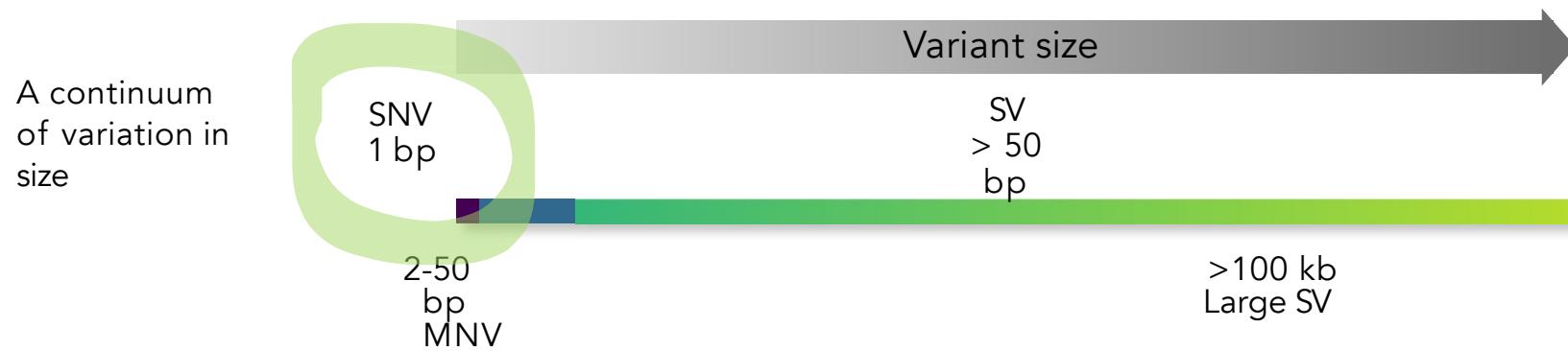


Structural Variants and adaptation

Day 4 - Lecture 1

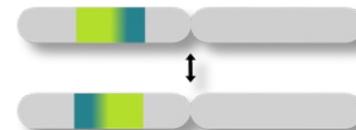
(adapted from Claire Mérot & Anna Tigano's slides)



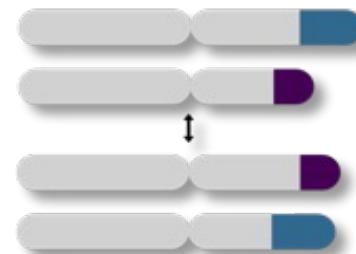
STRUCTURAL VARIANT
all changes in position or direction,
as well as gains or losses of sequence

Balanced SVs
(change in position or order)

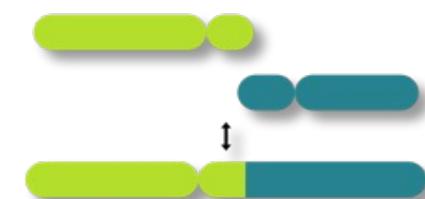
Inversion



Translocation

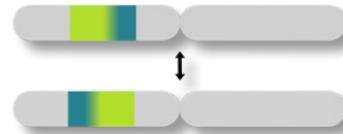


Fusion/fission

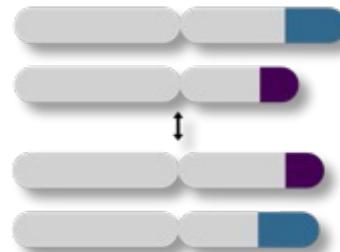


Balanced SVs
(change in position or order)

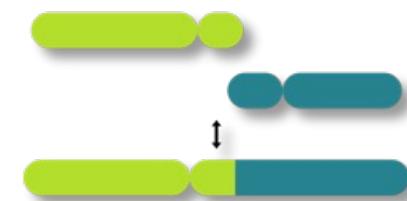
Inversion



Translocation

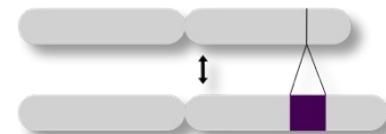


Fusion/fission

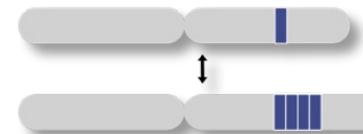


Unbalanced SVs
(gain/loss of DNA)

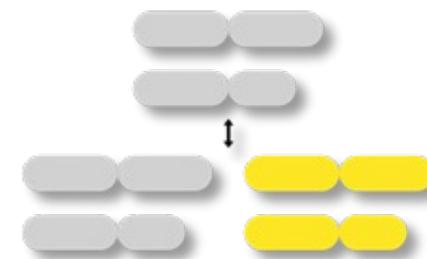
CNV:
Insertion/deletion



CNV:
duplications, tandem repeats

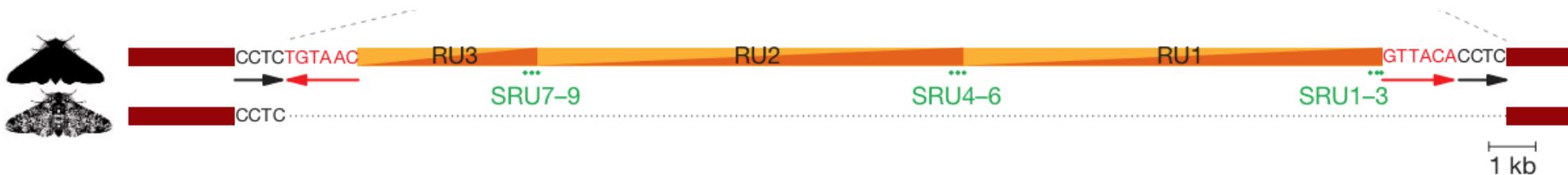


Polyplody



Why do we study them?

SVs are the genetic basis of several traits.



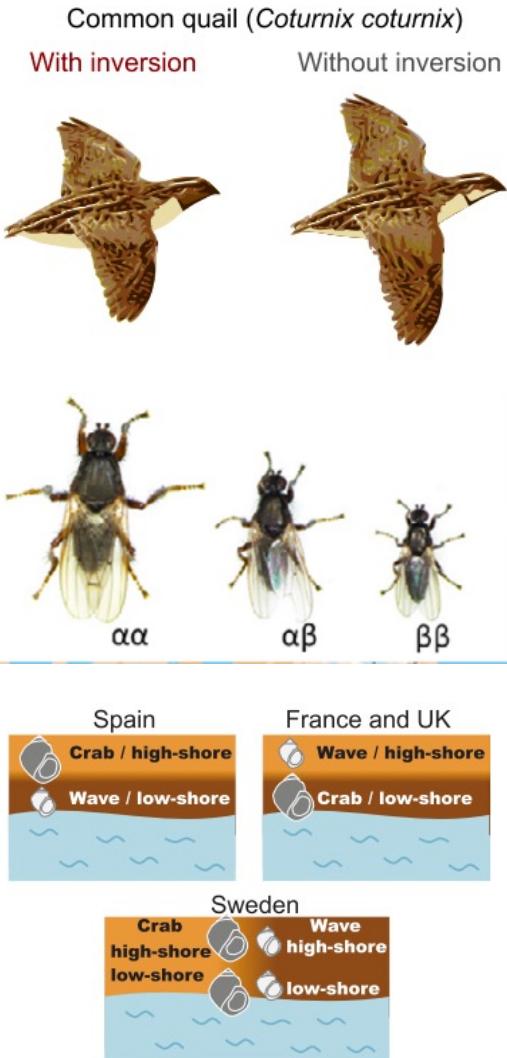
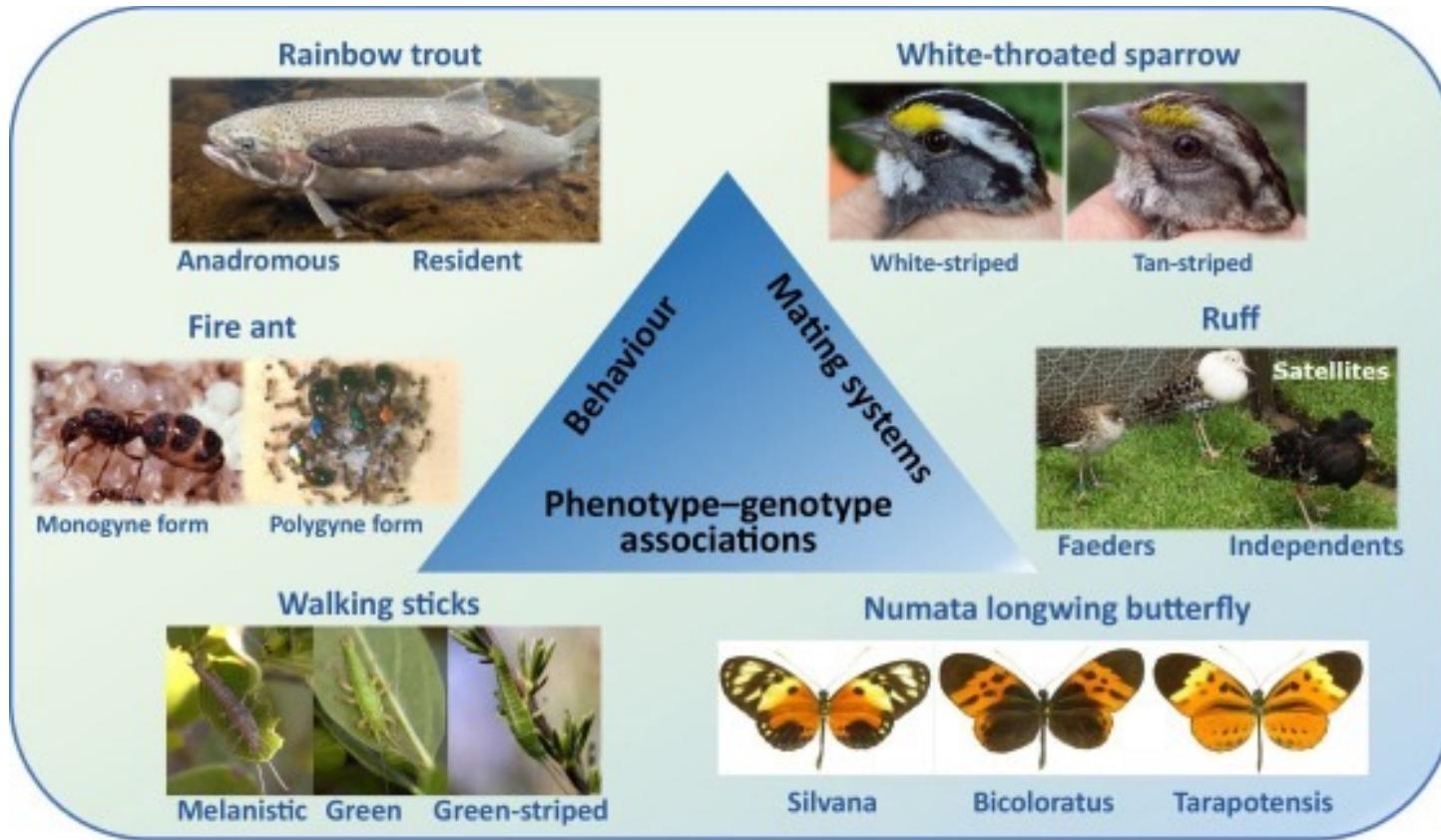
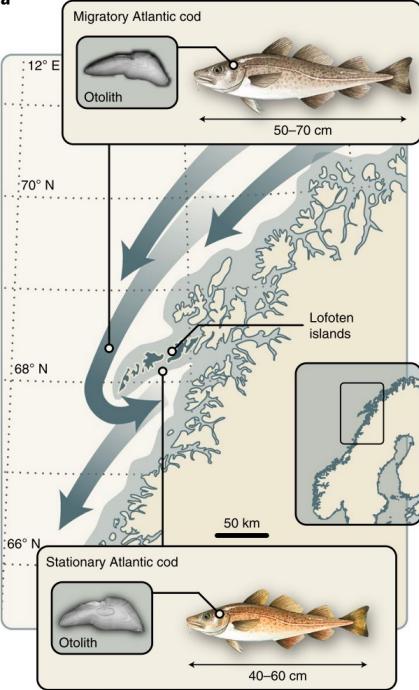
Van't Hof et al. 2016, Nature

