

Using RAD-seq for genomic and transcriptomic studies of non-model organisms

Julian Catchen
jcatchen@uoregon.edu



Cresko Lab
Institute for Ecology and Evolution
University of Oregon

Collaborators: Postlethwait & Bradshaw-Holzapfel Labs

What makes an organism a model?

classic models: fruit fly, mouse, zebrafish

1. Genetic Maps

(phenotypes, allozymes, microsats)

2. Physical Maps

(random, shotgun Sanger sequencing)

3. Transcriptomic Maps

(EST sequencing)

4. Gene Expression Analyses (microarrays, RNA-seq)



What makes an organism a model?

Organism	Markers	Organism	Markers
Bighorn Sheep	147	Silver carp	483
Shrimp	418	Guppy	790
Coral	420	Barramundi	240
Indian mustard	1,029	Catfish	331
Oilseed rape	13,551	Sea bass	368
Black spruce	1,111	Cichlid	204
Barley	2,111	Platyfish	290
Flounder	1,375	Halibut	604
Turbot	242	Sea bream	204
Human	250,000	Mouse	10,000

Sequence everything?

Outline

- Genetic marker analysis with RAD
- What if you have a reference genome?
 1. Parallel evolution in the threespine stickleback
- What if you **don't** have a reference genome?
 2. Building a genetic map of the spotted gar
 3. Assembling the genome of the platyfish
 4. Phylogeography of the pitcher plant mosquito

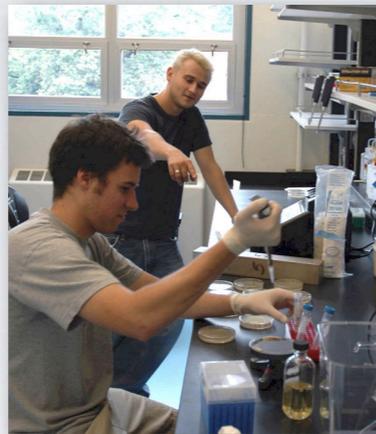
What is RAD-seq?

(Restriction-site Associated DNA)

Eric Johnson



Joe Dunham



Mike Miller



Illumina

2007

Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers

Michael R. Miller,¹ Joseph P. Dunham,² Angel Amores,³ William A. Cresko,² and Eric A. Johnson^{1,4}

¹Institute for Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA; ²Center for Ecology & Evolutionary Biology, University of Oregon, Eugene, Oregon 97403, USA; ³Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403, USA

2008

OPEN ACCESS Freely available online

PLoS one

Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers

Nathan A. Baird^{1,2}, Paul D. Etter^{1,2}, Tressa S. Atwood², Mark C. Currey³, Anthony L. Shiver¹, Zachary A. Lewis¹, Eric U. Selker¹, William A. Cresko³, Eric A. Johnson^{1*}

¹ Institute of Molecular Biology, University of Oregon, Eugene, Oregon, United States of America, ² Floragenex, Eugene, Oregon, United States of America, ³ The Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon, United States of America

What is RAD-seq?

(Restriction-site Associated DNA)



22,830 *SbfI* sites in Stickleback

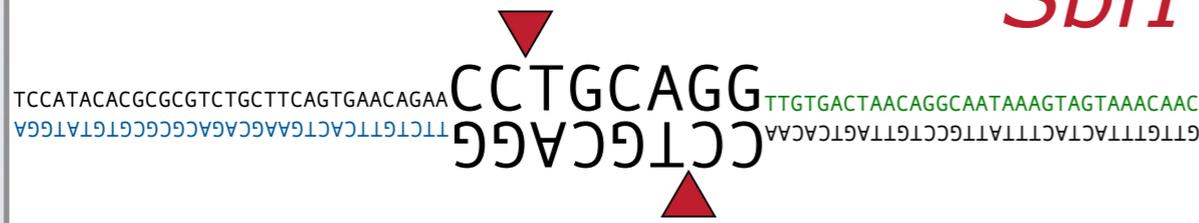
~ 45,000 RAD-Tags

HiSeq Illumina Lane:

100 million reads, 96 barcoded individuals

①

SbfI

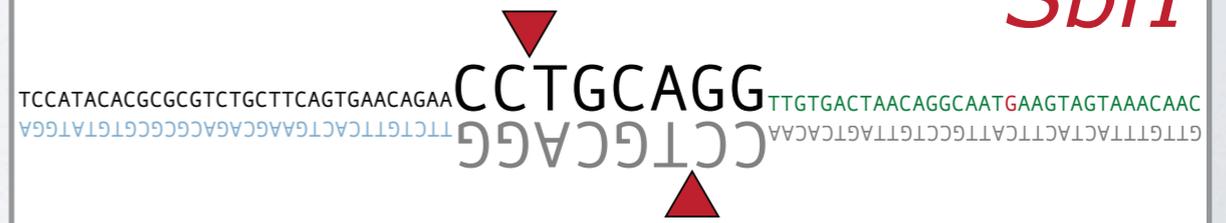


TGCAGG TTGTGACTAACAGGCAATAAAGTAGTAAACAAC

TGCAGG TTCTGTTCACTGAAGCAGACGCGGTGTATGGA

②

SbfI



TGCAGG TTGTGACTAACAGGCAAT ^{G/A} AAGTAGTAAACAAC

TGCAGG TTCTGTTCACTGAAGCAGACGCGGTGTATGGA

What if you have a reference genome?

Threespine Stickleback, *Gasterosteus aculeatus*

- **Ancestral Oceanic Populations**

Marine and Anadromous
Old (> 10 million years)

- **Derived Freshwater Populations**

Lake and stream
Young (< 15,000 years)



Ocean form



Freshwater form



Susan
Bassham



Paul
Hohenlohe

Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags

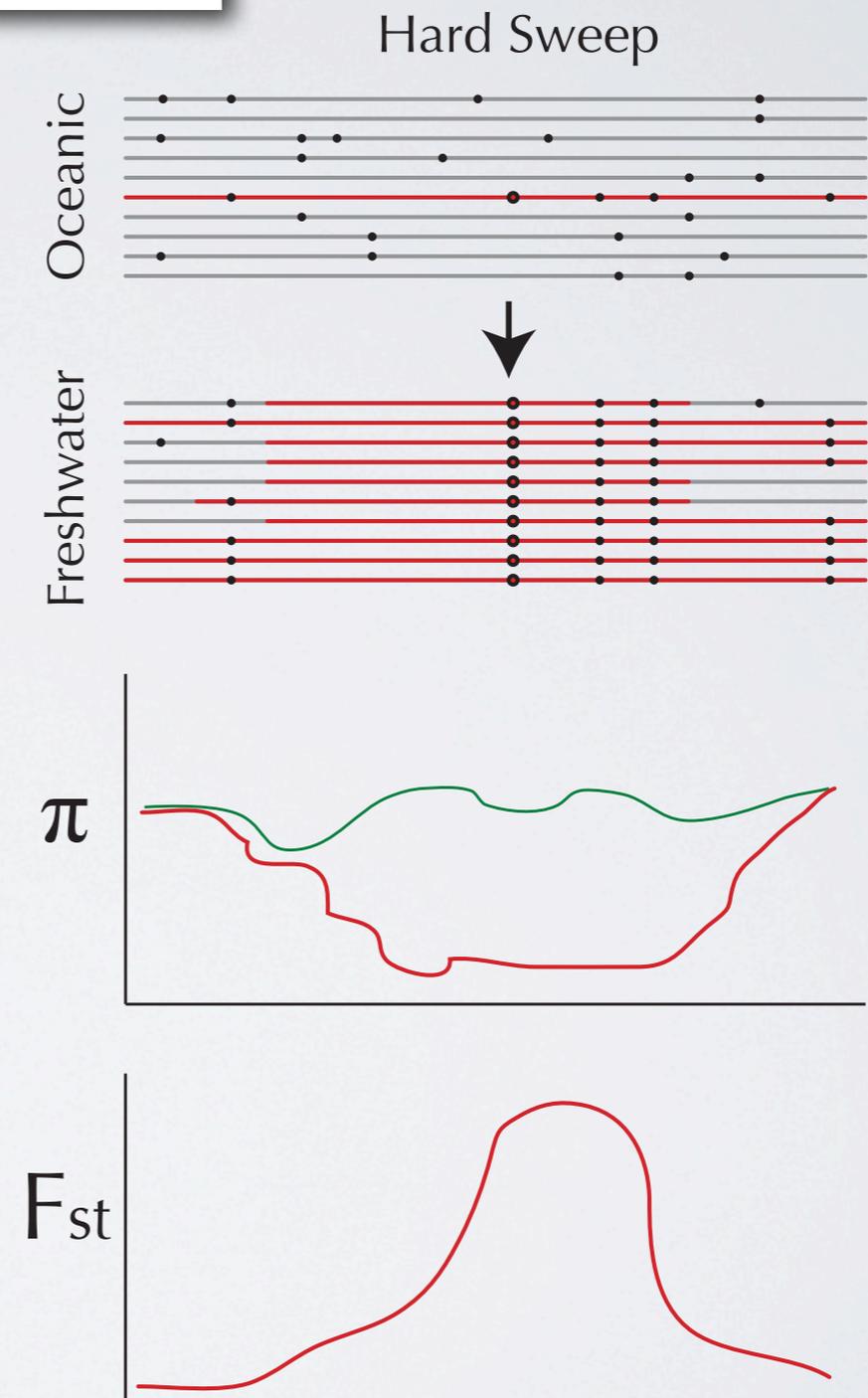
Paul A. Hohenlohe¹*, Susan Bassham¹*, Paul D. Etter², Nicholas Stiffler³, Eric A. Johnson², William A. Cresko^{1*}

¹ Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon, United States of America, ² Institute of Molecular Biology, University of Oregon, Eugene, Oregon, United States of America, ³ Genomics Core Facility, University of Oregon, Eugene, Oregon, United States of America

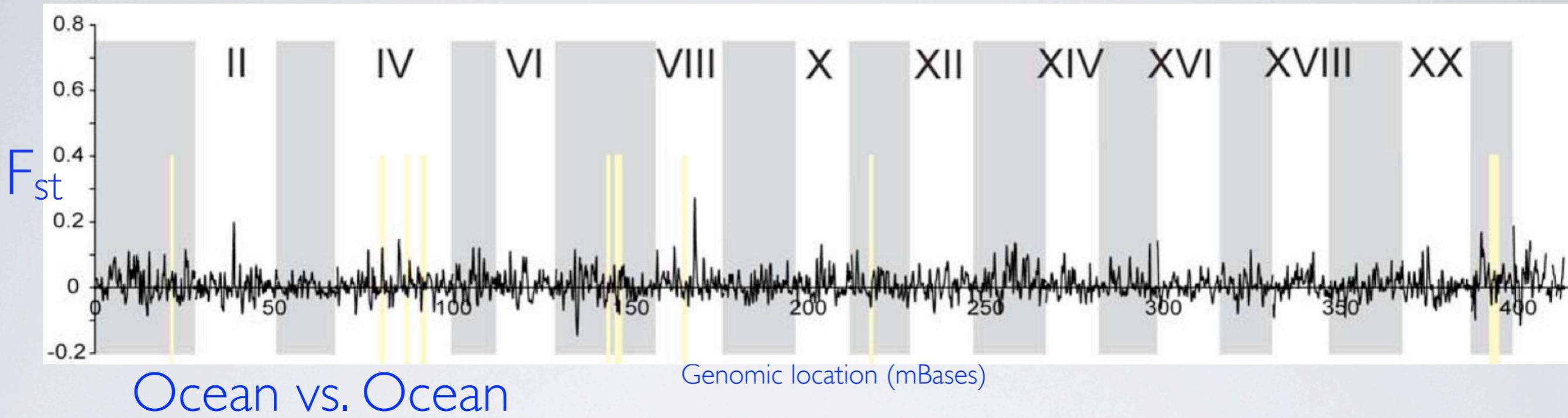
RAD-seq of 20 Individuals
from **2 marine** and
3 freshwater populations

Aligned reads against the
reference genome

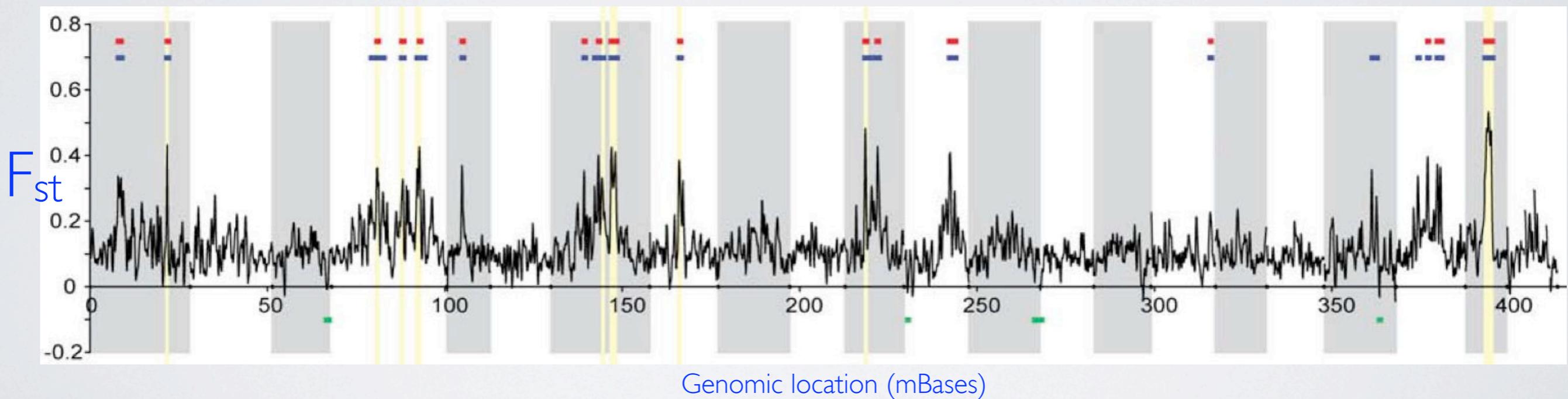
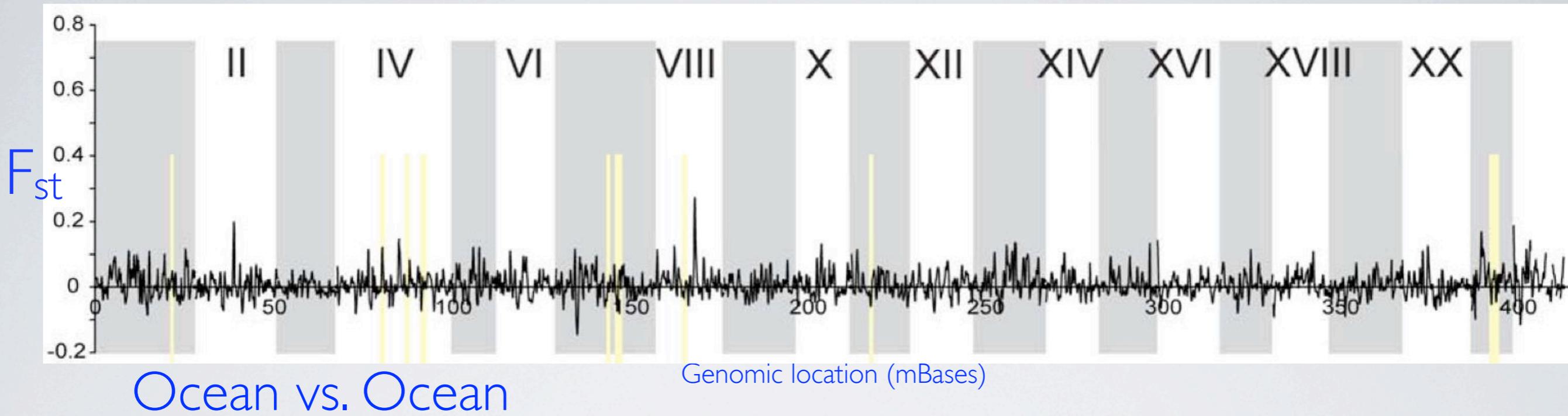
Looked for signatures of
selection seen in:
nucleotide diversity (π)
genetic divergence (F_{st})



