Ascidian and Amphioxus *Adh* Genes Correlate Functional and Molecular Features of the ADH Family Expansion During Vertebrate Evolution

Cristian Cañestro,¹ Ricard Albalat,¹ Lars Hjelmqvist,^{1,2} Laura Godoy,¹ Hans Jörnvall,²
Roser Gonzàlez-Duarte¹

¹ Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, E-08028 Barcelona, Spain

² Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17177 Stockholm, Sweden

Received: 10 April 2001 / Accepted: 23 May 2001

Abstract. The alcohol dehydrogenase (ADH) family has evolved into at least eight ADH classes during vertebrate evolution. We have characterized three prevertebrate forms of the parent enzyme of this family, including one from an urochordate (*Ciona intestinalis*) and two from cephalochordates (*Branchiostoma floridae* and *Branchiostoma lanceolatum*). An evolutionary analysis of the family was performed gathering data from protein and gene structures, exon–intron distribution, and functional features through chordate lines. Our data strongly support that the ADH family expansion occurred 500 million years ago, after the cephalochordate/vertebrate split, probably in the gnathostome subphylum line of the vertebrates. Evolutionary rates differ between the ancestral, ADH3 (glutathione-dependent formaldehyde dehydrogenase), and the emerging forms, including the classical alcohol dehydrogenase, ADH1, which has an evolutionary rate 3.6-fold that of the ADH3 form. Phylogenetic analysis and chromosomal mapping of the vertebrate *Adh* gene cluster suggest that family expansion took place by tandem duplications, probably concurrent with the extensive isoform burst observed before the fish/tetrapode split, rather than through the large-scale genome duplications also postulated in early vertebrate evolution. The absence of multifunctionality in lower chordate ADHs and the structures compared argue in favor of the acquisition of new functions in vertebrate ADH classes. Finally, comparison between *B. floridae*

and *B. lanceolatum* *Adhs* provides the first estimate for a cephalochordate speciation, 190 million years ago, probably concomitant with the beginning of the drifting of major land masses from the Pangea.

Key words: Alcohol dehydrogenase — Evolutionary rate — Gene duplication — Vertebrate genome evolution

Introduction

Major increases in the number of genes have delineated key steps during evolution (for review see Lundin 1999). It has been suggested that extensive gene duplications occurred by genome tetraploidization after the cephalochordate/vertebrate split and that this provided the flexibility to build the more complex vertebrate body plan (Garcia-Fernández and Holland 1994). Moreover, gene tandem duplications (isoform burst) have been postulated to have occurred at about the same time as or immediately before the cyclostome/gnathostome split (Sharman and Holland 1996; Suga et al. 1999). The expansion of single-copy genes in vertebrates has been dated in a variety of gene families (for reviews see Lundin 1993; Skrabanek and Wolfe 1998). Irrespective of the mechanism, duplications are responsible for the extant gene families and isoforms from fish to mammals (Ohno 1970; Lundin 1993) and are believed to be related to the gains in complexity which have allowed the acquisition of new functions through enzymogenesis

(Danielsson and Jörnvall 1992) and changes in gene regulation (for a review see Shimeld 1999).

Comparisons with extensive sets of sequences from several gene families are being made to elucidate how vertebrate genomes evolved (Wang et al. 1998; Kumar and Hedges 1998; Hughes 1999; Wang and Gu 2000). However, combination of the sequence data with the exon-intron structure, chromosomal location, expression domains, and other functional features would provide a more comprehensive scenario to understand gene and genome evolution.

The alcohol dehydrogenase (ADH) family belongs to a complex protein system with extensive multiplicity within the medium-chain dehydrogenase/reductase (MDR) superfamily. Different classes and isozymes of ADHs are found in the vertebrate lineage. In mammals, at least six ADH classes are known (Jörnvall and Höög 1995; Duester et al. 1999), of which five have been characterized in human. Of these, classes 1 and 3 have been most investigated in different animal clades. The class 1 enzyme, the "classical" ethanol-active form, has evolved rapidly and exhibits a considerable variability among species (Danielsson et al. 1994a). ADH class 3 [glutathione-dependent formaldehyde dehydrogenase (Koi-vusalo et al. 1989)] is present in prokaryotes and in all eukaryotes and appears to be the ancestral form from which the other vertebrate ADH classes emerged (Danielsson and Jörnvall 1992; Danielsson et al. 1994a; Cañestro et al. 2000) through gene duplications during early vertebrate radiation. Supporting evidence comes from the apparent absence of an ethanol-active class 1 form in lines originating before the bony fish (Danielsson et al. 1994b) and the presence of a class 1/3-mixed form in bony fish (Danielsson et al. 1992).

