BMC
Evolutionary Biology

Targeted approach to identify genetic loci associated with evolved dioxin tolerance in Atlantic Killifish (*Fundulus heteroclitus*)

Dina A Proestou[1,3*], Patrick Flight[2†], Denise Champlin[1†] and Diane Nacci[1]

Abstract

Background: The most toxic aromatic hydrocarbon pollutants are categorized as dioxin-like compounds (DLCs) to which extreme tolerance has evolved independently and contemporaneously in (at least) four populations of Atlantic killifish (*Fundulus heteroclitus*). Surprisingly, the magnitude and phenotype of DLC tolerance is similar among these killifish populations that have adapted to varied, but highly aromatic hydrocarbon-contaminated urban/industrialized estuaries of the US Atlantic coast. Multiple tolerant and neighboring sensitive killifish populations were compared with the expectation that genetic loci associated with DLC tolerance would be revealed.

Results: Since the aryl hydrocarbon receptor (AHR) pathway partly or fully mediates DLC toxicity in vertebrates, single nucleotide polymorphisms (SNPs) from 42 genes associated with the AHR pathway were identified to serve as targeted markers. Wild fish (N = 36/37) from four highly tolerant killifish populations and four nearby sensitive populations were genotyped using 59 SNP markers. Similar to other killifish population genetic analyses, strong genetic differentiation among populations was detected, consistent with isolation by distance models. When DLC-sensitive populations were pooled and compared to pooled DLC-tolerant populations, multi-locus analyses did not distinguish the two groups. However, pairwise comparisons of nearby tolerant and sensitive populations revealed high differentiation among sensitive and tolerant populations at these specific loci: AHR 1 and 2, cathepsin Z, the cytochrome P450s (CYP1A and 3A30), and the NADH dehydrogenase subunits. In addition, significant shifts in minor allele frequency were observed at AHR2 and CYP1A loci across most sensitive/tolerant pairs, but only AHR2 exhibited shifts in the same direction across all pairs.

Conclusions: The observed differences in allelic composition at the AHR2 and CYP1A SNP loci were identified as significant among paired sensitive/tolerant populations of Atlantic killifish with multiple statistical tests. The genetic patterns reported here lend support to the argument that AHR2 and CYP1A play a role in the adaptive response to extreme DLC contamination. Additional functional assays are required to isolate the exact mechanism of DLC tolerance.

Keywords: Adaptation, Ecotoxicology, Candidate gene scan, Killifish

* Correspondence: dina.proestou@ars.usda.gov
†Equal contributors
[1]US Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, 27 Tarzwell Drive, Narragansett, RI 02882, USA
[3]Current address: USDA Agricultural Research Service, 469 Center for Biotechnology and Life Sciences, 120 Flagg Road, Kingston, RI 02881, USA
Full list of author information is available at the end of the article

Background

Dioxin-like contaminants (DLCs), such as some polychlorinated biphenyls (PCBs), are highly toxic aromatic hydrocarbon pollutants whose ubiquitous occurrence presents global ecological and human health risks. The early life stages of fish are particularly sensitive to these toxic DLCs, and the Atlantic killifish, *Fundulus heteroclitus*, is one of the more sensitive fish species [1]. Despite this species' relative sensitivity to DLC exposure, several wild killifish populations residing in heavily contaminated North American Atlantic coast estuaries have recently and independently evolved dramatic, heritable, and adaptive tolerance to DLCs [2], for which the mechanistic basis has yet to be fully explained. To address this issue, a targeted, candidate gene scan was performed to reveal genetic variation associated with tolerance in four wild DLC-adapted killifish populations.

Considerable effort has been spent attempting to identify the genetic and biochemical mechanisms underlying inter- and intra-specific variation in DLC sensitivity in vertebrates. Multiple lines of evidence support the crucial role of the aryl hydrocarbon receptor (AHR) pathway in DLC toxicity in mammals. Polymorphisms in the ligand binding domain of the AHR among mouse strains result in differences in ligand binding affinity, and low binding affinity appears to protect against all toxic responses to DLCs. In rats, DLC tolerance is associated with variation in the transactivation domain of the AHR, yet functional consequences of the variation are less predictable [3]. A candidate gene approach effectively identified two amino acid substitutions in the AHR among avian wildlife and a consistent relationship between the amino acid residues present and DLC sensitivity was observed at the species level [4,5]. In fish, the striking difference in DLC sensitivity among wild Atlantic tomcod populations has been attributed to a six base pair deletion in the AHR2 gene that results in a five-fold decrease in ligand binding affinity and reduced ability to promote expression of detoxification enzymes targeted by the AHR pathway [6].

Within the killifish AHR pathway, several nonsynonymous single nucleotide polymorphisms (SNPs) have been identified in two AH receptor genes (AHR1 and AHR2), but patterns of genetic variation at these loci do not unequivocally reflect differences in DLC sensitivity among populations and no functional consequences (i.e., ligand binding and the ability to interact with xenobiotic metabolizing enzymes) were associated with AHR1 variants [7-9]. As an alternative to the candidate gene approach, Williams and Oleksiak [10,11] performed whole genome scans, whereby patterns of variation across hundreds of genetic loci were contrasted between killifish populations from polluted and reference sites in order to identify genes under selection with respect to DLC contamination. A handful of selectively important genetic markers were identified in each of three separate comparisons between populations residing in polluted habitats and their respective reference populations, and a single marker (in the CYP1A promoter) was found to be selectively important in all comparisons.

Both the single candidate gene approach and genomewide scans (typically with sets of anonymous genetic markers) have led to great success in elucidating the genetic basis for many adaptive phenotypes [12,13], but neither has offered a comprehensive link to the observed variation in DLC sensitivity among wild killifish populations. A 'candidate gene scan' approach, which targets a relatively large set of expressed genes with known physiological function, should increase the probability of isolating genes that are under selection and relevant with respect to traits of interest [14]. Thus, this approach was adopted to maximize efficiency in the identification of genes associated with DLC tolerance, and complement previous and ongoing work investigating the mechanism(s) involved in the repeated adaptation to DLCs in wild killifish populations.

In this study, patterns of genetic variation at SNP markers distributed across genes that are components of, or whose expression is affected by the AHR and interacting pathways were examined among eight killifish populations that vary genetically in their responsiveness to DLCs. Four relatively uncontaminated populations whose sensitivities to a prototypical DLC (3,3',4,4',5- pentachlorobiphenyl, PCB126) range from 20–38 ng/L (reviewed in [12]) were chosen as sensitive populations. Each of these populations is located near one of four EPA-designated urban/industrial estuarine Superfund sites. Killifish resident to these sites are dramatically tolerant to the effects of PCB126, displaying LC20 values ranging from ~400 to ~8000 times higher than those for the sensitive killifish (Table 1). A companion study [9] provides a fine-scaled examination of genetic variation in three AHR-related loci (AHR1, AHR2, and AHRR) among killifish populations residing in uncontaminated and polluted habitats of the North Atlantic coast of the US, including some of the same populations examined in this study.

The question of whether the genetic variants observed at the targeted SNP markers in this study could explain the stark phenotypic differences between DLC-adapted (–tolerant) and -sensitive killifish populations was addressed in several ways. Measures of genetic diversity were compared between DLC-sensitive and DLC-tolerant groups and patterns of genetic differentiation among populations were tested against expectations of isolation by distance. In addition, by contrasting the behavior of individual loci among DLC-sensitive and DLC-tolerant populations, specific loci under selection were identified.

Table 1 Killifish population locations and site characteristics

Site ID	Site location	Latitude × Longitude (W)	Distance from NBH, latitude, km	Latitudinal distance between pairs, km	LC20 (log), ng PCB126/L	DLC responsiveness
NBH	New Bedford, MA	41.6676 × 70.9159	0	56	4.6	Tolerant
BI	Block Island, RI	41.1818 × 71.5793	56		1.4	Sensitive
BP	Bridgeport, CT	41.1570 × 73.2189	63	22	4.1	Tolerant
FLAX	Flax Pond, NY	40.9637 × 73.1342	85		1.6	Sensitive
NWK	Newark, NJ	40.7006 × 74.1223	115	26	4.1	Tolerant
SH	Sandy Hook, NJ	40.4687 × 74.0113	141		1.3	Sensitive
ER	Portsmouth, VA	36.8078 × 76.2945	547	55	5.3	Tolerant
KC	Gloucester, VA	37.3016 × 76.4226	492		1.4	Sensitive

Taken from Nacci et al. 2010. LC20 values represent the concentration of PCB126 resulting in 20% lethality of killifish embryos.

