

Evidence for a Period of Directional Selection Following Gene Duplication in a Neurally Expressed Locus of Triosephosphate Isomerase

T. J. S. Merritt and J. M. Quattro

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

Manuscript received May 15, 2001
Accepted for publication July 9, 2001

ABSTRACT

A striking correlation between neural expression and high net negative charge in some teleost isozymes led to the interesting, yet untested, suggestion that negative charge represents an adaptation (via natural selection) to the neural environment. We examine the evolution of the triosephosphate isomerase (TPI) gene family in fishes for periods of positive selection. Teleost fish express two TPI proteins, including a generally expressed, neutrally charged isozyme and a neurally expressed, negatively charged isozyme; more primitive fish express only a single, generally expressed TPI isozyme. The TPI gene phylogeny constructed from sequences isolated from two teleosts, a single acipenseriform, and other TPI sequences from the databases, supports a single gene duplication event early in the evolution of bony fishes. Comparisons between inferred ancestral TPI sequences indicate that the neural TPI isozyme evolved through a period of positive selection resulting in the biased accumulation of negatively charged amino acids. Further, the number of nucleotide changes required for the observed amino acid substitutions suggests that selection acted on the overall charge of the protein and not on specific key amino acids.

TISSUE-SPECIFIC isozymes, multiple molecular forms of a specific enzyme within individuals, are fundamental components of cellular diversity, representing specialization of proteins for specific functions and distinct cellular environments. For this reason, researching isozyme evolution provides insight into the mechanisms through which biochemical and biological complexity has evolved. Several authors hypothesize that functional specialization within gene families is the result of natural selection (*e.g.*, LI 1983; OHTA 1991; HUGHES 1994), and the unique biochemical properties of highly specialized isozymes within gene families circumstantially support this selection hypothesis. For example, within the lactate dehydrogenase (LDH: EC 1.1.1.27) gene family, *Ldh-C* results from a relatively recent duplication of the *Ldh-B* locus in teleost fishes (QUATTRO *et al.* 1993). Although the duplication was relatively recent, LDH-C is expressed predominantly in the eye and brain and has pH optima, reaction kinetics, and thermal stability distinct from LDH-B (WHITT 1970; SHAKLEE *et al.* 1973). These unique attributes possibly represent adaptation of LDH-C to expression in the distinct, highly aerobic, neural environment (WHITT 1970).

Neural isozymes such as the teleost LDH-C are widely distributed among vertebrates (*e.g.*, PENHOET *et al.* 1966; FISHER *et al.* 1980; MARANGOS and SCHMECHEL 1987;

MORIZOT and SCHMIDT 1990). Most neural isozymes have not been as well characterized biochemically as LDH-C, but a striking number have a high net negative charge (*e.g.*, BURGER *et al.* 1963a,b; SHAKLEE *et al.* 1973; CHAMPION and WHIT 1976; FISHER and WHITT 1978; FISHER *et al.* 1980). FISHER *et al.* (1980) hypothesized that high net negative charge might represent an adaptation to the unique biochemistry of the neural environment.

Recently, positive selection has been shown to have been involved in adaptive radiation within a number of gene families (*e.g.*, TANAKA and NEI 1989; HUGHES and HUGHES 1993; ZHANG *et al.* 1998; DUDA and PALUMBI 1999; BISHOP *et al.* 2000; WILLETT 2000). In these cases, gene duplication was shown to be followed by a period of positive selection, a reflection of functional divergence within the gene family. Functional divergence of the neural, and generally expressed, fish LDH isozymes suggests that they might have evolved through a similar period of positive selection, perhaps subsequent to gene duplication. This would be especially interesting in that all examples of gene duplication and positive selection to date have involved reproductive isolation mechanisms, ligand binding, or recognition of non-self. If positive selection was involved in the functional divergence of the neural isozymes this would be the first example reported involving duplication, and specialization to a unique environment within an organism, of a housekeeping protein.

Triosephosphate isomerase (TPI: E.C. 5.3.1.1) is a dimeric glycolytic enzyme that catalyzes the interconversion of dihydroxyacetone phosphate and glyceraldehyde-

Corresponding author: Joseph M. Quattro, Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208. E-mail: quattro@mail.biol.sc.edu

3-phosphate. Two TPI proteins are expressed in teleost fish: a neutrally charged, generally expressed isozyme and a negatively charged neural isozyme (PONTIER and HART 1981; MORIZOT and SCHMIDT 1990); all other jawed vertebrates (gnathostomes), including the sarcopterygian fish, the coelocanth, express a single TPI protein (KOLB *et al.* 1974; MAQUAT *et al.* 1985; STRAUS and GILBERT 1985; OLD and MOHRENWEISER 1988; CHENG *et al.* 1990; KURAKU *et al.* 1999). In addition, only a single TPI is expressed in the lancelet, a primitive chordate (NIKOH *et al.* 1997), and in two agnathid vertebrates, lamprey and hagfish (KURAKU *et al.* 1999). Several phylogenetic studies of TPI sequences suggest that a TPI gene phylogeny is recoverable and likely well supported (NIKOH *et al.* 1997; KURAKU *et al.* 1999). The apparent simplicity of the TPI gene family makes it an attractive system with which to investigate the role of selection in the evolution of neurally expressed isozymes.

Recognition that gene families evolve episodically, with periods of positive selection followed by periods of purifying selection, has been crucial to the demonstration of the role of selection in the diversification of gene families. Positive selection can result in the rate of non-synonymous substitution exceeding that of synonymous substitution, whereas purifying selection generally results in the rate of synonymous substitution exceeding the nonsynonymous rate. The period of positive selection following gene duplication may be relatively brief, and if gene sequences are compared across too long a period of evolutionary time, evidence for positive selection might be obscured by subsequent substitutions (HUGHES 1994). Given that many gene duplications are quite old, 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 approach requires that taxa be selected such that ancestral sequences tightly bracket the period of evolutionary interest.

