SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 30, No. 2, pp. 283–289, 2011

Printed in the USA
DOI: 10.1002/etc.403

Omics and Environmental Science

GENOMIC APPROACHES WITH NATURAL FISH POPULATIONS FROM POLLUTED ENVIRONMENTS

GORAN BOZINOVIC† and MARJORIE F. OLEKSIAK*‡
†North Carolina State University, Raleigh, North Carolina, USA
‡University of Miami, Miami, Florida, USA

(*Submitted* 18 *December* 2009; *Returned for Revision* 1 *February* 2010; *Accepted* 28 *April* 2010)

Abstract—Transcriptomics and population genomics are two complementary genomic approaches that can be used to gain insight into pollutant effects in natural populations. Transcriptomics identify altered gene expression pathways, and population genomics approaches more directly target the causative genomic polymorphisms. Neither approach is restricted to a predetermined set of genes or loci. Instead, both approaches allow a broad overview of genomic processes. Transcriptomics and population genomic approaches have been used to explore genomic responses in populations of fish from polluted environments and have identified sets of candidate genes and loci that appear biologically important in response to pollution. Often differences in gene expression or loci between polluted and reference populations are not conserved among polluted populations, suggesting a biological complexity that we do not yet fully understand. As genomic approaches become less expensive with the advent of new sequencing and genotyping technologies, they will be more widely used in complementary studies. However, although these genomic approaches are immensely powerful for identifying candidate genes and loci, the challenge of determining biological mechanisms that link genotypes and phenotypes remains. Environ. Toxicol. Chem. 2011;30:283–289. © 2010 SETAC

Keywords—*Fundulus heteroclitus* Genomics Natural populations Pollution

GENOMICS WITH NATURAL POPULATIONS

The power of genomic tools is the ability to simultaneously query large portions of the genome of an organism. These tools are just beginning to be harnessed for studies of aquatic toxicology in natural populations. The advantage of genomic approaches used in aquatic toxicology studies of natural populations is that an a priori choice of biomarkers of an adverse effect is not necessary. This can be particularly important when organisms are exposed to mixtures of chemicals, as is often the case in environmental studies. Because the biological targets of pollutants often are not fully known, the global view gained with genomic approaches can be particularly valuable.

Two complementary genomic approaches that are becoming easier and cheaper to implement with non-model, natural populations are microarrays and population genomics. Microarrays allow expression levels of many thousands of genes to be measured simultaneously, giving information about what genes are expressed or repressed where (tissue-specific expression), when, and why (under what conditions). Although enzyme activities or protein concentrations in a cell are not regulated solely by messenger RNA levels, this knowledge can give insights regarding gene function, interacting genes, and gene pathways. Moreover, changes in cell type, development, or physiological states often are affected by changes in messenger RNA. Most microarray studies have examined physiological or dose responses in select species. Far fewer microarray studies have been used to measure changes in gene expression in natural populations.

Population genomic approaches more directly target the underlying variation found in the DNA among individuals and populations. Population genomics involves genome-wide sampling of sequence variation and is based on the premise that demography and the evolutionary history of populations will affect neutral loci similarly, whereas loci under selection will be affected differently [1]. Different population genomic approaches are just now being used with natural aquatic populations to identify adaptively important sequence variations [2]. As more genetic markers become available via high-throughput sequencing technologies and genotyping becomes less expensive, population genomics will be more widely used on a diversity of species.

Together, global gene expression analyses (transcriptomics) and population genomics have the potential to influence studies of diverse aquatic species and populations that offer many different ecological, physiological, and evolutionary questions with respect to pollutant exposure. Diverse fish species offer many different ecological, physiological, and evolutionary lines of query, and as such, provide powerful models for environmental genomics [3].

Although most fish genomes are not yet sequenced, expressed sequence tags (ESTs) have been developed and used with microarrays to provide genome-wide gene expression data within and among natural fish populations. Currently, the National Center for Biotechnology Information database contains EST collections with greater than 10,000 sequences for three dozen teleost species [4]. Most of these EST collections were developed from normalized complementary DNA (cDNA) libraries and sequenced using Sanger sequencing [5], which is relatively expensive compared with newer sequencing technologies. New sequencing technologies such as Roche (454) FLX, Illumina Genome Analyzer, ABI SOLiD, and HeliScope sequencing (see Morozova et al. [6] for a review of next-generation sequencing) can be used to sequence thousands of

Presented at the 30th Annual Meeting, SETAC North America, New Orleans, Louisiana, USA, November 19–23, 2009.
* To whom correspondence may be addressed (moleksiak@rsmas.miami.edu).
Published online 11 November 2010 in Wiley Online Library (wileyonlinelibrary.com).

transcripts at a time much more quickly and cost-effectively. In contrast to the ESTs from cDNA libraries, one does not gain a physical clone of each sequence obtained using high-throughput sequencing methods. However, the resulting sequences can and have been used to design microarrays for studies of natural populations [7]. Thus, high-throughput sequencing technologies make transcriptomic studies possible in virtually any species.