Methods

SNP marker development and preliminary screening

An extensive literature search was conducted to identify genes and biochemical pathways with demonstrated and potential involvement in the toxic responses to DLCs. A list of over 150 genes was compiled, which included components of the AHR pathway, nuclear receptors known to 'cross talk' with the AHR pathway (i.e., estrogen receptors, retinoic acid receptors, hypoxia inducible factors), cytochrome p450s, genes involved in cardiac development, cathepsins, and genes having oxidoreductase activity (e.g. [15], and references therein). The gene list was filtered at several stages of the marker development process. Sequence information for many, but not all, of the genes listed in Additional file 1 was retrieved from the *F. heteroclitus* unigene database in GenBank (http://www.ncib.nlm.nih.go/GenBank/). Additional unpublished sequences were kindly provided by Sibel Karchner (personal communication). Putative SNPs were detected with the QualitySNP pipeline [16]. In QualitySNP, sequences with > 95% similarity were assembled into contiguous sequences (contigs) using the sequence assembly program CAP3. Available contigs, those containing ≥ 4 overlapping sequences, were then evaluated for polymorphisms. Available contigs with SNPs were subjected to further scrutiny to assess the reliability of the putative SNPs identified. Specifically, SNPs were considered 'true' if they were represented in ≥ 2 ESTs, paralogous sequences within the cluster could be distinguished and haplotypes identified, and they were located in high quality sequence regions [16]. PCR primers and melting temperature (T_m)-shift genotyping assays [17] were then designed for the 'true' SNPs with suitable flanking sequence (e.g. no SNPs, low sequence complimentarity in priming sites). A test panel of eight killifish from two populations (BI and NBH) was genotyped at 128 of the loci deemed highly reliable by QualitySNP. A SNP maker was considered valid if amplification product of the appropriate size was generated and polymorp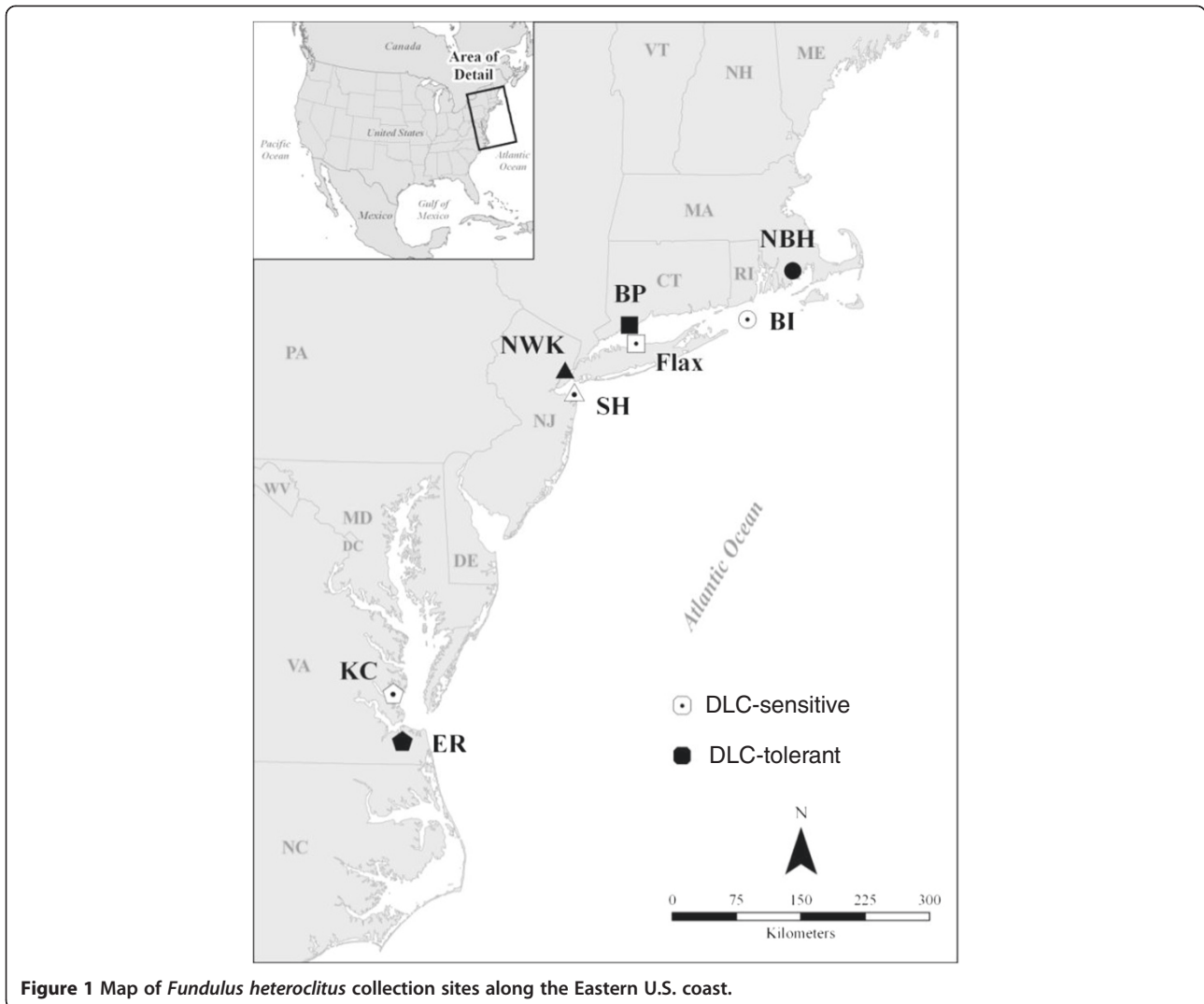hism was observed among the eight individuals in the test panel. Outcomes for each stage of the SNP marker development process are detailed in Additional file 1.

Fish collection

During the summer and fall months between 2008 and 2011, adult *Fundulus heteroclitus* were collected using baited minnow traps from eight estuarine sites spanning approximately 600 km of the Atlantic Coast of the USA (Figure 1). These specific killifish populations had already been characterized as either DLC-tolerant or DLC-sensitive, based on early life stage sensitivity to PCB126 [2,18,19]. To better assess genetic differences between tolerant and sensitive populations absent geographic influences, each DLC-tolerant population was paired with a nearby DLC-sensitive population. These pairs are located within the same or adjacent states but are separated by discontinuous shoreline and deep channels that limit fish migration and contaminant spread. Latitudinal distances between killifish residence sites (Table 1) were used to provide a consistent proxy to assess regional influences but were not intended to convey 'as-the-fish-swims' distances. Sixty to 100 live or sacrificed/frozen fish from each population were transported to the US Environmental Protection Agency (EPA) laboratory in Narragansett, RI, USA. Those that were transported live were sacrificed immediately upon return to the laboratory. Whole fish were stored at either –20 or –80°C prior to DNA extraction.

Sample preparation and population genotyping

Genomic DNA was extracted from approximately 20 mg of caudal fin tissue from 36 or 37 of the 60–100 archived individuals per population according to the QIAGEN DNeasy protocol for animal tissue (optional RNase treatment included), quantified with the PicoGreen dsDNA assay (Invitrogen), and diluted to a standard concentration of 20 ng/μl. Diluted DNA extracts were submitted to the University of Minnesota's BioMedical Genomics Center in a 96-well format for sample quality assessment and SNP genotyping using the Sequenom MassARRAY®

Figure 1 Map of *Fundulus heteroclitus* collection sites along the Eastern U.S. coast.

technology. Three multiplex assays (containing 32, 27, and 12 SNPs respectively) were designed using MassARRAY® Designer software. Sampled *F. heteroclitus* for which DNA was extracted were genotyped with the first two plexes (because all SNP containing genes were represented) following the iPLEX assay protocol. Genotypes for each individual at each locus were called using the Sequenom System Typer Analysis package.

Data analysis

Routine population genetics analyses were conducted using the freely available software package Arlequin ver. 3.5 [20]. Standard diversity indices, including the percentage of polymorphic loci (P_O), average observed and expected heterozygosities (H_O and H_E), and the within population fixation index (F_{IS}), were determined for each sampled population. Indices were based on data from loci with < 10% missing data. Loci were also tested for departures from Hardy-Weinberg equilibrium (HWE)

with 100,000 permutations and the percentage of loci in HWE was calculated for each population. The ability of this suite of SNP markers to detect genetic differentiation among populations was assessed by computing pairwise multi-locus F_{ST} estimates. To test whether the assumption of independence among loci would be violated by including multiple SNPs per gene in the analysis, data from a subset of SNPs (representing one SNP per gene) were also analyzed. Results did not differ significantly between the complete and limited datasets; therefore, only results from the complete dataset are reported here.

Genetic structure, where populations were assigned to groups defined by DLC sensitivity, was also included in the analysis and the pairwise multi-locus F_{ST} between the two groups was estimated with the 'Compute pairwise F_{ST}' option in Arlequin ver. 3.5. An analysis of molecular variance (AMOVA) was performed to better understand how the observed genetic variance was partitioned within populations, among populations within each group, and

between groups. In addition, as in [9], Student's T-tests were used to detect significant differences in genetic diversity measures (P_O and H_O) among the two groups.

For species like *F. heteroclitus* that are non-migratory, genetic differences at neutral loci should accumulate over time and generate a pattern of isolation by distance (IBD), whereby differentiation among populations increases with geographic distance [21]. Deviations from IBD patterns can be attributed to responses to local selection pressures [22]. Mantel tests and reduced major axis (RMA) regression analyses were performed in Isolation By Distance Web Service (IBDWS) [23] with 10,000 permutations to test for significant correlations between genetic distance (pairwise multi-locus F_{ST} values) and latitudinal distance. Mantel tests were also used to test for significant correlations between pairwise genetic differences and relative differences in sensitivity to PCB126 (LC20 as reported in [2]); moreover, partial Mantel tests were conducted to determine if there was a significant relationship between sensitivity to PCB126 and genetic divergence after taking latitudinal distance into account.

Although most of the genes included in this analysis were chosen based on prior work suggesting they might be responsive to DLCs, whether any or all contribute to the adaptive phenotype remains largely unresolved. To identify selectively important SNPs among the genes surveyed, F_{ST} values were calculated at each locus separately for each sensitive/tolerant comparison with the AMOVA function in Arlequin ver. 3.5. The statistical significance of each F_{ST} was determined through permutation testing with 10,000 iterations. An F_{ST} modeling approach similar to that described in [24] as implemented in Arlequin ver. 3.5 was also used to detect outliers. F_{ST} distributions conforming to a neutral model were simulated with 20,000 iterations, heterozygosities computed from empirical data, and assuming 10 demes. SNPs with F_{ST} values exceeding the 95th percentile of the null distribution were considered to be strong candidates for natural selection. In addition, as in [12], minor allele frequencies were calculated at each locus for each population with the expectation that loci associated with the adaptive phenotype would display shifts in allele frequency for each sensitive/tolerant pair. Consistent shifts in the same direction across all four comparisons were considered further evidence for selection acting at the loci.

Results
SNP markers
Killifish sequences representing 125 of the 150 candidate genes were mined from the GenBank nucleotide and unigene databases. CAP3 assembled those sequences into 183 contigs, of which 105 were available for further analysis (met the criterion of containing four or more overlapping sequences). Approximately 500 highly reliable putative

SNP loci distributed among 50 genes were identified with the QualitySNP pipeline. T_m-shift genotyping assays were designed for 128 of the highly reliable putative SNPs and the validity of each locus was tested by genotyping eight killifish collected from two populations (BI and NBH) (Additional file 1). Twenty-four of the 128 putative SNPs assayed (19%) did not amplify or produced uninterpretable melting curves. An additional 29 of the SNPs tested (23%) appeared to be monomorphic. Ultimately, 75 (59%) of the putative SNPs screened were polymorphic and 59 (those represented in the first two iplex assays) were used in the population genetic analysis (Table 2).

Standard diversity indices
Substantial genetic variation was observed in each killifish population examined. The percentage of polymorphic loci (P_O) ranged from 51% to 73% across populations, with a greater percentage of monomorphic loci in NBH and the Virginia populations. A similar pattern was detected when observed heterozygosity (H_O) was used to measure diversity; ER, KC and NBH populations were the least heterozygous. According to the average within population fixation index (F_{IS}), the loss of heterozygosity in NBH, ER, KC, and SH was statistically significant (Table 3).