Any changes in selective pressures acting on the two TPI isozymes presumably occurred directly following duplication of the *Tpi* locus. The presence of a second *Tpi* locus in teleost fishes, and only in teleost fishes despite the wide variety of chordates that have been surveyed, strongly suggests that the two fish isozymes are the products of a gene duplication sometime during the radiation of higher fishes. To bracket this period in the evolution of the TPI locus we have sequenced the entire coding region of the single TPI gene from the shortnose sturgeon, a primitive actinopterygian fish, and two TPI genes, each from two teleost species, the zebrafish and the southern platy. Comparisons of ancestral TPI nucleotide and amino acid sequences constructed using these sequences and other vertebrate TPI sequences from the data banks indicate a period of

directional selection and the specific accumulation of negatively charged amino acids following the TPI gene duplication.

MATERIALS AND METHODS

TPI cloning and sequencing: *Acipenser brevirostrum* (shortnose sturgeon) tissues were obtained from R. Chapman, South Carolina Department of Natural Resources (Charleston, SC); *Xiphophorus maculatus* (southern platy) and *Danio rerio* (zebrafish) samples were purchased from local pet stores. Tissues were dissected from single fish and processed immediately or stored at -70° . Total RNA was purified from tissue using a commercially available kit (RNAeasy, QIAGEN, Chatsworth, CA). Complementary DNA (cDNA) was synthesized from total RNA according to the manufacturer's instructions (Superscript Preamplification System, GIBCO BRL, Gaithersburg, MD).

Degenerate oligonucleotide primers were designed from an alignment of vertebrate TPI sequences in GenBank: TPI14F, 5'-GTN GGN GGN AAY TGG AAR ATG-3'; TPI16F, 5'-GGN AAY TGG AAR ATG AAY GG-3'; TPI225R, 5'-CAN ARR AAN CCR TCN AYR TC-3'; and TPI235R, 5'-WWN TCN ACR AAY TCN GGY TT-3'.

Degenerate positions are represented by the following ambiguity codes: *N* = A, G, C, T; *Y* = C, T; *R* = A, G; *W* = A, T. Numbers refer to the amino acid position populated by the 3' base of each primer; a preliminary ClustalW (THOMPSON *et al.* 1994) alignment of protein sequences from GenBank was used as a reference.

Combinations of these four oligonucleotide primers were used to amplify a 681- to 687-bp segment of the TPI 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° for 1 min, annealing at 48° for 1 min, and extension at 72° for 1 min. In some cases nested reamplifications were necessary to obtain sufficient PCR product for cloning. In these cases, a sample of the first amplification product (using primers 14F and 235R) was diluted 1:100 and used as template for a secondary amplification (using primers 16F and 225R) under identical PCR conditions. PCR products were cloned into pGEM T-vector (Promega, Madison, WI) and sequenced manually. Dideoxy DNA sequencing was performed with Sequenase (USB Biochemicals) and [35 S]dATP. At least three independent clones per PCR fragment were sequenced on both strands.

Individual TPI loci were targeted for PCR amplification by exploiting the tissue-specific expression of each isozyme (PONTIER and HART 1981; MORIZOT and SCHMIDT 1990). TPI-B was amplified from liver cDNA from both zebrafish and platy. TPI-A and TPI-B were amplified from zebrafish eye or ovary cDNA. Sequence analysis of individual clones allowed us to distinguish the two loci. Cloning of degenerate PCR products from platy eye, brain, or ovary cDNA yielded only TPI-B clones. However, direct sequencing of amplification products from platy eye or ovary cDNA allowed portions of the TPI-A sequence to be read as "background," secondary bands on the sequencing gel, behind the strong signal of the TPI-B sequence. TPI-A specific primers were designed from these regions. The extreme 5' and 3' coding and untranslated regions of each cDNA was amplified using the method of rapid amplification of cDNA ends (RACE; FROHMAN 1990). RACE amplifications used gene-specific primers designed from the initial gene fragment. Conditions for the PCR, cloning, and screening were as described above. Isoelectric point (pI) values were calculated for each predicted amino acid sequence using the EMBL isoelectric point service (<http://www.embl-heidelberg.de/cgi/pi-wrapper.pl>).

Phylogenetic analyses: In addition to the five fish TPI se-

quences reported here, the following gnathostome sequences were obtained from GenBank and used in phylogenetic analyses: human (GenBank accession no. M10036), rhesus monkey (X08023), rat (L36250), mouse (L31793), and chicken (M11941). A previous analysis of TPI gene evolution revealed no evidence for a gene duplication prior to the divergence of agnathid and gnathostome vertebrates; thus the single TPI cDNAs isolated from agnathans (lamprey, AB025327; hagfish, AB025322) were used to root phylogenetic trees describing TPI evolution in gnathostome vertebrates.

Amino acid sequences were aligned using the ClustalW multiple alignment program (THOMPSON *et al.* 1994). Minor adjustments were made to the alignment manually, resulting in an overall alignment of 246-amino-acid positions. Nucleotide sequences were then aligned using the amino acid alignment as a guide. A phylogeny was constructed by the neighbor-joining (NJ) method (SAITOU and NEI 1987) using the algorithm implemented in PAUP* (version 4.0 b2; SWOFFORD 1999). Positions in which alignment forced gaps within the sequences were removed because of uncertain homology. Missing data were restricted to the outgroup (agnathans); thus these sites were retained and treated as missing data in phylogenetic analyses. Ingroup (gnathostome) relationships were estimated first and this topology used as a backbone constraint in subsequent analyses (*e.g.*, see SWOFFORD *et al.* 1996).

Pairwise distances were calculated from the first and second position nucleotides using the Kimura two-parameter model (KIMURA 1980). Pairwise proportional distances were calculated from the amino acid sequences. Bootstrapping (FELSENSTEIN 1985) was used to evaluate the degree of support for particular groupings in the NJ analysis.

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. Reconstructions used the ANC-GENE computer program (ZHANG *et al.* 1998) with the Jones, Taylor, and Thornton (JTT; JONES *et al.* 1992) model of amino acid substitution. Only the TPI sequences from the jawed vertebrates were used in the sequence reconstructions because of missing data in the lamprey and hagfish sequences. The first three amino acids of each sequence could not be inferred because of gaps in the ClustalW alignment of extant TPI 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 for amino acid substitution, the Jukes-Cantor model of distance estimation (JUKE and CANTOR 1969), or a combination of either of these methods with the JTT model or proportional distances yielded results essentially identical to those reported here. The number of potential nonsynonymous sites (N) and synonymous sites (S) were also calculated for the TPI 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).

