Computational analyses of mammalian lactate dehydrogenases: Human, mouse, opossum and platypus LDHs

Author
Holmes, Roger, Goldberg, Erwin

Published
2009

Journal Title
Computational Biology and Chemistry

DOI
https://doi.org/10.1016/j.compbiolchem.2009.07.006

Copyright Statement
Copyright 2009 Elsevier. This is the author-manuscript version of this paper. Reproduced in accordance with the copyright policy of the publisher. Please refer to the journal's website for access to the definitive, published version.

Downloaded from
http://hdl.handle.net/10072/29281
In silico analyses of mammalian lactate dehydrogenases: human, mouse, opossum and platypus LDHs

Roger S Holmes1,3, and Erwin Goldberg2

1School of Biomolecular and Physical Sciences, Griffith University, Nathan, Qld, Australia; 2Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois, 60201, USA

Corresponding Author:
Roger S Holmes, D.Sc.
School of Biomolecular and Physical Sciences
Griffith University
Nathan, Qld, 4111, Australia.
Email: r.holmes@griffith.edu.au Phone: 61-7-3735-5077

Keywords: Mammals; data mining; sequence analyses; lactate dehydrogenase
Running Head: Mammalian Lactate Dehydrogenase Genes

Summary

Three major mammalian lactate dehydrogenase (LDH) genes and proteins have been extensively investigated, including LDHA (major muscle isozyme); LDHB (major heart isozyme); and LDHC (major sperm isozyme), and another (LDH6B) has been reported in humans. In this study, in silico methods were used to predict the amino acid sequences, structures and gene locations for LDH genes and proteins using genome sequence databanks for human, mouse, opossum and platypus mammalian species. Amino acid sequence alignments and predicted secondary and tertiary structures enabled observation (by similarity) of key residues previously reported for human and mouse LDHA, LDHB and LDHC subunits. The human genome contained at least 4 LDH genes encoding LDHA, B, C and 6B subunits, with the predicted LDH6B gene showing no evidence of introns. Two other human LDH6-like genes were also observed, including LDH6A (7 introns) and LDH6C (single exon). Human LDHA, LDHC and LDH6A genes were located in tandem on chromosome 11, while LDH6B and LDH6C genes were on chromosomes 15 and 12, respectively. In silico evidence was obtained for at least 13 human LDH pseudogenes located on 10 separate chromosomes of the human genome, of which seven were imbedded within introns of other genes involved in distinct but unrelated functions. Opossum LDHC and LDH6B genes were located in tandem with the opossum LDHA gene on chromosome 5 and contained 7 (LDHA and LDHC) or 8 (LDH6B) exons. An amino acid sequence prediction for the opossum LDH6B subunit gave an extended N-terminal sequence, similar to the human and mouse LDH6B sequences, which may support the export of this enzyme into mitochondria. The platypus genome contained at least 3 LDH genes encoding LDHA, LDHB and LDH6B subunits. Phylogenetic studies and sequence analyses indicated that LDHA, LDHB and LDH6B genes are present in all mammalian genomes examined, including a monotreme species (platypus), whereas the LDHC gene may have arisen more recently in marsupial mammals.

Keywords: Mammals; amino acid sequence; genomics; lactate dehydrogenase; opossum; platypus.

Running Head: Mammalian LDH Genes and Subunits

Published in:

Computational Biology and Chemistry (2009) 33:379-385
**Introduction**

Mammalian lactate dehydrogenase (LDH; E.C.1.1.1.27) comprises three major families of conserved enzymes that catalyse the reversible interconversion of pyruvate and lactate, a key metabolic step in glycolysis and other metabolic pathways (Everse & Kaplan, 1973). At least five LDH tetrameric isozymes are reported in somatic mammalian tissues, comprising LDHA and LDHB subunits, whereas the homotetrameric LDHC\textsubscript{4} isoform is found only in mature testis and spermatozoa (Goldberg & Hawtrey, 1967; Goldberg, 1973; Li et al., 1989). The LDHA, LDHB and LDHC families of mammalian LDH genes and subunits have been extensively investigated, with human and mouse LDHA and LDHC genes located in tandem on chromosomes 11 and 7 respectively (Edwards et al., 1989), as compared with the LDHB gene, on chromosomes 12 (human) and 6 (mouse) (Takeno & Li, 1989). Phylogenetic studies have indicated that the LDHC gene has arisen from independent gene duplication events during vertebrate evolution, including separate LDHB gene duplications in fish and birds (pigeon) (Zinkham et al., 1969; Markert et al., 1975; Hiraoka et al., 1989; Quattro et al., 1993; Mannen et al., 1997), and an LDHA gene duplication during mammalian evolution (Millan et al., 1987).

Transcription studies have reported two other human LDHA-like genes, designated as LDH6A and LDH6B, which are expressed in brain and testis respectively, and located on chromosome 11 (LDH6A in tandem with human LDHA and LDHC genes) (Ota et al., 2004) and chromosome 15 (LDH6B, an intronless gene) (Wang et al., 2005). In this study, we have identified and characterized in silico new forms of mammalian LDHs and described predicted amino acid sequences, protein subunit structures, gene locations and exon structures for human (LDH6C), mouse (LDH6B), opossum (LDHA; LDHB; LDHC; and LDH6B) and platypus (LDHA, LDHB and LDH6B) genes and proteins, as well as the phylogenetic relationships for mammalian LDH gene families. In silico evidence is also presented for N-terminal extensions of LDH6B subunit sequences which may support mitochondrial export and location of human, mouse and opossum LDH6B.

