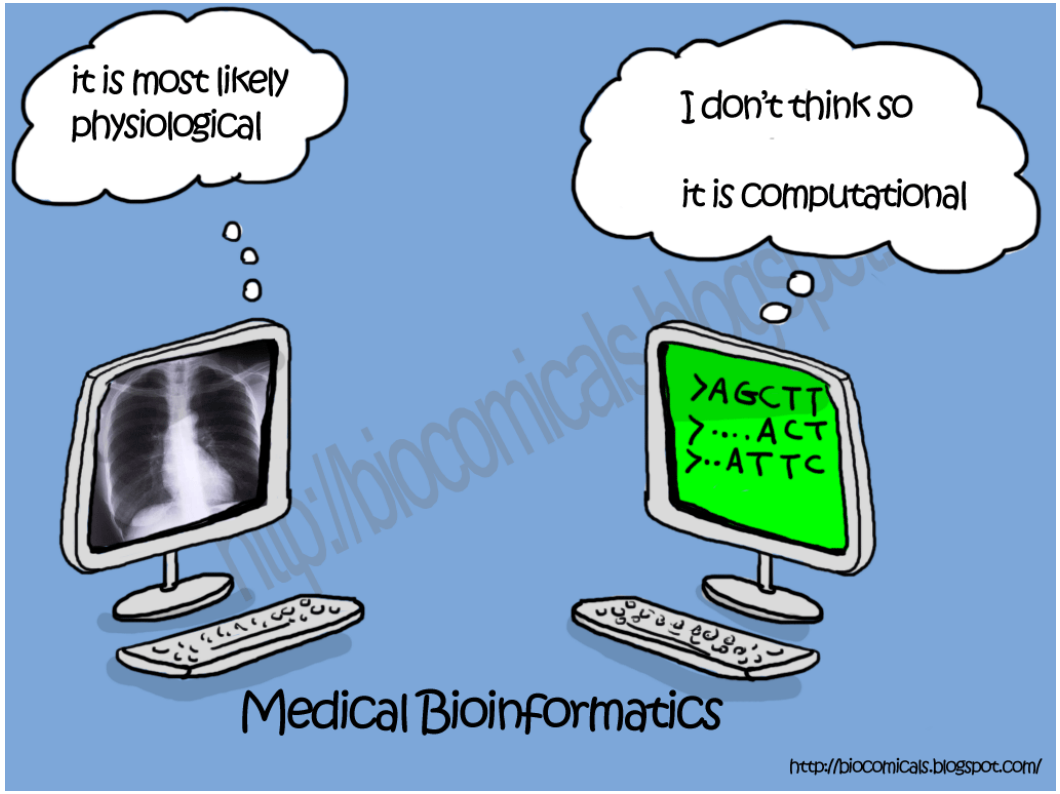
RAD/GBS approach
When biology means bio-informatics

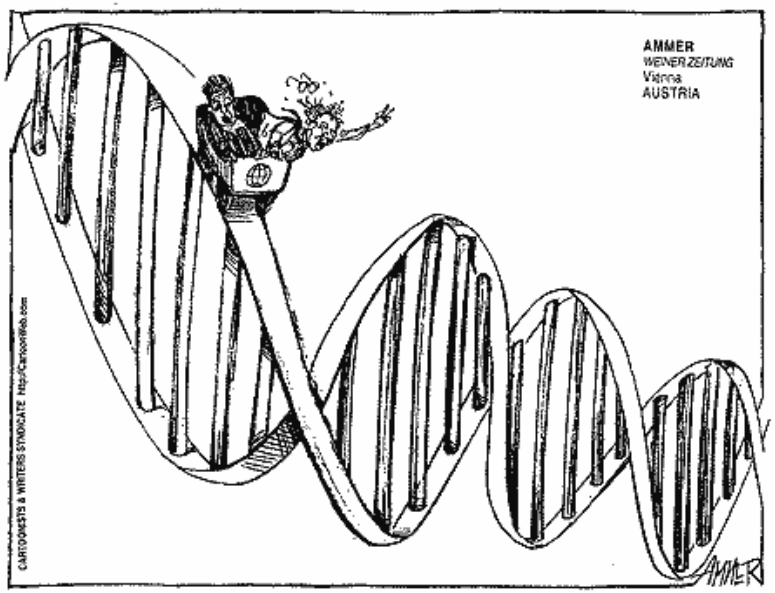


3 main steps

1. Library preparation and sequencing
2. **Identifying and genotyping genetic markers**
3. Analysing genetic structure

Library preparation

- Using digestion enzyme to sequence a genome subset
- Two methods :
  - Restricted Site Associated (RAD)
  - Genotyping By Sequencing (GBS)



