

The effect of genetic and environmental variation on metabolic gene expression

CINDA P. SCOTT,* DEAN A. WILLIAMS† and DOUGLAS L. CRAWFORD*

*Department of Marine Biology and Fisheries, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA, †Department of Biology, Texas Christian University, Winton Scott Room 401, 2800 South University Drive, Fort Worth, TX 76129, USA

Abstract

What is the relationship between genetic or environmental variation and the variation in messenger RNA (mRNA) expression? To address this, microarrays were used to examine the effect of genetic and environmental variation on cardiac mRNA expression for metabolic genes in three groups of *Fundulus heteroclitus*: (i) individuals sampled in the field (field), (ii) field individuals acclimated for 6 months to laboratory conditions (acclimated), or (iii) individuals bred for 10 successive generations in a laboratory environment (G10). The G10 individuals have significantly less genetic variation than individuals obtained in the field and had a significantly lower variation in mRNA expression across all genes in comparison to the other two groups ($P = 0.001$). When examining the gene specific variation, 22 genes had variation in expression that was significantly different among groups with lower variation in G10 individuals than in acclimated individuals. Additionally, there were fewer genes with significant differences in expression among G10 individuals vs. either acclimated or field individuals: 66 genes have statistically different levels of expression vs. 107 or 97 for acclimated or field groups. Based on the permutation of the data, these differences in the number of genes with significant differences among individuals within a group are unlikely to occur by chance ($P < 0.01$). Surprisingly, variation in mRNA expression in field individuals is lower than in acclimated individuals. Relative to the variation among individual within a group, few genes have significant differences in expression among groups (seven, 2.3%) and none of these are different between acclimated and field individuals. The results support the concept that genetic variation affects variation in mRNA expression and also suggests that temporal environmental variation associated with estuarine environments does not increase the variation among individuals or add to the differences among groups.

Keywords: *Fundulus heteroclitus*, gene expression evolutionary genomics, genetic variation, microarray

Received 1 May 2008; revision received 2 April 2009; accepted 5 April 2009

Introduction

Variation in messenger RNA (mRNA) expression should be a function of genetic and environmental variation. The quantification of variance due to the additive effects of genes is important as natural selection acts on this genetic component of variance (Falconer & Mackay 1996). In out-bred populations of the teleost fish *Fundulus heteroclitus*, there is substantial variation in mRNA expression within and between groups of individuals (Oleksiak *et al.* 2002;

Oleksiak *et al.* 2005; Whitehead & Crawford 2006b; Crawford & Oleksiak 2007). Although much of this variation appears to be due to random neutral evolutionary processes, a significant fraction of variation in expression is best explained by evolution by natural selection (Whitehead & Crawford 2006a). These evolutionarily adaptive patterns suggest that the variation in mRNA expression is biologically important because any trait evolving by natural selection must affect fitness (Whitehead & Crawford 2006a). Investigation into the functional importance of natural variation in mRNA expression revealed that variation in mRNA expression explains differential use of metabolic substrates among groups of individuals providing additional evidence

Correspondence: Douglas L. Crawford, Fax: (305) 421 4600, E-mail: dcrawford@rsmas.miami.edu

for the biological significance for otherwise seemingly chaotic patterns of expression (Oleksiak *et al.* 2005). While these studies assume that variation in mRNA expression among individuals is genetically based due to the rearing of fish in a common environment, no studies have been performed to test this hypothesis.

To provide evidence supporting this hypothesis, we examine the variation in metabolic mRNA expression among groups that have different levels of genetic and environmental variation. If mRNA expression is heritable and genotypic effects dominate the variation in expression, then more genetically similar individuals should have less variation and fewer significant differences in mRNA expression than among unrelated individuals. Notice, we are not asking whether mRNA expression differs in different environments, or between genotypes. Instead, we are asking whether the variation in mRNA expression is a function of environmental and genetic variation. We demonstrate that variation in mRNA expression is significantly lower among closely related individuals compared to outbred fish raised in similar environments. Surprisingly, increased environmental heterogeneity in unacclimated individuals sampled in the field did not increase the variation in mRNA expression among the outbred samples. These observations suggest that the normal environmental variation associated with tidal fluxes in estuarine environments does not substantially add to the variation in mRNA expression in *F. heteroclitus*.

Methods and materials

Organism

Fundulus heteroclitus were caught from wild populations in Beaufort, North Carolina, USA (34°43'34"N, 76°40'62"S) by allowing fish to swim into seine nets on an outgoing tide. Seine nets were used to catch a few fish (10–20) to minimize handling time and stress. Upon capture in the field, five males and five females were placed in a bucket, weighed, sexed, killed by cervical dislocation, their hearts removed and stored in RNA later (Ambion, Inc.). From the same location and tide, more fish were captured and transported to the University of Miami where they were acclimated to 20 °C and 15 ppt in laboratory aquaria for approximately 6 months. These fish were compared to fish trapped at the same location, and raised at 20 °C and 15 ppt and allowed to interbreed for 10 successive generations (G10) at the Aquatic Biotechnology and Environmental Laboratory at the University of Georgia. For the purposes of this experiment, five males and five females from each of the following groups were used: field caught (field), field caught then acclimated for 6 months at the University of Miami (acclimated) and fish raised for 10 generations (G10).

Genetic diversity

The G10 fish were started from a group of approximately 16 adults and were allowed to freely interbreed for 10 generations. In order to characterize levels of genetic diversity and pairwise relatedness within and between the G10 and field-caught (field) individuals, we genotyped 49 G10 and 109 field individuals (including individuals used in the microarray experiments) at 10 microsatellite loci for *F. heteroclitus* (Adams *et al.* 2005).

DNA was extracted from dried fin clips. The tissue was placed in 300 µL lysis buffer (75 mM NaCl, 25 mM EDTA, 1% SDS) and incubated with 0.1 mg proteinase K at 55 °C for 2 h. Proteins were precipitated by adding a half volume of 7.5 M ammonium acetate and centrifugation for 10 min. DNA was precipitated from the supernatant by adding 0.7 volumes of isopropanol and centrifugation for 15 min. The DNA pellet was washed with 70% ethanol then allowed to air dry for 30 min followed by re-suspension in 50 µL 10 mM Tris-HCl pH 8.5.

Microsatellite loci were amplified in four fluorescently labelled multiplex primer groups containing the following final concentrations: A–(0.15 µM CA-1, 0.07 µM CA-A3, 0.20 µM C-1), B–(0.10 µM ATG-18, 0.10 µM ATG-B4), C–(0.07 µM ATG-25, 0.07 µM ATG-6), D–(0.07 µM ATG-B128, 0.15 µM CA-21). Locus ATG-20 was amplified alone at a final concentration of 0.5 µM. The 10-µL reaction contained 2.5 mM MgCl₂, 1× polymerase chain reaction (PCR) buffer, 0.2 mM dNTPs, 0.4 U *Taq* DNA polymerase, 70 ng DNA, and one of the five primer combinations (see above for concentrations). The PCR thermal cycling profile consisted of 94 °C for 2 min, followed by 31 cycles of 94 °C for 15 s, 55 °C (A, C, ATG-20) or 55 °C (B and D) for 15 s, and 72 °C for 30 s, ending with a 5-min extension step at 72 °C. Following PCR amplification, the products from A, C, and D were coloaded, as were ATG-20 and B, before being electrophoresed on an ABI 3730XL Genetic Analyser (Applied Biosystems). GeneMapper version 4.0 (Applied Biosystems) was used to score the genotypes. All genotypes were checked by two individuals.

