

## Genetic variation at aryl hydrocarbon receptor (*AHR*) loci in populations of Atlantic killifish (*Fundulus heteroclitus*) inhabiting polluted and reference habitats

Reitzel *et al.*

RESEARCH ARTICLE

Open Access

# Genetic variation at aryl hydrocarbon receptor (*AHR*) loci in populations of Atlantic killifish (*Fundulus heteroclitus*) inhabiting polluted and reference habitats

Adam M Reitzel<sup>1,2†</sup>, Sibel I Karchner<sup>1†</sup>, Diana G Franks<sup>1†</sup>, Brad R Evans<sup>1,5</sup>, Diane Nacci<sup>3</sup>, Denise Champlin<sup>3</sup>, Verónica M Vieira<sup>4,6</sup> and Mark E Hahn<sup>1\*</sup>

## Abstract

**Background:** The non-migratory killifish *Fundulus heteroclitus* inhabits clean and polluted environments interspersed throughout its range along the Atlantic coast of North America. Several populations of this species have successfully adapted to environments contaminated with toxic aromatic hydrocarbon pollutants such as polychlorinated biphenyls (PCBs). Previous studies suggest that the mechanism of resistance to these and other “dioxin-like compounds” (DLCs) may involve reduced signaling through the aryl hydrocarbon receptor (AHR) pathway. Here we investigated gene diversity and evidence for positive selection at three *AHR*-related loci (*AHR1*, *AHR2*, *AHRR*) in *F. heteroclitus* by comparing alleles from seven locations ranging over 600 km along the northeastern US, including extremely polluted and reference estuaries, with a focus on New Bedford Harbor (MA, USA), a PCB Superfund site, and nearby reference sites.

**Results:** We identified 98 single nucleotide polymorphisms within three *AHR*-related loci among all populations, including synonymous and nonsynonymous substitutions. Haplotype distributions were spatially segregated and *F*-statistics suggested strong population genetic structure at these loci, consistent with previous studies showing strong population genetic structure at other *F. heteroclitus* loci. Genetic diversity at these three loci was not significantly different in contaminated sites as compared to reference sites. However, for *AHR2* the New Bedford Harbor population had significant  $F_{ST}$  values in comparison to the nearest reference populations. Tests for positive selection revealed ten nonsynonymous polymorphisms in *AHR1* and four in *AHR2*. Four nonsynonymous SNPs in *AHR1* and three in *AHR2* showed large differences in base frequency between New Bedford Harbor and its reference site. Tests for isolation-by-distance revealed evidence for non-neutral change at the *AHR2* locus.

**Conclusion:** Together, these data suggest that *F. heteroclitus* populations in reference and polluted sites have similar genetic diversity, providing no evidence for strong genetic bottlenecks for populations in polluted locations. However, the data provide evidence for genetic differentiation among sites, selection at specific nucleotides in *AHR1* and *AHR2*, and specific *AHR2* SNPs and haplotypes that are associated with the PCB-resistant phenotype in the New Bedford Harbor population. The results suggest that *AHRs*, and especially *AHR2*, may be important, recurring targets for selection in local adaptation to dioxin-like aromatic hydrocarbon contaminants.

**Keywords:** Local adaptation, Pollution, Molecular mechanism, Resistance, Tolerance, Convergent evolution, Population genetics

\* Correspondence: mhahn@whoi.edu

†Equal contributors

<sup>1</sup>Biology Department, Woods Hole Oceanographic Institution, 45 Water Street, Woods Hole, MA 02568, USA

Full list of author information is available at the end of the article

## Background

Understanding the molecular basis of adaptation to environmental change is an important goal in environmental biology. Animal populations adapt to a variety of natural environmental stressors through genetic and epigenetic changes that affect gene expression or protein structure and/or function. Anthropogenic stressors, including toxic chemicals, can also drive selection in natural populations. For example, evolved resistance of insects to the acute neurotoxicity of insecticides is well known and occurs through a variety of mechanisms involving reduced target site sensitivity or enhanced expression of proteins involved in biotransformation and excretion of the chemicals [1-3]. Thus, we have learned a great deal about adaptation to chemicals designed to be toxic to their target organisms. However, field examples are less frequent and adaptive mechanisms are not as well understood for broadly distributed industrial pollutants that produce unintended consequences in non-target organisms. Recent studies (reviewed in [4-6]) have provided strong evidence for adaptation of fish populations to aromatic hydrocarbons such as polychlorinated biphenyls (PCBs), polychlorinated dibenzo-*p*-dioxins (PCDDs), and polycyclic aromatic hydrocarbons (PAHs) that cause toxicity similar to that caused by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). These “dioxin-like compounds” (DLCs) are capable of interfering with embryonic development and eliciting acute and chronic effects on reproduction, immune function, and other essential processes [7,8] with population-level consequences [9].

Populations of the non-migratory Atlantic killifish *Fundulus heteroclitus* that persist in highly contaminated environments may provide insight into the molecular mechanisms by which natural populations adapt to long-term, multi-generational exposure to DLCs. *F. heteroclitus* is widely used as an environmental model [10] for studying adaptations to natural environmental variables such as temperature [11,12] and evolved tolerance to anthropogenic chemicals [13]. Several distinct and geographically distant populations of *F. heteroclitus* inhabiting highly contaminated Superfund sites have been demonstrated to possess enhanced tolerance or resistance to one or more DLCs as compared to reference populations. The most well-studied populations are found in Superfund sites at Newark Bay, NJ (EPA ID: NJD980528996, contaminated with TCDD) [14-17], the Elizabeth River, VA (EPA ID: VAD990710410, contaminated with PAHs from creosote) [18-20], and the Acushnet River Estuary (EPA ID: MAD980731335, New Bedford Harbor (NBH), MA, contaminated with PCBs [21]) [22,23]. Adaptation also has been demonstrated in *F. heteroclitus* inhabiting more moderately contaminated sites [6,24].

The molecular mechanism(s) underlying the DLC resistance are not known for any of these populations.

However, the characteristics of the resistant phenotype provide important clues, which can be illustrated using the NBH population as an example. First, killifish embryos from NBH are less sensitive to the developmental toxicity of 3,3',4,4',5-pentachlorobiphenyl (PCB-126) than embryos from a reference site [22]. Second, when exposed to DLCs, NBH larvae display poor inducibility of the well-known biomarker cytochrome P450 1A (CYP1A) [22]. Similar insensitivity to DLCs is found in adult killifish from NBH, in which the resistance to altered CYP1A gene expression occurs in all tissues and at the level of gene transcription [23]. Third, the altered sensitivity of NBH fish to the toxic and biochemical effects of DLCs is heritable through at least 2 generations, consistent with genetic adaptation rather than physiological acclimation [6,22,25]. Most of these phenotypic characteristics are shared among independent DLC-resistant populations, suggesting that similar mechanisms of DLC tolerance have evolved in parallel at multiple sites [6,26].

Theoretical considerations suggest that adaptation to extreme pollution such as that found in NBH and other contaminated sites is more likely to result from major gene effects rather than polygenic adaptation [27,28]. Killifish populations at NBH and other highly polluted sites experience strong selection intensity (exposure to PCB concentrations well above the LC50), exhibit a large phenotypic shift (differences in sensitivity of two orders of magnitude as compared to reference populations [22]), have large population sizes [29], and have gene flow from neighboring populations [30]—all features that favor adaptation via single genes with large effects [27,28,31]. Consistent with this, previous studies have shown that the PCB-resistant NBH population has similar levels of overall genetic diversity compared with nearby populations of PCB-sensitive fish [29,32,33]. In light of these theoretical and empirical considerations and in keeping with a desire to employ a mechanistic perspective [34], we have taken a candidate gene approach to investigate the molecular basis of adaptation to DLCs in killifish.

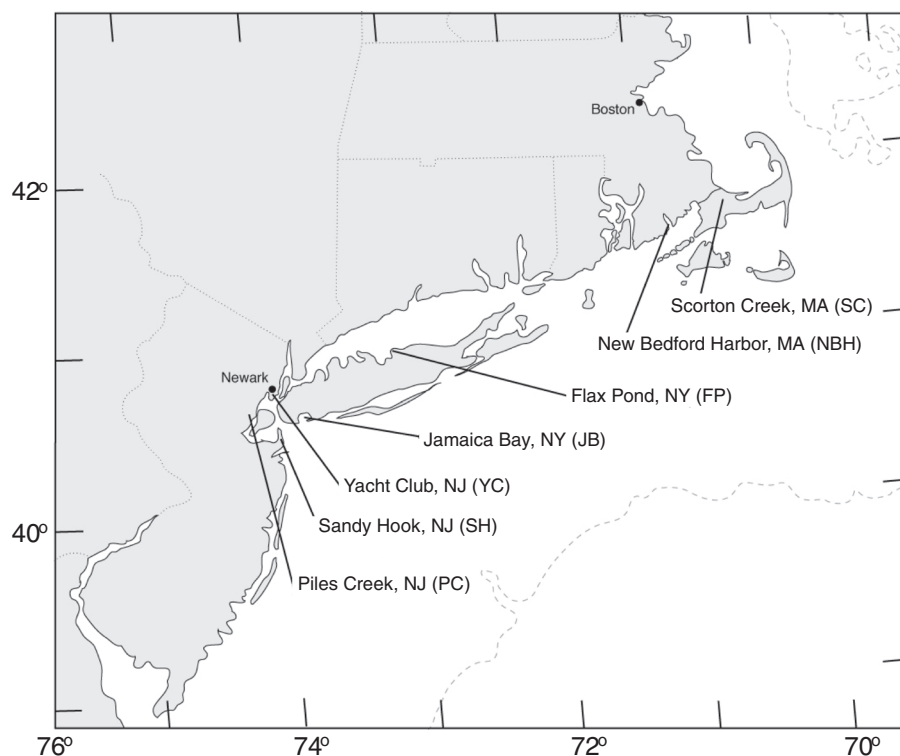
The most likely candidates for major genes affecting sensitivity to DLCs are those encoding proteins in the aryl hydrocarbon receptor (AHR)-dependent signaling pathway, the master regulator of responses to many of the most toxic DLCs, including TCDD and the PCBs with TCDD-like effects. The AHR is a ligand-activated transcription factor that exhibits high affinity for TCDD and other DLCs, regulates expression of a large set of genes in response to DLC exposure, and is required for TCDD or PCB toxicity in mammals [35,36] and fish [37,38]. Previously, we identified and cloned multiple components of the *F. heteroclitus* AHR pathway, including two AHR paralogs (AHR1, AHR2), an AHR nuclear translocator (ARNT2), and AHR repressor (AHRR) [39-42]. We hypothesized that allelic variation at one or more of these

loci resulted in proteins with reduced sensitivity to activation by DLCs. This hypothesis is consistent with results from other vertebrate species, where differences in the AHR pathway underlie many of the differences among species, strains, or cell lines in the sensitivity to DLC effects [43-46]. In some mammalian systems, reduced sensitivity to AHR agonists results from allelic variation at the *AHR* locus [47-50].

To identify common genetic loci associated with tolerance, we compared variation in AHR-related genes within and among fish populations resident to highly contaminated sites and nearby, less-contaminated reference sites. In an earlier study, we identified multiple *AHR1* alleles and observed differences in the frequency of these alleles between the PCB-sensitive (reference site) and -resistant (NBH) populations [51]. Here, we continue the focus on the NBH population and expand those studies to include additional populations of PCB-sensitive and -resistant fish as well as additional loci (*AHR2*, *AHRR*). The contaminated sites included two of the most well-studied DLC-tolerant populations of *F. heteroclitus*, NBH, MA, and Newark Bay, NJ (Yacht Club; YC) (Figure 1). We also included fish from two other highly contaminated sites suspected of tolerance to DLCs or other contaminants: Piles Creek, NJ (PC) [13,52], and

Jamaica Bay, NY (JB) [6], as well as three reference sites: Scorton Creek, MA (SC), Flax Pond, NY (FP), and Sandy Hook, NJ (SH) (Figure 1). A companion study [53] used a 'candidate gene scan' approach to investigate associations between DLC tolerance and SNP markers at 59 loci in four pairs of sensitive and tolerant populations of *F. heteroclitus*, including some of the same populations examined in the present study.

