MOLECULAR ECOLOGY

Molecular Ecology (2011) doi: 10.1111/j.1365-294X.2011.05360.x

Evolutionary and functional analyses of cytochrome P4501A promoter polymorphisms in natural populations

LARISSA M. WILLIAMS* and MARJORIE F. OLEKSIAK†,‡

*Department of Environmental and Molecular Toxicology, PO Box 7633, North Carolina State University, Raleigh, NC 27695-7633, USA, † Department of Marine Biology and Fisheries, Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA

Abstract

The functional importance of variable, transcriptional regulatory sequences within and among natural populations is largely unexplored. We analysed the cytochrome P4501A (CYP1A) promoter in three populations of the minnow, *Fundulus heteroclitus*, because two SNPs in the promoter and first intron of CYP1A are under selection in populations adapted to pollutants. To define the importance of these SNPs, 1630 bp of the CYP1A promoter and first intron and exon were sequenced in eight individuals from three populations: a population from a polluted environment resistant to some aromatic pollutants and two flanking reference populations. CYP1A is induced by many aromatic pollutants, but in populations adapted to pollutants, CYP1A has been shown to be refractory to induction. We were interested in understanding whether variation in the CYP1A promoter explains mechanism(s) of adaptation to these aromatic pollutants. The CYP1A promoter was extremely variable (an average of 9.3% of the promoter nucleotides varied among all populations) and exhibited no fixed differences between populations. As CYP1A is poorly inducible in adapted fish, we hypothesized that CYP1A promoter regions might vary functionally between populations. Unexpectedly, *in vitro* analysis showed significantly greater transcription from CYP1A promoters found in the population from the polluted environment relative to promoters found in both reference populations. Thus, despite extensive variation among populations and lack of fixed differences between populations, individuals from a polluted environment have significantly enhanced promoter activity. These data demonstrate that intraspecific variation, which provides the raw material for natural selection to act on, can occur while maintaining promoter function.

Keywords: CYP1A, *Fundulus heteroclitus*, genetic variation, New Bedford Harbor, selection, superfund

Received 13 July 2011; revision received 30 September 2011; accepted 7 October 2011

Introduction

The estuarine minnow, *Fundulus heteroclitus*, inhabits several of the most heavily polluted estuaries in the world (Wirgin & Waldman 2004), including New Bedford Harbor, Massachusetts, USA, a Superfund site polluted with high levels of polychlorinated biphenyls (PCBs) (Weaver 1983). While this site has only been contaminated for approximately 100 years, *F. heteroclitus* has adapted to its seemingly inhospitable conditions (Nacci *et al.* 1999, 2010). Recently, two SNPs in the proximal promoter of the hepatic monooxygenase, cytochrome P4501A (CYP1A) were shown to be under selection in the New Bedford Harbor population as compared to clean, reference populations (Williams & Oleksiak 2011) indicating a genetic underpinning for adaptation. CYP1A transcription is induced by some polychlorinated biphenyls and some halogenated and nonhalogenated aromatic hydrocarbons via the aryl

Correspondence: Marjorie F. Oleksiak, Fax: (305) 421 4600; E-mail: moleksiak@rsmas.miami.edu

‡Present address: Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA

hydrocarbon receptor (AHR) pathway (Whitlock *et al.* 1996). Ligand-activated AHRs bind to well-conserved xenobiotic response elements (XREs) upstream of the transcriptional start site in promoters to enhance gene transcription. Yet, *F. heteroclitus* individuals from New Bedford Harbor are refractory to CYP1A induction by prototypic inducers (Nacci *et al.* 1999; Bello *et al.* 2001) and have lower CYP1A enzyme activity as compared to reference populations (Bello 1999). The basis for this refractory phenotype has not been characterized, but could involve the SNPs under selection either directly or through linkage to areas responsible for modulating CYP1A transcription.

The mechanism of differential sensitivity to pollutants has been proposed to involve the AHR signalling pathway, whose activation *via* pollutant binding is signalled by CYP1A induction. The refractory nature of CYP1A induction in the New Bedford population supports the linkage between differential sensitivity and the AHR pathway, although the direct role of CYP1A in toxicity and/or tolerance is unclear (Hahn *et al.* 2004). Efforts have been made to explore the adaptive role of AHR and other AHR pathway components. AHR (1 and 2), aryl hydrocarbon receptor nuclear translocator (ARNT), and AHR repressor (AHRR) gene expression patterns do not differ between New Bedford Harbor and reference populations, nor does the methylation status of the AHR promoter (Powell *et al.* 2000; Karchner *et al.* 2002; Aluru *et al.* 2011). While the AHR1 and AHR2 loci are polymorphic and have different allele frequencies in New Bedford Harbor and reference populations that may have been driven by selective pressures related to polluted environments (Hahn *et al.* 2002, 2004), no variants have been shown to be functionally important *in vitro* (Hahn *et al.* 2004).

However, two polymorphisms in or near the CYP1A promoter are under selection or linked to areas of the genome under selection in *F. heteroclitus* from New Bedford Harbor (Williams & Oleksiak 2011). One selectively important SNP is located at −670 bp upstream of the +1 transcriptional start site and the other at 173 bp into the first intron of the CYP1A gene. These selectively important SNPs may alter transcription factor binding or be linked to areas affecting transcription and may be responsible for the refractory induction of CYP1A in the New Bedford Harbor population.

To further explore the CYP1A promoter and the role of these SNPs in resistance to anthropogenic contaminant mixtures, we sequenced 1630 bp of the promoter and the first (noncoding) exon and intron in multiple individuals from New Bedford Harbor and reference populations to assess promoter-wide patterns of variation and to conduct evolutionary analyses among and between populations. To assess the functional behaviour of the promoter, we measured promoter activity and inducibility in a fish cell line. The null hypothesis was that the variation in the promoter would not lead to functional differences in promoter activity between populations. Alternatively, promoters from the different populations could exhibit differences in inducible activity, perhaps attributable to nucleotide differences. Thus, we tested the inducibility of CYP1A promoters from 12 individuals to a prototypic PAH using a transient luciferase transfection assay in a fish cell line to determine CYP1A promoter activities within and between populations. Based on published *in vivo* studies, we expected that induction from the CYP1A promoters of the resistant New Bedford Harbor population would be lower than that from promoters of the reference populations.

Materials and methods

Genomic clones

Fundulus heteroclitus were collected using minnow traps during the spring of 2005. Fin clips were sampled from eight individuals from each of the three collection sites along the Atlantic Coast of the United States since we were only interested in the underlying genomic information and not epigenetic factors such as methylation (Fig. 1). One of the three collection sites was an Environmental Protection Agency (EPA) Superfund site: New

Fig. 1 *Fundulus heteroclitus* collection sites. Superfund site (New Bedford Harbor, MA, USA) is denoted by a star, and north and south clean, reference sites are denoted by circles.

