Insights into Cyclostome Phylogenomics: Pre-2R or Post-2R?

Shigehiro Kuraku*

Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany

Interest in understanding the transition from prevertebrates to vertebrates at the molecular level has resulted in accumulating genomic and transcriptomic sequence data for the earliest groups of extant vertebrates, namely, hagfishes (Myxiniformes) and lampreys (Petromyzontiformes). Molecular phylogenetic studies on species phylogeny have revealed the monophyly of cyclostomes and the deep divergence between hagfishes and lampreys (more than 400 million years). In parallel, recent molecular phylogenetic studies have shed light on the complex evolution of the cyclostome genome. This consists of whole genome duplications, shared at least partly with gnathostomes (jawed vertebrates), and cyclostome lineage-specific secondary modifications of the genome, such as gene gains and losses. Therefore, the analysis of cyclostome genomes requires caution in distinguishing between orthology and paralogy in gene molecular phylogeny at the gene family scale, as well as between apomorphic and plesiomorphic genomic traits in larger-scale analyses. In this review, we propose possible ways of improving the resolvability of these evolutionary events, and discuss probable scenarios for cyclostome genome evolution, with special emphasis on the hypothesis that two-round (2R) genome duplication events occurred before the divergence between cyclostomes and gnathostomes, and therefore that a post-2R state is a genomic synapomorphy for all extant vertebrates.

Key words: hagfish, lamprey, orthology, hidden paralogy, long branch attraction, whole genome duplication

INTRODUCTION

Hagfishes (Myxiniformes) and lampreys (Petromyzontiformes) hold basal phylogenetic positions as the earliest groups of extant vertebrates, and they have been analyzed from various viewpoints to understand the transition from prevertebrates to vertebrates at the molecular level (e.g., Kuratani et al., 2002; Lamb et al., 2007; Osorio and Retaux, 2008). Currently, the National Center for Biotechnology Information (NCBI) sequence database has 124,029 and 24,521 entries of nucleotide sequences, including expressed sequence tags (ESTs), for species in the orders Petromyzontiformes and Myxiniformes, respectively (and 3006 and 652 entries for annotated protein sequences, respectively, as of March 26, 2008). Although many of these database entries represent limited types of gene families (e.g., genes encoding homeodomain-containing transcription factors, antigen-recognition proteins involved in the adaptive immune system, and so on), the current collection of annotated cyclostome genes is providing a rough but insightful overview into understanding the evolutionary properties of cyclostome genomes.

In light of the theories and knowledge of molecular evolution, abundant sequence resources have enabled various kinds of evolutionary information to be extracted, such as phylogenetic relationships, evolutionary time scales, and gains/losses of genes. In particular, the evolution of gene repertoires has had an impact on comparative analyses of gene function, including the regulatory gene network that governs development, physiology, and other biological processes. Decades of molecular studies have shown that many well-studied genes have similar copies between species as well as within species. Indispensable terms to characterize these gene copies evolutionarily, namely, 'orthology' and 'paralogy,' were originally introduced in the early 1970s (Fitch, 1970) and later described as follows (Fitch, 2000):

> "Orthology is that relationship where sequence divergence follows speciation, that is, where the common ancestor of the two genes lies in the cenancestor of the taxa from which the two sequences were obtained. This gives rise to a set of sequences whose true phylogeny is exactly the same as the true phylogeny of the organisms from which the sequences were obtained. Only orthologous sequences have this property. Paralogy is defined as that condition where sequence divergence follows gene duplication. Such genes might descend and diverge while existing side by side in the same lineage."

In general, recognition of orthology and paralogy is not straightforward when the genomic evolution of a species in question has experienced a series of complicated events

* Corresponding author. Phone: +49-7531-88-2763;
  Fax   : +49-7531-88-3018;
  E-mail: shigehiro.kuraku@uni-konstanz.de

(Fitch, 2000). In discussing early vertebrate evolution, the closest attention should be paid to this, because two rounds of genome duplications occurred, which resulted, for example, in four Hox gene clusters observed in non-teleost gnathostomes, such as mammals, chicken, *Xenopus*, and chondrichthyans (reviewed in Kuraku and Meyer, 2008). It has also been proposed that a large-scale duplication event occurred in the cyclostome lineage (see below).

Importantly, the above definitions of orthology and paralogy do not include any properties of gene function. For some cyclostome genes, changes in expression patterns are described as possible factors explaining the morphological differences between lamprey and gnathostomes (e.g., Shigetani et al., 2002; Uchida et al., 2003; Hammond and Whitfield, 2006). Apart from cyclostomes, many more studies highlight dynamic changes in gene expression patterns among orthologs during vertebrate evolution (e.g., Locascio et al., 2002; Kuraku et al., 2005). To conduct reasonable evolutionary studies, any comparative analysis regarding gene expression patterns and functions should follow the solid characterization of the phylogenetic nature of genes: orthology/paralogy should be clarified independently of any functional property of genes.

In this review, from the viewpoint of molecular evolution/phylogeny and genome informatics, current knowledge and perspectives are summarized to provide a better link between genomic properties and phenotypic evolution.

PHYLOGENY

Monophyly of cyclostomes

Although the taxonomic term Cyclostomata was first introduced in the early 19th century (Duméril, 1806), many subsequent studies on morphology regarded only hagfishes as the earliest branching group, taking lampreys as the true sister taxon of the gnathostomes (Janvier, 1996; see also Ota et al., 2007). However, the monophyly of cyclostomes was first supported by molecular phylogenetics in the early 1990's (Stock and Whitt, 1992). Many molecular phylogenetic studies have since supported the monophyly of cyclostomes using ribosomal DNA (Mallatt and Sullivan, 1998; Mallatt and Winchell, 2007), mitochondrial DNA (Delarbre et al., 2002), and protein-coding genes in the nuclear genome (Kuraku et al., 1999; Takezaki et al., 2003; Blair and Hedges, 2005; Delsuc et al., 2006) (Fig. 1A). The monophyly of cyclostomes can be regarded as one of the most clear-cut examples in which molecular phylogenetics has succeeded in updating phylogenetic relationships based on nonmolecular traits (Meyer and Zardoya, 2003).

