ELSEVIER

Gene 259 (2000) 129–138

GENE
AN INTERNATIONAL JOURNAL ON GENES, GENOMES AND EVOLUTION

www.elsevier.com/locate/gene

Evolutionary history of the enolase gene family

Michael R. Tracy [1], S. Blair Hedges *

Department of Biology, Institute of Molecular Evolutionary Genetics, and Astrobiology Research Center, Pennsylvania State University, University Park, PA 16802, USA

Received 15 May 2000; received in revised form 15 August 2000; accepted 25 August 2000
Received by T. Gojobori

Abstract

The enzyme enolase [EC 4.2.1.11] is found in all organisms, with vertebrates exhibiting tissue-specific isozymes encoded by three genes: alpha (α), beta (β), and gamma (γ) enolase. Limited taxonomic sampling of enolase has obscured the timing of gene duplication events. To help clarify the evolutionary history of the gene family, cDNAs were sequenced from six taxa representing major lineages of vertebrates: *Chiloscyllium punctatum* (shark), *Amia calva* (bowfin), *Salmo trutta* (trout), *Latimeria chalumnae* (coelacanth), *Lepidosiren paradoxa* (South American lungfish), and *Neoceratodus forsteri* (Australian lungfish). Phylogenetic analysis of all enolase and related gene sequences revealed an early gene duplication event prior to the last common ancestor of living organisms. Several distantly related archaebacterial sequences were designated as 'enolase-2', whereas all other enolase sequences were designated 'enolase-1'. Two of the three isozymes of enolase-1, α- and β-enolase, were discovered in actinopterygian, sarcopterygian, and chondrichthian fishes. Phylogenetic analysis of vertebrate enolases revealed that the two gene duplications leading to the three isozymes of enolase-1 occurred subsequent to the divergence of living agnathans, near the Proterozoic/Phanerozoic boundary (approximately 550 Mya). Two copies of enolase, designated α_1 and α_2, were found in the trout and are presumed to be the result of a genome duplication event.

Keywords: Archaebacteria; Cambrian explosion; Gene duplication; Horizontal transfer; Polyploidization; Vertebrate evolution

1. Introduction

Gene and genome duplications are believed to have contributed significantly to the evolution of complex body plans in the metazoans (Holland et al., 1994; Brooke et al., 1998), especially vertebrates (Nadeau and Sankoff, 1997; Stock et al., 1997; Williams and Holland, 1998). The importance of such duplications for animal evolution relies on knowledge of the timing and order of the duplications. For example, if gene duplication occurred before the divergence of deuterostome and protostome animals, it would not be directly related to developmental and morphological events unique to vertebrates (Hughes, 1999), although it could be related through later recruitment (Ono-Koyanagi et al., 2000). Broad taxonomic sampling enhances any evolutionary analysis of gene relationships, and examples of well-studied gene families include hemoglobin (Hardison, 1996; Gorr et al., 1999), lactate dehydrogenase (Crawford et al., 1989; Stock and Whitt, 1992b; Mannen et al., 1997; Stock et al., 1997) and *Otx* (Williams and Holland, 1998).

Enolase (2-phospho-D-glycerate hydrolase) [EC 4.2.1.11], a dimeric enzyme of the glycolytic pathway responsible for catalyzing the conversion of 2-phosphoglycerate to phosphoenolpyruvate (Day et al., 1993; Brown and Doolittle, 1997), is another conserved gene family that allows for the evolutionary study of gene duplication in animals. It is found in all organisms and is a member of the enolase gene superfamily, which includes carboxyphosphonoenolpyruvate synthase, mandelate racemase, galactonate dehydratase, glucarate dehydratase, muconate-lactonizing enzymes, *N*-acylamino acid racemase, beta-methylaspartate ammonia-lyase, and *o*-succinylbenzoate synthase (Babbitt et al., 1996; Pegg and Babbitt, 1999). Vertebrates exhibit three,

Abbreviations: cDNA, complimentary DNA; MSE, muscle-specific enolase; Mya, million years ago; NNE, non-neuronal enolase; NSE, neuron-specific enolase; PCR, polymerase chain reaction; poly(A)$^+$ RNA, polyadenylated RNA.

* Corresponding author. Tel.: +1-814-865-9991; fax: +1-814-865-9131.
E-mail address: sbh1@psu.edu (S.B. Hedges)
[1] Present address: School of Medicine, University of Pennsylvania, Suite 100 Stemmler Hall, Philadelphia, PA 19104, USA.

0378-1119/00/$ - see front matter
PII: S0378-1119(00)00439-X

tissue-specific enolase isozymes: non-neuronal enolase (NNE), muscle-specific enolase (MSE), and neuron-specific enolase (NSE) (Rider and Taylor, 1975; Bishop and Corces, 1990). NNE is a homodimer of α-enolase subunits and is the only form found in liver tissue. MSE, a homodimer of β-enolase, and NSE, a homodimer of γ-enolase, are localized in adult muscle cells and neuronal and neuroendocrine cells, respectively (Oliva et al., 1989). An α heterodimer from neuronal tissue also has been reported (Schmechal et al., 1984).

Analyses of amino acid and nucleotide sequences have suggested that the paralogs rapidly arose from a single ancestral gene through a 'burst' event early in vertebrate evolution (Day et al., 1993). While numerous sequences are available for enolases, especially vertebrate α-enolases, sequences for vertebrate β- and γ-enolases are restricted to chicken, human, rat, and mouse (an amino acid sequence for rabbit β-enolase also is available). Based on intron lengths, Giallongo et al. (1990) proposed that β- and γ-enolase are the two most closely related genes. Because of the paucity of sequence data, the timing and order of the gene duplications have remained unresolved.

Several studies have placed the enolase gene duplications between 200 and 300 Mya (Rider and Taylor, 1975; Clark-Rosenberg and Marangos, 1980; Segil et al., 1988). However, Landrey et al. (1978) proposed that a gene duplication was responsible for the six electrophoretically distinct enolase isozymes that they found in *Coho salmon*. With the likely origin of actinopterygian fishes occurring approximately 450 Mya (Benton, 1997; Kumar and Hedges, 1998), this conclusion regarding *Coho salmon* should be re-examined. Landrey et al. (1978) suggested that the polyploidy of salmonids could be a complicating factor, but, without direct sequence information, it is impossible to determine whether the isozymes found by Landrey et al. (1978) are the result of a gene duplication in the vertebrate lineage, a genome duplication in the salmonid lineage, or both. Furthermore, it remains unclear which vertebrates might exhibit multiple copies of the enolase gene.