Signatures of natural selection across the genome

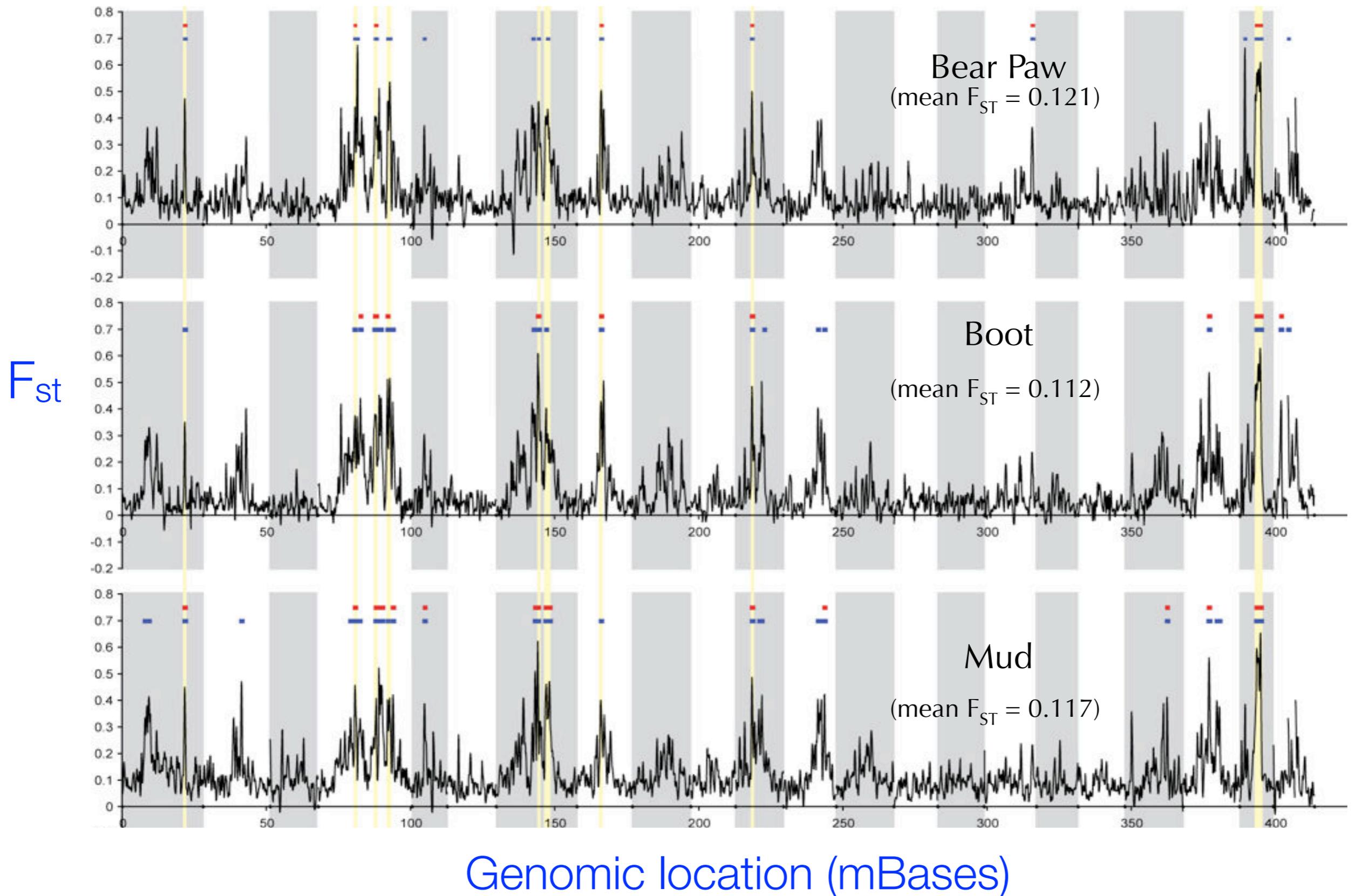


Signatures of natural selection across the genome

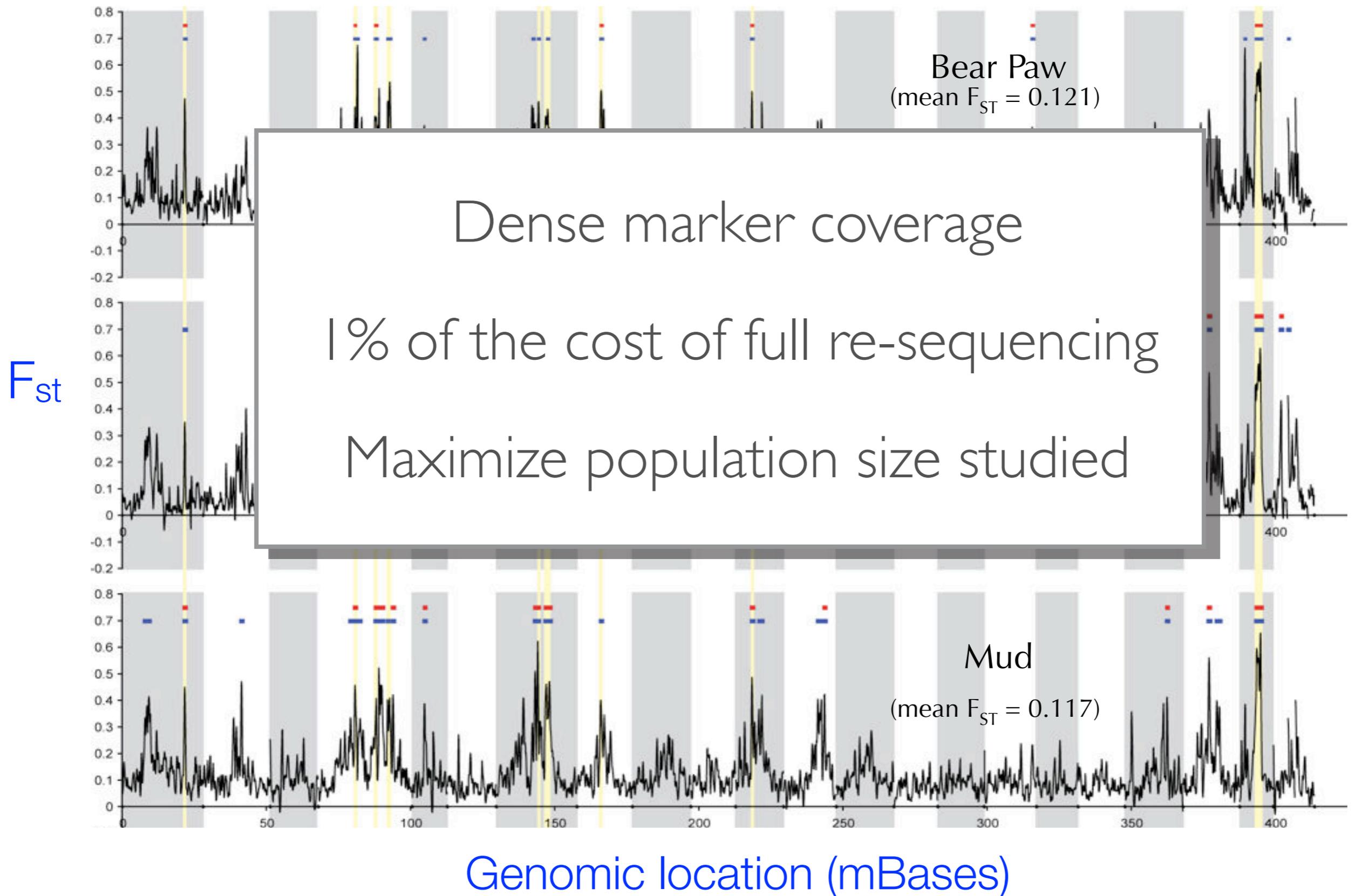


Freshwater Populations (grouped) vs. Ocean

Consistent pattern of parallel selection across independently evolved freshwater populations



Consistent pattern of parallel selection across independently evolved freshwater populations



What if you **don't** have a reference genome?



The image shows a browser window with the URL <http://creskolab.uoregon.edu/stacks/>. The page features the word "Stacks" in a large, bold, black font, with a stylized barcode graphic integrated into the letter 'S'. Below the title, a paragraph describes the software: "Stacks is a software pipeline for building loci out of a set of short-read sequenced samples. Stacks was developed for the purpose of building genetic maps from RAD-Tag Illumina sequence data, but can also be readily applied to population studies, and phylogeography." A prominent green button with a downward arrow icon and the text "Download Stacks" is centered on the page, with "Version 0.982" written below it. At the bottom of the page, there is a link for "Recent Changes [updated Mar 29, 2011]".

G3: Genes, Genomes, Genetics

Stacks: Building and Genotyping Loci *De Novo* From Short-Read Sequences

Julian M. Catchen,* Angel Amores,[†] Paul Hohenlohe,* William Cresko,* and John H. Postlethwait^{†,1}

*Center for Ecology and Evolutionary Biology and [†]Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403



...



Stack 1

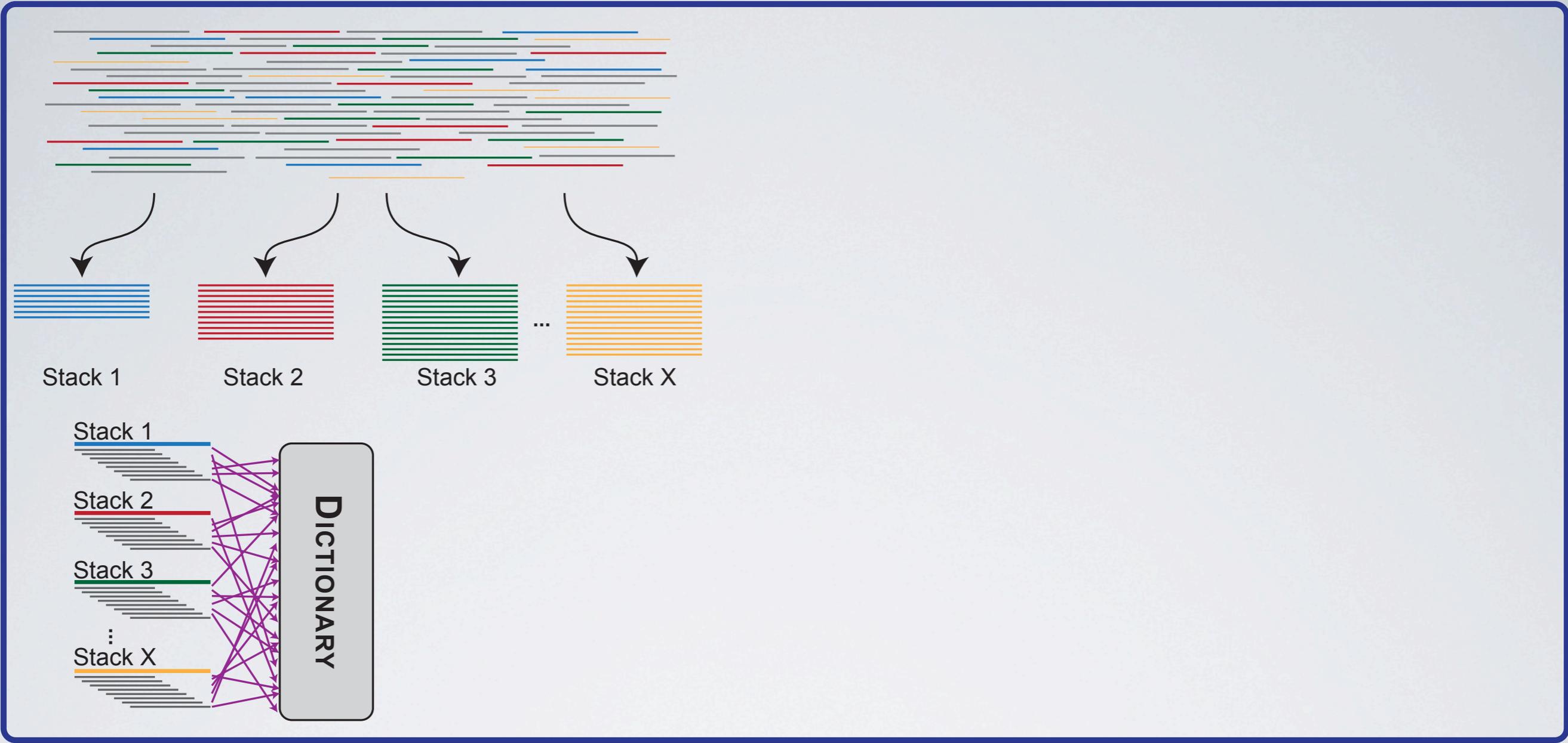
Stack 2

Stack 3

Stack X

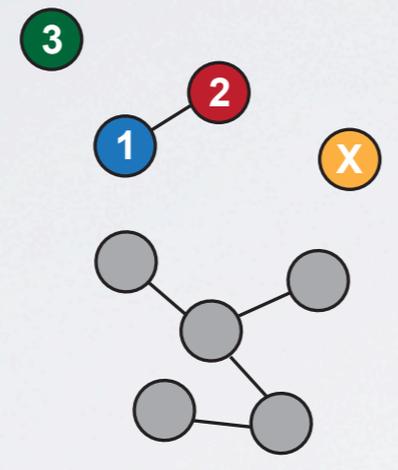
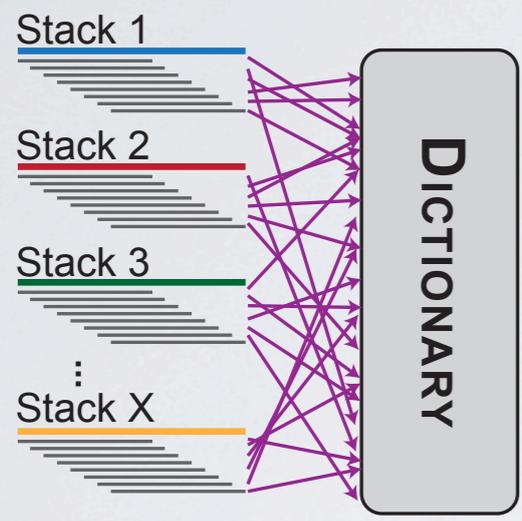
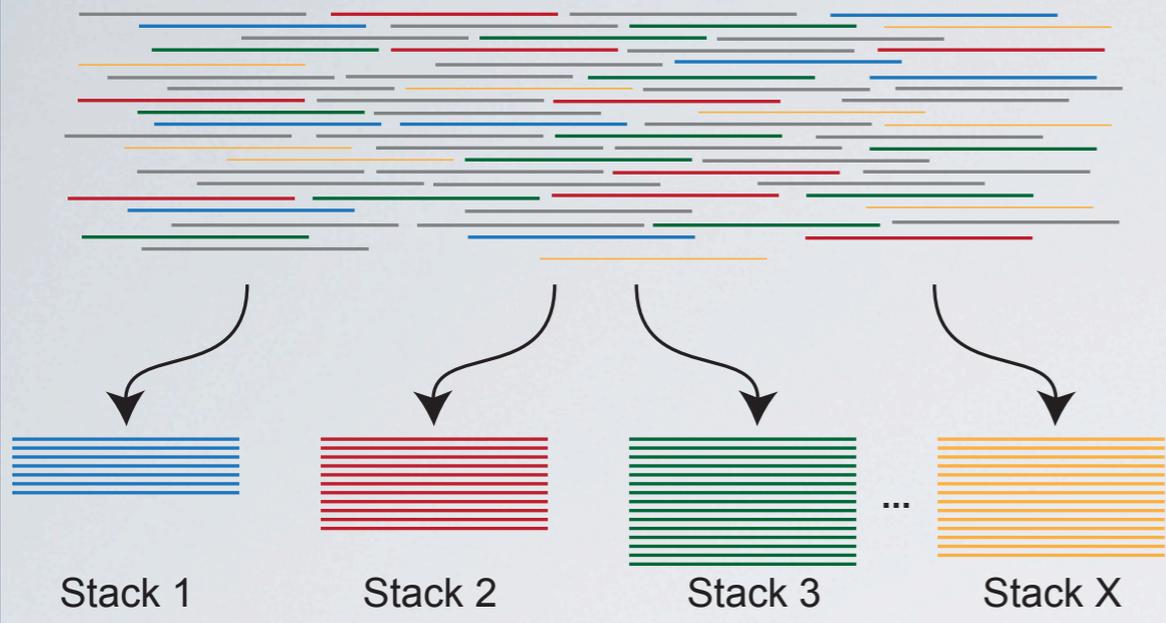
C

