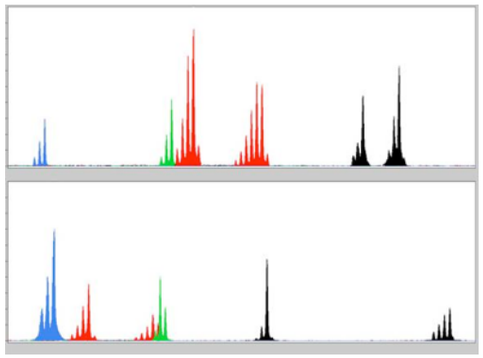
RAD sequencing

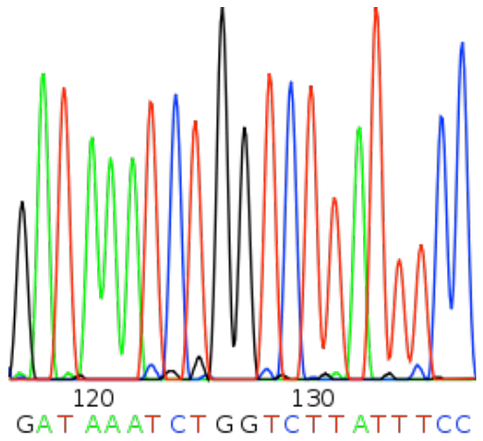
Methods and application

Christine Grossen, University of Zürich

How to bridge the gap?



microsatellites



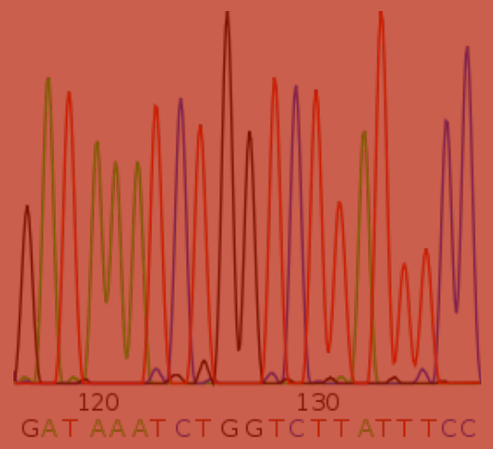
Sanger sequencing

whole genome

How to bridge the gap?

LOW INFORMATION & LOW COST per sample

microsatellites



Sanger sequencing

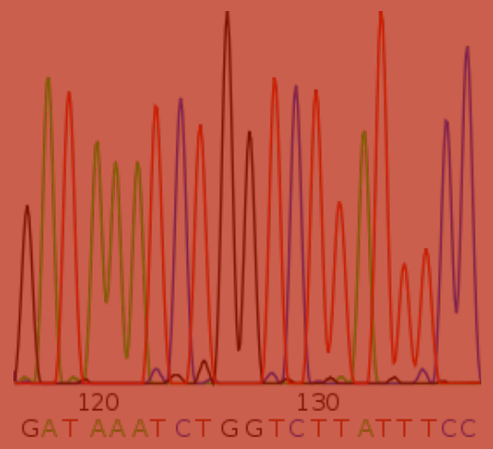
HIGH INFORMATION & HIGH COST per sample

whole genome

How to bridge the gap?

LOW INFORMATION & LOW COST per sample

microsatellites



Sanger sequencing

Reduced genome complexity using Restriction enzymes

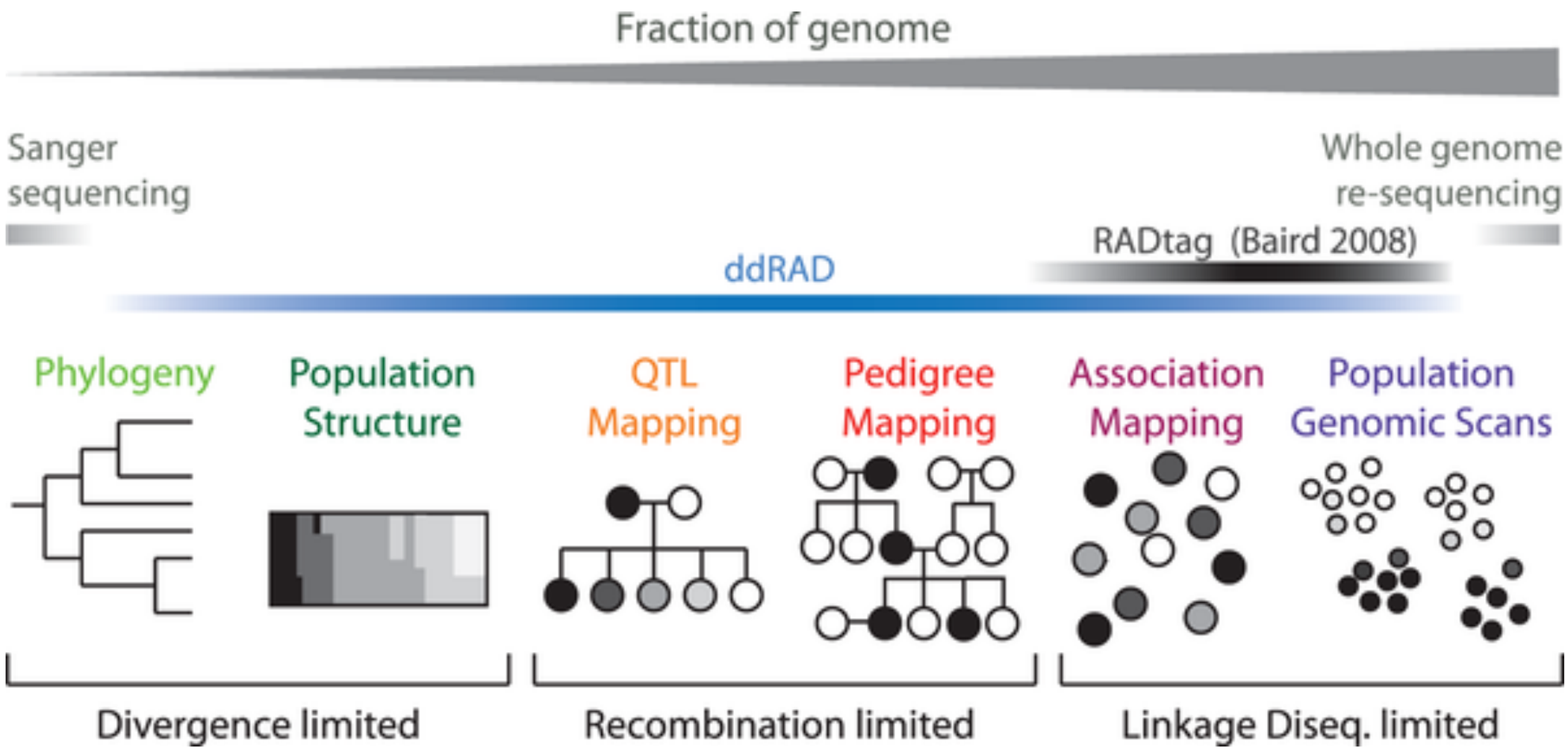
RAD, ddRAD, GBS, ...

SNP chips/arrays

HIGH INFORMATION & HIGH COST per sample

whole genome

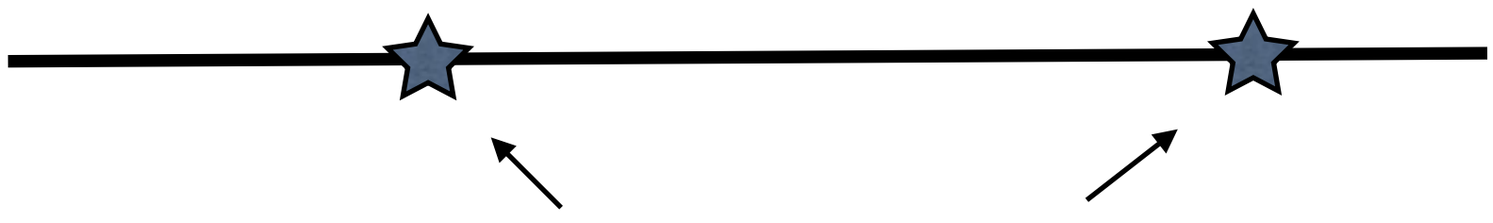
How to bridge the gap?