Time scale

In light of the robust support for cyclostome monophyly, one can estimate the divergence time between Myxiniformes and Petromyzontiformes (these two taxa form monophyletic groups). By incorporating some fossil records for extinct cyclostomes, the relaxed molecular clock analysis, which is currently frequently used to estimate divergence times, provides an estimate that this divergence occurred 520–430 million years ago (Hedges, 2001; Blair and Hedges, 2005; Kuraku and Kuratani, 2006; summarized in Kuraku et al., 2008a) (Fig. 1A). Estimates vary largely depending on the assumed date of the divergence between

cyclostomes and gnathostomes. However, it has consistently been reported that the ancestor of the Myxiniformes and Petromyzontiformes diverged shortly (up to 100 million years) after the cyclostome and gnathostome lineages split (summarized in Kuraku et al., 2008a). In terms of the evolutionary time that has elapsed, it would not be surprising even if we were to identify differences between the genomes of hagfishes and lampreys that were as large as those observed in a comparison between the genomes of Mammalia and Chondrichthyes.

BASIC GENOMIC PROPERTIES

Karyotypes

Many cytogenetic observations of cyclostome genomes were described in the 1970s (Potter and Rothwell, 1970; Potter and Robinson, 1971; Robinson et al., 1975). In contrast to the relatively small number of chromosomes of hagfishes (2n=14–48), most lamprey species have more than 150 chromosomes (Fig. 1B) (original data were retrieved from the Animal Genome Size Database, http://www.genomesize.com). C-values (genome sizes) of hagfishes range from 2.29 to 4.59 pg, whereas those of lampreys range from 1.29 to 2.44 pg (Fig. 1B). Judging from the very small size of chromosomes and normal C-values in lampreys, the uniqueness of lamprey karyotypes is thought to have been caused mainly by successive Robertsonian chromosome fissions in the lamprey lineage (Robertson, 1916; Sumner, 2003).

Noncoding and repetitive landscape

Although the paucity of genomic sequences, especially in hagfishes, prevents the analysis of general genomic properties, there are some implications based on transcriptomic data. By analyzing the GC-content of four-fold degenerate sites (GC_4) in protein-coding regions, it has been shown that lampreys (both northern- and southern-hemisphere species) have high levels of GC_4 (70–90%), whereas hagfishes have moderate levels of GC_4 (40–60%) (Kuraku and Kuratani, 2006; Kuraku et al., unpublished observations). Currently available genomic sequences of lampreys have revealed that the extraordinarily high GC-content in protein-coding regions, which is represented by GC_4, is not a reflection of global genomic base composition (Table 1); rather, it is probably because of highly biased codon usage in lampreys.

Other evidence obtained in transcriptome analysis shows the existence of a short genomic element that might have spread throughout the lamprey genome (designated 'lamprino'; GenBank accession number AB425244). This element was found in the 3' untranslated region of the homeobox-containing transcription factor gene *LjHox13α* (Kuraku et al., 2008b). Similar sequences have been identified in many untranslated regions of lamprey ESTs (however, not in those of hagfish ESTs). The lamprey genome also contains a transcribed and translated sequence (found in AF464190) that has high similarity to Tc1-like transposase identified in salmonids (de Boer et al., 2007). It is highly likely that this gene was horizontally transferred from the host by parasitism. Moreover, it has been shown that the lamprey genome contains at least dozens of the microRNAs (miRNAs) already reported for other model vertebrates (Heimberg et al., 2008).
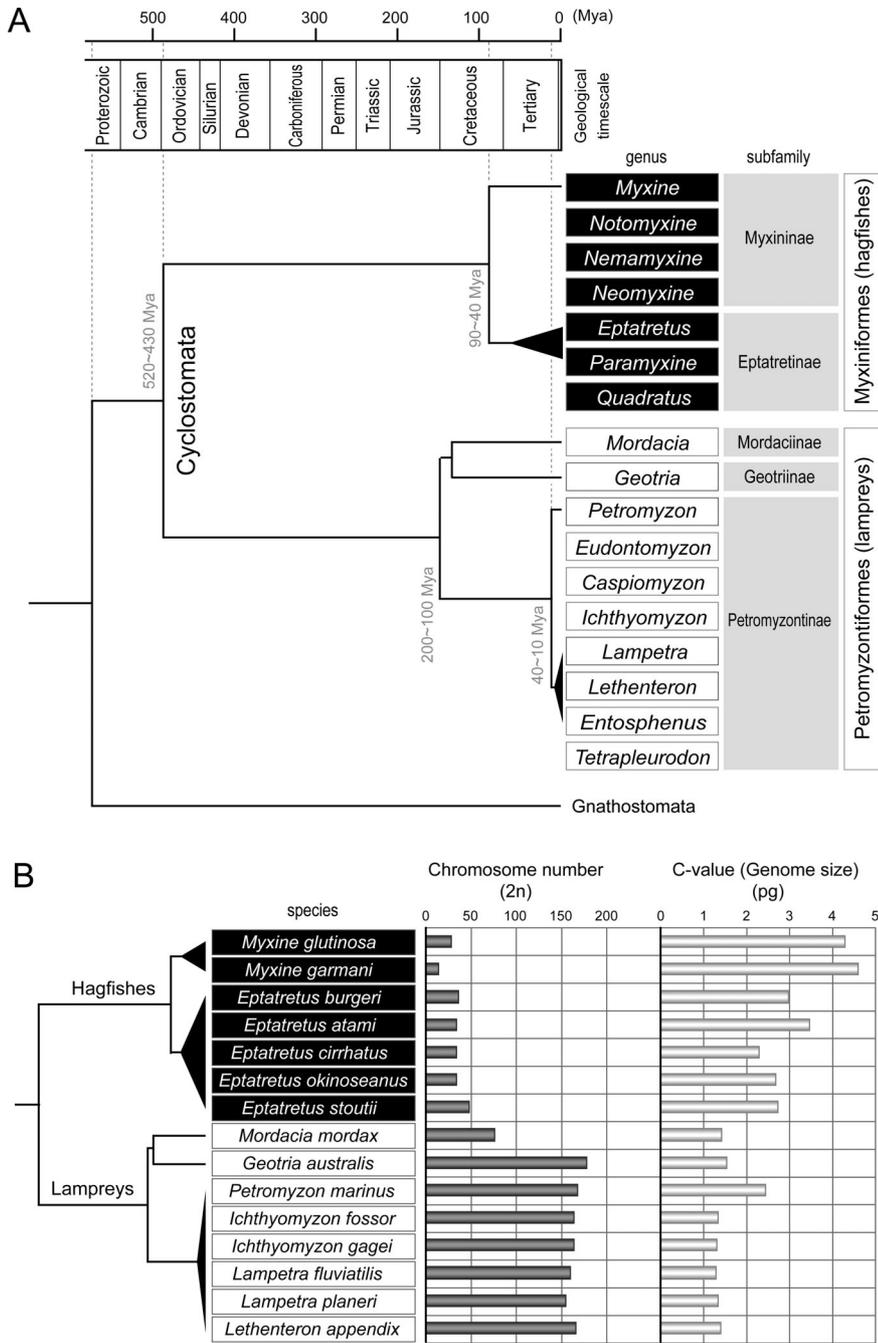
Fig. 1. Molecular-based phylogeny and genomic properties of cyclostomes. **(A)** Phylogenetic relationships of extant cyclostomes and estimated divergence times along the geological time scale. See Kuraku and Kuratani (2006) for details of divergence times within Myxiniformes and Petromyzontiformes. The divergence time between Myxiniformes and Petromyzontiformes (482 million years ago) is the average of values obtained from available reports (see Kuraku et al., 2008a for details). The phylogenetic relationships of southern hemisphere lampreys (*Geotria* and *Mordacia*) and their divergence times are based on unpublished observations by the author and collaborators. No molecular sequence data are available for reliable phylogenetic analyses of the hagfish genera *Notomyxine*, *Nemamyxine*, *Neomyxine*, and *Quadratus*, and the lamprey genera *Eudontomyzon*, *Caspiomyzon*, *Ichthyomyzon*, and *Tetrapleurodon*. **(B)** Karyotypes and C-values of cyclostomes. Data were retrieved from the Animal Genome Size Database (www.genomesize.com) and are shown only for species for which information was available. Tree topology and branch lengths are based on the phylogenetic relationships and time scale shown in A.

