

# Evolution of the Vertebrate Cytosolic Malate Dehydrogenase Gene Family: Duplication and Divergence in Actinopterygian Fish

T.J.S. Merrit,\* J.M. Quattro

Department of Biological Sciences, Program in Marine Science, Baruch Institute and School of the Environment, University of South Carolina, Columbia, SC 29208, USA

Received: 10 June 2002 / Accepted: 2 October 2002

**Abstract.** A general correlation between neural expression and negative charge in isozymes suggests charge represents an adaptation to the neural environment. Interestingly, a notable exception exists in teleost fish. Two cytosolic malate dehydrogenase (MDH) isozymes have different spatial expression patterns in certain fishes: one is expressed in all tissues and the second is expressed primarily in the eye and skeletal muscle. While the neural MDH isozyme is negatively charged, the difference in charge between the two isozymes is not as pronounced as that observed in other gene families (e.g., triosephosphate isomerase and lactate dehydrogenase). Most tetrapods express a single cytosolic MDH isozyme, and it has been demonstrated recently that the pair of isozymes found in teleosts results from a gene duplication sometime after the separation of teleosts and tetrapods, although the exact timing of this duplication has not been inferred. Phylogenetic analyses suggest that the duplication of teleost isozymes occurred during the radiation of actinopterygian fish, consistent with the timing of duplication at other loci. Using inferred amino acid sequences, we examine the

pattern of change following the duplication and across the rest of the MDH gene tree. Comparison between the MDH gene family and another gene family that shows a larger charge differential among members (triosephosphate isomerase) indicates that the smaller charge difference between MDH isozymes is best explained by greater constraint on amino acid change directly following the duplication, not greater constraint across the entire gene tree. This difference in constraint might result from the wider pattern of expression of the “neural” MDH isozyme.

**Key words:** Malate dehydrogenase — Gene duplication — Neural expression — Positive selection — Ancestral reconstruction — Teleost fish

## Introduction

Isozymes, tissue-specific enzymes, are examples of protein specialization to unique environments and functions within an organism. Many vertebrate gene families include a member with an unusual electrophoretic phenotype: neurally expressed isozymes are frequently negatively charged (e.g., Fisher et al. 1980; Merritt and Quattro 2001). This correlation between neural expression and net negative charge suggests that this phenotype might represent an adaptation to the neural environment (Fisher et al. 1980). Recently, we have shown that the high net negative charge of the neurally expressed triosephosphate isomerase isozyme characteristic of the

---

Sequences reported in this paper have been submitted to GenBank (accession numbers AY161291–AY161295)

\**Current address:* Department of Ecology and Evolution, State University of New York—Stony Brook, Stony Brook, NY 11794, USA

*Correspondence to:* Joseph M. Quattro, Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA; email: quattro@mail.biol.sc.edu

teleost retina (TPI-A) resulted from the biased accumulation of charged amino acids during a period of directional (positive) selection (Merritt and Quattro 2001). In contrast, the neural isozymes of some other gene families do not exhibit as large a net negative charge as that characterizing TPI-A (e.g., see Morizot and Schmidt 1990). By comparing the evolution of these contrasting examples, we can further understand the evolution, and perhaps function, of negative charge concomitant with neural expression.

Here we examine the evolution of a neurally expressed, but not highly negatively charged, isozyme of malate dehydrogenase (MDH; EC 1.1.1.37; L-malate:NAD oxidoreductase). MDH is a dimeric protein that functions in the citric acid cycle to catalyze the interconversion of malate and oxaloacetate. Teleost fish express two cytosolic MDH isozymes (Wheat et al. 1973; Rainboth and Whitt 1974; Fisher et al. 1980). One isozyme, MDH-A, is expressed in all tissues across various fish taxa, whereas the second isozyme, MDH-B, is expressed predominantly in the eye and skeletal muscle (Wheat et al. 1971). Given the pronounced expression of MDH-B in the eye, we, and others, refer to MDH-B as the “neural” MDH, although its pattern of expression is not as canalized as that of some other neural isozymes (e.g., TPI-A and LDH-C).

The two teleost cytosolic MDH proteins are the products of two unlinked loci (Wheat et al. 1973) and are distinct biochemically, differing in thermostability, thermal optima, and reaction kinetics (Schwantes and Schwantes 1982; Lin and Somero 1995; Lin et al. 2002). The two isozymes also differ in the timing of their expression during development: *Mdh-A* appears to be expressed throughout development, while *Mdh-B* appears to be first expressed concomitantly with development of the eye (Philipp et al. 1979) or ontogeny of muscle (Champion and Whitt 1976). Electrophoretic separation indicates that both MDH isozymes are relatively neutrally charged, although MDH-B does have a slightly more negative charge than does MDH-A (e.g., Wheat et al. 1971; Philipp et al. 1979). This difference in charge is not as pronounced as between some neural and nonneural isozymes in other gene families (e.g., triosephosphate isomerase [TPI] [Pontier and Hart 1981; Merritt and Quattro 2001] and lactate dehydrogenase [LDH] [Markert et al. 1975; Quattro et al. 1993]). An MDH isozyme restricted in expression to mitochondria is also present in all vertebrate taxa (Schwantes and Schwantes 1982; Wheat et al. 1971) but is not examined here.

Unlike teleost fish, invertebrates (Meizel and Markert 1967; Markert and Whitt 1968) and most tetrapods (Thorne and Cooper 1963; Davidson and Cortner 1967; Karig and Wilson 1971) express a single cytosolic MDH, suggesting that the two teleost

cytosolic MDH isozymes result from duplication after the divergence of teleosts and tetrapods. However, Fisher et al. (1980) report a second cytosolic MDH isozyme in cephalochordates and place the duplication of cytosolic MDH early in chordate evolution. Similarly, Schwantes and Schwantes (1982) report a second cytosolic MDH in some amphibia. They suggest that the pair of isozymes in teleost fish and amphibians is orthologous and results from duplication prior to the divergence of these two lineages. On the other hand, Philipp et al. (1983) report only a single cytosolic MDH in amphibians, which is consistent with the duplication of MDH early in the evolution of teleost fish. Recently, Lin et al. (2002) described both cytosolic MDH protein sequences from barracuda (*Sphyræna idiasstes*) and suggested that the gene duplication event most likely occurred after the separation of teleosts and tetrapods. However, barracuda are a recently derived group of fish, and without the inclusion of MDH sequences from more basal fish taxa in phylogenetic analyses, it is not possible to determine whether the duplication occurred early in actinopterygian evolution, early in teleost evolution, or late in teleost evolution. Pinpointing the duplication event is necessary to establish the orthology or paralogy of the various vertebrate (tetrapod and teleost) isozymes and is an essential first step in the investigation of their evolution. Further, the context of the duplication might shed light on the cause of the small charge differential observed between MDH isozymes. If this duplication occurred more recently in teleost evolution than those of the neural LDH or TPI, comparisons among these gene families might not be appropriate—sufficient time might not have elapsed for a more pronounced difference in charge to have evolved between MDH isozymes of fish relative to that observed electrophoretically and by inference in TPI and LDH.

Similarly, a well-supported gene tree is necessary to investigate the role of selection in the evolution of the MDH neural isozyme. Many gene families evolve episodically, with bursts of positive selection bracketed by periods of purifying selection (e.g., Messier and Stewart 1997; Zhang et al. 1998; Merritt and Quattro 2001). The period of positive selection following gene duplication might be brief and evidence for positive selection might be obscured by subsequent substitutions after long periods of evolutionary time (Hughes 1994). For this reason, comparison of modern gene sequences might not detect evidence for positive selection. Comparisons of inferred ancestral sequences, internal nodes on a gene tree that bracket a period of interest, circumvent this problem by isolating periods of evolutionary time (e.g., Messier and Stewart 1997; Zhang et al. 1998).