We hypothesized that there would be selection in favor of AHR variants that conferred reduced sensitivity to PCBs. This might be detected as purifying selection reducing diversity at one or more of the AHR loci, or as positive selection for certain SNPs or haplotypes in all polluted sites versus reference sites. Alternatively, evidence for selection might be population-specific. By sequencing the three AHR-related loci from individuals collected at these reference and polluted sites along the Atlantic coast of North America, we compare genetic and haplotype diversity and haplotype distribution, and use multiple methods to assess signatures of positive selection at particular nucleotides. Together, our data support previous studies indicating no loss of genetic diversity in populations at polluted sites, but suggest that particular nucleotides in each gene have a signature of selection that may underlie the differences in phenotype in killifish populations.



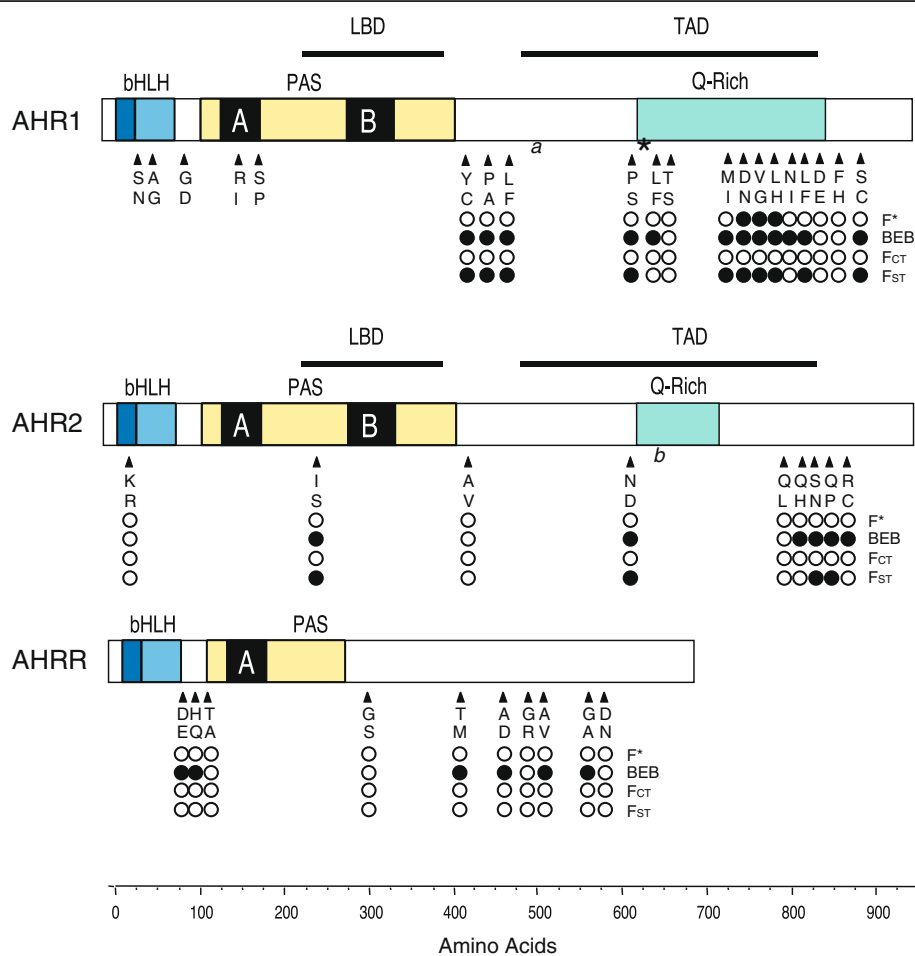
**Figure 1** Collection sites of *Fundulus heteroclitus* along the north Atlantic coast of the United States. Sample sizes for each population are provided in Table 1.

## Results

### AHR1, AHR2, and AHRR polymorphisms

**AHR1**—Previously, twenty-five single nucleotide polymorphisms (SNPs), nine of which were non-synonymous, were identified in killifish from SC and NBH [51]. Four of these nonsynonymous SNPs (ns-SNPs) occurred in the highly conserved and functionally important bHLH and PAS domains, but not at positions that are highly conserved among AHRs from different species. In the present study, full-length AHR1 cDNAs (2835 bp) were sequenced from 49 individuals from five new locations (JB, FP, PC, SH, and YC) and these data were combined with the exon 10 sequences of 52 individuals from SC and NBH determined earlier [51]. Overall, 44 SNPs, including 20 ns-

SNPs, were identified. Considering only exon 10 (the longest exon: 1540 bp), there were 31 SNPs, 15 of which were non-synonymous. Twelve of these ns-SNPs were clustered in the C-terminal half of the coding sequence, surrounding the Q-rich region that is involved in transactivation (Figure 2); the resulting amino acid replacements included a mixture of conservative and non-conservative changes. In addition to the SNPs, a 6-nucleotide deletion was also identified in some fish from JB, FP, PC, SH, and YC; none of the fish from SC or NBH had the deletion. Only sequences from exon 10, which contained most of the SNPs (Figure 2), were used in subsequent population comparisons, because for some of the fish only exon 10 sequences were available.



**Figure 2** Location of ns-SNPs in relation to AHR domain structure and results of three tests for selection at individual residues. For each position, the two amino acids are indicated at their position in the expressed protein. A six base-pair deletion identified in AHR1 is indicated by an asterisk. Significant results from tests for selection are indicated by a black dot, while non-significant positions are indicated by a white dot. The top row (F\*) indicates positions within each locus that were tested with Fu and Li's F\* statistic. The second row (BEB) indicates positively selected sites indicated by Bayesian posterior probabilities assessed in all unique haplotypes. The third row indicates sites with significantly different base frequencies in polluted versus reference populations, based on F-statistics (F<sub>CT</sub>). The bottom row indicates sites with significantly different base frequencies among all populations (F<sub>ST</sub>). In addition to the ns-SNPs in exon 10 that were used in the present analysis, the diagram for AHR1 includes ns-SNPs in the amino-terminal portion (exons 1–9) of AHR1 that were identified earlier [51] in the SC and NBH populations or in this study. The N-terminal ns-SNPs were not included in analyses for this study. The approximate locations of the AHR1\_1530 SNP (a) and AHR2\_1929 SNP (b) identified by Proestou et al. [53] are indicated.



**AHR2**—Although fish have multiple AHRs, we have observed that AHR2 is the predominant (most highly and widely expressed) form in many fishes, suggesting that AHR2 may have an important role in adaptive and toxic responses (as distinct from physiological responses) in fish [54,55]. Sequencing of 148 alleles (2860 bp each) from 74 fish at 7 locations revealed 29 SNPs, including 9 ns-SNPs. As observed for AHR1, the AHR2 ns-SNPs occurred primarily in the C-terminal half of the coding sequence, including within the Q-rich region (Figure 2); all of the C-terminal ns-SNPs resulted in non-conservative amino acid changes. Two ns-SNPs were found in the N-terminal half: a ns-SNP in the N-terminal basic region results in a conservative Lys to Arg change, and a ns-SNP in the PAS domain results in a non-conservative Ile to Ser change between the PAS-A and PAS-B repeats, in a region that is poorly conserved among species, but may have functional importance [56,57].

**AHRR**—AHRR sequences (2040 bp each) were obtained from 54 fish from the five Hudson River estuary sites. (Although AHRR is expressed in SC and NBH fish [42], no PCR products were obtained from SC or NBH fish. The reasons are not known, but could involve polymorphisms at primer target sites.) Thirty-eight SNPs, including 10 ns-SNPs, were found in the 54 sequenced fish (Table 1). The ns-SNPs occur in the C-terminal region of the coding sequence (6 ns-SNPs, 5 of which result in non-conservative aa changes), in a region after the PAS domain (1 ns-SNP), and in the region between the HLH and PAS domains (3 ns-SNPs, 2 of which result in conservative aa changes) (Figure 2).

**AHR1, AHR2, and AHRR haplotypes and genetic diversity**

**AHR1**—The 31 AHR1 SNPs were arranged in 76 distinct haplotypes (Table 1). Many of the haplotypes occurred

only once or a few times and frequently at only one location. The percentage of site-specific haplotypes (“private haplotypes”; black wedges in Figure 3) frequently exceeded that of haplotypes shared with other populations. The highest percentage of private haplotypes was from JB, where fish contained 14 total haplotypes of which 10 were unique to this location. A few haplotypes occurred at high frequency, and at multiple sites (Figure 3; Additional file 1: Figure S1). Overall genetic diversity for exon 10 of AHR1 among the sampled fishes was high ( $\pi = 0.00613$ ); however, there was no significant difference in nucleotide or haplotype diversity between polluted and reference locations (Table 2). The categorization of fish from JB as reference or polluted did not affect these results.

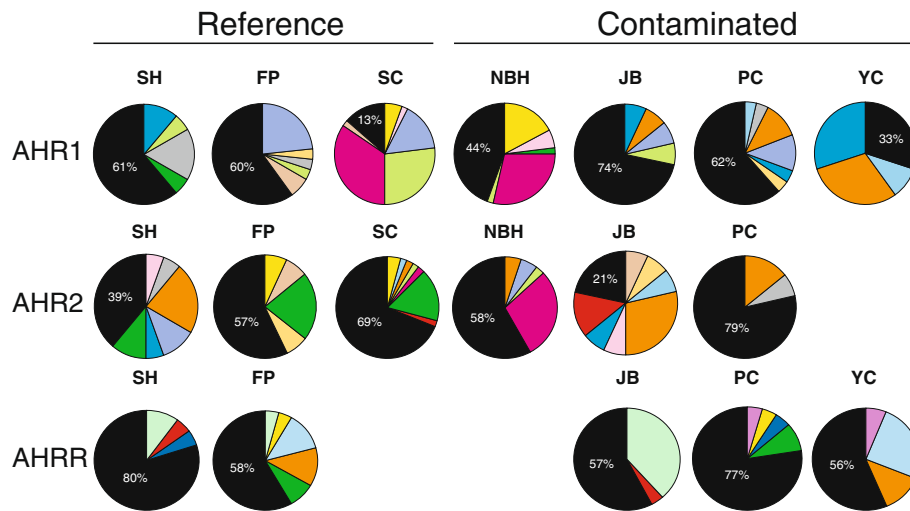
Pairwise  $F_{ST}$  comparisons of AHR1 among all populations identified approximately half (12 of 21) with significant differences between sites (Table 3).  $F_{ST}$  values varied from zero (PC vs SH) to 0.346 (YC vs NBH). There were no consistent patterns for genetic differentiation between reference and polluted sites. NBH and SC were significantly different when compared with all other populations (including the nearest other reference site FP) but not significantly different from one another.

**AHR2**—The 29 AHR2 SNPs were arranged in 65 haplotypes, many occurring only once or 2–3 times at a single location. As with AHR1, haplotypes unique to single locations accounted for more than half of the total, indicating a high proportion of private alleles (Figure 3). There were also a few shared, high-frequency haplotypes (e.g., at SC and FP; SH, JB, and PC) (Figure 3; Additional file 2: Figure S2). Compared to AHR1, overall genetic diversity for AHR2 among the sampled fishes was lower ( $\pi = 0.00238$ ). Similar to AHR1, there was no significant difference in nucleotide or haplotype diversity between

**Table 1 Data summary of sequencing results for *F. heteroclitus* AHR1, AHR2, and AHRR from seven locations along the Atlantic coast of the United States**

Population	AHR1					AHR2					AHRR				
	N	$\pi$	Hap	Hd	SNPs	N	$\pi$	Hap	Hd	SNPs	N	$\pi$	Hap	Hd	SNPs
NBH	52	0.0054	10	0.836	22	38	0.0026	18	0.902	22					
SC	52	0.0057	12	0.792	24	48	0.0024	23	0.916	25					
JB	14	0.0046	14	1	18	14	0.0009	10	0.923	7	26	0.0031	10	0.812	15
FP	30	0.0043	21	0.945	23	14	0.0015	11	0.956	16	24	0.0037	18	0.971	29
PC	26	0.0055	20	0.975	23	14	0.0020	8	0.912	12	22	0.0046	21	0.966	30
SH	18	0.0058	15	0.974	24	18	0.0015	11	0.928	12	20	0.0038	17	0.984	24
YC	10	0.0026	5	0.844	11	2	0	1	0	0	16	0.0033	10	0.925	20
<b>Summary</b>	<b>202</b>	<b>0.0061</b>	<b>76</b>	<b>0.948</b>	<b>31</b>	<b>148</b>	<b>0.0024</b>	<b>65</b>	<b>0.970</b>	<b>29</b>	<b>108</b>	<b>0.0042</b>	<b>65</b>	<b>0.976</b>	<b>38</b>

Note: N = number of alleles sequenced.  $\pi$  = nucleotide diversity. Hap = number of haplotypes found. Hd = Haplotype diversity. SNPs = number of single nucleotide polymorphisms found. No AHRR sequences were obtained from NBH or SC fish. The “Summary” line (bold text) includes the total number of distinct haplotypes and SNPs across all populations. SNP and haplotype counts for AHR1 include sequences from exon 10 only. The unique haplotypes for each AHR locus are available in the Dryad repository, at doi:10.5061/dryad.t2888.



