

Signatures of selection

Day 2 - Lecture 2

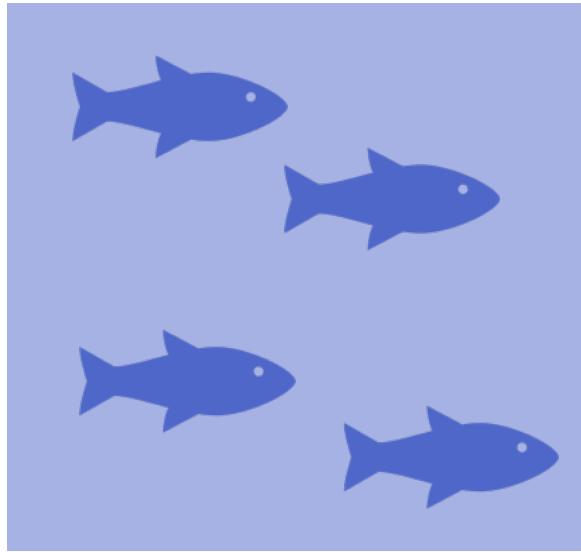
Anna Tigano, Ph.D.

Adaptation is the evolutionary response of a population to a selective force

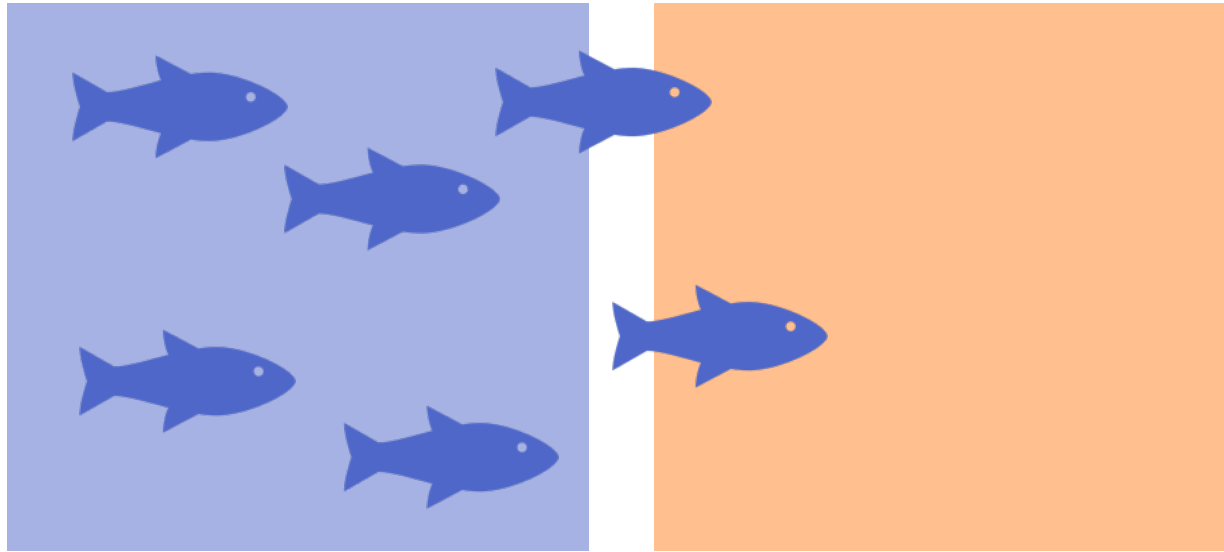
If we identify the loci underpinning the genetic basis of an adaptive trait we can better understand

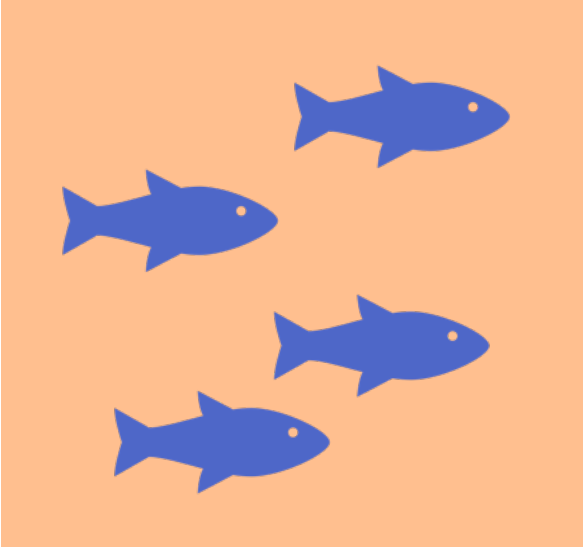
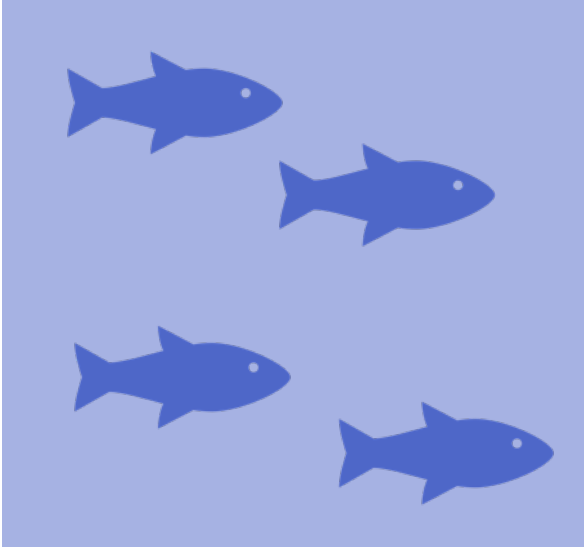
- the adaptive trait itself (e.g. its molecular and physiological basis)
- the process of adaptation (including the interplay with other evolutionary forces)

A population of cold-adapted fish

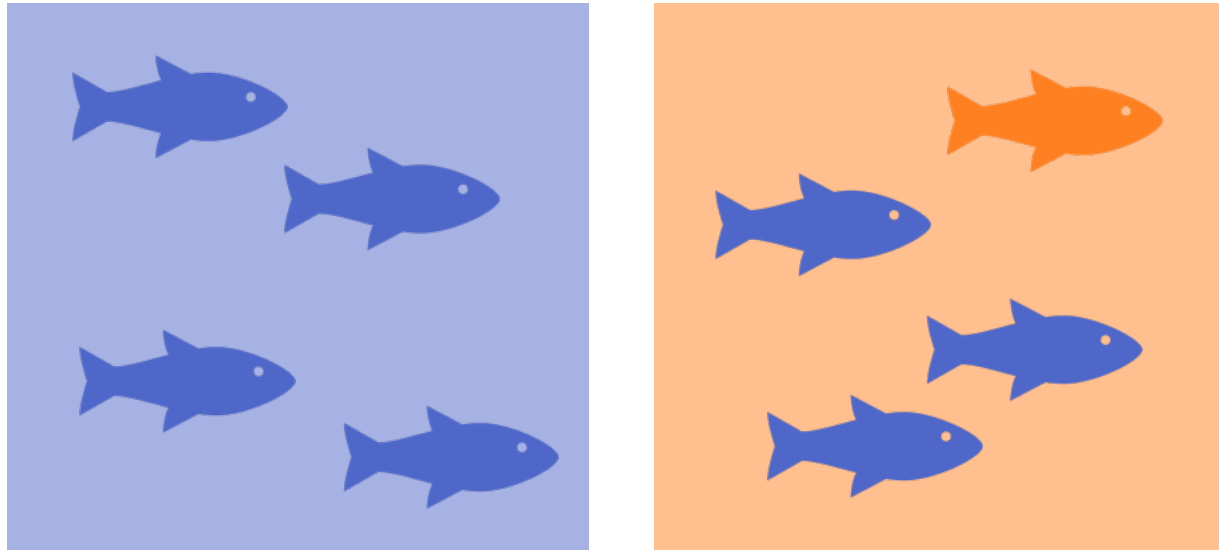


Cold-adapted fish migrate to warmer waters

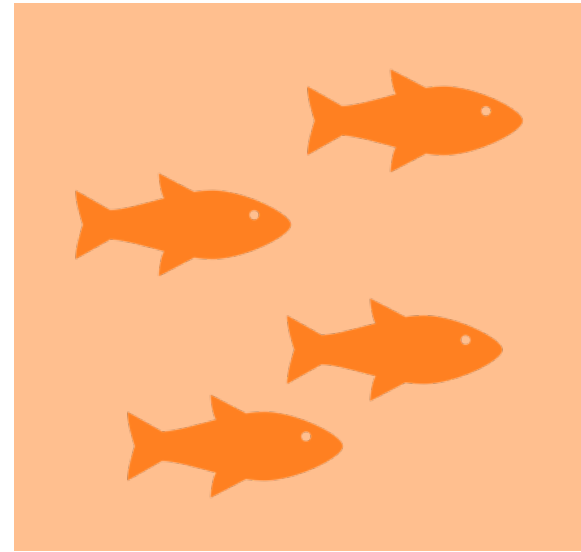
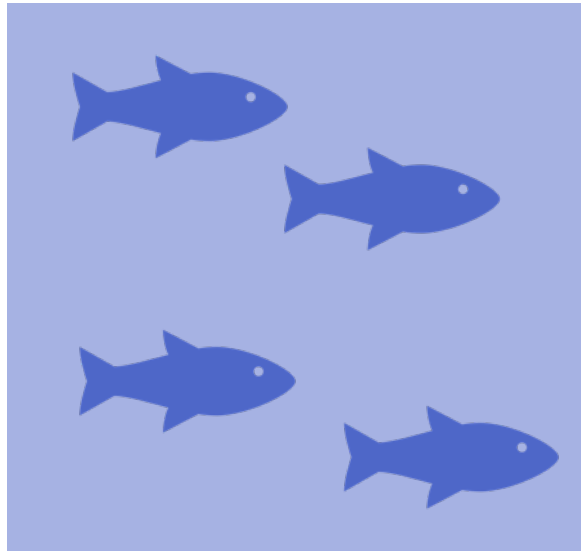