MICROARRAYS WITH POLLUTED FISH POPULATIONS

Gene expression often is altered as a result of toxicant exposure [8,9] and thus is a sensitive, measurable endpoint for toxicity that can serve as an early warning of compromised health. The challenges are to identify genes that respond to toxicant exposure, to discover novel gene interactions, and to improve knowledge of complex regulatory networks and cross-communication between different pathways in both health and disease. Microarrays that measure most, if not all, genes expressed in a tissue can be used to discover which of these genes respond.

Importantly, microarrays can be used to examine polymorphic responses among individuals. Because individuals have variation in gene expression [10,11], one must quantify individual variation in gene expression within and among populations. In so doing, genes consistently associated with exposure to xenobiotics can be identified. This knowledge serves two purposes: to identify genes that respond to and affect exposure to xenobiotics and to identify genes that have little variation within polluted populations and thus are of greater utility as biomarkers in future analyses of toxicant effects on gene expression. Finally, knowledge of individual variation within a natural population, which cannot be inferred using cell cultures or inbred strains, permits realistic extrapolations to highly variable human populations.

EUROPEAN FLOUNDER (*PLATICHTHYS FLESUS*)

Multiple transcriptomic studies have been done on natural populations of European flounder (*Platichthys flesus*) collected from polluted sites (Table 1). This species was chosen as a good candidate for pollution monitoring because of its wide distribution throughout estuarine environments in Western Europe, many of which are significantly impacted by toxicants including xenoestrogens, heavy metals, polycyclic aromatic hydrocarbons (PAHs), and dioxins [12]. The first microarray study used a targeted array on liver samples of fish collected from polluted and reference estuaries and identified 11 transcripts (7%) with significantly different expression between males from polluted and reference sites. Although a number of transcripts were differentially expressed between females from reference and polluted sites, the altered expression profiles were not significant because of high interindividual variation. The European flounder array has since been expanded to approximately 14,000 elements representing 3,336 unique expressed sequence tag clusters. This expanded array was used with male flounder collected from six sampling sites with differing levels of heavy metals, petroleum hydrocarbons, PAHs, and polychlorinated biphenols (PCBs) to determine gene expression differences related to pollutant impact [13]. Combined with expression data from acute laboratory exposure studies, a model was developed to identify the site of origin of field-collected fish. Surprisingly, only 16 clones predicted the site of origin of an independent sample of 15 fish from three of the sites with 100% accuracy. These data suggest that some expression changes

Table 1. Studies using microarrays with natural fish populations from polluted areas.

Species	Major pollutants	Tissue	Comparison	Major findings	References
European flounder (*Platichthys flesus*)	Heavy metals, petroleum hydrocarbons, PAHs[a], and PCBs[b]	Liver	One reference population versus any one of six polluted populations	Identified 16 clones predictive of the site of origin	[13]
Hornyhead turbot (*Pleuronichthys verticalis*)	Chemical and waste effluents	Liver	Fish from two different polluted sites compared with reference fish	Altered gene expression in genes with potential endocrine effects	[14]
Atlantic cod (*Gadus morhua*)	Cd, Zn, Pb, Cu, and Hg and chlororganic compounds PCBs and dioxin-like PCBs	Liver	Fish from two different polluted sites compared with reference fish, and farmed fish compared with reference fish	Most genes with altered expression were sex and site specific. Farmed fish were most different from reference fish.	[15]
Killifish (*Fundulus heteroclitus*)	Chromium	Liver	Fish from a chromium contaminated Superfund site *versus* reference fish	Gene expression patterns of fish captured at the Superfund site became increasingly similar to those of fish captured at the reference site during remediation of the Superfund site.	[19]
	PCBs, PAHs, pesticides, and heavy metals	Brain and liver	Three chronically polluted populations, each compared 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.	[20,21]
		Stage 31 embryos	Embryos from parents collected from two chronically polluted populations compared with three reference populations	No genes were significantly different, suggesting canalization of developmental gene expression and gene by environment effects.	[25]
Tomcod (*Microgadus tomcod*)	PCBs	Embryos	Embryos from parents collected one chronically polluted population compared with two reference populations	Altered expression of genes representing biomarkers of cardiomyopathy in mammals	[26]

[a] Polycyclic aromatic hydrocarbons.
[b] Polychlorinated biphenyls.

caused by short-term exposures can be used to infer longer-term exposures accumulated over months or years of exposure [13].