**Figure 3 Haplotype frequencies for each AHR locus among populations.** The black wedge indicates the percentage of site-specific haplotypes. Colored wedges indicate shared haplotypes. See Additional files 1, 2 and 3 for more detailed description of haplotypes. Overall, all populations have a higher proportion of unique haplotypes regardless of the locus or habitat type (i.e., reference or polluted).

polluted and reference locations overall (Table 2). The categorization of fish from JB as reference or polluted did not affect the results.

Pairwise  $F_{ST}$  comparisons among all populations revealed a similar number of significant relationships as for AHR1 (11 of 15), but with some different populations exhibiting significant genetic differentiation (Table 3). (YC could not be compared with other populations because only one individual was sequenced.)  $F_{ST}$  values varied from 0.0122 (PC vs SH) to 0.279 (JB vs FP). There were no consistent patterns for genetic differentiation between

reference and polluted sites when considered as a group. However, NBH (PCB-resistant population) had significant  $F_{ST}$  values in comparison to both SC and FP (the two nearest reference populations;  $F_{ST}$  values of 0.142 and 0.216, respectively), while SC did not differ significantly from FP ( $F_{ST} = 0.0468$ ).

**AHRR**—The 38 AHRR SNPs occurred in 65 haplotypes, most of which were low-frequency and site-specific (Figure 3; Additional file 3: Figure S3). Some higher-frequency AHRR haplotypes were present in most populations but these showed similar distribution among the contaminated and reference sites. Overall genetic diversity for AHRR was intermediate between AHR1 and AHR2 ( $\pi = 0.00417$ ). There was no significant difference in nucleotide or haplotype diversity between polluted and reference locations (Table 2). As observed for AHR1 and AHR2, the categorization of fish from Jamaica Bay as reference or polluted did not affect the results.

For AHRR, pairwise  $F_{ST}$  comparisons among all populations identified half (5 of 10) that were significantly different (Table 3).  $F_{ST}$  values varied from zero (PC vs FP) to 0.270 (JB vs FP). Similar to AHR1 and AHR2, there were no consistent patterns for genetic differentiation between reference and polluted sites.

**Table 2 Statistical comparison of nucleotide and haplotype diversity of *F. heteroclitus* from polluted and reference sites**

Character	Gene	Polluted	Reference	P-value
Nucleotide diversity	AHR1	0.00451 (0.00137)	0.00525 (0.000818)	P = 0.4098
	AHR2	0.00183 (0.00085)	0.00181 (0.00052)	P = 0.9695
	AHRR	0.00367 (0.00082)	0.00371 (0.000057)	P = 0.9365
Haplotype diversity	AHR1	0.9138 (0.0858)	0.9037 (0.0978)	P = 0.8937
	AHR2	0.9123 (0.0106)	0.9333 (0.0206)	P = 0.2158
	AHRR	0.9110 (0.0928)	0.9775 (0.00919)	P = 0.3400

For these comparisons JB was grouped with other polluted populations. However, when JB was considered a reference site, the relationships of nucleotide and haplotypes diversity remained unchanged. Values shown are the mean and (standard deviation).

#### Tests for selection

We used three methods for detecting candidate nucleotides undergoing selection: Tajima's D and Fu and Li's  $F^*$ , likelihood ratio tests, and position-specific F-statistics. Tajima's D was not significant for any of the loci (AHR1:  $D = 1.298$ ,  $p > 0.1$ ; AHR2:  $D = 0.841$ ,  $p > 0.1$ ; AHRR:  $D = 0.848$ ,  $p > 0.1$ ). However, Fu and Li's  $F^*$  test was significant for AHR1 ( $F = 2.069$ ,  $p < 0.02$ ) but not

**Table 3 Pairwise  $F_{ST}$  values between all populations for AHR1, AHR2, and AHRR**

	SH	PC	YC	JB	FP	NBH	SC
<b>AHR1</b>							
SH	0						
PC	-0.00467	0					
YC	0.26461	0.1819	0				
JB	0.08137	0.04123	0.14891	0			
FP	0.08411	0.04209	<b>0.26796</b>	<b>0.0717</b>	0		
NBH	<b>0.2316</b>	<b>0.24953</b>	<b>0.34609</b>	<b>0.32853</b>	<b>0.30805</b>	0	
SC	<b>0.16025</b>	<b>0.15137</b>	<b>0.2387</b>	<b>0.18917</b>	<b>0.14406</b>	0.06992	0
<b>AHR2</b>							
SH	0						
PC	0.01223	0					
JB	0.06531	0.1143		0			
FP	<b>0.11681</b>	<b>0.14203</b>		<b>0.27903</b>	0		
NBH	<b>0.16712</b>	<b>0.12756</b>		<b>0.15459</b>	<b>0.21634</b>	0	
SC	<b>0.12239</b>	<b>0.12998</b>		<b>0.17283</b>	0.04684	<b>0.14197</b>	0
<b>AHRR</b>							
SH	0						
PC	0.09077	0					
YC	<b>0.13945</b>	0.06193	0				
JB	0.09775	<b>0.24847</b>	<b>0.20896</b>	0			
FP	<b>0.10319</b>	-0.00318	0.06668	<b>0.26985</b>	0		

Bold values indicate statistically significant differences after correcting for multiple comparisons. YC was omitted from the AHR2 analysis because of insufficient data. Similarly, AHRR sequences were not available from NBH and SC; see methods for details.

for AHR2 ( $F = 1.024$ ,  $p > 0.1$ ) or AHRR ( $F = 0.319$ ,  $p > 0.1$ ). An F-statistic significantly greater than zero for AHR1 indicated an excess of intermediate frequency alleles. Using the sliding window approach for AHR1, three nucleotide positions, clustered in the Q-rich region of the transactivation domain (TAD), were significant (top row of AHR1 in Figure 2).

The likelihood ratio tests implemented with codeml in the PAML software suite provided evidence for positive selection shaping the frequency of nonsynonymous polymorphisms in all three loci. Both tests (M1a vs. M2a; M7 vs. M8) were highly significant ( $p < 0.001$ ) for each locus. For both comparisons, the Bayes Empirical Bayes (BEB) method identified the same set of nucleotides as under positive selection; these included 12 residues in AHR1, 6 in AHR2, and 6 in AHRR (second row in Figure 2). Each of these residues was inferred with high probability ( $p > 0.99$ ) to be under strong selection, with ratios of nonsynonymous substitutions to synonymous substitution ( $\omega$ ) greater than 9. The 12 residues identified in AHR1 represented 80% of the ns-SNPs sequenced from these populations and 2.2% of the total residues. These sites are dispersed throughout the sequenced region of exon 10 with intervening sites showing no evidence of positive selection. For AHR2, one identified residue

was in the region between the two PAS domains and the remaining residues were in the TAD or further in the C-terminus; four of these were successive substitutions in the C-terminus. The six positively selected sites for AHRR were scattered throughout this locus, representing 60% of ns-SNPs, and included two residues in the region between the bHLH and PAS domains. Over all three loci, 1.1% of codons and more than 70% of nonsynonymous substitutions were identified as being under positive selection.

Locus-by-locus AMOVA was used to test for significant differences in SNP frequencies between populations classified as polluted versus reference ( $F_{CT}$ , third row of Figure 2). These tests did not identify any ns-SNPs in AHR1 with significant differences in these two habitat types. These results were not affected by classification of JB as a polluted or reference site. However, nucleotide frequencies were significantly different among populations for 10 of 15 ns-SNPs when all geographic locations were included ( $F_{ST}$ ; fourth row of Figure 2) or for 9 sites within reference or polluted ( $F_{SC}$ ) (Additional file 4: Table S1). Synonymous substitutions, like the ns-SNPs, did not show significant variation between reference and polluted populations (Additional file 4: Table S1,  $F_{CT}$  column), but they did show some significant  $F_{ST}$  and  $F_{SC}$  values.



As observed for AHR1, F-statistics for AHR2 and AHRR identified no ns-SNPs as having significant differences in frequencies between populations classified as polluted versus reference. Four AHR2 ns-SNPs had significant  $F_{ST}$  values (Figure 2) and a single ns-SNP located in the PAS domain region of AHR2 (an I/S replacement) had a significant  $F_{SC}$  value (Additional file 4: Table S1). For AHRR, no ns-SNPs had significant  $F_{SC}$  or  $F_{ST}$  values. Synonymous substitutions in both AHR2 and AHRR were also not significant when populations were grouped as reference or polluted, but two synonymous SNPs showed significant  $F_{SC}$  and six showed significant  $F_{ST}$  (Additional file 4: Table S1).

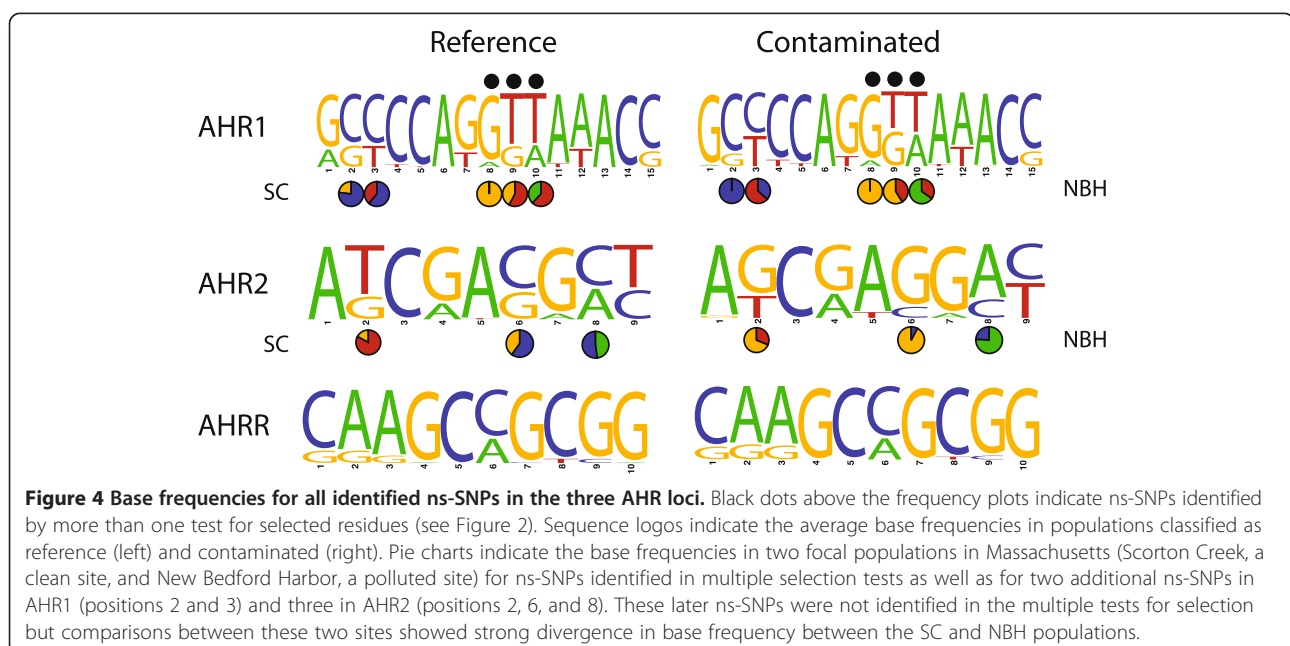
Comparing results of these different tests for ns-SNPs under selection, three ns-SNPs in the Q-rich domain of AHR1 were identified by three tests (Figure 2). An additional seven ns-SNPs in AHR1 and four ns-SNPs in AHR2 were identified by two tests. No ns-SNPs in AHRR were identified by more than one test.