Sequences reported in this article have been submitted to GenBank (accession nos: AF387818–AF387822).

RESULTS

Cloning: PCR products isolated from zebrafish liver cDNA consistently yielded a cDNA of 897 nucleotides, including an ATG start codon, TAA stop codon, 80 nucleotides of 5' untranslated sequence, and 70 nucleo-

tides of 3' untranslated sequence. The cDNA open reading frame codes for a putative 248-amino-acid protein with a predicted pI of 6.5 (Figure 1). PCR products from zebrafish eye cDNA included two cDNAs: the first identical to that found in liver samples and the second consisting of 1028 nucleotides, including an ATG start codon, a TGA stop codon, 79 nucleotides of 5' untranslated sequence, and 202 nucleotides of 3' untranslated sequence. The open reading frame of this second cDNA codes for a 248-amino-acid protein with a predicted pI of 4.7 (Figure 1). Both sequences are similar to TPI sequences in GenBank. On the basis of the tissues from which they were isolated, the cDNAs were tentatively identified as, respectively, the TPI-B and TPI-A proteins (herein referred to as *DrTPI-B* and *DrTPI-A*) reported by PONTIER and HART (1981). The predicted pI values for both proteins are consistent with PONTIER and HART's (1981) observation of a neutrally charged, generally expressed isozyme and a negatively charged neural isozyme.

To confirm that the two cDNAs we characterized corresponded to the tissue-specific isozymes reported by PONTIER and HART (1981) 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 in zebrafish. Primers specific for the *DrTPI-B* gene amplified a product from every tissue examined (eye, brain, ovary, liver, gut, and muscle), whereas primers specific for the *DrTPI-A* gene amplified a product only from eye, brain, and ovary (PCR products were sequenced and verified as *DrTPI-B* or *DrTPI-A*). This pattern of message expression matches the pattern of protein expression reported by PONTIER and HART (1981).

PCR products from southern platy tissue samples were similar to those from zebrafish. A single cDNA, consisting of 874 nucleotides, including an ATG start codon, TAA stop codon, 45 nucleotides of 5' untranslated sequence, and 85 nucleotides of 3' untranslated sequence, was isolated from liver samples (hereafter *XmTPI-B*). The cDNA open reading frame codes for a putative 247-amino-acid protein with a predicted pI of 7.7 (Figure 1). Sequence analysis of PCR products from platy eye revealed two cDNAs: the first identical to that found in liver samples and the second consisting of 929 nucleotides, including an ATG start codon, TAA stop codon, 90 nucleotides of 5' untranslated sequence, and 95 nucleotides of 3' untranslated sequence (hereafter *XmTPI-A*). The open reading frame of this second cDNA codes for a 247-amino-acid protein with a predicted pI of 4.4 (Figure 1). Again, both sequences show high similarities to TPI sequences in GenBank, and the predicted pI values of the two platy TPI proteins are consistent with previous observations (MORIZOT and SCHMIDT 1990) of a neutrally charged, generally expressed isozyme and a negatively charged neural isozyme.

We used the same strategy that we used in confirming the identity of the two zebrafish TPI cDNAs to confirm that the two cDNAs we characterized from platy corre-

<i>AbTPI</i>	MAPTRKFFVGGNWKLN	GDGKKS	LGELIHTLNSGKINADTDV	VCGAPTIIYLD	FVRSKLD	SKM	60
<i>T</i>	---S.....M.....	A.L...E.....	S.....A.....		P.I	
<i>DrTPI-B</i>	..SG.....M.....	IE..AN..A.L.P.E.....		YA.....	PNI	
<i>XmTPI-B</i>	..S.....M.....Q.....	VDSNVE.....	S.....A.....		A.F	
<i>B</i>	---S.....M.....	A.L...E.....	S.....A.....		P.I	
<i>DrTPI-A</i>	..SS.....M.....	E M	TASL DE	S.....YA.....		QRI	
<i>XmTPI-A</i>	..V.R.....M.N	E EQ..T..	TASL DDQ E.....	A.S.....A..L..		PRI	
<i>A</i>	---S.....M.....	E M	TASL DE E.....	S.....A.....		PRI	
<i>AbTPI</i>	GVAEQNCYKVAKGAF	TGEVSPAMIKDCG	VS	WVILGHSERRHVF	GESDELIGQKVA	HALQE	120
<i>T</i>I.....		N.....			ES	
<i>DrTPI-B</i>	D.....I.....		K.....			EN	
<i>XmTPI-B</i>S.....	I.....	N.....			T..ES	
<i>B</i>I.....		N.....			ES	
<i>DrTPI-A</i>P.....	I.....	ID			C.ES	
<i>XmTPI-A</i>	..S.....I.....		S AD			ES	
<i>A</i>I.....		D			ES	
<i>AbTPI</i>	GLGVIACIGEKLDERE	AGITEKVVFEQTKA	IADNVKDWSKV	VLAYEPVVAIGT	GKTAS	PQ	180
<i>T</i>V.....						
<i>DrTPI-B</i>A.F.....						
<i>XmTPI-B</i>G.....	A.V.....					
<i>B</i>A.V.....						
<i>DrTPI-A</i>	D E T D		V.....	TR.....		E	
<i>XmTPI-A</i>	DT.E.....DT..... E	IYA.....	QV.....E.T.G.....			T. E	
<i>A</i>	D E T E		V.....			E	
<i>AbTPI</i>	QAQEVHEKLRQWLK	THVSEEVGGAVRI	IYGGSVTAATCKEL	ASKPDVDGFLV	GGASLKPE	240	
<i>T</i>N.DA.ANS.....		GG.....	Q.....			
<i>DrTPI-B</i>D.....N..A.ANS.....		GG.....	QK.L.....			
<i>XmTPI-B</i>A.....N.DA.ANS.....		GG.....	QK.....			
<i>B</i>N.DA.ANS.....		GG.....	QK.....			
<i>DrTPI-A</i> ERAN.DA.ANS.....		GGN.....	AQA.....			
<i>XmTPI-A</i> AFTAN.DD.ADSL.....		G.N.R.....	QG.....			
<i>A</i> EAN.DA.ANS.....		GGN.....	QA.....			
<i>AbTPI</i>	FIDIINAFQ	249	pI = 6.3				
<i>T</i>KA		pI = 6.6				
<i>DrTPI-B</i>KA		pI = 6.5				
<i>XmTPI-B</i>KA		pI = 7.7				
<i>B</i>KA		pI = 7.7				
<i>DrTPI-A</i>	.V.....RS		pI = 4.7				
<i>XmTPI-A</i>	.V.....RA		pI = 4.4				
<i>A</i>	.V.....RA		pI = 4.7				

