Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 9392–9396, September 1994
Evolution

Evolutionary relationships of lactate dehydrogenases (LDHs) from mammals, birds, an amphibian, fish, barley, and bacteria: LDH cDNA sequences from *Xenopus*, pig, and rat

SOICHI TSUJI*†, MASOOD A. QURESHI*‡, ESTHER W. HOU*, WALTER M. FITCH§, AND STEVEN S.-L. LI*¶

*Laboratory of Genetics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709; and §Department of Ecology and Evolutionary Biology, School of Biological Sciences, University of California, Irvine, CA 92717

Contributed by Walter M. Fitch, June 9, 1994

ABSTRACT **The nucleotide sequences of the cDNAs encoding LDH (EC 1.1.1.27) subunits LDH-A (muscle), LDH-B (liver), and LDH-C (oocyte) from *Xenopus laevis*, LDH-A (muscle) and LDH-B (heart) from pig, and LDH-B (heart) and LDH-C (testis) from rat were determined. These seven newly deduced amino acid sequences and 22 other published LDH sequences, and three unpublished fish LDH-A sequences kindly provided by G. N. Somero and D. A. Powers, were used to construct the most parsimonious phylogenetic tree of these 32 LDH subunits from mammals, birds, an amphibian, fish, barley, and bacteria. There have been at least six LDH gene duplications among the vertebrates. The *Xenopus* LDH-A, LDH-B, and LDH-C subunits are most closely related to each other and then are more closely related to vertebrate LDH-B than LDH-A. Three fish LDH-As, as well as a single LDH of lamprey, also seem to be more related to vertebrate LDH-B than to land vertebrate LDH-A. The mammalian LDH-C (testis) subunit appears to have diverged very early, prior to the divergence of vertebrate LDH-A and LDH-B subunits, as reported previously.**

Lactate dehydrogenase (LDH; EC 1.1.1.27) catalyzes the interconversion of lactate and pyruvate with nicotinamide adenine dinucleotide (NAD^+) as coenzyme (1). LDH, a tetrameric enzyme, is widely distributed among vertebrates, plants, and bacteria. In vertebrates there are three different subunits of LDH isozymes: LDH-A, LDH-B, and LDH-C (2). The LDH-A_4 isozyme is best suited for pyruvate reduction in anaerobic tissues (muscle), whereas the LDH-B_4 isozyme is superior for lactate oxidation in aerobic tissues (heart). The tissue distribution of the LDH-C_4 isozyme varies from organism to organism. In lower teleost fishes, LDH-C has a generalized tissue distribution, but in advanced teleost fishes it is found either in the liver (e.g., in cod) or in the eye (e.g., in salmon). In mammals and columbid birds, the LDH-C_4 isozyme is expressed only in mature testes.

The LDH isozymes were first thought to have arisen from a single LDH-A-like locus in Agnatha, since lampreys have only a single LDH isozyme. This original locus was presumably duplicated to form LDH-A and LDH-B loci; then the LDH-B locus was duplicated to form LDH-B and LDH-C loci (2, 3). Phylogenetic analyses of amino acid sequences revealed that the mammalian testis-specific LDH-C subunit diverged prior to the split of the LDH-A and LDH-B subunits (4–6). A third possibility has also been suggested (7, 8), that the mammalian LDH-C gene originated from a duplication of the LDH-A gene after the A–B duplication. As a separate matter, the teleost LDH-C gene is derived from a duplication of the LDH-B gene (9). No LDH sequences of amphibians and reptiles have been reported, nor has the testis-specific LDH-C_4 isozyme from columbid birds.

We determined the nucleotide sequences of the cDNAs encoding *Xenopus* LDH-A, LDH-B, and LDH-C, pig LDH-A and LDH-B, and rat LDH-B and LDH-C.‖ We have used these seven LDH sequences deduced from their cDNA sequences and 22 other published LDH sequences, plus three unpublished fish LDH-A sequences, to analyze the evolutionary relationships of LDH isozymes from mammals, birds, an amphibian, fish, barley, and bacteria.

MATERIALS AND METHODS

Materials. The adults and tadpoles of *Xenopus laevis* (wild type) were obtained from Xenopus I (Ann Arbor, MI). A *Xenopus* oocyte cDNA library in λgt11 vector was kindly provided by T. R. Burglin (10). A *Xenopus* liver cDNA library in λUni-ZAP XR, a pig muscle cDNA library in λUni-ZAP, and rat testis cDNA in λZAPII vectors were obtained from Stratagene. Rat heart and testis cDNA libraries in λgt11 vectors were purchased from Clontech.

Gel Electrophoresis of *Xenopus* LDH Isozymes. The *Xenopus* LDH isozymes present in the soluble protein fraction from various tissues were separated on agarose gels in nondenaturing buffer (pH 8.2) and activities were visualized by nitroblue tetrazolium reduction to formazan according to the manufacturers' recommendations (Paragon electrophoresis system and LDH isozyme kit, Beckman).