To study ADH gene family expansion, we have compared the vertebrate gene classes with those found in two prevertebrates, ascidians and amphioxus. Ascidians, subphylum Urochordata, are considered the simplest chordates and a critical evolutionary link between invertebrates and vertebrates. They show compact genomes which are comparable in size to those of *Drosophila* and *Caenorhabditis elegans*. On the other hand, amphioxus, a member of the subphylum Cephalochordata, is thought to be the closest living relative to vertebrates (García-Fernández and Holland 1994; Holland 1996) and therefore to be an ideal outgroup for evolutionary and phylogenetic studies of vertebrate gene families. We have characterized the *Adh3* genomic structure of the urochordate *Ciona intestinalis*, and the cephalochordates *Branchiostoma floridae* and *Branchiostoma lanceolatum*. Analyses of gene structure, amino acid substitutions, biochemical data, and expression domains strongly suggest that class divergence took place after the vertebrate/cephalochordate split. We estimate the first gene duplication, leading to new forms of the enzyme, to have occurred about 500 million years ago (MYA), in the

early vertebrate evolution before the fish/tetrapode split. Changes in the evolutionary rate of the ancestral and the newly generated forms can be correlated with the functional evolution of vertebrate ADHs, where the ancestral ADH3 has remained biochemically constant from insects, via amphioxus to the human enzyme (Danielsson et al. 1994a; Cañestro et al. 2000), and the newly generated forms have acquired novel activities. Finally, the chromosomal location in human and mice supports extensive isoform generation by tandem duplications rather than by large-scale genome duplications for the expansion of the ADH family.

Materials and Methods

Library Screenings and Southern Blot Analysis

A *B. lanceolatum* genomic library (Cañestro et al. 2000) and a *B. floridae* genomic library (kindly provided by J. García-Fernández) were probed with [α - 32 P]dCTP. *B. floridae Adh3* cDNA labeled by random-hexamer priming. Total RNA from *C. intestinalis* was purified by the guanidinium thiocyanate method (Chomczynski and Sacchi 1987). Synthesis of cDNA was performed with MLV/RT (Promega) and random-hexamer priming. Degenerate oligonucleotides (GA(AG)GC(ATCG)TG(TC)CA(TC)AA(AG)GG(ATCG)TGG and CC(AG)AA(ATCG)GC(ATCG)GT(ATCG)CC(CT)TTCCA) were designed to amplify a *C. intestinalis* cDNA fragment of the *Adh3* gene, and this fragment was used to screen a genomic library (kindly provided by M. Levine) and a cDNA library (kindly provided by R. Di Lauro).

Hybridizations for library screenings were carried out in phosphate-sodium dodecyl sulfate (SDS) solution (Church and Gilbert 1984) at 65°C overnight. Washes were done at 65°C for 2 × 15 min in 2× standard saline citrate (SSC), 0.1% SDS, and 1 × 15 min in 1× SSC, 0.1% SDS. DNA fragments from positive recombinant phages were isolated, subcloned into a pUC18 vector, characterized by restriction mapping, and sequenced in an ABI-Prism 377 DNA sequencer from PE Biosystems.

Sequence Data

The evolutionary distance, k_{aa} , among different groups of organisms was calculated using the sequence data from the SWISS-PROT, TrEMBL, EMBL and GenBank data banks. Accession numbers of sequences are as follow.

ADH Class 3. Primates: *Homo sapiens* (P11766). Rodents: *Mus musculus* (P28474); *Rattus norvegicus* (P12711); *Oryctolagus cuniculus*, rabbit (O19053). Mammals: Primates + rodents + *Equus caballus*, horse (P19854). Reptiles: *Uromastix hardwickii*, lizard (P80467). Tetrapods: Mammals + reptiles. Fish: *Sparus aurata*, gilthead sea bream (P79896); *Gadus morhua callarias*, cod (P81601). Gnathostomes: Tetrapods + fish. Hagfish: *Myxine glutinosa* (P80360). Vertebrates: hagfish + gnathostomes. Cephalochordates: *Branchiostoma lanceolatum* (AF156698–AF156708); *Branchiostoma floridae* (AF154331, AF344170, AF344171). Urochordates: *Ciona intestinalis* (AF344173). Deuterostomes: Vertebrates + cephalochordates + urochordates.

ADH Class 1. Primates: *Homo sapiens* (A, P07327; B, P00325; C, P00326); *Papio hamadryas*, baboon (P14139); *Macaca mulatta*, rhesus

macaque (P28459). Rodents: *Mus musculus* (P00329); *Rattus norvegicus* (P06757); *Peromyscus maniculatus*, deer mouse (P41680); *Geomys bursarius* (Q64413); *Geomys knoxjonesi* (Q64415); *Oryctolagus cuniculus*, rabbit (Q03505). Mammals: Primates + rodents + *Equus caballus*, horse (E, P00327; S, P00328). Birds: *Gallus gallus*, chicken (P23991); *Coturnix coturnix japonica*, quail (P19631); *Apteryx australis*, kiwi (P49645); *Struthio camelus*, ostrich (P80338). Reptiles: *Uromastix hardwickii*, Indian spiny-tailed lizard (A, P25405; B, P25406); *Alligator mississippiensis* (P80222); *Naja naja*, cobra (P80512). Amphibians: *Rana perezi*, frog (P22797). Tetrapods: Mammals + birds + reptiles + amphibians. Fish: *Gadus morhua callarias*, cod (P26325).

