Genetics and Molecular Biology, 23, 4, 745-752 (2000)

Review Article

Mitochondrial genome organization and vertebrate phylogenetics

Sérgio Luiz Pereira

Abstract

With the advent of DNA sequencing techniques the organization of the vertebrate mitochondrial genome shows variation between higher taxonomic levels. The most conserved gene order is found in placental mammals, turtles, fishes, some lizards and *Xenopus*. Birds, other species of lizards, crocodilians, marsupial mammals, snakes, tuatara, lamprey, and some other amphibians and one species of fish have gene orders that are less conserved. The most probable mechanism for new gene rearrangements seems to be tandem duplication and multiple deletion events, always associated with tRNA sequences. Some new rearrangements seem to be typical of monophyletic groups and the use of data from these groups may be useful for answering phylogenetic questions involving vertebrate higher taxonomic levels. Other features such as the secondary structure of tRNA, and the start and stop codons of protein-coding genes may also be useful in comparisons of vertebrate mitochondrial genomes.

CHARACTERIZATION OF THE "CONSERVED VERTEBRATE MITOCHONDRIAL GENOME"

With the development of DNA sequencing methods and the extensive sequencing experiments undertaken in the last two decades in a wide variety of organisms, the order of the genes in the mitochondrial DNA molecule has begun to be disclosed. A great number of phylogenetic studies using mitochondrial gene sequences have been done and reported, some dealing with the use of mitochondrial genes in the establishment of different levels of phylogenetic relationships (Kumazawa and Nishida, 1993; Zardoya and Meyer, 1996). For example, the control region of the mitochondrial genome is frequently used in population studies due to the high variability in its nucleotide sequence, while protein-coding genes, such as cytochrome b (Cyt *b*), are generally used for phylogenetic analysis of taxa above the species level.

The success of mtDNA sequences in phylogenetic studies is due to several characteristics: a) compact gene packing, with little noncoding intergenic nucleotides and some nucleotide overlapping between genes encoded in opposite strands (Cantatore and Saccone, 1987; and other genomes completely sequenced by several researchers); b) lack of recombination (Clayton, 1982, 1992; Hayashi *et al*., 1985); c) mainly maternal inheritance (Kondo *et al*., 1990; Gyllestein *et al*., 1991); d) faster sequence evolution as compared to nuclear sequences, perhaps due to repair inefficiency (Brown *et al*., 1979), and e) multicopy status in a cell (Michaels *et al*., 1982; Robin and Wong, 1988).

The first complete sequencing studies of the whole mtDNA of animals showed that the majority of the invertebrate mitochondrial genome is made up of nearly the same number of genes as in vertebrates. Nevertheless, several rearrangements have been found in nematodes (Okimoto *et al*., 1991, 1992), arthropods (Boore *et al*., 1995), bivalves (Hoffman *et al*., 1992), pulmonate mollusks (Yamazaki *et al*., 1997), echinoderms (Cantatore *et al*., 1987, 1989; Jacobs *et al*., 1988; De Giorgi *et al*., 1991; Smith *et al*., 1989, 1990, 1993), and *Drosophila yakuba* (Clary and Wolstenholme, 1985).

This variation in gene order among invertebrates allowed Sankoff *et al*. (1992) to use the mitochondrial genome data in a phylogenetic approach. They compared the gene order of 16 taxa including fungi and several other eukaryotes, and obtained a tree highly compatible with that known for the relationships between fungi and metazoans. This kind of approach was recently used to establish the phylogeny of the three major groups of arthropods (atelocerates, crustaceans and chelicerates) and other higher taxa (Boore *et al*., 1995), including echinoderms (Smith *et al*., 1993) and gastropods (Yamazaki *et al*., 1997).

Until recently this approach was not used in vertebrate phylogeny due to the belief that there was a "conserved gene order" for the vertebrate mitochondrial genome. The reason for this is that the first complete mitochondrial genome sequences found in vertebrate taxa had no variation in the position of the genes along the molecule. This has been seen in taxa as diverse as *Xenopus laevis* (Roe *et al*., 1985) humans and other mammals (Bibb *et al*., 1981; Anderson *et al*., 1981,1982; Brown *et al*., 1982; Gadaleta *et al*., 1989; Árnason and Johnsson, 1992; Árnason and Gullberg, 1993), and some fish species

Departamento de Biologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, 277, sala 215, 05508-900 São Paulo, SP, Brasil. Fax: +55-11-818-7419. E-mail: sergiolp@ib.usp.br

(Johansen *et al*., 1990; Tzeng *et al*., 1992; Chang *et al*., 1994; Zardoya *et al*., 1995; Zardoya and Meyer, 1996). This "conserved gene order" has also been found in even more primitive vertebrates, including the dogfish (*Scyliorhinus caniculas*), a shark (*Mustela manazo*), the "living fossil" coelacanth (*Latimeria chalumnae*), and the bichir (*Polypterus ornatipinnis*) (Noack *et al*., 1996; Zardoya and Meyer, 1997; Delarbre *et al*., 1998; Cao *et al*., 1998). Thus, the typical vertebrate mitochondrial genome was assumed to have a noncoding control region, 13 protein-coding genes, two ribosomal RNAs (rRNA), and 22 transfer RNAs (tRNA) spread along the circular DNA molecule (Anderson *et al*., 1981; Bibb *et al*., 1981; Roe *et al*., 1985; Gadaleta *et al*., 1989; Johansen *et al*., 1990) (Figure 1).

The noncoding control region is responsible for size variation in the mitochondrial genomes due to differences in length and copy number of short repeats in various groups (e.g., Brown *et al*., 1996; Gissi *et al*., 1998; Härlid *et al*., 1998; Ursing and Arnason, 1998). Three clusters of tRNAs can be identified and they are named according to the one-letter amino acid code: IQM representing the tRNAs, respectively, for isoleucine (I), glutamate (Q) and methionine (M); WANCY for tryptophan (W), alanine (A), aspartic acid (N), cysteine (C) and tyrosine (Y); and, HSL, for histidine (H), serine (S) and leucine (L). Each mtDNA strand has its own replication origin and they form stable stem-and-loop structures. The replication origin of the H-strand (O_H) is located at the displacement loop region (D-loop) in the control region and that for the L-strand (O_L) is in the WANCY cluster between the tRNAs for asparagine and cysteine. The O_L has a 3'-GGCCG-5'-conserved sequence at the 5' base of the stem. This sequence seems to be necessary for the replication of the light strand *in vitro* (Hixson *et al*., 1986).

