The Main Features of the Craniate Mitochondrial DNA Between the ND1 and the COI Genes Were Established in the Common Ancestor with the Lancelet

Christiane Delarbre,* Véronique Barriel,† Simon Tillier,† Philippe Janvier,‡ and Gabriel Gachelin*‡

*Unité de Biologie Moléculaire du Géne, Institut National de la Santé et de la Recherche Médicale, Institut Pasteur, Paris, France; †Laboratoire de Systématic Moléculaire, Muséum National d’Histoire Naturelle, Paris, France; and ‡Laboratoire de Paléontologie, Muséum National d’Histoire Naturelle, Paris, France

We have cloned the mitochondrial DNA fragment extending from tRNA-Leu to the cytochrome oxidase subunit 1 (COI) genes of Branchiostoma lanceolatum, Myxine glutinosa, Lampetra fluviatilis, and Scyliorhinus caniculus and have determined their respective gene sequences and organization. In all four species, this region contains the ND1 and ND2 genes and the genes coding eight tRNAs, namely, tRNA-Ile, -Gln, -Met, -Trp, -Ala, -Asn, -Cys, and -Tyr. The gene order is the same in the hagfish, lamprey and dogfish. In the lancelet, the location of the tRNA genes is slightly different. The mitochondrial code of Myxine, Lampetra, and Scyliorhinus is identical to that of vertebrates. The code used by the lancelet is the same with the exception of AGA (a stop codon in vertebrates), which codes for glycine in the lancelet. From the comparison of the four maps with already published ones for other species, we propose that the main features of the craniate mtDNA between the ND1 and COI genes were established in the common ancestor to cephalochordates and vertebrates more than 400 MYA. The origin of replication of the light-strand (Ori-L), usually located between the tRNA-Asn and tRNA-Cys genes in vertebrates, was not found in the lancelet, hagfish, or lamprey (Lampetra). In contrast, it was found in the dogfish. Thus the position of Ori-L was established for the first time in the common ancestor to the Chondrichthyes and Osteichthyes and remains present in all later-emerging vertebrates.

Introduction

The phylogenetic relationship between the Recent representatives of the early-diverging prochordate and chordate lineages is still the subject of controversies (Forey and Janvier 1993). The study of the organization of the mitochondrial genome has been used in the past to answer similar questions concerning other species (see for example, Kumazawa and Nishida 1995). In the course of a cladistic analysis of the lancelet, hagfish, lamprey, and dogfish relationships, based on the sequences of the mitochondrial NADH dehydrogenase subunit 1 (ND1) and NADH dehydrogenase subunit 2 (ND2) genes (unpublished data), we have cloned and sequenced the region of the mitochondrial DNA which extends from Leu-tRNA to COI genes in the lancelet (Branchiostoma lanceolatum), hagfish (Myxine glutinosa), lamprey (Lampetra fluviatilis), and dogfish (Scyliorhinus caniculus) and encompasses the two above-mentioned genes. In the absence of information concerning a similar region in urochordates, we have compared the gene organization of these four species with those of the echinoderm Paracentrotus lividus (Cantatore et al. 1989) and the teleost Cyprinus carpio (Chang, Huang, and Lo 1994). In the hagfish, lamprey, and dogfish, the gene organization did not differ significantly from that of the mitochondrial genome of other vertebrates (Anderson et al. 1981; Chang, Huang, and Lo 1994; Lee and Kocher 1995). That of the lancelet, in spite of local differences, clearly also belongs to the craniate type. We thus propose that the organization of the mtDNA, as now present in vertebrates, was established very early in the course of chordate evolution, before the emergence of cephalochordates. On the other hand, the origin of replication of the light-strand (Ori-L), usually located between the tRNA-Asn and tRNA-Cys genes in vertebrates, was not found in the lancelet, hagfish, or lamprey (Lampetra). In contrast, it was found in the dogfish. Thus the position of Ori-L was established for the first time in the common ancestor to the Chondrichthyes and Osteichthyes and remains present in all later-emerging vertebrates.

Key words: lancelet, hagfish, lamprey, dogfish, mitochondrial DNA, molecular phylogeny.

