

Integrating population genomic data to understand mechanisms of chemical susceptibility and resistance

Mark Hahn^{1,2} and Adam Labadorf³

¹Woods Hole Oceanographic Institution, Woods Hole, MA

²Boston University Superfund Research Program

³Dept of Medicine, Boston University School of Medicine



Boston University
Superfund Research Program



early life exposures, later life consequences



**National Institute of
Environmental Health Sciences**
Superfund Research Program

Integrating population genomic data to understand mechanisms of chemical susceptibility and resistance

Boston Univ. and Duke Univ. SRPs

Mark Hahn, Neel Aluru, Sibel Karchner (WHOI and BUSRP), **Adam Labadorf (BU)**, Filisia Agus (BU); Noah Reid (UCONN)

Rich Di Giulio (Duke SRP), Mark DeLong, David Corcoran (Duke); Joshua Osterberg and Thomas Schultz (Duke Marine Lab)

Consultants/Collaborators: Andrew Whitehead (UC Davis); Wes Warren (U Missouri), Diane Nacci and Bryan Clark (U.S. EPA, Office of Research and Development)



Boston University
Superfund Research Program



Duke
UNIVERSITY

SUPERFUND
Research Center

early life exposures, later life consequences



**National Institute of
Environmental Health Sciences**
Superfund Research Program

Research Questions

- What are the impacts of **long-term exposure** to toxic chemicals on natural populations?
- Can populations **adapt** over time? What are the **mechanisms** of adaptation? What are the **fitness costs** of adaptation?
- How does genetic variation influence sensitivity to toxic chemicals (GxE)?

Population-based Model System

- Atlantic killifish (*Fundulus heteroclitus*) inhabiting estuaries (including Superfund sites) along the U.S. East Coast.



Atlantic killifish
(*Fundulus heteroclitus*)

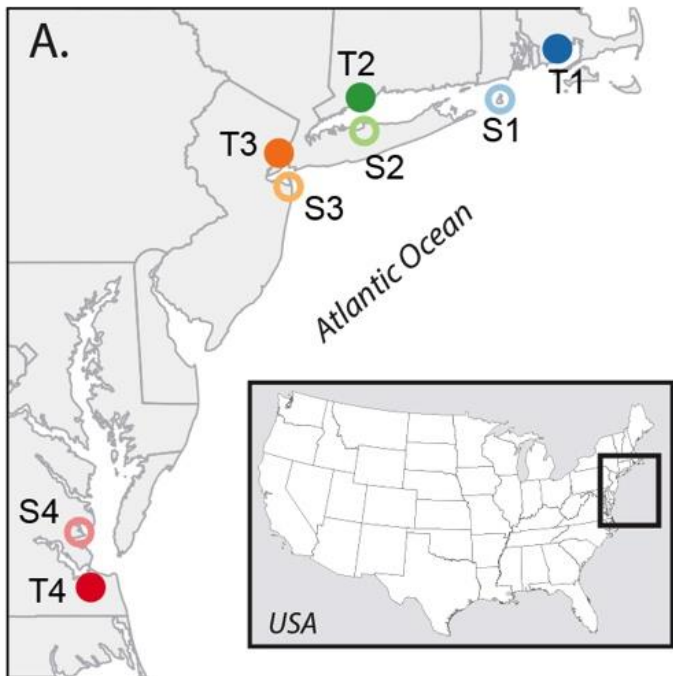
(photo courtesy of
Dr. Evan D'Alessandro, RSMAS)

Goals

- Integrate two large genome-sequencing data sets (and related data) from two SRP Centers that both study the genetic mechanisms underlying evolved resistance to
 - PCBs and dioxins (Boston University)
 - PAHs (Duke University)in Atlantic killifish populations living at Superfund sites.
- Enhance the findability, accessibility, interoperability, and reuse (FAIR) of the data sets.
- Enable comparison of these datasets to other data (zebrafish, mouse, human) on genetic variation associated with chemical pollution and disease.

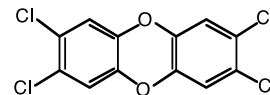
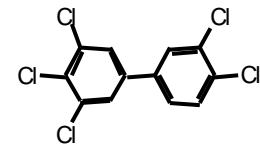


Tolerant (T) and Sensitive (S) populations of killifish: Whole-genome sequence data (Boston Univ SRP and collaborators)

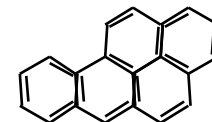


T=Tolerant S=Sensitive

- New Bedford Harbor, MA DL-PCBs
- Bridgeport, CT DL-PCBs, PAHs
- Newark, NJ Chlorinated dioxins



- Elizabeth River, VA PAHs



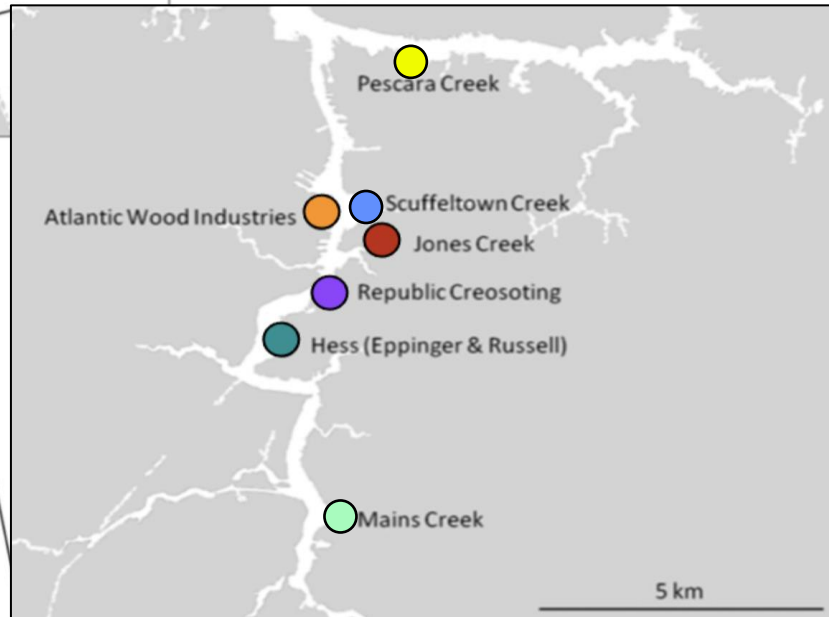
(DL-PCBs = dioxin-like PCBs)