RAD

Enzyme digestion

Barcoding

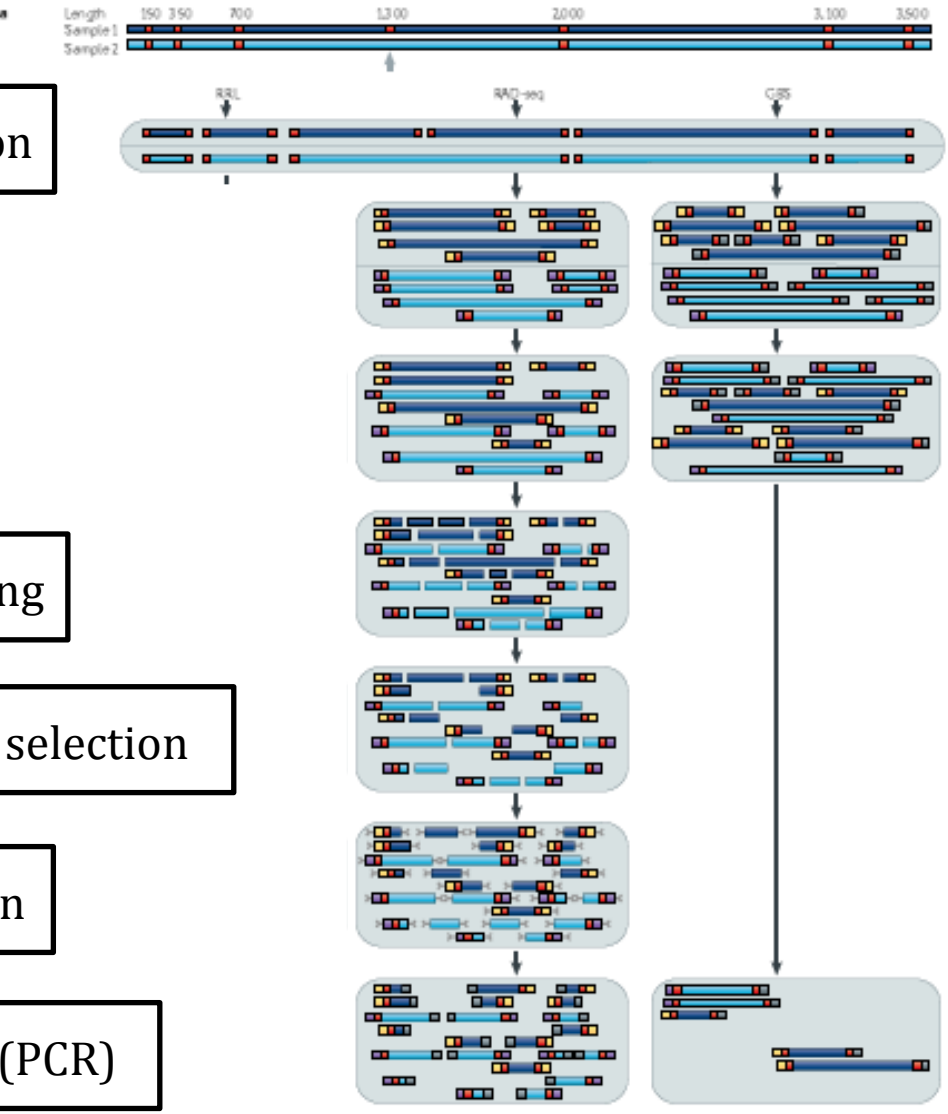
Pooling

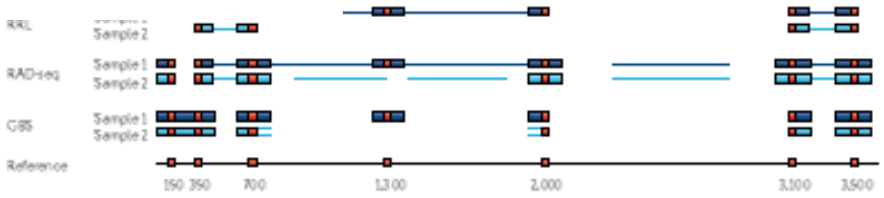
Random shearing

Fragment size selection

Adapter ligation

Amplification (PCR)





GBS

Enzyme digestion

Barcoding

Pooling

Amplification (PCR)

