

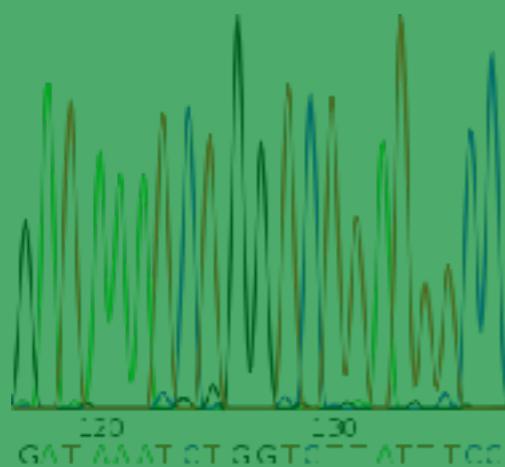
RAD sequencing

Methods and application

A large variety of “genomic” markers

LOW
INFORMATION &
LOW COST per
sample

microsatellites



Sanger sequencing

Reduced genome
complexity using
Restriction enzymes

RAD, ddRAD, GBS, ...

Amplicon sequencing
(e.g. eDNA)



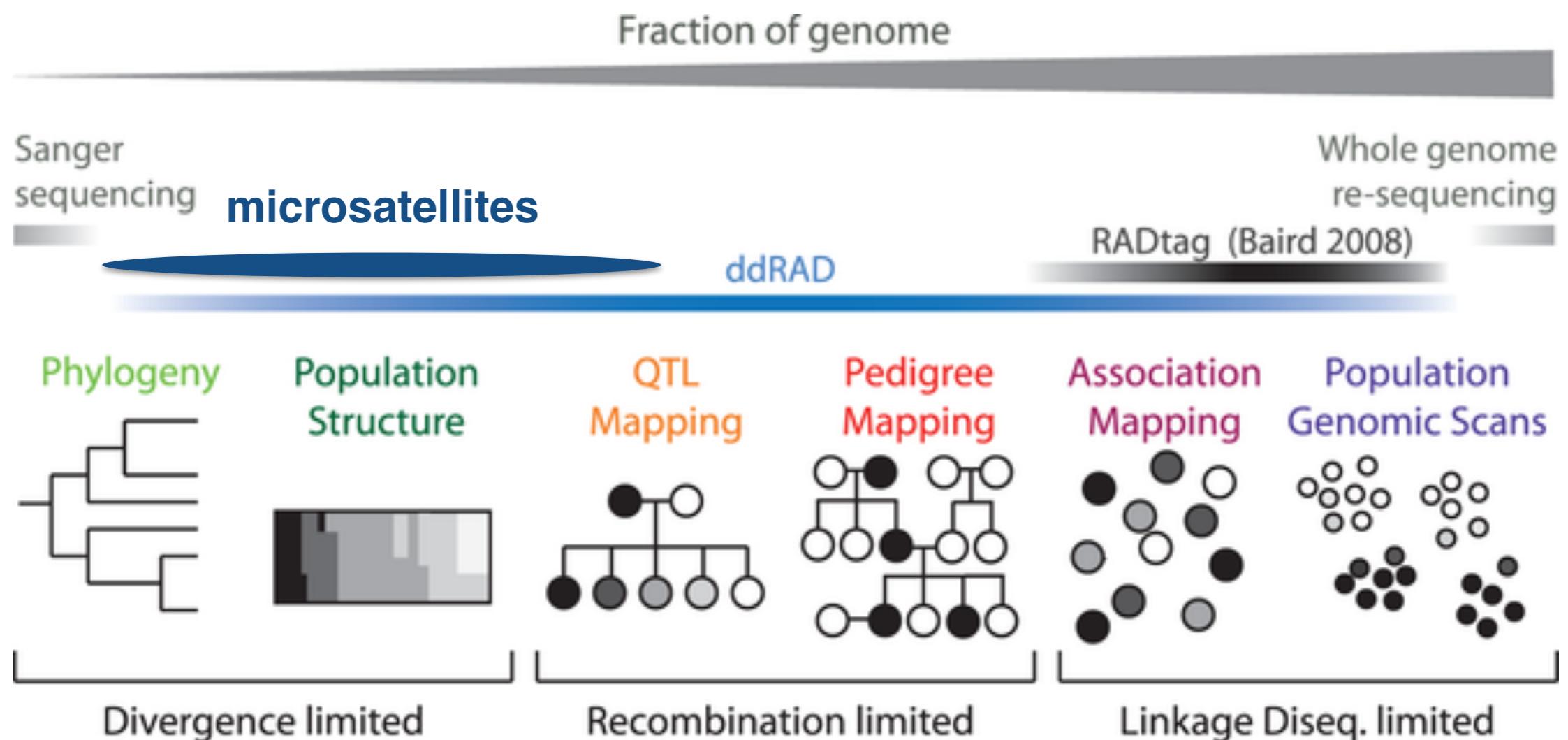
SNP chips/arrays

HIGH
INFORMATION & HIGH COST per sample



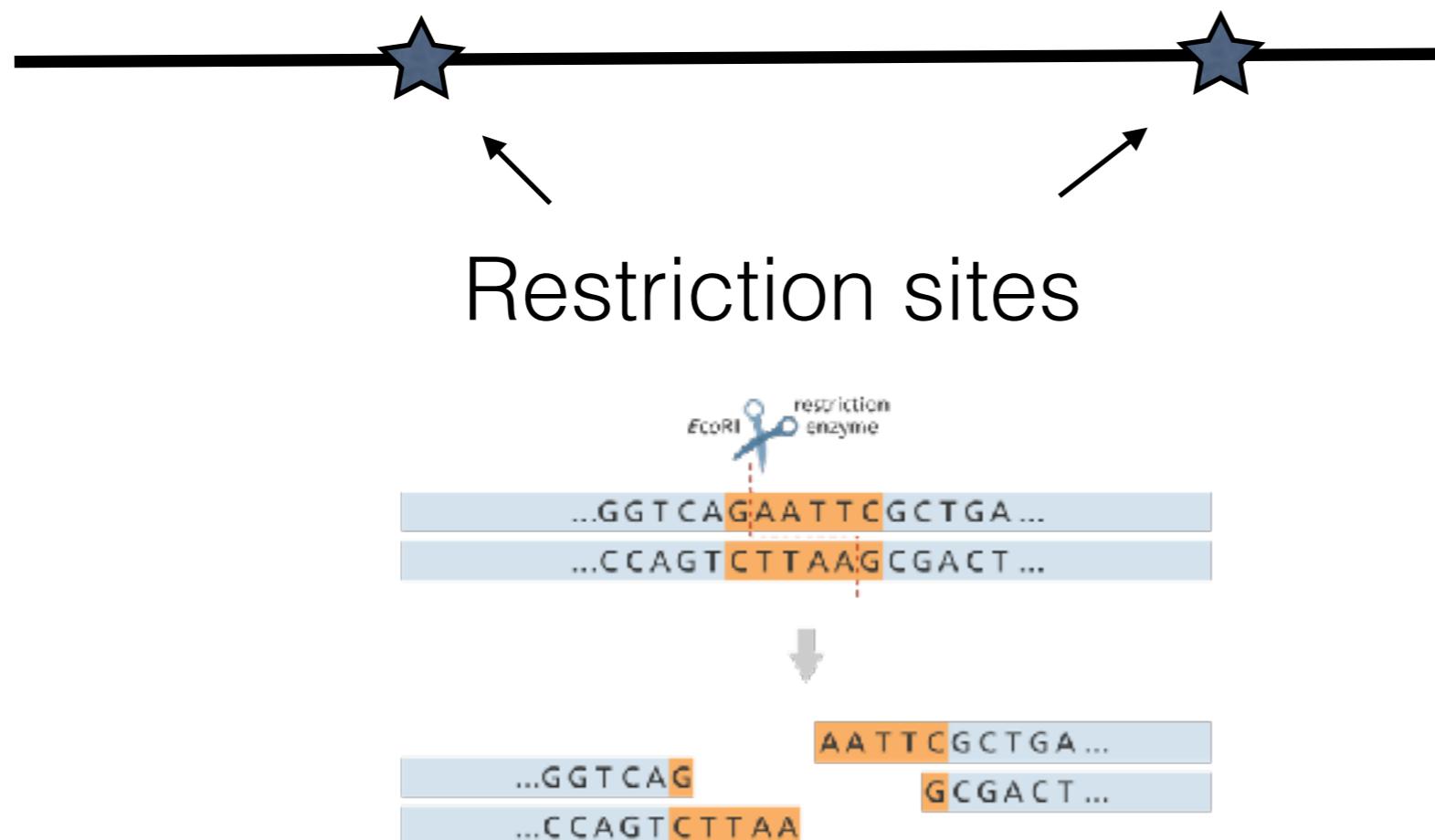
whole genome

A large variety of “genomic” markers



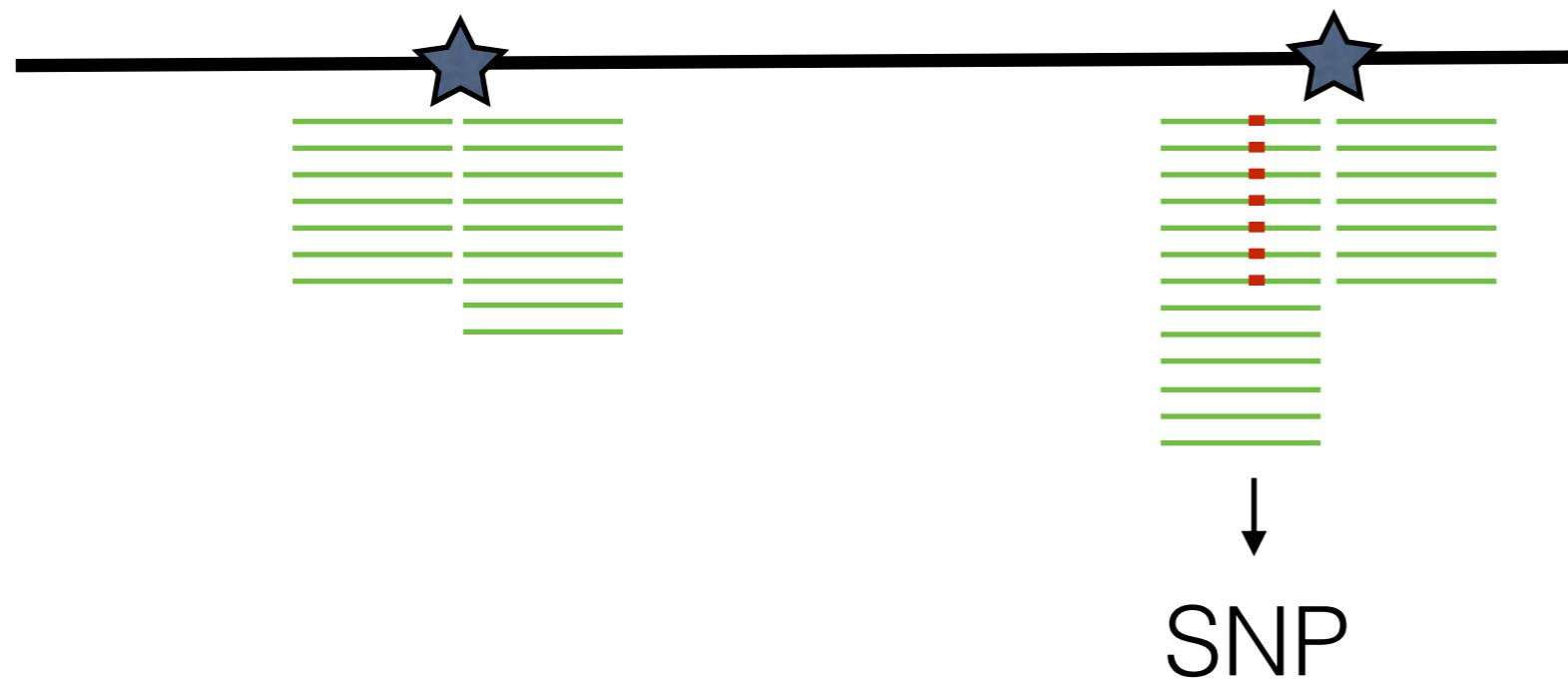
What is RAD?

RAD: Restriction site associated DNA sequencing:
= Sequencing of flanking sites of AFLP fragments

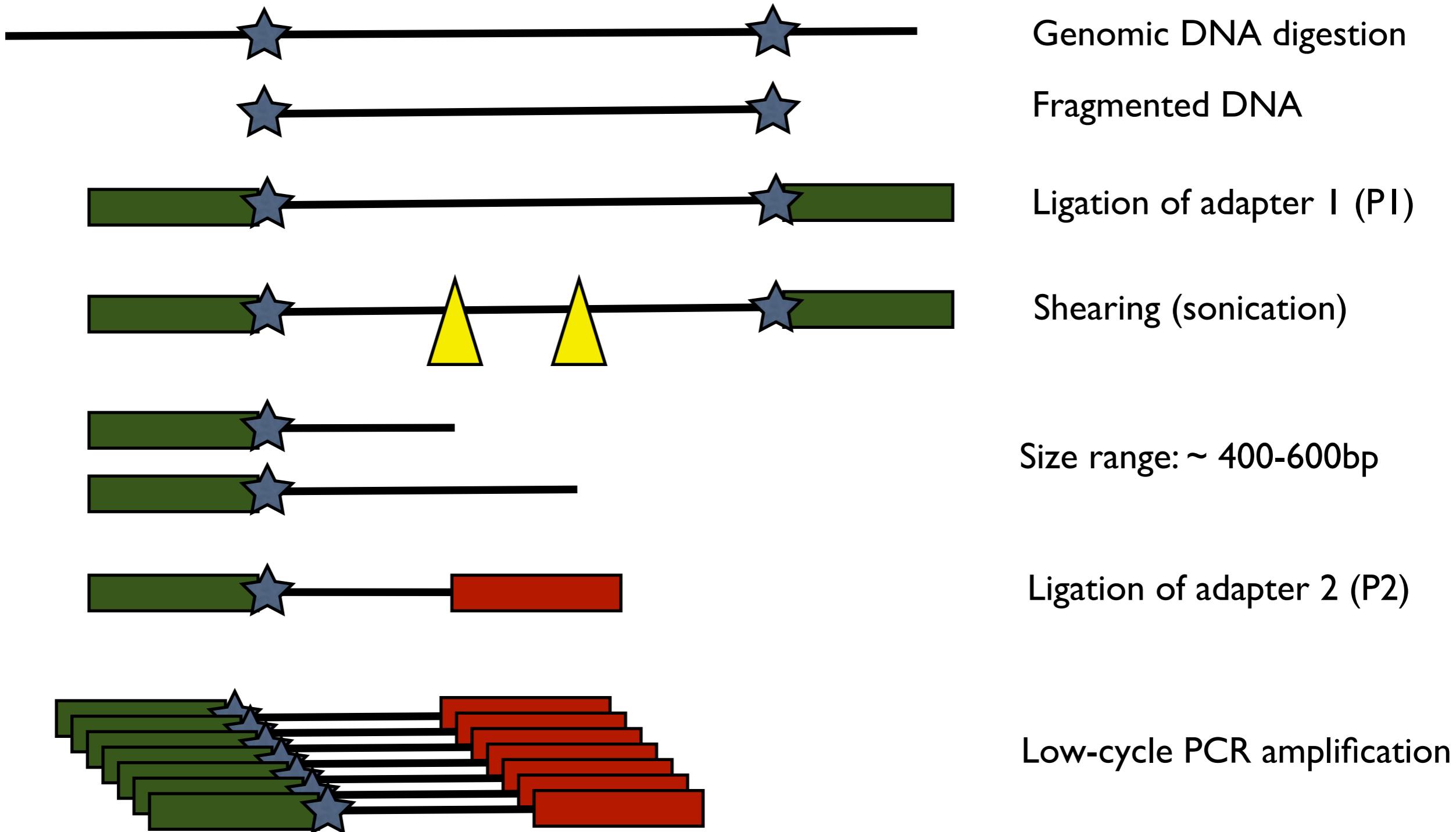


What is RAD?

RAD: Restriction site associated DNA sequencing:
= Sequencing of flanking sites of AFLP fragments



original RAD (or mbRAD, Miller et al. 2007; Baird et al. 2008)



More flavours: ddRAD (double digest, Peterson et al. 2012)



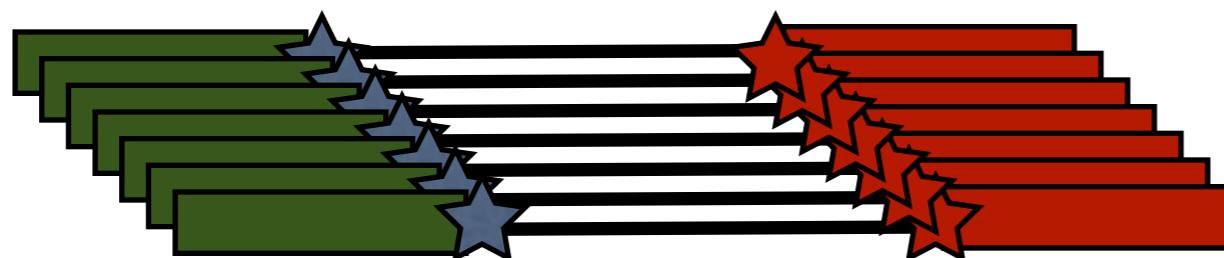
Genomic DNA digestion
with two enzymes



Fragmented DNA



Ligation of adapters



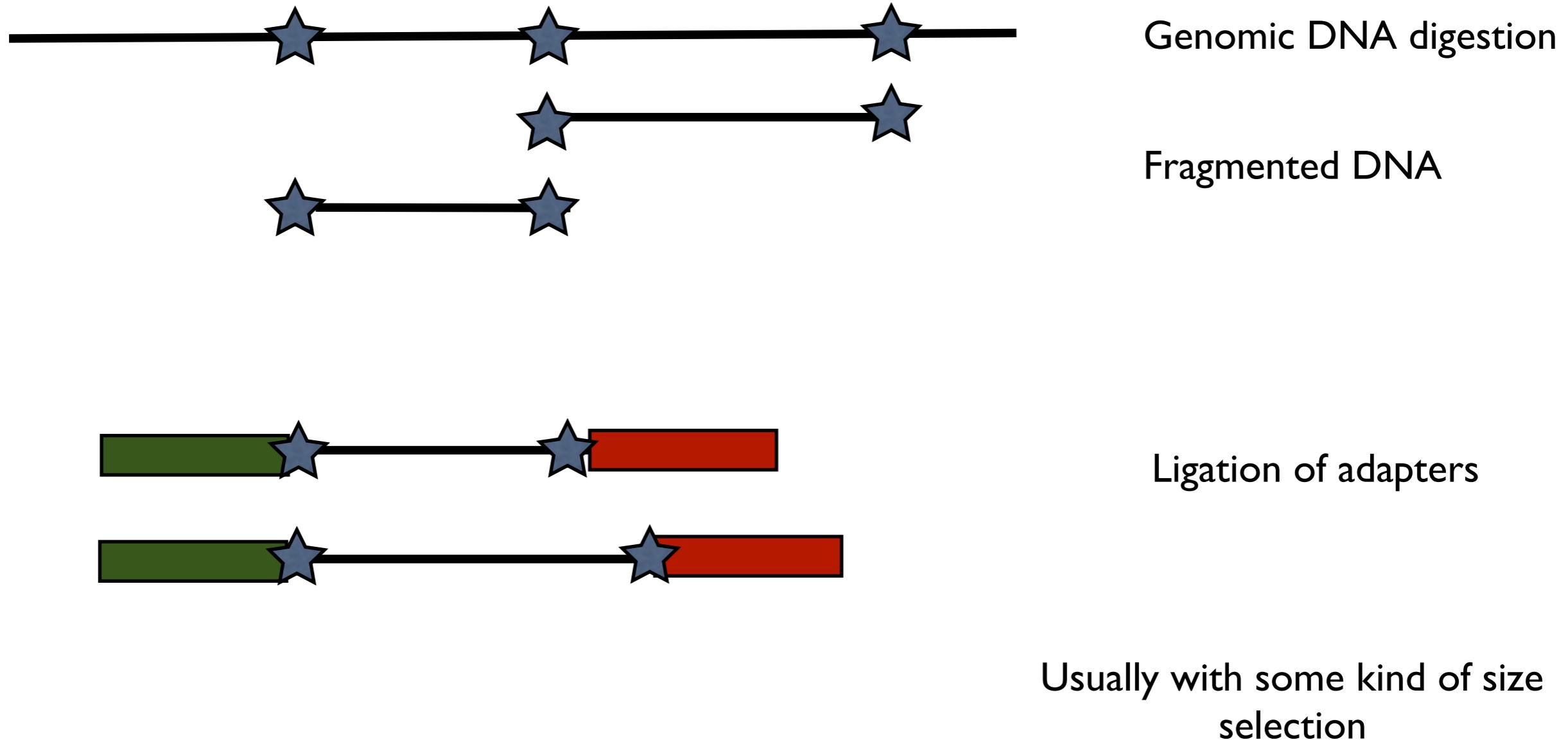
With or without size selection

Low-cycle PCR amplification

ddRAD allows flexibility in marker number

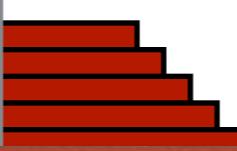


More flavours: GBS (single digest, Elshire et al. 2011)



Yet other RAD flavours:

cRAD (Tsoopon et al. 2012)



Low-cycle PCR amplification

License fee ([http://www.keygene.com/get-your-license-here/!](http://www.keygene.com/get-your-license-here/))

Experimental design: which method?

Study aim (number of markers versus number of individuals)?

Population genetics: e.g. differentiation along genome, selection detection

Genome-wide association (GWAS)

Phylogenetics

Study organism?

Available genomic resources (e.g reference genome)?

Genome size and estimated number of cut sites: coverage per individual, individuals per lane

Ploidy

Repetitive elements

Lab experience and knowhow

Budget

Experimental design: many decisions

Which restriction enzyme (frequent vs low cutter)?

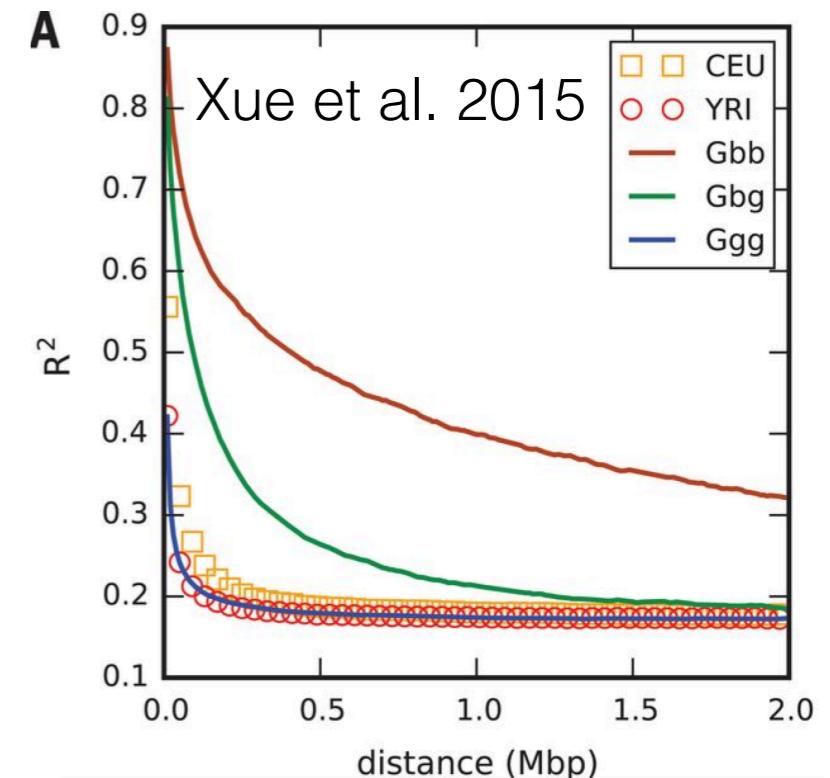
How many markers, marker density?

LD block size?

How many individuals?

Coverage needed (allele frequencies versus genotypes)?

Paired-end vs single-end sequencing?



Lab facilities in Zürich

GDC

Adapter aliquots (RAD & ddRAD)
Covaris (shearing)
Caliper (size selection)
BioAnalyzer (quality control and quantification)
Qubit (quantification)
qPCR (quality control and quantification)
MiSeq
Help & advice!

