Structural and Functional Differences in the Promoter and 5' Flanking Region of Ldh-B Within and Between Populations of the Teleost Fundulus heteroclitus

Patricia M. Schulte, Marta Gómez-Chiarri and Dennis A. Powers

Hopkins Marine Station, Stanford University, Pacific Grove, California 93950

Manuscript received June 10, 1996
Accepted for publication November 22, 1996

ABSTRACT

We have investigated the mechanisms underlying differences in the transcriptional regulation of lactate dehydrogenase-B (Ldh-B) between northern and southern populations of a teleost fish, Fundulus heteroclitus. A 1-kb region immediately 5' of the gene was sequenced from populations throughout the species range. There were two major allele classes in the sample, one containing alleles from Maine and another containing those from Florida. Populations from intermediate localities contained both allele classes. Some individuals from Georgia had sequences intermediate between the two classes, representing either ancestral alleles or recombinants. Tests of neutrality were applied to determine whether observed variation was consistent with neutral expectations. Significant deviations from neutral expectations were detected for the 5' flanking region, but not for other loci. The functional consequences of flanking sequence variation were assessed by transfection of reporter gene constructs into cultured cells and injection into living fish. Consistent with observed variation in Ldh-B transcription rate between populations, significant differences in reporter gene activity were driven by flanking regions from northern and southern populations both in cell culture and in vivo. This functional differentiation, coupled with departures from neutral expectations, suggests that selection may have acted on the regulation of Ldh-B in F. heteroclitus.

Populations of the killifish, Fundulus heteroclitus, occur in bays, estuaries and salt marshes along the Atlantic coast of North America from Newfoundland to Florida. Although the species is sometimes split into two subspecies based on the genetic, morphological and behavioral differences between northern and southern populations, there is a wide zone at intermediate latitudes characterized by clinal variation in these characters (reviewed in Powers et al. 1993). It has been suggested that concordances in the geographic distribution of these characters may be the result of secondary contact between previously isolated forms (Powers and Place 1978; Able and Felley 1986; Gonzalez-Villaseñor and Powers 1990; Bernardi et al. 1993).

When two differentiated populations of a species come into contact, the cline formed at the intergrade zone will begin to erode due to gene flow unless selection acts to maintain the differentiation. Estimates of gene flow between populations of F. heteroclitus suggest that while the characteristics that distinguish the northern and southern forms may have arisen in isolation, the clines are most likely maintained by selection (Brown and Chapman 1991). Since some allelic isozymes are not clinal, while others are clinally distributed with slopes varying from gradual to steep (reviewed in Powers et al. 1993), it seems likely that the intensity and type of selection, if any, varies from locus to locus.

Allelic isozymes of the glycolytic enzyme lactate dehydrogenase-B (LDH-B) vary from homozygous fixation for one allele at the northern extreme of the species range to fixation for an alternate allele in the south (Powers and Place 1978). These alleles differ in their kinetic properties and are associated with differences in hatching time (Dimichele and Powers 1982a), developmental rate (Dimichele et al. 1986), swimming performance (Dimichele and Powers 1982b) and differential mortality in selection experiments conducted at elevated temperatures (Dimichele and Powers 1991). Taken together, these studies strongly suggest that natural selection is operating on allelic variants of Ldh-B.

In addition to differences in LDH-B kinetics, northern populations have twice as much LDH-B in liver tissue as their southern counterparts (Crawford and Powers 1989). This twofold difference in LDH-B enzyme concentration between northern and southern populations is associated with twofold differences in both the abundance of the transcript and the transcription rate of the gene (Crawford and Powers 1992), suggesting that the differences in LDH-B enzyme concentration are the result of changes at the level of transcription. It has been suggested that the differences in enzyme concentration are functionally important (Crawford and Powers 1989) because twice as much enzyme is required to maintain the same reaction velocity for every 10° decrease in temperature, and the mean annual water temperature at the northern and southern extremes of the species range differs by ~12°.
Although detailed mechanistic analyses of the regulation of transcription have been undertaken for many genes, there have been relatively few attempts to study the evolution of differences in transcriptional regulation within or between species (reviewed in Dickinson 1991; but see Odgers 1995). In this report we describe structural and functional variation in the promoter and 5′ flanking regions of Ldh-B within and between populations of F. heteroclitus, and address some of the genetic and evolutionary mechanisms that may be responsible for this variation.

