

# Signatures of selection

Day 2 - Lecture 2

Mafalda Ferreira, Ph.D.

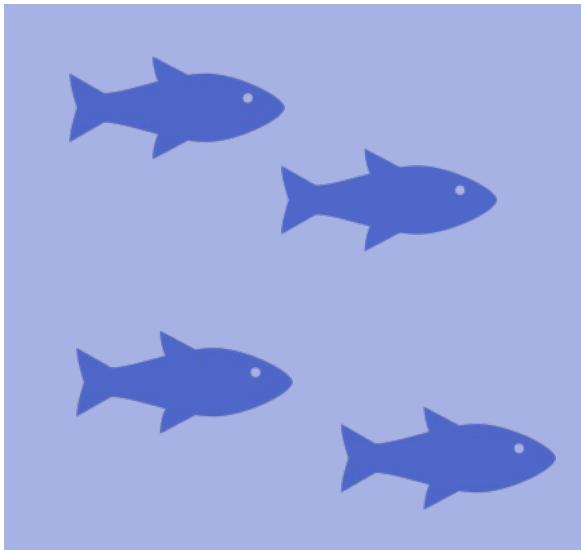
(adapted from Anna Tigano)

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

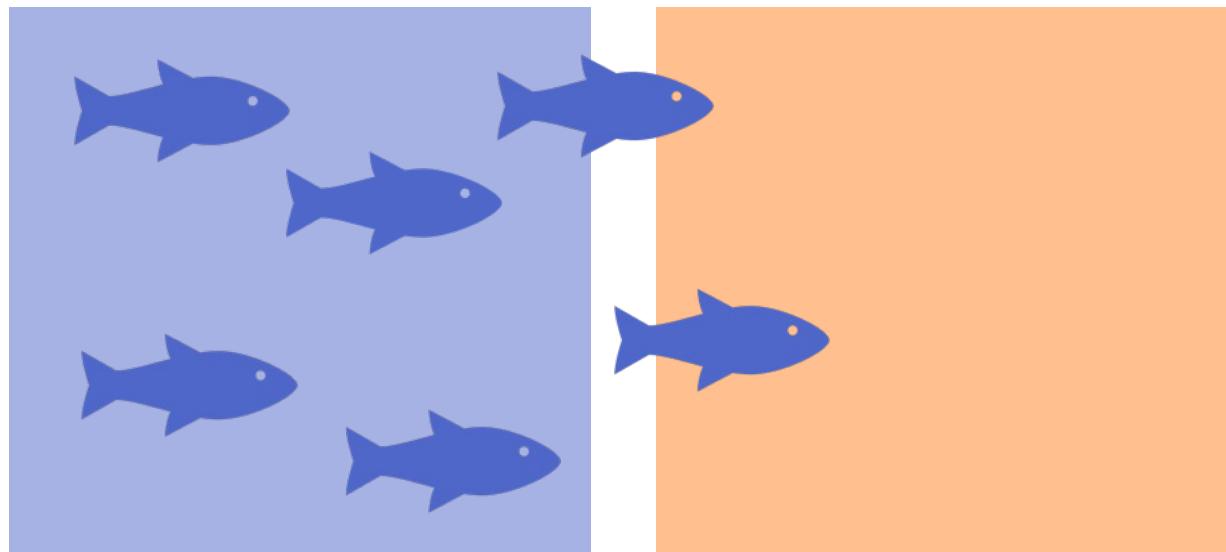
If we identify the loci underpinning the genetic basis of an adaptive trait we can better understand

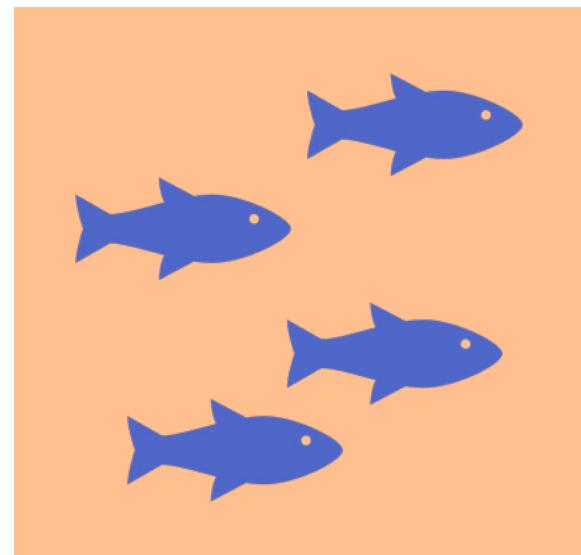
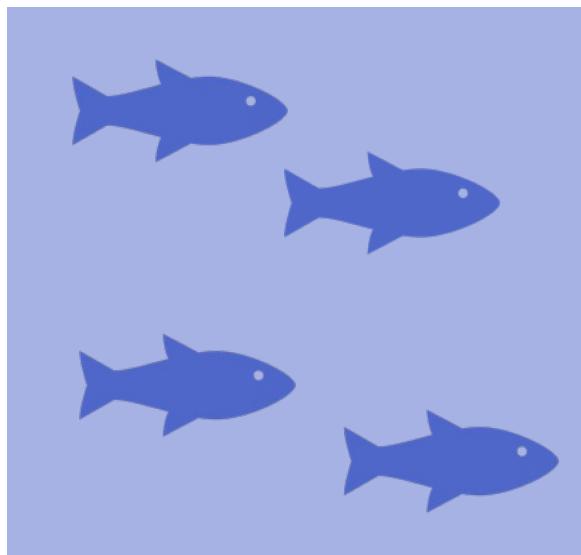
- the adaptive trait itself (e.g. its molecular and physiological basis)
- the process of adaptation (including the interplay with other evolutionary forces)

A population of cold-adapted fish

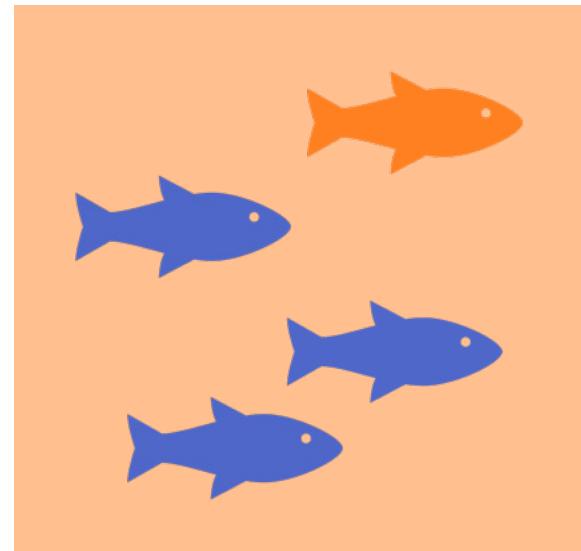
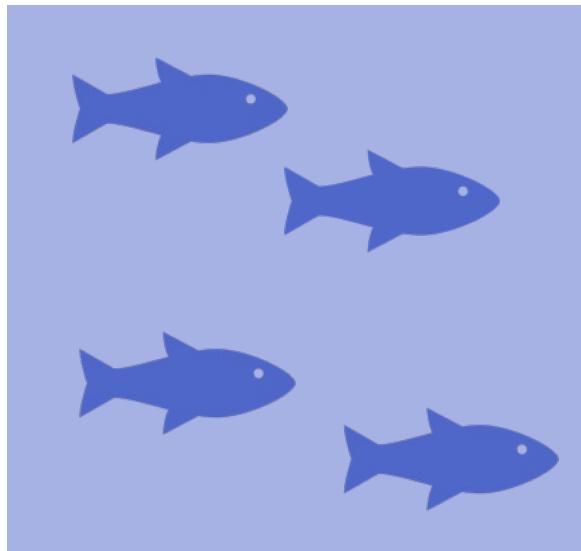


Cold-adapted fish migrate to warmer waters

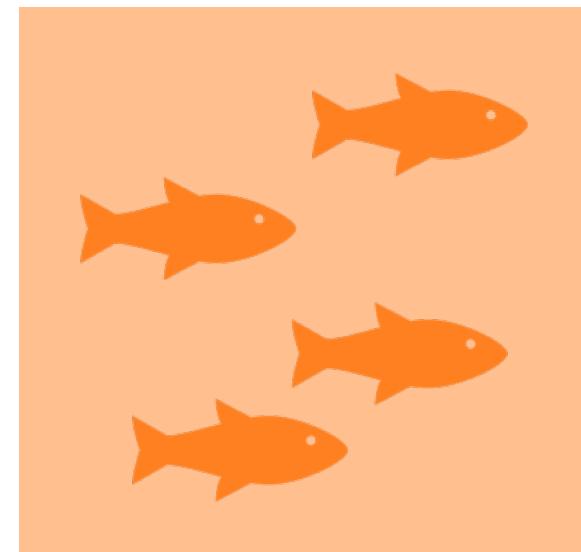
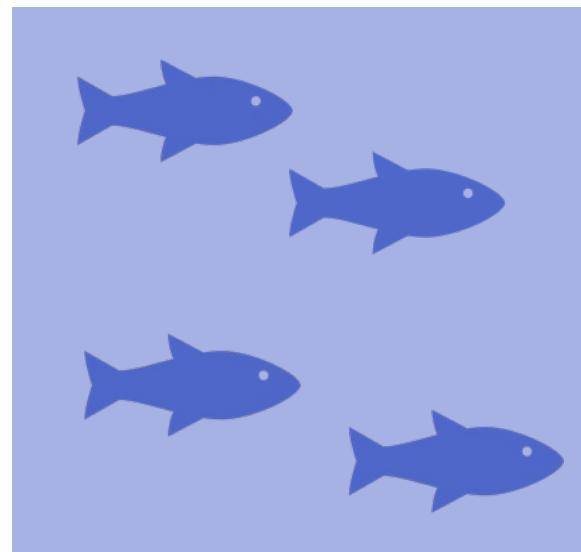




A new mutation confers a fitness benefit in warmer waters

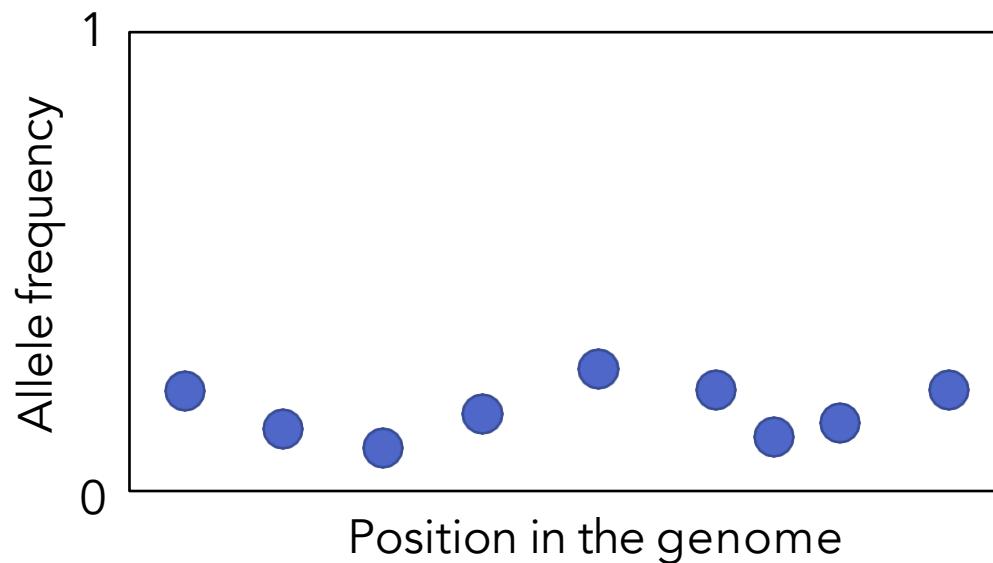


The new mutation spreads quickly in the population and reaches fixation



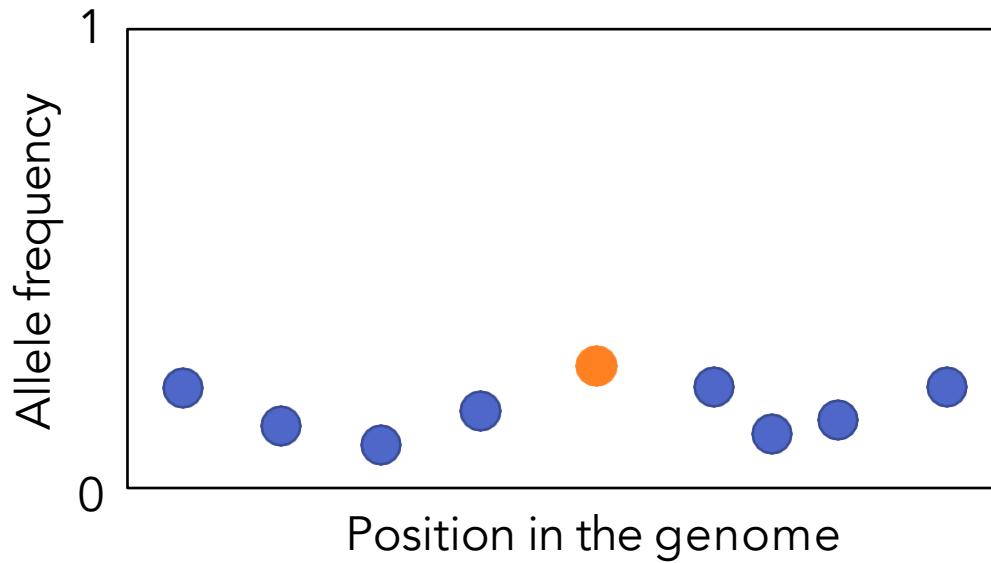
# Changes in allele frequencies

Cold-adapted fish migrate to warmer waters



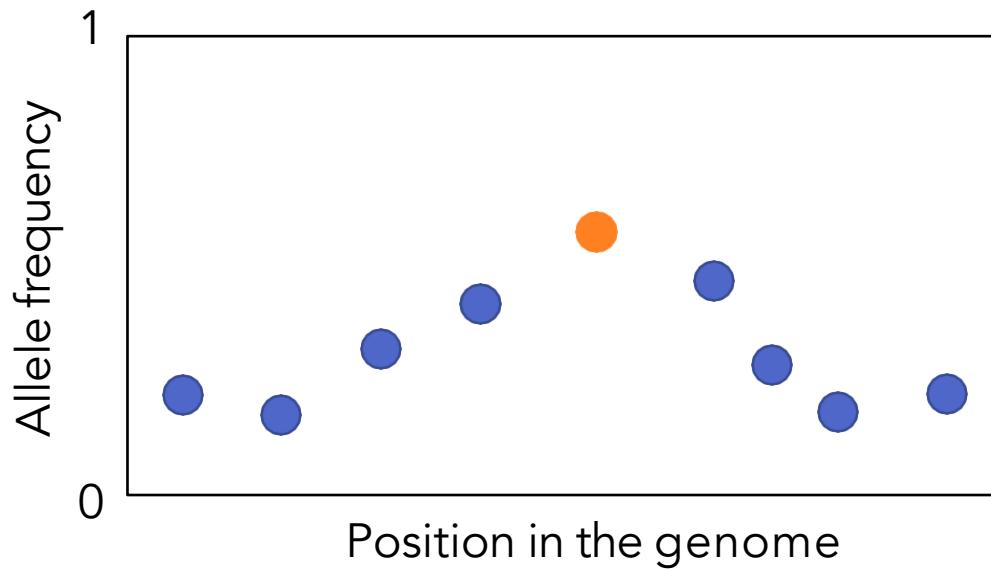
# Changes in allele frequencies

A new mutation confers a fitness benefit in warmer waters



# Changes in allele frequencies

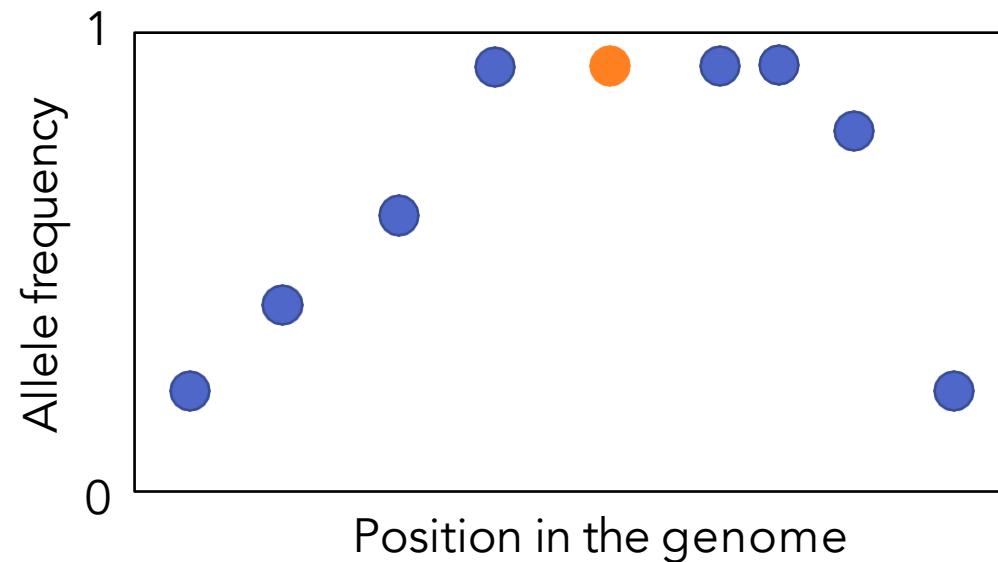
The mutation increases quickly in frequency, together with the variants in its proximity



# Changes in allele frequencies

The new mutation spreads quickly in the population and reaches fixation

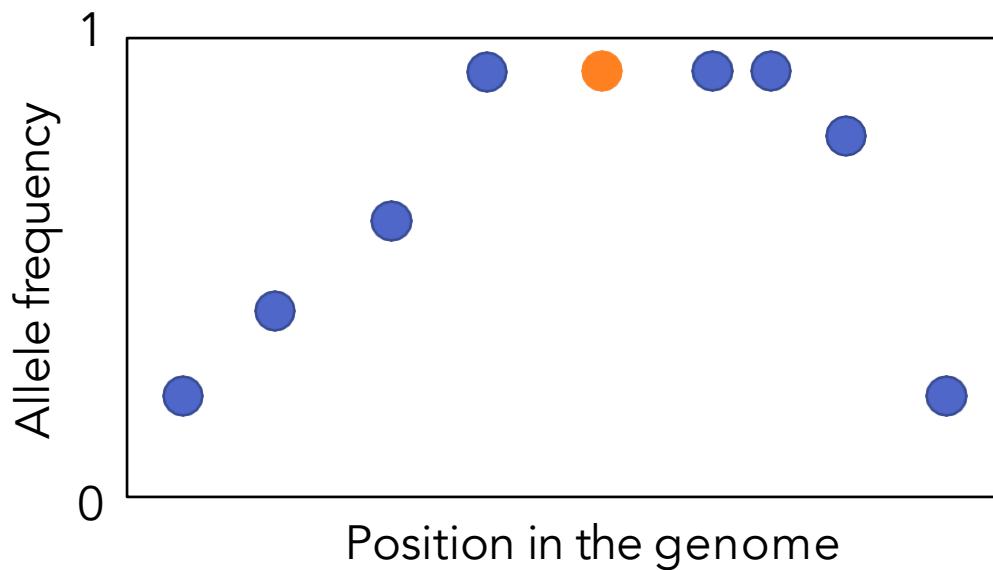
Selective sweep!



