

3-2011

Population Proteomics: Quantitative Variation Within and Among Populations in Cardiac Protein Expression

Bernard Rees

University of New Orleans, brees@uno.edu

Follow this and additional works at: https://scholarworks.uno.edu/biosciences_facpubs



Part of the [Biology Commons](#)

Recommended Citation

Rees, B.B., Andacht, T., Skripnikova, E., and Crawford, D.L. 2011. Population proteomics: quantitative variation within and among populations in cardiac protein expression. *Molecular Biology and Evolution* 28 (3): 1271-1279.

This Article is brought to you for free and open access by the Department of Biological Sciences at ScholarWorks@UNO. It has been accepted for inclusion in Biological Sciences Faculty Publications by an authorized administrator of ScholarWorks@UNO. For more information, please contact scholarworks@uno.edu.

Population Proteomics: Quantitative Variation Within and Among Populations in Cardiac Protein Expression

Bernard B. Rees,¹ Tracy Andacht,² Elena Skripnikova,¹ and Douglas L. Crawford^{*,3}

¹Department of Biological Sciences, University of New Orleans

²Proteomics Resource Facility, Integrated Biotech Laboratories, University of Georgia

³Marine Biology & Fisheries, Rosenstiel School of Marine and Atmospheric Sciences, University of Miami

*Corresponding author: E-mail: dcrawford@rsmas.miami.edu.

Associate editor: Willie Swanson

Abstract

Population analysis of gene expression is typically achieved by quantifying levels of mRNA; however, gene expression is also a function of protein translation and turnover. Therefore, a complete understanding of population variation in gene expression requires quantitative knowledge of protein expression within and among natural populations. We used two-dimensional fluorescence difference gel electrophoresis (2D-DIGE) to quantitatively compare expression of heart ventricle proteins among 18 individuals in three populations of the teleost fish *Fundulus*. Among populations, expressions between orthologous proteins and mRNAs were generally positively correlated. Additionally, similar to the pattern of cardiac mRNA expression for the same populations, we found considerable variation in protein expression both within and among populations: Of 408 protein features in 2D gels, 34% are significantly different ($P < 0.01$) among individuals within a population, 9% differ between populations, and 12% have a pattern of expression that suggests they have evolved by natural selection. Although similar to mRNA expression, the frequency of significant differences among populations is larger for proteins. Similar to mRNA expressions, expressions of most proteins are correlated to the expressions of many other proteins. However, the correlations among proteins are more extensive than the correlation for similar RNAs. These correlations suggest a greater coordinate regulation of protein than mRNA expression. The larger frequency of significant differences among populations and the greater frequency of correlated expression among proteins versus among RNAs suggest that the molecular mechanisms affecting protein expression enhance the differences among populations, and these regulatory steps could be a source of variation for adaptation.

Key words: 2D-DIGE, gene expression, evolution, *Fundulus*.

Introduction

Natural variation in gene expression and its causes are central questions of evolutionary biology (Crawford and Powers 1992; Enard et al. 2002; Oleksiak et al. 2002; Rockman and Wray 2002; Gibson 2003; Townsend et al. 2003; Wittkopp et al. 2004; Lemos et al. 2005; Whitehead and Crawford 2006b; Crawford and Oleksiak 2007; Fay and Wittkopp 2008; Wittkopp et al. 2009). Microarrays revealed that 15–25% of genes show significant variation in mRNA expression among individuals within healthy populations in humans, mice, fish, and flies (Jin et al. 2001; Pritchard et al. 2001; Oleksiak et al. 2002; Bray et al. 2003; Schadt et al. 2003; Whitney et al. 2003; Yan et al. 2003; Radich et al. 2004; Doss et al. 2005). In the teleost fish *Fundulus heteroclitus*, variation in cardiac gene expression is associated with differences in metabolism (Pierce and Crawford 1997; Podrabsky et al. 2000; Oleksiak et al. 2005) and appears to be evolving by natural selection (Oleksiak et al. 2002; Whitehead and Crawford 2006a, 2006b). These data suggest that much variation in mRNA expression is biologically meaningful. However, protein expression is more likely to determine an organism's phenotype. Thus, strong conclusions regarding population variation in gene expression require measurements of proteins. Yet, studies ad-

ressing population variation in broad-scale patterns of protein expression are lacking (Biron et al. 2006; Karr 2008; Nedelkov 2008).

Here, we present a quantitative analysis of cardiac protein expression in natural populations of the teleost fish in the genus *Fundulus*. This small fish is widely spread throughout marine, estuarine, and freshwater habitats of North America. Two closely related species, *F. heteroclitus* and *F. grandis*, occur in estuaries of the Atlantic and Gulf of Mexico, and the former has been well studied with respect to population biology, including measures of cardiac mRNA expression (Powers et al. 1993; Oleksiak et al. 2002, 2005; Burnett et al. 2007; Crawford and Oleksiak 2007). In this study, we use two-dimensional fluorescence difference gel electrophoresis (2D-DIGE) to quantitatively compare protein expression in the same tissue (heart ventricle), populations, and experimental design employed in previous cDNA microarray studies (Oleksiak et al. 2002). We describe significant within- and among-population variation in protein expression and provide evidence that among-population variation is consistent with evolution by natural selection. For proteins that are orthologous to specific mRNAs, population means of protein and mRNA expression were generally positively correlated.

Importantly, statistically significant differences in protein expression are more frequent than found previously for mRNA (Oleksiak et al. 2002), and unlike mRNA, protein expression is often a function of body mass. These observations suggest additional levels of control on protein expression compared with mRNA expression. These additional controls on protein expression could evolve to affect adaptive change.

