Published in final edited form as:

J Fish Biol. 2010 April ; 76(5): 1067–1093. doi:10.1111/j.1095-8649.2010.02563.x.

Genomic Approaches with Natural Fish Populations

M. F. Oleksiak¹

Rosenstiel School of Marine and Atmospheric Sciences University of Miami 4600 Rickenbacker
Cswy Miami, FL 33149

Abstract

Natural populations *versus* inbred stocks provide a much richer resource for identifying the effects of nucleotide substitutions because natural populations have greater polymorphism. Additionally, natural populations offer an advantage over most common research organisms because they are subject to natural selection, and analyses of these adaptations can be used to identify biologically important changes. Among fishes, these analyses are enhanced by having a wide diversity of species (> 28,000 species, more than any other group of vertebrates) living in a huge range of environments (from below freezing to > 46° C, in fresh water to salinities > 40 ppt.). Moreover, fishes exhibit many different life history and reproductive strategies and have many different phenotypes and social structures. While fishes provide numerous advantages over other vertebrate models, there is still a dearth of available genomic tools for fishes. Fish make up approximately half of all known vertebrate species, yet less than 0.2% of fish species have significant genomic resources. Nonetheless, genomic approaches with fishes have provided some of the first measures of individual variation in gene expression and insights in to environmental and ecological adaptations. Thus, genomic approaches with natural fish populations have the potential to revolutionize fundamental studies of diverse fish species that offer myriad ecological and evolutionary questions.

Keywords

genomics; natural populations; adaptation; microarrays; population genomics; *Fundulus heteroclitus*

Introduction

In general there have been two methods to study adaptation: experimental evolution (Gibbs *et al.*, 1997; Treves *et al.*, 1998; Ferea *et al.*, 1999; Wichman *et al.*, 2000; Riehle *et al.*, 2001; Archer *et al.*, 2007) and studies of natural populations (Koehn *et al.*, 1976; Watt, 1977; Pierce and Crawford, 1997). Experimental evolution studies produce clear demonstrations of adaptation because one can demonstrate a derived change that has arisen independently and affects fitness relative to the original population or a non-selected population. Even though these studies are powerful tools for understanding adaptive mechanisms, they rely on relatively small populations or little standing genetic variation. Natural populations can be much larger and tend to have much more genetic variation. Additionally, natural populations “suffer” from the effects of migrations, stochastic variation in population size, and unknown gene by environmental effects. These traits can enhance natural selection by providing greater genetic variation but also make adaptive changes more difficult to distinguish from stochastic differences. Thus, populations exposed to natural

¹Author to whom correspondence should be addressed. Tel.: 305-421-4341; Fax: 305-421-4600; moleksiak@rsmas.miami.edu..

selection in an appropriate ecological context or natural setting frequently exhibit more realistic responses than organisms exposed in the laboratory. Moreover, similar to laboratory replicates, comparisons among populations or species can provide the means to discern biologically relevant adaptations.

One of the reasons that inbred organisms are so often used in medical studies is that it is easier to study cause and effect when you can account for genetic variation. Yet, most populations of interest (endangered species, ecologically important species, species heavily impacted by anthropogenic pollutions, and humans) have large amounts of genetic variation (on average 33-50% of enzymes are polymorphic (Mitton, 1997)), and individuals have many heterozygotic loci (approximately 15% of loci are in a heterozygotic state (Mitton, 1997)). Among humans, it is estimated that greater than three million nucleotide differences occur between any two humans (Li, 1997), and the analysis of whole genome sequences for individual humans supports this number (Bentley *et al.*, 2008; Wang *et al.*, 2008). However, a compelling question throughout biology is how much inter-individual variation is biologically important? In other words, how much of this variation affects longevity, reproductive fitness, or probability of survival? The huge challenge in studying variation is that most of the DNA variations within and between populations and species are effectively neutral and have no phenotypic or fitness consequences. Moreover, most traits are influenced by multiple genes and environmental factors so that a particular trait will only manifest in the right condition (*e.g.* human height is a heritable trait that will only manifest with proper nutrition and sufficient food). Thus, identification of genetic variation underlying phenotypic traits is one of the most challenging tasks in biology, and finding causative polymorphisms that affect ecologically relevant traits in free-living organisms is even more challenging.

This review focuses on genomic approaches used with natural fish populations to discern biologically important variation. The power of genomic tools is just beginning to be harnessed for adaptive studies of natural fish populations. Two complementary approaches include microarrays and population genomics. While the vast majority of microarray studies have examined physiological or dose responses, far fewer microarray studies have looked at adaptive changes in gene expression in natural fish populations. Population genomic approaches more directly target the underlying DNA variation found among individuals and populations and are just now being used with natural fish populations to identify adaptively important sequence variations. Together, these two genomic approaches (adaptive gene expression and population genomics) have the potential to leverage studies of diverse fish species that offer many different ecological and evolutionary questions.

Why fish?

Fish are the most specious vertebrate (28,600 extant species *versus* 4,629 in mammals and 9,946 in birds (Wilson and Reeder, 1993)). While the mammalian lineage separated from fishes approximately 410 million years ago, divergence times among fish species can be less than 10,000 years, and the diversity of fish species, many of them closely related, provides for powerful phylogenetic comparisons (Felsenstein, 1985; Garland *et al.*, 1992; Garland and Carter, 1994; Crawford *et al.*, 1999, Figure 1). As vertebrates, fish share many development pathways, physiological mechanisms, and organ systems with humans and other mammals. Thus, studies with different fish species should further our understanding of vertebrate embryogenesis and physiological processes in general. This is exemplified by the plethora of gene knock-out and knock-in studies in zebrafish development aimed at understanding the mechanistic basis of different diseases as well as the many studies that target particular physiological processes such as respiration, cardiovascular metabolism,

osmoregulation, and endocrinology. Moreover, as experimental animals, fish are generally much less expensive to maintain in the laboratory than any other vertebrates.

While fish have many advantages for understanding fundamental vertebrate biological processes, they also comprise a highly diverse group. They inhabit a huge range of environmental conditions (Nelson, 1984), and the intimate contact which all fish species have with their aquatic environment (*i.e.* the direct contact of water in their environment with many tissues and internal compartments) has prompted the view that fish represent excellent models for “environmental genomics” (Cossins and Crawford, 2005). Fish also exhibit a variety of phenotypes, reproductive modes, behaviors, and morphologies.

Finally, many fish species have large populations. Thus, evolution by natural selection only requires relatively small selection coefficients (s): $s > 1/N_e$ to dominate evolution by random drift (where N_e is the effective population size (Li, 1997)). Many fish species also (at least in the laboratory) can have very large family sizes. These large family sizes allow association studies and heritability studies to be performed even without inbred strains. For instance, Atlantic salmon, *Salmo salar* (Linnaeus, 1758), infected with the bacterium *Aeromonas salmonicida* suffer high mortality in salmon due to furunculosis. A study using forty full siblings from each of 120 families of Atlantic salmon showed that survival was associated with particular MHC alleles, and the relative fitness difference between individuals carrying different MHC alleles was as high as 0.5 (Langefors *et al.*, 2001; Lohm *et al.*, 2002). This type of association study was only possible with large family sizes. Similarly, multiple large families were used to measure heritability of gene expression in *Fundulus heteroclitus* (Linnaeus, 1766). This study used thirteen families of sibling-parent analyses to show that 6.5% of genes had significant heritability in gene expression with a median h^2 of 0.86, similar to studies using inbred strains (D. Crawford, personal communication). Thus, fish have many attributes that make them good model organisms for both vertebrates in general as well as ecological, environmental, and behavioral questions.

“Good” Natural Populations

Attributes of a “good” species include species with a known population structure and phylogeny so that similarities and differences due to relatedness can be separated from those due to a stress, environment, or treatment. Similarly, a known population structure is needed to gauge the relatedness of individuals within species. Thus, one powerful approach to studying natural variation is to combine measures of genetic distance such as microsatellite analyses with measures of variation. For example, to identify genes evolving by natural selection in *F. heteroclitus*, Whitehead and Crawford used a phylogenetic comparative method to correct for genetic relatedness; they allocated among-population variation in gene expression to genetic distance and then examined the remaining variation relative to temperature (see Felsenstein (1985) and Harvey and Pagel (1991) for the reasoning behind phylogenetic comparative methods). Although much of the variation in gene expression fit a null model of neutral drift, the variation in expression for 22% of the temperature related genes (4% of all genes examined) regressed with habitat temperature and was far greater than could be accounted for by genetic distance alone suggesting evolution by natural selection (Whitehead and Crawford, 2006). Similarly, Larsen *et al.* combined gene expression and microsatellite analyses to explore adaptive differences in gene expression between North Sea and Baltic Sea flounders maintained in a long-term reciprocal transplantation experiment mimicking natural salinities. Several of the differentially regulated genes could be directly linked to fitness traits (Larsen *et al.*, 2007).

As well as the ability to measure relatedness within and among species, “good” natural populations also require an observable ecology with environmental attributes likely to affect

the fitness of individuals. For example, alcohol dehydrogenase polymorphism of *Drosophila melanogaster* has been studied in relation to environmental ethanol (within a winery), ethanol tolerance and alcohol dehydrogenase activity (Gibson and Wilks, 1988), armour in stickleback and *Daphnia* has long been studied in association with predation (Harvell, 1990; Peichel *et al.*, 2001a), *Saccharomyces paradoxus*, the closest known relative of the model eukaryote *Saccharomyces cerevisiae* (bakers' yeast), has been isolated mainly from the bark of oak trees and surrounding soil, while *S. cerevisiae* is commonly obtained from vineyards (Koufopanou *et al.*, 2006), and *Fundulus heteroclitus* have long been studied to understand temperature adaptation among populations (Powers *et al.*, 1991). This is exemplified by the groundbreaking work of Powers and his colleagues that explored temperature adaptations in *Fundulus* populations that inhabit the East coast of North America where they experience a steep thermal cline (Mied and Powers, 1978; Place and Powers, 1979; DiMichele and Powers, 1982b; a; Place and Powers, 1984; Van Beneden and Powers, 1985; Crawford and Powers, 1989; DiMichele *et al.*, 1991; Crawford and Powers, 1992; Bernardi *et al.*, 1993; Quattro *et al.*, 1993; Segal *et al.*, 1996; Schulte *et al.*, 2000).