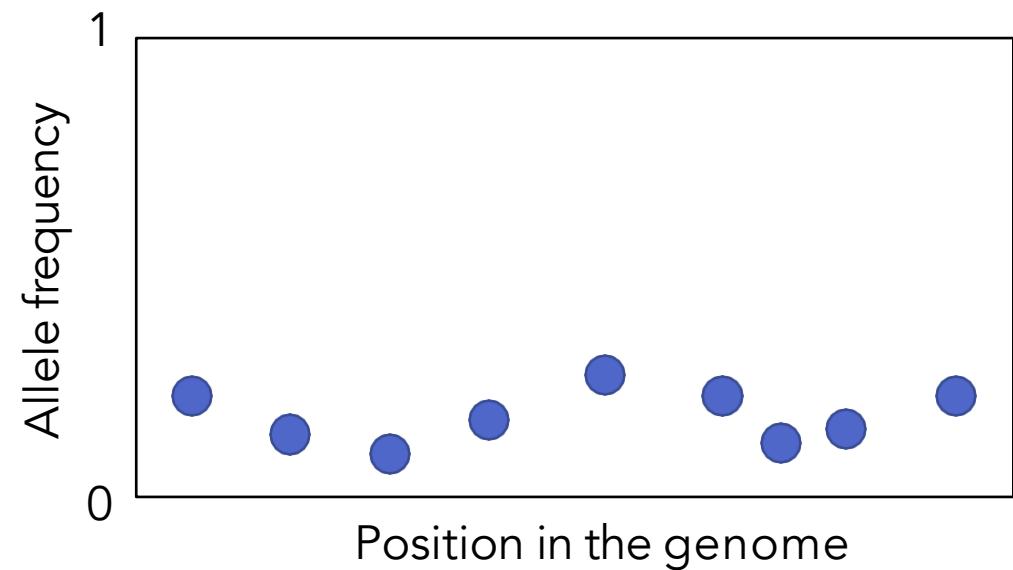
# Changes in allele frequencies



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

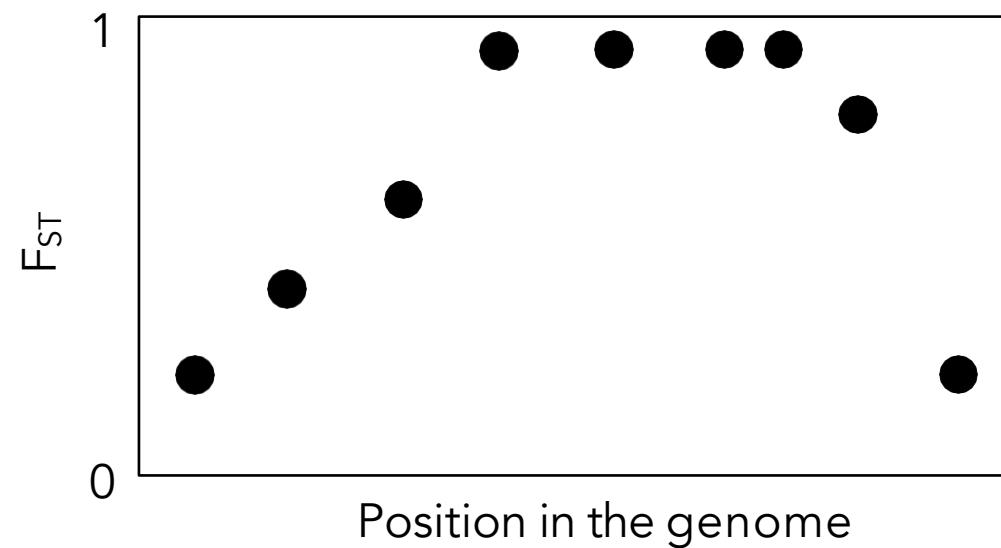
$\pi_S$  = expected heterozygosity in each population

Review on Fst calculation and interpretation:

Holsinger & Weir 2009, Nature Reviews Genetics

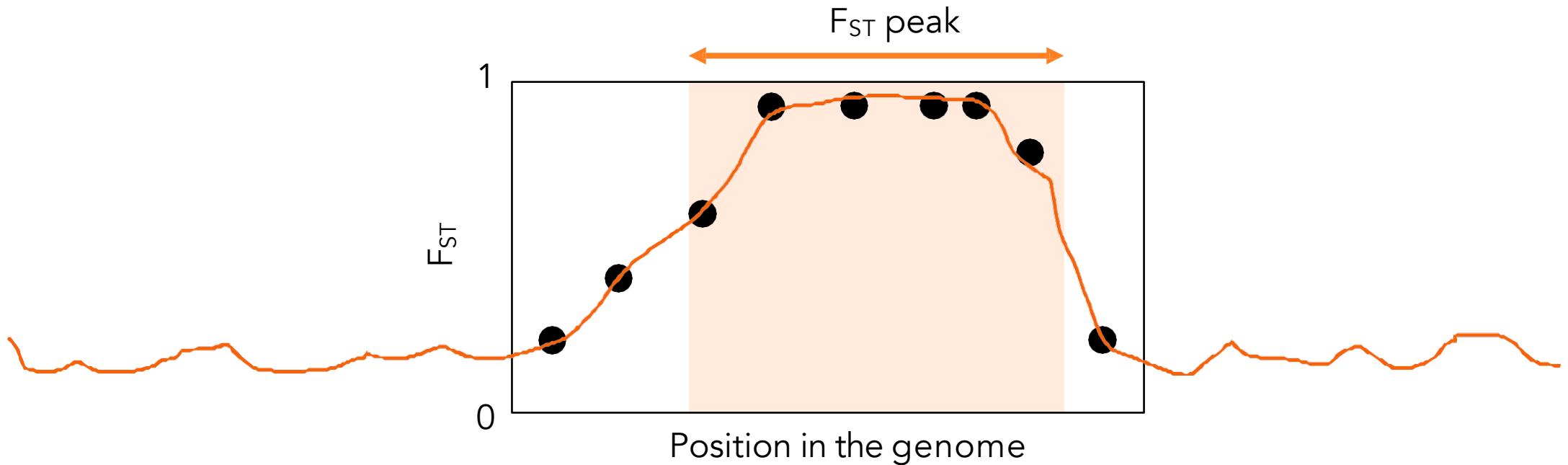
# Genetic differentiation

Differentiation between cold- and warm-adapted populations



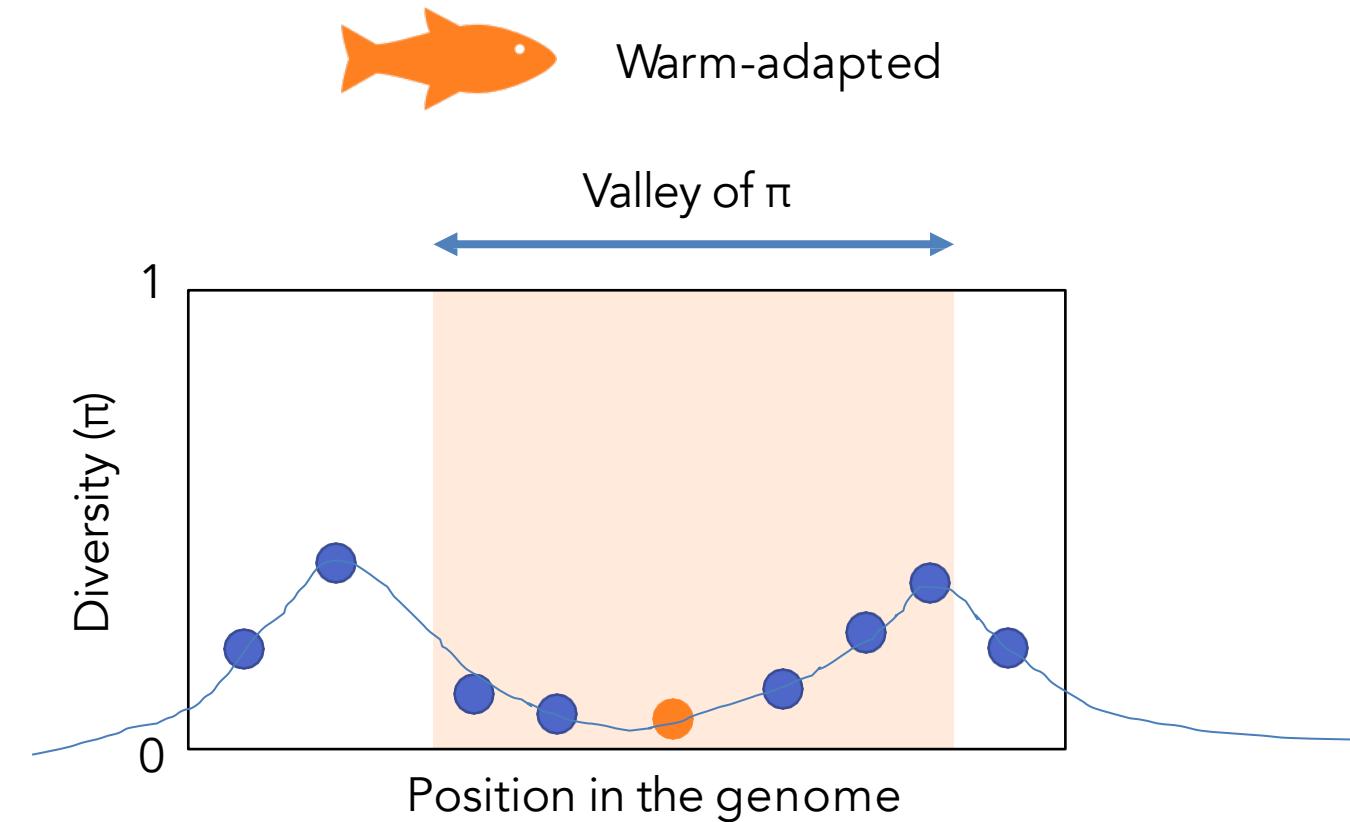
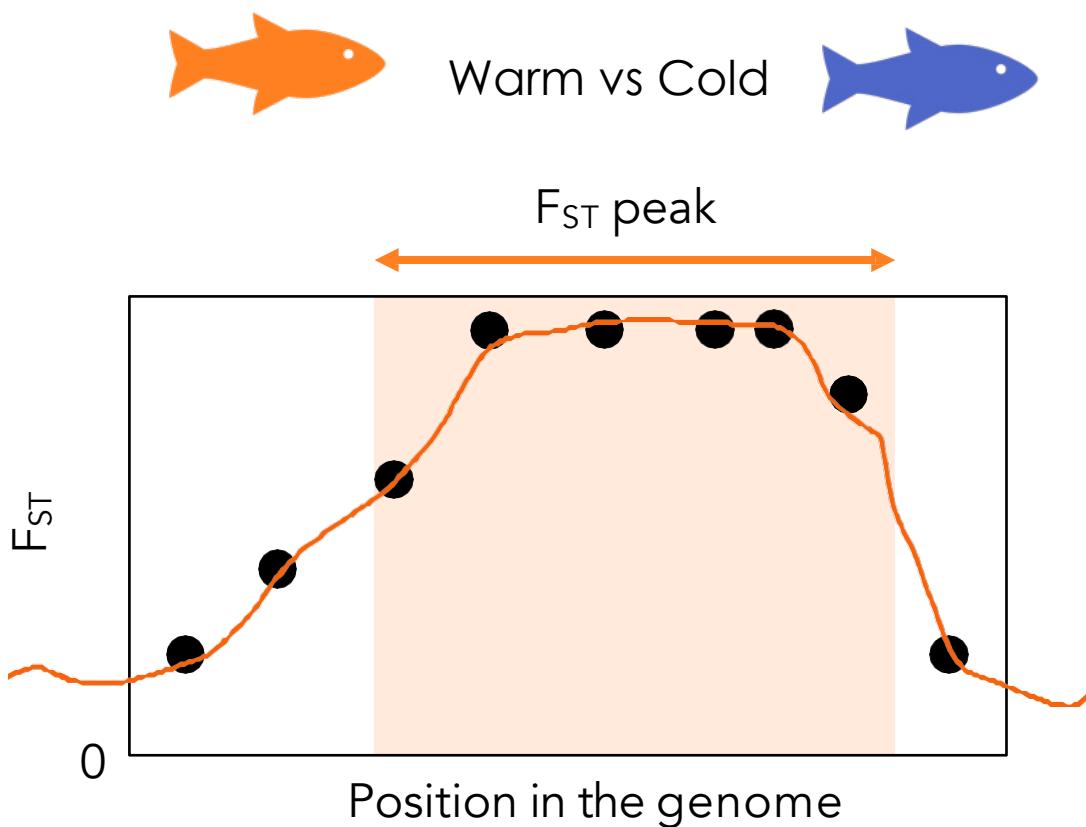
# Genetic differentiation