This paper investigates the evolution of the pair of cytosolic *Mdh* genes in fish. Five cytosolic MDH

cDNAs were sequenced from three modern Actinopterygian fish: a single cDNA from the shortnose sturgeon (*Acipenser brevirostrum*), a primitive actinopterygian; and two cDNAs each from two teleost fish species, zebrafish (*Danio rerio*) and Japanese medaka (*Oryzias latipes*). These sequences were chosen to flank closely the most likely location of the duplication that led to the pair of teleost cytosolic MDH isozymes. This paper has two main objectives: (1) to characterize more precisely the duplication event that gave rise to the two cytosolic MDH isozymes in teleost fish and (2) to compare and contrast the evolution of the MDH and TPI neural isozymes to investigate a possible explanation for the lack of a large charge differential between the MDH isozymes.

## Methods

**MDH Cloning and Sequencing.** *Acipenser brevirostrum* (shortnose sturgeon) tissues were obtained from R. Chapman, South Carolina (SC) Department of Natural Resources, Charleston; *Oryzias latipes* (Japanese medaka) samples were purchased from Carolina Biological Supply; and *Danio rerio* (zebrafish) samples were purchased from local pet stores. Tissues were dissected from single fish and processed immediately or stored at  $-70^{\circ}\text{C}$ . Total RNA was purified from tissue using a commercially available kit (RNeasy; Qiagen). Complementary DNA (cDNA) was synthesized from total RNA using the Superscript Preamplification System (Gibco BRL).

At the time this study was undertaken, complete and partial sequences for cytosolic malate dehydrogenase (MDH) were available from a number of vertebrate taxa, including partial sequences from two teleost fish, zebrafish and Japanese medaka. The partial teleost sequences were completed and then used, along with other vertebrate MDH sequences, to design degenerate oligonucleotide primers to amplify the second MDH locus from the two teleost species and a single locus from sturgeon. This strategy is outlined briefly below.

Two MDH expressed sequence tags (ESTs) from zebrafish, one from the 5' end and one from the 3' end of the cDNA (GenBank accession numbers AA566495 and AI477368, respectively), were used to design specific oligonucleotide primers.

DMDHF: 5'-TTG GCG CAC ATC TCA ACC TGC-3'

DMDHR: 5'-AGT TTG ACT CTA GCG TCT GCC-3'

These primers amplified a single 1147-base pair (bp) product from zebrafish liver cDNA by the polymerase chain reaction (PCR) (Saiki et al. 1988). PCR was carried out for 40 cycles under the following conditions: denaturation at  $95^{\circ}\text{C}$  for 1 min, annealing at  $48^{\circ}\text{C}$  for 1 min, and extension at  $72^{\circ}\text{C}$  for 1 min. PCR products were cloned into pGEM T-vector (Promega) and sequenced using a 377 automated sequencer (ABI). Dye terminator sequencing was performed with BigDye termination mix (Applied Biosystems) using the manufacturer's protocol. At least three independent clones per PCR fragment were sequenced on both strands. The extreme 5' and 3' coding and untranslated regions of the cDNA were amplified using the RACE (rapid amplification of cDNA ends) method (Frohman 1990). RACE amplifications used gene-specific primers designed from the initial gene fragment. Conditions for the PCR, cloning, and sequencing were as described above.

GenBank also contains a partial MDH sequence from the Japanese medaka (accession number AJ012193). This sequence is missing the 5' end of the coding region and the 5' untranslated region. The RACE method was used to amplify the 5' end of the

cDNA using primers designed from this partial MDH sequence and conditions as described for the zebrafish MDH cDNA above. A set of degenerate primers was designed using an alignment of the zebrafish and medaka MDH sequences and other vertebrate MDH sequences from GenBank.

FishMDH47F: 5'-CTS CTG GAY ATC MAY CCC ATG-3'

FishMDH55F: 5'-CTG CCG GTG CTG GAC GGA GTS GTC-3'

FishMDH299R: 5'-TTC CAG STC TTG TTC TTG AT-3'

FishMDH308R: 5'-TCG TTG ATR GRG AGM CCR TCA AC-3'

Degenerate positions are represented by the following ambiguity codes: M = A, C; R = A, G; S = C, G; and Y = C, T. Numbers refer to the amino acid position occupied by the 3' base of each primer using the mouse sequence (M29462) as reference.

Combinations of these four oligonucleotide primers were used to amplify a 755- to 803-bp segment of the MDH gene from shortnose sturgeon, zebrafish, and medaka by PCR with conditions as described above. In some cases nested reamplifications were necessary to yield sufficient PCR product for cloning. In these cases, an amplification using the primers FishMdh47F and FishMdh308R was diluted 1:100 and used as template in a reaction using the primers FishMdh55F and FishMdh299R under identical reaction conditions. The extreme 5' and 3' coding and untranslated regions of these genes were amplified, cloned, and sequenced using the RACE method, with PCR conditions as described above.

Combinations of these primers amplified segments of two distinct zebrafish MDH cDNAs but only one medaka cDNA (see Results). Another set of forward primers was therefore designed around amino acid substitutions unique to the second zebrafish locus.

MDHB 141F: 5'-ACH AAC TGY YTR ATT GCM G-3'

MDHB 259F: 5'-CAC ATG AGR GAY ATY TGG AC-3'

H = A, C, T and degenerate positions and numbering are as above. These primers were paired with a new reverse primer designed from the five fish MDH cDNA sequences already characterized.

MDH 320R: 5'-GCN GTR TCY CTY TCY TCM AC-3'

Degenerate sites and numbering are as above. A PCR amplification combining MDHB 141F and MDH 320R was performed, but failed to yield enough product for cloning. This product was diluted 1:100 and used as the template in a partially nested PCR reaction using MDHB 259F and MDH 320R. This reaction amplified a 225-bp segment of a second MDH gene from medaka. The extreme 5' and 3' coding and untranslated regions of this gene were amplified, cloned, and sequenced using the RACE method, with conditions as described above.

Isoelectric point (pI) values were calculated for each predicted amino acid sequence using the on-line Isoelectric Point service of EMBL (<http://www.embl-heidelberg.de/cgi/pi-wrapper.pl>).

**Phylogenetic Analyses.** In addition to the five fish MDH sequences reported here, the following cytosolic MDH sequences were obtained from GenBank and used in phylogenetic analyses: barracuda MDH-S (AF390559), barracuda MDH-L (AF390560), pig (U44846), human (U20352), rat (AF093773), mouse (M29462), alfalfa (AF020272), iceplant (X96539), and snail (AF218065).

Amino acid sequences were aligned using the ClustalW multiple alignment program (Thompson et al. 1994). Minor adjustments were made "by eye," resulting in an overall alignment of 333 amino acid positions. A phylogeny was constructed by the neighbor-joining (NJ) method (Saitou and Nei 1987) using the algorithm implemented in MEGA (version 1.0 [Kumar et al. 1993]). Pairwise gamma distances were calculated according to the formula of Nei et al. (1976), with the  $\alpha$  parameter set to 2.05 (Uzzel and Corbin 1971), and the complete deletion option for

gaps and missing data. Bootstrapping (Felsenstein 1985) was used to evaluate the degree of support for particular groupings in the NJ analysis. Maximum likelihood analyses were carried out using the PROTML program included in the MOLPHY version 2.2 software package (Adachi and Hasegawa 1994) using the Jones, Taylor, and Thornton (JTT; Jones et al. 1992) model of amino acid substitution.

**Test for Positive Selection.** Ancestral sequences were reconstructed using a distance-based method (Zhang et al. 1997) by first inferring the amino acid sequences and then inferring the nucleotide sequences under the restriction of the ancestral amino acid sequences. Ancestral amino acid reconstructions used the ANCGENE computer program (Zhang et al. 1998) with the JTT (Jones et al. 1992) model of amino acid substitution. The second-to-last amino acid in each sequence could not be inferred because of gaps in our alignment of extant MDH sequences. The reconstructed sequences were used to compute the number of synonymous (*s*) and nonsynonymous (*n*) substitutions per branch using proportional differences as implemented in the BN-BS computer program (Zhang et al. 1998). Reconstruction of ancestral sequences and estimation of the number of substitutions using a Poisson model of amino acid substitution, the Jukes–Cantor (1969) model of distance estimation, or a combination of either of these methods with the JTT model or proportional distances yielded results essentially identical to those reported here. The numbers of potential nonsynonymous sites (*N*) and synonymous sites (*S*) were also calculated for the MDH sequences using the BN-BS program. The ratio of transitions to transversions, required for calculation of *s*, *n*, *S*, and *N*, was calculated using MEGA (version 1.0 [Kumar et al. 1993]). A major goal of this research was to compare the patterns of DNA and amino acid change across evolutionary history between the TPI and the MDH gene families. To allow direct comparison between the two gene families, only sequences for which analogues were available for both TPI and MDH (e.g., cyprinid and atherinid neural and nonneural MDH but not the pair of MDH isozymes from barracuda) are included in the ancestral reconstruction, test for positive selection, and other tests of DNA and amino acid change presented here. Inclusion of the barracuda sequences in the reconstruction and test of positive selection for MDH only gave essentially identical results to those presented here (data not shown).

