

Merging protein, gene and genomic data: the evolution of the MDR-ADH family

R Gonzàlez-Duarte and R Albalat

Departament de Genètica, Universitat de Barcelona, Avda. Diagonal 645, Barcelona 08028, Spain

Multiple members of the MDR-ADH (MDR: Medium-chain dehydrogenases/reductases; ADH: alcohol dehydrogenase) family are found in vertebrates, although the enzymes that belong to this family have also been isolated from bacteria, yeast, plant and animal sources. Initial understanding of the physiological roles and evolution of the family relied on biochemical studies, protein alignments and protein structure comparisons. Subsequently, studies at the genetic level yielded new information: the expression pattern, exon–intron

distribution, *in silico*-derived protein sequences and murine knockout phenotypes. More recently, genomic and EST databases have revealed new family members and the chromosomal location and position in the cluster of both the first and new forms. The data now available provide a comprehensive scenario, from which a reliable picture of the evolutionary history of this family can be made.

Heredity (2005) **95**, 184–197. doi:10.1038/sj.hdy.6800723

Keywords: alcohol dehydrogenase; gene tandem duplications; gene and protein families; neofunctionalization; vertebrate ADH cluster

Introduction

Gene duplications have provided the genetic material for building complex organisms. Following such duplication, mutations cause gene copies to diverge. One copy might be silenced by deleterious mutations (pseudofunctionalization) or, alternatively, both copies may be preserved if substitutions result in novel capacities. Preservation can lead to one of three evolutionary fates: (i) both copies persist with perfect (or near perfect) sequence similarity if, simply, extra amounts of protein or RNA are required, (ii) each copy adopts some of the tasks of the ancestor (subfunctionalization) and (iii) one gene maintains the original function, while the other acquires a new role (neofunctionalization) (Force *et al*, 1999; Zhang, 2003). The recurrence of duplication and functional divergence has generated the extant gene families.

The medium-chain dehydrogenases/reductases (MDR) constitute a large superfamily of enzymes with almost 1000 members occurring in all types of organisms. There are at least 23 MDR in the human genome, 10 in *Drosophila melanogaster*, 13 in *Caenorhabditis elegans*, 38 in *Arabidopsis thaliana*, 15 in *Saccharomyces cerevisiae* and 17 in *Escherichia coli* (Jörnvall *et al*, 1999; Nordling *et al*, 2002). The MDR superfamily can be divided into at least eight families, of which the dimeric alcohol dehydrogenase (ADH) family has been most closely analysed to date.

MDR-ADH has been described in most major life forms, ranging from bacteria and archaea to plants, yeast and animals. Although multiplicity has been described

in many animal and plant species, here we focus on animal forms, as independent duplications gave rise to the family expansions in each kingdom. In vertebrates, eight distinct ADH classes, based upon sequence homology, catalytic features and gene expression pattern, have been defined (ADH1–8, class I–VII, according to Duester *et al* (1999) and the late class VIII following Peralba *et al* (1999)). However, a single class with only one representative encoding a glutathione-dependent formaldehyde dehydrogenase (FALDH) enzyme has been described in invertebrates (Luque *et al*, 1994). The invertebrate enzyme, being proorthologous to all vertebrate classes (Cañestro *et al*, 2002), should be named ADH1/2/3/4/5/6/7/8, but for simplicity and to reflect that the biochemical activity has been preserved in the vertebrate class III enzymes, it is referred to here as ADH3. Since ADH1 and ADH3 were first described, the number of ADH classes has grown continuously, as have the members of each class. The isolation of novel forms from animal tissues and the synthesis of recombinant enzymes have provided detailed knowledge at the biochemical level of the new forms (Jörnvall *et al*, 2000). Recently, genomic and EST databases have contributed greatly to the expansion of our understanding of this family. From all the data available, the MDR-ADH/*Adh* protein/gene system appears to be one of the most suitable models to illustrate the evolution of a large protein family, the fate of tandem duplications, the dynamics of intron gains and losses and the acquisition of new functions in vertebrates.

The ancestral MDR-ADH member: ADH3

Intron–exon structure

ADH illuminated the early debate on the evolutionary origin of introns at a time when very few correspon-

dences at the genetic and protein level were available. Protein–gene structure comparisons of the maize *Adh3* nucleotide-binding domain (Brändén *et al*, 1984) were believed to support the introns-early theory of eukaryotic genes (Gilbert, 1987), to which other *Adh* sequences and other genes encoding nucleotide-binding segments later contributed (Michelson *et al*, 1985; Stone *et al*, 1985; Duester *et al*, 1986; Quigley *et al*, 1988). However, as more examples became available and the exon theory of genes was assessed more fully, no correspondences between exons and units of protein structure were found for *Adh*, or for a collection of other genes (Stoltzfus *et al*, 1994; Cho and Doolittle, 1997; Dolferus *et al*, 1997).

Beyond the introns-early/introns-late debate, the variety of *Adh* structures found throughout the animal kingdom provides valuable clues to the evolution of the gene. Under the most parsimonious scenario, the *Adh* architecture of a variety of animal species (Figure 1) supports an ancestral metazoan *Adh3* with 10 introns of highly variable size (Cañestro *et al*, 2002). The urochordate *Ciona* shows all 10 positions fully conserved, which together with the nine introns shown by the echinoderm *Strongylocentrotus purpuratus* reinforces the view of a deuterostome ancestor with 10 introns. Of these, eight are fully conserved in all vertebrate species. Among protostomes, the nematode *Brugia malayi* with nine conserved introns (all except intron 6) and the arthropods (*D. simulans*, *D. subobscura* and *D. yakuba*) with intron 6 preserved, support the 10-intron organization and extend this gene structure to the ancestral metazoan. Lineage-specific gains could account for intron 5* in *C. elegans* (alternatively, it could be regarded as a 29-nt slippage of intron 6) and intron 8* in amphioxus. Intron slippage seems to have occurred in a few cases, intron

2 (12 nt) in *Drosophila* and *Ciona*, the latter with an additional 10-nt slippage in intron 4. Under the 10-intron hypothesis for the *Adh* ancestor, considerable intron losses have to be assumed in some protostome lineages (*Drosophila*, *Anopheles* and *Caenorhabditis*).

Evolution of coding regions

Adh3 encodes a dimeric, zinc-containing, NAD⁺-dependent enzyme (ADH3) with a subunit molecular mass of approximately 40 kDa. The number of amino-acid residues in the subunit ranges from 373 (ie human, the initiator Met is not considered; numbering for ADH3 and the other classes refers to the human forms) to 383 (ie *C. elegans*, the initiator Met is not considered), and the active form is structured in two domains: a coenzyme-binding domain (residues 177–322) at the dimer interface and a catalytic domain (residues 1–176 and 323–373) distal to the dimer interface (Yang *et al*, 1997). The active site is located in the cleft between the two domains.

Alignment of the 33 full-length proteins available (Supplementary information #1; the accession numbers for all sequences used in this study are shown in Table 1) supports the initial claim that no less than one-third of the positions were fully conserved (133 invariant residues out of 373) and that more than 45% exhibited only limited variability. The conservation pattern is uneven along the protein chain, variability clustering in two ADH3 segments, defined as V3a and V3b (around positions 240–270 and 330–350, respectively), which correspond to nonfunctional, superficial regions (Danielsen *et al*, 1994a; Cañestro *et al*, 2000). The evolutionary rate (v_{aa}) of the coding regions of ADH3 has been estimated to be 0.27×10^{-9} amino-acid substitutions per

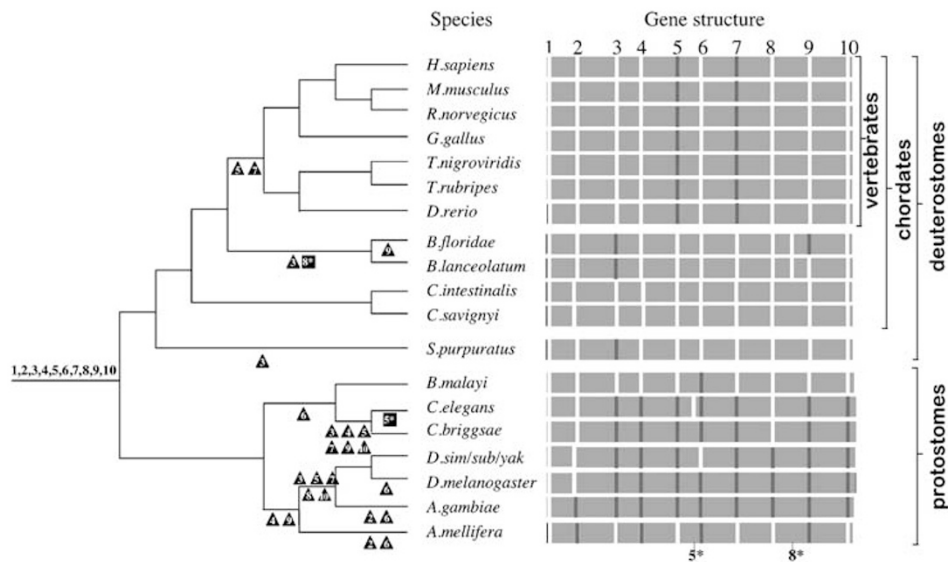


Figure 1 Exon–intron structure of animal *Adh3* genes. Schematic representation of *Adh3* exons (grey boxes) in 21 animal species. Each discontinuity in rectangles (white spaces) corresponds to an intron position. Introns are numbered (1–10) at the top of the scheme after an ancestral metazoan *Adh* with 10 introns is assumed under the most parsimonious view. The hypothetical localizations of the introns lost in the current gene structure are depicted with dark grey spaces. The position of lineage-specific introns 5* and 8* is indicated at the bottom of the scheme. Intron positions have been well preserved in protostomes and deuterostomes. Vertebrate *Adh3* genes show eight preserved positions, the urochordate *Ciona* shows full conservation of the predicted ancestral arrangement and the protostome *B. malayi* shows nine preserved positions. Lineage-specific losses (triangle) have been frequent in protostomes, and lineage-specific gains (box) would account for intron 5* in *C. elegans* and 8* in *B. floridae* and *B. lanceolatum*. The *D.sim/sub/yak* structure corresponds to the shared organization in *Drosophila simulans*, *D. subobscura* and *D. yakuba* species.