A new mutation confers a fitness benefit in warmer waters

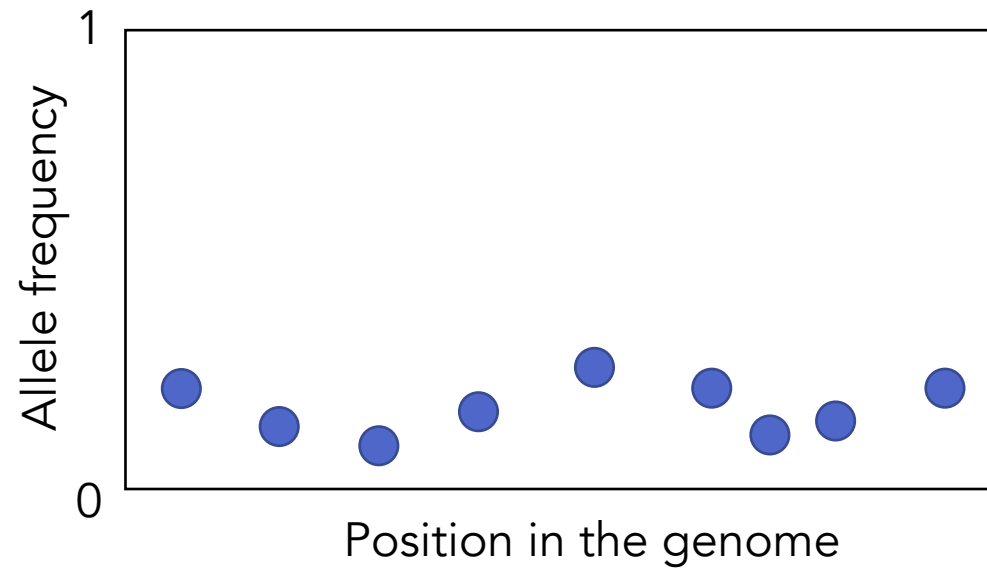


The new mutation spreads quickly in the population and reach fixation



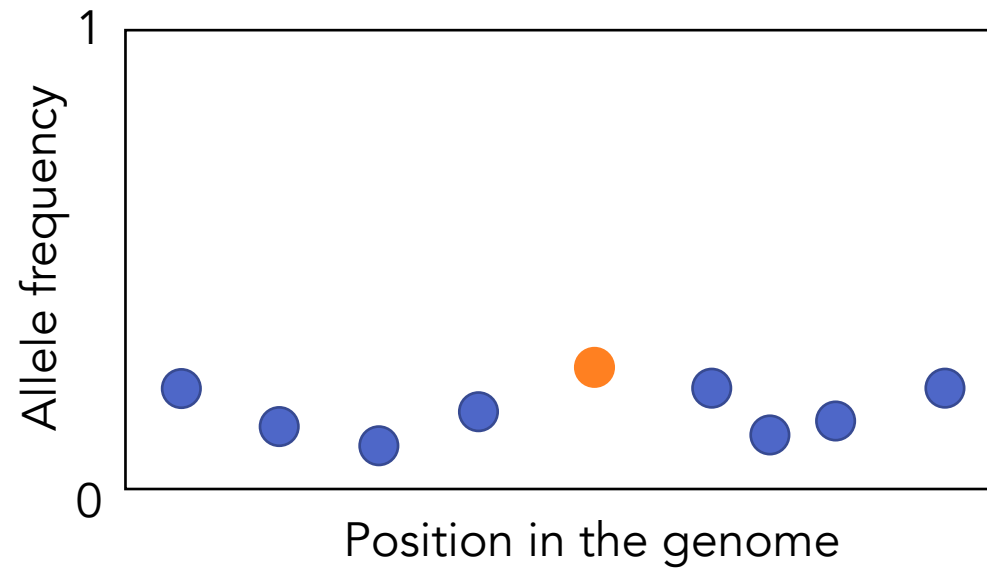
Changes in allele frequencies

Cold-adapted fish migrate to warmer waters



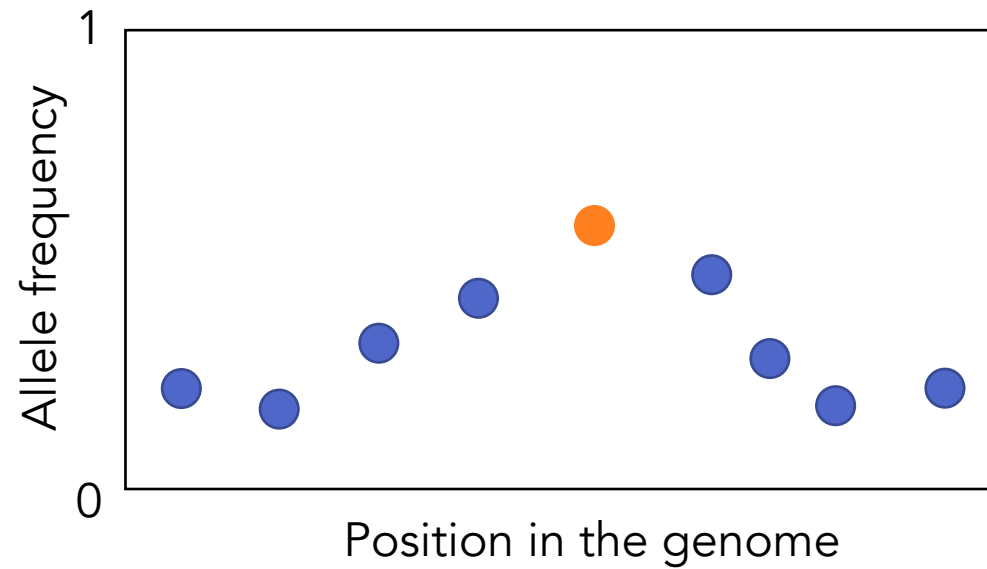
Changes in allele frequencies

A new mutation confers a fitness benefit in warmer waters



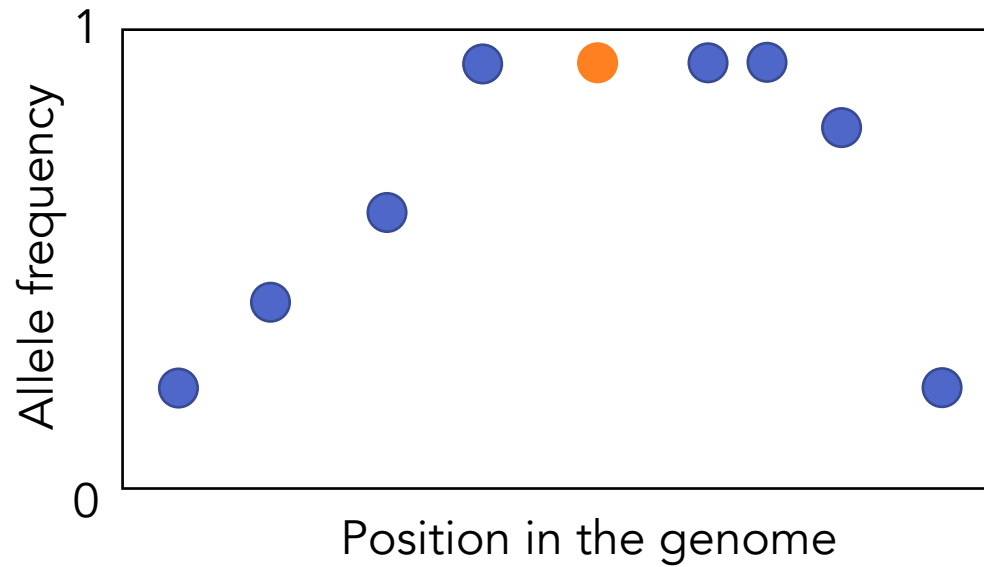
Changes in allele frequencies

The mutation increases quickly in frequency, together with the variants in its proximity



Changes in allele frequencies

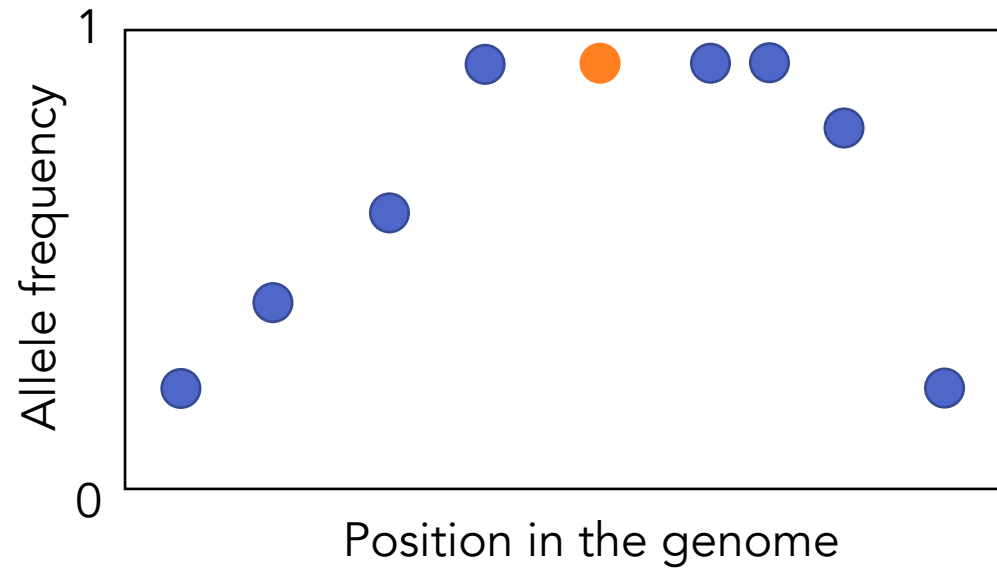
The new mutation spreads quickly in the population and reaches fixation.



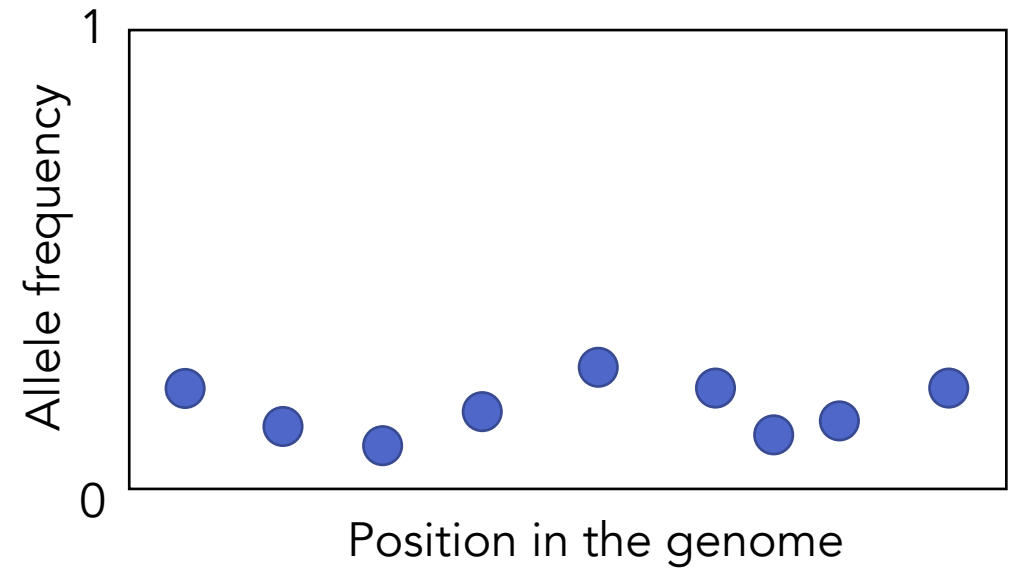
Changes in allele frequencies



Warm-adapted



Cold-adapted



How do we measure these changes?

F_{ST} is a normalized measure of allele frequency differences between populations.

$$F_{ST} = (\pi_T - \pi_S) / \pi_T$$

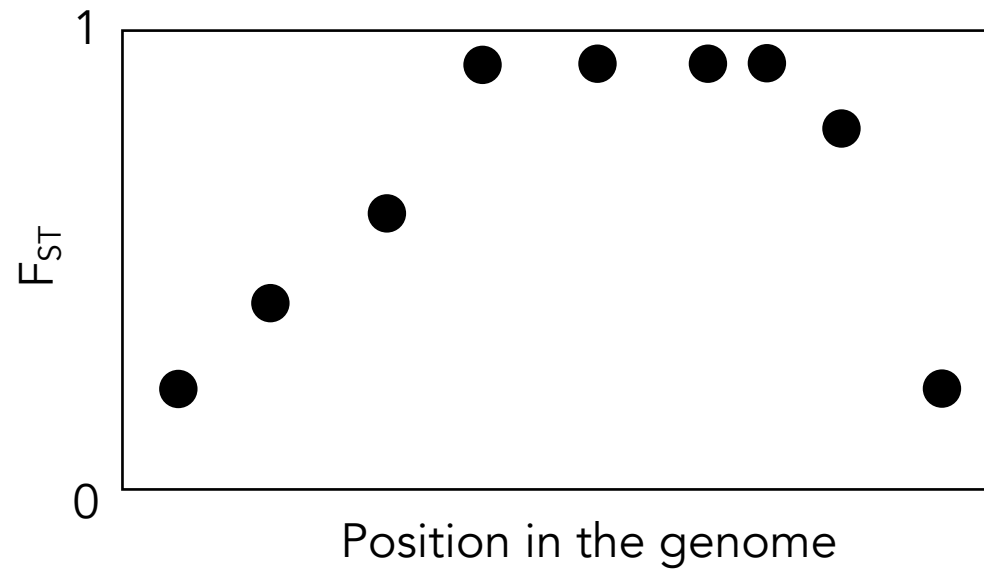
π_T = expected heterozygosity in the total sample

π_S = expected heterozygosity in each population

*Note that there are many ways to estimate F_{ST} *

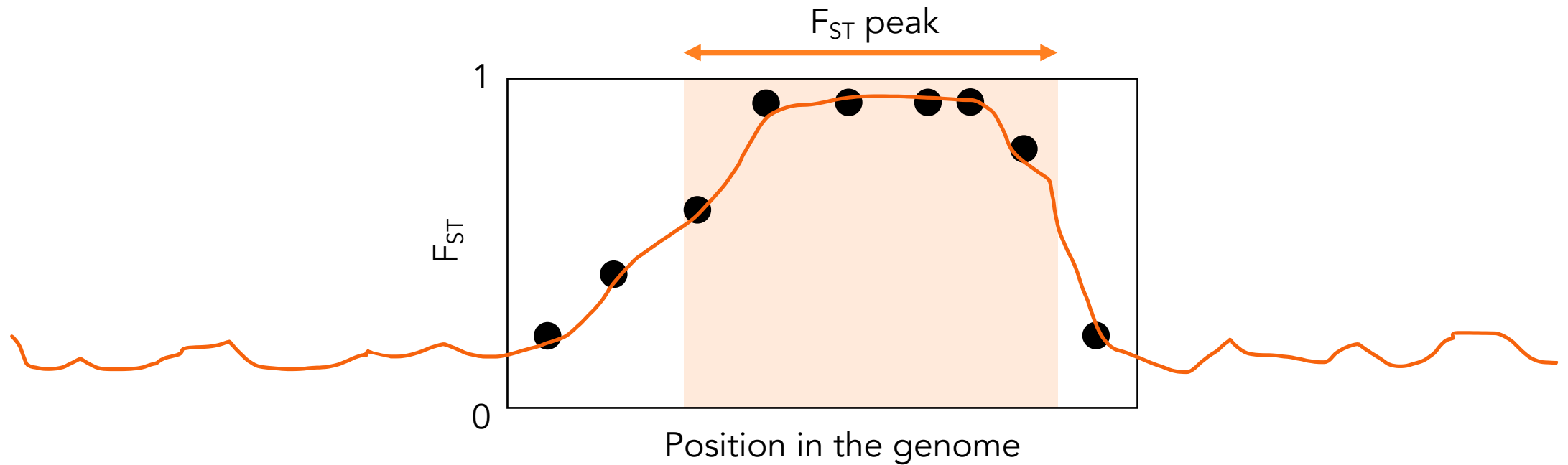
Genetic differentiation

Differentiation between cold- and warm-adapted populations



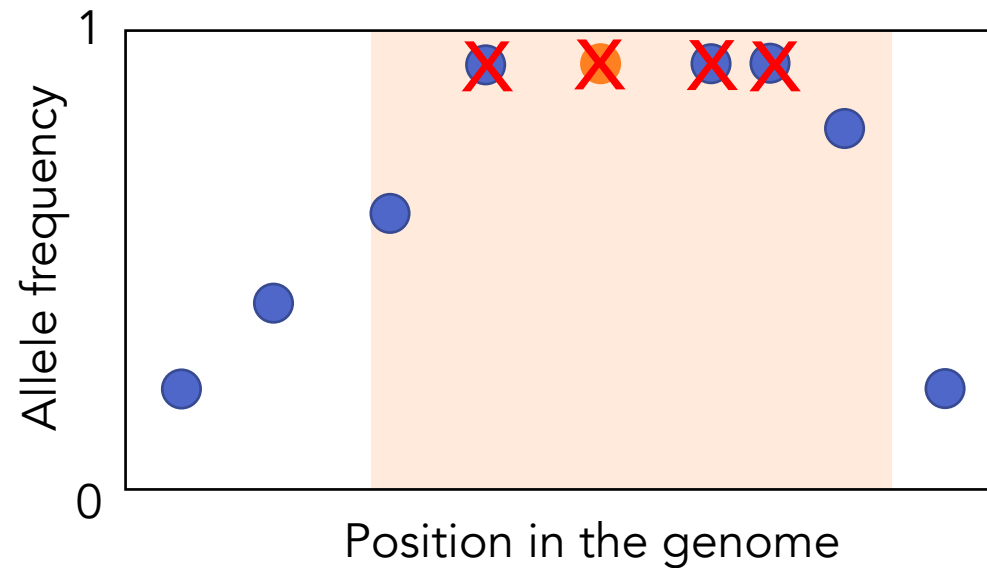
Genetic differentiation

Differentiation between cold- and warm-adapted populations



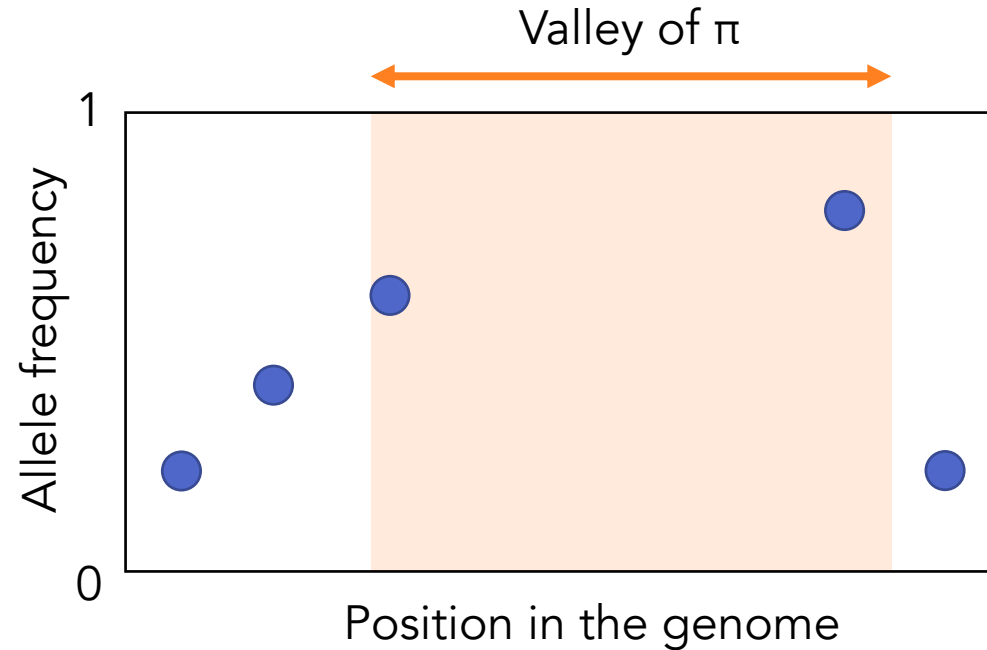
Reduction in diversity

Once the beneficial allele has been brought to fixation, variation at the target of selection and surrounding sites is eliminated from the population.