**Materials and Methods**

*In silico mammalian LDH gene and protein identification.*

BLAST (Basic Local Alignment Search Tool) studies were undertaken using web tools from the National Center for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1997). Protein BLAST analyses used previously reported human LDHA (Tsujibo et al., 1985), LDHB (Takeno and Li, 1989), LDHC (Millan et al., 1987) and LDH6B (Ota et al., 2004) amino acid sequences. Non-redundant protein sequence databases for several mammalian genomes were examined using the blastp algorithm, including the human (International Human Genome Sequencing Consortium, 2001); mouse (Mus musculus) (Mouse Sequencing Consortium, 2002); opossum (Mikkelsen et al., 2007); and platypus (Platypus Genome Sequencing Consortium, 2008). This procedure produced multiple BLAST ‘hits’ for each of the protein databases which were individually examined and retained in FASTA format, and a record kept of the sequences for predicted mRNAs and encoded CES-like proteins. These records were derived from annotated genomic sequences using the gene prediction method: GNOMON and predicted sequences with high similarity scores for mammalian LDH. With some exceptions, predicted LDHA, LDHB, LDHC and LDH6B protein subunit sequences were obtained in each case and subjected to in silico analyses of predicted protein and gene structures. Other LDH sequences were obtained following BLAT (BLAST-Like Alignment Tool) in silico analysis using the human LDHA, LDHB, LDHC and LDH6B sequences to interrogate human, mouse, opossum and platypus genome sequences using the UC Santa Cruz gene browser (http://genome.ucsc.edu/cgi-bin/hgBlat) (Kent et al. 2003) with the default settings to obtain Ensembl generated protein sequences by applying the method of Hubbard et al (2002) (http://www.ensembl.org/index.html).

BLAT analyses were subsequently undertaken for each of the predicted LDH amino acid sequences using the UC Santa Cruz gene browser (http://genome.ucsc.edu/cgi-bin/hgBlat) (Kent et al. 2003) with the default settings to obtain the predicted locations for each of the mammalian LDH genes, including predicted exon boundary locations and gene sizes. For a study of predicted human LDH pseudogenes, BLAT analyses were undertaken of the human genome using human LDHA, LDHB, LDHC and LDH6B-like subunit sequences in each case (see Table 1; Figure 1). Predicted human LDH pseudogene structures were deduced following corrections for changes in sequence and size, and details recorded for each pseudogene, including the BLAT
score, percentage of identity with the LDH subunit sequence used and its location within the human genome. Structures for human LDHA, LDHB and LDHC isoforms (splicing variants) were obtained using the AceView website (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html?human) to examine predicted gene and protein structures using this database of human mRNA sequences (Thierry-Mieg and Thierry-Mieg, 2006).

Predicted Structures and Properties for Mammalian LDH Subunits.

Predicted secondary and tertiary structures for human and other mammalian LDH-like subunits were obtained using the PSIPRED v2.5 web site tools provided by Brunel University [http://bioinf.cs.ucl.ac.uk/psipred/psiform.html] (McGuffin et al. 2000) and the SWISS MODEL web tools [http://swissmodel.expasy.org/], respectively (Guex & Pietzsch 1997; Kopp & Schwede 2004). Reported tertiary structures for human LDHA (PDB ID 1i10A), human LDHB (PDB ID 1i0zA) (Read et al., 2001) and mouse LDHC (PDB ID 9ldtA) (Hogrefe et al., 1987) served as references for predicted opossum LDHA and LDH6B; LDHB; and LDHC tertiary structures, respectively. Modeling ranges for the opossum LDH residues were as follows: 2 to 332 (LDHA); 2 to 333 (LDHB); 2 to 331 (LDHC); and 51-381 (LDH6B).

Theoretical isoelectric points and molecular weights for mammalian LDH subunits were obtained using Expasy web tools (http://au.expasy.org/tools/pi_tool.html). In silico prediction of an LDH N-terminal protein region that may support a mitochondrial targeting sequence and the identification of a potential cleavage site was conducted using MITOPROT web based methods (Claros and Vincens, 1996) (ftp://ftp.biologie.ens.fr/pub/molbio).

Phylogenetic Studies and Sequence Divergence

Phylogenetic trees were constructed using an amino acid alignment from a ClustalW-derived alignment of CES protein sequences, obtained with default settings and corrected for multiple substitutions (Chenna et al 2003; Larkin et al. 2007) [http://www.ebi.ac.uk/clustalw/]. An alignment score was calculated for each aligned sequence by first calculating a pairwise score for every pair of sequences aligned. The alignment ambiguous amino-terminus region was excluded prior to phylogenetic analysis yielding alignments of 332 residues for comparisons of mammalian LDHA, LDHB, LDHC and LDH6B sequences with chicken LDHA and LDHB sequences, which served as 'outgroup' sequences (see Table 1). Sequence identities for mammalian LDH subunits were determined using the SIM-Alignment tool for Protein Sequences [http://au.expasy.org/tools/sim-prot.html] (Pietsch 1995; Schwede et al. 2003).

Results and Discussion

Alignments of human LDHA, LDHB, LDHC, LDH6A, LDH6B and LDH6C amino acid sequences.

The amino acid sequences for human LDHA (Tsujibo et al., 1985), LDHB (Takeno and Li, 1989), LDHC (Millan et al., 1987) and LDH6B (Ota et al., 2004) and the in silico derived LDH6A and LDH6C human subunits are aligned in Figure 1 (see Table 1). The predicted mitochondrial N-terminal sequences were aligned prior to phylogenetic analysis yielding alignments of 332 residues for comparisons of mammalian LDHA, LDHB, LDHC and LDH6B sequences with chicken LDHA and LDHB sequences, which served as ‘outgroup’ sequences (see Table 1). Sequence identities for mammalian LDH subunits were determined using the SIM-Alignment tool for Protein Sequences (http://au.expasy.org/tools/sim-prot.html) (Pietsch 1995; Schwede et al. 2003).

