Time Scale for Cyclostome Evolution Inferred with a Phylogenetic Diagnosis of Hagfish and Lamprey cDNA Sequences

Authors: Shigehiro Kuraku, and Shigeru Kuratani
Source: Zoological Science, 23(12) : 1053-1064
Published By: Zoological Society of Japan
URL: https://doi.org/10.2108/zsj.23.1053
Time Scale for Cyclostome Evolution Inferred with a Phylogenetic Diagnosis of Hagfish and Lamprey cDNA Sequences

Shigehiro Kuraku* and Shigeru Kuratani

Laboratory for Evolutionary Morphology, RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minami, Chuo-ku, Kobe 650-0047, Japan

The Cyclostomata consists of the two orders Myxiniformes (hagfishes) and Petromyzoniformes (lampreys), and its monophyly has been unequivocally supported by recent molecular phylogenetic studies. Under this updated vertebrate phylogeny, we performed in silico evolutionary analyses using currently available cDNA sequences of cyclostomes. We first calculated the GC-content at four-fold degenerate sites (GC_{4}), which revealed that an extremely high GC-content is shared by all the lamprey species we surveyed, whereas no striking pattern in GC-content was observed in any of the hagfish species surveyed. We then estimated the timing of diversification in cyclostome evolution using nucleotide and amino acid sequences. We obtained divergence times of 470–390 million years ago (Mya) in the Ordovician–Silurian–Devonian Periods for the interordinal split between Myxiniformes and Petromyzoniformes; 90–60 Mya in the Cretaceous–Tertiary Periods for the split between the two hagfish subfamilies, Myxininae and Eptatretinae; 280–220 Mya in the Permian–Triassic Periods for the split between the two lamprey subfamilies, Geotriinae and Petromyzoninae; and 30–10 Mya in the Tertiary Period for the split between the two lamprey genera, Petromyzon and Lethenteron. This evolutionary configuration indicates that Myxiniformes and Petromyzoniformes diverged shortly after the common ancestor of cyclostomes split from the future gnathostome lineage. Our results also suggest that intra-subfamilial diversification in hagfish and lamprey lineages (especially those distributed in the northern hemisphere) occurred in the Cretaceous or Tertiary Periods.

Key words: cyclostome, hagfish, lamprey, GC-content, synonymous substitution, molecular clock

INTRODUCTION

Extant agnathans, the cyclostomes, comprise the hagfishes (Hyperotreti; order Myxiniformes) and lampreys (Hyperoarida; order Petromyzoniformes [often misspelled as “Petromyzontiformes”]) (Hardisty and Potter, 1971; Forey and Janvier, 1993; Jørgensen, 1998; see also Ota and Kuratani, 2006) (Fig. 1). After a long-standing controversy on the phylogenetic positions of hagfishes and lampreys, the monophyly of cyclostomes has been unequivocally supported by molecular phylogenetics using a triad of molecules frequently used for reconstruction of species phylogeny, namely, mitochondrial genes (mtDNA), nuclear ribosomal RNA genes (rDNA), and nuclear protein-coding genes (nuDNA) (Fig. 1; Stock and Whitt, 1992; Mallatt and Sullivan, 1998; Kuraku et al., 1999; Delarbre et al., 2002; Furlong and Holland, 2002; Takezaki et al., 2003; Blair and Hedges, 2005; Delsuc et al., 2006). Therefore, our interest in cyclostome evolution has shifted to the topological and temporal aspects of divergence patterns within this animal group.

* Corresponding author. Phone: +81-78-306-3064; Fax : +81-78-306-3370; E-mail: venezia@cdb.riken.jp
doi:10.2108/zjs.23.1053

The order Myxiniformes is thought to be monophyletic, based on molecular phylogenetic studies using mitochondrial 16S rDNA (Kuo et al., 2003; Chen et al., 2005). This order is divided into the two subfamilies Myxininae and Eptatretinae, based on morphological features (Fig. 1; Fernholm, 1998). The subfamily Myxininae consists of four genera, Myxine and three genera intrinsic to the southern hemisphere (Neomyxine, Nemamyxine, and Notomyxine). The other subfamily, Eptatretinae, consists of three genera, Eptatretus, Paramyxine, and Rubicundus. In contrast, the order Petromyzoniformes is composed of three subfamilies, Mordaciinae, Geotriinae, and Petromyzoninae, in accordance with morphology such as dentition (Hubbs and Potter, 1971; Gill et al., 2003). The subfamilies Mordaciinae and Geotriinae are endemic to the southern hemisphere, and each comprises a single genus, Mordacia and Geotria, respectively (Potter and Strahan, 1968). The subfamily Petromyzoninae is composed of at least six genera (Fig. 1; Hardisty and Potter, 1971; Hubbs and Potter, 1971; Potter and Gill, 2003). However, there are few detailed reports of molecular approaches to estimate divergence times in the cyclostome lineage. To address questions regarding the temporal pattern of cyclostome evolution, the accumulating nucleotide and amino acid sequences of hagfishes and lampreys will provide novel information.
In this study, we analyzed the GC-content in cDNA sequences of hagfishes and lampreys and calculated the divergence times of several branching points in cyclostome phylogeny using nucleotide and amino acid sequences, based on an updated version of vertebrate phylogeny representing the monophyly of cyclostomes.