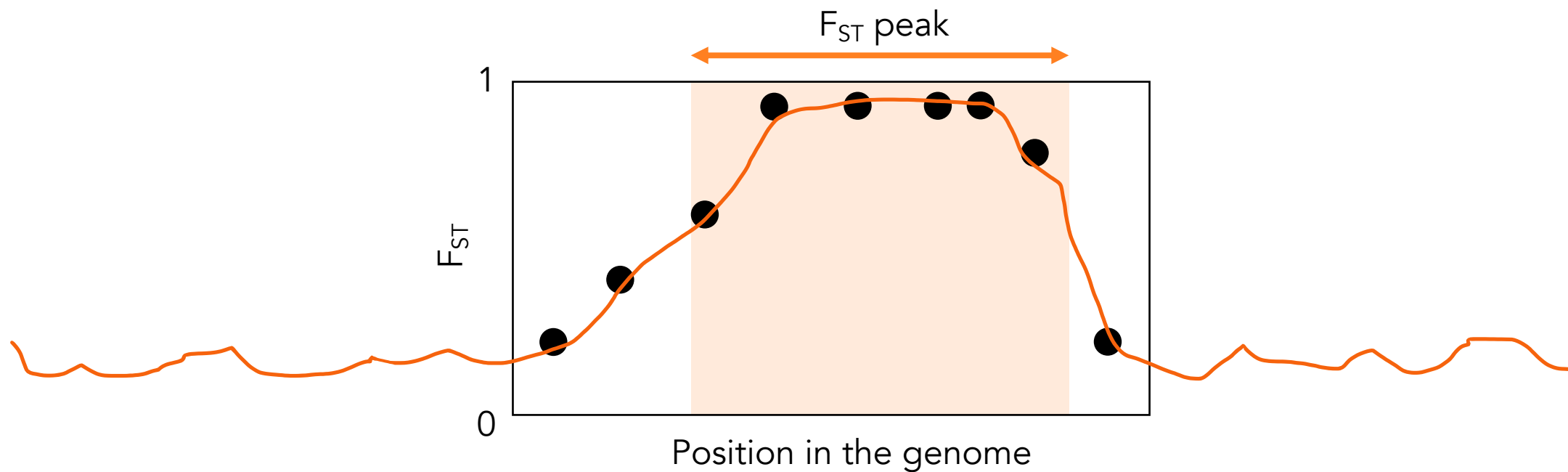


Reduction in diversity

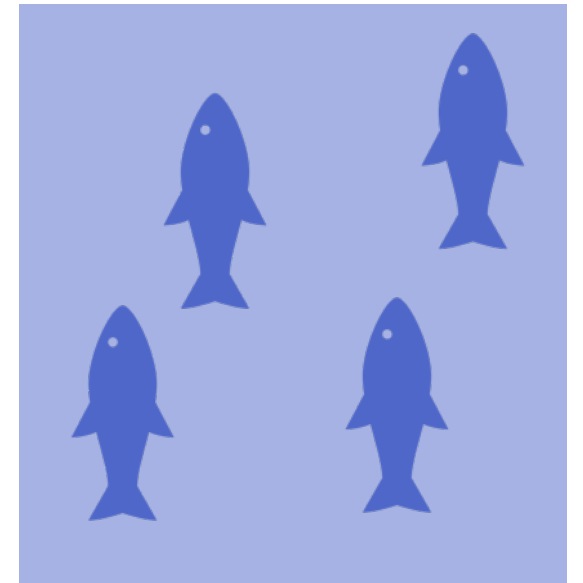
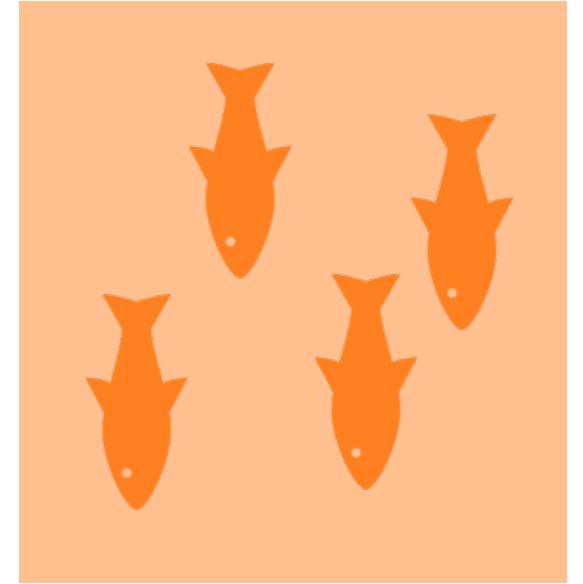
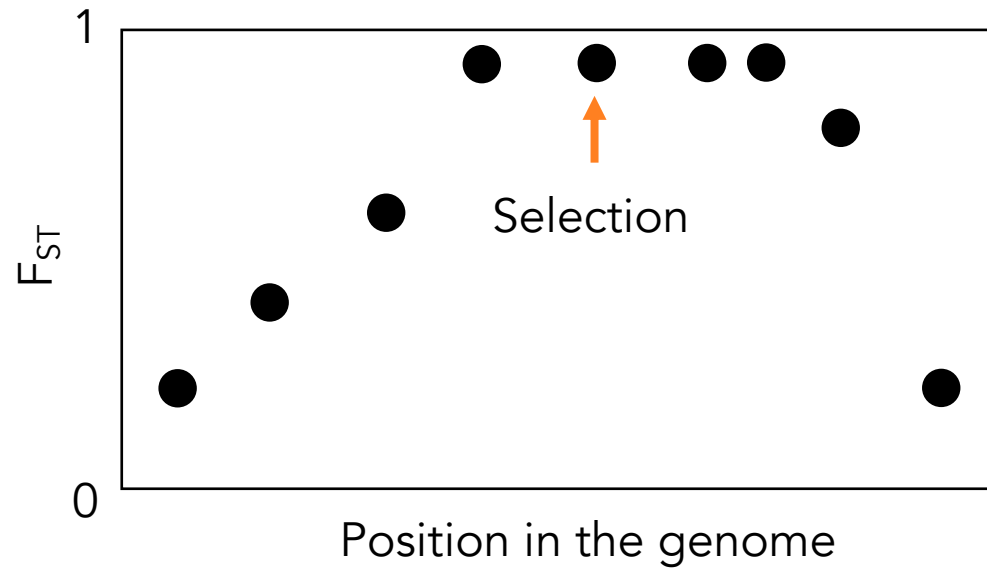
Once the beneficial allele has been brought to fixation, variation at the target of selection and surrounding sites is eliminated from the population.



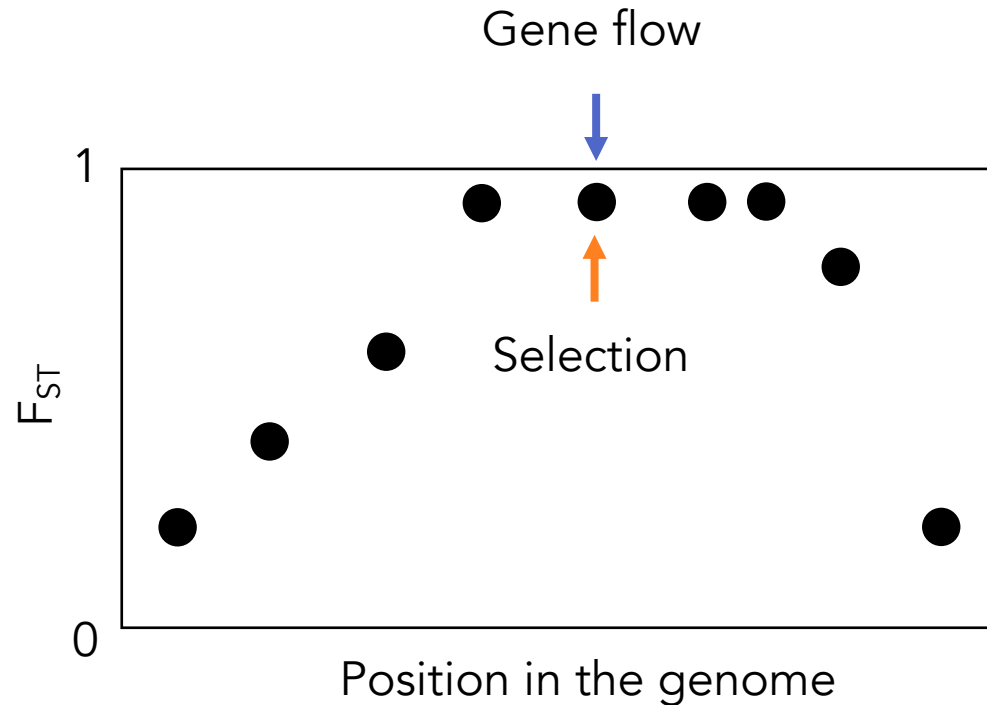
What determines the size of a F_{ST} peak?



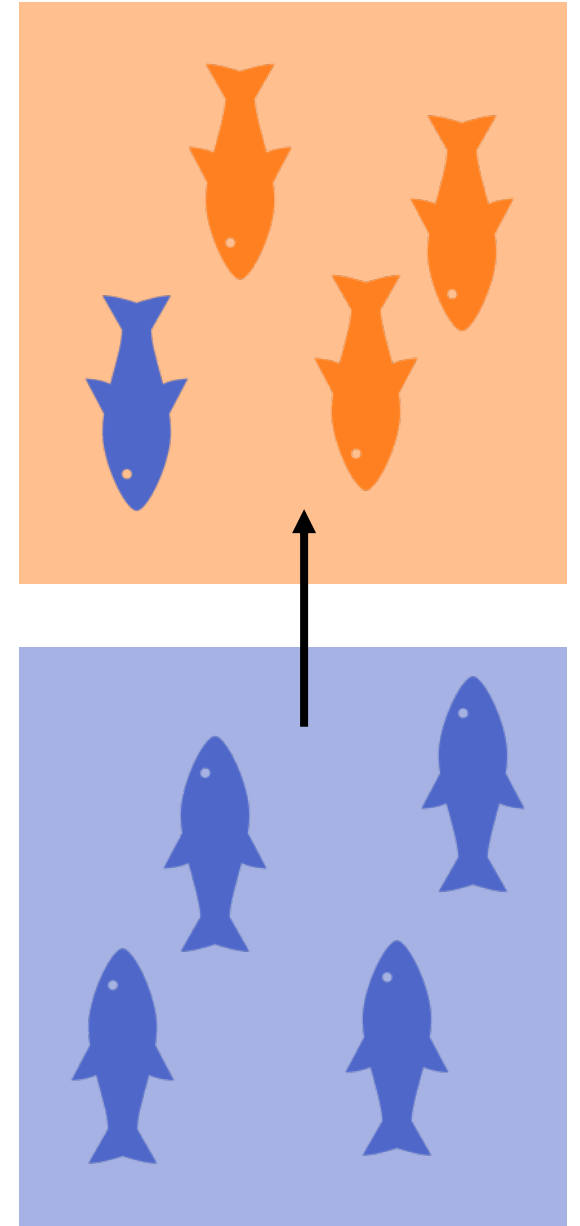
Selection vs. gene flow



Selection vs. gene flow

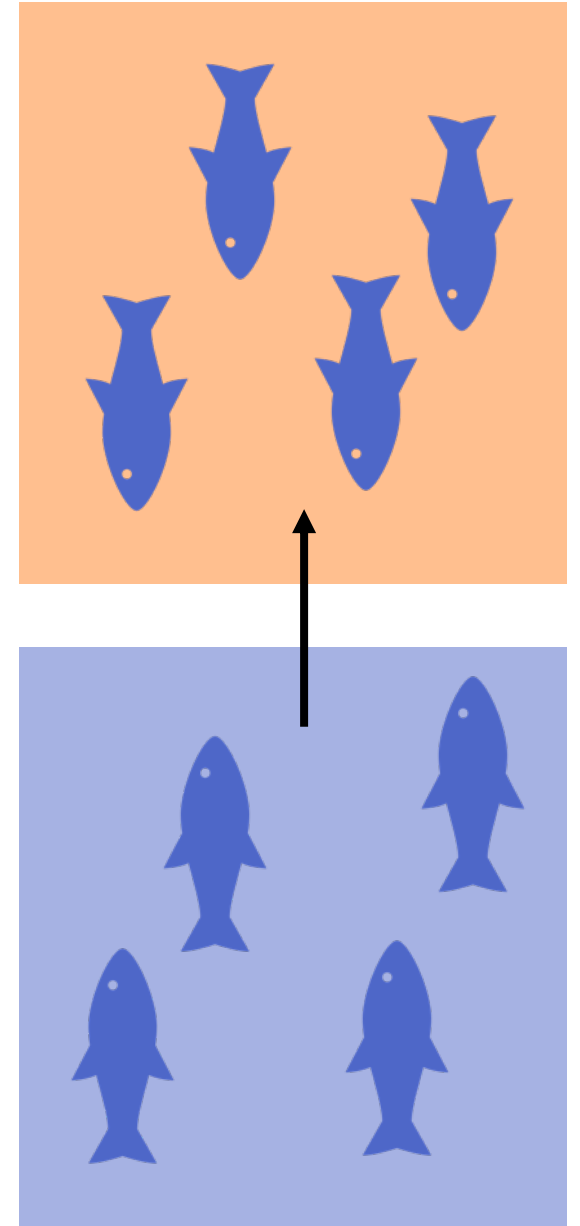


If gene flow introduces maladaptive alleles, the strength of selection s will determine whether the locally adaptive allele will be swamped.



Selection vs. gene flow

If selection is **weaker** than gene flow, the locally adapted alleles will be swamped and the local adaptation will be lost.