Address for correspondence and reprints: Gabriel Gachelin, Unité de Biologie Moléculaire du Géne, Département d’Immunologie, Institut Pasteur, 25 rue du Dr. Roux, 75015 Paris, France. E-mail: ggachel@pasteur.fr.

High-molecular-weight DNA was prepared from liquid nitrogen frozen samples of the same organs using a Proteinase-K-based technique (Hogan et al. 1994). The genomic DNA of the lancelet was prepared using several animals.

PCR Amplification

Sequences of mtDNA and mtcDNA were PCR amplified from high-molecular-weight DNA and single-strand cDNA, respectively, using a variety of primers and conditions, depending on the evolutionary distance separating the species under study. PCR amplification was carried out in 50 μl of reaction mixture with 2.5 units of the high-fidelity Pfu polymerase (Stratagene) for 3 min at 94°C followed by 50 cycles (1 min at 94°C, 1 min at 50–60°C, 4–8 min at 72°C). The last cycle was terminated by incubation at 72°C during 10 min. The primers were added at the final concentration of 0.5 μM; dNTP were used at a final concentration of 200 μM.

The sequence extending from the 3' end of the ND1 gene to the 5' end of the COI gene was amplified using primers located either in the ND1 gene: for Scyliorhinus, Lampetra, and Myxine: 5'-CCACGATTTCCGATATGATCAAC-3'; for Branchiostoma: 5'-TTTCTCTGGCA TCGGGGAGGGC-3'; and within the COI gene (common reverse primer): 5' -CCRATRTCYTTGTGWTAG TAGA-3', in which R stands for A, G; Y for C, T; and W for A, T.

The sequence extending from the tRNA-Leu gene into the ND1 gene was amplified using primers located either in the tRNA-leu gene (for Branchiostoma, Scyliorhinus, and Lampetra): 5' -TCCRMAAGRYTATA GCCYCTT-3' (in which M stands for A, C), or in the 16S mRNA gene for hagfish: 5' -CGTGATCTGAGTTGAGG-3'. The reverse primers were all located in the 3' end of the ND1 gene: Branchiostoma: 5' -CCGGTATTCAAGAAGAGCC-3'; Myxine: 5' -ATTAAATAAGAGGAGTGATAAGGC-3'; Scyliorhini- nus: 5' -GATAAATTGCTAGGTTAGATAG-3'; Lampetra: 5' -GAGTAAAAAGGCTAGGTTAGG-3'.

Isolation of Clones Derived from the ND1–COI Region

PCR fragments were electroeluted from agarose gels, 5' phosphorylated using polynucleotide kinase (Boehringer), and ligated with the T4 DNA ligase (Boehringer) at the dephosphorylated EcoRV site of the pBluescript KS vector (Stratagene) as described elsewhere (Delarbre et al. 1992).

Sequencing of DNA

The sequences of the clones were determined using the Sequenase kit (Amersham-USB) and a variety of primers, either internal, derived from partial nucleotide sequences of the clones, or primers in pBluescript KS vector (17 and reverse primer).

Results and Discussion

The nucleotide sequences of the DNA region extending from the tRNA-Leu gene to the COI gene were determined for B. lanceolatum, M. glutinosa, L. fluvia- tillis and S. caniculus. The accession numbers at EMBL/GenBank libraries are Y09524, Y09527, Y09528, and Y09526, respectively. The tRNAs were identified on the basis of their distinctive traits, and their extents were identified by modeling their expected secondary structures. The protein-coding sequences were identified by alignment of the encoded sequences with the conserved portions of the ND1, ND2, and COI proteins.