RNA isolation, labelling and hybridization

Total RNA was isolated from using 4.5 M guanidinium thiocyanate, 2% N-lauroylsarcosine, 50 mM EDTA, 25 mM Tris-HCl, 0.1 M β-Mercaptoethanol and 0.2% Antifoam A. The extracted RNA was further purified using a QIAGEN RNeasy Mini kit in accordance with the manufacturer's protocols. The quantity and quality of the RNA was determined using a spectrophotometer (Nanodrop, ND-1000 version 3.2.1) and by electrophoresis with the use of a bioanalyser (Agilent 2100). RNA was then converted into amino allyl-labelled RNA (aRNA) using the Ambion Amino Allyl MessageAmp II aRNA Amplification kit. This

method converts poly A RNA into complementary (cDNA) with a T7 RNA polymerase binding site and T7 is used to synthesize many new strands of RNA (*in vitro* transcription) (Eberwine 1996). During this *in vitro* transcription of aRNA, an amino allyl UTP (aaUTP) is incorporated into the elongating strand. The incorporations of aaUTPs allows for the coupling of Cy3 or Cy5 dyes (GE biosciences) onto aRNA for microarray hybridization.

Dye-labelled aRNA aliquots for each hybridization (30 pmol each of Cy3 and Cy5) were vacuum dried together and resuspended in 15 µL hybridization buffer (final concentration of each labelled sample = 2 pmol/µL). Hybridization buffer consisted of 5× SSPE, 1% SDS, 50% formamide, 1 mg/mL polyA, 1 mg/mL sheared herring sperm carrier DNA, and 1 mg/mL BSA. Slides were washed in sodium borohydride solution in order to reduce autofluorescence. Following rinsing, slides were boiled for 2 min and spin-dried in a centrifuge at 800 r.p.m. for 3 min. Samples (15 µL) were heated to 90 °C for 2 min, quick cooled to 42 °C, applied to the slide (hybridization zone area was 350 mm²), and covered with a cover slip. Slides were placed in an airtight chamber humidified with paper soaked in 5× SSPE and incubated 24–48 h at 42 °C.

Microarrays

The amount of gene-specific mRNA expression was measured using microarrays with four spatially separated replicates per gene on each array. Microarrays were printed using 384 *F. heteroclitus* cDNAs that included 329 cDNAs that encode essential proteins for cellular metabolism (Table 1; Paschall *et al.* 2004). Average lengths of cDNAs were 1.5 Kb with a majority including the N-terminal methionine. Table 1 provides a summary of expressed sequence tags (EST) used for printing where the most meaningful GO term is used to categorized the annotation (Paschall *et al.* 2004). These cDNAs were amplified with amine-linked primers and printed on 3-D Link Activated slides (Surmodics Inc., Eden Prairie, Minnesota) at the University of Miami core microarray facility.

Dye-coupled aRNA from 30 individuals from three groups were hybridized to slides using two loops (Kerr & Churchill 2001; Oleksiak *et al.* 2002); one for males and one for females (Fig. S1, Supporting information). The female loop is as follows; G1 → A1 → F1 → G2 → A2 → F3 → G7 → A5 → F6 → G4 → A10 → F15 → G5 → A12 → F18 → G1. The male loop is as follows; G12 → A6 → F2 → G14 → A7 → F4 → G15 → A8 → F5 → G6 → A9 → F7 → G10 → A11 → F8 → G12. Each arrow represents a single hybridization between an individual labelled with Cy3 at the base of the arrow and an individual labelled with Cy5 at the head of the arrow. The first letter represents either a G10 (G), acclimated (A) or field caught (F) individual. The number represents each

Table 1 Three hundred and eighty-four microarray pathways

Pathway	Number of cDNAs
Amino acid metabolism	28
ATP synthesis	27
Blood group glycolipid biosynthesis	3
Channel	3
Citrate cycle (TCA cycle)	24
Fatty acid metabolism/transport	36
Fructose and mannose metabolism	4
Galactose metabolism	2
Glutamate metabolism	7
Glutathione metabolism	10
Glycerolipid metabolism	7
Glycolysis/Gluconeogenesis	27
Inositol phosphate metabolism	14
Ox-Phos-ATPsyn	64
Pentose phosphate pathway	6
Purine and pyrimidine metabolism	9
Pyruvate metabolism	2
Signalling pathway	10
Starch and sucrose metabolism	2
Sterol biosynthesis	8
Synthesis and degrad. of ketone bodies	4
Tetrachloroethene degradation	3
Secondary	27
Total metabolic genes	329

individual used in the study. Three males (A6, F12 and G10) were removed from the analysis due to poor hybridization or too strong of signals (i.e. where too many different genes had signals that saturated the PMT = 65 535). Therefore, a total of 12 males (four per group) and 15 females (five per group) were used for the analysis.

The microarray slides were scanned using ScanArray Express. The raw TIFF-image data were quantified using Imagene (version 5). If a gene had a fluorescent signal that was too low or too high, it was eliminated from the analysis for all individuals. Fluorescent signals were considered too low if the average across all samples were within two standard deviations of the average signal from the *Ctenophore* negative controls. Fluorescent signals were considered too high if the average signal plus two standard deviations exceeded 55 000. This procedure is based on empirical analyses of data and removes fluorescent signals that saturate the photomultiplier tube (maximum signal is 65 565). Using these criteria, 100 genes were eliminated from all individuals leaving 284 genes.

Statistics

Microsatellite loci were tested for deviation from Hardy–Weinberg equilibrium and for linkage disequilibrium using GenePop version 3.3 (Raymond & Rousset 1995). The number of alleles (N_A), observed heterozygosity (H_O), and expected heterozygosity (H_E) were calculated using

Genalex 6 (Peakall & Smouse 2006). Allelic richness (A_R) for each group (G10 and field) was calculated with FSTAT version 2.9.3 (Goudet 1995) with a sample size adjustment of $n = 49$ individuals (the smallest sample size). We compared average measures of genetic diversity calculated across loci between G10 and field individuals by randomizing locus-specific values between groups and recalculating the difference in mean values 5000 times to generate a random distribution of mean values. The location of the observed mean difference within this random distribution was used to determine the probability that it was significantly different from the random distribution.

Genetic similarity between individuals within groups was estimated by the relatedness coefficient R of Queller and Goodnight using Relatedness 5.0 (Queller & Goodnight 1989). The allele frequencies used to calculate relatedness coefficients came from the entire sample of G10 and field individuals. Standard errors of the estimates were obtained by jackknifing over loci (Sokal & Rohlf 1995). We compared the average relatedness of G10 and field individuals by jackknifing over the unpaired R difference using Relatedness. We also estimated genetic similarity by the proportion of shared alleles (Bowcock *et al.* 1994). Significance was determined by permuting individuals between the groups and recalculating the mean proportion of alleles shared between individuals 1000 times to construct the 95% CI around the random expectation. Ninety-five per cent confidence intervals were also calculated around each mean by bootstrapping values within each group 1000 times.