Alignments of mammalian LDHA, LDHB, LDHC and LDH6B amino acid sequences.

The amino acid sequences for predicted mouse LDH6B, opossum LDHA, LDHB, LDHC and LDH6B, and platypus LDHA, LDHB and LDH6B subunits are aligned with previously reported sequences for the corresponding human and mouse subunits (Tsujibo et al., 1985; Takeno and Li, 1989; Millan et al., 1987; Fukasawa and Li, 1987; Sakai et al., 1987; Hiraoka et al., 1990) (Figure 2; see Table 1). The predicted opossum and platypus LDH sequences
showed higher levels of identity with homologue sequences from human and mouse sources, particularly for the LDHA and LDHB sequences, which were 89-93% identical and 80-97% identical, respectively. Mammalian LDHC and LDH6B sequences, however, exhibited lower levels of identity, showing 65-74% identity for human, mouse and opossum LDHC sequences and 59-75% for human, mouse and platypus LDH6B sequences, respectively (Table 2). Mammalian LDH6B sequences showed evidence of N-terminus extensions for the predicted mouse, opossum and platypus subunits in comparison with LDHA, LDHB and LDH6B sequences for all species examined (Figure 2). MITOPROT computer based analyses of these sequences predicted high probabilities for mouse and opossum LDH6B subunit export into mitochondria (0.98 and 0.79, respectively), as well as potential cleavage sites at residues 36 (mouse LDH6B) and 49 (opossum LDH6B) (Table 1; Figure 2). The platypus LDH6B sequence, however, differed significantly in this property, with the 53 residue N-terminus extension showing a lower probability as a mitochondrial signal peptide (0.27) (Table 1; Figure 2). Mitochondrial LDH (Brooks et al., 1999) has been previously proposed to play a role in the intracellular lactate shuttle and in lactate clearance by mitochondria, however the responsible LDH isozyme(s) have not been conclusively identified. The identification of a mitochondrial leader sequence for human, mouse and opossum LDH6B subunits may assist further investigations concerning a potential role for mammalian LDH in mitochondrial lactate clearance.

Each of the predicted mouse (LDH6B), opossum (LDHA; LDHB; LDHC; and LDH6B) and platypus (LDHA; LDHB; and LDH6B) sequences aligned closely with the corresponding human and mouse sequences, and all subunits (with one exception), showed sequence identity for the key active site residues previously described for human LDH (see Read et al., 2001). The predicted platypus LDH6B sequence, however, contained an Arg residue in place of the key LDHA coenzyme binding residue (Asn138), which may significantly alter the kinetic properties for this enzyme.

Differences in the theoretical isoelectric points (pI) for opossum and platypus LDHA and LDHB subunits were observed, with LDHA showing higher pI values (7.1 and 8.2) than for the LDHB subunits (5.7 and 7.1), which is consistent with pI differences observed for other mammalian LDHs (Table 1). LDH6B subunits showed higher pI values than for the LDHA and LDHB subunits, which may be explained by the high basic amino acid content for the N-terminus peptide extensions, whereas theoretical pI values for mammalian LDHC subunits were intermediate between LDHB (lower pI) and LDHA/LDH6B (higher pI). Human, mouse and opossum LDHA, LDHB and LDH6B subunits examined contained 331-334 amino acid sequence residues, whereas LDH6B subunits contained 381-385 amino acids due to the N-terminus extensions in each case.

Comparative Mammalian LDH Genomics

The NCBI AceView web browser currently defines the human LDHA gene by 7912 GenBank accessions isolated from a wide range of tissues (Thierry-Mieg and Thierry-Mieg, 2006). These human LDHA transcripts included 20 alternatively spliced variants (LDHA isoforms) which may result from differential 5’ or 3’ end truncations, exon shuffling, overlapping exons with different boundaries and alternative splicing (Figure 3). The human LDHB gene is also defined by a large number of GenBank accessions and the LDHB transcripts included 14 alternatively spliced variants also resulting from differential truncations of the 5’ and 3’ ends, exon shuffling and overlapping exons with different boundaries. In contrast, transcription of the human LDHC gene produced only 6 alternatively spliced mRNAs apparently resulting from exon shuffling (Figure 3). The differential roles for these splicing variants (LDH isoforms) for human LDHA, LDHB and LDH6B isoforms have not been established.

Figures 1 and 2 show the locations of the intron-exon boundaries for the mammalian LDH gene products examined, and compares them with previously reported human and mouse LDH gene structures (Chung et al., 1985; Fusakawa and Li, 1987; Takeno and Li, 1989a,b) and their positioning within the aligned amino acid sequences. The mammalian LDHA, LDHB and LDHC genes examined, and the predicted human LDH6A gene, contained 7 exons in each case, with intron-exon boundaries in identical or comparable positions. In contrast, the human and mouse LDH6B genes were without intronic sequences, confirming a report for the human LDH6B gene (Wang et al., 2005), for which expression was observed in human testis. The predicted LDH6B genes in the opossum and platypus genomes, however, contained 8 exons, with the first exon encoding the predicted N-terminus extensions for these gene products, whereas the other 7 exons were localized in similar or identical positions to other mammalian LDH genes.
Table 1 describes the predicted locations for the mammalian LDH genes examined which showed that human, mouse and opossum LDHA and LDHC genes are located together within respective genomes on chromosomes 11, 7 and 5, respectively. The human and mouse LDHA and LDHC genes are very closely located together being separated by < 7 kilobases of DNA. The predicted human LDH6A gene is also part of this gene cluster on chromosome 11, as is the opossum LDH6B gene on chromosome 5 of the opossum genome. In addition, the platypus LDHA and LDH6B genes are apparently located on or near the same contiguous piece of DNA (Contig3116) suggesting that these genes are also closely located on the platypus genome. In contrast, the human, mouse and opossum LDHB genes are on a separate chromosome to that of the LDHA-like gene cluster (Table 1).