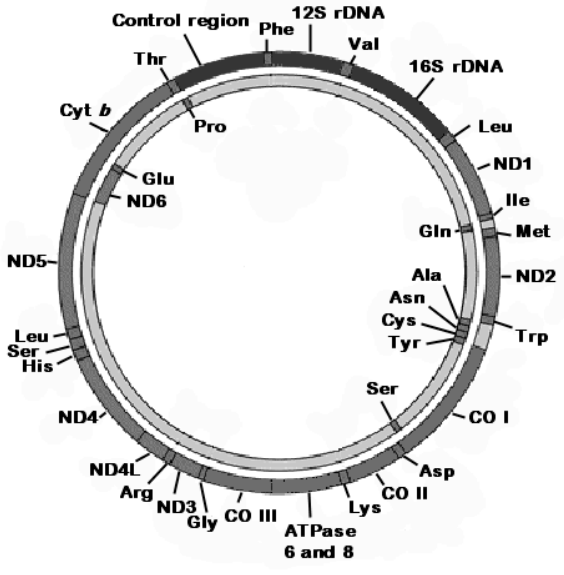

Figure 1 - Schematic representation of the circular molecule of the "conserved" vertebrate mitochondrial genome. Genes outside and inside the circle are transcribed in the H and L strands, respectively. Protein-coding genes are represented as follows: Cyt *b* - cytochrome *b*; CO I, CO II and CO III - subunits I, II and III of the cytochrome oxidase; ND1-6 - subunits 1 to 6 of the NADH reductase. tRNA are represented by their three-letter amino acid abbreviations.

The protein-coding genes are punctuated by at least one tRNA at their 5' or 3' end, and this may represent a recognition signal for RNA processing (Ojala *et al*., 1981). This is of extreme molecular importance in the mitochondrial DNA molecule as some genes overlap by a few nucleotides (nt) and some overlapping genes are transcribed in opposite strands. Except for the tRNAs of glutamic acid, alanine, aspartic acid, cysteine, tyrosine, serine (UCN), glutamate, proline, and the ND6 gene, all other sequences are transcribed from the H-strand.

VARIATION IN GENOME ORGANIZATION

What was once the "universal" gene order for living vertebrates was discarded as more data were obtained from several other vertebrates (Figure 2). When the mtDNA of birds were analyzed (Desjardins and Morais, 1990, 1991; Quinn and Wilson, 1993; Ramirez *et al*., 1993) a different organization of the genes near the control region, involving protein-coding genes, was found in this lineage. The sequence control region - $tRNA^{Pro}$ - $tRNA^{Thr}$ - cytochrome b - $tRNA^{Glu}$ - ND6 - ND5 became control region - $tRNA^{Glu}$ - ND6 - $tRNA^{Pro}$ - $tRNA^{Thr}$ - cytochrome b - ND5 in avian mitochondrial DNA. This event dates from the bird-crocodilian splitting, 254 million years ago (mya) (Janke and Árnason, 1997) and is common to Paleognathae (Struthioniformes and Tinamiformes; Harlid *et al*., 1997, 1998) and Neognathae birds (all other bird groups; Desjardins and Morais, 1990, 1991; Quinn and Wilson, 1993; Ramirez *et al*., 1993). Desjardins and Morais (1990) suggest that this type of avian mitochondrial genome organization confers a selective advantage in the propagation and expression of avian mtDNA genes, because the acquisition of flight demands more energy, although this correlation has not yet been demonstrated.

Mindell *et al*. (1998) found another avian rearrangement, apparently derived from the first avian rearrangement described in *Gallus*. The new gene order found in Falconiformes, some Cuculiformes, some Piciformes and the suboscines Passeriformes is 12S - $tRNA^{Phe}$ - noncoding sequence - $tRNA^{Glu}$ - ND6 - $tRNA^{Pro}$ - control region - $tRNA^{Thr}$ - cytochrome b - ND5. This new rearrangement is supposed to have occurred in parallel, since the four Avian orders where it is found do not comprise a monophyletic group. Other regions of the mtDNA in these birds were not checked for other rearrangements.

In crocodilians the "conserved" HSL cluster is arranged in the order SHL, and there are 1 to 29 spacer nucleotides between them. In birds, lizards, mammals and *Xenopus* these bases are absent. In one of the crocodilians (*Cai-*

