ECOLOGICAL GENOMICS

The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish

Noah M. Reid,[1] Dina A. Proestou,[2] Bryan W. Clark,[3] Wesley C. Warren,[4] John K. Colbourne,[5] Joseph R. Shaw,[5,6] Sibel I. Karchner,[7,8] Mark E. Hahn,[7,8] Diane Nacci,[9] Marjorie F. Oleksiak,[10] Douglas L. Crawford,[10] Andrew Whitehead[1]*

Atlantic killifish populations have rapidly adapted to normally lethal levels of pollution in four urban estuaries. Through analysis of 384 whole killifish genome sequences and comparative transcriptomics in four pairs of sensitive and tolerant populations, we identify the aryl hydrocarbon receptor–based signaling pathway as a shared target of selection. This suggests evolutionary constraint on adaptive solutions to complex toxicant mixtures at each site. However, distinct molecular variants apparently contribute to adaptive pathway modification among tolerant populations. Selection also targets other toxicity-mediating genes and genes of connected signaling pathways; this indicates complex tolerance phenotypes and potentially compensatory adaptations. Molecular changes are consistent with selection on standing genetic variation. In killifish, high nucleotide diversity has likely been a crucial substrate for selective sweeps to propel rapid adaptation.

The current pace of environmental change may exceed the maximum rate of evolutionary change for many species (*1*), yet little is known of the circumstances and mechanisms through which evolution might rescue species at risk of decline (*2*). The Atlantic killifish, *Fundulus heteroclitus*, is nonmigratory and abundant in U.S. Atlantic coast salt-marsh estuaries (*3*), including sites contaminated with complex mixtures of persistent industrial pollutants (Fig. 1A) that have reached lethal levels in recent decades (*4*). Some killifish populations resident in polluted sites exhibit inherited tolerance to normally lethal levels of these highly toxic pollutants (*5*) (Fig. 1B). To understand the genetics of rapid adaptation to radical environmental change in wild populations, we sequenced complete genomes from 43 to 50 individuals from each of eight populations (Fig. 1A and table S1): four tolerant (T) populations from highly polluted sites, each paired with a nearby reference [sensitive (S)] population. We combined these data with RNA sequencing (RNA-seq) to uncover unique and shared functional pathways and adaptive signatures of selection across populations.

Genomes from T1 and S1 populations were sequenced to 7-fold coverage per individual and the remaining populations, to 0.6-fold coverage (*6*). Genetic variation is strongly partitioned by geography (Fig. 1C); northern populations (T1, S1, T2, S2, T3, and S3) form a cluster distinct from southern populations (T4 and S4), consistent with their known phylogeography (*7*). In tolerant populations, nucleotide diversity is reduced genome-wide, and Tajima's D is shifted positive, relative to sensitive population counterparts (fig. S1); these indicate reduced effective population size in polluted sites. Tolerant-sensitive (T-S) population pairs share the most similar genetic backgrounds, and the fixation index (F_{ST}) is low between them (0.01 to 0.08) (fig. S2). We conclude that tolerant populations are recently and independently derived from local gene pools.

We identified genomic regions that are candidates for pollution tolerance (table S2 and fig. S3) by defining outlier regions as 5-kb windows that fell in the extreme 0.1% tails (for pi and Tajima's D) and 99.9% tails (for F_{ST}) of null distributions simulated from demographic models estimated from the data (*6*). Most outlier regions are small (52 to 69 kb), although a few are up to ~1.8 Mb (fig. S4). For each T-S population pair, signatures of selection are skewed in prevalence toward the tolerant population (fig. S5). Most outliers are specific to a tolerant population (0.5% of 5-kb outlier windows are shared) (fig. S6). However, loci showing the strongest signals of recent selection [highly ranked outliers (*6*)] are shared (Fig. 2A), suggesting convergent evolution for pollution tolerance. Within these shared outliers are key genes involved in the aryl hydrocarbon receptor (AHR) signaling pathway (*AHR2a*, *AHR1a*, *AIP*, and *CYP1A*) (Fig. 2B).

