HAL
open science

Evolutionary Toxicogenomics of the Striped Killifish (Fundulus majalis) in the New Bedford Harbor (Massachusetts, USA)

Paolo Ruggeri, Xiao Du, Douglas L Crawford, Marjorie F Oleksiak

To cite this version:

Paolo Ruggeri, Xiao Du, Douglas L Crawford, Marjorie F Oleksiak. Evolutionary Toxicogenomics of the Striped Killifish (Fundulus majalis) in the New Bedford Harbor (Massachusetts, USA). International Journal of Molecular Sciences, 2019, 20 (5), pp.1129. 10.3390/ijms20051129 . hal-02095321

HAL Id: hal-02095321

https://hal.sorbonne-universite.fr/hal-02095321

Submitted on 10 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

International Journal of
Molecular Sciences

MDPI

Article

Evolutionary Toxicogenomics of the Striped Killifish (*Fundulus majalis*) in the New Bedford Harbor (Massachusetts, USA)

Paolo Ruggeri [1,2,3,*], Xiao Du [1], Douglas L. Crawford [1] and Marjorie F. Oleksiak [1]

1 Marine Biology and Ecology, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA; xdu@rsmas.miami.edu (X.D.); dcrawford@rsmas.miami.edu (D.L.C.); moleksiak@rsmas.miami.edu (M.F.O.)
2 Dipartimento di Scienze della Vita e dell'Ambiente, Universita' Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy
3 Laboratory of Integrative Biology of Marine Models, CNRS-Sorbonne University, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France
* Correspondence: paolo.ruggeri@sb-roscoff.fr or ruggeripaolo@inwind.it; Tel.: +1-0033672331441

Received: 17 January 2019; Accepted: 23 February 2019; Published: 5 March 2019

Abstract: In this paper, we used a Genotyping-by-Sequencing (GBS) approach to find and genotype more than 4000 genome-wide SNPs (Single Nucleotide Polymorphisms) from striped killifish exposed to a variety of polychlorinated biphenyls (PCBs) and other aromatic pollutants in New Bedford Harbor (NBH, Massachusetts, USA). The aims of this study were to identify the genetic consequences of exposure to aquatic pollutants and detect genes that may be under selection. Low genetic diversity (H_E and π) was found in the site exposed to the highest pollution level, but the pattern of genetic diversity did not match the pollution levels. Extensive connectivity was detected among sampling sites, which suggests that balanced gene flow may explain the lack of genetic variation in response to pollution levels. Tests for selection identified 539 candidate outliers, but many of the candidate outliers were not shared among tests. Differences among test results likely reflect different test assumptions and the complex pollutant mixture. Potentially, selectively important loci are associated with 151 SNPs, and enrichment analysis suggests a likely involvement of these genes with pollutants that occur in NBH. This result suggests that selective processes at genes targeted by pollutants may be occurring, even at a small geographical scale, and may allow the local striped killifish to resist the high pollution levels.

Keywords: candidate outliers; migration; evolutionary genomics; GBS; SNPs; toxicant resistance

1. Introduction

Anthropogenic stressors are a major threat to ecological balance in both terrestrial and aquatic environments. Most anthropogenic stressors are related to several classes of chemical compounds that are directly or indirectly released in the environment by human activities [1]. Although most of these chemicals are released in terrestrial environments, they affect marine environments because of the tight relationship between terrestrial and marine ecosystems [2]. For example, more than 80% of terrestrial contaminants reach coastal and oceanic waters [3]. These pollutants can negatively impact ecosystem functioning of both marine coastal and pelagic areas by reducing the primary production and increasing respiration [4], and these negative effects contribute to demographic instability of several marine species, which induces strong ecological shifts in species composition as well as physiological and evolutionary changes in wild populations [5,6]. The mass-mortality or the shift in available niches generated by some pollutants may alter natural patterns of genetic structure, which

triggers selectively important micro-evolutionary events. Pollutants might act as a selective force, which eliminates individuals that cannot tolerate the pollutants and consequently reduces genetic variability [7,8].

The physiological and evolutionary consequences of several pollutants in wild populations are still not completely understood [9–12]. Finding pollutant-mediated evolutionary trajectories in aquatic populations is a complex task due to high local variability in pollutant mixtures, which can change the pattern and the target of genes affected by selection [13–17]. In this context, methods based on pools of pre-selected markers (i.e., a "candidate genes approach") may not clarify how natural selection is acting when pollutant mixtures occur. The current Next Generation Sequencing (NGS) methods may offer a broader survey of genetic variation and an unbiased opportunity to scan population genomes and identify genome-wide polymorphisms affected by selective pressures from anthropogenic pollutants [10,18]. Among the NGS methods, the use of Genotyping by Sequencing (GBS) is currently becoming a fast, effective, and inexpensive choice to target thousands of single nucleotide polymorphisms (SNPs) throughout the genome [19] and characterize population genetic variation and genomic selection in aquatic toxicology [10].

A promising location to examine the population genetic effects of anthropogenic pollution is in New Bedford Harbor, Massachusetts, USA (hereafter NBH). NBH is a US Superfund site, a polluted site designated for cleanup with funds from the responsible party, and contains hazardous chemicals discharged into it since the mid-1940s. These hazardous chemicals pose a risk to humans and the environment needs to be cleaned up. Despite the cleanup effort in NBH, there is still a persistent pollution gradient that runs from the upper harbor (where the River Acushnet discharge freshwater) to the lower harbor. The pollution level was tracked by a monitoring program run between 1993 and 2009, and data about PCB levels were recorded [20,21]. The NBH hosts a complex pollution mixture made up of polychlorinated biphenyls (PCBs), polychlorinated di-benzo-*p*-dioxins, polychlorinated dibenzofurans, polycyclic aromatic hydrocarbons (PAHs), and metals [20,21]. These pollutants are associated with local biodiversity loss and affect fisheries and shellfish catch because tissues are too contaminated for public consumption [21]. Previous analyses have shown that exposure to environmentally-relevant concentrations of PCB-like compounds (i) have detrimental effects to fish population reproduction [22], (ii) induce embryological developmental anomalies [23], (iii) produce cardiotoxicity [24], and (iv) result in natural selection in wild fish populations [9,25,26].

Two common NBH fish species are the striped killifish (*Fundulus majalis*) and the Atlantic killifish (*Fundulus heteroclitus*) [27]. Atlantic killifish (*F. heteroclitus*) have been used to identify genetic and physiological responses to pollutants in the wild [23,25,28,29]. Key features of *F. heteroclitus* include: (i) large population sizes (>10,000 individuals), (ii) rapid tolerance to new environmental features, (iii) a wide range (that spans from New Brunswick, Canada to the North Atlantic Florida Coast), (iv) non-migratory behavior with limited seasonal movements (in the range of a few meters), and (v) short generation times (essential to evaluate if natural selection affects variation in allele frequencies across generations) [30]. The Atlantic killifish is able to respond physiologically to a large set of environmental variations and stress [11,28,31–33]. The *F. heteroclitus* population in the NBH was found to be tolerant to contaminants and showed genetic differences when related to congeneric populations in unpolluted sites [9,16,30,33–35]. For precisely the reasons mentioned above, *F. heteroclitus* is often used as a model to test physiological and evolutionary responses to local contaminants [9,25,36].

The striped killifish *F. majalis* shares similar biological and ecological features with *F. heteroclitus.* Both killifish species have a sympatric range, but, unlike the more well-known congener (*F. heteroclitus*), the striped killifish is (i) seldom found high in the upper-inner tidal, (ii) infrequently found in fresher water, and (iii) has a greater mobility, which allows this fish to move for several Km and have greater population connectivity [37].

Despite the little information about the evolutionary and physiological responses to pollutants, the striped killifish population in the NBH area persists, which suggests the possibility that it has developed resistance to pollutants. In addition, the striped killifish lacks many resources available for

the Atlantic killifish (i.e., well characterized populations resistance to toxicants or extensive genomic resources) and is, therefore, a good model to assess the power of GBS methodologies and the effects of pollutants in the wild. The GBS approach has been infrequently used with non-model species in an ecotoxicological context and at a fine scale (in this study within a single harbor) [38]. The aim of this study is, therefore, to investigate the genetic variation in striped killifish samples collected along a steep pollution variation in the New Bedford Harbor (MA, USA). We apply an unbiased genomic approach (genotyping by sequencing, GBS) to simultaneously find and genotype thousands of SNPs in striped killifish and offer the possibility to add this species as a new model for genomic resistance to aquatic pollutants. We use multiple tests to detect candidate outliers and verify how they respond when used on non-model species exposed to steep variation in pollutants.

2. Results

2.1. NGS Sequencing

Illumina GBS sequencing using the TASSEL pipeline with the *Fundulus heteroclitus* reference genome (https://www.ncbi.nlm.nih.gov/genome/743) recovered 5403 SNPs found in at least 80% of individuals with 136 individuals having at least 70% of all SNPs. Among these SNPs, 1275 SNPs were in the Hardy Weinberg disequilibrium with observed heterozygosity significantly exceeding expected heterozygosity (HWE, $p < 0.01$). These SNPs were removed, and the final dataset of 4128 SNPs was subsequently used for statistical analyses.

The preliminary test for detecting SNPs under directional selection identified 564 SNPs as potential candidate outliers from all the six possible pairwise comparisons among sampling sites. Subsequent analyses, where neutral genetic variation was expected, were run using 2208 presumably neutral SNPs (excluding SNPs with significant linkage disequilibrium and candidate outliers).

2.2. Genetic Differentiation Between Populations and Gene Flow

Pairwise genetic differentiation (F_{ST} values) based on 2208 neutral SNP loci were all non-significant, which shows negative values that ranged between −0.0039 (PIL vs. FAH) and −0.0057 (HST vs. MAT, Table 1). When pairwise F_{ST} values were estimated using candidate outliers, 5 out of 6 pairwise population comparisons were significant and ranged between 0.0034 (PIL vs. HST) and 0.0303 (PIL vs. FAH, Table 1). The AMOVA test showed that most of the molecular variance was explained within populations (99.67%, FST: 0.0023) even though a significantly small portion was explained by groups (0.23%, FCT: 0.0032, Table S1).