'Missing heritability problem'

When sequence variation explains only for a portion of the variation in a given trait (or disease)

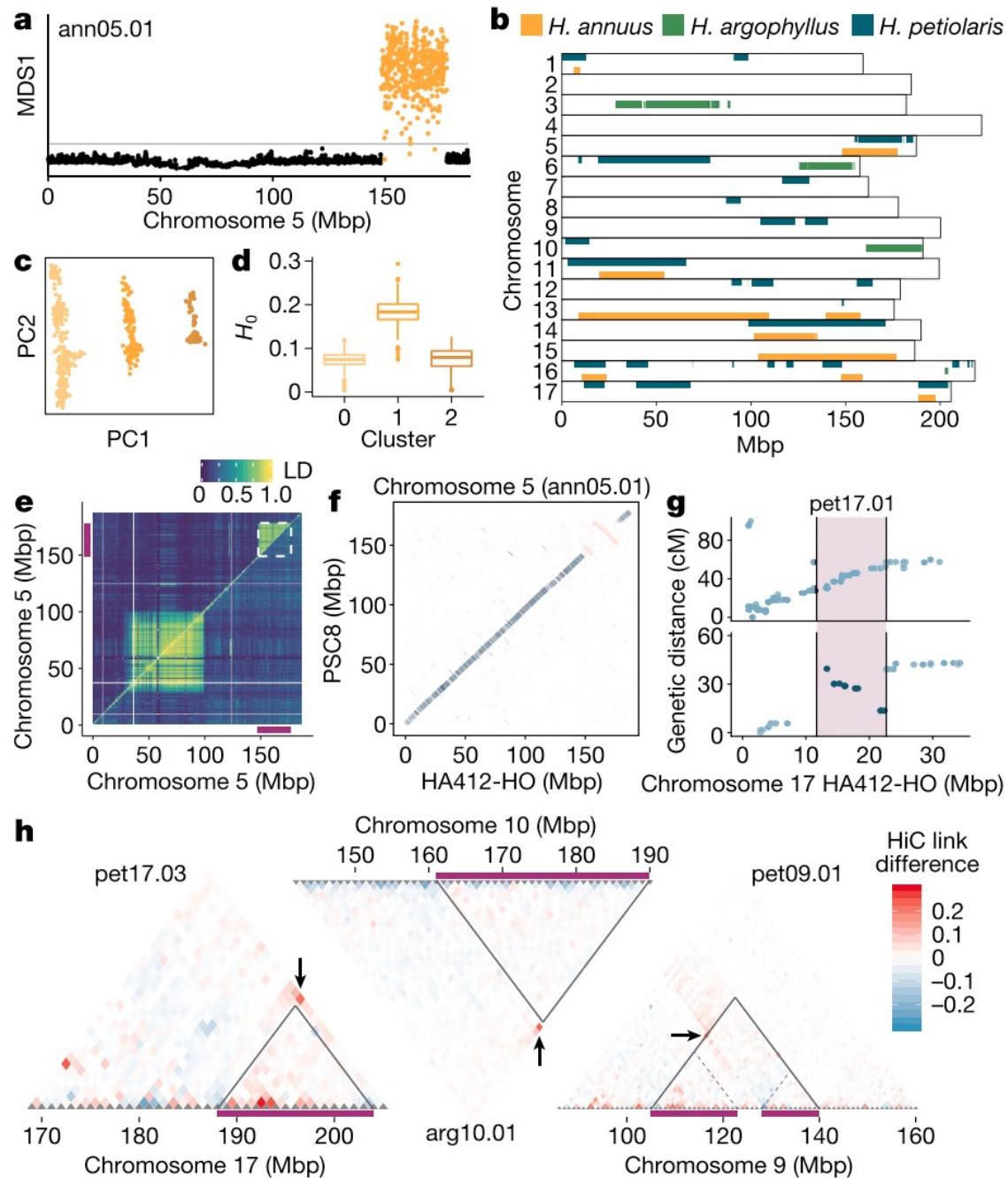
Inversions are the SV most commonly associated with adaptive differences

a



Maren Wellenreuther and Louis Bernatchez, TREE, 2018
 Sanchez-Donoso et al., Curr Bio, 2022
 Matschiner, et al., Nat. Ecol. Evol. 2022

Twyford & Friedman, Evolution, 2015
 Berdan et al., Evol Let, 2021
 Morales et al., Sci Adv, 2019



Sunflowers

Massive haplotypes underlie ecotypic differentiation in sunflowers

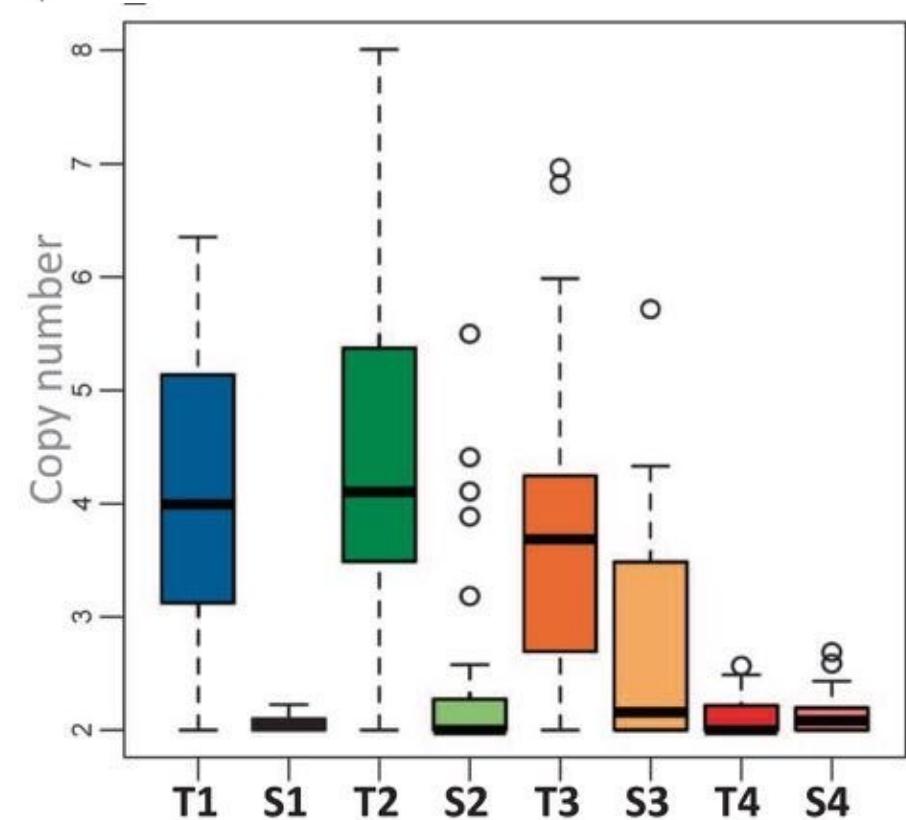
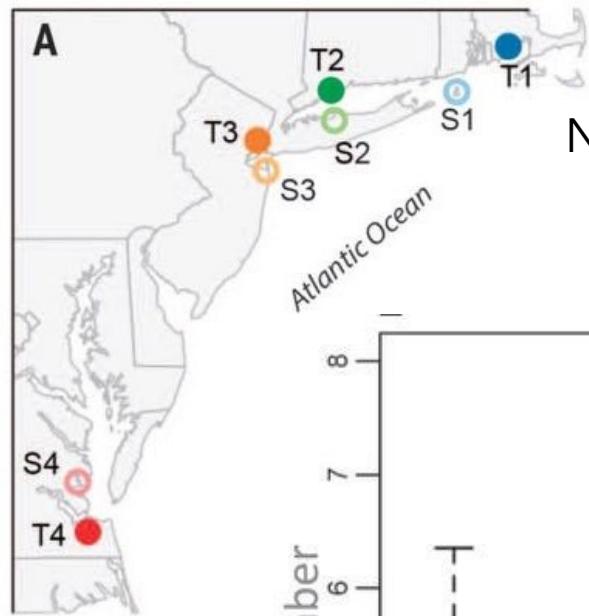
Copy Number Variants



Atlantic killifish

Copy Number Variants (CNVs)
associated with rapid adaptation
to chemical pollution

