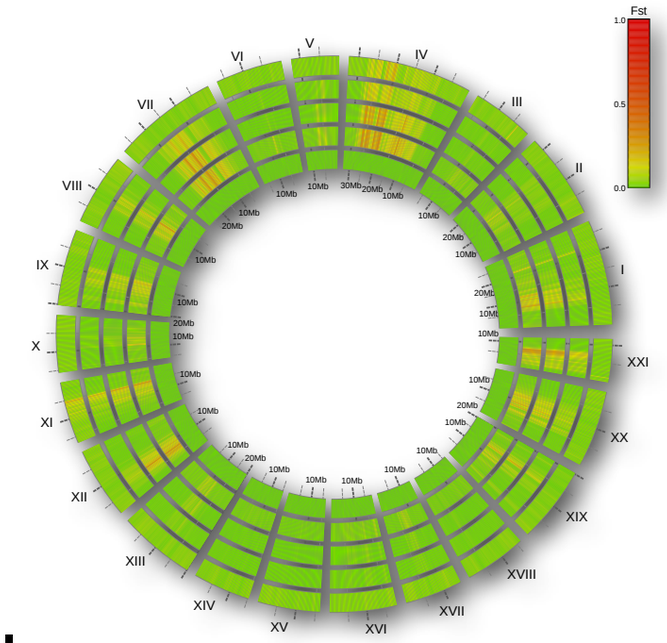


Ecological & evolutionary genomic analyses using RAD-seq

2015 Workshop on Genomics
Český Krumlov

Bill Cresko
Institute of Ecology and Evolution
Department of Biology
University of Oregon

Outline for today's lecture

RAD-seq for evolutionary genomic analyses

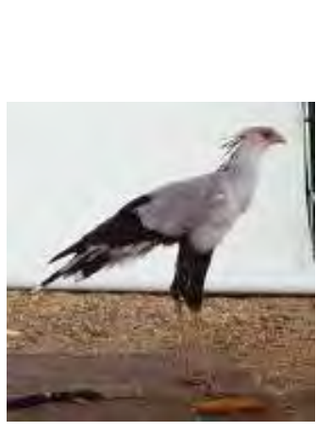
Evolutionary genomcis of stickleback fish

RAD-seq experimental and statistical considerations

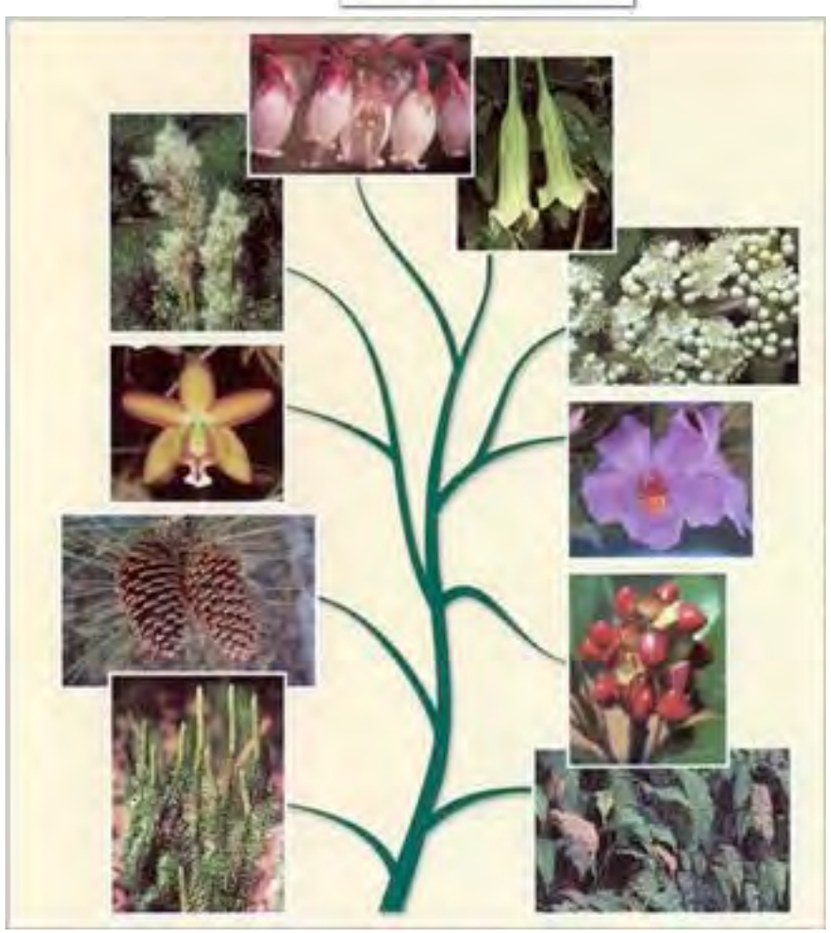
Genomically enabling pipefish

Stacks *software pipeline (this afternoon & evening)*

Why do species look the way that they do?

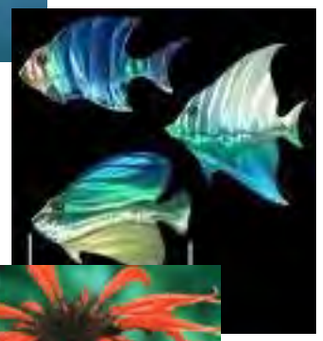








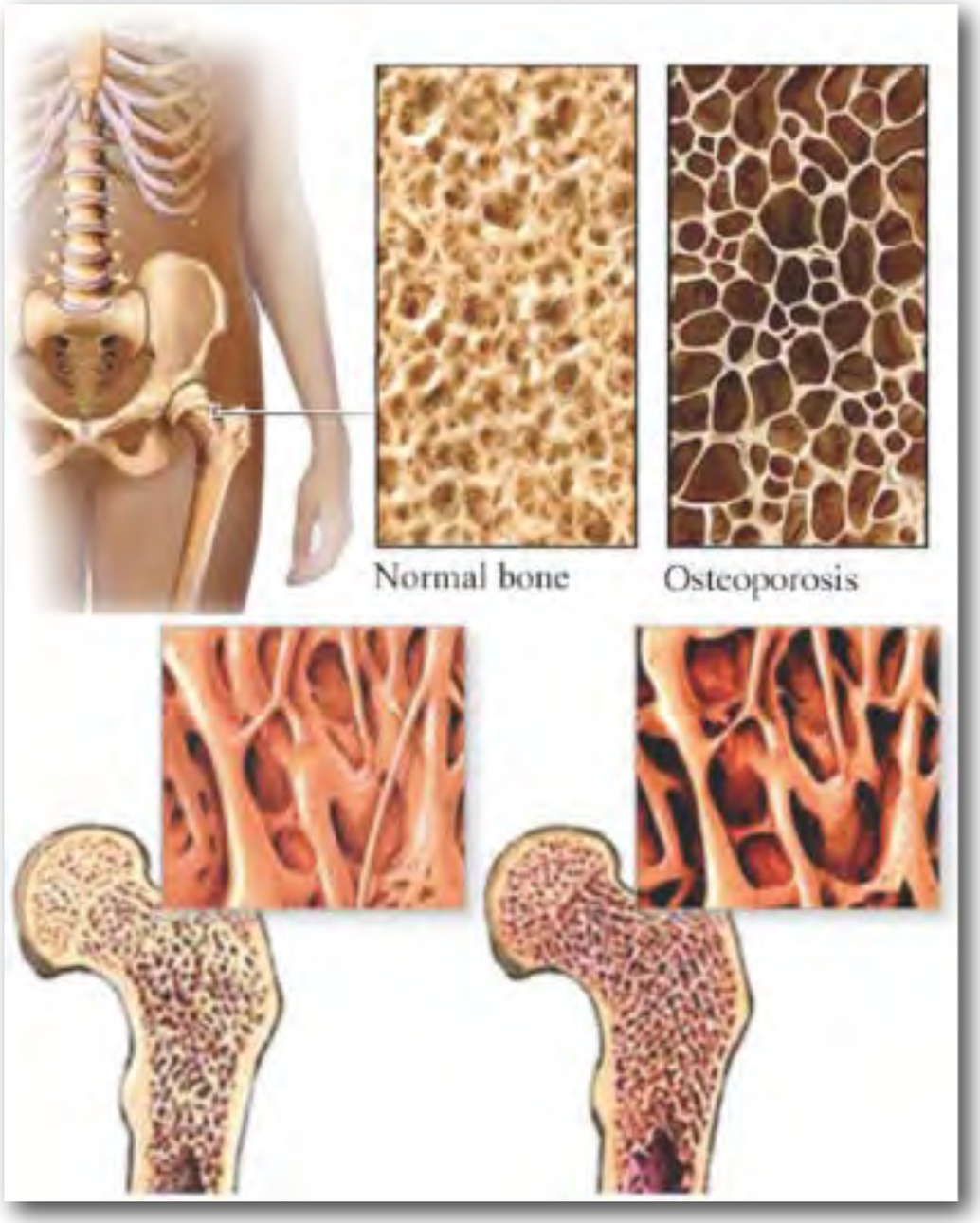




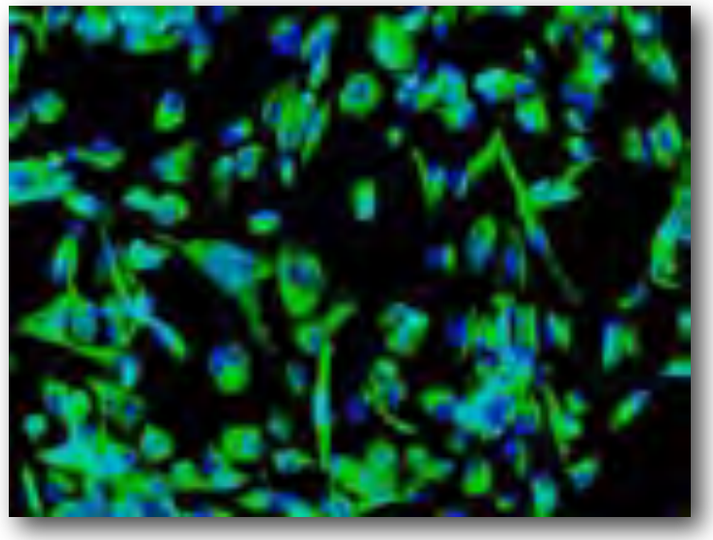


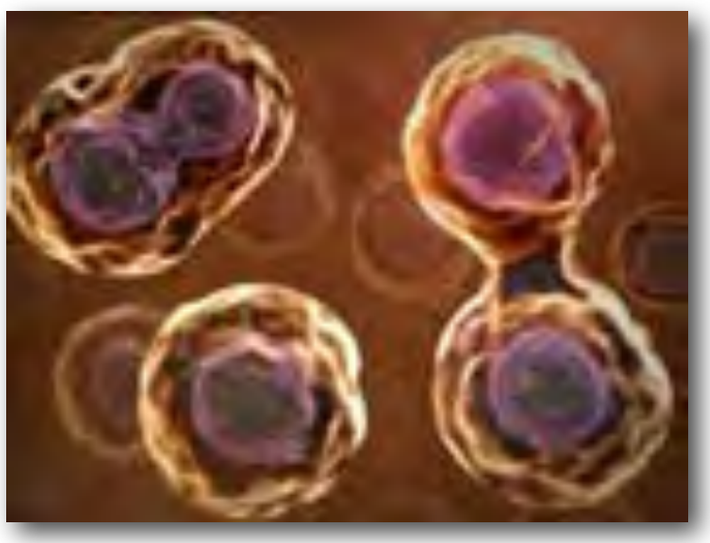


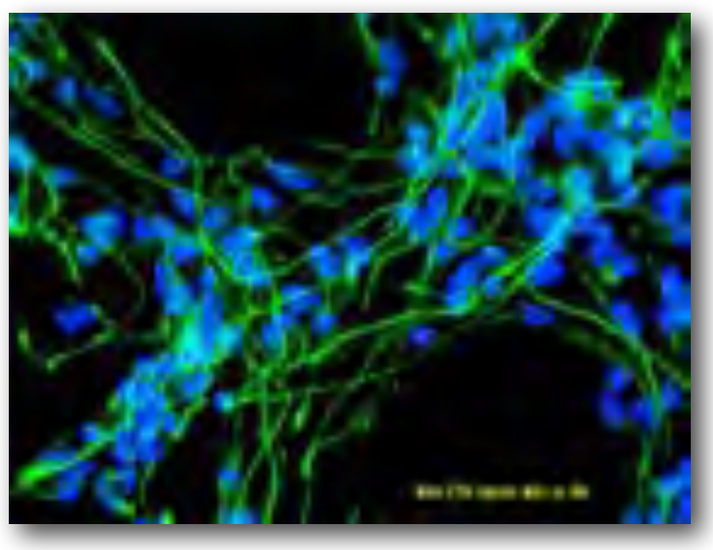

Why do organisms vary?

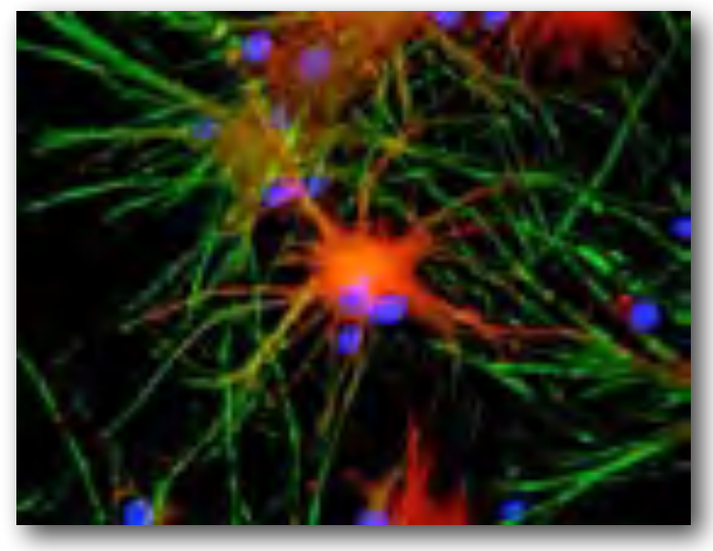


How is cellular functional diversity created?

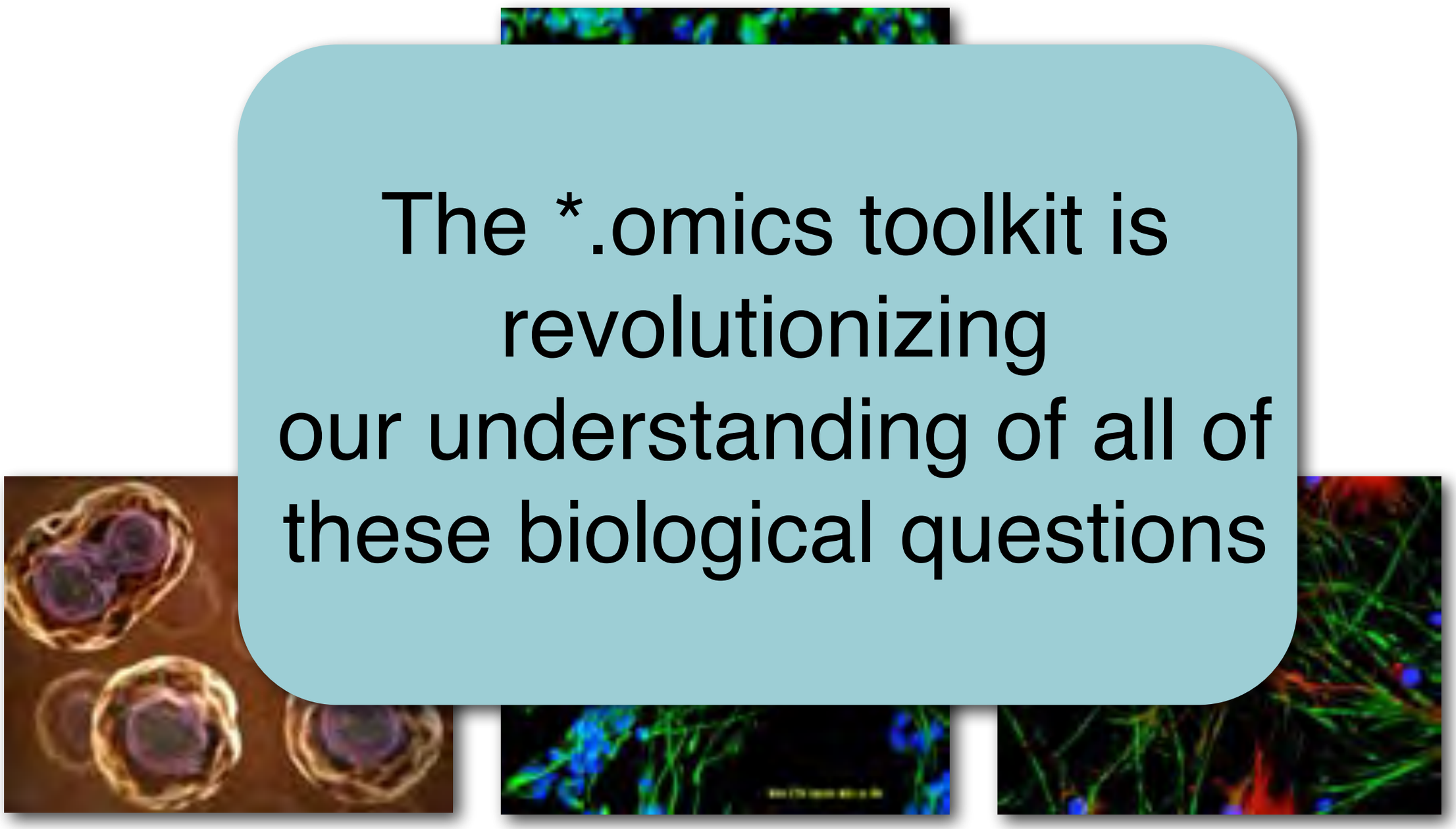








How is cellular functional diversity created?

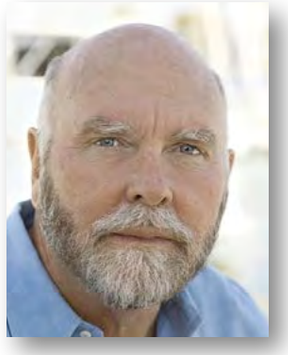


The *.omics toolkit is revolutionizing our understanding of all of these biological questions

(1998)

(2000)

(2002)



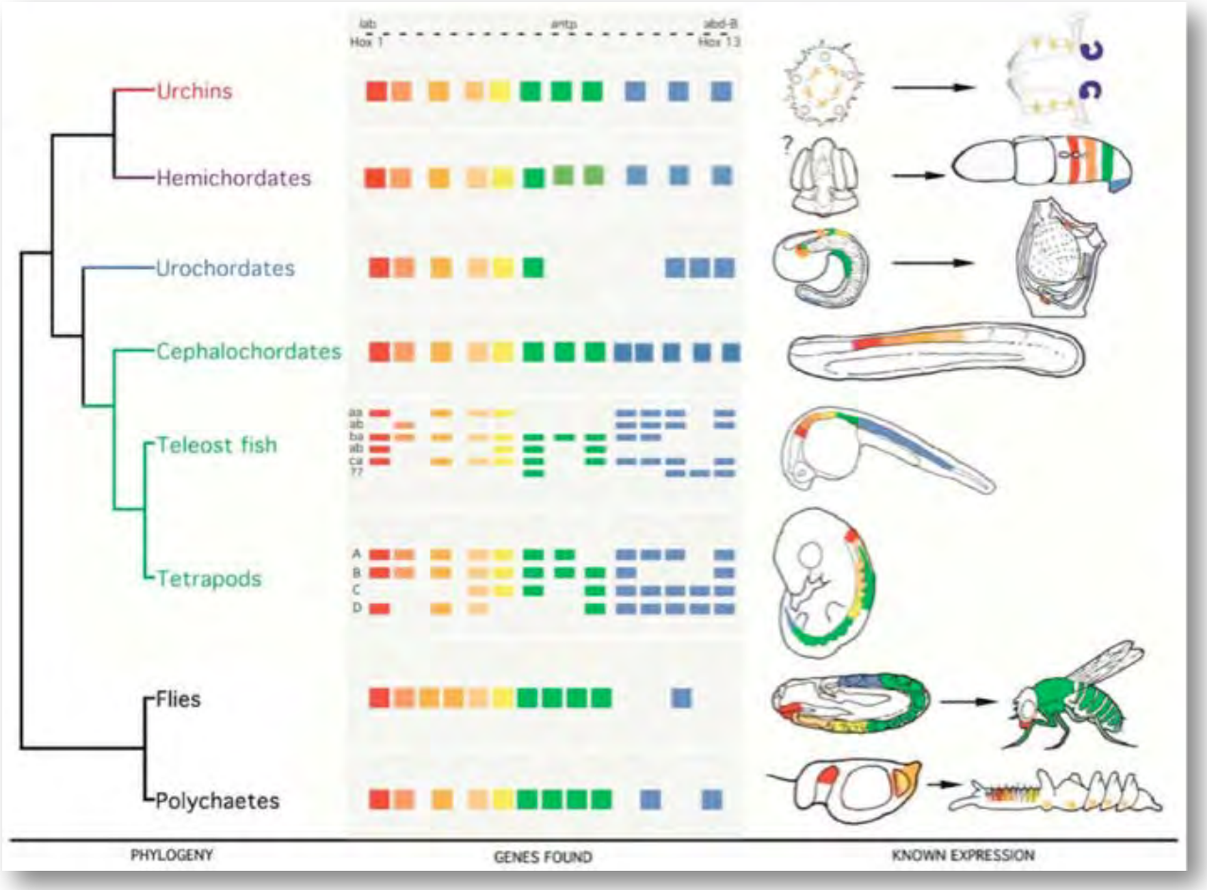
(2006)

(2006)

(2002)

(2007)

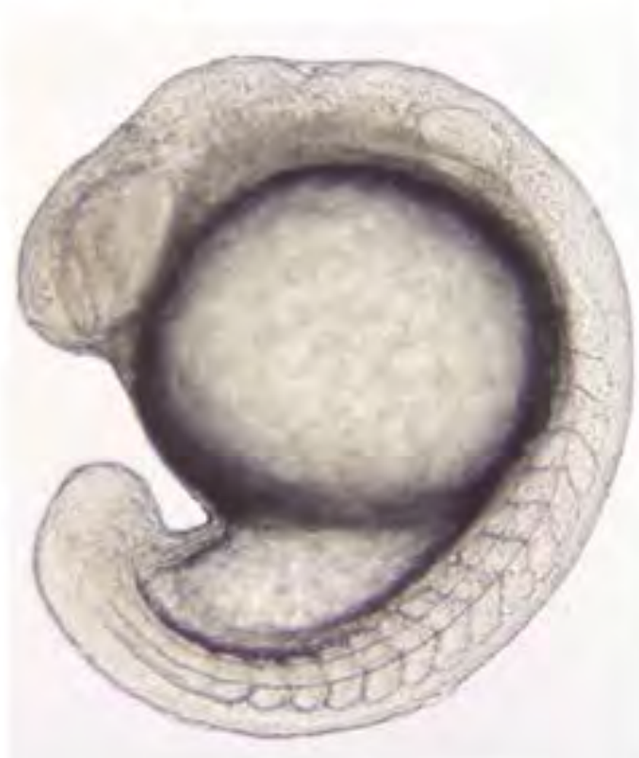
(2006)



Comparative Genomics

Vertebrate **zygotes** or embryos

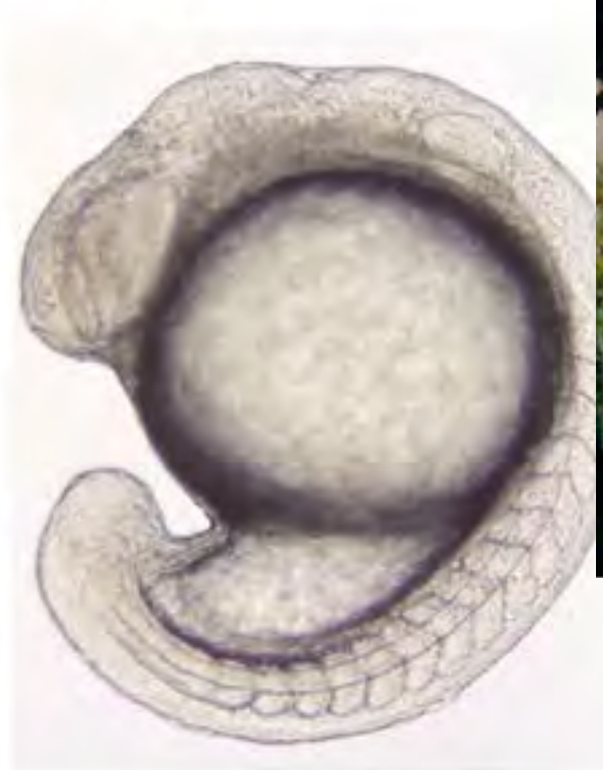
28 day human



19h zebrafish

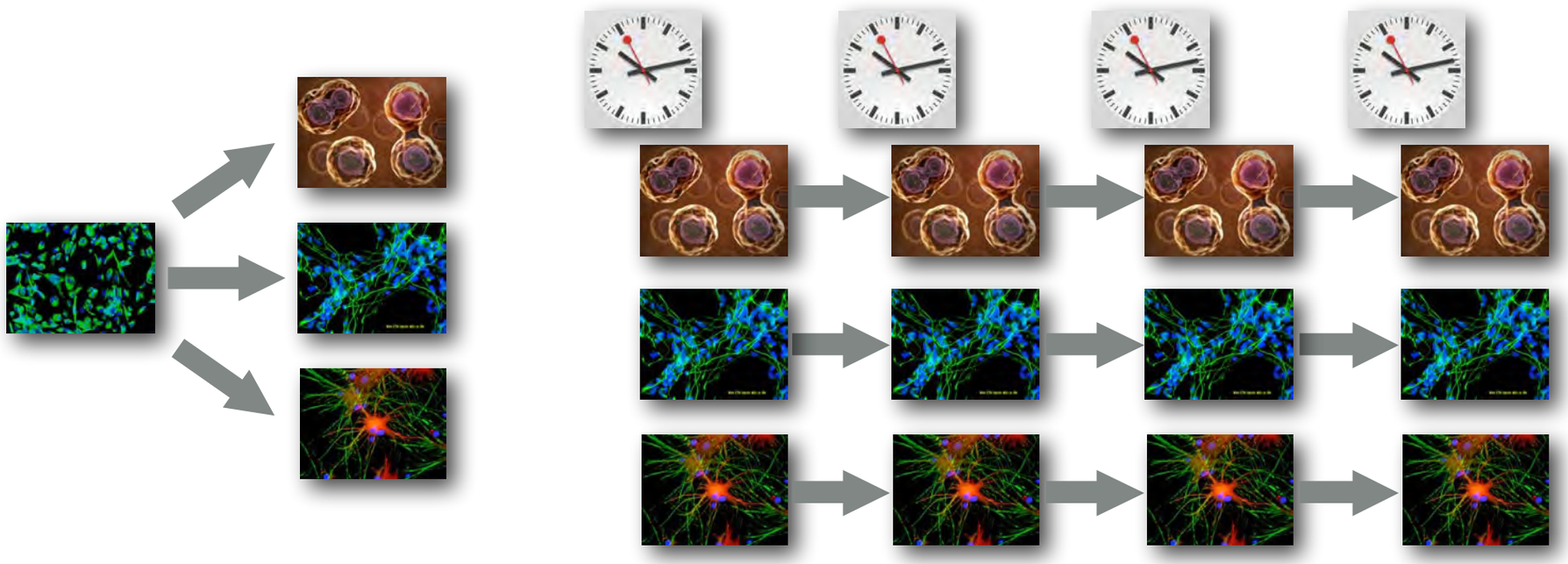
Vertebrate **zygotes** or embryos

28 day human



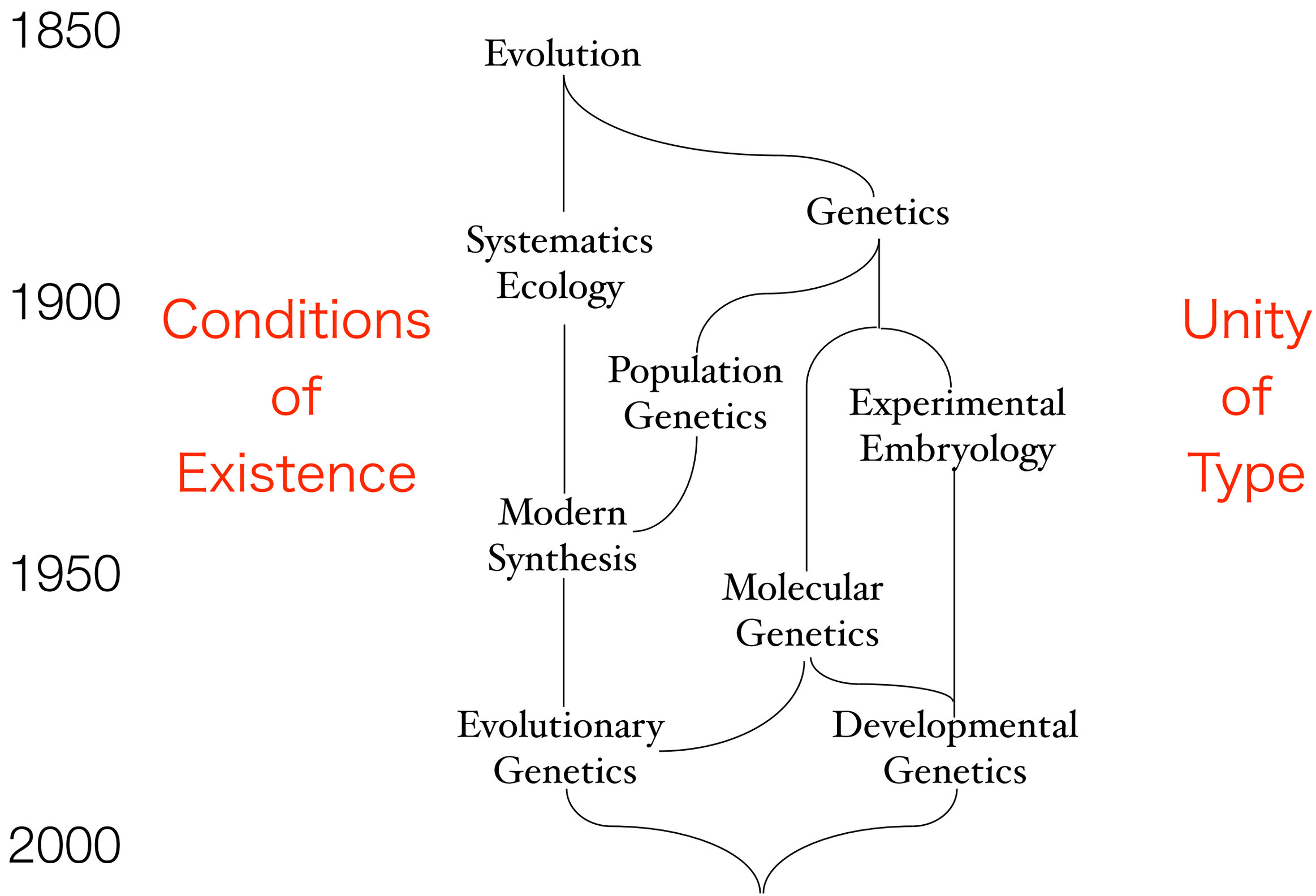
19h zebrafish

Dr. Catchen in his ‘following Phish Phase’



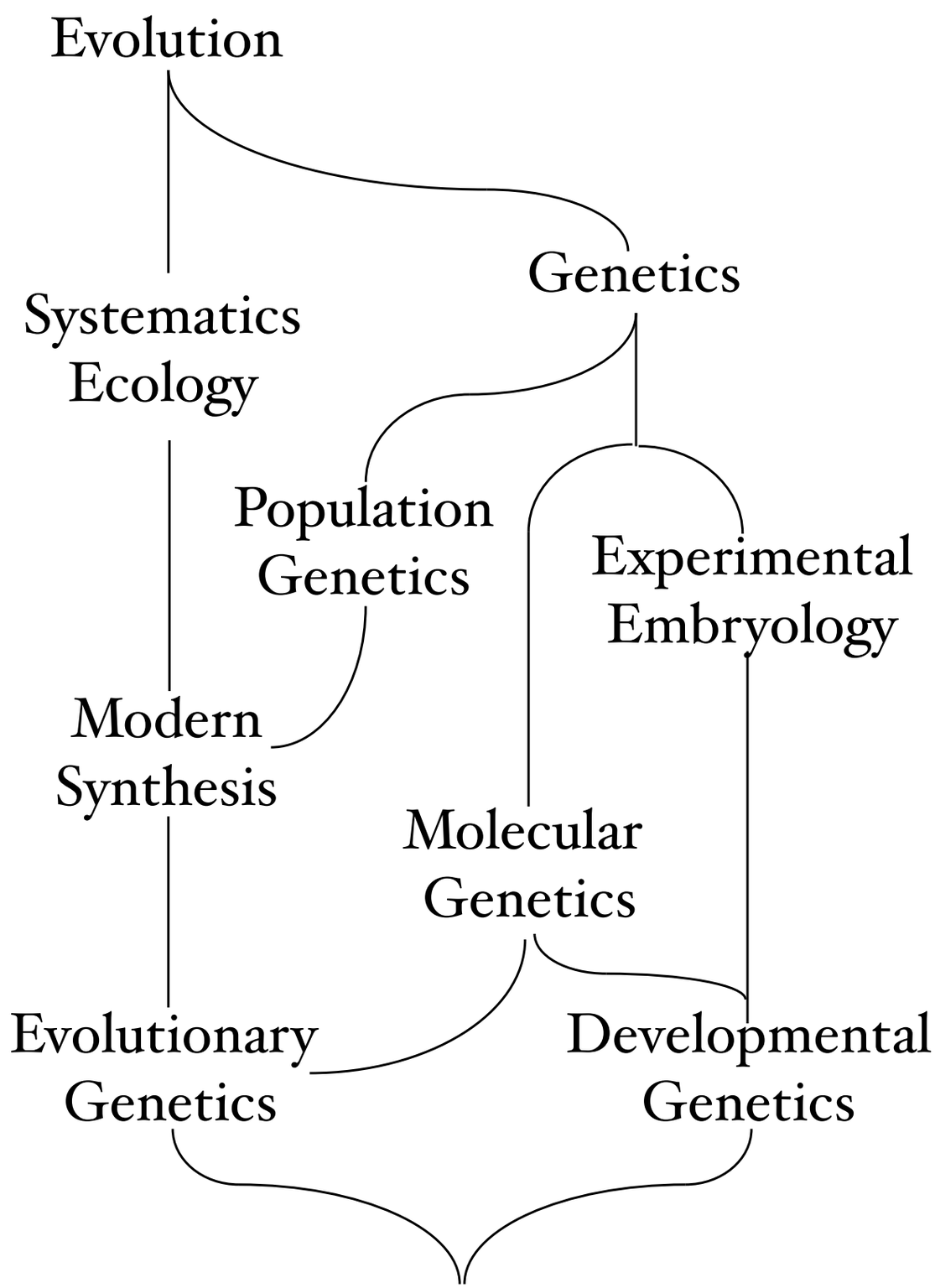
Functional Genomics

Population Genomics



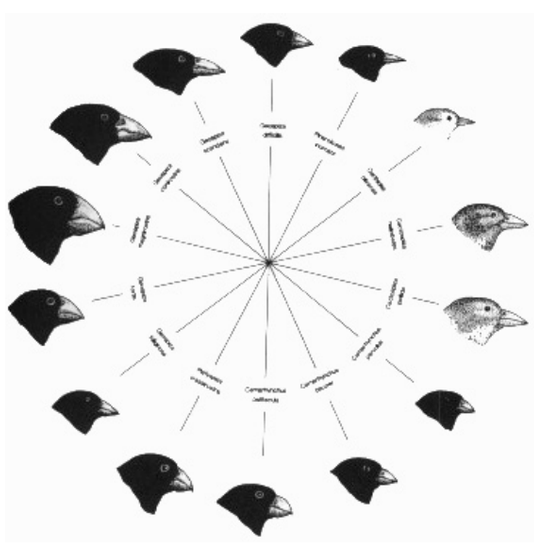
Modified from Gilbert 1998

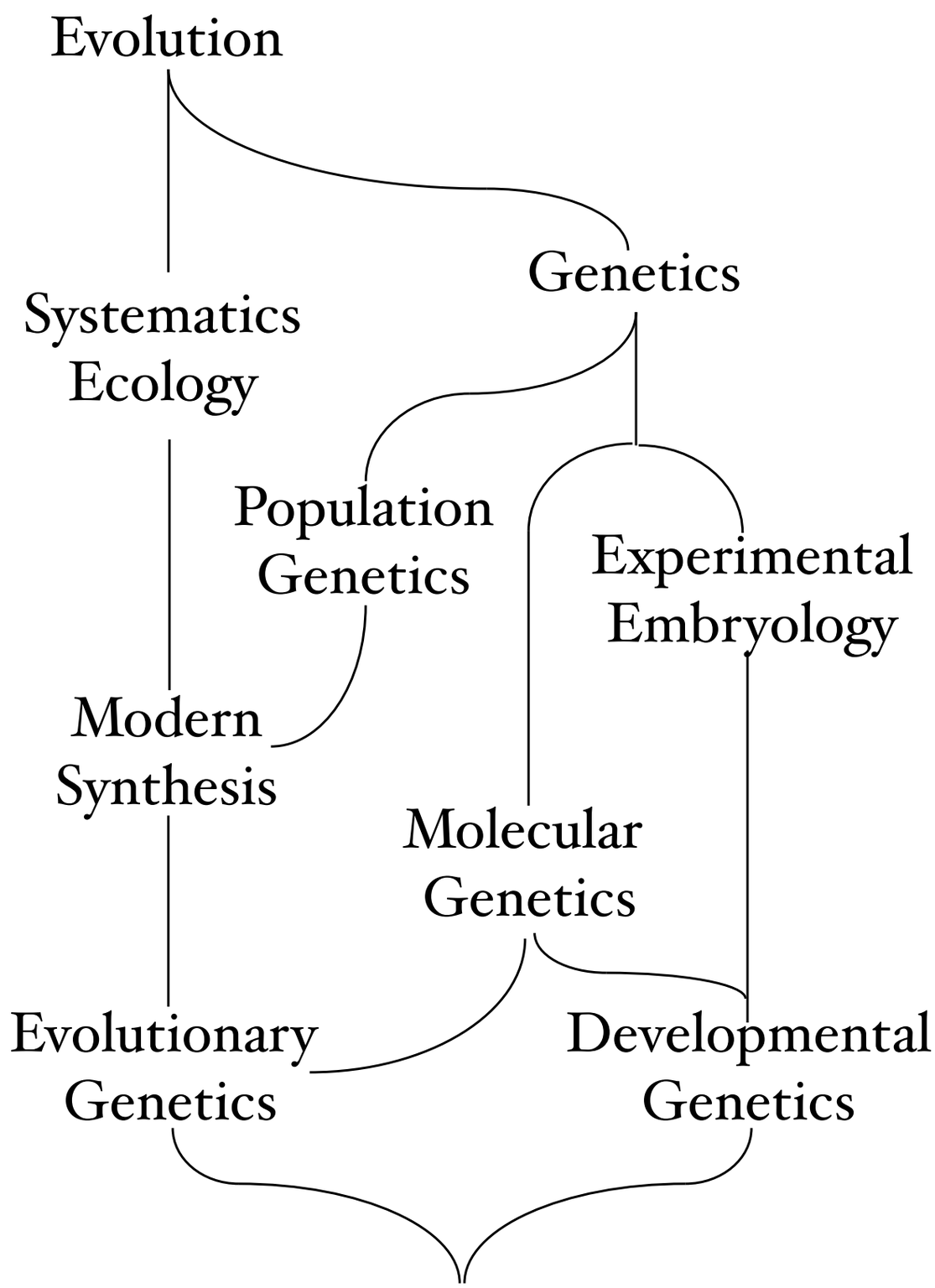
Conditions of Existence



Unity of Type

Conditions of Existence

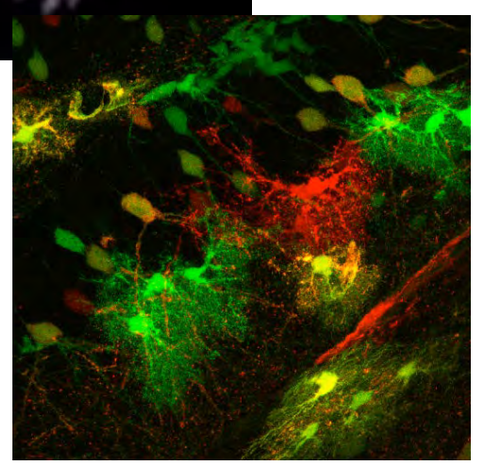




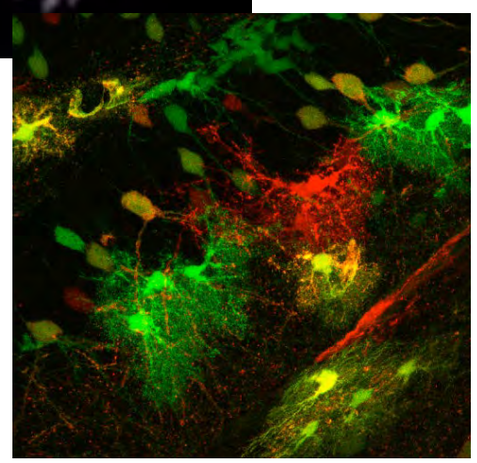
Unity of Type

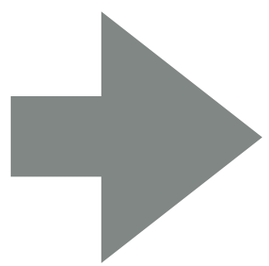
functional evolutionary genomics

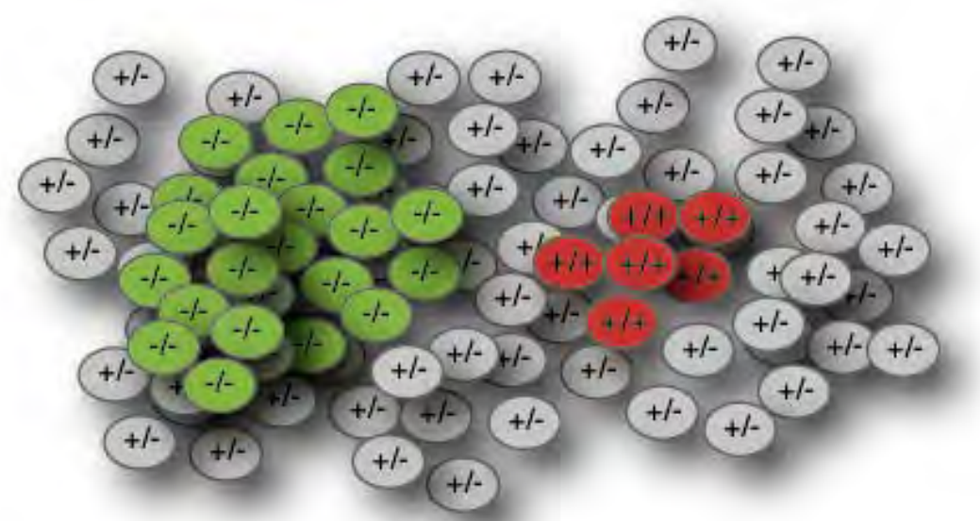
Studying cancer as an evolutionary process



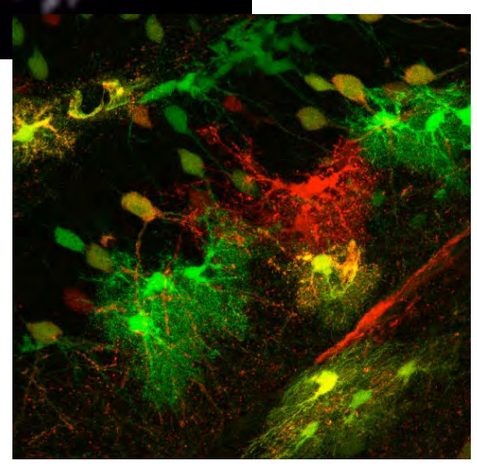
Studying cancer as an evolutionary process

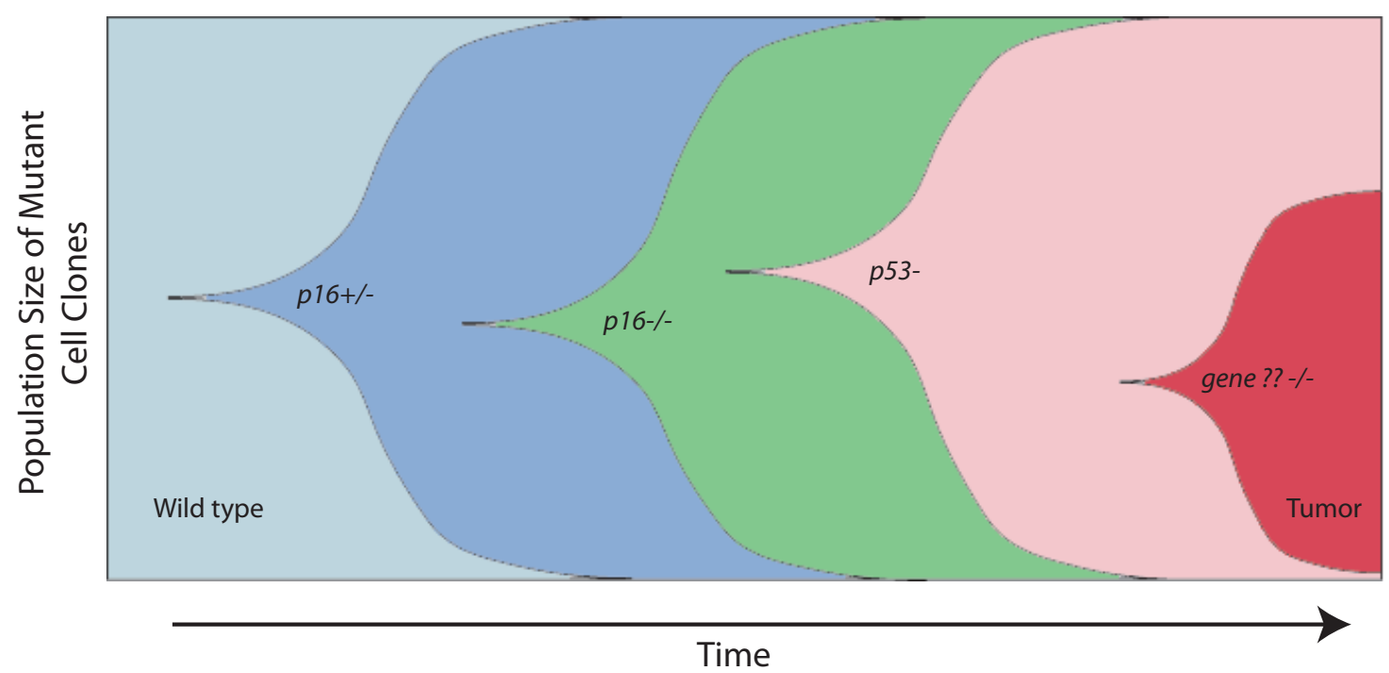




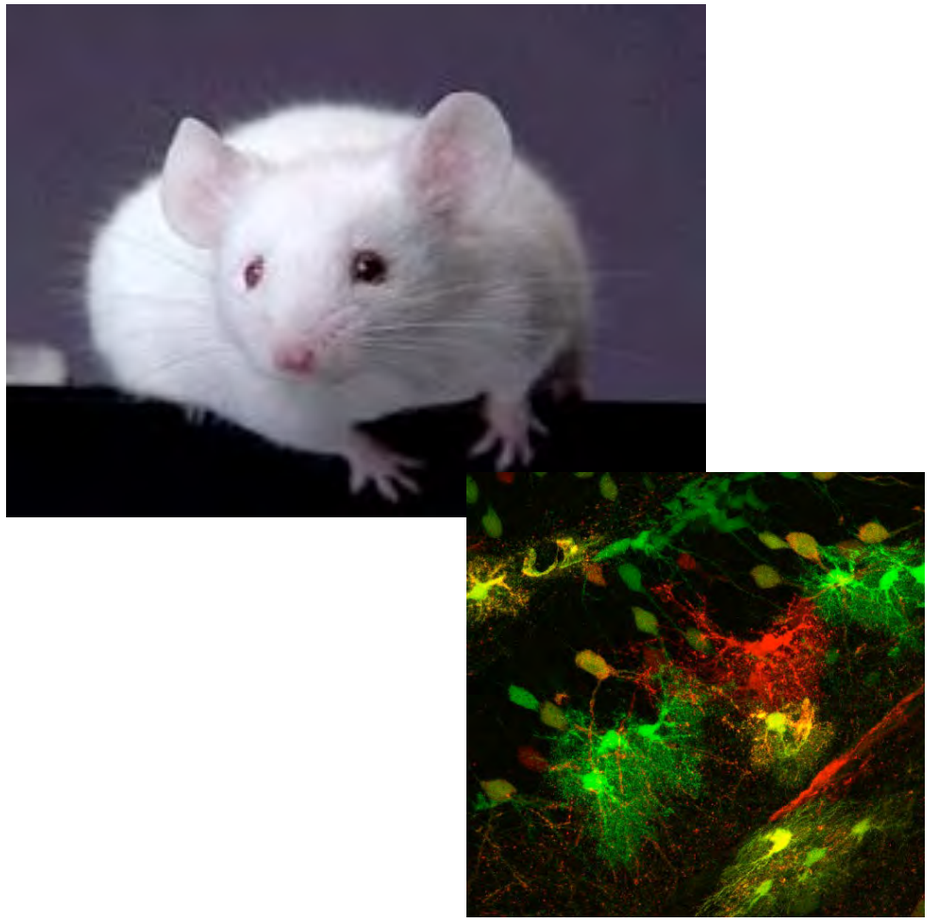


Studying cancer as an evolutionary process



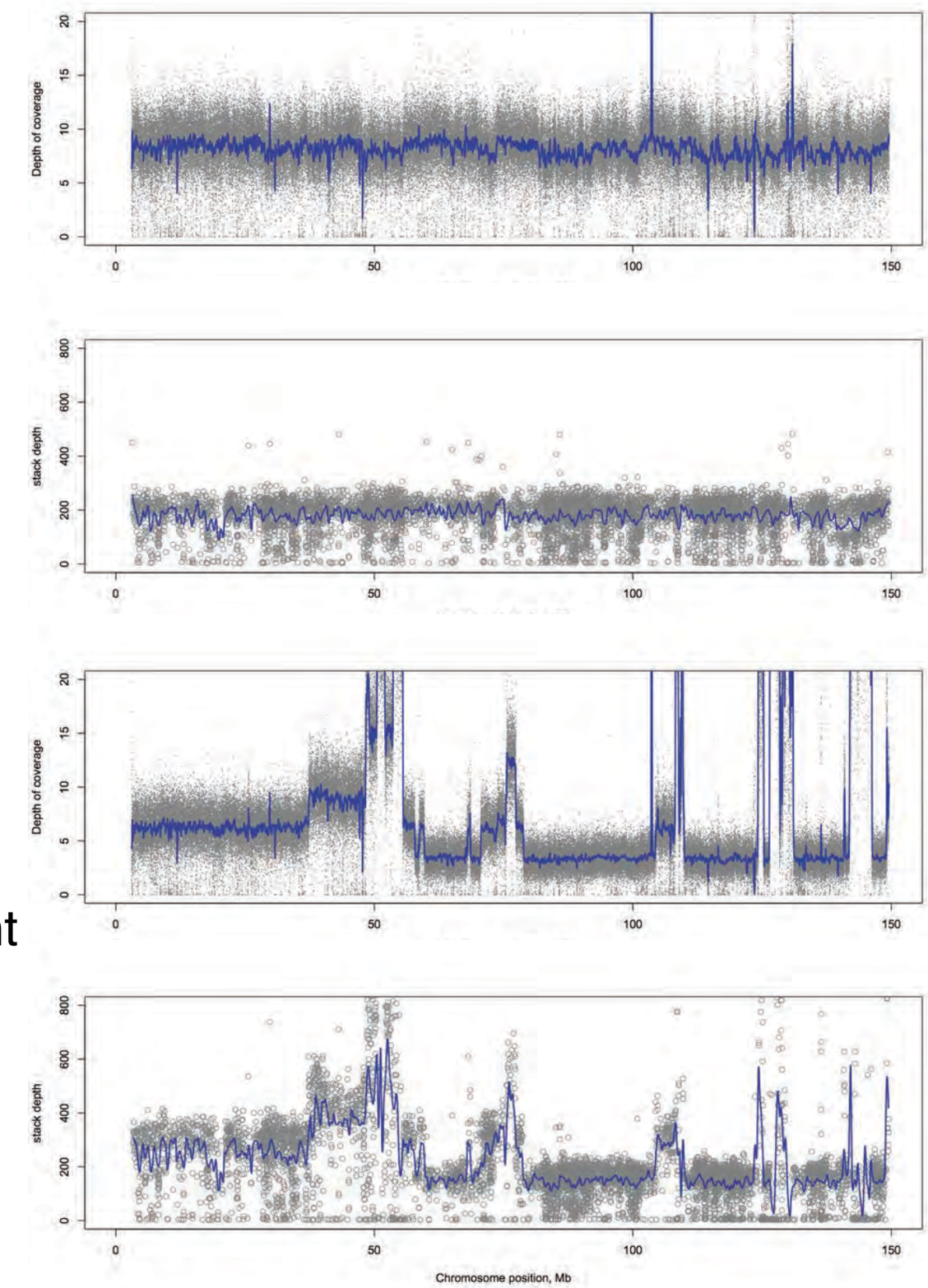


Studying cancer as an evolutionary process



Wild Type

Mutant



How do organisms adapt to novel environments?

from Grant and Grant. 2007. How and why species multiply: The radiation of Darwin's finches. Princeton University Press

How do organisms adapt to novel environments?

How is genetic diversity partitioned across individuals, populations and species?

What genomic regions are important for adaptation to novel environments?

How does genome architecture influence rapid evolution?

Where does the basis for evolutionary novelties reside in genomes?

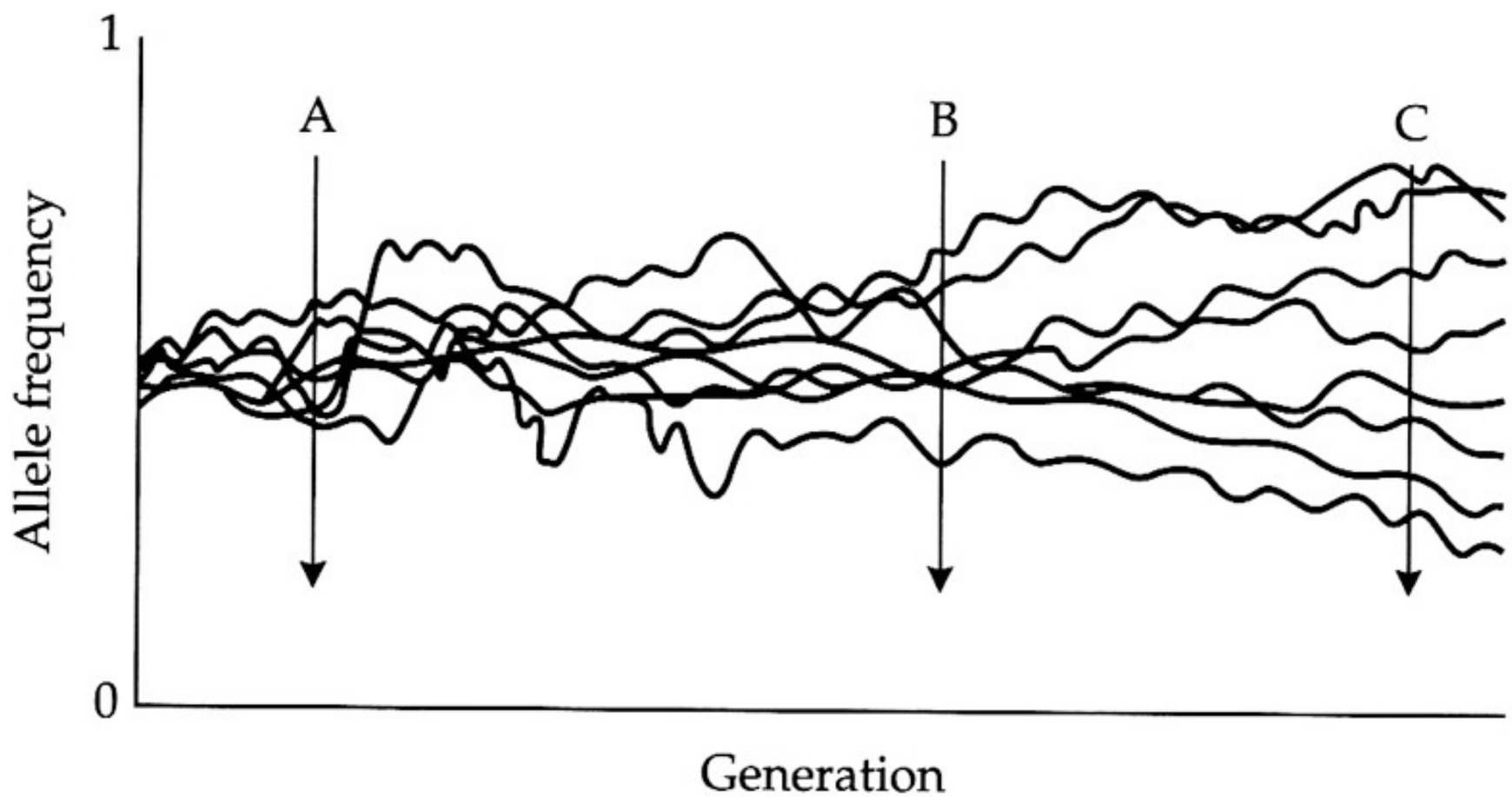
from Grant and Grant. 2007. How and why species multiply: The radiation of Darwin's finches. Princeton University Press

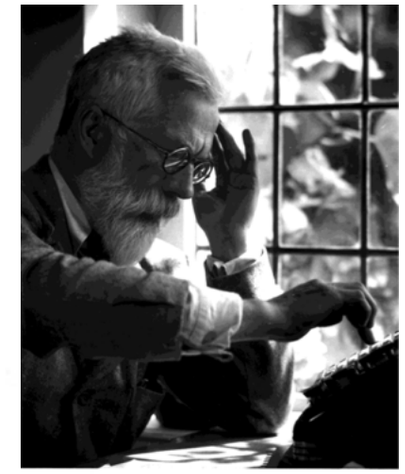
Four fundamental processes in evolution

Origin of genetic variation;
mutation
migration

Sorting of variation;
genetic drift
natural selection

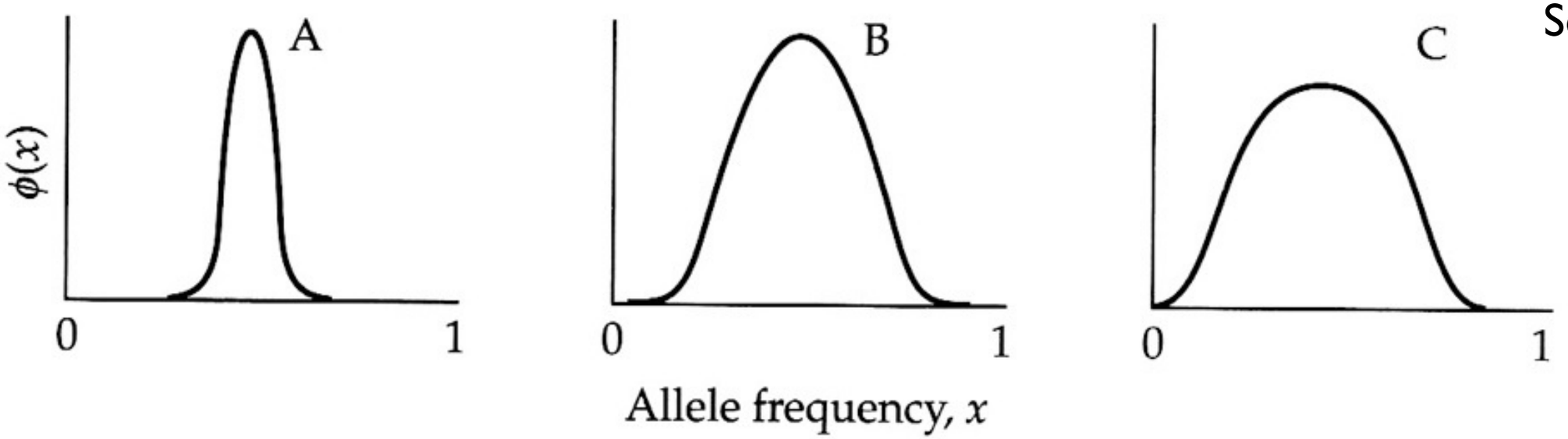
Genetic drift is a null model





R. A. Fisher

Sewall Wright



Population genomics

Simultaneous genotyping of **neutral** and **adaptive** loci

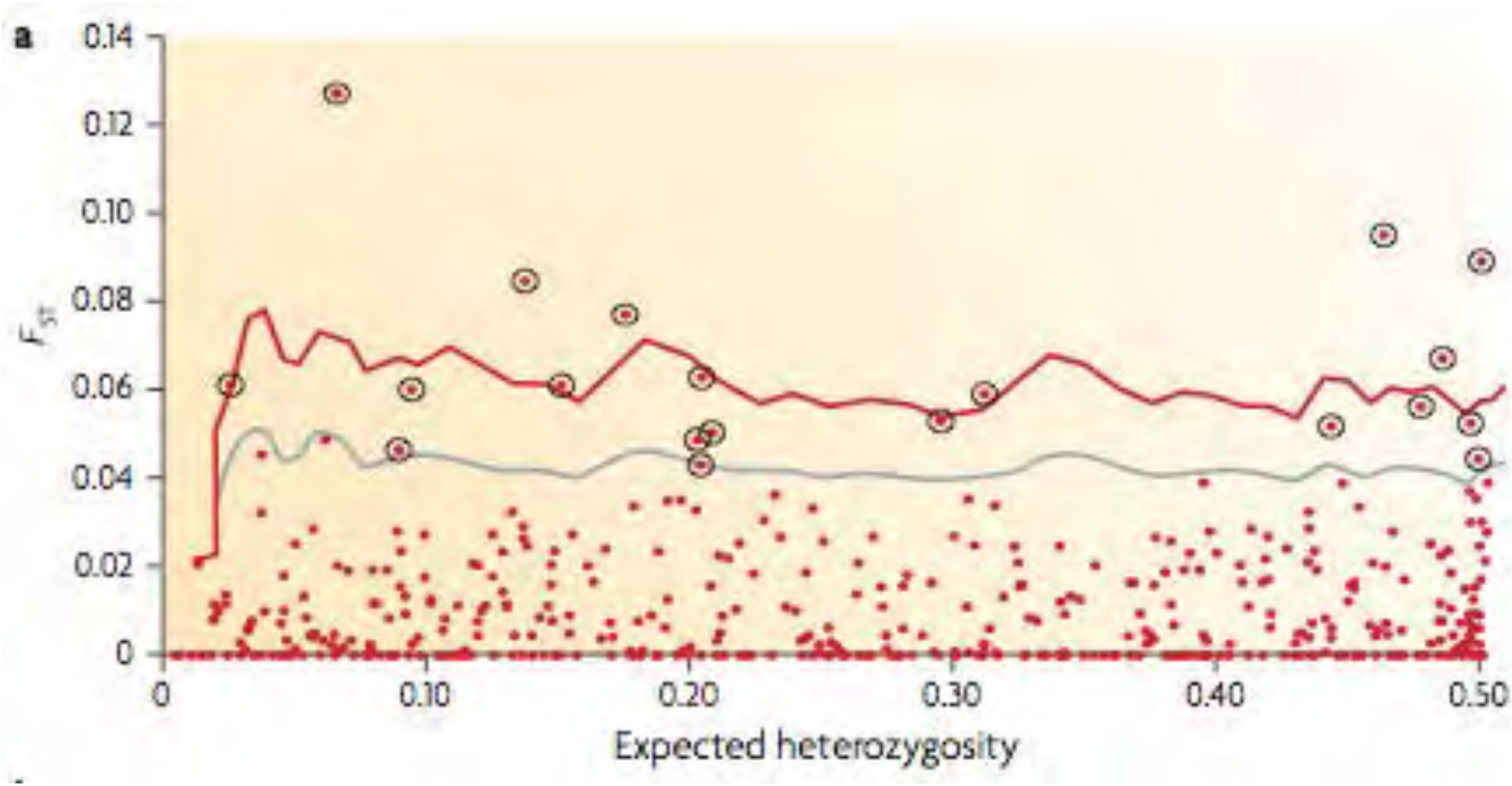
Genome-wide background provides more precise estimates:

- Demographic processes (e.g. N_e)
- Phylogeography

Outliers from background indicate:

- Selective sweeps
- Local adaptation

Population genomics of unordered markers

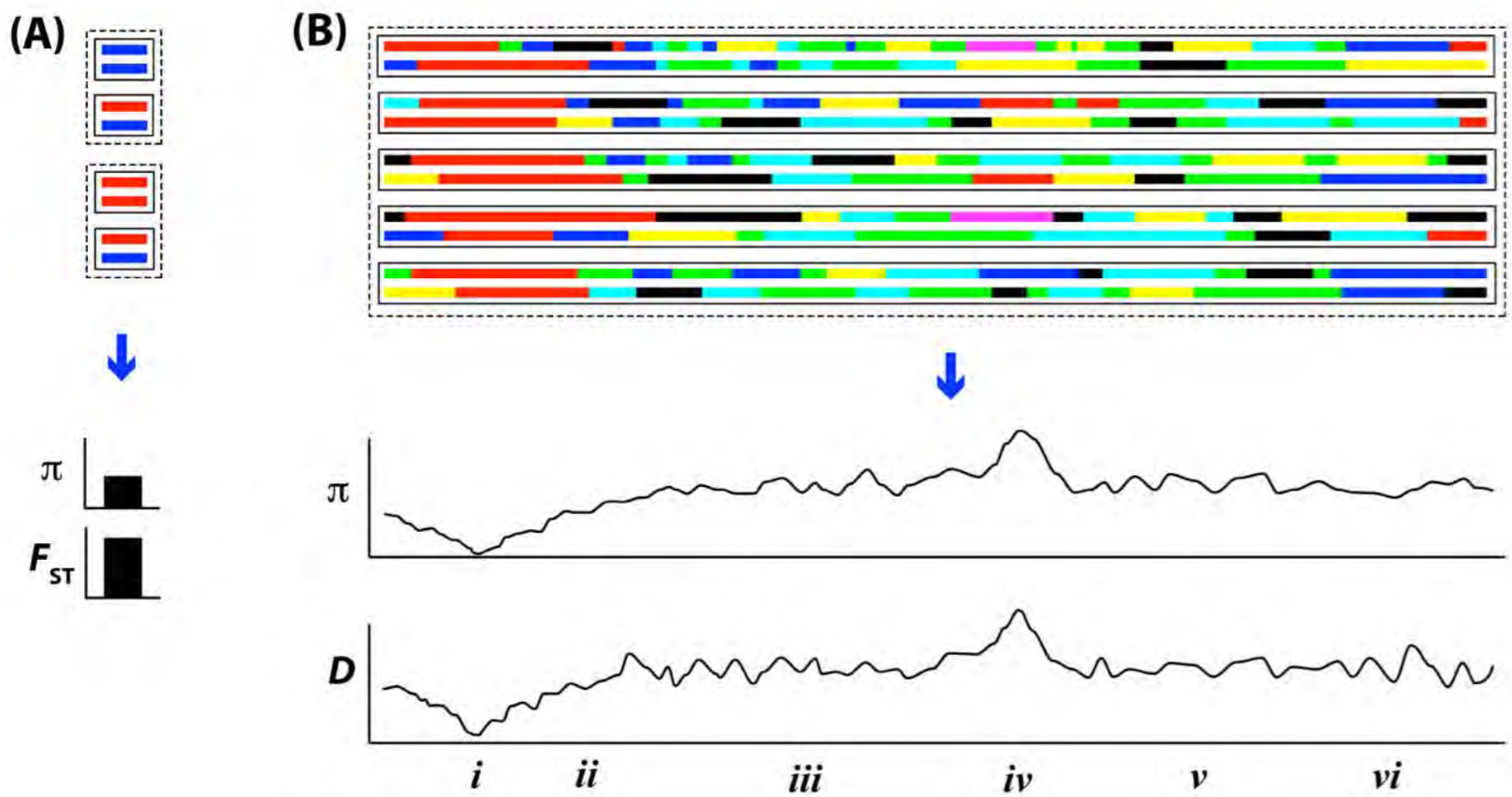


Namroud et al (2008) *Mol Ecol* 17:3599

Population genomics of ordered markers

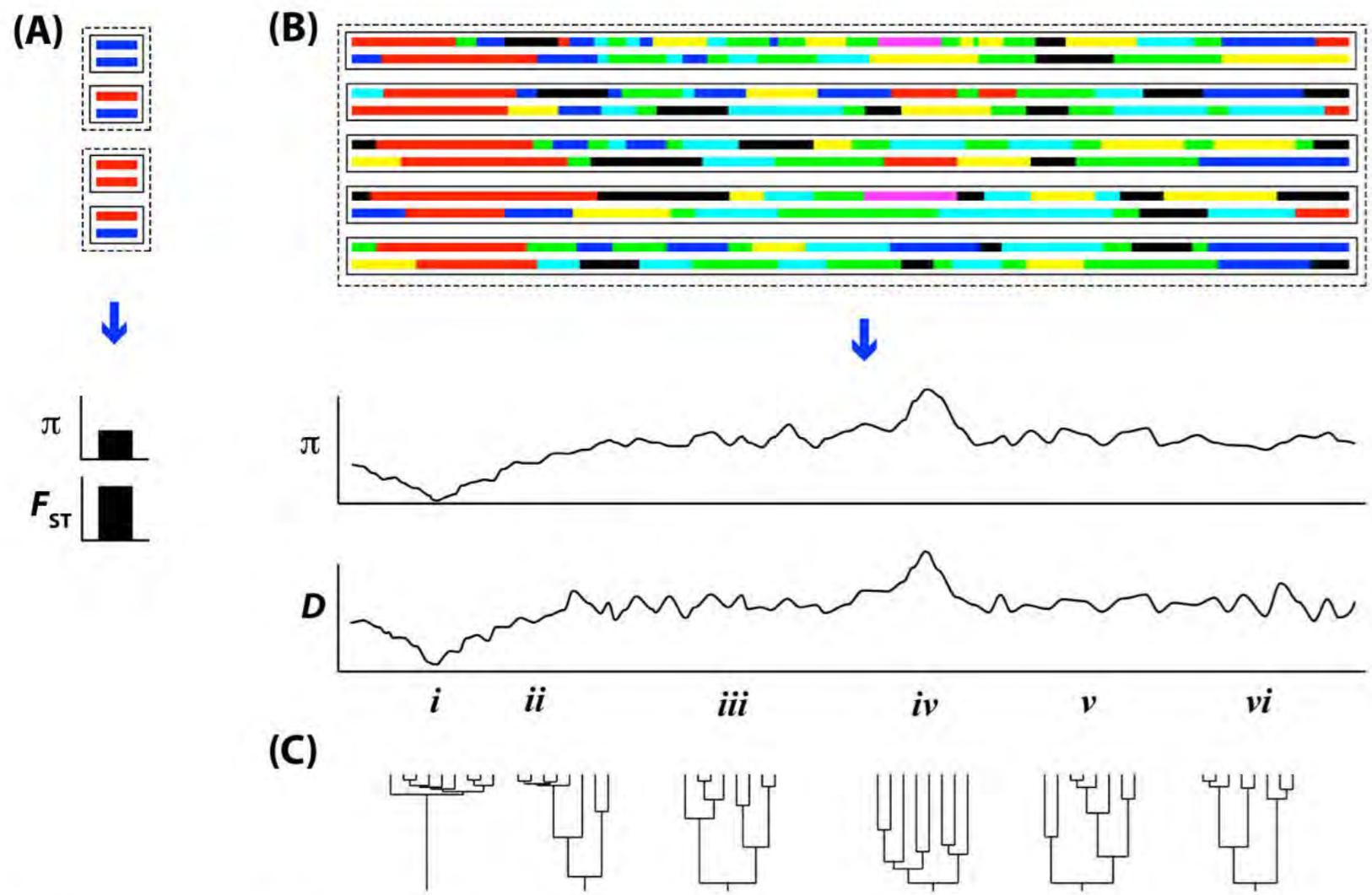
Hohenlohe et al. 2010. Int J Plant Sciences

Population genomics of ordered markers



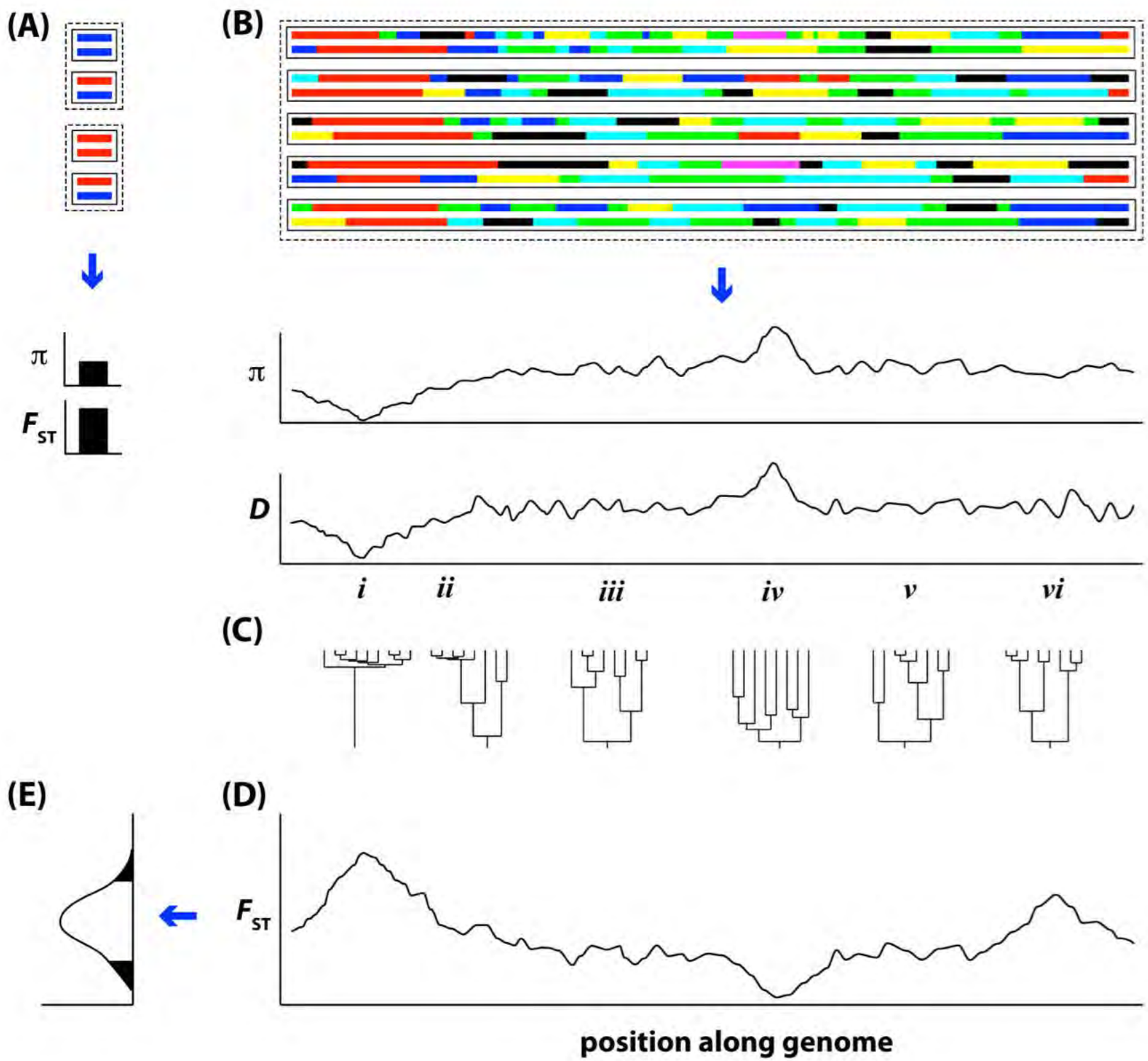
Hohenlohe et al. 2010. Int J Plant Sciences

Population genomics of ordered markers



Hohenlohe et al. 2010. Int J Plant Sciences

Population genomics of ordered markers

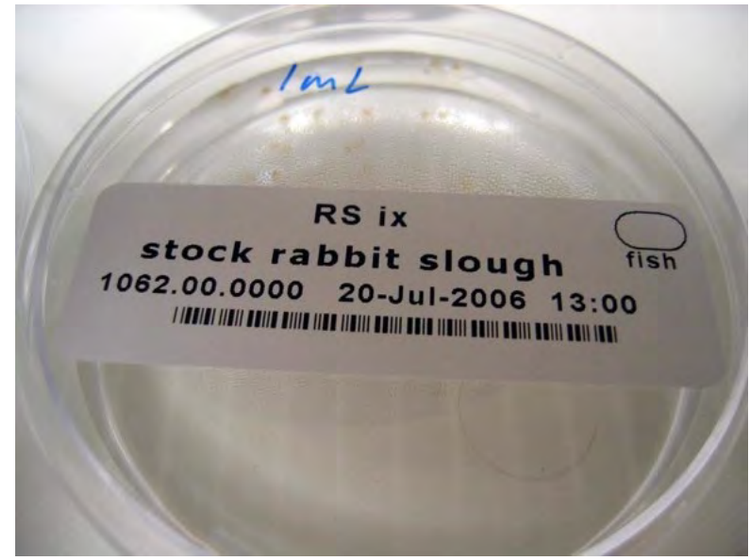


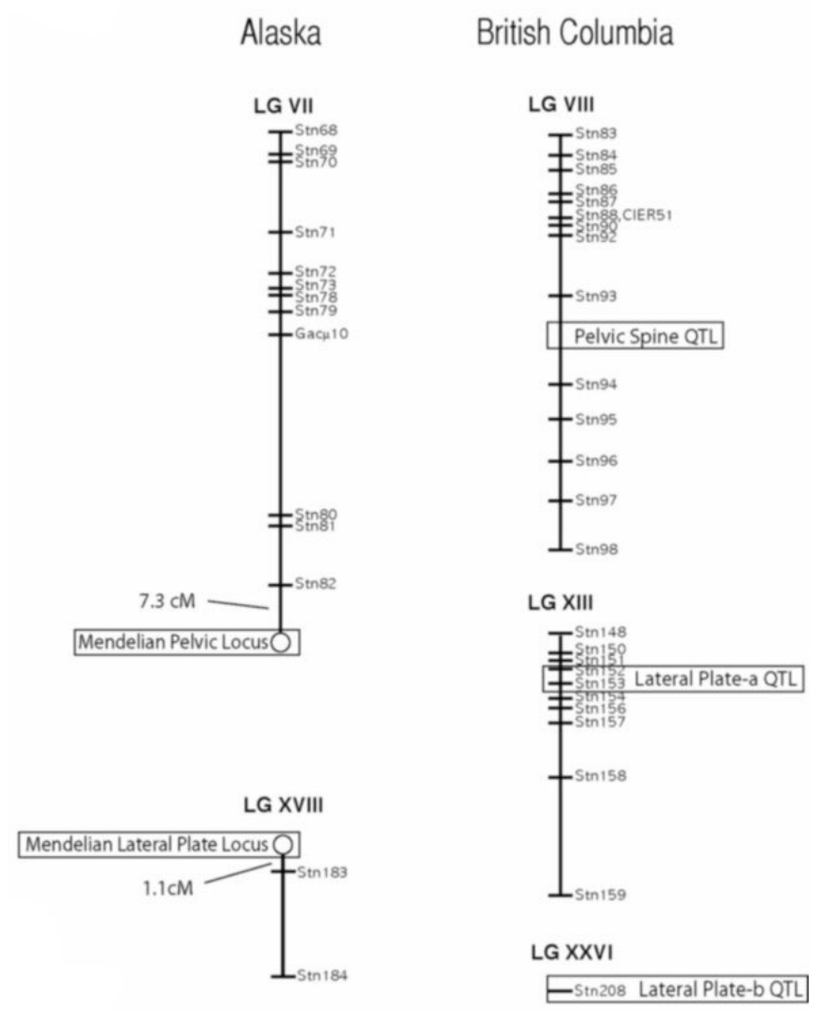
Hohenlohe et al. 2010. Int J Plant Sciences

How do we ‘genomically enable’ research studies of non-model organisms?

1. Genetic Markers & Genetic Maps
2. Physical Maps (genomes)
3. Transcriptomes
4. Gene Expression Analyses
5. Epigenetic analyses

In the field and in the lab until a few years ago....





Shouldn’t we just sequence everything?

(*note* - the answer to this question may be ‘yes’ soon, and if so I will stop at this slide. But until then….)

Why not sequence the entire genome??

- Still prohibitively expensive for many studies
    - Human height GWAS; over 15,000 individuals assayed
    - Identified many new regions contributing to the variation
    - Still only identified a fraction of the heritability

- For many studies a full sequence isn't necessary
    - genomes of many organisms are organized in linkage blocks
    - well spaced markers will provide the necessary coverage

- Genetic maps are very useful in genomic studies
    - a high density genetic map can facilitate genome assembly
    - genomes may be segregating a lot of structural variation

Alternative approach - Reduced representation NGS for genotyping

- Focus sequencing on homologous regions across the genome
- Simultaneous identification and typing of single nucleotide polymorphisms (SNPs)
- The cost will be a fraction of the cost of resequencing the genome
  - i.e. 1% genome coverage will be less than 1% the cost
  - often coverage is more even than whole genome sequencing
- Thousands of genomes to be assayed in just a few weeks
- WHY NOT - complete genomic sequence is often useful
  - when linkage disequilibrium blocks (LD) are very short
  - Inferring patterns of LD may be easiest with full sequences

Different flavors of Reduced Representation Library (RRL) Sequencing for genotyping

- Common acronyms
  - **RRL** - **R**educed **R**epresentation **L**ibrary
  - **GBS** - **G**enotyping **B**y **S**equencing
  - **CRoPS** - **C**omplexity **R**eduction **o**f **P**olymorphic **S**equences
  - **RAD** - **R**estriction site **A**ssociated **D**NA
- All rely on restriction enzyme digestion
- RRL, CRoPS, and GBS use one or two restriction enzymes only
- RAD-seq uses a shearing step to more efficiently capture all restriction sites
- Incorporation of barcodes on adaptors for multiplexing
- **Aligned** against a reference genome or **assembled *de novo***
- Statistical issues
  - new level of sampling variation (sequencing in addition to biological)
  - sequencing error and problems for aligning or clustering

What is RAD-seq?

(Restriction-site Associated DNA)

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]

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

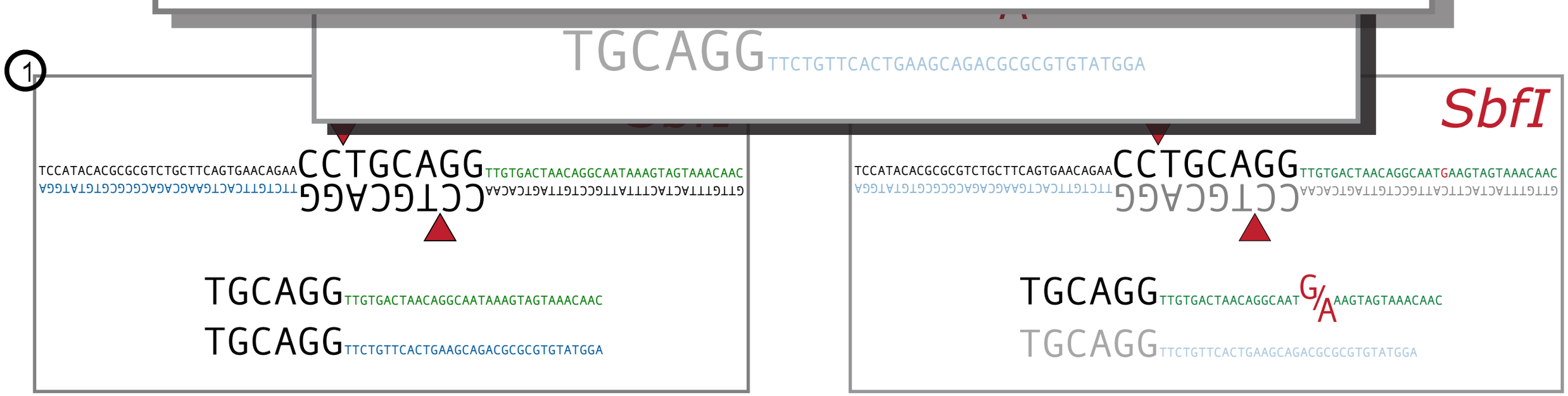
Nathan A. Baird[1¶], Paul D. Etter[1¶], Tressa S. Atwood[2], Mark C. Currey[2], Anthony L. Shiver[1], Zachary A. Lewis[1], Eric U. Selker[1], William A. Cresko[3], Eric A. Johnson[1*]

What is RAD-seq?

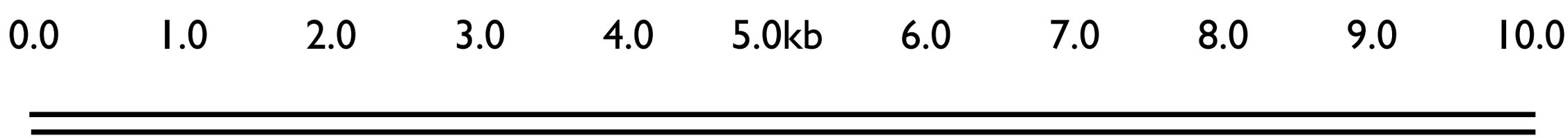
(Restriction-site Associated DNA)

22,830 *SbfI* sites in threespine stickleback
~ 45,000 RAD-Tags

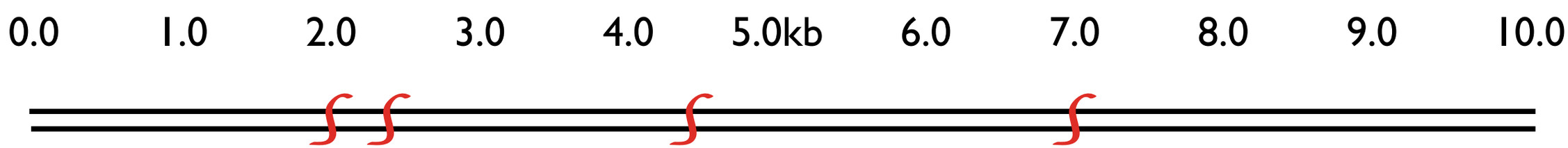
HiSeq2500 Illumina Lane:
160 million reads
HiSeq4000 Illumina Lane:
350 million reads



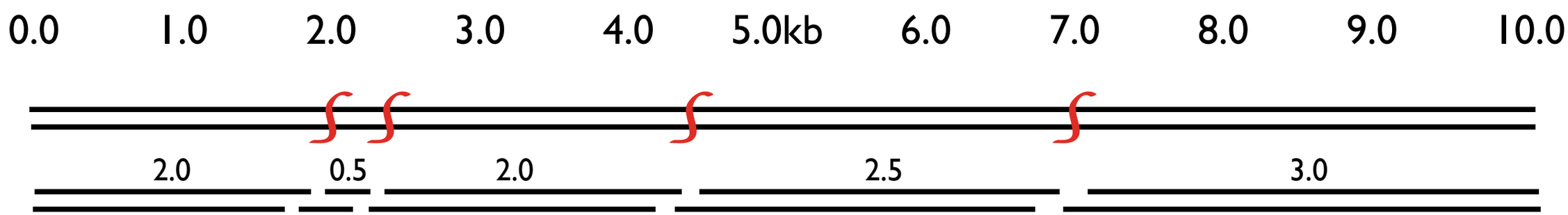
Restriction Enzyme (RE) digestion and first adaptor ligation



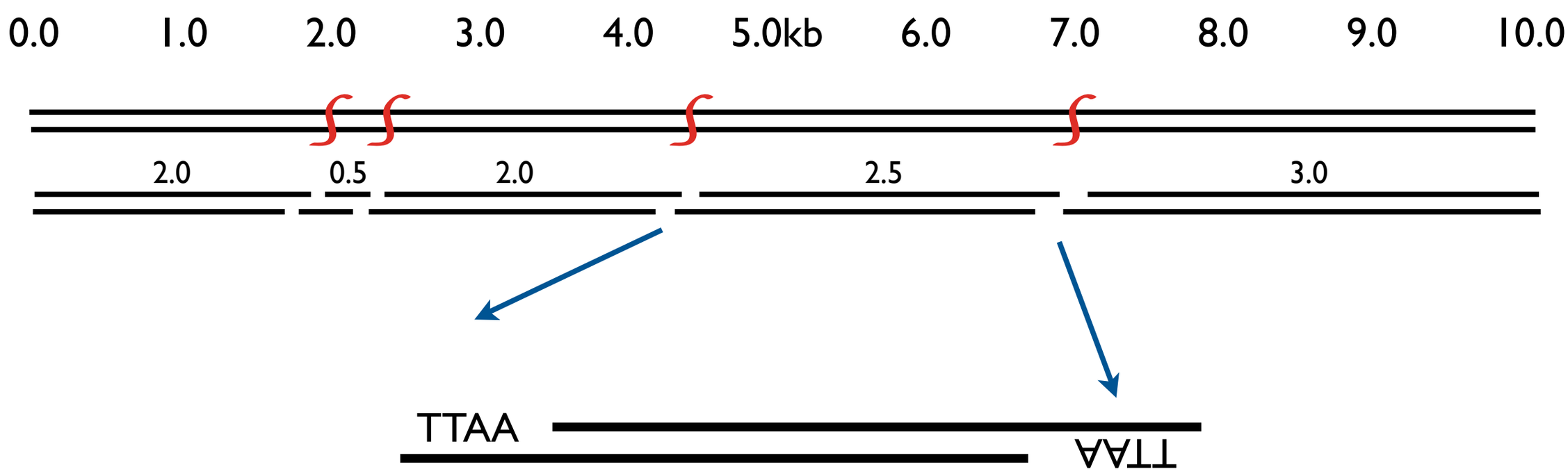
Restriction Enzyme (RE) digestion and first adaptor ligation



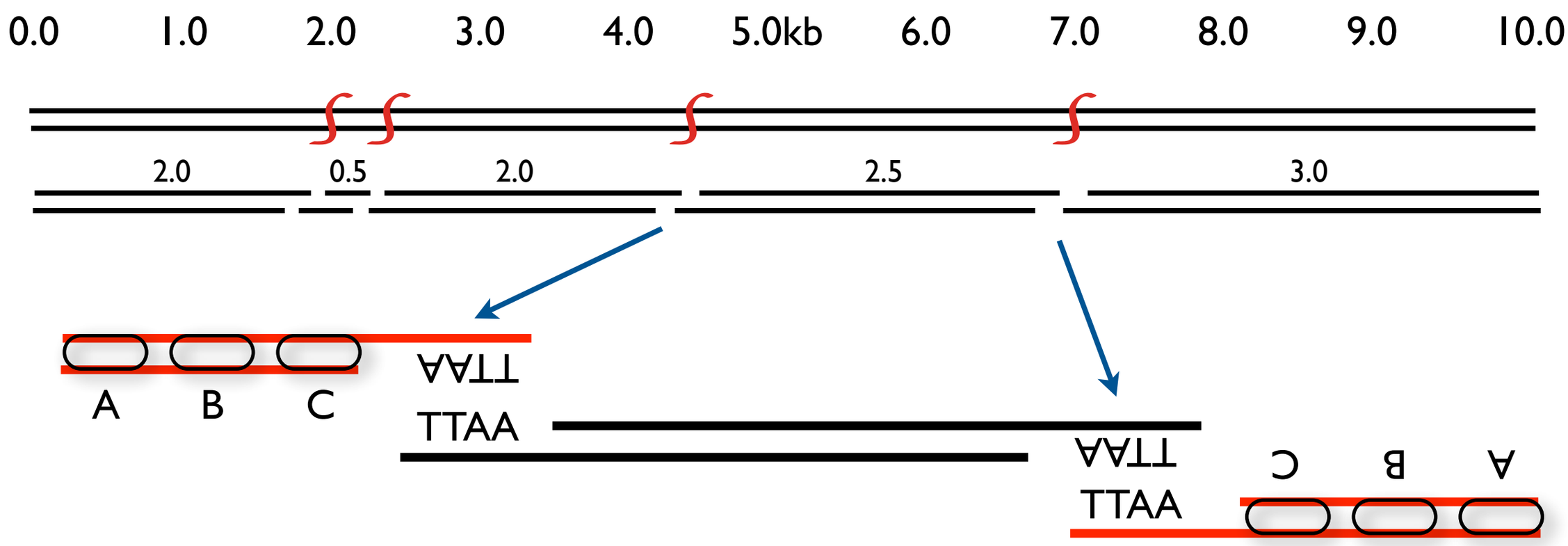
Restriction Enzyme (RE) digestion and first adaptor ligation



Restriction Enzyme (RE) digestion and first adaptor ligation

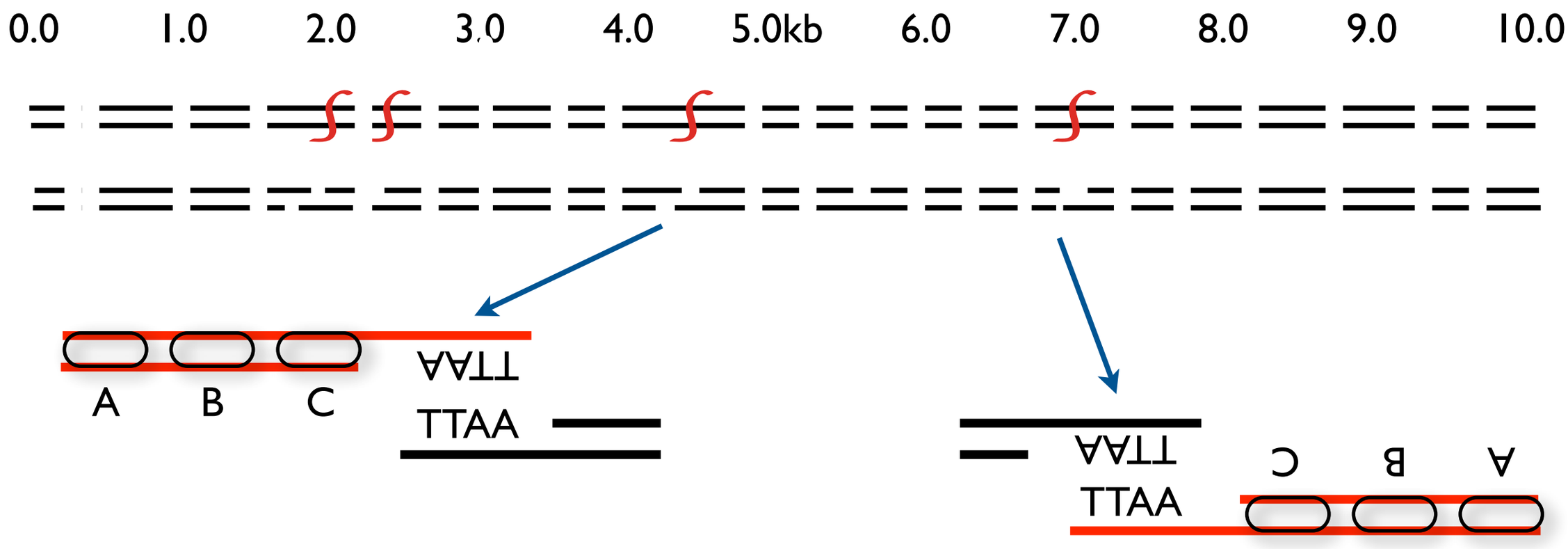


Restriction Enzyme (RE) digestion and first adaptor ligation



A = Amplification primer
B = Sequencing primer
C = Barcode

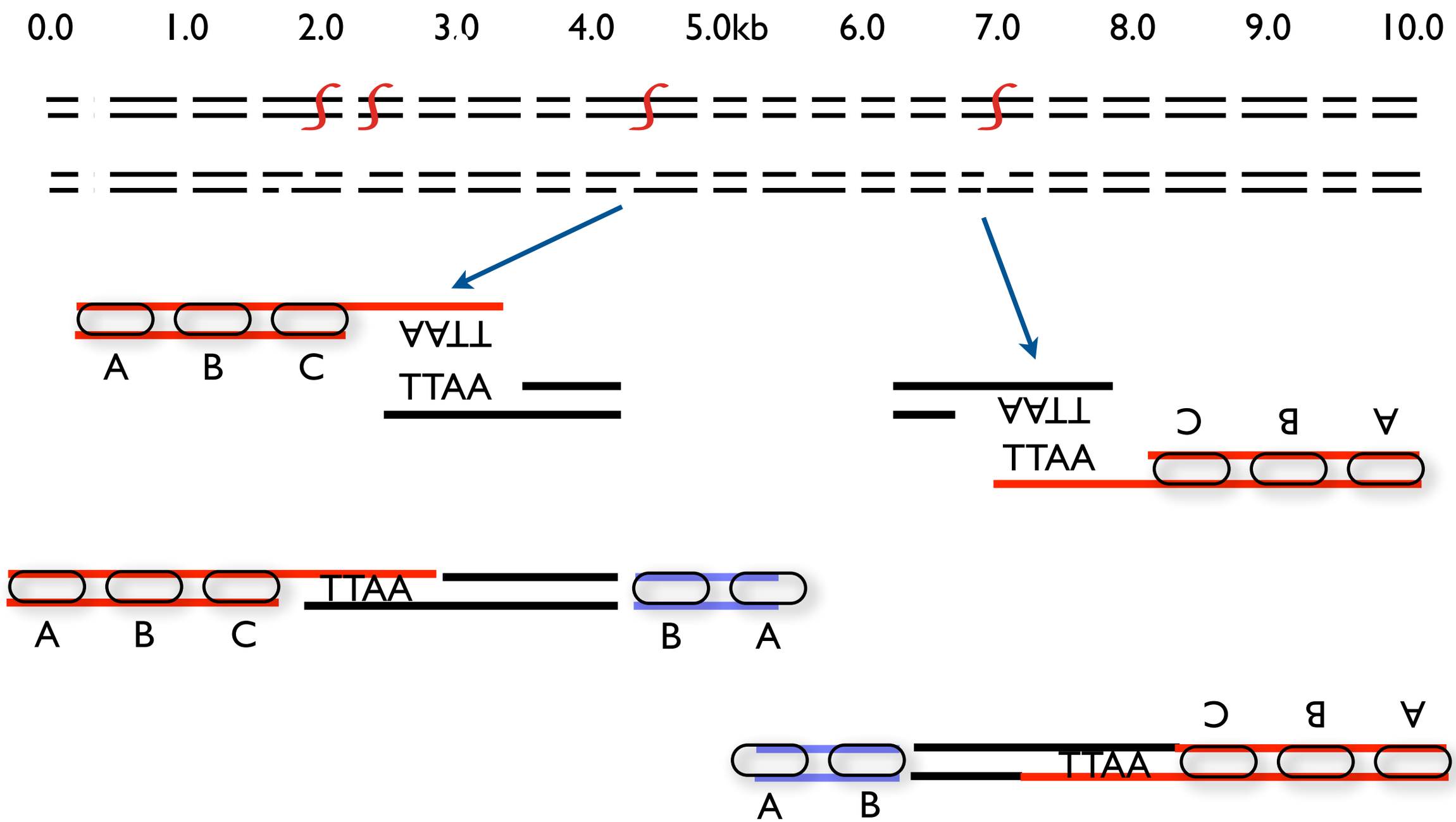
Shearing and second adaptor ligation



* Important step here*

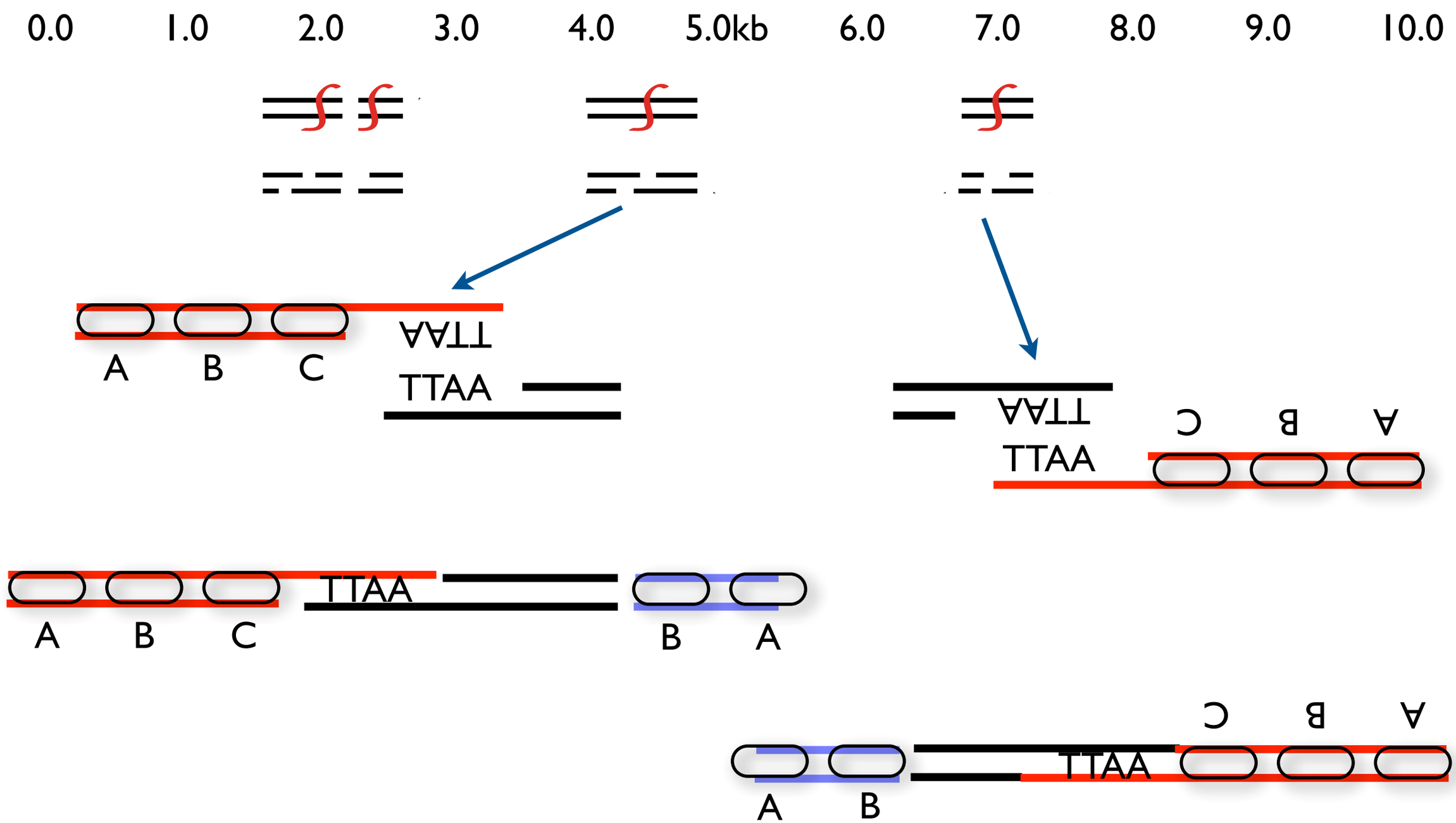
A = Amplification primer
B = Sequencing primer
C = Barcode

Shearing and second adaptor ligation



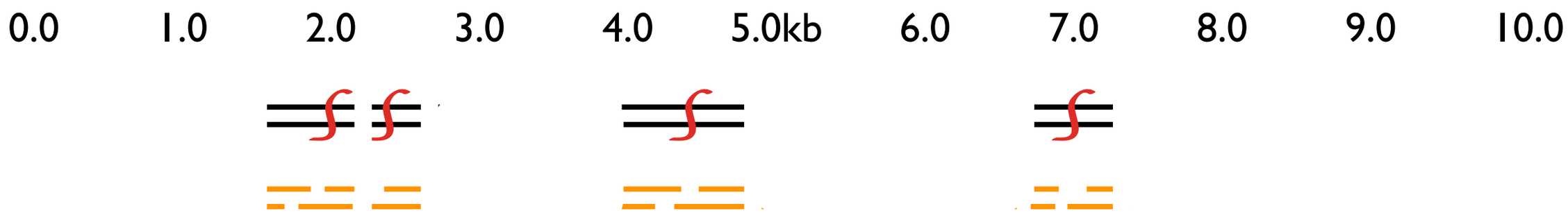
A = Amplification primer
B = Sequencing primer
C = Barcode

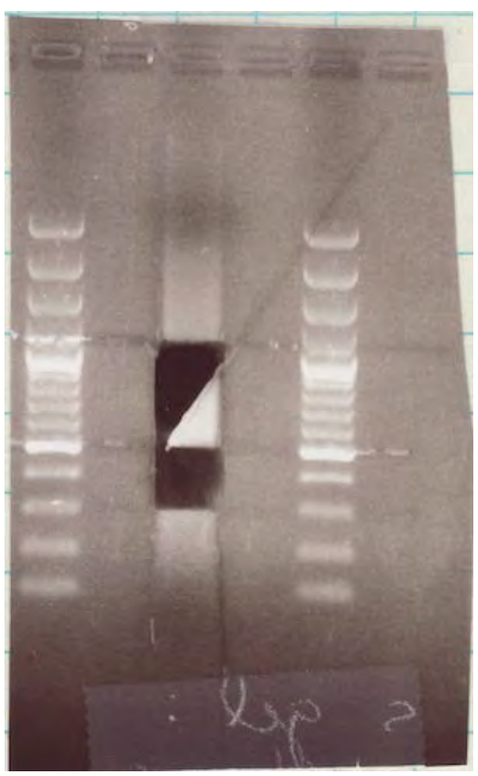
Shearing and second adaptor ligation

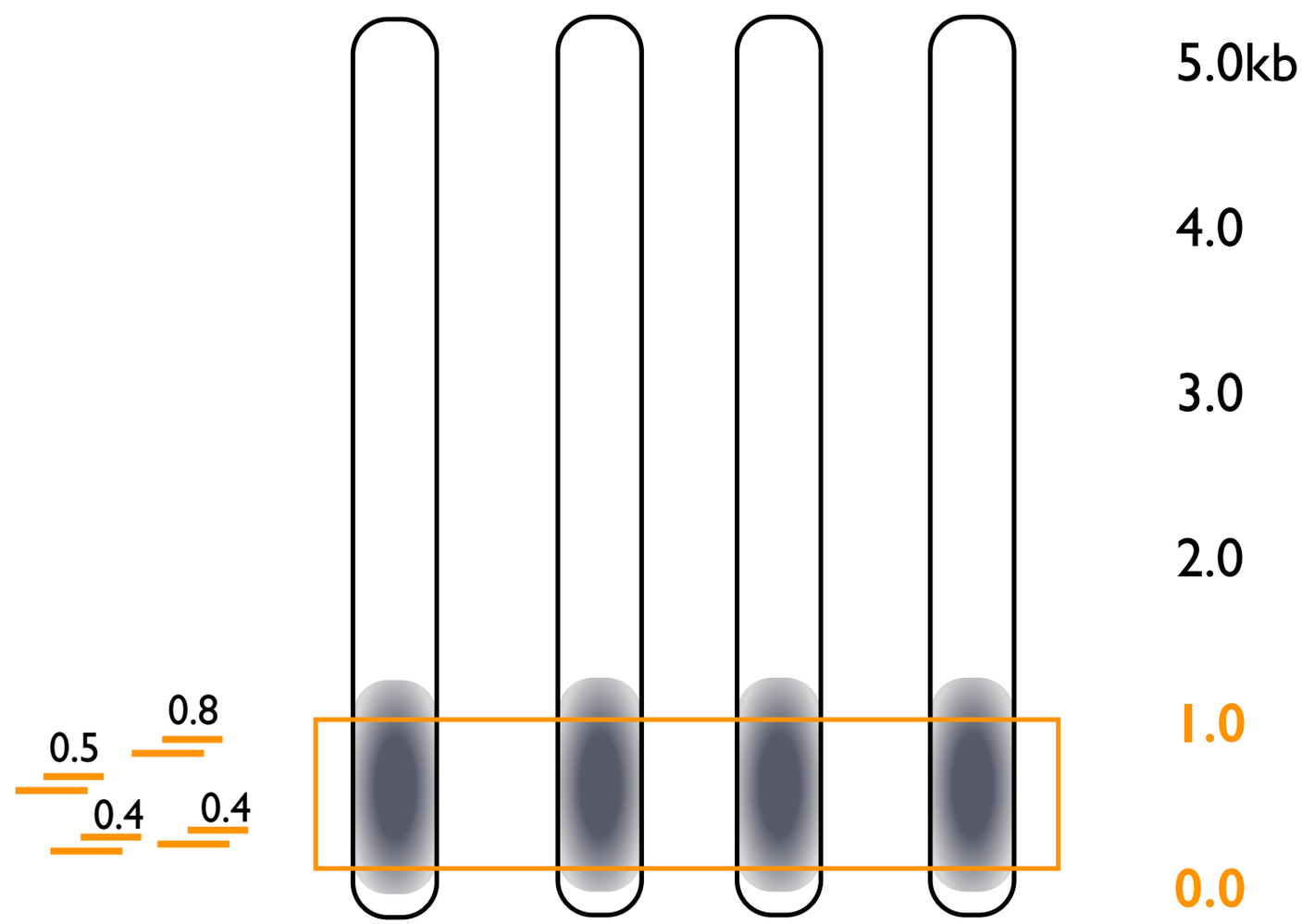


A = Amplification primer
B = Sequencing primer
C = Barcode

Shearing makes consistent fragments for sequencing

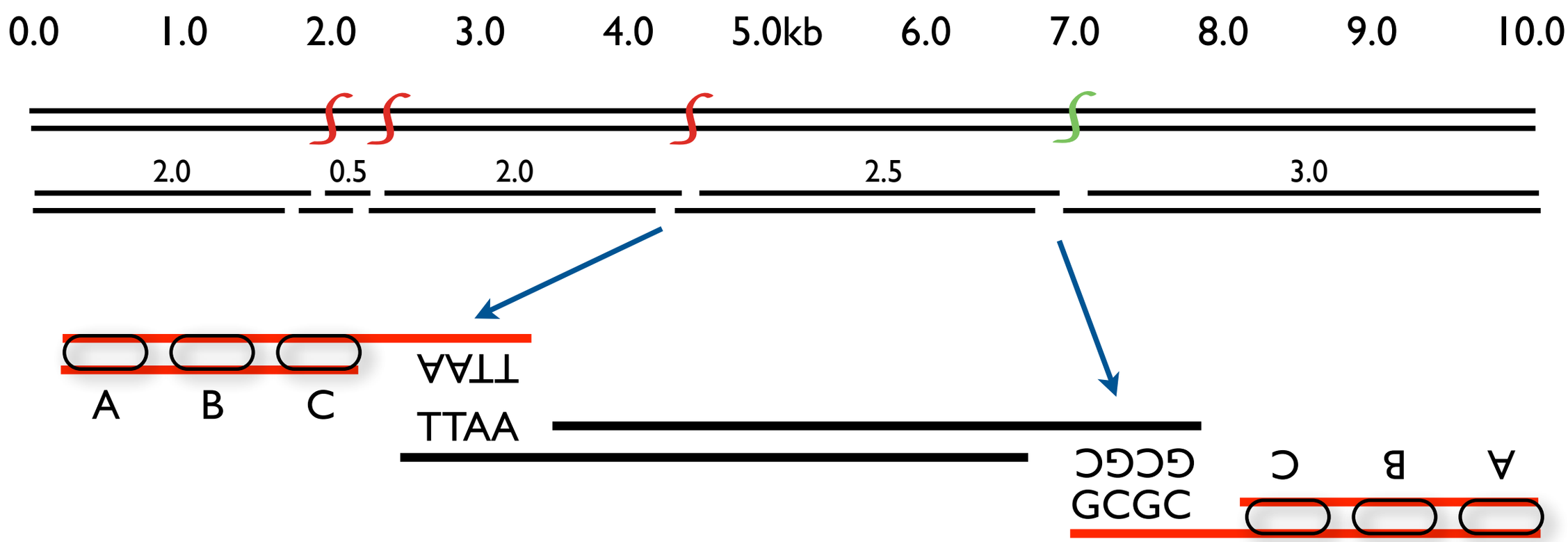






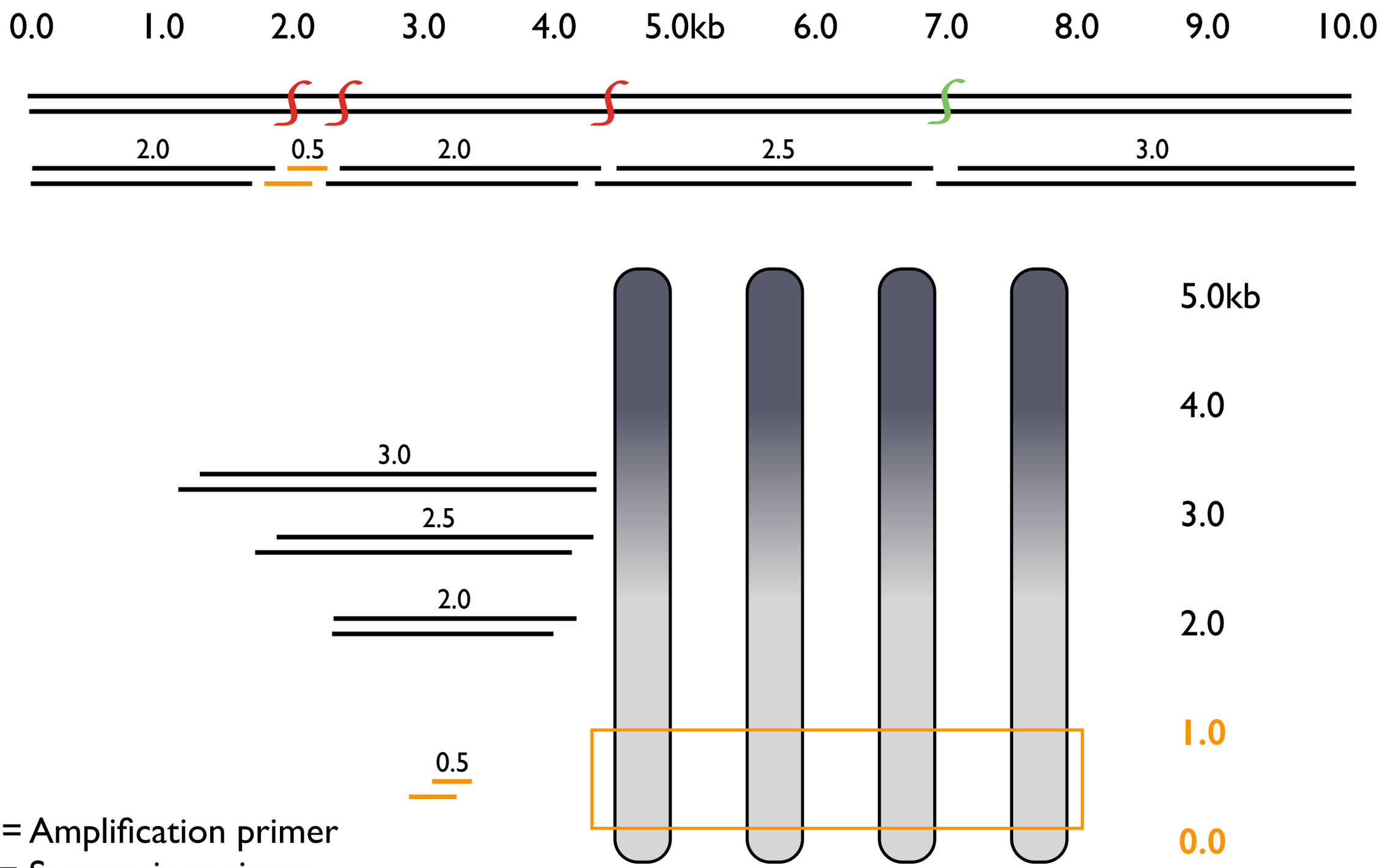
A = Amplification primer
B = Sequencing primer
C = Barcode

Single (GBS) or Double Digest RAD (ddRAD)



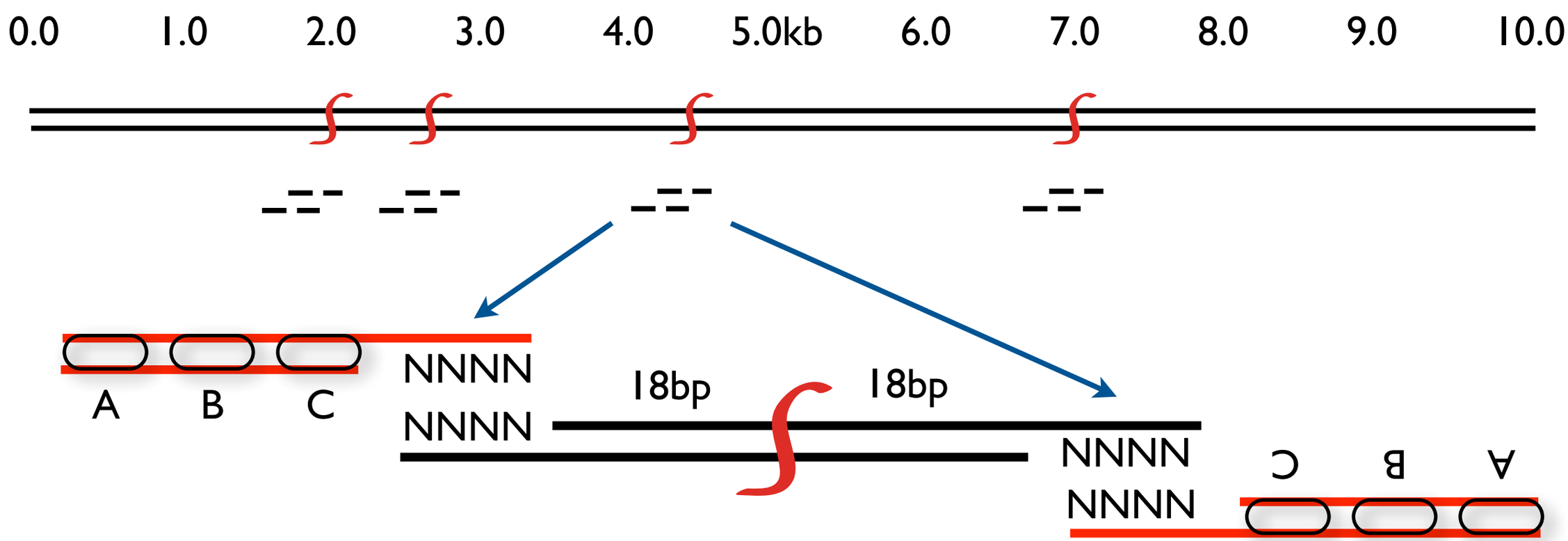
A = Amplification primer
B = Sequencing primer
C = Barcode

Size selection is more problematic without shearing



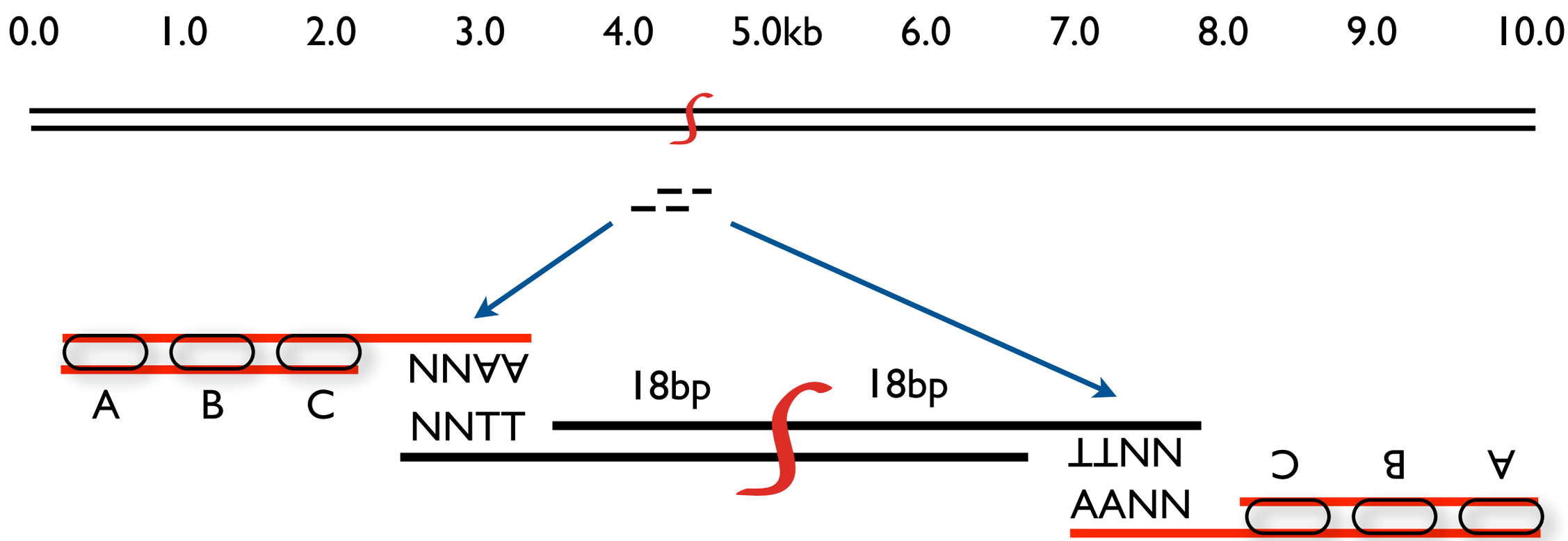
A = Amplification primer
B = Sequencing primer
C = Barcode

2bRAD - type 2b restriction enzyme



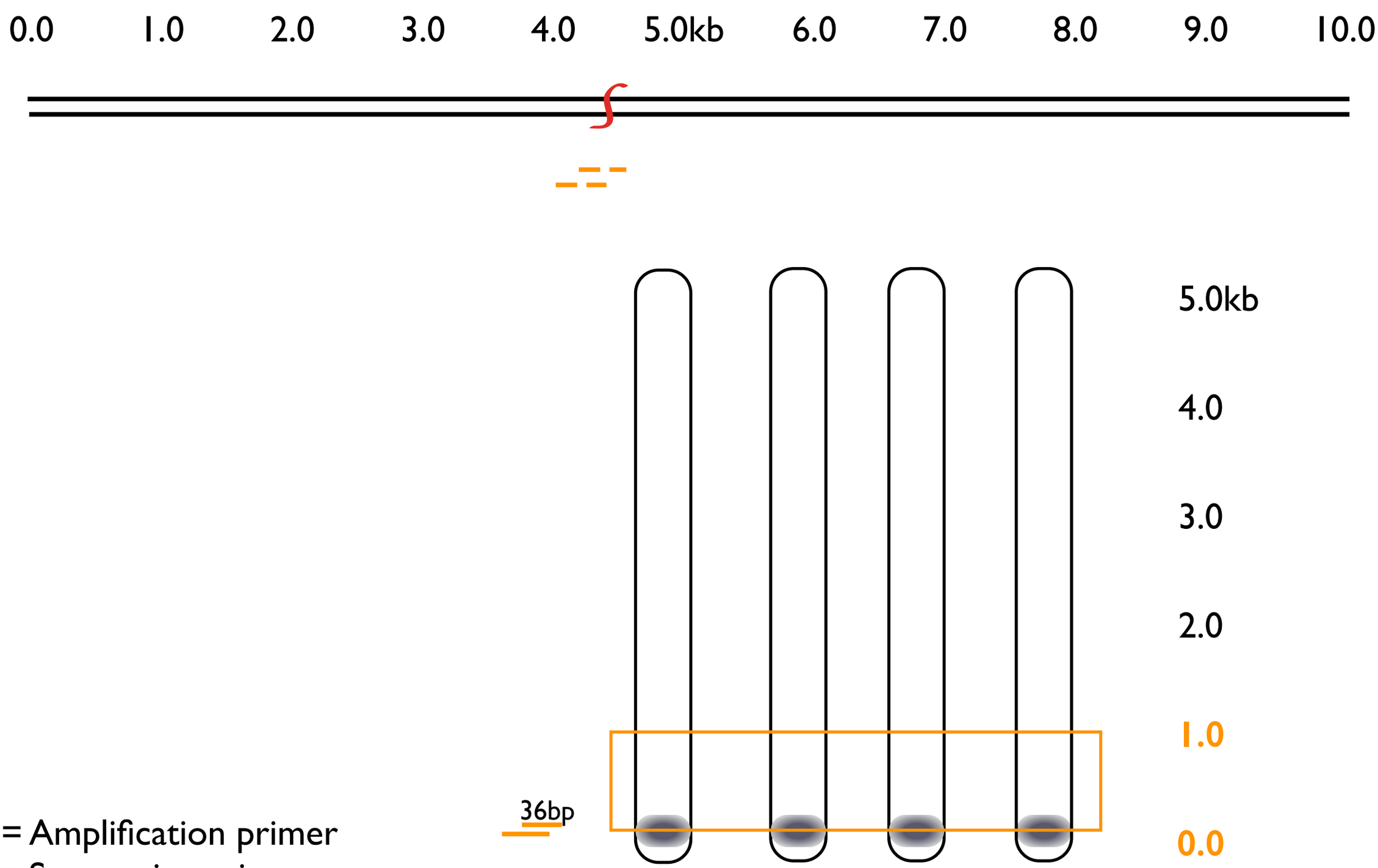
A = Amplification primer
B = Sequencing primer
C = Barcode

2bRAD - can scale number of markers easily



A = Amplification primer
B = Sequencing primer
C = Barcode

2bRAD - size selection is difficult

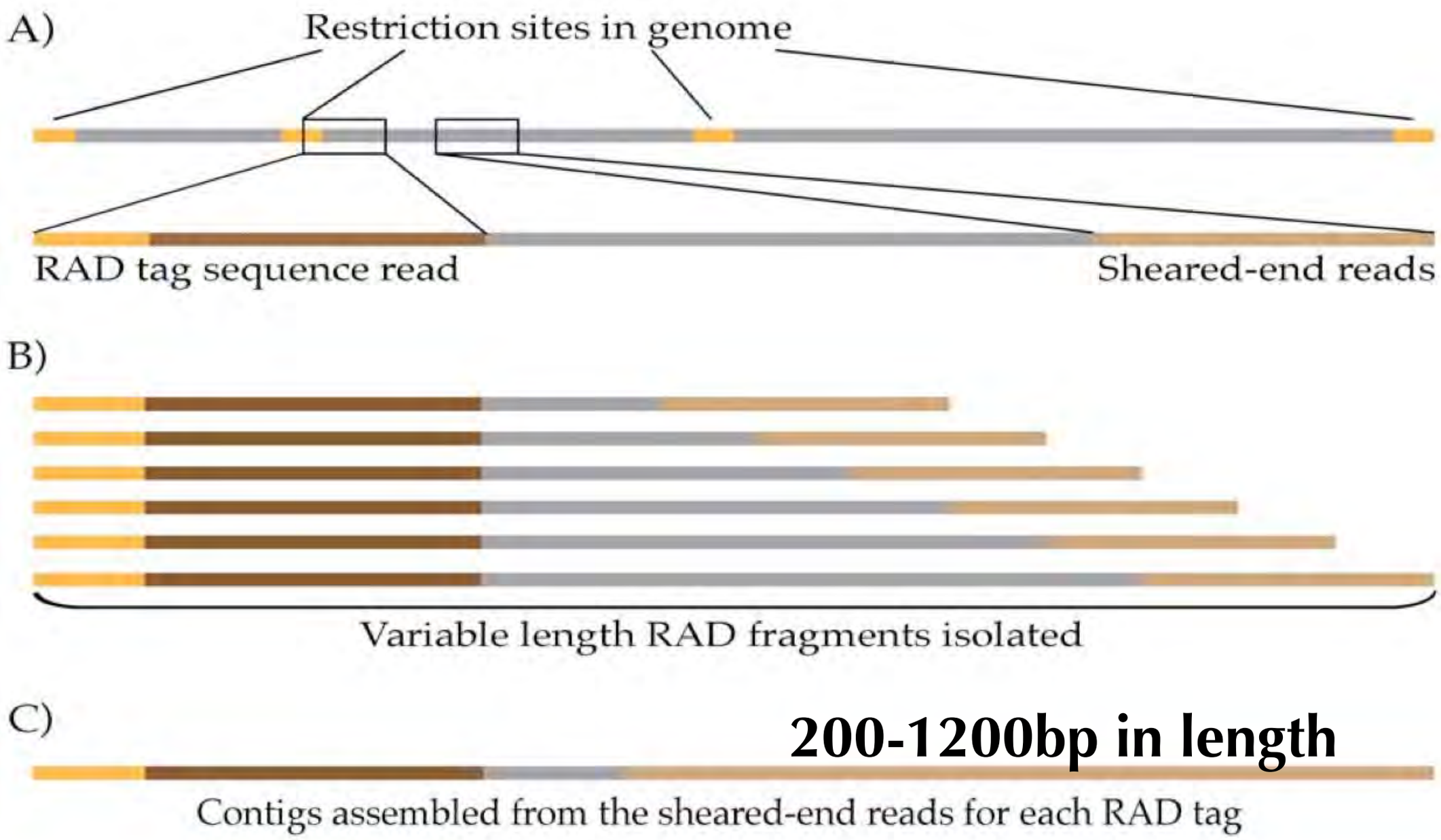


A = Amplification primer
B = Sequencing primer
C = Barcode

Summary of plusses and minuses of RAD family

	Sheared RAD	Single or ddRAD	2b-RAD
plusses	- Consistent reads - Local assemblies - Identify PCR duplicates	- Fewer steps - Easy marker scaling	- Fewest steps - Easy marker scaling
minuses	- Shearing step - Scaling requires different enzymes	- Multiple enzymes - Poor consistency - PCR duplicates	- Very short reads - PCR duplicates