A majority of the 59 SNP loci assayed (64%) were considered to be common SNPs, with minor allele frequencies greater than 0.1, when data from all populations were pooled. For each population, the percentage of common SNPs ranged from 54% to 76% with BP and ER at the two extremes. Within each population the minor allele frequency, averaged across all polymorphic loci, varied between 0.16 and 0.22. The mean minor allele frequency was lowest in BP and BI fish and highest in ER fish (Table 3).

Most of the SNP loci in this study were found to be in Hardy-Weinberg equilibrium within each population. Again, the population with the fewest loci conforming to Hardy-Weinberg expectations was ER (Table 3). In total, 14 loci deviated from HWE, but no single deviation was consistent across all eight populations or across the four tolerant populations.

Genetic differentiation
Mean pairwise F_{ST} values for all population comparisons were found to be statistically significant and suggest moderate to very high levels of genetic differentiation. Overall, markedly higher genetic differentiation was detected between the two Virginia populations and all others (Figure 2). A standard analysis of molecular variance (AMOVA) confirmed that a substantial proportion (23%) of the observed molecular variation can be attributed to differences among populations.

No significant differentiation emerged when populations were grouped by DLC sensitivity. The expectation that, on

Table 2 List of SNP containing candidate genes included in the population genetic analysis

Gene name	Unigene ID	Putative function	Reference
Aryl hydrocarbon receptor 1	1743961[uid]	transcription factor, response to xenobiotic stimulus	[7]
Aryl hydrocarbon receptor 2	1743966[uid]	transcription factor, response to xenobiotic stimulus	[8]
Aryl hydrocarbon receptor 2b	N/A	transcription factor, response to xenobiotic stimulus	Karchner & Hahn, unpublished;
Aryl hydrocarbon receptor repressor	1743962[uid]	signal transducer, response to xenobiotic stimulus	[25]
Atrial natriuretic peptide	1743159[uid]	receptor binding, cardiac muscle hypertrophy in response to stress	[26]
Aldehyde dehydrogenase family 9 member A1	1741825[uid]	oxidoreductase activity, retinoic acid metabolic process	[27]
Cardiac myosin light chain-1	1741781[uid]	calcium ion binding, cardiac muscle tissue development	
Cathepsin E precursor	1744072[uid]	endopeptidase, antigen processing	
Cathepsin F precursor	1741927[uid]	cysteine-type peptidase activity	
Cathepsin Z precursor	2476770[uid]	cysteine-type peptidase activity, angiogenesis	
Complement Component C3	1742328[uid]	protein binding, lipid binding, endopeptidase inhibitor, innate immune response, regulation of angiogenesis	[28]
Cytochrome p450 1A	2476796[uid]	oxidoreductase activity, dibenzo-p-dioxin catabolic process	[11,29-31]
Cytochrome p450 3A30	1742582[uid]	oxidoreductase activity, response to xenobiotic stimulus	[28]
Cytochrome B5	1743444[uid]	enzyme binding, heme binding, oxidation reduction process	[29-31]
Estrogen Receptor alpha	1743972[uid]	estrogen receptor activity, transcription factor binding, response to estradiol stimulus, regulation of retinoic acid receptor signalling pathway	[32]
Estrogen Receptor beta a	2301213[uid]	estrogen receptor activity, estrogen response element binding, response to estradiol stimulus, regulation of transcription factor activity	[32]
Glyceraldehyde 3 phosphate dehydrogenase	1742899[uid]	oxidation reduction process	
Hepcidin	1744101[uid]	hormone activity, defense response	
Hepcidin 2	N/A		Karchner & Hahn, unpublished
Heatshock protein 90 beta	1743580[uid]	protein binding, response to stress	
Plasma Kallikrein precursor	1742469[uid]	peptidase activity, proteolysis, hemostasis	
Kallikrein	N/A		Karchner & Hahn, unpublished
Myosin Light chain 2	1743984[uid]	actin monomer binding, heart development	[29]
Myosin Light chain alkali, smooth muscle	1743182[uid]	structural constituent of muscle, muscle contraction	
NADH [ubiquinone] dehydrogenase 1 alpha subcomplex subuint 4	1743852[uid]	electron transport chain	[28]
Myosin regulatory light chain 3	1741657[uid]	calcium ion binding	
NADH dehydrogenase subunit 1	3014035[uid]	oxidoreductase activity	
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10	1743800[uid]	electron transport chain	[33]
NADH dehydrogenase subunit 2	1742301[uid]	oxidoreductase activity	
NADH dehydrogenase subunit 3	1742764[uid]	oxidoreductase activity	
NADH dehydrogenase subunit 4	1741822[uid]	oxidoreductase activity	
NADH dehydrogenase subunit 6	3014152[uid]	oxidoreductase activity	
Platelet Coagulation Factor XI	3014242[uid]	peptidase activity, proteolysis, hemostasis	[28]
Retinoic Acid Receptor Responder Protein 1	1743266[uid]	negative regulation of cell proliferation	
Retinoic Acid Receptor Responder Protein 3	1741754[uid]	negative regulation of cell proliferation	

Table 2 List of SNP containing candidate genes included in the population genetic analysis (Continued)

Rho-class Glutathione S-Transferase	1743333[uid]	transferase activity	[30,31,34]
Serine Protease Inhibitor	N/A	negative regulation of endopeptidase activity	Karchner & Hahn, unpublished
TBT Binding Protein	1741555[uid]	response to xenobiotic stimulus	[28]
Thioredoxin	1743303[uid]	protein binding, oxidoreductase activity	[33]
Translation Initiation Factor 2	1744423[uid]	ligand-dependent nuclear receptor binding, estrogen receptor binding, retinoic acid receptor binding, signal transduction	
Telethonin (Titin cap protein)	1743784[uid]	heart development, cardiac muscle contraction, sarcomere organization, response to stress	[29]
Cardiac Troponin T2	1742146[uid]	actin binding, tropomyosin binding, troponin binding, cardiac muscle morphogenesis, sarcomere organization, muscle filament sliding	[26,29,35]
Ubiquitin A-52 residue ribosomal protein fusion product 1	1744070[uid]	structural constituent of ribosome, translation	

Note: Potential involvement of each gene in toxic DLC response was demonstrated in citations listed in 'Reference' column.
N/A = not available.

average, DLC-adapted populations are less diverse (as quantified by P_O and H_O) than DLC-sensitive populations, was not supported (Table 4). The genetic differentiation between the two groups as measured by F_{ST} was 0.0212; however, the hierarchical AMOVA results suggest that the groups were not significantly different (Table 5).

Isolation by distance

A regression of the genetic and latitudinal distance matrices when all populations were included in the analysis was statistically significant (r =0.8607, p < 0.001) reflecting a pattern of IBD (Figure 3A). Given that populations were sampled from three distinct geographic regions differentially impacted by the Pleistocene glacial retreat, the observed relationship is most likely driven by long-term history and demography rather than

contemporary forces [22]. Not surprisingly, the pattern persisted when only sensitive populations were considered (r = 0.9223, p = 0.0372). A clear positive relationship was also apparent when only tolerant populations were included in the analysis (r = 0.9443), but the trend was not significant (p = 0.0856) (Figure 3C). The lack of a significant IBD pattern among DLC-adapted populations could be indicative of local selection (in response to chemical contamination) counteracting the effects of history, demography, migration, and drift, but is more likely a consequence of small sample size. It was hypothesized that if selection is a major force shaping patterns of genetic variation among the sampled killifish populations, a strong correlation between pairwise genetic differences and relative differences in sensitivity to PCB126 should be evident. Additional Mantel tests found no such relationship (data not shown).
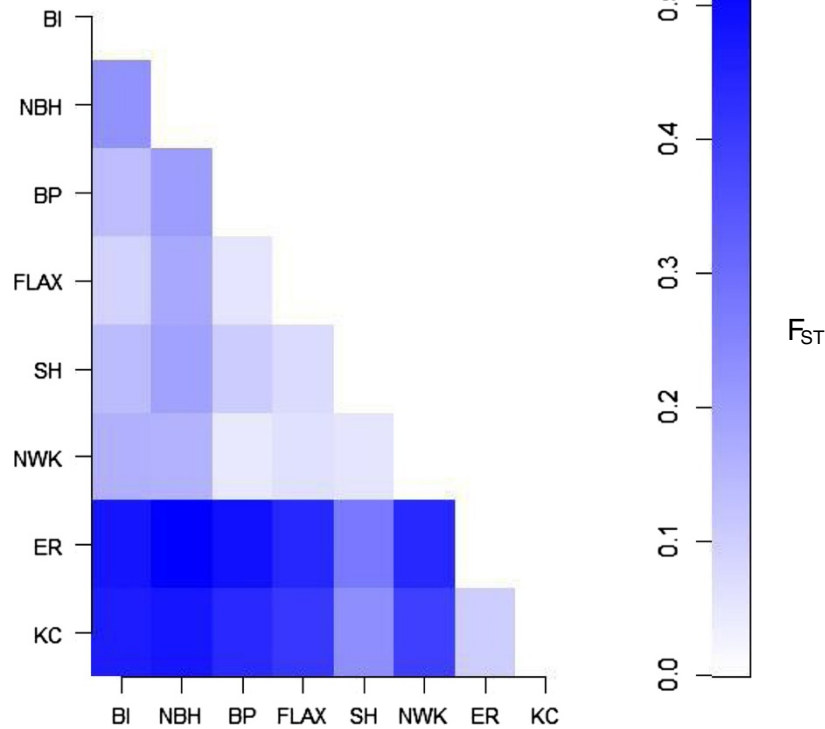
Loci under selection

The number of loci with significant F_{ST} values greater than 0.10, suggesting they are greatly differentiated among populations [36], varied among the four sensitive/tolerant population pairings. Of the four comparisons, the largest number of highly differentiated loci was detected between BI and NBH (Table 6). Included among these loci were the aryl hydrocarbon receptors (AHR) 1 and 2, cathepsin Z, the cytochrome P450s (CYP 1A and 3A30), and the NADH ubiquinone oxidoreductase MLRQ subunit. Only the CYP1A locus exhibited a consistently high F_{ST} value across all population pairs while both AHR genes were highly differentiated in all comparisons except for that between SH and NWK. While great differences were detected between paired populations at the CYP1A locus, a hierarchical AMOVA did not confirm that those differences are associated with DLC sensitivity (Table 5). In contrast, variation at the AHR2 locus could distinguish DLC-sensitive and tolerant groups (Table 5).

