

# Signatures of selection

Day 2 - Lecture 1

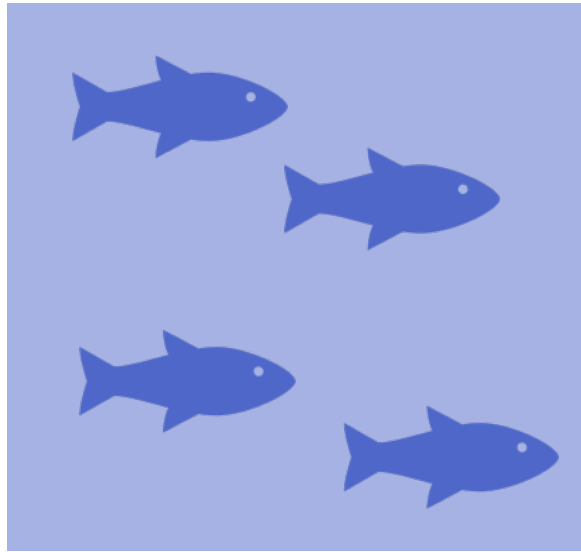
Anna Tigano, Ph.D.

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

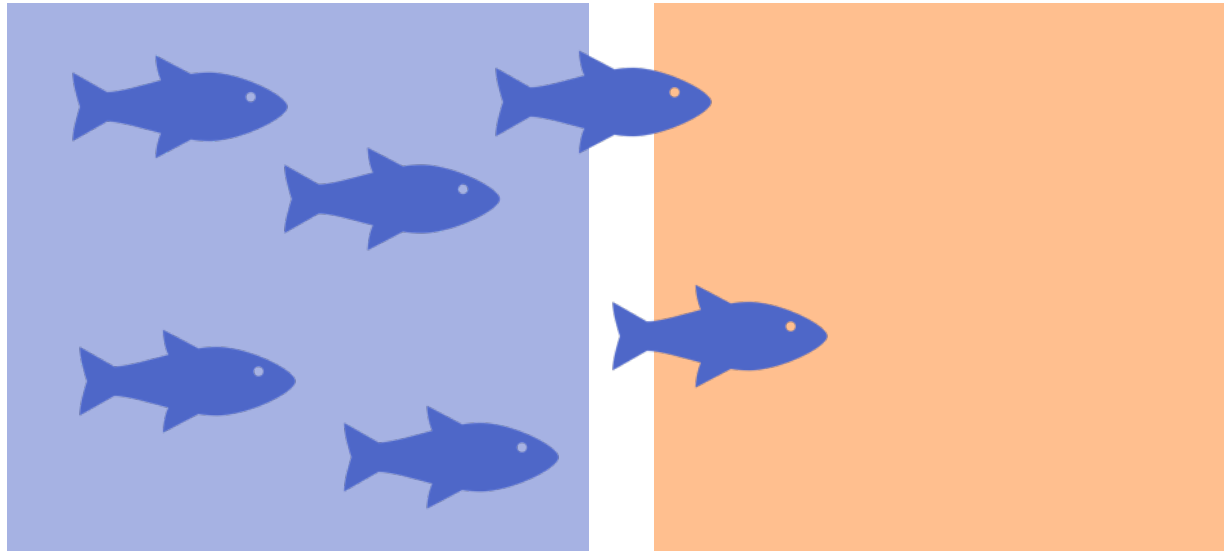
If we identify the loci underpinning the genetic basis of an adaptive trait we can better understand

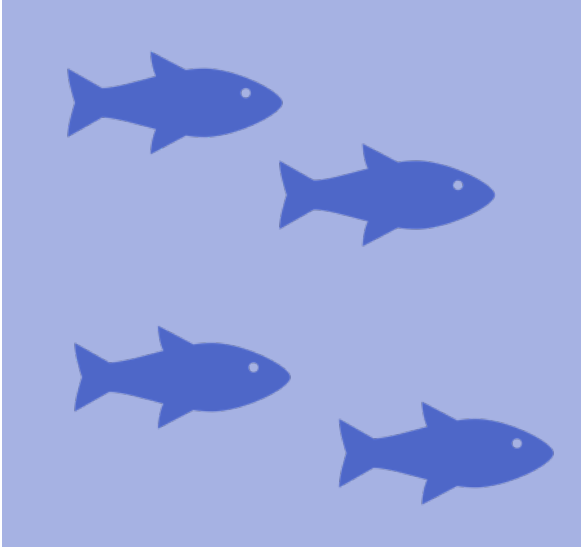
- the adaptive trait itself (e.g. its molecular and physiological basis)
- the process of adaptation (including the interplay with other evolutionary forces)

A population of cold-adapted fish

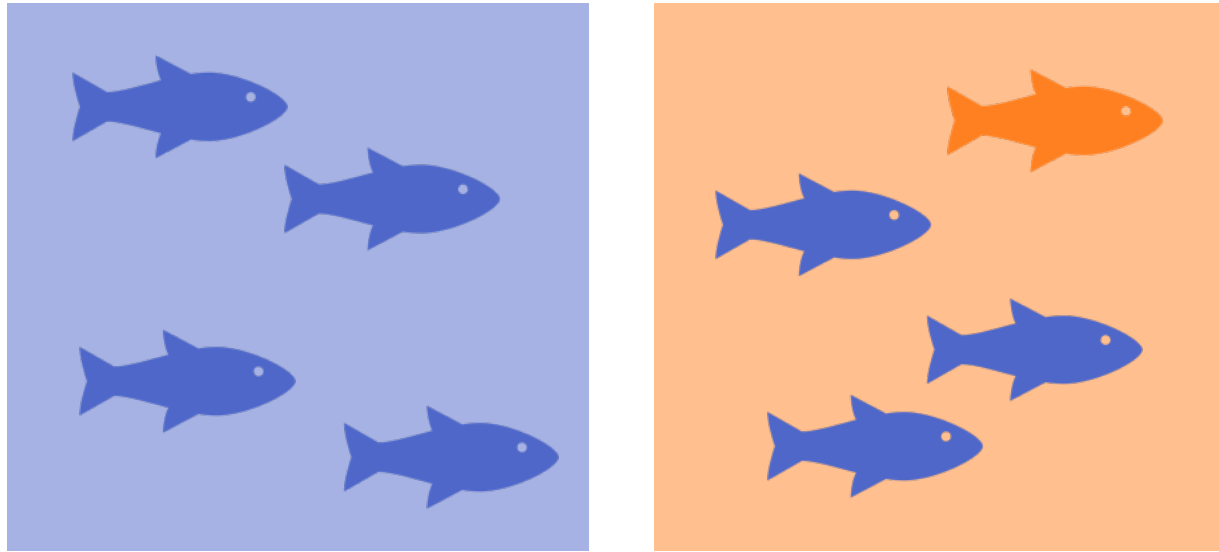


Cold-adapted fish migrate to warmer waters

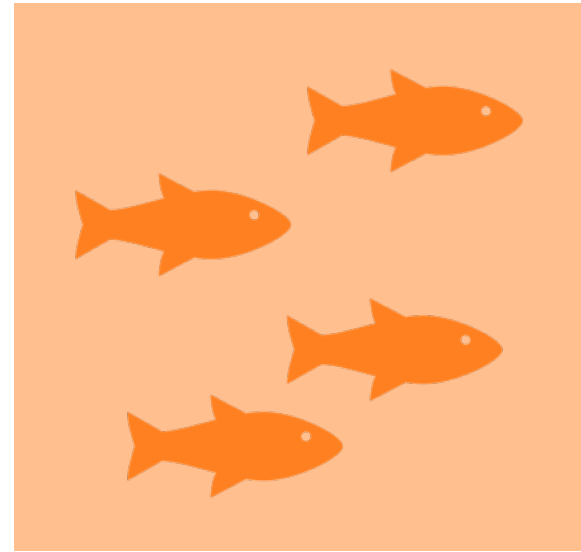
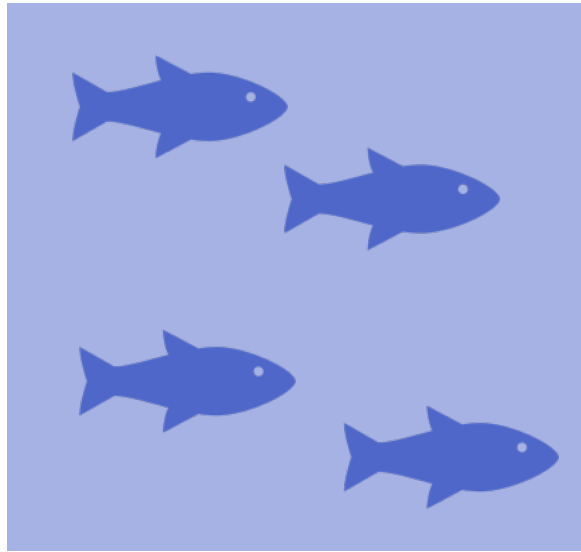




