Neutral and adaptive variation in gene expression

Andrew Whitehead*† and Douglas L. Crawford‡

*Department of Biological Sciences, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803; and ‡University of Miami, Rosenstiel School of Marine and Atmospheric Science, 4600 Rickenbacker Causeway, Miami, FL 33149

Variation among populations in gene expression should be related to the accumulation of random-neutral changes and evolution by natural selection. The following evolutionary analysis has general applicability to biological and medical science because it accounts for genetic relatedness and identifies patterns of expression variation that are affected by natural selection. To identify genes evolving by natural selection, we allocate the maximum among-population variation to genetic distance and then examine the remaining variation relative to a hypothesized important ecological parameter (temperature). These analyses measure the expression of metabolic genes in common-gardened populations of the fish *Fundulus heteroclitus* whose habitat is distributed along a steep thermal gradient. Although much of the variation in gene expression fits a null model of neutral drift, the variation in expression for 22% of the genes that regress with habitat temperature was far greater than could be accounted for by genetic distance alone. The most parsimonious explanation for among-population variation for these genes is evolution by natural selection. In addition, many metabolic genes have patterns of variation incongruent with neutral evolution: They have too much or too little variation. These patterns of biological variation in expression may reflect important physiological or ecological functions.

evolutionary analysis | *Fundulus* | microarray | phylogenetic comparative approach | genomics

Gene expression has been hypothesized to be of adaptive importance (1), and heritable variation that affects fitness is necessary for evolution by natural selection. Although adaptive differences in expression have been identified in single-gene studies (2–6; see ref. 7 for review), microarray approaches offer great promise to rigorously address this hypothesis because they assay many loci at once. Furthermore, it is generally agreed that much of variation in gene expression for a particular environmental condition has a genetic basis (8, 9) according to studies in yeast (10–13), *Drosophila* (14–18), mice (19), and humans (20–22).

Widespread heritable variation combined with extensive natural variation in gene expression revealed by microarray studies (11, 13, 14, 23–26) provides the substrates for evolution. However, two evolutionary forces govern the variance of traits among taxa: neutral drift and natural selection. Under a neutral drift model, the variation in a trait has little biological effect§ and is a function of genetic distance: Traits will be more similar among closely related taxa than among more distantly related taxa (27–29). If natural selection has occurred, the variation in a trait affects an organism's fitness and is a function of the ecological setting: Traits are conserved or diverge depending on the specific ecological pressures (30). Recent studies suggest that much of the extensive variation in gene expression among individuals and taxa is simply random neutral divergence (31, 32), whereas others have found extensive hallmarks of selection (16, 18, 33). However, no specific adaptive hypotheses have been applied to microarray data to refute or substantiate these claims. Here we apply a correction for genetic relatedness (i.e., the phylogenetic comparative method; refs. 34–36) to reveal variation most parsimoniously accounted for by neutral evolution. After removing the neutral phylogenetic effects, remaining variation significantly associated with an independent ecological factor, such as habitat temperature, would be most parsimoniously accounted for by natural selection. This approach is conservative because it attributes maximum among-population variation to genetic distance before considering natural selection.

This approach was applied to *Fundulus heteroclitus*, a teleost fish widely distributed along the Atlantic coast of North America, where there is a change of 1°C per degree latitude or ≈12.5°C between Maine and Georgia (37). In this species, there is evidence for local adaptation to this clinal variation in temperature (3, 37–39). Also, adaptation may be more prevalent in this species because of large local populations (census population size >10,000 for a single estuarine creek, $N_e > 10^5$; ref. 65) in which small selective pressures should dominate the effects of genetic drift. However, because many ecological factors may covary with temperature along the latitudinal gradient from Maine to Georgia, caution should be exercised in identifying only temperature as the specific agent of natural selection (30). Thus, temperature is considered a proxy for environmental variation that we predicted affects the evolution of gene expression.

Evolution by natural selection requires heritable variation in traits that affect fitness. Although the heritable variation among individuals in gene expression was not directly ascertained, all individuals were subject to common conditions and acclimated (see *Methods*) for >2 months. Acclimation is well studied (40, 41), and acclimating fish to a common environment minimizes physiological differences caused by differences in an animal's native habitat. Thus, much of the variation in gene expression is unlikely to be due to the native habitat temperature and more likely represents both genetic and other random biological sources of variation. Although raising animals at one common temperature reduces environmental influences, it ignores gene-by-environment interactions. Thus, heritable differences due to complex interactions are not ascertained.