FIGURE 1.—Amino acid sequences for the fish TPI genes reported here and amino acid sequences for the inferred ancestral teleost (*T*), ancestral TPI-A (*A*), and ancestral TPI-B (*B*) genes. Negative amino acids in the neural proteins (extant and inferred) are shown in boldface. Protein surface amino acids are shaded. Amino acids were considered to be on the protein surface if they had an accessible surface area $\geq 1 \text{ \AA}^2$ as determined by the Surface computer program (COLLABORATIVE COMPUTATIONAL PROJECT 1994) using the chicken TPI protein as a reference structure. Predicted *pI* values are shown at the 3' end of each protein.

sponded to the tissue-specific isozymes reported by MORIZOT and SCHMIDT (1990): 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 in platy. Primers specific for the *XmTPI-B* gene amplified a product from every tissue examined (eye, brain, ovary, liver, gut, and muscle), whereas primers specific for the *XmTPI-A* gene amplified a product only from eye, brain, and ovary (PCR products were sequenced and verified as *XmTPI-B* or *XmTPI-A*). This pattern of message expression matches the pattern of TPI message expression we found in zebrafish and the pattern of protein expression reported by MORIZOT and SCHMIDT (1990).

PCR products from shortnose sturgeon samples contained one cDNA consisting of 1063 nucleotides, includ-

ing an ATG start codon, TGA stop codon, 38 nucleotides of 5' untranslated sequence, and 216 nucleotides of 3' untranslated sequence. The cDNA open reading frame codes for a 249-amino-acid protein (herein referred to as *AbTPI*; Figure 1). Interestingly, sequence analysis of several sturgeon 3' RACE products revealed a second TPI cDNA. We sequenced 144 bp of the 3' end of coding region of this cDNA (including the stop codon) and 100 bp of the 3' untranslated sequence. The coding regions of the two sequences differ by 5% at the nucleotide level (7 changes in 133 bp) with no nonsynonymous changes (including termination codons). The 3' untranslated sequences, however, differ by 50%, implying that the gene sequences are not likely alleles at the same locus, but represent separate loci. In contrast, over this same portion of the coding region, the two zebrafish

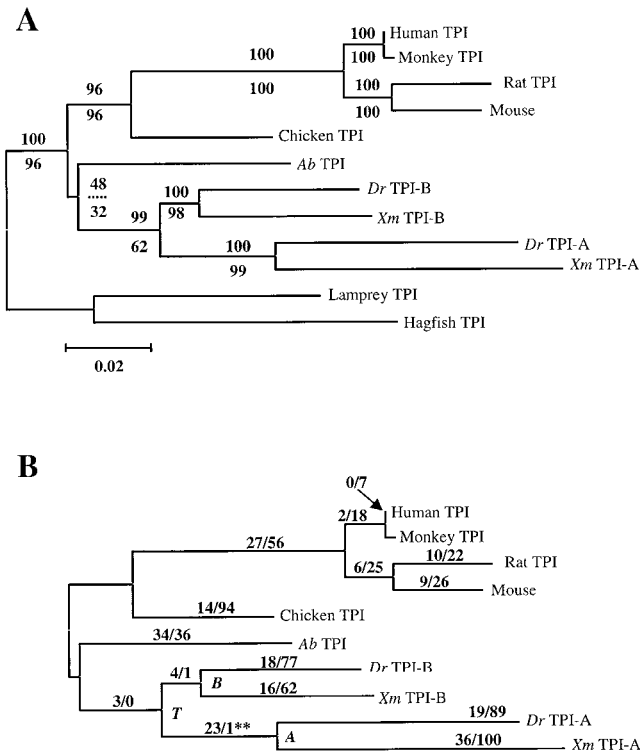


FIGURE 2.—Phylogenetic analysis of vertebrate TPI sequences. (A) Neighbor-joining phylogeny summarizing inferred evolutionary relationships among vertebrate TPI sequences. The tree was rooted with hagfish and lamprey TPI sequences (see MATERIALS AND METHODS). Numbers along each branch represent the proportion of 1000 bootstrap replicates supporting that node (analysis of TPI nucleotide sequences above, amino acid sequences below). (B) Estimated number of synonymous and nonsynonymous changes along branches of the TPI gene tree. 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*. ** indicates the only comparison in which *n/s* is significantly greater than *N/S* (Fisher's exact test, $P = 0.0005$). T, B, and A indicate the nodes corresponding to the ancestral single teleost TPI sequence, ancestral TPI-B sequence, and ancestral TPI-A sequence, respectively; see text and Figure 1 for details.

TPI sequences differ by 26% (35 changes) and the two platy TPI sequences by 25% (33 changes). Shortnose sturgeon are thought to be recent polyploids including a recent genome duplication event unique to the species (BLACKLIDGE and BIDWELL 1993), which explains the close sequence similarity between the two cDNAs. Given the lack of amino acid differences between the sturgeon TPI proteins, we have based our analysis of TPI evolution on the single completely sequenced TPI coding region from sturgeon.