A new mutation confers a fitness benefit in warmer waters

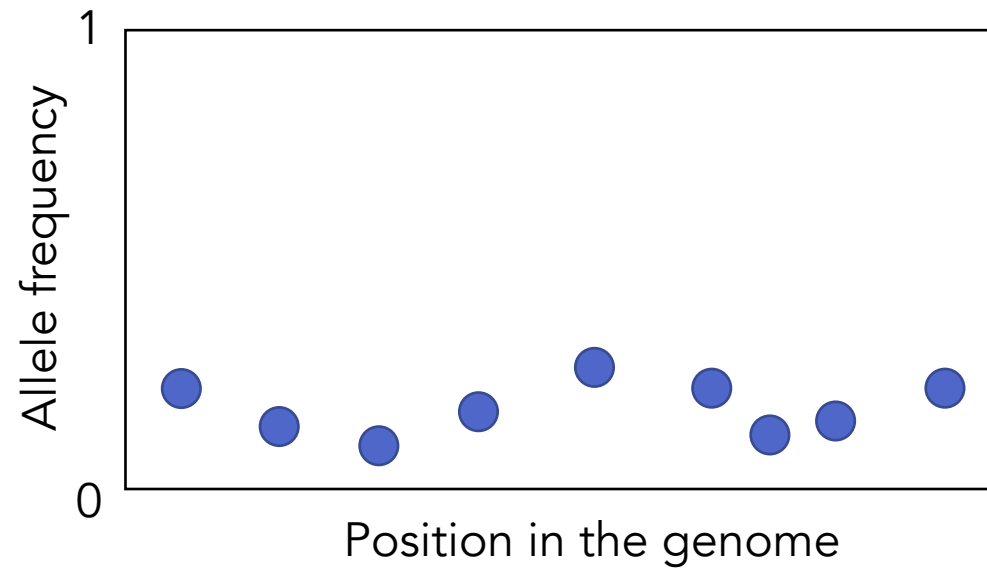


The new mutation spreads quickly in the population and reach fixation



# Changes in allele frequencies

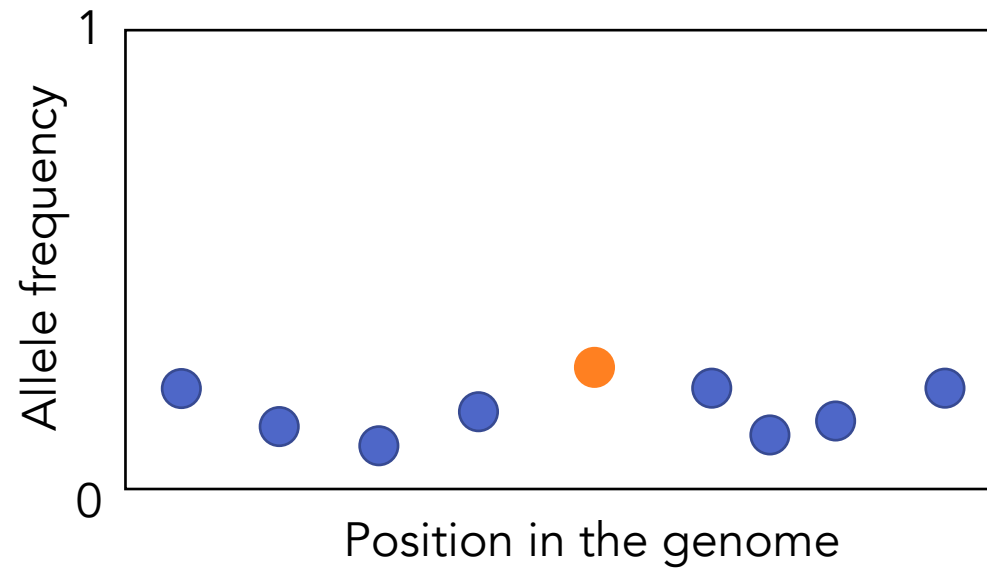
Cold-adapted fish migrate to warmer waters





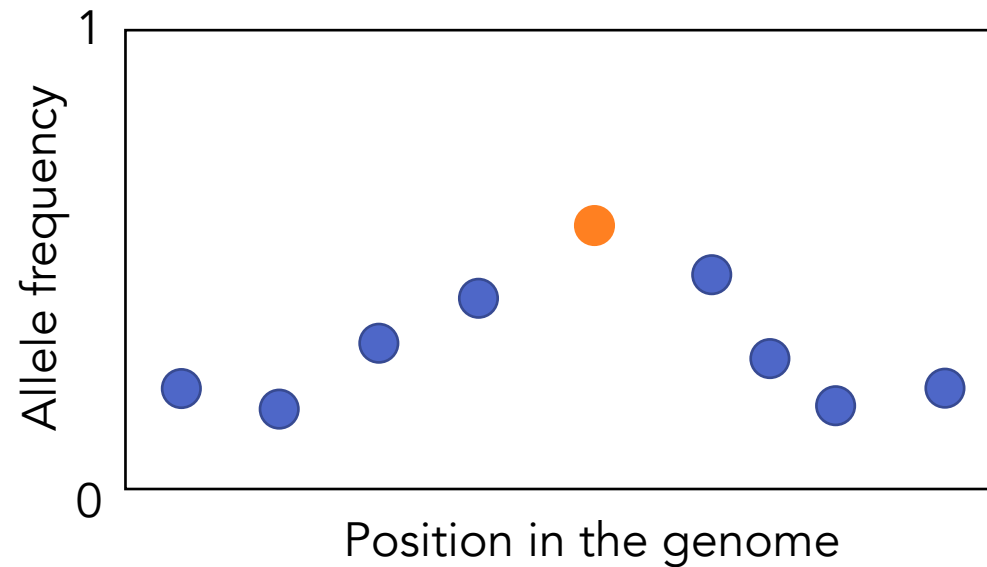
# Changes in allele frequencies

A new mutation confers a fitness benefit in warmer waters



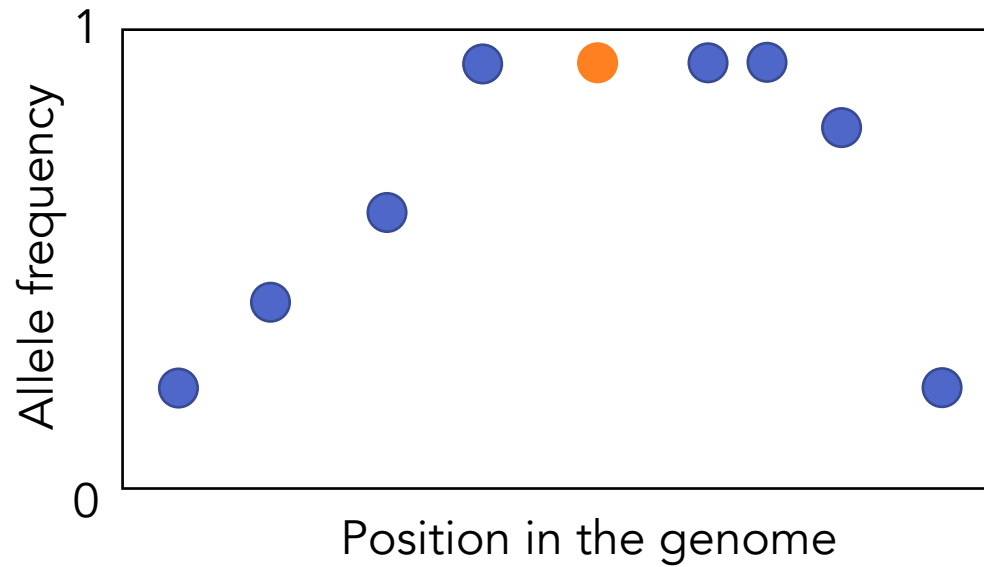
# Changes in allele frequencies

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



# Changes in allele frequencies

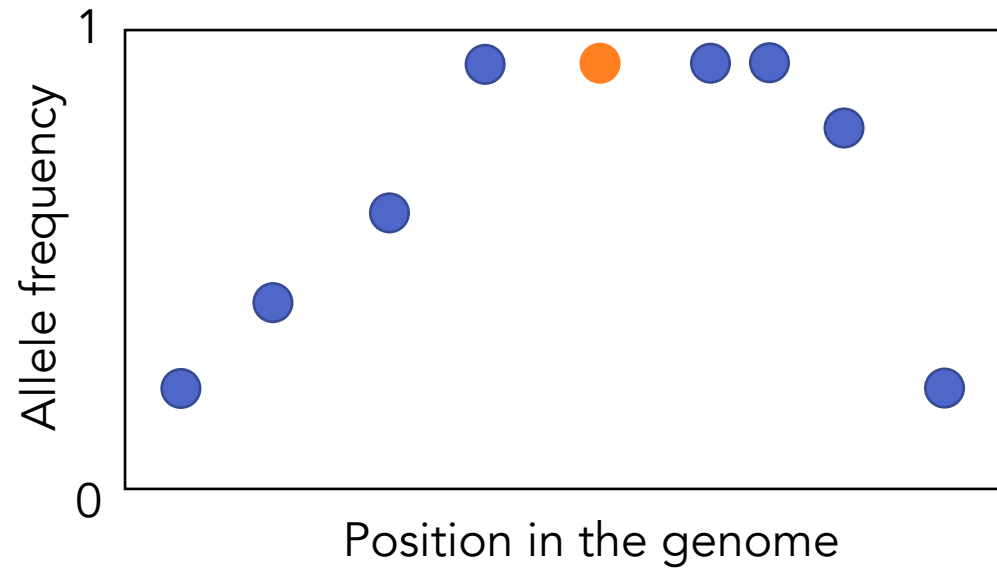
The new mutation spreads quickly in the population and reaches fixation.



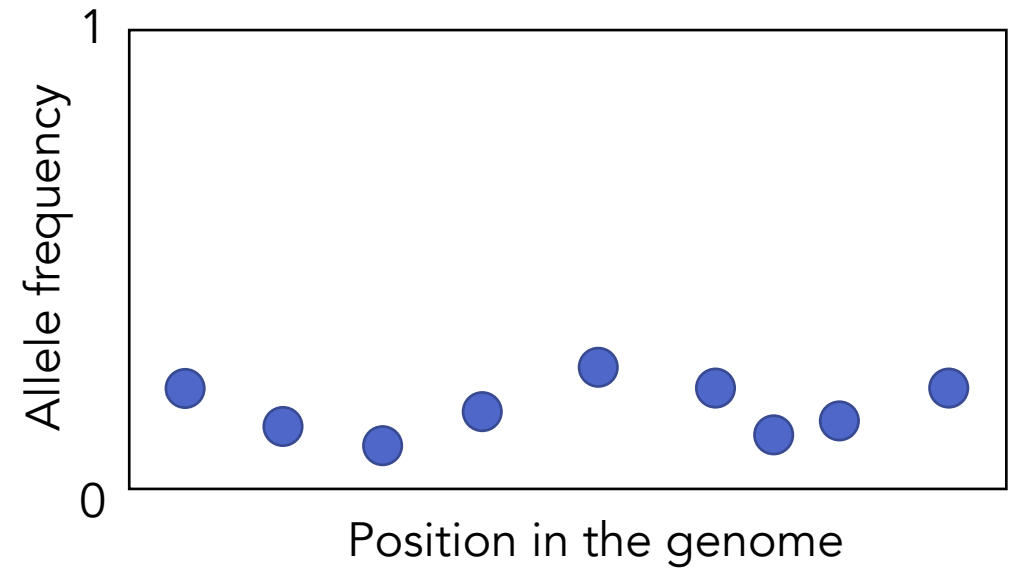
# Changes in allele frequencies



Warm-adapted



Cold-adapted



# How do we measure these changes?

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

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

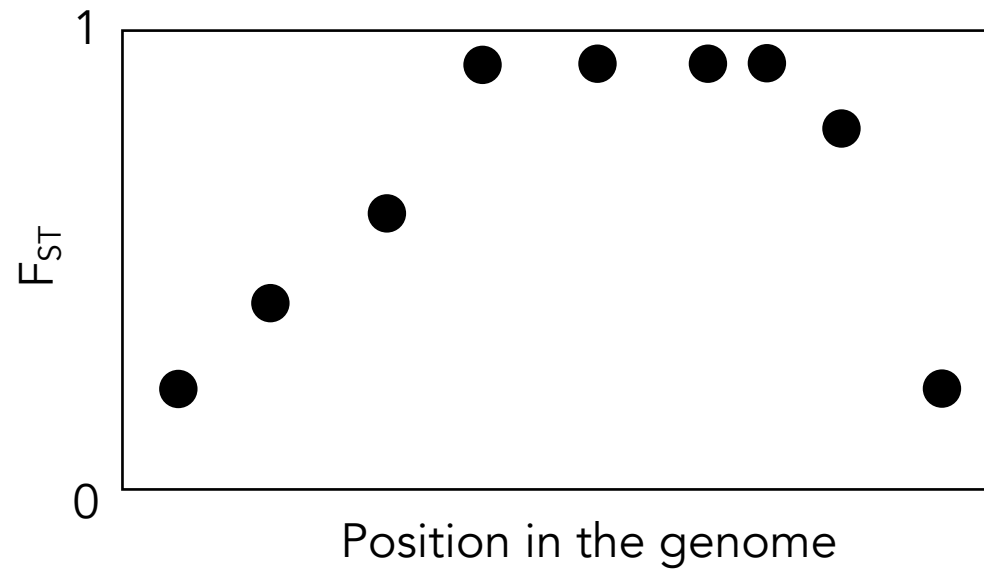
$\pi_T$  = expected heterozygosity in the total sample