FGCZ

HiSeq (8 lanes = 1 flowcell; up to 300 mio 150 bp read pairs = ~90Gb)
MiSeq (1 lane; 25-30 Mio 300 bp read pairs = ~18Gb)

May also consider to outsource
library prep and sequencing

Lab considerations

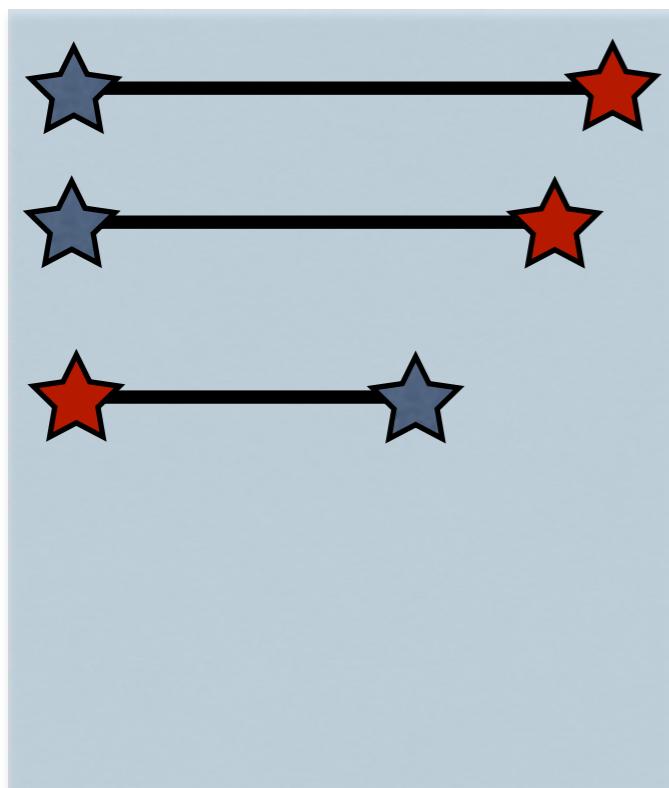
Need high quality DNA (preferably tissue)

Large quantities of DNA optimal, but possible with 50 to 100 ng

Taxa specific protocol optimisation

Once protocol works, ~1 week of work

Variation in size selection, most of all important for GBS, ddRAD



Lab considerations

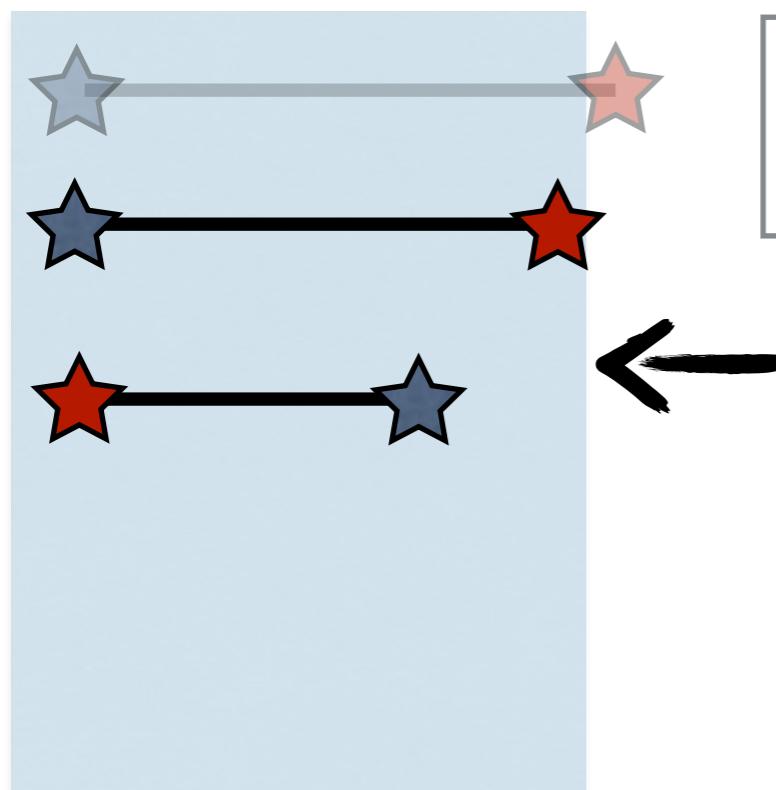
Need high quality DNA (preferably tissue)

Large quantities of DNA optimal, but possible with 50 to 100 ng

Taxa specific protocol optimisation

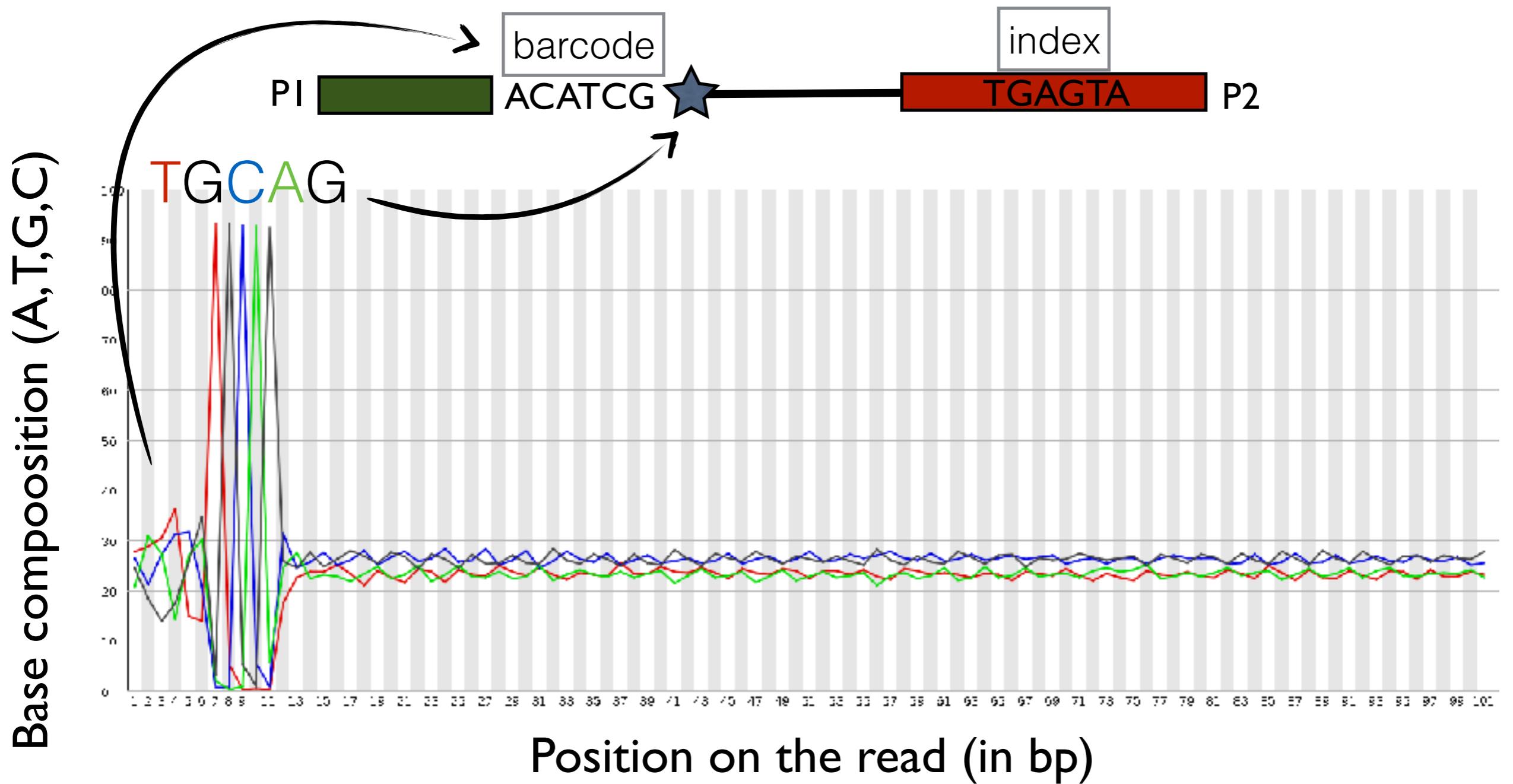
Once protocol works, ~1 week of work

Variation in size selection, most of all important for GBS, ddRAD



shift in selected size leads
to different set of markers

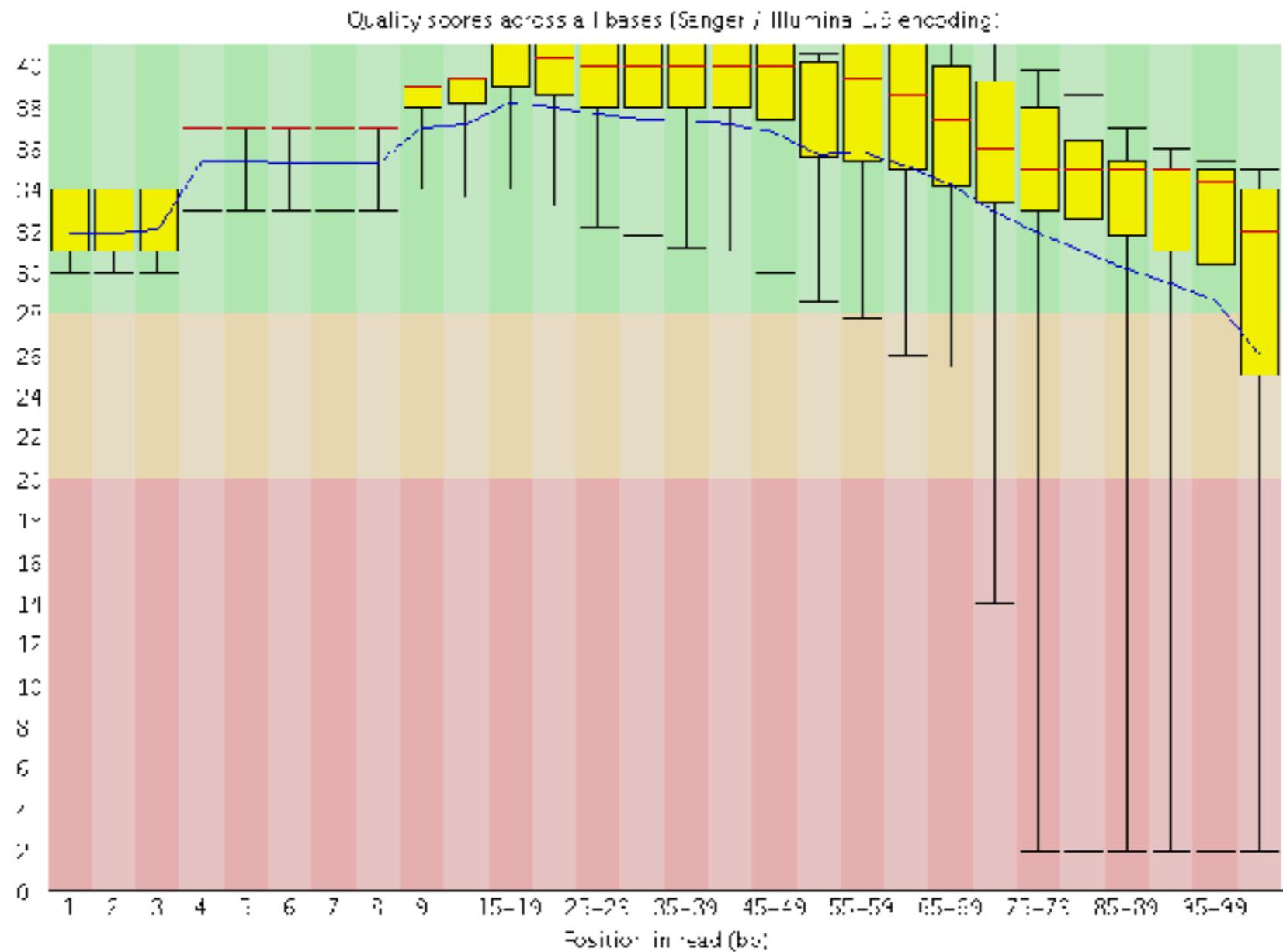
Problem of low complexity



Addition of PhiX alleviates problem

Data analysis: sequence quality

Adapter removal, quality trimming



Data analysis: barcode splitting

Adapter removal, quality trimming



(Demultiplexing), Barcode splitting



Barcode OR Barcode + cutsite

ACATCGTGCAGGAACCCAGTGTGGACCGGGCAGGCAGACCCGCACGACTGCTTACTGCTAA....

CACATCGTGCAGGAACCCAGTGTGGACCGGGCAGGCAGACCCGCACGACTGCTTACTGCTAA....

ACATGGTGCAGGAACCCAGTGTGGACCGGGCAGGCAGACCCGCACGACTGCTTACTGCTAA....

Allow some mismatch

Data analysis: align reads to reference

Adapter removal, quality trimming



Demultiplexing, Barcode splitting

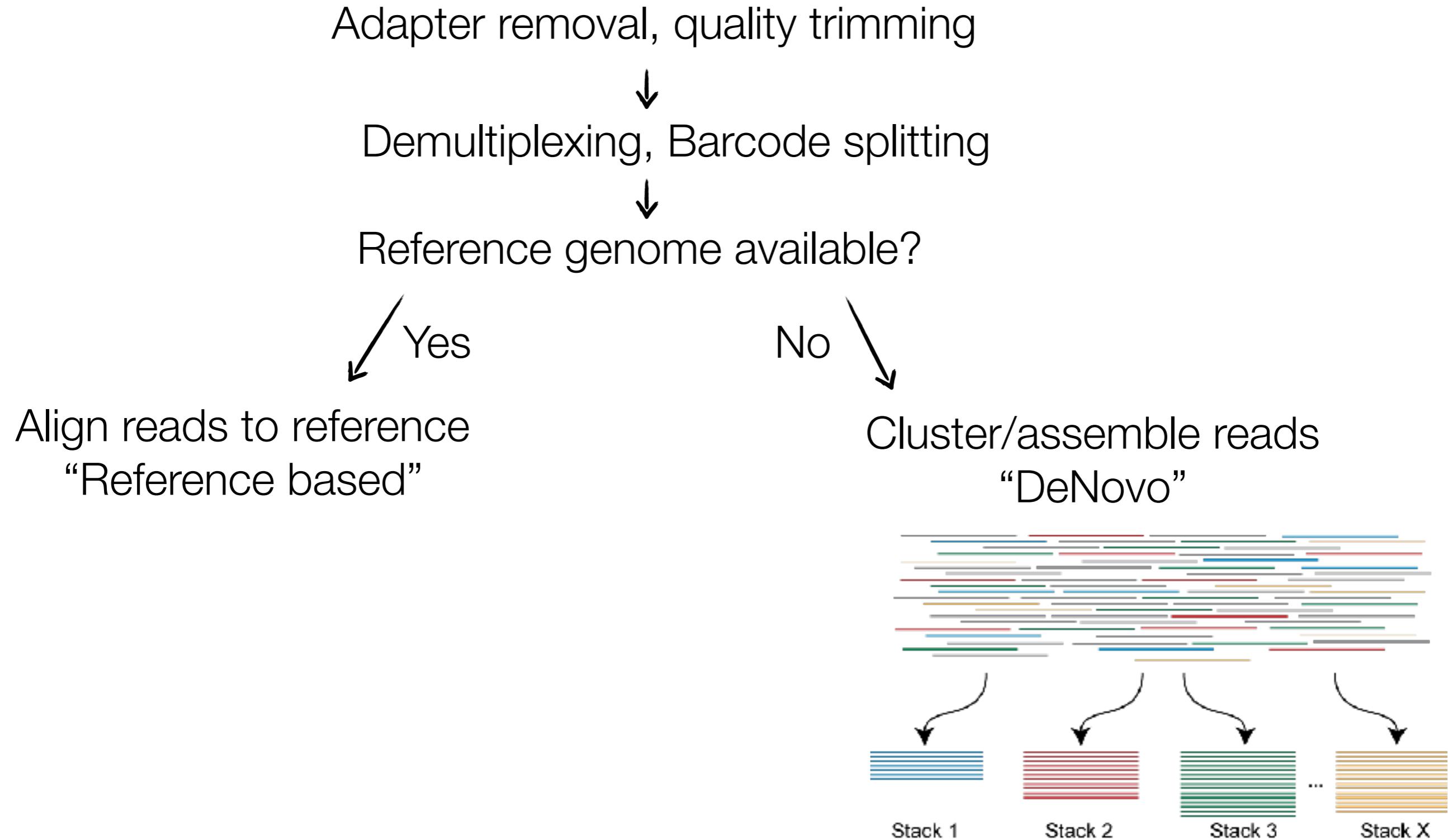


Reference genome available?

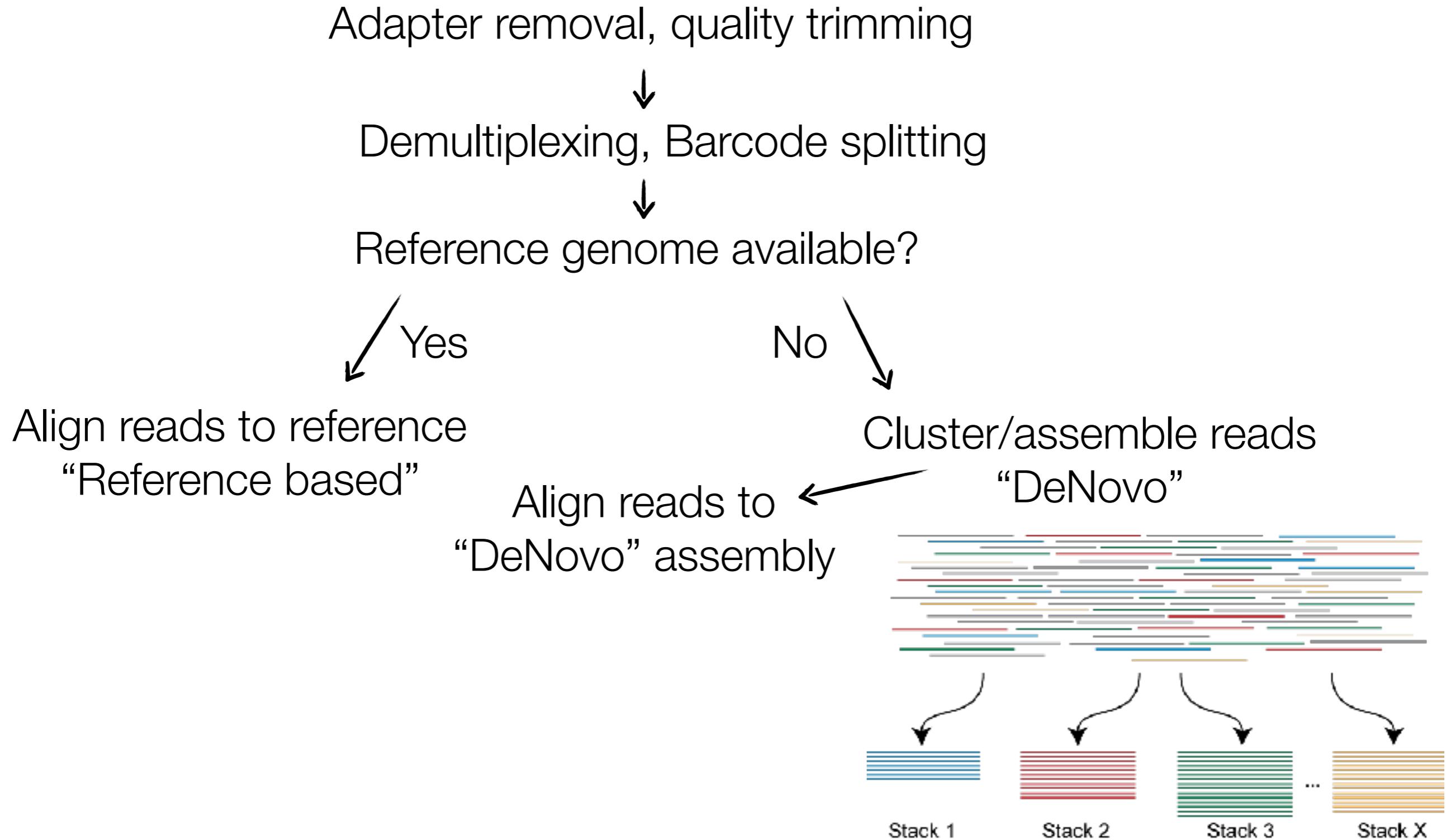
Yes

Align reads to reference
“Reference based”