ADH Class 2. *Homo sapiens* (P08319), *Mus musculus* (Q9QYY9).

ADH Class 4. *Homo sapiens* (P40394); *Mus musculus* (Q64437).

Alignments, Evolutionary Distances, and Relative Rate Tests

Alignments of amino acid sequences, from Ile at position 6 to Arg at position 369 of human ADH3, were generated with the ClustalX program (Thompson et al. 1997) using default gap penalties, with adjustments made by eye.

The k_{aa} value of amino acid substitutions per site (evolutionary distance) was calculated with the MEGA (Molecular Evolutionary Genetic Analysis) package (Kumar et al. 1993) for the amino acid γ distance, with an α parameter obtained with PUZZLE version 4.0.1 (Strimmer and Haeseler 1996). This same quartet puzzling procedure was used for the maximum likelihood calculations, following the JTT model for amino acid substitutions (Jones et al. 1992), estimating the amino acid frequencies from the data set and taking into account the among-site rate heterogeneity. When the evolutionary distances between groups were based on more than one species, the average value of k_{aa} was used.

Rate differences among lineages were analyzed by the “two-cluster” and “branch-length” tests provided in the Lintre package (Takezaki et al. 1995). Both tests were applied to a phylogenetic tree according to the widest accepted phylogeny (Kuma and Miyata 1994; Naylor and Brown 1997) and the γ -corrected distribution for multiple replacements.

Results

Characterization of *C. intestinalis*, *B. floridae*, and *B. lanceolatum* Adh3 Genes

After library screening, restriction map analysis of *B. lanceolatum*, *B. floridae*, and *C. intestinalis* showed that in each case all the phages overlapped the same *Adh3* genomic region. A *B. lanceolatum* 13-kb phage containing the complete *BlAdh3* coding region was analyzed further, and intron/exon boundaries were sequenced (AF156698–AF156708). The same analysis was performed with a 15-kb phage of *B. floridae* (AF266713–AF266719) and a 6-kb phage of *C. intestinalis* (AF344173), both encompassing the entire *BfAdh3* and *CiAdh3* coding regions, respectively.

Exon–Intron Structure of Ascidian and Amphioxus ADH3

The exon–intron organization of *CiAdh3*, *BlAdh3*, and *BfAdh3* was established by comparing the corresponding

genomic clones with the *C. intestinalis* (AF344172) and *B. floridae* (AF154331) cDNA (Fig. 1A). The *CiAdh3*, *BlAdh3*, and *BfAdh3* genes consisted of 11, 11, and 10 exons, respectively. Intron sizes differed greatly between the two subphyla; the average amphioxus intron length was 0.9 ± 0.6 kb, whereas it was shorter for the ascidian, 0.2 ± 0.1 kb. In comparisons of the *Adh3* exon/intron architecture of protostomes and deuterostomes, prevertebrates showed the highest number of introns (10 and 9 of the 12 introns identified; Fig. 1B). Introns 1, 2, 7, and 9 were conserved in at least one protostome and all deuterostome species. Intron 2 showed an identical 12-nucleotide (nt) slippage in the fruit fly (Luque et al. 1994) and, in ascidians, the latter with an additional 10-nt slippage in intron 4. *Drosophila subobscura* is the only protostome where intron 7 has been identified (Abad 2000). What had been considered a unique intron 6 in *C. elegans* could be regarded as a 29-nt slippage of intron 7. Introns 3, 4, 11, and 12 seemed to be restricted to the deuterostome lineage. *B. floridae* lacked intron 11, although it was present in *C. intestinalis*, *B. lanceolatum*, and all *Adh* vertebrate genes. Moreover, introns 5 and 8 appeared to be restricted to prevertebrates, as was intron 10 to the amphioxus lineage. Interestingly, vertebrate ADH classes 1, 2, 3, and 4 shared the same exon–intron organization (introns 1, 2, 3, 4, 7, 9, 11, and 12) (Fig. 1B). This, together with the evolutionary and functional data, suggests that class divergence took place after the vertebrate/cephalochordate split.

Evolutionary Distances and Amino Acid Substitution Rates of Chordate ADH3 and ADH1

The evolutionary distance k_{aa} was calculated for chordate ADH3 and ADH1 with a γ -corrected distribution. As this method takes into account the nonuniform among-site rate variation pattern of each enzyme, it is considered the most pertinent. The low values for the α parameter, 0.5 and 0.6 for class 3 and class 1, respectively, were in agreement with the structural pattern of ADHs, made of highly variable regions embedded in conserved segments along the polypeptide chains (Danielsson et al. 1994a). The approach to saturation of amino acid substitutions was analyzed for ADH3 by plotting the frequency of observed differences vs estimated substitutions (Philippe et al. 1994) (Fig. 2A). Although the curve starts to level off, it does not reach a plateau, meaning that the ADH3 sequences were informative and not mutationally fully saturated. Moreover, we compared the γ -distance vs ML estimates (Roger et al. 1999) and very similar values were obtained (Fig. 2B).