The importance of these outliers is supported by transcriptomics. When sensitive and tolerant populations were raised in a common clean environment for two generations and embryos were challenged with a model toxic pollutant, the polychlorinated biphenyl (PCB) 3,3′,4,4′,5-pentachlorobiphenyl (PCB 126)–tolerant populations exhibit reduced inducibility of AHR-regulated genes (Fig. 2C). The 70 genes up-regulated in response to pollutant challenge in sensitive populations, but not in tolerant populations (table S3), are enriched for those regulated by the AHR signaling pathway ($P < 0.0001$). Impaired AHR signaling is most apparent with the canonical transcriptional targets of AHR (Fig. 2C and table

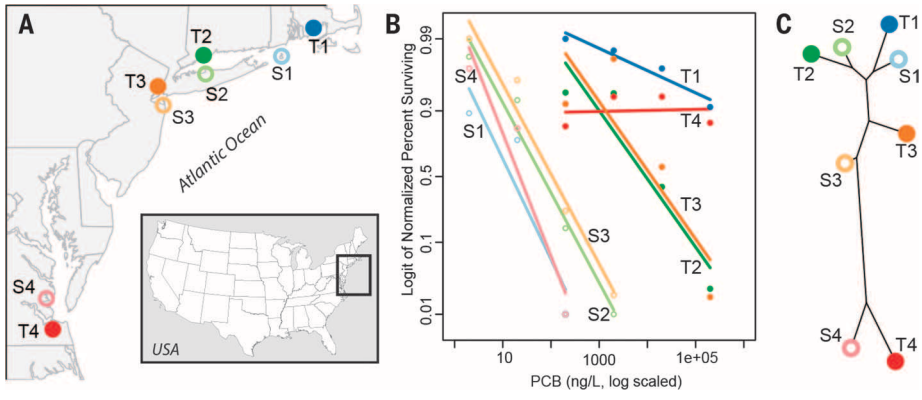

Fig. 1. Focal *F. heteroclitus* populations. (**A**) Locations of pollution-tolerant ("T"; bold tone, filled circles) and sensitive ("S"; pastel tone, open circles) population pairs numbered from north to south. (**B**) Population variation in larval survival (linear regression of logit survival to 7 days post hatch) after two generations reared in a common environment, when challenged with increasing log exposure concentrations of PCB 126. Populations from polluted sites exhibit tolerance to pollutants at concentrations hundreds to thousands of times normally lethal levels. (**C**) Phylogenetic tree, estimated from genome-wide biallelic single-nucleotide polymorphism (SNP) frequencies, showing that genetic differentiation is lowest between T-S population pairs [Phylogeny Inference Package (PHYLIP) Gene Frequencies and Continuous Characters Maximum Likelihood (CONTML) module, bootstrap supports are 100 for all branches].

[1]Department of Environmental Toxicology, University of California, Davis, CA 95616, USA. [2]Agricultural Research Service, U.S. Department of Agriculture, Kingston, RI 02881, USA. [3]Oak Ridge Institute for Science and Education, Office of Research and Development, U.S. Environmental Protection Agency, Narragansett, RI 02882, USA. [4]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA. [5]School of Biosciences, University of Birmingham, Edgbaston B15 2TT, UK. [6]School of Public and Environmental Affairs, Indiana University, Bloomington, IN 47405, USA. [7]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. [8]Boston University Superfund Research Program, Boston University, Boston, MA 02118, USA. [9]Office of Research and Development, U.S. Environmental Protection Agency, Narragansett, RI 02882, USA. [10]Department of Marine Biology and Ecology, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149, USA.
*Corresponding author. Email: awhitehead@ucdavis.edu

S4). Dominant pollutants at T sites include halogenated aromatic hydrocarbons (HAHs) and polycyclic aromatic hydrocarbons (PAHs) that bind AHR and initiate aberrant signaling that causes malformations during development and subsequent embryo and larval lethality, as well as toxicity in adults (*8*). Given that the AHR pathway is repeatedly desensitized in tolerant populations (Fig. 2C) (*9*) and top-ranked outliers contain AHR pathway genes, we conclude that the AHR signaling pathway is likely a key and repeated target of natural selection in tolerant populations. This convergence suggests that adaptive options are constrained to modifications of this signaling pathway that mediates the toxicity of many HAHs and PAHs.

