Nucleic Acids Research, 1998, Vol. 26, No. 13 **3279–3285**

Complete sequence of the amphioxus (*Branchiostoma lanceolatum*) mitochondrial genome: relations to vertebrates

Nathalie Spruyt, Christiane Delarbre[1], Gabriel Gachelin[1] and Vincent Laudet*

CNRS UMR 319, Mécanismes du Développement et de la Cancérisation, Institut de Biologie de Lille, Institut Pasteur, 1 rue Calmette, 59021 Lille, France and [1]Unité de Biologie Moléculaire du Gène, INSERM U277, Département d'Immunologie, Institut Pasteur, 25 rue du Dr Roux, 75015 Paris, France

Received February 4, 1998; Revised and Accepted May 18, 1998

DDBJ/EMBL/GenBank accession no. Y16474

ABSTRACT

The complete nucleotide sequence of the mitochondrial DNA of the amphioxus *Branchiostoma lanceolatum* has been determined. This mitochondrial genome is small (15 076 bp) because of the short size of the two rRNA genes and the tRNA genes. In addition, this genome contains a very short non-coding region (57 bp) with no sequence reminiscent of a control region. The organisation of the coding genes, as well as of the two rRNA genes, is identical to that of the sea lamprey. Some differences in the repartition of the tRNA genes occur when compared to the lamprey. The mitochondrial codon usage of the amphioxus is reminiscent of that of urochordates since the AGA codon is read as a glycine and not as a stop codon as in vertebrates. Moreover, the base composition at the wobble positions of the codon is strongly biased toward guanine. Altogether, these data clearly emphasise the close relationships between amphioxus and vertebrates, and reinforce the notion that prochordates may be viewed as the brother group of vertebrates.

INTRODUCTION

The pattern of mitochondrial genome organisation is informative in the phylogeny of distantly related taxa, because the rate of gene rearrangement is much slower than nucleotide substitutions (1; for a review see 2). Shifts in gene order define major lineages without evidence of parallelism or reversal. The numerous complete sequences of mitochondrial genomes determined to date have revealed that each phylum exhibits a common basic gene order, despite minor relocalisations of tRNA genes in some taxa (3–21 and references therein). In this respect, all vertebrates share a common organisation of the mitochondrial genome, although minor rearrangements have been found in chicken, reptile, amphibian and marsupial mitochondrial DNAs (mtDNAs). These differences can in most cases be explained by inversion of DNA fragments (9,22–25).

MtDNA genomic maps can be used as a tool to reconstruct the evolutionary history of distantly related taxa (1). This is particularly useful in the case of the phylogeny of deuterostomes, due to the large evolutionary gaps existing between present representatives of early chordates. For instance, it has been shown that several changes in the gene order are observed when the mitochondrial genomes of the sea urchin and of the sea lamprey are compared (7,8,14,17,26,27). Interestingly, the lamprey itself harbours a different gene order as compared with other vertebrates such as dogfish or various teleosts, although the difference in gene order is more subtle than between sea urchin and lamprey (14,16,20).

In this context it was particularly interesting to study the case of the lancelet (or amphioxus) *Branchiostoma lanceolatum*. Indeed, this prochordate has for many decades been considered as the most closely related to the vertebrates (28 and references therein). Many recent data, obtained by comparing the expression pattern of genes that are important for embryonic development, have emphasised the close relationship between prochordates and vertebrates (29). For example the Hox genes complex of the amphioxus, although existing as a single copy as compared to the four loci known in vertebrates, exhibits a highly similar organisation and expression pattern (30). The aim of the present study is to determine the prochordates/vertebrates relationship using the mtDNA gene order as a tool. We have thus determined the complete sequence of the amphioxus *B.lanceolatum*. This sequence reveals an organisation identical to the one of the lamprey with the exception of some differences in the location of tRNAs. Interestingly, despite this 'vertebrate like' organisation, the amphioxus mtDNA retains ancestral characteristics such as a codon usage reminiscent of that of the sea urchin. Taken together, these results clearly confirm the status of the prochordates as a brother group to the vertebrates.

MATERIALS AND METHODS

DNA isolation

Complete adult amphioxus (*B.lanceolatum*) were caught offshore at Roscoff (France) and frozen in liquid nitrogen. DNA was

*To whom correspondence should be addressed at present address: CNRS UMR 49, Laboratoire de Biologie Moléculaire et Cellulaire, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France. Tel: +33 4 72 72 81 90; Fax: +33 4 72 72 86 86; Email: vincent.laudet@ens-lyon.fr

extracted by the Proteinase K digestion method as in Escriva *et al.* (36). Sequencing was done with the specific methods recommended by the supplier of our automatic DNA sequencing apparatus.

PCR

In order to have a specific probe encompassing regions of the mtDNA, we performed PCR experiments using degenerated oligonucleotides in order to amplify a short and conserved region of the 16S rRNA gene. We used a 'touch-down' PCR cycle as described by Escriva *et al.* (36). The oligonucleotides used were as follows (I is an inosine):

A (+) 5′-(C/T)(A/T)A CC(g/C) (T/C)Ag ggA TAA CAg Cg-3′
B (+) 5′-(T/A)I(A/g) g(g/T)T T(A/g)C gAC CTC gAT gT-3′
C (–) 5′-g(g/T)T CT(A/g) AAC (C/T)CA (g/A)(A/C)T CAC gT-3′
D (–) 5′-(T/A)(T/C)(A/T) (g/C)I(A/g) (g/T)TC CTT TCg TAC TA-3′

These oligonucleotides allow the amplification of a fragment of 104–217 bp depending on the combination used.