"Ideal model species" have been further defined as those with a large community of investigators, available genomic sequence and chromosomal maps, forward and reverse genetic tools, and ability measure transcript and protein levels (Feder and Mitchell-Olds, 2003). By definition, this limits ideal model species to species with known genomic sequences. Among fishes, stickleback has the most characteristics of an ideal model species. However, with the advent of new sequencing technologies and decreasing costs, many more natural fish populations have the potential to join stickleback as "ideal model species."

Caveats for Studies of Natural Populations

Studying adaptation in natural populations is not trivial primarily due to two interacting factors: genetic variation and environmental variation. Thus, an organism (its genotype) responds to its environment, and differing genotypes in heterogeneous environments likely will respond differently. The genetic contribution to variation can be directly measured by using controlled breeding designs (if not inbred organisms), by measuring relatedness using genetic markers or by mapping the genetic factors using linkage and/or association studies. However, these approaches often are not possible with natural fish populations that lack crosses, pedigrees, and sufficient genetic markers to measure relatedness and linkage. Thus, a common strategy to differentiate genetic and environmental effects in natural populations is to hold the environment constant and vary the genetics (*e.g.*, common garden experiments where individuals from two or more populations are maintained in a common environment, most often a controlled laboratory environment). The opposite also can be done: reciprocal transplantations or one genetically related population placed in different environments. Endler provides a comprehensive summary of methods for detecting natural selection in the wild that is beyond the scope of this paper (Endler, 1986) and also discusses many practical problems in detecting natural selection. These problems lead to three broad outcomes: the lack of detection of natural selection when it is present (Type II error), apparent detection of natural selection when it is not present (Type I error), and misleading demonstrations of natural selection. Many of the problems derive from a lack of knowledge of the ecology and/or biology of the species or from poorly designed sampling and statistics (Endler, 1986).

Genomic Resources for Fish

Currently, there are five fish nuclear genome projects (*versus* mitochondrial genomes). These are zebrafish (*Danio rerio* (Hamilton, 1822)), Japanese Rice fish or Medaka (*Oryzias latipes* (Temminck & Schlegel, 1846)), two pufferfish (*Fugu rubripes* (Temminck & Schlegel, 1850) and *Tetraodon nigroviridis* (Marion de Procé, 1822)), and three-spine

stickleback (*Gasterosteus aculeatus* (Linnaeus, 1758)). Atlantic salmon, *Salmo salar* (Linnaeus, 1758) and trout, *Oncorhynchus mykiss*, (Walbaum, 1792) genome projects are underway and due to be completed by 2011. While none of these fish genomes has been completely assembled (sequences joined together to reassemble chromosomal order), fish genomic sequences were used to identify many human genes and were used as one of the primary pieces of evidence that the vertebrate genome had 30,000-40,000 genes and not 70,000-140,000 (Roest Croliius *et al.*, 2000). Although nuclear genome projects are lacking for most fish species, a wealth of sequence data exists for many more fish species. This sequence data is largely based on Expressed Sequence Tag (EST) projects. ESTs represent a partial sequence of the much longer RNA expressed in a cell. Because the mRNAs have been processed and edited in the cell, ESTs encode genes that are actively transcribed without intervening intron sequences and so can be more informative about the ultimate function of the gene. However, even though ESTs are transcribed genes and thus we should expect 30,000 to 40,000, there are hundreds of thousands of different ESTs in many EST databases. Alternative splicing is partially responsible for the hundreds of thousands of ESTs in different databases; *e.g.*, while humans have between 20,000 and 30,000 genes, the Gene Index Project human database (TGI) lists 1,083,935 total unique sequences for humans. The other reason for the multitude of ESTs is that the clustering of mRNA from the same gene is often incomplete, yielding more than one unigene per locus (Paschall *et al.*, 2004).

In teleost fishes, three-dozen species in diverse orders have EST collections that contain more than 10,000 sequences (Table I). Even when all fishes with listed ESTs are included (Table II), only half a dozen more species are added. This is a small number compared to the total number of fish species. Not surprisingly, zebrafish have the most ESTs, followed by medaka, then the salmoniformes (Atlantic salmon and rainbow trout) and finally three-spine stickleback. Three of these species also have whole genome sequences. The two pufferfish with whole genome sequences do not have large EST collections because the whole genomes were sequenced first. Among the Acanthopterygii (spiny finned fishes), which comprise 55% of all teleosts and over one-third of all vertebrates, only half of the orders have species with EST collections. At the species level, only 0.13% (19/14797) of the species in Acanthopterygii have significant EST collections. There is a higher percentage of species with ESTs in the less diverse Protacanthopterygii (2.5%, 9/367), mostly salmoniformes, but these small percentages have yet to exploit the full potential of diverse fish species (Figure 1). The conundrum remains: should resources be directed to select “ideal model species” or more widely dispersed among species which are perhaps more relevant to the question at hand? The hope is that cheaper sequencing costs will make this question moot.

While ESTs often are sequenced with the end goal of using them for gene expression analyses, ESTs also are a rich source for discovering microsatellites and SNPs. For instance, Ju *et al.* mined EST sequences from 49,430 unigenes representing a total of 692,654 ESTs from four model fish (medaka, *Fundulus*, zebrafish, and *Xiphophorus*) for their potential use in developing simple sequence repeats (SSRs), or microsatellites. They identified a total of 3,018 EST derived SSRs and simulation analysis suggested that a majority of these EST-SSRs have sufficient flanking sequences for polymerase chain reaction (PCR) primer design (Ju *et al.*, 2005). Another approach to find polymorphic markers is to compare ESTs within any one species that has multiple ESTs representing the same contig. If these multiple sequence were derived from more than one individual, then the differences among sequences can be used to identify putative SNPs. However, caution needs to be used because one cannot always be certain that a particular SNP in an EST is due to true polymorphism *versus* sequence error or clustering of different loci (*i.e.*, sequence paralogues). EST-derived microsatellites have been used for linkage mapping in turbot, salmon, trout (Rexroad *et al.*, 2005; Bouza *et al.*, 2008; Moen *et al.*, 2008), and recently, Kucuktas *et al.* combined both

microsatellites and SNPs derived from ESTs, to construct a genetic linkage map of the channel catfish genome (Kucuktas *et al.*, 2009). Other uses for these EST-derived microsatellites and SNPs include population genomic analyses ((Vasemagi *et al.*, 2005), below) and whole genome association studies.

EST Sequencing

For EST sequencing projects, typically mRNA is reverse transcribed into cDNA which is made into a cDNA library (Oleksiak *et al.*, 2001). Clones from the library then are randomly sequenced. Often, either the mRNA or cDNA library is normalized to reduce the copies of highly expressed genes and increase the efficiency of sequencing novel transcripts. Normalization is necessary because a few (10-20) genes make up at least 20% of the total mRNA mass (*i.e.*, there are many thousands of transcripts for a few genes) and a few hundred genes make up 40-60% of the total mRNA mass (Bonaldo *et al.*, 1996). Unless the copies of transcripts for each gene are equalized or normalized, the same gene will be sequenced many times, lessening the discovery of new gene sequences. For example, in the *Fundulus* EST collection, 33% of the cDNAs were redundant from the non-normalized cDNA library, but only 11% were redundant from the normalized library (Oleksiak *et al.*, 2001). Thus, normalization decreased the probability of sampling the same cDNA by 3-fold, and was an effective method for increasing productivity. This approach, of making a cDNA library, normalizing it and sequencing random clones, provides both sequence and the DNA in a plasmid. To be useful, the clones must be organized and stored for later retrieval. This requires considerable space and methods to precisely locate clones. However, these clones can be used to create DNA for printing microarrays, can be re-sequenced to verify gene identification or get greater sequence depth or for any future manipulations.

A faster sequencing method is 454 sequencing. Using 454 sequencing obviates the cloning step, and the cDNAs are directly sequenced using a highly parallel sequencing platform. Using 454 sequencing, up to 400,000 transcripts can be sequenced in a few hours. What took months just five years ago now can be accomplished in days. Among other uses, these sequences can and have been used to design microarrays for studies of natural populations (Vera *et al.*, 2008). This opens the door for transcriptomic studies in virtually any species. Notice that with 454 technology one only gains sequences, and thus, oligonucleotide arrays have to be synthesized in order to perform transcriptomic studies.

Illumina and SOLiD are two other high throughput, sequencing technologies. These technologies are widely used in species with sequenced genomes because the read lengths are relatively short. While the average read length for 454 sequencing is ~400 bp, Illumina sequences are typically 35-100 bp and SOLiD sequences are 35 and 50 bp. The longer 454 sequences can be more readily assembled and annotated; however, a genome scaffold is needed to most easily assemble and identify the shorter Illumina and SOLiD reads. (However, the large number of reads makes Illumina appropriate for *de novo* transcriptome studies). Thus, while 454 sequencing is more expensive per megabase of sequence and suffers from greater sequencing error (mostly at the end of polynucleotide stretches) than either Illumina or SOLiD sequencing, it currently is the most viable option for high throughput, *de novo* sequencing.