Differentiation between cold- and warm-adapted populations

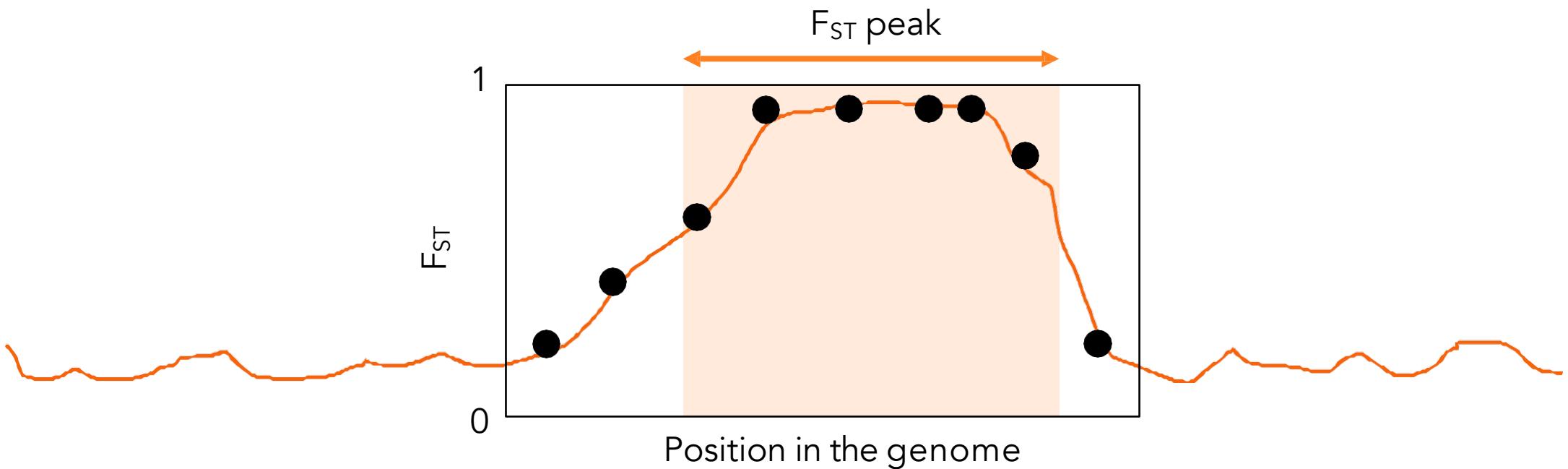


# Reduction in diversity

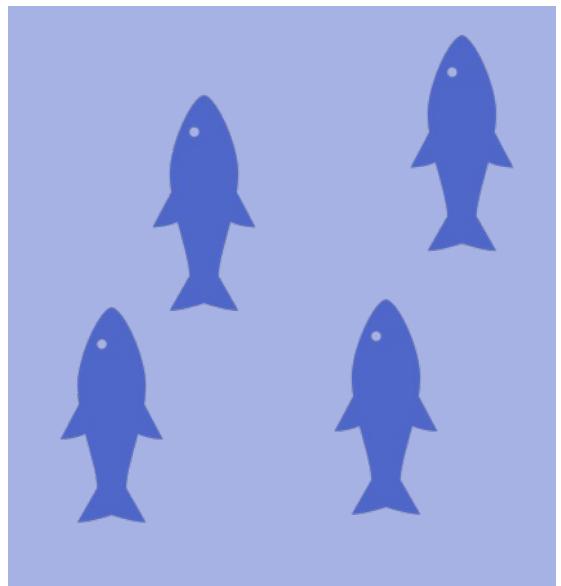
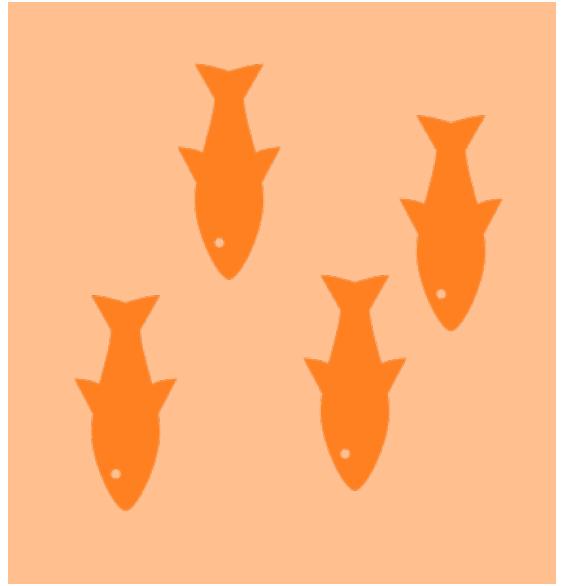
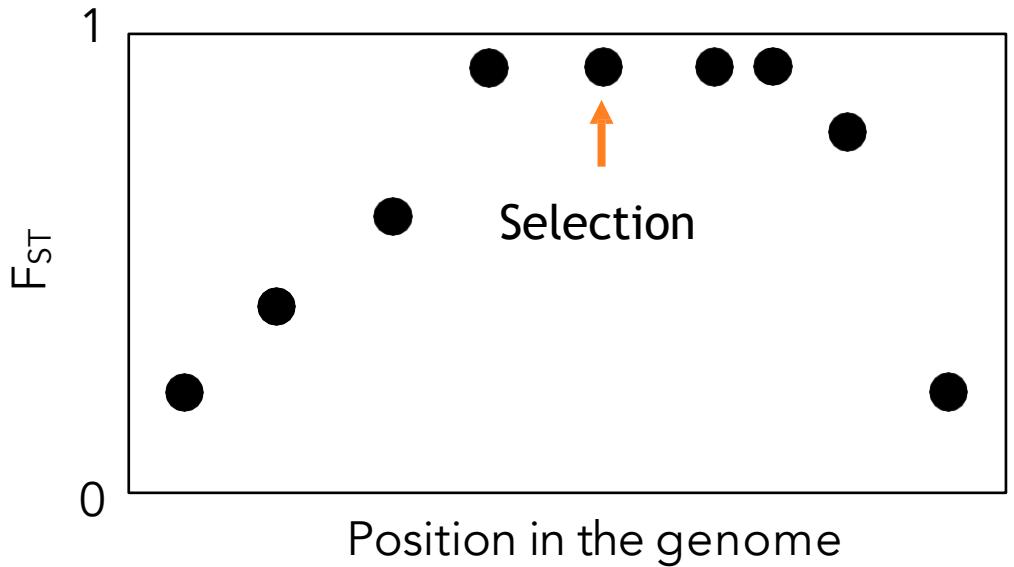
Once the beneficial allele is **fixed** (allele frequency = 1), variation at the target of selection and surrounding sites is eliminated from the population.



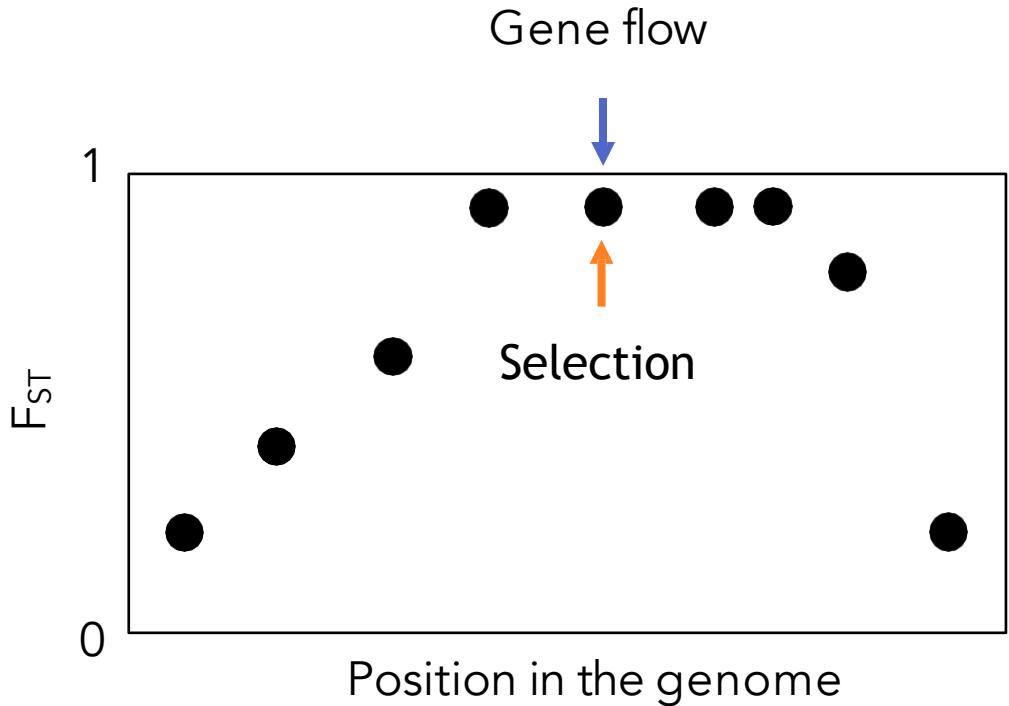
# What determines the size of a $F_{ST}$ peak?



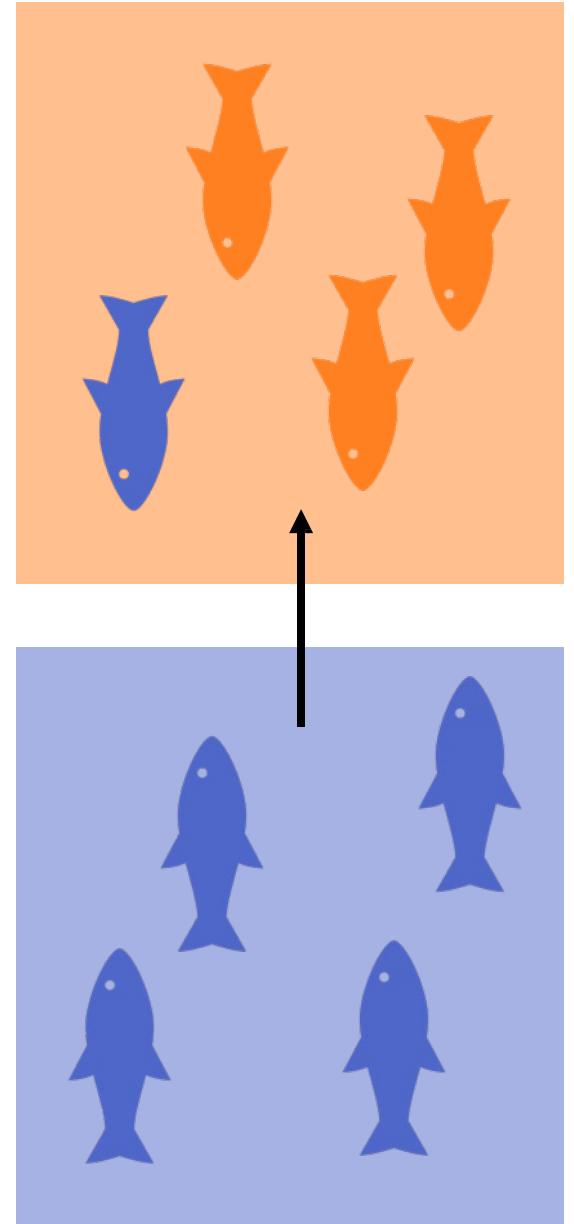
# Selection vs. gene flow



# Selection vs. gene flow

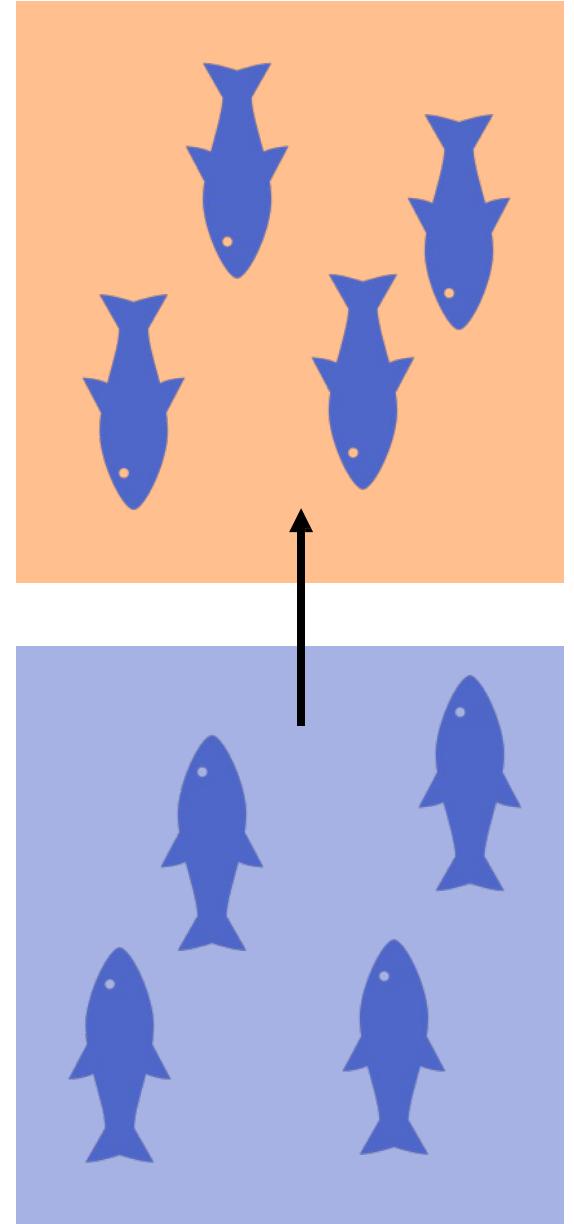


If gene flow introduces maladaptive alleles, the strength of selection  $s$  will determine whether the locally adaptive allele will not reach fixation.



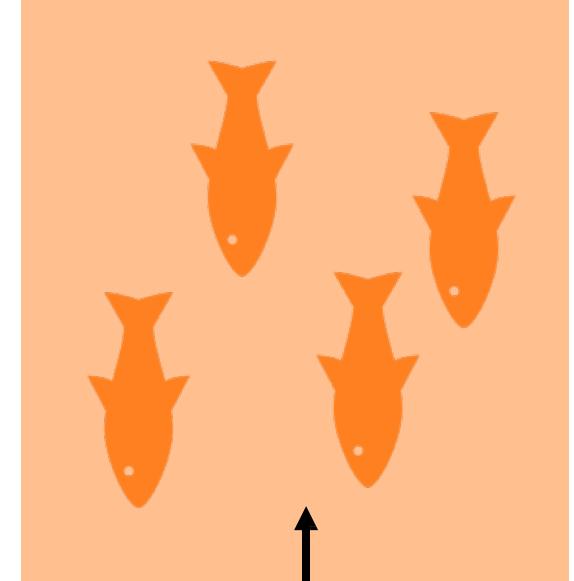
# Selection vs. gene flow

If selection is weaker than gene flow, the locally adapted alleles will have lower allele frequency and the local adaptation may be lost.

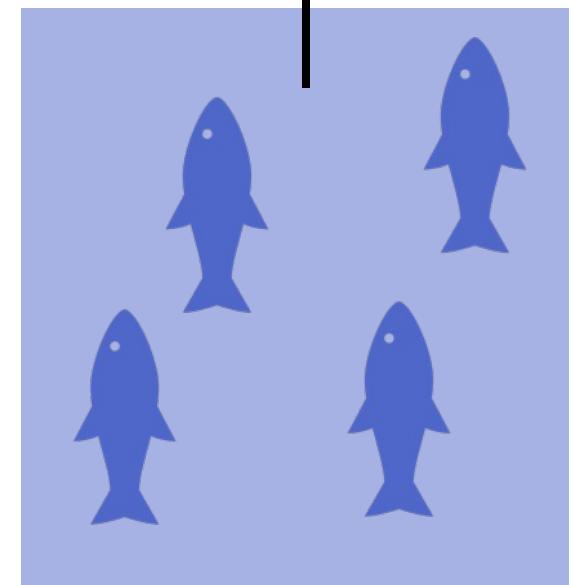


# Selection vs. gene flow

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



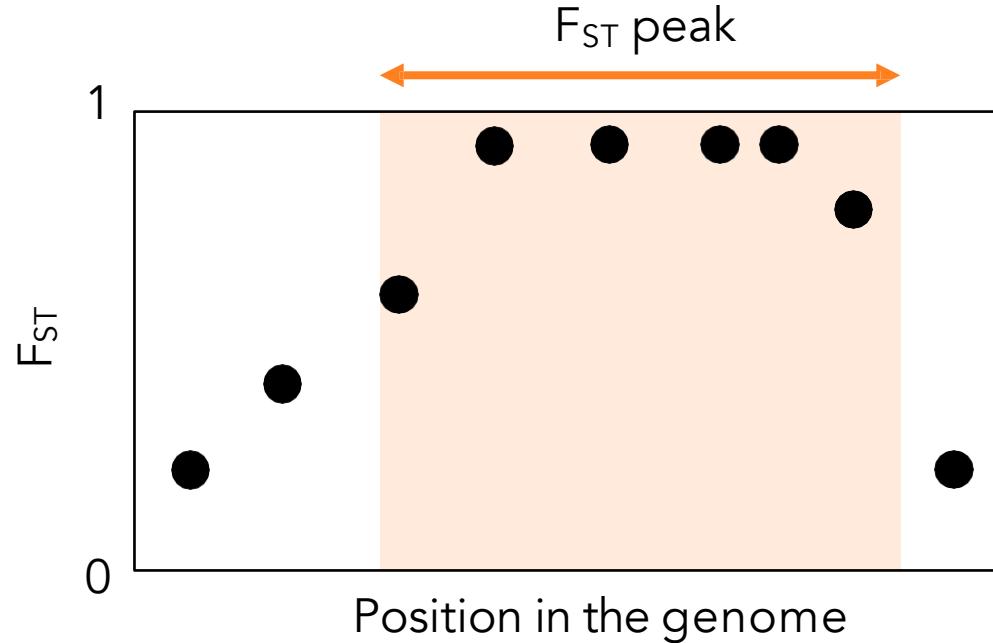
The locally adaptive allele will be maintained but overall population differentiation will be homogenized by gene flow through recombination.



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

# Physical linkage and recombination

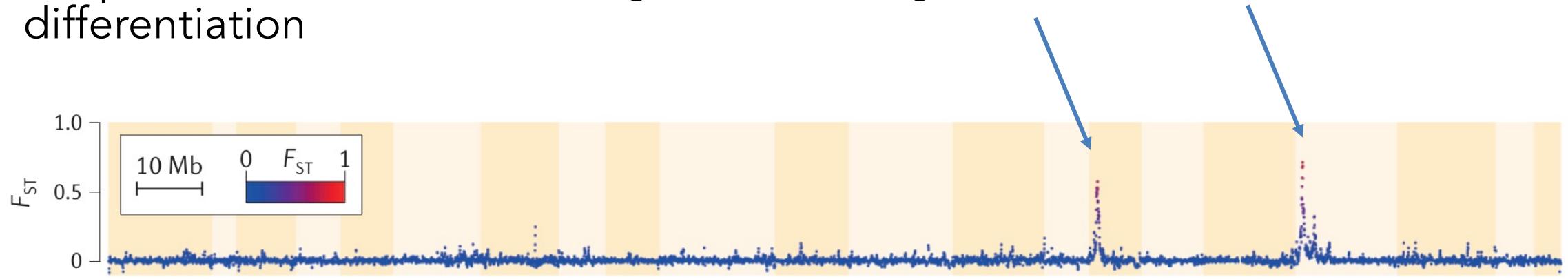
Recombination breaks linkage among sites at a rate that rapidly decays with distance.