To evaluate the constancy of the ADH class 3 amino acid substitution rate in vertebrates and cephalochordates, the “two-cluster” and the “branch-length” statistical tests were performed, using the ascidian ADH sequence as the outgroup. The χ^2 values that combined all

Fig. 1. Exon-intron architecture. **A** Exon (boxed) and intron (non-boxed) distribution in *B. floridae* (top), *B. lanceolatum* (middle), and *C. intestinalis* (bottom) class 3 *Adh*. The exon number is shown over the boxes and the figures below indicate the amino acid position before the intron. Intron sizes (kb) are shown above the intron line. **B** Schematic

comparison of the exon-intron architecture for ADH classes in animal species. Arrowheads indicate the intron positions relative to the *C. intestinalis* coding region. Introns are numbered sequentially (1 to 12). Intron slippage is shown by the asterisk.

interior nodes and sequences showed no significant rate differences within vertebrates and cephalochordates ($p > 0.05$) (Table 1). Relative rate tests were also performed for the ADH class 1 enzyme. In this case, the amphioxus class 3 was used as the outgroup and a neighbor-joining tree topology was built. From the initial 23 sequences, 19 showed a constant rate of evolution (Table 1).

Amino acid substitution rates for ADH3 and ADH1 were estimated (Table 2). Five pairs of groups were established: (a) primates/rodents, (b) reptiles/mammals, (c) amphibians/amniotes, (d) fish/tetrapods and (e) hagfish/gnathostomes. Divergence times (T) for each pair deduced from fossil records are 65–100 MYA for a, 288–310 MYA for b, 350 MYA for c, 400 MYA for d, and 555 MYA for e (Dayhoff 1978; Archibald 1996; Benton 1997; Lee 1999; Shu et al. 1999). Multiple calibration points were used to minimize errors (Lee 1999) because

too few sequences were available for some animal groups. The average evolutionary distance for each pair was plotted against the minimal divergence time. A significant linear dependence of k_{aa} on T was observed, with a correlation coefficient of 0.97 and 0.82 for ADH classes 3 and 1, respectively. From the slope of the regression line, the amino acid substitutions per site per year (v_{aa}) for ADH3 and ADH1 were estimated to be 0.27×10^{-9} and 0.98×10^{-9} , respectively. This 3.6-fold difference could not be assigned to differential mutation rates, since both genes showed similar rates of synonymous substitutions at the DNA level (data not shown).

Class 3/1 Duplication

The amino acid differences between class 3 and class 1 and their evolutionary rates were used to estimate the

Fig. 2. Estimates of the number of amino acid substitutions per site in ADH class 3 enzymes. **A** Substitution saturation curve. The y axis shows the frequency of observed differences between pairs of species sequences, and the x axis shows the inferred distance between the same two sequences determined by maximum likelihood. Each point therefore represents one pair of sequences, considering the observed over the inferred number of substitutions. The resulting curve lies between the diagonal line (no saturation) and a horizontal plateau line (fully saturated), which means that the ADH3 set is informative and only moderately saturated. **B** Comparison of the numbers of ADH3 substitutions per site estimated by the γ -distance correction (y axis) versus those estimated by the maximum likelihood method (x axis).

class duplication time. The amino acid distances between both classes were calculated using a γ -corrected distribution with an α of 0.8. The α value was slightly higher than that obtained for each particular class ($\alpha = 0.5$ for ADH3 and 0.6 for ADH1), due to the different arrangements of the conserved regions in the proteins compared. Since class 1 has not been found in hagfish, its ADH3 data were not used for the comparison. From all the comparisons of each class 3 vs class 1 from codfish to

Table 1. ADH class 3 and class 1 tests for homogeneity of divergence rates

		Test	χ^2 ^a	df	p value
ADH3	Vertebrates/cephalochordates	Two-cluster	10.08	14	0.69
		Branch-length	10.85	14	0.62
ADH1	Vertebrates ^b	Two-cluster	13.77	18	0.74
		Branch-length	23.85	18	0.16

^a Combines the rate differences for all the interior nodes (n) under the root, with $n - 1$ degrees of freedom (df).

^b The rat, deer mouse, and lizard are excluded.

human, an average k_{aa} was obtained. The duplication time, estimated as $T = k_{aa}/(v_{aa}\text{-ADH3} + v_{aa}\text{-ADH1})$, was 493–500 MYA. Therefore, the ADH family would have expanded after the cephalochordate/vertebrate split [682–694 MYA for ADH3 (our estimate) (Table 2) and 680–700 MYA for triosephosphate isomerase and 790–860 MYA for aldolase C (Nikoh et al. 1997)].