Bedford Harbor (EPA ID: MAD980731335) and the other two represent clean reference sites that flank the polluted site equidistant north and south, Sandwich, MA (41° 44′ N and 70° 23′) and Point Judith, RI (41° 21.7′ N and 71° 28.9′ W). These sites differ in their sensitivities to PCBs: fish from reference sites have a LC20 in the range of 20 ng PCB126/L and fish living in the polluted New Bedford Harbor site have a LC20 of 42 845 ng PCB126/L (Nacci *et al.* 2010). *Fundulus grandis*, a sister species to *F. heteroclitus* (Gonzalez *et al.* 2009), was collected from Port O'Connor, TX (28° 26′ N and 96° 24′ W) in the spring of 2008 and was used as an outgroup.

Genomic DNA from fin clips was extracted using a modified version of Aljanabi & Martinez (1997), and DNA was resuspended in 50 μL 0.1X TE buffer. 2 kb of the CYP1A promoter and first intron and two exons were amplified using PCR from the genomic DNA of each individual. The forward primer (5′-AGTTAT AGCCACAGTCCAGTCATTT-3′) was located 1575 bp upstream of the transcriptional start site and the reverse primer (5′-CAAGGCTATCAAACCCTCAGACAC-3′) was 526 bp downstream of the transcriptional start site in the second exon (Powell *et al.* 2004). PCR amplification products were ligated into a Promega pGEM-T vector overnight at 16 °C with T4 DNA ligase and transformed into electrocompetant JM109 *Escherichia coli* cells, which were grown on carbenicillin selective plates. Colonies were screened for insert with PCR using vector specific primers. Insert PCR products from three clones per individual were sequenced in both directions using ABI BigDye terminator chemistry providing 6× coverage of each nucleotide. Sequences were aligned in MacVector 8.0 (Olson 1994) and CAP3 (Huang & Madan 1999).

Sequence analysis

CYP1A sequences from each of eight individuals in New Bedford Harbor, Point Judith and Sandwich and one *Fundulus grandis* individual were analysed with DnaSP (Librado & Rozas 2009) to determine patterns of sequence variation along the promoter as well as in the first and second intron and first exon of the CYP1A gene. For Tajima's *D* and Fu and Li's *D*, the significance of *D* was evaluated with 10 000 coalescent simulations. Linkage disequilibrium (LD), G_{ST} and F_{ST} were extracted from DnaSP. For linkage disequilibrium, only polymorphic nucleotides were considered in the analysis. The degree of LD was estimated by the following parameters: *D* (Lewontin & Kojima 1960), *D*′ (Lewontin 1964), *R* and R^2 (Hill & Robertson 1968). Both the two-tailed Fisher's exact test and the chi-square test are computed to determine whether the associations between polymorphic sites are significant. DnaSP also performs the Bonferroni correction for multiple tests. F_{ST} ($F_{ST} = (H_T - H_S)/H_T$) was calculated in DnaSP using the equations of Lynch & Crease (1990) on a per locus basis (Lynch & Crease 1990). For G_{ST}, as the data are unphased, DnaSP first reconstructs the haplotype phases. This haplotype reconstruction is conducted using the algorithms provided by PHASE (Stephens *et al.* 2001; Stephens & Donnelly 2003), fastPHASE (Scheet & Stephens 2006) and HAPAR (Wang & Xu 2003). After reconstructing haplotypes, G_{ST} is calculated using the method of Hudson *et al.* (1992) where a permutation-based test is performed to detect genetic differentiation of subpopulations at different localities (Hudson *et al.* 1992). G_{ST} is defined $1-(H_S/H_T)$ where H_S is a weighted average of estimated haplotype diversities in the subpopulations, and H_T is an estimate of the haplotype diversity of the total population.

Functional sites were defined *in silico* using AliBaba 2 (Grabe 2002), a program which predicts transcription factor binding sites using TRANSFAC 7.0 (Wingender *et al.* 1996).

Reporter gene constructs

CYP1A was inserted into the pGL3-Basic vector upstream of the luciferase gene using restriction endonucleases. Specifically, CYP1A genomic DNA, originally cloned into pGEM-T vector, was digested with Apa I and the 5′overhang filled in with T4 DNA polymerase; CYP1A was then released from the pGEM-T vector by digestion with Sac I. Promega pGL3-Basic vector containing the Firefly luciferase gene was linearized by digesting with Kpn I and the 3′ overhang filled in with T4 DNA polymerase. This blunting was followed by digestion with Sac I. Digested insert DNA and pGL3-Basic vector were gel purified using Zymoclean Gel DNA recovery kit and ligated together overnight at 16 °C. Ligation product was transformed into electrocompetant JM109 *Escherichia coli* cells, and plated on carbenicillin selective plates. Colonies were screened for insert with PCR using vector specific primers.

Cell culture and transfection

Fish hepatoma cells PLHC-1 (*Poeciliopsis lucida* hepatoma cell) (Ryan & Hightower 1994) were grown in Gibco CO_2 independent media supplemented with 5% FBS and 50 U/mL of penicillin G/50 μg/mL streptomycin in a 30 °C incubator. PLHC-1 cells function similarly to *Fundulus heteroclitus* liver in that they have intact, catalytically active and AHR-mediated inducible CYP1A protein (Hahn *et al.* 1993). Cells were usually split every 4 days by dissociating with 0.05% (w/v) trypsin and 0.5 mM EDTA and subcultured at 2×10^7 cells per each 75 cm^2 tissue culture flask. Using Fugene 6

transfection reagent (Boehringer), cells at 50–80% confluence were transiently co-transfected in 12-well plates at a density of 2×10^6 cells per well with the reporter gene vector and an internal control vector expressing *Renilla* luciferase under the control of a cytomegalovirus promoter (pRL-CMV) (Promega). Cells were treated with 3-methylcholanthrene (3-MC) dissolved in DMSO, or with DMSO only. 3-MC is a prototypic inducer of CYP1A through the AHR pathway (Meyer *et al.* 2002). Although PCBs are the pollutants most represented in New Bedford Harbor, PCB-resistant New Bedford Harbor fish also are resistant to PAHs (specifically 3-MC), showing the same refractory phenotype (Nacci *et al.* 1999), and we treated cells with this PAH due to its higher toxicity. Luciferase and *Renilla* activities were measured using the Promega Dual-Luciferase Reporter Assay System.