Statistical analyses of the mRNA expression data were carried out using JMP genomics (SAS JMP Genomics version 7.0.2). All analyses used fluorescent measures that were \log_2 transformed and loess normalized. These normalized fluorescent measures showed nearly identical distributions among all individuals (Fig. S2, Supporting information). Standardization of data (mean signal with an average intensity equal to zero) or further normalization using ANOVA or a mixed model did not substantially affect the distribution of fluorescences nor did it affect the relative frequency of genes with significant differences in expression among individuals. Thus, the simpler of the normalizations (\log_2 -loess) were used for parsimony and clarity of results.

For each gene, an ANOVA for the significant differences in mRNA expression among individuals within each group used the linear mixed model (Kerr & Churchill 2001; Wolfinger *et al.* 2001; Yu *et al.* 2004; Patterson *et al.* 2006): $y_{ijk} = \mu + A_i + D_j + I_k + \varepsilon_{ijk}$, where y_{ijk} represents the fluorescence intensities on a log scale and μ is a constant. The fixed effect is I_k for k^{th} individual (for one of the nine individuals per group) and the random effects are D_j for the j^{th} dye of the two Cy dyes, A_i is the i^{th} array (for one of the 27 different arrays) and ε_{ijk} are random residual term. This analysis, to define difference among individuals, was applied to the nine individuals in each group. With nine individuals per

group and eight replicates (four replicates per array and two dyes), there are 8 and 52 degrees of freedom. To determine if the number of genes with significantly different expression among individuals in each group was statistically meaningful, this mixed model analysis was run on all 126 possible combinations of five out of nine individuals for each group (4 and 30 degrees of freedom). From these 126 ANOVAs, the average number of genes and the confidence intervals that were significantly different among individuals was calculated. In these and other statistical analyses of mRNA expression, a P value of 0.01 is used (with and without Bonferonni's correction). This arbitrary value is chosen because it provides more a conservative value than a P value of 5% and it allows the reader to quickly judge the 1% frequency of type 1 errors.

zStatistical analysis of group or sex effects used the least square means from the linear mixed model. This model, which provides a single measure of expression for each gene for each individual, is identical to the model described above except the linear model used all 27 individuals across all three groups (instead of the nine in each group; resulting in 26 and 161 degrees of freedom).

Using the least square means for each individual, an ANOVA was used to determine the significant differences in mRNA expression between groups or sex. Notice that with the least square means, there are no dye nor array effects (only one measure for each individual and each gene from the mixed model with dyes and arrays as random factors), thus, these factors are not included. For the ANOVA with sex as a fixed effect, there were 1 and 25 degrees of freedom. For the ANOVA with group as fixed effect there were 2 and 24 degrees of freedom.

Hierarchical clustering of gene expression uses Macintosh's version (104) of Eisen's Cluster and Treeview (105). Volcano plots are from SAS-JMP and use a t -test for comparisons between any two individuals or groups.

Results

The genetic variation and expression of mRNA was measured in three groups: *Fundulus* caught in the field and immediately sacrificed (field), *Fundulus* caught in the field and acclimated for 6 months to common laboratory conditions (acclimated) and *Fundulus* bred for 10 successive generations (G10) in a common laboratory environment. All fish originated from the same location in Beaufort, North Carolina. Most of the mRNAs quantified encode metabolic genes (Table 1).

Genetic diversity and relatedness

The microsatellite loci in the field and acclimated samples (outbred groups) were highly polymorphic and in Hardy-Weinberg ($P = 0.70$) and genotypic linkage equilibrium

Table 2 Genetic diversity values for laboratory bred (G10) individuals ($n = 49$) and field-caught individuals (F) ($n = 109$)

Locus	A_R G10	A_R F	H_O G10	H_O F	H_E G10	H_E F	F_{IS} G10	F_{IS} F
ATG-18	3.98	6.73	0.24	0.55	0.24	0.53	-0.02	-0.03
ATG-20	3.00	8.89	0.63	0.73	0.62	0.72	-0.02	-0.03
ATG-25	5.00	9.95	0.76	0.83	0.74	0.83	-0.02	0.00
ATG-6	2.98	5.29	0.55	0.64	0.44	0.64	-0.25	0.00
ATG-B4	5.00	19.79	0.57	0.88	0.53	0.91	-0.08	0.03
ATG-B128	3.00	7.79	0.59	0.69	0.51	0.75	-0.17	0.09
ATG-C1	4.96	11.38	0.47	0.83	0.57	0.78	0.17	-0.06
CA-1	6.96	12.91	0.90	0.78	0.78	0.77	-0.15	-0.01
CA-21	4.00	22.18	0.46	0.93	0.48	0.93	0.04	0.01
CA-A3	6.00	30.36	0.76	0.97	0.72	0.96	-0.05	-0.01
Average \pm SE	4.49 \pm 0.43	13.53 \pm 2.55	0.59 \pm 0.06	0.78 \pm 0.04	0.56 \pm 0.05	0.78 \pm 0.04	-0.06 \pm 0.04	0.00 \pm 0.01

A_R , allelic richness corrected for a sample size of 49 individuals; H_O , observed heterozygosity; H_E , expected heterozygosity; and F_{IS} , the inbreeding coefficient.

($P > 0.20$ in all cases). The G10 individuals had significantly lower mean genetic diversity values than the field caught individuals for three of our four measures (A_R $P = 0.001$; H_O $P = 0.001$; H_E $P = 0.001$; F_{IS} $P = 0.11$) (Table 2). The G10 fish were also in Hardy–Weinberg equilibrium ($P = 0.14$), but 13 of the 45 pairwise comparisons between loci had significant linkage disequilibrium after a Bonferroni’s correction for multiple comparisons. Relatedness among individuals was significantly higher than zero and close to the level of full siblings for the G10 individuals ($R = 0.43$; 95% CI 0.33–0.53). This contrasts with the relatedness of field individual that were not significantly different from zero ($R = 0.03$; 95% CI -0.004–0.066). The proportion of shared alleles between individuals revealed a similar pattern; individuals within the G10 group shared more alleles (mean = 0.545; 95% CI 0.536–0.553) than individuals within the field group (mean = 0.315; 95% CI 0.312–0.317).

We also compared the proportion of shared alleles between the individuals that were used only in the microarray experiment (Fig. 1). The G10 individuals used in the microarray experiment shared a significantly higher proportion of their alleles (0.46, 95% CI 0.42–0.51) than either the acclimated (0.36, 95% CI 0.34–0.38) or the field (0.30, 95% CI 0.28–0.33). Although the acclimated group had a slightly higher level of shared alleles than the field group, both fell within the 95% CI expected for random pairings of individuals (Fig. 1).