Table 1 Accession numbers of the ADH sequences used in the present study

Gene name	Species	Accession number
<i>Invertebrates</i>		
AgADH3	<i>Anopheles gambiae</i>	XM_314472, BX049473
AmADH3	<i>Apis mellifera</i>	XM_393266
BfADH3	<i>Branchiostoma floridae</i>	AF344171
BlADH3	<i>Branchiostoma lanceolatum</i>	AAF73255
BmADH3	<i>Brugia malayi</i>	Assembly 14992/CA:1384561 (TIGR database)
CbADH3	<i>Caenorhabditis briggsae</i>	CAE71050
CeADH3	<i>Caenorhabditis elegans</i>	Q17335
CiADH3	<i>Ciona intestinalis</i>	AAL72131
CsADH3	<i>Ciona savignyi</i>	AACT01001666
DmADH3	<i>Drosophila melanogaster</i>	AAA57187
OvADH3	<i>Octopus vulgaris</i>	P81431
ScADH3	<i>Schistosoma mansoni</i>	CD121084 CD077882 AI067204 CD080866 315491335 (Trace Archive database)
SmADH3	<i>Schmidtea mediterranea</i>	AY067494 and unpublished data
<i>Vertebrates</i>		
<i>Agnathan</i>		
MgADH3	<i>Myxine glutinosa</i>	P80360
<i>Fish</i>		
CcADH1A1	<i>C. carpio</i>	CF662249
CcADH1A2	<i>C. carpio</i>	CF662406
CcADH3H	<i>C. carpio</i>	CA964773 CA968782
DrADH1A1	<i>Danio rerio</i>	AAK97853
DrADH1A2	<i>D. rerio</i>	AY309075
DrADH1B1	<i>D. rerio</i>	NM_205642
DrADH1B2	<i>D. rerio</i>	NM_200455
DrADH3H	<i>D. rerio</i>	NP_571924
GcADH1A	<i>Gadus callarias</i>	P26325
GmADH3H	<i>G. morhua</i>	P81600
GmADH3L	<i>G. morhua</i>	P81601
GaADH3H	<i>Gasterosteus aculeatus</i>	CD503600 CD503599
IpADH1A	<i>Ictalurus punctatus</i>	BM438208
IpADH1B	<i>I. punctatus</i>	BM438601
OmADH1A	<i>Oncorhynchus mykiss</i>	CA350727 CA356649 CA357614 BX311144
OIADH1A	<i>Oryzias latipes</i>	AY682722
OIADH3H	<i>O. latipes</i>	AY512892
SaADH3H	<i>Sparus aurata</i>	JC4967
TrADH1A	<i>Takifugu rubripes</i>	CAAB01004230
TrADH1B	<i>T. rubripes</i>	CAAB01001615
TrADH3H	<i>T. rubripes</i>	CAAB01001615
TrADH3L	<i>T. rubripes</i>	CAAB01004230
TnADH3L	<i>Tetraodon nigroviridis</i>	96048271 95863608 100853635 97487377 97288213 99241634 95630447 (Trace Archive database)
<i>Amphibian</i>		
RpADH1	<i>Rana perezi</i>	P22797
RpADH8	<i>R. perezi</i>	O57380
XlADH1A1	<i>Xenopus laevis</i>	AAH61682
XlADH1A2	<i>X. laevis</i>	BC077635
XlADH1B1	<i>X. laevis</i>	BC074210
XlADH1B2	<i>X. laevis</i>	BC070669
XlADH1C1	<i>X. laevis</i>	BC070654
XlADH1C2	<i>X. laevis</i>	CF521684 CF547548
XlADH3	<i>X. laevis</i>	AY393843
XlADH4	<i>X. laevis</i>	AAH54260
XlADH8	<i>X. laevis</i>	BC041319
XtADH1A	<i>Xenopus tropicalis</i>	BC091025

Table 1 Continued

Gene name	Species	Accession number
XtADH1B	<i>X. tropicalis</i>	BC088070
XtADH1C	<i>X. tropicalis</i>	BC088050
XtADH2	<i>X. tropicalis</i>	BQ526692, Scaffold 785 (Joint Genome Institute)
XtADH3	<i>X. tropicalis</i>	CR762130
XtADH4	<i>X. tropicalis</i>	CR762130
XtADH7	<i>X. tropicalis</i>	BC088582
XtADH8A	<i>X. tropicalis</i>	Scaffold 785 (Joint Genome Institute)
XtADH8B	<i>X. tropicalis</i>	Scaffold 785 (Joint Genome Institute)
<i>Reptile</i>		
AmADH1	<i>Alligator mississippiensis</i>	P80222
NnADH1	<i>Naja naja</i>	P80512
UhADH1A	<i>Uromastyx hardwickii</i>	P25405
UhADH1B	<i>U. hardwickii</i>	P25406
UhADH3	<i>U. hardwickii</i>	P80467
<i>Avian</i>		
AaADH1	<i>Apteryx australis</i>	S78778
CjADH1	<i>Coturnix japonica</i>	P19631
GgADH1	<i>Gallus gallus</i>	P23991
GgADH3	<i>G. gallus</i>	XP_420657
GgADH7	<i>G. gallus</i>	NP_990423
GgADH7S1	<i>G. gallus</i>	XP_420658
GgADH7S2	<i>G. gallus</i>	XM_420659
ScADH1	<i>Struthio camelus</i>	P80338
ScADH2	<i>S. camelus</i>	P80468
<i>Mammal</i>		
BtADH3	<i>Bos taurus</i>	XM_581812
EcADH1E	<i>Equus caballus</i>	P00327
EcADH1S	<i>E. caballus</i>	P00328
EcADH3	<i>E. caballus</i>	P19854
GkADH1	<i>Geomys knoxjonesi</i>	Q64415
HsADH1A	<i>Homo sapiens</i>	P07327
HsADH1B	<i>H. sapiens</i>	P00325
HsADH1C	<i>H. sapiens</i>	P00326
HsADH2	<i>H. sapiens</i>	BC022319
HsADH3	<i>H. sapiens</i>	BC014665
HsADH4	<i>H. sapiens</i>	P40394
HsADH5	<i>H. sapiens</i>	NP_000663
MamADH1	<i>Macaca mulatta</i>	P28469
MmADH1	<i>Mus musculus</i>	P00329
MmADH2	<i>M. musculus</i>	CAB57455
MmADH3	<i>M. musculus</i>	NP_031436
MmADH4	<i>M. musculus</i>	Q64437
MmADH5ps	<i>M. musculus</i>	AC079832
MmADH6A	<i>M. musculus</i>	BAB25012
MmADH6B	<i>M. musculus</i>	XM_110439 AC079845
OcADH1	<i>Oryctolagus cuniculus</i>	S29343
OcADH2A	<i>O. cuniculus</i>	O46649
OcADH2B	<i>O. cuniculus</i>	O46650
OcADH3	<i>O. cuniculus</i>	O19053
PhADH1	<i>Papio hamadryas</i>	P14139
PmADH1	<i>Peromyscus maniculatus</i>	P41680
PmADH5	<i>P. maniculatus</i>	P41681
RnADH1	<i>Rattus norvegicus</i>	P06757
RnADH2	<i>R. norvegicus</i>	NP_058966
RnADH3	<i>R. norvegicus</i>	BC083724
RnADH4	<i>R. norvegicus</i>	NP_599156
RnADH5	<i>R. norvegicus</i>	XP_227746
RnADH6A	<i>R. norvegicus</i>	XM_215715
RnADH6B	<i>R. norvegicus</i>	XM_227745
SsADH3	<i>Sus scrofa</i>	AY609557

When the ADH structure has been derived from several sequences, all accession numbers are provided. Specific databases/archives used are also indicated.

site per year (Cañestro *et al*, 2002). This value is significantly lower than that of other ADH classes, in accordance with the classical view of a ‘constant’ enzyme in function, enzymological features and overall structure (Danielsson *et al*, 1994a). The evolutionary rate was estimated with the γ -corrected distribution, which takes into account the rate variation among sites. Under this model, the among-site rate variation depends on the α parameter, which is an index of the degree of variation: the lower the α value, the larger the rate variation. In our case, the low α value for ADH3 ($\alpha=0.5$) is consistent with the uneven distribution of amino-acid substitutions.

Functionally important residues have been identified from the crystallographic structure of the human ADH3 enzyme. A total of 22 residues have been associated with the substrate-binding site, catalytic domain, coenzyme-binding domain and active zinc coordination residues (Yang *et al*, 1997; Sanghani *et al*, 2002). In general, these positions have been strictly conserved throughout animal evolution (Table 2). Even the most variable, Ile93, shows conservative substitutions. Two animals, *B. malayi* (Nematoda) and *S. mediterranea* (Platyhelminthes) show an Ile93Thr change, previously reported in yeast, plant and bacterial enzymes. Four conserved Cys – positions 96, 99, 102 and 110 – bind the structural zinc in a superficial loop without secondary structure. The four residues directly involved in the active zinc coordination, Cys44, His66, Glu67 and Cys173, and Arg368, which

interacts with Glu67 and the cofactor, are strictly preserved (Sanghani *et al*, 2002). Two other residues interacting with the coenzyme, His45 and Thr46, are also fully conserved. The conformation of the catalytic domain appears to be stabilized by the interactions of Glu57 with Arg114 and Tyr49 with Ala294, which are preserved in all the sequences analysed. The structure of the catalytic domain in the apo-form is intermediate (‘semiopen’) with respect to the two alternative domain structures (‘open’ and ‘closed’) of ADH1. This semiopen conformation agrees with the poor ADH3 activity in the presence of small alcohols (Sanghani *et al*, 2003). A total of 18 positions have been related with substrate binding: Thr46, Tyr49, Asp55, Glu57, His66, Glu67, Tyr92, Ile93, Leu109, Gln111, Arg114, Met140, Lys283, Val293, Ala294, Val308, Thr309 and Ala317 (Yang *et al*, 1997; Sanghani *et al*, 2002). Of these, the eight strictly conserved in ADH3 – Thr46, His66, Glu67, Arg114, Val293, Ala294, Val308 and Ala317 – have been considered a signature for direct assignment of any novel sequence (Norin *et al*, 2004). Finally, a large hydrophobic segment around positions 270–320 constitutes the main subunit/subunit interaction domain of the dimer (Eklund *et al*, 1990). In this segment, more than 50% of the residues are strictly conserved and no less than two-thirds exhibit only limited variability in animal sequences. Evidence for little divergence in the segments of subunit interactions is the fact that human ADH3 crosshybridizes *in vitro* with ADH3 of other

Table 2 ADH3 alignment at relevant positions for function in animal sequences

	44	<u>45</u>	<u>46</u>	<u>49</u>	55	<u>57</u>	66	67	92	93	109	111	<u>114</u>	140	173	283	293	<u>294</u>	308	309	317	<u>368</u>
Human	C	H	T	Y	D	E	H	E	Y	I	L	Q	R	M	C	K	V	A	V	T	A	R
Pig																						
Horse																						
Cow																						
Rabbit																						
Mouse																						
Rat																						
Chicken																						
Lizard																						
<i>X. laevis</i>																						
<i>X. tropicalis</i>																						
Medaka													V									
Sea bream													V									
Zebrafish													V									
Cod (H)													V									
Cod (L)													V									
Fugu (H)													V									
Fugu (L)													V									
Pufferfish													V									
Hagfish													V									
<i>B. lanceolatum</i>													V									
<i>B. floridae</i>																						
<i>C. intestinalis</i>																						F
<i>C. savignyi</i>																						F
<i>A. gambiae</i>																						P
<i>D. melanogaster</i>																						R
<i>A. mellifera</i>																						V
<i>B. malayi</i>																						S
<i>C. briggsae</i>																						T
<i>C. elegans</i>																						V
<i>O. vulgaris</i>																						V
<i>S. mansoni</i>																						S
<i>S. mediterranea</i>																						T

Numbering is from the human enzyme. Empty spaces indicate identity to human ADH3. The substrate-binding positions are shown in bold. Italics denotes amino acids directly involved in the active zinc coordination. Single underlines depict residues interacting with the cofactor and double underlines those involved in the catalytic domain.

species (Danielsson *et al*, 1994b). Remarkably, this domain is not located in an overall conserved region, but surrounded by the two highly variable segments of the enzyme, V3a and V3b (Danielsson *et al*, 1994b).