Materials and Methods

Animals and Acclimations

Eighteen individuals from three teleost fish populations (six/population) were used in this study: northern *Fundulus heteroclitus*, southern *F. heteroclitus*, and *F. grandis* (sister taxa to *F. heteroclitus*). Northern *F. heteroclitus* were collected in Wiscasset, ME (lat 43°57'15"N, long 69°43'12"W). Southern *F. heteroclitus* were collected from the ferry dock to Sapelo Island, GA (lat 31°29'25"N, long 81°15'48"W). *Fundulus grandis* were collected on Santa Rosa Island, FL (lat 30°21'16"N, long 87°02'33"W). Fish were caught using minnow traps, transported back to the laboratory, and maintained in recirculating aquarium systems with a single shared water supply. Populations were kept in separate tanks, with shared water circulated through all tanks via a central sump. The system was maintained at a salinity of 15 ppt in artificial seawater made using Instant Ocean Sea Salt and municipal water dechlorinated via reverse osmosis. All fish were put through a pseudowinter cycle: Water temperature was maintained at 8 °C with a 10:14 h light:dark cycle. After 6 weeks of pseudowinter, temperatures were slowly increased to 24 °C, the lighting was changed to a 14:10 light:dark cycle, and the fish were allowed to spawn. Postreproductive fish were used in this study, and thus, fish had spent a minimum of 6 months in a common environment. Fish were fed OSI Marine Flake ad libitum once daily in the evening.

Protein Isolation, Labeling, and Annotation

Fundulus hearts were dissected, flash frozen in liquid nitrogen, and stored at -80 °C until proteins were extracted by sonication in 7 M urea, 2 M thiourea, 4% [(3-cholamidopropyl)-dimethyl-ammonio]1-propanesulfonate (CHAPS), and 15 mM Tris, pH 8.4 at 4 °C. Insoluble material was removed by centrifugation at 100,000 × g for 4 h at 20 °C. Ionic interfering substances were removed, and protein extracts were concentrated with three buffer exchanges of the extraction buffer using Microcon YM-10 centrifugal filter devices (Millipore). A pooled reference was created by combining 25 µg of each sample. All samples, including the reference, were diluted to 5 mg/ml using the extraction buffer.

Separate aliquots of 50 µg of each individual sample were labeled with Cy3 and Cy5 NHS ester (GE Healthcare) using a minimum ratio of 200 pmol dye:50 µg sample. The pooled reference sample was labeled in bulk with Cy2 NHS ester (at a ratio of 200 pmol dye:50 µg sample) at 4 °C for 30 min. The labeling reaction was quenched with 10 nmol lysine. Prior to electrophoresis, 50 µg of two individually

labeled samples (one with Cy3 one with Cy5) were combined with 50 µg of the Cy2-labeled pooled reference and diluted to a final volume in 7 M urea, 2 M thiourea, 2% CHAPS, 0.5% IPG buffer (GE HealthCare, immobilized pH gradient), 18.2 mM dithiothreitol (DTT), and 0.002% bromophenol blue. Protein samples were loaded on 18-cm pH 3–10 IPG strips by active rehydration at 30 V and separated by isoelectric focusing for 32,000 Vh using the IPGphor (GE Healthcare). After isoelectric focusing, IPG strips were equilibrated in 6 M urea, 2% sodium dodecyl sulfate (SDS), 65 mM DTT, 30% glycerol, 50 mM Tris, pH 8.8, and 0.002% bromophenol blue for 15 min at room temperature. IPG strips were then equilibrated with the above buffer, replacing DTT with 135 mM iodoacetamide, for 15 min at room temperature. Proteins were separated in the second dimension by SDS-polyacrylamide gel electrophoresis (Laemmli 1970) on 20 × 26-cm 8–15% polyacrylamide gels using the Dalt II (GE Healthcare).

Gels were fixed in 30% ethanol and 7.5% acetic acid overnight at room temperature. Gels were imaged by a Typhoon 9400 (GE Healthcare) after optimizing the gain on the photomultiplier tubes for each laser to achieve the broadest dynamic range. Gel images were matched, and spot intensities were analyzed based on the log₂ standardized volume with DeCyder 4.0 software (GE Healthcare).

Gels selected for protein identification were stained with Sypro Ruby (Molecular Probes) overnight, destained in 10% methanol and 6% acetic acid for 30 min at room temperature, imaged, and matched to the Cy images using DeCyder software. Protein spots of interest were excised as 2.0-mm discs, washed, reduced and alkylated, and digested with trypsin by the Spot Handling Workstation (GE Healthcare). Briefly, gel discs were washed twice with 50 mM ammonium bicarbonate/50% methanol for 20 min followed by a wash with 75% acetonitrile for 20 min and dried at 40 °C for 10 min. These gel discs were then incubated in 10 mM DTT/20 mM ammonium bicarbonate at 37 °C for 1 h. The DTT solution was removed and immediately replaced with 100 mM iodoacetamide/20 mM ammonium bicarbonate and incubated at room temperature in the dark for 30 min. Gel discs were then incubated with 200 ng sequencing grade trypsin (Promega) at 37 °C for 2 h. Peptide products were extracted from gel discs in two washes of 50% acetonitrile/0.1% TFA for 20 min at room temperature and concentrated by SpeedVac (Jouan). After concentration, approximately 25% of the resulting volume of each peptide digest was spotted onto targets for mass spectrometry (MS) with partially saturated α -cyano-4-hydroxy-cinnamic acid (Sigma).

Matrix-assisted/laser desorption-ionization time-of-flight mass spectrometry was performed on the 4700 Proteomics Analyzer (Applied Biosystems) using standard acquisition methods. Tandem MS (MS/MS) data were acquired in a data-dependent fashion, selecting the 15 most intense peaks from the initial MS spectrum. MS spectra were calibrated using two trypsin autolysis peaks (1045.5 and 2211.1 m/z). MS/MS spectra were calibrated using the instrument default processing method.