Genomic library screening

In order to isolate the complete mitochondrial genome we screened a *B.lanceolatum* genomic library kindly furnished by Peter Holland with our 16S rDNA specific probe. Among the 12 positive clones, three were characterised in more detail since they encompass most, if not all, of the genome. Library screening was done according to standard procedures.

Sequencing

DNA was prepared from these clones using standard procedures (37) and was then used as a template for sequencing reactions using the Sanger method and the PRISM kit from Applied Biosystems and various oligonucleotides for priming. The sequence of the oligonucleotides used for the DNA walking is available upon request to V. Laudet. The sequencing was done on both strands by a primer walking strategy done independently on each strand of the mtDNA molecule. Furthermore, each step was designed to largely overlap with the preceding one, in order to ensure that each base of each strand could be read using at least two different primers. Sequencing reactions were then run on an Applied Biosystem 377A automatic sequencer using the conditions recommended by the supplier. Sequencing was done with the specific methods recommended by the supplier of our automatic DNA sequencing apparatus.

Sequence analysis

The sequences were aligned, compared and translated using the software available on the Infobiogen network (www.infobiogen.fr) and the DNA strider software.

RESULTS AND DISCUSSION

Size of the amphioxus mitochondrial genome

The mtDNA of *B.lanceolatum* is 15 076 bp long (Fig. 1 and Table 1). To our knowledge, it is the shortest complete mitochondrial genome analysed to date in deuterostomes. The size of the mitochondrial genome is ~16 400–17 000 bp in mammals (Table 2; see 2 for a review; the platypus has a genome of 17 019 bp; 19), 16 700 bp in birds (9), 17 500 bp in amphibians (5), 16 500 bp in Osteichthyes (12,13,18), 16 700 bp in Chondrichthyes (21), 16 200 bp in lamprey (14) and 15 600 bp in Echinoderms (7,8,17; H.Himeno, H.Masaki, T.Kawai, T.Ohta, I.Kumagai, K.Miura and K.Watanabe, unpublished). The only shorter genomes found to date in metazoans are those of two nematodes, namely *Caenorhabditis elegans* (13 794 bp) and *Ascaris suum* (14 284 bp), in which the ATPase 8 gene is missing (11; see 2 for a review). In spite of this small size, the amphioxus mtDNA contains the same number of genes (13 protein coding genes, two rRNA genes, 22 tRNAs genes) as larger genomes and may even contain one more tRNA than the lamprey mtDNA

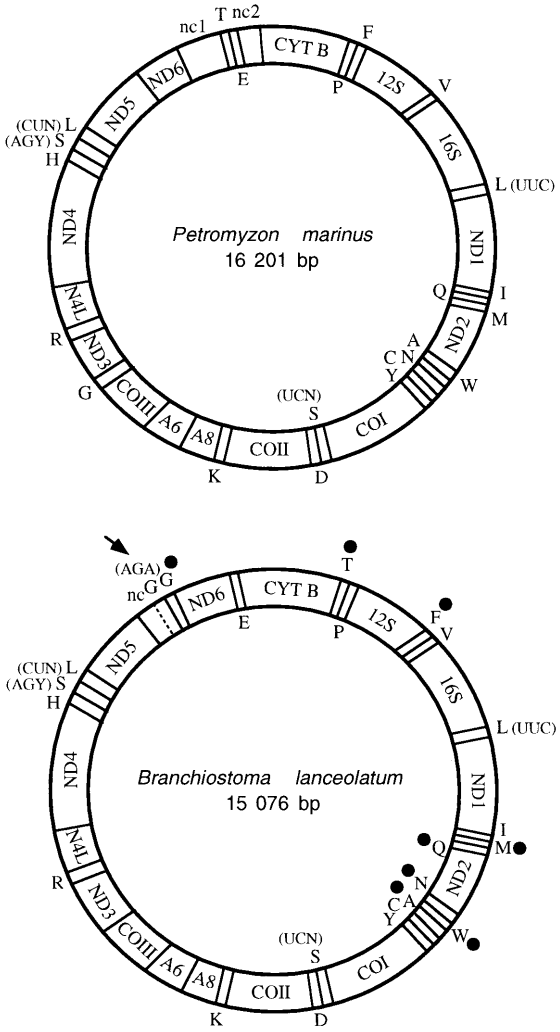

Figure 1. Genes map of the amphioxus (*B.lanceolatum*) mitochondrial genome (lower panel) compared to that of the sea lamprey *P.marinus* (upper panel). The positions of the 13 protein coding genes and of the two rRNA genes are indicated using the same abbreviations as in Table 1. All protein-coding genes except ND6 are encoded on the first strand (outer) with clockwise transcriptional polarity. tRNA genes are represented by the one letter amino acid code located either outside or inside the circles according to the coding strand. The codon families of each serine and leucine tRNA are presented in brackets. Those labelled outside the circle are encoded on the first strand with clockwise transcriptional polarity. The two non-coding regions of the lamprey genome are indicated as nc1 and nc2 whereas the single non-coding region of the amphioxus mitochondrial genome is indicated as nc. The extra tRNA is indicated by an arrow as a tRNA for glycine on the AGA codon. The tRNA genes located at different positions in amphioxus when compared to lamprey are indicated by a black spot.

