Comparative and Evolutionary Studies of ALDH1A1 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 ALDH1A1 genes encode a bifunctional mitochondrial enzyme, catalyzing a 2-step conversion of glutamate to glutamyl semialdehyde, subsequently converted into proline, ornithine and arginine. Bioinformatic analyses of vertebrate and invertebrate genomes were undertaken using known ALDH1A1 amino acid sequences. G5K (glutamyl kinase) and GPR (glutamyl phosphate reductase) domain sequences were identified for all vertebrate and invertebrate genomes examined, whereas bacterial sequences encoded separate enzymes. Vertebrate ALDH1A1 (also called P5CS) sequences were highly conserved throughout vertebrate evolution. A mechanism for generating two major vertebrate ALDH1A1 isoforms is proposed with ‘a’ isoform containing Asn239-Val240 with wide tissue expression, whereas the ‘b’ isoform lacking the dipeptide has been reported in gut tissues. Phylogenetic analyses describe the relationships and potential origins of the ALDH1A1 gene during vertebrate and invertebrate evolution and a proposal for generating the bifunctional vertebrate and invertebrate ALDH1A1 gene from a bacterial operon (proBA) encoding G5K and GPR. A more recent Aldh18a1 gene duplication event has apparently occurred with a primordial rat genome.

Keywords: Aldehyde dehydrogenases; bifunctional enzyme; ALDH1A1; glutamyl kinase; GK; glutamyl phosphate reductase; GPR; vertebrates; invertebrates; evolution; phylogeny; primordial gene; gene integration; transcription factor binding sites; CpG islands.

Abbreviations: ALDH: aldehyde dehydrogenase; P5CS: delta 1-pyrroline-5-carboxylate synthase; GK: gamma-glutamyl kinase; GPR: gamma-glutamyl phosphate reductase; EC: Enzyme Commission; BLAST: Basic Local Alignment Search Tool; BLAT: Blast-Like Alignment Tool; NCBI: National Center for Biotechnology Information; AceView: NCBI Based representation of public mRNAs; TFBS: transcription factor binding sites; UTR: Untranslated Gene Region; CpG: region of high density of guanine-cytosine dinucleotides; mRNA: messenger RNA.

INTRODUCTION

The aldehyde dehydrogenase 18A1 gene (ALDH1A1) is one of 19 human ALDH genes encoding enzymes with diverse metabolic roles in aldehyde metabolism in the body (http://aldh.org/superfamily.php) [1]. Human ALDH1A1 encodes a bifunctional enzyme, designated as delta 1-pyrroline-5-carboxylate synthase (P5CS), which catalyses the first two steps in proline, ornithine and arginine biosynthesis [2-3]:

Step 1: glutamate 5-kinase (domain G5K) EC=2.7.2.11

\[ \text{ATP + L-glutamate} = \text{ADP + L-glutamate 5-phosphate}^* \text{ (unstable intermediate)} \]

Step 2: gamma-glutamyl phosphate reductase (domain GPR) EC=1.2.1.41

\[ \text{L-glutamate 5-phosphate + NADPH + H}^+ = \text{L-glutamyl-5-semialdehyde + Pi + NADP}^+ \]
Subsequent GSA metabolism results in the formation of proline, via pyrroline-5-carboxylate reductase (P5CR) (EC=1.5.1.12) [4]; ornithine, via ornithine aminotransferase (OAT) (EC=2.6.1.13) [5]; and arginine, via arginase (ARG) (EC=3.5.3.1) [6].

The human ALDH18A1 and mouse Aldh18a1 genes encode two major isoforms which differ by a 2 amino acid deletion within the G5K domain (239Asn-240Val), designated as ALDH18A1long and ALDH18A1short [3,7,8]. These isoforms differ in tissue distribution and kinetic properties, with the long version being widely distributed in most tissues of the body and insensitive to product inhibition, whereas the short isoform is restricted to gut tissues and is inhibited by L-ornithine at physiological concentrations. Several mutations for human ALDH18A1 have been reported which confirm the key role for this gene and protein in the body. These include a rare autosomal recessive cutis laxa type 3A and Marburg Micro syndrome [9,10]; an autosomal dominant form of cutis laxa [11]; and an inherited spastic paraplegia condition [12].