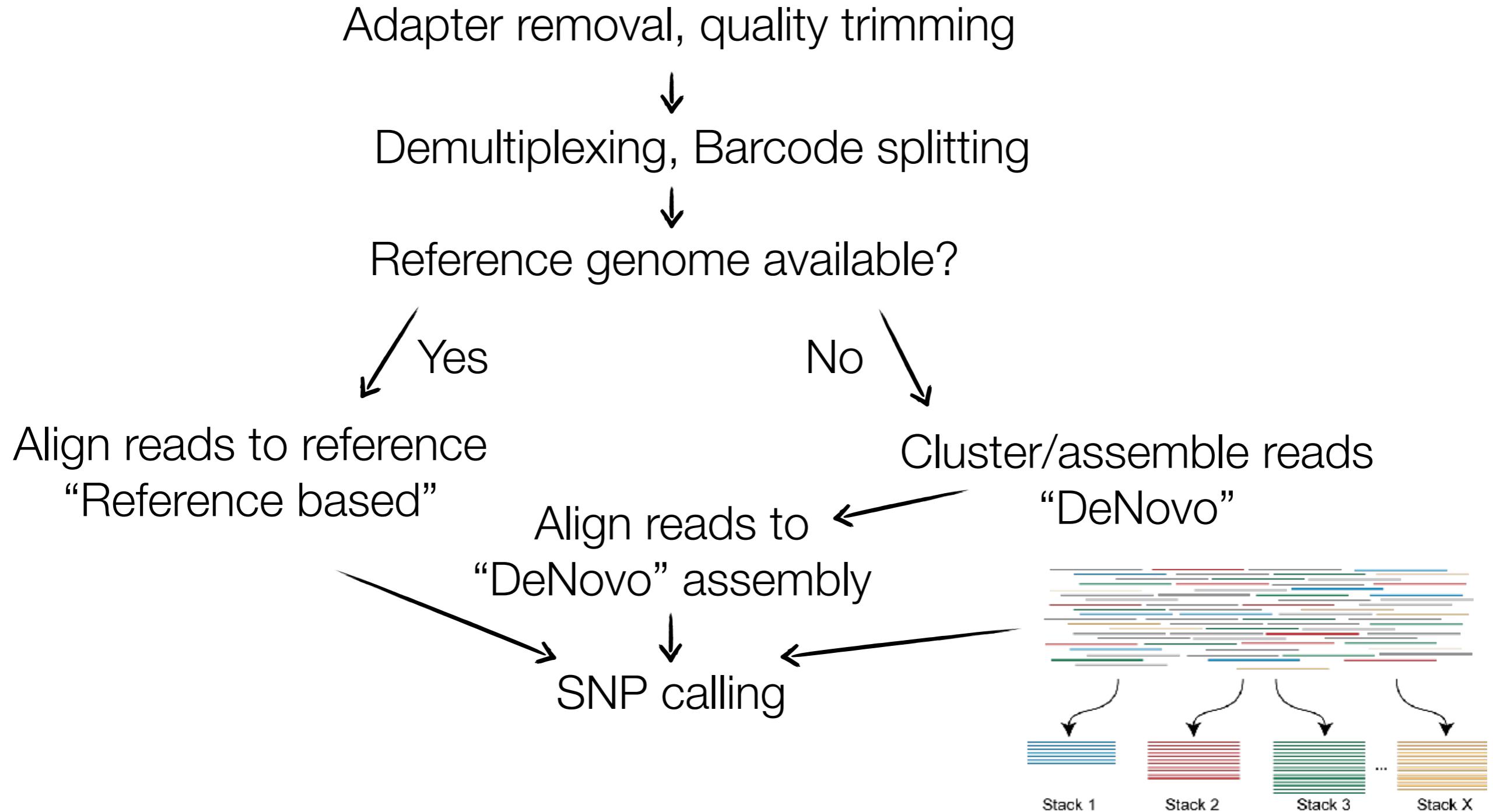
Data analysis: denovo analysis



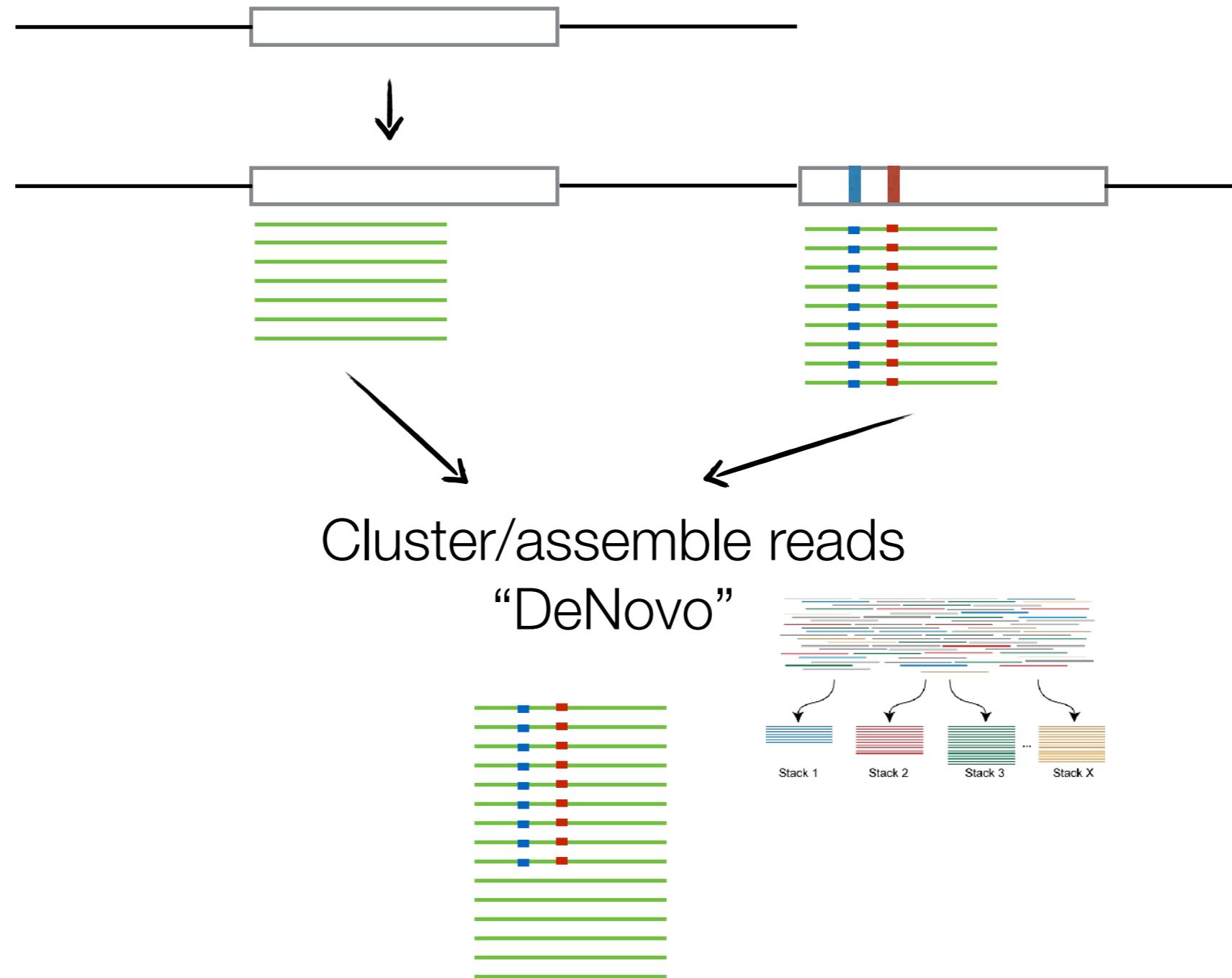
Data analysis: denovo analysis



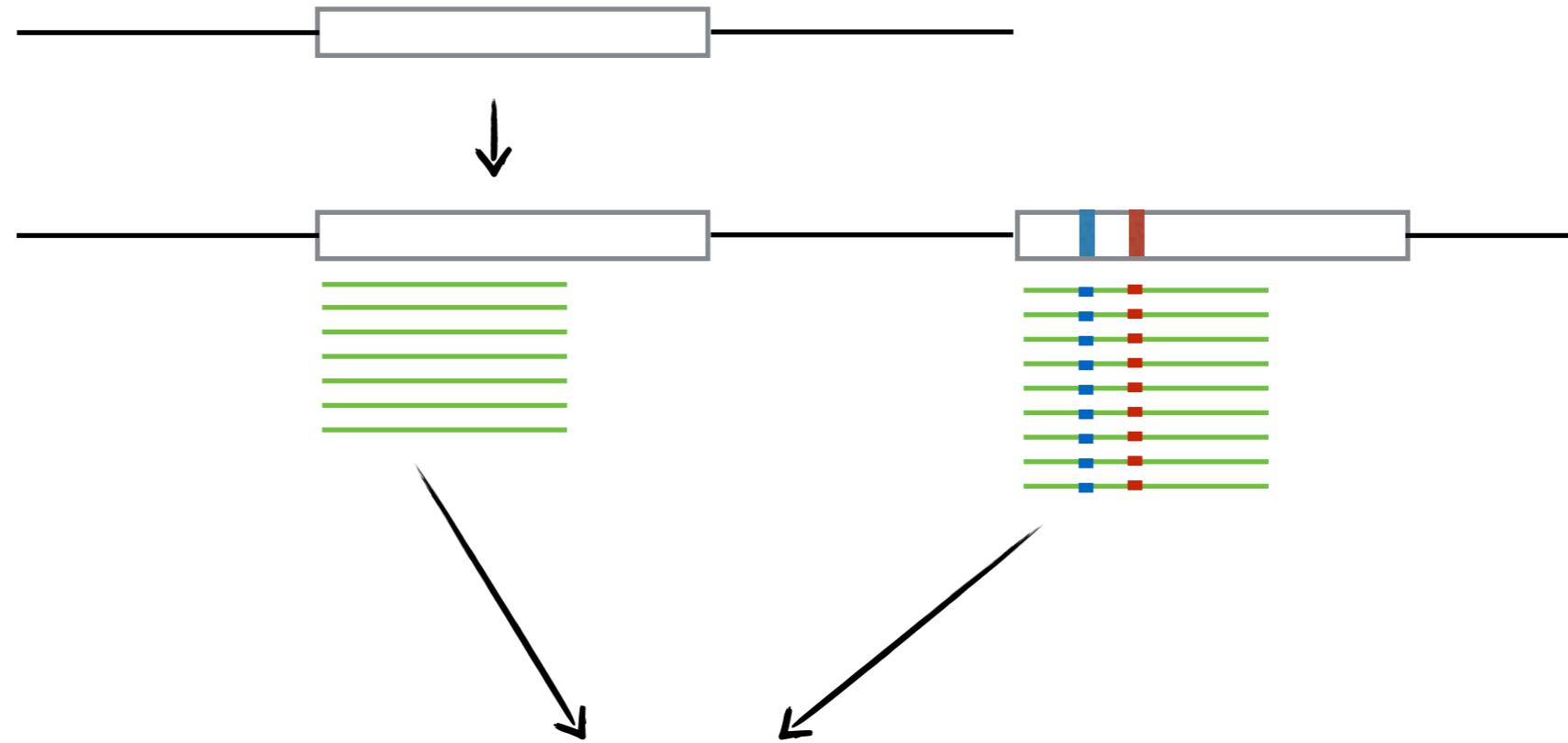
Data analysis: SNP calling



Avoid paralogs

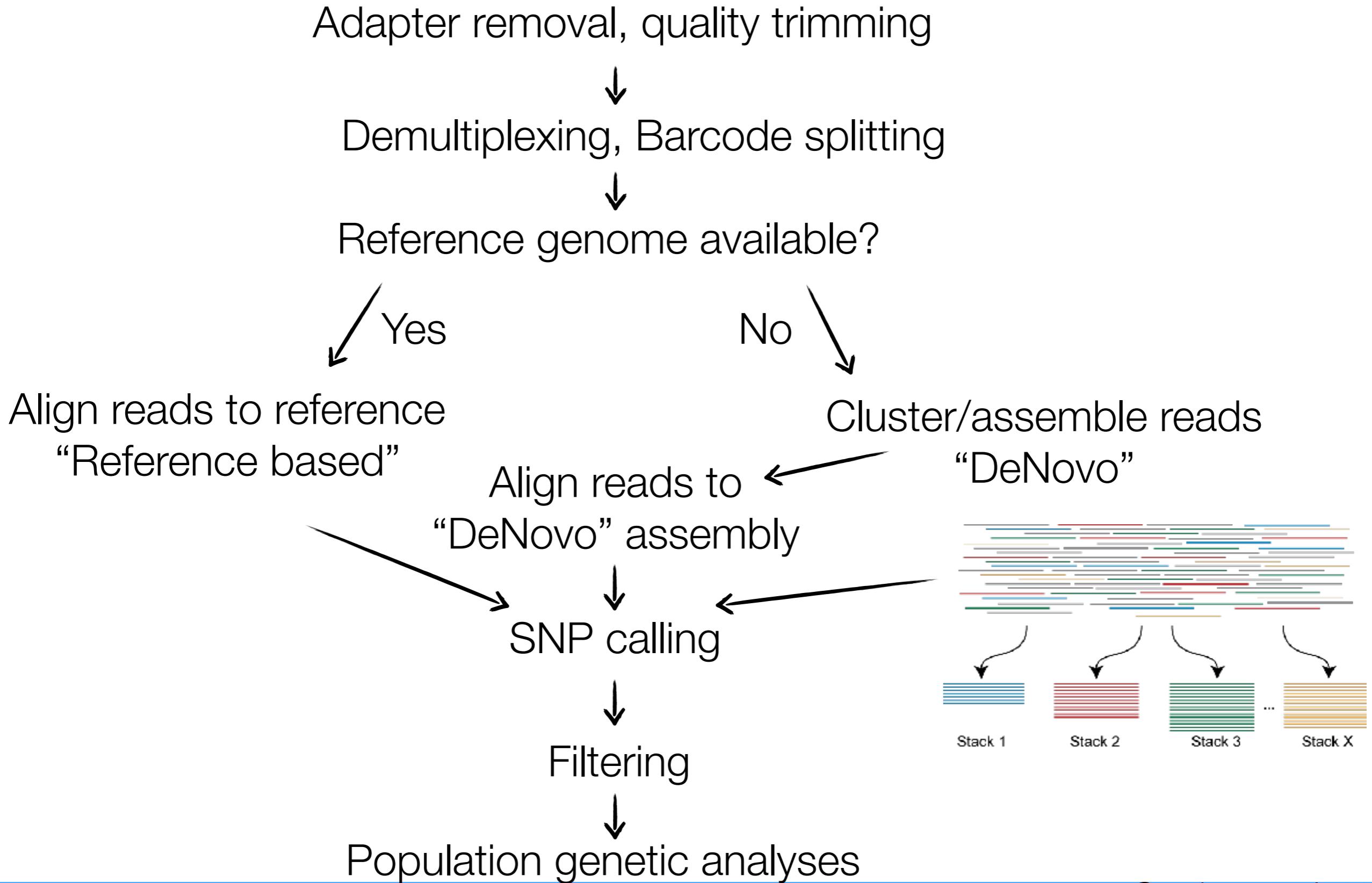


Avoid paralogs



SNP calling from aligned
reads (mapping quality)
reduces paralog problem

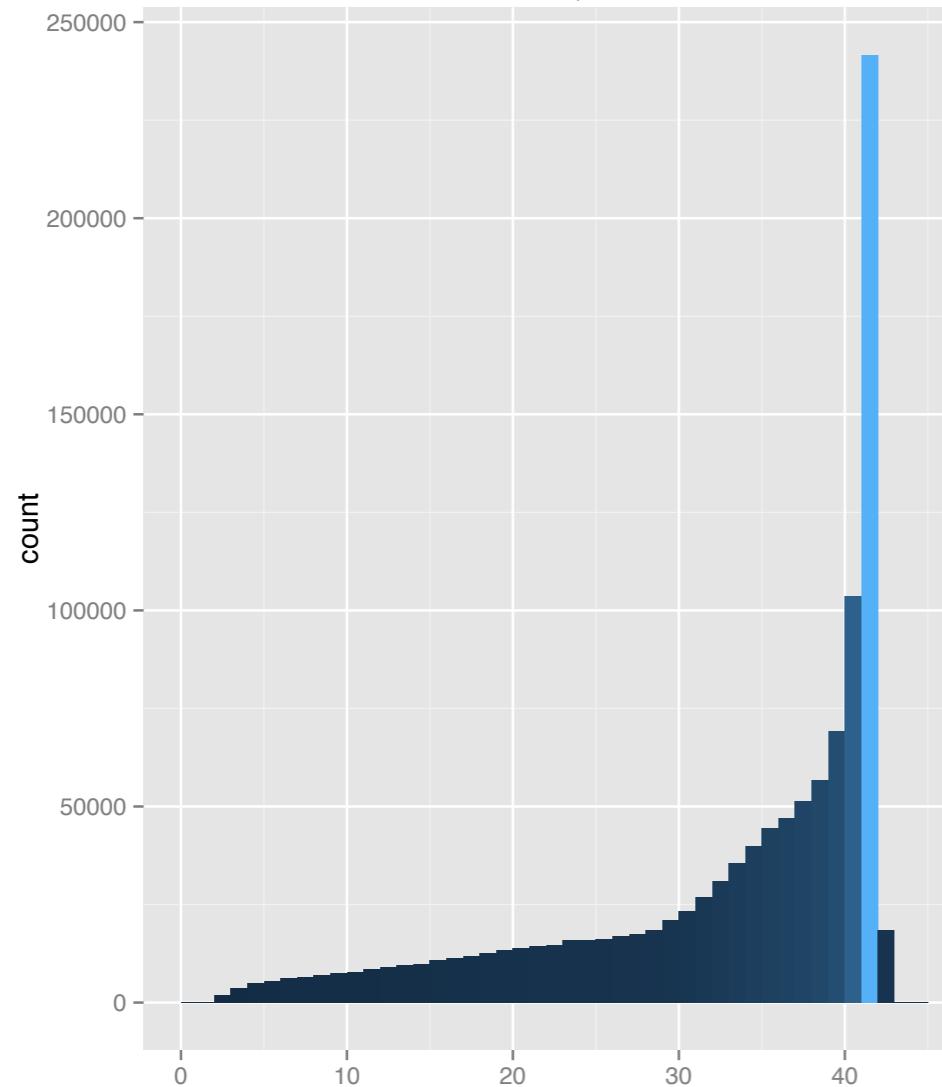
Data analysis: Filtering



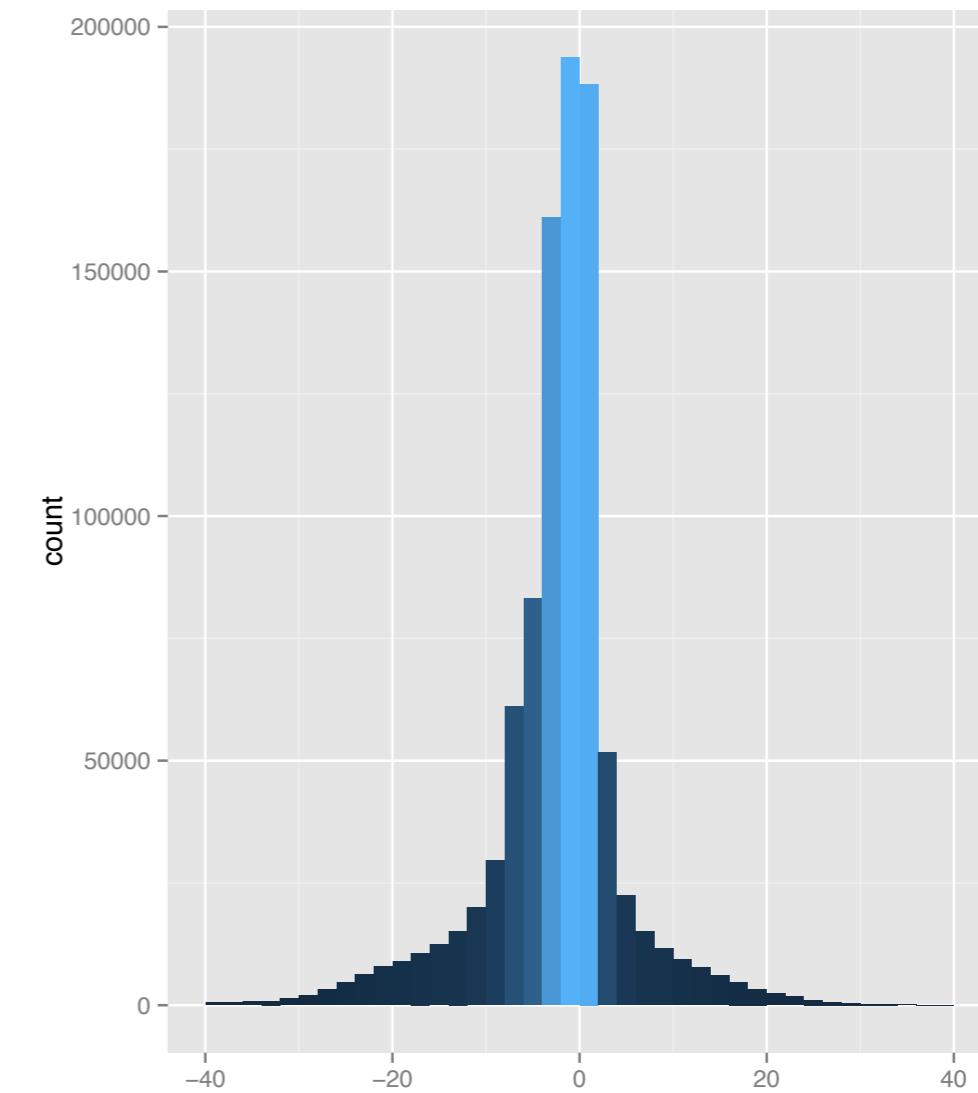
Data analysis: Filtering

Empirical cut-offs

Mapping quality



MQ rank sum



Data analysis: Filtering

Empirical cut-offs

Use caution when filtering for MAF and heterozygosity (HWE)

Consider only removing singletons and private doubletons

Available pipelines

Stacks (Catchen et al. 2013)

First available pipeline

Designed for RAD but handles other methods

dDocent (Puritz et al. 2014)

Simple, customisable bash backbone for bioinformatics

Designed for ddRAD & ezRAD

FreeBayes or GATK

pyRAD (Eaton 2014)

Analysis pipeline written in Python

Many different RAD types

Clustering using Usearch or Vsearch

UNEAK, Tassel

Designed for GBS data as produced in Cornell

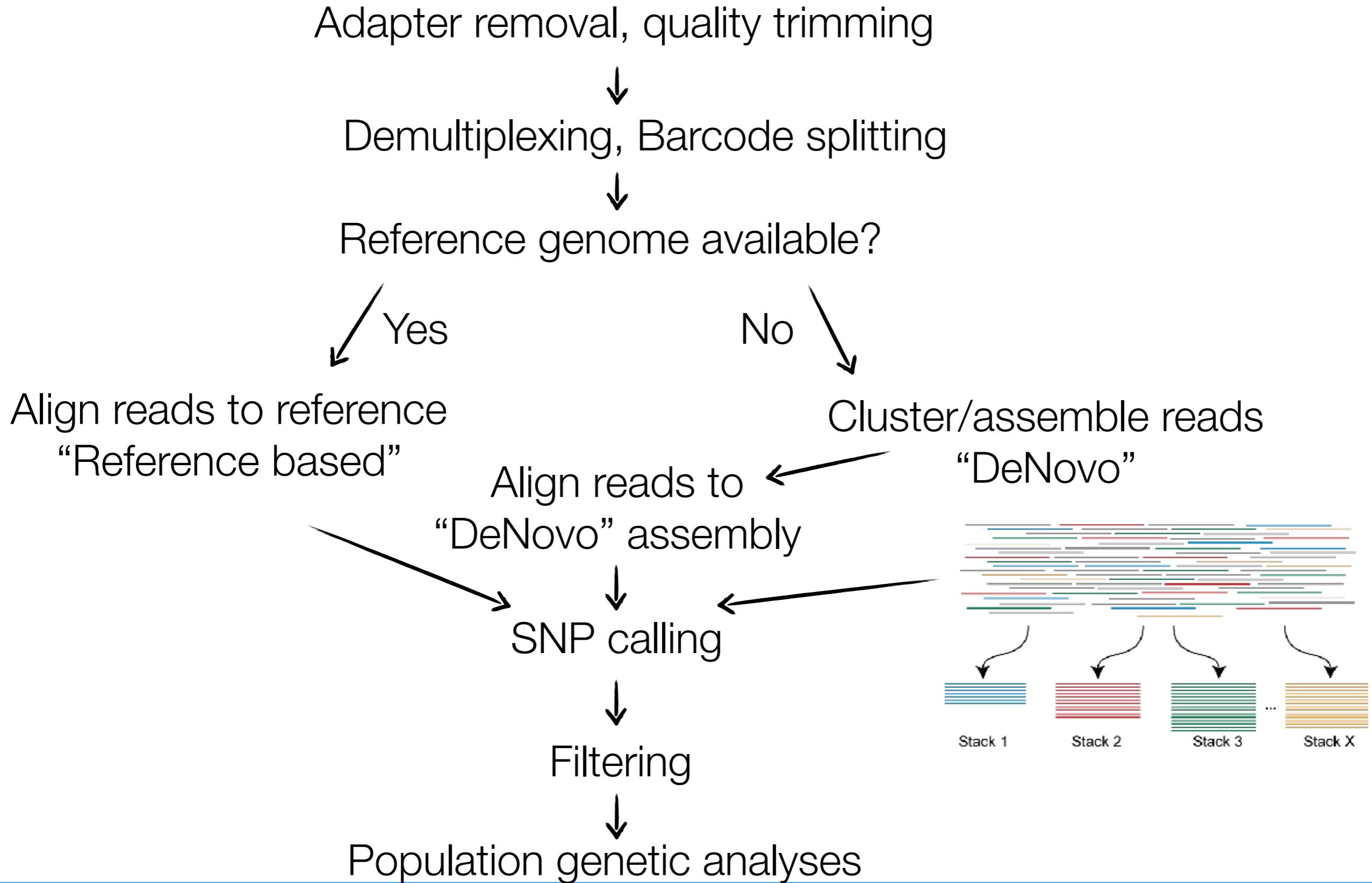
aftrRAD (Sovic et al. 2015)

Blend between stacks and pyrad

Fast-GBS (Torkamaneh et al. 2017)

Uses BWA and Platypus

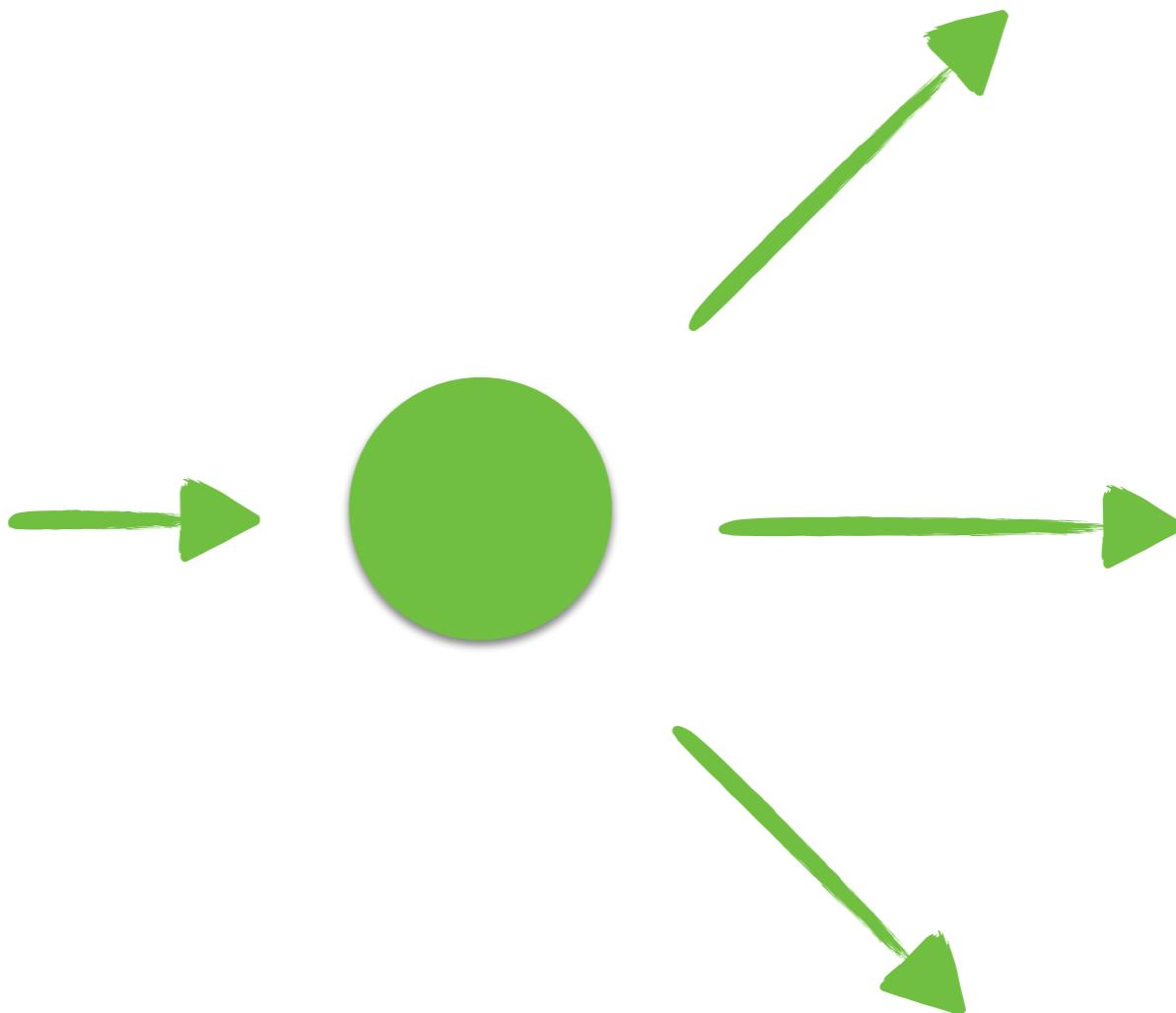
Data analysis with custom scripts



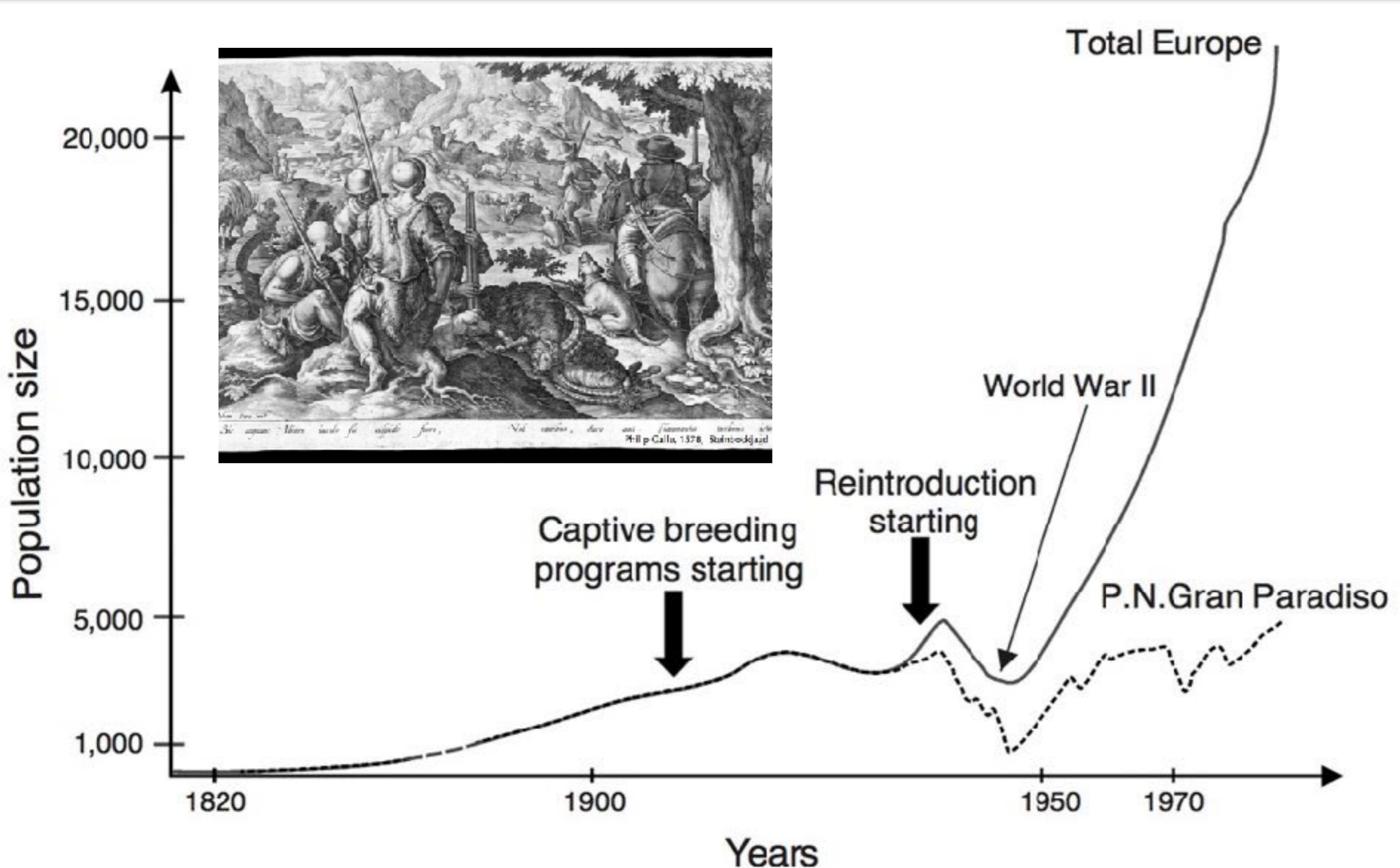
Population genomics of species reintroductions



