Title

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

Journal

Science, 354(6317)

ISSN

0036-8075

Authors

Reid, Noah M

Proestou, Dina A

Clark, Bryan W

et al.

Publication Date

2016-12-09

DOI

10.1126/science.aah4993

Peer reviewed

HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2017 June 09.

Published in final edited form as:

Science. 2016 December 09; 354(6317): 1305–1308. doi:10.1126/science.aah4993.

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], **Mark E. Hahn**[7], **Diane Nacci**[8], **Marjorie F. Oleksiak**[9], **Douglas L. Crawford**[9], and **Andrew Whitehead**[1,*]

[1]Department of Environmental Toxicology, University of California, Davis, CA 95616, USA

[2]United States Department of Agriculture, Agricultural Research Service, Kingston, RI 02881, USA

[3]Oak Ridge Institute for Science and Education at the United States Environmental Protection Agency, Office of Research and Development, 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, 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 and Boston University Superfund Research Program, Boston University, Boston, MA

[8]United States Environmental Protection Agency, Office of Research and Development, Narragansett, RI, 02882, USA

[9]Department of Marine Biology & Ecology, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149, USA

Abstract

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, indicating complex tolerance phenotypes and potentially compensatory adaptations. Molecular changes are consistent with selection on standing genetic variation. In

* awhitehead@ucdavis.edu.

Supplementary Materials

Materials and Methods
Tables S1 – S4
Figures S1 – S26
References (26 – 45)

killifish high nucleotide diversity has likely been a crucial substrate for selective sweeps to propel rapid adaptation.

One Sentence Summary

Convergent evolution of a key signaling pathway and connected pathways underlies repeated evolutionary rescue from a lethal human-altered environment.

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 non-migratory 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 43–50 individuals from each of eight populations (Fig. 1A, 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-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, S3) form a cluster distinct from southern populations (T4, 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), indicating reduced effective population size in polluted sites. Tolerant-sensitive (T-S) population pairs share the most similar genetic backgrounds and F_{ST} is low between them (0.01–0.08) (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, 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 (52–69 kb) though 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, 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 challenged with a model toxic pollutant (PCB-126), tolerant populations exhibit reduced inducibility of AHR-regulated genes (Fig. 2C). The seventy 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, Table 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 embryolarval lethality, as well as toxicity in adults (8). Given that the AHR pathway is repeatedly de-sensitized 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). Intriguingly, 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. 3A,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 also associates 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 (Fig. 2A, AIP). 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 break points, allowing identification of a core haplotype region (~100 kb) that coincides with peak differentiation (Fig. S10), within which we find aryl hydrocarbon receptor interacting protein (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, thereby influencing AHR signaling and regulating 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 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, Fig. S7, 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 following duplication. However, since 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), though 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 differently with AHR-induced CYP1A than 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.

Though 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–S14) directly interact with AHR protein, whereas the remaining three (CYP1C1/1C2, GFRP, GST-theta; Figs S15–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 reflect 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). Intriguingly, two genes whose products form the conductance pore of the voltage-gated potassium channel (KCNB2, KCNC3) are within top-ranking outlier windows in T4 (Fig. S17, 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 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, 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, 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 pre-existing 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 pre-exist, or arise through new mutation, at a rate that scales by population size. *F. heteroclitus* presently 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sequence data are archived at NCBI (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 United States National Science Foundation (collaborative research grants DEB-1265282, DEB-1120512, DEB-1120013, DEB-1120263, DEB-1120333, DEB-1120398 to JKC, DLC, MEH, SIK, MFO, JRS, WW, and AW). Further support was provided by the National Institutes of Environmental Health Sciences (1R01ES021934-01 to AW; P42ES007381 to MEH; R01ES019324 to JRS), and the National Science Foundation (OCE-1314567 to AW). BC was supported by the Postdoctoral Research Program at the US 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 US EPA.