Measuring Gene Expression with Microarrays

To enhance our understanding of genetic differences among populations or species, we need to be able to survey much of the genome. One of the more powerful ways to achieve this is to quantify most or all of gene expression. Currently, microarrays are the most cost effective means to measure all of gene expression in both model and non-model species. While

feasible, quantitative transcriptome sequencing using high throughput sequencing technologies is too expensive for more than a few individuals.

Microarrays are thousands of 150 to 250 micron spots of DNA bound to microscope slides in a precise and known pattern (Ramsay, 1998; Schena *et al.*, 1998). Each DNA spot quantitatively hybridizes to a specific mRNA so that expression of thousands of individual genes can be measured simultaneously. Importantly, microarray techniques are sensitive: typically 2-fold differences in mRNA concentration are determined, and each gene/DNA spot has a sensitivity of 15 attomoles (Schena *et al.*, 1995) or approximately one out of 300,000 transcripts can be measured (Hill *et al.*, 2000). Application of statistical analyses with appropriate replication has improved these analyses such that less than 1.5 fold differences are readily discerned (Jin *et al.*, 2001; Wolfinger *et al.*, 2001; Oleksiak *et al.*, 2002).

Microarray studies of mRNA expression ignore other processes that affect protein activity. Variation in protein activities due to allosteric activators, phosphorylation or other post-translational modifications is not measured. Although enzyme activities or protein concentrations in a cell are not regulated solely by mRNA levels, often changes in cell type, development or physiological states are affected by changes in mRNA. For example, most cancers have altered gene expression, and this has been used with microarray analyses as a way to classify cancers (Zhang *et al.*, 1997; Golub *et al.*, 1999; Elek *et al.*, 2000; Forozan *et al.*, 2000; Ono *et al.*, 2000). As well, stress responses resulting from infection by pathogenic organisms, exposure to xenobiotics, or altered environments (*e.g.* hypoxia) often result in altered gene expression (Lam *et al.*, 2006; Ju *et al.*, 2007; Oleksiak, 2008; Peatman *et al.*, 2008).

The huge advantage of microarrays is that they can be used to measure thousands of gene transcripts simultaneously. The basic idea is that knowledge of where, when, and under what conditions a gene or set of genes is induced or repressed will give insight into biological mechanisms. Gene expression studies have huge potential for non-model species by using 454 sequencing to quickly sequence thousands of transcripts: these sequences are used to direct the synthesis of oligonucleotide microarrays (Vera *et al.*, 2008).

Studies with Fish Microarrays

This Journal recently published a special volume devoted to fish microarrays, Journal of Fish Biology: Fish Microarrays (2008), edited by K. M. Miller and N. Maclean, and described arrays from 12 fish species: largemouth bass, *Micropterus salmoides* (Lacepède, 1802), goby, *Gillichthys mirabilis* (Cooper, 1864), fathead minnow, *Pimephales promelas* (Rafinesque, 1820), channel catfish, *Ictalurus punctatus* (Rafinesque, 1818), and the closely related blue catfish, *Ictalurus furcatus* (Valenciennes, 1840), European flounder, *Platichthys flesus* (Linnaeus, 1758), rainbow trout, *Oncorhynchus mykiss* (Walbaum, 1792), Atlantic salmon, *Salmo salar* (Linnaeus, 1758), common carp, *Cyprinus carpio* (Linnaeus, 1758), Atlantic halibut, *Hippoglossus hippoglossus* (Linnaeus, 1758), three-spined stickleback, *Gasterosteus aculeatus* (Linnaeus, 1758), and *Kryptolebias marmoratus* (Poey, 1880, *Rivulus marmoratus* until 2004). The array for the largemouth bass used 454 sequencing to quickly sequence transcripts for the array (Garcia-Reyero *et al.*, 2008). Interestingly, von Schalburg *et al.* investigated oligonucleotide designs based on distance from the polyA tail, the effects of mismatches and cross-species hybridization specificity. Notable for comparisons of different natural populations, they found that even 3 mismatches (95% identity) in a 70-mer oligonucleotide probe decreased the signal by 40% (von Schalburg *et al.*, 2008). Thus, apparent differences among populations may simply be due to variable sequences among individuals. While this should not be a problem for cDNA arrays given the

longer probe sizes, this can be mitigated for oligo arrays by averaging signal for each gene across multiple probes. Yet, this adds both to cost and complexity.

Microarrays also have been used with other fish species. In addition to the species mentioned above, there are multiple microarray studies in the model species, zebrafish and medaka. As well, gene expression studies have examined hypoxia in *Xiphophorus maculatus* (Günther, 1866) using a heterologous medaka array (Boswell *et al.*, 2009), osmoregulation in the European eel *Anguilla Anguilla* (Linnaeus, 1758) (Kalujnaia *et al.*, 2007), dominance in cichlids (Aubin-Horth *et al.*, 2007), limnetic niche use among Coregonine fishes (*Coregonus* spp. *Salmonidae*) (St-Cyr *et al.*, 2008), and both adaptive and physiological responses in the killifish, *Fundulus* (Oleksiak *et al.*, 2002; Oleksiak *et al.*, 2005; Whitehead and Crawford, 2006; Fisher and Oleksiak, 2007; Roling *et al.*, 2007; Oleksiak, 2008; Everett and Crawford, in press).

In fact, the first microarray experiments to measure inter-individual variation in gene expression in any species were done in *F. heteroclitus* (Oleksiak *et al.*, 2002). A subsequent experiment measured an astonishing amount of individual variation in metabolic gene expression: 95% of the metabolic genes varied significantly in expression among individuals (Oleksiak *et al.*, 2005). It is hard to believe that this frequency of inter-individual difference in gene expression is biologically interesting. Yet, the differences in gene expression are associated with changes in cardiac metabolism: the patterns of gene expression explain up to 81% of the variation in cardiac metabolism. Surprisingly, the patterns of gene expression that explain metabolism differ among individuals. Thus, for glucose metabolism, in 1/3 of the individuals, variation in glycolytic gene expression is most important (explains >80% of the variation in metabolism) but in other individuals it is the variation in oxidative-phosphorylation genes that explain metabolism (Oleksiak *et al.*, 2005; Crawford and Oleksiak, 2007).

The surprisingly large amount of individual variation that explains much of a physiological trait (substrate specific metabolism) has four critically important implications for human and other studies. First, examining one or a few inbred individuals could be misleading because a pathway or mechanism that is important for one individual is not necessarily important for another one. Second, many genes or different pathways can affect substrate specific metabolism. While enzymes involved in glycolysis might explain metabolic variation for a subset of individuals, enzymes involved in oxidative phosphorylation might be important for metabolic variation in a different subset of individuals. Thus, we cannot measure one or a few enzymes and expect it to always explain metabolic variation among many different individuals. Third, it is unlikely that one gene or one set of genes is responsible for the phenotypic variation among all individuals. The same phenotype might have different underlying causes. Thus, we should not expect a magic bullet that will cure everyone of a specific disease. Instead, different cures will be necessary for different individuals. Finally, the lack of correlation between gene expression and phenotype is not necessarily due to the lack of importance of gene expression. Instead, if the importance of gene expression is context dependent, then a significant relationship will only be discernable in a specific context.

Most gene expression studies measure treatment effects rather than individual variation. Additionally, the vast majority of microarray studies have examined physiological or dose responses such as responses to temperature, salinity, pollution, specific chemical treatments, stress, disease, and age to name a few. Far fewer microarray studies have looked at adaptive changes in gene expression in natural fish populations (or in any natural population, Table III).

Evolutionary Genomics

For adaptive studies, one can use an evolutionary or phylogenetic perspective to reveal changes that are produced by natural selection and therefore indicate whether the variation affects longevity, reproductive fitness, or probability of survival, *i.e.*, whether the variation is biologically important. Specifically, if the variation is neutral, the pattern of variation will be random and thus will be correlated with evolutionary distance among taxa rather than significantly correlated with important environmental parameters (Pierce and Crawford, 1997; Crawford *et al.*, 1999; Oleksiak *et al.*, 2002). In contrast, if the pattern of variation is nonrandom and most likely due to evolution by natural selection, it will correlate significantly with important environmental parameters and be independent of evolutionary distance between taxa. Thus, studying signatures of selection in natural populations has been used to discern biologically important variation in studies from flies to fish to men (Endler, 1986).

Adaptive Microarray Studies in Natural Fish Populations

Fundulus provide a powerful system to explore evolved changes in gene expression because many populations show adaptations (Burnett *et al.*, 2007), N_e in local population sizes is generally $> 10^5$, allowing evolution by natural selection with relatively small selection coefficients to dominate evolution by random drift (Li, 1997), and migration among populations is sufficient to minimize neutral divergence but does not appear to overcome local adaptation (Brown and W., 1991; Powers *et al.*, 1993; Fisher and Oleksiak, 2007). *Fundulus* also has a well described phylogeny, and this was used to explore adaptive changes in gene expression among taxa. Oleksiak *et al.* (2002) compared heart gene expression in a northern population of *F. heteroclitus* to that in a southern population of *F. heteroclitus* and the sister taxa, *F. grandis* (Baird & Girard, 1853). The expectation is that the northern and southern populations should share more similarities with each other than either has to the sister taxa, because they are more closely related. However, a certain subset of genes had expression patterns that were significantly different in the northern population from both the southern population and *F. grandis*, suggesting adaptive differences in gene expression, perhaps due to temperature (Oleksiak *et al.*, 2002).