$\pi_S$  = expected heterozygosity in each population

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

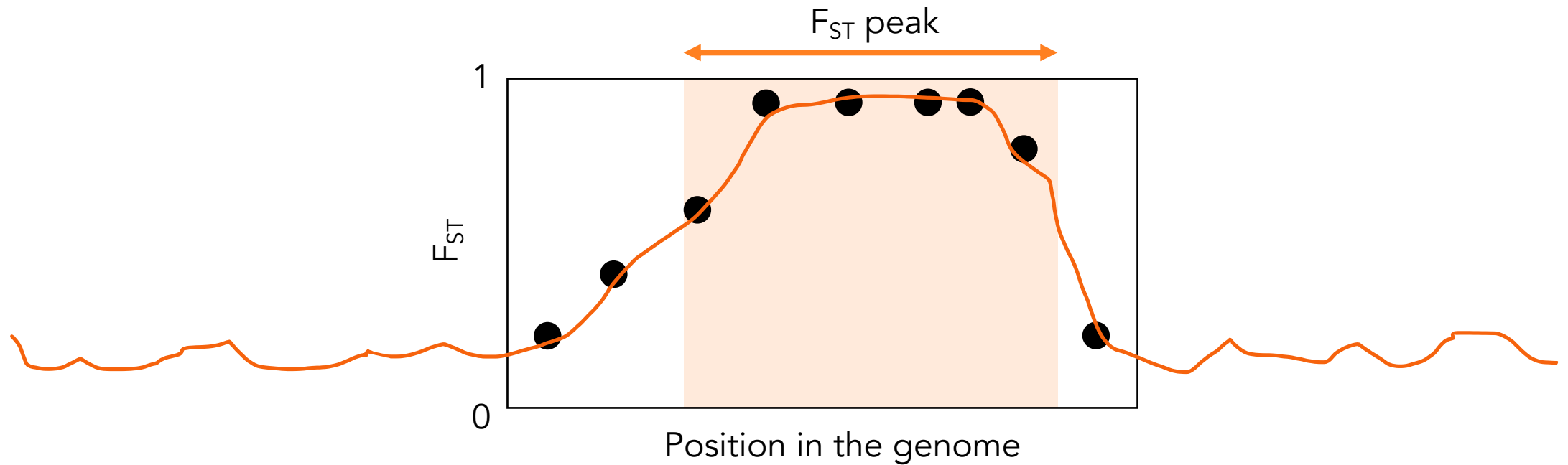
# Genetic differentiation

Differentiation between cold- and warm-adapted populations



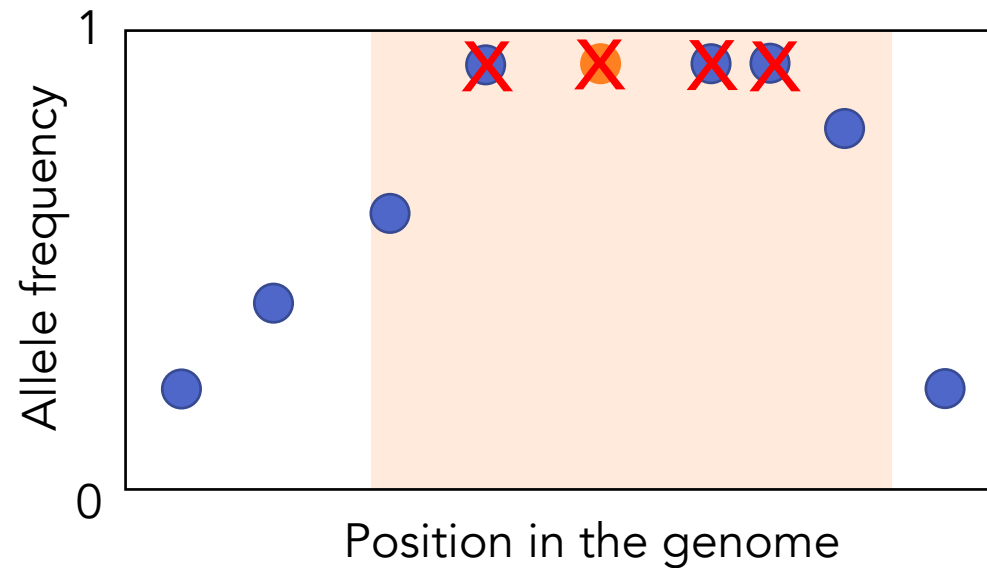
# Genetic differentiation

Differentiation between cold- and warm-adapted populations



# Reduction in diversity

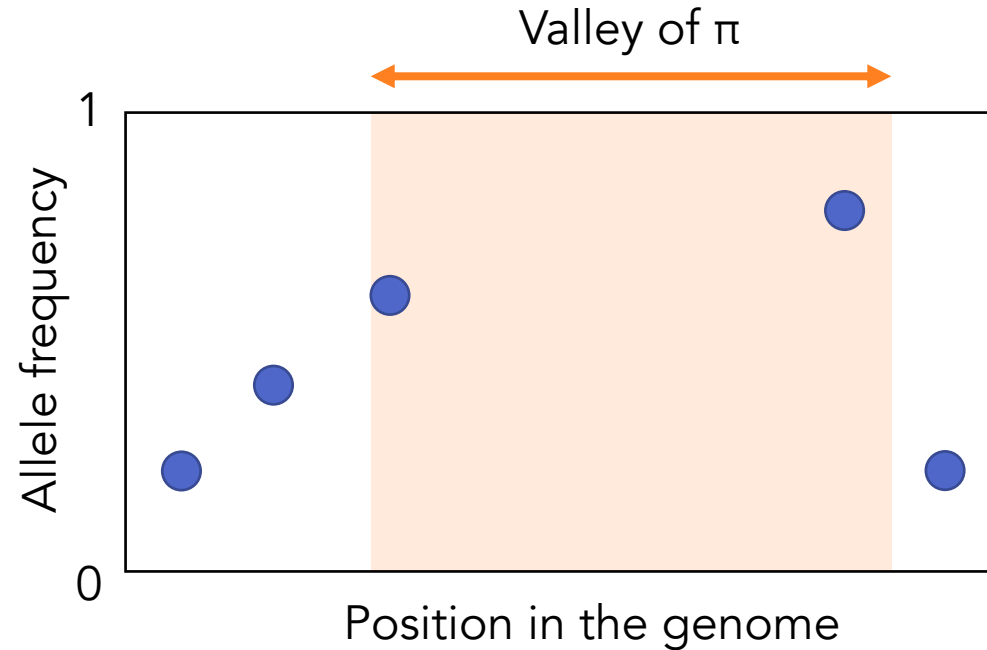
Once the beneficial allele has been brought to fixation, variation at the target of selection and surrounding sites is eliminated from the population.



