# Comparative and evolutionary studies of vertebrate ALDH1A-like genes and proteins


## Author

Holmes, Roger S



## Published

2015

## Journal Title

Chemico-Biological Interactions

## Version

Accepted Manuscript (AM)

## DOI

10.1016/j.cbi.2014.11.002

**Comparative and Evolutionary Studies of Vertebrate *ALDH1A*-Like Genes and Proteins**

**Roger S Holmes**

The Eskitis Institute for Drug Discovery and School of Natural Sciences, Griffith University, Nathan 4111 QLD AUSTRALIA
Email: r.holmes@griffith.edu.au

## ABSTRACT

Vertebrate *ALDH1A*-like genes encode cytosolic enzymes capable of metabolizing all-trans-retinaldehyde to retinoic acid, a molecular 'signal' guiding vertebrate development and adipogenesis. Bioinformatic analyses of vertebrate genomes were undertaken using known ALDH1A1, ALDH1A2 and ALDH1A3 amino acid sequences. Comparative analyses of the *ALDH1A*-like human genes provided evidence for distinct modes of gene regulation and expression with the identification of transcription factor binding sites (TFBS), CpG islands and micro-RNA binding sites. Phylogenetic analyses of a primitive vertebrate genome (lamprey) and representative vertebrate genomes indicated that three *ALDH1A*-like genes were present for most vertebrates. The *ALDH1A1* gene was absent in many bony fish genomes examined, with the *ALDH1A3* gene also absent in the medaka and tilapia genomes. Multiple *ALDH1A1*-like genes were identified in mouse, rat and marsupial genomes. Vertebrate ALDH1A1, ALDH1A2 and ALDH1A3 subunit sequences were highly conserved throughout vertebrate evolution. Comparative amino acid sequence studies showed that mammalian ALDH1A2 sequences were more highly conserved than for the ALDH1A1 and ALDH1A3 sequences. Phylogenetic analyses suggested that *ALDH1A2* is the likely primordial gene in early vertebrates. Subsequent gene duplication events have apparently generated *ALDH1A1*, *ALDH1A2* and *ALDH1A3* genes in most vertebrate genomes, with the exception of some bony fish.

**Keywords:** Aldehyde dehydrogenases; retinaldehyde dehydrogenase; retinal; retinoic acid; ALDH1A1; ALDH1A2; ALDH1A3; vertebrates; gene duplication; evolution; phylogeny; primordial gene; transcription factor binding sites; CpG islands; miRNA binding sites.

**Abbreviations:** ALDH: aldehyde dehydrogenase; BLAST: basic local alignment search tool; BLAT: blast-like alignment tool; NCBI: National Center for Biotechnology Information; KO: knock out; AceView: NCBI based representation of public mRNAs; TFBS: transcription factor binding sites; UTR: untranslated region; CpG: region of high density of guanine-cytosine dinucleotides; miRNA: small noncoding RNA miRNA binding site in the 3'UTR; mRNA: messenger RNA; RA: retinoic acid (or retinoate).

## INTRODUCTION

Vertebrate aldehyde dehydrogenase (*ALDH*; EC 1.2.1.3) genes encode many enzymes performing diverse roles in aldehyde metabolism in the body (http://aldh.org/superfamily.php) [1]. Several related *ALDH1* and *ALDH2* like genes and encoded enzymes have been described, including three retinaldehyde dehydrogenases, designated as ALDH1A1 (or RALDH1) [2,3], ALDH1A2 (or RALDH2) [4] and ALDH1A3 (or RALDH3 and ALDH6) [5-8]. These genes are separately localized on the human genome: *ALDH1A1* on chromosome 9, encoding cytosolic liver, red cell, adipose tissue, cornea, lens and stem cell ALDH1A1 [2,3,10-16]; *ALDH1A2* on chromosome 15, encoding embryonic forebrain cytosolic ALDH1A2, which plays a major role in

generating retinoic acid as a signal for guiding development [17-21]; and *ALDH1A3*, also on chromosome 15, encoding cytosolic ventral sensory gland ALDH1A3 functioning as a molecular signal during neural development [6-8].

This study describes the predicted sequences, structures and phylogeny of vertebrate *ALDH1A1, ALDH1A2* and *ALDH1A3* genes and enzymes and compares these results for those previously reported for the corresponding human (*Homo sapiens*) and mouse (*Mus musculus*) genes and proteins [2,4,5,22,23]. Sequences for vertebrate ALDH1A1, ALDH1A2 and ALDH1A3 enzymes and gene locations for *ALDH1A1, ALDH1A2* and *ALDH1A3* genes were obtained using data from respective genome sequences. This paper also describes evidence for distinct modes of gene regulation and expression with transcription factor binding sites (TFBS), CpG islands and micro-RNA binding sites identified for the human *ALDH1A1*, *ALDH1A2* and *ALDH1A3* genes. Phylogenetic analyses also describe the relationships and potential origins of *ALDH1A1, ALDH1A2* and *ALDH1A3* genes during vertebrate evolution as well as evidence for *ALDH1A2* serving as the primordial gene during early vertebrate evolution.

## MATERIALS AND METHODS

### Vertebrate *ALDH1A1, ALDH1A2* and *ALDH1A3* gene and enzyme identification

ALDH1A1, ALDH1A2 and ALDH1A3 sequences for representative vertebrate species were retrieved from two major databases (NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and ExPASy (http://www.expasy.org) [24]) using human ALDH1A1 ([2,22], ALDH1A2 [17] and ALDH1A3 [5] sequences to seed searches. Identification of these genes in vertebrate genomes was based on high predictive scores (>850) and sequence coverage (>98%) for ALDH proteome sequences listed by NCBI, in each case (Table 1 and Supplementary Tables 1s, 2s and 3s). BLAT searches were performed using relevant ALDH1A1, ALDH1A2 and ALDH1A3 protein sequences to confirm the presence or absence of these genes among the vertebrate species examined using the UCSC Genome Browser [25]. Gene locations, predicted gene structures and ALDH1A1, ALDH1A2 and ALDH1A3 protein subunit sequences were obtained for each ALDH examined showing identity with the respective ALDH sequences (Table 1; Tables 1s, 2s and 3s). Representations of human *ALDH1A1, ALDH1A2* and *ALDH1A3* gene structures were obtained using the AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/) web browser. Identification of potential *ALDH1A1*, *ALDH1A2* and *ALDH1A3* gene regulatory sites, including

transcription factor binding sites (TFBS) and CpG islands within the respective gene promoter regions, and miRNA binding sites within the 3'UTR regions, was undertaken using the UCSC Human Genome Browser [25].

**Amino acid sequence alignments and phylogenetic analyses**

All vertebrate ALDH1A1, ALDH1A2 and ALDH1A3 sequences were subjected to phylogenetic analysis using bioinformatic programs, coordinated using the http://www.phylogeny.fr/ portal, to enable alignment (MUSCLE), curation (Gblocks), phylogeny (PhyML) and tree rendering (TreeDyn) to reconstruct phylogenetic relationships [26]. Sequences were identified as members of one of the ALDH1A1, ALDH1A2 or ALDH1A3 groups, including the proposed primordial vertebrate *ALDH1A2* gene and protein identified for the primitive vertebrate, lamprey (*Petromyzon marinus*) [27].

## RESULTS AND DISCUSSION

**Alignments of vertebrate ALDH1A1, ALDH1A2 and ALDH1A3 amino acid sequences**

Amino acid sequence alignments for human (*Homo sapiens*) ALDH1A1 [22], ALDH1A2 [17], ALDH1A3 [5], tilapia or nile perch (*Oreochromis niloticus*) ALDH1A1, zebra fish (*Danio rerio*) ALDH1A2 and ALDH1A3 (see Table 1) are shown in Figure 1. Comparisons of these sequences with the sheep ALDH1A1 (PDB: 1BXSA) [28] and rat ALDH1A2 (PDB: 1BI9) [29] sequences, for which tertiary structures have been described, enabled identification of key residues contributing to catalysis, structure and function. Active site residues (human ALDH1A2 numbers used) binding the substrate (Glu286 and Cys320) or stabilizing the transition state for the catalyzed reaction (Asn187) were conserved for all of these ALDH1A1, ALDH1A2 and ALDH1A3 sequences. A dinucleotide-binding motif and $NAD^+$ binding domain (263Gly-Ser-Thr-Glu-Val-268Gly) [28,29] was conserved for these vertebrate ALDH1A-like sequences examined, with the exception of a 266Glu/Ala substitution for the tilapia ALDH1A1 sequence (Figure 1). Exon 1 N-terminal sequences were divergent and varied in length from 22 (human ALDH1A1) to 74 (tilapia ALDH1A1 for exons 1 and 2) residues (Figure 1). Percentage identities for these aligned human and teleost fish ALDH1A1, ALDH1A2 and ALDH1A3 sequences were in the range of 69-72% indicating that these sequences were strongly conserved throughout vertebrate evolution.

**Predicted gene locations, exonic structures and amino acid sequences for vertebrate *ALDH1A1, ALDH1A2* and *ALDH1A3* genes and proteins**

Table 1 and Tables 1s, 2s and 3s (supplementary data) summarize the predicted locations, sizes and numbers of coding exons for vertebrate *ALDH1A1, ALDH1A2* and *ALDH1A3* genes examined, and of encoded ALDH1A-like subunit amino acid sequences. These were based on BLAST interrogations of *ALDH1A1, ALDH1A2* and *ALDH1A3* databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using the reported sequences for human ALDH1A1 [2,22], ALDH1A2 [17] and ALDH1A3 [5] and BLAT analyses of vertebrate genomes using the UC Santa Cruz Genome Browser (http://genome.ucsc.edu/cgi-bin/hgBlat) [25]. In nearly all cases, there were 13 coding exons observed (Table 1; Tables 1s, 2s and 3s), whereas tilapia (*Oreochromis niloticus*) *ALDH1A1* contained an additional coding exon contributing to an extended N-terminal sequence for this protein (Figure 1). Other African lake fish ALDH1A1 sequences examined, zebra mbuna (*Maylandia zebra*) and red zuma (*Haplochromis nyererei*), also exhibited extended N-terminal amino acid sequences, with 553 residues. In contrast to all other vertebrate genomes, bonyfish genomes contained only one (*ALDH1A2*: medaka, *Oryzias latipes*) or two *ALDH1A*-like genes [*ALDH1A1* and *ALDH1A2*, for tilapia (*Oreochromis niloticus*); and *ALDH1A2* and *ALDH1A3,* for pufferfish (*Takifugu rupides*) and stickleback (*Gasterosteus aculeatus*)], whereas the primitive vertebrate genome (lamprey, *Petromyzon marinus*), contained only the *ALDH1A2* gene. This suggested that *ALDH1A2* served as the primordial vertebrate *ALDH1A*-like gene, from which *ALDH1A1* and *ALDH1A3* have been generated by subsequent gene duplication events.

Figure 2 gives a diagrammatic illustration of comparative locations for *ALDH1A*-like genes for several vertebrate genomes. For those genomes containing both *ALDH1A2* and *ALDH1A3* genes, these are located on the same chromosome in all cases, with the exception of the rat *ALDH1A2* and *ALDH1A3* genes, which are located on chromosomes 8 and 1, respectively (also see Tables 2s and 3s). The consistent colocation of these genes suggested that an unequal crossover event of an ancestral *ALDH1A2* gene resulted in the appearance of these genes. In addition, multiple *ALDH1A1*-like genes were observed on mouse chromosome 19 and rat chromosome 1, and on opossum chromosome 6, for which 4 *ALDH1A*-like genes were observed in closely located positions, which is suggestive of further unequal crossover gene duplication events of the ancestral *ALDH1A1* genes, in each case.

***ALDH1A1*, *ALDH1A2* and *ALDH1A3* gene regulation sites**

Predicted transcription factor binding sites (TFBS) identified in the promoter regions for human *ALDH1A1*, *ALDH1A2* and *ALDH1A3* genes are summarized in Table 2:

***ALDH1A1***: Cellular distribution studies for mammalian ALDH1A1 have reported high levels for this enzyme in liver [22,28,30] and erythrocytes [9]. At least 12 TFBS were observed for the *ALDH1A1* promoter region, including those contributing to a high level of gene expression in liver (*HLF* and *HNF3B*) erythroid cells (*GATA1*) and during neuronal development (LHX3; POU3F2; AREB6) ([17,21]. The promoter region for the human *ALDH1A1* gene has previously been shown to contain an ATA box and a CCAAT box upstream of the transcription initiation site (32 and 74bps, respectively) [2], which also play a role in up-regulating mouse adult and embryonic β-like globin genes. Moreover, the rat *ALDH1A1* promoter has been described, with TATA and CCAAT cis-acting elements identified, as well as Sp1, AP1 and OCT/POU2F1 sites shown to be critical for promoter activity [31]. Promoter polymorphisms for human *ALDH1A1* have been identified which may influence alcohol consumption behavior [32].