Cloning and Sequencing of LDH cDNAs. The protein-encoding regions of mouse LDH-A (muscle) and LDH-B (heart) cDNAs were used as probes to screen *Xenopus* liver and oocyte cDNA libraries. Pig LDH-A (muscle) and LDH-B (heart) cDNA clones were isolated from muscle cDNA library by using mouse LDH-B cDNA as a probe. A rat LDH-B (heart) cDNA clone was isolated from a heart cDNA library by using a mouse LDH-B cDNA probe, while rat LDH-C (testis) cDNA clones were obtained from testis cDNA libraries by using mouse LDH-C (testis) cDNA as a probe. The cDNA insert from plaque-purified λUni-Zap phages was excised to pBluescript SK (−) phagemids (Stratagene). The cDNA insert from the λgt11 phages was isolated and subcloned in pBluescript phagemid. The nucleotide sequence of the inserted cDNAs was determined by the dideoxynucleotide chain-termination method, using a T7 DNA polymerase and 2′-deoxyadenosine 5′-[α-[^{35}S]thio]-triphosphate (11). Some *Xenopus* cDNA inserts were also

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: LDH, lactate dehydrogenase.

†Present address: Laboratory of Animal Breeding and Genetics, Faculty of Agriculture, Kobe University, Kobe 657, Japan.
‡Present address: Department of Physiology, University of Karachi, Karachi, Pakistan.
¶To whom reprint requests should be addressed.

‖The seven cDNA sequences reported in this paper have been deposited in the GenBank data base (accession nos. U07175–U07181).

labeled with the Dye Terminator kit (Applied Biosystems), and their nucleotide sequences were determined by using an automated DNA sequencer (Applied Biosystems model 373A). Oligonucleotides synthesized according to the newly determined sequences were used as primers to obtain additional sequences.

Amino Acid Sequences of LDH Subunits. Amino acid sequences of pig LDH-A, pig LDH-B, and rat LDH-C subunits were deduced from their cDNA sequences, and they are consistent with the previously published protein sequences (12, 13). The amino acid sequences of rat LDH-B and *Xenopus* LDH-A, LDH-B, and LDH-C subunits were deduced only from their cDNA sequences. The unpublished amino acid sequence of the killifish LDH-A subunit was kindly made available for phylogenetic analysis by D. A. Powers (Stanford University, Stanford, CA) and the sequences of the teleost fish scorpaenid and barracuda LDH-A subunit, by G. N. Somero (Oregon State University, Corvallis). Twenty-two other LDH protein sequences were previously reported: human LDH-A (14, 15), mouse LDH-A (16, 17), rat LDH-A (18), bovine LDH-A (19), rabbit LDH-A (20), chicken LDH-A (21), dogfish (shark) LDH-A (22), lamprey LDH (23), human LDH-B (24, 25), mouse LDH-B (6), chicken LDH-B (26), duck LDH-B (8), killifish LDH-B (5), human LDH-C (7, 27, 28), mouse LDH-C (29, 30), killifish LDH-C (9), barley LDH (31), and five bacterial LDHs, from *Bacillus stearothermophilus* (32, 33), *Bacillus subtilis* (34), *Bifidobacterium longum* (35), *Thermus aquaticus* (36), and *Thermus caldophilus* (37).

The scientific names of the organisms are as follows: human, *Homo sapiens*; mouse, *Mus musculus*; rat, *Rattus norvegicus*; pig, *Sus scrofa*; bovine, *Bos taurus*; rabbit, *Oryctolagus cuniculus*; chicken, *Gallus domesticus*; duck, *Anas platyrhynchos*; African frog, *Xenopus laevis*; dogfish (shark), *Squalus acanthias*; killifish, *Fundulus heteroclitus*; teleost fish barracuda, *Sphyraena argentea*; teleost fish scorpaenid, *Sebastolobus alascanus*; lamprey, *Petromyzon marinus*; and barley, *Hordeum vulgare*.

Construction of Evolutionary Tree. The most parsimonious tree among amino acid sequences of 32 LDH subunits was constructed by the ANCESTOR program of Fitch (38). The amino acids are back-translated into ambiguous codons so that results are reported as nucleotide substitutions, those required for the amino acid replacements (39). Bootstrappings were used to give measures of support for individual clades.

RESULTS

Patterns of *Xenopus* LDH Isozymes. The electrophoretic patterns of *Xenopus* LDH isozymes from muscle, liver, heart, testis, oocyte, and tadpoles of different stages are presented in Fig. 1. In skeletal muscle, the predominant band moving fastest toward the cathode is likely to be homotetrameric LDH-A_4 isozyme. In liver, the five LDH isozymes are presumably formed by tetrameric combinations of

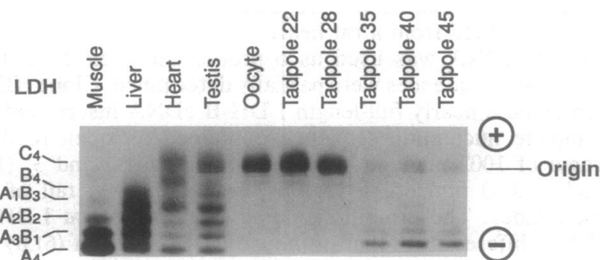

FIG. 1. Electrophoretic patterns of *Xenopus* LDH isozymes from adult tissues and tadpoles.