Table 1. Localisation of the mtDNA genes and non-coding regions of *B.lanceolatum*

Gene	from	to	size (bp)	size (aa)	Strand
CYT b	1	1143	1143	380	
tRNA-Thr	1143	1208	66		
tRNA-Pro	1207	1274	68		L
12S rRNA	1275	2118	844		
tRNA-Phe	2119	2185	67		
tRNA-Val	2188	2254	67		
16S rRNA	2255	3621	1367		
tRNA-Leu (TTR)	3622	3692	71		
NADH1	3696	4637	942	313	
tRNA-Ile	4636	4698	63		
tRNA-Met	4699	4765	67		
tRNA-Gln	4765	4833	69		L
NADH 2	4835	5875	1041	346	
tRNA-Asn	5868	5937	70		L
tRNA-Trp	5937	6005	69		
tRNA-Ala	6004	6066	63		L
tRNA-Cys	6067	6124	58		L
tRNA-Tyr	6125	6190	66		L
CO I	6201	7748	1548	515	
tRNA-Ser (TCN)	7746	7816	71		L
tRNA-Asp	7825	7889	65		
CO II	7890	8580	691	230	
tRNA-Lys	8581	8646	66		
ATPase 8	8647	8811	165	54	
ATPase 6	8805	9488	684	227	
CO III	9488	10276	789	262	
NADH 3	10277	10630	354	117	
tRNA-Arg	10628	10696	69		
NADH 4L	10698	10973	276	91	
NADH 4	10973	12331	1359	452	
tRNA-His	12331	12396	66		
tRNA-Ser (AGY)	12397	12462	66		
tRNA-Leu (CTN)	12463	12530	68		
NADH 5	12531	14327	1797	598	
NC sequence	14328	14456	129		
tRNA-Gly	14457	14524	68		
NADH 6	14506	15009	504	167	L
tRNA-Glu	15011	15075	65		L

(see below). The short size of the amphioxus mitochondrial genome is accounted for by the following. (i) A general compaction of the size of the regions coding for RNA molecules: 12S rRNA (844 bp in amphioxus; 976, 900 and 957 bp respectively in sea urchin, lamprey and dogfish), 16S rRNA (1367 bp in amphioxus; 1530, 1621 and 1670 bp respectively in sea urchin, lamprey and dogfish) and a reduced size of the tRNA genes (66.7 nt on average in amphioxus versus 72.8 nt in dogfish). By a comparison with the human genes, we observed that the small size of the 12S and 16S rRNA genes of amphioxus is mainly due to a reduction of the size of the loops (data not shown). Figure 2A presents an alignment of tRNA genes from amphioxus compared with their homologues in lamprey and sea urchin. It is clear from these alignments as well as from the clover leaf structure of the $tRNA^{Asp}$ presented in Figure 2B that the loops of the tRNA are shortened in amphioxus when compared to lamprey or sea urchin. Nevertheless, in some cases, we also observed tRNAs of identical sizes in the three species. In such cases (exemplified by the $tRNA^{Cys}$ in Fig. 2A), we observed that the space between two adjacent tRNAs is reduced in amphioxus when compared with other species. Thus, apparently several mechanisms play a role in the size reduction of the tRNA genes in the amphioxus mitochondrial genome. This clearly suggests that a selective pressure toward size reduction is effectively working in amphioxus. (ii) An apparent absence of a recognisable DNA replication control region. Indeed, we noticed that the only non-coding region found in the lancelet genome is a short stretch of 129 nt located between the NADH5 gene and the tRNA-Gly (positions 14328–14456). Since this non-coding sequence harbours in its 3′ part a region which may be folded in a non-conventional tRNA (see below), only 57 bp remain totally unassigned, and may correspond to a control region. However, this sequence does not harbour any signal known to be implicated in the replication of the mtDNA, such as conserved sequence blocks (CSB), termination associated sequences (TAS) or even palindromes. The only salient feature of this 57 bp sequence is the presence of a 9 bp inverted repeat (TTTTTTGGG, positions 14343–14361). (iii) The absence of a recognizable (i.e. independent of the tRNA) origin of replication of the L strand (31).

Table 2. Lengths of mitochondrial genes in deuterostomes

	Sea urchin	Sea star	Amphioxus	Lamprey	Dogfish	Carp	Frog	Human
Control region	121	446	129[a]	491	1050	928	2134	1043
12S rRNA	976	896	844	900	957	950	951	954
16S rRNA	1530	1530	1367	1621	1670	1680	1621	1559
Cytb	1157	1140	1143	1191	1144	1139	1140	1141
ND1	969	980	942	966	975	974	970	956
ND2	1059	1064	1041	1044	1047	1046	1039	1042
ND3	351	332	354	351	351	351	342	346
ND4	1380	1382	1359	1377	1381	1379	1384	1378
ND4L	294	296	276	291	297	296	297	297
ND5	1914	1931	1797	1797	1830	1823	1815	1811
ND6	495	488	504	519	522	518	513	528
COI	1554	1553	1548	1554	1554	1550	1549	1541
COII	690	687	691	690	691	689	688	684
COIII	783	783	789	786	786	785	781	784
ATP6	690	692	684	714	684	683	679	679
ATP8	168	164	165	168	168	164	168	207
Total	15 650	16 260	15 076	16 201	16 696	16 575	17 553	16 569

Data are from: sea urchin (*Strongylocentrotus purpuratus*) (7); sea star (*Asterina pectinifera*) (H.Himeno, H.Masaki, T.Kawai, T.Ohta, I.Kumagai, K.Miura and K.Watanabe, unpublished). Genbank accession number D16387; Amphioxus (*B.lanceolatum*) this study; Lamprey (*P.marinus*) (14); Dogfish (*S.canicula*) (Delarbre *et al*., in press). Carp (*Cyprinus carpio*) (13); Frog (*Xenopus laevis*) (5); Human (*Homo sapiens*) (3).