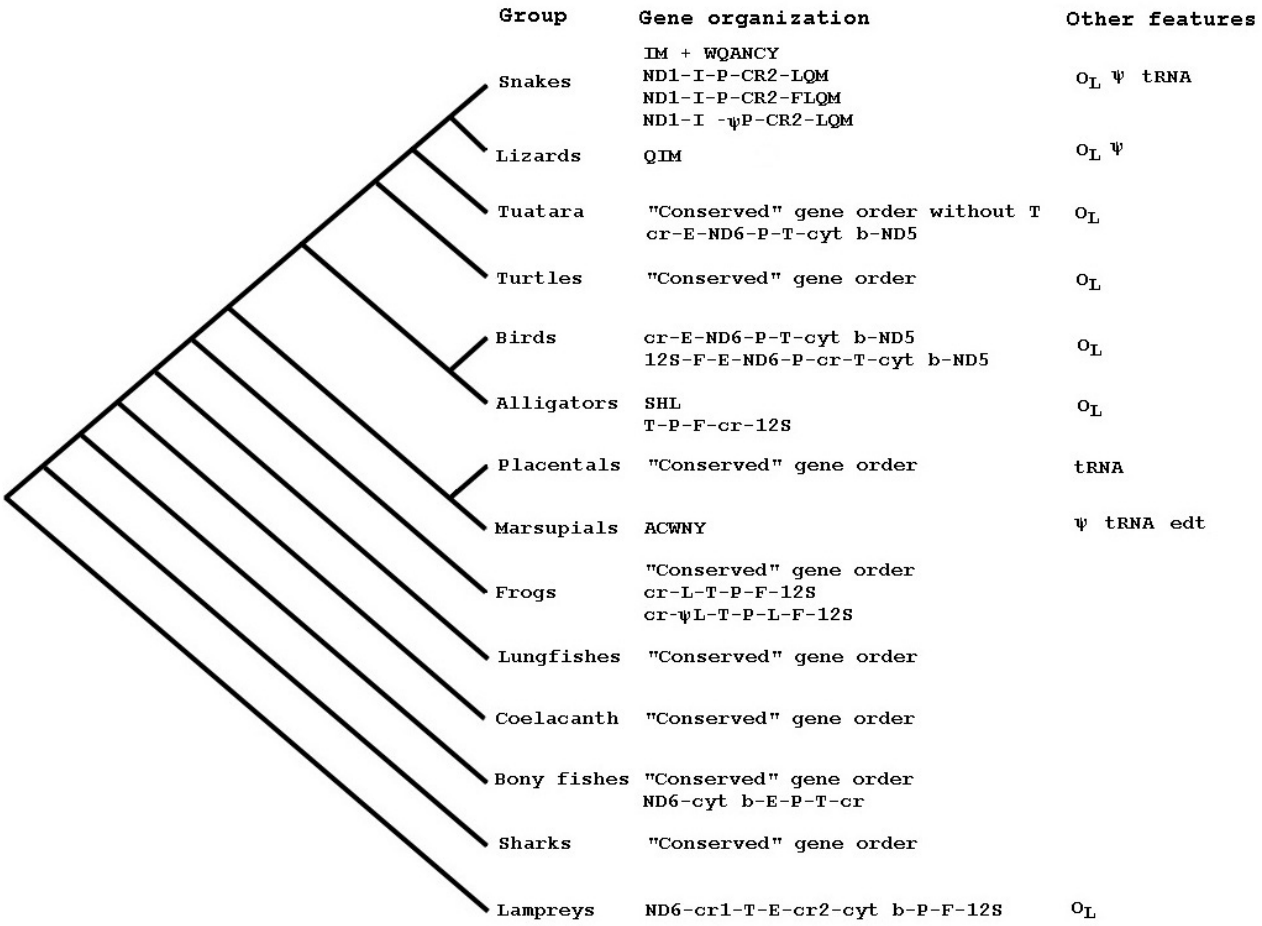

Figure 2 - A phylogenetic hypothesis among vertebrates. Gene rearrangements and other features of the vertebrate mitochondrial DNA found in each group are summarized. Abbreviations: cr - control region; 12S - 12S rDNA; ψ - tRNA pseudogene; other genes as in the legend to Figure 1, except tRNAs, represented by their one-letter code; O_L - loss of the origin of the light-strand replication; tRNA - unusual tRNA secondary structures; edt - tRNA edition to restore original function. Not all snakes lose the O_L in the WANCY cluster (see text).

man), there is a noncoding region of 264 nt between the rearranged SHL cluster and the upstream gene ND4, but only a 23-31-nt long region in the other crocodilians analyzed (Janke and Árnason, 1997; Kumazawa and Nishida, 1995). Another unique aspect of crocodilian mtDNA was found in *Alligator mississippiensis* (Janke and Árnason, 1997) and *Crocodylus porosus* (Quinn and Mindell, 1996). The $tRNA^{Phe}$ between the control region and the 12SrDNA region "moved" to the 5' end of the control region, forming a new cluster with the tRNAs for threonine and proline (Janke and Árnason, 1997).

Marsupials have a different arrangement in the WANCY cluster. The new order is ACWNY, with the O_L between the $tRNA^{Trp}$ and $tRNA^{Asn}$ region. Among marsupials the O_L hairpin structure is flanked by long nonconserved sequences, with a 3'-GGGCCC-5' sequence similar to the placental mammalian 3'-GGCCG-5' present at the 5' base of the stem. With this rearrangement, $tRNA^{Ala}$ overlaps three bases with the 3' end of the ND2 gene. A striking feature is the presence of 5-14 nt between the tRNAs of alanine, cysteine and tryptophan. In birds, fishes and mammals there are between 0-7 nt separating these tRNAs (Pääbo *et al*., 1991; Janke *et al*., 1994).

In amphibians, the gene order in *Xenopus* was the same as the "vertebrate conserved" order (Roe *et al*., 1985), but in two different species of the genus *Rana* two novel gene orders have been found in this group; both involved the transposition of tRNAs between the control region and the $tRNA^{Phe}$ region. The bull frog (*Rana catesbeiana*) has the gene order control region - $tRNA^{Leu(CUN)}$ - $tRNA^{Thr}$ - $tRNA^{Pro}$ - $tRNA^{Phe}$ - 12SrDNA (Yoneyama, 1987), and the rice frog (*Rana limnocharis*) has the order control region - $tRNA^{Thr}$ - $tRNA^{Pro}$ - $tRNA^{Leu(CUN)}$ - $tRNA^{Phe}$ - 12SrDNA (Macey *et al*., 1997). Interestingly there are 15 and 37 noncoding nucleotides flanking the leucine tRNA in the rice frog. This species also has a sequence similar to the $tRNA^{Leu(CUN)}$ sequence between the control region and the $tRNA^{Thr}$ sequence that has been considered to be a pseudogene. These authors did not investigate the 5' region of the control region and the HSL cluster from where these transposed tRNA genes

probably came from to determine how these regions are now arranged. The amphibian O_L region was found in the canonical position of the "consensus order".