Table 1. Pairwise F_{ST} values between sampling sites based on neutral genetic markers (2208 SNPs, below the diagonal) and non-neutral genetic markers (linked loci + candidate outliers, above the diagonal). * *p*-value < 0.05, ** *p* -value < 0.01, *** *p* -value < 0.001.

	PIL	FAH	HST	MAT
PIL	0	0.0303 **	0.0034	0.0111 *
FAH	−0.0039	0	0.0250 ***	0.0221 ***
HST	−0.0048	−0.0053	0	0.0158 **
MAT	−0.004	−0.0052	−0.0057	0

Results from DAPC were stable after retaining 28 out of 60 PCAs and showed the first two components as significant. In general, all four sampling sites were not well resolved. However, along the axis of the first significant component, a separation between PIL + FAH + HST against MAT appears (Figure 1). Similarly, along the axis of the second most significant component, a slight separation between PIL + FAH, MAT, and HST was detected (Figure 1).

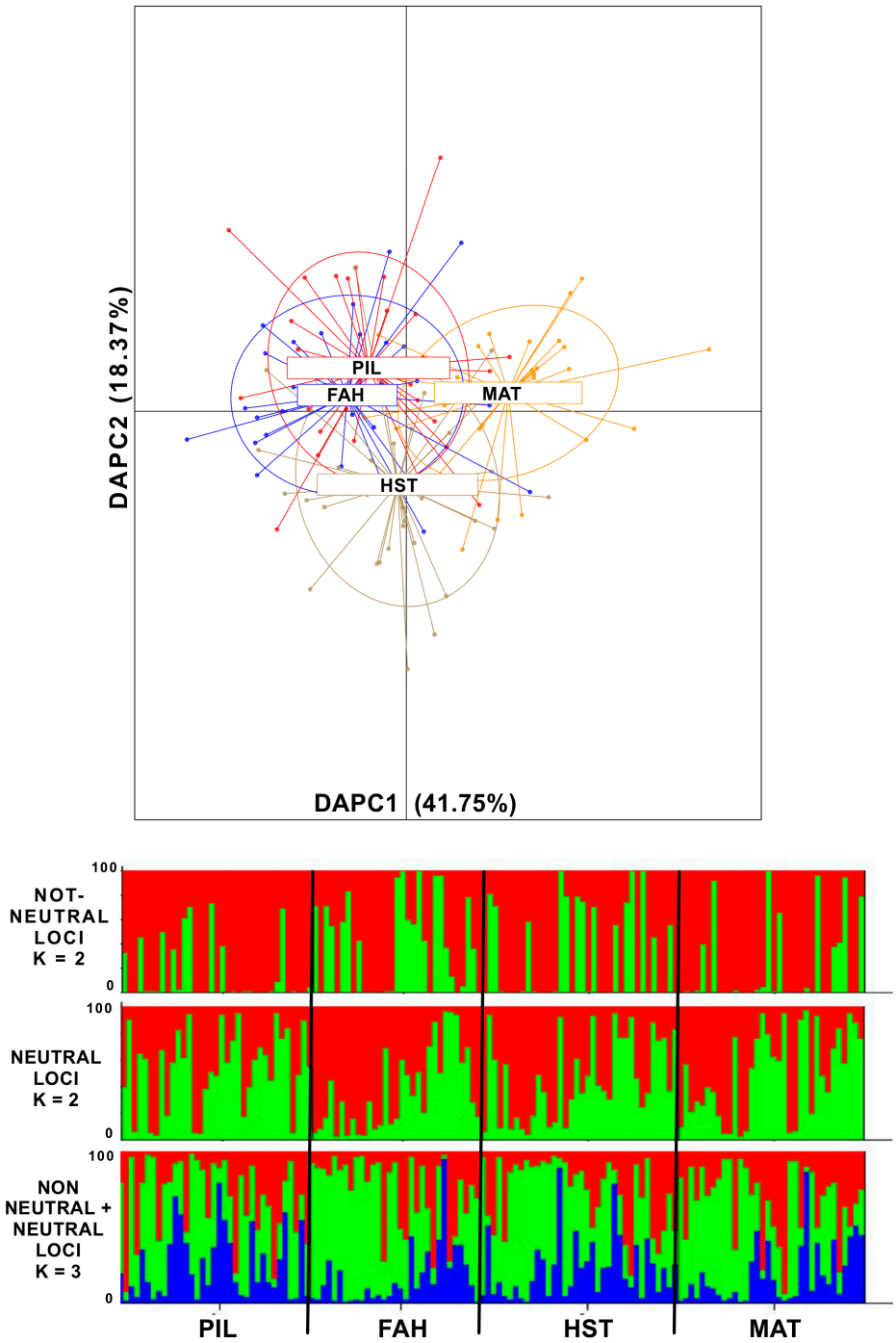


Figure 1. Figure 1 (above) DAPC (Discriminant Analysis of Principal Components) plot of individuals from the four sampling sites along the two most significant axes (PC1 41.75%, PC2 18.37% explained variance). In red are the individuals from Pilgrim Avenue (PIL) and, in blue, are the individuals from Fairhaven (FAH). Both PIL and FAH are located in the inner New Bedford Harbor (NBH) close to the pollution source. Depicted in light brown are individuals from Hacker Street (HST) and, in yellow, are individuals from Mattapoisett (MAT). Both these latter sampling sites are located in the outer NBH area and are exposed to lower pollution concentrations. The insert explains the percent of variance for the first three discrimination eigenvalues (in dark grey are the significant DAs). Figure 1 (below), STRUCTURE plots of the most likely K tested from the simulations with non-neutral (K = 2), only neutral (K = 2), and the complete set of markers analyzed (K = 3).

STRUCTURE simulations of 2208 neutral SNPs yielded after the Evanno method, the most likely structure of two genetic clusters (K = 2) (Ln[P(K)] = −29,0230.24 ± 68.46) and a potential substructure for four genetic clusters (K = 4) (Figure 1, Figure S1). A similar structure (K = 2 and a substructure for K = 4) was detected using 1920 loci that included candidate outliers and loci under potential linkage disequilibrium (Figure 1, Figure S1). The simulation carried out using all SNPs yielded a higher log likelihood for three main genetic clusters (K = 3) (Ln[P(K)] = −531,638.50 ± 51.84). The same scenario was confirmed after the Evanno method (Figure S1). However, the q value distribution in both simulations did not detect a spatially explicit genetic structure among the four sampling sites (Figure 1).

Estimation of the number of first generation migrants showed large connectivity among sampling sites. Table S2 provides the percentage of individuals assigned to the same sampling site). Migration ranges between 8.82% in MAT and 16.67% in HST. The source of first-generation migrants was mostly HST (representing 45.71% in PIL, 35.48% in FAH, and 32.35% in MAT) and PIL (representing 38.89% in HST and 32.35% in MAT, Table S2).

MIGRATE-n results showed values of the Θ parameter ranging from 0.00010 in MAT to 0.00057 in HST (Figure 2). The historical gene flow seemed to be symmetrical between every sampling location pair except between FAH and HST ($M_{FAH \rightarrow HST}$ = 61.7, $M_{HST \rightarrow FAH}$ =127.7, Figure 3). The M values ranged between 61.7 ($M_{FAH \rightarrow HST}$) to 131.8 ($M_{HST\text{-}MAT}$) (Figure 2).

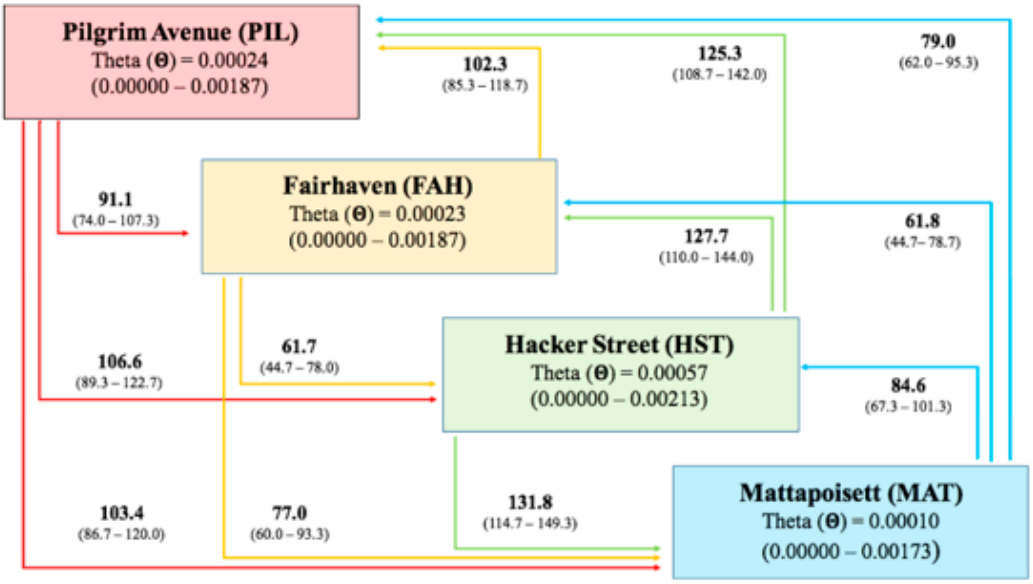


Figure 2. Pattern of historical gene flow from the Migrate-n simulation. The squares represent the four sampling locations and theta (Θ) values with 95% confidence intervals. Colored arrows refer to the sampling location of the same color and explain the directionality and rate of gene flow (M values are reported at the corner of each arrow with a 95% confidence interval).