**MATERIALS AND METHODS**

**DNA amplification and sequencing:** F. heteroclitus were obtained from St. Andrew's, New Brunswick; Wiscasset, Maine; Woods Hole, Massachusetts; Stone Harbor, New Jersey; Sapelo Island, Georgia; and Whitney Island, Florida. Spleens were dissected, and genomic DNA was isolated by proteinase K digestion and phenol/chloroform extraction. Approximately 1 kb of the Ldh-B 5′ flanking sequence was PCR amplified from genomic DNA with Tag DNA polymerase (Perkin-Elmer) using primers (FLK) AGATATATCCAGTTTTGCT and (UTR) CTGGGATCAGAGACTGAG (94° 1 min; 55° 1 min; 72° 2.5 min; 35 cycles). Primers were designed from a sequence of the Ldh-B 5′ flanking region obtained by screening a F. heteroclitus genomic library (Segal et al. 1996).

PCR products were end-polished with DNA polymerase I (Klenow fragment) and blunt-end cloned into pBluescript KS+ (Stratagene). At least five clones per individual were sequenced with internal primers on both strands using Sequenase 2.0 (Stratagene). Occasionally, three substantially different sequences were present in the original sample of five clones. In most cases this was the result of length variation in microsatellites contained within the cloned sequences. The probable cause of these errors is slippage of Tag polymerase during PCR, which has been shown to result in artefactual length variation in microsatellites (Holsbeck et al. 1994). If more than three alleles were detected, the PCR was repeated and treated as before. Majority rule consensus was used to choose the ‘correct’ sequence or sequences.

The sequence of the Ldh-B flanking region was obtained from a closely related congeneric species, F. grandis (obtained from Galveston Bay, Texas) for use as an outgroup. We were able to obtain the 5′ flanking region of F. grandis using the primers designed for F. heteroclitus. All sequences have been deposited in GenBank under accession numbers U59833-U59856. Sequence alignments are available from the authors on request.

**Phylogenetic analyses:** Sequences were aligned using GeneWorks (Intelligenetics) followed by alignment by eye. Because of the difficulty involved in aligning microsatellite sequences and the possibility of artefactual length variation (see above), all microsatellites and regions containing gaps were eliminated from genealogical analyses. Tree topologies were obtained using both maximum parsimony (MP) and neighbor-joining (NJ) methods. NJ trees were determined using the computer program MEGA (Kumar et al. 1993). MP trees were determined using both MEGA and PAUP (Swofford 1991). The degree of support for internal branches of the tree was further assessed by bootstrapping with 500 replicates (Felsenstein 1985) using PAUP for the MP trees and MEGA for NJ trees.

**Statistical analyses:** From the aligned sequences of the 23 F. heteroclitus alleles we estimated the parameter θ (Tajima 1983) excluding insertions and deletions. Tests of goodness of fit to an infinite sites models were applied (Tajima 1989; Fu and Li 1993), as well as a test of the concordance between polymorphism and divergence (HKA test, Hudson et al. 1987; Kreitman and Hudson 1991). To compare the levels of variation in mitochondrial and autosomal genes, the HKA test was modified so that the effective population size of mitochondrial genes was one-quarter that of autosomal genes (see e.g., Begun and Aquadro 1991; Ballard and Kreitman 1994). Modifications were also made to take account of the different number of alleles sampled for each locus (see e.g., Berry et al. 1991). Variation along the sequences was also investigated by making use of a “sliding window” approach (Hudson and Kaplan 1988). The recombination rate in the 5′ flanking sequence was estimated by calculating the recombination parameter (C) based on the variance in the distribution of numbers of sites at which pairs of alleles differ, using the method of Hudson (1987). Linkage disequilibrium between sites was estimated using the Fisher's exact test (Weir 1990).