This study describes the predicted sequences, structures and phylogeny of vertebrate and invertebrate ALDH18A1 genes and enzymes and compares these results with those previously reported for the human and mouse genes and proteins [2,3,7,8]. Evidence is presented concerning the generation of ALDH18A1 isoforms from several vertebrate species, and for distinct modes of gene regulation and expression with transcription factor binding sites (TFBS) and CpG islands identified for the human ALDH18A1 gene. Two Aldh18a1 genes were observed on chromosome 1 of the rat genome, apparently resulting from tandem duplication. Phylogenetic analyses also describe the relationships and potential origins of the ALDH18A1 gene during vertebrate and invertebrate evolution and a proposal for generating the bifunctional vertebrate and invertebrate ALDH18A1 gene from a bacterial operon (proBA) encoding G5K and GPR.

MATERIALS AND METHODS

ALDH18A1 gene and enzyme identification

ALDH18A1 amino acid sequences for vertebrate and invertebrate species were retrieved from databases, NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and ExPASy (http://www.expasy.org) [13], using human, mouse, fruit fly and worm ALDH18A1 sequences to seed searches. Identification of the genes was based on high predictive scores (>850) and sequence coverage (>98%) for ALDH18A1 proteome.
sequences (as listed by NCBI) in each case (Table 1 and Supplementary Table 1). BLAT searches were performed using relevant ALDH1A1 protein sequences to confirm the gene and enzyme sequences among the species examined using the UCSC Genome Browser [14]. Gene locations, predicted gene structures and protein subunit sequences were obtained for each ALDH1A1 examined showing identity with the respective ALDH sequences (Tables 1 and Supplementary Table 1). Representations of human ALDH1A1 gene structures were obtained using the AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/) web browser. Identification of potential ALDH1A1 gene regulatory sites, including transcription factor binding sites (TFBS) and CpG islands within the respective gene promoter regions, was undertaken using the UCSC Human Genome Browser [14].

**Predicted structures and properties of vertebrate ALDH1A1 subunits**

Predicted secondary and tertiary structures for vertebrate G5K and GPR domain sequences were obtained using SWISS MODEL web tools [15]. The *Escherichia coli* G5K (pd:2J5T) [16] and human GPR [17] tertiary structures served as references for obtaining these structures.

**Amino acid sequence alignments and phylogenetic analyses**

Alignments of identified vertebrate G5K and GPR domain sequences were undertaken using Clustal Omega, a multiple sequence alignment program [18]. Phylogenetic analyses used several bioinformatic programs, coordinated using the http://www.phylogeny.fr/ bioinformatic portal, to enable alignment (MUSCLE), curation (Gblocks), phylogeny (PhyML) and tree rendering (TreeDyn), to reconstruct phylogenetic relationships [19]. Sequences were identified as ALDH1A1 members, including the proposed primordial ALDH1A1 genes and ALDH1A1 proteins identified for the fruit fly, *Drosophila melanogaster*, and round worm, *Caenorhabditis elegans*.

