

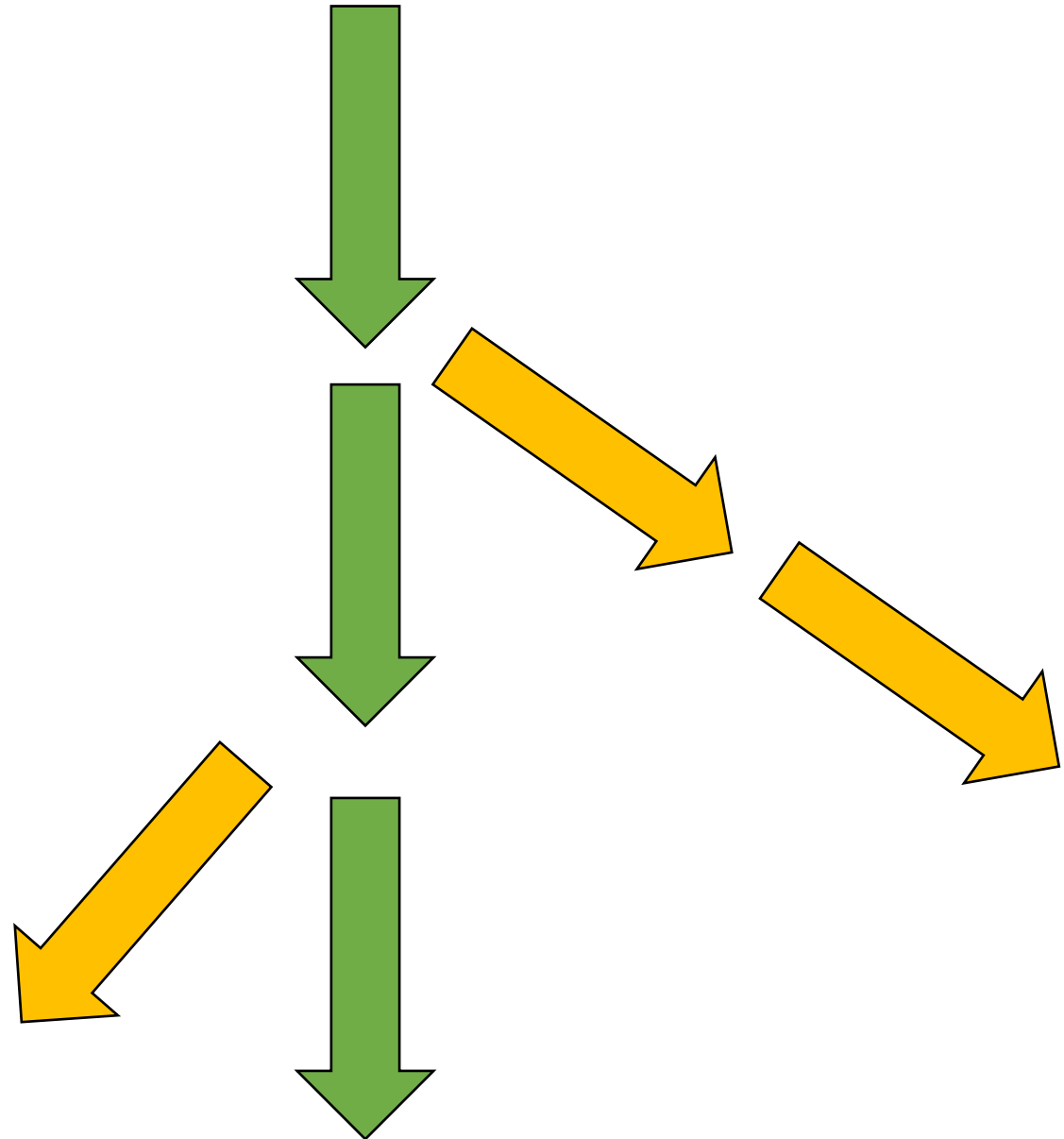
RAD Analysis

Niklaus Zemp
24 June 2020

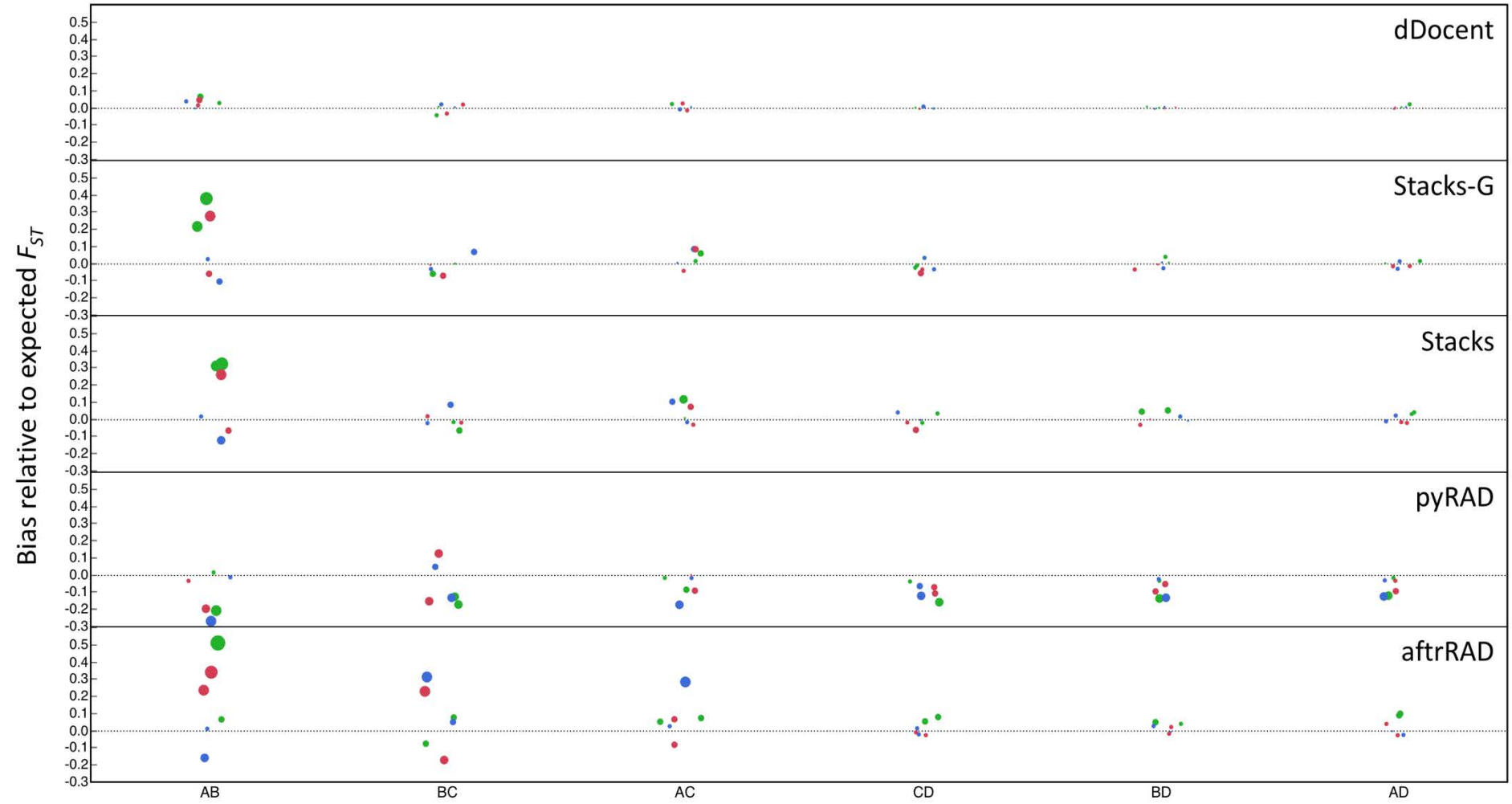
Genetic Diversity Centre (GDC)
Bioinformatics
ETH Zurich


Pipelines

- Stacks (Catchen et al. 2013)
- dDocent (Puritz et al. 2014)
- pyRAD (Eaton 2014)
- aftrRAD (Sovic et al. 2015)