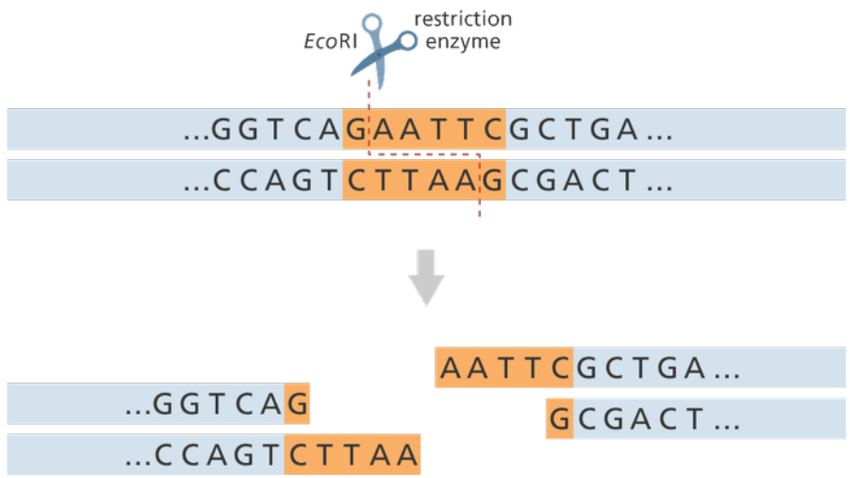
Peterson et al. 2012

What is RAD?

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



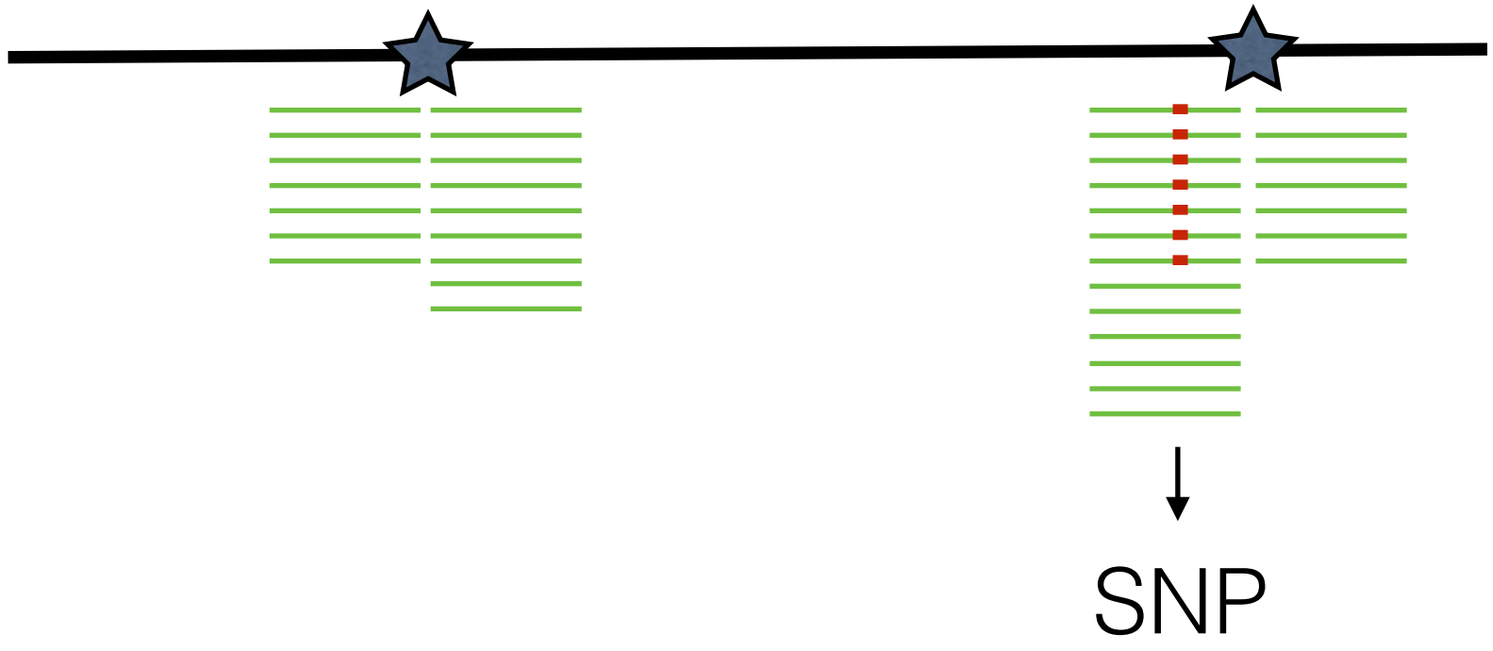
Restriction sites



picture http://www.yourgenome.org

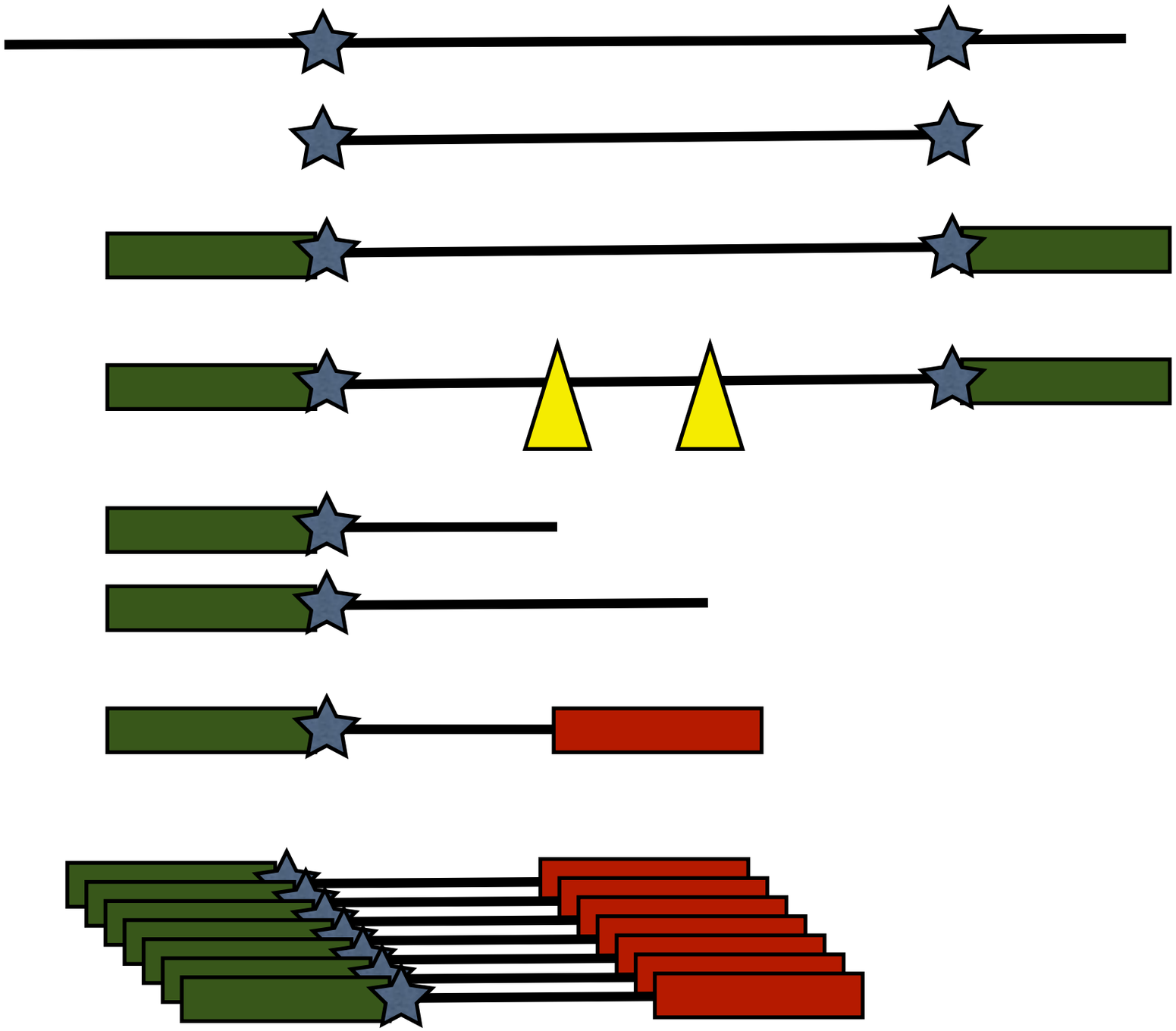
What is RAD?