**MATERIALS AND METHODS**

**GC$_4$ calculation**

Currently available annotated nucleotide sequences (as of February 15, 2006) were retrieved for each cyclostome species from NCBI Entrez Nucleotide (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide). Redundant sequences were manually removed. To avoid biased gene selection, in which a large proportion of the sequence population for one species is occupied by members of a limited number of gene families, cDNAs derived from variable leukocyte receptor genes in the sea lamprey, _Petromyzon marinus_ (accession numbers CK988414-CK988652 in NCBI dbEST; Pancer et al., 2004a), and the inshore hagfish, _Eptatretus burgeri_ (accession numbers AY964719-AY965612; Pancer et al., 2005), were excluded from our sequence collection. The nucleotide sequences were used to calculate the GC-content at four-fold degenerate sites (GC$_4$) with the Perl script, in which an open reading frame is automatically detected with an alignment generated by BLASTX (Altschul et al., 1997). Sequences of mtDNA and nuclear rRNA genes were excluded from this GC calculation.

**Molecular phylogenetic tree inference**

Sequences that showed significant similarity to a query in a BLASTP search (Altschul et al., 1997) were retrieved from databases: GenBank (release 151), NCBI-refseq (release 06-02-16), SWISSPROT (release 49.0), and PIR (release 80.0). An optimal multiple alignment of these amino acid sequences was constructed using the alignment editor XCED implemented in the MAFFT program (Katoh et al., 2002) in combination with manual inspection. Molecular phylogenetic trees were inferred by the neighbor-joining method (Saitou and Nei, 1987) with XED and the maximum-likelihood method (Felsenstein, 1981; Kishino et al., 1990) with PAML 3.1 (Yang, 1997), using amino acid sites at which the alignment was

---

**Fig. 1.** Taxonomy and molecular phylogeny of cyclostomes. Taxonomic information was based on NCBI Entrez Taxonomy, which is mainly consistent with the classification by Renaud (1997) and Fernholm (1998). Taxonomic ranks are indicated in the top row. References that suggest particular phylogenetic relationships are listed near the corresponding nodes in gray shades, with the type of molecule (nuDNA, rDNA, or mtDNA) employed in their analyses. Note that branch lengths do not correspond to evolutionary times. The numbers of nucleotide and amino acid sequences, respectively, found in NCBI Entrez Nucleotide and NCBI Entrez Protein (as of February 15, 2006) are shown in parentheses beside genus names. The hagfish genera _Eptatretus_ and _Paramyxine_ are not monophyletic to each other according to the molecular phylogenetic tree of the 16S rRNA gene (Kuo et al., 2003; Chen et al., 2005). The lamprey genera _Entosphenus_, _Lamptera_, and _Lethenteron_ are not monophyletic to one another according to the molecular phylogenetic tree of the cytochrome b and NADH dehydrogenase 3 genes (Docker et al., 1999; see DISCUSSION for details).
unambiguous with no gaps, with among-site rate heterogeneity taken into account (Yang, 1994).

Estimation of number of synonymous and nonsynonymous substitutions per site

Nucleotide sequences were prepared as described above in the procedures for GC content calculation. By inferring molecular phylogenetic trees, we selected genes with a homologue present as a single orthologue in a pair of species in question. Nucleotide sequences of the selected genes were aligned based on an alignment generated for their deduced amino acid sequences. The number of synonymous and nonsynonymous substitutions per site ($K_s$ and $K_a$, respectively) was calculated with the codon-based maximum-likelihood method (Goldman and Yang, 1994). Computation was accomplished using PAML 3.1 (Yang, 1997).

Amino acid sequence-based divergence time estimation

Estimation of divergence times was processed without assuming a global molecular clock, using the MULTIDIVTIME program in which Markov-chain Monte-Carlo (MCMC) procedures for Bayesian analysis are implemented (Kishino et al., 2001). The upper and lower limits of divergence times outside the cyclostomes were preset by referring to a set of fossil records (Young, 1962) used by Dickerson (1971), or by referring to molecular dating (Kumar and Hedges, 1998; Blair and Hedges, 2005). To estimate divergence times using mitochondrial genes, we used a modified version of the MULTIDIVTIME program as instructed on the developers’ web page (http://statgen.ncsu.edu/thorne/multidivtime.html). Results were confirmed with the program R8S, which enables penalized rate smoothing (data not shown; Sanderson, 2002, 2003).