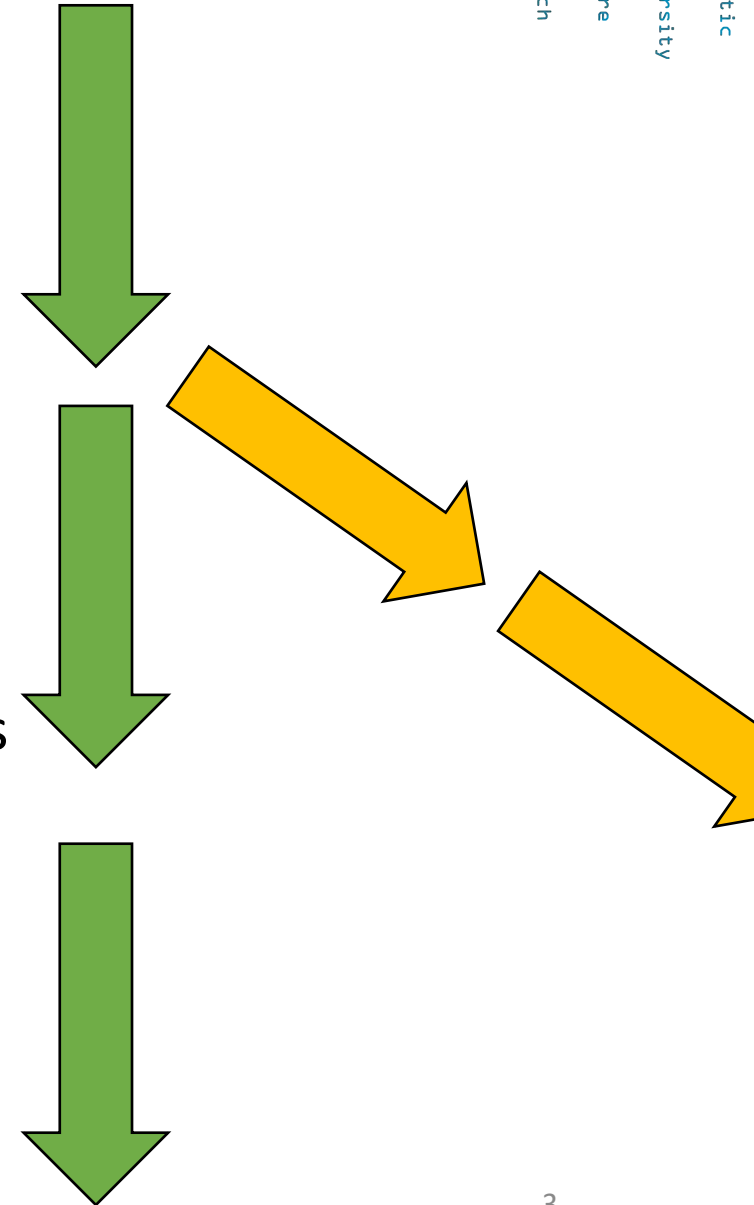
Population differentiation bias of different pipelines

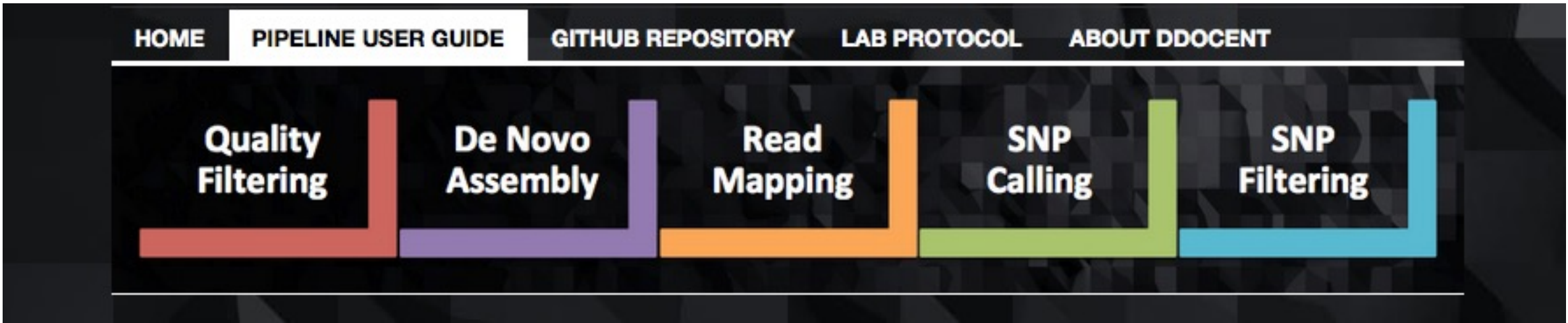




Pipelines

- Stacks (Catchen et al. 2013)
- dDocent (Puritz et al. 2014)
 - Can handle Indels
 - Simple customizable backbone for bioinformatics
- pyRAD (Eaton 2014)
 - Can handle many RADseq types, focused on phylogenetics