A similar approach was taken with *F. heteroclitus* populations that inhabit highly polluted Superfund sites (Superfund sites contain high levels of a variety of lipophilic, persistent and toxic contaminants worthy of being remediated using Federal funds) and have adapted to the pollutants in their environment. Instead of using phylogeny, this experiment used triads of comparisons: each polluted site was compared to two flanking reference sites (see Fisher and Oleksiak, 2007; Oleksiak, 2008). This is experimentally robust because the genetic distance between the two clean reference populations is greater than the genetic distance between the polluted population and either reference population. This triad comparison also controls for clinal effects. Because *F. heteroclitus* inhabit a temperature cline along the Atlantic seaboard, if only two populations were compared, differences between the two populations could be due to temperature differences rather than pollution or genetic drift. By comparing the polluted population to both a northern and a southern population, effects of pollution are no longer confounded with either temperature or genetic drift. Thus, divergence in a polluted population compared to both paired reference populations suggests that pollution might be causative. That is, if the differences are shared with both reference populations (*i.e.*, polluted *versus both* reference 1 and reference 2), those shared differences are more likely to be due to pollution. This experimental design was used to quantify gene expression in both brain and liver from *F. heteroclitus* collected from polluted sites and found a number of significantly altered gene expression patterns that likely evolved due to chronic exposure to pollution (Fisher and Oleksiak, 2007; Oleksiak, 2008).

Even relatively weak signals of population structure have been suggested to represent evolutionary units able to adapt to the local environment causing fine-scale spatial and temporal divergence (Conover *et al.*, 2006), and adaptive differences in gene expression were found among European flounder with genetically weakly differentiated populations (F_{ST} ranging from 0.005 to 0.006). European flounder populations distributed across the salinity gradient stretching from the brackish Baltic Sea to the high saline North Sea showed that 5% of genes expressed in the liver were significantly differently expressed in Baltic Sea and North Sea flounders under common garden conditions (Larsen *et al.*, 2007). This study was followed by tissue specific (gill and kidney) gene expression of four candidate genes. Hsp70 was up-regulated six-fold in kidney of North Sea flounder following long term acclimation to Baltic Sea salinities, and significant differences in expression of 5-aminolevulinic acid synthase (ALAS) in response to different salinities in gill and liver tissue were found between North Sea and Baltic Sea flounders (Larsen *et al.*, 2008). The identification of adaptive differences even among high gene flow marine organisms has important implications for management and protection of marine biodiversity.

Microarrays also have been used to explore life history traits in Atlantic salmon, cichlids and whitefish. Male Atlantic salmon either attain sexual maturity precociously (1-3 years) and become sneaker males or migrate out to sea and return 3-7 years later as large, anadromous animals (Fleming, 1998). A sneaker male invades the larger male's territory during spawning and ejects sperm toward the female's newly extruded eggs while the larger male is distracted by his own spawning. Although this tactic fertilizes eggs poorly in comparison to the large males, sneaker males have a greater probability of surviving to reproduce such that the lifetime fitness of smaller males is similar to that of the large, anadromous males (Gross, 1985). This developmental plasticity of sneaker males was explored by comparing gene expression profiles in mature sneaker males and age matched immature males and females. 15% of genes were significantly differently expressed in the brain of mature sneaker males compared with immature males. Not surprisingly, reproduction associated genes were almost exclusively sneaker biased. The importance of the gene expression differences was further evidenced by a cluster analysis that separated the different phenotypes (Aubin-Horth *et al.*, 2005).

In cichlids, microarrays were used with 14 established social groups of the Tanganyikan cichlid *Neolamprologus pulcher* (Trewavas & Poll, 1952) to profile brain gene expression of dominant females. Gene expression in dominant breeder females was most similar to that of the males (independent of social rank), and dominant breeder females also were masculinized at the hormonal level (Aubin-Horth *et al.*, 2007). Thus, while cichlids are an emerging evolutionary genomic model system for fundamental questions on the origin of phenotypic diversity, cooperatively breeding species such as *N. pulcher* provide a powerful model system to further our understanding of the evolution of social behaviour.

Dwarf and normal whitefish populations have morphological, behavioral, ecological and life-history traits that most likely represent adaptations towards exploiting distinct trophic resources (Fenderson 1964; Bernatchez *et al.* 1999; Lu & Bernatchez 1999; Rogers, Gagnon & Bernatchez 2002). These reproductively isolated, whitefish phenotypes occurred in many lakes following ice cap retreat after the Wisconsin glaciations 15,000 years ago. Gene expression was explored in dwarf and normal whitefish from two natural lakes, as well as from populations reared in controlled environments. Dwarf whitefish consistently showed significant over-expression of genes potentially associated with survival through enhanced activity (energy metabolism, iron homeostasis, lipid metabolism, detoxification) and down-regulation of genes associated with growth (protein synthesis, cell cycle, cell growth) in dwarf relative to normal whitefish (St-Cyr *et al.*, 2008). The authors suggest that enhanced survival via a more active swimming, necessary for increased foraging and predator

avoidance, entails energetic costs that result in slower growth rate and reduced fecundity in dwarf relative to normal whitefish. These three studies illustrate the power of microarrays for exploring questions of life history evolution.

Finally, a recent genomic, gene expression study was done *in silico*. Gene expression of the Antarctic notothenioid fish, *Dissostichus mawsoni* (Norman, 1937) was based on transcript abundance from an EST sequencing project of non-normalized cDNA libraries and compared to same tissue transcript abundances of five temperate/tropical fish species in the database. 177 notothenioid protein families were highly expressed compared to the temperate species. The genomic basis for these up-regulated genes was explored using comparative genomic hybridization of DNA from four pairs of Antarctic and basal non-Antarctic notothenioids. Importantly, significant Antarctic specific gene duplications occurred for 118 genes, many corresponding to the up-regulated genes identified *in silico* (Chen *et al.*, 2008).

Markers for Population Genetics/Genomics

Most genomic and genetic studies in natural populations use one or a few molecular markers to assess population parameters (Palumbi, 1994; McMillan and Palumbi, 1995; Williams *et al.*, 2001; Barber *et al.*, 2002; Crandall *et al.*, 2008). Allozymes have long been used due to the ease of use across species (Nevo, 1990). However, the limited number of loci and low variability of allozymes limits the statistical power of the resulting data. Mitochondrial DNA (mtDNA) was the first widely used DNA marker and has provided many insights into the demography of natural populations (Avice *et al.*, 1986; Saunders *et al.*, 1986; Bowen *et al.*, 2007), but because mtDNA is a single locus, its ability to resolve population structure is relatively limited (Avice, 1994). Most recent genetic studies of natural populations have used microsatellites. The high variation in many microsatellites provides high statistical power for population genetics, but they suffer from two drawbacks. First, they require species-specific marker development, and second, they suffer from a high potential for null alleles and are prone to genotyping errors due to their size-based nature (homoplasmy). These shortcomings and the fact that analysis of microsatellite loci requires experience and expertise make data comparisons among laboratories and with published data problematic (Hoffman and Amos, 2005).

Amplified fragment length polymorphisms (AFLPs) have been widely used since first described (Vos *et al.*, 1995) due to their ease of use in species with no prior sequence information: many AFLP markers can be easily amplified and scored. However, AFLP analyses require high quality DNA and provide dominant markers so that heterozygotes cannot be directly measured. Additionally, AFLPs are anonymous, and thus loci can be scored incorrectly due to their size-based nature (the same sized band can represent multiple loci).

Single nucleotide polymorphisms (SNPs) have recently attracted the interest of geneticists due to their potential for automation and high throughput as well as lower genotyping errors than microsatellites (Hauser and Seeb, 2008). In addition, SNP analyses can use DNA of varying quality. Further, as they provide sequence data, the data is directly transferable among different laboratories. Yet, like microsatellites, SNPs require species-specific development for non-model species. Thus, although SNPs are the most abundant marker in most genomes, discovery of useful SNPs can be challenging in organisms with no genomic information, such as many fish species. One approach to identify SNPs is to mine large, expressed sequence tag (EST), data collections for putative SNPs.

Population Genomic Approaches

While microarray approaches can identify significant changes in gene expression that are potentially adaptively important, a thorough understanding of the genomic basis underlying phenotypic adaptive divergence is still emergent (MacCallum and Hill, 2006). Population genomics is one approach to better understand the genomic basis of adaptive divergence. Population genomics is more than population genetic analyses on a large number of loci (Stinchcombe and Hoekstra, 2008) because this large number of loci makes statistical tests for non-random patterns possible (Nielsen, 2001; Schlotterer, 2002; Luikart *et al.*, 2003; Storz, 2005). The basic tenet is that local adaptation and directional selection should reduce genetic variability within populations and increase variation among populations at the locus under selection. One can use a variety of statistical tests to ask whether an individual locus acts as an outlier compared to loci in the rest of the genome (Storz, 2005). The application of these approaches is useful for understanding adaptation in genetically uncharacterized natural populations (Pogson *et al.*, 1995; Beaumont and Nichols, 1996; Yan *et al.*, 1999; Allendorf and Seeb, 2000; Bradshaw *et al.*, 2000; Frary *et al.*, 2000; Parsons and Shaw, 2001; Peichel *et al.*, 2001b; Mock *et al.*, 2002; Kohn *et al.*, 2003; Storz and Nachman, 2003; Storz and Dubach, 2004) and should yield insights into long-standing evolutionary questions.

Population genomics first requires genotyping a large number of loci. These loci might be microsatellites, AFLPs, SNPs or sequences. Without a genome, AFLPs are the easiest and cheapest method to genotype large numbers of loci. However, with new sequencing technologies, other markers are likely to be as available. The advantages of population genomics approaches are the ease of generating large numbers of genetic markers, the ability to scan the genome without measuring phenotypes, and the ability to sample individuals without knowledge of their breeding history (Stinchcombe and Hoekstra, 2008). However, a big disadvantage of AFLPs as well as many SNPs and microsatellites used in population genomics is that they are anonymous or in nonfunctional sequence, so linking a candidate locus to a causative gene or mutation is improbable without further analyses or experiments.