PITFALLS IN CYCLOSTOME GENE PHYLOGENY

Orthology/paralogy

Because of insufficient gene identification in cyclostomes and possible secondary gene gains/losses in their lineage (discussed below), orthology/paralogy between multiple cyclostome genes should be treated with caution. When more than two genes of the same cyclostome species cluster together, it is highly likely that they are paralogous (Fig. 2A). When a pair of hagfish and lamprey genes forms a cluster, this is regarded as a reflection of species phylogeny representing cyclostome monophyly (Fig. 2B). However, it is also possible that they are paralogous if we assume that a gene duplication has occurred in the cyclostome lineage, because of a potential 'hidden paralogy' (see Gribaldo and Philippe, 2002 for this term) (Fig. 2C). In contrast, even when a hagfish gene seems to have diverged first because of a gene duplication, it is still probable that this topology is caused by the rapid evolutionary rate of the hagfish sequence and that this rapidly evolving hagfish gene is orthologous to the lamprey gene (Fig. 2D). Otherwise, this is considered a result of secondary independent gene losses in the lineages leading to lampreys and gnathostomes (Fig. 2E). As exemplified here, the molecular phylogenies of cyclostome genes involve various difficult issues. Even when we extend our scope to the genome-wide level,

Table 1. GC-content of reported cyclostome genomic sequences over 10 kb in length.

Species	Annotation	Acc. No.	length (kb)	GC (%)
Hagfishes				
Mg	ParaHox region, BAC90C8	EU122193	145.0	46
Eb	ParaHox region, BAC7-H10	EU122194	103.4	44
Eb	VLR-A gene	AY965678	43.4	42
Eb	VLR-B gene	AY965679	92.0	42
Es	VLR-A gene	AY965680	81.7	42
Es	VLR-B gene	AY965681	76.8	45
Lampreys				
Pm	VLR gene	AY577941	57.3	46
Pm	VLR gene	AY577942	58.2	47
Pm	ABCB9-like gene	AH012171	14.5	46
Pm	HoxW10a gene	AF464190	29.6	51
Pm	CD45 gene	DQ008073	37.6	43
Pm	UNK clone CH303-4_3B5	AC182744	55.2	55
Pm	UNK clone CH303-4_3A4	AC182746	36.7	47
Pm	UNK clone CH303-4_3B4	AC182747	60.2	46
Pm	UNK clone CH303-4_3B11	AC182729	77.9	42
Pm	UNK clone CH303-4_3A12	AC182728	113.9	45
Pm	UNK clone CH303-4_3C11	AC182745	137.7	41
Pm	UNK clone CH303-4_3D5	AC182743	142.0	44
Pm	UNK clone CH303-4_3C12	AC182742	148.5	38
Pm	UNK clone CH303-4_3D9	AC182725	151.7	45
Pm	UNK clone CH303-4_3D4	AC182727	157.5	46
Pm	UNK clone CH303-4_3D2	AC182726	179.5	46
Lj	VLR gene 5' LRRCT segment	AB275449	10.6	50
Lj	VLR-B gene	AB272084	12.9	44
Lj	VLR-A gene	AB272083	40.2	45
Lj	MASP gene	AB078894	11.1	52

Abbreviations for species name: Mg, *Myxine glutinosa*; Eb, *Eptatretus burgeri*; Es, *Eptatretus stoutii*; Pm, *Petromyzon marinus*; Lj, *Lethenteron japonicum*.

gene orthology/paralogy should be recognized as an indispensable element with which to elucidate the history of cyclostome genome evolution (see Sicheritz-Ponten and Andersson, 2001; Huerta-Cepas et al., 2007 for attempts to practice genome-wide phylogenetic analysis, designated 'phylome').

Gene naming

Confusion in gene naming is a notorious artificial factor that can potentially prevent the reasonable integration of the phylogenetic nature of genes and their functional properties. It is highly recommended that newly identified genes be designated based on careful phylogenetic diagnosis. This procedure is achieved by a preliminary step involving homology search tools, such as BLAST (Altschul et al., 1997) (note that BLAST detects only 'similarity,' not 'homology') and subsequent elaborate phylogenetic analysis. For cyclostome genes, which often have elevated evolutionary rates, it is strongly recommended to employ molecular phylogenetic methods that are less susceptible to biases in among-lineage and also among-site rate variation (Felsenstein, 1978; Felsenstein, 1996; Philippe et al., 2005; summarized in Whelan et al., 2001; Holder and Lewis, 2003; Felsenstein, 2004). By taking into account the ambiguous signals of orthology between cyclostomes and gnathostomes that are frequently observed and the possibility of cyclostome lineage-specific gene duplications (discussed below), it is also rec-

ommended that cyclostome genes be named with care so that there is no confusion in understanding the phylogeny of these genes (e.g., the bone morphogenetic protein [BMP] genes *BMP2/4-A*, *BMP2/4-B*, and *BMP2/4-C* of lampreys in the BMP2/4 subfamily of the TGFβ gene family [McCauley and Bronner-Fraser, 2004]).