Phylogenetic analyses of enolase paralogs could be complicated by the possible effects of gene sharing. Gene sharing occurs when a gene, without duplication, acquires a second function while maintaining its original function. This places two different sets of selective constraints on mutational change within the shared gene (Piatigorsky, 1989; Piatigorsky and Wistow, 1989). The α-enolase gene also is expressed as the lens protein τ-crystallin in lampreys, other fishes, and reptiles (Wistow et al., 1988). Presently, it is not known if either β- or γ-enolase functions in this manner, but the possibility does exist that the degree of conservation and rate of evolution may differ markedly between the paralogs. Indeed, McAleese et al. (1988) found that γ-enolase is more stringently conserved than α-enolase. More recently, Day et al. (1993) reported that NNE appears to be evolving faster than either NSE or MSE.

To elucidate more clearly the evolutionary history of the enolase gene family, we have generated a phylogeny from all available, complete (or nearly so) enolase amino acid sequences. We have also sequenced α- and β-enolase cDNAs from primitive vertebrates, including chondrichthian, actinopterygian, and sarcopterygian fishes. The presence of multiple, identifiable forms of enolase in these organisms places a lower limit on the age of the duplications and allows predictions of the presence or absence of the duplicated genes in other vertebrates. Finally, we have addressed the timing of the enolase duplication events and the issue of evolutionary rate variation between the duplicated genes.

2. Materials and methods

The following is a list of amino acid sequences obtained from the public databases (all are enolase-1 unless designated otherwise). Archaebacteria: *Archaeoglobus fulgidus*, GenBank Accession No.= AAB90112 (enolase-2); *Aeropyrum pernix*, BAA81473; *Haloarcula marismortui*, P29201 (enolase-2); *Methanobacterium thermoautotrophicum*, AAB84550 (enolase-2); *Methanococcus jannaschii*, Q60173; *Pyrococcus abyssi*, CAB50622, CAB49458 (enolase-2); *Pyrococcus horikoshii*, BAA31069; F71042 (enolase-2). Eubacteria: *Aquifex aeolicus*, AAC06738; *Bacillus subtillus*, P37869; *Borrellia burgdorferi*, O51312; *Campylobacter jejuni*, CAB73659; *Chlamydia muridarum*, AAF39672; *Chlamydia trachomatis*, AAC68189; *Chlamydophila pneumoniae*, AAF38843; *Deinococcus radiodurans*, AAF12173; *Escherichia coli*, P08324, P31458 (galactonate dehydratase); *Haemophilus influenzae*, P43806; *Helicobacter pylori*, AAD05723; *Mycobacterium tuberculosis*, P96377; *Mycoplasma genitalium*, U39723; *Mycoplasma pneumoniae*, P75189; *Neisseria meningitidis*, AAF41661; *Nitrosomonas europaea*, AF061753; *Pseudomonas putida*, A28700 (mandelate racemase), P42206 (glucarate dehydratase); *Staphylococcus aureus*, AF065394; *Streptococcus intermedius*, BAA81815; *Streptococcus thermophilus*, AF027167; *Synechocystis* PCC6803, P77972; *Thermotoga maritima*, AAD35958; *Treponema pallidum*, P74934; *Ureaplasma urealyticum*, AAF30591; and *Zymomonas mobilis*, P33675. Protista: *Entameoba histolytica*, P51555; *Mastigamoeba balamuthi*, AAF13454; *Plasmodium falciparum*, Q27727; *Toxoplasma gondii*, AAD51128. Fungi: *Aspergillus oryzae*, Q12560; *Candida albicans*, P30575; *Cladosporium herbarum*, X78226; *Cunninghamella elegans*, Y17298; *Neocallimastix frontalis*, P42894; *Pneumocystis carinii*, AF063247; *Saccharomyces cerevisiae*, P00924; *Schizosaccharomyces pombe*, P40370. Viridiplantae: *Alnus glutinosa*, Q43321;

Arabidopsis thaliana, P25696; *Chlamydomonas reinhardtii*, P31683; *Lupinus luteus*, CAB75428; *Lycopersicon esculentum*, P26300; *Mesembryanthemum crystallinum*, Q43130; *Oryzae sativa*, Q42971; *Ricinus communis*, P42896; and *Zea mays*, U17973. Invertebrates: *Caenorhabditis elegans*, Q27527; *Drosophila melanogaster*, X17034; *D. pseudoobscura*, AF025805; *D. subobscura*, AF025806; *Homarus gammarus*, P56252; *Loligo pealei*, O02654; *Penaeus monodon*, AAC78141; *Schistosoma japonicum*, P33676; and *Schistosoma mansoni*, U30175. Mammals: *Oryctolagus cuniculus*, P25704 (β-enolase).

The following amino acid and cDNA nucleotide sequences for vertebrates were obtained from the public databases. α-enolases. Amphibians: *Xenopus laevis*, Y00718. Reptiles: *Alligator mississippiensis*, AF072586; *Caiman crocodilus*, AF115855; *Eumeces inexpectus*, AF115857; *Pelusios subniger*, AF115859; *Python regius*, AF072589; *Sceloporus undulatus*, AF072587; *Sphenodon punctatus*, AF115856; and *Trachemys scripta*, AF072588. Birds: *Anas platyrhynchos*, M20749; and *Gallus gallus*, D37900. Mammals: *Bos taurus*, AF149256; *Homo sapiens*, X66610; *Mus* sp., X52379; *Rattus norvegicus*, X02610. β-enolases. Birds: *Gallus gallus*, D37901. Mammals: *H. sapiens*, X56832; *Mus musculus*, X62667; and *R. norvegicus*, Y00979. γ-enolases. Birds: *G. gallus*, AB004291. Mammals: *H. sapiens*, M22349; *Mus* sp., X52380; and *R. norvegicus*, AF019973. The following cyclostome enolase sequences were included: *Eptatretus burgeri*, BAA88479; *Lampetra reissneri*-i, BAA88482; and *Lampetra reissneri*-ii, BAA88483 (the partial lamprey sequence of 'α-enolase', B32132, is identical to that 97-residue region within BAA88483). CLUSTAL X (Thompson et al., 1997) was used to align amino acid and nucleotide sequences.