**RESULTS AND DISCUSSION**

**Alignments of vertebrate and invertebrate G5K and GPR amino acid sequences**

Amino acid sequence alignments for human, zebra fish (*Danio rerio*), round worm (*Caenorhabditis elegans*), and fruit fly (*Drosophila melanogaster*) G5K domain sequences (see Table 1) are shown in Figure 1. Comparisons of these sequences with the bacterial (*Campylobacter jejuni*) G5K sequence, for which the tertiary structure has been described (template pdb:2AKO) [20], enabled prediction
of secondary structures and likely key residues contributing to catalysis, structure and function. Active site residues (human G5K numbers used) binding the substrate (Ser117, Asp223, Asn246) or nucleotide (Ala80, Ser266-Asp267-Val268, Leu271-Phe272, Met305-Gly-306, Gly307, Met308, Glu309-Ala310-Lys311, Lys348) were conserved or underwent conservative substitutions for all of these G5K sequences. Twelve predicted α-helices and 9 β-sheets were observed for the human G5K sequence, including α3 in the region for ATP binding, which was fully conserved among the sequences examined. (Figure 1). Percentage identities for these aligned vertebrate and invertebrate G5K domain sequences were >53% identical indicating that these sequences have been strongly conserved throughout evolution.

Amino acid sequence alignments for GPR domain sequences from these vertebrate and invertebrate species are shown in Figure 2. A comparison of these sequences with the human GPR sequence, for which the tertiary structure has been reported (template pdb:2H5G) [17], enabled identification of key residues for this domain (human GPR numbers used), including substrate binding (Glu581 and Cys612) or stabilizing the transition state for the catalyzed reaction (Asn499). In addition, a fully conserved dinucleotide-binding motif and NAD(P)$^+$ binding domain (Leu501-Leu-Leu-Lys-Gly-Gly-Lys-Glu-Ala509) was observed for these vertebrate and invertebrate GPR sequences. Eighteen α-helices and 15 β-sheets were observed for the human GPR sequence, including α19 and β13 in the NAD(P)$^+$ binding site, which was highly conserved among the sequences examined. (Figure 2). Overall percentage identities for these aligned vertebrate and invertebrate GPR domain sequences were >58% identical indicating that these were also strongly conserved.

**Predicted gene locations and exonic structures for **ALDH18A1** genes and proteins**

Table 1 and Supplementary Table 1 summarize the predicted locations, sizes and number of coding exons for vertebrate **ALDH18A1** genes examined, and of the encoded ALDH18A1 subunit amino acid sequences. These were based on BLAST interrogations of **ALDH18A1** databases (http://blast.ncbi.nlm.nih.gov/BLAST.cgi) using the reported sequences for human and mouse ALDH18A1 [2,3,8] and BLAT analyses of vertebrate genomes using the UC Santa Cruz Genome Browser (http://genome.ucsc.edu/cgi-bin/hgBlat) [13]. In all cases for the vertebrate **ALDH18A1** genes examined, there were 17 coding exons observed (Table 1 and Supplementary Table 1), whereas fruit fly and
roundworm ALDH1A1 genes contained fewer coding exons (12 and 9, respectively). This suggested that the primordial vertebrate ALDH1A1 gene encoded both G5K and GPR domain sequences, as for the invertebrate ALDH1A1 genes examined, which is in contrast to the corresponding genes (ProB and ProA) in a bacterial species (Escherichia coli), which are localized together as separate genes within the PRO operon [21].

Table 1 and Supplementary Table 1 show comparative locations and coding exon compositions for vertebrate and invertebrate ALDH1A1 genes. Two ALDH1A1 genes were observed on rat chromosome 1 in closely located positions and with the same orientations, which is suggestive of a gene duplication event of an ancestral rat ALDH1A1 gene.

Figure 3 presents the genomic sequences for the two major isoforms derived from vertebrate ALDH1A1 genes, confirming previous reports for human and mouse ALDH1A1 genes and proteins, for which a dipeptide (Asn239-Val240) is either present (ALDH1A1long or ALDH1A1a) or absent (ALDH1A1short or ALDH1A1b). As previously proposed [3,8], it would appear that these isoforms have resulted from slippage in gene splicing, with two GT sequences, separated by conserved AAAT/C nucleotide sequences, for all vertebrate ALDH1A1 genes examined. This high level of conservation is extended into the proximate intronic sequence for 5 or 6 nucleotides, following which sequence variability was observed until the AG dinucleotide sequence, required for directing splicing specificity. It is apparent that these multiple ALDH1A1 isoforms were present for all vertebrate genomes examined, which extends the biochemical rationale for these isoforms being differentially distributed and inhibited by ornithine to all vertebrate ALDH1A1 genes examined.