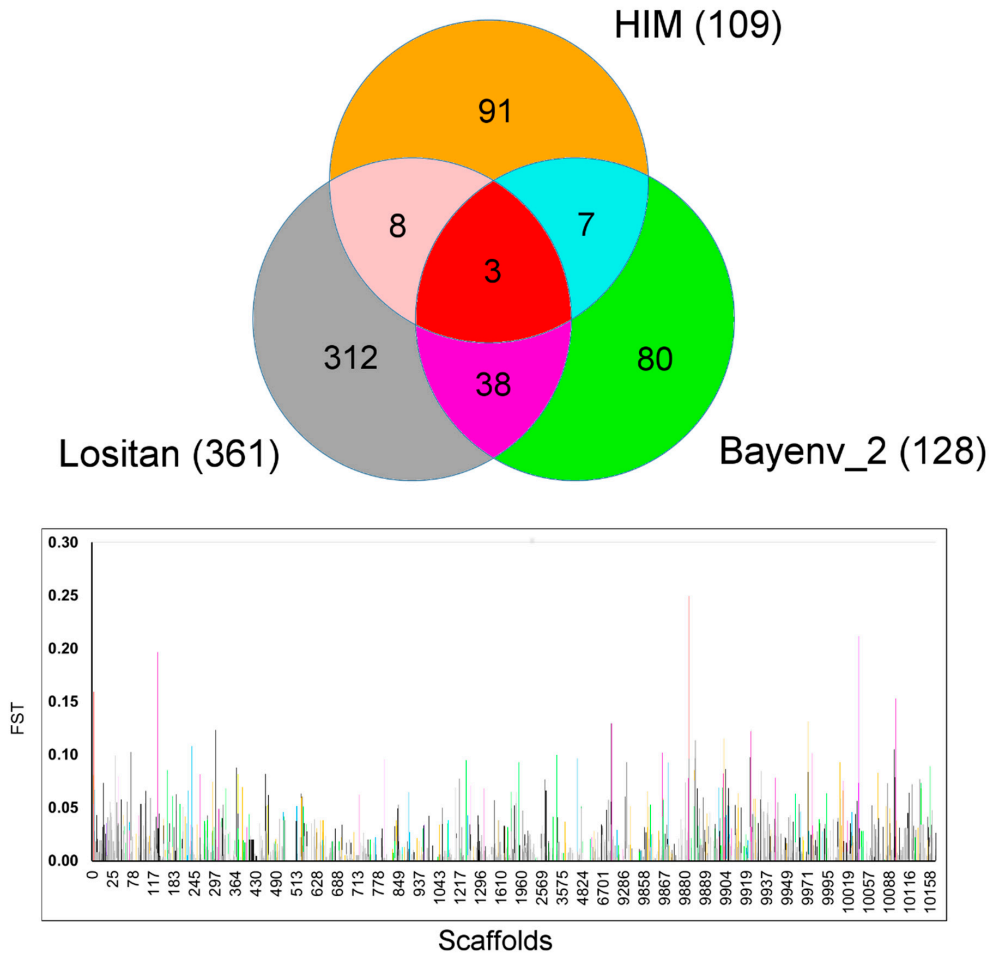


Figure 3. Venn diagram of the outlier loci interpolation among four candidate outlier detection tests and F_{ST} value distribution of each outlier locus along the scaffolds position. In the Venn diagram (above), the numbers encased represent the number of shared SNPs between and among the corresponding set of methodologies. The colors in the Venn diagram reflect the pattern of color used in the Manhattan plot (below) that represents the scaffold position of each outlier locus and its estimated F_{ST} value.

2.3. Tests for Detecting Signatures of Selection

The F_{ST}-based test to detect SNPs under directional selection in Lositan indicates 361 SNPs as candidate outliers. The pairwise comparison between PIL vs. MAT detected 117 candidate outliers (after 1% FDR correction), PIL vs. HST, 82 candidate outliers, FAH vs. MAT 115 candidate outliers, FAH vs. HST, 96 candidate outliers, and HST vs. MAT, 90 candidate outliers.

Identification of outlier SNPs using a Hierarchical Island Method (HIM) implemented in Arlequin 3.5.2.2 detected 109 candidate outliers (after 1% FDR correction).

A third analysis to detect SNPs evolving by natural selection used Bayenv 2 analysis, which defines SNPs correlated with an environmental parameter (pollution) after correcting for demography [39]. A total of 128 SNP_S [with $Log_{10}(BF) > 1$ and with empirical p-values < 0.05] were detected among the 10 replicates performed. Among them, 110 SNPs were detected only in one replicate, 13 SNPs were detected in two replicates, 3 SNPs were detected in three replicates, and 2 SNPs were found six times.

Outcomes from (i) Lositan, (ii) HIM, and (iii) Bayenv 2 revealed a total of 539 potential candidate SNP outliers. In all the tests performed, we detected outliers with a distribution across the whole *Fundulus* genome (Figure 3). There was overlap for three loci with all the three outlier's detection methods (S0_4352665, S0_4352669, S9887_384963) including 56 SNPs were found to overlap between at least two detection methods (Lositan, Bayenv2 and HIM) (Figure 3).

2.4. Tests for Functional Annotation of Candidate Outliers

The sequences (75 base pair sequences) for the 539 candidate outliers were aligned against any available GenBank resource using the blastn algorithm. A total of 237 SNPs out of 539 (28.01%) had hits with significant (E-value < 0.0001) annotations. A total of 99.75% of those hits matched Eukaryote sequences. Additionally, 68.77% of the Eukaryote hits were related to teleost fish species (Figure S2) and 36.23% of them belonged to the Cyprinodontiformes Order, of which 41.76% of them specifically referred to *Fundulus heteroclitus*, which is the most closely related species to *F. majalis* with available genomic resources (Figure S2).

A total of 151 SNPs loci out of the 237 SNPs with annotations were related to coding regions (functionally annotated SNPs) NCBI ID codes for these functionally annotated SNPs, which were converted into human UNIPROT ID codes and used for the enrichment analysis in DAVID 6.7. These 151 SNPs produced a total of 958 Uniprot hits associated with 429 different human genes (*Homo sapiens*) (Table S3). The analysis with DAVID 6.7 identified 44 different gene clusters and 14 out of 44 clusters showed significant associations ($p < 0.05$, Table S4). These clusters are associated with 35 functional pathways and 25 out of 35 pathways showed significant cellular/biochemical/physiological functions (Bonferroni p -values < 0.05, Table 2). A total of 1126 of the total 1147 Uniprot hits were also significantly (p -value < 0.001) related to 13 different disease classes that span from metabolic (167), cardiovascular (137), and cancer (103) to developmental (60) and reproductive (37) pathologies (Table 3).

Table 2. The list of 35 cellular/physiological pathways targeted by the enrichment analysis (KEGG Pathways) with DAVID 6.7. The analysis includes results from a list of 429 genes. N = number of genes participating to the KEGG Pathway. % = percentage of genes in the total of 429 genes. Category = represents the general cellular/physiological function to which the KEGG pathway is involved in. The Kegg pathway terms in red are not supported by significant P-values (P > 0.05). Categories: a = Metabolism. b = Cellular differentiation/survival. c = Cellular organization/adhesion. d = Cancer. e = Development. f = Immune response. g = Reproduction. h = Inflammatory processes. I = Neuronal transmission.

KEGG Pathway Terms	N	%	*p*-Value	Category
Nitrogen metabolism	11	2.5	3.40×10^{-12}	a
PPAR signaling pathway	11	2.5	1.30×10^{-5}	a
Mineral absorption	9	2.1	2.70×10^{-5}	a
Neurotrophin signaling pathway	14	3.2	2.40×10^{-5}	b
Regulation of actin cytoskeleton	18	4.2	6.30×10^{-5}	c
ErbB signaling pathway	11	2.5	1.30×10^{-4}	b, d
Adherens junction	8	1.9	3.30×10^{-3}	c
Ras signaling pathway	15	3.5	3.80×10^{-3}	b, c
MAPK signaling pathway	16	3.7	4.20×10^{-3}	b, h
Fc gamma R-mediated phagocytosis	8	1.9	8.30×10^{-3}	b, c, f
Proteoglycans in cancer	13	3	9.20×10^{-3}	b, c, d
Tight junction	8	1.9	1.00×10^{-2}	b, c
Epithelial cell signaling in *Helicobacter pylori* infection	7	1.6	1.00×10^{-2}	b, f, h
Pathogenic *Escherichia coli* infection	6	1.4	1.30×10^{-2}	c, f, h
Rap1 signaling pathway	13	3	1.30×10^{-2}	b, c
Leukocyte transendothelial migration	9	2.1	1.40×10^{-2}	c, f, h
Endocytosis	14	3.2	1.60×10^{-2}	a, c
cAMP signaling pathway	12	2.8	2.10×10^{-2}	b, c, e, f, g
Salmonella infection	7	1.5	4.60×10^{-2}	b, f, h
Renal cell carcinoma	6	1.4	3.50×10^{-2}	b, d
Arhythmogenic right ventricular cardiomyopathy (ARVC)	6	1.4	4.20×10^{-2}	c, f, h
Cell adhesion molecules (CAMs)	9	2.1	4.20×10^{-2}	c, f
Inflammatory mediator regulation of TRP channels	7	1.6	4.30×10^{-2}	h
Estrogen signaling pathway	7	1.6	4.60×10^{-3}	b, g
T cell receptor signaling pathway	7	1.6	5.80×10^{-2}	b, f, h
Choline metabolism in cancer	7	1.6	6.00×10^{-2}	b, c, h
Dopaminergic synapse	8	1.9	6.40×10^{-2}	i
Bacterial invasion of epithelial cells	6	1.4	6.40×10^{-2}	c, f, h
Non-small cell lung cancer	5	1.8	6.80×10^{-2}	b, d
Regulation of lipolysis in adipocytes	5	1.8	5.80×10^{-2}	a
Hepatitis C	8	1.9	7.50×10^{-2}	d, h
Apoptosis	5	1.2	9.10×10^{-2}	b, d
Shigellosis	5	1.2	1.00×10^{-1}	d, h
Central carbon metabolism in cancer	5	1.2	1.00×10^{-1}	d, h

Table 3. The list of 13 group of diseases targeted by the enrichment analysis (DAVID 6.7). The analysis includes two tests to assess significance of the genes participating to each disease category (Permutation test P-value and FDR correction). N = number of genes with implication in the categorized disease. % = percentage of representation in the total of 429 genes.