The nucleotide sequence alignment was then used to construct three sets of primers for reverse transcription and PCR amplification: non-specific (i.e. primers that did not distinguish between paralogous forms of the enolase gene), α-enolase specific primers, and β-enolase specific primers. Specific primers were created by placing the 3′ end of the primer on a site that was diagnostic between α- and β-enolase. Sites that were different between α and β sequences yet identical among α sequences and (separately) among β sequences were considered diagnostic. The non-specific primers and sequences (5′ to 3′) were:

E1:GACTTCAARTCYCCCGATGAYCCCAGCAGRTACAT,
E2:CTGAAGTTYYTKCCRGCAAAGCKGGCCTTGCTGCC,
E3:ATMTTTGAYTCYCGYGGGAAYCCYACWGTKGAGGT,
E4:ATGAAYTCYTGCATRGCCAGCTTRTTGCCAGCATG,
E5:GAGAAGGGWGTCCCMYTGTACCGYCACATYGCTGA,
E6:TGGTCAAAKGGRTCYTCRATRGAYACCACTGGRTA,
E18:GATGGAAAATATGAYCTGGACTTC,
E19:CAGTCTTGATCTGTCCAGTGC,
E20:AARGGTGTCCCRCTGTACCGYCACATTGC,
E21:CCAGTGATCCTGGTCAAAGGGRTCYTC,
E23:CATRAACTCYTGCATGGCCAGCTTGTTGCC,
E24:AARGGKGTCCCMCTGTAYCGYCACATTGC,
E25:ATCCTGGTCAAAGGGRTCTTCAATGG,
E26:AATCTTTGAYTCYCGTGGGAAYCCYAC, and
E27:TGTTGTGGAGCAGGAAAAGATTGA.

The alpha specific primers and sequences (5′ to 3′) were:

AE41:AACCCYACMGTGGAGGTHGACCTST,
AE42:AARGCMGGRACNGGSAGGATGAC,
AE43:TGACCTBGCTGGCAAYCCYGADG,
AE44:TTGTACAGGTCWSCCAGCTGRTCDGG,
AE45:AYGAYYCYAGCMGDTACATCACCCC,
AE46:TGKGCMAKCTTRCASGCCTSCAG,
AE47:CCATCTTGCGGGCAACACACAGG, and
AE49:AGGGTAATTTTTAATGAAACTCGA.

The beta specific primers and sequences (5′ to 3′) were:

BE3:AAYCCYACRGTGGAGGTSGACCTGC,
BE4:AARGCVGGRACVGGVAGGATGAG,
BE5:TGACCTBGCYGGAAAYAMRGAMC,
BE6:TTGTAVAGSTCKCCCAGYTKCTCSCC,
BE7:AYGAYCCYASYCGBYACATCWSYGG
BE9:ATGCGCATGGCCTCGCGGAACGA
BE10:GCGGCGTCCGAGTTCCACCGCAG.

Polyadenylated RNAs [poly(A)$^+$ RNA] were isolated from liver tissue of a brown spotted cat shark (*Chiloscyllium punctatum*; pet trade), South American lungfish (*Lepidosiren paradoxa*; pet trade), and brown trout (*Salmo trutta*; domestic). Polyadenylated RNAs were isolated from skeletal muscle tissue from a brown spotted cat shark, South American lungfish, Australian lungfish (*Neoceratodus forsteri*; locality unknown), coelacanth (*Latimeria chalumnae*; Virginia Institute of Marine Science 8118, Comoros Islands), and brown trout. Polyadenylated RNAs were isolated from a mixture of liver and smooth muscle tissue from a bowfin (*Amia calva*).

Isolation of poly(A)$^+$ mRNA from all samples except bowfin followed the protocol of the Micro-FastTrack mRNA Isolation Kit (Invitrogen) with minor modifications. Initial incubation of the tissue suspension was 30 min. Ethanol precipitates were prepared by adding 10 μl of glycogen (2 mg/ml), 60 μl of 2 M sodium acetate, and 1200 μl of 200 proof ethanol to 400 μl of

elutent containing the poly(A)$^+$ RNA. Isolation of bowfin mRNA followed the instruction protocol of the Oligotex Direct mRNA Isolation Kit (Qiagen).

The GeneAmp RNA PCR core kit (Perkin-Elmer) and downstream, non-specific enolase primers were used to synthesize enolase cDNAs from the poly(A)$^+$ RNA. PCR amplification of the cDNA using a non-specific upstream primer followed the protocol or was modified for the use of Ampli-Taq Gold (Perkin-Elmer). When Ampli-Taq Gold was used, a 10 min pre-PCR step at 95°C was added, and the final step at 72°C was increased from 7 to 10 min. The amplified cDNA was purified on a low melt agarose gel by a method described elsewhere (Hedges et al., 1991). The band core was melted in 300–500 μl of Tris–EDTA buffer (pH 8.0) at 93–95°C for 3 min.

One microliter of the non-specific cDNA was used as template for a double-stranded PCR reaction (25–45 cycles; 94°C for 15 s/55–72°C for 15 s/72°C for 45 s) performed in 25 μl of a solution containing each dNTP at 0.323 mM, bovine serum albumin, 25 mM $MgCl_2$, thermophilic DNA 10× Buffer (Promega), upstream and downstream primers each at 20 μM, and 0.1 unit of Ampli-Taq (*Thermus aquaticus* DNA polymerase; Promega). Alpha- and beta-specific primers were used to amplify the individual paralogs from the non-specific cDNAs. The product of the initial amplification was purified as described above, and 1 μl was used as template for a second amplification conducted under the same conditions in 50 μl of solution. The double-stranded product of the second amplification was purified with four passes in a 30 000 MW filter (Millipore). Cycle sequencing reactions were performed by the Nucleic Acids Facility, The Pennsylvania State University, using 3′ dye-labeled dideoxynucleotide triphosphates. Nucleotide sequences were generated for the forward and reverse strands of each fragment of each gene on an ABI PRISM 377 DNA Sequencer (Perkin-Elmer).