Discussion

Evidence in Favor of the ADH Family Expansion After the Cephalochordate/Vertebrate Split

Intron positions are relevant to elucidate gene duplication events. According to our analysis, all vertebrate ADH classes share an identical nine-exon structure (Zgombic-Knight et al. 1995; Strömberg and Höög 2000) and show marked differences from that of prevertebrates (Fig. 1B). Several hypothetical scenarios for the ancestral exon structure and the duplication events through chordate evolution have been considered (Fig. 3). Consistent with the most parsimonious process, the ancestral gene structure contained 11 exons and the duplication events leading to the various vertebrate classes took place after the cephalochordate/vertebrate split (Fig. 3A). Moreover, under this hypothesis, no recurrent events would have to be postulated. Concerning the *Adh3* architecture, ascidians and amphioxus show a higher number of introns but of a smaller size than those of vertebrates, in agreement with other reported genes (Boeddrich et al. 1999). With regard to protostomes, *Adh3* exhibits a lower intron number than deuterostomes and introns of a smaller size. More protostome *Adh* sequences should be characterized to infer the ancestral metazoan structure.

Estimates of the amino acid distances between the vertebrate ADH3 and the vertebrate ADH1 classes and of the evolutionary rates make it possible to infer the duplication event to be at about 500 MYA (Table 2), clearly after the cephalochordate/vertebrate split (682–694 MYA), probably subsequent to the hagfish divergence (555 MYA) and before the origin of tetrapods (400

Table 2. Evolutionary distances and divergence times among different animal groups estimated from the amino acid sequences of ADH3; Duplication time of the class 3 and class 1 enzymes

	Divergence time (MYA) ^a	γ distance (k_{aa})	
	Fossils (calibration points)	ADH3 ($\alpha = 0.5$)	ADH1 ($\alpha = 0.6$)
Animal groups			
Primates/rodents	65–100	0.06	0.21
Reptiles/mammals	288–310	0.12	0.41
Amphibians/amniotes	350	—	0.53
Fish/tetrapods	400	0.23	1.03
Hagfish/gnathostomes	555	0.30	—
	ADH3 (molecular estimations) ^b		
Cephalochordates/vertebrates	682–694	0.37	—
<i>B. floridae</i> / <i>B. lanceolatum</i>	186–189	0.10	—
Correlation coefficient ^c		0.97–0.96	0.82–0.81
ν_{aa} (10^{-9} /site/year) ^d		0.27–0.27	0.98–0.96
Duplication time, class 3 and class 1 ($\alpha = 0.8$)	493–500	0.62	

^a The two calibration points separated by a dash constitute the minimal confidence value from the fossil record and the widest divergence time accepted as a calibration point for each animal group (Lee 1999). Those calculations that were based on both values are separated by a dash.

^b The molecular estimates were derived from the constant rate of amino acid substitution of ADH3.

^c The regression line was constrained to pass through the origin.

^d The evolutionary rate ν_{aa} was estimated from the slope of the regression line.

MYA). Biochemical data also support this duplication to be restricted to the vertebrate lineage since no ethanol-active class 1 form has been detected in amphioxus (Cañestro et al. 2000) and hagfish (Danielsson et al. 1994b).

Amphioxus ADHs: B. floridae–B. lanceolatum Speciation

Remarkably, *Adh3* is more similar between human and mouse than between the two amphioxus species analyzed, *B. floridae* and *B. lanceolatum*. Concerning gene structure, vertebrate species share all intron positions but this is not the case for intron 11 in amphioxus, although the other introns appear to be fully conserved (Fig. 1). At the protein level, amino acid differences amounted to 9.3% for the amphioxus species, whereas they were only 7.2% for human and mouse. The constant evolutionary rate of ADH3 makes it possible to infer that the two amphioxus species may have diverged around 190 MYA (Table 2), during the midsuperior Jurassic. Although such a high value would be unexpected on morphological grounds, since the two species appear very similar, it would be consistent with the human and mouse divergence time of 65–100 MYA. Moreover, similar data were obtained with cytoplasmic and muscle actin (data not shown). Speciation of the two cephalochordates—*B. floridae* lives along the Atlantic coast of America, whereas *B. lanceolatum* is found along the Atlantic coast of Europe and in the Mediterranean sea (Poss and Boschung 1996)—could have been a consequence of the drifting of major land masses, beginning around 200 MYA

from the Pangea (Marzoli et al. 1999). Therefore, it could be argued that present similarities in the two amphioxus species would not be based on recent divergence but, rather, on the absence of major morphological changes during the last 200 million years (Myr) or thereabouts. In fact, this morphological stability supports the “archetypal” organization postulated for amphioxus (García-Fernández and Holland 1994).

Relationship Between ADH Family Expansion and Vertebrate Genome Evolution

How does the ADH family expansion fit into the predicted genome duplications (the 2R hypothesis) in early vertebrate evolution? Paralogous chromosomal locations of some gene family members of the human and mouse have been argued to constitute remnants of the two or three rounds of tetraploidization (Lundin 1993). However, genes encoding the ADHs are not spread over different chromosomes but cluster in the human genome at 4q21–25 and in the syntenic region of mouse chromosome 3 (Giri et al. 1989). Furthermore, the 2R hypothesis would predict a (AB)(CD) phylogenetic tree topology for the four resulting genes (Hughes 1999), although this does not always hold true for all gene families that were generated during early vertebrate evolution (Wang and Gu 2000). Phylogenetic analysis of the ADH family rendered an (A(B(CD))) topology, also in agreement with the physical map of the ADH members on the human chromosome 4 (Fig. 4). In this topology, the pattern of family expansion fits better with tandem duplication