Table 3 compares the predicted locations, sizes, exon number and percentage identities for 13 proposed human LDH pseudogenes. Ten of these predicted pseudogenes showed a higher degree of identity with the human LDHA exonic sequences, and were designated as LDHaps genes; two showed higher sequence identity with human LDHB exonic sequences (designated as LDHBps genes); and one was more closely aligned with human LDHC gene exonic sequences (LDHCPsi1). These predicted genes are apparently located on 9 different chromosomes, and several of these were localized within intron sequences for other genes, which encode proteins responsible for distinct functions in the body, such as the LYST gene, encoding a lysosome trafficking regulatory protein (Barbosa et al., 1996); the DYH6 gene, encoding dynen heavy chain 6 (Ota et al., 2004); the MYO1E gene, encoding myosin 1E (Bement et al., 1994); and the S4A4 gene, encoding a solute carrier protein family 4 (Burnham et al., 1997). It is possible that LDH pseudogenes may perform as yet unknown regulatory functions for a range of genes (eg LYST; DYH6; MYO1E; and S4A4) or may serve as passive genetic elements within intronic sequences for these and other genes of the human genome.

Secondary and Tertiary Structures for Mammalian (and Chicken) LDH Sequences

Figures 1 and 2 show the secondary structures previously reported for human LDHA and LDHB (Read et al., 2001) and for mouse LDHC (Hogrefe et al., 1987) or predicted for mammalian LDHA, LDHB, LDHC and LDH6B subunit sequences, together with human LDH6A and LDH6C sequences. Predicted secondary structures for chicken LDHA and LDHB sequences were also examined as these were used as ‘outgroup’ LDH sequences for comparative analyses of mammalian LDH gene and protein structures. Similar α-helix β-sheet structures were observed for all mammalian and chicken LDH subunits examined, particularly near key residues or functional domains, including active site residues such as the active site proton acceptor (His193), as well as coenzyme (Arg99 and Asn138) and substrate (Arg106; Arg169; Thr248) binding residues (Read et al., 2001; Hogrefe et al., 1987). The obvious major difference in mammalian LDH secondary structure related to the N-terminus extensions for human LDH6B and LDH6C, and for mouse and opossum LDH6B, which contained an additional amphiphilic α-helix at the amino terminus, which may support being exported into mitochondria via these potential mitochondrial leader sequences (see Table 1). Although the platypus LDH6C N-terminal sequence contained a predicted α-helix, this did not extend into regions containing basic amino acid residues which may explain the lower probability for this sequence as a mitochondrial signal peptide (Table 1; Figure 2). Predictions of LDH secondary structure, however, may not fully reflect structures in vivo and serve only as a guide as to the comparative structures for mammalian LDH subunits.

Predicted tertiary structures for opossum LDHA, LDHB, LDHC and LDH6B subunits were examined and compared with previously reported tertiary structures for human LDHA and LDHB (Read et al., 2001), and for mouse LDHC (Hogrefe et al., 1987) (Figure 4). The predicted tertiary structures for opossum LDHA (residues 2 to 332) and LDH6B (residues 51 to 381) were sufficiently similar to the human LDHA structure to be based on the previously reported human LDHA-NADH-oxamate complex structure (Read et al., 2001) (Figure 4). In addition, the predicted structures for opossum LDHB and LDHC were sufficiently similar to the previously reported human LDHB-NADH-oxamate complex (residues 2-333) (Read et al., 2001) and mouse LDHC (residues 2 to 331) (Hogrefe et al., 1987), respectively. It is apparent from these predictions that LDHA, LDHB, LDHC and LDH6B subunits are highly conserved in mammals, and it is likely that LDH subunits in the opossum will resemble the corresponding LDHs in human.

Phylogeny of Mammalian LDH Subunits

A phylogenetic tree (Figure 5) was calculated by the progressive alignment of human LDHA, LDHB, LDHC and LDH6B amino acid sequences with the corresponding LDH sequences from mouse, opossum and the
platypus. Chicken LDHA and LDHB sequences were also included and served as an ‘outgroup’ for this analysis of mammalian LDHs. Four major clusters of mammalian and chicken LDHs were observed: the mammalian (and chicken) LDHA and LDHB gene clusters; the LDH6C gene cluster of human, mouse and opossum; and the LDH6B cluster of human, mouse, opossum and platypus. This is consistent with the existence of four distinct mammalian LDH gene families: LDHA, encoding the major skeletal muscle isozyme; LDHB, encoding the major heart isozyme (Markert et al., 1975); LDHC, encoding the testis and sperm specific isozyme (Millan et al., 1987); and LDH6B, which awaits more detailed investigation. LDHA and LDHB have been described in all vertebrates examined and may be considered as the ‘ancestral’ genes for this enzyme (Holmes, 1972; Markert et al., 1975). In contrast, the LDH6C gene has arisen independently from the LDHB gene in both teleost fish (Quattro et al., 1993) and in some birds (eg. pigeon) (Zinkham et al., 1969; Mannen et al., 1997), while in mammals, the LDHC gene has been apparently formed from an LDHA gene duplication event (Millan et al., 1987; Mannen et al., 1997). Biochemical studies have previously shown that LDHA, LDHB and LDH6C isoforms are present in several Australian marsupials examined, including the pretty-faced wallaby (Macropus parryi), the koala (Phascolarctos cinereus) and the brush-tailed possum (Trichosurus vulpecula) (Holmes et al., 1973) whereas LDHC is apparently absent in monotreme mammals, the echidna (Tachyglossus aculeatus) and the platypus (Ornithorhynchus anatinus) (Baldwin and Temple-Smith, 1973). This study of LDH genes and proteins predicted from the South American gray short-tailed opossum (Monodelphis domestica) genome lends support to the distribution of LDHA, LDHB and LDH6C genes and proteins among marsupials from both Australia and South America. The absence of an LDHC-like gene in the monotreme (platypus) genome, however, suggests that the proposed LDH-A gene duplication event leading to the appearance of the marsupial LDHC gene may have occurred following the separation of marsupial and monotreme common ancestors. In contrast, the mammalian LDH6B gene is apparently present throughout eutherian, marsupial and monotreme mammalian evolution but is apparently absent in the chicken genome (Table 1; Figure 5). A further LDH6C gene duplication event is proposed forming the ancestral LDH6B gene at an earlier stage of mammalian evolution, prior to the separation of monoteromes from the marsupial and eutherian mammalian common ancestors. This is supported by the higher levels of sequence identities observed for LDHA and LDH6B subunits (65-71%) as compared with LDHB and LDH6B subunits (57-62%), and the close locations observed for LDHA and LDH6B genes for the mammalian genomes examined.