Phylogenetic analyses: Neighbor-joining analyses of chordate TPI nucleotide sequences produce a tree (Figure 2A) whose topology is consistent with previously reported TPI gene trees (NIKOH *et al.* 1997; KURAKU *et al.* 1999). Bootstrap analysis strongly supported mono-

phyly of the four teleost TPI sequences and, within this group, the two TPI-A and two TPI-B proteins. Bootstrap analysis provided no support for grouping either the two zebrafish TPI genes or the two platy genes (these sequences did not group in any of the bootstrap trees). This indicates that the pair of TPI sequences observed in teleost fish result from a gene duplication event that occurred sometime early in the radiation of ray-finned fishes, after the separation of the Acipenseriform and teleost fishes (~200 mya; GRANDE and BEMIS 1991), but before the radiation of teleost fishes (~100 mya; PATTERSON 1993).

Rooting the ingroup taxa with the two agnathid sequences places the root between actinopterygian fishes and tetrapods as would be expected, given currently held views on vertebrate phylogeny. Monophyly of the actinopterygian fishes (placement of the sturgeon TPI with the teleost TPI sequences), however, is only weakly supported using both nucleotide and amino acid characters. We are inclined to accept this topology in that other placements of the sturgeon TPI sequences within the TPI tree would require a greater number of gene duplication and loss events than required by this topology (see DISCUSSION).

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*) and synonymous (*S*) changes (ZHANG *et al.* 1997). Positive, diversifying selection can lead to an increase in *n/s* relative to *N/S*, whereas purifying selection can lead to a decrease in *n/s* relative to *N/S* (ZHANG *et al.* 1997). *n* and *s* were calculated directly for each branch of the 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% for the amino acids and from 83 to 98% for the nucleotides. Using a ratio of transitions to transversions of 1.1 (calculated from the data set), *N* and *S* were calculated to be 536 and 196, respectively ($N/S = 2.7$). The *n/s* ratio exceeds 2.7 along three branches of the tree (Figure 2B): the branch leading to the ancestral single teleost gene (*F-T* branch), the branch leading to the ancestral B gene (*T-B* branch), and the branch leading to the ancestral A gene (*T-A* branch). Using Fisher's exact test, the difference between the *n/s* and *N/S* values was statistically significant only along the *T-A* branch ($P = 0.005$).

Four sites of the inferred *T*, *B*, or *A* sequences (amino acids 27, 90, 194, and 242; Figure 1) were deemed ambiguous because the most likely amino acid was less than twice as probable as the next most likely amino acid (*e.g.*, ZHANG *et al.* 1998). If these four sites are excluded from the analysis, *n/s* still significantly exceeds *N/S* ($P = 0.003$) along the *T-A* branch. Additionally,

TABLE 1
Number of amino acid substitutions that alter or conserve a given amino acid property along branches of the TPI phylogeny

Property	Reconstructed changes along branch(s)		<i>P</i>
	T-A	Tree	
Charge			
Conserve	12(10)	133(121)	0.007(0.014)
Alter	8(7)	21(21)	
Polarity			
Conserve	15(13)	110(101)	0.484(0.426)
Alter	5(4)	44(42)	
Size			
Conserve	14(11)	97(90)	0.363(0.556)
Alter	6(6)	57(53)	

Branches are as defined in Figure 1; tree refers to summed changes along all branches, excluding T-A. *P* is the result of a Fisher's exact test that the observed number of changes within properties is homogeneous across branch designations. Results of a similar analysis excluding four ambiguously reconstructed amino acid positions (see Figure 1 legend for details) are shown in parentheses.

ancestral sequences were reconstructed using alternative placements of the sturgeon sequence (*e.g.*, basal to all other gnathostome TPI sequences). Because the methods of ZHANG *et al.* (1997) assume a basal polytomy, the reconstructed sequences, and *n* and *s*, were unaffected, and inference of *n* and *s*, *n/s* still significantly exceeds *N/S* ($P = 0.005$) along the branch leading to the ancestral TPI-A.

Nonrandom amino acid substitution: Different patterns of amino acid substitution across two evolutionary periods suggest differences in selective pressures between those periods. Predicted amino acid sequences (Figure 1) were used to compare the amino acid substitutions along the T-A branch with those across the rest of the TPI 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, size, or polarity. Of the 20 amino acid changes between nodes T and A, 8 result in a negatively charged amino acid. No changes along this branch resulted in positively charged amino acids. There are 154 amino acid changes across the rest of the TPI gene tree; 21 of these are to or from negative amino acids (data not shown). These ratios, 12/8 and 133/21, are significantly different (Fisher's exact test, $P = 0.007$; Table 1). Excluding the four ambiguous ancestral sites (see above), 17 amino acid changes occur along the T-A branch, 7 of which are to negative amino acids; 142 changes occur across the rest of the tree, 21 of which are to or from negative or positive amino acids. The ratio of changes along the T-A branch is still significantly different from the ratio ob-

served for the rest of the tree (10/7 *vs.* 121/21; $P = 0.014$); substitutions to a negative amino acid occur significantly more often along the T-A branch than across the rest of the tree.

DISCUSSION

Gene phylogeny and selective pressures: Phylogenetic analysis of vertebrate TPI cDNA sequences confirms that the TPI gene duplicated early in the evolution of teleost fish and indicates that the two TPI isozymes found in zebrafish and southern platy are products of this duplication. Monophyly of the actinopterygian fish sequences (placement of the sturgeon TPI sequence with the teleost TPI sequences) is only weakly supported by both the TPI nucleotide and amino acid data sets (Figure 2A). Placement of the sturgeon TPI elsewhere in the tree, however, would require a greater number of duplications and would lose events to reconcile the gene tree topology with currently accepted views on vertebrate phylogeny. For example, placement of the sturgeon TPI outside of both tetrapod and teleost sequences would require duplication of the TPI gene prior to the radiation of jawed vertebrates, a subsequent loss of one locus in the sturgeon lineage, either a single loss of the other locus prior to the divergence of teleosts and tetrapods or two independent losses, one in teleosts and the other in tetrapods, and, finally, duplication of the TPI gene in teleosts. Other placements of the sturgeon sequence require similar numbers of additional duplication and loss events. Given that the topology shown in Figure 2A requires the fewest number of duplication and loss events to explain the pattern of expression of TPI across taxa we are inclined to accept this topology, even though bootstrap support for placement of the sturgeon sequence is low. In any event, reconstruction of ancestral sequences (below) and inference of patterns of sequence change over evolutionary time rely on an *unrooted* tree. Placement of the sturgeon TPI sequence as sister to the teleost TPI sequences, with the tetrapod TPI sequences basal to all other gnathostome sequences, does not change the results of this analysis.