Nucleotide sequences were added to the alignment and were used to deduce amino acid sequences. Neighbor-joining trees (Saitou and Nei, 1987) were constructed with MEGA (Kumar et al., 1993) using the Kimura 2-parameter transitions + transversions (s + v) and transversions only (v) distances (Kimura, 1980) for nucleotides. For amino acid analyses, a Poisson-corrected distance and a gamma distance ($\alpha = 1.05$; Wang et al., 1999) were used with neighbor-joining. A maximum likelihood analysis was conducted on the restricted vertebrate protein data with MOLPHY (Adachi and Hasegawa, 1996) using the JTT-F option. The statistical significance of groups in the neighbor-joining trees was assessed by the bootstrap probability (Felsenstein, 1985) with 2000 replications (Hedges, 1992).

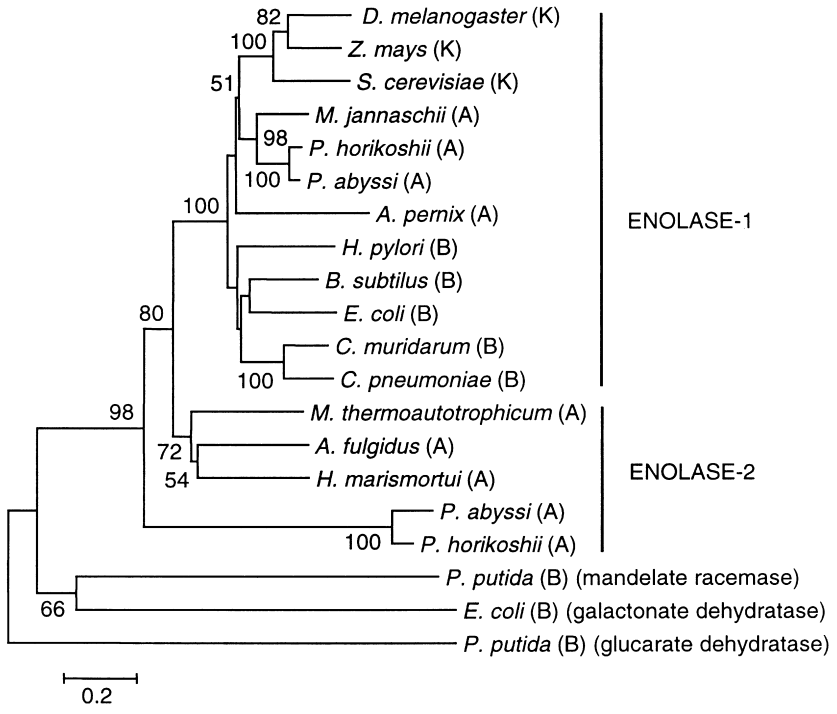

Fig. 1. Phylogenetic tree (neighbor-joining) of enolase and closely related proteins. Archaebacterial enolase amino acid sequences, and representative eubacterial and eukaryote sequences, are included along with other genes of the enolase gene superfamily. Letters following taxon labels indicate the three domains: Archaebacteria (A), Eubacteria (E), and Eukaryota (K). Numbers at nodes are bootstrap *P*-values (≥50%). The tree is rooted by the midpoint of the longest branch.

3. Results

3.1. Global amino acid analyses

It was clear from initial analyses that several archaebacterial enolase sequences were not closely related to the majority of enolase sequences, including other archaebacterial sequences. Therefore, we aligned and compared all archaebacterial sequences with representatives of eubacterial and eukaryote enolases, and with several other genes in the enolase gene superfamily (Fig. 1). The well-defined (100% bootstrap confidence) cluster containing the majority of enolase sequences from all three domains is designated here as 'enolase-1'. The remaining five enolase sequences, all archaebacterial, are more distantly related and form two clusters in the tree. However, the two clusters are not separated by a significant confidence value and may belong to a single gene. Therefore, they are conservatively designated as 'enolase-2' (additional sequences and analyses may resolve an additional gene). The fact that the archaebacterial representatives of enolase-1 belong to two kingdoms (Crenarchaeota and Euryarchaeota) and that both genes (enolase-1 and enolase-2) have been sequenced from the same species (*Pyrococcus abyssi* and *P. horikoshii*) suggests that gene duplication is involved rather than horizontal transfer, although the latter cannot be ruled out. These highly divergent (and, in some cases, shorter) sequences were not included in later analyses to maximize the number of aligned sites.

The tree generated from amino acid sequences of enolase-1 (Fig. 2) shows monophyly of the vertebrate sequences (99%) and of each enolase-1 paralog: α-enolase (93%), β-enolase (95%), and γ-enolase (99%). However, the branching order of the three genes is not significantly resolved. The remaining animals (invertebrates), except the squid (basal), form a cluster that groups with the vertebrate cluster. Animal enolases are monophyletic (<50%). Plant (58%) and fungal (50%) enolases also form monophyletic clusters, and the three kingdoms form a monophyletic group (<50%). The four protists are basal on the monophyletic (99%) eukaryote branch (Fig. 2). The four archaebacteria (95%) and the 24 eubacteria (85%) each form monophyletic groups. The tree was rooted with eubacteria because the eukaryotes appear to be derived by normal vertical evolution from archaebacteria and not by horizontal transfer from proteobacteria (see Doolittle, 1999). Virtually identical topologies were obtained with either a Poisson-corrected distance or a gamma-corrected ($\alpha = 1.05$) distance.

3.2. New vertebrate sequences

Enolase-1 cDNA sequences and deduced amino acid sequences were generated for six taxa: *C. punctatum* (shark) α (GenBank Accession No. AY005159) and β (AY005160), A. calva (bowfin) α (AY005152) and β (AY005153), *S. trutta* (trout) α_1 (AY005161) and α_2 (AY005162), *L. chalumnae* (coelacanth) α (AY005154), L. paradoxa (South American lungfish) β (AY005157), and *N. forsteri* (Australian lungfish) β (AY005158). Partial *L. paradoxa* α (AY005156) and *L. chalumnae* β (AY005155) sequences (not used in analyses) were also generated.