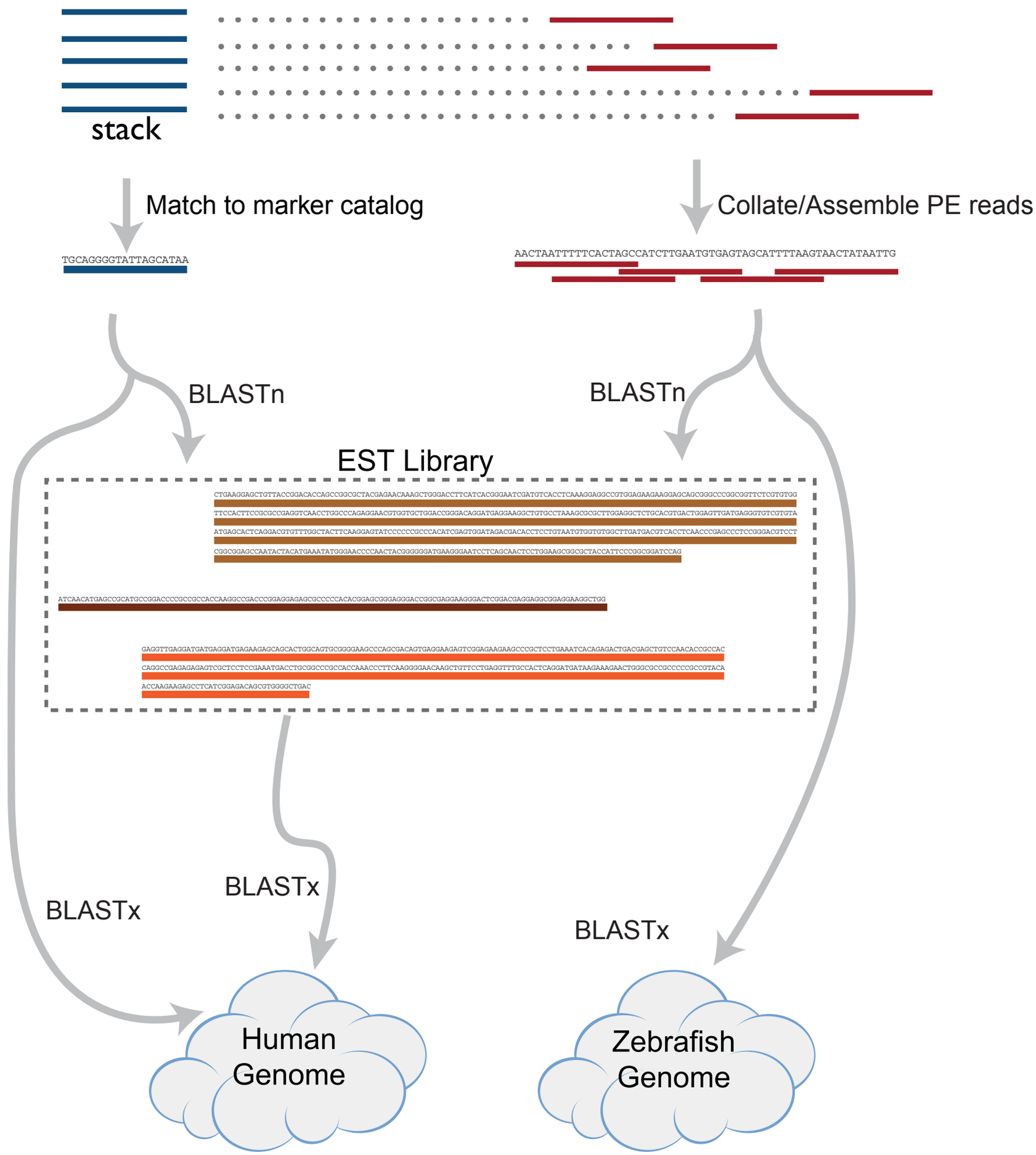
Additional benefits of random shearing in RAD



Acquire paired-end sequence

Associate markers / PE contigs with ESTs

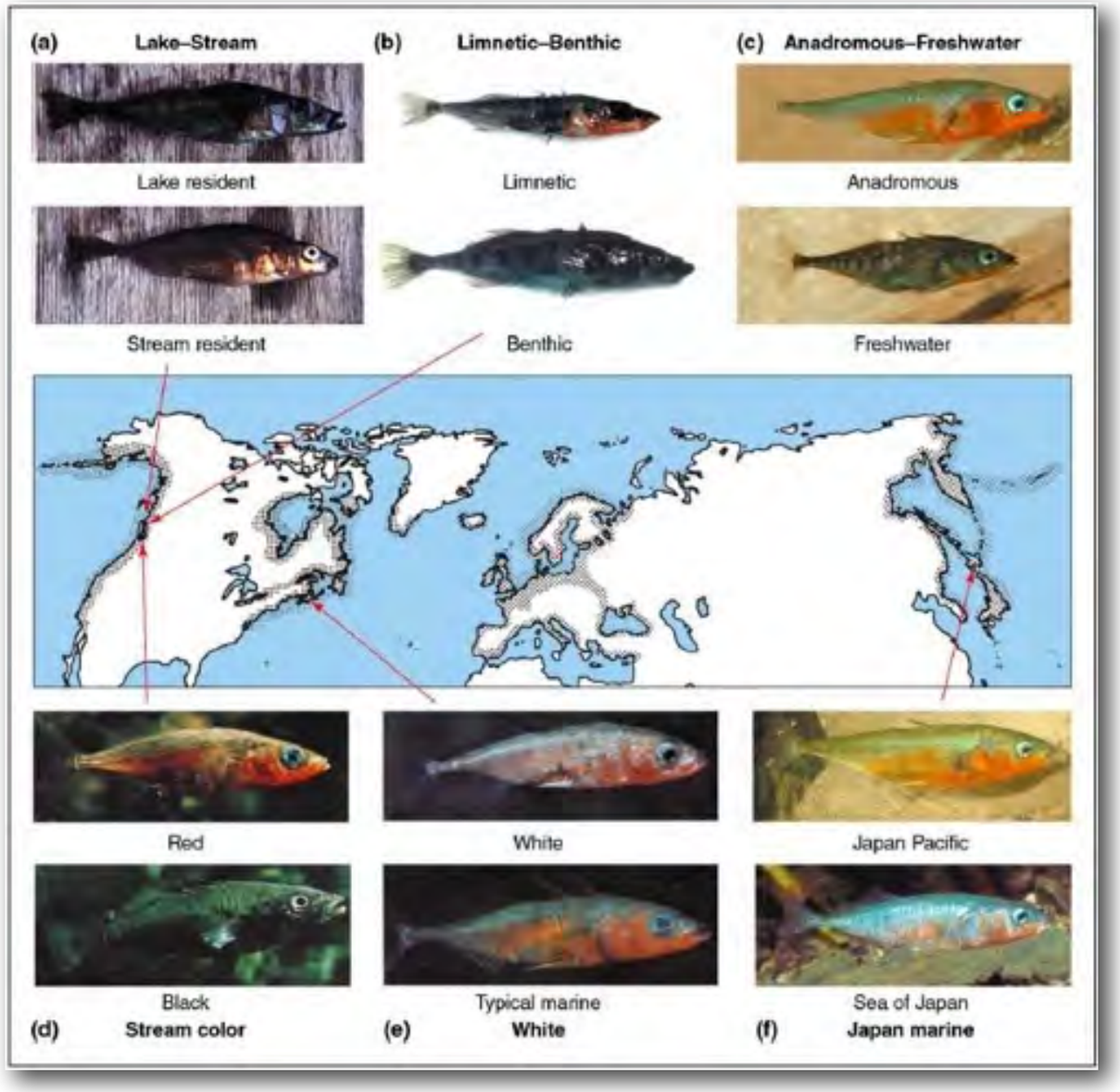
Assign orthology to:
- markers
- PE contigs
- ESTs



Case studies of using RAD for an organism with a reference genome: population genomics of threespine stickleback

Threespine stickleback, *Gasterosteus aculeatus*

Threespine stickleback, *Gasterosteus aculeatus*

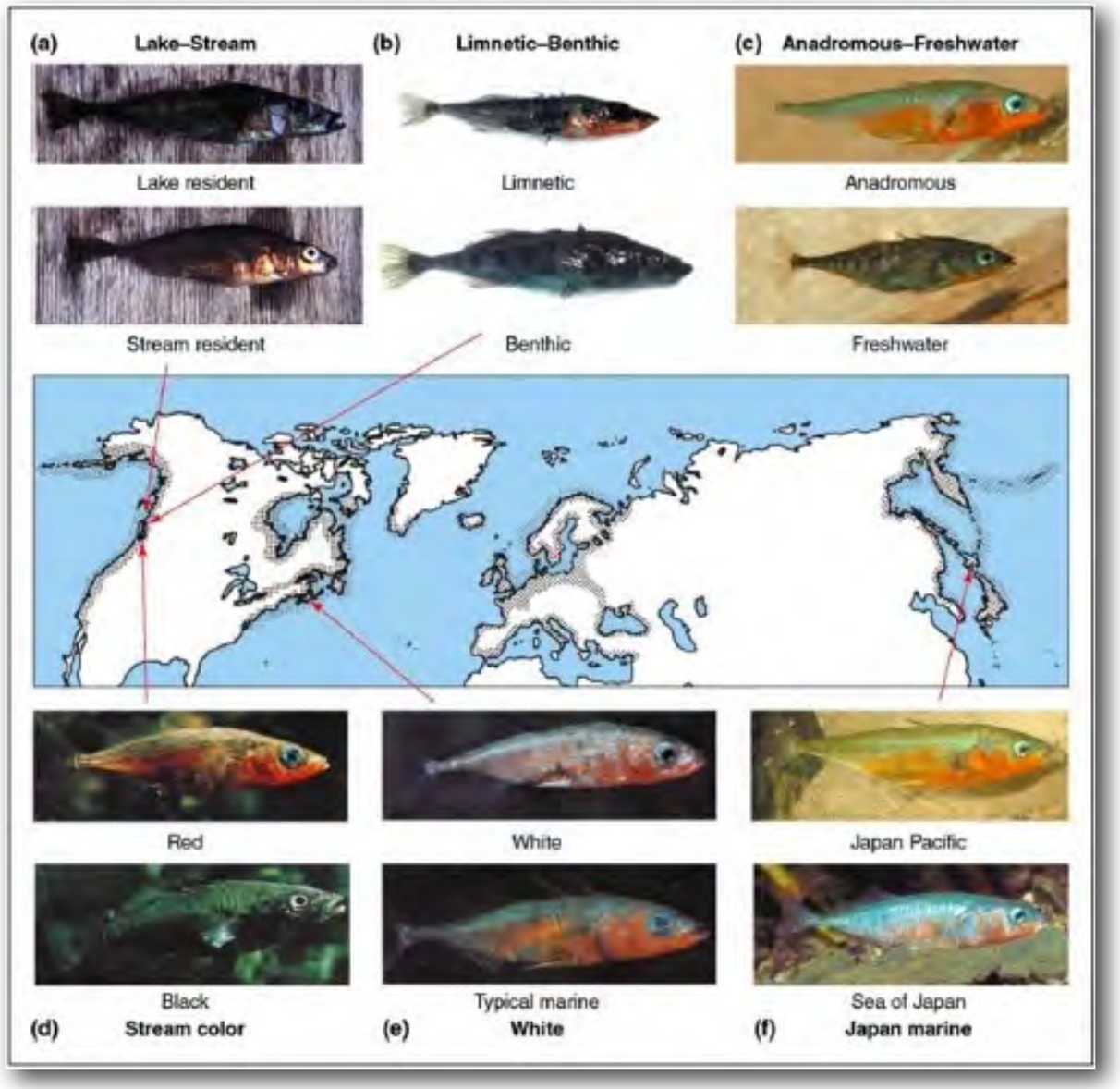


Rundle and McKinnon 2002

Threespine stickleback, *Gasterosteus aculeatus*

Pelvic Structure

Lateral Plates

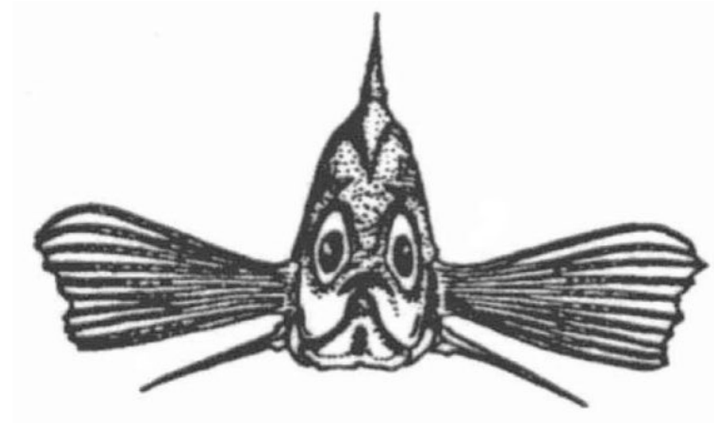


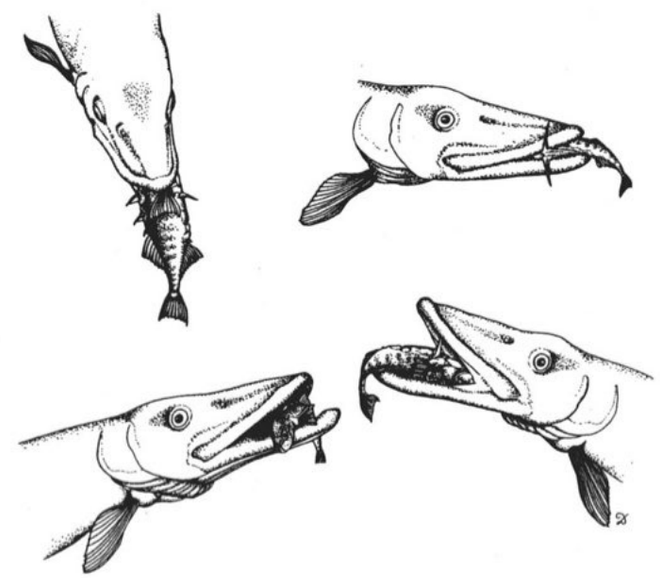
Rundle and McKinnon 2002

Threespine stickleback, *Gasterosteus aculeatus*

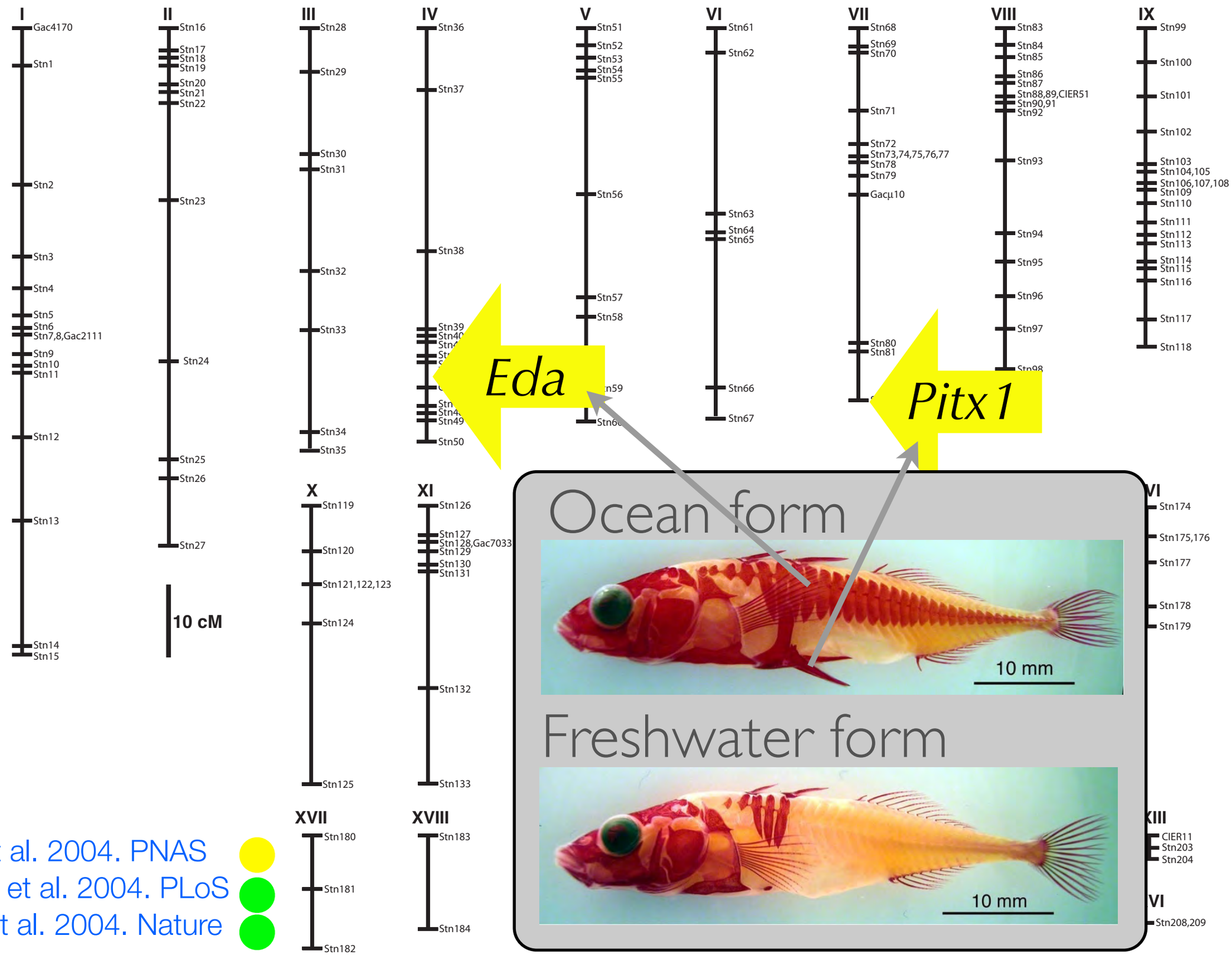
Pelvic Structure

Lateral Plates

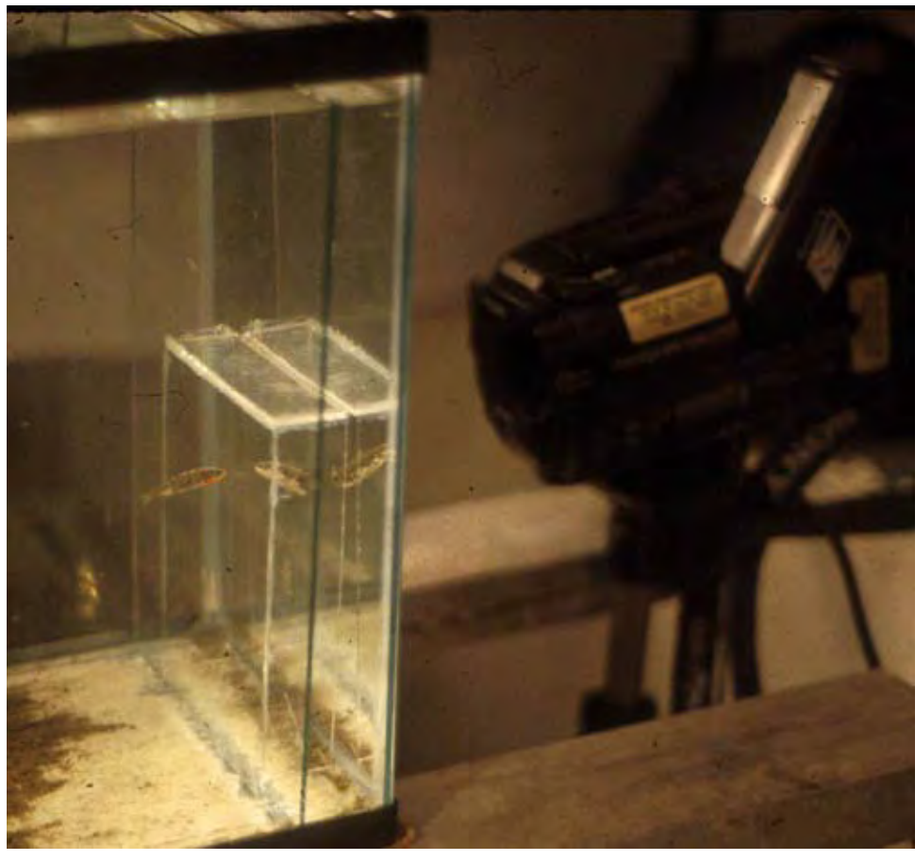




Laboratory mapping of large effect loci



Cresko et al. 2004. PNAS
Colosimo et al. 2004. PLoS
Shapiro et al. 2004. Nature



Stickleback phenotypes mapped in the lab so far.....

Pelvic structure size and shape *** (*Eda*)
Lateral plate number *** (*Pitx1*)
Body coloration *** (*KitL*)
Opercle bone shape
Pelvic spine length
Body shape
Courtship behavior
Gill raker size
Dorsal spine length

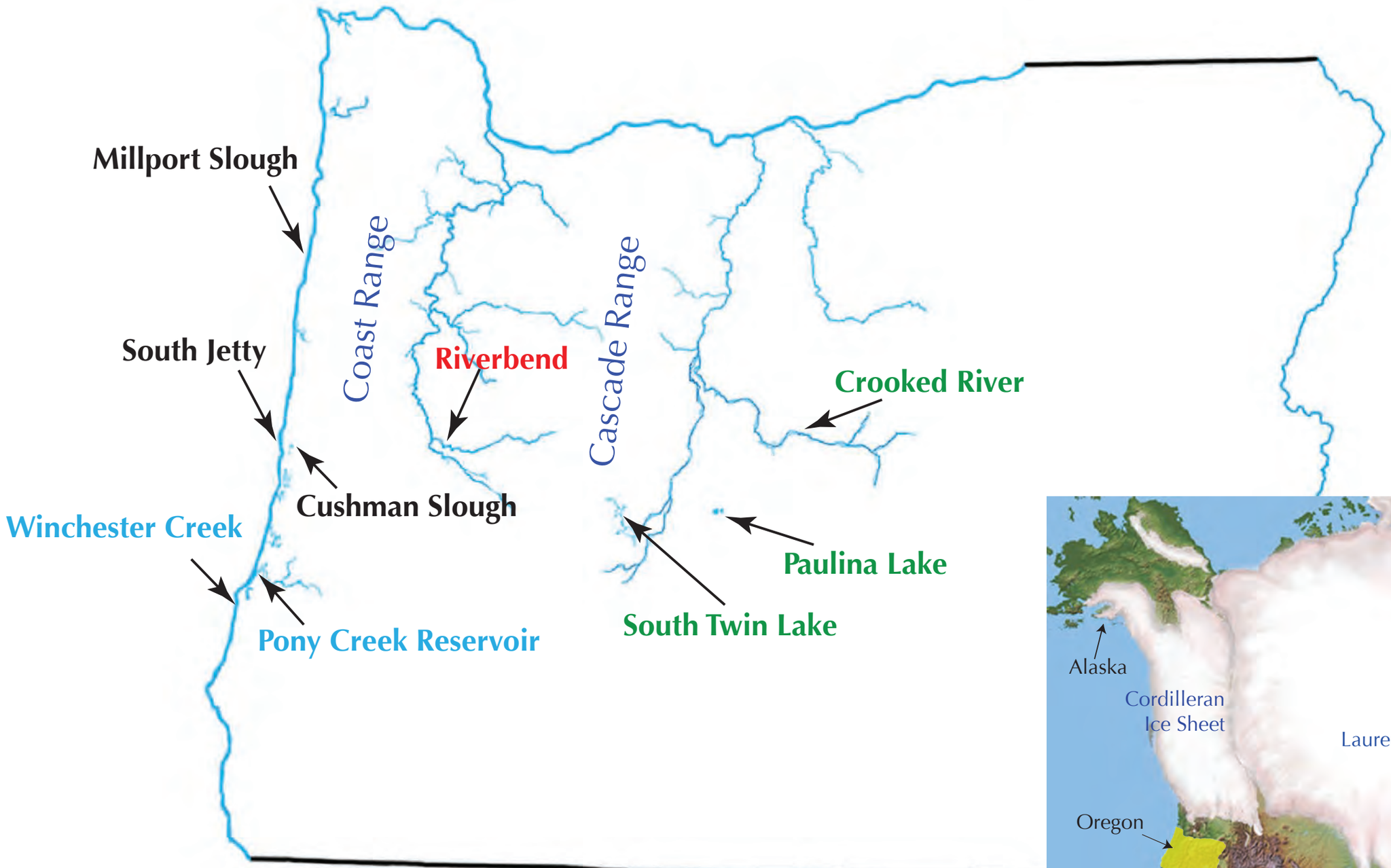
- A trend of large effect loci identified in the laboratory

- Similar genomic regions and sometimes alleles mapped in independent populations

- A problem is that laboratory mapping approaches are under-powered in stickleback

- A question is whether population genomics studies can provide complementary and more complete information.

Population genomic structure of Oregon stickleback

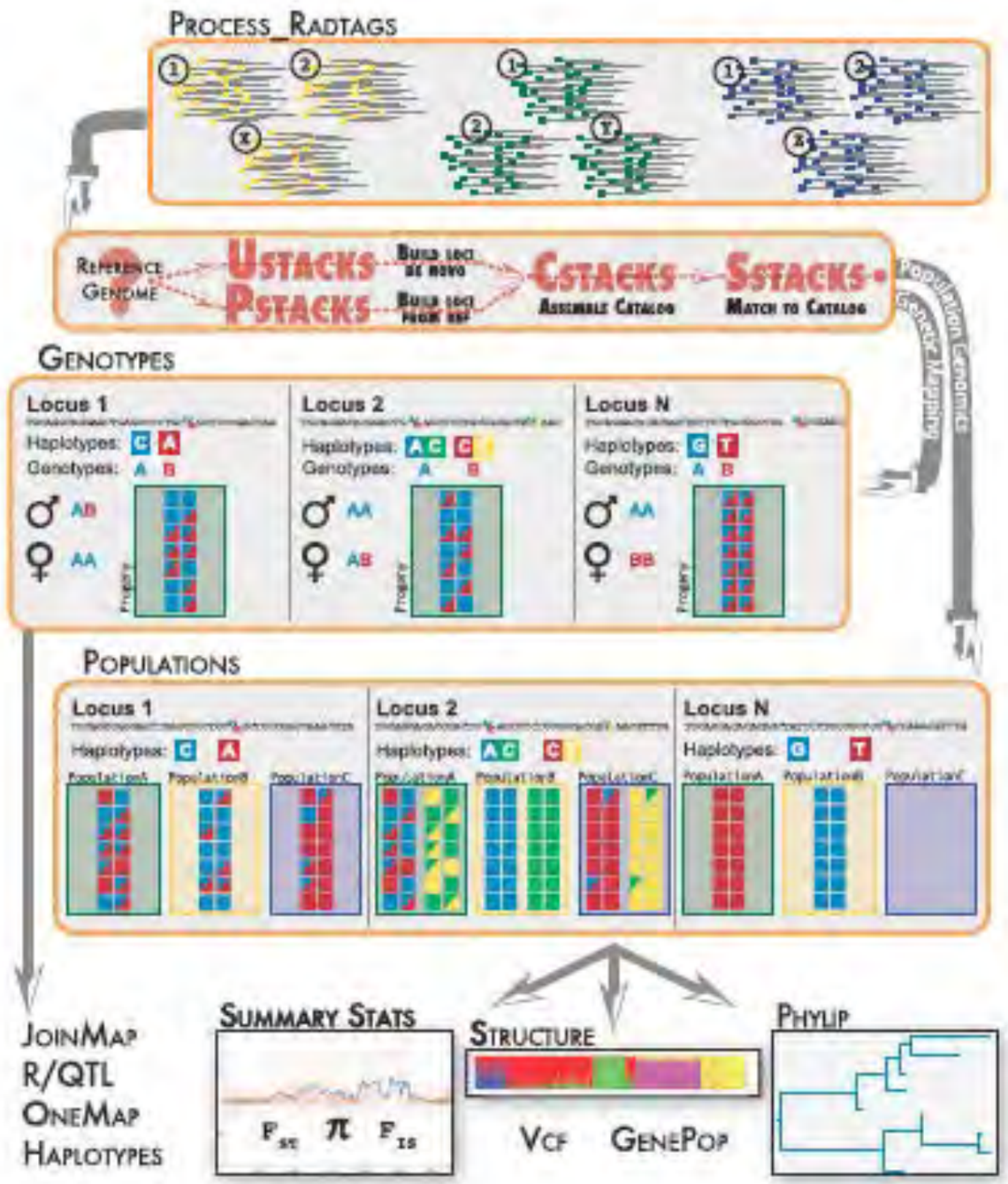


590 Individuals

115,000 SNPs each

Catchen et al. 2013, Molecular Ecology

Stacks analysis pipeline for RAD-seq



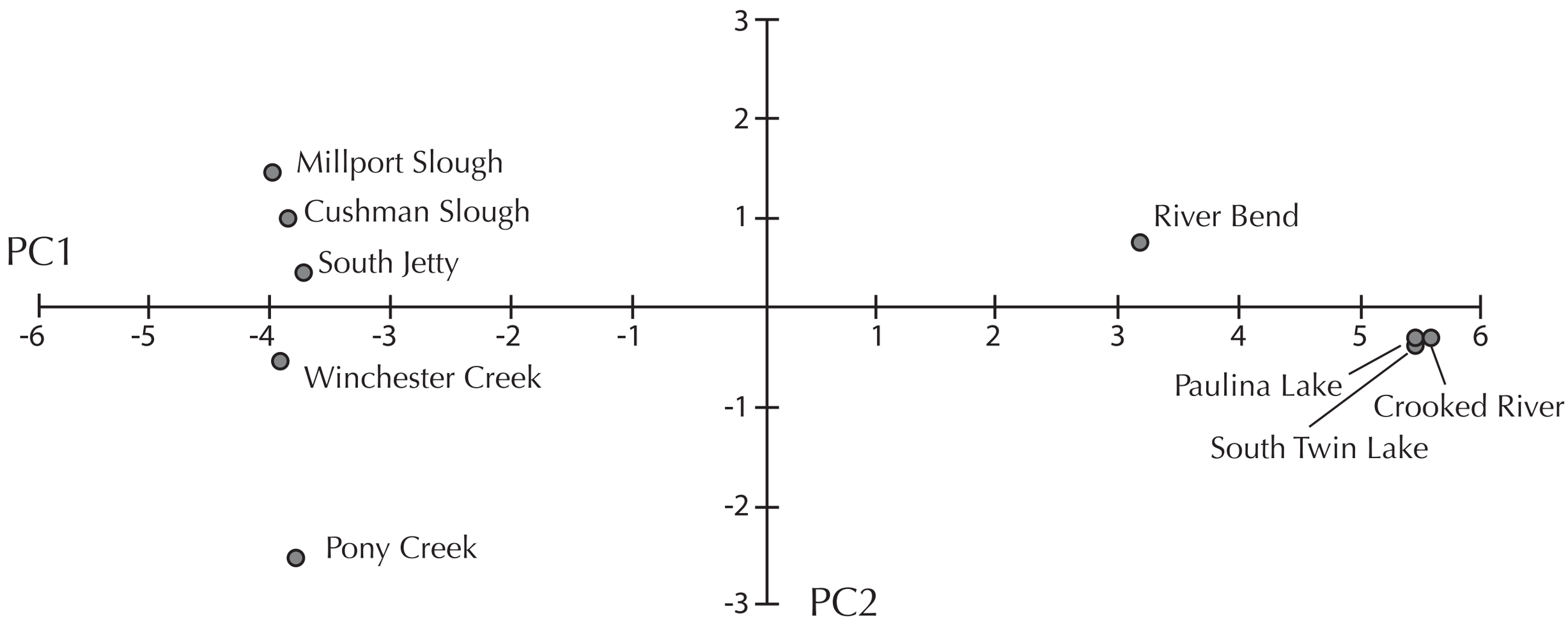
Stacks: Building and Genotyping Loci *De Novo* From Short-Read Sequences

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

Stacks: an analysis tool set for population genomics

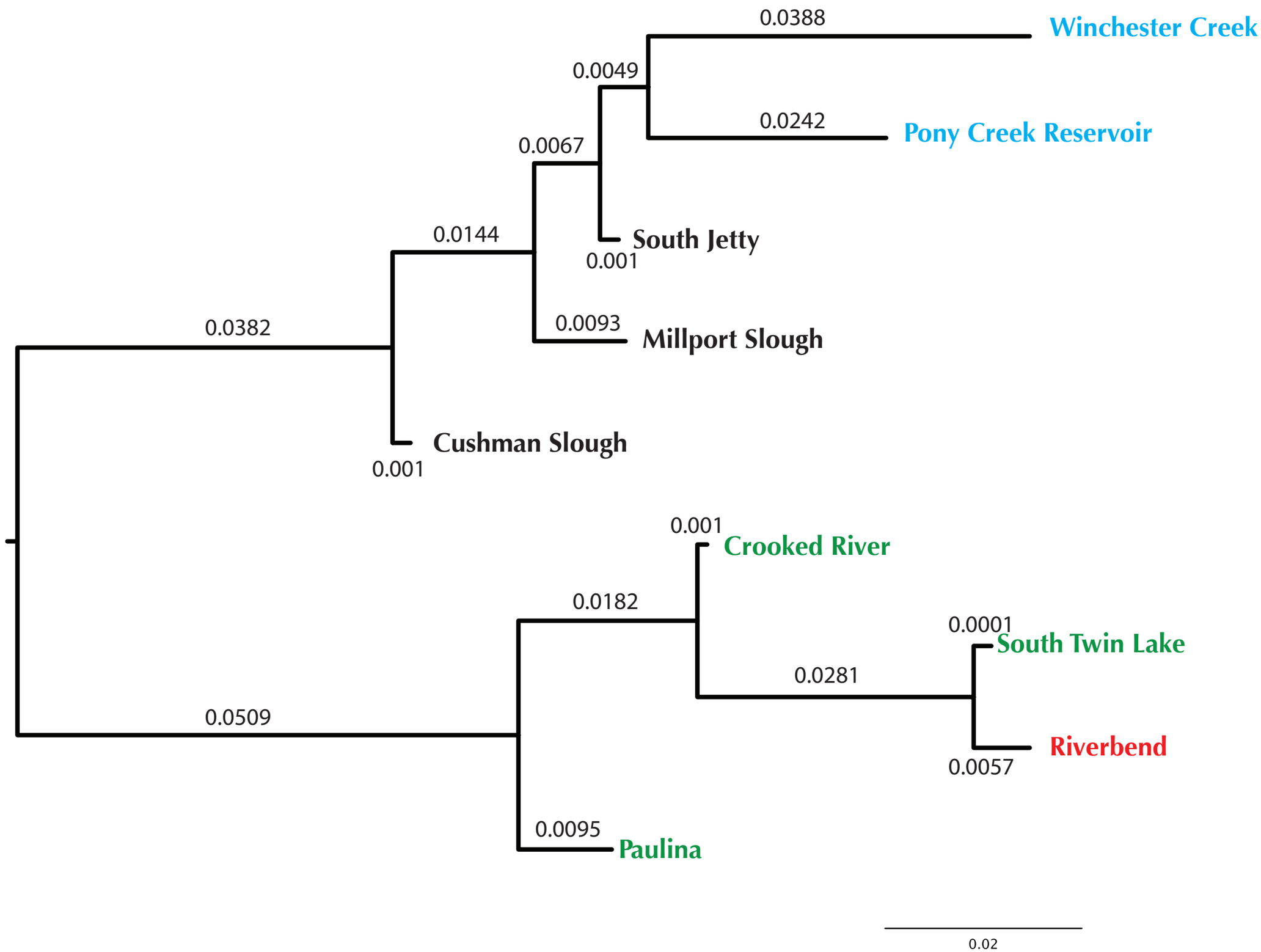
JULIAN CATCHEN,* PAUL A. HOHENLOHE,*† SUSAN BASSHAM,* ANGEL AMORES‡ and WILLIAM A. CRESKO*
*Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403-5289, USA, †Biological Sciences, University of Idaho, Moscow, ID 83844-3051, USA, ‡Institute of Neuroscience, University of Oregon, Eugene, OR 97403-1254, USA

Population structure using PCA

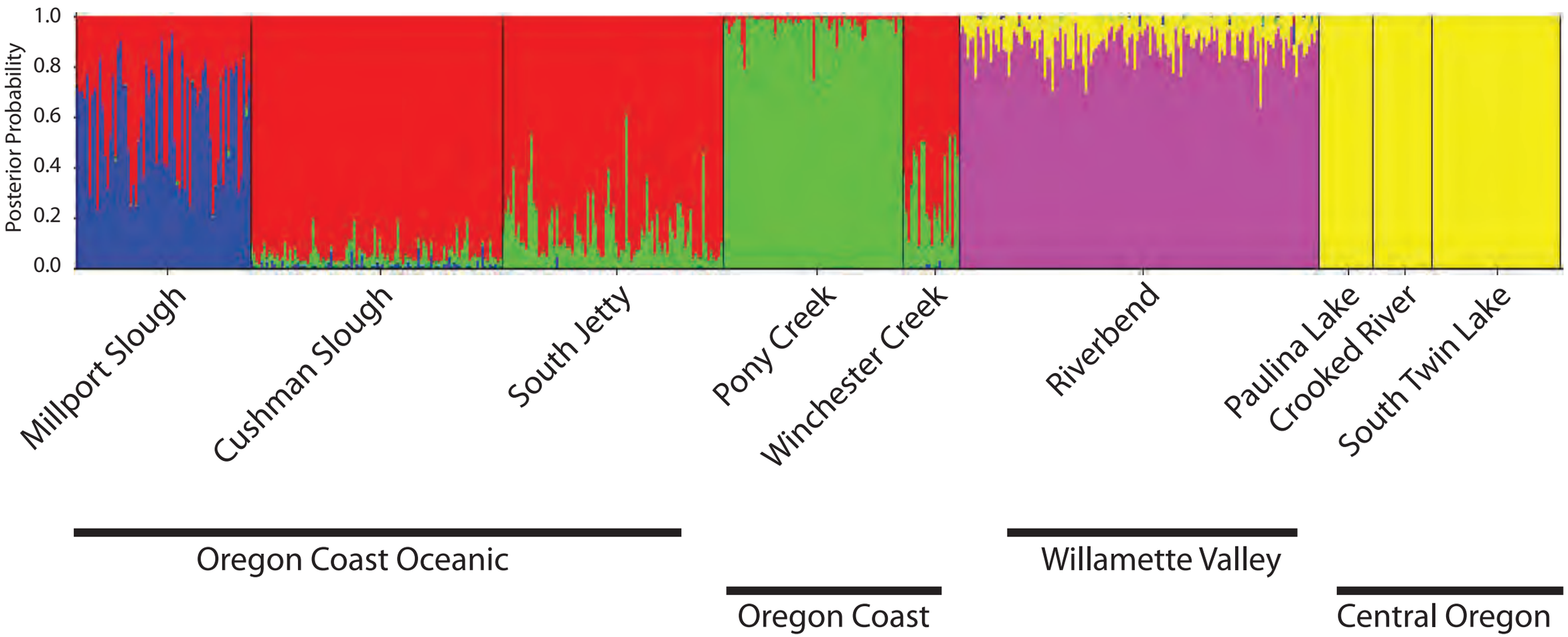


PC 1 explains 89% of the overall variance

Phylogenetic relationship among populations

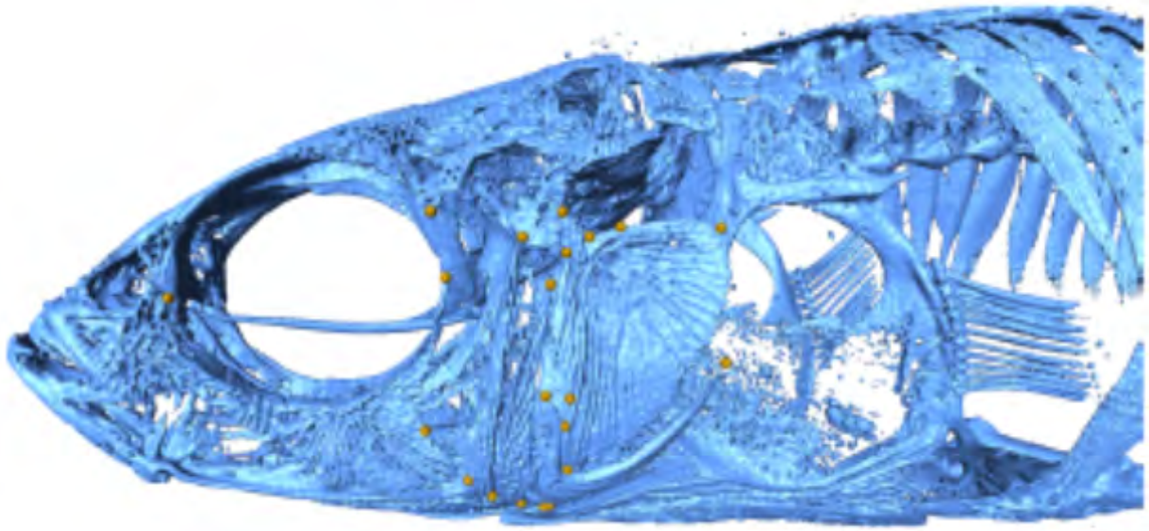


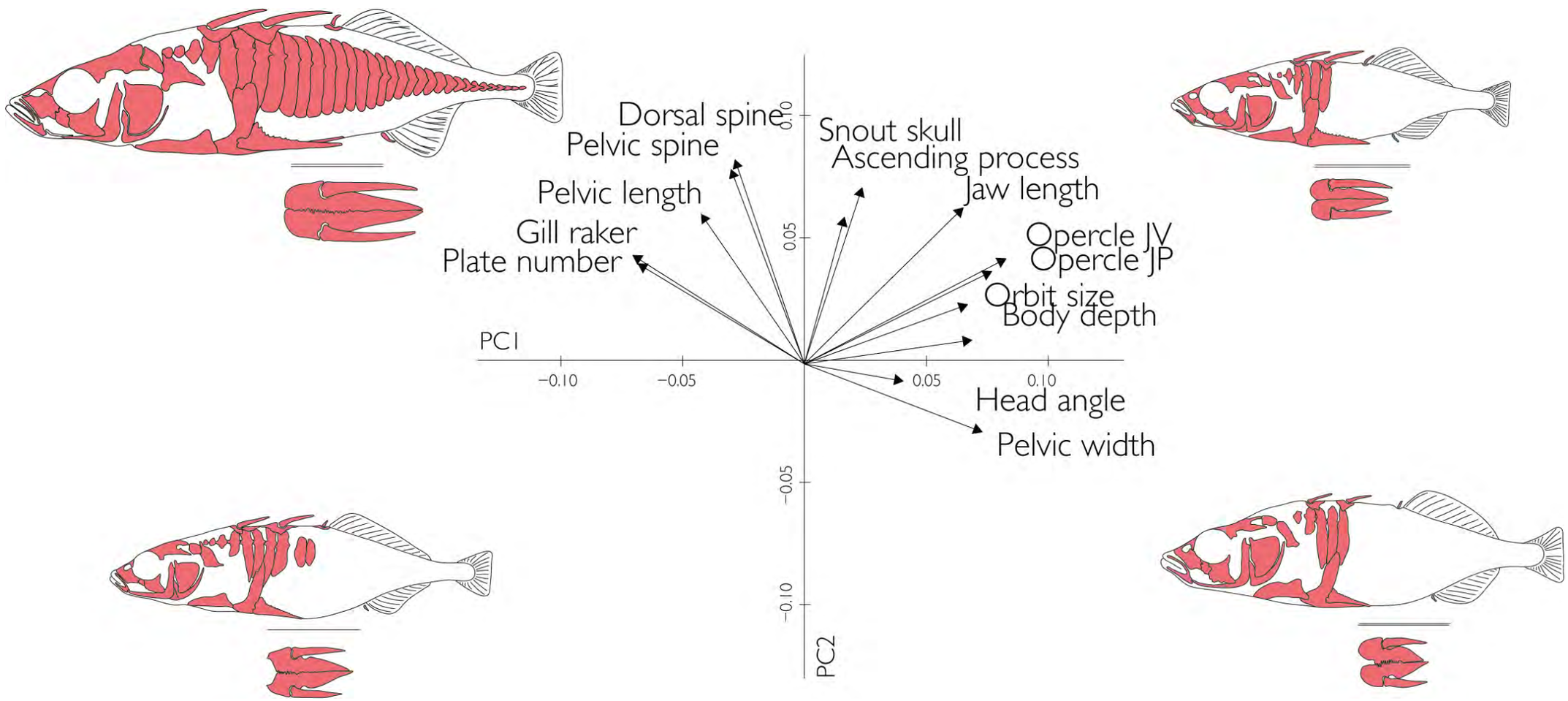
Population structure using Bayesian analysis (*Structure*)



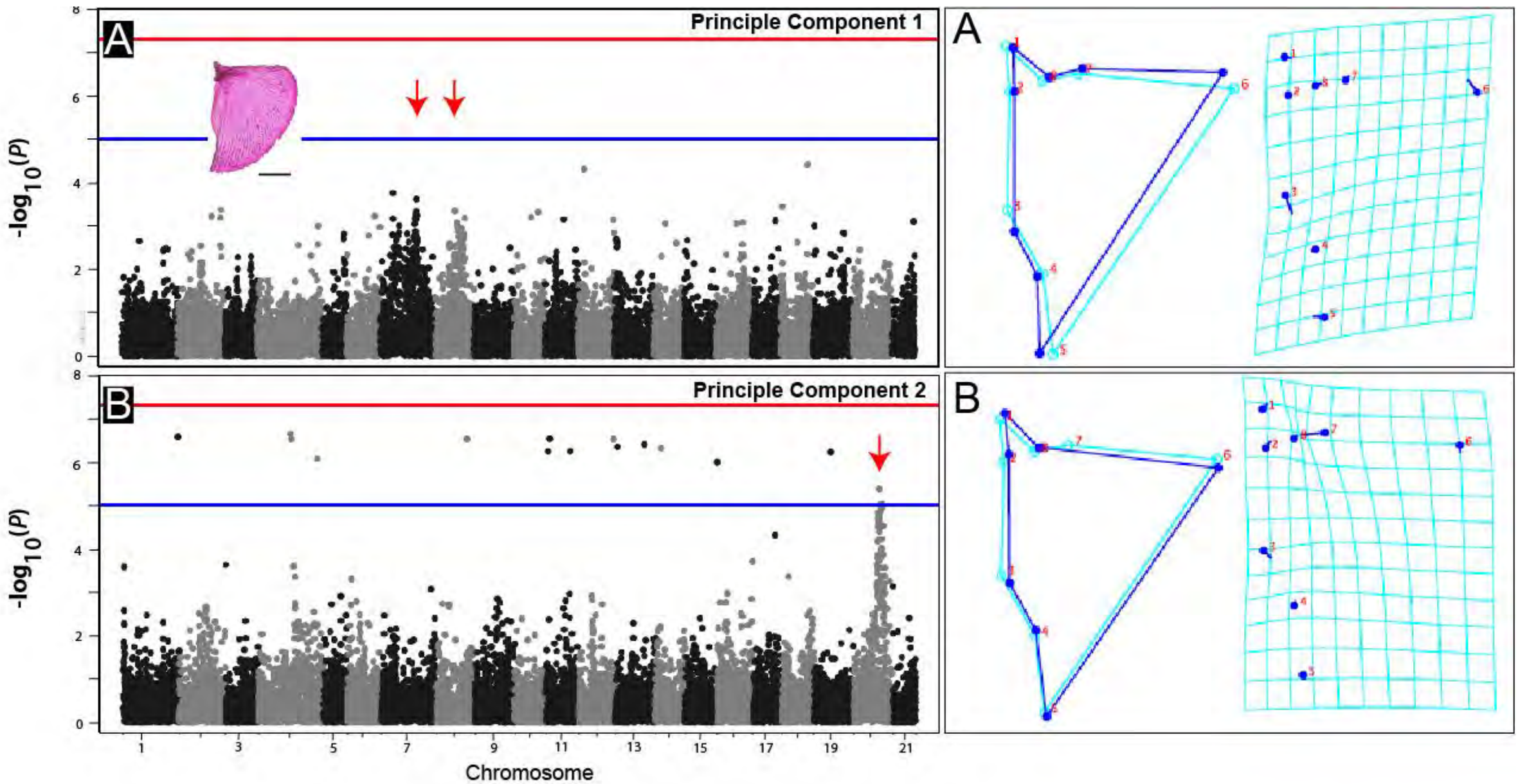
Genome-Wide Association Study (GWAS) using RAD

Kristin Alligood and
Mark Currey

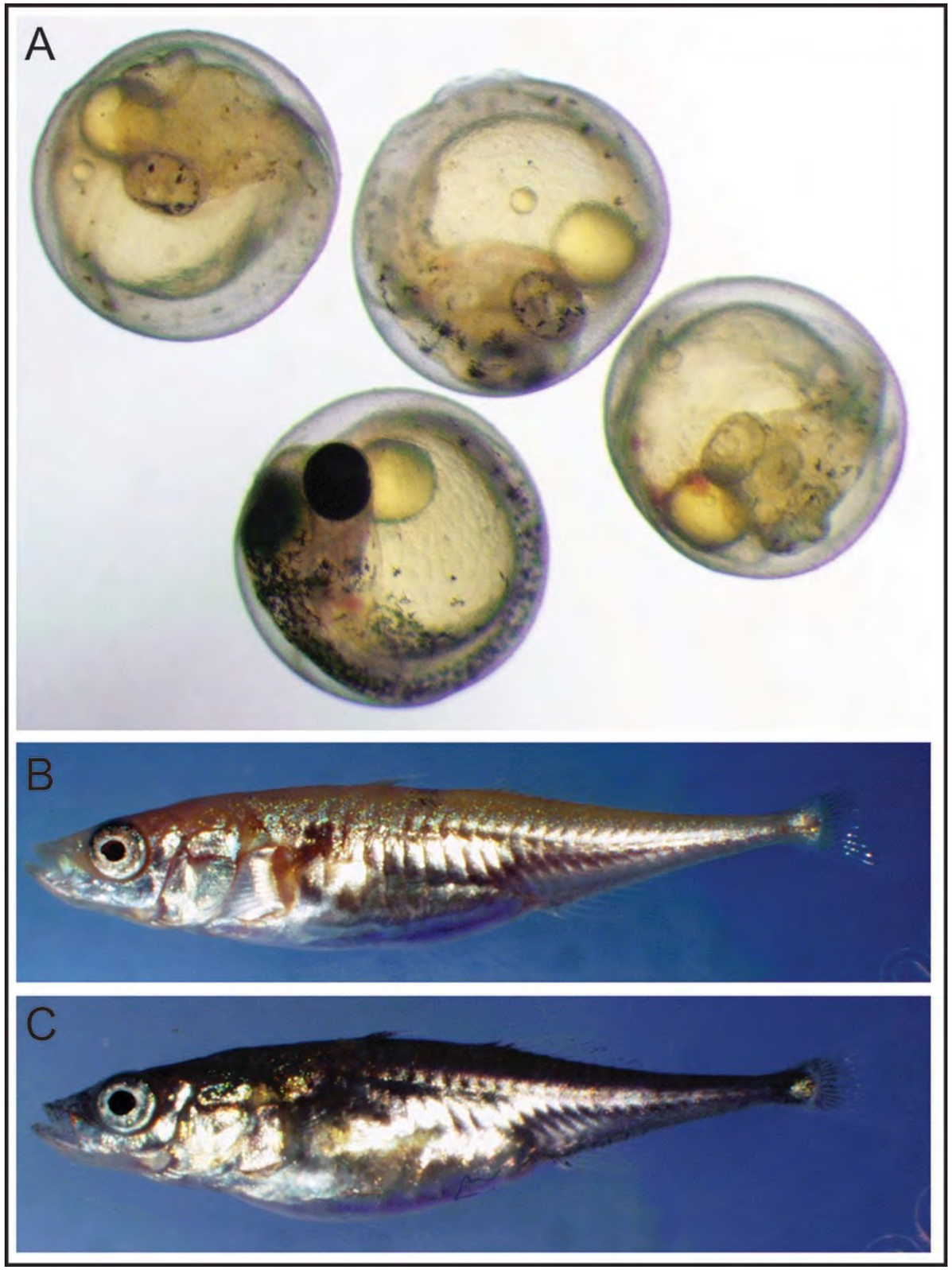




Genome-Wide Association Study (GWAS) using RAD



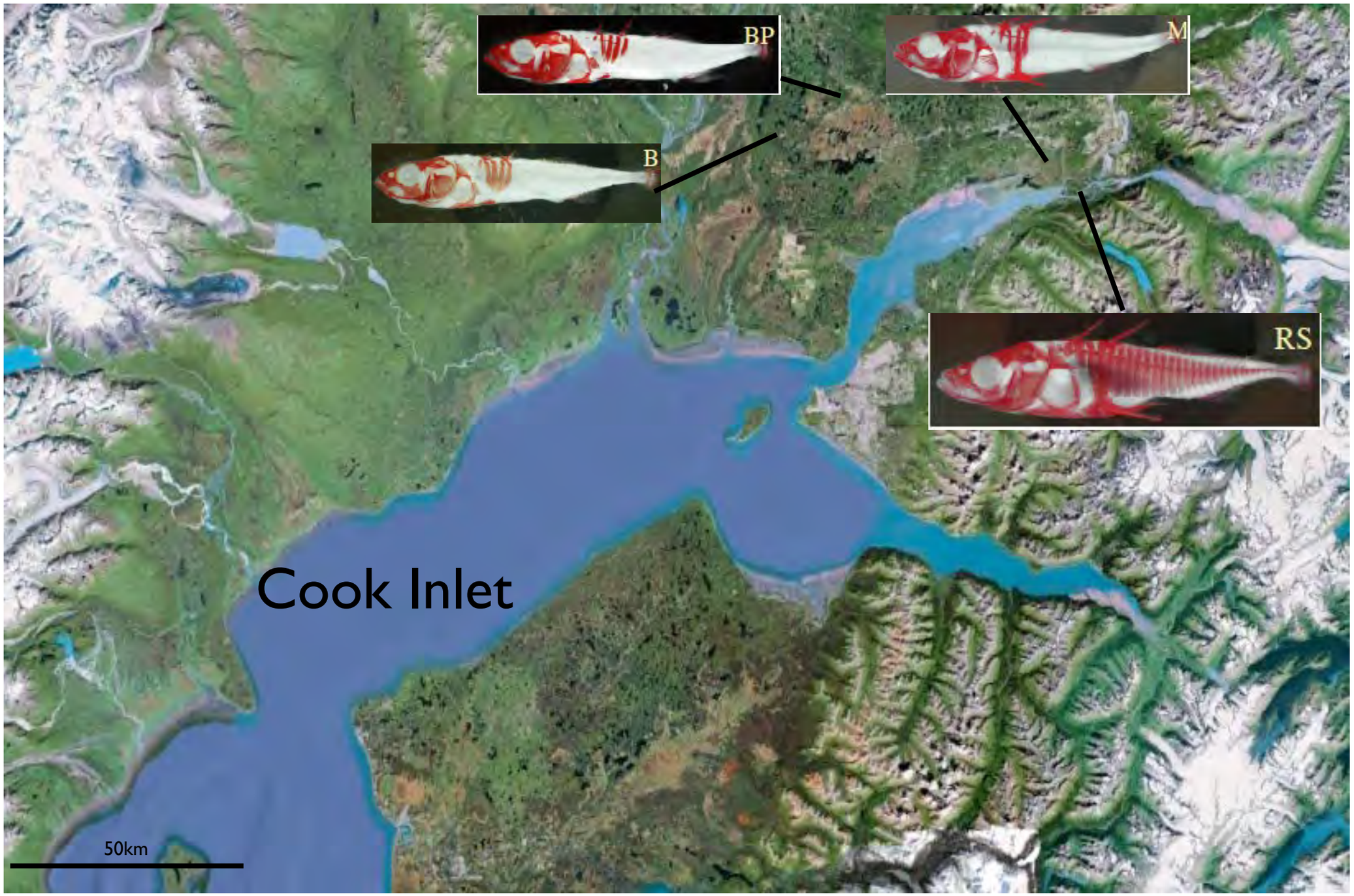
CRISPR gene editing in stickleback to test associations



What genomic regions are associated with the different habitats?

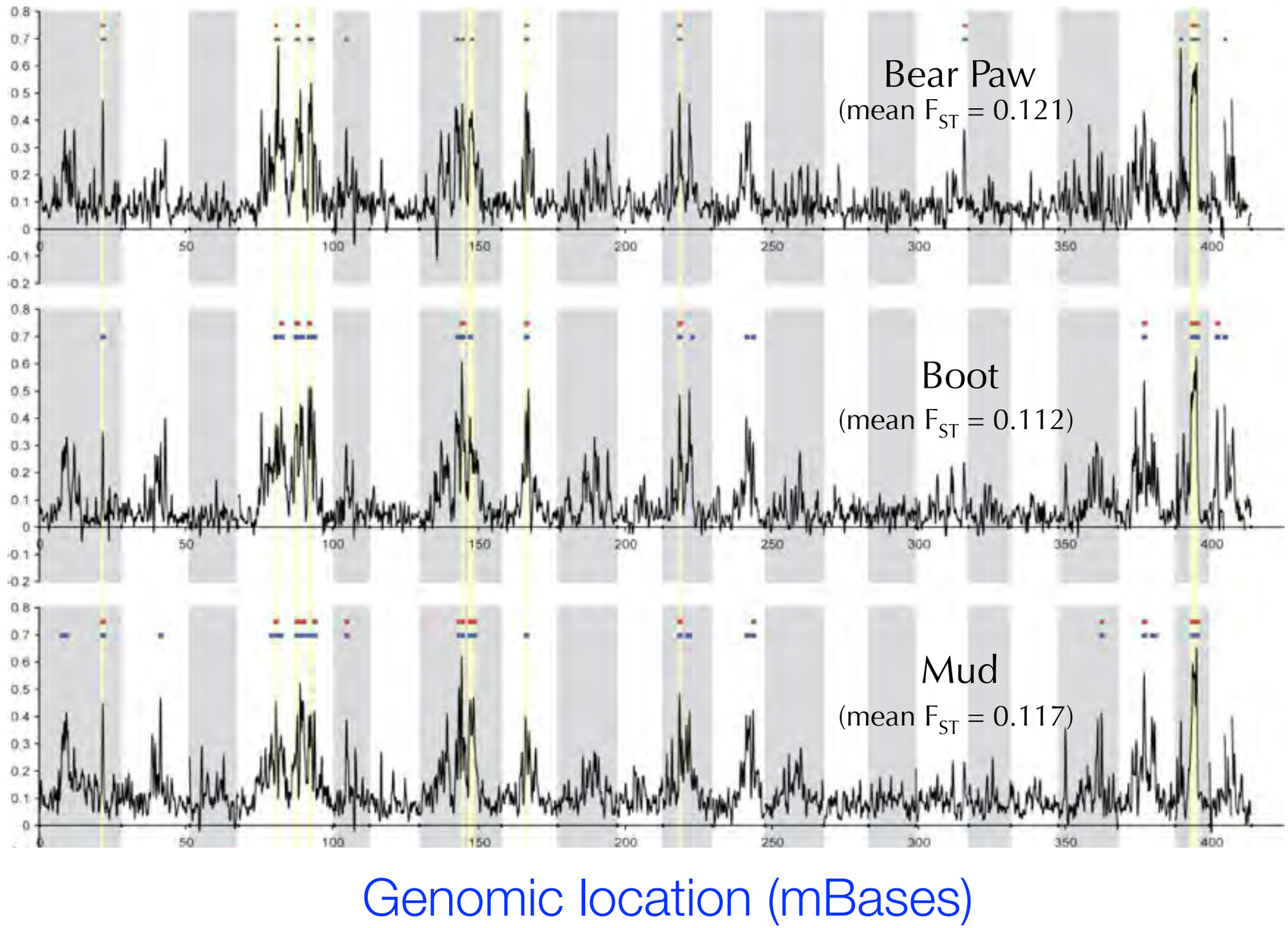
How quickly can the allele frequencies change?