**Transfection assays:** Portions of the 5′ flanking sequence were obtained by PCR amplification from the cloned 5′ flanking sequences ME4, FL1, GA2, GA4 and GA5 using primer UTRb (AGATCAGAAGAAAAGTCTGGGATCAGAGACTGAG), in the 5′ untranslated region of the mRNA, and one of the following primers: FLK; 5′- GATATATTCCAGTTTTGCT (94° 1 min; 55° 1 min; 72° 2.5 min; 35 cycles). Primers were designed from a sequence of the Ldh-B 5′ flanking region obtained by screening a F. heteroclitus genomic library (Segal et al. 1996). PCR products were end-polished with DNA polymerase I (Klenow fragment) and blunt-end cloned into pBluescript KS+. The sequence of the Ldh-B flanking region was obtained by screening a F. heteroclitus genomic library (Segal et al. 1996). PCR products were end-polished with DNA polymerase I (Klenow fragment) and blunt-end cloned into pBluescript KS+. The sequence of the Ldh-B flanking region was obtained by screening a F. heteroclitus genomic library (Segal et al. 1996).

**Plasmid characterization:** Plasmids were introduced into F. heteroclitus using the calcium phosphate method (Sambrook et al. 1989). All transfections utilized 10 μg of the luciferase reporter gene plasmid (pCMVlucZ, containing the β-galactosidase reporter gene) as an internal control for transfection efficiency. Three days after transfection, cells were lysed and luciferase activity in the lysate was detected using the Enhanced Luciferase Assay Kit (Analytical Luminescence Laboratory). Light emission was measured with a single photon count program in a TM Analytic (Mark V) liquid scintillation counter. Galactosidase activity was measured spectrophotometrically using standard techniques (Sambrook et al. 1989). Results are expressed as fg luciferase per unit β-galactosidase activity. Differences in activity between constructs were tested using a two-way ANOVA followed by multiple comparisons among means (Sokal and Rolf 1981).

**The functional activity of representative full length constructs from Maine and Florida was also assessed by direct injection of reporter gene constructs into F. heteroclitus in vivo. Similar approaches have been utilized to analyze the expression of genes in both large and small mammals (Wolff et al. 1990; Von Harsdorff et al. 1993). F. heteroclitus obtained from Whitney Island, Florida were acclimated to 50% seawater at 20° for a minimum of 4 months. Fish were lightly anesthetized with 3-aminobenzolic acid ethyl ester (MS-222) and 50 μl of transfected solution was injected into the ventral trunk musculature between the dorsal fin and lateral line. Transfection solution consisted of phosphate buffered saline containing 50 μg of total plasmid DNA (full length constructs ME4 or FL1 see above), 10 μg of control plasmid (CMVtk β-galactosidase), and India ink to localize the site of injection. After 24 hr the fish were sacrificed by rapid decapitation and the region
RESULTS AND DISCUSSION

Structure of the Ldh-B flanking region: Figure 1 shows the general structure of the F. heteroclitus Ldh-B 5' flanking region. The transcription start sites for the gene are located in a cluster between sites 1095 and 1128, with the two major start sites at 1122 and 1128 (Segal et al. 1996). The proximal promoter is located between sites 950 and 1150, and contains several highly conserved sequence motifs (sites 1040–1069) that are similar to the Sp1 transcription factor binding site (Dykan and Tjian 1983). The major transcription start sites are associated with consensus initiator motifs (Roeder 1991). Both the initiator and Sp1 binding sites are functionally important components of the F. heteroclitus proximal promoter (Segal et al. 1996). No TATA box is present, typical of the promoters of many housekeeping genes (Dykan 1986).