For the vertebrate enolase nucleotide alignment, 1083 (625 variable) sites were aligned across all species, beginning at site 85 of the complete *D. melanogaster* cDNA sequence. The deduced amino acid sequence alignment for the vertebrate enolases began at residue 29 of the complete *D. melanogaster* sequence and contained 360 (217 variable) sites.

The nucleotide and amino acid sequences for *Gallus gallus* β-enolase were obtained from GenBank, but these sequences differed qualitatively from other sequences. At numerous positions in the amino acid alignment, this sequence has an amino acid residue that differs from either all other enolases (except *Drosophila melanogaster*) or all other vertebrate enolases. Furthermore, at site 202, *G. gallus* β-enolase has a histidine (H) residue, which is exceptional among enolases. At this site, *D. melanogaster* and all vertebrate α-enolases share a lysine (K) residue, β-enolases share an asparagine (N) residue, and γ-enolases share a serine (S) residue. It is possible that some of these differences represent sequence errors. Also, the nucleotide sequence for *R. norvegicus* β-enolase from GenBank appeared to contain a sequencing and/or alignment error when placed in the overall enolase alignment. To correct the problem, the adenine at position 471 was removed and replaced by an ambiguity at position 489.

The percentage sequence identity was calculated for all cases in which more than one paralog had been completely sequenced for an individual species. Both nucleotide and amino acid sequences showed a high degree of identity (excluding the *S. trutta* paralogs: 73–80% for nucleotides and 80–86% for amino acids). The trout paralogs showed an exceptionally high identity for both nucleotides (86%) and amino acids (90%).

3.3. Amino acid analyses

A neighbor-joining tree was constructed using amino acid sequences and a gamma-corrected ($\alpha = 1.05$) distance (Fig. 3). A virtually identical topology was obtained with a Poisson-corrected distance. The two trout sequences clustered with other α-enolase sequences, and each paralog was found to be monophyletic: α-enolase (72%), β-enolase (<50%), and γ-enolase (99%). The maximum likelihood analysis resulted in a single ML tree ($\ln L = -5614.7$), also supporting the monophyly of each paralog, with the exception of fowl

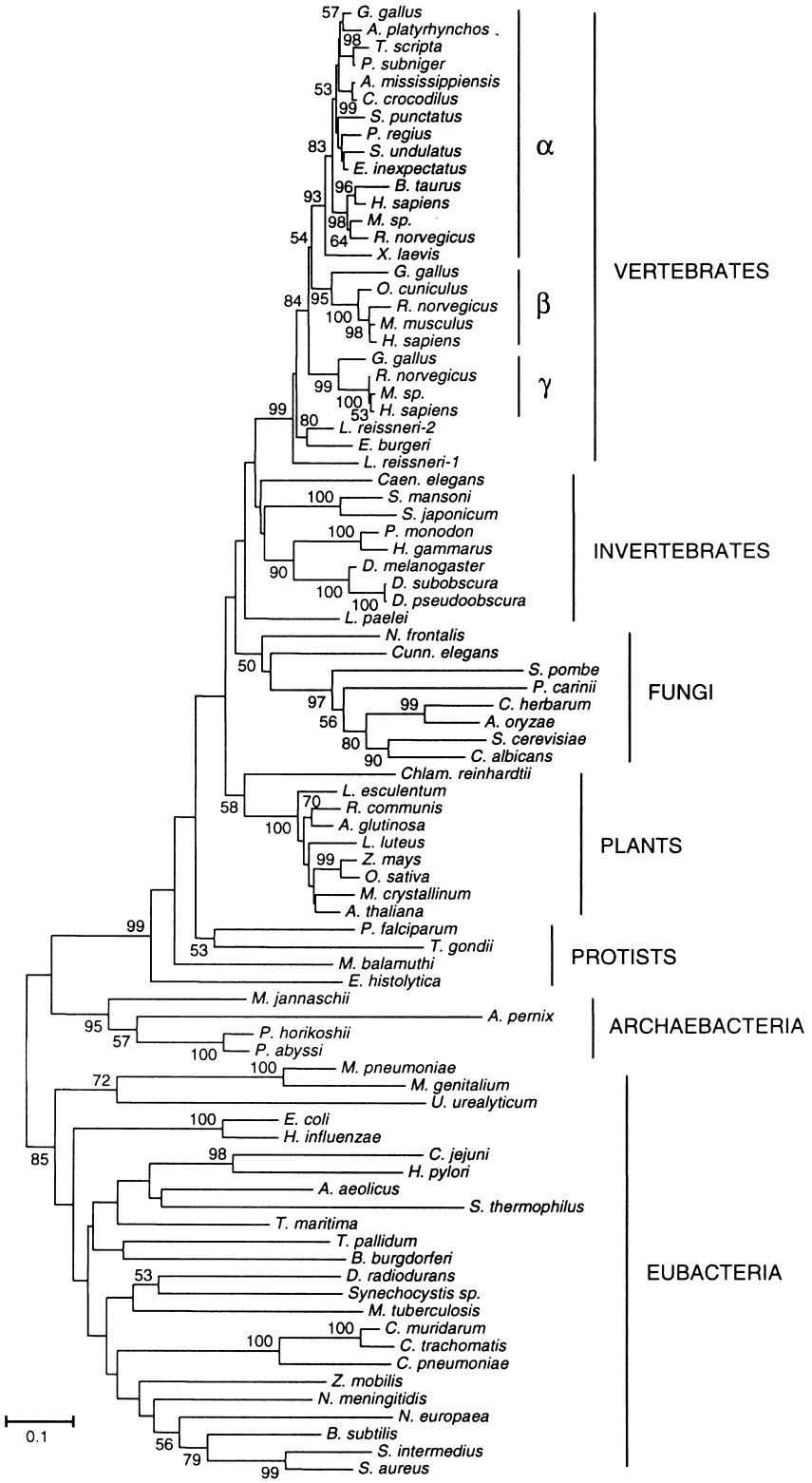

Fig. 2. Phylogenetic tree of enolase-1 amino acid sequences. A gamma-corrected distance ($\alpha = 1.05$) is used with the neighbor-joining method. The tree is rooted with Eubacteria.