HORNYHEAD TURBOT (*PLEURONICHTHYS VERTICALIS*)

Similar pollutant monitoring approaches using microarrays have been implemented with other fish species. Baker et al. [14] designed a microarray as a diagnostic tool to screen the endocrine effects of environmental chemicals in fish: microarray probes were based on sequences from Tetraodoniformes (*Fugu* and *Tetraodon*), Perciformes (cichlid, tilapia, sea bass, sea bream), available sequences from medaka, stickleback, and zebrafish, as well as hornyhead turbot-specific cDNA sequences obtained by degenerate polymerase chain reaction cloning [14]. The goal of the study was to quantify gene expression in wild hornyhead turbot (*Pleuronichthys verticalis*) collected from polluted (chemical and waste effluent sites) and clean coastal waters in Southern California, USA. Similar to European flounder, the bottom-feeding hornyhead turbot are a sentinel fish because they are at higher risk of exposure to chemicals that accumulate in sediments. These fish also have a restricted range, which allows one to localize the site of chemical pollution. Hornyhead turbot sampled at two polluted sites exhibited strong increases in the expression of cytochrome P4503A, retinoid X receptor, estrogen receptor β, vitellogenin 2, and aldosterone receptor relative to control fish. The vitamin D receptor also was up-regulated compared with control fish, whereas transcripts encoding the farnesoid X receptor, peroxisome proliferator-activated receptor α, and thyroid receptors α and β were down-regulated [14]. These data suggest that endocrine disruptors are affecting hepatic gene expression in the pollution-impacted hornyhead turbots.

ATLANTIC COD (*GADUS MORHUA*)

Atlantic cod (*Gadus morhua*) has been used for environmental monitoring and as a model species to investigate biological effects of contaminants emitted from industry and the petroleum sector in Norwegian waters [15]. A custom cDNA array, the CodStress array, which contains 746 stress-responsive genes and genes relevant to immunity, was used to quantify gene expression in two natural populations of Atlantic cod from western Norway. Fish were collected from two sites: one site was highly polluted with a wide range of metals (Cd, Zn, Pb, Cu, and Hg) and chlororganic compounds; a second site had high levels of PCBs and dioxin-like PCBs [15]. Similar to the European flounder study, most altered genes were sex-specific, and no genes were significantly altered in both sexes and in both sites. The authors also compared gene expression profiles of farmed fish with gene expression profiles of fish from the reference site; surprisingly, the largest number of significantly differentially expressed genes was detected in the farmed group compared with those from the reference site [15]. These results highlight the importance of distinguishing impacts of contaminants from biogeographic structuring or evolutionary history.

KILLIFISH (*FUNDULUS HETEROCLITUS*)

Targeted arrays also have been used with *Fundulus heteroclitus. Fundulus heteroclitus* is an important model organism found in estuarine habitats along the east coast of the United States. Extensive background information is available on the ecology, reproduction, physiology, biochemistry, and pathology of this fish [16]. Importantly, *F. heteroclitus* has been the subject of extensive studies involving the fate and effects of environmental chemicals (reviewed in Burnett et al. [16]).

Fundulus heteroclitus populations within major bays and estuaries are genetically distinct, and adaptive differences among populations are not uncommon [17]. Particularly striking are populations of *F. heteroclitus* that have adapted to highly polluted environments. These populations show evolved tolerance or resistance to the pollutants in their environment, in contrast to individuals from reference (unpolluted) sites: resistant individuals have been shown to be more than three orders of magnitude more tolerant to environmental pollutants (most notably to dioxin-like compounds) compared with individuals from nearby reference populations [18].

The first arrays used to examine polluted populations of *F. heteroclitus* were targeted arrays. Roling et al. [19] used a cDNA array containing genes differentially expressed on exposure to PAHs and several metals to monitor a Cr-contaminated Superfund site (Shipyard Creek, Charleston, SC, USA) undergoing remediation [19]. As remediation progressed, the gene expression patterns of *F. heteroclitus* captured at the Superfund site became increasingly similar to those of individuals captured at a nearby reference site. Furthermore, the gene expression profiles associated with chromium bioavailability and body burden demonstrate the use of arrays as multidimensional biomarkers to monitor site mitigation.

In addition to use for monitoring purposes, targeted metabolic arrays were used to discern evolved responses to pollutants from adaptive ones in three independent, polluted populations of *F. heteroclitus* [20,21]. In these studies, gene expression was quantified in individuals collected from three Superfund sites (New Bedford Harbor, MA, Newark Bay, NJ, and Elizabeth River, VA, USA; Fig. 1). New Bedford Harbor is highly polluted with PCBs and other halogenated aromatic hydrocarbons [22], Newark Bay is polluted with a variety of pollutants including PAHs, PCBs, pesticides, and metals [23],