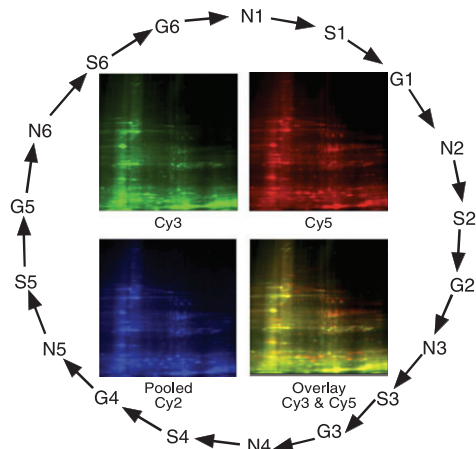


Fig. 1 Experimental design and fluorescent images. Loop design used to compare 18 individuals. Individuals were labeled with Cy3 or Cy5 fluorescent dyes and electrophoresed together with a pooled sample labeled with Cy2. Individuals compared in a gel are shown with connecting arrows where the bases of the arrows are Cy3 labeled and the heads of the arrows are the Cy5-labeled samples. Fg: *Fundulus grandis*, S: southern *F. heteroclitus*, and N: northern *F. heteroclitus*. Fluorescent images: gels with Cy3, Cy5, and pooled Cy2 samples.

Several databases were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/entrez/>), including NCBI nr and all the available Actinopterygii (ray-finned fishes) UniGene databases (*Fundulus heteroclitus*, *Danio rerio*, *Gasterosteus aculeatus*, *Oryzias latipes*, *Salmo salar*, and *Takifugu rubripes*). The Actinopterygii UniGene databases were concatenated into one database, and NCBI nr, Actinopterygii UniGene, and *Fundulus heteroclitus* UniGene databases were incorporated into Mascot 1.9.05 (<http://www.matrixscience.com/>). Peptide mass fingerprints and MS/MS data were searched against these databases considering fixed cysteine carbamidomethylation and methionine oxidation modifications, one missed tryptic cleavage, and 25 ppm mass accuracy. Identifications were cross-examined using mass accuracy, molecular weight, and pI (supplementary table 1, Supplementary Material online).

Experimental Design and Statistics

A loop design was used to determine sample loading for gel electrophoresis. Each gel contained Cy3- and Cy5-labeled samples originating from different fish populations and a pooled protein sample labeled with Cy2 (fig. 1). The latter served as an internal control across all gels to facilitate gel matching and spot quantification. Every tissue extract was labeled with both Cy3 and Cy5, resulting in a total of 18 gels and 54 fluorescent images (18 Cy3, 18 Cy5, and 18 Cy2 images). Thus, the experimental design was balanced with respect to population of origin and fluorescent labels used for quantification.

Analyses of protein spot intensity used \log_2 of the Cy3/Cy2 or Cy5/Cy2 fluorescent ratios (hereafter referred to as “Cy ratio”) and employed SAS JMP-Genomics (3.2) and MatLab (7.1), a matrix algebra-programming environment (MathWorks, Natick, MA). We removed systematic bias in the data by applying a mixed-model normalization to all

36 measures of protein expression (18 individuals, two dyes per individuals). We controlled for the effect of body mass by using the residual of \log_2 Cy ratio regressed against \log_{10} body mass. We used these residuals for each protein spot in an analysis of variance (ANOVA) to determine differences in protein expression among individuals within each population using the linear mixed model previously described for microarrays (Kerr and Churchill 2001; Wolfinger et al. 2001; Yu et al. 2004; Patterson et al. 2006). The general model was $y_{ijkpo} = m \pm G_i \pm D_j \pm S_p \pm V_k(P_o)_p \pm P_{op} \pm e_{ijkpo}$. In this model, y_{ijkpo} is the signal from the i th gel with dye j for protein in spot p in individual k among population o . The variable m represents the mean signal of all the proteins across the entire experiment. The overall variation in gel and dyes is represented by the terms for gels (G_i), and dye (D_j). The term S_p is the average signal for protein spot p across gels, dyes, and populations. The term e_{ijkpo} , or residual, represents measurements of experimental error. Finally, the two parameters of interest are $V_k(P_o)_p$, the variation for each protein p among individuals k within each population o and P_{op} , the variation among populations. For the evolutionary analysis, a similar linear analysis was performed except that variation is between two groups (northern *F. heteroclitus* vs. the joint variation in southern *F. heteroclitus* plus *F. grandis*), with individuals nested within these two groups. We calculated the F statistics separately for each protein and compared them with the tabulated F distribution at a P value ≤ 0.01 significance level (supplementary table 1, Supplementary Material online). The degrees of freedom (df) depends on the number of individuals for which the protein spot was resolved by 2-DDIGE (three to six individuals per protein for each population). Ideally, there would be two dyes, six individuals, and three populations per spot producing 2, 15 df for ANOVA among populations; 15, 17 df for ANOVA among individuals within populations, and 1, 16 for ANOVA testing the evolutionary hypothesis. However, not all proteins were resolved on every gel, and for these, the df reflected the lower number of individuals used in these analyses. These F statistics were computed on a per-protein basis, and thus, there is no need to assume a common error variance across spots.

Protein–RNA Homologies

To verify orthologies, proteins and cDNAs were compared with each other and to the teleost fish *Danio rerio* ref-seq database (supplementary table 2, Supplementary Material online). First, we used BlastP to compare the intact protein or individual peptide sequences from MASCOT analysis to the ref-seq protein database for *D. rerio*. Similarly, we used BlastX to compare the *F. heteroclitus* expressed sequence tags (EST) database (for CN952504:CN956229 used to build the microarray) with the same *D. rerio* ref-seq protein database.