Disease Category	N	%	P-Value	FDR
Metabolic	167	34.9	1.80×10^{-3}	4.60×10^{-3}
Cardiovascular	137	28.7	4.50×10^{-3}	1.00×10^{-2}
Chemo-dependency	115	24.1	1.90×10^{-2}	2.80×10^{-2}
Pharmacogenomic	109	22.8	2.40×10^{-6}	4.30×10^{-5}
Neurological	106	22.2	1.80×10^{-4}	8.00×10^{-4}
Cancer	103	21.5	1.70×10^{-2}	2.80×10^{-2}
Psych	81	16.9	4.10×10^{-5}	2.50×10^{-4}
Unknown	66	13.8	3.10×10^{-4}	1.10×10^{-3}
Developmental	60	12.6	1.10×10^{-3}	3.40×10^{-3}
Other	58	12.1	1.10×10^{-2}	1.90×10^{-2}
Renal	54	11.3	5.70×10^{-3}	1.10×10^{-2}
Reproduction	37	7.7	2.10×10^{-2}	2.90×10^{-2}
Normal-variation	33	6.9	7.00×10^{-6}	6.30×10^{-5}

3. Discussion

Evolutionary toxicology, which is the study of evolutionary responses of populations to human-mediated stressors, is still in an early stage [40]. It aims to understand how pollutants contribute to (i) changes in genome-wide genetic diversity, (ii) population differentiation, and (iii) local selection at genes that modify physiological pathways associated with pollutant resistance [40]. Genomic approaches provide unbiased genome-wide analyses that help clarify populations' molecular evolutionary responses to pollutants.

In this study, we investigated how natural selection affects local population exposed to contaminants in an estuarine, non-model species: the striped killifish (*F. majalis*). Particularly, this method successfully allowed us to obtain more than 4000 SNP markers from *Fundulus majalis*, that were used to detect allelic shifts in loci potentially connected to the exposure to different levels of pollutants. Similar approaches were previously tested comparing genomes of individuals collected from contrasting environments [12,41,42]. This strategy should help define loci under strong selection, typically in isolated environments. An important caveat is that we are assuming that pollutants, primary PCBs, are important selective forces. Yet, other abiotic (salinity) and biotic (community composition) factors could also co-vary with PCBs and may also be selectively important. The focus on pollution in this study seems reasonable because we are investigating a single population with samples separated by few Km of coastal area where the pollutant concentration has the largest environmental variation relative to many other abiotic factors. Additionally, this approach is supported by the observation that more than half of the annotated outlier loci were related to genes known to be influenced by PCBs. The factors include (a) regulation of apoptosis and cell lifespan, (b) cancer, and (c) hormones and neuropeptides signaling. In this case, we examine the genetic variation of a population of *Fundulus majalis* in New Bedford Harbor (Massachusetts, USA) exposed to a large variation in aquatic contaminants.

3.1. Genetic Variability and Population Differentiation for F. majalis in the New Bedford Harbor

Along the pollution gradient, our data indicate consistently high genetic variation. This goes against basic assumptions in evolutionary toxicology and the "genetic erosion hypothesis" postulated by van Straalen and Timmermans [43] that suggests genetic variability erosion due to rapid demographic collapse when populations are faced with contaminants. Such an event can decrease genetic variability promoted by genetic drift and selection [40]. This model predicts that "genetic erosion" should be proportional to the proximity of individuals to the contaminant source

and should result in a progressive loss of genetic variation with lower diversity in the most polluted sites and higher diversity in the unpolluted ones. Yet, in this study, there is no apparent progressive reduction in genetic variation (Figure 4, expected heterozygosity and nucleotide diversity) along the pollution gradient. The same result was observed in a genotyping-by-sequencing analysis for a rove beetle (*Staphylinus erythropterus*) population along a heavy metals pollution gradient in Poland where there was no clear pattern of reduced genetic variation measured from over three thousand SNPs in beetles located close to the pollution source [38]. This beetle study also found weak genetic differentiation between sampling sites and suggested that it was due to extensive gene flow among sites. Therefore, high beetle mobility may counteract genetic variation loss associated with pollution by continuous migration from nearby, less impacted sites. Similar reasoning could explain the data for the NBH striped killifish population. A lack of population structure in the striped killifish collected in the NBH area and high connectivity among sampling sites were detected. This connectivity is likely related to the mobility of the species, which covers similar distances to those that separate the sampled sites (<20 Km). Boorse and Storlie [44] suggest that striped killifish daily movements can enhance the connectivity of local groups of individuals. This observation suggests that connectivity within populations may contribute to counteract the local genetic variability loss [45,46]. Yet, even in the face of high connectivity, our data suggest that evolutionary divergence affects allele frequencies at many SNP loci. This is in agreement with new theories that suggest the need for high genetic variation in order to maintain adaptive responses in populations chronically exposed to stressful environments [47]. In this context, selective divergence with high gene flow could be advantageous by increasing population size and introducing new alleles that may be locally beneficial [46,47]. The evidence for and the problems with identifying selective divergence in the striped killifish are discussed below.

3.2. What Tests for Candidate Outliers can Tell Us about Selective Response to Pollutants: A Matter of Relative Performance

One of the main challenges in ecological genomics is to identify genes underlying functional traits, such as toxicant resistance [18,48]. Currently, tests to detect selectively important genetic variation are based on different assumptions that affect their statistical power and accuracy [49–51]. In general, comparing the results from several tests is a good practice to assess the occurrence of functionally relevant genetic variation [49,52]. In this study, we used three different tests for detecting outlier loci based on (i) F_{ST}-based pairwise comparisons of most divergent loci, (ii) an F_{ST}-based approach that can account for a hierarchical structure, and (iii) a Bayesian correlation test between pollutant concentrations and the genetic variation at loci. The proportions of candidate outliers ranged between 2.64% and 8.75% of the total SNPs markers we retrieved. These percentages are roughly in accordance with the relative number of SNPs showing significant genetic variation (<5–10% of loci screened) found in other studies [48–53].

There was limited overlap among the methods applied (Lositan, HIM and Bayenv2) with only three overlapping loci. Limited overlap is a frequent issue when comparing genome-environment association methods (i.e., Bayenv2) with others that rely on divergence in F_{ST} (i.e., Lositan and HIM) [49–51] and depends on the assumptions made by the different statistical tools that could not perform in the same manner when dealing with (i) polygenic traits, (ii) panmictic/highly connected populations, and/or (iii) when the selective agent is actually represented by a mix of multiple stressors/features.

For instance, several investigations demonstrated that the polygenic nature of phenotypic traits poorly accord with the basic assumptions of most genome-wide scan approaches [12,54,55]. With polygenic selection, allele frequency changes only need to occur at a subset of loci that could alter an evolutionary trait. Thus, signals of selection are distributed across many loci and are not necessarily the same in all individuals, which leads to weak individual-locus detection for selection [54,55]. It could be supposed that genes functionally relevant for tolerance to pollutants might be polygenic traits and, hence, vary in the degree of detectability by the different methods used in this scenario.

Demography also affects the power of outlier tests producing a high type II error (falsely accepting the null hypothesis) [49–51]. This is frequently found when there is a lack of population structure or in the case of strong connectivity [56]. In this context, F_{ST}-related estimates between sampling sites would be lowered by higher gene flow and, in turn, affect the power of tests to detect loci under selection [49,57]. A similar outcome was proposed in a study on the panmictic American eel (*Anguilla rostrata*) that also exhibited less than 1% of candidate outliers from a genome-wide scan performed at a larger spatial scale than this study [58]. This scenario fits well with the high level of connectivity we detected in *F. majalis* from NBH and could be an explanation for the small overlap of candidate outliers obtained.

Similarly, the effects generated by a gradual variation of the agent that is supposed to generate local selection can lead to analogous outcomes. In fact, other studies have pointed out how environmental gradients break the assumption of population isolation and exposure and lead the genome scan methods to bias the expected number of outlier loci [56,57].

The role of multiple interacting stressors could also lead to variation in the relative performance produced by different tests to detect outliers. The New Bedford Harbor (NBH) pollution source comes mainly from PCBs, but they are not the exclusive contaminants in the area [20,21].

Although we assumed that all contaminants in NBH follow the same pattern of concentration, we can speculate either that contaminants other than PCBs (to which information about local concentrations are not available) could be more greatly affecting (i) the selective responses in the striped killifish or (ii) the responsiveness of the environmental correlation methods employed (i.e., Bayenv2).

3.3. Genome-Environment Interaction in F. majalis from NBH

Although the outliers SNPs obtained from four different methods did not consistently overlap, enrichment analysis of the functionally annotated SNPs suggest the possibility that these loci relate to important traits that respond to pollution in the striped killifish. The enrichment analysis showed that most of these markers can be clustered in 14 different functional groups involved in: (i) antioxidant, xenobiotic, and fatty acid metabolisms, (ii) control of the cell fate (i.e., apoptosis, cell proliferation), (iii) control of the innate immunity, and (iv) neuromodulation. Similarly, these genetic clusters reflect 25 functional pathways that control analogous physiological responses. Most of the functionally annotated SNPs were also related to pathologies that span from cancer, cardiovascular, metabolic, and neurological to reproductive diseases. All these disease categories are known to be associated with physiological cellular interactions with xenobiotics, such as PCBs and other aromatic compounds [59–61]. These toxicants are usually highly lipophilic, and their functions are connected to membranes and the vesicles structure as well as to their transmembrane molecular crosstalk [61].

Several targeted SNPs are in genes involved in ATP-GTP binding activities (i.e., Ras-related protein RAB38 and syntrophin alpha) or with calcium-dependent molecules (i.e., susd1, cacng5, PLCD4, and stx2). These genes can be connected with the cardiovascular activities, and, in general, the cardiovascular system has been identified as a main target of dioxin-like compounds toxicity in all vertebrates [61] and more specifically in *F. heteroclitus* [62]. One of the calcium-dependent genes, syntaxin 2, was also identified to be involved in the acrosomal reaction and with sperm mobility [63]. This function may be put in relation with the estrogen-mimicking behavior that many xenobiotics (i.e., PCBs) assume and their role as endocrine disruptors for many organisms, including fish [22,64]. PCBs act as estrogen-like molecules in both males and females, inducing cancer and reproductive alterations, like induced feminization in males [65]. The reproductive consequences of PCBs are supported by the significant allele frequency change for the estrogen receptor 1 (esr1). This gene receptor is one of the main targets for synthetic xenoestrogens [22,66] and it plays a protective role from PCB-induced DNA damage in human breast cancer cells and apoptotic cycle alterations [67]. It also has been shown to be differentially regulated in NBH *F. heteroclitus* compared to reference fish [68,69].