Reid *et al.* 2016 *Science* **354**: 1305

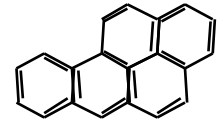
Tolerant and Sensitive populations of killifish: RAD-seq data (Duke SRP)



Elizabeth River, VA

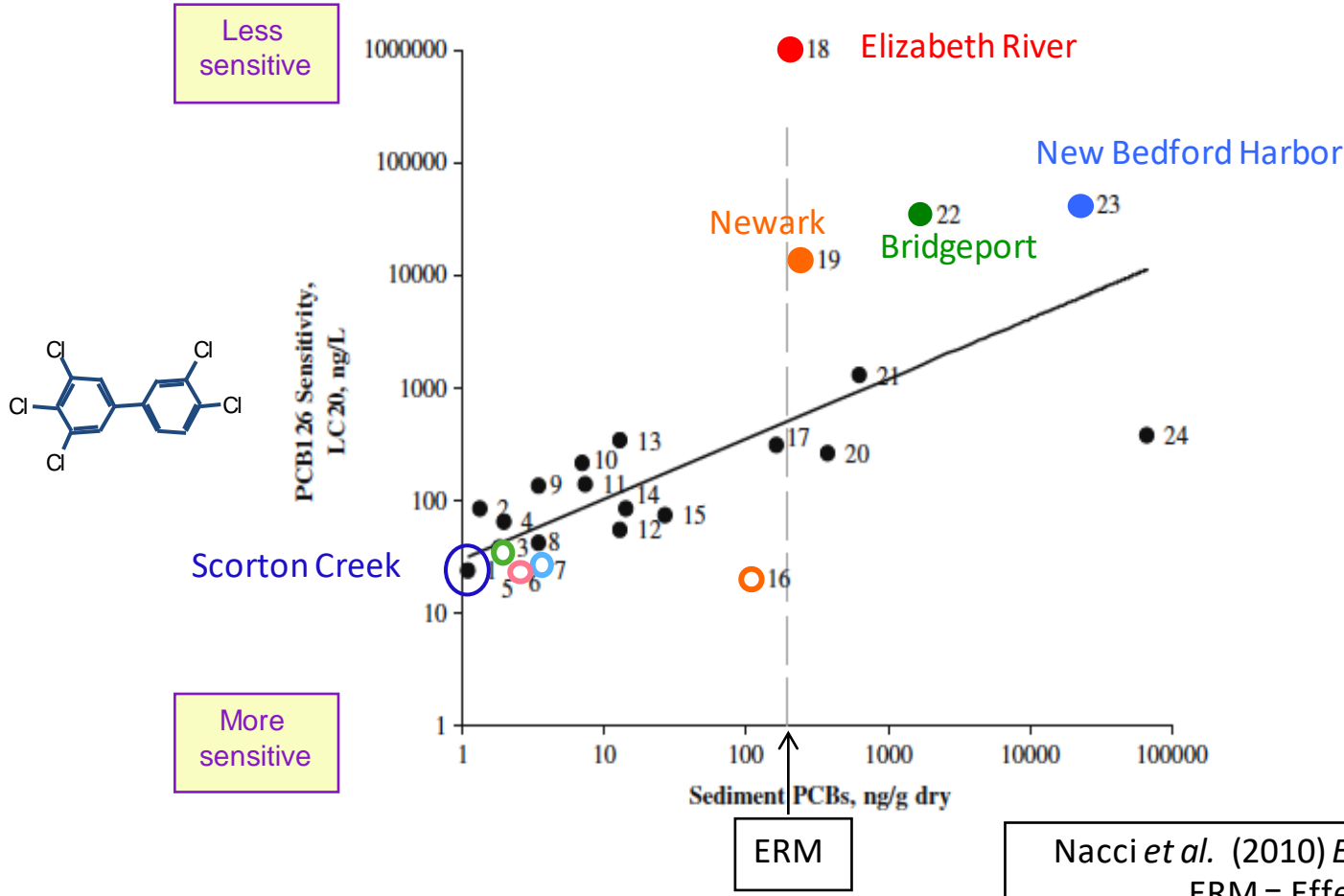


PAHs



Osterberg *et al.* 2018 *Aquat Toxicol*
200: 73, and unpublished

Populations vary in embryo sensitivity to PCB-126 (up to 8,000-fold)