Predicted transcription factor binding sites (TFBS) located in the promoter region for the human ALDH1A1 gene were observed: two N-myc proto-oncogene binding sites located together in sequence, which have been implicated in activating growth-related genes [22]; three cyclic AMP-dependent transcription factors (ATF-1, ATF-2 and ATF-6 alpha) which trigger cell proliferation and transformation [23]; and GATA-1 or erythroid transcription factor, which serves as a general switch factor for erythroid development [24]. These binding sites are in addition to two previously reported p53-binding consensus sequences in the promoter region and in intron 1 of the human ALDH1A1 gene [8]. The ALDH1A1 promoter also contains a large CpG island (CpG94: genome size of 866 nucleotides) which may contribute
to the high level and wide tissue expression of this gene previously reported (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/).

**Phylogeny and evolution of vertebrate and invertebrate ALDH18A1 sequences**

A phylogram (Figure 4) was calculated by the progressive alignment of vertebrate ALDH18A1 amino acid sequences, using invertebrate ALDH18A1 sequences (from *D. melanogaster* and *C. elegans*) (Table 1; Supplementary Table 1) to ‘root’ the tree. ALDH18A1 sequences were identified for all mammalian, bird, lizard, frog, bony fish and cartilaginous fish genomes examined. The phylogram demonstrates separation of these sequences into distinct groups consistent with their relatedness during vertebrate evolution, and suggests that these genes have been derived from an ancestral invertebrate *ALDH18A1* gene. The two rat *Aldh18a1* genes reported earlier in this paper encoded identical rat ALDH18A1 proteins with almost identical genomic sequences located in tandem on rat chromosome one, suggesting that these genes are a product of a recent gene duplication event not shared with the mouse *Aldh18a1* gene.

Figure 5 summarizes a working hypothesis for the evolution of vertebrate and invertebrate *ALDH18A1* genes:

1. A proposed primordial *ALDH18A1*-like gene was derived from a bacterial ancestor, for which G5K and GPR domains were encoded together on a *PRO* operon, which was responsible for proline biosynthesis in bacteria [21].

2. A gene fusion step for the G5K and GPR encoding genes took place within the primordial invertebrate genome, resulting in the formation of the *G5K-GPR* (or *ALDH18A1*) gene, which was integrated into subsequent invertebrate and vertebrate genes. This has been also described for *ALDH1L*-like genes and proteins, encoding mitochondrial and cytosolic folate dehydrogenases [26]; and for Frizzled-like CRD (cysteine-rich domain) [27] and ATP citrate lyase genes and proteins [28].

3. A subsequent gene duplication event occurred for an ancestral rat *Aldh18a1* gene, apparently via tandem duplication, resulting with the presence of consecutive *Aldh18a1* genes, on rat
CONCLUSIONS

BLAST and BLAT analyses of several vertebrate and invertebrate genome databases were undertaken using amino acid sequences reported for human and mouse ALDH18A1 for interrogation of vertebrate and invertebrate genome sequences. Predicted amino acid sequences for these vertebrate and invertebrate ALDH18A1 subunits showed a high degree of sequence identities and the presence, in each case, of encoded G5K and GPR domain sequences. Secondary structure and key residue identification were undertaken using previous reports for bacterial G5K and human GPR. Bioinformatic analyses enabled the identification of putative gene regulation sites, including transcription factor binding sites (TFBS) and a CpG island within the human ALDH18A1 gene promoter, in addition to the p53-binding consensus sequences, previously reported [8].

