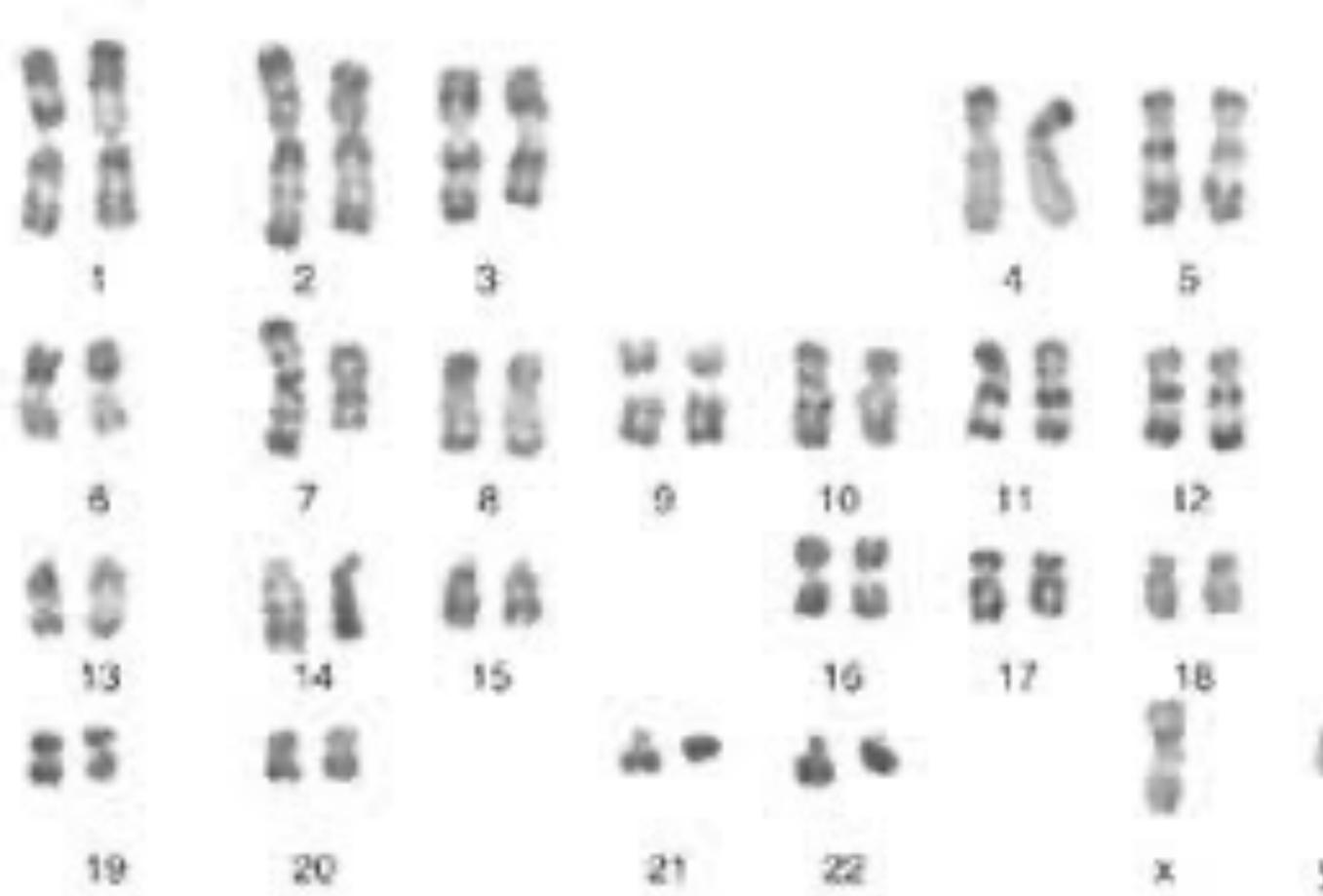
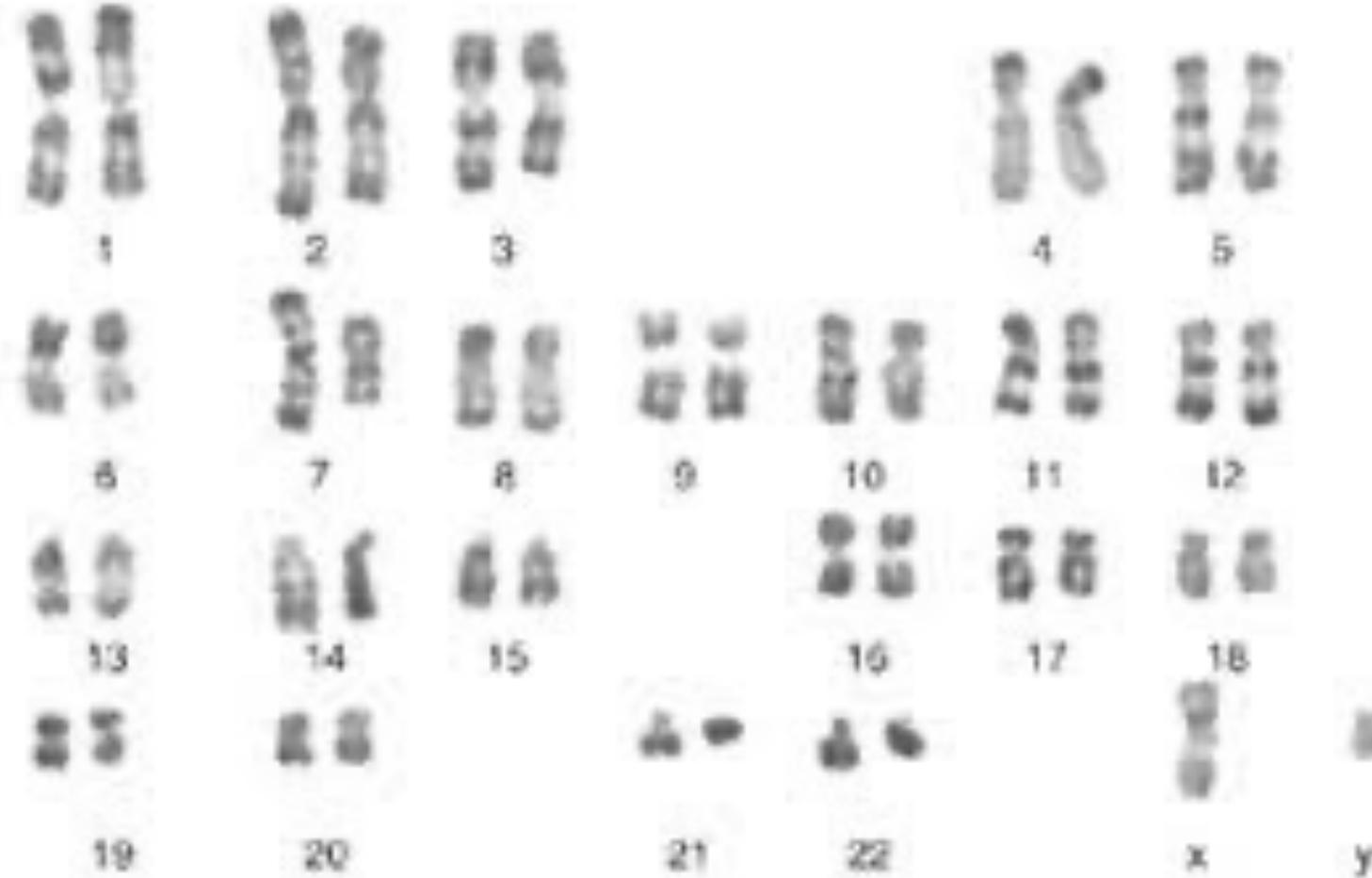


Genome Biology & Genomics



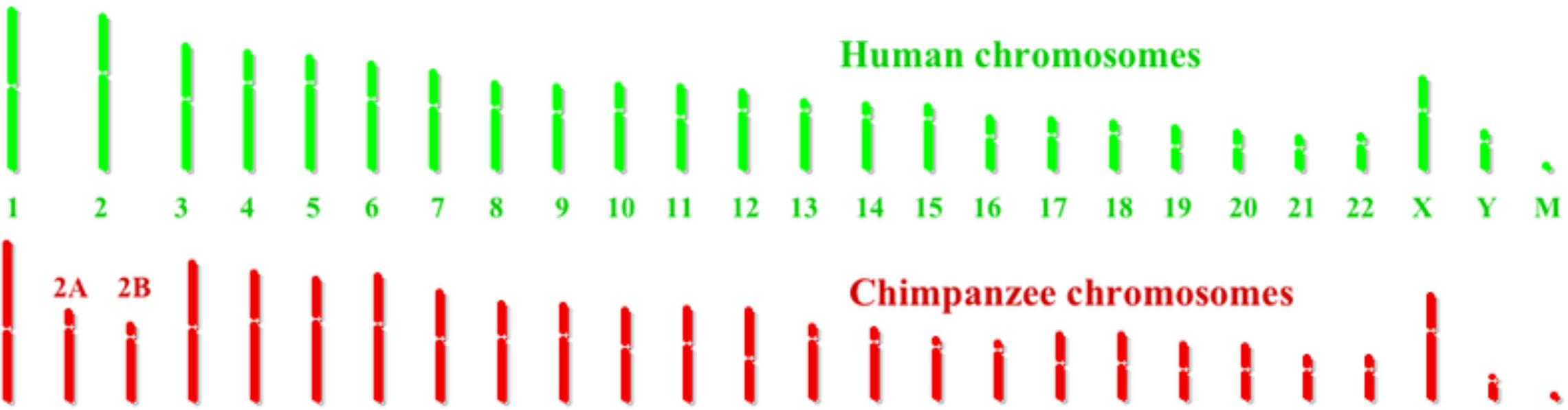
DNA Sequencing Revolution
July 21st, 2022

The complement of chromosomes inside a cell forms the karyotype



[Human Karyotype](#)

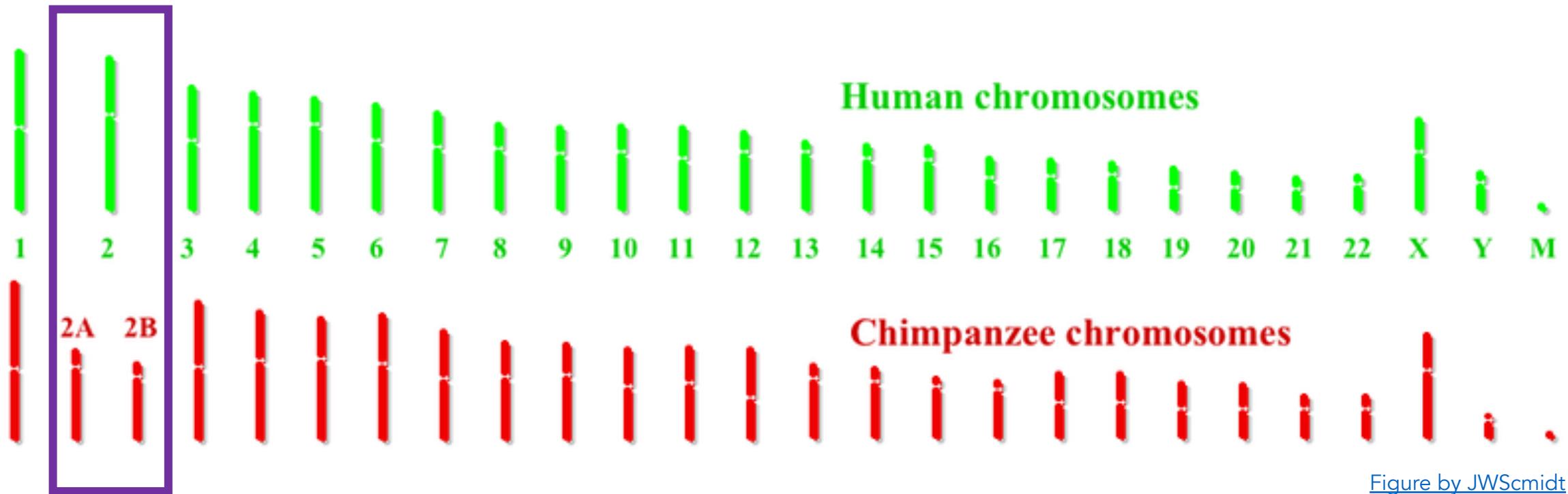
Chromosomal rearrangements alter **karyotype**



[Figure by JWScmidt](#)

Karyotype – The complement of chromosomes inside a cell

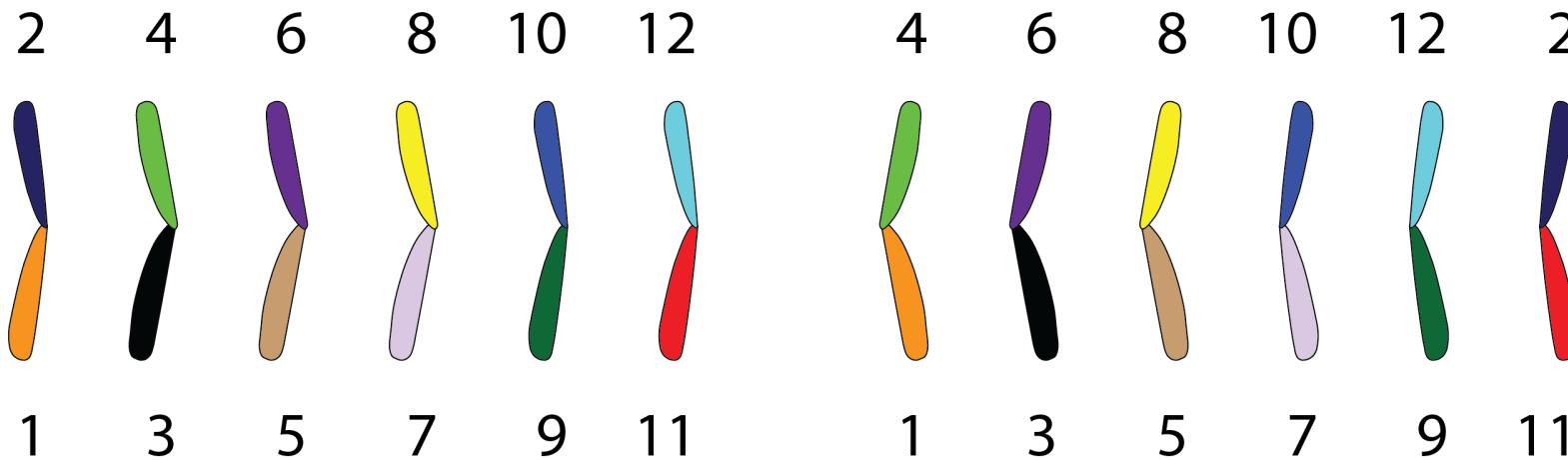
Chromosomal rearrangements alter **karyotype**



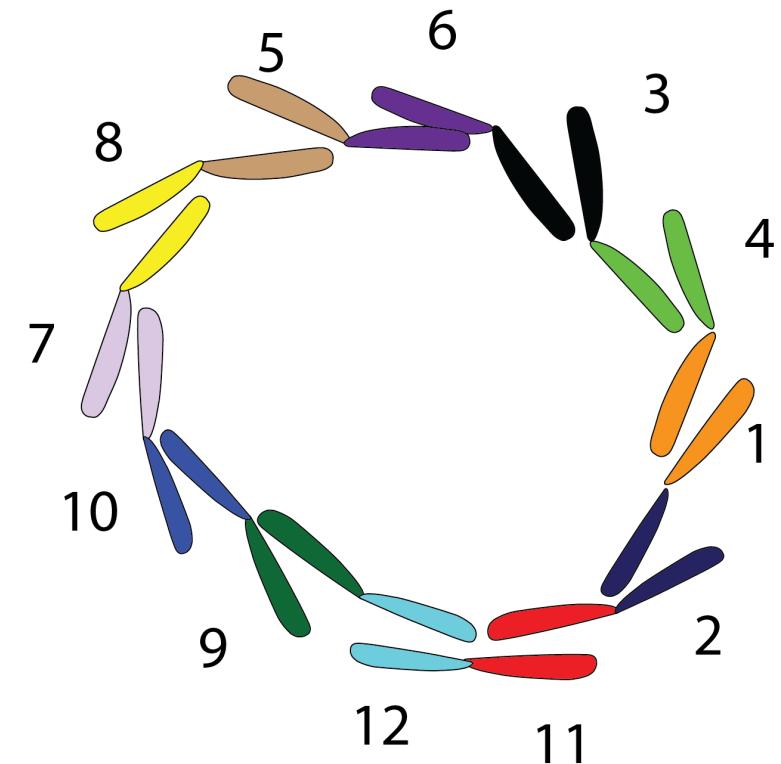
[Figure by JWScmidt](#)

Karyotype – The complement of chromosomes inside a cell

Chromosomal rearrangements alter karyotype



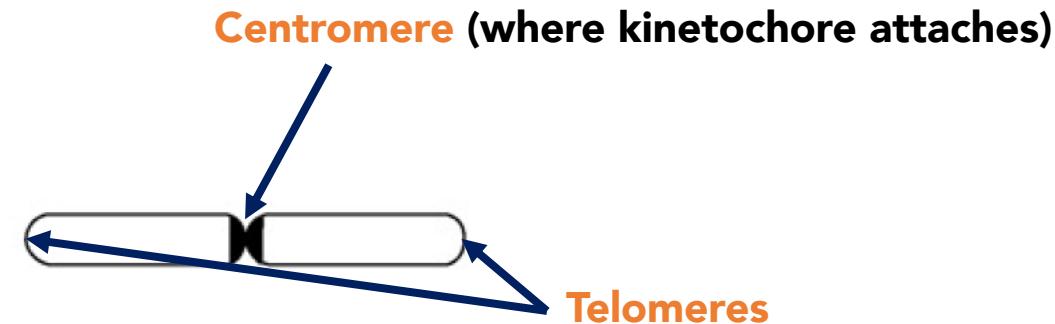
Permanent translocation heterozygosity
in *Oenothera* (evening primroses)