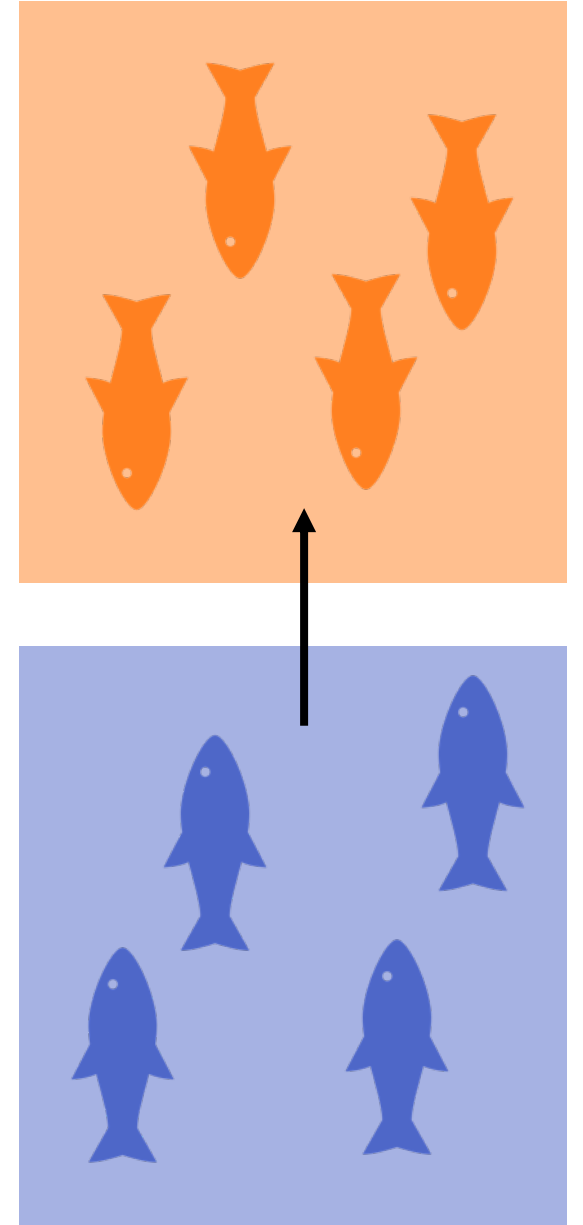


Selection vs. gene flow

If selection is **stronger** than gene flow, the locally adapted alleles will be maintained despite gene flow.

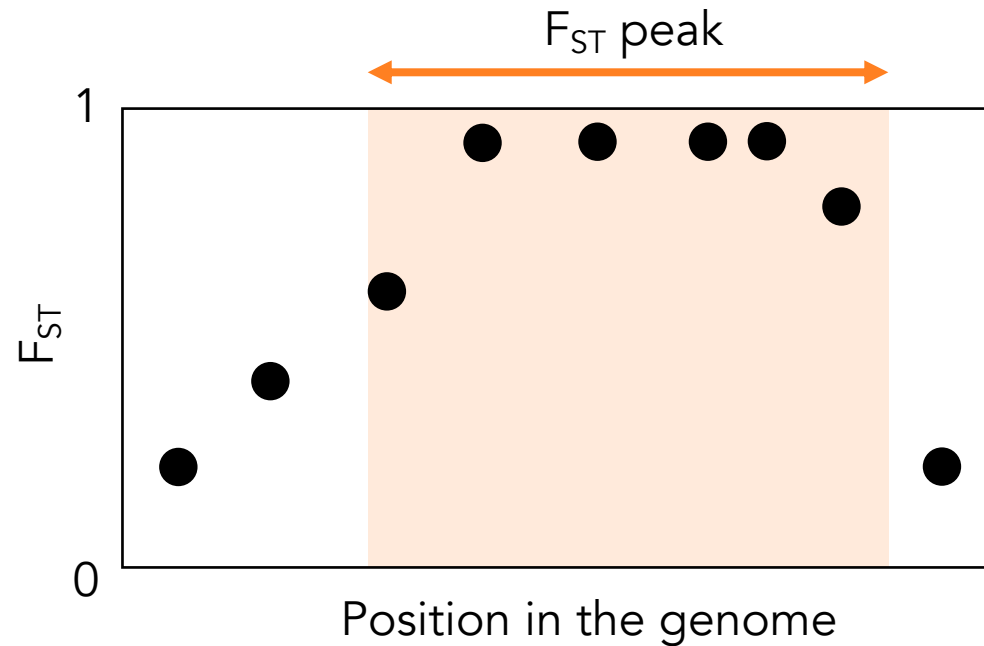
The locally adaptive allele will be maintained but overall population differentiation will be homogenized by gene flow through recombination.

Strong selection → strong linked selection → larger F_{ST} peak



Physical linkage and recombination

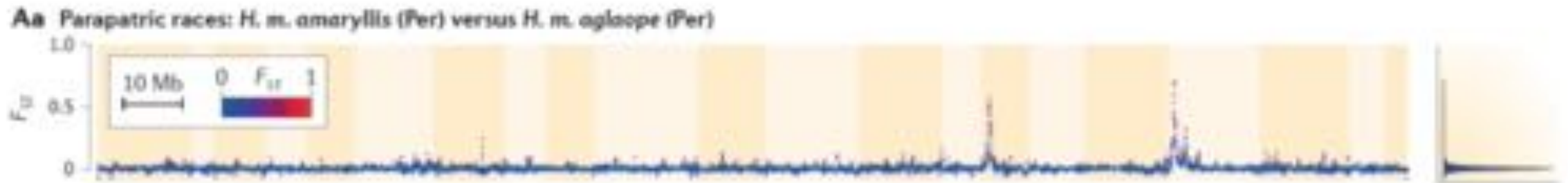
Recombination breaks linkage among sites at a rate that rapidly decays with distance.



Physical linkage and recombination

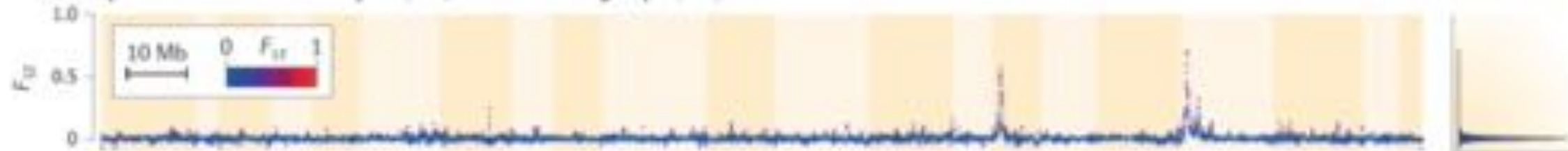
Isolating and identifying the loci underpinning local adaptations is generally easier when we can compare populations with different local adaptations that are connected by gene flow.

So that adaptive loci can stand out from a genomic background of low differentiation



Seehausen et al. 2014, Nat. Gen. Rev.

Aa Parapatric races: *H. m. amaryllis* (Per) versus *H. m. ogloope* (Per)



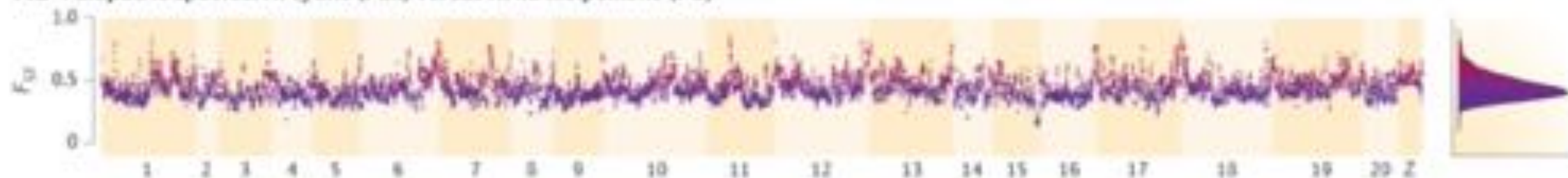
Ab Allopatric races: *H. m. rosino* (Pan) versus *H. m. melpomene* (FG)



Ac Sympatric species: *H. cydno* (Pan) versus *H. m. rosino* (Pan)

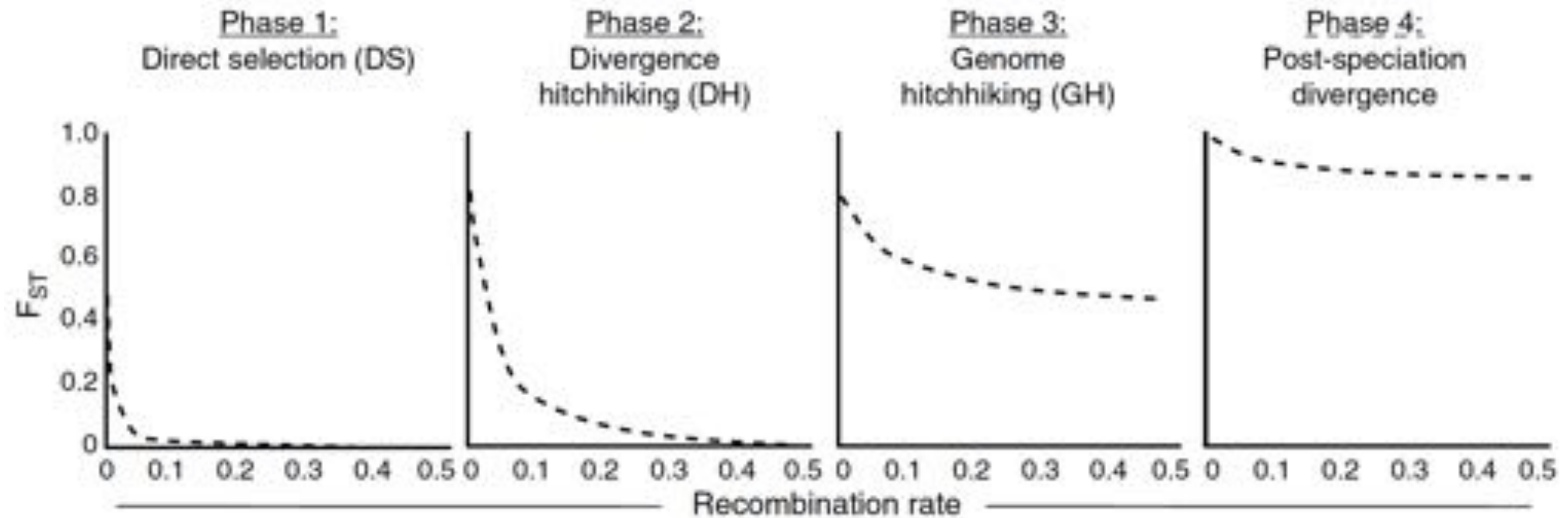


Ad Allopatric species: *H. cydno* (Pan) versus *H. m. melpomene* (FG)



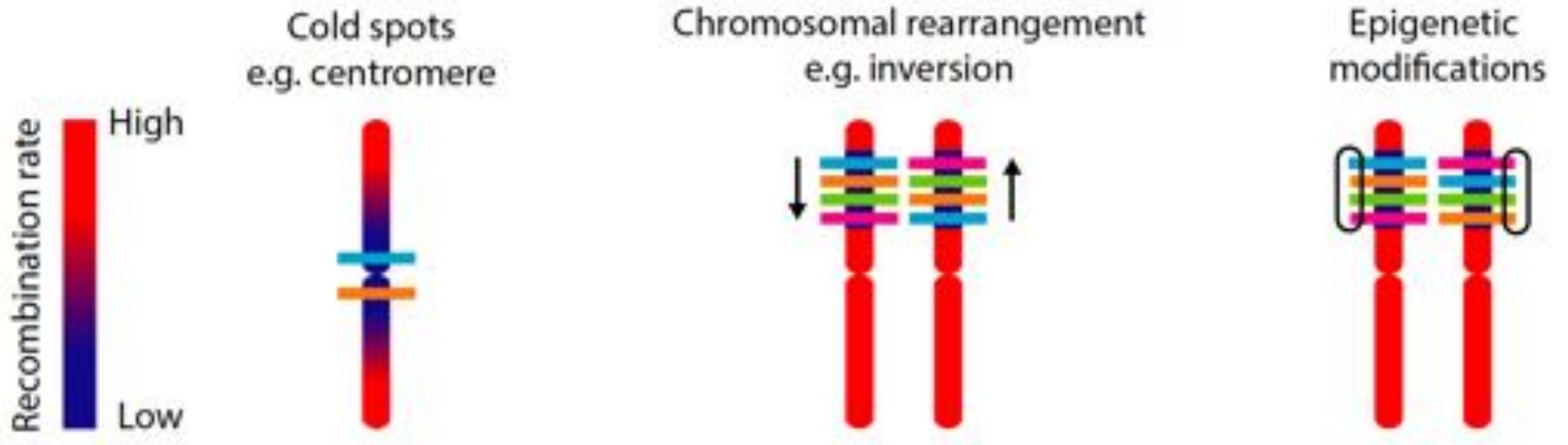
The combination of gene flow and recombination reduces the size of the peak and genome-wide differentiation.

In isolation (no gene flow), divergence at locally adapted loci builds up over long genomic distances.



