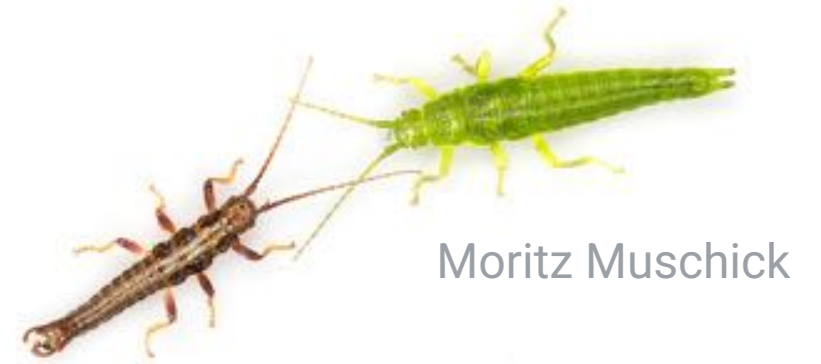# Confounding factors

Day 3 - Lecture 1
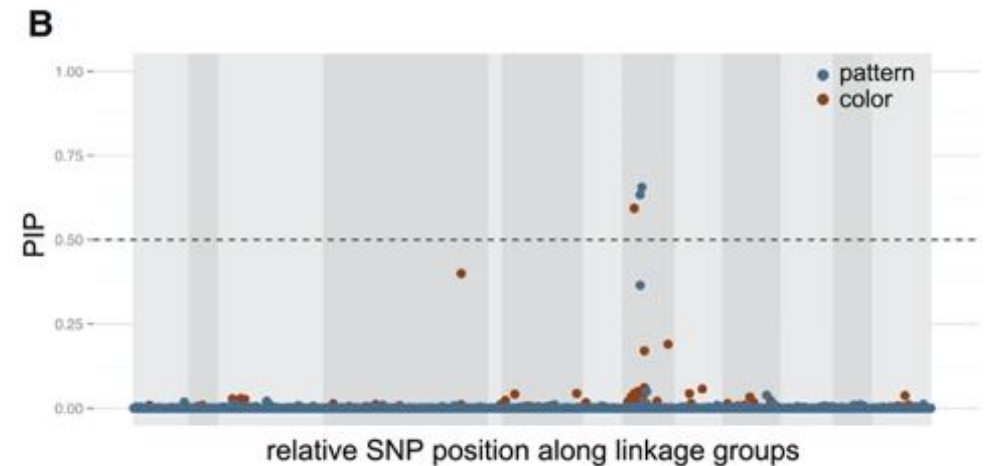
Anna Tigano, Ph.D.

Expectation

Selection increases the frequency of beneficial mutations, so that when we compare two populations adapted to different local conditions we should see a difference in frequency at the locus underpinning that adaptive trait.


Moritz Muschick

Comeault et al. 2015, Current Biology

Yesterday I presented on the factors that contribute to our ability to detect signatures of a selective sweep in the genome.

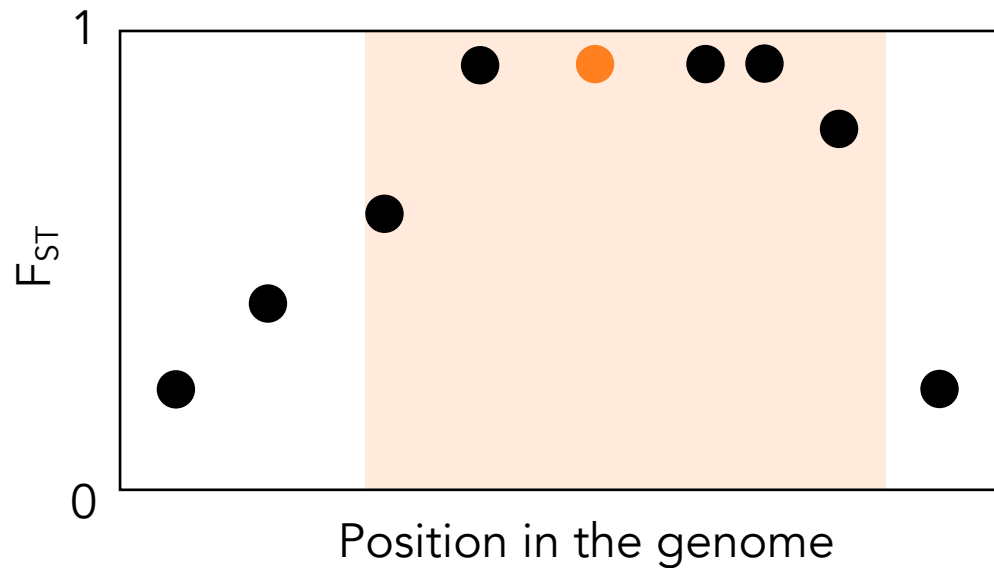However, that was a simple scenario, only one of many possible, and often more complex, scenarios.

Linked selection

Due to linked selection, several sites in close proximity to the real target of selection could show high F_{ST}.
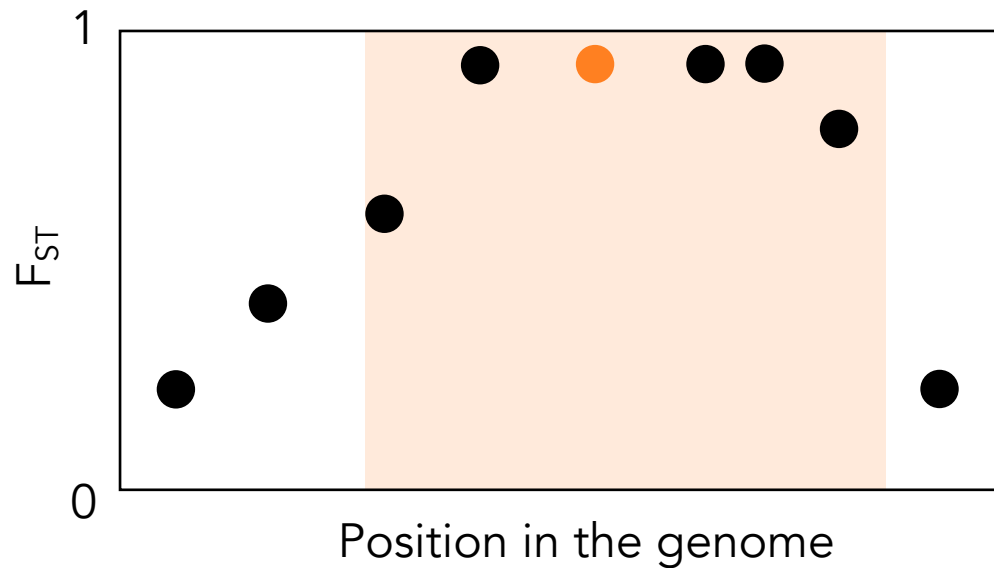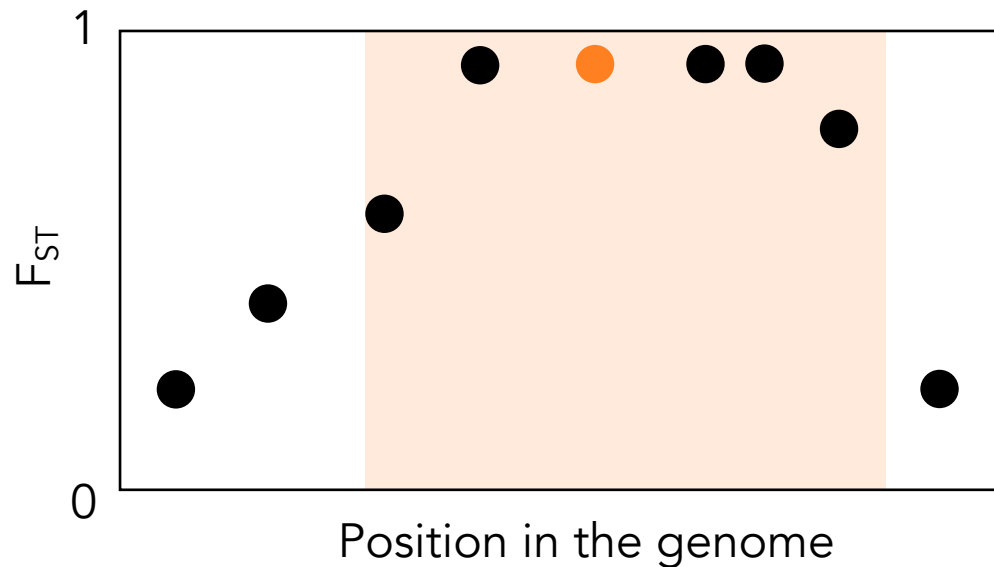
Linked selection

Due to linked selection, several sites in close proximity to the real target of selection could show high F_{ST}.

With whole genome data, it is likely that you will genotype all, or most, of these sites with high F_{ST}.

Linked selection

Due to linked selection, several sites in close proximity to the real target of selection could show high F_{ST}.