***ALDH1A2***: cellular distribution and development studies for mammalian ALDH1A2 have reported high levels of this enzyme in many tissues of the body, particularly the embryonic trunk region, where it appears in the mouse at E7.5-E10.5, and transiently at E8.5-E9.5 in the ventral optic vesicle and surrounding frontonasal region [17-20]. Studies using *ALDH1A2* knock out mice have shown that it is essential for the development of posterior pharyngeal arches and the enteric nervous system [21]. At least 4 TFBS were identified in the human *ALDH1A2* gene promoter region, including the general transcription factor TATA box, as well as binding regions for early growth response proteins 2-4 (Table 2). *HOXA13* also regulates mouse *Aldh1a2* gene expression in association with retinoic acid signaling and limb morphogenesis [33]. Bone morphogenetic protein (BMP) plays a critical role in skeletogenesis, with *ALDH1A2* serving as a major target of BMP signaling [34]. ALDH1A2 mediated retinoic acid synthesis also regulates early mouse embryonic forebrain development by fibroblast growth factor and sonic hedgehog signaling pathways [35].

***ALDH1A3***: cellular distribution and development studies for mammalian ALDH1A3 have reported high levels of this enzyme in the ventral region of the retina [6-8]; the urogenital tract and sex gland epithelia; and the differentiating intestinal lamina propria [21]. Studies using *ALDH1A3* knock out mice reported a

suppression of retinoic acid synthesis, malformations in ocular and nasal regions and choanal atresia, which is responsible for respiratory distress and subsequent death at birth [36]. At least 7 TFBS were identified in the human *ALDH1A3* gene promoter region, including *HOXA3* and *HOXA5,* which assist with the positioning on the anterior-posterior axis during development; and *PPARG*, forming a dimer with the retinoic acid receptor (Table 2).

Table 3 summarizes the locations of other potential gene regulation sites for the human *ALDH1A1*, *ALDH1A2* and *ALDH1A3* genes, including CpG islands and miRNA binding sites, located in the promoter and 3'-UTR regions respectively. Large CpG islands were located in the *ALDH1A2* (CpG146) and *ALDH1A3* (CpG189) gene promoter regions which may play a role in regulating gene expression during development or participate in gene inactivation via hypermethylation, as reported for tumor suppressor genes in cancer cells [37]. In contrast, human *ALDH1A1* lacked a CpG island in its promoter region which may reflect the housekeeping roles proposed for this gene in coordinating hepatic gluconeogenesis and lipid metabolism [38], participating in the metabolism of neural catecholamines [39] and serving as a marker for stem cells in the body [13]. This is in contrast to the specialist roles established for *ALDH1A2* and *ALDH1A3* in directing limb, neuronal, retinal and skeletal development in the body [6,7,17,21].

Multiple miRNA binding sites were observed for the 3'-UTR regions for human *ALDH1A2* and *ALDH1A3* genes, whereas the 3'-UTR region for the *ALDH1A1* gene contained no detectable mi-RNA binding site (Table 3). Mi-RNA binding sites usually play key roles in the transcriptional and translational control of gene expression, principally via mRNA cleavage of target genes or translational repression by base pairing with 3'UTRs of target mRNAs [40]. The mi-R137 binding site identified for the *ALDH1A2* gene have been previously investigated in relation to impacts on nervous system development [41]. It is a recognized candidate for influencing the functional connectivity of dorsolateral prefrontal-hippocampal regions during neural development and has been designated as a 'schizophrenia candidate'.

**Phylogeny and evolution of vertebrate ALDH1A1, ALDH1A2 and ALDH1A3 sequences**

A phylogram (Figure 3) was calculated by the progressive alignment of vertebrate ALDH1A1, ALDH1A2 and ALDH1A3 amino acid sequences, using the primitive vertebrate lamprey (*Petromyzon marinus)* ALDH1A2 sequence [27] to 'root' the tree. ALDH1A-like sequences were identified for all

mammalian, bird, lizard and frog genomes examined, whereas fish genomes displayed a more restricted distribution pattern for *ALDH1A1* and *ALDH1A3* genes. This is predominantly consistent with a previous report on vertebrate *ALDH1A*-like gene evolution [42]. The phylogram demonstrated separation of these sequences into three distinct groups during vertebrate evolution (ALDH1A1, ALDH1A2 and ALDH1A3 sequences) and suggested that these genes have been derived from an ancestral vertebrate *ALDH1A2* gene. Cañestro and coworkers [42] have concluded from their study that the *ALDH1A1* gene has been lost from teleost fish genomes. Evidence is presented here, however, for a restricted distribution of this gene among African river or lake teleost species, including tilapia (or nile perch) (*Oreochromis niloticus*), zebra mbuna (*Maylandia zebra*) and red zuma (*Haplochromis nyererei*) (Figure 3; Table 1s). It would appear that any teleost *ALDH1A1* gene loss has not taken place for all teleost genomes. Moreover, the teleost *ALDH1A3* gene was absent from the medaka (*Oryzias latipes*) and tilapia (*Oreochromis niloticus*) genomes which is consistent with previous studies [42].

Figures 1s, 2s and 3s (supplementary data) show phylograms derived from the progressive alignment of all vertebrate ALDH1A-like, ALDH1A2-like and ALDH1A3-like sequences examined respectively, including those derived from African river and lake teleost fish genomes, and from a cartiliginous species (elephant shark-*Callorhynchus milii*). The tree was 'rooted' using the lamprey ALDH1A2 sequence (also see Table 1s). Multiple *ALDH1A1*-like genes were observed in mouse and rat genomes (designated *ALDH1A1* and *ALDH1A7*), and in marsupial genomes (designated *ALDH1A1*, *ALDH1A8*, *ALDH1A9* and *ALDH1A10*), confirming previous reports [43-45]. The *ALDH1A7* gene has been shown to be phenobarbital-induced in rat liver, as compared with the constitutively expressed *ALDH1A1* gene, and does not function biologically in retinoic acid synthesis during development in the mouse [43]. The rat ALDH1A1 and ALDH1A7 sequences showed 84% identity [29], confirming their ALDH1A1-like status. The mouse and rat *ALDH1A1* and *ALDH1A7* genes were closely localized on chromosomes 9 and 1 respectively, in opposed transcription orientations, which suggests that these genes have arisen via an unequal crossover gene duplication mechanism. In addition, the *ALDH1A1*-like phylogenetic tree (Figure 1s) suggested that this duplication has taken place prior to the separation of the mouse and rat lineages. Similarly, several marsupial *ALDH1A1*-like gene duplication events have apparently occurred prior to the separation of the South American (opossum-*Monodelphis domestica*) and

Australian (Tasmanian devil-*Sarcophilus harrisii*) lineages, which generated the *ALDH1A9-ALDH1A8-ALDH1A1-ALDH1A10* gene cluster on opossum chromosome 6 (Figure 1s, Figure 2, Table 1s).

Figure 4 summarizes a working hypothesis for the evolution of vertebrate *ALDH1A*-like genes:

1. A proposed primordial *ALDH1A2*-like gene was derived from an invertebrate ancestor and was present in an early vertebrate genome (lamprey-*Petromyzon marinus*).

2. Successive *ALDH1A2* gene duplication events, either by polyploidy or unequal gene duplication events, previously reported for teleost genomes [42,46], generated ancestral teleost fish *ALDH1A1*, *ALDH1A2* and *ALDH1A3* genes.

3. Subsequent gene loss events occurred for the *ALDH1A1* gene for most teleost species, with exception of African lake species genomes examined, and for the *ALDH1A3* gene within some fish genomes.

4. *ALDH1A1*, *ALDH1A2* and *ALDH1A3* genes were retained throughout subsequent vertebrate evolution, but with a 4-7 times slower divergence rate observed for the *ALDH1A2* gene.

5. Unequal crossover gene duplication events occurred for the marsupial and rodent ancestral *ALDH1A1* genes, generating multiple *ALDH1A*-like genes, including the *ALDH1A7* gene within mouse and rat genomes, and several *ALDH1A1*-like genes in opossum and Tasmanian devil genomes.

**CONCLUSIONS**

BLAST and BLAT analyses of several vertebrate genome databases were undertaken using amino acid sequences reported for human ALDH1A1, ALDH1A2 and ALDH1A3 for interrogation of vertebrate genome sequences. Predicted amino acid sequences for these vertebrate ALDH1A-like subunits showed a high degree of sequence identities (69-100%). Bioinformatic analyses identified potential gene regulation sites, including transcription factor binding sites (TFBS) and CpG islands within the human *ALDH1A1*, *ALDH1A2* and *ALDH1A3* gene promoter regions, as well as mi-RNA binding sites in the 3'-UTR regions for these genes. Distinct gene regulatory sites were identified in

each case. Phylogenetic analyses supported an hypothesis concerning the molecular evolution of vertebrate *ALDH1A1, ALDH1A2* and *ALDH1A3* genes, being derived from gene duplication events from a single *ALDH1A2* gene observed in the primitive vertebrate (lamprey: *Petromyzon marinus*) genome; followed by gene loss of the *ALDH1A1* gene within most teleost genomes, as well as the *ALDH1A3* gene within some teleost species; and subsequent gene duplication events of ancestral rodent and marsupial *ALDH1A1* genes, generating multiple *ALDH1A1* genes within the rodent and marsupial genomes examined. These studies also suggested that *ALDH1A2* is the likely primordial *ALDH1A*-like gene originating in an invertebrate ancestor.

**ACKNOWLEDGEMENTS**

I thank Dr Laura Cox from the Texas Biomedical Research Institute and Dr Tom Hurley from the Indiana University School of Medicine for their advice.

**REFERENCES**

[1] Marchitti SA, Brocker C, Stagos D, Vasiliou V (2008) Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. Expert Opin. Drug Toxicol. 4:697-720.
[2] Hsu LC, Chang WC, Yoshida A (1989) Genomic structure of the human cytosolic aldehyde dehydrogenase gene. Genomics 5:857-865.
[3] Yoshida A, Hsu LC, Davé V (1992) Retinal oxidation activity and biological role of human cytosolic aldehyde dehydrogenase. Enzyme 46:239-244.
[4] Zhao D, McCaffrey P, Ivins KJ, Neve RL, Hogan P, Chin WW, Drager UC (1996) Molecular identification of a retinoic-acid-synthesizing enzyme, a retinaldehyde-specific dehydrogenase. Eur. J. Biochem. 240:15-22.
[5] Hsu LC, Chang WC, Hiraoka L, Hsieh CL (1994) Molecular cloning, genomic organisation, and chromosomal localization of an additional human aldehyde dehydrogenase gene, ALDH6. Genomics 24:333-341.
[6] Suzuki R, Shintani T, Sakuta H, Kato A, Ohkawara T, Osumi N, Noda M (2000) Identification of RALDH3, a novel retinaldehyde dehydrogenase, expressed in the ventral region of the retina. Mech. Dev. 98:37-50.
[7] Mic FA, Molotkov A, Fan X, Cuenca AE, Deuster G (2000) RALDH3, a retinaldehyde dehydrogenase that generates retinoic acid, is expressed in the ventral retina, otic vescile and olfactory pit during mouse development. Mech. Dev. 97:227-230.
[8] Grün F, Hirose Y, Kawauchi S, Ogura T, Umesono K (2000) Aldehyde dehydrogenase 6, a cytosolic retinaldehyde dehydrogenase prominently expressed in sensory neuroepithelia during development. J. Biol. Chem. 275:41210-41218.
[9] Agarwal DP, Cohn P, Goedde HW, Hempel J (1989) Aldehyde dehydrogenase from human erythrocytes: structural relationship to the liver cytosolic enzyme. Enzyme 42:47-52.
[10] King G, Holmes RS (1998) Human ocular aldehyde dehydrogenase isozymes: distribution and properties as major soluble proteins in cornea and lens. J. Exp. Zool. 282:12-17.
[11] Jester JV, Moller-Pedersen T, Huang J, Sax CM, Kays WT, Cavangh HD, Petroll WM, Piatigorsky J (1999) The cellular basis of corneal transparency: evidence for 'corneal crystallins'. J. Cell Sci. 112:613-622.
[12] Lassen N, Bateman JB, Estey T, Kuszak JR, Nees DW, Piatigorsky J, Deuster G, Day BJ, Huang J, Hines LM, Vasiliou (2007) Multiple and additive functions of ALDH3A1 and ALDH1A1:

cataract phenotype and ocular oxidative damage in Aldh3a1(-/-)/Aldh1a1(-/-) knock-out mice. J. Biol. Chem. 282:25668-25676.
[13] Levi BP, Yilmaz OH, Duester G, Morrison SJ (2009) Aldehyde dehydrogenase 1a1 is dispensable for stem cell function in the mouse hematopoietic and nervous systems. Blood 113:1670-1680.
[14] Xiao T, Shoeb M, Siddiqui MS, Zhang M, Ramana KV, Srivastava SK, Vasiliou V, Ansari NH (2009) Molecular cloning and oxidative modification of human lens ALDH1A1: implication in impaired detoxification of lipid aldehydes. J. Toxicol. Environ. Health Part A 72:577-584.
[15] Morimoto K, Kim SJ, Tanei T, Shimazu K, Tanji Y, Taguchi T, Tamaki Y, Terada N, Noguchi S (2009) Stem cell marker aldehyde dehydrogenase 1-positive breast cancers are characterized by negative estrogen receptor, positive human epidermal growth factor receptor type 2, and high Ki67 expression. Cancer Sci. 100:1062-1068.
[16] Kiefer FW, Orasanu G, Nallamshetty S, Brown JD, Wang H, Luger P, Qi NR, Burant CF, Duester G, Plutzky J. (2012) Retinaldehyde dehydrogenase 1 coordinates hepatic gluconeogenesis and lipid metabolism. Endocrinology 153:3089-3099.
[17] Haselbeck RJ, Hoffmann I, Duester G (1999) Distinct functions for Aldh1 and Raldh2 in the control of ligand production for embryonic retinoid signaling pathways. Devel. Genetics 25:353-364.
[18] Ulven SM, Gundersen TE, Weedon MS, Landaas VO, Sakhi AK, Fromm SH, Geronimo BA, Moskaug JO, Blomhoff R. (2000) Identification of endogenous retinoids, enzymes, binding proteins, and receptors during early postimplantation development in mouse: important role of retinal dehydrogenase type 2 in synthesis of all-trans-retinoic acid. Devel. Genetics 220: 379-391.
[19] Mic FA, Haselbeck RJ, Cuenca AE, Duester G (2002) Novel retinoic acid generating activities in the neural tube and heart identified by conditional rescue of Raldh2 null mutant mice. Development 129:2271-2282.
[20] Mic FA, Molotkov A, Molotkova N, Duester G (2004) Raldh2 expression in optic vesicle generates a retinoic acid signal needed for invagination of retina during optic cup formation. Developmental Dynamics 231:270-277.
[21] Niederreither K, Vermot J, Le Roux I, Schuhbaur B, Chambon P, Dollé P (2003) The regional pattern of retinoic acid synthesis by RALDH2 is essential for the development of posterior pharyngeal arches and the enteric nervous system. Development 130:2525-2534.
[22] Hempel J, von Bahr-Lindström H, Jörnvall H (1984) Aldehyde dehydrogenase from human liver. Primary structure of the cytoplasmic isoenzyme. Eur. J. Biochem. 141:21-35.
[23] Rongnoparut P, Weaver S (1991) Isolation and characterization of a cytosolic aldehyde dehydrogenase-encoding cDNA from mouse liver. Gene 101:261-265.
[24] Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H (2012) ExPASy: SIB bioinformatics portal. Nucleic Acids Res. 40:W597-603.
[25] Karolchik D, Bejerano G, Hinrichs AS, Kuhn RM, Miller W, Rosenbloom KR, Zweig AS, Haussler D, Kent WJ (2009) Comparative genomic analysis using the UCSC genome browser. Methods Mol. Biol. 395:17-34.
[26] Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36:W465-469.
[27] Moore SA, Baker HM, Blythe TJ, Kitson KE, Kitson TM, Baker EN (1998) Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. Structure 6:1541-1551.
[28] Lamb AL, Newcomer ME (1999) The structure of retinal dehydrogenase type II at 2.7Å resolution: implications for retinal specificity. Biochemistry 38:6003-6011.
[29] Kathmann EC, Lipsky JJ (1997) Cloning of a cDNA encoding a constitutively expressed rat liver cytosolic aldehyde dehydrogenase. Biochem. Biophys. Res. Commun. 236:527-531.
[30] Guimond J, Devost D, Brodeur H, Mader S, Bhat PV (2002) Characterization of the rat ALDH1 promoter. A functional CCAAT and octamer motif are critical for basal promoter activity. Biochim. Biophys. Acta 1579:81-91.

[31] Spence JP, Liang T, Eriksson CJ, Taylor RE, Wall TL, Ehlers CL, Carr LG (2003) Evaluation of aldehyde dehydrogenase 1 promoter polymorphisms identified in human populations. Alcohol Clin. Exp. Res. 27:1389-94.
[32] Shou S, Carlson HL, Perez WD, Stadler HS (2013) HOXA13 regulates Aldh1a2 expression in the autopod to facilitate interdigital programmed cell death. Dev. Dyn. 242:687-698.
[33] Hoffman LM, Garcha K, Karamboulas K, Cowan MF, Drysdale LM, Horton WA, Underhill TM (2006) BMP action in skeletogenesis involves attenuation of retinoid signaling. J. Cell Biol. 174:101-113.
[34] Ribes V, Wang Z, Dollé P, Niederreither K (2006) Retinaldehyde dehydrogenase 2 (RALDH2)-mediated retinoic acid synthesis regulates early mouse embryonic forebrain development by controlling FGF and sonic hedgehog signaling. Development 133: 351-361.
[35] Dupé V, Matt N, Garnier JM, Chambon P, Mark M, Ghyselinck NB (2003) A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment. Proc. Natl. Acad. Sci. U S A 100:14036-14041.
[36] Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nature Rev. Genetics 14:204-220.
[37] Kiefer FW, Orasanu G, Nallamshetty S, Brown JD, Wang H, Luger P, Qi NR, Burant CF, Duester G, Plutzky J (2012) Retinaldehyde dehydrogenase 1 coordinates hepatic gluconeogenesis and lipid metabolism. Endocrinology 153:3089-3099.
[38] Anderson DW, Schray RC, Deuster, Schneider JS (2011) Functional significance of aldehyde dehydrogenase ALDH1A1 in the nigrostriatal dopamine system. Brain Res. 1408:81-87.
[39] Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Ce;; 136:215-233.
[40] Strazisar M, Cammaerts S, van der Ven K, Forero DA, Lanaerts AS, Nordin A, Almeida-Souza L, Genovese G, Timmerman V, Liekens A, De Rijk P, Adolfsson R, Callaerts P, Del-Favero J (2014) MiR137 variants identified in psychiatric patients affect synaptogenesis and neuronal transmission gene sets. Mol. Psychiatry doi: 10.1038/mp.2014.53.
[41] Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, Morgan JR, Buxbaum JD, Sachidanandam R, Sims C, Garruss AS, Cook M, Krumlauf R, Wiedemann LM, Sower SA, Decatur WA, Hall JA, Amemiya CT, Saha NR, Buckley KM, Rast JP, Das S, Hirano M, McCurley N, Guo P, Rohner N, Tabin CJ, Piccinelli P, Elgar G, Ruffier M, Aken BL, Searle SM, Muffato M, Pignatelli M, Herrero J, Jones M, Brown CT, Chung-Davidson YW, Nanlohy KG, Libants SV, Yeh CY, McCauley DW, Langeland JA, Pancer Z, Fritzsch B, de Jong PJ, Zhu B, Fulton LL, Theising B, Flicek P, Bronner ME, Warren WC, Clifton SW, Wilson RK, Li W (2013) Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insight into vertebrate evolution. Nature Genetics 45:415-421.
[42] Cañestro C, Catchen JM, Rodríguez-Marí A, Yokoi H, Postlethwait JH (2009) Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. PloS Genet. 5:e1000496.
[43] Hsu LC, Chang WC, Hoffmann I, Duester G (1999) Molecular analysis of two closely related mouse aldehyde dehydrogenase genes: identification of a role for Aldh1, but not Aldh-pb, in the biosynthesis of retinoic acid. Biochem. J. 339:387-395.
[44] Kathmann EC, Naylor S, Lipsky JJ (2000) Rat liver constitutive and phenobarbital-inducible cytosolic aldehyde dehydrogenases are highly homologous proteins that function as distinct isozymes. Biochemistry 39:11170-11176.
[45] Holmes RS (2009) Opossum aldehyde dehydrogenases: evidence for four ALDH1A1-like genes on chromosome 6 and ALDH1A2 and ALDH1A3 genes on chromosome 1. Evidence is presented for six opossum *ALDH1A* genes, including four *ALDH1A1*-like genes on chromosome 6 and *ALDH1A2*- and *ALDH1A3*-like genes on chromosome 1. Biochem. Genet. 47:609-624.
[46] Yang L, Zou M, Fu B, He S (2013) Genome-wide identification, characterization and expression analysis of lineage-specific genes within zebrafish. BMC Genomics 14:65.

**LEGENDS FOR FIGURES**

**Figure 1: Amino acid sequence alignments for human ALDH1A1, ALDH1A2, ALDH1A3, tilapia ALDH1A1, zebra fish ALDH1A2 and ALDH1A3 sequences**
**(colored version)**

See Table 1 for details of ALDH1A1, ALDH1A2 and ALDH1A3 genes and proteins; * shows identical residues for ALDH subunits; : similar alternate residues; . less similar alternate residues; active site residues (based on [27,28]) are shown as Asn; Glu; and Cys; **bold** font shows known or predicted exon junctions; exon numbers refer to the human *ALDH1A1* gene; coenzyme binding domain is in green.

**Figure 1: Amino acid sequence alignments for human ALDH1A1, ALDH1A2, ALDH1A3, tilapia ALDH1A1, zebra fish ALDH1A2 and ALDH1A3 sequences (black and white version)**

See Table 1 for details of ALDH1A1, ALDH1A2 and ALDH1A3 genes and proteins; * shows identical residues for ALDH subunits; : similar alternate residues; . less similar alternate residues; active site residues (based on [27,28]) are shown as Asn; Glu; and Cys; **bold** font shows known or predicted exon junctions; exon numbers refer to the human *ALDH1A1* gene; coenzyme binding domain is shown.

**Figure 2: Comparative *ALDH1A*-like vertebrate gene locations on vertebrate chromosomes (colored and black and white versions)**

See Tables 1, 1s, 2s and 3s for sources of ALDH1A-like gene locations; arrows give transcription directions; chromosomal lengths are not drawn to scale; GL refers to the scaffold identification for lamprey DNA.

**Figure 3: Phylogenetic tree for vertebrate ALDH1A-like sequences (colored and black and white versions)**

The tree is labeled with the gene name and the name of the vertebrate. Note the major clusters for the vertebrate ALDH1A1, ALDH1A2 and ALDH1A3 sequences. The tree is 'rooted' with the lamprey ALDH1A2 sequence. See Tables 1, 1s, 2s and 3s for details of *ALDH1A*-like genes and proteins. Note the absence of ALDH1A1 and ALDH1A3 sequences for some fish species, whereas ALDH1A2 sequences were observed for all vertebrate genomes examined. A genetic distance scale is shown. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates is represented as a fraction out of 100. Only replicate values of 0.9 or more are highly significant, with 100 bootstrap replicates performed in each case.

**Figure 4: A proposal for the evolutionary appearance of the vertebrate *ALDH1A*-like genes (colored and black and white versions)**

Proposed evolutionary appearance of vertebrate *ALDH1A*-like genes and proteins. Note the presence of at least three *ALDH1A*-like genes throughout vertebrate evolution with the exception of teleost fish and the 4-7 times slower rate of sequence divergence for vertebrate ALDH1A2 proteins.

## LEGENDS FOR SUPPLEMENTARY FIGURES

**Figure 1s (supplementary data): Phylogenetic tree for vertebrate ALDH1A1 sequences**

The tree is labeled with the gene name and the name of the vertebrate. Note the two major clusters for the vertebrate ALDH1A1 sequences: African river and lake fish species; and other vertebrates, including mammals, birds, lizard and frog. The tree is 'rooted' with the lamprey ALDH1A2 sequence. See Table 1s for details of *ALDH1A1* genes and proteins. Note the absence of ALDH1A1 for some fish species. A genetic distance scale is shown. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates is represented as a fraction out of 100. Only replicate values of 0.9 or more are highly significant with 100 bootstrap replicates performed in each case.