RESULTS

GC-content in cyclostome cDNAs

GC$_s$ was calculated for cDNA sequences derived from nuclear protein-coding genes for each cyclostome species. The GC$_s$ of annotated cDNAs exhibited a unimodal distribution, with peaks at 40–60% in hagfish species (Figs. 2A–C) and at 70–90% in lamprey species (Figs. 2D–F). Non-annotated abundant cDNAs of Eptatretus burgeri (Suzuki et al., 2004) and Petromyzon marinus (Pancer et al., 2004b) showed similar GC$_s$ distributions to those of annotated cDNAs for hagfish and lamprey species, respectively (Fig. 2G). The results of the GC$_s$ calculation for genera or species with a small number of available cDNAs were as follows: Paramyxine, 41–61% (n=5); Ichthyomyzon, 73–87% (n=7); Entosphenus, 75–83% (n=2); Mordacia mordax, 72–82% (n=3); Geotria australis, 73–90% (n=9).

Estimated number of synonymous substitutions

We selected genes that existed as single orthologues in a pair of species in question with more than 600 bp of aligned nucleotide stretches. Estimation of $K_s$ was processed using the maximum-likelihood method (Goldman and Yang, 1994). The average $K_s$ between the two hagfish genera, Myxine and Eptatretus, was 0.24 (standard deviation (SD), 0.11; n=11; Table 1). In contrast, the average $K_s$ for Petromyzon–Lethenteron and Geotria–Lethenteron pairs was 0.15 (SD, 0.09; n=19; Table 2A) and 1.03 (SD, 0.39; n=2; Table 2B), respectively. Sequence comparison between a hagfish species and a lamprey species always yielded an apparently saturated $K_s$ (>3; data not shown).

We also estimated the $K_s$ of mitochondrial protein-coding genes in the Myxine–Eptatretus and Petromyzon–Lam-
Table 1. Estimated numbers of synonymous and non-synonymous substitutions between Myxininae and Eptatretinae.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Species</th>
<th>Aligned length (nt)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>blood 5-aminolevulinate synthase</td>
<td>Mg</td>
<td>603</td>
<td>0.20</td>
<td>0.01</td>
</tr>
<tr>
<td>creatine kinase</td>
<td>Eb</td>
<td>831</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>eukaryotic translation elongation factor 1 gamma</td>
<td>Mg</td>
<td>1137</td>
<td>0.30</td>
<td>0.05</td>
</tr>
<tr>
<td>eukaryotic elongation factor-1 alpha</td>
<td>Mg</td>
<td>630</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>fructose-bisphosphate aldolase, muscle type (EJM8)</td>
<td>Mg</td>
<td>900</td>
<td>0.20</td>
<td>0.02</td>
</tr>
<tr>
<td>guanine nucleotide-binding protein, beta polypeptide 2-like 1</td>
<td>Mg</td>
<td>1092</td>
<td>0.24</td>
<td>0.01</td>
</tr>
<tr>
<td>eukaryotic translation elongation factor 1 gamma</td>
<td>Mg</td>
<td>819</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>ribosomal protein S2</td>
<td>Mg</td>
<td>603</td>
<td>0.31</td>
<td>0.00</td>
</tr>
<tr>
<td>ribosomal protein S3</td>
<td>Mg</td>
<td>633</td>
<td>0.23</td>
<td>0.00</td>
</tr>
<tr>
<td>ribosomal protein S4</td>
<td>Mg</td>
<td>714</td>
<td>0.49</td>
<td>0.03</td>
</tr>
<tr>
<td>ribosomal protein L5</td>
<td>Mg</td>
<td>888</td>
<td>0.23</td>
<td>0.04</td>
</tr>
<tr>
<td>ribosomal protein L7a</td>
<td>Mg</td>
<td>723</td>
<td>0.26</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Species names are indicated as abbreviations: Mg, Myxine glutinosa; Eb, Eptatretus burgeri; Es, Eptatretus stoutii.

Table 2. Estimated numbers of synonymous and non-synonymous substitutions between lamprey species.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Species</th>
<th>Aligned length (nt)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>bone morphogenetic protein (BMP) 2/4a</td>
<td>Pm</td>
<td>831</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>cytoplasmic actin (LCA1)</td>
<td>Lj</td>
<td>1128</td>
<td>0.22</td>
<td>0.00</td>
</tr>
<tr>
<td>Dlx1/6 (DlxD)</td>
<td>Lj</td>
<td>804</td>
<td>0.14</td>
<td>0.00</td>
</tr>
<tr>
<td>enolase-2</td>
<td>Lj</td>
<td>1165</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>eukaryotic elongation factor-1 alpha</td>
<td>Lj</td>
<td>1389</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>fructose-bisphosphate aldolase, muscle type (EJM8)</td>
<td>Lj</td>
<td>1068</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>guanine nucleotide binding protein, beta polypeptide 2-like 1</td>
<td>Lj</td>
<td>951</td>
<td>0.12</td>
<td>0.00</td>
</tr>
<tr>
<td>HMG 1/2/3</td>
<td>Lj</td>
<td>624</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>ikaros-like transcription factor IKLF1</td>
<td>Lj</td>
<td>1197</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>intermediate filament protein type III</td>
<td>Lj</td>
<td>1419</td>
<td>0.26</td>
<td>0.01</td>
</tr>
<tr>
<td>low molecular mass polypeptide (LMPX)</td>
<td>Lj</td>
<td>615</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>myosin heavy chain (LMHC3)</td>
<td>Lj</td>
<td>786</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Pax2/5/8</td>
<td>Lj</td>
<td>834</td>
<td>0.24</td>
<td>0.00</td>
</tr>
<tr>
<td>phosphoglycerate kinase (PGK)</td>
<td>Lj</td>
<td>828</td>
<td>0.22</td>
<td>0.00</td>
</tr>
<tr>
<td>ribosomal protein S2</td>
<td>Lj</td>
<td>603</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>ribosomal protein L7a</td>
<td>Lj</td>
<td>672</td>
<td>0.36</td>
<td>0.01</td>
</tr>
<tr>
<td>SH3-domain GRB2-like (endophilin SH3p4)</td>
<td>Lj</td>
<td>627</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>TATA-box binding protein</td>
<td>Lj</td>
<td>867</td>
<td>0.09</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 3. Divergence times and numbers of synonymous substitutions reported for closely related organism pairs.