CT



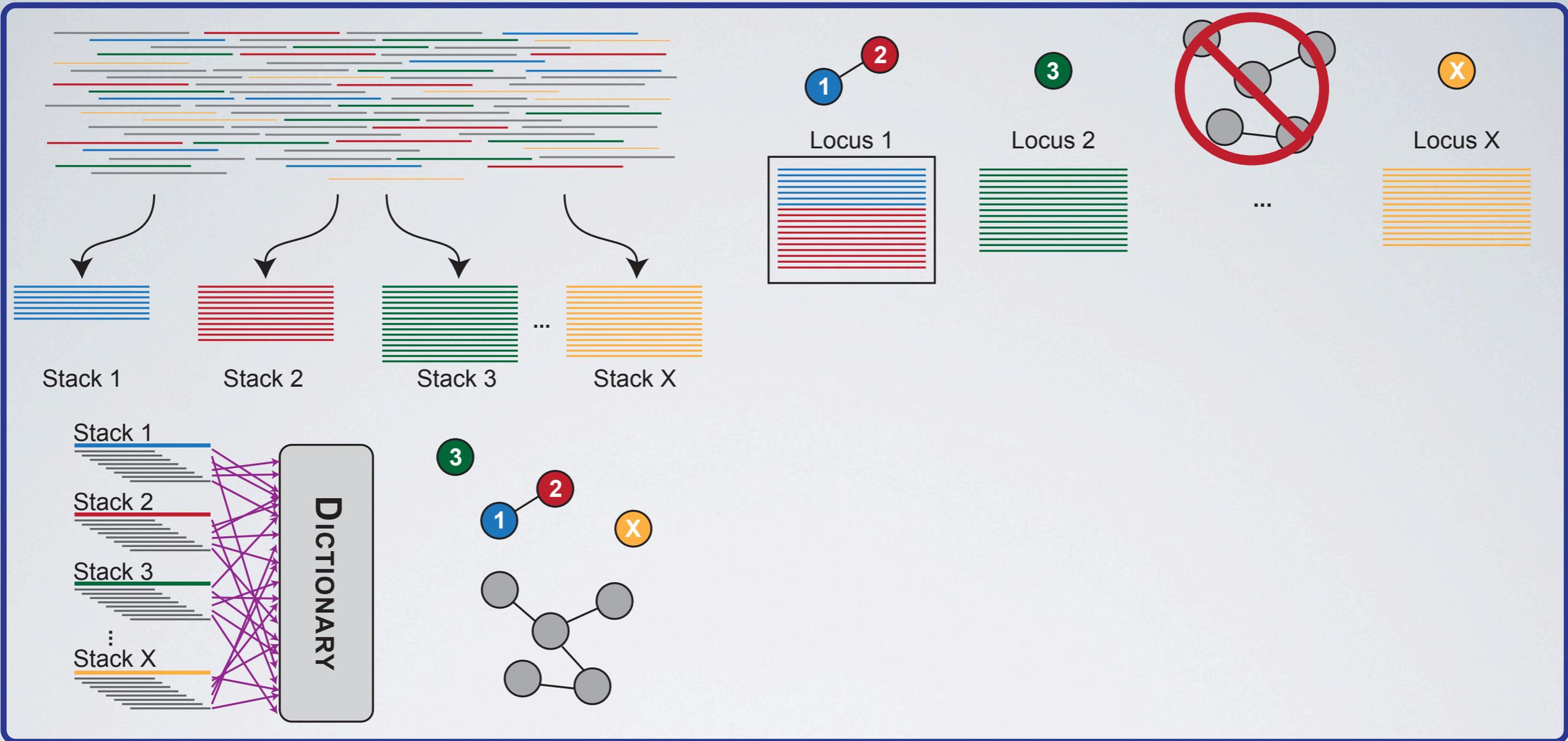
c

CT



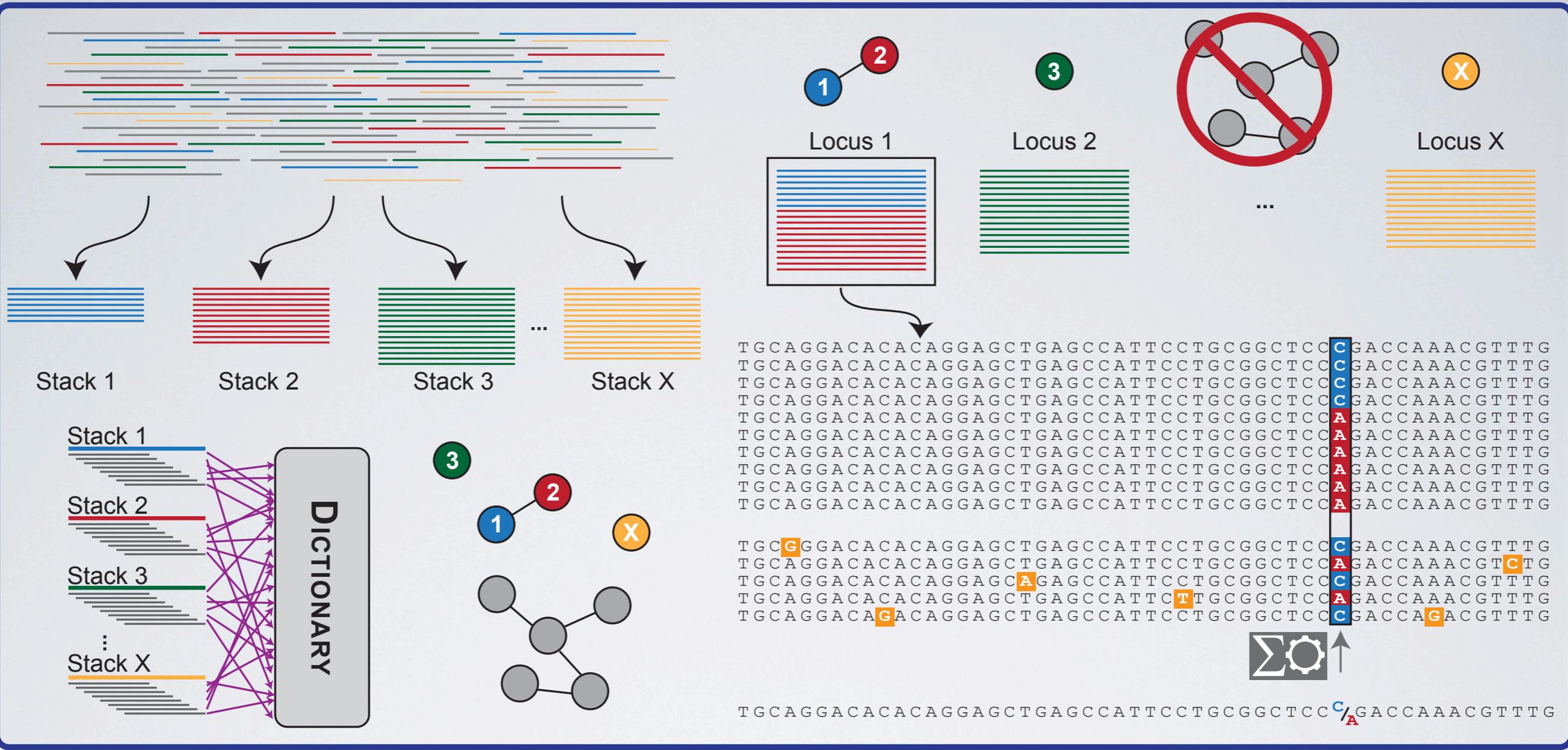
C

CT



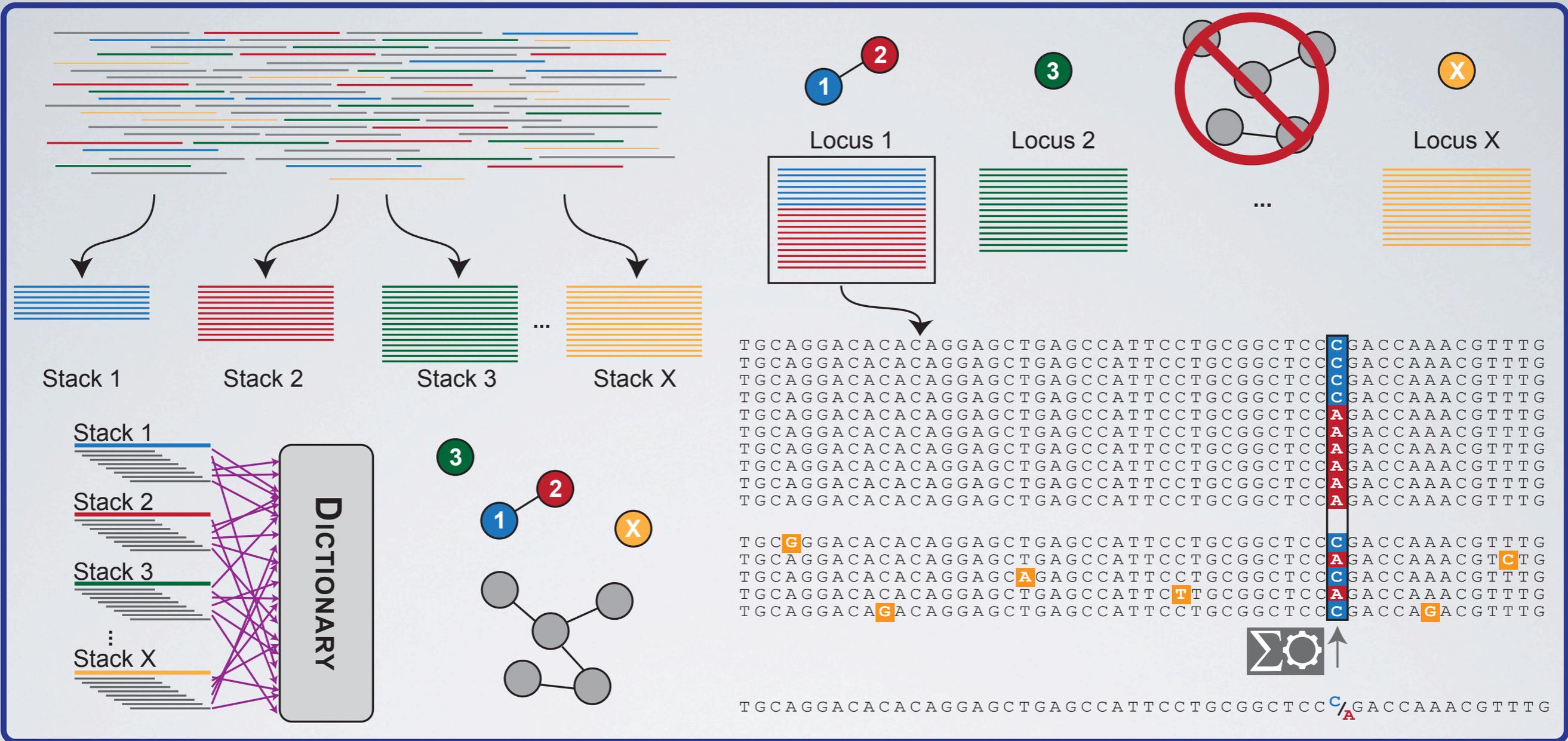
c

CT

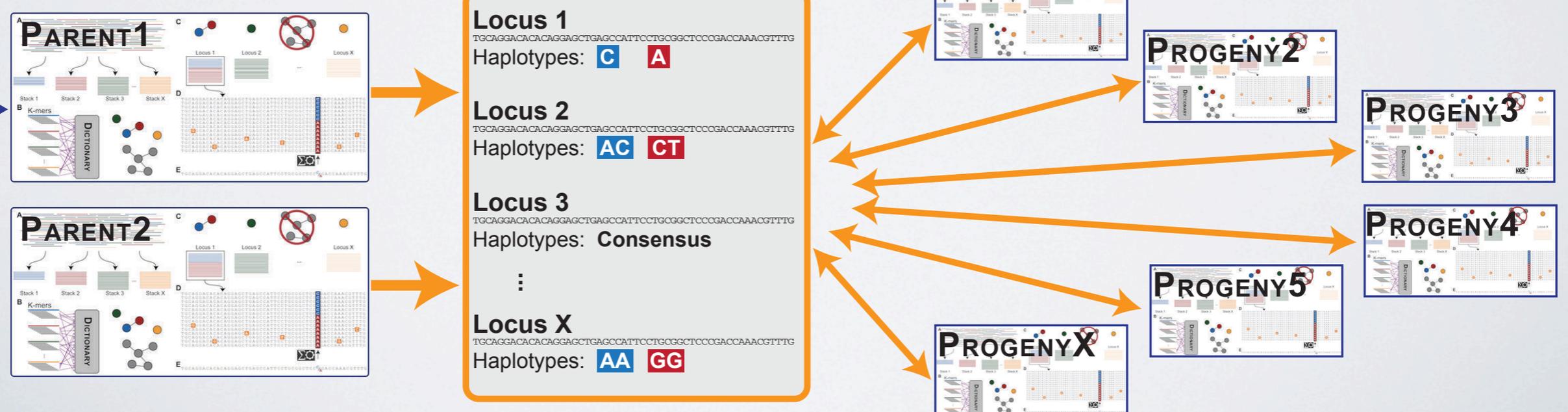


C

CT



CATALOG



Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication

Angel Amores,* Julian Catchen,¹ Allyse Ferrara,² Quenton Fontenot,² and John H. Postlethwait*¹

*Institute of Neuroscience and ¹Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon 97403, and ²Nicholls State University, Department of Biological Sciences, Thibodaux, Louisiana 70310



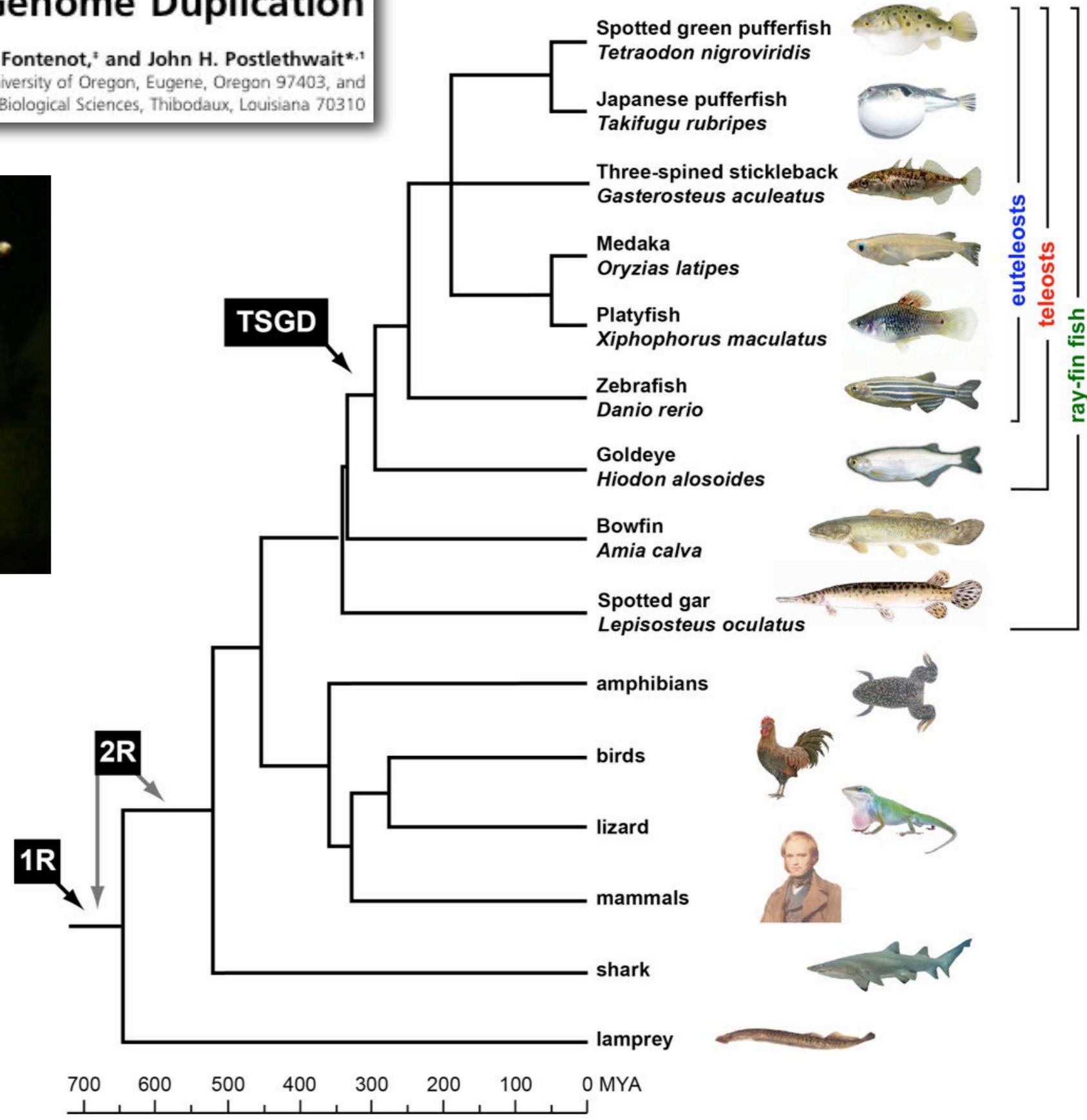
Spotted Gar



Angel Amores



John Postlethwait



Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication

Angel Amores,* Julian Catchen,¹ Allyse Ferrara,² Quenton Fontenot,² and John H. Postlethwait*¹

*Institute of Neuroscience and ¹Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon 97403, and ²Nicholls State University, Department of Biological Sciences, Thibodaux, Louisiana 70310