Reid et al. 2016, Science

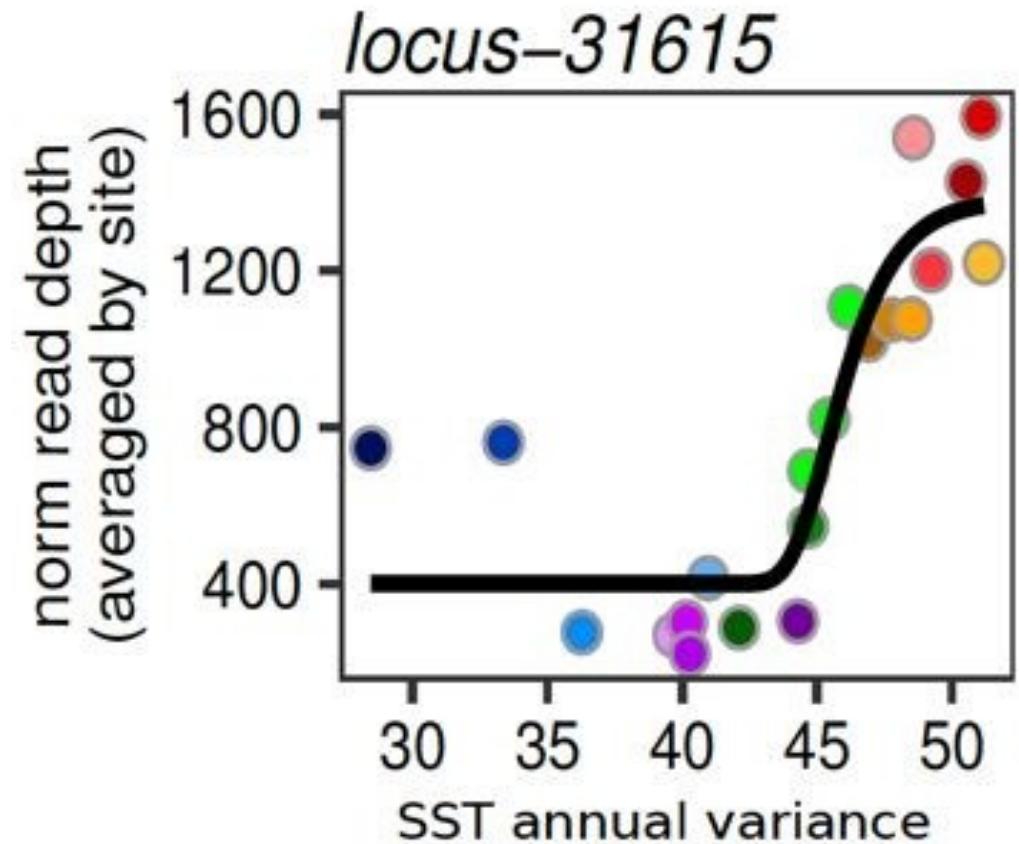


Copy Number Variants



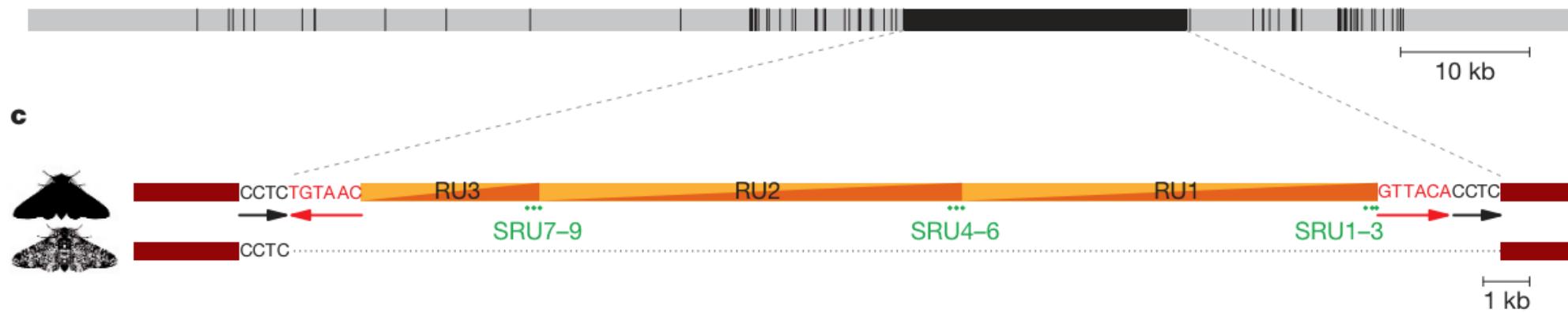
Lobster

Variation in copy number associated with environmental variance



Transposable elements

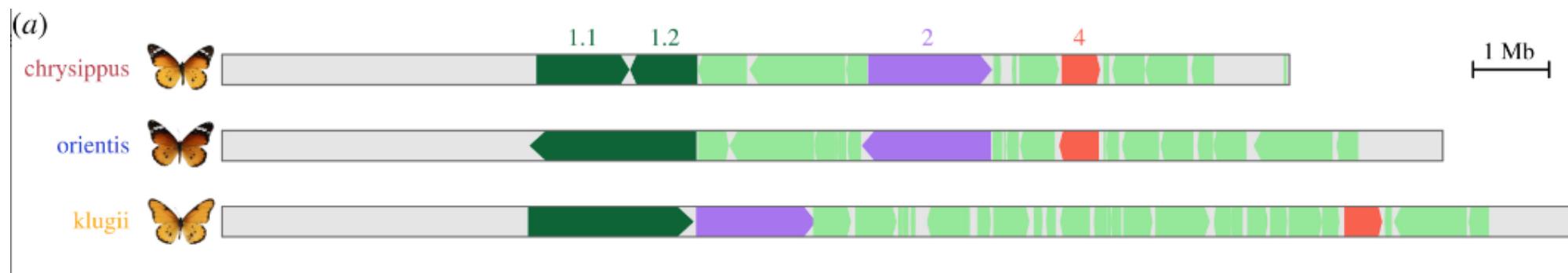
Insertion of a **transposon** associated with melanism in peppered moth



Duplications and inversions

Successive duplications and inversions distinguish color morphs of African monarch butterfly

Supergene contains color patterning loci



How do SVs affect phenotypes?

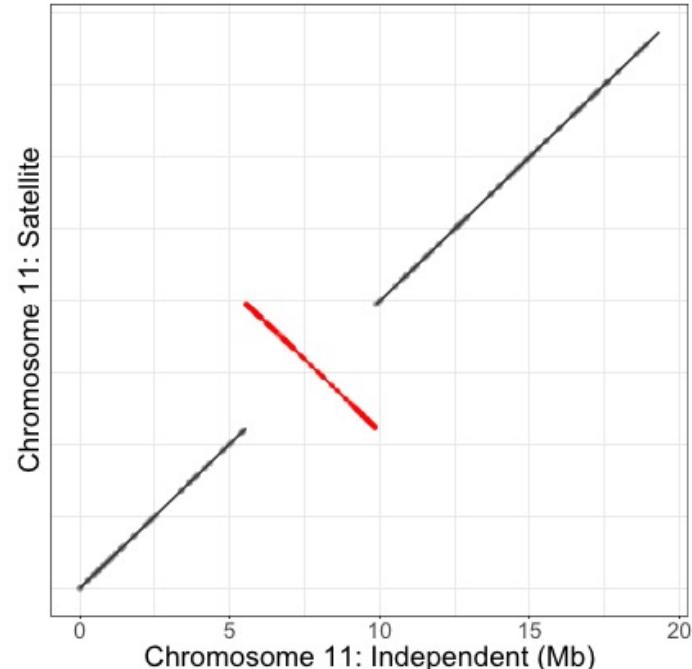
(a) Independent



Faeder



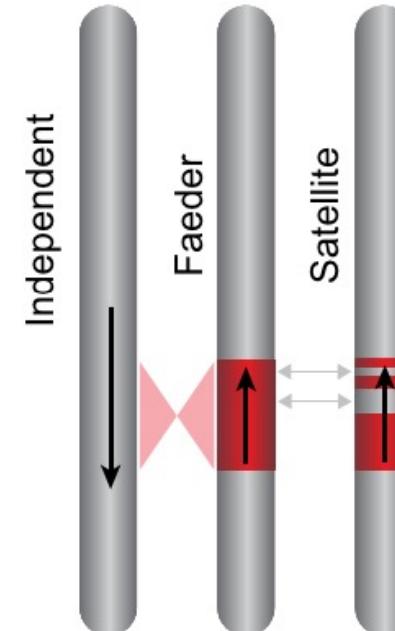
Satellite



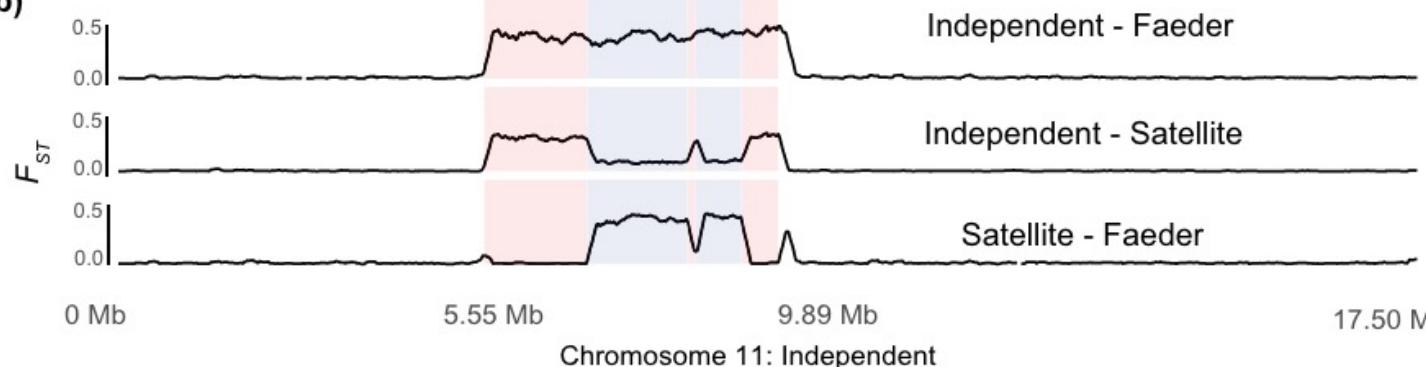
Independent

Faeder

Satellite



(b)