Signatures of natural selection in 13,000 years



Hohenlohe, Bassham et al. 2010. PLoS Genetics

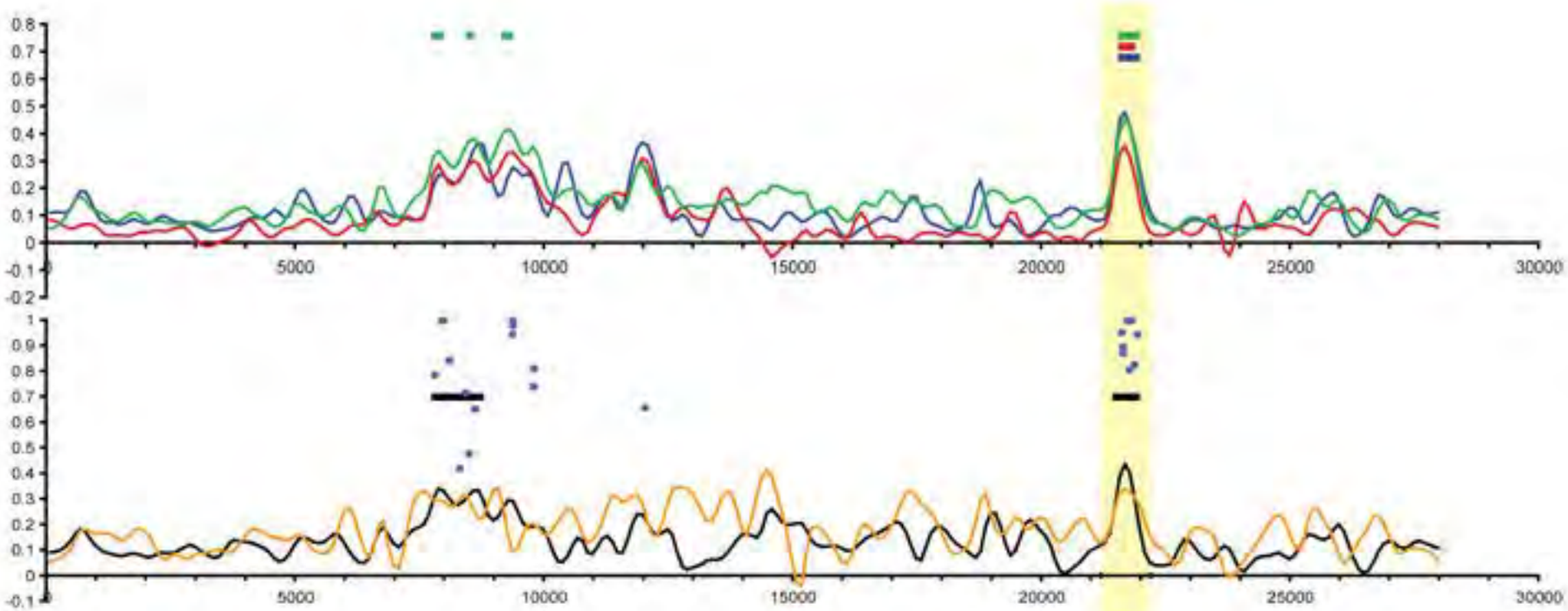
Signatures of natural selection in 13,000 years



Hohenlohe, Bassham et al. 2010. PLoS Genetics

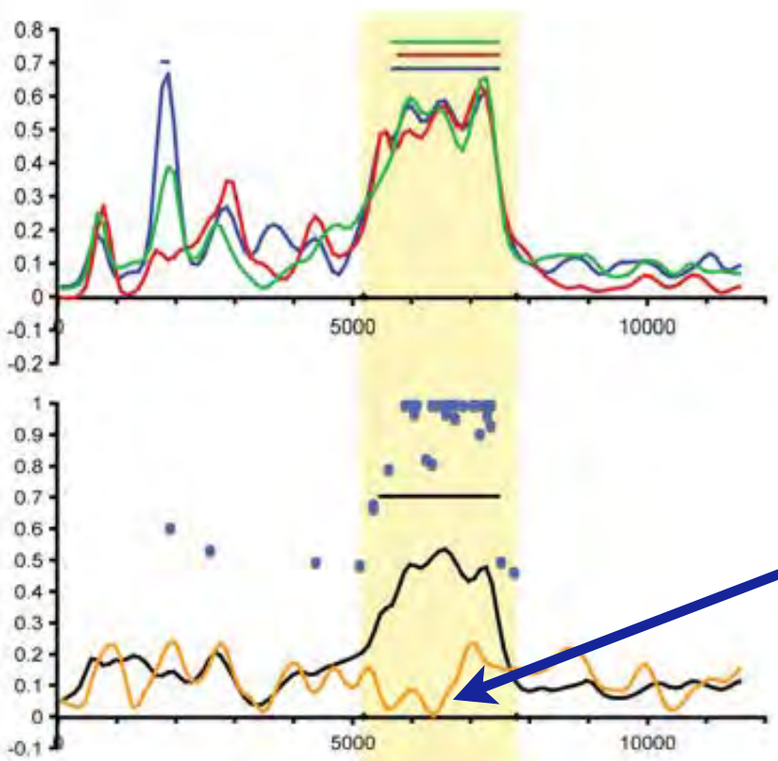
Numerous novel regions identified

LGI



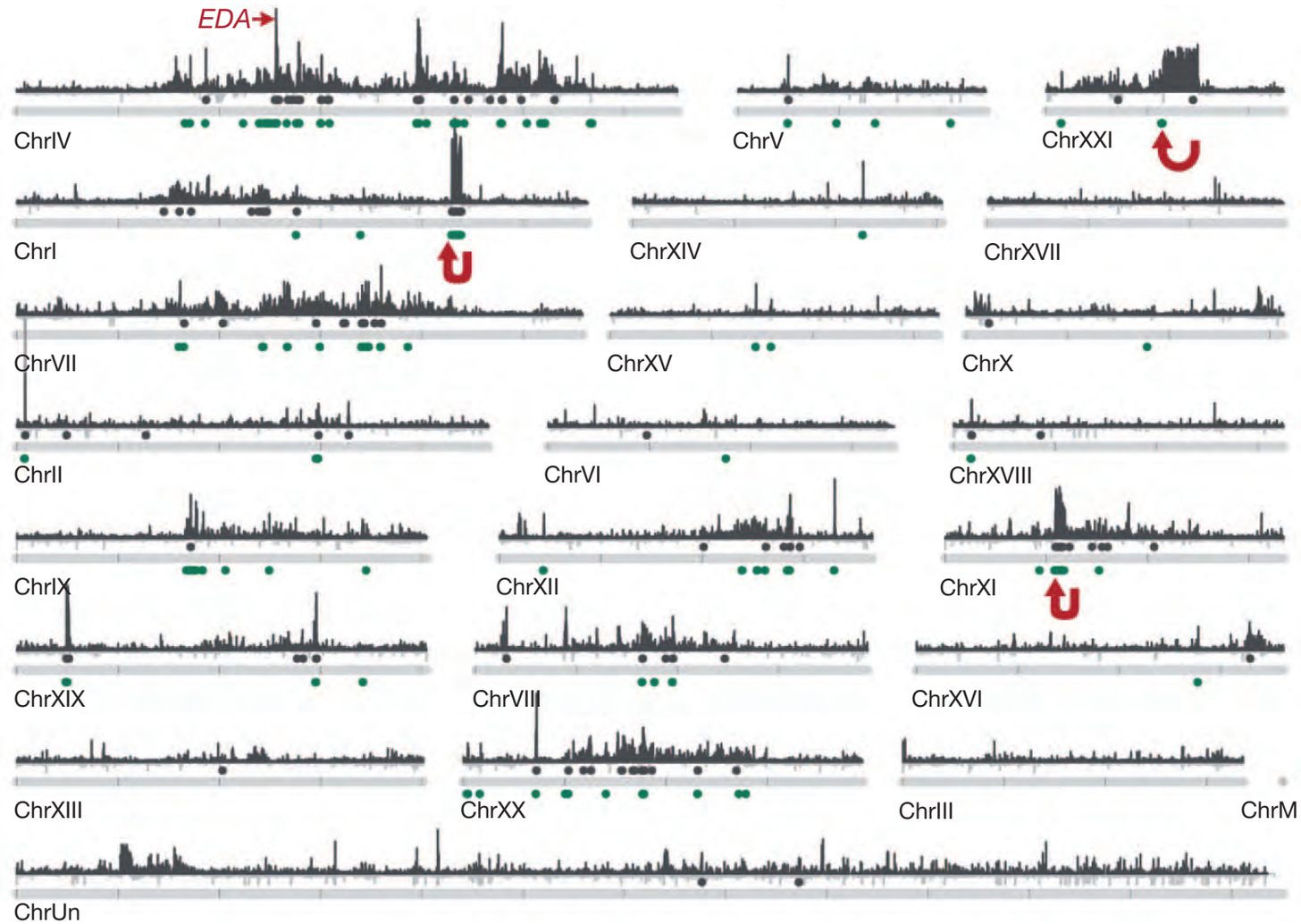
Different alleles

LGXXI



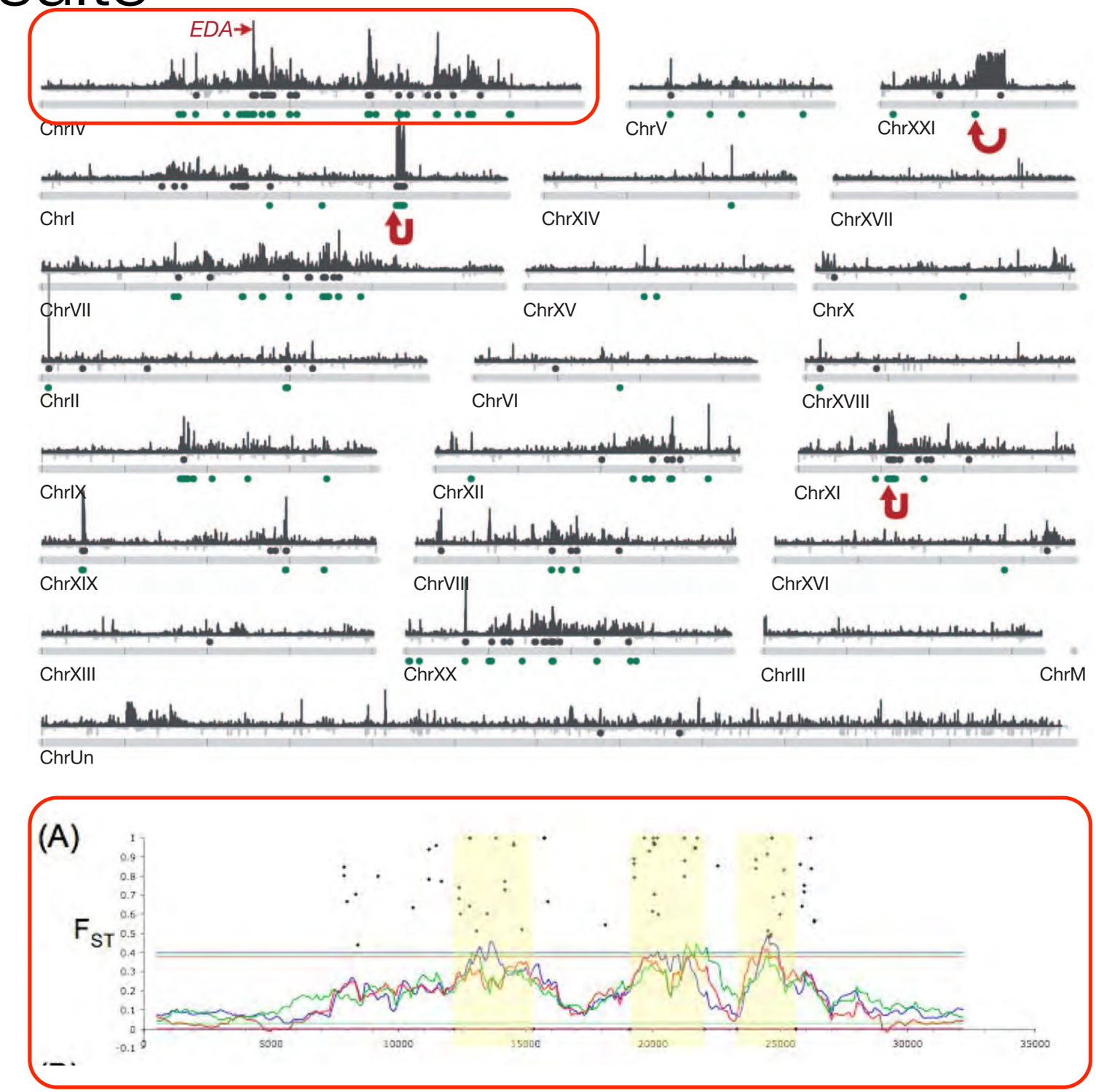
More often the same alleles

Global analysis of complete sequencing consistent with the Alaskan results



Jones et al. 2012. Nature

Global analysis of complete sequencing consistent with the Alaskan results



Jones et al. 2012. Nature

Shake rattle and evolve in 50 years

team earthquake

Susan Bassham

Emily Lescak

Mary Sherbick

Julian Catchen

Frank von Hippel

Middleton Island

1955

2008

Photo Credits: BLM, E-Terra

Middleton Island

1955

2008

Photo Credits: BLM, E-Terra

Middleton Island - 50 year old populations

1955

2008

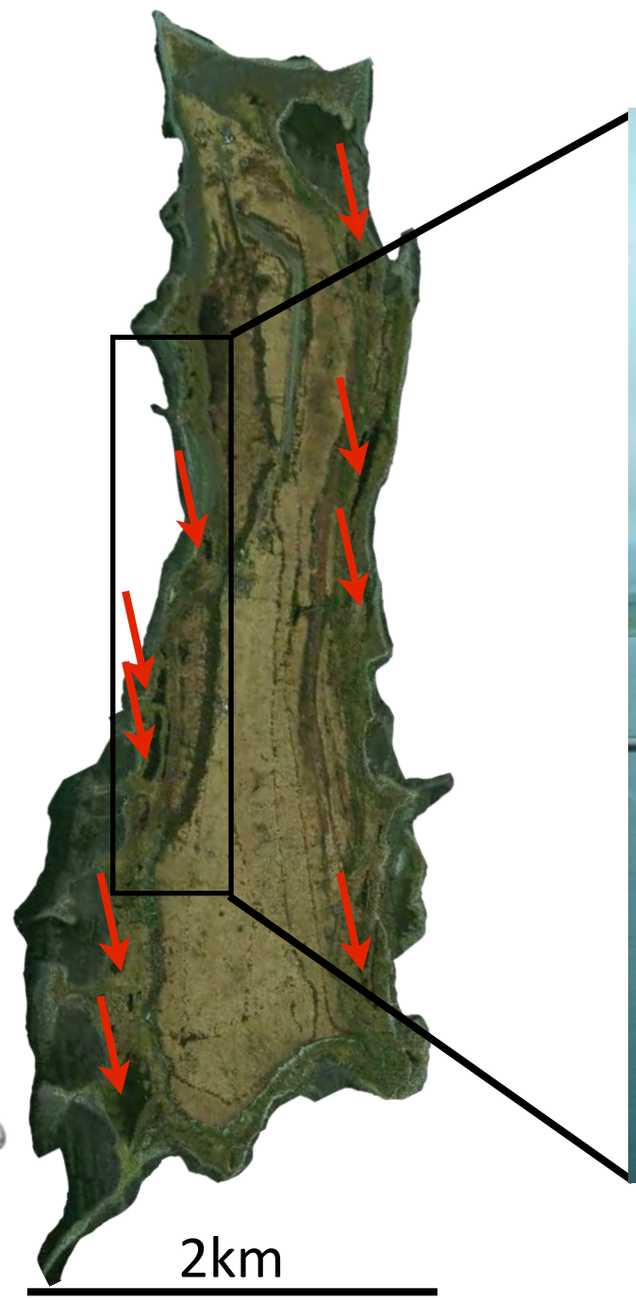
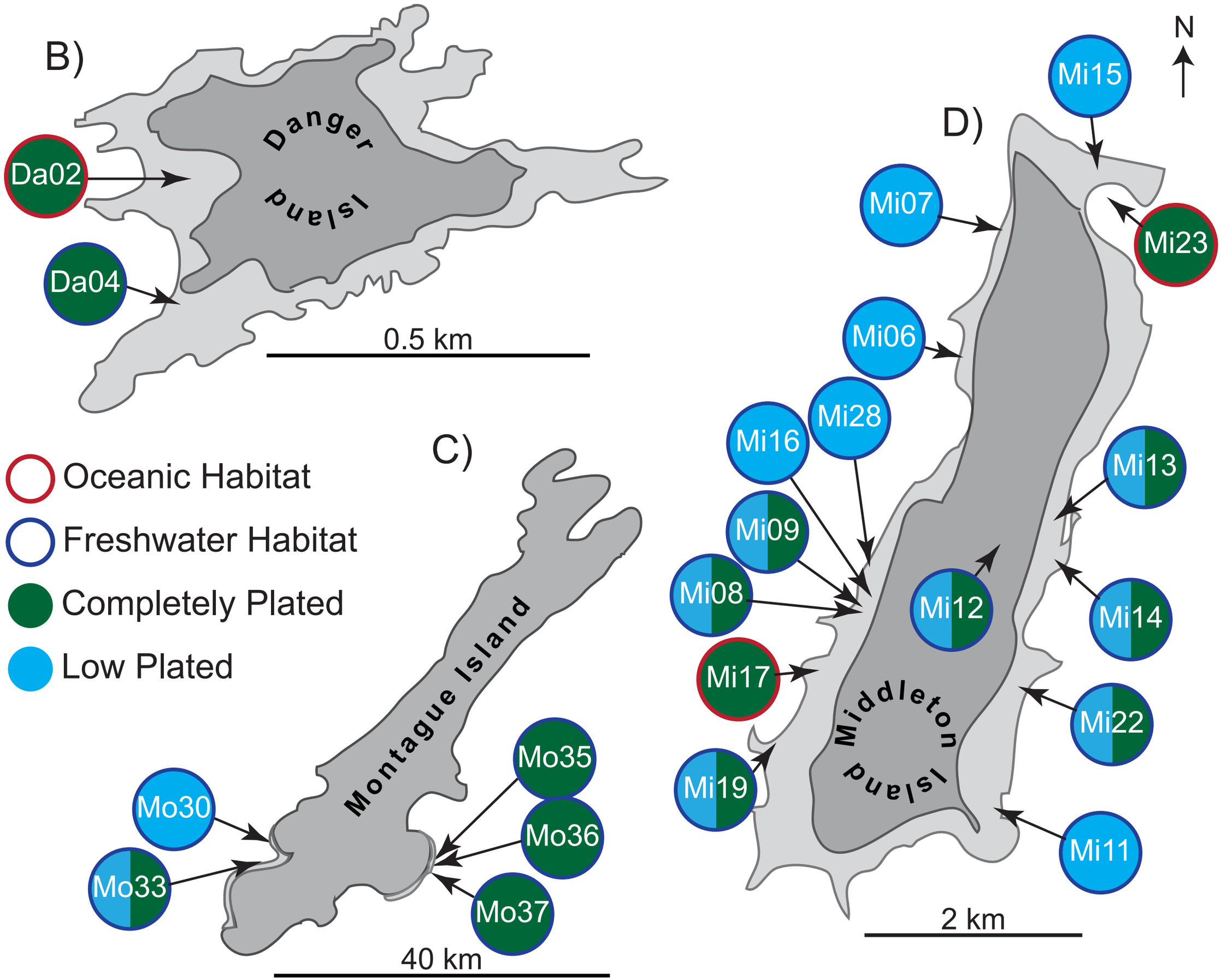


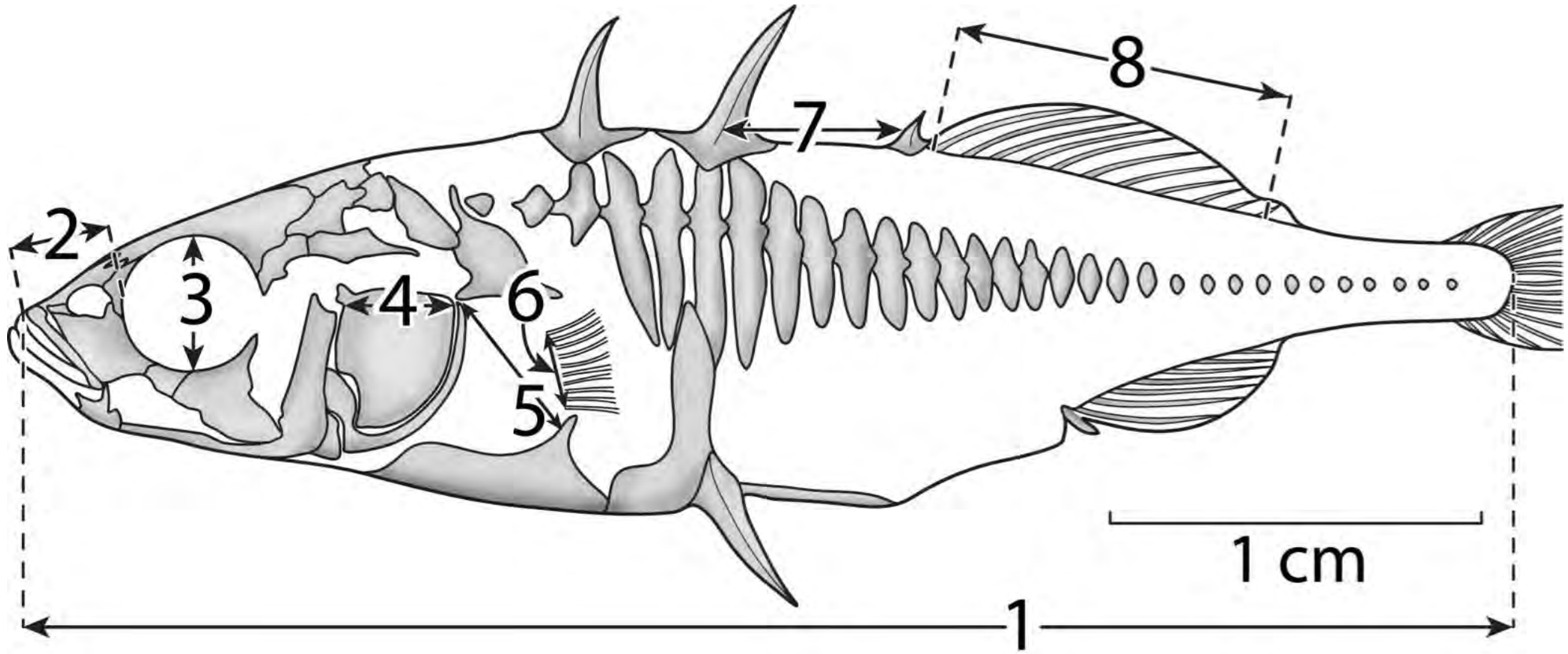
Photo Credits: BLM, E-Terra

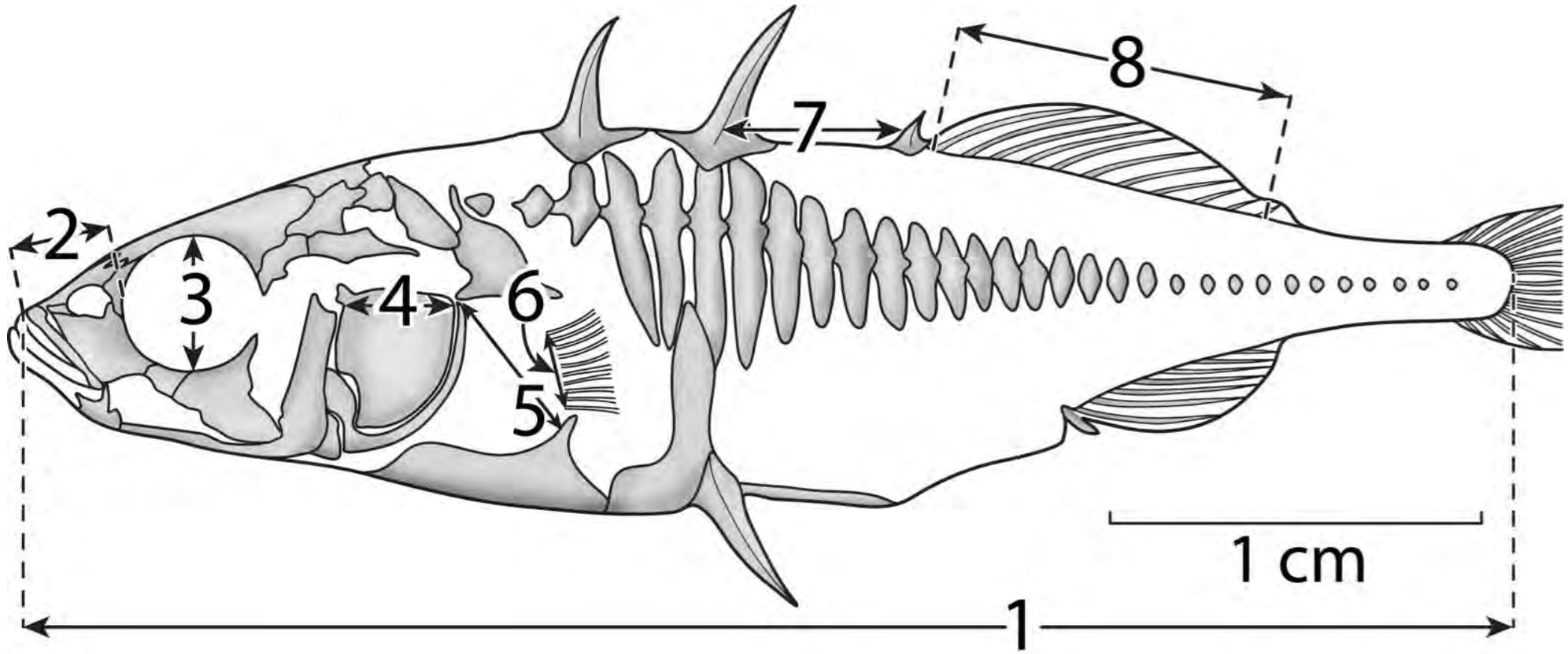

MDO SITE 13 BAR

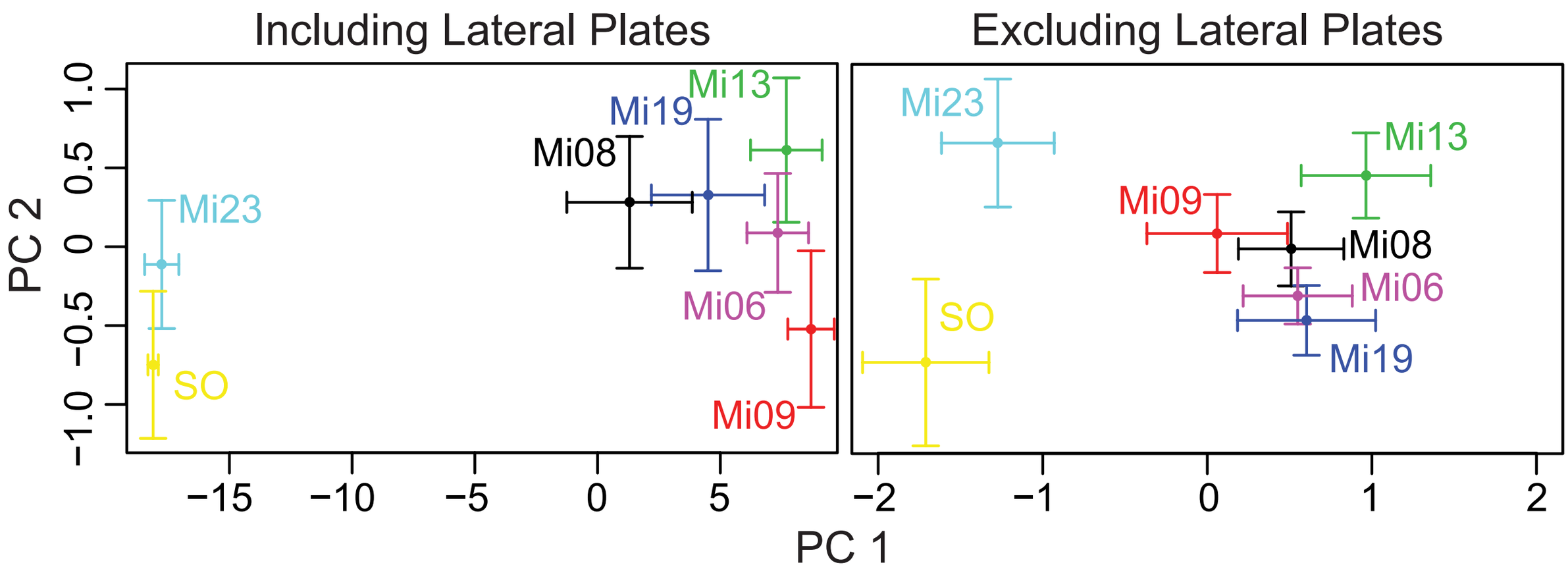

N34WM

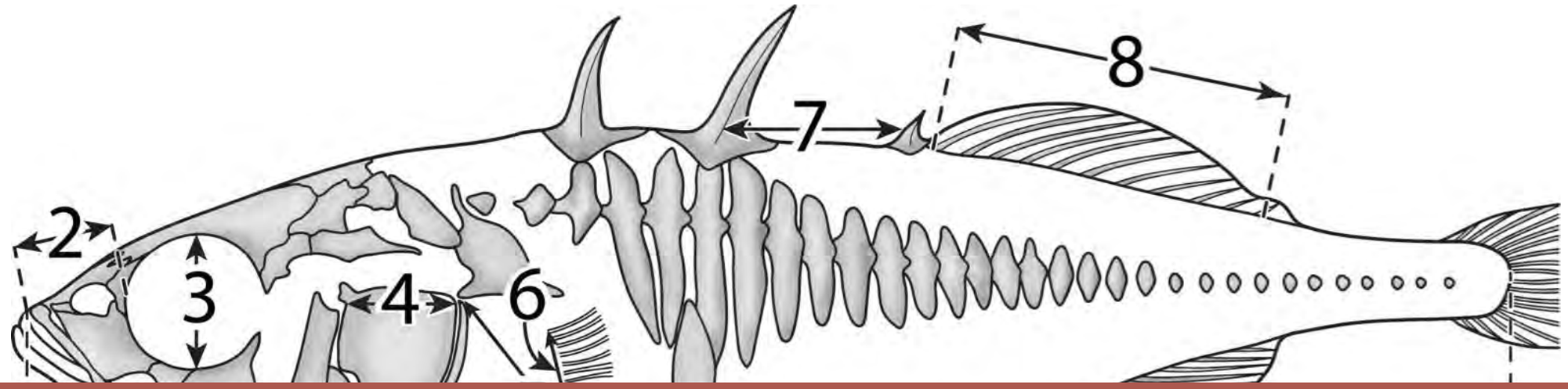

DANGER
HIGH
VOLTAGE


B)
Danger Island
Da02
Da04
0.5 km
C)
Montague Island
Mo30
Mo33
Mo35
Mo36
Mo37
40 km
Oceanic Habitat
Freshwater Habitat
Completely Plated
Low Plated
D)
N
Mi15
Mi07
Mi23
Mi06
Mi28
Mi16
Mi13
Mi09
Mi08
Mi12
Mi14
Mi17
Mi22
Mi19
Mi11
Middleton Island
2 km


8
7
2
3
4
6
5
1 cm
1

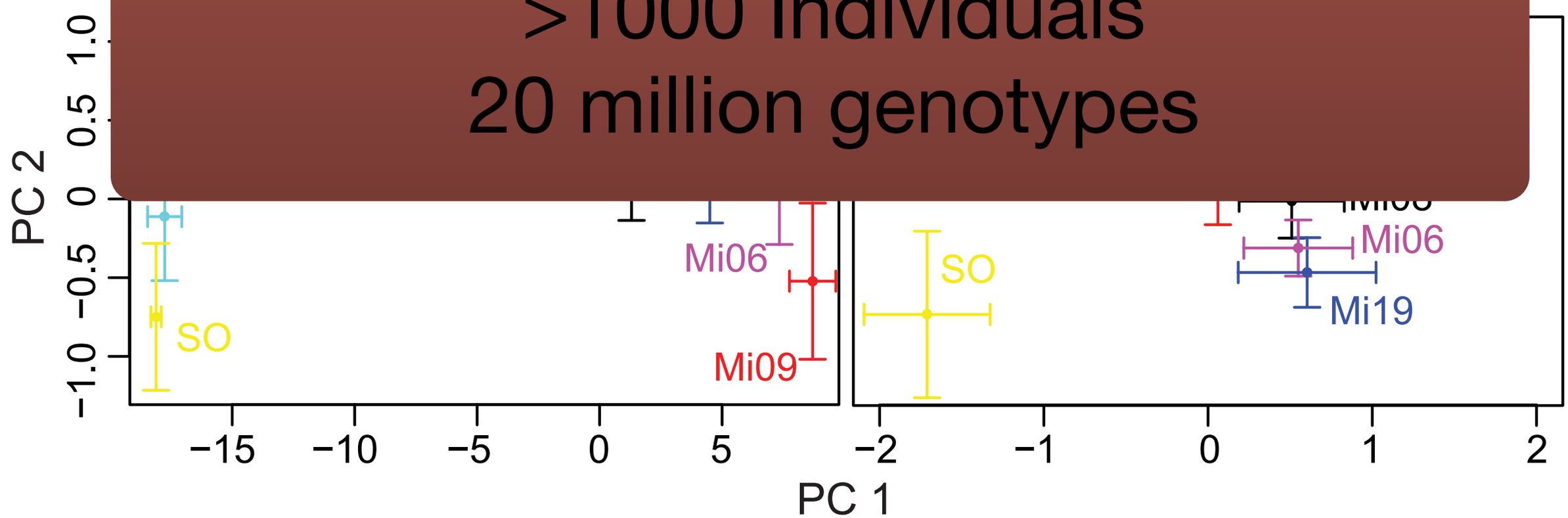

8
7
2
3
4
6
5
1 cm
1


Including Lateral Plates
Excluding Lateral Plates
PC 2
PC 1
1.0
0.5
0
−0.5
−1.0
−15
−10
−5
0
5
−2
−1
0
1
2
Mi13
Mi19
Mi08
Mi23
Mi06
SO
Mi09
Mi23
Mi13
Mi09
Mi08
Mi06
SO
Mi19

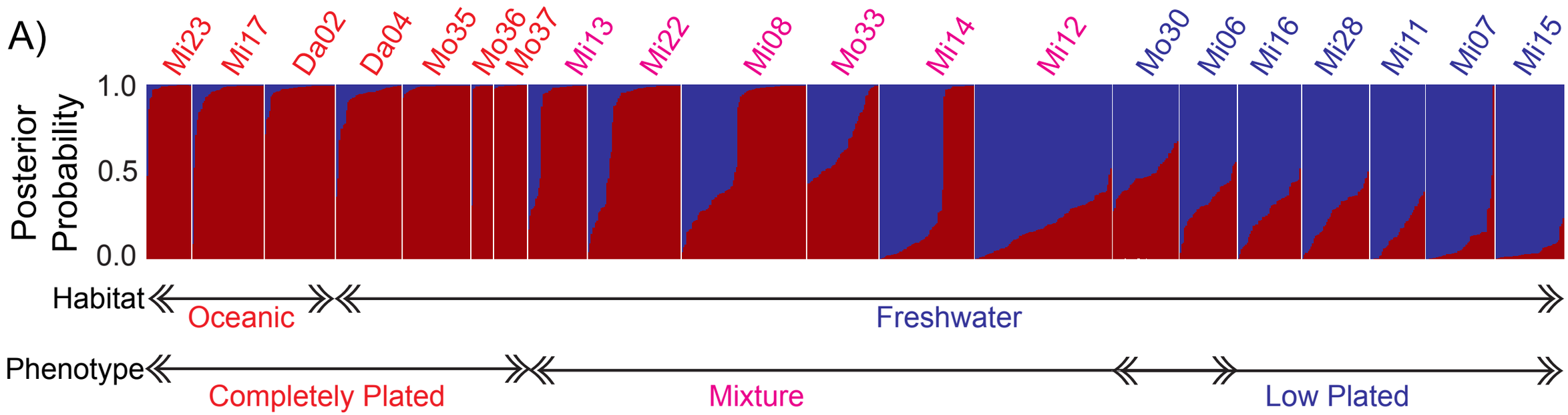


RAD-seq analysis

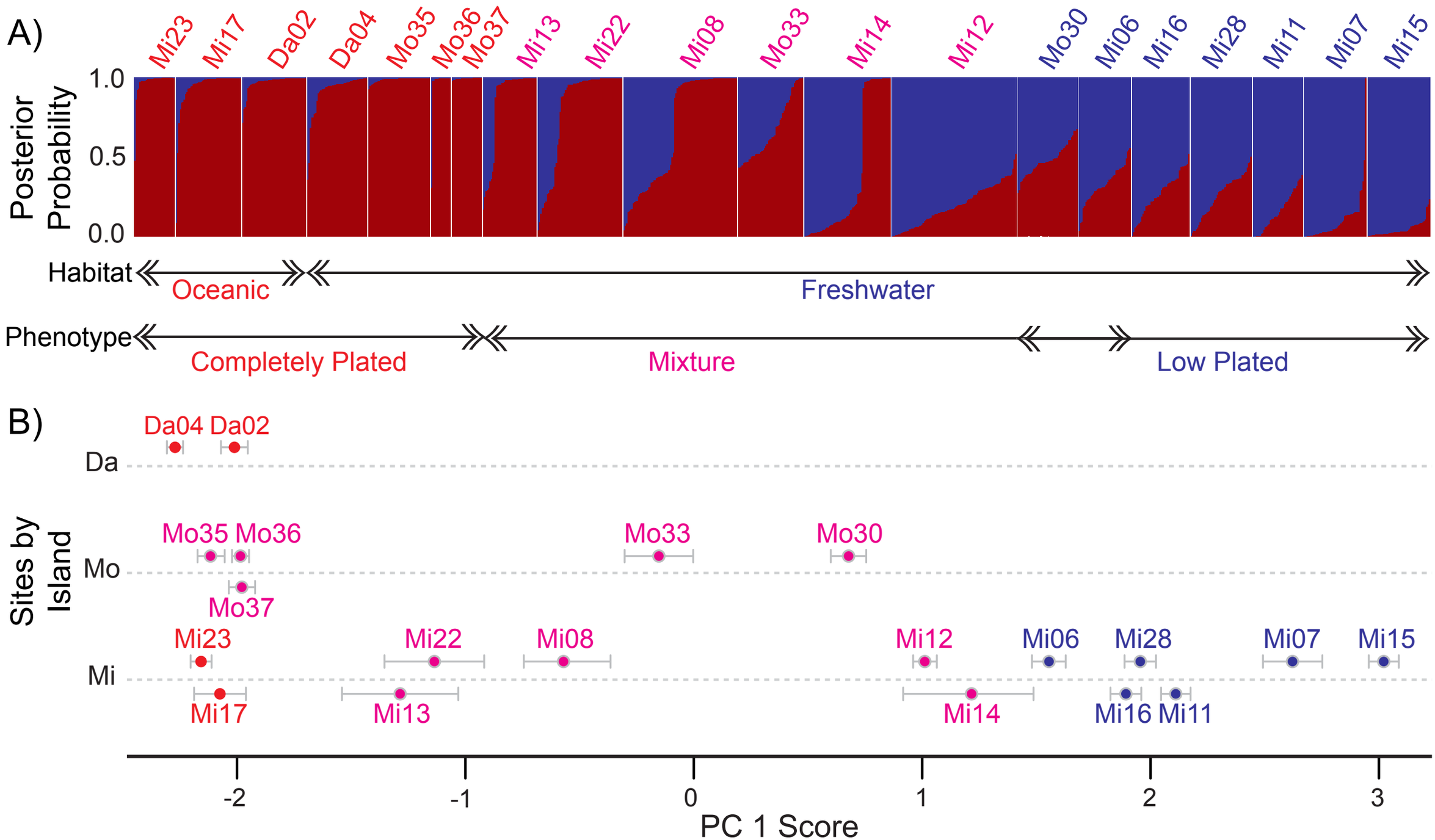
110,000 SNPs per individual
>1000 Individuals
20 million genotypes



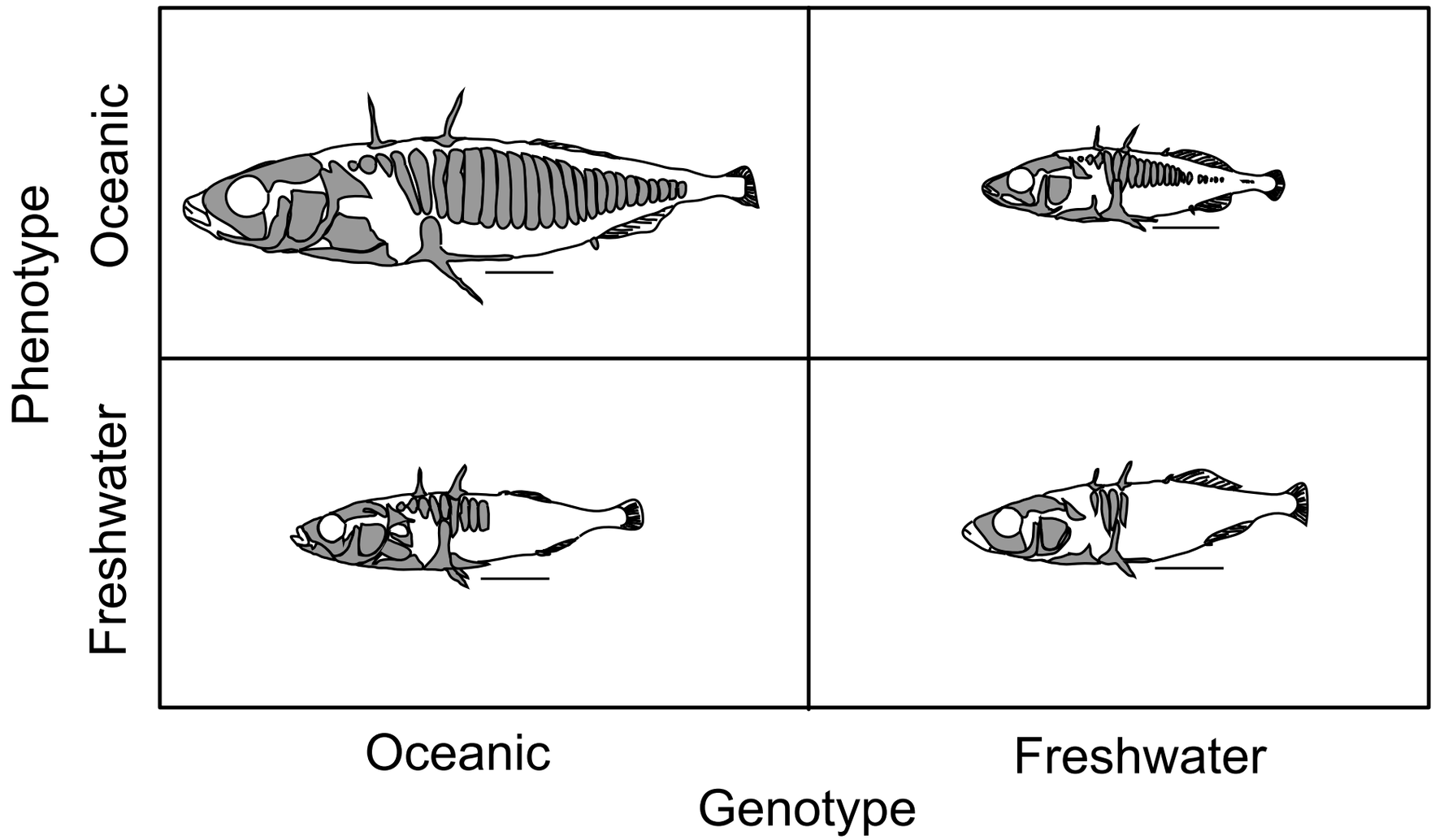
The majority of the genetic variation is partitioned between oceanic and freshwater fish



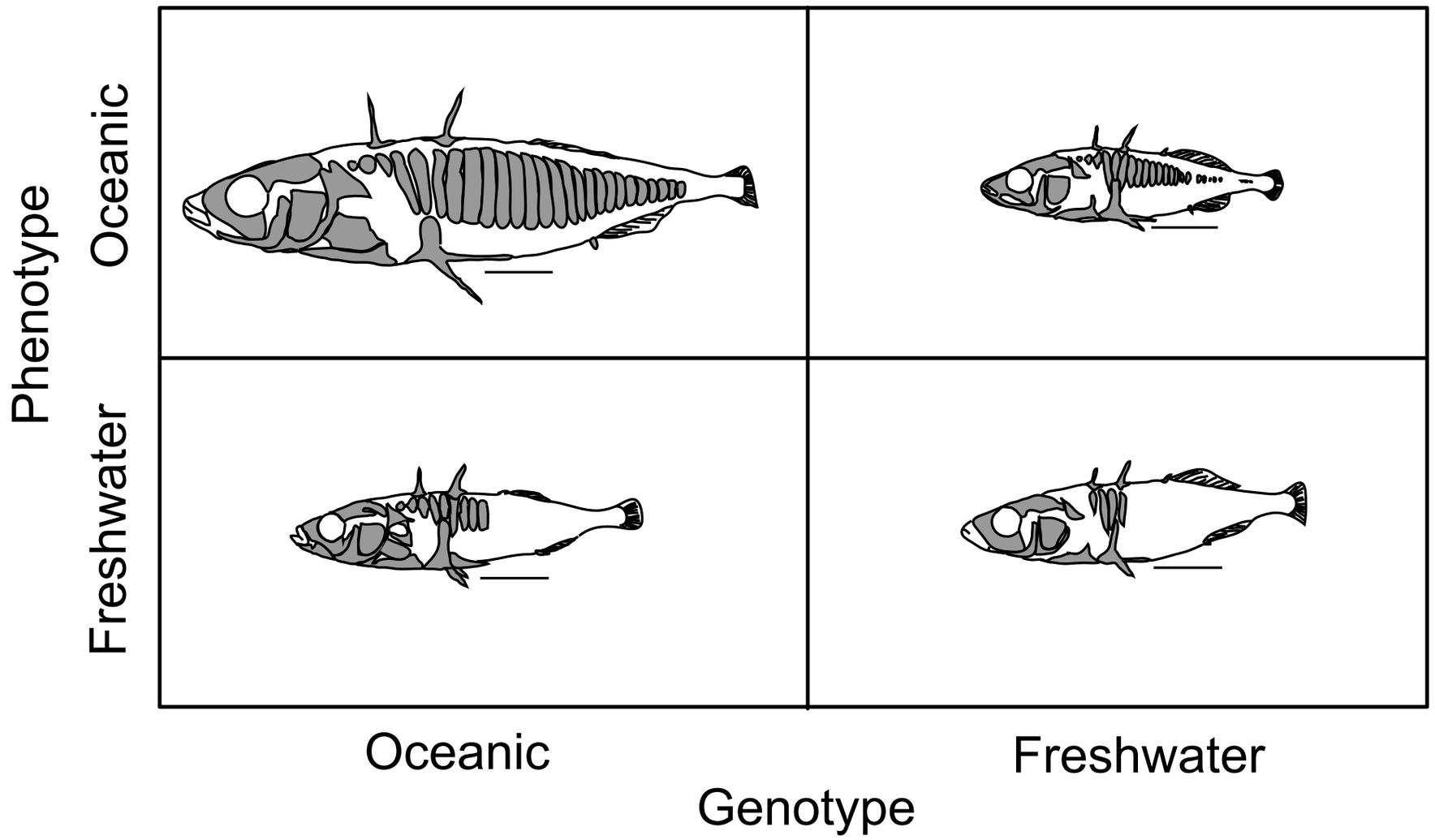
The majority of the genetic variation is partitioned between oceanic and freshwater fish

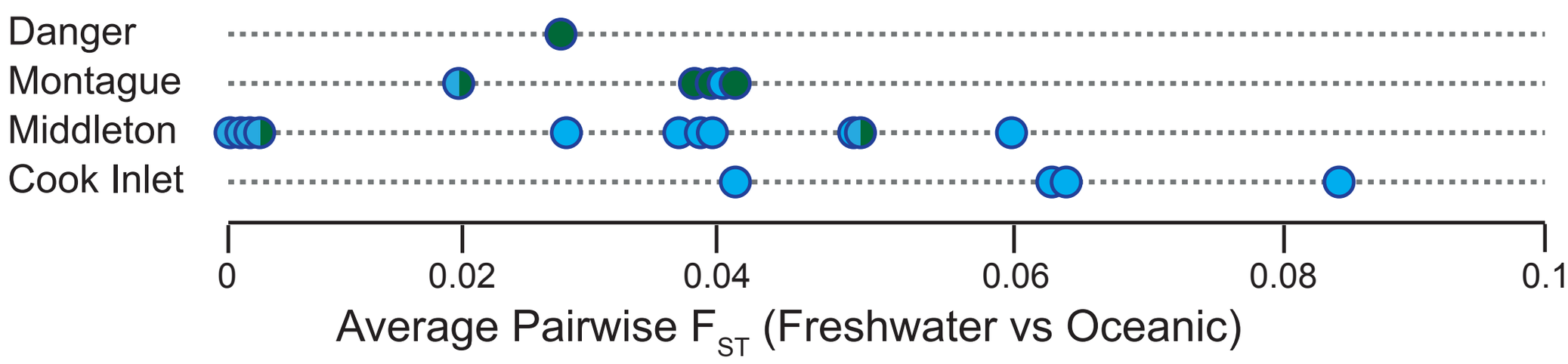


Extensive phenotypic change occurs even in populations with little overall differentiation

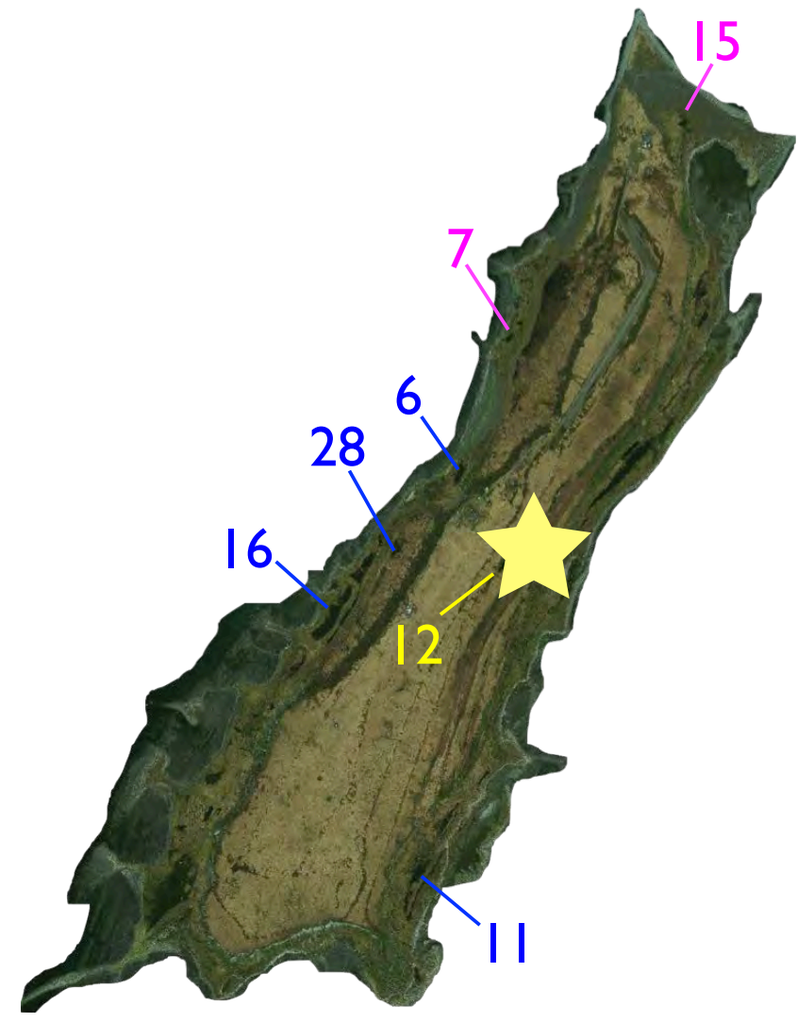


Extensive phenotypic change occurs even in populations with little overall differentiation

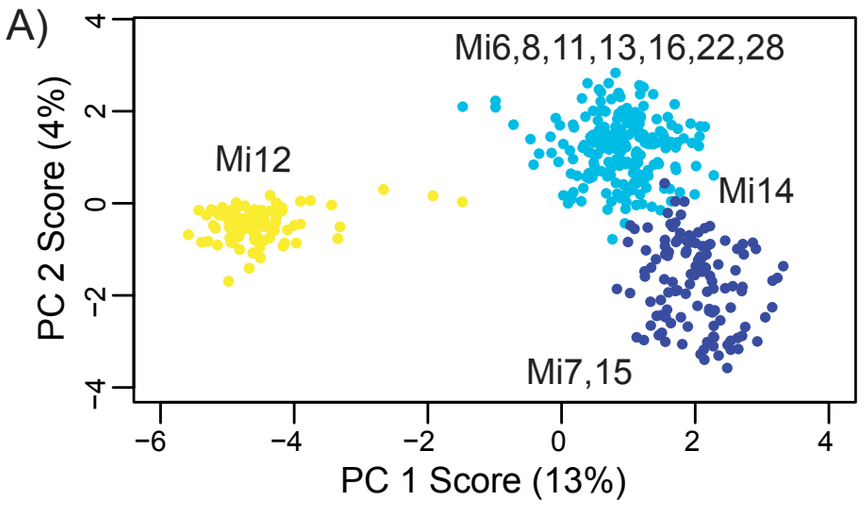


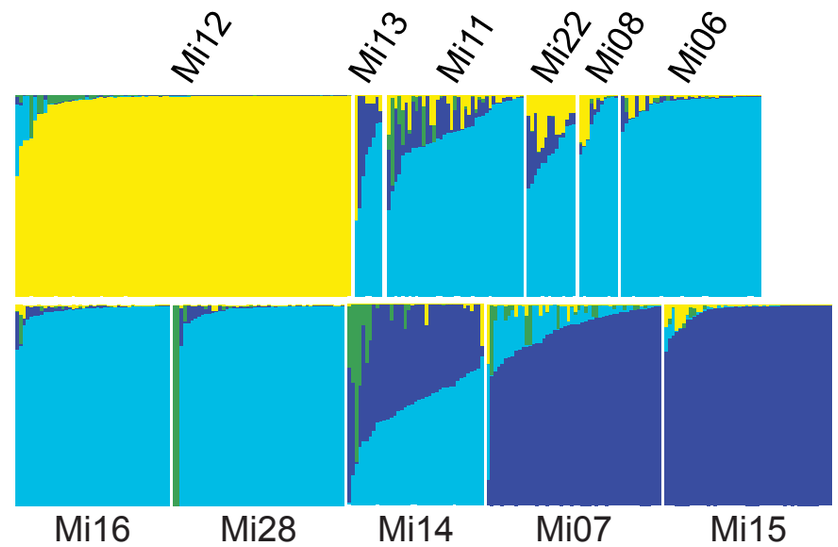


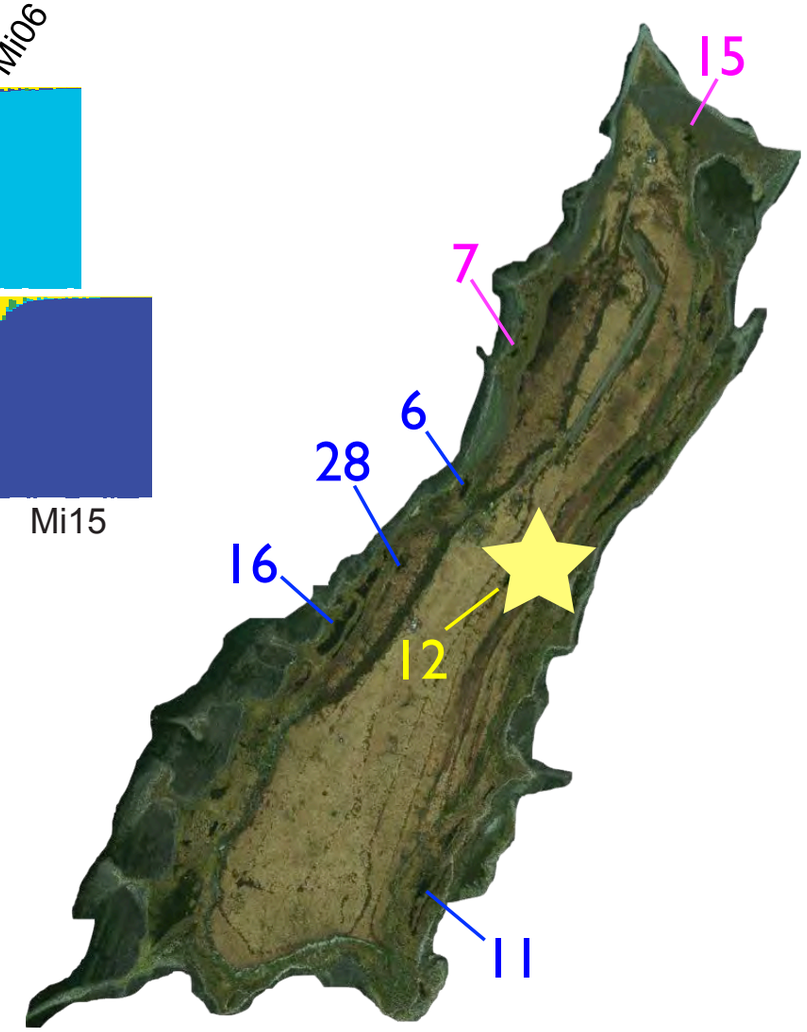
Structure analysis shows independent evolution even among populations on a single island



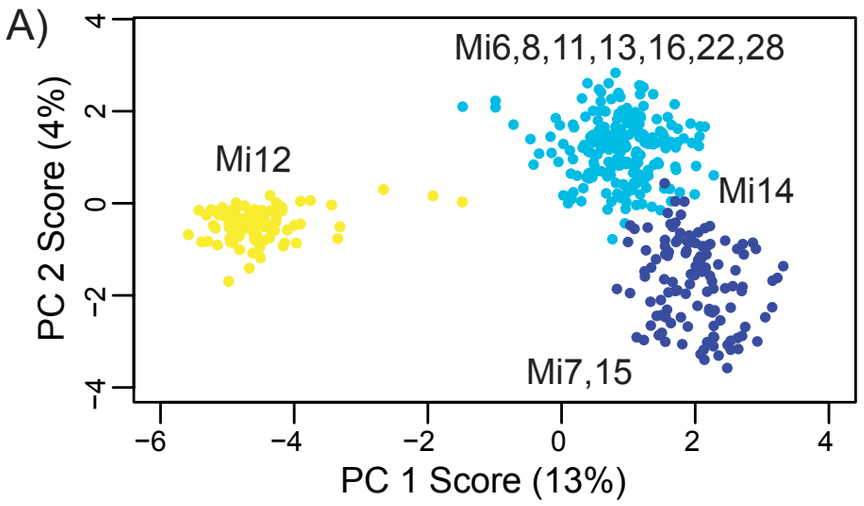
Structure analysis shows independent evolution even among populations on a single island

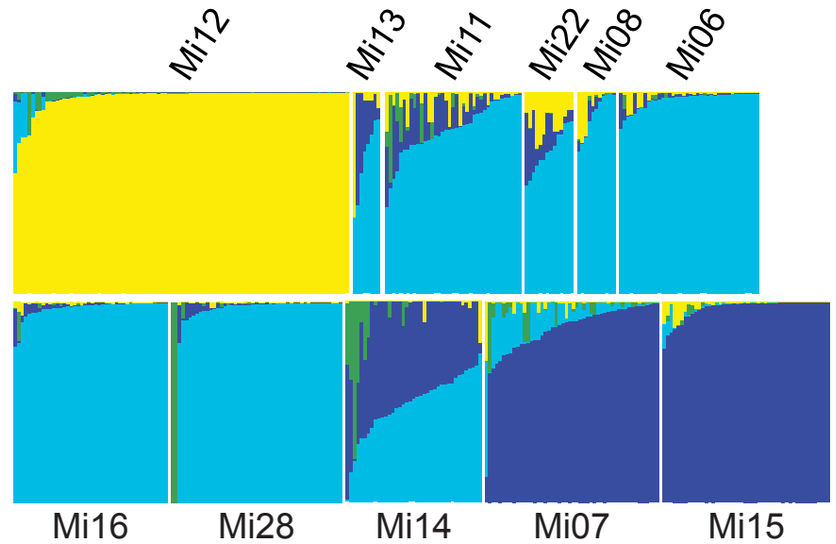


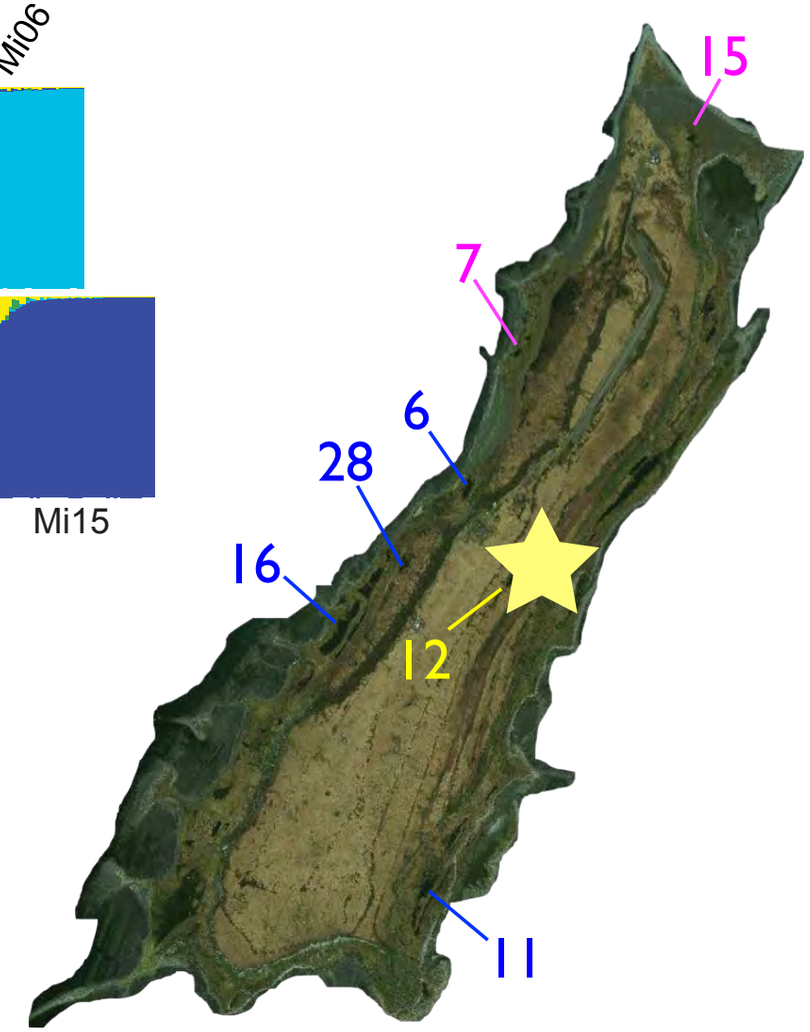


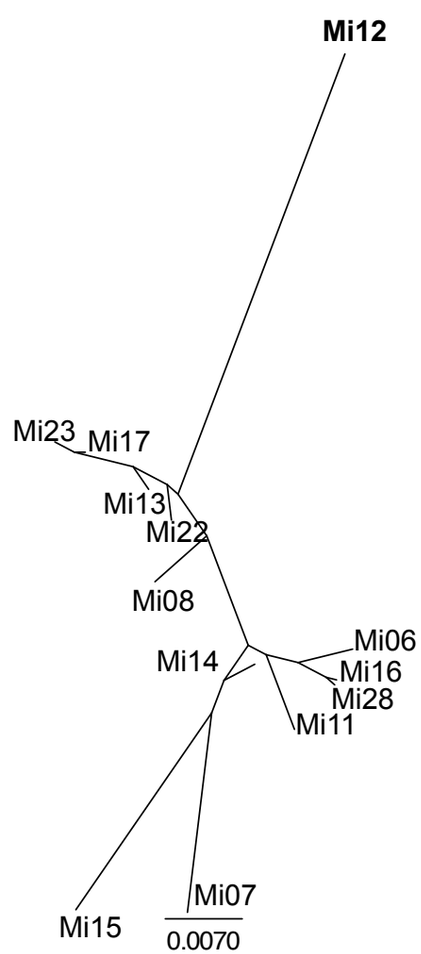


Structure analysis shows independent evolution even among populations on a single island

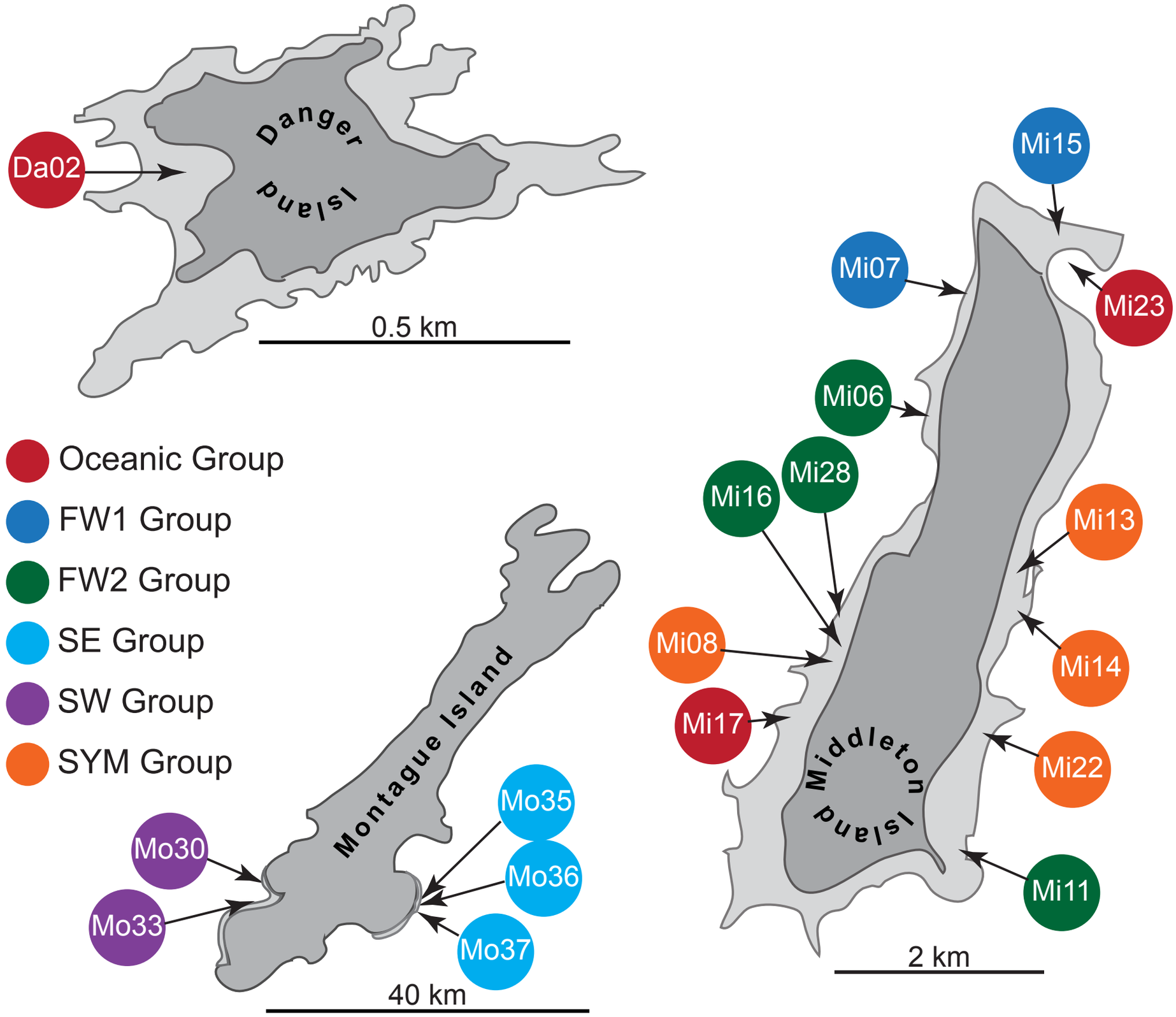








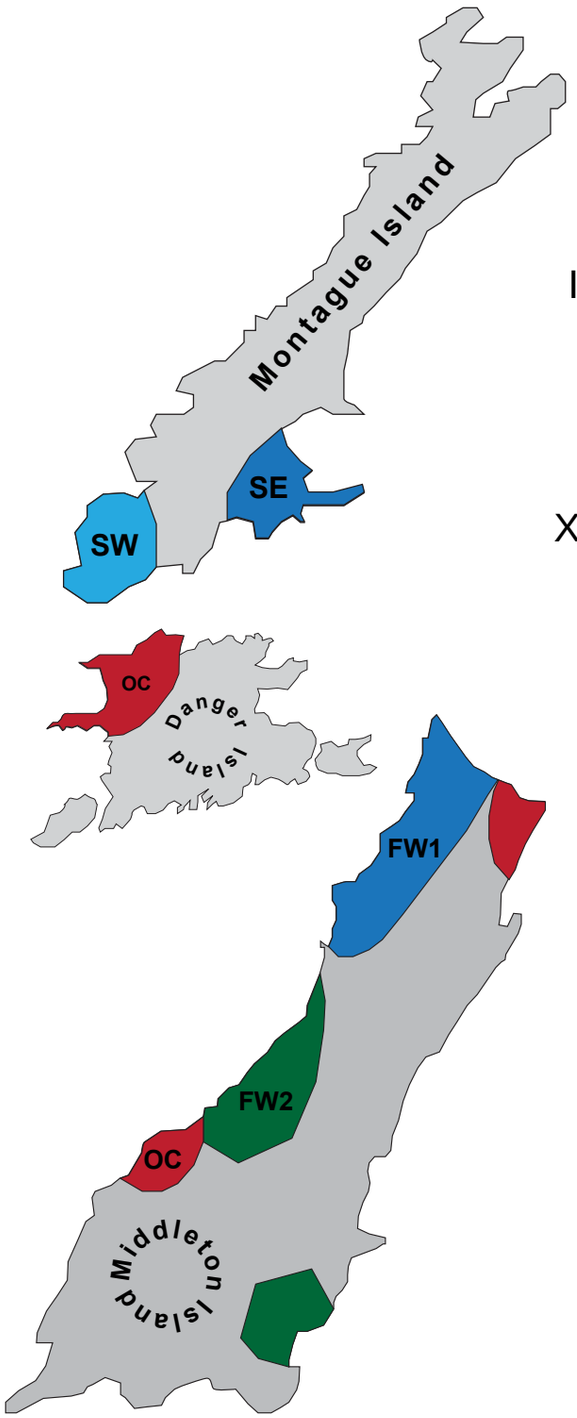
At least six independent evolutionary events in freshwater in the last 50 years

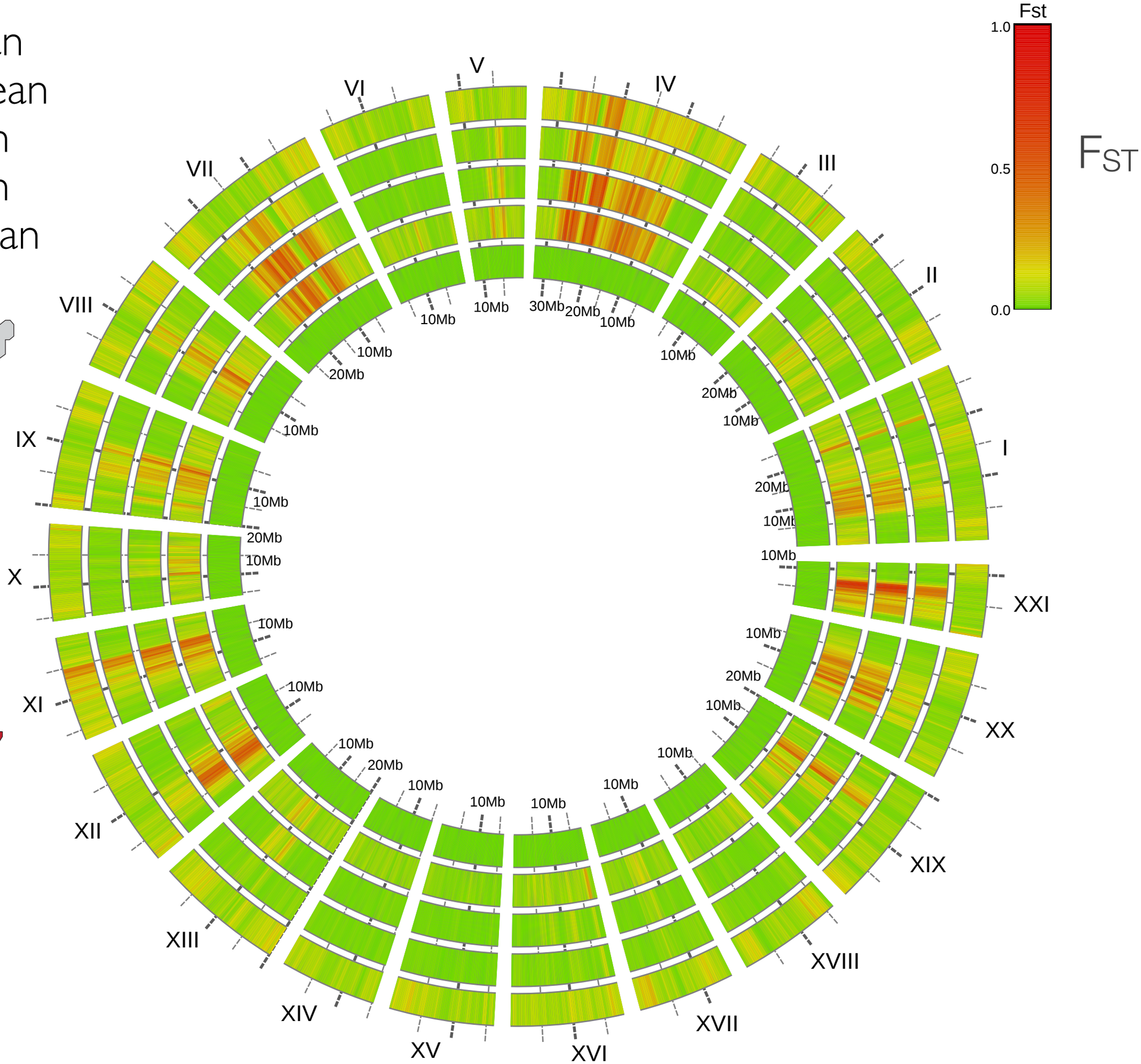


How much of the genome is differentiated?