(*Gallus*) beta, which appeared in a basal position on the tree (see comments above about anomalies of this sequence). However, the relationships of the three paralogs could not be resolved. The hagfish sequence clustered with the lamprey-2 sequence, as was found by Kuraku et al. (1999), which provides additional molecu-

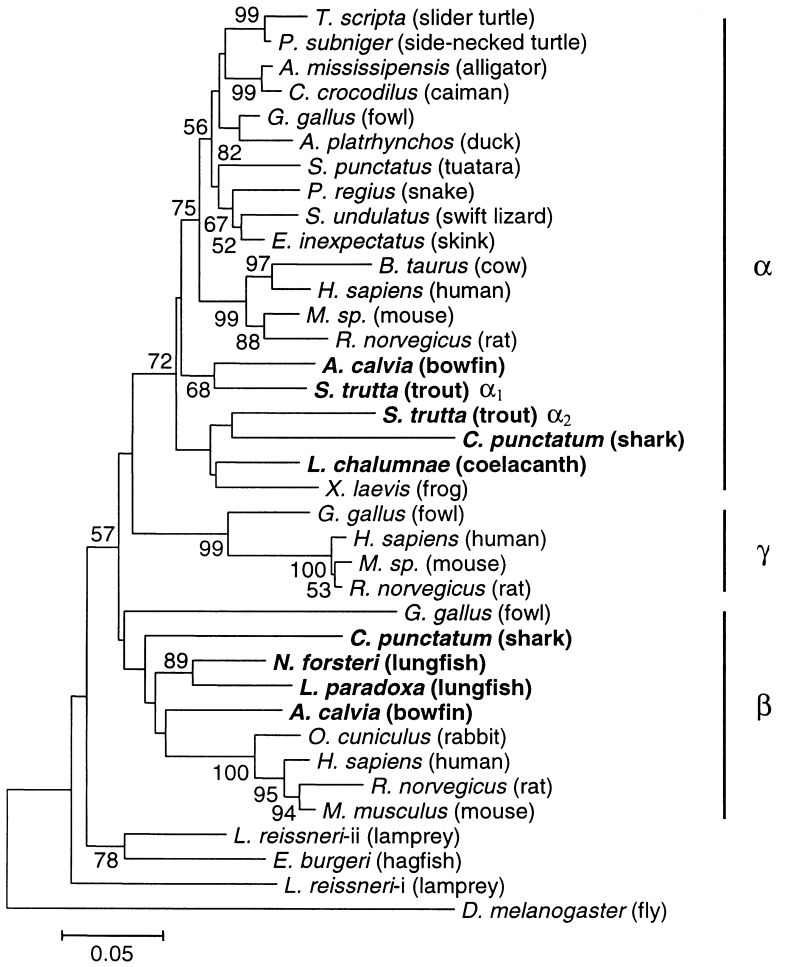

Fig. 3. Phylogenetic tree of vertebrate enolase-1 amino acid sequences. A gamma-corrected distance ($\alpha = 1.05$) is used with the neighbor-joining method. Newly sequenced enolases (this study) are indicated in bold. The tree is rooted with *Drosophila melanogaster*.

lar support for cyclostome monophyly (e.g. Stock and Whitt, 1992a; Mallatt and Sullivan, 1998; Hedges, 2001). Also, their basal position among vertebrate sequences agrees with the phylogenetic position of agnathans and suggests that they diverged prior to the two gene duplications that led to α-, β-, and γ-enolases. However, the presence of two lamprey enolase sequences indicates that an additional gene duplication occurred early in chordate or vertebrate evolution (Kuraku et al., 1999). Although phylogenetic analyses consistently place the agnathan sequences basal to the divergence of the three isozymes (Kuraku et al., 1999; Fig. 2), a possibility exists that they are misplaced on the trees and that the multiple copies of lamprey enolase represent early branches of the isozymes. Relationships among the fishes (alpha and beta paralogs) are unconventional, but none of those nodes have a significant bootstrap support. These details may never be resolved in a statistical sense because of the limitations of gene length, but additional sequencing of basal vertebrates and chordates should improve the resolution of enolase phylogeny.

3.4. Nucleotide analyses

The emphasis in this study was on amino acid analysis because saturation often occurs in nucleotide analyses of distantly-related sequences. None the less, one advantage of nucleotide data is the increased number of sites for analysis, which may provide a better resolution of some aspects of the phylogeny. As in the amino acid tree (Fig. 3), analysis of the nucleotide sequences showed that each of the three paralogs was monophyletic. The two *S. trutta* sequences clustered together significantly and were renamed *S. trutta* α_1 and α_2.

4. Discussion

4.1. Time of enolase gene duplications within vertebrates

A single enolase gene gave rise to the three enolase paralogs through two closely spaced duplication events. The discovery of multiple isozymes of enolase in acti-

nopterygian, sarcopterygian, and chondrichthian fishes constrains the timing of the duplication events. These duplications apparently occurred in the chordate lineage subsequent to the cyclostome-gnathostome divergence (564 Mya) and prior to the chondrichthian–osteichthian divergence (528 Mya) (Kumar and Hedges, 1998; Hedges, 2001). This would place these duplications during a relatively short period of evolutionary time near the Proterozoic/Phanerozoic boundary. A much more recent timing for the duplications, 200–300 Mya, was estimated previously (Rider and Taylor, 1975; Clark-Rosenberg and Marangos, 1980; Segil et al., 1988).

Although the trichotomy remains unresolved, β- and γ-enolase are presumed to be most closely related based on intron lengths (Giallongo et al., 1990). If true, and considering the presence of α- and β-enolase in chondrichthians, all living gnathostomes should possess the three vertebrate enolase paralogs, regardless of the branching pattern for these paralogs.

Besides enolase, isozymes of other genes also show a pattern of gene duplication around the time of the cyclostome–gnathostome divergence (Ono-Koyanagi et al., 2000). It has been proposed that this burst in gene duplication led to functionally similar isozymes, whereas an earlier period of extensive gene duplication, near the origin of animals, led to different functional subtypes of genes (Ono-Koyanagi et al., 2000). Those authors also proposed that recruitment of pre-existing gene duplicates was the molecular mechanism for the Cambrian explosion. However, an additional factor not considered by Ono-Koyanagi et al. (2000) is the possible effect of planetary-scale environmental changes on organismal and molecular evolution. Global glaciation events ('Snowball Earth's') occurred during the Neoproterozoic (750–600 Mya) and almost certainly reduced populations of organisms (including animals) to small numbers of individuals in refugia. Such events may have accelerated adaptive evolution and speciation (Hoffman et al., 1998; Hedges, 2001). A direct connection between these global glaciations and gene duplication, through fixation of chromosomal changes in small populations, is also a possibility. A sufficiently large and diverse data set on the timing of gene duplications in animals is not yet available to test these hypotheses.