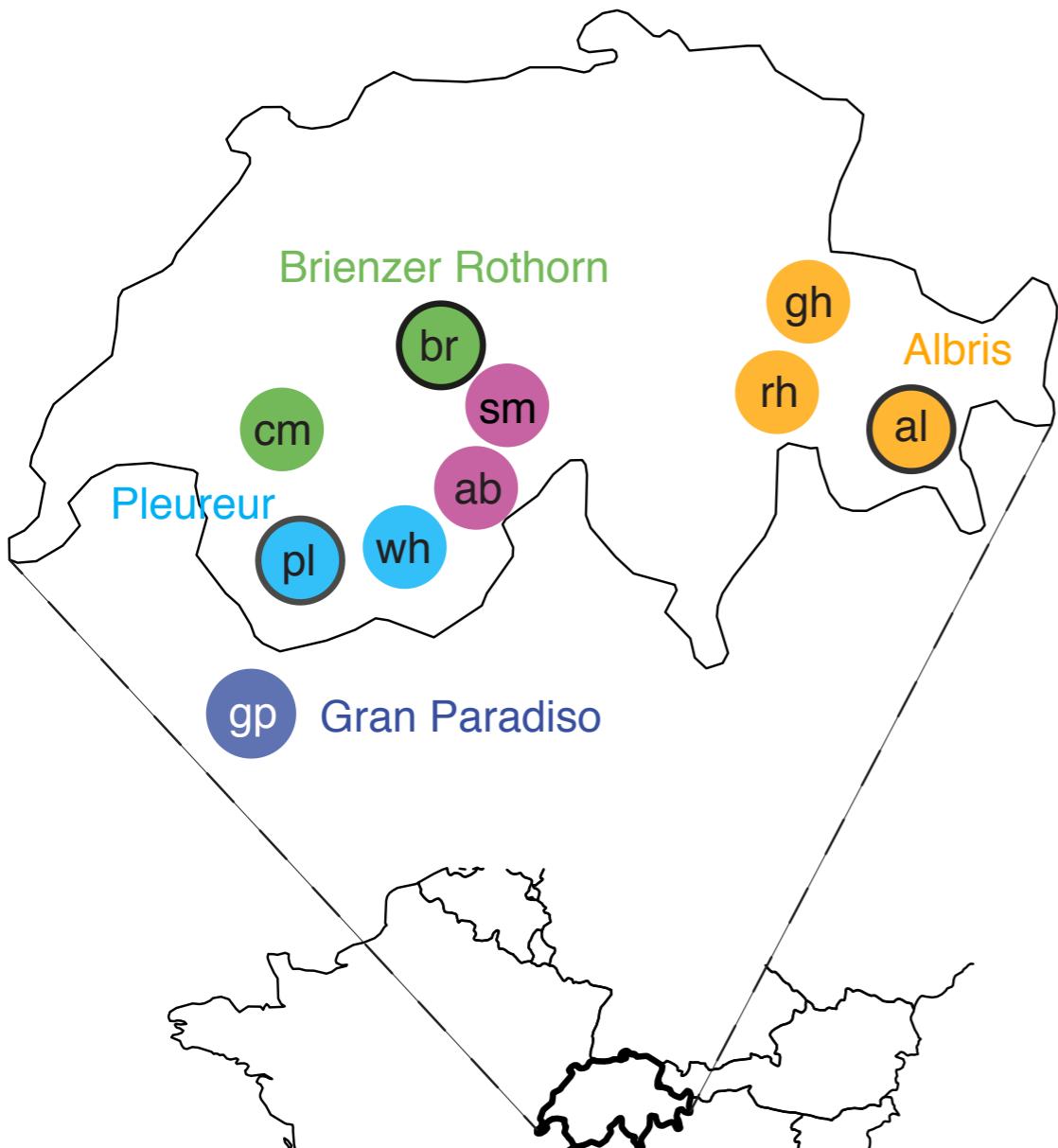
The problem of small population sizes



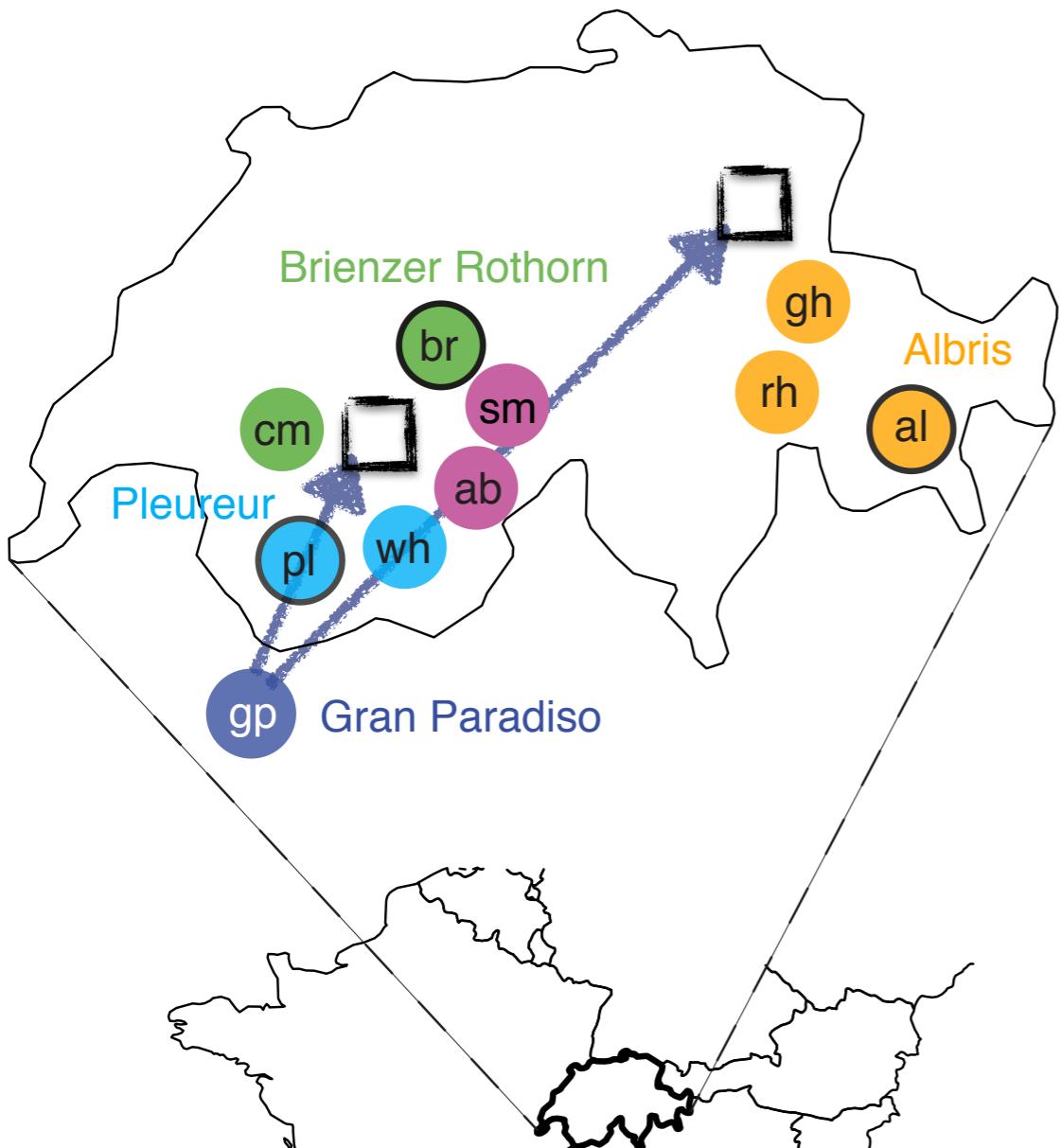
Alpine ibex went through a series of bottlenecks



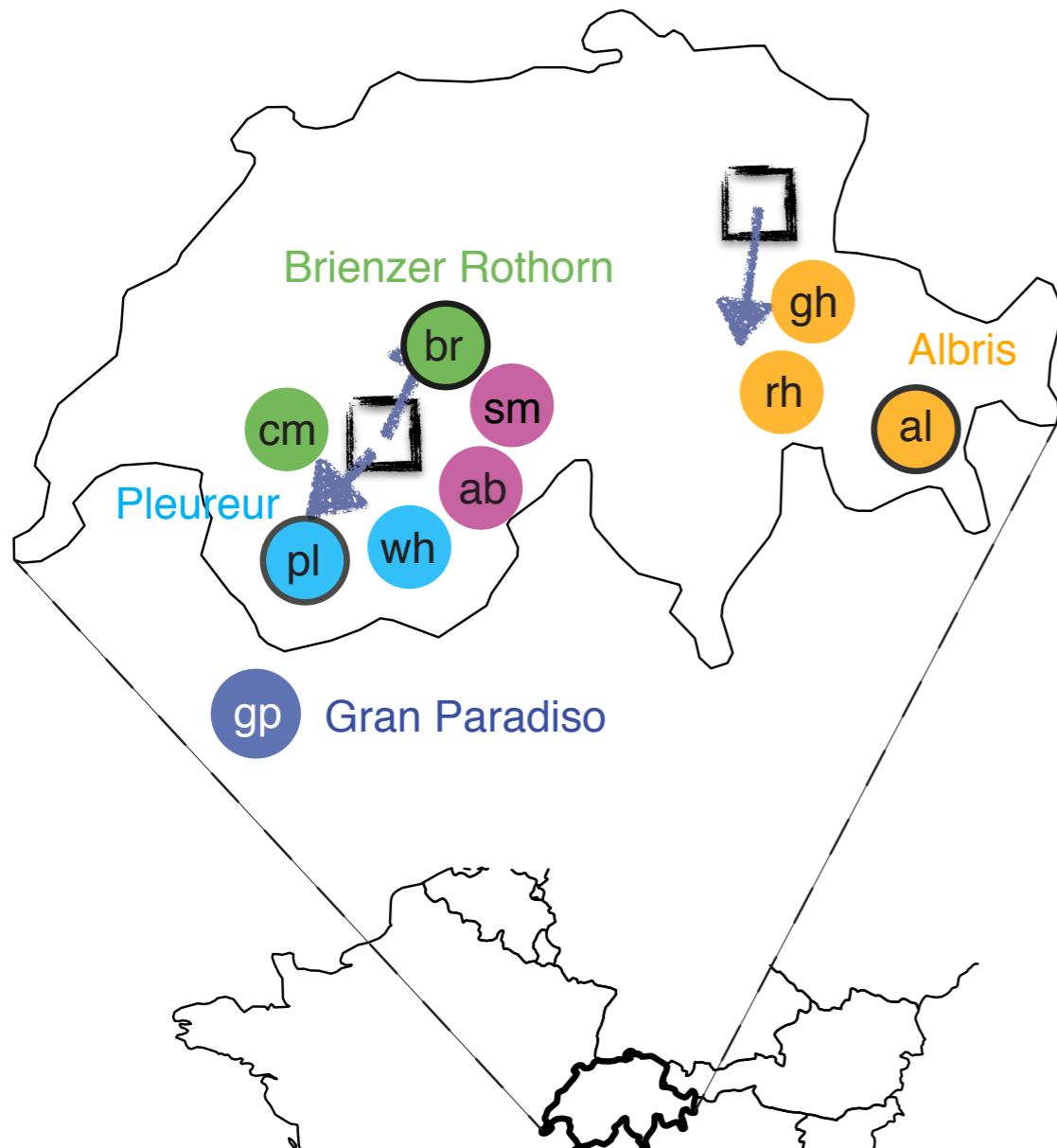
Genomic footprint of reintroductions



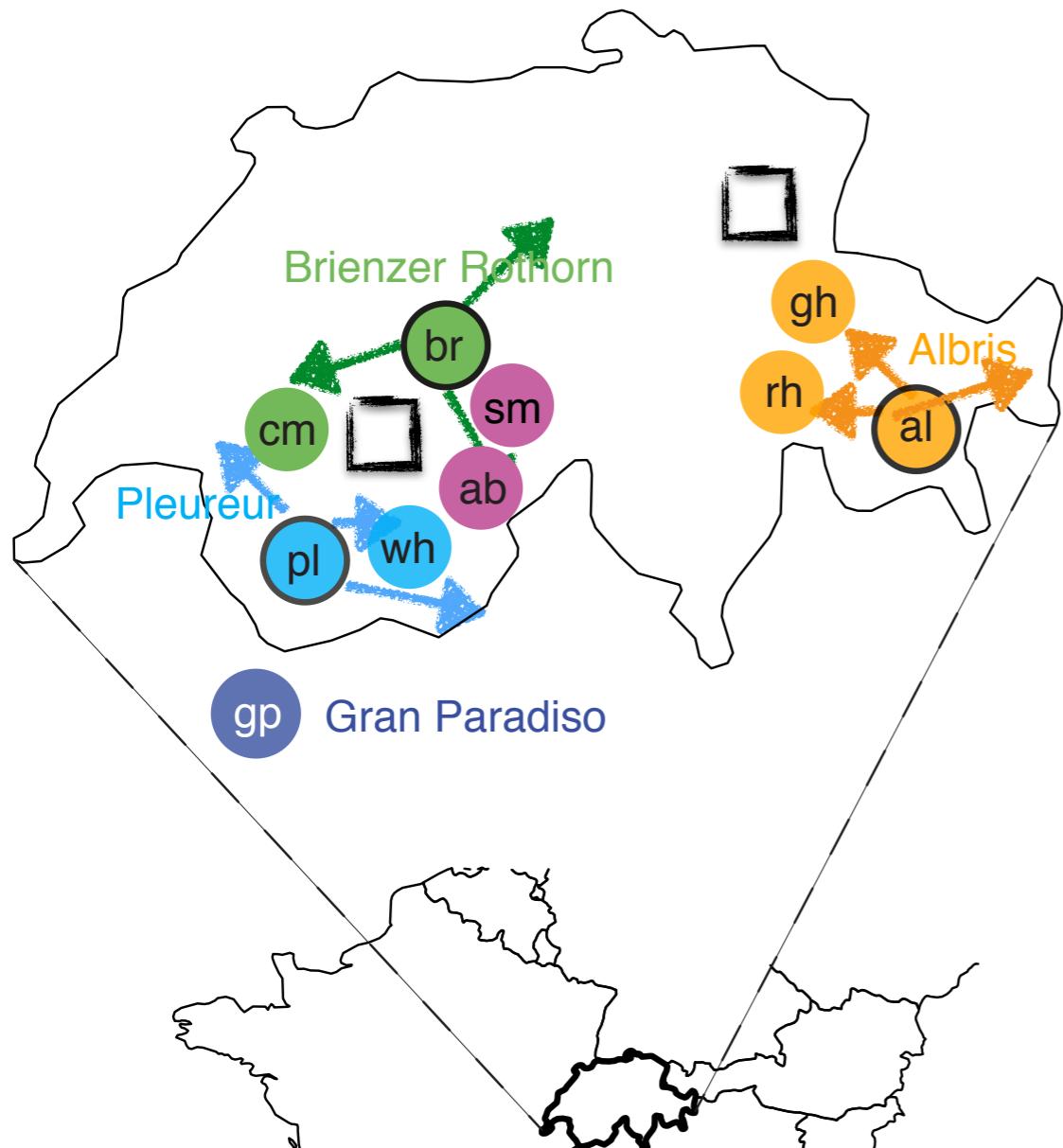
Captive breeding



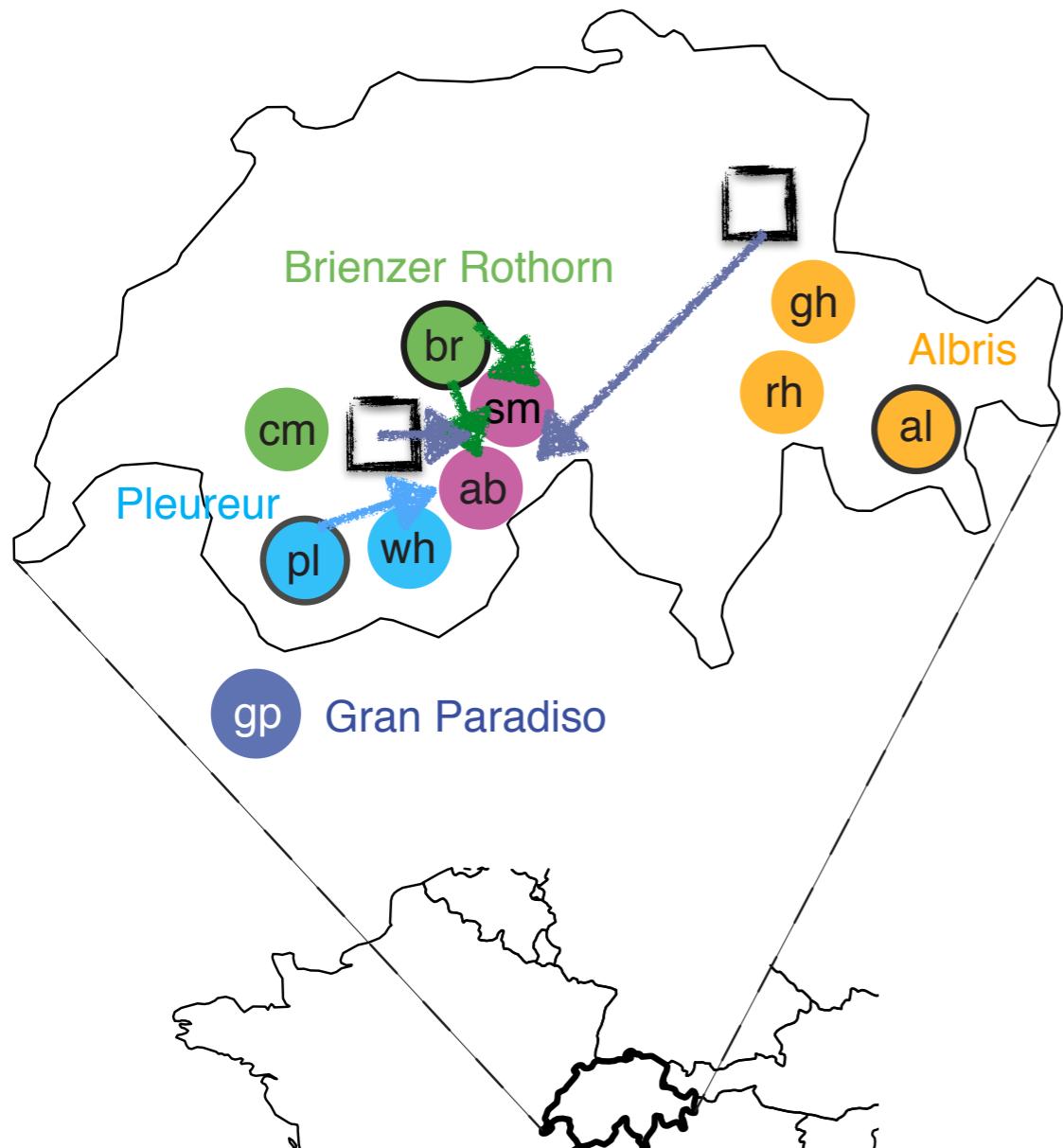
Founding of primary reintroduced populations



Founding of secondary reintroduced populations

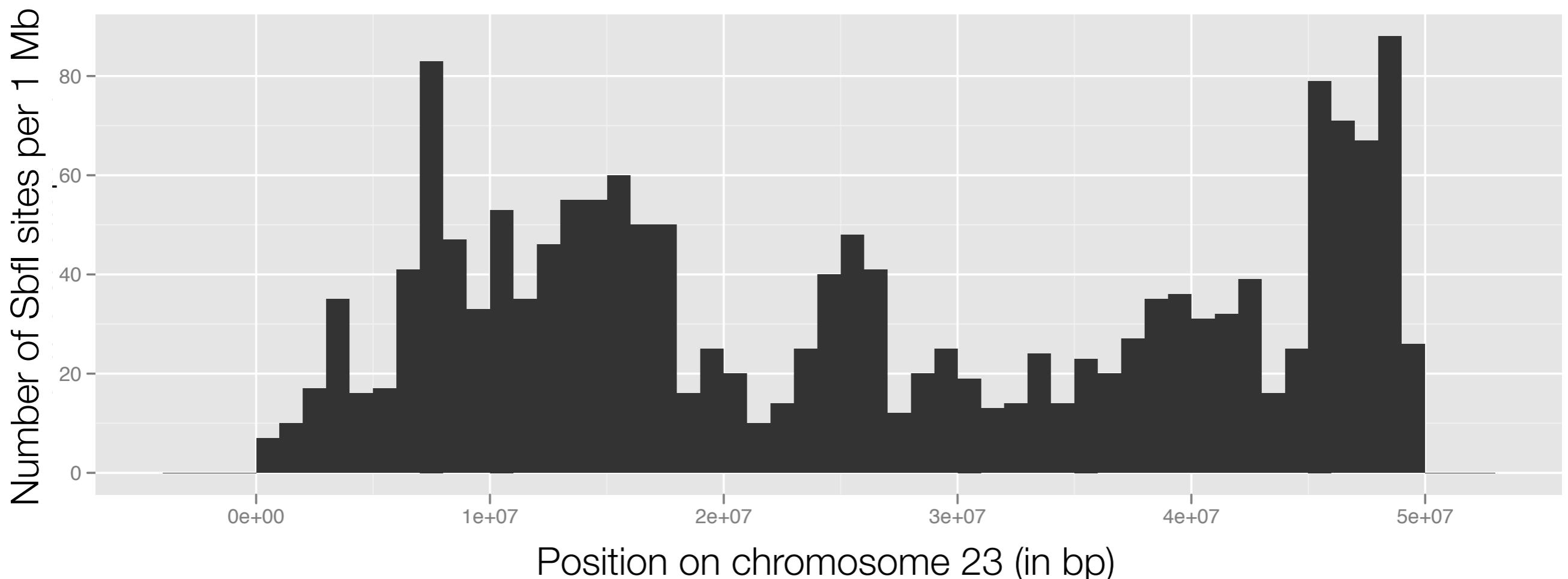


Founding of secondary reintroduced populations

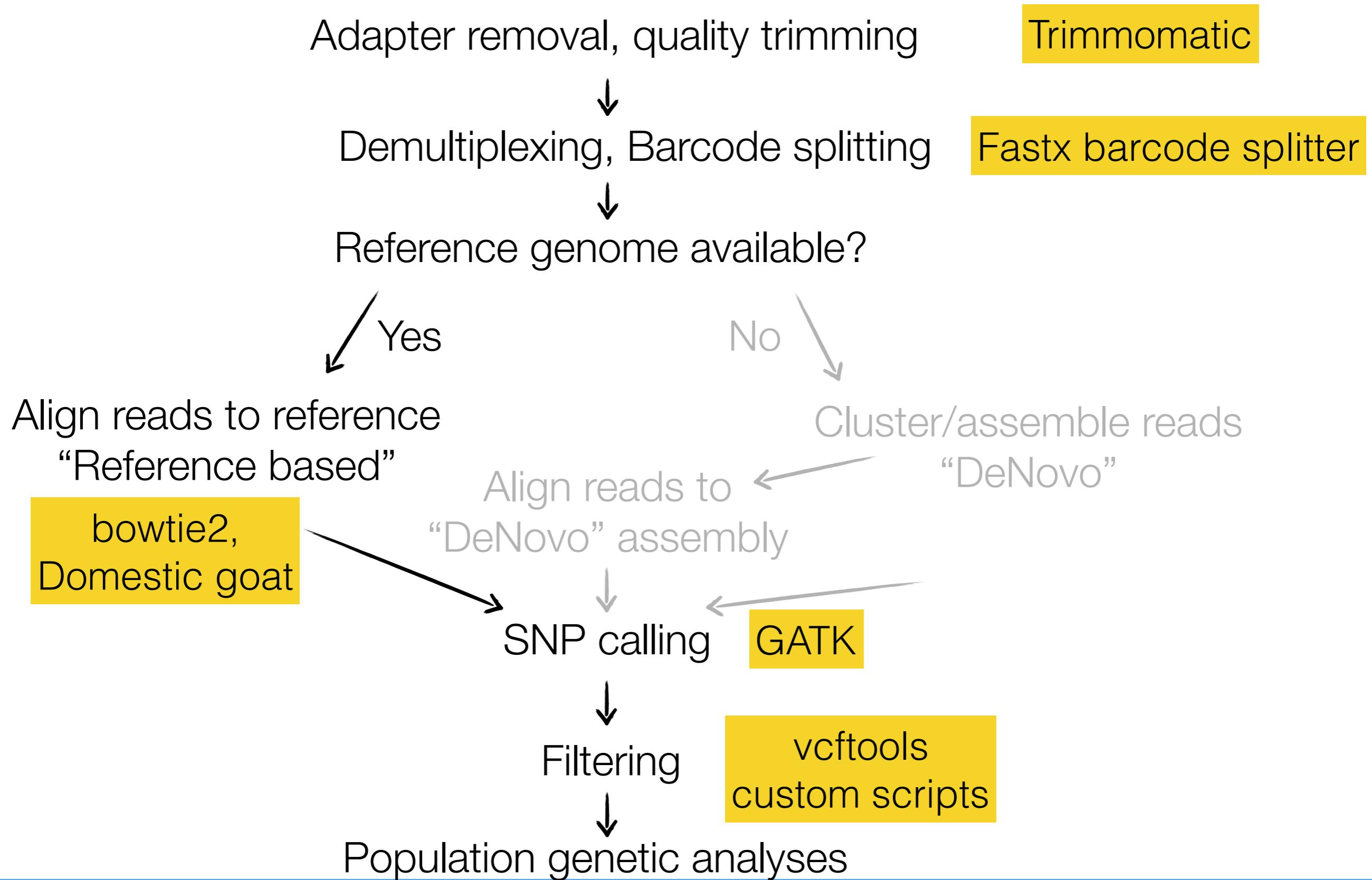


“original” RAD with Alpine ibex

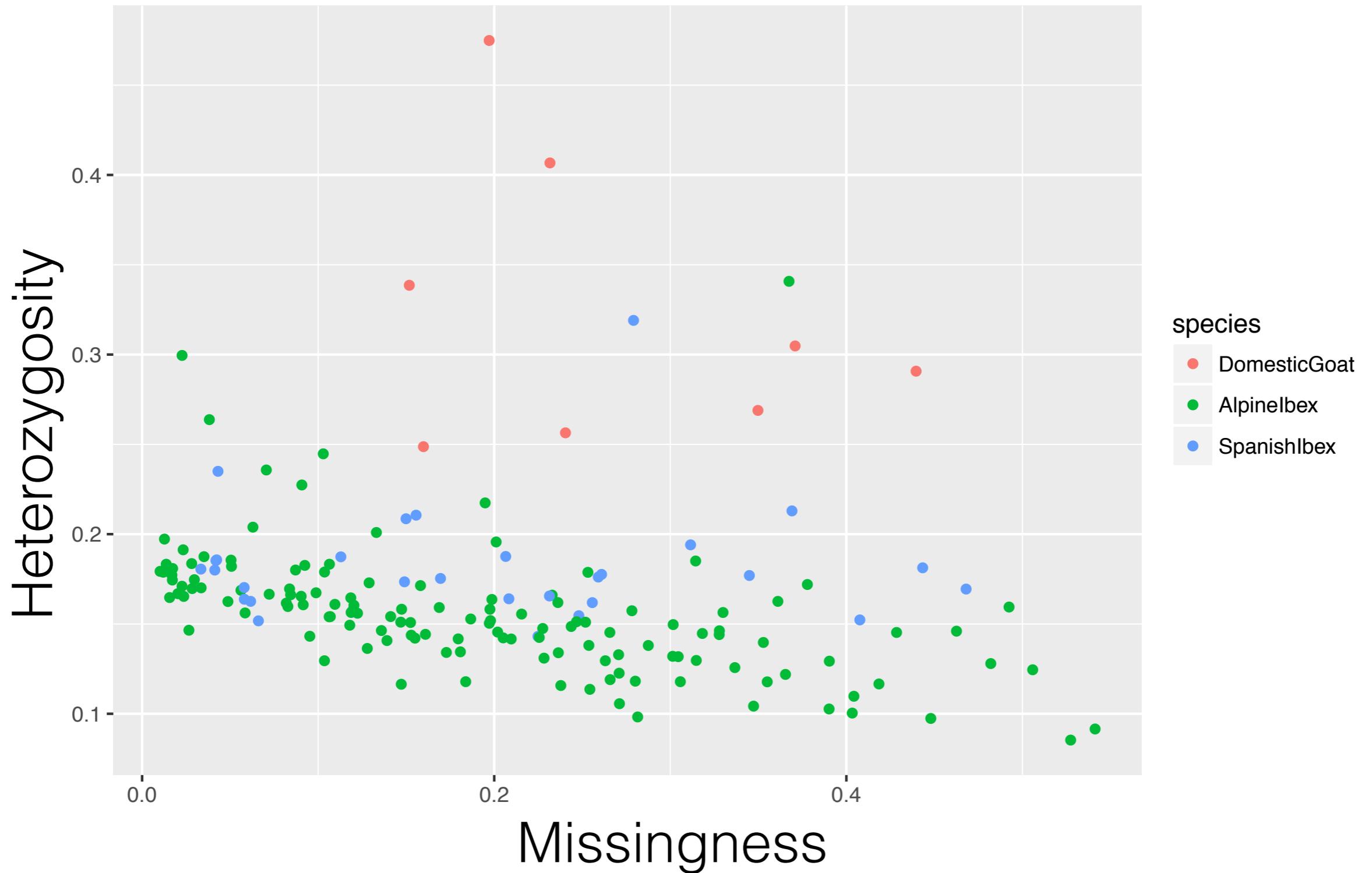
SbfI: 68'000 restriction sites on domestic goat reference genome