LDH-A (muscle) and LDH-B (liver) subunits. In oocytes and tadpoles of stages 22 and 28, a single broad band of LDH isozyme is presumably the homotetramer (LDH-C_4) of a third LDH subunit, which is encoded by a gene tentatively designated LDH-C. In heart, the presence of more than five isozyme bands may be derived from tetrameric combinations of three different LDH subunits. In testis, the isozyme pattern differed from that of heart, and seven LDH isozyme bands were observed. The possibility of a testis-specific LDH isozyme remains to be investigated. In tadpoles of stages 35, 40, and 45, LDH-A and LDH-B genes are both expressed, but the LDH-A_4 isozyme is predominant, while the LDH-C_4 isozyme has been greatly reduced. These gel patterns of *Xenopus* LDH isozymes, which are similar to those of previous reports (40, 41), indicate the differential expression of three LDH genes (42).

***Xenopus* LDH-A, LDH-B, and LDH-C Sequences.** *Xenopus* LDH-A and LDH-B cDNA clones were isolated by screening a liver cDNA library, using mixed probes of mouse LDH-A and LDH-B cDNAs, and the cDNA inserts from five positive clones were sequenced. Clones X906, X904, and X908 possessed cDNA inserts of 1.3, 1.1, and 0.8 kb, respectively, and these cDNA sequences appeared to be derived from the same mRNA. Clones X902 and X901 contained cDNA inserts of 1.4 and 0.6 kb, respectively, and their cDNA sequences resulted from the same transcript, which is different from that of clones X906, X904, and X908. The cDNA sequence of clone X906 contained an open reading frame of 1002 nucleotides, 5′ (18 nucleotides) and 3′ (101 nucleotides) noncoding regions, and a poly(A) tail of 8 nucleotides. The deduced sequence of 333 amino acids (Fig. 2) from the cDNA of clone X906 has a net charge of −2 at pH 7.0 and a pI of 6.90. The cDNA sequence of clone X902 contained an open reading frame of 1002 nucleotides, 5′ (18 nucleotides) and 3′ (119 nucleotides) noncoding regions, and a poly(A) tail of 11 nucleotides. The deduced sequence of 333 amino acids (Fig. 2) from the cDNA of clone X902 has a net charge of −3 at pH 7.0 and a pI of 6.71. As described above, the five bands of LDH isozymes present in *Xenopus* liver as well as muscle resulted from tetrameric combinations of LDH-A and LDH-B subunits, with the LDH-A_4 isozyme as the band moving fastest to the cathode. On the basis of pI of the deduced amino acid sequences, clone X906 cDNA contains the coding sequence of LDH-A subunit, while clone X902 cDNA encodes LDH-B subunit.

