2009-03-26

The Genetic Basis of Evolved Differences in Gene Expression in Fundulus heteroclitus

Cinda Pitts Scott
University of Miami, cscott@rsmas.miami.edu

UNIVERSITY OF MIAMI

THE GENETIC BASIS OF EVOLVED DIFFERENCES IN GENE EXPRESSION IN FUNDULUS HETEROCLITUS

By

Cinda Pitts Scott

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

May 2009

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

THE GENETIC BASIS OF EVOLVED DIFFERENCES IN GENE EXPRESSION IN FUNDULUS HETEROCLITUS

Cinda Pitts Scott

Approved:

Douglas L. Crawford, Ph.D.
Professor of Marine Biology
and Fisheries

Terri A. Scandura, Ph.D.
Dean of the Graduate School

Marjorie F. Oleksiak, Ph.D.
Assistant Professor of Marine Biology
and Fisheries

Gary Hitchcock, Ph.D.
Professor of Marine Biology
and Fisheries

Alexandra Worden, Ph.D.
Assistant Professor of Marine Biology
and Fisheries
Monterey Bay Aquarium Research Institute

Martin Grosell, Ph.D.
Associate Professor of Marine
Biology and Fisheries

SCOTT, CINDA PITTS (Ph.D., Marine Biology and Fisheries)
The Genetic Basis of Evolved Differences in (May 2009)
Gene Expression in Fundulus heteroclitus

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Douglas L. Crawford.
No. of pages in text. (115)

This dissertation explores the genetic basis of gene expression in *Fundulus heteroclitus* by focusing on the role of the environment and its effects on gene expression and by making direct estimates of heritability using cDNA microarrays. The second chapter describes the utility of *F. heteroclitus* cDNA microarrays for studies of *F. heteroclitus* which seek to understand the genetic variation in gene expression. Measurements of mRNA fluorescence and concentration as well as differences in sample preparation and sampling of blood from a single individual over time demonstrate that *F. heteroclitus* cDNA microarrays are quantitative, reproducible and consistent. The third chapter examines the effect of the environment and genetic factors on the variation of gene expression. *F. heteroclitus* cDNA microarrays are used to determine whether a genetic component of gene expression can describe the variation in gene expression between inbred and outbred individuals from the same population.

The results show that variation in mRNA expression is related to the genetic variation among individuals within a group. While chapter three reveals that there is a genetic component of variation in gene expression, the percentage of genes that are significantly heritable was not known. In the fourth chapter, the heritability of the variation in gene expression is estimated to determine the genetic basis of gene expression in F_1 individuals from natural, outbred populations of *F. heteroclitus*. The data presented in chapter 4 are the first to formally estimate the genetic component of gene expression in *F. heteroclitus*. The estimates of heritability range from 0.25 to 0.86 depending on the estimation method with approximately 6.5% of genes having significant heritability. The results presented in this dissertation support the concept that genetic variation affects variation in mRNA expression among natural populations of *F. heteroclitus*. Natural, heritable variation in gene expression is important for understanding evolutionary adaptation and the role of natural selection in evolutionary processes.

For my family.

ACKNOWLEDGEMENTS

I must first extend a sincere thank you to my advisor, Douglas Crawford, for extending his hand to me at a time when I was unsure whether continuing my graduate education was the best option for me. Doug, you provided me with a place to learn and to cultivate ideas and most importantly you had faith in me and my abilities from the very beginning. I thank you for introducing me to evolutionary genomics, for all of your advice and most importantly for remaining passionate about your work and bringing out that passion in me. Thank you to Marjorie Oleksiak for your understanding and guidance. You inspired me to focus and work hard when I was uncertain of my true capabilities. I also thank Martin Grosell for always listening and for being a source of support and guidance. I am also grateful to Alexandra Worden who provided much wisdom and insight and for instilling confidence in me when it was lacking. Thank you to Gary Hitchcock who guided me through tough terrain and who has been extremely supportive throughout this entire experience. I am also thankful for Lora Fleming who has also instilled in me greater confidence and who continues to inspire me everyday. I also thank Danielle McDonald whose support and advice I will always remember. Thank you to David Letson for your guidance and for being an extra source of support. Last, but certainly not least, thank you to Bob Cowen for always having faith in me and for all of your support.

I would also like to acknowledge all of the students and faculty who have made a significant impact on me both as a scholar and a person. The members of the Oleksiak and Crawford labs both past and present have pushed me to do my best, work hard and have fun. Special thanks to Larissa Williams and Meredith Everett who are always there to lend a helping hand and who are definitely wise beyond their years. I also thank Dean Williams, who has the patience of a saint, for his never ending support and guidance and for teaching me how to genotype. I especially thank Kristin Kleisner and Leopoldo Llinas for their friendship, love and support and for making my experience in Miami memorable and for always making me smile.

Most importantly, I thank my family who love me unconditionally and who have been there every step of the way, picking me up when I was down and cheering me all the way across the finish line. I love you dearly, you mean the world to me and none of this would have been possible without your love.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

CHAPTER 1: AN INTRODUCTION TO THE GENETICS OF GENE EXPRESSION IN FUNDULUS HETEROCLITUS

Background

Evolution by natural selection requires heritable variation among traits that affect fitness. One important source of variation is the difference among individuals in gene expression. A full understanding of the role of differential gene expression among individuals within populations, and the ability of members of a population to adapt to changing environments, remains incomplete. Variation in gene expression in the teleost fish *Fundulus heteroclitus* from Northern and Southern populations explains differences in cardiac metabolism (Oleksiak *et al.*, 2005), and many genes have expression patterns indicative of evolution by natural selection (Oleksiak *et al.*, 2002; Pierce, Crawford, 1997). Previous studies of *F. heteroclitus* assumed that a majority of this biologically important variation is genetically based. However, only common gardened adults were used in these studies, therefore the assumption that variation in mRNA expression is genetically based is unsubstantiated and needs further examination. The research presented in this dissertation examines the genetic basis of differences in gene expression in *Fundulus heteroclitus*.

Functional Genetic Variation in Fundulus heteroclitus

F. heteroclitus provides a model system to investigate the evolutionary importance of gene expression. This species is found in estuaries, bays and inlets along the Atlantic coastline of the United States from Maine to Georgia (Wiley, 1986). *F. heteroclitus* resides in water temperatures ranging from 8.5ºC in the North to 20.4ºC in the South. Therefore, this species is exposed to a steep thermal cline with a 1ºC change in temperature per degree latitude (Pierce, Crawford, 1997; Powers *et al.*, 1993). *F.*

heteroclitus is a poikilotherm and is strongly influenced by its surrounding environment. One of the most important environmental variables affecting this species is temperature as is evidenced by studies showing differential measures of specific metabolic enzyme activity from Northern and Southern populations (Pierce, Crawford, 1997; Powers *et al.*, 1991). Based on enzymatic studies and protein sequence analyses, it was hypothesized that populations of *F. heteroclitus* ranging from Maine to Georgia are genetically divergent.

Fundulus heteroclitus is highly polymorphic at a number of enzyme encoding loci. An examination of 50 loci from fish collected along the Atlantic Coast found that 45% of these loci were polymorphic (Powers *et al.*, 1993). Heritability studies have shown that 16 out of 17 polymorphic loci in *F. heteroclitus* F1 generations segregate as autosomally inherited codominant alleles (Powers *et al.*, 1993). Some of these polymorphic loci have allelic isozymes that change in frequency with latitude and type of aquatic environment. Therefore, the degree of genetic diversity in a particular population can be attributed to latitude and clinal variation (Powers *et al.*, 1991). Why are alleles more or less prevalent at different latitudes? To better understand the relationship between directional changes in genetic characters and latitude, analysis of mitochondrial DNA fragments was completed on coastal and bay populations of *F. heteroclitus*.

A cline in gene frequencies exists between Northern and Southern populations of *F. heteroclitus.* Where this cline divides Northern and Southern populations of *F. heteroclitus* was determined using mitochondrial restriction fragment length polymorphisms (RFLPs). Based on mitochondrial fragment mobility on electrophoretic gels, it was shown that populations from Maine to Vince Lombardi, New Jersey share mitochondrial DNA sequences and populations from Stone Harbor, New Jersey to Georgia share mitochondrial DNA sequences (Powers *et al.*, 1993). It was suggested that the discontinuity of gene frequencies between Northern and Southern populations was consistent with the last glacial maximum that began 15,000 years ago near the Hudson River (Cronin *et al.*, 1981; Powers *et al.*, 1993). However, nucleotide differences in mitochondrial haplotypes suggested that the Northern and Southern populations diverged before this last glacial event. A study using five restriction enzymes on mitochondrial DNA sequences from 740 individuals collected from 29 locations along the Chesapeke and Delaware bays was completed to determine whether Northern mitochondrial DNA haplotypes could be detected as remnants of a preglaciation distribution (Avise, 1989; Powers *et al.*, 1993). Fifty mitochondrial DNA haplotypes were found with Northern, Southern and intermediate groupings. These findings suggest that an intergrade zone exits at 41ºN latitude, and supports the idea that populations adapted to their local environments which led to genetic differences along a gradient (Powers *et al.*, 1993).

Genetic divergence in *F. heteroclitus* is also attributed to biochemical differences which ultimately affect changes in physiological processes and gene expression. Biochemical analyses of lactate dehydrogenase-B_4 (LDH- B_4) in populations of *F. heteroclitus* demonstrated that functional differences exist between allelic isozymes of *LDH-* B_4 which are genetically divergent between Northern and Southern populations (Powers *et al.*, 1991). *LDH-* B_4^a *and LDH-* B_4^b are codominant allelic isozymes of LDH-B_4 which is primarily responsible for converting lactate to pyruvate during aerobic respiration for the production of ATP and in gluconeogenesis (Powers *et al.*, 1991). The *LDH-* B_4^b allele is predominant in the North and the *LDH-* B_4^a allele is predominant in the South while a mixture of both alleles is found at mid-latitudes (41ºN) (Powers *et al.*, 1991; Powers *et al.*, 1993).

The alleles vary with latitude and functional differences exist between populations in the North which express *LDH-* B_4^b and populations in the South which express the *LDH-* B_4^a allele. These include differences in catalytic efficiency, substrate and product inhibition, stability of LDH-B_4 allelic isozymes, enzyme concentration, structure, swimming endurance, hatching time, development and survivial (Powers *et al.*, 1993). For example, the *LDH-* B_4^a allele has better efficiency at high temperatures than low temperatures whereas the *LDH-* B_4^b allele has better efficiency at low temperatures than high which is consistent with their geographic distribution. Heat denaturation studies showed that the structural stability of *LDH-* B_4^b is greater than *LDH-* B_4^a (Powers *et al.*, 1991). LDH- B_4 enzyme concentration was found to be two times as high in Maine *F. heteroclitus* liver than in Georgia livers even after fish were acclimated to the same temperatures in the laboratory (Powers *et al.*, 1991). Maine and Georgia fish have the same number of gene copies, but they were found to have different concentrations of messenger RNA (Powers *et al.*, 1993). Therefore, rates of LDH- B_4 transcription were measured and it was found that Northern *F. heteroclitus* have increased rates of transcription than Southern populations of *F. heteroclitus*. However, there was no reported difference in the overall total rate of transcription (Powers *et al.*, 1993). Increased rates of transcription in the Northern population suggest that these fish elevate rates of transcription to compensate in colder waters. Is the difference between Northern and Southern populations genetically driven? Is the transcriptional regulation of particular genes a more important causation of variation than variation in enzymatic processes?

Variation in gene expression among *Fundulus* species is hypothesized to be evolutionarily important (Oleksiak *et al.*, 2005; Pierce, Crawford, 1997; Powers *et al.*, 1993). It is apparent that populations of *F. heteroclitus* residing in different thermal regimes compensate in part by regulating their rates of transcription. Therefore, the variation in the concentration of an enzyme may be selectively important (Pierce, Crawford, 1997). A study examining the influence of temperature on enzyme concentration for 15 *Fundulus* taxa revealed that among eleven enzymes (lactate dehydrogenase and ten glycolytic enzymes), three enzymes (lactate dehydrogenase (LDH), glyceraldehydes-3-phosphoglycerokinase (GAPDH) and pyruvate kinase(PYK)) had concentrations that correlated negatively with temperature after phylogenetic effects were removed. The most parsimonious reason for this pattern among 15 *Fundulus* taxa, is that evolution by natural selection affects enzyme concentration. As temperature increased, the concentration of LDH, GAPDH and PYK enzymes decreased suggesting that enzyme concentration has evolved in response to temperature for these three enzymes, that the existing variation between other enzymes is not a direct adaptation to temperature, and most importantly, these three enzymes affect metabolic flux (Pierce, Crawford, 1997). Specifically, these three enzymes explain much of the variation in glycolytic flux within and between *F. heteroclitus* populations (Podrabsky *et al.*, 2000).

For one of the loci (*Ldh-B*) the adaptive difference in gene expression is due to nucleotide variation in the proximal promoter (Crawford *et al.*, 1999). All fixed differences in proximal promoter sequence between populations of *F. heteroclitus* were found to reside in the functional regions (regions which affect transcription and bind transcription factors) of the *Ldh-B* proximal promoter (Crawford *et al.*, 1999). The number of fixed differences was found to be higher than the expected number for neutral evolutionary processes. Therefore, variation in *Ldh-B* proximal promoter sequences is best explained by natural selection (Crawford *et al.*, 1999). Indeed, transcriptional regulation appears to play a major role in the variation of enzyme function. Since enzyme function can be the result of mRNA expression, the focus of several studies has shifted from protein and biochemical analyses to the essential initial steps of transcription via global gene expression profiling.

Microarrays

Microarrays provide a powerful means for the simultaneous examination of expression for thousands of genes. DNA microarrays are classified into two types; 1) oligonucleotide arrays and 2) cDNA based microarrays. Oligonucleotide microarrays consist of short or long, nucleotide probes that range from 25-150 nucleotides or longer (Antipova *et al.*, 2002; Lettieri, 2006; Li *et al.*, 2002). These probes are synthesized rather than cloned. Unlike oligonucleotide probes which are synthesized, cDNA microarrays are created by printing cDNA clones onto glass slides after amplification by PCR. Oligonucleotide microarrays are often the platform of choice because they avoid problems such as cross-hybridization between mRNAs which may lead to false positive signals and inaccurate measures of changes in gene expression (Li *et al.*, 2002). Oligonucleotide microarrays containing thousands of genes can be purchased or custom made for model organisms whose genome sequences are complete (Affymetrix, Inc.). However, cDNA microarrays are the method of choice for many laboratories whose research involves the use of non-model organisms with incomplete genome sequences because it is the most cost-effective approach due to the relatively low cost of robotic spotting (Auburn *et al.*, 2005; Gracey *et al.*, 2001; Lettieri, 2006; Renn *et al.*, 2004; Williams *et al.*, 2003).

The decision to use a cDNA microarray platform rather than oligonucleotide platforms for *Funudlus heteroclitus* is based on the lack of availability of a fully sequenced *F. heteroclitus* genome, the cost of synthesizing oligonucleotide arrays versus creating cDNA arrays in our own laboratory and the ease of isolation, normalization and sequencing of cDNAs in our laboratory (Oleksiak *et al.*, 2001). In addition, the questions addressed in this dissertation require the use of hundreds of individuals using an experimental design that allows for the simultaneous monitoring of hundreds of genes. Without the use of microarrays, the depth of gene expression analyses would be limited to one or a few genes using Northern blots, quantitative PCR, Serial Analysis of Gene Expression (SAGE) and Massively Parallel Signature Sequencing (MPSS) (Auburn *et al.*, 2005; Brenner *et al.*, 2000; Velculescu *et al.*, 2000). New technologies such as 454 pyrosequencing to assemble transcriptomes of various organisms have been successfully applied to both non-model and model species (Andreas *et al.*, 2007; Moore *et al.*, 2006; Vera *et al.*, 2008; Wicker *et al.*, 2006). While this technology is accessible to non-model organisms, such as the Glanville fritillary butterfly, *Melitaea cinxia*, performing 454 sequencing on *F. heterclitus* at this juncture would only serve to enrich the coverage and numbers of genes on our microarrays rather than address the specific goals outlined in this dissertation.

The construction of cDNA microarrays is accomplished by the covalent bonding of cDNAs onto microscope slides in a specific manner. cDNAs are isolated and sequenced using cDNA libraries followed by normalization of the cDNA library. Colonies from *F. heteroclitus* cDNA libraries are then randomly picked, amplified by PCR and sequenced (Oleksiak *et al.*, 2001). These sequences are then subtracted from the normalized library and the process is repeated until there is a desired number of cDNAs or genes. The length of each cDNA printed on a slide is approximately 1.5 kb. Microarray studies in *F. heteroclitus* have utilized a variety of microarrays including heart ventricle and metabolic arrays (Oleksiak *et al.*, 2002; Oleksiak *et al.*, 2005; Whitehead, Crawford, 2006a). In these studies, RNA isolated from *F. heteroclitus* individuals is amplified and labeled with either Cy3 or Cy5 fluorescent dye. Upon hybridization to the microarray, the quantification of thousands of genes (specific to metabolism or to other processes) can be determined. Microarray analyses have provided evidence in *F. heteroclitus* that patterns of gene expression are associated with differential use of particular metabolic substrates which ultimately affect performance (Oleksiak *et al.*, 2005).

Recent microarray studies have focused on understanding differences in gene expression among populations of *F. heteroclitus*. Variation in gene expression explains the variation in cardiac metabolism in different groups of individuals (Oleksiak *et al.*, 2005). Using microarrays with 192 cDNAs from a *F. heteroclitus* cDNA library encoding proteins for cellular metabolism, 84% (112/119 genes included in analysis; $p<0.01$, range 104-117) and 79% (94/119 genes included in analysis; $p<0.01$, range 82-106) of genes were differentially expressed in individuals from Northern and Southern populations, respectively. Although this is a large amount of variation, the results showed that differential expression of genes resulted in the use of the same substrate for individuals within a group to carry out metabolism. Therefore, different combinations of genes explain substrate-specific metabolism suggesting that gene expression is biologically meaningful (Oleksiak *et al.*, 2005). The divergence in variation in gene expression between Northern and Southern populations of *Fundulus* is perhaps more important than variation in protein sequence because it can describe the evolutionary forces which act to cause changes in metabolic processes.

Many studies have shown that microarrays are equivalent in quantitative ability to northern blot and quantitative PCR techniques (Auburn *et al.*, 2005; Ding *et al.*, 2007; Draghici *et al.*, 2006; Schena *et al.*, 1995). Studies have shown the utility of microarrays for quantifying and examining gene expression. Microarray quantification is best realized if there is a linear relationship between fluorescence of the dye labeled sample and the RNA concentration of the sample. For *Fundulus* arrays, 100 fold variation in the concentration of mRNA provides a linear signal for a vast majority of genes (92.9% or 197/212 genes) (Scott *et al.*, 2009). Microarrays are an integral part of studies of heritability and will be used for the purposes of the work completed in this dissertation.

Except for studies that identify DNA changes responsible for the phenotypic changes in gene expression or other physiological processes, the genetic basis for variation in gene expression is not well studied. However, the use of microarrays in conjunction with studies of heritability can provide evidence for genetic variation in gene expression (Jansen, Nap, 2001; Li, Burmeister, 2005). This dissertation seeks to address the genetic basis for evolved differences in gene expression in *F. heteroclitus* by investigating the heritability of gene expression with the use of microarrays.

Heritability of Gene Expression

Measures of the heritability of gene expression have been documented in *Drosophila*, yeast and humans (Brem, Kruglyak, 2005; de Koning, Haley, 2005; Monks *et al.*, 2004; Wayne *et al.*, 2004). Among humans (unrelated, siblings and monozygotic twins) there is greater similarity in gene expression between monozygotic twins than between siblings or unrelated individuals (Cheung *et al.*, 2003a; Stamatoyannopoulos, 2004). For humans, twin-studies (Sharma *et al.*, 2005; Tan *et al.*, 2005) and replicate measures of the same individuals over time (Cobb *et al.*, 2005; Eady *et al.*, 2005; Radich *et al.*, 2004; Whitney *et al.*, 2003) suggest that there is a strong genetic component to the natural variation in mRNA expression. Therefore, phenotypic expression in humans is familial and thus genetic.

Comparisons of gene expression between parents and offspring in lymphoblastoid cell lines for fifteen families of the Centre d'Etude du Polymorphisme Humain (CEPH) found that 31% (762 genes/2,430 genes) of differentially expressed genes were significantly heritable with a median of 0.34 (Monks *et al.*, 2004). Whole body transcript levels of 10 heterozygous *Drosophila similans* cross progeny males revealed that 8% (663 genes/7886 genes) had significant genetic variation with an h^2 median of 0.47(Wayne *et al.*, 2004). In studies of yeast, individual strains of *Saccharomyces cerevisiae* revealed that in 1528 out of 6215 genes, 84% were highly heritable indicating that variation is genetic in this system (Brem *et al.*, 2002). In another study investigating 112 segregants of *S. cerevisiae*, 3,546 of 5,727genes showed high heritability with h^2 values higher than 69% (Brem, Kruglyak, 2005). These studies of heritability show that for each system heritability can be readily measured.

In *F. heteroclitus* approximately 18% (161 of 907 genes) of gene expression is statistically different between individuals from the same population which is concordant with other studies in yeast (24%) and *Drosophila* (25%) (Brem *et al.*, 2002; Jin *et al.*, 2001; Oleksiak *et al.*, 2002; Stamatoyannopoulos, 2004). Studies of *F. heteroclitus* focusing on differential expression of cardiac metabolic genes revealed that an astounding 94% (112 of 119 genes) of gene expression is significantly different among individuals (Oleksiak *et al.*, 2005). Although there was large variation among individuals, these differences in gene expression were related to variation in cardiac metabolism. For example, 84% of the variation in cardiac glucose utilization could be explained by patterns of gene expression (Oleksiak *et al.*, 2005). Whether this variation is primarily due to genetic or environmental factors will help answer some fundamental questions about the purpose of such high levels of variation. For example, how much of this variation has an affect on cardiac metabolism? And, do genes whose variation in expression is highly heritable determine cardiac metabolic output more so than the expression of genes that are not as heritable?