Sequence variation in the flanking region: One striking difference between alleles is the presence of a number of microsatellites and small insertion-deletions in the promoter region (Figure 1). Two sets of TCC repeats of variable length (sites 1017–1040 and 1070–1088; Figure 1) are located surrounding the putative Sp1 sites (sites 1041–1069). The basic repeat unit of the Sp1 site (CCGGGCC) is present once in alleles NB, ME1-ME5, WH1, NJ1 and NJ2; twice in alleles WH2, NJ3, NJ4, GA1, GA2, GA3, GA5, GA6, FL1-FL5 and F. grandis; and three times in allele GA4. We have previously shown that these microsatellites are functionally important components of the F. heteroclitus Ldh-B promoter, as assessed by transfection into a heterologous cell line (Segal et al. 1996). Two additional microsatellites are located within 200 bp upstream of the promoter: a small ATC repeat (sites 882–893) and a highly variable sequence composed almost entirely of thymine and adenine bases (sites 798–823). While no microsatellites are located within 1 kb upstream of the transcription start sites, there are at least 10 additional small insertions or deletions (Figure 1).

As illustrated in Figure 2, sites that are variable within populations are fairly evenly distributed throughout the flanking region, whereas sites that vary between populations are unevenly distributed. The promoter (sites 950–1150) is the least variable part of the flanking region, both within and between populations. It has only 0.89 variable sites per 100 bp, and is less variable than silent sites within the coding region (which has 1.4 variable sites per 100 bp; Bernardi et al. 1993). Low variability has been reported for the promoter regions of other genes (Kreitman and Hudson 1991; O’dgers et al. 1995), consistent with functional constraints. In contrast to the low variability in the promoter region, there is a peak of variability between the Maine and Florida populations near the central region of the 5' flanking sequences (Figure 2). In this central region, the average pairwise difference between alleles from Maine and Florida populations is approximately five times that of the fourfold degenerate sites in the coding region (or 7.46 per 100 bp), but this region is not highly variable within populations (Table 1; Figure 2).
Geographical population structure: Most alleles of the 5' flanking region fall into one of two distinct groups, one containing all of the alleles from Maine and another containing all of the alleles from Florida (Figure 3). Alleles from Maine and Florida populations are almost as divergent from each other as they are from the outgroup, F. grandis. The average pairwise number of variable sites is 39.2 between Maine and Florida populations of F. heteroclitus, whereas the average pairwise number of variable sites between F. grandis and F. heteroclitus from Maine or Florida is 41.6 and 42.6, respectively. There is higher variability within the Florida population than within the Maine population (average pairwise difference within Florida and Maine populations is 12.9 and 0.64 sites, respectively; Figure 3; Table 1). Geographically intermediate populations from New Jersey and Woods Hole contain both Maine-like and Florida-like alleles. As a result, these populations have the highest level of sequence variability (average pairwise difference within the New Jersey sample is 22.9 sites).

Results from previous studies using both nuclear and mitochondrial markers suggest a sharp disjunction between northern and southern F. heteroclitus genotypes (Gonzalez-Villaseñor and Powers 1990; Bernardi et al. 1993). However, in phylogenetic analyses of the 5' flanking sequences, several of the Georgia alleles (GA1, GA2, GA3, GA4) appear to be intermediate between Maine and Florida alleles (Figure 3). These alleles could represent ancestral forms or recombinants between Maine and Florida alleles. Putative recombinants have previously been observed in samples of the Ldh-B coding region from geographically intermediate populations (Powell et al. 1992; Bernardi et al. 1993).

Recombination: The relative rates of recombination and mutation in a gene region can be estimated based on the variance in the distribution of numbers of sites at which pairs of alleles differ (Hudson 1987). Taking into account only point mutations, we estimate the recombination parameter (C) to be 22.0 for the 5' flanking region (θ = 30.87; Sx = 242; N = 23; Hudson 1987). Since C = 4Nε (where ε is recombination rate) and θ = 4Nu (μ = mutation rate), the ratio Cθ should be an index of the relative rates of recombination and mutation. The Cθ ratio for the Ldh-B flanking region in F. heteroclitus is ~0.71, suggesting that recombination and mutation are about equally frequent on a per nucleotide basis. However, this estimate assumes that the alleles were drawn from a single randomly mating population. Because populations of F. heteroclitus throughout the species range may not satisfy this assumption, we also estimated the parameter C using only alleles from the Georgia and Florida populations. In this case, the Cθ is ~1.3, suggesting again that recombination and mutation are about equally frequent in the 5' flanking region (θ = 30.58; Sx = 206; n = 11; C = 40).