Xenopus LDH-C cDNAs were isolated by screening an oocyte cDNA library (10), using mixed probes of mouse LDH-B and *Xenopus* LDH-A and LDH-B cDNAs. Fourteen positive clones were purified after the third screening, and the sizes of their cDNA inserts were estimated by PCR amplification using two vector primers and agarose gel electrophoresis. Three clones, X851, X843, and X801, were found to contain the cDNA insert of 1.2 kb, and their partial sequences were obtained. The nucleotide sequences of both strands from clone X851 were completely determined. The first two nucleotides of the initiation Met codon from the cDNA of clone X851 were truncated, and 61 nucleotides were found at its 3′ noncoding region without polyadenylylation signal and poly(A) tail. The deduced sequence of 333 amino acids (Fig. 2) from *Xenopus* LDH-C cDNA of clone X851 has a net charge of −3 at pH 7.0 and pI of 6.70. The net charge and pI of *Xenopus* LDH-C as well as LDH-A and LDH-B subunits are consistent with their electrophoretic mobilities on agarose gel (Fig. 1). The deduced amino acid sequence of *Xenopus* LDH-C is 6.6% and 2.1% different from the sequences of *Xenopus* LDH-A and *Xenopus* LDH-B proteins, respectively. The deduced amino acid sequences from clones X906 (LDH-A) and X902 (LDH-B) are 5.7% different. Therefore, the amino acid sequence of *Xenopus* LDH-C is more similar to *Xenopus* LDH-B than to *Xenopus* LDH-A. More-

```
Xenopus A  ASVQEKLITC VCQDKPAKPT NKITIVGVGQ VGMACAVSVL LKELADELAL   50
Xenopus B  STVQEKLITN VCQDKAAKPT NKITIVGVGQ VGMACAVSVL LKELADELAL   50
Xenopus C  SSVQENLITN VCQDKAAKPT NKITIVGVGQ VGMACAVSVL LKELADELAL   50
Pig     A  ATLKDQLIHN LLKEEH.VPH NKITVVGVGA VGMACAISIL MKELADEIAL   49
Pig     B  ATLKEKLIAP VAEEETTIPN NKITVVGVGQ VGMACAISIL GKSLTDELAL   50
Rat     B  ATLKEKLIAP VADDETAVPN NKITVVGVGQ VGMACAISIL GKSLADELAL   50
Rat     C  STVKEQLIQN LAPDEKQSRC .KITVVGVGN VGMACAISIL LKGLADELAL   49

Xenopus A  VDILEDKLKG EMMDLQHGSL FLKTPTIVAD KDYSVTANSR IVVVTGGVRQ  100
Xenopus B  VDILEDKLKG EVMDLQHGSL FLKTPTIVAD KDYSVTANSR IVVVTGGVRQ  100
Xenopus C  VDILEDKLKG EVMDLQHGSL FLKTPTIVAD KDYSVTANSR IVVVTGGVRQ  100
Pig     A  VDVMEDKLKG EMMDLQHGSL FLRTPKIVSG KDYNVTANSR LVVITAGARQ   99
Pig     B  VDVLEDKLKG EMMDLQHGSL FLQTPKIVAD KDYSVTANSK IVVVTAGVRQ  100
Rat     B  VDVLEDKLKG EMMDLQHGSL FLQTPKIVAD KDYSVTANSK IVVVTAGVRQ  100
Rat     C  VDADENKLKG EALDLLHGSL FLSTPKIVFG KDYSVSANSK LVIITAGARM   99

Xenopus A  QEGESRLNLV QRNVNIFKFI IPQIVKYSPD CIILVVSNPV DILTYVTWKL  150
Xenopus B  QEGESRLNLV QRNVNVFKFI IPQVVKYSPD CIIIVVSNPV DILTYVTWKL  150
Xenopus C  QEGESRLNLV QRNVNVFKFI IPQVVKYSPD CIIIVVSNPV DILTYVTWKL  150
Pig     A  QEGESRLNLV QRNVNIFKFI IPNIVKYSPN CKLLVVSNPV DILTYVAWKI  149
Pig     B  QEGESRLNLV QRNVNVFKFI IPQIVKYSPD CIIIVVSNPV DILTYVTWKL  150
Rat     B  QEGESRLNLV QRNVNVFKFI IPQIVKYSPD CTIIVVSNPV DILTYVTWKL  150
Rat     C  VSGESRLALL QRNVTSMKAI VPGVIQNSPD CKIMIVTNPV DILTYVVWKI  149

Xenopus A  SGLPQHRIIG SGTNLDSARF RHLIAEKLGV HPTSCHGFIL GEHGDSSVAV  200
Xenopus B  SGLPQHRIIG SGTNLDSARF RHLISEKLGV HPSSCHGFIL GEHGDTSVAV  200
Xenopus C  SGLPQHRIIG SGTNLDSARF RHLISEKLGV HPSSCHGFIL GEHGDTSVAV  200
Pig     A  SGFPKNRVIG SGCNLDSARF RYLMGERLGV HPLSCHGWIL GEHGDSSVPV  199
Pig     B  SGLPKHRVIG SGCNLDSARF RYLMAEKLGV HPSSCHGWIL GEHGDSSVAV  200
Rat     B  SGLPKHRVIG SGCNLDSARF RYLMAEKLGI HPSSCHGWIL GEHGDSSVAV  200
Rat     C  SGLPVSSVIG SGCNLDSARF RYLIGEKLGV NPSSCHGWVL GEHGDSSVPI  199

Xenopus A  WSGVNVAGVS LQSLKPDIGT DEDCCKWKEV HKQVVDSAYE VIKLKGYTNW  250
Xenopus B  WSGVNVAGVS LQSLKPEIGT DQDSCNWKEV HKKVVDSAYE VIKLKGYTNW  250
Xenopus C  WSGVNVAGVS LQSLKPEIGT DQDSCNWKEV HKKVVDSAYE VIKLKGYTNW  250
Pig     A  WSGVNVAGVS LKNLHPELGT DADKEHWKAV HKQVVDSAYE VIKLKGYTSW  249
Pig     B  WSGVNVAGVS LQELNPEMGT DNDSENWKEV HKMVVESAYE VIKLKGYTNW  250
Rat     B  WSGVNVAGVS LQELNPEMGT DNDSENWKEV HKMVVDSAYE VIKLKGYTNW  250
Rat     C  WSGVNIAGVT LKSLNPAIGS DSDKEQWKTV HKQVVDGGYE VLNLKGYTSW  249

Xenopus A  AIGFSVAEIV ESITKNLGRV HPVSTMVKGM YGIETEVFLS LPCVLNGNGL  300
Xenopus B  AIGFSVAEIV ESITKNLGRV HPVSTMVKGM YGIETEVFLS LPCVLNGNGL  300
Xenopus C  AIGFSVAEIV ESITKNLGRV HPVSTMVKGM YGIETEVFLS LPCVLNGNGL  300
Pig     A  AIGLSVADLA ESIMKNLRRV HPISTMIKGL YGIKEDVFLS VPCILGQNGI  299
Pig     B  AIGLSVADLI ESMLKNLSRI HPVSTMVQGM YGIENEVFLS LPCVLNARGL  300
Rat     B  AIGLSVADLI ESMLKNLSRI HPVSTMVKGM YGIENEVFLS LPCILNARGL  300
Rat     C  AIALSVTDIA ASILKNLKRV HAVTTLVKGL YGIKEEIFLS IPCVLGQSGI  299

Xenopus A  TSVINQKLKD NEVGQLQKSA ETLWSIQKDL KDL                   333
Xenopus B  TSVINQKLKD DEVGQLQKSA ETLWGIQKDL KDL                   333
Xenopus C  TSVISQKLKD DEVGQLQKSS ETLWGIQKDL QVL                   333
Pig     A  SDVVKVTLTP EEEAHLKKSA DTLWGIQKEL QF                    331
Pig     B  TSVINQKLKD DEVAQLKNSA DTLWGIQKDL KDL                   333
Rat     B  TSVINQKLKD DEVAQLRKSA DTLWDIQKDL KDL                   333
Rat     C  TDLVKVNMNT EEEALFKKSC DILWNIQKDL QL                    331
```