AFLP analyses have long been used to explore genetic diversity, but it is only recently that they have been used to identify loci indicative of adaptation in natural fish populations. To date, only a handful of studies using anonymous genome-wide markers have analyzed multiple populations to test for conserved outlier loci (Wilding *et al.*, 2001; Campbell and Bernatchez, 2004; Vasemagi *et al.*, 2005; Bonin *et al.*, 2006; Williams and Oleksiak, 2008). Three of these were in natural fish populations: Atlantic salmon from fresh, brackish, and saltwater habitats in the Barents and White Seas, dwarf and normal ecotypes of whitefish, and pollution adapted populations of *F. heteroclitus*. Although shared differences among populations are suggestive of non-random patterns, they still could be demographic (*e.g.*, due to a bottleneck). Population genomics allows one to test whether the loci that are significantly different in population A *versus* population B and also are significantly different from all other loci in population A. That is, with hundreds of loci, one makes the assumption that a majority of differences among loci are either non-functional (a polymorphism with no phenotypic effect) or neutral (variation whose phenotypic effect has no or little fitness consequences, *i.e.*, selective advantage is $< 1/2N_e$). Thus one looks for outliers among these loci: loci that have significantly different divergence *versus* a random permutation or modeling of the data.

Genomic and EST-derived microsatellites were screened for patterns of divergent selection among eight Atlantic salmon populations from fresh, brackish, and saltwater habitats in the Barents and White Seas. This study found a high percentage of differentiated loci (12%, 9 of

78 EST-linked loci) suggesting that EST associated markers could improve the efficiency of whole genome scans (Vasemagi *et al.*, 2005). AFLPs were used to compare dwarf and normal ecotypes in whitefish, *Coregonus clupeaformis* (Mitchill, 1818), and found 1.4% differentiated loci (Campbell and Bernatchez, 2004). AFLPs also were used to compare pollution adapted populations of *F. heteroclitus* in a similar experimental design to that used for microarray studies discussed above. After correcting for multiple comparisons, 1% to 6% of loci were implicated as being under selection or linked to areas of the genome under selection (Williams and Oleksiak, 2008).

Note however, that even when both phenotypes and selective environments are very similar among populations, this does not mean that the same loci will be fixed in response to similar environmental conditions (Hoekstra and Nachman, 2003; Hoekstra, 2006). On an evolutionary scale, this is similar to the physiological results where different patterns of gene expression explain substrate specific metabolism (Oleksiak *et al.*, 2005). This suggests a biological complexity that we are only beginning to understand.

Conclusions and Future Directions

In 1966, Lewontin and Hubby first quantified genetic variation in different fly populations and were surprised to find 30% of loci had two or more alleles (Lewontin and Hubby, 1966). This level of polymorphism was unexpected. With the advent of whole genome sequencing projects, we can more definitively quantify genetic variation and do it in a diversity of species to reveal evolutionary adaptive patterns. Even among organisms without sequenced genomes, genomic approaches including microarray studies and SNP identification can reveal substantial inter-individual variation. Again, a critical question is: how much of this variation is biologically important? The first vertebrate studies using microarrays to measure individual variation in gene expression found 20-95% of genes had significant variation among individuals (Oleksiak *et al.*, 2002; Oleksiak *et al.*, 2005). A recent study in flies corroborated this amount of individual variation in gene expression and showed further, that 68% of it was heritable (Ayroles *et al.*, 2009). With natural populations, natural selection is one way to assess biological importance, and discerning evolution by natural selection is more readily accomplished using the rich species diversity among teleost fishes and the depth of environmental knowledge about these species.

Having a full-genome sequence provides many resources to better understand the genetics and molecular mechanisms affecting change in gene frequency or gene expression. Yet, with the advances of newer high throughput sequencing approaches, “genomic” approaches can be applied to most species to yield genome wide information. As discussed, high throughput sequencing technologies have been used to sequence transcripts for gene expression studies but also can be used for amplicon resequencing for cost-effective SNP discovery (Bundock *et al.*, 2009), genomic DNA SNP discovery (Wiedmann *et al.*, 2008), microsatellite discovery by deep sequencing of microsatellite enriched genomic libraries (Santana *et al.*, 2009), mitochondrial sequencing (Cui *et al.*, 2009), and targeted resequencing of particular genes or gene regions (Hodges *et al.*, 2007; Okou *et al.*, 2007). Resulting tools coupled with genomic analyses can give insights into how individuals and populations evolve. Thus, while most fishes do not have full genome sequences and will not until NIH's \$1,000 genome goal becomes reality, the potential for researchers to exploit powerful genomic tools in their studies of fish species is ever expanding.

Acknowledgments

Thanks to Douglas L. Crawford at the University of Miami for helpful comments on this manuscript. Funding for this work was received from NIH 5 RO1 ES011588, 2P42 ES010356, and 2 P42 ES007381.

References

- Allendorf FW, Seeb LW. Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. *Evolution* 2000;54:640–651. [PubMed: 10937239]
- Archer MA, Bradley TJ, Mueller LD, Rose MR. Using experimental evolution to study the physiological mechanisms of desiccation resistance in *Drosophila melanogaster*. *Physiol Biochem Zool* 2007;80:386–398. Epub 2007 May 2008. [PubMed: 17508334]
- Aubin-Horth N, Desjardins JK, Martei YM, Balshine S, Hofmann HA. Masculinized dominant females in a cooperatively breeding species. *Mol Ecol* 2007;16:1349–1358. [PubMed: 17391260]
- Aubin-Horth N, Landry CR, Letcher BH, Hofmann HA. Alternative life histories shape brain gene expression profiles in males of the same population. *Proc Biol Sci* 2005;272:1655–1662. [PubMed: 16087419]
- Avise, J. *Molecular Markers, Natural History, and Evolution*. Chapman and Hall; New York: 1994.
- Avise JC, Helfman GS, Saunders NC, Hales LS. Mitochondrial DNA differentiation in North Atlantic eels: Population genetic consequences of an unusual life history pattern. *Proc Natl Acad Sci U S A* 1986;83:4350–4354. [PubMed: 16593712]
- Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF. Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet* 2009;41:299–307. [PubMed: 19234471]
- Barber PH, Palumbi SR, Erdmann MV, Moosa MK. Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Mol Ecol* 2002;11:659–674. [PubMed: 11972755]
- Beaumont M, Nichols R. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London* 1996;263:1619–1626.
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara ECM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczky C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovskiy Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Mullikin JC, Hurler ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008;456:53–59. [PubMed: 18987734]
- Bernardi G, Sordino P, Powers DA. Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. *Proc Natl Acad Sci U S A* 1993;90:9271–9274. [PubMed: 8105474]

- Bonaldo MF, Lennon G, Soares MB. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res* 1996;6:791–806. [PubMed: 8889548]
- Bonin A, Taberlet P, Miaud C, Pompanon F. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol* 2006;23:773–783. [PubMed: 16396915]
- Boswell MG, Wells MC, Kirk LM, Ju Z, Zhang Z, Booth RE, Walter RB. Comparison of gene expression responses to hypoxia in viviparous (*Xiphophorus*) and oviparous (*Oryzias*) fishes using a medaka microarray. *Comp Biochem Physiol C Toxicol Pharmacol* 2009;149:258–265. Epub 2008 Nov 2021. [PubMed: 19049829]
- Bouza C, Hermida M, Millan A, Vilas R, Vera M, Fernandez C, Calaza M, Pardo BG, Martinez P. Characterization of EST-derived microsatellites for gene mapping and evolutionary genomics in turbot. *Anim Genet* 2008;39:666–670. [PubMed: 18786152]
- Bowen BW, Grant WS, Hillis-Starr Z, Shaver DJ, Bjorndal KA, Bolten AB, Bass AL. Mixed-stock analysis reveals the migrations of juvenile hawksbill turtles (*Eretmochelys imbricata*) in the Caribbean Sea. *Mol Ecol* 2007;16:49–60. [PubMed: 17181720]
- Bradshaw H, Ceulemans R, Davis J, Stettler R. Emerging model systems: poplar (*Populus*) as a model forest tree. *Journal of Plant Growth and Regulation* 2000;19:306–313.
- Brown BL, W. CR. Gene Flow and Mitochondrial Dna Variation in the Killifish *Fundulus-Heteroclitus*. *Evolution* 1991;45:1147–1161.
- Bundock PC, Elliott FG, Ablett G, Benson AD, Casu RE, Aitken KS, Henry RJ. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant Biotechnol J* 2009;7:347–354. [PubMed: 19386042]
- Burnett KG, Bain LJ, Baldwin WS, Callard GV, Cohen S, Di Giulio RT, Evans DH, Gomez-Chiarri M, Hahn ME, Hoover CA, Karchner SI, Katoh F, Maclatchy DL, Marshall WS, Meyer JN, Nacci DE, Oleksiak MF, Rees BB, Singer TD, Stegeman JJ, Towle DW, Van Veld PA, Vogelbein WK, Whitehead A, Winn RN, Crawford DL. *Fundulus* as the Premier Teleost Model in Environmental Biology: Opportunities for New Insights Using Genomics. *Comp Biochem Physiol Part D Genomics Proteomics* 2007;2:257–286. [PubMed: 18071578]
- Campbell D, Bernatchez L. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol Biol Evol* 2004;21:945–956. [PubMed: 15014172]
- Chen Z, Cheng CH, Zhang J, Cao L, Chen L, Zhou L, Jin Y, Ye H, Deng C, Dai Z, Xu Q, Hu P, Sun S, Shen Y. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc Natl Acad Sci U S A* 2008;105:12944–12949. [PubMed: 18753634]
- Conover DO, Clarke LM, Munch SB, Wagner GN. Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology* 2006;69:21–47.
- Cossins AR, Crawford DL. Fish as models for environmental genomics. *Nat Rev Genet* 2005;6:324–333. [PubMed: 15803200]
- Crandall ED, Frey MA, Grosberg RK, Barber PH. Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Mol Ecol* 2008;17:611–626. [PubMed: 18179436]
- Crawford DL, Oleksiak MF. The biological importance of measuring individual variation. *J Exp Biol* 2007;210:1613–1621. [PubMed: 17449827]
- Crawford DL, Pierce VA, Segal JA. Evolutionary physiology of closely related taxa: analyses of enzyme expression. *American Zoologist* 1999;32:389–400.
- Crawford DL, Powers DA. Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish *Fundulus heteroclitus*. *Proc Natl Acad Sci U S A* 1989;86:9365–9369. [PubMed: 2594773]
- Crawford DL, Powers DA. Evolutionary adaptation to different thermal environments via transcriptional regulation. *Mol Biol Evol* 1992;9:806–813. [PubMed: 1528107]
- Cui Z, Liu Y, Li CP, You F, Chu KH. The complete mitochondrial genome of the large yellow croaker, *Larimichthys crocea* (Perciformes, Sciaenidae): unusual features of its control region and the phylogenetic position of the Sciaenidae. *Gene* 2009;432:33–43. Epub 2008 Dec 2003. [PubMed: 19100818]