[a]Irrespective of the presence of an extra-tRNA which further reduces the size of the non-coding region to 57 bp.

A

```
tRNA Ala (6004-6066)
              10        20        30        40        50        60        71
AMPH  TAAAATCTG-AGGAGACC-----ACATTTTTTGATTGCAAATCAAATAGTTTAT--TTAACTTAAGATTTT   63
LAMP  ......T..C.A..T.T.CTC--....C..C...AC...G..GC....AA-....GC...A.....   68
URCH  .GT.GCT.CT.A..CTTTAACTT.G.AC.....C...G....CGT...TCAA....AC...A.CCAC   71

tRNA Ile (4636-4698)
              10        20        30        40        50        60        70
AMPH  GAAATGTAGTTTAAAAAAAACGCTGTTT---TGATAGGGCAGAGATGCA-----TAAAATGTCATTTTTA   62
LAMP  ...G.A.GTCCG.C..T.GGAA.CAC..---.....A.--TG.C.A..GGGGA..TT.CCC.C...C.T   65
URCH  ..GC.TGCCCC....GTT..G.-...C.A.C.....ATT-...TT.--.AGGG-.T....-..CC..CA.   64

tRNA Cys (6067-6124)
              10        20        30        40        50        60      68
                                                                  Y
AMPH  AGGCTTTCAAA-AGGTCTACTTTTACTTTCTGGGTTGCAACCAGAATTTTTATAAAGCCTAGTAAAAG   67
LAMP  ..A.C.CA.C.G.AC.GCTTC.A..TC...A.AT.....T.TA.CA.GAAC.TC.C.A..AGGTC..   68
URCH  .A.....GGC.-.C...CTAG..G.TC.AGG.AT.....T.C...A.G...GTT.CA...CA...CC   67

tRNA Asp (7825-7889)
              10        20        30        40        50        60        71
AMPH  GAGGTATTAGGTTAA-TTAAAACTTACAGTTTGTCAGGTTGTAAATCTTAG---TCTTA---AGATATCTC   64
LAMP  A..A..C...TAA..-...TT..ACT-.CC.....A.G.AAT.T.A.G..CT-..C..CTC.TG.....T   68
URCH  AGA.GCC...T.A..C...T.....T-G.G.....CCAA..T.GC.G.TTAAAC.CCAGC.GCT...G   70
```

B

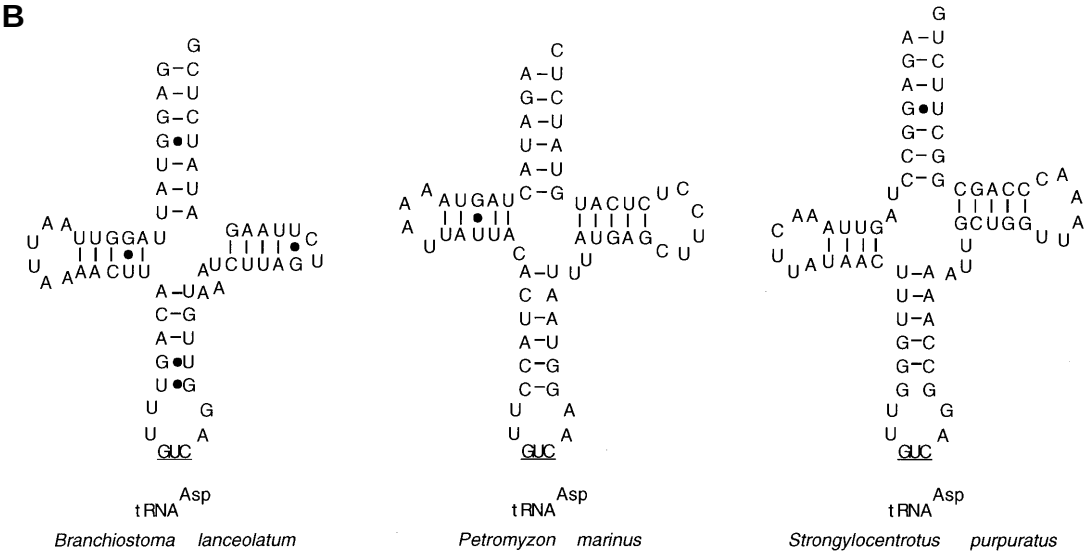

Figure 2. (**A**) Alignment of four tRNAs genes of amphioxus with their homologues in lamprey and sea urchin (*S.purpuratus*). The anticodon is underlined. Identical nucleotides are shown by points; gaps are depicted by dashes. The position of the tRNAs in the amphioxus mitochondrial genome are indicated. (**B**) comparison of the clover leaf structure of the tRNAAsp in amphioxus (left), lamprey (middle) and sea urchin (right). Classical Watson–Crick pairings are indicated by dashes whereas G–T pairings are indicated by dots.

