Evolutionary Analysis of TATA-less Proximal Promoter Function

Douglas L. Crawford, Jeff A. Segal,[1] *and J. Lynn Barnett*

Division of Molecular Biology and Biochemistry, University of Missouri, Kansas City

Many molecular studies describe how components of the proximal promoter affect transcriptional processes. However, these studies do not account for the likely effects of distant enhancers or chromatin structure, and thus it is difficult to conclude that the sequence variation in proximal promoters acts to modulate transcription in the natural context of the whole genome. This problem, the biological importance of proximal promoter sequence variation, can be addressed using a combination of molecular and evolutionary analyses. Provided here are molecular and evolutionary analyses of the variation in promoter function and sequence within and between populations of *Fundulus heteroclitus* for the lactate dehydrogenase-B (*Ldh-B*) proximal promoter. Approximately one third of the *Ldh-B* proximal promoter contains interspersed regions that are functionally important: (1) they bind transcription factors in vivo, (2) they effect a change in transcription as assayed by transient transfection into two different fish cell lines, and (3) they bind purified transcription factors in vitro. Evolutionary analyses that compare sequence variation in these functional regions versus the nonfunctional regions indicate that the changes in the *Ldh-B* proximal promoter sequences are due to directional selection. Thus, the *Ldh-B* proximal promoter sequence variations that affect transcriptional processes constitute a phenotypic change that is subject to natural selection, suggesting that proximal promoter sequence variation affects transcription in the natural context of the whole genome.

Introduction

Does natural sequence variation in protein/DNA-binding sites play an important role in modulating transcription rates? Evolutionary analyses can help address this question. Molecular traits are biologically important if they evolve by natural selection, because natural selection only acts on a trait if it affects fitness. Evolutionary analyses can thus supplement experimental analyses and provide a means to determine if subtle variations in protein/DNA-binding sites are biologically important (i.e., effecting a phenotypic change that affects an organism's longevity, reproductive fitness, or probability of survival). One method used to discern if a DNA sequence is evolving by natural selection is to examine the pattern of sequence variation among silent substitutions versus substitutions that alter function or structure (McDonald and Kreitman 1991). This method has been applied to the sequence variation in the enhancer that regulates the developmental expression of the *Drosophila* even-skipped homeobox gene (*eve* [Ludwig and Kreitman 1995]). Molecular analyses suggest that protein/DNA-binding sequences within the *eve* enhancer should have few substitutions within and among closely related species, because even a few changes would severely alter developmental processes. However, the substantial sequence variation in these regions is not different from expected neutral patterns. Ludwig and Kreitman (1995) suggest that the random variation in the *eve* enhancer is tolerated because of the redundancy of enhancer elements. These results, unlike those of the molecular analyses, suggest that some sequence variation is acceptable and thus emphasize the contribution that evolutionary analyses can make toward understanding the significance of regulatory sequence variation. Reported here is an analysis of the *Fundulus heteroclitus* lactate dehydrogenase-B (*Ldh-B*) proximal promoter (DNA sequences within a few hundred base pairs of the start of transcription [Ernst and Smale 1995*b*]). We seek to determine if proximal promoter sequence variation is biologically important by establishing whether this regulatory region is evolving by natural selection.

DNA sequences in the proximal promoter bind transcription factors, accurately position RNA polymerase II, and direct the rate of polymerase initiation (Roeder 1991; Smale 1994; Ernst and Smale 1995*a*; Yean and Gralla 1997). One of these sequences in the proximal promoter is TATA at −30 bp (relative to the start of transcription), which binds the transcription factor TFIID; an essential factor for all mRNA transcription (Roeder 1991; Smale 1994). However, many promoters lack a TATA sequence (e.g., approximately 50% in *Drosophila* [Arkhipova 1995]). Among TATA-less promoters, there is little similarity in TFIID-binding sequences (Roeder 1991; Azizkhan et al. 1993; Smale 1994), and transcription may depend more on initiator sequences (Inr's), Sp1-binding sites, or interactions among binding sites within the proximal promoter (Martinez et al. 1994; Smale 1994; Emami, Navarre, and Smale 1995; Ernst and Smale 1995*b*; Weis and Reinberg 1997). Many studies demonstrate how different components of the proximal promoter interact and affect transcriptional processes, but these studies do not account for the likely effects of distant enhancers and chromatin structure. Thus, it is difficult to conclude that the sequence variation in proximal promoters acts to modulate transcription in the natural context of the whole genome. This question, the biological importance of proximal promoter sequence variation, can be addressed using a combination of molecular and evolutionary analyses.

The *Ldh-B* locus in the teleost fish *Fundulus heteroclitus* provides a unique opportunity to apply both molecular and evolutionary approaches to proximal promoter function. In vivo transcription rates (measured by

[1] Present address: Lilly Research Laboratories, CNS Discovery, Indianapolis, Indiana.

Key words: transcription, TATA-less, directional selection, Sp1.

Address for correspondence and reprints: Douglas L. Crawford, Division of Molecular Biology and Biochemistry, University of Missouri, Kansas City, Missouri 64110. E-mail: crawd@cctr.umkc.edu.

Mol. Biol. Evol. 16(2):194–207. 1999
© 1999 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

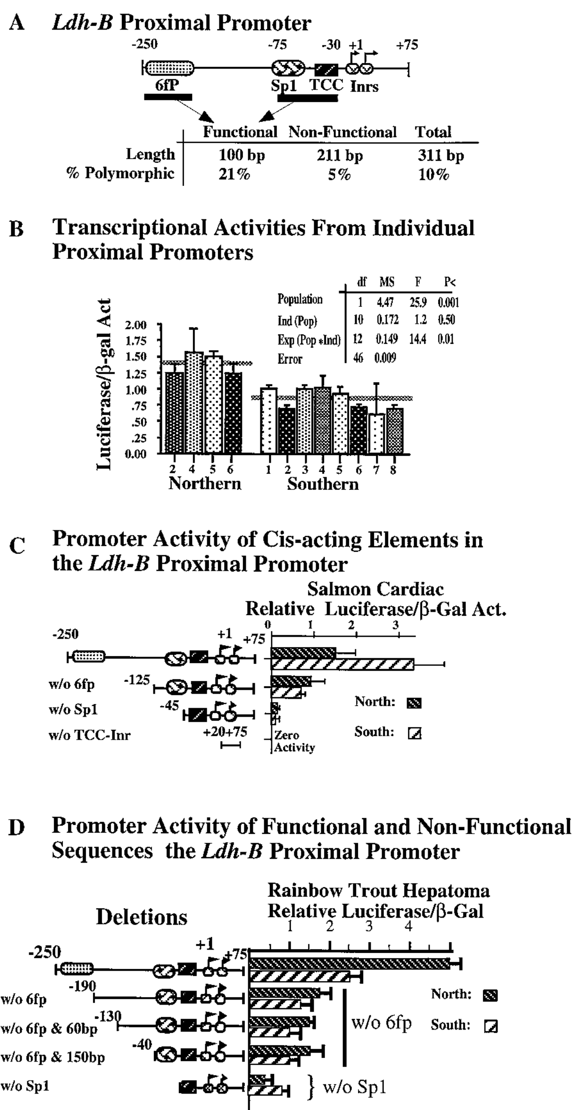

FIG. 1.—Functional properties of the *Ldh-B* proximal promoter. *A,* The *Ldh-B* proximal promoter is diagrammed, and the length and percentages of polymorphic sites are listed for functional (protein/DNA-binding sites) and nonfunctional regions. The *Ldh-B* proximal promoter has TCC repeats instead of a TATA box, initiation of transcription sequences (Inr's) near the two major transcriptional start sites (bent arrows), and Sp1-like binding sequences (Segal et al. 1996). Underlined with a thick black line are the three regions that bind proteins in vivo in *Fundulus* livers (''functional sequences'': 6fp, Sp1-like sequence, and TCC repeats; Segal et al. 1996). *B,* Variation in transcriptional activity among *F. heteroclitus* individual promoters (as depicted in fig. 1*A*) was assayed with rainbow trout hepatoma cells (RTH-149). The line extending from the boxes represents the standard error of means. Gray lines behind boxes are means for each population. Populations and experimental replicates are significant factors, but variation among individuals is not a significant factor (nested ANOVA: $P < 0.001$, $P < 0.01$, and $P > 0.5$, respectively). *C,* Promoter activity of functional regions, 6fp, Sp1, and TCC repeats, as assayed for a salmon cardiac cell line (CHH-1). The three in vivo footprinted regions were serially deleted, and promoter strength was assayed. Constructs with regions deleted are significantly different within a population (two-way ANOVA: $P < 0.02$). *D,* Promoter activity of 6fp and regions 5′ of the Sp1-like binding sequence. Deletions are 1 bp, approximately 60 bp, and 150 bp 3′ of 6fp. The 150-bp deletion is 4 bp 5′ of the Sp1-like sequence. The last construct contains only the TCC repeat and Inr (without the Sp1 sequences) and was constructedfrom synthesized oligonucleotides.

←

There is no significant difference between constructs with deletions between the 6fp and Sp1 sequences ($P > 0.25$).

nuclear run-on assays) of the *Ldh-B* locus are twofold greater in livers from northern versus southern populations of *F. heteroclitus* (Crawford and Powers 1992), resulting in a twofold greater LDH-B_4 enzyme concentration, which compensates for the colder northern temperatures (Crawford and Powers 1989, 1990, 1992; Segal and Crawford 1994). The *Ldh-B* proximal promoter is TATA-less, has Inr sequences associated with multiple start sites, a cluster of Sp1-like sequences (Segal et al. 1996), and considerable DNA sequence variation (fig. 1*A*). In vivo footprinting experiments using *F. heteroclitus* liver cells revealed that this sequence variation contributes to differential protein binding in three regions: the putative TFIID-binding site (TCC repeats), Sp1 sites, and a region immediately upstream, 6fp (Segal et al. 1996). The sequence variation within the Sp1 sites affects the SP1-protein-binding affinities: northern Sp1 sequences have higher affinities for recombinant SP1 protein than do the southern Sp1 sequences (unpublished data). Moreover, this sequence variation affects transcriptional processes (Segal et al. 1996; unpublished data). Similar Sp1-like sequences in other proximal promoters also affect transcription by altering interactions among different transcription factors (Bigger, Melnikova, and Gardner 1997). Does this type of sequence variation matter? That is, if sequence variation in the proximal promoter has measurable effects on molecular assays, will it also alter gene transcription in the natural context of the genome, and if so, does this variation in transcription matter to the organism's fitness?