Most of the lizards analyzed had the same "vertebrate consensus" gene order (e.g. Kumazawa and Nishida, 1999), but in five groups (Agamidae, Amphibaenia, Scincidae, Xantusidae (Macey *et al*., 1997) and Rhyncocephalia (Seutin *et al*., 1994)) there was loss of the O_L region. Representatives of the acrodonts, Agamidae and Chamaelonidae, have the gene order of the IQM cluster changed to QIM (Macey *et al.*, 1997). Before they diverged from other vertebrates 310 mya (Benton, 1990), birds and crocodilians also lost the origin of light-strand replication, O_L, in the WANCY region (Desjardins and Morais, 1990, 1991; Ramirez *et al*., 1993; Seutin *et al*., 1994; Kumazawa and Nishida, 1995; Janke and Árnason, 1997).

Loss of the O_L region has also happened on other two occasions: 550 mya in the lampreys (Lee and Kocher, 1995; Delarbre *et al*., 2000) and, more recently, after the blind snake diverged from other snakes 140 mya (Kumazawa and Nishida, 1995; Kumazawa *et al*., 1996). A singular feature of the blind snake is the migration of the functional $tRNA^{Gln}$ from the IQM cluster to the WANCY cluster, between the $tRNA^{Trp}$ and $tRNA^{Ala}$ regions. This tRNA is located 1.2 kb away from its original position. Migration of mtDNA sequences to distant positions is known to occur in lizards (Stanton *et al*., 1994). Macey *et al*. (1997) concluded that the movement of the tRNA of the WANCY region and the loss of the O_L region are not independent phenomena. The larger spacer regions found between "transposed" genes is assumed to be a relict of tRNAs that were deleted when the rearrangement occurred (Janke *et al*., 1994).

Kumazawa *et al*. (1996, 1998) reported new mtDNA rearrangements in snakes of the families Viperidae, Colubridae and Boidae. The species analyzed show two control regions. One of the control regions is located in the "canonical" position. The other is inserted in the IQM cluster together with some tRNAs.

Both control regions found in the snakes probably were present 70 mya in the common ancestor of these snakes. It seems unlikely that the second control region would be inserted in the same distant position in five different species of three families. Surprisingly there is more similarity within copies than between different taxa. For such a sequence high divergence between copies would be expected. This is also a good example of concerted evolution. Kumazawa *et al.* (1998) hypothesized a mechanism for such uniformity between copies in the colubrid akamata, invoking tandem duplication or gene conversion.

All snakes analyzed by Kumazawa and co-workers had a new tRNA cluster, the LQM cluster. The viper himehabu also has a $tRNA^{Phe}$ and a spacer between the second control region and the LQM cluster. All species, except the boid python, has a $tRNA^{Pro}$ 3' end of the second control region. This tRNA seems to be a pseudogene in the akamata. Its functional $tRNA^{Pro}$ is in the "conserved" position while himehabu has a functional $tRNA^{Pro}$ at the 3' end of the second control region and a $tRNA^{Pro}$ pseudogene at the 3' end of the first control region.

So the basic rearrangement of the genes in snake mitochondrial genome is 16SrRNA - ND1 - $tRNA^{Iso}$ - $tRNA^{Pro}$ - control region - LQM cluster - ND2. The akamata and himehabu WANCY region has been sequenced, and it has been shown to have the typical organization for this region, including an O_L region.

The mitochondrial genome of turtles (Kumazawa and Nishida, 1999) has the same gene organization as placental mammals, fishes and *Xenopus*, and the coding sequences have similar lengths to the coding sequences of placental homologs.

Heteroplasmy for two gene orders was found in the tuatara *Sphenodon punctatus* (Quinn and Mindell, 1996). One of the gene orders is similar to consensus gene organization without the $tRNA^{Thr}$, while the other is similar to the first gene order found in birds (Desjardins and Moraes, 1990).

More recently, Miya and Nishida (1999) reported the first example of tRNA gene rearrangement in bony fishes between the ND6 and the control region (ND6 - cyt *b* - $tRNA^{Glu}$ - $tRNA^{Pro}$ - $tRNA^{Thr}$ - control region), changing the "conserved" status of fishes.

OTHER ASPECTS OF THE MITOCHONDRIAL GENOME

Some tRNAs of vertebrates have unusual features. The tRNA for serine with the anticodon AGY lacks a DHU arm, a common finding in all vertebrates. Placental and marsupial mammals share a $tRNA^{Lys}$ region with a reduced D arm and a $tRNA^{Ser(UCN)}$ region with six instead of five nucleotides at the anticodon stem. Birds, reptiles, ray- and lobe-fishes and amphibians have normal cloverleaf structures for these two tRNAs. This unusual tRNA structure probably appeared after the Mammalia diverged from other vertebrate groups, about 300 mya (Benton, 1990). Surprisingly, Monotremata, the sister-group to marsupials according to mtDNA sequences (Janke *et al*. 1997), does not have these unusual tRNA structures (Janke *et al*., 1996).

Snakes seem to have very unusual tRNA structures. The T arm of akamata and himehabu has less than five base pairs in the T stem. It seems that snakes may have evolved a simpler T arm structure (as in nematodes; Wolstenholme *et al*., 1987; Watanabe *et al*., 1994), associated with the loss of tertiary interaction between the T and D loops (Kumazawa *et al*., 1996, 1998).

There are, generally, up to three nt separating mtDNA genes. All three monotremes studied have this separation between the genes, but there are an additional 88 base pairs separating the $tRNA^{Leu(UUR)}$ region and the ND1 gene. This sequence has no similarity with any other known gene, and is apparently not the product of duplication. No stable secondary structure could be assigned, but a stable transcript whose function remains unknown has been detected by Northern blotting (Janke *et al*., 1996).

In marsupials there seems to be editing of tRNA af-

ter transcription. The $tRNA^{Asp}$ region has the anticodon GCC instead of the normal GTC. The second position of the codon is modified to restore the original function of the tRNA, by cytosine deamination (Janke and Pääbo, 1993; Janke *et al*., 1994, 1997). Monotremes, the sister taxon of Marsupialia, also have the normal GTC anticodon, as do other vertebrates. Thus, the unusual GCC anticodon seems to have appeared 116 mya in the lineage leading to the marsupials (Janke *et al*., 1994, 1997).