We compared the base composition between reference and contaminated sites for each of the ns-SNPs in all three loci (Figure 4). For the three ns-SNPs identified by three tests for selection (black dots above the nucleotide position), we compared the frequency of each nucleotide in the SC and NBH populations. Despite overall similarity in base frequency between reference and polluted sites, the base frequencies varied considerably between SC and NBH for two of these ns-SNPs in AHR1: ns-SNPs 9 and 10. Two other positions (ns-SNPs 2 and 3) also show base frequency differences between SC and NBH and in the comparisons of all populations were

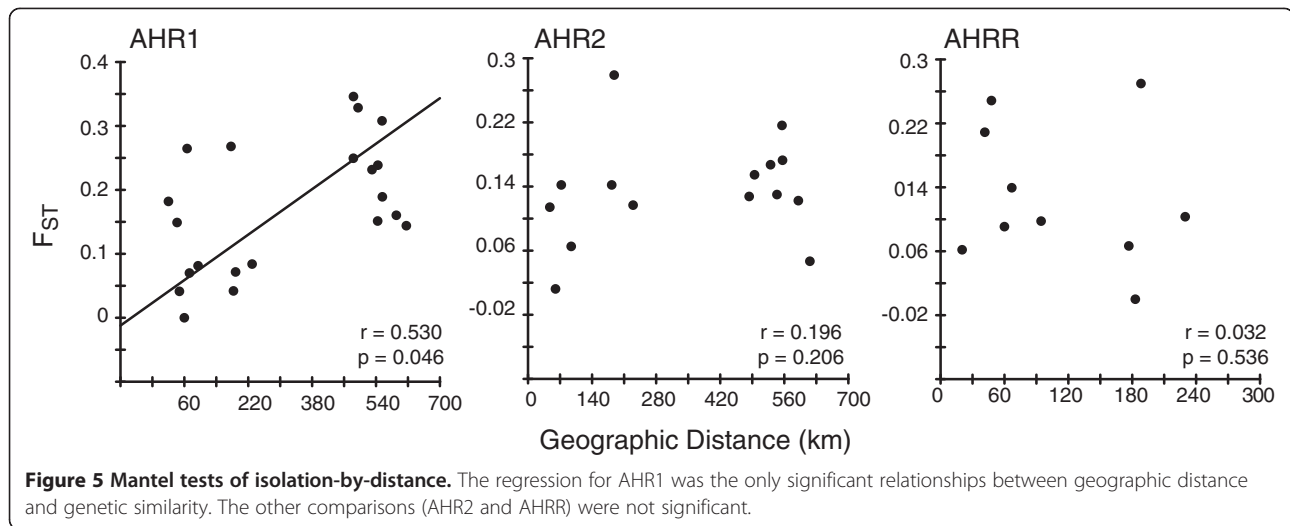
identified as under selection by both likelihood ratio tests and AMOVA- $F_{ST}$ . Three positions in AHR2 showed dramatic differences in base composition between SC and NBH fish: ns-SNP 2 (NBH: 31.5% T, SC: 80% T), ns-SNP 6 (NBH: 8% C; SC: 60% C), and ns-SNP 8 (NBH: 76% A; SC: 48% A) (Figure 4). Two of these (ns-SNPs 2 and 8) were also identified by both likelihood ratio tests and AMOVA- $F_{ST}$  as under selection in the comparisons of all populations.

#### Isolation by distance

The sampled populations in this analysis spanned a total geographic distance of more than 600 km along the Atlantic coastline of the United States (see Additional file 5: Table S2 for geographic distances). Statistical tests for isolation-by-distance resulted in a significant regression when AHR1 diversity was compared among all sampled fishes ( $r = 0.530$ ,  $p = 0.046$ ), a relationship consistent with a neutral expectation, although distance explains only part of the difference in genetic diversity. The regressions for AHR1 remained positive when fish were separated by type of habitat (i.e., reference vs. polluted); however these statistical tests were not significant (data not shown). In contrast to the results for AHR1, results for AHR2 and AHRR indicated non-significant relationships (possible non-neutral evolution) both when comparing among all estuaries (Figure 5) and when sites were separated by habitat type (data not shown). An alternative possibility, perhaps likely for AHRR, is that sample sizes were too small to detect isolation by distance.



**Figure 4** Base frequencies for all identified ns-SNPs in the three AHR loci. Black dots above the frequency plots indicate ns-SNPs identified by more than one test for selected residues (see Figure 2). Sequence logos indicate the average base frequencies in populations classified as reference (left) and contaminated (right). Pie charts indicate the base frequencies in two focal populations in Massachusetts (Scorton Creek, a clean site, and New Bedford Harbor, a polluted site) for ns-SNPs identified in multiple selection tests as well as for two additional ns-SNPs in AHR1 (positions 2 and 3) and three in AHR2 (positions 2, 6, and 8). These later ns-SNPs were not identified in the multiple tests for selection but comparisons between these two sites showed strong divergence in base frequency between the SC and NBH populations.



## Discussion

The repeated evolution of resistance to DLCs in widely separated populations of *F. heteroclitus* along the U.S. east coast provides an opportunity to understand the mechanistic basis for rapid adaptation to anthropogenic environmental change. There is strong evidence—initially from the widespread loss of inducibility of AHR-regulated CYP1A [reviewed in 6] and subsequently confirmed by gene expression profiling [26,58,59]—that this adaptation involves altered sensitivity of the AHR-dependent signaling pathway. Thus, we used a candidate gene approach and focused on three known AHR-related genes in seven populations of *F. heteroclitus*. Our analysis of the sequence data from all seven locations reveals a complex pattern of selection at the three loci. Because our primary focus has been on the NBH Superfund site [23,42,51,59-63], we also examined the patterns of variation at NBH and its two nearest reference sites, SC and FP.

### Comparisons of seven populations from polluted and reference sites

Three AHR-related loci (AHR1, AHR2, AHRR) from *F. heteroclitus* inhabiting seven estuaries along the U.S. east coast contain a large number of polymorphisms, many of which result in changes in the encoded amino acids. Overall, 1.5% of the nucleotide positions were variable among the sequences analyzed in this study, and 38% of the SNPs were nonsynonymous. In contrast, AHR2 in tomcod sampled from three sites (60 alleles total) showed very low nucleotide variability (0.1%) [64]. For comparison, 3.5% of the nucleotide positions were variable in AHR coding sequences from 13 inbred strains of mice (*Mus musculus*) [65]. By contrast, the human AHR (0.4%) exhibits much less variability than either mouse or killifish AHRs [66,67].

In previous studies examining inter-specific and intra-specific variability in AHR sequences [45,65,68], the most highly conserved region is the basic-helix-loop-helix (bHLH) domain, which is involved in DNA binding and protein dimerization [69,70]. The Per-Arnt-Sim (PAS) domains, required for ligand-binding and protein-protein interactions [69,70], are also well conserved, whereas the C-terminal half of the protein that harbors transactivation domains [71,72] is more variable [45,65,68]. Our results showing that the majority of ns-SNPs are in the C-terminal half of the sequences or between bHLH and PAS domains (Figure 2) are consistent with these earlier results.

The genetic diversity of these *F. heteroclitus* populations at these three loci is strongly partitioned among locations, but there are no significant differences in nucleotide diversity between populations inhabiting polluted habitats versus those at relatively clean habitats. Similarly, each locus is represented by dozens of haplotypes that exhibit a high degree of location-specific distribution but, again, with no consistent differences in haplotype diversity in polluted versus reference habitats. In previous studies, examination of other sequence-based markers, microsatellites, and anonymous markers has led to the conclusion that there is restricted gene flow among these populations (i.e., genetic structure) and that populations inhabiting pollutant-impacted sites show no strong signature for a genetic bottleneck (i.e., loss of genetic diversity) [29,32,33,53,73]. Our results show that these conclusions also pertain to the three AHR-related gene loci. Thus, the populations exhibit strong genetic structure at these loci but no loss of nucleotide or haplotype diversity in populations classified as “polluted.”

Despite an overall similarity in genetic diversity between reference and polluted populations, a suite of tests suggested that some loci and certain polymorphisms may

be under selection. The three statistical methods we used test for deviations from neutrality using different metrics. Tajima's D and Fu and Li's  $F^*$  statistics test for a statistically significant excess or reduction of allele frequencies among sampled individuals. The likelihood/BEB method tests for an excess of nonsynonymous substitutions compared to synonymous substitutions along particular haplotype lineages, at each position. The comparative F-statistics test for significant differences in base-frequencies among populations ( $F_{ST}$ ) or between groups of populations classified as reference or polluted ( $F_{CT}$ ). These tests may identify different sets of nucleotides potentially under selection. However, when the same nucleotide is identified by more than one test, it increases confidence that it has been shaped by selection. The most consistently identified three SNPs, identified by three tests, were located in the Q-rich region of AHR1, in a region of exon 10 associated with the transcriptional activation function [71,72]. Other ns-SNPs in exon 10 of AHR1 and in AHR2 were identified by two tests (Figure 2).

The position-specific likelihood ratio tests identified more residues under selection when compared with the other tests. Generally, ratio comparisons of nonsynonymous and synonymous SNPs are considered conservative tests for positive selection [74]. Additionally, our results were inferred through analysis of haplotypes in a phylogenetic framework, thus allowing a more accurate representation of the evolution of particular nucleotides within the lineage of an allele. Whether all identified SNPs represent true positives is uncertain, in part due to the limitations of our data set. One limitation is that recombination within loci can hamper interpretation of residues under selection in tree-based analyses by increasing the proportion of false positives [75]. An initial test for recombination at each of the three loci using the GARD test [76] suggests potential recombination events in the sampled individuals from this study. These analyses indicate that one (AHR1, AHRR) or two (AHR2) recombination events likely have occurred in our sampled sequences. Such a low frequency of recombination is unlikely to cause false positives in nucleotide-specific tests for selection [75]. Likelihood ratio tests are also sensitive to data sets in which polymorphic sites are not independent, and because of linkage the BEB analysis may over-represent residues under selection [77]. Thus, some SNPs identified as undergoing selection may represent SNPs in linkage disequilibrium with neighboring nucleotides for which selection was operating. Discerning among these possibilities would be assisted by sequencing additional loci from these individuals.

More broadly, our results provide a mixed assessment of which AHR locus may represent the best candidate for explaining evolved resistance in natural populations

of *F. heteroclitus*. Similar to population genetics studies of other *F. heteroclitus* loci [32,53,73,78-80], at each AHR locus we observed a high number of polymorphisms that segregate among populations, with many haplotypes restricted to individual locations. Such a large proportion of geographically restricted genetic diversity reflects this species' large population sizes and relatively limited migration between adjacent locations [10]. AHR1 and AHR2 are possible candidates for explaining the mechanism of molecular adaptation by populations to polluted environments. For the statistical tests for selection, AHR1 had proportionally larger numbers of ns-SNPs with evidence of selection. However, AHR1 diversity showed a significant (though weak) relationship with geographic distance, a result consistent with either neutral evolution (isolation by distance) or selection pressure that correlates with latitude (e.g., temperature, photoperiod) [81]. However, DLC contamination is not correlated with latitude, i.e., the polluted sites in this portion of *F. heteroclitus*' range, as well as along the Atlantic coast, are interspersed among clean sites. Thus, the Mantel test result for AHR2, showing the lack of a relationship of genetic diversity with geographic distance, is more consistent with adaptation to local environments. On the other hand, we found fewer AHR2 ns-SNPs to be under selection, although some of these variable positions showed significant differences when comparing all populations ( $F_{ST}$ ) and in comparison of NBH and SC (see below). Experimental tests to empirically determine functional characteristics (e.g., PAH binding, protein-protein interactions) of the diverse AHR1 and AHR2 allelic variants would help to discern the role of SNPs in adaptation in these populations and to develop hypotheses about the role of particular haplotypes in polluted and reference populations of *F. heteroclitus*.