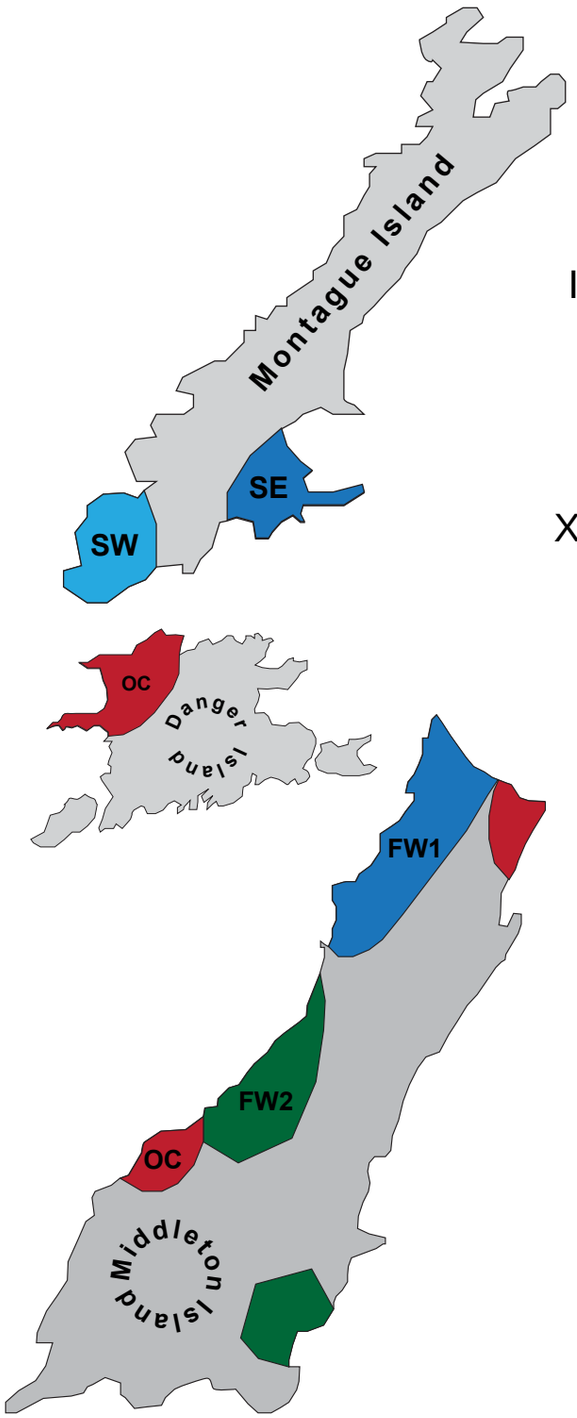
How similar are the genomic patterns of differentiation?

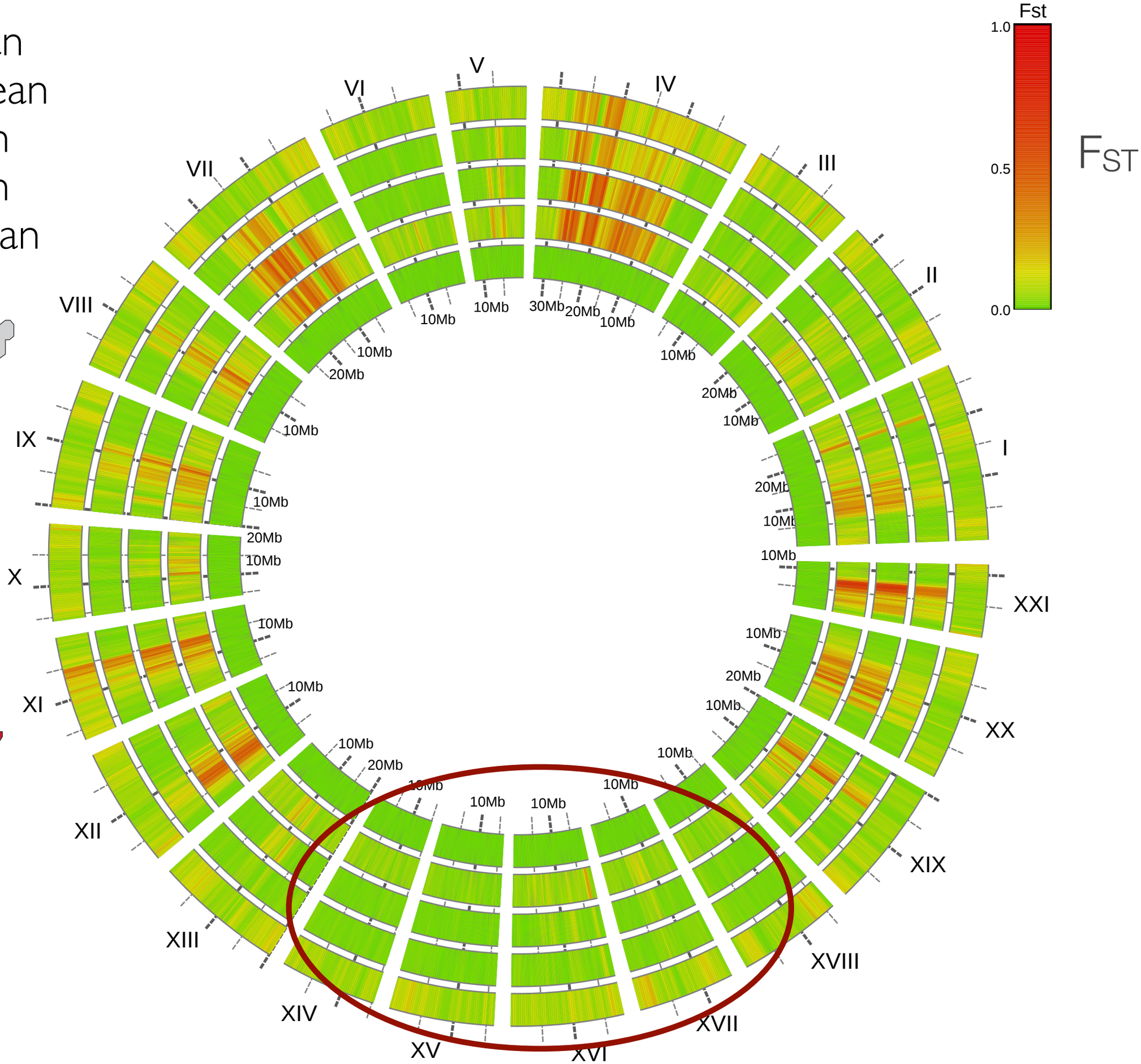
MoSe vs. Ocean
MoSW vs. Ocean
FW2 vs. Ocean
FW1 vs. Ocean
Ocean vs. Ocean



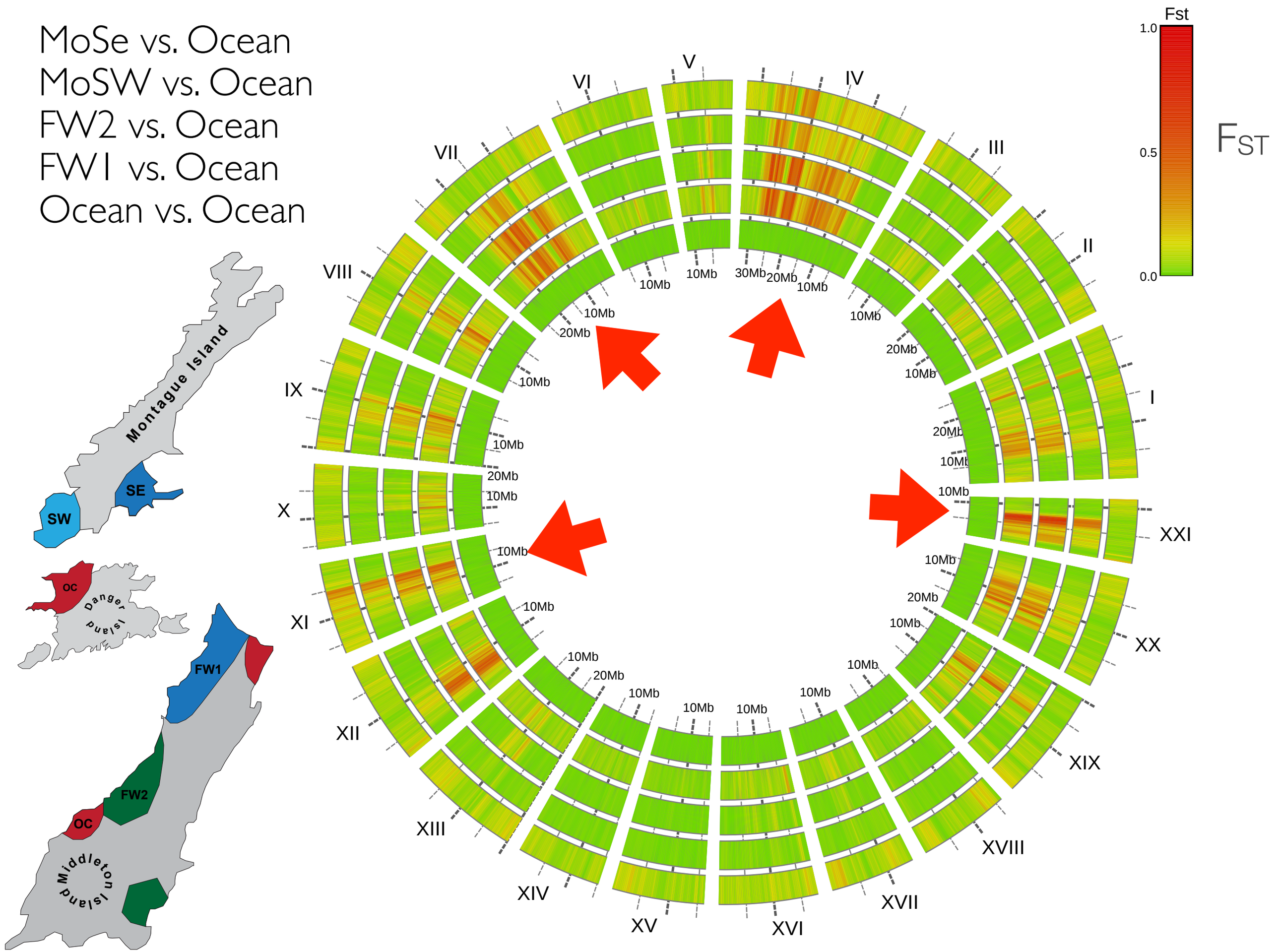


MoSe vs. Ocean
MoSW vs. Ocean
FW2 vs. Ocean
FW1 vs. Ocean
Ocean vs. Ocean



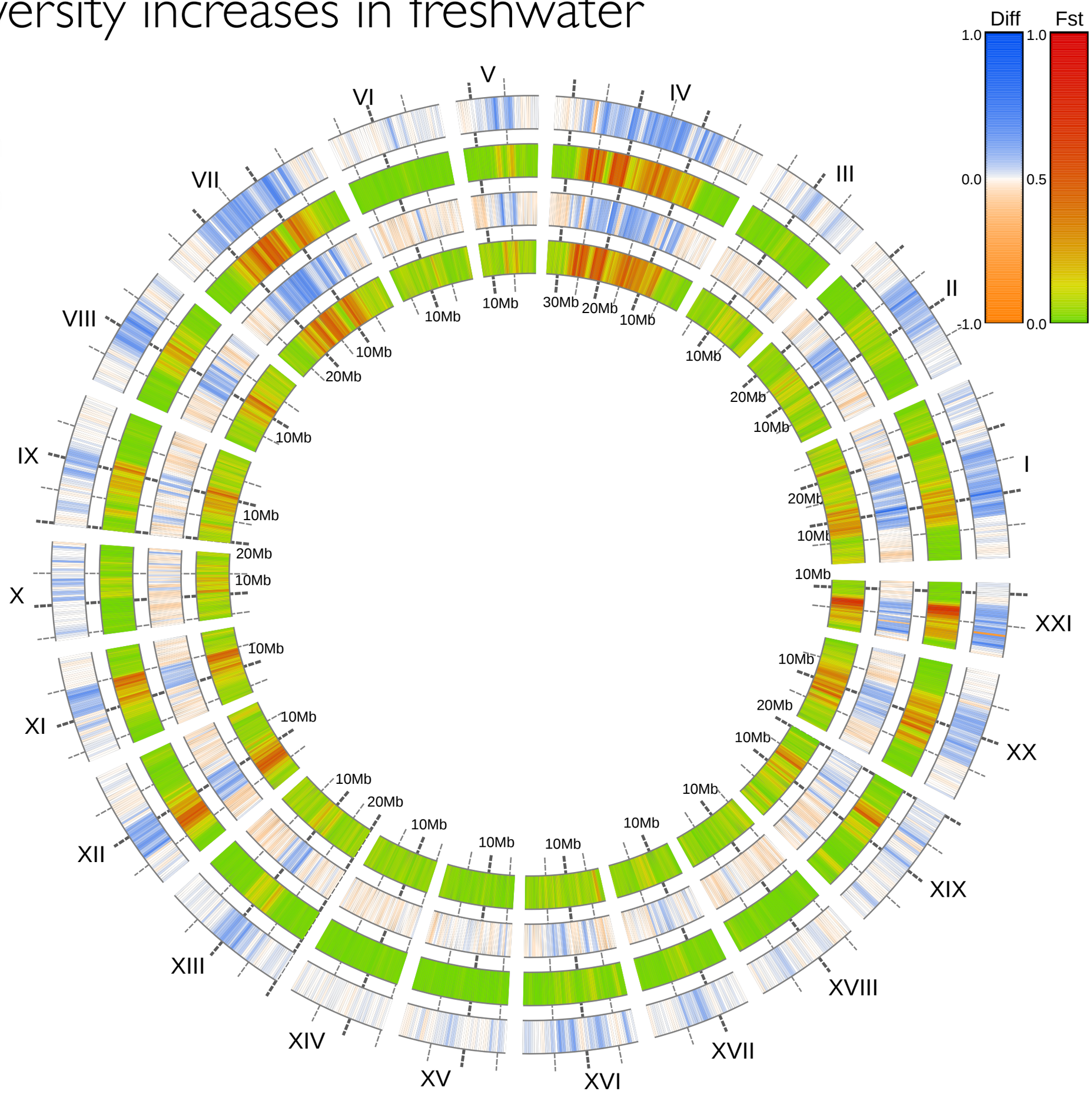


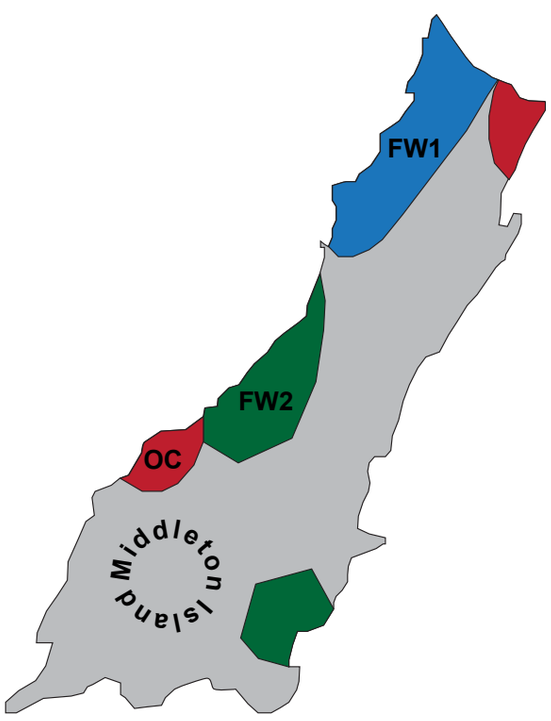
MoSe vs. Ocean
MoSW vs. Ocean
FW2 vs. Ocean
FW1 vs. Ocean
Ocean vs. Ocean



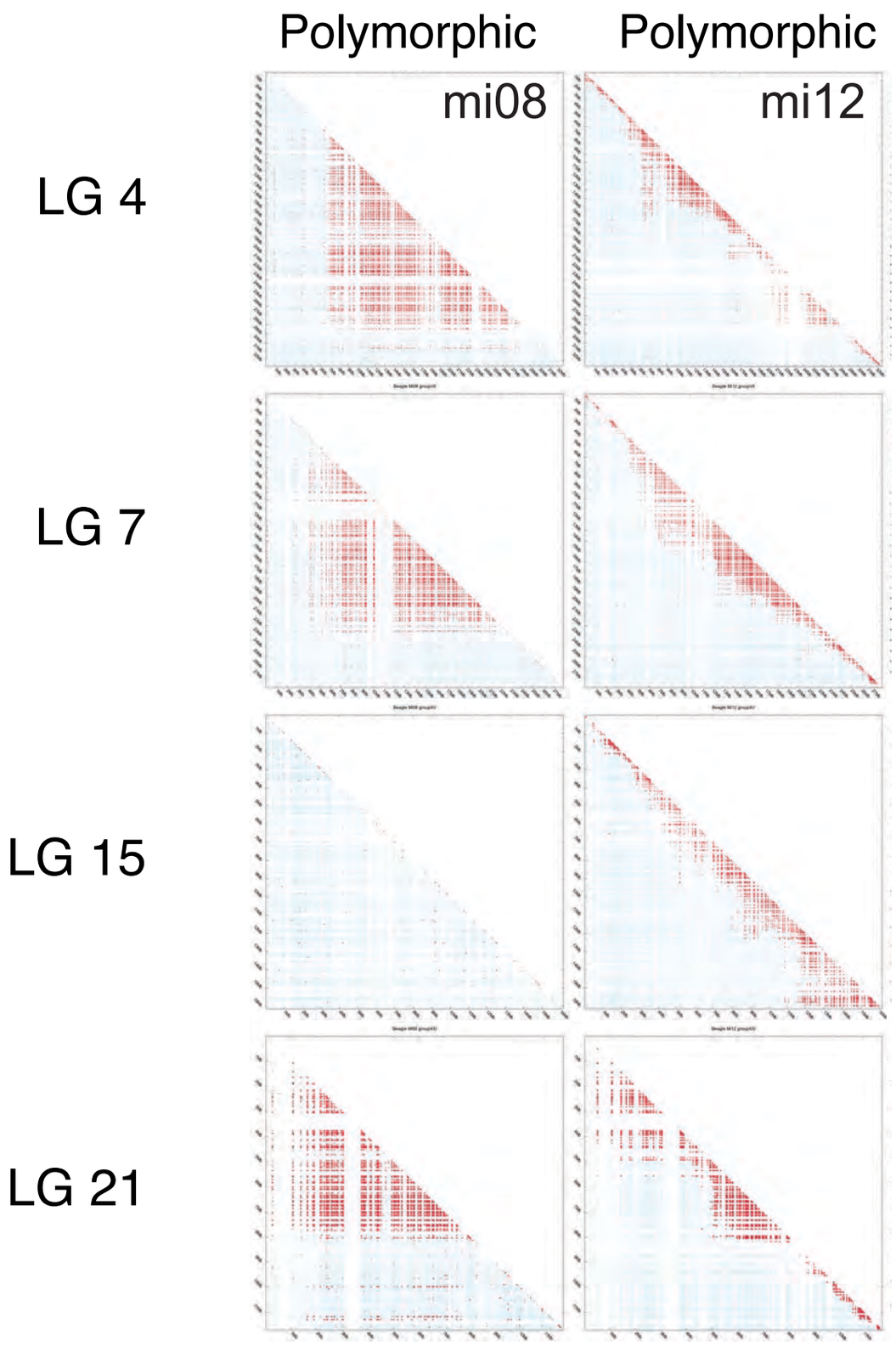
Haplotype diversity increases in freshwater

FW1 vs. Ocean
FW2 vs. Ocean

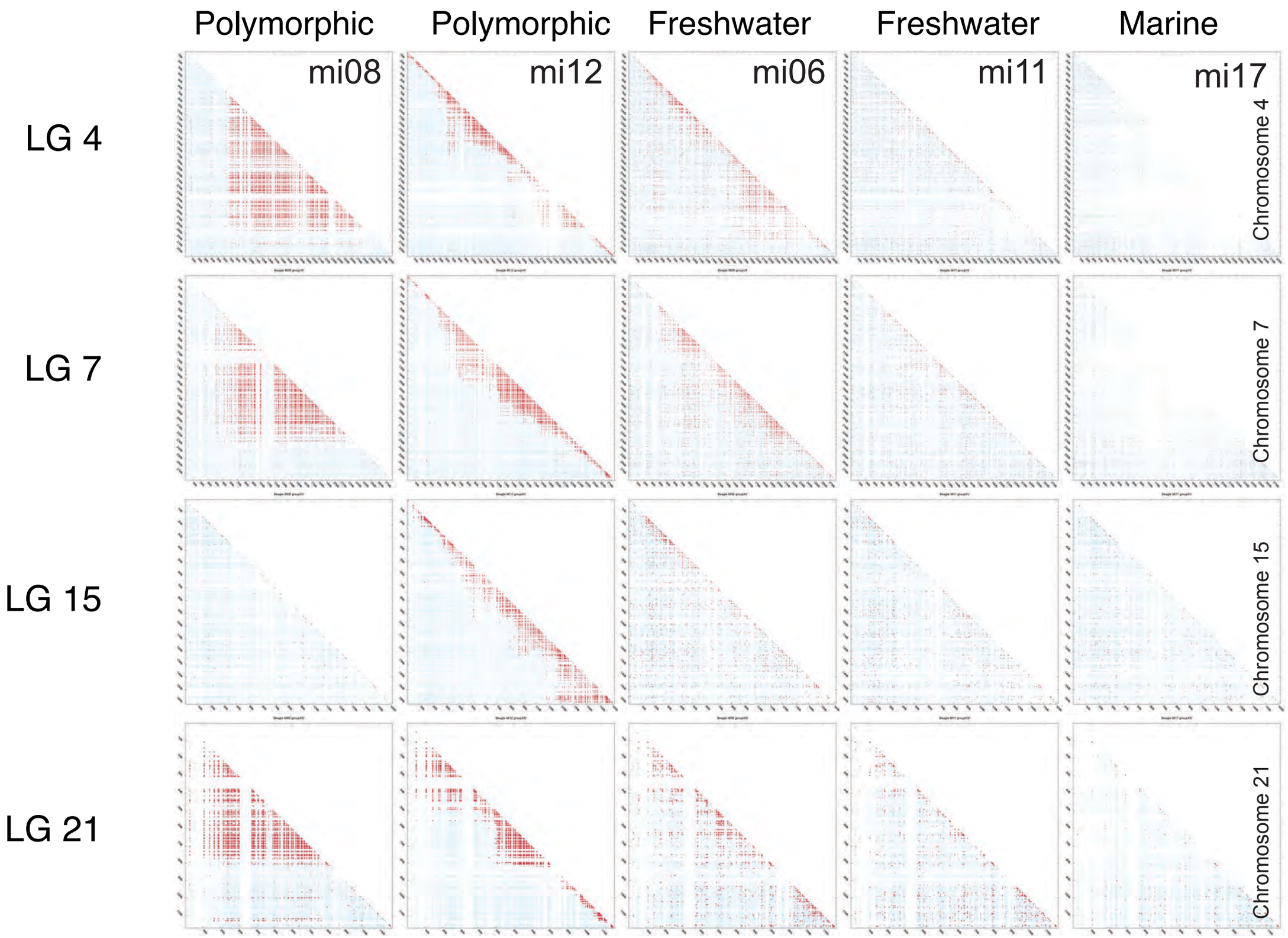




Linkage disequilibrium is extensive between oceanic and freshwater stickleback, but not within

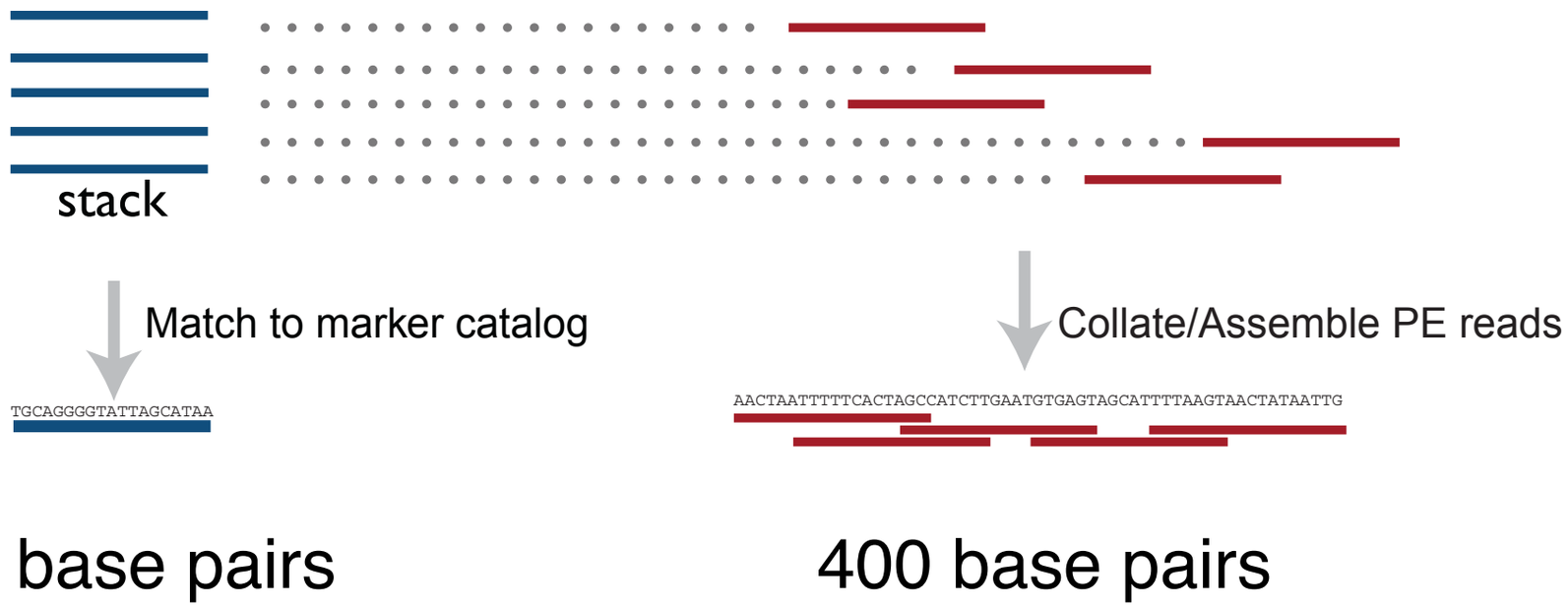


Linkage disequilibrium is extensive between oceanic and freshwater stickleback, but not within



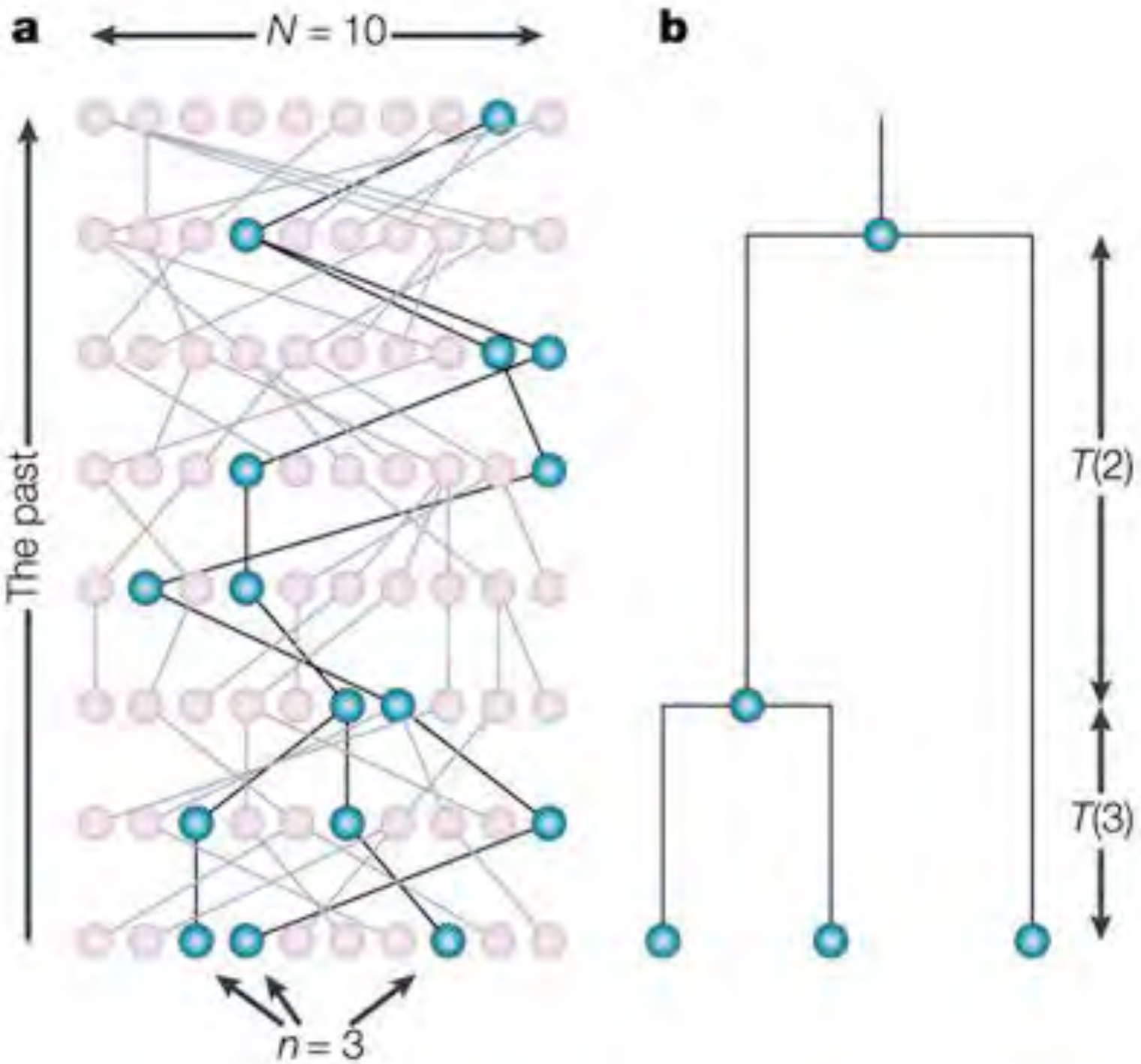
Are these genomic blocks habitat specific?

From SNPs to haplotypes



- SNPs can be ordered into haplotypes
- Haplotypes provide deep & shallow evolutionary information
- Phasing genotypes within and among RAD sites
- Genotype imputation for missing SNPs

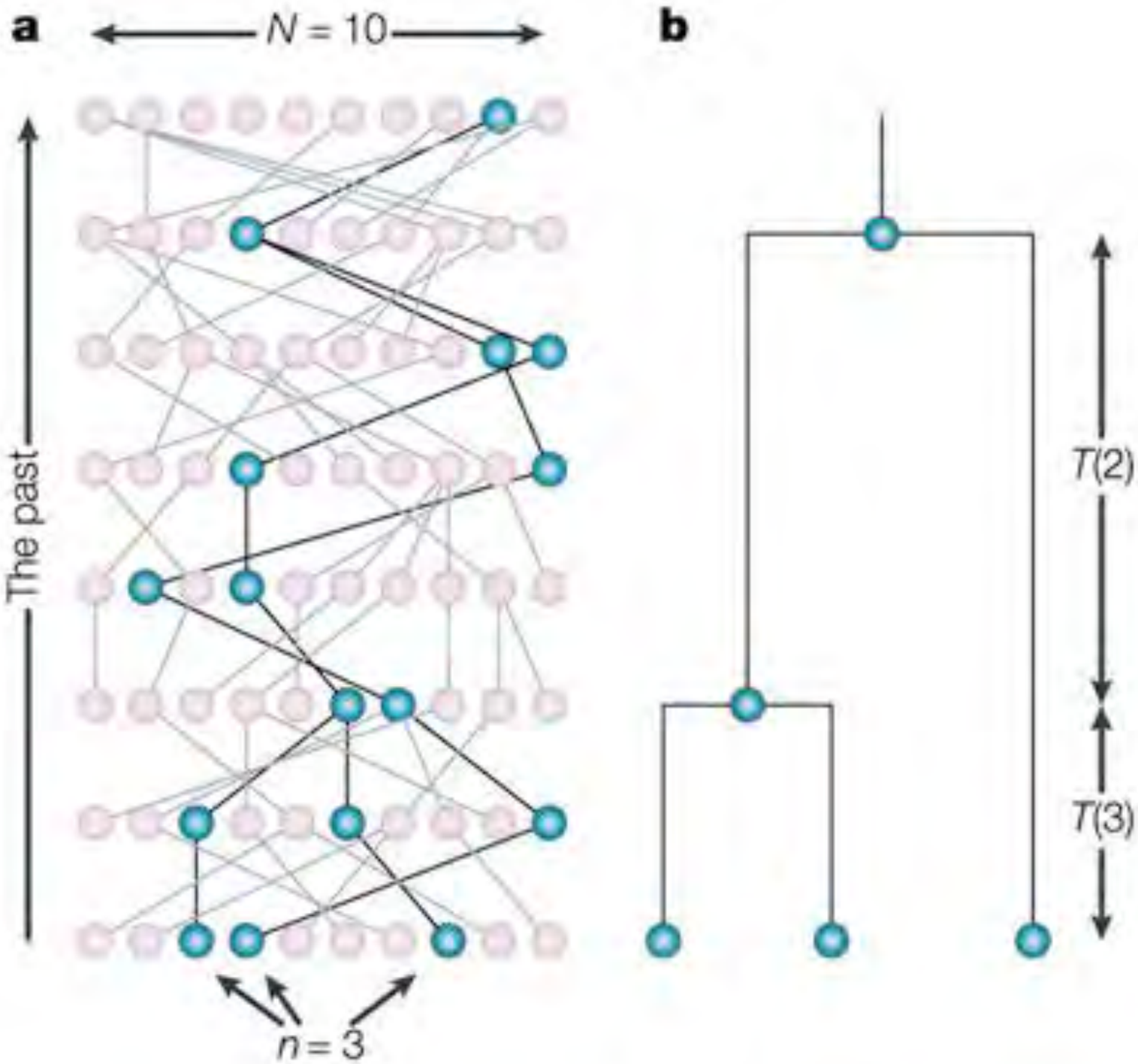
Coalescent analysis using RAD-seq data



Noah A. Rosenberg & Magnus Nordborg
Nature Reviews Genetics **3**, 380-390 (May 2002)

Nature Reviews | Genetics

Coalescent analysis using RAD-seq data



35000

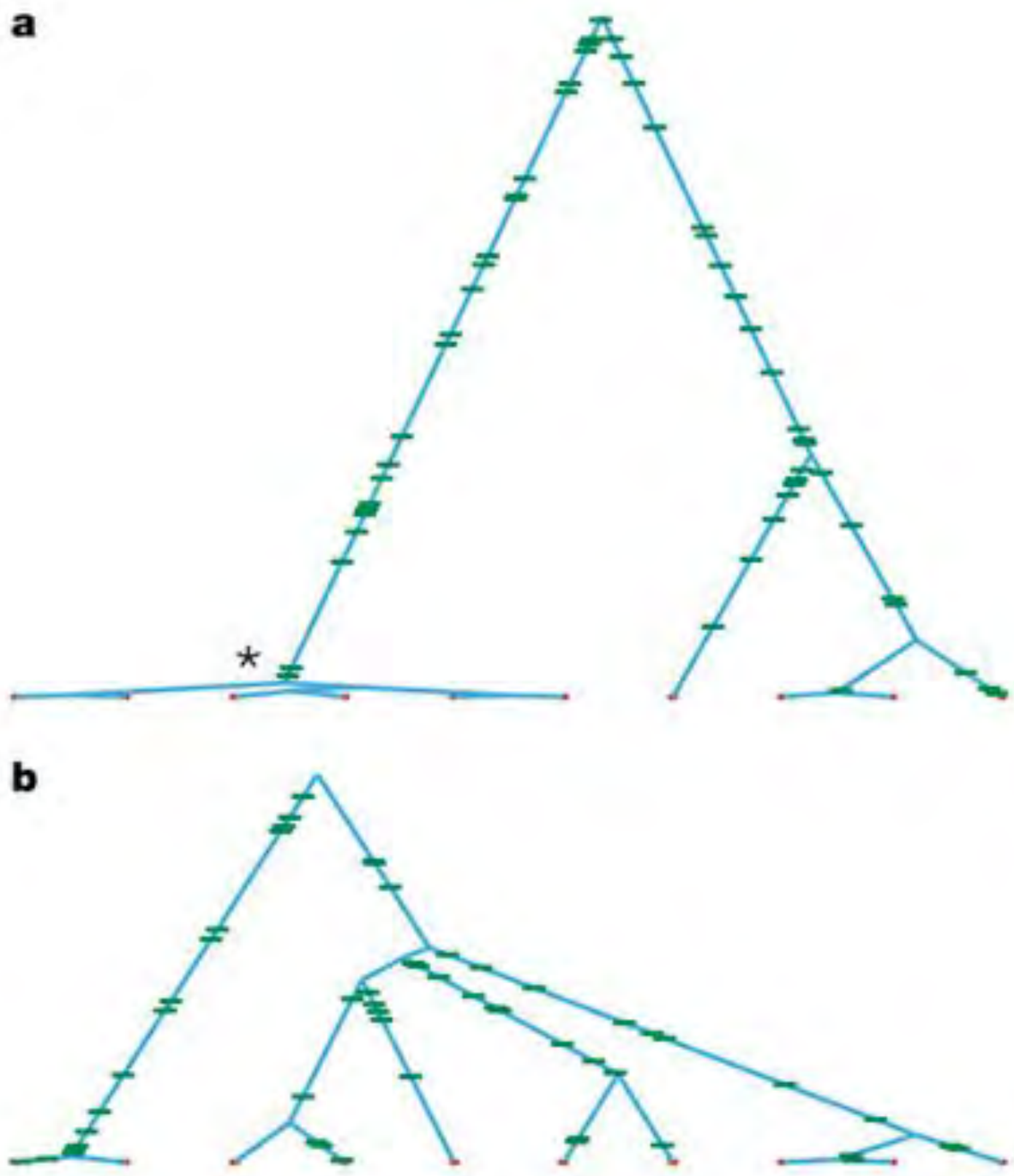
Noah A. Rosenberg & Magnus Nordborg
Nature Reviews Genetics **3**, 380-390 (May 2002)

Nature Reviews | Genetics

Neutral coalescent expectations

Noah A. Rosenberg & Magnus Nordborg
Nature Reviews Genetics **3**, 380-390 (May 2002)

Natural selection and the coalescent



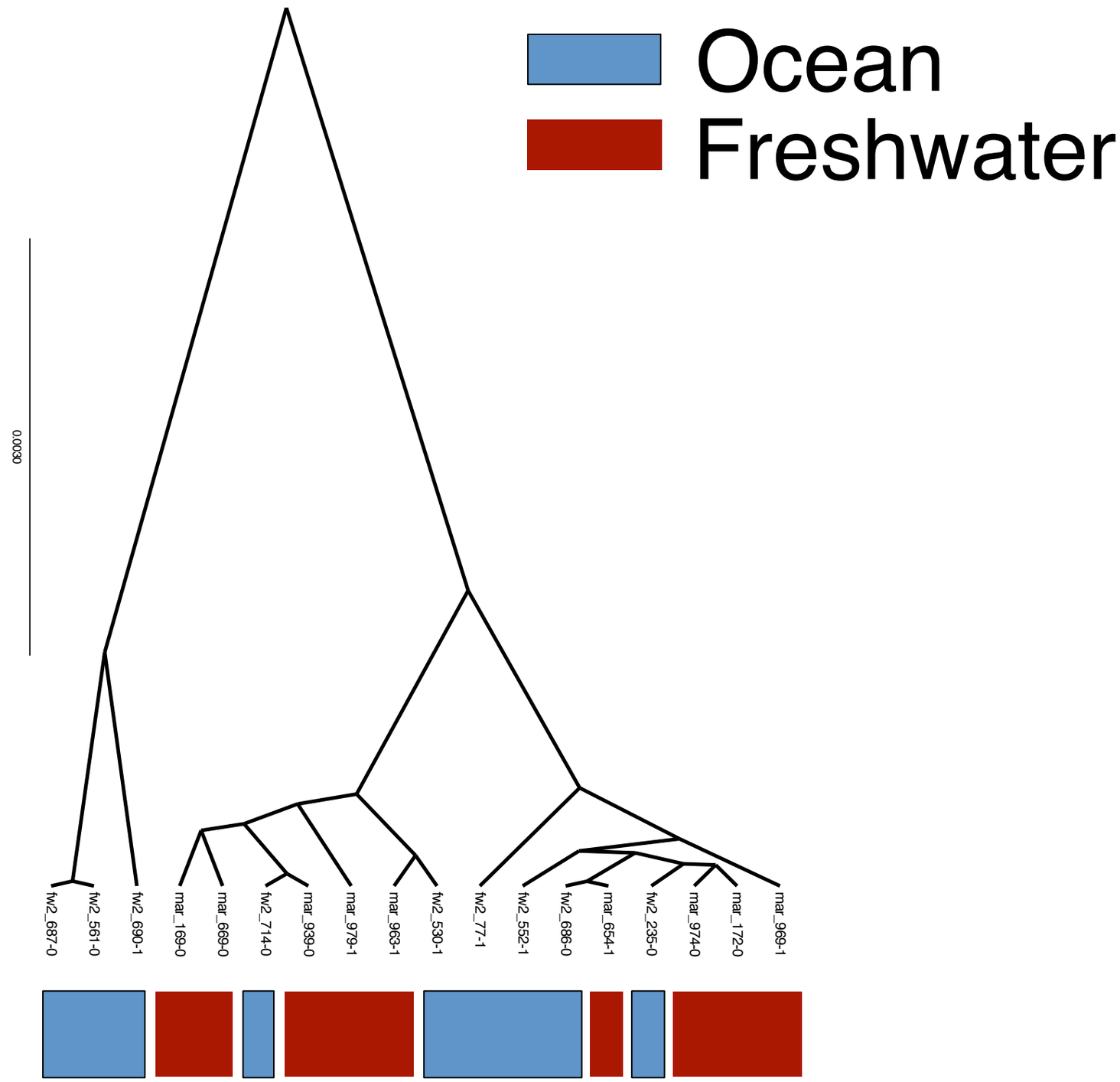
Divergent Selection

Balancing Selection

Noah A. Rosenberg & Magnus Nordborg
Nature Reviews Genetics **3**, 380-390 (May 2002

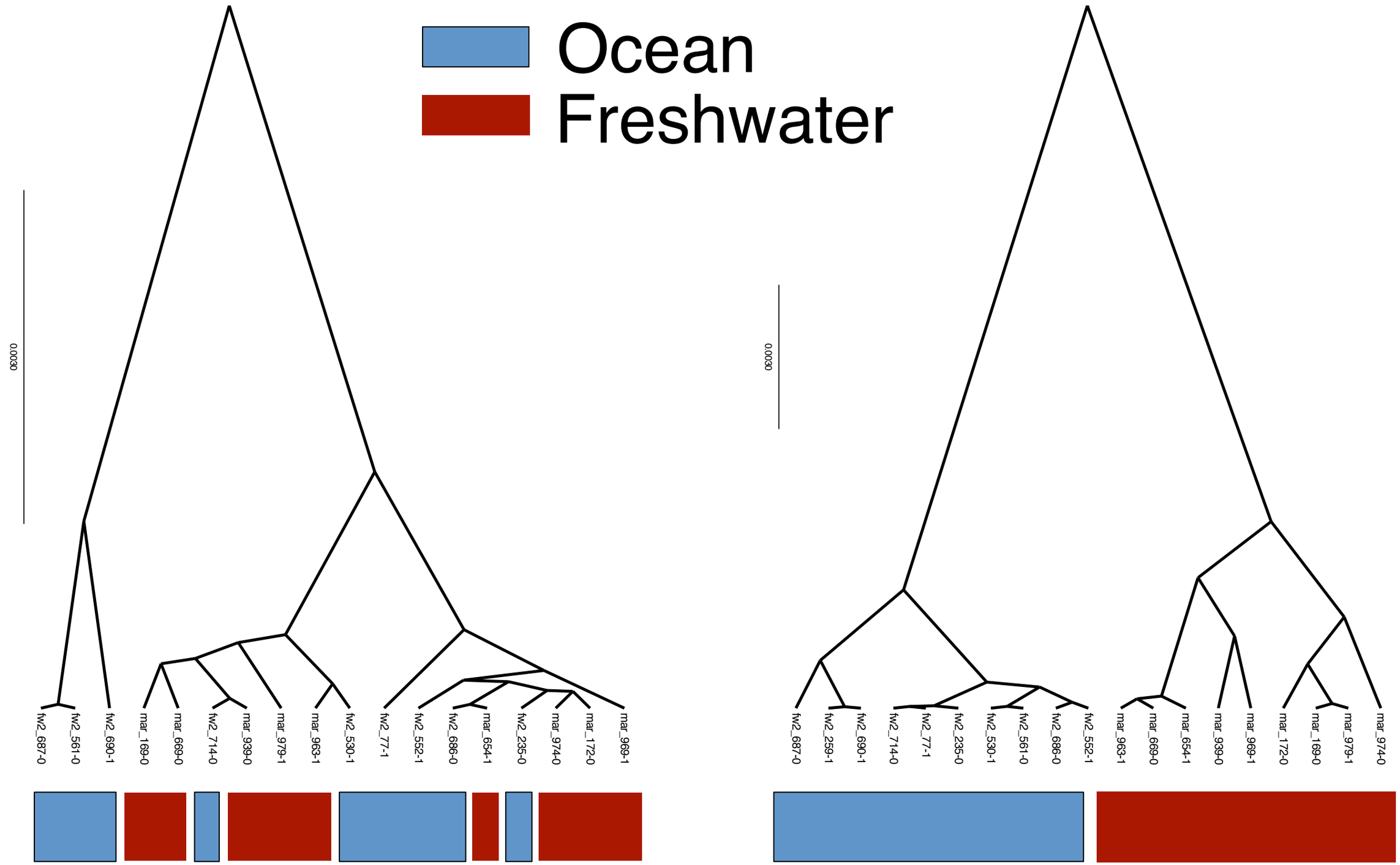
Nature Reviews | Genetics

RAD-seq coalescent in stickleback

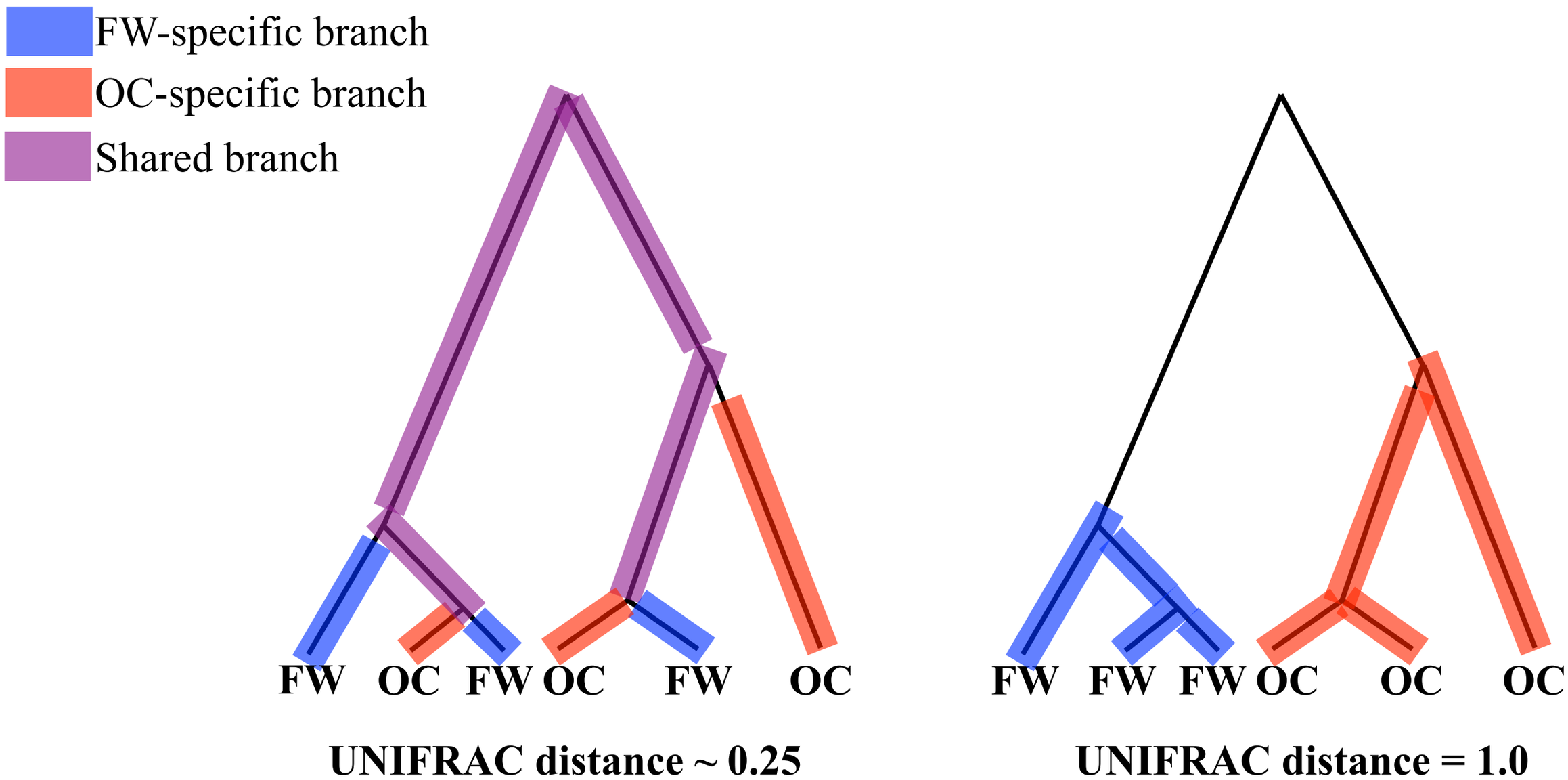


Thom Nelson &
Julian Catchen

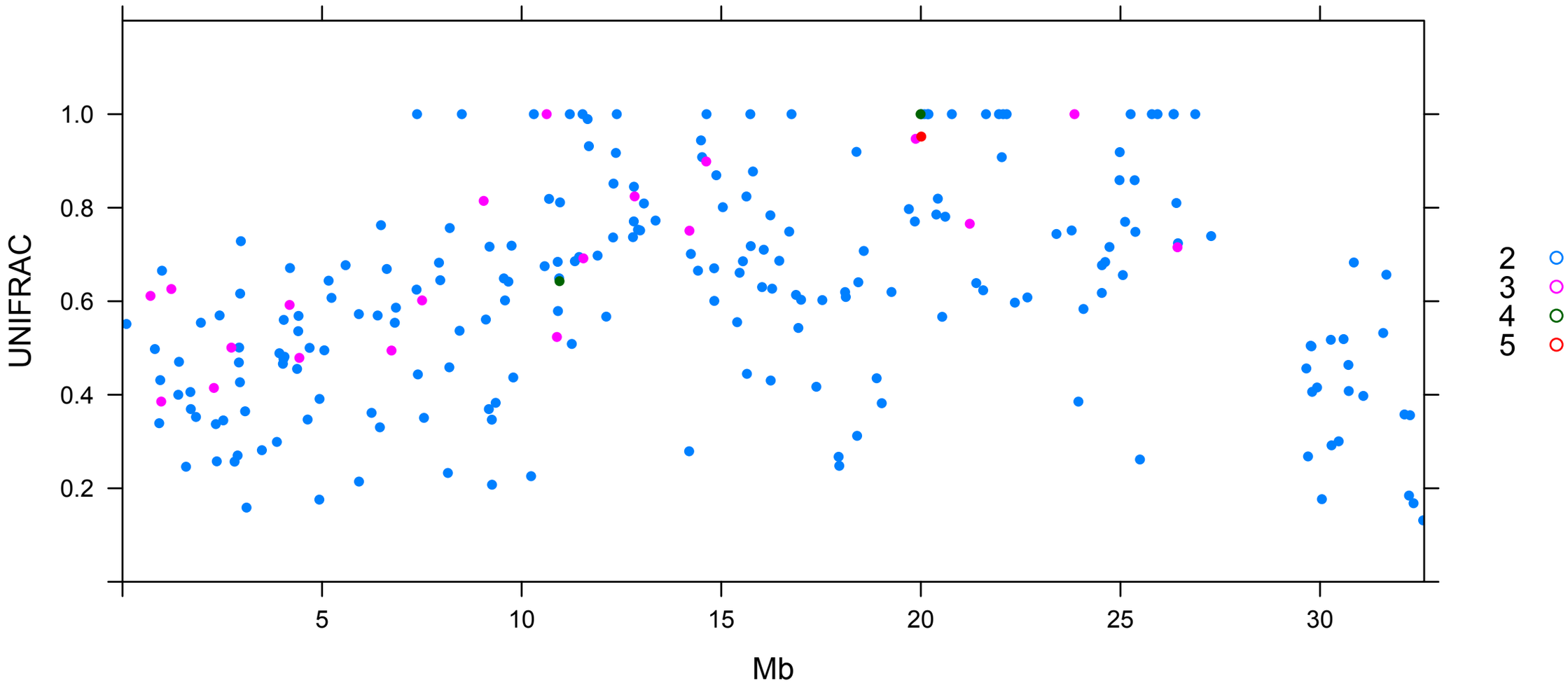
RAD-seq coalescent in stickleback



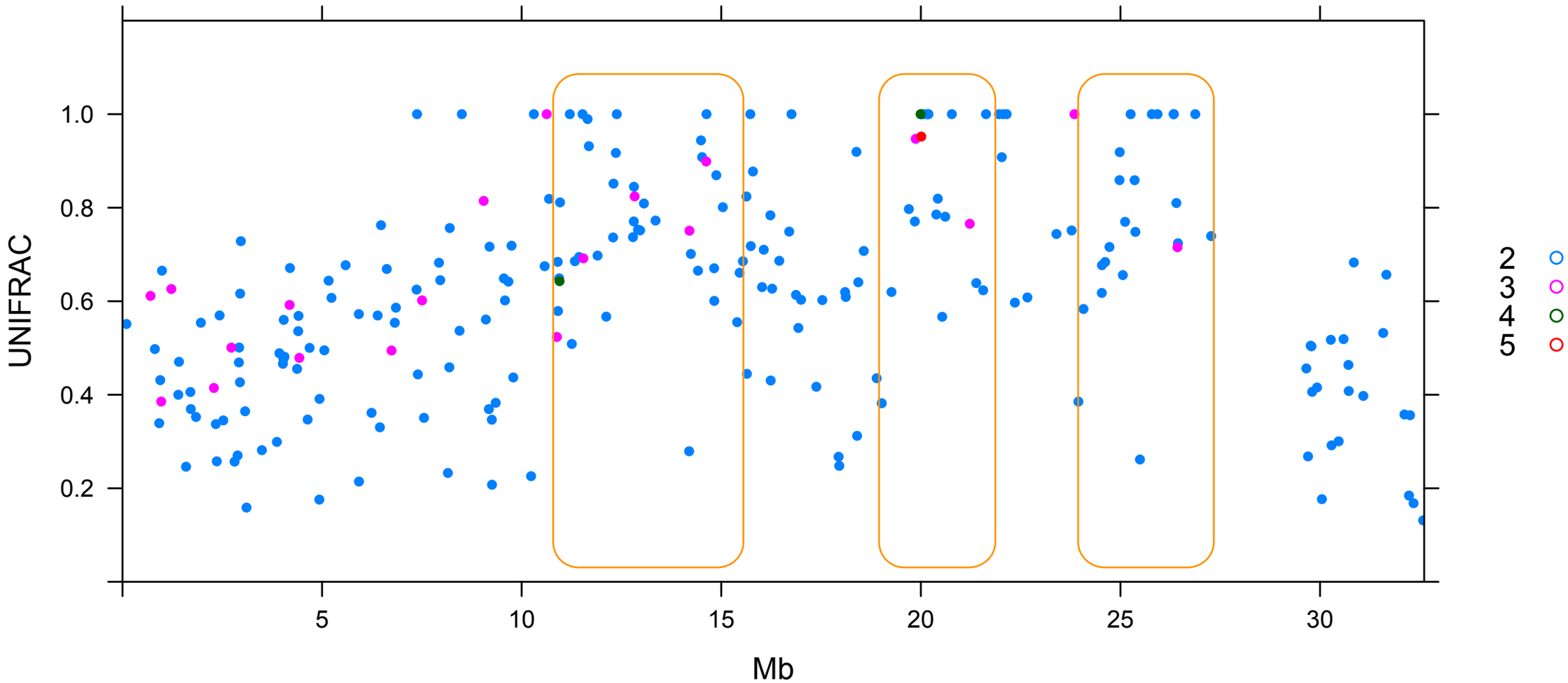
RAD-seq coalescent in stickleback - **UNIFRAC**



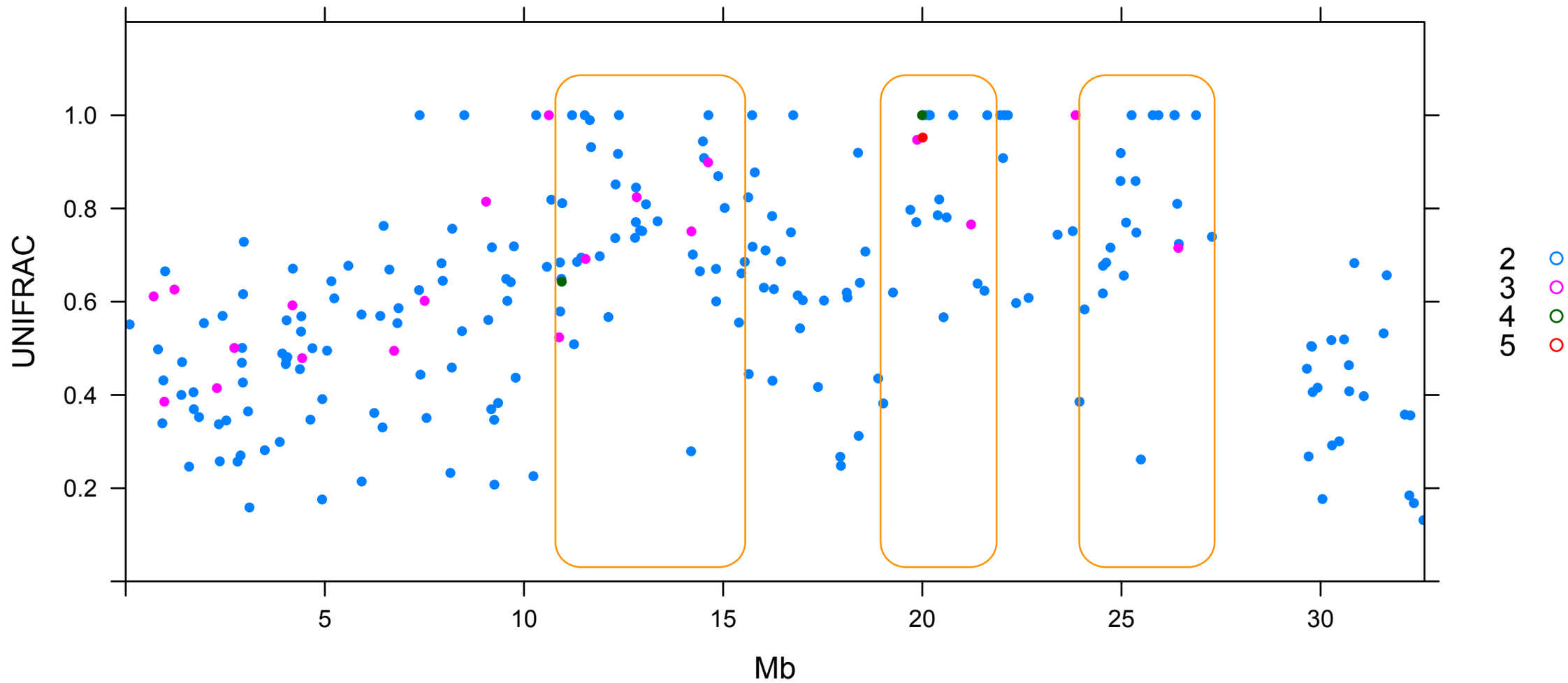
RAD-seq coalescent in stickleback - **UNIFRAC**

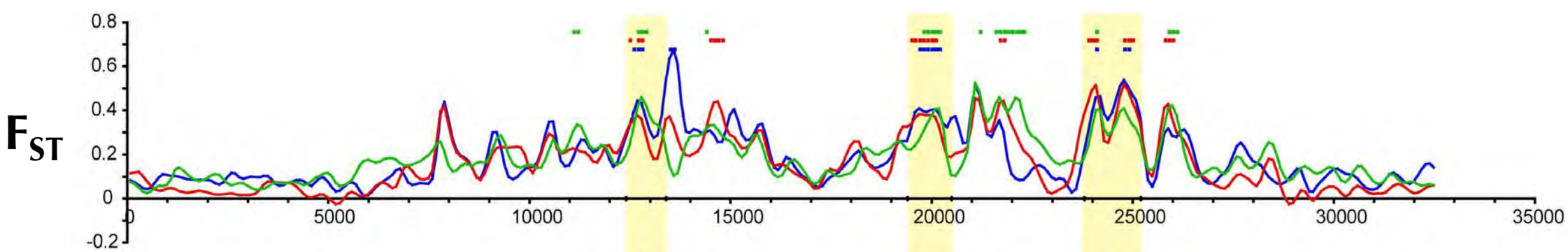


Many haplotypes are habitat specific

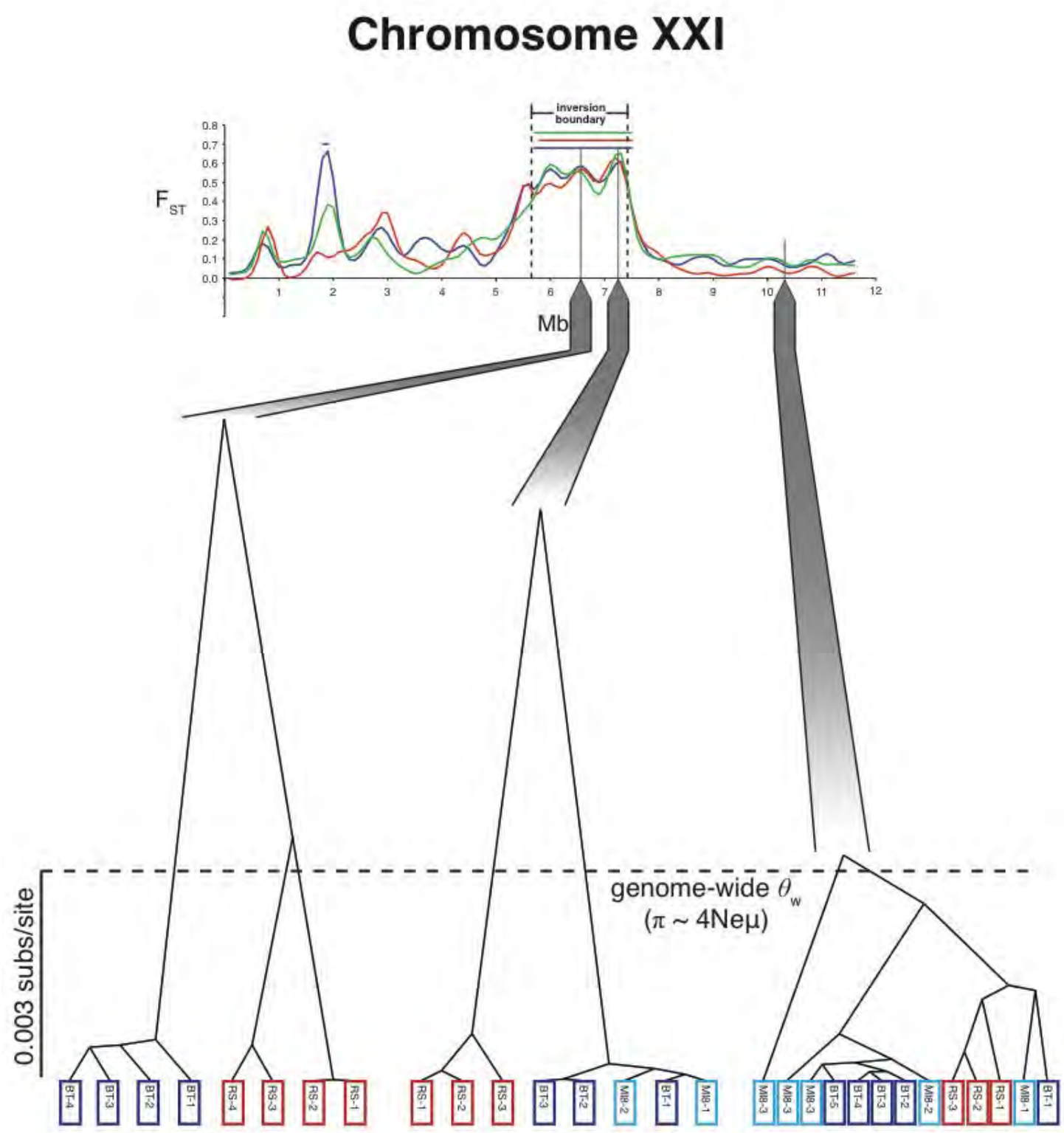


And correspond with signatures of divergent selection





Many haplotypes in regions subject to divergent section are quite old

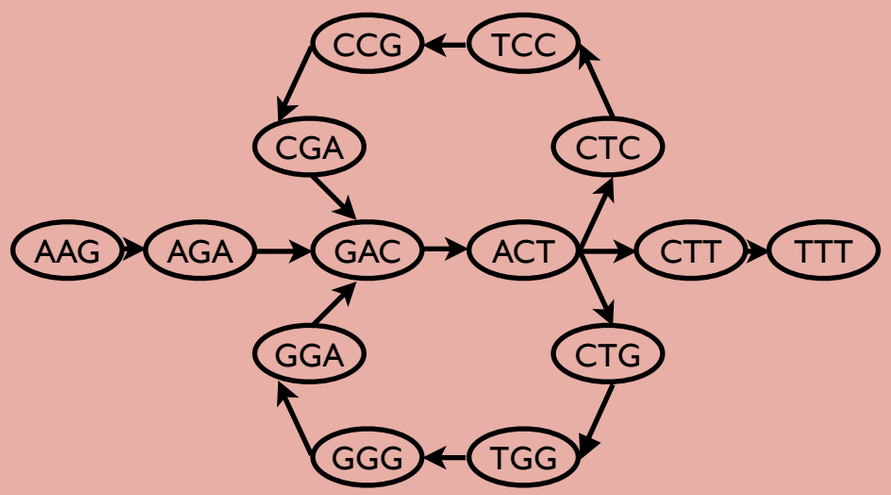


What can explain such rapid evolution and haplotype structure?