Expression pattern and activity

ADH3 has been described as a ubiquitous enzyme in vertebrates (Ang *et al*, 1996; Funkenstein and Jakowlew, 1996), although differences in activity in mammalian tissues (up to 30-fold) and, at the transcriptional level, during fish development (up to five-fold) have been reported (Funkenstein and Jakowlew, 1996; Uotila and Koivusalo, 1997). Indeed, from our data, the widespread expression during zebrafish development was not ubiquitous, but consistent with spatio-temporal regulation (Dasmahapatra *et al*, 2001; Cañestro *et al*, 2003). At the subcellular level, in contrast to other ADHs of strictly cytosolic localization, ADH3 has also been reported in the cell nucleus (Iborra *et al*, 1992; Fernández *et al*, 2003), presumably related to DNA protection.

The ADH3 expression pattern in invertebrates, far from being widespread, has mainly been found in digestive tissues: in the deuterostomes amphioxus and ascidian, restricted to the posterior portion of the developing gut in the former, and to the anterior endoderm in the latter, which forms the gastric cavity after metamorphosis (Cañestro *et al*, 2000, 2003). In the protostome *Drosophila*, expression was first uniformly distributed in the fertilized egg but later restricted to the fat body (Cañestro *et al*, 2003). The comparable restricted pattern in protostomes and deuterostomes suggests ancestral tissue-specific expression for ADH3 that later expanded in the vertebrate lineage.

ADH3 has classically been considered a glutathione-dependent formaldehyde dehydrogenase that catalyses the NAD^+ -dependent oxidation of *S*-hydroxymethylglutathione (HMGS) to *S*-formylglutathione (Koivusalo *et al*, 1989) (Figure 2). The enzyme can also promote the reversible reaction, the reduction of *S*-formylglutathione to formaldehyde, using the same cofactor with even higher velocity at pH below 8 (Uotila and Koivusalo, 1974). The early origin of this activity and its preservation in the major life forms have suggested a crucial physiological contribution of ADH3 in formaldehyde

metabolism, significantly in detoxification (Uotila and Koivusalo, 1989), among other roles, not yet fully explored, in ω -hydroxy fatty acid (Wagner *et al*, 1984) and leukotriene metabolism (Gotoh *et al*, 1990). However, the specific regulation of ADH3 in invertebrates challenges its presumed 'housekeeping' detoxification role. Although far from clear, but compatible with its primary site of activity, ADH3 could regulate the formaldehyde levels for the serine- and folate-dependent one-carbon metabolism, mainly operating in the liver and required for fetal and postnatal growth and development (Thompson *et al*, 2001).

Data on the biochemical capabilities of ADH3 have related this enzyme with nitric oxide (NO) homeostasis and retinoic acid (RA) metabolism (Figure 2). In the case of the former, the fact that ADH3 can effectively reduce *S*-nitrosoglutathione (GSNO) has supported its contribution to the control of the intracellular levels of GSNO and *S*-nitrosylated proteins (Liu *et al*, 2001, 2004), crucial constituents in signal transduction, host defence and nitrosative stress pathways in vertebrates. This activity, initially characterized in mammals (Jensen *et al*, 1998), is conserved in cephalochordates (unpublished), plants (Sakamoto *et al*, 2002) and yeast (Fernández *et al*, 2003). The link between ADH3 and NO/GSNO homeostasis could be supported by the enzyme tissue-specific expression and NO accumulation *in vivo* in digestive tissues. However, the diffusible nature of NO makes it difficult to demonstrate this relationship.

Finally, the hypothesis that ADH3 contributes to RA metabolism was based on its generalized expression pattern in vertebrates and the finding that mouse ADH3 oxidizes all-*trans*-retinol *in vitro* at a rate comparable to that of 'conventional' retinol dehydrogenases (Molotkov *et al*, 2002b). Although challenging, this hypothesis would be difficult to reconcile with recent findings that zebrafish and ascidian embryos use retinal as the main RA precursor during development (Costaridis *et al*, 1996; Irie *et al*, 2003; Lampert *et al*, 2003), and with the developmental strategies of *Drosophila*. Moreover, the viability of *Adh3*-knockout mice, the difficulties in dissecting ADH3 specificity from the overlapping activities of other ADH classes – ADH1 and ADH4, both of which oxidize retinal – and the difficulty of distinguishing the phenotypic effects of RA deficiency from

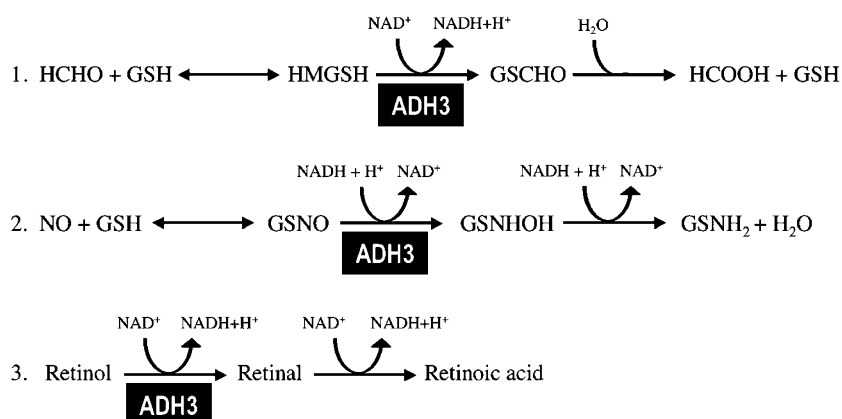


Figure 2 Main metabolic reactions catalysed by ADH3: (1) The classical role as glutathione-dependent formaldehyde dehydrogenase, (2) as a GSNO reductase and (3) the retinol oxidation activity.

nitrosative-homeostasis alterations, all question a significant contribution of ADH3 to RA metabolism.

The generation of new ADH

ADH1

The first expansion of the ADH family involved a class III tandem duplication around 500 million years ago (MYA), near the agnathan/gnathostome divergence (555 MYA), but before the distinct gnathostome lineages were established (400 MYA) (Cederlund *et al.*, 1991; Cañestro *et al.*, 2002). One of the copies acquired novel metabolic capacities defining a novel class, the ethanol-active class I, and evolved at a rate 3.6-fold higher than the other, which retained the ancestral ADH3 activity. Differences in evolutionary rates often mislead phylogenetic relationships and highlight the need to integrate structural and molecular data to avoid such biases. The main argument in favour of the expansion of the ADH family restricted to vertebrates from an ancestral ADH3 form is that it is the only representative in all invertebrates examined to date at the biochemical, genetic and genomic level (Table 1; Kaiser *et al.*, 1993; Danielsson *et al.*, 1994b; Luque *et al.*, 1994). Moreover, the identical intron organization of a collection of vertebrate *Adhs* further reinforces this assumption (Cañestro *et al.*, 2002). A similar intron distribution analysis led to the view that plant ethanol-active *Adh* genes (class P) also arose by a duplication of an *Adh3* ancestor but following an independent pathway (Dolferus *et al.*, 1997). Because genes encoding ADH are not spread over different chromosomes but cluster in the human genome at 4q21–25 and in the syntenic regions of mouse (at 3G3) and rat (at 2q44) (Figure 3), it was proposed that the family expansion proceeded by tandem duplications during vertebrate evolution (Cañestro *et al.*, 2002).

Structural differences between ADH3 and ADH1 enzymes account for the functional differences. In ADH class I, the classical liver ethanol dehydrogenase, the binding of the coenzyme induces the catalytic domain to approach the coenzyme-binding domain and to narrow the active site cleft. The two domain conformations are thus described as ‘open’ in the apoenzyme, and ‘closed’

in the binary and ternary complexes. This domain closure promotes the binding of small alcohols since it effectively decreases the alcohol-binding site. These conformations could account for the different substrate specificity and kinetic mechanisms of ADH1 and ADH3. The coenzyme-induced conformational change is consistent with the ordered kinetics of ADH1, while the random mechanism of ADH3 is coherent with its ‘semiopen’ domain conformation (Sanghani *et al.*, 2000). The proton relay pathway is also significantly different in the two classes. In class I, the components are the Thr/Ser48 that hydrogen binds with the alcohol hydroxyl group, the hydroxyl groups of nicotinamide ribose, and His51, a general base in contact with the solvent (Ramaswamy *et al.*, 1994). However, in the ADH3 enzyme, His51 is not found, which suggests that proton transfer proceeds directly to the solvent (Sanghani *et al.*, 2002). Besides the Thr/Ser48, class I enzymes share three positions, His67, Glu68 and Phe140, thus far strictly conserved. This triad has been proposed as a signature for class assignment in further gene analyses (Norin *et al.*, 2004), although preservation of these positions does not necessarily imply ethanol oxidizing activity (Reimers *et al.*, 2004), as discussed below. When the distribution of the constant and variable segments of class I and III was compared, further differences were found (Danielsson *et al.*, 1994a). In class I, three segments stand out as variable. They lie near the substrate-binding pocket and participate in the subunit interactions. In contrast, these regions are among the most conserved segments of ADH3 (Cañestro *et al.*, 2000).