LDH-C, the negatively charged LDH neural isozyme of teleost fishes, results from a gene duplication that occurred during roughly this same time period (QUATTRO *et al.* 1993, 1995; STOCK *et al.* 1997). Studies of the evolution of the *Hox* (AMORES *et al.* 1998) and *Dlx* (STOCK *et al.* 1996; NEIDERT *et al.* 2001) genes also suggest duplications in both these gene families early in the radiation of ray-finned fishes. The duplication of these last two gene families early in the evolution of higher fishes is thought to result from duplication of entire chromosomes or perhaps the entire genome (STOCK *et al.* 1996; AMORES *et al.* 1998; and NEIDERT *et al.* 2001). It is possible that the TPI and LDH neural isozymes result from this large-scale duplication event.

Comparison of the observed number of nonsynony-

mous and synonymous changes (n and s ; Figure 2B) with the expected number of nonsynonymous and synonymous changes (N and S) across each branch of the TPI gene tree shows that, in general, n/s is less than N/S . This indicates that the TPI gene has evolved under purifying selection throughout much of its evolutionary history, as is expected for a protein coding sequence. However, the ratio of n/s is significantly greater than N/S across the branch between the ancestral single teleost TPI protein (T) and the ancestral TPI-A protein (A). This suggests that, following duplication of the ancestral gene, TPI-A evolved through a period of positive selection. Along the terminal TPI-A branches n/s is again less than N/S , indicating a return to purifying selection.

Pattern of amino acid substitution: The high net negative charge predicted for the TPI-A ancestral protein results from the number of changes (eight) along the T - A branch from amino acids with positive or neutral net charges to amino acids with negative net charges (Figure 1). Eight substitutions that alter amino acid charge is particularly striking in that charge changes are generally rare in protein evolution (PEETZ *et al.* 1986; XIA and LI 1998). Accumulation of negatively charged amino acids could have resulted from an overall acceleration in the rate of amino acid change during positive selection along this branch (diversifying selection) or, more interestingly, from an acceleration specifically in the rate of change to negative amino acids along this branch (directional selection).

To distinguish between diversifying and directional selection in TPI-A evolution, the amino acid substitutions occurring along the T - A branch were compared with all other substitutions across the TPI gene tree. Forty percent of the amino acid substitutions along the T - A branch (8 of 20) were from either neutral or positive amino acids to negative amino acids. There are no changes along this branch from negative amino acids to neutrally or positively charged residues. In comparison, significantly fewer substitutions (14%, or 21 of 154) along the other branches are to or from negative amino acids. When substitutions are grouped by other physicochemical characteristics (size or polarity), the amino acid changes along the T - A are indistinguishable from those along the other branches. These two points indicate accelerated accumulation of negative amino acids by the TPI-A protein during a period of positive selection. The TPI-A gene evolved through a period of directional selection, resulting in a more negatively charged protein.

The amino acid substitutions observed along the T - A branch occur across the surface of the protein (Figure 1) and avoid the catalytic center and dimer interface residues (as identified by WIERENGA *et al.* 1991). That the changes should occur across the surface of the protein is not unexpected; internal changes, especially those involving changes in amino acid charge, would be more

likely to alter the three-dimensional structure of the protein and to disrupt protein function than changes across the surface. That the changes do not involve the active site or dimerization interface is also expected for a protein evolving within functional constraints, and the TPI-A protein does form dimers and function at least well enough to allow *in vitro* staining (PONTIER and HART 1981). This apparent avoidance of change at critical amino acids is also consistent with theoretical models that suggest that following duplication daughter genes do not radically change their function but act to specialize within the broader function of the parent gene (*e.g.*, HUGHES 1994).

During the course of these analyses we observed a pattern of nucleotide substitution along the T - A branch that is more difficult to explain than the aforementioned periods of positive and directional selection. Interestingly, all changes to negatively charged amino acids along the T - A branch require single nucleotide changes (*e.g.*, a lysine-to-glutamic acid substitution requires a single nucleotide change), suggesting that amino acid substitution might have been a function of the number of evolutionary steps (nucleotide changes) required during the evolution of the neural isozymes. To test if this apparent correlation was due simply to chance, we calculated the expected number of mutational events required to yield the observed charge profile, assuming that amino acid substitution occurred at random along the surface of the protein. One hundred sets of eight amino acids were created by random sampling without replacement from the pool of all non-negative surface amino acids of the TPI-T ancestral protein (104 possible amino acids; Figure 1). The minimum number of evolutionary steps required to change each amino acid within a set to a negative amino acid was then calculated. For example, lysine requires two nucleotide changes to convert into an aspartic acid, but only one to convert into a glutamic acid; therefore, a lysine requires only one step to become a negative amino acid. The minimum number for each amino acid was then summed within each 8-amino-acid set to give a set total (with a possible range from 8 to 16 steps). Across the 100 replicates, the average number of evolutionary steps required was ~ 11 , with a range of 8–14 steps. Only a single replicate required the minimum of 8 steps. It is unlikely then ($P = 0.01$) that substitutions requiring the observed number of evolutionary steps would have occurred due to chance, indicating that the substitutions to negative amino acids observed along the T - A branch are correlated with the number of evolutionary steps required to make those substitutions.

This is not surprising if selection acted to fix any substitution to a negative amino acid across the surface of the protein; amino acid substitutions that require the fewest possible number of steps (one) are most likely to occur and, therefore, should be observed more often than those requiring a greater number of steps. This

correlation between observed amino acid substitution and number of evolutionary steps is surprising, however, if selection acted to fix negative amino acids at specific key points across the protein; this would require that key positions across the protein all happened to require only a single evolutionary step to change to a negative amino acid. This analysis, then, suggests that following the duplication of the TPI gene selection acted on the overall net charge of the TPI-A protein and not on key amino acids. Evolution of the TPI-A protein seems to have followed one of the "shortest paths" to overall high net negative charge.