Provided here are data (1) on the sequence variation of the *Ldh-B* proximal promoter within and between species, (2) establishing how promoter strength varies within and between populations, and (3) confirming the functional importance of the three footprinted regions in the *Ldh-B* proximal promoter (6fp, Sp1, and TCC repeats). These data are used to examine the biological importance of the sequence variation by providing a more specific hypothesis for evolutionary analyses of the proximal promoter: functional regions will have evolved a greater number of fixed differences between populations (a fixed difference is a nucleotide site at which all individuals within a population share a common nucleotide that is different from the nucleotide in the other population). The evolutionary analyses presented here indicate that the three interspersed functional regions are evolving due to directional selection. Thus, the sequence variation in the *Ldh-B* proximal promoter is biologically important, suggesting that this variation effects a phenotypic change by altering transcription rates.

Materials and Methods

Genomic DNA and Constructs

Genomic DNAs were isolated from 6 northern (Wiscasset, Maine) and 10 southern (Meridian, Ga.) *F.*

Table 1
Variable Sites in *Fundulus Ldh-B* Proximal Promoter

```
                 NUCLEOTIDE POSITION
                            11111111111111111111122222222222
              3444555668888900111133566677777789901111267799
TAXON         0123059695679768145968548901238921501345731201

G1 ......     TAT-AAGT-CAGTCAC-CCTTTTT2-3C3CCC-C4CC--6CCCGAG
G2 ......     TGT1CAGT-CAGTCAA-CTTTTTT--3C3CCC-C4CC--6CCCGAG
G3 ......     TAT-AAGT-CAGTCAC-CTTTTTT---C3CCC-C4CC--5CCCGAG
G4 ......     AAT-AAGT-CAGTCAC-CTTTTTC----3CCC-C4CC--6CCCGAG
G5 ......     TAT-AAGT-CAGTCAC--TGTTTC--3C3CCC-C4CC---CCCGAG
G6 ......     AAT1CAGT-CAGTCAC-CTTTTTT-----CCC-C4CC555CCCGAG
G7 ......     AAT1CAGC-CAGTCGC-CTTTTTC-----CCC-C4CC--6CCCGAG
G8 ......     AAT-AAGT-CAGTCAC-CTTTTTT-----CCC-C4CC-55CCCGAG
G9 ......     TAT-AAGT-CAGTCAC-CTTTTTC----3CCC-C4CC556CACGAG
G10 .....     TAT-AAGT-CAGTCAC-CTTTTTT----3CCC-C4AC-56TCCGAG
M1 ......     TAG-CGAT-CAGTCAC-CTTTTTC-C3C3CCCC5-CT-56CCCGAG
M2 ......     TAG-CGAT-CAGTCAC-CTTTTTC-C3C3CCCCG-CT-56CCCGAG
M3 ......     TAG-CGAT-CAGTCAC-CTTTTTC-C3C3CCCCG-CT-56CCCGAG
M4 ......     TAG-CGAT-CAGTCAC-CTTTTTC-C3C3CCCCG-CT-56CCCGAG
M5 ......     TAG-CGAT-CAGTCAC-CTTCTTC-C3C3CTCCG-CT-56CCCGAG
M6 ......     TAG-CGAT-CAGTCAC-CTTTTTC-C3C3CCCCG-CT-56CCCGAG
Fg20 .....    TAT-TAGTT--A--AC-CTTTTTT--3C3CCC-C4CC---CCGCAG
Fg13 .....    TA--TAGTT--A-CACACTTTCAT------C--C4CC---CCCGGA
— .......     00F00FF0******00*0000**00F000*0*FFF0F00000****
SGP .....     00F00FF|      0| ||||  00F000 0 00F|F000||
```

NOTE.—Variable sites among northern (Maine) and southern (Georgia) *F. heteroclitus* and *F. grandis.* Positions of variable nucleotides are given above sequences. Numbers within sequences refer to specific gaps (see gap code below). Total variable sites = 46. Within *F. heteroclitus,* 31 sites are variable (12 of these are gaps), and 8 are fixed differences between populations. The symbols below the sequences indicate (0) polymorphisms, (F) fixed differences, and (*) variation within *F. grandis* or between species. The second line of symbols (SGP) indicates whether homologous sites were variable in Schulte, Gomezchiarri, and Powers (1997); symbols are the same as above, and "|" indicates that a site was not variable in Schulte, Gomezchiarri, and Powers (1997). Comparisons of variation within *F. grandis* were not made because only a single individual was used by Schulte, Gomezchiarri, and Powers (1997). Gap code: 1 = CAT, 43–45; 2 = CCCTCCC, 167–173; 3 = CCT, 178–179, 181–183; 4 = GGCCA, 202–206; 5 = TCC, 229–231, 232–234, 236–237; 6 = TGC, 236–237.

heteroclitus individuals, as well as from two *Fundulus grandis* individuals (Panacea, Fla.). DNA was prepared as described previously (Segal et al. 1996), and detailed protocols are available at http://sgi.bls.umkc.edu/funnylab/DLCprotocols.html.

Genomic DNA isolated from different individuals was used to amplify approximately 300 bp (fig. 1*A*) of the *Ldh-B* proximal promoter. Polymerase chain reaction (PCR) used *Taq* polymerase and primers specific to the *Ldh-B* 5′ mRNA leader and a conserved region that is 5′ of 6fp (an in vivo footprinted region [Segal et al. 1996]). Each primer had either a *Kpn*I or a *Xho*I site added to the 5′ end. Each of these promoters was digested with *Kpn*I and *Xho*I, ligated into the *Kpn*I and *Xho*I site of pBSKS$^-$ (Stratagene) and sequenced twice in both directions with Sequenase (U.S. Biochemical). To verify individual polymorphisms, genomic DNA was reamplified and directly sequenced with an ABI automatic sequencer. All proximal promoter sequences are available from GenBank (accession numbers AF082687–AF082704).

Sequences were aligned with CLUSTAL V (Higgins, Bleasby, and Fuchs 1992). The initial alignment was modified to reduce the number of differences between populations: introducing a single 5-bp gap plus a second single-base-pair gap removed eight fixed differences within the Sp1-binding sequence. This alignment is the most conservative: modification greatly reduced the number of polymorphic characters, introduced fewer fixed differences between populations, and thus reduced the maximum-parsimony tree length. Every gap was further reduced to a single character (see table 1). The length of aligned sequences was reduced from 333 to 311 bp by these modifications. Alignments with and without reduced gaps are available at http://sgi.bls.umkc.edu/funnylab/dlc/ppalign.html.

To quantify promoter activity, individual proximal promoters were subcloned with the use of *Xho*I and *Kpn*I sites into a luciferase reporter plasmid, pGL3-basic (Promega, Madison). Two proximal promoters from each population did not have intact *Kpn*I sites, and therefore neither were subcloned or assayed. All *Ldh-B*-luciferase constructs were sequenced using an ABI automatic sequencer at the University of Missouri, Kansas City.

Deletion Analyses

Two sets of deletions were made. In the first set of deletions (fig. 1*C*), using the original genomic northern and southern clones (Segal et al. 1996), the three in vivo footprinted regions, 6fp, Sp1, and TCC repeats, were serially deleted with judicious use of restriction sites or PCR reactions. In the second set of deletions (fig. 1*D*), using M2 (Maine) and G6 (Georgia) DNA, unidirectional deletions of 6fp between 6fp and Sp1 sequences and the Sp1 sequences were completed with PCR using

Ldh-B-specific primers. For these two deletion experiments, there is no variation between the northern proximal promoters, and only a few nucleotide substitutions occur between the southern proximal promoters in the nonfunctional region. Deletions were ligated into *Xho*I-*Kpn*I or *Xho*I-*Hin*dIII-blunt sites of a luciferase reporter plasmid (pGL3 basic) and were sequenced in both directions. *Xho*I-*Hin*dIII-blunt subcloning was used for some Sp1-TCC-Inr- 5′ leader clones and for all TCC-Inr- 5′ leader clones. There was no significant difference in promoter strength between constructs that were ligated into *Kpn*I sites and those that were ligated into *Hin*dIII-blunt sites.

Transfections

Two fish cell lines were used for determinations of promoter activity: rainbow trout hepatoma cells (RTH-149, ATCC, Rockville, Md.; fig. 1*B* and *D*) and salmon cardiac cells (CHH-1, ATCC; fig. 1*C*). Transfections were performed as in Segal et al. (1996). Each individual promoter or deletion was transfected in triplicate on different days, and each transfection was assayed in triplicate. Promoter activity is expressed as net relative light units (luciferase activity minus the negative control) normalized for transfection activity (net β-galatosidase activity, β-galatosidase activity minus the negative control). Promoter activity was further standardized by dividing by a constant: the mean activity for a southern proximal promoter (for data in fig. 1*B*) or a southern promoter containing Sp1-TCC-Inr and 5′ leader (for data in fig. 1*C* and *D*). This does not alter the statistical variance and makes it simpler to compare relative activity in different cell lines.

Footprinting

A northern (M2) and a southern (G6) proximal promoter cloned into pBSKS$^-$ (Stratagene, La Jolla, Calif.) was digested with *Xho*I (5′ overhang) and *Kpn*I (3′ overhang), gel-purified, and end-labeled with ^{32}P-dCTP using the large subunit of DNA polymerase I (Klenow). Experiments typically used 0.5 footprinting units (fpu's) of purified recombinant human SP1 protein (rhSP1, expressed in HeLa cells; Promega) per reaction. Controls lacked rhSP1. The rhSP1 and binding buffer (4% glycerol, 1 mM $MgCl_2$, 0.5 mM EDTA, 0.5 mM DTT, 50 mM NaCl, 10 mM Tris-HCl [pH 7.5], 0.05 μg/μl poly dI·dC, and 1 mg BSA) were incubated on ice for 10 min before the addition of freshly labeled fragment. After the addition of labeled fragment, the reaction was incubated for 20 more min on ice, followed by the addition of 0.01 unit of RQ1-DNase (Promega) in $MgCl_2$-$CaCl_2$ (final concentrations 10 mM and 2.5 mM, respectively) and incubated for 2 min at room temperature. DNase reactions were stopped by the addition of an equal volume of 30 mM EDTA, 2% SDS, and 200 mM NaCl at 37°C. Reactions were phenol-chloroform-extracted and ethanol-precipitated. Ethanol precipitates were resuspended in 4 μl dH_2O and 6 μl of loading buffer (0.1M NaOH, formamide [1:2 v/v], 0.2% bromophenol blue, and 0.2% xylene cyanol). All reactions were heated to approximately 80°C for 2 min and then were electrophoresed in a 7 M urea 5% acrylimide gel. The gel was subsequently dried and subjected to autoradiography overnight with an intensifying screen. Varying $MgCl_2$ or poly dI·dC had no substantial effect.