**Figure 2s (supplementary data): Phylogenetic tree for vertebrate ALDH1A2 sequences**

The tree is labeled with the gene name and the name of the vertebrate. Note two major clusters for the vertebrate ALDH1A2 sequences: teleost fish species; and other vertebrates, including mammals, birds, lizard and frog. The tree is 'rooted' with the lamprey ALDH1A2 sequence. See Table 2s for details of *ALDH1A2* genes and proteins. A genetic distance scale is shown. Note the zero or very small genetic distances observed, particularly for mammalian ALDH1A2 protein sequences. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates is represented as a fraction out of 100. Only replicate values of 0.9 or more are highly significant with 100 bootstrap replicates performed in each case.

**Figure 3s (supplementary data): Phylogenetic tree for vertebrate ALDH1A3 sequences**

The tree is labeled with the gene name and the name of the vertebrate. Note the two major clusters for the vertebrate ALDH1A3 sequences: teleost fish species; and other vertebrates, including mammals, birds, lizard and frog. The tree is 'rooted' with the lamprey ALDH1A2 sequence. See Table 3s for details of *ALDH1A3* genes and proteins. A genetic distance scale is shown. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates is represented as a fraction out of 100. Only replicate values of 0.9 or more are highly significant with 100 bootstrap replicates performed in each case.

## LEGENDS FOR TABLES

**Table 1: Vertebrate aldehyde dehydrogenase (*ALDH1A1, ALDH1A2* and *ALDH1A3*) genes and enzymes**

RefSeq refers to the NCBI reference sequence; [1]predicted NCBI sequence; na-not available; *gene scaffold ID; * refers to Contig ID; bps refers to base pairs of nucleotide sequence.

**Table 2: Human *ALDH1A1*, *ALDH1A2* and *ALDH1A3* gene promoter transcription factor binding sites (TFBS)**

Transcription factor binding sites (TFBS) were identified within human *ALDH1A1*, *ALDH1A2* and *ALDH1A3* gene promoters using the UCSC human genome browser [25]; data and references concerning each TFBS are described using the UNIPROT website (www.uniprot.org).

**Table 3: Human *ALDH1A1*, *ALDH1A2* and *ALDH1A3* CpG islands and miRNA binding sites**

The UCSC human genome browser [25] was used to identify CpG islands within the human *ALDH1A2* and *ALDH1A3* gene promoters; and miRNA binding sites within the human *ALDH1A2* and *ALDH1A3* gene 3'-UTR regions.

## LEGENDS FOR SUPPLEMENTARY TABLES

**Table 1s (supplementary data): Vertebrate aldehyde dehydrogenase *ALDH1A1* genes and enzymes**

RefSeq refers to the NCBI reference sequence; *predicted NCBI sequence; na-not available; ^gene scaffold ID; bps refers to base pairs of nucleotide sequence.

**Table 2s (supplementary data): Vertebrate aldehyde dehydrogenase *ALDH1A2* genes and enzymes**

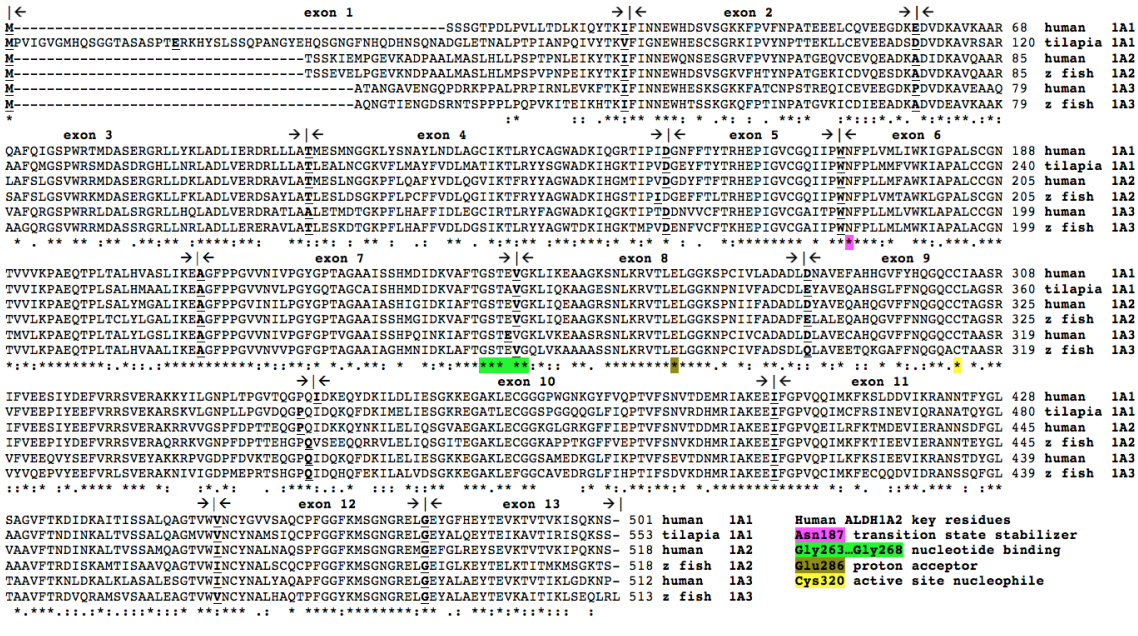
RefSeq refers to the NCBI reference sequence; *predicted NCBI sequence; na-not available; ^gene scaffold ID; bps refers to base pairs of nucleotide sequence.

**Table 3s (supplementary data): Vertebrate aldehyde dehydrogenase *ALDH1A3* genes and enzymes**

RefSeq refers to the NCBI reference sequence; *predicted NCBI sequence; na-not available; ^gene scaffold ID; bps refers to base pairs of nucleotide sequence.

## Figures

## Figure 1 colored




## Figure 1 black and white

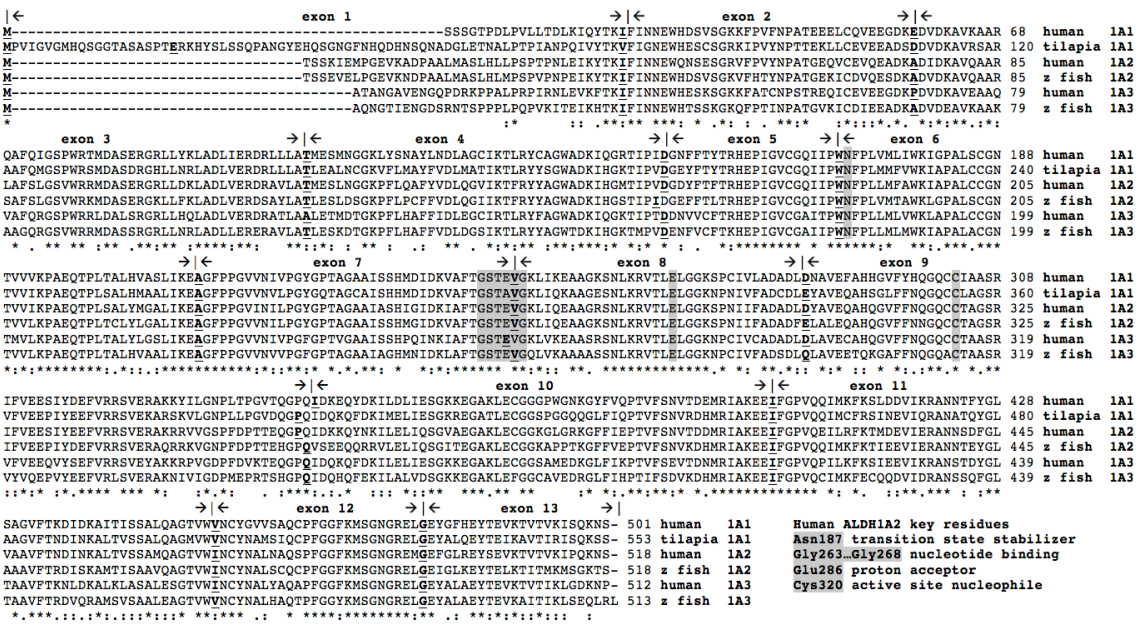


**Figure 2**

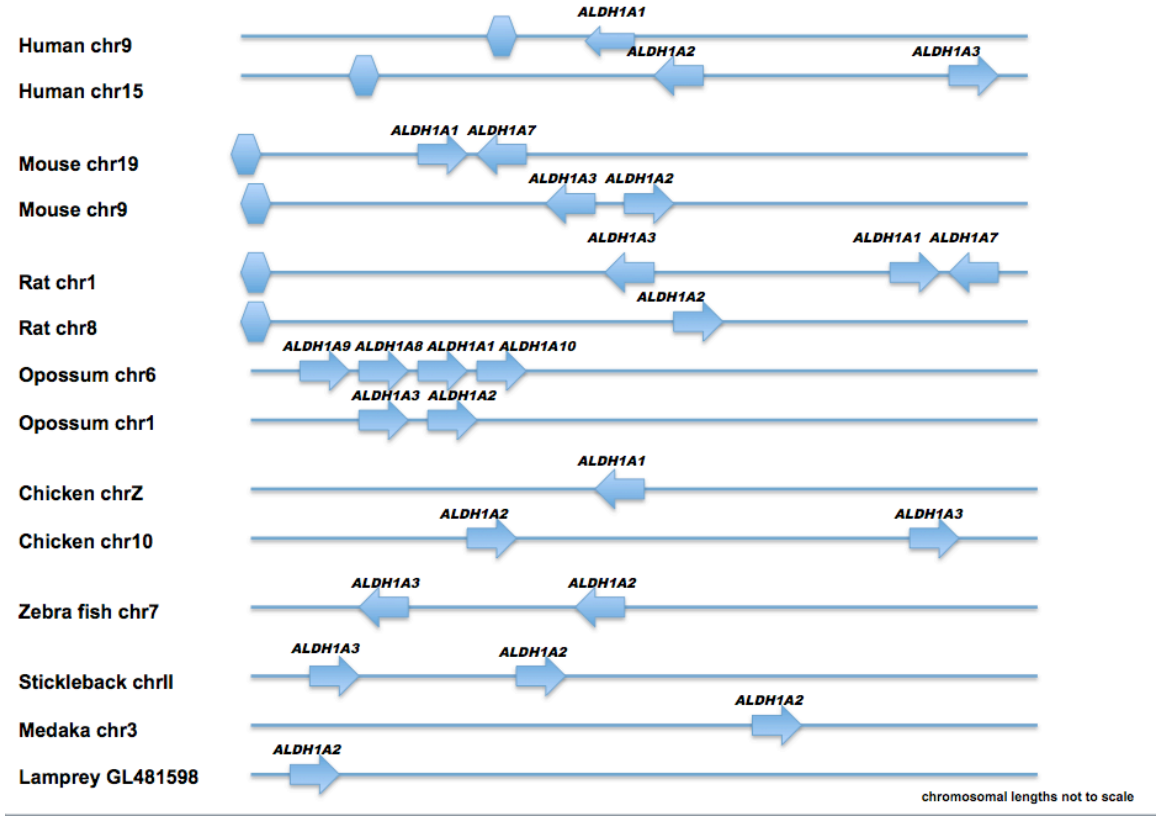

Human chr9
Human chr15
Mouse chr19
Mouse chr9
Rat chr1
Rat chr8
Opossum chr6
Opossum chr1
Chicken chrZ
Chicken chr10
Zebra fish chr7
Stickleback chrII
Medaka chr3
Lamprey GL481598
ALDH1A1
ALDH1A2
ALDH1A3
ALDH1A1 ALDH1A7
ALDH1A3 ALDH1A2
ALDH1A3
ALDH1A1 ALDH1A7
ALDH1A2
ALDH1A9 ALDH1A8 ALDH1A1 ALDH1A10
ALDH1A3 ALDH1A2
ALDH1A1
ALDH1A2
ALDH1A3
ALDH1A3
ALDH1A2
ALDH1A3
ALDH1A2
ALDH1A2
ALDH1A2
chromosomal lengths not to scale