The results of the Fisher's exact test (Weir 1990) suggest that the variable sites in the Ldh-B flanking sequence are not in strong linkage disequilibrium (Figure 4). However, there is some indication that the flanking region can be divided into three major recombining blocks: one containing the promoter region, another in the central region of the sequence, and a third group upstream. We subdivided the sequences into these blocks, and constructed three separate phylogenetic trees (Figure 5). Trees constructed for the upstream and central regions suggest different but reasonably well supported genealogies, which is consistent with recombination. For example, alleles GA1–GA4, which appear intermediate between Maine and Florida alleles in Figure 3, group with some of the Florida alleles when only the upstream region is considered (Figure 5) and with the Maine alleles and F. grandis when only the central region is considered (Figure 5). While the hypothesis of recombination between Maine and Georgia alleles is reasonable, the possibility of recombination with a F. grandis allele cannot be dismissed a priori, since the ranges of F. heteroclitus and F. grandis overlap in northeastern Florida (Lee et al. 1980), and hybridization within the genus is common (Newman 1998).

As well as the variable positioning of alleles GA1–GA4, there are additional differences between the phylogenies illustrated in Figure 5 that suggest other recombination events may have occurred. The genealogy of the region containing the promoter (sites 655–1169)
<table>
<thead>
<tr>
<th>Site #</th>
<th>ME1</th>
<th>ME2</th>
<th>ME3</th>
<th>NB</th>
<th>WH1</th>
<th>WH2</th>
<th>NJ1</th>
<th>NJ2</th>
<th>NJ3</th>
<th>NJ4</th>
<th>GA1</th>
<th>GA2</th>
<th>GA3</th>
<th>GA4</th>
<th>GA5</th>
<th>GA6</th>
<th>FL1</th>
<th>FL2</th>
<th>FL3</th>
<th>FL4</th>
<th>FL5</th>
<th>F. grandis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TTACATGGTGCAACTGACCTGCTGGCAAAATTGAAACTCCTGACCTTGACTGAGCTGCTGCGAAATTAGGGCCCTCTCGGTGCGGCCTGGCCCTGACCTCTCATAGTGACAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 1**

Polymorphic sites within the 5'-flanking region and sites that have diverged between *F. heteroclitus* and *F. grandis*.
is particularly difficult to reconstruct. Although the three major groups of alleles (maine-like, FL1-like and FL5-like) are still apparent, no well supported genealogy can be determined for many of the sequences and a large number of conflicting phylogenies are equally good. Since results of this kind are expected when reconstucting the genealogies of sequences that have undergone recombination (HEIN 1993), these results are consistent with frequent recombination in the promoter region, although it should be noted that when the data are partitioned, statistical power to detect linkage disequilibrium may be very weak.

The genealogies in Figure 5 are based only on point mutations. However, the majority of the variation in the promoter is due to length variation in microsatellites. To further address the possibility of recombination within the promoter, we performed Fisher’s exact tests to determine whether the alleles at the microsatellites are linked. There was no significant evidence of linkage disequilibrium (data not shown), suggesting that recombination may have occurred, although the possibility of parallel mutations cannot be ruled out. It has been suggested that the presence of microsatellites may increase the likelihood of polymerase slippage, gene conversion and recombination (see DOVER 1993 for review), which provides a potential mechanism for recombination in the microsatellite-rich Ldh-B promoter of F. heteroclitus.