Table 3 Genetic parameters for sampled *F. heteroclitus* populations

Population	P_O	H_O	H_E	% HWE	F_{IS}	MAF
BI	65.31 (49)	0.15	0.15	71.88	0.042	0.16
NBH	53.19 (47)	0.14	0.16	76.00	0.171*	0.20
BP	64.58 (48)	0.15	0.15	93.55	0.001	0.16
FLAX	61.22 (49)	0.16	0.17	90.00	0.059	0.19
SH	73.33 (45)	0.16	0.21	78.79	0.240*	0.19
NWK	64.58 (48)	0.15	0.16	80.65	0.031	0.19
ER	51.02 (49)	0.12	0.15	60.00	0.108*	0.21
KC	57.14 (49)	0.14	0.17	82.14	0.122*	0.22
Mean	61.30	0.15	0.17	79.12	0.097	0.19
Standard Deviation	7.28	0.01	0.02	10.46	0.080	0.02

P_O = number of polymorphic loci/number of usable loci (in parentheses); H_O = sum of observed heterozygosity for each usable locus/number of usable loci; H_E = sum of expected heterozygosity for each usable locus/number of usable loci; % HWE = number of loci in Hardy-Weinberg equilibrium/number of polymorphic loci; F_{IS} = population specific fixation index; MAF = average Minor Allele Frequency, calculated for polymorphic loci only. *p ≤ 0.05 based on 10,000 permutations.

	BI	NBH	BP	FLAX	SH	NWK	ER	KC
BI	*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
NBH	0.22483	*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BP	0.1351	0.20125	*	0.0000	0.0000	0.0000	0.0000	0.0000
FLAX	0.08803	0.17894	0.05391	*	0.0000	0.0000	0.0000	0.0000
SH	0.13978	0.19243	0.10443	0.07516	*	0.0000	0.0000	0.0000
NWK	0.16307	0.15865	0.04881	0.0647	0.05414	*	0.0000	0.0000
ER	0.4817	0.52344	0.49028	0.44478	0.27278	0.44235	*	0.0000
KC	0.46434	0.47866	0.43884	0.41061	0.23224	0.39375	0.10278	*

Figure 2 Pairwise F_{ST} comparisons demonstrating genetic differences between all populations. F_{ST}s calculated with data from all SNP loci. Values below the diagonal represent genetic distances and values above the diagonal represent significance level. BI = Block Island, RI, NBH = New Bedford Harbor, MA, BP = Bridgeport, CT, FLAX = Flax Pond, NY, SH = Sandy Hook, NJ, NWK = Newark, NJ, ER = Elizabeth River, VA, KC = Kings Creek, VA.

Table 4 T-test results

Diversity measure	DLC-sensitive	DLC-tolerant	P-value
P_O	0.64 (0.07)	0.58 (0.07)	0.28
H_O	0.15 (0.01)	0.14 (0.01)	0.23
H_E	0.18 (0.03)	0.16 (0.01)	0.18
MAF	0.15 (0.02)	0.15 (0.01)	1.0

Comparison of mean genetic diversity measures between DLC-sensitive and DLC-tolerant *F. heteroclitus* populations. Standard deviations are in parentheses. P_O = average percentage of polymorphic loci within groups; H_O = average observed heterozygosity within groups; H_E = average expected heterozygosity within groups; MAF = average minor allele frequency within groups. Results are based on genotype information gathered from 36/37 individuals per population at 59 SNP loci representing 42 genes.

Further exploration of associations between specific loci and DLC sensitivity using an F_{ST} modeling approach produced results similar to those derived from the locus-by-locus AMOVA. Again, the CYP1A locus was identified as a significant outlier with respect to the simulated F_{ST} null distribution in three out of the four pairwise comparisons between each sensitive population and its DLC-tolerant counterpart. Moreover, when data for all sensitive populations were pooled and compared to the pooled data for all tolerant populations, the AHR2 locus emerged as a significant outlier (Figure 4). These findings lend support to the hypothesis that the CYP1A and AHR2 loci may be involved in the evolution of DLC tolerance.

Evidence of selection can also be gleaned from subtle shifts in allele frequencies in response to environmental

Table 5 AMOVA results (populations grouped according to DLC sensitivity); data from all loci, CYP1A, and AHR2

Source of variation	df	Sum of squares	Variance components	Percent variation
All Loci				
Among Groups	1	31.573	−0.19211 (V_a)	−4.38
Among Pops w/in Groups	6	528.228	1.15127 (V_b)	26.27
W/in Pops	580	1985.789	3.42377 (V_c)	78.12
Total	587	2545.590	4.38293	
CYP1A				
Among Groups	1	2.726	−0.01613 (V_a)	−6.52
Among Pops w/in Groups	6	44.501	0.09971 (V_b)	40.32
W/in Pops	574	93.977	0.16372 (V_c)	66.20
Total	581	141.204	0.24730	
AHR2				
Among Groups	1	12.822	0.03671 (V_a)	13.50
Among Pops w/in Groups	6	13.271	0.02775 (V_b)	10.20
W/in Pops	570	118.296	0.20754 (V_c)	76.30
Total	577	144.389	0.27200	

variables after controlling for population structure [37]. By reviewing differences in minor allele frequency between sensitive and DLC-adapted population pairs, a strong signal was apparent for two SNPs: AHR2_1929 and CYP1A_2140 (Table 7). Upon further examination, the shift in minor allele frequency for the AHR2 SNP varied in magnitude but was consistently in the same direction for all four population comparisons, providing yet another line of evidence linking AHR2 with the DLC-tolerant phenotype. This was not the case at the CYP1A locus. Although substantial differences in minor allele frequency were observed between each sensitive and tolerant pair, the frequency shift between KC and ER was in the direction opposite of the other three pairings (Figure 5).
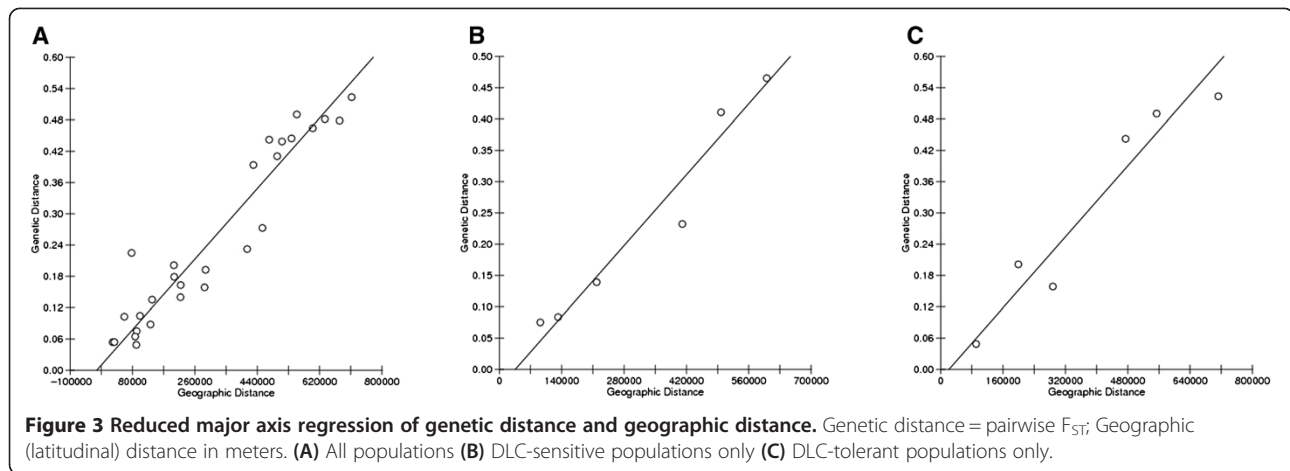
Discussion

Four independent DLC-adapted killifish populations were contrasted with neighboring sensitive counterparts in an attempt to reveal genetic loci associated with intra-specific DLC tolerance in the wild. The AHR pathway is known to mediate toxic responses to DLCs in all vertebrates and several studies involving killifish have implicated altered AHR pathway function in eliciting the tolerant phenotype [29,30,37]. Therefore, a 'candidate gene scan' approach, focused on components and targets of the AHR and interacting pathways represented among available killifish genetic resources, was applied to gain a better understanding of the genetic basis for the observed phenotypic variation in DLC sensitivity among killifish populations.

Genetic diversity and population structure of tolerant killifish populations were examined with the expectation that patterns of historical stress would be revealed. Levels of genetic diversity and population structure in killifish have been evaluated in several studies; some with the specific aim of addressing the impacts of pollutant-driven selection [9,22,38-41]. Results of the multi-locus analyses conducted here are in agreement with previous reports. Comparably high levels of genetic variation were estimated for all populations, rendering the hypothesis that a genetic bottleneck facilitated the emergence of the DLC-adaptive phenotype highly unlikely. With respect to population structure, each population included in this study was moderately to highly different from all others and observed F_{ST} values were comparably higher than earlier measures for killifish populations spanning a similar geographic range (e.g. F_{ST} ranged from 0.03 - 0.2 in [38] and from 0.01 - 0.24 in [41]). The difference in the extent of population structure detected among independent studies could be a consequence of the type of markers used (targeted SNPs vs. putatively neutral microsatellites); however, the overall patterns are consistent: genetic differences appear to be driven by geographic distance (latitudinal or shoreline) rather than DLC sensitivity. Moreover, a hierarchical AMOVA did not attribute any of the existing molecular variation to differences between DLC-sensitive and DLC-tolerant groups.

The failure to detect a significant relationship between multi-locus measures of genetic diversity/differentiation and increased tolerance to DLCs may be because a majority of the markers used in the current and previous analyses are selectively neutral. It has been suggested that neutral markers best reflect the effects of anthropogenically-mediated environmental change when populations are in decline and genetic exchange among populations is restricted [42]. There is no evidence that either sensitive or DLC-tolerant killifish populations have experienced a reduction in population size [22,38,39]. A more plausible explanation for the rapid adaptation to DLC contamination is that the trait in question is controlled by a small number of loci [39,43]. This explanation is consistent with theoretical models that predict single allelic differences of large effect dominate adaptive shifts when environmental change is sudden and selection is intense [42]. The prediction has been tested and verified in insect populations reacting to pesticides [44], benthic marine invertebrates subject to heavy metal toxicity [45], and fish exposed to DLCs [6].

