp8 atp6 cox3 G nad3 A R N S1 E F nad5 H nad4 nad4L T P nad6 cob S2 nad1 L2 L1 rrnL V rrnS
1. Duplication and non-random loss of genes
TI
I Q M nad2 W C Y cox1 cox2 K D atp8 atp6 cox3 G nad3 A R N S1 E F nad5 H nad4 nad4L T P nad6 cob S2 nad1 L2 L1 rrnL V rrnS
rrnS V rrnL L1 L2 nad1 S2 cob nad6 P T nad4L nad4 H nad5 F E S1 N R A nad3 G cox3 atp6 atp8 D K cox2 cox1 Y C W nad2 M Q I
TT
2. Translocation of trnT
I M nad2 W C cox1 cox2 K D atp8 atp6 cox3 G nad3 A R N S1 E T nad6 cob S2 Q Y F nad5 H nad4 nad4L P nad1 L2 L1 rrnL V rrnS
Millipede arrangement (Narceus and Thyropygus)
I M nad2 W C cox1 cox2 K D atp8 atp6 cox3 G nad3 A R N S1 E nad6 cob S2 T Q Y F nad5 H nad4 nad4L P nad1 L2 L1 rrnL V rrnS



N. annularus



Thyropygus sp.

Figure 3