PHYLOME

Cyclostome lineage-specific gen(om)e duplication(s)

In the process of tree building using molecular sequences, cyclostome genes, which often have elevated evolutionary rates, potentially tend to form clusters because of a technical artifact called long branch attraction (LBA) (Felsenstein, 1978; Philippe et al., 2005). The idea of cyclostome lineage-specific genome duplication(s) has been mainly based on analyses of Hox genes, which suggests that at least one of the multiple Hox gene clusters in cyclostomes is derived from a cyclostome lineage-specific cluster duplication event that the gnathostome lineage did not experience (Force et al., 2002; Irvine et al., 2002; Fried et al., 2003; Stadler et al., 2004). This is because we often observe a flock of cyclostome genes in molecular phylogenetic trees (designated here as 'cyclogs'), in addition to a flock of multiple gnathostome paralogs. Possible cyclostome lineage-specific gene duplications have also been observed in other homeobox-containing genes, such as *Dlx* and
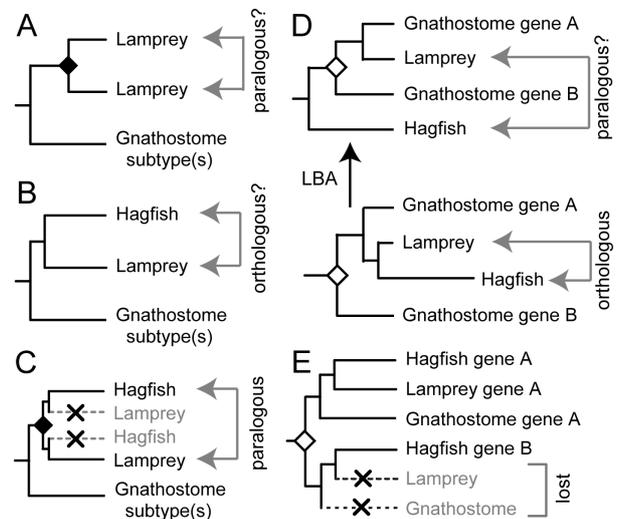

Fig. 2. Possible examples of orthology/paralogy between cyclostome genes. **(A)** Probable paralogy between two lamprey genes ('cyclog'; see text). It is also possible that this topology is an artifact produced by long-branch attraction if the two genes that duplicated before the cyclostome-gnathostome split evolved rapidly. **(B)** Probable orthology between hagfish and lamprey genes. **(C)** Hidden paralogy between hagfish and lamprey genes. If we assume that a gene duplication occurred in the cyclostome lineage in B, the hagfish-lamprey cluster can be regarded as a pair of paralogous genes. **(D)** Artificial paralogy between hagfish and lamprey genes. This can be caused by an elevated evolutionary rate due to long-branch attraction (LBA) of the hagfish sequence, even if these hagfish and lamprey genes are orthologous to each other. **(E)** Another explanation for early branching of cyclostome genes. The topology is similar to that of D; however, the topology in E is due to the loss of lamprey and gnathostome genes.

Engrailed (Neidert et al., 2001; Matsuura et al., 2008). Homeobox-containing genes, however, do not provide a robust conclusion in molecular phylogeny because of the short length of their alignable regions. In this sense, other gene families that can potentially provide more robust phylogeny will be better references. For example, gene duplications specific to the cyclostome lineages are also observed in the BMP2/4 gene subfamily (McCauley and Bronner-Fraser, 2004). The ongoing sea lamprey genome sequencing project will give rise to more conclusive data on whether these possible cyclostome lineage-specific gene duplications are technical artifacts, and if not, whether they were caused by a genome-wide duplication event. The sea lamprey genome sequencing project should also confirm whether the transposable elements mentioned above largely contributed to the expansion of 'cyclogs'.

Importantly, even if there was at least one whole genome duplication (WGD) in the cyclostome lineage, it is most likely that the evolutionary path leading to gnathostomes independently experienced two-round (2R) genome duplications. This has been repeatedly confirmed by the existence of multiple (usually four, as in Hox gene clusters) similar arrays of genes along chromosomes in humans and mice (reviewed in Kasahara, 2007). It should be emphasized that multiple gene copies observed in gnathostomes cannot be imputed to cyclostome lineage-specific events, and thus cyclostome lineage-specific gen(om)e duplications should be treated separately from the WGDs that resulted in redundant gene repertoires in gnathostomes.

Timing of WGDs: pre-2R or post-2R?

It remains controversial whether the mode of genome duplication in the cyclostome-gnathostome split was a simple '2R' pattern (Furlong and Holland, 2002; see also

Hughes, 1999; Friedman and Hughes, 2001; Hughes and Friedman, 2003 for a negative view about the WGDs per se; see also Gregory, 2005 for alternative modes of large-scale duplication). However, under the assumption that there were 2R WGDs, as has been more recently reported (Dehal and Boore, 2005), there are three possible scenarios for explaining the timing of WGDs relative to the cyclostome-gnathostome split (Fig. 3). Hypothesis A has not been explicitly proposed to date, whereas there are some reports supporting hypotheses B (Pendleton et al., 1993; Sharman and Holland, 1998; Escriva et al., 2002; Force et al., 2002; Stadler et al., 2004) and C (Fried et al., 2003; Furlong et al., 2007). Because of the high level of inconsistency in gene phylogeny between gene families, there is no consensus on which of these three scenarios is correct (reviewed in Kuraku and Meyer, 2008). This difficulty is thought to have been caused by a combination of possible complicating factors, as listed below.