FIG. 2. Amino acid sequence comparison of *Xenopus* LDH-A, LDH-B, and LDH-C, pig LDH-A and LDH-B, and rat LDH-B and LDH-C subunits. Gaps at residue 17 of pig LDH-A and residue 21 of rat LDH-C are denoted by dots. The initiation Met is post-translationally removed in LDH isozymes.

over, in the 3′ noncoding sequences, *Xenopus* LDH-C cDNA is 53.3% and 40.8% different from the sequences of *Xenopus* LDH-A and LDH-B, respectively, while the sequence of *Xenopus* LDH-A is 39.6% different from that of LDH-B. These 3′ sequence results demonstrate that *Xenopus* LDH isozymes are indeed encoded by three different genes and are not simple alleles.

Pig LDH-A and LDH-B Sequences. From a screening of a pig muscle cDNA library using mouse LDH-B cDNA as a probe, 16 clones exhibited a strong hybridization signal, while 40 clones showed a weak hybridization signal. These 40 clones were rescreened with a mouse LDH-A cDNA probe, and 12 clones were isolated and partially characterized. Eight of these 12 clones contained inserts of about 1.7 kb, and the nucleotide sequences of two clones, pA117 and pA315, were determined. Both clones contained the same cDNA sequence except that the 5′ noncoding region of clone pA315 was 29 nucleotides shorter than that of clone pA117. The cDNA insert of clone pA117 contained an open reading frame of 996 nucleotides, 5′ (84 nucleotides) and 3′ (584 nucleotides) noncoding regions, and a poly(A) tail of 18 nucleotides. The deduced sequence of 331 amino acids (Fig. 2) was identical, except for one acid/amide difference at residue no. 232, to the previously published pig LDH-A sequence obtained by amino acid sequencing (12).

The cDNA inserts from 10 of 16 strongly positive clones were partially analyzed, and 8 of them were shown to contain cDNA inserts of about 1.4 kb. The nucleotide sequences of the inserted cDNAs from two strongly positive clones, pB12 and pB15, are identical and contain an open reading frame of 1002 nucleotides, the 5′ (94 nucleotides) and 3′ (201 nucleotides) noncoding regions, and a poly(A) tail of 20 nucleotides. The deduced sequence of 333 amino acids (Fig. 2) was identical to the previously published sequence of pig LDH-B isozyme determined by direct protein sequencing (12), except five acid/amide differences at residue nos. 13, 14, 80, 130, and 213, probably due to deamination during peptide purification and sequencing.

Rat LDH-B and LDH-C Sequences. Two LDH-B cDNAs were identified from a rat heart cDNA library when mouse LDH-B cDNA was used as a probe, and their 5′ and 3′ nucleotide sequences were partially determined. Clone rB217 contained a nearly full-length LDH-B cDNA insert, and its complete nucleotide sequence contained an open reading frame of 1002 nucleotides, 5′ (18 nucleotides) and 3′ (182 nucleotides) noncoding regions, and a poly(A) tail of 25 nucleotides. The protein-encoding sequence of rat LDH-B cDNA differed from that of mouse LDH-B cDNA (6) by 64 nucleotides. However, these differences result in only two amino acid replacements, at residues 10 and 16 within the variable amino-terminal arm of the LDH molecule (4).