<table>
<thead>
<tr>
<th>Taxonomic group</th>
<th>Species pairs</th>
<th>Ks</th>
<th>Ks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammalia</td>
<td>1, human-mouse</td>
<td>0.56</td>
<td>ML</td>
</tr>
<tr>
<td></td>
<td>2, mouse-rat</td>
<td>0.19</td>
<td>ML</td>
</tr>
<tr>
<td></td>
<td>3, human-pig</td>
<td>0.31</td>
<td>ML</td>
</tr>
<tr>
<td>Amniota</td>
<td>4, human-chicken</td>
<td>1.66</td>
<td>ML</td>
</tr>
<tr>
<td>Sauropoda</td>
<td>5, chicken-turkey</td>
<td>0.12</td>
<td>ML</td>
</tr>
<tr>
<td></td>
<td>6, chicken-turtle</td>
<td>0.96</td>
<td>ML</td>
</tr>
<tr>
<td>Teleostei</td>
<td>7, Tetradon nigroviridis-Fugu rubripes</td>
<td>0.42</td>
<td>PBL</td>
</tr>
<tr>
<td>Cephalochordata</td>
<td>8, Branchiostoma belcheri-B. florbidae</td>
<td>0.38</td>
<td>ML</td>
</tr>
</tbody>
</table>

ML, Calculated with the method of Goldman and Yang (1994); PBL, Calculated with the method by Pamilo and Bianchi (1993) and Li (1993). ICGSC, International Chicken Genome Sequencing Consortium. RGSPC, Rat Genome Sequencing Project Consortium. Numbers 1–8 in the column of species pairs correspond to those in Fig. 3. See Supplemental Table S2.
and nuclear genes (7.9-fold for the Petromyzon–Lampetra pair) was roughly consistent with previous observations in mammals and amphibians (Miyata et al., 1982; Crawford, 2003).

Rough estimation of divergence times based on number of synonymous substitutions

The number of synonymous substitutions per site has been reported for some pairs of chordate species (Table 3). In the present study, to supplement pre-existing data, we preliminary estimated the $K_s$ between two species of the cephalochordate genus Branchiostoma (Supplemental Table S2, http://dx.doi.org/10.2108/zsj.23.1053). The divergence times and $K_s$ values in Table 3, including the data for Branchiostoma, were plotted two-dimensionally in Fig. 3. The overall rate of this putative clock was $2.4 \times 10^{-9}$/site/year (Clock A; Fig. 3A), whereas the rate of the clock for selected species pairs (for details, see legend for Fig. 3) was $1.9 \times 10^{-9}$/site/year (Clock B; Fig. 3B). By applying these clocks tentatively to cyclostome taxon pairs, we obtained divergence times of the inter-subfamilial split between Myxininae and Eptatretinae in the hagfish lineage at 93–28 Mya, the inter-generic split between Petromyzon and Lethenteron at 57–15 Mya, and the inter-subfamilial split between Geotriinae and Petromyzoninae in the lamprey lineage at 383–136 Mya (Table 4).

Estimation of divergence times using amino acid sequences

We selected 10 nuclear protein-coding genes with relatively long alignment lengths (>150 amino acids; total length, 2947 amino acids; Table 5) in which no gene duplication was detected in major vertebrate lineages, as shown in Fig. 4 for the gene GNB2L1. The upper and lower limits of divergence times for branching points outside the cyclostomes were preset as shown in Table 6, and the tree topology shown in Fig. 5 was assumed. By executing the MULTIDIVTIME program (Kishino et al., 2001), we obtained divergence times for Myxiniformes and Petromyzoniformes at 671–391 Mya, Myxininae and Eptatretinae at 162–63 Mya, and Petromyzon and Lethenteron at 30–2 Mya (Table 7).