Results

Population Genetics. Individuals for this study were collected from Maine, Connecticut, New Jersey, North Carolina, and Georgia, and extensive structure existed among populations (Fig. 1). Population genotyping by using microsatellite markers yielded pairwise F_{ST} estimates ranging from 0.01 to 0.24 (all statistically significant, $P < 0.05$). The neighbor-joining tree (Fig. 1) supports previous analyses of these populations, which show a break between northern and southern groups at the Hudson River (43–45). These data also indicate significant isolation-by-distance (Mantel test, 1,000 permutations: $P = 0.024$, $r = 0.65$). Thus, the spatially separated groups appear to follow indepen-
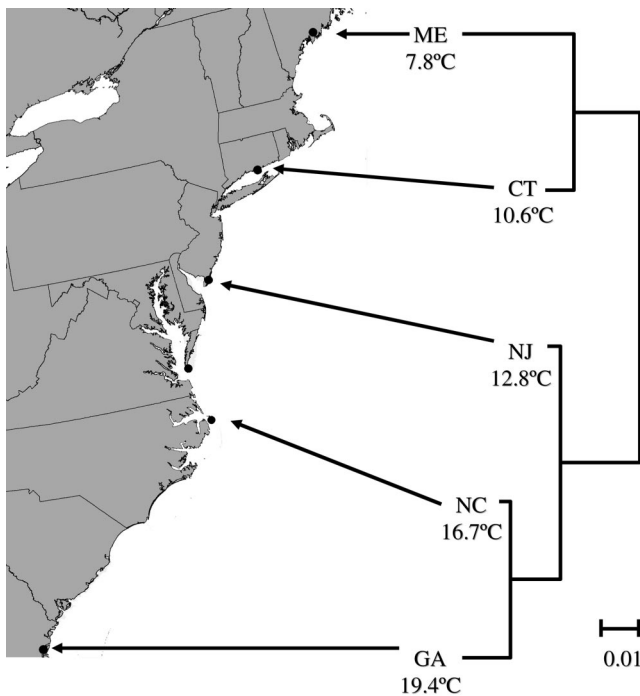
EVOLUTION

Fig. 1. Fundulus habitat sites and microsatellite-derived neighbor-joining tree. Collection sites along the United States Atlantic seaboard: ME, Wiscasset, ME; CT, Clinton, CT; NJ, Stone Harbor, NJ; NC, Roanoke Island, NC; GA, Sapelo Island, GA; corresponding median annual habitat temperatures (°C) are averaged over 30 years. Dendrogram is a neighbor-joining tree constructed from pairwise Cavalli-Sforza and Edwards' chord distances (42) calculated from microsatellite allele frequencies.

dent demographic trajectories, allowing application of phylogenetic analyses (35).

Gene Expression Variation Among Individuals and Populations. Expression for 329 genes involved in central metabolic pathways was measured by using eight technical replicates for each of five males per population sampled from five populations ($n = 25$). The mean coefficient of variation (CV; equals SD/mean) for technical replication was 2.54%, with >95% of genes with CV < 5%. A list of all genes and results from all statistical tests can be found as Table 2, which is published as supporting information on the PNAS web site. All fish used appeared healthy and unstressed. Because fish were raised in a common laboratory setting for 2 months, differences are likely not due to acclimatization to native habitat (40, 41). Differences in body weight or sex are unlikely sources of significant variation because body weight does not affect gene expression in this species (25) and only postreproductive males were used. Additionally, when blood was sampled four times over a 6-week period from five different individuals, there was little additional variation (D.L.C., unpublished data). These results demonstrate that RNA isolation, array hybridization, and random biological differences (e.g., temporal variation in the physiological state of individuals) contribute little additional variation. Instead we suggest that much of the variation is more likely due to genetic sources of variation. Alternatively, epigenetic or early developmental effects could establish an irreversible phenotype and cannot be ruled out. However, given that much of gene expression variation is heritable (8), it seems more parsimonious to assume a genetic basis (although this basis remains to be verified). Other sources of variation (e.g., random biological differences, differences in social status, and general well being)

seem unlikely to covary with temperature and, thus, would make any significant relationship more difficult to discern.

Gene expression varied extensively among individuals within populations (69%, 227 of 329 genes at $P < 0.05$; false discovery rate <1%). False discovery rate (FDR) estimates the proportion of tests that reject the null hypothesis that are false; thus with an FDR of <1%, ≈2 of the 227 significant differences are false positives. False discovery rates were estimated as described by Storey and Tibshirani (46) with the application of QVALUE software provided at http://faculty.washington.edu/~jstorey/qvalue/index.html.

Among populations, 12%, or 41 of 329 genes, were significantly different ($P < 0.05$; FDR = 22%). This relatively large FDR suggests that many of the 41 differences may be false positives, and, thus, the difference among populations may be as low as 10% of all measured genes. Regression analyses suggest that habitat temperature (see *Methods*) accounted for significant variation for 18% of genes (58 of 329, FDR = 21%; average $r^2 = 0.68$). Of these 58 genes, approximately equal numbers increase (28) or decrease (30) in expression with colder latitudes (Fig. 2*A*). Among these genes that regress significantly with temperature, the clustering of populations (top of Fig. 2*A*) is similar to the topology of the genetic distance neighbor-joining tree (Fig. 1). The apparent covariance between habitat temperature differences and genetic distances illustrates the problem of relatedness among taxa: Either of these factors could be responsible for the divergence in gene expression among populations.