Although there have been several studies suggesting the evolutionary significance of adaptation and gene expression in *Fundulus* (Crawford *et al.*, 1999; Oleksiak *et al.*, 2005; Pierce, Crawford, 1997; Powers *et al.*, 1991; Powers *et al.*, 1993), to date, the heritability of gene expression phenotypes in *F. heteroclitus* has not been examined. Heritability and genetic studies in *F. heteroclitus* are important for understanding how many loci contribute to natural genetic variation. My dissertation seeks to define narrow sense heritability for gene expression. Narrow sense heritability (h^2) is the amount of phenotypic variation that is explained by additive genetic variation. Heritability will be measured by regressing offspring values of gene expression versus mid-parent values (Cheung *et al.*, 2003b; Falconer, Mackay, 1996). Knowing the heritability of gene expression provides an important means for examining the genetics of gene expression.

Though not explored in this dissertation, expressed quantitative trait loci (eQTL), the loci which control differences in gene expression (de Koning, Haley, 2005), are important to discuss as they represent genomic regions for the genetic control of gene expression (Petretto *et al.*, 2006). eQTL studies are used to determine whether the location of a transcript within the genome in comparison to the location of an eQTL is *cis* or *trans*-acting (de Koning, Haley, 2005; Petretto *et al.*, 2006). The amount of *cis*-acting versus *trans*-acting variation is of recent interest as differences in eQTL position (*cis* or *trans*) are thought to contribute substantially to differences in gene expression and transcript levels.

eQTL studies are carried out using linkage analysis whereby the loci thought to control transcript levels of particular genes are mapped. Variation is *cis*-acting if DNA variations in a gene directly influence transcript levels of that gene and *trans*-acting if a gene is influenced by other genetic variations (Doss *et al.*, 2005). eQTL studies in yeast, mice and humans suggest that both *cis* and *trans*-acting regulators affect patterns of gene expression. In liver tissue gene expression studies in mice, for example, *cis*-acting eQTL with LOD scores greater than 4.3 account for 25% of the variation in transcription (Schadt *et al.*, 2003). The percentage increases to approximately 50% for eQTL with LOD scores greater than 7.0 (Schadt *et al.*, 2003). High LOD scores indicate tight linkage between a marker and an eQTL. In a follow-up analysis to verify that these *cis*-acting eQTL were true positives, it was confirmed that approximately 64% of the *cis*-acting eQTL with LOD scores greater than 4.3 were indeed true positives (Doss *et al.*, 2005). In general, it is thought that up to one-third of eQTLs are *cis*-acting (Gibson, Weir, 2005).

In a study monitoring *trans*-acting regulatory variation in *S. cerevisiae*, it was found that of 2,294 differentially expressed genes linked to locations within the genome and 75% of those genes were not found to show self-linkage (Yvert *et al.*, 2003). This indicates that the majority of loci are *trans*-acting and that these regulators are responsible for the majority of gene expression variation (Yvert *et al.*, 2003). What is the contribution of *cis* or *trans* regulation to genetic variation in *F. heteroclitus*? This proposed research will investigate the percentage of genes in *F. heteroclitus* that are regulated by *cis*-acting mechanisms. However, differences in gene expression may not solely be attributed to either *cis* or *trans* acting regulators. Variation must be viewed as complex interactions between many different loci some of which are *cis*-regulated and some of which are *trans*-regulated (Chesler *et al.*, 2005). Knowing what type of regulation occurs at particular loci will uncover whether adaptive differences between populations of *F. heteroclitus* are genetically controlled.

The concept of eQTL and linkage analysis can be extended to *F. heteroclitus* in the future by measuring the association between genetic markers (such as single nucleotide polymorphisms (SNPs) and microsatellites) and microarray gene expression phenotypes. Analysis of the linkage between molecular markers and quantitative traits will reveal polymorphisms among individuals which can explain the genetic basis of variation. Linkage between genetic markers and quantitative traits in *F. heteroclitus* are expected to be similar to studies done in yeast, mice, rats and humans(de Koning, Haley, 2005; Hubner *et al.*, 2005; Schadt *et al.*, 2003). eQTL studies in rats identified approximately 1000 eQTL associated with the regulation in gene expression in fat and kidney tissue which are important for understanding the inheritance of metabolic syndrome (Hubner *et al.*, 2005). Furthermore, in human studies, CEPH pedigree studies found that for several *cis*-acting eQTLs identified by linkage mapping between parents and offspring, promoter polymorphisms which affect transcriptional regulation showed association with differences in transcript abundance (Gibson, Weir, 2005; Morley *et al.*, 2004). The association of polymorphisms within regulatory regions and the degree of transcript abundance will provide evidence that particular genes or groups of genes are responsible for a given phenotype. Upon the sequencing of the *F. heteroclitus* genome and the development of inbred lines in the future, eQTL studies in *F. heteroclitus* will provide insight into the underlying genetic causes of variation in gene expression and evolved differences in physiological functions.

Objectives

The specific aims of my dissertation are as follows:

1. To determine the quantitative capacity, technical and biological variation of *Fundulus heteroclitus* cDNA microarrays.
2. To test the hypothesis that there is a genetic basis for the inter-individual variation in mRNA expression by studying the effect of environmental variation on gene expression.

3. To ascertain the heritability of gene expression by parent-offspring analyses using 13 families from NJ and *Fundulus heteroclitus* cDNA microarrays.

Significance

Each aim is discussed in the three subsequent chapters. Chapter 2 discusses the importance of microarrays as tools for understanding the expression of hundreds to thousands of genes at one time. In this chapter, I analyze the relationship between the amount of mRNA hybridized to a microarray and the resultant fluorescent signal. These hybridizations demonstrate that for a wide range of mRNA concentrations, the fluorescent signal is a linear function of the amount of mRNA. Additionally, the separate isolation, labeling or hybridization of RNA does not add significant amounts of variation in measures of gene expression using *F. heteroclitus* cDNA microarrays. However, single or double rounds of RNA amplification for amino allyl labeling do have small but significant affects on 10% of genes, but this source of technical variation is easy to avoid by using either one or two rounds of amplification, but not both in the same experiment. To examine both technical and stochastic biological variation, mRNA expression was measured from the same five individuals over a six-week time course. I concluded that there were few, if any, meaningful differences in gene expression among time points. Thus, microarray measures using standard laboratory procedures can be precise and quantitative and are not subject to significant random biological noise.

The hypothesis that variation in gene expression is a function of genetic variation is explored in Chapter 3 by altering both environmental and genetic backgrounds of *F. heteroclitus*. The effect of genetic and environmental variation on cardiac mRNA expression was examined using microarrays for three groups of *Fundulus heteroclitus* from the same North Carolina population: (1) individuals sampled in the field (field), (2) field individuals acclimated for six months to common laboratory conditions (acclimated) or (3) individuals bred for ten successive generations in a common laboratory environment (G10). Microsatellite analyses indicated that G10 individuals have significantly less genetic variation than individuals obtained in the field. Similar to the genetic variation, G10 individuals had a significantly lower variation in mRNA expression across all genes in comparison to the other two groups ($p \leq 0.001$).

When examining the gene specific variation among individuals, twenty-two of 284 genes (7.7%) had significant differences in the variation in expression among groups. Of these, seventeen (75%) have lower variation in G10 individuals than in acclimated individuals and this is unlikely to occur by chance (χ^2 $p < 0.01$). Additionally, there were fewer genes with significant differences in expression in G10 versus either acclimated or field individuals: 66 genes have statistically different levels of expression *versus* 107 or 97 for acclimated or field groups, respectively (23% vs. 38%, 34%). 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 ($p <0.01$). Although, many genes differ among individuals 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 among these natural populations.

Chapter 4 describes the approaches used to determine whether gene expression is related to the genetic variation among individuals. Microarray analyses were combined with genetics approaches to estimate heritability. Two methods were used for estimating heritability; 1) Regression analyses and 2) Components of variance analyses. The regression analysis provides an estimation of narrow sense heritability and the components of variance analysis provide an estimation of broad sense heritability. Though the two methods differ in their ability to dissect genetic variance, they both nonetheless are an important means for genetic determination of mRNA expression.

To estimate narrow sense heritability, parent-offspring regressions were performed where the mean gene expression values of the male and female parents (mid-parent gene expression values), the female parent expression values or the male parent expression values were regressed on the gene expression values of the F1 offspring. Broad sense heritability was estimated by calculating the components of variance for sib-sib analyses using gene expression values from 13 F1 offspring families. For the narrow sense heritability estimates, approximately 6.5%, 2.67% and 6.12% of genes had a significant heritability at $p \leq 0.05$ for the mid-parent, female and male on offspring regressions, respectively. For the mid-parent, female and male on offspring regressions, the median h^2 was .861, .729 and .875, respectively. The broad sense heritability analysis found that 8.86% of genes had significant heritability at $p \leq 0.05$ with a median H^2 of 0.25. Though the approaches for estimating heritability differ, the results confirm that microarrays are a useful and effective tool for determining the genetic basis of gene expression. In addition, these estimates of heritability provide the first data indicating that there is a genetic basis of gene expression in *F. heteroclitus*.

Summary

The dissertation concludes with a summary of the findings throughout each chapter. Microarrays are useful for understanding differential gene expression which in turn is important for the methods used to estimate heritability. The integration of these three chapters provides the first data suggesting that microarrays can be used to successfully determine the genetic basis of gene expression in *Fundulus heteroclitus.* In addition to these findings, recommendations for future research are discussed.

CHAPTER 2: TECHNICAL ANALYSIS OF cDNA MICROARRAYS

Background

Microarrays simultaneously quantify several hundred to thousands of genes on a single glass slide and their use has greatly expanded the breadth of quantified gene expression (Brem *et al.*, 2002; de Koning, Haley, 2005; Enard *et al.*, 2002; Gibson *et al.*, 2004; Monks *et al.*, 2004; Oleksiak *et al.*, 2002; Oleksiak *et al.*, 2005; Schadt *et al.*, 2003; Townsend *et al.*, 2003; Yvert *et al.*, 2003). Yet the preparation of RNA affects the precision of microarray measures and therefore the ability to accurately quantify the content of an RNA sample (Baugh *et al.*, 2001). Additionally, differences in microarray platforms, laboratory procedures and post-quantification analyses affect the precision among arrays (Bloom *et al.*, 2004; Irizarry *et al.*, 2005; Larkin *et al.*, 2005; Quackenbush, 2002). Thus, technical variation can substantially affect the interpretation of microarrays.

For the teleost fish *Fundulus heteroclitus* variation among individuals in mRNA expression is extensive: > 60% of genes have significant differences in expression among individuals within a population (Oleksiak *et al.*, 2002; Oleksiak *et al.*, 2005; Whitehead, Crawford, 2006a; Whitehead, Crawford, 2006b). Many of these differences in gene expression are associated with variation in cardiac metabolism (Oleksiak *et al.*, 2005). However, the accuracy and biological relevance of these differences in expression depends on the technical variation inherent to microarray processing (Oleksiak *et al.*, 2002).

Accurate microarray quantification is best realized when there is a linear relationship between fluorescence and RNA concentration. This linear relationship fails when the dynamic range of microarrays are exceeded. For any microarray, there are two parameters that define its dynamic range: the range of fluorescence that can be measured and the range of RNA concentrations that can bind to a specific array feature. These two components of the dynamic range reflect the two types of saturation that can occur on a microarray: photomultiplier tube (PMT) saturation and biological saturation. A linear relationship between fluorescence and RNA concentration can only occur if the cDNA on the microarray captures proportional amounts of RNA and if the PMT is not saturated. The PMT measures the number of photons from the fluorescently labeled RNA that are excited by the lasers. PMT saturation is a result of the photomultiplier tube becoming oversaturated due to an overabundance of converted electrons by the analog to digital (A/D) converter. The A/D converter can only convert the PMT signal into a value less than or equal to 2^{16}-1 or 65,535 and thus any fluorescent photons captured at this value of 65,535 are not discernable (Yang *et al.*, 2002). This type of saturation can be avoided by reducing the PMT voltage and laser power. Alternatively, the specific activity of the mRNA (number of fluorescent molecules per message) can be reduced. However, the reduction of the PMT voltage, power of the lasers, or reduced labeling, does not address the question of whether or not a particular cDNA on a microarray is biologically saturated.

Biological saturation occurs when the amount of mRNA that can hybridize to the DNA on a microarray reaches a maximum binding capacity of the printed DNA. If biological saturation is reached, then the amount of a mRNA will be underestimated and differences among arrays or experiments can not be appropriately determined. To avoid biological saturation, the amount of target RNA must be present in quantities less than the amount that the cDNA on the microarray slide can bind. To determine the range and linear response of increasing amounts of mRNA, we hybridized a 500-fold concentration range of labeled RNA from cardiac tissue to the *F. heteroclitus* 384 cDNA metabolic microarray.

Sources of technical variation, other than PMT and biological saturation, come from methods used to fluorescently label the mRNA, the day on which the RNA is processed and varying amounts of available tissue (Gold *et al.*, 2004; van Haaften *et al.*, 2006). One of the most common approaches to fluorescently label mRNA for microarray studies is to amplify the RNA by synthesizing cDNA with a T7 RNA polymerase binding site. RNA is then synthesized *in vivo* by using the T7 RNA polymerase to incorporate amino allyls followed by covalent binding of fluorescent molecules to the incorporated amino allyls (Vangelder *et al.*, 1990). For small amounts of starting mRNA, the synthesis of RNA using T7 can be repeated to double the amplification. To understand the effect of a single round versus a double round of linear amplification we compared the quantification of RNA using both methods.

The day and process used to isolate mRNA are two additional sources of technical variation. Variation in the preparation of mRNA could alter its quality affecting how well the RNA amplifies, is fluorescently labeled, and the signal observed on the microarray. The day on which a tissue is sampled is not strictly technical but can introduce a second type of variation: biological variation. That is, isolating tissues on different days could introduce technical variation because of the precision of dissection and the quality of tissue or RNA preparation. However, because tissues are sampled on different days, the organisms may be biologically different (under more or less stress, healthier, or just one day older). To examine technical variation due only to RNA isolation, a single blood sample was divided into four, RNA was separately isolated from each sample and, gene expression was quantified. Biological variation was examined in a separate experiment where five fish were bled every two weeks for a total of six weeks in order to collect four separate samples from each individual. Gene expression was quantified for these four temporally separate samples.

These experiments indicate that for a wide range of experimental conditions, microarray experiments using the *Fundulus* array are both accurate and precise.

Materials and Methods

Organism

Fundulus heteroclitus were caught from wild populations in Wiscasset, Maine, USA (43º57’41”N, 69º42’45”W) by minnow trap. Fish were transported to the Rosenstiel School of Marine and Atmospheric Science at the University of Miami and acclimated to 20ºC and 15ppt for approximately 6 months.

Blood Sampling

Fundulus heteroclitus ($N = 20$) were anesthetized with MS222 ($0.1\ g{\cdot}l^{-1}$) and given tags with subdermal latex markers. Whole blood samples from each fish were taken every two weeks by caudal puncture using a 50 µl Hamilton syringe rinsed with heparinzed saline (50 i.u. $\cdot ml^{-1}$). Samples were immediately frozen in liquid N_2 and stored at -80ºC. Only individuals that had all four serial samples taken ($N = 5$) were used in the present study.

RNA isolation and amino allyl labeling

Total RNA was isolated using 4.5M guanidinium thiocyanate, 2% N-lauroylsarcosine, 50mM EDTA, 25mM Tris-HCl, 0.1M β-Mercaptoethanol and 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 V3.2.1) and a bioanalyzer (Agilent 2100). RNA was then converted into amino allyl labeled RNA (aRNA) using the Ambion Amino Allyl MessageAmp II aRNA Amplification kit. This method converts poly-A RNA into cDNA with a T7 RNA polymerase binding site; T7 is then used to synthesize 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. aaUTP incorporation allows for the coupling of Cy3 or Cy5 dyes (GE biosciences) onto aRNA for microarray hybridization.

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

Microarrays

mRNA expression was measured using microarrays where each array had four spatially separated replicates per gene. The 384 *F. heteroclitus* cDNA microarrays were printed using 55 control genes and 329 cDNAs which encode essential proteins for cellular metabolism (Table 2.1). The annotation of genes and related pathways used *FunnyBase* (Paschall *et al.*, 2004) and these were manually compared to KEGG pathway designations. Because many genes belong to more than one pathway, central metabolic pathways were preferentially used if the gene coded for a protein that was a catabolic or anabolic enzyme (*versus* acting in a signaling pathway that affected metabolism). Controls include DNA spots labeled with Cy5 (positive control for position and gridding) and *Ctenophore* cDNA as negative controls.

Microarrays were created by printing cDNAs amplified with amine-linked primers onto 3-D Link Activated slides (Surmodics Inc., Eden Prairie, MN) at the University of Miami's microarray facility. All printed cDNAs were re-sequenced from the same source used for printing. The microarray slides were scanned using ScanArray Express. The raw TIFF-image data was quantified using Imagene (v5).

All experiments used a loop design for hybridization of dye labeled aRNA (Kerr, Churchill, 2001). In a loop design (Kerr, Churchill, 2001) each individual is labeled with Cy3 and Cy5. Each dye labeled sample is then hybridized on different arrays with another individual (Oleksiak *et al.*, 2002). Thus, each individual is hybridized to two arrays with four replicates per array for a total of eight technical replicates per individual. This experimental design is a more efficient use of resources, providing more data per array and is thus statistically more powerful than a reference design.

To test for the relationship between fluorescence and the quantity of RNA, five concentrations of fluorescently labeled RNA were used: 1.2 to 700 pmol of Cy3 or Cy5 labeled mRNA where pmol are for the amount of incorporated dye (Table 2.2). A 15 µl hybridization using the 384 cDNA array corresponds to 0.09 to 47 µM of Cy dye. Cy5 dye labeled RNA was used at concentrations 18% less than Cy3 because the Cy5 dye is a more efficient fluorophore (greater fluorescence per photon) than the Cy3 dye. The average of eight fluorescence values for each gene was normalized to the original concentration of RNA added.

Criteria for Inclusion

For a gene to be included in an analysis, the average signal among all arrays and dyes had to exceed background but not exceed 95% of PMT saturation (65,535). Background signal was determined as the amount of fluorescence in negative control array elements. Not all genes met these criteria and therefore were not included in the analysis.

Statistics

To adjust for systematic variation, gene expression values were first sum normalized, log2 transformed, and then loess normalized using Microarray Data Analysis System Software (MIDAS) (Dudoit *et al.*, 2003; Quackenbush, 2002) and SAS JMP Genomics v.6.0.2. For every gene, eight fluorescence values were captured; four Cy3 values and four Cy5 values. Analysis of variance (ANOVA) was performed using SAS JMP Genomics v.6.0.2. To look for differences between single and double rounds of amplification the following ANOVA model was applied: $y_{ijkl} = \mu + A_i + D_j + T_k + R_l + \varepsilon_{ijkl}$ where μ is the sample mean, A_i is the effect of the i^{th} array (i=1-18), D_j is the effect of the j^{th} dye (Cy3 or Cy5), T_k is the effect of the number of rounds of amplification (single or double, k= 2), R_l is the effect of the day on which samples were prepared (l=3), and epsilon is stochastic effects. The number of rounds of amplification (single or double) and channel variables were treated as fixed effects and array, and day on which samples were prepared were treated as random effects. Statistical analyses of replicate blood samples or repetitive measures of the same five individuals were applied to a separate ANOVA for each individual. The ANOVA model for this comparison was as follows: $y_{mnp} = \mu + A_m + D_n + T_p + \varepsilon_{mnp}$ where μ is the sample mean, A_m is the effect of the *m*th array (m=1-4 for both replicate and repetitive samples), D_n is the effect of the *n*th dye (Cy3 or Cy5), T_p is the treatment effect and epsilon is stochastic effects. Sample, representative of either one of four temporal samples from an individual or one of four replicate blood samples, and channel were treated as fixed effects. Array was treated as a random effect. Significant differences were evaluated with a p-value cut-off of 0.01.

Results

Biosaturation

The concentration of fluorescently labeled RNA (0.09 to 47 μM of Cy dye) represents 0.1X, 1X, 5X, 10X, 50X the concentration of RNA typically used on *F. heteroclitus* cDNA microarrays (Crawford, Oleksiak, 2007; Oleksiak *et al.*, 2002; Oleksiak *et al.*, 2005; Whitehead, Crawford, 2005; Whitehead, Crawford, 2006a) (Table 2.2, MIAME GSE12858). Among the 329 metabolic genes on the array, 212 of these genes met our criteria of being less than 95% of the PMT saturation and more than two standard deviations above the negative controls (*Ctenophore* cDNA with no similarity to vertebrate genes).

The linear relationship between the amount of RNA and relative fluorescence is shown in Figure 2.1. To remove the gene specific differences in expression, the fluorescence at each concentration was divided by the mean fluorescence for that specific gene (Fig. 2.1). The linear relationship between the amount of total fluorescent RNA added and the measures of gene specific fluorescence was determined for each gene. Most genes (176/212 or 83%) had an R^2 > 95% and 78 genes had a nearly perfect R^2 (> 0.995; Fig. 2.1B; Table 2.3). Examining the 18 genes with the lowest R^2 values (less than 0.8) revealed a non-linear relationship that can be explained by an apparent saturation at the 50X concentration of RNA (Fig. 2.1C). The relationship disappears if the fluorescence values for the 50X concentrations of RNA are removed and the 0.1 to 10X are plotted (Fig 2.1D-F). In the 100-fold range (0.1 to 10X) only three genes (1.4%) had R^2 values less than 0.8 (Table 2.3). Examination of the higher concentrations (1.0 to 50X) revealed 19 genes (9%) with R^2 less than 0.8 (Table 2.3). These data suggest that for most genes there is a linear relationship for a 500-fold range of RNA, however some cDNAs on the microarray will reach biological saturation at the highest RNA concentration.

Variation in RNA preparation

To determine how RNA preparation affects variation, cardiac RNA from three individuals were combined, and then evenly divided and amino allyl and dye labeled on three separate days using single and double rounds of amplification (MIAME, GSE12858). Only 110 genes met our criteria for inclusion because many genes were below the low cut-off (*Ctenophore* negative control cDNAs). In this experiment fewer genes met our criteria of above background and below saturation due to sample RNA being divided for separate labeling using either single or double rounds of amplification. An ANOVA was performed to measure differences between single and double rounds of amino allyl labeled RNA amplification. Twelve of the 110 genes (11%) used in this analysis were significantly different between single and double rounds of amplification at

$p < 0.01$. The majority of genes (59%) had a higher fluorescence signal when only one round of amplification was performed.