In addition, the timing of the above branching points was estimated using 12 mitochondrial protein-coding genes that had relatively long alignment lengths (total length, 3320 amino acids). ATP synthase F0 subunit 8 was excluded from this analysis because of its short alignment length. As a result, divergence times were estimated to be 728–459 Mya for the Myxiniformes–Petromyzoniformes split, 72–39 Mya for the Myxininae–Eptatretinae split, and 37–18 Mya for the Petromyzon–Lampetra split (Table 7).

Table 4. Divergence times estimated with tentative synonymous substitution clocks.

<table>
<thead>
<tr>
<th>Node (Taxon pair)</th>
<th>e (Myxininae-Eptatretinae)</th>
<th>f (Petromyzon-Lethenteron)</th>
<th>g (Geotriinae-Petromyzoninae)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_s$ [0.24±0.11 (n=11)]</td>
<td>$K_s$ [0.14±0.07 (n=19)]</td>
<td>$K_s$ [1.03±0.39 (n=2)]</td>
</tr>
</tbody>
</table>

| Clock          | A: 51±23 | B: 65±28 | A: 30±15 | B: 38±19 | A: 219±83 | B: 278±105 |

Clocks A and B correspond to those in Fig. 3. Nodes e-g correspond to those in Fig. 5.

Fig. 3. Divergence times and number of synonymous substitutions per site. (A) The tentative clock calibrated by all of the divergences of species pairs (black circles) in Table 3. (B) The tentative clock calibrated by nodes 3, 5, 6, and 8 in Table 3. In this "Clock B", nodes 1 and 2 were excluded because of an elevated evolutionary rate in the rodent lineage. Node 4 was excluded because of its unreliable high $K_s$ value (>1). Node 7 was excluded because $K_s$ for this species pair was estimated with a different method from the others. In both clocks, regression lines indicated as broken lines pass through the origin. For nodes 6 and 7, the intermediate values of estimated ranges of divergence times in Table 3 were used as divergence times in this Fig. Divergences in cyclostomata (e–g) that were dated with these clocks are indicated as open circles.
The GC-content at synonymous sites in a protein-coding gene is expected to positively correlate with the global GC-content of the genomic region where the gene is located (Clay et al., 1996; Musto et al., 1999; Kuraku et al., 2006). Therefore, we focused on GC\(_4\) in currently available cDNAs reported for cyclostomes (Fig. 2). Hagfish and lamprey cDNAs show similar levels of heterogeneity in GC-content (Fig. 2). However, there is a striking difference in the level of GC-content between hagfish and lamprey: every lamprey species we analyzed showed a high GC\(_4\) (70–90%; Figs. 2A–C), whereas every hagfish species we analyzed showed a relatively moderate GC\(_4\) (40–60%; Figs. 2D–F).

Cytogenetic studies have revealed that hagfishes have a relatively moderate number of relatively moderate-sized chromosomes compared to other vertebrates (2n=14–36 in somatic cells), whereas lampreys possess a much greater number of small, dot-like chromosomes (2n=76–178; Potter and Rothwell, 1970; Potter and Robinson, 1971; Robinson et al., 1975; Nakai et al., 1995; Animal Genome Size Database, http://www.genomesize.com). The contrast in chromosome size, chromosome number, and GC-content between hagfishes and lampreys is reminiscent of the intra-genomic difference between macrochromosomes and microchromosomes seen in sauropsids (Burt, 2002; Kuraku et al., 2006).

**DISCUSSION**

**Table 5.** Genes on nuDNA used for amino acid-based estimation of divergence times.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene symbols of a human ortholog</th>
<th>Species(^a)</th>
<th>Aligned length (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLR2A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POLR3A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EEF1G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GNB2L1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP5A1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPS2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPS3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPL5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPL7A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPI1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Species names: Mg, Myxine glutinosa; Eb, Eptatretus burgeri; Pm, Petromyzon marinus; Lj, Lethenteron japonicum; Lr, Lethenteron reissneri.

**Table 6.** Divergence times for constrained nodes.

<table>
<thead>
<tr>
<th>Node</th>
<th>Constraint set</th>
<th>Min</th>
<th>Max</th>
<th>Min</th>
<th>Max</th>
<th>Min</th>
<th>Max</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
<td>300</td>
<td>300</td>
<td>400</td>
<td>400</td>
<td>500</td>
<td>500</td>
<td>Dickerson, 1971</td>
</tr>
<tr>
<td>II</td>
<td></td>
<td>310</td>
<td>310</td>
<td>415</td>
<td>485</td>
<td>489</td>
<td>639</td>
<td>Kumar and Hedges, 1998</td>
</tr>
<tr>
<td>III</td>
<td></td>
<td>310</td>
<td>370</td>
<td>425</td>
<td>495</td>
<td>605</td>
<td>742</td>
<td>Blair and Hedges, 2005</td>
</tr>
</tbody>
</table>

Max, Upper limit of divergence time. Min, lower limit of divergence time. Letters indicating nodes (a–c) correspond to those in Fig. 5.