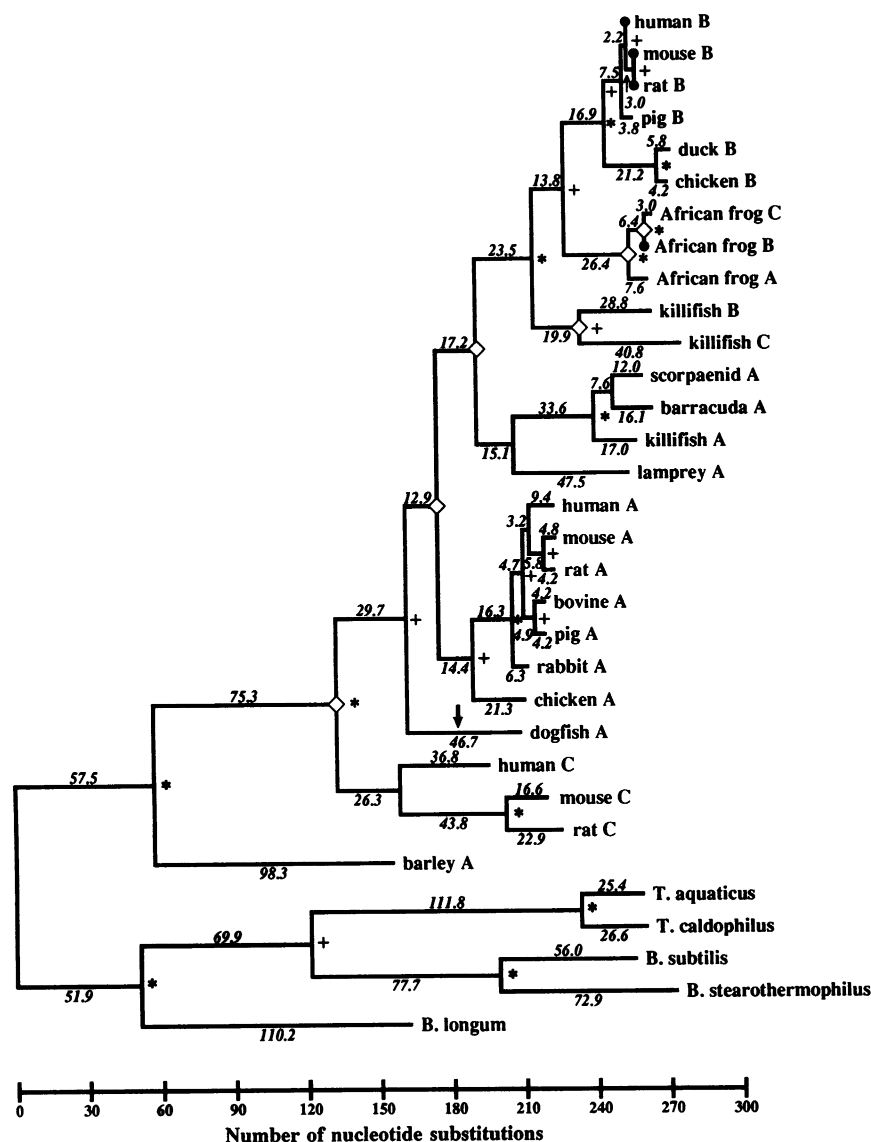

FIG. 3. Evolutionary tree of LDH subunits. The most parsimonious tree of 32 LDH amino acid sequences is presented. The numbers on the branches are nucleotide substitutions required to account for the amino acid replacements. Bootstrap support is shown by an asterisk (99–100%) or a plus sign (80–98%) at nodes representing the cenancestor (the most recent ancestor common to all members) of the clade. Nodes with an open diamond denote required gene duplications. The arrow shows where the lamprey sequence can be moved to produce a tree requiring only one additional substitution. The length of the tree is 1574 substitutions. Two types of vertebrate LDH subunits present in somatic tissues are denoted as A (muscle) and B (heart or liver), while the third type of LDH subunit, found in mammalian testis and *Xenopus* oocyte, is denoted as LDH-C.

Three LDH-C cDNA clones were isolated from a rat testis cDNA library when mouse LDH-C cDNA was used as a probe, and their nucleotide sequences were determined. The cDNA insert of clone rC105 contained an open reading frame of 996 nucleotides, and 5′ (93 nucleotides) and 3′ (46 nucleotides) noncoding regions. The deduced amino acid sequence of rat LDH-C (Fig. 2) is consistent with the earlier data obtained from direct protein sequencing, although several errors were found in the tentative amino acid sequence based on homology (13). The 3′ noncoding region of the rat LDH-C cDNA clone contains a putative polyadenylylation signal, AATAAA, but its poly(A) tail was truncated. The 3′ noncoding sequence of the rat LDH-C cDNA contains 16 nucleotide differences and a deletion of 45 nucleotides compared with that of mouse (29).

Evolutionary Tree. The phylogenetic relationships among amino acid sequences of 32 LDH subunits from mammals, birds, amphibian, fish, barley, and bacteria were analyzed, and the most parsimonious tree, along with bootstrap and gene duplication information, is presented in Fig. 3.

DISCUSSION

Phylogeny. The most parsimonious tree, shown in Fig. 3, is generally consistent with many previous molecular analyses of LDH, including those of Li *et al.* (4), Crawford *et al.* (5), Hiraoka *et al.* (6), and Quattro *et al.* (9). This includes the mammalian LDH-C gene arising from the earliest vertebrate gene duplication. That duplication is supported by the next node down (which is not a duplication), which is the cenancestor (the most recent common ancestor) of all LDHs being studied except the mammalian LDH-Cs. It has 88% bootstrap support. While one might wish it were higher, it is within the range (>75%) where Hillis and Bull (43) have shown, for simulated data, that the probability of getting the correct tree is substantially higher than the bootstrap value.

The tree corroborates the conclusion of Quattro *et al.* (9) that the killifish C is a duplicate of the B form, the duplication occurring after the divergence of the tetrapods and fish.