Consistency of Quantitative Determination

In teleost fish, red blood cell (RBCs) nuclei are transcriptionally active (Currie, Tufts, 1997; Koldkjaer *et al.*, 2004), and these cells can be sampled without sacrificing the fish. Thus to assess the consistency of microarray determinations, two experiments were performed on blood gene expression: 1) to examine technical variation a single sample of blood was divided into four samples; RNA isolations, amino allyl and dye labeling, hybridization and quantitative analyses were performed on each sample and 2) to examine biological variation, RNA isolated from blood from the five individuals were each sampled four times over a 6 week period (two weeks between samples).

A one-way ANOVA was used to test for the technical variation in gene expression between the four RNA samples isolated from a single blood sample (Fig. 2.2). Among all 252 genes (eight replicates per gene per sample) only 6 genes were significantly different for the four isolates at a critical p-value of 0.01. Three false-positives are expected at a p-value of 0.01 and thus with only 6 significant differences (Fig. 2.2) there is little evidence that separate RNA isolation, labeling and hybridization has much affect on measures of gene expression. The lack of differences is not due to high technical variation: CV (standard deviation/mean) among the eight replicates was 4% and, only three genes had a CV of > 10%. Nor was it due to the low p-value of 0.01 *versus* 0.05 (Fig. 2.2); the number of significant differences simply reflects the p-values. Random biological variation can contribute to differences in expression. We tested for random biological variation by bleeding the same five individuals four times with two weeks between bleedings (Fig. 2.3). For each of the 304 genes that met our criteria, an ANOVA tested for differences in expression among the four different time periods for each individual (four sample periods with eight replicates per gene per sample period). Among the four temporal samples, there were between one and seven genes that had a significant difference in expression at a p-value of 0.01 (Fig. 2.3). Only one individual had more than the expected number of false positives at the critical p-value: individual-00 had 7 (2%) significant genes at p-value 0.01 for 304 genes.

Discussion

Understanding sources of variation in gene expression is important for determining the biological importance of measured differences in mRNA expression. The analyses of technical variation in the metabolic *F. heteroclitus* cDNA microarray suggest that measures of gene expression using the *F. heteroclitus* 384 cDNA microarray are quantitative and precise. This conclusion is based on the observation that there is a linear increase in fluorescence with increasing mRNA (Fig. 2.1), and that there is little additional variation due to RNA processing (Fig. 2.2) or the day on which RNA is isolated (Fig. 2.3).

There is a linear increase in fluorescence with increasing mRNA for 98.5% of genes between 0.1X to 10X concentrations (0.09 pmol/ul to 9.3 pmol/ul) and 95% of genes between 0.1X to 50X (0.09 pmol/ul to 47pmol/ul). The linear relationship between RNA and fluorescence is quite strong for RNA concentrations of 0.1X to 10X having average R^2 values of 0.97, and most genes (88%) have R^2 values greater than 0.95 for these four concentrations. The genes most affected by biological saturation do not have a high fluorescence; if anything, they are less than the average (genes with $R^2 < 0.8$ for 1X to 50X have a mean that is 60% of the mean for all other genes). The two possible explanations for biological saturation with low fluorescence are that the synthesis of amino allyl labeled RNA for these genes is strongly truncated or that there is less DNA printed on the array for these genes. Truncation of amino allyl labeling would produce many more short probes with few labels per probe. Thus, to produce a similar fluorescence many more molecules would be necessary and these would saturate the DNA on the array. These problems can be avoided by using moderate amounts of probe (< 10 pmol/ul). We typically avoid this problem by using 0.7 to 2 pmol/ul. Using concentrations of RNA up to 50X (47pmol/ul) is feasible, but our data suggest that at this high of a concentration some genes will biologically saturate the cDNA on the array and therefore should be avoided.

If RNA samples are amino allyl labeled using one round of T7-RNA synthesis (Eberwine, 1996) versus two rounds of T7-RNA synthesis, 11% of genes have significant differences in fluorescence at a p-value of 0.01. Although this difference in gene expression for single *versus* double labeling is not large, it may be unacceptably high. Thus, we would suggest that for any one experiment that a researcher uses only single or double labeling procedures but not both within an individual experiment. Approximately half (59%) of genes with a significant difference between single and double labeling were greater for single labeling. The greater fluorescence for single labeling than that for double labeling would occur if cDNA or RNA synthesis was truncated with each round of labeling. Truncation would occur if the synthesis of cDNA or RNA were incomplete forming shorter nucleotide sequences with less fluorescence per RNA.

We used blood to test the effect of different RNA isolations, amino allyl labeling and hybridizations. The first experiment used a single blood isolation that was divided into four equal samples. There are few differences in expression, 2.4% at a p-value of 0.01 (i.e., six *versus* the expected three false positives). If a Bonferroni correction was applied none of these genes would be significant. Therefore, technical errors do not necessarily contribute significant amounts of variation. Similar conclusions were made about microarray results among laboratories: many different laboratories yielded similar results using different varieties of platforms (Bammler *et al.*, 2005; Beissbarth *et al.*, 2000; Bloom *et al.*, 2004; Irizarry *et al.*, 2005; Larkin *et al.*, 2005; Tan *et al.*, 2003; Yauk *et al.*, 2004). However, a few laboratories yielded different results. Together these data suggest that good experimental practice can minimize the effect of technical variation.

In a separate experiment, five individuals were bled once every two weeks during a six-week period, resulting in four serial blood samples from each individual. Any differences in expression among sampling times could be due to technical variation, of which there is very little as shown by the previous experiment, or biological variation. That is, although fish appeared healthy, had normal blood glucose, and the stress hormone, cortisol, did not vary significantly ($p > 0.1$), gene expression could vary significantly for unknown biological reasons. Yet, for the five individuals there are few, if any, meaningful differences in gene expression (only one individual had more than the expected number of false positives, Fig. 2.3). These data confirm the observation that technical errors do not necessarily affect microarray measures. Importantly, these data also suggest that for a tissue or blood sample there is little random stochastic variation in gene expression. These data are in contrast to other publications suggesting that mRNA expression is noisy and has large stochastic variation (Blake *et al.*, 2003; Raj *et al.*, 2006). The important distinction is that for a single cell, transcription is pulsatile, occurring in bursts (Blake *et al.*, 2003; Raj *et al.*, 2006), and for an individual cell this creates large stochastic variation in mRNA expression. However, our results demonstrate that for millions of cells, this variation is not apparent across a 6-week time course. We suggest that if there is a large stochastic variation in each cell, sampling of millions of cells masks this variation such that the amount of expression from any one gene is stable over time.

The microarrays used here have array elements for essential metabolic genes (Table 1) and are similar to the array elements used in previous work demonstrating larger inter-individual variation in gene expression (Oleksiak *et al.*, 2002; Oleksiak *et al.*, 2005; Whitehead, Crawford, 2006a; Whitehead, Crawford, 2006b). While the data presented here addresses the sources of variation in many microarray experiments, the lack of temporal variation in gene expression in our study may only reflect the expression of the metabolic genes. However, these results are similar to studies of gene expression in humans where the same individuals were sampled over a time period of 24 hours to four weeks (Cobb *et al.*, 2005; Eady *et al.*, 2005; Whitney *et al.*, 2003). These studies also found relative stable expression of a more diverse set of genes when the same individuals were sampled over time. Thus, although there are good biochemical or molecular reasons to expect stochastic variation in gene expression, this variation is not necessarily observed using routine sampling methods.

Microarrays are a useful technology for observing differences in gene expression and data extracted from microarrays can be reliably reproduced. With reasonable care, any experiment involving microarrays is capable of obtaining biological data that is not masked by technical variation thereby providing a true representation of the transcriptome under a particular set of conditions. However, caution is required before making conclusions about the biological nature of the data until the sources of technical variation are understood.

Table 2.1 384 Microarray Metabolic Pathways

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 & Pyrimidine metabolism	9
Pyruvate metabolism	2
Signaling	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

Table 2.2 Concentrations of Cy3 and Cy5 dye labeled RNA used for hybridization.

	50X	10X	5X	1X	.1X
Cy 3	700 pmol	140 pmol	70 pmol	14 pmol	1.4 pmol
Cy 5	583.3 pmol	116.6 pmol	58.3 pmol	11.6 pmol	1.2 pmol

Table 2.3. Number of genes and corresponding R^2 for various ranges of RNA concentrations.

R^2	0.1 -50X	0.1-10X	1.0-50X
>0.9	176	199	178
<0.8	18	3	19

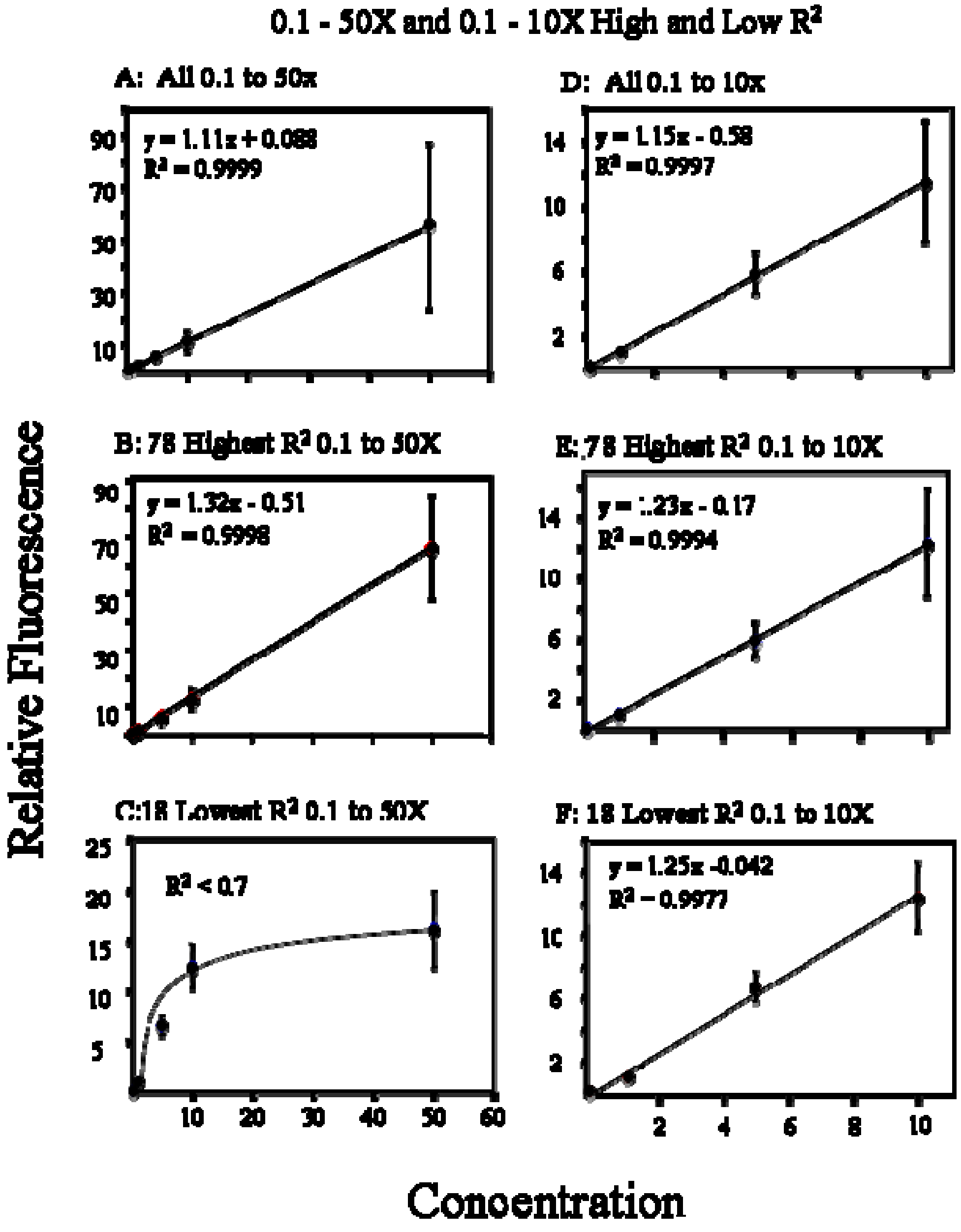
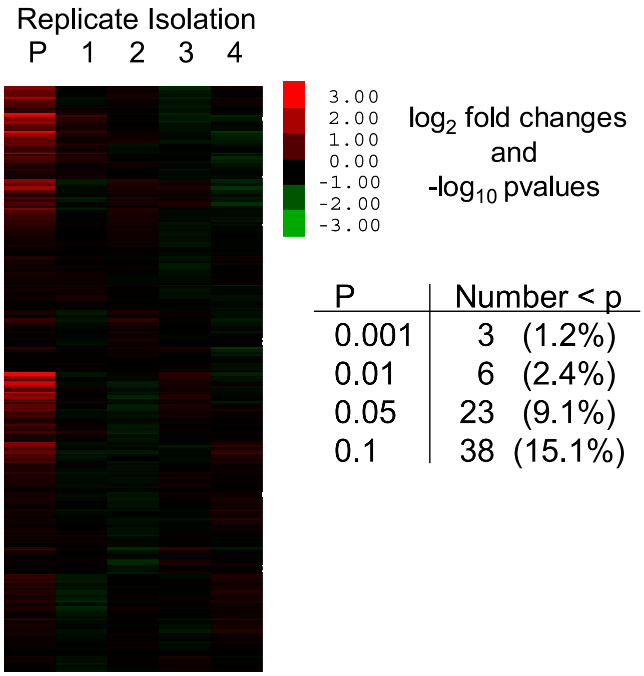
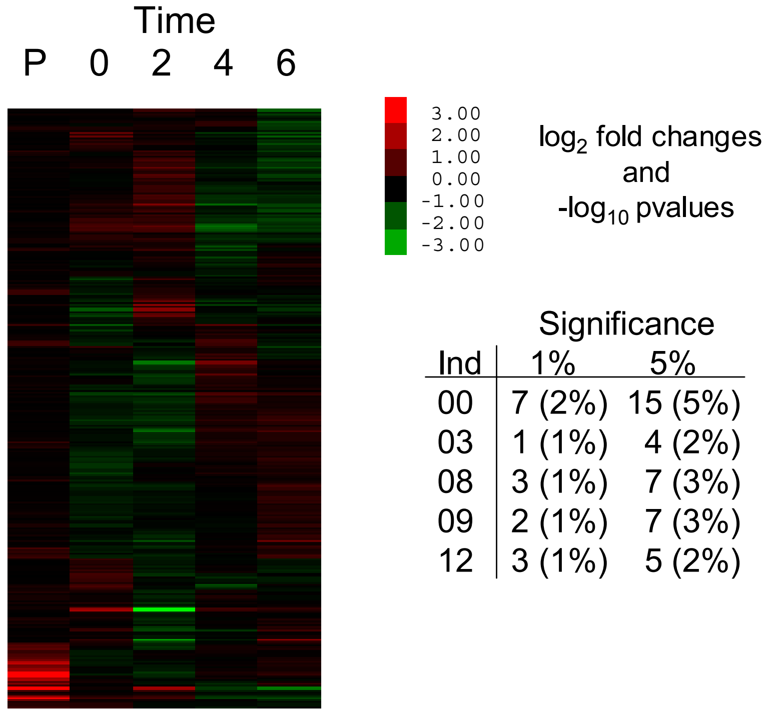


Figure 2.1 Linear relationship of RNA concentration to relative fluorescence. Graphs show linear relationship between concentrations of RNA (0.1-50X, A-C, and 0.1-10X, D-F) and relative fluorescence. Relative fluorescence is a normalized measure of fluorescence divided by the gene specific mean. 1X RNA is equal to 0.9 pmol μl^{-1}. Shown are the RNA concentrations *versus* fluorescence for 0.1 to 50X (A-C) and for 0.1X to 10X (D-F); for all genes (A and D), for the 78 genes with the highest R^2 values (B and E), and for the 18 with lowest R^2 values (C and F).



P	Number < p
0.001	3 (1.2%)
0.01	6 (2.4%)
0.05	23 (9.1%)
0.1	38 (15.1%)

Figure 2.2 Gene expression for single blood isolate. Heat map for single blood isolate that was divided into four. RNA was purified, labeled and hybridized separately for each sample. Red is greater and green is less than the average gene specific fluorescence. First column (P) is the p-value from a one-way ANOVA. Only 6 genes (2.3%) out of 252 are significant at a critical p-value of 0.01. P-values (-log10) shown in the heat map are from an ANOVA for significant differences among samples using the 8 replicates for each separate RNA isolation. Color bar gives fold difference for log$_2$ gene expression (e.g., 2=4x) and negative log$_{10}$ p-value (e.g., 2 = p-value 0.01).



	Significance	
Ind	1%	5%
00	7 (2%)	15 (5%)
03	1 (1%)	4 (2%)
08	3 (1%)	7 (3%)
09	2 (1%)	7 (3%)
12	3 (1%)	5 (2%)

Figure 2.3 Individuals sampled over time. Heat map for one individual (00) that was sampled 4 times over a total of 6 weeks. Numbers above the heat map are time points (0, 2, 4 and 6 weeks) and the "P" is for p-value (-log$_{10}$). P-values are from the ANOVA that tested for differences among separate blood isolations within an individual (4 isolations and 8 replicates per isolation). For gene expression, red is greater and green is lower expression than the mean expression for each gene. Table provides number of significant genes and percent (rounded up) out of the total of 304. Color bar gives fold difference for log$_2$ gene expression (e.g., 2=4x) and negative log$_{10}$ p-value (e.g., 2 = p-value 0.01).

CHAPTER 3: THE EFFECT OF GENETIC AND ENVIRONMENTAL VARIATION ON GENE EXPRESSION

Background

Variation in mRNA expression is 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 outbred populations of the teleost fish *Fundulus heteroclitus* there is substantial variation in mRNA expression within and between groups of individuals (Crawford, Oleksiak, 2007; Oleksiak *et al.*, 2002; Oleksiak *et al.*, 2005; Whitehead, Crawford, 2006a). 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 apparent adaptive patterns suggest that the variation in mRNA expression is biologically important because it is the result of natural selection (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 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 data to test this hypothesis, we examine the variation in 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 among individuals, then more genetically similar individuals should have less variation and fewer significant differences in mRNA expression than among unrelated individuals. 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 from the field did not increase the variation in mRNA expression among these outbred samples. These observations suggest that the normal environmental variation associated with tidal fluxes in estuarine environments does not substantially add to the differences in mRNA expression in *F. heteroclitus.*

Methods and Materials

Organism

Fundulus heteroclitus were caught from wild populations in Beaufort, North Carolina, USA (34º43’34’’N, 76º40’62’’S) by baiting commercially available minnow traps with dry dog food. Upon capture in the field, five males and five females were killed by cervical dislocation, their hearts removed and stored in 1 ml of RNA later at -20ºC (Ambion, Inc.). The remaining fish were acclimated to 20ºC and 15 ppt of artificial seawater in laboratory aquaria at the University of Miami for approximately 6 months (Instant Ocean, Inc.). These fish were compared to fish trapped at the same location, and raised at 20ºC and 15 ppt of artificial seawater and allowed to interbreed for ten successive generations (G10) at the Aquatic Biotechnology and Environmental Laboratory at the University of Georgia. For the purposes of this experiment, 5 males and 5 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 ten 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 pair wise 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). Briefly, microsatellite primers were constructed by using *F. heteroclitus* genomic fragments (350-700 bp long) that were ligated into a pUC19 plasmid and electroporated into *Escherichia coli* (Adams *et al.*, 2005). DNA sequencing was performed on inserts which were PCR amplified resulting in unique microsatellite-containing sequences from which primer sets were specifically optimized for 17 loci (Adams *et al.*, 2005).

DNA was extracted from dried fin clips. The tissue was placed in 300 µL lysis buffer (75 mM NaCl, 25mM EDTA, 1%SDS) and incubated with 0.1 mg *Proteinase K* at 55ºC for 2 hours. Proteins were precipitated by adding a half volume of 7.5 M ammonium acetate and centrifugation for 10 minutes at 16,000 g at room temperature. DNA was precipitated from the supernatant by adding 0.7 volumes of isopropanol and centrifugation for 15 minutes at 15,000 g at room temperature. The DNA pellet was washed with 70% ethanol then allowed to air dry for 30 minutes followed by re-suspension in 50 µL 10 mM Tris-HCl pH 8.5.

Microsatellite loci were amplified in four fluorescently labeled 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*) (Adams *et al.*, 2005). Locus *ATG-20* was amplified alone at a final concentration of 0.5 µM. The 10 µL reactions contained 2.5 mM $MgCl_2$, 1X PCR buffer (500mM Tris-HCl, pH 9.2, 160mM $(NH_4)_2SO_4$, 22.5 mM $MgCl_2$, 20% (v/v) DMSO, 1% (v/v) Tween T 20, water to 10 ml volume), 0.2 mM dNTPs, 0.4 units Taq DNA polymerase (Promega), 70 ng DNA, and one of the five primer combinations (see above for concentrations). The PCR thermal cycling profile consisted of 94ºC for 2 minutes, followed by 31 cycles of 94ºC for 15 seconds, 55ºC (A, C, *ATG-20*) or 55ºC (B and D) for 15 seconds, and 72ºC for 30 seconds, ending with a 5 minute extension step at 72ºC. Following PCR amplification, the products from A, C, and D were co-loaded, as were ATG-20 and B, before being subject to electrophoresis on an ABI 3730XL Genetic Analyzer (Applied Biosystems). GENEMAPPER v4.0 (Applied Biosystems) was used to score the genotypes. All genotypes were checked by two members of our laboratory to ensure the proper scoring of genotypes.

RNA isolation, labeling 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 (Sigma). 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 V3.2.1) and by capillary electrophoresis with the use of a bioanalyzer (Agilent 2100). RNA was then converted into amino allyl labeled RNA (aRNA) using the Ambion Amino Allyl

MessageAmp II aRNA Amplification kit. This method converts poly-A RNA into 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. aaUTP incorporation allows for the coupling of Cy3 or Cy5 dyes (GE biosciences) onto aRNA for microarray hybridization.