Alternative male mating
strategies in ruff

Supergene contains 100 genes

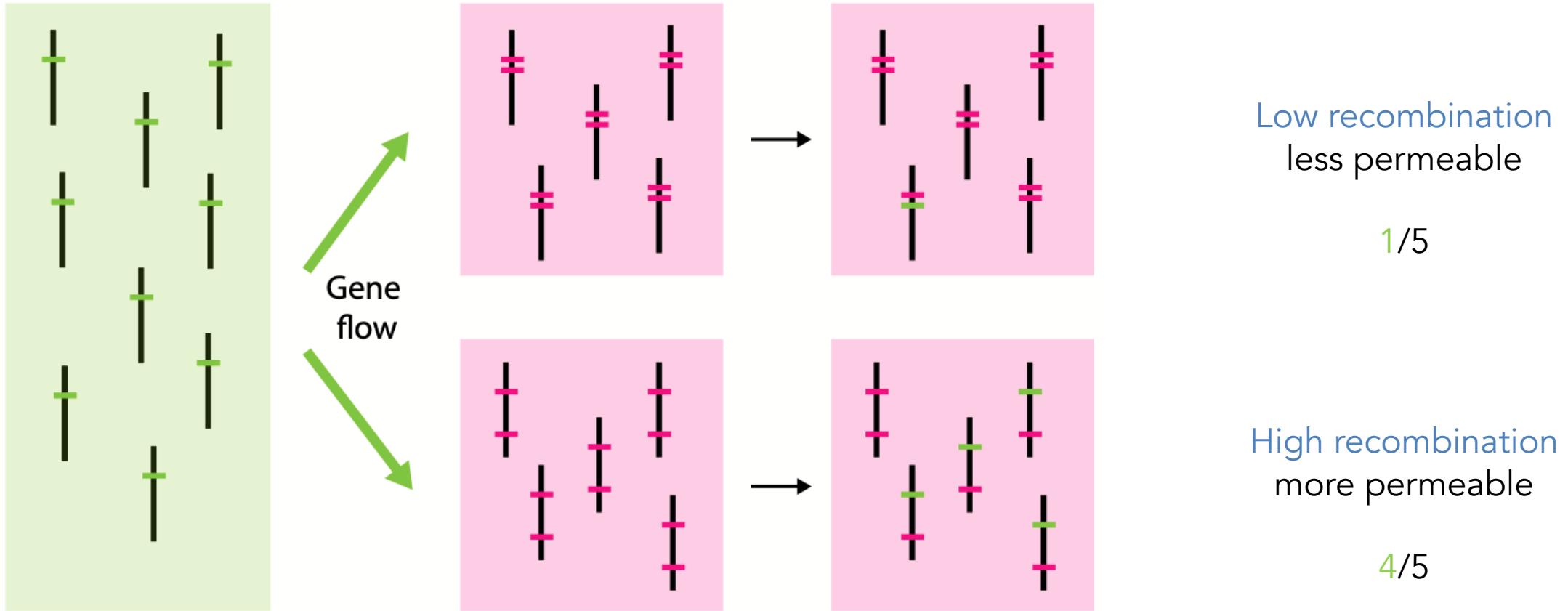
- MC1R (pigmentation gene)

High F_{st} in inversion region

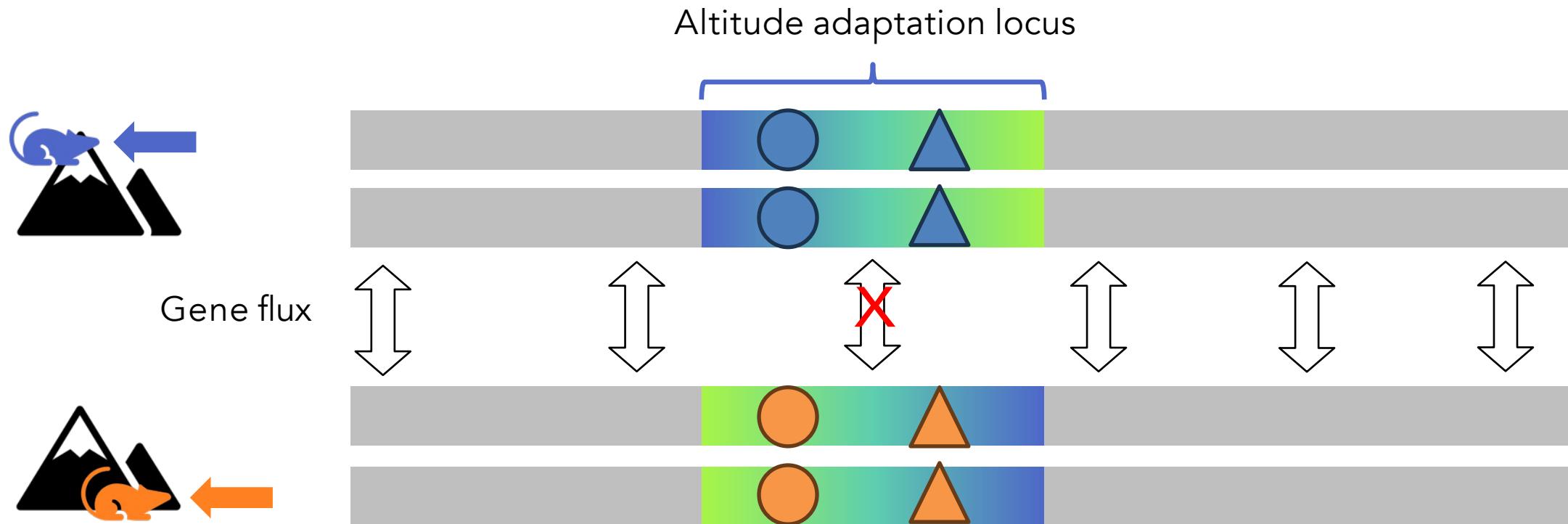
Structural variants that suppress recombination, like inversions, maintain complexes of co-adapted alleles in the face of gene flow.

Supergenes!

Tight genetic architectures are more resistant to gene flow

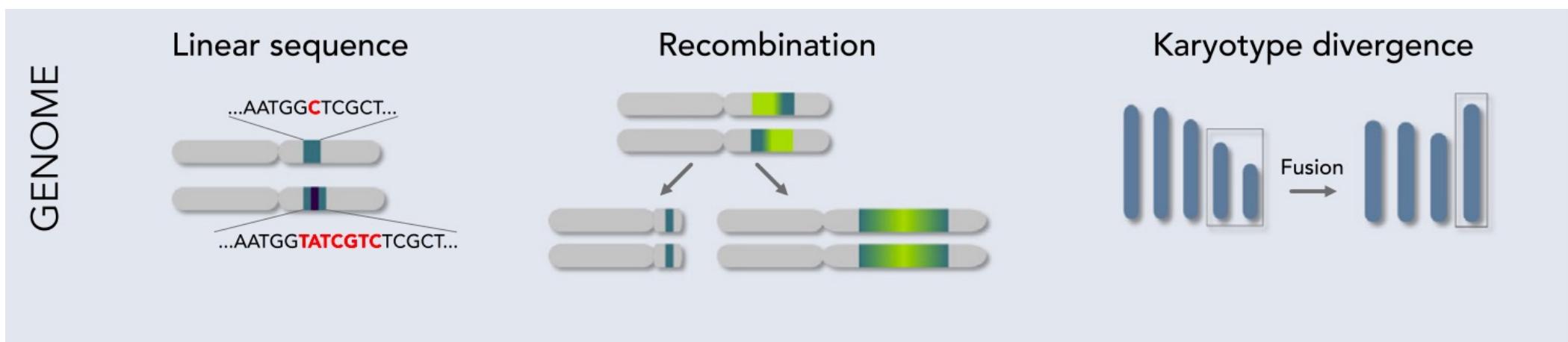


Structural variants (inversions, indels) reduce recombination
maintain **combinations** of adaptive alleles in the face of **gene flow**

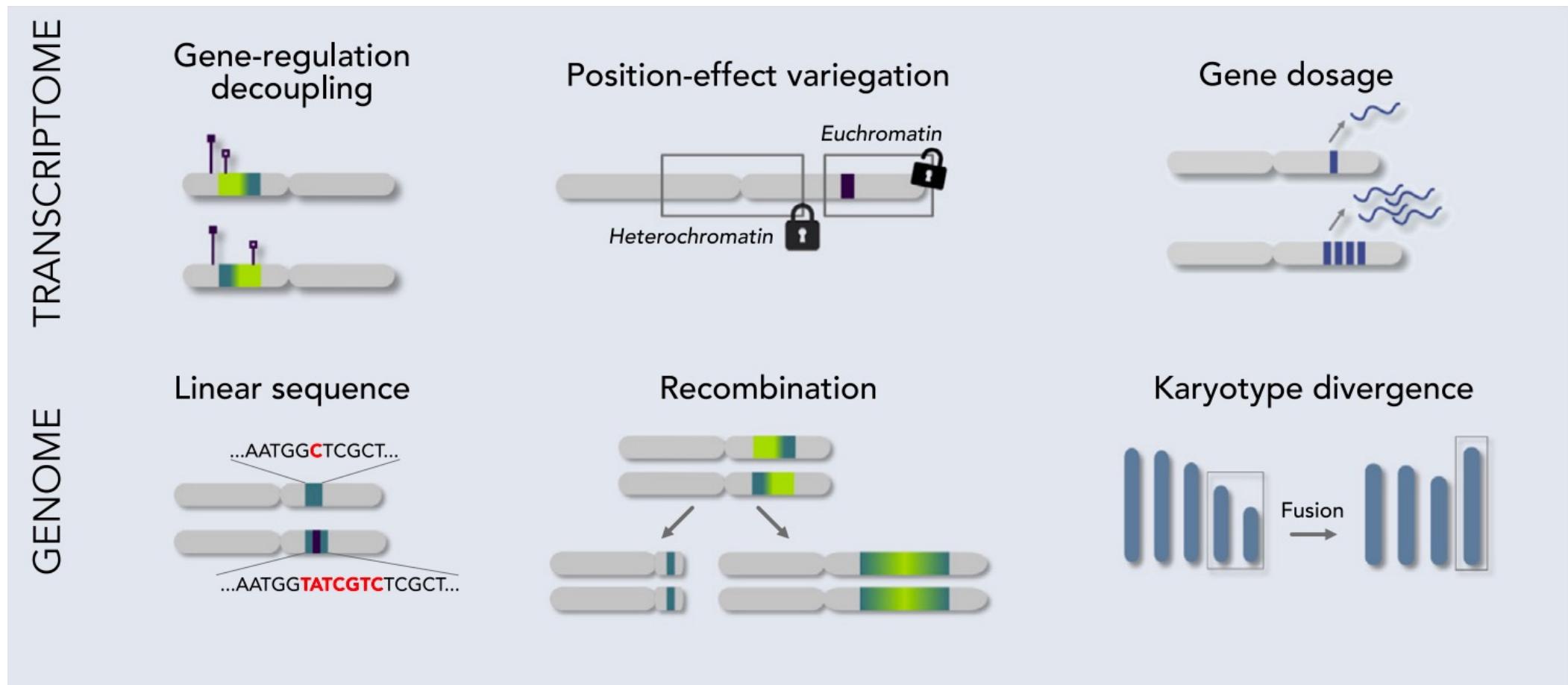


How do SVs affect phenotypes?