- The WGDs, plus the cyclostome-gnathostome divergence, occurred within a short period of time (Larhammar et al., 2002; Horton et al., 2003).
- A certain amount of absolute time (more than 500 million years) has elapsed since the WGDs. This is relatively ancient compared with other well-studied WGD events (e.g., those in the lineages of fungi, land plants, teleosts, and *Xenopus*; reviewed in Wolfe, 2001).
- Many cyclostome genes studied to date are not ideal genes for molecular phylogenetic analysis (short [e.g., homeobox genes], rapidly evolving, or highly susceptible to selection or small-scale gene duplications).
- Cyclostome genes (especially hagfish genes) are often rapidly evolving and thus can potentially prevent a robust molecular phylogenetic analysis.
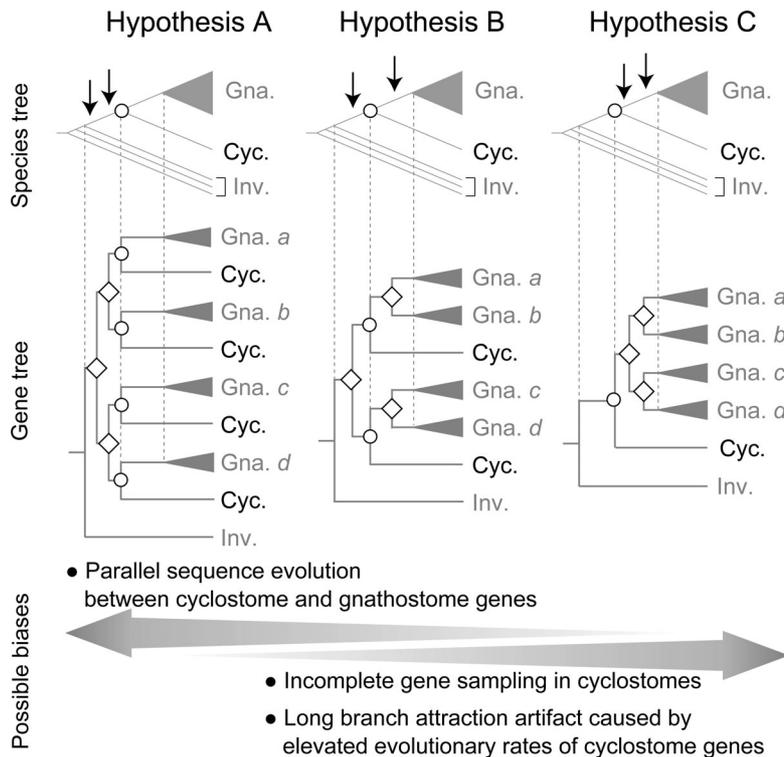

Fig. 3. Possible scenarios for two-round whole genome duplications (WGDs). Under the assumption that there were two rounds of WGDs, three possible scenarios with different timings of WGDs (arrows) are shown (hypotheses A–C). Expected tree topologies for gene phylogeny are illustrated for an imaginary gene family comprising one invertebrate outgroup (Inv.), four gnathostome (Gna.) genes (paralogs *a–d*), and an intact set of cyclostome (Cyc.) genes. Gene duplications are represented by white diamonds. The cyclostome-gnathostome divergence is represented by white circles. At the bottom, possible artificial biases are shown with the direction of their misleading conclusion. Cyclostome lineage-specific gene duplications are not shown in this figure (see text).
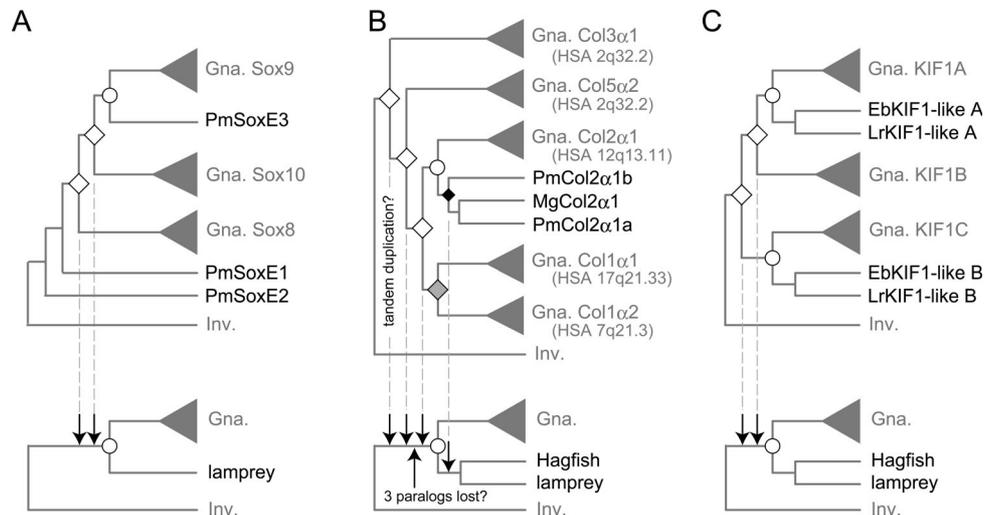
Fig. 4. Previously reported cyclostome gene phylogenies and deduced evolutionary scenarios. **(A)** HMG-box containing SoxE genes (Zhang and Cohn, 2006; Zhang et al., 2006). **(B)** Clade A fibril collagen genes (Iwabe and Miyata, 2002). **(C)** Kinesin KIF1 A/B/C genes. Gene duplications that gave rise to gnathostome paralogs are represented by white diamonds. The cyclostome-gnathostome divergence is represented by a white circle. A gene duplication thought to have occurred in the cyclostome lineage is represented by a black diamond. A gene duplication whose timing cannot be determined by the tree topology is represented by a gray diamond. The timing of gene duplications observed in the upper trees are indicated by arrows in the species phylogeny at the bottom. Gnathostome genes are represented by the names of paralogs, and cyclostome genes are shown as combinations of abbreviations of species names and gene names. Abbreviations of species names: Lj, *Lethenteron japonicum*; Pm, *Petromyzon marinus*; Mg, *Myxine glutinosa*; Eb, *Eptatretus burgeri*; Lr, *Lethenteron reissneri*. In C, chromosomal localizations of human genes are also shown. For B, the phylogeny of cyclostome genes is based on McCauley and Bronner-Fraser (2006) (see also Zhang et al., 2006).

- Gene sampling in gnathostomes usually concentrates on bony vertebrates, and thus the earliest branching group in extant gnathostomes, namely, Chondrichthyes, is not included. Therefore, genes in basal jawed vertebrates are usually represented by *Xenopus laevis* and teleost fishes, which underwent an additional genome duplication in each lineage (Wolfe, 2001).
- Many cyclostome genes remain to be identified.
- Species sampling for cyclostomes is not thorough (usually only one species is included in an analysis).
- Intensive phylogenetic analysis (model selection [e.g. consideration of among-site rate heterogeneity], probabilistic tree search methods [e.g., maximum-likelihood, Bayesian methods], and statistical evaluation of alternative tree topologies obtained; see Holder and Lewis, 2003) has not yet been thoroughly applied to phylogenetic studies of cyclostome genes.
- The results of different molecular phylogenetic analyses are evaluated by different criteria. In many cases, only the best tree is analyzed, without any statistical examination.