Tests of neutrality: The hypothesis that all mutations in the 5′ flanking region are selectively neutral can be tested in a variety of ways. The Tajima test for data from a single locus (Tajima 1989) compares the estimate of θ based on the average number of pairwise differences to that based on the number of segregating sites. The test of Fu and Li (1993) with outgroup compares estimates of the number of segregating sites to the number of mutations on external branches expected under neutrality. The results of these tests as applied to the flanking region and to previously published Ldh-B cDNA and
mitochondrial cyt-b (cyt-b) data (Bernardi et al. 1993) are presented in Table 2. There is no evidence for departure from neutrality when these tests are applied to pooled data from all populations, although both the Fu and Li (1993) and Tajima (1989) tests tend toward slightly negative values. Slightly negative values are to be expected when one or a few sequences are drawn from several different geographic regions, since such a sample is likely to have an excess of unique or low frequency polymorphisms. When the approach of Tajima (1989) is applied to each population separately, there is no evidence for departures from neutrality for any of the loci (Table 2). On the other hand, the Fu and Li (1993) test rejects the null hypothesis that all sites are neutral in the 5' flanking region for the Maine and New Jersey populations, but not for the Ldh-B coding region or cyt-b (Table 2). The difference between the two tests may reflect the low sensitivity of the Tajima (1989) analysis when sample sizes are small (see Stmonsens et al. 1995 for discussion).

The positive D-statistic in the Fu and Li (1993) test for the flanking region in New Jersey population (Table 2) suggests that there are two ancient alleles within the population. This result is consistent with either secondary intergradation between northern and southern forms, or balancing selection (Fu and Li 1993). Because data are not available for the other loci in the New Jersey population, we cannot distinguish between these two possibilities, but previous studies (Gonzalez-Villalasenor and Powers 1990) have also suggested that clines in F. heteroclitus may be the result of secondary intergradation.

The negative D-statistic for the flanking region of the Maine population (Table 2) suggests that the alleles in this population are of recent origin, which is consistent with a number of interpretations including purifying selection, background selection (Charlesworth et al. 1993), selective sweep, population bottleneck, or rapid population expansion (Fu and Li 1993). Purifying selection and background selection are unlikely in this case because they would be expected to influence variation in all populations. Since Maine populations have lower variability than southern populations for most molecular markers (Powers et al. 1993), a population bottleneck or rapid population expansion are more reasonable explanations for the departures from neutrality in

<table>
<thead>
<tr>
<th>Population</th>
<th>5' flanking region</th>
<th>Ldh-B cDNA</th>
<th>Cytochrom-b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tajima Fu and Li</td>
<td>Tajima Fu and Li</td>
<td>Tajima Fu and Li</td>
</tr>
<tr>
<td>All</td>
<td>-0.82 -0.72</td>
<td>-0.18 -0.17</td>
<td>-0.29 -0.73</td>
</tr>
<tr>
<td>North</td>
<td>-0.62 -1.99*</td>
<td>-0.59 -0.85</td>
<td>-0.54 0.12</td>
</tr>
<tr>
<td>New Jersey</td>
<td>0.27 1.91*</td>
<td>NA NA</td>
<td>NA NA</td>
</tr>
<tr>
<td>South</td>
<td>-0.12 -0.75</td>
<td>-0.44 -0.77</td>
<td>-0.80 -0.89</td>
</tr>
</tbody>
</table>

Tajima D statistics are calculated as in Tajima (1989). Fu and Li D-statistics (with outgroup) are calculated as in Fu and Li (1993) using F. grandis as outgroup for the 5' flanking region and F. parvipinnus as outgroup for Ldh-B cDNA and cytochrome-b. Ldh-B cDNA and cytochrome-b sequences from Bernardi et al. (1993). Populations are as follows: North includes Maine and New Brunswick for the 5' flanking region; Maine and Nova Scotia for cDNA and cytochrome-b, South includes all individuals from Georgia and Florida for all loci. *Significant D-statistics, P < 0.05. NA indicates data not available.
Within and between species (Kreitman and Hudson prediction of the neutral theory that levels of polymorphism within a species should be correlated with levels of divergence between species across all loci. This test is sufficiently sensitive to detect a population bottleneck or rapid changes in population size or selection. However, the results of HKA testing for the Ldh-B coding region or cyt-b in the Maine population (Table 2) as would be expected if a population bottleneck or rapid population expansion had occurred. If the Fu and Li test is sufficiently sensitive to detect a population bottleneck at cyt-b and the coding region, this discrepancy between loci suggests the possibility of a selective sweep at the 5' flanking region in the northern population.