Feder et al. 2012, TREE

Mechanisms that reduce recombination

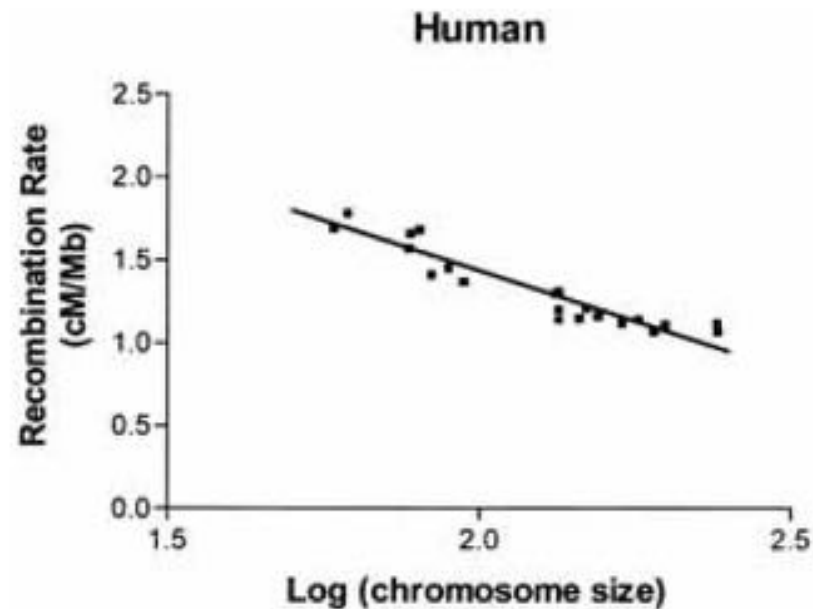




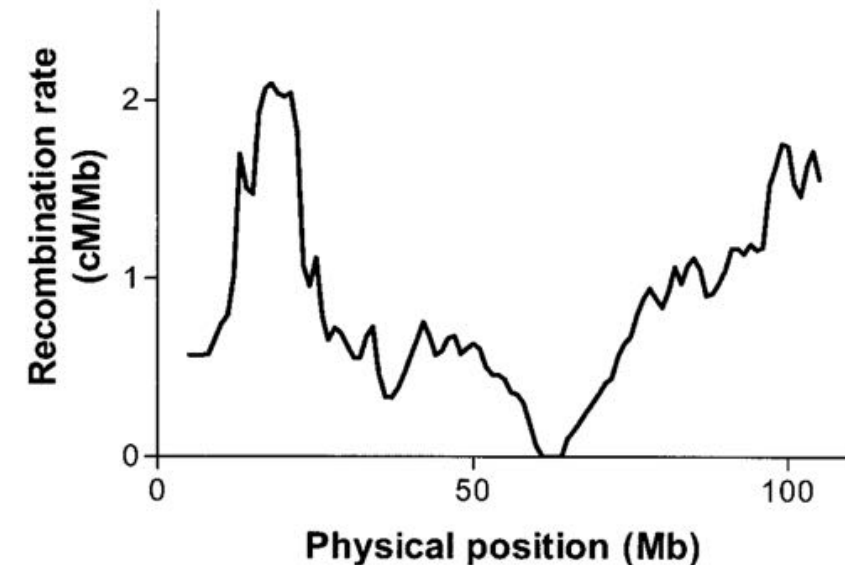
Cold spots of recombination

Recombination rates are not even across the genome.

Recombination rates can
vary among chromosomes



...and within chromosomes

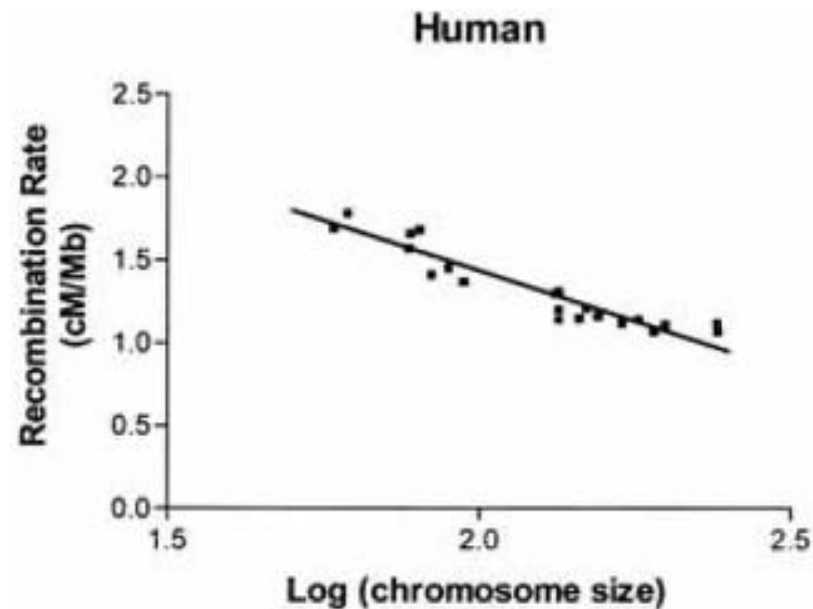




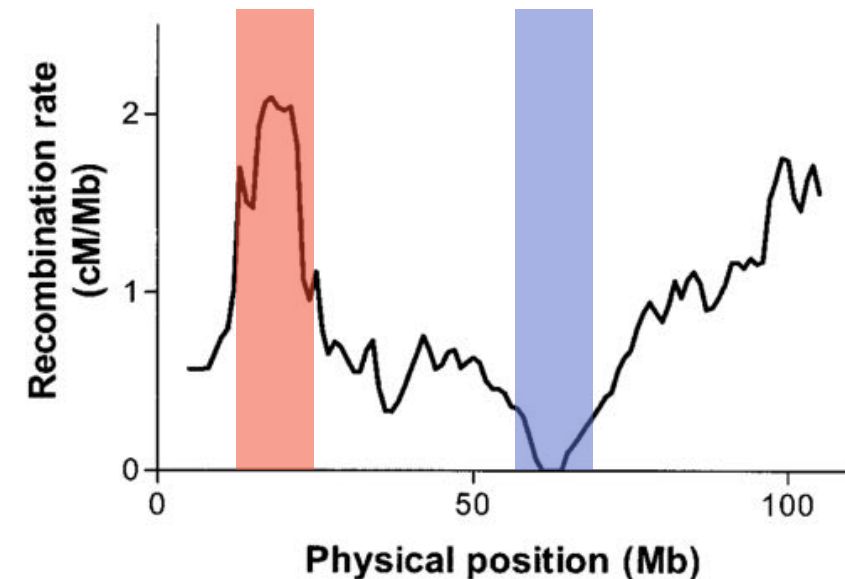
Cold spots of recombination

Recombination rates are not even across the genome.

Recombination rates can
vary among chromosomes

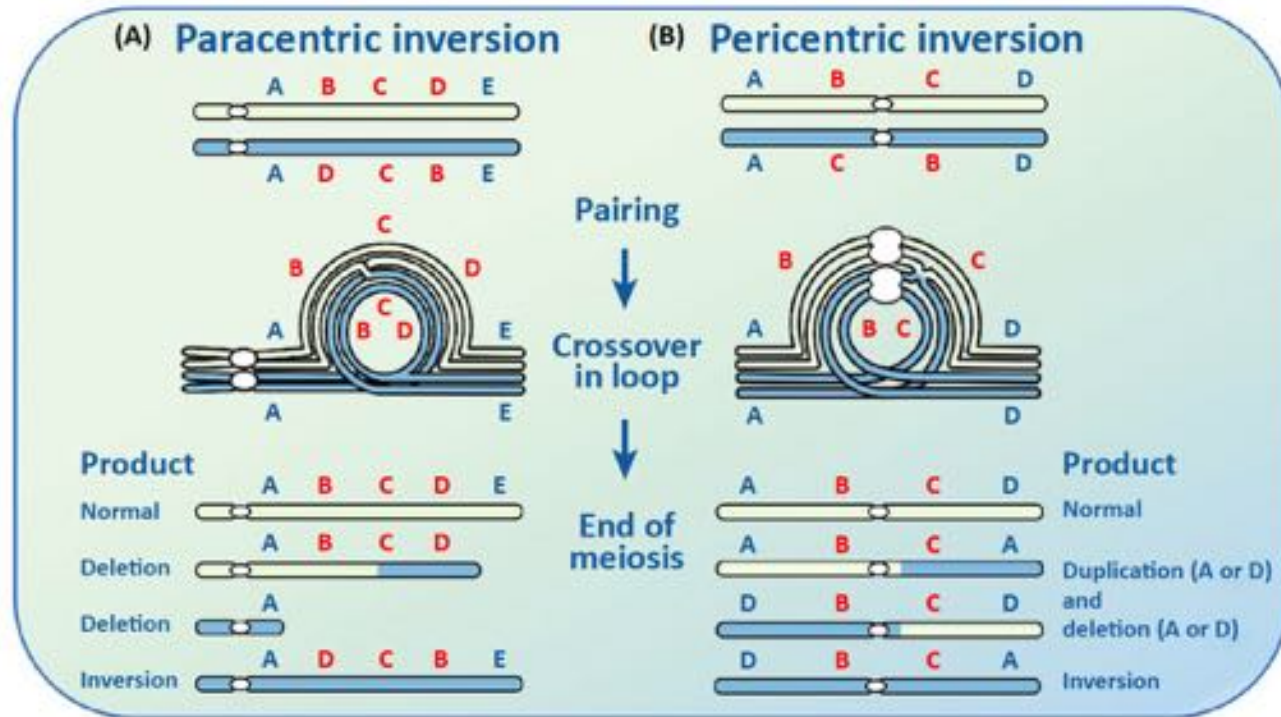
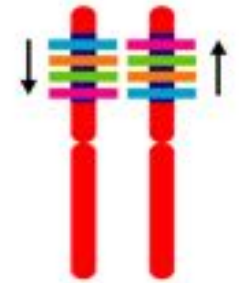


...and within chromosomes



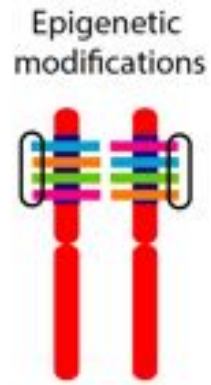
Chromosomal rearrangements

Chromosomal rearrangement
e.g. inversion



When chromosomes with different structural rearrangements recombine, arrangements different from the parental ones will likely carry deleterious deletions or duplications, thereby suppressing recombination in the areas affected by inversions or other structural variants.

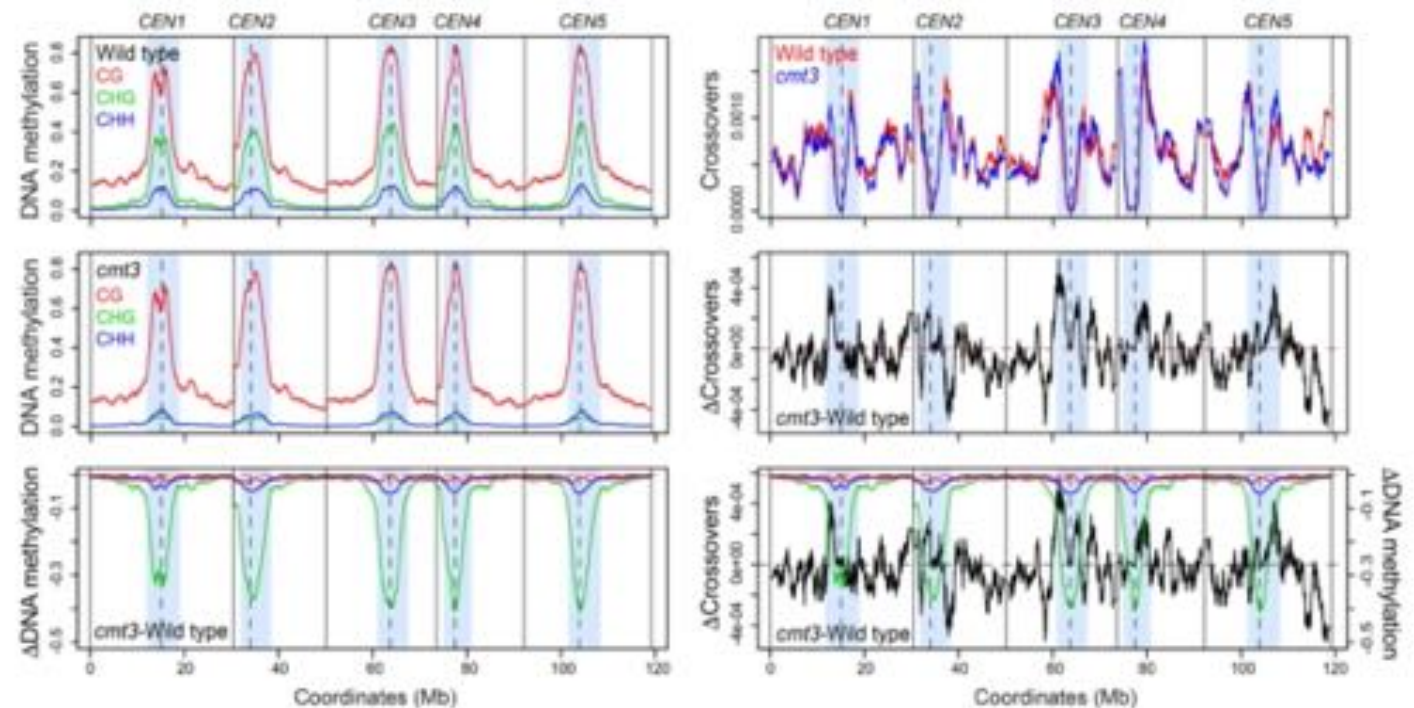
Wellenreuther & Bernatchez 2018, TREE



Epigenetic modifications

Methylation can suppress recombination locally, for example around centromeres.

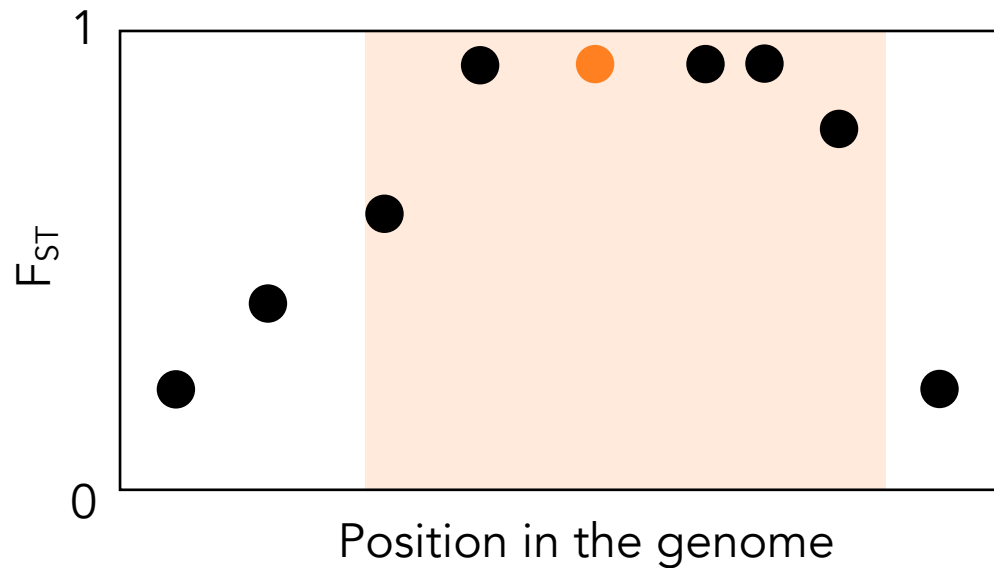
In *Arabidopsis thaliana* a mutation of the methyltransferase gene *cmt3* increases recombination at the centromeres.