## Results

**Cloning and Sequencing.** Amplification from zebrafish liver cDNA using the DMDH-F and DMDH-R primers yielded a 1147-bp product. RACE amplification, cloning, and sequencing of the extreme 5' and 3' ends and cloning and sequencing of the entire coding region indicated a cDNA consisting of 1348 bp, including a start codon, a TGA stop codon, 50 bp of 5' untranslated region (UTR), and 295 bp of 3' UTR. The cDNA open reading frame codes for a 333-amino acid protein with a predicted pI of 7.2 (Fig. 1). This sequence, referred to hereafter as *DrMDH-A*, is similar to other MDH sequences in GenBank and, across the region of overlap, identical to the zebrafish MDH ESTs in GenBank.

Amplification from medaka liver cDNA produced a single product consisting of 1383 bp, including 69 bp of 5' untranslated sequence, a start codon, a TGA stop codon and 311 bp of 3' untranslated sequence. The cDNA open reading frame coded for a 333-

amino acid protein with a predicted pI of 7.7 (Fig. 1). The partial medaka MDH sequence in GenBank contained a few minor inconsistencies compared to MDH sequences in GenBank: an apparent deletion of nucleotides 397, 398, and 399 (using the mouse MDH as reference), an apparent single base insertion at position 431, and an apparent single base deletion at position 449. In comparison with other MDH sequences, deletion of nucleotides 397–399 removes a glutamine from the amino acid sequence, the insertion at position 431, followed by the deletion at position 449, alters, and then recovers, the open reading frame of the sequence. All of the clones we examined contained the three nucleotides at that same position missing from the partial medaka sequence (nucleotides 397, 398, and 399) and had neither the insertion at position 431 nor the deletion at position 449, suggesting that the unusual features of the partial medaka sequence in GenBank are artifacts. With these exceptions, the medaka sequence reported here, referred to hereafter as *O/MDH-A*, was identical to the partial medaka sequence in GenBank and similar to other MDH sequences reported in GenBank.

PCR products from zebrafish eye cDNA amplified using the “FishMDH” primers yielded two products: one that was identical to the sequence amplified from liver samples (*DrMDH-A*) and one that differed from this sequence by approximately 20%. Cloning and sequencing of the 5' and 3' ends of the second zebrafish MDH coding and untranslated regions indicated a cDNA consisting of 1615 bp, including an ATG start codon, a TGA stop codon, 96 bp of 5' untranslated sequence, and 517 nucleotides of 3' untranslated sequence. The open reading frame of this second zebrafish cDNA codes for a protein of 333 residues with a predicted pI of 6.0 (Fig. 1). This sequence, hereafter referred to as *DrMDH-B*, is similar to *DrMDH-A* and other MDH sequences in GenBank.

PCR amplification from medaka eye cDNA using the FishMDH primers yielded only a single product, identical to the product amplified from medaka liver, *O/MDH-A*. PCR amplification from medaka eye samples using the “MDHB” primers yielded a single PCR product that differed by approximately 20% from *O/MDH-A*. Subsequent cloning and sequencing of the 5' and 3' ends and coding and untranslated regions indicates a cDNA consisting of 1346 bp, including a start codon, a TGA stop codon, 33 bp of 5' UTR, and 311 bp of 3' UTR. This second cDNA codes for a 333-amino acid protein with a predicted pI of 6.0 (Fig. 1). This sequence, hereafter referred to as *O/MDH-B*, is similar to *O/MDH-A* and other MDH sequences in GenBank.

To confirm that the cDNAs we characterized corresponded to the tissue-specific MDH isozymes reported by earlier authors (and thus our assignment

<i>AbMDH</i>	MSDFIRVLVT	GAAGQIAYSL	LYNIAKGDVF	GKDQPLVLVL	LDITPMLPVL	DGVVMELQDC	60
<b>MDH-T</b>	.AE.....	.....S.....	.....II...	.....	.....	.....	
<i>DrMDH-A</i>	.AE.....	.....S.....	.....II...	.....	.....	.....	
<i>O/MDH-A</i>	.AE...V..	.....FS.....	.....II.I..	...P.....	.....	.....	
<b>MDH-A</b>	.AE.....	.....S.....	.....II...	.....	.....	.....	
<i>DrMDH-B</i>	.E.....	.....G.....	.....L.....	.....E.....	.....	.....	
<i>O/MDH-B</i>	.AE..S....	.....FS.....	.....II.L..	.....E.....	.....	.....	
<b>MDH-B</b>	.AE.....	.....S.....	.....II.L..	.....E.....	.....	.....	
<i>AbMDH</i>	ALPLLREVIP	TDKEEVAFKD	LDAAILVGSM	PRKEGMERKD	LLKANVKIFK	SQGAALDKYA	120
<b>MDH-T</b>	.....	.....	.....	.....A...	...E.....	.....	
<i>DrMDH-A</i>	.....	...V..G...	.....	.....A...	T..E..E...	.....	
<i>O/MDH-A</i>	.....	...V..G...I	.....	.....A...	T..Q.....	.....	
<b>MDH-A</b>	.....	...V..G...	.....	.....A...	T..E.....	.....	
<i>DrMDH-B</i>	.....	...D.....	...R.....	.....A...	...E.....	.....	
<i>O/MDH-B</i>	.....A	.....	.....	.....A..R	.....	.....	
<b>MDH-B</b>	.....	.....	.....	.....A...	...E.....	.....	
<i>AbMDH</i>	KKTVKVLVVG	NPANTNCLIA	SKSAPSIPKE	NFSCLTRLDH	NRARSQVAIR	VGVPADNVKD	180
<b>MDH-T</b>	.....	.....	.....	.....M.	C.....N	.....	
<i>DrMDH-A</i>	.....	.....	.....	.....M.	...S..S..N	.....	
<i>O/MDH-A</i>	.....I...	.....	.....	...C...M.	C...SS.K..N	.....	
<b>MDH-A</b>	.....	.....	.....	.....M.	C...S...N	.....	
<i>DrMDH-B</i>	.....	A.....	.....	...S...L.	C.IAPN...N	.....	
<i>O/MDH-B</i>	...H.....	A.....	.....	...C...M.	C...TH..N	.....	
<b>MDH-B</b>	.....	A.....	.....	...C...M.	C...N...N	.....	
<i>AbMDH</i>	VIIWGNHSST	QYPDVHHAHV	KVQGKEIGAY	EAVKDDGWLK	GDFISTVQQR	GAIVIKARKL	240
<b>MDH-T</b>	.....	N.....A.F	D.....S...	.....	.....	.....	
<i>DrMDH-A</i>	.T.....	.....I..	TRN...A.F	D..N.ES...	.....	.....	
<i>O/MDH-A</i>	.....	.....	N.S.S.VS..	D...N.A..R	.....L.	.....	
<b>MDH-A</b>	.....	.....	N.N...A.F	D.....S...	.....	.....	
<i>DrMDH-B</i>	.....	...V..C..	N...DET.F	D.....A...	.E.....	.....	
<i>O/MDH-B</i>	.....	...C..	SMSSG.LACF	D.....S...	...A.....	.....	
<b>MDH-B</b>	.....	...C..	N.....A.F	D.....S...	.....	.....	
<i>AbMDH</i>	SSAMSAAKAI	SDHMRDIWFG	TPEGEFISMG	VISDGNAYGI	PSDLIYSFPV	VIKNTWNIV	300
<b>MDH-T</b>	.....	C.....	.....Y.T..S..V	.E.....	Q.....K..	.....	
<i>DrMDH-A</i>	.....	C.....	...D..WV...	IY.S..S..V	.D..M....	K...S.KV.	
<i>O/MDH-A</i>	.....	C.....	...K.....	.YAA..S...	.E.....I	Q.....KV.	
<b>MDH-A</b>	.....	C.....	.....Y.T..S..V	.E.....	Q.....KV.	.....	
<i>DrMDH-B</i>	.....	C.....T.	.....Y.T..S..V	.E.....I	S...D.N.K..	.....	
<i>O/MDH-B</i>	.....	C.....S.	.....Y.T..P..V	.E.....	Q...D.A.K..	.....	
<b>MDH-B</b>	.....	C.....S.	.....Y.T..S..V	.E.....	Q...D...K..	.....	
					*		
<i>AbMDH</i>	DGLAINDFSR	GKMDATAKEL	VEERDTALEF	I-A	333		
<b>MDH-T</b>	...P.....	.....A..	.....A..	L-		pl= 7.3	
<i>DrMDH-A</i>	...S.....	.....A..	.....T.	LSA		pl= 6.6	
<i>O/MDH-A</i>	...P.....	A.....A..	.....MD.	LSQ		pl= 7.2	
<b>MDH-A</b>	...P.....	.....A..	.....A..	L-		pl= 7.7	
<i>DrMDH-B</i>	...P....K	A..E..A..	.....S.	LGV		pl= 7.2	
<i>O/MDH-B</i>	.S.....K	S.....A..	M.....A.	L.A		pl= 6.0	
<b>MDH-B</b>	...P....K	A.....A..	.....A..	L-		pl= 6.0	
						pl= 6.3	