Dye labeled 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 labeled sample = 2 pmol μl^{-1}). Hybridization buffer consisted of 5X SSPE, 1% SDS, 50% formamide, 1mg ml^{-1} polyA, 1 mg ml^{-1} sheared herring sperm carrier DNA, and 1mg ml^{-1} BSA (Botwell, Sambrook, 2003). Slides were washed in sodium borohydride solution in order to reduce autofluorescence. Following rinsing, slides were boiled for 2 minutes and spin-dried in a centrifuge at 14 g for 3 minutes at room temperature. Samples (15 µl) were heated to 90°C for 2 minutes, quick cooled to 42°C, applied to the slide (hybridization zone area was 350 mm^2), and covered with a cover slip. Slides were placed in an airtight chamber humidified with paper soaked in 5X SSPE and incubated 24-48 hours 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 *Fundulus heteroclitus* cDNAs that included 329 cDNAs that encode essential proteins for cellular metabolism (Table 3.1, (Paschall *et al.*, 2004)). Average lengths of cDNAs were 1.5 Kb with a majority including the N-terminal methionine. Table 3.1

provides a summary of the ESTs used for printing where the most meaningful gene ontology (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, MN) at the University of Miami core microarray facility (http://www.rsmas.miami.edu/groups/ohh/genomics/genomics_core.htm).

The microarray slides were scanned using ScanArray Express. The raw TIFF-image data was quantified using Imagene (v5). If a gene had a fluorescent signal that was too low or too high, then it was eliminated from the analysis for all individuals. Fluorescent signals were considered too low if the average across all samples were within 2 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) (Scott *et al.*, 2009). Using these criteria 100 genes were eliminated from all individuals leaving a total of 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 5,000 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 1,000 times to construct the 95% CI around the random expectation. Ninety-five percent confidence intervals were also calculated around each mean by bootstrapping values within each group 1,000 times.

Statistical analyses of the mRNA expression data were carried out using JMP genomics (SAS JMP Genomics v.7.0.2). All analyses used fluorescent microarray measures that were $\log_2$ transformed and loess normalized (Quackenbush, 2002). These normalized fluorescent measures showed nearly identical distributions among all individuals. Standardization of data (mean signal with an average intensity equal to zero) or further normalization using ANOVA or mixed model did not substantially affect the distribution of fluorescence nor did it affect the relative frequency of genes with significant differences in expression among individuals. Thus, the simpler of the two normalization methods ($\log_2$-loess) was used for parsimony and clarity of results.

To test for the significant differences in gene expression among individuals within each group we used a linear mixed model ANOVA (Kerr, Churchill, 2001; Patterson *et al.*, 2006; Wolfinger *et al.*, 2001; Yu *et al.*, 2004):

$$y_{ijk}=\mu + A_i + D_j + I_k+\varepsilon_{ijk} \qquad (1)$$

where y_{ijk} represents the fluorescence intensities on a log scale and μ is a constant. The fixed effect is I_k for the k^{th} individual (for one of the nine individuals per group), the random effects are D_j for the j^{th} dye of the two Cy dyes and A_i is the i^{th} array (for one of the 27 different arrays) and ε_{ijk} is a random residual term. With nine individuals per group (acclimated, field or G10) 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 individual in each group was statistically different, a mixed model analysis was performed using equation (1) on all 126 possible combinations of 5 out of nine individuals for each group. From these 126 ANOVAs, the confidence intervals and the average number of genes that were significantly different among individuals were calculated.

For all other analyses, the linear mixed model was used to define the least square means, providing a single measure of expression for each gene for each individual. This model is identical to the model described above but all individuals were modeled without regard to group (26 and 161 degrees of freedom). The PROC MIX statement used in the

SAS analysis was: class Dye Sex Treatment Array indiv; model response= indiv; random Dye Array; lsmeans Indiv;.

To find significant differences in mRNA expression between groups or sex, we performed an ANOVA with the least square means for each individual. Notice that when using the least square means there are no dye or array effects (only one measure per gene results from the mixed model), thus, these factors are not included in the model. 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.

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 six months to common laboratory conditions (Acclimated) and *Fundulus* bred for ten successive generations (G10) in a common laboratory environment. All fish originated from the same location in Beaufort, North Carolina.

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 ($p > 0.20$ 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 3.2). The G10 fish were also in Hardy Weinberg equilibrium ($p = 0.14$), but 13 of the 45 pair-wise 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. 3.1). 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 3.1).

Significant Differences between sexes in mRNA expression

Gene expression was measured in a total of 12 males and 15 females using two separate hybridization loops. Yet only three genes have significant differences ($p < 0.01$) in mRNA expression between males and females. None are significant with Bonferroni's correction for multiple comparisons ($p < 3.5 * 10^{-5}$). With so few differences in gene expression between the sexes, males and females were analyzed together.

Significant Differences within Groups in mRNA expression

One measure of variation in mRNA expression is the number of genes that differ among individuals within each group (Fig. 3.2). 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 3.3). G10 individuals had approximately two-thirds the number of significantly different genes (66 or 23% $p < 0.01$, 23 significant with Bonferroni's corrected $p < 3.5 * 10^{-5}$) as did acclimated individuals (107 or 38% $p < 0.01$, 46 with Bonferroni's corrected $p < 3.5 * 10^{-5}$) even though they were raised in similar laboratory conditions. Surprisingly, field individuals had an intermediate number of significant genes (97 or 34% $p < 0.01$, 29 with Bonferroni's corrected $p < 3.5 * 10^{-5}$). Examples of the magnitude and associated p-values with these differences are shown in the volcano plots (Fig. 3.3). Although only three of the possible 36 paired comparisons in each group are shown in figure three, these differences are representative samples and suggest that the difference among acclimated individuals tends to be larger (x-axis, $\log_2$ differences in the least square mean) and they have 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 nine individuals were examined (Table 3.3). Among these combinations, the average numbers of genes with a significant difference in expression share the same pattern as the analysis of all nine individuals: acclimated > field > G10 for the number of genes with significant differences in mRNA expression. For each group, the mean number of genes with significant differences in expression among these 126 combinations has confidence intervals that do not overlap (Table 3.3) and are statistically different (Kruskal-Wallis non-parametric test $p < 0.001$). Thus, there is statistical support that there are a greater number of genes with significant differences in expression in acclimated versus field or G10, and field versus G10.

Variance in mRNA expression Across Genes

Another test of how the variation in mRNA differs among groups is to examine the mean variation among individuals for all genes. That is, rather than testing for the quantitative differences in mRNA expression, we tested the variation in mRNA expression among individuals for each group across all 284 genes. 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: $\text{lsmean}_{ij}/(\text{Avg}_j)$ where the lsmean_j is the least square mean for the i_{th} individual and the j_{th} gene. These measures are divided by Avg_j, the average least square mean for the j_{th} gene. The variance from these normalized values among the nine individuals within each group was calculated. The variance across all 284 genes is significantly different (Kruskal-Wallis test, $p < 0.001$) with a mean variance of 0.575 (stdev = 0.346), 0.423 (stdev = 0.256) and 0.386 (stdev = 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 Barlett's test for homogeneity of variance among groups using the least square means for each individual (Table 3.4) (Snedecor, Cochran, 1991). 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, acclimated individuals had larger variation than G10 individuals 77% of the time. This bias of larger variance in mRNA expression in the acclimated group *versus* G10 is unlikely to occur by chance (χ^2 $p < 0.01$). For the acclimated *versus* field, or field *versus* G10 there are the same (50%) or nearly the same (59%) number of genes with greater variance in mRNA expression (Table 3.4).

Significant differences between groups in mRNA expression

We expect a difference in the variance in mRNA expression between groups, but this will not necessarily be associated with a difference in the mean of mRNA expression. Using a p-value of 0.01, 7 genes (2.5%) have a significant difference in mRNA expression among groups (Fig. 3.4). None of these are significant with a Bonferroni's corrected p-value of $3.5 * 10^{-5}$). Using a t-test between the 3 pairs of comparisons (Acclimated vs. Field, Acclimated vs. G10 and Field vs. G10), there are no significant differences between the acclimated and the field groups (Fig. 3.4A). Significant differences in mRNA expression are only between either the acclimated or field versus G10 group (Fig. 3.4B).

Discussion

The genetic basis for the variation in mRNA expression among natural populations, including *F. heteroclitus,* is not well understood. In other species, our understanding of the genetics of mRNA expression has relied on the study of inbred strains (Gibson, Weir, 2005; Schadt *et al.*, 2003; Wayne *et al.*, 2004) 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 QTLs and has narrow sense heritability (h^2) greater than 30% (Cheung *et* *al.*, 2003b; Fu *et al.*, 2009; Gibson, Weir, 2005; Rockman, Kruglyak, 2006; Sharma *et al.*, 2005; Tan *et al.*, 2005). Heritability of mRNA expression has been measured in a variety of organisms. For example, in ten 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 *Sacchoromyces cerevisiae* segregants, 3,546 out of 5,727 measured genes (62%) had a $h^2 > 0.69$ (Brem, Kruglyak, 2005). Using lymphoblast human cell lines, among 15 families, 762 out of 2,430 (31%) of differentially expressed genes had a significant h^2 with median of 0.34 (Monks *et al.*, 2004; Williams *et al.*, 2007). Thus, it appears that much of the variation in gene expression has a substantial genetic component.

These studies on inbred lines or cell culture are informative and they provide the foundation for understanding 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 (Cobb *et al.*, 2005; Eady *et al.*, 2005; Radich *et al.*, 2004; Whitney *et al.*, 2003) 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 (Crawford, Oleksiak, 2007; Oleksiak *et al.*, 2002; Oleksiak *et al.*, 2005; Whitehead, Crawford, 2006a) 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 G10 individuals (Fig. 3.1). 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 interbred for ten 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-Weinburg equilibrium (Table 3.2).

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 3.3, Fig. 3.2 and 3.3), and the mean variance across all genes is significantly less. Additionally when examining the gene specific variation in mRNA expression, 22 genes have significant difference in the variation between individuals (within a group) 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 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 a 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, for genes with a significant difference in the variation in expression, 50% are larger in acclimated individuals and thus for these 22 genes there is no significant difference between the field and acclimated groups in the frequency of genes with significant variation in mRNA expression. These data indicate that acclimated individuals 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 appear to have a major affect on the variation in gene expression in this particular investigation.

Among the three groups (Acclimated, Field and G10) there are few statistically significant differences in expression: seven genes with a critical p-value of 1% (no genes with Bonferroni's corrected p-value). For all differences that do exist, these differences in expression are only significant between G10 and the two outbred groups (Fig. 3.4). We could not resolve what environmental factors might be held in common between the field and acclimated groups that could explain this observation. 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 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 random genetic drift due to a recent bottleneck caused by the successive breeding of the G10 individuals over ten generations. These patterns of variation in mRNA expression in acclimated, field and G10 individuals are consistent with the hypothesis that much of the variation among individuals is due to genetic variation. Specifically, there are fewer differences among G10 individuals that share 43% of their alleles *versus* completely outbred individuals (acclimated) even though both were subjected to similar laboratory conditions for at least six months.

We found little support for environmental affects on mRNA expression: acclimated individuals *versus* field individuals, that suffer the daily inundation of environmental variation associated with estuarine environments, have equal or more variation in mRNA expression. Added to these observations is the fact that there is little significant variation in mRNA expression when the same individual is repetitively measured over a six-week period (i.e. mRNA expression from blood sample every 2 weeks over six-week period; (Scott *et al.*, 2009). Together, these data strongly support the hypothesis that the large inter-individual variation in gene expression measured here and elsewhere (Crawford, Oleksiak, 2007; Oleksiak *et al.*, 2002; Oleksiak, Crawford, 2006; Oleksiak *et al.*, 2005; Whitehead, Crawford, 2005; Whitehead, Crawford, 2006a) is unlikely to reflect large environmental differences and is more reasonably 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 (Crawford, Powers, 1989; Hochachka, Somero, 1984; Pierce, Crawford, 1997; Prosser, 1986; Schmidt-Neilsen, 1990; Segal, Crawford, 1994).

However, these observations do not refute that other non-heritable mechanisms can effect 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 temperatures, 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 G10 individuals whose parents all experienced the same environment, affected mRNA expression differently than acclimated individuals whose parents experience a wide range of environments. However, there are few differences in expression among G10, acclimated or field individuals. Thus, epigenetic effects causing a difference in gene expression are not supported.

Alternatively, one could suggest that the variation in mRNA expression (but not a difference in the mean expression) is related to the environmental variation experienced by the parents or the developing embryo. That is, the more variable the paternal or developmental environment, the greater variation there is in mRNA expression. To test the hypothesis that variation in gene expression is related to the environmental variance experienced by the parent or developing embryo, gene expression could be measured at various time points during the growth of embryos. However, to explain most of the data, the hypothesis that the variation in expression is a function of parental or developmental environmental variation would require that greater environmental variation produce many different adult phenotypes and thus have a greater variation in mRNA expression. This hypothesis is unlike any currently available data, and cannot be readily rejected. Other genes may not adhere to this conclusion, however, it seems more prudent to suggest that the larger inter-individual 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 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 mRNA expression is either greater or at the very least similar among field and acclimated individuals. These data indicate that for the genes examined in this analysis, much of the variation in mRNA expression is related to genetic variation and less of the variation is in response to environmental change.

Table 3.1 384 Microarray Metabolic Pathways

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 & Pyrimidine metabolism	9
Pyruvate metabolism	2
Signaling	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

Table 3.2. Genetic diversity values for laboratory bred (G10) individuals (n = 49) and field caught individuals (F) (n = 109). 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.

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 ± SE	4.49 ± 0.43	13.53 ± 2.55	0.59 ± 0.06	0.78 ± 0.04	0.56 ± 0.05	0.78 ± 0.04	-0.06 ± 0.04	0.00 ± 0.01

Table 3.3 Number of Genes with Significantly Different mRNA Expression. Significantly different number of genes among all individuals without regard to group (All) and for acclimated (Acc), Field (Fld), and G10 groups. Average number of significant genes and 95% confidence intervals for all possible 126 combinations of 5 out of 9 individuals.

	Number of Significant Genes for each Group			
	Across All	Acc	Fld	G10
p <0.01	281	107	97	66
Avg. # for 5 out of 9 combinations		88	81	41
95% CI		90-86	87-75	45-37

Table 3.4. Homogeneity of Variance. Number of genes with unequal variances (Bartlett's test for homogeneity of variance).

	Total	Acc > G10	Fld> G10	Acc>Fld	Acc & Fld > G10
Number of Genes	284	210	154	184	131
Number Significant Genes	22	17	13	11	11
% of Significant Genes	100.0%	77.3%	59.1%	50.0%	50.0%

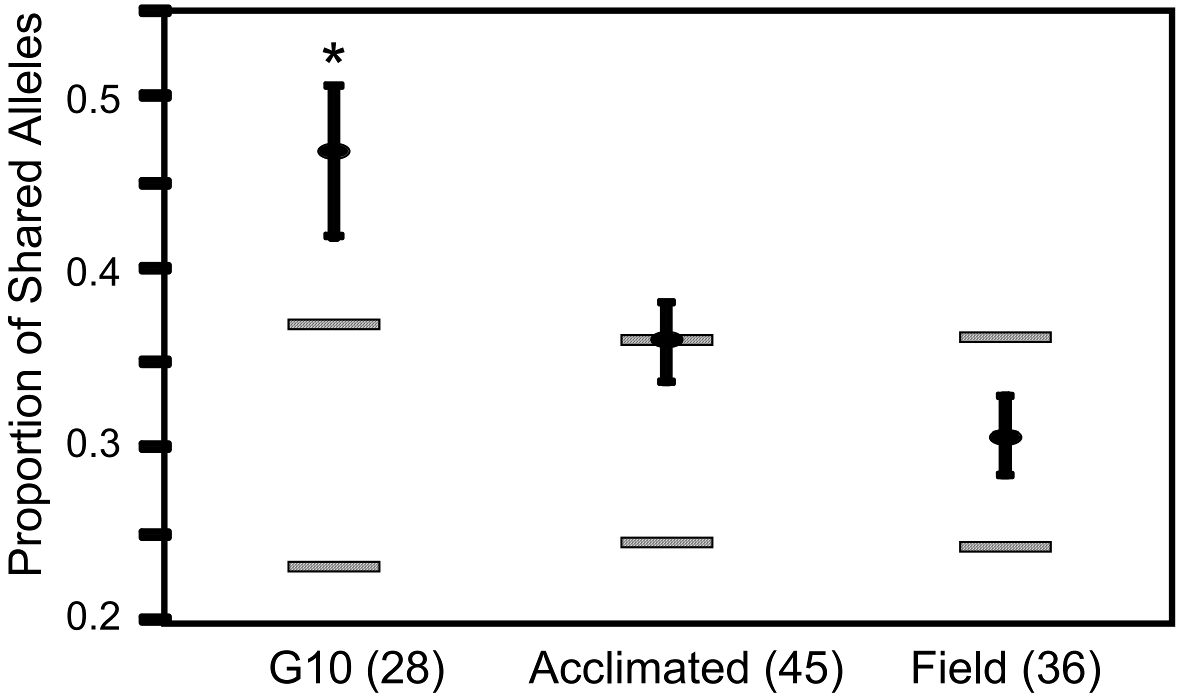


Figure 3.1. Average proportion of shared alleles within groups. Proportion of shared alleles is shown for inbred (G10), outbred acclimated (Acclimated), and outbred field (Field) groups. Numbers in parentheses are the number of pairwise comparisons in each group. Vertical lines with ellipses are 95% bootstrapped confidence intervals around each calculated mean value. 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.

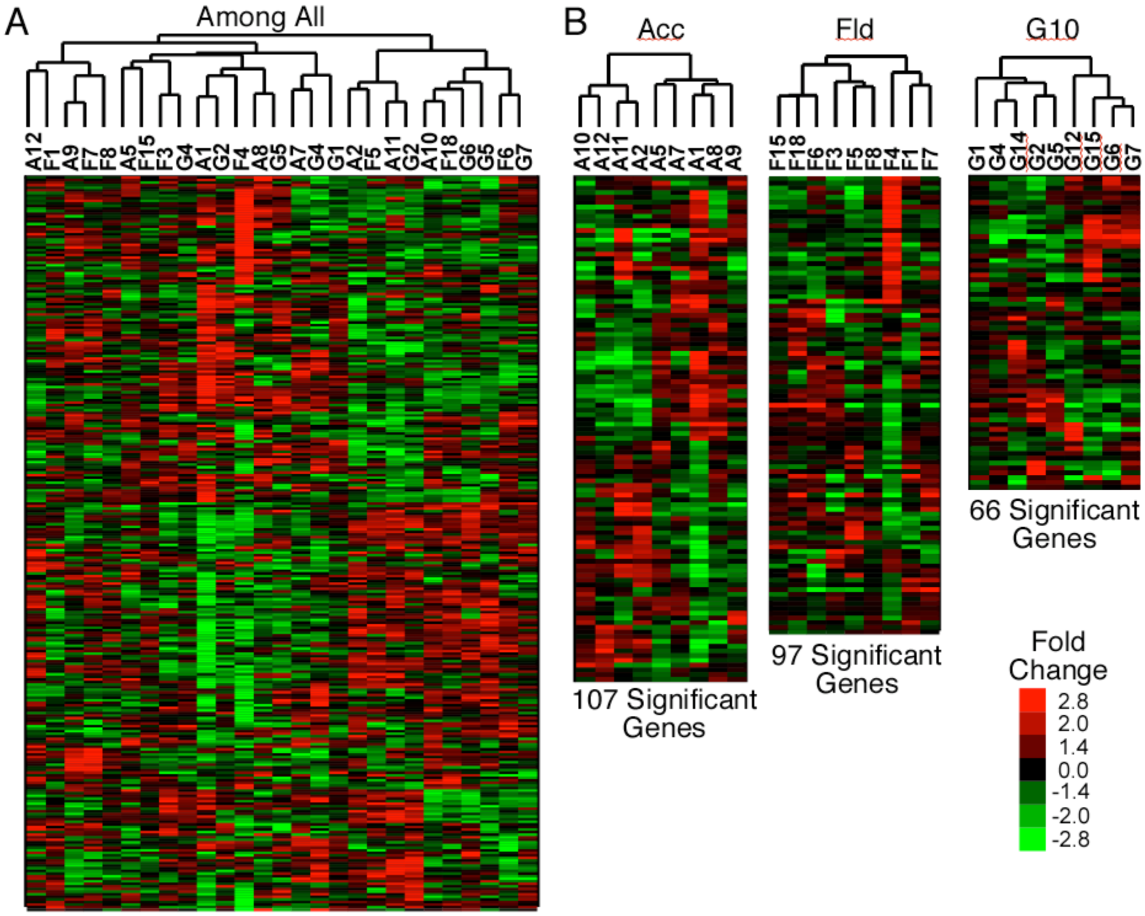


Figure 3.2. Heat maps of individual gene expression. Clustering is based on the correlations of least square means relative to the gene specific means. A. Relative expression among all individuals ignoring groups. B. Relative expression within each group.

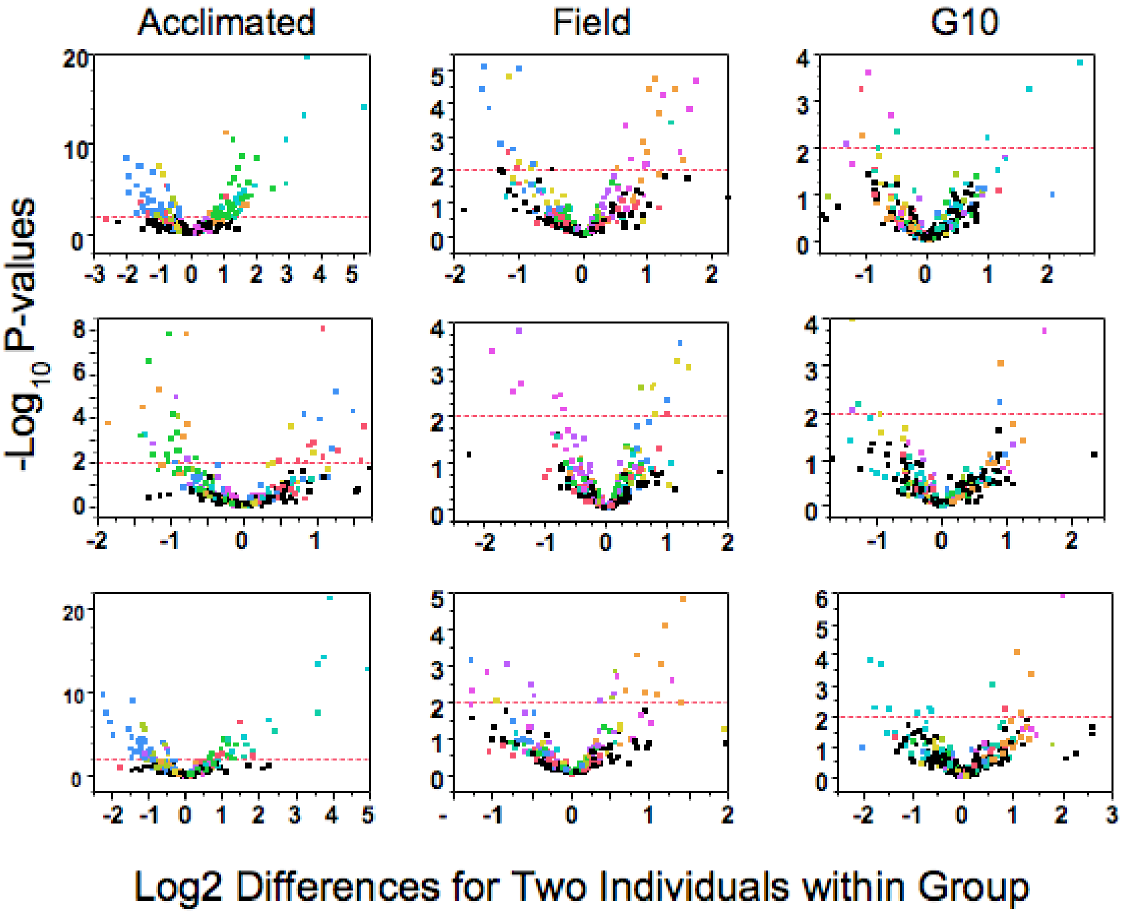


Figure 3.3 Volcano plots among three individuals within each group. Negative log10 p-values (e.g. p-value of 0.01 = 2) versus the difference of log2 expression values (difference of 1 = two-fold). Three of the possible 36 possible comparisons within each treatment are displayed. The differences among three representative individuals are displayed for acclimated, field and G10 groups (1-2, 2-3 & 1-3). Notice the axes are different for each different comparison.