AHR deletions are found in tolerant populations. Four paralogs of *AHR* exist in the *F. heteroclitus* genome (*10*). Knockdown of *AHR2a* is protective of toxicity from many HAHs and PAHs [e.g., (*11*)]. Tandem paralogs *AHR2a* and *AHR1a* are within a highly ranked outlier region in all tolerant populations (Fig. 2A). Note that three tolerant populations have deletions (fig. S7) spanning *AHR2a* and *AHR1a* (Fig. 3A). In T4, a deletion is found in a single haplotypic background (fig. S8) that segregates at high frequency (81%) but is absent in S4 (Fig. 3B). In T4 individuals, RNA-seq data reveal expression of a chimeric transcript (joining exon 10 of *AHR2a* and exon 7 of *AHR1a*). In T1 and T3, different deletions spanning *AHR2a* and *AHR1a* (Fig. 3, A and B) occur in two and one haplotypic backgrounds, respectively (fig. S9). A deletion is present in at least one sensitive population (Fig. 3B), but no deletion was found in T2. Variation in this region is also associated with sensitivity to PCB toxicity in T1 (*12*) and in PCB-adapted tomcod (*13*). We thus conclude that *AHR* genes are likely common loci of selection for multiple genetic variants, including deletions, where a single deletion-associated haplotype has swept in the southern tolerant population.

The strongest signal of selection we observed is in a window that is a shared outlier in all tolerant populations [aryl hydrocarbon receptor–interacting protein (AIP) in Fig. 2A]. In northern tolerant populations, a single large (650-kb) haplotype has swept to high frequency, accompanied by reduced pi. In T4, a different haplotype has swept to high frequency (Fig. 3C). In T1 (sequenced to higher coverage), we detect recombination breakpoints, allowing identification of a core haplotype region (~100 kb) that coincides with peak differentiation (fig. S10), within which we find AIP. Variation near this locus also associates with sensitivity to PCB toxicity in T1 (*12*). AIP regulates cytoplasmic stability and cytoplasmic-nuclear shuttling of the AHR protein and thereby influences AHR signaling and regulates toxicity (*14*).

A key transcriptional target of AHR, the biotransformation gene *CYP1A*, is within a top-ranking outlier region shared by all tolerant populations (Fig. 2A). Genotypes from tolerant populations are highly differentiated from sensitive populations (Fig. 3D) and *CYP1A* single-nucleotide polymorphism (SNP) variants are linked with tolerance (*15*). In northern tolerant populations, *CYP1A* duplications have swept to high frequency, where individuals have up to eight copies of the *CYP1A* gene (Fig. 3E and figs. S7 and S11), and duplicates are present in some sensitive populations. *CYP1A* expression is not increased in northern tolerant populations (embryos) (table S4), as one might expect after duplication. However, because *AHR* knockout in rodents decreases basal *CYP1A* expression (*16*) and AHR signaling is impaired in tolerant killifish, we hypothesize that *CYP1A* duplication has been favored as a compensatory, dosage-compensating adaptation for impaired AHR signaling in northern tolerant fish. In contrast, we find no evidence of duplication in T4 (Fig. 3E), although this region retains a strong signature of selection (Fig. 2A) and is highly differentiated from S4 (Fig. 3D). PAHs primarily contaminate T4, and these chemicals interact