The only deletion of genes as yet found in a vertebrate genome was reported by Janke *et al*. (1994, 1997) in one marsupial. The typical tRNA of lysine is missing. In the corresponding region of the $tRNA^{Lys}$, there is a region showing a sequence lacking the anticodon for the amino acid lysine. RNA edition is known to occur in marsupials (Janke *et al*., 1994), but this does not seem to be the case here. Multiple sites are usually edited to restore the original function of $tRNA^{Lys}$. Janke *et al*. (1997) evoked importation of $tRNA^{Lys}$ from the cytoplasm to explain the functional translation system in marsupial mitochondria.

In Amphisbaena addition of functional genes was not found, but some tRNA sequences were thought to be pseudogenes (Macey *et al*., 1998). This has also been reported in snakes (Kumazawa *et al*., 1996, 1998).

EARLY VERTEBRATE GENE ORDER

The lamprey is known to be one of the closest relatives of tetrapods and its time of divergence is assumed to be 550 mya (Lee and Kocher, 1995). Its mitochondrial gene organization is quite different from the "conserved" gene order (Lee and Kocher, 1995; Delarbre *et al*., 2000). Two major noncoding control region sequences were found. One of them, albeit having only 491 bp, seems to be the canonical control region. It possesses an A+T rich region of 28 bp at the 5' end. The conserved sequence blocks CSB-II and CSB-III, and other vertebrate control regions. The second major noncoding control region is separated from the first by glutamate and threonine tRNAs, and is basically formed by seven copies of an A+T rich 27 bp repeat.

In lampreys the gene order around the control region is ND6 - the first control region - $tRNA^{Thr}$ - $tRNA^{Glu}$ - the second control region - cyt *b* - $tRNA^{Pro}$ - $tRNA^{Phe}$ - 12Sr DNA. Only lampreys and birds have a novel organization involving modification of the position of protein-coding genes.

Although it can be said that the "consensus" vertebrate organization was established early in vertebrate evolutionary history (Lee and Kocher, 1995), this new gene organization makes the ancestral state unknown. Has the lamprey the ancient gene organization and the "vertebrate consensus" the apomorphic condition? Or does the lamprey arrangement represent a unique feature of this group?

MECHANISM FOR THE REARRANGEMENTS

All the rearrangements of mitochondrial genes involve tRNA genes. It seems that regions like tRNA can "move" along the molecule due to their capacity to form stable stem-and-loop structures (Stanton *et al*., 1994). The absence of a typical O_L region is not so dramatic. An alternative stem-and-loop site, such as that of tRNAs, may be used as the initiation point for the replication of the L-strand (Clary and Wolstenholme, 1985).

Several mechanisms, such as inversion, transposition, intramolecular recombination, tandem duplication and deletion were proposed to explain the rearrangements. Tandem duplication and multiple deletion mechanisms might explain the origin of the majority of the rearrangements found so far. However, some of the observed rearrangements cannot be explained by the above mechanisms, and other unknown mechanisms must be acting (Stanton *et al*., 1994).

Complementary base pairing and the formation of tRNA stem-and-loop structures may be recognized by the enzyme that initiates L-strand synthesis, in the absence of a typical O_L region. As replication is asymmetrical when the new H-strand synthesis is completed the nascent light-strand detaches and slips ahead to another stem-and-loop site. When the L-strand in synthesis is completed, the termination point is ahead of the initiation point. So a tandem duplication occurs. A new site acting as an O_L region is then formed, preventing other duplication events. This new site is probably produced because of structural affinities between replication proteins and stem-and-loop structure. In order for the rearrangement to occur at least two deletion events must happen after the in tandem duplication event, each one in a different copy of the tandem duplication, thus eliminating the repeated genes (Pääbo *et al*., 1991; Kumazawa and Nishida, 1995; Macey *et al*., 1997).

The reading frame of the genes has not changed and terminal repeats, typical of transposable elements, have not been found in the majority of the rearrangements. Because of this the hypothesis of transposition may be rejected, although integration of many different genetic elements occurs in the tRNAs of bacteria (Reiter *et al*., 1989), slime molds (Marschalek *et al*., 1989) and yeasts (Chalker and Sandmeyer, 1990). In *Rana limnocharis,* transposition may explain the new organization, since terminal direct repeats were found (Macey *et al*., 1997).

Intramolecular recombination could generate new rearrangements (Thyagarajan *et al*., 1996; Lunt and Hyman, 1997) but it seems to be very rare in the mitochondrial genome. Inversions do not explain the novel gene orders since none of the rearrangements caused inversion of the coding polarity of the genes, except in echinoderms (Smith *et al*., 1993) in which the 16SrDNA gene has its polarity inverted.

GENE ORGANIZATION AND PHYLOGENY OF DEEPLY BRANCH GROUPS

Gene order may be used to study the constraints on the evolution of the mitochondrial genome, the mechanism

responsible for new gene rearrangements and deep-branch phylogeny. Because of the unlikely possibility of rearrangement convergence and the certainty of mtDNA gene homology Boore and Brown (1998) highlighted the suitability of gene order as applied to phylogeny.

The tRNA clusters should be the focus of more attention since they are always involved in new rearrangements and seem to be involved in divergences occurred more than 100 million years old (Kumazawa and Nishida, 1993, 1995; Smith *et al*., 1993). Mitochondrial DNA sequences are generally not used in deep branch phylogeny because of the saturation of changes and difficulties in the alignment of the loops of the rRNA sequences. Gene rearrangements may overcome these problems (Quinn, 1997; Boore and Brown, 1998), since they are rare and no alignment hypothesis is need. However, one has to be careful because greater divergence times may also hide back mutations at the genome level (Sankoff *et al*., 1992) and some of the features of the mitochondrial genome represent parallel events. Indeed, the loss of the O_L region occurred independently in several vertebrate lineages (Desjardins and Morais, 1990, 1991; Ramirez *et al.*, 1993; Seutin *et al.*, 1994; Kumazawa and Nishida, 1995; Janke and Arnason, 1997; Macey *et al*., 1997) as did the second gene order found in non-monophyletic bird groups (Mindell *et al*., 1998).