The reasons that may explain the short size of the amphioxus mtDNA as well as the lack of recognisable DNA replication control regions are as yet unclear. A comparatively shorter size may be important to allow a rapid replication of the genome, but there is no indication that the amphioxus may need a replication of its mtDNA more rapidly than any other deuterostome. Nevertheless, the reduction in size of all the tRNAs, as well as in the two ribosomal rRNA genes, suggests that a selective pressure is acting in order to reduce the size of this genome. The lack of discernible control region is even more fascinating. The mitochondrial genome of the amphioxus obviously has to be replicated and the DNA replication machinery should find a place to initiate this process. It is tempting to speculate that the 57 bp non-coding sequence may play a role in the initiation of DNA replication, but the precise mechanisms by which this occurs remain mysterious. However, this lack of control region may be viewed as a derived character of the amphioxus, since such blocks of sequence identity have been found in sea urchin (CSB-3; 7,8,17; H.Himeno, H.Masaki, T.Kawai, T.Ohta, I.Kumagai, K.Miura and K.Watanabe, unpublished) and in lamprey (CSB-2 and CSB-3;14,37). The amphioxus may have lost these sequences during evolution. Whether this is linked to the reduction of genome size remains to be determined. It is not clear at present if the primary determinant for these structural specificities was the need to reduce the size of the genome or the need to build up a specific type of control region.

Genomic organization of the mtDNA of the lancelet

As shown in Figure 1, the amphioxus mtDNA displays the same basic gene organisation as the lamprey. Moreover, as for lamprey or dogfish, the only protein-coding gene located on the L strand is the ND6 gene. From these data, we conclude that the rearrangements that occurred between the common ancestor of all deuterostomes and the common ancestor of vertebrates were in fact completed before the split of prochordates. The amphioxus mitochondrial genome may thus be viewed as a 'proto-vertebrate' genome. In this respect the prochordates can really be viewed as the brother group to the vertebrates. This allowed us to hypothesise that the hagfish should harbour the same gene organisation as that of the amphioxus and of the lamprey, a conclusion that is supported by the sequencing of the 16S rRNA-NAD3 region of the hagfish mtDNA (C.Delarbre and G.Gachelin, unpublished). It would be of great interest in this context to determine the complete sequence of the mtDNA genome of a urochordate, in order to better monitor the events that could have arisen during the evolution of the early chordates.

The tRNA gene order observed in amphioxus is completely different from that known in the sea urchin or the sea star, two species in which most of the tRNAs are clustered in a complex of genes near the 12S rRNA gene (7,8,17; H.Himeno, H.Masaki, T.Kawai, T.Ohta, I.Kumagai, K.Miura and K.Watanabe, unpublished). Furthermore, the amphioxus mitochondrial genome contains some differences with respect to the lamprey at the level of the tRNA location. First, some tRNAs are found at different positions in the two genomes: the tRNA-Gly is located between the ND5 and ND6 genes in *Branchiostoma* whereas it is located between the CO III and ND3 genes in *Petromyzon*. The tRNA-Thr and Phe have moved from the 5′ part of a gene to its 3′ part between the two species (tRNA-Thr is located before the Cytb gene in the lamprey and after in the amphioxus; tRNA-Phe has moved on the same way around the 12S rRNA gene). Finally, in five cases (tRNA-Met, tRNA-Gln, tRNA-Trp, tRNA-Asn and tRNA-Ala), there was inversion of order inside groups of three or five tRNAs. For example in lamprey, between the ND1 and ND2 genes the tRNAs for Ile, Gln and Met are found whereas in amphioxus we observed the order Ile, Met, Gln. Cantatore *et al.* (33) suggested that the rearrangement of tRNA genes might occur by illicit priming of replication. It appears highly improbable that the movements observed in the location of the tRNA between lamprey and amphioxus may be explained by such a phenomenon. Interestingly, we never noticed a change of the strand encoding a tRNA when lamprey, dogfish or amphioxus are compared. Recently, a mechanism for the movement of tRNA genes based on errors in light strand replication and tandem duplication followed by gene loss was proposed (24,25). However, it is difficult to know whether such a model could be responsible for the change in the tRNA gene order that we observed in the lancelet mitochondrial genome, since the position of the light strand origin of replication is not known.

Features of the tRNA and codon usage

The classical 22 tRNAs were identified on the basis of their location, sequence and the ability of the coded RNAs to fold into a cloverleaf structure. As mentioned above, the tRNA of the amphioxus mtDNA are smaller overall than those of the dogfish or lamprey genomes. As shown in Table 1 and Figure 1, the majority of the tRNA genes are located on the H strand. In fact, nine of them (tRNA Pro, Gln, Asn, Ala, Cys, Tyr, Ser TCN, Glu) are located on the L strand. Interestingly, we found that a 74 bp region (positions 14385–14458) in the 3′ part of the non-coding region may form a degenerated cloverleaf structure suggesting that it could represent a twenty-third tRNA (Fig. 3). It is noteworthy that this putative tRNA contains the TCT anti-codon that is thus able to recognise the AGA codon. This finding is interesting since the AGA codon is read as a glycine in the mitochondrial genetic code of the amphioxus as is the case in tunicates, and not as a stop codon as in all vertebrates. We can thus consider that in the amphioxus mitochondrial genome two tRNA can lead to the inclusion of a glycine residue: the classical tRNA-Gly GGN and the extra tRNA-Gly AGA. The classical tRNA-Gly GGN is located just 3′ to this tRNA-Gly AGA, between positions 14 457 and 14 524, and we can speculate that a duplication of this tRNA may have created the extra tRNA-Gly AGA. It is not clear whether this extra tRNA has any function, since some conserved features in the anticodon loop (such as an uridine in the 5′ side and a purine in the 3′ side) are conserved even in divergent tRNAs lacking the T or D stem and are absent in the extra-tRNA. It is not yet known whether such an extra tRNA exists in the mitochondrial tRNA of tunicates.