Results and Discussion

Variation Within and Among Populations

Similar to an earlier microarray study examining gene expression (Oleksiak et al. 2002), natural variation in protein

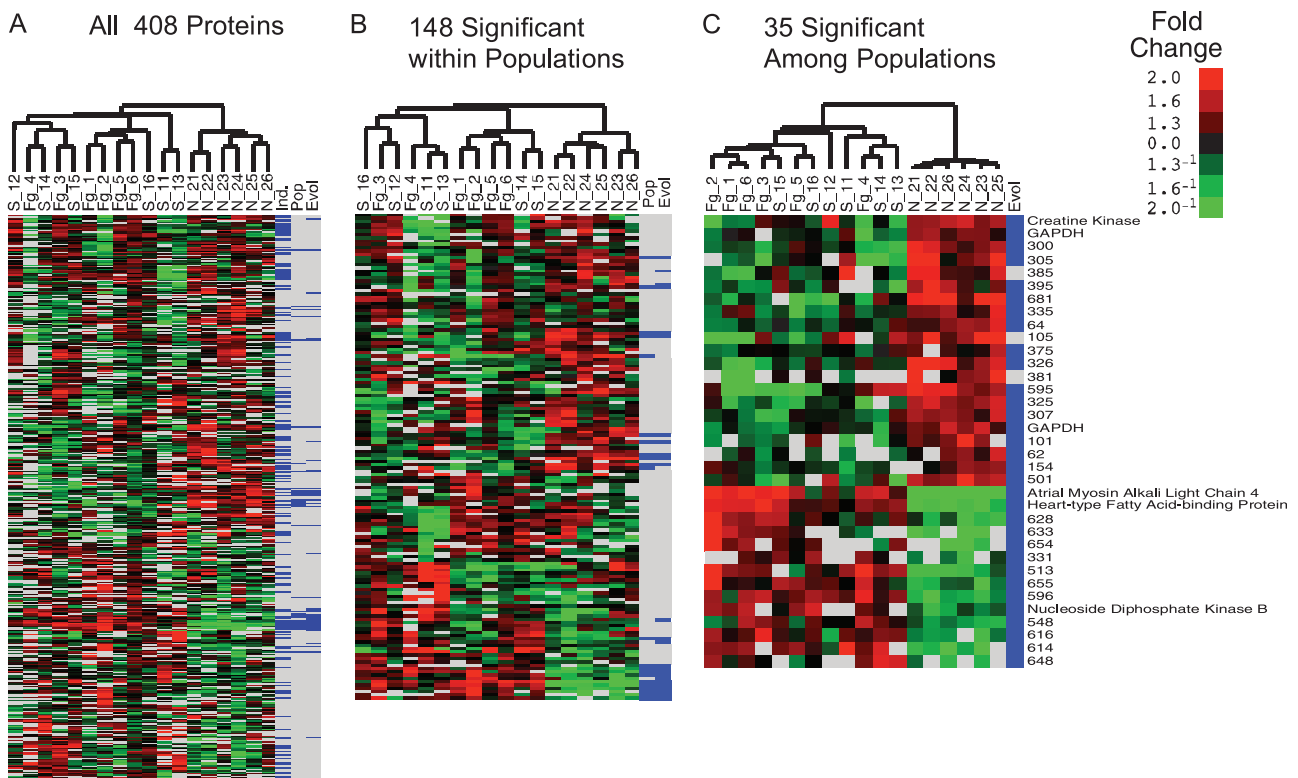


FIG. 2 Relative protein expression. Protein expressions (residuals from log body mass regressions) relative to the grand mean across all 18 individuals for each protein are shown. Bright red is 2-fold higher, and bright green is 2-fold lower than the grand mean for each protein. Gray cells indicate proteins not resolved for that individual. Letters and numbers define individuals (N and S are for northern and southern *F. heteroclitus* respectively, and Fg stands for *F. grandis*). (A) Patterns of protein expression for all 408 proteins. (B) Patterns of protein expression for 139 proteins that are significantly different ($P < 0.01$) among individuals within populations. (C) Patterns of protein expression for the 35 proteins that are significantly different ($P < 0.01$) among populations. Blue bars indicate proteins that are significantly different ($P < 0.01$) among individuals within populations (IND.), among populations (Pop), or for the evolutionary analysis (EvoI; where the variance in the northern population is compared with the joint variation in the southern *F. heteroclitus* and *F. grandis* populations). Annotations are provided only for C for clarity.

expression was determined in cardiac tissue among six fish from each of three *Fundulus* populations (northern and southern *F. heteroclitus* and sister taxa *F. grandis*) maintained in a common environment for greater than 6 months. Using 2D-DIGE, 408 proteins were quantified in a minimum of three individuals in each population (supplementary table 1, Supplementary Material online). The ratio of maximum to minimum protein expression among individuals ranged from 1.5-fold to 49.2-fold, with an average of 6.3-fold difference across all proteins. These data suggest large variation in protein expression among individuals. Our central questions are: how is this variation partitioned within and among populations and is it biologically meaningful.

Unlike mRNA expression (Oleksiak et al. 2002, 2005; Crawford and Oleksiak 2007), expression of many proteins is related to body size. Using a liberal critical P value of 0.10, protein expression significantly regresses with \log_{10} body mass for 32% of the proteins (132 of 408). Using more typical (0.05) or conservative (0.01) P values, 21% (84) or 6% (25) of proteins significantly regress with body mass. Consequently, to remove effects of body mass, differences in protein expression used residuals from these regressions.

Among individuals within populations, 34% (139/408) of proteins have significantly different expression ($P < 0.01$; fig. 2; supplementary table 1, Supplementary Material online). With a critical P value of 0.01, the number of false positives is only four proteins. Clearly, the determination of 139 proteins with significant differences in fluorescence ratio far exceeds the rate of false positives. Among the three populations (*F. heteroclitus* north and south and *F. grandis*), 35 of 408 proteins (9%) are significantly different among population with a critical P value of 0.01 (fig. 2C). This number of differences is 9-fold greater than the expected 1% false positive rates. Many proteins that differed significantly among the three populations also differed significantly among individuals within populations (22 of 35, fig. 2). Because the ANOVA had a maximum of 2 and 15 df, this result requires that the variance among populations be at least 6.3-fold greater than that within populations. For individuals raised in a common environment, this degree of variation in protein expression among different populations is surprisingly large.