# 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.

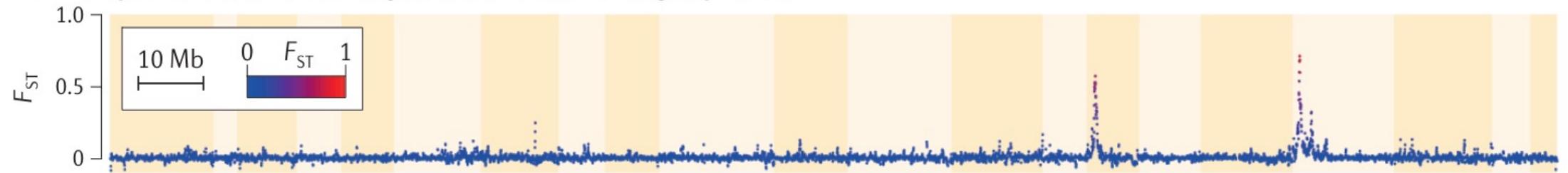
Adaptive loci stand out from a genomic background of low differentiation



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

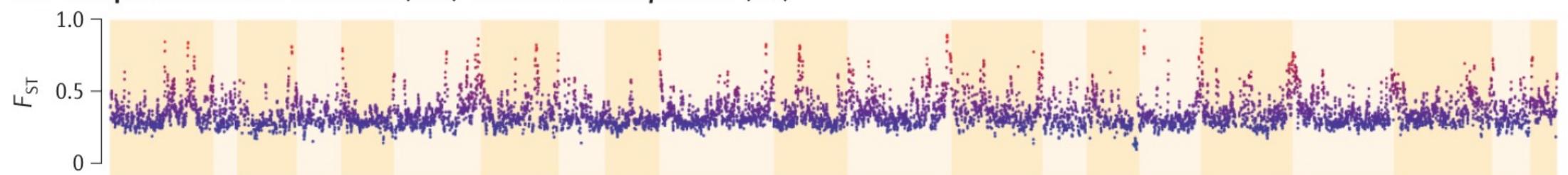
Parapatric = "near" distribution

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



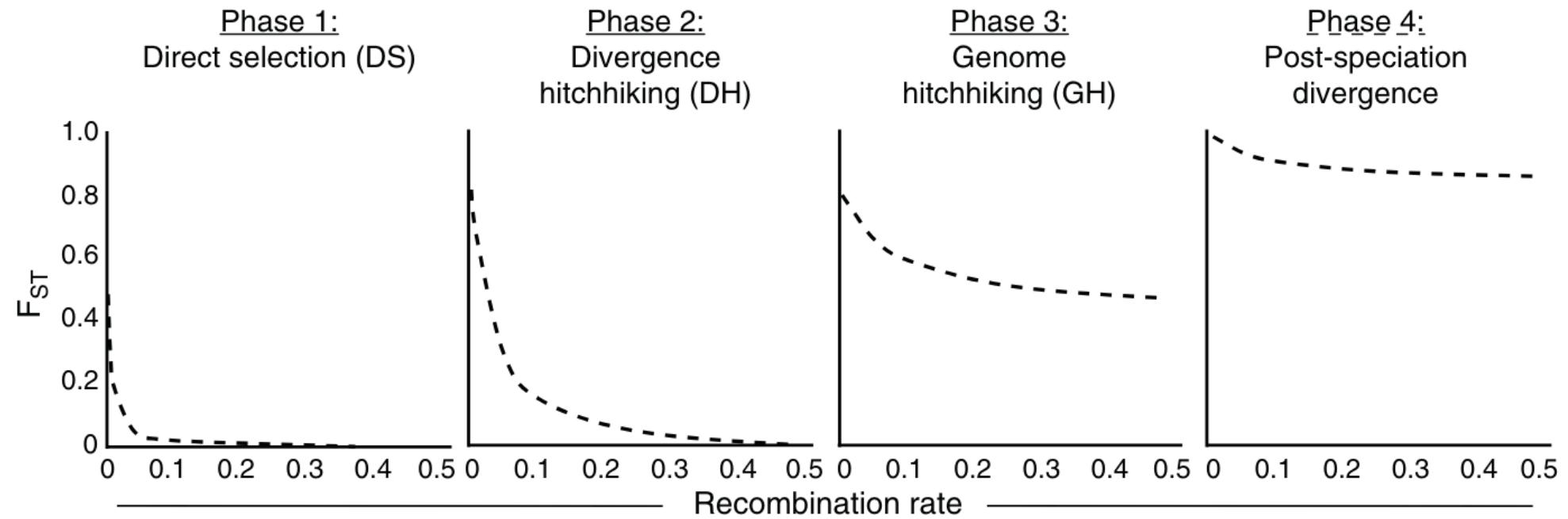
Allopatric = non-overlapping

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



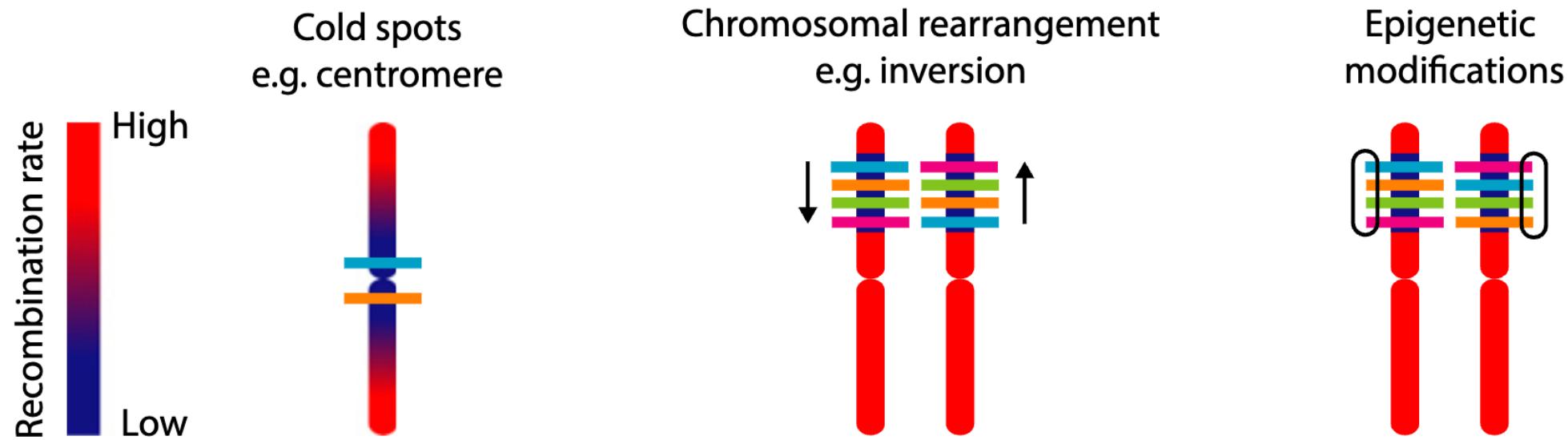
The combination of gene flow and recombination reduces the size of the peak and genome-wide differentiation.

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



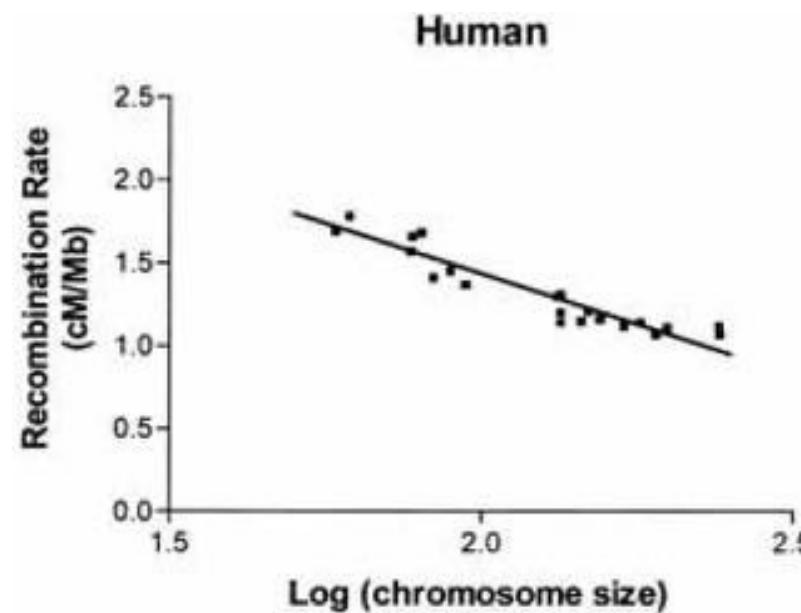
Cold spots  
e.g. centromere



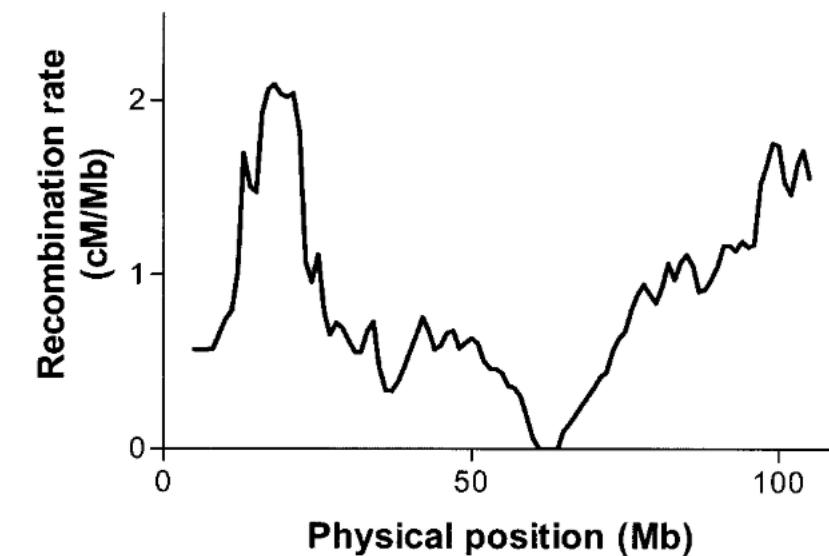
# Cold spots of recombination

Recombination rates are not even across the genome.

Recombination rates can  
vary among chromosomes



...and within  
chromosomes



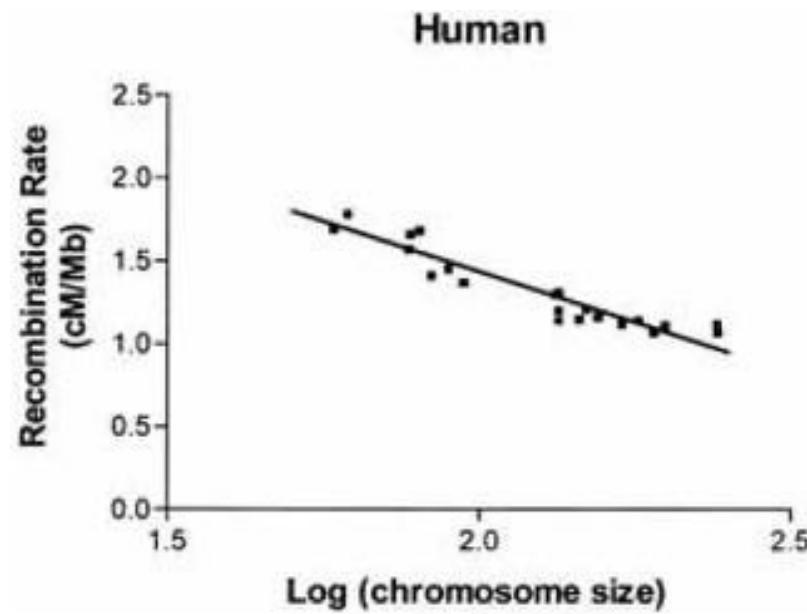
Cold spots  
e.g. centromere



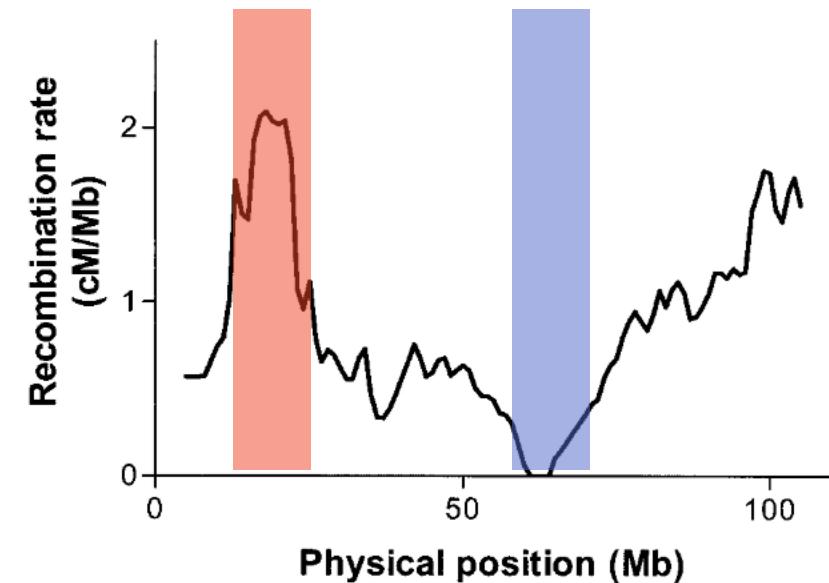
# Cold spots of recombination

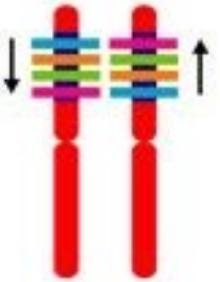
Recombination rates are not even across the genome.

Recombination rates can vary among chromosomes



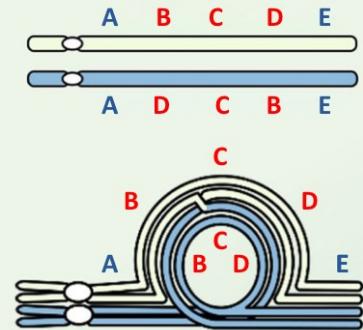
...and within chromosomes



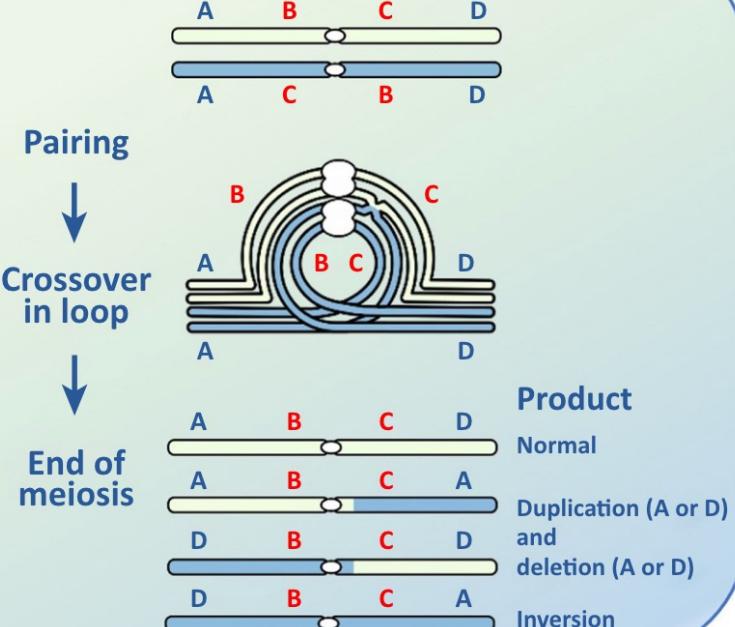


# Chromosomal rearrangements

(A) Paracentric inversion



(B) Pericentric inversion



Structural variants, like inversions, impair meiosis in heterozygotes suppressing recombination.

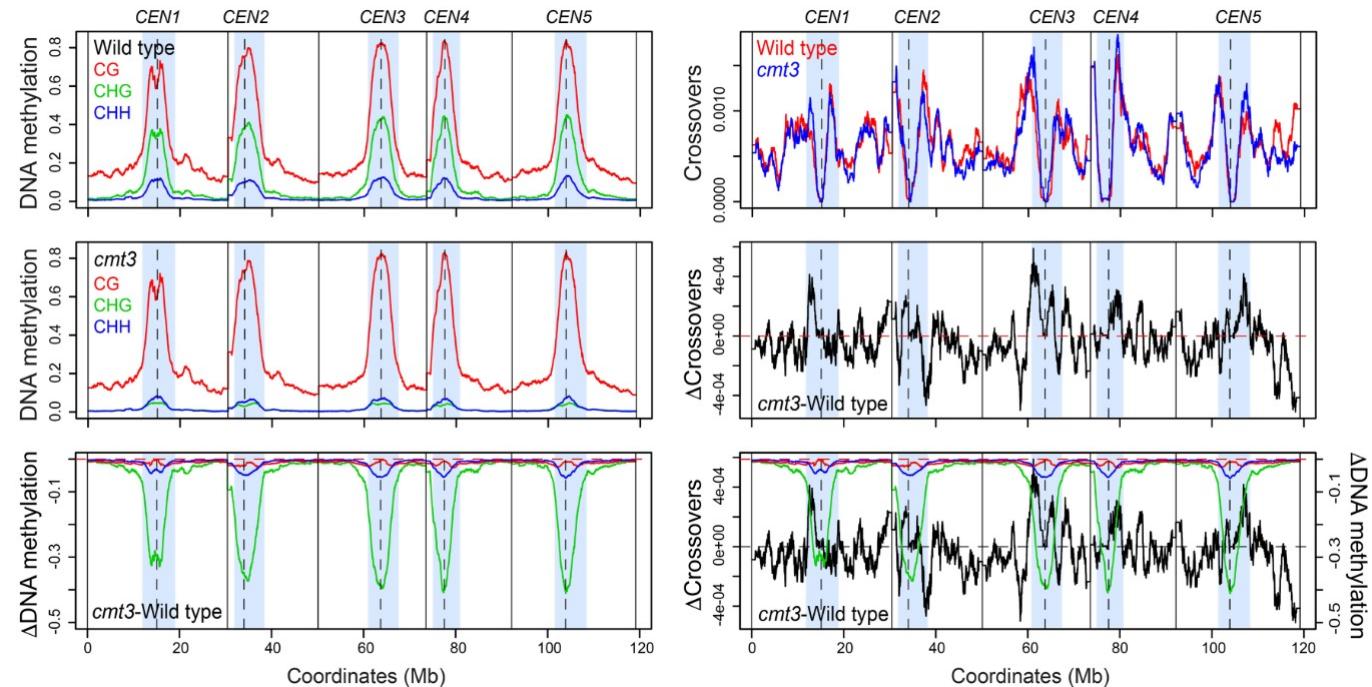
Wellenreuther & Bernatchez  
2018 TREE



# Epigenetic modifications

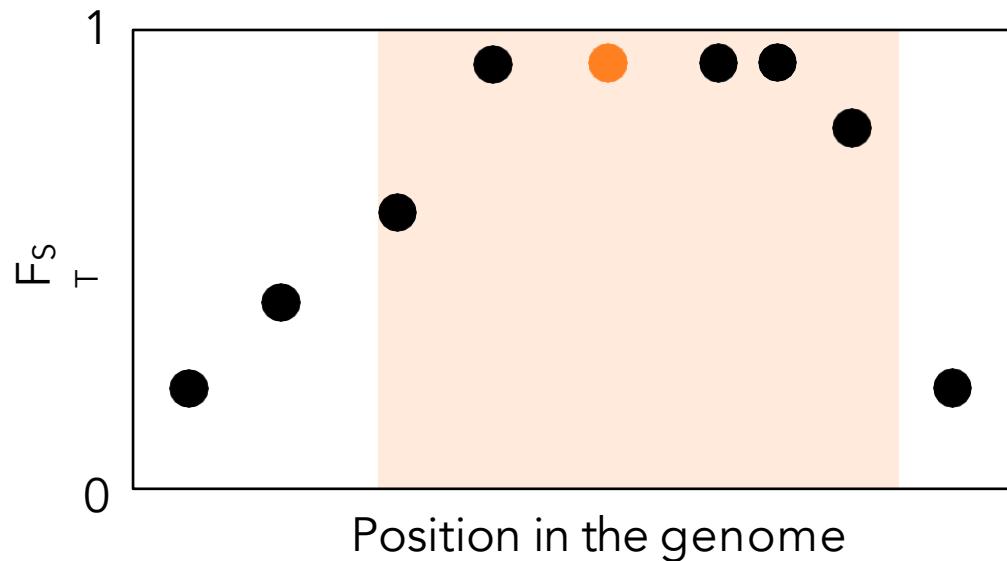
Methylation can suppress recombination locally, for example around centromeres.