**Fig. 1.** Amino acid sequences for fish MDH genes reported here and for the inferred ancestral teleost (MDH-T), ancestral MDH-A (MDH-A), and ancestral MDH-B (MDH-B) proteins. The single unique negative amino acid shared by both teleost MDH-B proteins is shown in *boldface* and indicated by the *asterisk*. Predicted isoelectric point (pI) values are shown at the 3' end of each protein.

of MDH-A versus MDH-B), we designed oligonucleotide primers specific for each gene and used RT-PCR amplification to assay for the presence of message across a variety of tissues (eye, brain, ovary, liver, gut, and muscle) in zebrafish and medaka (data not shown). The identities of the various PCR amplification products were verified by direct sequencing, to eliminate the possibility of undetected cross amplification between paralogues within specific tissues. Primers specific to either the *DrMDH-A* or the *O/MDH-A* cDNA amplified a product from every

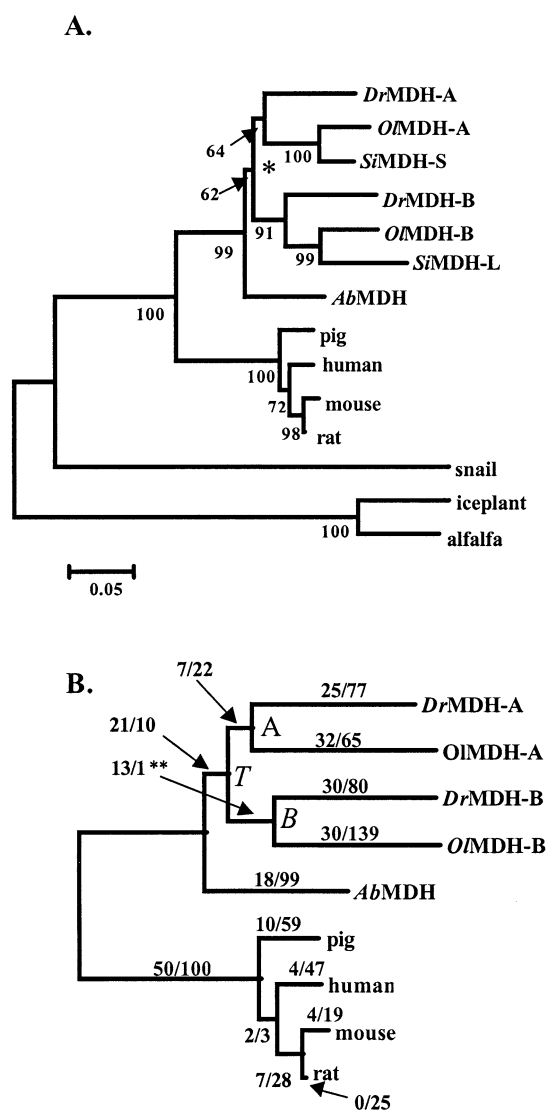
tissue examined in both species. Primers specific for either the *DrMDH-B* or the *O/MDH-B* gene yielded strong amplification products (bright ethidium bromide-stained bands on a 2% agarose gel) from eye and muscle of both species. In contrast, amplifications using the MDH-B primers yielded weak amplification products (fainter bands on the same ethidium bromide-stained gel) from brain, gill, and ovary samples, suggesting lower levels of MDH-B mRNA in these tissues and consistent with the pattern of MDH-A and MDH-B protein expression

reported by Champion and Whitt (1976) and Philipp et al. (1979) in other teleost species.

PCR products amplified from shortnose sturgeon samples using the FishMDH primers yielded a single homogeneous product consisting of 1261 nucleotides, including an ATG start codon, a TGA stop codon, 60 nucleotides of 5' untranslated sequence, and 202 nucleotides of 3' untranslated sequence. The open reading frame codes for a 332-amino acid protein, hereafter referred to as *AbMDH*, with a predicted pI of 7.3 (Fig. 1).

Sequence analysis of 3' RACE products from shortnose sturgeon revealed a second MDH cDNA sequence. A total of 271 bp of this "second" cDNA was sequenced, including 59 nucleotides of coding region (including the stop codon) and 212 nucleotides of 3' untranslated sequence. The coding regions of the two sequences differed at three positions (5%); two substitutions were synonymous. The one nonsynonymous substitution involved a chemically conservative alanine-to-serine change (using the criteria of Grantham 1974). In contrast, over this same portion of the coding region, the two zebrafish MDH cDNAs differ by 14% (55 changes) and the two medaka MDH sequences by 23% (44 changes). The untranslated regions of the two sturgeon sequences one differ by 14%, with 29 differences across 212 bp, including two insertion-deletion events, one of 2 nucleotides and one of 10 nucleotides. Shortnose sturgeons are thought to be recent polyploids (Blackledge and Bidwell 1993). In analysis of the evolution of the TPI gene family in fish, multiple TPI cDNAs were also found in this species (Merritt and Quattro 2001). It seems likely that the multiple cDNAs found from both gene families represent paralogues, genes resulting from a polyploidization event unique to sturgeons. Given the limited amino acid differences between the sturgeon MDH proteins, we have based our analysis of MDH evolution on the single completely sequenced sturgeon MDH coding region,

**Phylogenetic Analyses.** The tree topology from the neighbor-joining (NJ) analysis is shown in Fig. 2A. Analysis using parsimony or maximum likelihood methods yielded identical topologies (parsimony data not shown, maximum likelihood analyses are explored further below). The NJ analysis indicates that the MDH neural and nonneural isozymes of barracuda, medaka, and zebrafish are homologous: 640 of 1000 bootstrap replicates grouped barracuda MDH-S (*SiMDH-S*), *O*/MDH-A, and *DrMDH-A*, 910 of 1000 bootstrap replicates grouped barracuda MDH-L (*SiMDH-L*), *O*/MDH-B, and *DrMDH-B*, while none grouped *SiMDH-S* and *SiMDH-L*, *O*/MDH-A and *O*/MDH-B, or *DrMDH-A* and *DrMDH-B*. Further, the NJ tree topology indicates that the duplication that gave rise to these two genes occurred within the