Summary and Conclusions

Mammalian LDHs comprise at least four gene families encoding distinct subunits (A; B; C; 6B) which form tetrameric enzymes and catalyze a key step in carbohydrate metabolism in all tissues of the body. LDH genes are differentially expressed in mammalian tissues, with LDHA and LDHB genes exhibiting high expression levels in skeletal and heart muscle respectively, but with wide tissue expression patterns (Everse and Kaplan, 1973; Markert et al., 1975). In contrast, the LDH6C gene is expressed predominantly in spermatoocytes and the mature testis, and is required for male fertility (Odet et al., 2008). This isozyme plays an essential role in ATP production by glycolysis in spermatozoa. The human LDH6B gene has not been extensively studied, but has been shown to lack introns and to be expressed in testis (Wang et al., 2005).

In this study, we report in silico predictions for the amino acid sequences, structures and gene locations for LDH genes and proteins of four mammalian species, the human, mouse, opossum (a South American marsupial) and platypus (an Australian monotreme). The human genome contained at least 4 LDH genes encoding LDH A, B, C and 6B subunits, with the predicted LDH6B gene showing no evidence of introns. Two other human LDH6-like genes were observed, including an intronless LDH6C gene and a proposed LDH6A gene, which contained 7 introns. Human LDHA, LDHC and LDH6A genes were located in tandem on chromosome 11, while LDH6B and LDH6C genes were located on chromosomes 15 and 12, respectively. Several LDH pseudogenes were located elsewhere on the human genome, of which seven were apparently located within introns of other genes involved in distinct but unrelated functions. Opossum LDHC and LDH6B genes were located in tandem with the opossum LDHA gene on chromosome 5 and contained 7 (LDHA and LDHC) or 8 (LDH6B) exons. An amino acid sequence prediction for the opossum LDH6B subunit yielded an extended N-terminal sequence, similar to the human and mouse LDH6B sequences, which are proposed to support the export of these enzymes into mitochondria. The platypus genome contained at least 3 LDH genes encoding LDHA, LDHB and LDH6B subunits. Phylogenetic studies analyses indicated that LDHA, LDHB and LDH6B genes are present in all mammalian genomes examined, including a monotreme (platypus), whereas the
**LDHC** gene may have arisen more recently in marsupial mammals prior to the appearance of eutherian mammals.

**Acknowledgements:**

This project was supported in part by NIH HD05863 to E.G.

**REFERENCES**


Figure Legends:

**Figure 1:** Amino acid sequence alignments for human LDHA, LDHB, LDHC, LDH6A, LDH6B and LDH6C sequences
See Table 1 for sources of LDH sequences; * shows identical residues; Residues identified by MITOPROT as high probability mitochondrial leader sequences; conserved active site residues Arg99 and 106; Asn138; Arg169; His193; and Thr248
Helix (Human LDHA and LDHB or predicted helix); Sheet (Human LDHA and LDHB or predicted sheet). Bold underlined font shows known or predicted exon junctions (). A, B, C, 6A, 6B and 6C refer to the corresponding human LDH subunits.

**Figure 2:** Amino acid sequence alignments for human, mouse, opossum, platypus and chicken LDH sequences
See Table 1 for sources of LDH sequences; * shows identical residues; Residues identified by MITOPROT as high probability mitochondrial leader sequences; conserved active site residues Arg99 and 106; Asn138; Arg169; His193; and Thr248
Helix (Human LDHA and LDHB or predicted helix); Sheet (Human LDHA and LDHB or predicted sheet). Bold underlined font shows known or predicted exon junctions (). LDHs examined included human (hu); mouse (mo); opossum (op); platypus (pl); and chicken (ch). A, B, C and 6B refer to the corresponding LDH subunits.

**Figure 3:** Gene structures and splicing variants for human LDHA, LDHB and LDHC genes
 Isoform variants (a, b, c etc) are shown with capped 5'- and validated 3'-ends for the predicted mRNA sequences. NM numbers refer to annotated RefSeq sequences for human LDHA, LDHB and LDHC genes. Scale refers to base pairs of nucleotide sequences.