Sequence Analyses

Six northern and 10 southern *F. heteroclitus* individual promoters were analyzed. The pattern of sequence variation in protein/DNA-binding regions (functional regions) versus that in nonfunctional regions was analyzed by examining the number of polymorphisms and fixed differences between *F. heteroclitus* populations. To determine the evolutionary processes affecting the *Ldh-B* proximal promoter, a modification of the McDonald-Kreitman test (McDonald and Kreitman 1991) was used. Specifically, the patterns of sequence variation among functional versus nonfunctional regions were compared instead of synonymous versus nonsynonymous sites. A similar application was used in the analysis of the *eve* enhancer (Ludwig and Kreitman 1995). This statistical test has few assumptions. It distinguishes between natural selection and neutral evolutionary processes by examining the ratio of fixed functional substitutions (nonsynonymous, as originally used in the analysis of coding regions) to fixed nonfunctional substitutions (synonymous) between taxa relative to the ratio of polymorphisms within a taxon for these functional and nonfunctional regions. Unlike in the original application, two populations, not two species, were used. A two-tailed Fisher's exact test was use to test the null hypothesis that the sequence variation is random.

Phylogenetic analysis was applied to proximal promoter sequences of 16 *F. heteroclitus* and 2 *F. grandis* individuals. The phylogenetic relationship among these taxa is well-established (Wiley 1986; Cashner, Rogers, and Grady 1992; Bernardi and Powers 1995), and the use of these three taxa is not for systematic purposes, but to determine which characters are derived. Phylogenetic analysis of proximal promoter sequences used both maximum-parsimony (PAUP 3.1) and maximum-likelihood methods (Felsenstein 1981, 1996). Three tests were used to determine whether there is significant structure in the phylogenetic trees: the g-test, the PTP test, and the tPTP test (Faith 1991; Faith and Cranston 1991; Huelsenbeck and Hillis 1992). The g-test measures the skewness of tree lengths: tree lengths with a significant left skewness indicate that relatively few solutions exist. PTP and tPTP are randomization tests for cladistic structure, comparing the observed tree length to the frequency of trees that have equal or shorter lengths from a randomization of the original data. The original data were randomized (100 times) among taxa, and the shortest tree for each randomization was determined. The relative frequency at which a random tree is as short or shorter is the probability. tPTP tests a specific tree-branching pattern or topology.

Results

Proximal Sequences

The proximal promoter is approximately 300 bp long, with approximately 240 bp 5′ of the start of tran-

Table 2
Variable Sites from Schulte, Gomezchiarri, and Powers (1997)

Taxon	Nucleotide Position
	11111111111122222222222
	11123333444455555667888899000267789999111112677 88
	13079002712340456809056780867897121045623456160625
FHnb1	CGGACTTCCAG-CCTGT-AT-C-GAGCAT-GCC3CG---TC6--ACCA
FHme5 ...	CGAACTTCTAG-CCTGT-AT-G-CAGCAG-GCC3CG---TC6-7--ACAA
FHme4 ...	CGAACTTCTAG-CCTGT-ATTG-CAGCAG-GCC3CG---TC6-7--ACCA
FHme3 ...	CGAACCTCTAG-CCTGT-AT-G-CAGCAG-GCC3CG---TC6-7--ACCA
FHme2 ...	CGAACTTCTAG-CCTGT-AT-C-GAGCAG-GCC3CG---TC6-7--ACCA
FHme1 ...	CGAACTTCTAG-CCTGT-AT-C-GAGCAG-GCC3CG---TC6-7--ACCA
FHwh2 ...	CGAACTTCTAT-CAAGG2GT---CAG-G--GT--CGC-5CC-----ACCA
FHwh1 ...	CGAACTTCTAG-CCTGT-AT-G-CAG-AG-GCC3CG---TC6-7--ACCA
FHga6	GGAACTACTAT1TCTAT-GT-G-CAGCAG-AC---C--5CC----AA-CG
FHga5	CGATCTTTTGT1CCTAT-GT-G-CAACAG-GT-3CGC-5CC--7--ACCG
FHga4	CGAACTACTAT-CATAT-GT-GCTAG-AGTGT-3CC--5CC--7--CGGA
FHga3	TAAACTACTAT1CCTAT-GT-G-CAGCAG-GC-3CGC45CC--7A-ACCA
FHga2	AGAACTTCTAT-CATAT-GT-G-CAGCAG-GC--TGC-5CC--7--ACCA
FHfA1	CGAACTACTAT1CCTAT-GT-G-CAG-AG-AC--CGC-5CC66---ACCG
FHfl5	CGAACTACTAT1CCTAT-GT-G-CAGCAG-GC---C--5CC667--AC-A
FHfl4	CGAACTACTAT1CCTAT-GT-G-CAGCAG-GC---C--5CC667--AC-A
FHfl3	CGAACTACTAT1CCTAT-GT-G-CAG-AG-AC--CGC-5CC66---ACCG
FHfl2	CGAACTACTAT1CCTAT-G--G-CAGCAG-AC--CGC-5CC66---ACCG
FHfl1	CGAACTACTAT1CCTAT-GT-G-CAG-AG-AC--CGC-5CC66---ACCG
FGtx1	CGAAATTCTAT-CTTAT-GT-GCCGG-AG-GT-3CGC-5CA-----A-CG
—	0000*00000F0000F00F00000*0000000F00000FF*000000000

NOTE.—Sequences were retrieved from GenBank and include only homologous proximal promoters. Total variable sites = 50. Within *Fundulus heteroclitus,* 47 sites are variable; 19 of these are gaps, and 6 are fixed differences between populations. The symbols below the sequences are as in table 1. Gap code: 1 = CAT; 2 = TTAGACCAT; 3 = CCTCCCTC; 4 = GGCCCC; 5 = GGCCA; 6 = TCC; 7 = TGC. Gaps, 1, 2+3, 4, 5, and 6 are homologous to Gaps 1, 3, 5, 6, and 7 (table 1), respectively.

scription. When all gaps are reduced to a single character (table 1), the overall lengths of the aligned sequences are 311 bp. The *Ldh-B* proximal promoter has a considerable amount of DNA sequence variation in (fig. 1*A*). Within the *F. heteroclitus* proximal promoters, 31 sites are variable, and among *F. heteroclitus* and *F. grandis,* 15 additional sites are variable (table 1). Thus, in *F. heteroclitus,* 9.9% (31/311, or, if excluding gaps, 19/311 [6.1%]) of the sites are polymorphic. Among *F. heteroclitus* and excluding gaps (number of sites = 297), the nucleotide diversity is 0.018 (Jukes and Cantor 1969), or θ per site is 0.019 ($\theta = 4Nu$, where N is the effective population size and u is the neutral mutation rate). This represents θ per sequence of 5.73 for both populations and ranges from 0.876 (for the northern population) to 4.242 (for the southern population). This amount of variation is approximately 10-fold higher (9.9% vs. 0.9% polymorphic sites) than the amount of variation for the *F. heteroclitus Ldh-B* proximal promoter reported by Schulte, Gomezchiarri, and Powers (1997). Although these results appear incongruent, nearly all of the variable sites described here, and several more, were found in this other study (tables 1 and 2). Differences in alignment, spacing, and how sequences surrounding gaps (but not gaps themselves) were handled may account for the difference.

Eight fixed differences occur between *F. heteroclitus* populations (table 1). Excluding gaps, the average nucleotide differences between populations (Dxy) is 7.93, or a Dxy per site of 0.0267 (Jukes and Cantor 1969). This is the minimal number of fixed differences: the original CLUSTAL alignment yielded 14 fixed differences. The number of fixed differences was reduced to 8 by the introduction of a single-base-pair gap into all southern sequences at position 188 (relative to G1) and a 5-bp gap in all northern sequences at position 192. Thus, 8 of the original 14 fixed nucleotide differences were reduced to two fixed gaps. Six of the eight fixed differences were also found in an independent analysis of the *F. heteroclitus Ldh-B* promoter (tables 1 and 2; Schulte, Gomezchiarri, and Powers 1997). The two other fixed differences provided in this study were polymorphic in the southern population as described by Schulte, Gomezchiarri, and Powers (1997). One of these sites is the consensus Sp1 sequence (GGGCGG/CCGCCC [Dynan and Tjian 1983]), found only in the southern population. In both studies, all northern individuals have G→C transversions (GGGCCG). In this study, the consensus Sp1 site is found in all southern individuals: 10 southern individuals had only a single C forming the consensus Sp1 site. All DNAs were originally sequenced four times per base pair, and northern and southern individuals' sequences were electrophoresed next to each other. These procedures demonstrated that the spacing and nucleotides for this Sp1 region were consistently different between populations. In summary, a large percentage of sites in the *Ldh-B* proximal promoter are variable based on the sequences from the two populations in this study and in five populations provided by Schulte, Gomezchiarri, and Powers (1997). Approximately 20%–26% (8/31 reported here or 6/31) of these polymorphic sites represent fixed differences be-

tween northern (>44°N) and southern (<32°N) populations.

Promoter Activity and Sequence Variation

Does natural sequence variation affect promoter activity? To answer this question, individual *Ldh-B* proximal promoters (as in fig. 1*A* and table 1) were isolated from northern and southern fish and ligated into luciferase reporter plasmids, and their promoter strengths were determined by transient transfection into a liver-derived fish cell line (rainbow trout hepatoma, RTH-149, fig. 1*B*). The promoters from northern populations had significantly greater transcriptional activity than did the southern promoters (fig. 1*B*). However, despite the extensive DNA sequence variation, especially within the southern population, individual promoters within a population are not a significant source of variation (nested ANOVA, fig. 1*B*). Thus, only the fixed sequence differences between populations effect a difference in transcription when assayed in this cell line.