Phenotypic similarities among tolerant populations exposed to a prototypical DLC (e.g. LC20) [2] supported an expectation of common loci under selection; however, genetic loci associated with tolerance might vary across populations due to differences in the selective agents present in their native habitats. The presumed

Figure 3 Reduced major axis regression of genetic distance and geographic distance. Genetic distance = pairwise F_{ST}; Geographic (latitudinal) distance in meters. **(A)** All populations **(B)** DLC-sensitive populations only **(C)** DLC-tolerant populations only.

selective agents include large classes of aromatic hydrocarbons whose toxic effects are mediated fully (DLCs) or partly (polycyclic aromatic hydrocarbons, PAHs) through the AHR pathway. While urban contamination includes moderate levels of both PAHs and PCBs, toxic levels of DLCs have been measured at NBH and NWK, PAHs at ER, and both at BP [2,46]. Therefore, genetic analyses were conducted to identify common loci under selection in tolerant populations, and differences between sensitive/tolerant paired killifish populations.

Since functional variation in AHR-ligand binding initiates the AHR pathway cascade (and largely explains some intra- and inter-species differences in DLC sensitivity, e.g. [15] and references therein), evidence for selection acting on AHR loci was of obvious interest in comparing tolerant versus sensitive killifish populations. As in other fish species, killifish possess at least two distinct AH receptor genes, AHR1 and AHR2, and the expression of AHR2 predominates in most tissues [47]. Strong signals of selection were detected for an AHR1 (AHR1_1530) and AHR2 SNP (AHR2_1929) included in this analysis in three of the four pairwise comparisons (SH/NWK excluded) when locus-by-locus F_{ST} were considered. Although these two SNPs are synonymous and do not result in amino acid changes, they are both located in the transactivation domain of their respective AHR genes and in close proximity to nonsynonymous SNPs found to be under selection in Reitzel et al. [9]. Minor allele frequencies (equivalent to base frequencies referred to in [9]) of the AHR2_1929 SNP were consistently higher in the DLC-tolerant populations for all four population pairs. Although the F_{ST} modeling approach identified the AHR1_1530 SNP as a significant outlier in only one of the comparisons (BI/NBH), when genotype data from all DLC-sensitive populations were pooled and compared to that of all DLC-tolerant populations AHR2_1929 was the only locus found to deviate from neutral expectations. Similar patterns of genetic variation in AHR1 and AHR2 loci is not surprising given that these

two genes are arranged in tandem (and therefore linked) within the killifish genome [9,48]. To determine whether variation in both loci, a single locus, or neither locus underlies the DLC-tolerant phenotype, population genetic data must be accompanied by functional assays. A focused examination of allelic variation in killifish AHR1 and AHR2 revealed SNPs under positive selection in both genes; however, AHR1 variants were not responsible for alterations in receptor function and AHR2 variations have yet to be tested [7,9]. Studies investigating the genetic basis for DLC-tolerance in zebrafish and Atlantic tomcod have isolated an AHR2 gene as a key player in mediating DLC sensitivity [6,48]; while AHR1 and AHR2 seem to play functional roles in dioxin toxicity in red seabream [35]. Taken together, these results suggest that AHR2 variation likely plays a strong role in DLC sensitivity and tolerance in killifish, but more complex interactions may be revealed as new AHR paralogs are being identified and characterized [9].

Consistent with the refractory induction patterns for CYP1A (an early and sensitive marker of AHR pathway activation) observed among DLC-tolerant killifish populations in laboratory studies [2,29,49], the CYP1A SNP emerged as a strong candidate for selection in all tolerant populations. High F_{ST} values and large differences in minor allele frequencies were observed at the CYP1A locus in all four population comparisons and the F_{ST} modeling approach identified CYP1A as a significant outlier in three out of the four comparisons (BI/NBH excluded). However, the shift in allele frequency, although significant, was not in the same direction across all comparisons. A full genome scan analysis of three of the same tolerant killifish populations (excluding BP) and their sensitive counterparts also identified a SNP marker in the CYP1A promoter region as the only locus (out of 354 screened) under selection in all three DLC-tolerant populations surveyed, but again, the direction of the allele frequency shift between DLC-sensitive

Table 6 Single locus F$_{ST}$ values demonstrating high differentiation among population pairs

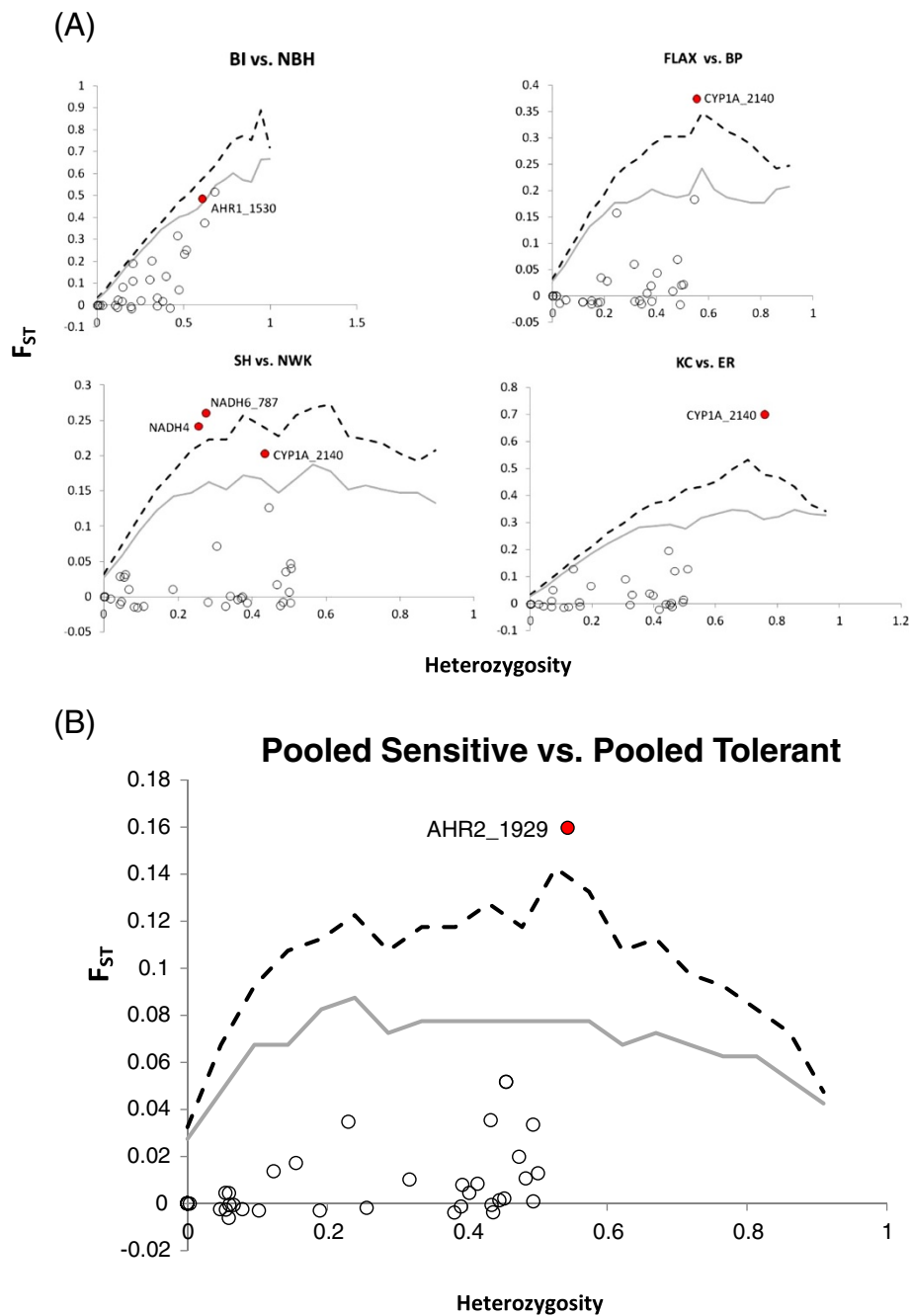|	Population comparison			
Locus	BI/NBH	FLAX/BP	SH/NWK	KC/ER
Aryl hydrocarbon receptor 1_1530	**0.48393**	**0.15677**	−0.00349	**0.12191**
Aryl hydrocarbon receptor 1_161	**0.26498**	0.00028	0.00929	N/A
Aryl hydrocarbon receptor 1_2289	**0.21749**	N/A	−0.00376	N/A
Aryl hydrocarbon receptor 1_948	**0.35061**	−0.0137	−0.01388	N/A
Aryl hydrocarbon receptor 2_1929	**0.25477**	**0.1825**	0.04692	**0.19729**
Aryl hydrocarbon receptor 2_792	**0.36916**	N/A	0.00417	N/A
Cathepsin F_653	**0.18941**	−0.0137	−0.01271	0.03401
Cathepsin Z_624	**0.22978**	0.04241	−0.00759	−0.00325
Cytochrome P450 1A_2140	**0.37453**	**0.3733**	**0.20304**	**0.70083**
Cytochrome P450 3A_1166	**0.51614**	−0.01332	−0.00179	−0.01449
Hepcidin2_399	**0.14131**	0.06457	0.00719	0.04197
Heat shock protein 90_775	0.09438	N/A	0.02932	**0.12676**
NADH ubiquinone oxidoreductase MLRQ subunit_363	**0.32115**	−0.01268	**0.12679**	−0.01516
Myosin regulatory light chain 3_372	**0.11663**	−0.00843	−0.01529	0.0674
NADH dehydrogenase subunit 4_347	0.0137	N/A	**0.24148**	N/A
NADH dehydrogenase subunit 4_586	0.01315	N/A	**0.13504**	0.01408
NADH dehydrogenase subunit 4_669	0.01259	N/A	**0.25714**	N/A
NADH dehydrogenase subunit 6_787	0.01486	N/A	**0.26028**	N/A
Serine protease inhibitor_938	0.00078	N/A	0.03176	**0.10229**
TBT binding protein_635	**0.10673**	−0.00786	−0.00228	N/A
Thioredoxin_582	**0.19931**	0.01625	−0.00785	0.1248

F$_{ST}$ values for highly differentiated loci across all four DLC-sensitive/-tolerant population pairs. F$_{ST}$s were calculated using the locus by locus AMOVA function in Arlequin v. 3.5. Values exceeding an F$_{ST}$ threshold of 0.1 are noted in bold and are significant at $p < 0.05$. Comparisons that were not made because the markers were monomorphic in some populations are noted as N/A. BI = Block Island, NBH = New Bedford Harbor, Flax = Flax Pond, BP = Bridgeport, SH = Sandy Hook, NWK = Newark, KC = Kings Creek, and ER = Elizabeth River.

and tolerant populations was not uniform [11]. In a follow-up study, Williams and Oleksiak [50] found that CYP1A promoter variants derived from a DLC-tolerant population (NBH) resulted in elevated expression of CYP1A *in vitro*, contradicting the well-documented refractory response of CYP1A to DLC exposure among tolerant populations *in vivo*. These previously reported results, coupled with the knowledge that the CYP1A SNP included in this analysis is located in the 3' untranslated region of the gene [51] make the exact functional role (if any) of CYP1A SNPs in the DLC-tolerant phenotype unclear. It may be that additional, yet to be discovered factors associated with CYP1A regulation are also important.