That selection acted on the overall charge, and not on key amino acids, is also suggested by slight differences in the negative amino acids present in *Df*TPI-A and *Xm*TPI-A (Figure 1). While both neural proteins share a strongly net negative charge and the majority of negative amino acids, there are negative amino acids that are unique to the neural isozyme of either species. The selective constraints that are present (purifying selection predominates along each terminal node; Figure 2B) seem to conserve the overall charge of the molecule, but allow some variation in the amino acids that contribute to the charge. This is in direct contrast to cases such as the opsin gene family (reviewed in YOKOYAMA 1997) in which selection appears to act at specific amino acid sites promoting specific substitutions.

Analysis of vertebrate TPI gene coding regions supports earlier proposals that the high net negative charge of neural isozymes is a protein level adaptation. We have not attempted to address the exact nature of the selective pressures involved, but the large number of gene families that include a negatively charged neural isozyme suggests that these pressures are a general feature of the neural environment, not a phenomenon specific to the TPI gene family. The physiological or biochemical advantage of negative neural isozymes is not immediately apparent. However, a high intracellular, relative to extracellular, concentration of negatively charged organic molecules is necessary for maintenance of the resting potential of neurons (NICHOLLS *et al.* 1992). Accumulation of a large net negative charge by multiple neural proteins might in some way be associated with this requirement. This is consistent with the conclusion that selection appears to have acted on the overall net charge of the protein, not on changes at specific amino acids. Further examination of other gene families that include neural isozymes may shed further light on this topic.

We thank David Stock, Joseph Staton, Robbie Young, Jacqueline Litzgus, Jim Grady, Robert Friedman, and Austin Hughes for helpful comments on this manuscript, and Lirong Shi for technical assistance. We are also grateful to Lukasz Lebioda for assistance with the amino acid surface area analysis. 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)], and the National Science Foundation (OCE-9814172).

LITERATURE CITED

- AMORES, A., A. FORCE, Y.-L. YAN, L. JOLY, C. AMEMIYA *et al.*, 1998 Zebrafish *hox* clusters and vertebrate genome evolution. *Science* **282** (27): 1711–1714.
- BISHOP, J. G., A. M. DEAN and T. MITCHELL-OLDS, 2000 Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. USA* **97**: 5322–5327.
- BLACKLIDGE, K. H., and C. A. BIDWELL, 1993 Three ploidy levels indicated by genome quantification in *Acipenseriformes* of North America. *J. Hered.* **84**: 427–430.
- BURGER, A., M. EPPENBERGER, U. WIESMANN and R. RICHTERICH, 1963a Isozyme der Creatin-Kinase. *Helv. Physiol. Acta* **21**: C6–C10.
- BURGER, A., R. RICHTERICH and H. AEBI, 1963b Die Heterogenität der Creatin-Kinase. *Biochem. Z.* **339**: 305–314.
- CHAMPION, M. J., and G. S. WHITT, 1976 Differential gene expression in multilocus isozyme systems of the developing green sunfish. *J. Exp. Zool.* **196**: 263–282.
- CHENG, J., L. M. MIELNICKI, S. C. PRUITT and L. E. MAQUAT, 1990 Nucleotide sequence of murine triosephosphate isomerase cDNA. *Nucleic Acids Res.* **18**: 4261.
- COLLABORATIVE COMPUTATIONAL PROJECT, 1994 Surface. *Acta Crystallogr.* **50**: 760–763.
- DUDA, T. F., JR., and S. R. PALUMBI, 1999 Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl. Acad. Sci. USA* **96**: 6820–6823.
- FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.
- FISHER, S. E., and G. S. WHITT, 1978 Evolution of isozyme loci and their differential tissue expression. Creatine kinase as a model system. *J. Mol. Evol.* **12**: 25–55.
- FISHER, S. E., J. B. SHAKLEE, S. D. FERRIS and G. S. WHITT, 1980 Evolution of five isozyme systems in the chordates. *Genetica* **52**: 73–85.
- FROHMAN, M. A., 1990 RACE: rapid amplification of complementary DNA ends, pp. 28–38 in *PCR Protocols: A Guide to Methods and Applications*, edited by M. A. INNIS, D. H. GELFAND, J. J. SNINSKY and T. J. WHITE. Academic Press, New York.
- GRANDE, L., and W. E. BEMIS, 1991 Osteology and phylogenetic relationships of fossil and recent paddlefish (polyodontidae) with comments on the interrelationships of the acipenseriformes. *J. Vert. Paleo. Memoir* **1**: 1–121.
- GRANTHAM, R., 1974 Amino acid difference formula to help explain protein evolution. *Science* **185**: 862–864.
- HUGHES, A. L., 1994 The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. Ser. B* **256**: 119–124.
- HUGHES, A., and M. K. HUGHES, 1993 Adaptive evolution in the rat olfactory receptor gene family. *J. Mol. Evol.* **36**: 249–254.
- JONES, D. T., W. R. TAYLOR and J. M. THORTON, 1992 The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**: 275–282.
- JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–123 in *Mammalian Protein Metabolism*, edited by H. N. MUNRO. Academic Press, New York.
- KIMURA, M., 1980 A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
- KOLB, E., J. I. HARRIS and J. BRIDGEN, 1974 Triosephosphate isomerase from the coelacanth. An approach to the rapid determination of an amino acid sequence with small amounts of material. *Biochem. J.* **137**(2): 185–197.
- KUMAR, S., K. TAMURA and M. NEI, 1993 MEGA: molecular evolutionary genetics analysis, version 1.0. University Park: Pennsylvania State University.
- KURAKU, S., D. HOSHIYAMA, K. KATOH, H. SUGA and T. MIYATA, 1999 Monophyly of lampreys and hagfish supported by nuclear DNA-coded genes. *J. Mol. Evol.* **49**: 729–735.
- LI, W.-H., 1983 Evolution of duplicate genes and pseudogenes, pp. 14–37 in *Evolution of Genes and Proteins*, edited by M. NEI and R. K. KOEHN. Sinauer Associates, Sunderland, MA.
- MAQUAT, L. E., R. CHILCOTE and P. M. RYAN, 1985 Human triosephosphate isomerase cDNA and protein structure. Studies of triosephosphate isomerase deficiency in man. *J. Biol. Chem.* **260**: 3748–3753.