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



picture http://www.yourgenome.org

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



Genomic DNA digestion

Fragmented DNA

Ligation of adapter 1 (P1)

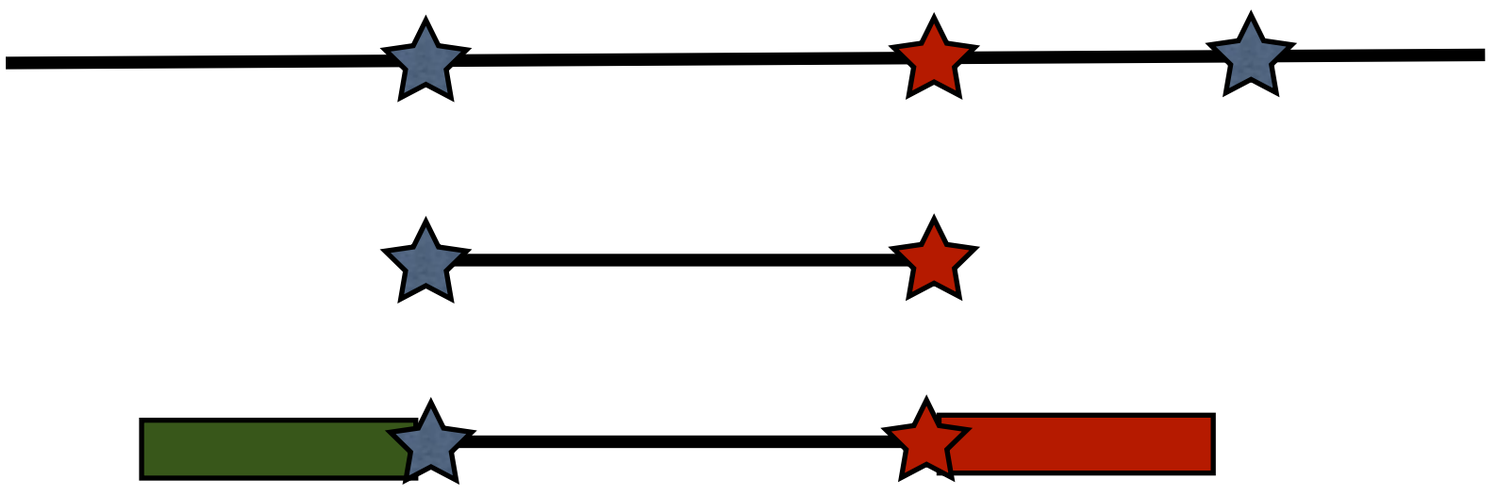
Shearing (sonication)

Size range: ~ 400-600bp

Ligation of adapter 2 (P2)