Results

Proximal and first exon and intron sequences of CYP1A

A total of 2013 nucleotides (Fig. S1, Supporting Information; 1630 in the promoter (5′ to the start of transcription) and 383 from the first exon and intron) were sequenced from eight individuals from each of three populations (Sandwich, New Bedford Harbor, and Point Judith) and one *Fundulus grandis* individual (sister species). SNPs (927 and 1890 bp) were found to be under selection in a previous study (Williams & Oleksiak 2011). Among the 1630 bp of promoter sequence are 16 putative transcription factor binding sites that make up 9.2% of the promoter sequence (150 bp) including an intact TATA box at 1594 bp. Among the 16 binding sites, only the three xenobiotic response elements (XRE, at 799, 848 and 1403 bp) have been defined experimentally by their ability to bind AHR *in vitro* (Powell *et al.* 2004).

For the 1630 bp of the promoter, 10.3% of the nucleotides are variable among all individuals in both species. Within *Fundulus heteroclitus*, 9.3% of promoter sites are variable, with 3.7, 6.3, and 0.4% within Sandwich, New Bedford Harbor, and Point Judith individuals, respectively. The promoter has 106 parsimony informative sites (shared derived), and 22 fixed differences between *F. heteroclitus* and its sister species *F. grandis* (Fig. S1, Supporting Information). Two of the three xenobiotic response elements (XRE2 and 3 (Powell *et al.* 2004)) had three polymorphisms across populations and species (Fig. S1, Supporting Information). XRE3 had two polymorphisms present only in the New Bedford Harbor population at nucleotides 1404 and 1405, where New Bedford Harbor individuals NBH6 and NBH13 have a TT rather than an AC. The average nucleotide diversity, defined as the average number of nucleotide differences per site between two sequences, for the proximal promoter for *F. heteroclitus* among all three populations is 0.0150 (excluding gaps; $N = 1538$ bp). Heterozygosity or theta (θ) is 0.027 where θ is defined as $4Nu$, N is the effective population size and u is the mutation rate per nucleotide (or per sequence) and per generation, per site (Jukes & Cantor 1969). Per population, the average nucleotide diversity is 0.012 for Sandwich ($N = 1528$ bp), 0.021 for New Bedford Harbor ($N = 1538$ bp), and 0.0016 for Point Judith ($N = 1529$ bp).

The first exon and intron (383 bp) also were sequenced, as one of the SNPs under selection (Williams & Oleksiak 2011) was located in the first intron (at nucleotide 1890). For the 383 bp of the first intron and exon, 11% of the nucleotides are variable among all individuals in both species. Within *F. heteroclitus* 10.2% of sites are variable, with 5.2, 7.3 and 0.5% within Sandwich, New Bedford Harbor and Point Judith individuals, respectively. The first exon contained two informative sites, and the first intron contained 24 informative sites. All polymorphisms were synonymous substitutions and thus did not alter the amino acid sequence. The average nucleotide diversity for the first exon was 0.018 for Sandwich ($N = 111$ bp) and 0.009 for New Bedford Harbor ($N = 111$ bp); Point Judith contained no polymorphisms. The average nucleotide diversity for the first intron was 0.017 for Sandwich ($N = 268$), 0.040 for New Bedford Harbor ($N = 268$), and 0.005 for Point Judith ($N = 267$).

Based on the sequence variation, populations are distinct with significant genetic distances (Table 1). Between the two most geographically proximal sites, Sandwich and New Bedford Harbor, the G_{ST} was 0.028 and F_{ST} was 0.054. Between New Bedford Harbor and

Table 1 Measures of population divergence between *Fundulus heteroclitus* populations

Population 1	Population 2	K_{xy}	G_{ST}	F_{ST}
Sandwich	New Bedford Harbor	42.11	0.028	0.054
Sandwich	Point Judith	41.14	0.040	0.410
New Bedford Harbor	Point Judith	46.09	0.037	0.405

$K_{xy} = \pi D$ (average number of pairwise nucleotide differences). G_{ST} is Nei's coefficient of gene variation and defined as $1-(H_S/H_T)$ where H_S is a weighted average of estimated haplotype diversities in the subpopulations, and H_T is an estimate of the haplotype diversity of the total population. F_{ST} is Wright's inbreeding coefficient and is defined as $(H_T-H_S)/H_T$.

Point Judith, the G_{ST} was 0.037 and F_{ST} was 0.405. Sandwich and Point Judith, the two furthest sampling sites from each other, had a G_{ST} of 0.040 and a F_{ST} of 0.410. The disparity in the population differentiation calculations (G_{ST} and F_{ST}) is a result of the way in which each is calculated. For G_{ST}, haplotypes were reconstructed using several different methods. However, almost every single individual had a different haplotype. The basis for the differences between the values may lie in the fact that G_{ST} is examining haplotypes and F_{ST} is examining differentiation between populations on a per locus basis. Thus, the haplotypes are unable to accurately differentiate the populations, but the locus by locus F_{ST} method is able to detect this differentiation, a measure that was previously reported using AFLPs (Williams & Oleksiak 2008). The average number of pairwise nucleotide differences (K_{xy}) were similar in all three pairwise comparisons, with 42.11 for Sandwich vs. New Bedford Harbor, 41.14 for Sandwich vs. Point Judith, and 46.09 for New Bedford Harbor vs. Point Judith (Table 1).

Linkage disequilibrium (LD) was compared for all polymorphic nucleotides along the promoter and the first exon and intron. For Sandwich, 558 of the total 12 483 pairwise comparisons of polymorphic loci (4.5%) were in significant LD (Fig. S2, Supporting Information). New Bedford Harbor had 2175 significant pairwise comparisons in LD, of a total of 33 871 pairwise combinations (6.4%) (Fig. S3, Supporting Information). Point Judith had the fewest significant pairwise comparisons (158) but had the highest percentage of significant comparisons (46.76%) among the 338 pairwise comparisons (Fig. S4, Supporting Information). Sandwich sequences showed strong LD among nucleotides in the 5′ end of the promoter. In New Bedford Harbor, the sequence 'ACTT' spanning nucleotides 1972–1975 at the 3′ end of the first intron was in LD with a large portion (40%) of the other polymorphic nucleotides within the promoter, first exon and first intron. This 'ACTT' sequence was also present in a Sandwich individual as well as all Point Judith individuals, but not in strong LD with other polymorphic nucleotides within the promoter. Point Judith sequences also have a large block of LD in the 5′ end of the promoter where an ancestral insertion occurs. There is no LD among the SNPs (927 and 1890 bp) found to be under selection in the previous study (Williams & Oleksiak 2011) with the remainder of the promoter, intron one or exon one. There was no significant LD among any known or putative transcription factor binding sites with the exception of XRE3 in the New Bedford Harbor population. XRE3 (1403–1408) had significant LD to nucleotides 1334–1359, which are just upstream of the XRE3 binding site.