The noncoding spacer nucleotides associated with regions where rearrangements occur seem to be a good indicator of the time of the transposition event. Mitochondrial DNA is an economical molecule and it tends not to accumulate noncoding sequences, except the control region and the O_L region because of their role in the duplication of mtDNA and the transfer and ribosomal RNAs necessary for translation.

The number of possible new mitochondrial genome organizations is great. Few of the possible rearrangements are found in vertebrates. This is not indicative that other gene organization has not occurred but that the organization as seen today may have functional constraints. Stable structural gene rearrangement may be due to constraints imposed on gene movement, selection on potential differences in transcription of genes, and low probability of transposition to specific gene junctions (Smith *et al.*, 1993). Due to their rarity, it is not plausible that different groups will have the same rearrangement (Boore *et al*., 1995). For example, birds and lampreys have a different specific gene organization to that which is found in other groups.

When a rearrangement occurs, the evolutionary constraints in the regions involved in the new organization may also change (Kumazawa and Nishida, 1993). This could be exemplified by the evolution of the $tRNA^{Glu}$ region. In birds this gene is not associated with an O_L region and its sequence is more variable when compared to groups such as mammals where it is associated with the O_L (Quinn and Wilson, 1993). Moreover, some regions could be under selective constraints, keeping their organization unchanged, such as the $tRNA^{Tyr}$ of marsupials. Pääbo *et al*. (1991) reported that this gene is more highly conserved among marsupials than among placental mammals thought to be as old as marsupials. Heavy-strand transcript processing was invoked to explain the constraint in marsupials (Pääbo *et al*., 1991).

Quinn (1997) proposed that with the increasing amount of mitochondrial and nuclear sequence data available new outcomes for phylogenetic studies at several levels will emerge. Data such as codon usage, start and termination codons, and overlapping genes may be helpful in the future as the knowledge of molecular evolution increases. Even the secondary structure of the control region and ribosomal genes should be checked out for use in phylogeny as soon as the understanding of these sequences and their role in the mitochondria genome improves.

ACKNOWLEDGMENTS

I am grateful to Cintia Fridman, Anita Wajntal, Flora M.C. Fernandes-Matioli and two anonymous reviewers for helpful suggestions. Publication supported by FAPESP.

RESUMO

Com o advento das técnicas de seqüenciamento de DNA, a organização do genoma mitocondrial de vertebrados mostrou variação entre níveis taxonômicos superiores. A organização "conservada" é encontrada em mamíferos placentários, tartarugas, peixes, alguns lagartos e *Xenopus*. As aves, outras espécies de lagartos, crocodilianos, mamíferos marsupiais, cobras, tuatara, lampréia, uma espécie de peixe e alguns outros anfíbios apresentam ordens que diferem do que foi inicialmente considerado a "ordem gênica conservada". O mecanismo mais provável de novos rearranjos gênicos parece ser duplicação *in tandem* e múltiplos eventos de deleção, sempre associados com seqüências de tRNAs. Alguns novos rearranjos parecem ser típicos de grupos monofiléticos e o uso destes dados pode ser útil em questões de filogenia envolvendo níveis taxonômicos superiores. Outros aspectos como a estrutura secundária de tRNA e códons de início e de parada de transcrição de genes codificantes para proteínas podem ser úteis para comparações entre os genomas mitocondriais de vertebrados.

REFERENCES

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijin, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Standen, R. and **Young, I.G.** (1981). Sequence and organization of the human mitochondrial genome. *Nature 290*: 457-465.

Anderson, S., de Bruijin, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and **Young, I.G.** (1982). Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *J. Mol. Biol. 156*: 683-717.

Árnason, U. and **Gullberg, A.** (1993). Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. *J. Mol. Evol. 37*: 312-322.

Árnason, U. and **Johnsson, E.** (1992). The complete mitochondrial DNA sequences of the harbor seal, *Phoca vitulina. J. Mol. Evol. 34*: 493-505.

Benton, M. (1990). Phylogeny of the major tetrapod groups: morphologi-

cal data and divergence dates. *J. Mol. Evol. 30*: 409-424.

Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and **Calyton, D.A.** (1981). Sequence and organization of mouse mitochondrial DNA. *Cell 26*: 167-180.

Boore, J.L. and **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., Collins, T.M., Stanton, D., Daehler, L.L. and **Brown, W.M.** (1995). Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature 376*: 163-165.

Brown, J.R., Beckenback, K., Beckenback, A.T. and **Smith, M.J.** (1996). Length variation, heteroplasmy, and sequence divergence in the mitochondrial DNA of four species of sturgeon (*Acipenser*). *Genetics 142*: 525-535.

Brown, W.M., George, M. and **Wilson, A.C.** (1979). Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci USA 76*: 1967-1971.

Brown, W.M., Prager, E.M., Wang, A. and **Wilson, A.C.** (1982). Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol. 18*: 225-239.

Cantatore, P. and **Saccone, C.** (1987). Organization, structure, and evolution of mammalian mitochondrial genes. *Int. Rev. Cytol. 108*: 149-208.

Cantatore, P., Roberti, M., Morisco, P., Rainaldi, G., Gadaleta, M. and **Saccone, C.** (1987). A novel gene order in the *Paracentrotus lividus* mitochondrial genome. *Gene 53*: 41-54.

Cantatore, P., Roberti, M., Rainaldi, G., Gadaleta, M. and **Saconne, C.** (1989). The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. *J. Biol. Chem. 264*: 10965-10975.

Cao, Y., Waddell, P.J., Okada, N. and **Hasegawa, M.** (1998). The complete mitochondrial DNA sequence of the shark *Mustelus manazo*: evaluating rooting contradictions to living bony vertebrates. *Mol. Biol. Evol. 15*: 1637-1646.