- DiMichele L, Paynter KT, Powers DA. Evidence of lactate dehydrogenase-B allozyme effects in the teleost, *Fundulus heteroclitus*. *Science* 1991;253:898–900. [PubMed: 1876847]
- DiMichele L, Powers DA. LDH-B genotype-specific hatching times of *Fundulus heteroclitus* embryos. *Nature* 1982a;296:563–564. [PubMed: 7070498]
- DiMichele L, Powers DA. Physiological basis for swimming endurance differences between LDH-B genotypes of *Fundulus heteroclitus*. *Science* 1982b;216:1014–1016. [PubMed: 7079747]
- Elek J, Park KH, Narayanan R. Microarray-based expression profiling in prostate tumors. *In Vivo* 2000;14:173–182. [PubMed: 10757075]
- Endler JA. *Natural Selection in the Wild*. Princeton University Press; Princeton: 1986.
- Everett MV, Crawford DL. Adaptation versus allometry: Population and body mass effects on hypoxic metabolism in *Fundulus grandis*. *Physiological and Biochemical Zoology*. in press.
- Feder ME, Mitchell-Olds T. Evolutionary and ecological functional genomics. *Nat Rev Genet* 2003;4:651–657. [PubMed: 12897776]
- Felsenstein J. Phylogenies and the comparative method. *American Naturalist* 1985;125:1–15.
- Ferea TL, Botstein D, Brown PO, Rosenzweig RF. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Natl Acad Sci U S A* 1999;96:9721–9726. [PubMed: 10449761]
- Fisher MA, Oleksiak MF. Convergence and divergence in gene expression among natural populations exposed to pollution. *BMC Genomics* 2007;8:108. [PubMed: 17459166]
- Fleming IA. Pattern and variability in the breeding system of Atlantic salmon, with comparisons to other salmonids. *Can. J. Fish. Aquat. Sci* 1998;55:59–76.
- Forozan F, Mahlamaki EH, Monni O, Chen Y, Veldman R, Jiang Y, Gooden GC, Ethier SP, Kallioniemi A, Kallioniemi OP. Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. *Cancer Research* 2000;60:4519–4525. [PubMed: 10969801]
- Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 2000;289:85–88. [PubMed: 10884229]
- Garcia-Reyero N, Griffitt RJ, Liu L, Kroll KJ, Farmerie WG, Barber DS, Denslow ND. Construction of a robust microarray from a non-model species largemouth bass, *Micropterus salmoides* (Lacépède), using pyrosequencing technology. *Journal of Fish Biology* 2008;72:2354–2376. [PubMed: 19936325]
- Garland T Jr. Carter PA. Evolutionary physiology. *Annual Review of Physiology* 1994;56:579–621.
- Garland T Jr. Harvey PH, Ives AR. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology* 1992;41:18–32.
- Gibbs AG, Chippindale AK, Rose MR. Physiological mechanisms of evolved desiccation resistance in *Drosophila melanogaster*. *J Exp Biol* 1997;200:1821–1832. [PubMed: 9225453]
- Gibson JB, Wilks AV. The alcohol dehydrogenase polymorphism of *Drosophila melanogaster* in relation to environmental ethanol, ethanol tolerance and alcohol dehydrogenase activity. *Heredity* 1988;60(Pt 3):403–414. [PubMed: 3138207]
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–537. [PubMed: 10521349]
- Gross M. Disruptive selection for alternative life histories in salmon. *Nature* 1985;313:47–48.
- Harvell CD. The Ecology and Evolution of Inducible Defenses. *The Quarterly Review of Biology* 1990;65:323–340. [PubMed: 2236483]
- Harvey, PH.; Pagel, MD. *The comparative method in evolutionary biology*. Oxford University Press; New York: 1991.
- Hauser L, Seeb JE. Advances in molecular technology and their impact on fisheries genetics. *Fish and Fisheries* 2008;9:473–486.
- Hill AA, Hunter CP, Tsung BT, Tucker-Kellogg G, Brown EL. Genomic Analysis of Gene Expression in *C. elegans*. *Science* 2000;290:809812.

- Hodges E, Xuan Z, Baliya V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 2007;39:1522–1527. [PubMed: 17982454]
- Hoekstra HE. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 2006;97:222–234. [PubMed: 16823403]
- Hoekstra HE, Nachman MW. Different genes underlie adaptive melanism in different populations of rock pocket mice. *Mol Ecol* 2003;12:1185–1194. [PubMed: 12694282]
- Hoffman JI, Amos W. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Mol Ecol* 2005;14:599–612. [PubMed: 15660949]
- Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genetics* 2001;29:389–395. [print]. [PubMed: 11726925]
- Ju Z, Wells MC, Heater SJ, Walter RB. Multiple tissue gene expression analyses in Japanese medaka (*Oryzias latipes*) exposed to hypoxia. *Comp Biochem Physiol C Toxicol Pharmacol* 2007;145:134–144. Epub 2006 Jun 2030. [PubMed: 16916623]
- Ju Z, Wells MC, Martinez A, Hazlewood L, Walter RB. An in silico mining for simple sequence repeats from expressed sequence tags of zebrafish, medaka, *Fundulus*, and *Xiphophorus*. In *Silico Biol* 2005;5:439–463. [PubMed: 16268789]
- Kalujnaia S, McWilliam IS, Zaguinaiko VA, Feilen AL, Nicholson J, Hazon N, Cutler CP, Cramb G. Transcriptomic approach to the study of osmoregulation in the European eel *Anguilla anguilla*. *Physiol Genomics* 2007;31:385–401. Epub 2007 Jul 2031. [PubMed: 17666525]
- Koehn RK, Milkman R, Mitton JB. Population genetics of marine pelecypods. 4. Selection, migration and genetic differentiation in the blue mussel *Mytilus edulis*. *Evolution* 1976;30:2–32.
- Kohn MH, Pelz HJ, Wayne RK. Locus-specific genetic differentiation at *Rw* among warfarin-resistant rat (*Rattus norvegicus*) populations. *Genetics* 2003;164:1055–1070. [PubMed: 12871915]
- Koufopanou V, Hughes J, Bell G, Burt A. The spatial scale of genetic differentiation in a model organism: the wild yeast *Saccharomyces paradoxus*. *Philos Trans R Soc Lond B Biol Sci* 2006;361:1941–1946. [PubMed: 17028086]
- Kucuktas H, Wang S, Li P, He C, Xu P, Sha Z, Liu H, Jiang Y, Baoprasertkul P, Somridhivej B, Wang Y, Abernathy J, Guo X, Liu L, Muir W, Liu Z. Construction of Genetic Linkage Maps and Comparative Genome Analysis of Catfish Using Gene-associated Markers. *Genetics* 2009;25:25.
- Lam SH, Winata CL, Tong Y, Korzh S, Lim WS, Korzh V, Spitsbergen J, Mathavan S, Miller LD, Liu ET, Gong Z. Transcriptome kinetics of arsenic-induced adaptive response in zebrafish liver. *Physiol Genomics* 2006;27:351–361. Epub 2006 Aug 2001. [PubMed: 16882884]
- Langefors A, Lohm J, Grahn M, Andersen O, von Schantz T. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc Biol Sci* 2001;268:479–485. [PubMed: 11296859]
- Larsen PF, Nielsen EE, Williams TD, Hemmer-Hansen J, Chipman JK, Kruhoffer M, Gronkjaer P, George SG, Dyrskjot L, Loeschcke V. Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Mol Ecol* 2007;16:4674–4683. [PubMed: 17922814]
- Larsen PF, Nielsen EE, Williams TD, Loeschcke V. Intraspecific variation in expression of candidate genes for osmoregulation, heme biosynthesis and stress resistance suggests local adaptation in European flounder (*Platichthys flesus*). *Heredity* 2008;101:247–259. [PubMed: 18560442]
- Lewontin RC, Hubby JL. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 1966;54:595–609. [PubMed: 5968643]
- Li, W-H. *Molecular Evolution*. Sinauer Associate, Inc.; Sunderland, MA: 1997.
- Lohm J, Grahn M, Langefors A, Andersen O, Storset A, von Schantz T. Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. *Proc Biol Sci* 2002;269:2029–2033. [PubMed: 12396502]
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* 2003;4:981–994. [PubMed: 14631358]
- MacCallum C, Hill E. Being positive about selection. *PLoS Biol* 2006;4:e87. [PubMed: 16524343]