The interpretation of CYP1A as a candidate for selection must also take into account that tolerance has evolved in response to different AHR ligands. Specifically, the role of CYP1A in AHR-mediated toxicity varies by ligand class. Classically described, AHR agonists induce the production of enzymes (predominantly CYP1A) that metabolize PAHs, but not DLCs. Unlike DLC toxicity, PAH toxicity is self-limiting, due to AHR-enhanced PAH elimination, and

includes components that are not AHR-mediated. For example, CYP1A knockdown studies in zebrafish embryos have demonstrated the protective value of CYP1A during developmental PAH exposures [52]. That tolerant killifish populations from varied selective environments (PAH- versus DLC-dominated pollution profiles) show similar, poorly-inducible CYP1A phenotypes warrants the consideration of the adaptive value of the CYP1A recalcitrant phenotype more broadly, e.g., as an energy conservation strategy associated with chronic pollution exposures. As in other species [53,54], variation in the CYP1A sequence among killifish populations may be beneficial if associated with conditional fitness under transient, low level PAH exposures. Alternatively, variation in CYP1A may be related to its position 'downstream' in the AHR pathway, and secondary to changes in 'upstream' loci, causally associated with tolerance. Regardless of mechanism, variation at the CYP1A locus differentiates tolerant from sensitive killifish.

Among nearby sensitive/tolerant killifish population comparisons, the BI/NBH pairing was the most genetically differentiated based on multi-locus F$_{ST}$; many

Figure 4 Results of an F_{ST} modeling approach to detect loci under selection. Empirical F_{ST} values have been plotted against heterozygosity for each pairing of DLC-sensitive and DLC-tolerant sites **(A)** and for the pooled DLC-sensitive/-tolerant comparison **(B)**. Gray lines represent the 95th percentile and black, dotted lines represent the 99th percentile. Red data points indicate outlier SNPs with respect to the F_{ST} null distribution.

additional loci were identified as candidates for selection (i.e., $F_{ST} > 0.1$ and significant allele frequency shifts). In addition to the AHRs and CYP1A, these populations appear to be highly differentiated at cathepsin F, cathepsin Z, CYP3A30, and the NADH ubiquinone oxidoreductase MLRQ subunit loci. Cathepsins are a large group of proteolytic enzymes that have been implicated in cardiomyopathies

and cardiovascular disease, ultimately resulting in impaired pump function [55,56]. Given that the cardiovascular system is a main target of DLC toxicity in all vertebrates [48], it is reasonable to propose that alterations in the cathepsin coding sequence could contribute to existing differences in DLC sensitivity. CYP3A30 is an abundant xenobiotic metabolizing enzyme in killifish livers responsible for the breakdown and

Table 7 Minor allele frequencies of *F. heteroclitus* SNP markers calculated for each population

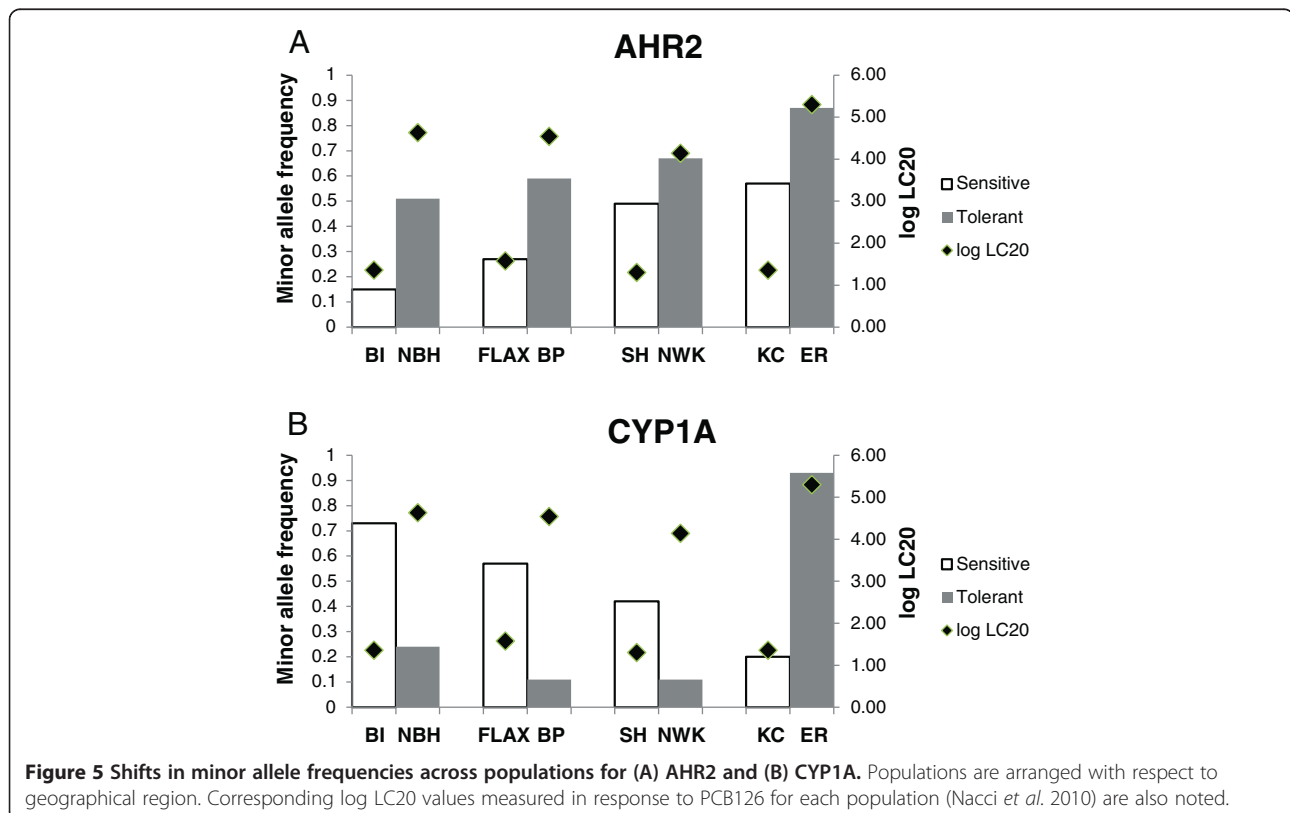| Locus	Minor allele	Population							
		BI	NBH	FLAX	BP	SH	NWK	KC	ER
Aglobin_89	A/**G**	0	0	0	0	0	0	0	0
AHR1_1530	C/**G**	0.08	0.62	0.23	0.03	0.26	0.2	0.44	0.19
AHR1_161	C/**G**	0.14	0.5	0.04	0.01	0.03	0.08	0	0
AHR1_2289	G/**T**	0.07	0.64	0	0	0.03	0.05	0	0
AHR1_948	C/**T**	0.07	0.49	0.01	0.01	0.03	0.03	0	0
AHR2_1929	**C**/T	0.15	0.51	0.27	0.59	0.49	0.67	0.57	0.87
AHR2_792	C/**T**	0.07	0.5	0	0	0.03	0.07	0	0
AHR2B_992	**A**/G	0	0	0.05	0.07	0.05	0.01	0	0
AHRR_1095	C/**T**	0	0	0.11	0.05	0	0	0	0.06
AHRR_1299	C/**T**	0	0	0.15	0.05	0	0	0.01	0.06
ANP_521	C/**T**	0.24	0.44	0.7	0.41	0.29	0.23	0.38	0.28
BAD_213	G/**T**	0.15	0.28	0.24	0.27	0.21	0.28	0.06	0.07
Bglobin_193	C/**T**	0	0	0	0.01	0.03	0.01	0.1	0.07
Cardiac_MLC1_250	C/**T**	0	0	0	0	0	0	0	0
CathE_730	A/**C**	0	0	0	0	0	0	0	0
CathF_653	A/**T**	0.2	0	0.09	0.09	0.21	0.2	0.14	0.26
CathZ_624	G/**T**	0.5	0.15	0.19	0.34	0.43	0.38	0.31	0.38
CC3_571	A/**G**	0	0	0	0	0	0	0	0
CYP1A_2140	**A**/**T**	0.73	0.24	0.57	0.11	0.42	0.11	0.2	0.93
CYP3A_1166	**A**/G	0.78	0.19	0.42	0.43	0.21	0.28	0.06	0.06
CytB5_626	**G**/T	0.01	0	0.01	0.01	0.06	0.05	0.07	0
ERa_1497	**G**/T	0	0	0	0	0	0	0	0
ERba_1373	**C**/T	0.19	0.25	0.51	0.38	0.68	0.47	0.04	0.03
GP3D_530	**A**/G	0	0	0	0.01	0	0	0	0
Hepcidin_69	C/**T**	0.11	0.08	0.07	0.16	0.01	0.03	0	0
Hepcidin2_399	A/**G**	0.12	0.37	0.26	0.11	0.39	0.49	0.68	0.82
HSP90_775	A/**G**	0.01	0.14	0	0	0.04	0	0.14	0
Kallikrein_502	G/**T**	0	0	0	0	0	0	0	0
KallikreinS_309	A/**G**	0	0	0	0	0	0	0	0
MLC2_436	**C**/T	0	0	0.07	0.05	0.14	0.07	0.19	0.33
MLC2_535	A/**C**	0	0	0.03	0.01	0.07	0	0.32	0.35
MLCA_577	**C**/T	0.09	0.03	0.08	0.08	0.26	0.09	0.03	0.01
MLRQ_363	A/**T**	0.46	0.07	0.11	0.09	0.42	0.17	0.7	0.7
MRCL3_372	**A**/G	0.27	0.14	0.22	0.18	0.05	0.04	0.17	0.04
NADH10_107	A/**T**	0.19	0.03	0.15	0.24	0.14	0.05	0.13	0.31
NADH10_439	**C**/T	0.31	0.28	0.23	0.19	0.25	0.18	0.51	0.64
NADH2_349	**A**/G	0	0	0.11	0.05	0	0.05	0.03	0
NADH3_165	C/**G**	0	0	0	0	0	0	0	0
NADH4_347	C/**T**	0.03	0	0	0	0.26	0	1	1
NADH4_586	A/**G**	0.03	0	0	0	0.27	0.05	0.03	0
NADH4_669	G/**T**	0.03	0	0	0	0.28	0	1	1
NADH6_787	**A**/G	0.08	0.05	0	0	0.28	0	1	1
PCFXI_511	**G**/T	0	0	0	0	0	0	0	0