**Figure 3**

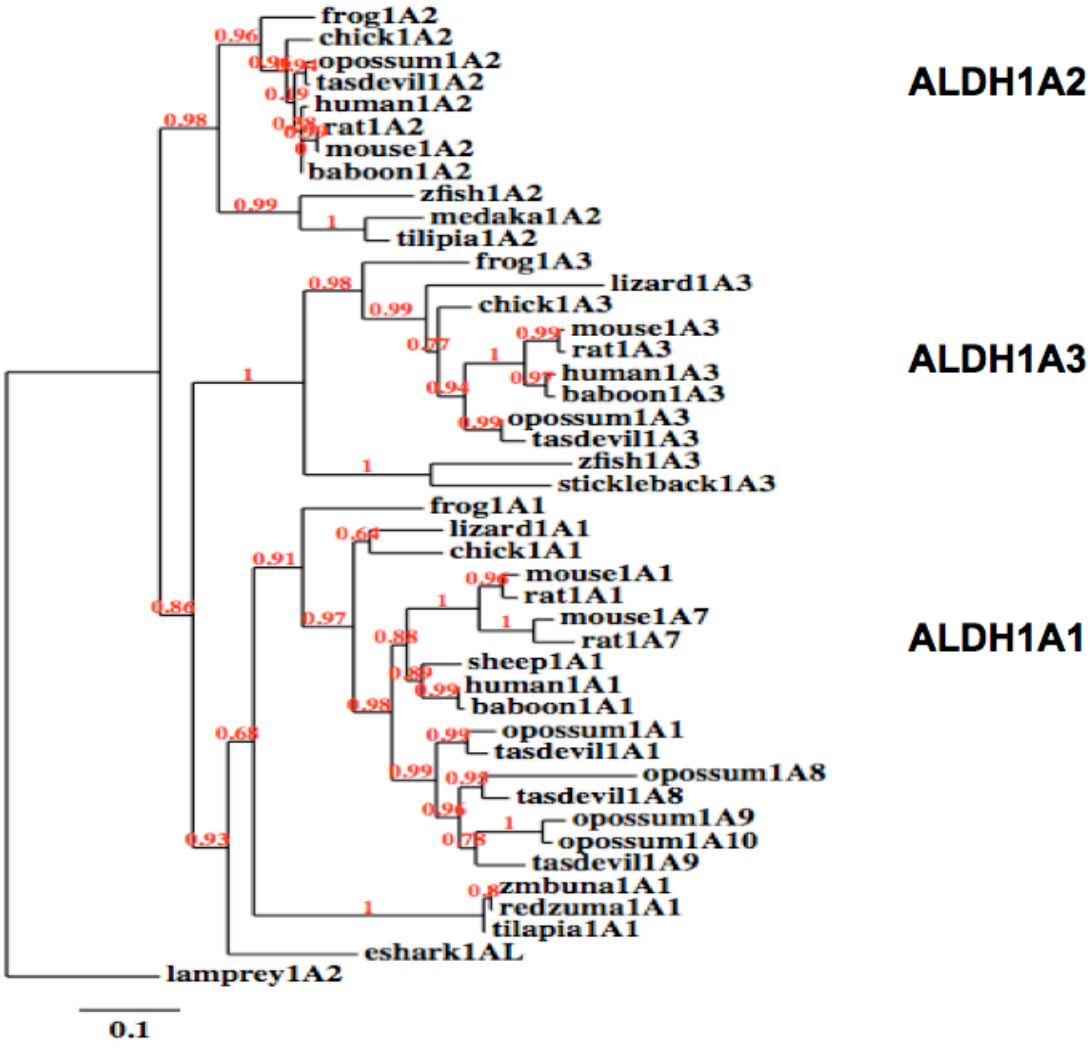
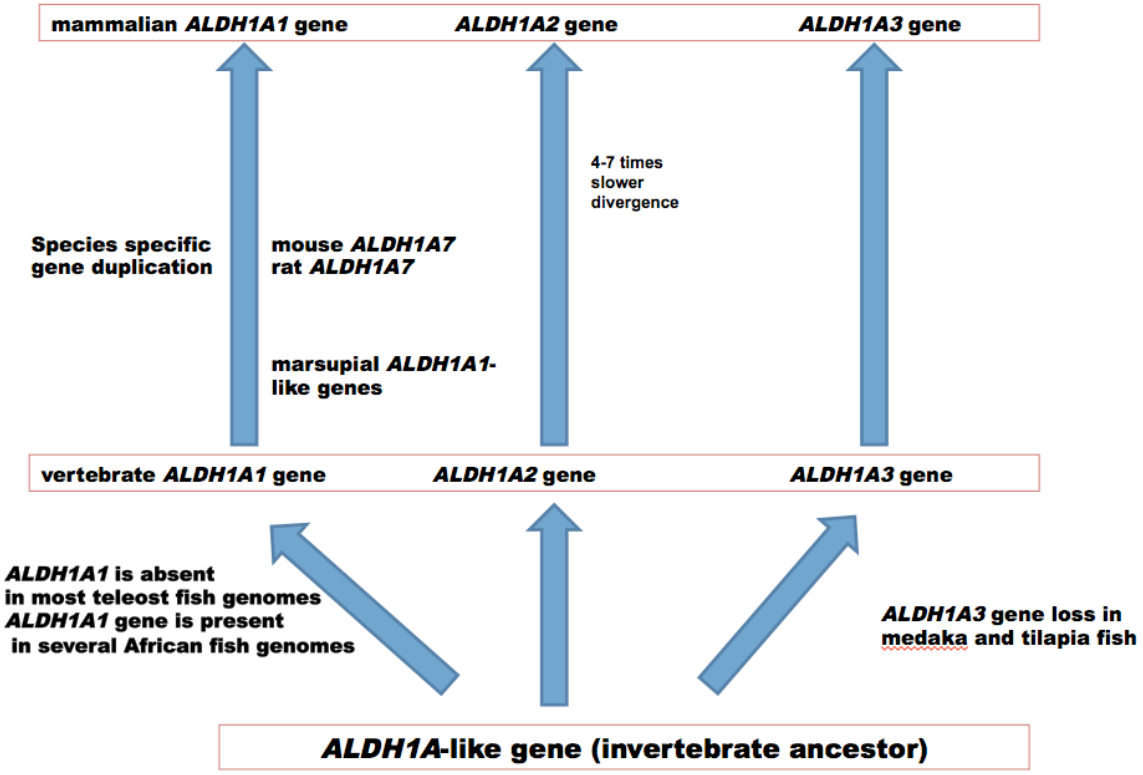

frog1A2
chick1A2
opossum1A2
tasdevil1A2
human1A2
rat1A2
mouse1A2
baboon1A2
zfish1A2
medaka1A2
tilipia1A2
frog1A3
lizard1A3
chick1A3
mouse1A3
rat1A3
human1A3
baboon1A3
opossum1A3
tasdevil1A3
zfish1A3
stickleback1A3
frog1A1
lizard1A1
chick1A1
mouse1A1
rat1A1
mouse1A7
rat1A7
sheep1A1
human1A1
baboon1A1
opossum1A1
tasdevil1A1
opossum1A8
tasdevil1A8
opossum1A9
opossum1A10
tasdevil1A9
zmbuna1A1
redzuma1A1
tilapia1A1
eshark1AL
lamprey1A2
0.1
ALDH1A2
ALDH1A3
ALDH1A1

**Figure 4**


mammalian ALDH1A1 gene
ALDH1A2 gene
ALDH1A3 gene
Species specific gene duplication
mouse ALDH1A7
rat ALDH1A7
marsupial ALDH1A1-like genes
4-7 times slower divergence
vertebrate ALDH1A1 gene
ALDH1A2 gene
ALDH1A3 gene
ALDH1A1 is absent in most teleost fish genomes
ALDH1A1 gene is present in several African fish genomes
ALDH1A3 gene loss in medaka and tilapia fish
ALDH1A-like gene (invertebrate ancestor)

## Table 1

| Vertebrate | Species | Gene | RefSeq ID Prediction[1/2] | GenBank ID | Chromosome location | Coding Exons (strand) | Gene Size (bps) | UNIPROT ID | Amino acids | Subunit MW |
|---|---|---|---|---|---|---|---|---|---|---|
| Human | *Homo Sapiens* | *ALDH1A1* | NM_000689 | AK293176 | 9:74,705,947-74,757,736 | 13 (-ve) | 52,656 | P00352 | 501 | 54,862 |
| Mouse | *Mus musculus* | *Aldh1a1* | NM_013467 | BC054386 | 19:20,601,982-20,643,462 | 13 (+ve) | 41,481 | P24549 | 501 | 54,468 |
| Mouse | *Mus musculus* | *Aldh1a7* | NM_011921 | BC046315 | 19:20,692,953-20,727,556 | 13 (-ve) | 34,604 | O35945 | 501 | 54,588 |
| Opossum | *Monodelphis domestica* | *ALDH1A1* | XP_001373199[1] | na | 6:92,999,463-93,042,087 | 13 (+ve) | 42,625 | na | 508 | 55,506 |
| Chicken | *Gallus gallus* | *ALDH1A1* | NM_204577 | X58869 | Z:35,753,405-35,787,275 | 13 (-ve) | 33,871 | P27463 | 509 | 55,809 |
| Frog | *Xenopus tropicalis* | *ALDH1A1* | NM_001032315 | BC096010 | *GL172794:265,843-296,760 | 13 (+ve) | 30,918 | Q9PWJ3 | 502 | 55,216 |
| Tilapia | *Oreochromis niloticus* | *ALDH1A1* | XP_003445644[1] | na | LG12:10,049,008-10,053,564 | 14 (+ve) | 4,557 | I3JXA1 | 553 | 60,452 |
| Human | *Homo Sapiens* | *ALDH1A2* | NM_003888 | BC030589 | 15:58,247,398-58,357,848 | 13 (-ve) | 110,451 | O94788 | 518 | 56,724 |
| Mouse | *Mus musculus* | *Aldh1a2* | NM_009022 | BC075704 | 9:71,063,654-71,141,482 | 13 (+ve) | 77,829 | Q62148 | 518 | 56,626 |
| Opossum | *Monodelphis domestica* | *ALDH1A2* | XP_001368154[1] | na | 1:163,077,208-163,186,017 | 13 (+ve) | 108,810 | F6PGG8 | 518 | 56,663 |
| Chicken | *Gallus gallus* | *ALDH1A2* | NM_204995 | AF181680 | 10:6,607,160-6,659,397 | 13 (+ve) | 52,238 | O93344 | 518 | 56,732 |
| Frog | *Xenopus tropicalis* | *ALDH1A2* | NM_001045731 | BC157513 | *GL172944:1,126,660-1,157,292 | 13 (+ve) | 30,633 | F6QDK2 | 518 | 56,743 |
| Zebra fish | *Danio rerio* | *ALDH1A2* | NM_131850 | AF315691 | 7:31,664,192-31,703,960 | 13 (-ve) | 39,769 | Q90Y03 | 518 | 56,580 |
| Human | *Homo Sapiens* | *ALDH1A3* | NM_000693 | BC069274 | 15:99,237,636-99,272,498 | 13 (+ve) | 34,863 | P47895 | 512 | 56,108 |
| Mouse | *Mus musculus* | *Aldh1a3* | NM_053080 | BC058277 | 7:66,392,692-66,427,412 | 13 (-ve) | 34,721 | Q9JHW9 | 512 | 56,157 |
| Opossum | *Monodelphis domestica* | *ALDH1A3* | XP_003340034[1] | na | 1:139,416,365-139,472,530 | 13 (+ve) | 56,166 | F7DI08 | 512 | 56,663 |
| Chicken | *Gallus gallus* | *ALDH1A3* | NM_204669 | AF152358 | 10:16,932,409-16,964,722 | 13 (+ve) | 32,214 | Q9DD46 | 512 | 56,312 |
| Frog | *Xenopus tropicalis* | *ALDH1A3* | XP_002939310[1] | na | *GL173099:391,077-414,061 | 13 (-ve) | 22,985 | F7BV06 | 512 | 56,278 |
| Zebra fish | *Danio rerio* | *ALDH1A3* | NM_001044745 | DQ105978 | 7:10,543,820-10,624,872 | 13 (-ve) | 81,053 | Q0H2G3 | 513 | 55,880 |