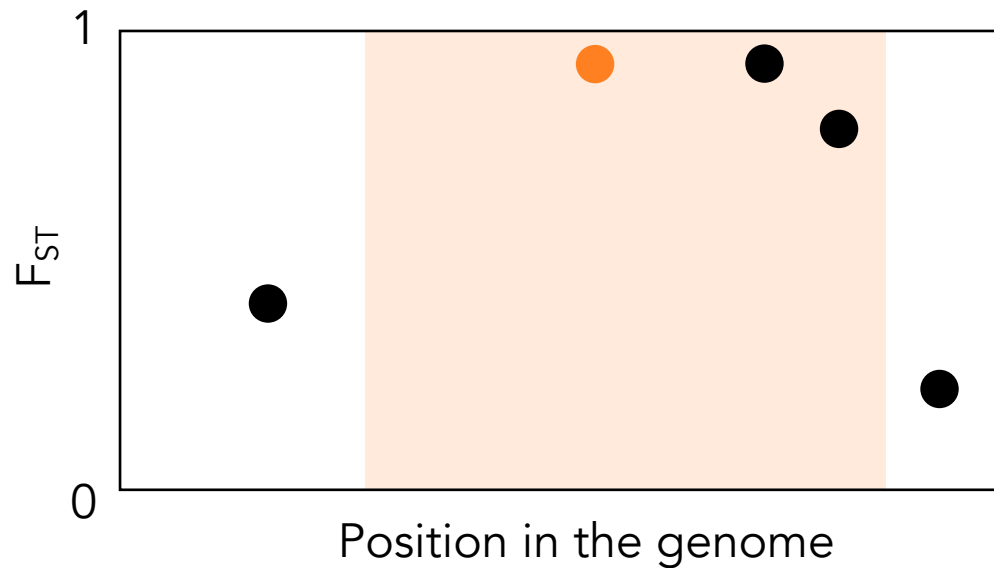F_{ST} vs Position in the genome

With whole genome data, it is likely that you will genotype all, or most, of these sites with high F_{ST}.

The tighter linkage is, the harder is to disentangle target from linked sites.

Linked selection

Due to linked selection, several sites in close proximity to the real target of selection could show high F_{ST}.

With reduced-representation approaches, you won't.
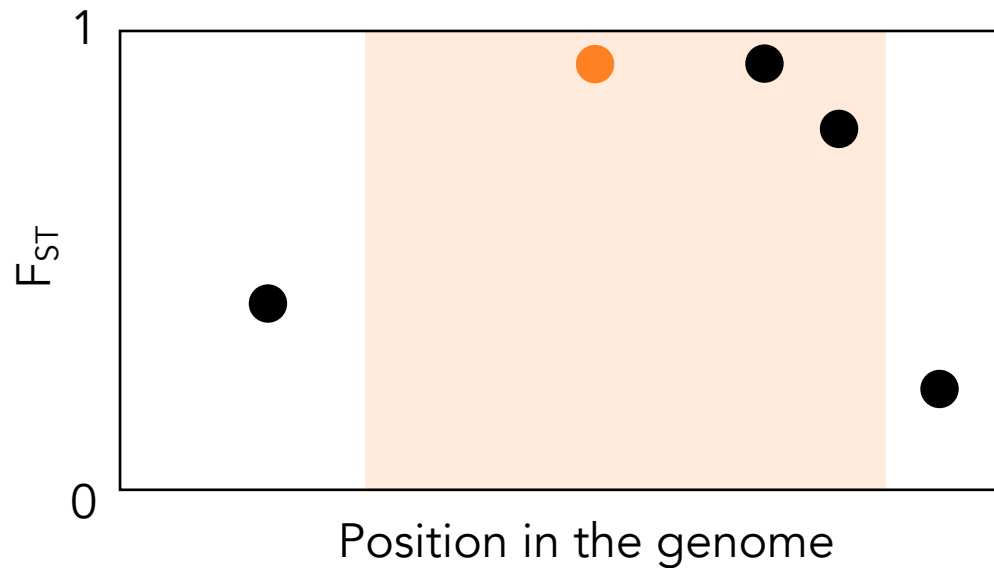
Linked selection

Due to linked selection, several sites in close proximity to the real target of selection could show high F_{ST}.

With reduced-representation approaches, you won't.

Your F_{ST} outlier could be the real target of selection, or a neutral sites linked to it.
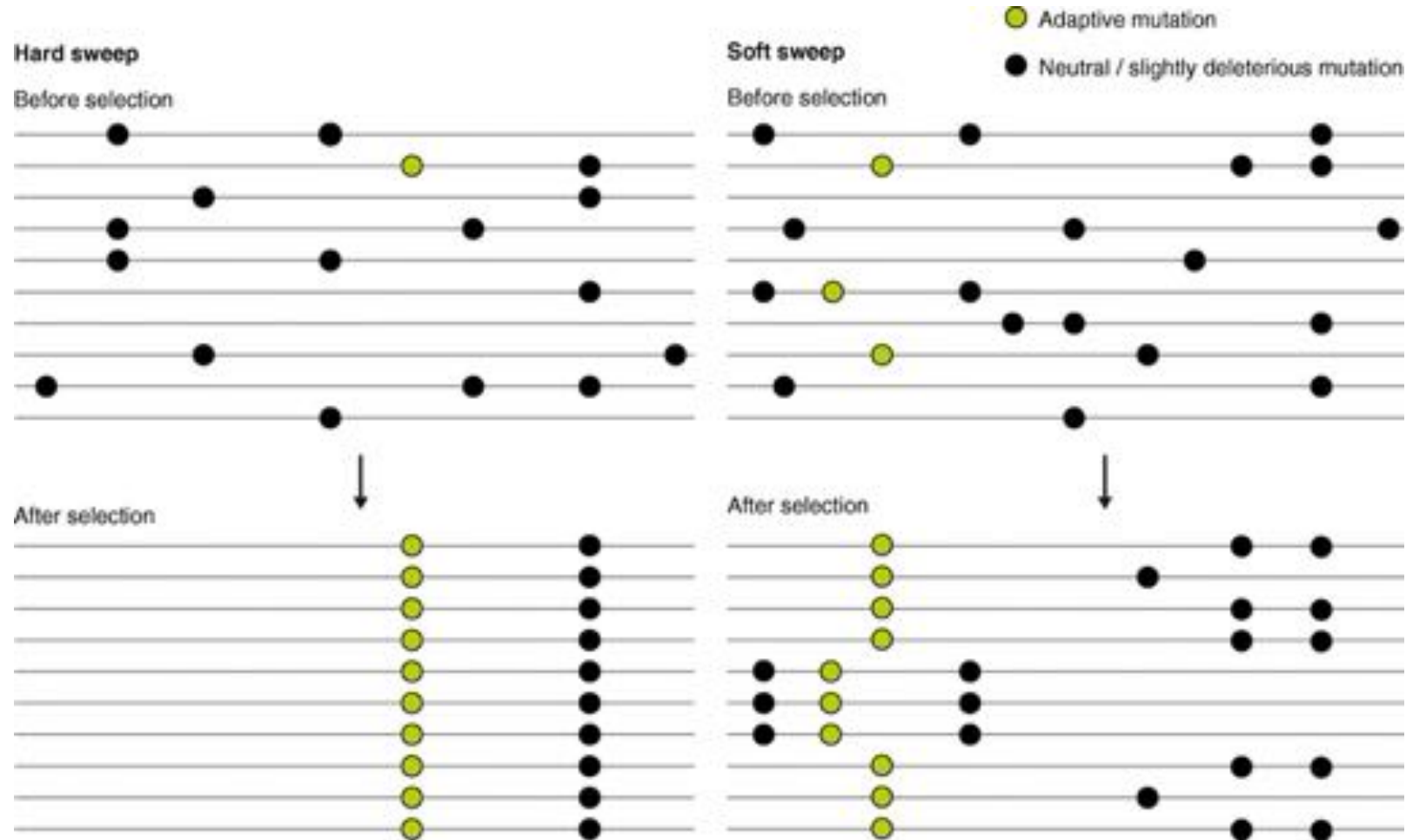
Positive selection

Natural selection = positive selection

→ Increases frequency of beneficial mutations

One mutation on one genomic background ← **Hard vs. soft sweeps** → Different mutations and/or different genomic backgrounds