Low-cycle PCR amplification

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

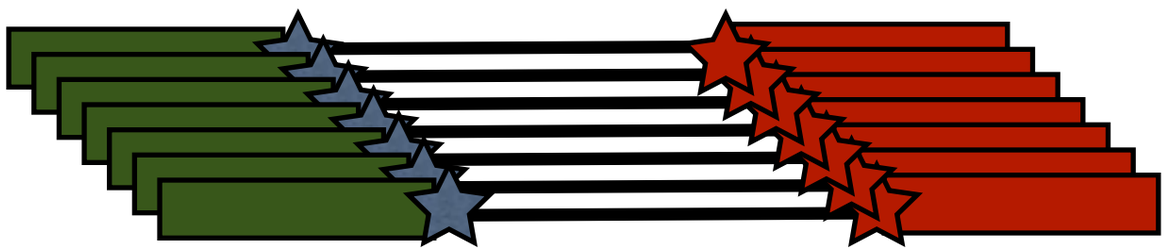


Genomic DNA digestion with two enzymes

Fragmented DNA

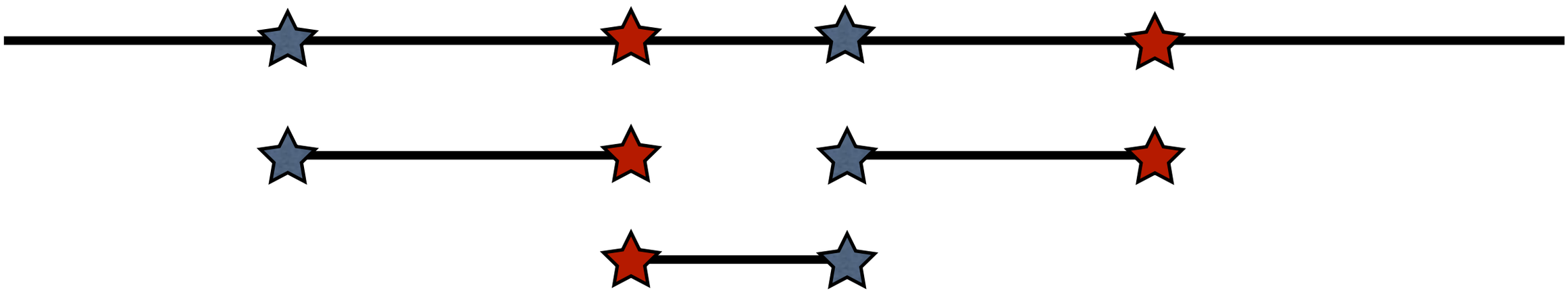
Ligation of adapters

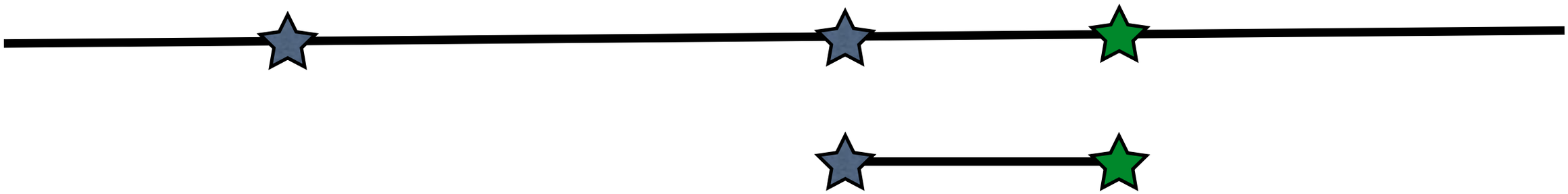
With or without size selection



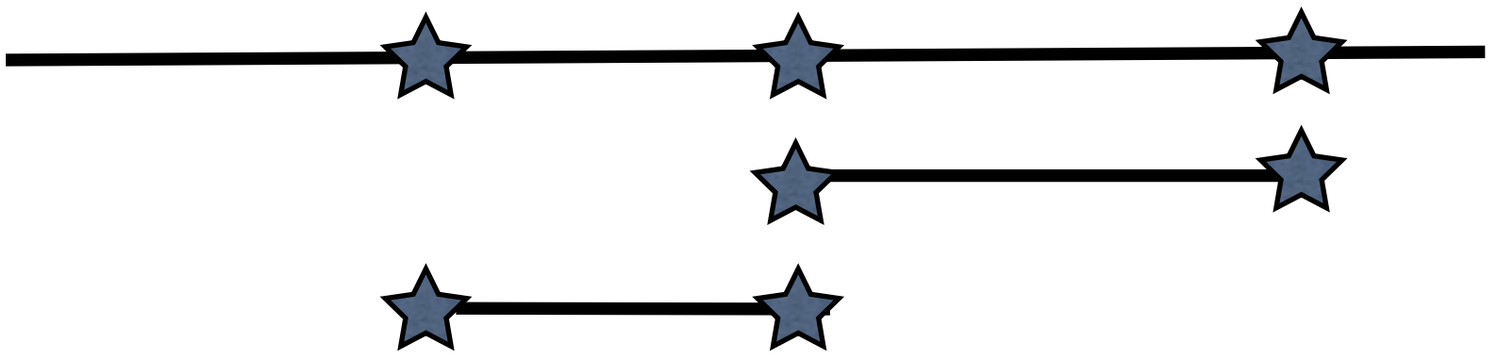
Low-cycle PCR amplification

ddRAD allows flexibility in marker number



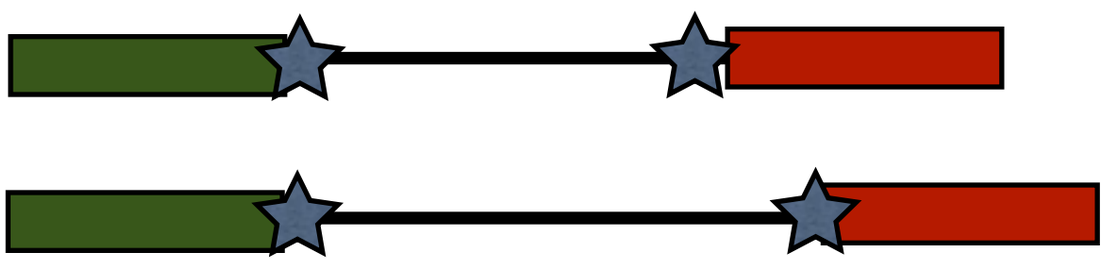


More flavours: GBS (single digest, Elshire et al. 2011)



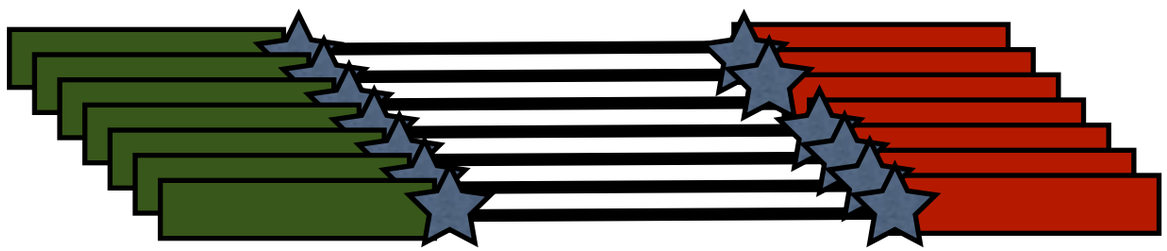
Genomic DNA digestion

Fragmented DNA



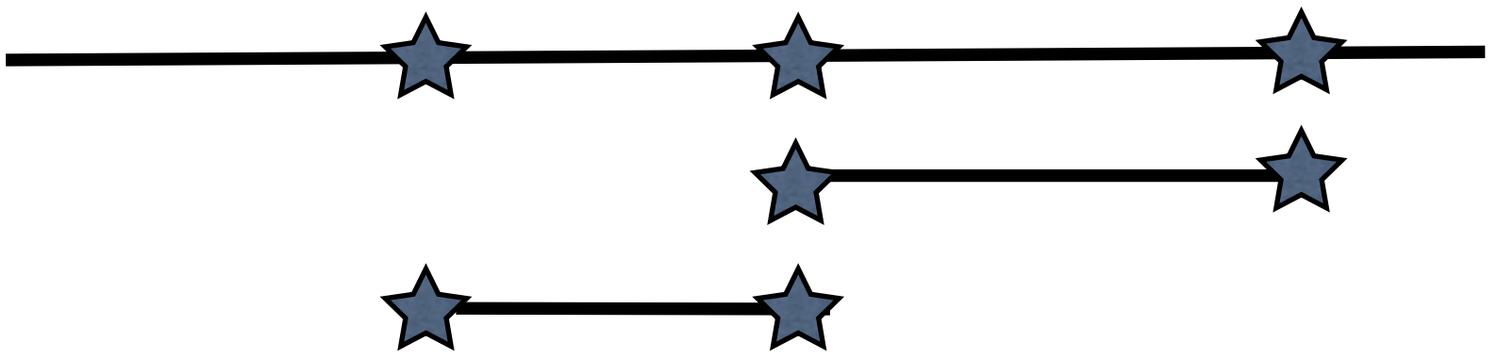
Ligation of adapters

Usually with some kind of size selection



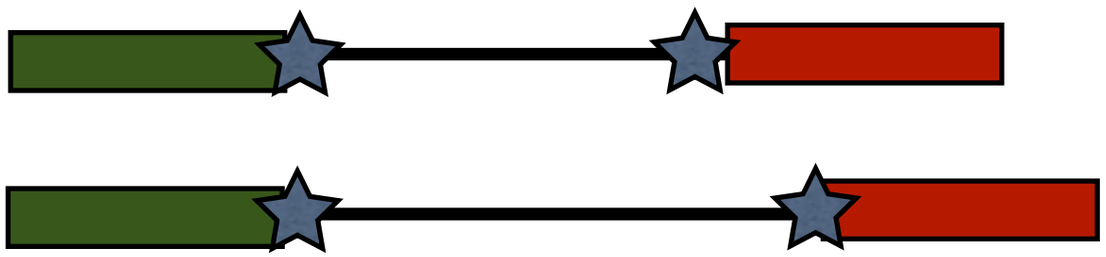
Low-cycle PCR amplification

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



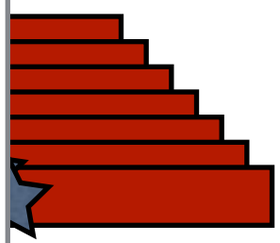
Genomic DNA digestion

Fragmented DNA



Ligation of adapters

Usually with some kind of size selection



Low-cycle PCR amplification

Yet other RAD flavours:
ezRAD (Toonen *et al.* 2013)
2bRAD (Wang *et al.* 2012)

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)?

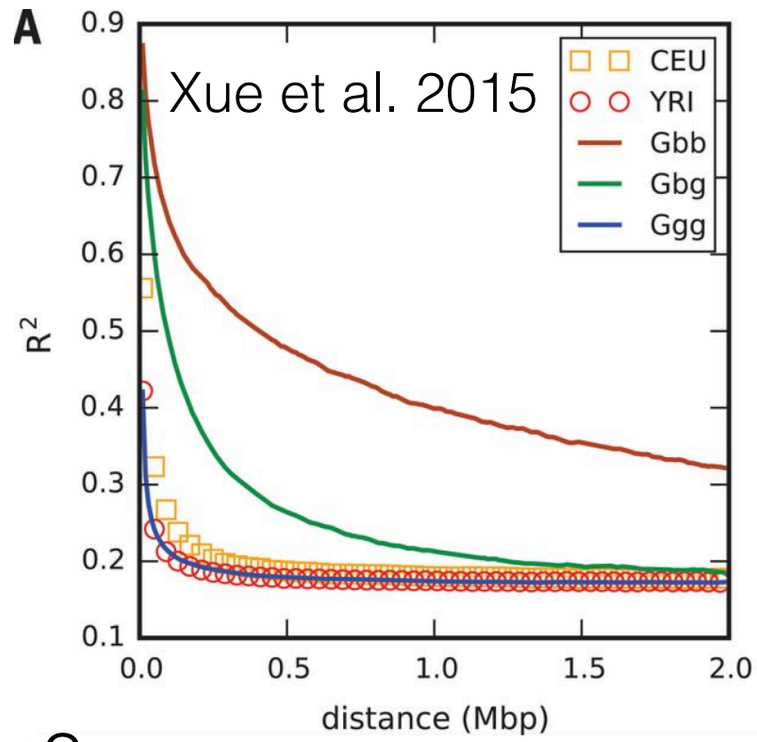
How many markers, marker density?