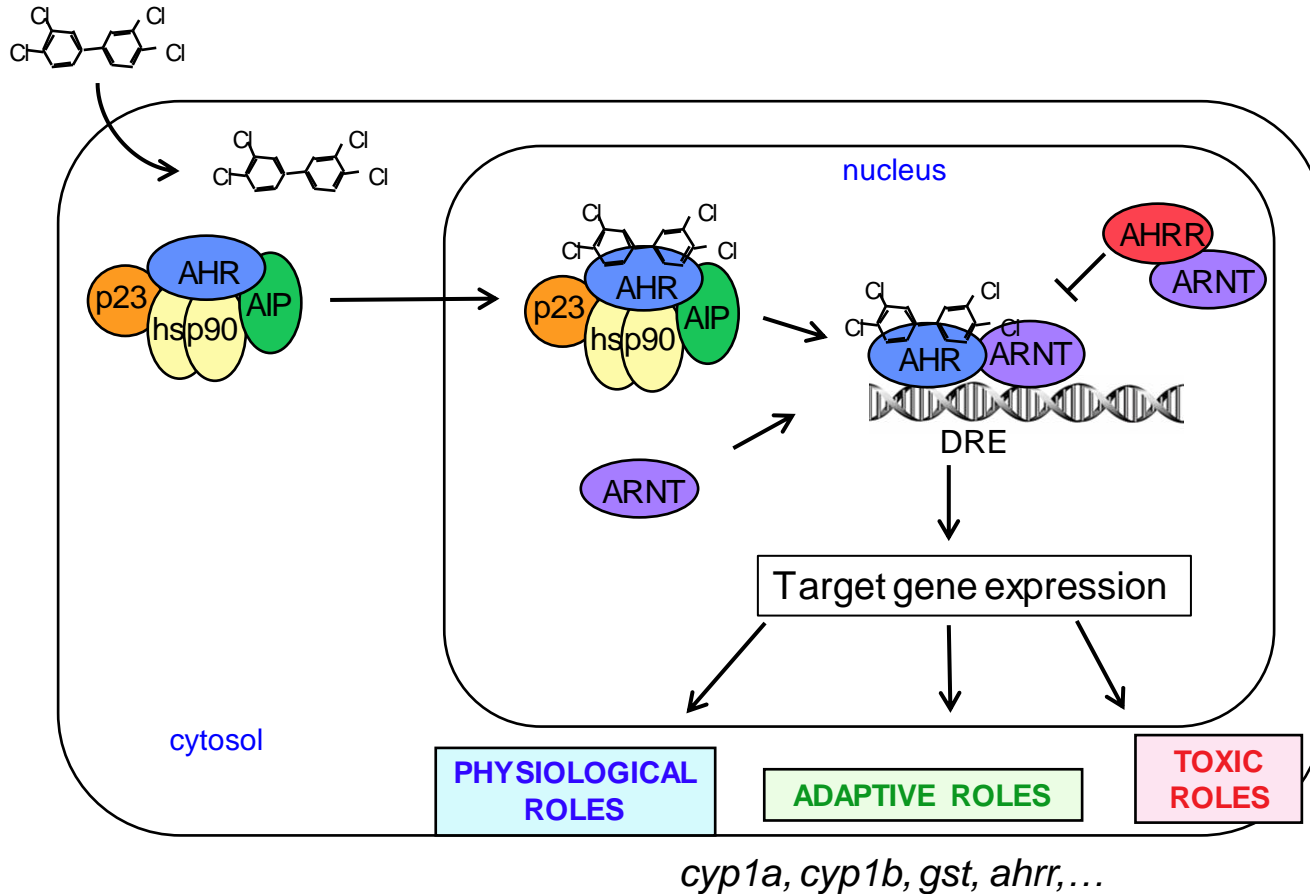


Nacci *et al.* (2010) *Estuaries and Coasts* **33**: 853
ERM = Effects Range Median

Genome-wide approaches to understand the mechanism of resistance

- **Gene expression profiling**
(Whitehead *et al.* (2012) *Proc Royal Soc Biol Sci* **279**: 427-433)
- **Quantitative genetics** [quantitative trait loci (QTL)]
(Nacci *et al.* (2016) *Mol Ecol* **25**: 5467-5482)
- **Restriction-site Associated DNA (RAD)-sequencing (RAD-seq)**
(Osterberg *et al.* (2018) *Aquat Toxicol* **200**: 73-82)
- **Whole Genome Sequencing (WGS)**
(Reid *et al.* (2016) *Science* **354**: 1305-1308)

Aryl Hydrocarbon Receptor (AHR) pathway as a Shared Target of Selection



Existing data sets

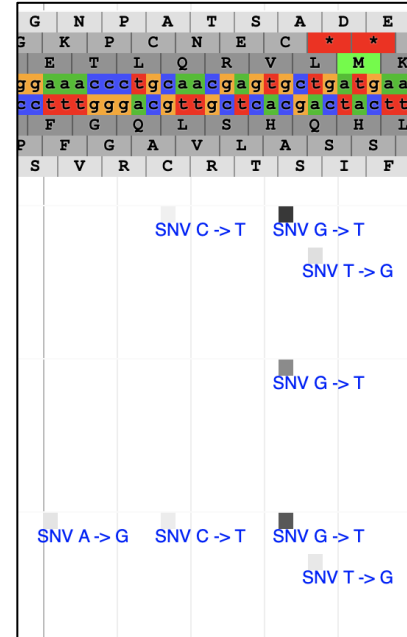
- [Whole-genome re-sequencing \(WGS\) data](#) for 384 individual killifish in eight populations (four pairs of tolerant and sensitive populations) [Reid, *et al.* (2016) *Science* **354**: 1305-1308].
- [RAD-seq data](#) (genome sequence samples) from 288 individual fish from nine populations [Osterberg, *et al.* (2018) *Aquat Toxicol* **200**: 73-82].
- [RNA-seq](#) data.
- *F. grandis* [WGS data](#) from nine populations [Oziolor, *et al.* (2019) *Science* **364**: 455-457].
- *Fundulus heteroclitus* [New genome assembly \(NCBI\)](#) [Warren, Whitehead, *et al.* (2020) PRJNA615222]