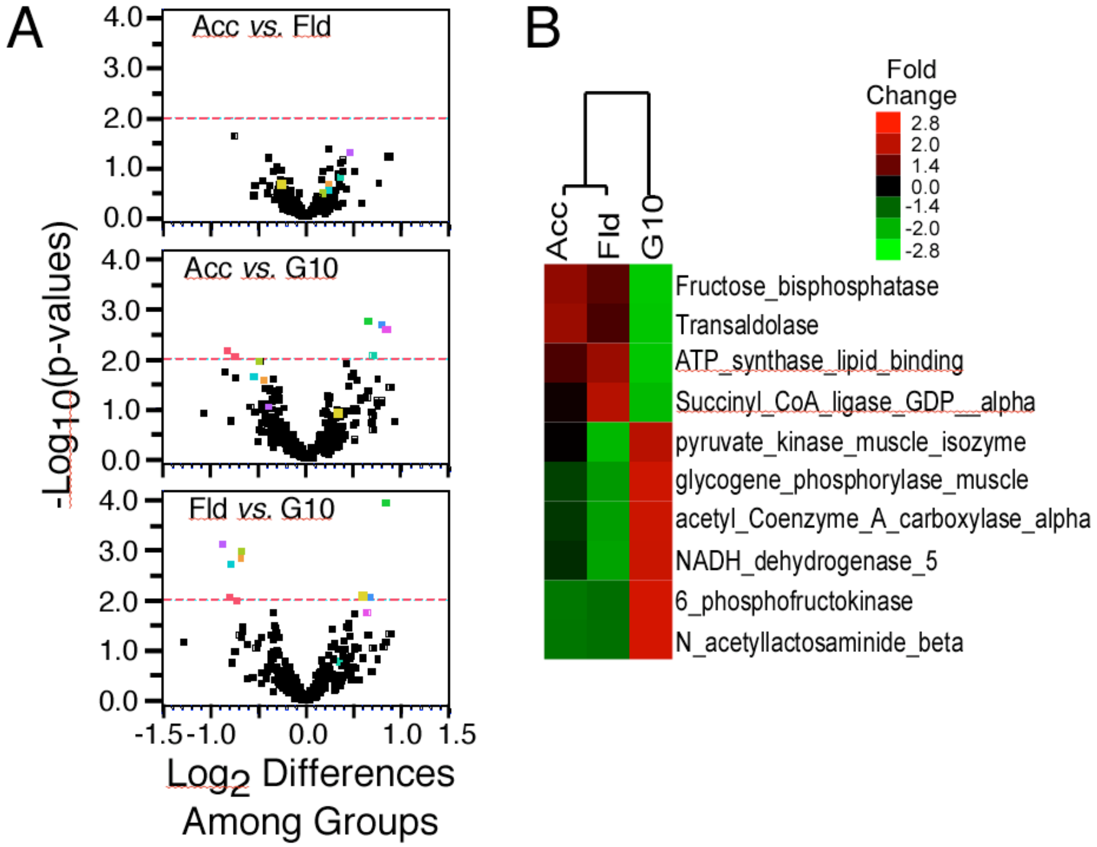


Figure 3.4. Differences Among Groups. Among any pair of groups there are ten genes that are significantly different (p-value < 0.01). A. Volcano plots of log2 differences in expression versus the negative log10 of p-value (2 = 0.01). B. Heat map of patterns of gene expression. Acc = acclimated, Fld = field and G10 = inbred population.

CHAPTER 4: THE HERITABILITY OF THE VARIATION IN GENE EXPRESSION IN FUNDULUS HETEROCLITUS

Background

Understanding the genetic basis of mRNA expression is important for understanding health and disease and the role of evolutionary adaptation in shaping patterns of gene expression. Though environmental contributions to gene expression may be important, it is the genetic component of the variation in gene expression that helps address questions pertaining to natural selection and evolution. Variation in the expression of genes is environmentally influenced, genetically determined or a combination of both. The ability to partition this variance is of interest as differential mRNA expression can discern the heritable components of variance and with further study, uncover whether the determinants of heritable variation are *cis* or *trans* acting (Petretto *et al.*, 2006; Schadt *et al.*, 2003; Yvert *et al.*, 2003). By treating each cDNA on a microarray as an independent quantitative trait, heritability in a variety of organisms including yeast, humans, mice and *Drosophila* has been estimated (Brem *et al.*, 2002; Cheung *et al.*, 2003a; Schadt *et al.*, 2003).

Heritability (h^2 or H^2) is the ratio of additive genetic variance to phenotypic variance (V_A/V_P) or the amount of phenotypic variation that is attributed to genetic differences. The concept was first put forth by two evolutionary theorists with competing views on the estimation of heritability. Sewall Wright (1917) used correlation and regression of parents on offspring to estimate heritability while R.A. Fisher (1918) used analysis of variance (ANOVA) to partition the components of variance (Dempster, Lerner, 1950). Since then, estimates of heritability using both methods have been used to understand stature in man, milk-yield in cows, litter size in pigs, egg production in poultry, body weight in mice, abdominal bristle number in *Drosophila melanogaster*, immune response to diphtheria-tetanus vaccine in blue tits and radiosensitivity of human lymphocytes which are used to determine cancer susceptibility (Barker, Robertson, 1966; Clayton *et al.*, 1957; Emsley *et al.*, 1977; Falconer, Mackay, 1996; Finnon *et al.*, 2008; Raberg *et al.*, 2003; Roberts *et al.*, 1978; Rutledge *et al.*, 1973; Strang, Smith, 1979). Just as the ability to compartmentalize genetic variance has helped breeders and farmers select for the most productive individuals via estimates of heritability, so too has the ability to use thousands of cDNAs as independent traits enabled scientists to understand differential gene expression and its genetic components (Cheung *et al.*, 2003a; Schadt *et al.*, 2003; Schena *et al.*, 1995). Rather than measuring heritability for only a single quantitative trait, such as fur coat or quality in silver foxes, microarrays provide thousands of readily observable quantitative traits on a single glass slide (Wierzbicki *et al.*, 2004). In other words, gene expression can be used as a quantitative trait. Microarrays have therefore greatly broadened the concept of heritability by using each gene as a quantitative trait to estimate heritability.

The heritability of quantitative traits, such as cDNA microarray measures, has been documented in studies using yeast, drosophila and humans (Brem *et al.*, 2002; Monks *et al.*, 2004; Wayne *et al.*, 2004). In human lymphoblast cell lines, 31% of 2,430 differentially expressed genes among 15 families had a significant heritability with a median estimate of 0.34 (Monks *et al.*, 2004). In yeast, 3,546 out of 5,727 (62%) differentially expressed genes measured among 112 *Sacchoromyces cerevisiae* segregants, had $h^2 > 0.69$ (Wayne *et al.*, 2004). Furthermore, a study looking at the heritability of gene expression in ten lines of *Drosophila simulans* found that 8% (663 /

7886) of differentially expressed genes had a median heritability of 0.34 (Brem *et al.*, 2002). These studies provide important information regarding the genetic basis of gene expression. It is evident that genetic estimates of gene expression greatly vary depending on the size of the study and the environment in which heritability is estimated. An estimate of heritability is only applicable for a particular population, in a specific environment and at a certain point in time. However, estimating heritability remains a worthwhile endeavor as it provides a foundation for understanding evolved differences in gene expression among individuals.

In *Fundulus heteroclitus*, approximately 18% (161/907 genes) of gene expression has been shown to be statistically different between individuals from the same population (Oleksiak *et al.*, 2002). A later study found that 94% (112/119 genes) of cardiac metabolic gene expression is significantly different among individuals within a population (Oleksiak *et al.*, 2005). This is a large amount of variation, however it is unknown whether the observed differences between individuals are genetic or if they are caused by environmental factors. Determination of the evolutionary significance of this large variation in gene expression requires an understanding of whether this variation is primarily due to genetic or environmental factors.

The goal of this chapter is to estimate the heritability of gene expression in *F. heteroclitus*. Estimation of heritability uses two methods; 1) regression analysis and 2) components of variance analysis. This study provides the first known estimates of heritability in metabolic gene expression for *F. heteroclitus*.

Methods and Materials

Organism

Fundulus heteroclitus were caught from wild populations in Stone Harbor, New Jersey, USA (39° 3' 3" N, 74° 45' 29" W) by baiting commercially available minnow traps with dry dog food in May, 2007. Upon capture in the field, males and females were sorted into separate bins for ease in the pairing of males and females for breeding. Gravid females and males exhibiting mating color (bright yellow) were paired together for a total of sixty breeding pairs. Eggs were physically expelled from females and placed onto moist paper towels in Petri dishes. The male from each breeding pair was then milted and the sperm was dispersed over the eggs of the respective female. The sixty breeding pairs and fertilized eggs were transported to the University of Miami's Rosenstiel School of Marine and Atmospheric Science (RSMAS). The adult breeding pairs were acclimated to 20ºC and 15 ppt artificial seawater in laboratory aquaria at RSMAS for approximately 2 months (Instant Ocean, Inc.). Adult pairs were sacrificed for their hearts which were then stored in 1 ml RNA*later* (Ambion, Inc.) for future use in microarray experiments. Fertilized eggs were hatched approximately two weeks after fertilization by submersing the eggs in 15 ppt artificial seawater and applying a flow of air bubbles for approximately 5 minutes. The fish larvae from each breeding pair were placed in separate aquaria and allowed to grow for approximately 5 months. The juveniles were then moved to a re-circulating aquarium system where all individuals shared a common aqueous environment for approximately 6 months. F1 individuals from each breeding pair were sacrificed for their hearts which were placed in 1 ml RNA*later*

(Ambion, Inc.) for use in microarray experiments. Thirteen of the sixty breeding pairs and their offspring were used for heritability studies.

Genotyping-Confirmation of Unique Families

To confirm the relationship of siblings to parents, all F1 individuals from each of the 13 breeding pairs were genotyped at 4 microsatellite loci for *F. heteroclitus* (Adams *et al.*, 2005). DNA was extracted from dried fin clips for all adults and all F1 individuals. The tissue was placed in 300 µL lysis buffer (75 mM NaCl, 25mM EDTA, 1%SDS) and incubated with 0.1 mg *Proteinase K* at 55ºC for 2 hours. Proteins were precipitated by adding a half volume of 7.5 M ammonium acetate and centrifugation at 16,000 g for 10 minutes at room temperature. DNA was precipitated from the supernatant by adding 0.7 volumes of isopropanol and centrifugation at 16,000 g for 15 minutes at room temperature. The DNA pellet was washed with 70% ethanol then allowed to air dry for 30 minutes followed by re-suspension in 50 µL 10 mM Tris-HCl pH 8.5.

Microsatellite loci were amplified using fluorescently labeled primers containing the following final concentrations: 1- 0.10 µM *ATG-B4*, 2- 0.50 µM *ATG-20*, 3- 0.07 µM *ATG-25*, 4- 0.07 µM *ATG-6* (Adams *et al.*, 2005). The 10 µL PCR reactions contained 2.5 mM $MgCl_{2,}$ 1X PCR buffer (500mM Tris-HCl, pH 9.2, 160mM $(NH_4)_2SO_4$, 22.5 mM $MgCl_2$, 20% (v/v) DMSO, 1% (v/v) Tween T 20, water to 10 ml volume), 0.2 mM dNTPs, 0.4 units Taq DNA polymerase (Promega), 50 ng DNA, and one of the four primers (see above for concentrations). The PCR thermal cycling profile consisted of 94ºC for 2 minutes, followed by 31 cycles of 94ºC for 15 seconds, 55ºC (ATG-25, *ATG-20, ATG-6*) or 50ºC (ATG-B4) for 15 seconds, and 72ºC for 30 seconds, ending with a 5 minute extension step at 72ºC. Following PCR amplification, the products from reactions 1 and 3, and the products from reactions 2 and 4 were diluted with 50 µl and 20µl of distilled water, respectively. Products from reactions 1 and 2 and products from reactions 3 and 4 were then co-loaded before electrophoresis on an ABI 3730XL Genetic Analyzer (Applied Biosystems). GENEMAPPER v4.0 (Applied Biosystems) was used to score the genotypes.

RNA isolation, labeling and hybridization

Total RNA was isolated from adult breeding pairs and F1 offspring using 4.5M guanidinium thiocyanate, 2% N-lauroylsarcosine, 50 mM EDTA, 25 mM Tris-HCl, 0.1M β-Mercaptoethanol and 0.2% Antifoam A (Sigma). The extracted RNA was further purified using RNAClean in accordance with the manufacturer’s protocols (Agencourt). The quantity and quality of the RNA was determined using a spectrophotometer (Nanodrop, ND-1000 V3.2.1) and a bioanalyzer (Agilent 2100). RNA was then converted into amino allyl labeled RNA (aRNA) using the Ambion Amino Allyl MessageAmp II aRNA Amplification kit. This method converts poly-A RNA into 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. aaUTP incorporation allows for the coupling of Cy3 or Cy5 dyes (GE biosciences) onto aRNA for microarray hybridization.

Dye labeled aRNA aliquots for each hybridization (30 pmol each of Cy3 and Cy5) were ethanol precipitated in the presence of 10µg µl^{-1} of herring sperm and -20ºC overnight. After centrifugation at 16,000 g at 4ºC, the pellets were re-suspended in 10 µl of hybridization buffer (final concentration of each labeled sample = 2 pmol μl^{-1}). Hybridization buffer consisted of 5X SSPE, 1% SDS, 50% formamide, 1mg ml^{-1} polyA, 1mg ml^{-1} sheared herring sperm carrier DNA, and 1mg ml^{-1} BSA. Microarray slides were first blocked with 5% ethanolamine, 100mM Tris pH 7.8 and 0.1% SDS for 30 minutes at room temperature. After blocking, slides were washed with 4X SSC and 0.1% SDS at 50ºC for 60 minutes. Following rinsing, slides were boiled for 2 minutes and spin-dried in a centrifuge at 14 g for 3 minutes at room temperature. Samples (10 μl) were heated to 90°C for 2 minutes, quick cooled to 42°C, applied to the slide (hybridization zone area was 10 mm x 20 mm), and covered with a cover slip. Microarray slides were placed in an airtight chamber humidified with paper soaked in 5X SSPE and incubated for 48 hours at 42°C.

Microarrays

The amount of gene specific mRNA expression was measured using *Fundulus heteroclitus* microarrays (Paschall, et al.; Oleksiak 2002, 2005, Crawford & Oleksiak, 2007). Microarrays were printed using 384 cDNAs which encode essential proteins for cellular metabolism and include 12 controls (genomic DNA and cDNA from a *Ctenophore* library with no known similarity to any vertebrate gene). All ESTs with enzyme commission numbers or associated with central metabolic pathways from a *F. heteroclitus* EST collection of over 42,000 expressed sequences were included on the array (Paschall *et al.*, 2004). The approximate average length for the printed cDNAs is 1.5 Kb. These cDNAs were amplified with amine-linked primers and printed on epoxide slides (Corning) at the University of Miami core microarray facility using inkjet technology (ArrayJet printer). Each glass slide had six separate hybridization zones (six arrays) and each array had three spatially separated replicates per gene. The experimental design used a loop pattern for hybridization of samples to the microarray (as opposed to reference design, (Kerr, Churchill, 2001)).

Dye coupled aRNA from adult breeding pairs were hybridized to slides using a single loop; the adult breeding pair loop includes 19 individuals as follows; 5M→ 7F→ 53M→ 2F→ 39M→ 55F→ 9M→ 37F→ 5F→ 39F→ 45F→ 2M → 4F→ 23M→ 3F→ 7M→ 32F→ 3M→ 9M→ 5M. Each arrow represents a single hybridization between the individual labeled with Cy5 at the head of the arrow and the individual labeled with Cy3 at the tail of the arrow. The number represents the family and the letter represents either male (M) or female (F). Note that due to the loss of several adults prior to sacrificing the fish for the extraction of hearts, only 19 of the 26 individuals from the breeding pairs were hybridized.

The F1 individuals from each breeding pair (128 individuals) were hybridized to a total of 20 slides encompassing 118 microarrays. Due to the loss of some cy3 samples and some cy5 samples, one large loop was used along with pairings of individuals with only cy3 or cy5 samples (118 individuals); 53-12→ 5-3→ 37-6→ 45-3→ 7-5→ 55-18→ 53-17→ 9-8→ 55-4→ 53-22→ 9-24→ 7-1→ 2-14→ 53-2→ 55-19→ 37-9→ 2-9→ 32-2 → 39-10→ 37-1→ 2-1→ 5-5→ 4-4→ 45-15→ 53-21→ 55-5→ 4-2→ 7-6→ 55-21→ 2-8→ 45-11→ 4-6→ 2-2→ 53-15→ 45-1→ 9-22→ 55-3→ 4-13→ 23-10→ 37-11→ 45-12→ 55-10→ 7-4→ 32-3→ 4-10→ 45-4→ 2-7→ 37-8→ 5-6 → 9-7→ 55-9→ 3-1→ 7-16→ 55-11→ 45-2→ 53-14→ 5-4→ 39-9→ 5-9→ 7-19→ 53-24→ 55-15→ 5-8→ 9-23→ 7-7→ 3-14→ 53-3→ 2-4→ 7-15→ 45-8→ 9-9→ 4-15 → 45-5→ 53-13→ 37-10→ 4-3→ 32-19→ 39-6→ 5-11→ 3-5→ 7-14→ 45-7→ 39-5, 53-4→ 32-1 → 3-8 → 53-10→

5-10→ 53-11→ 45-17→ 55-6 → 53-16→ 3-2→ 5-14→ 55-17→ 7-17→ 32-14 → 37-7→ 45-14→ 7-3→ 32-18→ 45-10→ 53-5 → 4-11→ 7-8→ 23-4→ 9-21→ 32-20→ 53-12, 2-15→ 3-3, 23-3→ 9-2, 5-2→ 23-12, 55-20→ 4-5, 7-18→ 39-12, 9-1→ 2-10, 9-10→37-15, 9-20→ 2-3, 9-4→ 23-5, 9-5→ 32-10. Each arrow represents a single hybridization between the individual labeled with Cy3 at the head of the arrow and the individual labeled with Cy5 at the tail of the arrow. The first number represents the family (breeding pair number) and the second number is the F1 individual's number. The microarray slides were scanned using ScanArray Express. The raw TIFF-image data was quantified using Imagene (v5). Control genes (*Ctenophore* negative controls) were eliminated from the analysis. Of the 384 genes printed on each microarray, 363 genes were included in the analysis for each individual for both the breeding pairs and F1 individuals.

Statistics

Statistical analyses of the mRNA expression data were carried out using SAS JMP genomics (SAS JMP Genomics v.7.02). Gene expression data for parents and offspring were first $\log_2$ transformed and then subject to loess normalization (Figure 1). The $\log_2$, loessed data were fit to a gene-by-gene mixed model ANOVA;

$$y_{ijk}= \mu +A_i+D_j+I_k+(AD)_{ij}+ \varepsilon_{ijk} \quad (2)$$

where y_{ijk} represents the fluorescence intensities on a log scale pertaining to the kth biological sample hybridized to the ith array and labeled with jth dye and μ is a constant (Wolfinger *et al.*, 2001). The term I_k is a fixed effect where I is treatment or individual effect. The terms D_j, A_i, $(AD)_{ij}$, and ε_{ijk} are random effects where A represents array effects, D represents dye effect, (AD) represents array by dye interactions and ε_{ijk} are random residual terms. The PROC Mixed statement was as follows: class Array Dye Individual Slide Family; model response= Individual; random Dye Array Dye*Array; lsmeans Individual. All analyses used the least squares means (LSmeans) from a mixed model analysis with Individuals (ignoring Family) as fixed effect and Array, Dye and Array*Dye as random effects. Two sets of analyses were performed to determine the heritability of gene expression. One set of analyses used regression based methods to determine heritability and the second set of analyses used analysis of variance to determine heritability, based on the calculation of the components of variance.

An ANOVA was used to determine the statistical significance of the variance among families. The LSmeans from the gene-by-gene mixed model were then used in a one-way ANOVA, where variance among families and F1s were the numerator and denominator, respectively to calculate the F-statistic. The following model, $y_m = \mu + F_m + \varepsilon_m$ where y_m represents the fluorescence intensities on a log scale pertaining to the mth biological sample and μ is a constant. The term F_m is a fixed effect where F represents families and the term ε represents random residual terms. The LSmeans family-based partitioning of variances provides broad (H^2) estimates of the genetic variation (V_G/V_T) for gene expression among all siblings.