Adaptive mutation

Neutral / slightly deleterious mutation

Hard sweep

Before selection

After selection

Soft sweep
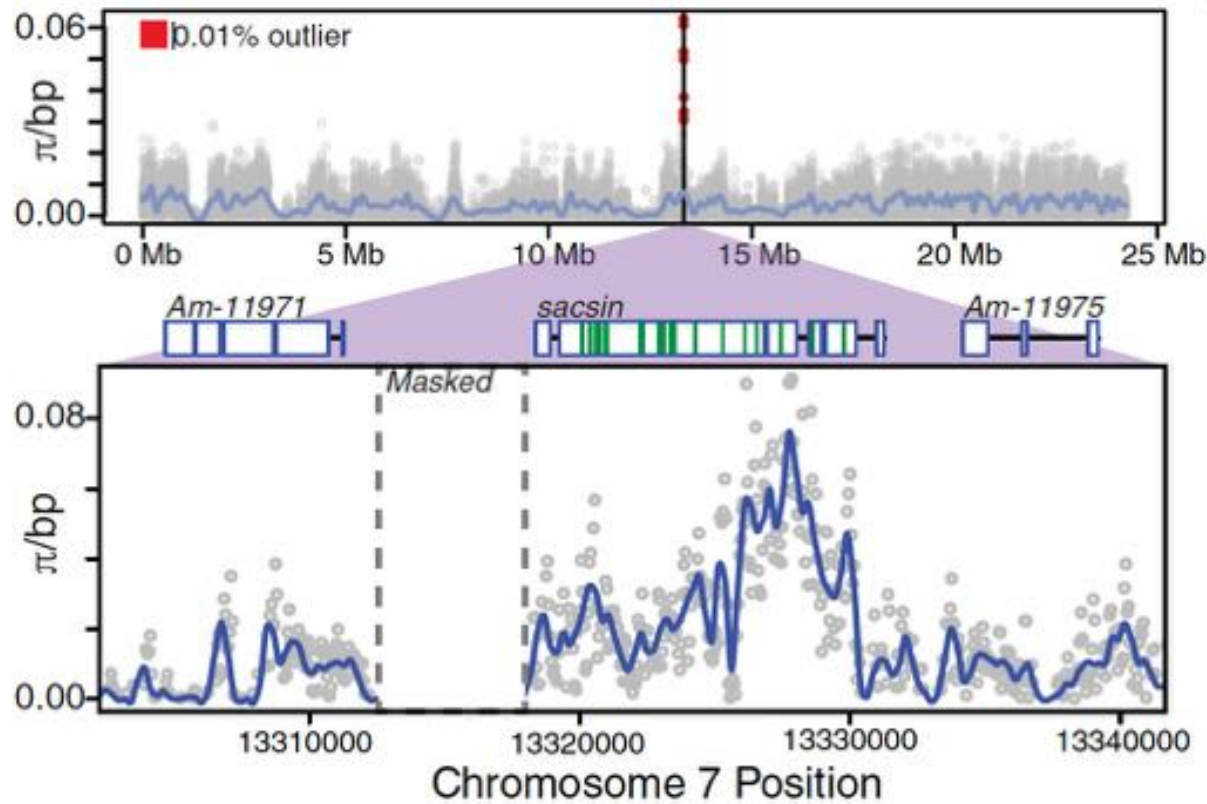
Before selection

After selection

Balancing selection

Balancing selection maintains multiple alleles in the gene pool of a population at frequencies larger than expected from genetic drift alone.

Mechanisms:

- Heterozygote advantage

- Frequency-dependent selection

- Spatially and/or temporally varying selection

Signature of balancing selection

Higher genetic diversity at the locus under balancing selection

Fuller et al. 2020, Science

Background selection

Negative selection = purifying selection = background selection
→ Removes alleles that are deleterious

Similarly to positive selection, negative selection reduces diversity around the deleterious mutations under selection.

Protein-coding regions, for example, are under strict functional constraints and purifying selection is what keeps them functional.
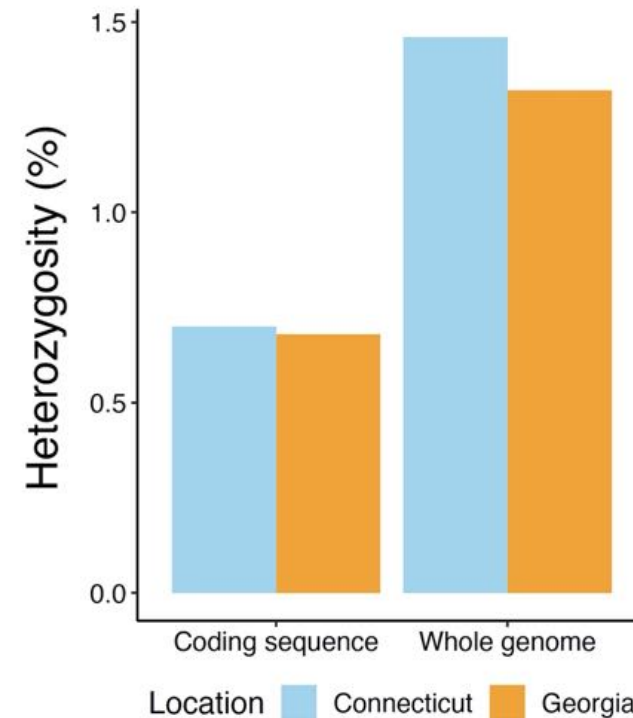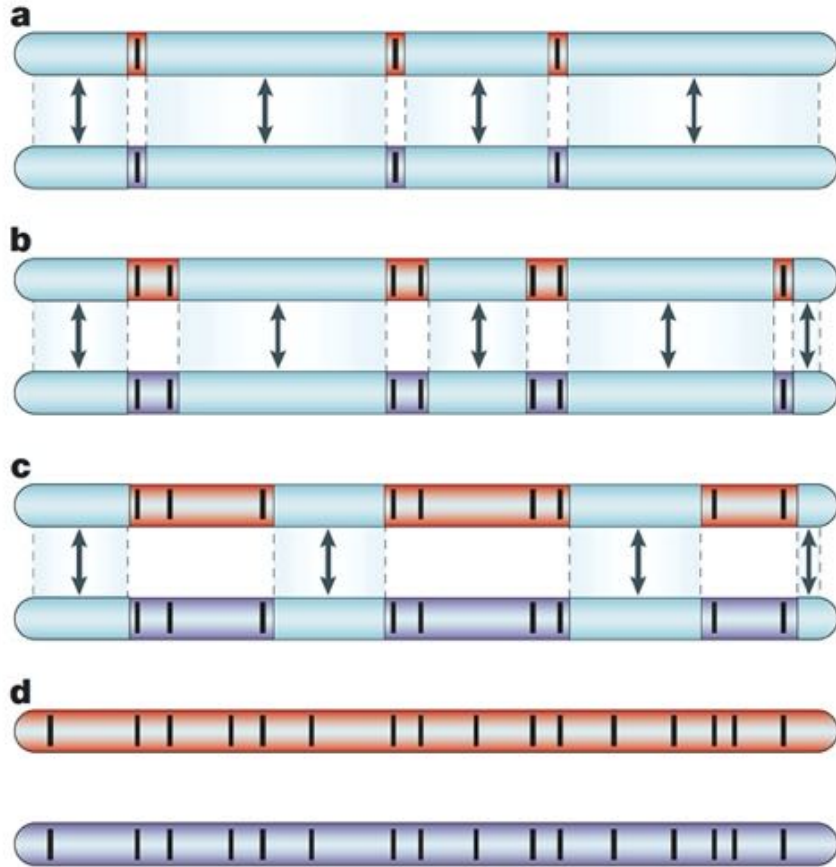

Tigano et al., in prep

Could background selection generate the same genomic islands of differentiation as positive selection?
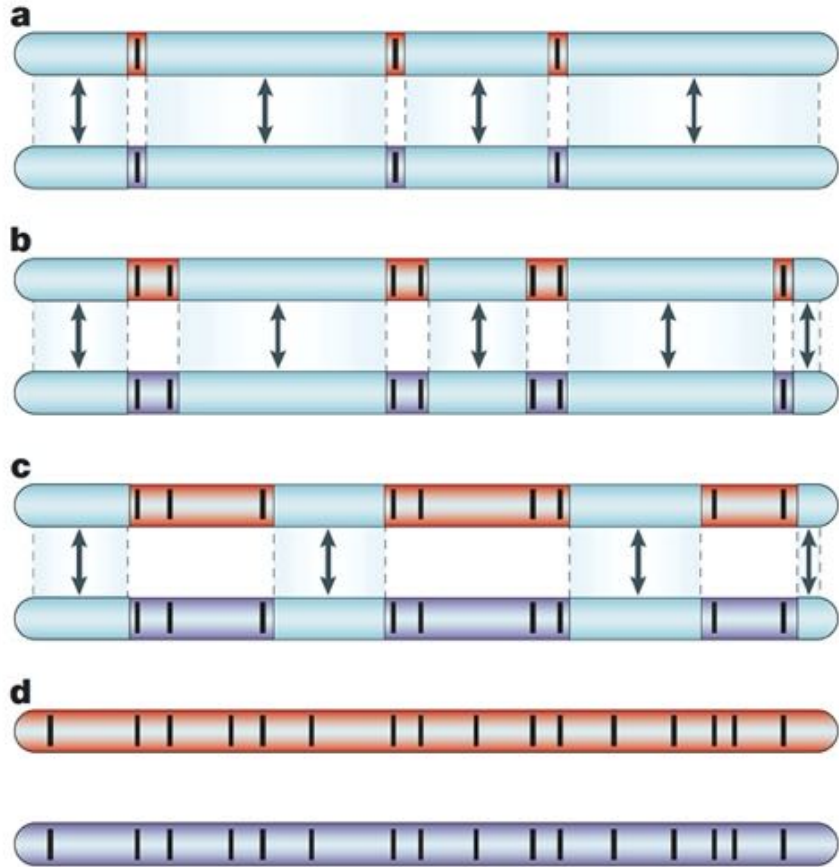
Genomic islands of differentiation