Chromosome biology



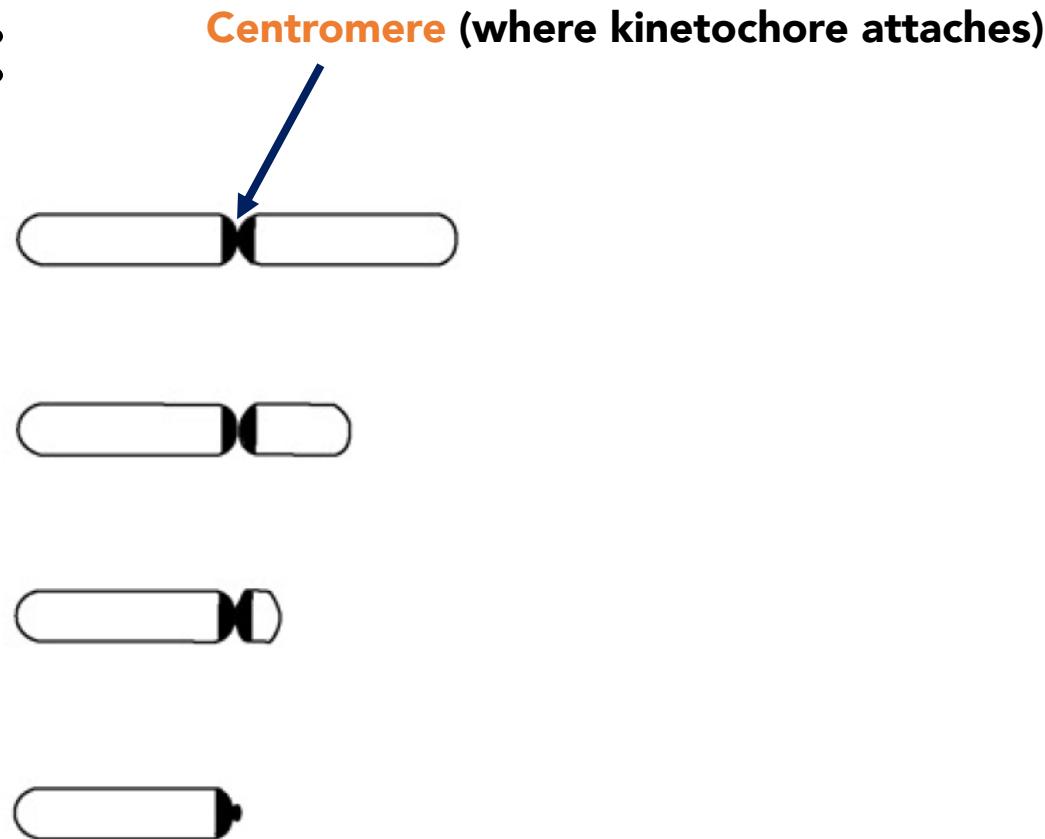
Chromosome biology



Chromosome biology

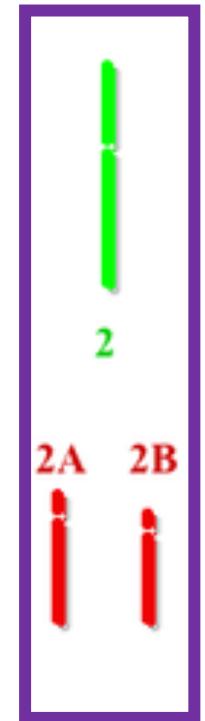
Major Chromosomal Types:

- Metacentric
- Submetacentric
- Acrocentric
- Telocentric



Chromosomal rearrangements:

- Fusion/Fission
- Translocation
- Insertion/Deletion
- Inversion
- Repeat Expansion
- Chromothripsis (chromosome shattering)
- Duplications (4/11/22)



[Figure by JWScmidt](#)

Inversions

- DNA flips in orientation relative to the rest of the chromosome

Inversions between

Human
Chr 5 Chimpanzee
Chr 4



Griffiths et al., *Introduction to Genetic Analysis*, 7th Edition, W.H. Freeman Company

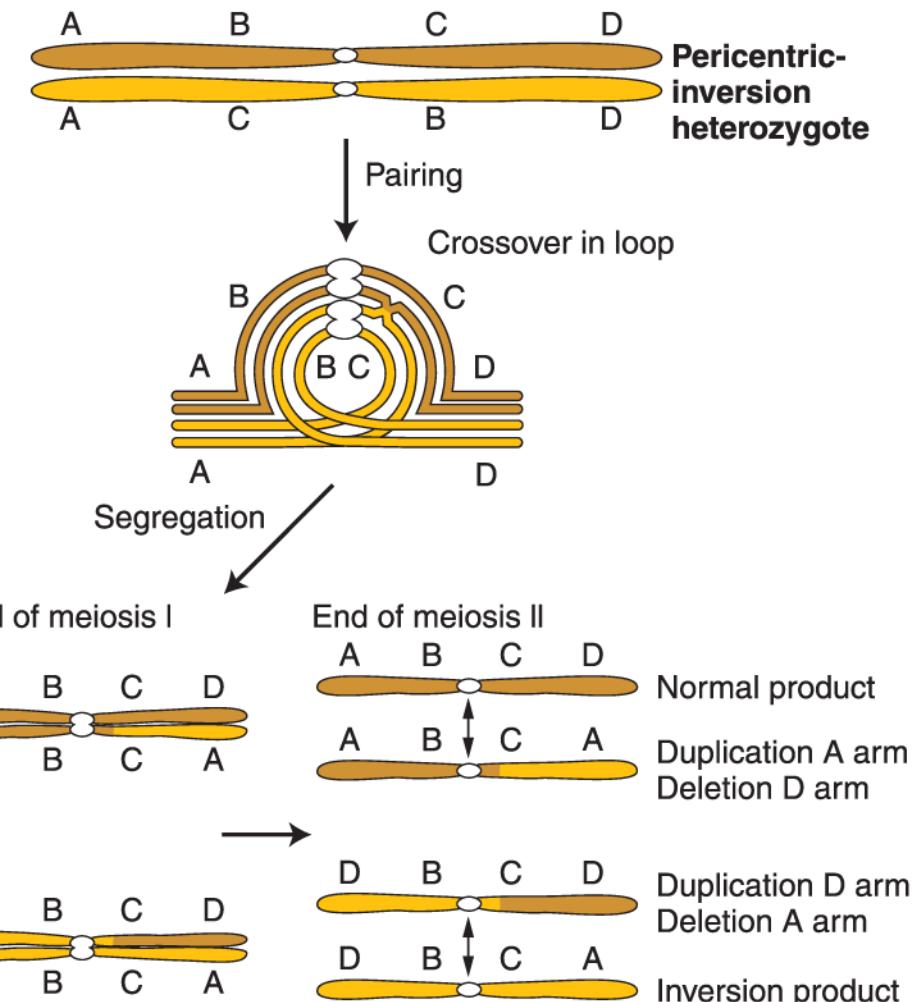
Inversions

- Recombination within inversion loop...
- Can prevent hybridization!



Rhagoletis pomonella

Pericentric inversions can lead to duplication-and-deletion products

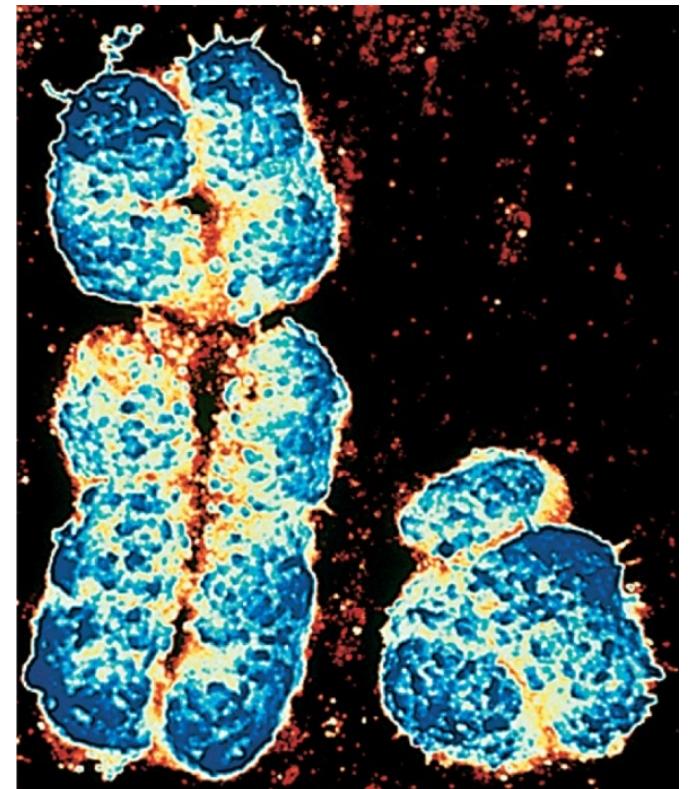


Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

Epigenetic (“above genetics”) inheritance

- Modifications to the **chromatin, histones, and DNA** that affect gene expression
- Can be inherited across generations
- Do not change DNA sequence*

*Except in the case of cytosine deamination

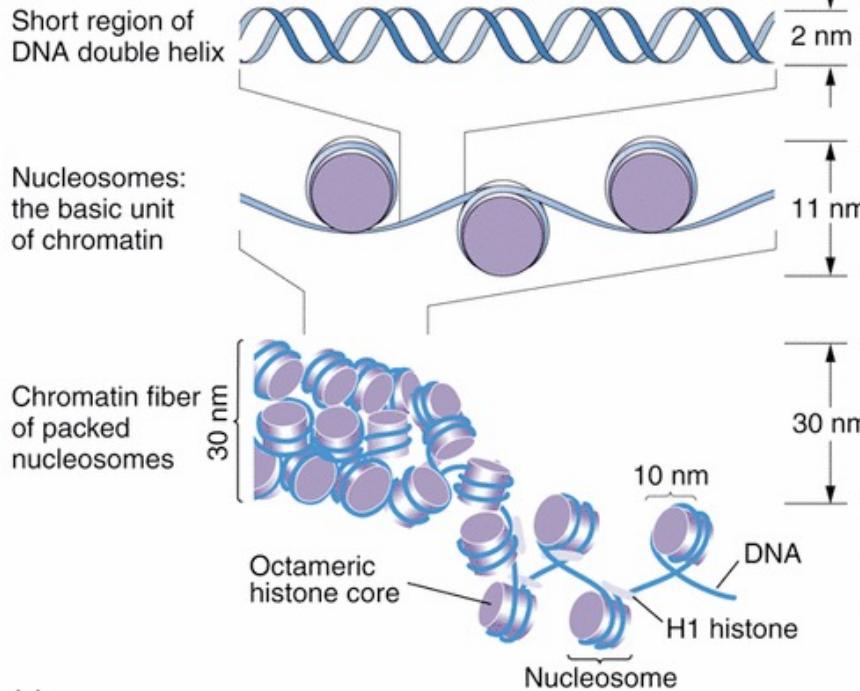


Chromatin

- Euchromatin (open) vs. heterochromatin (closed)

The structure of chromatin

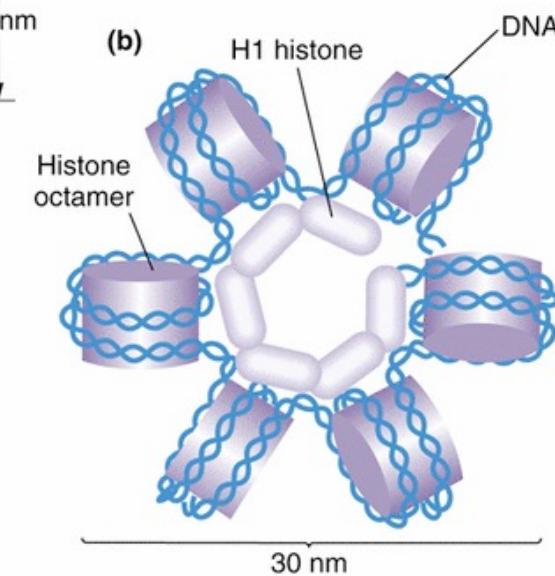
(a)



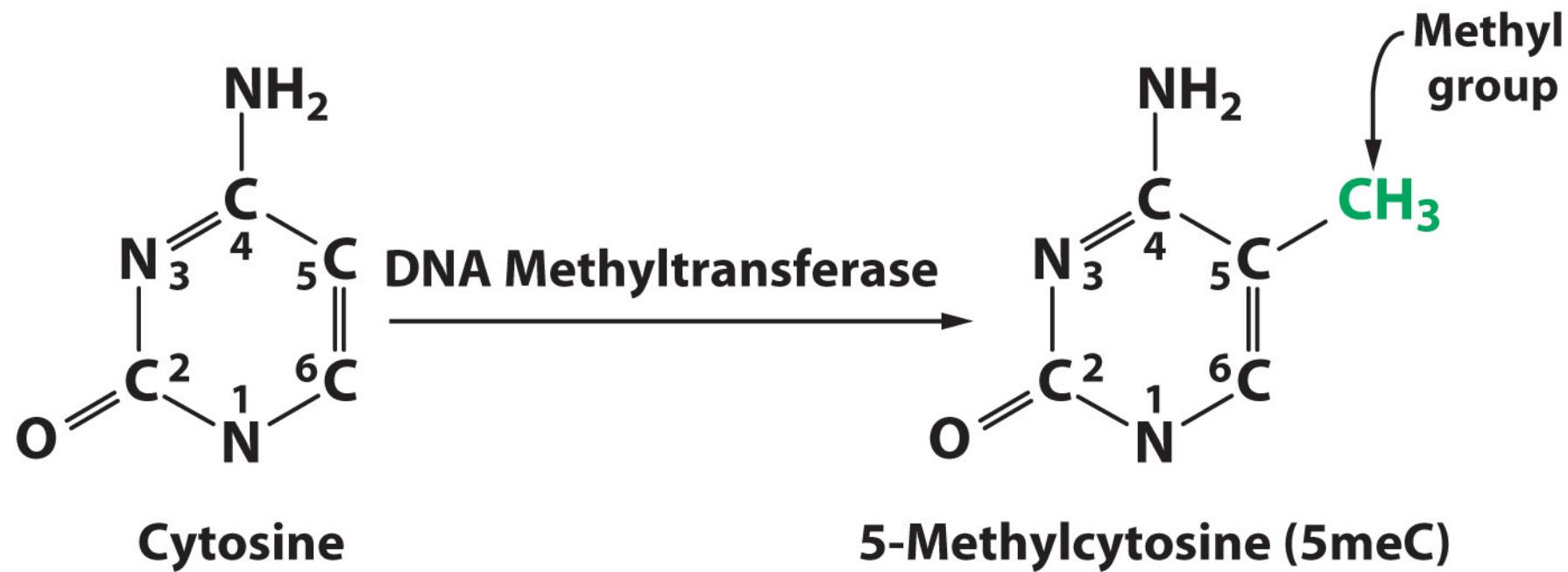
(c)



(b)



DNA modifications



Unnumbered 12 p449

Introduction to Genetic Analysis, Eleventh Edition
© 2015 W. H. Freeman and Company