Distinctive Features of the Genes

The ND-1 and ND-2 subunits are components of Complex I (NADH-ubiquinone reductase) of the electron transport chain. They are encoded by mitochondrial genes. The coding sequences of the ND1 and ND2 genes of the four animals studied were determined using the vertebrate mitochondrial code, which generates the best amino acid alignment among themselves and with the sequences of a variety of animals from echinoderms to mammals. ND1 amino acid sequences are far more conserved than those of ND2, presumably because of stronger structural constraints of ND1 in Complex I. Upon comparison of the amino acid sequences of the lancelet, hagfish, lamprey and dogfish with themselves and with the sequence of the bichir, a basal actinopterygian (Noack, Zardoya, and Meyer 1996), the lancelet appears to be equally distant from all the other animals studied (61%–65% identity of the amino acid sequence for ND1 and 30%–34% identity for ND2). The amino acid sequences of the dogfish are far more similar to those of the bichir (73% and 59% for ND1 and ND2, respectively). Finally, the similarities between Lampetra and Petromyzon (Lee and Kocher 1995) are 96% for ND1 and 90.5% for ND2.

Some amino acids play important roles in the folding and function of proteins. These residues retain the exact same position among phylogenetically distant species. Considering the triplets coding for those amino acids, one can thus deduce the genetic code used by the four animals studied. Under this assumption, all four species studied appear to use the vertebrate mitochondrial code. The only exception is the codon AGA, a stop codon in all vertebrates (Jukes and Osawa 1990), found at position 79 (counting the first methionine of the sequence as 1) of the ND2 gene of the lancelet. No other AGA was found elsewhere in the genes we sequenced. The presence of this unique AGA was confirmed by sequencing several genomic clones. The AGA was also found in the cDNA and thus was not altered by mRNA editing. AGA is thus unlikely to be a stop codon in the ND2 gene of the lancelet. The amino acid found at this position is not conserved in other species, and its identity in the lancelet cannot be ascertained on this basis. AGA is used by echinoderms to code for serine (Himen et al. 1987). In the fragments of sequences published, the codon AGA is used in tunicates (along with the vertebrate mitochondrial code) to code for glycine (Yokobori, Ueda, and Watanabe 1993). We have found a polymorphism at the aa position 84 of the ND2 protein of the lancelet in which AGG is present in a clone and GGG in five additional clones: AGG is also used for
glycine in tunicates (Yokobori, Ueda, and Watanabe 1993), whereas GGG unambiguously codes for glycine. We thus propose the assignment of AGA for glycine in the lancelet. The use of AGA (and perhaps AGG) in association with an otherwise vertebrate-type mitochondrial code would constitute an ancestral trait shared by the lancelet and tunicates. The rarity of its usage in the lancelet and its absence in the hagfish, lamprey, and dogfish may suggest the time of the disappearance of AGA as a coding codon during the evolution of the mitochondrial code (Osawa et al. 1992).

The codon usage was essentially the same for the ND1 and ND2 genes within a given species, but differed markedly between species (table 1). Overall, the codon usage of the dogfish is similar to that of the lamprey. The codon usage in the hagfish shares the most features with the latter two species but displays a preferential usage of some codons (like GGG and CCT), as in the lancelet. The codon usage of the latter is significantly different. In the lamprey and dogfish, as in vertebrates, the third letter of the four-fold degenerate codons is rarely G (3%-5%). Its frequency goes up to 12% in the hagfish and 24% in the lancelet.

The initiation codon was usually ATG, with the exception of ATA in the ND1 gene of the lancelet. The initiation codon was GTG in the COI gene of all animals with the exception of the ATG of the COI gene of the hagfish. It is generally agreed that GTG is the initiation codon in the genes which are located immediately downstream of a tRNA (Gadaleta et al. 1989). This rule is unlikely to be applied universally, since the codon GTG was used as initiation codon by Lampetra, Scyliorhinus, and Branchiostoma, although it was separated from the tRNA-Tyr by one, one, and nine nucleotides, respectively.

We have also determined the 3’ sequences of the mRNA coding for ND1 of the lancelet and ND2 of the lamprey and the hagfish. The comparison of the cDNA and gene sequences showed that the termination codon was always TAA in the cDNAs. The primary transcripts are thus cleaved within TAG, which is converted into TAA by addition of a stretch of A after T or TA, as in the case of human mitochondrial mRNAs (Anderson et al. 1981; Ojala, Montoya, and Attardi 1981).