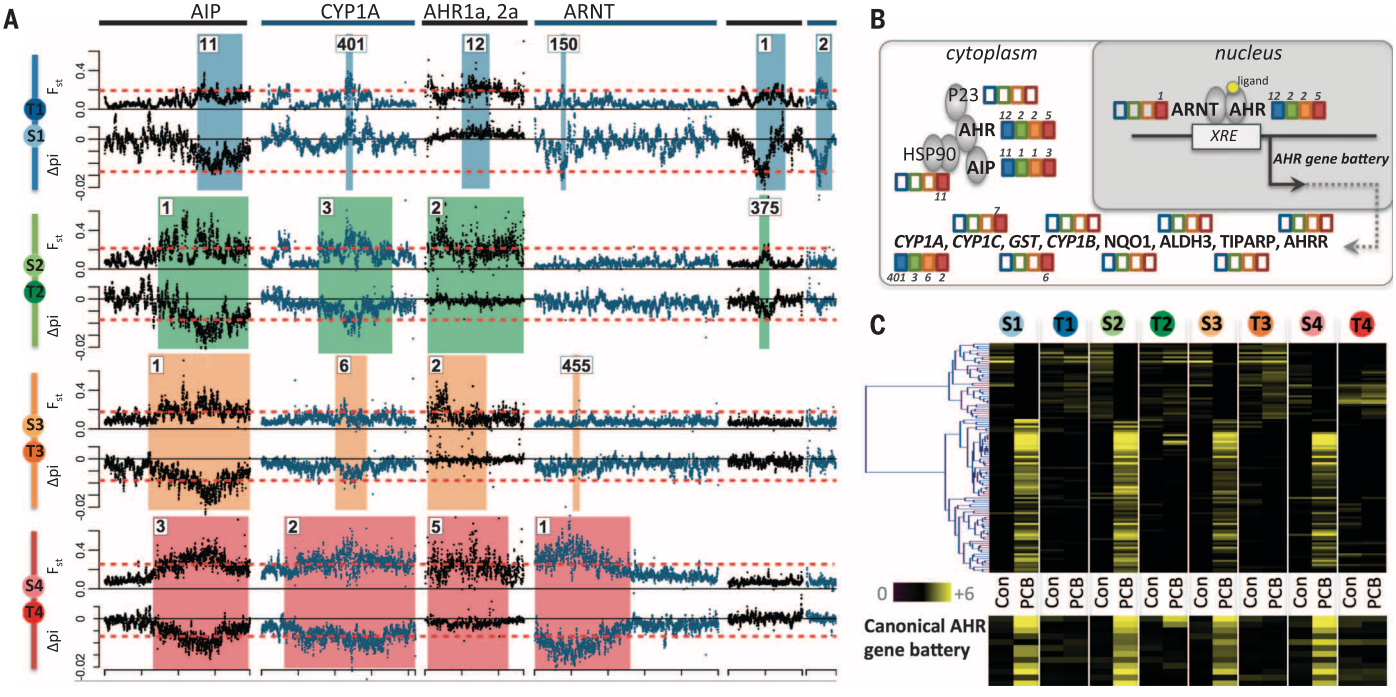

Fig. 2. Patterns of structural and functional genomic divergence. (**A**) Allele frequency differentiation (F_{ST}, top) and nucleotide diversity (pi, bottom) difference (tolerant pi – sensitive pi) for each population pair studied for top-ranking outlier regions (including the top two per pair). Colored panels span the outlier region of each respective population comparison where number indicates outlier rank for each tolerant-sensitive pair. Red dashed lines indicate outlier thresholds. Each tick on *x* axis is at the 500-kb position on the scaffold, and each candidate gene name is indicated (top) for each outlier region. Top outlier regions are not colocalized in the genome (fig. S3). (**B**) Model of key molecules in the AHR signaling pathway, including regulatory genes and transcriptional targets (*AHR* gene battery). Boxes next to genes are color-coded by population pair; filled boxes indicate the gene is within a top-ranking outlier region for that pair, and number indicates ranking of the outlier region as in (A). Top-ranking outlier regions contain AHR pathway genes and tend to be outliers in all population pairs, although some significant outliers are population-specific. (**C**) Gene-expression (of developing embryos) heat map shows up-regulated genes in response to PCB 126 exposure ("PCB"; 200 ng/liter) compared with control exposure ("Con") for sensitive populations, most of which are unresponsive in tolerant populations. The bottom panel highlights genes characterized as transcriptionally activated by ligand-bound AHR (table S4).