References and Notes

1. Hendry AP, Farrugia TJ, Kinnison MT. Mol Ecol. 2008 Jan.17:20. [PubMed: 18173498]
2. Bell G. Philosophical transactions of the Royal Society of London. Series B, Biological sciences. 2013 Jan 19.368:20120080. [PubMed: 23209162]
3. Valiela I, Wright JE, Teal JM, Volkmann SB. Marine Biology. 1977; 40:135.
4. Nacci D, et al. Marine Biology. 1999 Jun.134:9.
5. Nacci D, Champlin D, Jayaraman S. Estuaries and Coasts. 2010; 33:853.
6. Materials and methods are available as supplementary materials at the Science website.
7. Duvernell DD, Lindmeier JB, Faust KE, Whitehead A. Mol Ecol. 2008; 17:1344. [PubMed: 18302693]
8. Pohjanvirta, R. The AH receptor in biology and toxicology. Hoboken, N.J.: Wiley; 2012. p. xiiip. 533
9. Whitehead A, Pilcher W, Champlin D, Nacci D. P Roy Soc B-Biol Sci. 2012 Feb 7.279:427.
10. Reitzel AM, et al. Bmc Evol Biol. 2014 Jan 14.14
11. Clark BW, Matson CW, Jung D, Di Giulio RT. Aquat. Toxicol. 2010 Aug 15.99:232. [PubMed: 20605646]
12. Nacci D, Proestou DA, Champlin D, Martinson J, Waits ER. Mol Ecol. **in press**, (available at http://onlinelibrary.wiley.com/doi/10.1111/mec.13848/abstract.
13. Wirgin I, et al. Science. 2011 Mar 11.331:1322. [PubMed: 21330491]
14. Nukaya M, et al. Journal of Biological Chemistry. 2010 Nov 12.285:35599. [PubMed: 20829355]
15. Proestou DA, Flight P, Champlin D, Nacci D. Bmc Evol Biol. 2014 Jan 14.14
16. Schmidt JV, Su GHT, Reddy JK, Simon MC, Bradfield CA. P Natl Acad Sci USA. 1996 Jun 25.93:6731.
17. Denison MS, Soshilov AA, He G, DeGroot DE, Zhao B. Toxicol Sci. 2011 Nov.124:1. [PubMed: 21908767]
18. Incardona JP, et al. Environ Health Persp. 2005 Dec.113:1755.
19. Brette F, et al. Science. 2014 Feb 14.343:772. [PubMed: 24531969]
20. Beischlag TV, Morales JL, Hollingshead BD, Perdew GH. Crit Rev Eukar Gene. 2008; 18:207.
21. Fernandez-Salguero PM, Hilbert DM, Rudikoff S, Ward JM, Gonzalez FJ. Toxicol Appl Pharm. 1996 Sep.140:173.
22. Greytak SR, Tarrant AM, Nacci D, Hahn ME, Callard GV. Aquat. Toxicol. 2010 Aug 15.99:291. [PubMed: 20570371]
23. Berg JJ, Coop G. Genetics. 2015 Oct.201:707. [PubMed: 26311475]
24. Wilson B, Pennings P, Petrov D. bioRxiv. (2016-01-01 00:00:00, 2016).
25. Leffler EM, et al. PLoS Biology. 2012 Sep.10:e1001388. [PubMed: 22984349]
26. Nacci DE, Champlin D, Coiro L, McKinney R, Jayaraman S. Environmental Toxicology and Chemistry. 2002 Jul.21:1525. [PubMed: 12109755]
27. Langmead B, Salzberg SL. Nature Methods. 2012 Apr.9:357. [PubMed: 22388286]
28. Li H. arXiv preprint arXiv:1303.3997. 2013