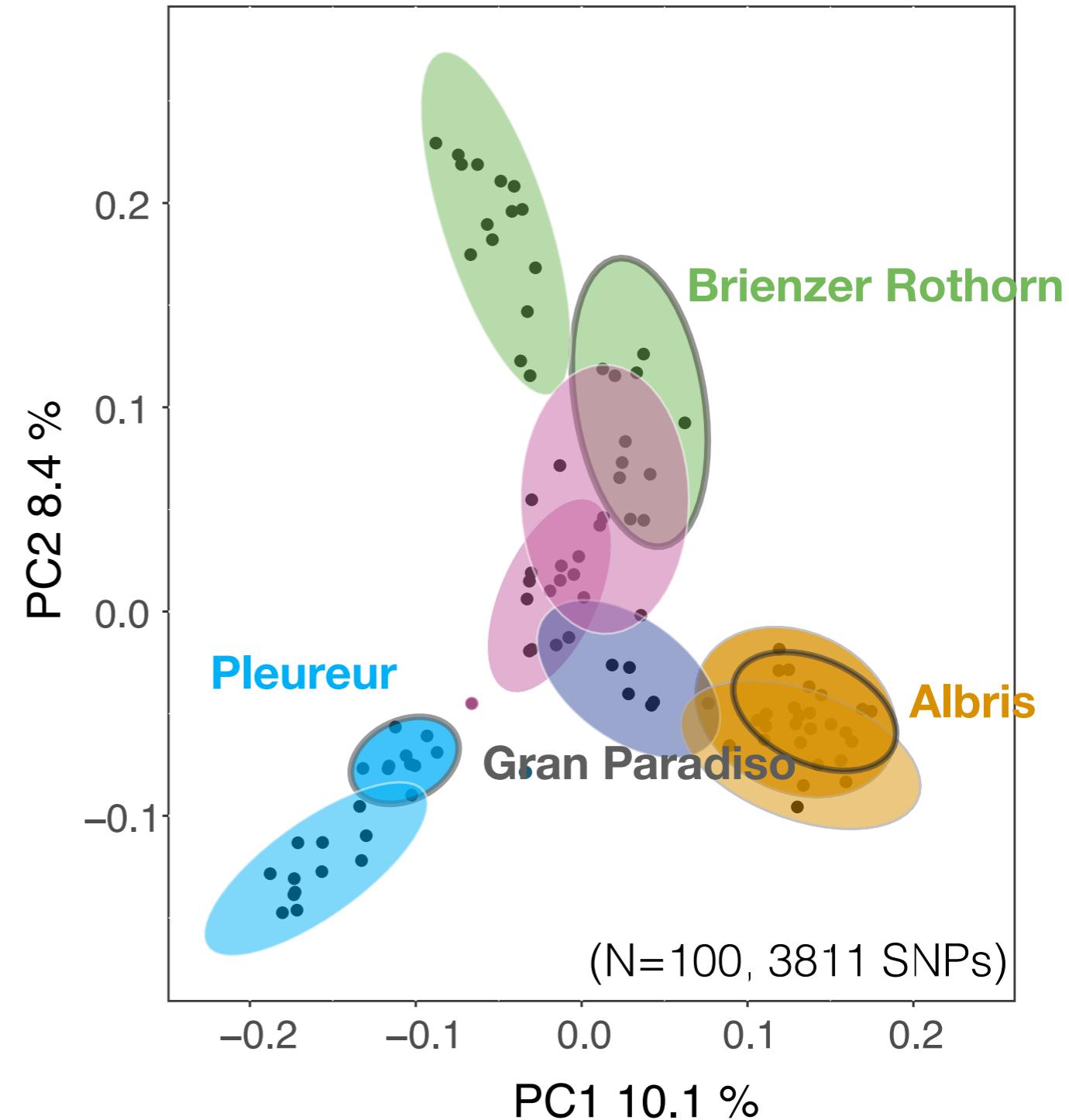
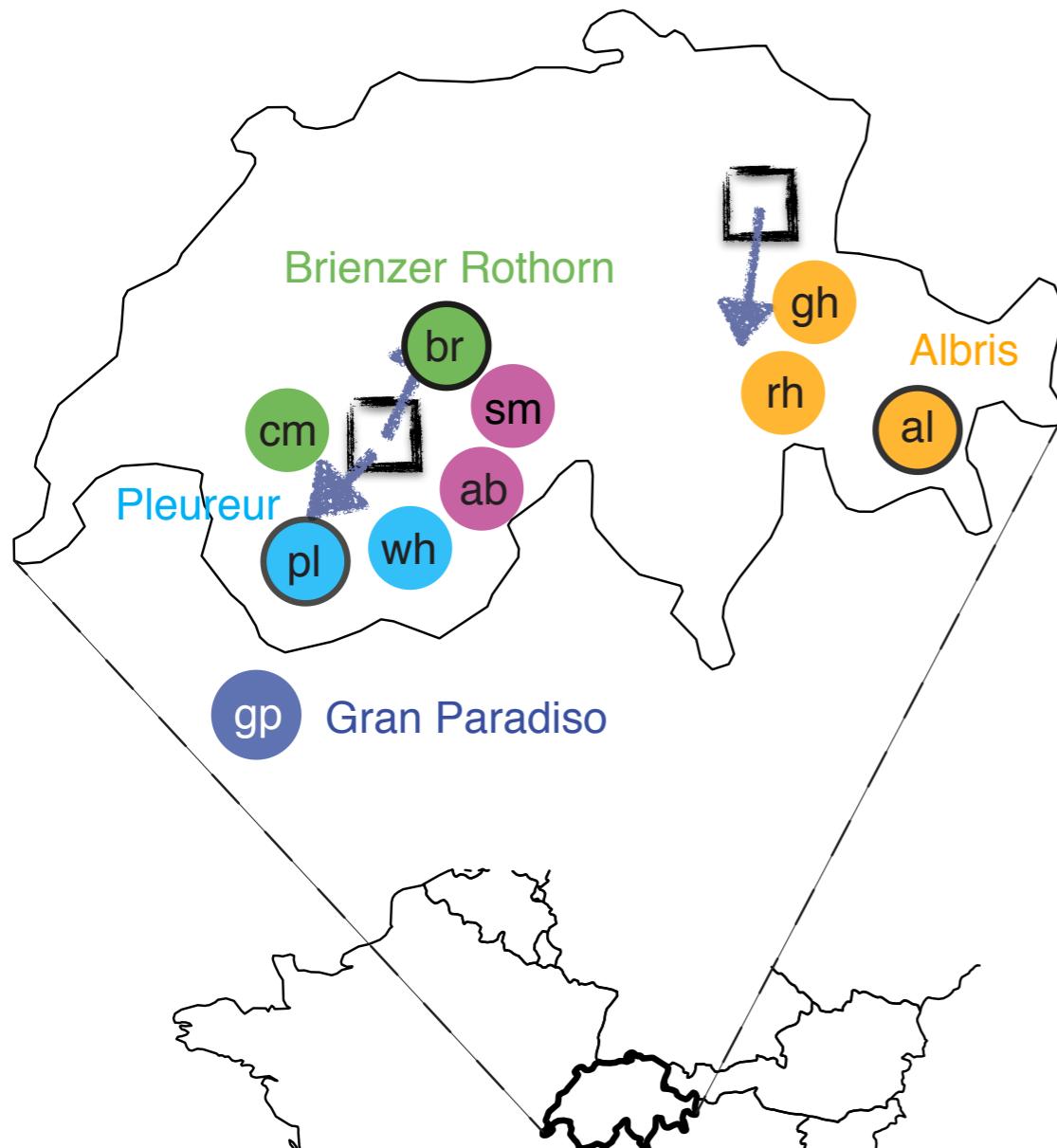
Data analysis with custom scripts



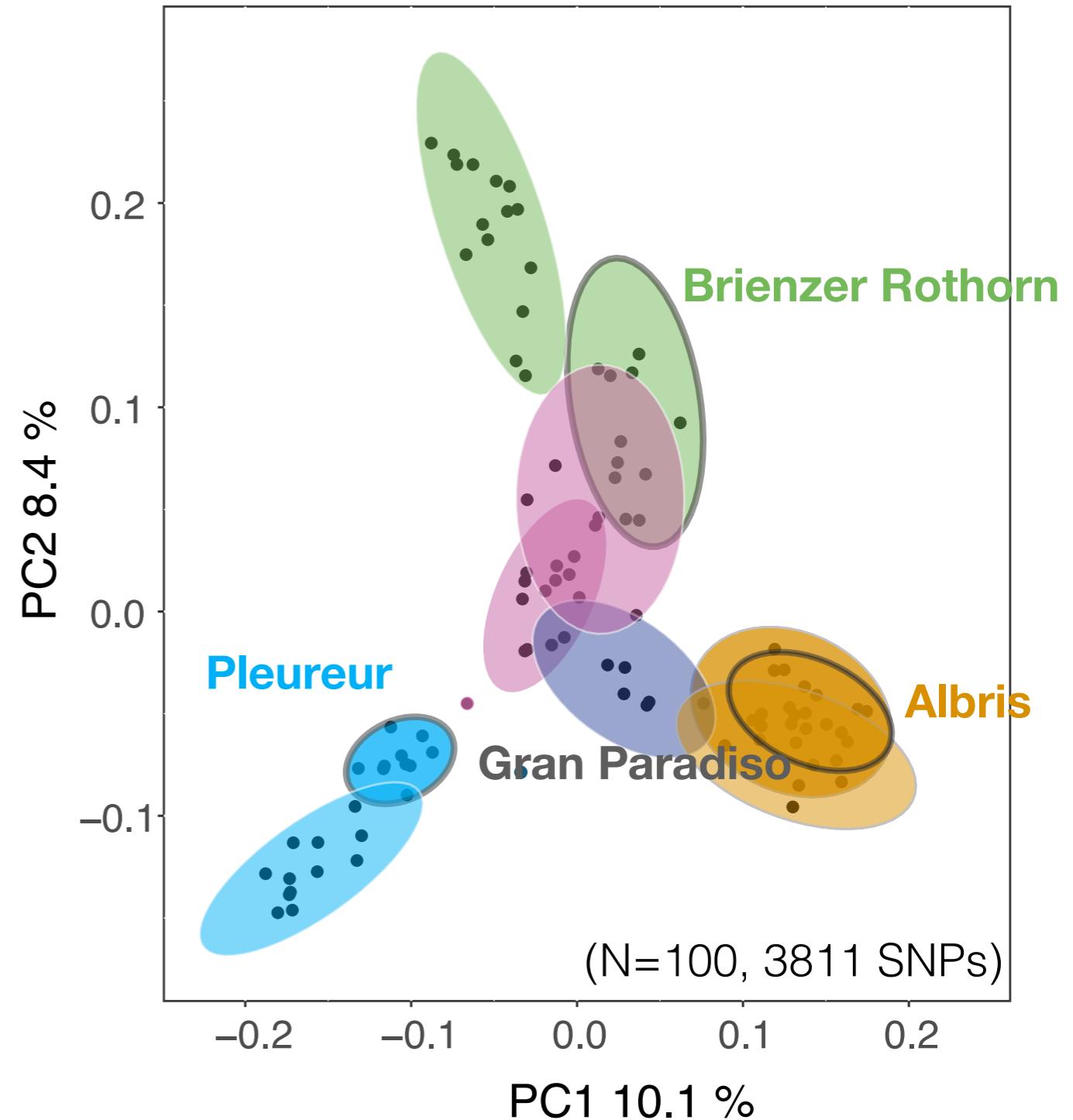
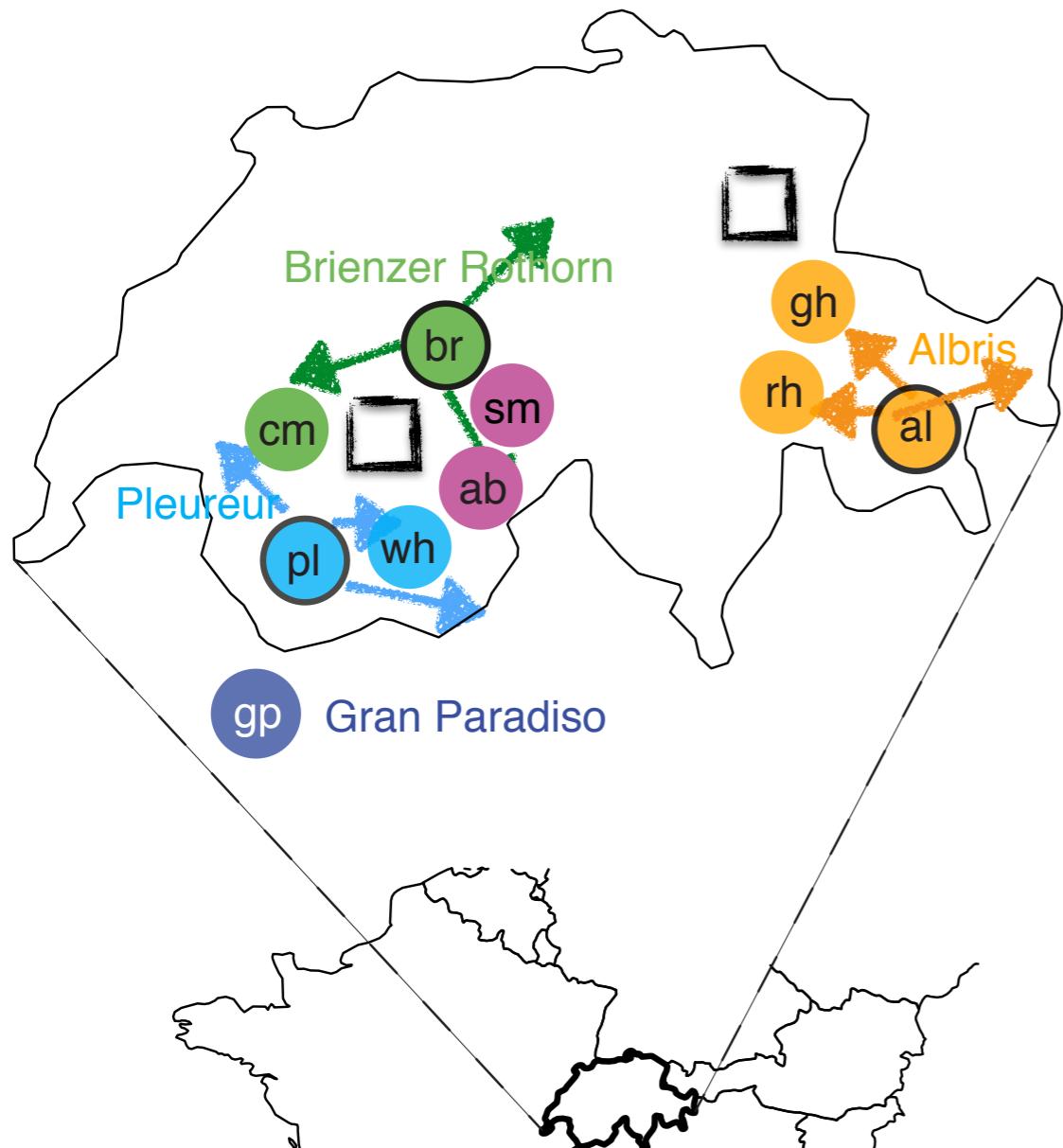
Diversities and coverage (missingness)



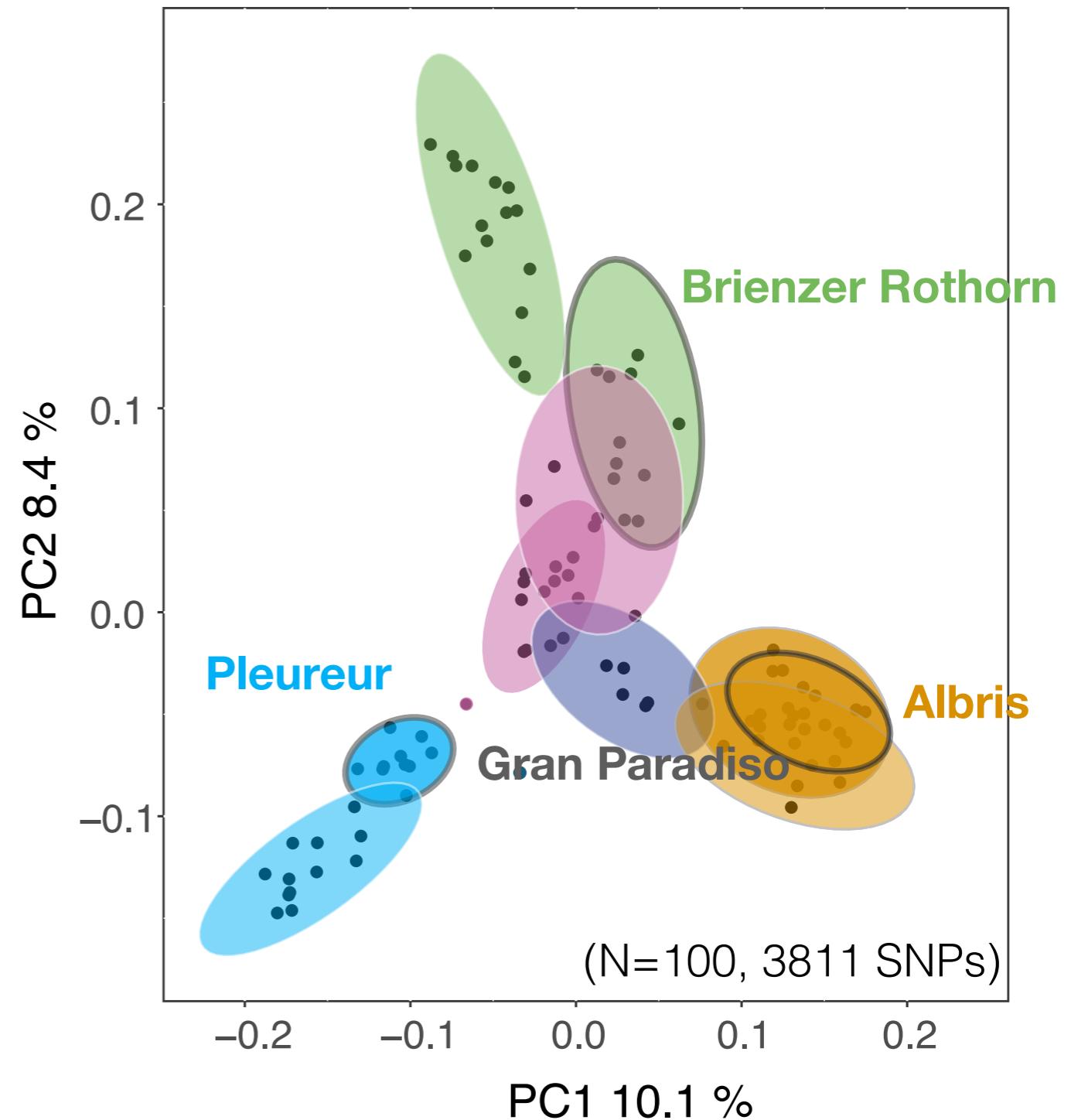
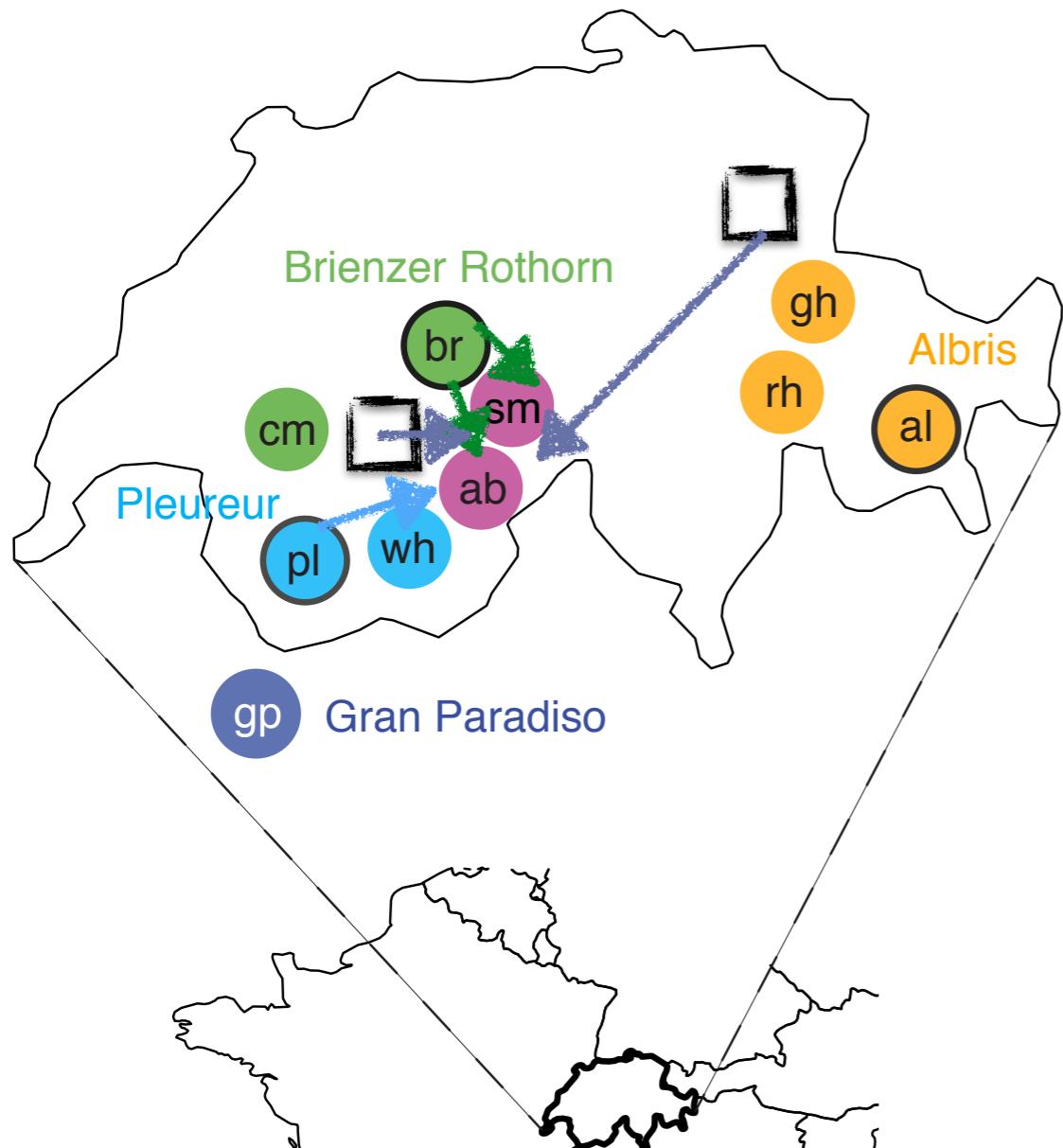
Founding of primary reintroduced populations