The variance for each gene from the ANOVA model was used as an estimate of the heritable variation for gene expression among all siblings. The calculation for heritability is the family variance divided by the total calculated variance (V_G/V_T) (Falconer, Mackay, 1996). To estimate heritability, the mean square values for each gene from the one-way ANOVA were calculated. The following equations were used to calculate each component of variance: $\sigma_F^2 = (MS_F\text{-}MS_I)/(K_1)$, $\sigma_I^2 = MS_I$, $\sigma_T^2 = \sigma_F^2 +$

$\sigma_I^2 = (MS_F\text{-}MS_I)/(K_1) + MS_I$, where σ_F^2 represents the family component of variance (variance among families), MS_F represents the mean squares for the families, σ_I^2 represents the component of variance for all F1 individuals (siblings), MS_I represents the mean squares for the siblings, σ_T^2 represents the total component of variance and K_1 is a coefficient. Mean squares were calculated by dividing the sums of squares for families by the degrees of freedom for families and by dividing the sums of squares for siblings by the degrees of freedom for siblings. The coefficient, K_1, was calculated as follows for families with an unequal number of individuals per family: $K_1 = 1/(S-1) * [N-(\Sigma n_i^2/N)]$, where S represents the total number of families, N represents the total number of individuals and Σn_i^2 is the sum of all of the squares for the number of individuals per family (Shuster, 2006). For the sibling analysis, a total of 13 families (13 groups of F1 individuals each with different parents) were used with numbers of individuals per family ranging in size; family 2 (9), family 3 (6), family 4 (9), family 5 (7), family 7 (13), family 9 (8), family 23 (4), family 32 (7), family 37 (7), family 39 (4), family 45 (16), family 53 (15), family 55 (13); the number in parentheses are the number of individuals belonging to the family. A total of 118 individuals were used in the analysis and the value of K_1 is 8.944. Finally, heritability was calculated by dividing the family variance component by the total variance component and multiplying by 2; $h^2 = 2*(\sigma_F^2)/(\sigma_T^2)$, where h^2 is the heritability of gene expression among all siblings.

The LSmeans for each gene for siblings and parents were used for the regression analysis to estimate heritability. For all midparent-offspring regression analyses, the slope of the regression was recorded as the value for heritability (Falconer, Mackay, 1996). In addition to midparent-offspring regression analyses, the LSmeans of the female parents from 11 families and the male parents from 8 families were regressed against their respective offspring's LSmeans expression values. For female-offspring and male-offspring regression analyses, the slope of the line was multiplied by 2 to obtain the value for the heritability of gene expression. Multiplication of the slope by two is necessary because siblings have half of the genetic information from each parent. For offspring regressed to a single parent, $b=1/2\ h^2$ where b is the slope of the line or regression and h^2 is the narrow-sense heritability (Falconer, Mackay, 1996).

Additional Analyses

In addition to estimating broad and narrow sense heritability, the variation in gene expression among parents (19 individuals) was examined using ANOVA. Furthermore, to determine whether differences in age cause significant differences between individuals, variation in gene expression was examined between two groups; parents (P) and offspring (S).

Results

The expression of mRNA and genetic variation were measured for all parents and offspring. Heritability was measured using two separate analyses: 1- Components of variance analysis and 2- Parent-offspring regression analyses.

Components of Variance Analysis

Among all F1, 158 genes had significant variation in mRNA expression among individuals. Of these 158 genes, there was a significant broad sense heritability (H^2) determined for 14 genes (8.9%) ($p \leq 0.05$). The H^2 values, based on full-sibs, ranged from .19 to .43 (Table 4.1). The median heritability for the 14 genes is 0.25. Figure 4.2 provides a summary of the proportion of genes per estimate of H^2 for the 158 differentially expressed genes. The majority of these heritability estimates, 77.8% (122/158 genes), were less than or equal to 0.1 and only 1.3% (2/158 genes) of genes had heritability estimates greater than or equal to 0.4 (Figure 4.2).

Hierarchical clustering of the 13 families comprising all siblings revealed a correlation among families 2, 5, 23, 9, 3 and 39 and among families 4, 32, 7, 53, 37, 45 and 55 (Figure 4.3). The families that make up these two groups share patterns of mRNA expression for the 14 genes with significant heritability (Figure 4.3). The hierarchical clustering of all individuals among the 13 families for the set of 158 differentially expressed genes showed that genes with shared patterns of expression can be grouped according to individuals in families that group together as seen in Figure 3 (Figure 4.4).

Parent-Offspring Regression Analysis

Narrow sense heritability (h^2) estimates were obtained by executing the following regressions: 1) mid-parent (average of the two parent's LSmeans gene expression values) on offspring (6 families, 47 offspring), 2) male parent on offspring (8 families, 66 offspring), and 3) female parent on offspring (11 families, 98 offspring). Out of a total of 363 genes, 200, 187 and 196 had positive slopes for the mid-parent on offspring, female parent on offspring and male parent on offspring regressions, respectively (Table 4.2). Of these genes with detectable heritability, 13 (6.5%), 5 (2.7%) and 12 (6.1%) genes had significant h^2 at $p \leq 0.05$ for the mid-parent, female and male regressions, respectively (Table 4.2).

The power to detect heritability at 0.4 for the three parent-offspring regressions are: 28.9% (mid-parent- offspring regression), 22.9% (male-offspring regression) and 37.4% (female-offspring regression). The mid-parent- offspring regression analysis had the largest number of genes with significant heritability (13 genes) followed by the male-offspring analysis (12 genes) and lastly, the female-offspring analysis (5 genes) (Table 4.3). For the set of genes with significant heritability, the median heritability estimates were quite high; .861 (mid-parent analysis), .729 (male parent analysis) and .875 (female parent analysis) (Table 4.3). The gene, NADH-ubiquinone oxidoreductase chain 2, was the only gene found to have significant heritability in all three analyses (Table 4.3, Figure 4.5). Acetyl Co-A carboxylase 1, Heterogeneous nuclear ribonucleoprotein A/B, NADH-ubiquinone oxidoreductase chain 2 and Nuclear factor erythroid 2 related factor, were found to have significant heritability in both the mid-parent- offspring and male-offspring regression analyses (Table 4.3, Figure 4.5). Three genes were found to have a shared, significant heritability in both the female-offspring and mid-parent offspring analyses; Glutathione S-transferase, NADH-ubiquinone oxidoreductase chain 2 and NADH-ubiquinone oxidoreductase chain 4 (Table 4.3, Figure 4.5). There are no shared genes between the male-parent and female-parent on offspring regressions. Figure 6 provides each of the significant regressions at $p \leq 0.05$ for mid-parent (13), male (12) and female (5) versus offspring. The parent LSmeans gene expression is represented on the x-axis and the offspring LSmeans gene expression is on the y-axis. Confidence interval shading at 95% is in pink. The standard errors for each gene are located in Table 4.3.

Although the regression analysis and the components of variance analysis use two very different approaches, the expectation is that those genes with highly significant, high heritability would overlap between the two approaches. However, very few of the genes were found to overlap between the regression analysis and the components of variance analyses. Only three genes overlapped between the two analyses; Heterogeneous nuclear ribonucleoprotein A/B, Hypoxia-inducible factor 1 alpha and NADH-ubiquinone oxidoreductase chain 2 (Table 4.1 and Table 4.3). These 3 genes were found to have rather low heritability in the components of variance analysis; Heterogeneous nuclear ribonucleoprotein A/B (H^2=0.25), Hypoxia-inducible factor 1 alpha (H^2=0.22) and NADH-ubiquinone oxidoreductase chain 2 (H^2=0.21) (Table 4.1). However, NADH-ubiquinone oxidoreductase chain 2 had h^2 of 0.92, 0.95 and 0.92 in the mid-parent, male parent and female parent versus offspring regression analysis, respectively (Table 4.3). Heterogeneous nuclear ribonucleoprotein A/B had h^2 of 0.9 and 0.72 for the mid-parent and male parent versus offspring regressions, respectively (Table 4.3). Finally, Hypoxia-inducible factor 1 alpha, was found to have a significant heritability in the male-offspring regression analysis of 4.84 (standard error 0.95) (Table 4.3). This is an extremely high estimate of heritability and is also beyond the normal values for heritability which range from 0.0-1.0.

Additional Analyses

In addition to estimating broad and narrow sense heritability, the variation in gene expression among parents (19 individuals) was examined. Statistically significant differences in expression between adults (parents) for 38% of 363 genes were observed at $p \leq 0.05$. To determine whether differences in age cause significant differences between individuals, variation in gene expression was examined between two groups; parents (P) and offspring (S). The 19 parents were compared to the 118 offspring using age and of 363 genes 167 (46%) were differentially expressed.

Discussion

The majority of our understanding of the genetics of mRNA expression has relied on the study of inbred strains (Gibson, Weir, 2005; Schadt *et al.*, 2003; Wayne *et al.*, 2004) or cell culture (Monks *et al.*, 2004). Though heritability has been measured in a variety of organisms, the genetic basis for the variation in mRNA expression among natural populations, including *F. heteroclitus,* is not well understood. Estimates of heritability using inbred strains or cell culture are often criticized for being inflated in comparison to heritability estimated from natural populations due to differences in environmental heterogeneity (Astles *et al.*, 2006; Hoffmann, Merila, 1999; Weigensberg, Roff, 1996). Therefore, studies of heritability using natural populations are preferred to those carried out in the laboratory. Care should be taken when estimating heritability as the methods used to measure genetic correlations (components of variance analysis using ANOVA or regression of gene expression means of parents on offspring) each have inherent weaknesses. These weaknesses include, but are not limited to, the power to detect heritability and the possibility of obtaining negative estimates or zero values for components of variance (Windig, 1997). These factors can influence the heritability estimation, thereby decreasing the significance of the findings for natural populations (Hoffmann, Merila, 1999; Windig, 1997). However, with appropriate experimental design, heritability estimates can be successfully measured. Despite these problems, estimates of heritability were successfully obtained for *F. heteroclitus*.

The data presented here suggest that there is a significant heritability for variation in mRNA expression in *Fundulus heteroclitus*. Heritability as estimated from parent on offspring regressions and components of variance analysis provide similar percentages of genes with significant heritability. The relatedness of individuals from the same family in the full-sib components of variance analysis was assumed to add a significant amount to the genetic component of gene expression. However, of the 158 differentially expressed genes from this analysis, 8.9% are heritable with a median H^2 value of 0.25 when $p \leq 0.05$ (Table 4.1). Broad sense heritability is the genetic variance divided by the phenotypic variance (V_G / V_P). V_G includes one-half additive genetic variance (V_A) for full siblings plus some epistatic effects (epistatic effects are assumed to be small, if any). However, dominance variance (V_D) and maternal effects were not measured and therefore were not included as variance components because this confounds the true value of V_G. Therefore, broad sense rather than narrow sense heritability is estimated for the full-sibling analysis.

In contrast, narrow-sense heritability estimates from the regression analyses found that of the genes detected to have a positive slope, 6.5%, 2.7% and 6.1% were heritable for the mid-parent, female and male on offspring regressions, respectively (Table 4.2). The mid-parent- offspring regression estimates of heritability had the greatest percentage of genes (6.5%) with significant heritability where the median $h^2 = 0.86 \pm 0.19$ (Table 4.3). This is similar to results found in *Drosophila* where 8.4% of differentially expressed genes had a median h^2 of 0.47 (Wayne *et al.*, 2004). Of the three regression analyses, the mid-parent on offspring analysis is the most powerful as the expression values for both parents are represented in the regression.

Understanding the inherent difficulties of using a particular method with a particular experimental design can help to evaluate any introduced bias (Hoffmann, Merila, 1999; Windig, 1997). Regression of mid-parent lsmeans expression on offspring lsmeans expression gives better precision than just a single parent under most circumstances (Falconer, Mackay, 1996). In terms of variance, estimating heritability using full-sib analyses are twice as precise as those estimated from half-sib analyses (Falconer, Mackay, 1996). Though many studies prefer to use one method over the other, both methods can provide precise estimates of heritability as long as the most optimal experimental design is constructed (i.e. large sample size, power, distribution of data) (Astles *et al.*, 2006; Falconer, Mackay, 1996; Windig, 1997).

Of the methods used to estimate h^2, the mid-parent on offspring regression analysis heritability estimates are more precise because the relationship between mid-parent and offspring is least likely to be influenced by the environment or dominance. Thus, the mid-parent analysis is a more appropriate method for estimating heritability. In addition, the mid-parent offspring relationship uses the variance in gene expression from both parents thereby providing a more precise estimate of heritability than either parent on offspring (Falconer, Mackay, 1996).

The ability to obtain estimates of heritability is important because the choice of the environment in which heritability is measured can affect the reliability of these estimates. Interestingly, we have seen, as described in Chapter 3, that the variance in individuals acclimated to laboratory conditions is greater in comparison to the variance of those individuals captured in the field therefore suggesting that environmental variation in the field (tidal changes, spatial and temporal changes in salinity, food availability, oxygen, etc.), does not have a major affect on the variation in gene expression (Marshall, 2003; Marshall *et al.*, 2005). This is opposite to the assumption that the environment increases variation in gene expression.

Heritability is dependent upon the environment in which it is measured. Even for traits where the variation is fully under genetic control, there can be little or no h^2 because the environment suppresses the expression of the phenotypic variation. For example, human height and thorax length in *Drosophila melanogaster* are heritable traits affected by nutrition (Bubliy *et al.*, 2001). In both the offpring-parent regression and estimates of genetic variation analyses, the inbred, mixed population and mixed population F1 individuals that experienced limited-food conditions had decreased thorax length, but no significant differences in the heritability between the two environments were found (Bubliy *et al.*, 2001).

The full sib analysis measured heritability for individuals raised in a common environment. While only one environment was used in this study, small differences in environmental conditions between aquaria cannot be ruled out. By controlling for only one environment, one that perhaps is more conducive to growth and survival, the genetic variance may have been unknowingly lowered, thereby influencing the true value of H^2. This may explain why the heritability estimates from the full-sib components of variance analysis were much lower in comparison to the regression analyses. The same F1 individuals that were used in the components of variance analysis were also used in the regression analyses. All F1 individuals were reared in laboratory aquaria in a common environment; however, the parents were caught in the field and were only acclimated to laboratory conditions for approximately 2 months. As mentioned above, the added environmental influence due to the difference in acclimation time between parents and offspring may explain the high estimates of heritability for the parent-offspring regression analyses.

Though the heritability estimates from the regression analyses were high and were also found to have high error margins, they are most likely more reliable than the full-sib analysis which suffers from potential decreases in variance estimates and potentially large amounts of dominance variance which was not measured (Tables 4.1 and 4.3). Narrow-sense heritability estimates are indeed much stronger estimates than broad-sense estimates because the numerator for narrow-sense heritability has only additive genetic effects whereas broad-sense heritability has both additive and non-additive effects in the numerator (Devlin *et al.*, 1997). Therefore, broad sense heritability should be larger than narrow-sense heritability. The data here shows otherwise, but can be explained by the observation of low measures of variance due to individuals raised in a common environment. Despite these problems, the estimates of heritability that were measured using both methods, components of variance and regression analysis, are significant as no study has attempted to measure the heritability of gene expression in *Fundulus heteroclitus.* It is important to note, however, that studies of heritability using non-standard techniques, such as protein level comparisons between parents, F1, F2 and F3 generations in *F. heteroclitus* have been performed (Meyer *et al.*, 2002).

Though the estimate of heritability for the components of variance analysis is lower (median $H^2 = 0.25$), is not unlike other studies of heritability. For example, in ten lines of *Drosophila,* 663 out of 7886 measured genes (8.4%) had significant genetic variation with a median h^2 of 0.47 (Wayne *et al.*, 2004). In studies where sample sizes are much larger, the percentage of genes with detectable heritability greatly increases. For example, among 112 *Sacchoromyces cerevisiae* segregants, 3,546 out of 5,727 measured genes (62%) had $h^2 > 0.69$ (Brem, Kruglyak, 2005). It is evident that with an increase in sample size, there is also greater power to detect h^2. Reasons for the low percentages of genes detected to have significant heritability in this study can be attributed to sample size and power.

The power to detect heritability at 0.4 was only 28.9% for the mid-parent on offspring regression whereas the power to detect heritability at 0.4 using 13 full-sib families and an average sample size of 9 individuals per family was 46%. Our null hypothesis is that there is no heritability of gene expression; therefore, we will reject the null hypothesis approximately 46% of the time which leaves us unable to detect heritability for 54% of the genes. Thus, many more genes may have significant, heritable variation, but due to small sample sizes, we lacked the power to detect these genes. The number of genes detected with significant heritability between the two analyses was almost equivalent (14 genes-full-sib, 13 genes mid-parent on offspring regression), but the actual heritability estimates were quite different (Tables 4.1 and 4.3). Again, effects of dominance and environment could have influenced the actual broad-sense estimates. Interestingly, the male on offspring regression estimates of heritability are stronger than the female on offspring estimates. This can be explained by the influence of maternal effects which are not present in the male on offspring estimates.

The 14 genes with significant heritability found in the full-sib analysis have 2 genes that are encoded by the mitochondria: Cytochrome c oxidase polypeptide II and NADH-ubiquinone oxidoreductase chain 2 (Table 4.1). These 2 genes encode proteins that form the many subunits of Complex I and Complex IV of the electron transport chain. These genes play an important role in the transfer of electrons in cellular respiration. The remaining 11 of the 14 genes with significant heritability produce proteins that are involved in many facets of metabolism, most importantly, ATP synthesis.

In all three parent on offspring regression analyses and in the full-sib analysis, NADH-ubiquinone oxidoreductase chain 2, was found to have significant heritability (Tables 4.1 and 4.3). Additionally, NADH-ubiquinone oxidoreductase chain 4 was found to have significant heritability in the mid-parent and female on offspring regression analyses (Table 4.3). Finally, NADH-ubiquinone oxidoreductase chain 3 was found to have significant heritability in the mid-parent-offspring regression analysis (Table 4.3). These findings are significant because NADH-ubiquinone oxidoreductase (Complex I), the largest of the membrane bound respiratory chain enzymes, is responsible for catalyzing the first step in the electron transport chain (Saraste, 1999). Analysis of Complex I isolated from bovine hearts revealed that it is comprised of 46 different subunits of which 7 are encoded by the mitochondrial DNA (Carroll *et al.*, 2002). Immunopurification of human NADH dehydrogenase (Complex I) confirmed the presence of 42 of the 46 NADH subunits found in bovine hearts (Murray *et al.*, 2003). Therefore, the finding that these Complex I genes have significant heritability is of importance since this enzyme is a critical step in the electron transport chain. Deficiencies or mutations in this enzyme are known to cause Mitochondrial Encephalopathy (MELAS Syndrome), myopathy and fatal infantile multisystem disorder (Ravn *et al.*, 2001). It is unknown how many NADH subunits comprise complex I in fish mitochondria or whether deficiencies in this complex cause similar diseases.

Another gene worth examining that has significant heritability in both the mid-parent-offspring and female-offspring regressions is Glutathione-S transferase. This gene is involved in the metabolism of xenobiotics. *Fundulus heteroclitus* live in estuarine areas along the eastern seaboard of the United States. Due to the proximity of this habitat to human influences, such as dumping and contamination, upregulation of Glutathione-S transferase is important when fish are in contact with chemicals such as polycyclic aromatic hydrocarbons (PAHs)(Escartin, Porte, 1999).

Cluster diagrams of the 190 differentially expressed genes and the 14 genes with significant heritability share a similar pattern (Figures 4.2 and 4.3). The cluster diagram for genes with significant differences among individuals has four rather distinct quadrants of gene expression (Figure 4.3). Upon closer examination, the individuals that cluster together to form these quadrants are the same individuals from the families that cluster together to create the pattern seen in Figure 4.2. Though not all the individuals from each of the families that grouped together in Figure 4.2 clustered together in Figure 4.3, it is important to note that many of those individuals did cluster and that the pattern of expression for all individuals is a function of differences between families.

The present study is limited by sample size, however, the data presented here are the first to formally estimate the genetic component of gene expression in *F. heteroclitus*. Using the most reliable analysis of the heritability of gene expression, mid-parent-offspring regression, 6.5% of genes have significant heritability with a median h^2 of 0.86. Though this represents only a small proportion of genes included in the analysis, it is a significant finding. Improvements to the design of this study would include larger sample sizes and perhaps performing the study on F_2 and possibly F_3 generation fish. In addition, controlling for environmental differences between parents and offspring should be addressed. Gene expression is determined by both genetic and environmental factors.

Estimates of heritability in this study indicate that the environment has more of a contribution to gene expression than do genetics. *F. heteroclitus* are exposed to variable environments day to day and even hour to hour. This requires the ability of a genotype to produce different phenotypes when exposed to different environments (Gibson, 2008). Therefore, it is interesting that in this study, some of the genes that are heritable are those that are involved in the response to hypoxia and xenobiotics: Glutathione-S transferase, Hypoxia-inducible factor 1-alpha.

The genetic basis for differences in the variation of gene expression varies from organism to organism and is dependent upon the environment in which gene expression is measured. Perhaps with greater sample sizes, the ability to partition the genetic component of variance would be stronger thereby allowing for a more robust estimate of heritability. Nevertheless, our findings are similar to those reported in studies using inbred strains. This study is the first to estimate the heritability of gene expression for *Fundulus heteroclitus.* More importantly, this study provides the foundation for understanding the genetic contribution of differential gene expression. Natural, heritable, variation in gene expression is important for understanding the evolution of the genes that control gene expression. Heritable variation is the raw material for evolutionary processes and thus, measures of heritability are critical for understanding how differences in mRNA expression evolve.

Table 4.1. Estimates of variance components and broad-sense heritability for 14 significant genes (both differentially expressed and significantly heritable) at $p \leq 0.05$.

		Variance Components				Heritability	Probability (Family)
Gene	F	Family (V_G)	Individual(V_W)	Total (V_P)	K	H^2	p
60S ribosomal protein L28	1.87	0.01	0.12	0.13	8.94	0.19	0.05
ADP,ATP carrier protein, isoform T2	2.26	0.03	0.18	0.21	8.94	0.26	0.01
Cytochrome c oxidase polypeptide II	2.16	0.02	0.14	0.16	8.94	0.25	0.02
Cytochrome c oxidase polypeptide VIa, mitochondrial precursor	3.13	0.04	0.16	0.20	8.94	0.40	<0.001
Dihydropyrimidinase	3.27	0.03	0.13	0.16	8.94	0.43	<0.001
Glutaminase, liver isoform, mitochondrial precursor	1.99	0.01	0.09	0.11	8.94	0.22	0.03
Heterogeneous nuclear ribonucleoprotein A/B	2.10	0.01	0.08	0.09	8.94	0.25	0.02
Hypoxia-inducible factor 1 alpha	1.98	0.01	0.11	0.12	8.94	0.22	0.03
Inositol-3-phosphate synthase	2.07	0.01	0.07	0.08	8.94	0.25	0.02
NADH-ubiquinone oxidoreductase chain 2	2.02	0.06	0.51	0.57	8.94	0.21	0.03
Platelet-activating factor acetylhydrolase precursor	2.33	0.01	0.06	0.08	8.94	0.30	0.01
Tenascin X precursor	1.97	0.02	0.20	0.22	8.94	0.21	0.03
Transaldolase	2.14	0.01	0.09	0.10	8.94	0.25	0.02
Troponin C, slow skeletal and cardiac muscles	1.86	0.02	0.15	0.16	8.94	0.19	0.05

Table 4.2. The percentage of genes from the regression analyses with significant h^2. The percentage was calculated by dividing the number of genes with significant h^2 by the number of genes with positive slopes.