#### **NBH versus reference population comparisons**

Examination of AHR diversity in multiple populations, including several exhibiting resistance to DLCs [6], revealed evidence for AHR loci and specific SNPs under selection, but the population genetic data are complex and their interpretation is not straightforward. A limitation of this multi-population approach is that resistance is likely to have evolved independently in the different resistant populations and may involve different loci or different SNPs or haplotypes under selection. In addition, our classification of locations as "polluted" combined locations with very different types of pollution (PCBs, dioxins, PAHs, metals), and included a population (PC) for which DLC resistance has not yet been assessed. It is useful, then, to also take a more focused look at the population of greatest interest in our studies, the one inhabiting the NBH Superfund site, and the two nearest reference

populations, SC and FP, thus minimizing effects of geographic distance on the genetic data.

Consistent with the Mantel test showing isolation-by-distance for AHR1 across all populations, the NBH and SC populations did not show strong genetic differentiation at this locus, but each had significant pairwise  $F_{ST}$  values in comparison to all of the more distant populations (including FP). By contrast, for the AHR2 locus NBH had significant  $F_{ST}$  values in pairwise comparisons to SC and FP (Table 3), while the two reference sites did not differ significantly from each other, despite the fact that they are farther apart from each other than either is from NBH. These results were supported by the distinct pattern of haplotype frequencies at NBH as compared with SC or FP (Figure 3) and by the identification of three AHR2 ns-SNPs for which NBH and SC differ substantially (Figure 4). Thus, examination of these three populations points to specific AHR2 SNPs and haplotypes as being associated with the PCB-resistant phenotype.

One interesting result is that, for both AHR1 and AHR2, neither specific haplotypes nor the SNPs exhibiting evidence for selection were fixed in DLC-resistant fish populations, raising questions about their contribution to the resistant phenotype. One possibility is that these loci individually have relatively small effect and are part of a larger polygenic adaptation response [82]. Alternatively, there could be multiple haplotypes at one of these loci (e.g., AHR2) that confer resistance. Such a situation could arise from selection on pre-existing (standing) genetic variation, in which one or more SNPs conferring reduced AHR function exists in multiple haplotypes in the population prior to environmental change, and selection leads to fixation of multiple alleles (soft sweep [83-85]). Population genomic studies will help to distinguish between these possibilities.

#### **Role of AHR2 in controlling susceptibility of fish to DLC effects**

Fish have multiple AHR genes, classified in two clades, AHR1 and AHR2 [86]. The functions of AHR1 and AHR2 are not completely understood, but AHR2 is the most likely candidate for a resistance locus, based on several lines of evidence. First, studies using gene-specific knock-down in zebrafish embryos have shown that AHR2 controls the induction of CYP1A and sensitivity to developmental toxicity of TCDD, PCBs, and PAHs in this species [37,38,87]. Second, AHR2 was one of the candidate genes emerging from a genome-wide QTL screen for genes controlling PCB cardiotoxicity in zebrafish embryos [88]. Third, and more directly relevant to the species of interest in the current study, knock-down of AHR2 in embryos of *F. heteroclitus* provided partial protection against the teratogenic effects of PAHs and PCBs [89].

In addition to the experimental studies cited above, two recent population-level studies suggest AHR2 as a resistance locus. In an independent analysis being published as a companion paper in this journal [53], a 'candidate gene scan' investigation of associations between DLC resistance and SNP markers at 59 loci in four pairs of sensitive and tolerant populations of *F. heteroclitus* identified AHR2 as one of two loci under selection (the other was CYP1A) [53]. There is partial overlap in the populations studied by Proestou et al. [53] and in the present paper (NBH, FP, SH, YC/NWK) but the other populations examined were specific to each study (us: SC, JB, PC; Proestou et al.: BL, BP, ER, KC). In Proestou et al. [53], SNPs in both AHR1 (AHR1\_1530) and AHR2 (AHR2\_1929) exhibited evidence of selection (significant  $F_{ST}$  values) in 3 of 4 population pairs, including NBH and its reference site. In our study, AHR1\_1530 also had a significant  $F_{ST}$  value in a locus-by-locus AMOVA and it is located just downstream from three ns-SNPs also exhibiting evidence for selection (Figure 2; Additional file 4: Table S1). Although the AHR2\_1929 SNP did not have a significant  $F_{ST}$  value in our study (Additional file 4: Table S1), it was near a ns-SNP that did (AHR2\_1813; N/D amino acids in Figure 2). Additional evidence for selection at the AHR2\_1929 SNP in Proestou et al. [53] came from patterns of minor allele frequencies between pairs of populations and the identification of this SNP as the only outlier after  $F_{ST}$  modeling of pooled sensitive and tolerant populations [53].

A second study, in another fish species showing population-specific evolution of PCB resistance, also implicated the AHR2 locus. Atlantic tomcod (*Microgadus tomcod*) inhabiting the PCB-contaminated Hudson River were nearly monomorphic for an AHR2 variant with reduced capacity to bind and be activated by halogenated AHR ligands such as TCDD or PCB-126 [64]. The AHR2 variant in Hudson River fish was characterized by a 2-amino acid deletion, just downstream from the PAS domain, that was proposed to alter the ligand-binding affinity or stability of the AHR2 protein in these fish. A similar deletion was not found in the AHR2 variants of NBH killifish in our study, but a SNP within the PAS domain and several near the C-terminal transactivation domain emerged as potentially under selection and with distinct patterns in NBH fish as compared to the reference sites (Figures 2, 4).

Based on our results and those described above [53,64,88,89], we suggest that evolution of resistance to PCBs in fish may converge on a common target gene, AHR2, but that the specific molecular changes may differ between species, and perhaps also within a species among populations that have independently evolved the resistant phenotype (for other examples, see [90,91]). Nevertheless, changes in other loci—including paralogous AHR loci



(see below) as well as other loci encoding proteins involved in the mechanism of dioxin toxicity—may also play a role in conferring the resistant phenotype. Population genomic surveys currently underway will help illuminate such possibilities.

Since completion of this work, through transcriptome sequencing [59], we have identified two additional AHR loci in *F. heteroclitus*, paralogs of the AHR loci studied here. (The differences between paralog sequences are sufficiently large so that the paralogs could not have interfered with the sequencing or SNP determinations reported in this paper.) Multiple AHRs, often occurring as pairs of paralogous AHR1 and/or AHR2 forms, have been identified in other species of fish including *Danio rerio* (zebrafish), *Takifugu rubripes* and *Tetraodon nigroviridis* (pufferfishes), *Oryzias latipes* (medaka), and salmonids (reviewed in [86]). Consistent with phylogenetic relationships (unpublished analysis) and the nomenclature we have used for other fish AHRs [86], the original killifish AHR genes (the focus of this paper) have been designated AHR1a and AHR2a; the novel AHR genes are AHR1b and AHR2b. The function and expression patterns of these new AHRs are not known, but are under active investigation in our laboratory. Sequencing and assembly of the *F. heteroclitus* genome has revealed that AHR1a and AHR2a occur in tandem (~14 kb apart), as do AHR1b and AHR2b (~4 kb apart), as we have described for other fish AHR1/AHR2 pairs [86,92]. Linkage of AHR1a and AHR2a may have influenced the patterns of diversity and evidence for selection obtained in our study and that of Proestou et al. [53], for example by causing both AHR1 and AHR2 to display evidence for selection even if only one of these genes may be involved in the mechanism of resistance. Clearly, additional research will be needed to determine the function of the new AHRs and the possible role of all four AHR genes in evolved resistance to PCBs and related chemicals.

## Conclusion

The data presented here suggest that *F. heteroclitus* populations in reference and polluted sites have similar genetic diversity, with no evidence for genetic bottlenecks in populations inhabiting polluted locations. However, the populations exhibit strong genetic structure at all three AHR-related loci, and for AHR2 the NBH population exhibits significant genetic differentiation from its two nearby reference sites. In addition, the data revealed positive selection at specific nucleotides in AHR1 and AHR2, and specific AHR2 SNPs and haplotypes that are associated with the PCB-resistant phenotype in the NBH population. The results suggest that AHRs, and especially AHR2, may be recurring targets for selection during local adaptation of fish to dioxin-like aromatic hydrocarbon contaminants, although the specific molecular changes

may vary among independently adapting populations or species.

## Methods

### Site selection, fish collection, and sample processing

*F. heteroclitus* (26 fish per site) were collected from New Bedford Harbor, MA, USA (NBH; PCB-contaminated site) and Scorton Creek, Sandwich, MA, USA (SC; reference site for NBH) in May-June, 2003 as part of a previous study on AHR1 alleles [51]. Additional *F. heteroclitus* (15 fish per site) were collected between June and October 2002 from five sites within or near the lower Hudson River ecosystem (Figure 1). The polluted sites were: Newark Bay, NJ [Roanoke Yacht Club (YC) [14,15]], Piles Creek, NJ (PC) [13,52], and Jamaica Bay, NY (JB) [6] (Figure 1). The additional reference sites were: Sandy Hook, NJ (SH) [6] and Flax Pond, NY (FP) [6,15]. These sites were chosen because the PCB sensitivities of most of their *F. heteroclitus* populations have been characterized and sediment PCB levels have been measured [6,22-24,32], allowing us to classify them as polluted or reference. The exception was Piles Creek, a highly contaminated site with killifish that have evolved resistance to methyl mercury [13,52] but also show some abnormalities [93,94]; DLC resistance has not yet been assessed for this population. Fish were collected and tissues sampled using protocols approved by the Woods Hole Oceanographic Institution's Animal Care and Use Committee (Animal Welfare Assurance Number A3630-01).

Note on number of alleles analyzed: A formal power analysis was not performed prior to conducting these studies. Although the number of alleles sampled was sufficient to detect selection despite the high genetic diversity, sampling of a greater number of alleles from each site may have allowed us to identify additional SNPs potentially under selection. The number of alleles sampled here (10–52 per population) is in line with numbers used in other studies seeking evidence for adaptive genetic change, for example in color patterns in beach mice (8–40 alleles per population [95]), tomcod exhibiting resistance to PCBs (20–124 alleles per population [64]), and rats evolving resistance to warfarin (variable number of alleles per population [96]).

### Oligonucleotide primers

Primers were synthesized by Midland Certified Reagent Company, Inc. (Midland, Texas), Life Technologies, Inc. (Rockville, MD) or Integrated DNA Technologies, Inc. (Coralville, IA). Primer sequences are listed in Additional file 6: Table S3.

### RNA isolation, RT-PCR, and DNA sequencing

Total RNA was isolated from combined soft tissue of individual fish using RNA STAT-60 (Tel-Test B, Inc.;



Friendswood, TX). RNA quality was assessed by gel electrophoresis. PolyA<sup>+</sup> RNA was purified with the MicroPoly (A) Purist kit (Ambion, Grand Island, NY). First-strand cDNA was synthesized from 2 µg of polyA<sup>+</sup> RNA using the Omniscript Reverse Transcription kit (Qiagen, Valencia, CA). When possible, we amplified the full coding sequences using a single pair of primers (Additional file 6: Table S3); in some cases, we used two pairs of oligonucleotide primers to produce overlapping fragments of ~1500 bp each. For these PCR reactions, 1 µl of undiluted cDNA was used with the amplification primers indicated in Additional file 6: Table S3, using Advantage 2 polymerase mix (Clontech, Mountain View, CA). PCR conditions were: For AHR1: 95°C, 1 min.; 5 cycles of [95°C, 5 sec., 73°C, 5(1.5\*) min.], 5 cycles of [95°C, 5 sec., 71°C, 5(1.5\*) min.], 35 cycles of [95°C, 5 sec., 69°C, 5(1.5\*) min.], 72°C, 7 min., For AHR2: 95°C, 1 min., 5 cycles of [95°C, 5 sec., 72°C, 3(2.5\*) min.]; 5 cycles of [95°C, 5 sec., 70°C, 3(2.5\*) min.]; 40(35\*) cycles of [95°C, 5 sec., 68°C, 3(2.5\*) min.]; 72°C, 7 min. For AHRR: 95°C, 1 min.; 5 cycles of [95°C, 5 sec., 72°C, 2.5 min.]; 5 cycles of [95°C, 5 sec., 70°C, 2.5 min.]; 35 cycles of [95°C, 5 sec., 68°C, 2.5 min.]; 72°C, 7 min (\* indicates program used for amplification from New Bedford Harbor and Scorton Creek samples). PCR products were initially confirmed by gel electrophoresis, and then purified with the MinElute PCR Purification Kit (Qiagen). After purification, PCR products were sequenced directly on an ABI 3730 capillary sequencer (Marine Biological Laboratory, Woods Hole, MA) using gene specific oligonucleotide primers (Additional file 6: Table S3). For unknown reasons, sequencing was not successful for all individuals from each site. This was particularly true for AHR2 sequences from YC fish, and AHRR sequences from SC and NBH fish.