Significant difference between sexes in mRNA expression

Gene expression was measured in a total of 12 males and 15 females using two separate hybridization loops. To test for difference between the sexes or the confounding affect of loops, we applied an ANOVA using the least square means from the linear model with individual as fixed effect and array and dyes as random affect. Only three genes have

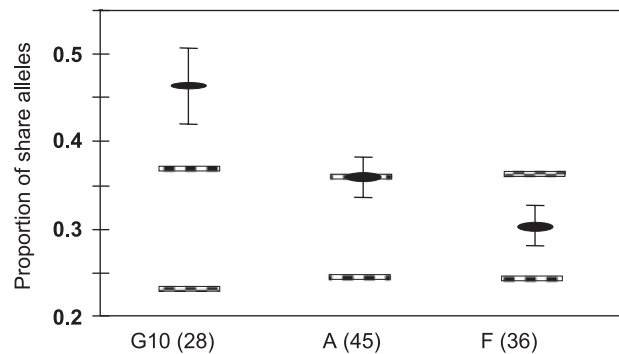


Fig. 1 Average proportion of shared alleles within groups. Proportion of shared alleles is shown for inbred (G10), outbred acclimated (acclimated), and outbred field (field) groups. Numbers in parentheses are the number of pairwise comparisons in each group. Vertical lines with ellipse are 95% bootstrapped confidence intervals around each calculated mean value (ellipse). Striped horizontal lines are the 95% confidence intervals around the random expectation calculated by permuting individuals between groups. G10 differ significantly at $P = 0.001$.

significant differences ($P < 0.01$) in mRNA expression between the two sexes. None are significant with Bonferroni’s correction for multiple comparisons ($P < 3.5 \times 10^{-5}$). With so few differences in gene expression between the sexes, we analysed males and females together.

Significant differences within groups in mRNA expression

Among all individuals (ignoring group effects) 281 of the 284 genes have significant differences in expression (Fig. 2a, $P < 0.01$). Within each group the number genes with significant differences in mRNA expression is a measure of variation and thus, there is an expectation that the

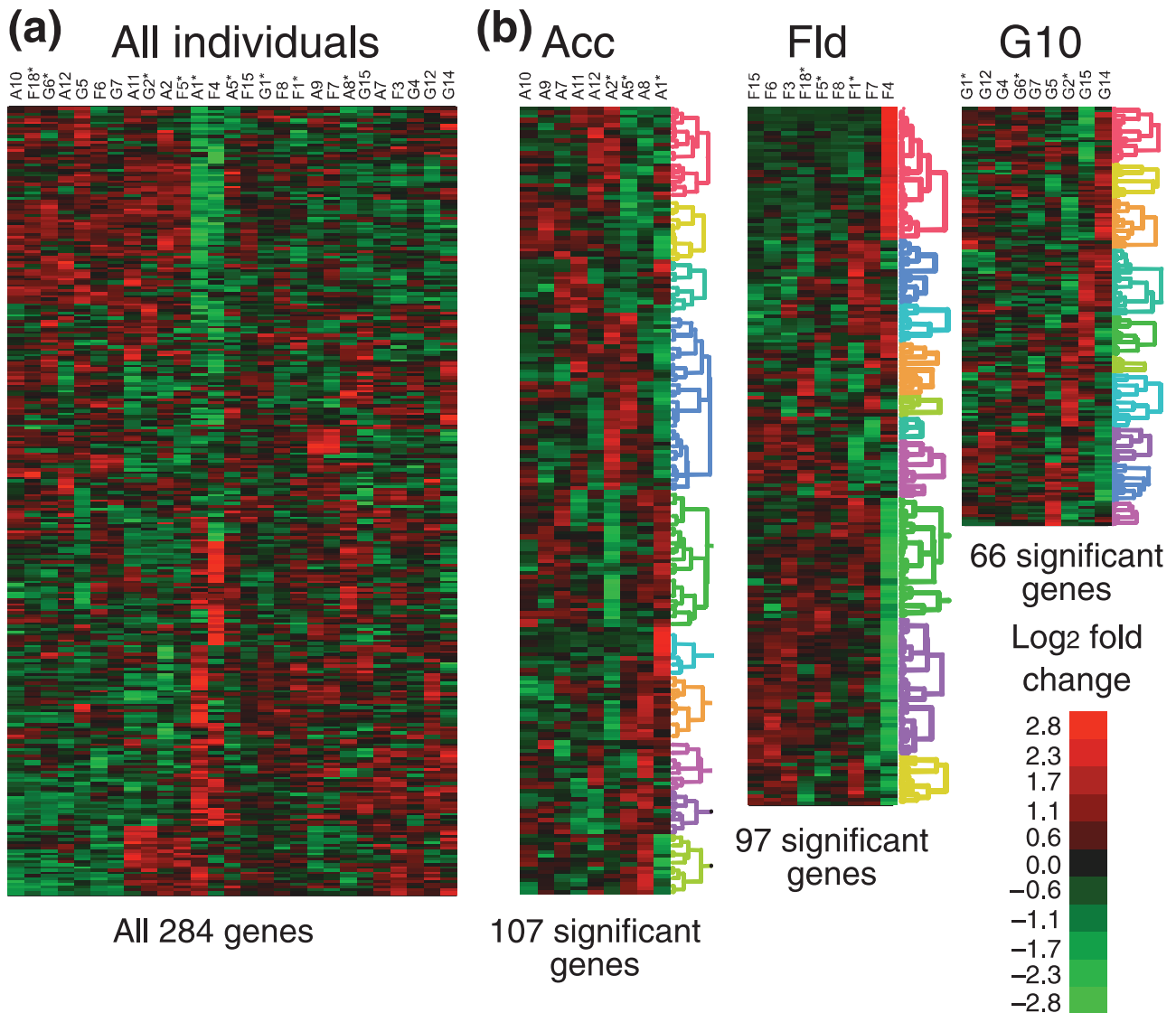


Fig. 2 Hierarchical clustering of individual gene expression. Clustering is based on the correlations of least square means of mRNA expression. Colour panel in lower right corner illustrates relationship between fold change and colour. (a) Relative expression among all individuals ignoring groups for all 284 measure of gene expression (Table 1). (b) Relative expression for genes with significantly different expression among individuals within each group: Acc, acclimated; Fld, field; and G10, 10th generation inbred. Individual used in volcano plots (Fig. 3) are indicated by '*'. Colour dendrograms indicate genes with similar patterns of expression.

number of significant genes will be related to genetic and environmental variation (Fig. 2, Table S1, Supporting information). A mixed-model ANOVA was used to test if there are significant differences in mRNA expression among individuals within each group ($P < 0.01$, Table 3). G10 individuals had approximately two-thirds the number of significantly different genes (66 or 23%, 23 significant with Bonferroni's corrected $P < 3.5 \times 10^{-5}$) as did acclimated individuals (107 or 38%, 46 with Bonferroni's corrected $P < 3.5 \times 10^{-5}$) even though both groups were raised in similar laboratory conditions. Surprisingly, field individuals had an

intermediate number of significant genes (97 or 34%, 29 with Bonferroni's corrected $P < 3.5 \times 10^{-5}$).