29. Faust GG, Hall IM. Bioinformatics. 2014 Sep 1.30:2503. [PubMed: 24812344]
30. Li H, et al. Bioinformatics. 2009; 25:2078. [PubMed: 19505943]
31. Garrison E, Marth G. arXiv preprint arXiv:1207.3907. 2012
32. Weir BS, Cockerham CC. Evolution. 1984:1358.
33. Browning SR, Browning BL. The American Journal of Human Genetics. 2007; 81:1084. [PubMed: 17924348]
34. Purcell S, et al. American Journal of Human Genetics. 2007 Sep.81:559. [PubMed: 17701901]
35. Korneliussen TS, Albrechtsen A, Nielsen R. BMC bioinformatics. 2014; 15:356. [PubMed: 25420514]
36. Hudson RR. Bioinformatics. 2002 Feb.18:337. [PubMed: 11847089]
37. Yi X, et al. Science. 2010; 329:75. [PubMed: 20595611]
38. Amores A, et al. Genetics. 2014 Jun.197:625. [PubMed: 24700104]
39. Revell LJ, Chamberlain SA. Methods in Ecology and Evolution. 2014; 5:976.
40. Bolger AM, Lohse M, Usadel B. Bioinformatics. 2014:btu170.
41. MacManes MD. bioRxiv. 2014:000422.
42. Trapnell C, Pachter L, Salzberg SL. Bioinformatics. 2009; 25:1105. [PubMed: 19289445]
43. Liao Y, Smyth GK, Shi W. Bioinformatics. 2013:btt656.
44. Lund SP, Nettleton D, McCarthy DJ, Smyth GK. Statistical applications in genetics and molecular biology. 2012; 11:8.
45. Robinson MD, McCarthy DJ, Smyth GK. Bioinformatics. 2010; 26:139. [PubMed: 19910308]

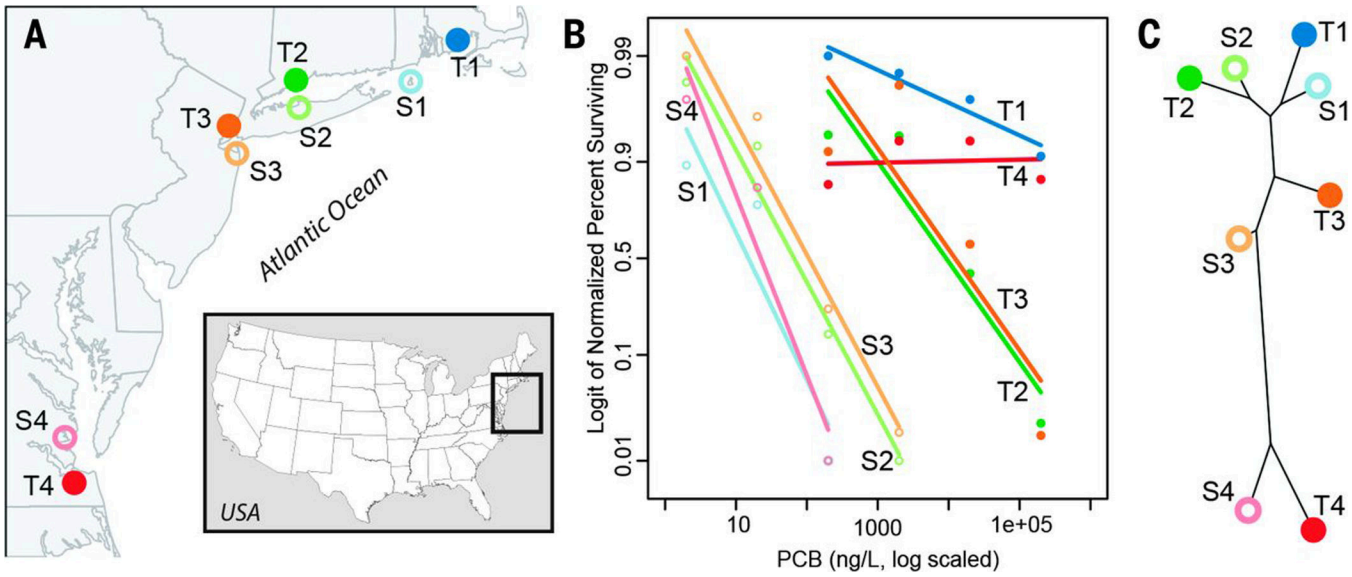

Fig. 1.