Functional Regions in the Ldh-B Proximal Promoters

Three regions (TCC repeats, Sp1, and 6fp) in the *Ldh-B* proximal promoter bind nuclear proteins in vivo (Segal et al. 1996). These three variable regions contribute significantly to promoter activity when assayed in fish cell lines that express *Ldh-B* (salmon cardiac and rainbow trout hepatoma lines [fig. 1*C* and *D*] and salmon embryonic lines [Segal et al. 1996]).

In salmon cardiac cells, promoters that lack 6fp, Sp1, or TCC-Inr have significantly reduced transcription rates (two-way ANOVA—promoters: $P < 0.005$; population: $P > 0.1$). In the southern proximal promoter, 6fp and TCC repeats had a significant effect on promoter activity (Fisher's LSD, $P < 0.02$). In the northern proximal promoter, Sp1 and TCC repeats had a significant effect on promoter activity (Fisher's LSD, $P < 0.02$). These salmon cardiac data indicate that all three protein/DNA-binding regions affect transcription. The effect of these deletions is different between northern and southern promoters, suggesting that sequence variation can affect transcriptional processes.

Rainbow trout hepatoma cells were used to further investigate the effect of 6fp and Sp1 sequences and to more narrowly define these protein/DNA-binding regions (fig. 1*D*). These deletions support the locations of the 3′ end of 6fp and the 5′ end of Sp1-protein-binding regions: within a population, there is no significant difference in promoter activity among deletions between 6fp and Sp1 sequences (ANOVA, $P > 0.25$ in each population). These analyses (fig. 1*D*) also demonstrate that sequence differences between populations affect the function of 6fp and Sp1 (two-way ANOVA, $P < 0.001$ for populations and deletion constructs). Analysis of deletions within each population indicated that 6fp sequences have a significant effect on promoter activity (Fisher's LSD, $P < 0.01$) in both populations, and deletion of Sp1 sequences has a significant effect on northern populations (Fisher's LSD, $P < 0.01$). However, in rainbow trout hepatoma cells, the southern Sp1 sequences did not have a significant effect (Fisher's LSD, $P > 0.05$; fig. 1*D*). The difference in the effect of deleting the Sp1 sequence may reflect the effect of the TCC-TFIID-binding-site polymorphism (southern TCC repeat vs. northern TCC-TTC repeats differentially affect transcription depending on Sp1 sequences; unpublished data).

Notice that these deletions replace the natural promoter sequences with identical plasmid sequences. Increasing the proximity of this plasmid sequence to the start of transcription had no effect on transcription (fig. 1*D*). In both salmon cardiac and rainbow trout cell lines (fig. 1*C* and *D*), replacement of protein/DNA-binding sites reduces transcription by an amount that depends on promoter sequence. The plasmid sequence does not have any promoter activity (see last deletion in fig. 1*C*). However, it is possible that deletions and plasmid sequence could interact in unexpected ways. This was not seen with deletions between 6fp and Sp1 sequences (fig. 1*D*). Additionally, it does not explain why deletion of Sp1 has a larger effect on the northern promoters than on southern promoters in both cell lines, or why there are differences between populations in how 6fp affects transcription.

In TATA-less promoters, Sp1 binding has a pivotal role in initiating transcription (Zenzie-Gregory et al. 1993; Zhang et al. 1994; Colgan and Manley 1995; Block, Shou, and Poncz 1996; Chen et al. 1997). In the *F. heteroclitus Ldh-B* promoter, the Sp1-like binding sites (fig. 2*A*) consist of several heterogeneous sequences that match at least five of the six nucleotides for the canonical Sp1 binding site (GGGCGG [Dynan and Tjian 1983]). All Sp1 sequences are on the antisense strand and are capable of activating transcription in this orientation (Smale et al. 1990; Zenzie-Gregory, O'Shea-Greenfield, and Smale 1992). The Sp1 sequences in the *Ldh-B* promoter can be divided into two regions: a GC-rich region and a region represented by TCCC repeats. Within the GC-rich region is the only consensus Sp1 sequence (a 6-of-6-nucleotide match), and this consensus Sp1 site is found only in southern promoters.

In vitro footprinting (fig. 2*B*) was used to determine where SP1 is capable of binding within the *Ldh-B* proximal promoter. Specifically, does SP1 bind to both the TCCC and GC regions in the Sp1-like sequences? rhSP1 appears to bind most avidly to the TCCC repeat, as evidenced by this region having the greatest protection. Surprisingly, the least protection is provided to the perfect Sp1 sequence in the southern promoter. This observation is consistent with the differences in binding affinity between rhSP1 and the *Ldh-B* Sp1 sequences: the northern sequence has a lower Kd (greater affinity) than the southern Sp1 sequences (unpublished data). Additionally, when assayed in rainbow trout hepatoma cells, deletion of the northern Sp1 sequence has a significant effect on transcription, whereas the deletion of the southern Sp1 sequence is not significant (Fisher's LSD, $P > 0.05$; fig. 1*D*). These data indicate that the complete Sp1-like sequences are capable of binding SP1 protein, and the sequence variation can affect protein/DNA interaction.

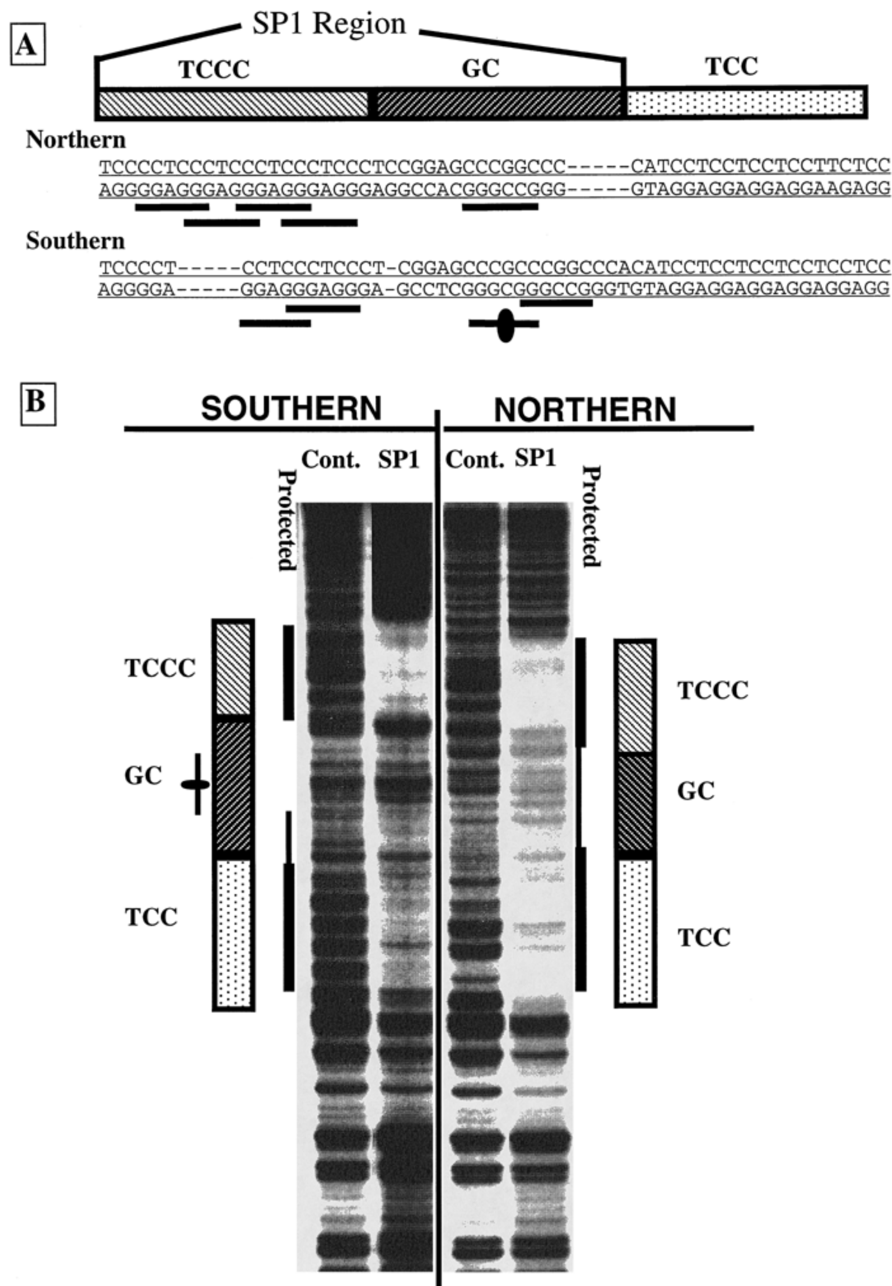

FIG. 2.—*Ldh-B* Sp1 sequences and protein interactions. *A,* Sp1 and the putative TFIID-binding site (TCC) of the southern and northern *Ldh-B* proximal promoters. In *Ldh-B,* Sp1 sequences are on the complimentary strand. Boxes above sequences highlight different regions in the proximal promoter: Sp1-like sequences that consist of either TCCC repeats or are GC-rich and the putative TFIID-binding sequence that consists of TCC repeats. Bars below sequences indicate potential Sp1-binding sites at which at least five of six nucleotides match the consensus SP1 protein recognition sequence. The bar with a closed circle is a consensus Sp1 site (6 of 6 nucleotides). *B,* In vitro footprinting reactions for a representative southern proximal promoter and a northern proximal promoter (approximately 325 bp each). Boxes represent regions of the proximal promoter given in *A.* The bar nearest the autoradiogram represents protected regions. Notice that the consensus Sp1 site (single bar with closed circle) is not well protected. Lanes labeled ''CONT'' are proximal promoters without rhSP1 protein. Lanes labeled ''SP1'' contain the rhSP1 protein.

Relevance of Molecular Assays

The use of human SP1 protein and heterologous fish cell lines may seem at first to generate more questions than answers: does *Fundulus* SP1 ''behave'' like human Sp1? Do the SP1 and other *trans*-acting factors from the salmon cardiac and rainbow trout cell lines demonstrate the same properties as the equivalent proteins from *Fundulus*? Does the proximal promoter activity measured in either cell line mechanistically reflect the activity that occurs in vivo? These are important questions, but they are not addressed above. In this study, the purpose of the molecular assays was to identify specific regions that bind a transcription factor or contribute to promoter activity. The three sequences within the *F. heteroclitus Ldh-B* proximal promoter that affect transcription were originally defined by in vivo

footprinting (Segal et al. 1996). All of the molecular studies indicate that the three regions are functionally important: they affect transcriptional processes.