The patterns of variation among individuals in mRNA expression are similar in all three groups (Fig. 2b). This supposition is supported by hierarchical cluster of all individual ignoring groups (Fig. 2a): although groups differ in the number of significant genes, the patterns of variation are shared among individuals from different groups. Additionally, these patterns (Fig. 2) suggest that the significant differences in mRNA expression among individuals are not due to just one or a few individuals (Fig. 2b). This

Fig. 3 Volcano plots among three individuals within each group. Negative $\log_{10} P$ values (e.g. red dotted line is for P value of $0.01 = 2$) vs. the difference of \log_2 expression values (difference of $1 =$ twofold). Three of the possible 36 comparisons within each group are displayed (individuals with '*' in figure two). The differences among three representative individuals are displayed for acclimate, field and G10 groups (Acc: A1 vs. A2, A2 vs. A5, and A1 vs. A5; Fld: F18 vs. F1, F1 vs. F5, and F18 vs. F5; G10: G2 vs. G6, G1 vs. G6, and G1 vs. G2). Notice axes are different for different comparison. P values are from SAS-JMP t -test of differences among any pair of individuals. Colours signify genes that are significantly different among individuals and have shared pattern of expression in the hierarchical clustering (Fig. 2).

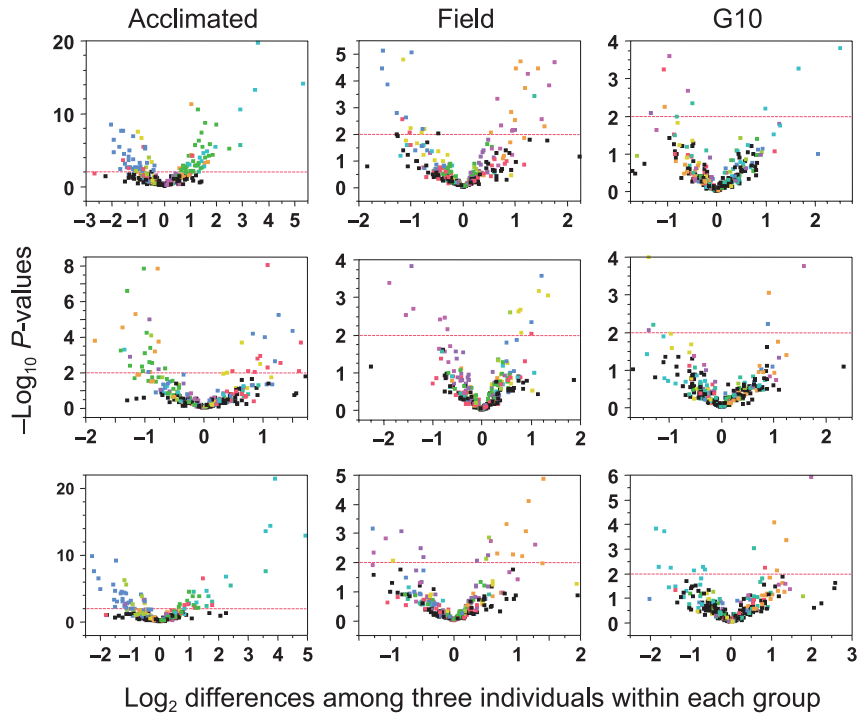


Table 3 Number of genes with significantly different mRNA expression. Significant different ($P < 0.01$) number of genes among individuals within acclimated (Acc), Field (Fld), and G10 groups and among all individuals ignoring groups (All). Average number of significant genes and 95% confidence interval for all possible 126 combination of five out of nine individuals. Table S1 provides a complete list of genes

	Number of significant genes within group			
	Across All	Acc	Fld	G10
$P < 0.01$	281	107	97	66
Av. no. for 5 out of 9 individuals		88	81	41
95% CI		90–86	87–75	45–37

later supposition is supported by the permutation analysis (see below) where all possible five out of nine individuals have similar patterns of differential mRNA expression (see below). Examples of the magnitude and associated P values with these differences are shown in the volcano plots (Fig. 3). Although only three of the possible 36 paired comparison in each group are shown in Fig. 3, these differences are a representative samples and suggest that the difference among acclimated individuals tend to be larger (x -axis, \log_2 differences in the least squared mean) and are more significant [y -axis, negative $\log_{10}(P$ values)].

To test if the number of genes with significant differences in expression were meaningful, all 126 possible combinations

of five out of nine individuals per group were examined (Table 3). This permutation of the data allows us to calculate means, variance and test for differences between groups for the number of genes with significantly different mRNA expression. Among these combinations, the average number of genes with a significant difference in expression share the same pattern as the analysis of all nine individuals: for the number of genes with significant differences in mRNA expression acclimated > field > G10. Notice that the overall number of significant genes is less because only five individuals were examined at a time and thus there are fewer degrees of freedom. For each group, mean numbers of genes with significant differences in expression among these 126 combinations have confidence intervals that do not overlap (Table 3) and these means are statistically different (Kruskal–Wallis nonparametric test $P < 0.001$). Thus, there is statistical support that there are a greater number of genes with significant difference in expression in acclimated vs. field or G10, and field vs. G10.

Variance in mRNA expression across genes

Another test of how the variation in mRNA differs among groups is to examine and compare the mean of the variance estimates across all genes among individuals. That is, rather than testing for the quantitative differences in mRNA expression, we tested whether the mean variation in mRNA expression across all 284 genes among individuals was different among groups. Because the variance is a function of the magnitude of the mean, the measures of expression

of all 284 genes was normalized so that the average expression for each gene was equal to one: $lsmean_{pg}/(Avg_g)$ where the $lsmean_p$ is the least square mean for the p_{th} individual and the g_{th} gene, and these measures are divided by Avg_g , the average least square mean for the g_{th} gene. For each individual, the mean variance for all 284 measures of expression using these normalized values was calculated. This mean variation in expression among the nine individuals is significantly different between groups (Kruskal–Wallis test, $P < 0.001$) with mean variance of 0.575 (SD = 0.346), 0.422 (SD = 0.256) and 0.386 (SD = 0.132) for acclimated, field and G10, respectively. It is interesting that the standard deviation of the variance is greatest in the acclimated group, suggesting greater differences in the variation among individuals in this group.

Homogeneity of variance

A third test of how the variation among individuals for mRNA expression differs among the three groups is to examine the similarity of the variance for each gene. We tested the similarity for each gene by applying the Bartlett's test for homogeneity of variance among groups using the least squared means for each individual (Table 4). Of the 284 measures of mRNA expression used in this experiment, 22 (7%), had an unequal variance among groups ($P < 0.01$). Among these 22 genes with significant differences in the individual variation in mRNA expression, 17 (77%) of genes had greater expression variation in acclimated individuals than G10 individuals. This bias of larger variance in mRNA expression in the acclimated group vs.