Sequences for vertebrate ALDH18A1 genes were examined in the region enabling two isoforms to be generated, which play key roles in the regulation of proline and ornithine biosynthesis in the body [3,8]. The nucleotide sequence conservation observed in the region of the splicing junction responsible for the formation of the 2 major isoforms suggested that ALDH18A1 isoform formation is subject to selection during vertebrate evolution to enable the ALDH18A1 short isoform (minus the dipeptide sequence) to be expressed only in gut cells. Intestinal epithelial cells play a major role in glutamine and glutamate metabolism, particularly following heavy protein ingestion in the diet [29]. Given that the ALDH18A1 short isoform is subject to feedback inhibition by ornithine [3,8], its presence in the gut during periods of high protein feeding, may contribute to increased availability of glutamate for circulation within the body.

Phylogenetic analyses also suggested that vertebrate and invertebrate ALDH18A1 genes were derived from a gene integration event, during which G5K (ProB) and GPR (ProA) genes, located together on a bacterial operon [21], underwent gene fusion to form the primordial invertebrate ALDH18A1 gene. A subsequent tandem gene duplication event of an ancestral rat ALDH18A1 gene, apparently generated two rat Aldh18a1 genes on rat chromosome one.

ACKNOWLEDGEMENTS
REFERENCES


**LEGENDS FOR FIGURES**

**Figure 1:** Amino acid sequence alignments for human, zebra fish, worm ({C. elegans}) and fruit fly ({D. melanogaster}) G5K domain (ALDH18A1) sequences

G5K refers to the glutamate kinase domain; see Table 1 for details of {ALDH18A1} genes and proteins; “*•” shows identical residues for G5K domain sequences; “*•” similar alternate residues; “.” less similar alternate residues; ATP binding residues (based on [16]) are shown; **bold** font shows known or predicted exon junctions; exon numbers refer to the human {ALDH18A1} gene; * designates residues involved in nucleotide binding; # for substrate binding residues; & designates dipeptide (Asn239-Val240) residues involved in isoform differences; predicted α-helices (white enclosed in black) and β-sheets (shaded) are numbered from the amino-terminus end.

**Figure 2:** Amino acid sequence alignments for human, zebra fish, worm ({C. elegans}) and fruit fly ({D. melanogaster}) GPR domain (ALDH18A1) sequences

GPR refers to the glutamyl phosphate reductase domain; see Table 1 for details of {ALDH18A1} genes and proteins; “*•” shows identical residues for GPR domain sequences; “*•” similar alternate residues; “.” less similar alternate residues; NAD(P)H binding residues and likely active site (●) residues (Asn499; Lys504; 581Glu; and 612Cys) (based on [17])
are shown; **bold** font shows known or predicted exon junctions; exon numbers refer to the human ALDH18A1 gene; predicted α-helices (white enclosed in black) and β-sheets (shaded) are numbered from the amino-terminus end.

**Figure 3: Sequences for vertebrate ALDH18A1 gene regions encoding isoforms a (ALDH18A1.long) and b (ALDH18A1.short)**

See Tables 1 and Supplementary Table 1 for sources of vertebrate ALDH18A1 gene sequences; isoform a encodes the longer sequence containing the dipeptide (Asn239-Val240); the isoform b sequence lacks these residues; nucleotide sequences are aligned, showing identical residues (*) for the exonic (shown in capitals) and intronic (in lower case) sequences; ‘splicing junction’ specific nucleotide residues (GT (or gt)----AG) are shaded; human and coelacanth ALDH18A1 amino acid sequences are shown, with differences from human sequence shaded.

**Figure 4: Phylogenetic tree for vertebrate and invertebrate ALDH18A1 sequences**

The tree is labeled with the gene name and the name of the vertebrate or invertebrate. The tree is ‘rooted’ with the invertebrate ALDH18A1 sequences. See Table 1 and Supplementary Table 1 for details of ALDH18A1 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 (shown at each node). Only replicate values of 90 or more are highly significant, with 100 bootstrap replicates performed in each case.

**Figure 5: A proposal for the evolutionary appearance of vertebrate and invertebrate ALDH18A1 genes derived from the fusion of bacterial ProB and ProA genes**

Proposed evolutionary appearance of vertebrate and invertebrate ALDH18A1 genes and proteins resulting from a fusion of bacterial ProB (encoding G5K) and ProA (encoding GPR) genes. Gene details are shown in Table 1 (based on [21]).