differently with *AHR*-induced *CYP1A* than with HAHs, which dominate northern sites (*17*). We propose that different chemical pollutants acting as selective agents may govern the fate of different *CYP1A* variants between HAH- and PAH-polluted sites.

Although AHR pathway genes are among shared outliers, they are also within population-specific outlier regions. Tandem paralogs *AHR1b* and *AHR2b* are within an outlier region in T3 and T4 (fig. S12) so that all four *AHR* paralogs are within outlier regions for one or more tolerant populations. Five additional AHR pathway genes are significant outliers for only T4. Two of these (*ARNT1c* and *HSP90*) (figs. S13 and S14) directly interact with AHR protein, whereas the remaining three (*CYP1C1/1C2*, *GFRP*, and *GSTT1*) (figs. S15 and S16) are PAH biotransformation genes that are also key transcriptional targets of *AHR* (Fig. 2C). The inclusion of PAH biotransformation genes among outliers specific to T4 (primarily polluted with PAHs) likely reflects differences between cellular effects of PAHs and HAHs (*17*).

Other selective targets include genes outside of AHR signaling. Some PAHs, particularly those that are abundant only at T4, cause cardiotoxicity independent of AHR (*18*) through disruption of voltage-gated potassium channels and regulation of intracellular calcium (*19*). Note that two genes whose products form the conductance pore of the voltage-gated potassium channel (*KCNB2* and *KCNC3*) are within top-ranking outlier windows in T4 (figs. S17 and S18). Similarly, ryanodine receptor (RYR) regulates intracellular calcium, and RYR3 is within an outlier window in T4 (fig. S19). We conclude that components of the adaptive phenotype are underpinned by genes that are both related and unrelated to AHR signaling, consistent with complex adaptations to complex chemical mixtures.

Our results also suggest compensatory adaptation associated with the (potential) costs of evolved pollution tolerance. AHR signaling has diverse functions and interacts with multiple pathways, including estrogen and hypoxia signaling, regulation of cell cycle, and immune system function (*20*). Estrogen receptor 2b is within an outlier region in T2 (fig. S20), and estrogen receptor–regulated genes are enriched within outlier gene sets for all tolerant populations ($P < 0.001$) (fig. S21). Estrogen receptor is also inferred as a significant upstream regulator for genes differentially expressed between tolerant and sensitive populations ($P < 0.05$) (e.g., genes in Fig. 2C). Hypoxia-inducible factor 2α is within an outlier window in T3 (fig. S22). Interleukin and cytokine receptors are in outlier windows in T4 (fig. S23). We conclude that some components of the adaptive phenotype in polluted sites may be due to compensation for the altered AHR signaling that underlies the primary pollutant-tolerance phenotype. Selection for compensatory changes may be common following rapid adaptive evolution.

In animal models, single gene (*AHR*) knockout can protect from toxicity of some HAH or PAH compounds [e.g., (*21*)]. However, in wild killifish populations, adaptive genotypes appear complex, including multiple AHR signaling pathway elements and other genes. We suggest that this complexity arises from two primary factors. First, tolerant sites are contaminated with complex mixtures of hydrocarbons. Mixture components may interact in subtly different ways with AHR (*17*), and some exert toxicity through pathways other than AHR (*18*), such that adaptations in multiple pathways are required. Second, because many of the AHR signaling pathway genes identified here as targets of selection interact with multiple regulatory pathways (*20*), changes to their function may have deleterious consequences