Notice that the differences among the three populations described above include two populations of *F. heteroclitus* (northern and southern) and one population of *F. grandis*

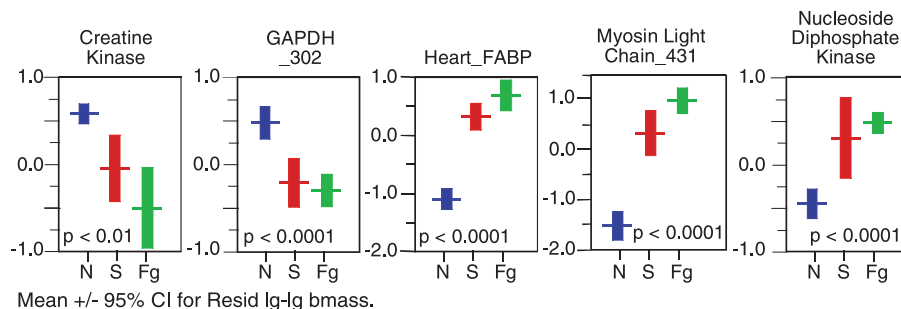


Fig. 3 Protein expression (residuals from body mass regressions) and associated P values for evolutionary analysis (H_1 : $N \neq S$ and Fg). N: northern *F. heteroclitus*, S: southern *F. heteroclitus*, and Fg: *F. grandis*. Mean \pm 95% confidence interval. GAPDH: glyceraldehyde-3 dehydrogenase and Heart_FABP: heart FABP. Number following annotation refers to specific protein spot.

(similar to original microarray publication; Oleksiak et al. 2002). Based on the neutral evolutionary theory, one expects populations within a species to be more similar to one another than either is to a congeneric species. Accordingly, we compared the protein expression in the two populations of *F. heteroclitus* with one another and, in a separate analysis, the protein expression in both *F. heteroclitus* populations with *F. grandis*. Between populations of *F. heteroclitus*, 31 proteins are significantly different, whereas between *F. heteroclitus* and *F. grandis*, only 12 proteins differ in expression (supplementary table 1, Supplementary Material online). These results contrast with neutral expectation where one expects populations within a species to be more similar to one another than either is to a congeneric species. Instead, these results are consistent with natural selection to different thermal environments where the southern *F. heteroclitus* and *F. grandis* populations occur in similar thermal habitats. If selection for their native habitats is important, then one might expect southern *F. heteroclitus* to be more similar to *F. grandis* than to northern *F. heteroclitus* (Pierce and Crawford 1997; Oleksiak et al. 2002; Crawford and Oleksiak 2007).

One can apply an evolutionary analyses based on the neutral hypothesis described above. For this evolutionary analysis, the variation within the northern *F. heteroclitus* population and within both southern taxa is compared with that among these two groups (i.e., northern *F. heteroclitus* vs. southern *F. heteroclitus* and *F. grandis*). The neutral expectation is that populations within a species will be more similar to each other than to populations in different species. Thus, the large joint variation across species will minimize the detection of difference in protein expression. The adaptive hypothesis is that southern taxa sharing similar native environments will be more similar than within the species. Thus, the variation among taxa sharing similar environments should be relatively small. This evolutionary analysis demonstrated that of 408 proteins, 47 (12%) differed significantly between northern *F. heteroclitus* and the two southern taxa ($P < 0.01$). There are fewer differences in expression between species (12 proteins) than between northern versus southern populations from both species (47 proteins). This similarity of protein expression measured in hearts of southern *F. heteroclitus* and *F. grandis*

accounts for the nonrandom clustering of protein intensities observed across all three data sets: Individuals from northern *F. heteroclitus* populations form a group distinct from southern *F. heteroclitus* and *F. grandis* (dendrogram at top of fig. 2A–2C). We suggest that these observations are most parsimoniously explained by evolution by natural selection.

We suggest that this evolutionary analysis is indicative of adaptive differences because the variation among *F. heteroclitus* populations is greater than that among species and more readily explained by native habitat temperature. These differences occur after all individuals were acclimated to common conditions for greater than 6 months and thus are not due to a reversible physiological response to temperature. We are tentatively assuming that these differences have a heritable basis and thus are of evolutionary importance based on the heritability of gene expression in other organisms and our previous analyses of evolved differences in gene expression (Pierce and Crawford 1997; Schadt et al. 2003; Whitney et al. 2003; Yan et al. 2003; Radich et al. 2004; Lemos et al. 2005; Crawford and Oleksiak 2007; Ayroles et al. 2009; Scott et al. 2009). Other explanations include irreversible developmental or epigenetic effects, and if operative, would reduce the frequency of evolved differences.

Among 30 full-length annotated proteins (excluding protein fragments; supplementary table 1, Supplementary Material online) identified by MS, there are at least 17 unique proteins (proteins with the same annotation that likely represent one or more loci; supplementary table 2, Supplementary Material online). Of the 30 annotated proteins, 7 differ ($P < 0.01$) in the evolutionary analysis. These 7 proteins correspond to five unique annotations of 17 uniquely annotated proteins (29%; fig. 3; heart-type fatty acid-binding protein (FABP), myosin light chain, nucleoside diphosphate kinase B, GAPDH, and creatine kinase). Two of these five proteins (GAPDH and myosin light chain) are represented by multiple protein spots. All three spots annotated as GAPDH have similar patterns of expression (northern *F. heteroclitus* > southern *F. heteroclitus* and *F. grandis*), but only two are significant ($P < 0.01$; supplementary table 1, Supplementary Material online). Two of the four protein spots annotated as myosin light chain have

a similar pattern of expression and are significant ($P < 0.01$). Not all proteins share the same patterns of expression among the three populations: For the first two proteins in [figure 3](#), northern *F. heteroclitus* have higher expression than southern *F. heteroclitus* and *F. grandis*, whereas for the other three proteins, expression in northern *F. heteroclitus* is lowest. These different patterns of expression argue against a systematic experimental bias producing these patterns.