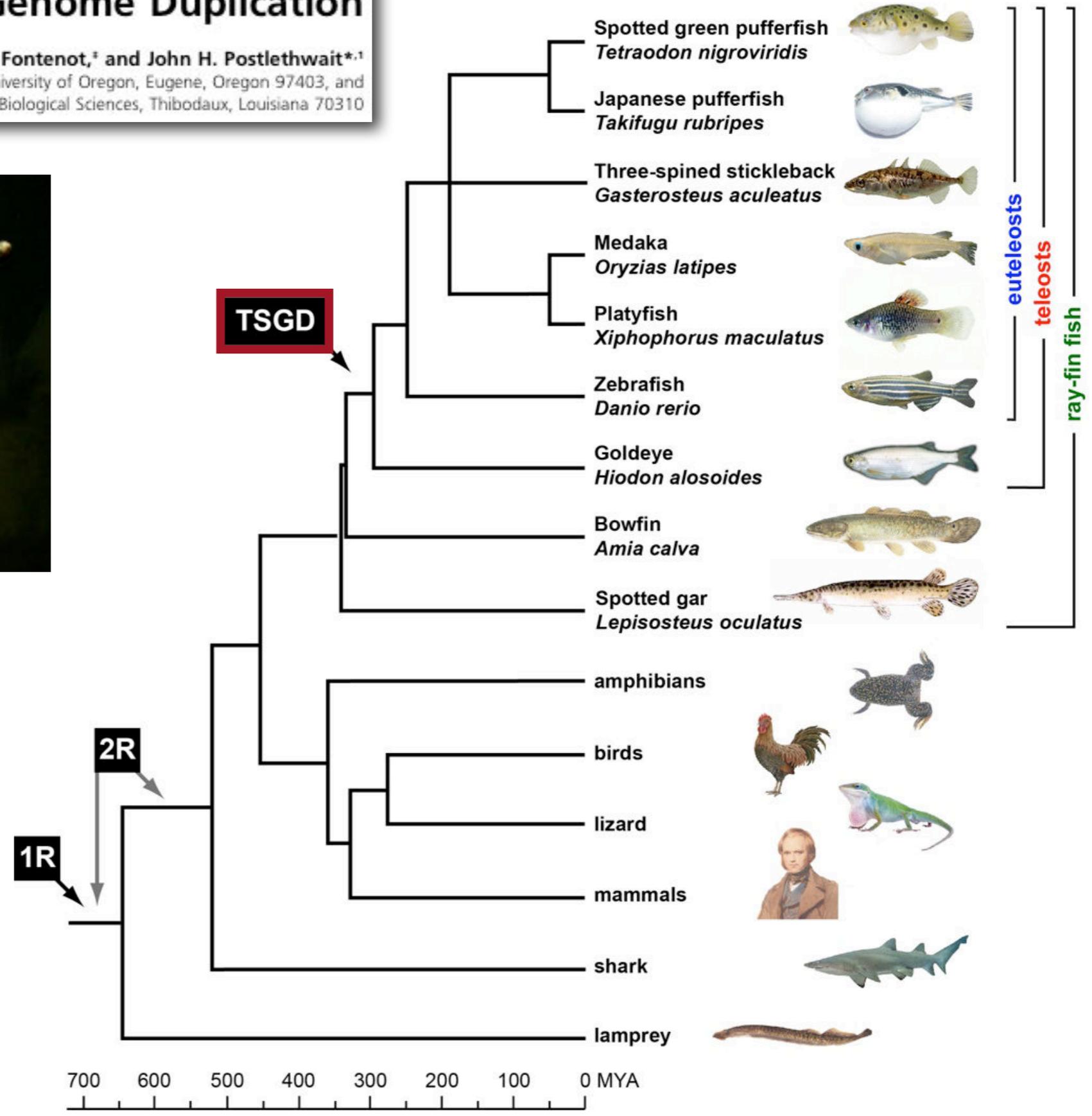
Spotted Gar



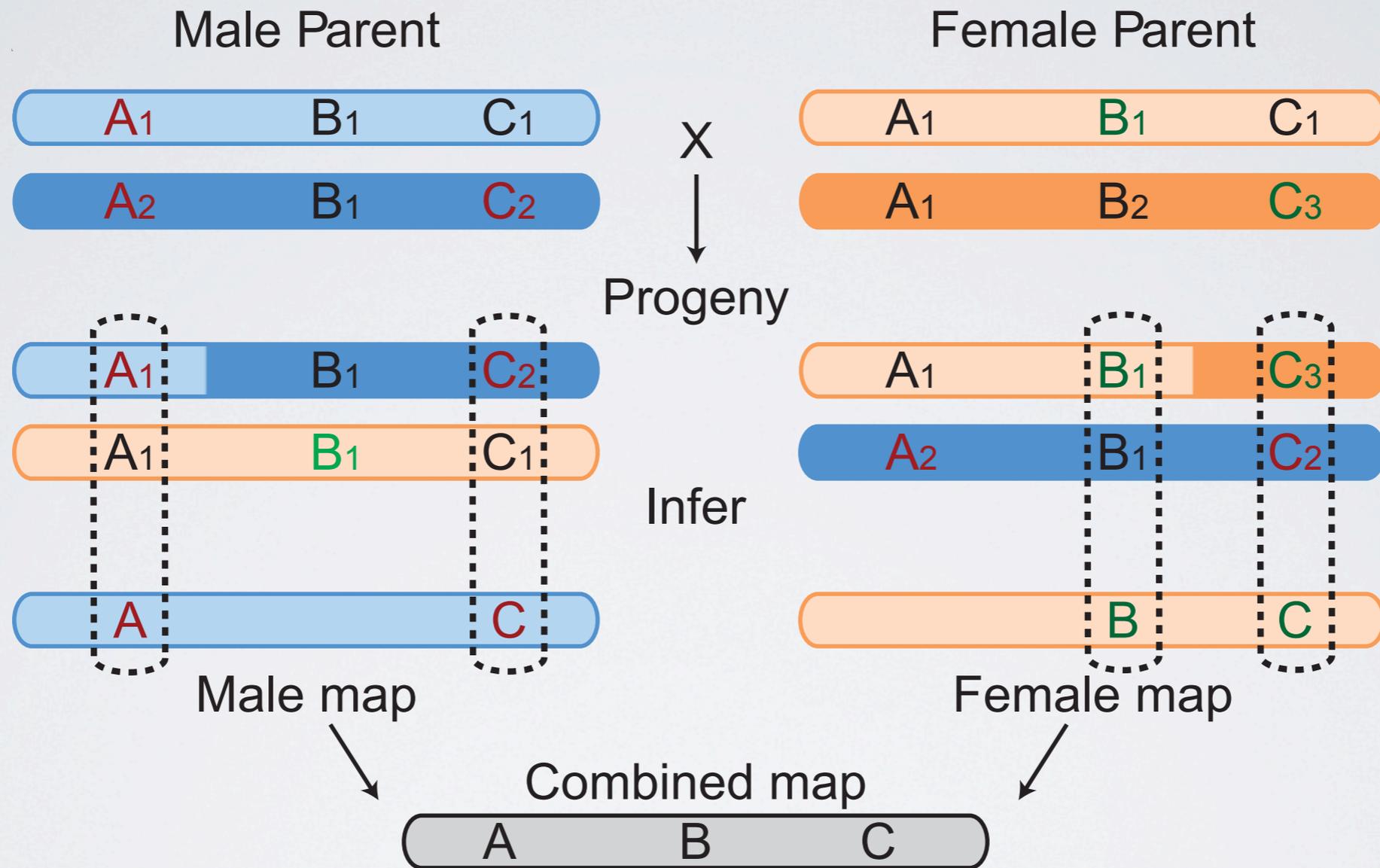
Angel Amores

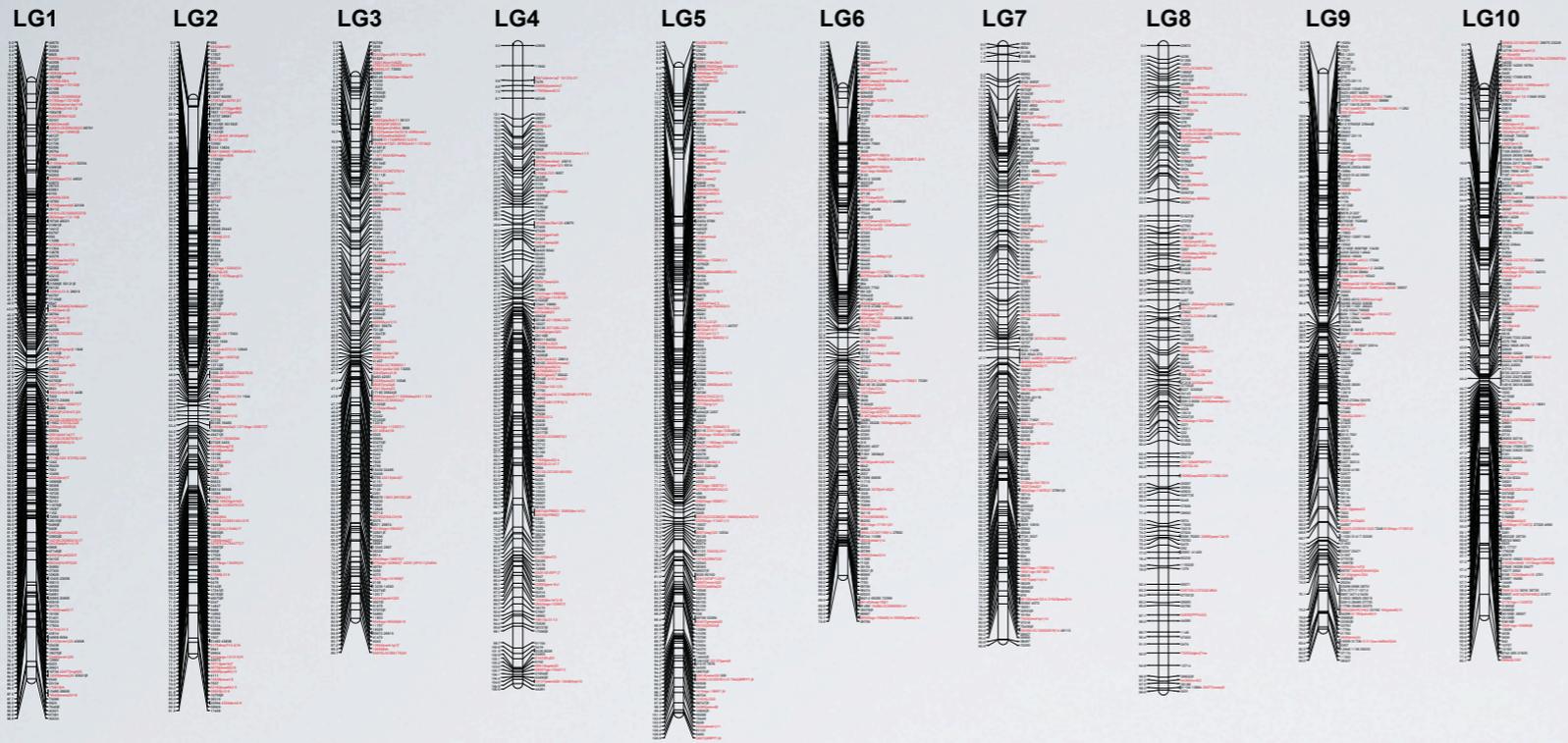


John Postlethwait

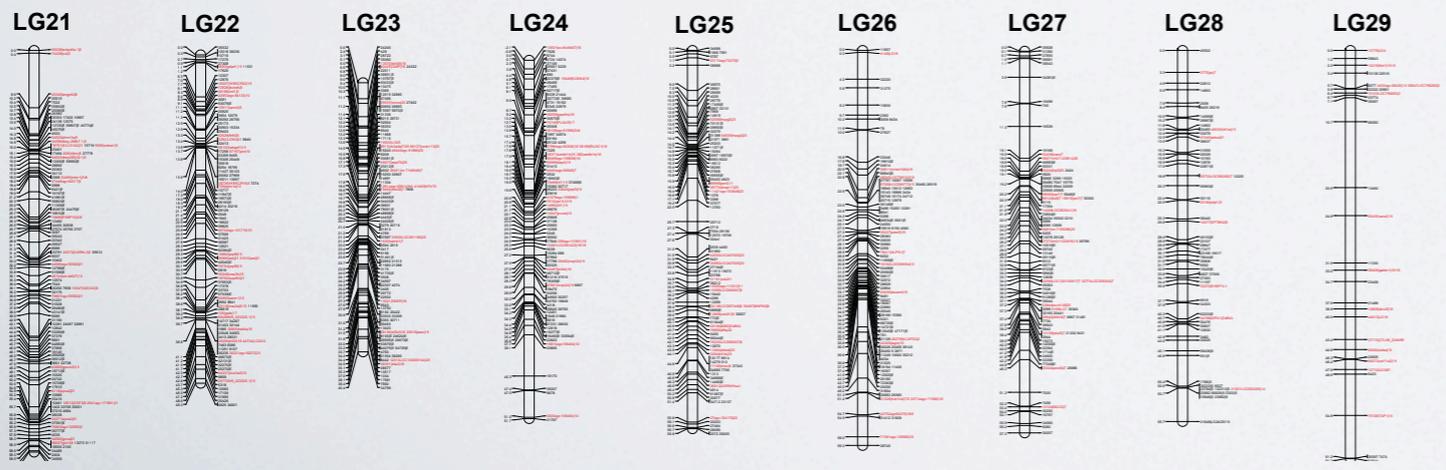
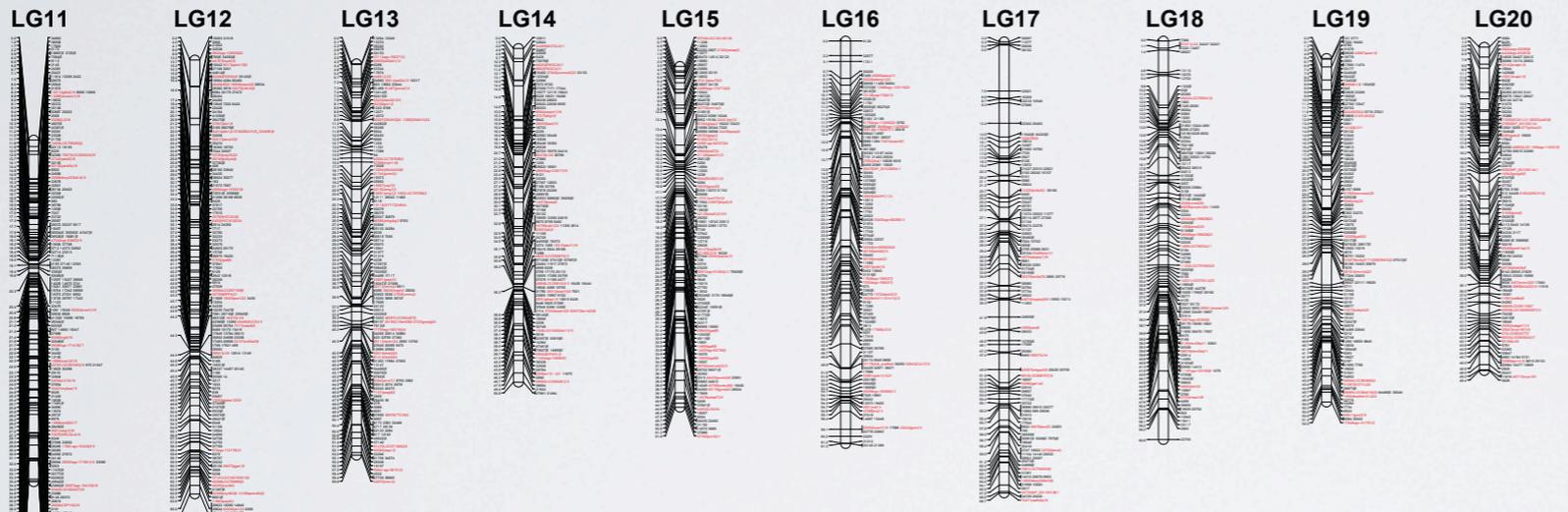