To optimize molecular phylogenetic analysis for addressing this issue, one is expected to select appropriate gene families with unbiased taxon sampling. Apart from large-scale duplication events, such as WGDs, rampant small-scale gene duplications, such as tandem gene duplications and retropositions, would also have occurred. These events occur in a gene family-specific manner. For example, the globin gene family has undergone a series of tandem gene duplications during vertebrate evolution (Goodman et al., 1975), and glycoprotein hormones (e.g., gonadotropin) have also undergone gene family-specific, small-scale gene
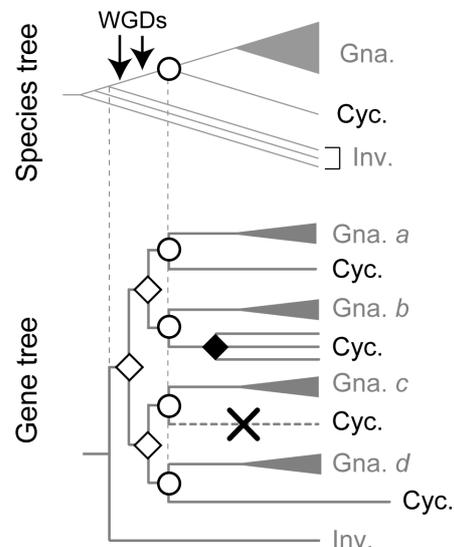

Fig. 5. Schematic illustration of the pan-vertebrate tetraploidization (PV4) hypothesis. A generalized gene tree of an imaginary gene family (lower) is shown with a species tree on which hypothesized events in the evolution of gene repertoires are mapped (upper). Gene duplications that gave rise to gnathostome paralogs are represented by white diamonds. The cyclostome-gnathostome split (white circles) is thought to have occurred before the duplicates generated in WGDs became fully neo- or subfunctionalized. Subsequently, possibly because of less functional constraint, cyclostome paralogs were secondarily duplicated to produce 'cyclogs' (represented by a black diamond), lost (X), or neofunctionalized (an elongated branch). These kinds of secondary modifications also occurred in the gnathostome lineage, but the secondary modifications in the cyclostome lineage are often neglected because of excessive expectation that cyclostomes should have retained a genomic architecture ancestral for vertebrates.

Table 2. Molecular phylogenetic properties of the gene families shown in Fig. 4.

Gene families	Property	# of gnathostome paralogs	# of cyclostome paralogs identified so far[*]	Supported hypothesis (A–C)	Cyclostome lineage-specific gene duplications observed ?
SoxE (8/9/10)	HMG-box containing TF	3	3	A	No
Clade A fibril collagen	collagen	5	2	A or B	Yes
KIF 1 A/B/C	kinesin	3	2	A	No

* This count is based on a parsimonious estimate, with an assumption of cyclostome monophyly. TF, transcription factor.

duplications (Sower et al., 2006). Gene families that have undergone small-scale gene duplications should be treated with caution when assessing the timing of genome-wide duplication events.

For estimating the timing of WGDs, gene numbers in cyclostomes have been repeatedly used as markers. For example, a lower number of *Dlx* genes in lampreys than in gnathostomes was regarded as evidence that lampreys diverged before at least one round of WGDs (Neidert et al., 2001; Donoghue and Purnell, 2005). However, lampreys have at least six *Dlx* genes, as do non-teleost gnathostomes (Kuraku et al., 2008c). The many cases in which a lower number of genes has been identified in cyclostomes compared with gnathostomes are expected to be updated by the identification of many more genes in whole genome sequencing. We should also take into account the possibility of secondary gene losses in the cyclostome lineage (see Furlong et al., 2007 for an example of the hagfish ParaHox gene cluster). Even in gnathostomes, after the quadruplication of a single ancestral gene, many gene families have lost at least one paralog produced in 2R WGDs. The most frequently observed ratio of gene numbers per gene family, which is derived from a single ancestral gene in the chordate ancestor, between invertebrates and non-teleost gnathostomes is 1:2 (Furlong and Holland, 2002). For example, most subfamilies in the nuclear receptor gene families have a ratio of 1:3 (e.g., retinoic acid receptor [*RAR*α, *RAR*β, and *RAR*γ), whereas most subfamilies in the *Wnt* gene family have a ratio of 1:2 (e.g., *Wnt7a* and *Wnt7b*; interestingly, there are no subfamilies with the ratios of 1:3 or 1:4 in this gene family). Nevertheless, genome-wide identification of similar arrays of genes along multiple (usually up to four) chromosomal regions (Dehal and Boore, 2005) and detailed analysis of conserved landmark gene clusters have led to the idea of WGDs and the ratio of 1:4 to parsimoniously explain the difference in gene numbers between invertebrates and non-teleost gnathostomes (reviewed in Kasahara, 2007). To obtain a clear image of genomic status in cyclostomes, this line of evidence on conserved syntenies to analyze 'how' intra-genome redundancy was introduced should be accompanied by large-scale, elaborate molecular phylogenetic analyses to clarify 'when' such intra-genome redundancy was created. A lower number of gene repertoires in cyclostomes does not necessarily mean that they diverged before the genome expansion: a tree-based approach is indispensable.

Molecular phylogenetic studies have also been performed for a number of non-homeobox gene families. Based on observations in recent studies (Fig. 4), hypothesis A (Fig. 3) is proposed here as an alternative scenario and is designated as the 'pan-vertebrate tetraploidization (PV4)' hypoth-

esis (Fig. 5). For all three gene families shown in Fig. 4, gene duplications were observed more frequently before the cyclostome-gnathostome split than after it, suggesting hypothesis A as a scenario for explaining the timing of WGDs relative to the cyclostome-gnathostome split (see Table 2). These gene families have longer alignment lengths and exhibit a greater level of sequence divergence among orthologs and paralogs than most gene families on which hypotheses B and C are based (see Fig. 3). It is likely that the hypotheses suggested to date (hypotheses B and C) might have been produced by the technically difficult factors mentioned above. This could be confirmed with elaborate molecular phylogenetic analyses using larger data sets.