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 PCB126. 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 bi-allelic SNP frequencies, showing genetic differentiation is lowest between T-S population pairs (Phylip, CONTML module, bootstrap supports are 100 for all branches).

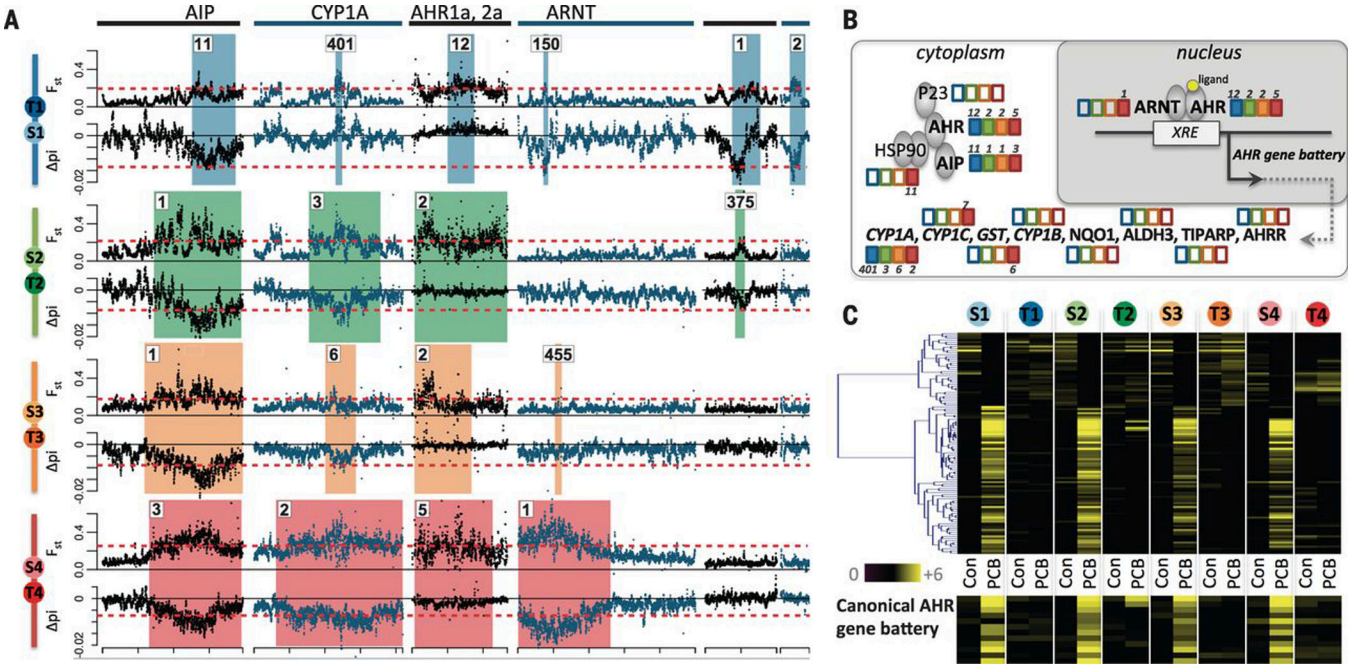

Fig. 2.

Patterns of structural and functional genomic divergence. A) Allele frequency differentiation (F_{ST}, top) and nucleotide diversity (pi, lower) difference (Tolerant pi – Sensitive pi) for each population pair studied for top-ranking outlier regions (including the top 2 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 line indicates outlier thresholds. Each tick on X axis is 500 kb position on scaffold and candidate gene name is indicated (top) for each outlier region. Top outliers regions are not co-localized 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 panel A. Top-ranking outlier regions contain AHR pathway genes and tend to be outliers in all population pairs, though some significant outliers are population-specific. C) Gene expression (developing embryos) heatmap shows up-regulated genes in response to PCB126 exposure ("PCB"; 200 ng/L) compared to 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).