The possibility that various regions of the genome are subject to different evolutionary forces can be examined using the HKA test (Hudson et al. 1987). By comparing the ratio of intraspecific to interspecific nuclotide variation at several loci, the HKA approach tests a prediction of the neutral theory that levels of polymorphism within a species should be correlated with levels of divergence between species across all loci. This test should be able to distinguish between a population bottleneck, which would affect all loci, and a selective sweep, which should only affect the locus under selection. We compared variation at three loci: the Ldh-B coding region, mitochondrial cyt-b (Bernardi et al. 1993) and the Ldh-B 5' flanking region, using a modified HKA test that compares average pairwise difference within and between species (Kreitman and Hudson 1991). When polymorphism in northern and southern populations is compared to divergence between them, there is significant deviation from neutral expectations only in comparisons involving the Ldh-B 5' flanking region (Table 3). This suggests that the deviations from neutrality are the result of variation in the 5' flanking region rather than in one of the other two loci, possibly as a result of selection. However, the results of HKA tests should be interpreted with caution since the violation of any one of the underlying assumptions may result in rejection of the null hypothesis. The assumptions that may be violated by the data presented here are:

(1) no selection, (2) large constant population size, (3) no linkage between loci, (4) no recombination within a locus, and (5) no ancestral polymorphism. Violations of assumptions 3–5 are thought to make the test more censorative (see for example Hudson et al. 1987; Ford et al. 1994 for discussion) and thus are unlikely to account for the deviations from neutrality. Therefore these results are most likely accounted for by either changes in population size or selection.

**Functional analyses**: If selection is acting on the 5' flanking region, the minimal condition that must be met is evidence of some form of functional differentiation between alleles. To determine whether any of the sequence variability in the 5' flanking region of F. heteroclitus has functional consequences for transcriptional regulation, we compared the ability of several randomly selected alleles from northern and southern populations to promote the transcription of a luciferase reporter gene in human embryonic kidney (HEK 293) cells. We chose to perform the initial functional experiments using HEK 293 due to the lack of cell lines from F. heteroclitus and the fact that most fish cell lines and primary cultures tested showed low and variable transfection efficiencies (M. Gómez-Chiarri, unpublished results). HEK 293 has extremely good growth and high transfection efficiencies that maximize the sensitivity of the functional assays. This may be critical since activity differences in natural populations appear to be small (for example, see Oggers et al. 1995), and the twofold differences expected for F. heteroclitus Ldh-B might be essentially undetectable under many conditions.

We first assessed the luciferase activity promoted by one of the Maine alleles (ME4) and several of the most divergent alleles from southern populations (GA2, GA4, GA5, FL.I), including two of the putative recombinants. Length variation in the microsatellites that alter the spacing between conserved, functionally important sites in the promoter does not appear to have strong effects on transcription since there are no differences in the activity of some constructs that differ in microsatellite length (e.g., GA5 and ME4; GA2 and GA4; Figure 6).
contrast, the Maine allele (ME4) promotes significantly greater transcription than the mean level promoted by the southern alleles (GA2, GA4, GA5, FL1) (20.62 ± 2.00 vs. 15.21 ± 7.41 fg luc/µg β-gal; P < 0.05) and significantly greater transcription than all but one southern allele (GA5). This is consistent with the differences in Ldh-B transcription rate between populations (CRAWFORD and POWERS 1992). To localize the sequences responsible for the observed differences in transcriptional activity between alleles, we compared the activity of various deletions of FL1 and ME4. As can be seen in Figure 7, there are significant differences between ME4 and FL1 for the 400-bp construct. There are no significant differences in activity among northern constructs, but for southern constructs the shortest construct gives significantly greater activity than either of the two longer constructs (Figure 7). These results suggest that there are sequences repressing transcription in the region deleted from the shortest of the constructs (sites 585–709) in FL1 and that this repressor activity is absent in ME4.

We examined the region between sites 585 and 709 in FL1 for known transcription factor binding sites using the program Signal Scan (PRESTIDGE 1993). A sequence located between sites 704 and 710 in the FL1 allele is identical to the MTV-GRE negative regulatory motif (LANGER and OSTROWSKI 1988). This putative repressor motif is altered in ME4 due to a polymorphism at site 707 (Table 1). GA5 also lacks this motif (Table 1), which might account for the high level of transcription promoted by this allele (Figure 6). However, further experiments are required to assess this possibility since GA5 also differs from FL1 at positions other than the putative repressor.