**Fig. 2.** **A** Neighbor-joining tree depicting phylogenetic relationships among MDH amino acid sequences. The asterisk indicates the duplication that gave rise to the two TPI genes in higher fish. Numbers below each node represent the proportion of 1000 bootstrap replicates supporting that node. **B** The portion of the tree in **A** that includes the vertebrate MDH sequences used in the reconstruction of ancestral sequences. Numbers above each branch represent the number of nonsynonymous (*n*) and synonymous (*s*) substitutions calculated for each branch of the tree, shown as *n/s*. The double-asterisk indicates the only comparison in which *n/s* is significantly greater than *N/S*. *T*, *A*, and *B* indicate the nodes corresponding to the ancestral single teleost MDH sequence, ancestral MDH-A sequence, and ancestral MDH-B sequence, respectively; see text for details.

actinopterygian fish lineage, after the separation of the sturgeon (acipenseriform) and teleost lineages: 990 of 1000 bootstrap replicates grouped the five fish sequences together as a single clade, 620 of 1000 replicates grouped the three pair of teleost MDH sequences within this fish clade. Bootstrap analysis does not support the hypothesis that the two teleost MDH sequences result from a duplication prior to the separation of the teleost and sturgeon lineages: only

60 of 1000 replicates grouped the sturgeon MDH sequence with the MDH-A and MDH-S sequences and none of the replicates grouped the sturgeon sequence with the MDH-B and MDH-L sequences. In sum, the results of our distance-based phylogenetic analysis support the hypothesis that the two teleost cytosolic MDH isozymes result from a duplication event that occurred during the radiation of actinopterygian fish, not earlier in vertebrate evolution as suggested by other authors (e.g., Fisher et al. 1980; Schwantes and Schwantes 1982).

We used maximum likelihood analysis to examine further alternative hypotheses that place the MDH duplication early in the evolution of vertebrates, prior to the separation of the teleost and tetrapod lineages. For the hypothesis of a duplication preceding the separation of teleosts and tetrapods to be correct, vertebrates in which only a single cytosolic MDH protein has been observed (pig, human, rat, mouse, and sturgeon) must be assumed to have lost at least one copy of the gene. Thus, we constrained the tree topology such that the sturgeon MDH and tetrapod MDH sequences grouped with either the teleost MDH-A clade or the MDH-B clade and compared the log likelihood ( $\ln L$ ) of the constrained topology with that of our NJ tree. Although this hypothesis requires the assumption that sturgeon and tetrapods have both lost one cytosolic *Mdh* locus, we cannot assume that sturgeon and tetrapods have lost the same locus. To account for this uncertainty, we examined both topologies, i.e., alternatives that do or do not assume orthology of the sturgeon and tetrapod sequences. Using a criterion of two standard errors of the difference in the likelihood, the topology from our NJ analysis (placing the duplication within fish) has a significantly higher likelihood than that of any of the alternative hypotheses (difference in likelihood between these hypotheses and the NJ tree topology  $\pm 2$  SE ranged from  $\Delta \ln L = -38.7 \pm 13.7$  to  $-46.5 \pm 14.1$ ). We can therefore reject the hypothesis that the pair of MDH proteins found in teleost fish results from a gene duplication early in vertebrate evolution. The pairs of cytosolic MDH isozymes reported in cephalochordates (Fisher et al. 1980) and some amphibians (Schwantes and Schwantes 1982) must be the products of gene duplication events independent from that giving rise to the two fish cytosolic *Mdh* genes.

The above likelihood analysis does not rule out the possibility that the duplication occurred after the separation of tetrapods from the acipenseriform and teleost lineage (actinopterygians) but prior to the separation of the acipenseriformes and teleosts. Under this second alternative hypothesis, the observed sturgeon MDH is orthologous to either teleost MDH-A or teleost MDH-B, a requirement supported by a small fraction (60 of 1000) of bootstrap repli-

cates in our NJ analysis. This second hypothesis is also consistent with the pair of MDH sequences that we found in shortnose sturgeon resulting from the same duplication as the teleost MDH isozymes, and not from the polyploid nature of this species. To test this hypothesis, we constrained the tree topology such that the sturgeon sequence grouped with either the teleost MDH-A clade or the MDH-B clade and compared these topologies with the topology in Fig. 2A. Again using the criterion of two standard errors of the difference in likelihood, we can rule out the constrained tree topologies as consistent with the MDH data set (the difference in likelihood from the NJ tree topology  $\pm 2$  SE ranged from  $\Delta \ln L = -15.6 \pm 7.6$  to  $-23.8 \pm 11.2$ ). The results of this likelihood-based analysis, like those of the distance-based analysis, support the hypothesis of a duplication of the MDH gene within the actinopterygians, after the separation of acipenseriform and teleost fish. These results also support our decision to use the single completed sturgeon sequence as representative of the preduplication fish MDH (see below).

*Reconstructing Ancestral Sequences and Statistical Tests for Selection.* In the absence of selective pressure, the ratio of observed nonsynonymous changes ( $n$ ) to synonymous changes ( $s$ ) will equal the ratio of potential nonsynonymous ( $N$ ) to potential synonymous ( $S$ ) changes (Kimura 1983). Positive, diversifying selection will increase  $n/s$  relative to  $N/S$ , whereas purifying selection will decrease  $n/s$  relative to  $N/S$  (Zhang et al. 1997). Observed changes were calculated for each branch of the MDH gene tree by direct pairwise comparisons of all nodes using the methods of Zhang et al. (1998). Ancestral sequences (internal nodes) were calculated from extant sequences (terminal nodes). The posterior probability of the inferred ancestral states ranged from 98 to 99% (amino acids) and 83 to 97% (nucleotides). Using a ratio of transitions to transversions ( $R$ ) of 1.0 (calculated from the data set),  $N$  and  $S$  were calculated to average 730 and 263, respectively, across the MDH tree. The  $n/s$  ratio exceeds the  $N/S$  ratio along two branches of the tree (Fig. 2B): the branch leading to the ancestral single teleost gene and the branch leading to the ancestral B gene ( $T-B$  branch). Using Fisher's exact test, the difference between the  $n/s$  and the  $N/S$  values was significant only along the  $T-B$  branch ( $p = 0.08$ ). Acceptance of a significance level of  $\alpha = 0.08$  is permissive but is justified here by the fact that our analysis of the TPI gene family (Merritt and Quattro 2001) gave us an *a priori* reason to expect an increase in  $n/s$  relative to  $N/S$  along this *specific* branch of the gene tree. Seven sites of the inferred  $T$ ,  $B$ , or  $A$  sequences (amino acids 164, 176, 203, 207, 210, 274, and 329; Fig. 1) were deemed ambiguous because the most likely amino acid was

**Table 1.** Number of amino acid substitutions that alter or conserve a given amino acid property along branches of the MDH phylogeny.

Property	Reconstructed changes <sup>a</sup>		<i>p</i> <sup>b</sup>
	<i>T–B</i> branch	Tree	
<i>Charge</i>			
Conserve	8 (8) <sup>c</sup>	164 (149)	0.33 (0.59)
Alter	3 (1)	36 (27)	
<i>Polarity</i>			
Conserve	8 (7)	140 (127)	0.57 (0.60)
Alter	3 (2)	60 (49)	
<i>Size</i>			
Conserve	8 (7)	135 (121)	0.50 (0.44)
Alter	3 (2)	65 (55)	

<sup>a</sup> Branches are as defined in Fig. 2; “tree” refers to changes along all branches summed excluding the MDH *T–B* branch.  
<sup>b</sup> The result of a Fisher’s exact test that the observed number of changes within properties is homogeneous across branch designations.  
<sup>c</sup> Results of a similar analysis excluding seven ambiguously reconstructed amino acid positions (see Results) are shown in parentheses. See Methods for details.

less than twice as probable as the next most likely amino acid (per Zhang et al. 1998). If these seven sites are excluded from the analysis, *n/s* still exceeds *N/S* along the *T–B* branch, although the  $\alpha$  level is reduced (*n/s* = 11/1; *p* = 0.13).