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)

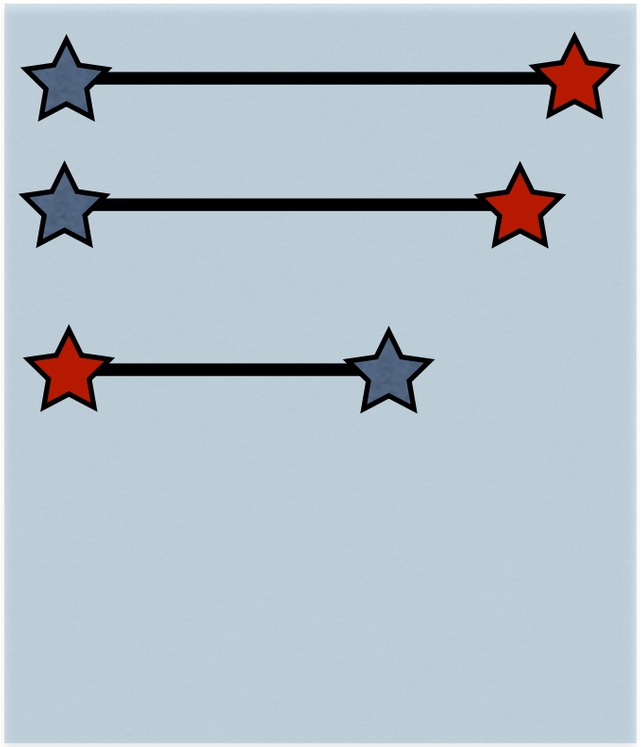
May also consider to outsource library prep and sequencing; e.g. Genomic Diversity Facility, Cornell

Lab considerations

Need high quality DNA (preferably tissue)
Large quantities of DNA
Taxa specific protocol optimisation

Once protocol works, ~1 week of work

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

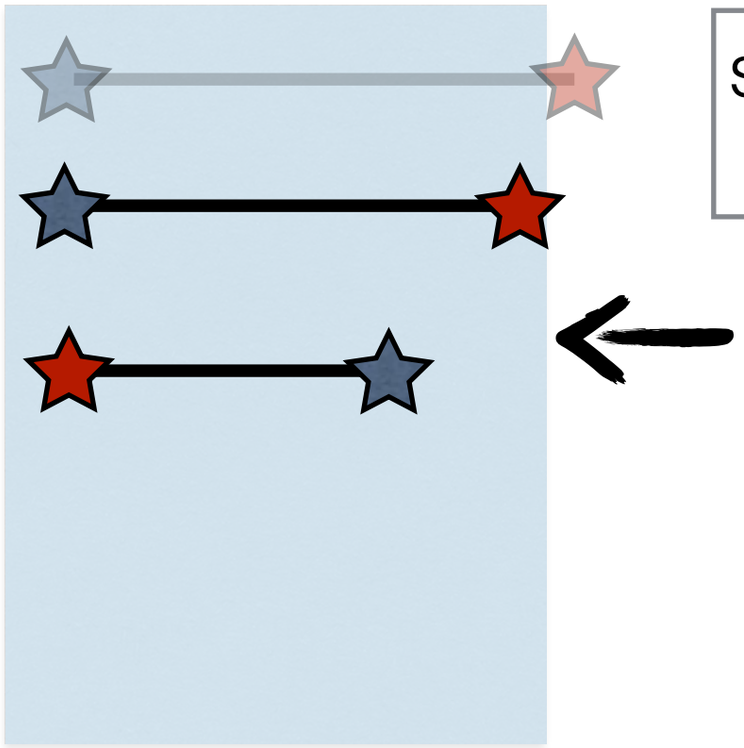


Lab considerations

Need high quality DNA (preferably tissue)
Large quantities of DNA
Taxa specific protocol optimisation