Concerning the biological function, in addition to the oxidation of ethanol, ADH1 has been implicated in other physiological pathways, for example, norepinephrine, dopamine, serotonin and bile acid metabolism (Höög *et al.*, 2001b and references therein). Furthermore, it can catalyse the oxidation of retinol *in vitro* (Boleda *et al.*, 1993) and *in vivo* (Deltour *et al.*, 1999b). However, analysis of *Adh1*-null mutant mice challenged a major role in retinol metabolism and, instead, suggested a protective function against vitamin A toxicity (Molotkov *et al.*, 2002a, b).

The *Adh1* gene is expressed at a very high level in the liver and also at a significant degree in the small and

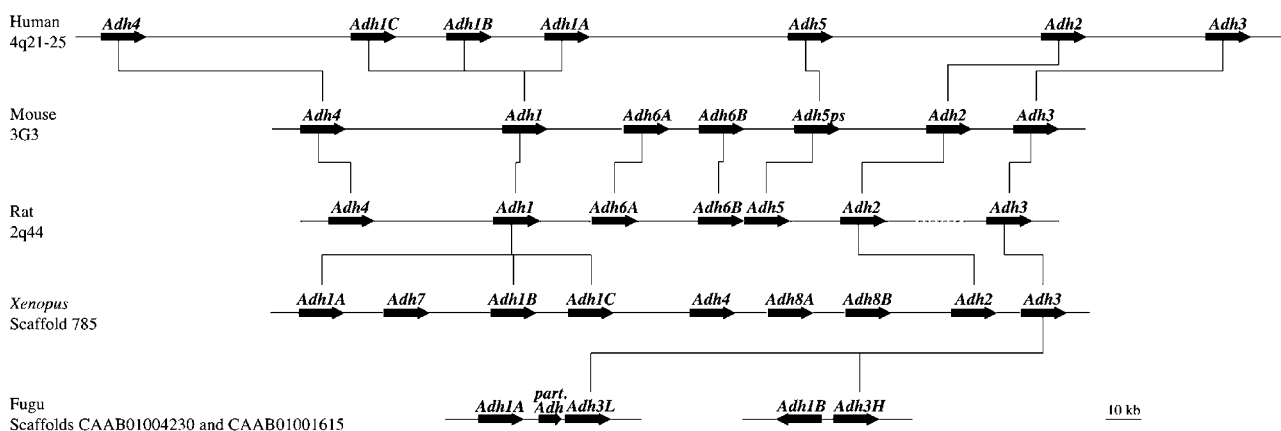


Figure 3 Structural organization of the *Adh* cluster in vertebrates. Human and rat schemes have been drawn after the Map Viewer website from NCBI. The rat *Adh3* (dotted line), not yet located, has been assumed at the 3' end of the cluster. The mouse complex is as in Szalai *et al.* (2002). The *X. tropicalis* cluster has been based on the JGI Assembly v3.0 of the genome database. The *Adh* sequences from fugu have been derived from scaffolds CAAB01001615 and CAAB01004230. Arrows indicate transcriptional orientation.

large intestine, kidney, adrenal, testis, epididymis, uterus and ovary of adult mice (reviewed in Duester, 2000). *Adh1* tissue-specific expression, in contrast to the widespread distribution of *Adh3* in vertebrates, resembles the specific pattern of invertebrate *Adh3* (Cañestro *et al*, 2000, 2003) and would be in agreement with the preservation of some regulatory sequences of the ancestor in the copy that acquired the new specificities.

The ADH1/ADH3 tandem arrangement appears widespread in gnathostomes but its evolutionary fate has been diverse in the different lineages, with independent duplication events. Thus, in what follows, we will analyse the status of the family in the distinct vertebrate lines.

ADH1 and ADH3 in fish

Several lines of evidence point to additional duplications in fish leading to a multiplicity of ADH1 and ADH3 forms in several lineages. Two ADH3 isoforms were first described in *Gadus morhua* (Atlantic cod) and were named after their high and low activity, ADH3H and ADH3L, respectively (Danielsson *et al*, 1996). Later, in *Takifugu rubripes* (fugu), two forms were reported (Hjelmqvist *et al*, 2003), and as each clusters with one cod counterpart (Figure 4 and Supplementary information #2), we have named them H and L. Moreover, searches in EST and genomic databases have rendered ADH3L sequences in *Tetraodon nigroviridis* (pufferfish), and ADH3H in *Cyprinus carpio* (carp), *Danio rerio* (zebrafish), *Gasterosteus aculeatus* (three-spined stickleback), *Oryzias latipes* (Japanese medaka) and *Sparus aurata* (sea bream) (Table 1). Concerning ADH1 multiplicity, to the A forms initially reported in *Gadus callarias* (Baltic cod) (Danielsson *et al*, 1996), *T. rubripes* (Hjelmqvist *et al*, 2003) and *D. rerio* (ADH8) (Reimers *et al*, 2004), new sequences have now been added from genome searches: *C. carpio*, *I. punctatus*, *O. mykiss*, *O. latipes*. Also, after database searches, a new ADH1B subclass can be described in *D. rerio*, *I. punctatus* and *T. rubripes* (Table 1 and Figure 4). The constancy of the H/L and A/B forms in several distant teleost species supports an early duplication event. Indeed, the two-tandem arrangement of *T. rubripes* ADH1B–ADH3H and ADH1A–ADH3L (Figure 3) seems to be indicative of an early cluster that probably expanded with the genomic duplication assumed before the radiation of teleosts (>235 MYA) (Amores *et al*, 1998) but after the two major radiations of jawed vertebrates, the ray-finned fish (Actinopterygia) and the lobe-finned fish/tetrapod (Sarcopterygia), 400 MYA.

In *D. rerio*, two ADH1A (ADH1A1 or ADH8a, and ADH1A2 or ADH8b) and two ADH1B (ADH1B1 and ADH1B2) were identified. Moreover, the close structural relationship of *D. rerio* ADH1A1 and ADH1A2 with *C. carpio* ADH1A1 and ADH1A2, respectively, supports an A1/A2 duplication before the zebrafish–carp divergence. Finally, in fugu, a partial ADH1A sequence was identified within the ADH1A–ADH3L cluster.

Two evolutionary pathways could explain the pattern observed for the ADH family in the piscine species (Figure 5a and b), starting with either one or two tandem duplications before the teleost radiation. In the first case, the duplication (as a consequence of the genomic duplication in teleosts) of an early two-gene cluster,

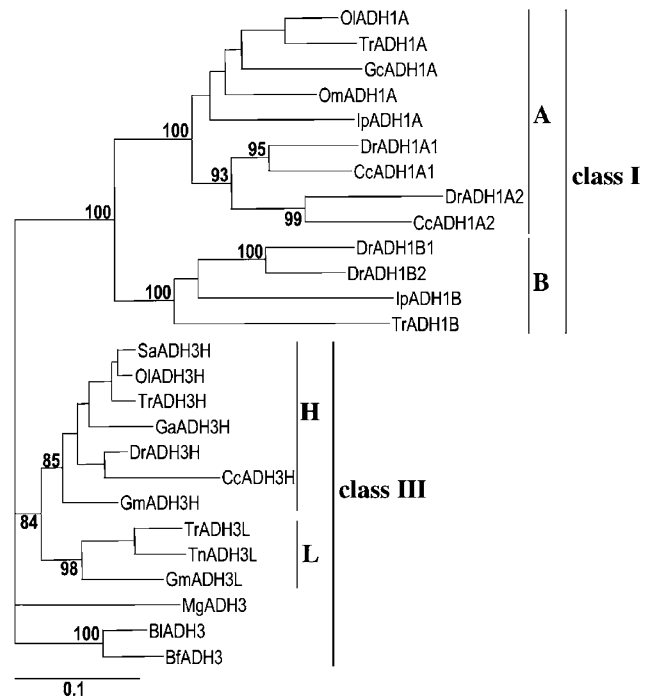


Figure 4 Neighbour-joining (NJ) phylogenetic tree of piscine ADH after protein alignment using ClustalX and adjusted by eye. Figures at nodes are the scores from 1000 bootstrap resampling of the data. The tree was rooted using amphioxus ADH3 as outgroup. The alignment was performed with 194 amino acids of the N-terminal segment, as full-length sequences were unavailable for some species. Two independent clades of class III, H and L (after the high- and low-activity forms reported in cod, respectively) seem to be of common occurrence in gnathostomes. Class I shows variants of A and B subclasses. ADH1A or 1B or both have been found in *O. latipes*, *T. rubripes*, *G. callarias*, *O. mykiss* and *I. punctatus*. Additional duplications generated ADH1A1 and ADH1A2 forms in *D. rerio* and *C. carpio*, and ADH1B1 and ADH1B2 in *D. rerio*. Nomenclature is as in Table 1.

followed by at least three lineage-specific tandem duplications, could account for the present structures in zebrafish and fugu (Figure 5a). Alternatively, the same genomic duplication involving, in this case, an initial segment with three gene copies, followed by gene conversion and the deletion of one copy in fugu, could have led to the same arrangements (Figure 5b). Gene conversion has been reported in every species and at every locus that has been examined in detail (Graur and Li, 1999), and mainly involves genes that are close to each other. In evolutionary studies, gene conversion erodes the record of molecular divergence and misleads phylogenetic reconstructions (Gogarten and Olendzenski, 1999; White and Crother, 2000). Although present data do not allow clear discrimination between the two evolutionary pathways, the biochemical data lend more support to the three-copy/gene-conversion scenario. While cod ADH1A and zebrafish ADH1A1 are class I enzymes, zebrafish ADH1A2 is phylogenetically class I but functionally class III (Reimers *et al*, 2004), compatible with an ADH3 copy that incorporated ADH class I segments by conversion.