Table 7 Minor allele frequencies of *F. heteroclitus* SNP markers calculated for each population *(Continued)*

RARR1_759	A/**C**	0.08	0.14	0.22	0.21	0.52	0.35	0.7	0.39
RGST_397	A/**T**	0.1	0.12	0.19	0.3	0.14	0.19	0.52	0.64
SPI_938	**A**/G	0.04	0.08	0	0	0.06	0	0.27	0.08
TBTbp_635	**A**/T	0.03	0.18	0.04	0.02	0.02	0	0	0
Thioredoxin_582	**A**/G	0.3	0.04	0.53	0.65	0.47	0.5	0.75	0.49
TIF2_727	**C**/T	0	0	0	0	0.04	0.04	0.11	0.06
Titin_478	**C**/G	0.19	0.1	0.46	0.26	0.42	0.3	0.66	0.64
TroponinT2_653	**C**/T	0.3	0.19	0.28	0.19	0.38	0.4	0.17	0.23

Zero values indicate populations where the SNP marker was monomorphic. Minor alleles in bold.

clearance of a wide array of anthropogenically derived pollutants [57] and has been identified as a target of the AHR pathway [58]. In killifish, expression of this gene was found to be significantly higher in field caught ER females relative to those collected from KC [31]; however, it was not differentially expressed in killifish embryos derived from the same DLC-tolerant and sensitive populations included in this study when exposed to PCB126 under controlled laboratory conditions [29,37]. Mutations in the NADH ubiquinone oxidoreductase MLRQ subunit are associated with metabolic diseases [28]. Again, expression of this gene was found to be significantly higher in field caught ER females relative to those collected from KC [31]. Moreover, expression of another component of the NADH ubiquinone oxidoreductase enzyme, NDUB2, was found to be higher in the brains of NBH, NWK, and ER adult fish [59]. The repeated association of NADH ubiquinone oxidoreductase with DLC-tolerance suggests that biochemical pathways other than the AHR could be involved in this chemically-induced stress response.

It is not known whether the identification of additional candidate loci for strong selection in the BI/NBH pair only suggests unique tolerance-related or biologically-relevant differences (BI is the only oceanic rather than estuarine site

Figure 5 Shifts in minor allele frequencies across populations for (A) AHR2 and (B) CYP1A. Populations are arranged with respect to geographical region. Corresponding log LC20 values measured in response to PCB126 for each population (Nacci *et al.* 2010) are also noted.

included in this analysis) or technical artifacts (statistical power). While similar phenotypes among the four tolerant populations examined suggest a conserved biochemical basis for intra-specific DLC tolerance in killifish, whether that similarity is constrained to identical nucleotide changes remains to be seen. The genetic mechanisms of adaptation could vary among DLC-tolerant populations. Alternatively, the differences detected may be a reflection of additional unique stressors encountered by each population pair.

Conclusions

The purpose of this study was to identify genetic polymorphisms associated with DLC sensitivity in Atlantic killifish from a suite of candidate loci involved in the AHR and interacting pathways; whether the polymorphisms in and of themselves are responsible for the drastic differences in DLC sensitivity among populations was beyond the scope of this work. Two loci, AHR2 and CYP1A, displayed patterns of variation consistent with selection in each of four pairwise comparisons. Additional loci with specific alleles significantly overrepresented in some, but not all, DLC tolerant populations underscore the possibility that the genetic variation in each population may have been shaped by similar yet unique selection pressures. Although the intention was to include all components of the AHR and interacting pathways in this population screen, the underrepresentation or absence of key genes (e.g., aryl hydrocarbon receptor nuclear translocator (ARNT)) in the *F. heteroclitus* unigene database resulted in an enriched but incomplete set of candidate genetic markers for DLC toxicity and (presumably) tolerance. Genetic resources being made available through the *Fundulus* genome consortium as well as a parallel, unbiased, Quantitative Trait Locus (QTL) approach to the discovery of the genetic mechanism of DLC tolerance in killifish will greatly increase our understanding of this dramatic example of anthropogenically induced, rapid adaptation in the wild.

Availability of supporting data

The sequences for the SNPs analyzed in this article are available through the Dryad Digital Depository and can be accessed by searching for the data package title *Data from: Targeted approach to identify genetic loci associated with evolved dioxin tolerance in Atlantic Killifish* (Fundulus heteroclitus), doi:10.5061/dryad.2355t Data files: Proestou_etal_Killifish_SNP_sequences.

Additional file

Additional file 1: List of all genes considered for SNP analysis and their fate during the marker development process. Sequences provided by Sibel Karchner were derived from either 26 (AHR1, AHR2, and AHRR) or 10 individuals (not cDNA libraries) collected from Scorton Creek, MA. In some cases, Qualitiy SNP identified multiple available

contigs from a single unigene entry. The number of SNPs genotyped does not equal the number of SNPs in the analysis because 8 of the markers failed to provide quality genotype information. N = No, Y = Yes, N/A = Not available, U = Unknown.

Author details
[1]US Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, 27 Tarzwell Drive, Narragansett, RI 02882, USA. [2]Department of Ecology and Evolutionary Biology, Brown University, 80 Waterman Street, Providence, RI 02912, USA. [3]Current address: USDA Agricultural Research Service, 469 Center for Biotechnology and Life Sciences, 120 Flagg Road, Kingston, RI 02881, USA.

References
1. Van Veld PA, Nacci D: **Chemical tolerance: Acclimation and adaptations to chemical stress.** In *The Toxicology of Fishes.* Edited by Diguilio RT, Hinton DE. Washington, DC: Taylor & Francis; 2008:597–644.
2. Nacci DE, Champlin D, Jayaraman S: **Adaptation of the Estuarine Fish *Fundulus heteroclitus* (Atlantic Killifish) to Polychlorinated Biphenyls (PCBs).** *Estuaries Coast* 2010, **33**:853–864.
3. Pohjanvirta R, Korkalainen M, Moffat ID, Boutros PC, Okey AB: **Role of the AHR and its structure in TCDD toxicity.** In *The AH Receptor in Biology and Toxicology.* 1st edition. Edited by Pohjanvirta R. New Jersey: John Wiley & Sons; 2012:181–196.
4. Head JA, Hahn ME, Kennedy SW: **Key amino acids in the aryl hydrocarbon receptor predict dioxin sensitivity in avian species.** *Environ Sci Technol* 2008, **42**:7535–7541.
5. Farmahin R, Manning GE, Crump D, Wu D, Lukas JM, Jones SP, Hahn ME, Karchner SI, Giesy JP, Bursian SJ, Zwiernik SJ, Frederick TB, Kennedy SW: **Amino acid sequence of the ligand-binding domain of the aryl hydrocarbon receptor 1 predicts sensitivity of wild birds to effects of dioxin-like compounds.** *Toxicol Sci* 2013, **131**:139–152.
6. Wirgin I, Roy NK, Loftus M, Chambers RC, Franks DG, Hahn ME: **Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson river.** *Science* 2011, **331**:1322–1325.

7. Hahn ME, Karchner SI, Franks DG, Merson RR: **Aryl hydrocarbon receptor polymorphisms and dioxin resistance in Atlantic killifish (***Fundulus heteroclitus***).** *Pharmacogenetics* 2004, **14**:1–13.

8. Hahn ME, Karchner SI, Franks DG, Evans BR, Nacci D, Champlin D, Cohen S: *Mechanism of PCB-and Dioxin-Resistance in Fish in the Hudson River Estuary: Role of Receptor Polymorphisms.* Final Report, Hudson River Foundation Grant. Final Report, *Hudson River Foundation Grant 004/02A* http://hudsonriver.org/ls/ 2005.

9. Reitzel AM, Karchner SI, Franks DG, Evans BR, Nacci D, Champlin D, Vieira VM, Hahn M: **Genetic diversity at aryl hydrocarbon receptor loci in PCB sensitive and PCB resistant populations of Atlantic killifish (Fundulus heteroclitus).** *BMC Evol Biol* 2014, **14**:6.

10. Williams LM, Oleksiak MF: **Signatures of selection in natural populations adapted to chronic pollution.** *BMC Evol Biol* 2008, **8**:282.

11. Williams LM, Oleksiak MF: **Ecologically and evolutionarily important SNPs identified in natural populations.** *Mol Biol Evol* 2011, **28**:1817–1826.

12. Rosenblum EB, Hoekstra HE, Nachman MW: **Adaptive reptile color variation and the evolution of the MclR gene.** *Evolution* 2004, **58**:1794–1808.

13. Jaquiery J, Stoeckal S, Nouhaud P, Mieuzet L, Mahéo F, Legeai F, Bernard N, Bonvoisin A, Vitalis R, Simon JC: **Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex.** *Mol Ecol* 2012, **21**:5251–5264.

14. Hemmer-Hansen J, Nielsen EE, Meldrup D, Mittelholzer C: **Identification of single nucleotide polymorphisms in candidate genes for growth and reproduction in a nonmodel organism; the Atlantic cod Gadus morhua.** *Mol Ecol Resour* 2011, **11**(Suppl. 1):71–80.

15. Denison MS, Soshilov AA, He G, DeGroot DE, Zhao B: **Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the Aryl Hydrocarbon (Dioxin) Receptor.** *Toxicol Sci* 2011, **124**:1–22.