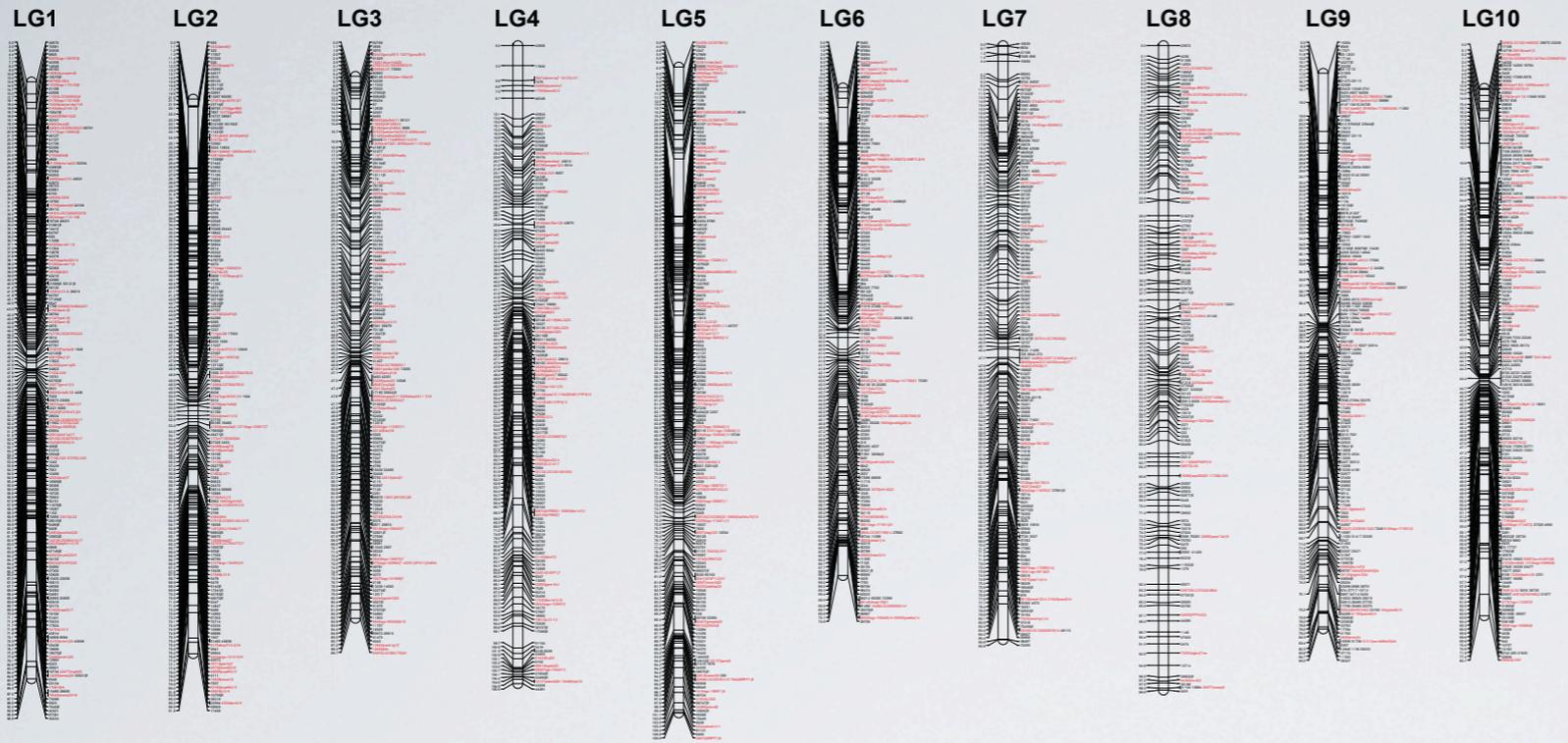
F1 Pseudo-Test Cross



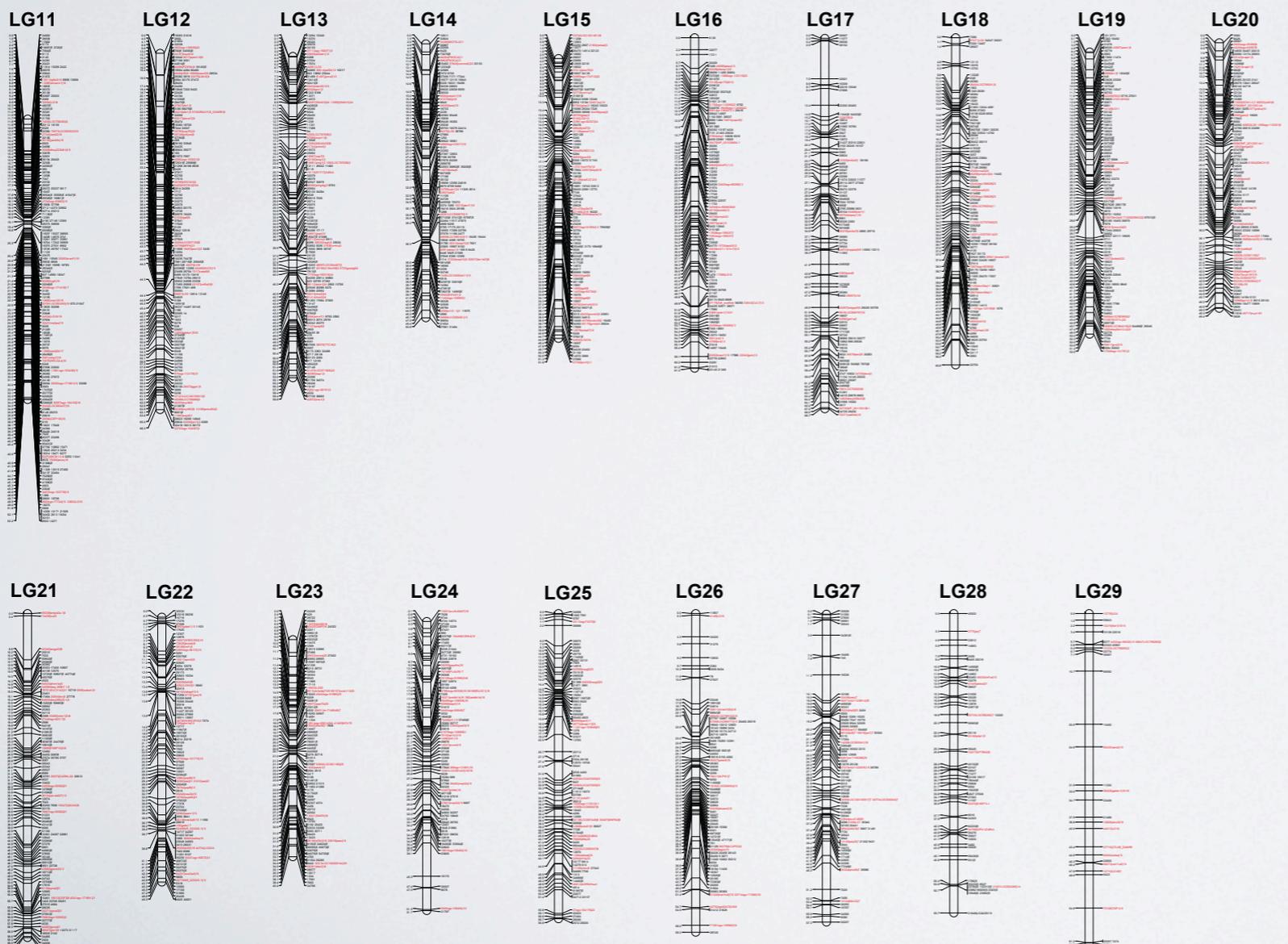


94 Individuals
15,076 Markers
 8,046 Mapped
 974 In Genes

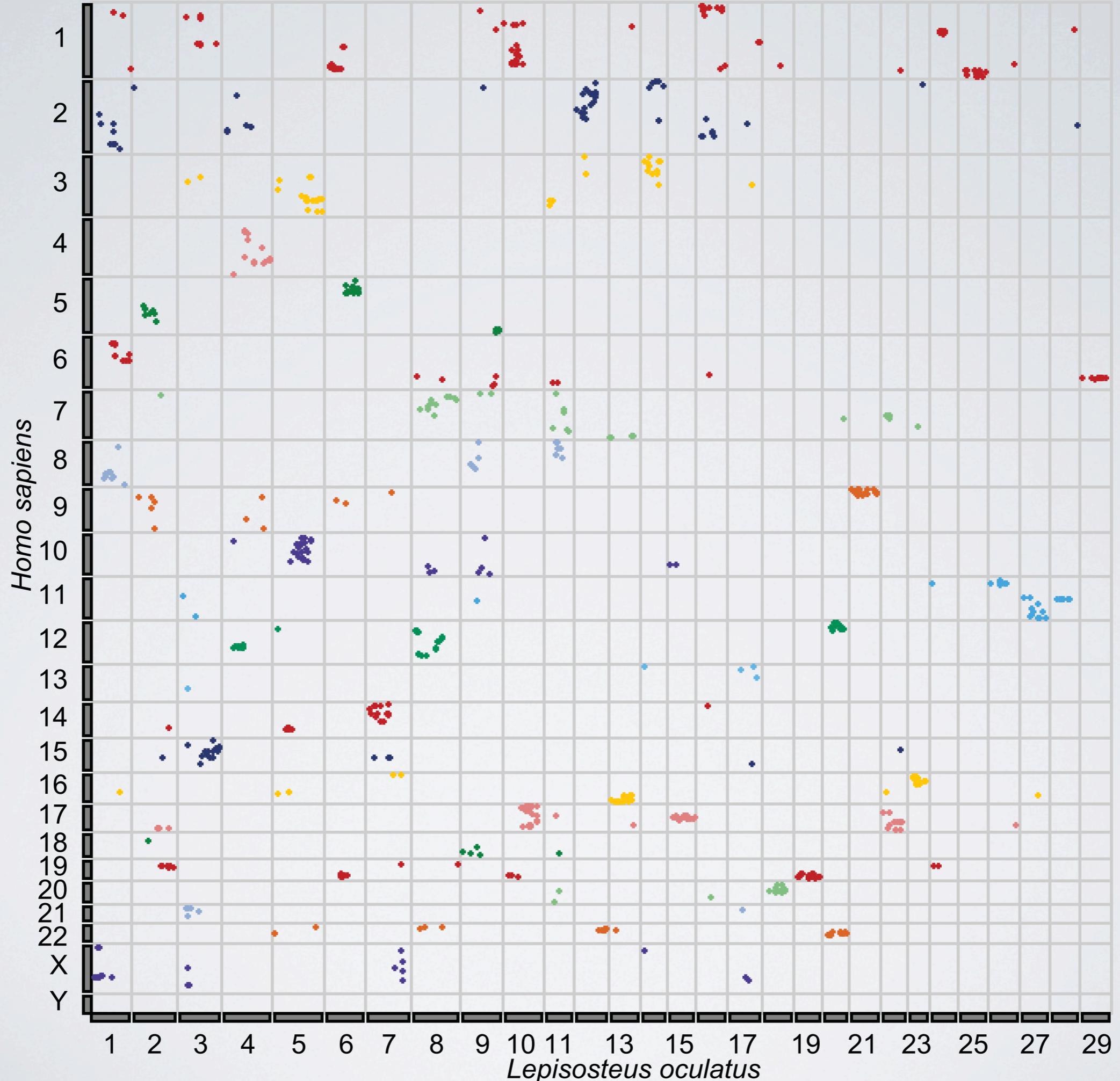


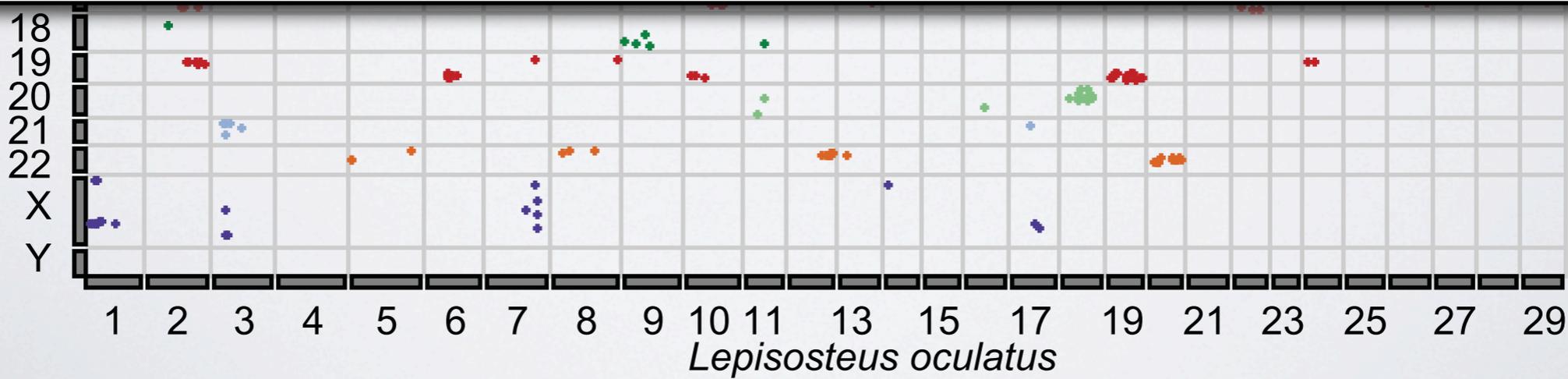
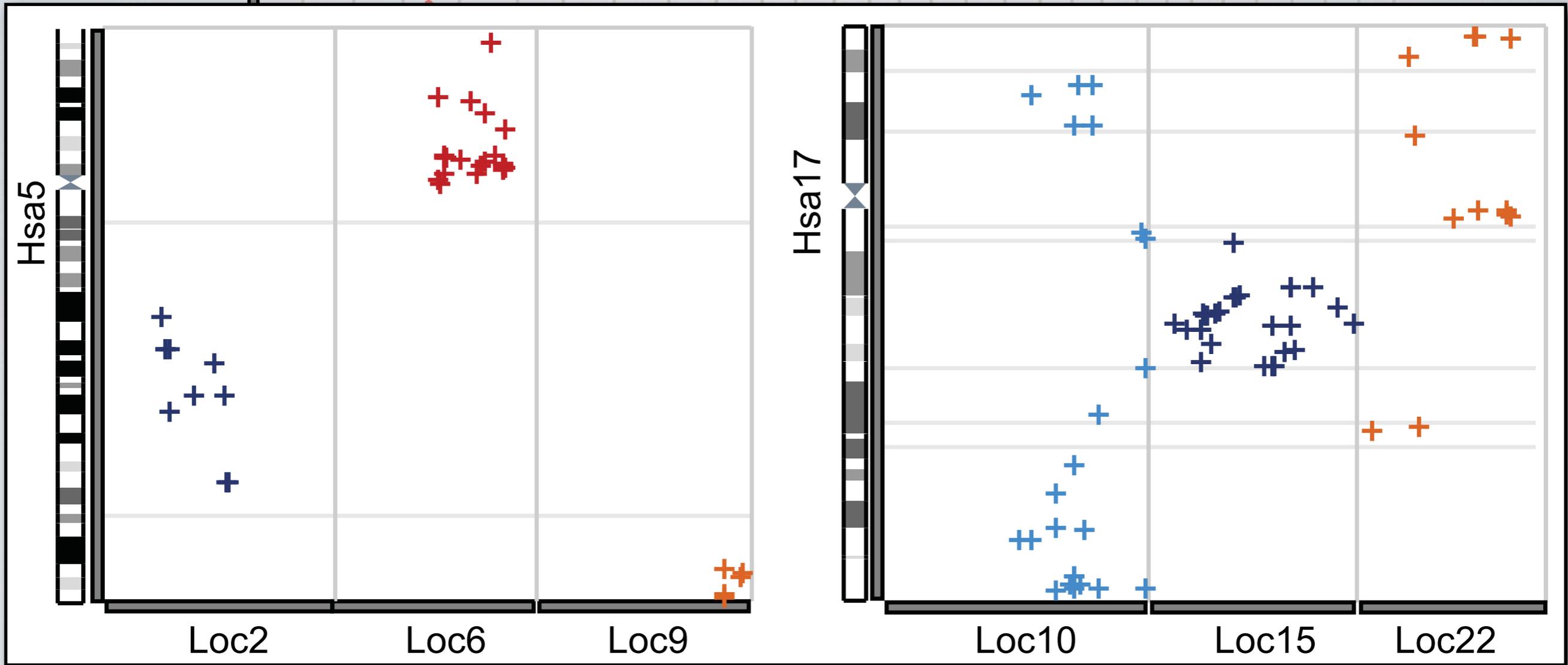
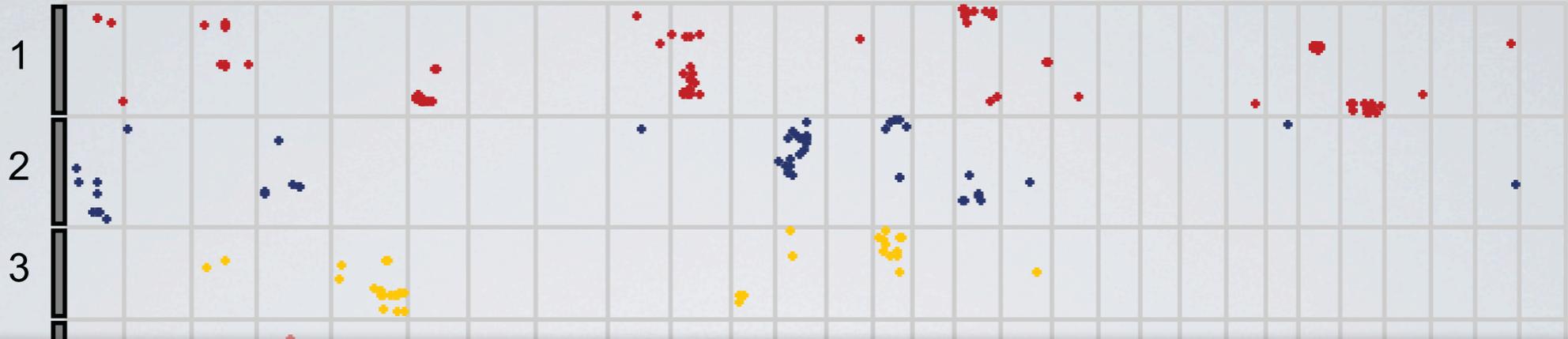


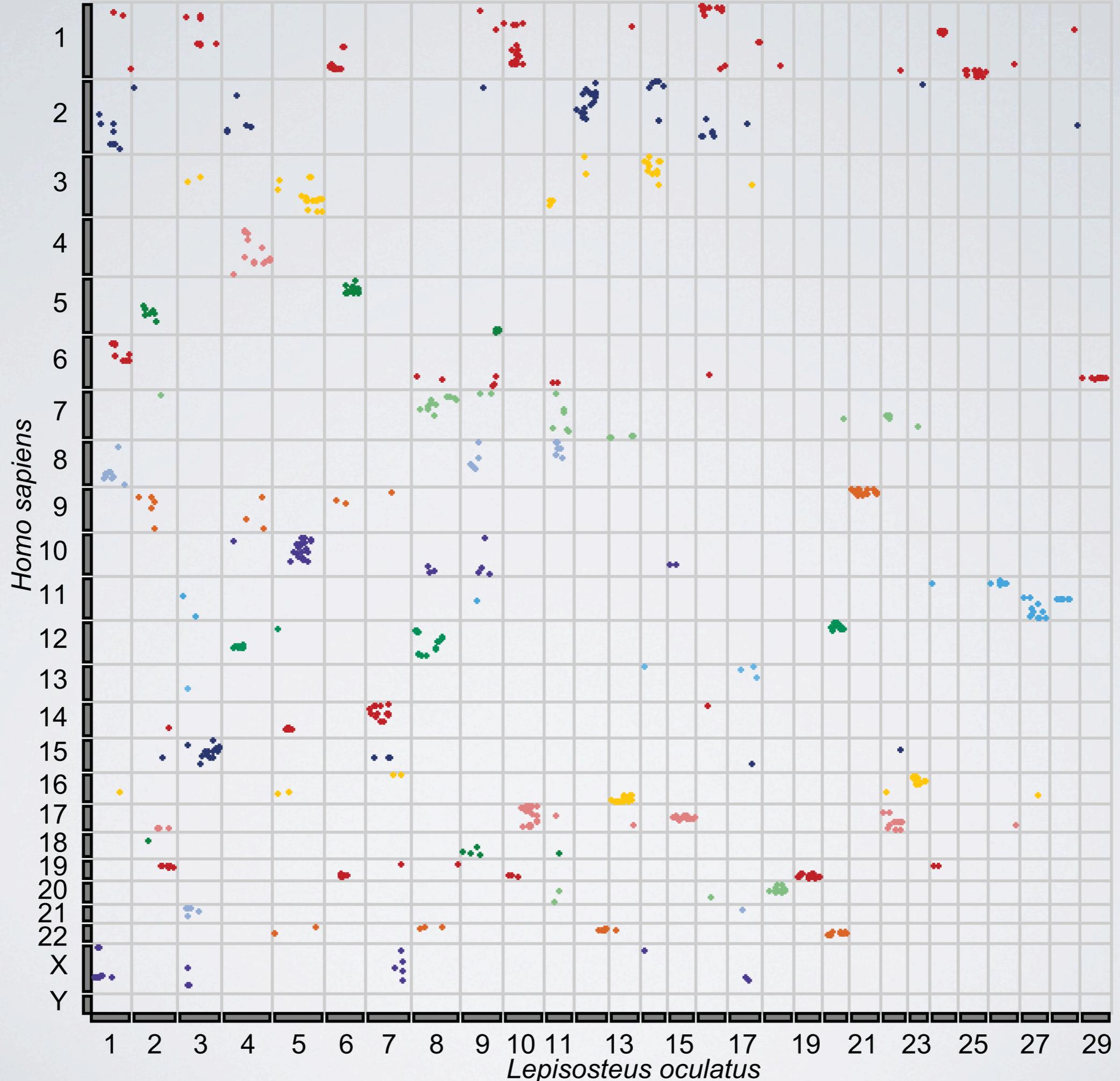
94 Individuals
15,076 Markers
 8,046 Mapped
 974 In Genes