Proteomic Compared With Microarrays

The experimental design used to quantify protein expression is similar to earlier mRNA analysis (Oleksiak et al. 2002). In that study of mRNA, data from five individuals from each of the same three populations indicated that 18% of nearly 1,000 genes had significantly different mRNA expression among individuals within populations (Oleksiak et al. 2002). The present study finds that nearly twice as many proteins (34%) vary significantly within populations even though there were fewer technical replicates and greater variation among replicates. However, caution should be taken interpreting this higher value because a protein from a single locus may be represented by multiple spots on 2D gels. If proteins with multiple spots tend to show greater variation in intensity among individuals, this would inflate the frequency of significance. To address this, we excluded protein fragments and normalized for the number of proteins with the same annotation: For each protein, we divided the number of significant differences by the total number of spots annotated as that protein ([supplementary table 1, Supplementary Material](#) online). For example, two of the three features on the 2D gels identified as full-length GAPDH are significantly different, yielding a normalized score of 0.667 (2/3). Using only the corrected normalized data, the frequency of significant differences among individuals is 71%. This seemingly large frequency is similar to the 68% of genes with significant and heritable differences in expression among inbred lines of *Drosophila melanogaster* (Ayroles et al. 2009). Importantly, both analyses of protein expression (34% of all 408 protein spots or 71% of 15 unique, full-length proteins) show a greater frequency of significant variation among individuals than does mRNA expression (18%) in the same populations (Oleksiak et al. 2002).

In previous analysis of mRNA expression (Oleksiak et al. 2002), 1.7% of genes differed significantly among all three populations and 3% were significant for evolutionary analysis comparing northern *F. heteroclitus* to two southern populations. The same comparisons for protein expression are higher, where 9% and 12% are significantly different among the three populations and for evolutionary analysis, respectively. However, this protein study used six individuals (not five as in mRNA study), and each protein spot was analyzed separately, ignoring that a single locus could be represented by more than one protein. To correct for the df, we examined all unique permutations of five out of six individuals. Among the approximately 100 proteins with five individuals measured in each population, on

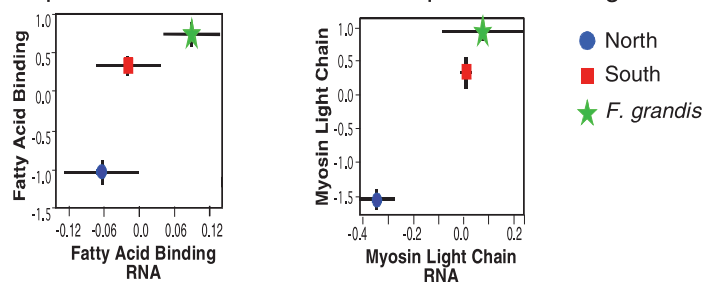
average, 12.7% and 17.7% were significantly different among the three populations or for the evolutionary analyses, respectively. To remove the effect of multiple proteins per locus, we used a weighted normalization (as above). For all permutations, an average of 27% of unique proteins (range: 18–33% among all permutations) is significantly different among populations. Similar analyses were done for the evolutionary analyses where the average frequency of significant variation in expression was 32% (range: 18–53%).

These analyses suggest many more significant differences among populations in protein expression compared with mRNA expression (Oleksiak et al. 2002), and this is not due to difference in sample size nor to multiple proteins per locus. Moreover, expression of many proteins is related to body mass, whereas mass did not affect mRNA expression (Oleksiak et al. 2002, 2005). These two observations, a greater frequency of significant differences among the three populations in protein versus mRNA expression and body mass effects on protein expression, suggest that the additional mechanisms regulating protein expression lead to larger differences among populations. If there is heritable variation in these mechanisms, then the regulation of protein expression could be a source for adaptive evolution

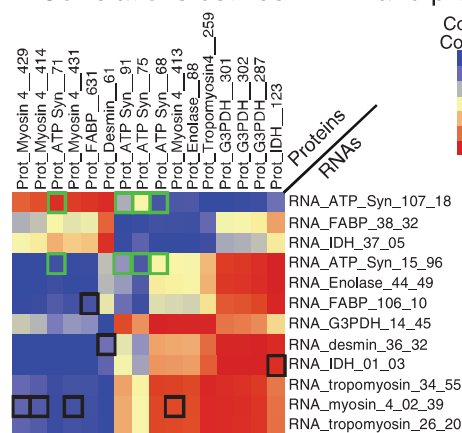
Correlation Between Specific Proteins and mRNAs

We also compared the mean population expression of specific proteins with expression of their corresponding mRNAs. To verify orthology, we used a series of BLAST comparisons for peptides, proteins, and cDNAs used to print the microarray ([supplementary table 2, Supplementary Material](#) online). We restricted our analyses to full-length or nearly full-length proteins ([supplementary table 1, Supplementary Material](#) online) because these peptides are more likely to be a function of mRNA abundance. Of the 15 unique full-length proteins, five proteins had orthologous mRNAs quantified in the earlier microarray study (Oleksiak et al. 2002), six proteins had paralogous mRNA measured (products of related but different loci), two proteins (actin and tropomyosin) had a matrix of similarities among homologs that made it difficult to define orthologs from paralogs, and two had no corresponding EST represented in the microarray study. Thus, of the 15 unique full-length proteins, 5 are orthologous to previously quantified mRNAs (Oleksiak et al. 2002): ATP synthase β subunit, desmin, FABP, isocitrate dehydrogenase-2 (IDH-2), and myosin light chain. For two proteins whose expression was significantly different among populations, FABP and myosin light chain ([fig. 4A](#)), the correlations were large and positive (0.82 and 0.85, respectively), whereas for three proteins that were not significantly different among populations, ATP synthase β , desmin, and IDH-2, the correlations ranged from -0.92 to 0.87 (boxes [fig. 4B](#)). Yet, among all annotated full-length proteins (excluding actin), most proteins have one form that has a large correlation ($\rho > 0.80$) with their orthologous mRNAs ([fig. 4B](#)). Notice, for the three populations, the correlation coefficient has to exceed 0.997 to be significant at a P value of 0.05. Although none of the orthologous pairs of protein and mRNA have correlations