- McMillan WO, Palumbi SR. Concordant evolutionary patterns among Indo-West Pacific butterflyfishes. *Proc Biol Sci* 1995;260:229–236. [PubMed: 7784441]
- Mied PA, Powers DA. Hemoglobins of the killifish *Fundulus heteroclitus*. Separation, characterization, and a model for the subunit compositions. *J Biol Chem* 1978;253:3521–3528. [PubMed: 25889]
- Mitton, JB. Selection in Natural Populations. Oxford University Press, Inc.; New York: 1997.
- Mock KE, Theimer TC, Rhodes OE Jr. Greenberg DL, Keim P. Genetic variation across the historical range of the wild turkey (*Meleagris gallopavo*). *Mol Ecol* 2002;11:643–657. [PubMed: 11972754]
- Moen T, Hayes B, Baranski M, Berg PR, Kjolglum S, Koop BF, Davidson WS, Omholt SW, Lien S. A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. *BMC Genomics* 2008;9:223. [PubMed: 18482444]
- Nelson, GJ. Fishes of the World. John Wiley and Sons; New York: 1984.
- Nevo E. Molecular evolutionary genetics of isozymes: pattern, theory, and application. *Prog Clin Biol Res* 1990;344:701–742. [PubMed: 1975442]
- Nielsen R. Statistical tests of selective neutrality in the age of genomics. *Heredity* 2001;86:641–647. [PubMed: 11595044]
- Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME. Microarray-based genomic selection for high-throughput resequencing. *Nat Methods* 2007;4:907–909. [PubMed: 17934469]
- Oleksiak MF. Changes in gene expression due to chronic exposure to environmental pollutants. *Aquat Toxicol* 2008;90:161–171. [PubMed: 18929415]
- Oleksiak MF, Churchill GA, Crawford DL. Variation in gene expression within and among natural populations. *Nature Genetics* 2002;32:261–266. [PubMed: 12219088]
- Oleksiak MF, Kolell KJ, Crawford DL. Utility of natural populations for microarray analyses: Isolation of genes necessary for functional genomic studies. *Marine Biotechnology* 2001;3:S203–S211. [PubMed: 14961317]
- Oleksiak MF, Roach JL, Crawford DL. Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat Genet* 2005;37:67–72. Epub 2004 Nov 2028. [PubMed: 15568023]
- Ono K, Tanaka T, Tsunoda T, Kitahara O, Kihara C, Okamoto A, Ochiai K, Takagi T, Nakamura Y. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. *Cancer Research* 2000;60:5007–5011. [PubMed: 11016619]
- Palumbi SR. Genetic Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology and Systematics* 1994;25:547–572.
- Parsons YM, Shaw KL. Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Mol Ecol* 2001;10:1765–1772. [PubMed: 11472543]
- Paschall J, Oleksiak M, VanWye J, Roach J, Whitehead J, Wyckoff G, Kolell K, Crawford D. FunnyBase: a systems level functional annotation of *Fundulus* ESTs for the analysis of gene expression. *BMC Genomics* 2004;20:96. [PubMed: 15610557]
- Peatman E, Terhune J, Baoprasertkul P, Xu P, Nandi S, Wang S, Somridhivej B, Kucuktas H, Li P, Dunham R, Liu Z. Microarray analysis of gene expression in the blue catfish liver reveals early activation of the MHC class I pathway after infection with *Edwardsiella ictaluri*. *Mol Immunol* 2008;45:553–566. Epub 2007 Jun 2027. [PubMed: 17599411]
- Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM. The genetic architecture of divergence between threespine stickleback species. *Nature* 2001a; 414:901–905. [PubMed: 11780061]
- Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM. The genetic architecture of divergence between threespine stickleback species. *Nature* 2001b; 414:901–905. [PubMed: 11780061]
- Pierce VA, Crawford DL. Phylogenetic analysis of glycolytic enzyme expression. *Science* 1997;276:256–259. [PubMed: 9092475]
- Place AR, Powers DA. Genetic variation and relative catalytic efficiencies: lactate dehydrogenase B allozymes of *Fundulus heteroclitus*. *Proc Natl Acad Sci U S A* 1979;76:2354–2358. [PubMed: 36616]

- Place AR, Powers DA. Kinetic characterization of the lactate dehydrogenase (LDH-B4) allozymes of *Fundulus heteroclitus*. *J Biol Chem* 1984;259:1309–1318. [PubMed: 6693387]
- Pogson GH, Mesa KA, Boutilier RG. Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics* 1995;139:375–385. [PubMed: 7705638]
- Powers DA, Lauerman T, Crawford D, DiMichele L. Genetic mechanisms for adapting to a changing environment. *Annu Rev Genet* 1991;25:629–659. [PubMed: 1812817]
- Powers, DA.; Smith, M.; Gonzalez-Villasenor, I.; DiMichelle, L.; Crawford, DL.; Bernardi, G.; Lauerman, T. A multidisciplinary approach to the selectionist/neutralist controversy using the model teleost, *Fundulus heteroclitus*.. In: Futuyma, D.; Antonovics, J., editors. *Oxford Surveys in Evolutionary Biology*. Oxford University Press; New York, NY: 1993. p. 43-108.
- Quattro JM, Woods HA, Powers DA. Sequence analysis of teleost retina-specific lactate dehydrogenase C: evolutionary implications for the vertebrate lactate dehydrogenase gene family. *Proc Natl Acad Sci U S A* 1993;90:242–246. [PubMed: 8419929]
- Ramsay G. DNA chips: state-of-the art. *Nature Biotechnology* 1998;16:40–44.
- Rexroad CE 3rd, Rodriguez MF, Coulibaly I, Gharbi K, Danzmann RG, Dekoning J, Phillips R, Palti Y. Comparative mapping of expressed sequence tags containing microsatellites in rainbow trout (*Oncorhynchus mykiss*). *BMC Genomics* 2005;6:54. [PubMed: 15836796]
- Riehle MM, Bennett AF, Long AD. Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc Natl Acad Sci U S A* 2001;98:525–530. [PubMed: 11149947]
- Roest Crolius H, Jaillon O, Bernot A, Dasilva C, Bouneau L, Fischer C, Fizames C, Wincker P, Brottier P, Quetier F, Saurin W, Weissenbach J. Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nat Genet* 2000;25:235–238. [PubMed: 10835645]
- Roling JA, Bain LJ, Gardea-Torresdey J, Key PB, Baldwin WS. Using mummichog (*Fundulus heteroclitus*) arrays to monitor the effectiveness of remediation at a superfund site in Charleston, South Carolina, U.S.A. *Environ Toxicol Chem* 2007;26:1205–1213. [PubMed: 17571687]
- Santana Q, Coetzee M, Steenkamp E, Mlonyeni O, Hammond G, Wingfield M, Wingfield B. Microsatellite discovery by deep sequencing of enriched genomic libraries. *Biotechniques* 2009;46:217–223. [PubMed: 19317665]
- Saunders NC, Kessler LG, Avise JC. Genetic Variation and Geographic Differentiation in Mitochondrial DNA of the Horseshoe Crab, *LIMULUS POLYPHEMUS*. *Genetics* 1986;112:613–627. [PubMed: 17246319]
- Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW. Microarrays: biotechnology's discovery platform for functional genomics. *Trends In Biotechnology* 1998;16:301–306. [PubMed: 9675914]
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–470. [see comments]. [PubMed: 7569999]
- Schlotterer C. Towards a molecular characterization of adaptation in local populations. *Curr Opin Genet Dev* 2002;12:683–687. [PubMed: 12433582]
- Schulte PM, Glemet HC, Fiebig AA, Powers DA. Adaptive variation in lactate dehydrogenase-B gene expression: role of a stress-responsive regulatory element. *Proc Natl Acad Sci U S A* 2000;97:6597–6602. [PubMed: 10841559]
- Segal JA, Schulte PM, Powers DA, Crawford DL. Descriptive and functional characterization of variation in the *Fundulus heteroclitus* Ldh-B proximal promoter. *J Exp Zool* 1996;275:355–364. [PubMed: 8691188]
- St-Cyr J, Derome N, Bernatchez L. The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). *Mol Ecol* 2008;17:1850–1870. [PubMed: 18312278]
- Stinchcombe JR, Hoekstra HE. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 2008;100:158–170. [PubMed: 17314923]
- Storz JF. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 2005;14:671–688. [PubMed: 15723660]