DNA methylation is heritable

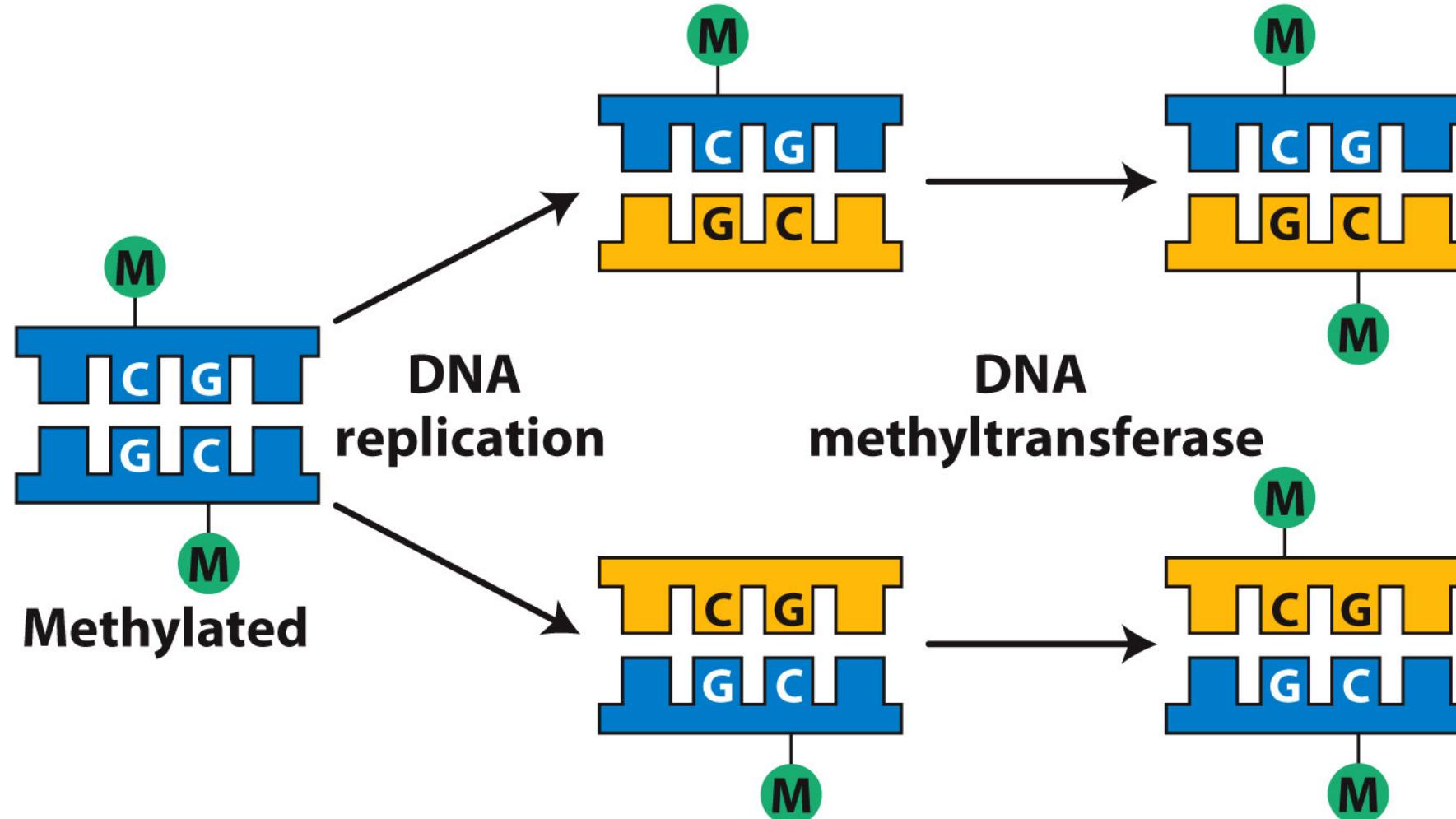
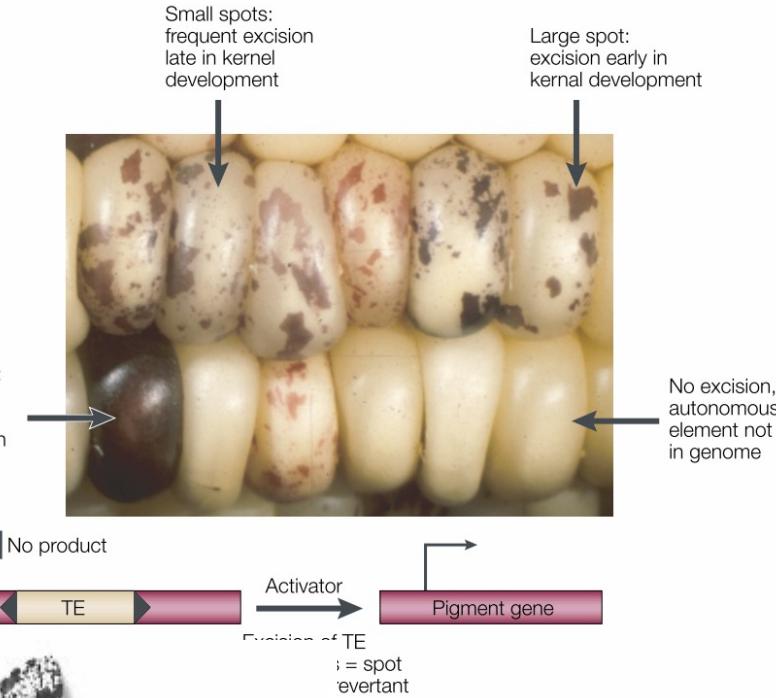


Figure 12-17

Introduction to Genetic Analysis, Eleventh Edition
© 2015 W. H. Freeman and Company

Transposable elements (TEs)

- “Genomic parasites” or “jumping genes”
- Discovered by Barbara McClintock in 1940s
- Can have potentially harmful consequences on host genome
- Growing appreciation for important but complex role in evolution
 - Genome architecture
 - Adaptation



LETTER

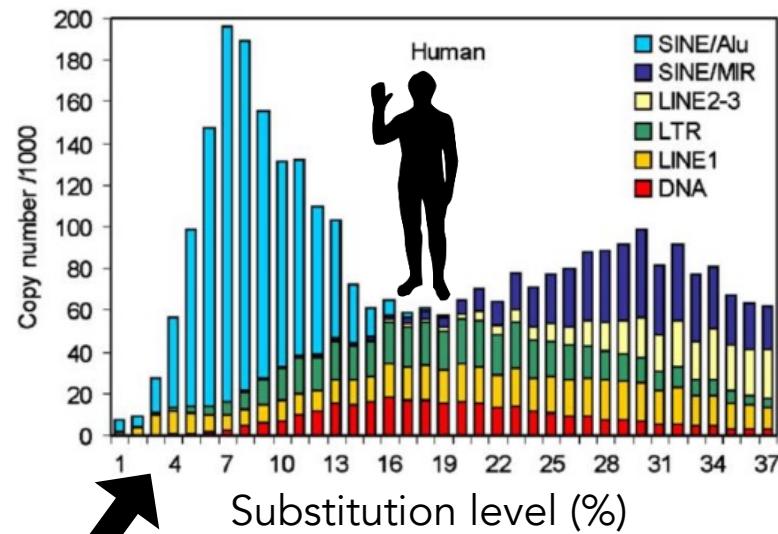
doi:10.1038/nature17951

The industrial melanism mutation in British peppered moths is a transposable element

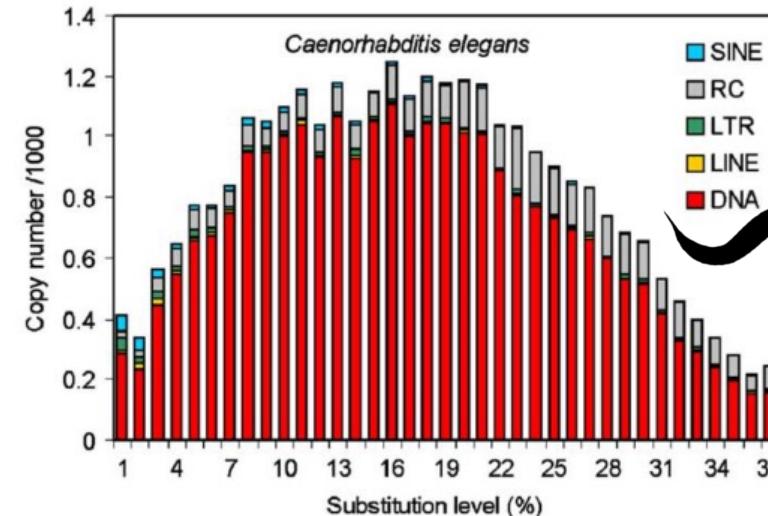
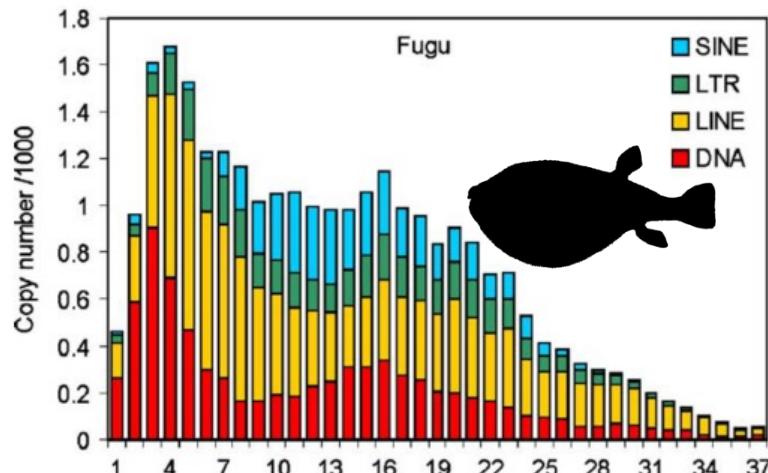
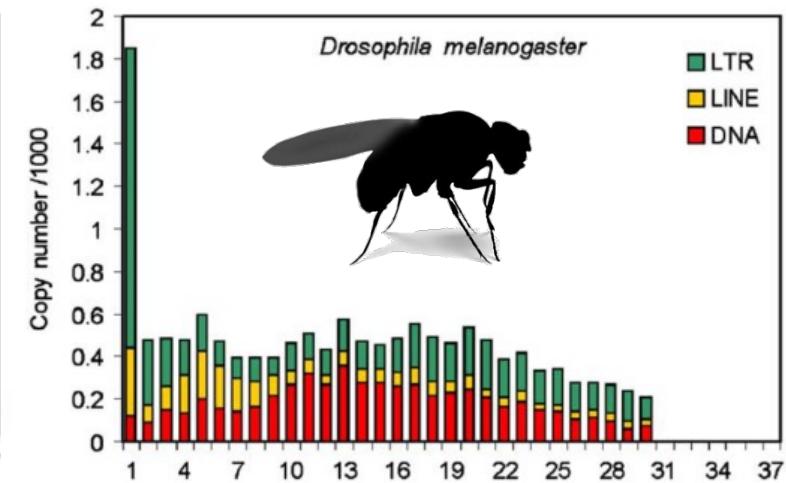
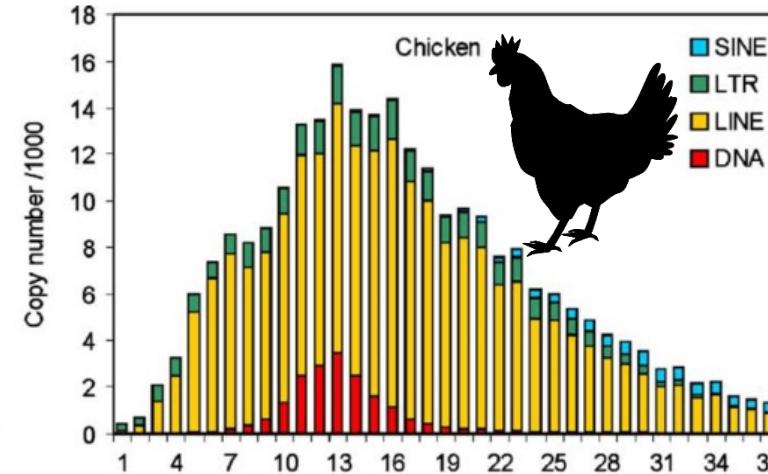
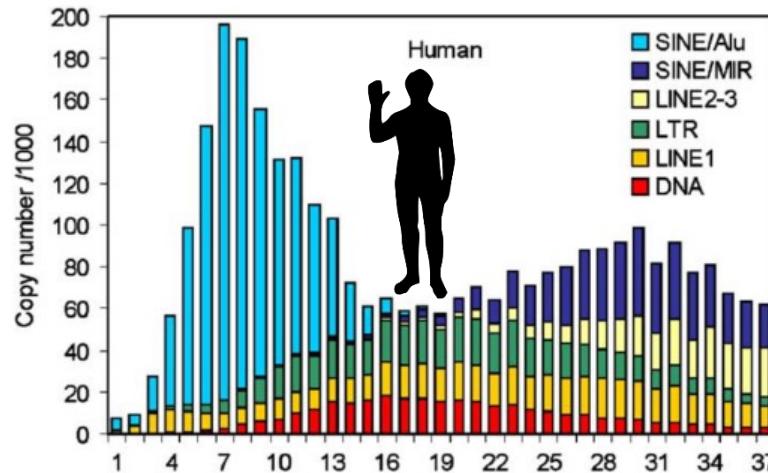
Arjen E. van't Hof^{1,*}, Pascal Campagne^{1,*}, Daniel J. Rigden¹, Carl J. Yung¹, Jessica Lingley¹, Michael A. Quail², Neil Hall¹, Alistair C. Darby¹ & Iliik J. Saccheri¹

Feschotte et al. 2002 Nat Rev Genet

Evolutionary dynamics and diversity of TEs vary greatly across species



Evolutionary dynamics and diversity of TEs vary greatly across species



Movement of TEs

DNA transposon: "cut and paste"



Retrotransposon: "copy and paste"



Movement of TEs

DNA transposon: "cut and paste"



Retrotransposon: "copy and paste"



Movement of TEs

DNA transposon: "cut and paste"



Retrotransposon: "copy and paste"



Movement of TEs

DNA transposon: "cut and paste"



Retrotransposon: "copy and paste"



Movement of TEs

DNA transposon: "cut and paste"

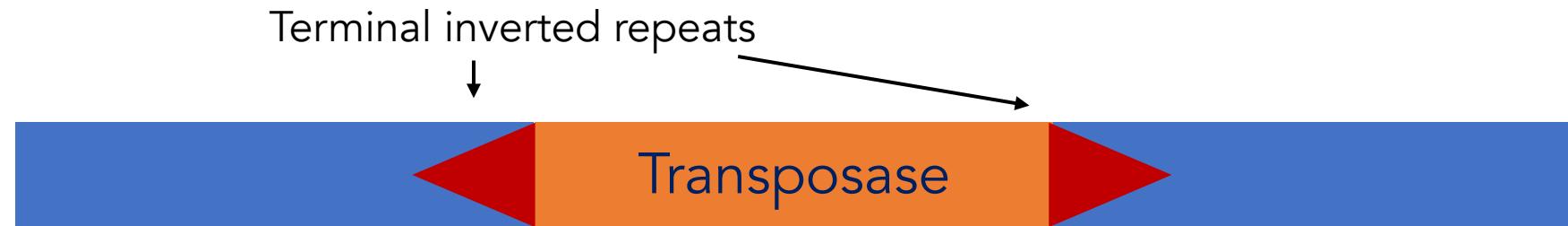


Retrotransposon: "copy and paste"

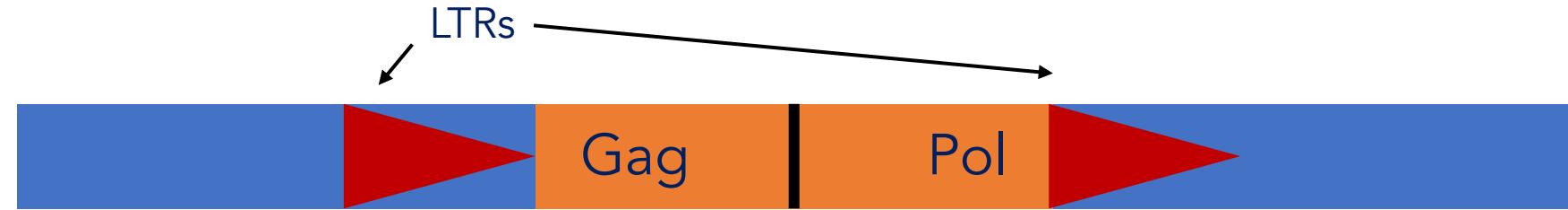


Classes of TEs

DNA Transposon



Retrotransposons



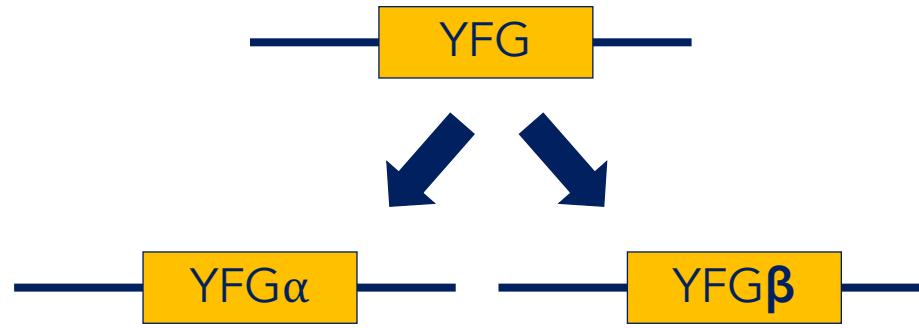
Long terminal repeats



Non-long terminal repeats

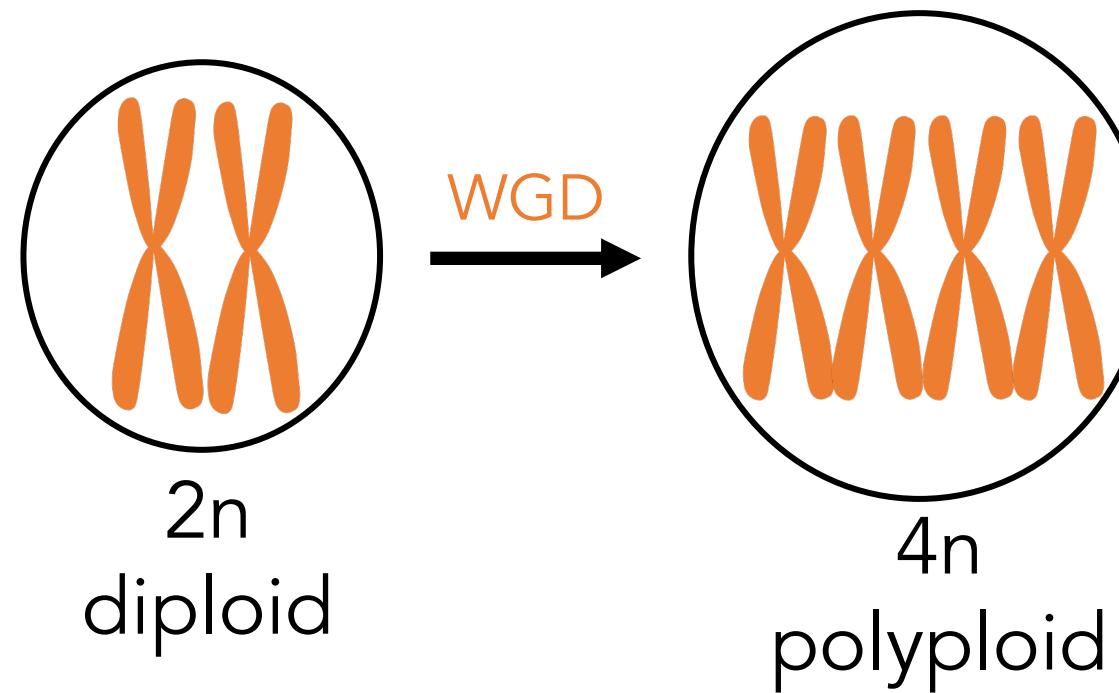
Gene/Genome Duplication – The fuel for evolution

- Gene duplication
 - Unequal crossover/
ectopic recombination
 - Transposition
 - Polymerase slippage