ADH2 and ADH7, two new ADH classes in tetrapods

Two additional classes, II and VII, seem to have emerged before the amphibian–amniota split, 350 MYA, during

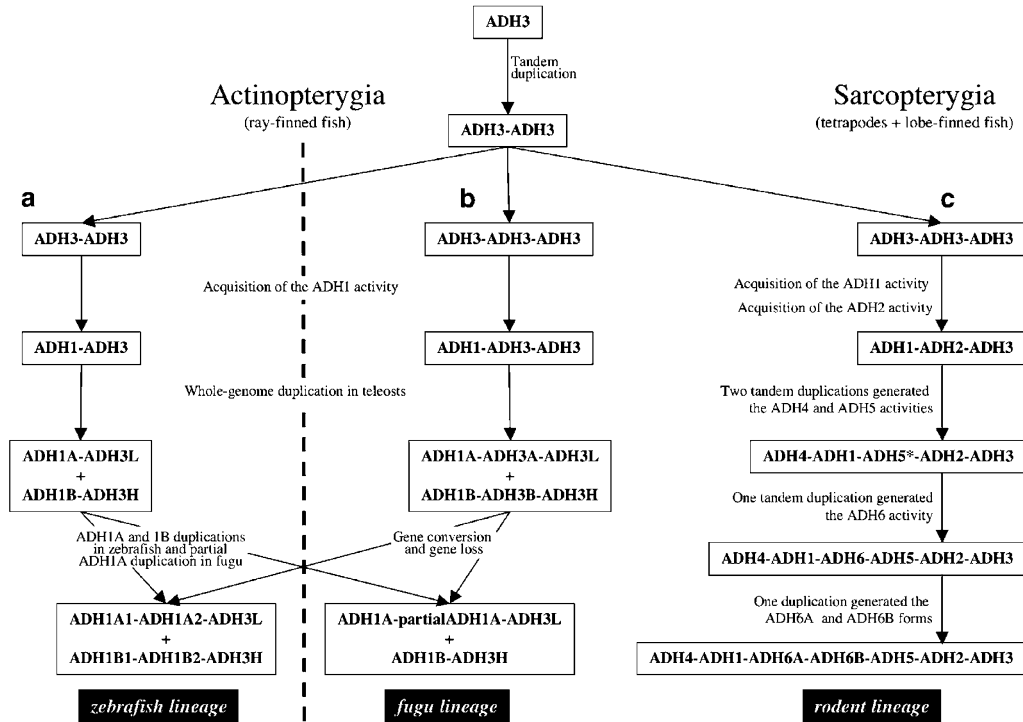


Figure 5 Hypothetical evolutionary pathways leading to vertebrate ADH multiplicity. At the base of vertebrate radiation, an initial tandem duplication of the ancestral ADH3 led to a two-gene cluster. In actinopterygia, two alternative pathways (a and b) may have occurred. Probably after the acquisition of ADH1 activity by the most 5' member of the cluster, whole-genome duplication in teleosts comprising a two- or three-ADH cluster generated the four or six ADH copies depicted in an intermediate step. In the last step, either (a) independent duplications or (b) gene conversion to ADH1 forms combined with one partial and one whole-gene deletion in fugu could account for the present arrangement found in the zebrafish and fugu lineages. (c) In sarcopterygia, subsequent tandem duplications and nucleotide substitutions led to class multiplicity. ADH1 and ADH2 found in all tetrapods were the first new forms, followed by ADH4 and ADH5 (mammals) and, subsequently, by ADH6A and ADH6B (rodents). If mammalian ADH5 was orthologous to the amphibian and avian class VII, class V/VII would have arisen early in tetrapods. The cluster ADH4-ADH1-ADH6A-ADH6B-ADH5-ADH2-ADH3 maps at rat 2q44 and mouse 3G3, the latter showing an ADH5 pseudogene.

tetrapod evolution, extending the metabolic capabilities derived from class III and I.

ADH2: Class II forms, previously only identified in mammalian and avian/reptilian lineages, have now been retrieved from the *Xenopus tropicalis* genome database (Table 1). Although not formally settled, class II enzymes were assumed to derive from class I, to which they are usually compared, since they are mainly found in the liver and contribute to ethanol metabolism (Ditlow *et al*, 1984; Estonius *et al*, 1996; Svensson *et al*, 1999). Moreover, phylogenetic analyses suggested that ADH2 is the sister group of the tetrapod nonclass III enzymes (Hjelmqvist *et al*, 1995; Hoffmann *et al*, 1998; Philippe and Lopez, 2001). However, the faster evolutionary rate of class II enzymes – two- to six-fold higher than class I and III, respectively – could have blurred the branching among these classes.

At the amino-acid level, ADH2 enzymes are more similar to class III than they are to class I forms (Hjelmqvist *et al*, 1995). Moreover, the semiopen conformation of the catalytic domain and the mechanism of catalysis that does not require full domain closure resemble ADH3 (Svensson *et al*, 2000). Besides, ADH2 lacks His51, the assumed catalytic base for proton relay in ADH1, which suggests that, as in ADH3, proton transfer proceeds directly to the solvent (Davis *et al*, 1994;

Svensson *et al*, 1999). Moreover, the ADH2 substrate-binding pocket contains several insertions and deletions. Among them, the four-residue insertion around position 119, preserved in all class II ADHs (Svensson *et al*, 1999), adds an appendix that makes the substrate-binding pocket larger than that of ADH1, although more closed than that of ADH3 (Svensson *et al*, 2000). Although it was initially assumed that class II enzymes contribute to ethanol metabolism, rodent ADH2 recognizes ethanol only poorly. Also, the human counterpart is only active, albeit at much lower efficiencies than human ADH4 and ADH1C, at high ethanol concentrations (Höög *et al*, 2001b). Therefore, based on the biochemical, structural and catalytic features, we hypothesize that a tetrapod ADH class III was the ancestor of class II (Figure 5c).

ADH7: Class VII forms were initially described in birds (Kedishvili *et al*, 1997). In chicken, ADH7 (or ADH-F) is flanked by two closely related sequences, annotated as 'similar to ADH-F'. Although the latter could derive from recent tandem duplication events (Figure 6), the strikingly high sequence similarities with ADH7, not only at the coding regions, but also at the intronic segments, challenge their true gene identity and raise the possibility of assembling artefacts. The catalytic properties of the chick ADH7 showed that it may act as a steroid/retinoid dehydrogenase (Kedishvili *et al*,

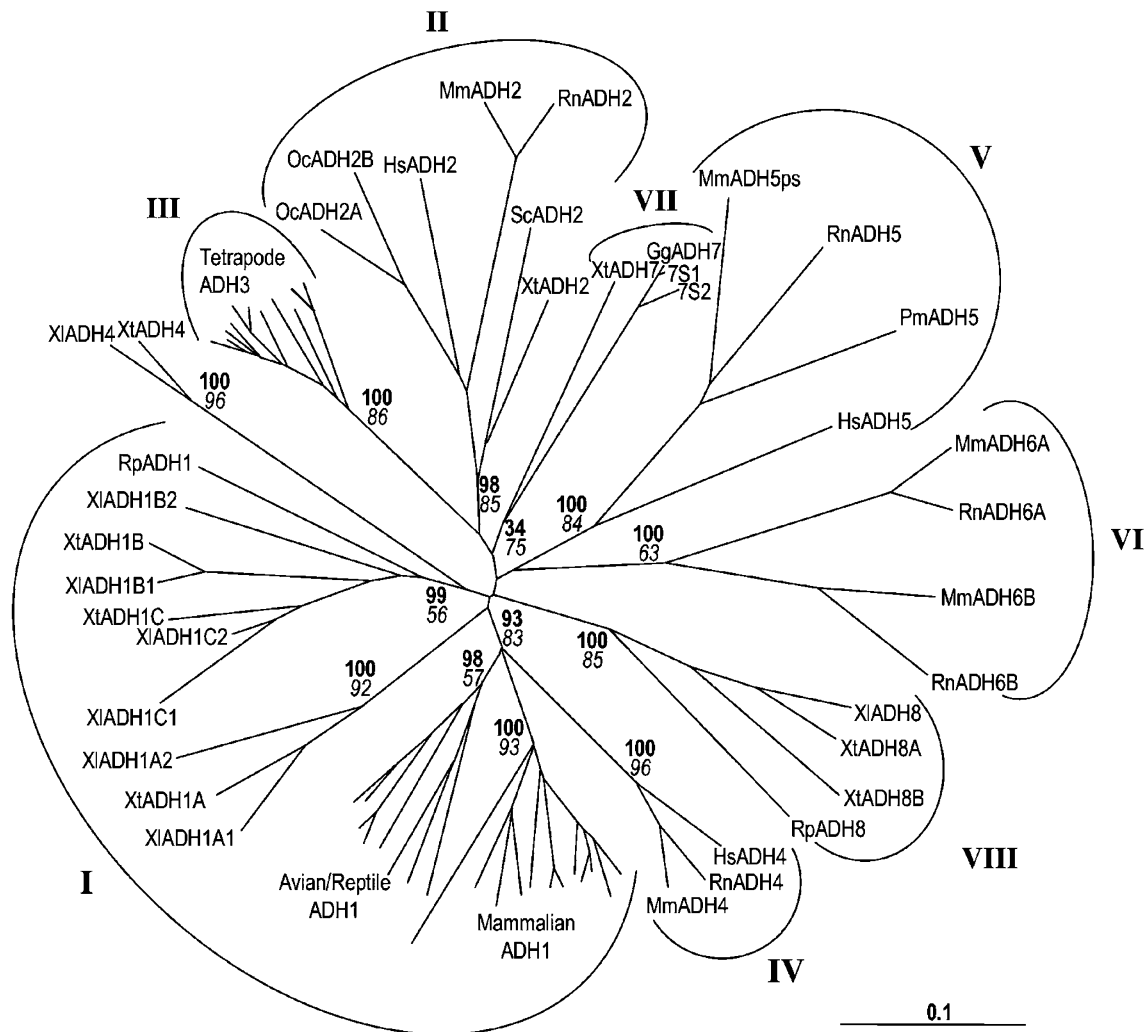


Figure 6 Phylogeny of ADH in tetrapods. The eight biochemical classes of ADH (I–VIII) can be differentiated phylogenetically, although their relationships are obscured by significant differences in evolutionary rates and lineage-specific duplications. The tree illustrates that each class is evolving at a different rate, the ancestral ADH3 being the most conserved class. *XIADH4* and *XtADH4*, the only enzymes not ascribed to any class, appear to be closer to amphibian class I than to mammalian class IV (see text for details). For NJ tree construction, amino-acid sequences extracted from databases were aligned using ClustalX and adjusted by eye. The reliability of the inferred phylogeny has been assessed by 1000 bootstrap repetitions. A second tree constructed following the maximum-likelihood (ML) method produced a similar topology. In this case, the confidence of each internal branch has been estimated by the quartet-puzzling method, a fast tree search algorithm implemented in PUZZLE 4.0.1. Figures at nodes are the scores from bootstrap resampling of the data (NJ), in bold, and quartet puzzling support values (ML), in italics. Tetrapod ADH3 includes *EcADH3*, *SsADH3*, *HsADH3*, *BtADH3*, *MmADH3*, *RnADH3*, *OcADH3*, *UhADH3*, *GgADH3*, *XIADH3* and *XtADH3* (clockwise). Avian/Reptile ADH1 includes *UnADH1B*, *UnADH1A*, *NnADH1*, *AmADH1*, *GgADH1*, *CjADH1*, *ScADH1* and *AaADH1* (clockwise). Mammalian ADH1 includes *HsADH1A*, *MamADH1*, *HsADH1B*, *PhADH1*, *HsADH1C*, *EcADH1E*, *EcADH1S*, *OcADH1*, *MmADH1*, *RnADH1*, *PmADH1*, *GkADH1* (clockwise). Nomenclature is as in Table 1.