16. Tang J, Vosman B, Voorrips RE, van der Linden CG, Leunissen JAM: **QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species.** *BMC Bioinforma* 2006, **7**:438.

17. Wang J, Chuang K, Ahluwalia M, Patel S, Umblas N, Mirel D, Higuchi R, Germer S: **High-throughput SNP genotyping by single-tube PCR with Tm-shift primers.** *Biotechniques* 2005, **39**:885–892.

18. Nacci D, Coiro L, Champlin D, Jayaraman S, McKinney R, Gleason TR, Munns WR Jr, Specker JL, Cooper KR: **Adaptations of wild populations of the estuarine fish Fundulus heteroclitus to persistent environmental contaminants.** *Mar Biol* 1999, **134**:9–17.

19. Nacci DE, Champlin D, Coiro L, McKinney R, Jayaraman S: **Predicting the occurrence of genetic adaptation to dioxin-like compounds in populations of the estuarine fish Fundulus heteroclitus.** *Environ Toxicol Chem* 2002, **21**:1525–1532.

20. Excoffier L, Lischer HEL: **Arlequin suite ver 3.5: a new series of programs to perform population genetics analysis under Linux and Windows.** *Mol Ecol Resour* 2010, **10**:564–567.

21. Wright S: **Isolation by distance.** *Genetics* 1943, **28**:114–138.

22. Duvernell DD, Lindmeier JB, Faust KE, Whitehead A: **Relative influences of historical and contemporary forces shaping the distribution of genetic variation in the Atlantic killifish, Fundulus heteroclitus.** *Mol Ecol* 2008, **17**:1344–1360.

23. Jensen JL, Bohonak AJ, Kelley ST: **Isolation by distance, web service.** *BMC Genet* 2005, **6**:13.

24. Beaumont MA, Nichols RA: **Evaluating loci for use in the genetic analysis of population structure.** *Proc R Soc Lond B Biol Sci* 1996, **263**:1619–1626.

25. Karchner SI, Franks DG, Powell WH, Hahn ME: **Regulatory interactions among three members of the vertebrate aryl hydrocarbon receptor family: AHR repressor, AHR1 and AHR2.** *J Biol Chem* 2002, **277**:6949–6959.

26. Carlson EA, Roy NK, Wirgin II: **Microarray analysis of polychlorinated biphenol mixture-induced changes in gene epression among Atlantic tom cod populations displaying differential sensitivity to halogenated aromatic hydrocarbons.** *Environ Toxicol* 2009, **28**:759–771.

27. Sartor MA, Schnekenburger M, Marlowe JL, Reichard JF, Wang Y, Fan Y, Ma C, Karyala S, Halbleib D, Liu X, Medvedovic M, Puga A: **Genomewide analysis of aryl hydrocarbon receptor binding targets reveals an extensive array of gene clusters that control morphogenic and developmental programs.** *Environ Health Perspect* 2009, **117**:1139–1146.

28. Meyer JN, Volz DC, Freedman JF, DiGiulio RT: **Differential display of hepatic mRNA from Fundulus heteroclitus inhabiting a superfund estuary.** *Aquat Toxicol* 2005, **73**:327–341.

29. Whitehead A, Triant DA, Champlin D, Nacci D: **Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population.** *Mol Ecol* 2010, **19**:5186–5203.

30. Whitehead A, Pilcher W, Champlin D, Nacci D: **Common mechanism underlies repeated evolution of extreme pollution tolerance.** *Proc R Soc Lond B Biol Sci* 2012, **279**:427–433.

31. Carney SA, Chen J, Burns CJ, Xiong KM, Peterson RE, Heideman W: **Aryl hydrocarbon receptor activation produces heart-specific transcriptional and toxic responses in developing zebrafish.** *Mol Pharmacol* 2006, **70**:549–561.

32. Greytak SR, Calllard GV: **Cloning of three estrogen receptors (ER) from killifish (Fundulus heteroclitus): differences in populations from polluted and reference environments.** *Gen Comp Endocrinol* 2007, **150**:174–188.

33. Fisher MA, Oleksiak MF: **Convergence and divergence in gene expression among natural populations exposed to pollution.** *BMC Genomics* 2007, **8**:108.

34. Paetzold SC, Ross NW, Richards RC, Jones M, Hellou J, Bard SM: **Up-regulation of hepatic ABCC2, ABCG2, CYP1A1, and GST in multixenobiotic-resistant killifish (fundulus heteroclitus) from the Sydney Tar ponds, nova Scotia, Canada.** *Mar Environ Res* 2009, **68**:37–47.

35. Waits ER, Nebert DW: **Genetic architecture of susceptibility to PCB-126 induced developmental cardiotoxicity in zebrafish.** *Toxicol Sci* 2011, **122**:466–475.

36. Wright S: *Evolution and the genetics of populations. Variability within and among natural populations. Vol. 4.* Chicago: University of Chicago Press; 1978.

37. Hancock AM, Alkorta-Aranburu G, Witonsky DB, Di Rienzo A: **Adaptations to new environments in humans: the role of subtle allele frequency shifts.** *Philos Trans R Soc Lond B Biol Sci* 2010, **365**:2459–2468.

38. Roark SA, Nacci D, Coiro L, Champlin D, Guttman SI: **Population genetic structure of a nonmigratory estuarine fish (Lindulus heteroclitus) across a strong gradient of polychlorinated biphenyl contamination.** *Environ Toxicol and Chem* 2005, **24**:717–725.

39. Adams SM, Lindmeier JB, Duvernell DD: **Microsatellite analysis of the phylogeography, Pleistocene history and secondary contact hypotheses for the killifish, fundulus heteroclitus.** *Mol Ecol* 2006, **15**:1109–1123.

40. McMillan AM, Bagley MJ, Jackson SA, Nacci DE: **Genetic diversity and structure of an estuarine fish (Fundulus heteroclitus) indigenous to sites associated with a highly contaminated urban harbor.** *Ecotoxicology* 2006, **15**:539–548.

41. Williams LM, Ma X, Boyko AR, Bustamante CD, Oleksiak MF: **SNP identification, verification and utility for population genetics in a non-model genus.** *BMC Genet* 2010, **11**:32.

42. Whitehead A, Crawford DL: **Neutral and adaptive variation in gene expression.** *Proc Natl Acad Sci U S A* 2006, **103**:5425–5430.

43. Oleksiak MF, Karchner SI, Jenny MJ, Franks DG, Welch DBM, Hahn ME: **Transcriptomic assessment of resistance to effects of an aryl hydrocarbon receptor (AHR) agonist in embryos of Atlantic killifish (Fundulus heteroclitus) from a marine superfund site.** *BMC Genomics* 2011, **12**:263.

44. Hoffmann AA, Willi Y: **Detecting genetic responses to environmental change.** *Nat Rev Genet* 2008, **9**:421–432.

45. Gahan L, Gould F, Heckel DG: **Identification of a gene associated with Bt resistance in Heliothis virescens.** *Science* 2001, **293**:857–860.

46. Martinez DE, Levinton J: **Adaptation to heavy metals in the aquatic oligochaete Limnodrilus hoffmeisteri: evidence for control by one gene.** *Evolution* 1996, **50**:1339–1343.

47. Meyer JN, Di Giulio RT: **Heritable adaptation and associated fitness costs in killifish (Fundulus heteroclitus) inhabiting a contaminated estuary.** *Ecol Appl* 2003, **13**:490–503.

48. Hahn ME, Karchner SI, Evans BR, Franks DG, Merson RR, Lapseritis JM: **Unexpected diversity of aryl hydrocarbon receptors in non-mammalian vertebrates: Insights from comparative genomics.** *J Exp Zool A Ecol Genet Physiol* 2006, **305A**:693–706.

49. Bak S-M, Iilda M, Hirano M, Iwato H, Kim E-Y: **Potencies of Red seabream AHR1- and AHR2-mediated transactivation by dioxins: implicationsof both AHRs in dioxin toxicity.** *Environ Sci Technol* 2013, **47**:2877–2885.

50. Mandel PK: **Dioxin: a review of its environmental effects and its aryl hydrocarbon receptor biology.** *J Comp Physiol B Biochem Syst EnvironPhysiol* 2005, **175**:221–230.

51. Williams LM, Oleksiak MF: **Evolutionary and functional analyses of cytochrome P4501A promoter polymorphisms in natural populations.** *Mol Ecol* 2011, **20**:5236–5247.

52. Morrison HG, Weil EJ, Karchner SI, Sogin ML, Stegeman JJ: **Molecular cloning of CYP1A from the estuarine fish Fundulus heteroclitus and**

phylogenetic analysis of CYP1 genes: update with new sequences. *Comp Biochem Physiol Part C Toxicol Pharmacol* 1998, **121**:231–240.

53. Billiard SM, Timme-Laragy AR, Wessenberg DM, Cockman C, Di Giulio RT: **The role of the aryl hydrocarbon receptor pathway in mediating synergistic developmental toxicity of polycyclic aromatic hydrocarbons to zebrafish.** *Toxicol Sci* 2006, **92**:526–536.

54. Goldstone H, Stegeman J: **A revised evolutionary history of the CYP1A subfamily: gene duplication, gene conversion and positive selection.** *J Mol Evol* 2006, **62**:708–717.

55. Zhou H, Qu Y, Wu H, Liao C, Zheng J, Diao X, Xue Q: **Molecular phylogenies and evolutionary behavior of AhR (aryl hydrocarbon receptor) pathway genes in aquatic animals: implications for the toxicology mechanism of some persistent pollutants (POPs).** *Chemosphere* 2010, **87**:193–205.

56. Lutgens SPM, Cleutjens KBJM, Daemen MJAP, Heeneman S: **Cathepsin cysteine proteases in cardiovascular disease.** *FASEB J* 2007, **21**:3029–3041.

57. Reiser J, Adair B, Reinheckel T: **Specialized roles for cysteine cathepsins in health and disease.** *J Clin Investig* 2010, **120**:3421–3431.

58. Hegelund T, Celandar MC: **Hepatic versus extrahepatic expression of CYP3A30 and CYP3A56 in adult killifish (Fundulus heteroclitus).** *Aquat Toxicol* 2003, **64**:277–291.

59. Kim JW, Lee Y, Kang HB, Choe YK, Chung TW, Chang SY, Kwang SL, Choe IS: **Cloning of the human cDNA sequence encoding the NADH: ubiquinone oxidoreductase MLRQ subunit.** *Biochem Mol Biol Int* 1997, **43**:669–675.