Additionally, our approach identified genes directly connected with the metabolism of xenobiotic like the cytochrome P450 genes (CYP2F2, CYP2K4, CYP2P2) and the Aryl Hydrocarbon Receptor

Nuclear Translocator 2 (ARNT2). A total of 137 genes encoding P450s were recently identified in fish, and their general activities relate to detoxifying systems against pollutants [70]. CYP2F2 and CYP2K4 are specifically associated with epoxygenase and hydroxylase activities of arachidonic acid, which is a polyunsaturated fatty acid involved with cellular signaling and inflammatory responses [71]. A significant transcriptional alteration of a member of the CYP2 family (CYP2N2), which is also involved with molecular mobilization of arachidonic acid, was also detected in the NBH *F. heteroclitus* population where the lower CYP2 expression in the polluted populations could be related to a limitation in energy storage and other metabolic functions [31]. The change in CYP2 allele frequency could have a similar role for *F. majalis* from the same area. The Aryl Hydrocarbon Receptor Nuclear Translocator 2 (ARNT2) was also demonstrated to be one of the genes that transcriptionally respond to exposure from PCBs and other aromatic compounds in *F. heteroclitus* embryos [72]. This gene resists pollutants developed by *F. heteroclitus* in the NBH, and it can be speculatively proposes that it has the same phenotype results for the *F. majalis* population we examined.

Furthermore, we detected SNPs loci associated with metallothioneins (MTs), which is a group of functionally diversified proteins with metal binding and redox capabilities that confer tolerance to metal pollutants. An increase in MTs is frequently associated with hepatotoxicity due to PCB exposure in fish. MTs could be similarly involved in toxicant resistance of *F. majalis* in NBH.

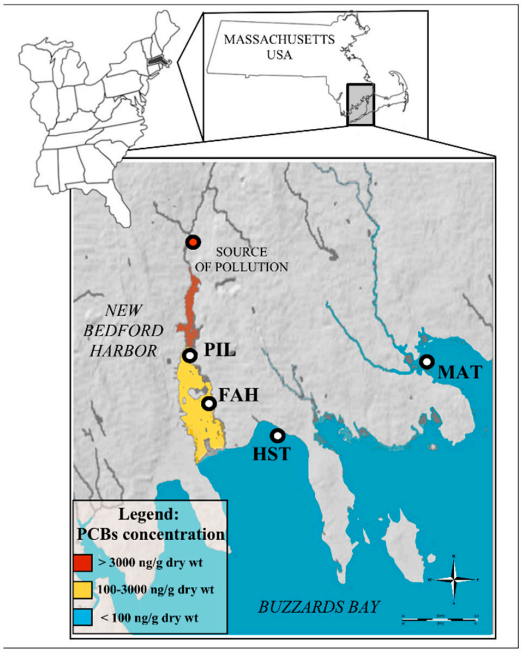
Another interesting result comes from the cluster that contains the Leucine-rich repeat genes (LRRs, genomic components that are involved in the innate immune response) and the functional pathways connected with the immunological response to bacterial infections. In fact, it is not so surprising that the existing connection between chronic exposure to contaminants like PCBs and metals and the development of transgenerational immunological suppression effects. Nacci et al. found that both adults and embryo *F. heteroclitus* from NBH are producing an immune response comparable to control specimens, and, therefore, they proposed that this killifish population might have evolved mechanisms that minimize the immuno-suppressive effects of PCBs [73]. A similar consideration could be proposed for the striped killifish population that we examined. The emerging genes from our GBS analysis that relate to the immunological response to bacterial infections could suggest that *F. majalis* from NBH have developed a genomic resistance that contrasts the immunosuppressive effect of PCBs or that, in general, genes related to the immune system are playing a role in the NBH population.

These findings suggest that local selection of genes that resist toxicants may occur in *F. majalis*. However, other relevant genes found in the adaptation of *F. heteroclitus* to PCB-contaminated environments [35,62] were not identified. In particular, we found no association between nucleotide variation and PCB-tolerant phenotype at the aryl hydrocarbon receptors 1 and 2 (AHR1, AHR2) [35,62], cathepsin Z, the cytochrome P450s (CYP1A and CYP3A30), and the NADH dehydrogenase subunits genes [62]. This points out one of the short-comings of GBS. GBS samples of approximately 0.1% of the *Fundulus* genome often does not assay all relevant genes. Thus, although we observed potential selection acting on previously described genes (i.e., the estrogen receptor 1 and CYP450 genes), the GBS approach is not comprehensive in detecting all loci implicated with *F. majalis'* selective response to local pollutants.

4. Materials and Methods

One-hundred and fifty-eight striped killifish were collected from the New Bedford Harbor (NBH, MA, USA). These individuals were collected in four sampling sites representative of a polychlorinated *biphenyl* (PCB) gradient measured as standard dry mass weight of the sediment, as reported in Roark et al. [36]. Two collection sites were in the Acushnet River estuary within New Bedford Harbor (Figure 4, PIL and FAH). These sites represent the areas closest to the pollution source where striped killifish naturally can occur. The other two collection sites have lower pollution levels. One site was in the New Bedford Outer Harbor (Figure 4, HST) while the most distant sampling site was in the Mattapoisett Harbor (Figure 4, MAT). The latter site was defined as a reference site for our analysis. This is because it: (i) experiences lower pollution levels, (ii) belongs to a different embayment system

than NBH, and (iii) represents the farthest geographical site from the pollution source compared to the other sampled sites. Fish were collected using minnow traps. All fish were returned to their environment after removal of small fin clips (<10 mm^2). Fin clips were stored in 320 µL of Chaos buffer (4.5 M guanadinium thiocynate, 2% N-lauroylsarcosine, 50 mM EDTA, 25 mM Tris-HCL pH 7.5, 0.2% antifoam, 0.1M β-mercaptoethanol) at 4 °C prior to being processed. Genomic DNA was isolated using a silica column protocol [74] and the DNA quality was assessed via agarose gel electrophoresis. DNA concentrations were quantified and standardized using AccuBlue broad range quantitation assays, according to the manufacturer's instructions (Biotium Inc., Fremont, CA, USA; https://biotium.com/).



Location	Code	N	Lat-Long	[PCBs]	D (km)	MAF	H_E	π	Mono
Pilgrim Avenue, Fairhaven, MA, USA	PIL	38	41◦38.8'N; 70◦54.7'W	874	3.5	0.733 (0.729-0.737)	0.350 (0.345-0.354)	0.342 (0.340-0.345)	21
Fairhaven boat launch, Fairhaven, MA, USA	FAH	40	41◦38.4'N; 70◦54.4'W	541	5	0.730 (0.725-0.734)	0.353 (0.349-0.357)	0.351 (0.349-0.353)	24
Hacker Street, Fairhaven, MA, USA	HST	40	41◦37.9'N; 70◦53.1'W	13	8	0.733 (0.728-0.737)	0.351 (0.346-0.355)	0.346 (0.343-0.348)	13
Mattapoisett Bay, Mattapoisett, MA, USA	MAT	40	41◦38.9'N; 70◦49.4'W	27	24	0.734 (0.729-0.738)	0.349 (0.345-0.353)	0.346 (0.344-0.349)	23
Overall						0.732 (0.730-0.734)	0.351 (0.349-0.351)	0.346 (0.343-0.350)	0

Figure 4. Geographical area and the sampling locations covered by this study. The main map refers to New Bedford Harbor and Buzzards Bay (URL: http://mapmaker.education.nationalgeographic.org). The colored coastal areas refer to the pollution gradient described in the figure legend. Map ruler (URL: https://upload.wikimedia.org/wikipedia/commons/1/1a/Scale_kilometres_miles.svg) and compass (URL: https://commons.wikimedia.org/wiki/File:Simple_compass_rose.svg). The map of Eastern USA was obtained from URL: http://www.clker.com/cliparts/m/3/H/f/K/e/eastern-u-s-map-md.png). The map representing the Massachusetts outline was downloaded from Worldatlas.com at the URL: http://www.worldatlas.com/webimage/countrys/namerica/usstates/outline/ma.htm). The table below the figure summarize the main pollution and genetic diversity observed. Code = collection site codes, N = sample size, Lat-Long = geographic coordinates of sampled points, [PCBs] = PCBs concentration in (ng/g dry weight), D = distance in Km from the source of pollution. Genetic diversity by population: averaged Major Allele Frequencies (MAF), averaged expected heterozygosity (H_E), and averaged nucleotide diversity (π), and number of monomorphic loci (Mono). Values between brackets refer to 95% confidence intervals.

The GBS library was prepared as described in Elshire et al. [19] using the restriction enzyme *Ase* I. Adaptors (0.4 pmol/sample) were ligated to 50 ng of gDNA. The GBS library was sequenced on a single lane using Illumina HiSeq 2500 with a 75 bp single-end reads (Elim Biopharmaceuticals, Inc., Hayward, CA, USA; https://www.elimbio.com/). Sequence reads were aligned and putative SNPs were retrieved using the TASSEL 5.0 pipeline [75] with the *F. heteroclitus* genome as a reference [76]; https://www.ncbi.nlm.nih.gov/genome/743]. Discovered SNPs were filtered with TASSEL 5.0 retaining individuals with at least 70% of SNPs and SNPs found in at least 80% of individuals. Allele frequencies and heterozygosities (H_E, H_O) from filtered data were calculated per sampling site and per locus using Arlequin 3.5.2.2 [77]. SNPs with observed heterozygosity (H_O) exceeding expected heterozygosity (H_E, $H_O > H_E$) and significantly different from Hardy-Weinberg expectation ($p < 0.01$) were excluded. This latter filter is used to remove potential SNPs that represent differences between paralogs versus true allelic variants for a single locus [78,79]. Overall and by-population estimates of nucleotide diversity, (π) were computed using DNAsp v5 [80]. Initially, SNPs under neutral and non-neutral expectations were evaluated with Lositan [81]. Lositan settings consisted of 500,000 iterations based on an Infinite Allele model. A list of candidate outliers detected under positive selection was annotated from each possible pairwise comparison. Each test between localities (for a total of six pairwise comparisons) was replicated three times, and the final list of candidate outliers (presumably non-neutral markers) was obtained by merging lists from the three replicates. A False Discovery Rate (FDR, 1%) correction method was applied after evaluating significances with p-values < 0.01. This strategy was applied with the aim to use a more stringent criterion to detect significant, non-neutral SNPs.