Recombination suppression is only one of many ways



How do SVs affect phenotypes?



Identification and characterization of SVs

Indirect methods

Patterns of differentiation and LD

Direct methods

Explicit mining of sequencing data

Objectives

Develop best practices to allow synthesis and comparison among studies

Assess sequencing/bioinformatic tools to select appropriate approaches depending on available resources

Account for population/species standing variation by enabling detection in large datasets

Understanding evolution and function of SVs

Population genomics

SV distribution & frequency

Differences among populations/species

Associations with environment and/or phenotype

Evolutionary simulations

Simulate effects on genome evolution

Population genetic simulations of SVs distribution and evolution

Experimental validation

Common garden and reciprocal transplant experiments

Experimental evolution

Differential gene expression

Gene knockdown/editing

Objectives

Acknowledge SVs as standing genetic variation

Describe evolutionary processes

Identify candidate SVs for adaptation & speciation

Objectives

Predict evolutionary trajectory

Disentangle effects of neutral and selective processes on SVs

Objectives

Validate candidate SVs for adaptation

Understand effects on phenotypes/recombination

Comparisons and meta-analyses among studies and across taxa

Comparative genomics - Phylogenomics

Synthesis of similarities and differences among studies and taxa

Objectives

Increase reproducibility and reliability of findings

Characterize SVs (e.g., number, position, size, breakpoints)

Assess frequency & distribution of SVs within and across species

Understand how SVs form, evolve, and persist

Ecological and evolutionary applications

New markers for genetic structure, environmental DNA, ancient DNA

Delineating evolutionarily significant units for conservation and management

Predicting population and species responses to global change

Agriculture and aquaculture breeding program design

We still know very little about SVs and lack of standardized protocols hamper progress.

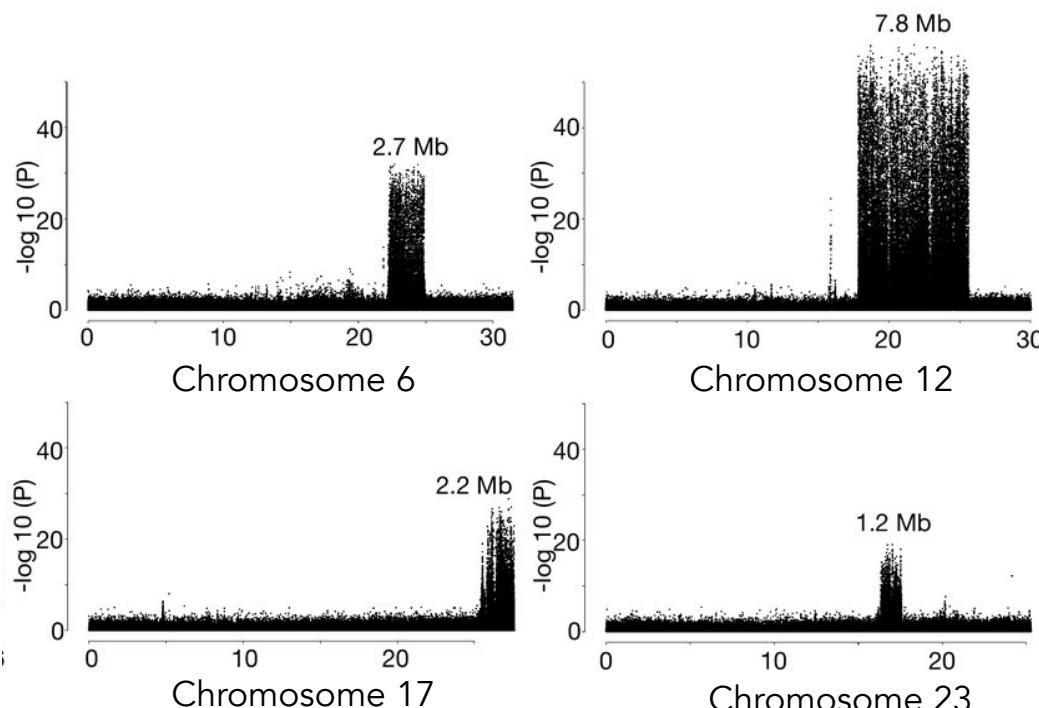
Roadmap to guide future studies focusing on the role of SVs in adaptation and speciation

Indirect methods of detection

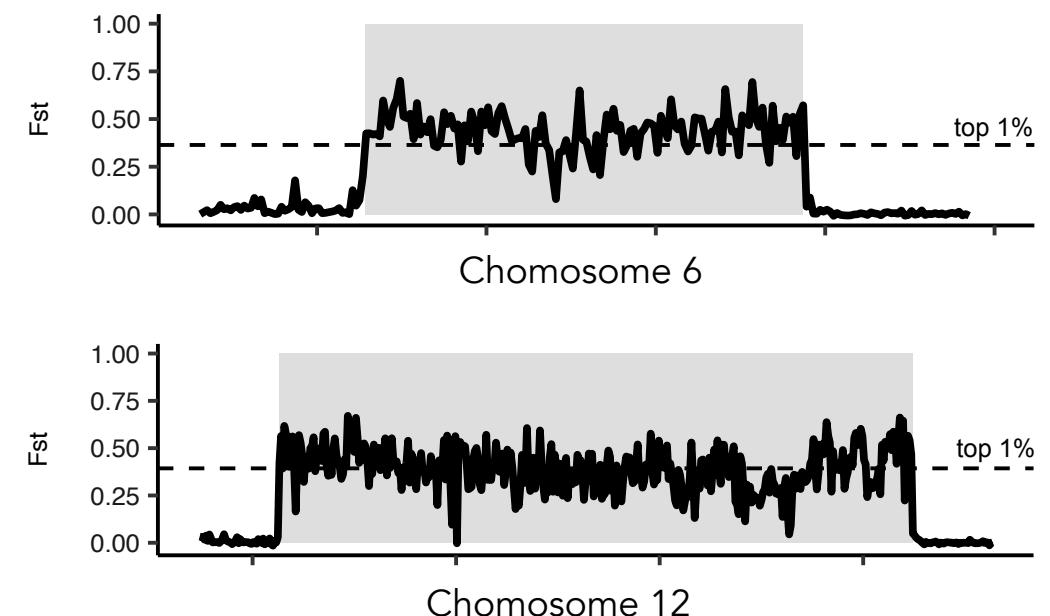
Linked SNPs from RAD-seq, WG-seq

Strong signature of differentiation

Allele frequency differences



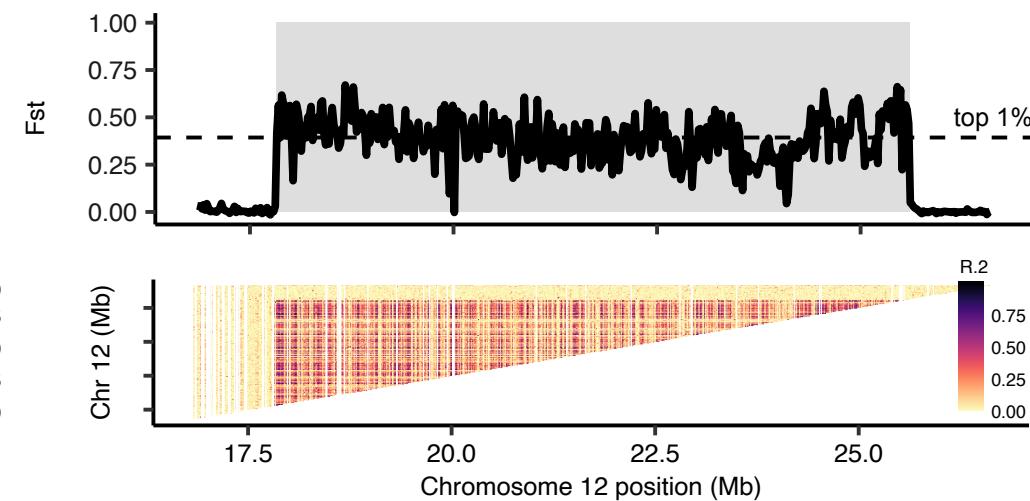
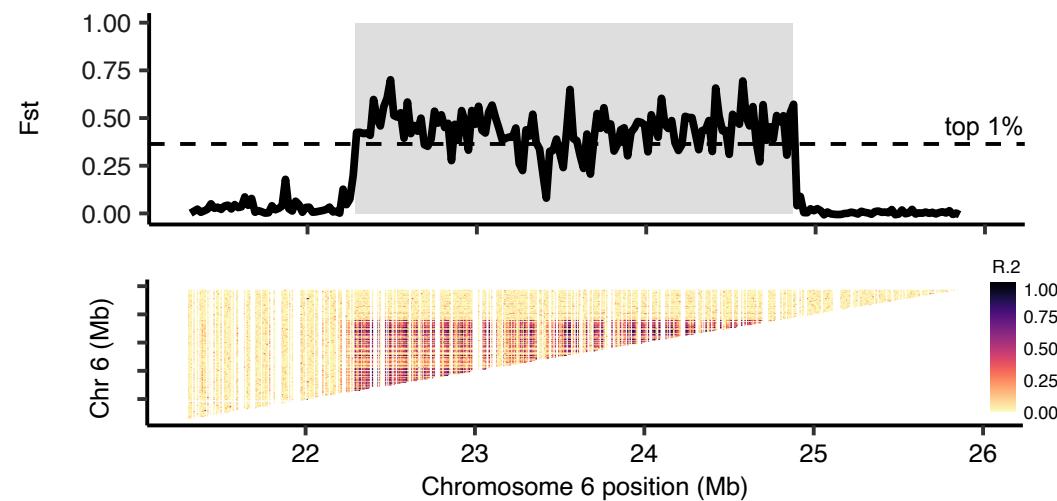
F_{st} differentiation between homozygotes