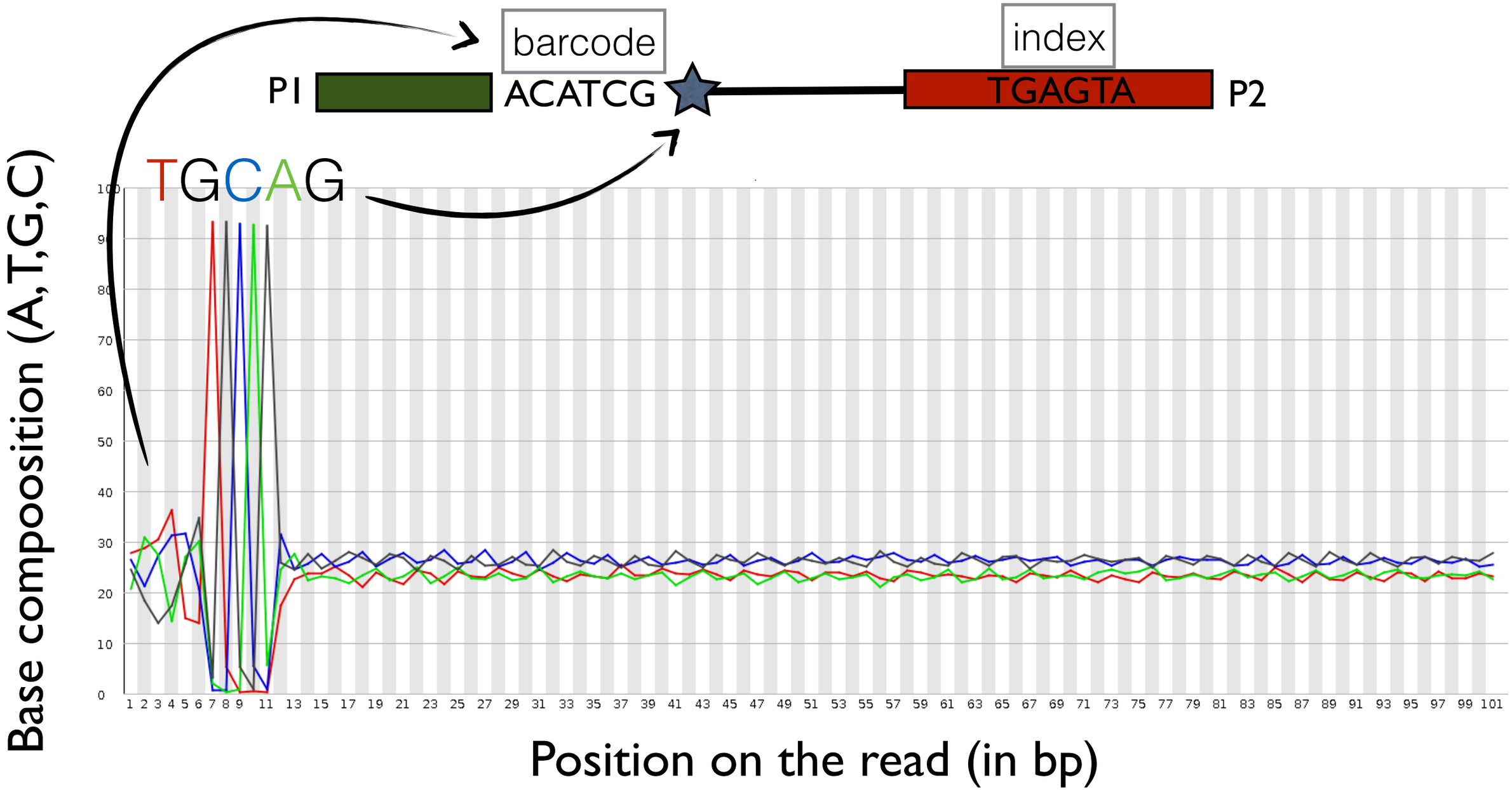
Once protocol works, ~1 week of work

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



shift in selected size leads to different set of markers

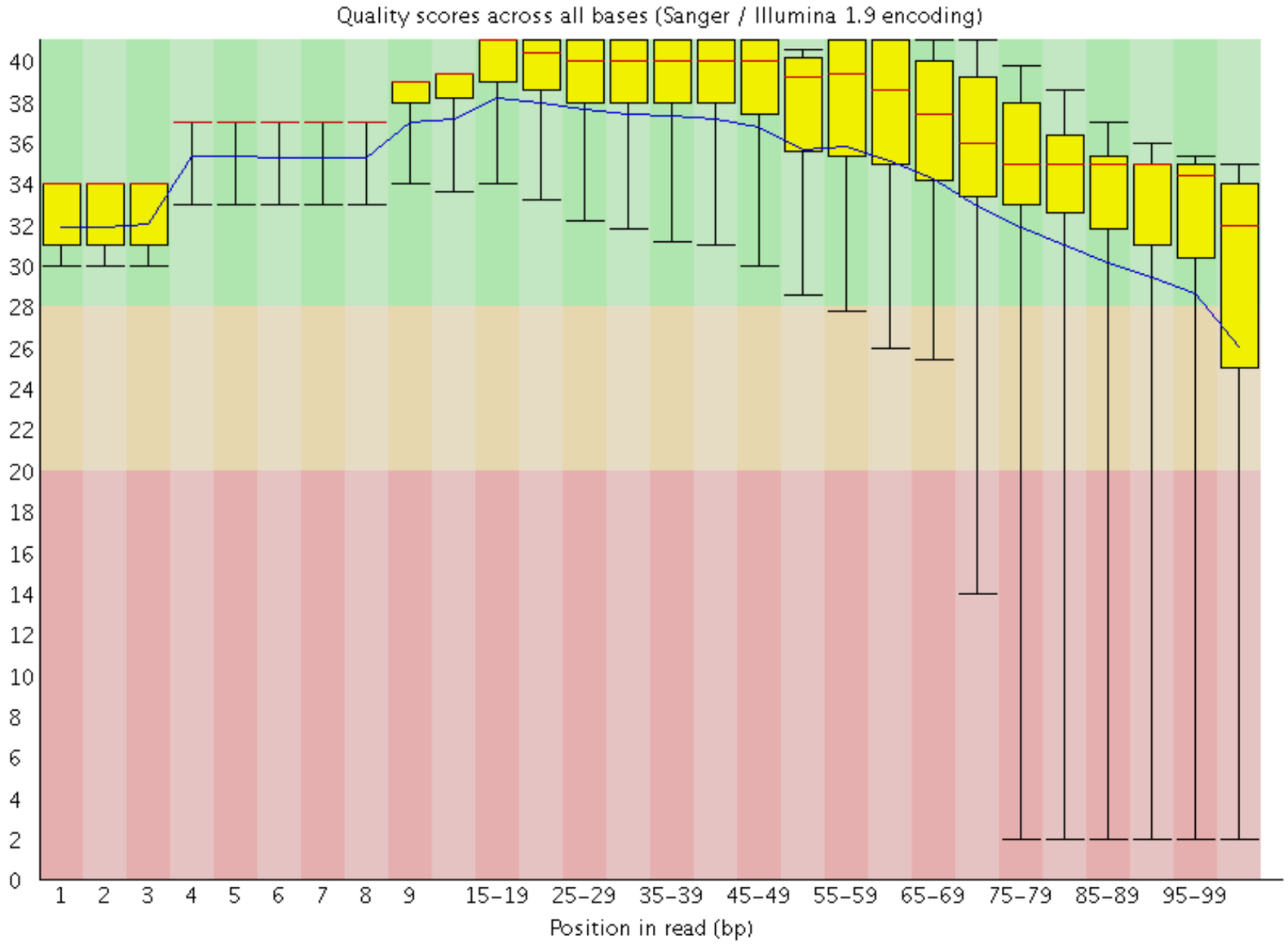
Problem of low complexity



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

ACATCGTGCAGGAACCCAGTGTGGACCGGCAGGCAGACCCGCACGACTGCTTACTGCTAATGGC
AGCTCCTGCCCACAGCTCAGCCCCCGACACAGGCCTTACCAGCTTGTCTCTGAAAG

CACATCGTGCAGGTGGATTATTTACCACTGAGCCACGTGGGAAGTCCTCTTACACAACAGGCACTC
AATAAACATCATCCCTTCAGCAGACATTTATCAAGCTTTTACTGGGAACCCAACCTGCAT

ACATGGTGCAGGAACCCAGTGTGGACCGGCAGGCAGACCCGCACGACTGCTTACTGCTAATGGC
AGCTCCTGCCCACAGCTCAGCCCCCGACACAGGCCTTACCAGCTTGTCTCTGAAAG

Allow some mismatch

Data analysis: SNP calling

Adapter removal, quality trimming

↓

Demultiplexing, Barcode splitting

↓

Reference genome available?

Yes

No

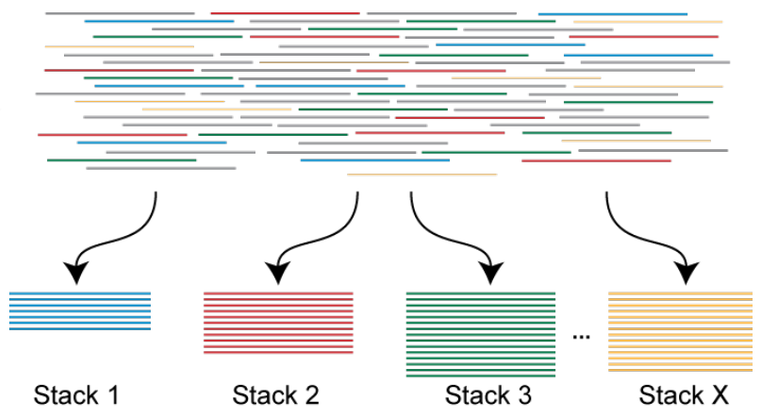
Align reads to reference
"Reference based"

Cluster/assemble reads
"DeNovo"