Number of data sets

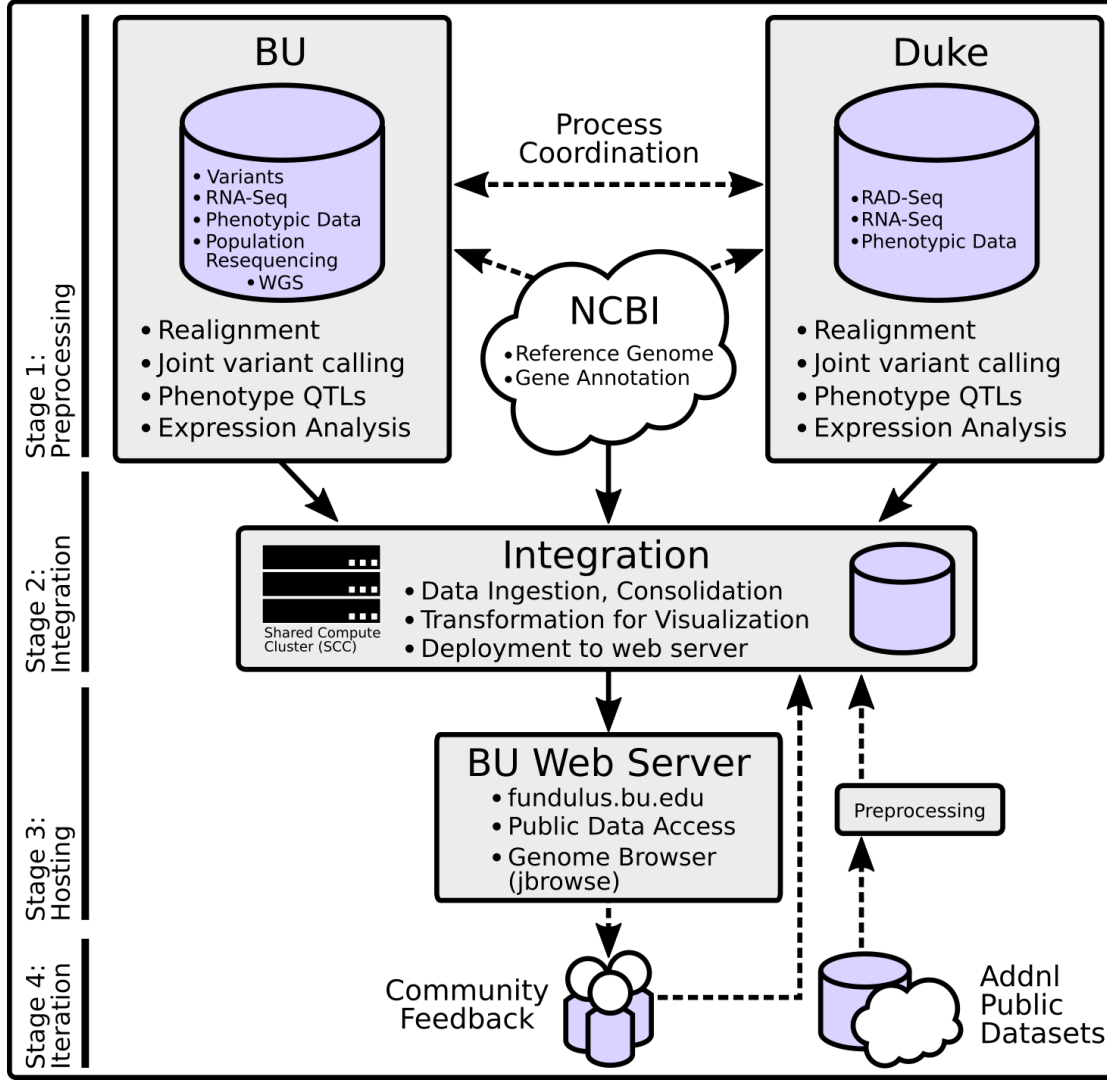
WGS	RAD-seq	RNA-seq	TOTAL
672	288	75	1035

Challenges

- How to integrate these disparate data sets (WGS, RAD-seq, RNA-seq) to allow analysis and reuse
- How to retain access to population, individual, and molecular-level information on genetic variants
- How to make the data easily accessible, expandable, and usable by other researchers
- How to enable comparison of these data with GxE data in other species