## Table 2

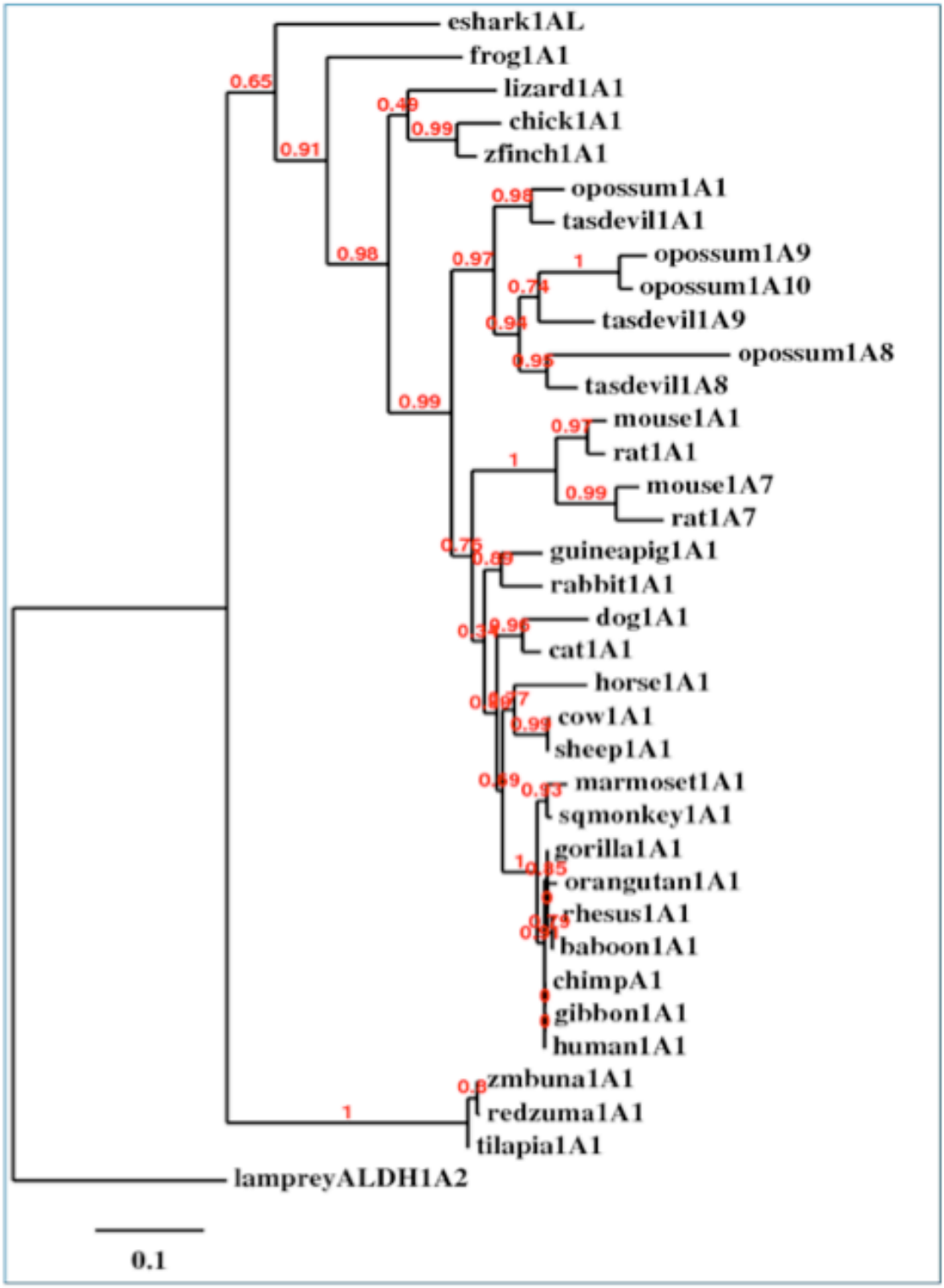
| Human *ALDH1A* Gene Chromosomal Location | 5' TFBS | 5' TFBS Chromosomal Location | TFBS Role | UNIPROT ID |
|---|---|---|---|---|
| ***ALDH1A1*** 9:75,515,578-75,568,233 | ***CREBP1*** | 9:75,569,684-91 | Cyclic AMP-dependent transcription factor ATF-2 regulates cell growth genes | P15336 |
| | ***HLF*** | 9:75,569,683-92 | Hepatic leukemia factor: highly expressed in liver and functioning as a gene promoter | Q16534 |
| | ***E4BP4*** | 9:75,569,682-93 | Nuclear factor interleukin-3-regulated protein acts as a transcriptional regulator | Q16649 |
| | ***HNF3B*** | 9:75,569,624-38 | Hepatocyte nuclear factor 3-beta: involved in embryonic development and gene regulation | Q9Y61 |
| | ***EVI1*** | 9:75,568,896-904 | Oncogene playing a role in development, cell proliferation and differentiation | Q03112 |
| | ***LHX3*** | 9:75,568,192-201 | LIM/homeobox protein LHX3: involved in motor neuron development | Q9UBR4 |
| | ***POU3F2*** | 9:75,568,189-202 | POU domain, class 3, transcription factor 2: neuroectodermal cell lineage expression | P20265 |
| | ***POU2F1*** | 9:75,568,017-29 | POU domain, class 2, transcription factor 1: gene promoter | P14859 |
| | ***EVI1*** | 9:75,568,032-46 | Oncogene playing a role in development, cell proliferation and differentiation | Q03112 |
| | ***OCT/POU2F1*** | 9:75,568,017-29 | POU domain, class 2, transcription factor 1: gene promoter | P14859 |
| | ***GATA1*** | 9:75,567,996-8,005 | Erythroid transcription factor: may serve as a general switch for erythroid development | P15976 |
| | ***AREB6*** | 9:75,567,936-48 | Zinc finger E-box-binding homeobox 1: positively regulates neuronal differentiaition | P37275 |
| ***ALDH1A2*** 15:58,245,611-58,306,295 | ***TATA*** | 15:58,357,933-47 | TATA-box-binding protein: general transcription factor | P20226 |
| | ***EGR4*** | 15:58,357,756-67 | Early growth response protein 4: transcriptional regulator required for differentiation | Q05215 |
| | ***EGR3*** | 15:58,357,756-67 | Early growth response protein 3: probable transcription factor during growth and development | Q06889 |
| | ***EGR2*** | 15:58,357,756-67 | Early growth response protein 2: probable transcription factor during growth and development | P11161 |
| ***ALDH1A3*** *15:101,420,009-101,456,830* | ***PPARG*** | 15:101,420,191-213 | Peroxisome proliferator-activated receptor gamma: heterodimer with retinoic acid receptor | P37231 |
| | ***GATA2*** | 15:101,420,178-87 | Endothelial transcription factor GATA2: regulates endothelin-1 gene expression in endothelial cells | P23769 |
| | ***HOX13/HOXA5*** | 15:101,420,111-40 | Homeobox protein HOX-A5: transcription factor assisting with positioning on the anterior-posterior axis | P20719 |
| | ***PAX4*** | 15:101,419,984-20,004 | Paired box protein PAX4: transcriptional regulator in pancreatic islet beta cells during development | O43316 |
| | ***HOXA3*** | 15:101,419,965-73 | Homeobox protein HOX-A3: transcription factor assisting with positioning on the anterior-posterior axis | O43365 |
| | ***S8/NF-E2*** | 15:101,418,463-73 | Transcription factor NF-E2: component of NF-E2 complex regulating erythroid differentiation | Q16621 |
| | ***PPARG*** | 15:101,419,738-54 | Peroxisome proliferator-activated receptor gamma: heterodimer with retinoic acid receptor | P37216 |

**Table 3**

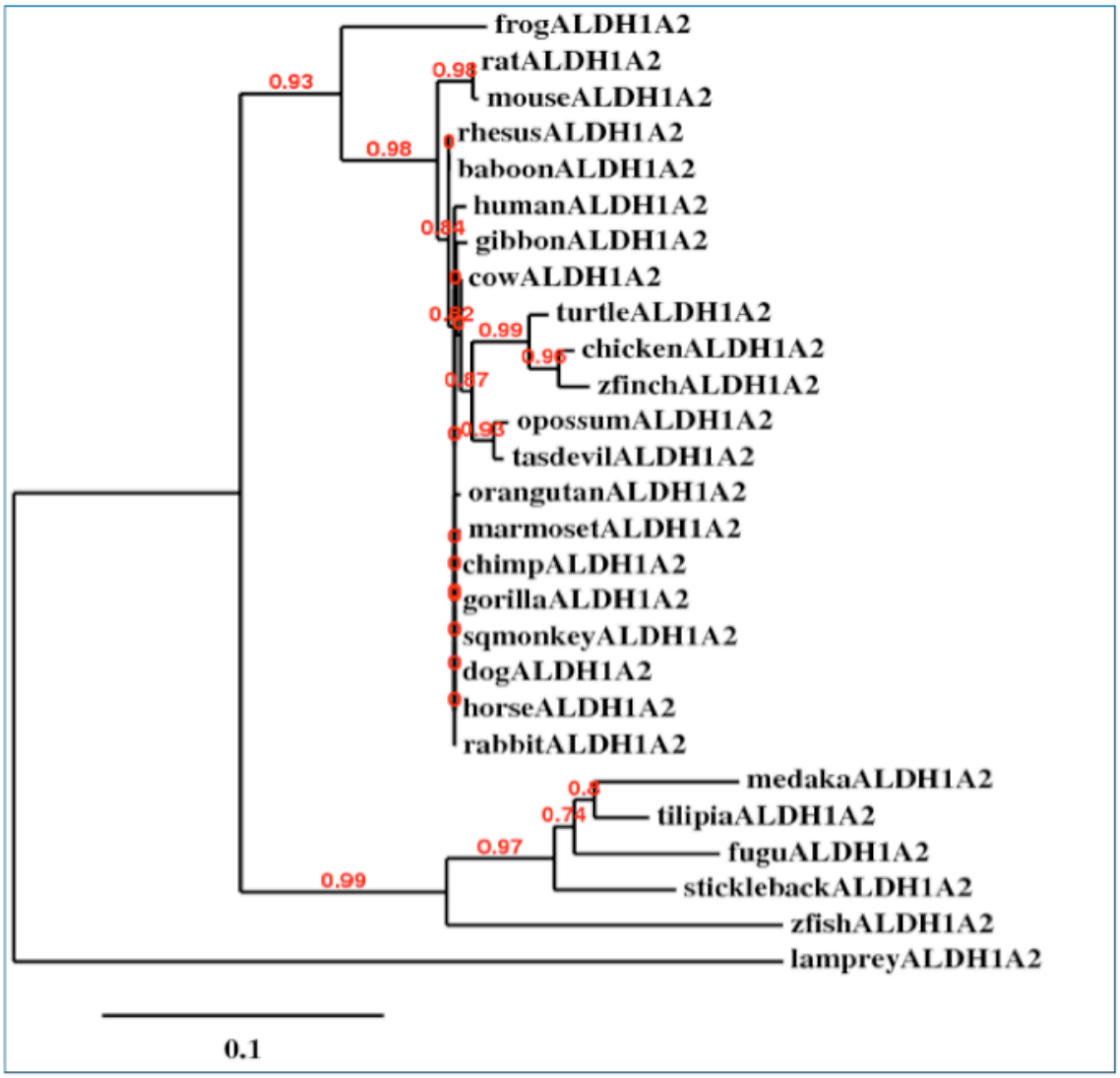
| Human *ALDH1A* Gene | *ALDH1A* Gene Chromosomal Location | 5' CpG Island (Size) | 5' CpG Island Chromosomal Location | miRNA Binding Site | miRNA Binding Site Chromosomal Location |
|---|---|---|---|---|---|
| | | | | | |
| ***ALDH1A1*** | 9:75,515,578-75,568,233 | na | na | na | na |
| ***ALDH1A2*** | 15:58,245,611-58,306,295 | **CpG146** (1,467) | 15:58,357,310-8,776 | **miR-137** | 15:58,245,644-5,651 |
| | | | | **miR-23ab** | 15:58,245,700-707 |
| | | | | **miR-340/340-5p** | 15:58,245,904-910 |
| ***ALDH1A3*** | *15:101,420,009-101,456,830* | **CpG189** (1,872) | 15:101,419,262-21,133 | **miR-125/351** | 15:101,455,261-67 |
| | | | | **miR-340/340-5p** | 15:101,456,591-97 |
| | | | | **miR-144** | 15:101,456,666-73 |
| | | | | **miR-101** | 15:101,456,667-73 |