**Figure 4:** Three dimensional structures for human LDHA and mouse LDHC subunits and predicted three dimensional structures for opossum LDHA, LDHB, LDHC and LDH6B subunits
Human LDHA and mouse LDHC 3-D structures and predicted opossum LDHA, LDHB, LDHC and LDH6B structures were obtained using the SWISS MODEL web site http://swissmodel.expasy.org/workspace/index.php? (see Table 1). The rainbow color code describes the 3-D structures from the N- (blue) to C-termini (red color). The structures are based on known 3-D structures for human LDHA and LDHB (Read et al., 2001) and mouse LDHC (Hogrefe et al., 1987) complexed with NADH and oxamate. Modeling ranges for the opossum LDH sequences were: LDHA: 2 to 332 based on PDB template 1i10A; LDHB: 2 to 333 based on PDB template 1i0zA; LDHC: 2 to 331 based on PDB template 9ldtA ; and LDH6B: 51 to 381 based on template 1i10A.

**Figure 5:** Phylogenetic tree of mammalian CES6 and of human CES1, CES2, CES3 and CES5 sequences.
The tree is labeled with the LDH family number and the species name. Note the separation of the LDH genes into four LDH family clusters: LDHA; LDHB; LDHC; and LDH6B.

**LEGENDS FOR TABLES**

**Table 1:** Mammalian and chicken lactate dehydrogenase (LDH) genes and enzymes examined
GenBank mRNA (or cDNA) IDs identify previously reported sequences (see http://www.ncbi.nlm.nih.gov/Genbank/); N-scan and SGP IDs identify gene predictions using gene structure prediction software provided by the Computational Genomics Lab at Washington University in St. Louis, MO, USA (see http://genome.ucsc.edu/); UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual LDH subunits (see http://kr.expasy.org/); Mitochondrial export probabilities and predicted signal peptides were based on MITOPROT web based tools (see Methods); Contig ID for platypus genome sequences; Prediction software based ENSOANT IDs; Sources for LDH sequences were provided by the above sources.
Table 2: Percentage identities for mammalian and chicken LDH amino acid sequences
Numbers show the percentage of amino acid sequence identities. Numbers in **bold** show higher sequence identities for eutherian mammalian LDH sequences.

Table 3: Predicted human LDH pseudogenes
Predicted human LDH pseudogenes are named LDHAps, LDHBps or LDHCps in numerical order according to the subunit showing highest sequence identity and BLAT score using the UC Santa Cruz human genome web browser (http://genome.ucsc.edu). ¹BLAT score determined by using the relevant human LDH subunit sequence (A, B or C) to interrogate the human genome; ²percentage identity of the derived human pseudogene sequence with the relevant LDH subunit sequence; ³range of LDH subunit sequence corresponding to the derived LDH pseudogene sequence; ⁴number of relevant LDH residues obtained for the derived LDH pseudogene sequence; ⁵predicted pseudogene size (nucleotides); ⁶predicted exon sequences observed; ⁷GenBank or prediction software based ENSOANT IDs for the pseudogene sequence; ⁸predicted colocation of the LDH pseudogene with another known human gene; ⁹colocated gene function identified.