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



# Genetic architecture of a trait

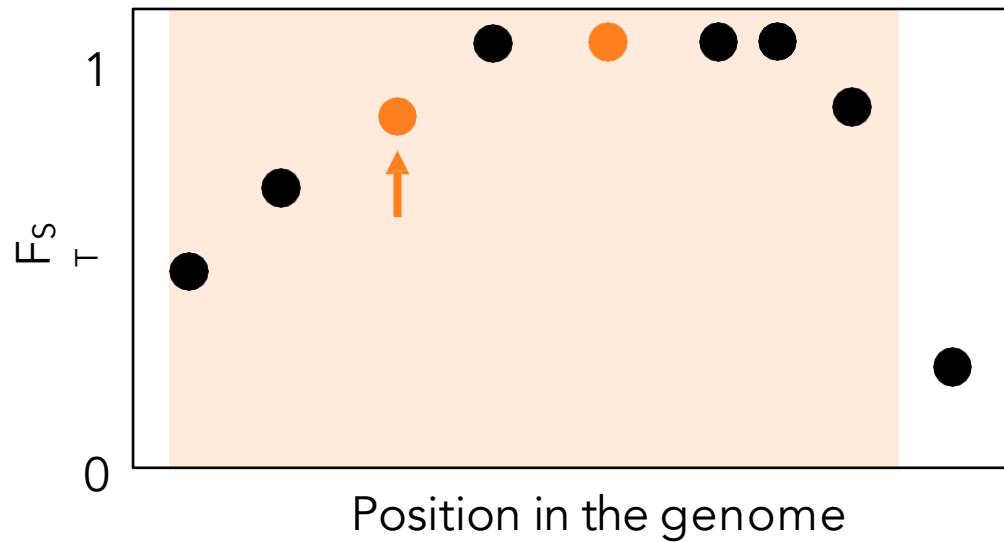
So far: **single locus** scenario, where one locus of large effect underlies the genetic basis of an adaptive trait.



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.



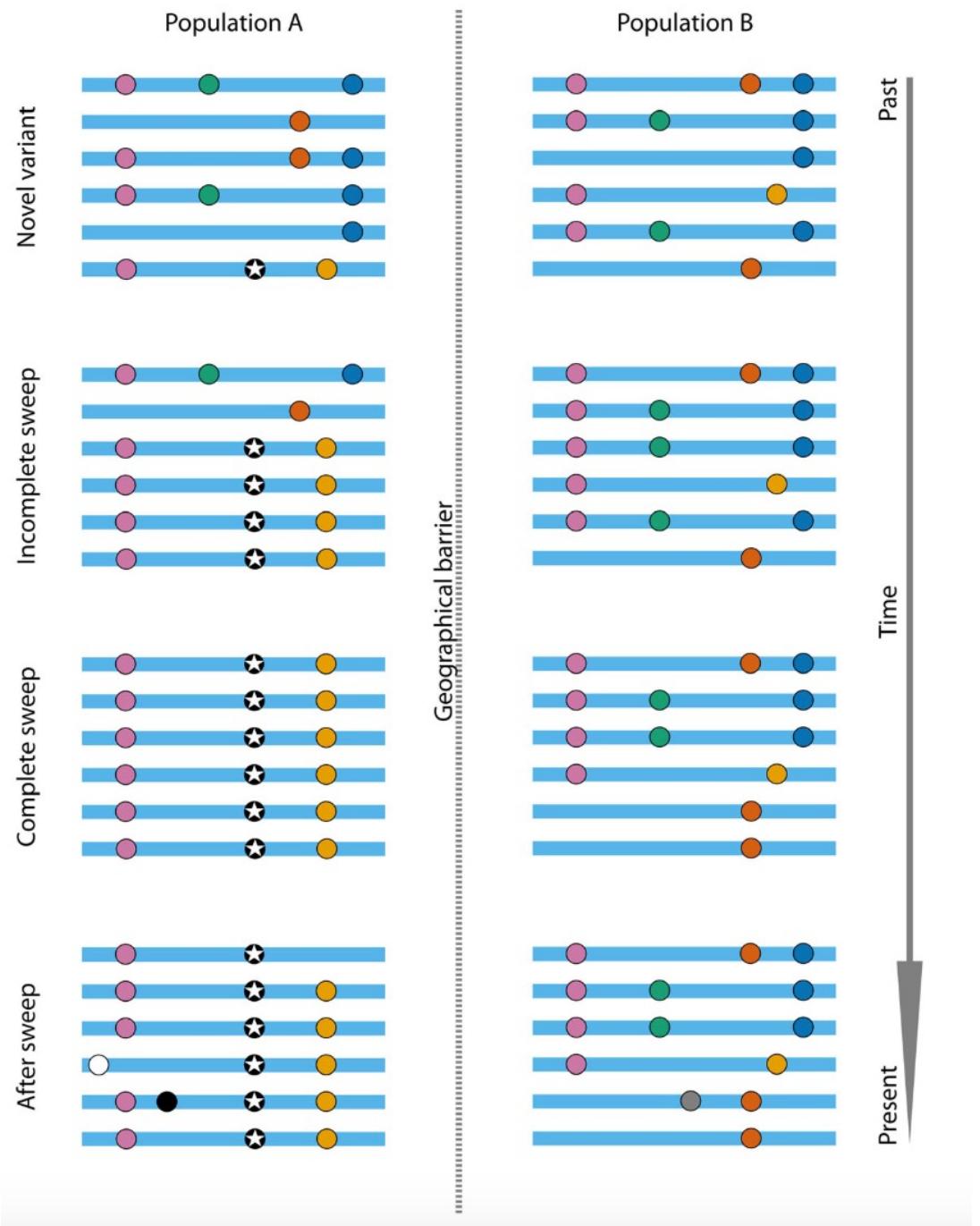
This leads to the evolution of clusters of adaptive loci showing large peaks of differentiation in the genome.

# 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.

Population A: Selective sweep

Population B: Only genetic drift and mutation

# 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

Research Article

[Evolutionary Biology, Genetics and Genomics](#)

## Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci

Fan Han, Minal Jamsandekar, Mats E Pettersson, Leyi Su, Angela P Fuentes-Pardo, Brian W Davis, Dorte Bekkevold, Florian Berg, Michele Casini ... Leif Andersson [see all »](#)

New Results

 [Follow this preprint](#)

### The origin and maintenance of supergenes contributing to ecological adaptation in Atlantic herring

 Minal Jamsandekar,  Mafalda S. Ferreira,  Mats E. Pettersson, Edward Farrell, Brian W. Davis,  Leif Andersson  
**doi:** <https://doi.org/10.1101/2023.10.23.562618>

This article is a preprint and has not been certified by peer review [what does this mean?].



[Abstract](#)

[Full Text](#)

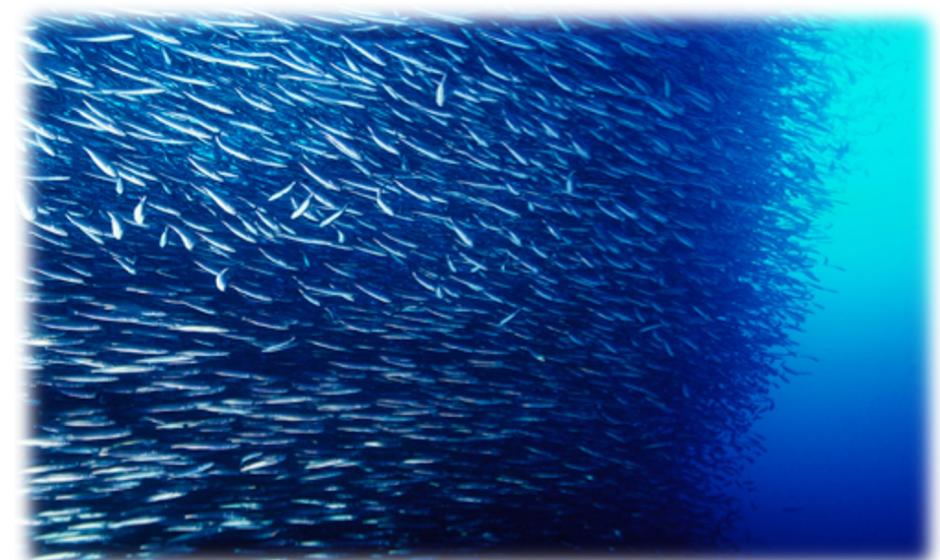
[Info/History](#)

[Metrics](#)

 [Preview PDF](#)



**Atlantic herring**  
North Atlantic species  
Large effective population size  
Selection drives divergence





Distribution of Atlantic herring



Two major life history strategies:

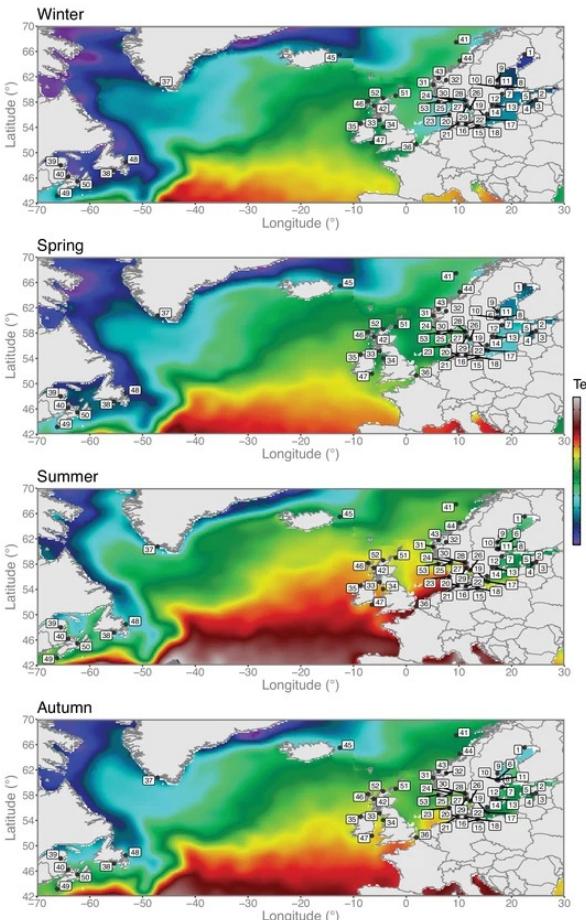
Spring spawning  
Autumn spawning

Hill et al 2019, PNAS

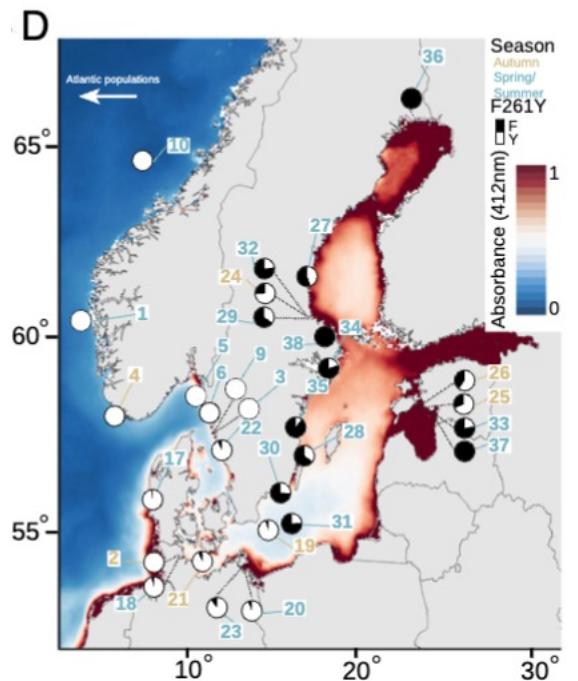
Han et al 2020, eLife

Jamsandekar et al 2023, bioRxiv

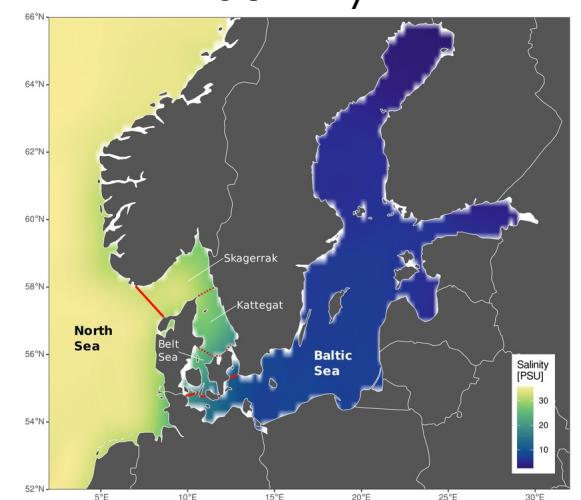
## Seasonal temperature gradients



## Optic environments



## Salinity



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

Genetic basis of traits:

What genes?



Genetic architecture:

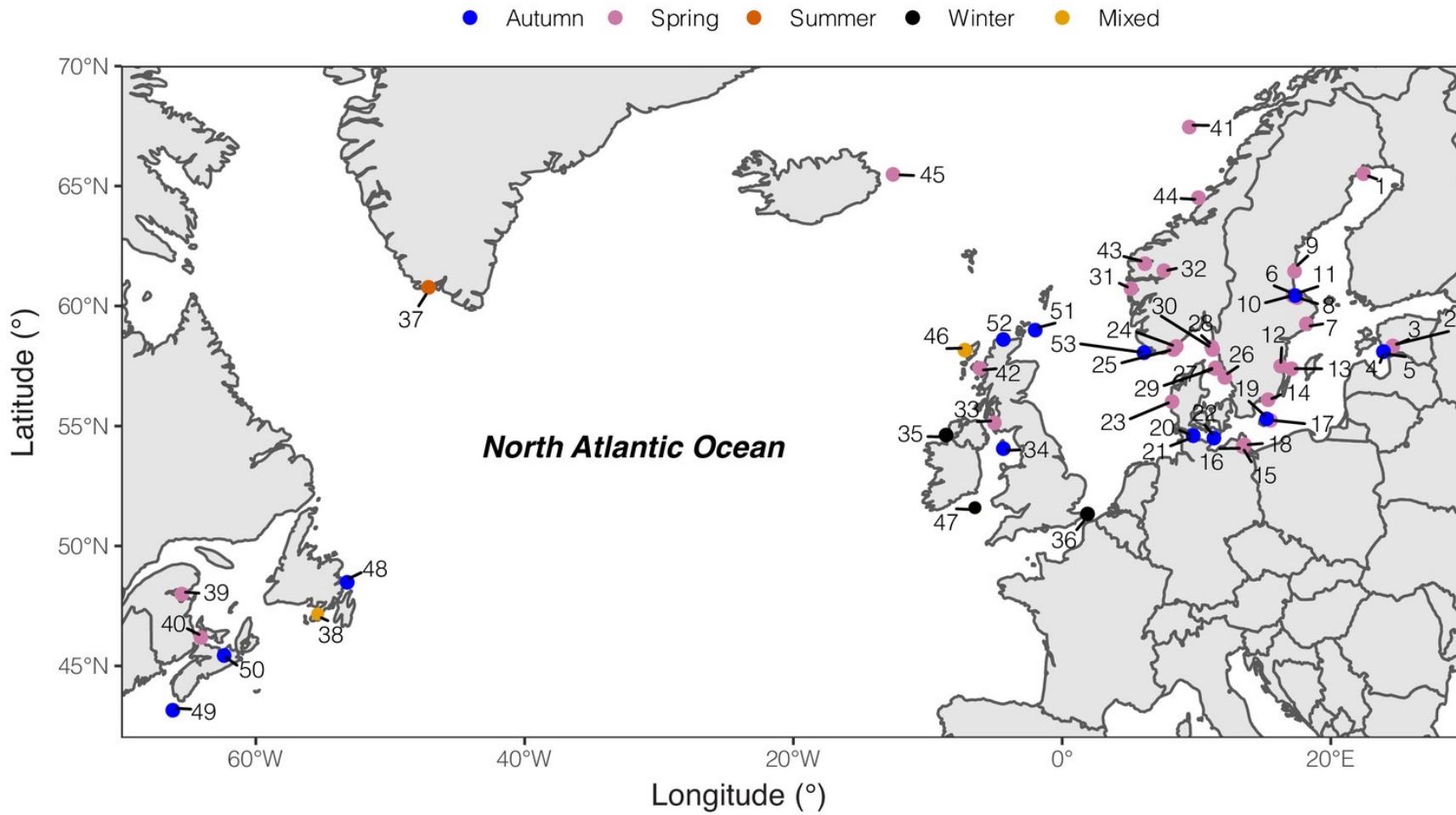
How many genes?