Approach



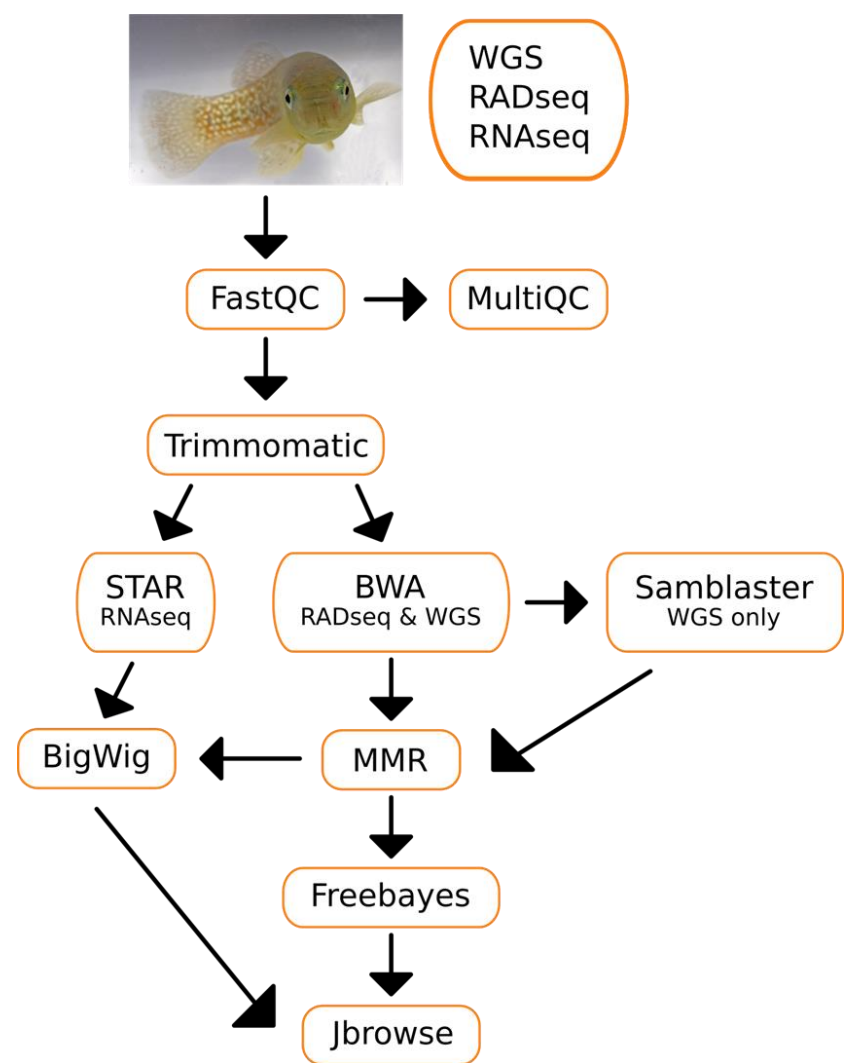
Pipeline

WGS	RAD-seq	RNA-seq	TOTAL
672	288	75	1035

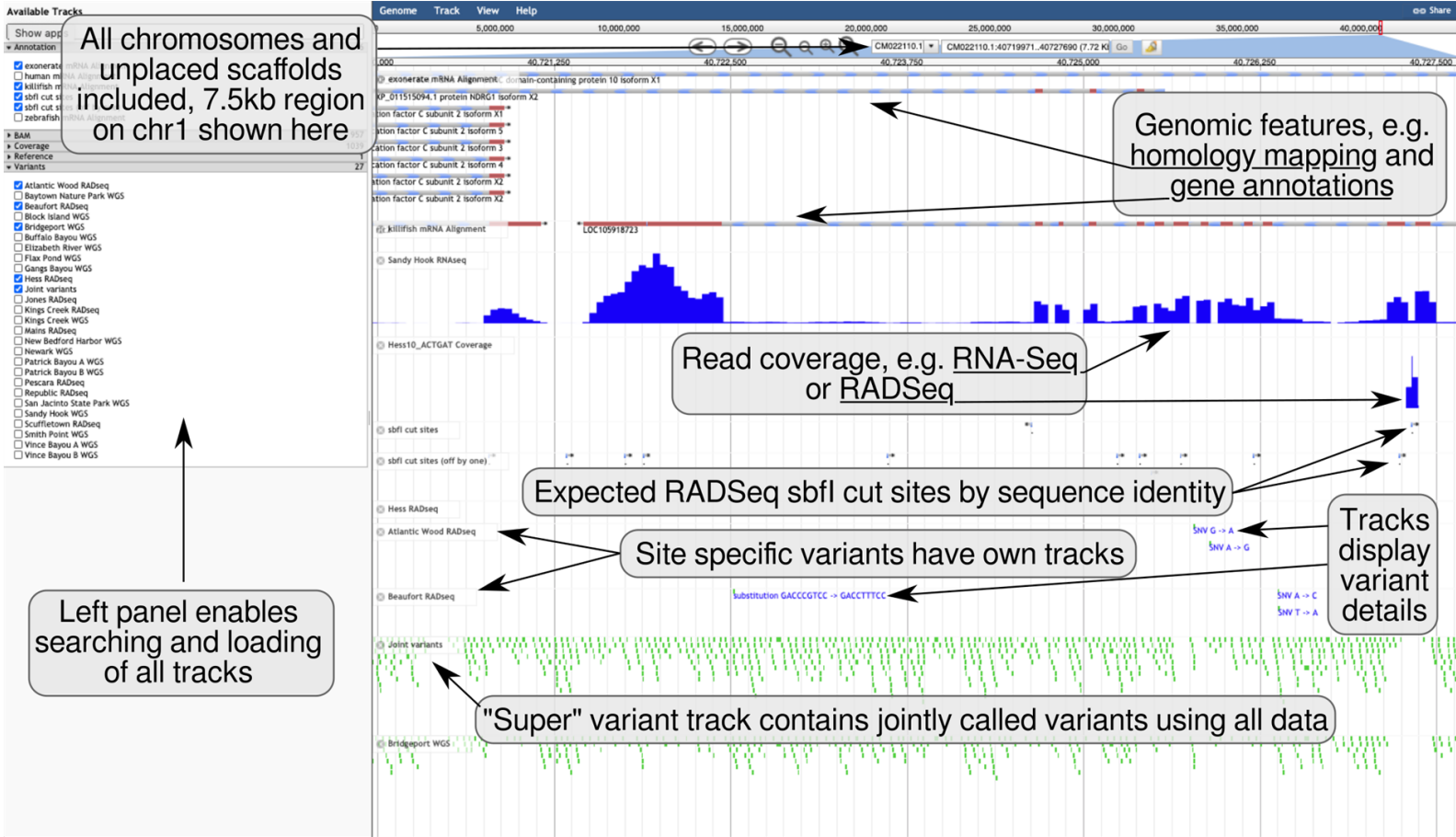
- **Harmonized** bioinformatics analysis pipelines
- **Reanalyzed** data using new *Fundulus heteroclitus* reference genome assembly
- Loaded data and metadata into the open-source genome browser **JBrowse** [Buels et al. (2016) *Genome Biology* **17**:66]
- Performed **joint variant calling**

SuperFunBase

<http://fundulus.bu.edu>



SuperFunBase: JBrowse Interface



SuperFunBase: Joint variant (SNV) calling

Available Tracks

filter tracks

Annotation 5

- human exonerate protein alignment
- killifish reference
- sbfl cut sites
- sbfl cut sites (off by one)
- zebrafish exonerate protein alignment

BAM 957

Coverage 1031

Reference 1

- Reference sequence

Variants 27

- Atlantic Wood RADseq
- Baytown Nature Park WGS
- Beaufort RADseq
- Block Island WGS
- Bridgeport WGS
- Buffalo Bayou WGS
- Elizabeth River WGS
- Flax Pond WGS
- Gangs Bayou WGS
- Hess RADseq
- Joint variant calling
- Jones RADseq
- Kings Creek RADseq
- Kings Creek WGS
- Mains RADseq
- New Bedford Harbor WGS
- Newark WGS
- Patrick Bayou A WGS
- Patrick Bayou B WGS
- Pescara RADseq
- Republic RADseq
- San Jacinto State Park WGS
- Sandy Hook WGS

Genome Track View Help Share

0 5,000,000 10,000,000 15,000,000 20,000,000 25,000,000 30,000,000 35,000,000 40,000,000

NC_046367.1 NC_046367.1:938053..938513 (461 b) Go

938,125 938,250 938,375 938,500

Reference sequence

killifish reference

LOC105937229 aryl hydrocarbon receptor (AHR1a)