1997), while no biochemical information is available for the other two similar forms. A sequence with moderate resemblance to the avian ADH7 has been retrieved from *X. tropicalis* database (Table 1, Figure 6 and Supplementary information #3). If class VII was present in amphibians, an old occurrence in tetrapods should be assumed, although no obvious representatives have been identified in the human, mouse or rat genomes. Either this class was lost or, alternatively, it could correspond to one of the mammalian-specific classes, whose relationship has been masked by high evolutionary rates. Unfortunately, the information gathered to date does not suggest an unequivocal answer.

Amphibian ADH

Amphibians exhibit the greatest ADH complexity reported to date in the animal kingdom (Table 1 and Figure 6). Searches on EST and genome databases have rendered at least nine *Adh* sequences in each *X. laevis* and in *X. tropicalis*. Moreover, two ADH enzymes have been reported in *Rana perezyi*. Two *Xenopus* sequences, one from each species, shared over 85% similarity with anniota ADH3 enzymes and showed preservation of the 22 conserved residues of class III (Table 2), identifying ADH3 members with reasonable confidence. Moreover, in *X. tropicalis*, members of the ADH2 and ADH7 classes have also been identified.

Concerning class I, an ADH enzyme with functional homologies to this class was reported in *R. perezy* (Cederlund *et al*, 1991). In *X. tropicalis*, we retrieved three ADH sequences from database searches, which we named ADH1A, ADH1B and ADH1C. All showed highest similarity with class I enzymes, in particular with *R. perezy* ADH1 (within the range of 67–77%). Each *X. tropicalis* ADH1 form appeared to be related to two *X. laevis* sequences, which were denominated following the homologies: ADH1A1 and ADH1A2 after *X. tropicalis* ADH1A (89 and 81% similarity, respectively), ADH1B1 and ADH1B2 after *X. tropicalis* ADH1B (94 and 74% similarity, respectively) and ADH1C1 and ADH1C2 after *X. tropicalis* ADH1C (84 and 89% similarity, respectively) (Figure 6). Gene family studies with globin (Hosbach *et al*, 1983) and α -actin sequences (Stutz and Spohr, 1987), and the genetic map of eight linkage groups, including the *Adh* locus (Graham, 2000), all support the allotetraploid origin of the *X. laevis* genome (reviewed in Tymowska, 1991), which we assume to be the basis of the multiplicity. Structurally, the presumptive amphibian ADH1 enzymes show the Thr/Ser48 proton relay component, the signature class I residues, His67, Glu68 and Phe140, and all, except *X. laevis* ADH1C1, His51 (Thr 51 in *X. laevis* ADH1C1). Therefore, from sequence comparisons, we propose assigning the *Xenopus* enzymes to class I.

Based on the expression pattern (present in the stomach, oesophagus and skin but absent in the liver), a *X. laevis* ADH sequence was classified as a class IV enzyme (Hoffmann *et al*, 1998). However, no data about its biochemical activity were available, and phylogenies with the ADH members were not unequivocal. In fact, *Xenopus* ADH4 shows highest similarity with other amphibian ADH1 (65.4% with ADH1C2 and 64.5% with ADH1B2) (Figure 6) and although these values are still too low for definite ascription, they are higher than those obtained with mammalian ADH4 (around 55%). There is no strong evidence in favour of an orthologous relationship of the amphibian enzyme with the mammalian class IV forms. Instead, an amphibian-specific duplication followed by subfunctionalization (Force *et al*, 1999) would be a plausible outcome. More information on the biochemical properties of the *Xenopus* ADH4 and on the expression pattern of the ADH members in lobe-finned fishes (the tetrapod sister group) should shed light on this issue.

Finally, a *R. perezy* ADH (termed ADH8) found to be active against retinoids (Peralba *et al*, 1999) added more complexity to the amphibian family. Although initially ascribed to class IV on the basis of its substrate specificity and gastric localization, as mammalian class IV forms, the three-dimensional structure (Rosell *et al*, 2003b; Valencia *et al*, 2003), *in vitro* behaviour – more as a retinal reductase than alcohol dehydrogenase – and its unique preference among vertebrate ADHs for NADP⁺ rather than NAD⁺, all suggested a separate class, named class VIII. It has been proposed that the triad Gly223–Thr224–His225, together with Leu200 and Lys228, may account for the cofactor preference, also verified by site-directed mutagenesis at the triad segment (Rosell *et al*, 2003a). Interestingly, in *Xenopus*, we have derived sequences that are highly similar to *R. perezy* ADH8: XIADH8 (73.1% similarity), XtADH8A (71.7%) and XtADH8B (66.7%) (Table 1 and Figure 6). Preservation

or conservative substitutions (Ser224 and Gln225, the latter only in *X. laevis*) point to a similar cofactor preference and biochemical role for these enzymes. Until now, class VIII ADH has been restricted to the amphibian lineage.

ADH in amniotas

ADH5 and ADH6: ADH5 and ADH6 are the least analysed classes and, thus far, only identified at the DNA level. *Adh5* genes have been reported in human (Yasunami *et al*, 1991; Stromberg and Höög, 2000), deer mouse (*Peromyscus maniculatus*) (Zheng *et al*, 1993) and rat (Höög and Brandt, 1995). Multiple mRNAs are produced by alternative polyadenylation and splicing events (Stromberg and Höög, 2000; Höög *et al*, 2001a). The longest transcript (3.3 kb) was found to be the most prominent. The highest transcription levels were detected in the liver, and weaker signals were observed in the small intestine and kidney, although discrepancies in the relative abundances in fetal and adult tissues have been reported (Estonius *et al*, 1996; Stromberg and Höög, 2000). The predicted protein sequences of human and rodent ADH5 showed important substitutions with respect to the other classes – including Gly47, in contrast to Arg or His in most ADH forms, and Lys51 instead of His in class I enzymes –, which call into question its ethanol oxidizing activity (Höög *et al*, 2001a).

In the human ADH cluster, the class 5 gene was located between *Adh1* and *Adh2*, whereas two *Adh* plus a pseudogene were found between mouse *Adh1* and *Adh2* (Szalai *et al*, 2002) (Figure 3). The sequence similarity of the three murine copies, *Adh5A*, *Adh5B* and *Adh5ps*, with the human *Adh5* and the equivalent position in the cluster suggested a close relationship. However, mouse *Adh5A* and *Adh5B* are closely related with rat *Adh6A* (92.2% similarity) and *Adh6B* (80.1%), respectively, while *Adh5ps* clustered with deer mouse, rat and human *Adh5* (75.9, 74.1 and 64.3% similarity, respectively) (Figure 6). Moreover, the mouse and rat *Adh* clusters show the same physical arrangement for these genes (Figure 3). We suggest that mouse *Adh5A* would be orthologous to rat *Adh6A* and mouse *Adh5B* to rat *Adh6B*, and hence they should be renamed *Adh6*, while mouse *Adh5ps* would be orthologous to rat, deer mouse and human *Adh5*. No information has been gathered about the expression pattern of *Adh6*. Interestingly, the rat and mouse enzyme has Gly at position 47, like ADH5 enzymes, but His at position 51, like ADH1. Moreover, all ADH6, and also rodent ADH5, lack Phe140, an otherwise strictly conserved residue in ethanol-active enzymes. Therefore, ADH6 activity against ethanol is still an open question.

In conclusion, ADH6 appears to conform a new class, hitherto restricted to rodents, probably generated by an additional tandem duplication before the rat–mouse divergence, 23 MYA (Adkins *et al*, 2001). The origin of the class V remains uncertain. If mammalian class V corresponds to amphibian–avian class VII, class V/VII enzymes would be common in tetrapods, while if class V is restricted to mammals, it would have a more recent origin, after the appearance of mammals, 300 MYA.

ADH4: The high ability of ADH4 to oxidize retinoids *in vitro* (Boleda *et al*, 1993; Yang *et al*, 1994; Allali-Hassani *et al*, 1998) prompted the view that this enzyme was a retinol dehydrogenase (Ang *et al*, 1996; Duester, 1996).

However, the weak phenotypic effects observed on *Adh4*-null mutant mice (mainly increased postnatal lethality during gestational vitamin A deficiency (VAD); Deltour *et al*, 1999a) challenged the hypothesis of the systemic contribution of ADH4 to retinol metabolism (Molotkov *et al*, 2002b) and suggested that ADH4 contributed to survival only during gestational VAD and in tissues with high RA requirements (Molotkov *et al*, 2002b).

Based on comparative analyses, we propose that the *Adh4* gene arose from a mammalian-specific duplication of the adjacent *Adh1* copy (Figure 5c), since no *Adh4* forms have been detected in the avian/reptilian line and amphibian *Adh4* does not seem to be orthologous to the mammalian form (see above). In mammals, the evolution of a high-activity retinol dehydrogenase enzyme could have been relevant for the developing embryo, which, in contrast to oviparous animals that store RA precursors in the egg, relies on maternal retinoids that cross the maternal–fetal barrier. As a result, depending on the maternal diet, embryonic retinoid concentrations can vary significantly and, if severely restricted, lead to the ‘VAD syndrome’. The evolution of a high-activity retinol dehydrogenase that could compensate for VAD, a condition that may be common in the wild, would have been selectively advantageous.

Final considerations and perspectives

The ADH-MDR family has been studied in depth, due to the widespread occurrence of the ancestral member in bacteria, yeast, animal and plant species and to the inherent interest of the ‘ethanol related’ enzymes. Indeed, the horse liver alcohol dehydrogenase, a class I ADH, was the first oligomeric enzyme for which a primary (Jörnvall, 1970) and tertiary structure (Eklund *et al*, 1974, 1976) were established. These data, together with the catalytic mechanism of the enzyme (Theorell and McKee, 1961), prepared the way for future work and encouraged further studies, mainly at the biochemical level. More recently, genomic and EST data provided extremely valuable tools for uncovering new family members, which, when combined with the structural and biochemical data, allow a comprehensive picture of the evolution of the family.