Genomic architecture:

Where are the genes? Structural variants?

# Genomic basis of adaptation to different environments?



53 populations

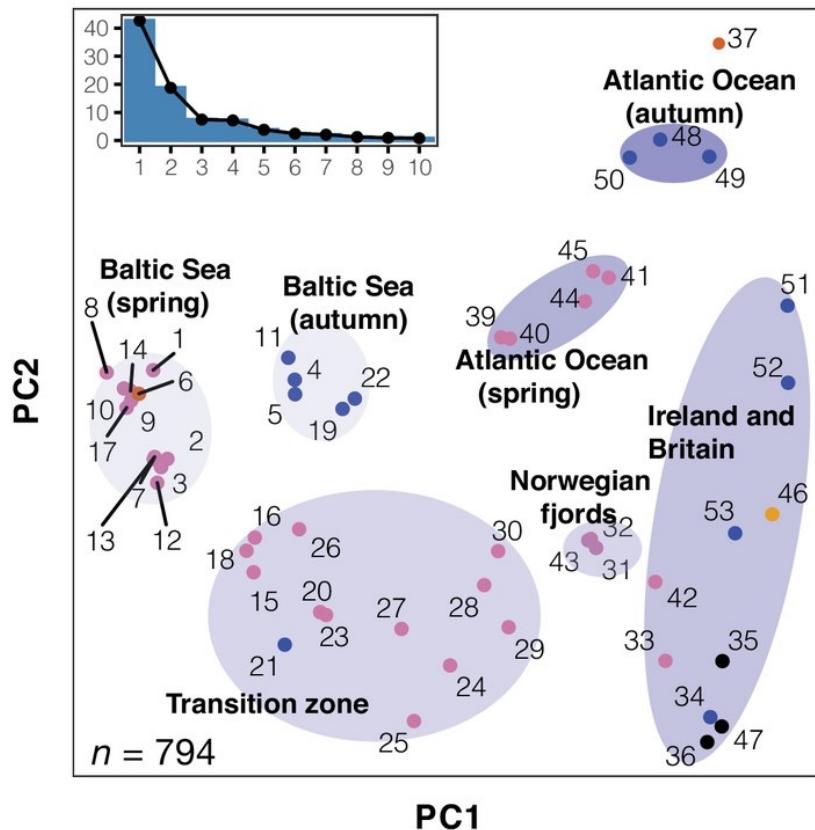
35-100 individuals

Pool-seq

# Population genomics

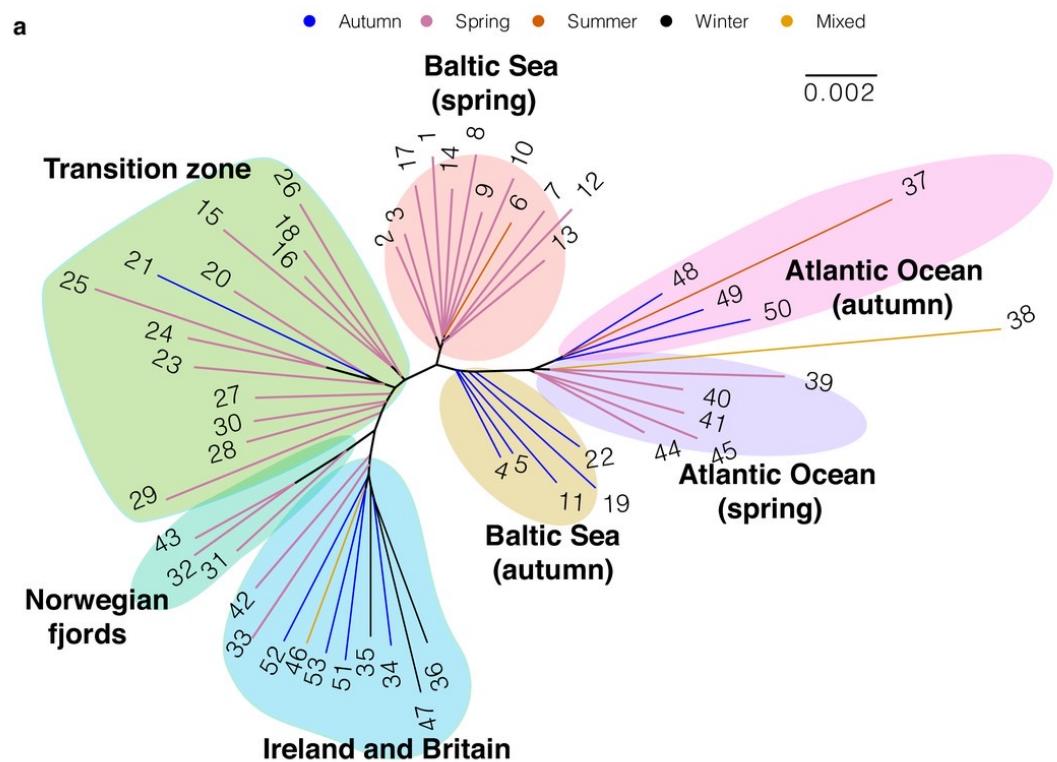
Principle component analysis

b



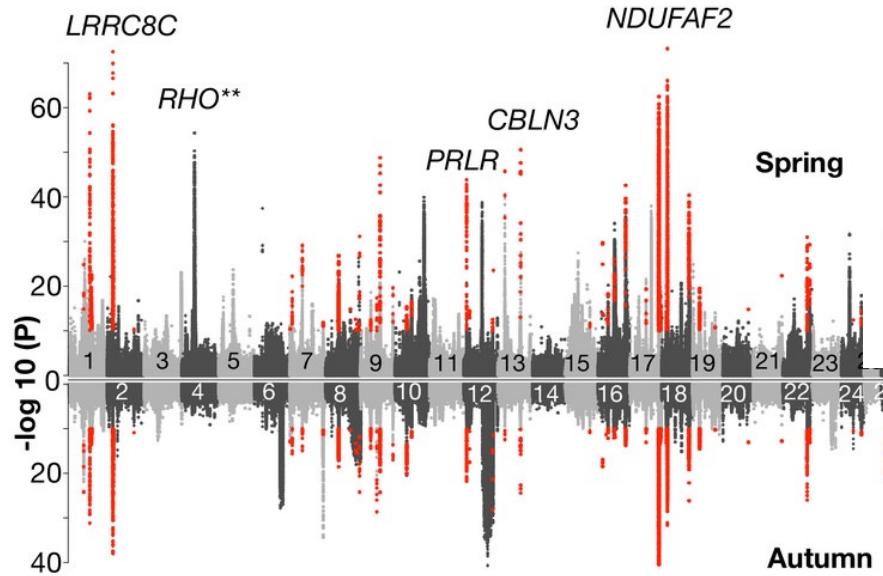
Neighbor joining tree

a

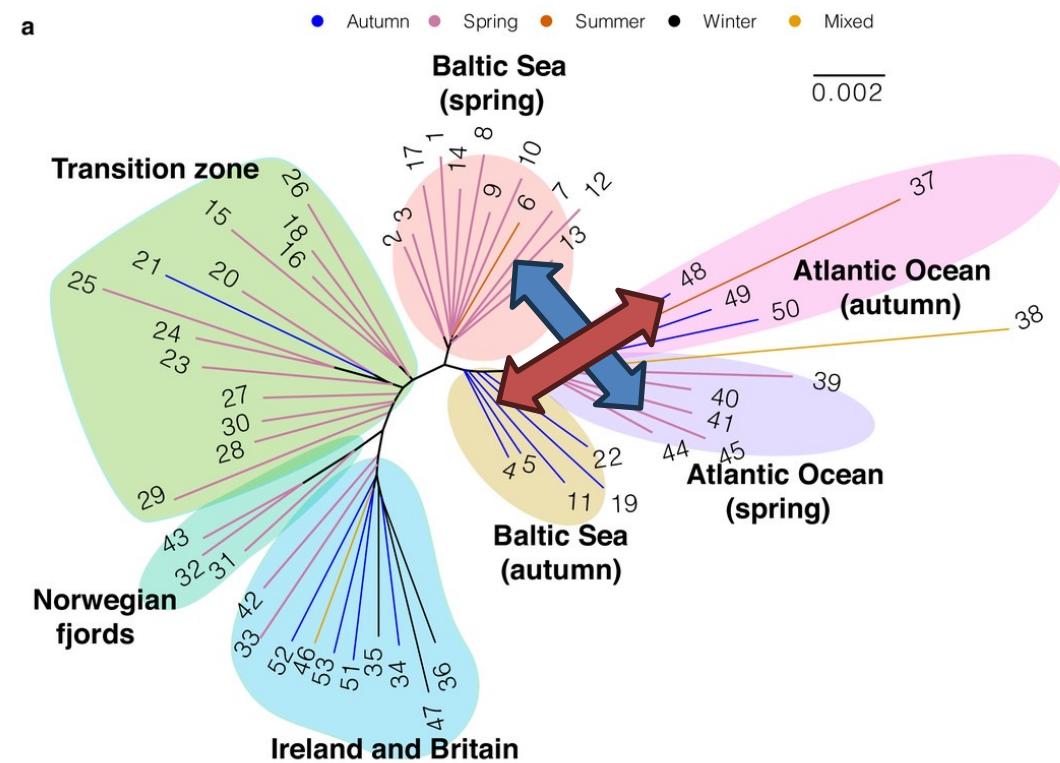


Using highly differentiated SNPs

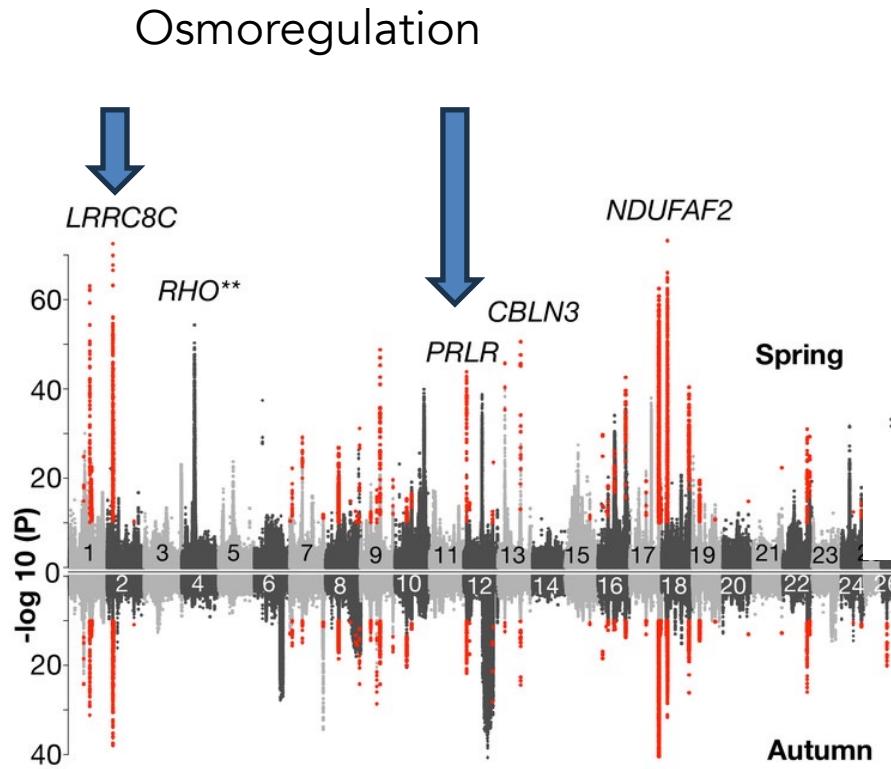
# Baltic vs Atlantic populations – Salinity contrast



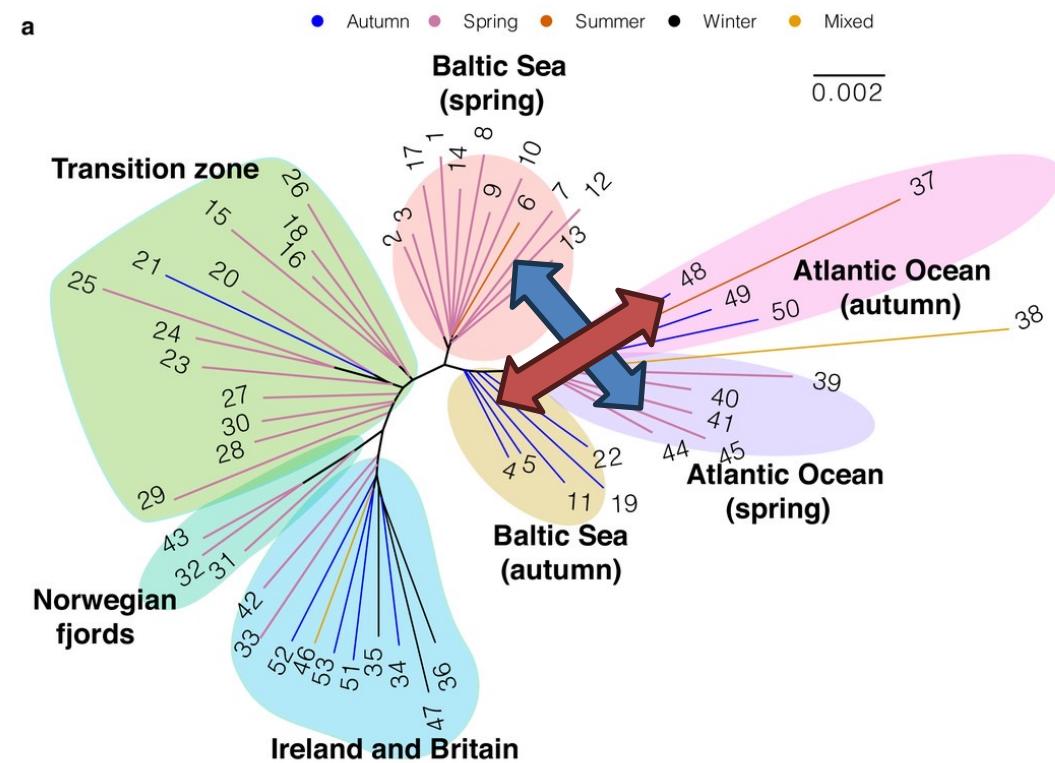
Allele frequency differences  
Chi<sup>2</sup> test for p-value



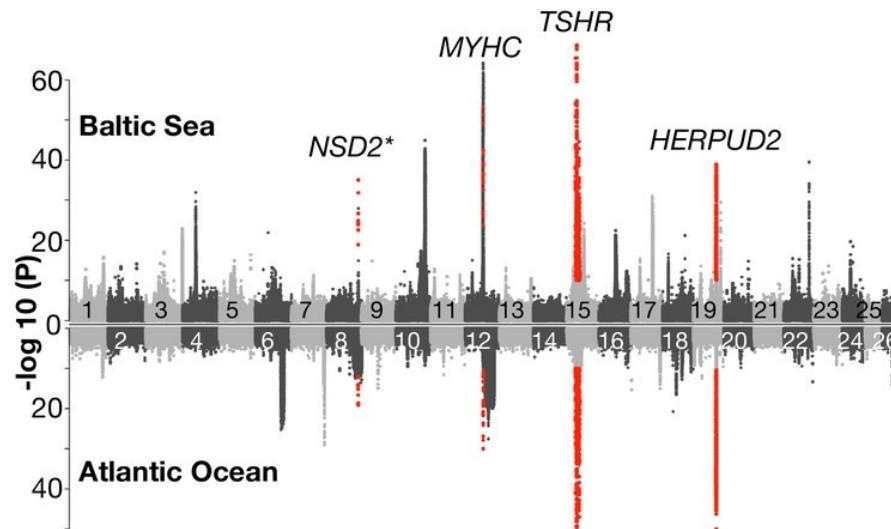
# Baltic vs Atlantic populations – Salinity contrast