Evolutionary analyses of promoters

The CYP1A proximal promoter sequence variation between *Fundulus heteroclitus* populations can be visualized in sliding window comparisons (Fig. 2). In the comparison of Sandwich to New Bedford Harbor, similar areas of the promoter are equally variable between populations except between nucleotides 1551–1617 (upstream of the TATA box at 1594 bp), where there is a difference between the nucleotide substitutions per site between the two populations. In this area, there are

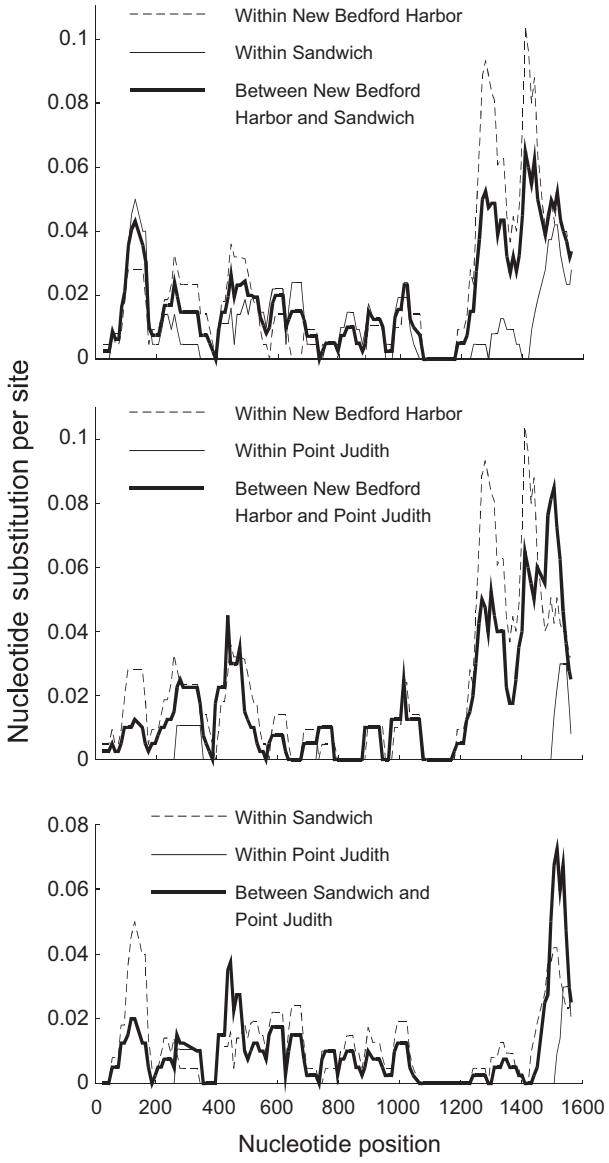

Fig. 2 Sliding window comparisons of average nucleotide substitutions per site within *Fundulus heteroclitus* populations (Pi1 and Pi2) and between populations (D_{xy}). Plots for the sliding window use a 50 bp-wide window and 10 bp step size.

several polymorphisms that only exist in one individual in one population, increasing the Dxy value. In the comparison of New Bedford Harbor to Point Judith, Dxy exceeds the nucleotide substitution per site among populations in the 3′ end of the promoter spanning nucleotides 1464–1553. In this area, Point Judith individuals do not have any sequence variation, and there is moderate variation in three individuals (4, 6, 13) in the New Bedford Harbor population. When clean, reference sites Sandwich and Point Judith are compared, the same area that had a large Dxy value in the Point Judith vs. New Bedford Harbor comparison (nucleotides 1464–1553) exhibits the same pattern due to the lack of sequence variation in the Point Judith population and moderate variation in the Sandwich population. There are no clear patterns of variation near predicted functional regions of the promoter between New Bedford Harbor and its clean, reference sites.

A similar lack of a clear pattern in variation along the CYP1A promoter can be visualized by plotting Tajima's *D* and Fu and Li's *D* and *F* statistics by population (Fig. 3). The Tajima test for data from a single locus (Tajima 1989) compares the estimate of theta (θ) based on the number of segregating sites to that based on the average number of pairwise differences. Tajima's *D* was −0.202 for Sandwich, 0.034 for New Bedford Harbor and 0.496 for Point Judith populations. None of these values were significant at a *P*-value of 0.01. The test of Fu and Li's *D* (Fu & Li 1993) with an outgroup (*F. grandis* in this case) compares estimates of the number of segregating sites to the number of mutations on external branches expected under neutrality and was 2.10 for Sandwich ($P < 0.02$), 2.04 for New Bedford Harbor ($P < 0.02$) and 0.96 for Point Judith ($P > 0.10$) populations. The Fu & Li's *F* test statistic (Fu & Li 1993) is based on the differences between the total number of mutations in external branches of the genealogy and the average number of nucleotide differences between pairs of sequences. The *F*-test statistic is 0.83 for Sandwich, 0.97 for New Bedford Harbor and 0.23 for Point Judith populations, and none of these values were significant at a *P*-value of 0.01. Sandwich, New Bedford Harbor and Point Judith sequences have generally positive Tajima's *D* and Fu and Li's *F* statistics along the promoter signifying low levels of both low and high frequency polymorphisms and indicating a decrease in population size and/or balancing selection. Fu and Li's *D* statistic was significantly positive (at a *P*-value of 0.01) in the New Bedford Harbor population at the sliding windows of nucleotide regions 1249–1340 bp. This region contains a predicted Oct2 transcription factor binding site, an enhancer that has been found to regulate immunoglobulin heavy-chain (IgH) genes in other teleosts (Cioffi *et al.* 2002).