**LEGENDS for TABLES**

**Table 1: Vertebrate and invertebrate ALDH18A1 genes and enzymes**

RefSeq refers to the NCBI reference sequence; *predicted NCBI sequence; na-not available; ^gene scaffold ID

**Supplementary Table 1: Vertebrate ALDH18A1 genes and enzymes**

RefSeq refers to the NCBI reference sequence; *predicted NCBI sequence; **mRNA sequence; ^gene scaffold ID
Fig. 1. Amino acid sequence alignments for human, zebra fish, worm (C. elegans) and fruit fly (D. melanogaster) GSK domain (ALDH1B1) sequences. GSK refers to the glutamate kinase domain. See Table 1 for details of ALDH1B1 gene and protein: *** shows identical residues for GSK domain sequences; ** similar alternate residues; * less similar alternate residues; ATP binding residues (based on (8)) are shown; bold shows known or predicted exons; exons numbers refer to the human ALDH1B1 gene. Designates residues involved in nucleotide binding: form a nucleotide binding pocket; 2 designates residues involved in substrate binding: 3 designates residues involved in substrate binding (Asn339-Val340) residues involved in isomer difference; predicted α-helices (white enclosed in black) and β-sheets (shaded) are numbered from the amino-terminus end.
Fig. 2. Amino acid sequence alignments for human, zebrafish, worm (C elegans) and fruit fly (D melanogaster) GPR domain (ADHRMM) sequences. GPR refers to the glutamyl phosphate reductase domain; see Table 1 for details of ADHRMM genes and proteins. "**" shows identical residues for GPR domain sequences; "*" similar alternate residues. NAD(P)H binding residues and likely acvriate (A) residues (Ahr499, Lpc004, Sbt142, and O124y) based on [18] are shown; bold font shows known or predicted exon junctions; exon numbers refer to the human ADHRMM gene; predicted α-helices (white enclosed in black) and β-sheets (shaded) are numbered from the amino-terminus end.
Fig. 1. Sequences for vertebrate ALDH1A1 gene regions encoding isoforms a (ALDH1A1(knot)) and b (ALDH1A1(ex19)). See Table 1 for sources of vertebrate ALDH1A1 gene sequences; isoform a encodes the longer sequence containing the dippeptide (Arg210-Val211); the isoform b sequence lacks these residues; nucleotide sequences are aligned, showing identical residues (*) for the exonic (shown in capital) and intronic (in lower case) sequences; "splicing junctions" specific mark nucleotide residues (GT(exon))——(AG) are shaded; human and coelacanth ALDH1A1 amino acid sequences are shown, with differences from human sequence shaded.

Fig. 4. Phylogenetic tree for vertebrate and invertebrate ALDH1A1 sequences.
Fig. 5. A proposal for the evolutionary appearance of vertebrate and invertebrate ALDH1A1 genes derived from the fusion of bacterial ProB and ProA genes. Proposed evolutionary appearance of vertebrate and invertebrate ALDH1A1 genes and proteins resulting from a fusion of bacterial ProB (encoding G2A) and ProA (encoding GPK) genes. Gene details are shown in Table 1 (based on [27]).