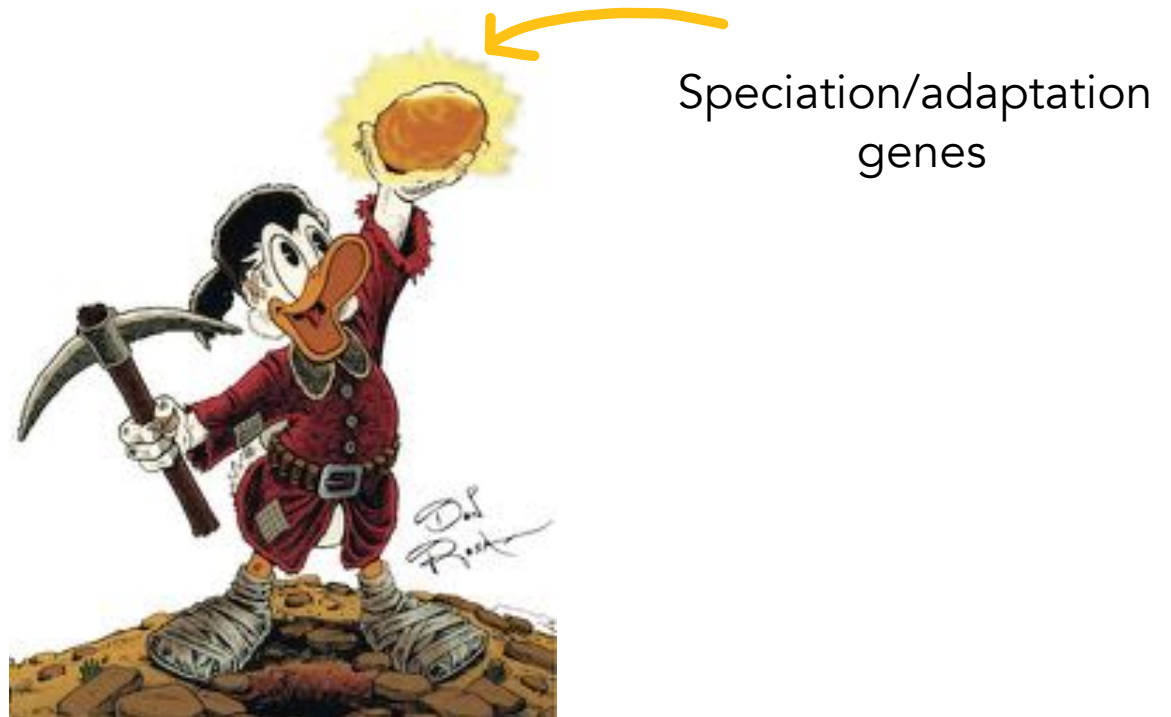The speciation-with-gene-flow model is based on the genic view of speciation by Wu (2001)

Wu and Ting 2004, Nat. Gen. Rev.

Genomic islands of differentiation

The speciation-with-gene-flow model is based on the genic view of speciation by Wu (2001)

Wu and Ting 2004, Nat. Gen. Rev.

Speciation/adaptation genes

Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds

Kira E. Delmore,[1,2] Juan S. Lugo Ramos,[1] Benjamin M. Van Doren,[3] Max Lundberg,[4] Staffan Bensch,[4] Darren E. Irwin,[5] and Miriam Liedvogel[1]

Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa

Nagarjun Vijay[1,2] | Matthias Weissensteiner[1,3] | Reto Burri[1,4] | Takeshi Kawakami[1,5] | Hans Ellegren[1] | Jochen B. W. Wolf[1,3]

Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers

Reto Burri,[1] Alexander Nater,[1] Takeshi Kawakami,[1] Carina F. Mugal,[1] Pall I. Olason,[2] Linnea Smeds,[1] Alexander Suh,[1] Ludovic Dutoit,[1] Stanislav Bureš,[3] Laszlo Z. Garamszegi,[4] Silje Hogner,[5,6] Juan Moreno,[7] Anna Qvarnström,[8] Milan Ružić,[9] Stein-Are Sæther,[5,10] Glenn-Peter Sætre,[5] Janos Török,[11] and Hans Ellegren[1]

Correlated patterns of genetic diversity and differentiation across an avian family

BENJAMIN M. VAN DOREN,*† LEONARDO CAMPAGNA,*† BARBARA HELM,‡ JUAN CARLOS ILLERA,§ IRBY J. LOVETTE*† and MIRIAM LIEDVOGEL¶

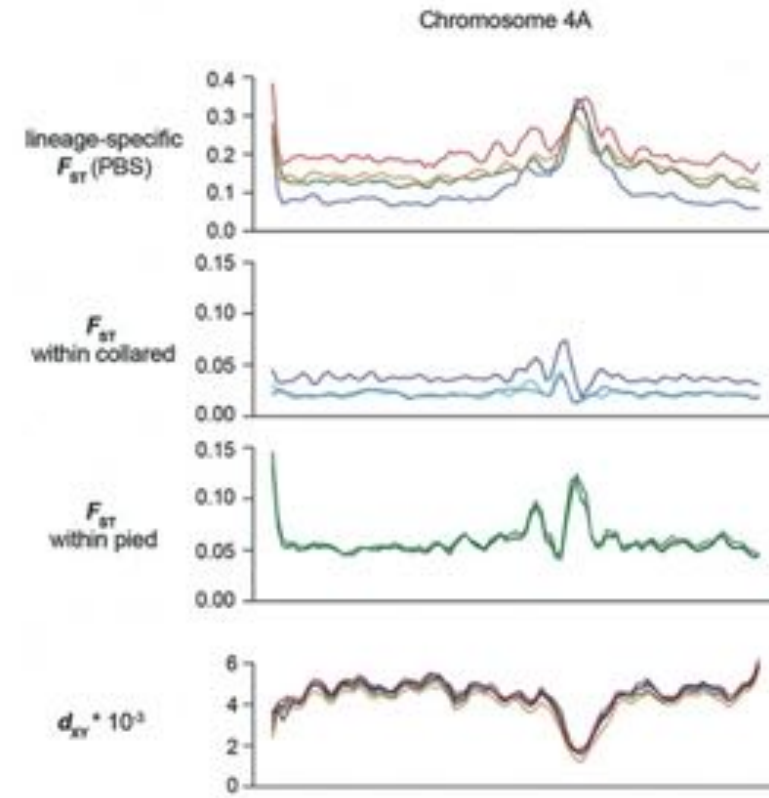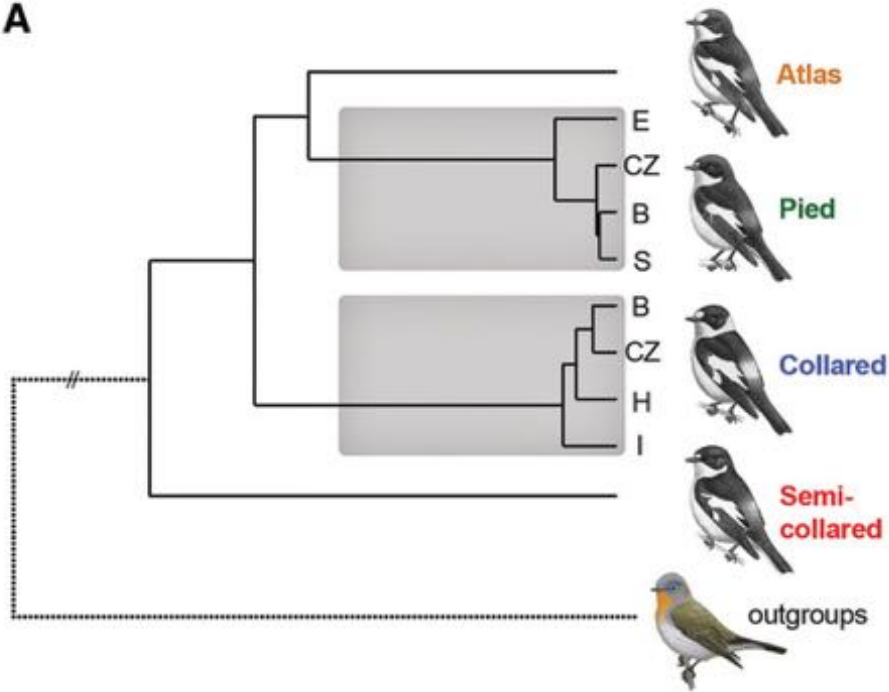# A comparison of genomic islands of differentiation across three young avian species pairs

Darren E. Irwin[1] | Borja Milá[2] | David P. L. Toews[1,3] | Alan Brelsford[1,4] | Haley L. Kenyon[1,5] | Alison N. Porter[1] | Christine Grossen[1,6] | Kira E. Delmore[1,7] | Miguel Alcaide[1,8] | Jessica H. Irwin[1]

Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers

Reto Burri,[1] Alexander Nater,[1] Takeshi Kawakami,[1] Carina F. Mugal,[1] Pall I. Olason,[2] Linnea Smeds,[1] Alexander Suh,[1] Ludovic Dutoit,[1] Stanislav Bureš,[3] Laszlo Z. Garamszegi,[4] Silje Hogner,[5,6] Juan Moreno,[7] Anna Qvarnström,[8] Milan Ružić,[9] Stein-Are Sæther,[5,10] Glenn-Peter Sætre,[5] Janos Török,[11] and Hans Ellegren[1]
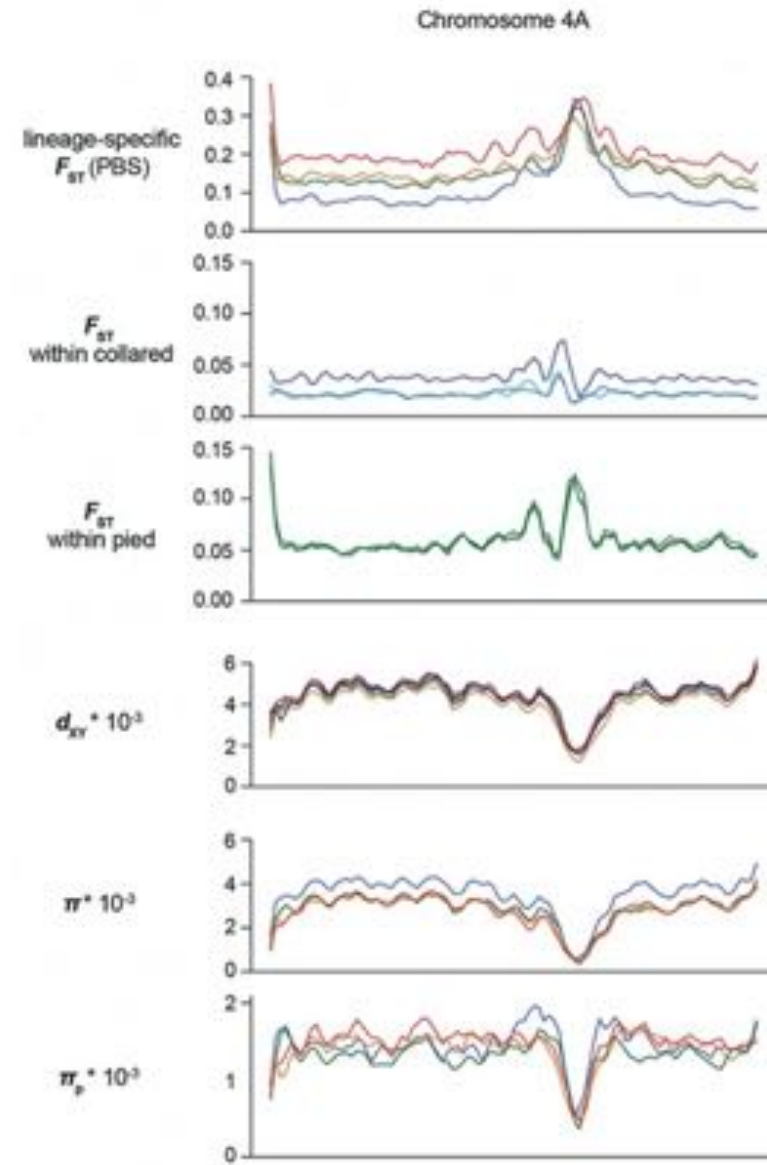
A

Does this mean that all these species/populations comparisons share the same 'speciation genes"?

Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers

Reto Burri,[1] Alexander Nater,[1] Takeshi Kawakami,[1] Carina F. Mugal,[1] Pall I. Olason,[2] Linnea Smeds,[1] Alexander Suh,[1] Ludovic Dutoit,[1] Stanislav Bureš,[3] Laszlo Z. Garamszegi,[4] Silje Hogner,[5,6] Juan Moreno,[7] Anna Qvarnström,[8] Milan Ružić,[9] Stein-Are Sæther,[5,10] Glenn-Peter Sætre,[5] Janos Török,[11] and Hans Ellegren[1]
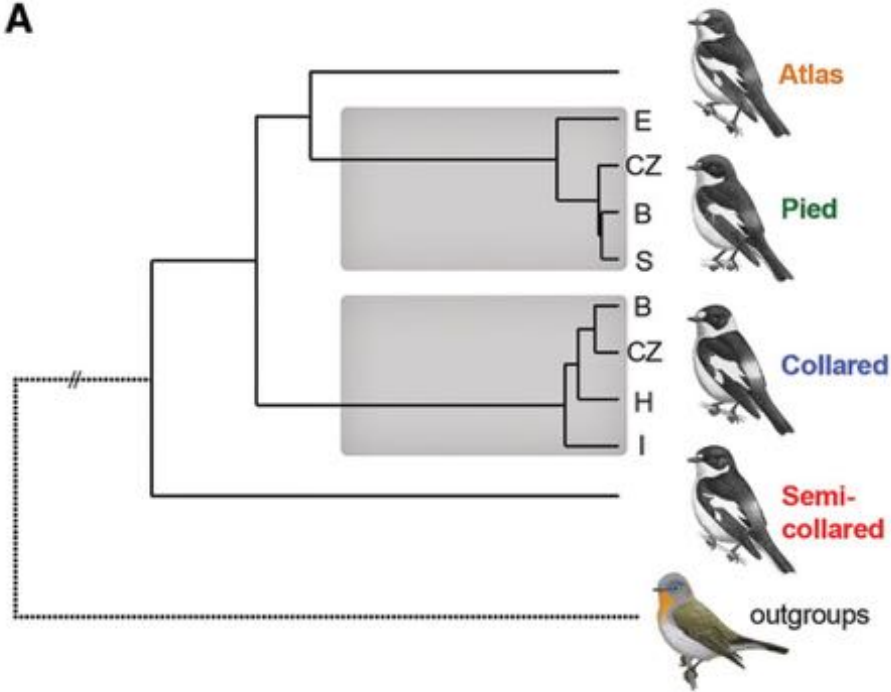
Chromosome 4A

But sequence divergence is actually reduced here…

Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers

Reto Burri,[1] Alexander Nater,[1] Takeshi Kawakami,[1] Carina F. Mugal,[1] Pall I. Olason,[2] Linnea Smeds,[1] Alexander Suh,[1] Ludovic Dutoit,[1] Stanislav Bureš,[3] Laszlo Z. Garamszegi,[4] Silje Hogner,[5,6] Juan Moreno,[7] Anna Qvarnström,[8] Milan Ružić,[9] Stein-Are Sæther,[5,10] Glenn-Peter Sætre,[5] Janos Török,[11] and Hans Ellegren[1]
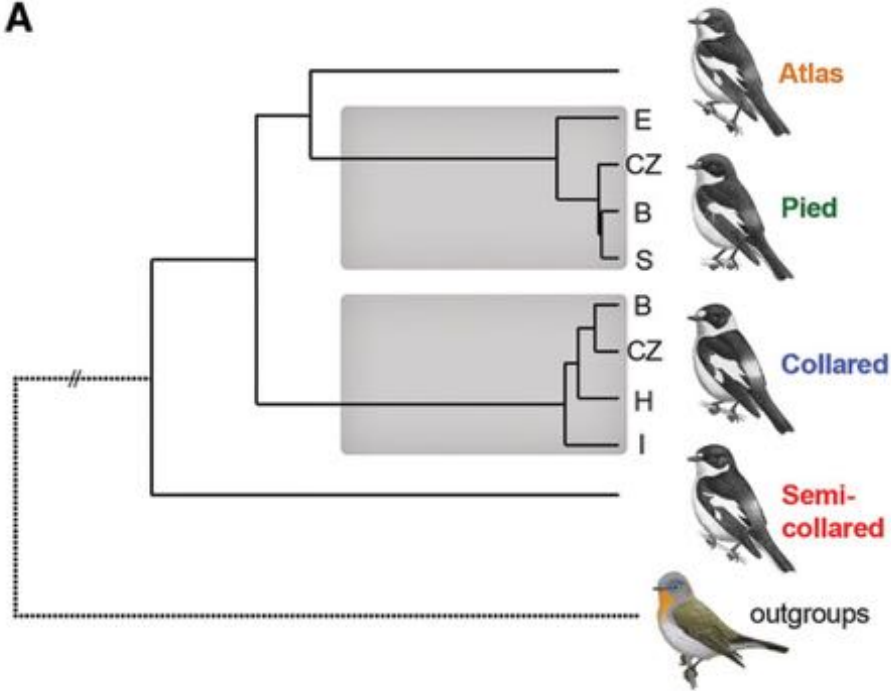
And nucleotide diversity, too.

Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers

Reto Burri,[1] Alexander Nater,[1] Takeshi Kawakami,[1] Carina F. Mugal,[1] Pall I. Olason,[2] Linnea Smeds,[1] Alexander Suh,[1] Ludovic Dutoit,[1] Stanislav Bureš,[3] Laszlo Z. Garamszegi,[4] Silje Hogner,[5,6] Juan Moreno,[7] Anna Qvarnström,[8] Milan Ružić,[9] Stein-Are Sæther,[5,10] Glenn-Peter Sætre,[5] Janos Török,[11] and Hans Ellegren[1]
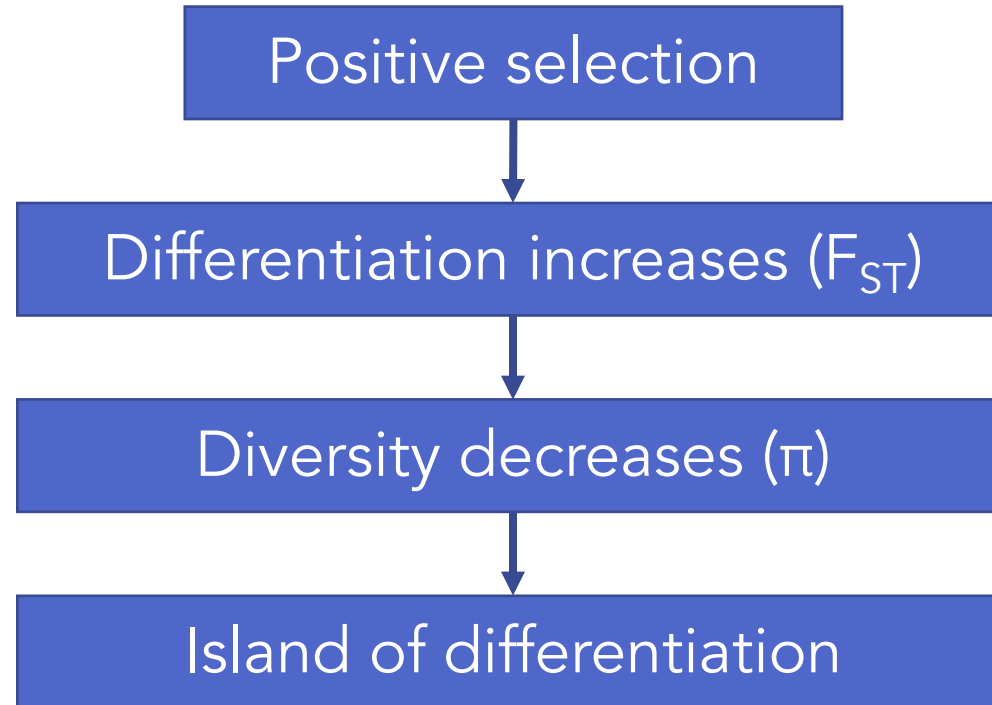
Recombination is reduced in the same area.
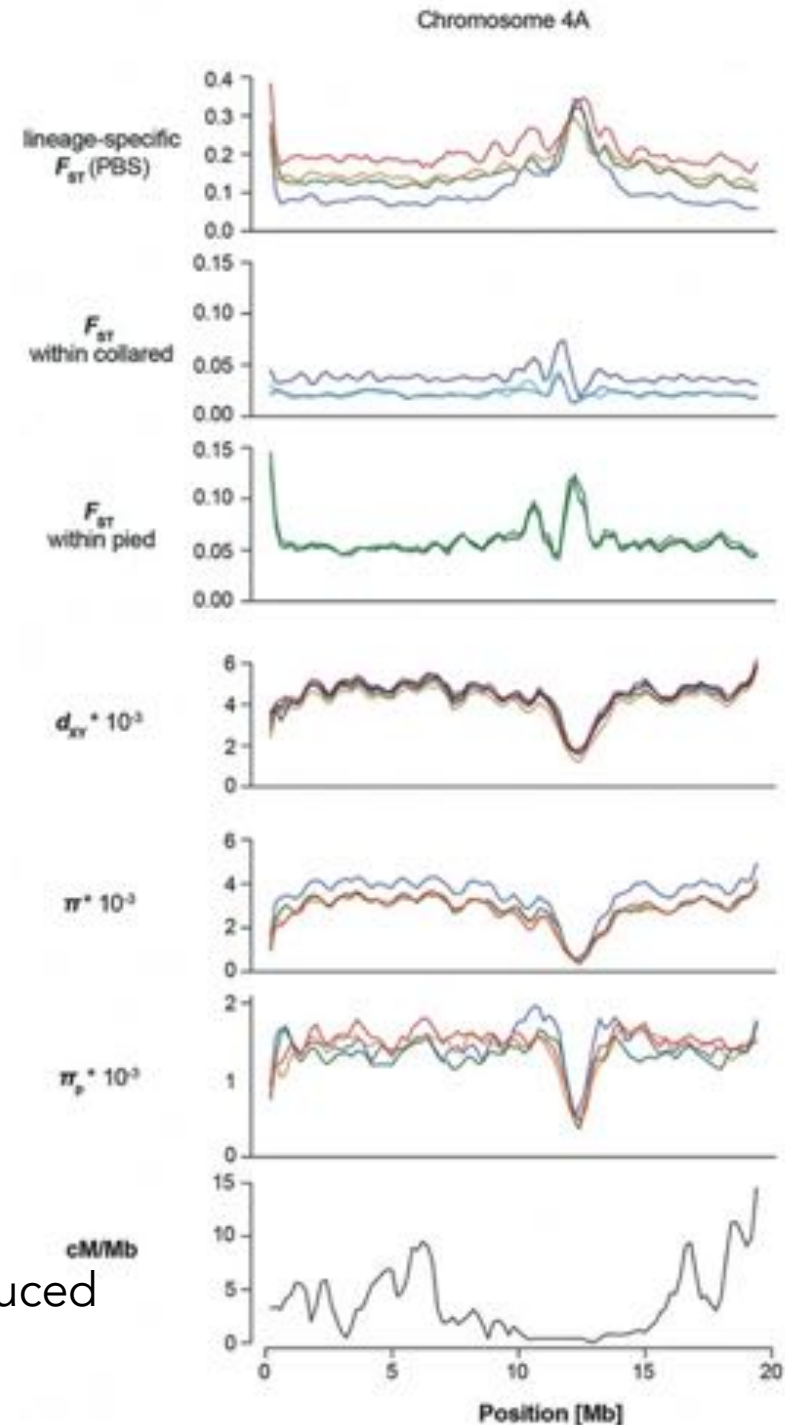
Adaptation/speciation with gene flow

Positive selection

↓

Differentiation increases (F_{ST})

↓

Diversity decreases (π)

↓

Island of differentiation

Or background selection without gene flow?

Positive selection?

Differentiation increases (F_{ST})

Diversity decreases (π)

Island of differentiation

Low recombination reduces diversity:
1. Recombination introduces diversity
2. Recombination reduces the diversity-reducing effect of positive and background selection

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

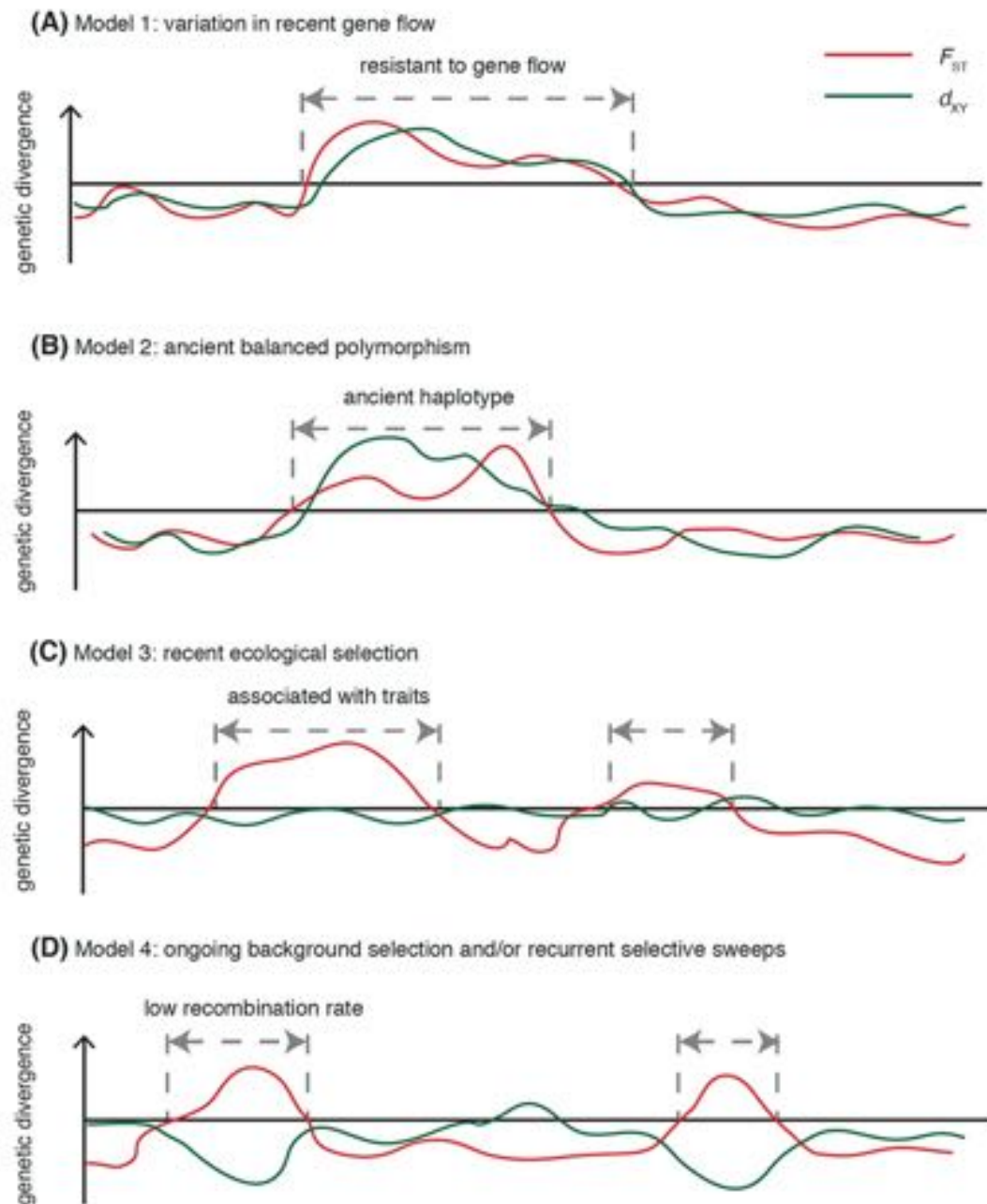F_{ST} is affected by levels of standing genetic variation:
- It is inflated in areas of low diversity
- It is low when diversity is high