Table 1: Codon Usage for ND1 and ND2 Proteins of R. lanceolatum (B), M. glutinosa (M), L. flavovittis (L), and S. cyclurus (S)

<table>
<thead>
<tr>
<th>Codon Usage for ND1 (B)</th>
<th>M</th>
<th>L</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phc...</td>
<td>TTT</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Leu...</td>
<td>TTA</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Ile...</td>
<td>ATC</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Leu...</td>
<td>CTT</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Val...</td>
<td>GTT</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Met...</td>
<td>ATA</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Ser...</td>
<td>CTA</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Thr...</td>
<td>ACC</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Lys...</td>
<td>AGA</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>His...</td>
<td>CAT</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Tyr...</td>
<td>TAC</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Trr...</td>
<td>TTT</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Ser...</td>
<td>CTA</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Thr...</td>
<td>ACC</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Lys...</td>
<td>AGA</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>His...</td>
<td>CAT</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Tyr...</td>
<td>TAC</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Trr...</td>
<td>TTT</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Ser...</td>
<td>CTA</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Thr...</td>
<td>ACC</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Lys...</td>
<td>AGA</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>His...</td>
<td>CAT</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Tyr...</td>
<td>TAC</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

Concerning the frequency of the amino acids encoded by the two genes, the lancelet differs from the other animals studied. Some amino acids are more frequently used by the lancelet than by the other species. The frequency of glycine is about 10% in the lancelet and about 5% elsewhere; that of valine is 9% versus 3.5%. Conversely, threonine and isoleucine are less frequently used by the lancelet: 3% versus 7% and 5% versus 9%, respectively.

Finally, we have found two different sequences for the ND2 gene of the lancelet, with a total of 42 mismatches spread at random along the sequence. Most mutations are synonymous. Eight are nonsynonymous. At amino acid position 92, H was mutated into N; at position 157, L into F; at position 191, A into V; at position 199, S into L; at position 217, S into C; at position 221, M into T; at position 229, H into Q; and at position...
FIG. 1.—Molecular organization of the tRNA-Leu—COI region of mitochondrial DNA of *B. lanceolatum* (A), *M. glutinosa* (B), *L. fluviatilis* (C), and *S. caniculus* (D). The sequences coding for ND1 and ND2 have been removed except at the 5' and 3' ends, whose initiation and stop codons (*) are underlined. The bars indicate the limits of the DNA coding for the tRNAs. The anticodons of the tRNAs are underlined. The numbers show the first and the last nucleotides of the sequenced DNA fragments.
FIG. 2.—Comparative organization of the tRNA-Leu-COI regions. The genes are aligned on the ND1 gene and are identified in each box. The unmarked heavily dotted boxes correspond to the 16S rRNA gene. The origin of replication of the L-strand is indicated by a hatched narrow box.

324, M into T. This polymorphism may be explained by the existence of a second mitochondrial or nuclear copy of the ND2 gene. The possible existence of additional copies of the ND2 gene is also suggested by the fact that we have not observed the same level of variation in the other genes of the lancelet we have sequenced.

The genes of the tRNAs were identified on the basis of their locations, their sequences, and the putative secondary structures of the encoded tRNAs. The total lengths of the eight tRNAs were 524, 537, 550, and 561 bp in the lancelet, hagfish, lamprey, and dogfish, respectively. Differences in the size of the tRNA arms account for the observed differences in length. All tRNAs display a canonical structure, with the exception of the tRNA-Cys of the lancelet, which has a much shorter D-arm (no stem), a structure similar to that found for tRNA-Cys of the lepidosaurian reptile Sphenodon (tuatara) (Seutin et al. 1994) and some lizards (Macey et al. 1997).

In addition to being shorter in their overall length, the tRNA genes of the lancelet often overlap with each other or with protein-encoding genes. In particular, the tRNA-Asn gene overlaps the 3' end of the ND2 gene by eight nucleotides. Overlapping is less marked in the other species studied.

Molecular Organization of the tRNA-Leu-COI Region

The length of the sequences extending from the tRNA-Leu gene to the initiation codon of the COI gene gradually increases from 2,504 bp in the lancelet to 2,563 bp in the hagfish, 2,603 bp in the lamprey, and 2,626 bp in the dogfish. The mitochondrial genome is known to be compact. In this respect, it is worth noting that this region of the genome seems more compact in earlier emerging lineages.