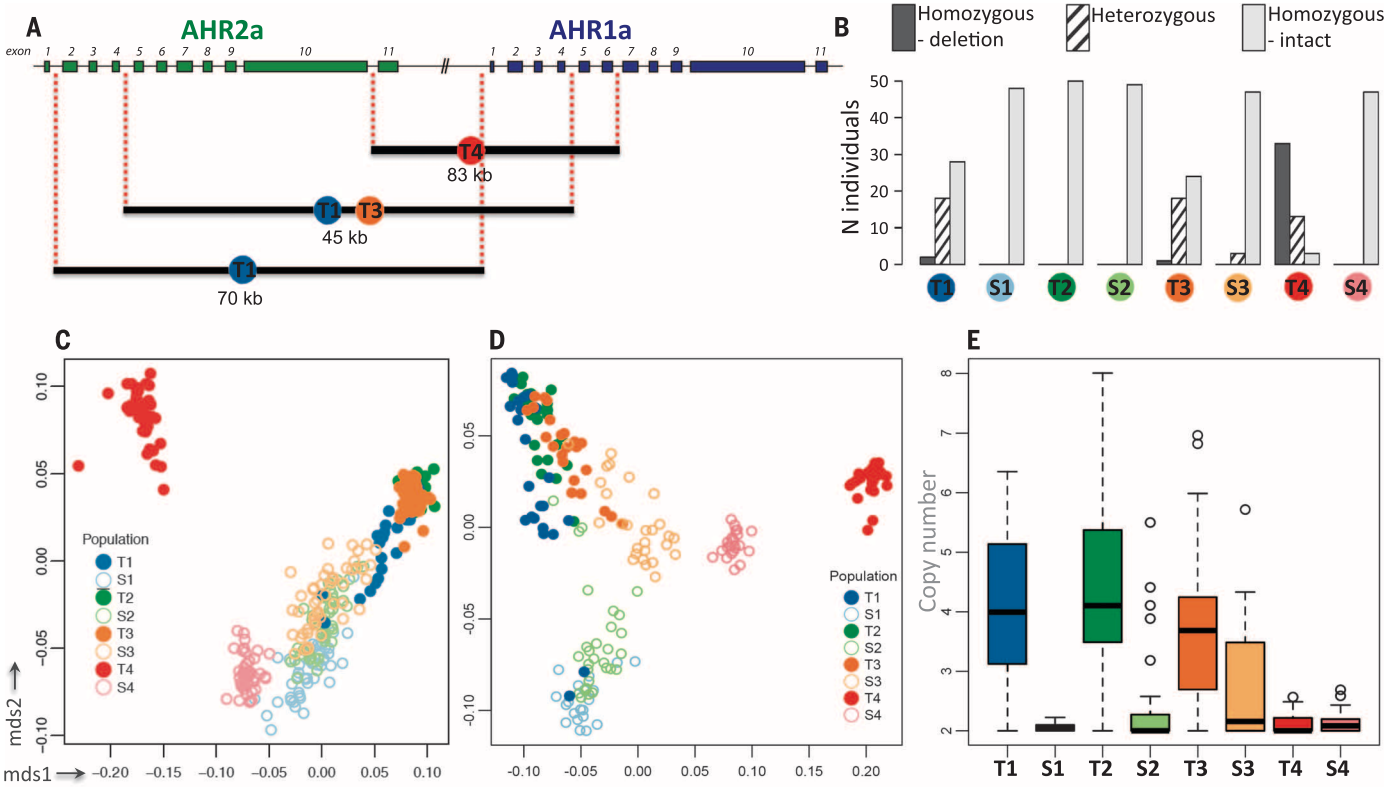

Fig. 3. Patterns of adaptive genetic variation for top-ranking and shared outliers. (**A**) Gene model of *AHR2a* and *AHR1a* (green or blue squares represent exons). Black bars indicate deleted regions present within tolerant populations. (**B**) The number of individuals homozygous for specific deletions (gray bar), heterozygous (hatched gray bar), or homozygous wild type (light bar) within each population. (**C**) Multidimensional scaling (MDS) plot of genotypic variation on the scaffold containing the *AIP* gene. (**D**) MDS plot of genotypic variation on the scaffold containing the *CYP1A* gene. (**E**) Bar plot of copy number of the duplications around *CYP1A*, where boxes, whiskers, and dots represent interquartile range, 1.5× interquartile range, and the remainder, respectively (the background diploid state includes two copies). Although the *CYP1A* region is highly differentiated in all tolerant populations (D), *CYP1A* duplications are found only in northern tolerant populations (E).

that may result in selection for compensatory change. Other changes in these highly altered estuaries may also exert selection pressures [e.g., estrogenic pollutants (*22*), hypoxia, or altered species diversity].

A fundamental question in evolutionary biology pertains to the nature and number of variants recruited by natural selection. The relative contributions of de novo variants, standing variation, and the number of competing beneficial variants depend in part on the strength of selection, its spatial patterning, existing genetic diversity and the beneficial mutation rate. Although modes of evolution can be difficult to distinguish (*23*), our data are revealing. We observe signals of convergence and divergence. Genes in the AHR pathway are repeated targets of selection, even in populations exposed to distinct chemical mixtures and separated by substantial genetic distance. This suggests adaptive constraint. Yet, different variants are often favored in different tolerant populations (e.g., *AHR* and *CYP1A*), some of which are present in sensitive populations, and common variants (e.g., large *AIP* haplotype) have rapidly swept in multiple populations of this low-dispersal fish. This suggests that selection on preexisting variants was important for rapid adaptation in killifish and that multiple molecular targets were available for selective targeting of a common pathway. The prevalence of soft sweeps is predicted to be high during rapid adaptation (*24*).

Evolutionary change relies on genetic variation that may preexist, or arise through new mutation, at a rate that scales by population size. *F. heteroclitus* at present has large population sizes (*3*) and a range of standing genetic variation (nucleotide diversity up to 0.016 for T3 and T4) that places them as one of the most diverse vertebrates (*25*). These factors suggest that Atlantic killifish have been unusually well positioned to evolve the necessary adaptations to survive in radically altered habitats.