Table 4 Homogeneity of variance. Number of genes with unequal variances (Bartlett's test for homogeneity of variance). Last column is the percentage of genes with attributes defined by in the first column (e.g. Acc > G10, the variance is larger in Acc vs. G10 for 77.3% of the 17 genes)

	Number of genes	Number of significant genes	Percentage of significant genes
Total	284	22	100
Acc > G10	210	17	77.3
Fld > G10	154	13	59.1
Acc > Fld	184	11	50.0
Acc & Fld > G10	131	11	50.02

G10 is unlikely to occur by chance ($\chi^2 P < 0.01$). For the acclimated vs. field, or field vs. G10 there are the same (50%) or nearly the same (59%) number of genes with greater variance in mRNA expression.

Significant differences between groups in mRNA expression

We expected a difference in the variance in mRNA expression between groups, but did not necessarily expect a difference in the mean of mRNA expression. Using a P value of 0.01, seven genes (2.5%) have a significant difference in mRNA expression among groups (Fig. 4). None of these are significant with a Bonferroni's corrected P value of 3.5×10^{-5} . To determine which groups differ in their gene-specific

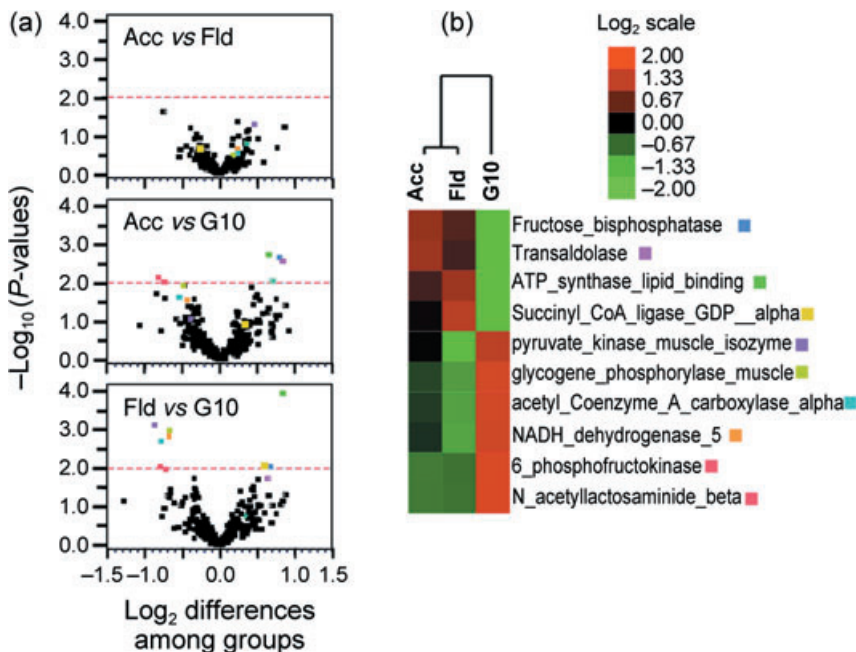


Fig. 4 Differences among groups. Among all pairs of groups there are 10 genes that are significantly different (t -test, red dotted line is a P value = 0.01). (a) Volcano plots of \log_2 differences in expression vs. the negative \log_{10} of P value ($2 = 0.01$). P values are from SAS-JMP t -test of differences among any pair of groups. Colours signify genes with shared pattern of expression in the hierarchical clustering in 4b. (b) Hierarchical clustering of patterns of gene expression. Acc, acclimate; Fld, field; and G10, inbred population.

mRNA expression, *t*-tests were applied between the three pairs of comparisons (Acc vs. Fld, Acc vs. G10 and Fld vs. G10). There are no significant differences between the acclimated and the field groups (Fig. 4a) and there are 10 significant differences in mRNA expression between either the acclimated or field vs. G10 group (Fig. 4b).

Discussion

The genetic basis for the variation in mRNA expression among natural populations, including *Fundulus heteroclitus*, is not well understood. In other species, our understanding of the genetics of mRNA expression has relied on the study of inbred strains (Schadt *et al.* 2003; Wayne *et al.* 2004; Gibson & Weir 2005) or cell culture (Monks *et al.* 2004). Using these systems, the variation in mRNA expression measured by microarrays appears to be genetically based: it differs between inbred lines, is associated with quantitative trait loci and has narrow sense heritability (h^2) greater than 30% (Cheung *et al.* 2003; Gibson & Weir 2005; Sharma *et al.* 2005; Tan *et al.* 2005; Rockman & Kruglyak 2006). Heritability of mRNA expression has been measured in a variety of organisms. For example, in 10 lines of *Drosophila*, 663 of 7886 measured genes (8%) had significant genetic variation with a medium $h^2 = 0.47$ (quartile range 0.39–0.60) (Wayne *et al.* 2004). Among 112 *Saccharomyces cerevisiae* segregants, 3546 out of 5727 measured genes (62%) had a $h^2 > 0.69$ (Brem & Kruglyak 2005). Using lymphoblast human cell lines, among 15 families, 762 out of 2430 (31%) of differentially expressed genes had a significant h^2 with median of 0.34 (Monks *et al.* 2004; Williams *et al.* 2007).

These studies on inbred lines or cell culture are informative. However, they do not provide data on the genetics of the variation in mRNA expression in outbred species. For humans, twin-studies (Sharma *et al.* 2005; Tan *et al.* 2005) and replicate measures of the same individuals over time (Whitney *et al.* 2003; Radich *et al.* 2004; Cobb *et al.* 2005; Eady *et al.* 2005) suggest a strong genetic component to the natural variation in mRNA expression. For natural populations of *Fundulus heteroclitus*, it is unclear if differences within and among populations (Oleksiak *et al.* 2002; Oleksiak *et al.* 2005; Whitehead & Crawford 2005a, b; Crawford & Oleksiak 2007) are a function of genetic variation or other less evolutionarily important parameters. The data presented here supports the hypothesis that much of the variation in mRNA expression is a function of genetic variation.

The genetic variation based on microsatellite markers in *F. heteroclitus* from a single North Carolina population is greater in the outbred groups (acclimated and field) than in the inbred G10 individuals (Fig. 1). Among G10 individuals, they have half the allelic richness and 75% of the heterozygosity of the outbred group. Additionally, the

relatedness among G10 individuals is nearly equal to full sibs ($R = 0.43$), but among outbred individuals the relatedness is not different from zero. The reduced genetic variation is expected in the G10 individuals because they originated from fewer than 16 individuals and were inbred for 10 generations; whereas the field caught individuals have effective population sizes that exceed 10^5 (Adams *et al.* 2006). The only measure of genetic variation that is not different for G10 is F_{IS} where F_{IS} is the fixation index relative to individuals within a subpopulation or group. The lack of a difference in F_{IS} is reasonable because each generation of siblings of the G10 group was allowed to breed randomly which allowed for the re-establishment of Hardy–Weinberg equilibrium (Table 1).

Among G10 individuals, there is also a lower variation in mRNA expression relative to acclimated individuals: fewer genes have significant differences in mRNA expression among individuals (Table 3, Figs 2 and 3), and the mean variance across all genes is significantly less. Additionally, when examining the gene specific variation in mRNA expression among individual within groups, 22 genes have significant differences in the variation among groups and for 77% of these genes, the variations in mRNA expression are lower in G10 than in acclimated individuals. These differences are found among individuals raised or acclimated to a laboratory environment with constant food, salinity, temperature, oxygen and lack of predators. These data support the supposition that outbred acclimated individuals have greater variation in mRNA expression than the inbred G10 individuals even though both groups share common stable environment.