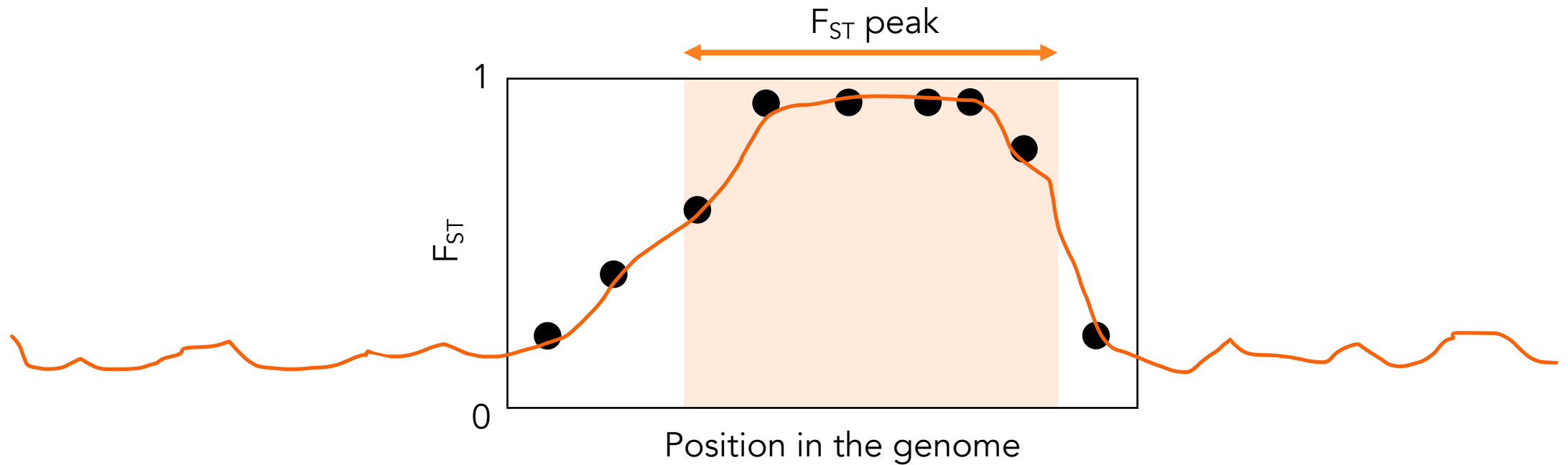


# Reduction in diversity

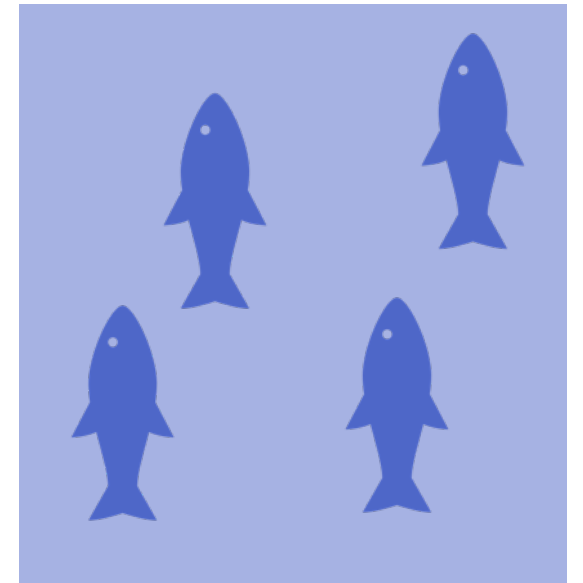
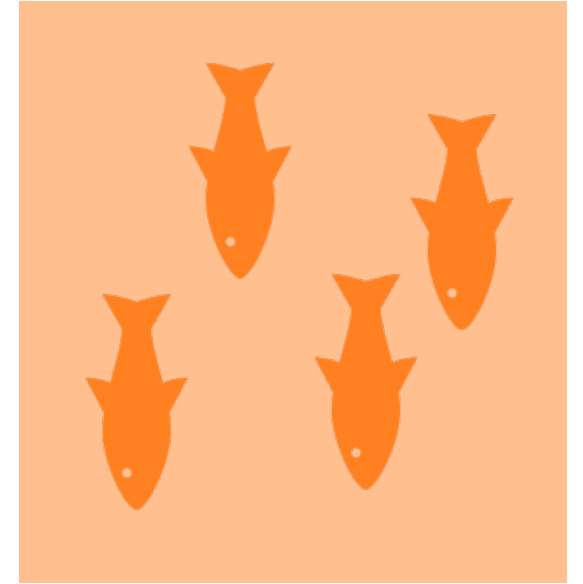
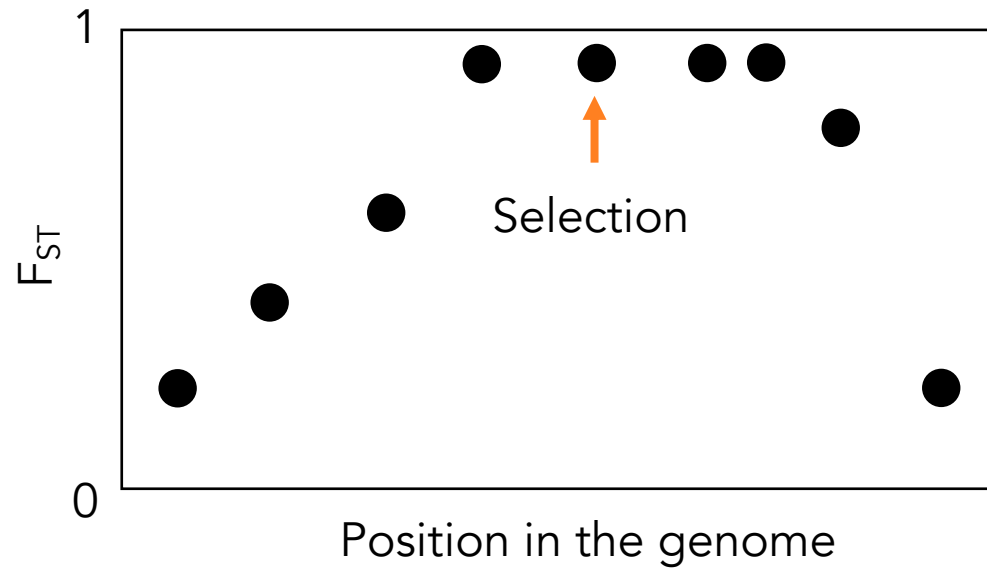
Once the beneficial allele has been brought to fixation, variation at the target of selection and surrounding sites is eliminated from the population.



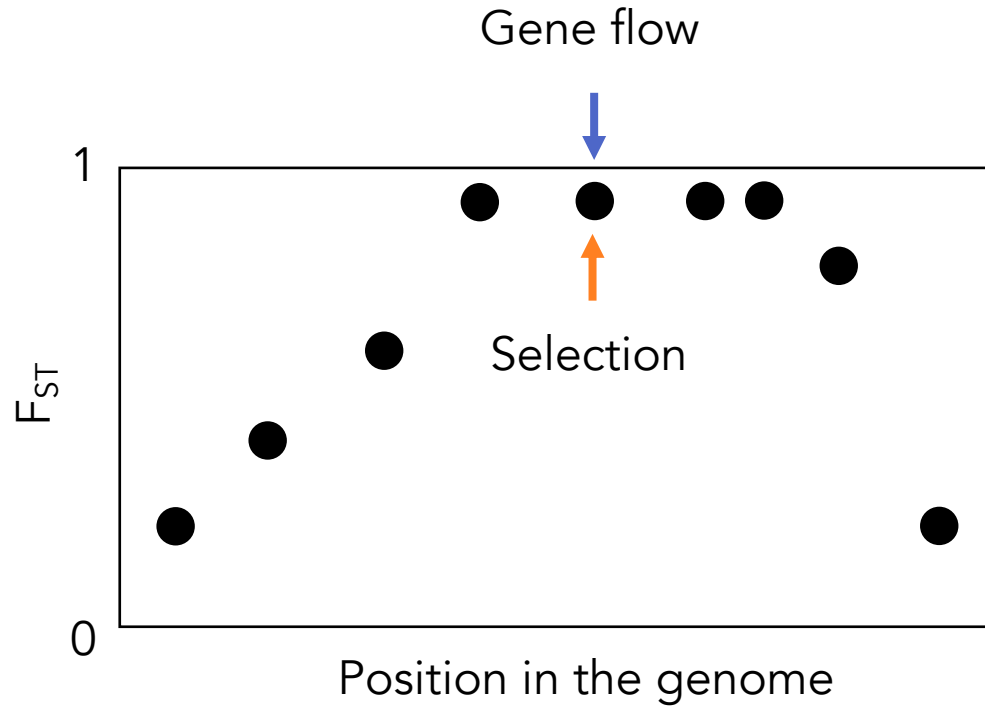
# What determines the size of a $F_{ST}$ peak?



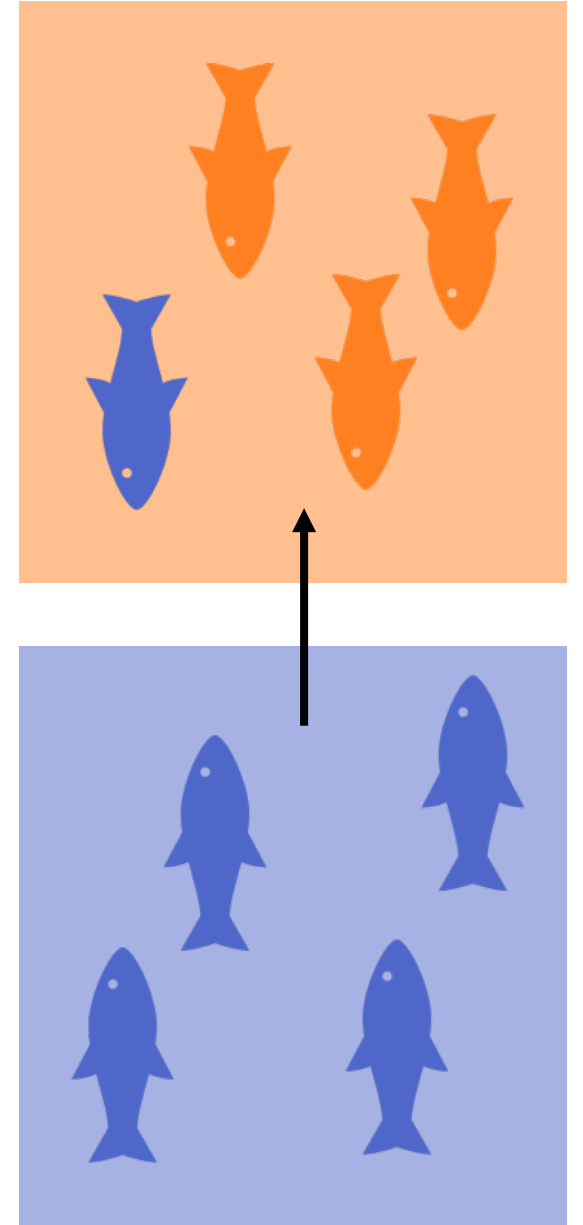
# Selection vs. gene flow



# Selection vs. gene flow

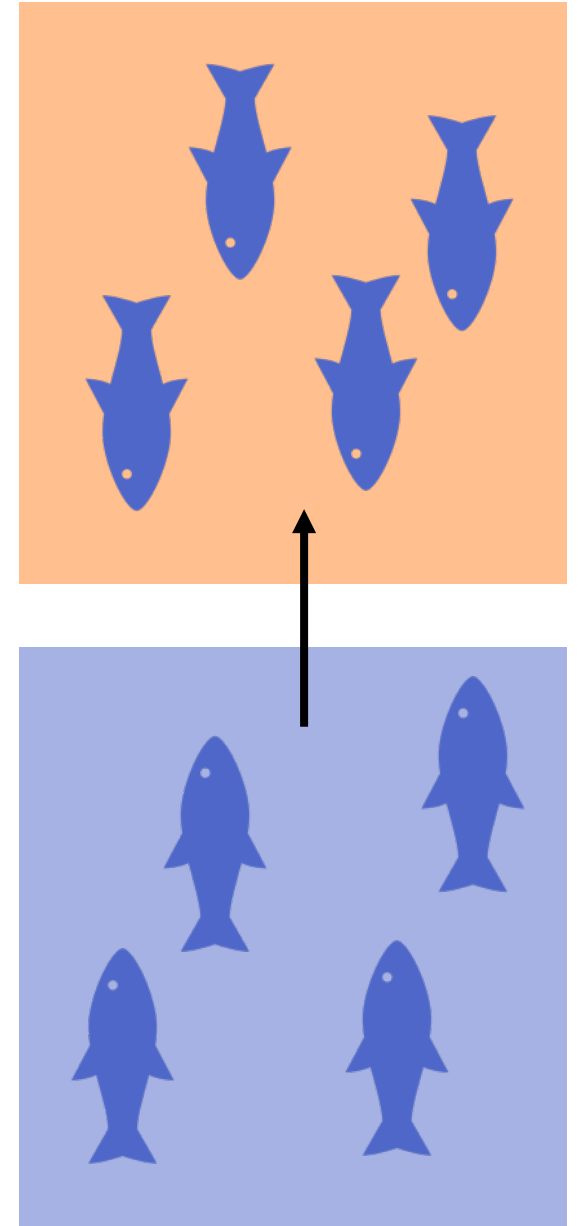


If gene flow introduces maladaptive alleles, the strength of selection  $s$  will determine whether the locally adaptive allele will be swamped.