Gene/Genome Duplication – The fuel for evolution

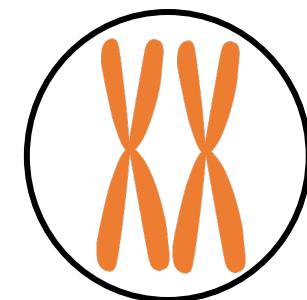
- Whole-genome duplication
 - Ploidy is increased



Whole genome duplications result from autopolyploidy or allopolyploidy

Autopolyploidy

WGD within species



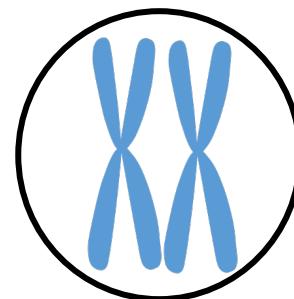
WGD



Allopolyploidy

Hybridization induced WGD

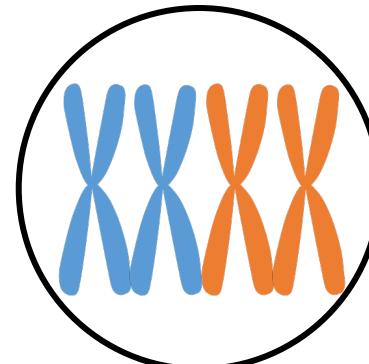
Diploid



WGD



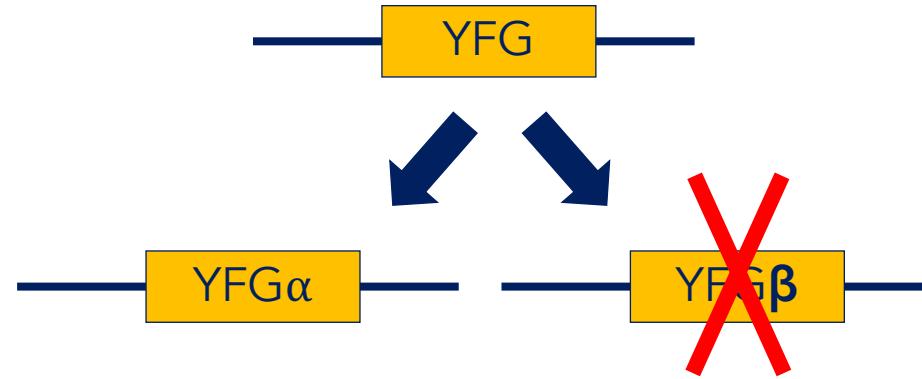
Polyplloid



Gene/Genome Duplication – The fuel for evolution

Evolutionary fate of gene duplicates

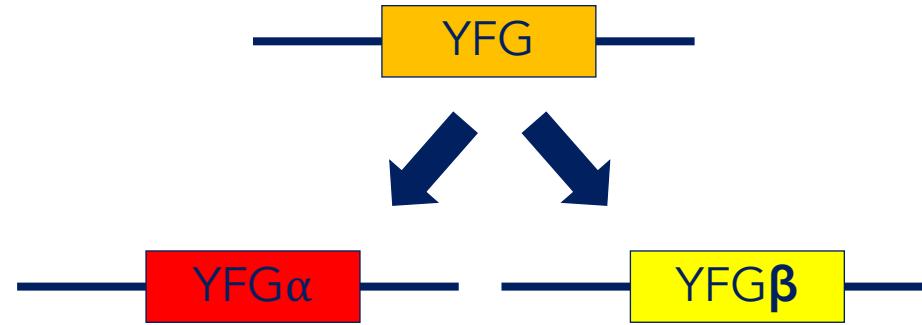
- Pseudogenization
 - Gene is lost
 - Most common fate



Gene/Genome Duplication – The fuel for evolution

Evolutionary fate of gene duplicates

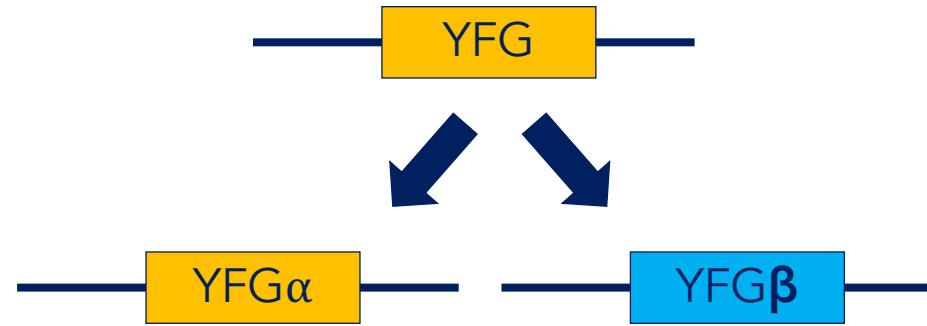
- Pseudogenization
- Subfunctionalization
 - Ancestral function is partitioned



Gene/Genome Duplication – The fuel for evolution

Evolutionary fate of gene duplicates

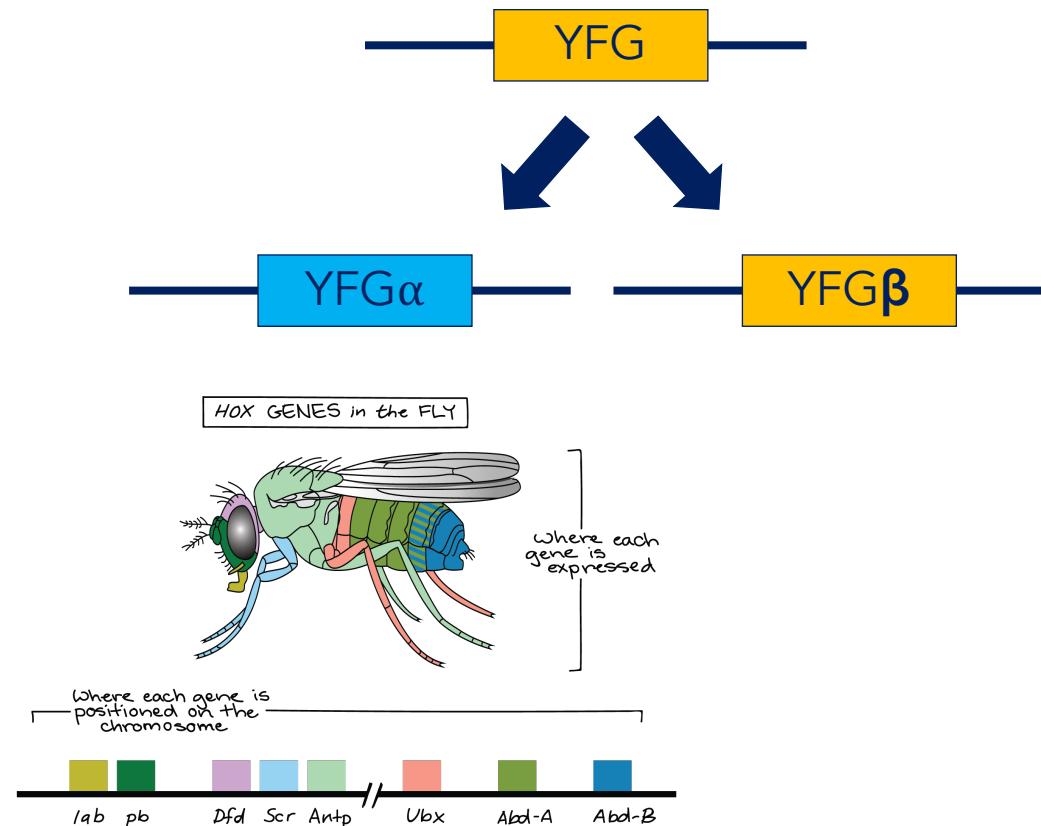
- Pseudogenization
- Subfunctionalization
- Neofunctionalization
 - Gain of novel function



Gene/Genome Duplication – The fuel for evolution

Evolutionary fate of gene duplicates

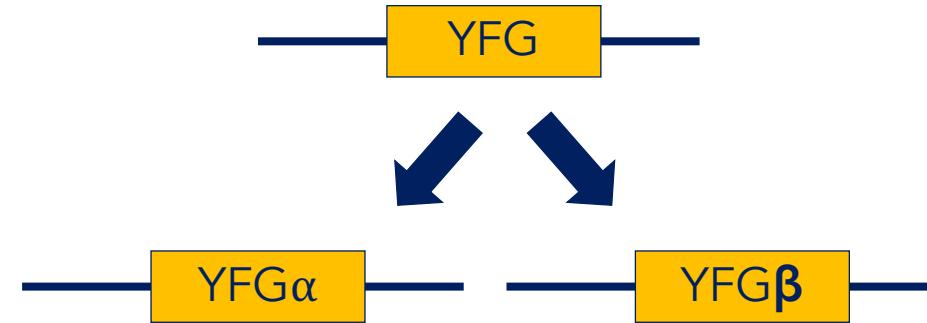
- Pseudogenization
- Subfunctionalization
- Neofunctionalization
- Exaptation/Co-option
 - Re-purposing to other function



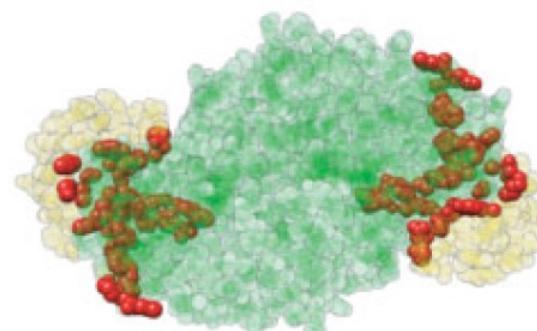
Gene/Genome Duplication – The fuel for evolution

Evolutionary fate of gene duplicates

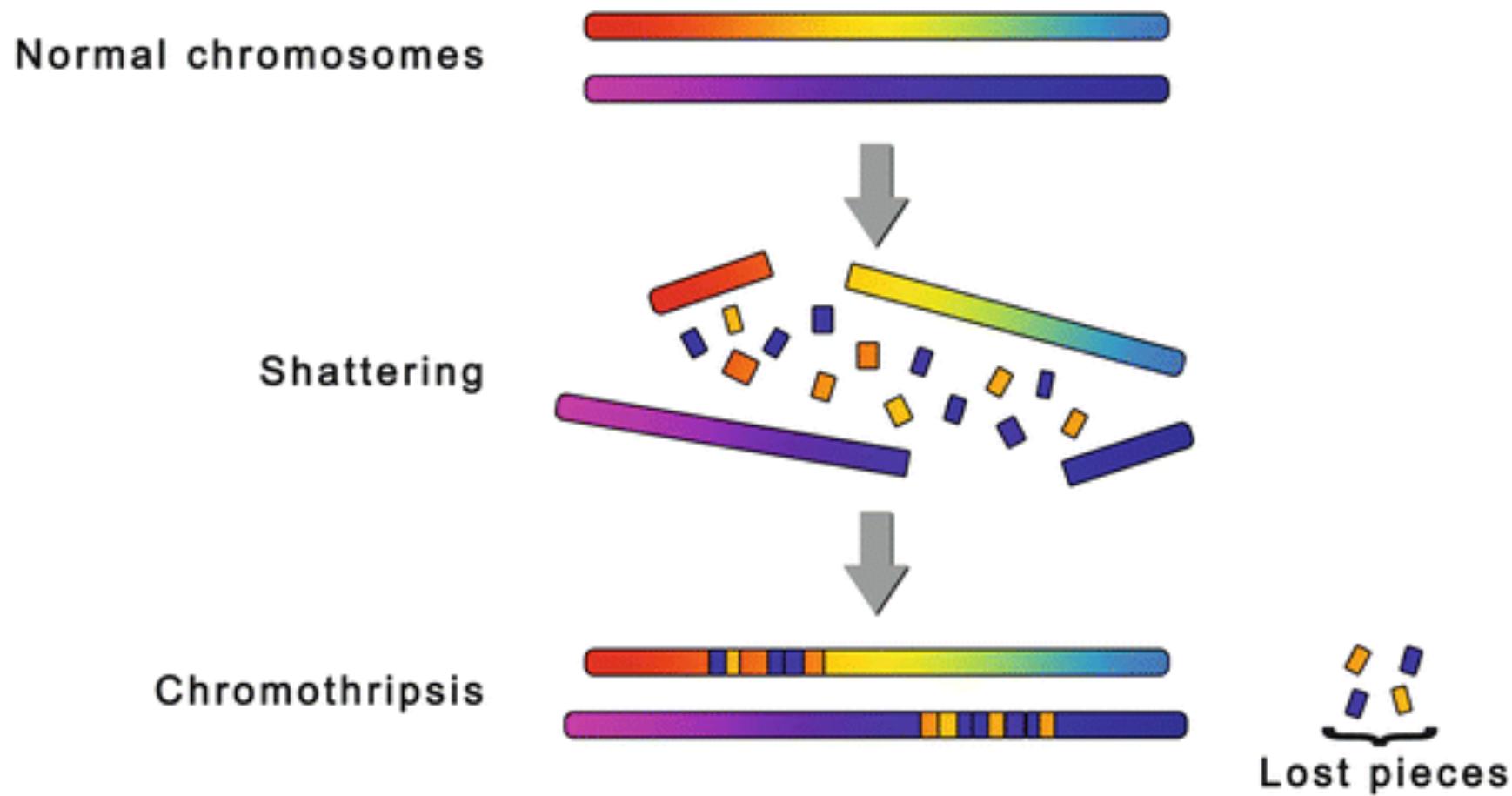
- Pseudogenization
- Subfunctionalization
- Neofunctionalization
- Exaptation/Co-option
- Maintenance



Rubisco

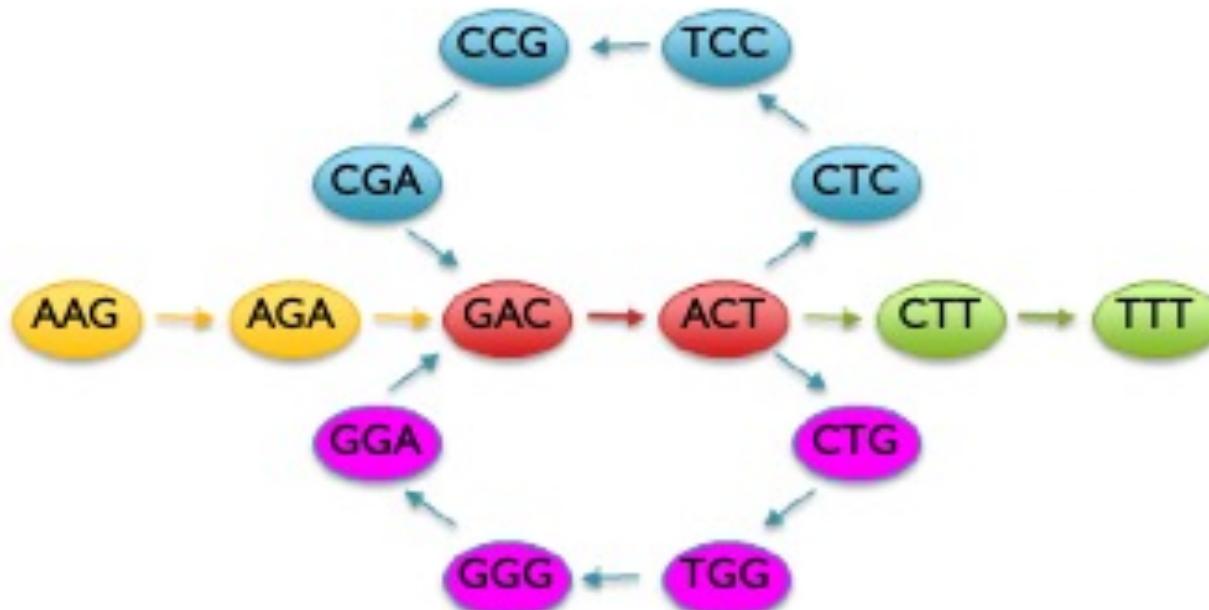


Chromothripsy (chromosome shattering)



Applications in Genomics

Genome and transcriptome assembly



Two primary assembly methods: Overlap-Layout-Consensus vs. de Bruijn Graphs

A

ATATATACTGGCGTATCGCAGTAAACGCGCCG

R1: ACTGGCGTAT

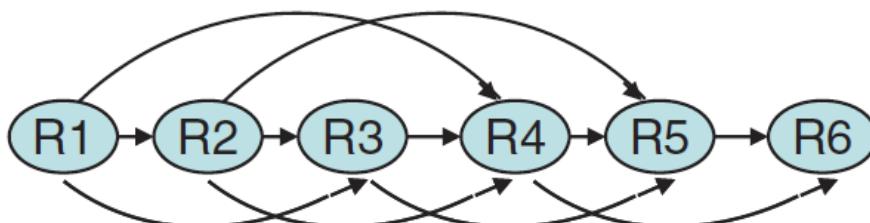
R2: TGGCGTATCG

R3: GGC GTATCGC

R4: CGTATCGCAG

R5: TATCGCAGTA

R6: CGCAGTAAAC



B

ATATATACTGGCGTATCGCAGTAAACGCGCCG

K1: ACTGG

K2: CTGGC

K3: TGGCG

K..:

K14: AGTAA

K15: GTAAA

K16: TAAAC



Overlap-layout-consensus

de Bruijn Graph

Pros:

- Simple, intuitive
- Easy traversal across het sites, repeats
- Flexible based on datatype

Cons:

- Computationally expensive
- Does not scale well with short reads

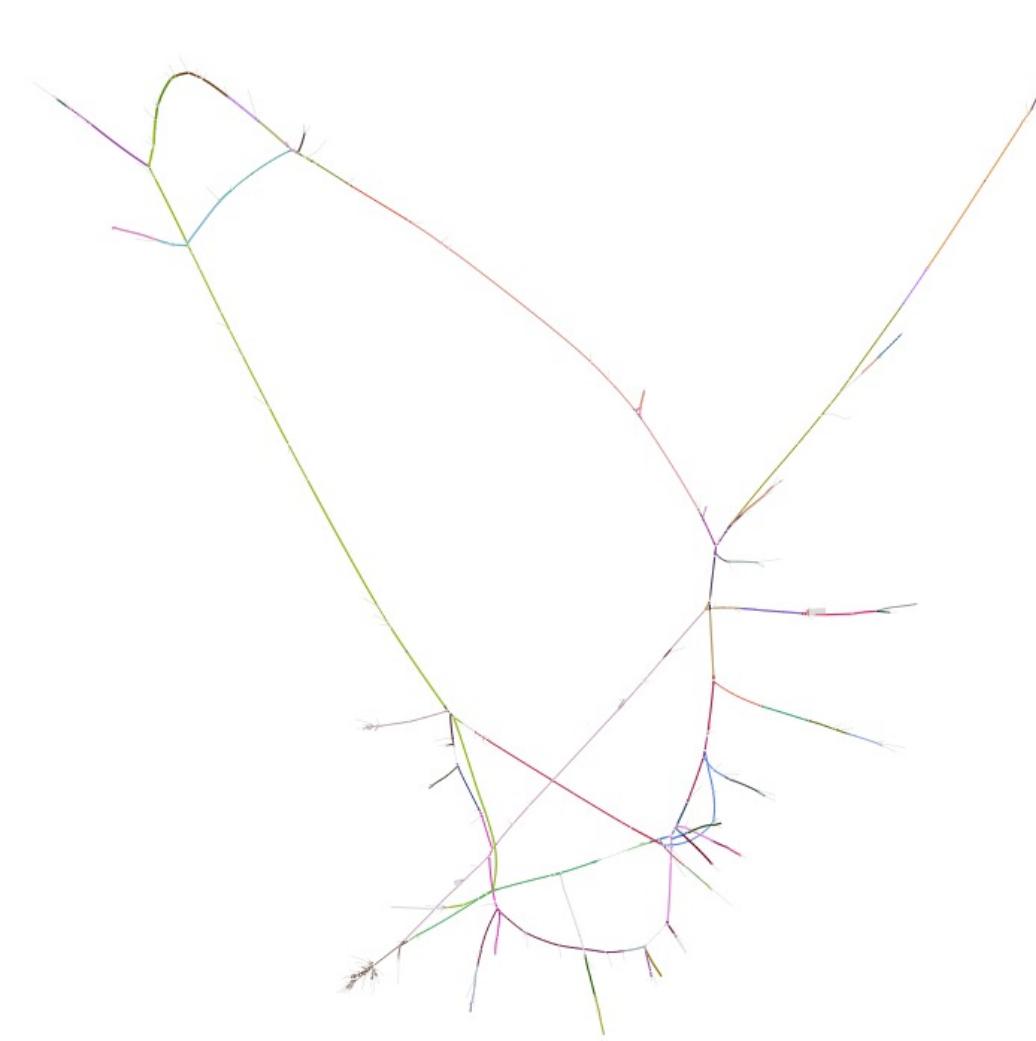
Pros:

- Fast, useful for short reads
- Can assemble a whole genome

Cons:

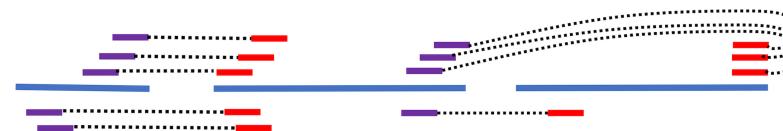
- Bubbles, het sites
- Dealing with sequencing errors a challenge with long reads

Visualizing assemblies w/ Bandage



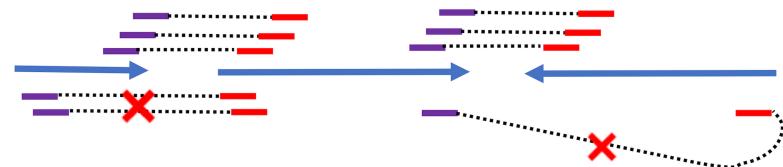
Scaffolding genomic contigs

- Map reads to contigs (e.g., using BWA)



Alignment of reads to contigs

- Paired reads that connect contigs?



Orienting and Ordering contigs based on alignments

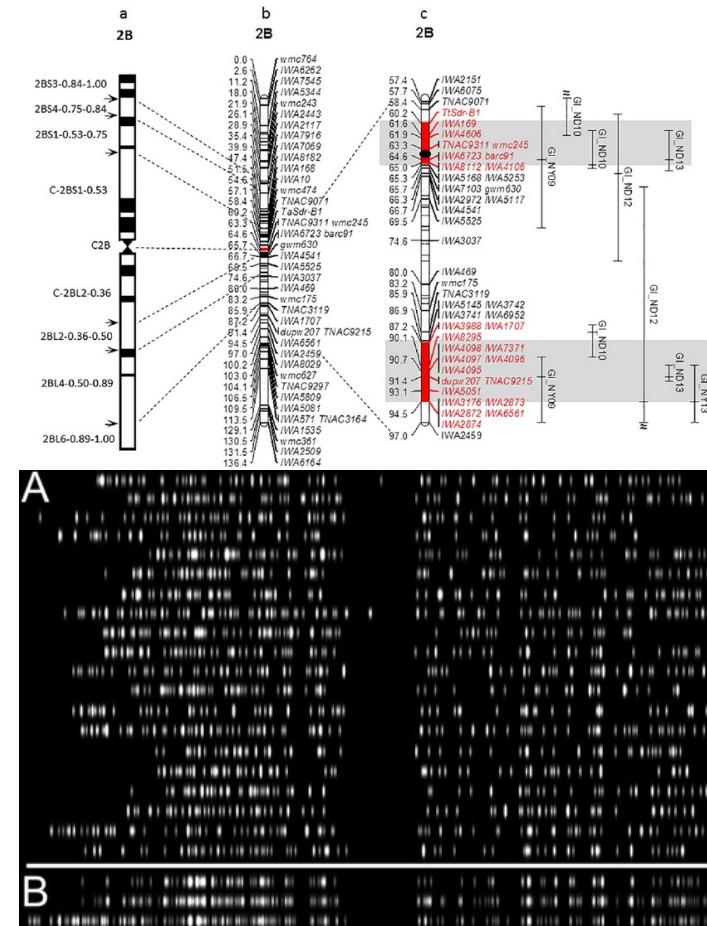
- Mate Pairs, chromatin-level barcoding, long reads



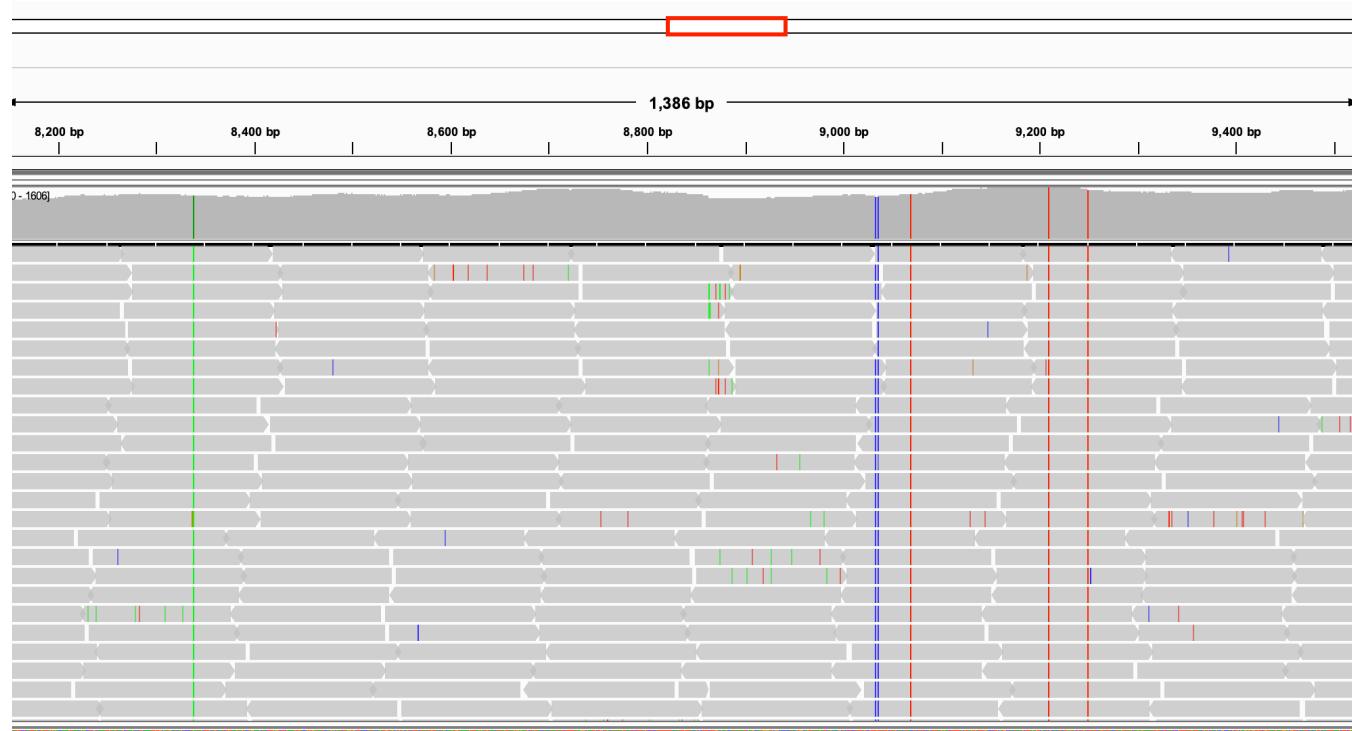
Chromosome-scale assemblies

Generally will require at least one of:

- **linkage map**
 - Need F2 recombinants
- **Long reads**
 - Pacbio/Nanopore
- **optical mapping (e.g., BioNano)**
 - Expensive



Variant Calling

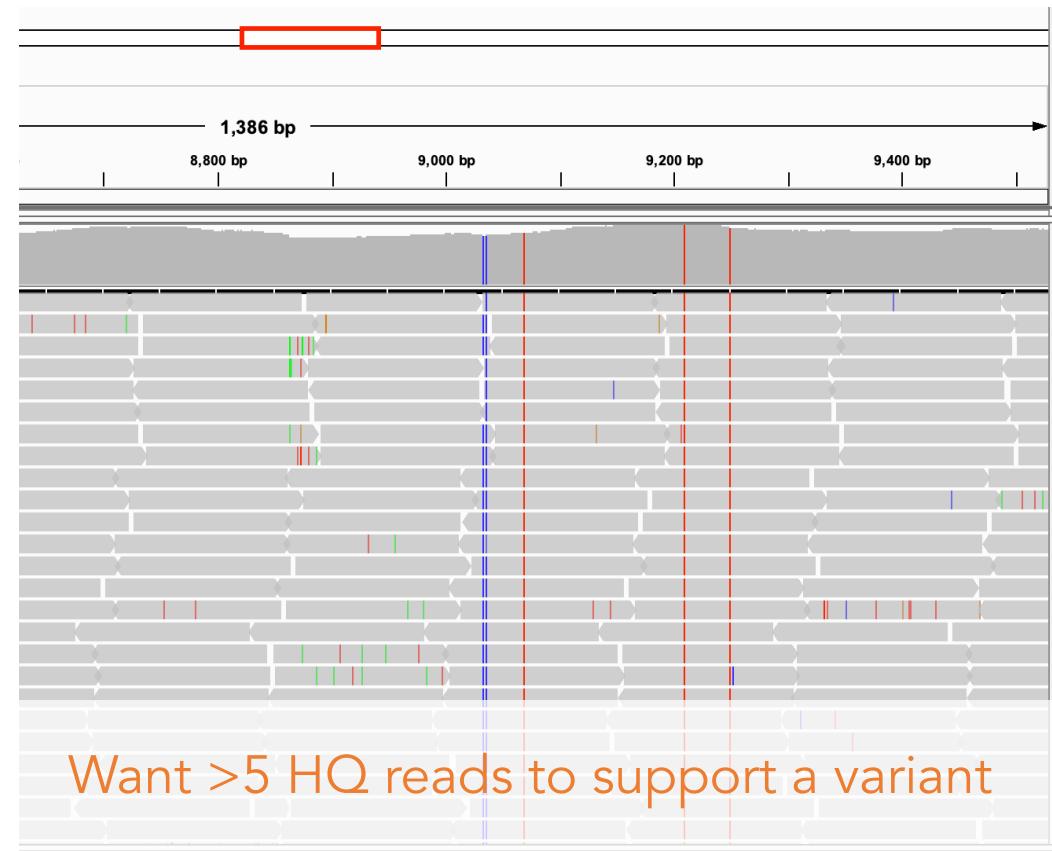


How to confidently identify true variants

Primary considerations

- Sequencing method/approach
- Depth of coverage
- Variant type
- Inheritance mode
- Reference quality
- Bioinformatic tool