Among outbred individuals (acclimated and field), there are also differences in the variation in mRNA expression: acclimated individuals had more genes with significant differences in mRNA expression (107 vs. 97) and greater variation in mRNA expression across all 284 genes. However, when comparing the variance among individual per gene, there is not a difference in the magnitude of the gene specific variation in the acclimated and field groups (Table 3). These data indicate that acclimated individuals appear to have greater or nearly equal variation as field individuals. Thus, these data support a surprising conclusion: the environmental variation in the field [tidal changes, spatial and temporal changes in salinity, food availability, oxygen, etc. (Marshall 2003; Marshall *et al.* 2005)] does not have a major affect on the variation in gene expression. Notice, we are not addressing whether changes in the environment alters mRNA expression. Instead, our data suggest that variation in environment does not have a substantial effect on the variation in mRNA expression.

Among the three groups (acclimated, field and G10), there are few differences in expression: seven genes have significant difference in mRNA expression with a critical *P* value of 1%, none with Bonferroni's correct *P* value.

This supposition of few differences among groups is supported by patterns of mRNA expression (Fig. 2a): individuals in the three groups do not cluster together. Instead, individuals in each group share common patterns of expression with individuals in other groups (Fig. 2a). With *t*-test between each pair of groups (which suffer from high false-positive rate), the differences in expression are only significant between G10 and the two outbred groups (Fig. 4). These data suggest that the laboratory vs. the field environments does not have a large effect on mRNA expression. If we propose that few significant differences between G10 and the outbred groups are important, it is difficult to imagine a common environmental factor for the field and acclimated groups that could explain the difference between these groups vs. the G10 group. Alternatively, if much of the variation in mRNA expression is genetically based, as suggested by the correlation between genetic variation and the variation in expression, then one would have to speculate that the G10 individuals have different or less frequent genotypes that affect the expression of these seven mRNAs. This difference is most parsimoniously explained by due random drift due to recent bottleneck.

These patterns of variation in mRNA expression in acclimated, field and G10 individuals are consistent with the hypothesis that much of the variations among individuals are due to genetic variation. Specifically, there are fewer differences among G10 individuals that share 43% of their alleles vs. completely outbred individuals (acclimated) even though both were subjected to similar laboratory condition for at least 6 months. Additionally, there is little support for environmental variation effect on mRNA expression: acclimated individuals have equal or more variation in mRNA expression vs. field individuals that suffers the slings and arrows of environmental variation associated with estuarine environments. Added to these observations is that there is little significant variation in mRNA expression when the same individual is repetitively measured over a 6-week period (i.e. mRNA expression from blood sample every 2 weeks over a 6-week period; Scott *et al.* 2009). That is, adult mRNA expression has little temporal variation. Together, these data strongly support the hypothesis that the large interindividual variations in gene expression measure here and elsewhere (Oleksiak *et al.* 2002; Oleksiak *et al.* 2005; Whitehead & Crawford 2005, 2006a, b; Crawford & Oleksiak 2007) are unlikely to reflect environmental variation and is more reasonable assigned to genetic variation.

One of the assumptions in this work is that acclimation removes most, if not all, of the physiological differences among individuals. Clearly, acclimation to a common environment can remove many physiological differences especially differences in enzyme expression (Hochachka & Somero 1984; Prosser 1986; Crawford & Powers 1989; Schmidt-Neilsen 1990; Segal & Crawford 1994; Pierce &

Crawford 1997). However, these observations do not refute that other nonheritable mechanisms can affect mRNA expression. For example in clones of sea anemones, metabolic rates are not affected by acclimation to a common temperature. Instead, metabolisms among genetically identical individuals reflect an individual's developmental temperature. Similarly, the maximum expression of heat shock proteins in sea urchins (*Strongylocentrotus purpuratus*) was unaffected by acclimation temperature but appeared to be influenced by irreversible acclimation at early life stages (Osovitz & Hofmann 2005). Additionally, an individual's phenotype can be influenced by maternal and other epigenetic effects. Thus, one could suggest that the inbred individuals (G10) whose parents all experience the same environment affected mRNA expression differently than acclimated individual whose parents experience a wide range of environments. However, there are few differences in expression among G10, acclimated or field individuals. Thus, epigenetic affects causing a difference in gene expression are not supported. Alternatively, one could suggest that the variation in mRNA expression (but not a different in the mean expression) is related to the environmental variation experience by parents or developing embryo. That is, the more variable the paternal or developmental environment, the greater the variation in mRNA expression. To explain most of the data, this hypothesis that the variation in expression is a function of parental or developmental environmental variation would require that greater environmental variation produces many different adult phenotypes, and thus there is a greater variation in mRNA expression. This hypothesis is unlike any currently available data, and cannot be readily rejected. However, it seems more parsimonious to suggest that the larger interindividual variation in mRNA expression is related to genetic variation, rather than a novel epigenetic mechanism that does not affect the mean mRNA expression but instead creates greater individual variation.

To summarize, our data support the hypothesis that variation in metabolic mRNA expression is primarily related to the genetic variation among individuals. For G10 individuals, high amounts of relatedness and low levels of allelic richness are associated with less variation in mRNA expression. Surprisingly, the variation in metabolic mRNA expression is either lower or, at the very least, similar among field vs. acclimated individuals. These data indicate that much of the variation in mRNA expression is related to genetic variation and less of the variation is in response to environmental variation.

Acknowledgements

The authors thank Richard Winn and Michelle Norris at the Aquatic Biotechnology and Environmental Laboratory at the University of Georgia for providing the G10 fish. We thank Dr

Marjorie Oleksiak for the creation of *Fundulus* microarrays, and Jeffrey VanWye for EST sequencing and printing of the microarray. This project was supported by NSF (OCE 0308777 BE/GEN-EN Functional Genomics), National Institutes of Health grant (NHLBI R01 HL065470) to Douglas L. Crawford and National Institutes of Environmental Health Sciences (NIEHS) through NIEHS ARCH grant S11 ES 11181.