Figure 1

Figure 2
Figure 3
### Table 1

<table>
<thead>
<tr>
<th>Species</th>
<th>LDH</th>
<th>GenBank ID</th>
<th>UniProt ID</th>
<th>NCBI RefSeq</th>
<th>Chromosome</th>
<th>Strand</th>
<th>Amino Acid Size</th>
<th>pI</th>
<th>MW (kDa)</th>
<th>Probability (Residues)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>LDHA</td>
<td>B0072222</td>
<td>P00228</td>
<td>NP00557</td>
<td>11:18,374,965-18,285,901</td>
<td>positive</td>
<td>332</td>
<td>10,426</td>
<td>8.4</td>
<td>26,639</td>
</tr>
<tr>
<td></td>
<td>LDHB</td>
<td>B0071500</td>
<td>P002195</td>
<td>NP002191</td>
<td>12:21,679,745-11,695,073</td>
<td>negative</td>
<td>334</td>
<td>19,127</td>
<td>5.7</td>
<td>35,620</td>
</tr>
<tr>
<td></td>
<td>LDHC</td>
<td>B0043362</td>
<td>P07064</td>
<td>NP002301</td>
<td>11:18,390,811-18,429,247</td>
<td>positive</td>
<td>332</td>
<td>38,407</td>
<td>7.1</td>
<td>36,281</td>
</tr>
<tr>
<td></td>
<td>LDHA</td>
<td>B0014340</td>
<td>P009409</td>
<td></td>
<td></td>
<td>positive</td>
<td>332</td>
<td>22,167</td>
<td>7.5</td>
<td>35,407</td>
</tr>
<tr>
<td></td>
<td>LDHB</td>
<td>B0022034</td>
<td>Q99712</td>
<td>NP19972</td>
<td>15:57,228,432-17,574,956</td>
<td>positive</td>
<td>381</td>
<td>1,148</td>
<td>8.9</td>
<td>39,643</td>
</tr>
<tr>
<td></td>
<td>LDHC</td>
<td>B0075512</td>
<td>Q95678</td>
<td></td>
<td></td>
<td></td>
<td>373</td>
<td>1,124</td>
<td>8.6</td>
<td>41,157</td>
</tr>
<tr>
<td>Mouse</td>
<td>LDHA</td>
<td>B0046396</td>
<td>P06151</td>
<td>NP008219</td>
<td>7:54,102,990-64,110,568</td>
<td>positive</td>
<td>332</td>
<td>7,519</td>
<td>7.6</td>
<td>36,499</td>
</tr>
<tr>
<td></td>
<td>LDHB</td>
<td>B0046755</td>
<td>P16125</td>
<td>NP032510</td>
<td>6:142,435,860-122,454,660</td>
<td>positive</td>
<td>334</td>
<td>15,101</td>
<td>7.8</td>
<td>37,572</td>
</tr>
<tr>
<td></td>
<td>LDHC</td>
<td>B0049602</td>
<td>Q51682</td>
<td>NP028608</td>
<td>7:54,617,160-64,192,244</td>
<td>positive</td>
<td>382</td>
<td>16,105</td>
<td>8.4</td>
<td>36,912</td>
</tr>
<tr>
<td>Oopossum</td>
<td>LDHA</td>
<td>B0015410</td>
<td>P070055</td>
<td>NP00055</td>
<td>17:5,417,512-5,418,057</td>
<td>negative</td>
<td>302</td>
<td>1,145</td>
<td>9.3</td>
<td>42,049</td>
</tr>
<tr>
<td>Platyopus</td>
<td>LDHA</td>
<td>A056832</td>
<td>Q95678</td>
<td></td>
<td></td>
<td></td>
<td>332</td>
<td>6,890</td>
<td>7.1</td>
<td>38,582</td>
</tr>
<tr>
<td></td>
<td>LDHB</td>
<td>A056896</td>
<td>Q95678</td>
<td></td>
<td></td>
<td></td>
<td>334</td>
<td>22,956</td>
<td>5.7</td>
<td>36,537</td>
</tr>
<tr>
<td></td>
<td>LDHC</td>
<td>A056896</td>
<td>Q95678</td>
<td></td>
<td></td>
<td></td>
<td>249,097</td>
<td>6.8</td>
<td>36,920</td>
<td>0.11 (N)</td>
</tr>
<tr>
<td>Chicken</td>
<td>LDHA</td>
<td>P00340</td>
<td>P009515</td>
<td></td>
<td></td>
<td></td>
<td>332</td>
<td>1,373</td>
<td>7.9</td>
<td>36,514</td>
</tr>
<tr>
<td></td>
<td>LDHB</td>
<td>P00357</td>
<td>P009516</td>
<td></td>
<td></td>
<td></td>
<td>332</td>
<td>9,059</td>
<td>7.1</td>
<td>38,229</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Human A</td>
<td>100</td>
<td>93</td>
<td>90</td>
<td>91</td>
<td>84</td>
<td>75</td>
<td>74</td>
<td>63</td>
<td>74</td>
<td>75</td>
<td>72</td>
<td>74</td>
<td>71</td>
<td>65</td>
<td>65</td>
<td>74</td>
</tr>
<tr>
<td>Mouse A</td>
<td>93</td>
<td>100</td>
<td>92</td>
<td>91</td>
<td>84</td>
<td>75</td>
<td>75</td>
<td>64</td>
<td>73</td>
<td>72</td>
<td>75</td>
<td>72</td>
<td>67</td>
<td>64</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Opossum A</td>
<td>90</td>
<td>92</td>
<td>100</td>
<td>89</td>
<td>82</td>
<td>77</td>
<td>75</td>
<td>75</td>
<td>64</td>
<td>73</td>
<td>72</td>
<td>71</td>
<td>71</td>
<td>67</td>
<td>65</td>
<td>73</td>
</tr>
<tr>
<td>Platypus A</td>
<td>92</td>
<td>91</td>
<td>89</td>
<td>100</td>
<td>85</td>
<td>75</td>
<td>75</td>
<td>74</td>
<td>74</td>
<td>73</td>
<td>77</td>
<td>71</td>
<td>75</td>
<td>70</td>
<td>66</td>
<td>65</td>
</tr>
<tr>
<td>Chicken A</td>
<td>84</td>
<td>84</td>
<td>82</td>
<td>85</td>
<td>100</td>
<td>74</td>
<td>74</td>
<td>63</td>
<td>71</td>
<td>70</td>
<td>68</td>
<td>72</td>
<td>65</td>
<td>60</td>
<td>62</td>
<td>68</td>
</tr>
<tr>
<td>Human B</td>
<td>75</td>
<td>75</td>
<td>77</td>
<td>75</td>
<td>74</td>
<td>100</td>
<td>97</td>
<td>97</td>
<td>91</td>
<td>90</td>
<td>69</td>
<td>65</td>
<td>67</td>
<td>65</td>
<td>62</td>
<td>58</td>
</tr>
<tr>
<td>Mouse B</td>
<td>74</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>74</td>
<td>97</td>
<td>100</td>
<td>96</td>
<td>81</td>
<td>89</td>
<td>69</td>
<td>64</td>
<td>64</td>
<td>62</td>
<td>57</td>
<td>65</td>
</tr>
<tr>
<td>Opossum B</td>
<td>74</td>
<td>74</td>
<td>75</td>
<td>74</td>
<td>74</td>
<td>97</td>
<td>100</td>
<td>80</td>
<td>89</td>
<td>89</td>
<td>64</td>
<td>67</td>
<td>64</td>
<td>62</td>
<td>57</td>
<td>64</td>
</tr>
<tr>
<td>Platypus B</td>
<td>63</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>81</td>
<td>81</td>
<td>80</td>
<td>100</td>
<td>77</td>
<td>59</td>
<td>57</td>
<td>59</td>
<td>55</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>Chicken B</td>