4.2. Early evolution of enolase

Horizontal gene transfer can explain many of the complex patterns seen in gene and gene-family relationships involving the three domains of life (Feng et al., 1997; Doolittle, 1999). However, another explanation that should be considered in some cases is a simple lack of phylogenetic resolution caused by the inescapable limitation of protein length. None the less, two patterns commonly observed are (1) a close relationship between eukaryotes and archaebacteria and (2) a close relationship between eukaryotes and eubacteria (α-proteobacteria). The first pattern has been attributed to vertical transmission, whereas the second pattern has been attributed to horizontal gene transfer subsequent to the symbiotic origin of mitochondria from an α-proteobacterium (Doolittle, 1999). Although not universally excepted, this has formed a working hypothesis for the further study of organismal genomes. The relationships of the three domains are not statistically resolved in the global analyses of enolase sequences (Figs. 1 and 2), but the fact that eubacteria are monophyletic and do not cluster closely with eukaryotes favors a vertical transmission for enolase-1. In addition, the finding of two enolase genes (each) in the archaebacteria *Pyrococcus abyssi* and *P. horikoshii* supports the interpretation here of an early gene duplication, prior to the last common ancestor of living organisms, leading to enolase-1 and enolase-2. If this interpretation is correct, enolase-2 was subsequently lost in eubacteria and in some archaebacteria. Thus, the polyphyly of archaebacteria in previous enolase phylogenies (Hannaert et al., 2000; Keeling and Palmer, 2000) is interpreted here as an artefact of rooting.

It has been proposed (Rivera and Lake, 1992), based on some genes showing vertical transmission, that eukaryotes are closer to some archaebacteria (Crenarchaeota) than to other archaebacteria (Euryachaeota). Only one crenarchaeotan enolase sequence is available (*Aeropyrum pernix*), and this clusters closely (95%) with other archaebacteria (Fig. 2). However, it has an unusually long branch, and therefore, additional crenarchaeotan sequences are needed to confirm this phylogenetic position.

4.3. Origin of salmonid enolases

Because the two copies of enolase from trout (a salmonid fish) group with α-enolases in the amino acid analysis (Fig. 3) and cluster with each other in nucleotide analyses suggests that they are α-enolases and the result of a genome duplication event. Salmonid fishes are known to have undergone a polyploidization event with subsequent retention of functional gene duplicates (Bailey et al., 1978). An additional line of evidence suggesting that these two trout enolases are the result of genome duplication is that they have a relatively high percentage sequence identity. The possibility that they represent γ-enolase can be eliminated because neuronal tissue was not used as a source for poly(A)$^+$ mRNA in this study. Likewise, β-enolase seems unlikely because the sequences were obtained from liver tissue where only α-enolase is believed to be expressed.

The identification of these trout enolase sequences as belonging to the α gene and being the result of a polyploidization event explains the results obtained by

Landrey et al. (1978). A minimum of three paralogs is necessary for an organism to exhibit six isoforms of a dimeric enzyme (e.g. αα, αβ, αγ, ββ, βγ, and γγ for enolase). Even though all gnathostomes have been predicted to possess the three standard enolase paralogs, a βγ heterodimer is unlikely, given the tissue-specific localization of these isoforms. Therefore, the result obtained by Landrey et al. (1978) is best explained by genome duplication.

Acknowledgements

We thank Carla A. Hass, Sudhir Kumar, Laura L. Poling, Stephen W. Schaeffer, Marcel van Tuinen, Billie J. Swalla, and Daniel Wang for comments and discussion; Maureen Lyons-Weiler for laboratory assistance; Linda R. Maxson for donating some taxa, including *Latimeria* (courtesy of J.A. Musick, R.W. Chapman, The Explorer's Club, and Virginia Institute of Marine Science) and *Neoceratodus* (courtesy of William Bemis); and Chris T. Amemiya for samples. This work was supported by grants from NSF and NASA (S.B.H.).

References

Adachi, J., Hasegawa, M., 1996. MOLPHY Version 2.3: Programs for Molecular Phylogenetics based on Maximum Likelihood. Institute of Statistical Mathematics, Tokyo.

Babbitt, P.C., Hasson, M.S., Wedekind, J.E., Palmer, D.R., Barrett, W.C., Reed, G.H., Rayment, I., Ringe, D., Kenyon, G.L., Gerlt, J.A., 1996. The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids. Biochemistry 35, 16489–16501.

Bailey, G.S., Poulter, R.T., Stockwell, P.A., 1978. Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. Proc. Natl. Acad. Sci. USA 75, 5575–5579.

Benton, M.J., 1997. Vertebrate Paleontology. second ed., Chapman and Hall, London.

Bishop, J.G., Corces, V.G., 1990. The nucleotide sequence of a *Drosophila melanogaster* enolase gene. Nucleic Acids Res. 18, 191

Brooke, N.M., Garcia-Fernandez, J., Holland, P.W., 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392, 920–922.

Brown, J.R., Doolittle, W.F., 1997. Archea and the prokaryote-to-eukaryote transition. Micro. Mol. Biol. Rev. 61, 456–502.

Clark-Rosenberg, R.L., Marangos, P.J., 1980. Phylogenetic distribution of neuron-specific enolase. J. Neurochem. 35, 756–759.

Crawford, D.L., Constantino, H.R., Powers, D.A., 1989. Lactate dehydrogenase-B cDNA from the teleost *Fundulus heteroclitus*: evolutionary implications. Mol. Biol. Evol. 6, 369–383.

Day, I.N., Peshavaria, M., Quinn, G.B., 1993. A differential molecular clock in enolase isoprotein evolution. J. Mol. Evol. 36, 599–601.

Doolittle, W.F., 1999. Phylogenetic classification and the universal tree. Science 284, 2124–2128.