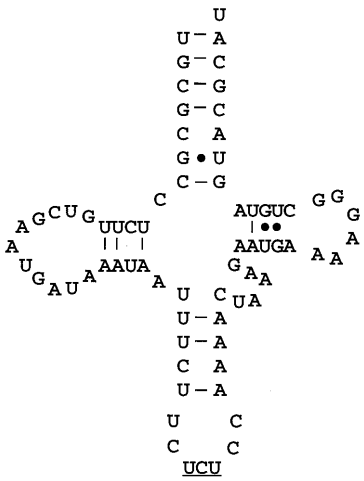

Figure 3. Structure of the extra tRNA of the amphioxus mitochondrial genome. This tRNA is encoded by the heavy strand between positions 14 385 and 14 458. It exhibits a clear but unusual cloverleaf structure and contains a TCT anticodon which could thus recognise the AGA codon read as a glycine in the amphioxus mtDNA. Classical Watson–Crick pairings are indicated by dashes whereas G–T pairings are indicated by dots.

We have compared the codon usage of *B.lanceolatum* with that of vertebrates such as *Petromyzon marinus* and *Scyliorhinus canicula* (data not shown). As mentioned above, the major difference between amphioxus and the vertebrates is that the AGA codon is not read as a stop codon (as in vertebrates), but as a glycine codon (as in tunicates). This observation has already been made by Delarbre *et al.* (31), who determined the sequence of a fragment of 2.4 kb of the amphioxus mtDNA encompassing the NAD1 and NAD2 genes. The additional reasons for that conclusion are the following. First, 12 AGAs are in frame in the mtDNA of the lancelet. Thus, they could not be used as stop codons as in the mtDNA of vertebrates. Second, the codon 13 884–13 886 in NAD5 codes for an amino acid which is a

glycine in all animals studied so far. These observations strengthen our previous conclusion that AGA specifies glycine in the lancelet, as it does in the tunicates (31,34). Interestingly enough, excepting the AGA present in the NAD2 genes, all AGAs are clustered in the NAD3 to NAD5 genes, and predominantly in the NAD4 and NAD5 genes. It is worth noting that AGG is not used by the lancelet, whereas it is used to specify glycine in the tunicates. Thus, the use of the AGA codon as a glycine codon is not a synapomorphy of the tunicates, but rather an ancestral character common to urochordates and prochordates that has been modified during the transition to vertebrates. Interestingly, this codon is apparently read as a stop codon in myxine (31), suggesting that this may represent a synapomorphy of the craniates.

We also observed clear bias in the codon usage when compared to *P.marinus* and *S.canicula* (data not shown). For example, by studying the relative frequency of the bases used at position 3 of 4-fold degenerate codons we noticed that there is a strong over-representation of the G in amphioxus (20.6%) when compared to *Petromyzon* (3.5%), *Scyliorhinus* (5.8%) or *Cyprinus* (6.7%). This over-representation of the G is detrimental mostly to the Cs that are present in only 9.2% of the 4-fold degenerate codons of amphioxus whereas it is present in 20–26% of these codons for *Petromyzon*, *Scyliorhinus* or *Cyprinus*. It is noteworthy that some codons preferentially used in early vertebrates such as the TTT codon for Phe are also predominantly used in *Branchiostoma*, whereas in *Cyprinus* a preferential use of TTC was observed. In several cases (codon for Ser AGT and AGC) the amphioxus exhibits a codon preference similar to that of lamprey. The initiation codons are ATG except ATA for ND1 and GTG for COI. The GTG codon for COI is conserved in all other chordates studied so far but not in echinoderms. Taken together, these observations suggest that at the level of the codon usage, the amphioxus mtDNA contains clear vertebrate-like characteristics, but retains some ancestral traits such as the use of the AGA codon as a Glycine codon. Again this emphasised the position of the amphioxus as a brother group to the vertebrates.

Relationship of the amphioxus mtDNA sequence with other deuterostomes

We then computed the percentage of amino acid identity between the various mtDNA coding genes of the lancelet with those of other deuterostomes (Table 3). These data clearly indicate that the lancelet is equally distant from vertebrates and from echinoderms. Indeed, the average percent identity between all the protein coding genes of lancelet and sea urchin is 51%, whereas this value is 51.4–56 % with vertebrates (52% with the lamprey, the maximal value, 56% being with the carp; see Table 3). When original genes are considered, the same picture appears. For example ND2, COI, COIII, ND4 and ND6 exhibit the same levels of identity between amphioxus and *Paracentrotus* than between amphioxus or any vertebrates. Some other proteins such as ND1, ATPase6, ND3 and Cytb appear to be a little more conserved between amphioxus and vertebrates than between amphioxus and *Paracentrotus*. Finally, the amphioxus COII appears more conserved with the sea urchin COII (62.6%) than with any other vertebrate including lamprey (60.4%). Recent phylogenetical analyses have emphasised the difficulties in reconstructing trees with distant organisms (35). Indeed, a phylogeny containing *Branchiostoma floridae* mitochondrial protein coding sequences as well as a wide range of protostome and deuterostome sequences failed to give a congruent tree. The correct topology (Fig. 4) was only revealed by the analysis of a subset of sites associated with residues important for protein folding such as charged amino acids. This analysis also revealed that aliphatic amino acids carry very few congruent phylogenetic information. Thus the fitness of a given mtDNA encoded protein to give the correct topology may be related to its richness in informative sites. This may explain why the various proteins of the lancelet mitochondrial genome do not give identical results in a phylogenetic analysis (data not shown). A detailed phylogenetic analysis will be published elsewere.