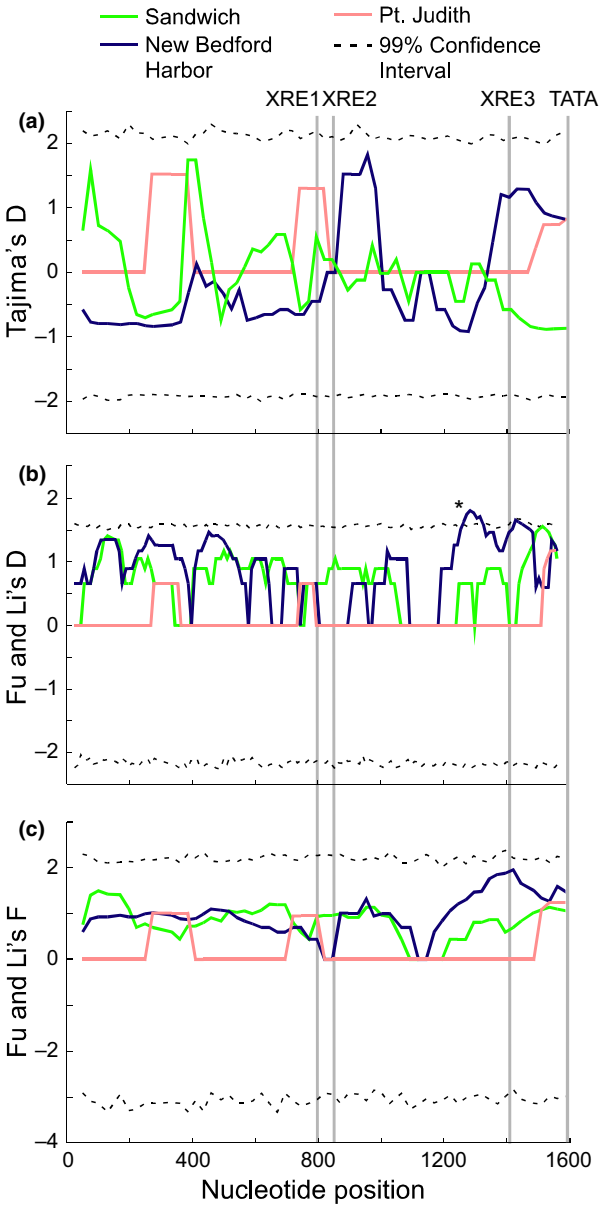

Fig. 3 Sliding window plots of Tajima's D, Fu and Li's *D* and Fu and Li's F. (a) Tajima's *D* calculations, (b) Fu and Li's *D* calculations, and (c) Fu and Li's *F* calculations using a sliding window with a 50 bp-wide window and a 10 bp step size for each *Fundulus heteroclitus* population. Solid and dashed lines represent the estimated values (*D* or *F*) and 99% confidence interval, respectively, obtained through 10 000 coalescent simulations. Green line: Sandwich population. Blue line: New Bedford Harbor population. Red line: Pt. Judith population. Asterisk (*) indicates the significant deviations of *D*.

Functional assays

To assess whether luciferase expression increased in a dose-dependent manner in cells transfected with the CYP1A proximal promoter, induction of luciferase

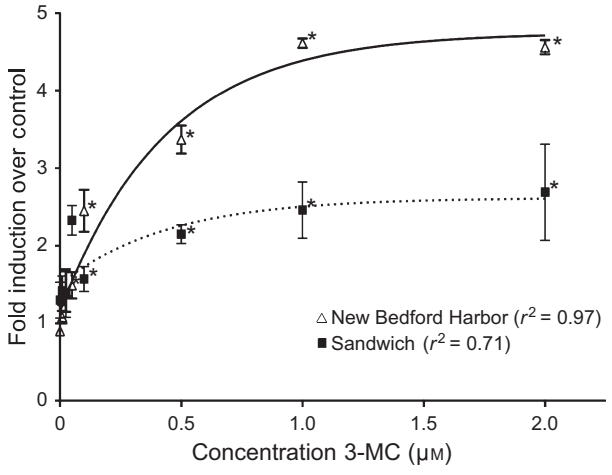

Fig. 4 Induction of luciferase activity from *Fundulus heteroclitus* CYP1A reporter gene constructs from a New Bedford Harbor and Sandwich individual. CYP1A promoter strength was assayed in PLHC-1 cells treated with 3-MC. Asterisks (*) indicate a significant fold induction over control ($P < 0.05$).

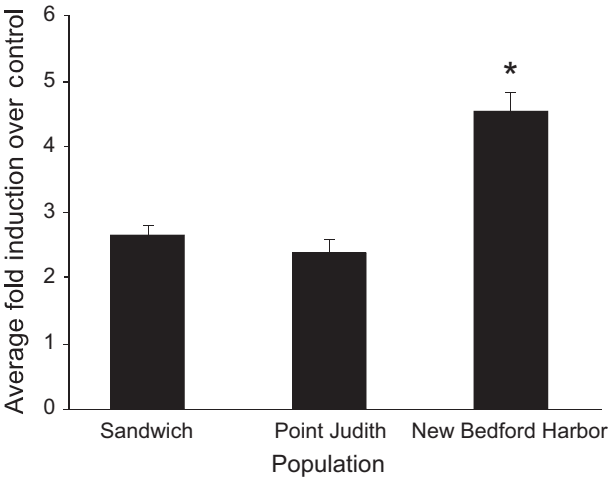

Fig. 5 Average fold induction of the CYP1A promoter over control for Sandwich, New Bedford Harbor and Point Judith populations. $N = 4$/population at a 1 μM concentration of 3-MC. A one-way ANOVA found a significant difference in the average fold induction between populations ($P < 0.0001$) and a Tukey's HSD *post hoc* test determined that New Bedford Harbor was significantly different from Sandwich ($P < 0.0001$) and Point Judith ($P < 0.0001$), but Sandwich and Point Judith were not significantly different from each other ($P = 0.258$). Significant induction in the New Bedford Harbor population compared to the reference populations is denoted by an asterisk (*).

expression was measured in PLHC-1 cells transfected with CYP1A proximal promoter reporter gene constructs from a clean (Sandwich individual S27) and polluted (New Bedford Harbor individual NBH6) individual and treated with 3-MC (Fig. 4). The promoter construct was 6831 bp in length, of which 2011 bp was CYP1A promoter, first intron and first exon, with the start of transcription at 1628 bp (Fig. S1, Supporting Information). There was an intact TATA box at 1594 bp. Upstream of the transcriptional start site, there were three XRE sites at 799, 848 and 1403 bp.

Transiently transfected cells were treated with 3-MC or treated with vehicle control (DMSO) as described in Williams *et al.* (Williams *et al.* 2000) at concentrations ranging from 0.001 to 2 μM. At lower concentrations (<0.05 μM), luciferase expression was similar for both constructs and not significantly different from vehicle-treated control (Fig. 4). At 0.05 μM 3-MC, the New Bedford Harbor construct had significant induction over control. At concentrations of 0.1 μM and above, both Sandwich and New Bedford Harbor constructs had significant induction over control. The New Bedford Harbor construct also had a significantly higher ($P \leq 0.05$) fold induction in comparison to the Sandwich construct, ranging from 1.5 to 1.9 times greater. A horizontal asymptote occurs at the 1 μM concentration, indicating that maximal induction of the promoter occurred at or around that concentration.

To determine if there are individual and population differences between the inducibility of CYP1A promoters, luciferase activities from four individual promoter constructs were compared within and between the three populations (Fig. 5) at a concentration of 1 μM 3-MC (maximal induction). The pattern observed was very similar to that of the dose–response curve: promoter constructs from reference populations Sandwich and Point Judith had on average a fold induction over control of 2.66 and 2.39, respectively, which was significantly lower than the New Bedford Harbor population promoter constructs, with an average fold induction over control of 4.54. Within population average fold induction over control variability was minimal, ranging from 0.14 to 0.30. A one-way ANOVA determined a significant difference between populations ($P < 0.0001$), and a Tukey's HSD *post hoc* test determined that New Bedford Harbor promoter constructs were significantly different from Sandwich ($P < 0.0001$) and Point Judith promoter constructs ($P < 0.0001$), but Sandwich and Point Judith promoter constructs were not significantly different from each other ($P = 0.258$).