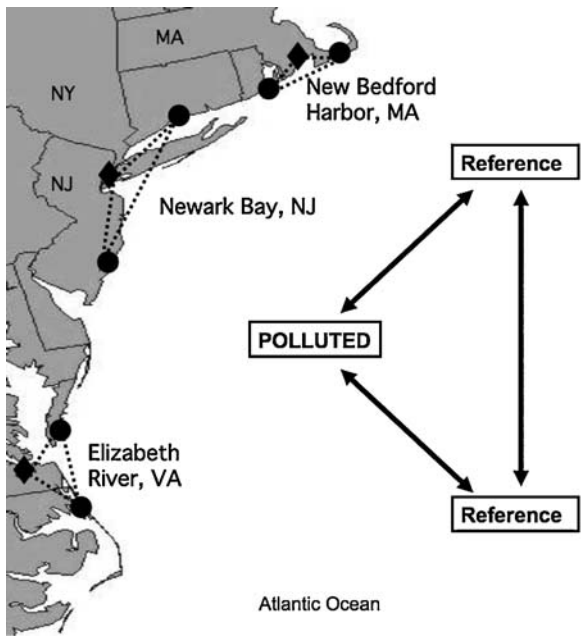

Fig. 1. Sampling sites along the East Coast of the United States and experimental contrasts for *Fundulus heteroclitus* populations. Diamonds are polluted (Superfund) sites, and circles are reference sites. Variation among the two reference sites captures the random/neutral variation. Thus, differences in a polluted population versus both reference populations most likely are attributable to evolved responses to pollution.

and the Elizabeth River is contaminated with creosote, a complex mixture of PAHs [24]. In contrast to the *F. heteroclitus* collected in Shipyard Creek, fish from New Bedford Harbor, Newark Bay, and Elizabeth River populations were common-gardened in the laboratory for four months to one year to minimize physiologically induced differences. Gene expression patterns in individuals from each polluted site were compared with gene expression patterns in individuals from flanking reference sites, one north and one south of each polluted site (Fig. 1). This experimental design was used because differences shared between each polluted site and both reference populations (i.e., polluted vs both reference 1 and reference 2) are more likely to be attributable to pollution (i.e., the effects of pollution are no longer confounded with either temperature or genetic drift). Using this design, gene expression was quantified in both liver and brain from *F. heteroclitus* collected from New Bedford Harbor, Newark Bay, and Elizabeth River, and compared with gene expression levels in individuals from flanking reference sites. A number of genes (5–32%) showed significantly altered gene expression patterns that likely evolved because of chronic exposure to pollution [20,21]. Although many genes (up to 32%) had significantly altered gene expression patterns, few genes were similarly altered in all three polluted populations (Fig. 2). The lack of concordance in genes with altered expression patterns among polluted populations is perhaps not surprising given the complexity of the polluted environments as well as the different pollutants found at each site. These results are similar to results in European flounder and Atlantic cod, in which different polluted populations showed largely unique patterns of altered gene expression.

A similar experimental design to that used with the nine adult *F. heteroclitus* populations was used on stage 31 *F. heteroclitus* embryos collected from parents from Newark Bay, Elizabeth River, and reference populations. This study used a larger array representing approximately 7,000 unique expressed sequence tag clusters. In surprising contrast to the large number of differentially expressed genes found in the adult studies, no genes were differentially expressed between embryos from polluted and reference populations [25]. This lack of difference in gene expression in a common garden environment suggests that gene expression is canalized during development and that gene-by-environment interactions may be necessary to clarify differences among embryos from polluted and reference populations.