Goldilocks Principle of Read Depth

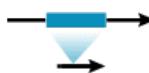


Structural Variant Detection

Single Nucleotide
Variant



Deletion



Insertion



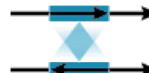
Tandem
Duplication



Interspersed
Duplication



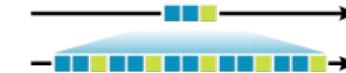
Inversion



Translocation



Copy Number
Variant



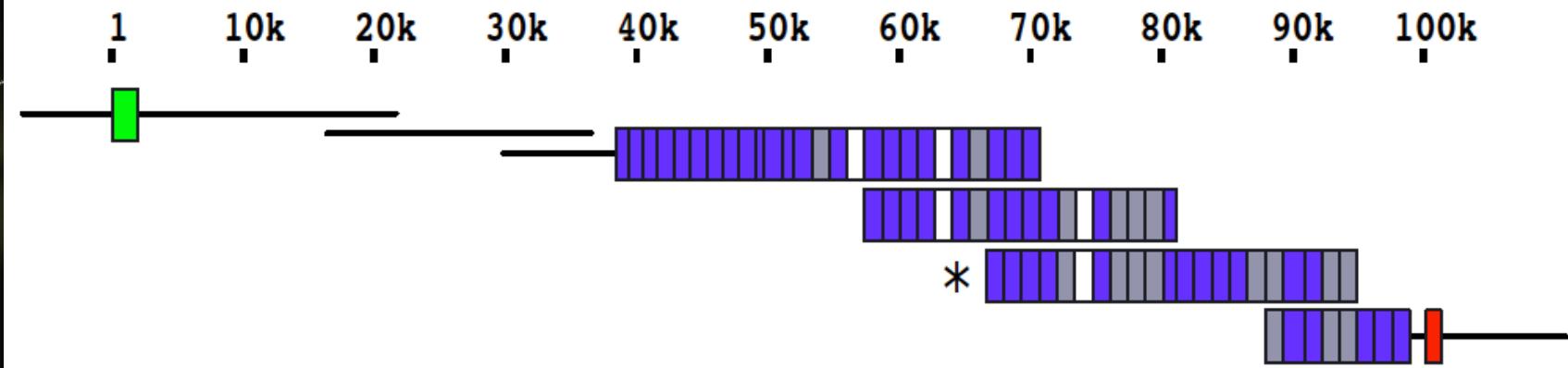
Types of Variants

Reconstructing tandem duplications – problems with consensus

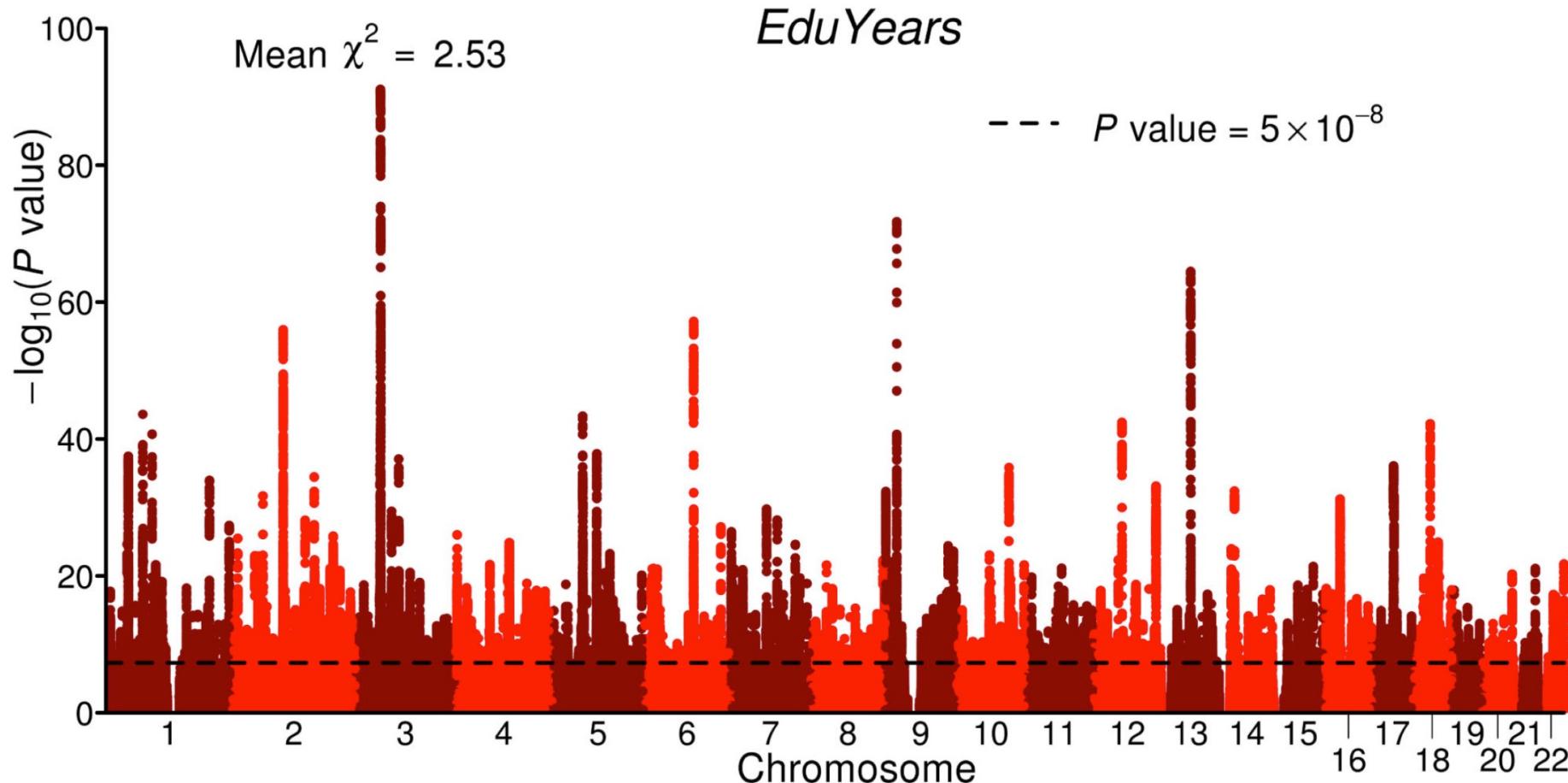


Nanopore Read 1	CCCCGGGAGCTGGTCCCTGATTTGGTTACCGGATATGTCCTTCGATCTGACTTGACT
Nanopore Read 2	CTGGGGGAGCTGGTCCCGATTTGGTTACCGGATATGTCCTTCGATCTGACTTGACT
Nanopore Read 3	CTGGGGGAGCTGGTCCCTGATTTGGTTACCGGATATGTCCTTCGATCTGACTTGACT
Nanopore Read 4	CTGGGGGAGCTGGTCCCTGATTTGGTTACCGGATATGTCCTTCGATCTGACTTGACT
Nanopore Read 5	CTGGGGGAGCTGGTCCCTGATTTGGTTACCGGATATGTCCTTCGATCTGACTTGACT
Nanopore Consensus	CTGGGGGAGCTGGTCCCTGATTTGGTTACCGGATATGTCCTTCGATCTGACTTGACT
Illumina Consensus	GTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 1	CTGGGGGAGCTGGTCCCTGATTTGGTTACCGGATATGTCCTTCGATCTGACTTGACT
Illumina Read 2	CTGGGGGAGCTGGTCCCTGATTTGGTTACCGGATATGTCCTTCGATCTGACTTGACT
Illumina Read 3	CTGGGGGAGCTGGTCCCTGATTTGGTTACCGGATATGTCCTTCGATCTGACTTGACT
Illumina Read 4	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 5	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 6	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 7	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 8	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 9	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 10	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 11	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 12	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 13	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 14	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 15	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 16	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 17	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 18	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 19	CTGGTGAGCTGGGCCCTGATTCTGGTTAACGGGATCTTCCTCGATATTGACTTGCGT
Illumina Read 20	CTGGTGAGCTGGGCCCTGATTCTGGTTACCGGATCTTCCTCGATATTGACTTGCGT

Reconstructing tandem duplications – problems with consensus



Genome-wide association studies



Genome-wide association studies

- Identify SNPs that are associated with phenotype(s) of interest

H_0 : Phenotype \perp SNP

A

H_A : Phenotype \sim SNP

ALIGNED SEQUENCES

LEFT SNP	RIGHT SNP	
... A C AT G C C G A C A T T C A T A G G C C ...		180
... A C AT G C C G A C A T T C A T A A G C C ...		175
... A C AT G C C G A C A T T C A T A G G C C ...		170
... A C AT G C C G A C A T T C A T A A G C C ...		165
... A C AT G C C G A C A T T C A T A G G C C ...		160
... A C AT G C C G A C A T T C A T A G G C C ...		145
... A C AT G C C G A C A T T C A T A A G C C ...		140
... A C AT G C C G A C A T T C A T A A G C C ...		130
... A C AT G T C G A C A T T C A T A G G C C ...		120
... A C AT G T C G A C A T T C A T A G G C C ...		115
... A C AT G T C G A C A T T C A T A A G C C ...		110
... A C AT G T C G A C A T T C A T A G G C C ...		110
... A C AT G T C G A C A T T C A T A A G C C ...		105
... A C AT G T C G A C A T T C A T A A G C C ...		100

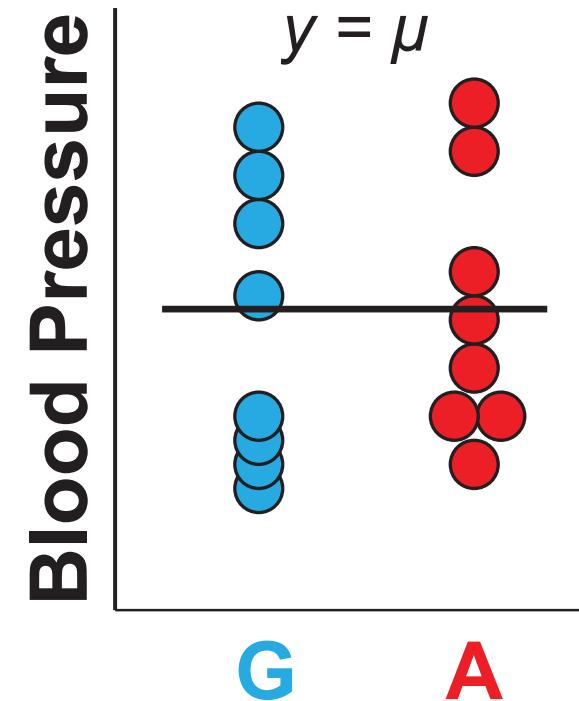
blood pressure

Genome-wide association studies

- Identify SNPs that are associated with phenotype(s) of interest

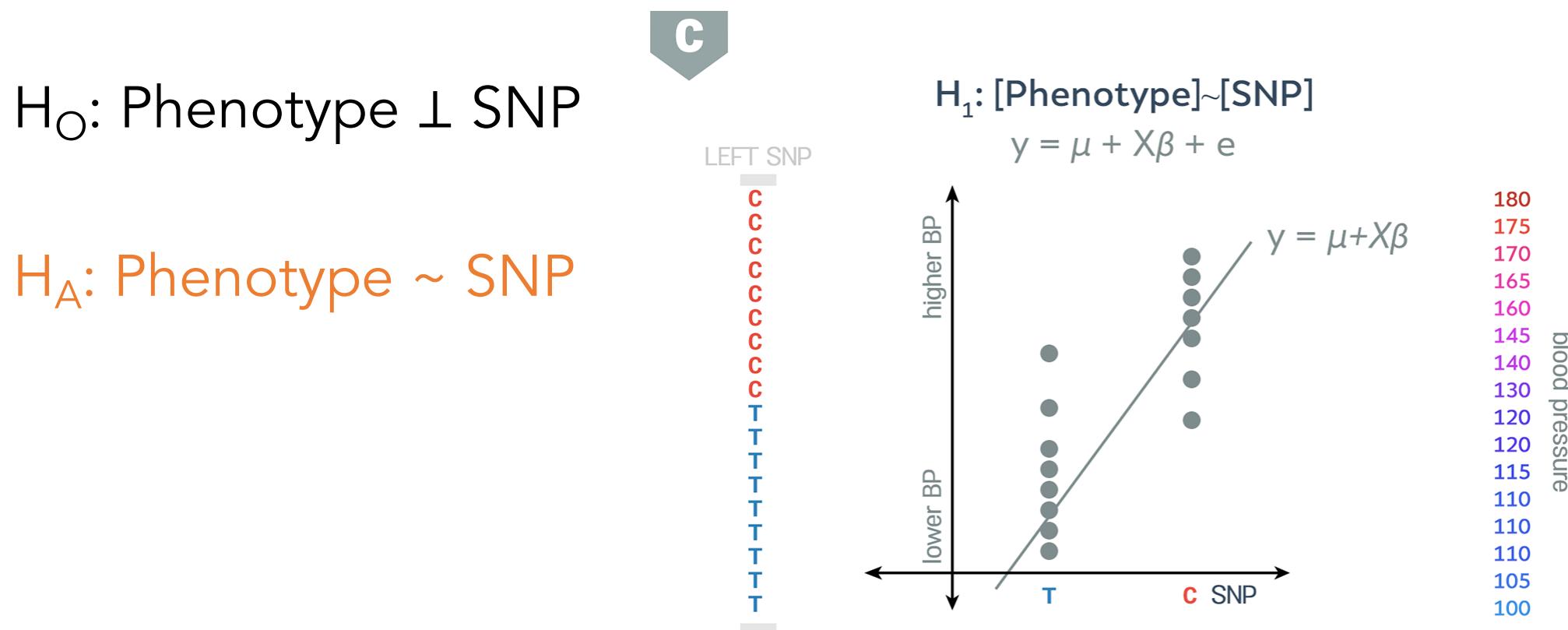
H_0 : Phenotype \perp SNP

H_A : Phenotype \sim SNP



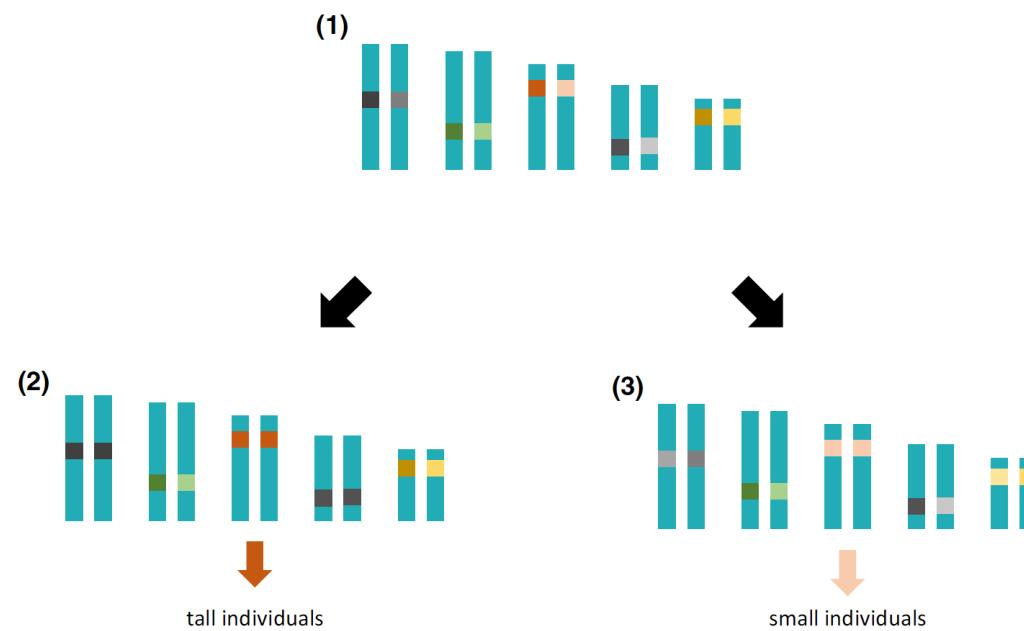
Genome-wide association studies

- Identify SNPs that are associated with phenotype(s) of interest



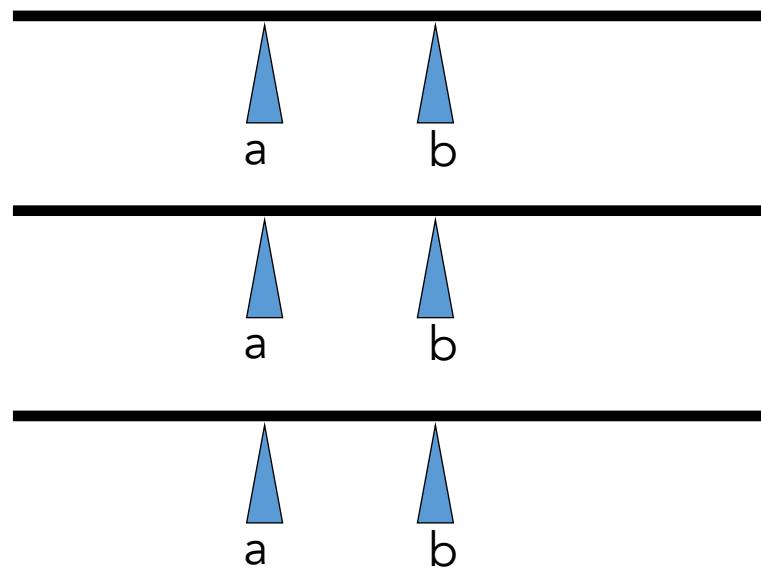
SNP associations may not be due to causal effect

Population structure (differential relatedness among individuals) means that SNPs resulting from Isolation by Descent (IBD) will be associated with the phenotype of interest

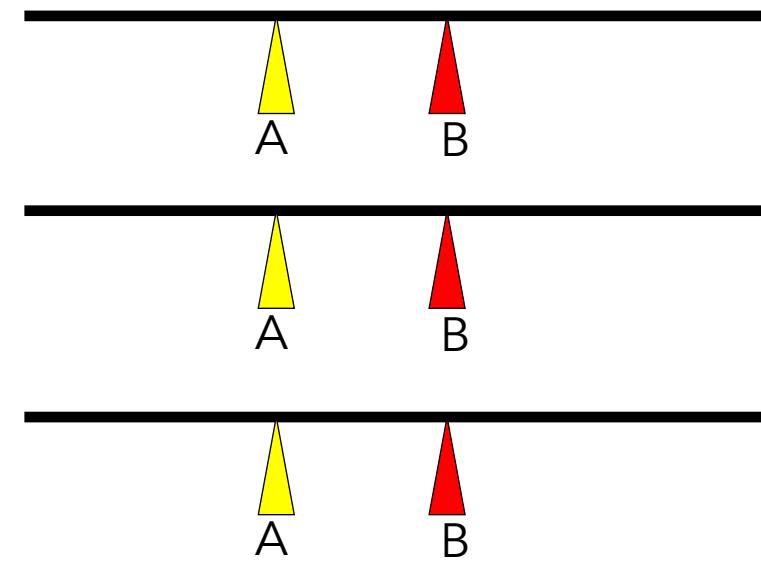


Population structure causes linkage disequilibrium (LD)