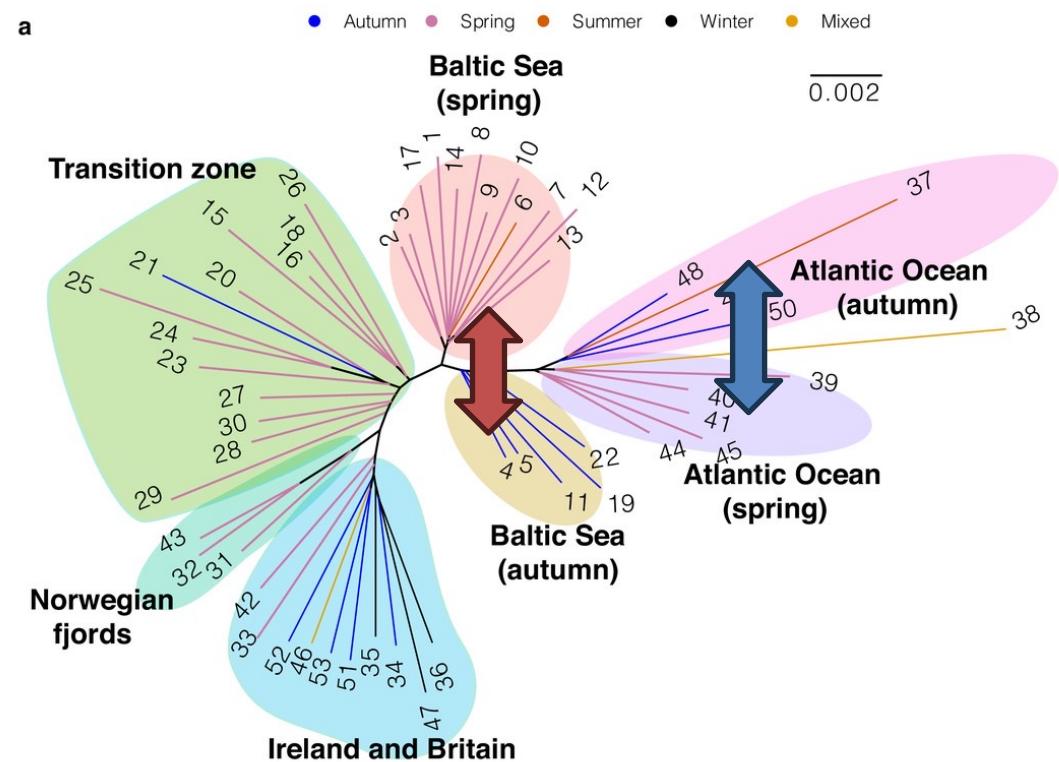
Allele frequency differences  
Chi<sup>2</sup> test for p-value



# Spring vs Autumn spawning populations

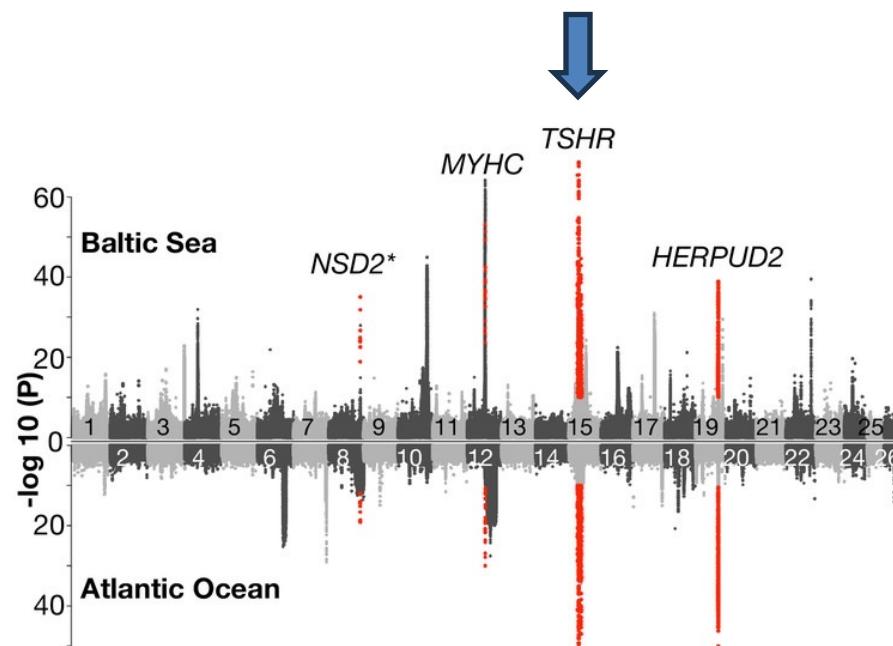


Allele frequency differences  
Chi<sup>2</sup> test for p-value

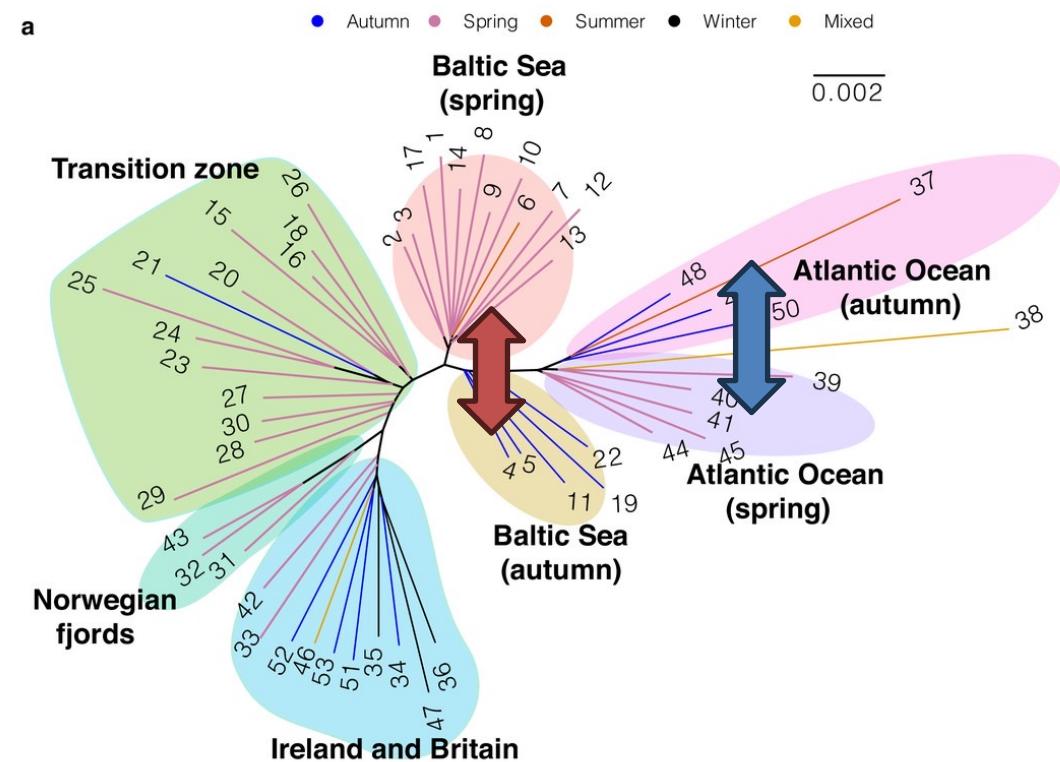


# Spring vs Autumn spawning populations

# Thyroid signaling hormone receptor



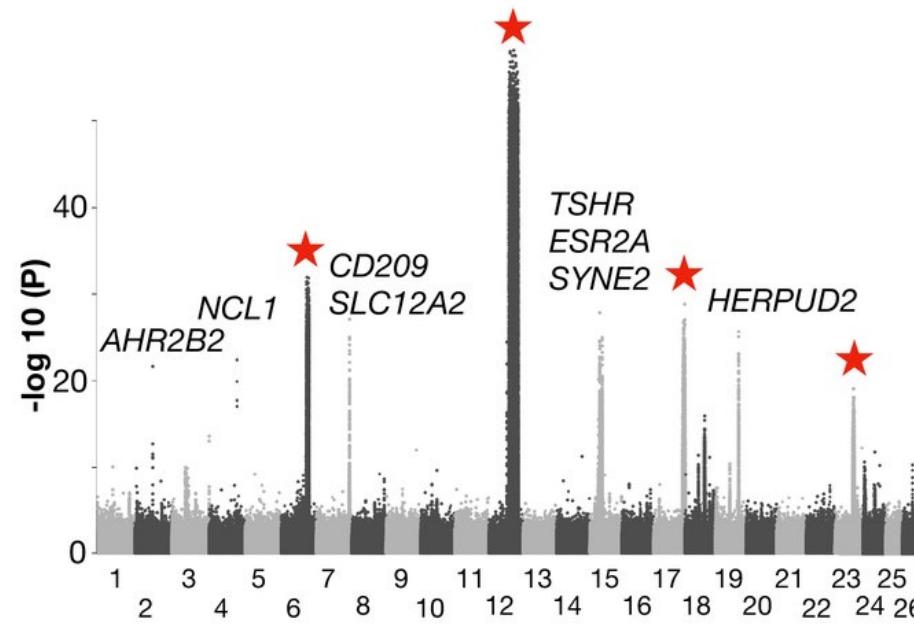
## Allele frequency differences Chi<sup>2</sup> test for p-value