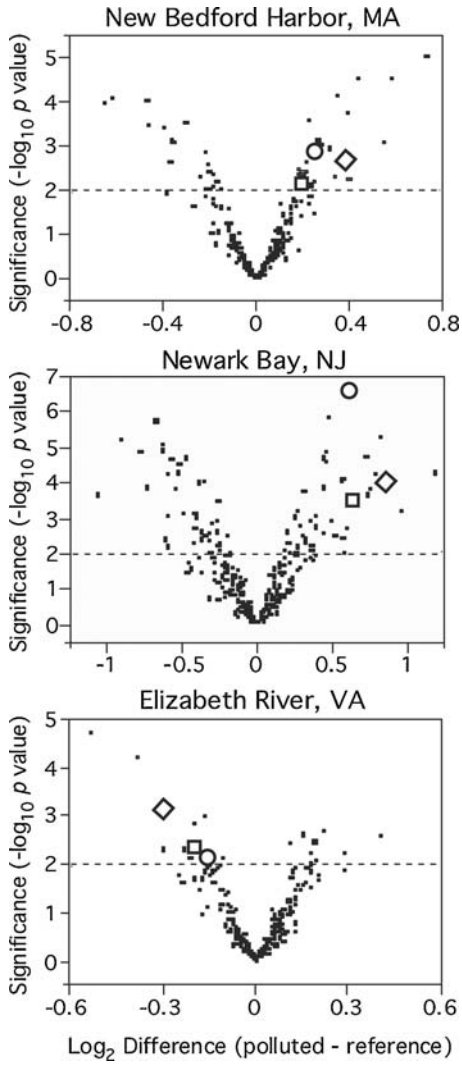

Fig. 2. Gene expression differences between polluted and reference populations. Significances of differences as $-\log_{10}(p$ values) are plotted against $\log_2$ differences in expression for the New Bedford Harbor, Newark Bay, and Elizabeth River triads (USA). The only three genes with significantly different expression in all three polluted populations show opposite patterns of expression in New Bedford Harbor and Newark Bay *Fundulus heteroclitus* as compared with Elizabeth River *F. heteroclitus*. Open diamonds: Acyl C-A binding protein; open squares: NADH ubiquinone oxidoreductase MNLL subunit; open circles: Thioredoxin.

TOMCOD (*MICROGADUS TOMCOD*)

Similar to *F. heteroclitus*, populations of tomcod (*Microgadus tomcod*) show differential sensitivity to dioxin-like compounds; embryos from the Hudson River (NY, USA) population are resistant to PCB-induced cytochrome P4501A (CYP1A) expression and early life stage toxicity compared with embryos from reference sites. The basis of this resistance was explored using microarrays made from a cardiac cDNA library [26]. Although the authors did not find any genes with a similar expression profile as cytochrome P4501A on exposure to PCBs (i.e., PCB induced expression in reference but not polluted fish), they did find that several transcripts representing biomarkers of cardiomyopathy in mammals (cardiac troponin T2, cathepsin L, and atrial natriuretic peptide) were differentially altered by PCBs among polluted and reference tomcod populations.

Thus, microarray studies have been used to detect both adaptive and evolved responses to pollutants. Further studies often are needed to understand the mechanistic implications of changes in gene expression. Even so, microarray studies broaden our understanding of biological processes affected by pollution, provide new directions for future studies, and are useful in environmental impact assessments.

TRANSCRIPTOME ANALYSIS USING HIGH-THROUGHPUT SEQUENCING

Currently, microarrays are the least expensive means to interrogate the transcriptome of an organism, but microarrays suffer from two major drawbacks. First, quantification of transcript abundance is indirect: similar to Northern analyses of gene expression, quantification of gene transcripts using microarrays is based on hybridization intensity and is not a direct measure of transcript abundance. Second, unless a gene is on the microarray, it will be missed. Thus, expression of novel genes will be missed. These drawbacks are overcome by new, high-throughput sequencing technologies that can be used for

transcriptome analyses. High-throughput sequencing allows sufficient sequencing depth to gain an adequate representation of all of the expressed transcripts, and Roche (454) FLX, Illumina Genome Analyzer, and ABI SOLiD sequencing technologies have been used for whole transcriptome sequencing [6]. Unlike microarrays, whole transcriptome sequencing data provide sequences from any location within all expressed transcripts. In addition to gene expression profiling, the sequencing data allows analysis of coding sequences of the expressed transcripts; these sequences can give insights into protein polymorphisms, mutations, and altered splice sites.

Whole transcriptome sequencing has many advantages compared with microarray analyses, yet cost still remains a major factor in using this approach. This is particularly the case in natural populations in which the transcriptomes of more than a few individuals would need to be sequenced to get an adequate representation of the population. Moreover, the data analysis of short sequence reads is greatly facilitated by the presence of a reference genome sequence against which to align the short sequences, and these reference genomes are still lacking for most organisms. Even for model species, more sophisticated data analyses are needed to better understand the massive amounts of sequence data generated by new sequencing methods. Thus, until cost is significantly decreased and data analysis is optimized for non-model species, whole transcriptome sequencing is unlikely to be widely used for quantitative gene expression studies of natural populations.

GENOME SCANS

Whole transcriptome approaches can identify significant changes in gene expression that are potentially biologically important. Yet, the genomic basis underlying these altered patterns of gene expression is mostly unknown. Population genomics is one approach to better understand the genomic basis of adaptive divergence, and it requires large numbers of loci. These loci, genetic markers in the genome, are statistically analyzed for nonrandom patterns [1]. At loci under selection, local adaptation and directional selection should reduce genetic variability within populations and increase variation among populations. Thus, statistical tests are used to determine whether an individual locus acts as an outlier compared with loci in the rest of the genome [27]. Population genomic approaches are useful for understanding adaptation in genetically uncharacterized natural populations [1], because genetic markers can be generated without a sequenced genome. Additional advantages of population genomics approaches include 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 [28].