<td>74</td>
<td>73</td>
<td>75</td>
<td>73</td>
<td>71</td>
<td>90</td>
<td>89</td>
<td>89</td>
<td>77</td>
<td>100</td>
<td>67</td>
<td>64</td>
<td>64</td>
<td>61</td>
<td>59</td>
<td>64</td>
</tr>
<tr>
<td>Human C</td>
<td>75</td>
<td>74</td>
<td>73</td>
<td>77</td>
<td>70</td>
<td>69</td>
<td>69</td>
<td>69</td>
<td>59</td>
<td>67</td>
<td>100</td>
<td>73</td>
<td>67</td>
<td>61</td>
<td>59</td>
<td>65</td>
</tr>
<tr>
<td>Mouse C</td>
<td>72</td>
<td>72</td>
<td>71</td>
<td>71</td>
<td>68</td>
<td>65</td>
<td>64</td>
<td>64</td>
<td>57</td>
<td>64</td>
<td>74</td>
<td>100</td>
<td>65</td>
<td>61</td>
<td>61</td>
<td>65</td>
</tr>
<tr>
<td>Opossum C</td>
<td>74</td>
<td>75</td>
<td>74</td>
<td>75</td>
<td>72</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>59</td>
<td>67</td>
<td>73</td>
<td>85</td>
<td>100</td>
<td>67</td>
<td>61</td>
<td>62</td>
</tr>
<tr>
<td>Human 68</td>
<td>71</td>
<td>72</td>
<td>71</td>
<td>70</td>
<td>65</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>56</td>
<td>64</td>
<td>67</td>
<td>65</td>
<td>67</td>
<td>100</td>
<td>75</td>
<td>63</td>
</tr>
<tr>
<td>Mouse 68</td>
<td>65</td>
<td>67</td>
<td>67</td>
<td>66</td>
<td>60</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>55</td>
<td>61</td>
<td>61</td>
<td>61</td>
<td>75</td>
<td>100</td>
<td>60</td>
<td>62</td>
</tr>
<tr>
<td>Opossum 68</td>
<td>65</td>
<td>64</td>
<td>65</td>
<td>65</td>
<td>62</td>
<td>58</td>
<td>57</td>
<td>57</td>
<td>50</td>
<td>59</td>
<td>59</td>
<td>59</td>
<td>61</td>
<td>63</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Platypus 68</td>
<td>74</td>
<td>75</td>
<td>73</td>
<td>77</td>
<td>68</td>
<td>65</td>
<td>65</td>
<td>64</td>
<td>57</td>
<td>64</td>
<td>65</td>
<td>65</td>
<td>64</td>
<td>66</td>
<td>62</td>
<td>59</td>
</tr>
<tr>
<td>Pseudogene</td>
<td>Score</td>
<td>Identity</td>
<td>Range</td>
<td>No. of Residues</td>
<td>Location</td>
<td>Strand</td>
<td>Gene Descriptions</td>
<td>GenBank ID</td>
<td>Gene Colocation</td>
<td>Possible Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>----------</td>
<td>-------</td>
<td>-----------------</td>
<td>----------</td>
<td>--------</td>
<td>-------------------</td>
<td>------------</td>
<td>----------------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp1</td>
<td>882</td>
<td>91</td>
<td>2-331</td>
<td>331</td>
<td>9:14,911,338-14,912,329</td>
<td>negative</td>
<td>992</td>
<td>2</td>
<td>ENST0879581</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp2</td>
<td>739</td>
<td>88</td>
<td>2-331</td>
<td>316</td>
<td>1:233,867,929-233,868,809</td>
<td>positive</td>
<td>981</td>
<td>4</td>
<td>ENST067815</td>
<td>LYST: part of intron for lysosome trafficking regulator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp3</td>
<td>705</td>
<td>88</td>
<td>22-232</td>
<td>310</td>
<td>2:41,900,453-41,901,395</td>
<td>positive</td>
<td>945</td>
<td>3</td>
<td>ENST0394996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp4</td>
<td>702</td>
<td>86</td>
<td>1-332</td>
<td>317</td>
<td>4:49,46,839-49,47,801</td>
<td>negative</td>
<td>964</td>
<td>3</td>
<td>ENST1000077</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp5</td>
<td>638</td>
<td>88</td>
<td>55-328</td>
<td>271</td>
<td>2:84,657,928-84,658,721</td>
<td>negative</td>
<td>913</td>
<td>1</td>
<td>ENST2853945</td>
<td>DHX5: part of intron for dynamin heavy chain 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp6</td>
<td>498</td>
<td>90</td>
<td>128-332</td>
<td>203</td>
<td>10:120,682,174-120,683,785</td>
<td>negative</td>
<td>612</td>
<td>3</td>
<td>ENST020787</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp7</td>
<td>459</td>
<td>76</td>
<td>27-327</td>
<td>301</td>
<td>15:57,286,657-57,287,559</td>
<td>positive</td>
<td>903</td>
<td>1</td>
<td>ENST2887135</td>
<td>MYO1E: part of intron for myosin 1E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp8</td>
<td>383</td>
<td>73</td>
<td>27-296</td>
<td>272</td>
<td>6:1,520,274-1,521,005</td>
<td>negative</td>
<td>812</td>
<td>2</td>
<td>BC029130</td>
<td>BC029130: part of intron for unidentified gene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp10</td>
<td>312</td>
<td>70</td>
<td>27-292</td>
<td>298</td>
<td>12:61,683,812-61,684,723</td>
<td>positive</td>
<td>899</td>
<td>2</td>
<td>ENST21426244363301</td>
<td>part of intron for testis mRNA-encoding gene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp11</td>
<td>731</td>
<td>87</td>
<td>2-334</td>
<td>331</td>
<td>X:75,471,647-75,472,645</td>
<td>negative</td>
<td>999</td>
<td>3</td>
<td>ENST0366646</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp12</td>
<td>199</td>
<td>77</td>
<td>106-306</td>
<td>119</td>
<td>13:112,979,604-112,979,964</td>
<td>positive</td>
<td>361</td>
<td>2</td>
<td>ENST0402271</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHAp13</td>
<td>210</td>
<td>65</td>
<td>22-230</td>
<td>242</td>
<td>4:72,518,995-72,522,295</td>
<td>negative</td>
<td>3301</td>
<td>6</td>
<td>BC030977</td>
<td>SAA4: part of intron for solute carrier protein family 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>