Discussion

This study sought to determine variation within the CYP1A promoter between populations from clean and polluted environments and establish whether the promoter plays a functional role in the CYP1A phenotype

that has been described in the New Bedford Harbor population: in the New Bedford Harbor population CYP1A is refractory to prototypic CYP1A inducers (Nacci *et al.* 1999; Bello *et al.* 2001).

The CYP1A proximal promoter, cloned from eight individuals from each of three populations, is extremely variable: up to 9.3% of the nucleotides in the promoter vary among *Fundulus heteroclitus* individuals. This level of variation is at least fourfold larger than variation found in promoters and enhancers: CCR5 5′ *cis*-regulatory region (2.1%; (Bamshad *et al.* 2002) and G protein-coupled receptor kinase 4 promoter [0.94%;(Hasenkamp *et al.* 2008)] in humans; the first intron of *Adh* [(1.7%; (Kreitman 1983)] and the *eve* enhancer [0.87%; (Ludwig & Kreitman 1995)] in *Drosophila melanogaster*; the chalcone synthetase promoter in *Arabidopsis thaliana* (1%; (de Meaux *et al.* 2005)). This level of variation is similar to the promoter variation in other loci that are evolving by natural selection: the *Ldh-B* proximal promoter in *F. heteroclitus* (6.1%; (Crawford *et al.* 1999); and the *endo16* promoter in the purple sea urchin (10%; (Balhoff & Wray 2005). The percentage of polymorphisms in predicted transcription factor binding sites (8.8% or 13 polymorphisms in 148 bp of *in silico* predicted functional nucleotides among all individuals and populations) is similar to the percentage of polymorphisms found within the promoter in the putative nonfunctional nucleotides (9.4%, nucleotides that do not bind transcription factors *in silico*). Yet, even though functional binding sites are not well characterized in the CYP1A promoter, it is difficult to explain this level of polymorphism. These sequence data alone suggest the evolutionary maintenance of polymorphisms thus suggesting a heterozygote advantage. Analysis of SNPs in the CYP1A promoter (Williams & Oleksiak 2011) suggest that natural selection is responsible for the increased frequency in the population from the polluted environment. Thus, in combination with earlier work, these data strongly suggest that natural selection may be acting to maintain CYP1A proximal promoter polymorphisms.

Fundulus heteroclitus is a nonmigratory species (Skinner *et al.* 2005), living in large subpopulations (Adams *et al.* 2006). These large subpopulations maintain high-standing genetic variation (Oleksiak 2010), allowing adaptation through natural selection to changing environmental conditions. Plots of variation between populations (Fig. 2) reveal that each population surveyed had a substantial and similar amount of variation. It is likely that this variation is entirely neutral. However, it also has been hypothesized that stabilizing selection on transcriptional output allows slightly deleterious mutations to persist, compensated for by adaptive changes elsewhere in the promoter and resulting in continuous binding-site turnover (Ludwig *et al.* 2000; Balhoff & Wray 2005). This alternative hypothesis is supported for Sandwich and New Bedford populations by the Fu and Li's *D* test for selection, which found significantly positive values for the promoter. Significantly positive values for this statistic reflect an excess of intermediate-frequency alleles, which can result from population bottlenecks, structure and/or balancing selection. Conversely, Tajima's *D* and Fu and Li's *F* for each population were not significant. Tajima's *D* has been found to be more powerful than Fu and Li's *D* test to test for very strong, recent directional selection (Braverman *et al.* 1995; Simonsen *et al.* 1995). However, it has been shown that Tajima's *D* and Fu and Li's *D* test have very low power to detect balancing selection (Charlesworth *et al.* 1995). Fu and Li's *F*-test (Fu & Li 1993), which has a slightly higher power to detect balancing selection, found no significant patterns of variation along the CYP1A promoter in any population. As only one of the three tests for selection was significant, the evidence for balancing selection acting on the CYP1A promoter is not overwhelming. Signatures of selection may be more obvious in functional sequences. However, because the majority of the polymorphisms in putative functional sequences were in one or a few individuals across populations, there was no significant selection signal on those areas in Tajima's *D* or Fu and Li's *D* and *F* statistics; significant values for these tests would require shared polymorphisms between individuals within populations. In addition, if balancing selection were maintaining those sites, one would also expect to find high linkage disequilibrium between the nucleotides of each site and other functional sites (Wall 1999; Charlesworth 2006); this was not the case. Given these data, the most parsimonious explanation for the unusually large nucleotide variation in the CYP1A promoter is that it is most likely maintained by neutral processes including large population sizes and migration. However, the power of these tests to detect selection is relatively weak (Wayne & Simonsen 1998), especially given the lack of fixed differences in the population from the polluted site and the extensive sequence variation among individuals.

The evolutionary analyses of nucleotide variation suggest that the most parsimonious explanation for the promoter variation among populations is that it is due to neutral processes. This differs from SNP analyses that demonstrated that two SNP in the CYP1A promoter and first intron were evolving by natural selection (Williams & Oleksiak 2011) and may reflect a lack of power of the evolutionary analyses of nucleotide variation. In addition, we show here a functional difference in promoter activity: among all four individuals from New Bedford Harbor, there is greater induction in

response to a prototypic PAH (Fig. 5). The observation that these individuals have better induction (vs. constitutive expression) suggests that the observed differences are not due to experimental artefact. That is, the fact that the constitutive expression was similar among individuals and the expression of reporter genes linked to New Bedford Harbor promoters showed a statistically significant increase indicates that these promoter sequences are responsible. Yet there is a conundrum: there is no one nucleotide or set of nucleotides shared by New Bedford Harbor individuals that explains this significantly greater induction. There was a polymorphism in the XRE3, where the XRE site was ''C**TT**GCGA'' rather than the consensus sequence ''C**AC**GCGA'' in New Bedford Harbor individuals 6 and 13. However, there is no difference among New Bedford Harbor individuals with or without this polymorphism in AHR-ARNT binding sites. While the XREs described in the CYP1A promoter do bind the AHR-ARNT complex *in vitro* (Powell *et al.* 2004), the functional significance of each of these XREs is still unclear with respect to inducing CYP1A transcription.