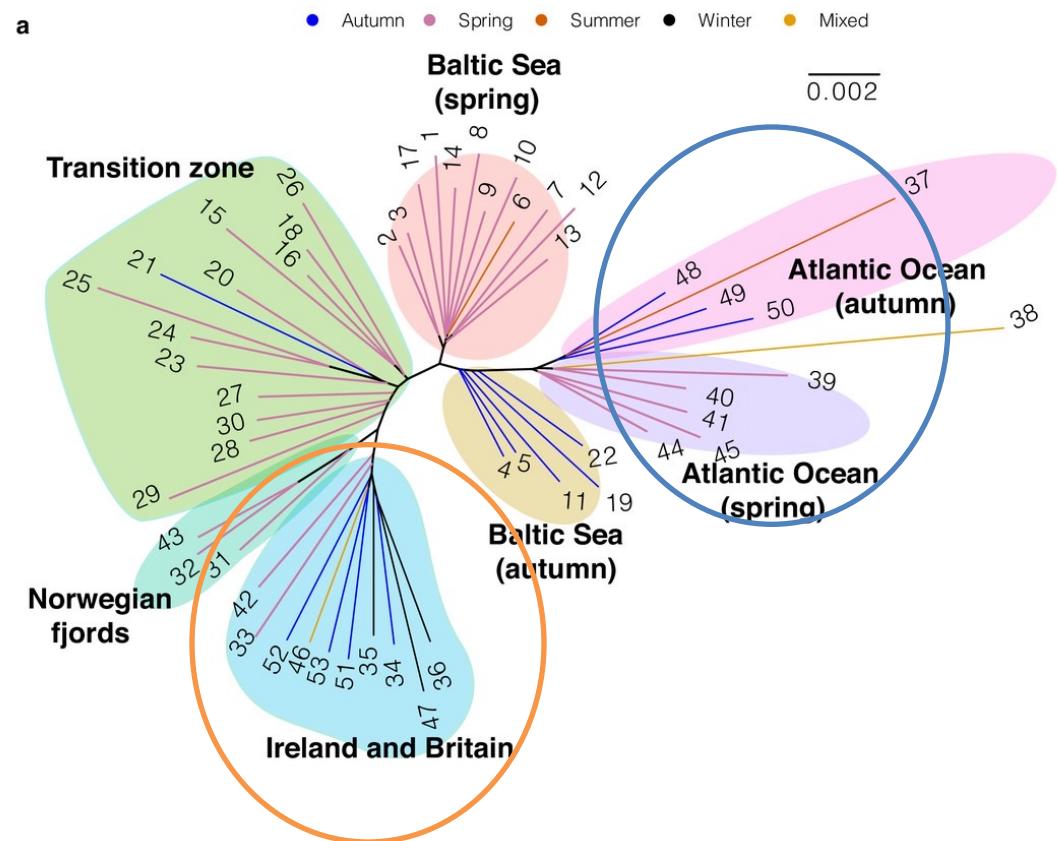
Colder

Warmer

# North vs South populations



Allele frequency differences  
Chi<sup>2</sup> test for p-value

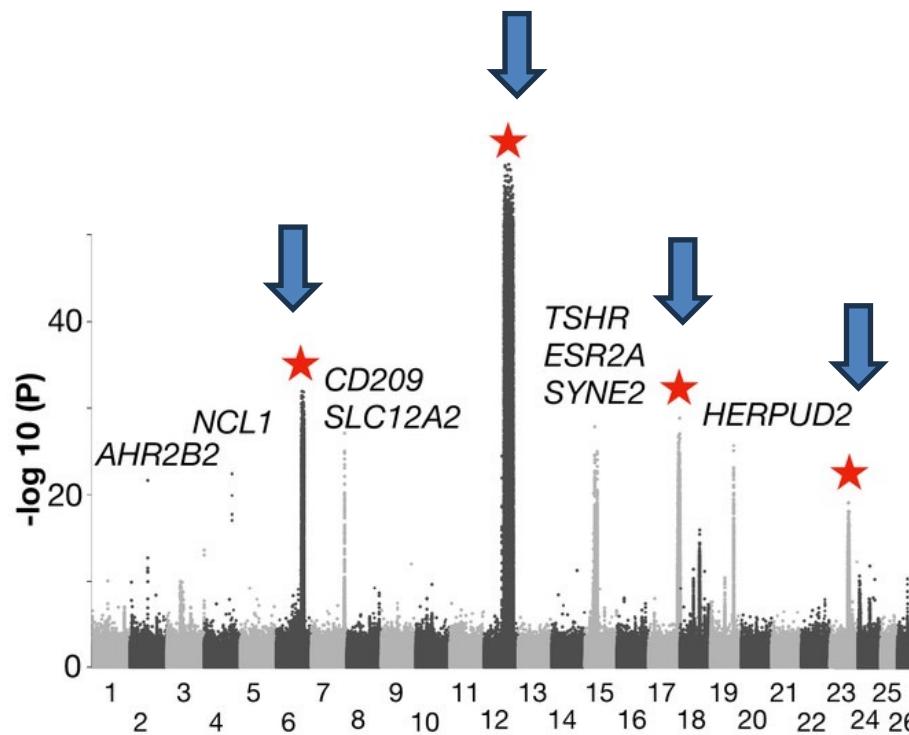


Colder

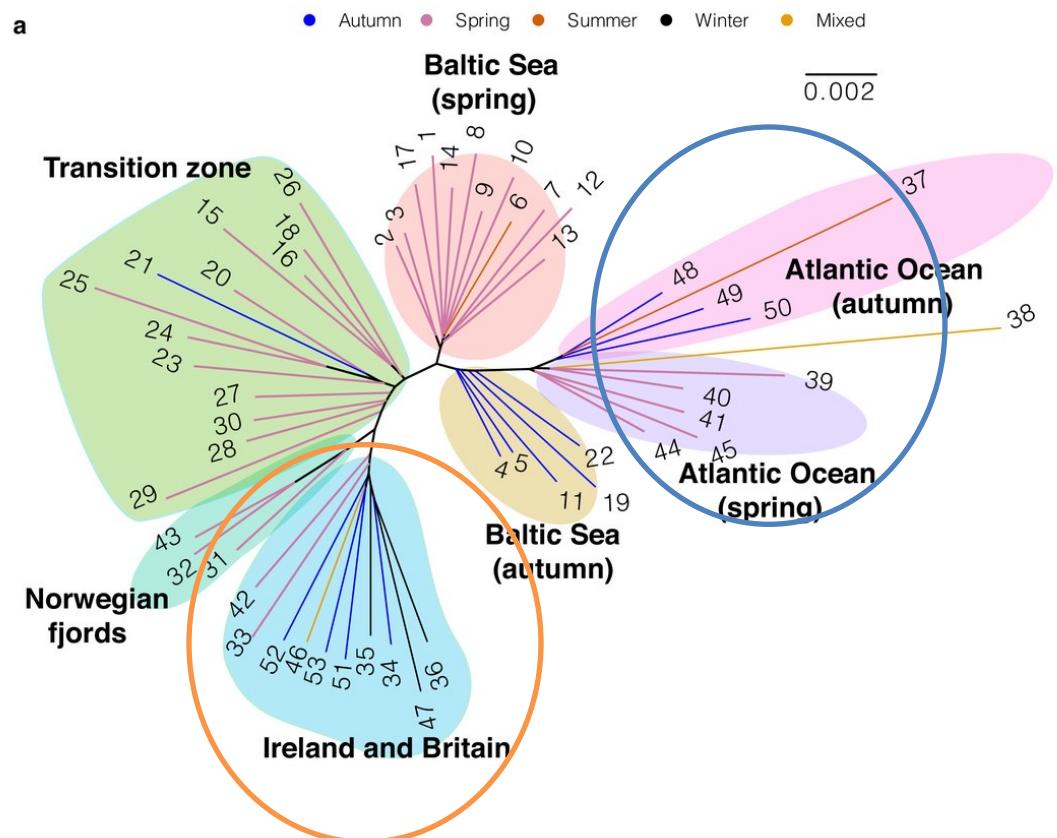
Warmer

# North vs South populations

Inversions



Allele frequency differences  
Chi<sup>2</sup> test for p-value

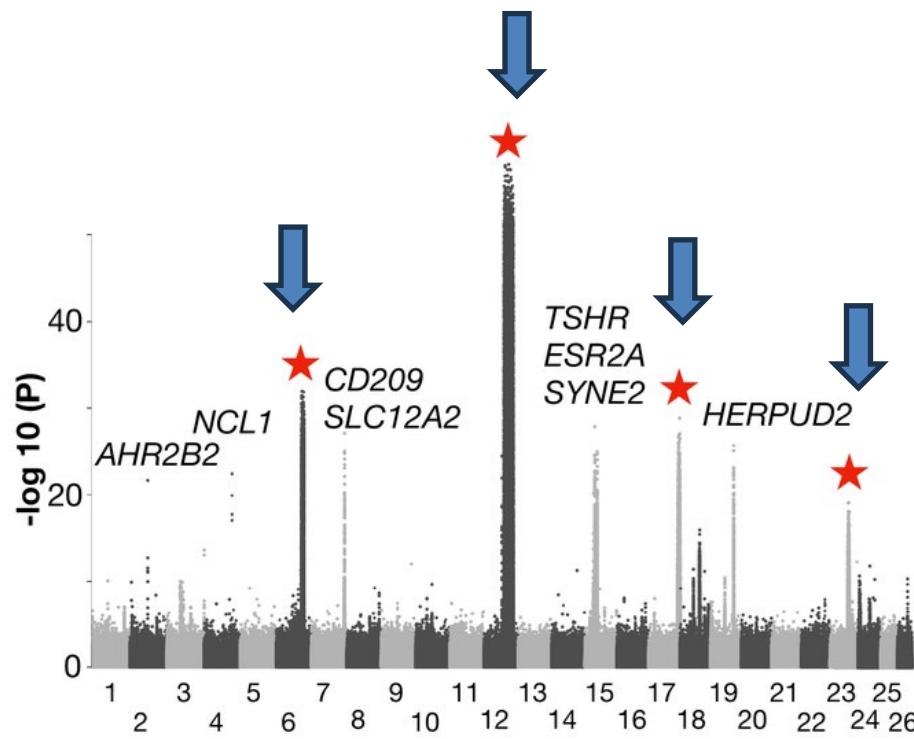


Colder

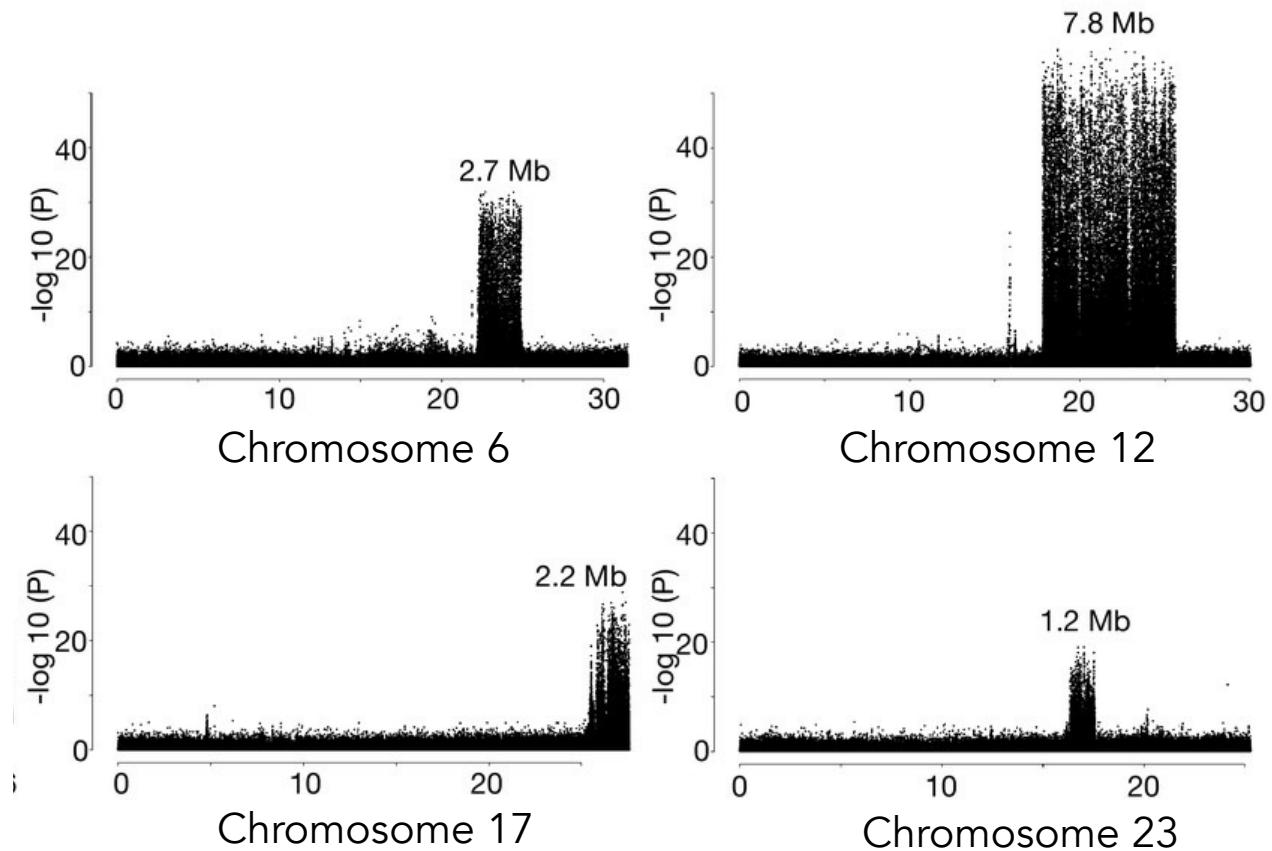
Warmer

# North vs South populations

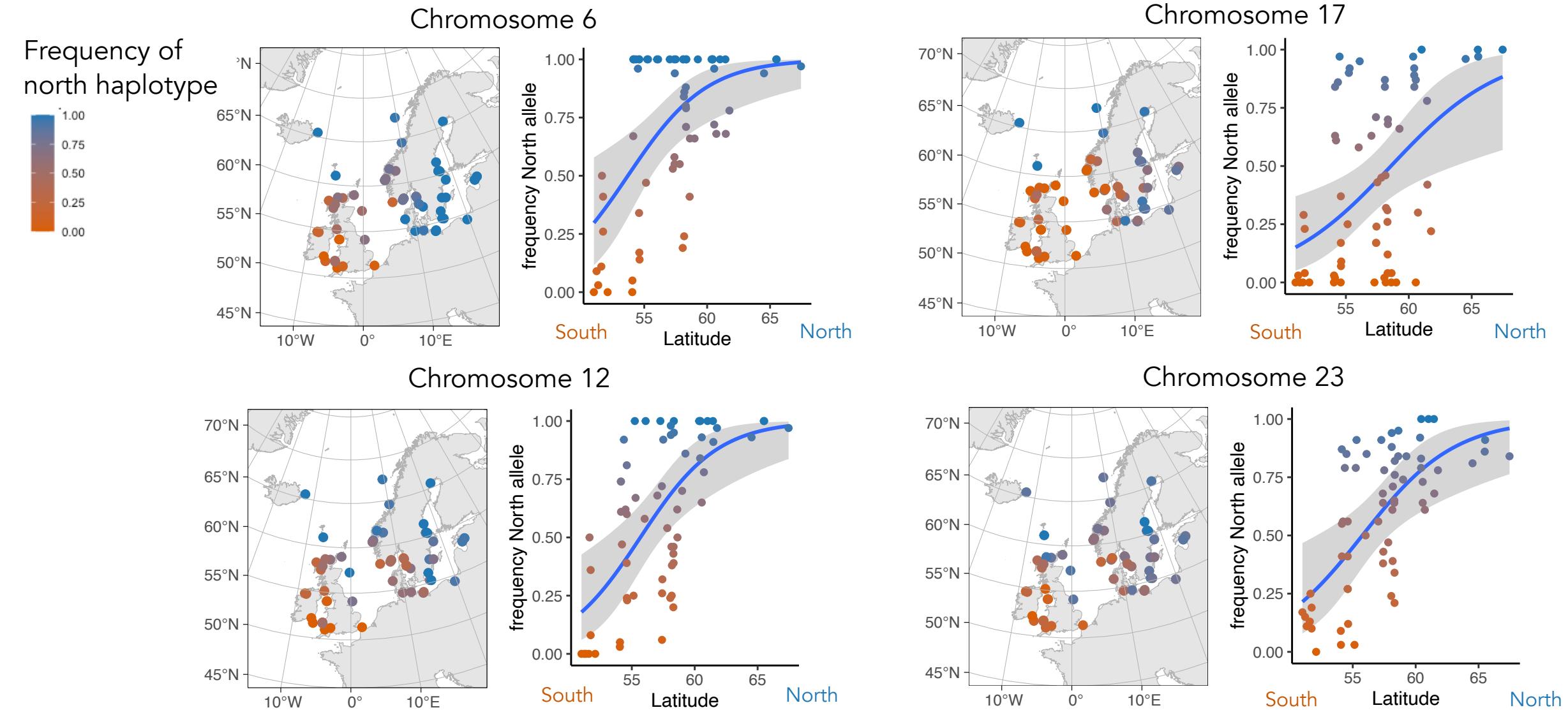
Inversions



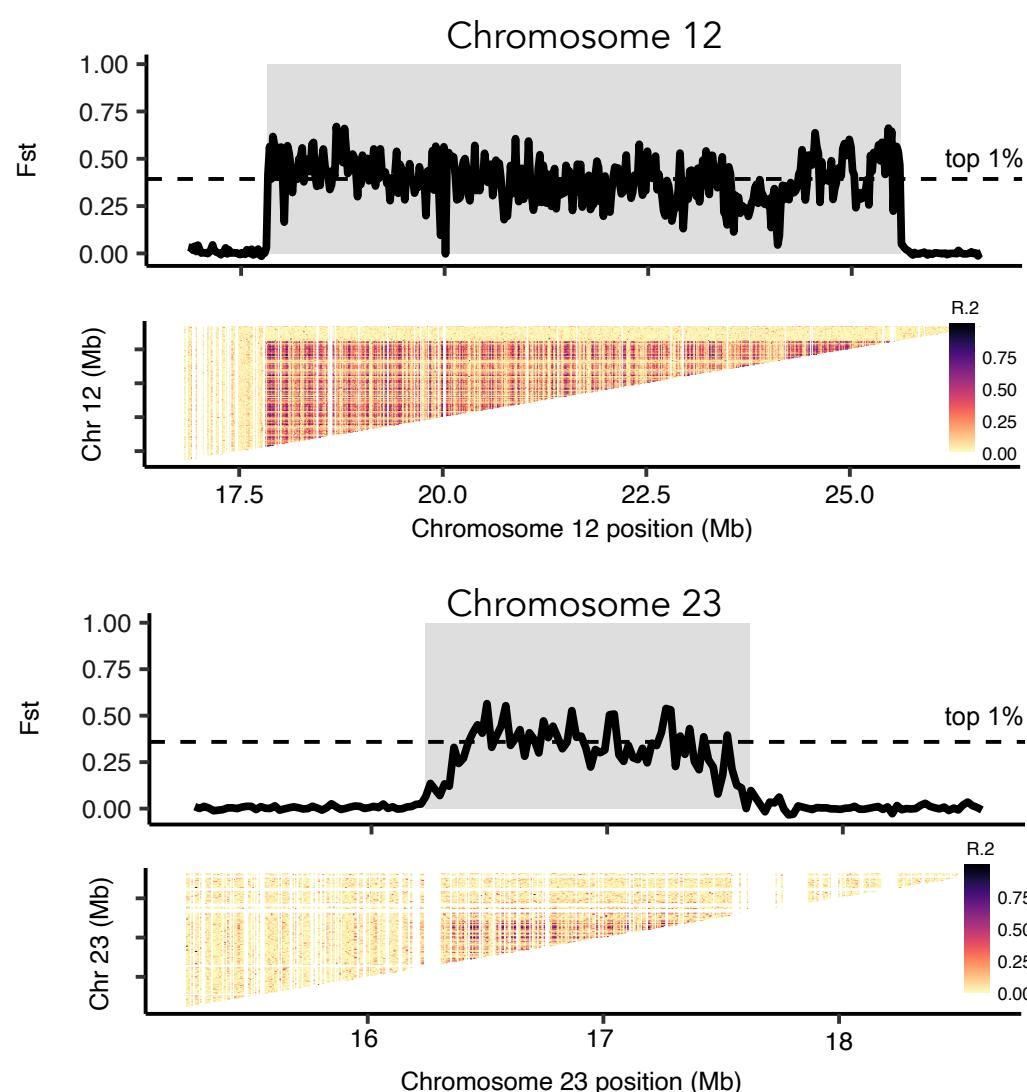
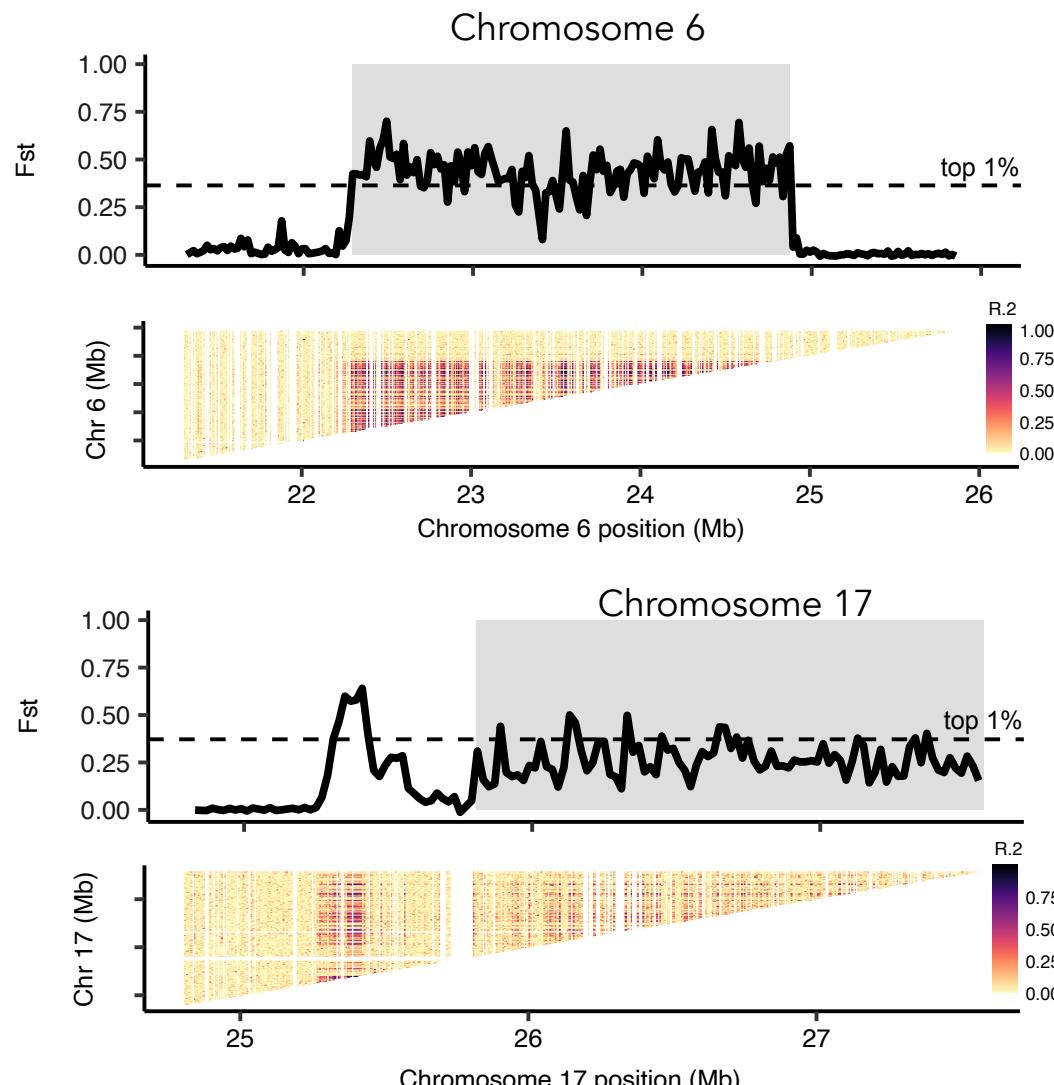
Allele frequency differences  
Chi<sup>2</sup> test for p-value



# High frequency of North and South haplotypes at the extremes of a latitudinal gradient



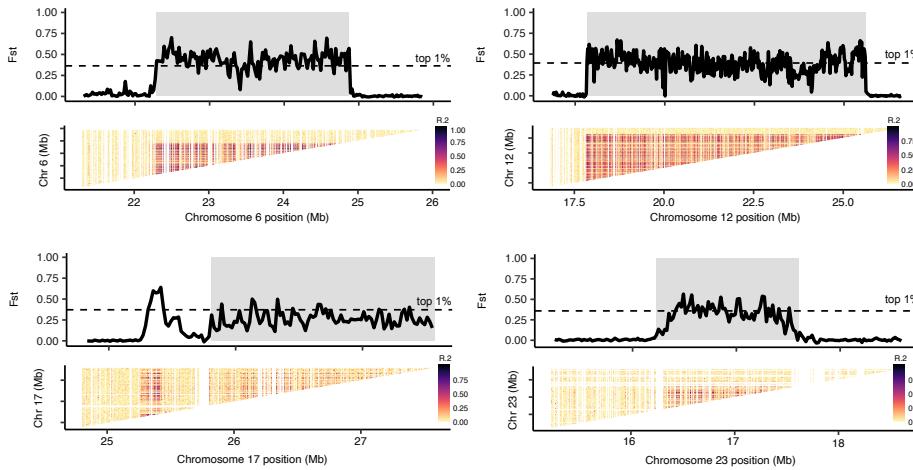
# High differentiation ( $F_{st}$ ) and linkage disequilibrium between South and North homozygotes





Atlantic herring

North Atlantic species  
Large effective population size



$F_{st}$  inversions >>  $F_{st}$  average WG

Population migration rate ( $Nm$ ) is strong enough ( $> 1$ ) to homogenize differentiation

Divergent 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 but also strong selection > clusters of differentiated loci

The end.