Indirect methods of detection

Linked SNPs from RAD-seq, WG-seq

Signature of recombination suppression



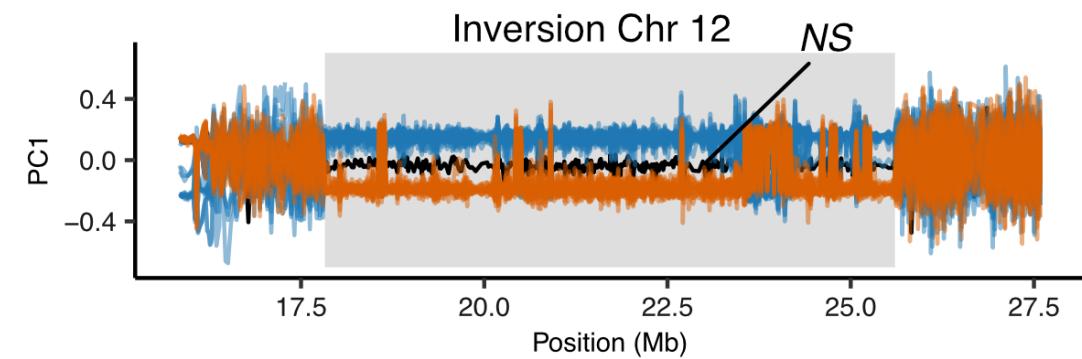
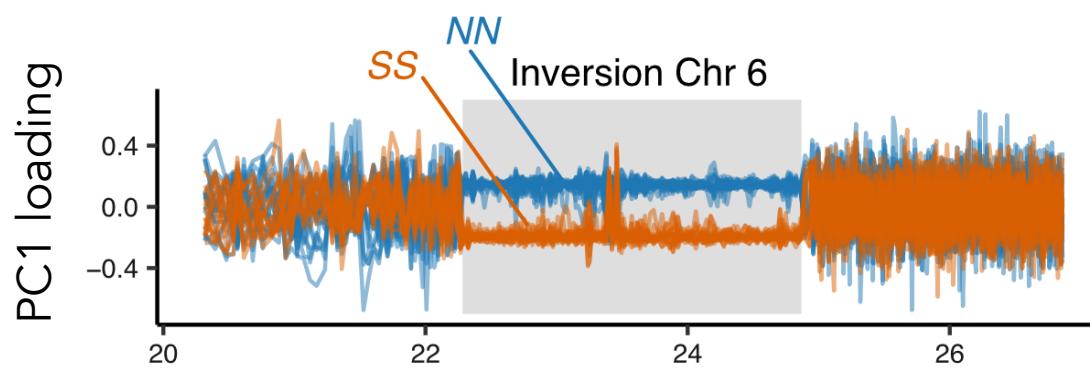
High R^2 among genotypes

Indirect methods of detection

Linked SNPs from RAD-seq, WG-seq

Local PCA

SNPs within structural variant will usually cluster in three groups belonging to three genotype classes



Lostruct Package:

Li & Ralph. 2019 Genetics <https://doi.org/10.1534/genetics.118.301747>

Detection of 7 inversions in *Helianthus* with Rad-seq data:

Huang et al. Mol Ecol. 2020. <https://doi.org/10.1111/mec.15428>

Jamsandekar, Ferreira, et al. 2023

Indirect methods of detection

Linked SNPs from RAD-seq, WG-seq

- Non-recombinating blocks (inversions and others)
- Long regions (> 100 kb)
- Highly divergent and polymorphic

Similar signatures:

- Introgression
- Linked selection

Direct methods of detection

Short-reads (100-150 bp)

- Read depth
- Mapping aberrations
- Proportion of heterozygotes

Detecting SVs with RADseq data

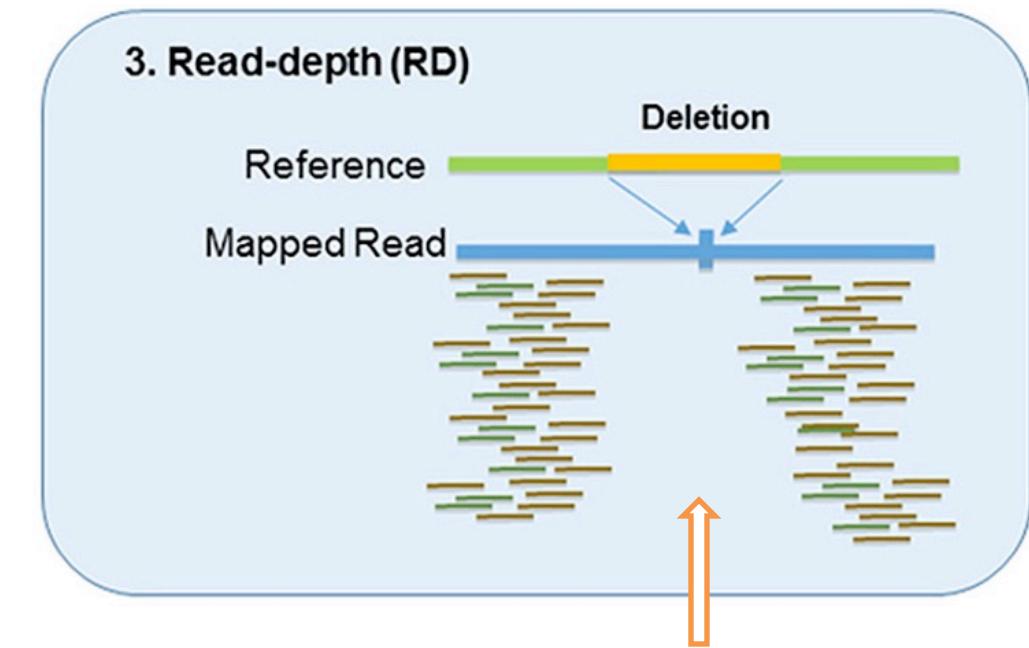
Method: McKinney et al 2016 Molecular Ecology Resources

Case study: Dorant et al 2020 Molecular Ecology

Direct methods of detection

Short-reads (100-150 bp)

- Read depth
- Mapping aberrations
- Proportion of heterozygotes



Detecting SVs with short read data

McKinney et al 2016 Molecular Ecology Resources

Pirooznia et al 2015 frontiers in Genetics

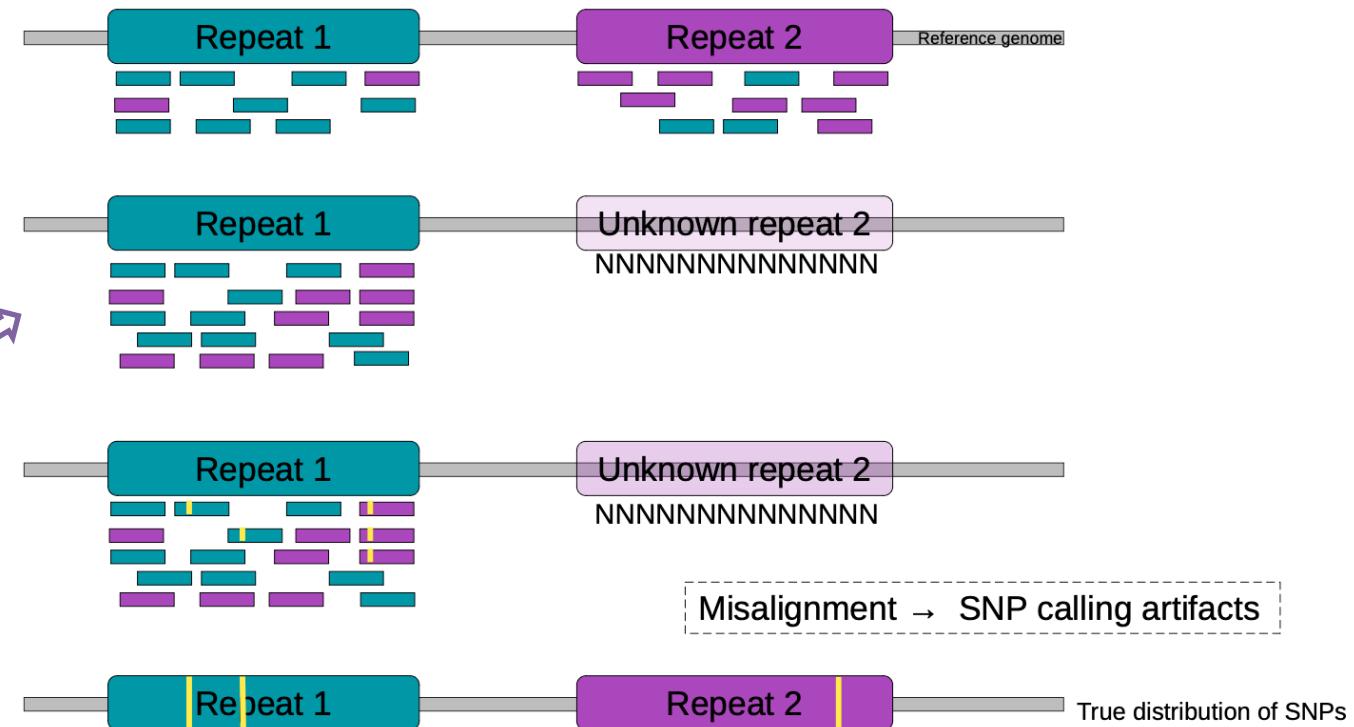
Dorant et al 2020 Molecular Ecology

Direct methods of detection

Short-reads (100-150 bp)