Underwood et al. 2020, Genome Research

Genetic architecture of a trait

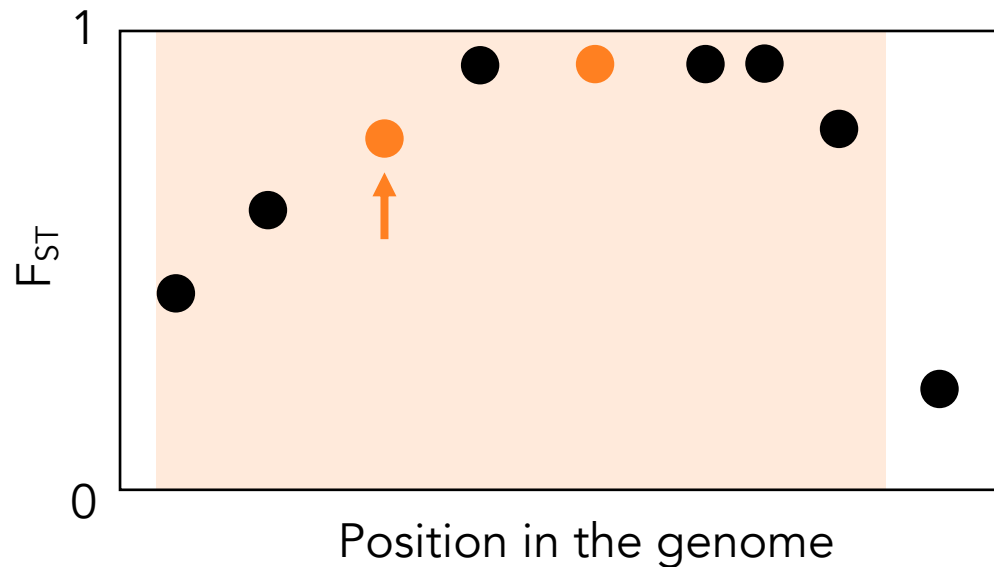
What we've seen so far was based on a single locus scenario, where one locus of large effect underlies the genetic basis of an adaptive trait.



One allele at one site increases in frequency due to selection, the others follow due to linked selection even though they are neutral.

Genetic architecture of a trait

If another nearby locus affects the same (or another adaptive trait), the probability of that beneficial mutation to increase in frequency and establish in the population is higher due to linked selection from the first locus.

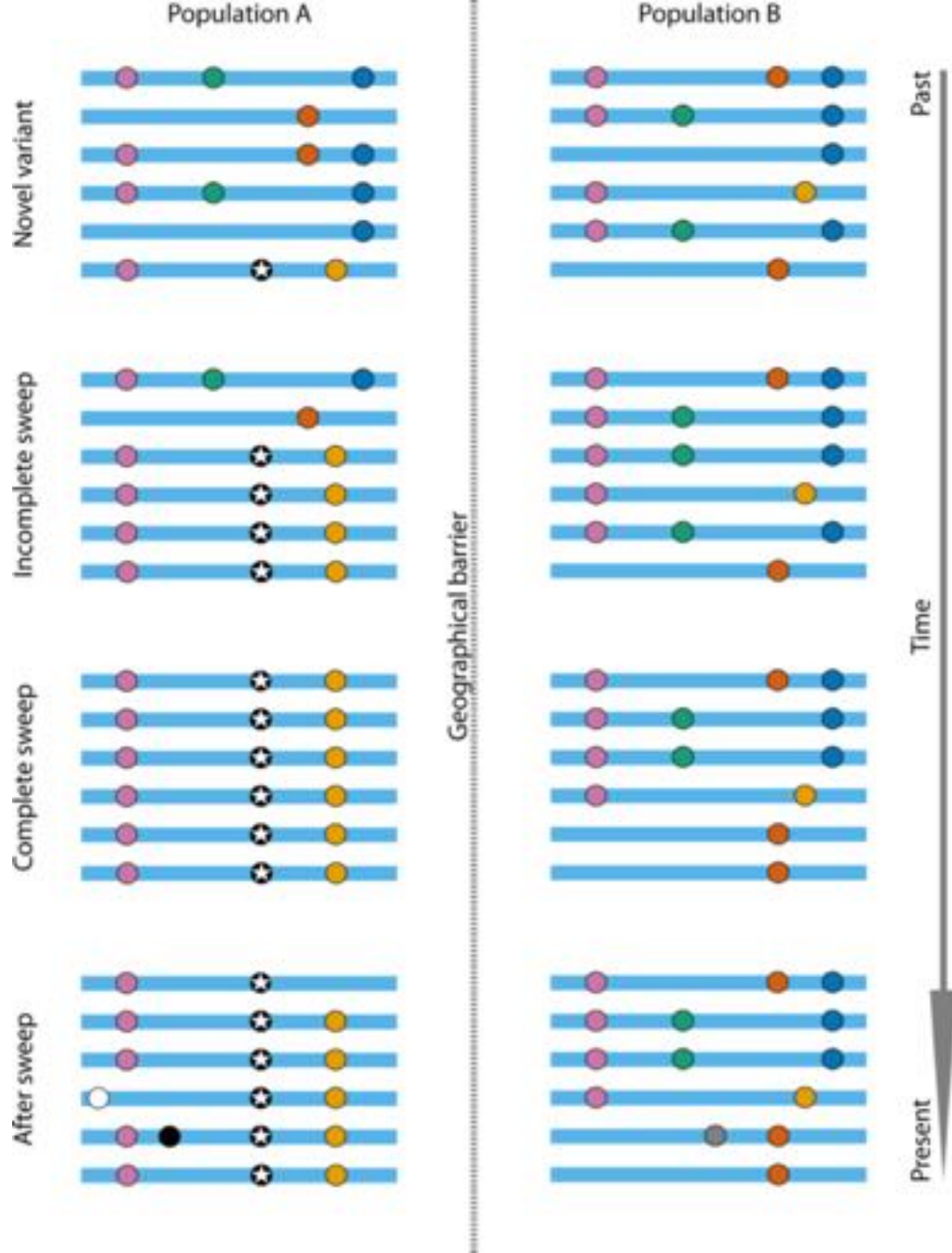


This leads to the evolution of clusters of adaptive loci showing large peaks of differentiation in the genome.

Genetic architecture of a trait

On the other hand, the genetic basis of a phenotype could be polygenic, i.e. controlled by many alleles of small effect.

In this case, the signature of selection won't be as evident as those typical of a selective sweeps.



Time

Adaptation is a process. Too early or too late in the process you may not be able to detect the diagnostic signature in the genome.

Factors to consider for your study design

- Gene flow
- Effective population size
- Recombination rate
- Genetic architecture
- Structural Variation
- Population differentiation
- Age of local adaptation
- Genome structure

Adaptation genomics

The main goal of adaptation genomics is to understand the genomic basis and architecture of adaptive traits

- Genetic basis of traits = loci that control the adaptive trait
- Genetic architecture = the interactions among alleles (dominance, epistasis, pleiotropy, polygeny)
- Genomic architecture = position of alleles and structural variants associated

Case study

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

The genomic basis of reproductive and migratory behaviour in a polymorphic salmonid

Anna Tigano  | Michael A. Russello 

SALMON OF NORTH AMERICA

ILLUSTRATED BY JOSEPH R. TORRESANI



Ecotypic differentiation in salmonids

The evolution of different ecotypes is common in Salmonids, making them great systems to study adaptation and speciation.

Ecotypic differentiation in sockeye

Oncorhynchus nerka



Migratory ecotypes

Anadromous (sockeye)

vs.

Resident (kokanee)

Ecotypic differentiation in sockeye



Migratory ecotypes

Anadromous (sockeye)

vs.

Resident (kokanee)

Spawning ecotypes

River/stream-spawning

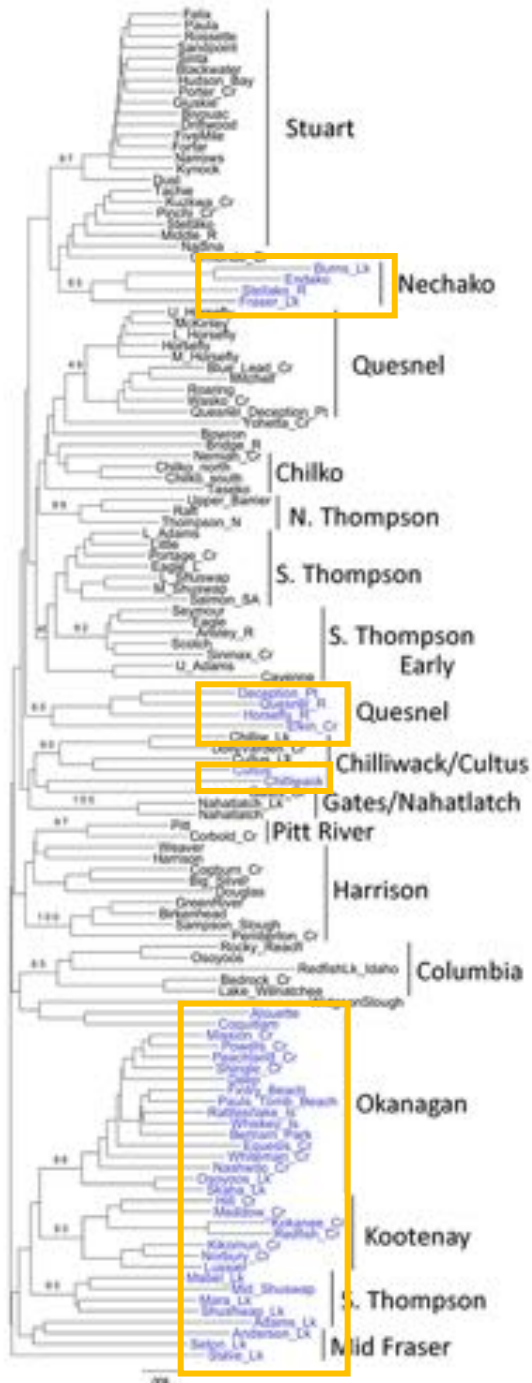
vs.

Beach/shore-spawning

Complex phylogenetic history

Kokanee has evolved from sockeye
multiple times.

Less is known about the evolution of
the spawning ecotypes.



What is the genomic basis and architecture of this phenotypic variation?

Genetic basis of traits:

several genes with different function

Genetic architecture:

polygenic, concentrated architectures

Genomic architecture:

structural variants maintaining clusters of differentiated loci, genome structure?



-  Okanagan Lake - shore
-  Okanagan Lake - stream
-  Okanagan River - Sockeye

Focus on Okanagan Lake/River

	Migratory	Spawning
--	-----------	----------

n=12



**Shore-spawning
kokanee**

No

Shore - late

n=12



**Stream-spawning
kokanee**

No

River - early

n=12



Sockeye

Yes

River - early



Sequence
variation

SNPs called from whole genome resequencing short read data (MGI, 7.5X) in *ANGSD* and *bcftools*

Structural
variation

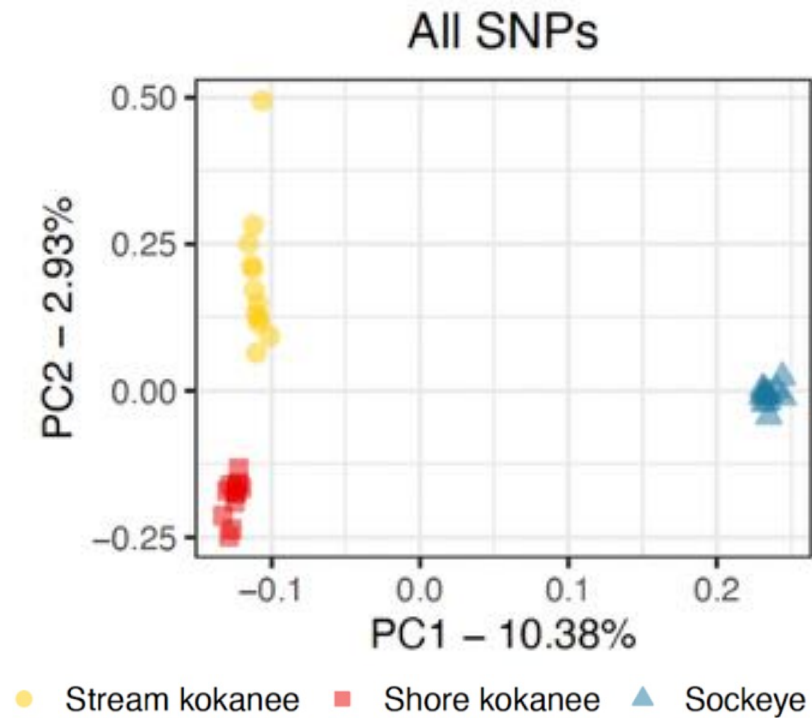
SVs called from whole genome resequencing short read data and Nanopore long reads for one shore- and one stream-spawning individual.