In contrast to the patterns of the temperature genes indicating ecological effect, genetic distance relationships (phylogeny) accounted for a significant proportion of among-population expression variation for 15% of genes (50 of 329, FDR = 18%; Mantel test). Identification of genes with a phylogenetic component by the PHYSIG program (47, 48) (a permutation test that examines correlation between trait values and phylogeny branch lengths) corroborated Mantel results (Pearson's correlation coefficient = 0.84; $P < 0.001$). The most parsimonious explanation for the among-population variation in expression for this subset of genes is random genetic drift because patterns of gene expression correlate with genetic distance. Not surprisingly, hierarchical clustering of the genes with a significant phylogenetic signal (Fig. 2*B*) produces a dendrogram identical to the genetic distance neighbor-joining tree topology (Fig. 1). Natural selection also may contribute to this phylogenetic pattern because habitat temperature covaries with genetic distance, but as in other analyses (31, 32), the most parsimonious explanation for this pattern is neutral drift.

Evolutionary Analysis. To determine whether natural selection is affecting patterns of gene expression requires evidence of departure from neutrality. This evidence would include a significant covariation between gene expression and native habitat temperature after removing gene-specific phylogenetic effects. Temperature was chosen because of the steep thermal cline in habitat temperature and previous research indicating adaptation to temperature (3, 37–39). Approximately half of the 58 genes that have a significant habitat temperate component also have a significant phylogenetic component (Fig. 3). To correct among-population expression data for nonindependence due to phylogeny for each gene separately (48, 49), a matrix of expected covariances among populations was constructed by using branch lengths of the microsatellite-derived dendrogram. After removing phylogenetic effects, the expression levels of 13 genes (22% of the 58 temperature genes) significantly regressed with habitat temperature (Fig. 3) (1,000 random permutations; FDR = 3%). Many of these genes appear related to temperature variation in other data sets. For example, 6-phosphogluconate dehydrogenase (*6PGD*), cold-inducible RNA-binding protein (*CIRP*), cytochrome C oxidase (*COXE*), δ-1-pyrroline-5-carboxylate dehydrogenase (*PUT2*), glucose-6-phosphate isomerase (*GPI*),
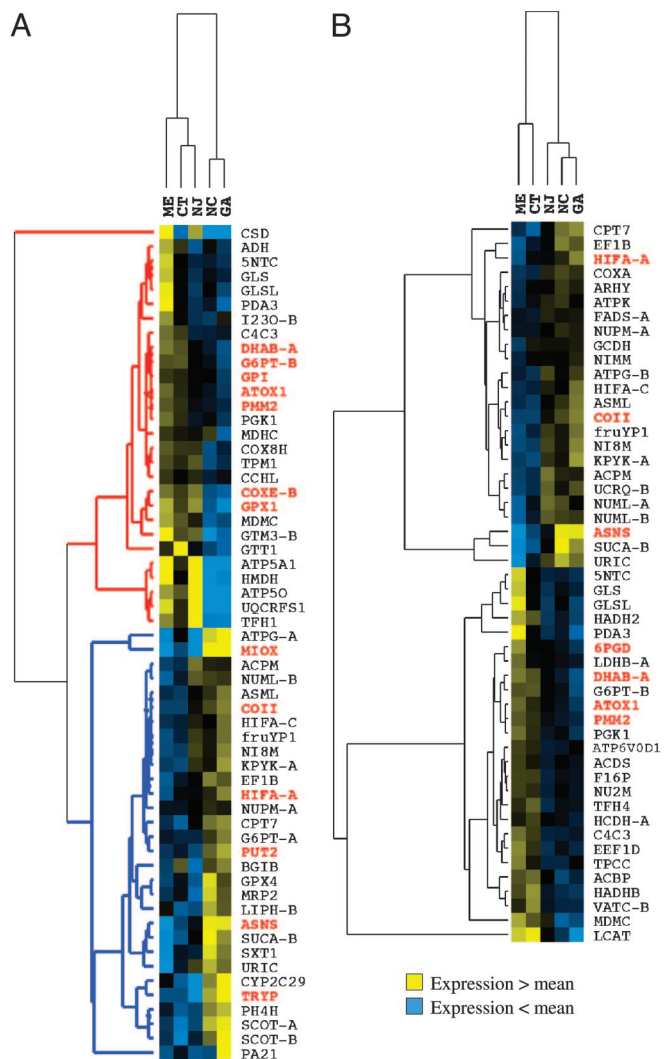
Fig. 2. Hierarchical clustering indicating genes with ecological (habitat temperature) and phylogenetic components to among-population variation in expression. Patterns of expression [yellow and blue represent expression greater than (yellow) and less than (blue) grand mean, respectively] for genes that correlate with habitat temperature (*A*) and that correlate with genetic distance (*B*). In *A*, gene clusters reflect decreases (red dendrogram) or increases (blue dendrogram) in expression with habitat temperature. Red bolded gene names are those that regress with habitat temperature after correction of observations for expected nonindependence due to phylogeny (phylogenetic generalized least squares method) and, thus, appear to be evolving by natural selection.