The population structure was evaluated using AMOVA in Arlequin [77], multivariate discrimination analysis (DAPC) in the R package ADEGENET 1.3-4 [82] (http://adegenet.r-forge.r-project.org/), and STRUCTURE version 2.3.4 [83] (https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/html/structure.html). Migration rates were evaluated as implemented in GENECLASS 2.0 [84] as well as MIGRATE-n v. 3.5.1 [85] (https://peterbeerli.com/migrate-html5/index.html). The genetic distance among populations was estimated using F_{ST} values [86], as implemented in Arlequin [77] (http://cmpg.unibe.ch/software/arlequin35/Arl35Downloads.html). Among the four sampling sites, the variation within and among populations was determined by analyzing molecular variance (AMOVA, Arlequin) [77]. In the AMOVA, a hierarchical structure based on both the PCB concentration and geographical distance (measured as the shortest linear sea distance among sampling sites) was provided. The two most polluted sites were grouped together in the first cluster (PIL and FAH) while the remaining two sampling sites (HST and MAT) represented two other distinct clusters. A thousand permutations were applied to test the significance of molecular variance partitions, and a threshold of 1% (p-value < 0.01) was assumed.

Discriminant analysis of principal components was performed using the software DAPC [82] implemented in the ADEGENET 1.3-4 package [87] since R DAPC is a multivariate approach that describes the genetic differences among groups while minimizing the differences within groups using principal components of genetic variation (PCA). A set of 60 principal components was retained as predictors for discriminant analysis and cross-validation using the default setting was performed.

A Bayesian method performed with STRUCTURE version 2.3.4 [83] was used to cluster individuals and investigate the population structure. An admixture model and correlated allele frequencies were assumed. Simulation settings were based on 10,000 discarded iterations (burn-in) and 50,000 MCMC (Monte Carlo Marcov Chain) retained replicates. The population structures were comprised of between one and six clusters (K) were simulated, and each K was tested five times. Two separate simulations were performed based respectively on (i) exclusively neutral and (ii) all markers. The online software STRUCTURE HARVESTER 0.6.92 [88] (http://taylor0.biology.ucla.edu/structureHarvester/) was used to evaluate results under the Evanno method expectations [89] while CLUMPAK server (http://clumpak.tau.ac.il) [90] was used to obtain graphical plots.

Gene flow and connectivity between sampling sites were evaluated by two separate methods using the same 2208 presumably neutral SNPs (as tested from Lositan runs in a pre-screening trial).

The first method allowed the calculation of current gene flow levels by estimating the number of first generation migrants using the Bayesian method of Rannala and Mountain [91] implemented in GENECLASS 2.0 [57] (http://www1.montpellier.inra.fr/URLB/GeneClass2/Help/). A set of 1000 MCMC iterations was produced to test the most probable site of each individual's origin based on their multi-locus neutral genotype assets. A threshold p-value of < 0.01 was used to assess the significance.

The second method to detect gene flow and connectivity was based on the coalescent calculation of historical migration rates between sampling locations using MIGRATE-n v. 3.5.1 [85]. A Bayesian method was applied [85] and F_{ST} estimates among localities were used as a baseline for calculating the demographic parameter Θ ($\Theta = 4N_e\mu$, where N_e is the effective population size and μ is the mutation rate) and the historical migration rate, M ($M = m/\mu$, where m is the immigration rate per generation and μ is the mutation rate). Negative F_{ST} were set to zero. A Brownian motion model was used, and mutation was considered constant over a set of 1000 neutral SNPs. The MCMC procedure consisted of one long chain with 500,000 recorded genealogies for each locus, with 10,000 genealogies discarded as the burn-in.

Two F_{ST}-based tests were performed to detect SNPs under positive selection. The first F_{ST}-based test was carried out using Lositan [81] (https://popgen.net/soft/lositan/), which extracted potentially selected SNPs from pairwise comparisons (PIL vs. MAT, PIL vs. HST, FAH vs. MAT, FAH vs. HST, and HST vs. MAT) of PCB polluted sites in NBH against the reference sampling site in Mattapoisett Harbor (MAT). The Lositan settings were the same used for the preliminary analysis in which neutral and non-neutral SNPs were identified. SNPs under potentially positive natural selection (candidate outliers) were considered only if SNPs were significant in every replicate. The second F_{ST}-based test was carried out in Arlequin [77,92] using the Hierarchical Island method (HIM). This method differs from the Lositan test by evaluating the impact of a hierarchical structure among the collected samples. The hierarchical structure considered in this test took into account site-by-site variation in PCBs. The first cluster included samples that are experiencing high pollution (PIL and FAH), the second cluster included a sample (HST) with low PCB concentrations, and the third cluster was represented by the reference samples collected in the Mattapoisett Harbor (MAT). HIM settings were characterized by simulating 100 demes with 10,000 coalescent iterations under the assumption of the hierarchical structure mentioned above. A significance threshold of 1% (p-value < 0.01) and a 1% FDR were applied to increase the level of accuracy in avoiding false positive tests.

The last test to identify candidate outliers was based on a statistical procedure in Bayenv2 [39] (https://gcbias.org/bayenv/). This method assesses genomic evidence for natural selection to specific environments by performing locus-by-locus Bayesian analysis to detect linear relationships between allele frequencies and environmental variables, assuming non-linear allele frequency changes [39]. The variance-covariance matrix was built up using 500,000 MCMC iterations in a simulation that involved only presumed neutral SNPs (candidate outliers and linked SNPs were excluded from the dataset). In order to apply a more rigorous approach, the last 200 variance-covariance matrices were averaged, and the resulting matrix was used to test environmental correlation against all SNPs (4128). A set of 10 replicates of environmental correlation were produced, and logarithmic values of Bayesian Factors (BF) were averaged in each locus. Only SNPs that showed average logarithmic values of Bayesian Factors (BF) larger than one ($Log_{10}BF > 1$) were considered significant and potentially evolving by natural selection.

An intersection analysis was performed to find candidate outliers jointly identified by all these methods. A Venn diagram, that represents the candidate outlier's intersections, was produced.

To annotate potentially selectively important genes, a FASTA file containing candidate outlier sequences was loaded in the online software BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastnandPAGE_TYPE=BlastSearchandLINK_LOC=blasthome) [93] good matches (E-values < 0.0001) between the queries and any similar sequence in GenBank were searched using the *blastn* algorithm. The annotated sequences from candidate outliers were further used for a UniProt (http://www.uniprot.org) search. The search was aimed to find Human UniProt code associated with candidate outliers found in *Fundulus* and other organisms' annotations. These UniProt codes

were subsequently used in DAVID 6.7 [94] (https://david.ncifcrf.gov/content.jsp?file=citation.htm) to produce Functional Annotation Clustering and provide further information about the molecular and biological pathways in which annotated genes are involved.

5. Conclusions

We found that samples collected within a radius of 20 km (shoreline distance) in New Bedford Harbor experiencing a steep variation in aquatic pollution, exhibited ~530 SNPs with unexpected genetic variation patterns. The high mobility and connectivity within this striped killifish population made it difficult to identify changes in overall genetic variation (H_E or π) associated with the role of pollutants. The enrichment analysis of loci likely under selective processes revealed that most of these SNPs occur in genes that are presumably associated with pollution resistance and are involved in physiological processes that can determine an inherited tolerance of this population to pollutants. The unbiased approach of GBS that detected functionally-relevant loci suggests that, even in a species with large migration rates, pollution can significantly affect allele frequencies. Yet, the lack of agreement among different tests for detecting SNPs under selection might suggest how these tests are sensitive to different evolutionary forces and/or might lack the power to identify important loci (type II error) or mis-identify genes (type I error). In addition, there is the possibility that other environmental features may have reduced the power of the genomic scan analyses, which limits the ability to detect more loci associated with pollutant resistance. Overall, the evidence still suggests non-neutral divergence among samples that occurs in spite of a small geographic scale, a panmictic population, and large migration rates.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/5/1129/s1.

Author Contributions: This research was designed jointly by M.F.O. and D.L.C. These Authors contributed toward developing reagents, sequencing costs, analytical tools, analyzing data, and writing the manuscript. P.R. contributed to this research by performing the computational analyses and writing the manuscript. X.D. contributed to the paper by performing lab and computational analyses.

Funding: National Science Foundation (NSF) grants MCB-1158241 and IOS-1147042. Data analysis was aided by reference to a draft *F. heteroclitus* genome, which was supported by funding from NSF (collaborative research grants DEB-1120512, DEB-1265282, DEB-1120013, DEB-1120263, DEB-1120333, DEB-1120398).

Acknowledgments: We wish to thank Vincenzo Caputo Barucchi and the Universita' Politecnica delle Marche (Ancona, Italy) for opening the opportunity to start a collaboration with the University of Miami and providing the funding for the postdoc position granted to Paolo Ruggeri to work on this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cope, W.G.; Leidy, R.B.; Hodgson, E. Classes of toxicants: Use classes. In *A Textbook of Modern Toxicology*; Hodgson, E., Ed.; John Wiley and Sons Inc.: New York, NY, USA, 2004; pp. 49–74.
2. Wisdom, H.L. Contamination of the marine environment from land–based sources. *Mar. Pollut. Bull.* **1992**, *25*, 32–36.
3. GESAMP. Reducing environmental impacts of coastal aquaculture. *Rep. Stud.* **1991**, *47*.
4. Johnston, E.L.; Mayer–Pinto, M.; Crowe, T.P.; Frid, C. Chemical contaminant effects on marine ecosystem functioning. *J. Appl. Ecol.* **2015**, *52*, 140–149. [CrossRef]
5. Hamilton, P.B.; Cowx, I.G.; Oleksiak, M.F.; Grahn, M.; Stevens, J.R.; Carvalho, G.R.; Nicol, E.; Tyler, C.R. Population-level consequences for wild fish exposed to sublethal concentrations of chemicals–a critical review. *Fish Fish.* **2015**, *17*, 545–566. [CrossRef]
6. Oziolor, E.M.; Matson, C.W. Evolutionary toxicology: Population adaptation in response to anthropogenic pollution. In *Extremophile Fishes: Ecology, Evolution, and Physiology of Teleosts in Extreme Environments*; Riesch, R., Tobler, M., Plath, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 247–277.
7. Ribeiro, R.; Baird, D.J.; Soares, A.M.; Lopes, I. Contaminant driven genetic erosion: A case study with *Daphnia longispina*. *Environ. Toxicol. Chem.* **2012**, *31*, 977–982. [CrossRef] [PubMed]