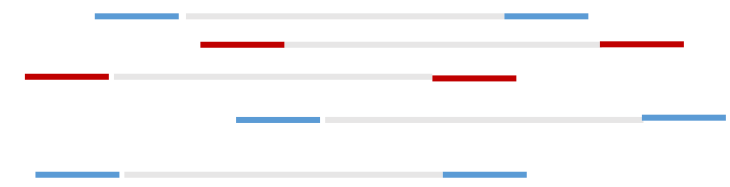
Tutorials: <https://github.com/jpuritz/dDocent>



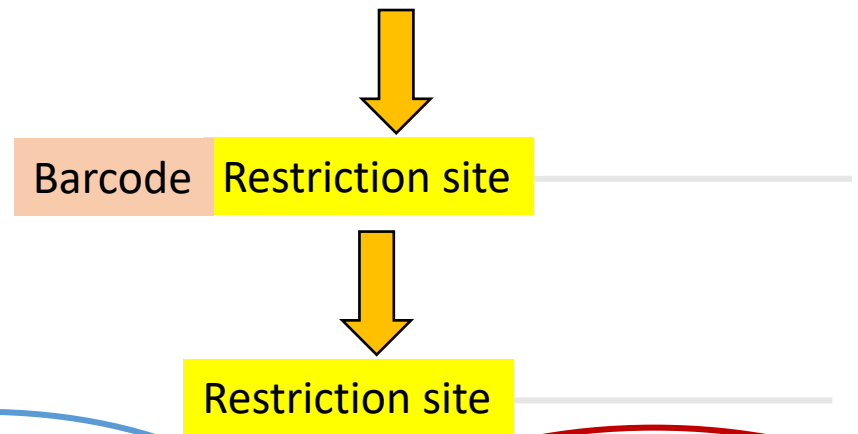
Demultiplexing and quality filtering



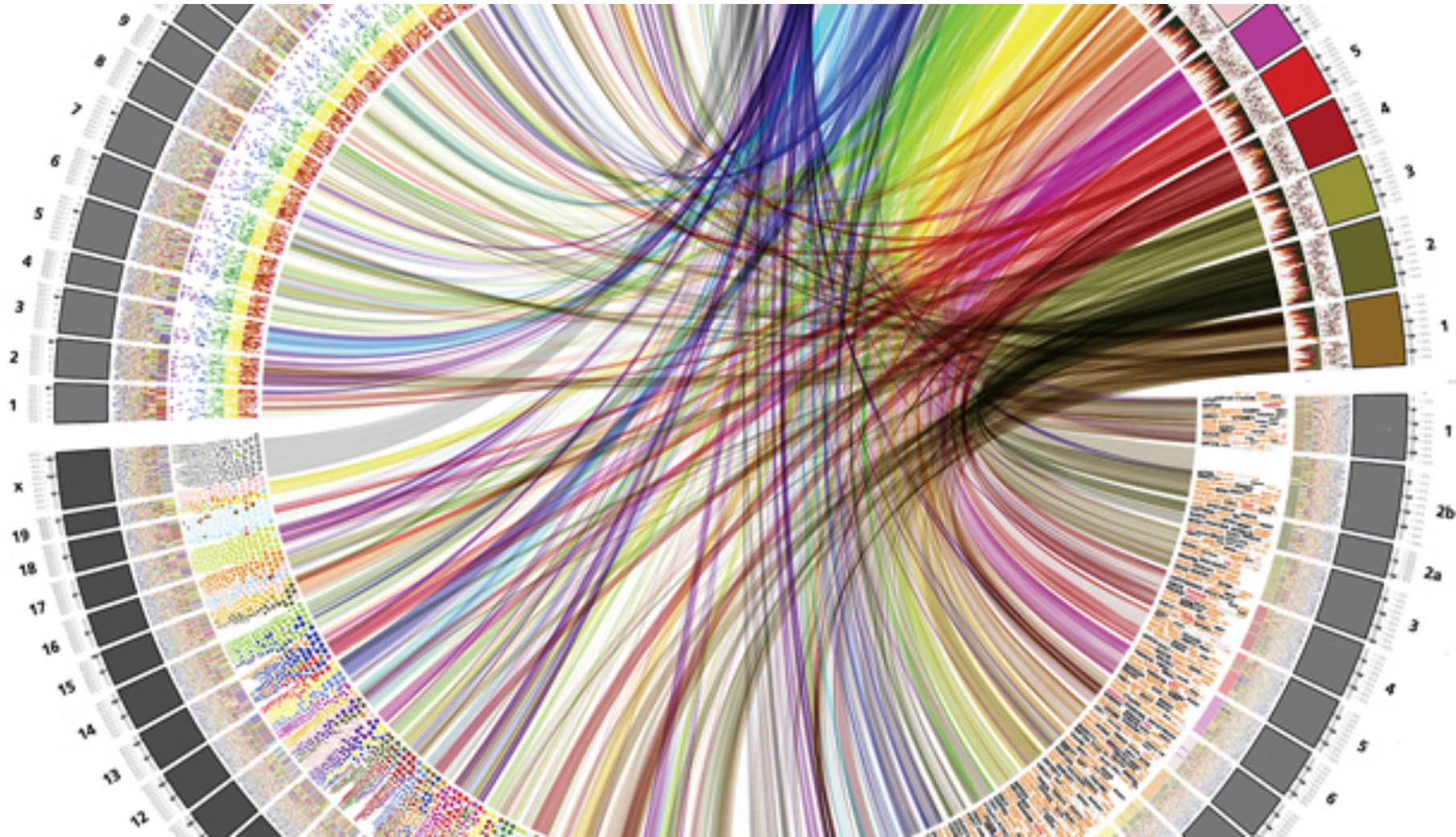
- Demultiplex reads
process_radtags (Stacks)