Fig. 3. Relationships between ecological and phylogenetic effects on among-population variation in gene expression. For each gene, the explained variation (r^2) for phylogeny (genetic distance based on Cavalli-Sforza and Edwards' chord distances (42) calculated from microsatellite allele frequencies) is plotted against the explained variation (r^2) for habitat temperatures. Venn diagram is for the numbers of genes that have significant regression with habitat temperature (blue), phylogeny (green), or both temperature and phylogeny (orange). Colors of spots in graph correspond to Venn diagram. Enlarged spots are the 13 genes that regress significantly with habitat temperature after correcting for phylogeny (red circle; Venn diagram) by using the phylogenetic generalized least squares (PGLS) approach and, thus, appear to be evolving by natural selection.

glutathione peroxidase (*GPX1*), hypoxia-inducible factor (*HIFA), NADH-ubiquinone oxidoreductase (NUAM*), nucleoside diphosphate kinase (*NDKB*), and phosphomannomutase (*PMM2*) appear to be involved in temperature acclimation responses in carp (50), and glucose-6-phosphate 1-dehydrogenase (G6PT) and NUAM are associated with adaptive patterns of variation within and among *Fundulus* species (24). Whether these 13 loci have evolved independently or, in the extreme, are governed by a single variant at a transacting locus that only affects this subset of genes, is not addressed in this study.

Discussion

Using *Fundulus* microarray data, we describe an approach to distinguish between neutral and adaptive evolutionary processes affecting gene expression. Our approach is different from other comparative analyses o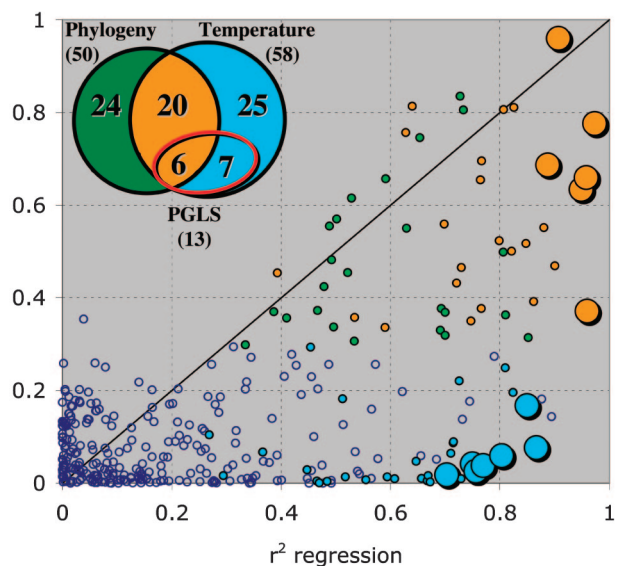f microarray data. It has recently been argued that higher variation among versus within taxa (equivalent to the *F* statistic) is indicative of natural selection (18). We extend this proposal by suggesting that variation among taxa accounted for by genetic distance is most parsimoniously explained by neutral drift, and only variation that exceeds this phylogenetic variance may be considered indicative of natural selection. Others have corrected the ratio of variance among versus within taxa for divergence time (33) or population size (16) with an arbitrary cutoff to identify the influence of natural selection. The analyses reported here differ from other attempts to define directional selection because the experimental design statistically tests for selection in an ecological context after accounting for the nonindependence of samples due to relatedness (34). Notice here that the detection of natural selection does not depend on a constant, an arbitrary cutoff value, or a rate. Instead, after accounting for the contribution of genetic distance for each gene separately, the residual variation must significantly regress with an ecological factor to reject of the neutral drift null hypothesis.

The identification of 22% of temperature-related genes as adaptive is conservative. Phylogenetic correction is not a technical improvement of comparative analyses *per se*, but rather a conceptual decision to prioritize random genetic drift over ecology as the correlate for trait variation (51). The comparative approach used here to identify these temperature-related genes assumes that ecological forces act only on the remaining residual variation after accounting for phylogenetic divergence and, thus, is highly conservative for identifying traits evolving by directional selection (51). Although half of the genes that correlate with habitat temperature also correlate with genetic distance, for the 13 genes where the variation accounted for by habitat temperature is far greater than can be accounted for by genetic distance alone, evolution by natural selection is the most parsimonious explanation.
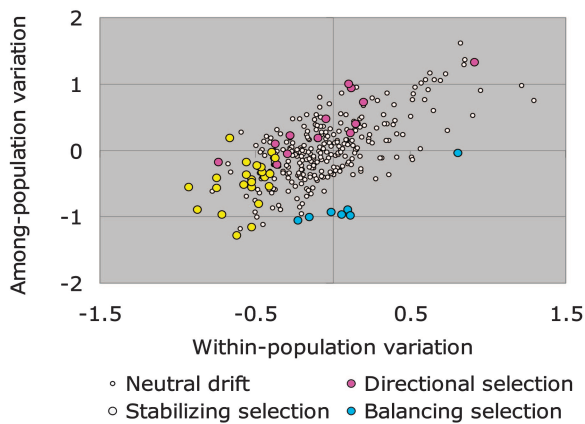
EVOLUTION

Fig. 4. Gene expression variation within and among populations indicating different patterns of evolutionary divergence. Plotted are the log of within- and among-population variation for gene expression. The ratios of these values are often used in statistical tests (e.g., ANOVA). Under neutral drift, within-population variation is correlated with among population variation (open circles). For other genes, different forms of selection overwhelm the general patterns indicated by drift and reject this null model. Genes under directional selection (pink circles) were identified as divergent along a habitat temperature gradient after correcting for variance due to phylogeny (phylogenetic generalized least squares method), and have higher variation among populations than within. Genes most influenced by stabilizing selection (yellow circles) have lower variation both within and among populations than most genes (*F* test), and genes under balancing selection (blue circles) have higher variation within than among populations (inverted *F* test).