In the case of human SP1 in vitro binding, the specific question addressed is if and where an SP1-like protein binds. Specifically, does SP1 bind to both the GC and TCCC Sp1 sequences in *F. heteroclitus*? The only assumption is that the *Fundulus* SP1-like protein has sequence specificity similar to that of the human SP1 protein used here. This assumption is justifiable. Experimental analyses indicate that *F. heteroclitus* SP1-like protein binds the same sequences as rhSP1 and has the same specificity (unpublished data). Additionally, the three different SP1-like proteins (SP1, SP3, and SP4) have the same sequence specificity (Kingsley and Winoto 1992; Hagen et al. 1994). There may be differences in binding affinities among the different SP1 proteins, but the sequence specificities should be similar.

For the cell culture studies, all of the transfections indicate that the three regions affect transcriptional processes, albeit not quantitatively the same among cell lines. The quantitative difference in promoter activity could reflect differences in the quantity or posttranslational modification of transcription factors (e.g., glycosylation, phosphorylation) (Jackson and Tjian 1988; Jackson, Gottlieb, and Hartley 1993; Jane et al. 1993; Leggett et al. 1995; Apt et al. 1996; Daniel et al. 1996). It is tempting to suggest that the results from hepatoma are more relevant because the greater northern *Ldh-B* transcription rate was measured with hepatocytes in vivo. Alternatively, evolutionary analyses indicating that the sequence variation effects a phenotypic change that is subject to natural selection may be more persuasive.

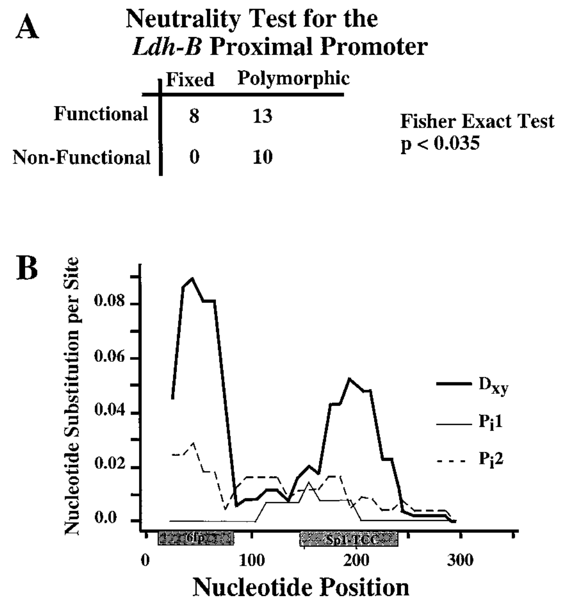

FIG. 3.—Evolutionary analyses of *Ldh-B* proximal promoter sequence variation. *A,* Neutrality test for the pattern of sequence variation (McDonald and Kreitman 1991). Probability that this pattern of sequence variation is due to random or neutral evolutionary processes is rejected by a two-sided Fisher's exact test. Functional sequences were defined before analysis as those interspersed regions which bind proteins in vivo (6fp, Sp1, and TCC repeats; fig. 1). *B,* Sliding-window comparison excluding insertions and deletions. Average nucleotide substitutions per site within the northern *Fundulus heteroclitus* population ($P_1$1), within the southern *F. heteroclitus* population ($P_1$2), and between *F. heteroclitus* populations (Dxy). Plots for the sliding window use a 50-bp-wide window with a 10-bp step size. Gray bars on the horizontal axis indicate the locations of windows whose midpoints include protein/DNA-binding sites. Thus, gray bars are 25 bp longer on each side than the actual binding site. Plots and computations used DnaSP software (Rozas and Rozas 1997).

Evolutionary Analyses

Evolution by natural selection is responsible for the sequence differences between populations in the *Ldh-B* proximal promoter (fig. 3*A*). This is deduced by examining the pattern of promoter sequence variation between protein/DNA-binding sequences that affect transcription (functionally important) and those interspersed sequences with no known function (nonfunctional) (McDonald and Kreitman 1991; Ludwig and Kreitman 1995). For the *Ldh-B* proximal promoter, functional sequences were defined prior to analysis as those regions which bind proteins in vivo (6fp, Sp1, and TCC repeats; fig. 1), affect transcription (fig. 1), and bind a transcription factor in vitro (fig. 2). The functional and nonfunctional regions have approximately the same number of polymorphisms. However, in the *F. heteroclitus Ldh-B* proximal promoter, all fixed differences between populations are in the functional region (fig. 3*A*). This is most readily visualized by plotting the average number of nucleotide differences per site between populations (fig. 3*B*). This pattern of variation is significantly different from the pattern expected by neutral evolutionary processes (two-tailed Fisher's exact test, $P < 0.035$) and thus is best explained by the *Ldh-B* proximal promoter evolving by natural selection. A similar result is achieved by applying this analysis to the proximal promoter sequences (Schulte, Gomezchiarri, and Powers 1997) from two northern and two southern populations (Fisher's exact test, $P < 0.04$).

All gaps, regardless of length, were scored as single characters. Spacing within the proximal promoter can affect the interactions between transcription factors, and small spacing differences can affect transcriptional processes (O'Shea-Greenfield and Smale 1992). Thus, reduction of each gap to a single character reduces the importance of gap length but retains the presence or absence of gaps as a single and potentially evolutionarily significant trait. However, even if all gaps are excluded from the analysis, the pattern of sequence variation is still significantly different from the pattern expected by random chance ($P < 0.05$). Cloning and sequence artifacts are not the source of polymorphisms, because all unique substitutions (singletons) were confirmed by reamplifying proximal promoters from genomic DNA and directly sequencing the PCR products. However, two of the fixed differences found in this study were not found by Schulte, Gomezchiarri, and Powers (1997). The pattern of sequence variation is still significantly different from random expectation ($P < 0.05$) if only the fixed differences found in both studies are used. Thus, the pattern of variation among functional and nonfunctional regions of the proximal promoter most likely is due to natural selection.

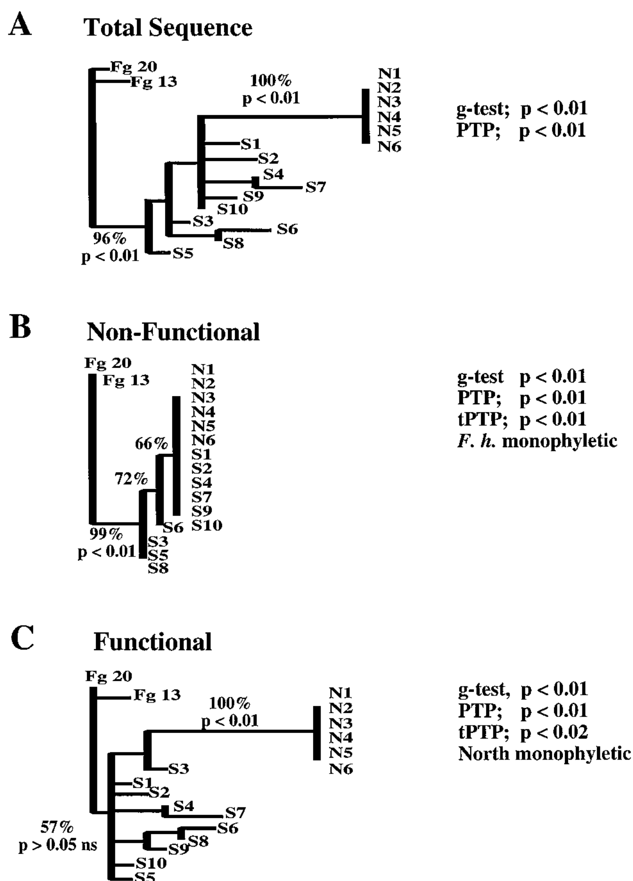

FIG. 4.—Phylogenetic analyses of the *Ldh-B* proximal promoter. *A,* Phylogenetic tree using all of the *Ldh-B* proximal promoter. *B,* Phylogenetic tree using only nonfunctional regions of the *Ldh-B* proximal promoter. *C,* Phylogenetic tree using only functional *Ldh-B* proximal promoter sequences. N and S refer to northern and southern populations, respectively. Horizontal branch lengths are proportional to the number of character changes (apomorphy; PAUP). Vertical distance is arbitrary. Phylogenetic trees represent consistent topology between DNAml (Felsenstein 1981, 1996) and parsimony analysis (Swofford 1989). Minor topological differences between methods exist in branches with less than 50% bootstrap value or with insignificant maximum-likelihood values. Percentage values (%) on phylogenetic trees represent bootstrap values ($N = 500$). All bootstrap values greater than 50% are listed. Probabilities are maximum-likelihood values. The g-test and PTP probabilities provide measurements of the significance of topological structure (Faith 1991; Faith and Cranston 1991; Huelsenbeck and Hillis 1992). The tPTP probability tests for specific monophyly, which is listed under each value (Faith 1991).

An indication of both the propensity to change and the direction of change can be determined by phylogenetic analysis (fig. 4). The phylogenetic relationship between *F. heteroclitus* and *F. grandis* is established (Wiley 1986; Cashner, Rogers, and Grady 1992; Bernardi and Powers 1995), allowing comparison among the three taxa: two populations of *F. heteroclitus,* with its sister taxon, *F. grandis,* as an outgroup. That is, the phylogenetic analyses provided here are not used to examine the well-established relationship between these sister taxa, but to examine how the functional and nonfunctional sequences affect the phylogenetic topology and to determine a derived state.

Analyses of the evolutionary relationships among *Fundulus* species indicate that the northern *Ldh-B* proximal promoter is evolving due to directional selection (fig. 4*A–C*). This conclusion is based on phylogenetic analyses using parsimony (Swofford 1989) and maximum-likelihood methods (Felsenstein 1981, 1996) on three data sets: (1) the complete proximal promoter, (2) only the interspersed 100 bp that were defined a priori as functionally important (fig. 1*A*), and (3) only nonfunctional sequences (approximately 211 bp). When using all of the proximal promoter sequence (fig. 4*A*), the topology is similar those from to previous molecular analyses of *F. heteroclitus* populations (Powers et al. 1991; Bernardi, Sordino, and Powers 1993; Bernardi and Powers 1995) and agrees with the previously established phylogeny of the two species (Wiley 1986; Cashner, Rogers, and Grady 1992; Bernardi and Powers 1995). When only the nonfunctional sequence is used (fig. 4*B*), *F. heteroclitus* forms a monophyletic group (tPTP, $P < 0.01$; Faith 1991) without any significant differences between *heteroclitus* populations: most southern sequences form a polytomy with all the northern sequences. With the nonfunctional sequences, the northern population is not a monophyletic group (tPTP, $P > 0.1$; Faith 1991). This pattern of variation for the nonfunctional sequences is similar to the expectation of neutral evolution between species with populations that are not well isolated.