# 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.

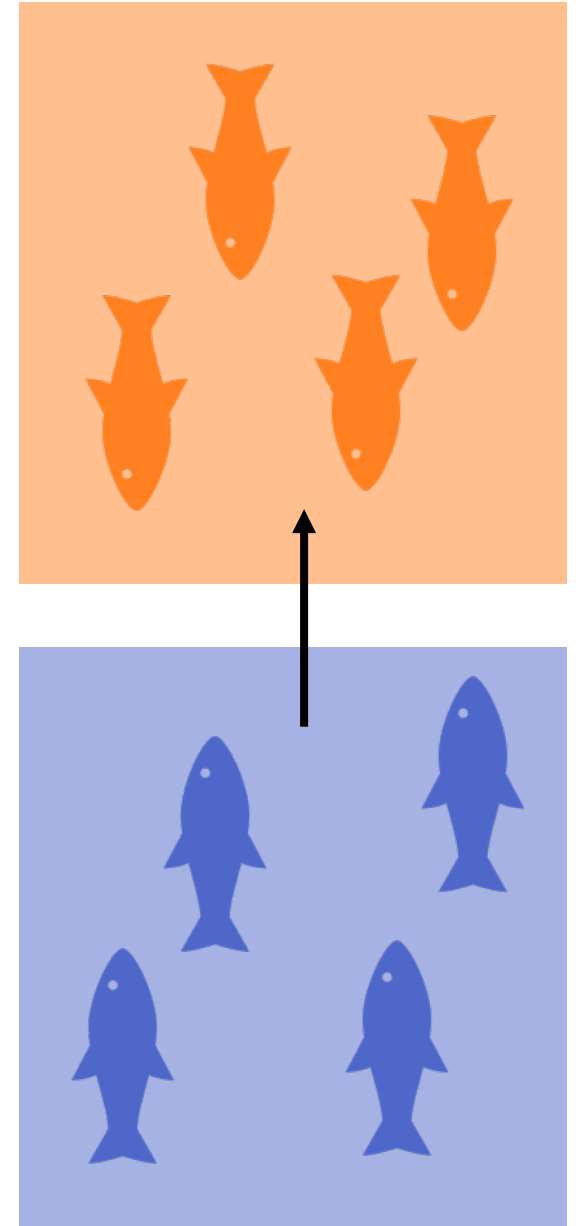


# Selection vs. gene flow

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

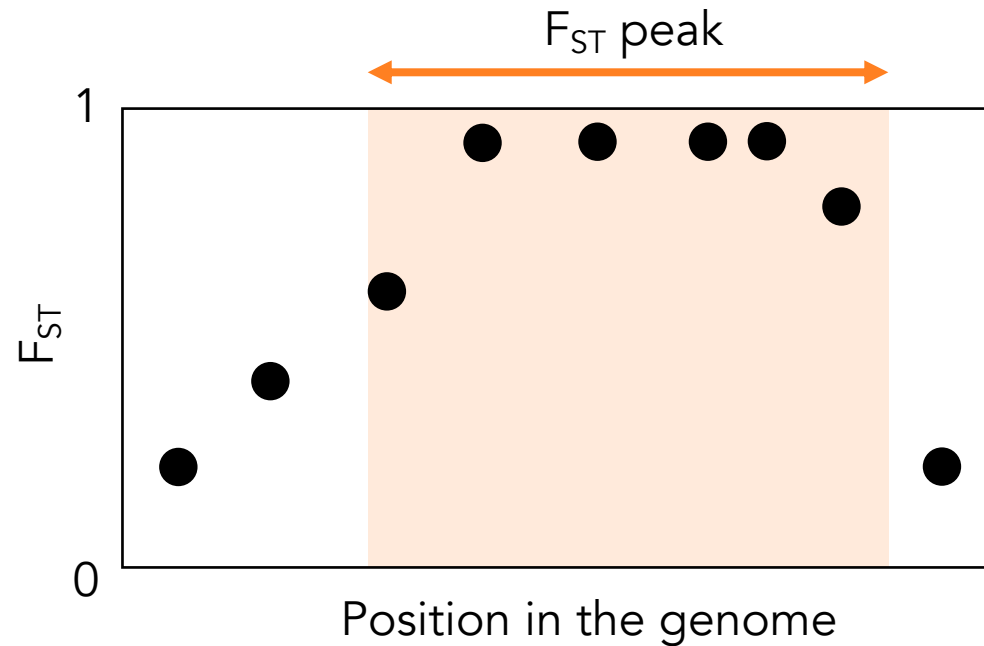
The locally adaptive allele will be maintained but overall population differentiation will be homogenized by gene flow through recombination.

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



# Physical linkage and recombination

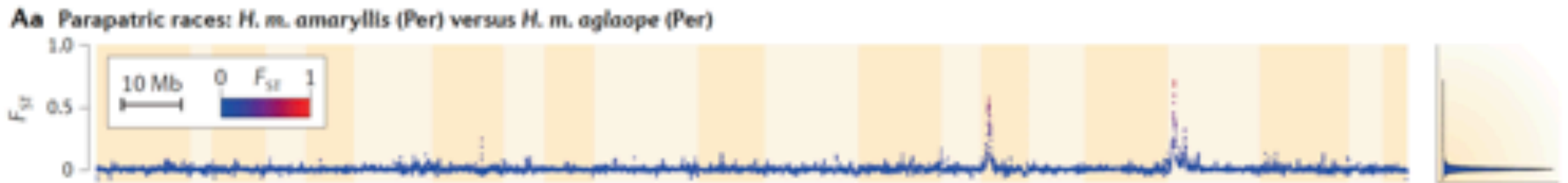
Recombination breaks linkage among sites at a rate that rapidly decays with distance.



# 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.

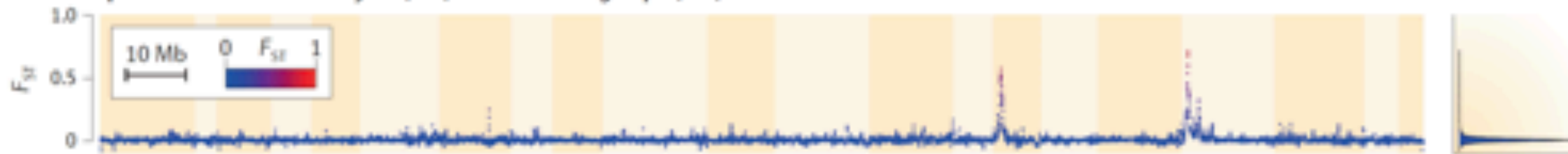
So that adaptive loci can stand out from a genomic background of low differentiation



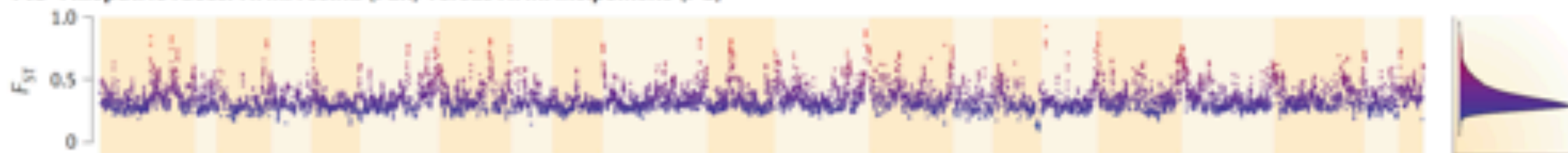
Seehausen et al. 2014, Nat. Gen. Rev.



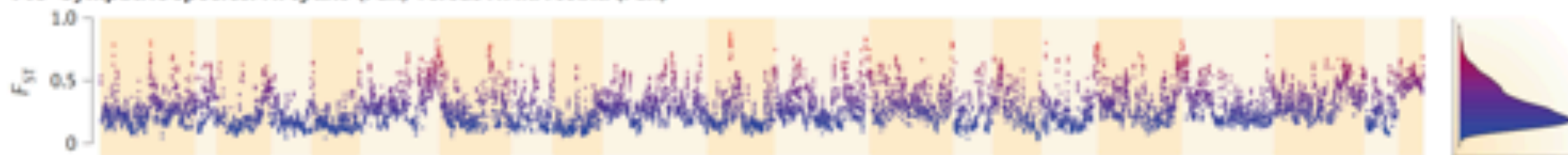
**Aa** Parapatric races: *H. m. amaryllis* (Per) versus *H. m. aglaope* (Per)



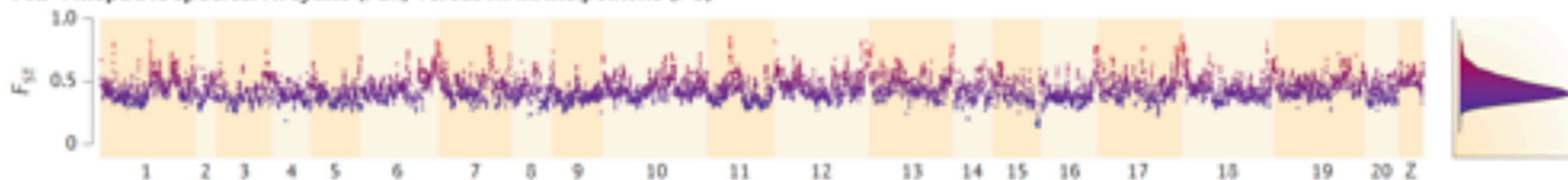
**Ab** Allopatric races: *H. m. rosina* (Pan) versus *H. m. melpomene* (FG)