- MARANGOS, P. J., and D. E. SCHMECHEL, 1987 Neuron specific enolase, a clinically useful marker for neurons and neuroendocrine cells. *Annu. Rev. Neurosci.* **10**: 269–295.
- MESSIER, W., and C.-B. STEWART, 1997 Episodic adaptive evolution of primate lysozymes. *Nature* **385**: 151–154.
- MORIZOT, D. C., and M. E. SCHMIDT, 1990 Starch gel electrophoresis and histochemical visualization of proteins, pp. 23–80 in *Applications of Electrophoresis and Isoelectric Focusing in Fisheries Management*, edited by D. H. WHITMORE. CRC Press, Boca Raton, FL.
- NEIDERT, A. H., V. VIRUPANAVAR, G. W. HOOVER and J. A. LANGE-LAND, 2001 Lamprey Dlx genes and early vertebrate evolution. *Proc. Natl. Acad. Sci. USA* **98** (4): 1665–1670.
- NICHOLLS, J. G., A. R. MARTIN and B. G. WALLACE, 1992 *From Neuron to Brain: A Cellular and Molecular Approach to the Function of the Nervous System*, Ed. 3. Sinauer Associates, Sunderland, MA.
- NIKOH, N., N. IWABE, K. KUMA, M. OHNO, Y. SUGIYAMA *et al.*, 1997 An estimate of divergence time of parazoa and eumetazoa and that of cephalochordata and vertebrata by aldolase and triosephosphate isomerase clocks. *J. Mol. Evol.* **45**: 97–106.
- OHTA, T., 1991 Multigene families and the evolution of complexity. *J. Mol. Evol.* **33**: 34–41.
- OLD, S. E., and H. W. MOHRENWEISER, 1988 Nucleotide sequence of the triosephosphate isomerase gene from *Macaca mulatta*. *Nucleic Acids Res.* **16**: 9055.
- PATTERSON, C., 1993 An overview of the early fossil record of acanthomorphs. *Bull. Mar. Sci.* **52** (1): 29–59.
- PEETZ, E. W., G. THOMSON and P. W. HEDRICK, 1986 Charge changes in protein evolution. *Mol. Biol. Evol.* **3**: 84–94.
- PENHOET, E., T. RAJKUMAR and W. J. RUTTER, 1966 Multiple forms of fructose diphosphate aldolase in mammalian tissues. *Proc. Natl. Acad. Sci. USA* **56** (4): 1275–1282.
- PONTIER, P. J., and N. H. HART, 1981 Developmental expression of glucose and triose phosphate isomerase genes in teleost fishes (*Brachydanio*). *J. Exp. Zool.* **217**: 53–71.
- QUATTRO, J. M., H. A. WOODS and D. A. POWERS, 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.
- QUATTRO, J. M., D. D. POLLOCK, M. POWELL, H. A. WOODS and D. A. POWERS, 1995 Evolutionary relations among vertebrate muscle-type lactate dehydrogenases. *Mol. Mar. Biol. Biotech.* **4**: 224–231.
- SAIKI, R. K., D. H. GELFAND, S. STOEFFEL, S. J. SCHARF, R. HIGUCHI *et al.*, 1988 Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.
- SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- SHAKLEE, J. B., K. L. KEPES and G. S. WHITT, 1973 Specialized lactate dehydrogenase isozymes: the molecular and genetic basis for the unique eye and liver LDHs of teleost fishes. *Exp. Zool.* **185**: 217–240.
- STOCK, D. W., D. L. ELLIES, Z. ZHAO, M. EKKER and F. H. RUDDLE, 1996 The evolution of the vertebrate Dlx gene family. *Proc. Natl. Acad. Sci. USA* **93**: 10858–10863.
- STOCK, D. W., J. M. QUATTRO, G. S. WHITT and D. A. POWERS, 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.
- STRAUS, D., and W. GILBERT, 1985 Genetic engineering in the precambrian: structure of the chicken triosephosphate isomerase gene. *Mol. Cell. Biol.* **5**: 3497–3506.
- SWOFFORD, D. L., 1999 PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods, version 4.0b). Sinauer Associates, Sunderland, MA.
- SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL and D. M. HILLIS, 1996 Phylogenetic inference, pp. 407–514 in *Molecular Systematics*, Chap. 11, edited by D. M. HILLIS, C. MORITZ and B. K. MABLE. Sinauer Associates, Sunderland, MA.
- TANAKA, T., and M. NEI, 1989 Positive Darwinian selection observed at the variable region genes of immunoglobins. *Proc. Natl. Acad. Sci. USA* **76**: 4521–4525.
- THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 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.
- WHITT, G. S., 1970 Developmental genetics of the lactate dehydrogenase isozymes of fish. *J. Exp. Zool.* **175**: 1–36.
- WIERENGA, R. K., M. E. M. NOBLE, G. VRIEND, S. NAUCHE and W. G. J. HOL, 1991 Refined 1.83 Å structure of trypanosomal triosephosphate isomerase crystallized in the presence of 2.4 M ammonium sulphate. *J. Mol. Biol.* **220**: 995–1015.
- WILLETT, C. S., 2000 Evidence for directional selection acting on pheromone-binding proteins in the genus *Choristoneura*. *Mol. Biol. Evol.* **17**: 553–562.
- XIA, X., and W.-H. LI, 1998 What amino acid properties affect protein evolution? *J. Mol. Evol.* **47**: 557–564.
- YOKOYAMA, S., 1997 Molecular genetic basis of adaptive selection: examples from color vision in vertebrates. *Annu. Rev. Genet.* **31**: 315–336.
- ZHANG, J., S. KUMAR and M. NEI, 1997 Small-sample tests of episodic evolution: a case study of primate lysozymes. *Mol. Biol. Evol.* **14**: 1335–1338.
- ZHANG, J., H. F. ROSENBERG and M. NEI, 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**: 3708–3713.

Communicating editor: S. YOKOYAMA