Several partial sequences of mtDNA genes of tunicates are available in the literature. For example, partial sequences of the COI gene from *Halocynthia roretzi* and from *Ascidia mentula*, and a complete sequence from COIII of *Pyura stolonifera* have been published. Interestingly, our amphioxus mtDNA sequence reveals that amphioxus is more closely related to vertebrates than to tunicates. Indeed, the partial COI protein from tunicates exhibits 70.6% identity with the lancelet protein, whereas the homologous regions of lancelet and lamprey COI share 83% amino acid identity. For COIII, the values are also very clear: 54% amino acid identity between *Pyura* and *Branchiostoma* versus 69.5% amino acid identity between *Branchiostoma* and *Petromyzon*. We also found a sequence of the 16S rRNA gene from *P.stolonifera* that exhibits 60% identity with the amphioxus gene versus 66% with the lamprey.

In conclusion, although the amphioxus mitochondrial genome exhibits some intriguing specificities such as its reduced size and the lack of obvious control region sequence signature, its organisation and the conservation of its proteins clearly confirm that the amphioxus should be viewed as the brother group of vertebrates (Fig. 4).

Table 3. Percent amino acid identities between *Branchiostoma* mitochondrial-encoded proteins with homologous proteins of other deuterostomes

Branchiostoma compared to:	Cytb	ND1	ND2	ND3	ND4	ND4L	ND5	ND6	COI	COII	COIII	ATP6	ATP8	Average
Paracentrotus	63	57	33	46	45.3	43	39	22.7	75	62.6	70	37	29.5	51
Petromyzon	64.5	61	30.5	55.5	44.5	43	39.5	22	76.7	60.4	69.4	48.5	26	52
Scyliorhinus	68.5	63	34	54	47.5	50.5	45	24.5	76.7	59.5	73.7	52	26	55
Cyprinus	70	64.5	35	57	48.5	50.5	47	25	77.7	60.4	74	51.5	29.5	56
Xenopus	69	59	32	51	45.6	32	44	19	75	58	73	46.7	31.5	53
Gallus	68	59	31	50.4	43.3	40.6	42	24	77	51.7	67.5	51	31.5	52
Didelphis	66	56	28	47	43	41.7	43	21.5	76	54	72	46	24	51.4

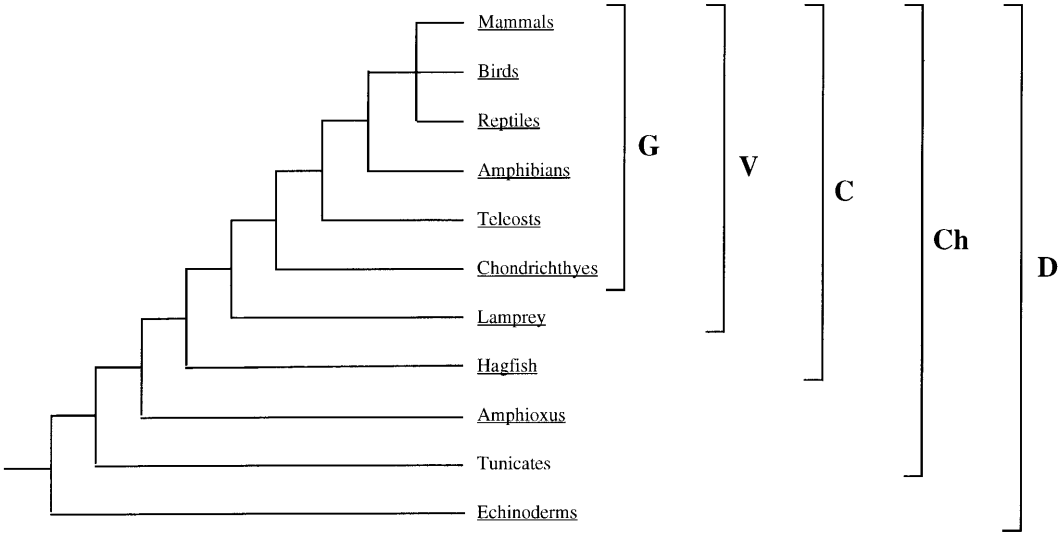

Figure 4. Simplified phylogenetic tree showing the relationships between the major deuterostome phyla. This tree represents the most consensual view of the relationships between these animals whereas the placement of some groups (e.g. tunicates) is still a matter of discussion. The hagfish is located as a sister taxa to all vertebrates (38). The relationships between the various fishes as well as between mammals/birds and the reptiles (which is not a monophyletic group) are not resolved and should be viewed only as a possible illustration. The hemichordates (acorn worms) which are deuterostomes were not included in this tree since their systematic affinities are not solved. Underlined groups are those for which at least one complete mitochondrial genome is known. G, gnathostomes; V, vertebrates; C, craniates; Ch, chordates; D, deuterostomes.

ACKNOWLEDGEMENTS

We thank Agnès Bègue for help and advice during the course of this work, Peter Holland for the generous gift of a *B.lanceolatum* genomic library, Catherine Hänni for collecting amphioxus specimens, Monique Monnerot for helpful discussion and Jean-Marc Vanacker for critical reading of the manuscript. We are grateful to anonymous reviewers for comments and suggestions. This work was financed by CNRS, ARC and Institut Pasteur de Lille.