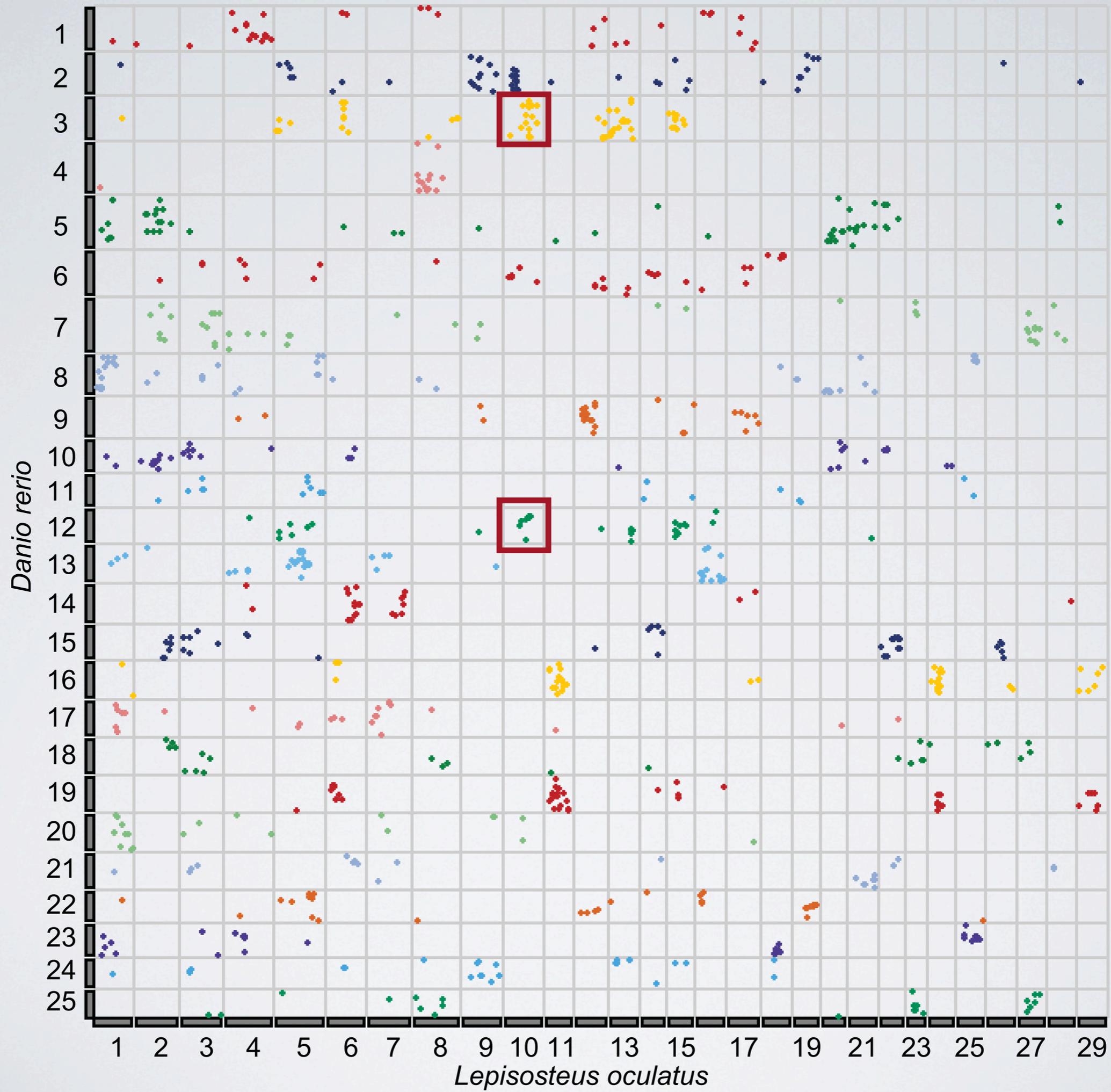


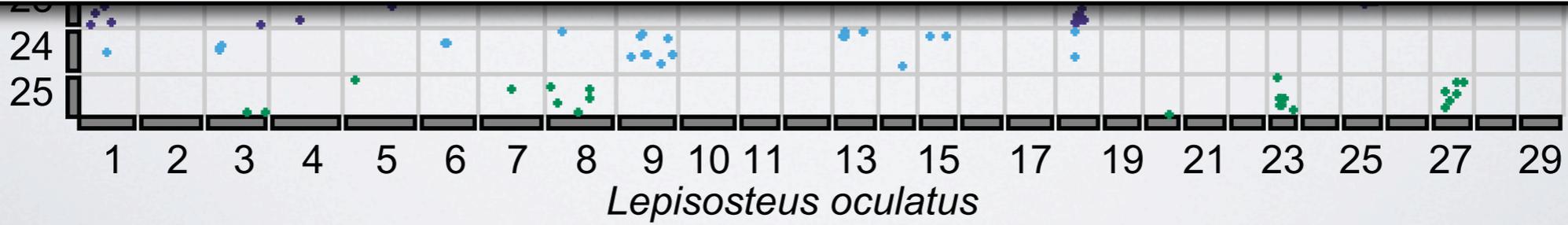
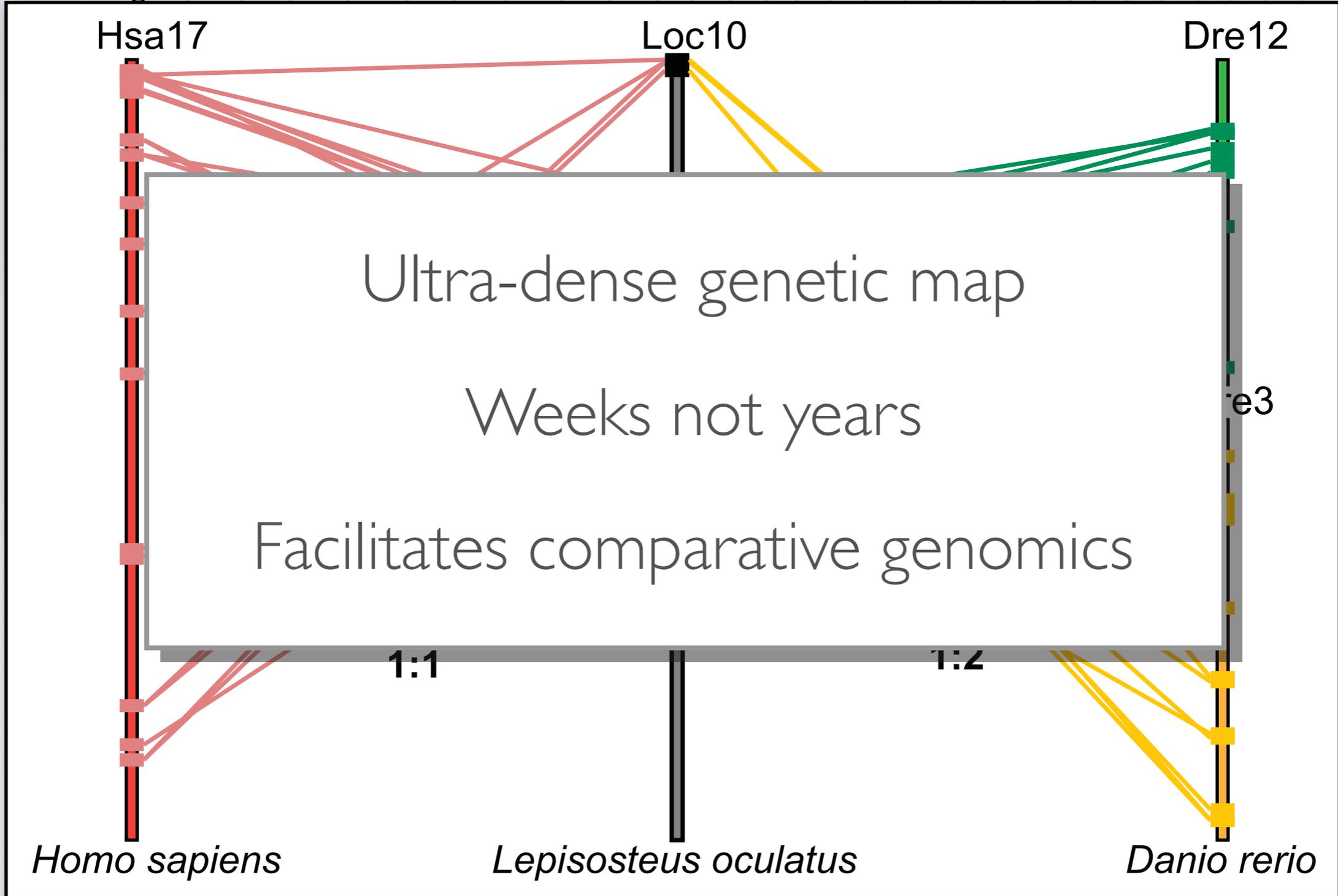
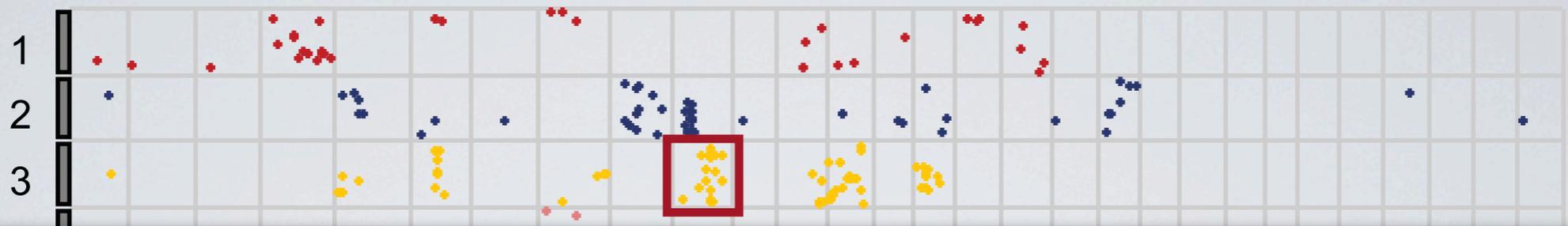
Organism	Markers
Silver carp	483
Guppy	790
Barramundi	240
Catfish	331
Sea bass	368
Cichlid	204
Platyfish	290
Halibut	604
Sea bream	204





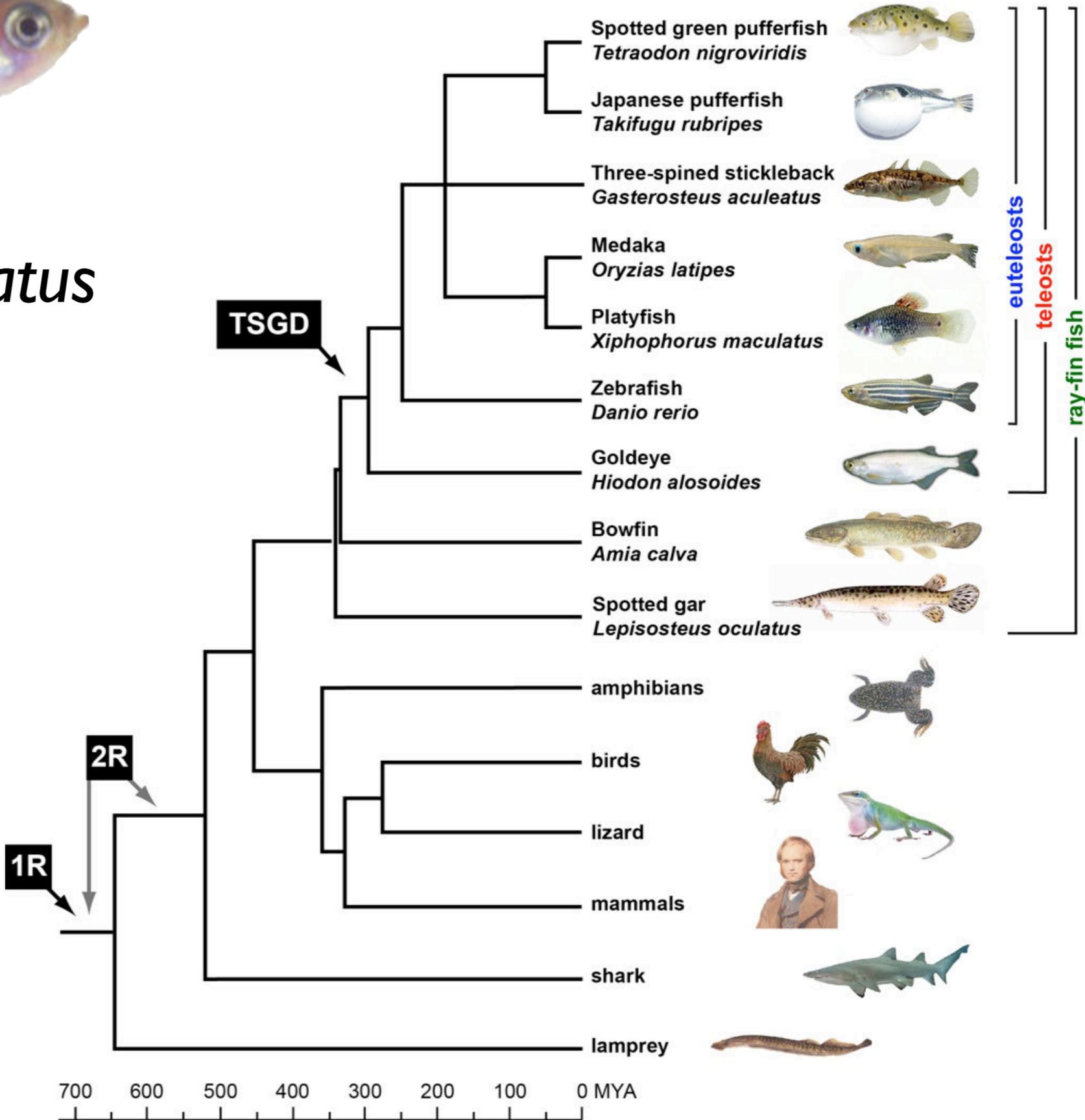








Xiphophorus maculatus



Angel Amores



John Postlethwait

Version 1.0

Copyright © 2004 by the Genetics Society of America
DOI: 10.1534/genetics.103.019349

A Microsatellite Genetic Linkage Map for *Xiphophorus*

**R. B. Walter,^{*,1} J. D. Rains,* J. E. Russell,* T. M. Guerra,* C. Daniels,* Dennis A. Johnston,[†]
Jay Kumar,[†] A. Wheeler,* K. Kelnar,* V. A. Khanolkar,* E. L. Williams,* J. L. Hornecker,*
L. Hollek,* M. M. Mamerow,* A. Pedroza* and S. Kazianis***

**Molecular Biosciences Research Group, Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas 78666 and
[†]Science Park Research Division, MD Anderson Cancer Center, University of Texas, Smithville, Texas 78957*