**Supplementary Figures**

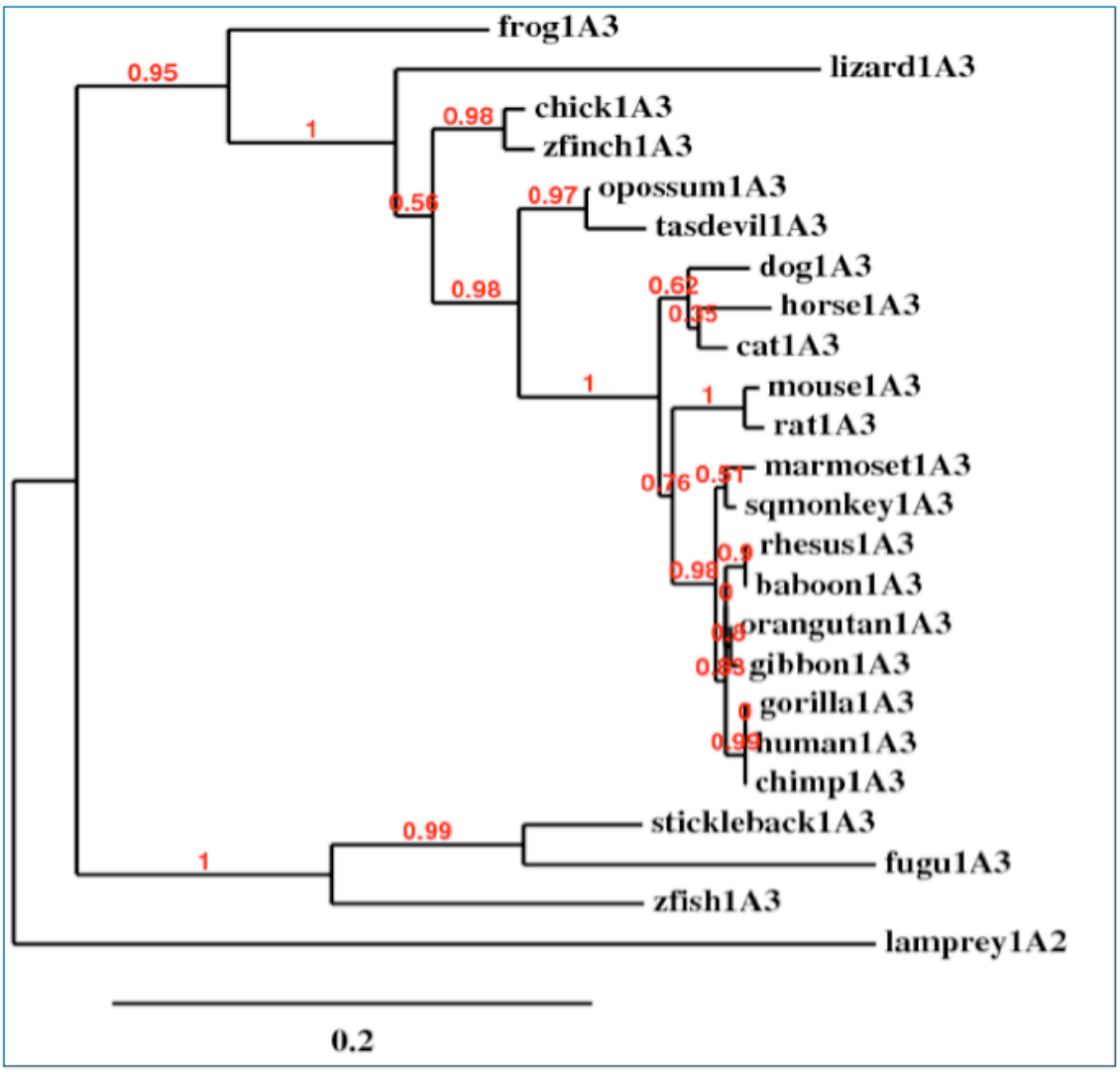
**Supplementary Figure 1s**



**Supplementary Figure 2s**



**Supplementary Figure 3s**



## Supplementary Table 1s

| Animal | Species | *ALDH1A1* Like Gene | RefSeq ID *Prediction | GenBank ID | Chromosome location (coding exons) | Coding Exons (strand) | UNIPROT ID | Amino acids | % Identity Human 1A1 |
|---|---|---|---|---|---|---|---|---|---|
| Human | *Homo sapiens* | ***1A1*** | NM_000689 | AK293176 | 9:75,515,578-75,568,233 | 13 (-ve) | P00352 | 501 | 100.0 |
| Chimpanzee | *Pan troglodytes* | ***1A1*** | NM_001246547 | AK306936 | 9:71,454,106-71,506,191 | 13 (-ve) | G2HJM2 | 501 | 100.0 |
| Gorilla | *Gorilla gorilla* | ***1A1*** | *XP_004048178 | na | 9:54,474,056-54,527,666 | 13 (-ve) | na | 501 | 99.8 |
| Orangutan | *Pongo pygmaeus abelii* | ***1A1*** | NM_001134137 | CR860845 | 9:67,652,586-67,705,567 | 13 (-ve) | Q5R5L2 | 501 | 99.0 |
| Gibbon | *Nomascus leucogenys* | ***1A1*** | *XP_003267459 | na | 1a:62,182,640-62,234,258 | 13 (+ve) | G1QKG9 | 501 | 99.8 |
| Rhesus | *Macaca mulatta* | ***1A1*** | *XP_001097604 | EHH24175 | 15:82,577,303-82,635,649 | 13 (-ve) | F6YY71 | 501 | 98.8 |
| Baboon | *Papio anubis* | ***1A1*** | *XP_003911858 | na | 15:79,690,145-79,748,760 | 13 (-ve) | na | 501 | 99.2 |
| Squirrel monkey | *Saimiri boliviensis boliviensis* | ***1A1*** | *XP_00395663 | na | ^JH378181:1,427,603-1,477,107 | 13 (+ve) | na | 501 | 97.8 |
| Marmoset | *Callithrix jacchus* | ***1A1*** | *XP_002742893 | JAB02181 | 1:96,110,361-96,170,206 | 13 (+ve) | F7IE41 | 501 | 96.2 |
| Mouse | *Mus musculus* | ***1A1*** | NM_013467 | BC044729 | 19:20,676,502-20,717,453 | 13 (+ve) | P24569 | 501 | 87.0 |
| | | ***1A7*** | NM_011921 | BC032880 | 19:20,693,466-20,727,512 | 13 (-ve) | O35945 | 501 | 84.4 |
| Rat | *Rattus norvegicus* | ***1A1*** | NM_022407 | BC061526 | 1:223,731,703-223,772,728 | 13 (+ve) | P51647 | 501 | 86.4 |
| | | ***1A7*** | NM_017272 | M23995 | 1:223,822,536-223,875,238 | 13 (-ve) | P13601 | 501 | 84.2 |
| Guinea pig | *Cavia porcellus* | ***1A1*** | *XP_003472266 | na | ^sc21:24,314,670-24,360,663 | 13 (-ve) | na | 501 | 91.0 |
| Bovine | *Bos taurus* | ***1A1*** | NM_174239 | BC105193 | 8:51,199,516-51,252,053 | 13 (-ve) | P48644 | 501 | 91.8 |
| Horse | *Equus caballus* | ***1A1*** | *XP_001489398 | na | 23:18,267,750-18,307,245 | 13 (+ve) | P15437 | 501 | 90.6 |
| Sheep | *Ovis aries* | ***1A1*** | NM_001009778 | U12761 | 2:67,700,675-67,747,206 | 13 (+ve) | P51977 | 501 | 91.6 |
| Dog | *Canis familiaris* | ***1A1*** | NM_001286977 | na | 1:84,940,455-84,994,244 | 13 (+ve) | J9NS92 | 501 | 89.0 |
| Rabbit | *Oryctolagus cuniculus* | ***1A1*** | NM_001082013 | AY038801 | 1:59,419,969-59,477,825 | 13 (-ve) | Q8MI17 | 496 | 91.5 |
| Cat | *Felis catus* | ***1A1*** | *XP_003995456 | na | D4:14,991,454-150,497,711 | 13 (-ve) | M3WAV4 | 501 | 92.8 |
| Microbat | *Myotis lucifugus* | ***1A1*** | *XP_006086871 | na | ^GL429798:2,625,399-2,671,968 | 13 (-ve) | G1P8S4 | 501 | 93.8 |
| Squirrel | *Spermophilus tridecemlineatus* | ***1A1*** | *XP_005324163 | na | ^JH393313:8,886,595-8,918,323 | 13 (-ve) | I3M206 | 501 | 92.4 |
| Opossum | *Monodelphis domestica* | ***1A1*** | *XP_001373199 | na | 6:92,999,463-93,042,087 | 13 (+ve) | na | 508 | 83.4 |
| | | ***1A8*** | *XP_001373180 | na | 6:92,858,372-92,918,445 | 14 (+ve) | na | 521 | 78.4 |
| | | ***1A9*** | *XP_001373128 | na | 6:92,781,115-92,828,272 | 13 (+ve) | na | 489 | 86.3 |
| | | ***1A10*** | *XP_001373217 | na | 6:93,064,310-93,110,122 | 13 (+ve) | na | 508 | 83.8 |
| Tasmanian devil | *Sarcophilus harrisii* | ***1A1*** | *XP_003761680 | na | ^GL841510:944,679-1,004,465 | 13 (+ve) | na | 510 | 85.8 |
| | | ***1A8*** | *XP_003761681 | na | ^GL841510:1,019,525-1,058,360 | 13 (+ve) | na | 507 | 84.8 |
| | | ***1A9*** | *XP_003761682 | na | ^GL841510:1,110,807-1,135,923 | 13 (+ve) | na | 508 | 85.4 |
| Chicken | *Gallus gallus* | ***1A1*** | NM_204577 | X58869 | Z:35,753,405-35,787,275 | 13 (-ve) | P27463 | 509 | 83.4 |
| Zebra finch | *Taeniopygia guttata* | ***1A1*** | *XP_002191253 | na | Z:16,187,146-16,218,424 | 13 (+ve) | H0YRB4 | 509 | 83.8 |
| Lizard | *Anolis caroliensis* | ***1A1*** | *XP_003216527 | na | 2:47,938,559-47,973,618 | 13 (+ve) | G1KMC0 | 511 | 83.2 |
| Painted turtle | *Chrysemys picta bellii* | ***1A1*** | *XP_005297807 | na | ^JH584686:7,804,796-7,850,048 | 13 (+ve) | na | 511 | 80.6 |
| Clawed frog | *Xenopus tropicalis* | ***1A1*** | NM_001032315 | BC096010 | ^GL172794:265,843-296760 | 13 (+ve) | Q4VBE1 | 502 | 77.6 |
| Nile tilapia | *Oreochromis niloticus* | ***1A1*** | *XP_003445644 | na | ^LG12:10,049,008-10,053,564 | 14 (+ve) | na | 553 | 71.9 |
| Zebra mbuna | *Maylandia zebra* | ***1A1*** | *XP_004544286 | na | na | na | na | 553 | 71.5 |
| Red zuma | *Haplochromis nyererei* | ***1A1*** | *XP_005729529 | na | na | na | na | 553 | 71.5 |
| Coelacanth | *Latimeria chalumnae* | ***1A1*** | *XP_006001724 | na | ^JH127281:26,249-83,125 | 13 (+ve) | na | 514 | 75.9 |