	No. genes with positive slopes/h^2	**No. genes with significant h^2**	**% genes with significant h^2**
Midparent	200	13	6.5
Female	187	5	2.7
Male	196	12	6.1

Table 4.3. Genes with significant heritability for each of the three regression analyses. Midparent-offspring and male-offspring regression analyses had four genes with significant heritability in common. Midparent-offspring and female-offspring analyses had three genes with significant heritability in common. Female-offspring and male-offspring regression analyses had only one gene in common.

Table 4.3 (cont.)

	Genes with significant heritability	**Slope**	**h2**	**p≤0.05**	**Std Error**	**95% Confidence**
Midparent	Acetyl-CoA carboxylase 1	0.77	0.77	0.01	0.30	0.59
	ADP-ribosylation factor	0.91	0.91	0.02	0.38	0.74
	AMBP protein precursor [Contains: Alpha-1-microglobulin; Inter-alpha-trypsin inhibitor light chain]	1.00	1.00	0.04	0.47	0.92
	Cold-inducible RNA-binding protein	1.90	1.90	0.01	0.67	1.32
	Glutathione S-transferase	0.72	0.72	0.02	0.29	0.57
	Heterogeneous nuclear ribonucleoprotein A/B	0.90	0.90	0.04	0.43	0.84
	NADH-ubiquinone oxidoreductase chain 2	0.92	0.92	0.03	0.40	0.79
	NADH-ubiquinone oxidoreductase chain 3	0.80	0.80	0.01	0.29	0.57
	NADH-ubiquinone oxidoreductase chain 4	1.10	1.10	0.03	0.49	0.95
	Nuclear factor erythroid 2 related factor	0.86	0.86	<0.001	0.19	0.37
	Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 protein	0.82	0.82	0.02	0.35	0.68
	Phosphoglycerate kinase 1	0.66	0.66	0.01	0.25	0.49
	Phosphomannomutase	0.77	0.77	0.02	0.33	0.64
Male	Acetyl-CoA carboxylase 1	0.24	0.48	0.03	0.11	0.21
	Alcohol dehydrogenase	0.50	0.99	0.05	0.25	0.49
	Betaine--homocysteine S-methyltransferase	0.26	0.52	0.003	0.09	0.17
	Calmodulin	0.40	0.80	0.01	0.15	0.30
	Heterogeneous nuclear ribonucleoprotein A/B	0.36	0.72	0.01	0.14	0.27
	Hypoxia-inducible factor 1 alpha	2.42	4.84	0.01	0.95	1.86
	NADH-ubiquinone oxidoreductase chain 2	0.48	0.95	0.04	0.23	0.45
	Nuclear factor erythroid 2 related factor	0.29	0.58	0.01	0.10	0.20
	O-methyltransferase	0.95	1.90	0.02	0.40	0.78
	StAR-related lipid transfer protein 13	0.29	0.58	0.02	0.12	0.24
	Telomerase-binding protein p23	0.10	0.19	0.04	0.05	0.09
	Vacuolar ATP synthase subunit C	0.37	0.74	0.03	0.17	0.33
Female	Glutathione S-transferase	0.27	0.53	0.05	0.13	0.26
	NADH-ubiquinone oxidoreductase 20 kDa subunit, mitochondrial precursor	0.44	0.87	0.03	0.20	0.40
	NADH-ubiquinone oxidoreductase B17 subunit	0.28	0.57	0.04	0.14	0.27
	NADH-ubiquinone oxidoreductase chain 2	0.46	0.92	0.02	0.20	0.38
	NADH-ubiquinone oxidoreductase chain 4	0.55	1.11	0.01	0.20	0.40

Figure 4.1. Kernal density overlay plot. Shows the univariate distributions of genes for all siblings for log transformed, loessed data. Each line represents a single individual. Loess normalization removes intensity dependent effects.

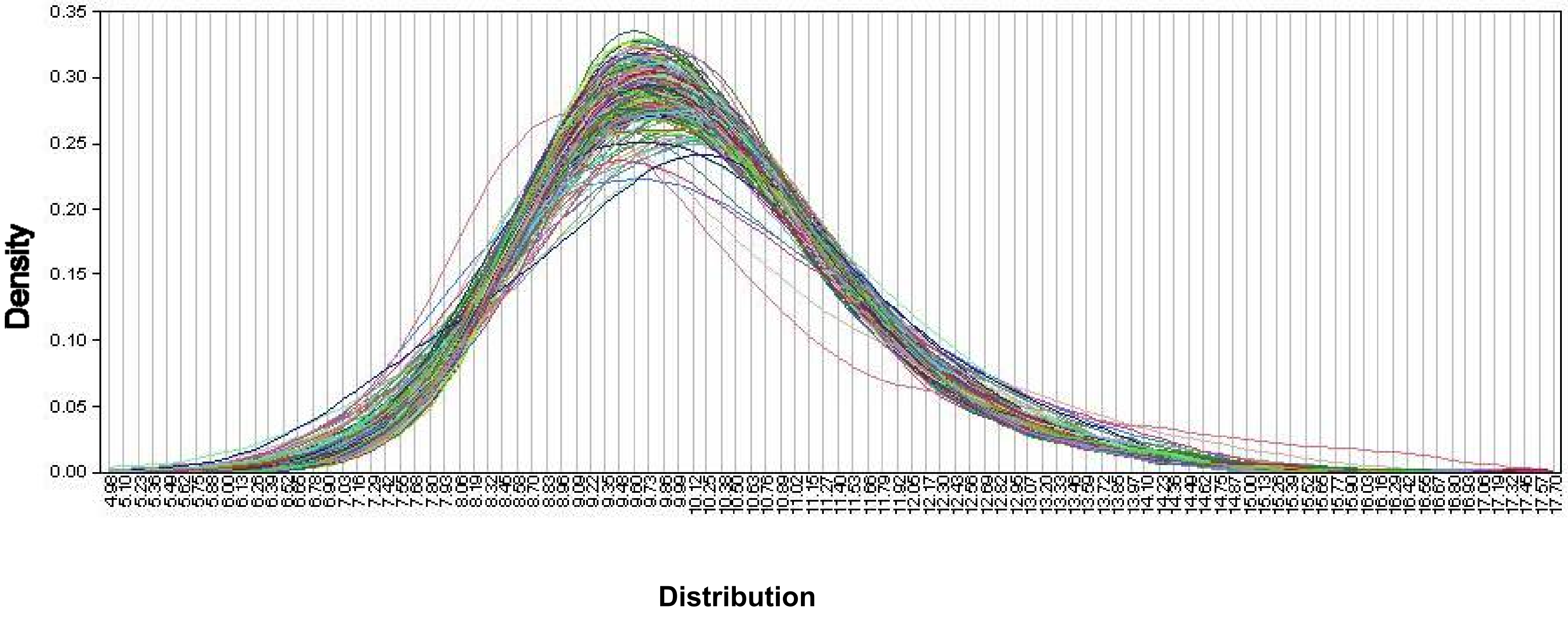
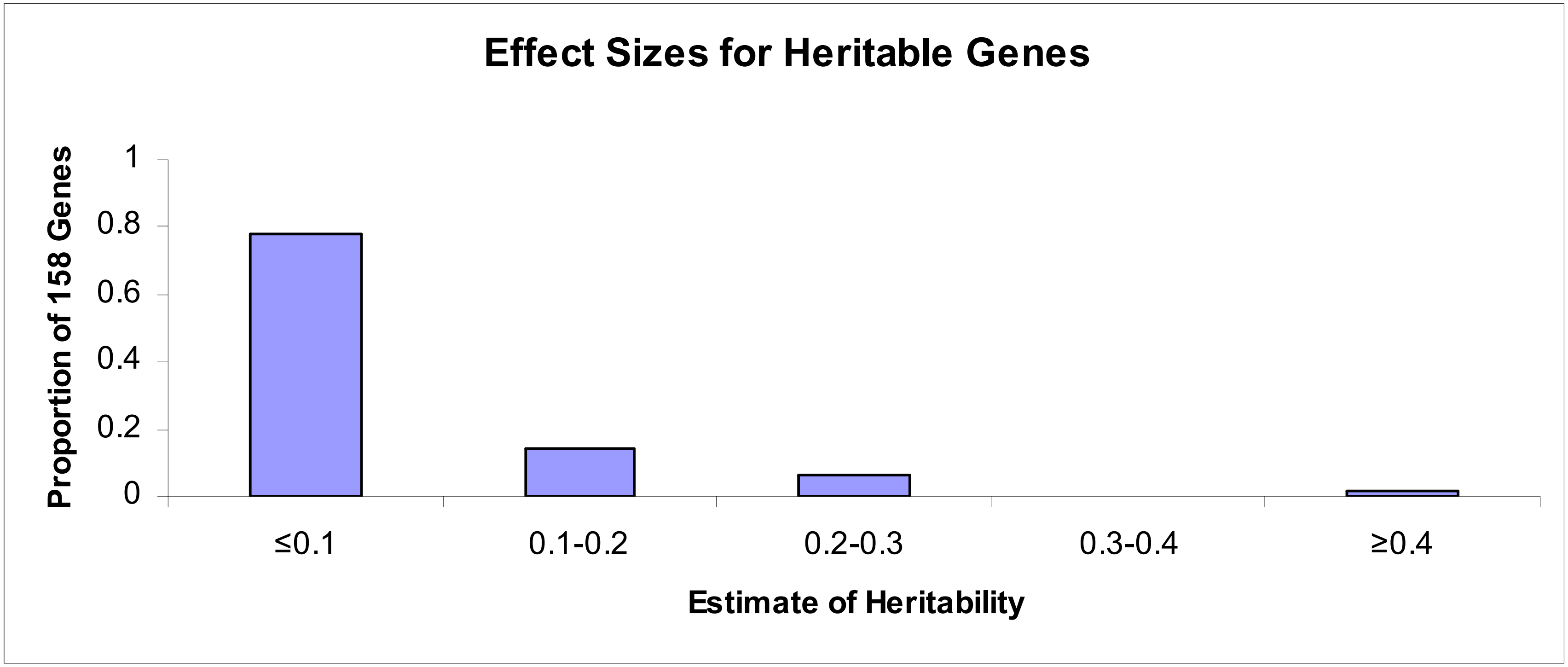


Figure 4.2. Proportion of genes per estimate of heritability. The figure shows the distribution of heritability for 158 genes.



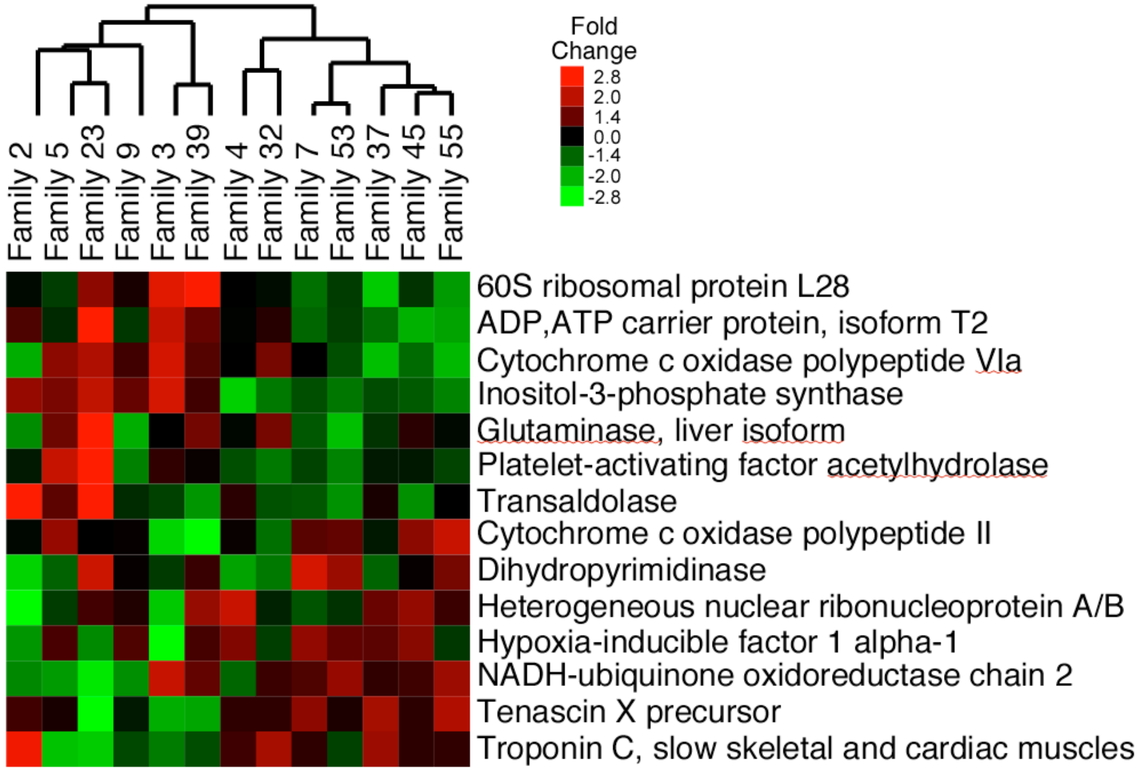


Figure 4.3. Genes with significant differences among families. Cluster diagram of 13 families and the 14 genes that are both differentially expressed and significantly heritable at $p \leq 0.05$.

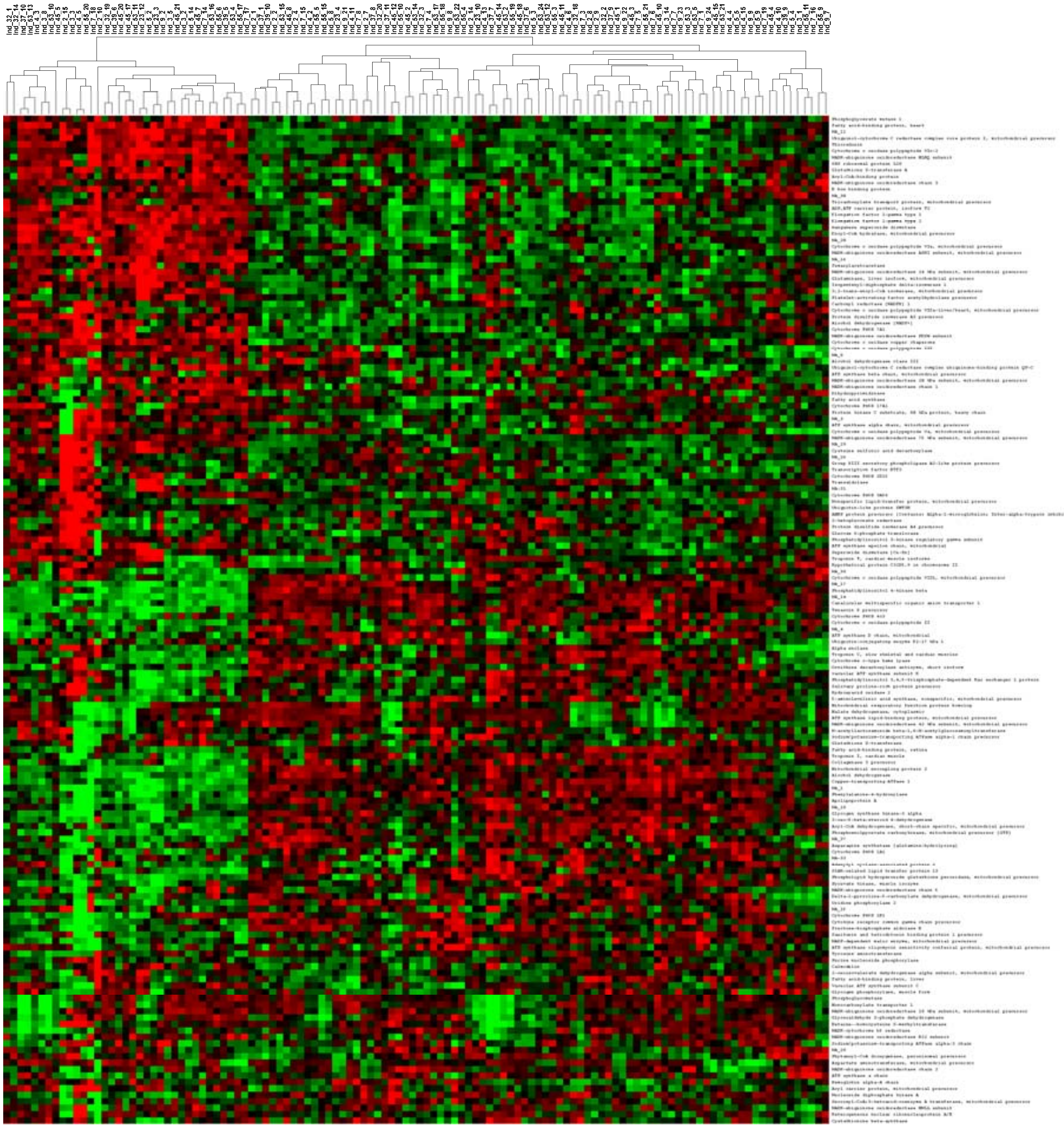


Figure 4.4. Hierarchical clustering of standardized LSmeans with significant differences among individuals. Cluster diagram of all individuals for 13 families and the 158 genes that differ significantly at $p \leq 0.05$.

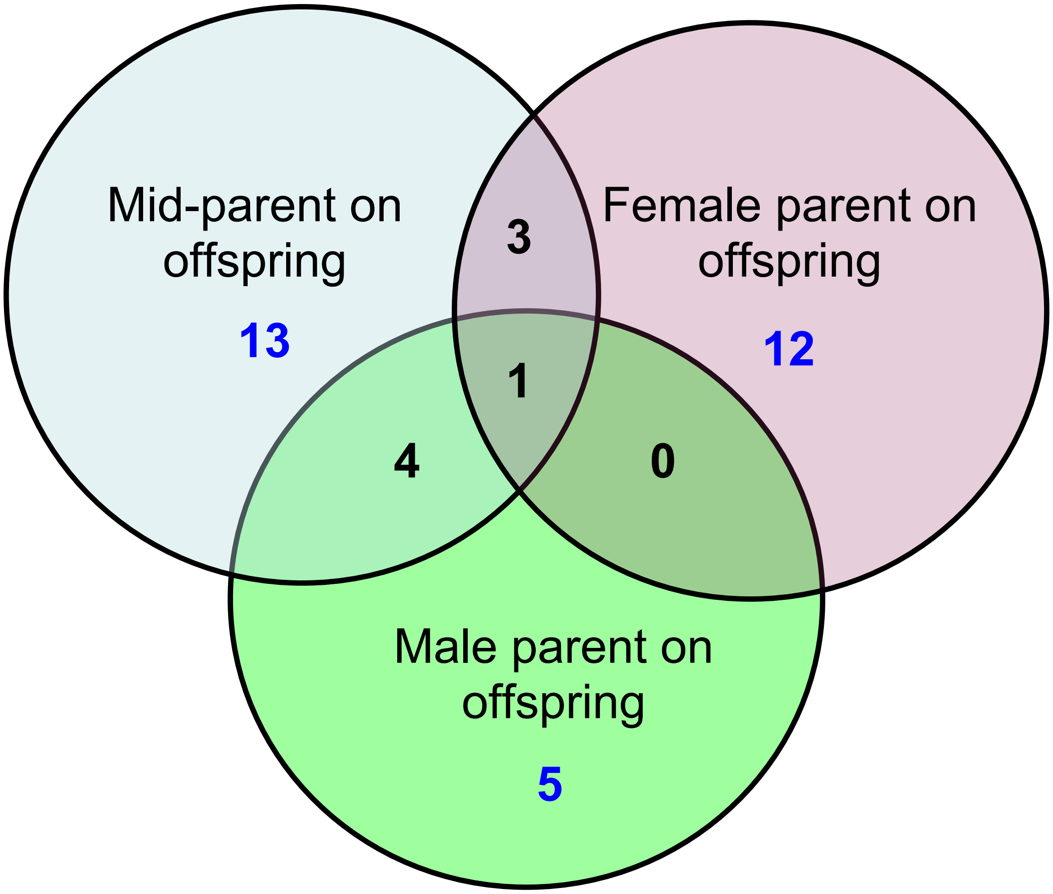
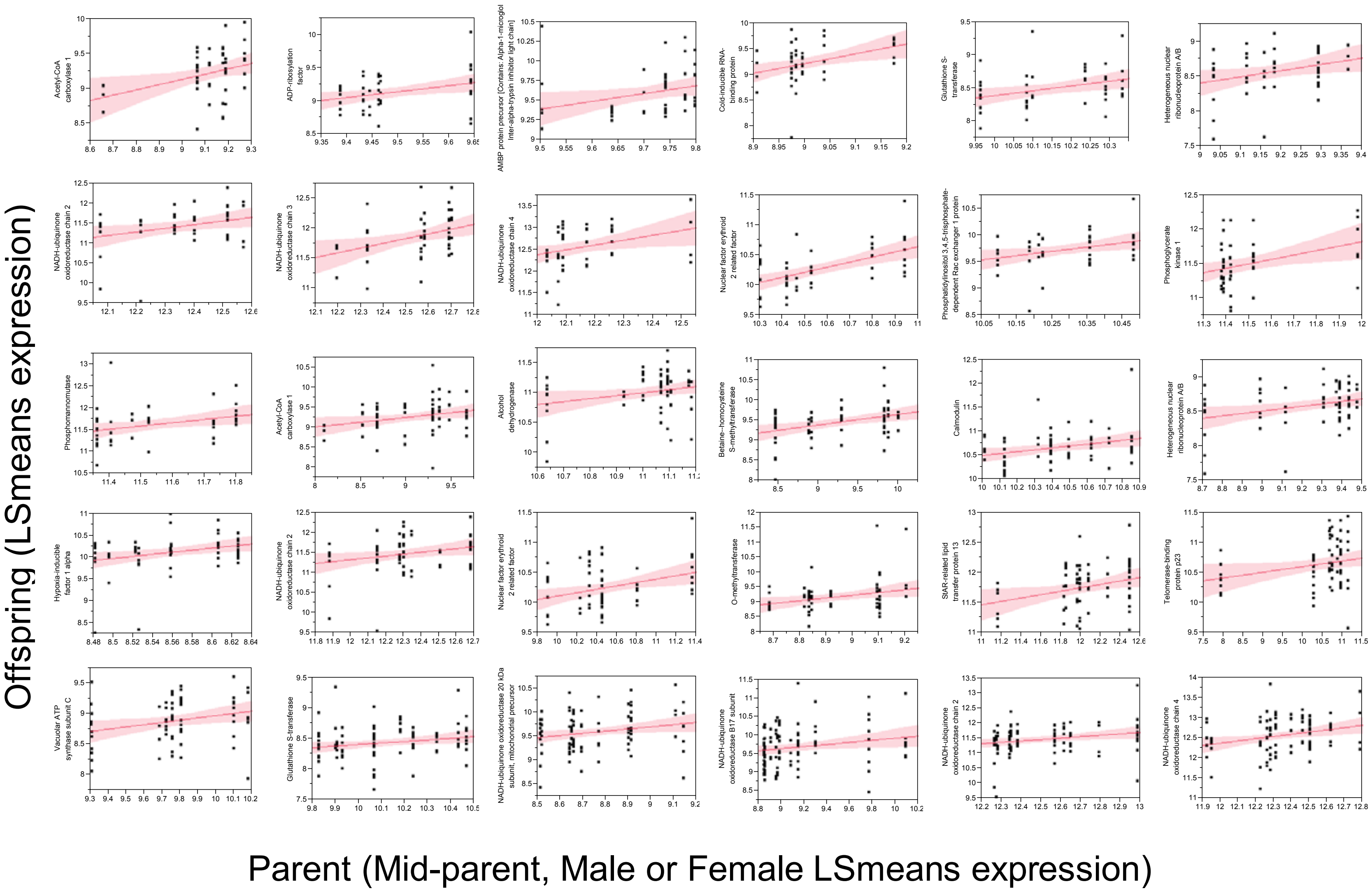


Figure 4.5. Venn diagram showing the number of significantly heritable genes for each of the three regression analyses (mid-parent on offspring, male parent on offspring and female-parent on offspring; numbers in blue) and the number of those genes that are shared between each analysis (numbers in black).