**Fig. 4.** Molecular phylogenetic tree of a guanine nucleotide-binding protein, beta polypeptide 2-like 1 (GNB2L1) protein. The tree was inferred with the neighbor-joining method using 300 amino acid residues and assuming among-site rate heterogeneity (shape parameter \(\alpha=0.64\); see MATERIALS AND METHODS for details). The maximum-likelihood method yielded similar tree topologies, although they are not statistically significant. For amino acid-based divergence time estimation, we selected 10 genes with no gene duplication, as shown in this tree.
Further investigation will be required to understand the putative relationships among these genomic features.

**Ks as a tool to standardize evolutionary distances**

The number of synonymous substitutions per site in a protein-coding region serves as an ideal standard for evolutionary distance when comparing closely related species, as long as the gene in question has evolved in a neutral manner (Miyata and Yasunaga, 1980; Perler et al., 1980). In this study, we estimated \( K_s \) with the maximum-likelihood method (Goldman and Yang, 1994) because this method is expected to produce relatively appropriate estimates, even in species with highly biased base compositions, such as lampreys (Fig. 2). For all the genes analyzed in this study, \( K_s \) was larger than \( K_a \), indicating that these genes have evolved neutrally without experiencing positive selection. In selecting pairs of cyclostome species for \( K_s \) estimation, we treated the lamprey genus *Lampetra* as having the same distance the genus *Lethenteron* has from *Petromyzon*. This is based on a previous phylogenetic study using the cytochrome \( b \) and NADH dehydrogenase subunit 3 (ND3) genes (Docker et al., 1999), which is consistent with the classification by Potter (1980), who formerly considered that the three subgenera, *Entosphenus*, *Lethenteron*, and *Lampetra*, compose the single genus *Lampetra*. Despite the small number of genes sampled, similar \( K_s \) values were consistently obtained for *Petromyzon–Lampetra* and *Petromyzon–Lethenteron* pairs (Table 2). In contrast to inter-generic and inter-subfamilial comparisons in cyclostomes, our preliminary \( K_s \) estimates between two tunicate species in the same genus (*Ciona intestinalis* and *C. savignyi*) and two amphioxus species in the same genus (*Branchiostoma belcheri* and *B. floridæ*) yielded relatively larger \( K_s \) values (see Supplemental Table S2, http://dx.doi.org/10.2108/zsj.23.1053 for *Branchiostoma*; \( K_s > 3 \) for *Ciona*; data not shown), despite their close taxonomic distances.

Theoretically, the \( K_s \) value is determined by the time that has elapsed since the divergence of two species in question, as long as neutrality holds. This feature of \( K_s \) is useful in judging orthology between genes (especially members of a gene family prone to gene duplications) of closely related species. If this idea is tentatively applied to previous cross-species comparisons in lamprey studies, for example, orthology between *L. japonicum* Hox6w and *L. fluvialis* HoxL6 is again confirmed with an extremely low \( K_s \) value (\( K_s = 0.039 \)), a reasonable estimate for intrasubfamilial comparison, as suggested previously (Takio et al., 2004).

**Methodological aspects of molecular dating**

Proposing a constant rate of nucleotide and amino acid substitutions was one of the milestones for molecular dating of divergence times ("molecular clock"); Zuckerkandl and Pauling, 1962, 1965; also see Donoghue et al., 2003; Bromham and Penny, 2003; and Kumar, 2005, for review). As shown in Fig. 3, the numbers of synonymous substitutions per site and divergence times behave in a clock-like manner, at least within the chordates, indicating that this may serve as a rough molecular clock for species pairs with an unsaturated \( K_s \). However, this clock was calibrated by branching points in different lineages outside the cyclostomes (Table 3), because no branching point with a known divergence time was available within cyclostomes. In addition, variation of evolutionary rate among lineages, such as rate elevation in rodents (Kikuno et al., 1985; Wu and Li, 1985; Rat Genome Sequencing Project Consortium, 2004), may confuse divergence time estimation. Therefore, our results need to be verified by refinement of this silent clock with reexamination of divergence times and estimation of \( K_s \) for more pairs of organisms.

We utilized amino acid sequences to obtain more robust estimates. Especially in our \( K_s \) analysis, synonymous substitutions between Myxiniformes and Petromyzoniformes were apparently saturated (data not shown), suggesting that this split would need to be dated with amino acid sequences rather than nucleotide sequences. However, as exemplified in Fig. 4, phylogenetic trees including cyclostome species often show a high degree of rate heterogeneity because of the accelerated evolutionary rate in hagfishes. To minimize the undesirable influence of this rate heterogeneity on divergence time estimation, we employed a non-parametric