Absolute measures of differentiation

D_{xy} is the average number of pairwise differences between sequences from two populations, excluding all comparisons between sequences within populations.

It is independent of the levels of diversity within the two populations examined.

However, it is affected by ancestral levels of diversity and mutation rate.
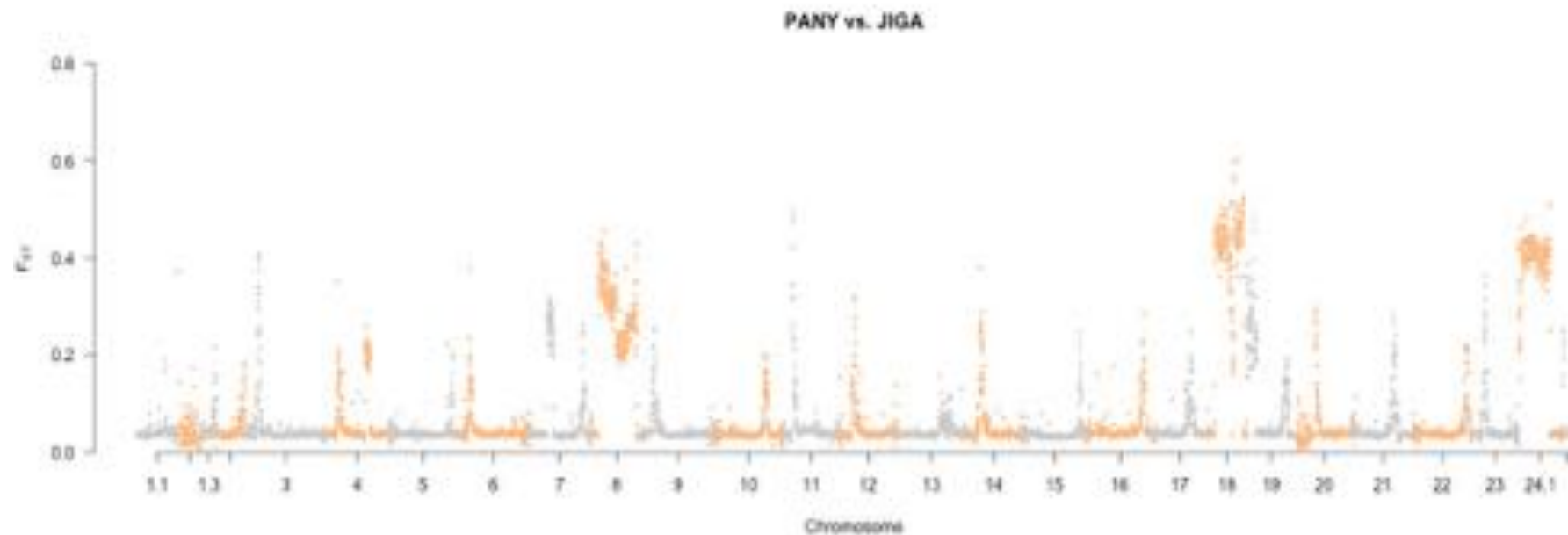
(A) Model 1: variation in recent gene flow
genetic divergence
resistant to gene flow
F_{ST}
d_{XY}

(B) Model 2: ancient balanced polymorphism
genetic divergence
ancient haplotype

(C) Model 3: recent ecological selection
genetic divergence
associated with traits

(D) Model 4: ongoing background selection and/or recurrent selective sweeps
genetic divergence
low recombination rate

Han et al. 2017, Genome Research

Ideally, knowing recombination rates across the genome would help greatly to disentangle different scenarios.

The joint assessment of different estimators/statistics can help us understand drivers of apparent differentiation.

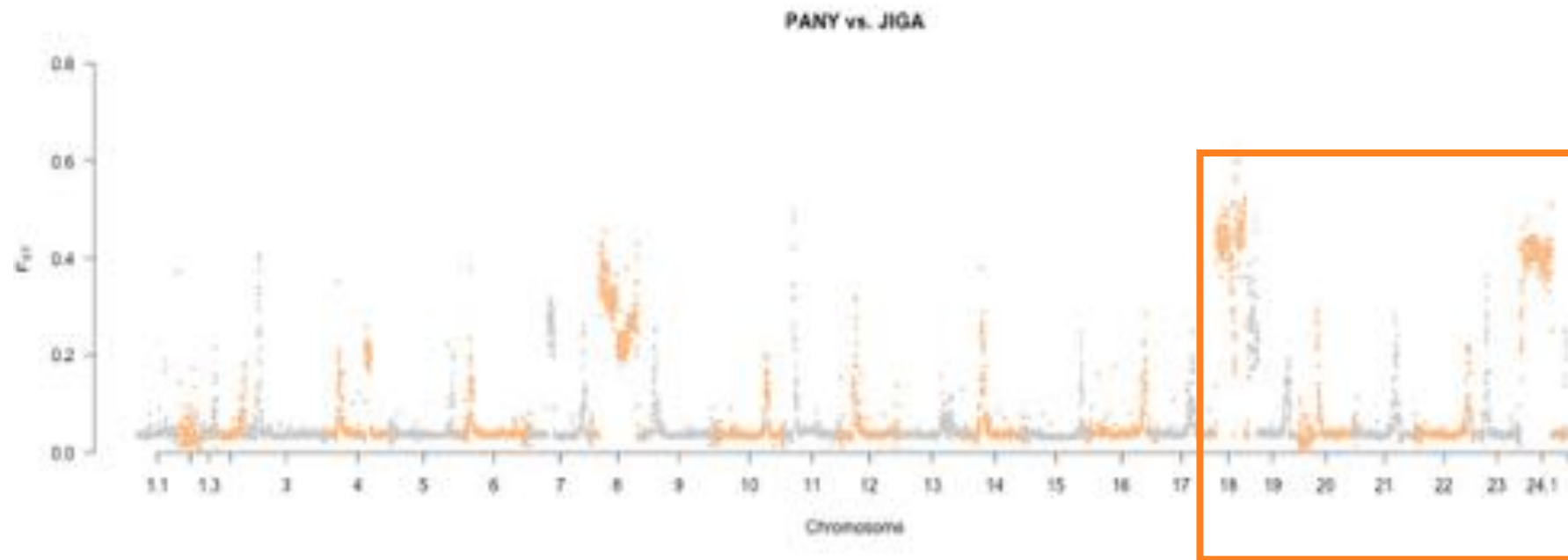PANY

JIGA
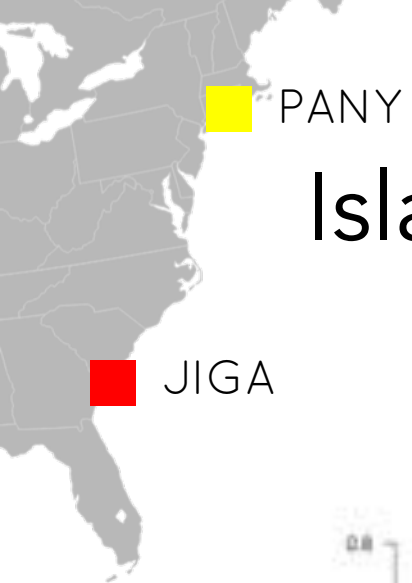
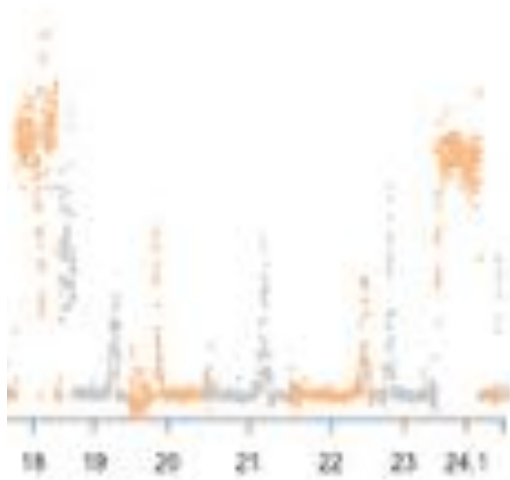# Islands of divergence in the Atlantic silverside

Relative differentiation - F_{ST}

PANY vs. JIGA

Tigano et al., in prep.