8. Dallinger, R.; Hockner, M. Evolutionary concepts in ecotoxicology: Tracing the genetic background of differential cadmium sensitivities in invertebrate lineages. *Ecotoxicology* **2013**, *22*, 767–778. [CrossRef] [PubMed]
9. Williams, L.M.; Oleksiak, M.F. Signatures of selection in natural populations adapted to chronic pollution. *BMC Evol. Biol.* **2008**, *8*, 282. [CrossRef] [PubMed]
10. Bozinovic, G.; Oleksiak, M.F. Genomic approaches with natural fish populations from polluted environments. *Environ. Toxicol. Chem.* **2011**, *30*, 283–289. [CrossRef] [PubMed]
11. Whitehead, A.; Galvez, F.; Zhang, S.; Williams, L.M.; Oleksiak, M.F. Functional genomics of physiological plasticity and local adaptation in killifish. *J. Hered.* **2011**, *102*, 499–511. [CrossRef] [PubMed]
12. Laporte, M.; Pavey, S.A.; Rougeux, C.; Pierron, F.; Lauzent, M.; Budzinki, H.; Labadie, P.; Geneste, E.; Couture, P.; Baudrimont, M.; Bernatchez, L. RAD sequencing reveals within–generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Mol. Ecol.* **2016**, *25*, 219–237. [CrossRef] [PubMed]
13. Narum, S.R.; Buerkle, C.A.; Davey, J.W.; Miller, M.R.; Hohenlohe, P.A. Genotyping–by–sequencing in ecological and conservation genomics. *Mol. Ecol.* **2013**, *22*, 2841–2847. [CrossRef] [PubMed]
14. Savolainen, O.; Lascoux, M.; Merila, J. Ecological genomics of local adaptation. *Nat. Rev. Genet.* **2013**, *14*, 807–820. [CrossRef] [PubMed]
15. Ellegren, H. Genome sequencing and population genomics in non–model organisms. *Trends Ecol. Evol.* **2014**, *29*, 51–63. [CrossRef] [PubMed]
16. Reid, N.M.; Whitehead, A. Functional genomics to assess biological responses to marine pollution at physiological and evolutionary timescales: Toward a vision of predictive ecotoxicology. *Brief Funct. Genom.* **2015**, *15*, 358–364. [CrossRef] [PubMed]
17. Svingen, T.; Vinggaard, A.M. The risk of chemical cocktail effects and how to deal with the issue. *J. Epidemiol. Community Health* **2016**, *70*, 322–323. [CrossRef] [PubMed]
18. Tiffin, P.; Ross–Ibarra, J. Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.* **2014**, *29*, 673–680. [CrossRef] [PubMed]
19. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping–by–sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [CrossRef] [PubMed]
20. Pruell, R.J.; Norwood, C.B.; Bowen, R.D.; Boothman, W.S.; Rogerson, P.F.; Hackett, M.; Butterworth, B.C. Geochemical study of sediment contamination in New Bedford Harbor, Massachusetts. *Mar. Environ. Res.* **1990**, *29*, 77–101. [CrossRef]
21. Nelson, W.G.; Bergen, B.J. The New Bedford Harbor Superfund site long–term monitoring program (1993–2009). *Environ. Monit. Assess.* **2012**, *184*, 7531–7550. [CrossRef] [PubMed]
22. Kidd, K.A.; Blanchfield, P.J.; Mills, K.H.; Palace, V.P.; Evans, R.E.; Lazorchak, J.M.; Flick, R.W. Collapse of a fish population after exposure to a synthetic estrogen. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8897–8901. [CrossRef] [PubMed]
23. Oleksiak, M.F.; Karchner, S.I.; Jenny, M.J.; Franks, D.G.; Welch, D.B.M.; Hahn, M.E. 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 Genom.* **2011**, *12*, 263. [CrossRef] [PubMed]
24. Incardona, J.P.; Scholtz, N.L. The influence of heart developmental anatomy on cardiotoxicity–based adverse outcome pathways in fish. *Aquat. Toxicol.* **2016**, *177*, 515–525. [CrossRef] [PubMed]
25. Nacci, D.E.; Champlin, D.; Jayaraman, S. Adaptation of the estuarine fish *Fundulus heteroclitus* (Atlantic Killifish) to polychlorinated biphenyls (PCBs). *Estuar. Coast.* **2010**, *33*, 853–864. [CrossRef]
26. Wirgin, I.; Roy, N.K.; Loftus, M.; Chambers, R.C.; Franks, D.G.; Hahn, M.E. Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson River. *Science* **2011**, *331*, 1322–1325. [CrossRef] [PubMed]
27. Maguire_Group. Essential Fish Habitat (EFH) Assessment New Bedford/Fairhaven Harbor Massachusetts March 2002. *Mass. Off. Coastl Zone Manag.* **2002**.
28. Du, X.; Crawford, D.L.; Oleksiak, M.F. Effects of anthropogenic pollution on the oxidative phosphorylation pathway of hepatocytes from natural populations of *Fundulus heteroclitus*. *Aquat. Toxicol.* **2015**, *165*, 231–240. [CrossRef] [PubMed]
29. Du, X.; Crawford, D.L.; Nacci, D.; Oleksiak, M.F. Heritable adaptation of oxidative phosphorylation pathway in pollutant resistant *Fundulus heteroclitus* population. *Aquat. Toxicol.* **2016**, *177*, 44–50. [CrossRef] [PubMed]

30. Burnett, K.G.; Bain, L.J.; Baldwin, W.S.; Callard, G.V.; Di Giulio, R.T.; Evans, D.H.; Gomez-Chiarri, M.; Hahn, M.E.; Hoover, C.A.; Karchner, S.I.; et al. Fundulus as the premier teleost model in environmental biology: Opportunities for new insights using genomics. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2007**, *2*, 257–286. [CrossRef] [PubMed]
31. Oleksiak, M.F. Changes in gene expression due to chronic exposure to environmental pollutants. *Aquat. Toxicol.* **2008**, *90*, 161–171. [CrossRef] [PubMed]
32. Whitehead, A.; Roach, J.L.; Zhang, S.; Galvez, F. Genomic mechanisms of evolved physiological plasticity in killifish distributed along an environmental salinity gradient. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6193–6198. [CrossRef] [PubMed]
33. Baris, T.Z.; Crawford, D.L.; Oleksiak, M.F. Acclimation and acute temperature effects on population differences in oxidative phosphorylation. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2016**, *310*, R185–R196. [CrossRef] [PubMed]
34. Williams, L.M.; Oleksiak, M.F. Evolutionary and functional analyses of cytochrome P4501A promoter polymorphisms in natural populations. *Mol. Ecol.* **2011**, *20*, 5236–5247. [CrossRef] [PubMed]
35. Reitzel, A.M.; Karchner, S.I.; Franks, D.G.; Evans, B.R.; Nacci, D.; Champlin, D.; Vieira, V.M.; Hahn, M.E. Genetic variation at aryl hydrocarbon receptor (AHR) loci in populations of Atlantic killifish (*Fundulus heteroclitus*) inhabiting polluted and reference habitats. *BMC Evol. Biol.* **2014**, *14*, 1–17. [CrossRef] [PubMed]
36. Roark, S.A.; Nacci, D.; Coiro, L.; Champlin, D.; Guttman, S.I. Population genetic structure of a nonmigratory estuarine fish (*Fundulus heteroclitus*) across a strong gradient of polychlorinated biphenyl contamination. *Environ. Toxicol. Chem.* **2005**, *24*, 717–725. [CrossRef] [PubMed]
37. Abraham, B.J. Species profiles, life histories and environmental requirements of coastal fishes and invertebrates (Mid–Atlantic). MUMMICHOG AND STRIPED KILLIFISH. DTIC Document (1985).
38. Giska, I.; Babik, W.; van Gestel, C.A.; van Straalen, N.M.; Laskowski, R. Genome–wide genetic diversity of rove beetle populations along a metal pollution gradient. *Ecotox. Environ. Safe* **2015**, *119*, 98–105. [CrossRef] [PubMed]
39. Günther, T.; Coop, G. Robust identification of local adaptation from allele frequencies. *Genetics* **2013**, *195*, 205–220. [CrossRef] [PubMed]
40. Bickham, J.W. The four cornerstones of evolutionary toxicology. *Ecotoxicology* **2011**, *20*, 497–502. [CrossRef] [PubMed]
41. Narum, S.R.; Campbell, N.R.; Meyer, K.A.; Miller, M.R.; Hardy, R.W. Thermal adaptation and acclimation of ectotherms from differing aquatic climates. *Mol. Ecol.* **2013**, *22*, 3090–3097. [CrossRef] [PubMed]
42. Ravinet, M.; Westram, A.; Johannesson, K.; Butlin, R.; Andre, C.; Panova, M. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol. Ecol.* **2016**, *25*, 287–305. [CrossRef] [PubMed]
43. van Straalen, N.; Timmermans, M. Genetic variation in toxicant–stressed populations: An evaluation of the "genetic erosion" hypothesis. *Hum. Ecol. Risk Assess.* **2002**, *8*, 983–1002. [CrossRef]
44. Boorse, D.; Storlie, C. Diel migration of invertebrates and fishesin Dean Creek, Sapelo Island, GA (1993).
45. Sexton, J.P.; Hangartner, S.B.; Hoffmann, A.A. Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution* **2014**, *68*, 1–15. [CrossRef] [PubMed]
46. Sexton, J.P.; Strauss, S.Y.; Rice, K.J. Gene flow increases fitness at the warm edge of a species' range. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11704–11709. [CrossRef] [PubMed]
47. Bijlsma, R.; Loeschcke, V. Genetic erosion impedes adaptive responses to stressful environments. *Evol. Appl.* **2012**, *5*, 117–129. [CrossRef] [PubMed]
48. Stinchcombe, J.R.; Hoekstra, H.E. Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity* **2008**, *100*, 158–170. [CrossRef] [PubMed]
49. Lotterhos, K.E.; Whitlock, M.C. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol. Ecol.* **2014**, *23*, 2178–2192. [CrossRef] [PubMed]
50. Whitlock, M.C. Modern approaches to local adaptation. *Am. Nat.* **2015**, *186*, S1–S4. [CrossRef] [PubMed]
51. Whitlock, M.C.; Lotterhos, K.E. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* **2015**, *24*, 1031–1046.
52. Narum, S.R.; Hess, J.E. Comparison of F(ST) outlier tests for SNP loci under selection. *Mol. Ecol. Res.* **2011**, *11*, 184–194. [CrossRef] [PubMed]
53. Nosil, P.; Funk, D.J.; Ortiz–Barrientos, D. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **2009**, *9*, 375–402. [CrossRef] [PubMed]