REFERENCES AND NOTES

1. A. P. Hendry, T. J. Farrugia, M. T. Kinnison, *Mol. Ecol.* **17**, 20–29 (2008).
2. G. Bell, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120080 (2012).
3. I. Valiela, J. E. Wright, J. M. Teal, S. B. Volkmann, *Mar. Biol.* **40**, 135–144 (1977).
4. D. Nacci *et al.*, *Mar. Biol.* **134**, 9–17 (1999).
5. D. Nacci, D. Champlin, S. Jayaraman, *Estuaries Coasts* **33**, 853–864 (2010).
6. Materials and methods are available as supplementary materials on *Science* Online.
7. D. D. Duvernell, J. B. Lindmeier, K. E. Faust, A. Whitehead, *Mol. Ecol.* **17**, 1344–1360 (2008).
8. R. Pohjanvirta, *The AH Receptor in Biology and Toxicology* (Wiley, Hoboken, NJ, 2012).
9. A. Whitehead, W. Pilcher, D. Champlin, D. Nacci, *Proc. Biol. Sci.* **279**, 427–433 (2012).
10. A. M. Reitzel *et al.*, *BMC Evol. Biol.* **14**, 6 (2014).
11. B. W. Clark, C. W. Matson, D. Jung, R. T. Di Giulio, *Aquat. Toxicol.* **99**, 232–240 (2010).
12. D. Nacci, D. Proestou, D. Champlin, J. Martinson, E. R. Waits, *Mol. Ecol.* **25**, 5467–5482 (2016).
13. I. Wirgin *et al.*, *Science* **331**, 1322–1325 (2011).
14. M. Nukaya *et al.*, *J. Biol. Chem.* **285**, 35599–35605 (2010).
15. D. A. Proestou, P. Flight, D. Champlin, D. Nacci, *BMC Evol. Biol.* **14**, 7 (2014).
16. J. V. Schmidt, G. H. T. Su, J. K. Reddy, M. C. Simon, C. A. Bradfield, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6731–6736 (1996).
17. M. S. Denison, A. A. Soshilov, G. He, D. E. DeGroot, B. Zhao, *Toxicol. Sci.* **124**, 1–22 (2011).
18. J. P. Incardona *et al.*, *Environ. Health Perspect.* **113**, 1755–1762 (2005).
19. F. Brette *et al.*, *Science* **343**, 772–776 (2014).
20. T. V. Beischlag, J. Luis Morales, B. D. Hollingshead, G. H. Perdew, *Crit. Rev. Eukaryot. Gene Expr.* **18**, 207–250 (2008).
21. P. M. Fernandez-Salguero, D. M. Hilbert, S. Rudikoff, J. M. Ward, F. J. Gonzalez, *Toxicol. Appl. Pharmacol.* **140**, 173–179 (1996).
22. S. R. Greytak, A. M. Tarrant, D. Nacci, M. E. Hahn, G. V. Callard, *Aquat. Toxicol.* **99**, 291–299 (2010).
23. J. J. Berg, G. Coop, *Genetics* **201**, 707–725 (2015).
24. B. Wilson, P. Pennings, D. Petrov, *bioRxiv* 10.1101/052993 (2016).
25. E. M. Leffler *et al.*, *PLOS Biol.* **10**, e1001388 (2012).

ACKNOWLEDGMENTS

Sequence data are archived at the National Center for Biotechnology Information (BioProject PRJNA323589). Phylogenetic tree data are archived at Dryad (doi: 10.5061/dryad.68n87). We thank G. Coop, B. Counterman, D. Champlin, I. Kirby, and A. Bertrand for their valuable input. Primary support was from the NSF (collaborative research grants DEB-1265282, DEB-1120512, DEB-1120013, DEB-1120263, DEB-1120333, DEB-1120398 to J.K.C., D.L.C., M.E.H., S.I.K., M.F.O., J.R.S., W.C.W., and A.W.). Further support was provided by the National Institutes of Environmental Health Sciences (1R01ES021934-01 to A.W.; P42ES007381 to M.E.H.; R01ES019324 to J.R.S.), and the National Science Foundation (OCE-1314567 to A.W.). B.W.C. was supported by the Postdoctoral Research Program at the U.S. Environmental Protection Agency (EPA) administered by the Oak Ridge Institute for Science and Education (agreement DW92429801). The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the EPA.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6317/suppl/DC1
Materials and Methods
Figs. S1 to S26
Tables S1 to S4
References (*26*–*45*)

6 July 2016; accepted 31 October 2016
10.1126/science.aah4993

Science

The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish

Noah M. Reid, Dina A. Proestou, Bryan W. Clark, Wesley C. Warren, John K. Colbourne, Joseph R. Shaw, Sibel I. Karchner, Mark E. Hahn, Diane Nacci, Marjorie F. Oleksiak, Douglas L. Crawford and Andrew Whitehead

Science **354** (6317), 1305-1308.
DOI: 10.1126/science.aah4993

Mapping genetic adaptations to pollution

Many organisms have evolved tolerance to natural and human-generated toxins. Reid *et al.* performed a genomic analysis of killifish, geographically separate and independent populations of which have adapted recently to severe pollution (see the Perspective by Tobler and Culumber). Sequencing multiple sensitive and resistant populations revealed signals of selective sweeps for variants that may confer tolerance to toxins, some of which were shared between resistant populations. Thus, high genetic diversity in killifish seems to allow selection to act on existing variation, driving rapid adaptation to selective forces such as pollution.

Science, this issue p. 1305; see also p. 1232

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.