**Nonrandom Amino Acid Substitution.** Comparison of modern and inferred amino acid sequences indicates that following the gene duplication that gave rise to the neural TPI protein, this protein accumulated negative amino acids at an accelerated rate (Merritt and Quattro 2001). Here we use a similar strategy to investigate the pattern of amino acid substitution following the duplication of the MDH gene in fish. Predicted amino acid sequences were used to compare the amino acid substitutions along the *T–B* branch (Fig. 2) with those across the rest of the MDH gene tree (Table 1). Amino acids were grouped by physical and chemical properties (using the criterion of Grantham 1974) to determine whether observed amino acid substitutions conserved or altered charge. Three of the 11 amino acid changes between node *T* and node *B* result in a change in amino acid charge: one change is from a neutral amino acid to a negatively charged amino acid, one is from a positively charged amino acid to a neutral amino acid, and one is from a negatively charged amino acid to a neutrally charged amino acid (Fig. 1). There are 200 amino acid changes across the rest of the MDH gene tree; 36 of these are to, or from, charged amino acids (data not shown). These ratios, 8/3 and 164/36, are not significantly different (Fisher’s exact test, *p* = 0.33; Table 1). Excluding the seven ambiguous ancestral sites (see above), 9 amino acid changes occur along the *T–B* branch, 1 of which alters charge (from neutral to negatively charged), and 176 changes occur across the rest of the tree, 27 of which are to, or from, negative or positive amino acids. The ratio of changes along the

*T–B* branch is still not significantly different from the ratio observed for the rest of the tree (8/1 vs 149/27; *p* = 0.59). Similarly, if amino acid changes were grouped by size or polarity, the *T–B* branch did not differ significantly from the rest of the gene tree (Table 1).

**Discussion**

Phylogenetic analysis of vertebrate MDH cDNA sequences, using both distance- and likelihood-based methods, indicated that the pair of MDH isozymes found in teleost fish is the result of duplication of the cytosolic *Mdh* gene early in the evolution of actinopterygian fish. Phylogenetic analysis provided no support for the hypothesis (e.g., Fisher et al. 1980; Schwantes and Schwantes 1982) that the two MDH isozymes found in teleost fish result from a gene duplication earlier in the evolution of vertebrates. These results confirm and extend the conclusion of Lin et al. (2002) that the two teleost MDH isozymes are the products of a duplication of the cytosolic *Mdh* gene after the separation of teleosts and tetrapods. The presence of pairs of MDH isozymes in both amphibians and teleosts has led to the proposal that they might represent an adaptation to ectothermy (Schawntes and Schwantes 1982; Lin and Somero 1995). Our analysis indicates that these pairs are not in fact orthologous but, rather, are the products of independent duplication events. Any similar adaptive function between these isozyme pairs must therefore be the product of convergent evolution.

Two other teleost isozymes expressed primarily in neural tissue, TPI-A and LDH-C, also result from duplications that occurred early in the radiation of teleost fish (Quattro et al. 1993; Stock et al. 1997; Merritt and Quattro 2001). Further, studies of the evolution of the *Hox* (Amores et al. 1998) and *Dlx* (Stock et al. 1996; Neidert et al. 2001) gene families



indicate duplication events during this same time period which have been interpreted as evidence for duplication of entire chromosomes, or perhaps the entire genome (Stock et al. 1996; Amores et al. 1998; Neidert et al. 2001). The similarity in timing of the duplication of the *Hox* and *Dlx* genes and the duplication(s) responsible for the teleost MDH, TPI, and LDH neural isozymes suggests that these duplications are all the result of a single large-scale duplication event early in the radiation of teleost fish. It is interesting to note that just such a large-scale duplication was originally predicted almost 20 years ago (Buth 1983) to explain the pair of teleost cytosolic MDH isozymes.

The neurally expressed TPI-A and LHD-C exhibit pronounced net negative charges (exceptions to this exist for LDH-C but appear to be correlated with changes in expression pattern [see Whitt 1987]). Based on the correlation of negative charge and neural expression across a number of gene families, the negative charge of the neural isozymes was proposed to be an adaptation to the neural environment (Fisher et al. 1980). In support of this proposal, we have shown that directly following the duplication of the *Tpi* gene in teleost fish, the neural TPI isozyme evolved through a period of directional selection, specifically accumulating negatively charged amino acids (Merritt and Quattro 2001). The predicted isoelectric points (pI) of the neural MDH isozymes from both zebrafish and medaka indicate that, like the neural TPI isozyme, the neural MDH isozyme (MDH-B) is more negatively charged than its generally expressed counterpart (MDH-A; Fig. 1). The differential in charge is not, however, as pronounced between MDH isozymes as it is between TPI isozymes. In modern teleost fish, the generally expressed TPI isozyme (TPI-B) has a predicted pI of approximately 7.0, whereas the neural TPI isozyme (TPI-A) has a predicted pI of approximately 4.5. In contrast, the generally expressed teleost MDH has a predicted pI of approximately 7.0, while the predicted pI of the neural isozyme is approximately 6.0 (Fig. 1). Comparison of amino acid sequence indicates that *DrMDH-B* and *O/MDH-B* isozymes share only a single negatively charged amino acid not found in the generally expressed isozyme (amino acid 294; Fig. 1). In contrast, the teleost neural TPI isozymes share seven unique negatively charged amino acids (Merritt and Quattro 2001).

The MDH and TPI isozyme pairs have similar patterns of expression, suggesting that they might be under similar selective regimes. Since our phylogenetic analysis indicates that the MDH and TPI neural isozymes likely result from a single duplication event, the neural MDH isozyme has had the same amount of time as the TPI neural isozyme to respond to selective pressures and accumulate a change in net

charge. The smaller charge differential in MDH, therefore, reflects either differences in selective pressure upon the MDH and TPI isozymes or differences in the abilities of these gene families to respond to similar selective pressures.

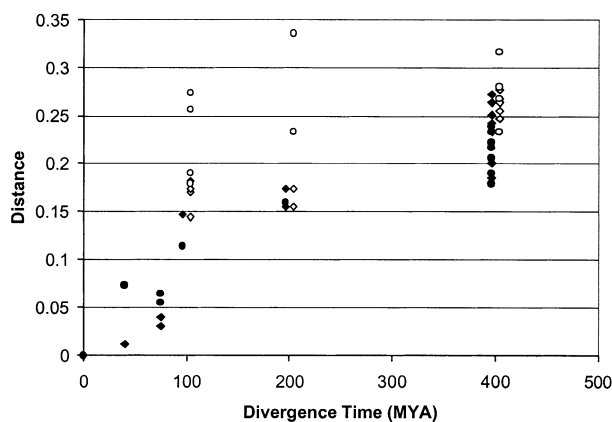
Any major response to changes in selective pressures is most likely to occur directly following the duplication event. Comparison of the observed number of nonsynonymous and synonymous changes ( $n$  and  $s$ ) with the potential number of such changes ( $N$  and  $S$ ) across the *T-B* branch of the MDH gene tree (Fig. 2B) indicates a relative excess of nonsynonymous change. The excess of amino acid substitutions suggests that directly following the duplication of the ancestral gene, MDH-B evolved through positive selection. Comparison of  $n/s$  with  $N/S$  across all other branches of the MDH gene tree shows that, in general,  $n/s$  is less than  $N/S$  (Fig. 2B), suggesting purifying selection, as is expected for a protein coding sequence. Notably,  $n/s$  is again less than  $N/S$  along the terminal MDH-B branches, indicating purifying selection subsequent to a period of positive selection. This episodic pattern of positive selection, preceded and followed by periods of purifying selection, is similar to the pattern seen in the evolution of the TPI neural isozyme (Merritt and Quattro 2001). The observed difference between  $n/s$  and  $N/S$  along the branch leading to the ancestral neural isozyme is, however, smaller in MDH ( $n/s = 4.7 \times N/S$ ) than in TPI ( $n/s = 8.5 \times N/S$ ).