- Remove adaptors and low quality bases
Trimmomatic/fastp



Reference assembly





De novo assembly

Merge reads in case of overlaps

PEAR

Remove all identical reads

Pool all individuals together

customized scripts

Single-end:

Cluster the non-redundant sequences based on similarity

cd-hit-est

Paired-end:

Assembly the non-redundant sequences and then using paired-end information

rainbow, cd-hit-est



Locus 1

Locus 2

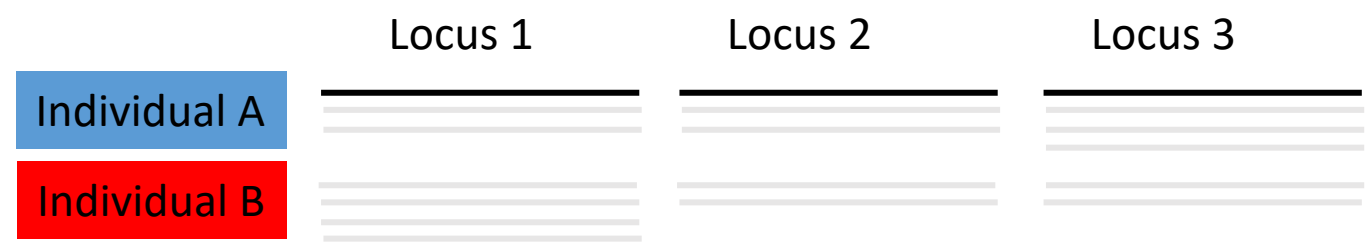
Locus 3




Read mapping

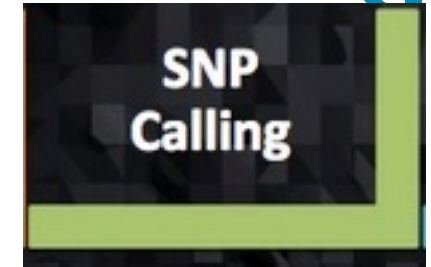
- Mapping reads against the reference catalogue

BWA





SNP calling



FreeBayes

Locus 1

Locus 2

Locus 3

Individual A

AATGC**A**GGG
 AATGC**A**GGG
 AATGC**A**GGG

AATGCT**T**GGGA
 AATGC**A**GGGA
 AATGCT**T**GGGA

AATGCT**TT**GGGA
 AATGCT**TA**GGGA
 AATGCT**TT**GGGA

Individual B

AATGCT**T**GGGA
 AATGCT**T**GGGA
 AATGCT**T**GGGA

AATGCT**T**GGGA
 AATGCT**T**GGGA
 AATGCT**T**GGGA

AATGCT**T** GGGA
 AATGCT**T** GGGA
 AATGCT**T** GGGA

SNP filtering



Filter only for good SNPs
VCFTools, vcfliib

Criteria:

Mapping quality

Coverage

Missing genotypes

Minor allele frequency

Balanced alleles