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

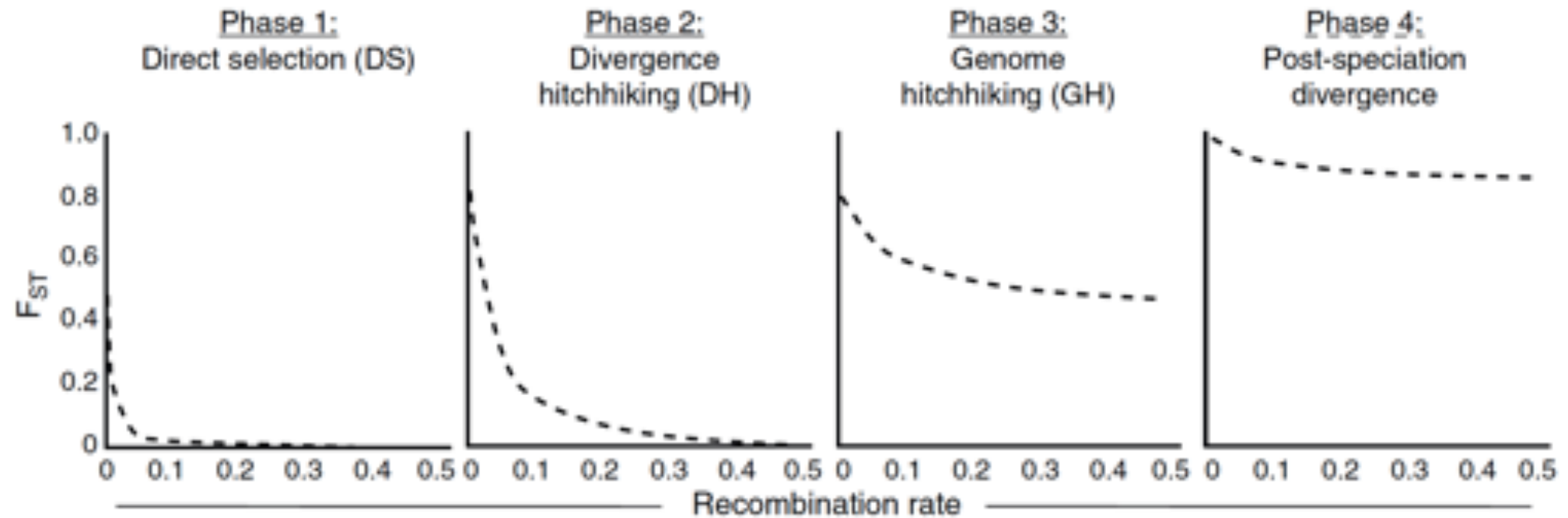


**Ad** Allopatric species: *H. cydno* (Pan) versus *H. m. melpomene* (FG)



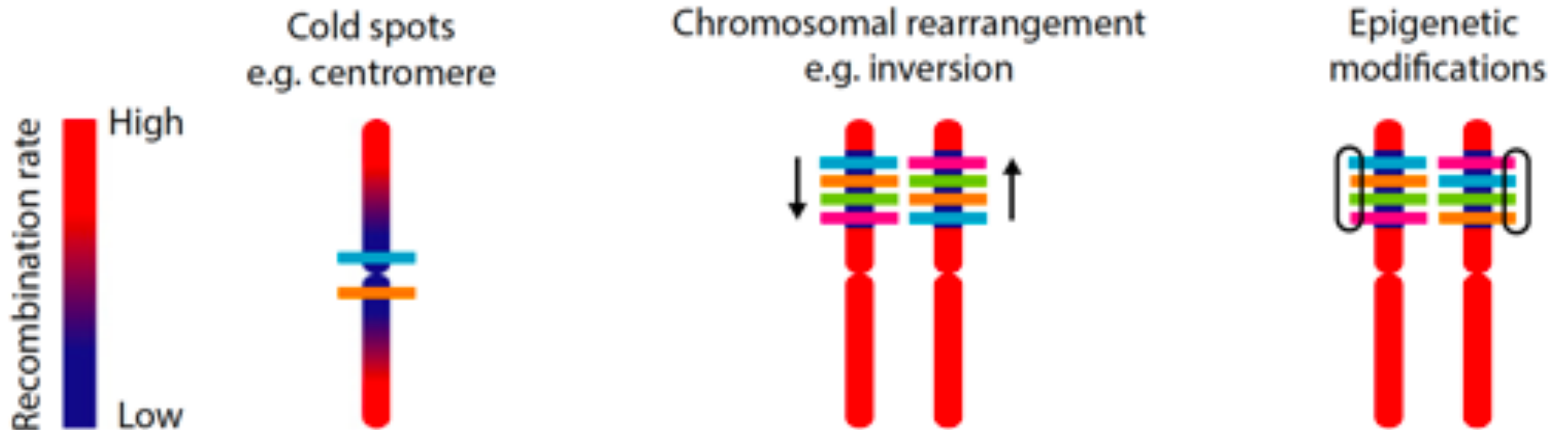
The combination of gene flow and recombination reduces the size of the peak and genome-wide differentiation.

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

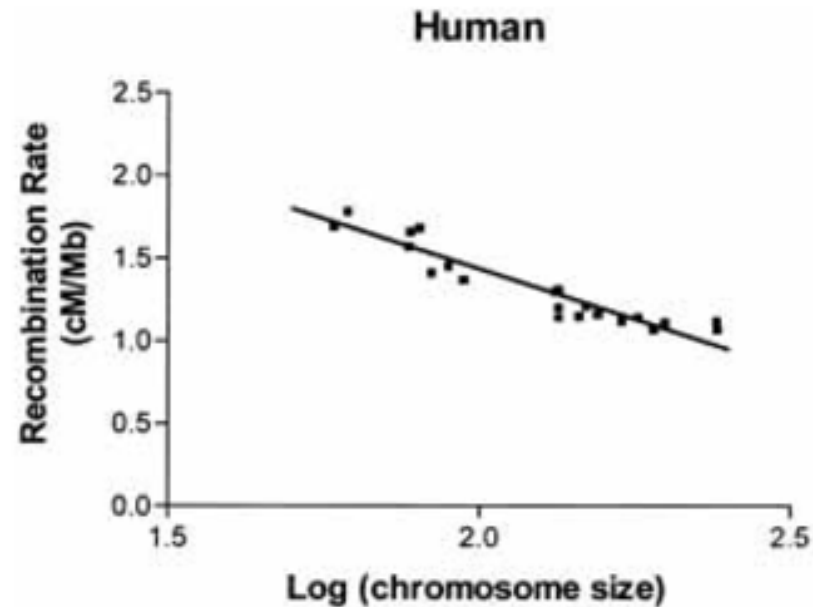




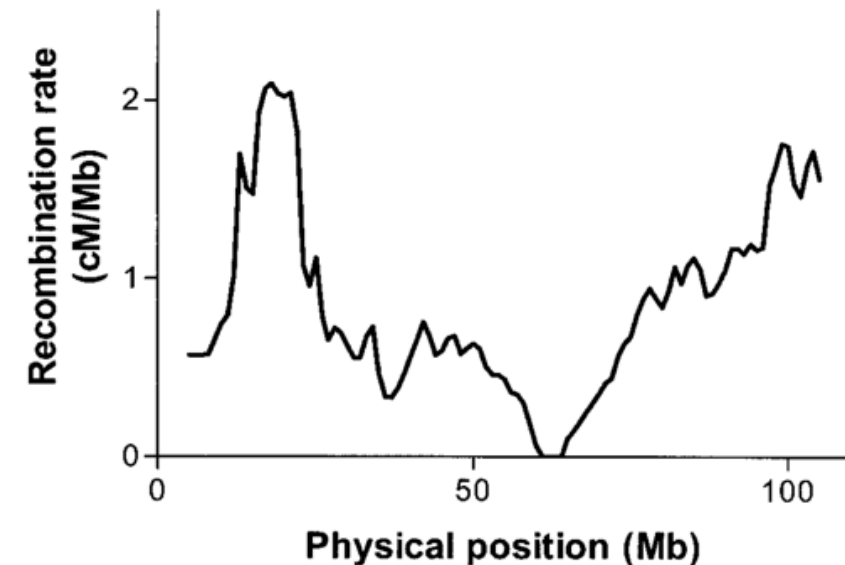
# Cold spots of recombination

Recombination rates are not even across the genome.

Recombination rates can  
vary among chromosomes



...and within chromosomes

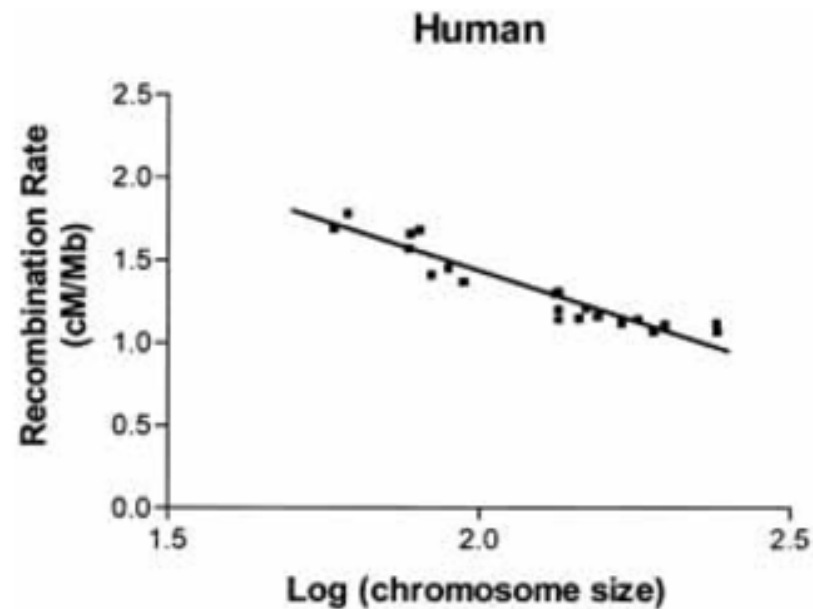




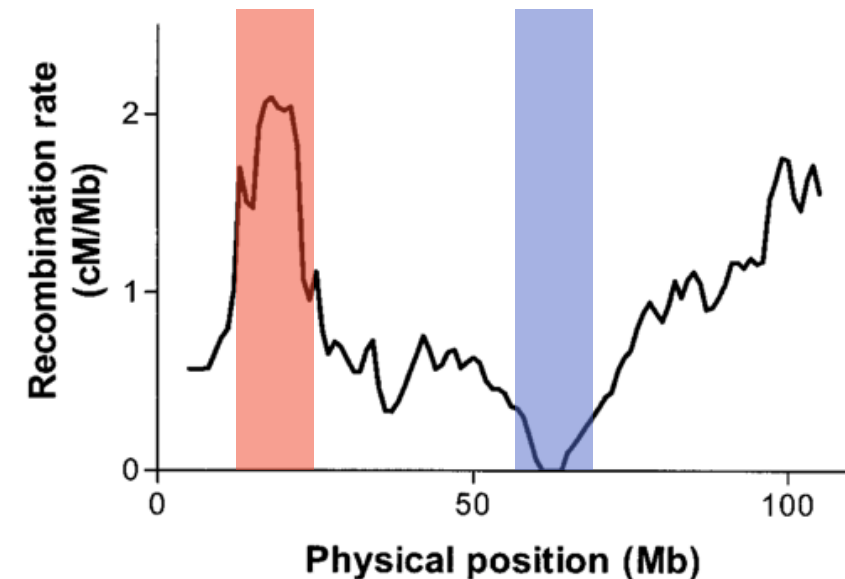
# Cold spots of recombination

Recombination rates are not even across the genome.