- Read depth
- Mapping aberrations
- Proportion of heterozygotes

2X read depth



Detecting SVs with short read data

McKinney et al 2016 Molecular Ecology Resources

Pirooznia et al 2015 frontiers in Genetics

Dorant et al 2020 Molecular Ecology

Direct methods of detection

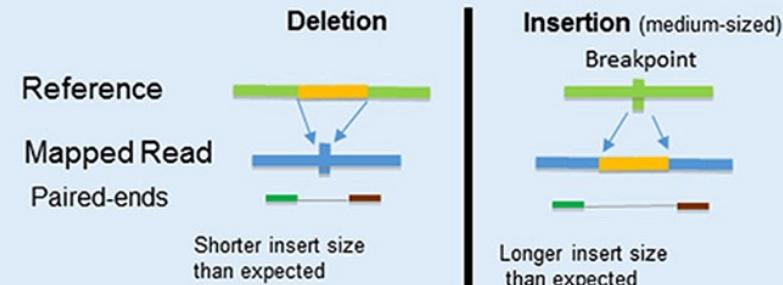
- Read depth
 - **Mapping aberrations**
 - Proportion of heterozygotes
-
- BreakDancer
 - Delly
 - Lumpy
 - MOPline

Detecting SVs with short read data

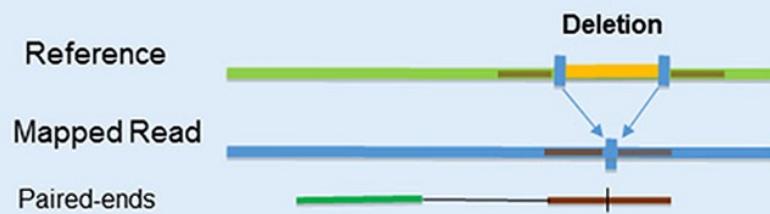
McKinney et al 2016 Molecular Ecology Resources
Pirooznia et al 2015 frontiers in Genetics
Mahmoud et al 2019 Genome Biology
Dorant et al 2020 Molecular Ecology

Short-reads (100-150 bp)

1. Read-Pair (RP)



2. Split-read (SR)



Insertion



Direct methods of detection

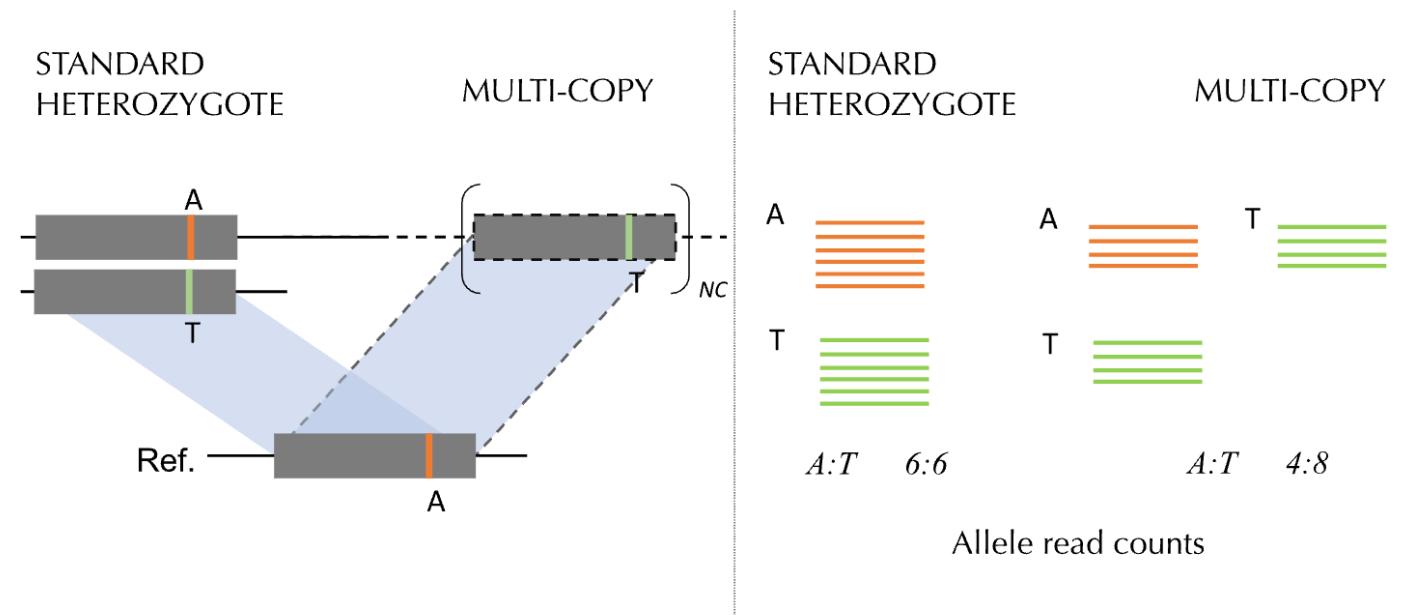
- Read depth
- Mapping aberrations
- Proportion of heterozygotes

Higher heterozygote proportion

Deviation of 1:1 read ratio in heterozygotes

- rCNV (CNVs from SNPs)*

Short-reads (100-150 bp)



Direct methods of detection

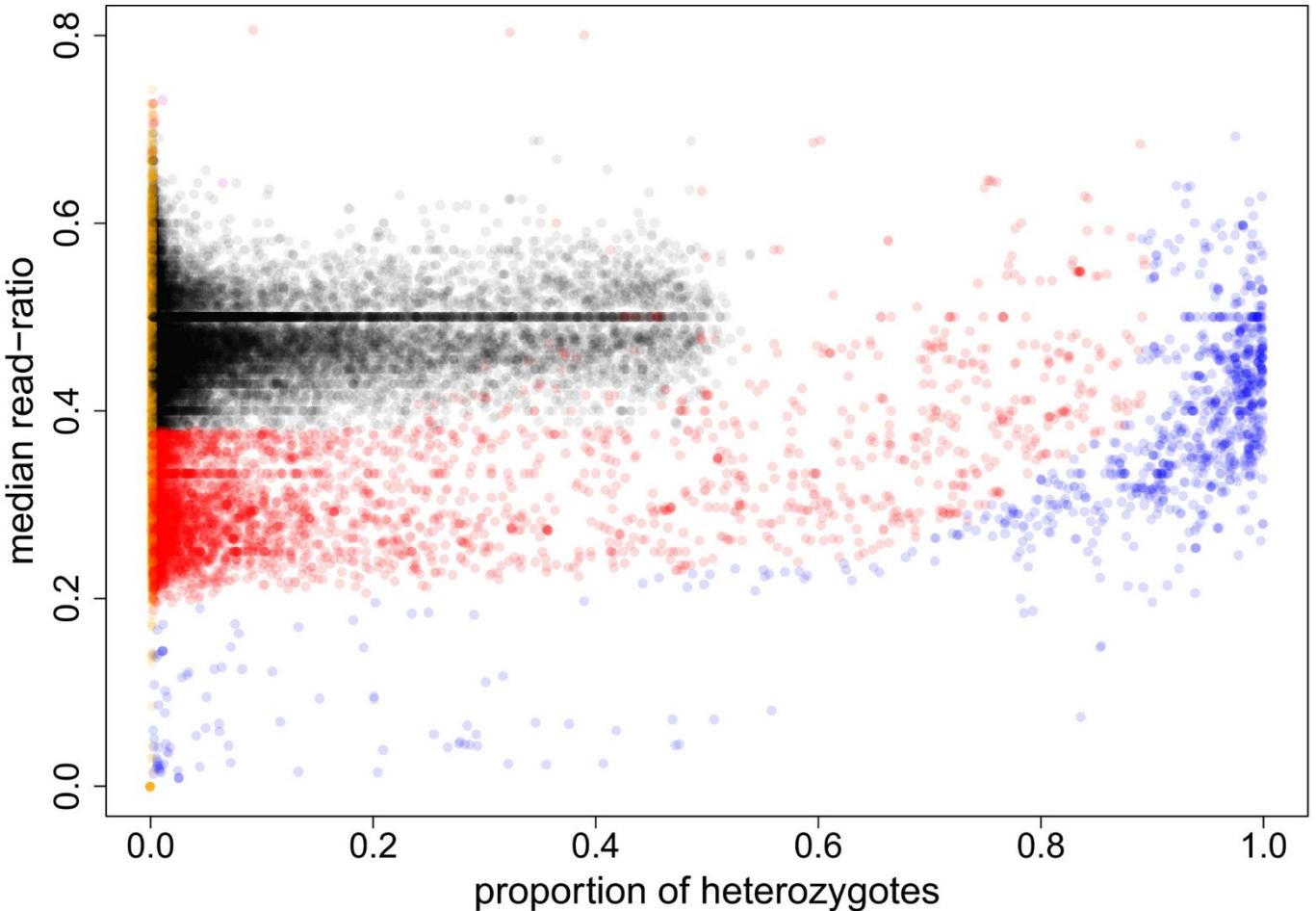
Short-reads (100-150 bp)