Sub-population 1

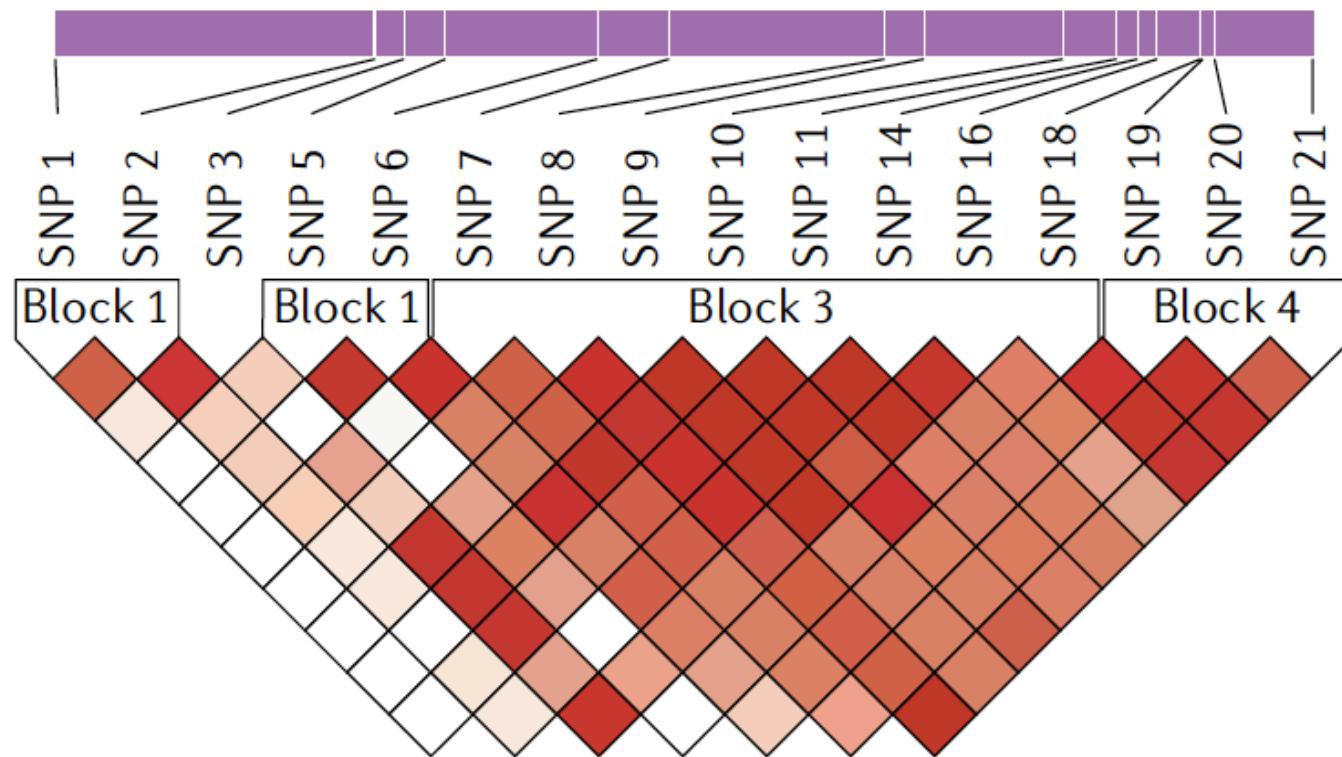


Sub-population 2

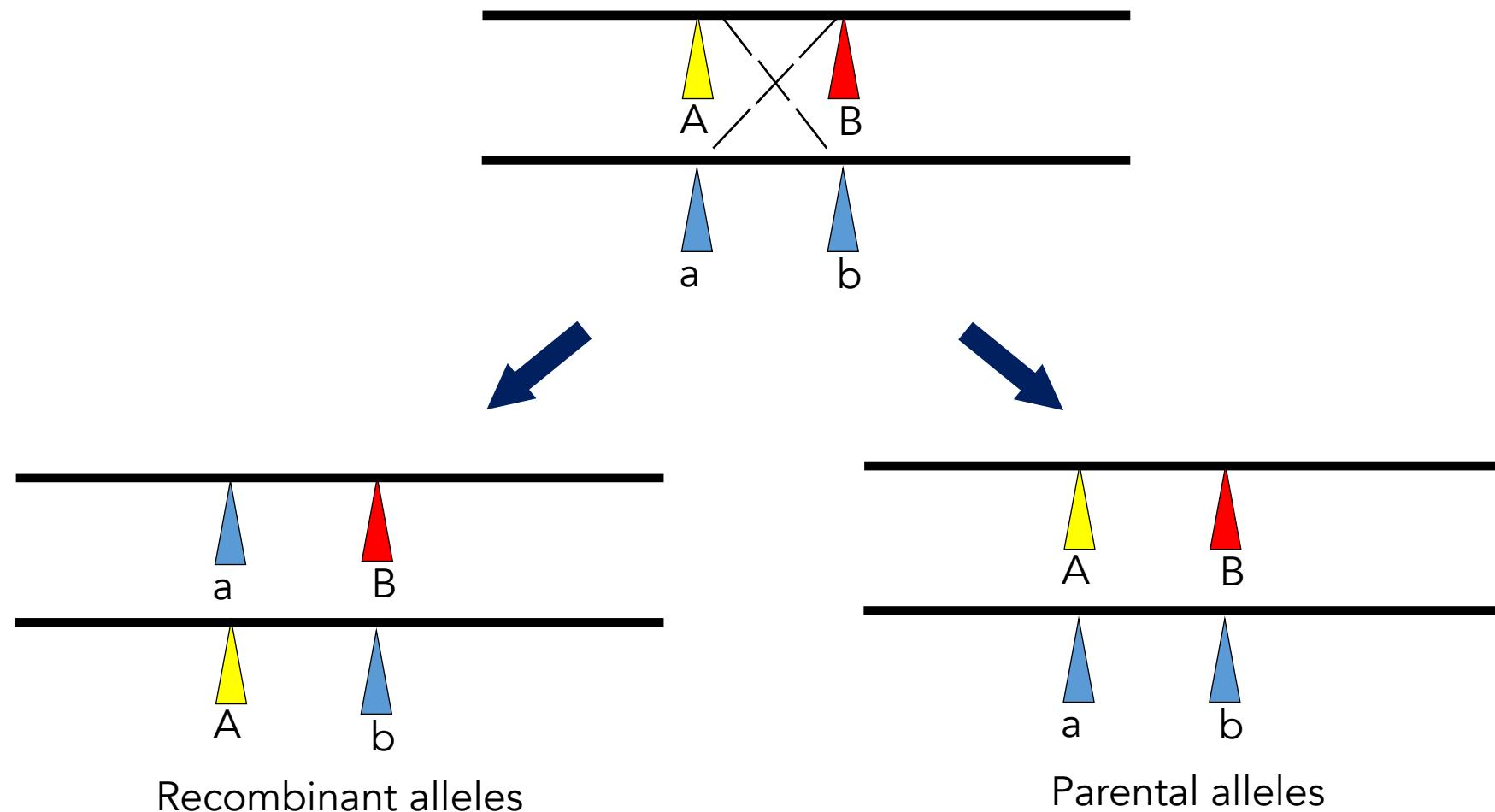


Population structure causes linkage disequilibrium (LD)

Linkage
disequilibrium

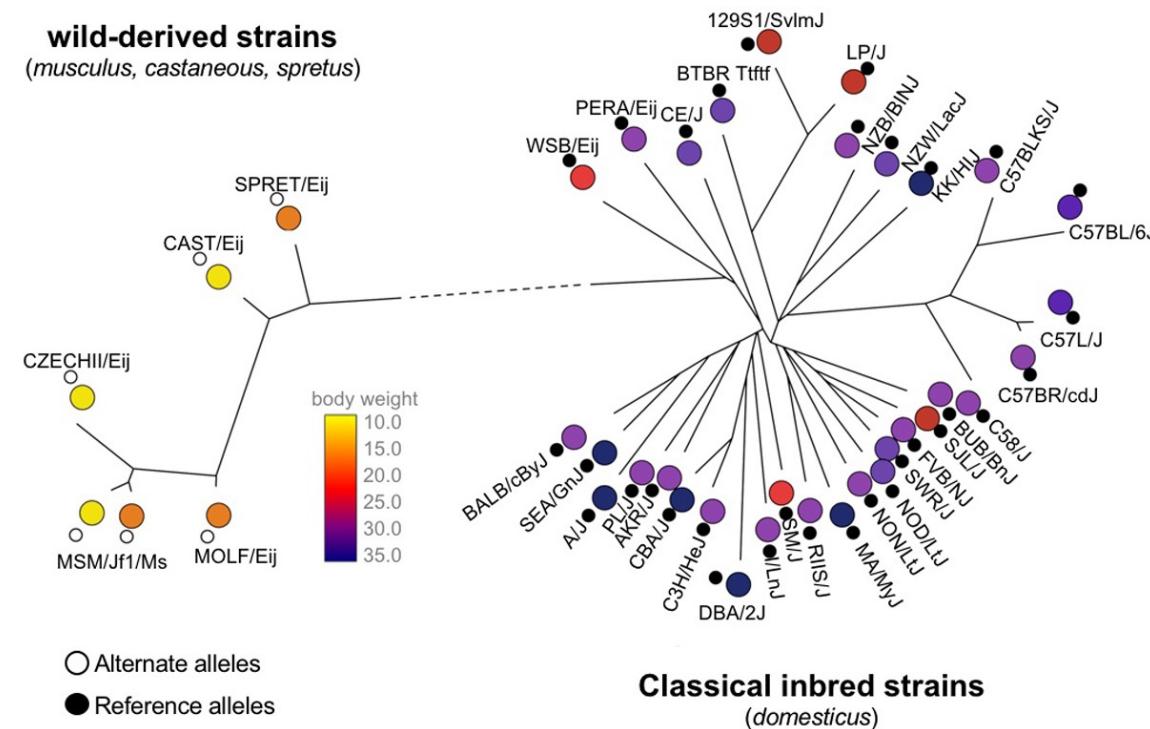


Recombination can break down LD, but it takes time, secondary contact



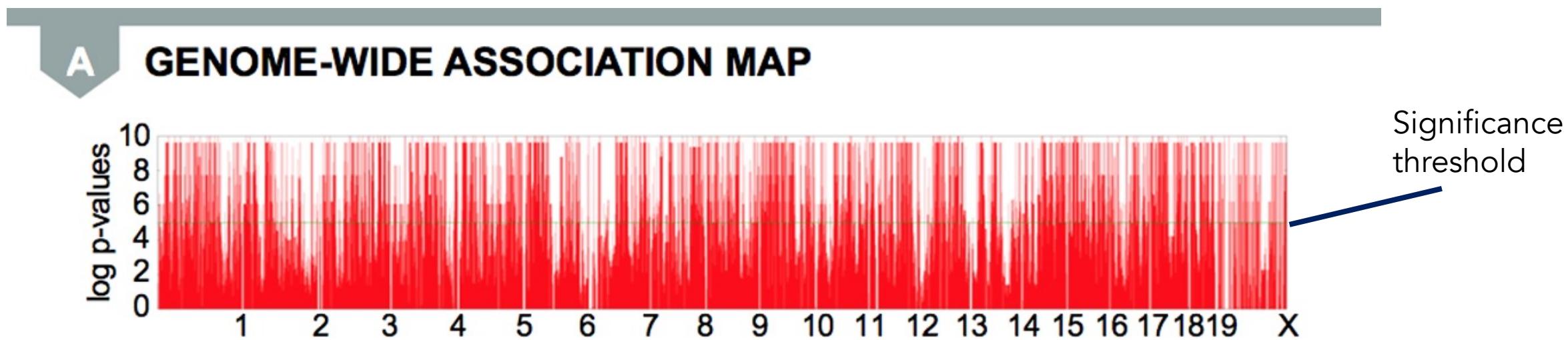
SNP associations may not be due to causal effect

Population structure (differential relatedness among individuals) means that SNPs resulting from Isolation by Descent (IBD) will be associated with the phenotype of interest

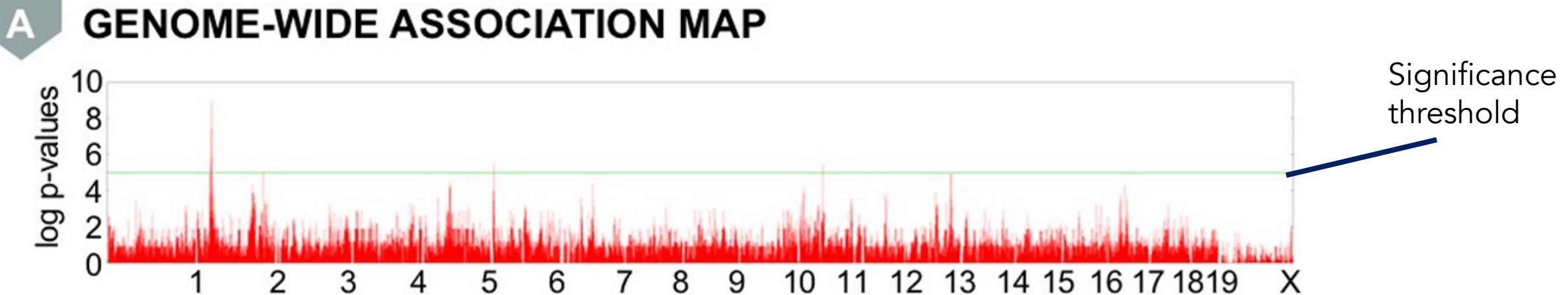


SNP associations may not be due to causal effect

Population structure (differential relatedness among individuals) means that SNPs resulting from Isolation by Descent (IBD) will be associated with the phenotype of interest

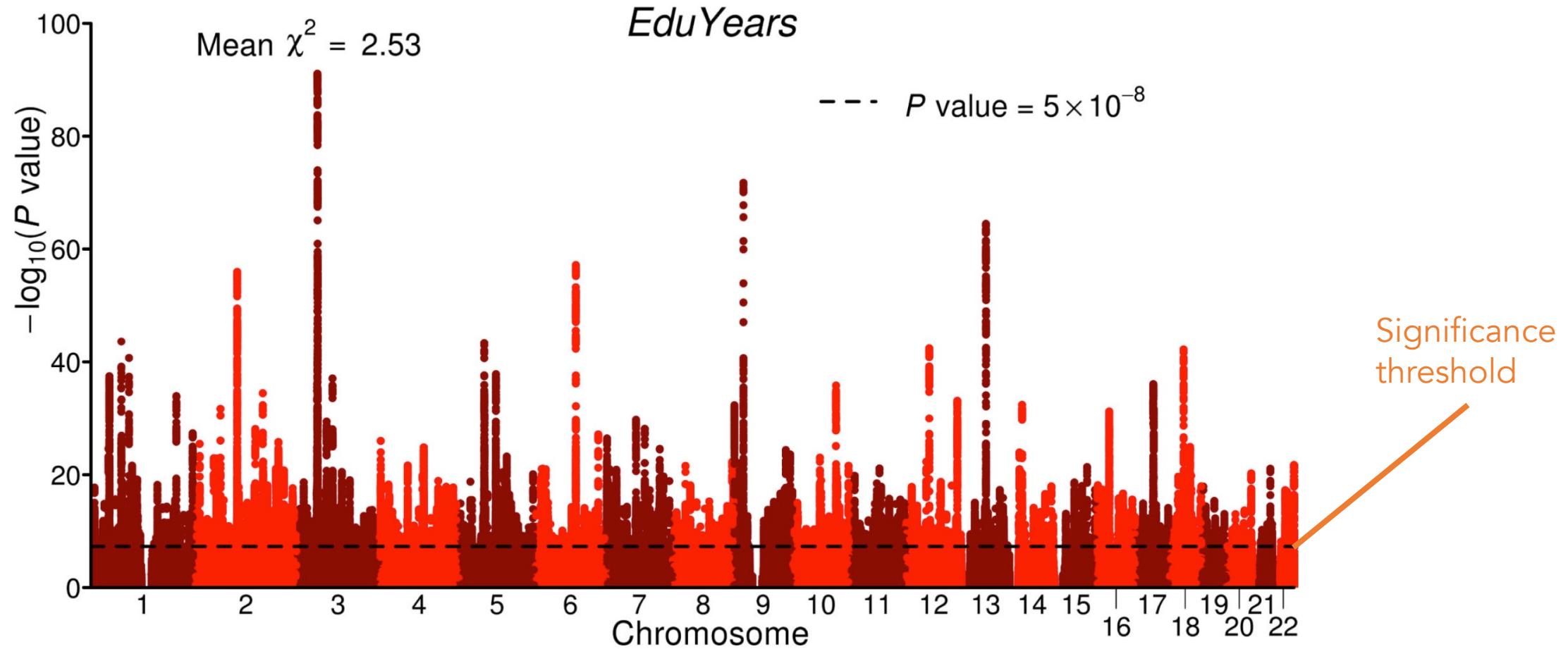


Accounting for population structure can remove spuriously associated SNPs



Sul et al 2018

Doing GWAS with only European populations is a problem

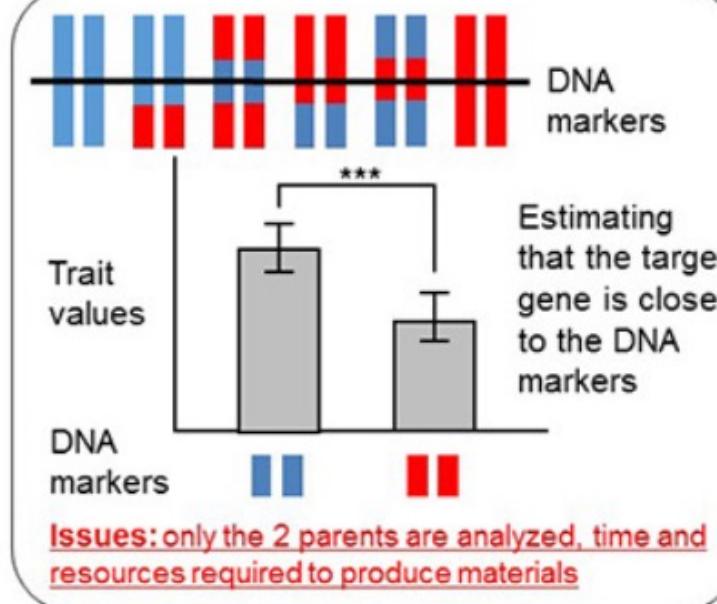


Genetic mapping can take several forms

Quantitative Trait Locus (QTL) Mapping
Genome Wide Association Studies (GWAS)

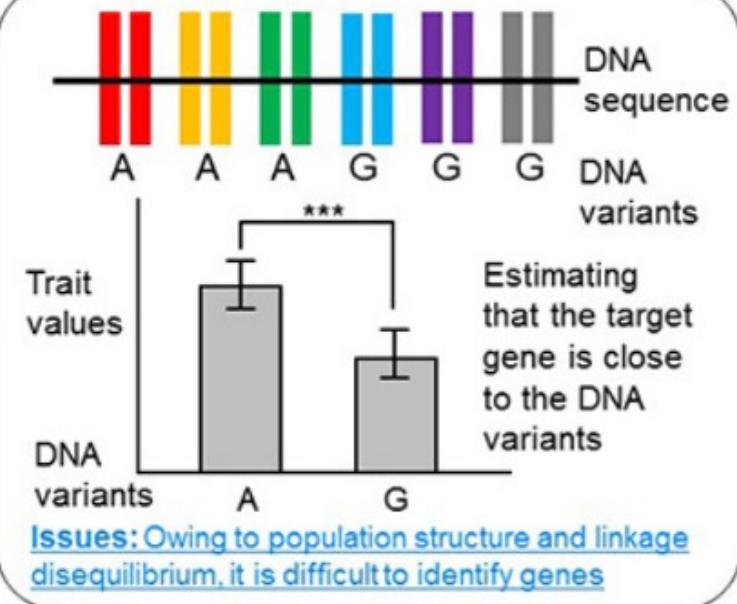
QTL analysis

Crossed population
(F1 intercross for example)