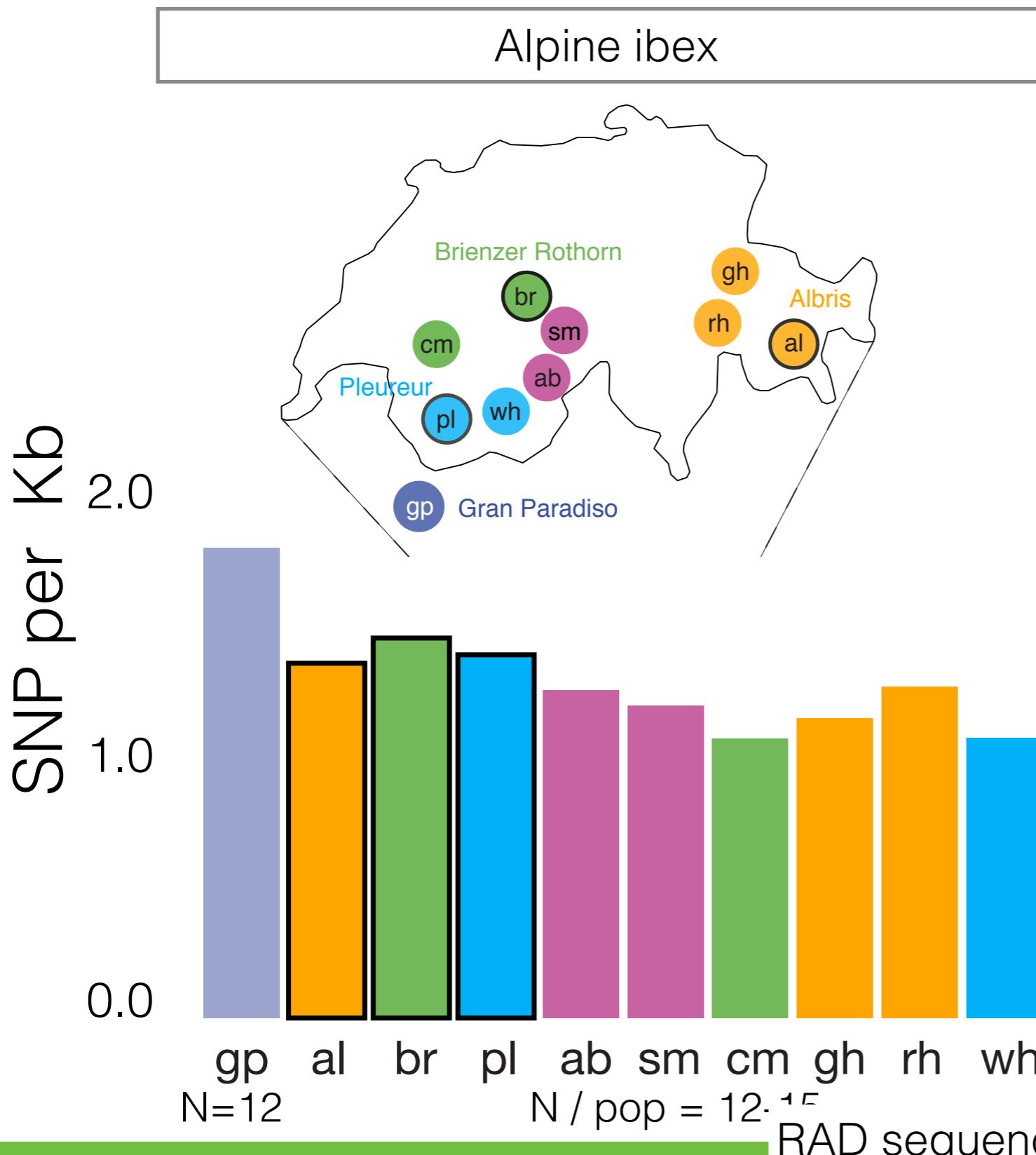
Founding of secondary reintroduced populations



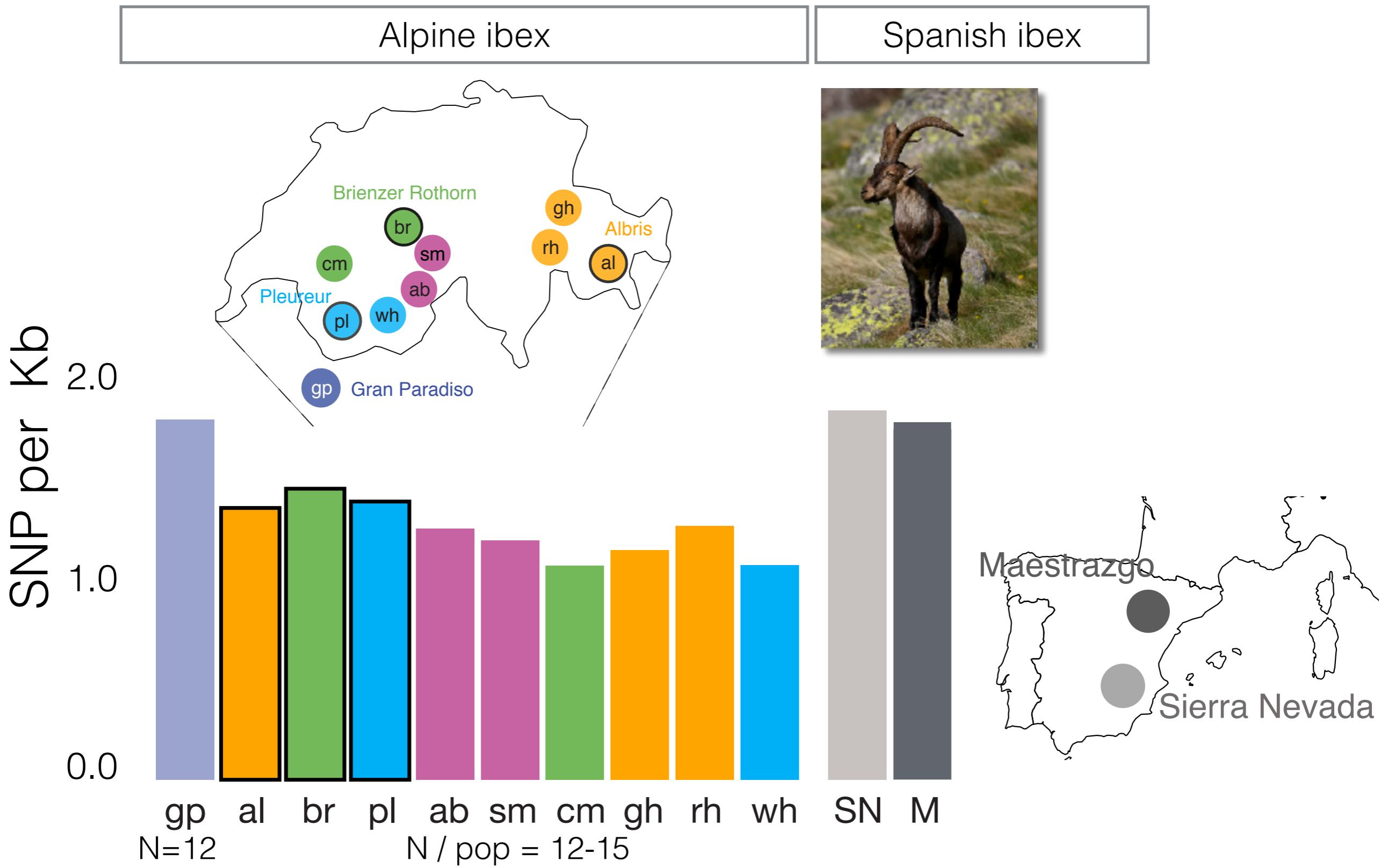
Founding of secondary reintroduced populations



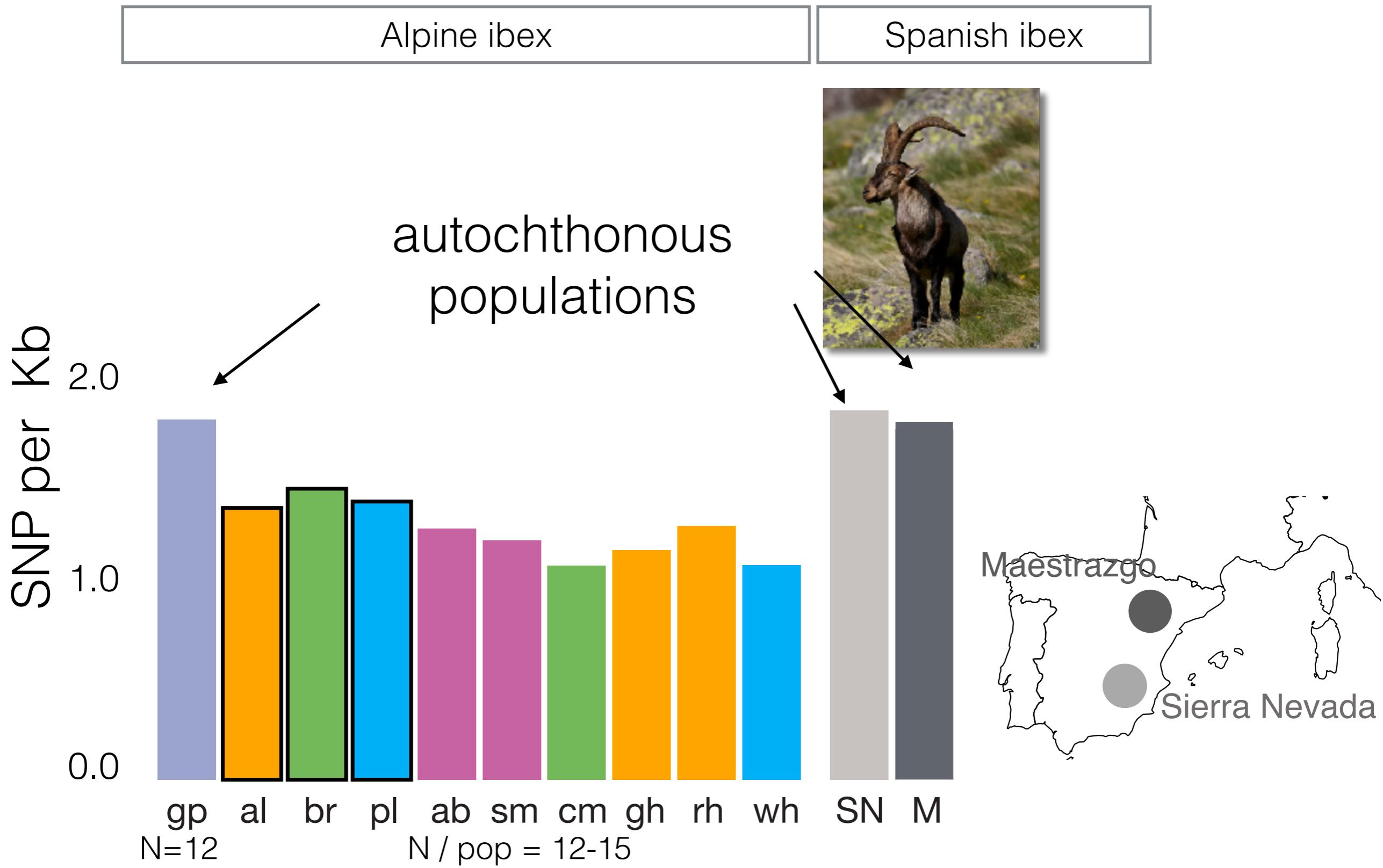
Higher diversity in resident populations



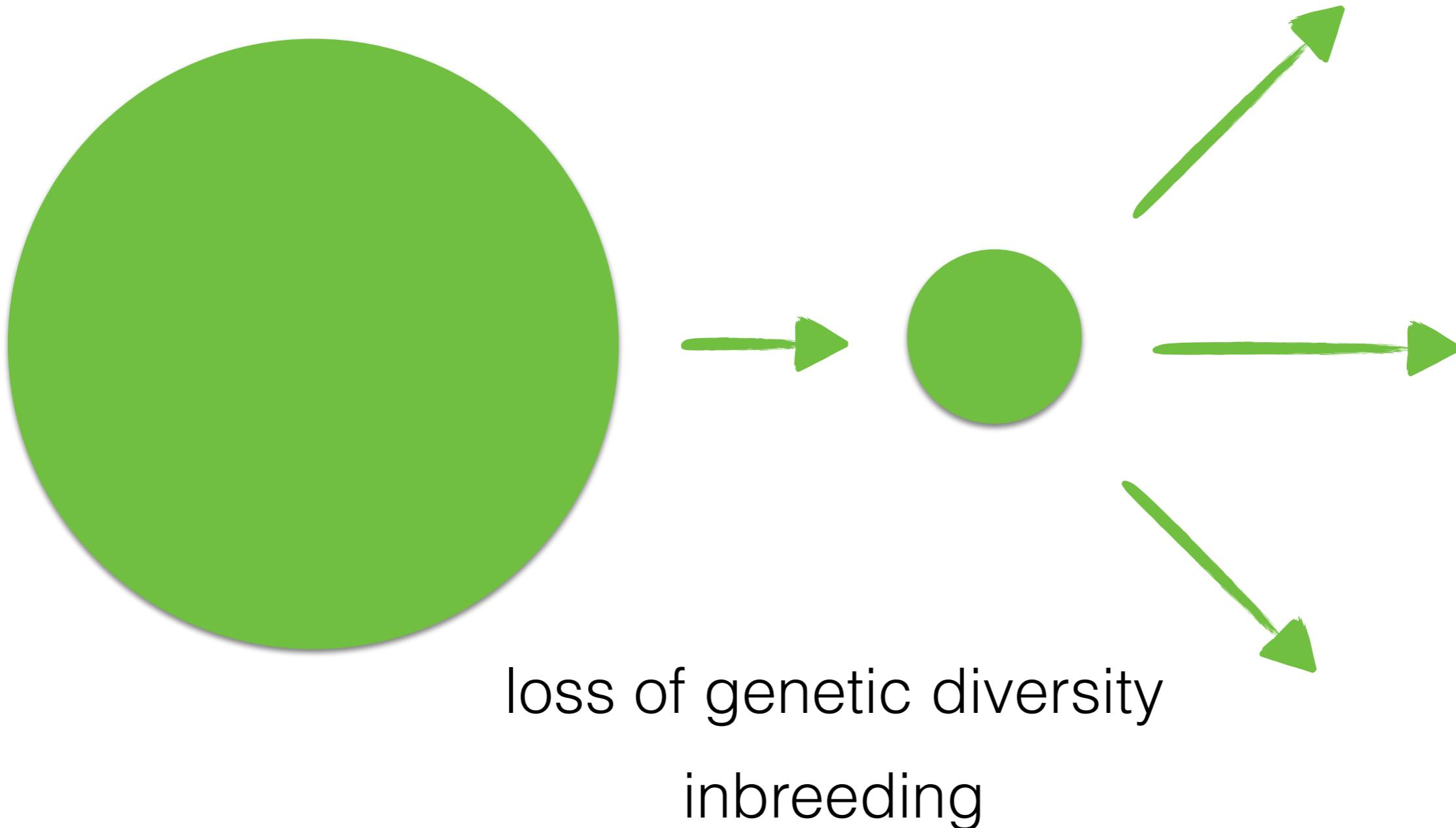
Higher diversity in resident populations



Higher diversity in autochthonous populations

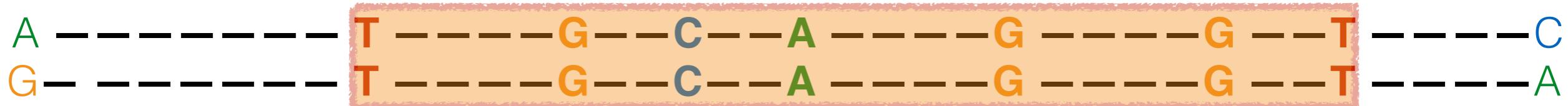


Consequences of small population size

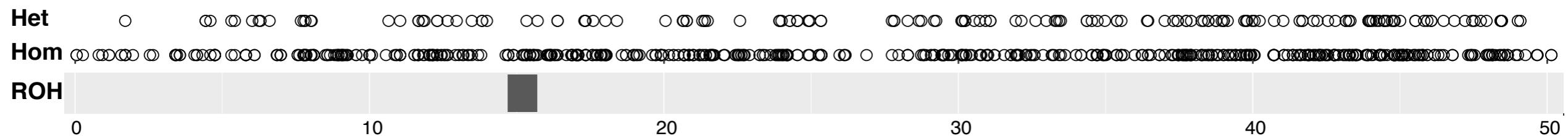


Inbreeding leads to long ROHs

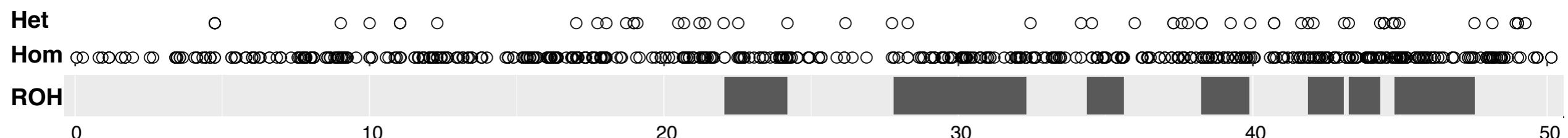
Run Of Homozygosity



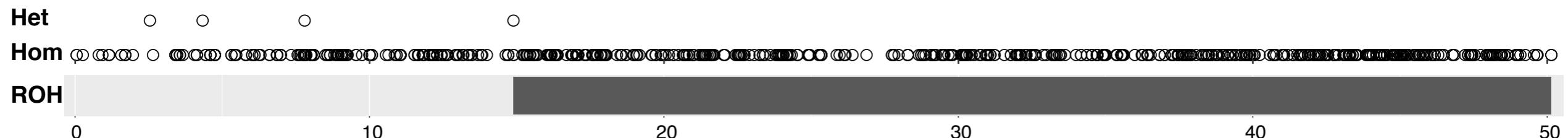
SGB.A10 (Domestic goat)



Z23 (Iberian ibex)

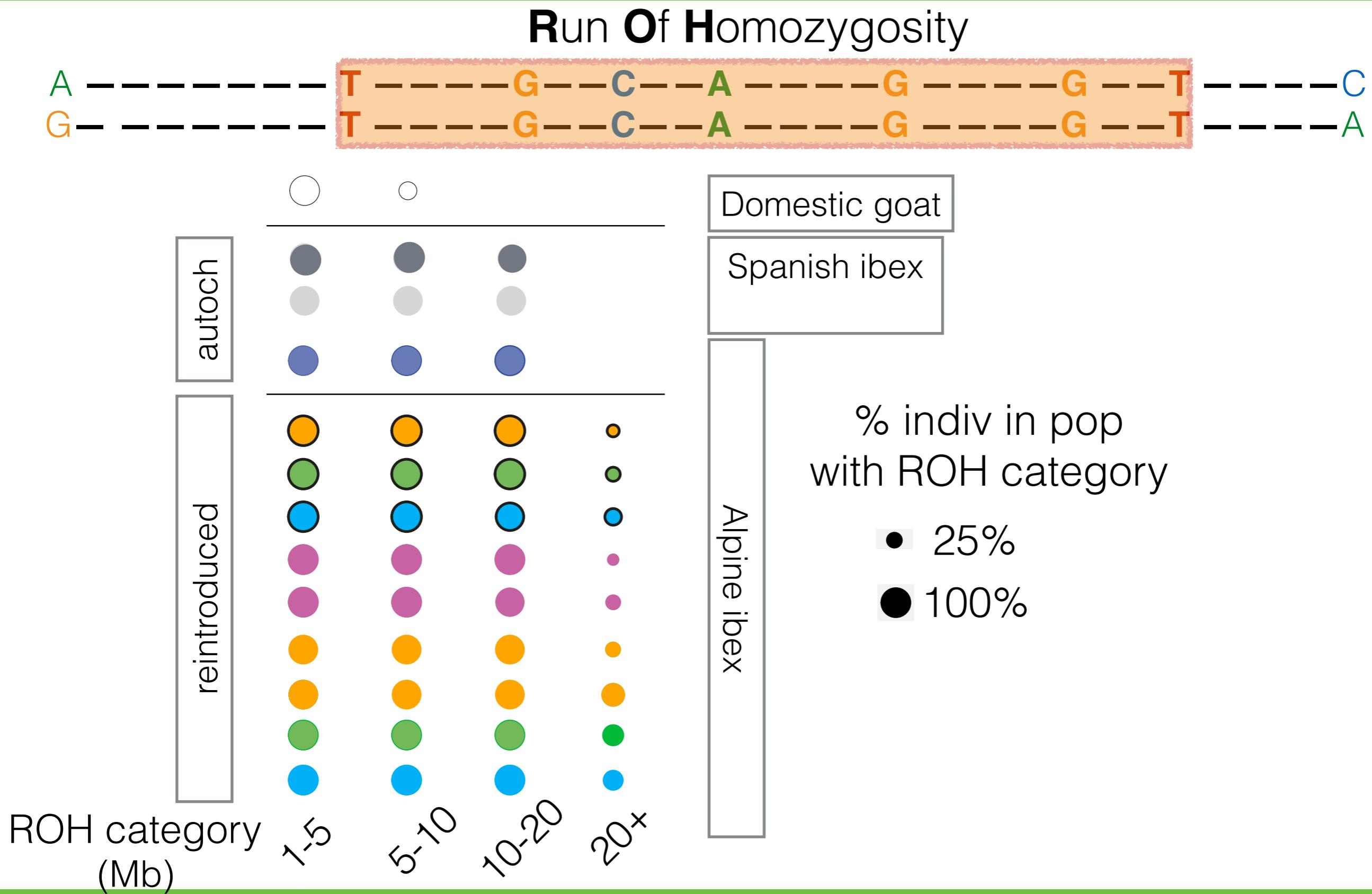


VS0034 (Alpine ibex)