Genome
structure

Enabled by sockeye reference genome.
Areas of low recombination can show patterns resembling selective sweeps (e.g. centromeres)

Sequence
variation

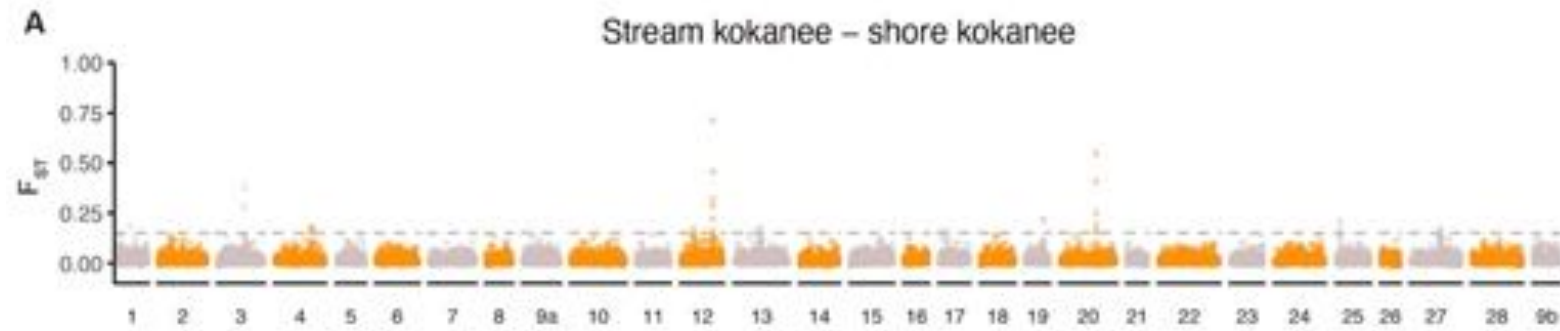
Differentiation among ecotypes



Differentiation based on 7.4
million SNPs called with
ANGSD

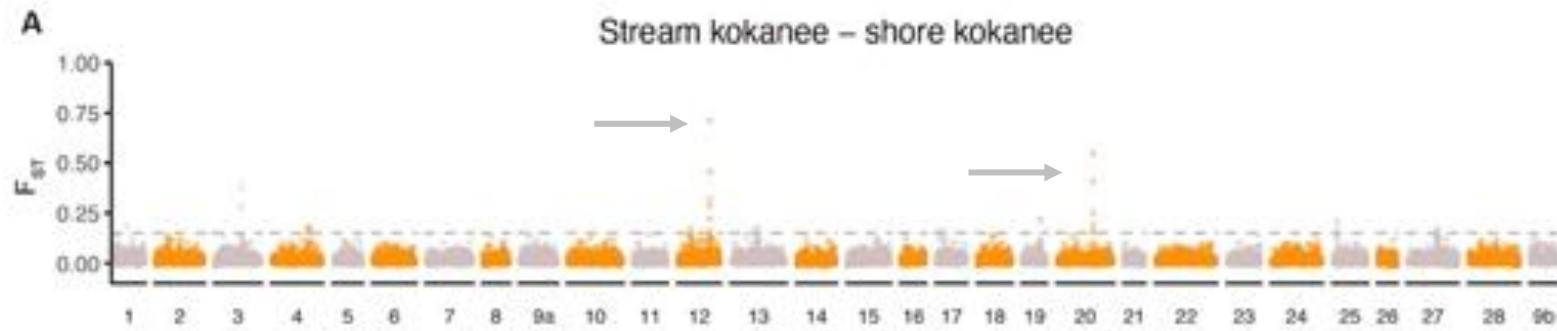
Genomic landscape of differentiation

$$F_{ST}=0.016$$



Genomic landscape of differentiation

$$F_{ST}=0.016$$



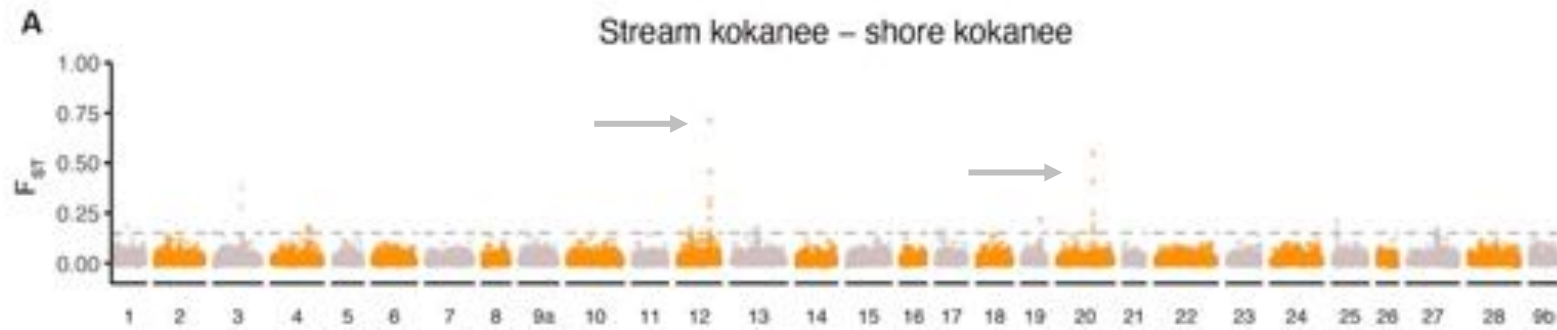
Outliers putatively associated with
differences between ecotype pairs

=

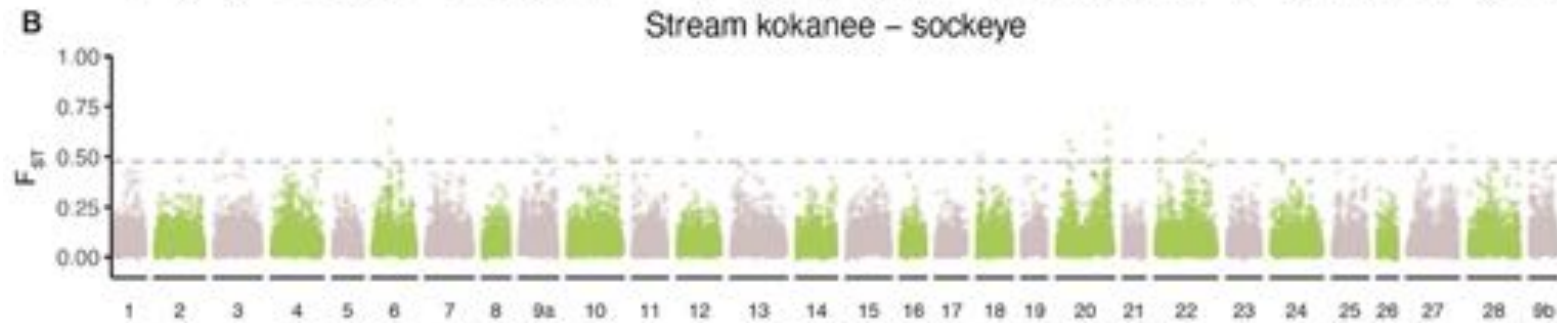
50 kb windows above the 99.9th percentile
of the empirical distribution

Genomic landscape of differentiation

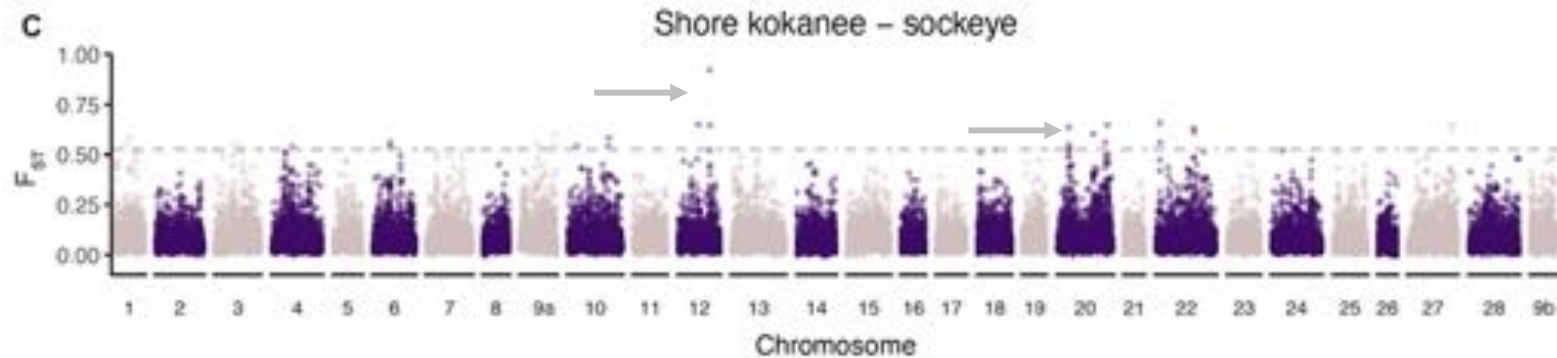
$$F_{ST}=0.016$$



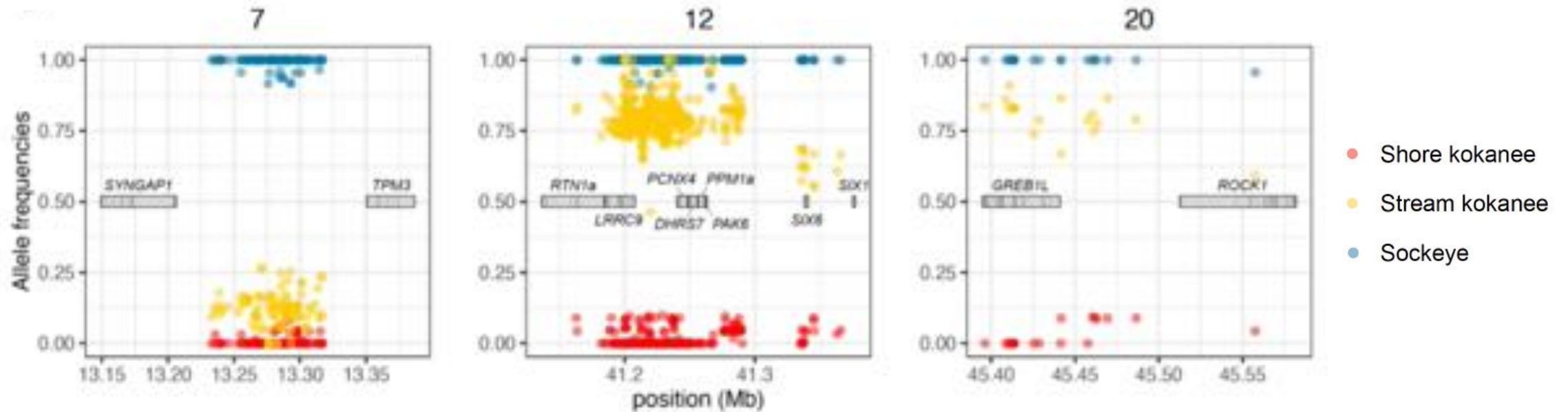
$$F_{ST}=0.077$$



$$F_{ST}=0.085$$

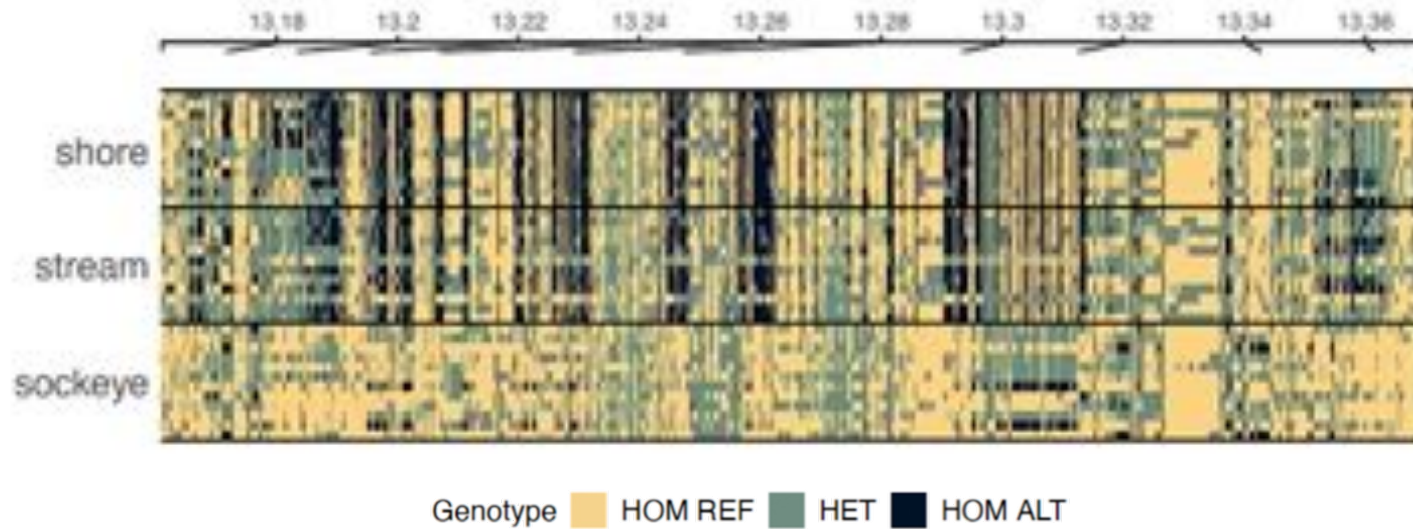


Nearly fixed differences

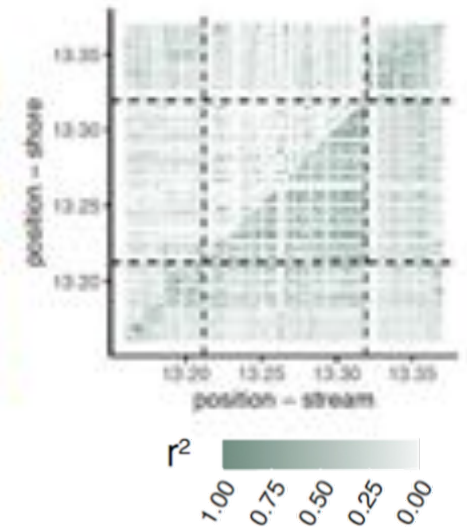


Fixed (or nearly fixed) differences only between shore kokanee and sockeye with highest concentration in 3 areas on 3 different chromosomes.

Area of differentiation on chr7 - migration

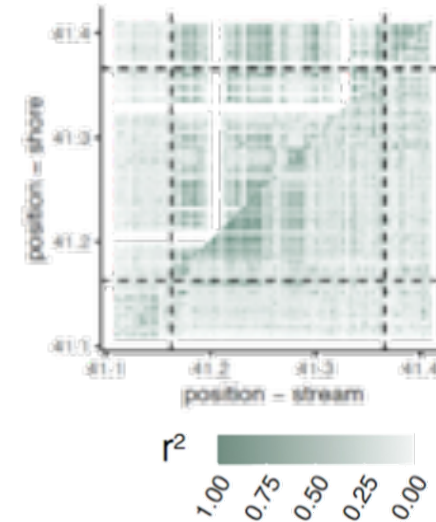
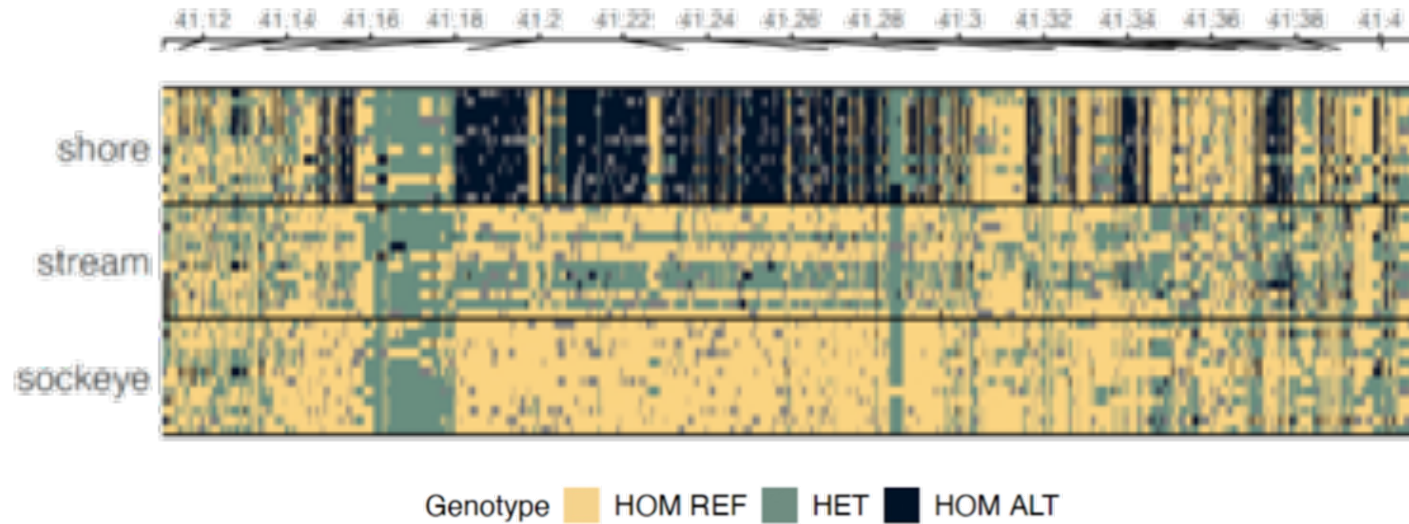