Loci used in population genomics studies include microsatellites, single nucleotide polymorphisms (SNPs), amplified fragment length polymorphisms (AFLPs), randomly amplified polymorphic DNA (RAPDs), and sequences (e.g., whole genome sequences). Microsatellites often are used for genetic studies because they exhibit high levels of variation. This high variation provides high statistical power. However, microsatellites require species-specific marker development, which can be labor intensive. Microsatellites also suffer from a high potential for null alleles and, because of their size-based nature, are prone to genotyping errors. These limitations, combined with the fact that experience and expertise are required for analysis of microsatellite loci, make data comparisons among laboratories and with published data challenging [29].

Single nucleotide polymorphisms (SNPs) have recently attracted the interest of both geneticists and biotechnology companies; although SNPs also require species-specific marker development, they have potential for automation and high throughput as well as lower genotyping errors than microsatellites [30]. Single nucleotide polymorphisms analyses can use DNA of varying quality, and because SNPs provide sequence data, the data are directly transferable among different laboratories. In humans, SNPs occur on average every 1,000 to 2,000 bases when two individuals are compared [31], but discovering SNPs is not trivial in organisms with no genomic information. One straightforward method to identify SNPs is to mine EST data collections for putative SNPs [32]. As long as the ESTs were derived from multiple individuals, ideally from multiple populations that encompass all populations of interest, putative SNPs can be identified in sequence alignments. One advantage of this approach to SNP discovery is that any SNPs identified as important through population genomic approaches are already associated with a sequenced gene transcript. Because generating EST data as well as genomic sequences for any species is relatively simple and inexpensive with high-throughput sequencing technologies, SNPs and microsatellites now are relatively easy to isolate.

Current, high throughput, genotyping technologies examine genetic variation in the form of SNPs because of ease of automation. New genotyping platforms (for a review of different platforms, see Ragoussis [33]) can assay more than one million SNPs at a time and most often are used for genome-wide association studies. Because of the loss of power attributable to multiple test corrections (significance for one million tests is quite low), such large genotyping efforts often are followed by genotyping of smaller marker sets (10–20,000), often using larger sample sizes than the original study. High-throughput genotyping assays capable of assaying millions of markers at the same time are available off the shelf only for model organisms. For nonmodel organisms, lower-throughput, custom platforms are available [34]. Although these platforms are robust and can be used for high sample (vs genotype) throughput, because of their custom nature, they are much more expensive (in the region of 100- to 1,000-fold [33]) than off-the-shelf, high-throughput platforms. Yet even with the additional cost, custom SNP genotyping platforms provide a means to quickly and accurately genotype hundreds to thousands of individuals [32]. Genotypes from large numbers of individuals allow one to use population genomic approaches with multiple populations to discover biologically important sequence variations.

Compared with other genetic markers, the big advantage of AFLPs and RAPDs is that no prior sequence information is required. Thus, without a genome, these markers provide the easiest and cheapest method to genotype large numbers of loci, and they have been widely used since first described [35] because of their ease of use in any species. Shortcomings of AFLP and RAPD analyses are that they require high-quality DNA and provide only dominant markers, so heterozygotes cannot be directly measured. Similar to microsatellites, AFLP and RAPD loci can be scored incorrectly because of their size-based nature (the same-sized band can represent multiple loci). One big disadvantage of AFLPs and RAPDs as well as many SNPs and microsatellites used in population genomics is that they often are anonymous or in nonfunctional areas of the genome, so linking a candidate locus to a causative gene or mutation requires further experiments.

Many studies have used different genetic markers to examine the effect of environmental contaminants on the genetic structures of populations (see Theodorakis et al. [36] and references therein). However, although population genomic approaches using high-density genetic markers have been used with a variety of fish species to detect selectively important loci [32,37–40], these approaches have rarely been used with polluted fish populations. We used AFLP analyses with polluted and reference *F. heteroclitus* populations to identify loci indicative of adaptation in the polluted populations [41]. Thus, similar to the gene expression studies with *F. heteroclitus*, AFLPs were used to contrast resistant populations of *F. heteroclitus* from three polluted environments with six reference or nonpolluted populations (Fig. 1). A population genomics modeling approach was used to test for outlier loci [41]. Briefly, one looks for outliers within hundreds of loci, assuming that a majority of differences among loci are either nonfunctional (a polymorphism with no phenotypic effect) or neutral (variation whose phenotypic effect has no or little fitness consequences; i.e., the selective advantage is $< 1/2N_e$). The outlier loci have significantly different divergence versus a random permutation or modeling of the data. Statistical modeling of the AFLP data implicated 1 to 6% of loci as being under selection or linked to areas of the genome under selection in the three polluted *F. heteroclitus* populations [41]. However, similar to the gene expression analyses in these same populations, few loci were shared outliers between and among the polluted populations [41]. This suggests a complexity in response to these highly polluted environments that we are just beginning to understand.