- Read depth
- Mapping aberrations
- Proportion of heterozygotes

Higher heterozygote proportion

Deviation of 1:1 read ratio in heterozygotes

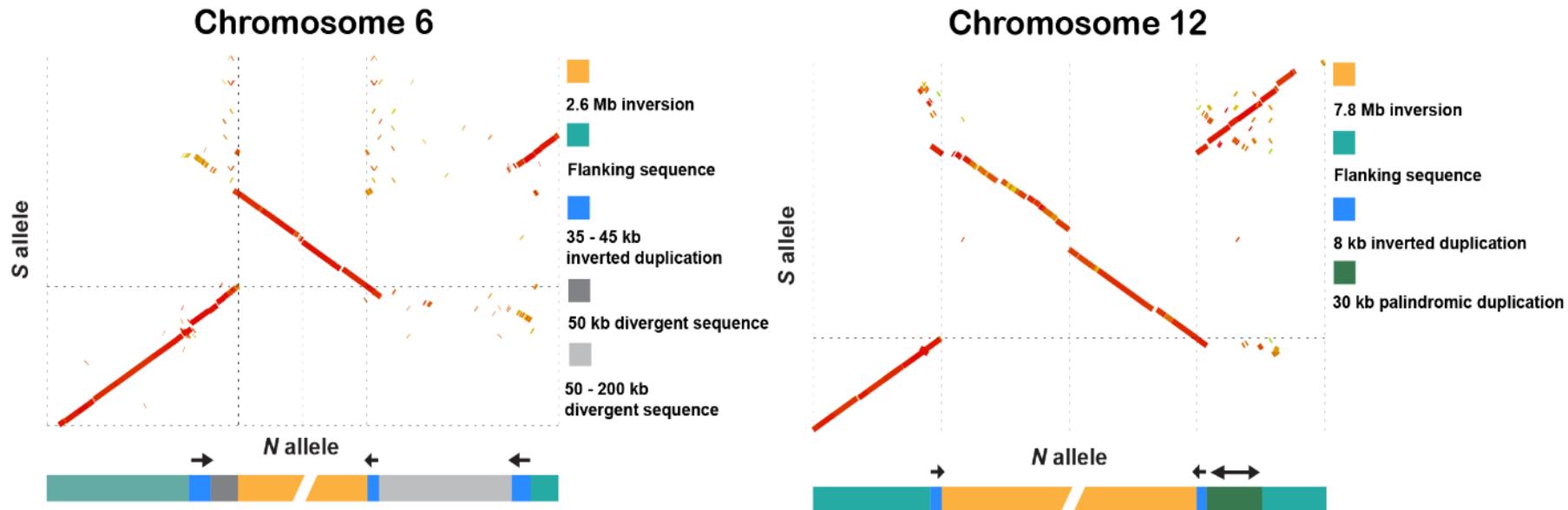
- rCNV (CNVs from SNPs)*



Direct methods of detection

Long Reads

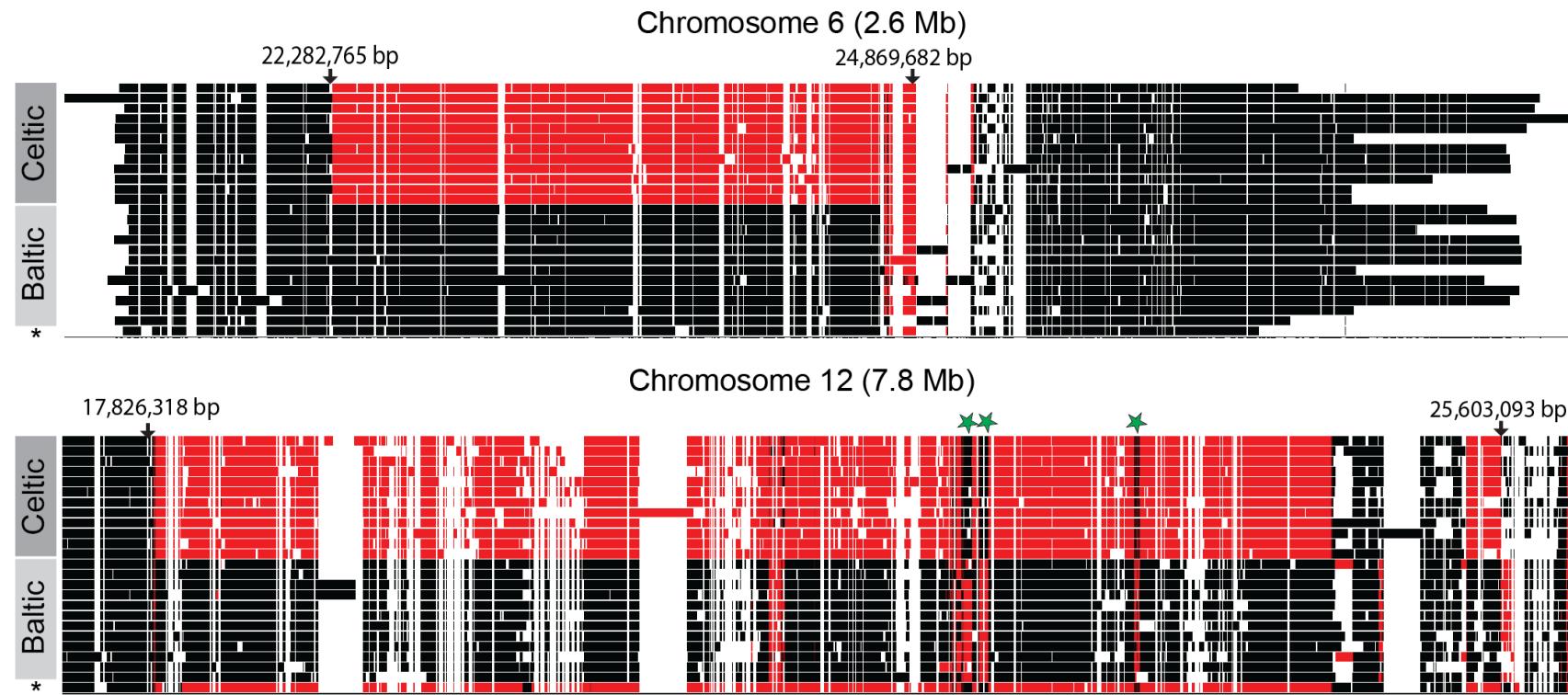
Direct mapping between assembled genomes



Direct methods of detection

Long Reads

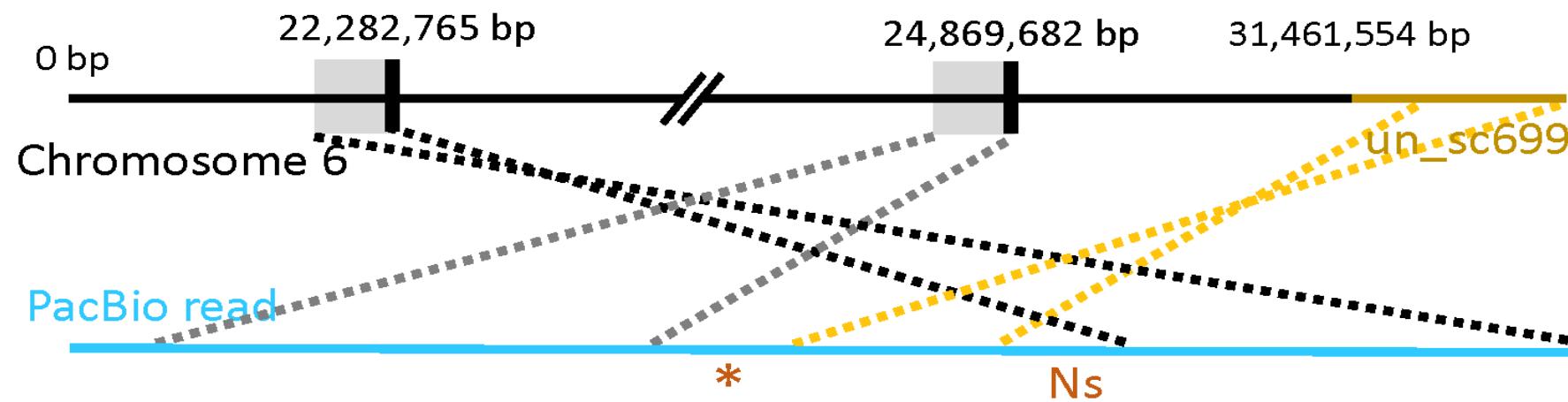
Direct mapping between assembled genomes [Pangenomes](#)



Direct methods of detection

Long Reads

Pattern of PacBio read mapping

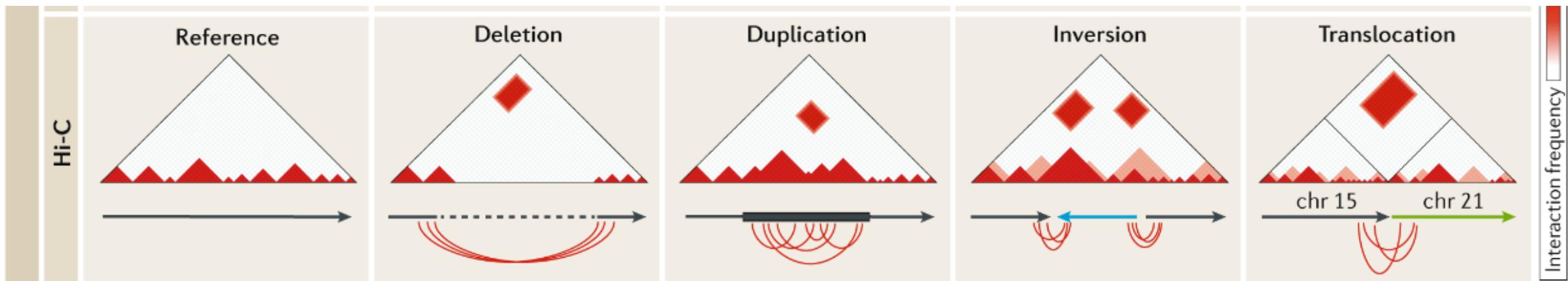


Read is split between the two breakpoints of the inversion

Direct methods of detection

Hi-C data

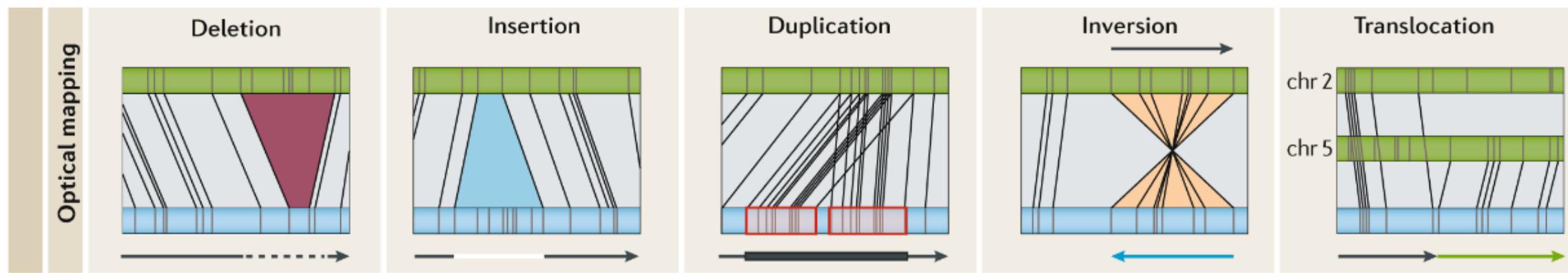
Spatial organization of chromatin in the cell



Direct methods of detection

BioNano

Maps the location of restriction enzyme sites along the chromosomes



Direct methods of detection

Good reviews:

- Ho et al 2020 Nature Review Genetics
- Ashan 2023 Nature Methods
- Merot, Oomen, Tigano, Wellenreuther 2020, TREE

Practical

Haploblocks – Indirect method to study structural variant 

- Local PCA
- Signature of differentiation
- Signature of suppressed recombination

Structural variant calling - Direct method to study structural variant 

- PCA analysis