‘Resident’ allele appears
dominant over ‘migratory’ allele



Strong LD in AOD
and beyond in
heterozygous stream
kokanee

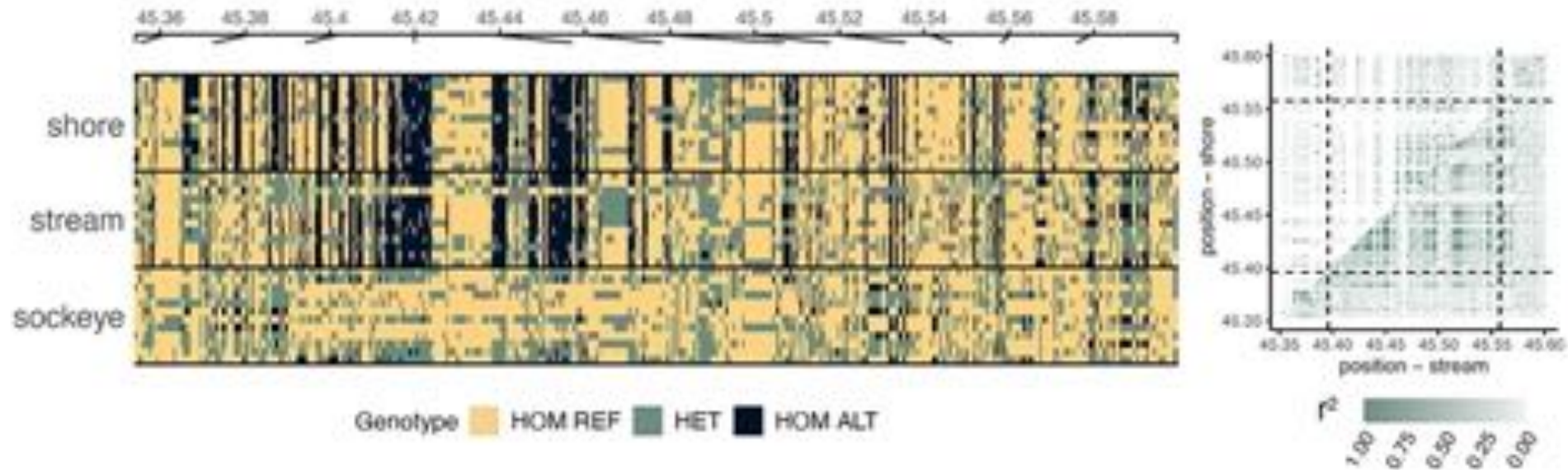
Area of differentiation on chr12 - spawning



‘Stream-spawning’ allele appears **dominant** over ‘shore-spawning’ allele

Strong LD in heterozygous stream kokanee, deserts of variation in shore kokanee
→ **strong selective sweep**

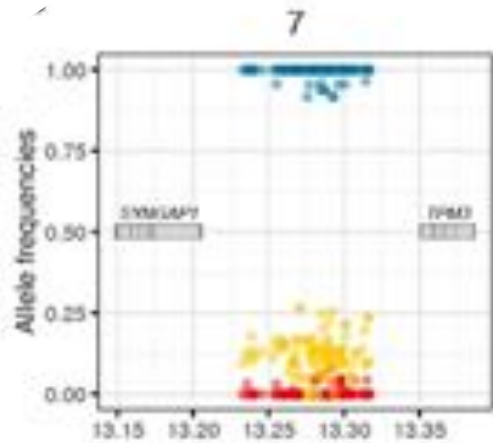
Area of differentiation on chr20 - spawning



Mosaic of 'shore-spawning'
and 'sockeye' alleles →
different aspects of spawning?

Strong LD in stream
kokanee, deserts of
variation in shore
kokanee → **strong
selective sweep**

Chr 7



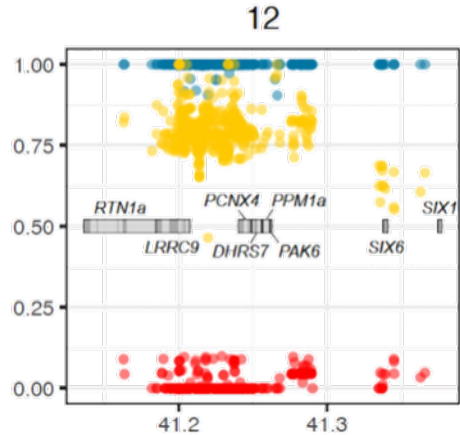
Candidates for migratory behaviour:

Expression regulation?

SYNGAP1 → working and reference spatial memory

TPM3 → swimming and muscle endurance

Chr 12



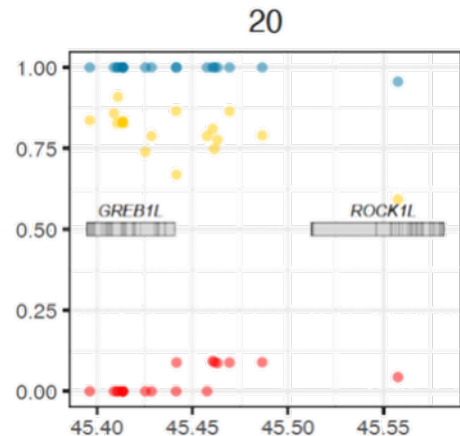
Candidates for spawning location:

Polygenic adaptation

LRRC9 previously associated with shore/stream spawning.

SIX6 and other genes are associated with eye and brain development (visual habitat?)

Chr 20



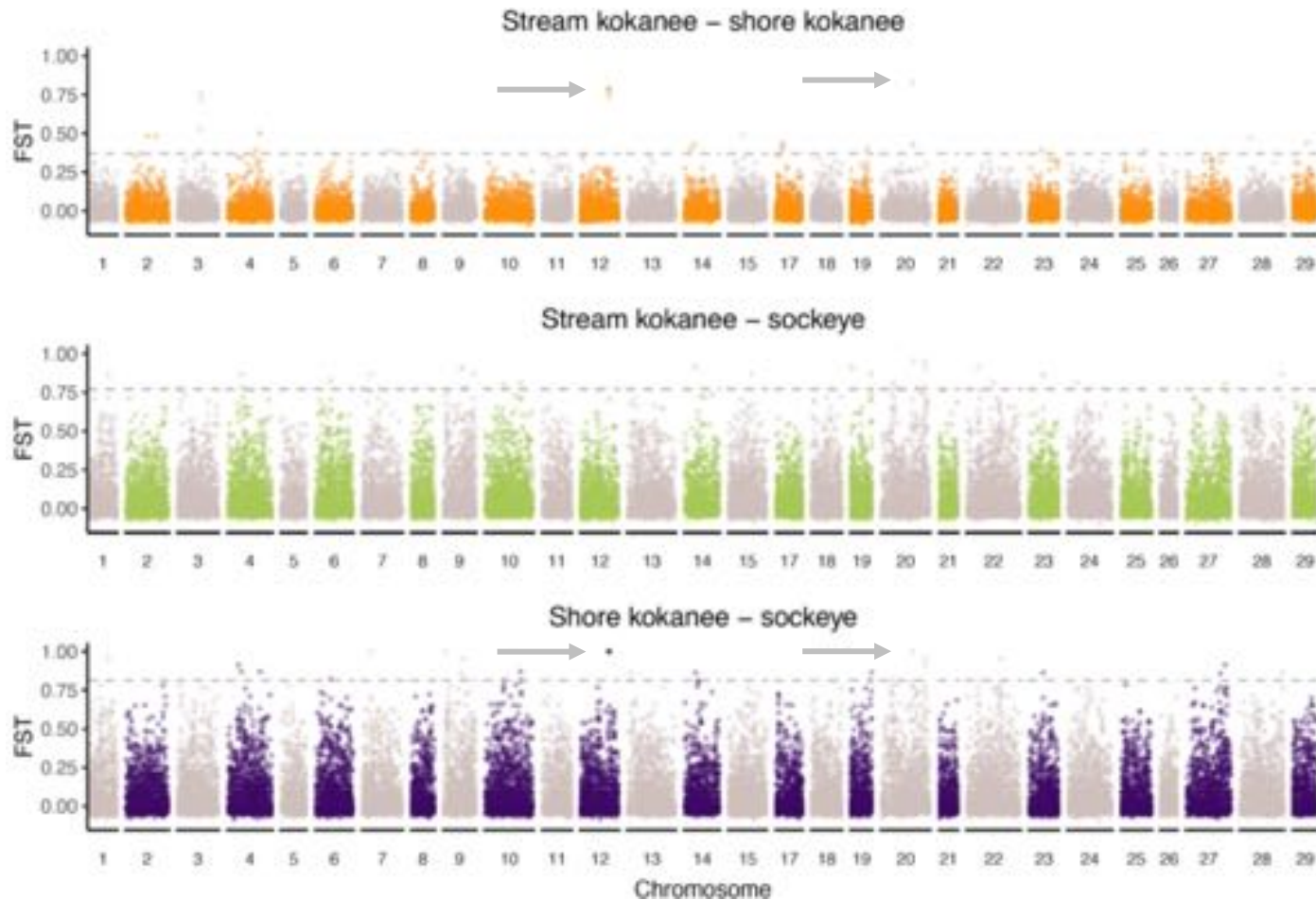
Candidates for timing of migration/spawning:

Coding and/or regulatory mechanisms

GREB1L and *ROCK1* are associated with timing of migration in steelhead trout and Chinook salmon.

Structural variation

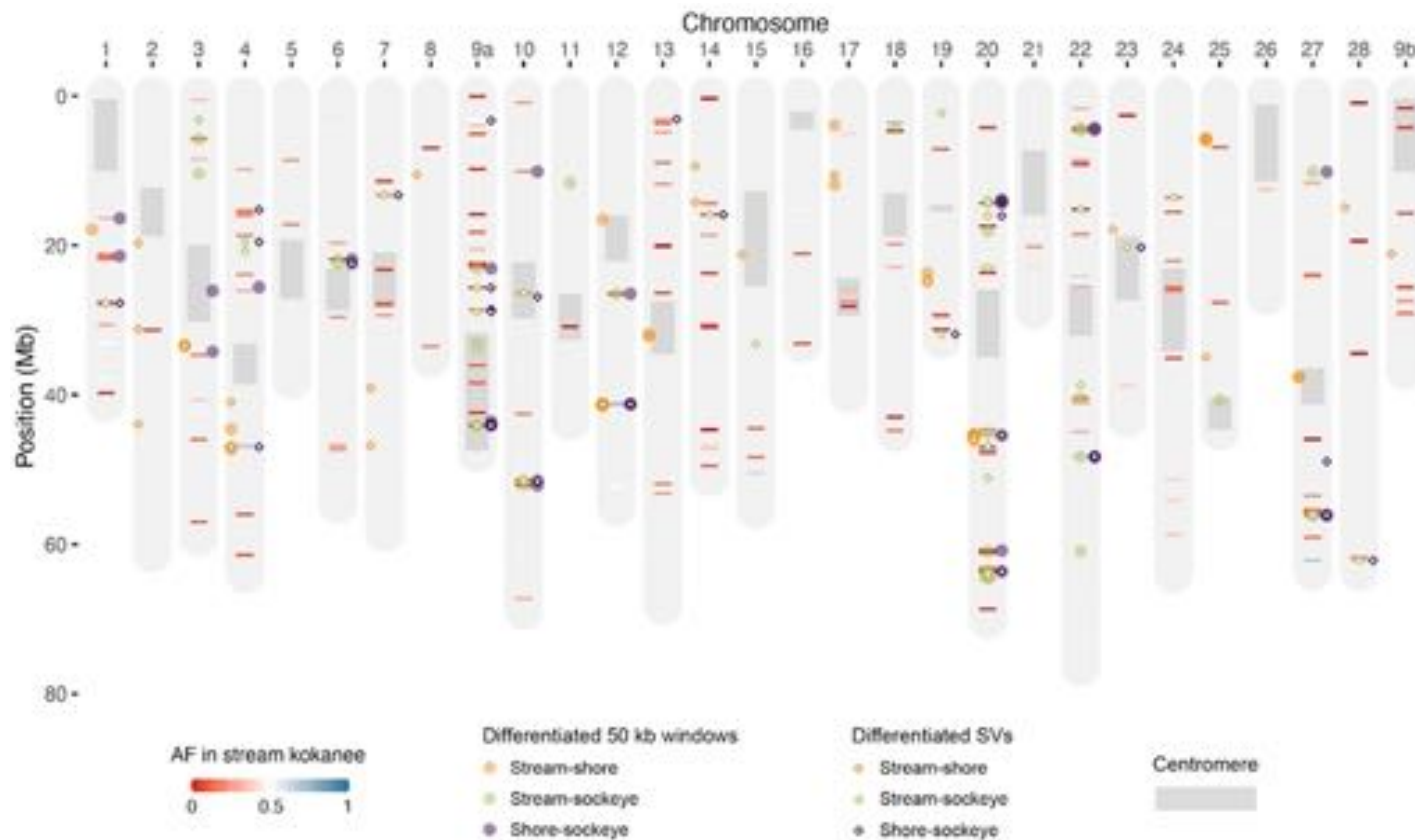
Many small SVs (mostly INDELS) associated with peaks of differentiation, but not large inversions.



- Causal variants?
 - Recombination suppression?
- Due to tight association between SNPs and SVs it is not possible to discern the two

Genome structure

Genome structure affects distribution of diversity, but centromeres are not enriched for candidate variants



→ 'true' outliers

→ strong LD among SNPs in areas of differentiation is due to SVs and/or strong selection

What is the genomic basis and architecture of this phenotypic variation?

Genetic basis of traits:

several genes with different function

Genetic architecture:

polygenic, concentrated architectures

Genomic architecture:

structural variants maintaining clusters of differentiated loci, genome structure



The end!