Block Island WGS

New Bedford Harbor WGS

Kings Creek WGS

Elizabeth River WGS

Kings Creek RADseq

Atlantic Wood RADseq

SNV G -> A

SNV G -> T

SNV G -> A

SNV G -> T

Population tracks

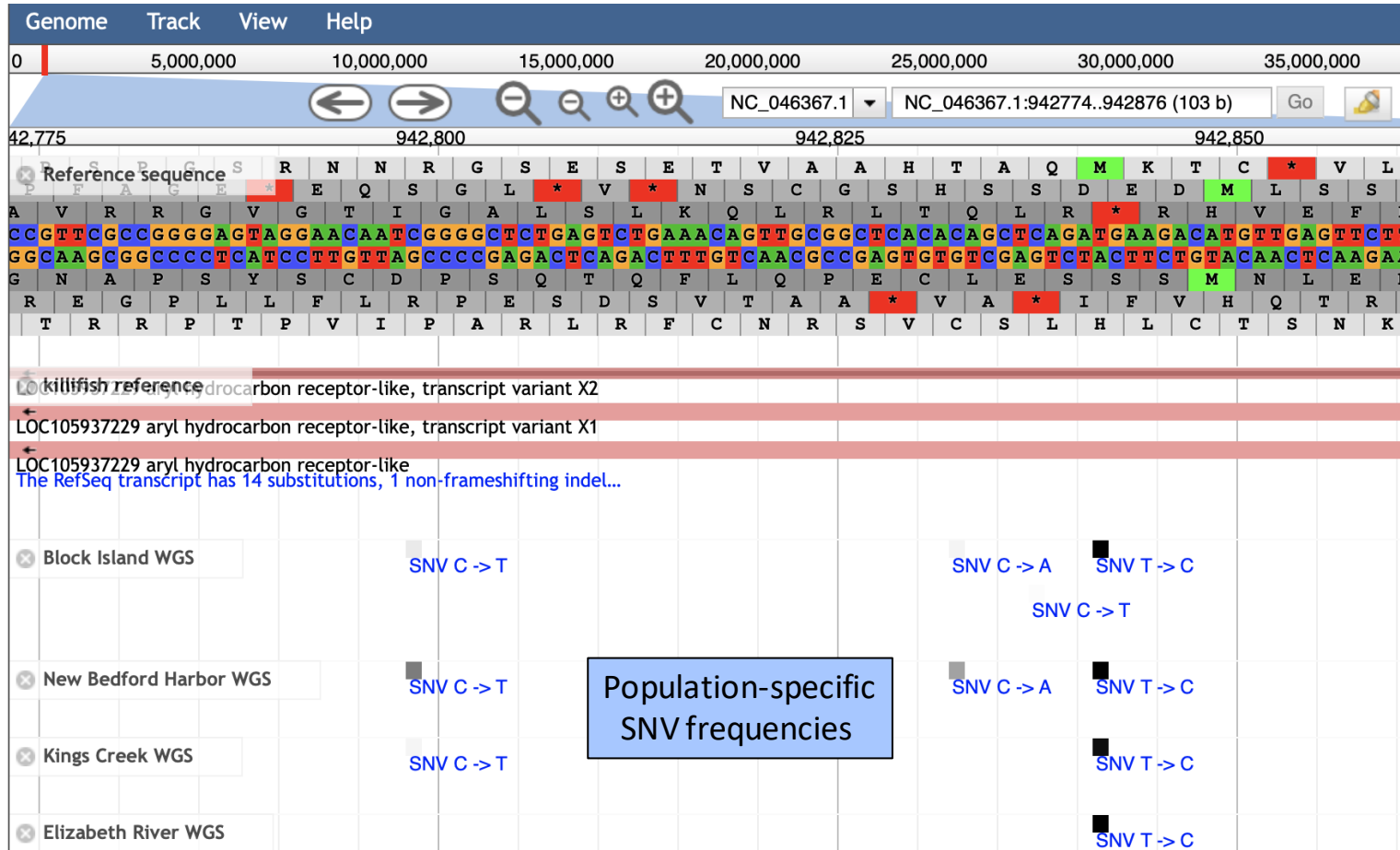
Killifish Reference Sequence

Annotated genes

WGS SNVs

RAD-seq SNVs

SuperFunBase: Population-specific variants



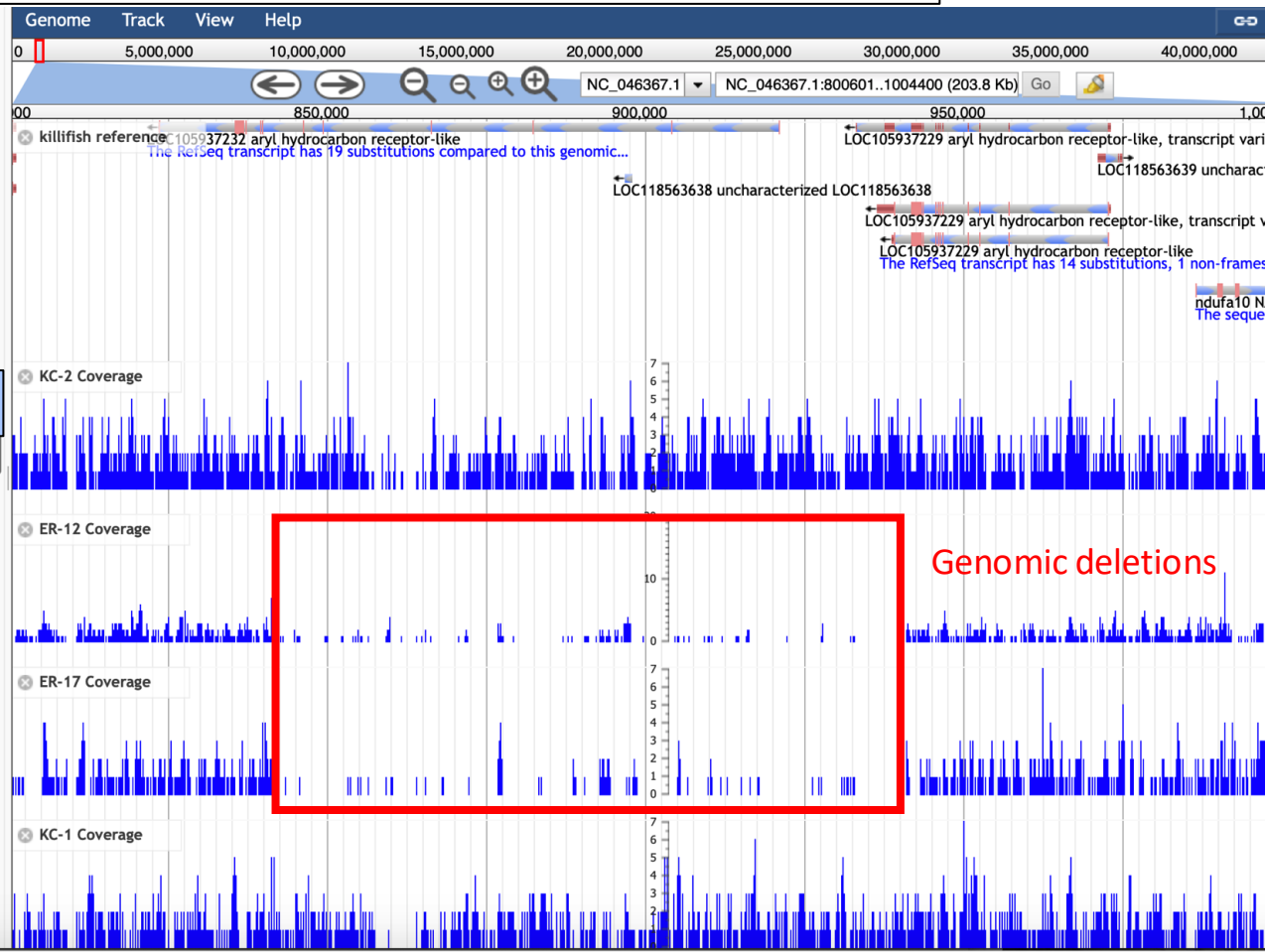
SuperFunBase: Larger genomic features

- Bft32_TTGCAG Coverage
- Bft3_GCCGAG Coverage
- Bft4_TGATCG Coverage
- Bft5_AGCTCC Coverage
- Bft6_GAATCT Coverage
- Bft7_GCTTTA Coverage
- Bft8_TAGCCT Coverage
- Bft9_CGCGGA Coverage
- ER-11 Coverage
- ER-12 Coverage
- ER-13 Coverage
- ER-14 Coverage
- ER-15 Coverage
- ER-16 Coverage
- ER-17 Coverage
- ER-18 Coverage
- ER-19 Coverage
- ER-20 Coverage
- ER-21 Coverage
- ER-22 Coverage

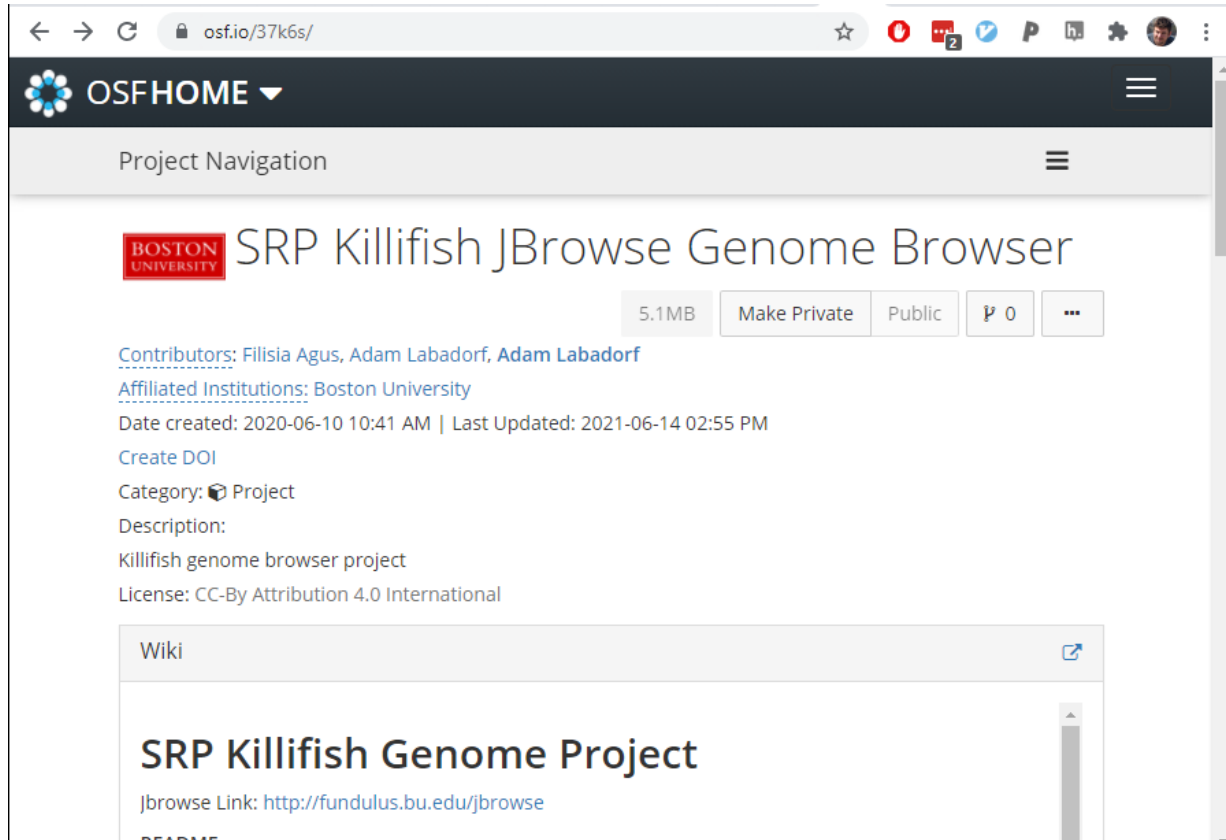
Individual fish sequence tracks

- ER-27 Coverage
- ER-28 Coverage
- ER-29 Coverage
- ER-30 Coverage
- ER-31 Coverage
- ER-32 Coverage
- ER-33 Coverage
- ER-34 Coverage
- ER-35 Coverage
- ER-36 Coverage
- ER-38 Coverage
- ER-39 Coverage
- ER-40 Coverage
- ER-41 Coverage
- ER-42 Coverage
- ER-43 Coverage
- ER-44 Coverage
- ER-45 Coverage
- ER-46 Coverage
- ER-47 Coverage
- ER-48 Coverage
- ER-49 Coverage
- ER-50 Coverage
- ER-51 Coverage
- ER-52 Coverage
- ER-53 Coverage
- ER-54 Coverage
- ER-55 Coverage
- ER-56 Coverage

Screenshot




Open Science Framework (OSF.io)



The screenshot shows a web browser window with the URL osf.io/37k6s/. The page header includes the OSFHOME logo and a navigation menu. The main content area displays the project title "SRP Killifish JBrowse Genome Browser" with a Boston University logo. Below the title, there are buttons for "5.1MB", "Make Private", "Public", "0", and a more options menu. The page also lists contributors (Filisia Agus, Adam Labadorf), affiliated institutions (Boston University), creation and update dates, a DOI link, category (Project), description (Killifish genome browser project), and license (CC-BY Attribution 4.0 International). A "Wiki" section is visible at the bottom, containing the title "SRP Killifish Genome Project" and a link to the JBrowse genome browser.

