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.
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 was 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 - tRNAPro - tRNAThr - cytochrome b - tRNAAsn - ND6 - ND5 became control region - tRNAAsp - ND6 - tRNAPro - tRNAThr - 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; Harlild 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 Piciformes and some Passeriformes is 12S - tRNAPro - noncoding sequence - tRNAAsp - ND6 - tRNAPro - control region - tRNAThr - 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-
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_CUN 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} - 12S rDNA (Yoneyama, 1987), and the rice frog (*Rana limnocharis*) has the order control region - tRNA_{Thr} - tRNA_{Pro} - tRNA_{Leu(CUN)} - tRNA_{Phe} - 12S rDNA (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_{Pro} 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

**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).
probably came from to determine how these regions are now arranged. The amphibian O$_1$ 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, Amphibiaenia, Scincidae, Xantusidae (Macey et al., 1997) and Rhynochoccephalia (Seutin et al., 1994)) there was loss of the O$_1$ region. Representatives of the acrodonts, Agamidae and Chamaeleonidae, 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 crocodileians 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$_1$ 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$^{	ext{Gln}}$ from the IQM cluster to the WANCY cluster, between the tRNA$^{	ext{Top}}$ and tRNA$^{	ext{Ala}}$ 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$_1$ 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 akama, 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$^{	ext{Ser(UUR)}}$ and a spacer between the second control region and the LQM cluster. All species, except the boid python, has a tRNA$^{	ext{Pro}}$ 3’ end of the second control region. This tRNA seems to be a pseudogene in the akama. Its functional tRNA$^{	ext{Pro}}$ is in the “conserved” position while himehabu has a functional tRNA$^{	ext{Pro}}$ at the 3’ end of the second control region and a tRNA$^{	ext{Phe}}$ 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$^{	ext{Asp}}$ - tRNA$^{	ext{Pro}}$ - control region - LQM cluster - ND2. The akama and himehabu WANCY region has been sequenced, and it has been shown to have the typical organization for this region, including an O$_1$ 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 tuataras Sphenodon punctatus (Quinn and Mindell, 1996). One of the gene orders is similar to consensus gene organization without the tRNA$^{	ext{Thr}}$, 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$^{	ext{Glu}}$ - tRNA$^{	ext{Pro}}$ - tRNA$^{	ext{Thr}}$ - 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$^{	ext{Asp}}$ region with a reduced D arm and a tRNA$^{	ext{Ser(UCN)}}$ 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 akama 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$^{	ext{Leu(UUR)}}$ 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\textsubscript{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 \textit{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 \textit{et al.}, 1994, 1997).

The only deletion of genes as yet found in a vertebrate genome was reported by Janke \textit{et al.} (1994, 1997) in one marsupial. The typical tRNA of lysine is missing. In the corresponding region of the tRNA\textsubscript{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 \textit{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\textsubscript{Lys} from the cytoplasm to explain the function of the tRNA, by cytosine deamination (Janke and Pääbo, 1993; Janke \textit{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 \textit{et al.}, 1994, 1997).

The only deletion of genes as yet found in a vertebrate genome was reported by Janke \textit{et al.} (1994, 1997) in one marsupial. The typical tRNA of lysine is missing. In the corresponding region of the tRNA\textsubscript{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 \textit{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\textsubscript{Lys} from the cytoplasm to explain the function of the tRNA, by cytosine deamination (Janke and Pääbo, 1993; Janke \textit{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 \textit{et al.}, 1994, 1997).

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 \textit{et al.}, 1994). The absence of a typical O\textsubscript{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 \textit{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\textsubscript{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\textsubscript{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 \textit{et al.}, 1991; Kumazawa and Nishida, 1995; Macey \textit{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 \textit{et al.}, 1989), slime molds (Marschalek \textit{et al.}, 1989) and yeasts (Chalker and Sandmeyer, 1990). In \textit{Rana limnocharis}, transposition may explain the new organization, since terminal direct repeats were found (Macey \textit{et al.}, 1997).

Intramolecular recombination could generate new rearrangements (Thyagarajan \textit{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 \textit{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 tRNA 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 and seem to be involved in new 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 O₂ 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 O₂ 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 are great. This could be exemplified by the evolution of the tRNA^{Glu}_{1} region. In birds this gene is not associated with an O₂ region and its sequence is more variable when compared to groups such as mammals where it is associated with the O₂ (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.

REFERENCES


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


(Received January 12, 2000)