**Table 7.** Divergence times estimated with amino acid sequences.

<table>
<thead>
<tr>
<th>Molecule type</th>
<th>Genes on nuDNA (n=10)</th>
<th>Genes on mtDNA (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Node c d e f</td>
<td></td>
</tr>
<tr>
<td>Constraint set I</td>
<td>500± 0 432±41 88±25 11± 9</td>
<td>500± 0 478±19 44±5 21±3</td>
</tr>
<tr>
<td>Constraint set II</td>
<td>629± 9 524±53 107±30 14±11</td>
<td>629± 8 602±24 55±6 27±4</td>
</tr>
<tr>
<td>Constraint set III</td>
<td>730±11 612±59 127±35 17±13</td>
<td>725±15 698±30 65±7 32±5</td>
</tr>
</tbody>
</table>

Letters indicating nodes (c–f) correspond to those in Fig. 5. For details of constraint sets (I–III), see Table 5. Divergence times are shown with standard errors representing 95% confidence intervals.
molecular dating method that does not assume rate constancy (Kishino et al., 2001; Thorne and Kishino, 2002; see also Hasegawa et al., 2003). Moreover, we paid close attention to the orthologous/paralogous relationships of multiple members of gene families between hagfishes, lampreys, and gnathostomes, because putative genome duplications in early vertebrate evolution (Ohno, 1970; McLysaght et al., 2002) often confuse orthology identification; Kuraku et al. (1999) provided an example in the enolase gene family. In addition, inclusion of genes prone to gene duplications may also result in misleading estimates of divergence time, possibly because of an accelerated evolutionary rate caused by neofunctionalization or subfunctionalization of duplicates. For these reasons, we deliberately selected genes for which no gene duplication was detected in major vertebrate lineages (Table 5).

As calibration points outside the cyclostomes, we used three sets of divergence time constraints (Table 6). One of the three constraint sets was based on the fossil records used by Dickerson (1971) (constraint set I), whereas the other two were based on previous studies using molecular data (constraint set II, Kumar and Hedges, 1998; constraint set III, Blair and Hedges, 2005). However, these molecular studies included genes that underwent duplication events early in vertebrate evolution and do not represent a 1:1 relationship between a cyclostome gene and a gnathostome counterpart (e.g., bone morphogenetic protein (BMP) 2/4, enolase-2). Although fossil records inherently tend to yield more recent divergence times because of potential incomplete fossil sampling in more ancient eras, it is possible that inappropriate gene selection might have yielded much more ancient estimates in these studies (constraint sets II and III) compared with commonly accepted fossil records (constraint set I). This discrepancy emphasizes again that precision in gene selection cannot be sacrificed, even in the name of high-throughput analysis using large data sets. For this reason, we summarize our results below, along a temporal axis based on the fossil records.

Temporal reconstruction of cyclostome phylogeny

The monophyly of cyclostomes has resulted in a dispute over when hagfishes and lampreys split from each other in the cyclostome lineage. To answer this question, our analyses using mitochondrial and nuclear genes consistently showed that Myxiniformes and Petromyzoniformes diverged from each other 30–110 million years after the cyclostome lineage split from the future gnathostome lineage (Table 7). When calibration by fossil records was applied, this Myxiniformes–Petromyzoniformes split dated back to 470–390 Mya in the Ordovician–Silurian–Devonian Periods (Fig. 6), when fossil agnathans are thought to have diversified (Foley and Janvier, 1993; Janvier, 1996). Although we still do not know the precise branching pattern among these agnathans, hagfishes and lampreys represent two distinct agnathan groups that diverged early in vertebrate evolution and have survived thereafter for more than 400 million years. Our results indicate that, although both groups are classified as cyclostomes, the distance between hagfishes and lampreys is similar to that between humans and cartilaginous fishes, in terms of the geological time that has elapsed since their divergence.

Later in the hagfish lineage, our molecular dating with amino acid-based relaxed clocks and synonymous substitution clocks indicated that there was no branching of extant taxa until the two hagfish subfamilies, Myxinidae and Eptatretinae, split from each other 90–60 Mya in the Cretaceous–Tertiary Periods (Fig. 6). However, there is a paleontological report of a fossil species, Myxinikela siroka, from the Carboniferous fauna (~300 Mya) that is regarded as a putative outgroup of extant hagfishes (Bardack, 1991). In our divergence time estimate using relaxed molecular clocks, even when we tentatively constrained the upper limit of the divergence time of the Myxinidae–Eptatretinae split to 300 Mya, we obtained an identical result (data not shown), suggesting that this fossil species, Myxinikela siroka, actually is an outgroup of extant hagfishes (Fig. 6).