FUTURE DIRECTIONS

The challenges of Omics technologies are similar to the challenges of basic biological research: inferring biological importance from the data. However, instead of a single gene, protein, or polymorphism, one potentially has thousands of genes, proteins, and polymorphisms to examine. The biggest shortcoming in using natural populations with Omics studies is the lack of biological replicates. Yet one of the biggest advantages of natural populations is the use of biologically realistic, genetically variable individuals to quantify individual response because susceptibility to toxic substances released into the environment depends on the organism's genotype and the manner in which it interacts with the polluted environment [42]. Thus, Omic approaches that analyze genetically variable individuals allow one to extrapolate a population response rather than a strain-specific response.

Natural populations introduce additional challenges by inhabiting uncontrolled environments. In natural environments, pollutants most often are present as complex mixtures rather than discrete chemicals. A more focused approach, using known inducers and inhibitors present in such mixtures and analyzing responses of specific targets, such as tissues or cell types, can provide insights into mechanisms of toxicity seen in natural populations. Thus, just as the integration of laboratory and field studies has provided insights into studies of particular genes and proteins, so too will the integration of laboratory and field studies be critical for Omic approaches. For example, gene expression profiles of the European flounder collected from multiple polluted environments were interpreted using comparisons with gene expression patterns measured in flounder exposed to selected toxicants under controlled laboratory conditions [13].

Just as combinations of field and laboratory studies are valuable for understanding the biological significance of omic data, so too are combinations of different traits. For instance, correlations between gene expression and altered phenotypes have been used to implicate several hundred candidate genes that affect different phenotypes in 40 unexposed fly strains derived from one natural population [43]. Thus, gene expression profiles accompanied by toxicant-induced changes in morphology, histopathology, physiology, biochemistry, and behavior during pollutant exposure can be a powerful tactic to identify candidate genes that play a critical role during stress response.

Finally, the results of different Omics technologies are complementary. Population genomics approaches provide insights into the sequence variations that underlie differences in gene expression and protein polymorphisms. Transcriptomics provide insights into altered protein levels (proteomics). Hence, analyses that combine multiple Omics approaches give a better understanding of biological effects at multiple levels. Undoubtedly, multiple genomic approaches will be invaluable for studies of aquatic toxicology in natural populations.

Acknowledgement—This work was supported by the National Institutes of Health (RO1 ES011588, P42 ES010356, and P42 ES007381).

REFERENCES

1. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: From genotyping to genome typing. *Nat Rev Genet* 4:981–994.
2. Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D. 2009. Population genomics of marine fishes: identifying adaptive variation in space and time. *Mol Ecol* 18:3128–3150.
3. Cossins AR, Crawford DL. 2005. Fish as models for environmental genomics. *Nat Rev Genet* 6:324–333.
4. Oleksiak MF. 2010. Genomic approaches with natural fish populations. *J Fish Biol* 76:1067–1093.
5. Oleksiak MF, Kolell KJ, Crawford DL. 2001. Utility of natural populations for microarray analyses: isolation of genes necessary for functional genomic studies. *Mar Biotechnol (NY)* 3:S203–211.
6. Morozova O, Hirst M, Marra MA. 2009. Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* 10:135–151.
7. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol* 17:1636–1647.
8. Thomas RS, Rank DR, Penn SG, Zastrow GM, Hayes KR, Pande K, Glover E, Silander T, Craven MW, Reddy JK, Jovanovich SB, Bradfield CA. 2001. Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol Pharmacol* 60:1189–1194.
9. Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, Sieber S, Bennett L, Tannant R, Stoll R, Barrett JC, Blanchard K, Paules RS, Afshari CA. 2002. Gene expression analysis reveals chemical-specific profiles. *Toxicol Sci* 67:219–231.
10. Oleksiak MF, Churchill GA, Crawford DL. 2002. Variation in gene expression within and among natural populations. *Nat Genet* 32:261–266.
11. Oleksiak MF, Roach JL, Crawford DL. 2005. Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat Genet* 37:67–72.
12. Williams TD, Gensberg K, Minchin SD, Chipman JK. 2003. A DNA expression array to detect toxic stress response in European flounder (*Platichthys flesus*). *Aquat Toxicol* 65:141–157.
13. Falciani F, Diab AM, Sabine V, Williams TD, Ortega F, George SG, Chipman JK. 2008. Hepatic transcriptomic profiles of European flounder (*Platichthys flesus*) from field sites and computational approaches to predict site from stress gene responses following exposure to model toxicants. *Aquat Toxicol* 90:92–101.
14. Baker ME, Ruggeri B, Sprague LJ, Eckhardt-Ludka C, Lapira J, Wick I, Soverchia L, Ubaldi M, Polzonetti-Magni AM, Vidal-Dorsch D, Bay S, Gully JR, Reyes JA, Kelley KM, Schlenk D, Breen EC, Sasik R, Hardiman G. 2009. Analysis of endocrine disruption in Southern California coastal fish using an aquatic multispecies microarray. *Environ Health Perspect* 117:223–230.