#### Sequence analysis

Sequences were initially scanned with Editview 1.0.1 and imported into Sequencher 4.1, which aligns the sequences and allows for the direct comparison of each electrophoretogram. The nucleotide and codon number were noted for each polymorphic site, and whether the base change resulted in a change in the amino acid at that site (nonsynonymous SNPs).

#### Haplotype reconstruction and data analysis

Haplotypes were inferred using PHASE v.2.02 [97,98], which implements a Bayesian statistical method to reconstruct haplotypes from unphased genotype data. The Bayesian approach used in PHASE is more accurate than the widely used Expectation-Maximization (EM) algorithm and other methods [98,99]. The output includes a summary of results with an estimate of population haplotype frequency and lists of the most probable haplotype pairs for each individual. Several PHASE runs (4 or 5) using

different values for the seed of the random number generator (-S function) were performed for each gene. The number of iterations, thinning interval, and burn-in values were increased to 1000, 10, and 1000, respectively, from the default values. The results from the multiple runs were compared with respect to the allele frequencies to check for consistency. Also, the goodness of fit outputs from different runs were compared by single-factor ANOVA to assess variation among runs.

TCS software [100] was used to estimate the genealogical relationships among the haplotypes. TCS uses the method of Templeton *et al.* [101] to reconstruct phylogenies while taking into account recombination events. Haplotype frequency data were incorporated into the TCS output.

#### Genetic diversity and data analysis

Genetic diversity of sampled populations was assessed by comparing nucleotide and haplotype diversity within and between populations. Both measures of diversity were calculated with DnaSP v.5 [102]. Genetic diversity measures were statistically compared with t-tests (JMP) by categorizing fish populations into two types: clean (reference; SC, FP, SH) and polluted (NBH, JB, PC, YC). While JB killifish have been shown to be sensitive to DLCs [6], site contamination and the presence of some PCB "hot spots" confounds the categorization as a reference site. Thus, alternate statistical comparisons were performed with this population categorized as either polluted or reference. To test for population genetic structure, pairwise measures of genetic differentiation among populations were calculated with F-statistics for each gene (Arlequin v.3 [103]). Significant relationships were assessed at  $p = 0.002$  for AHR1 and AHR2 and  $p = 0.005$  for AHRR to account for multiple comparisons ( $p = 0.05/\text{number of comparisons}$ ).

#### Tests for selection

Three tests using all SNPs (synonymous and nonsynonymous) were conducted to assess potential signatures of selection among these three loci. First, summary-statistic based methods (i.e., Tajima's D, Fu and Li's F\*) were analyzed in DnaSP. For these analyses, full length, aligned sequences for AHR2 and AHRR and exon 10 of AHR1 were used as input. [For NBH and SC fish, only exon 10 sequences were available for AHR1; exon 10 contains the majority of the SNPs at this locus [51].] The two methods differ in that Fu and Li's F\* is based on the difference between the number of singleton polymorphisms and the number expected under neutrality, given the number of segregating positions, while Tajima's D takes into account the difference between average pairwise diversity between sequences. Fu and Li's F\* test can therefore account for some degree of population structure [104], which is

expected for this species. In either test, a value that significantly exceeds zero indicates an excess of intermediate-frequency alleles that could result from balancing selection, while negative values indicate an excess of low frequency alleles, which may indicate purifying selection. For each test, we used the sliding window (100 bp window, 25 bp step) implemented in DnaSP to investigate whether particular regions of each gene showed significant signatures for differentiation among the sampled populations. Significance was assessed at  $p < 0.05$ .

Second, tree-based methods to test for positive selection were implemented in PAML v.4 (codeml, [105]). For these analyses, all unique haplotype sequences for each locus were used. For these position-specific tests for selection, a phylogenetic tree was required. For each locus, best trees for AHR1, AHR2, and AHRR were produced for all unique haplotypes with maximum likelihood analyses (RAxML, [106]) using the best model for nucleotide substitutions (jModelTest [107]). Support for nodes was determined with 1000 bootstrap replicates. Each analysis resulted in a single best tree with low bootstrap ( $< 50$ ) for most nodes. Two pairs of likelihood ratio tests [105] [108] were used to test for evidence of positive selection. In the first pair we compared the null model of nearly neutral evolution (M1a) to the alternate model of positive selection (M2a). The second test compares a model of a beta-distributed variable selection pressure (M7) to the alternate, which includes positive selection (M8). Codons under selection were determined with posterior probabilities determined by the Bayes Empirical Bayes (BEB) method [108]. For tests of each locus we performed the repeated comparisons with different codon frequency models to see if the results were influenced by this parameter. The results from these sensitivity tests found that this parameter did not change the nucleotide positions inferred to be under selection.

Third, we used analysis of molecular variance (AMOVA) to determine differentiation among populations and between sets of populations classified as polluted or reference. All nucleotide variants in our analyses are from the coding region of each transcript and represent a combination of synonymous and nonsynonymous polymorphisms, which may display variable signals of population structure and be under different degrees of selection. Selection may result in significant differences in F-statistics for particular nucleotides if they are not evolving under neutral conditions. The locus-by-locus AMOVA feature of Arlequin [103] was used to determine F-statistics for each variable position. We were especially interested in nucleotide positions that were significantly different between the set of populations classified as polluted as compared to those classified as reference ( $F_{CT}$ ) but also calculated variation among all populations ( $F_{ST}$ ) and among populations within each

class ( $F_{SC}$ ). To reduce Type 1 errors, we assessed significant differences at  $p < 0.0001$ .

We compared the frequency of polymorphisms identified in at least one test for selection in fish collected from New Bedford Harbor and Scorton Creek, Massachusetts. For these positions, we constructed sequence logos (Weblogo v. 3 [109]) to display the frequency of each base from all fishes categorized by reference or polluted population. We then calculated the frequency of these bases in the two focal populations at these positions as well as other positions at which there were large differences in base frequency.

### Isolation by distance

Statistical tests of isolation-by-distance were carried out for each gene among all populations and by studying populations from reference and polluted sites separately. Geographic distances between populations were determined by calculating a smoothed coastal distance between locations that ignored small inlets, when present. Pairwise genetic distances (multilocus  $F_{ST}$ ) were calculated between each population pair with Arlequin. Geographic and genetic distances were regressed with a web-implementation of Isolation By Distance v3.16 [110] with 1000 randomizations. Regressions were additionally completed using the linearized value of  $F_{ST}/(1-F_{ST})$  in place of  $F_{ST}$ . The relationships were unchanged and we only report results using  $F_{ST}$  for genetic distance.

### Availability of supporting data

The data supporting the results of this article (unique haplotypes for each AHR locus) are available in the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.t2888>.

### Additional files

**Additional file 1: Figure S1.** AHR1 haplotype frequencies and distributions. Haplotypes were reconstructed using PHASE and genealogical relationships among the haplotypes were estimated using TCS software, as described in *Materials and Methods*. In the top panel, circles refer to unique haplotypes, wedges are colored by sampled population, and numbers within wedges refer to the number of alleles with that haplotype in the population represented by that color. Numbers outside of the circles refer to the haplotype number as shown in the bottom panel. In the bottom panel, black wedges indicate the percentage of site-specific (SS) haplotypes at each site and colored wedges indicate haplotypes shared among populations.

**Additional file 2: Figure S2.** AHR2 haplotype frequencies and distributions. Haplotypes were reconstructed using PHASE and genealogical relationships among the haplotypes were estimated using TCS software, as described in *Materials and Methods*. For additional description, see legend to Additional file 1: Figure S1.

**Additional file 3: Figure S3.** AHRR haplotype frequencies and distributions. Haplotypes were reconstructed using PHASE and genealogical relationships among the haplotypes were estimated using TCS software, as described in *Materials and Methods*. For additional description, see legend to Additional file 1: Figure S1.

**Additional file 4: Table S1.** Locus-by-locus F-statistics for AHR-related loci from *Fundulus heteroclitus*. The locus-by-locus AMOVA feature of Arlequin was used to determine F-statistics for each variable nucleotide. Nucleotide positions that were significantly different among populations within each class ( $F_{SC}$ ), among all populations ( $F_{ST}$ ), and between the set of populations classified as polluted as compared to those classified as reference ( $F_{CT}$ ) were assessed using a significant cutoff of  $p < 0.0001$ . SS=synonymous site; NS=nonsynonymous site; nonsynonymous positions are also shaded. Nucleotide number (nt#) refers to the position with respect to the ATG translational start (A=1) unless otherwise specified.

**Additional file 5: Table S2.** Pairwise geographic distances (km) for locations where *Fundulus heteroclitus* were collected for this study.

**Additional file 6: Table S3.** PCR primers.

#### Abbreviations

TCDD: 2,3,7,8-tetrachlorodibenzo-p-dioxin; PCB-126: 3,3',4,4',5-pentachlorobiphenyl; ARNT: AHR nuclear translocator; AHRR: AHR repressor; AMOVA: Analysis of molecular variance; AHR: Aryl hydrocarbon receptor; BEB: Bayes Empirical Bayes; CYP1A: Cytochrome P450 1A; DLC: Dioxin-like compound; FP: Flax Pond; JB: Jamaica Bay; NBH: New Bedford Harbor; YC: Newark Yacht Club; ns-SNPs: Nonsynonymous SNPs; PC: Piles Creek; PCB: Polychlorinated biphenyl; PCDD: Polychlorinated dibenzo-p-dioxin; PAHs: Polycyclic aromatic hydrocarbon; SH: Sandy Hook; SC: Scorton Creek; SNPs: Single nucleotide polymorphisms; TAD: Transactivation domain.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

AMR carried out the population genetic and statistical analyses and drafted the manuscript. SIK participated in the design of the study, sequencing and sequence analysis, carried out the haplotype analyses, and edited the manuscript. DGF participated in sequencing and sequence analysis and edited the manuscript. BRE participated in the design of the study, collection of fish, and sequencing. DN participated in the design of the study, led the collection of fish, and edited the manuscript. DC participated in the collection of fish. VMV participated in the data analysis involving geographic distances. MEH conceived of the study, participated in its design and coordination, helped to draft the manuscript, and edited the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

We thank Sarah Cohen for helpful discussions and Dina Proestou and three anonymous reviewers for valuable comments on the original manuscript. This work was supported in part by the Hudson River Foundation (grant 004/02A; final report available at <http://www.hudsonriver.org/lis/>), by National Institute of Environmental Health Sciences (NIEHS) grant P42ES007381 (Superfund Basic Research Program at Boston University), by grant F32HD062178 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD), and by the National Science Foundation (DEB-1120263). Data interpretation was aided by reference to a preliminary draft of the *F. heteroclitus* genome sequence, which was supported by funding from the National Science Foundation (collaborative research grants DEB-1120512, DEB-1265282, DEB-1120013, DEB-1120263, DEB-1120333, DEB-1120398). The funding agencies were not involved in study design or performance or in the decision to publish the manuscript. The U.S. Government is authorized to produce and distribute reprints for governmental purposes notwithstanding any copyright notation that may appear hereon.