Recombination rates can  
vary among chromosomes

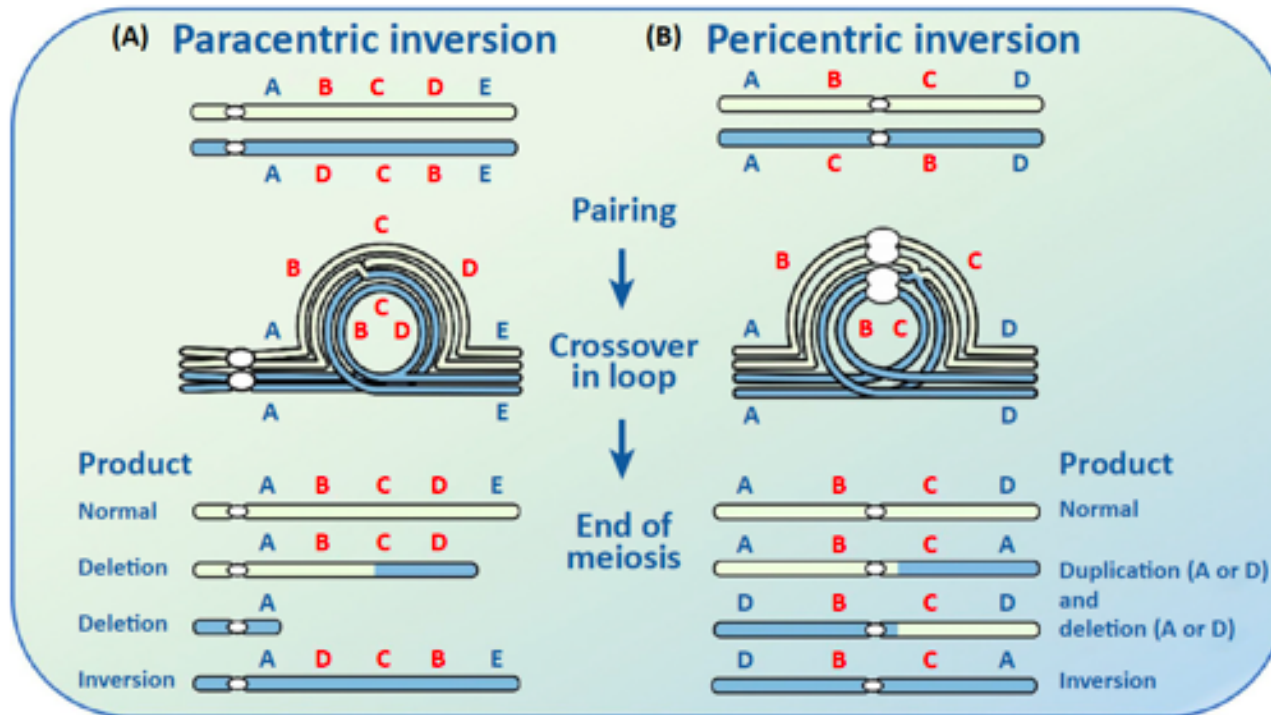
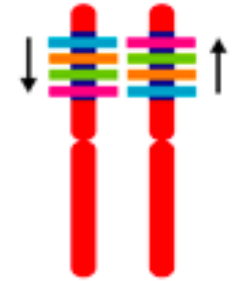


...and within chromosomes



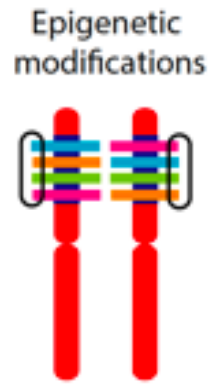
# Chromosomal rearrangements

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.

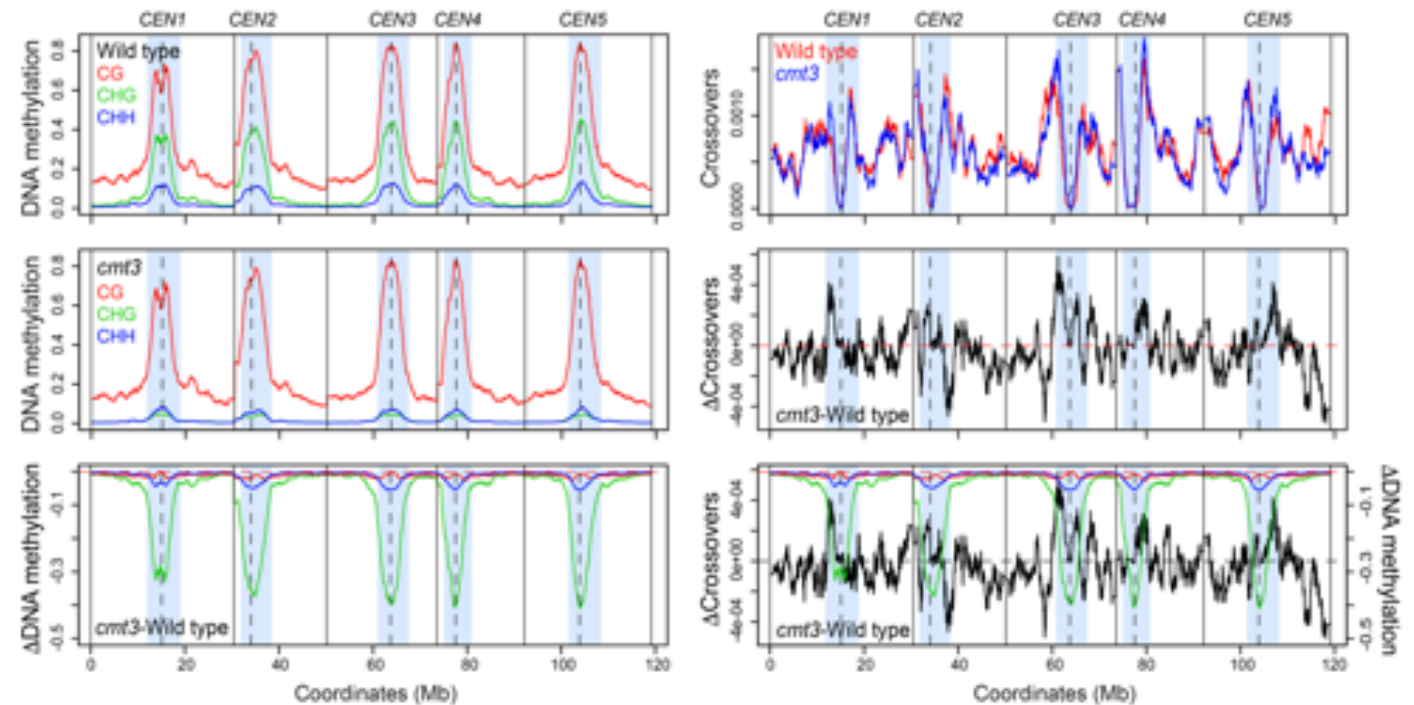
Wellenreuther & Bernatchez 2018, TREE



# 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.

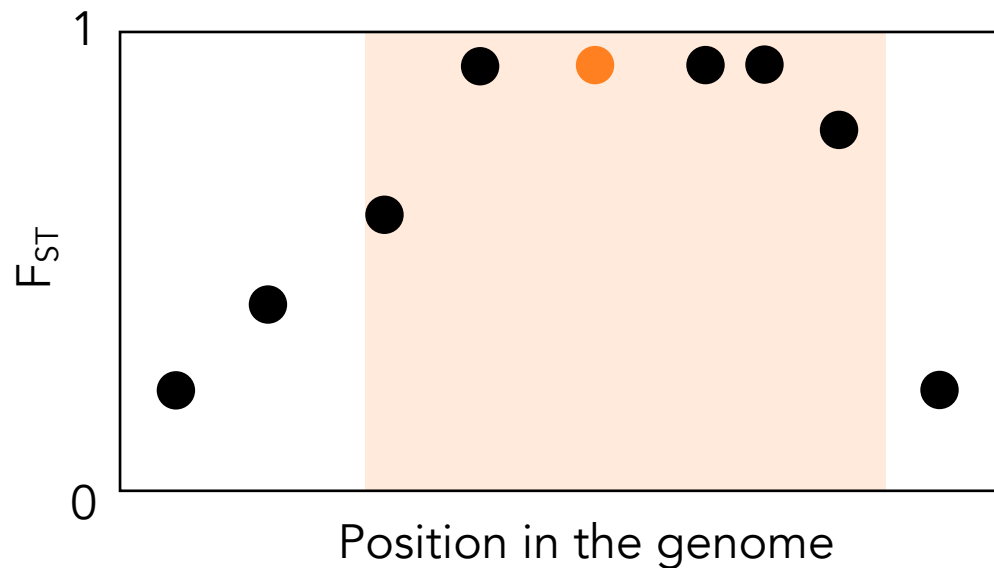


Underwood et al. 2020, Genome Research



# 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.

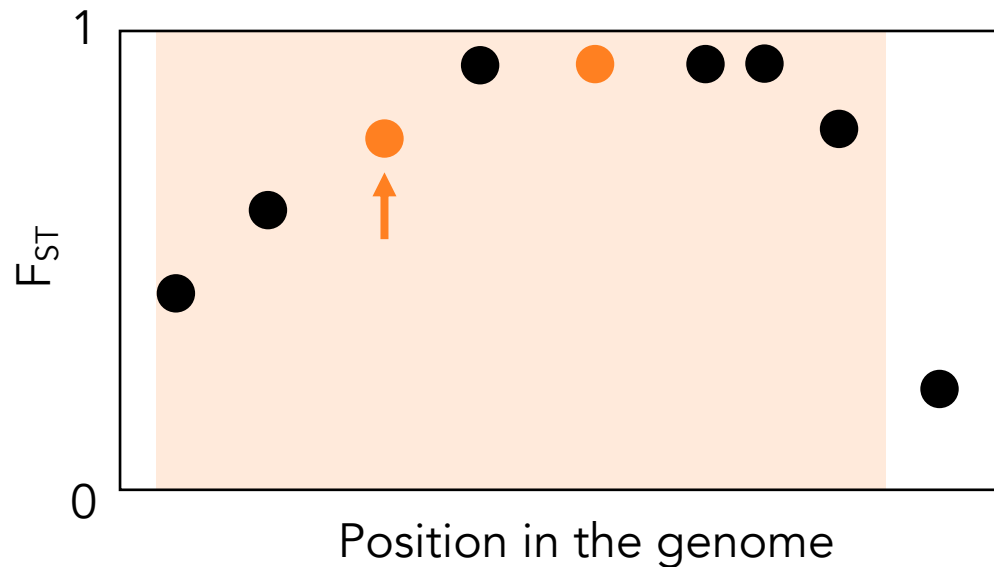


One allele at one site increases in frequency due to selection, the others follow due to linked selection even though they are neutral.