Eighty-two genes (25%) have patterns of variation accounted for by phylogeny, habitat temperature, or both. In contrast, what (if anything) accounts for patterns of expression variation in the remaining 75% of genes? The adaptive pattern referred to above for 13 temperature genes suggests directional or divergent selection, yet as others have pointed out (16, 18, 33), stabilizing and balancing selection also can influence patterns of gene expression. The neutral theory suggests that with greater constraints (stabilizing selection; ref. 52), one should expect low variation among individuals and among populations (18). A formal method to identify genes most influenced by stabilizing selection would calculate variation among all individuals across all populations, then test which genes have significantly less variation compared with all other genes. We do this calculation by comparing the variation for a gene versus the variation for all other genes: an *F* ratio with variance in expression among all individuals for all genes as the numerator, and variance in expression among individuals for a single gene as the denominator (expression levels are standardized such that the mean expression for each gene is equal). An additional criterion would be no correlation between among-population differences in expression and genetic distance. We identify 24 genes that have disproportionately low variance in expression among individuals and populations (Bonferroni-corrected $P < 0.01$; Fig. 4). These genes with the least variation in expression are disproportionately represented by genes involved in oxidative phosphorylation (10 of the 24 significant genes with low variance, $P < 0.05$; Fisher's exact test), suggesting the biological importance of maintaining tight regulation for expression in this pathway. These central metabolism genes are medically important. For example, variation in oxidative phosphorylation genes contributes to inherited human diseases (53, 54).

In addition to directional selection, another pattern in conflict with the predictions of neutral evolution are genes for which within-population variation is higher than among-population variation (18). We propose testing for this pattern by calculating the inverse *F* statistic (the variation within populations versus variation among populations). Seven genes have significantly greater variation within a population than among the population (Fig. 4), indicative of balancing selection.

In microarray studies, genes with little biological variance in expression (suggestive of stabilizing selection) or genes with high variance within populations but little variance among populations (suggestive of balancing selection) are rarely considered. Two other attempts to recognize stabilizing selection examined the variation among taxa versus within (16, 33) while taking into account either divergence time or population size, and both indicated extensive stabilizing selection (up to 100% of all genes). However, stabilizing selection interacts with drift to influence the variation of traits along a continuum from high constraints minimizing the variation caused by drift (dominance of stabilizing selection) to a high variation allowable by fewer constraints (dominance of neutral drift). As such, binning traits as those influenced by drift or stabilizing selection is unnecessarily arbitrary. Rather, we propose the application of statistics that identify traits more or less affected by these evolutionary forces along the continuum. The application of evolutionary theory to identify genes with expression variation under strong constraints could be useful in medical genetics, and identifying genes under balancing selection could contribute to the debate over how variation is maintained in populations (55).

Evolutionary analyses provide a powerful approach for identifying genes with expression patterns that are of general interest to biologists or medical science. For the data presented here, among-population variation was positively correlated with within-population variation ($P < 0.001$; product-moment correlation), supporting the neutral prediction of gene expression evolution (32). Certainly, because of the costs of selection (summarized in the concept of genetic load; ref. 56), it seems unreasonable to expect that many, or even a significant minority of, genes can be subject to the effects of natural selection. Studies have indicated that most expression variation between humans and apes and mice is selectively neutral (31, 32), but these comparisons lacked an ecological context and, thus, could not test alternative hypotheses for the small subset of genes that may be affected by natural selection. In contrast, within the ecological context of ocean-depth gradients, natural selection has been important for shaping metabolic variation among species, genera, families, and phyla of marine organisms (57, 58), and phylogeny appears to account for little metabolic variation among these same groups. Notice also that detecting phylogenetic signals in comparative gene expression data may be best accomplished by comparing closely related taxa because the character space in which traits may vary is finite, and the influence of drift may become less obvious as divergence increases. For example, we detect expression levels that vary >2-fold among populations due to drift [e.g., asparagine synthetase (*ASNS*), phosphatidylcholine-sterol acyltransferase (*LCAT*), and succinyl-CoA ligase (*SUCA*)], but we would not expect variance to continuously increase with phylogenetic divergence because of eventual functional constraints. That is, ever-increasing differences in gene expression will not occur if for no other reason than because there is a limit to the amount of mRNA that can be produced. With increasing genetic distance, phenotypic divergence among taxa may become nonlinear (59).