- Storz JF, Dubach JM. Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution Int J Org Evolution* 2004;58:1342–1352.
- Storz JF, Nachman MW. Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. *Evolution Int J Org Evolution* 2003;57:2628–2635.
- Treves DS, Manning S, Adams J. Repeated evolution of an acetate- crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol Biol Evol* 1998;15:789–797. [PubMed: 9656481]
- Van Beneden RJ, Powers DA. The isozymes of glucose-phosphate isomerase (GPI-A2 and GPI-B2) from the teleost fish *Fundulus heteroclitus* (L.). *J Biol Chem* 1985;260:14596–14603. [PubMed: 4055792]
- Vasemagi A, Nilsson J, Primmer CR. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in atlantic salmon (*Salmo salar* L.). *Mol Biol Evol* 2005;22:1067–1076. [PubMed: 15689532]
- Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol* 2008;17:1636–1647. [PubMed: 18266620]
- von Schallburg KR, Cooper GA, Leong J, Robb A, Lieph R, Rise ML, Davidson WS, Koop BF. Expansion of the genomics research on Atlantic salmon *Salmo salar* L. project (GRASP) microarray tools. *Journal of Fish Biology* 2008;72:2051–2070.
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, et al. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 1995;23:4407–4414. [PubMed: 7501463]
- Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Zhang G, Yang B, Yu C, Liang F, Li W, Li S, Ni P, Ruan J, Li Q, Zhu H, Liu D, Lu Z, Li N, Guo G, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S, Chen F, Li L, Zhou K, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Yang H. The diploid genome sequence of an Asian individual. *Nature* 2008;456:60–65. [PubMed: 18987735]
- Watt WB. Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: Biochemical and population aspects. *Genetics* 1977;87:177–194. [PubMed: 914029]
- Whitehead A, Crawford DL. Neutral and adaptive variation in gene expression. *Proc Natl Acad Sci U S A* 2006;103:5425–5430. [PubMed: 16567645]
- Wichman HA, Scott LA, Yarber CD, Bull JJ. Experimental evolution recapitulates natural evolution. *Philos Trans R Soc Lond B Biol Sci* 2000;355:1677–1684. [PubMed: 11127915]
- Wiedmann RT, Smith TP, Nonneman DJ. SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genet* 2008;9:81. [PubMed: 19055830]
- Wilding C, Butlin R, Grahame J. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* 2001;14:611–619.
- Williams LM, Oleksiak MF. Signatures of selection in natural populations adapted to chronic pollution. *BMC Evol Biol* 2008;8:282. [PubMed: 18847479]
- Williams ST, Knowlton N, Weigt LA, Jara JA. Evidence for three major clades within the snapping shrimp genus *Alpheus* inferred from nuclear and mitochondrial gene sequence data. *Mol Phylogenet Evol* 2001;20:375–389. [PubMed: 11527465]
- Wilson, DE.; Reeder, DM. *Mammal Species of the World*. Smithsonian Institution Press; 1993.
- Wolfinger, RD.; Gibson, G.; Wolfinger, ED.; Bennet, L.; Hamadeh, H.; Bushel, P.; Afshari, C.; Paules, RS. Assessing gene significant from cDNA microarray expression data via mixed models. 2001. <http://statgen.ncsu.edu/ggibson/Publications/WGetc.pdf>
- Yan G, Romero-Severson J, Walton M, Chadee DD, Severson DW. Population genetics of the yellow fever mosquito in Trinidad: comparisons of amplified fragment length polymorphism (AFLP) and restriction fragment length polymorphism (RFLP) markers. *Mol Ecol* 1999;8:951–963. [PubMed: 10434416]
- Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW. Gene Expression Profiles in Normal and Cancer Cells. *Science* 1997;276:1268–1272. [PubMed: 9157888]

Figure 1. Phylogenetic relationships of teleosts with significant EST collections. Numbers in parentheses are numbers of families, genus, species and ESTs (blue), respectively, in a particular order. Boxed orders contain species with sequenced genomes (or genomes in the process of being sequenced). After Nelson, 1984.

Table I

Teleosts with large (>9,000) EST collections. Numbers in parentheses represent the total number of ESTs in a particular super order or order.

Super Order	Order	Species	Common Name	# ESTs
Acanthopterygii (1,304,844)				
	Scorpaeniformes (23,668)			
		<i>Sebastes caurinus</i>	copper rockfish	23,668
	Tetraodontiformes (26,069)			
		<i>Takifugu rubripes</i>	torafugu	26,069
	Perciformes (270,476)			
		<i>Sparus aurata</i>	porgie	50,930
		<i>Dicentrarchus labrax</i>	European Seabass	44,363
		<i>Dissostichus mawsoni</i>	Antarctic toothfish	37,104
		<i>Paralabidochromis chilotes</i>	African cichlid	21,652
		<i>Hippoglossus hippoglossus</i>	Atlantic halibut	20,836
		<i>Lipochromis</i> sp. 'matumbi hunter'	African cichlid	16,795
		<i>Ptyochromis</i> sp. 'redtail sheller'	African cichlid	14,073
		<i>Psetta maxima</i>	turbot	12,427
		<i>Sebastes rastrelliger</i>	grass rockfish	11,207
		<i>Solea senegalensis</i>	Senegalese sole	10,631
		<i>Haplochromis burtoni</i>	African cichlid	10,312
		<i>Thunnus thynnus</i>	Northern bluefin tuna	10,163
		<i>Paralichthys olivaceus</i>	olive flounder	9,983
	Gasterosteiformes (276,922)			
		<i>Gasterosteus aculeatus</i>	three-spine stickleback	276,992
	Cyprinodontiformes (90,970)			
		<i>Fundulus heteroclitus</i>	killifish	74,755
		<i>Poecilia reticulata</i>	guppy	16,215
	Beloniformes (616,739)			
		<i>Oryzias latipes</i>	medaka	616,739
Paracanthopterygii (206,507)				
	Gadiformes (206,507)			
		<i>Gadus morhua</i>	Atlantic cod	206,507
Protacanthopterygii (867,412)				
	Esociformes (32,833)			
		<i>Esox lucius</i>	Northern pike	32,833
	Salmoniformes (834,579)			
		<i>Salmo salar</i>	Atlantic salmon	494,146
		<i>Oncorhynchus mykiss</i>	rainbow trout	281,335
		<i>Oncorhynchus tshawytscha</i>	chinook salmon	13,965

Super Order	Order	Species	Common Name	# ESTs
		<i>Oncorhynchus nerka</i>	sockeye salmon	11,389
		<i>Thymallus thymallus</i>	grayling	10,975
		<i>Coregonus clupeaformis</i>	lake whitefish	10,312
		<i>Salvelinus fontinalis</i>	brook trout	10,047
	Osmeriformes (36,926)			
		<i>Osmerus mordax</i>	rainbow smelt	36,028
Otocephala (2,307,644)				
	Siluriformes (493,853)			
		<i>Ictalurus punctatus</i>	channel catfish	354,377
		<i>Ictalurus furcatus</i>	blue catfish	139,475
	Cypriniformes (1,813,791)			
		<i>Misgurnus anguillicaudatus</i>	Oriental weatherfish	22,169
		<i>Danio rerio</i>	zebrafish	1,473,159
		<i>Pimephales promelas</i>	fathead minnow	249,941
		<i>Cyprinus carpio</i>	common carp	32,046
		<i>Rutilus rutilus</i>	roach minnow	18,470
		<i>Carassius auratus</i>	goldfish	9,230

Table II

Other fishes with significant EST collections (> 8,000 ESTs). Numbers in parentheses represent the total number of ESTs in a particular class or order.

Class	Order	Species	Common Name	# ESTs
Chondrichthyes (76,075)				
	Squaliformes (32,562)			
		<i>Squalus acanthias</i>	spiny dogfish	32,562
	Rajiformes (31,167)			
		<i>Leucoraja erinacea</i>	little skate	31,167
	Torpediniformes (10,185)			
		<i>Torpedo californica</i>	Pacific electric ray	10,185
	Chimaeriformes (27,944)			
		<i>Callorhynchus milii</i>	elephant fish	27,944*
Petromyzontida (129,520)				
	Petromyzontiformes (129,520)			
		<i>Petromyzon marinus</i>	lamprey	120,731
		<i>Lethenteron japonicum</i>	Arctic lamprey	8,788
Myxini (23,886)				
	Myxiniformes (23,886)			
		<i>Eptatretus burgeri</i>	Inshore hagfish	23,886

* Elephant fish sequences are GSS rather than ESTs. GSS are Genome Survey Sequences. GSS are genomic in origin, rather than cDNA (mRNA). The elephant fish sequences were derived from a whole genome shotgun library.

Table III
Examples of studies using microarrays or ESTs to identify adaptively important changes in gene expression.

Species	Array type	Tissue	Comparison	Major Findings	References
<i>Fundulus heteroclitus</i> (killifish)	cDNA	heart	Northern population versus a southern population and sister taxa from the Gulf. <i>F. grandis</i>	Gene expression for 20% of genes varied significantly between individuals. 15 genes (1.5%) with adaptive signature perhaps due to temperature.	(Oleksiak <i>et al.</i> , 2002)
	cDNA	brain & liver	Three chronically polluted populations, each with two flanking reference populations	Gene expression for 8-32% of genes showed potentially adaptive changes in gene expression. Few significant genes shared among different polluted populations.	(Fisher and Oleksiak, 2007; Oleksiak, 2008)
	cDNA	heart	Clinal variation among 5 populations	Much of variation in gene expression was neutral, but 22% of gene expression was potentially adaptive.	(Whitehead and Crawford, 2006)
<i>Platichthys flesus</i> (European flounder)	cDNA	liver	Baltic Sea <i>versus</i> North Sea population	5% (158) of genes were significantly differently expressed in North Sea and Baltic Sea flounders subjected to the same salinities under controlled conditions.	(Larsen <i>et al.</i> , 2007)
<i>Salmo salar</i> (Atlantic salmon)	cDNA	brain	Sneaker males <i>versus</i> age matched females and immature males	15% of genes differed between sneaker and immature males.	(Aubin-Horth <i>et al.</i> , 2005)
<i>Neolamprologus pulcher</i> (Tanganyikan cichlid)	cDNA	brain	Dominant males & females <i>versus</i> subordinate males and females	Dominant breeder females were masculinized at the molecular and hormonal level.	(Aubin-Horth <i>et al.</i> , 2007)
<i>Coregonus</i> sp. (whitefish)	cDNA	liver	Dwarf <i>versus</i> normal whitefish populations	Significant over-expression of genes potentially associated with enhanced activity, and down-regulation of genes associated with growth in dwarf whitefish.	(St-Cyr <i>et al.</i> , 2008)
<i>Dissostichus mawsoni</i> (Antarctic toothfish)	<i>In silico</i>	Brain, liver, head kidney, & ovary	Antarctic fish <i>versus</i> five temperate/tropical species	177 notothenioid protein families highly expressed compared to the temperate species. Significant Antarctic specific gene duplications.	(Chen <i>et al.</i> , 2008)