15. Lie KK, Lanzen A, Breilid H, Olsvik PA. 2009. Gene expression profiling in Atlantic cod (*Gadus morhua* L.) from two contaminated sites using a custom-made cDNA microarray. *Environ Toxicol Chem* 28: 1711–1721.
16. 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. 2007. Fundulus as the premier teleost model in environmental biology: Opportunities for new insights using genomics. *Comp Biochem Physiol Part D Genomics Proteomics* 2:257–286.
17. Powers DA, Ropson I, Brown DC, van Beneden R, Cashon R, Gonzalez-Villasenor LA, DiMichele JA. 1986. Genetic variation in *Fundulus heteroclitus*: Geographic distribution. *Am Zool* 26:131–144.
18. Van Veld PA, Nacci DE. 2008. Toxicity resistance. In Di Giulio RT, Hinton DE, eds, *The Toxicology of Fishes*. Taylor and Francis, Boca Raton, FL, USA, pp 597–641.
19. Roling JA, Bain LJ, Gardea-Torresdey J, Key PB, Baldwin WS. 2007. 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* 26:1205–1213.
20. Fisher MA, Oleksiak MF. 2007. Convergence and divergence in gene expression among natural populations exposed to pollution. *BMC Genomics* 8:108.
21. Oleksiak MF. 2008. Changes in gene expression due to chronic exposure to environmental pollutants. *Aquat Toxicol* 90:161–171.
22. Pruell RJ, Norwood CB, Bowen RD, Boothman WS, Rogerson PF, Hackett M, Butterworth BC. 1990. Geochemical study of sediment contamination in New Bedford Harbor, Massachusetts. *Mar Environ Res* 29:77–102.
23. Wolfe DA, Long ER, Thursby GB. 1996. Sediment toxicity in the Hudson-Raritan Estuary: Distribution and correlations with chemical contamination. *Estuaries* 19:901–912.
24. Bieri RH, Hein C, Huggett RJ, Shou P, Slone H, Smith H, Su C-W. 1986. Polycyclic Aromatic Hydrocarbons in Surface Sediments from the Elizabeth River Subesturary. *Int J Environ Anal Chem* 26:97–113.
25. Bozinovic G, Oleksiak MF. 2010. Embryonic gene expression among pollutant resistant and sensitive *Fundulus heteroclitus* populations. *Aquat Toxicol* 98:221–229.
26. Carlson EA, Roy NK, Wirgin II. 2009. Microarray analysis of polychlorinated biphenyl mixture-induced changes in gene expression among Atlantic tomcod populations displaying differential sensitivity to halogenated aromatic hydrocarbons. *Environ Toxicol Chem* 28:759–771.
27. Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 14:671–688.
28. Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100:158–170.
29. Hoffman JI, Amos W. 2005. Microsatellite genotyping errors: Detection approaches, common sources and consequences for paternal exclusion. *Mol Ecol* 14:599–612.
30. Hauser L, Seeb JE. 2008. Advances in molecular technology and their impact on fisheries genetics. *Fish Fish* 9:473–486.
31. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928–933.
32. Moen T, Hayes B, Nilsen F, Delghandi M, Fjalestad KT, Fevolden SE, Berg PR, Lien S. 2008. Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC Genet* 9:18.
33. Ragoussis J. 2009. Genotyping technologies for genetic research. *Annu Rev Genomics Hum Genet* 10:117–133.
34. De la Vega FM, Lazaruk KD, Rhodes MD, Wenz MH. 2005. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP genotyping assays and the SNPlex genotyping system. *Mutat Res* 573:111–135.
35. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Homes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407–4414.
36. Theodorakis CW, Lee KL, Adams SM, Law CB. 2006. Evidence of altered gene flow, mutation rate, and genetic diversity in redbreast sunfish from a pulp-mill-contaminated river. *Environ Sci Technol* 40: 377–386.
37. Makinen HS, Cano JM, Merila J. 2008. Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Mol Ecol* 17: 3565–3582.
38. Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger GL, Ogden R, Carvalho GR. 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evol Biol* 9:276.
39. Rogers SM, Bernatchez L. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. *Salmonidae*) species pairs. *Mol Biol Evol* 24:1423–1438.
40. Vasemagi A, Nilsson J, Primmer CR. 2005. 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* 22:1067–1076.
41. Williams LM, Oleksiak MF. 2008. Signatures of selection in natural populations adapted to chronic pollution. *BMC Evol Biol* 8:282.
42. Jelinek R. 2005. The contribution of new findings and ideas to the old principles of teratology. *Reprod Toxicol* 20:295–300.
43. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF. 2009. Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet* 41:299–307.