Although phylogeny accounted for much of the variation among populations of *F. heteroclitus* distributed along a strong habitat gradient, variation for 13 of 58 genes that regress with habitat temperature exceeded that which could be accounted for by phylogeny alone and is most parsimoniously explained by directional selection. In total, our data suggest that natural selection is acting on the expression of 44 of the 329 genes (directional selection acting on 13 genes, stabilizing selection acting on 24 genes, and balancing selection acting on 7 genes). This conclusion required analysis within the appropriate ecological context among closely

Whitehead and Crawford

Table 1. Sources of variance for nested ANOVA and regression

Source of variance	df	Sum of squares	F ratio
Among populations: Pop	4	$80 \times \Sigma\,(PM - GM)^2$	MS_Pop/MS_Ind
Regression: Reg	1	$80 \times \Sigma((T - TM) \times (PM - GM))^2/(80 \times \Sigma\,(T - TM)^2)$	MS_Reg/MS_DevReg
Deviations from regression: DevReg	3	SS_Pop − SS_Reg	
Among individuals within population: Ind	20	$16 \times \Sigma(IM - DM)^2$	MS_Ind/MS_Dye
Among dyes within individual: Dye	25	$8 \times \Sigma\,(DM - RM)^2$	

Where populations = 5, individuals per population = 5, dyes = 2, replicate hybridizations per dye = 2, replicate spots per hybridization = 4. PM, mean expression for population; GM, grand mean expression; T, habitat temperature (dependent variable in regression); TM, mean of habitat temperatures; IM, mean expression for individual within population; DM, mean expression for dye within individual; RM, mean expression for replicate hybridization within dye.

related taxa. The effect of natural selection along an environmental gradient in this organism is exemplary of what one would expect to find in any organism, including humans. That is, similar to the influence of malaria on human hemoglobin (60) or G6PDH (61), defining the functional importance of variation is difficult to achieve without evolutionary analyses.

Methods

Animals and Maintenance. The teleost fish *F. heteroclitus* were collected from the field in June 2003 and acclimated in the laboratory to common controlled conditions (20°C, 15 parts per thousand salinity) in recirculating 100-gallon tanks for at least 2 months before experiments. The acclimation temperature is experienced by all populations from spring to fall and, thus, is ecologically relevant. Fish were killed by cervical dislocation, and livers were excised and stored in RNAlater (Ambion, Austin, TX) at −20°C. Fish were collected from the following five populations: Wiscasset, ME; Clinton, CT; Stone Harbor, NJ; Roanoke Island, NC; and Sapelo Island, GA. Only healthy adult male fish were used for the following experiments.

Habitat temperatures were derived from a minimum of 30 years mean annual surface temperatures. Coastal surface temperatures were considered more appropriate estimators of shallow estuarine habitat temperatures than temperatures logged by nearby open-ocean buoys.

Population Genetics. Individuals from Georgia ($n = 209$), North Carolina ($n = 54$), New Jersey ($n = 98$), Connecticut ($n = 50$), and Maine ($n = 50$) were genotyped by using five microsatellite markers. Markers B101 (trimer), B128 (trimer), B4 (trimer), B7 (trimer), and C1 (tetramer) (62) had, respectively, 19, 13, 24, 21, and 21 alleles per locus among the 461 individuals. Weir and Cockerham's theta (θ) estimator of F_{ST} was calculated for each population pair from genotype frequencies in GENETIX (www. univ-montp2.fr/~genetix/genetix/genetix.htm), and significance of θ was tested by using 10,000 random permutations. Neighbor-joining trees were constructed in NTSYS-PC (Exeter Software, Setauket, NY) by using Cavalli-Sforza and Edwards' chord distance (CSE) as the genetic distance metric (42). Isolation by distance was tested by using a Mantel test (1,000 permutations) for correlation between CSE and shoreline distance (gapped across major bays) matrices.

Microarrays. Microarrays were printed by using 329 cDNAs that encode essential proteins for cellular metabolism. All ESTs with enzyme commission numbers or associated with central metabolic pathways from a *F. heteroclitus* EST collection of >42,000 expressed sequences (http://genomics.rsmas.miami.edu/funnybase/ super_craw4) were included on the array. These cDNAs were amplified with amine-linked primers and printed on 3D-Link activated slides (Surmodics, Eden Prairie, MN) at the University of Miami core microarray facility. The suite of 329 amplified cDNAs was printed as a group in four spatially separated replicates.

Microarray analyses were applied to livers from five individ-

uals collected from each of the five populations of *F. heteroclitus*. RNA was extracted from tissue homogenate in a chaotropic buffer by using phenol/chloroform/isoamyl alcohol, and purified RNA was amplified to make amplified RNA (aRNA) by using a modified Eberwine protocol as described in ref. 26. Each labeled aRNA sample was suspended in 1.5 pmol/μl hybridization buffer, applied to the slide, and incubated 12–18 h at 42°C. Each of the 25 samples was hybridized twice, once with Cy3 and once with Cy5. Because a hybridization zone covered four replicate printed arrays, total experimental replication per sample per gene was 8-fold. A total of 50 hybridizations (25 × 2) were balanced among replicate individuals and populations in a loop design. Slides were scanned by using the Packard Bioscience ScanArray Express microarray scanner (PerkinElmer), and spot identification and intensity quantifications were collected by using IMAGENE software (Biodiscovery, Marina Del Ray, CA).