Interestingly, while single-copy gene status was preserved during invertebrate evolution, the initial tandem repeat in vertebrates provided the genetic material for, probably by unequal crossover events, future independent extensions in fish, amphibian and amniota lineages, which led to the extant ADH classes. Moreover, additional multiplicity within classes was generated by further duplications, mainly ADH1A, B and C isoforms in human, 1E and S in horse, 1A and B in lizard, 1A (A1 and A2), B (B1 and B2) and C (C1 and C2) in *Xenopus*, 1A (A1 and A2) and B (B1 and B2) in fish, ADH2A and B in rabbit, ADH3 H and L in fish, ADH6A and B in rodents, and ADH8A and B in *Xenopus*. The ADH family expansion exemplifies a neofunctionalization process with reiterative duplication events leading to new activities. Not surprisingly, class ascription of the iterated members is restrained by overlaps in the biochemical capacities, differences in the evolutionary rates and gene conversion events. Besides, the high multiplicity has constrained the biological meaning of the individual forms. Beyond the *in vitro* biochemical assays, better

knowledge on substrate recognition could be attained with cells overexpressing a single ADH form coupled with RNAi inhibition studies to minimize redundancies. Further, the *in vivo* function and phenotypic effects could be approached taking advantage of the cluster arrangement of all ADH forms through the construction of full and partial knockout mouse models. Finally, an important point to be addressed is how redundancy is maintained and to what extent it affects the viability of the organism. This study highlights the relevance of merging data at the protein, gene and genomic level to understand the mechanisms that underlie the generation of new functions and opens new avenues for evaluating the impact of redundancy at the evolutionary level.

Acknowledgements

We thank Robin Rycroft and the Servei d'Assessorament Lingüístic for revising the English. This study was funded by Ministerio de Ciencia y Tecnología, Grant BMC2003-05211 to RG-D.

References

- Adkins RM, Gelke EL, Rowe D, Honeycutt RL (2001). Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. *Mol Biol Evol* **18**: 777–791.
- Allali-Hassani A, Peralba JM, Martras S, Farrés J, Parés X (1998). Retinoids, omega-hydroxyfatty acids and cytotoxic aldehydes as physiological substrates, and H2-receptor antagonists as pharmacological inhibitors, of human class IV alcohol dehydrogenase. *FEBS Lett* **426**: 362–366.
- Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A *et al* (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**: 1711–1714.
- 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.
- Boleda MD, Saubi N, Farrés J, Parés X (1993). Physiological substrates for rat alcohol dehydrogenase classes: aldehydes of lipid peroxidation, omega-hydroxyfatty acids, and retinoids. *Arch Biochem Biophys* **307**: 85–90.
- Brändén CI, Eklund H, Cambillau C, Pryor AJ (1984). Correlation of exons with structural domains in alcohol dehydrogenase. *EMBO J* **3**: 1307–1310.
- Cañestro C, Albalat R, Hjelmqvist L, Godoy L, Jörnvall H, González-Duarte R (2002). Ascidian and amphioxus *Adh* genes correlate functional and molecular features of the ADH family expansion during vertebrate evolution. *J Mol Evol* **54**: 81–89.
- Cañestro C, Godoy L, González-Duarte R, Albalat R (2003). Comparative expression analysis of *Adh3* during arthropod, urochordate, cephalochordate and vertebrate development challenges its predicted housekeeping role. *Evol Dev* **5**: 157–162.
- Cañestro C, Hjelmqvist L, Albalat R, Garcia-Fernández J, González-Duarte R, Jörnvall 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.
- Cederlund E, Peralba JM, Parés X, Jörnvall H (1991). Amphibian alcohol dehydrogenase, the major frog liver enzyme: relationships to other forms and assessment of an early gene duplication separating vertebrate class I and class III alcohol dehydrogenases. *Biochemistry* **30**: 2811–2816.

- Cho G, Doolittle RF (1997). Intron distribution in ancient paralogs supports random insertion and not random loss. *J Mol Evol* **44**: 573–584.
- Costaridis P, Horton C, Zeitlinger J, Holder N, Maden M (1996). Endogenous retinoids in the zebrafish embryo and adult. *Dev Dyn* **205**: 41–51.
- 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, El Ahmad M, Jörnvall H (1996). Isozyme multiplicity with anomalous dimer patterns in a class III alcohol dehydrogenase: effects on the activity and quaternary structure of residue exchanges at ‘nonfunctional’ sites in a native protein. *Biochemistry* **35**: 14561–14568.
- Danielsson O, Shafqat J, Estonius M, Jörnvall H (1994b). Alcohol dehydrogenase 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.
- Dasmahapatra AK, Doucet HL, Bhattacharyya C, Carvan III MJ (2001). Developmental expression of alcohol dehydrogenase (ADH3) in zebrafish (*Danio rerio*). *Biochem Biophys Res Commun* **286**: 1082–1086.
- Davis GJ, Carr LG, Hurley TD, Li TK, Bosron WF (1994). Comparative roles of histidine 51 in human beta 1 beta 1 and threonine 51 in pi pi alcohol dehydrogenases. *Arch Biochem Biophys* **311**: 307–312.
- Deltour L, Foglio MH, Duester G (1999a). Impaired retinol utilization in *Adh4* alcohol dehydrogenase mutant mice. *Dev Genet* **25**: 1–10.
- Deltour L, Foglio MH, Duester G (1999b). Metabolic deficiencies in alcohol dehydrogenase *Adh1*, *Adh3*, and *Adh4* null mutant mice – overlapping roles of *Adh1* and *Adh4* in ethanol clearance and metabolism of retinol to retinoic acid. *J Biol Chem* **274**: 16796–16801.
- Ditlow CC, Holmquist B, Morelock MM, Vallee BL (1984). Physical and enzymatic properties of a class II alcohol dehydrogenase isozyme of human liver: pi-ADH. *Biochemistry* **23**: 6363–6368.
- Dolferus R, Osterman JC, Peacock WJ, Dennis ES (1997). Cloning of the *Arabidopsis* and rice formaldehyde dehydrogenase genes: implications for the origin of plant ADH enzymes. *Genetics* **146**: 1131–1141.
- Duester G (1996). Involvement of alcohol dehydrogenase, short-chain dehydrogenase/reductase, aldehyde dehydrogenase, and cytochrome P450 in the control of retinoid signaling by activation of retinoic acid synthesis. *Biochemistry* **35**: 12221–12227.
- Duester G (2000). Families of retinoid dehydrogenases regulating vitamin A function: production of visual pigment and retinoic acid. *Eur J Biochem* **267**: 4315–4324.
- Duester G, Farrés J, Felder MR, Holmes RS, Höög JO, Parés X *et al* (1999). Recommended nomenclature for the vertebrate alcohol dehydrogenase gene family. *Biochem Pharmacol* **58**: 389–395.
- Duester G, Jörnvall H, Hatfield GW (1986). Intron-dependent evolution of the nucleotide-binding domains within alcohol dehydrogenase and related enzymes. *Nucleic Acids Res* **14**: 1931–1941.
- Eklund H, Müller-Wille P, Horjales E, Futer O, Holmquist B, Vallee BL *et al* (1990). Comparison of three classes of human liver alcohol dehydrogenase: emphasis on different substrate binding pockets. *Eur J Biochem* **193**: 303–310.
- Eklund H, Nordström B, Zeppezauer E, Söderlund G, Ohlsson I, Boiwe T *et al* (1974). The structure of horse liver alcohol dehydrogenase. *FEBS Lett* **44**: 200–204.
- Eklund H, Nordstrom B, Zeppezauer E, Soderlund G, Ohlsson I, Boiwe T *et al* (1976). Three-dimensional structure of horse liver alcohol dehydrogenase at 2–4 Å resolution. *J Mol Biol* **102**: 27–59.
- 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, Biosca JA, Parés X (2003). S-nitrosoglutathione reductase activity of human and yeast glutathione-dependent formaldehyde dehydrogenase and its nuclear and cytoplasmic localisation. *Cell Mol Life Sci* **60**: 1013–1018.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.
- Funkenstein B, Jakowlew SB (1996). Molecular cloning of fish alcohol dehydrogenase cDNA. *Gene* **174**: 159–164.
- Gilbert W (1987). The exon theory of genes. *Cold Spring Harb Symp Quant Biol* **52**: 901–905.
- Gogarten JP, Olendzenski L (1999). Orthologs, paralogs and genome comparisons. *Curr Opin Genet Dev* **9**: 630–636.
- Gotoh Y, Sumimoto H, Minakami S (1990). Formation of 20-oxoleukotriene B4 by an alcohol dehydrogenase isolated from human neutrophils. *Biochim Biophys Acta* **1043**: 52–56.
- Graham A (2000). The evolution of the vertebrates-genes and development. *Curr Opin Genet Dev* **10**: 624–628.
- Graur D, Li W-H (1999). *Fundamentals of Molecular Evolution*. Sinauer Associates Inc.: Sunderland, MA.
- Hjelmqvist L, Estonius M, Jörnvall H (1995). The vertebrate alcohol dehydrogenase system: variable class II type form elucidates separate stages of enzymogenesis. *Proc Natl Acad Sci USA* **92**: 10904–10908.
- Hjelmqvist L, Norin A, El-Ahmad M, Griffiths W, Jörnvall H (2003). Distinct but parallel evolutionary patterns between alcohol and aldehyde dehydrogenases: addition of fish/human betaine aldehyde dehydrogenase divergence. *Cell Mol Life Sci* **60**: 2009–2016.
- Hoffmann I, Ang HL, Duester G (1998). Alcohol dehydrogenases in *Xenopus* development: conserved expression of ADH1 and ADH4 in epithelial retinoid target tissues. *Dev Dyn* **213**: 261–270.
- Höög JO, Brandt M (1995). Mammalian class VI alcohol dehydrogenase. Novel types of the rodent enzymes. *Adv Exp Med Biol* **372**: 355–364.
- Höög JO, Brandt M, Hedberg JJ, Stromberg P (2001a). Mammalian alcohol dehydrogenase of higher classes: analyses of human ADH5 and rat ADH6. *Chem Biol Interact* **130–132**: 395–404.
- Höög JO, Hedberg JJ, Stromberg P, Svensson S (2001b). Mammalian alcohol dehydrogenase – functional and structural implications. *J Biomed Sci* **8**: 71–76.
- Hosbach HA, Wyler T, Weber R (1983). The *Xenopus laevis* globin gene family: chromosomal arrangement and gene structure. *Cell* **32**: 45–53.
- Iborra FJ, Renau Piqueras J, Portoles M, Boleda MD, Guerri C, Parés X (1992). Immunocytochemical and biochemical demonstration of formaldehyde dehydrogenase (class III alcohol dehydrogenase) in the nucleus. *J Histochem Cytochem* **40**: 1865–1878.
- Irie T, Kajiwara S, Seki T (2003). Storage of retinal in the eggs of the ascidian, *Halocynthia roretzi*. *Comp Biochem Physiol B* **134**: 221–230.
- Jensen DE, Belka GK, Du Bois GC (1998). S-nitrosoglutathione is a substrate for rat alcohol dehydrogenase class III isoenzyme. *Biochem J* **2**: 659–668.
- Jörnvall H (1970). Horse liver alcohol dehydrogenase. The primary structure of the protein chain of the ethanol-active isoenzyme. *Eur J Biochem* **16**: 25–40.
- Jörnvall H, Höög JO, Persson B (1999). SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Lett* **445**: 261–264.
- Jörnvall H, Höög JO, Persson B, Parés X (2000). Pharmacogenetics of the alcohol dehydrogenase system. *Pharmacology* **61**: 184–191.