Because these functional assays were performed using mammalian cells in culture, these results should be interpreted with caution. However, the known behavior of this repressor is consistent with that of F. heteroclitus Ldh-B. The MTV-GRE repressor is known to function only at low glucocorticoid concentration (LANGER and OSTROWSKI 1988), and differences in Ldh-B gene regulation between populations of F. heteroclitus are observed only at higher temperatures (SEGAL and CRAWFORD 1994) when glucocorticoid levels are low (LEACH and TAYLOR 1977). This suggests that a similar steroid-mediated mechanism might be responsible for the differences in Ldh-B regulation between northern and southern F. heteroclitus.

Because the action of regulatory sequences in heterologous cell cultures is not always equivalent to their action in vivo, we also assayed the function of the reporter gene constructs from Maine and Florida in F. heteroclitus muscle. When injected into fish obtained from the Florida population, the Maine construct promotes twofold greater luciferase activity than the Florida construct (Figure 8). While we have yet to directly test the function of these constructs in liver, the results of both the cell culture and intramuscular injections are consistent with the differences in Ldh-B transcription observed between northern and southern populations (CRAWFORD and POWERS 1992).
There is extensive geographic variation in the 5’ flanking sequences of the Ldh-B gene in different populations of Fundulus heteroclitus. Maine and Florida populations can be clearly distinguished, based on phylogenetic analyses of their alleles. Geographically intermediate populations are the most variable, and contain both northern and southern allele classes, which is consistent with a secondary intergradation model of cline formation in Fundulus heteroclitus. Various tests of neutrality indicate that the 5’ flanking region is not behaving according to neutral expectations. These deviations from neutrality could be the result of a number of processes, including changes in population size and selection. In vitro transfection assays and in vivo injections show that there are functional differences between northern and southern alleles consistent with the differences in Ldh-B transcription between populations (Crawford and Powers 1992). This functional differentiation, coupled with departures from neutral expectations in the 5’ flanking region, suggests the possibility that selection has acted on the regulation of Ldh-B in Fundulus heteroclitus.

Conclusions: There is extensive geographic variation in the 5’ flanking sequences of the Ldh-B gene in different populations of Fundulus heteroclitus. Maine and Florida populations can be clearly distinguished, based on phylogenetic analyses of their alleles. Geographically intermediate populations are the most variable, and contain both northern and southern allele classes, which is consistent with a secondary intergradation model of cline formation in Fundulus heteroclitus. Various tests of neutrality indicate that the 5’ flanking region is not behaving according to neutral expectations. These deviations from neutrality could be the result of a number of processes, including changes in population size and selection. In vitro transfection assays and in vivo injections show that there are functional differences between northern and southern alleles consistent with the differences in Ldh-B transcription between populations (Crawford and Powers 1992). This functional differentiation, coupled with departures from neutral expectations in the 5’ flanking region, suggests the possibility that selection has acted on the regulation of Ldh-B in Fundulus heteroclitus.

We thank D. Busher, D. Crawford, L. Dimichele, B. Marshall, V. Pence, J. Segal and T. Singer for help in obtaining various fish samples. Thanks to J. Quattro, G. Bernardi, K. Callenotte, S. Clemenchen, C. Reeb, W. Eanes and the anonymous reviewers of this manuscript for their helpful comments and suggestions. Particular thanks to M. Kreitman for his assistance with the HKA analyses. This work was supported by National Science Foundation grants BSR-9022648 and DEB-96-29572 to D.P., and National Science Foundation dissertation improvement grant DEB-95-20567 to P.S.

LITERATURE CITED


KUMAR, S., K. TAMURA and M. NEI, 1993 MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park, PA.


NEWMAN, H. H., 1908 The process of heredity as exhibited by the development of Fundulus hybrids. J. Exp. Zool. 5: 503-571.


Communicating editor: W. F. EAMES