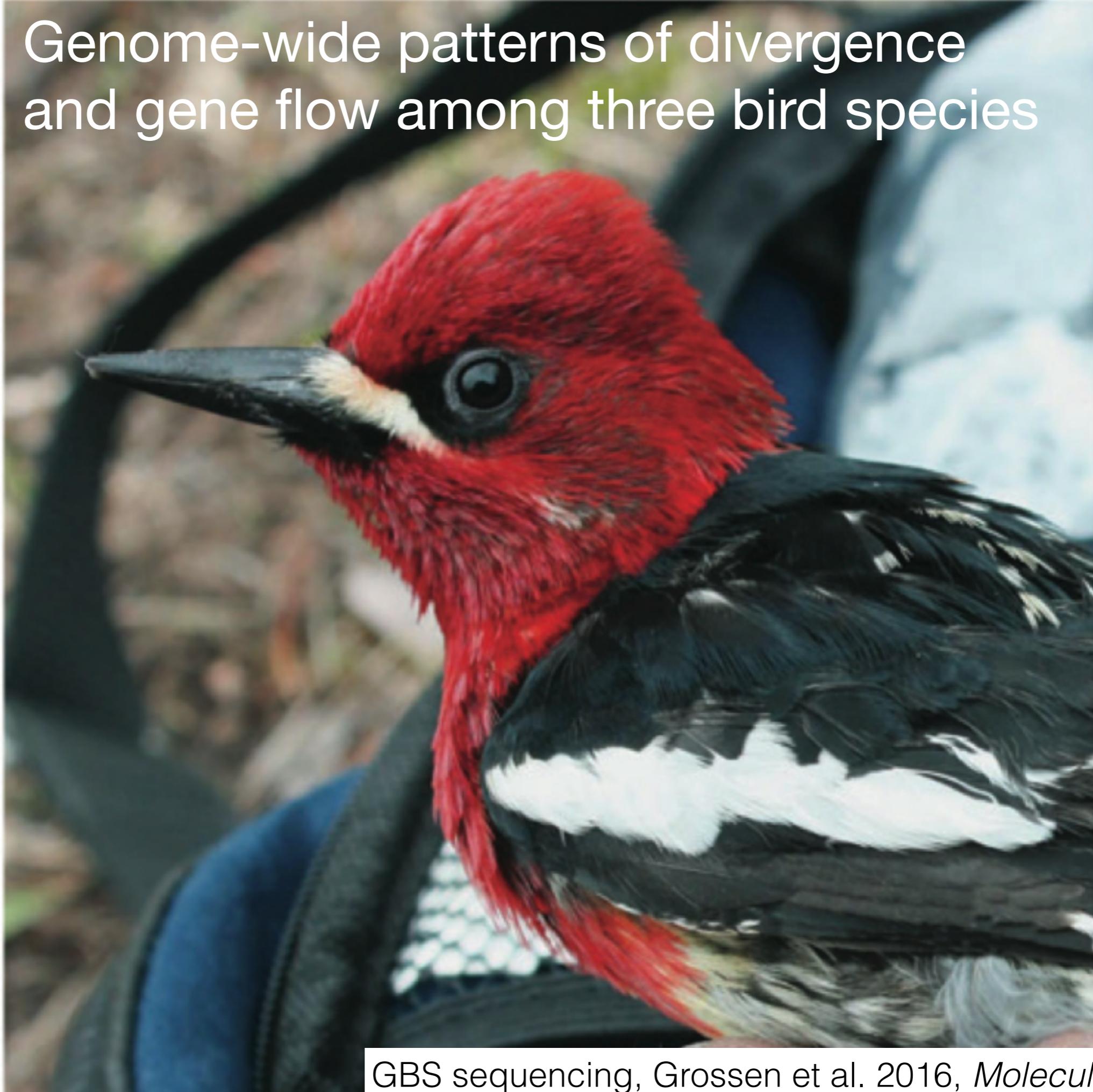


Chromosomal position in Mb

Longest ROH in reintroduced populations



Genome-wide patterns of divergence and gene flow among three bird species

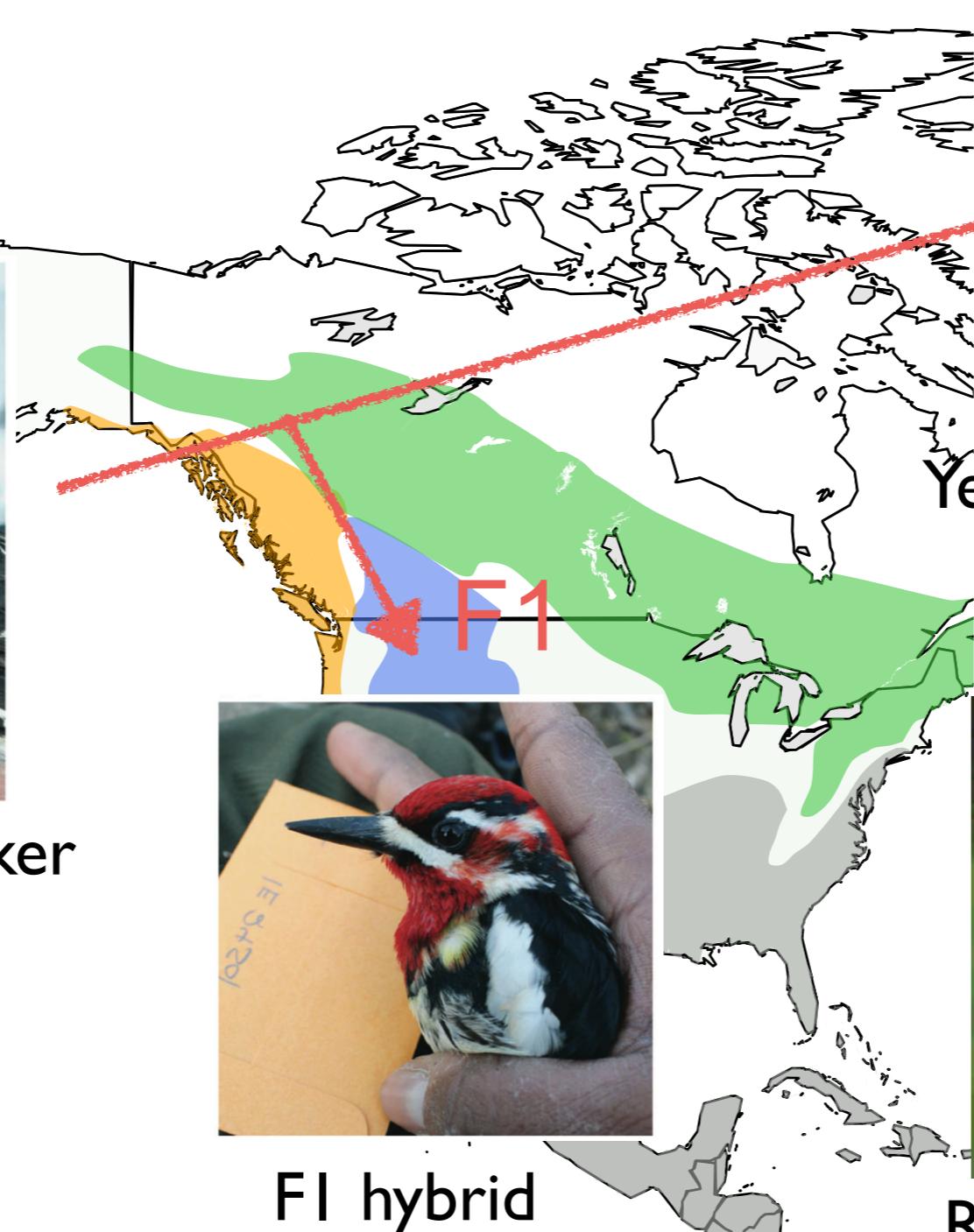


GBS sequencing, Grossen et al. 2016, *Molecular Ecology*

Three interbreeding bird species



Red-breasted sapsucker



F1 hybrid
Yellow-bellied x Red-breasted



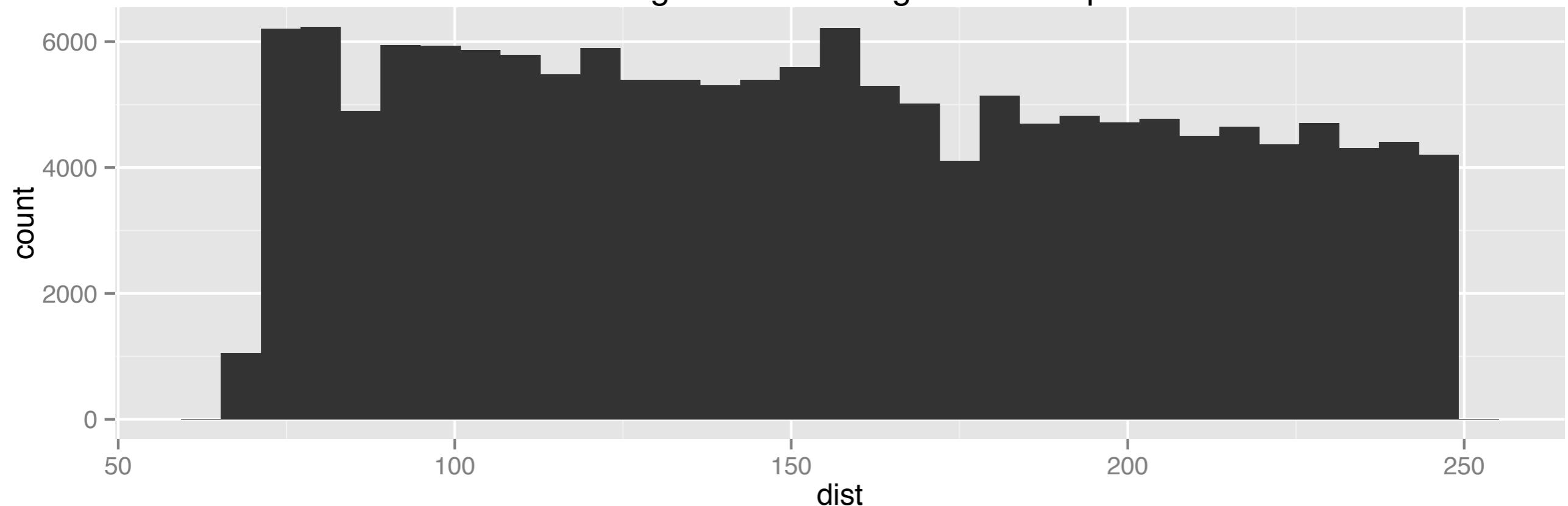
Yellow-bellied sapsucker



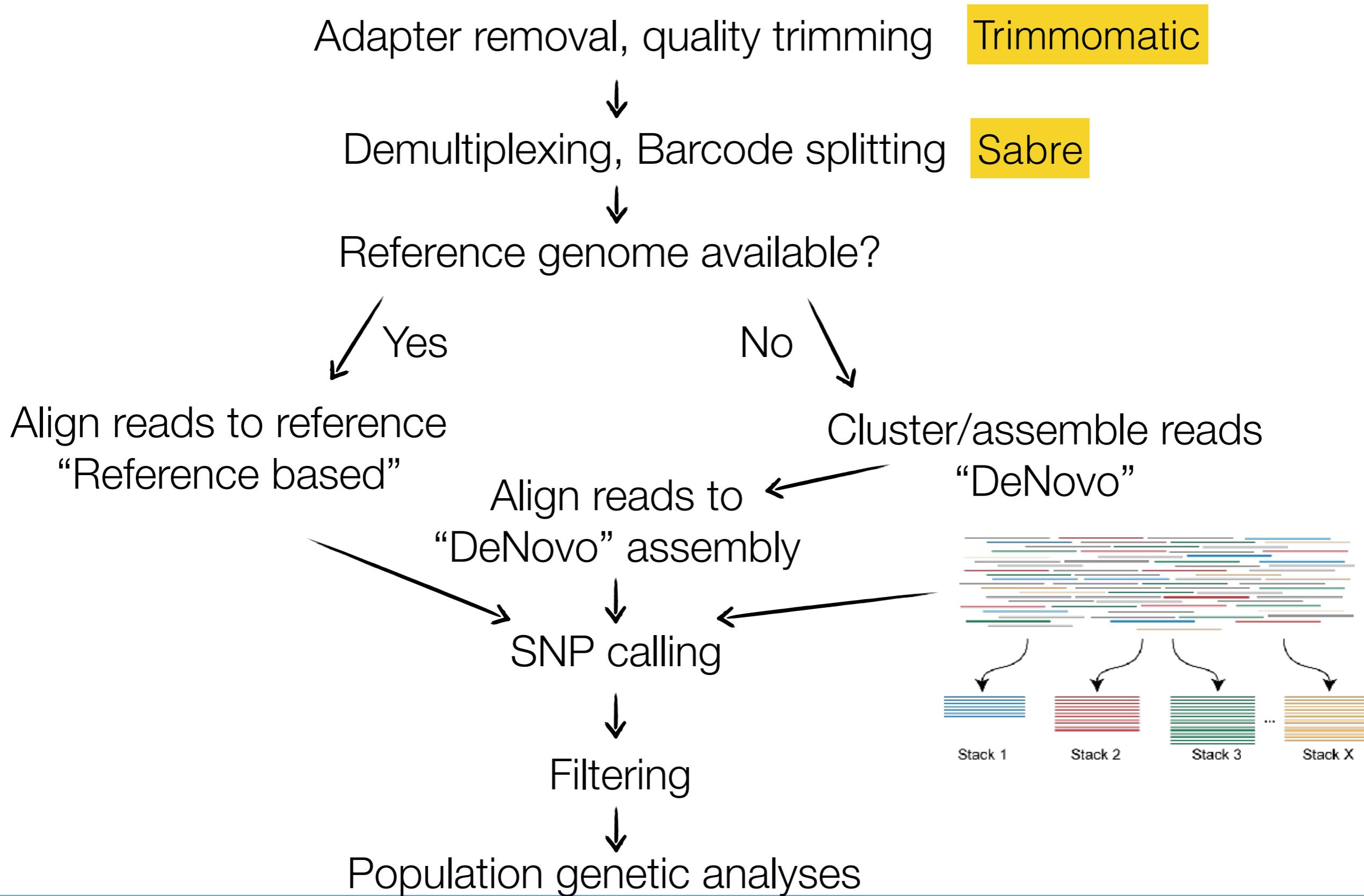
Red-naped sapsucker

GBS (Elshire et al. 2011) in Sapsuckers

Number of fragments with length 70–250bp: 156330



Data analysis with custom scripts



Data analysis with custom scripts



Downy woodpecker

14'000 contigs

Align reads to reference
“Reference based”

Adapter removal, quality trimming



Demultiplexing, Barcode splitting



Reference genome available?

Yes

No

Align reads to

Cluster/assemble reads

“DeNovo”

What if we would still be interested in genomic positions

Filtering



Population genetic analyses

Stack 1

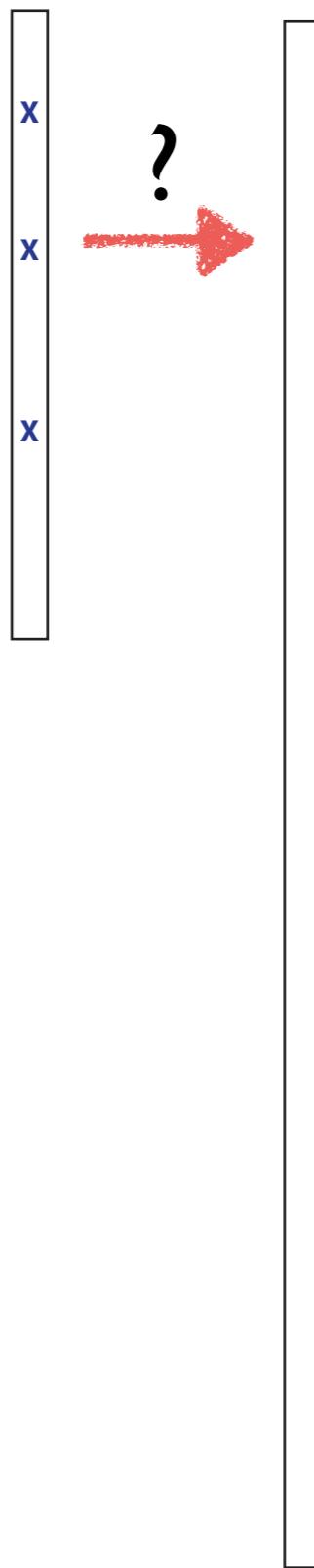
Stack 2

Stack 3

Stack X

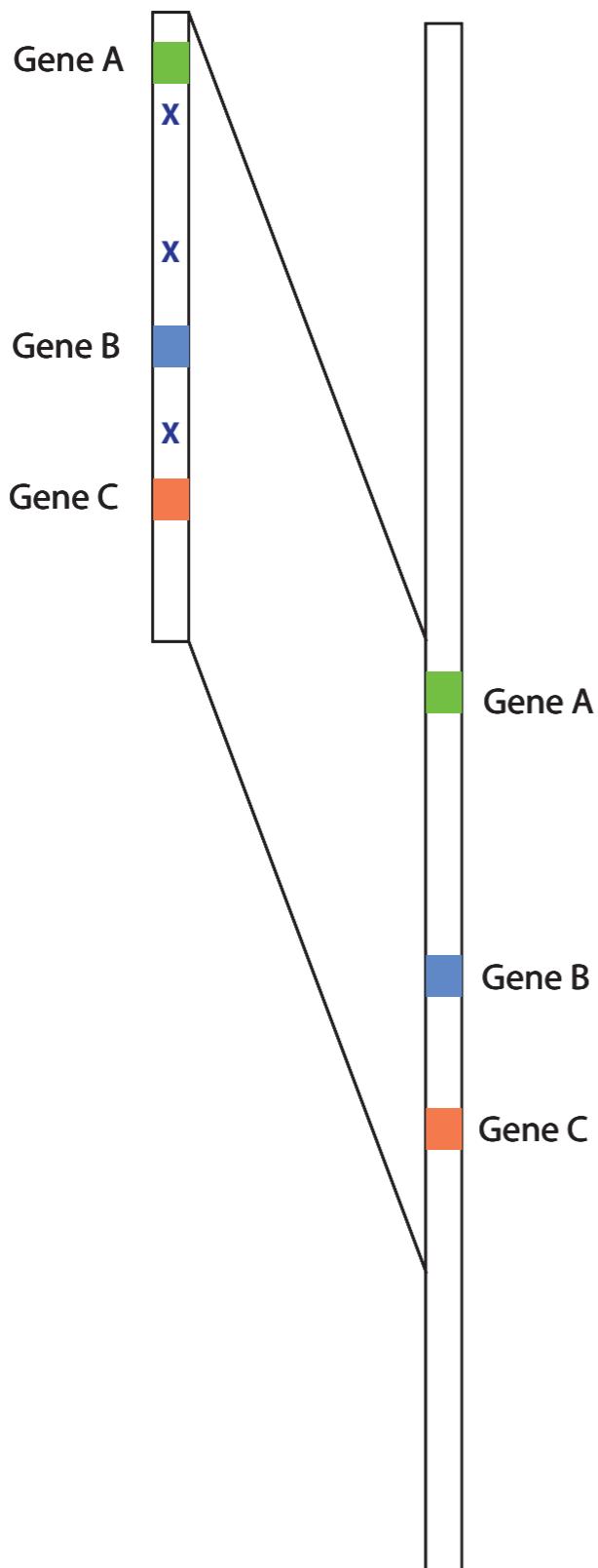
Estimate chromosomal positions

Downy woodpecker Zebra finch
scaffold XY chromosome

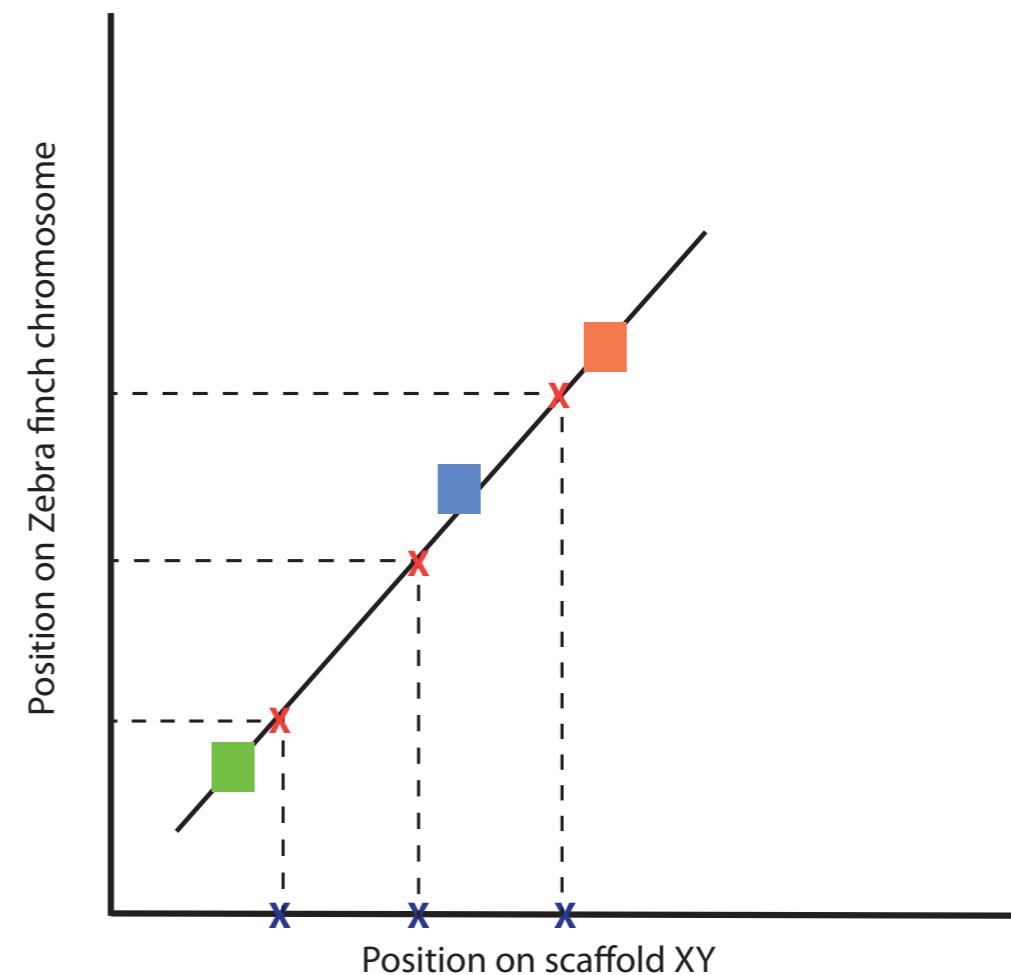
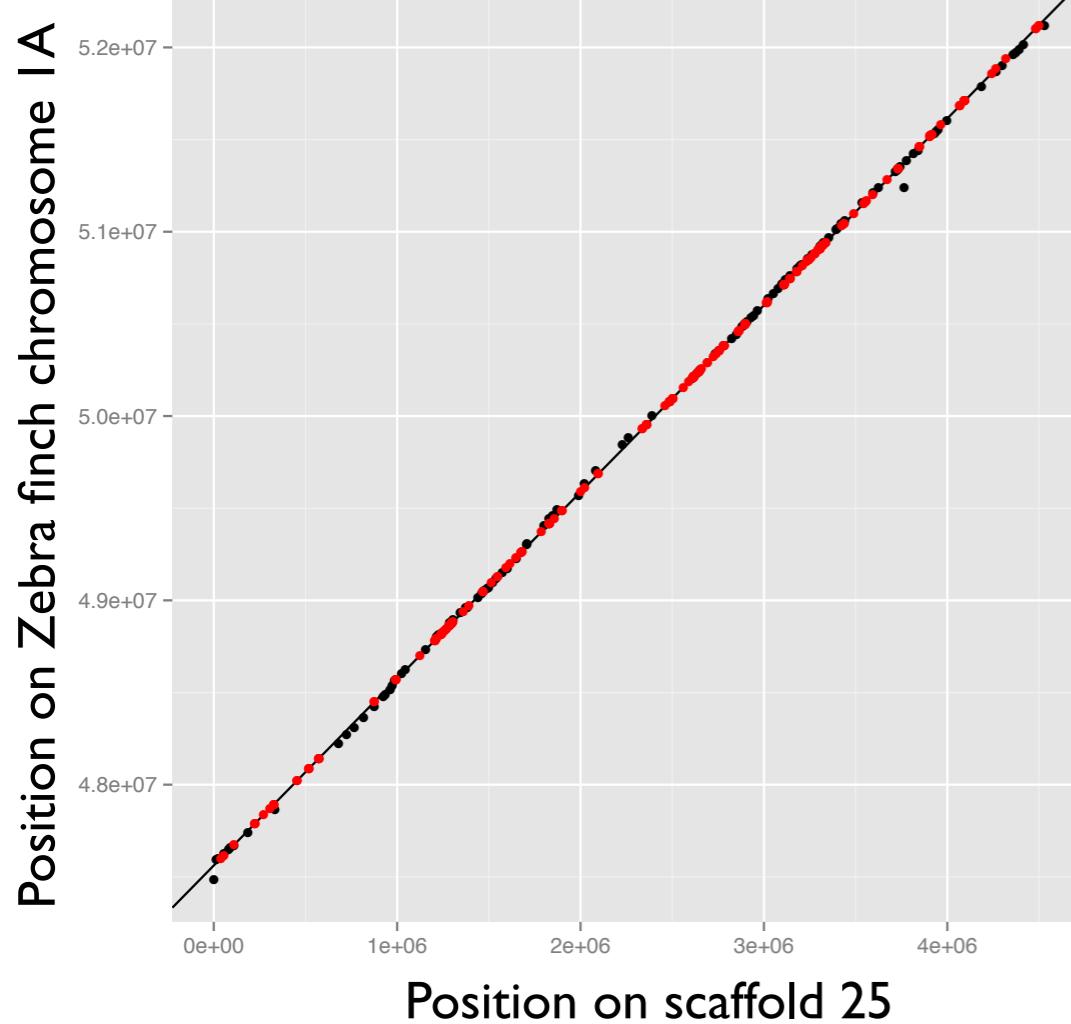


Estimate chromosomal positions

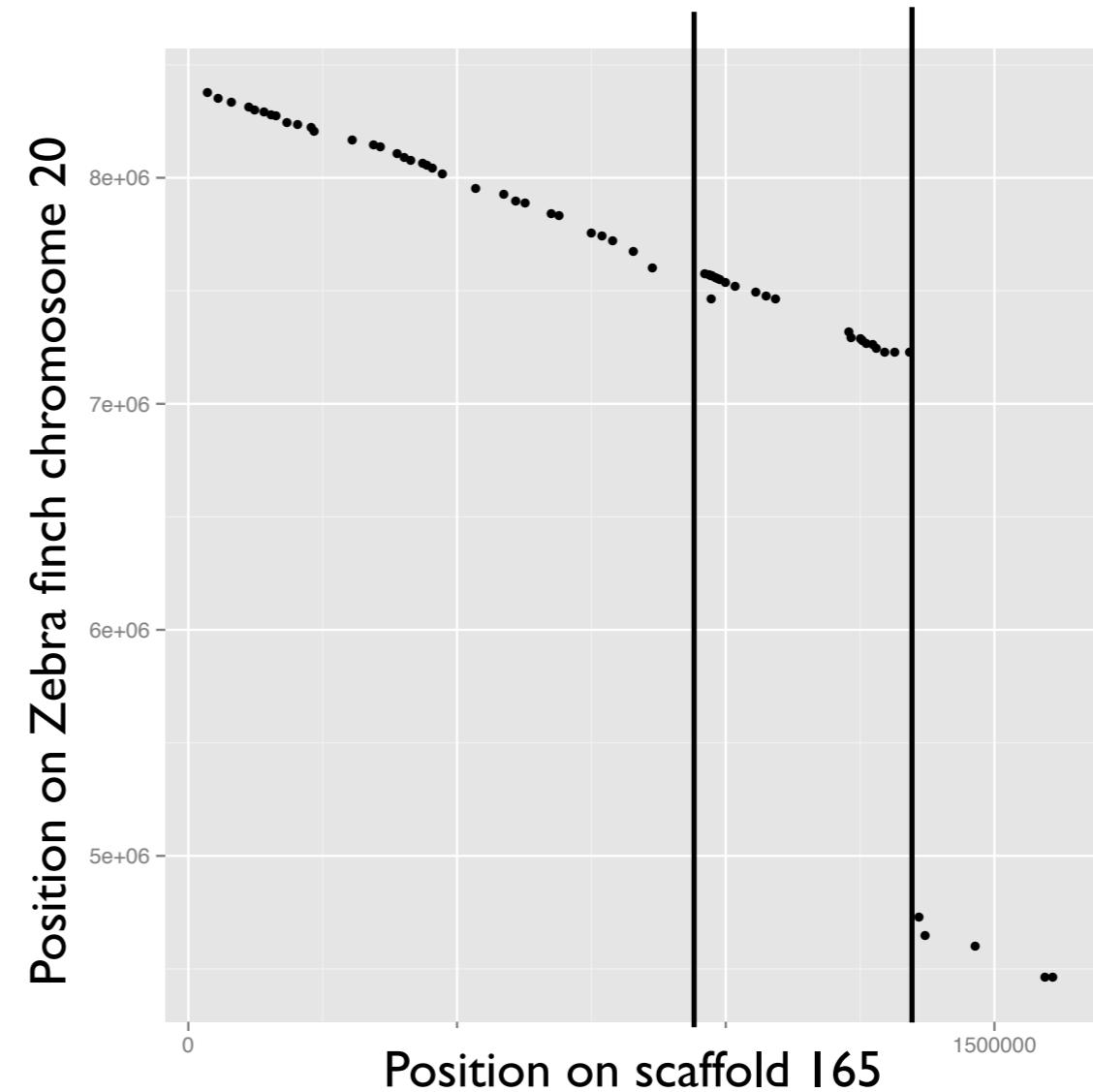
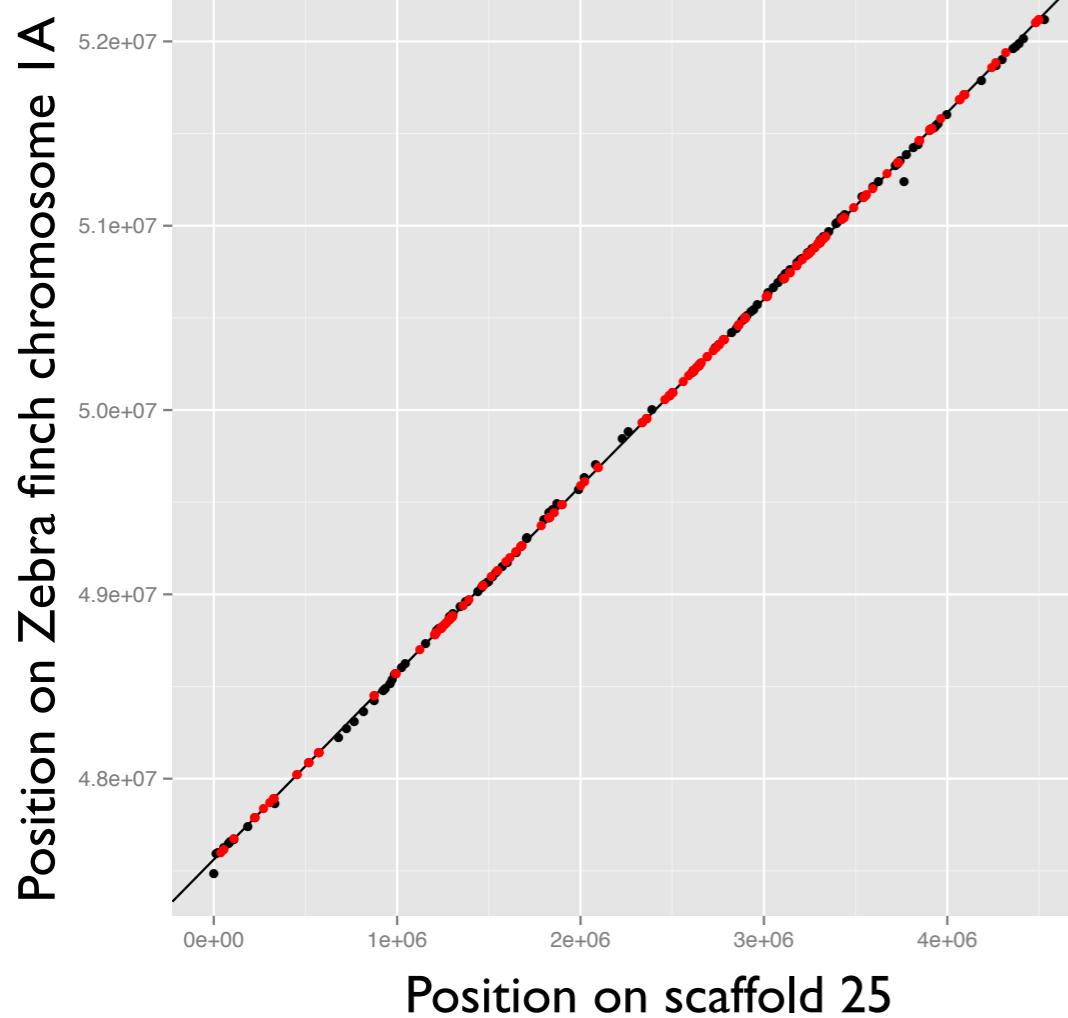
Downy woodpecker Zebra finch
scaffold XY chromosome