OSFHOME

Project Navigation

 SRP Killifish JBrowse Genome Browser

5.1MB Make Private Public 0 ...

Contributors: Filisia Agus, Adam Labadorf, Adam Labadorf

Affiliated Institutions: Boston University

Date created: 2020-06-10 10:41 AM | Last Updated: 2021-06-14 02:55 PM

Create DOI

Category: Project

Description:

Killifish genome browser project

License: CC-BY Attribution 4.0 International

Wiki

SRP Killifish Genome Project

Jbrowse Link: <http://fundulus.bu.edu/jbrowse>

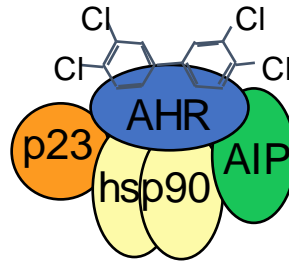
<https://osf.io/37k6s/>

What's next?

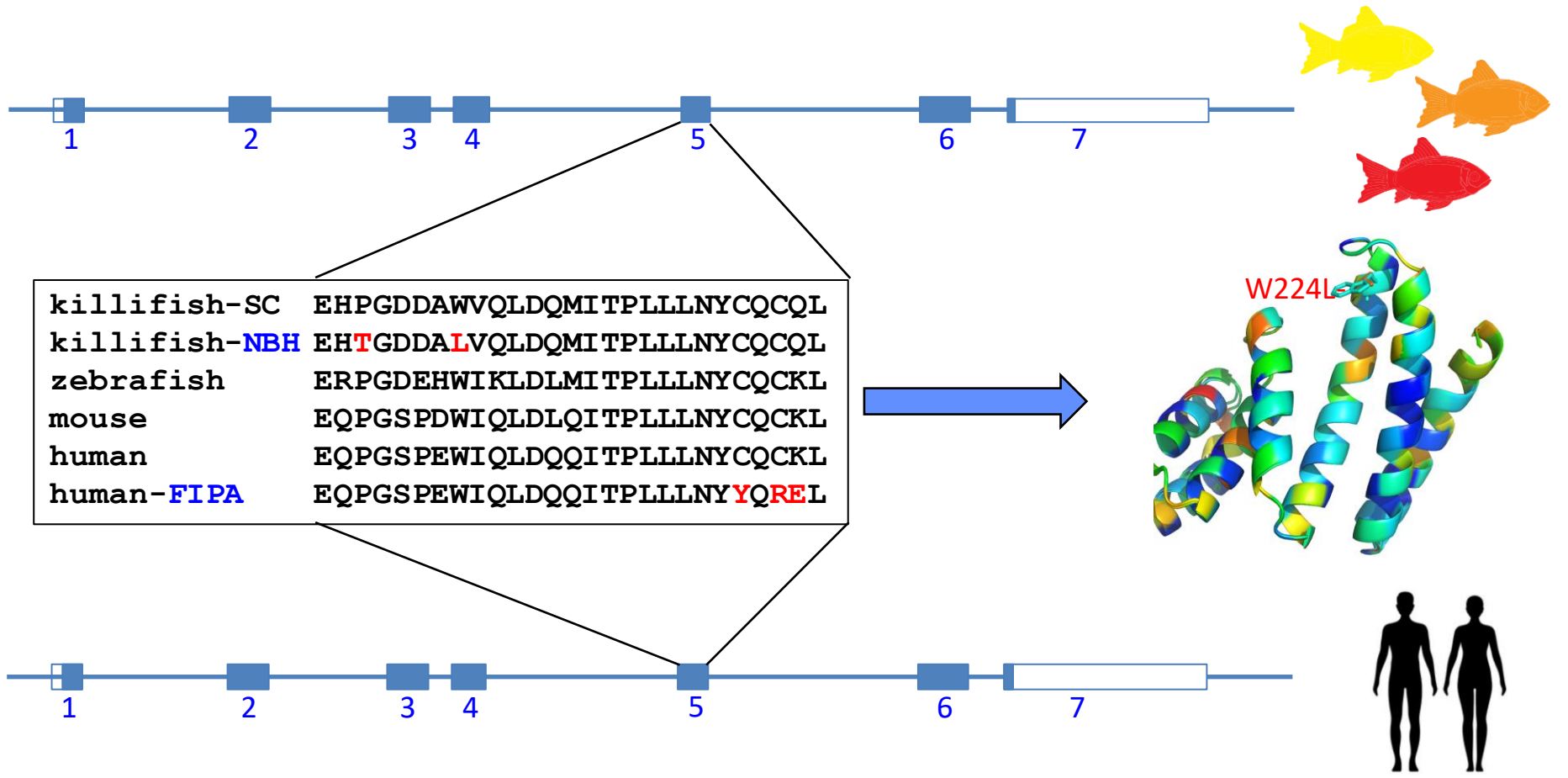
- Add population-specific **epigenomic** data (esp. DNA methylation)
- Call SNVs from **RNA-seq** data
- Strategy for linking killifish SNV data with SNV data for **orthologous genes in other species** (humans, rodents, zebrafish)

Example:

AIP (AHR-Interacting Protein)



Comparing SNVs in Orthologous Genes Across Populations and Species



Other Comparative Databases

- **Comparative Toxicogenomics Database (CTD)**
“Chemical–gene/protein interactions, chemical–disease and gene–disease relationships”
<http://ctdbase.org>
- **SeqAPASS** (Sequence Alignment to Predict Across Species Susceptibility)
“Predict across-species susceptibility to chemicals with known molecular targets”
<https://www.epa.gov/chemical-research/sequence-alignment-predict-across-species-susceptibility>
- **Monarch Initiative** of the Global Alliance for Genomics and Health (GA4GH)
“Connecting phenotypes to genotypes across species”
<https://monarchinitiative.org>

Integrating population genomic data to understand mechanisms of chemical susceptibility and resistance



SuperFunBase

<http://fundulus.bu.edu>

- Contact: mhahn@whoi.edu; richd@duke.edu; labadorf@bu.edu

P42ES007381-23S1

P42ES010356-17S1



Boston University
Superfund Research Program



Duke
UNIVERSITY

SUPERFUND
Research Center

early life exposures, later life consequences



**National Institute of
Environmental Health Sciences**
Superfund Research Program