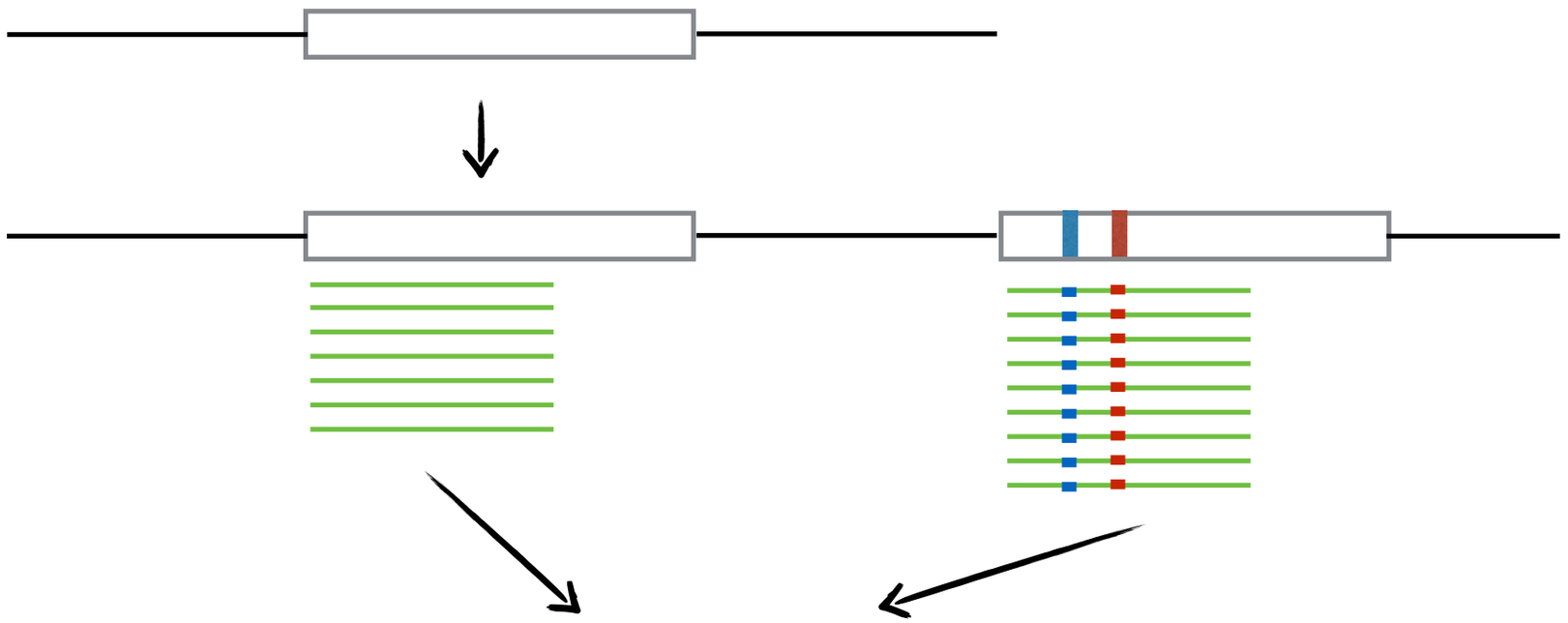
Align reads to
"DeNovo" assembly

↓

SNP calling

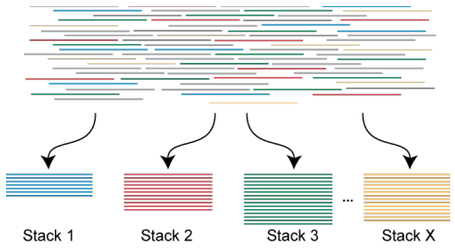


Avoid paralogs



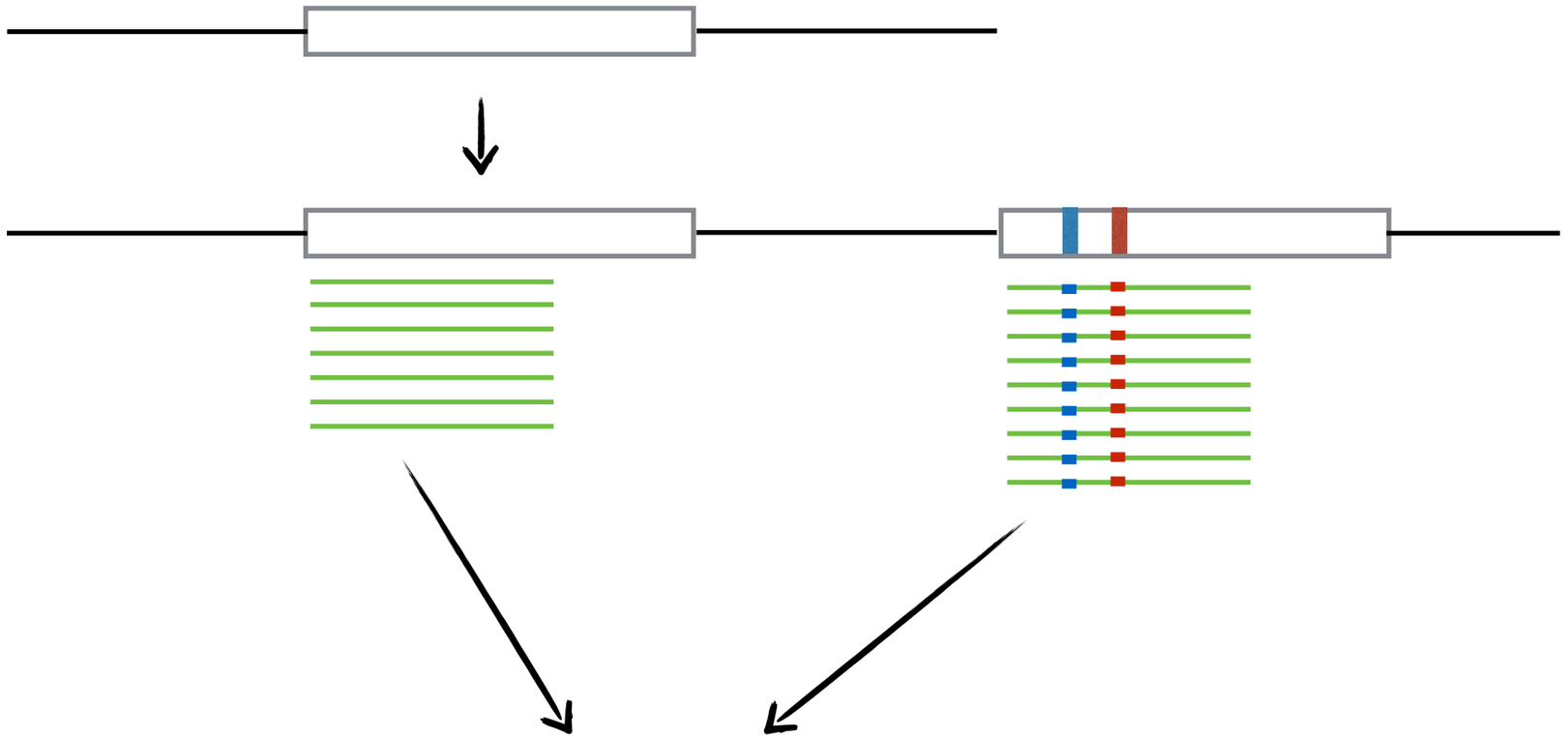
Cluster/assemble reads

“DeNovo”



2 heterozygous sites

Avoid paralogs



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

Data analysis: Filtering

Adapter removal, quality trimming

↓

Demultiplexing, Barcode splitting

↓

Reference genome available?