Fig. 3. *Adh* intron gains and losses assuming that the family expansion took place after (A, C, E), or before (B, D, F), the cephalochordate/vertebrate split. The putative ancestral exon structure is shown at the root of each tree [A and B, 11 exons (ascidians); C and D, 12 exons (maximum number); E and F, 9 exons (vertebrates)] and the actual structure at the end of each branch. The duplication event and the

cephalochordate/vertebrate split are shown (filled circles and bold lines, respectively). The assumed intron gains (▼) and losses (▲) are indicated (boxed), and the exons involved are labeled (shaded boxes and bold face numbers). Recurrent losses/gains in different lineages are marked with asterisks.

events than with whole or partial genome rearrangements assumed under the 2R hypothesis. Extensive gene duplications are thought to be responsible for the burst of isoforms during early vertebrate evolution, before the fish/tetrapode split, and probably these duplications were not directly linked to the Cambrian explosion (Suga et al. 1999).

Functional Divergence and Evolutionary Changes

Gene duplications allow the generation of new functions (Ohno 1970; Hughes 1994). In the ADH family, the data

gathered on the structural and enzymological features of the different isoforms support the orthologous relationships derived from the phylogenetic analysis and provide valuable clues to discriminate between the ancestral and the new functions. The amphioxus ADH is a typical class 3 enzyme (Cañestro et al. 2000), as are the octopus (Fernandez et al. 1993), the *Drosophila* (Luque et al. 1994), and the hagfish (Danielsson et al. 1994b) forms of the enzyme. Hence, all direct analyses provide evolutionary evidence that the ancestral form of the family was endowed with ADH3 activity. Therefore, we have named

Fig. 4. Phylogenetic representation and chromosomal mapping of the *Adh* family. **Left:** γ -distance neighbor-joining tree for human/mouse *ADH* classes 1, 2, 3 and 4, rooted with the amphioxus *Adhs*. Bootstrap values are indicated for each branch. **Center:** (A(B(CD))) phylogenetic pattern of the *Adh* family. **Right:** Map position of the *ADH* gene cluster on human chromosome 4: (A) *ADH3*(χ), 420,780–404,222 bp; (B) *ADH2*(π), 476,032–453,229 bp; (C) *ADH1A*(α), 622,843–608,229 bp; *ADH1B*(β), 684,498–638,355 bp; (D) *ADH4*(μ), 761,336–744,028 bp on the reverse strand of the contig NT_022863. Information gathered at <http://www.ncbi.nlm.nih.gov/AceView/> ADH* search. The ADH nomenclature is according to Duester et al. (1999).

the ascidian and amphioxus *Adh* genes *Adh3*. The enzymatic activity expressed as k_{cat}/K_m for this enzyme remains constant—within a factor of 3—in insects, amphioxus, and human. Moreover, the amino acid substitution rate for *ADH3* remains unchanged after the duplication event, whereas *ADH1* shows a 3.6-fold increase, also supported by their different branch lengths in the neighbor-joining tree (Fig. 4). Thus, the “functional density,” defined as the fraction of a polypeptide chain involved in specific functions, which is inversely related to the evolutionary rate (Zuckerandl 1976), could be assumed to be higher in *ADH3* than in *ADH1*, in accordance with the classical view of a “variable” class 1 and a “constant” class 3 enzyme (Danielsson et al. 1994a). Moreover, while the segments of maximal variation of class 1 are at functionally important regions, class 3 variability is restrained to superficial, less critical positions (Persson et al. 1993; Danielsson et al. 1994a). In summary, the present data strengthen the concept of an ancestral position for the *ADH3* activity.

Some of the newly generated forms deserve further consideration. *ADH4* has been assumed to be crucial in vertebrates for retinoic acid synthesis via retinol oxidation (Duester 1996). Retinoic acid is thought to be involved in the anterior/posterior axis formation in both ascidian and amphioxus and, also, in vertebrates (reviewed by Shimeld 1996). But the fact that amphioxus *ADH3* is not a multifunctional *ADH*, correlated with the

absence of *ADH4* in lower chordates and with data on the microsomal retinol dehydrogenase of amphioxus (Dalfó et al. 2001), points to a candidate other than *ADH* for carrying out the retinoic acid synthesis in animal lines in general.

Finally, in spite of the biochemical and evolutionary constancy of *ADH3*, changes in the promoter elements are evident when the amphioxus and vertebrate *ADH3* forms are compared. Whereas a ubiquitous *Adh3* expression has been shown in mouse and human (Giri et al. 1989; Ang et al. 1996; Estonius et al. 1996), a specific expression, restricted to the midgut region, has been shown in amphioxus by *in situ* hybridization (Cañestro et al. 2000). Similarly, we have found, by *in situ* hybridization, an *Adh3* tissue specificity in *Drosophila* and ascidian (data not shown), much like that which has been demonstrated for other *ADH* classes in vertebrates.