- Kaiser R, Fernández MR, Parés X, Jörnvall H (1993). Origin of the human alcohol dehydrogenase system: implications from the structure and properties of the octopus protein. *Proc Natl Acad Sci USA* **90**: 11222–11226.
- Kedishvili N, Gough W, Chernoff E, Hurley T, Stone C, Bowman K *et al* (1997). cDNA sequence and catalytic properties of a chick embryo alcohol dehydrogenase that oxidizes retinol and 3-beta,5-alpha-hydroxysteroids. *J Biol Chem* **272**: 7494–7500.
- 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.
- Lampert JM, Holzschuh J, Hessel S, Driever W, Vogt K, von Lintig J (2003). Provitamin A conversion to retinal via the beta,beta-carotene-15,15'-oxygenase (bcx) is essential for pattern formation and differentiation during zebrafish embryogenesis. *Development* **130**: 2173–2186.
- Liu L, Hausladen A, Zeng M, Que L, Heitman J, Stamler JS (2001). A metabolic enzyme for S-nitrosothiol conserved from bacteria to humans. *Nature* **410**: 490–494.
- Liu L, Yan Y, Zeng M, Zhang J, Hanes MA, Ahearn G *et al* (2004). Essential roles of S-nitrosothiols in vascular homeostasis and endotoxic shock. *Cell* **116**: 617–628.
- Luque T, Arrian S, Danielsson O, Jörnvall H, González-Duarte R (1994). Structure of the *Drosophila melanogaster* glutathione-dependent formaldehyde dehydrogenase/octanol dehydrogenase gene (class III alcohol dehydrogenase). Evolutionary pathway of the alcohol dehydrogenase genes. *Eur J Biochem* **225**: 985–993.
- Michelson AM, Blake CC, Evans ST, Orkin SH (1985). Structure of the human phosphoglycerate kinase gene and the intron-mediated evolution and dispersal of the nucleotide-binding domain. *Proc Natl Acad Sci USA* **82**: 6965–6969.
- Molotkov A, Deltour L, Foglio MH, Cuenca AE, Duester G (2002a). Distinct retinoid metabolic functions for alcohol dehydrogenase genes *Adh1* and *Adh4* in protection against vitamin A toxicity or deficiency revealed in double null mutant mice. *J Biol Chem* **277**: 13804–13911.
- Molotkov A, Fan X, Deltour L, Foglio MH, Martras S, Farrés J *et al* (2002b). Stimulation of retinoic acid production and growth by ubiquitously expressed alcohol dehydrogenase *Adh3*. *Proc Natl Acad Sci USA* **99**: 5337–5342.
- Nordling E, Jörnvall H, Persson B (2002). Medium-chain dehydrogenases/reductases (MDR). Family characterizations including genome comparisons and active site modeling. *Eur J Biochem* **269**: 4267–4276.
- Norin A, Shafqat J, El-Ahmad M, Alvelius G, Cederlund E, Hjelmqvist L *et al* (2004). Class III alcohol dehydrogenase: consistent pattern complemented with the mushroom enzyme. *FEBS Lett* **559**: 27–32.
- Peralba JM, Cederlund E, Crosas B, Moreno A, Julia P, Martínez SE *et al* (1999). Structural and enzymatic properties of a gastric NADP(H)-dependent and retinal-active alcohol dehydrogenase. *J Biol Chem* **274**: 26021–26026.
- Philippe H, Lopez P (2001). On the conservation of protein sequences in evolution. *Trends Biochem Sci* **26**: 414–416.
- Quigley F, Martin WF, Cerff R (1988). Intron conservation across the prokaryote-eukaryote boundary: structure of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. *Proc Natl Acad Sci USA* **85**: 2672–2676.
- Ramaswamy S, Eklund H, Plapp BV (1994). Structures of horse liver alcohol dehydrogenase complexed with NAD⁺ and substituted benzyl alcohols. *Biochemistry* **33**: 5230–5237.
- Reimers MJ, Hahn ME, Tanguay RL (2004). Two zebrafish alcohol dehydrogenases share common ancestry with mammalian class I, II, IV, and V ADH genes but have distinct functional characteristics. *J Biol Chem* **279**: 38303–38312.
- Rosell A, Valencia E, Ochoa WF, Fita I, Parés X, Farrés J (2003a). Complete reversal of coenzyme specificity by concerted mutation of three consecutive residues in alcohol dehydrogenase. *J Biol Chem* **278**: 40573–40580.
- Rosell A, Valencia E, Parés X, Fita I, Farrés J, Ochoa WF (2003b). Crystal structure of the vertebrate NADP(H)-dependent alcohol dehydrogenase (ADH8). *J Mol Biol* **330**: 75–85.
- Sakamoto A, Ueda M, Morikawa H (2002). *Arabidopsis* glutathione-dependent formaldehyde dehydrogenase is an S-nitrosoglutathione reductase. *FEBS Lett* **515**: 20–24.
- Sanghani PC, Bosron WF, Hurley TD (2002). Human glutathione-dependent formaldehyde dehydrogenase. Structural changes associated with ternary complex formation. *Biochemistry* **41**: 15189–15194.
- Sanghani PC, Robinson H, Bennett-Lovsey R, Hurley TD, Bosron WF (2003). Structure–function relationships in human Class III alcohol dehydrogenase (formaldehyde dehydrogenase). *Chem Biol Interact* **143–144**: 195–200.
- Sanghani PC, Stone CL, Ray BD, Pindel EV, Hurley TD, Bosron WF (2000). Kinetic mechanism of human glutathione-dependent formaldehyde dehydrogenase. *Biochemistry* **39**: 10720–10729.
- Stoltzfus A, Spencer DF, Zuker M, Logsdon Jr JM, Doolittle WF (1994). Testing the exon theory of genes: the evidence from protein structure. *Science* **265**: 202–207.
- Stone EM, Rothblum KN, Schwartz RJ (1985). Intron-dependent evolution of chicken glyceraldehyde phosphate dehydrogenase gene. *Nature* **313**: 498–500.
- Stromberg 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.
- Stutz F, Spohr G (1987). A processed gene coding for a sarcomeric actin in *Xenopus laevis* and *Xenopus tropicalis*. *EMBO J* **6**: 1989–1995.
- Svensson S, Höög JO, Schneider G, Sandalova T (2000). Crystal structures of mouse class II alcohol dehydrogenase reveal determinants of substrate specificity and catalytic efficiency. *J Mol Biol* **302**: 441–453.
- Svensson S, Stromberg P, Höög JO (1999). A novel subtype of class II alcohol dehydrogenase in rodents. Unique Pro(47) and Ser(182) modulates hydride transfer in the mouse enzyme. *J Biol Chem* **274**: 29712–29719.
- Szalai G, Duester G, Friedman R, Jia H, Lin S, Roe BA *et al* (2002). Organization of six functional mouse alcohol dehydrogenase genes on two overlapping bacterial artificial chromosomes. *Eur J Biochem* **269**: 224–232.
- Theorell H, McKee J (1961). Mechanism of action of liver alcohol dehydrogenase. *Nature* **192**: 47–50.
- Thompson HR, Jones GM, Narkewicz MR (2001). Ontogeny of hepatic enzymes involved in serine- and folate-dependent one-carbon metabolism in rabbits. *Am J Physiol Gastrointest Liver Physiol* **280**: G873–878.
- Tymowska J (1991). Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. In: Green DM, Sessions SK (eds) *Amphibian Cytogenetics and Evolution*. Academic Press: San Diego.
- Uotila L, Koivusalo M (1974). Formaldehyde dehydrogenase from human liver. Purification, properties, and evidence for the formation of glutathione thiol esters by the enzyme. *J Biol Chem* **249**: 7653–7663.
- Uotila L, Koivusalo M (1989). *Glutathione-dependent Oxidoreductases: Formaldehyde Dehydrogenase*. John Wiley & Sons: New York.
- Uotila L, Koivusalo M (1997). Expression of formaldehyde dehydrogenase and S-formylglutathione hydrolase activities in different rat tissues. *Adv Exp Med Biol* **414**: 365–371.
- Valencia E, Rosell A, Larroy C, Farrés J, Biosca JA, Fita I *et al* (2003). Crystallization and preliminary X-ray analysis of

- NADP(H)-dependent alcohol dehydrogenases from *Saccharomyces cerevisiae* and *Rana perezi*. *Acta Crystallogr D* **59**: 334–337.
- Wagner FW, Parés X, Holmquist B, Vallee BL (1984). Physical and enzymatic properties of a class III isozyme of human liver alcohol dehydrogenase: chi-ADH. *Biochemistry* **23**: 2193–2199.
- White ME, Crother BI (2000). Gene conversions may obscure actin gene family relationships. *J Mol Evol* **50**: 170–174.
- Yang Z, Davis G, Hurley T, Stone C, Li T, Bosron W (1994). Catalytic efficiency of human alcohol dehydrogenases for retinol oxidation and retinal reduction. *Alcohol Clin Exp Res* **18**: 587–591.
- Yang ZN, Bosron WF, Hurley TD (1997). Structure of human chi chi alcohol dehydrogenase: a glutathione-dependent formaldehyde dehydrogenase. *J Mol Biol* **265**: 330–343.
- Yasunami M, Chen CS, Yoshida A (1991). A human alcohol dehydrogenase gene (ADH6) encoding an additional class of isozyme. *Proc Natl Acad Sci USA* **88**: 7610–7614.
- Zhang J (2003). Evolution by gene duplication: an update. *Trends Ecol Evol* **18**: 292–298.
- Zheng YW, Bey M, Liu H, Felder MR (1993). Molecular basis of the alcohol dehydrogenase-negative deer mouse. Evidence for deletion of the gene for class I enzyme and identification of a possible new enzyme class. *J Biol Chem* **268**: 24933–24939.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)