Statistical Analysis. Raw microarray data were sum normalized (63), intensity bias on each array was smoothed by using a Lowess transformation in R/MAANOVA 0.93–2 (www.jax.org/staff/ churchill/labsite), and $\log_2$ values of Lowess-transformed sum-normalized data were used for all subsequent analyses. MIAME (minimum information about a microarray experiment) compliant data (64) are available upon request. Technical variance and variance among individuals and populations were quantified in a nested ANOVA framework (Table 1) by using scripts written in MATLAB, (Version 6; MathWorks, Natick, MA). Scripts are available upon request.

For each gene, we tested whether expression significantly correlated with median annual habitat temperature (least squares regression) or with genetic distance (Mantel test, 1,000 permutations), and we quantified the proportion of among-population expression variation accounted for by habitat temperature and genetic distance (r^2). A second measure of correlation between genetic distance and gene expression was included to corroborate Mantel results; the PHYSIG program (47, 48) used branch lengths of the microsatellite-derived neighbor-joining tree (computed in NTSYS-PC; Exeter Software) and 10,000 random permutations of the tree structure to test whether a phylogenetic signal was present in among-population gene expression patterns for each gene by using MATLAB scripts provided by Theodore Garland, Jr. (University of California, Riverside).

Regression of gene expression against habitat temperature was corrected for nonindependence of observations due to phylogeny by applying the phylogenetic generalized least squares method (48, 49) by using MATLAB scripts provided by Theodore Garland, Jr. (available upon request) and the MULREG module in NTSYS-PC. A neighbor-joining tree was created in NTSYS-PC based on microsatellite-derived genetic distances (Cavalli-Sforza and Edwards' chord distance; ref. 42), and the tree structure and branch lengths were used to produce a matrix of expected variances and covariances of traits between taxa based on a Brownian motion model of

EVOLUTION

evolution by using the COPH routine (PhyloCov method) in NTSYS-PC. The resulting covariance matrix was used to correct regression of gene expression against habitat temperature for expected lack of independence due to phylogeny.