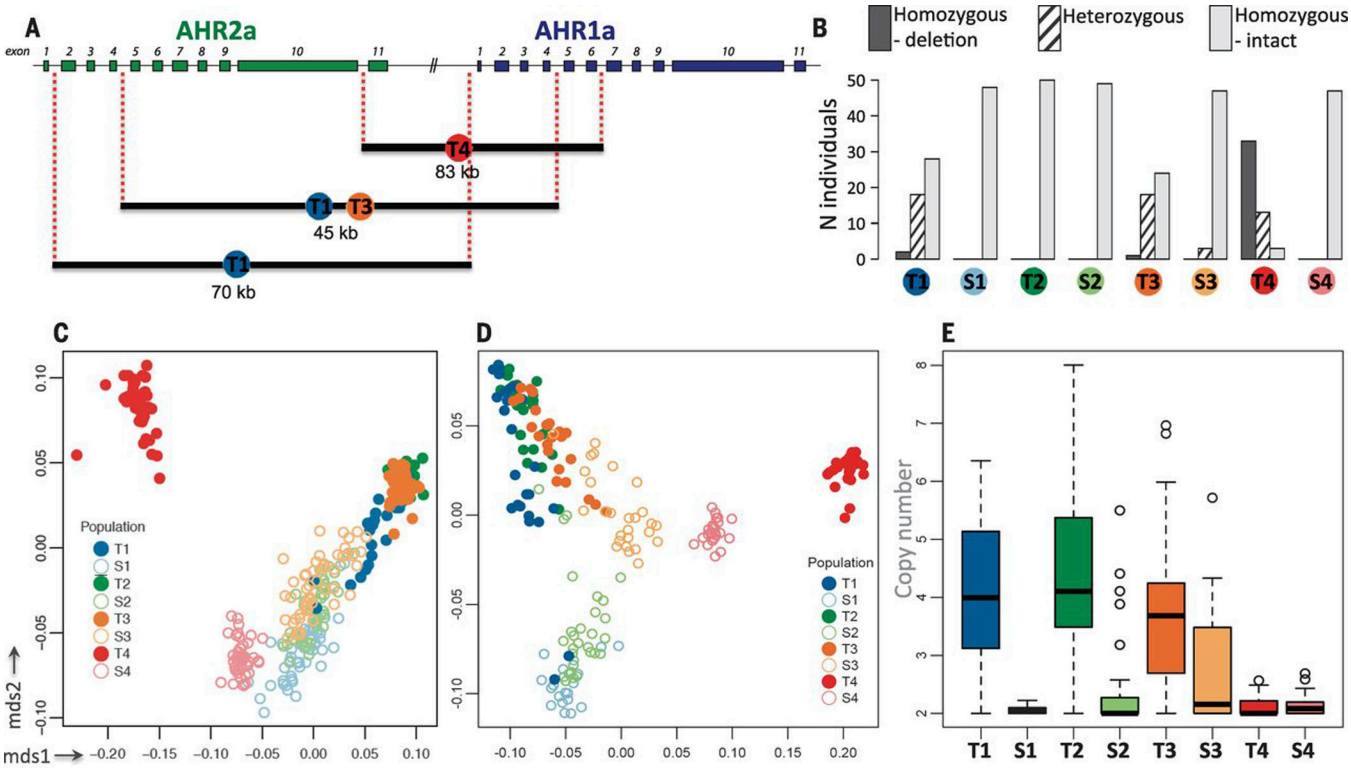

Fig. 3.

Patterns of adaptive genetic variation for top-ranking and shared outliers. A) Gene model of AHR2a and AHR1a (green/blue squares represent exons). Black bars indicate deleted regions present within tolerant populations. B) The number of individuals homozygous for specific deletions (black bar), heterozygous (hatched gray bar), or homozygous wildtype (light bar) within each population. C) Multi-dimensional 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). Though the CYP1A region is highly differentiated in all tolerant populations (D), CYP1A duplications are found only in northern tolerant populations (E).