54. Kemper, K.E.; Saxton, S.J.; Bolormaa, S.; Hayes, B.J.; Goddard, M.E. Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics* **2014**, *15*, 246. [CrossRef] [PubMed]
55. Le Corre, V.; Kremer, A. The genetic differentiation at quantitative trait loci under local adaptation. *Mol. Ecol.* **2012**, *20*, 1548–1566. [CrossRef] [PubMed]
56. De Mita, S.; Thuillet, A.C.; Gay, L.; Ahmadi, N.; Manel, S.; Ronfort, J.; Vigouroux, Y. Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol. Ecol.* **2013**, *22*, 1383–1399. [CrossRef] [PubMed]
57. Stankowski, S.; Sobel, J.M.; Streisfeld, M.A. Geographic cline analysis as a tool for studying genome–wide variation: A case study of pollinator–mediated divergence in a monkeyflower. *Mol. Ecol.* **2017**, *26*, 107–122. [CrossRef] [PubMed]
58. Babin, C.; Gagnaire, P.A.; Pavey, S.A.; Bernatchez, L. RAD–Seq reveals patterns of additive polygenic variation caused by spatially–varying selection in the American Eel (Anguilla rostrata). *Genome Biol. Evol.* **2017**, *9*, 2974–2986. [CrossRef] [PubMed]
59. Gascon, M.; Morales, E.; Sunyer, J.; Vrijheid, M. Effects of persistent organic pollutants on the developing respiratory and immune systems: A systematic review. *Environ. Int.* **2013**, *52*, 51–65. [CrossRef] [PubMed]
60. Taylor, K.W.; Novak, R.F.; Anderson, H.A.; Birnbaum, L.S.; Blystone, C.; DeVito, M.; Jacobs, D.; Lee, D.H. Evaluation of the association between persistent organic pollutants (POPs) and diabetes in epidemiological studies: A national toxicology program workshop review. *Environ. Health Persp.* **2013**, *121*, 774–783. [CrossRef] [PubMed]
61. Perkins, J.T.; Petriello, M.C.; Newsome, B.J.; Hennig, B. Polychlorinated biphenyls and links to cardiovascular disease. *Environ. Sci. Pollut. Res. Int.* **2016**, *23*, 2160–2172. [CrossRef] [PubMed]
62. Proestou, D.A.; Flight, P.; Champlin, D.; Nacci, D. Targeted approach to identify genetic loci associated with evolved dioxin tolerance in the Atlantic Killifish (*Fundulus heteroclitus*). *BMC Evol. Biol.* **2014**, *14*, 1–17. [CrossRef] [PubMed]
63. Hutt, D.M.; Baltz, J.M.; Ngsee, J.K. Synaptotagmin VI and VIII and syntaxin 2 are essential for the mouse sperm acrosome reaction. *J. Biol. Chem.* **2005**, *208*, 20197–20203. [CrossRef] [PubMed]
64. Souza, M.S.; Hallgren, P.; Balseiro, E.; Hansson, L.A. Low concentrations, potential ecological consequences: Synthetic estrogens alter life–history and demographic structures of aquatic invertebrates. *Environ. Pollut.* **2013**, *178*, 237–243. [CrossRef] [PubMed]
65. Jarque, S.; Quiros, L.; Grimalt, J.O.; Gallego, E.; Catalan, J.; Lackner, R.; Piña, B. Background fish feminization effects in European remote sites. *Sci. Rep.* **2015**, *5*, 11292. [CrossRef] [PubMed]
66. Salama, J.; Chakraborty, T.R.; Ng, L.; Gore, A.C. Effects of Polychlorinated Biphenyls on Estrogen Receptor–ß Expression in the Anteroventral Periventricular Nucleus. *Environ. Health Persp.* **2003**, *111*, 1278–1282.
67. Lin, C.H.; Huang, C.L.; Chuang, M.C.; Wang, Y.J.; Chen, D.R.; Chen, S.T.; Lin, P.H. Protective role of estrogen receptor–alpha on lower chlorinated PCB congener–induced DNA damage and repair in human tumoral breast cells. *Toxicol. Lett.* **2009**, *188*, 11–19. [CrossRef] [PubMed]
68. Greytak, S.R.; Callard, G.V. Cloning of three estrogen receptors (ER) from killifish (*Fundulus heteroclitus*): Differences in populations from polluted and reference environments. *Gen. Comp. Endocr.* **2007**, *150*, 174–188. [CrossRef] [PubMed]
69. Greytak, S.R.; Tarrant, A.M.; Nacci, D.; Hahn, M.E.; Callard, G.V. Estrogen responses in killifish (*Fundulus heteroclitus*) from polluted and unpolluted environments are site– and gene–specific. *Aquat. Toxicol.* **2010**, *99*, 291–299. [CrossRef] [PubMed]
70. Uno, T.; Ishizuka, M.; Itakura, T. Cytochrome P450 (CYP) in fish. *Environ. Toxicol. Pharmacol.* **2012**, *34*, 1–13. [CrossRef] [PubMed]
71. Calder, P.C. Polyunsaturated fatty acids and inflammatory processes: New twists in an old tale. *Biochimie* **2009**, *91*, 791–795. [CrossRef] [PubMed]
72. Powell, W.H.; Bright, R.; Bello, S.M.; Hahn, M.E. Developmental and Tissue–Specific Expression of AHR1, AHR2, and ARNT2 in Dioxin–Sensitive and –Resistant Populations of the Marine Fish *Fundulus heteroclitus*. *Toxicoll. Sci.* **2000**, *57*, 229–239. [CrossRef]
73. Nacci, D.; Huber, M.; Champlin, D.; Jayaraman, S.; Cohen, S.; Gauger, E.; Fong, A.; Gomez–Chiari, M. Evolution of tolerance to PCBs and susceptibility to a bacterial pathogen (*Vibrio harveyi)* in Atlantic killifish (*Fundulus heteroclitus*) from New Bedford (MA, USA) harbour. *Environ. Poll.* **2009**, *157*, 857–864. [CrossRef] [PubMed]

74. Ivanova, N.V.; Dewaard, J.R.; Hebert, P.D.N. An inexpensive, automation–friendly protocol for recovering high–quality DNA. *Mol. Ecol. Notes* **2006**, *6*, 998–1002. [CrossRef]
75. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevenssss, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef] [PubMed]
76. Reid, N.M.; Jackson, C.E.; Gilbert, D.; Minx, P.; Montague, M.J.; Hampton, T.H.; Helfrich, L.W.; King, B.L.; Nacci, D.E.; Aluru, N.; et al. The Landscape of Extreme Genomic Variation in the Highly Adaptable Atlantic Killifish. *Genome Biol. Evol.* **2017**, *9*, 659–676. [CrossRef] [PubMed]
77. Excoffier, L.; Lischer, H.E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* **2010**, *10*, 564–567.
78. Nunez, J.C.; Seale, T.P.; Fraser, M.A.; Burton, T.L.; Fortson, T.L.; Hoover, D.; Travis, J.; Oleksiak, M.F.; Crawford, D.L. Population Genomics of the Euryhaline Teleost *Poecilia latipinna*. *PLoS ONE* **2015**, *10*, e0137077. [CrossRef] [PubMed]
79. Crawford, D.L.; Oleksiak, M.F. Ecological population genomics in the marine environment. *Brief Funct. Genomics* **2016**, elw008. [CrossRef] [PubMed]
80. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [CrossRef] [PubMed]
81. Antao, T.; Lopes, A.; Lopes, R.J.; Beja–Pereira, A.; Luikart, G. LOSITAN: A workbench to detect molecular adaptation based on a Fst–outlier method. *BMC Bioinformatics* **2008**, *9*, 323. [CrossRef] [PubMed]
82. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics* **2010**, *11*, 1–15. [CrossRef] [PubMed]
83. Pritchard, J.K.; Stephens, M.; Donnely, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959. [PubMed]
84. Piry, S.; Alapetite, A.; Cornuet, J.M.; Paetkau, D.; Baudouin, L.; Estoup, A. GENECLASS2: A software for genetic assignment and first–generation migrant detection. *J. Hered.* **2004**, *95*, 536–539. [CrossRef] [PubMed]
85. Beerli, P. Comparison of Bayesian and maximum–likelihood inference of population genetic parameters. *Bioinformatics* **2006**, *22*, 341–345. [CrossRef] [PubMed]
86. Weir, B.; Cockerham, C.C. Estimating F–statistics for the analysis of population structure. *Evolution* **1984**, *38*, 1358–1370. [PubMed]
87. Jombart, T.; Ahmed, I. Adegenet 1.3–1: New tools for the analysis of genome–wide SNP data. *Bioinformatics* **2011**, *27*, 3070–3071. [CrossRef] [PubMed]
88. Earl, D.A.; vonHoldt, B.M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **2011**, *4*, 359–361. [CrossRef]
89. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2011**, *14*, 2611–2620. [CrossRef] [PubMed]
90. Kopelman, N.M.; Mayzel, J.; Jakobsson, M.; Rosenberg, N.A.; Mayrose, I. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Res.* **2015**, *15*, 1179–1191. [CrossRef] [PubMed]
91. Rannala, B.; Mountain, J.L. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9197–9201. [CrossRef]
92. Excoffier, L.; Hofer, T.; Foll, M. Detecting loci under selection in a hierarchically structured population. *Heredity* **2009**, *103*, 285–298. [CrossRef] [PubMed]
93. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. "BLAST+: Architecture and applications. *BMC Bioinformatics* **2008**, *10*, 421. [CrossRef] [PubMed]
94. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat. Protoc.* **2009**, *4*, 44–57. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).