A very different evolutionary relationship results from using only the functional 100 bp (fig. 4*C*). When using the functional sequences, the northern population is significantly different from both the southern populations and its sister taxon, *F. grandis:* individuals from the northern population can be excluded from all other taxa (tPTP, $P < 0.01$). However, few nucleotides distinguish the southern population from its sister taxon *F. grandis: F. grandis* cannot be excluded from all southern *F. heteroclitus* individuals (tPTP, $P > 0.2$). Similarly, the branch between the southern node and the northern node is the only branch with a significant maximum-likelihood length and whose confidence interval does not include zero. Thus, the southern population has undergone few changes among the protein/DNA-binding sites (100 bp) that distinguish it from its sister species (*F. grandis*). Yet, the same 100 bp in the northern population represents a derived condition, with many unique nucleotides not shared by its southern counterpart or its sister species.

Discussion

Proximal Promoter Variation

The sequence variation in the *Ldh-B* proximal promoter is relatively large for a region of DNA that must bind transcription factors to initiate transcription. Approximately 10% of sites in the *Ldh-B* proximal promoter are variable based on the sequences from the two populations in this study. Nearly all of the variable sites described here, and several more, were found in a separate study of this same region (tables 1 and 2; Schulte, Gomezchiarri, and Powers 1997). Using only polymor-

phic sites found in both studies (i.e., excluding all unique polymorphic sites), 4.8% (15/311) of the sites are polymorphic. This is a lot of variation relative to the percentages of polymorphic sites among different populations of *Drosophila melanogaster* for the first *Adh* intron (1.7%; Kreitman 1983), *eve* enhancer (0.87%; Ludwig and Kreitman 1995), and G6pd (1.0%; Eanes, Kirchner, and Yoon 1993) and in *F. heteroclitus* for the fourfold-degenerate sites of the *Ldh-B* coding regions (1.5%; Bernardi, Sordino, and Powers 1993).

Approximately 20%–26% (8/31 reported here or 6/31 [Schulte, Gomezchiarri, and Powers 1997]) of the proximal promoter polymorphic sites are fixed differences between northern (>44°N) and southern (<32°N) populations. The percentages of fixed nucleotide differences (8/311 = 2.6%) and of fixed differences per polymorphic site (8/31 = 26%) are similar to those found among species of *Drosophila* in the *Adh* coding region (3.6% and 55%, respectively; Kreitman 1983) and to the percentage of fixed differences in the *eve* enhancer (1.8%; Ludwig and Kreitman 1995). That is, there is as much variation between populations of *F. heteroclitus* as there is among species of *Drosophila.* These fixed differences between populations are not dispersed evenly; there are no fixed differences in regions that do not bind transcription factors (i.e., nonfunctional regions; fig. 3*B* and *C*). This pattern of variation for the nonfunctional sequences is similar to the expectation of neutral evolution among populations that are not well isolated. Thus, the large difference within a species does not necessarily indicate isolation of northern *F. heteroclitus* populations or insipid speciation. Instead, we argue that this large difference is in response to directional selection acting on regions that affect transcription.

Sequence variation in proximal promoters in general, and in the *F. heteroclitus Ldh-B* proximal promoter specifically, may not have a significant effect on transcription (Schulte, Gomezchiarri, and Powers 1997). The data presented here indicates the *Ldh-B* sequence variation could affect transcriptional processes: proximal promoters from northern populations effect a greater transcription rate than do southern proximal promoters when assayed in a teleost liver-derived cell line, and sequence variation in individual elements affects transcriptional processes (fig. 1*B–D*). Not all of the proximal promoter sequence variation affects transcriptional processes. Instead, only the variation in three protein/DNA-binding regions affects molecular assays of promoter strength (fig. 1). Of this variation, only the fixed differences between populations appear to be important when proximal promoter strength is assayed by transient transfection into a liver-derived teleost cell line (fig. 1*B*). These molecular assays are relevant in that they identify sequences that bind transcription factors (see *Results: Relevance of Molecular Assays*). However, the effect of this sequence variation should be questioned, because (1) these assays ignore other upstream enhancer or transcription sites, and (2) the magnitude of the difference between populations is dependent on the cell culture system (fig. 1*B* and *D* vs. *C*).

Proximal Promoter Evolution

How does one know if proximal promoter sequence variation is important in its natural context? Evolutionary analyses that indicate that a molecular trait is evolving by natural selection could provide some evidence. Evolution by natural selection distinguishes between variation that produces biologically important phenotypic differences (differences that affect longevity, probability of survival, or reproductive fitness) and variation that produces functional but unimportant phenotypic differences (Endler 1986).

The molecular analyses support the supposition that there are three protein/DNA-binding regions and that the interspersed regions do not affect transcription. These data were used to test a specific a priori evolutionary hypothesis: that the sequence variation in protein/DNA-binding regions would be significantly different from those predicted by neutral evolutionary processes. Evolutionary and phylogenetic analyses of the *Ldh-B* proximal promoter in *F. heteroclitus* indicate that different evolutionary forces are acting on the functional and nonfunctional regions. The pattern of sequence variation (too many fixed differences in the functional regions between populations; fig. 3), combined with experimental assays, indicates that the three protein/DNA-binding regions are evolving by natural selection. The alternative explanation, that the nonfunctional regions are under stabilizing selection that reduces differences between populations but has allowed this region to change between species, is much less justifiable. That is, one would have to argue that the sequence variation in the protein/DNA-binding regions is effectively neutral and that the intervening (nonfunctional) regions are conserved to maintain function only within a species. The experimental evidence does not support this. Additionally, the large number of fixed differences between *F. heteroclitus* and *F. grandis* in the nonfunctional regions but not in the functional regions is difficult to reconcile with stabilizing selection. Thus, the proximal promoter sequence differences that can affect transcriptional processes (figs. 1 and 2) have a nonrandom pattern of sequence variation, and this pattern of variation is most likely due to natural selection.

Importantly, the sequences in the functional regions are a derived condition in northern proximal promoters (fig. 4). This suggests that the northern proximal promoter has evolved by directional selection. We can speculate on the factors causing the directional selection on the northern *Ldh-B* proximal promoter. The most obvious factor is environmental temperature: the northern population is consistently subjected to temperatures more than 12°C colder than those for its southern counterpart. Temperature is an important selective factor affecting the expression of *Ldh-B* and other glycolytic enzymes among *Fundulus* species (Pierce and Crawford 1997*a*). Thus, the pattern of sequence variation in the *Ldh-B* proximal promoter may be explained by directional selection on the northern proximal promoter effecting a change in enzyme expression, possibly due to the colder environmental temperatures.

These findings that demonstrate that variation in the *Ldh-B* proximal promoter is affected by directional selection are indicative that this variation produces functional and important phenotypic differences. Most likely, this sequence variation affects the transcription rate, which regulates enzyme concentration (Crawford and Powers 1992). This leads to an interesting prediction: populations at intermediate latitudes that have intermediate concentrations of *Ldh-B* (Crawford, Place, and Powers 1990) should have proximal promoters with intermediate promoter strength with different combinations of the eight fixed nucleotides.

Populations and Evolutionary Change

Two evolutionary mechanisms that influenced how natural selection affected the proximal promoter sequence divergence could be (1) clinal selection along the coast or (2) selection on isolated northern populations and a breakdown of this barrier, followed by reintegration of the population to create the clinal variation in *Ldh-B* expression (Crawford, Place, and Powers 1990). The lack of derived characters in the nonfunctional regions is not consistent with the latter hypothesis (isolation/reintegration). That is, if northern populations were isolated for a considerable amount of time and then were reintegrated with more southern populations, one would expect to find some fixed differences in the nonfunctional regions due to neutral evolutionary processes. This pattern of sequence variation is not seen; all northern individuals share the same polymorphisms with most southern individuals (fig. 4*B*). However, mitochondrial and *Ldh-B* cDNA data (Bernardi, Sordino, and Powers 1993) suggest that northern populations were isolated and that this isolation contributed much to the divergence between the extreme northern and southern populations. These two mechanisms are not mutually exclusive. There could be selection along the thermal cline. Isolation of the northern population for a few generations, along with directional selection, would then greatly increase the differences between northern and southern populations (Garcia Ramos and Kirkpatrick 1997). Thus, the combination of historical isolation and selection could explain the variations in divergence among biochemical traits.

Evolutionary Importance of TATA-less Promoters

Subtle variation in enzyme activity is adaptively important (Hochachka and Somero 1984). For example, there is an adaptive increase of 22%–35% in the Michaelis-Menten constant, K_m, among barracuda (Graves and Somero 1982). In *F. heteroclitus,* the combination of kinetic variation and evolutionary and physiologically induced differences in expression alters *Ldh-B* activity by approximately threefold (Crawford and Powers 1989). Most of this adaptive change in *Ldh-B* activity is due to a change in enzyme expression (Crawford and Powers 1989). Adaptive change in enzyme expression may have a greater probability of occurring than change in enzyme structure that alters activity, because mutations in promoter regions are not constrained by codon usage or the effect of amino acid replacements on tertiary or quartenary structure of a protein. That is, amino acid substitutions that could have a positive effect on enzyme activity also may have negative effects on other attributes of an enzyme function. Promoter sequence variation affects only the linear array of nucleotides that may affect protein/DNA binding. Thus, there may be many more acceptable nucleotide substitutions that do not have negative epistatic interactions, and, individually or in combination, these substitutions could affect transcription. This hypothesis is supported by analysis of adaptive change among *Escherichia coli* isolates subjected to glucose-limited environments (Treves, Manning, and Adams 1998). All six independent replicate populations of *E. coli* that evolved acetate cross-feeding polymorphisms adapted by overexpression of the enzyme acetyl CoA synthetase. Changes in the 5′ regulatory region were responsible for all of these changes. The independent evolution of adaptive enzyme expression suggests that changes in transcription are evolutionarily more probable.