A. Population measure of RNA and protein orthologs



B. Correlations between RNA and proteins



C. Correlations among proteins

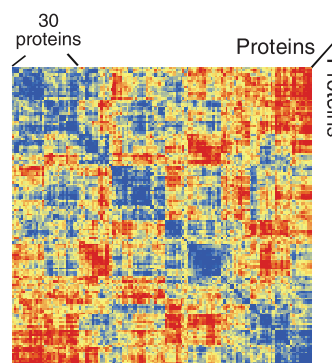


FIG. 4 Relationships between RNAs and proteins. (A) The three populations means for protein expression versus orthologous mRNA expression (Oleksiak et al. 2002) for proteins significantly different among populations plus and minus standard error. (B) Correlation matrix for proteins and RNA. Correlations are among the three populations. Boxes outline protein and mRNA orthologs. (C) Correlations among proteins with at least 17 individuals per protein. With a minimum of 15 df, R must be >0.6 to be significant ($P < 0.01$). Protein expression is the residual from \log_{10} body mass regression. The mRNA expression is normalized to a mean of 0.

that meet this criterion, these correlations support the hypothesis that mRNA expression affects protein abundance. However, there is complexity in these data. For example, there are eight correlations among four proteins and two mRNAs that are homologous to two ATP synthase β loci in *D. rerio* (supplementary Table 2, Supplementary Material online). These correlations range from -0.92 to 0.87 (supplementary table 3, Supplementary Material online). ATP synthase β spot 71 and cDNA 107–18 are most similar, yet the protein expression for spot 71 is negatively correlated with this most similar cDNA 107–18 (-0.92) and is positively correlated to a the less similar cDNA 15–96 (0.87). This, or any, negative correlation only makes molecular sense if the measures of mRNA and protein expression were from different loci. Yet for negative correlations for IDH-2 and ATP synthase β , the similarities among proteins and cDNAs do not clearly support this. We are left with a conundrum: Homology based on similarity suggests a different orthology than the patterns of expression between mRNA and protein.

Another interesting complexity is that expression levels of many mRNAs are correlated with many different proteins: 64 of the 150 (56%) mRNA–protein correlations have absolute values that exceed 0.80 (fig. 4B). With only three populations, the statistical support for these

correlations is weak. However, we have 139 measures of protein expression in at least 17 individuals, and correlations among proteins using all individuals have similar complexity (fig. 4C, supplementary table 4, Supplementary Material online): 8.7% have a statistically significant correlation ($P < 0.01$) with roughly the same number of positive and negative correlations. This pattern of positive and negative correlations for protein expression is similar to that of mRNA expression where many genes are positively or negatively correlated (Crawford and Oleksiak 2007). These patterns of correlated gene expression are not limited to *Fundulus*; among 40 strains of *Drosophila*, 68% of genes have significant heritable variation in mRNA expression with complex patterns of correlations (Ayroles et al. 2009).

Surprisingly, if one examines correlations among proteins or separately for mRNAs that form orthologous pairs, 33% of the proteins are significantly correlated to the other proteins ($p < 0.01$, correlation coefficients 0.63 – 0.93), but none of the mRNAs have a significant correlation (supplementary table 6, Supplementary Material online). Although speculative, if these correlations reflect meaningful coordinate regulation among genes, then the added regulatory mechanisms associated with protein expression enhance this coordination.

Conclusions

There are two important observations from these data: 1) 12% of proteins have evolutionary significant patterns of expression and 2) the variation in expression for proteins exceeds that observed for mRNAs.

Twelve percent of proteins that seem to have an evolutionary significant pattern of expression is larger than we would expect, given the potential cost of selection on all these loci (i.e., genetic load; Haldane 1957; Lewontin and Hubby 1966). Assuming that much of the variation in gene expression is heritable (Ayroles et al. 2009), the two most likely explanations for this are that 1) the genetic variation is not additive and 2) few loci control the expression of many loci. Clearly, the phenotypic effect of altered expression of one metabolic locus can be dependent on the expression of other loci (Segre et al. 2005). Additionally, in *F. heteroclitus*, the importance of expression of metabolic genes appears to be context dependent: The quantitative variation in expression of specific metabolic pathways depends on the background expression of other metabolic pathways (Oleksiak et al. 2005; Crawford and Oleksiak 2007). Additionally, the observation that 12% of loci have adaptive pattern of expression does not necessarily mean that each locus is evolving independently. The significant correlations among proteins suggest that fewer loci may control the expression of these genes, and thus, fewer loci may be evolving by natural selection. Thus, if the effect of altered expression of proteins has an epistatic component or if fewer loci control the expression of many, then there would be a much lower cost of selection.

In *Fundulus* cardiac tissue, both protein and mRNA measures of gene expression have a large variation among individuals that may be important for evolutionary adaptation. However, the frequency of significant differences in expression among the three populations as well as for the evolutionary analyses is larger for protein than for mRNA expression, and there is a higher frequency of significant correlations among proteins than among mRNAs. If the mechanisms affecting the enhanced differences in protein expression among populations are heritable, then these added regulatory mechanisms could enhance adaptive divergence. It is interesting to speculate that each added layer of biological organization (DNA → RNA → protein → networks → cell → organ systems → organisms) could enhance adaptive evolution.