#### Author details

<sup>1</sup>Biology Department, Woods Hole Oceanographic Institution, 45 Water Street, Woods Hole, MA 02568, USA. <sup>2</sup>Department of Biology, University of North Carolina at Charlotte, Charlotte, NC 28223, USA. <sup>3</sup>NHEERL, Atlantic Ecology Division, U.S. Environmental Protection Agency, Narragansett, RI 02882, USA. <sup>4</sup>Department of Environmental Health, Boston University School of Public Health, Boston, MA 02118, USA. <sup>5</sup>Current address: Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY 10029, USA. <sup>6</sup>Current address: University of California, Program in Public Health, Irvine, CA 92697, USA.

Received: 7 August 2013 Accepted: 18 December 2013

Published: 14 January 2014

#### References

1. McKenzie JA: *Ecological and Evolutionary Aspects of Insecticide Resistance*. Austin, TX: R.G. Landes Company and Academic Press; 1996.
2. Taylor M, Feyereisen R: **Molecular biology and evolution of resistance to toxicants.** *Mol Biol Evol* 1996, **13**(6):719–734.
3. Li X, Schuler MA, Berenbaum MR: **Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics.** *Annual Rev Entomol* 2007, **52**:231–253.
4. Wirgin I, Waldman JR: **Resistance to contaminants in North American fish populations.** *Mut Res* 2004, **552**(1–2):73–100.
5. van Veld PA, Nacci DE: **Chapter 13. Toxicity Resistance.** In *The Toxicology of Fishes*. Edited by Di Giulio RT, Hinton DE. Boca Raton: Taylor & Francis; 2007.
6. Nacci DE, Champlin D, Jayaraman S: **Adaptation of the estuarine fish *Fundulus heteroclitus* (Atlantic killifish) to polychlorinated biphenyls (PCBs).** *Estuar Coast Shelf Sci* 2010, **33**:853–864.
7. Billiard SM, Meyer JN, Wassenberg DM, Hodson PV, Di Giulio RT: **Nonadditive effects of PAHs on Early Vertebrate Development: mechanisms and implications for risk assessment.** *Toxicol Sci* 2008, **105**(1):5–23.
8. Goldstone HM, Stegeman JJ: **Molecular mechanisms of 2,3,7,8-tetrachlorodibenzo-p-dioxin cardiovascular embryotoxicity.** *Drug Metab Rev* 2006, **38**(1–2):261–289.
9. Cook PM, Robbins JA, Endicott DD, Lodge KB, Guiney PD, Walker MK, Zabel EW, Peterson RE: **Effects of aryl hydrocarbon receptor-mediated early life stage toxicity on lake trout populations in Lake Ontario during the 20th century.** *Environ Sci Technol* 2003, **37**(17):3864–3877.
10. Burnett KG, Bain LJ, Baldwin WS, Callard GV, Cohen S, Di Giulio RT, Evans DH, Gómez-Chiarri M, Hahn ME, Hoover CA, Karchner SJ, Katoh F, MacLatchy DL, Marshall WS, Meyer JN, Nacci DE, Oleksiak MF, Rees BB, Singer TP, Stegeman JJ, Towle DW, Veld PAV, Vogelbein WK, Whitehead A, Winn RN, Crawford DL: ***Fundulus* as the premier teleost model in environmental biology: Opportunities for new insights using genomics.** *Compar Biochem Physiol Part D: Genom Proteom* 2007, **2**:257–286.
11. Powers DA, Schulte PM: **Evolutionary adaptations of gene structure and expression in natural populations in relation to a changing environment: a multidisciplinary approach to address the million-year saga of a small fish.** *J Exper Zool* 1998, **282**(1–2):71–94.
12. Schulte PM: **Environmental adaptations as windows on molecular evolution.** *Compar Biochem Physiol - Part B: Biochem Mol Biol* 2001, **128**(3):597–611.
13. Weis JS, Weis P: **Tolerance and stress in a polluted environment.** *Bioscience* 1989, **39**:89–95.
14. Prince R, Cooper KR: **Comparisons of the effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on chemically impacted and nonimpacted subpopulations of *Fundulus heteroclitus*: I: TCDD toxicity.** *Environ Toxicol Chem* 1995, **14**(4):579–587.
15. Elskus AA, Monosson E, McElroy AE, Stegeman JJ, Woltering DS: **Altered CYP1A expression in *Fundulus heteroclitus* adults and larvae: a sign of pollutant resistance?** *Aquat Toxicol* 1999, **45**:99–113.
16. Arzuaga X, Elskus A: **Polluted-site killifish (*Fundulus heteroclitus*) embryos are resistant to organic pollutant-mediated induction of CYP1A activity, reactive oxygen species, and heart deformities.** *Environ Toxicol Chem* 2010, **29**(3):676–682.
17. Prince R, Cooper KR: **Comparisons of the effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on chemically impacted and nonimpacted subpopulations of *Fundulus heteroclitus*: II: Metabolic considerations.** *Environ Toxicol Chem* 1995, **14**(4):589–595.
18. Van Veld PA, Westbrook DJ: **Evidence for depression of cytochrome P4501A in a population of chemically resistant mummichog (*Fundulus heteroclitus*).** *Environ Sci* 1995, **3**(4):221–234.
19. Meyer JN, Nacci DE, Di Giulio RT: **Cytochrome P4501A (CYP1A) in killifish (*Fundulus heteroclitus*): heritability of altered expression and relationship to survival in contaminated sediments.** *Toxicol Sci* 2002, **68**(1):69–81.
20. Ownby DR, Newman MC, Mulvey M, Vogelbein WK, Unger MA, Arzayus LF: **Fish (*Fundulus heteroclitus*) populations with different exposure histories differ in tolerance of creosote-contaminated sediments.** *Environ Toxicol Chem* 2002, **21**(9):1897–1902.
21. Weaver G: **PCB contamination in and around New Bedford, Mass.** *Environ Sci Technol* 1984, **18**:22A–27A.



22. Nacci D, Coiro L, Champlin D, Jayaraman S, McKinney R, Gleason T, Munns WR Jr, Specker JL, Cooper K: Adaptation of wild populations of the estuarine fish *Fundulus heteroclitus* to persistent environmental contaminants. *Mar Biol* 1999, **134**(1):9–17.
23. Bello SM, Franks DG, Stegeman JJ, Hahn ME: Acquired resistance to aryl hydrocarbon receptor agonists in a population of *Fundulus heteroclitus* from a marine Superfund site: In vivo and in vitro studies on the induction of xenobiotic-metabolizing enzymes. *Toxicol Sci* 2001, **60**(1):77–91.
24. Nacci DE, Champlin D, Coiro L, McKinney R, Jayaraman S: Predicting the occurrence of genetic adaptation to dioxinlike compounds in populations of the estuarine fish *Fundulus heteroclitus*. *Environ Toxicol Chem* 2002, **21**(7):1525–1532.
25. Bello SM: Characterization of resistance to halogenated aromatic hydrocarbons in a population of *Fundulus heteroclitus* from a marine superfund site. Ph.D. Thesis. Woods Hole: Woods Hole Oceanographic Institution/Massachusetts Institute of Technology; 1999.
26. Whitehead A, Pilcher W, Champlin D, Nacci D: Common mechanism underlies repeated evolution of extreme pollution tolerance. *Proc Royal Soc B* 2012, **279**(1728):427–433.
27. Woods R, Hoffman A: Chapter 9: Evolution in Toxic Environments: Quantitative Versus Major Gene Approaches. In *Demography in Ecotoxicology*. Edited by Kammenga J, Laskowski R. Wiley; 2000.
28. Macnair MR: Why the evolution of resistance to anthropogenic toxins normally involves major gene changes: the limits to natural selection. *Genetica* 1991, **84**:213–219.
29. 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**(6):539–548.
30. Brown B, Chapman R: Gene flow and mitochondrial DNA variation in the killifish: *Fundulus heteroclitus*. *Evolution* 1991, **45**(5):1147–1161.
31. Hoffmann AA, Willi Y: Detecting genetic responses to environmental change. *Nat Rev Genet* 2008, **9**(6):421–432.
32. Cohen S: Strong positive selection and habitat-specific amino acid substitution patterns in MHC from an estuarine fish under intense pollution stress. *Mol Biol Evol* 2002, **19**(11):1870–1880.
33. Roark SA, Nacci D, Coiro L, Champlin D, Guttman SI: Population genetic structure of a nonmigratory estuarine fish (*Fundulus heteroclitus*) across a strong gradient of polychlorinated biphenyl contamination. *Environ Toxicol Chem* 2005, **24**(3):717–725.
34. Dalziel AC, Rogers SM, Schulte PM: Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. *Mol Ecol* 2009, **18**(24):4997–5017.
35. Fernandez-Salguero P, Hilbert DM, Rudikoff S, Ward JM, Gonzalez FJ: Aryl-hydrocarbon receptor-deficient mice are resistant to 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced toxicity. *Toxicol Appl Pharmacol* 1996, **140**:173–179.
36. Mimura J, Yamashita K, Nakamura K, Morita M, Takagi T, Nakao K, Ema M, Sogawa K, Yasuda M, Katsuki M, et al: Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor. *Genes Cells* 1997, **2**(10):645–654.
37. Prasch AL, Teraoka H, Carney SA, Dong W, Hiraga T, Stegeman JJ, Heideman W, Peterson RE: Aryl Hydrocarbon Receptor 2 mediates 2,3,7,8-Tetrachlorodibenzo-p-dioxin developmental toxicity in zebrafish. *Toxicol Sci* 2003, **76**:138–150.
38. Jönsson ME, Jenny MJ, Woodin BR, Hahn ME, Stegeman JJ: Role of AHR2 in the expression of novel cytochrome P450 1 family genes, cell cycle genes, and morphological defects in developing zebra fish exposed to 3,3',4,4',5-pentachlorobiphenyl or 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol Sci* 2007, **100**(1):180–193.
39. Hahn ME, Karchner SI, Shapiro MA, Perera SA: Molecular evolution of two vertebrate aryl hydrocarbon (dioxin) receptors (AHR1 and AHR2) and the PAS family. *Proc Natl Acad Sci USA* 1997, **94**(25):13743–13748.
40. Powell WH, Karchner SI, Bright R, Hahn ME: Functional diversity of vertebrate ARNT proteins: Identification of ARNT2 as the predominant form of ARNT in the marine teleost, *Fundulus heteroclitus*. *Arch Biochem Biophys* 1999, **361**(1):156–163.
41. Karchner SI, Powell WH, Hahn ME: Identification and functional characterization of two highly divergent aryl hydrocarbon receptors (AHR1 and AHR2) in the teleost *Fundulus heteroclitus*: Evidence for a novel subfamily of ligand-binding basic helix-loop-helix Per-ARNT-Sim (bHLH-PAS) factors. *J Biol Chem* 1999, **274**(47):33814–33824.
42. 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**(9):6949–6959.
43. Hahn ME: Mechanisms of innate and acquired resistance to dioxin-like compounds. *Rev Toxicol* 1998, **2**(5,6):395–443.
44. Lavine JA, Rowatt AJ, Klimova T, Whittington AJ, Dengler E, Beck C, Powell WH: Aryl Hydrocarbon Receptors in the frog *Xenopus laevis*: Two AhR1 paralogs exhibit low affinity for 2,3,7,8-Tetrachlorodibenzo-p-Dioxin (TCDD). *Toxicol Sci* 2005, **88**(1):60–72.
45. Karchner SI, Franks DG, Kennedy SW, Hahn ME: The molecular basis for differential dioxin sensitivity in birds: Role of the aryl hydrocarbon receptor. *Proc Natl Acad Sci USA* 2006, **103**:6252–6257.
46. Farmahin R, Manning GE, Crump D, Wu D, Mundy LJ, Jones SP, Hahn ME, Karchner SI, Giesy JP, Bursian SJ, et al: Amino acid sequence of the ligand binding domain of the aryl hydrocarbon receptor 1 (AHR1) predicts sensitivity of wild birds to effects of dioxin-like compounds. *Toxicol Sci* 2013, **131**:139–152.
47. Poland A, Palen D, Glover E: Analysis of the four alleles of the murine aryl hydrocarbon receptor. *Mol Pharmacol* 1994, **46**:915–921.
48. Pohjanvirta R, Wong JMY, Li W, Harper PA, Tuomisto J, Okey AB: Point mutation in intron sequence causes altered carboxyl-terminal structure in the aryl hydrocarbon receptor of the most 2,3,7,8-tetrachlorodibenzo-p-dioxin-resistant rat strain. *Mol Pharmacol* 1998, **54**:86–93.
49. Sun W, Zhang J, Hankinson O: A mutation in the aryl hydrocarbon receptor (AHR) in a cultured mammalian cell line identifies a novel region of AHR that affects DNA binding. *J Biol Chem* 1997, **272**(50):31845–31854.
50. Ema M, Ohe N, Suzuki M, Mimura J, Sogawa K, Ikawa S, Fujii-Kuriyama Y: Dioxin binding activities of polymorphic forms of mouse and human aryl hydrocarbon receptors. *J Biol Chem* 1994, **269**:27337–27343.
51. Hahn ME, Karchner SI, Franks DG, Merson RR: Aryl hydrocarbon receptor polymorphisms and dioxin resistance in Atlantic killifish (*Fundulus heteroclitus*). *Pharmacogenetics* 2004, **14**:131–143.
52. Weis JS, Heber M, Weis P, Vaidya S: Methylmercury tolerance of killifish (*Fundulus heteroclitus*) embryos from a polluted vs non-polluted environment. *Mar Biol* 1981, **65**:283–287.
53. Proestou DA, Flight P, Champlin D, Nacci D: Targeted Approach to Identify Genetic Loci Associated with Evolved Dioxin Tolerance in Atlantic Killifish (*Fundulus heteroclitus*). *BMC Evol Biol* 2014, **14**:7.
54. Hahn ME: Dioxin toxicology and the aryl hydrocarbon receptor: Insights from fish and other non-traditional models. *Mar Biotechnol* 2001, **3**(1):S224–S238.
55. Hahn ME: Aryl hydrocarbon receptors: diversity and evolution. *Chem-Biol Inter* 2002, **141**(1/2):131–160.
56. Chapman-Smith A, Lutwyche JK, Whitelaw ML: Contribution of the Per/Arnt/Sim (PAS) domains to DNA binding by the basic helix-loop-helix PAS transcriptional regulators. *J Biol Chem* 2004, **279**(7):5353–5362.
57. McGuire J, Okamoto K, Whitelaw ML, Tanaka H, Poellinger L: Definition of a dioxin receptor mutant that is a constitutive activator of transcription: delineation of overlapping repression and ligand binding functions within the PAS domain. *J Biol Chem* 2001, **276**(45):41841–41849.
58. Whitehead A, Triant DA, Champlin D, Nacci D: Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population. *Mol Ecol* 2010, **19**(23):5186–5203.
59. Oleksiak MF, Karchner SI, Jenny MJ, Franks DG, Mark Welch DB, 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**(1):263.
60. Powell WH, Bright R, Bello SM, Hahn ME: Developmental and tissue-specific expression of AHR1, AHR2, and ARNT2 in dioxin-sensitive and -resistant populations of the marine fish, *Fundulus heteroclitus*. *Toxicol Sci* 2000, **57**:229–239.
61. Harbeitner RC, Hahn ME, Timme-Laragy AR: Differential sensitivity to pro-oxidant exposure in two populations of killifish (*Fundulus heteroclitus*). *Ecotoxicology* 2013, **22**:387–401.
62. Aluru N, Karchner SI, Hahn ME: Role of DNA methylation of AHR1 and AHR2 promoters in differential sensitivity to PCBs in Atlantic Killifish, *Fundulus heteroclitus*. *Aquat Toxicol* 2011, **101**(1):288–294.
63. Greytak SR, Tarrant AM, Nacci D, Hahn ME, Callard GV: Estrogen responses in killifish (*Fundulus heteroclitus*) from polluted and unpolluted environments are site- and gene-specific. *Aquat Toxicol* 2010, **99**:291–299.