GWAS analysis

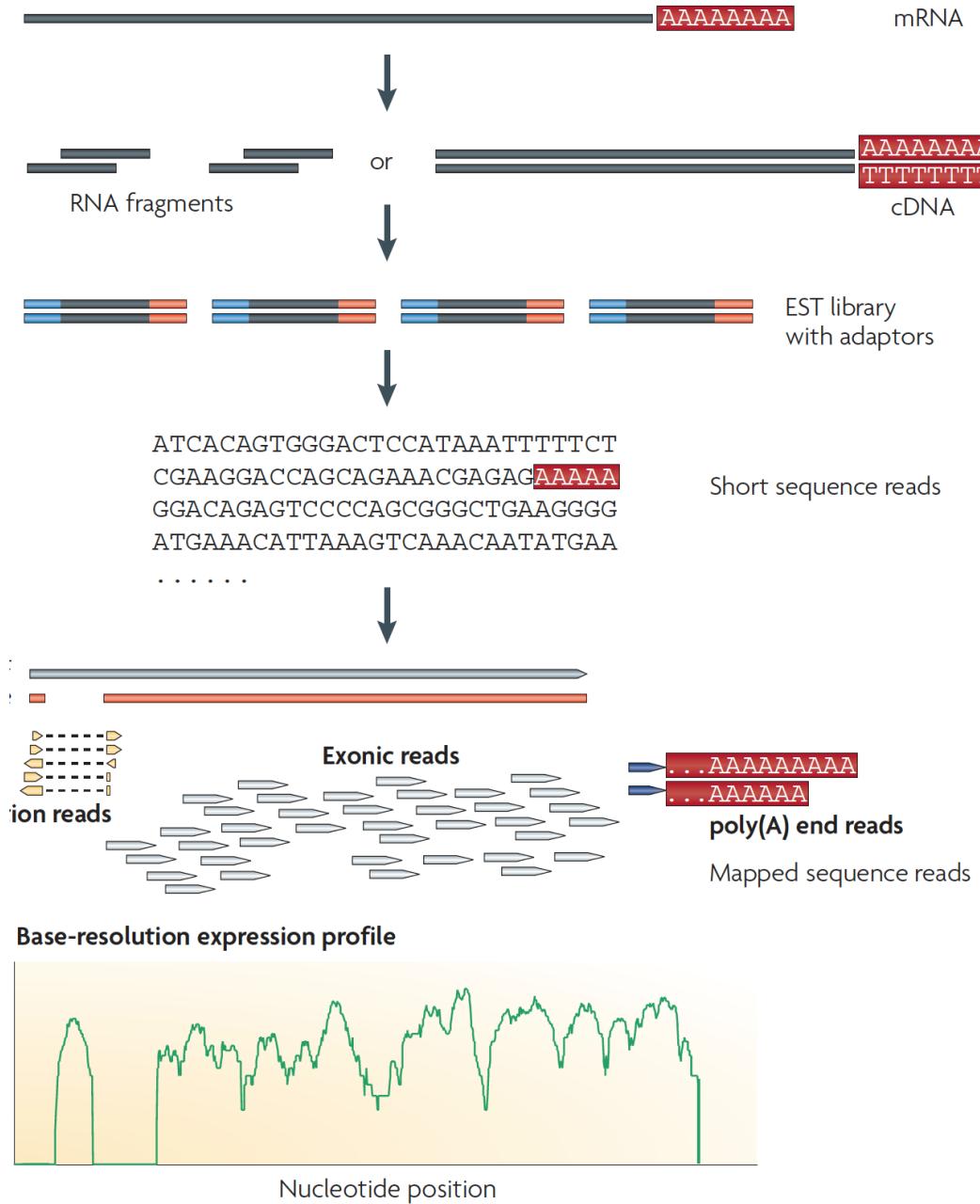
An interbreeding population of individuals of varying relatedness



RNA sequencing and the “transcriptome”

Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project

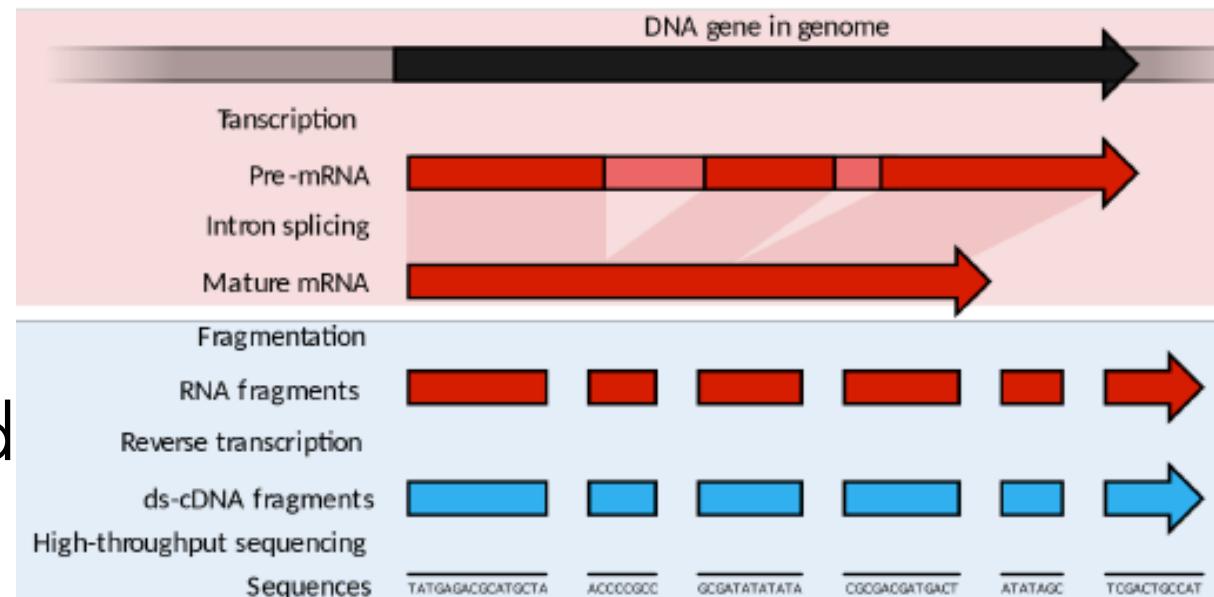
MARK D. ADAMS, JENNY M. KELLEY, JEANNINE D. GOCAYNE, MARK DUBNICK,
MIHAEL H. POLYMEROPoulos, HONG XIAO, CARL R. MERRIL, ANDREW WU,
BJORN OLDE, RUBEN F. MORENO, ANTHONY R. KERLAVAGE,
W. RICHARD McCOMBIE, J. CRAIG VENTER*



Sequencing RNA vs. DNA

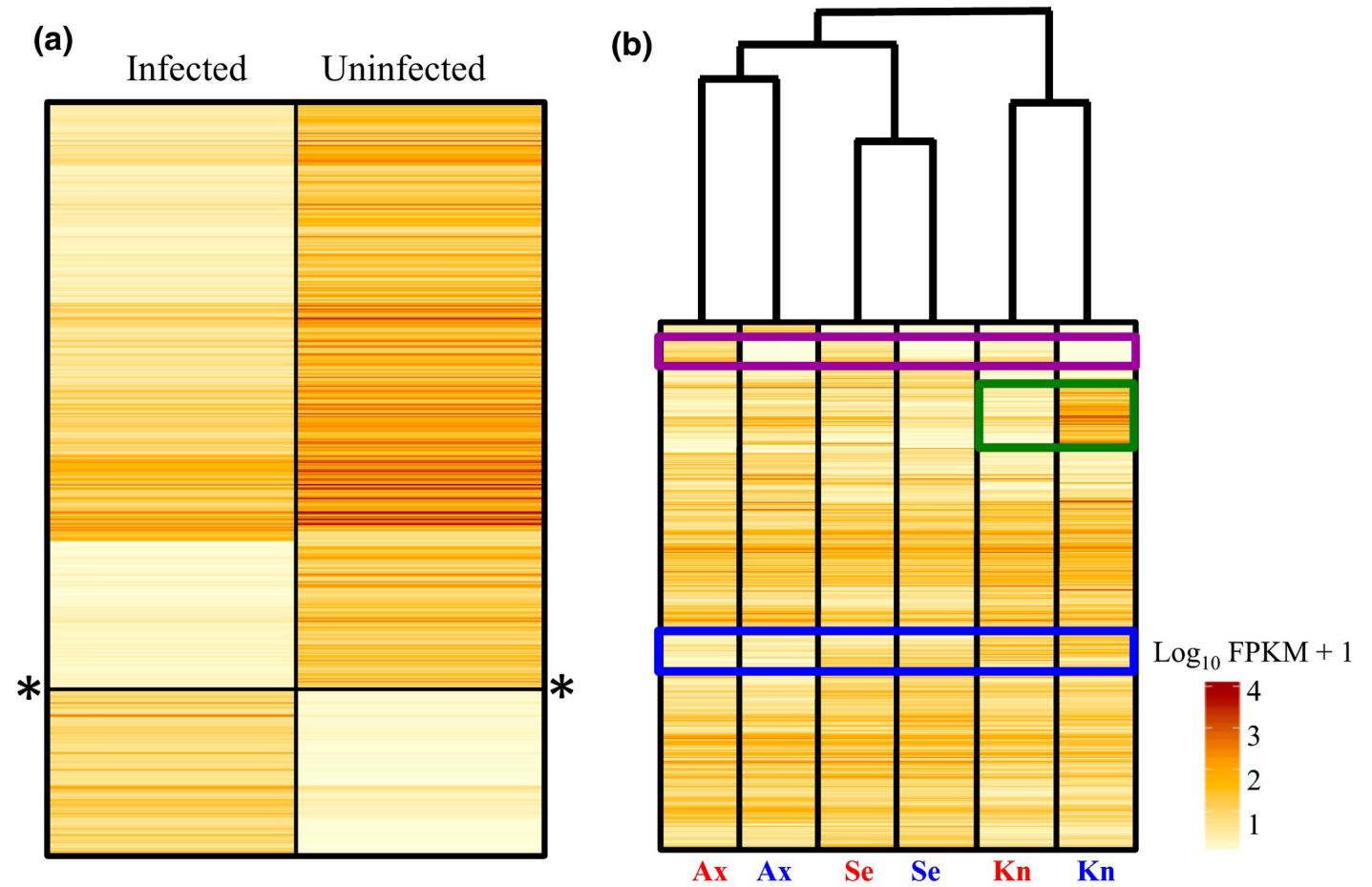
Important differences

- Single-stranded
- Variable copy number
- Post transcriptionally modified
 - Intron splicing
 - RNA editing
 - Polyadenylation (Poly-A tail)



RNAseq Applications

- Quantify gene expression



RNAseq Applications

- Quantify gene expression
- Determine gene sequences in non-model species
(de novo transcriptome assembly)

A



Unisexual *Ambystoma*
(LTTi)



Ambystoma laterale
(LL)



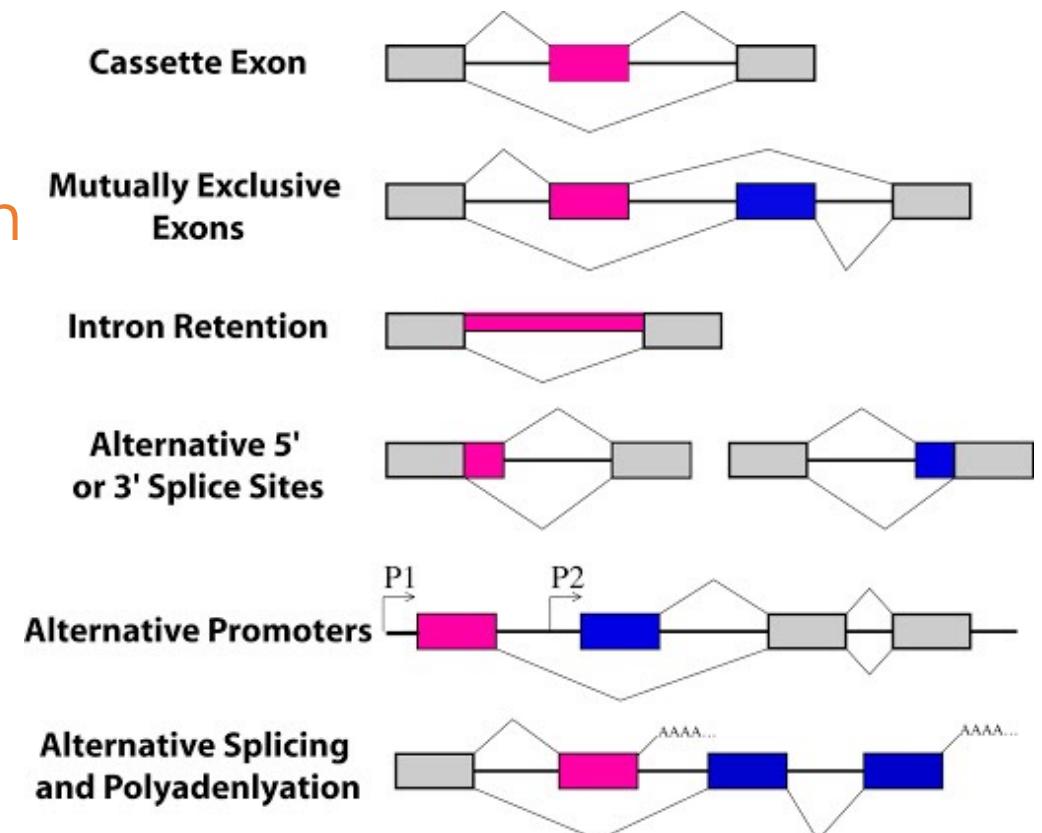
Ambystoma texanum
(TT)



Ambystoma tigrinum
(TiTi)

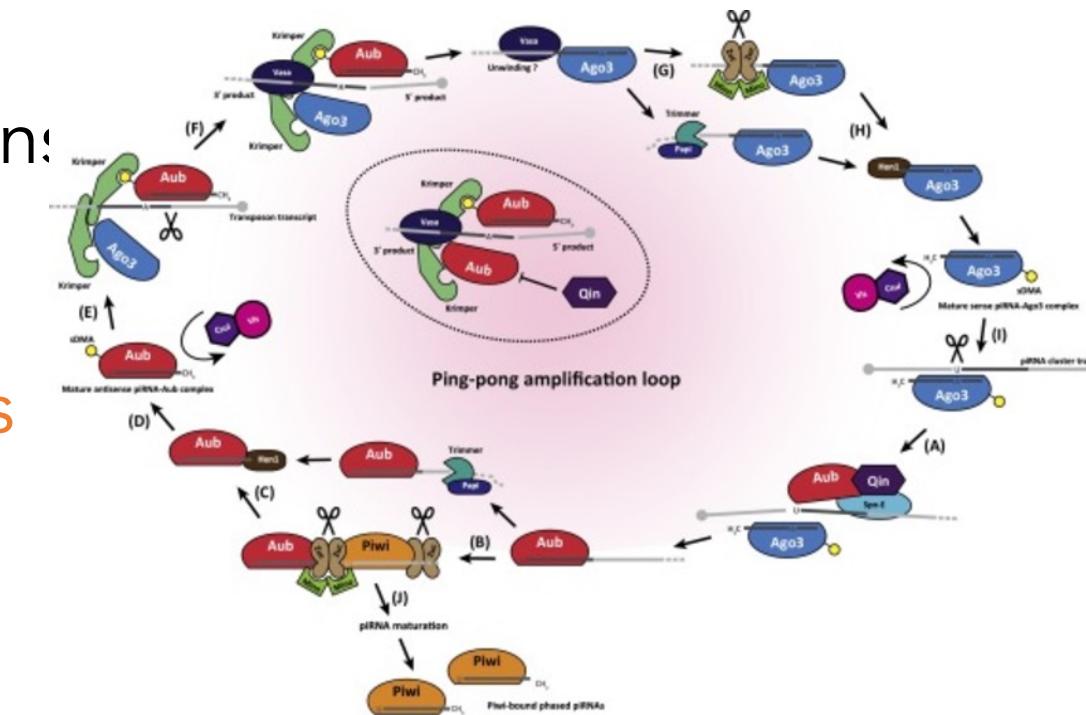
RNAseq Applications

- Quantify gene expression
- Determine gene sequences in non-model species (de novo transcriptome assembly)
- Identify post-transcriptional modification (alternative) splicing, RNA editing



RNAseq Applications

- Quantify gene expression
- Determine gene sequences in non-model species (de novo transcriptome assembly)
- Identify post-transcriptional modifications: (alternative) splicing, RNA editing
- Identify functional/transcribed elements (non-coding RNAs)

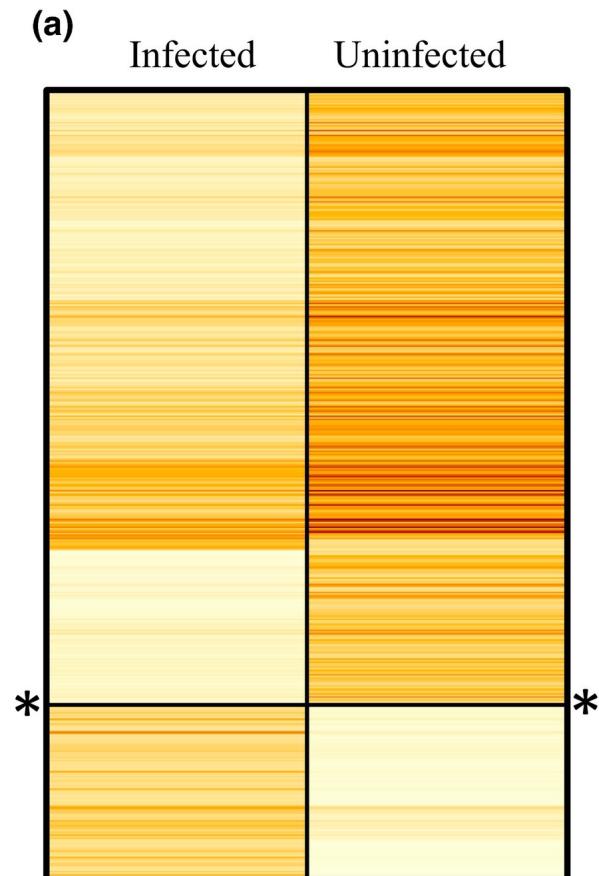


RNAseq Applications

- Quantify gene expression
- Determine gene sequences in non-model species
(de novo transcriptome assembly)
- Identify post-transcriptional modifications: (alternative) splicing, RNA editing
- Identify functional/transcribed elements
(non-coding RNAs)
- Identify targets or footprints of RNA binding proteins

RNAseq Applications

- Quantify gene expression
- Determine gene sequences in non-model species (de novo transcriptome assembly)
- Identify post-transcriptional modifications: (alternative) splicing, RNA editing
- Identify functional/transcribed elements (non-coding RNAs)
- Identify targets or footprints of RNA binding proteins



Bankers et al., 2017 Mol Ecol