All the three subfamilies (Mordaciinae, Geotriinae, and Petromyzoninae) in the lamprey lineage are thought to have diverged from one another in a considerably short period of time (Conlon et al., 2001; Gill et al., 2003), as indicated by ambiguous phylogenetic relationships between these three taxa in recent molecular studies (Baldwin et al., 1988; Silver et al., 2004; Takahashi et al., 2006). Our synonymous substitution clock indicates that the Geotriinae–Petromyzoninae split occurred in the Permian–Triassic Periods (280–220 Mya; Fig. 6). Although this estimate needs to be reinforced with more robust analyses using amino acid sequences, this divergence time may coincide with the break-up of Gondwanaland (inhabited by species in Mordaciinae and Geotriinae) from Laurasia (inhabited by species in Petromyzoninae). Later, in the lineage of Petromyzoninae, an inter-generic split between Petromyzon and Lethenteron/Lampetra occurred 30–10 Mya in the Tertiary Period (Fig. 6). This result is consistent with the rough estimate by Docker et al. (1999), who simply assumed that a 2% divergence in mtDNA sequence corresponds to one million years (Brown et al., 1979). To further confirm phylogenetic relationships in Petromyzoninae, reported previously based on morphological features (Gill et al., 2003), molecular sequence data from other genera need to be included. Fossils of the lampreys Hardistiella montanensis, Mayomyzon pieckoensis, and Pipiscius zangerlii were found in the Carboniferous fauna (~280 Mya) and have been treated as outgroups of all extant lampreys (Bardack and Zangerl, 1971; Bardack and Richard-son, 1977; Janvier and Lund, 1983). In nuDNA-based and mtDNA-based analyses using relaxed molecular clocks, adding these paleontological data did not produce any substantial differences in results (data not shown), indicating that these fossil lampreys should still be regarded as outgroups of all extant lamprey species in Petromyzoniformes (Fig. 6).

Perspectives

Thanks to the efforts of researchers in various fields of biology (e.g., Kuratani et al., 2002), nucleotide and amino acid sequences of hagfishes and lampreys are accumulating in public databases. However, information at the molecular level is still far from satisfactory for cyclostomes (Fig. 1), in two aspects. First, in terms of the coverage of species diversity, there is a paucity of molecular data for southern hemisphere species, and these data are crucial for inferences of phylogenetic relationships and divergence times. For exam-
ple, in Myxiniformes, no molecular sequence data have been reported for Notomyxine, Nemamyxine, or Neomyxine, which are thought to belong in the subfamily Myxininae (Jørgensen, 1998). Similarly, in lampreys, the unavailability of appropriate nucleotide sequences for Mordacia hindered inclusion of Mordaciinae in our silent clock analysis, and the unavailability of appropriate amino acid sequences for Mordacia and Geotria did not allow us to include Mordaciinae and Geotriinae in our amino acid-based relaxed clock analysis.

Second, there are few reports of genomic DNA sequences for hagfishes and lampreys, and most of the reported nucleotide sequences are derived from mRNAs. In this study, based on the nucleotide sequences of protein-coding exons, we estimated accumulated levels of synonymous substitutions for inter-subfamilial and inter-generic species pairs, which highlighted a relatively low level of neutral nucleotide changes in the Myxine–Eptatretus and Petromyzon–Lethenteron pairs (Tables 1 and 2A). If the synonymous substitution rate in coding regions roughly cor-
responds to the neutral substitution rate in intergenic or intronic regions without regulatory functions, our estimates imply that these taxon pairs might be too phylogenetically close to discern potentially functional sequences, such as cis-regulatory elements or non-coding genes.

This principle is referred to as “phylogenetic footprinting” (Gumucio et al., 1992; see also Zhang and Gerstein, 2003, for review), and selecting multiple species with appropriate levels of nucleotide substitution facilitates an efficient in silico detection of potentially functional genomic sequences (Uchikawa et al., 2003; Johnson et al., 2004; Kusakabe, 2005). In this context, using lampreys as an example, comparison of non-coding genomic sequences among multiple species in Petromyzontinae alone will not provide a sufficient level of resolution to highlight functional fractions. Instead, judging from the almost saturated $K_s$ level between Geotriinae and Petromyzontinae (Table 2B), inclusion of southern hemisphere lampreys (Mordacia or Geotria) would be highly promising in comparisons with species in the northern hemisphere subfamily Petromyzontinae, such as Petromyzon marinus, whose genome sequencing project is now underway.

Despite the scarcity of sequence information, we attempted to overview the general features of base composition and evolutionary distance in cyclostomes, and obtained results that will serve as standards for future evolutionary and genomic studies of cyclostomes. We propose that, in any taxa, this sort of succinct evolutionary analysis should be a prerequisite for any biological studies involving multi-species comparisons. Even if the amount of available sequence information is limited, general trends embedded in sequence information can be extracted in light of the theories of molecular evolution.

ACKNOWLEDGMENTS

We are very grateful to Kazutaka Katoh for computational advice and to Kinya Ota, David McCauley, and Yoichi Matsuda for insightful discussion. This work was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

REFERENCES


Axelsson E, Webster MT, Smith NG, Burt DW, Ellefgen H (2005) Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. Genome Res 15: 120–125


Pancer Z, Mayer WE, Klein J, Cooper MD (2004b) Prototypic T cell receptor and CD4-like coreceptor are expressed by lymphocytes in the agnathan sea lamprey. Proc Natl Acad Sci USA 101: 13273–13278


Potter IC, Strahan R (1968) The taxonomy of the lampreys, Geotria

(Received May 1, 2006 / Accepted July 4, 2006)