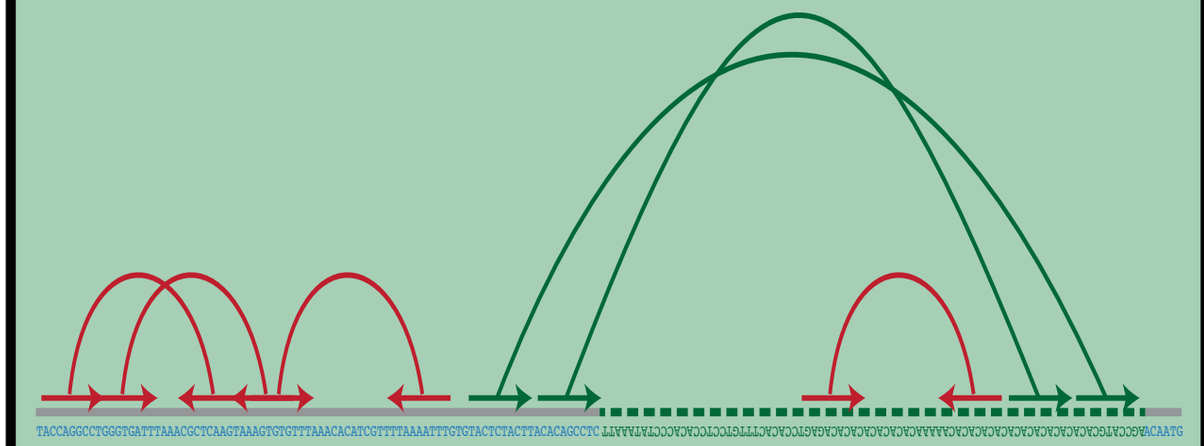
Is the stickleback genome architecture partly responsible?

Julian Catchen, Susie Bassham and Thom Nelson

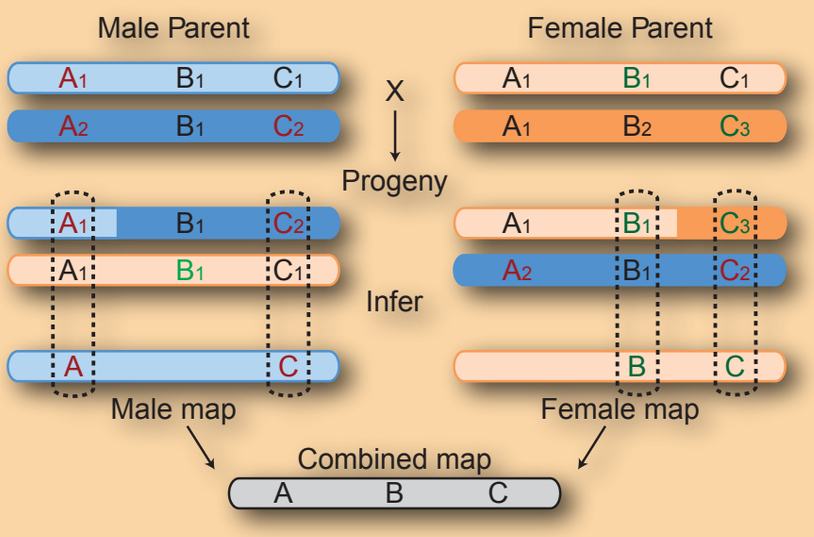
Genome Assembly



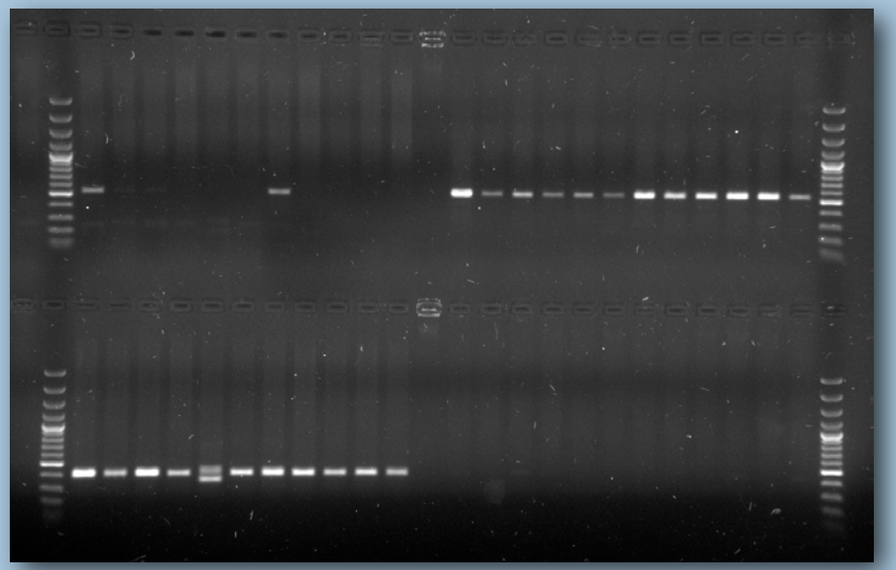
Paired-end Alignments



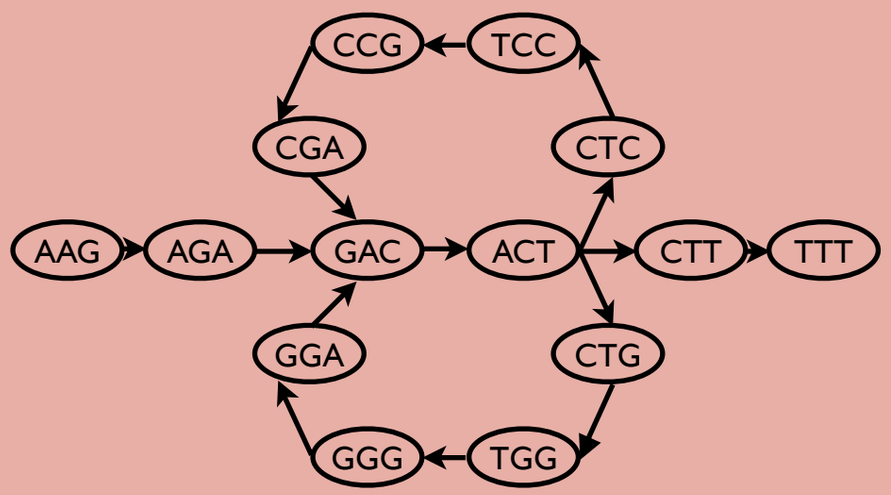
Genetic Map Construction



PCR Screening of Breakpoints

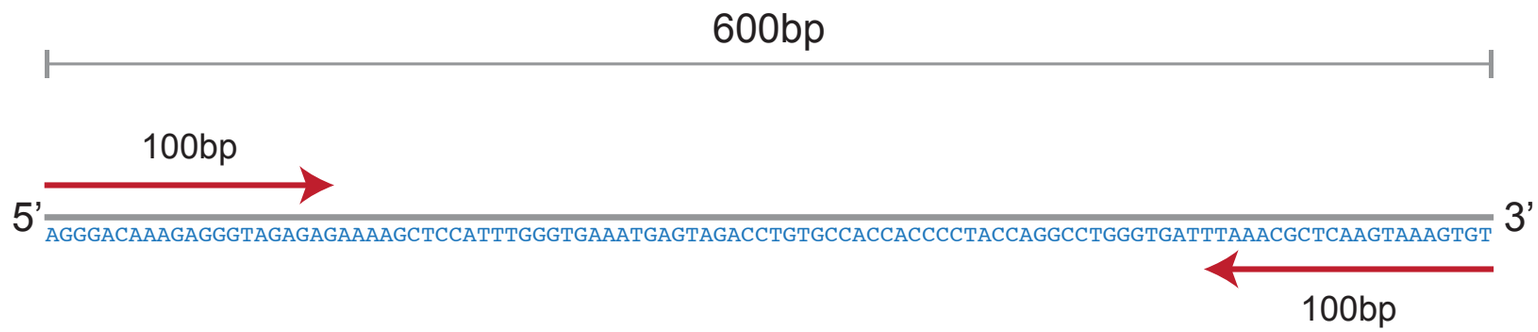


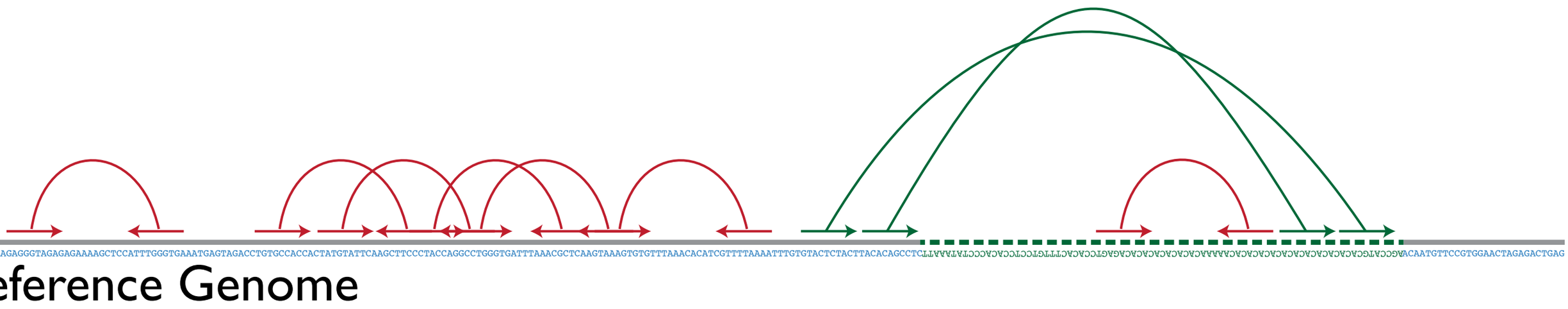
Genome Assembly



	♂ 	♀ 	♂ 	♀ 
N50	17,417 bp	18,982 bp	15,555 bp	15,534 bp
Max	199,905 bp	192,283 bp	238,768 bp	254,734 bp
Total	488.8 Mb	472.5 Mb	456.4 Mb	473.4 Mb
Median Coverage	24.6x	26.5x	24.1x	25.8x

Illumina Paired-end Reads

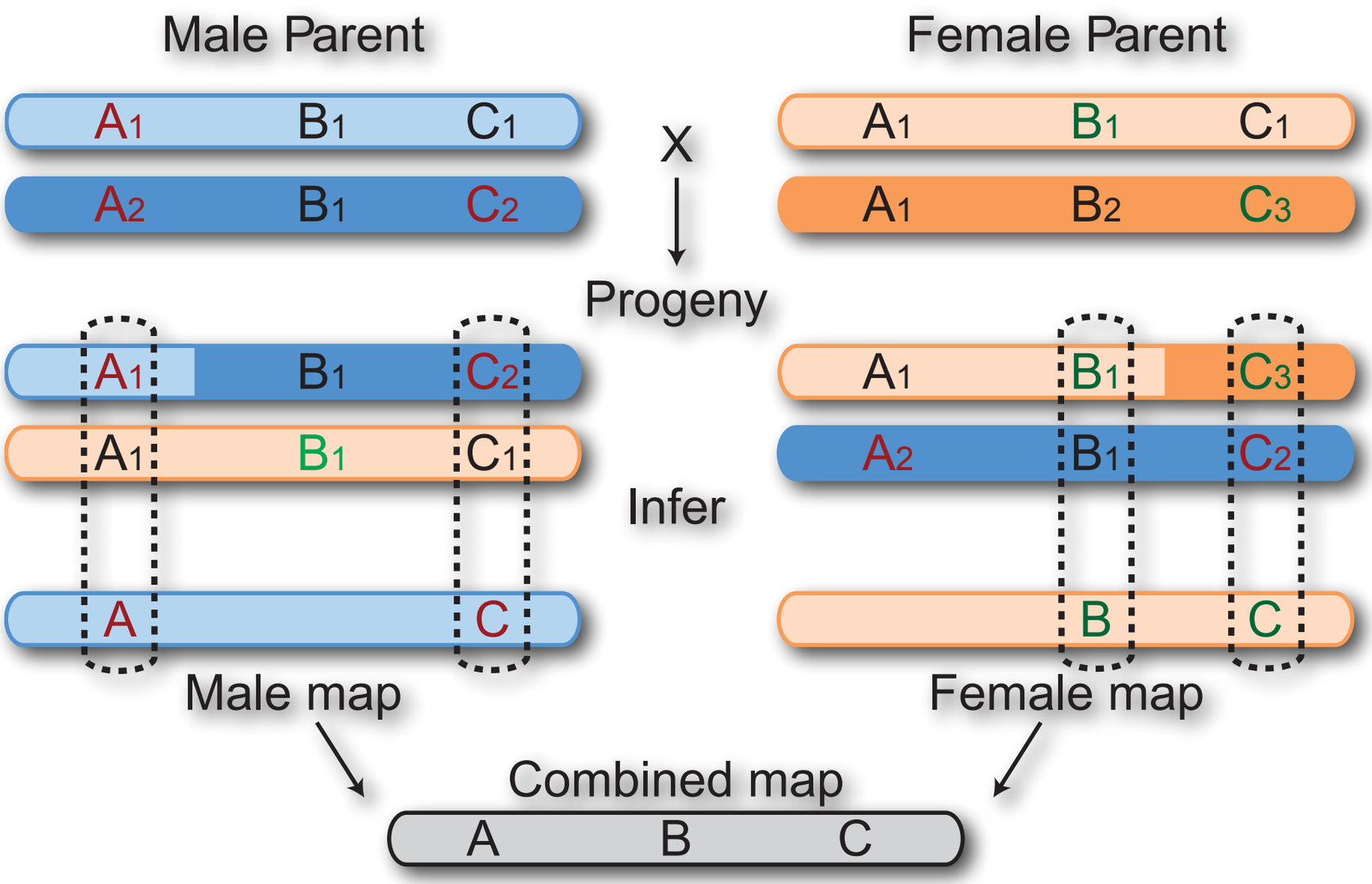


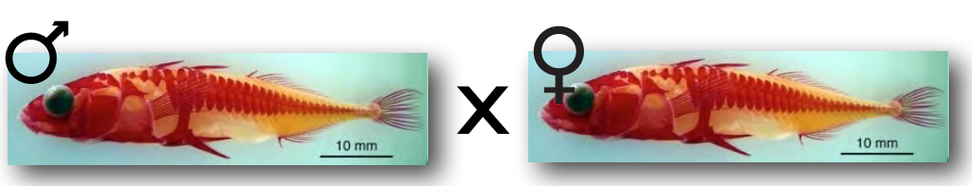


Reference Genome

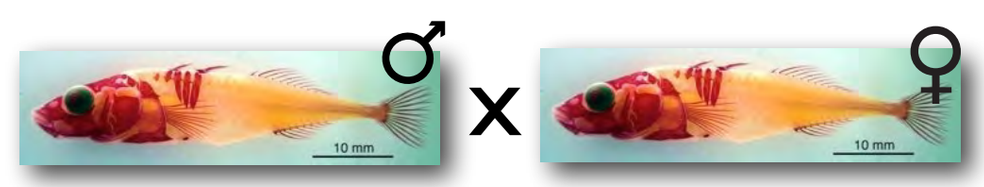
161,305,595 pairs	144,396,898 pairs	131,471,548 pairs	150,786,462 pairs

F1 Pseudo-testcross using RAD-seq



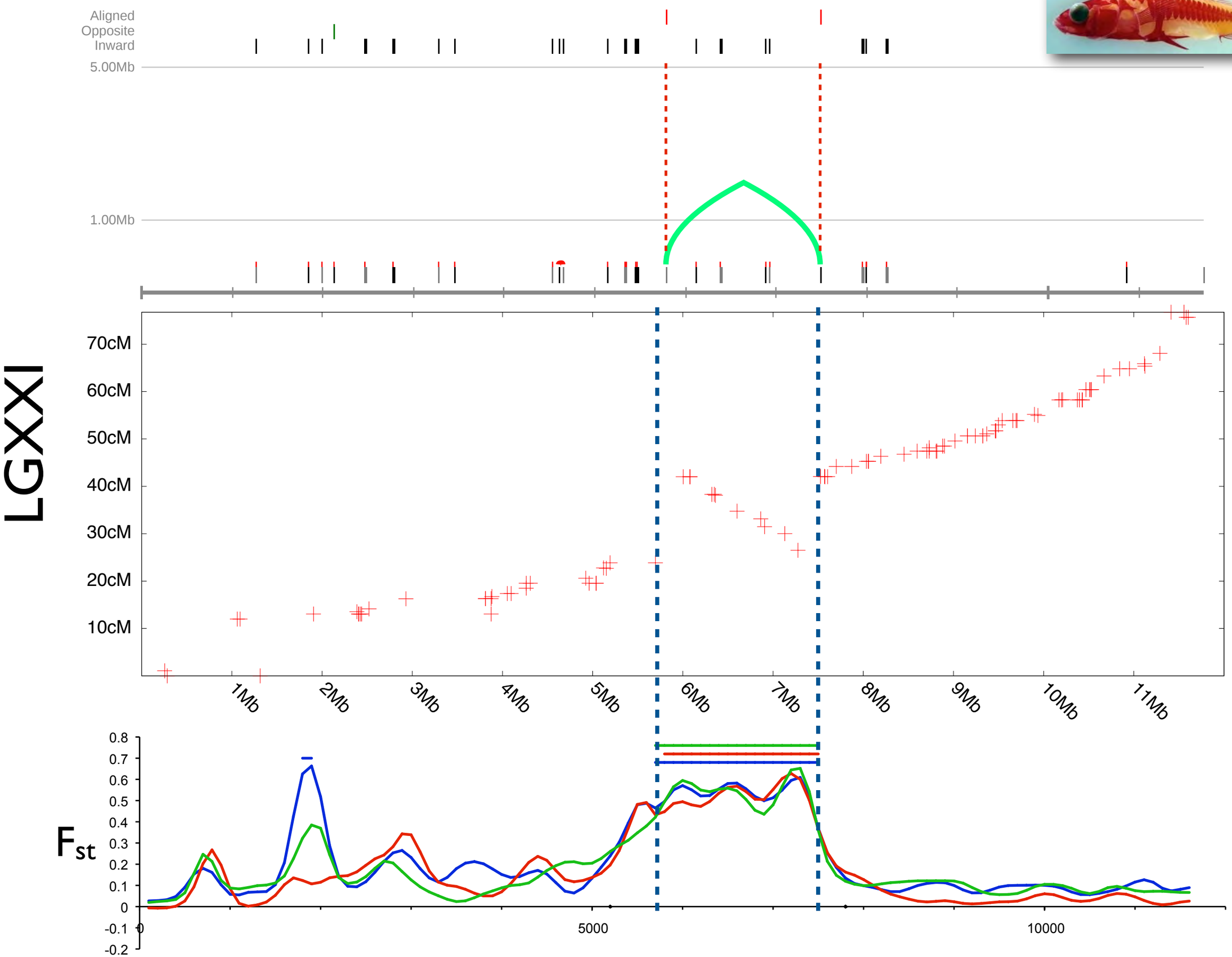


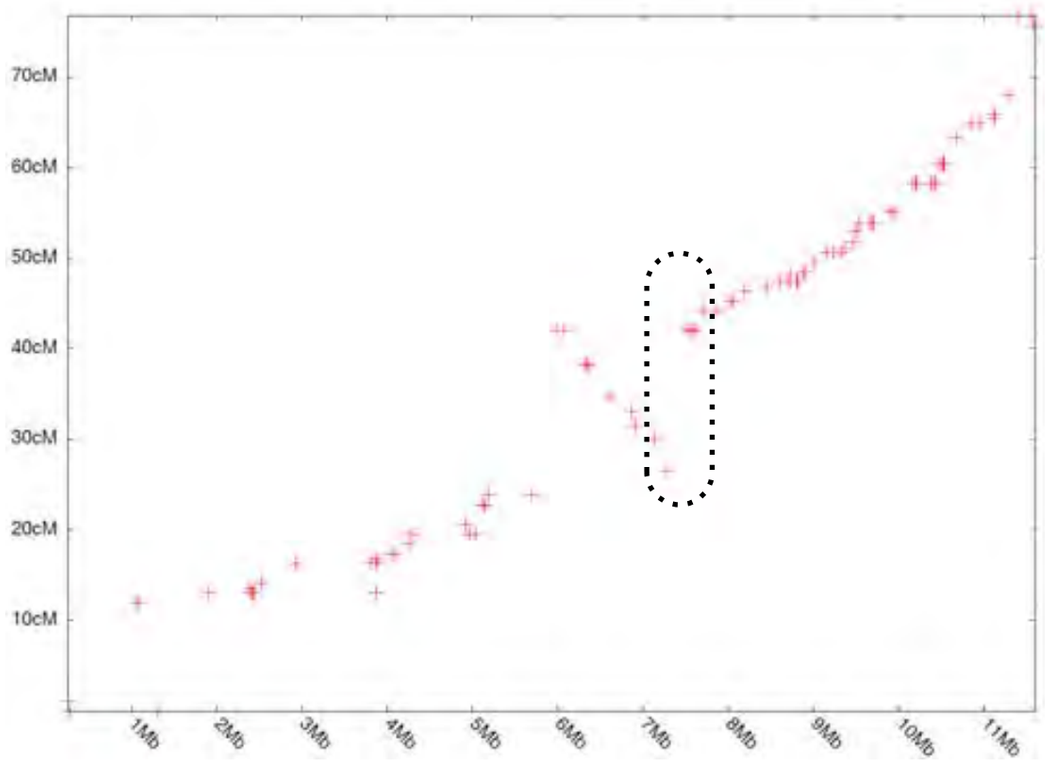
93 progeny
66,071 loci
5,351 markers



93 progeny
45,301 loci
3,927 markers


10 mm


LGXI
Aligned
Opposite
Inward
5.00Mb
1.00Mb
70cM
60cM
50cM
40cM
30cM
20cM
10cM
1Mb
2Mb
3Mb
4Mb
5Mb
6Mb
7Mb
8Mb
9Mb
10Mb
11Mb
Fst
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
0
5000
10000

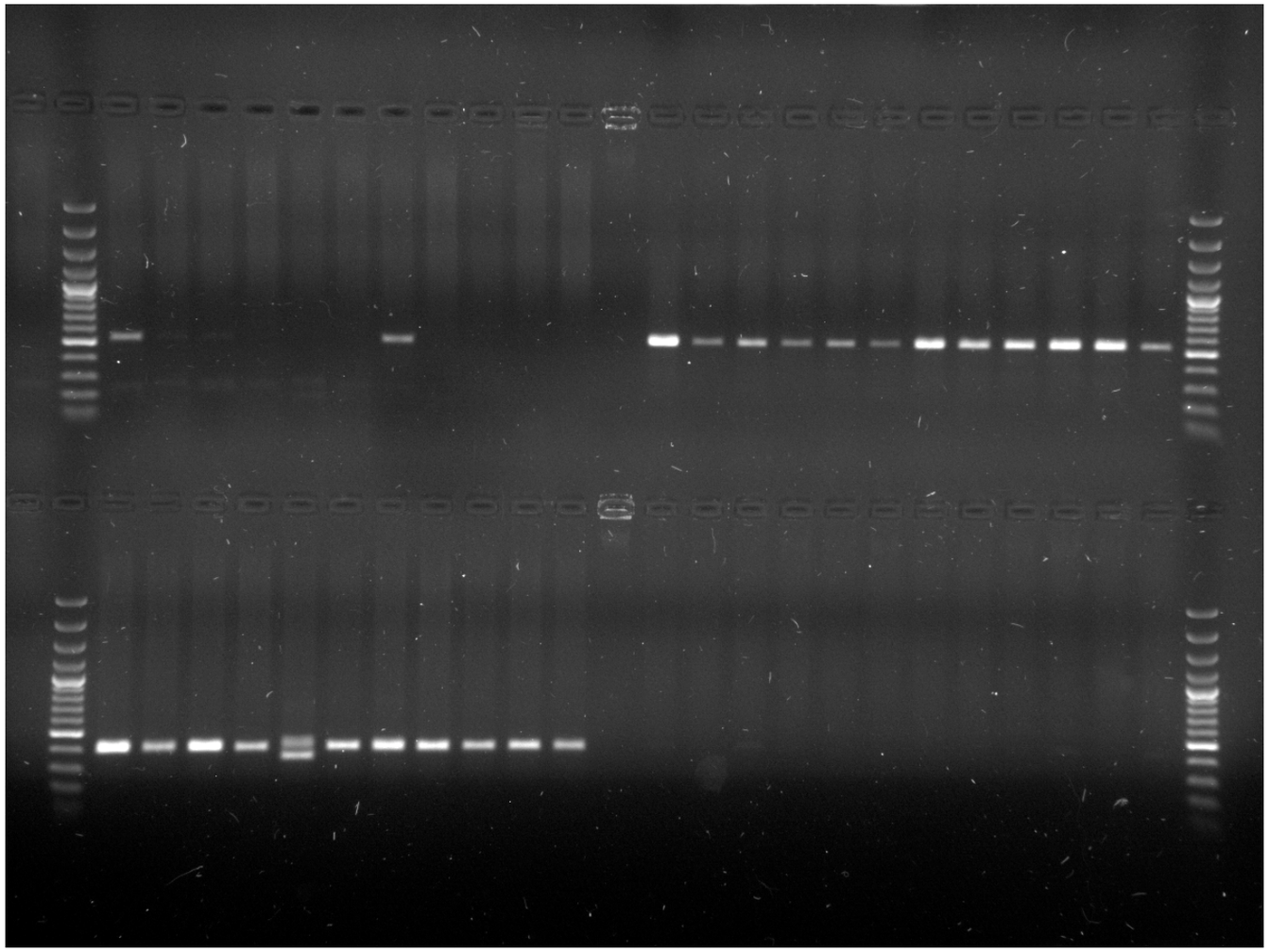


Linkage Group XXI

RS (Marine) Boot (Freshwater)

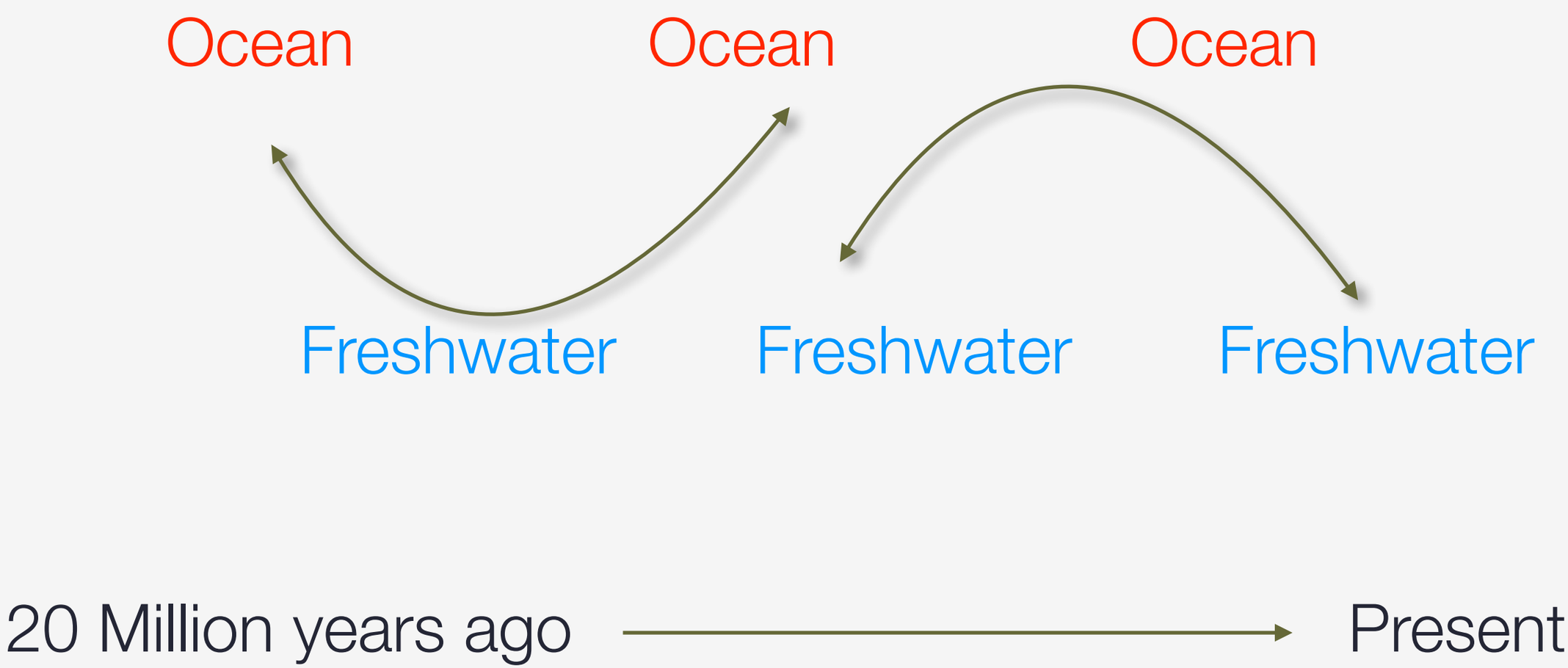
Genome Arrangement

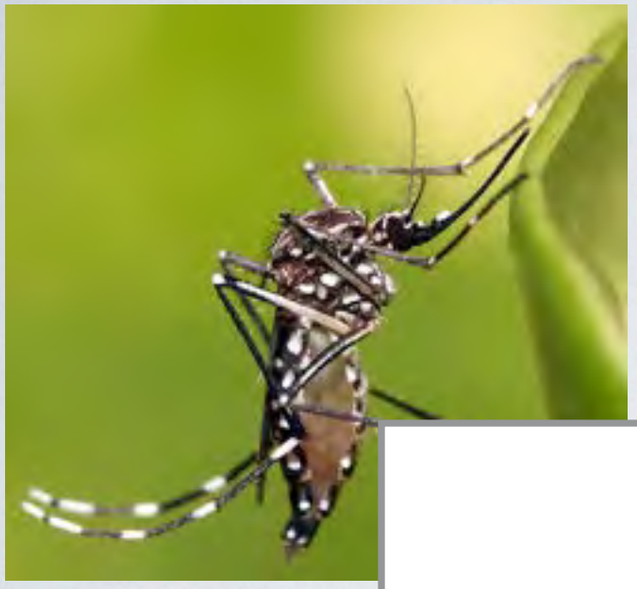
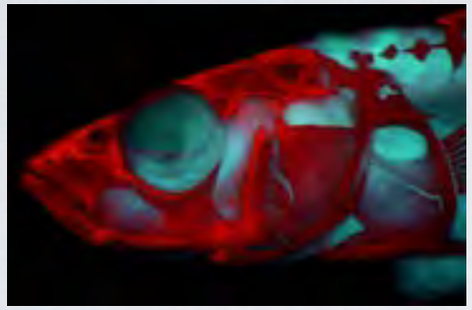
Inverted



1. Previous work has shown that the freshwater genomes evolve in 13,000 years.
2. These new Middleton Island data shows that the phenotype can appear in as little as *50 years*.
3. Much of the divergence involves soft sweeps.
4. This could represent thousands of haplotypes reassembling, but the genome appears *chunkier*.
5. Many haplotypes are habitat specific and are quite ancient.
6. Haplotypes often coincide with structural variation across the stickleback genome

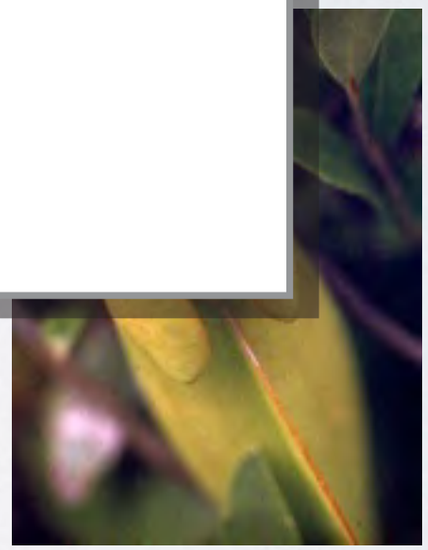
Hypothesis: Old genomic architecture variation is a product of the metapopulation structure of stickleback, and this architecture strongly influences subsequent rapid evolution.



Considerations for RAD-seq studies





Experimental design considerations for RAD

Tradeoffs:
Number of sites versus **Depth** of sequencing per site versus **Number of samples**

Experimental design considerations for RAD

Tradeoffs:
Number of sites versus **Depth** of sequencing per site versus **Number of samples**

```
raw reads / samples / sites = coverage at each RAD locus

1,000,000 / 100 / 1,000 = 10x coverage

25 to 50x average coverage per RAD locus is a good goal
```

Experimental design considerations for RAD

Tradeoffs:
Number of sites versus **Depth** of sequencing per site versus **Number of samples**

How many tags do I need?

Things to consider

Choice of enzyme and genome size $(0.25)^n \times$ genome size = expected # sites

Genomes are biased:

expect 112,300 six-cutter sites in stickleback (460 Mb)	actual ***EcoRI*** sites = 90,000
expect 7000 eight-cutter sites in stickleback	actual ***SbfI*** sites = 22, 800
expect 32,900 six-cutter sites in *C. remanei* (135 Mb)	actual ***EcoRI*** sites = 73,200

Experimental design considerations for RAD

Tradeoffs:
Number of sites versus **Depth** of sequencing per site versus **Number of samples**

How many tags do I need?

Things to consider

Choice of enzyme and genome size

Polymorphism and read length

Nucleotide polymorphism rate = 0.01 to 0.001 for most vertebrates

Stickleback populations: 0.01 to 0.02. At least 1 SNP every 100 bp, on average

Experimental design considerations for RAD

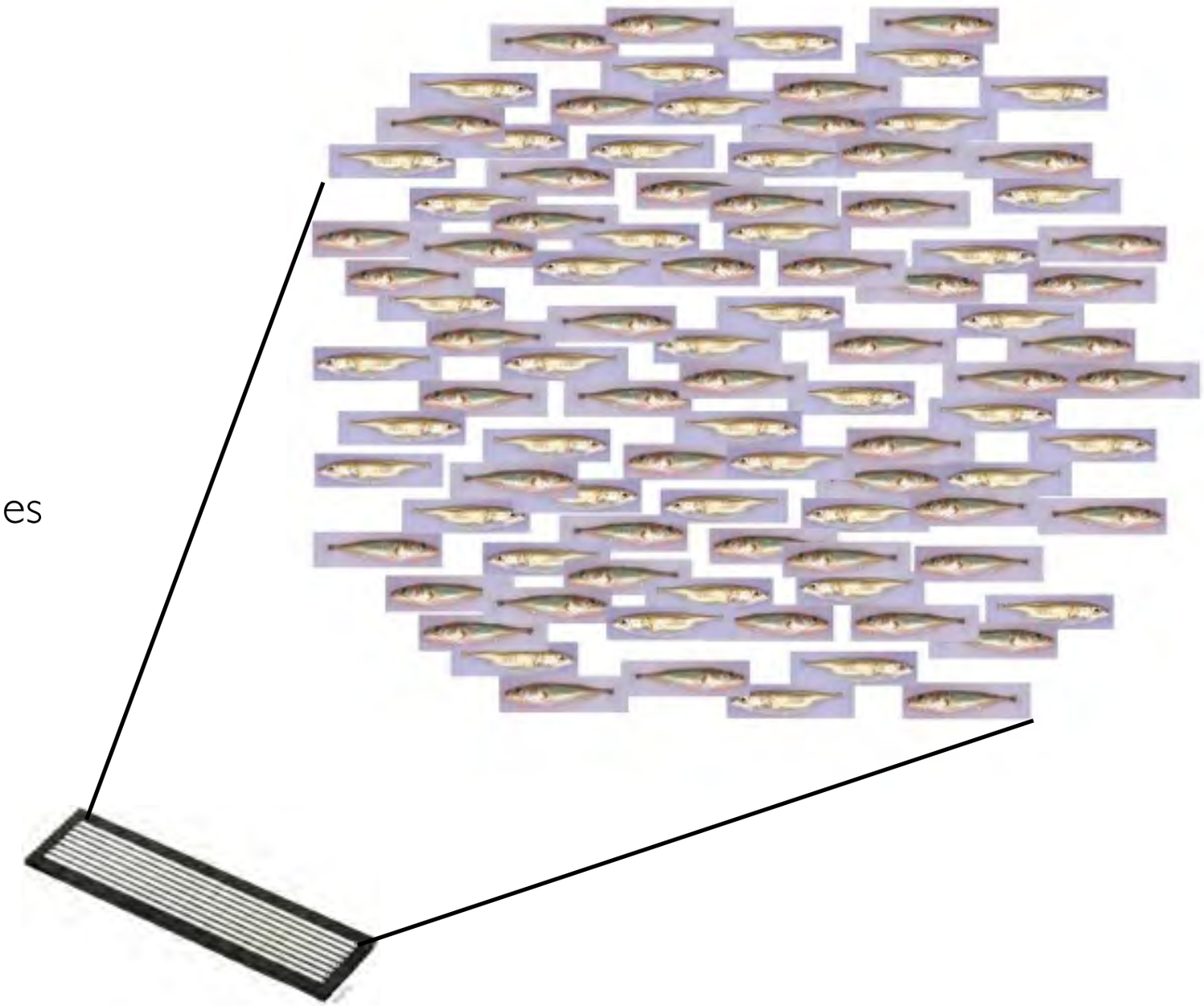
Tradeoffs:
Number of sites versus **Depth** of sequencing per site versus **Number of samples**

How many samples should be multiplexed?

Things to consider

Barcoded adapters
- 5 to 8nt barcodes
- Variable length barcodes
- Combinatorial barcodes (PE)

Barcode distance - two mismatches



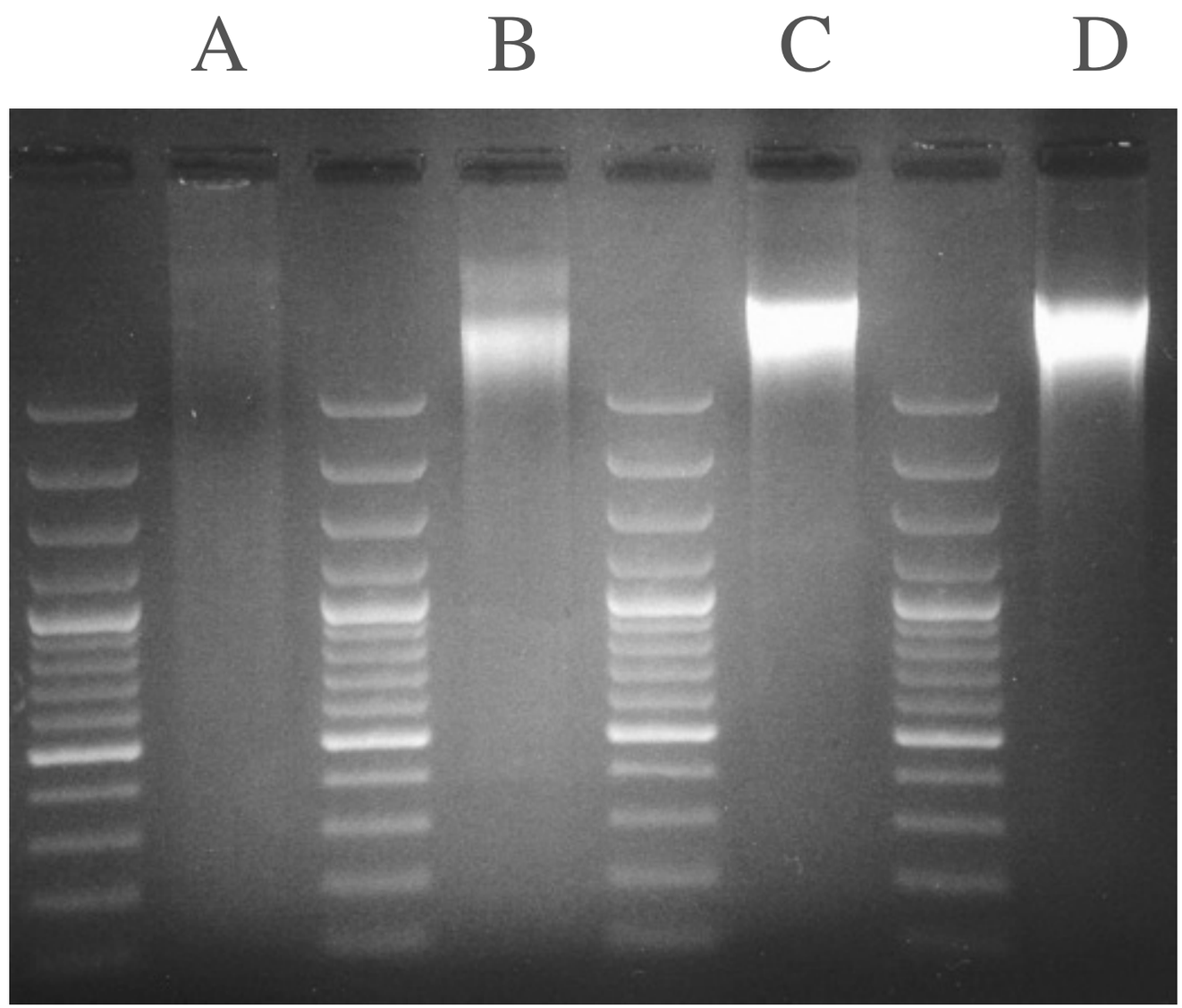
Molecular considerations in library building

How many samples should be multiplexed?

Things to consider

DNA Quality

Multiplex only like samples to help equalize representation of poor quality samples



Molecular considerations in library building

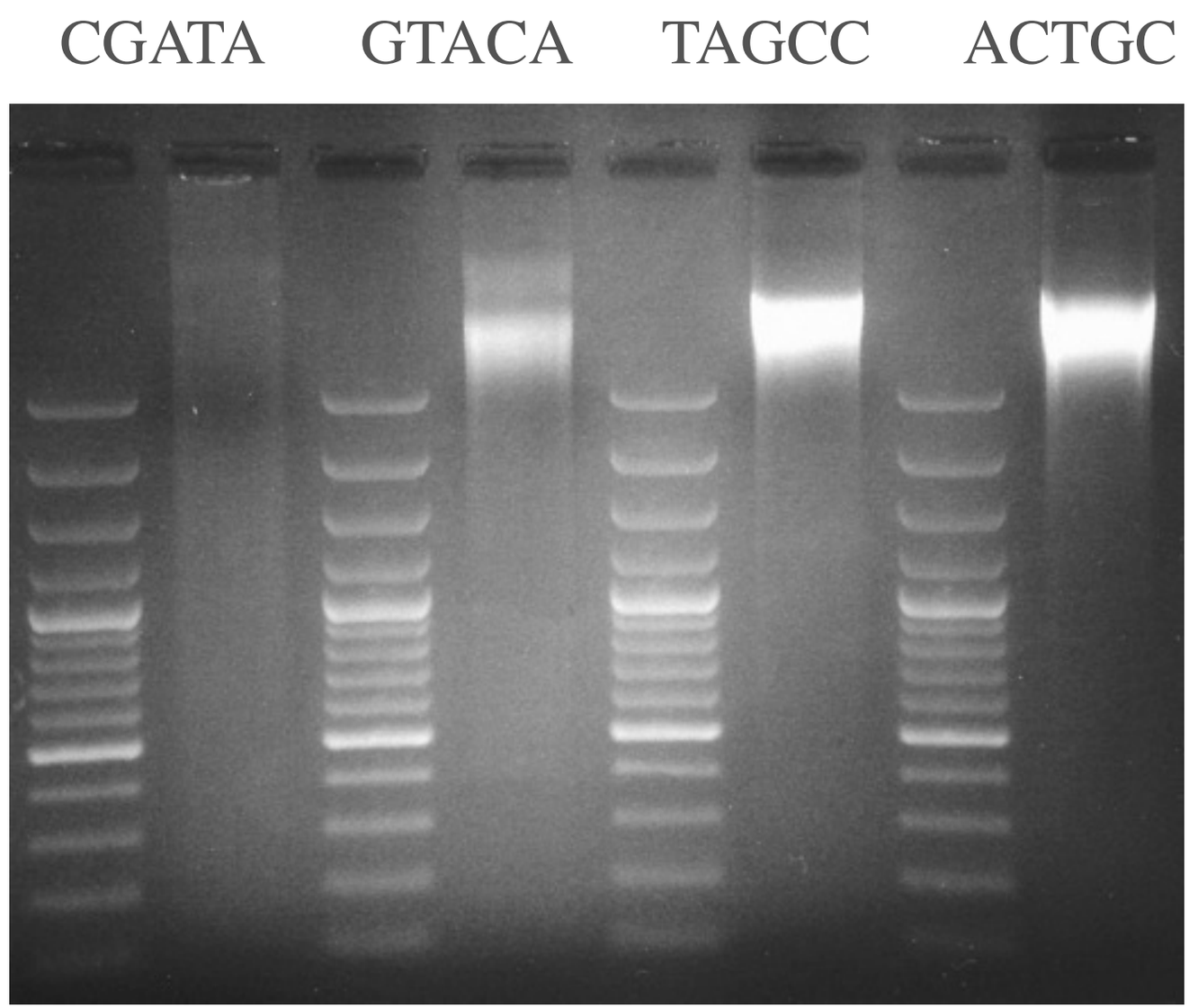
How many samples should be multiplexed?

Things to consider

DNA Quality

Diversify barcodes

- Illumina cluster calling is confused by repetition in first 4 bases - can offset barcodes



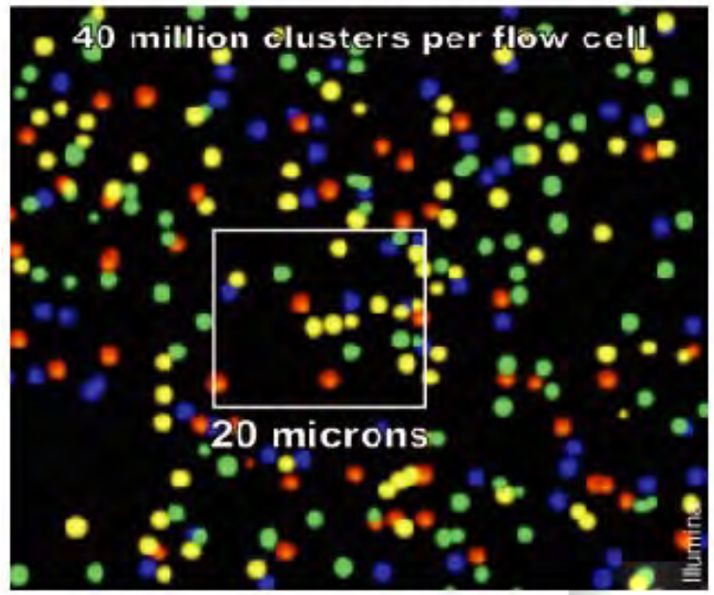
Molecular considerations in library building

How can I get the best depth of coverage?

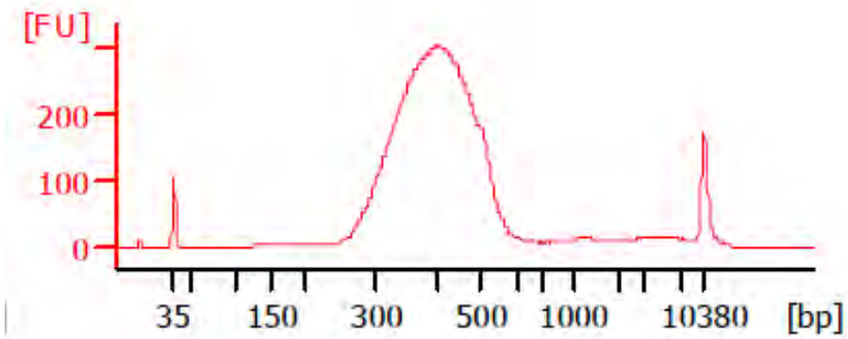
Things to consider

Fragment size

Smaller/tighter is better



Agilent Bioanalyzer



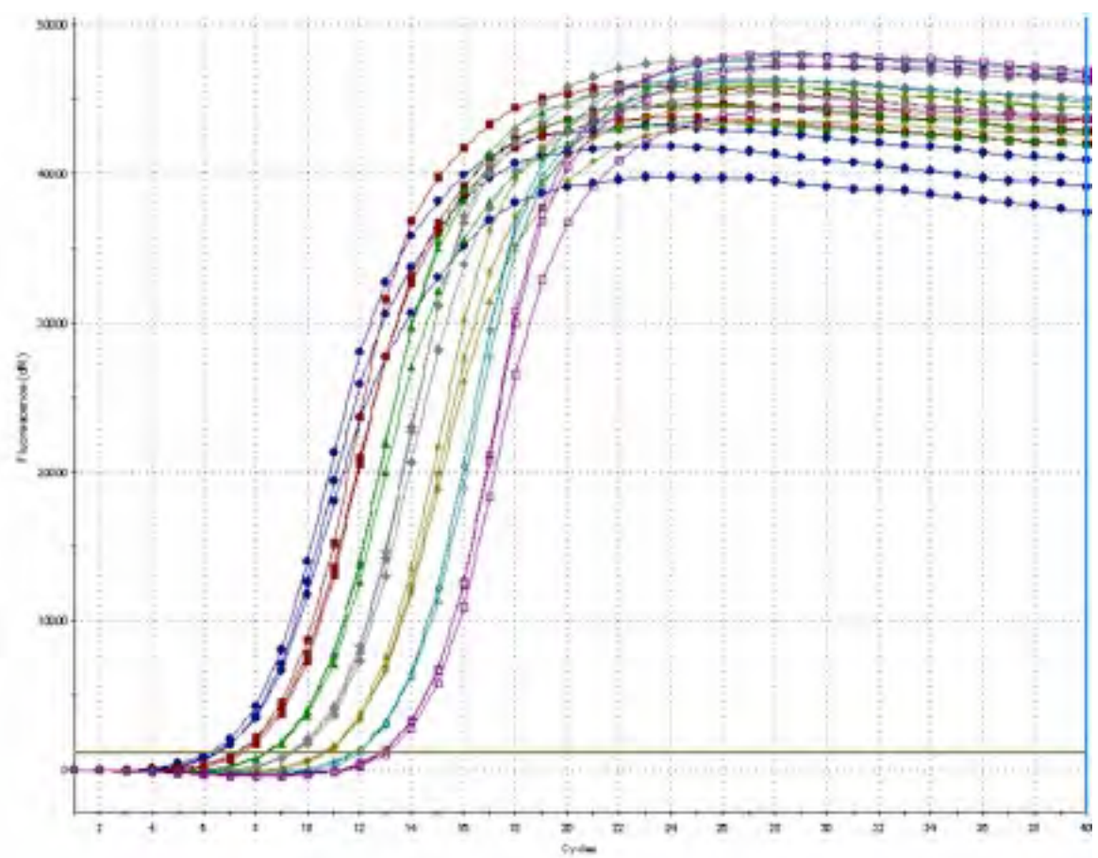
Molecular considerations in library building

How can I get the best depth of coverage?

Things to consider

Fragment size
Library quality
qPCR

qPCR control should be similar to measured sample:



Molecular considerations in library building

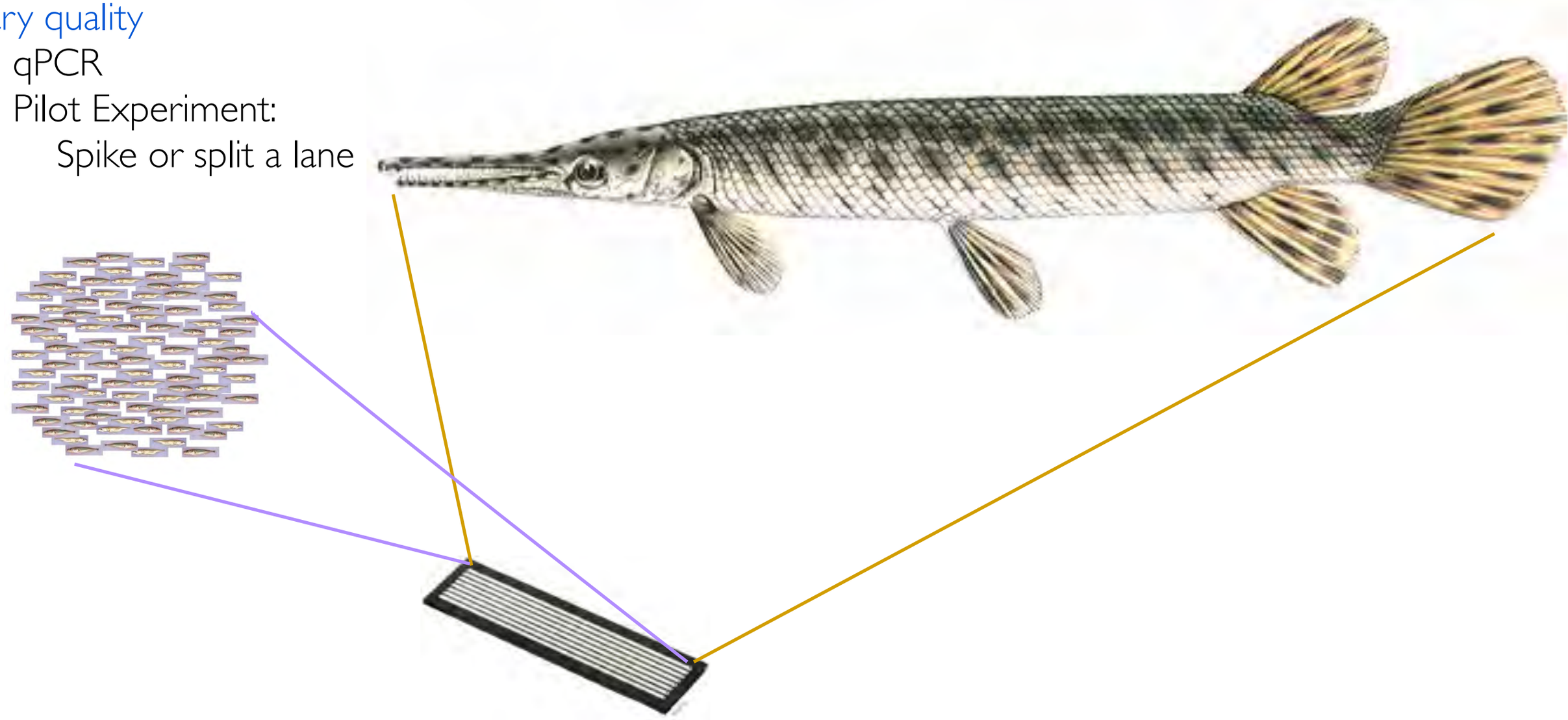
How can I get the best depth of coverage?