One region of DNA that could effect a subtle change in enzyme expression is the proximal promoter. Most genes have a single proximal promoter that must be used for all mRNA expression. Protein/DNA-binding sites in proximal promoters are responsible for formation of the preinitiation complex that is necessary for RNA polymerase activity. One of these binding sites, the TATA box, binds the essential TFIID protein complex and is critical for all mRNA transcription (Roeder 1991; Ernst and Smale 1995*b*). For example, a point mutation in the TATA box is thought to be responsible for some cases of β-thalasemia (Antonarakis et al. 1984). In TATA-less promoters, interaction among TFIID, SP1 proteins, and other proximal promoter transcription factors can effect subtle variation in expression (Wiley, Kraus, and Mertz 1992; Azizkhan et al. 1993; al-Asadi, Yi, and Merchant 1995; Yean and Gralla 1997). This is in contrast to upstream enhancers/promoter elements responsible for developmental or tissue-specific patterns of expression. Upstream enhancers or tissue-specific regulatory elements are important in determining when or where a gene is expressed (Guarente and Bermingham-McDonogh 1992; Patel 1994; Ernst and Smale 1995*a,* 1995*b*; Novina and Roy 1996). Typically, these regulatory elements effect a large (>10×) difference in transcription, or ''on/off,'' scenarios. Sequence variation in these temporal and spatial control elements either results in disruption of function (Vincent and Wilson 1989; Gumucio et al. 1993, 1994, 1996; McKenzie, Hu, and Brennan 1994; Patel 1994; Ross, Fong, and Cavener 1994; Lowe and Wray 1997; Singh, Barbour, and Berger 1998) or, because of redundancy in these elements, has no appreciable effect (Fang and Brennan 1992; McKenzie, Hu, and Brennan 1994; Ludwig and Kreitman 1995). This redundancy in regulatory regions may be a general phenomenon for enhancer or tissue-specific regulatory regions, as exemplified by the *Drosophila Adh* locus (Fang and Brennan 1992; McKenzie, Hu, and Brennan 1994) and the mammalian globin gene (Gumucio et al. 1993). The presence of a single proximal promoter at which sequence variation can ef-

fect small changes in transcriptional processes may more likely evolve variation that affects subtle change in enzyme expression.

Other *Ldh-B* Regulatory Elements

Recently, an independent analysis of approximately 1.1 kb of the *F. heteroclitus Ldh-B* 5′ regulatory regions (including the proximal promoter region) suggests that this 1.1-kb region is evolving by natural selection (Schulte, Gomezchiarri, and Powers 1997). This conclusion is based on a comparison of the 1.1-kb upstream region with mitochondrial cyt-b and *Ldh-B* coding regions (HKA test [Hudson, Kreitman, and Aguade 1987]) that indicated that the sequence variation significantly departs from random expectation. Although this analysis suggests that the 5′ regulatory region is affected by natural selection, it does not support our conclusion that the variation in the proximal promoter affects transcription. Instead, Schulte, Gomezchiarri, and Powers (1997) suggest that a repressor (a glucocorticoid response element) upstream of the proximal promoter is responsible for the naturally occurring difference in transcription rates and thus is the target of selection. Their conclusion is based on molecular analyses of cell transfections in human kidney cell lines and injection of plasmid constructs into *F. heteroclitus* muscle (which expresses *Ldh-A* and not *Ldh-B*). They suggest that the difference in transcription is caused by a glucocorticoid repressor element in southern populations that is active at 20°C. They also predict that, due to the nature of the glucocorticoid repressor, at low temperatures the southern promoter would be derepressed and the difference in transcription would disappear (as suggested by acclimation affects on *Ldh-B* mRNA concentrations; Segal and Crawford 1994). However, at low temperatures, the equal expression of *Ldh-B* mRNA between southern and northern populations is due to a reduction in the northern population's *Ldh-B* mRNA concentration (Segal and Crawford 1994) and not to a derepression of the southern promoter (Crawford, Pierce, and Segal 1998). Their suggestion that the variation in expression is due to temperature affects on the glucocorticoid element also fails to explain why cardiac *Ldh-B* expression is different between populations but unresponsive to thermal acclimation (Pierce and Crawford 1997*a,* 1997*b*). Although there are substantial data to suggest that the glucocorticoid repressor can be functionally important (Schulte, Gomezchiarri, and Powers 1997), it is less clear under what circumstances it affects transcription. We argue above that different molecular analyses (e.g., transcription in different cell lines) may yield different results, and although these techniques provide evidence that a specific sequence can be functionally important, they do not demonstrate whether a sequence is biologically important.

Assuming there is a biologically important enhancer upstream of the proximal promoter in which the sequence variation affects transcription, how can one reconcile this with the observed biological importance of the proximal promoter presented here? One answer is that enhancers and proximal promoters affect transcription differently (Yean and Gralla 1997). Enhancers may act to increase the probability of forming transcription complexes, while proximal promoters affect the longevity of the transcriptional complex and rate of reinitiation of transcription (Yean and Gralla 1997). Thus, the biological importance of the proximal promoter does not exclude the importance of other enhancers or regulatory elements, but it does suggest that sequence variation in the proximal promoter can modulate the effects of these upstream elements.

Conclusions

The data provided here demonstrate that *Ldh-B* proximal promoter sequence variation (1) arose due to directional selection and (2) affects promoter activity and transcription-factor-binding affinity. Thus, the proximal promoter sequence variation that can affect transcription processes (figs. 1 and 2) also effects a phenotypic change that is subject to natural selection, and therefore this variation in the proximal promoter is biologically important. These results also suggest that natural selection acting on northern populations of *F. heteroclitus* may explain some of the clinal variation in this species.

Acknowledgments

This project was aided by discussions with V. A. Pierce (CUNY, Manhattan), M. Oleksiak (UMKC), and J. Finnerty and M. Kreitman (University of Chicago). The authors wish to thank C. C. Laurie (Duke University) for critical review and helpful comments on a earlier draft of this manuscript and D. Hudson (University of California–Irvine) for insights into sequence analysis and critical review of the preliminary statistical analyses. This work was supported by NSF grant IBN-9419781 (to D.L.C.). J.A.S. was supported by NIH predoctoral training grant GM 07183.

LITERATURE CITED

AL-ASADI, R., E. C. YI, and J. L. MERCHANT. 1995. Sp1 affinity for GC-rich elements correlates with ornithine decarboxylase promoter activity. Biochem. Biophys. Res. Commun. **214**:324–330.

ANTONARAKIS, S. E., S. H. IRKIN, T. C. CHENG, A. F. SCOTT, J. P. SEXTON, S. P. TRUSKO, S. CHARACHE, and H. H. KAZAZIAN JR. 1984. Beta-thalassemia in American Blacks: novel mutations in the "TATA" box and an acceptor splice site. Proc. Natl. Acad. Sci. USA **81**:1154–1158.

APT, D., R. M. WATTS, G. SUSKE, and H. U. BERNARD. 1996. High Sp1/Sp3 ratios in epithelial cells during epithelial differentiation and cellular transformation correlate with the activation of the HPV-16 promoter. Virology **224**:281–291.

ARKHIPOVA, I. R. 1995. Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. Genetics **139**: 1359–1369.

AZIZKHAN, J. C., D. E. JENSEN, A. J. PIERCE, and M. WADE. 1993. Transcription from TATA-less promoters: dihydrofolate reductase as a model. Crit. Rev. Eukaryot. Gene Expr. **3**:229–254.

BERNARDI, G., and D. A. POWERS. 1995. Phylogenetic relationships among nine species from the genus *Fundulus* (Cy-

prinodontiformes, Fundulidae) inferred from sequences of the cytochrome B. Copeia **1995**:469–473.

BERNARDI, G., P. SORDINO, and D. A. POWERS. 1993. Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus.* Proc. Natl. Acad. Sci. USA **90**:9271–9274.

BIGGER, C. B., I. N. MELNIKOVA, and P. D. GARDNER. 1997. Sp1 and Sp3 regulate expression of neuronal nicotinic acetylcholine receptor B4 subunit gene. J. Biol. Chem. **272**: 25979–25982.

BLOCK, K. L., Y. P. SHOU, and M. PONCZ. 1996. An Ets/Sp1 interaction in the 5′-flanking region of the megakaryocyte-specific alpha-Iib gene appears to stabilize Sp1 binding and is essential for expression of this TATA-less gene. Blood **88**:2071–2080.

CASHNER, R. C., J. S. ROGERS, and J. M. GRADY. 1992. Phylogenetic studies of the genus *Fundulus.* Pp. 421–437 *in* R. L. MAYDEN, ed. Systematic, historical ecology and North American freshwater fishes. Stanford University Press, Stanford, Calif.

CHEN, S., P. C. SUPAKAR, R. L. VELLANOWETH, C. S. SONG, B. CHATTERJEE, and A. K. ROY. 1997. Functional role of a conformationally flexible homopurine/homopyrimidine domain of the androgen receptor gene promoter interacting with Sp1 and a pyrimidine single strand DNA-binding protein. Mol. Endocrinol. **11**:3–15.

COLGAN, J., and J. L. MANLEY. 1995. Cooperation between core promoter elements influences transcriptional activity in vivo. Proc. Natl. Acad. Sci. USA **92**:1955–1959.

CRAWFORD, D. L., V. A. PIERCE, and J. A. SEGAL. 1998. Evolutionary physiology of closely related taxa: analyses of enzyme expression. Am. Zool. (in press).

CRAWFORD, D. L., A. R. PLACE, and D. A. POWERS. 1990. Clinal variation in the specific activity of lactate dehydrogenase-B. J. Exp. Zool. **255**:110–113.

CRAWFORD, D. L., and D. A. POWERS. 1989. Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish *Fundulus heteroclitus.* Proc. Natl. Acad. Sci. USA **86**:9365–9369.

———. 1990. Molecular adaptation to the thermal environment: genetic and physiological mechanisms. Pp. 213–222 *in* M. T. CLEGG and S. J. O'BRIEN, eds. Molecular evolution. Vol. 122. Wiley-Liss, New York.

———. 1992. Evolutionary adaptation to different thermal environments via transcriptional regulation. Mol. Biol. Evol. **9**:806–813.

DANIEL, S., S. Y. ZHANG, A. A. DEPAOLIROACH, and K. H. KIM. 1996. Dephosphorylation of Sp1 by protein phosphatase 1 is involved in the glucose-mediated activation of the acetyl-Coa carboxylase gene. J. Biol. Chem. **271**:14692–14697.

DYNAN, W. S., and R. TJIAN. 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–87.