REFERENCES

1 Boore, J.L., Collins, T.M., Stanton, D., Daehler, L.L. and Brown, W.M. (1995) *Nature*, **376**, 163–165.
2 Wolsthenholme, D.R. (1992) *Int. J. Cytol*., **141**, 173–216.
3 Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R. and Young, I.G. (1981) *Nature*, **290**, 457–465.
4 Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) *Cell*, **26**, 167–180.
5 Roe, B.A., Ma, D.P., Wilson, R.K. and Wong, J.F. (1985) *J. Biol. Chem*., **260**, 9759–9774.
6 Clary, D.O. and Wolstenholme, D.R. (1985) *J. Mol. Evol*., **22**, 252–271.
7 Jacobs, H.T., Elliott, D.J., Math, V.B. and Farquharson, A. (1988) *J. Mol. Biol*., **202**, 185–217.
8 Cantatore, P., Roberti, M., Rainaldi, G., Gadaleta, M.N. and Saccone, C. (1989) *J. Biol. Chem*., **264**, 10965–10975.
9 Desjardins, P. and Morais, R. (1990) *J. Mol. Biol*., **212**, 599–634.
10 Arnason, U., Gullberg, A. and Widegren, B. (1991) *J. Mol. Evol*., **33**, 556–568.
11 Okimoto, R., Macfarlane, J.L., Clary, D.O. and Wolstenholme, D.R. (1992) *Genetics*, **130**, 471–498.
12 Tzeng, C.S., Hui, C.F., Shen, S.C. and Huang, P.C. (1992) *Nucleic Acids Res*., **20**, 4853–4858.
13 Chang, Y.S., Huang, F.L. and Lo, T.B. (1994) *J. Mol. Evol*., **38**, 138–155.
14 Lee, W.J. and Kocher, T.D. (1995) *Genetics*, **139**, 873–887.
15 Zardoya, R., Garrido-Pertierra, A. and Bautista, J.M. (1995) *J. Mol. Evol*., **41**, 942–951.
16 Zardoya, R. and Meyer, A. (1997) *Genetics*, **146**, 995–1010.
17 De Giorgi, C., Martiradonna, A., Lanave, C. and Saccone, C. (1996) *Mol. Phyl. Evol*., **5**, 323–332.
18 Johansen, S. and Bakke, I. (1996) *Mol. Mar. Biol. Biotechnol*., **5**, 203–214.
19 Janke, A., Gemmell, N.J., Feldmaier-Fuchs, G., von Haeseler, A. and PSSbo, S. (1996) *J. Mol. Evol*., **42**, 153–159.
20 Noack, K., Zardoya, R. and Meyer, A (1996) *Genetics*, **144**, 1165–1180.
21 Delarbre, C., Spruyt, N., Delmarre, C., Janvier, P., Laudet, V. and Gachelin, G. (1998) *Genetics*, in press.
22 Pääbo, S., Thomas, W.K., Whitfield, K.M., Kumazawa, Y. and Wilson, A.C. (1991) *J. Mol. Evol*., **33**, 426–430.
23 Kumazawa, Y., Ota, H., Nishida, M. and Ozawa, T. (1996) *Mol. Bio. Evol*., **13**, 1242–1254.
24 Macey, J.R., Larson, A., Ananjeva, N.B. and Papenfuss, T.J. (1997) *J. Mol. Evol*., **44**, 660–674.
25 Macey, J.R., Larson, A., Ananjeva, N.B., Fang, Z. and Papenfuss, T.J. (1997) *Mol. Biol. Evol*., **14**, 91–104.
26 Jacobs, H.T., Asakawa, S., Araki, T., Miura, K.I., Smith, M.J. and Watanabe, K. (1989) *Curr. Genet*., **15**, 193–206.
27 Smith, M.J., Arndt, A., Gorski, S. and Fajber, E. (1993) *J. Mol. Evol*., **36**, 545–554.
28 Gans, C., Kemp, N. and Poss, S. (1996) *Israel J. Zool*., **42**, Suppl. S1–S446.
29 Holland, P.W. and Garcia-Fernandez, J. (1996) *Dev. Biol*., **173**, 382–395.
30 Garcia-Fernandez, J. and Holland, P.W. (1994) *Nature*, **370**, 563–566.
31 Delarbre, C., Barriel, V., Tillier, S., Janvier, P. and Gachelin, G. (1997) *Mol. Biol. Evol*., **14**, 807–813.
32 Lee, W.J., Conroy, J., Howell, W.H. and Kocher, T.D. (1995) *J. Mol. Evol*., **41**, 54–66.
33 Cantatore, P. Gadaleta, M.N., Roberti, M., Saccone, C. and Wilson, A.C. (1987) *Nature*, **329**, 853–855.
34 Yokobori, S.I., Ueda, T. and Watanabe, K. (1993) *J. Mol. Evol*., **36**, 1–8.
35 Naylor, G.J.P. and Brown, W.M. (1997) *Nature*, **388**, 527–528.
36 Escriva, H., Safi, R., HSnni, C., Langlois, M.C., Saumitou-Laprade, P., Stéhelin, D., Capron A., Pierce, R. and Laudet, V. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 6803–6808.
37 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, NY.
38 Rasmussen, A.S., Janke, A. and Arnason U. (1998) *J. Mol. Evol*., **46**, 382–388.