Chalker, D.L. and **Sandmeyer, S.B.** (1990). Transfer RNA genes are genomic targets for *de novo* transposition of the yeast retrotransposons Ty*3*. *Genetics 126*: 837-845.

Chang, Y.-S., Huang, F.-C. and **Lo, T.-B.** (1994). The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. *J. Mol. Evol. 38*: 138-155.

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

Clayton, D.A. (1982). Replication of animal mitochondrial DNA. *Cell 28*: 693-705.

Clayton, D.A. (1992). Structure and function of the mitochondrial genome. *J. Inherited Metab. Dis. 15*: 439-447.

De Giorgi, C., Lanave, C., Musci, D. and **Saconne, C.** (1991). Mitochondrial DNA in the sea urchin *Arbacia lixula*: evolutionary inferences from nucleotide sequence analysis. *Mol. Biol. Evol. 8*: 515-529.

Delarbre, C., Spruyt, N., Delmare, C., Gallut, C., Barriel, V., Janvier, P., Laudent, V. and **Gachelin, G.** (1998). The complete nucleotide sequence of the mitochondrial DNA of the dogfish, *Scyliorhinus canicula*. *Genetics 150*: 331-344.

Delarbre, C., Escriva, H., Gallut, C., Barriel, V., Kourilsky, P., Janvier, P., Laudet, V. and **Gachelin, G.** (2000). The complete nucleotide sequence of the mitochondrial DNA of the agnathan *Lampreta fluviatilis*: bearings on the phylogeny of cyclostomes. *Mol. Biol. Evol. 17*: 519-529.

Desjardins, P. and **Morais, R.** (1990). Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. *J. Mol. Biol. 212*: 599-634.

Desjardins, P. and **Morais, R.** (1991). Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. *J. Mol. Evol. 32*: 153-161.

Gadaleta, G., Pepe, D., De Candia, G., Quagliariello, C., Sbisà, E. and **Saccone, C.** (1989). The complete nucleotide sequence of the *Rattus novergicus* mitochondrial genome: cryptical signals revealed by comparative analysis between vertebrates. *J. Mol. Evol. 28*: 497-516.

Gissi, C., Gullberg, A. and **Ulfur, A.** (1998). The complete mitochondrial DNA sequence of the rabbit, *Oryctolagus cuniculs*. *Genomics 50*: 161-169.

Gyllestein, U., Wharton, D., Josefsson, A. and **Wilson, A.C.** (1991). Paternal inheritance of mitochondrial DNA in mice. *Nature 352*: 255-257.

Härlid, A., Janke, A. and **Árnason, U.** (1997). The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. *Mol. Biol. Evol. 14*: 754-761.

Härlid, A., Janke, A. and **Árnason, U.** (1998). The complete mitochondrial genome of *Rhea americana* and early avian divergences. *J. Mol. Evol. 46*: 669-679.

Hayashi, J.I., Tagashira, Y. and **Yoshida, M.C.** (1985). Absence of extensive recombination between interspecies and intraspecies mitochodrial DNA in mammalian cells. *Exp. Cell Res. 160*: 387-395.

Hixson, J.E., Wong, T.W. and **Clayton, D.A.** (1986). Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand replication. *J. Biol. Chem. 261*: 2384-2390.

Hoffman, R.J., Boore, J.L. and **Brown, W.M.** (1992). A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics 131*: 397-412.

Jacobs, H., Elliot, D., Math, V. and **Farguharson, A.** (1988). Nucleotide sequence and gene organization of sea urchin mtDNA. *J. Mol. Biol. 201*: 185-217.

Janke, A. and **Árnason, U.** (1997). The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent Archosauria (birds and crocodiles). *Mol. Biol. Evol. 14*: 1266-1272.

Janke, A. and **Pääbo, S.** (1993). Editing of tRNA anticodon in marsupial mitochondria changes its codon recognition. *Nucleic Acids Res. 21*: 1523-1525.

Janke, A., Feldmaier-Fuchs, G., Thomas, W.K., von Haeseler, T. and **Pääbo, S.** (1994). The marsupial mitochondrial genome and the evolution of placental mammals. *Genetics 137*: 243-256.

Janke, A., Gemmell, N.J., Feldmaier-Fuchs, G., von Haeseler, A. and **Pääbo, S.** (1996). The mitochondrial genome of a monotreme - the platypus (*Ornithorhyncus anatinus*). *J. Mol. Evol. 42*: 153-159.

Janke, A., Xu, X. and **Árnason, U.** (1997). The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria. *Proc. Natl. Acad. Sci. USA 94*: 1276-1281.

Johansen, S., Guddal, P.H. and **Johansen, T.** (1990). Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. *Nucleic Acids Res. 18*: 411-419.

Kondo, R., Satta, Y., Matsuura, E.T., Ishiwa, H., Takahata, N. and **Chigusa, S.I.** (1990). Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. *Genetics 126*: 657-663.

Kumazawa, Y. and **Nishida, M.** (1993). Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *J. Mol. Evol. 37*: 380-398.

Kumazawa, Y. and **Nishida, M.** (1995). Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. *Mol. Biol. Evol. 12*: 759-772.

Kumazawa, Y. and **Nishida, M.** (1999). Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. *Mol. Biol. Evol. 16*: 784-792.

Kumazawa, Y., Ota, H., Nishida, M. and **Ozawa, T.** (1996). Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Mol. Biol. Evol. 13*: 1242-1254.

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

Lee, W.-J. and **Kocher, T.D.** (1995). Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. *Genetics 139*: 873-887.

Lunt, D.H. and **Hyman, B.C.** (1997). Animal mitochondrial DNA recombination. *Nature 387*: 247.

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

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

Marschalek, R., Brechner, T., Amon-Bohm, E. and **Dingermann, T.** (1989).

Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. *Science 244*: 1493-1496.

Michaels, G.S., Hauswirth, W.W. and **Laipis, P.J.** (1982). Mitochondrial DNA copy number in bovine oocytes and somatic cells. *Dev. Biol. 94*: 246-251.