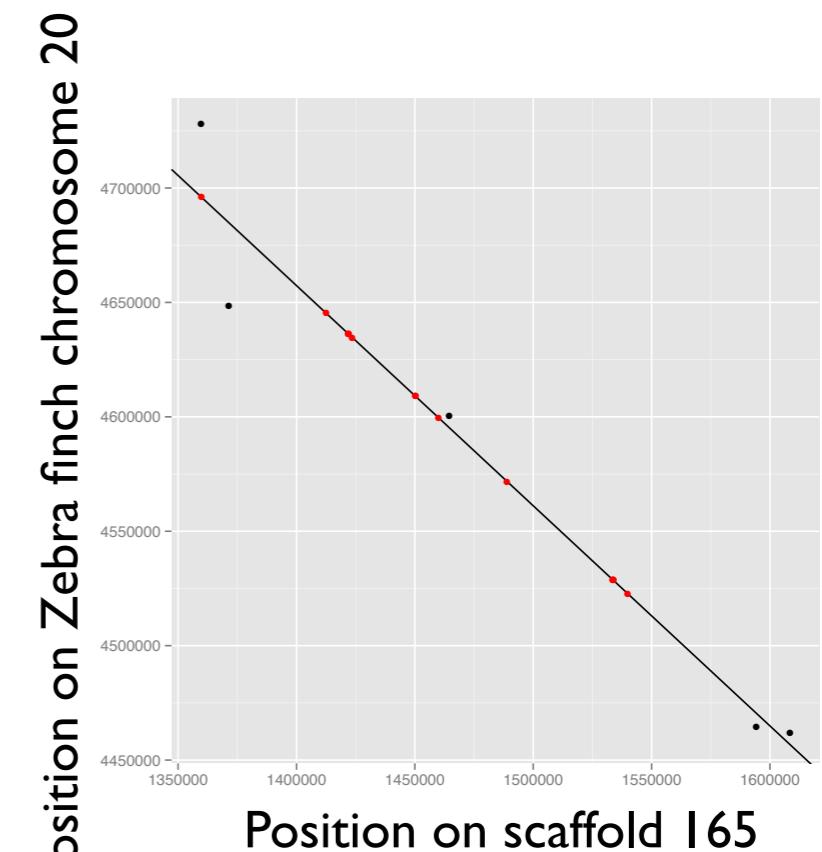
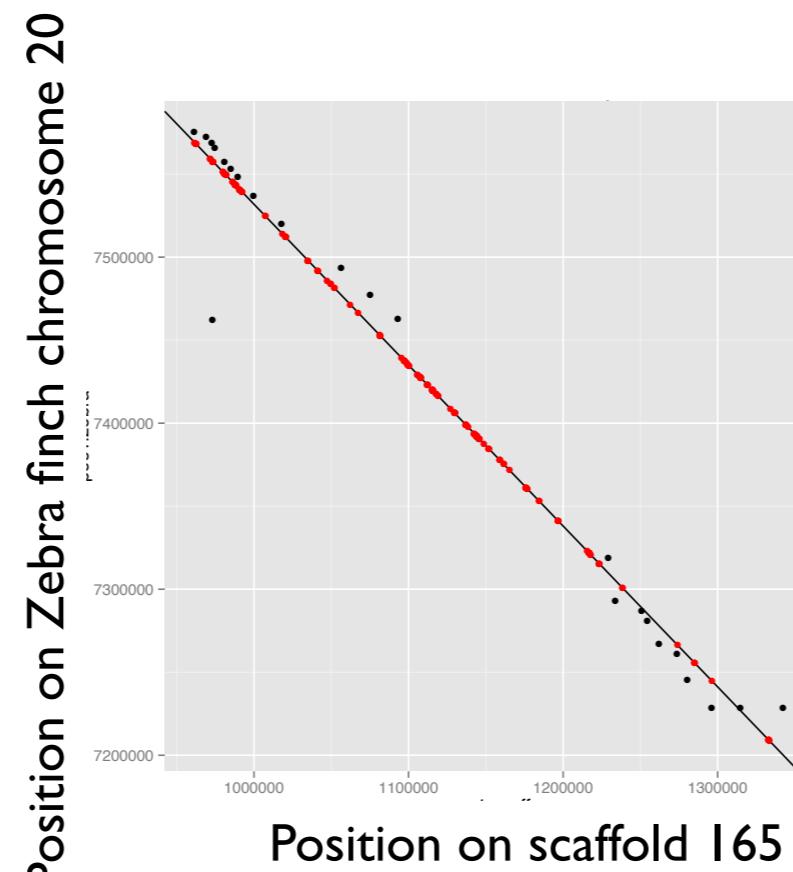
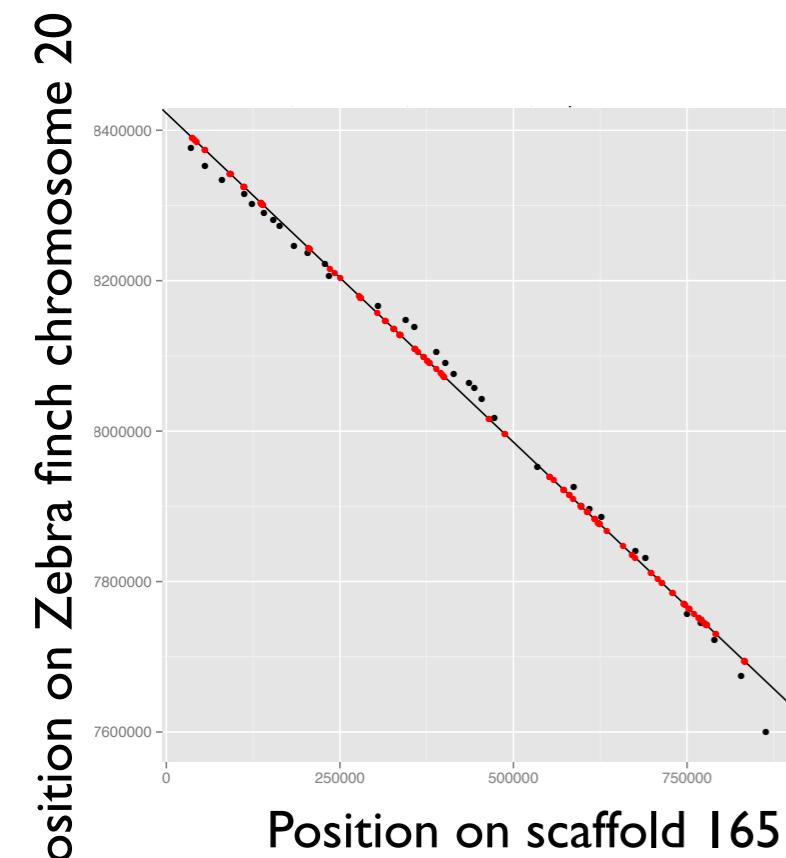
Collinearity



Some scaffolds had to be split



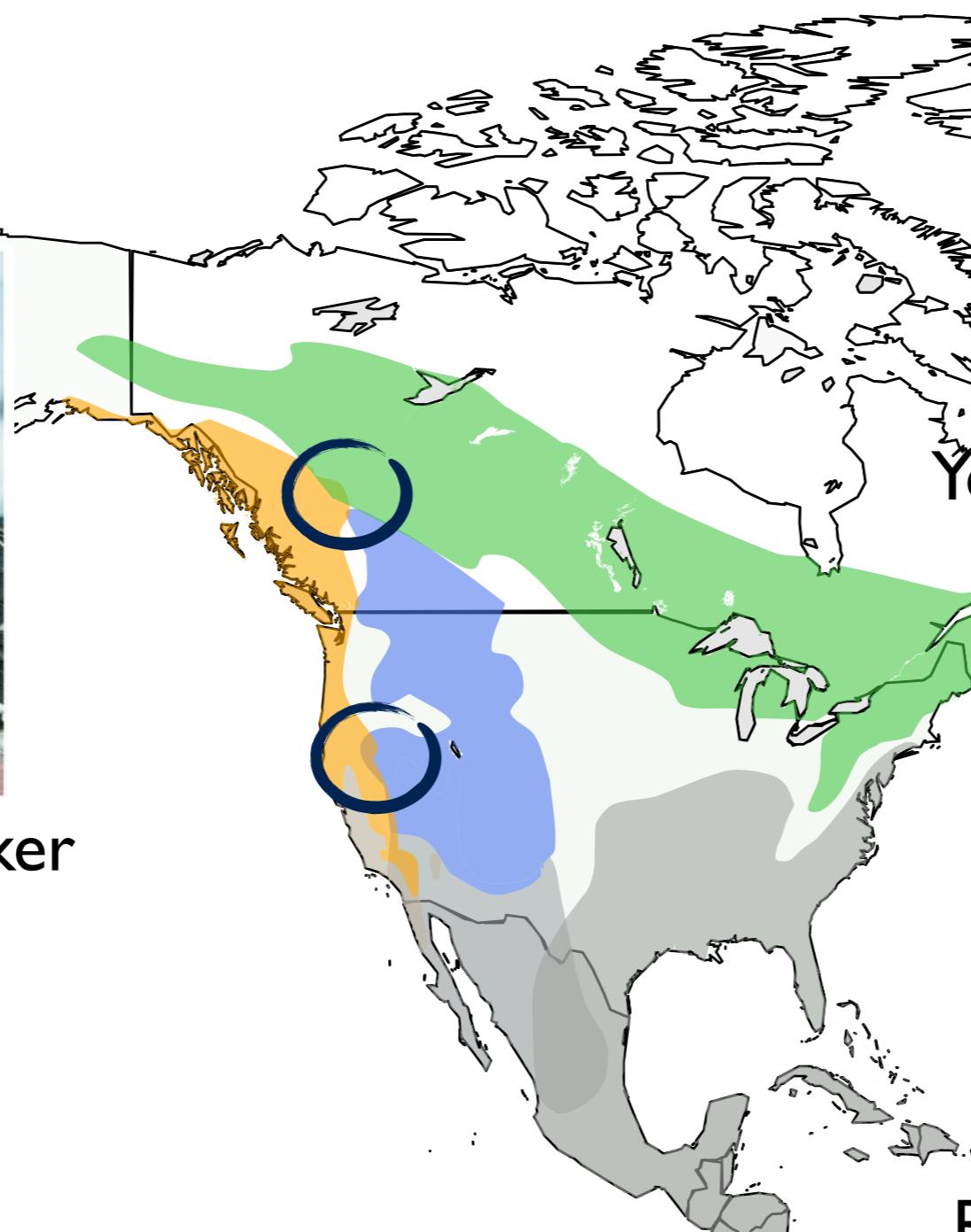
Some scaffolds had to be split



Hybrid zones allow the study of reproductive isolation



Red-breasted sapsucker

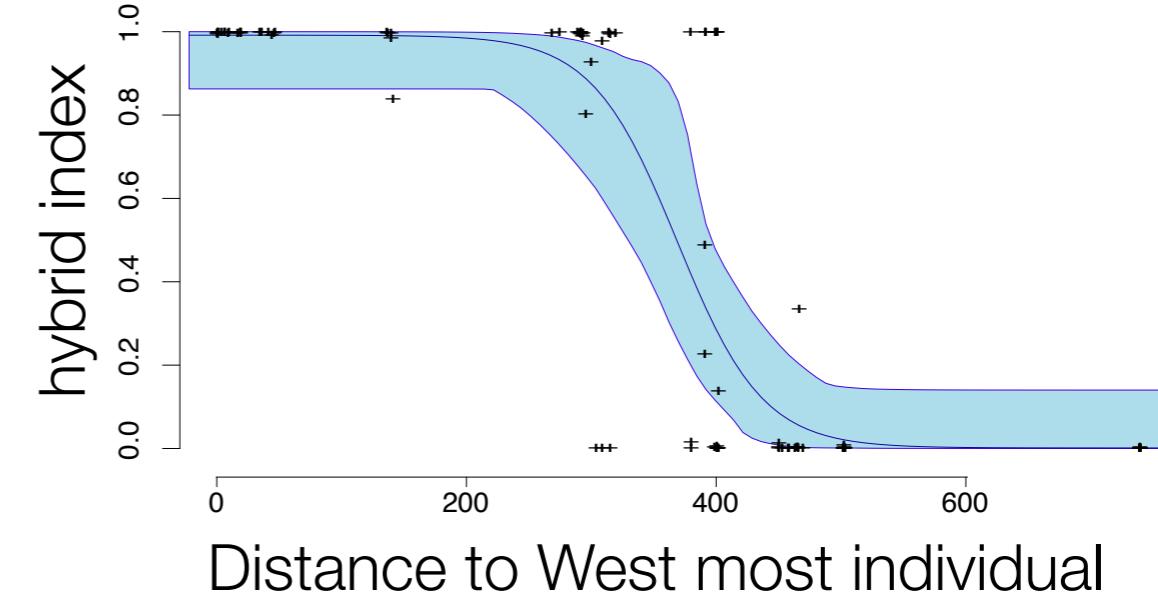
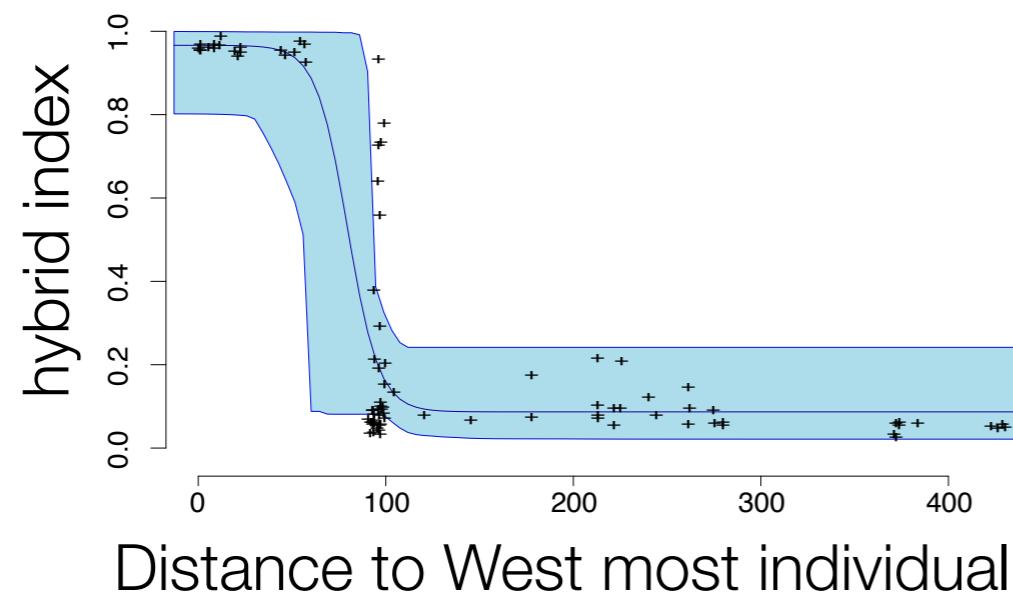
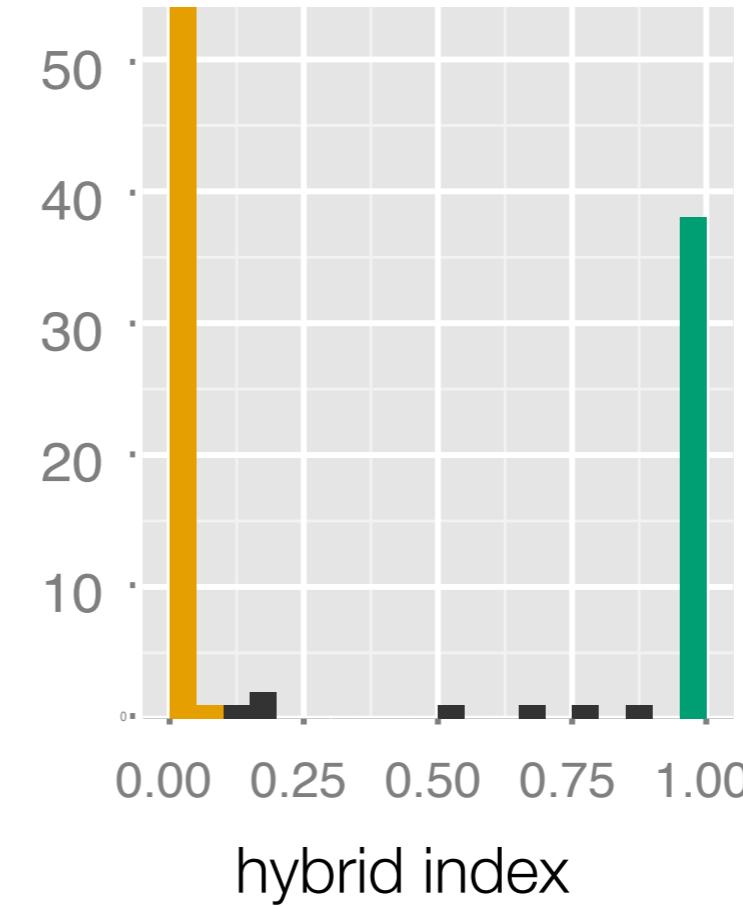
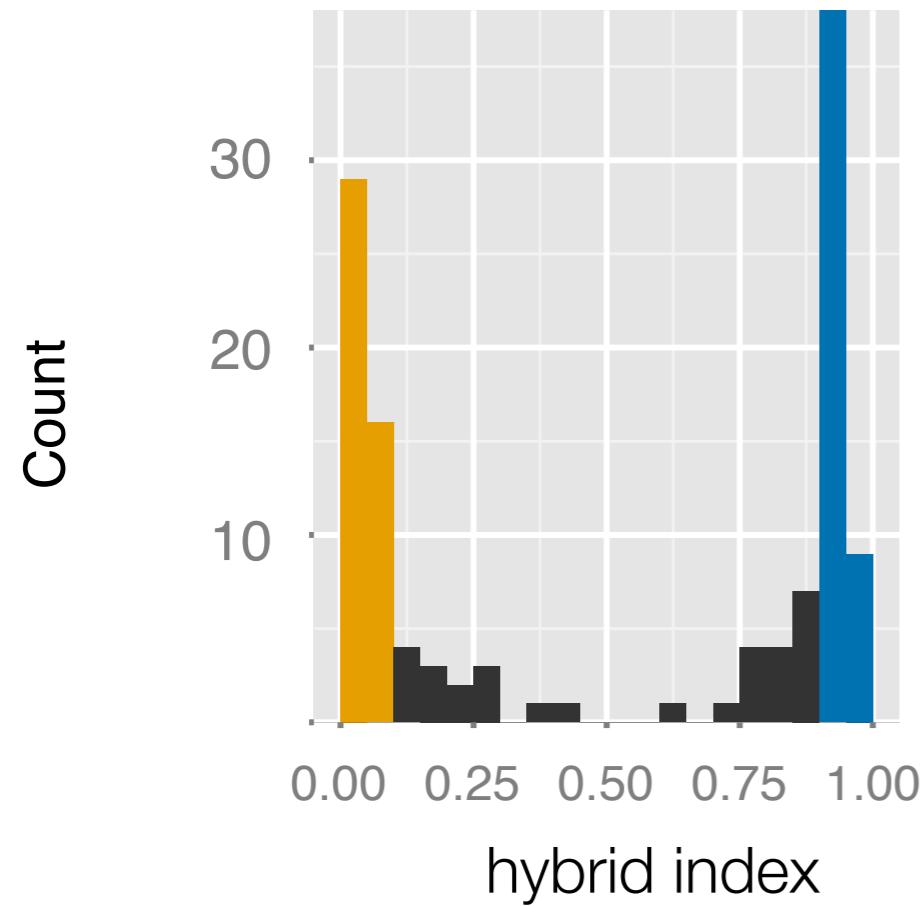


Yellow-bellied sapsucker

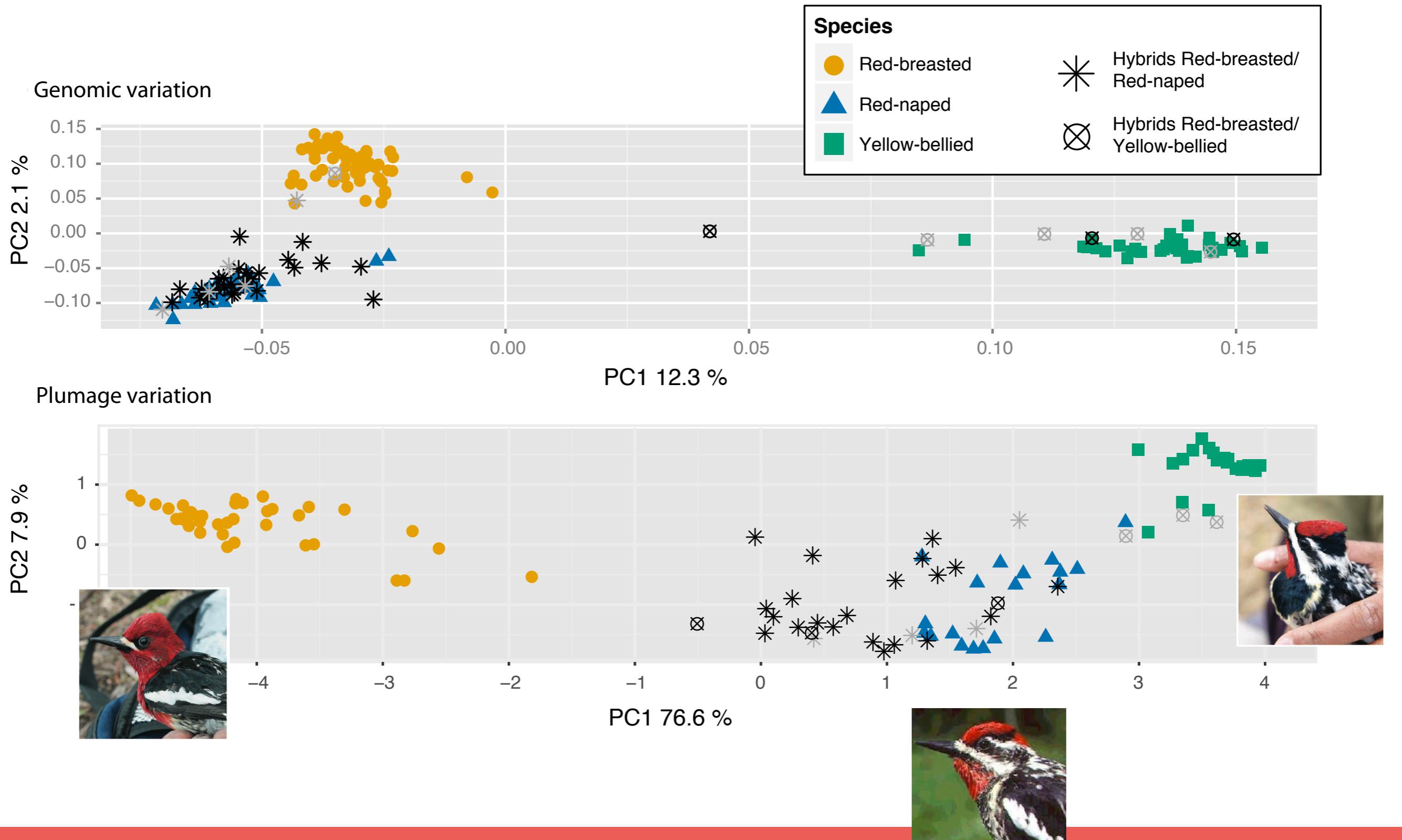


Red-naped sapsucker

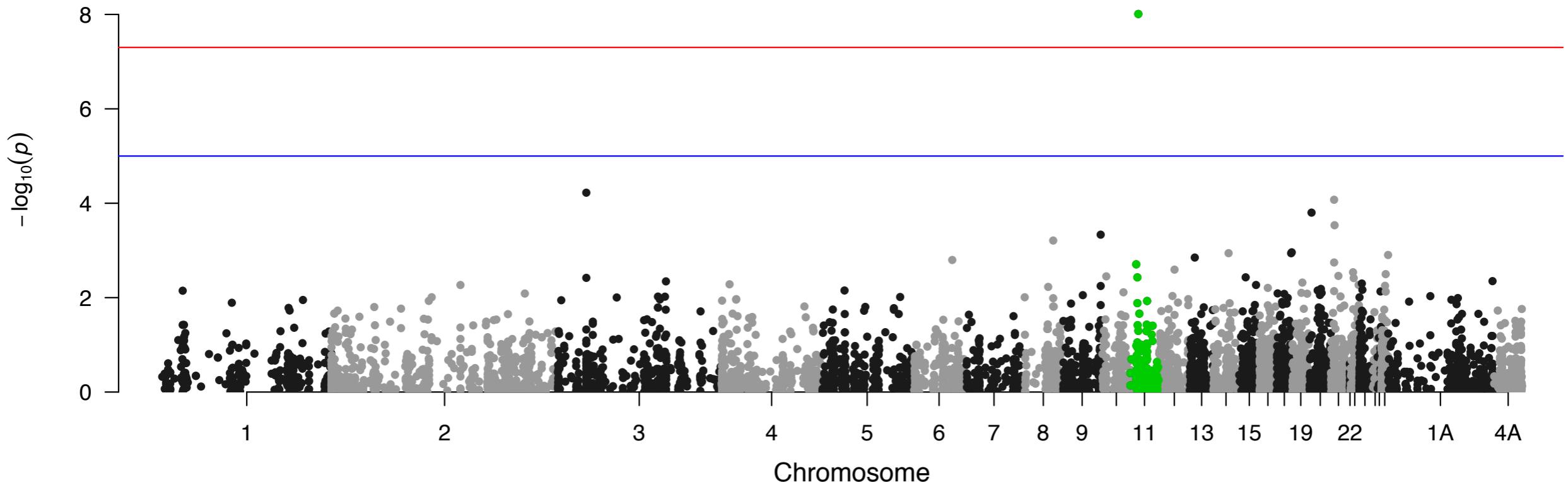
Strong reproductive isolation



Phenotype-genotype mismatch



One locus associated with plumage



more red



Questions?