Figure 4.6. Mid-parent, female parent or male parent and offspring regressions for genes with significant heritability at $p \leq 0.05$. Graph shows regression patterns for significantly heritable genes. Reading left to right, the first 13 graphs are mid-parent-offspring regressions. The following 10 graphs are male-offspring regressions. The last 5 graphs are female-offspring regressions. The y-axis is the gene name as listed in Table 4.3.

Figure 4.6. (contd.)



CHAPTER 5: SUMMARY AND CONCLUSIONS

The focus of this dissertation was to determine the importance of heritable factors which explain the differentially expressed genes among *F. heteroclitus* individuals. The information discussed in the preceding chapters provides a foundation for understanding the genetic basis of gene expression in *Fundulus heteroclitus*. The second chapter discussed in detail the technical variation of cDNA microarrays. To measure the random biological variation found among individuals, we used *F. heteroclitus* cDNA microarrays. The use of these arrays to quantify mRNA expression is relatively precise and has a large dynamic range. There is a linear relationship between fluorescence of RNA samples and RNA concentrations ranging from 0.1X to 10X for the majority of genes (88%). Examination of the technical variation in gene expression between the four RNA samples isolated from a single blood sample found that the CV (standard deviation/mean) among eight replicates was 4% and, only three genes had a CV > 10%. Therefore, when quantifying gene expression in the same individuals there are few differences in mRNA expression, suggesting that adult mRNA expression is stable.

In chapter 3, there was an explicit test of the importance of genetic and environmental variation on mRNA expression. Three groups of individuals from the same population were compared: individuals inbred for ten generations (G10) with significantly less genetic variation than outbred individuals; outbred individuals acclimated (Acclimated) to the same environment as the inbred individuals and outbred individuals sampled directly in the field (Field). The analyses of these data indicate that there is little difference in the magnitude of individual variation among field and acclimated individuals. In contrast, G10 individuals had significantly less variation in mRNA expression. 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. The most important finding for the study conditions in chapter 3 was that environmental factors do not necessarily add to differences in the variance found in gene expression between groups.

To provide quantitative support for the experimental observation made in chapter 3, the heritability of mRNA expression was measured. Two approaches were used to estimate heritability 1) parent-offspring regression and 2) components of variance analyses. The parent-offspring regressions and the components of variance analyses had similar percentages of genes with significant heritability, 6.5% and 8.9% for mid-parent on offspring and components of variance analyses, respectively. However, the median heritability estimate for the mid-parent on offspring regression was 0.86 and the median for the components of variance analysis was 0.25. The difference in the two estimates is not cause for alarm as two separate methods were used, each with their own sets of strengths and weaknesses as discussed in chapter 4.

Overall, three major conclusions can be drawn from the work presented in this dissertation: 1) There is little technical variation between *F. heteroclitus* microarray measures 2) Influences from the environment (including but not limited to: tidal changes, spatial and temporal changes in salinity, food and oxygen availability) account for few differences in the variation in gene expression and 3) There is some but not a lot of significant genetic variation based on the quantitative estimates of heritability.

Future directions

Our studies are limited by sample size. Increased sample sizes would allow for a greater power to detect heritability and therefore increase our understanding of the genetic components of variance. Our studies were also limited to a laboratory environment. Although it appears that the natural environment does not have a large affect on the variation of differential gene expression, it would be worthwhile to subject F1 individuals and their parents to natural environments and repeat the heritability estimates. Again, heritability is only relevant to the environment in which it is measured. Obtaining estimates of heritability from both natural environments and laboratory environments can provide a more cohesive understanding of the genetic basis of gene expression across environments.

Fish from Maine and Georgia have evolved differences in gene expression patterns due to exposure to different thermal regimes. Since heritability estimates are dependent upon the environment in which they are measured, estimates of heritability would be expected to vary depending on whether fish are exposed to cold or warm temperatures. Measuring the heritability of gene expression in fish from Maine and Georgia would be worthwhile as differences in the estimates of heritability could help to explain evolutionary adaptation.

Conclusions

The research presented in this dissertation confirms that there is a genetic component to the variation in gene expression. However, it is evident that the environment in which these genetic differences are found is an important factor in determining the extent to which significantly heritable genes can be detected. Differential gene expression varies greatly between individuals from populations of *Fundulus heteroclitus*. The quantification of the variation in gene expression between individuals is important for understanding how much of this variation in expression is explained by evolution by natural selection. More importantly, estimates of the genetic component of the variation in gene expression allows for more specific detection of deficiencies or mutations in particular regions of a genome. Studies using natural populations of *Fundulus heteroclitus* remain important to the overall understanding of genetic variation and its biological importance.

REFERENCES

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.

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.

Andreas P, Weber KL, Weber K, Carr CW, Ohlrogge JB (2007) Sampling the Arabidopsis Transcriptome with Massively Parallel Pyrosequencing. *Plant Physiology* **144**, 32-42.

Antipova AA, Tamayo P, Golub TR (2002) A strategy for oligonucleotide microarray probe reduction. *Genome Biology* **3**.

Astles PA, Moore AJ, Preziosi RF (2006) A comparison of methods to estimate cross-environment genetic correlations. *Journal of Evolutionary Biology* **19**, 114-122.

Auburn RP, Kreil DP, Meadows LA*, et al.* (2005) Robotic spotting of cDNA and oligonucleotide microarrays. *Trends in Biotechnology* **23**, 374-379.

Avise JC (1989) Gene trees and organismal histories - A phylogenetic approach to population biology. *Evolution* **43**, 1192-1208.

Bammler T, Beyer RP, Bhattacharya S*, et al.* (2005) Standardizing global gene expression analysis between laboratories and across platforms. *Nature Methods* **2**, 351-356.

Barker JSF, Robertson A (1966) Genetic and phenotypic parameters for the first three lactations in Friesian cows. *Animal Production* **8**, 221-240.

Baugh L, Hill A, Brown E, Hunter C (2001) Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Research* **29**.

Beissbarth T, Fellenberg K, Brors B*, et al.* (2000) Processing and quality control of DNA array hybridization data. *Bioinformatics* **16**, 1014-1022.

Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* **422**, 633-637.

Bloom G, Yang IV, Boulware D*, et al.* (2004) Multi-platform, multi-site, microarray-based human tumor classification. *American Journal of Pathology* **164**, 9-16.

Botwell D, Sambrook J (2003) *DNA Microarrays: A Molecular Cloning Manual* Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

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 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1572-1577.

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752-755.

Brenner S, Johnson M, Bridgham J*, et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* **18**, 630-634.

Bubliy OA, Loeschcke V, Imasheva AG (2001) Genetic variation of morphological traits in Drosophila melanogaster under poor nutrition: isofemale lines and offspring-parent regression. *Heredity* **86**, 363-369.

Carroll J, Shannon RJ, Fearnley IM, Walker JE, Hirst J (2002) Definition of the nuclear encoded protein composition of bovine heart mitochondrial complex I - Identification of two new subunits. *Journal of Biological Chemistry* **277**, 50311-50317.

Chesler EJ, Lu L, Shou S*, et al.* (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function.[see comment]. *Nature Genetics* **37**, 233-242.

Cheung VG, Conlin LK, Weber TM*, et al.* (2003a) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nature Genetics* **33**, 422-425.

Cheung VG, Jen KY, Weber TM*, et al.* (2003b) Genetics of quantitative variation in human gene expression. *Cold Spring Harbor Symposia on Quantitative Biology* **68**, 403-407.

Clayton GA, Morris JA, Robertson A (1957) An experimental check on quantitative genetical theory.I.Short-term responses to selection. *Journal of Genetics* **55**, 131-151.

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 of the United States of America* **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 of the United States of America* **86**, 9365-9369.

Crawford DL, Segal JA, Barnett JL (1999) Evolutionary analysis of TATA-less proximal promoter function. *Molecular Biology & Evolution* **16**, 194-207.

Cronin TM, Szabo BJ, Ager TA, Hazel JE, Owens JP (1981) Quaternary climates and sea levels of the United States Atlantic coastal plain. *Science* **211**, 233-240.

Currie S, Tufts BL (1997) Synthesis of stress protein 70 (Hsp70) in rainbow trout (Oncorhynchus mykiss) red blood cells. *Journal of Experimental Biology* **200**, 607-614.

de Koning DJ, Haley CS (2005) Genetical genomics in humans and model organisms. *Trends in Genetics* **21**, 377-381.

Dempster ER, Lerner M (1950) Heritability of threshold characters. *Genetics* **35**, 103-103.

Devlin B, Fienberg S, Resnick D, Roeder K (1997) *Intelligence, Genes, and Success: Scientists Respond to "the Bell Curve"(Statistics for Social Science and Public Policy)* Springer-Verlag New York, Inc., New York.

Ding Y, Xu L, Jovanovich BD*, et al.* (2007) The methodology used to measure differential gene expression affects the outcome. *Journal of Biomolecular Techniques* **18**, 321-330.

Doss S, Schadt EE, Drake TA, Lusis AJ (2005) Cis-acting expression quantitative trait loci in mice. *Genome Research* **15**, 681-691.

Draghici S, Khatri P, Eklund AC, Szallasi Z (2006) Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet* **22**, 101-109.

Dudoit S, Gendeman RC, Quackenbush J (2003) Open source software for the analysis of microarray data. *Biotechniques*, 45-51.

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.

Emsley A, Dickerson GE, Kashyap TS (1977) Genetic parameters in progeny-test selection for field performance of strain cross layers. *Poultry Science* **56**, 121-146.

Enard W, Khaitovich P, Klose J*, et al.* (2002) Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340-343.

Escartin E, Porte C (1999) Hydroxylated PAHs in bile of deep-sea fish. Relationship with xenobiotic metabolizing enzymes. *Environmental Science & Technology* **33**, 2710-2714.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*, Fourth edn. Pearson Education Limited, London.

Finnon P, Robertson N, Dziwura S*, et al.* (2008) Evidence for significant heritability of apoptotic and cell cycle responses to ionising radiation. *Human Genetics* **123**, 485-493.

Fu J, Keurentjes JJB, Bouwmeester H*, et al.* (2009) System-wide molecular evidence for phenotypic buffering in Arabidopsis. *Nature Genetics* **41**, 166-167.

Gibson G (2008) The environmental contribution to gene expression profiles. *Nature Reviews Genetics* **9**, 575-581.

Gibson G, Riley-Berger R, Harshman L*, et al.* (2004) Extensive sex-specific nonadditivity of gene expression in Drosophila melanogaster. *Genetics* **167**, 1791-1799.

Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends in Genetics* **21**, 616-623.

Gold D, Coombes K, Medhane D*, et al.* (2004) A comparative analysis of data generated using two different target preparation methods for hybridization to high-density oligonucleotide microarrays. *Bmc Genomics* **5**.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.

Gracey AY, Troll JV, Somero GN (2001) Hypoxia-induced gene expression profiling in the euryoxic fish Gillichthys mirabilis. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 1993-1998.

Hochachka PW, Somero GN (1984) *Biochemical Adaptation* Princeton University Press, Princeton, NJ.

Hoffmann AA, Merila J (1999) Heritable variation and evolution under favourable and unfavourable conditions. *Trends in Ecology & Evolution* **14**, 96-101.

Hubner N, Wallace CA, Zimdahl H*, et al.* (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease.[see comment]. *Nature Genetics* **37**, 243-253.

Irizarry RA, Warren D, Spencer F*, et al.* (2005) Multiple-laboratory comparison of microarray platforms. *Nature Methods* **2**, 345-349.

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. *Trends in Genetics* **17**, 388-391.

Jin W, Riley RM, Wolfinger RD*, et al.* (2001) The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster. *Nature Genetics* **29**, 389-395.

Kerr MK, Churchill GA (2001) Statistical design and the analysis of gene expression microarray data. *Genetical Research* **77**, 123-128.

Koldkjaer P, Pottinger TG, Perry SF, Cossins AR (2004) Seasonality of the red blood cell stress response in rainbow trout (Oncorhynchus mykiss). *Journal of Experimental Biology* **207**, 357-367.

Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J (2005) Independence and reproducibility across microarray platforms. *Nature Methods* **2**, 337-343.

Lettieri T (2006) Recent applications of DNA microarray technology to toxicology and ecotoxicology. *Environmental Health Perspectives* **114**, 4-9.

Li J, Burmeister M (2005) Genetical genomics: combining genetics with gene expression analysis. *Human Molecular Genetics* **14**, R163-R169.

Li J, Pankratz M, Johnson JA (2002) Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays. *Toxicological Sciences* **69**, 383-390.

Marshall WS (2003) Rapid regulation of NaCl secretion by estuarine teleost fish: coping strategies for short-duration freshwater exposures. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1618**.

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.

Meyer JN, Nacci DE, Di Giulio RT (2002) Cytochrome P4501A (CYP1A) in killifish (Fundulus heteroclitus): Heritability of altered expression and relationship to survival in contaminated sediments. *Toxicological Sciences* **68**, 69-81.

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.

Moore MJ, Dhingra A, Soltis PS*, et al.* (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *Bmc Plant Biology* **6**.

Morley M, Molony CM, Weber TM*, et al.* (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743-747.

Murray J, Zhang B, Taylor SW*, et al.* (2003) The subunit composition of the human NADH dehydrogenase obtained by rapid one-step immunopurification. *Journal of Biological Chemistry* **278**, 13619-13622.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature Genetics* **32**, 261-266.

Oleksiak MF, Crawford DL (2006) Functional genomics in fishes; Insights into physiological complexity. In: *The Physiology of Fishes* (eds. Evan D, Claiborne J), pp. 523-550. CRC Press, Boca Raton.

Oleksiak MF, Crawford DL, Kolell KJ (2001) Utility of natural populations for microarray analyses: Isolation of genes necessary for functional genomic studies. *Marine Biotechnology* **3**, S203-S211.

Oleksiak MF, Crawford DL, Roach JL (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**, 134-143.

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**.

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.

Petretto E, Mangion J, Dickens NJ*, et al.* (2006) Heritability and tissue specificity of expression quantitative trait loci. *Plos Genetics* **2**, 1625-1633.

Pierce VA, Crawford DL (1997) Phylogenetic analysis of thermal acclimation of the glycolytic enzymes in the genus Fundulus. *Physiological Zoology* **70**, 597-609.

Podrabsky JE, Javillonar C, Hand SC, Crawford DL (2000) Intraspecific variation in aerobic metabolism and glycolytic enzyme expression in heart ventricles. *American Journal of Physiology - Regulatory Integrative & Comparative Physiology* **279**, R2344-2348.

Powers DA, Crawford D, Lauerman T, DiMichele L (1991) Genetic mechanisms for adapting to a changing environment. *Annual Review of Genetics* **25**, 629-659.

Powers DA, Crawford DL, Smith M*, et al.* (1993) A multidisciplinary approach to the selectionist/neutralist controversy using the model teleost, *Fundulus heteroclitus*. *Oxford Surveys in Evolutionary Biology* **9**, 43-107.

Prosser CL (1986) *Adaptational Biology: From Molecules to Organisms* Wiley, New York.

Quackenbush J (2002) Microarray data normalization and transformation. *Nature Genetics* **32**, 496-501.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution* **43**, 258-275.

Raberg L, Stjernman M, Hasselquist D (2003) Immune responsiveness in adult blue tits: heritability and effects of nutritional status during ontogeny. *Oecologia* **136**, 360-364.

Radich JP, Mao M, Stepaniants B*, et al.* (2004) Individual-specific variation of gene expression in peripheral blood leukocytes. *Genomics* **83**, 980-988.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *Plos Biology* **4**, 1707-1719.

Ravn K, Wibrand F, Hansen FJ*, et al.* (2001) An mtDNA mutation, 14453G -> A, in the NADH dehydrogenase subunit 6 associated with severe MELAS syndrome. *European Journal of Human Genetics* **9**, 805-809.

Raymond M, Rousset F (1995) Genepop (Version-1.2) Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Renn SCP, Aubin-Horth N, Hofmann HA (2004) Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *Bmc Genomics* **5**.

Roberts DF, Billewicz WZ, Mcgregor IA (1978) Heritability of stature in a West African population. *Annals of Human Genetics* **42**, 15-24.

Rockman MV, Kruglyak L (2006) Genetics of global gene expression. *Nature Reviews Genetics* **7**, 862-872.

Rutledge JJ, Eisen EJ, Legates JE (1973) Experimental evaluation of genetic correlation. *Genetics* **75**, 709-726.

Saraste M (1999) Oxidative phosphorylation at the fin de siecle. *Science* **283**, 1488-1493.

Schadt EE, Monks SA, Drake TA*, et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297-302.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470.

Schmidt-Neilsen K (1990) *Animal Physiology: Adaptation and Environment*, Fourth edn. Cambridge University Press, New York.

Scott CP, VanWye JD, McDonald D, Crawford DL (2009) Technical analysis of cDNA microarrays. *PLoS One* **4(2),** 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**, R1150-1153.

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.

Snedecor GW, Cochran WG (1991) *Statistical Methods*, Eighth edn. Blackwell Publishing.

Sokal RR, Rohlf FJ (1995) *Biometry: The Principles and Practice of Statistics in Biological Research* W.H. Freeman and Company, New York.

Stamatoyannopoulos JA (2004) The genomics of gene expression. *Genomics* **84**, 449-457.

Strang GS, Smith C (1979) Note on the heritability of litter traits in pigs. *Animal Production* **28**, 403-406.

Tan PK, Downey TJ, Spitznagel EL*, et al.* (2003) Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Research* **31**, 5676-5684.

Tan QH, 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.

Townsend JP, Cavalieri D, Hartl DL (2003) Population genetic variation in genome-wide gene expression. *Molecular Biology & Evolution* **20**, 955-963.

van Haaften RIM, Schroen B, Janssen BJA*, et al.* (2006) Biologically relevant effects of mRNA amplification on gene expression profiles. *Bmc Bioinformatics* **7**.

Vangelder RN, Vonzastrow ME, Yool A*, et al.* (1990) Amplified Rna Synthesized from Limited Quantities of Heterogeneous Cdna. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 1663-1667.

Velculescu VE, Vogelstein B, Kinzler KW (2000) Analysing uncharted transcriptomes with SAGE. *Trends in Genetics* **16**, 423-425.

Vera JC, Wheat CW, Fescemyer HW*, et al.* (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* **17**, 1636-1647.

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.

Weigensberg I, Roff DA (1996) Natural heritabilities: Can they be reliably estimated in the laboratory? *Evolution* **50**, 2149-2157.

Whitehead A, Crawford DL (2005) Variation in tissue-specific gene expression among natural populations. *Genome Biology* **6**.

Whitehead A, Crawford DL (2006a) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **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 of the United States of America* **100**, 1896-1901.

Wicker T, Schlagenhauf E, Graner A*, et al.* (2006) 454 sequencing put to the test using the complex genome of barley. *Bmc Genomics* **7**.

Wierzbicki H, Filistowicz A, Jagusiak W (2004) Breeding value evaluation in Polish fur animals: Statistical description of fur coat and reproduction traits - relationship and inbreeding. *Czech Journal of Animal Science* **49**, 16-27.

Wiley EO (1986) A study of the evolutionary relationships of Fundulus topminnows (Teleostei, Fundulidae). *American Zoologist* **26**, 121-130.

Williams RBH, Chan EKF, Cowley MJ, Little PFR (2007) The influence of genetic variation on gene expression. *Genome Research* **17**, 1707-1716.

Williams TD, Gensberg K, Minchin SD, Chipman JK (2003) A DNA expression array to detect toxic stress response in European flounder (Platichthys flesus). *Aquatic Toxicology* **65**, 141-157.

Windig JJ (1997) The calculation and significance testing of genetic correlations across environments. *Journal of Evolutionary Biology* **10**, 853-874.

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.

Yang YH, Buckley MJ, Dudoit S, Speed TP (2002) Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics* **11**, 108-136.

Yauk CL, Berndt ML, Williams A, Douglas GR (2004) Comprehensive comparison of six microarray technologies. *Nucleic Acids Research* **32**.

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**.

Yvert G, Brem RB, Whittle J*, et al.* (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nature Genetics* **35**, 57-64.