There is a second complexity in the data: when whole organisms are exposed to PAHs, CYP1A transcription has little or lower induction in New Bedford Harbor fish (Nacci *et al.* 1999; Bello *et al.* 2001). Thus, it was expected that the CYP1A promoter for this population would be less inducible by prototypic PAHs as compared to populations without environmental contamination histories, yet the opposite was found. Two immediate explanations for this discrepancy between *in vivo* and *in vitro* results are that (i) the regulatory sequence responsible for the *in vivo* repression of transcriptional phenotype lies further upstream or downstream of the CYP1A promoter region that was sequenced (1630 bp) or (ii) regulation involves proteins (*trans*- acting factors), which were not present in the cell line in which the transfection assays were completed. Either explanation needs to address the observation that the proximal promoter (approx, 1 kb 5′ to start site) from the polluted population effect a greater induction and has a SNP that appears to be evoluting by natural selection. Epigentic factors such as methylation status or microRNA-based regulation also are possible. However, no differences in CYP1A methylation were found in two other Superfund populations (Elizabeth River, VA and Newark Bay, NJ, USA) sharing the same refractory phenotype (Arzuaga *et al.* 2004; Timme-Laragy *et al.* 2005), and to date, there is little known about microRNA regulation in *Fundulus heteroclitus*. Compound specific differences in induction, which have been observed in other resistant fish populations (Courtenay *et al.* 1999), might be found if a PCB were used rather than a PAH. However, *in vivo*, New Bedford Harbor fish are refractory to CYP1A induction by PAHs as well as PCBs (Nacci *et al.* 1999; Bello *et al.* 2001). Regardless of the discrepancy between *in vivo* and *in vitro* results and despite extremely high levels of sequence variation in the CYP1A promoter within and between populations, we found significant differences in CYP1A promoter strength between the polluted Superfund population (NBH) and two flanking reference populations. These data and SNP analyses suggest that the simplest explanation of the sequence variation (that the promoter variation among populations is due to neutral processes) is incorrect. Instead, we would suggest that a selectively important change in the NBH population is effecting a change in CYP1A expression.

We can only speculate what the molecular mechanism is that would explain the *in vivo, in vitro* and evolutionary results. Because the SNP nucleotide variation in the proximal promoter is adaptively important and has a functional effect (although the opposite of what is found *in vivo*), our data suggest that the transcription factor(s) binding to the proximal promoter *in vivo* is different than that found in cell culture. Alternatively, there could be an upstream element that interacts with the proximal promoter to alter the enhancer effect of this site from stimulatory to repressive. Resolving this problem would require defining all the binding sites that are affected by PCBs or PAHs and the transcription factors that bind these sites.

Conclusion

The CYP1A promoter is extremely variable in both predicted functional (8.8% variation) and nonfunctional sequence (9.4% variation). Patterns of variation differ among the three tested populations; however all are equally variable. One experimentally verified functional site, XRE3, which differed in two New Bedford Harbor individuals, did not affect the inducibility of the CYP1A promoter *in vitro* to a prototypic inducer. Despite the extreme variability of the CYP1A promoter within and among populations, we found significant differences in promoter strength among the population from the polluted environment (New Bedford Harbor) and flanking reference populations, corroborating previous results (Williams & Oleksiak 2011), which identified selectively important SNPs in the CYP1A promoter of individuals from chronically polluted populations. The fact that our subsequent analyses of promoter sequences showed no signature of selection in promoters from individuals from the polluted environment likely reflects a lack of statistical power. In contrast to *in vivo* data, promoters from individuals from the polluted environment induced gene expression more strongly than promoters

from individuals from the reference sites. This result suggests that the large nucleotide divergence affects CYP1A expression, but other cis or trans acting factors must also exist to reproduce gene expression found in nature. These data demonstrate that intraspecific variation, which provides the raw material for natural selection to act on, can occur while maintaining promoter function.

Acknowledgements

Partial funding for this work was received from NIH 5 RO1 ES011588, ES007046 and NSF OCE 1008542. The Authors thank Douglas L. Crawford for helpful discussions and review of the manuscript and the University of Miami's Molecular Core for sequencing.

References

Adams SM, Lindmeier JB, Duvernell DD (2006) Microsatellite analysis of the phylogeography, pleistocene history and secondary contact hypotheses for the killifish, *Fundulus heteroclitus*. *Molecular Ecology*, **15**, 1109–1123.

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692–4693.

Aluru N, Karchner SI, Hahn ME (2011) Role of DNA methylation of AHR1 and AHR2 promoters in differential sensitivity to PCBs in Atlantic Killifish, *Fundulus heteroclitus*. *Aquatic Toxicology*, **101**, 288–294.

Arzuaga X, Calcano W, Elskus A (2004) The DNA de-methylating agent 5-azacytidine does not restore CYP1A induction in PCB resistant Newark Bay killifish (*Fundulus heteroclitus*). *Marine Environmental Research*, **58**, 517–520.

Balhoff JP, Wray GA (2005) Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8591–8596.

Bamshad MJ, Mummidi S, Gonzalez E *et al.* (2002) A strong signature of balancing selection in the 5′ cis-regulatory region of CCR5. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 10539–10544.

Bello SM (1999) *Characterization of Resistance to Halogenated Aromatic Hydrocarbons in a Population of* Fundulus heteroclitus *from a Marine Superfund Site*. Woods Hole Oceanographic Institution, Massachusetts Institute of Technology, Woods Hole, MA.

Bello SM, Franks DG, Stegeman JJ, Hahn ME (2001) Acquired resistance to Ah receptor agonists in a population of Atlantic killifish (*Fundulus heteroclitus*) inhabiting a marine superfund site: in vivo and in vitro studies on the inducibility of xenobiotic metabolizing enzymes. *Toxicological Sciences*, **60**, 77–91.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W (1995) The Hitchhiking effect on the site frequency-spectrum of DNA polymorphisms. *Genetics*, **140**, 783–796.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *Plos Genetics*, **2**, 379–384.

Charlesworth D, Charlesworth B, Morgan MT (1995) The pattern of neutral molecular variation under the background selection model. *Genetics*, **141**, 1619–1632.

Cioffi CC, Pollenz RS, Middleton DL *et al.* (2002) Oct2 transcription factor of a teleost fish: activation domains and function from an enhancer. *Archives of Biochemistry and Biophysics*, **404**, 55–61.

Courtenay S, Grunwald C, Kreamer G *et al.* (1999) A comparison of the dose and time response of CYP1A1 mRNA induction in chemically treated Atlantic tomcod from two populations. *Aquatic Toxicology*, **47**, 43–69.

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

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Gonzalez I, Levin M, Jermanus S, Watson B, Gilg M (2009) Genetic assessment of species ranges in *Fundulus heteroclitus* and *F. grandis* in Northeastern Florida Salt Marshes. *Southeastern Naturalist*, **8**, 227–243.

Grabe N (2002) AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biology*, **2**, S1–S15.