Things to consider

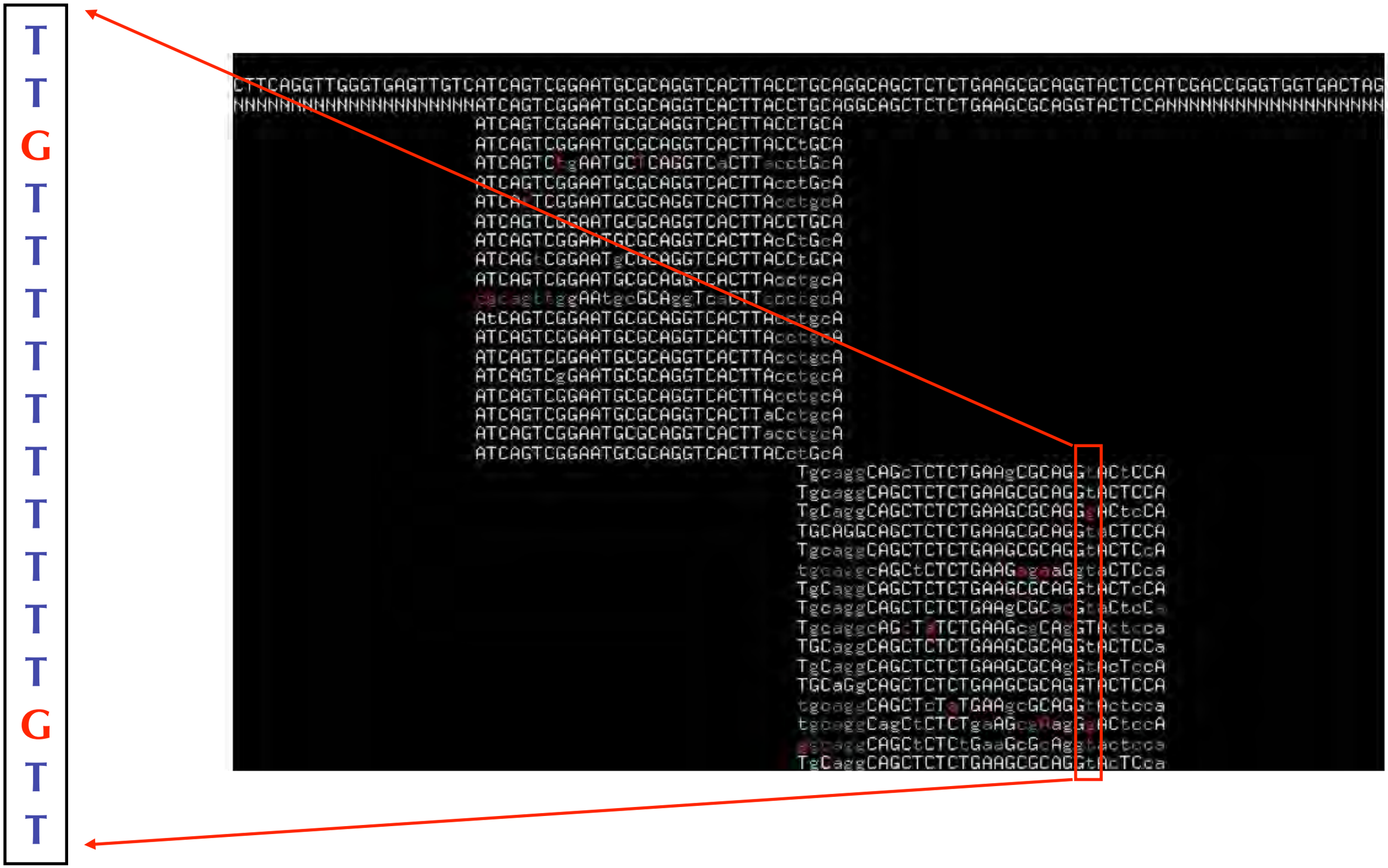
Fragment size

Library quality

- qPCR
- Pilot Experiment:
  - Spike or split a lane



Statistical considerations in RAD-seq


T
T
G
T
T
T
T
T
T
T
T
T
T
T
T
T
G
T
T

T
T
G
T
T
T
T
T
T
T
T
T
T
T
G
T
T

The reads are 14 T and 2 G:

GT heterozygote?
GG homozygote with error?
AA homozygote with lots of error?

Needed a rigorous method to call genotypes

T
T
G
T
T
T
T
T
T
T
T
T
T
T
T
G
T
T

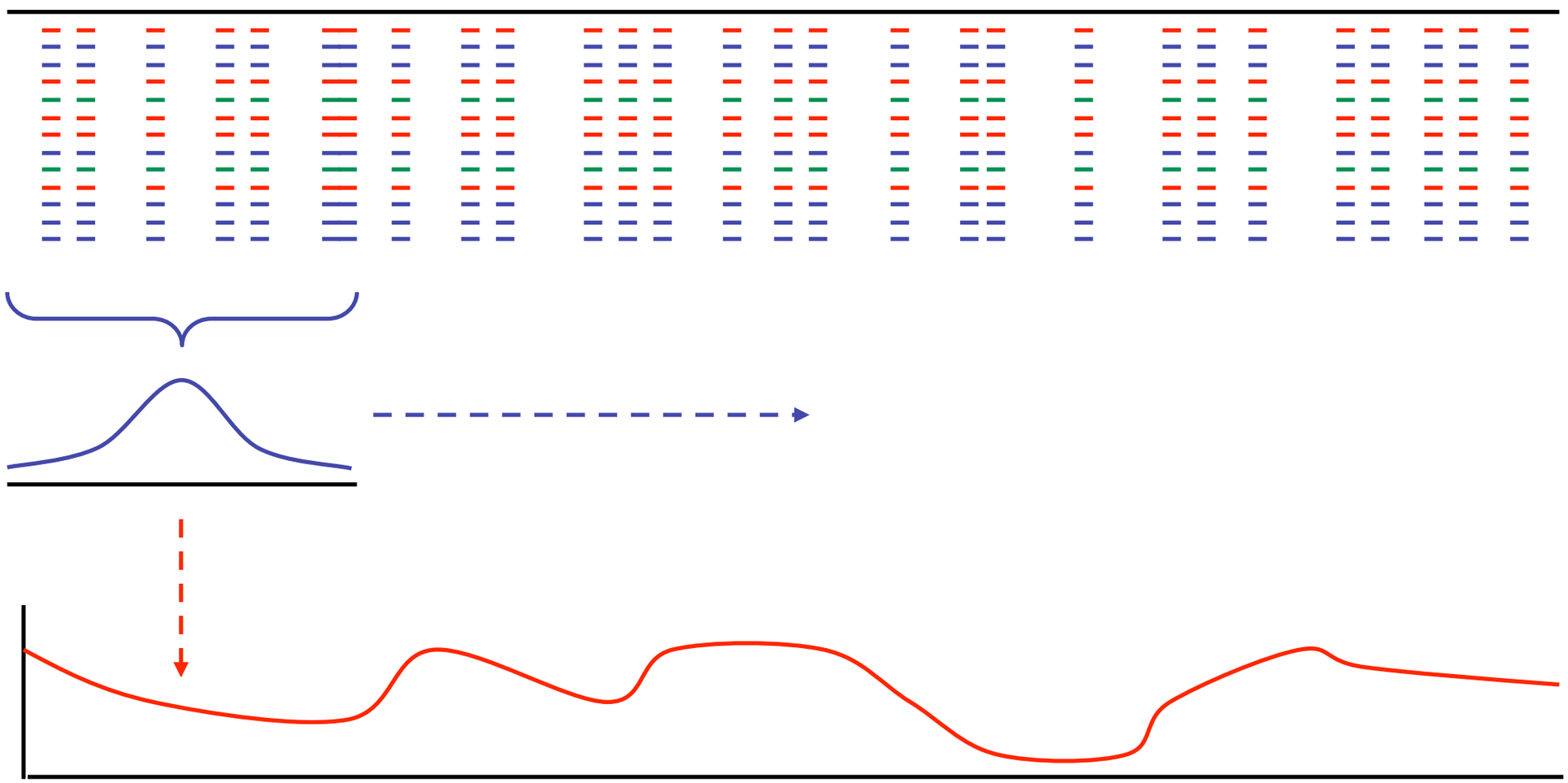
$$L(n_1 \text{hom}) = P(n_1, n_2, n_3, n_4) = \frac{n!}{n_1! n_2! n_3! n_4!} \left(1 - \frac{3\varepsilon}{4}\right)^{n_1} \left(\frac{\varepsilon}{4}\right)^{n_2} \left(\frac{\varepsilon}{4}\right)^{n_3} \left(\frac{\varepsilon}{4}\right)^{n_4}$$

$$L(n_1 n_2 \text{het}) = P(n_1, n_2, n_3, n_4) = \frac{n!}{n_1! n_2! n_3! n_4!} \left(0.5 - \frac{\varepsilon}{4}\right)^{n_1} \left(0.5 - \frac{\varepsilon}{4}\right)^{n_2} \left(\frac{\varepsilon}{4}\right)^{n_3} \left(\frac{\varepsilon}{4}\right)^{n_4}$$

Maximum likelihood genotyping based on multinomial distribution of nucleotide reads

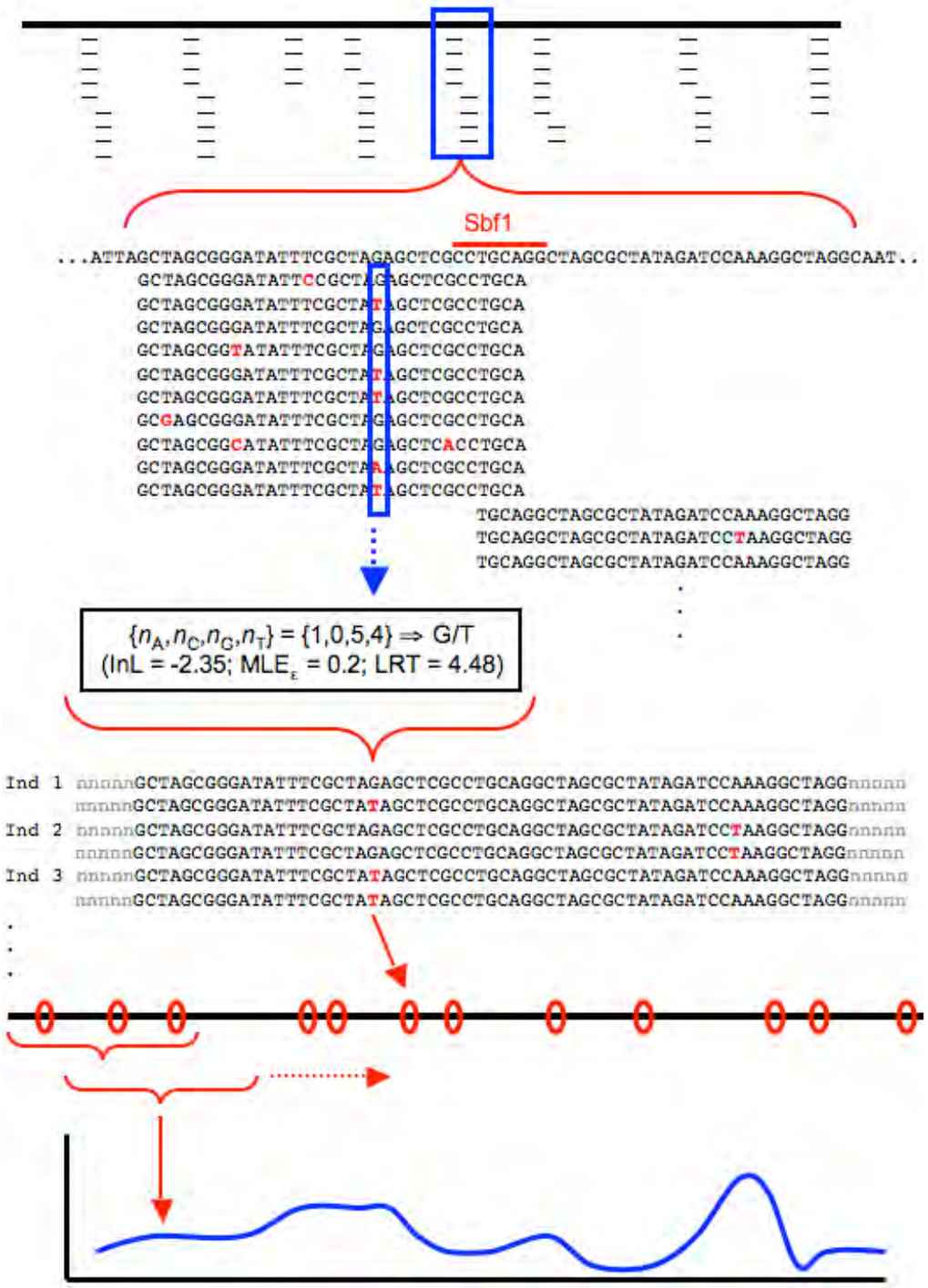
Making statistics continuous across the genome

Kernel-smoothing average of summary statistics along genome

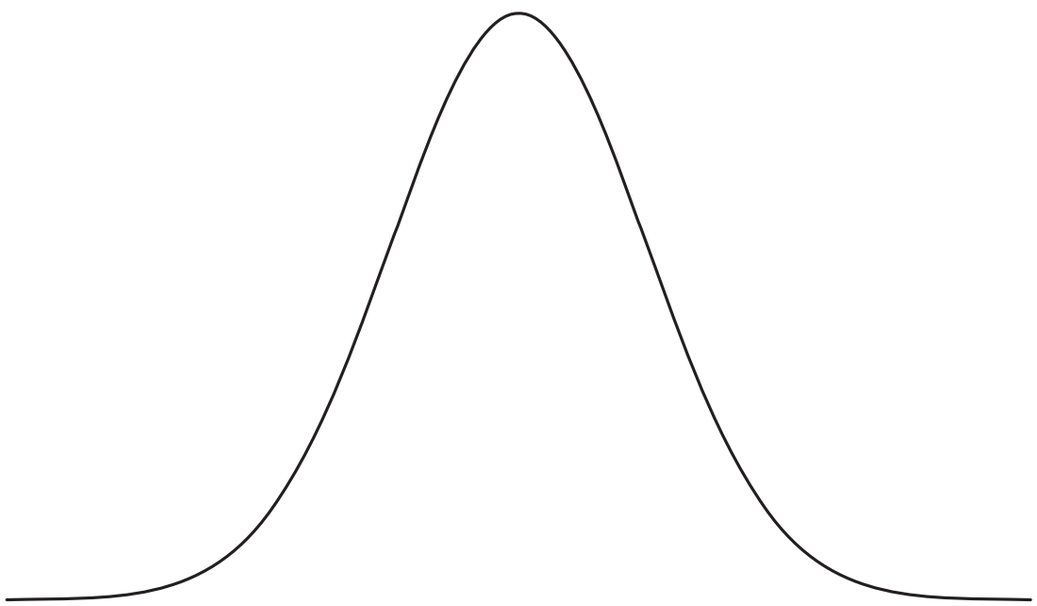


Bootstrap re-sampling to estimate significance of moving average

Overall pipeline



‘Bias’ in RAD-sequencing

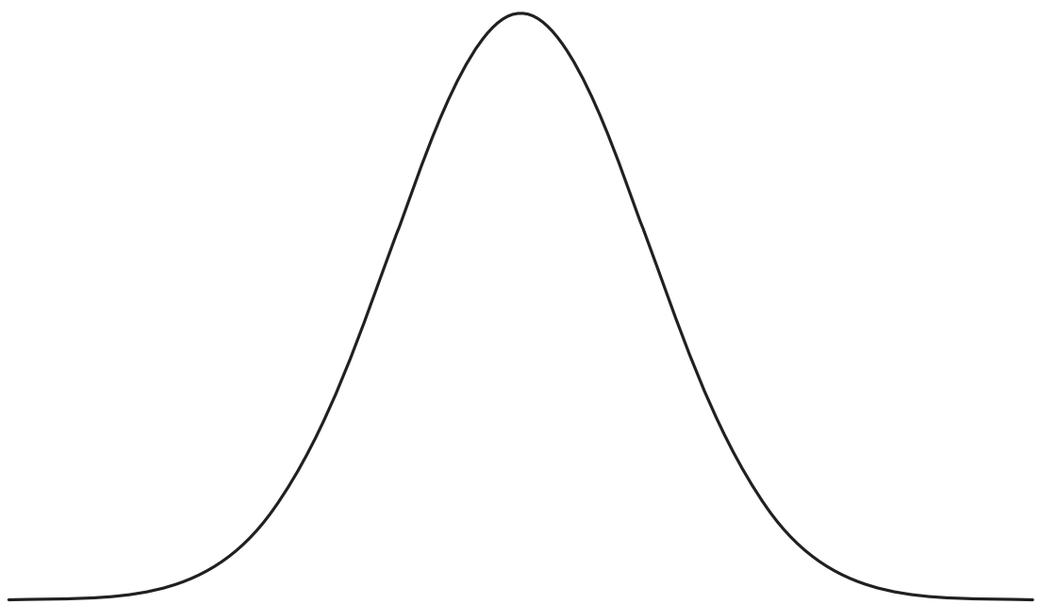


$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$e = 2.7182....$$

$$\pi = 3.1415....$$

‘Bias’ in RAD-sequencing



$$\bar{x} = \frac{1}{n}\sum_{i-1}^{n} x_i$$

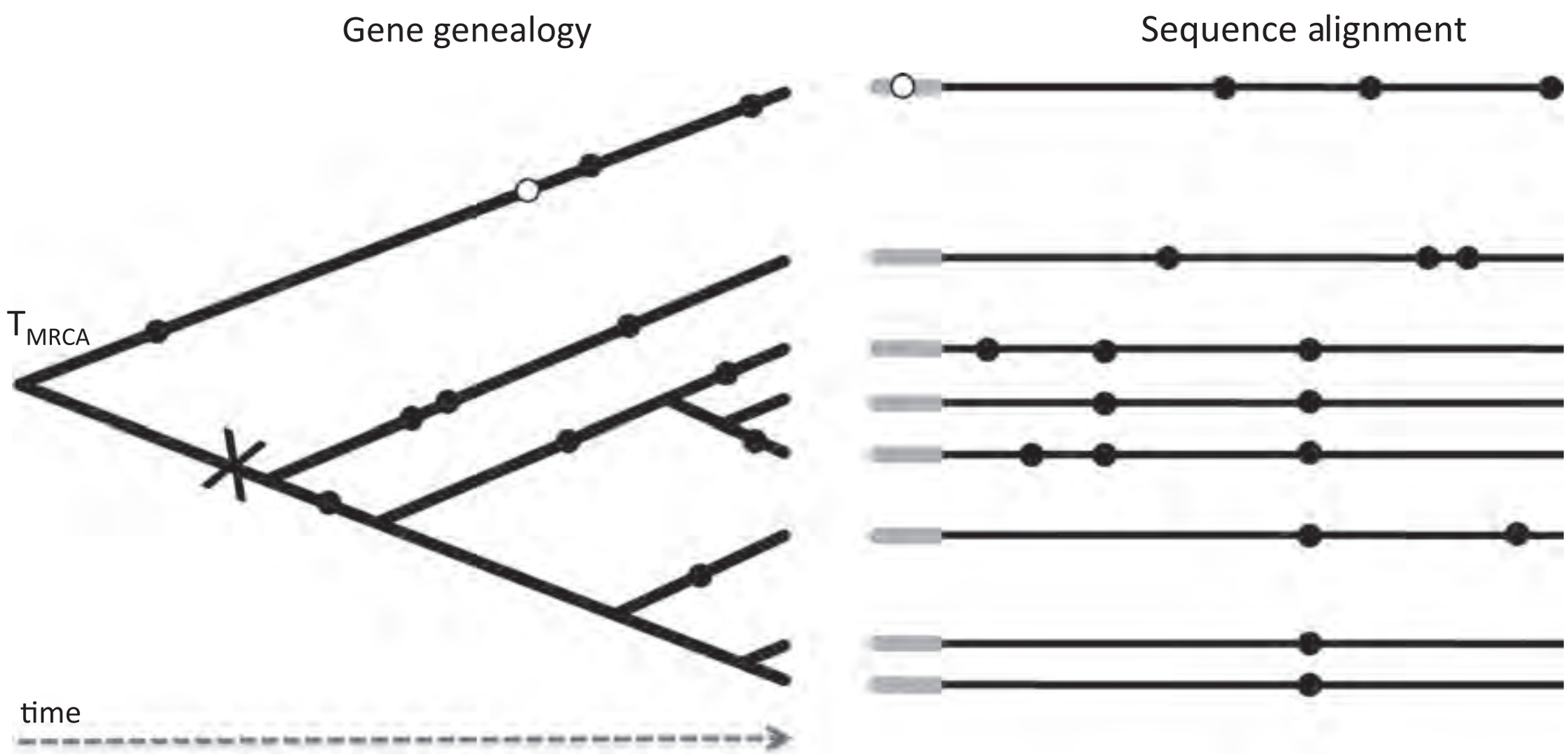
$$s^2 = \frac{1}{n-1}\sum_{i-1}^{n}(y_i - \bar{y})^2$$

Bias in RAD-sequencing

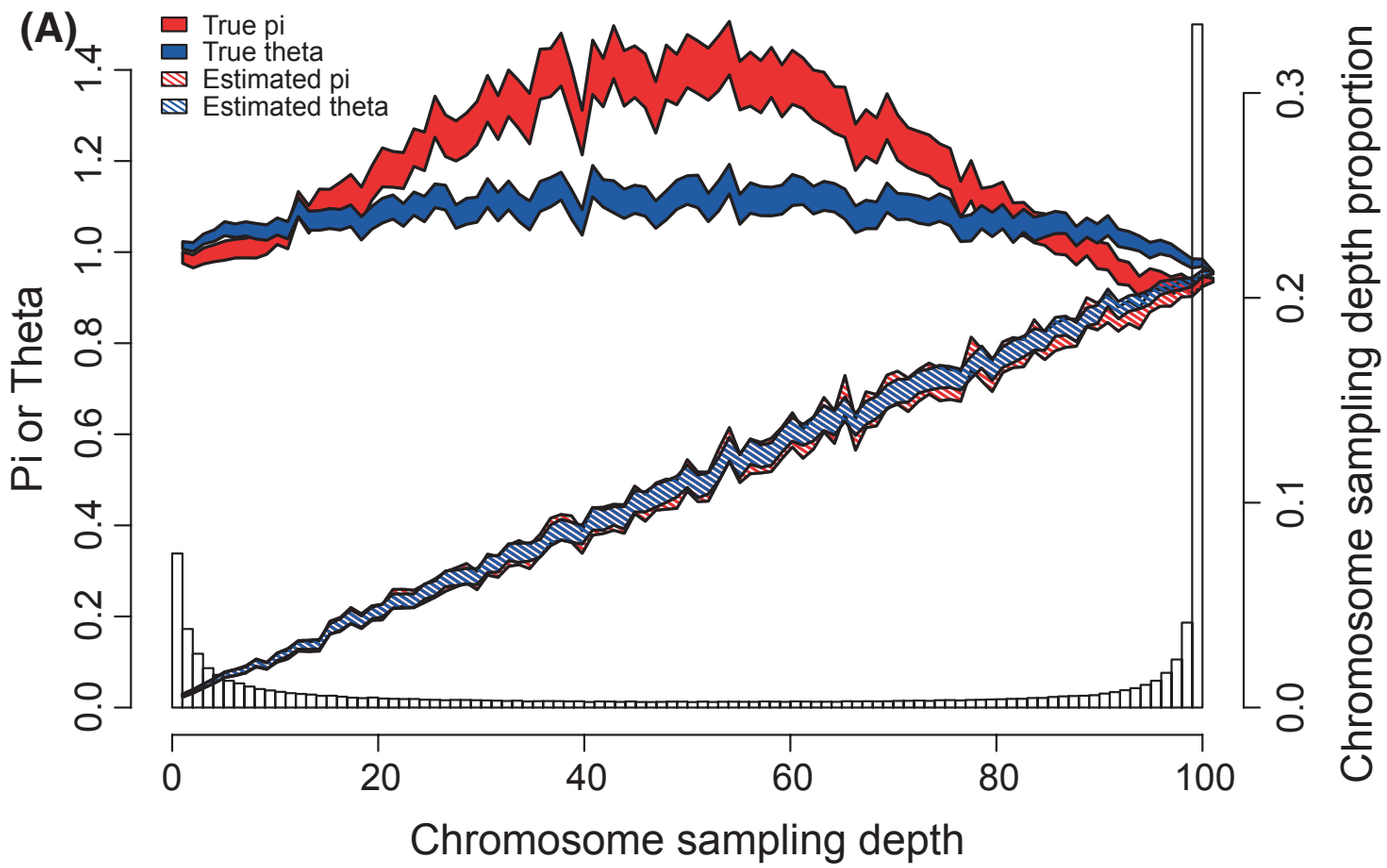
MOLECULAR ECOLOGY
Molecular Ecology (2013) **22**, 3179–3190 doi: 10.1111/mec.12

RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling

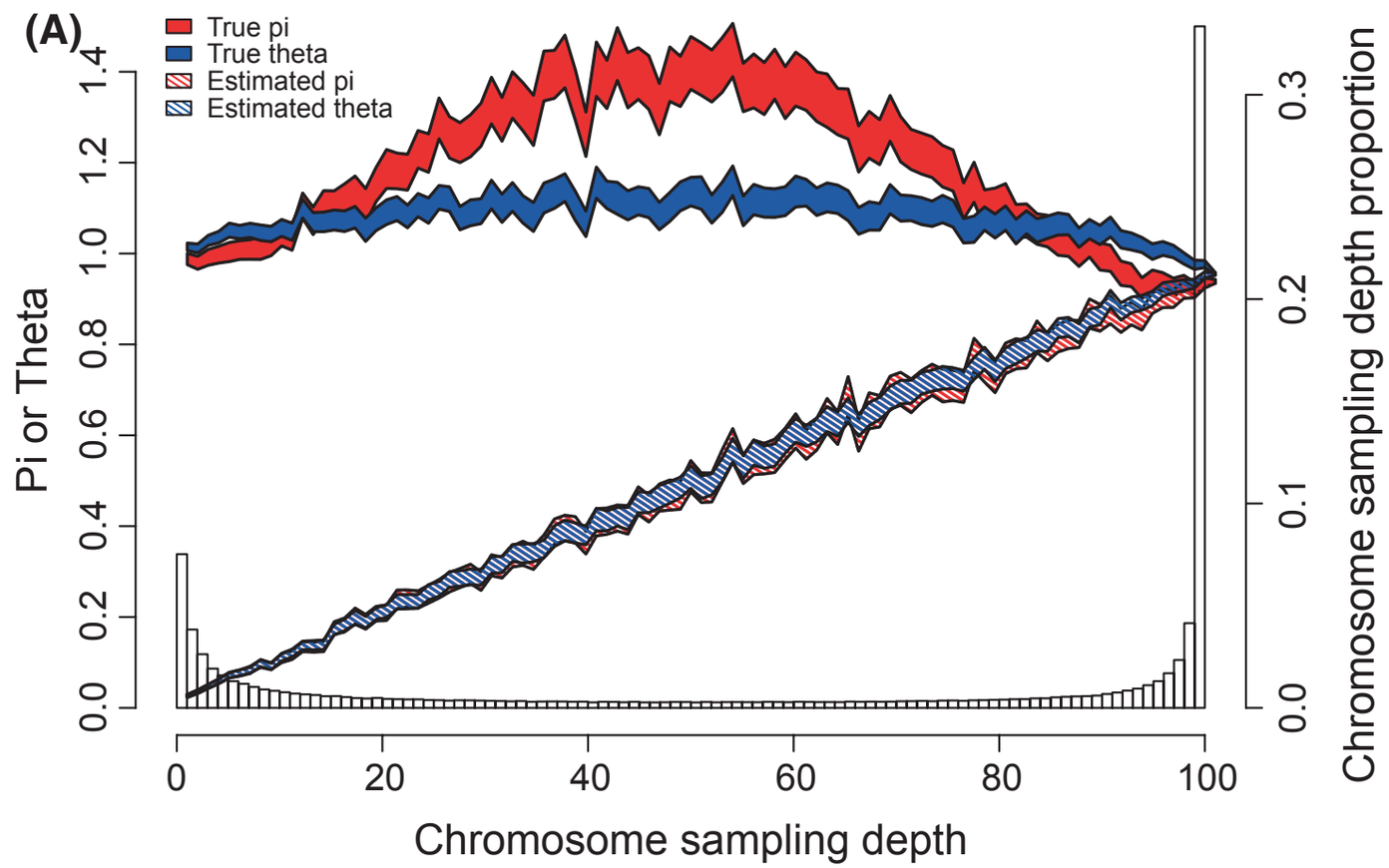
B. ARNOLD,[1] R. B. CORBETT-DETIG,[1] D. HARTL and K. BOMBLIES
Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

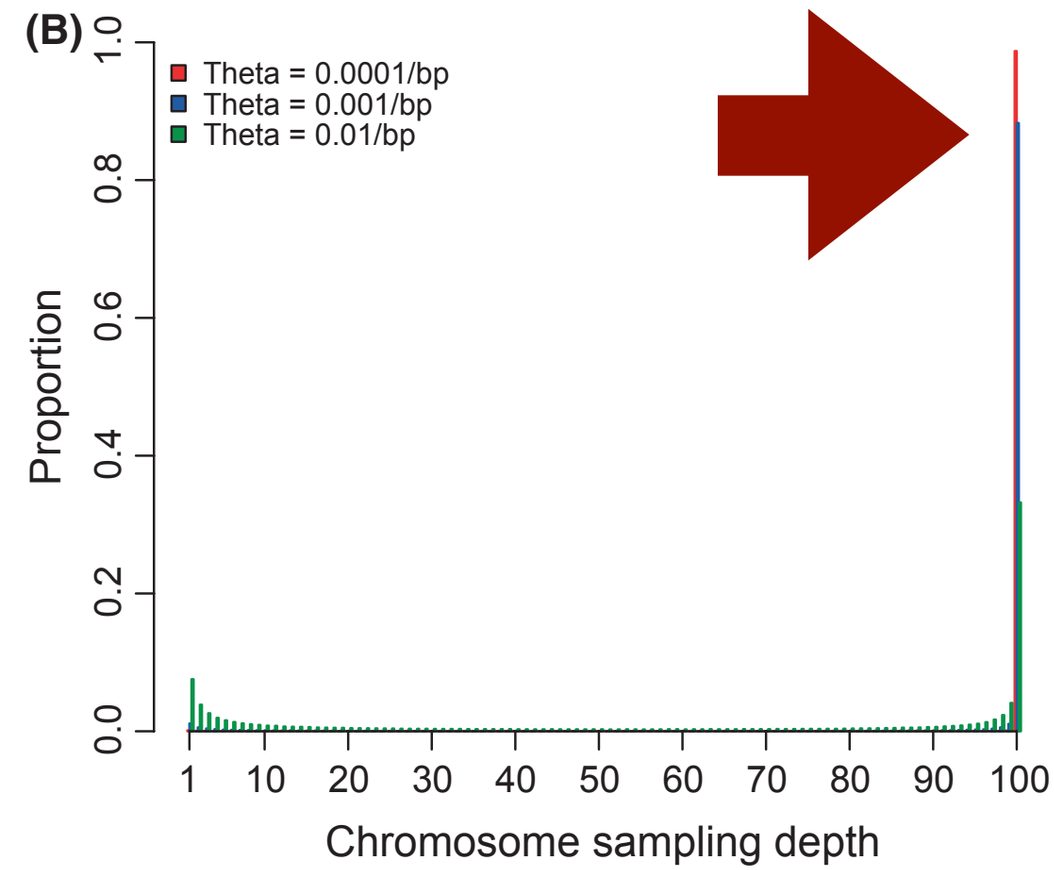


Bias in RAD-sequencing; genetic diversity

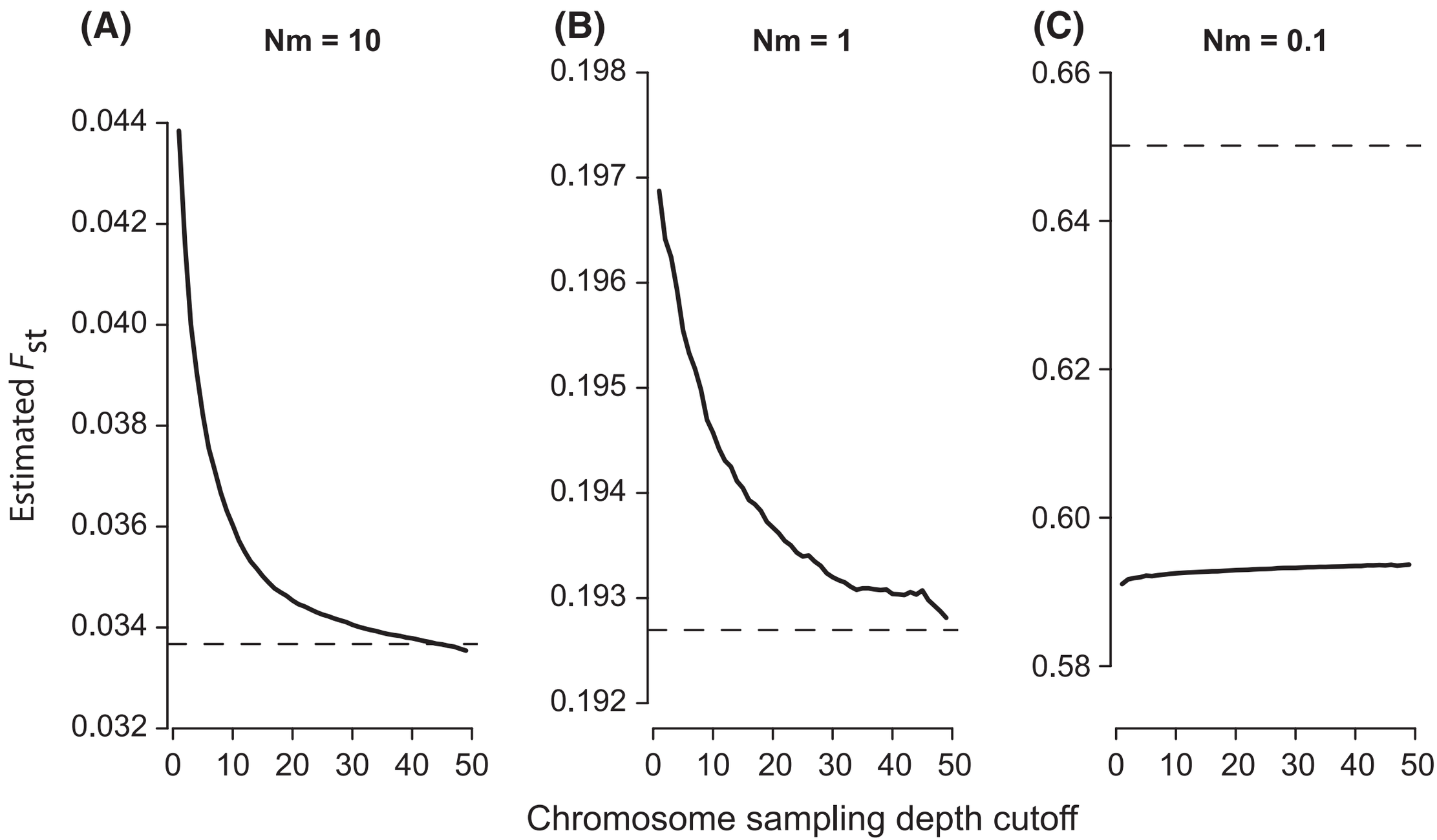


Bias in RAD-sequencing; genetic diversity





Bias in RAD-sequencing; Fst



Bias in RAD-sequencing summary

		Mean	
		Recombination	
Protocol	θ per bp	θ_{we}/θ_{wa}	π_e/π_a
Standard	0.0001	0.994	0.995
	0.001	0.987	0.982
	0.01	0.956	0.933
Double digest	0.0001	0.835	0.836
	0.001	0.858	0.851
	0.01	0.829	0.797

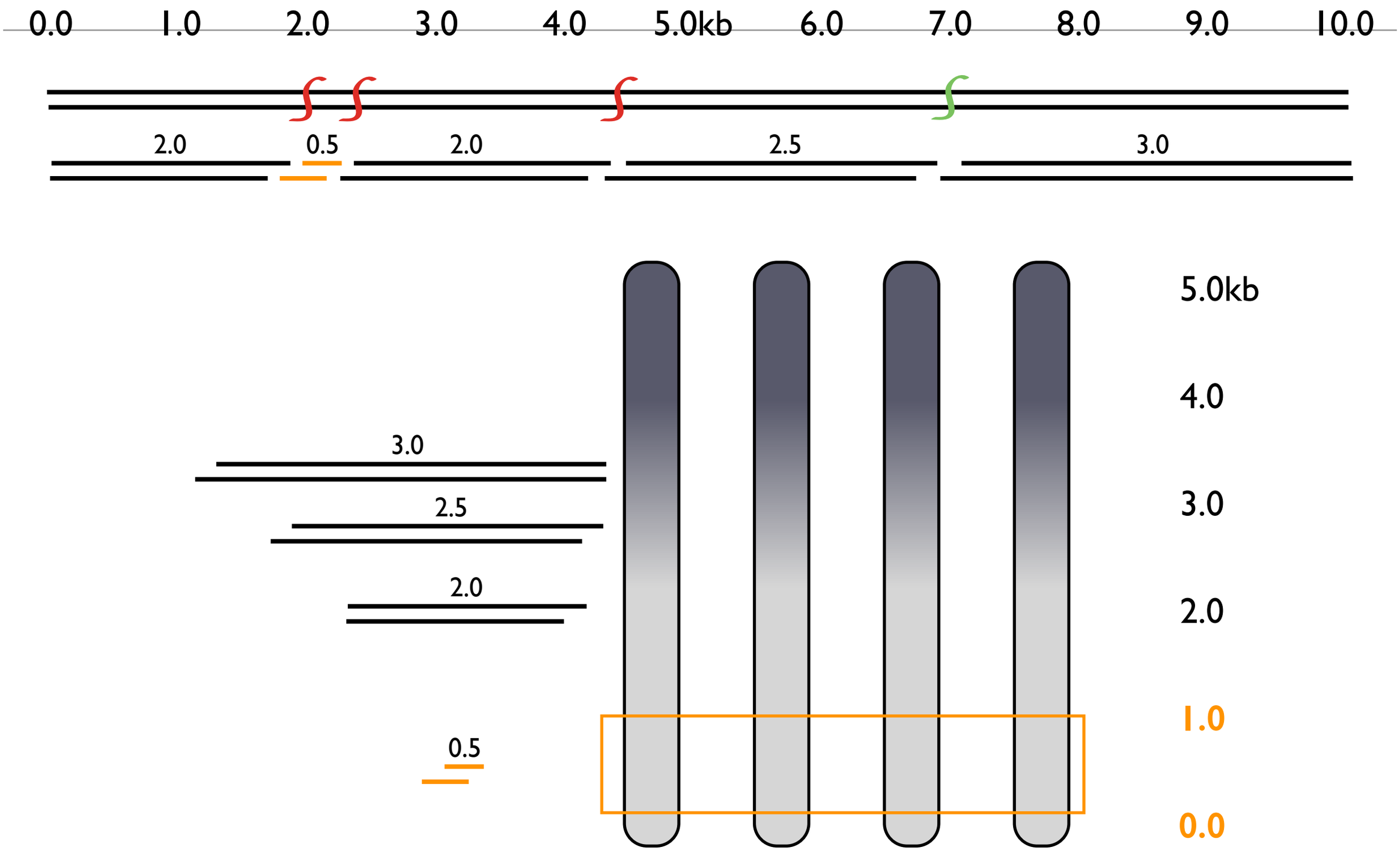
Bias in RAD-sequencing summary

		Mean	
		Recombination	
Protocol	θ per bp	θ_{we}/θ_{wa}	π_e/π_a
Standard	0.0001	0.994	0.995
	0.001	0.987	0.982
	0.01	0.956	0.933
Double digest	0.0001	0.835	0.836
	0.001	0.858	0.851
	0.01	0.829	0.797

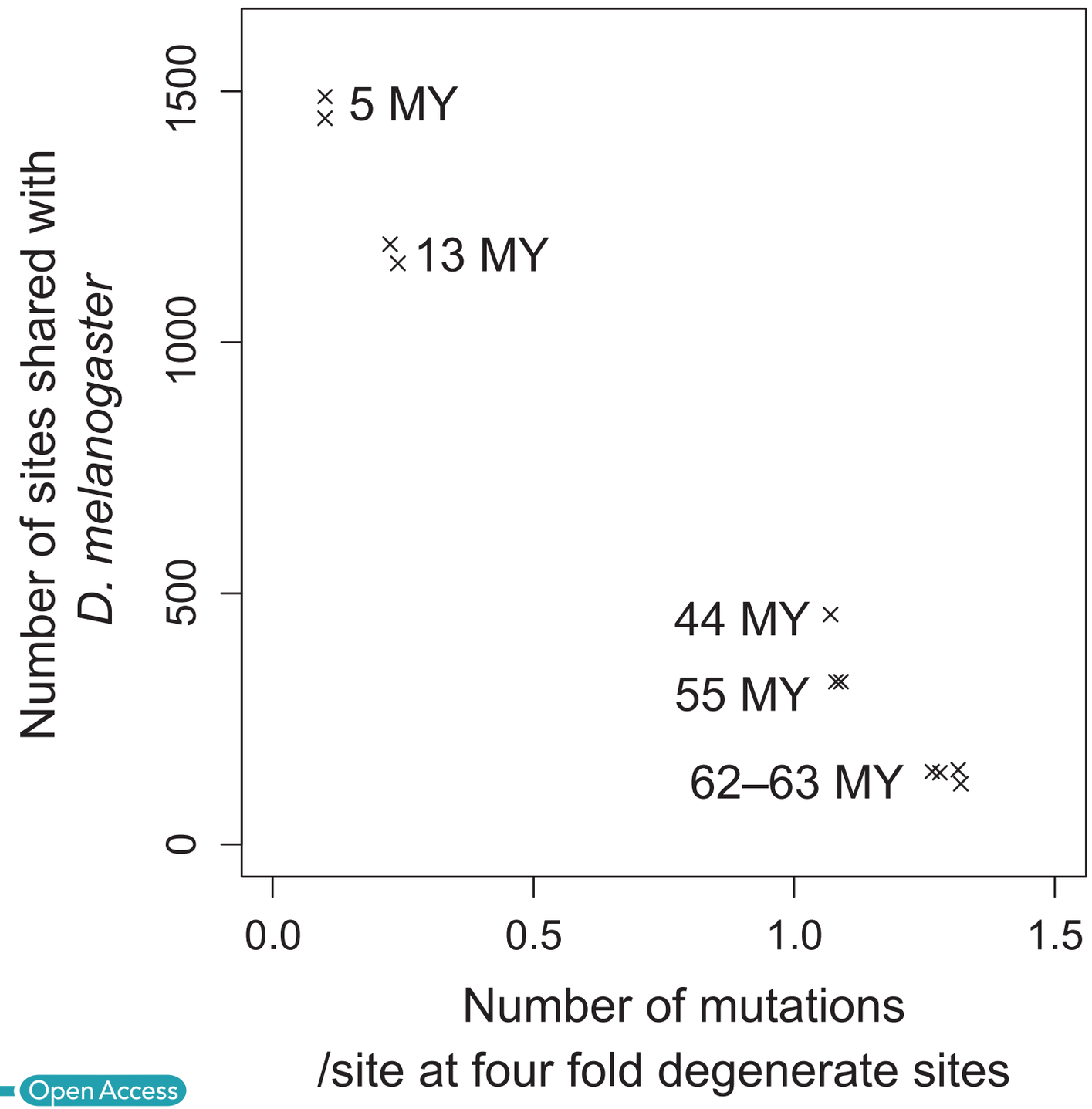
Bias in RAD-sequencing summary

		Mean	
		Recombination	
Protocol	θ per bp	θ_{we}/θ_{wa}	π_e/π_a
Standard	0.0001	0.994	0.995
	0.001	0.987	0.982
	0.01	0.956	0.933
Double digest	0.0001	0.835	0.836
	0.001	0.858	0.851
	0.01	0.829	0.797

Why is ddRAD so much more biased?



RAD-seq and phylogenetics of divergent species



Ecology and Evolution

Open Access

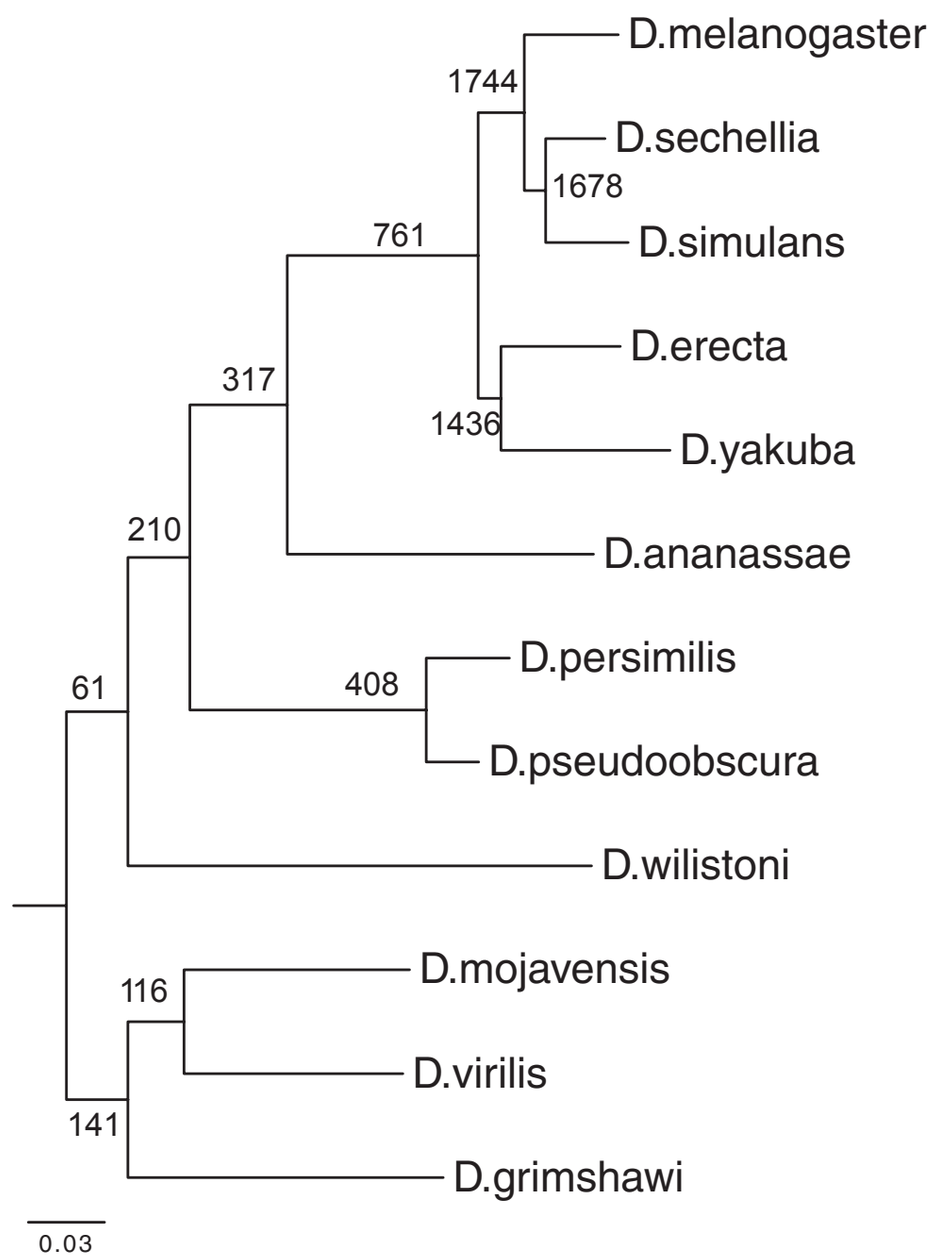
Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization

Marie Cariou, Laurent Duret & Sylvain Charlat

RAD-seq and phylogenetics of divergent species

Species pair *D. melanogaster*	Node depth (My)	Orthologous tags	Retrieved orthologous tags (%)	In clusters including paralogs (%)
D.sechellia	5.4	2978	99	5
D.simulans	5.4	2892	99	4
D.erecta	12.6	2390	97	3
D.yakuba	12.8	2314	97	8
D.ananassae	44.2	916	68	9
D.persimilis	54.9	648	65	9
D.pseudoobscura	54.9	648	66	9
D.wilistoni	62.2	242	49	6
D.grimshawi	62.9	290	60	8
D.virilis	62.9	286	59	5
D.mojavensis	62.9	298	59	8

RAD-seq and phylogenetics



Ecology and Evolution

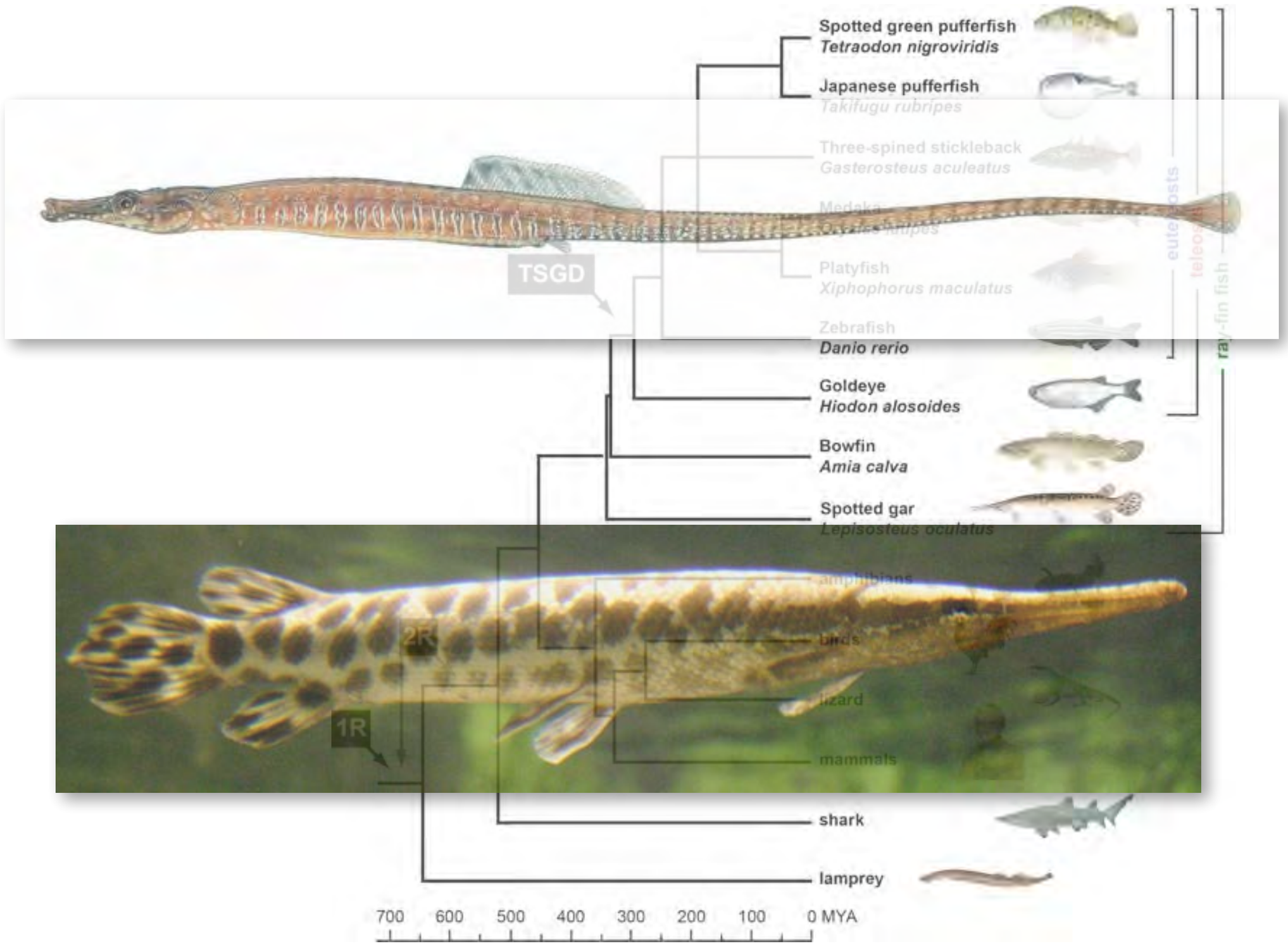
Open Access

Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization

Marie Cariou, Laurent Duret & Sylvain Charlat

What if you don’t have a genome sequence?

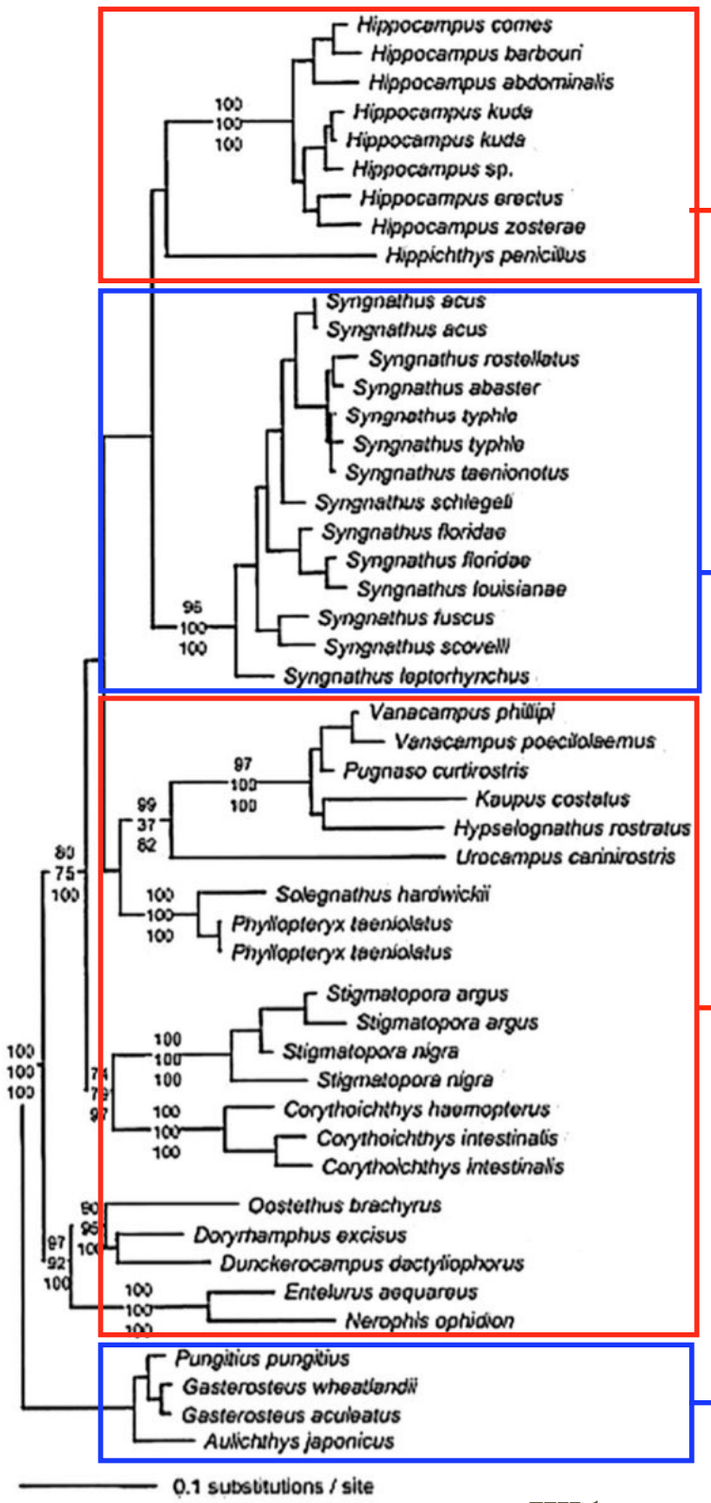
Genomically enabling very non-model organisms:
RAD-seq can help



Julian Catchen, Allison Fuiten, Susie Bassham, Clay Small and Adam Jones

Seahorses, sea dragons and pipefishes

Gasterosteidae and Syngnathidae are historically considered to be closely related



Seahorses

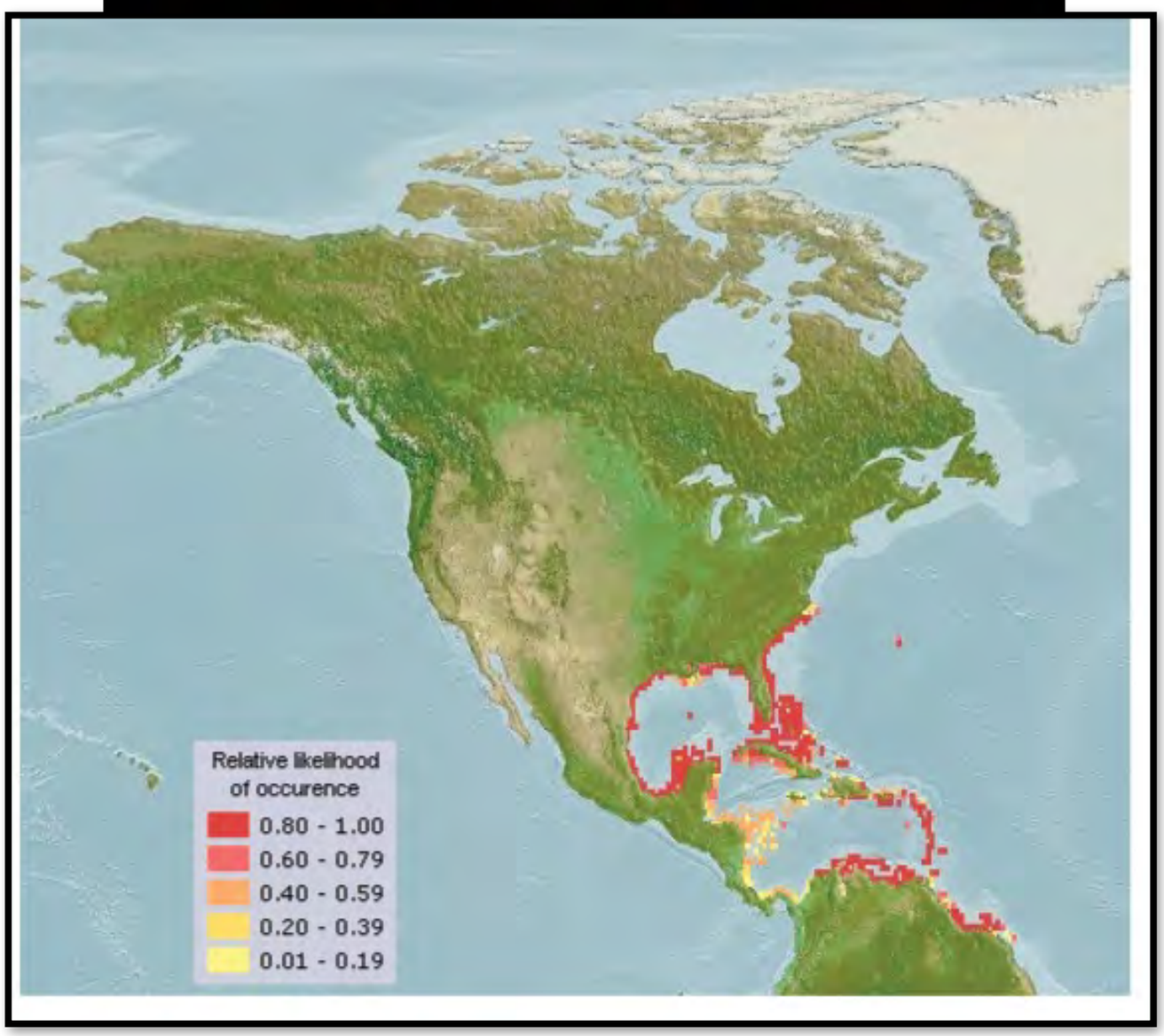
Pipefish

Seadragons

Stickleback

Wilson et al. 2003

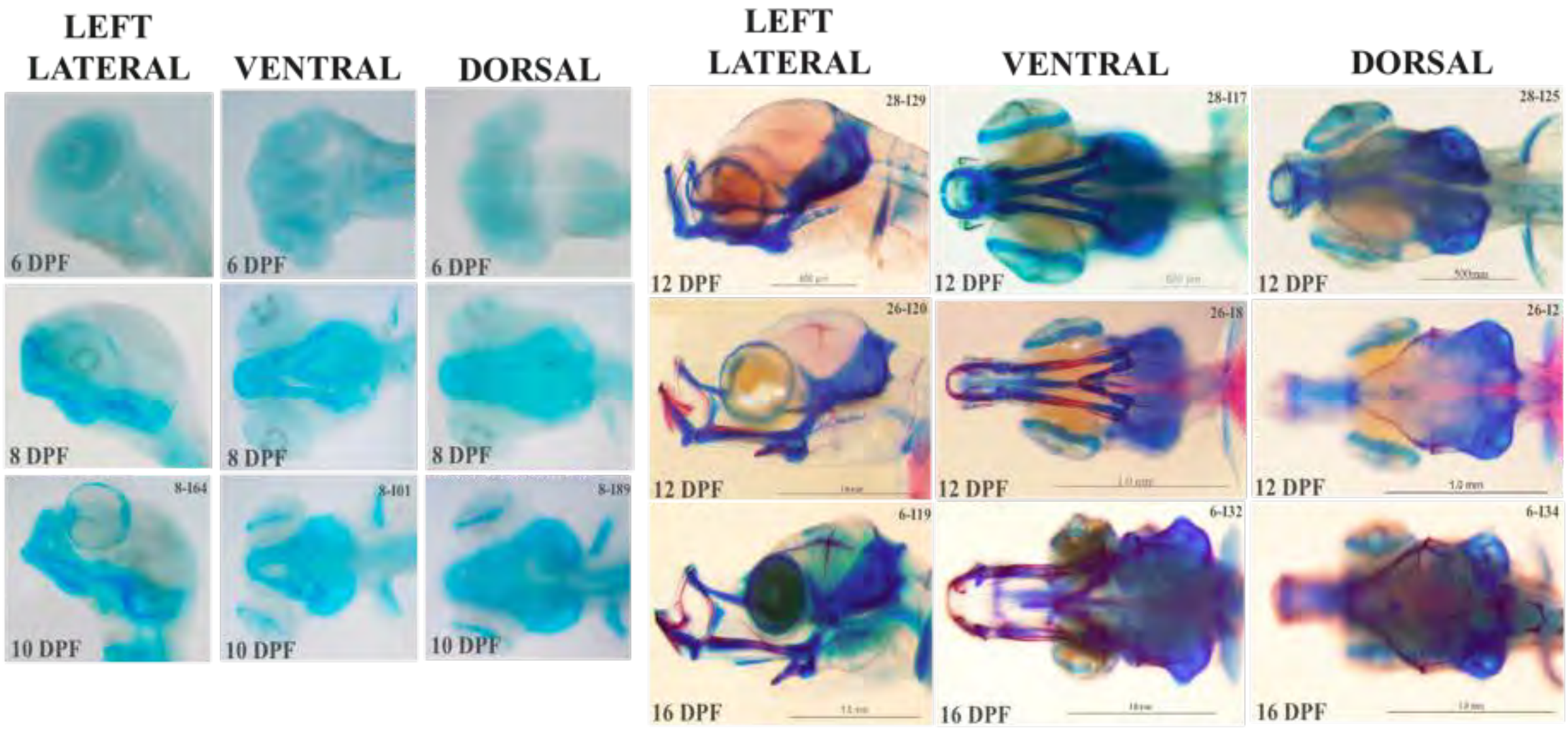
Gulf Pipefish
Syngnathus scovelli



-160 mm (6.3”)
-reversed sex roles
-sexual dimorphism
-specialized
suction feeding
-no sequences in
international databases

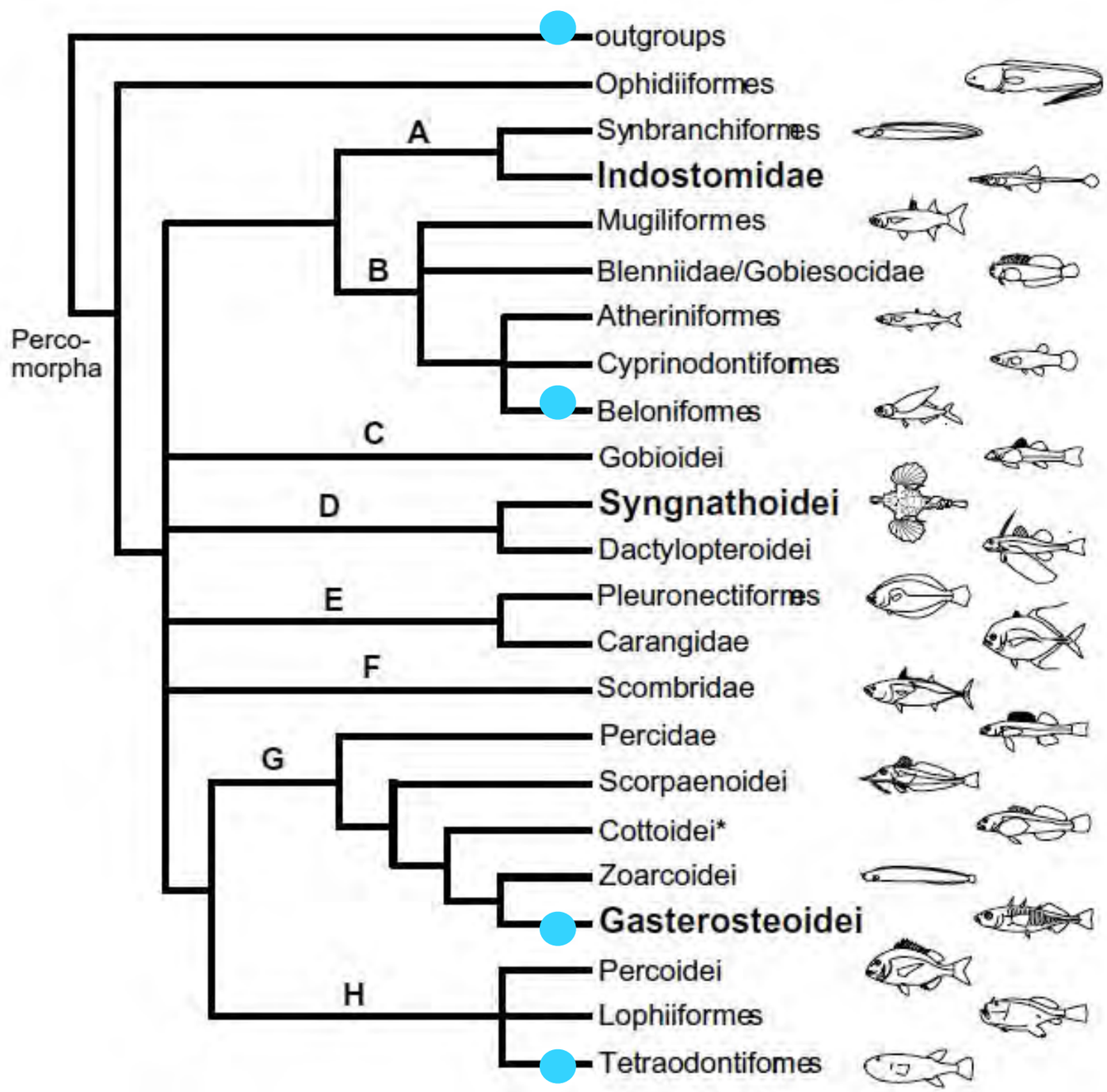
www.bio.tamu.edu/USERS/ajones/charlyn.html
Fishbase.org