The pattern of amino acid substitution along the branch leading to the ancestral neural MDH isozyme is not significantly different from the pattern of substitution across the rest of the MDH gene tree (Table 1). This uniformity in the pattern of amino acid substitution is unexpected given the evidence for positive selection across this same branch and is distinctly different from the situation observed in TPI (Merritt and Quattro 2001). Across the corresponding branch of the TPI gene tree, the rate of substitution to negatively charged amino acids was almost three times the rate of *all* types of charge across the rest of the tree. This significant bias in the accumulation of negatively charged amino acids is evidence that the selective pressure on the TPI gene was specifically for negatively charged amino acids (Merritt and Quattro 2001). Without a similar bias in the pattern of amino acid substitution apparent in the evolution of the MDH-B isozymes, the nature of the selective pressure responsible for the excess of nonsynonymous change in this isozyme is unclear.

The small charge differential observed between the MDH isozymes could reflect a constraint on charge changes in the *Mdh* gene family not present, or present to a lesser degree, in gene families which possess larger charge differences between isozymes.

Under such a constraint, the neural MDH isozymes responded to the same selective pressure as TPI, but MDH did so at a slower rate than the less constrained TPI. Comparison of charge changes following the duplication in each gene family shows that the pattern of amino acid substitution observed in MDH (3/8; Table 1) is not significantly different from that observed in TPI (8/12 [from Merritt and Quattro 2001]; Fisher's exact test,  $p = 0.38$ ). The lack of a significant difference might result from the small number of changes along this branch of the *Mdh* gene tree. However, comparison of charge changes across each gene tree as a whole leads to the same conclusion. Across the *Mdh* gene tree, there are 36 amino acid substitutions that alter charge and 164 that do not. Across the *Tpi* gene tree there are 21 changes that alter charge and 133 that do not. These ratios are not significantly different (Fisher's exact test,  $p = 0.17$ ). This similarity in the relative frequency of amino acid substitutions that change charge across the whole of the *Mdh* and *Tpi* gene trees suggests that both gene families have evolved under a similar level of constraint against charge change. The larger charge differential in the *Tpi* gene family is, then, unlikely to be the result of greater constraint on charge changes across the *Mdh* gene family.

The limited charge differential could instead result from an overall slower rate of amino acid substitution in the *Mdh* gene family (rate of all amino acid changes, not just changes involving amino acid charge) than in gene families with a larger differential in charge between general and neural isozymes. Under this hypothesis, the neural MDH and TPI isozymes are both responding to the same selective pressures, but MDH is doing so at a slower rate as a result of an overall slower rate of change across this gene family. To address this possibility, we compared the rate of amino acid substitution in the *Mdh* gene family with that in the *Tpi* gene family. The eight amino acid sequences available for both gene families (single sequences from human, mouse, rat, and sturgeon, a pair of cyprinid sequences, and a pair of atherinid sequences) were used to calculate all possible pairwise distances for each gene family using the same parameters used in tree construction. These distances were then plotted against divergence times taken from the literature (Fig. 3). While the species of atherinid used in the MDH and TPI studies are different, this analysis focuses on the divergence across large-scale taxonomic groups (e.g., primates and atherinids or acipensiformes and cyprinids) and is unlikely to be influenced by differences within the species compared. Across all divergence times, MDH (diamonds) shows somewhat smaller distances than TPI (circles), suggesting a slower rate of change in MDH than in TPI. However, this apparent difference is largely the result of large distance values involving



**Fig. 3.** Plot of distance versus time of divergence from pairwise comparisons of vertebrate TPI and cytosolic MDH amino acid sequences. Values for TPI are shown as circles; values for MDH, as diamonds. Open symbols indicate values involving the two neural isozymes (TPI-A and MDH-B), whereas filled symbols indicate values for all other comparisons. Open and filled symbols are shown offset, bracketing the value on the X axis to aid in visualization. Divergence times are from the literature: rat/mouse, 40 MYA; primate/rodent, 75 MYA (Dayhoff 1978); cyprinid/atherinid, 100 MYA (Patterson 1993); acipensiforme/teleost, 200 MYA (Grande and Bemis 1996); and tetrapod/actinopterygian fish, 400 MYA (Dayhoff 1978).

the neural TPI. If distances involving the neural isozymes (Fig. 3; open symbols) are ignored, the two gene families appear to be accumulating change at very similar rates. This suggests that the difference in magnitude of the charge differential cannot simply be explained by an overall slower rate of change in MDH than TPI and, instead, implicates the neural isozymes themselves as the source of the difference in the charge differential.

There are proportionately fewer amino acid substitutions along the branch leading to the neural MDH (11 of 333, or 3%; compare Fig. 1, *T* and *A*) than along the same branch in the TPI tree (20 of 249, or 8% [from Merritt and Quattro 2001, Fig. 1]; Fisher's  $p = 0.01$ ). This difference suggests a more limited response to changes in selective pressures from MDH than TPI. To examine further the difference in response between the two gene families, we performed a relative rate test (Takezaki et al. 1995) on each gene family to test for constancy of evolutionary rate after duplication. Using the same set of eight sequences as above, we compared the amount of change among neural isozymes, generally expressed isozymes, and other vertebrate sequences. Tests were conducted using the two-cluster method of Takezaki et al. (1995) as implemented in PHYLTEST (version 2.0 [Kumar 1996]). Distances were calculated as in the NJ analysis. Our comparison of  $n/s$  with  $N/S$  indicates that the rate of amino acid substitution is elevated above neutral expectations for both gene families (indicated by the bias toward nonsynonymous changes), specifically along the branch leading

to the neural isozyme. In TPI, comparison of the relative rate of change between the neural and the general isozymes indicates that this bias is pronounced enough to lead to violation of rate constancy ( $p < 0.05$ ); the neural TPI isozymes have evolved at a significantly faster rate than their generally expressed counterparts. In contrast, the acceleration in amino acid substitution of MDH is not pronounced enough to provide evidence for any difference in the rate of evolution between the neural and the generally expressed isozymes ( $p > 0.05$ ). Following the duplication event, MDH does not show the same degree of acceleration in amino acid substitution as observed for TPI.

We conclude that following a duplication event both the neural TPI and the neural MDH isozymes evolved through a period of positive selection. In TPI-A, this period was accompanied by a large number of amino acid substitutions with a strong bias toward the accumulation of negatively charged amino acids, thereby increasing the overall net negative charge of the protein. No such bias in amount or type (based on charge, size, or polarity) of amino acid change is apparent in MDH-B. In the absence of any apparent bias, the focus of selection in the evolution of the neural MDH remains unclear.

The different patterns of amino acid change between the MDH and the TPI neural isozymes might reflect different biochemical functions or different expression patterns. Different biochemical functions, different pathways, or different classes of chemical reactions might reasonably be assumed to put different constraints on the proteins involved. The selective pressure that promotes the accumulation of negative charge by neural isozymes might be more constrained by the biochemical role of MDH than by the roles of isozymes from gene families that exhibit a larger charge differential. Further, the neural MDH has a broader pattern of expression than either TPI-A or LDH-C, being expressed in muscle as well as neural tissue and at lower levels in other tissues as well. This broader pattern of expression might increase the constraint upon the MDH isozyme and thereby limit the accumulation of negative charge by this protein. The relatively small charge differential observed between the MDH isozymes might represent the maximum amount of negative charge that the MDH protein can accumulate and still maintain function across a broad range of expression.

A better understanding of the selective pressure driving the accumulation of the negative charge by neural isozymes might shed light on the reason for varying charge differential among isozymes within gene families. That the correlation between net negative charge and neural expression exists across a variety of gene families suggests that the selective advantage provided by the negative charge is not

specific to a particular biochemical pathway or type of reaction. We have suggested that the increase in negative charge might play a role in maintenance of the resting potential of neural tissue (Merritt and Quattro 2001). Other possibilities are almost certainly as likely, perhaps relating to enzymatic function and some unique feature of the neural cellular environment. Examination of a larger number of gene families that include a neural isozyme (negatively charged and otherwise) and the specific environment in which they function (e.g., cell type) will likely shed more light on the biological significance of the strong charge bias.

**Acknowledgments.** We thank Joseph Staton, Jacqueline Litzgus, Efe Sezgin, and Luciano Matzkin for helpful comments on the manuscript. This work was supported by the Research and Productive Scholarship Fund of the University of South Carolina, the Cooperative Institute for Fisheries Molecular Biology (FISHTEC; NOAA/NMFS [RT/F-1]), the National Science Foundation (OCE-9814172), and SC SeaGrant (R/MT-5).