Supplementary Material

Supplementary tables 1–6 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This research was supported by National Science Foundation Biocomplexity grant OCE OCE-0221879 and OCE-0308777. The authors would like to thank M. F. Oleksiak for assistance in experimental design, sample collection, data access, and critical review of this manuscript.

References

- Ayroles JF, Carbone MA, Stone EA, et al. (11 co-authors). 2009. Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet.* 41:299–307.
- Biron DG, Loxdale HD, Ponton F, Moura H, Marche L, Brugidou C, Thomas F. 2006. Population proteomics: an emerging discipline to study metapopulation ecology. *Proteomics* 6:1712–1715.
- Bray NJ, Buckland PR, Owen MJ, O'Donovan MC. 2003. Cis-acting variation in the expression of a high proportion of genes in human brain. *Hum Genet.* 113:149–153.
- Burnett KG, Bain LJ, Baldwin WS, et al. (26 co-authors). 2007. *Fundulus* as the premier teleost model in environmental biology: opportunities for new insights using genomics. *Comp Biochem Physiol Part D Genomics Proteomics.* 2:257–286.
- Crawford DL, Oleksiak MF. 2007. The biological importance of measuring individual variation. *J Exp Biol.* 210:1613–1621.
- Crawford DL, Powers DA. 1992. Evolutionary adaptation to different thermal environments via transcriptional regulation. *Mol Biol Evol.* 9:806–813.
- Doss S, Schadt EE, Drake TA, Lusis AJ. 2005. Cis-acting expression quantitative trait loci in mice. *Genome Res.* 15:681–691.
- Enard W, Khaitovich P, Klose J, et al. (13 co-authors). 2002. Intra- and interspecific variation in primate gene expression patterns. *Science* 296:340–343.
- Fay JC, Wittkopp PJ. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity.* 100:191–199.
- Gibson G. 2003. Population genomics: celebrating individual expression. *Heredity* 90:1–2.
- Haldane JBS. 1957. The cost of natural selection. *J Genet.* 55:511–524.
- Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G. 2001. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster* [see comment]. *Nat Genet.* 29:389–395.
- Karr TL. 2008. Application of proteomics to ecology and population biology. *Heredity* 100:200–206.
- Kerr MK, Churchill GA. 2001. Experimental design for gene expression microarrays. *Biostatistics* 2:183–201.
- Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685.
- Lemos B, Meiklejohn CD, Caceres M, Hartl DL. 2005. Rates of divergence in gene expression profiles of primates, mice and flies: stabilizing selection and variability among functional categories. *Evolution* 59:126–137.
- Lewontin RC, Hubby JL. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54:595–609.
- Nedelkov D. 2008. Population proteomics: investigation of protein diversity in human populations. *Proteomics* 8:779–786.
- Oleksiak MF, Churchill GA, Crawford DL. 2002. Variation in gene expression within and among natural populations. *Nat Genet.* 32:261–266.
- Oleksiak MF, Roach JL, Crawford DL. 2005. Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat Genet.* 37:67–72.
- Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, et al. (18 co-authors). 2006. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol.* 24:1140–1150.
- Pierce VA, Crawford DL. 1997. Phylogenetic analysis of glycolytic enzyme expression. *Science* 276:256–259.
- Podrabsky JE, Javillonar C, Hand Steven C, Crawford D. 2000. Intraspecific variation in aerobic metabolism and glycolytic enzyme expression in heart ventricles. *Am J Physiol.* 279: R2344–R2348.

- Powers DA, Smith M, Gonzalez-Villasenor I, DiMichelle L, Crawford DL, Bernardi G, Lauerma T. 1993. A multidisciplinary approach to the selectionist/neutralist controversy using the model teleost, *Fundulus heteroclitus*. In: Futuyma D, Antonovics J, editors. *Oxford surveys in evolutionary biology*. New York: Oxford University Press. p. 43–108.
- Pritchard CC, Hsu L, Delrow J, Nelson PS. 2001. Project normal: defining normal variance in mouse gene expression. *Proc Natl Acad Sci U S A*. 98:13266–13271.
- Radich JP, Mao M, Stepaniants S, et al. (11 co-authors). 2004. Individual-specific variation of gene expression in peripheral blood leukocytes. *Genomics* 83:980–988.
- Rockman MV, Wray GA. 2002. Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol*. 19:1991–2004.
- Schadt EE, Monks SA, Drake TA, et al. (14 co-authors). 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422:297–302.
- Scott CP, Williams DA, Crawford Douglas L. 2009. The effect of genetic and environmental variation on gene expression. *Mol Ecol*. 18:2832–2843.
- Segre D, DeLuna A, Church GM, Kishony R. 2005. Modular epistasis in yeast metabolism. *Nat Genet*. 37:77–83.
- Townsend JP, Cavalieri D, Hartl DL. 2003. Population genetic variation in genome-wide gene expression. *Mol Biol Evol*. 20:955–963.
- Whitehead A, Crawford DL. 2006a. Neutral and adaptive variation in gene expression. *Proc Natl Acad Sci U S A*. 103:5425–5430.
- Whitehead A, Crawford DL. 2006b. Variation within and among species in gene expression: raw material for evolution. *Mol Ecol*. 15:1197–1211.
- Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO. 2003. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci U S A*. 100:1896–1901.
- Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. *Nature* 430:85–88.
- Wittkopp PJ, Stewart EE, Arnold LL, Neidert AH, Haerum BK, Thompson EM, Akhras S, Smith-Winberry G, Shefner L. 2009. Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. *Science* 326:540–544.
- Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol*. 8:625–637.
- Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. 2003. Allelic variation in human gene expression. *Science* 297:1143.
- Yu X, Chu TM, Gibson G, Wolfinger RD. 2004. A mixed model approach to identify yeast transcriptional regulatory motifs via microarray experiments. *Stat Appl Genet Mol Biol*. 3: Article22.