Table 1

<table>
<thead>
<tr>
<th>Animal/Organism</th>
<th>Species</th>
<th>Gene</th>
<th>RefSeq ID</th>
<th>ReSeq ID Prediction</th>
<th>Chromosome Location</th>
<th>Coding Strand</th>
<th>UNIPROT ID</th>
<th>Amino acids (f)</th>
<th>Amino acids (l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Homo sapiens</td>
<td>ALDH1A1</td>
<td>NM_00107423</td>
<td>10:57,369,522-57,413,134</td>
<td>17 (-)</td>
<td>P54886</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orangutan</td>
<td>Pongo abelii</td>
<td>ALDH1A1</td>
<td>NM_00107424</td>
<td>10:54,518,106-54,566,982</td>
<td>17 (-)</td>
<td>Q5SR4M</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baboon</td>
<td>Papio anubus</td>
<td>ALDH1A1</td>
<td>XP_002121311</td>
<td>9:87,392,288-9:87,426,030</td>
<td>17 (-)</td>
<td>AA9066PSR</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macaques</td>
<td>Macaca fascicularis</td>
<td>ALDH1A1</td>
<td>XP_00006311</td>
<td>12:82,657,023-12:82,693,191</td>
<td>17 (-)</td>
<td>P71R5S</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>Mus musculus</td>
<td>ALDH1A1b</td>
<td>NM_019998</td>
<td>19:40,551,151-19:40,585,577</td>
<td>17 (-)</td>
<td>Q29IT1O</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>Rattus norvegicus</td>
<td>ALDH1A1a</td>
<td>NM_01108524</td>
<td>12:66,506,497-12:66,535,023</td>
<td>17 (-)</td>
<td>D3B1O</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cow</td>
<td>Bovus</td>
<td>ALDH1A1a</td>
<td>NM_01108525</td>
<td>12:67,086,300-12:67,114,928</td>
<td>17 (-)</td>
<td>D3B1O</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dog</td>
<td>Canis familiaris</td>
<td>ALDH1A1</td>
<td>XP_005561701</td>
<td>28:5,162,413-28:5,144,309</td>
<td>17 (-)</td>
<td>G5SSQ2</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit</td>
<td>Oryctolagus cuniculus</td>
<td>ALDH1A1</td>
<td>XP_002896460</td>
<td>18:4,403,355-18:4,435,877</td>
<td>17 (-)</td>
<td>G1SOQ2</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taumanian devil</td>
<td>Sarcophilus harrisii</td>
<td>ALDH1A1</td>
<td>XP_012935806</td>
<td>18:4,403,355-18:4,435,877</td>
<td>17 (-)</td>
<td>GPWXX1</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opossum</td>
<td>Monodelphis domestica</td>
<td>ALDH1A1</td>
<td>XP_007241270</td>
<td>17:5,928,061-17:5,968,388</td>
<td>17 (-)</td>
<td>F71R5S</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicken</td>
<td>Gallus gallus</td>
<td>ALDH1A1</td>
<td>XP_004941346</td>
<td>6:34,845,525-6:34,879,155</td>
<td>17 (-)</td>
<td>E1QG3</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lizard</td>
<td>Anolis carolinensis</td>
<td>ALDH1A1</td>
<td>XP_001800631</td>
<td>5:36,000,129-5:36,021,712</td>
<td>17 (-)</td>
<td>G4R82</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frog</td>
<td>Xenopus tropicalis</td>
<td>ALDH1A1</td>
<td>XP_006389317</td>
<td>7:173,698:173,698:173,698:173,698</td>
<td>17 (-)</td>
<td>Q6QQG</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zebra fish</td>
<td>Danio rerio</td>
<td>ALDH1A1</td>
<td>XP_003305061</td>
<td>12:30,674,720-12:30,692,083</td>
<td>17 (-)</td>
<td>A4R4C8</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roundworm</td>
<td>Caenorhabditis elegans</td>
<td>ALDH1A1</td>
<td>XM_00777132</td>
<td>X:17,792,332-17,792,917</td>
<td>17 (-)</td>
<td>P54889</td>
<td>802 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit fly</td>
<td>Drosophila melanogaster</td>
<td>ALDH1A1</td>
<td>NM_001299848</td>
<td>3:21,892,022-3:21,882,788</td>
<td>17 (-)</td>
<td>G4R82</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterium</td>
<td>E. coli</td>
<td>ALDH1A1</td>
<td>NM_001299848</td>
<td>3:21,892,022-3:21,882,788</td>
<td>17 (-)</td>
<td>G4R82</td>
<td>795 793</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>