The tree does not support mammalian LDH-C as the result of a duplication of LDH-A after A's divergence from B as suggested by Millan *et al.* (7) and Hendriks *et al.* (8), nor does it support the alternative that LDH-C arose from LDH-B after B's divergence from A as suggested by Holmes (3) and Markert *et al.* (2).

The major clades are well supported. Twelve of the 29 clades are supported at 99% or better and another 11 at 80% or better. There are three relations of importance that are not well supported. One is the joining of the lamprey LDH to the fish LDH-A (bootstrap value = 24%). The second is the relation between the fish A and the LDH-B group (bootstrap value = 40%). This second one involves a gene duplication. The third is the preceding group, LDH-B plus fish A, to the mammalian LDH-A (bootstrap value = 11%) which also involves a gene

duplication. Taken collectively, the uncertainty would permit an alternative relationship that would permit the order of divergence among the LDH-A sequences to be lamprey, dogfish, teleosts, birds, and mammals, a conventional arrangement. That would save one gene duplication but at high cost. That tree requires an additional 11 replacement substitutions, and there was only 18% bootstrap support for the lamprey in that position, only 21% support for the dogfish next, and only 2% support for the fish A with the mammal–bird A.

While gene duplications seem to occur often enough for us not to be put off by an extra gene duplication (6 vs. 5; see next section), there are other problems with the most parsimonious tree. If indeed lamprey has only one LDH gene, but there were three earlier gene duplications (see the diamonds on Fig. 3), then the lamprey must have lost three of its four copies. *Xenopus* has six duplications in its history, and it must have lost four of the genes if there are no other LDH genes to be recovered. Similar arguments apply to every organism in the tree. Even more disconcerting, yet possibly true, is the implication that three of the six gene duplications occurred prior to the divergence of the vertebrates as shown by the lamprey's connection to the other members in the tree. This suggests that examining higher invertebrates might be illuminating.

On the other hand, the second most parsimonious tree is only one replacement substitution longer and is interesting in that it relocates the lamprey LDH to join the dogfish as its sister taxon at the position of the arrow on Fig. 3. This does not solve all the problems just mentioned, but it does mean that only one of the gene duplications had to have occurred prior to the divergence of the agnathans and the other fish. This reduces considerably the number of gene losses, and so we regard it as the more correct tree.

Gene Duplications. There are six duplications in the tree. There must be at least two duplications simply to account for the fact that humans, mice, rats, African frogs, and killifish each have three genes. That lower bound is unrealistic for several reasons. The three African frog sequences are so close to each other that we need two just for them. We need a third for the killifish B and C sequences. We need a fourth to join the fish A group containing the other two killifish and we need a fifth to account for the mammalian LDH-Cs. The only doubtful duplication is the one accounting for the divergence of the mammalian A and B LDHs, the one discussed in the preceding section.

To avoid the sixth duplication requires placing the fish A divergence after the dogfish divergence and the lamprey divergence before it. This costs too much, and so we believe a sixth is required even if some minor details of the tree prove different from those shown.

Nomenclature of LDHs. As more and more sequences are determined, there is an accompanying increase in the number of gene duplications required, and there are probably others awaiting discovery. This means that the suffixes A, B, and C cannot connote three paralogous sets of orthologous genes. The three African frog genes sequenced here make that particularly clear. If the A, B, and C suffixes are to be meaningful, then their application must not be according to the designation already applied to its closest known relative. We suggest that, until the tissue of major expression is known, no letter be assigned and the letter A be reserved for the case where the tissue of major expression is muscle, B be reserved for heart and/or liver, and C be reserved for gonadal tissue. If other LDH genes are found expressed mainly in some other tissue, their designation may best be decided at that time. But, in accordance with that philosophy, we have here reversed the assignments of B and C given by Wolff and Kobel (40) to *Xenopus* LDH.

We thank Profs. G. N. Somero and D. A. Powers for providing the unpublished amino acid sequences of three fish LDH-A subunits, Ms. Helene Van for computer analyses and figure drawing, Drs. F. M. Johnson and J. E. Welch for reading the manuscript, and Ms. N. Gore for preparing it. This investigation was supported in part by the Ministry of Education, Science, and Culture of Japan (S.T.) and Fulbright Scholar Program (M.A.Q.). This work was partially supported by National Science Foundation Grant DEB9096152 to W.M.F.