## Supplementary Table 2s

| Animal | Species | *ALDH1A2* Like Gene | RefSeq ID *Prediction | GenBank ID | Chromosome location (coding exons) | Coding Exons (strand) | UNIPROT ID | Amino acids | % Identity Human 1A2 |
|---|---|---|---|---|---|---|---|---|---|
| Human | *Homo sapiens* | *1A2* | NM_003888 | BC030589 | 15:58,247,398-58,357,848 | 13 (-ve) | O94788 | 518 | 100 |
| Chimp | *Pan troglodytes* | *1A2* | *XP_001172159 | na | 15:55,297,531-55,408,118 | 13 (-ve) | H2Q9I2 | 518 | 99.6 |
| Gorilla | *Gorilla gorilla* | *1A2* | *XP_004056304 | na | 15:37,075,785-37,192,713 | 13 (-ve) | G3QXF4 | 518 | 99.6 |
| Orangutan | *Pongo pygmaeus abelii* | *1A2* | NM_001134137 | CR860845 | 15:54,823,180-54,939,280 | 13 (-ve) | H2NNC7 | 518 | 99.4 |
| Gibbon | *Nomascus leucogenys* | *1A2* | *XP_003267079 | na | 6:77,050612-77,184,830 | 13 (-ve) | na | 518 | 99.2 |
| Rhesus | *Macaca mulatta* | *1A2* | *XP_001090504 | JV635656 | 7:35,112,260-35,226,090 | 13 (-ve) | I2CVL1 | 518 | 99.4 |
| Squirrel monkey | *Saimiri boliviensis boliviensis* | *1A2* | *XP_003929025 | na | ^JH378136:16,995,995-17,110,411 | 13 (-ve) | na | 518 | 99.4 |
| Marmoset | *Callithrix jacchus* | *1A2* | *XP_003929025 | JAB07275 | 10:1107,556-215,404 | 13 (-ve) | F7H8D6 | 518 | 99.4 |
| Mouse | *Mus musculus* | *1A2* | NM_009022 | BC075704 | 9:71,063,654-71,141,482 | 13 (+ve) | Q62148 | 518 | 97.3 |
| Rat | *Rattus norvegicus* | *1A2* | NM_053896 | BC098910 | 8:75,692,108-75,768,930 | 13 (+ve) | Q63639 | 518 | 97.5 |
| Hamster | *Mesocricetus auratus* | *1A2* | *XP_005075520 | na | na | na | na | 518 | 97.5 |
| Naked Mole rat | *Heterocephalus glaber* | *1A2* | *XP_004855794 | na | ^JH602094:13,756,483-13,843,533 | 13 (+ve) | G3X6U1 | 518 | 97.3 |
| Bovine | *Bos taurus* | *1A2* | *XP_002690901 | na | na | 13 (-ve) | G3X6U1 | 518 | 99.2 |
| Horse | *Equus caballus* | *1A2* | *XP_001500420 | na | 1:133,332,590-133,421,757 | 13 (+ve) | F7CVL4 | 518 | 100 |
| Sheep | *Ovis aries* | *1A2* | *XP_004010617 | na | 7:49,204,039-49,320,319 | 13 (+ve) | na | 518 | 99.0 |
| Dog | *Canis familiaris* | *1A2* | NM_001287089 | na | 30:23,076,195-23,170,341 | 13 (-ve) | F1PGT3 | 518 | 100 |
| Rabbit | *Oryctolagus cuniculus* | *1A2* | *XP_002717750 | na | 17:14,307,940-14,432,553 | 13 (+ve) | G1TCP3 | 518 | 100 |
| Pika | *Ochotona princeps* | *1A2* | *XP_004578103 | na | na | na | na | 518 | 99.8 |
| Tenrec | *Echinops telfairi* | *1A2* | *XP_004709506 | na | ^JH980330:16,131,533-16,243,924 | 13 (-ve) | na | 518 | 98.1 |
| Squirrel | *Spermophilus tridecemlineatus* | *1A2* | *XP_005316673 | na | ^JH393281:19,177,744-19,262,919 | 13 (-ve) | na | 518 | 99.0 |
| Panda | *Ailuropoda melanoleuca* | *1A2* | *XP_002920715 | na | ^GL192773:196,483-289,428 | 13 (+ve) | G1LVH9 | 518 | 99.4 |
| Rhinoceros | *Ceratotherium simum simum* | *1A2* | *XP_004421679 | na | ^JH767728:38,290,721-38,382,169 | 13 (-ve) | na | 518 | 99.6 |
| Ferret | *Mustela putorius furo* | *1A2* | *XP_004777346 | JP005442 | ^GL897139:1,244,668-1,335,844 | 13 (+ve) | M3XZ73 | 518 | 99.2 |
| Shrew | *Tupaia chinensis* | *1A2* | *XP_006141843 | na | na | na | na | 518 | 99.4 |
| Alpaca | *Vicugna pacos* | *1A2* | *XP_006211033 | na | ^KB632553:3,396,561-3,494,485 | 13 (-ve) | na | 518 | 99.0 |
| Orca | *Orcinus orca* | *1A2* | *XP_004274720 | na | na | na | na | 518 | 98.8 |
| Mole | *Chrsochloris asiatica* | *1A2* | *XP_006831667 | na | na | na | na | 518 | 99.2 |
| Manatee | *Trichechus manatus latirostris* | *1A2* | *XP_004374664 | na | ^JH594636:3,662,704-3,757,042 | 13 (-ve) | na | 518 | 98.1 |
| Megabat | *Pteropus vampyrus* | *1A2* | *XP_006920386 | na | 116:321,334-406,795 | 13 (-ve) | na | 518 | 98.8 |
| Opossum | *Monodelphis domestica* | *1A2* | *XP_001368154 | na | 1:163,077,208-163,186,017 | 13 (+ve) | F6PGG8 | 518 | 97.5 |
| Tasmanian devil | *Sarcophilus harrisii* | *1A2* | *XP_003755761 | na | ^GL834490:990,311-1,101,789 | 13 (+ve) | G3WFU4 | 518 | 97.5 |
| Chicken | *Gallus gallus* | *1A2* | NM_204995 | AF181680 | 10:6,607,160-6,659,397 | 13 (+ve) | O03354 | 518 | 96.0 |
| Zebra finch | *Taeniopygia guttata* | *1A2* | NP_001070153 | na | 10:6,786,462-6,838,408 | 13 (+ve) | na | 518 | 94.6 |
| Budgerigar | *Melopsittacus undulatus* | *1A2* | *XP_005145967 | na | ^JH556567:3,463,590-3,512,251 | 13 (-ve) | na | 518 | 96.0 |
| Pigeon | *Columbia livia* | *1A2* | *XP_005511200 | na | na | na | na | 518 | 96.1 |
| Goose | *Anser anser* | *1A2* | na | HQ260987 | na | na | na | 518 | 95.6 |
| Falcon | *Falco peregrinus* | *1A2* | *XP_005230668 | na | na | na | na | 518 | 96.0 |
| Fly catcher | *Ficedula albicollis* | *1A2* | *XP_005052056 | na | na | na | U3KCL4 | 518 | 95.8 |
| Painted turtle | *Chrysemys picta* | *1A2* | *XP_005309446 | na | JH584895:651,699-730,152 | 13 (+ve) | na | 518 | 96.5 |
| Alligator | *Alligator mississippiiensis* | *1A2* | *XP_006264050 | na | JH733284:440,063-517,529 | 13 (+ve) | na | 518 | 95.2 |
| Toad | *Xenopus tropicalis* | *1A2* | NM_001045731 | CR942326 | GL172944:1,126,660-1,157,292 | 13 (+ve) | F6QDK2 | 518 | 91.1 |
| Clawed Toad | *Xenopus laevis* | *1A2* | NM_001090775 | BC128678 | na | na | Q9DEX5 | 518 | 90.3 |
| Zebra fish | *Danio rerio* | *1A2* | NM_131850 | AF339837 | 7:31,664,192-31,703,960 | 13 (-ve) | B3DKM0 | 518 | 79.3 |
| Medaka | *Oryzias latipes* | *1A2* | NM_001104821 | DQ897366 | 3:20,047,026-20,069,315 | 13 (+ve) | Q09HW0 | 518 | 80.9 |
| Nile tilapia | *Oreochromis niloticus* | *1A2* | *XP_003442313 | na | ^LG1:12,722,665-12,750,058 | 13 (+ve) | I3JEN0 | 518 | 82.1 |
| Pufferfish | *Takifugu rubripes* | *1A2* | NM_001033639 | AB197675 | 13:15,928,438-15,945,015 | 13 (+ve) | na | 502 | 82.3 |
| Eel | *Polypterus senegalus* | *1A2* | na | HQ876171 | na | na | na | 518 | 89.4 |
| Lamprey | *Petromyzon marinus* | *1A2* | na | FJ536260 | ^GL481598:469-3317 | na | na | 523 | 67.0 |

## Supplementary Table 3s

| Animal | Species | *ALDH1A3* Like Gene | RefSeq ID *Prediction | GenBank ID | Chromosome location (coding exons) | Coding Exons (strand) | UNIPROT ID | Amino acids | % Identity Human 1A3 |
|---|---|---|---|---|---|---|---|---|---|
| Human | *Homo sapiens* | *1A3* | NM_000693 | BC069274 | 15:58,99,237,636-99,272,498 | 13 (+ve) | P47895 | 512 | 100 |
| Chimp | *Pan troglodytes* | *1A3* | *XP_001142882 | na | 15:98,414,516-98,449,731 | 13 (+ve) | H2Q9I2 | 512 | 99.8 |
| Gorilla | *Gorilla gorilla* | *1A3* | *XP_004056883 | na | 15:80,999,651-81,037,527 | 13 (+ve) | G3QXF4 | 512 | 99.4 |
| Orangutan | *Pongo pygmaeus abelii* | *1A3* | *XP_002825935 | na | 15:98,093,520-98,129,114 | 13 (+ve) | na | 512 | 98.6 |
| Gibbon | *Nomascus leucogenys* | *1A3* | *XP_004093199 | na | 6:120,108,751-120,137,590 | 13 (+ve) | na | 512 | 98.4 |
| Rhesus | *Macaca mulatta* | *1A3* | NM_001266556 | JV636026 | 7:80,359,733-80,393,939 | 13 (+ve) | F7H3Z0 | 512 | 97.9 |
| Squirrel monkey | *Saimiri boliviensis boliviensis* | *1A3* | *XP_003840611 | na | ^JH378234:2,758,968-2,793,253 | 13 (+ve) | na | 512 | 97.9 |
| Marmoset | *Callithrix jacchus* | *1A3* | *XP_002749112 | na | 6:7,030,295-7,067,269 | 13 (-ve) | F7IH81 | 512 | 97.1 |
| Mouse | *Mus musculus* | *1A3* | NM_053080 | BC038646 | 7:66,392,692-66,427,412 | 13 (-ve) | Q9JHW9 | 512 | 94.0 |
| Rat | *Rattus norvegicus* | *1A3* | NM_1153300 | BC166415 | 1:120,847,749-120,881,883 | 13 (-ve) | Q8K4D8 | 512 | 93.6 |
| Hamster | *Mesocricetus auratus* | *1A3* | *XP_005084482 | na | na | na | na | 512 | 93.4 |
| Squirrel | *Ictidomys tridecemlineatus* | *1A3* | *XP_005320140 | na | ^JH393293:2,932,768-2,967,215 | 13 (-ve) | na | 512 | 95.5 |
| Horse | *Equus caballus* | *1A3* | *XP_001490605 | na | 1:106,035,852-106,071,058 | 13 (+ve) | F6Y0Q6 | 512 | 93.4 |
| Dog | *Canis familiaris* | *1A3* | XP_003639013 | na | 3:40,132,844-40,168,603 | 13 (-ve) | E2R543 | 512 | 93.0 |
| Pika | *Ochotona princeps* | *1A3* | *XP_004593209 | na | 324:292,359-320,976 | 13 (+ve) | na | 512 | 94.1 |
| Tenrec | *Echinops telfairi* | *1A3* | *XP_004717516 | na | ^JH980445:884,717-926,472 | 13 (-ve) | na | 512 | 91.8 |
| Rhinoceros | *Ceratotherium simum simum* | *1A3* | *XP_004431947 | na | ^JH767761:4,566,113-4,601,420 | 13 (-ve) | na | 512 | 94.1 |
| Ferret | *Mustela putorius furo* | *1A3* | *XP_004759897 | na | ^GL896971:5,877,137-5,912,173 | 13 (+ve) | M3XZ73 | 512 | 93.4 |
| Manatee | *Trichechus manatus latirostris* | *1A3* | *XP_004372703 | na | ^JH594625:12,268,931-12,318,233 | 13 (-ve) | na | 512 | 93.6 |
| Opossum | *Monodelphis domestica* | *1A3* | *XP_003340034 | na | 1:139,416,365-139,472,530 | 13 (+ve) | F7DI08 | 512 | 88.7 |
| Tasmanian devil | *Sarcophilus harrisii* | *1A3* | *XP_003755635 | na | ^GL834481:548,516-593,494 | 13 (+ve) | G3W6D5 | 512 | 86.3 |
| Chicken | *Gallus gallus* | *1A3* | NM_204669 | AF152358 | 10:16,932,409-16,964,722 | 13 (+ve) | Q9DD46 | 512 | 86.1 |
| Zebra finch | *Taeniopygia guttata* | *1A3* | *XP_002199949 | na | 10:6,786,462-6,838,408 | 13 (+ve) | H0ZF04 | 512 | 86.1 |
| Fly catcher | *Ficedula albicollis* | *1A3* | *XP_005051790 | na | na | na | U3K6Q3 | 512 | 86.5 |
| Painted turtle | *Chrysemys picta* | *1A3* | *XP_005294883 | na | ^JH584649:994,983-1,040,200 | 13 (-ve) | na | 512 | 86.7 |
| Toad | *Xenopus tropicalis* | *1A3* | *XP_002939310 | na | ^GL173099:391,077-414,061 | 13 (+ve) | F7BV06 | 512 | 80.9 |
| Clawed Toad | *Xenopus laevis* | *1A3* | NM_001095605 | BC169603 | na | na | Q58YE0 | 512 | 79.9 |
| Zebra fish | *Danio rerio* | *1A3* | NM_001044745 | EF375713 | 7:10,543,820-10,624,872 | 13 (-ve) | Q0H2G3 | 513 | 69.3 |
| Stickleback | *Gasterosteus aculeatus* | *1A3* | *ENSGACT00000018580 | na | II:251,902-276,710 | 13 (+ve) | G3PLR3 | 511 | 67.9 |
| Pufferfish | *Takifugu rupides* | *1A3* | NM_001160185 | AB435248 | ^HE592038:6,317-14,928 | 13 (-ve) | na | 511 | 66.3 |