We're really interested in the head and body axis



Few teleost genomes are available

Gasterosteiformes: only stickleback



Zebrafish

Medaka

Stickleback

Two pufferfish

Solution: 'genomically enable' pipefish

1) A high quality transcriptome

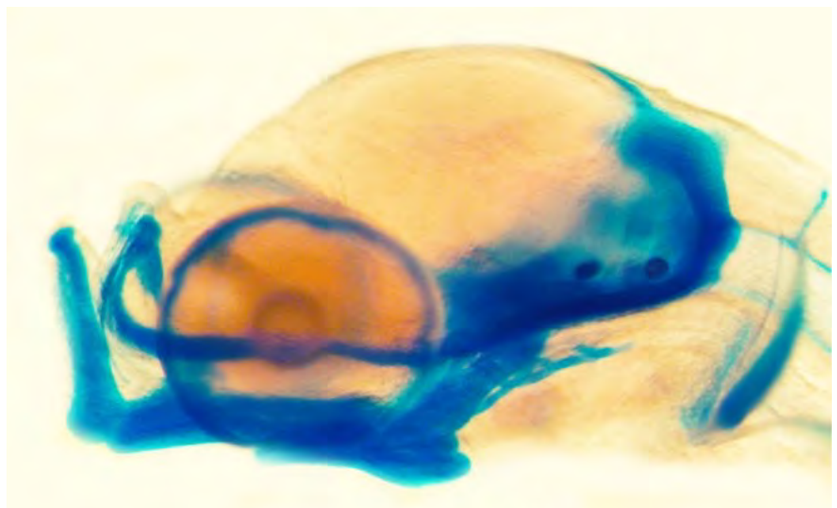
2) Very dense RAD genetic map

3) Deep coverage shotgun sequencing of genome

4) Order genomic and transcriptomic contigs against the RAD reference map

Pipefish Transcriptome

Building an EST database in pipefish



Pipefish embryonic mRNA

↓

Illumina sequencing:
100 nt, paired-end

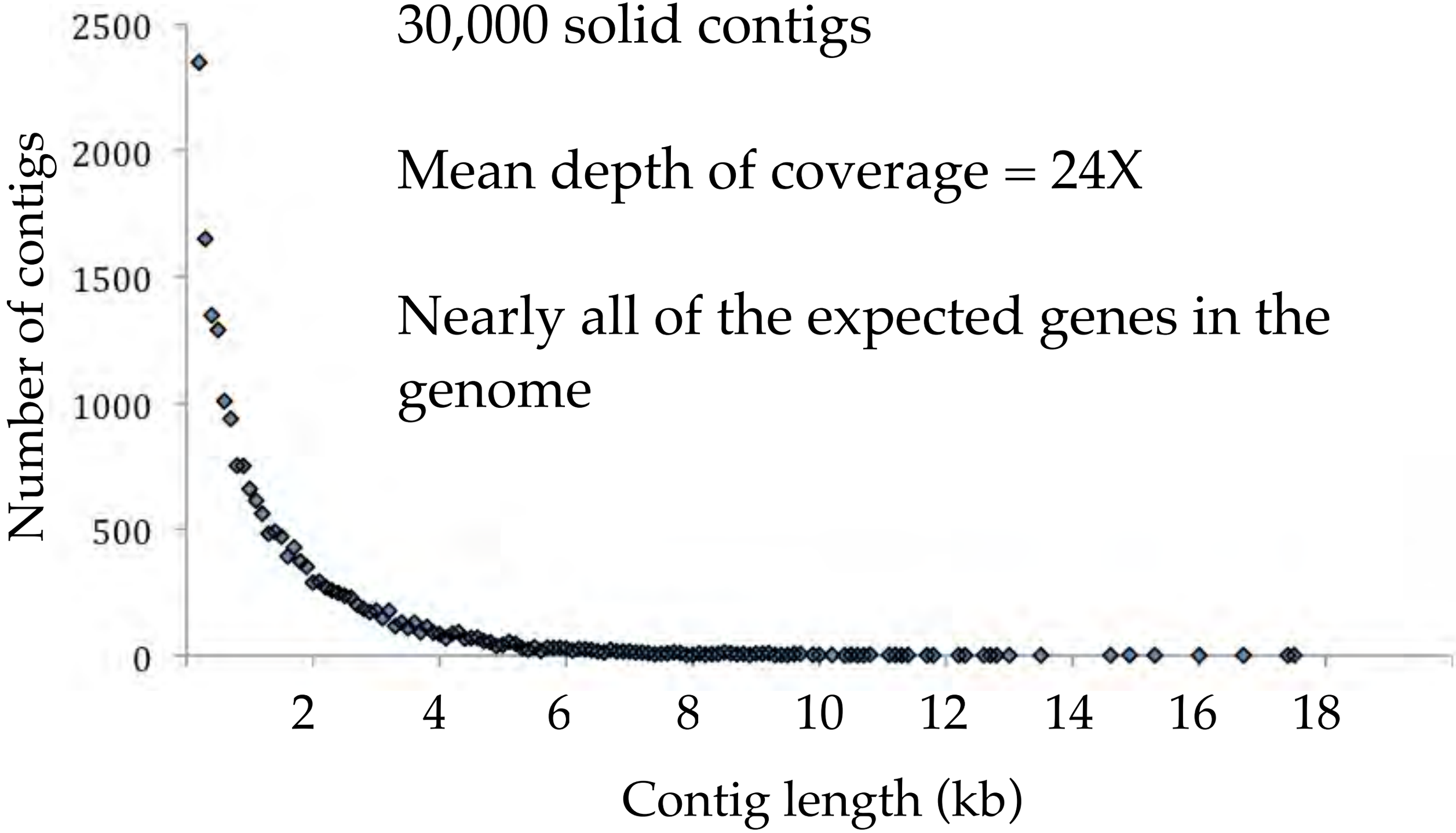
↓

200 million reads (two lanes)

↓

Assembly of transcripts

Transcriptome



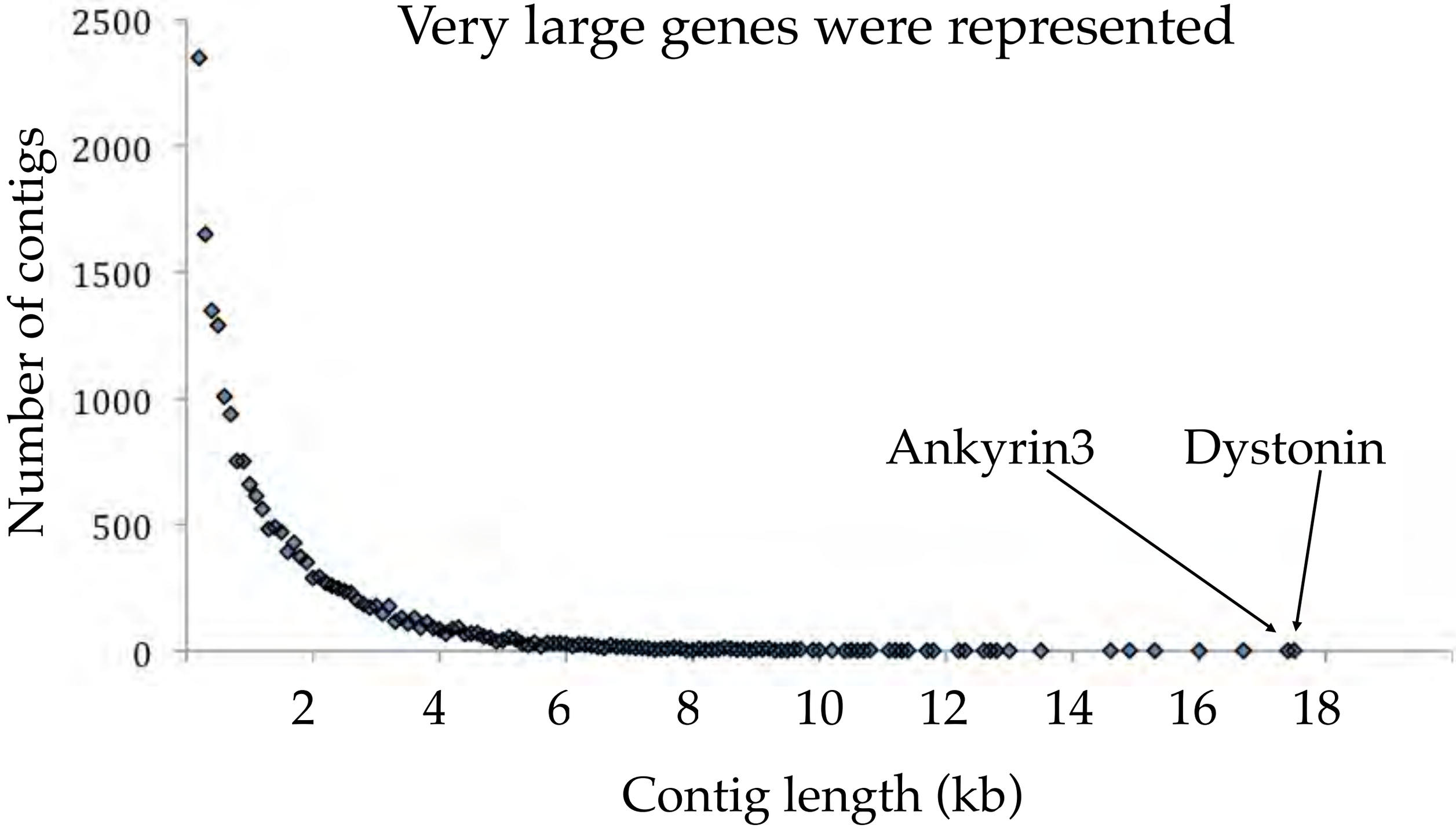
30,000 solid contigs

Mean depth of coverage = 24X

Nearly all of the expected genes in the genome

Transcriptome

Very large genes were represented



Pipefish Genetic Map

Genetic map workflow

Generated an F1 family of 103 individuals

RAD sequenced the parents and offspring

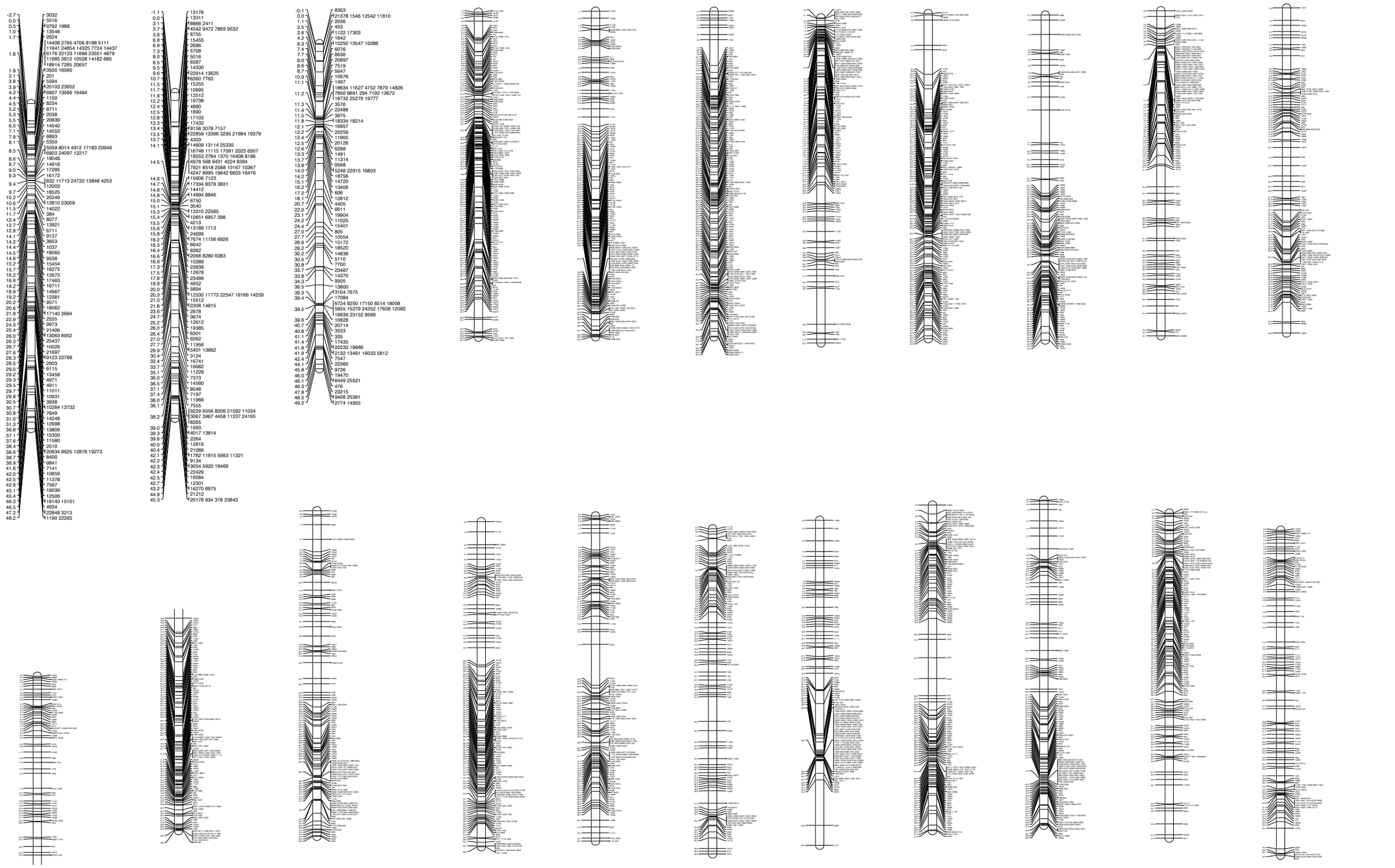
Analyzed the data using *Stacks*

Paired end local assemblies

Output to JoinMap format

Created Linkage map

The pipefish genetic map is closed; 22 LGs
6000 segregating SNPs; 30,000 RAD sites



Pipefish Genome Project

Genome workflow

Generated DNA from a single individual

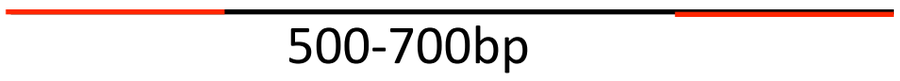
Random Illumina shotgun sequencing

Removed highly repetitive kmers

Produced *several* different genome assemblies

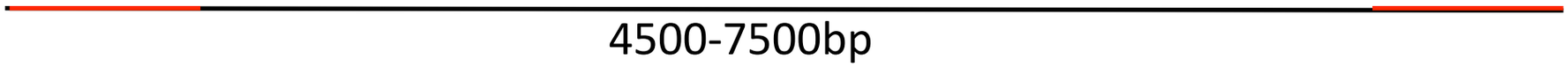
Illumina genomic libraries for pipefish genome

paired end 101bp



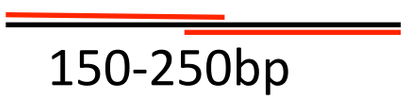
25x

mate pair



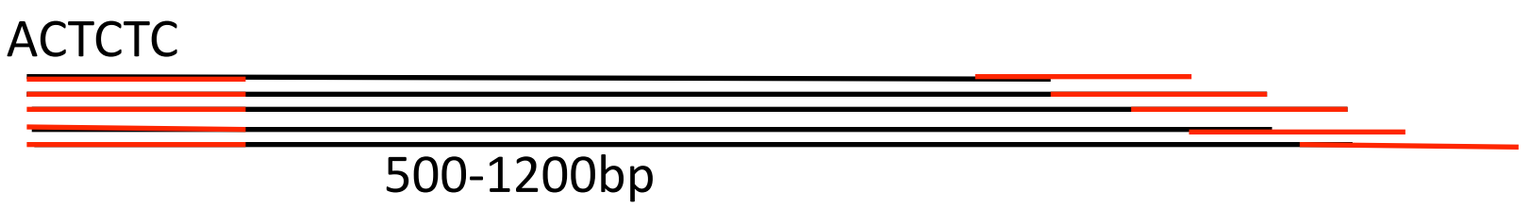
2x

overlapping



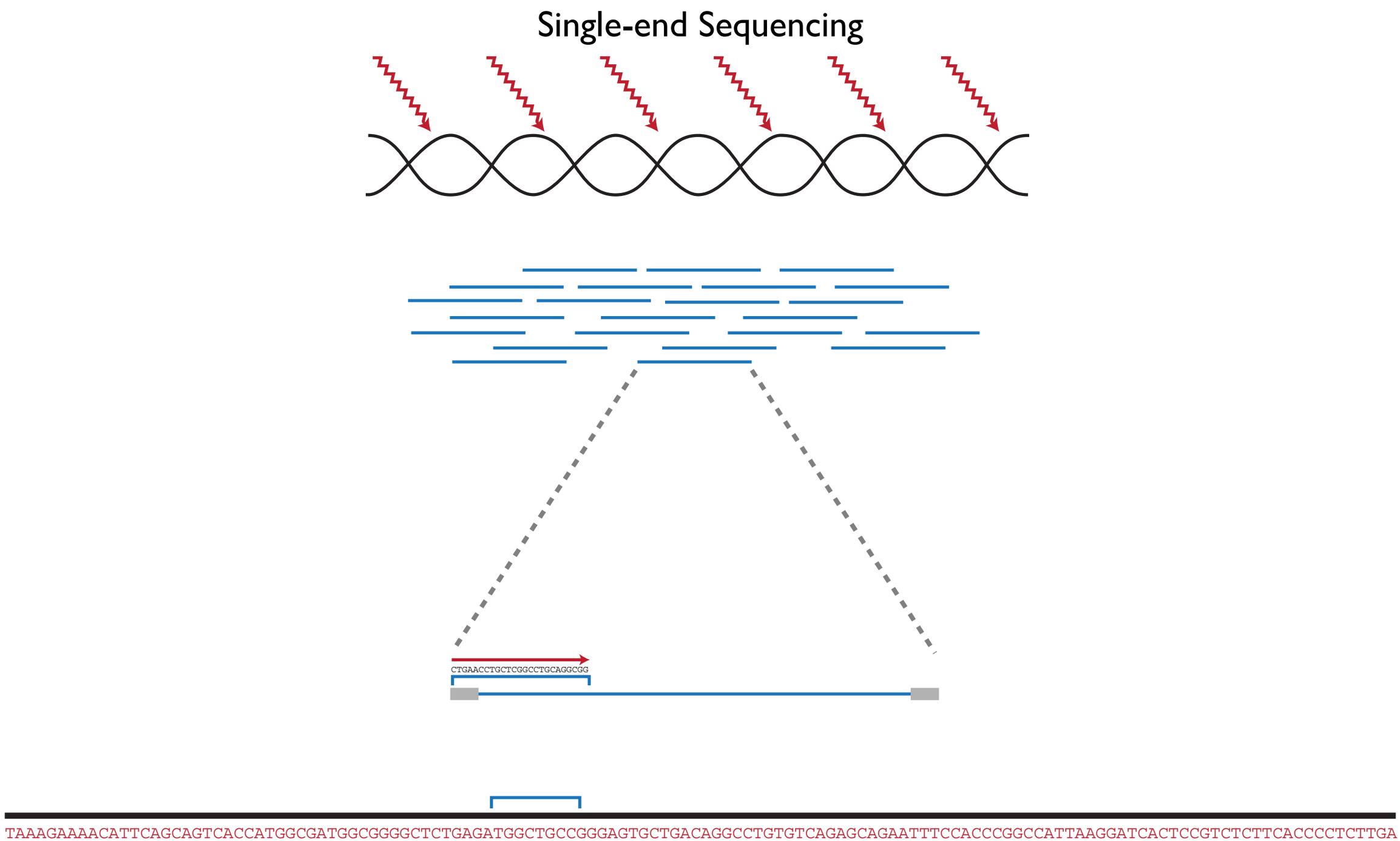
40x

paired end RAD

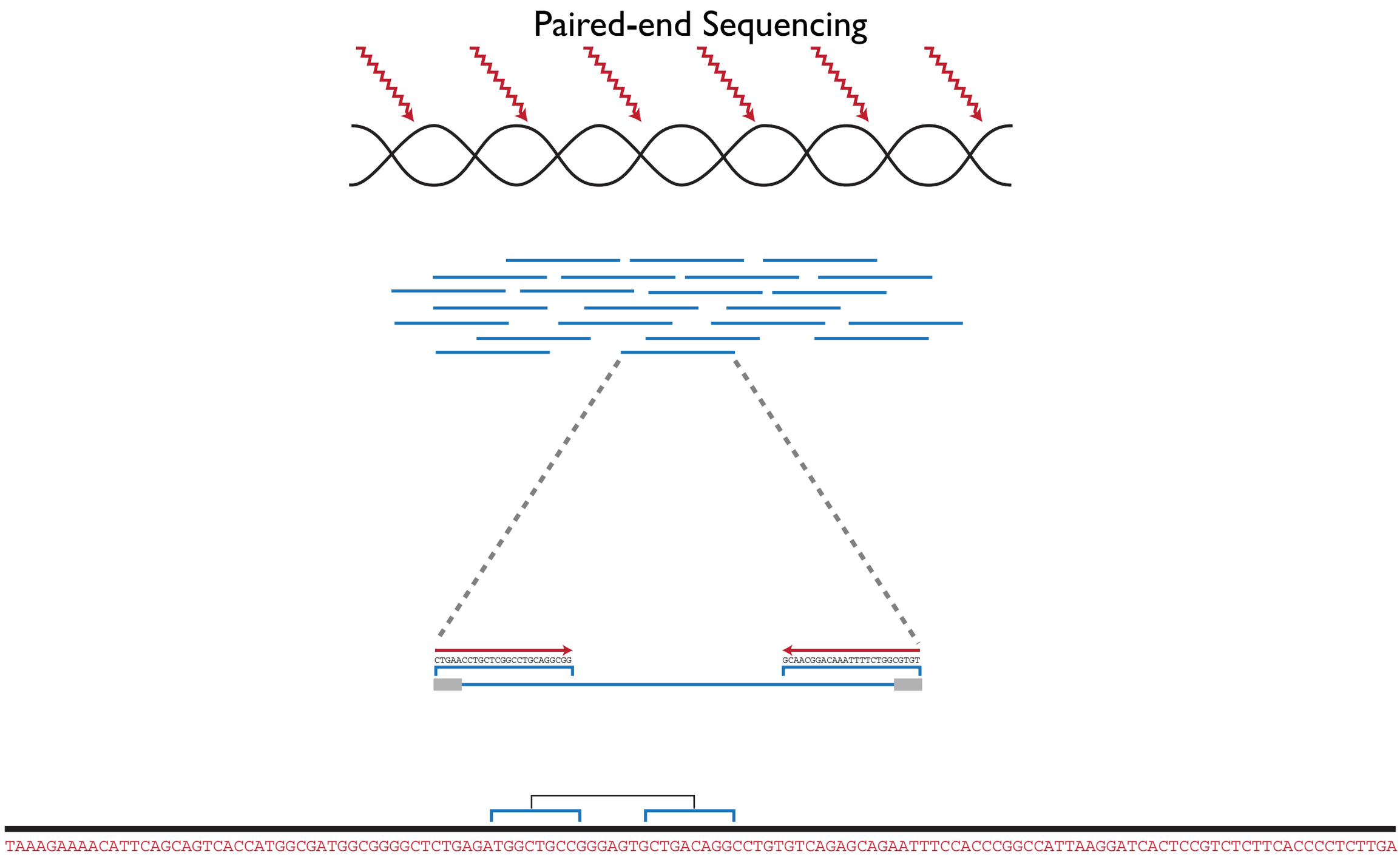


15-25x of 3% of the genome

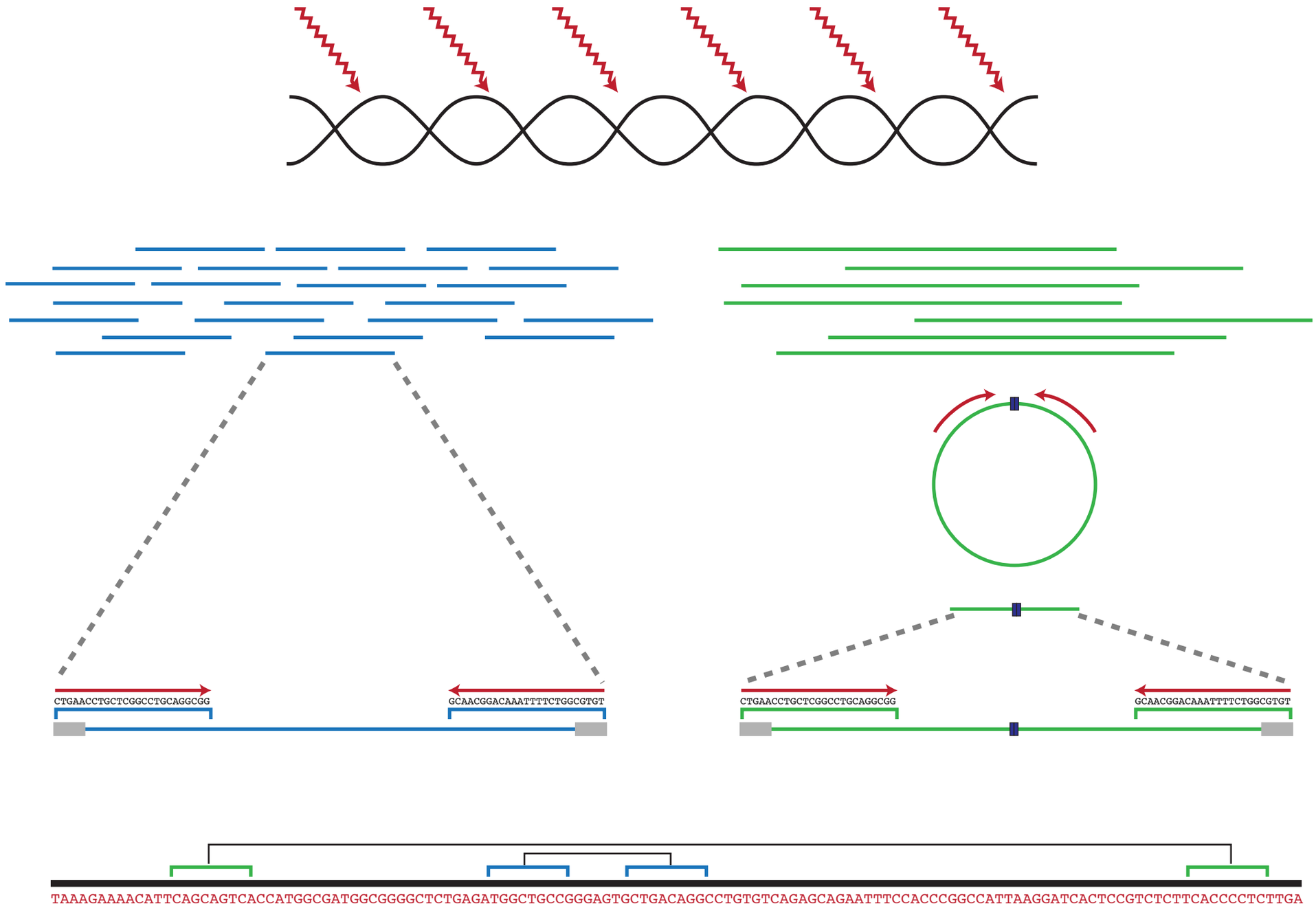
de novo Genome Sequencing



de novo Genome Sequencing



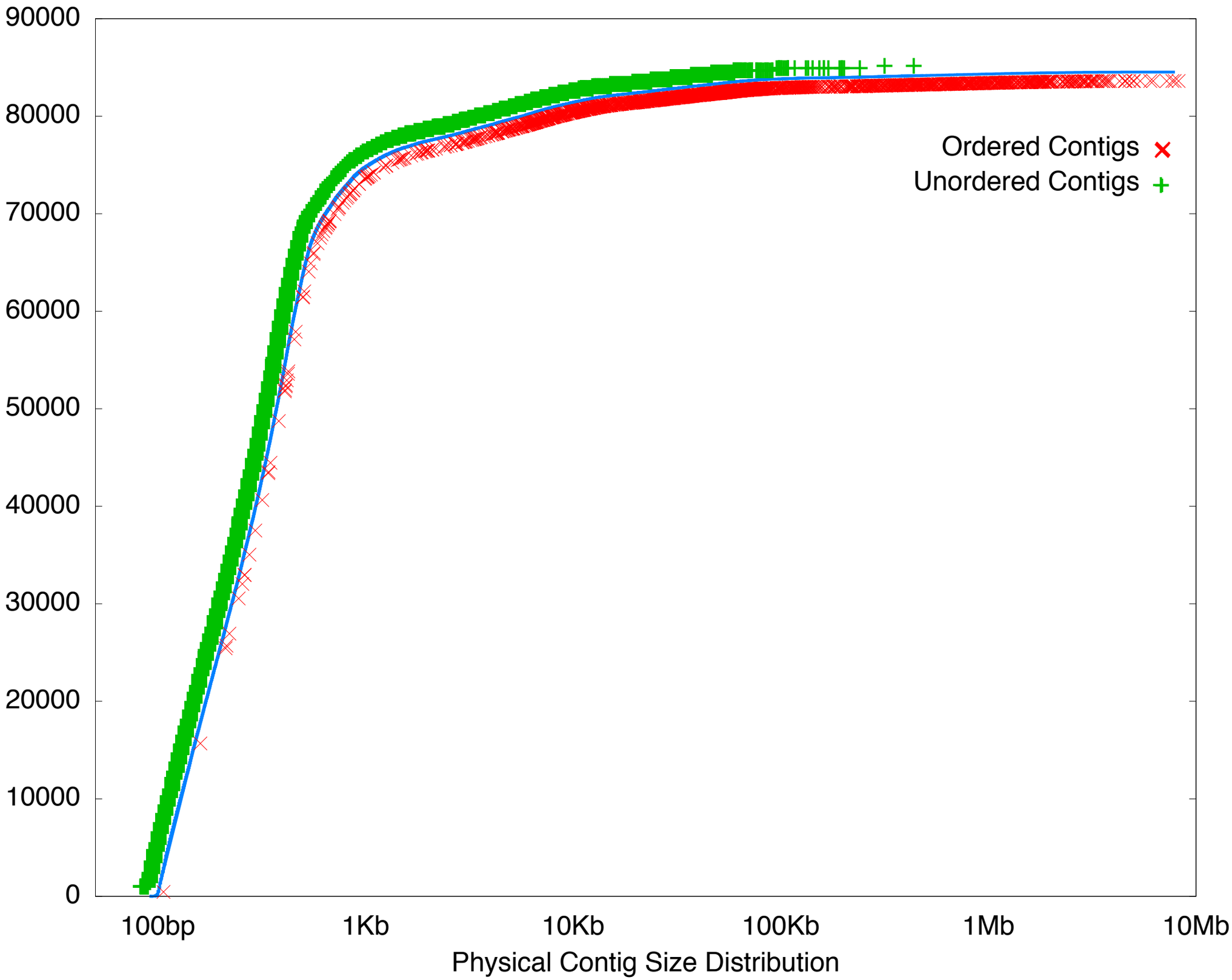
Mate-pair Sequencing



Pipefish genome assembly version 0.99
Nearly the whole genome is covered

Coverage	Scaffolds	Contigs	Scaffold N50	Contig N50
All (66.6x)	1,820	25,075	796,183	23,012

Max	Average Length	Total Length	Gap Length	%
5,676,603	162,928	296,529,585	27,303,839	(8.39%)


90000
80000
70000
60000
50000
40000
30000
20000
10000
0
100bp
1Kb
10Kb
100Kb
1Mb
10Mb
Physical Contig Size Distribution
Ordered Contigs
Unordered Contigs

Overall Conclusions

Genomics can be a tool for enabling new research in models & nonmodels

- RAD-seq can be used for SNP identification and genotyping
- documenting patterns of genetic variation
- identifying the molecular genetic basis of important phenotypic variation
- assessing how ecological processes structure this genetic variation in genomes
- analytical and computational approaches are challenging but manageable

Not your father's genome assembly

- a mixture of data types can be efficiently combined
- a genetic map is extremely useful for pulling it all together
- having a tiled genome is good enough - it doesn't have to be completely closed

Open Source Genomics provides a suite of breakthrough technologies

- the molecular approaches are not as daunting as they first appear
- analytical and computational approaches are challenging
  - **New software tools can help, but knowledge of Unix and Scripting is essential**
  - **Also essential to be comfortable with classical and modern statistics**

Acknowledgments

- *Past and present lab members* **Paul Hohenlohe**, **Thom Nelson**, Joe Dunham, Nicole Nishimura & **Mark Currey**
- *Collaborators* **Eric Johnson**, Patrick Phillips, **Chuck Kimmel**, **John Postlethwait**
- *Funding* from NSF & NIH, as well as Keck & Murdock Foundations

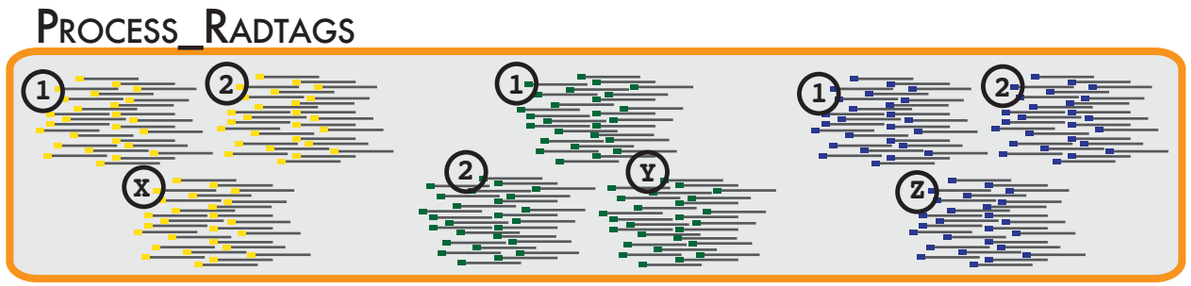
TUTORIAL - USING *STACKS*

G3: Genes, Genomes, Genetics

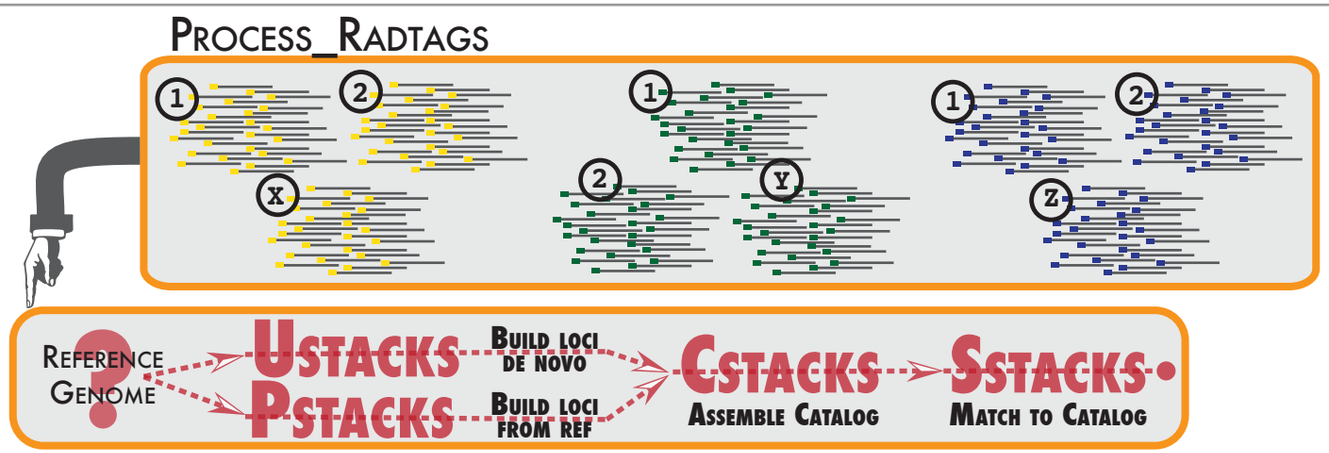
***Stacks*: Building and Genotyping Loci *De Novo* From Short-Read Sequences**

Julian M. Catchen,* Angel Amores,[1] 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

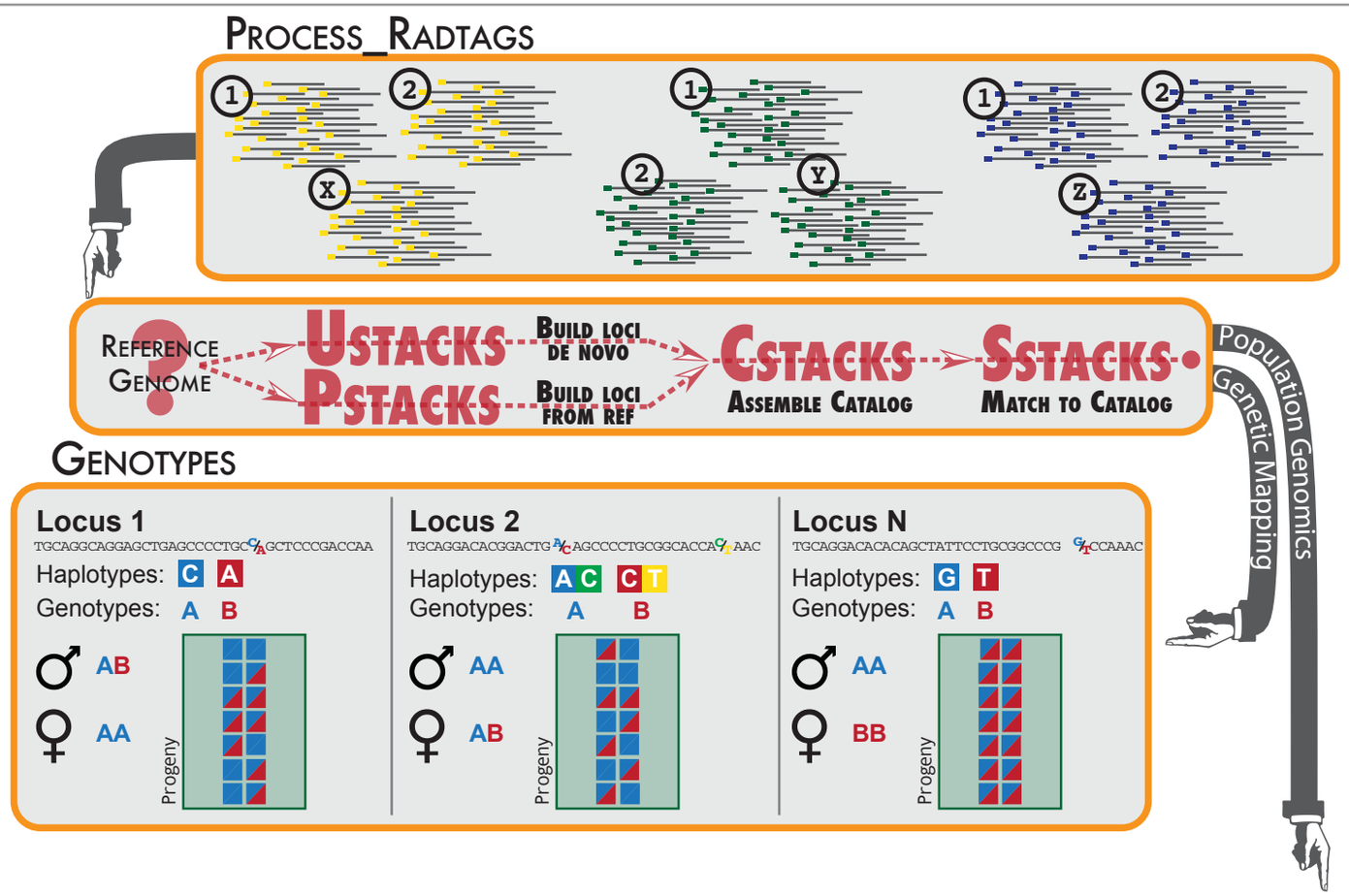
Stacks workflow



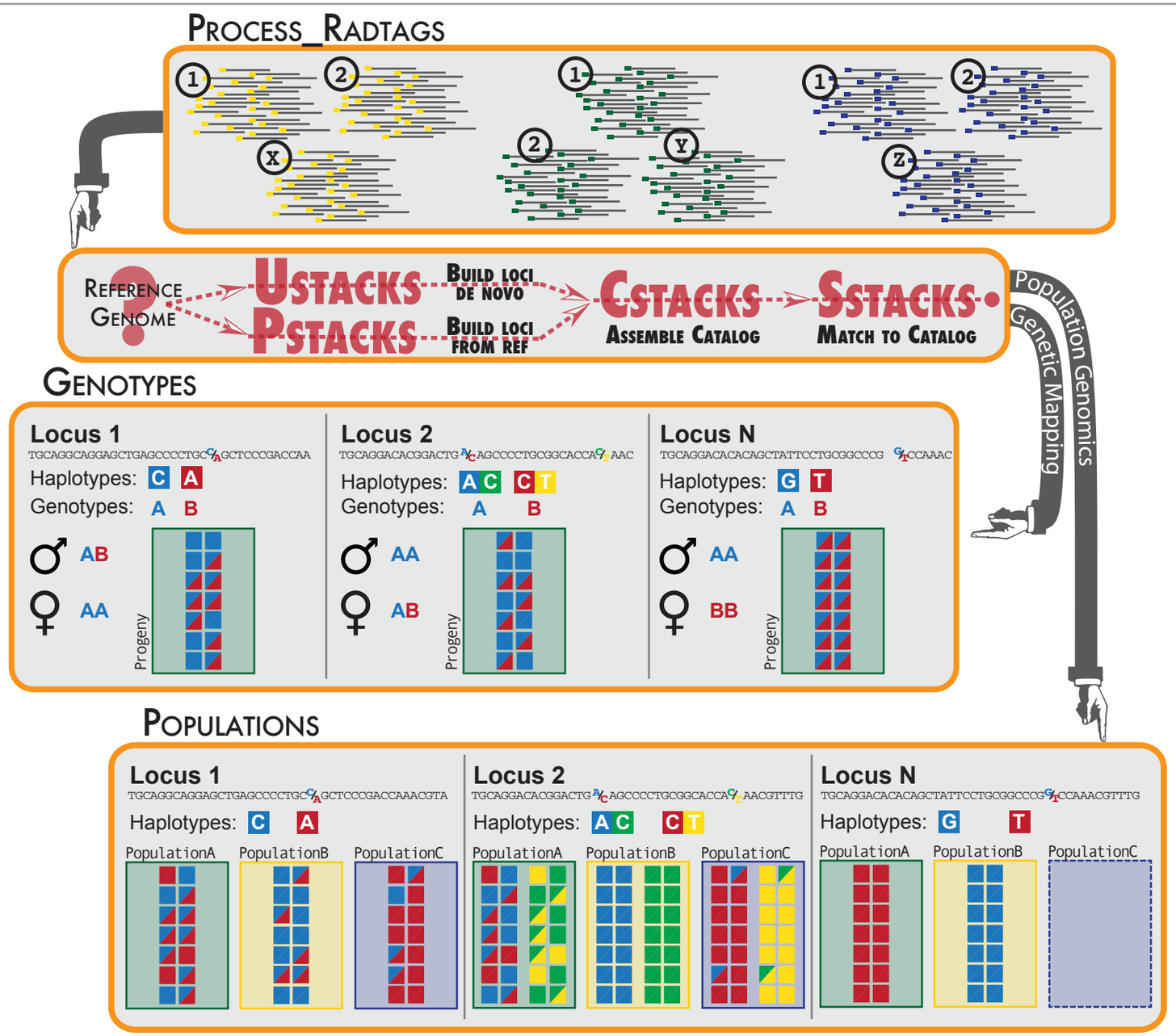
Stacks workflow



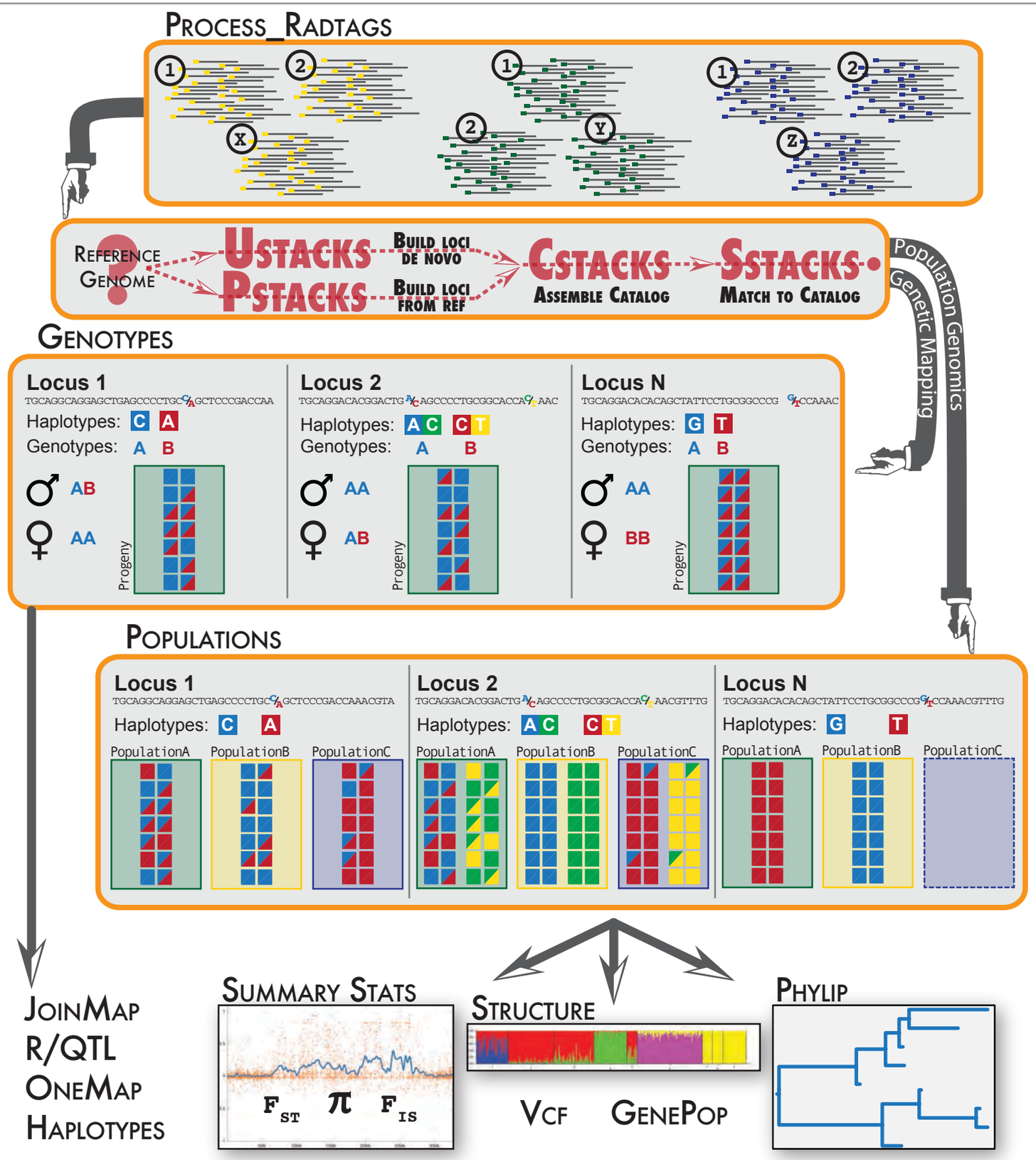
Stacks workflow



Stacks workflow



Stacks workflow



Stacks Analysis Pipeline: RAD-Tag Catalog Viewer

http://genome.uoregon.edu/stacks/catalog.php?id=1&db=gartut_radtags&p=1&pp=10&filter_type[]=cata&filter_cata=103&filter_alle_l=1&

1 (1 tags) tags per page 10

Id	SNP	Consensus	Matching Parents	Progeny	Marker	Ratio	Genotypes
~103 annotate	Yes [2nuc]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC	2	92 / 91	ab/ac	aa: 25 (27.5%) ab: 24 (26.4%) ac: 18 (19.8%) bc: 24 (26.4%)	91

SNPs

Column: 52; G/A
Column: 70; T/G

Alleles

a : GT
b : GG
c : AG

Matching Samples

View: ☑ Haplotypes ☐ Allele Depths ☐ Genotypes

Male GT / GG	Female AG / GT	Progeny 1 GT	Progeny 2 AG / GG	Progeny 3 GG / AG	Progeny 4 GG / GT	Progeny 5 GG / AG	Progeny 6 AG	Progeny 7 GT / GG	Progeny 8 AG / GT
Progeny 9 GT	Progeny 10 GT	Progeny 11 GG / GT	Progeny 12 GT / AG	Progeny 13 GG / AG	Progeny 14 GT / AG	Progeny 15 GT / GG	Progeny 16 GG / GT	Progeny 17 GG / AG	Progeny 18 GT
Progeny 19 GT / AG	Progeny 20 AG / GG	Progeny 21 GT / AG	Progeny 22 AG / GT	Progeny 23 GG / AG	Progeny 24 GG / AG	Progeny 25 GT	Progeny 26 GG / GT	Progeny 27 GG / AG	Progeny 28 GG / GT
Progeny 29 GT / GG	Progeny 31 GT	Progeny 32 GT	Progeny 33 GT	Progeny 34 GT	Progeny 35 GT / GG	Progeny 36 GT	Progeny 37 GT / AG	Progeny 38 GT	Progeny 39 AG / GT
Progeny 40 GT	Progeny 41 GT	Progeny 42 GT	Progeny 43 GT / GG	Progeny 44 GG / GT	Progeny 45 GT	Progeny 46 GG / GT	Progeny 47 GG / AG	Progeny 48 GT	Progeny 49 GT / GG
Progeny 50 GT	Progeny 51 GT	Progeny 52 GT / AG	Progeny 53 GG / GT	Progeny 54 GT / GG	Progeny 55 AG / GG	Progeny 56 GT	Progeny 57 AG / GT	Progeny 58 GT / AG	Progeny 59 GG / GT
Progeny 60 GT / GG	Progeny 61 GT / GG	Progeny 62 GT / AG	Progeny 63 GG / AG	Progeny 64 GG / GT	Progeny 65 GT	Progeny 66 GT	Progeny 67 GG / GT	Progeny 68 GT	Progeny 70 GG / AG
Progeny 71 GG / AG	Progeny 72 AG / GG	Progeny 73 GT	Progeny 74 GG / AG	Progeny 75 GT / GG	Progeny 76 GT	Progeny 77 GG / AG	Progeny 78 GG / AG	Progeny 79 GT / GG	Progeny 80 GT
Progeny 81 GT / AG	Progeny 82 GT / AG	Progeny 83 GG / AG	Progeny 84 GT	Progeny 85 GT / GG	Progeny 86 GT / GG	Progeny 87 GT	Progeny 88 GG / AG	Progeny 89 GT	Progeny 90 GG / AG
Progeny 91 AG / GG	Progeny 92 GT / AG	Progeny 93 AG / GG	Progeny 94 GG / AG						

1 (1 tags) tags per page 10

Stacks Analysis Pipeline: RAD-Tag Catalog Viewer

http://genome.uoregon.edu/stacks/catalog.php?id=1&db=gartut_radtags&p=1&pp=10&filter_type[]=cata&filter_cata=103&filter_alle_l=1&

1 (1 tags) tags per page 10

Id	SNP	Consensus	Matching Parents	Progeny	Marker	Ratio	Genotypes
~103 annotate	Yes [2nuc]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC	2	92 / 91	ab/ac	aa: 25 (27.5%) ab: 24 (26.4%) ac: 18 (19.8%) bc: 24 (26.4%)	91

SNPs

Column: 52; G/A
Column: 70; T/G

Alleles

a : GT
b : GG
c : AG

Matching Samples

View: ☑ Haplotypes ☑ Allele Depths ☐ Genotypes

Male GT / GG 34 / 13	Female AG / GT 12 / 14	Progeny 1 GT 7	Progeny 2 AG / GG 8 / 16	Progeny 3 GG / AG 26 / 14	Progeny 4 GG / GT 15 / 11	Progeny 5 GG / AG 14 / 8	Progeny 6 AG 29	Progeny 7 GT / GG 22 / 11	Progeny 8 AG / GT 12 / 5
Progeny 9 GT 25	Progeny 10 GT 23	Progeny 11 GG / GT 32 / 14	Progeny 12 GT / AG 22 / 7	Progeny 13 GG / AG 7 / 8	Progeny 14 GT / AG 7 / 8	Progeny 15 GT / GG 2 / 3	Progeny 16 GG / GT 19 / 14	Progeny 17 GG / AG 9 / 4	Progeny 18 GT 15
Progeny 19 GT / AG 6 / 3	Progeny 20 AG / GG 6 / 9	Progeny 21 GT / AG 18 / 9	Progeny 22 AG / GT 4 / 5	Progeny 23 GG / AG 7 / 6	Progeny 24 GG / AG 8 / 10	Progeny 25 GT 7	Progeny 26 GG / GT 10 / 16	Progeny 27 GG / AG 3 / 3	Progeny 28 GG / GT 4 / 5
Progeny 29 GT / GG 8 / 5	Progeny 31 GT 11	Progeny 32 GT 10	Progeny 33 GT 17	Progeny 34 GT 20	Progeny 35 GT / GG 7 / 3	Progeny 36 GT 8	Progeny 37 GT / AG 12 / 4	Progeny 38 GT 9	Progeny 39 AG / GT 12 / 7
Progeny 40 GT 9	Progeny 41 GT 5	Progeny 42 GT 9	Progeny 43 GT / GG 9 / 12	Progeny 44 GG / GT 3 / 6	Progeny 45 GT 6	Progeny 46 GG / GT 4 / 11	Progeny 47 GG / AG 3 / 7	Progeny 48 GT 18	Progeny 49 GT / GG 5 / 6
Progeny 50 GT 18	Progeny 51 GT 9	Progeny 52 GT / AG 8 / 5	Progeny 53 GG / GT 10 / 8	Progeny 54 GT / GG 5 / 6	Progeny 55 AG / GG 8 / 10	Progeny 56 GT 22	Progeny 57 AG / GT 17 / 16	Progeny 58 GT / AG 23 / 24	Progeny 59 GG / GT 25 / 13
Progeny 60 GT / GG 12 / 18	Progeny 61 GT / GG 22 / 29	Progeny 62 GT / AG 7 / 23	Progeny 63 GG / AG 15 / 11	Progeny 64 GG / GT 13 / 20	Progeny 65 GT 44	Progeny 66 GT 27	Progeny 67 GG / GT 23 / 17	Progeny 68 GT 30	Progeny 70 GG / AG 14 / 13
Progeny 71 GG / AG 15 / 7	Progeny 72 AG / GG 9 / 6	Progeny 73 GT 42	Progeny 74 GG / AG 31 / 29	Progeny 75 GT / GG 15 / 22	Progeny 76 GT 41	Progeny 77 GG / AG 14 / 17	Progeny 78 GG / AG 25 / 17	Progeny 79 GT / GG 29 / 14	Progeny 80 GT 34
Progeny 81 GT / AG 17 / 29	Progeny 82 GT / AG 29 / 24	Progeny 83 GG / AG 16 / 25	Progeny 84 GT 41	Progeny 85 GT / GG 14 / 24	Progeny 86 GT / GG 6 / 4	Progeny 87 GT 15	Progeny 88 GG / AG 5 / 11	Progeny 89 GT 18	Progeny 90 GG / AG 5 / 17
Progeny 91 AG / GG 14 / 13	Progeny 92 GT / AG 12 / 6	Progeny 93 AG / GG 7 / 7	Progeny 94 GG / AG 3 / 2						

1 (1 tags) tags per page 10

Stacks Analysis Pipeline: RAD-Tag Catalog Viewer

http://genome.uoregon.edu/stacks/catalog.php?id=1&db=gartut_radtags&p=1&pp=10&filter_type[]=cata&filter_cata=103&filter_alle_l=1&

1 (1 tags)

tags per page 10

Id	SNP	Consensus	Matching Parents	Progeny	Marker	Ratio	Genotypes
~103 annotate	Yes [2nuc]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC	2	92 / 91	ab/ac	aa: 25 (27.5%) ab: 24 (26.4%) ac: 18 (19.8%) bc: 24 (26.4%)	91

SNPs

Column: 52; G/A
Column: 70; T/G

Alleles

a : GT
b : GG
c : AG

Matching Samples

View: ☑ Haplotypes ☑ Allele Depths ☑ Genotypes

Male GT / GG 34 / 13	Female AG / GT 12 / 14	Progeny 1 GT 7 aa	Progeny 2 AG / GG 8 / 16 bc	Progeny 3 GG / AG 26 / 14 bc	Progeny 4 GG / GT 15 / 11 ab	Progeny 5 GG / AG 14 / 8 bc	Progeny 6 AG 29 AC	Progeny 7 GT / GG 22 / 11 ab	Progeny 8 AG / GT 12 / 5 ac
Progeny 9 GT 25 aa	Progeny 10 GT 23 aa	Progeny 11 GG / GT 32 / 14 ab	Progeny 12 GT / AG 22 / 7 ac	Progeny 13 GG / AG 7 / 8 bc	Progeny 14 GT / AG 7 / 8 ac	Progeny 15 GT / GG 2 / 3 ab	Progeny 16 GG / GT 19 / 14 ab	Progeny 17 GG / AG 9 / 4 bc	Progeny 18 GT 15 aa
Progeny 19 GT / AG 6 / 3 ac	Progeny 20 AG / GG 6 / 9 bc	Progeny 21 GT / AG 18 / 9 ac	Progeny 22 AG / GT 4 / 5 ac	Progeny 23 GG / AG 7 / 6 bc	Progeny 24 GG / AG 8 / 10 bc	Progeny 25 GT 7 AC	Progeny 26 GG / GT 10 / 16 ab	Progeny 27 GG / AG 3 / 3 bc	Progeny 28 GG / GT 4 / 5 ab
Progeny 29 GT / GG 8 / 5 ab	Progeny 31 GT 11 aa	Progeny 32 GT 10 aa	Progeny 33 GT 17 aa	Progeny 34 GT 20 aa	Progeny 35 GT / GG 7 / 3 ab	Progeny 36 GT 8 aa	Progeny 37 GT / AG 12 / 4 ac	Progeny 38 GT 9 aa	Progeny 39 AG / GT 12 / 7 ac
Progeny 40 GT 9 aa	Progeny 41 GT 5 aa	Progeny 42 GT 9 aa	Progeny 43 GT / GG 9 / 12 ab	Progeny 44 GG / GT 3 / 6 ab	Progeny 45 GT 6 AC	Progeny 46 GG / GT 4 / 11 ab	Progeny 47 GG / AG 3 / 7 bc	Progeny 48 GT 18 aa	Progeny 49 GT / GG 5 / 6 ab
Progeny 50 GT 18 aa	Progeny 51 GT 9 aa	Progeny 52 GT / AG 8 / 5 ac	Progeny 53 GG / GT 10 / 8 ab	Progeny 54 GT / GG 5 / 6 ab	Progeny 55 AG / GG 8 / 10 bc	Progeny 56 GT 22 aa	Progeny 57 AG / GT 17 / 16 ac	Progeny 58 GT / AG 23 / 24 ac	Progeny 59 GG / GT 25 / 13 ab
Progeny 60 GT / GG 12 / 18 ab	Progeny 61 GT / GG 22 / 29 ab	Progeny 62 GT / AG 7 / 23 ac	Progeny 63 GG / AG 15 / 11 bc	Progeny 64 GG / GT 13 / 20 ab	Progeny 65 GT 44 aa	Progeny 66 GT 27 aa	Progeny 67 GG / GT 23 / 17 ab	Progeny 68 GT 30 aa	Progeny 70 GG / AG 14 / 13 bc
Progeny 71 GG / AG 15 / 7 bc	Progeny 72 AG / GG 9 / 6 bc	Progeny 73 GT 42 aa	Progeny 74 GG / AG 31 / 29 bc	Progeny 75 GT / GG 15 / 22 ab	Progeny 76 GT 41 aa	Progeny 77 GG / AG 14 / 17 bc	Progeny 78 GG / AG 25 / 17 bc	Progeny 79 GT / GG 29 / 14 ab	Progeny 80 GT 34 aa
Progeny 81 GT / AG 17 / 29	Progeny 82 GT / AG 20 / 24	Progeny 83 GG / AG 16 / 25	Progeny 84 GT 41	Progeny 85 GT / GG 14 / 24	Progeny 86 GT / GG 6 / 4	Progeny 87 GT 15	Progeny 88 GG / AG 5 / 11	Progeny 89 GT 18	Progeny 90 GG / AG 5 / 17

Stacks Analysis Pipeline: RAD-Tag Viewer

http://genome.uoregon.edu/stacks/tag.php?db=gartut_radtags&batch_id=1&sample_id=2&tag_id=73

Stacks

version 0.998

Batch #1 [2011-08-10; 80bp Lepisosteus oculatus F1 Genetic Map RAD-Tag Samples]

RAD-Tag Sample #2 [female]

Sequence #73

Catalog ID	Depth	SNPs		Alleles		Deleveraged?	Lumberjackstack?	Blacklisted?
#103	26x	Column: 52	G/A	AG	46.15%	False	False	False
		Column: 70	T/G	GT	53.85%			

	Relationship	Seq ID	Sequence
	consensus		TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
	model		OOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO
1	primary	CAGTC_2_0018_768_1365_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACTTCACAGTCCC
2	primary	CAGTC_2_0029_1628_1751_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACTTCACAGTCCC
3	primary	CAGTC_2_0053_1692_1388_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACTTCACAGTCCC
4	primary	CAGTC_2_0058_1588_1038_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACTTCACAGTCCC
5	primary	CAGTC_2_0059_1524_1186_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACTTCACAGTCCC
6	primary	CAGTC_2_0094_1356_1854_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACTTCACAGTCCC
7	primary	CAGTC_2_0096_1791_1246_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACTTCACAGTCCC
8	primary	CAGTC_2_0021_877_296_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
9	primary	CAGTC_2_0024_307_735_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
10	primary	CAGTC_2_0025_108_810_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
11	primary	CAGTC_2_0039_1252_1764_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
12	primary	CAGTC_2_0061_596_159_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
13	primary	CAGTC_2_0068_1310_997_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
14	primary	CAGTC_2_0070_644_2040_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
15	primary	CAGTC_2_0074_328_659_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
16	primary	CAGTC_2_0075_1668_1862_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
17	primary	CAGTC_2_0079_1481_505_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
18	primary	CAGTC_2_0084_805_1974_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
19	primary	CAGTC_2_0100_481_1043_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATTCCC
20	secondary	CAGTC_2_0014_728_1008_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACTTCACAGACCC
21	secondary	CAGTC_2_0016_86_1022_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCAACACTTCACATACCC
22	secondary	CAGTC_2_0042_426_1001_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACCAAGCAACACTTCACAGTCCC
23	secondary	CAGTC_2_0052_867_1387_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCCGTGGACCGAGAGCACAAAGCAACACTTCACAGTCCC
24	secondary	CAGTC_2_0012_221_1043_1[35245]	TGCAGGAGCCCTCCCACTAGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACTTCACAGTCCC
25	secondary	CAGTC_2_0095_120_1067_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCGCAAAGCCACACTTCACATCCCC
26	secondary	CAGTC_2_0077_1003_356_1[35245]	TGCAGGAGCCCTCCCACTCGCTGATGGCCACTCCATTCAGTGGACCGAGAGCACAAAGCAACACCTCACAGTCCC

last updated: Sun Jan 8 08:59:35 PST 2012