290 markers

Version 2.0

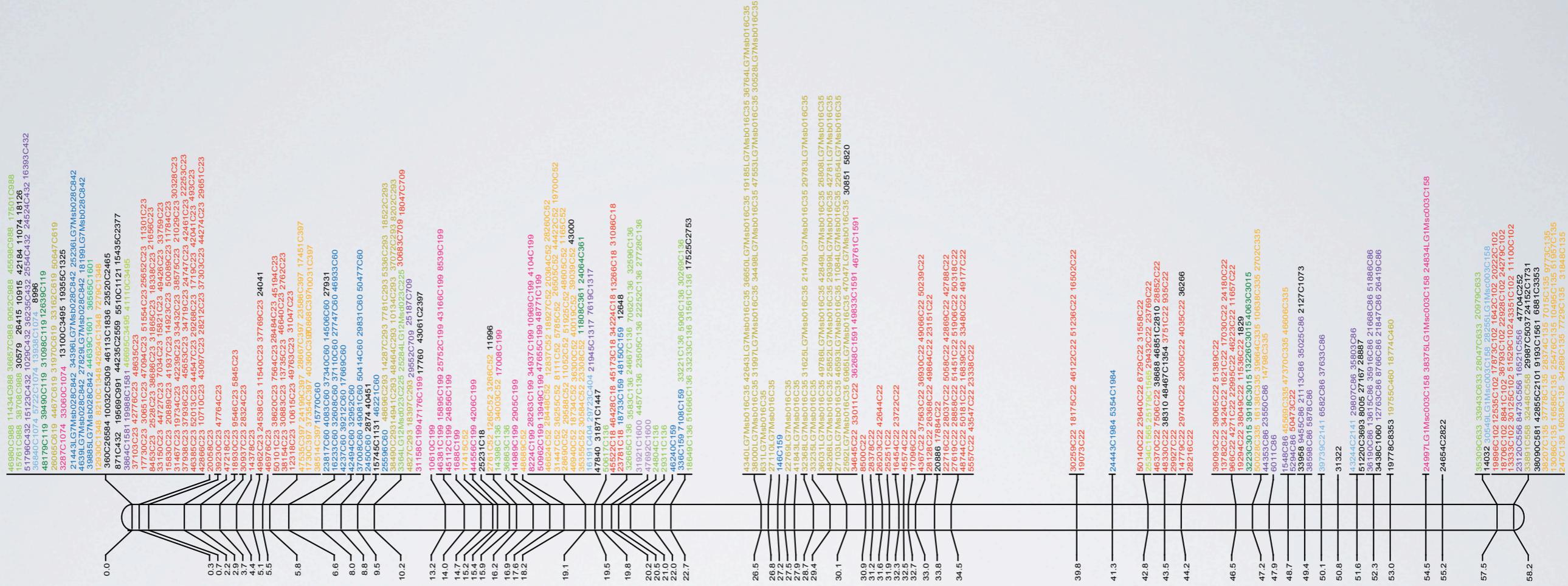
269 total fish

828,373,617 raw reads

613,162,521 incorporated reads

22,144 markers in at least 90 progeny

LG7



Given 80,000 contigs: 86% of genome ordered

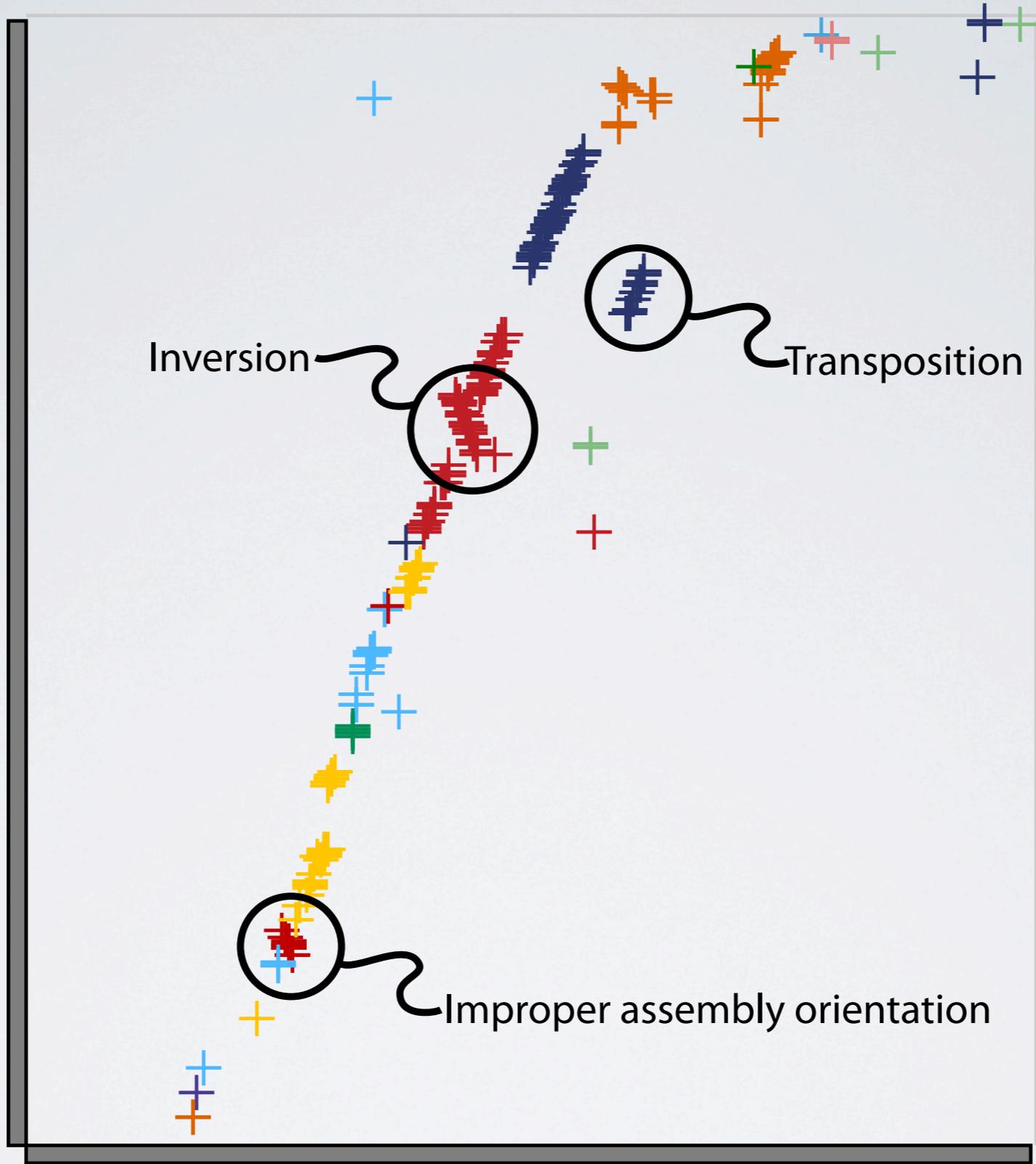


groupXXI





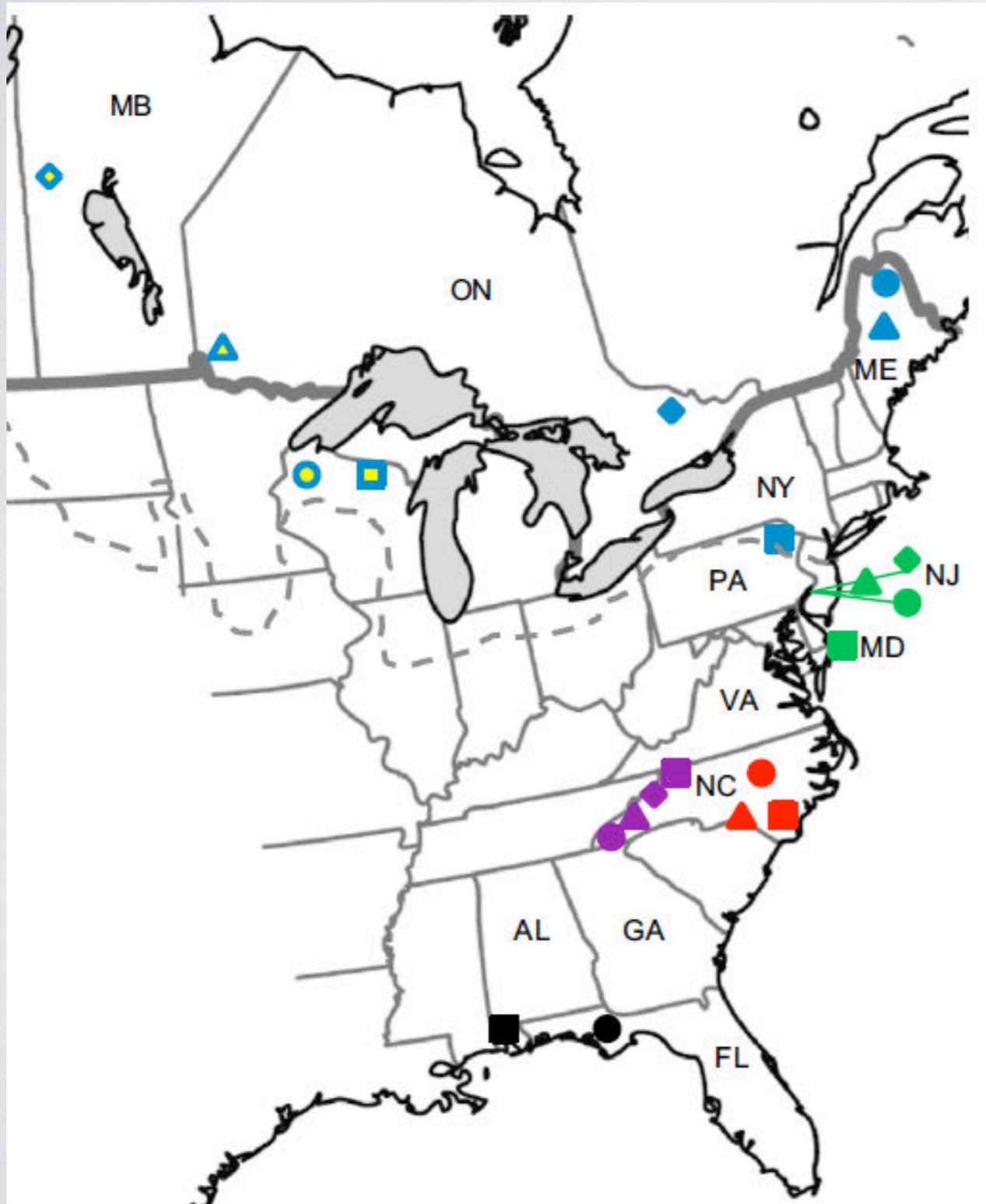
groupXXI



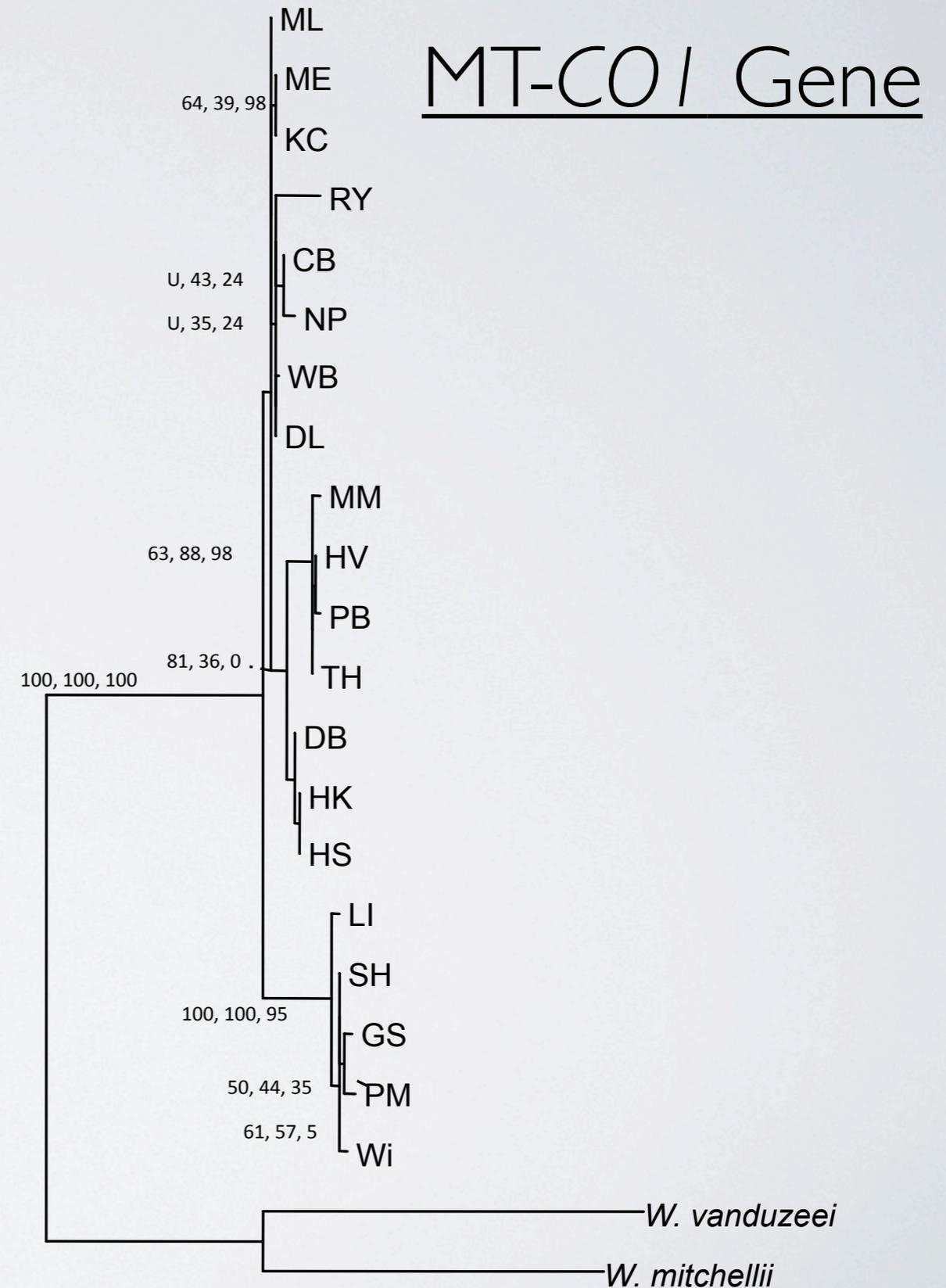
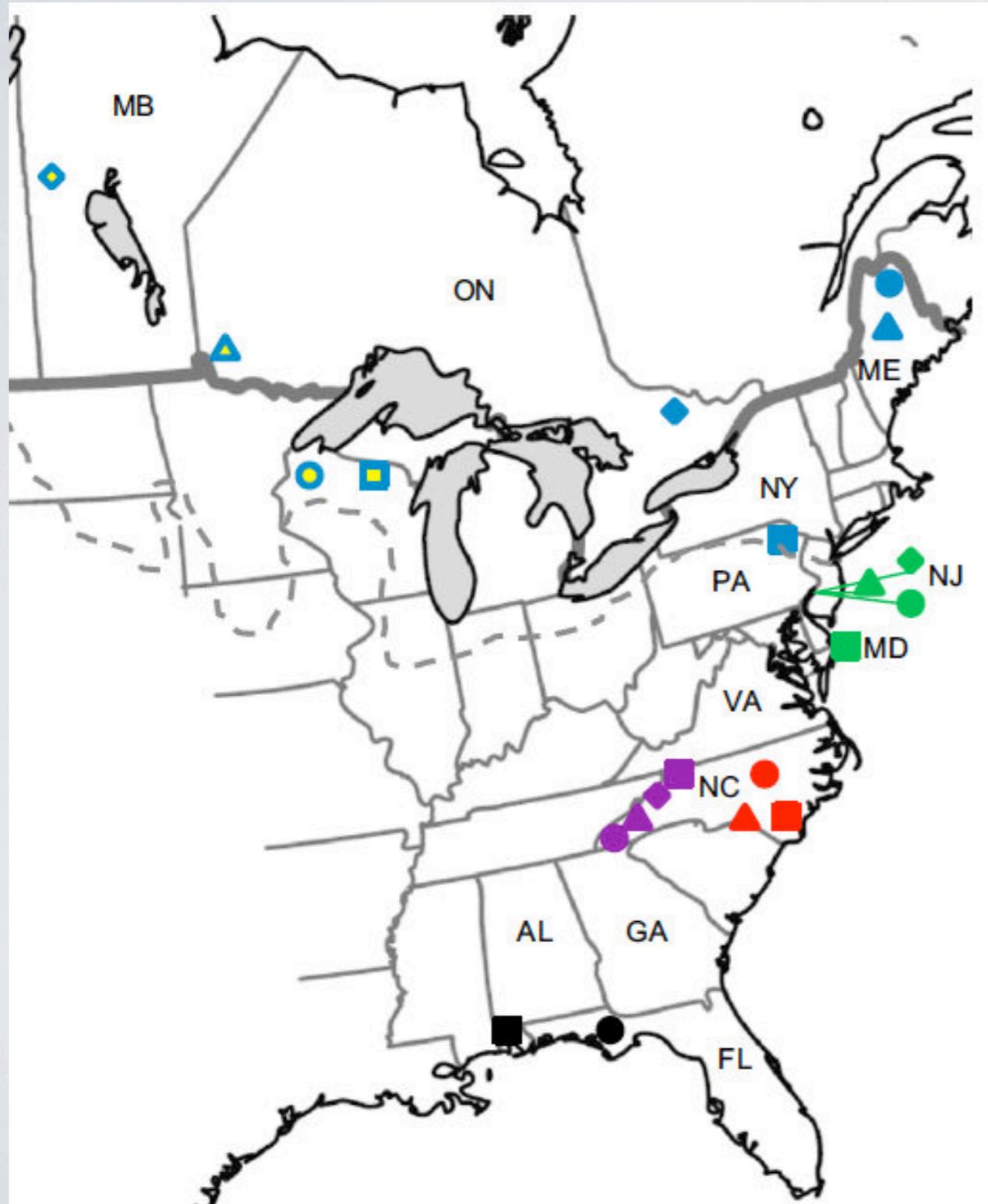
Ig21



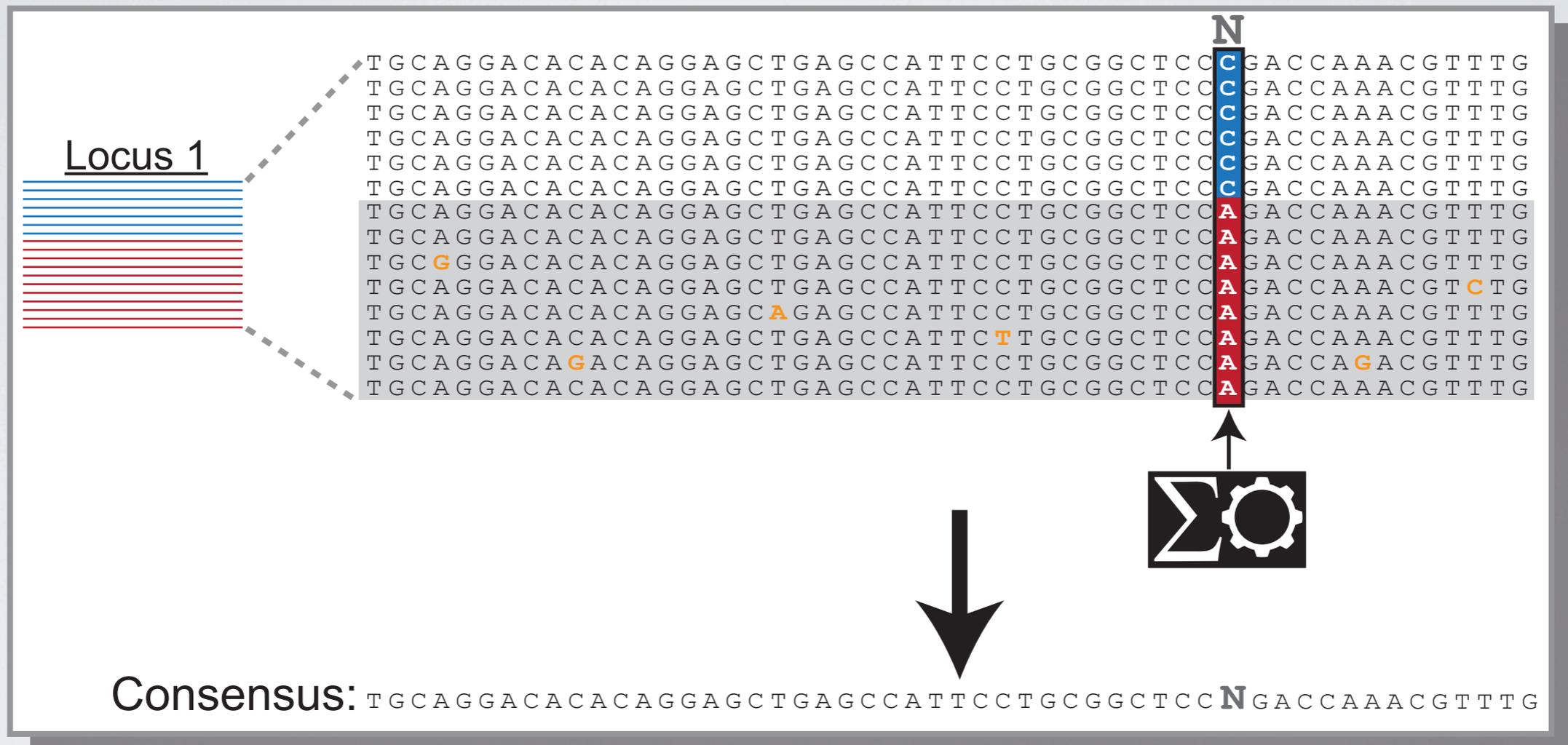
The pitcher plant mosquito, *Wyeomyia smithii* a model for evolution due to climate change