SUMMARY

When we take into account the difficult mentioned factors above regarding (1) the absolute and relative timing of evolutionary events, (2) technical limitations, and (3) insufficient data sets, cyclostome genomes can be recognized as very challenging targets for evolutionary genomics. Whether cyclostomes diverged from the future gnathostome lineage before or after WGDs, cyclostome genomes seem to show relatively degenerate gene repertoires with unique features in some respects (e.g., high GC-content in protein-coding regions and high chromosome number in lampreys). It would not be surprising, however, even if cyclostomes have experienced a large amount of secondary modification to their genomes, to discover large differences in genomic content between cyclostomes and gnathostomes, which diverged more than 500 million years ago, and also between hagfishes and lampreys, diverged more than 400 million years ago. In contrast, stable genomes of chondrichthyans have been estimated by analyses of partial genomic sequences of the ghost shark, *Callorhinchus milii* (Venkatesh et al., 2007; Yu et al., 2008; see also Kuraku and Meyer [2008] for the possible retention of the ancestral structure of the Hox gene clusters of gnathostomes by the horn shark, *Heterodontus francisci*). Under the assumption that the post-WGD state is a genomic synapomorphy of all extant vertebrates (hypothesis A in Fig. 3; Fig. 5), this contrast might reflect differences between the taxon that diverged before genes produced in WGDs became fully sub- or neofunctionalized (namely, Cyclostomata) and the taxon that diverged after many of them acquired the functions that are shared by diverse living gnathostomes (namely, Chondrichthyes). The differences between these two taxa in levels of constraints acting on genes might have led to the difference in the frequency of gene retentions and duplicability. Thus, determining orthology/paralogy based on solid molecular phylogenetic evidence is a prerequisite for linking the evolution of gene repertoires with the evolution of gene functions

and phenotypes. Currently, a genome sequencing project for the sea lamprey is underway. In terms of deep divergence time within cyclostomes, it will be interesting to compare this genome with the hagfish genome. Without their genomes, we cannot discuss what defines vertebrate genomes.

ACKNOWLEDGMENTS

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402

Blair JE, Hedges SB (2005) Molecular phylogeny and divergence times of deuterostome animals. Mol Biol Evol 22: 2275–2284

de Boer JG, Yazawa R, Davidson WS, Koop BF (2007) Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics 8: 422

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3: e314

Delarbre C, Gallut C, Barriel V, Janvier P, Gachelin G (2002) Complete mitochondrial DNA of the hagfish, *Eptatretus burgeri*: the comparative analysis of mitochondrial DNA sequences strongly supports the cyclostome monophyly. Mol Phylogenet Evol 22: 184–192

Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439: 965–968

Donoghue PC, Purnell MA (2005) Genome duplication, extinction and vertebrate evolution. Trends Ecol Evol 20: 312–319

Duméril AMC (1806) Zoologie Analytique, ou Méthode Naturelle de Classification des Animaux, Rendue plus Facile a l'Aide de Tableaux Synoptiques par. Allais, Paris

Escriva H, Manzon L, Youson J, Laudet V (2002) Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. Mol Biol Evol 19: 1440–1450

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27: 401–410

Felsenstein J (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol 266, 418–427

Felsenstein J (2004) Inferring Phylogenies. Sinauer Associates, Sunderland, MA

Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst Zool 19: 99–113

Fitch WM (2000) Homology: a personal view on some of the problems. Trends Genet 16: 227–231

Force A, Amores A, Postlethwait JH (2002) Hox cluster organization in the jawless vertebrate *Petromyzon marinus*. J Exp Zool 294: 30–46

Fried C, Prohaska SJ, Stadler PF (2003) Independent Hox-cluster duplications in lampreys. J Exp Zool B 299: 18–25

Friedman R, Hughes AL (2001) Pattern and timing of gene duplication in animal genomes. Genome Res 11: 1842–1847

Furlong RF, Holland PW (2002) Were vertebrates octoploid? Philos Trans R Soc Lond B 357: 531–544

Furlong RF, Younger R, Kasahara M, Reinhardt R, Thorndyke M, Holland PW (2007) A degenerate ParaHox gene cluster in a degenerate vertebrate. Mol Biol Evol 24: 2681–2686

Goodman M, Moore GW, Matsuda G (1975) Darwinian evolution in the genealogy of hemoglobin. Nature 253: 603–608

Gregory TR (2005) The Evolution of the Genome. Elsevier, San Diego

Gribaldo S, Philippe H (2002) Ancient phylogenetic relationships. Theor Popul Biol 61: 391–408

Hammond KL, Whitfield TT (2006) The developing lamprey ear closely resembles the zebrafish otic vesicle: *otx1* expression can account for all major patterning differences. Development 133: 1347–1357

Hedges SB (2001) Molecular evidence for the early history of living vertebrates. In "Major Events in Early Vertebrate Evolution" Ed by PE Ahlberg, Taylor & Frances, London, pp 119–134

Heimberg AM, Sempere LF, Moy VN, Donoghue PC, Peterson KJ (2008) MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci USA 105: 2946–2950

Holder M, Lewis PO (2003) Phylogeny estimation: traditional and Bayesian approaches. Nat Rev Genet 4: 275–284

Horton AC, Mahadevan NR, Ruvinsky I, Gibson-Brown JJ (2003) Phylogenetic analyses alone are insufficient to determine whether genome duplication(s) occurred during early vertebrate evolution. J Exp Zoolog B Mol Dev Evol 299: 41–53

Huerta-Cepas J, Dopazo H, Dopazo J, Gabaldon T (2007) The human phylome. Genome Biol 8: R109

Hughes AL (1999) Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J Mol Evol 48: 565–576

Hughes AL, Friedman R (2003) 2R or not 2R: testing hypotheses of genome duplication in early vertebrates. J Struct Funct Genomics 3: 85–93

Irvine SQ, Carr JL, Bailey WJ, Kawasaki K, Shimizu N, Amemiya CT, Ruddle FH (2002) Genomic analysis of Hox clusters in the sea lamprey *Petromyzon marinus*. J Exp Zool 294: 47–62

Iwabe N, Miyata T (2002) Kinesin-related genes from diplomonad, sponge, amphioxus, and cyclostomes: divergence pattern of kinesin family and evolution of giardial membrane-bounded organella. Mol Biol Evol 19: 1524–1533

Janvier P (1996) Early Vertebrates. Clarendon Press, Oxford

Kasahara M (2007) The 2R hypothesis: an update. Curr Opin Immunol 19: 547–552

Kuraku S, Kuratani S (2006) Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. Zool Sci 23: 1053–1064

Kuraku S, Meyer A (2008) The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. Int J Dev Biol: in press

Kuraku S, Hoshiyama D, Katoh K, Suga H, Miyata T (1999) Monophyly of lampreys and hagfishes supported by nuclear DNA-coded genes. J Mol Evol 49: 729–735

Kuratani S, Kuraku S, Murakami Y (2002) Lamprey as an evo-devo model: lessons from comparative embryology and molecular phylogenetics. Genesis 34: 175–183