EANES, W. F., M. KIRCHNER, and J. YOON. 1993. Evidence for adaptive evolution of the G6pd gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc. Natl. Acad. Sci. USA **90**:7475–7479.

EMAMI, K. H., W. W. NAVARRE, and S. T. SMALE. 1995. Core promoter specificities of the Sp1 and VP16 transcriptional activation domains. Mol. Cell. Biol. **15**:5906–5916.

ENDLER, J. A. 1986. Natural selection in the wild. Princeton University Press, Princeton, N.J.

ERNST, P., and S. T. SMALE. 1995*a.* Combinatorial regulation of transcription II: the immunoglobulin mu heavy chain gene. Immunity **2**:427–438.

———. 1995*b.* Combinatorial regulation of transcription. I: general aspects of transcriptional control [Review]. Immunity **2**:311–319.

FAITH, D. P. 1991. Cladistic permutation test for monophyly and nonmonophyly. Syst. Zool. **40**:366–375.

FAITH, D. P., and P. S. CRANSTON. 1991. Could a cladogram this short have arisen by chance alone? On the permutation tests for cladistic structure. Cladistics **7**:1–28.

FANG, X. M., and M. D. BRENNAN. 1992. Multiple cis-acting sequences contribute to evolved regulatory variation for Drosophila Adh genes. Genetics **131**:333–343.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1996. PHYLIP (phylogeny inference package). Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

GARCIA RAMOS, G., and M. KIRKPATRICK. 1997. Genetic models of adaptation and gene flow in peripheral populations. Evolution **51**:21–28.

GRAVES, J. E., and G. N. SOMERO. 1982. Electrophoretic and functional enzyme evolution in four species of eastern Pacific Barracudas from different thermal environments. Evolution **36**:97–106.

GUARENTE, L., and O. BERMINGHAM-MCDONOGH. 1992. Conservation and evolution of transcriptional mechanisms in eukaryotes [Review]. Trends Genet. **8**:27–32.

GUMUCIO, D. L., D. A. SHELTON, W. J. BAILEY, J. L. SLIGHTOM, and M. GOODMAN. 1993. Phylogenetic footprinting reveals unexpected complexity in trans factor binding upstream from the epsilon-globin gene. Proc. Natl. Acad. Sci. USA **90**:6018–6022.

GUMUCIO, D. L., D. A. SHELTON, K. BLANCHARD-MCQUATE, T. GRAY, S. TARLE, H. HEILSTEDT-WILLIAMSON, J. L. SLIGHTOM, F. COLLINS, and M. GOODMAN. 1994. Differential phylogenetic footprinting as a means to identify base changes responsible for recruitment of the anthropoid gamma gene to a fetal expression pattern. J. Biol. Chem. **269**: 15371–15380.

GUMUCIO, D. L., D. A. SHELTON, W. ZHU, D. MILLINOFF, T. GRAY, J. H. BOCK, J. L. SLIGHTOM, and M. GOODMAN. 1996. Evolutionary strategies for the elucidation of cis and trans factors that regulate the developmental switching programs of the beta-like globin genes. Mol. Phylogenet. Evol. **5**:18–32.

HAGEN, G., S. MULLER, M. BEATO, and G. SUSKE. 1994. Sp1-mediated transcriptional activation is repressed by Sp3. EMBO J. **13**:3843–3851.

HIGGINS, D. G., A. J. BLEASBY, and R. FUCHS. 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8**:189–191.

HOCHACHKA, P. W., and G. N. SOMERO. 1984. Biochemical adaptation. Princeton University Press, Princeton, N.J.

HUDSON, R. R., M. KREITMAN, and M. AGUADE. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **120**:153–159.

HUELSENBECK, J. P., and D. M. HILLIS. 1992. Signal, noise, reliability in molecular phylogenetic analyses. J. Hered. **83**: 189–195.

JACKSON, S., T. GOTTLIEB, and K. HARTLEY. 1993. Phosphorylation of transcription factor Sp1 by the DNA-dependent protein kinase. Adv. Second Messenger Phosphoprotein Res. **28**:279–286.

JACKSON, S. P., and R. TJIAN. 1988. O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. Cell **55**:125–133.

JANE, S. M., D. L. GUMUCIO, P. A. NEY, J. M. CUNNINGHAM, and A. W. NIENHUIS. 1993. Methylation-enhanced binding of Sp1 to the stage selector element of the human gamma-globin gene promoter may regulate development specificity of expression. Mol. Cell. Biol. **13**:3272–3281.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KINGSLEY, C., and A. WINOTO. 1992. Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. Mol. Cell. Biol. **12**:4251–4261.

KREITMAN, M. 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster.* Nature **304**:412–417.

LEGGETT, R. W., S. A. ARMSTRONG, D. BARRY, and C. R. MUELLER. 1995. Sp1 is phosphorylated and its DNA binding activity down-regulated upon terminal differentiation of the liver. J. Biol. Chem. **270**:25879–25884.

LOWE, C. J., and G. A. WRAY. 1997. Radical alterations in the roles of homeobox genes during echinoderm evolution. Nature **389**:718–721.

LUDWIG, M. Z., and M. KREITMAN. 1995. Evolutionary dynamics of the enhancer region of even-skipped in *Drosophila.* Mol. Biol. Evol. **12**:1002–1011.

MCDONALD, J. H., and M. KREITMAN. 1991. Adaptive protein evolution at the Adh locus in *Drosophila.* Nature **351**:652–654.

MCKENZIE, R. W., J. HU, and M. D. BRENNAN. 1994. Redundant cis-acting elements control expression of the Drosophila affinidisjuncta Adh gene in the larval fat body. Nucleic Acids Res. **22**:1257–1264.

MARTINEZ, E., C. M. CHIANG, H. GE, and R. G. ROEDER. 1994. TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. EMBO J. **13**:3115–3126.

NOVINA, C. D., and A. L. ROY. 1996. Core promoters and transcriptional control. Trends Genet. **12**:351–355.

O'SHEA-GREENFIELD, A., and S. T. SMALE. 1992. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. J. Biol. Chem. **267**:6450.

PATEL, N. H. 1994. The evolution of arthropod segmentation: insights from comparisons of gene expression patterns. Dev. Suppl. 201–207.

PIERCE, V. A., and D. L. CRAWFORD. 1997*a.* Phylogenetic analysis of glycolytic enzyme expression. Science **275**:256–259.

———. 1997*b.* Phylogenetic analysis of thermal acclimation of the glycolytic enzymes in the genus *Fundulus.* Physiol. Zool. **70**:597–609.

POWERS, D. A., T. LAUERMAN, D. CRAWFORD, and L. DIMICHELE. 1991. Genetic mechanisms for adapting to a changing environment. Annu. Rev. Genet. **25**:629–659.

ROEDER, R. G. 1991. The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. Trends Biochem. Sci. **16**:402–408.

ROSS, J. L., P. P. FONG, and D. R. CAVENER. 1994. Correlated evolution of the cis-acting regulatory elements and developmental expression of the Drosophila Gld gene in seven species from the subgroup melanogaster. Dev. Genet. **15**:38–50.

ROZAS, J., and R. ROZAS. 1997. DnaSP version 2.0: a novel software package for extensive molecular population genetic analysis. Comput. Appl. Biosci. **13**:307–311.

SCHULTE, P. M., M. GOMEZCHIARRI, and D. A. POWERS. 1997. Structural and functional differences in the promoter and 5′ flanking region of Ldh-B within and between populations of the teleost Fundulus heteroclitus. Genetics **145**:759–769.

SEGAL, J. A., and D. L. CRAWFORD. 1994. LDH-B enzyme expression: the mechanisms of altered gene expression in acclimation and evolutionary adaptation. Am. J. Physiol. **267**:R1150–R1153.

SEGAL, J. A., P. M. SCHULTE, D. A. POWERS, and D. L. CRAWFORD. 1996. Descriptive and functional characterization of variation in the *Fundulus heteroclitus Ldh-B* proximal promoter. J. Exp. Zool. **275**:355–364.

SINGH, N., K. W. BARBOUR, and F. G. BERGER. 1998. Evolution of transcriptional regulatory elements within the promoter of a mammalian gene. Mol. Biol. Evol. **15**:312–315.

SMALE, S. T. 1994. Core promoter architecture for eukaryotic protein-coding genes. Pp. 63–81 *in* R. C. CONAWAY, and J. W. CONAWAY, ed. Transcription: mechanisms and regulation. Raven Press, New York.

SMALE, S. T., M. C. SCHMIDT, A. J. BERK, and D. BALTIMORE. 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. USA **87**:4509–4513.

SWOFFORD, D. L. 1989. PAUP: phylogenetic analysis using parsimony. Version 3.0b. Illinois Natural History Survey, Champaign.

TREVES, D. S., S. MANNING, and J. ADAMS. 1998. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli.* Mol. Biol. Evol. **15**:789–797.

VINCENT, K. A., and A. C. WILSON. 1989. Evolution and transcription of old world monkey globin genes. J. Mol. Biol. **207**:465–480.

WEIS, L., and D. REINBERG. 1997. Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. Mol. Cell. Biol. **17**:2973–2984.

WILEY, E. O. 1986. A study of the evolutionary relationships of *Fundulus* topminnows (Teleostei: Fundulidae). Am. Zool. **26**:121–130.

WILEY, S. R., R. J. KRAUS, and J. E. MERTZ. 1992. Functional binding of the "TATA" box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. Proc. Natl. Acad. Sci. USA **89**:5814–5818.

YEAN, D., and J. GRALLA. 1997. Transcription reinitiation rate: a special role for the TATA box. Mol. Cell. Biol. **17**:3809–3816.

ZENZIE-GREGORY, B., A. KHACHI, I. P. GARRAWAY, and S. T. SMALE. 1993. Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. Mol. Cell. Biol. **13**:3841–3849.

ZENZIE-GREGORY, B., A. O'SHEA-GREENFIELD, and S. T. SMALE. 1992. Similar mechanisms for transcription initiation mediated through a TATA box or an initiator element. J. Biol. Chem. **267**:2823–2830.

ZHANG, D. E., C. J. HETHERINGTON, S. TAN, S. E. DZIENNIS, D. A. GONZALEZ, H. M. CHEN, and D. G. TENEN. 1994. Sp1 is a critical factor for the monocytic specific expression of human CD14. J. Biol. Chem. **269**:11425–11434.

STEPHEN PALUMBI, reviewing editor

Accepted October 13, 1998