# 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.

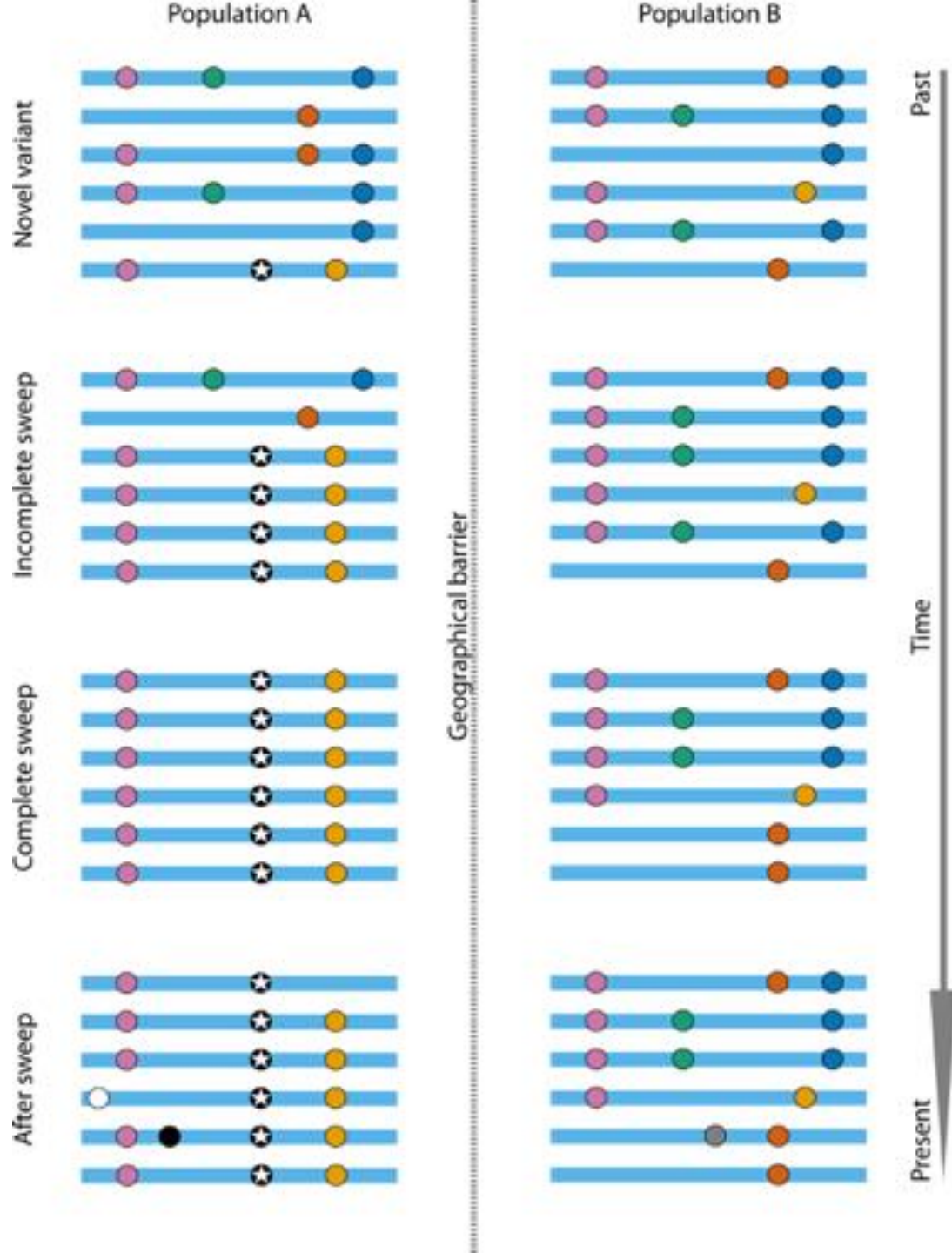


This leads to the evolution of clusters of adaptive loci showing large peaks of differentiation in the genome.

# 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

# Case study

## CORAL GENOMICS

# Population genetics of the coraal *Acropora millepora*: Toward genomic prediction of bleaching

Zachary L. Fuller<sup>1\*</sup>, Veronique J. L. Mocellin<sup>2</sup>, Luke A. Morris<sup>2,3,4</sup>, Neal Cantin<sup>2</sup>, Jihanne Shepherd<sup>1</sup>,  
Luke Sarre<sup>1</sup>, Julie Peng<sup>5</sup>, Yi Liao<sup>6,7</sup>, Joseph Pickrell<sup>8</sup>, Peter Andolfatto<sup>1</sup>, Mikhail Matz<sup>6†</sup>,  
Line K. Bay<sup>2\*†</sup>, Molly Przeworski<sup>1,9,10\*†</sup>



Phenotypic variation in bleaching  
and heat stress among colonies of  
*Acropora millepora*

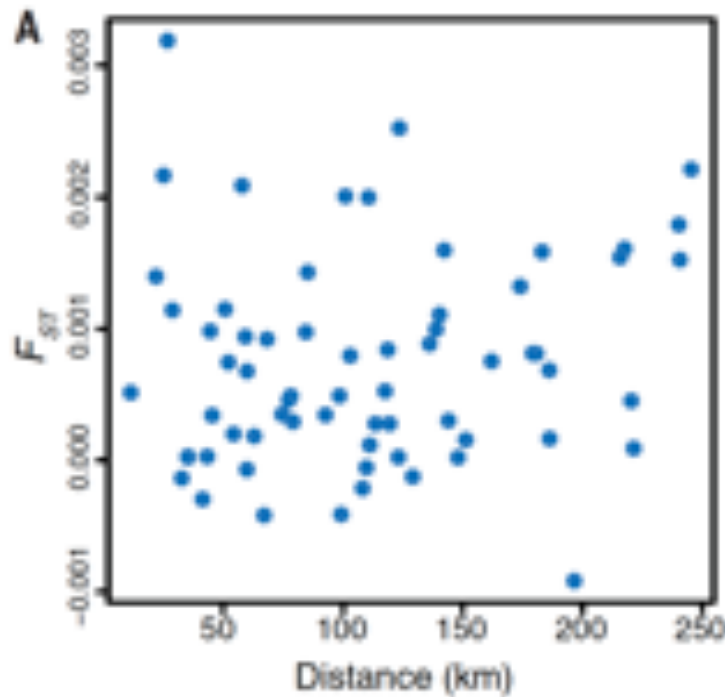
→ What is the genomic basis and  
architecture of resistance to  
bleaching and thermal adaptation?

## Sequencing approach

A mixture of few high coverage genomes ( $n=48$ ) and many more low coverage genome ( $n=237$ ) for population genomics and GWAS

# First thing first: population structure

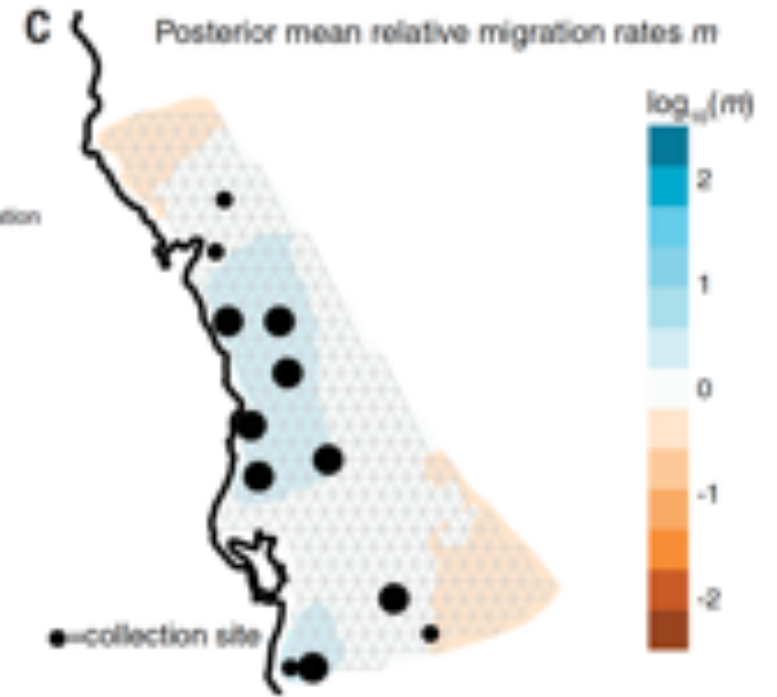
Lack of population structure and high levels of gene flow are promising to isolate loci under selection



No IBD



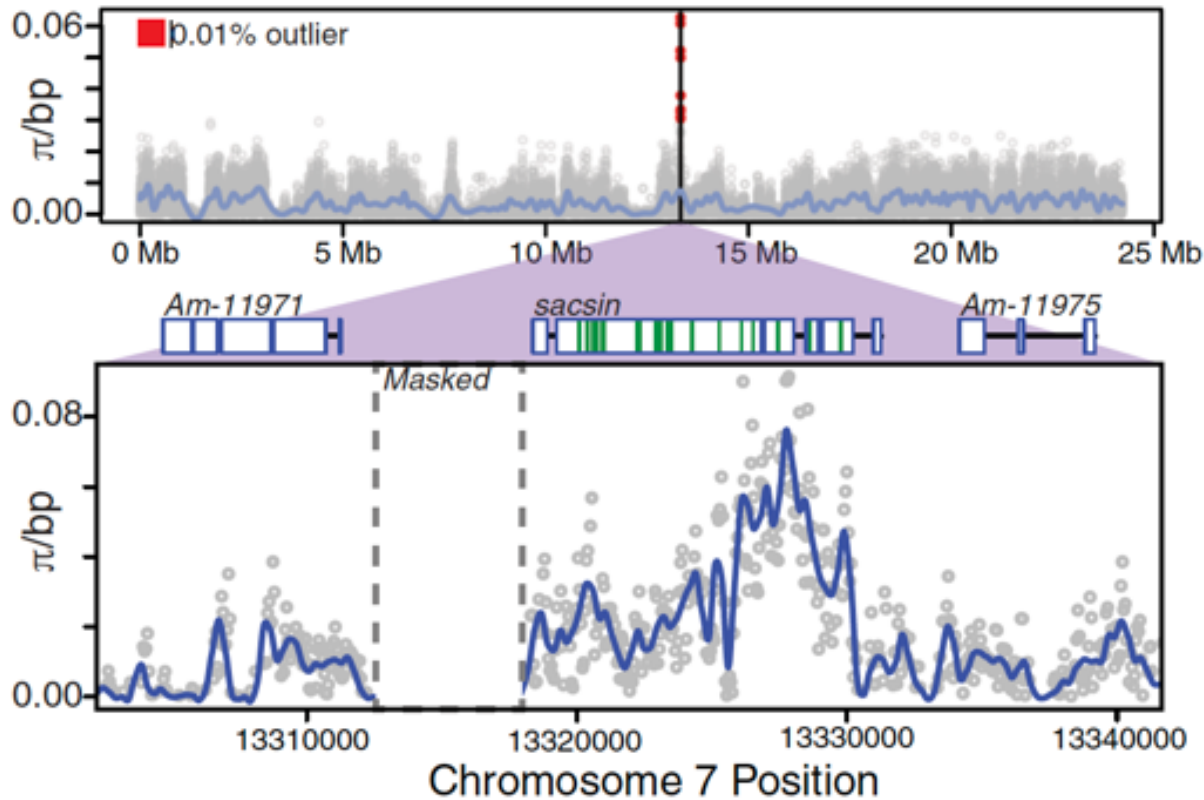
No pop. structure



High gene flow



# Scan for BALANCING selection



Contrary to natural selection, balancing selection maintains different alleles at intermediate frequencies in the population and is promoted by spatially varying selection.

The signature of balancing selection in high nucleotide diversity  $\pi$ .

# Functional annotation of candidate gene

$\pi$  peak is closest to the gene *sacsin*, associated with response to heat stress.

→ however, the source of balancing selection is not clear

# GWAS (n=213)

Linear mixed model to test for additive effects of SNPs associated with

- environmental variables,
- population structure,
- batch effects collection depth,
- proportion of symbiont associated with thermal tolerance.

# GWAS (n=213)

Linear mixed model to test for additive effects of SNPs associated with

- environmental variables,
- population structure,
- batch effects collection depth,
- proportion of symbiont associated with thermal tolerance.

→ No SNPs below significance threshold suggest polygenic basis of adaptation and low power of study design.

# Development of polygenic score (PGS)

Based on LD and effect sizes estimated by GWAS of top SNPs

- Model explains 62% of the variation in bleaching.
- Genetic basis of resistance to bleaching and thermal stress is highly polygenic

The end!