64. 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.
65. Thomas RS, Penn SG, Holden K, Bradfield CA, Rank DR: **Sequence variation and phylogenetic history of the mouse Ahr gene.** *Pharmacogenetics* 2002, **12**(2):151–163.
66. Harper PA, Wong JMY, Lam MSM, Okey AB: **Polymorphisms in the human AH receptor.** *Chem-Biol Inter* 2002, **141**:161–187.
67. Rowlands CJ, Staskal DF, Gollapudi B, Budinsky R: **The human AHR: identification of single nucleotide polymorphisms from six ethnic populations.** *Pharmacogen Genom* 2010, **20**(5):283–290.
68. Carver LA, Hogenesch JB, Bradfield CA: **Tissue specific expression of the rat Ah-receptor and ARNT mRNAs.** *Nucl Acids Res* 1994, **22**:3038–3044.
69. Dolwick KM, Swanson HI, Bradfield CA: **In vitro analysis of Ah receptor domains involved in ligand-activated DNA recognition.** *Proc Natl Acad Sci USA* 1993, **90**:8566–8570.
70. Fukunaga BN, Probst MR, Reiszporszasz S, Hankinson O: **Identification of functional domains of the aryl hydrocarbon receptor.** *J Biol Chem* 1995, **270**(49):29270–29278.
71. Jain S, Dolwick KM, Schmidt JV, Bradfield CA: **Potent transactivation domains of the Ah receptor and Ah receptor nuclear translocator map to their carboxyl termini.** *J Biol Chem* 1994, **269**:31518–31524.
72. Rowlands JC, McEwan IJ, Gustafsson JA: **Trans-activation by the human aryl hydrocarbon receptor and aryl hydrocarbon receptor nuclear translocator proteins: direct interactions with basal transcription factors.** *Mol Pharmacol* 1996, **50**(3):538–548.
73. 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**(4):1109–1123.
74. Nielsen R: **Statistical tests of selective neutrality in the age of genomics.** *Heredity* 2001, **86**(Pt 6):641–647.
75. Anisimova M, Nielsen R, Yang Z: **Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites.** *Genetics* 2003, **164**(3):1229–1236.
76. Delpont W, Poon AFY, Frost SDW, Kosakovsky Pond SL: **Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology.** *Bioinformatics* 2010, **26**(19):2455–2457.
77. Bustamante CD, Wakeley J, Sawyer SA, Hartl DL: **Directional selection and the site-frequency spectrum.** *Genetics* 2001, **159**(4):1779–1788.
78. Mulvey M, Newman MC, Vogelbein WK, Unger MA, Ownby DR: **Genetic structure and mtDNA diversity of *Fundulus heteroclitus* populations from polycyclic aromatic hydrocarbon-contaminated sites.** *Environ Toxicol Chem* 2003, **22**(3):671–677.
79. Williams LM, Oleksiak MF: **Evolutionary and functional analyses of cytochrome P4501A promoter polymorphisms in natural populations.** *Mol Ecol* 2011, **20**(24):5236–5247.
80. Cohen CS, Tirindelli J, Gomez-Chiarri M, Nacci D: **Functional implications of Major Histocompatibility (MH) variation using estuarine fish populations.** *Integr Comp Biol* 2006, **46**(6):1016–1029.
81. Whitehead A, Crawford DL: **Neutral and adaptive variation in gene expression.** *Proc Natl Acad Sci USA* 2006, **103**(14):5425–5430.
82. Pritchard JK, Di Rienzo A: **Adaptation - not by sweeps alone.** *Nat Rev Genet* 2010, **11**(10):665–667.
83. Hermisson J, Pennings PS: **Soft sweeps: molecular population genetics of adaptation from standing genetic variation.** *Genetics* 2005, **169**(4):2335–2352.
84. Barrett RD, Schluter D: **Adaptation from standing genetic variation.** *Trends Ecol Evol* 2008, **23**(1):38–44.
85. Pritchard JK, Pickrell JK, Coop G: **The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation.** *Cur Biol: CB* 2010, **20**(4):R208–215.
86. 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 Exper Zool A: Comp Exp Biol* 2006, **305A**(9):693–706.
87. Billiard SM, Timme-Laragy AR, Wassenberg 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**(2):526–536.
88. Waits ER, Nebert DW: **Genetic architecture of susceptibility to PCB126-induced developmental cardiotoxicity in zebrafish.** *Toxicol Sci* 2011, **122**(2):466–475.
89. Clark BW, Matson CW, Jung D, Di Giulio RT: **AHR2 mediates cardiac teratogenesis of polycyclic aromatic hydrocarbons and PCB-126 in Atlantic killifish (*Fundulus heteroclitus*).** *Aquat Toxicol* 2010, **99**:232–240.
90. Gerstein AC, Lo DS, Otto SP: **Parallel genetic changes and nonparallel gene-environment interactions characterize the evolution of drug resistance in yeast.** *Genetics* 2012, **192**(1):241–252.
91. Gompel N, Prud'homme B: **The causes of repeated genetic evolution.** *Dev Biol* 2009, **332**(1):36–47.
92. Karchner SI, Franks DG, Hahn ME: **AHR1B, a new functional aryl hydrocarbon receptor in zebrafish: tandem arrangement of *ahr1b* and *ahr2* genes.** *Biochem J* 2005, **392**:153–161.
93. Schmalz WF Jr, Hernandez AD, Weis P: **Hepatic histopathology in two populations of the mummichog, *Fundulus heteroclitus*.** *Mar Environ Res* 2002, **54**(3–5):539–542.
94. Zhou T, John-Alder HB, Weis P, Weis JS: **Thyroidal status of mummichog (*Fundulus heteroclitus*) from a polluted versus a reference habitat.** *Environ Toxicol Chem* 1999, **18**(12):2817–2823.
95. Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP: **A single amino acid mutation contributes to adaptive beach mouse color pattern.** *Science* 2006, **313**(5783):101–104.
96. Pelz HJ, Rost S, Hunerberg M, Fregin A, Heiberg AC, Baert K, MacNicol AD, Prescott CV, Walker AS, Oldenburg J, et al: **The genetic basis of resistance to anticoagulants in rodents.** *Genetics* 2005, **170**(4):1839–1847.
97. Stephens M, Donnelly P: **A comparison of bayesian methods for haplotype reconstruction from population genotype data.** *Amer J Hum Genet* 2003, **73**(6):1162–1169.
98. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data.** *Amer J Hum Genet* 2001, **68**(4):978–989.
99. Niu T, Qin ZS, Xu X, Liu JS: **Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms.** *Amer J Hum Genet* 2002, **70**(1):157–169.
100. Clement M, Posada D, Crandall KA: **TCS: a computer program to estimate gene genealogies.** *Mol Ecol* 2000, **9**(10):1657–1659.
101. Templeton AR, Crandall KA, Sing CF: **A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data: III: Cladogram estimation.** *Genetics* 1992, **132**(2):619–633.
102. Librado P, Rozas J: **DnaSP v5: A software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25**:1451–1452.
103. Excoffier L, Laval G, Schneider S: **Arlequin ver. 3.0: An integrated software package for population genetics data analysis.** *Evol Bioinform Online* 2005, **1**:47–50.
104. Guinand B, Lemaire C, Bonhomme F: **How to detect polymorphisms undergoing selection in marine fishes? A review of methods and case studies, including flatfishes.** *J Sea Res* 2004, **51**(4):167–182.
105. Yang Z: **PAML 4: a program package for phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24**:1586–1591.
106. Stamatakis A: **RAXML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**(21):2688–2690.
107. Posada D: **jModelTest: phylogenetic model averaging.** *Mol Biol Evol* 2008, **25**:1253–1256.
108. Yang Z, Wong WS, Nielsen R: **Bayes empirical bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22**(4):1107–1118.
109. Crooks GE, Hon G, Chandonia J-M, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14**(6):1188–1190.
110. Jensen J, Bohonak A, Kelley S: **Isolation by distance, web service.** *BMC Genet* 2005, **6**(1):13.

doi:10.1186/1471-2148-14-6

**Cite this article as:** Reitzel et al.: Genetic variation at aryl hydrocarbon receptor (AHR) loci in populations of Atlantic killifish (*Fundulus heteroclitus*) inhabiting polluted and reference habitats. *BMC Evolutionary Biology* 2014 **14**:6.