Mindell, D.P., Sorenson, M.D. and **Dimcheff, D.E.** (1998). Multiple independent origins of mitochondrial gene order in birds. *Proc. Natl. Acad. Sci. USA 95*: 10693-10697.

Miya, R. and **Nishida, M.** (1999). Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): first example of transfer RNA gene rearrangements in bony fishes. *Mar. Biotechnol. 1*: 416-426.

Noack, K., Zardoya, R. and **Meyer, A.** (1996). The complete mitochondrial DNA sequence of the bichir (*Polypterus ornatipinnis*), a basal ray-finned fish: ancient establishment of the consensus vertebrate gene order. *Genetics 144*: 1165-1180.

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

Okimoto, R., Chamberlin, H., Macfarlane, J. and **Wolstenholme, D.** (1991). Repeated sequence sets in mitochondrial DNA molecules of root-knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host-race identification. *Nucleic Acids Res. 19*: 1619-1626.

Okimoto, R., Macfarlane, J.L., Clary, D.O. and **Wolstenholme, D.R.** (1992). The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics 130*: 471-498.

Pääbo, S., Thomas, W.K., Whitfield, K.M., Yumazawa, Y. and **Wilson, A.C.** (1991). Rearrangements of mitochondrial transfer RNA genes in marsupials. *J. Mol. Evol. 33*: 426-430.

Quinn, T.W. (1997). Molecular evolution of the mitochondrial genome. In: *Avian Molecular Evolution and Systematics (*Mindell, D.P., ed). Academic Press, San Diego, CA, USA, pp: 3-28.

Quinn, T.W. and **Mindell, D.P.** (1996). Mitochondrial gene order adjacent to the control region in crocodile, turtle, and tuatara. *Mol. Phylogenet. Evol. 5*: 344-351.

Quinn, T.W. and **Wilson, A.C.** (1993). Sequence evolution in and around the mitochondrial control region in birds. *J. Mol. Evol. 37*: 417-425.

Ramirez, V., Savoie, P. and **Morais, R.** (1993). Molecular characterization and evolution of a duck mitochondrial genome. *J. Mol. Evol. 37*: 296-310.

Reiter, W., Palm, P. and **Yeats, S.** (1989). Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res. 17*: 1907-1914.

Robin, E.D. and **Wong, R.** (1988). Mitochondrial DNA molecules and virtual number of mitochondrial per cell in mammalian cells. *J. Cell. Physiol. 136*: 507-513.

Roe, B.A., Ma, D.-P., Wilson, R.K. and **Wong, F.-H.** (1985). The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem. 260*: 9759-9774.

Sankoff, D., Leduc, G., Antoine, N., Paquin, B., Lang, B.F. and **Cedergren, R.** (1992). Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. *Proc. Natl. Acad. Sci. USA 89*: 6575-6579.

Seutin, G., Lang, B.F., Mindell, D.P. and **Morais, R.** (1994). Evolution of the WANCY region in amniote mitochondrial DNA. *Mol. Biol. Evol. 11*: 329-340.

Smith, M.J., Banfield, D.K., Doteval, K., Gorski, S. and **Kowbel, D.** (1989). Gene arrangement in sea star mtDNA demonstrates a major inversion event during echinoderm evolution. *Gene 76:* 181-185.

Smith, M.J., Banfield, D.K., Doteval, K., Gorski, S. and **Kowbel, D.** (1990). Nucleotide sequence of nine protein-coding genes and 22 tRNAs in the mtDNA of the sea star *Pisaster ochraceus*. *J. Mol. Evol. 31*: 195-204.

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

Stanton, D.J., Daehler, L.L., Moritz, C.C. and **Brown, W.M.** (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.

Thyagarajan, B., Padua, R.A. and **Campbell, C.** (1996). Mammalian mitochondria possess homologous DNA recombination activity. *J. Biol. Chem. 217*: 27536-27543.

Tzeng, C.-S., Hui, C.-F., Shen, S.-C. and **Huang, P.C.** (1992). The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates. *Nucleic Acids Res. 20*: 4853-4858.

Ursing, B.M. and **Arnason, U.** (1998). The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). *J. Mol. Evol. 47*: 302-306.

Watanabe, Y., Tsuri, H., Ueda, T., Furushima, R., Takamiya, S., Kita, K., Nishikawa, K. and **Watanabe, K.** (1994). Primary and higher order structures of nematode (*Ascaris suum*) mitochondrial tRNAs lacking their T or D stem. *J. Biol. Chem. 269*: 22902-22906.

Wolstenholme, D.R., Macfarlane, J.L., Okimoto, R., Clary, D.O. and **Wahleithner, J.A.** (1987). Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proc. Natl. Acad. Sci. USA 84*: 1324-1328.

Yamazaki, N., Ueshima, R., Terrett, J.A., Yokobori, S.-I., Kaifu, M., Segawa, R., Kobayashi, T., Numachi, K.-I., Ueda, T., Nishikawa, K., Watanabe, K. and **Thomas, R.H.** (1997). Evolution of Pulmonate Gastropod mitochondrial genomes: comparisons of gene order organizations of *Euhadra, Cepaea* and *Albinaria* and implications of unusual tRNA secondary structures. *Genetics 145*: 749-758.

Yoneyama, Y. (1987). The nucleotide sequence of the heavy and light strand replication origins of the *Rana castabeiana* mitochondrial genome. *J. Nippon Med. Sch. 54*: 429-440.

Zardoya, R. and **Meyer, A.** (1997). The complete DNA sequence of the mitochondrial genome of a "living fossil", the coelacanth (*Latimeria chalumnae*). *Genetics 146*: 995-1010.

Zardoya, R. and **Meyer, A.** (1996). The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolli*) supports its phylogenetic position as a close relative of land vertebrates. *Genetics 142*: 1249-1263.

Zardoya, R., Garrido-Pertierra, A. and **Bautista, J.M.** (1995). The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*. *J. Mol. Evol. 41*: 942-951.

(Received January 12, 2000)