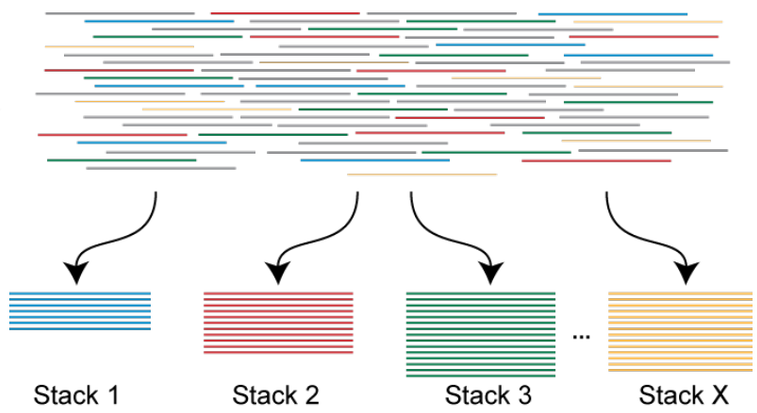
Yes

No

Align reads to reference
"Reference based"

Cluster/assemble reads
"DeNovo"

Align reads to
"DeNovo" assembly



↓

SNP calling

↓

Filtering

↓

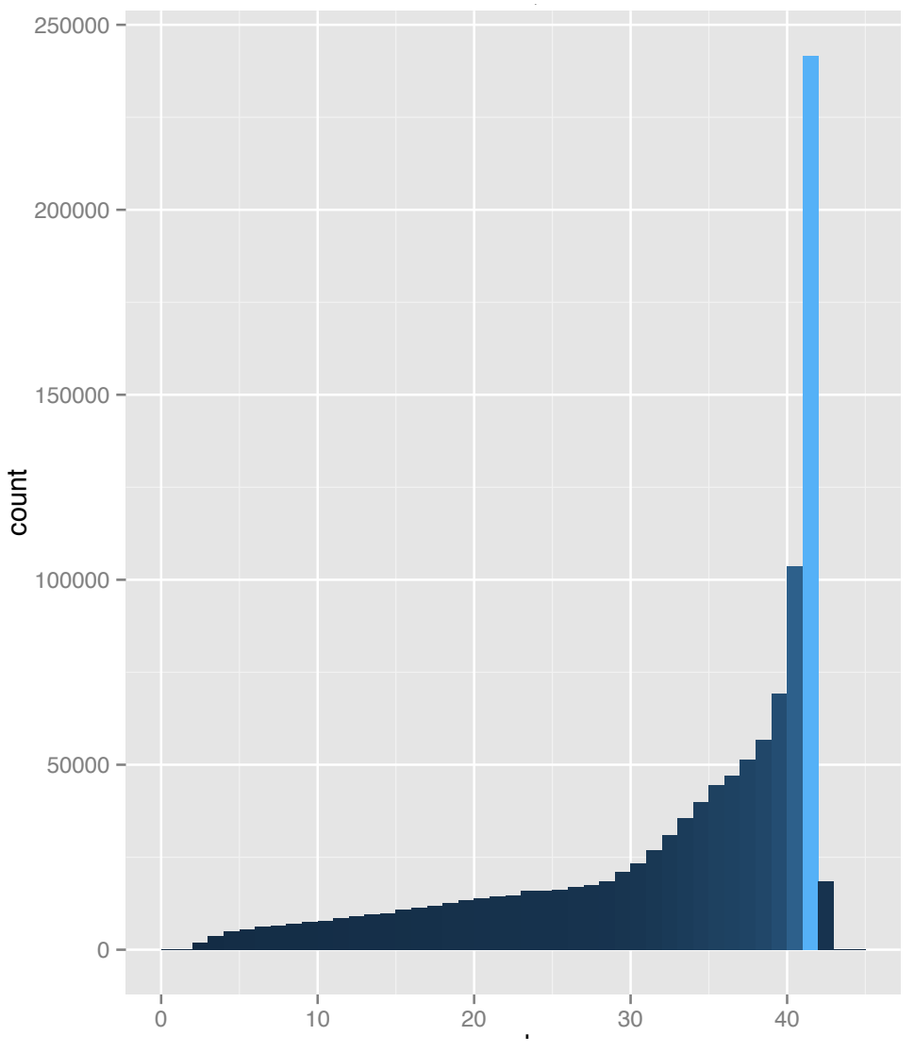
Population genetic analyses

Catchen et al. 2011

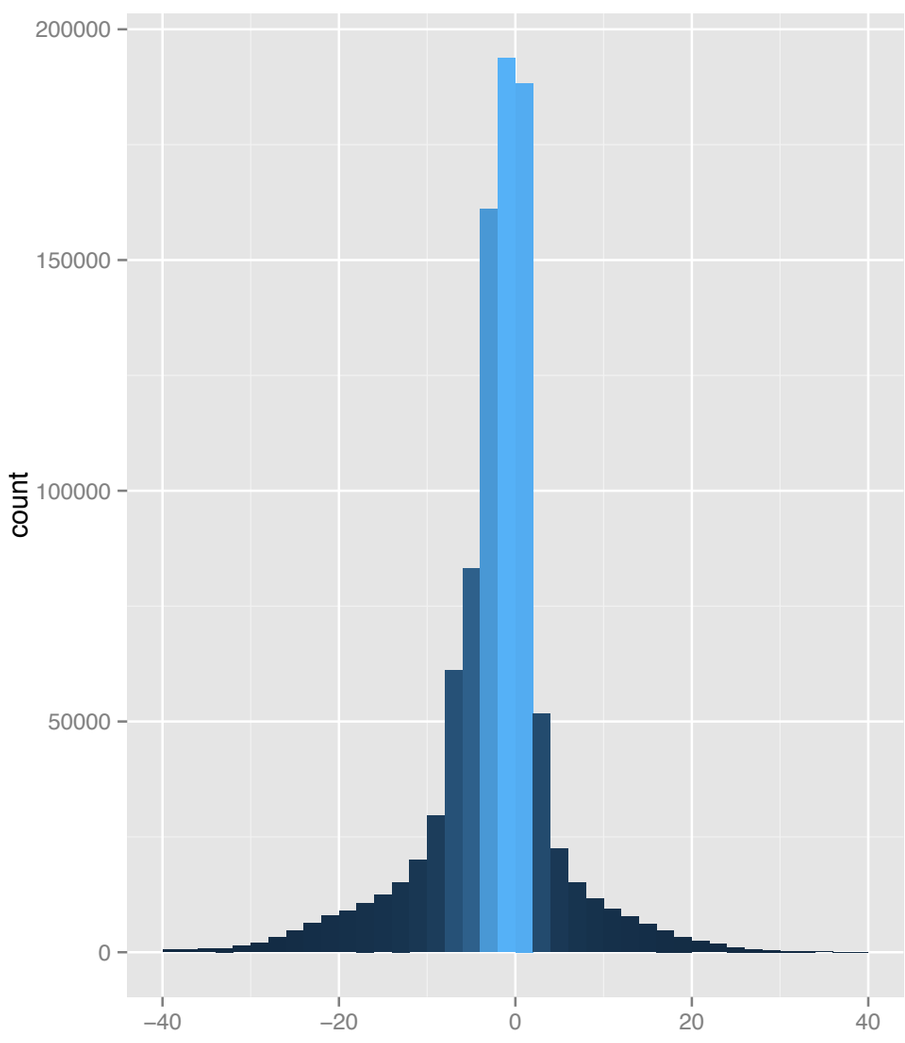
Data analysis: Filtering

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

Data analysis with custom scripts

Adapter removal, quality trimming

↓

Demultiplexing, Barcode splitting

↓

Reference genome available?

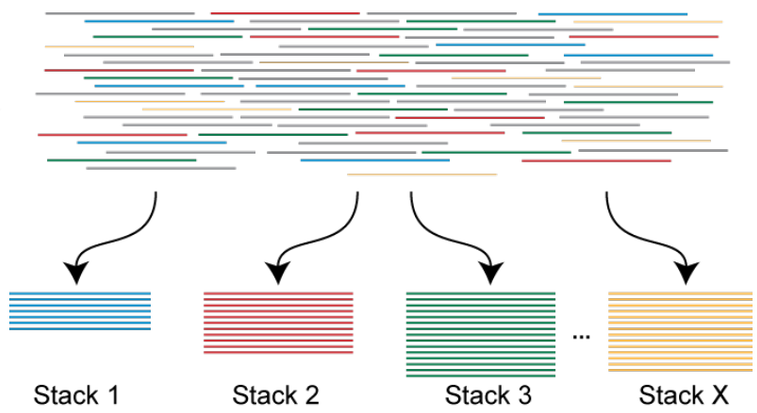
Yes

No

Align reads to reference
"Reference based"

Cluster/assemble reads
"DeNovo"

Align reads to
"DeNovo" assembly



↓

SNP calling

↓

Filtering

↓

Population genetic analyses

Questions?