Felsenstein, J., 1985. Confidence limits in phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Feng, D.F., Cho, G., Doolittle, R.F., 1997. Determining divergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. USA 94, 13028–13033.

Giallongo, A., Oliva, D., Cali, L., Barba, G., Barbieri, G., Feo, S., 1990. Structure of the human gene for α-enolase. Eur. J. Biochem. 190, 567–573.

Gorr, T.A., Mable, B.K., Kleinschmidt, T., 1999. Phylogenetic analysis of reptilian hemoglobins: trees, rates and divergences. J. Mol. Evol. 47, 471–485.

Hannaert, V., Brinkmann, H., Nowitzki, U., Lee, J.A., Albert, M.A., Sensen, C.W., Gaasterland, T., Muller, M., Michels, P., Martin, W., 2000. Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi* and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. Mol. Biol. Evol. 17, 989–1000.

Hardison, R.C., 1996. A brief history of hemoglobins: plant, animal, protist and bacteria. Proc. Natl. Acad. Sci. USA 93, 5675–5679.

Hedges, S.B., Bezy, R.L., Maxson, L.R., 1991. Phylogenetic relationships and biogeography of the xantusiid lizards, inferred from mitochondrial DNA sequences. Mol. Biol. Evol. 8, 767–780.

Hedges, S.B., 1992. The number of replications needed for accurate estimation of the bootstrap p-value in phylogenetic studies. Mol. Biol. Evol. 9, 366–369.

Hedges, S.B., 2000. Molecular evidence for the early history of living vertebrates. In: Ahlberg, E.P. (Ed.), Major Events in Early Vertebrate Evolution: Paleontology, Phylogeny and Development. Taylor and Francis, London, in press

Hoffman, P.F., Kaufman, A.J., Halverson, G.P., Schrag, D.P., 1998. A Neoproterozoic snowball Earth. Science 281, 1342–1346.

Holland, P.W.H., Garcia-Fernandez, J., Williams, N.A., Sidow, A., 1994. Gene duplications and the origins of vertebrate development. Dev. Suppl. 1994, 125–133.

Hughes, A.L., 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.

Keeling, P.J., Palmer, J.D., 2000. Parabasalian flagellates are ancient eukaryotes. Nature 405, 635–636.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Kumar, S., Tamura, K., Nei, M., 1993. MEGA: Molecular Evolutionary Genetic Analysis Version 1.01. The Pennsylvania State University, University Park, PA.

Kumar, S., Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. Nature 392, 917–919.

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.

Landrey, S.R., Applegate, R., Cardenas, J.M., 1978. Enolase isozymes in *Coho salmon*. Comp. Biochem. Physiol. 60, 383–388.

McAleese, S.M., Dunbar, B., Fothergill, J.E., Day, I.N.M., 1988. Complete amino acid sequence of the neuron-specific gamma isozyme of enolase (NSE) from human brain and comparison with the non-neuronal alpha form (NNE). Eur. J. Biochem. 178, 413–417.

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

Mannen, H., Tsoi, S.C.M., Krushkal, J.K., Li, W.H., Li, S.S.-L., 1997. The cDNA cloning and molecular evolution of reptile and pigeon lactate dehydrogenase isozymes. Mol. Biol. Evol. 14, 1081–1087.

Nadeau, J.H., Sankoff, D., 1997. Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics 147, 1259–1266.

Oliva, D., Barba, G., Barbieri, G., Giallongo, A., Feo, S., 1989. Cloning, expression and sequence homologies of cDNA for human gamma enolase. Gene 79, 355–360.

Ono-Koyanagi, K., Suga, H., Katoh, K., Miyata, T., 2000. Protein tyrosine phosphatases from amphioxus, hagfish, and ray: diver-

gence of tissue-specific isoform genes in the early evolution of vertebrates. J. Mol. Evol. 50, 302–311.

Pegg, S.C., Babbitt, P.C., 1999. Shotgun: getting more from sequence similarity searches. Bioinformatics 15, 729–740.

Piatigorsky, J., 1989. Lens crystallins and their genes: diversity and tissue-specific expression. FASEB J. 3, 1933–1940.

Piatigorsky, J., Wistow, G.J., 1989. Enzyme/crystallins: gene sharing as an evolutionary strategy. Cell 57, 197–199.

Rider, C.C., Taylor, C.B., 1975. Enolase isoenzymes. II. Hybridization studies, developmental and phylogenetic aspects. Biochim. Biophys. Acta 405, 175–187.

Rivera, M.C., Lake, J.A., 1992. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science 257, 74–76.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Schmechal, D.E., Brightman, M.W., Marango, P.J., 1984. Neurons switch from noneuronal enolase to neuron-specific enolase during differentiation. Brain Res. 190, 195–214.

Segil, N., Shrutkowski, A., Dworkin, M.B., Dowrkin-Rastl, E., 1988. Enolase isozymes in adult and developing *Xenopus laevis* and characterization of a cloned enolase sequence. Biochem. J. 251, 31–39.

Stock, D.W., Whitt, G.S., 1992a. Evidence from 18S ribosomal RNA sequences that lampreys and hagfishes form a natural group. Science 257, 787–789.

Stock, D.W., Whitt, G.S., 1992b. Evolutionary implications of the cDNA sequence of the single lactate dehydrogenase of a lamprey. Proc. Natl. Acad. Sci. USA 89, 1799–1803.

Stock, D.W., Quattro, J.M., Whitt, G.S., Powers, D.A., 1997. Lactate dehydrogenase (LDH) gene duplication during chordate evolution: the cDNA sequence of the LDH of the tunicate *Styela plicata*. Mol. Biol. Evol. 14, 1273–1284.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Wang, D.Y.C., Kumar, S., Hedges, S.B., 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals, and fungi. Proc. R. Soc. Lond. B 266, 163–171.

Williams, N.A., Holland, P.W.H., 1998. Gene and domain duplication in the chordate *Otx* gene family: insights from amphioxus *Otx*. Mol. Biol. Evol. 15, 600–607.

Wistow, G.J., Lietman, T., Williams, L.A., Stapel, S.O., de Jong, W.W., Horwitz, J., Piatigorsky, J., 1988. τ-crystallin/α-enolase: one gene encodes both an enzyme and a lens structural protein. J. Cell. Biol. 107, 2729–2736.