References

- Adams SM, Oleksiak MF, Duvernell DD (2005) Microsatellite primers for the Atlantic coastal killifish, *Fundulus heteroclitus*, with applicability to related *Fundulus* species. *Molecular Ecology Notes*, **5**, 275–277.
- Adams SM, Lindmeier JB, Duvernell DD (2006) Microsatellite analysis of the phylogeography, Pleistocene history and secondary contact hypotheses for the killifish, *Fundulus heteroclitus*. *Molecular Ecology*, **15**, 1109–1123.
- Bowcock AM, Ruizlinares A, Tomfohrde J *et al.* (1994) High-resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455–457.
- Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences, USA*, **102**, 1572–1577.
- Cheung VG, Jen KY, Weber T *et al.* (2003) Genetics of quantitative variation in human gene expression. *Cold Spring Harbor Symposia on Quantitative Biology*, **68**, 403–407.
- Cobb JP, Mindrinos MN, Miller-Graziano C *et al.* (2005) Application of genome-wide expression analysis to human health and disease. *Proceedings of the National Academy of Sciences, USA*, **102**, 4801–4806.
- Crawford DL, Oleksiak MF (2007) The biological importance of measuring individual variation. *Journal of Experimental Biology*, **210**, 1613–1621.
- Crawford DL, Powers DA (1989) Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish *Fundulus heteroclitus*. *Proceedings of the National Academy of Sciences, USA*, **86**, 9365–9369.
- Eady JJ, Wortley GM, Wormstone YM *et al.* (2005) Variation in gene expression profiles of peripheral blood mononuclear cells from healthy volunteers. *Physiological Genomics*, **22**, 402–411.
- Eberwine J (1996) Amplification of mRNA populations using aRNA generated from immobilized oligo (dT) -T7 primed cDNA. *BioTechniques*, **20**, 584–&.
- Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*, 4th edn. Pearson Education Limited, Harlow, UK.
- Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends in Genetics*, **21**, 616–623.
- Goudet J (1995) FSTAT (version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.
- Hochachka PW, Somero GN (1984) *Biochemical Adaptation*. Princeton University Press, Princeton, New Jersey.
- Kerr MK, Churchill GA (2001) Experimental design for gene expression microarrays. *Biostatistics*, **2**, 183–201.
- Marshall WS (2003) Rapid regulation of NaCl secretion by estuarine teleost fish: coping strategies for short-duration freshwater exposures. *Biochimica et Biophysica Acta (BBA) — Biomembrana*, **1618**, 95.
- Marshall WS, Ossum CG, Hoffmann EK (2005) Hypotonic shock mediation by p38 MAPK, JNK, PKC, FAK, OSR1 and SPAK in osmosensing chloride secreting cells of killifish opercular epithelium. *Journal of Experimental Biology*, **208**, 1063–1077.
- Monks SA, Leonardson A, Zhu H *et al.* (2004) Genetic inheritance of gene expression in human cell lines. *American Journal of Human Genetics*, **75**, 1094–1105.
- Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature Genetics*, **32**, 261–266.
- Oleksiak MF, Roach JL, Crawford DL (2005) Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nature Genetics*, **37**, 67–72.
- Osovitz CJ, Hofmann GE (2005) Thermal history-dependent expression of the hsp70 gene in purple sea urchins: biogeographic patterns and the effect of temperature acclimation. *Journal of Experimental Marine Biology and Ecology*, **327**, 134.
- Paschall JE, Oleksiak MF, VanWye JD *et al.* (2004) FunnyBase: a systems level functional annotation of *Fundulus* ESTs for the analysis of gene expression. *BMC Genomics*, **5**, 96.
- Patterson TA, Lobenhofer EK, Fulmer-Smentek SB *et al.* (2006) Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nature Biotechnology*, **24**, 1140–1150.
- Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
- Pierce VA, Crawford DL (1997) Phylogenetic analysis of thermal acclimation of the glycolytic enzymes in the genus *Fundulus*. *Physiological Zoology*, **70**, 597–609.
- Prosser CL (1986) *Adaptational Biology: From Molecules to Organisms*. Wiley, New York.
- Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258–275.
- Radich JP, Mao M, Stepaniants S *et al.* (2004) Individual-specific variation of gene expression in peripheral blood leukocytes. *Genomics*, **83**, 980–988.
- Raymond M, Rousset F (1995) GenePop (version 1.2) — population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Rockman MV, Kruglyak L (2006) Genetics of global gene expression. *Nature Reviews Genetics*, **7**, 862–872.
- Schadt EE, Monks SA, Drake TA *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, **422**, 297–302.
- Schmidt-Neilsen K (1990) *Animal Physiology: Adaptation and Environment*, 4th edn. Cambridge University Press, New York.
- Scott CP, VanWye JD, McDonald D, Crawford Douglas L (2009) Technical analysis of *Fundulus heteroclitus* cDNA microarrays. *PLoS ONE*, **4**, e4486.
- Segal JA, Crawford DL (1994) LDH-B enzyme expression: the mechanisms of altered gene expression in acclimation and evolutionary adaptation. *American Journal of Physiology*, **267**, R1150–R1153.
- Sharma A, Sharma VK, Horn-Saban S *et al.* (2005) Assessing natural variations in gene expression in humans by comparing with monozygotic twins using microarrays. *Physiological Genomics*, **21**, 117–123.
- Sokal RR, Rohlf FJ (1995) *Biometry: The Principles and Practice of Statistics in Biological Research*, 3rd edn. W.H. Freeman, New York.
- Tan Q, Christensen K, Christiansen L *et al.* (2005) Genetic dissection of gene expression observed in whole blood samples of elderly Danish twins. *Human Genetics*, **117**, 267–274.

- Wayne ML, Pan YJ, Nuzhdin SV, McIntyre LM (2004) Additivity and trans-acting effects on gene expression in male *Drosophila simulans*. *Genetics*, **168**, 1413–1420.
- Whitehead A, Crawford D (2005) Variation in tissue-specific gene expression among natural populations. *Genome Biology*, **6**, R13.11–R13.13.14.
- Whitehead A, Crawford DL (2006a) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences, USA*, **103**, 5425–5430.
- Whitehead A, Crawford DL (2006b) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*, **15**, 1197–1211.
- Whitney AR, Diehn M, Popper SJ *et al.* (2003) Individuality and variation in gene expression patterns in human blood. *Proceedings of the National Academy of Sciences, USA*, **100**, 1896–1901.
- Williams RBH, Chan EKF, Cowley MJ, Little PFR (2007) The influence of genetic variation on gene expression. *Genome Research*, **17**, 1707–1716.
- Wolfinger RD, Gibson G, Wolfinger ED *et al.* (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology*, **8**, 625–637.
- Yu X, Chu TM, Gibson G, Wolfinger RD (2004) A mixed model approach to identify yeast transcriptional regulatory motifs via microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, Article22.

Dr. Cinda Scott has just received her Ph.D. (University of Miami) and is interested in the genetics and functional genomics of evolutionary adaptation. Dr. Dean Williams is an Assistant Professor at Texas Christian University in the Department of Biology and is interested in evolutionary genomics, conservation genetics, behavioral ecology and ecology and genetics of invasive species. Dr. Douglas Crawford is Professor at the Rosenstiel School of Marine and Atmospheric Sciences in the Marine Biology & Fisheries Division. Dr. Crawford's research interest focuses on the evolution of quantitative variation in gene expression and using evolutionary analysis to better understand health, disease, physiology and biochemistry.

Supporting information

Additional supporting information may be found in the online version of this article:

Fig. S1 Male and female loop designs used in microarray studies.

Fig. S2 Box plots of log₂-loess normalized data for all samples (individuals for both dyes).

Table S1 List of all Genes. *P*-values for differences among all individuals, if significant (= 1), Gene names, Treatment group (Acclimated, Field, G10, *P*-values for difference among individuals within acclimated treatment, if significant (= 1), *P*-values for difference among individuals within Field treatment, if significant (= 1), *P*-values for difference among individuals within G10 treatment, if significant (= 1).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.