1. King, M. C. & Wilson, A. C. (1975) *Science* **188,** 107–116.
2. Crawford, D. L. & Powers, D. A. (1992) *Mol. Biol. Evol.* **9,** 806–813.
3. Crawford, D. L., Segal, J. A. & Barnett, J. L. (1999) *Mol. Biol. Evol.* **16,** 194–207.
4. Hahn, M. W., Rockman, M. V., Soranzo, N., Goldstein, D. B. & Wray, G. A. (2004) *Genetics* **167,** 867–877.
5. Lerman, D. N., Michalak, P., Helin, A. B., Bettencourt, B. R. & Feder, M. E. (2003) *Mol. Biol. Evol.* **20,** 135–144.
6. Rockman, M. V., Hahn, M. W., Soranzo, N., Goldstein, D. B. & Wray, G. A. (2003) *Curr. Biol.* **13,** 2118–2123.
7. Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V. & Romano, L. A. (2003) *Mol. Biol. Evol.* **20,** 1377–1419.
8. Gibson, G. & Weir, B. (2005) *Trends Genet.* 21, 616–623.
9. Stamatoyannopoulos, J. A. (2004) *Genomics* **84,** 449–457.
10. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. (2002) *Science* **296,** 752–755.
11. Cavalieri, D., Townsend, J. P. & Hartl, D. L. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 12369–12374.
12. Ferea, T. L., Botstein, D., Brown, P. O. & Rosenzweig, R. F. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9721–9726.
13. Townsend, J. P., Cavalieri, D. & Hartl, D. L. (2003) *Mol. Biol. Evol.* **20,** 955–963.
14. Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G. & Gibson, G. (2001) *Nat. Genet.* **29,** 389–395.
15. Gibson, G., Riley-Berger, R., Harshman, L., Kopp, A., Vacha, S., Nuzhdin, S. & Wayne, M. (2004) *Genetics* **167,** 1791–1799.
16. Rifkin, S. A., Kim, J. & White, K. P. (2003) *Nat. Genet.* **33,** 138–144.
17. Wayne, M. L., Pan, Y. J., Nuzhdin, S. V. & McIntyre, L. M. (2004) *Genetics* **168,** 1413–1420.
18. Nuzhdin, S. V., Wayne, M. L., Harmon, K. L. & McIntyre, L. M. (2004) *Mol. Biol. Evol.* **21,** 1308–1317.
19. Pritchard, C. C., Hsu, L., Delrow, J. & Nelson, P. S. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 13266–13271.
20. Cheung, V. G., Conlin, L. K., Weber, T. M., Arcaro, M., Jen, K. Y., Morley, M. & Spielman, R. S. (2003) *Nat. Genet.* **33,** 422–425.
21. Sharma, A., Sharma, V. K., Horn-Saban, S., Lancet, D., Ramachandran, S. & Brahmachari, S. K. (2005) *Physiol. Genomics* **21,** 117–123.
22. Tan, Q., Christensen, K., Christiansen, L., Frederiksen, H., Bathum, L., Dahlgaard, J. & Kruse, T. A. (2005) *Hum. Genet.* **117,** 267–274.
23. Enard, W., Khaitovich, P., Klose, J., Zoellner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., *et al.* (2002) *Science* **296,** 340–343.
24. Oleksiak, M. F., Churchill, G. A. & Crawford, D. L. (2002) *Nat. Genet.* **32,** 261–266.
25. Oleksiak, M. F., Roach, J. L. & Crawford, D. L. (2005) *Nat. Genet.* **37,** 67–72.
26. Whitehead, A. & Crawford, D. (2005) *Genome Biol.* **6,** R13.
27. Kreitman, M. (1996) *BioEssays* **18,** 678–683, discussion 683.
28. Takahata, N. (1996) *Curr. Opin. Genet. Dev.* **6,** 767–772.
29. Orr, H. A. (1998) *Genetics* **149,** 2099–2104.
30. Endler, J. A. (1986) *Natural Selection in the Wild* (Princeton Univ. Press, Princeton).
31. Yanai, I., Graur, D. & Ophir, R. (2004) *OMICS* **8,** 15–24.
32. Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W. & Paabo, S. (2004) *PLoS. Biol.* **2,** E132.
33. Lemos, B., Meiklejohn, C. D., Caceres, M. & Hartl, D. L. (2005) *Evolution* **59,** 126–137.
34. Felsenstein, J. (1985) *Amer. Nat.* **125,** 1–15.
35. Garland, T., Jr., Harvey, P. H. & Ives, A. R. (1992) *Syst. Biol.* **41,** 18–32.
36. Harvey, P. H. & Pagel, M. D. (1991) *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, New York).
37. Powers, D. A., Smith, M., Gonzalez-Villasenor, I., DiMichelle, L., Crawford, D. L., Bernardi, G. & Lauerman, T. (1993) in *Oxford Surveys in Evolutionary Biology*, eds. Futuyma, D. & Antonovics, J. (Oxford Univ. Press, New York), Vol. 9, pp. 43–108.
38. Crawford, D. L. & Powers, D. A. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9365–9369.
39. Pierce, V. A. & Crawford, D. L. (1997) *Science* **276,** 256–259.
40. Hochachka, P. W. & Somero, G. N. (1984) *Biochemical Adaptation* (Princeton Univ. Press, Princeton).
41. Prosser, C. L. (1986) *Adaptational Biology: From Molecules to Organisms* (Wiley, New York).
42. Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967) *Evolution* **32,** 550–570.
43. Cashon, R. E., Van Beneden, R. J. & Powers, D. A. (1981) *Biochem. Genet.* **19,** 715–728.
44. Able, K. W. & Fewlley, J. D. (1986) *Am. Zool.* **25,** 145–157.
45. Bernardi, G., Sordino, P. & Powers, D. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9271–9274.
46. Storey, J. D. & Tibshirani, R. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 9440–9445.
47. Garland, T., Jr., & Ives, A. R. (2000) *Am. Nat.* **155,** 346–364.
48. Blomberg, S. P., Garland, T., Jr., & Ives, A. R. (2003) *Evolution* **57,** 717–745.
49. Rohlf, F. J. (2001) *Evolution* **55,** 2143–2160.
50. Gracey, A. Y., Fraser, E. J., Li, W., Fang, Y., Taylor, R. R., Rogers, J., Brass, A. & Cossins, A. R. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 16970–16975.
51. Westoby, M., Leishman, M. R. & Lord, J. M. (1995) *J. Ecol.* **83,** 531–534.
52. Kimura, M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5969–5973.
53. Shoubridge, E. A. (2001) *Hum. Mol. Genet.* **10,** 2227–2284.
54. Taylor, R. W. & Turnbull, D. M. (2005) *Nat. Rev.* **6,** 389–406.
55. Mitton, J. B. (1997) *Selection in Natural Populations* (Oxford Univ. Press., New York).
56. Lewontin, R. C. (1997) in *Evolution*, ed. Ridley, M. (Oxford Univ. Press, Oxford), pp. 79–88.
57. Childress, J. J. (1995) *Trends in Ecology & Evolution* **10,** 30–36.
58. Seibel, B. A., Thuesen, E. V. & Childress, J. J. (2000) *Biol. Bull. (Woods Hole, MA)* **198,** 284–298.
59. Whitehead, A. & Crawford, D. (2006) *Mol. Ecol.*, in press.
60. Dean, A. M. (1998) *Am. Sci.* **86,** 26–37.
61. Tishkoff, S. A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., *et al.* (2001) *Science* **293,** 455–462.
62. Adams, S. M., Oleksiak, M. F. & Duvernell, D. D. (2005) *Mol. Ecol. Notes* **5,** 275–277.
63. Quackenbush, J. (2002) *Nat. Genet.* **32,** 496–501.
64. Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., *et al.* (2001) *Nat. Genet.* **29,** 365–371.
65. Adams, S. M., Lindmeier, J. B. & Duvernell, D. D. (2006) *Mol. Ecol.* **15,** 1109–1124.