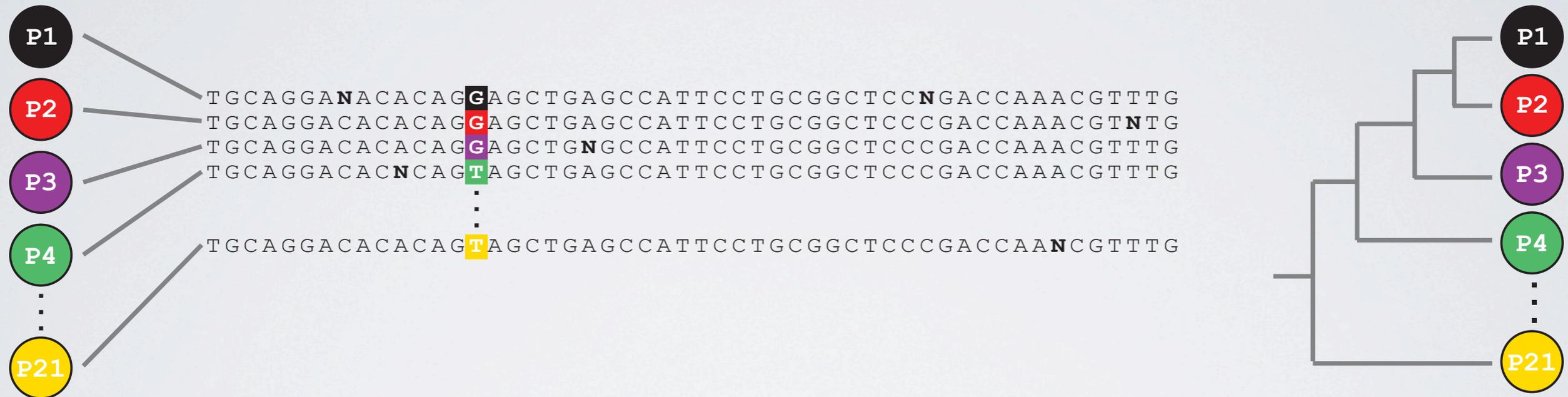
Resolving post-glacial phylogeography in *Wyeomyia smithii*



Sampling RAD-Tag markers for population-level phylogenetic analysis



Sampling RAD-Tag markers for population-level phylogenetic analysis

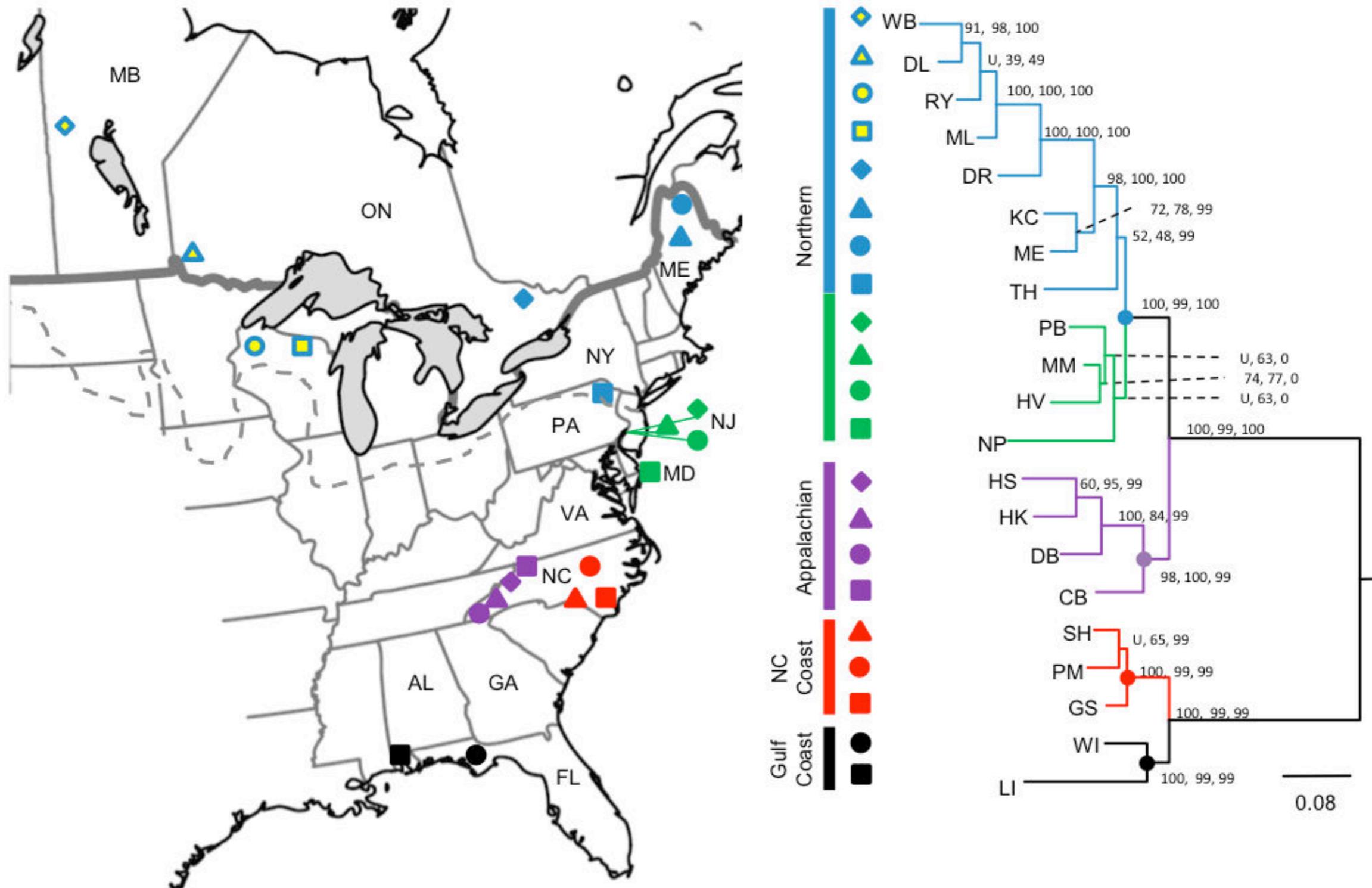


Resolving postglacial phylogeography using high-throughput sequencing

PNAS

Kevin J. Emerson¹, Clayton R. Merz, Julian M. Catchen, Paul A. Hohenlohe, William A. Cresko, William E. Bradshaw, and Christina M. Holzapfel

Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR 97403-5289



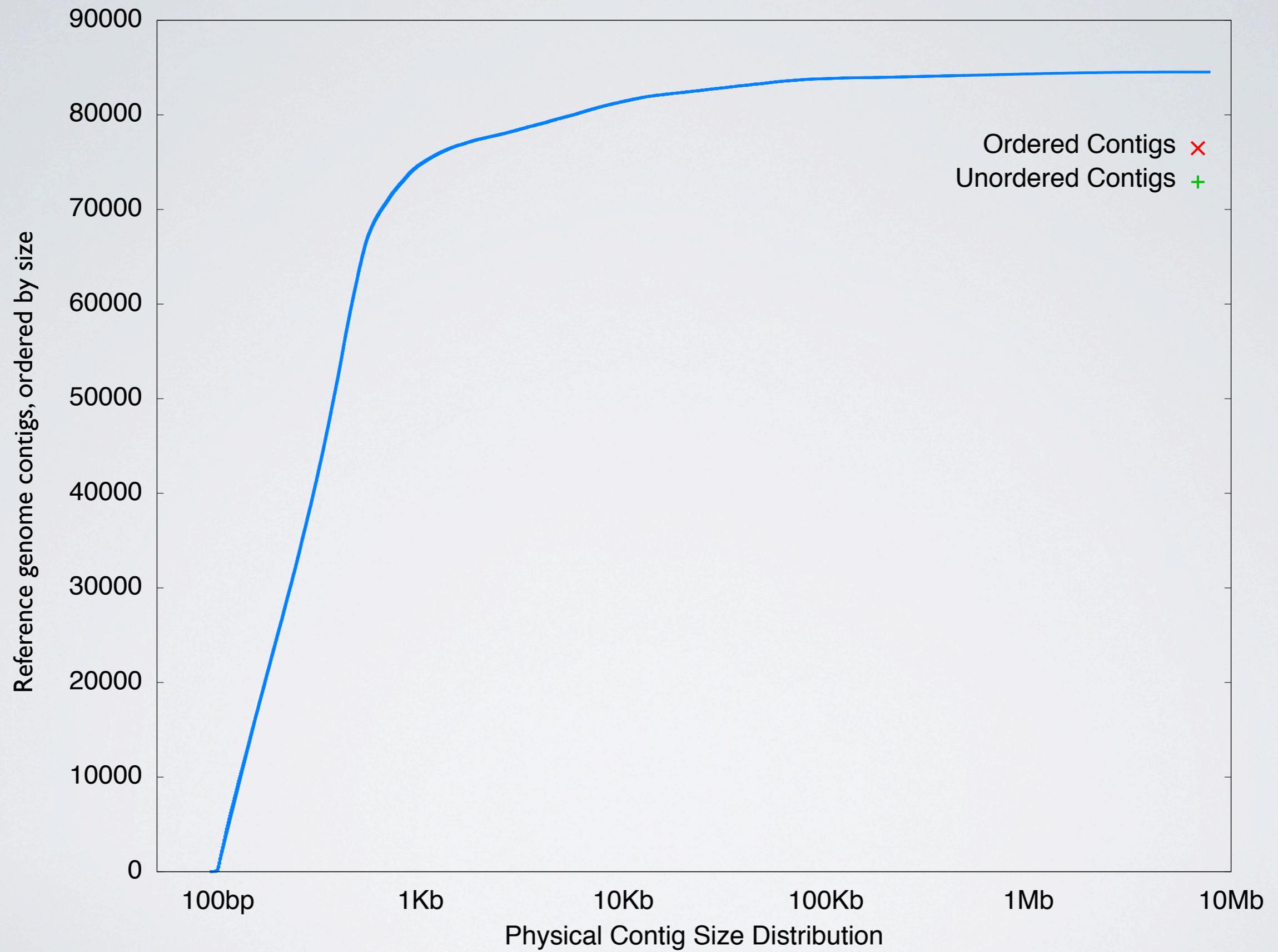
SPECIAL THANKS TO

- *Present lab members:*

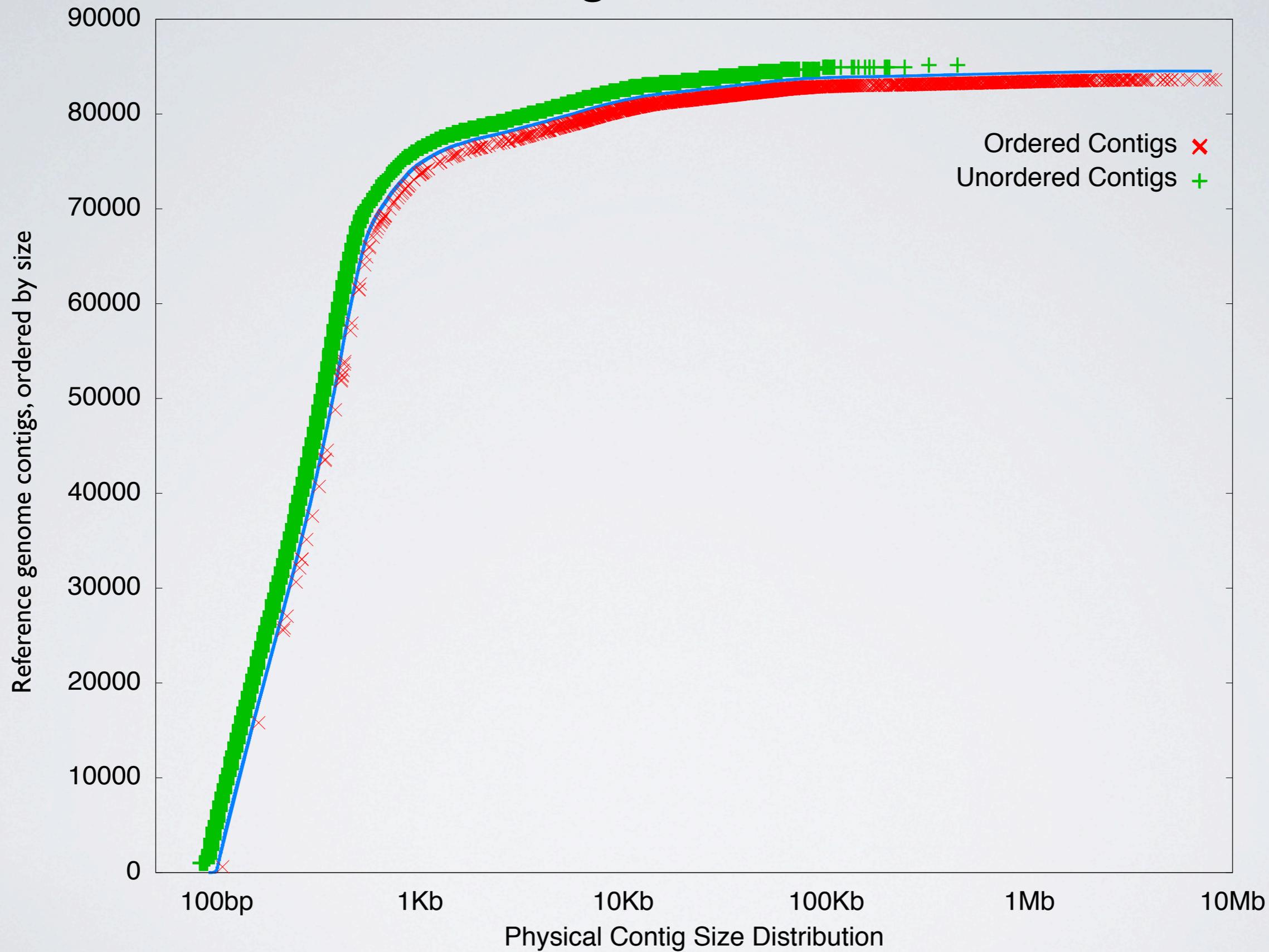


- *Cresko lab members* Susan Bassham, Paul Hohenlohe
- *Collaborators* Angel Amores, John Postlethwait, Kevin Emerson, William Bradshaw, Christina Holzapfel, Eric Johnson, Patrick Phillips, Paul Etter
- *Funding* from NSF & NIH, as well as Keck & Murdock Foundations





1,814 contigs, 651Mb / 746Mb





Acquire
paired-end
sequence



Acquire
paired-end
sequence



↓ Match to marker catalog

TGCAGGGGTATTAGCATAA

↓ Collate/Assemble PE reads

AACTAATTTTCACTAGCCATCTTGAATGTGAGTAGCATTTTAAGTAACTATAATTG

Acquire
paired-end
sequence



Match to marker catalog

TGCAGGGGTATTAGCATAA

Collate/Assemble PE reads

AACTAATTTTTCACTAGCCATCTTGAATGTGAGTAGCATTTTAAGTAACTATAATTG

Associate
markers / PE
contigs with
ESTs

BLASTn

BLASTn

EST Library



Acquire
paired-end
sequence



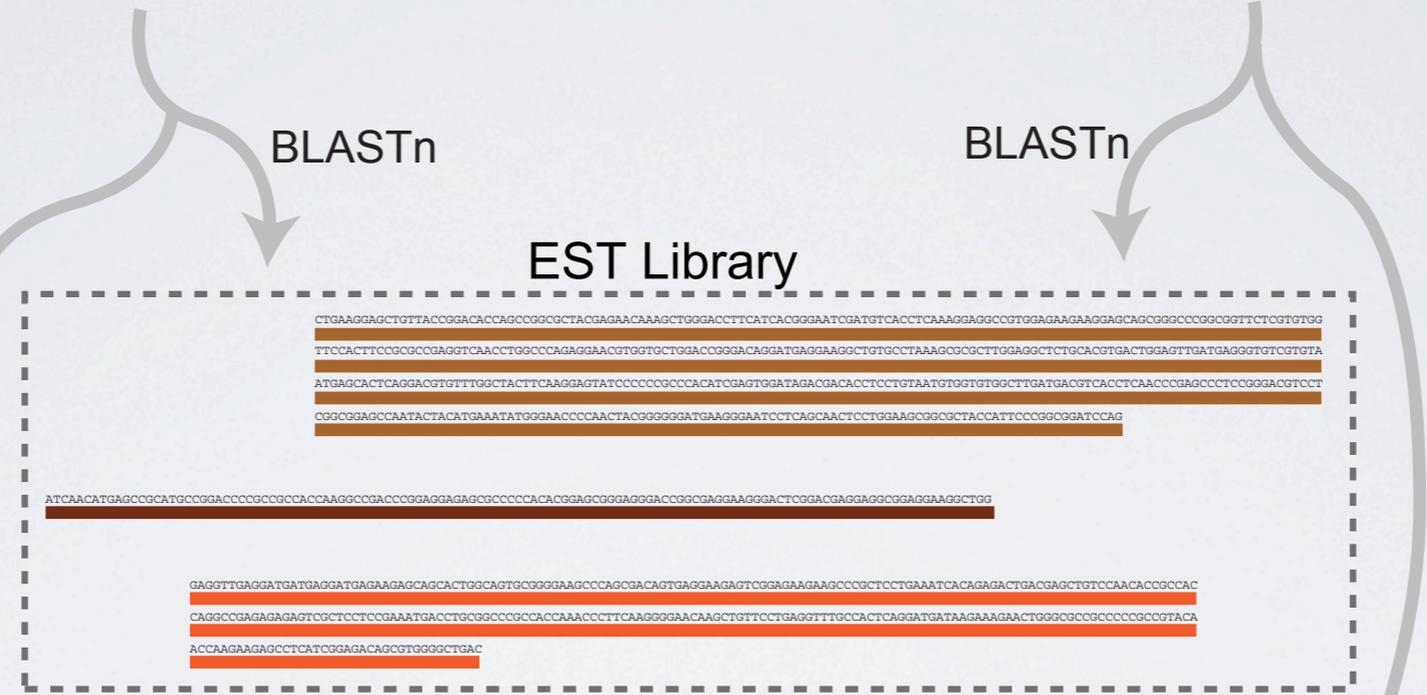
Match to marker catalog

TGCAGGGGTATTAGCATAA

Collate/Assemble PE reads

AACTAATTTTTCACTAGCCATCTTGAATGTGAGTAGCATTTTAAAGTAACTATAATTG

Associate
markers / PE
contigs with
ESTs



Assign
orthology to:
markers
PE contigs
ESTs

BLASTx

BLASTx

BLASTx