Acknowledgments. We thank M. Riutort for helpful discussions on the phylogenetic analysis, W. Stone for helpful advice, M. Levine and A. Di Gregorio for the *C. intestinalis* genomic library, R. Di Lauro for the cDNA *C. intestinalis* library, and J. García-Fernández for the *B. floridae* genomic library. This work was supported by Ministerio de Ciencia y Tecnología (Spain) Grants BMC2000-0536 and UE98-0014, EU Contract BIO4 CT 97-2123, the Swedish MRC (Project 3532), CIRIT (Generalitat de Catalunya, 1997FI00007), and an FPI fellowship to C.C. We are indebted to the Serveis Científico-Tècnics (UB) for DNA sequence analysis.

References

- Abad L (2000) Utilización del gen octanol deshidrogenasa (*Odh*) como marcador molecular para el análisis de poblaciones de *D. subobscura*. Master thesis, Department of Genetics, University of Barcelona, Barcelona, Spain.
- Ang HL, Deltour L, Hayamizu TF, Zgombic-Knight M, Duester G (1996) Retinoic acid synthesis in mouse embryos during gastrulation and craniofacial development linked to class IV alcohol dehydrogenase gene expression. *J Biol Chem* 271:9526–9534
- Archibald JD (1996) Fossil evidence for a late cretaceous origin of “hoofed” mammals. *Science* 272:1150–1153
- Benton MJ (1997) Vertebrate paleontology. Chapman & Hall, New York, *Curr Opin Genet Dev* 8:694–700
- Boeddrich A, Burgdorf C, Hennig S, Panopoulou C, Steffens C, Borzym K, Lehrach H (1999) Sequence analysis of an amphioxus cosmid containing a gene homologous to members of the aldo-keto reductase gene superfamily. *Gene* 230:207–214
- Cañestro C, Hjelmqvist L, Albalat R, García-Fernández J, González-Duarte R, Jönvall H (2000) Amphioxus alcohol dehydrogenase is a class 3 form of single type and of structural conservation but with unique developmental expression. *Eur J Biochem* 267:6511–6518
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 116:156–159
- Church GM, Gilbert W (1984) Genomic sequencing. *Proc Natl Acad Sci USA* 81:1991–1995
- Dalfó D, Cañestro C, Albalat R, González-Duarte R (2001) Characterization of a microsomal retinol dehydrogenase gene from amphioxus: retinoid metabolism before vertebrates. *Chem-Biol Interact* 130:359–370
- Danielsson O, Jönvall H (1992) “Enzymogenesis”: Classical liver alcohol dehydrogenase origin from the glutathione-dependent form-