Hahn ME, Lamb TM, Schultz ME, Smolowitz RM, Stegeman JJ (1993) Cytochrome-P4501a induction and Inhibition by 3,3′,4,4′-tetrachlorobiphenyl in an Ah receptor-containing fish hepatoma-cell line (Plhc-1). *Aquatic Toxicology*, **26**, 185–208.

Hahn ME, Karchner SI, Franks DG (2002) Aryl hydrocarbon receptor polymorphisms and dioxin resistance in Atlantic killifish (*Fundulus heteroclitus*). *Marine Environmental Research*, **54**, 409.

Hahn ME, Karchner SI, Franks DG, Merson RR (2004) Aryl hydrocarbon receptor polymorphisms and dioxin resistance in Atlantic killifish (*Fundulus heteroclitus*). *Pharmacogenetics*, **14**, 131–143.

Hasenkamp S, Telgmann R, Staessen JA *et al.* (2008) Characterization and functional analyses of the human G protein-coupled receptor kinase 4 gene promoter. *Hypertension*, **52**, 737–746.

Hill W, Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, **38**, 226–231.

Huang XQ, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research*, **9**, 868–877.

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, **9**, 138–151.

Jukes TH, Cantor CR (1969) *Evolution of Protein Molecules*. Academic Press, New York.

Karchner SI, Franks DG, Powell WH, Hahn ME (2002) Regulatory interactions among three members of the vertebrate aryl hydrocarbon receptor family: AHR repressor, AHR1, and AHR2. *Journal of Biological Chemistry*, **277**, 6949–6959.

Kreitman M (1983) Nucleotide polymorphism at the alcohol-dehydrogenase locus of drosophila-melanogaster. *Nature*, **304**, 412–417.

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, **49**, 49–67.

Lewontin RC, Kojima K (1960) The evolutionary dynamics of complex polymorphisms. *Evolution*, **14**, 458–472.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Ludwig MZ, Kreitman M (1995) Evolutionary dynamics of the enhancer region of even-skipped in drosophila. *Molecular Biology and Evolution*, **12**, 1002–1011.

Ludwig MZ, Bergman C, Patel NH, Kreitman M (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature*, **403**, 564–567.

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution*, **7**, 377–394.

de Meaux J, Goebel U, Pop A, Mitchell-Olds T (2005) Allele-specific assay reveals functional variation in the chalcone synthase promoter of Arabidiopsis thaliana that is compatible with neutral evolution. *Plant Cell*, **17**, 676–690.

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

Nacci D, Coiro L, Champlin D *et al.* (1999) Adaptations of wild populations of the estuarine fish *Fundulus heteroclitus* to persistent environmental contaminants. *Marine Biology*, **134**, 9–17.

Nacci DE, Champlin D, Jayaraman S (2010) Adaptation of the Estuarine Fish *Fundulus heteroclitus* (Atlantic Killifish) to polychlorinated biphenyls (PCBs). *Estuaries and Coasts*, **33**, 853–864.

Oleksiak MF (2010) Genomic approaches with natural fish populations. *Journal of Fish Biology*, **76**, 1067–1093.

Olson SA (1994) MacVector: Sequence Comparisons Using a Matrix Method. pp. 215–225.

Powell WH, Bright R, Bello SM, Hahn ME (2000) Developmental and tissue-specific expression of AHR1, AHR2, and ARNT2 in dioxin-sensitive and -resistant populations of the marine fish *Fundulus heteroclitus*. *Toxicological Sciences*, **57**, 229–239.

Powell WH, Morrison HG, Weil EJ *et al.* (2004) Cloning and analysis of the CYP1A promoter from the Atlantic Killifish (*Fundulus heteroclitus*). *Marine Environmental Research*, **58**, 119–124.

Ryan JA, Hightower LE (1994) Evaluation of heavy-metal ion toxicity in fish cells using a combined stress protein and cytotoxicity assay. *Environmental Toxicology and Chemistry*, **13**, 1231–1240.

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, **78**, 629–644.

Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, **141**, 413–429.

Skinner MA, Courtenay SC, Parker WR, Curry RA (2005) Site fidelity of mummichogs (*Fundulus heteroclitus*) in an Atlantic Canadian estuary. *Water Quality Research Journal of Canada*, **40**, 288–298.

Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Timme-Laragy AR, Meyer JN, Waterland RA, Di Giulio RT (2005) Analysis of CpG methylation in the killifish CYP1A promoter. *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology*, **141**, 406–411.

Wall JD (1999) Recombination and the power of statistical tests of neutrality. *Genetical Research*, **74**, 65–79.

Wang L, Xu Y (2003) Haplotype inference by maximum parsimony. *Bioinformatics*, **19**, 1773–1780.

Wayne ML, Simonsen KL (1998) Statistical tests of neutrality in the age of weak selection. *Trends in Ecology & Evolution*, **13**, 236–240.

Weaver G (1983) PCB contamination in and around New Bedford, Mass. *Environmental Science & Technology*, **18**, 22A–27A.

Whitlock JP, Okino ST, Dong LQ *et al.* (1996) Cytochromes P450 .5. Induction of cytochrome P4501A1: a model for analyzing mammalian gene transcription. *Faseb Journal*, **10**, 809–818.

Williams LM, Oleksiak MF (2008) Signatures of selection in natural populations adapted to chronic pollution. *BMC Evolutionary Biology*, **8**, 282.

Williams LM, Oleksiak MF (2011) Ecologically and evolutionarily important SNPs identified in natural populations. *Molecular Biology and Evolution*, **28**, 1817–1826.

Williams TD, Lee JS, Sheader DL, Chipman JK (2000) The cytochrome P450 1A gene (CYP1A) from European flounder (Platichthys flesus), analysis of regulatory regions and development of a dual luciferase reporter gene system. *Marine Environmental Research*, **50**, 1–6.

Wingender E, Dietze P, Karas H, Knuppel R (1996) TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Research*, **24**, 238–241.

Wirgin I, Waldman JR (2004) Resistance to contaminants in North American fish populations. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*, **552**, 73–100.

L.W.'s research is broadly focused on understanding the molecular mechanisms involved in the response to toxicant exposure in aquatic organisms and the role of genetic variation in adaptation to toxicant exposure. M.O.'s research addresses the biological importance of individual variation.

Data accessibility

DNA sequences have been deposited in Genbank: Genbank accessions JN166936-JN166960

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Full length CYP1A proximal promoter, exon 1 and intron 1.

Fig. S2 Linkage disequilibrium (LD) among polymorphic nucleotides along the proximal promoter and first exon and intron of CYP1A for Sandwich.

Fig. S3 Linkage disequilibrium (LD) among polymorphic nucleotides along the proximal promoter and first exon and intron of CYP1A for New Bedford Harbor.

Fig. S4 Linkage disequilibrium (LD) among polymorphic nucleotides along the proximal promoter and first exon and intron of CYP1A for Point Judith.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.