Islands of divergence in the Atlantic silverside

PANY

JIGA

Relative differentiation - F_{ST}

PANY vs. JIGA

Tigano et al., in prep.

Differentiation

Diversity

F_{ST}

π

D_{xy}

Tajima's D

Tigano et al., in prep.

Differentiation

Diversity

F_{ST}

Inversions

π
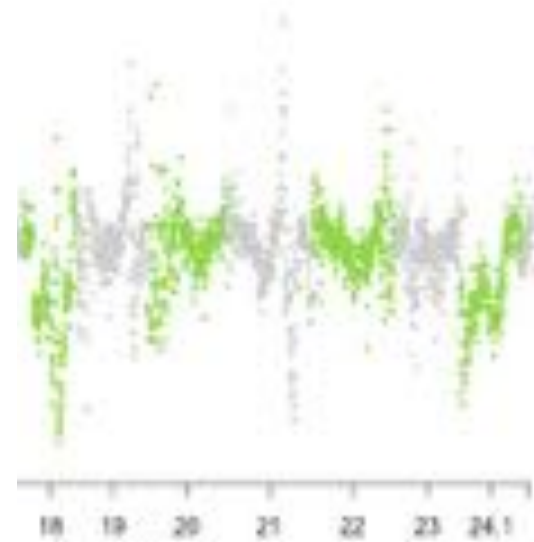
D_{xy}

Tajima's D

Differentiation

Diversity

F_{ST}

Centromeres?

π

D_{xy}

Tajima's D

Tigano et al., in prep.
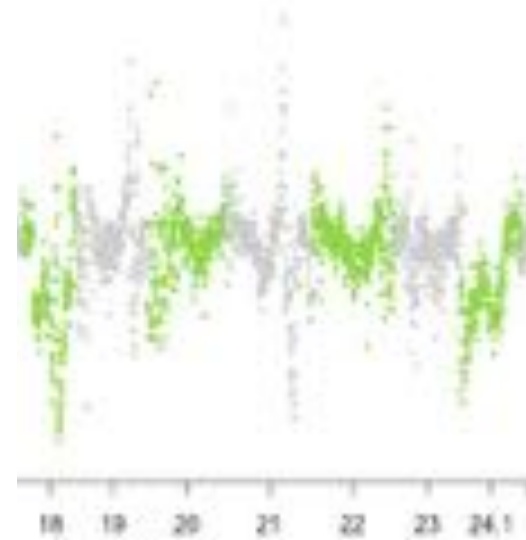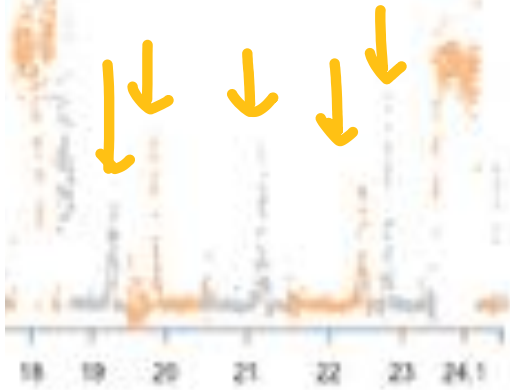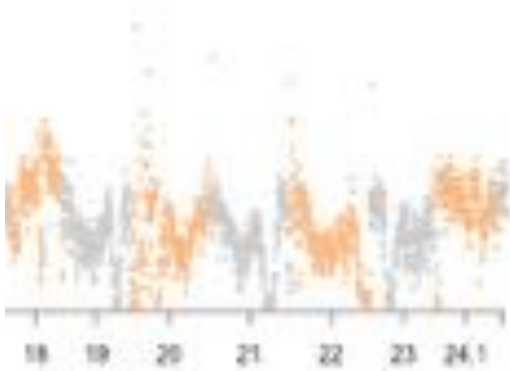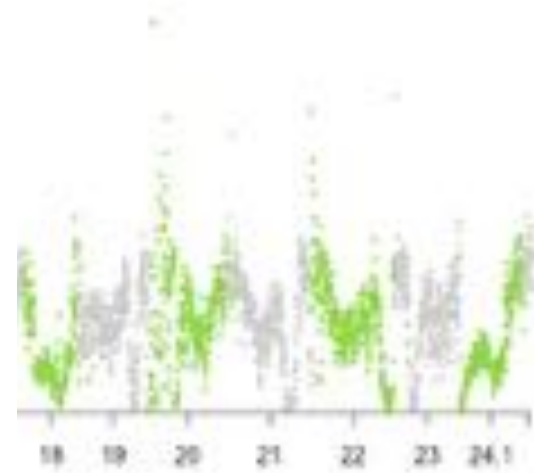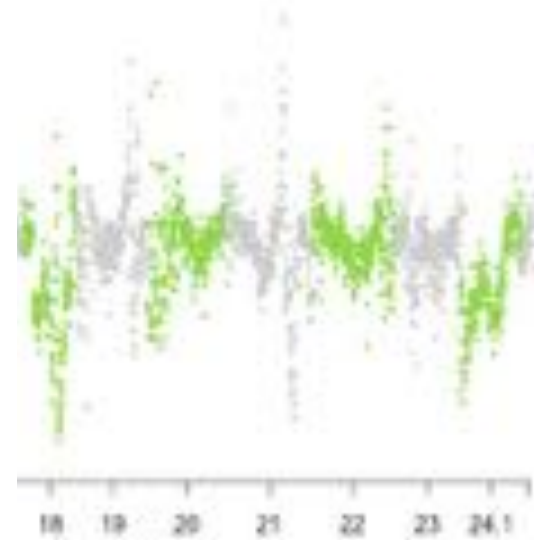
Structural variation

Though we have evidence from clinal analyses and correspondence between phenotypic differences and frequency of those inversions, SVs (including inversions) can be maintained as polymorphism within populations or species by neutral processes.

Structural variation

How structural variation can confound signatures of selection.

- Recombination suppression → elevated differentiation due to low recombination rather than selection

Structural variation

How structural variation can confound signatures of selection.

- Recombination suppression → elevated differentiation due to low recombination rather than selection

- Collapse of paralog sequences (Copy Number Variants) → detection of excess of polymorphism and heterozygosity

Structural variation

How structural variation can confound signatures of selection.

- Recombination suppression → elevated differentiation due to low recombination rather than selection

- Collapse of paralog sequences (Copy Number Variants) → detection of excess of polymorphism and heterozygosity

- Important variants could get filtered out → low quality assembly and mapping
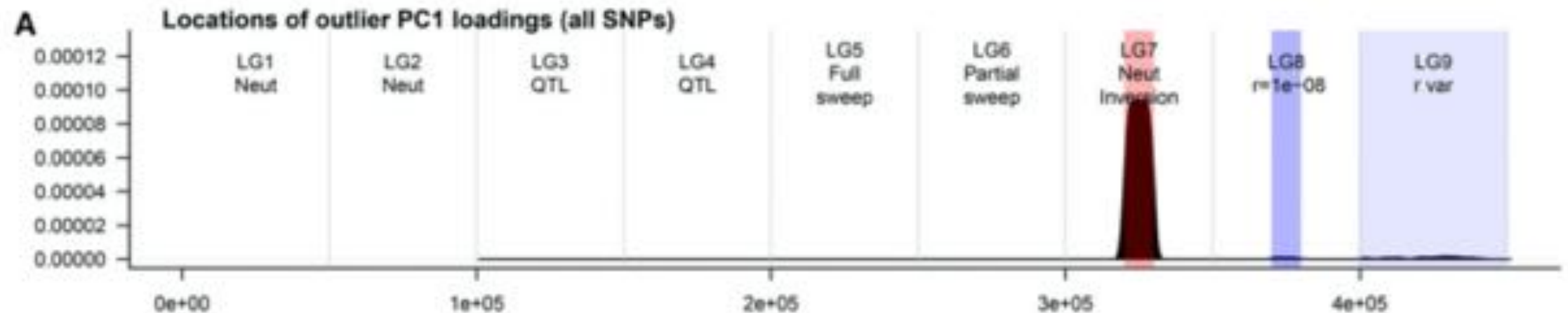
Structural variation

How structural variation can confound signatures of selection.

- Recombination suppression → elevated differentiation due to low recombination rather than selection

- Collapse of paralog sequences (Copy Number Variants) → detection of excess of polymorphism and heterozygosity

- Important variants could get filtered out → low quality assembly and mapping

- Missing heritability → if only sequence variation is screened but the variant associated with phenotype/environment in a SV

Structural variation

How structural variation can confound signatures of selection.

- Neutral demographic model for outlier methods

Lotterhos 2019, G3

The end!