- aldehyde dehydrogenase line. *Proc Natl Acad Sci USA* 89:9247–9251
- Danielsson O, Eklund H, Jörnvall H (1992) The major piscine liver alcohol dehydrogenase has class-mixed properties in relation to mammalian alcohol dehydrogenases of classes I and III. *Biochemistry* 31:3751–3759
- Danielsson O, Atrian S, Luque T, Hjelmqvist L, González-Duarte R, Jörnvall H (1994a) Fundamental molecular differences between alcohol dehydrogenase classes. *Proc Natl Acad Sci USA* 91:4980–4984
- Danielsson O, Shafqat J, Estonius M, Jörnvall H (1994b) Alcohol dehydrogenases class III contrasted to class I. Characterization of the cyclostome enzyme, the existence of multiple forms as for the human enzyme, and distant cross-species hybridization. *Eur J Biochem* 225:1081–1088
- Dayhoff MO (1978) Survey of new data & computer methods of analysis. eds. Dayhoff MO vol 5, suppl 3, National Biochemical Research Foundation, Washington, DC
- Duester G (1996) Involvement of alcohol dehydrogenase, short-chain dehydrogenase/reductase, aldehyde dehydrogenase, and cytochrome P-450 in the control of retinoid signaling by activation of retinoic acid synthesis. *Biochemistry* 35:12221–12227
- Duester G, Farrés J, Felder VC, Höög JO, Parés X, Plapp B, Yin SJ, Jörnvall H (1999) Recommended nomenclature for the vertebrate alcohol dehydrogenase gene family. *Biochem Pharmacol* 58:389–395
- Estonius M, Svensson S, Höög JO (1996) Alcohol dehydrogenase in human tissues: Localisation of transcripts coding for five classes of the enzyme. *FEBS Lett* 397:338–342
- Fernández MR, Jörnvall H, Moreno A, Kaiser R, Parés X (1993) Cephalopod alcohol dehydrogenase: Purification and enzymatic characterization. *FEBS Lett* 328:235–238
- García-Fernández J, Holland PWH (1994) Archetypal organization of the amphioxus *Hox* gene cluster. *Nature* 370:563–566
- Giri PR, Krug JF, Kozak C, Moretti T, O'Brien SJ, Seanez HN, Goldman D (1989) Cloning and comparative mapping of a human class III (chi) alcohol dehydrogenase cDNA. *Biochem Biophys Res Commun* 164:453–460
- Holland PWH (1996) Molecular biology of lancelets: Insights into development and evolution. *Israel J Zool* 42:247–272
- Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci* 256:119–124
- 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
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *CABIOS* 8:275–282
- Jörnvall H, Höög J-O (1995) Nomenclature of alcohol dehydrogenases. *Alcohol Alcohol* 30:153–161
- Koivusalo M, Baumann M, Uotila L (1989) Evidence for the identity of glutathione-dependent formaldehyde dehydrogenase and class III alcohol dehydrogenase. *FEBS Lett* 257:105–109
- Kuma K, Miyata T (1994) Mammalian phylogeny inferred from multiple protein data. *Jpn J Genet* 69:555–566
- Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. *Nature* 392:917–920
- Kumar S, Tamura K, Nei M (1993) *MEGA: Molecular evolutionary genetics analysis*, v. 1.02. Pennsylvania State University, University Park
- Lee MSY (1999) Molecular clock calibrations and metazoan divergence dates. *J Mol Evol* 49:385–391
- Lundin LG (1993) Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* 16:1–19
- Lundin LG (1999) Gene duplication in early metazoan evolution. *Semin Cell Dev Biol* 10:523–530
- Luque T, Atrian S, Danielsson O, Jörnvall H, González-Duarte R (1994) Structure of the *Drosophila melanogaster* glutathione-dependent formaldehyde dehydrogenase/octanol dehydrogenase gene. *Eur J Biochem* 225:985–993
- Marzoli A, Renne PR, Piccirillo EM, Ernesto M, Bellieni G, De Min A (1999) Extensive 200-million-year-old continental flood basalts of the central Atlantic magmatic province. *Science* 284:616–618
- Naylor GJP, Brown WM (1997) Structural biology and phylogenetic estimation. *Nature* 388:527–528
- Nikoh N, Iwabe N, Kuma K, Ohno M, Sugiyama T, Watanabe Y, Yasui K, Shi-Cui Z, Hori K, Shimura Y, Miyata T (1997) An estimate of divergence time of parazoa and eumetazoa and that of cephalochordata and vertebrate by adolase and triose phosphate isomerase clocks. *J Mol Evol* 45:97–106
- Ohno S (1970) Evolution by gene duplication. Springer-Verlag, Heidelberg
- Persson B, Bergman T, Keung WM, Waldenström U, Holmquist B, Vallee BL, Jörnvall H (1993) Basic features of class-I alcohol dehydrogenase: variable and constant segments coordinated by inter-class and intra-class variability. *Eur J Biochem* 216:49–56
- Philippe H, Chenuil A, Adoutte A (1994) Can the cambrian explosion be inferred through molecular phylogeny? *Development* 120 (Suppl):15–25
- Poss SG, Boschung HT (1996) Lancelets (Cephalochordata: Branchiostomatidae): How many species are valid? *Israel J Zool* 42:13–66
- Roger AJ, Sandblom O, Doolittle WF, Philippe H (1999) An evaluation of elongation factor 1 α as a phylogenetic marker for eukaryotes. *Mol Biol Evol* 16:218–233
- Sharman AC, Holland PWH (1996) Conservation, duplication and divergence of developmental genes during chordate evolution. *Neth J Zool* 46:47–67
- Shimeld S (1996) Retinoic acid, Hox genes and the anterior-posterior axis in chordates. *BioEssays* 18:613–616
- Shimeld S (1999) Gene function, gene networks and the fate of duplicated genes. *Semin Cell Dev Biol* 10:549–553
- Shu D-G, Luo H-L, Morris SC, Hu X-L, Chen L, Han J, Zhu M, Li Y, Chen LZ (1999) Lower Cambrian vertebrates from south China. *Nature* 402:42–46
- Skrabanek L, Wolfe KH (1998) Eukaryote genome duplication—Where's the evidence? *Curr Opin Genet Dev* 8:694–700
- Strimmer K, von Haeseler A (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol Biol Evol* 13:964–969
- Strömberg P, Höög JO (2000) Human class V alcohol dehydrogenase (ADH5): A complex transcription unit generates C-terminal multiplicity. *Biochem Biophys Res Commun* 278:544–549
- Suga H, Hoshiyama D, Kuraku S, Katoh K, Kubokawa K, Miyata T (1999) Protein tyrosine kinase cDNAs from amphioxus, hagfish and lamprey: Isoform duplications around the divergence of cyclostomes and gnathostomes. *J Mol Evol* 49:601–608
- Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol* 12:823–833
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Wang DY, Kumar S, Hedges SB (1999) Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc R Soc Lond B Biol Sci* 266:163–171
- Wang Y, Gu X (2000) Evolutionary patterns of gene families generated in the early stage of vertebrates. *J Mol Evol* 51:88–96
- Zgombic-Knight M, Foglio MH, Duester G (1995) Genomic structure and expression of the ADH7 gene encoding human class IV alcohol dehydrogenase, the form most efficient for retinol metabolism in vitro. *J Biol Chem* 270:4305–4311
- Zuckerkandl E (1976) Evolutionary process and evolutionary noise at the molecular level. I. Functional density in proteins. *J Mol Evol* 7:167–183