Kuraku S, Usuda R, Kuratani S (2005) Comprehensive survey of carapacial ridge-specific genes in turtle implies co-option of some regulatory genes in carapace evolution. Evol Dev 7: 3–17

Kuraku S, Ota KG, Kuratani S (2008a) Cyclostomata. In "Timetree of Life" Ed by SB Hedges, S Kumar, Oxford University Press, New York, in press

Kuraku S, Takio Y, Tamura K, Aono H, Meyer A, Kuratani S (2008b) Noncanonical roles of Hox14 revealed by its expression patterns in lamprey and shark. Proc Natl Acad Sci USA 105: 6679–6683

Kuraku S, Meyer A, Kuratani S (2008c) Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before, or after? Mol Biol Evol (doi:10.1093/molbev/msn222)

Lamb TD, Collin SP, Pugh EN Jr (2007) Evolution of the vertebrate

eye: opsins, photoreceptors, retina and eye cup. Nat Rev Neurosci 8: 960–976

Larhammar D, Lundin LG, Hallbook F (2002) The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. Genome Res 12: 1910–1920

Locascio A, Manzanares M, Blanco MJ, Nieto MA (2002) Modularity and reshuffling of *Snail* and *Slug* expression during vertebrate evolution. Proc Natl Acad Sci USA 99: 16841–16846

Mallatt J, Sullivan J (1998) 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. Mol Biol Evol 15: 1706–1718

Mallatt J, Winchell CJ (2007) Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star. Mol Phylogenet Evol 43: 1005–1022

Matsuura M, Nishihara H, Onimaru K, Kokubo N, Kuraku S, Kusakabe R, Okada N, Kuratani S, Tanaka M (2008) Identification of four *Engrailed* genes in the Japanese lamprey, *Lethenteron japonicum*. Dev Dyn 273: 1581–1589

McCauley DW, Bronner-Fraser M (2004) Conservation and divergence of BMP2/4 genes in the lamprey: expression and phylogenetic analysis suggest a single ancestral vertebrate gene. Evol Dev 6: 411–422

Meyer A, Zardoya R (2003) Recent advances in the (molecular) phylogeny of vertebrates. Annu Rev Ecol Evol Syst 34: 311–338

Neidert AH, Virupannavar V, Hooker GW, Langeland JA (2001) Lamprey *Dlx* genes and early vertebrate evolution. Proc Natl Acad Sci USA 98: 1665–1670

Osorio J, Retaux S (2008) The lamprey in evolutionary studies. Dev Genes Evol. 218: 221–235

Ota KG, Kuraku S, Kuratani S (2007) Hagfish embryology with reference to the evolution of the neural crest. Nature 446: 672–675

Pendleton JW, Nagai BK, Murtha MT, Ruddle FH (1993) Expansion of the Hox gene family and the evolution of chordates. Proc Natl Acad Sci USA 90: 6300–6304

Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F (2005) Heterotachy and long branch attraction in phylogenetics. BMC Evol Biol 5: 50

Potter IC, Robinson ES (1971) The chromosomes. In "The Biology of Lampreys" Ed by MW Hardisty, IC Potter, Academic Press, London, pp 279–293

Potter IC, Rothwell B (1970) The mitotic chromosomes of the lamprey, *Petromyzon marinus* L. Experientia 26: 429–430

Robertson WMRB (1916) Chromosome studies. I. Taxonomic relationships shown in the chromosomes of Tettegidae and Acrididiae: V-shaped chromosomes and their significance in Acrididiae, Locustidae and Grillidae: chromosomes and variation. J Morphol 27: 179–331

Robinson ES, Potter IC, Atkin NB (1975) The nuclear DNA content of lampreys. Experientia 31: 912–913

Sharman AC, Holland PW (1998) Estimation of Hox gene cluster number in lampreys. Int J Dev Biol 42: 617–620

Shigetani Y, Sugahara F, Kawakami Y, Murakami Y, Hirano S, Kuratani S (2002) Heterotopic shift of epithelial–mesenchymal interactions in vertebrate jaw evolution. Science 296: 1316–1319

Sicheritz-Ponten T, Andersson SG (2001) A phylogenomic approach to microbial evolution. Nucleic Acids Res 29: 545–552

Sower SA, Moriyama S, Kasahara M, Takahashi A, Nozaki M, Uchida K, Dahlstrom JM, Kawauchi H (2006) Identification of sea lamprey GTHbeta-like cDNA and its evolutionary implications. Gen Comp Endocrinol 148: 22–32

Stadler PF, Fried C, Prohaska SJ, Bailey WJ, Misof BY, Ruddle FH, Wagner GP (2004) Evidence for independent Hox gene duplications in the hagfish lineage: a PCR-based gene inventory of *Eptatretus stoutii*. Mol Phylogenet Evol 32: 686–694

Stock DW, Whitt GS (1992) Evidence from 18S ribosomal RNA sequences that lampreys and hagfishes form a natural group. Science 257: 787–789

Sumner AT (2003) Chromosomes: organization and function. Blackwell, Malden, MA

Takezaki N, Figueroa F, Zaleska-Rutczynska Z, Klein J (2003) Molecular phylogeny of early vertebrates: monophyly of the agnathans as revealed by sequences of 35 genes. Mol Biol Evol 20: 287–292

Uchida K, Murakami Y, Kuraku S, Hirano S, Kuratani S (2003) Development of the adenohypophysis in the lamprey: evolution of epigenetic patterning programs in organogenesis. J Exp Zoolog B 300: 32–47

Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, et al. (2007) Survey sequencing and comparative analysis of the elephant shark (*Callorhinchus milii*) genome. PLoS Biol 5: e101

Whelan S, Lio P, Goldman N (2001) Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet 17: 262–272

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet 2: 333–341

Yu W, Rajasegaran V, Yew K, Loh W, Tay B, Amemiya C, Brenner S, Venkatesh B (2008) Elephant shark sequence reveals unique insights into the evolutionary history of vertebrate genomes: a comparative analysis of the protocadherin cluster. Proc Natl Acad Sci USA 105: 3819–3824

Zhang G, Cohn MJ (2006) Hagfish and lancelet fibrillar collagens reveal that type II collagen-based cartilage evolved in stem vertebrates. Proc Natl Acad Sci U S A 103: 16829–16833

Zhang G, Miyamoto MM, Cohn MJ (2006) Lamprey type II collagen and Sox9 reveal an ancient origin of the vertebrate collagenous skeleton. Proc Natl Acad Sci USA 103: 3180–3185