## References

- Adach J, Hasegawa M (1994) MOLPHY: Maximum likelihood inference of protein phylogeny. Version 2.2. Department of Statistical Science, Graduate University for Advanced Study, Japan
- Amores A, Force A, Yan YL, et al. (1998) Zebrafish *hox* cluster and vertebrate genome evolution. *Science* 282(27):1711–1714
- Blackledge KH, Bidwell CA (1993) Three ploidy levels indicated by genome quantification in *Acipenseriformes* of North America. *J Hered* 84:427–430
- Buth DG (1983) Duplicate isozyme loci in fishes: Origins, distribution, phylogenetic consequences, and locus nomenclature. In: Rattazzi MC, Scandalios JG, Whitt GS (eds) *Isozymes: Current topics in biological and medical research*. Alan R. Liss, New York, Vol 10, pp 381–400
- Champion MJ, Whitt GS (1976) Differential gene expression in multilocus isozyme systems of the developing green sunfish. *J Exp Zool* 196:263–282
- Davidson RG, Cortner JA (1967) Genetic variant of human erythrocyte malate dehydrogenase. *Nature* 215:761–762
- Dayhoff MO (1978) Survey of new data and computer methods of analysis. In: Dayhoff MO (ed) *Atlas of protein sequence and structure*, Vol 5, Suppl 3. National Biochemical Research Foundation, Washington, DC pp 1–8
- Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791
- Fisher SE, Shaklee JB, Ferris SD, Whitt GS (1980) Evolution of five isozyme systems in the chordates. *Genetica* 52/53:73–85
- Frohman MA (1990) RACE: Rapid amplification of complementary DNA ends. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) *PCR protocols: A guide to methods and applications*. Academic Press, New York, pp 28–38
- Grande L, Bemis WE (1996) Interrelationships of *Acipenseriformes* with comments on “chondrostie.” In: Stiassny MLJ, Parenti LR, Johnson GD (eds) *Interrelationships of fishes*. Academic Press, San Diego, Chap 5, pp 85–115
- Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185:862–864
- Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B* 256:119–124

- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *CABIOS* 8:275–282
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) *Mammalian protein metabolism*. Academic Press, New York, pp 21–123
- Karig LM, Wilson AC (1971) Genetic variation in supernatant malate dehydrogenases of birds and reptiles. *Biochem Genet* 5:211–221
- Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge
- Kumar S (1996) PHYLTEST: A program for testing phylogenetic hypothesis, version 2.0. Pennsylvania State University, University Park
- Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis, version 1.0. Pennsylvania State University, University Park
- Lin J-J, Somero G (1995) Thermal adaptation of cytoplasmic malate dehydrogenases of eastern pacific barracuda (*Sphyræna* spp.): The role of differential isozyme expression. *J Exp Biol* 198:551–560
- Lin JJ, Yang TH, Wahlstrand BD, Fields PA, Somero GN (2002) Phylogenetic relationships and biochemical properties of the duplicated cytosolic and mitochondrial isoforms of malate dehydrogenase from a teleost fish, *Sphyræna idiatæ*. *J Mol Evol* 54(1):107–117
- Markert CL, Whitt GS (1968) Molecular varieties of isozymes. *Experientia* 24:977–991
- Markert CL, Shaklee JB, Whitt GS (1975) Evolution of a gene. *Science* 189:102–114
- Meizel S, Markert CL (1967) Malate dehydrogenase isozymes of the marine snail, *Ilyanassa obsoleta*. *Arch Biochem Biophys* 122:753–765
- Merritt TJS, Quattro JM (2001) Evidence for a period of directional selection following gene duplication in a neurally-expressed locus of triosephosphate isomerase. *Genetics* 159: 689–697
- Messier W, Stewart C-B (1997) Episodic adaptative evolution of primate lysozymes. *Nature* 385:151–154
- Morizot DC, Schmidt ME (1990) Starch gel electrophoresis and histochemical visualization of proteins. In: Whitmore DH (ed) *Electrophoretic isoelectric focusing techniques in fisheries management*. CRC Press, Boca Raton, FL
- Nei M, Chakabarty R, Fuerst PA (1976) Infinite allele model with varying mutation rate. *Proc Natl Acad Sci USA* 73:4164–4168
- Neidert AH, Virupannavar V, Hoover GW, Langeland JA (2001) Lamprey Dlx genes and early vertebrate evolution. *Proc Natl Acad Sci USA* 98(4):1665–1670
- Paterson C (1993) An overview of the early fossil record of acanthomorphs. *Bull Mar Sci* 52(1):29–59
- Philipp DP, Childers WF, Whitt GS (1979) Evolution of patterns of differential gene expression: A comparison of the temporal and spatial patterns of isozyme locus expression in two closely related fish species (northern largemouth bass, *Micropterus salmoides salmoides*, and smallmouth bass, *Micropterus dolomieu*). *J Exp Zool* 210:473–488
- Philipp DP, Parker HR, Whitt GS (1983) Evolution of gene regulation: Isozymic analysis of patterns of gene expression during hybrid fish development. In: Rattazzi MC, Scandalios JG, Whitt GS (eds) *Isozymes: Current topics in biological and medical research*. Alan R. Liss, New York, Vol 10, pp 193–237
- Pontier, PJ, Hart NH (1981) Developmental expression of glucose and triose phosphate isomerase genes in teleost fishes (*Brachydanio*). *J Exp Zool* 217:53–71
- Quattro JM, Woods HA, Powers DA (1993) Sequence analysis of teleost retina-specific lactate dehydrogenase C: Evolutionary implications for the vertebrate lactate dehydrogenase gene family. *Proc Natl Acad Sci USA* 90:242–246
- Rainboth WJ, Whitt GS (1974) Analysis of evolutionary relationships among shiners of the subgenus *Luxilus* (Teleostei, Cypriniformes, *Notropis*) with the lactate dehydrogenase and malate dehydrogenase isozymes systems. *Comp Biochem Physiol* 49B:241–252
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Ehrlich HA (1988) Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Schwantes MLB, Schwantes AR (1982) Adaptive features of ectothermic enzymes. I. Temperature effects on the malate dehydrogenase from a temperate fish *Leiostomus xanthurus*. *Comp Biochem Physiol* 72B:49–50
- Stock DW, Ellies DL, Zhao Z, Ekker M, Ruddle FH (1996) The evolution of the vertebrate Dlx gene family. *Proc Natl Acad Sci USA* 93(20):10858–10863
- Stock DW, Quattro JM, Whitt GS, Powers DA (1997) Lactate dehydrogenase (LDH) gene duplication during chordate evolution: The cDNA sequence of the LDH of the tunicate *Styela plicata*. *Mol Biol Evol* 14:1273–1284
- Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol* 12:823–833
- Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680
- Thorne CJR, Cooper PM (1963) Preparation of pig heart-supernatant malate dehydrogenase. *Biochim Biophys Acta* 81:397–399
- Uzzel T, Corbin KW (1971) Fitting discrete probability distribution to evolutionary events. *Science* 172:1089–1096
- Wheat TE, Childers WI, Miller ET, Whitt GS (1971) Genetic and in vitro molecular hybridization of malate dehydrogenase isozymes in interspecific bass (*Micropterus*) hybrids. *Anim Blood Grps Biochem Genet* 2:3–14
- Wheat TE, Whitt GS, Childers WF (1973) Linkage relationships of six enzyme loci in interspecific sunfish hybrids (genus *Lepomis*). *Genetics* 74:343–350
- Whitt GS (1987) Species differences in isozyme tissue patterns: Their utility for systematic and evolutionary analyses. In: Rattazzi MC, Scandalios JG, Whitt GS (eds) *Isozymes: Current topics in biological and medical research*, Vol 15. Alan R. Liss, New York, pp 2–26
- Zhang J, Kumar S, Nei M (1997) Small-sample tests of episodic evolution: A case study of primate lysozymes. *Mol Biol Evol* 14:1335–1338
- Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA* 95:3708–3713