1. Holbrook, J. J., Liljas, A., Steindel, S. J. & Rosmann, M. G. (1975) in *The Enzymes: Oxidation-Reduction*, eds. Boyer, P. (Academic, New York), Vol. 11, Part A, pp. 191–192.
2. Markert, C. L., Shaklee, J. B. & Whitt, G. S. (1975) *Science* **189**, 102–114.
3. Holmes, R. S. (1972) *FEBS Lett.* **28**, 51–55.
4. Li, S. S.-L., Fitch, W. M., Pan, Y.-C. E. & Sharief, F. S. (1983) *J. Biol. Chem.* **258**, 7029–7032.
5. Crawford, D. L., Consteantino, H. R. & Powers, D. A. (1989) *Mol. Biol. Evol.* **6**, 369–383.
6. Hiraoka, B. Y., Sharief, F. S., Yang, Y.-W., Li, W.-H. & Li, S. S.-L. (1990) *Eur. J. Biochem.* **189**, 215–220.
7. Millan, J. L., Driscoll, C. E., LeVan, K. M. & Goldberg, E. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5311–5315.
8. Hendriks, W., Mulders, J. W. M., Bibby, M. A., Slingsby, C., Bloemendal, H. & de Jong, W. W. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7114–7118.
9. Quattro, J. M., Woods, H. A. & Powers, D. A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 242–246.
10. Burglin, T. R., Mattaj, I. W., Newmeyer, D. D., Zeller, R. & De Robertis, E. M. (1987) *Genes Dev.* **1**, 97–107.
11. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
12. Kiltz, H. H., Keil, W., Griesbach, M., Petry, K. & Meyer, H. (1977) *Hoppe Seyler Z. Physiol. Chem.* **358**, 123–127.
13. Pan, Y.-C. E., Sharief, F. S., Okabe, M., Huang, S. & Li, S. S.-L. (1983) *J. Biol. Chem.* **258**, 7005–7016.
14. Tsujibo, H., Tiano, H. F. & Li, S. S.-L. (1985) *Eur. J. Biochem.* **147**, 9–15.
15. Chang, S.-M. T., Lee, C.-Y. & Li, S. S.-L. (1979) *Biochem. Genet.* **17**, 715–729.
16. Li, S. S.-L., Tiano, H. F., Fukasawa, K. M., Nakashima, Y. & Pan, Y.-C. E. (1985) *Eur. J. Biochem.* **149**, 215–225.
17. Fukasawa, K. M. & Li, S. S.-L. (1987) *Genetics* **116**, 99–105.
18. Matrisian, L. M., Rautmann, G., Magun, B. E. & Breathnach, R. (1985) *Nucleic Acids Res.* **13**, 711–726.
19. Ishiguro, N., Osame, S., Kagiya, R., Ichijo, S. & Shinagawa, M. (1990) *Gene* **91**, 281–285.
20. Sass, C., Briand, M., Benslimane, S., Renaud, M. & Briand, Y. (1989) *J. Biol. Chem.* **264**, 4076–4081.
21. Hirota, Y., Katsumata, A. & Takeya, T. (1990) *Nucleic Acids Res.* **18**, 6432.
22. Taylor, S. S. (1977) *J. Biol. Chem.* **252**, 1799–1806.
23. Stock, D. W. & Whitt, G. S. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1799–1803.
24. Sakai, I., Sharief, F. S., Pan, Y.-C. E. & Li, S. S.-L. (1987) *Biochem. J.* **248**, 933–936.
25. Takano, T. & Li, S. S.-L. (1989) *Biochem. J.* **257**, 921–924.
26. Torff, H. J., Becker, D. & Schwarzwalder, J. (1977) in *Pyridine Nucleotide Dependent Dehydrogenase*, ed. Sund, H. (de Gruyter, Berlin), pp. 31–42.
27. Takano, T. & Li, S. S.-L. (1989) *Biochem. Biophys. Res. Commun.* **159**, 579–583.
28. Wu, K. C. & Li, S. S.-L. (1990) *J. Genet. Mol. Biol.* **1**, 72–76.
29. Sakai, I., Sharief, F. S. & Li, S. S.-L. (1987) *Biochem. J.* **242**, 619–622.
30. Wu, K. C., Chan, K., Lee, C.-Y. G. & Lau, Y.-F. (1987) *Biochem. Biophys. Res. Commun.* **146**, 964–970.
31. Hondred, D. & Hanson, A. D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7300–7304.
32. Barstow, D. A., Clarke, A. R., Chia, W. N., Wigley, D., Sharman, A. F., Holbrook, J. J., Atkinson, T. & Minton, N. P. (1986) *Gene* **46**, 47–55.
33. Wirz, B., Suter, F. & Zuber, H. (1983) *Hoppe Seyler Z. Physiol. Chem.* **363**, 893–909.
34. Hediger, M. A., Frank, G. & Zuber, H. (1986) *Biol. Chem. Hoppe Seyler* **367**, 891–903.
35. Minowa, T., Iwata, S., Sakai, H., Masaki, H. & Ohta, T. (1989) *Gene* **85**, 161–168.
36. Ono, M., Matsuzawa, H. & Ohta, T. (1990) *J. Biochem. (Tokyo)* **107**, 21–26.
37. Kunai, K., Machida, M., Matsuzawa, H. & Ohta, T. (1986) *Eur. J. Biochem.* **160**, 433–440.
38. Fitch, W. M. (1971) *Syst. Zool.* **20**, 406–416.
39. Fitch, W. M. & Farris, J. S. (1974) *J. Mol. Evol.* **3**, 263–278.
40. Wolff, J. & Kobel, H. R. (1982) *J. Exp. Zool.* **223**, 203–210.
41. Hornby, J. E., Munitz, L. & Gough, P. (1989) *Comp. Biochem. Physiol.* **93B**, 85–92.
42. Graf, J.-D. (1989) *Genetics* **123**, 389–398.
43. Hillis, D. M. & Bull, J. J. (1993) *Syst. Biol.* **42**, 182–192.