A detailed molecular organization of the tRNA-Leu-COI region was drawn from the nucleotide sequences (fig. 1A–D). The maps generated for the four species are very similar. They harbor the same genes and the same tRNAs as in fishes, amphibians, birds, and
mammals (Anderson et al. 1981; Roe et al. 1985; Desjardins and Morais 1990; Noack, Zardoya, and Meyer 1996; Zardoya and Meyer 1996). The tRNA-Leu gene is located 5' to ND1; ND2 is separated from ND2 by tRNA-Ile, -Gln, and -Met (IQM). The ND2 gene is separated from the COI gene by tRNA-Trp, -Ala, -Asn, -Cys, and -Tyr (WANCY) (fig. 2). The orientation of the ND1, ND2, and COI genes and of the tRNA genes is the same in all four species, and they are transcribed from the same strand as in vertebrates. The order of the tRNAs is the same in the dogfish, hagfish, and lamprey as in fishes, birds, and mammals. Although the general pattern of organization is basically the same in the lancelet as in the other animals studied, the order of the same tRNA genes is slightly different in the lancelet (IMQ and NWACY). The two differences can each easily be explained by the translocation of a single tRNA.

In Osteichthyes and most other vertebrates, the origin of replication of the L-strand of mitochondrial DNA forms a characteristic loop, located between the tRNA-Asn and tRNA-Cys genes. The Ori-L loop was found in the dogfish at this very location. The structure of the Ori-L of the dogfish is similar to that of carp (Chang, Huang, and Lo 1994) (fig. 3), and, as in Xenopus (Roe et al. 1985), mouse (Bibb et al. 1981), and human (Anderson et al. 1981), it is a T-rich loop. Stretches of DNA which could fold into a loop of similar characteristic structure were not present in the ND1–COI region of the lancelet, hagfish and lamprey. The absence of Ori-L had already been noted in the sea lamprey (Lee and Kocher 1995).

Finally, an extended map of the mitochondrial genome (from the Cytochrome b gene, accession number Y09849, to the ATPase 8/6 gene, accession number Y09525) of the lancelet shows it very similar to that of the sea lamprey Petromyzon marinus (data not shown). Such an organization is found neither in nonchordate deuterostomes like echinoderms (Cantatore et al. 1989) nor in protostomes (Clary and Wolstenholme 1985).

Conclusion

The results reported in the present paper are in good agreement with the phylogenetic relationships among early vertebrates that are currently accepted (discussed in Janvier 1996). The lancelet is considered to be the sister group of vertebrates (Garcia-Fernandez and Holland 1994; Gee 1994). We show that, in spite of some distinctive and presumably ancestral traits, the lancelet also shares its mitochondrial gene organization (at least for the region we have studied) with craniates but not with echinoderms. The ancestral organization of this region of mtDNA was most probably from the origin of chordates, close to the one observed in recent lineages, and under went minor changes since more than 400 MYA. The chordate common ancestor most likely also lacked an identifiable Ori-L in the ND1–COI region. This looped structure, still absent in Agatha, is likely to have appeared later, in the common ancestor of Chondrichthyes and Osteichthyes, as it has been first detected in Chondrichthyes and has remained since then in most other vertebrates. Its absence in crocodiles and birds is likely to be a secondary loss in the common ancestor of all archosauromorphs (crocodiles, birds, and a large number of fossil taxa, including “dinosaurs”). Considering current morphology-based phylogenies of archosauromorphs, it may be inferred that dinosaurs lacked Ori-L as well. The fact that the Ori-L is also lacking in the blind snake (Typhlopsidae) (Kumazawa and Nishida 1995), the tuatara (Seutin et al. 1994), and some lizards (Macey et al. 1997), although most other lizards and snakes studied possess it, is more puzzling. This absence can be regarded as secondary losses independent from that in archosauromorphs.

LITERATURE CITED


ROSS H. CROZIER, reviewing editor

Accepted April 23, 1997