Identifying and genotyping genetic markers

1. STACKS (Catchen et al 2013)

-> https://groups.google.com/forum/#!forum/stacks-users

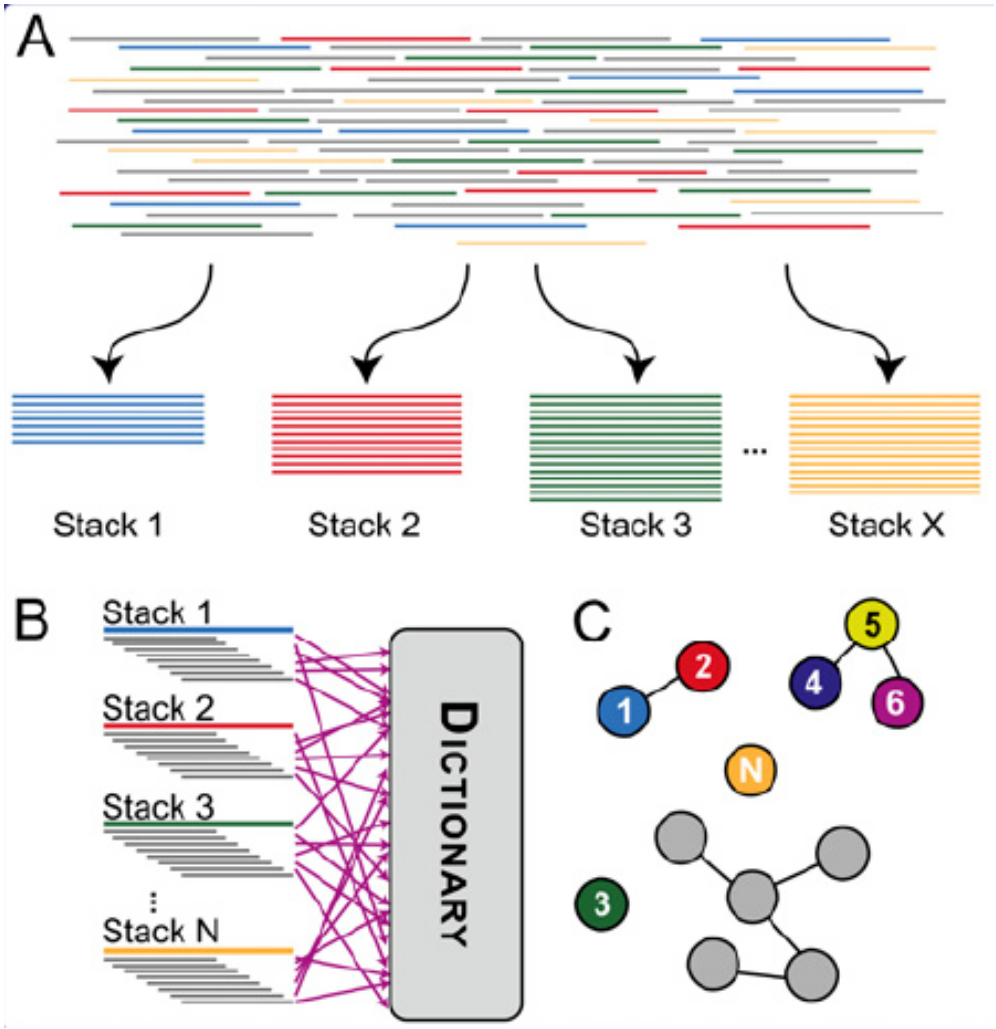
2. PyRAD (Eaton et al. 2014)
3. GBSx (Herten et al 2015)
4. dDocent (Putitz et al 2014)
5. AftrRAD (Sovic et al 2015)

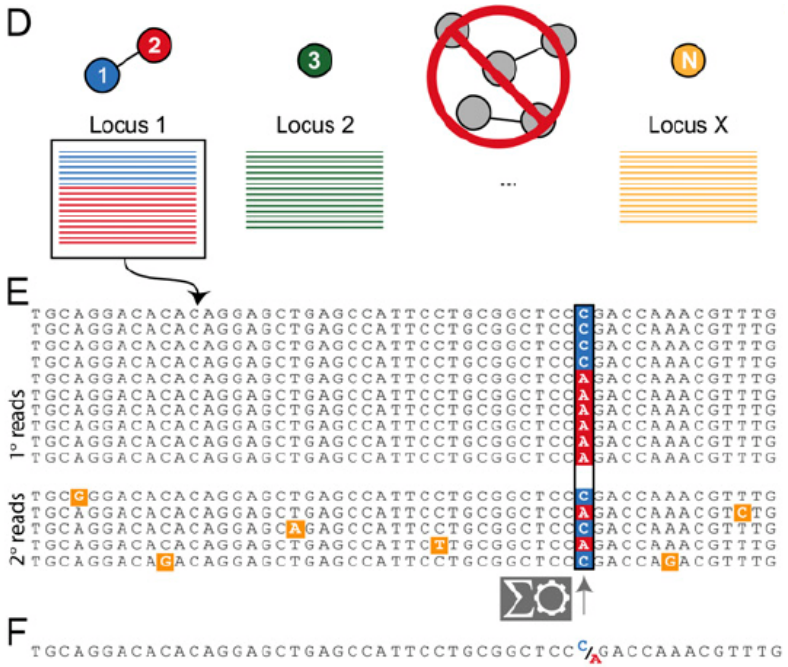
Following STACKs tutorial

https://github.com/enormandeau/stacks_workflow

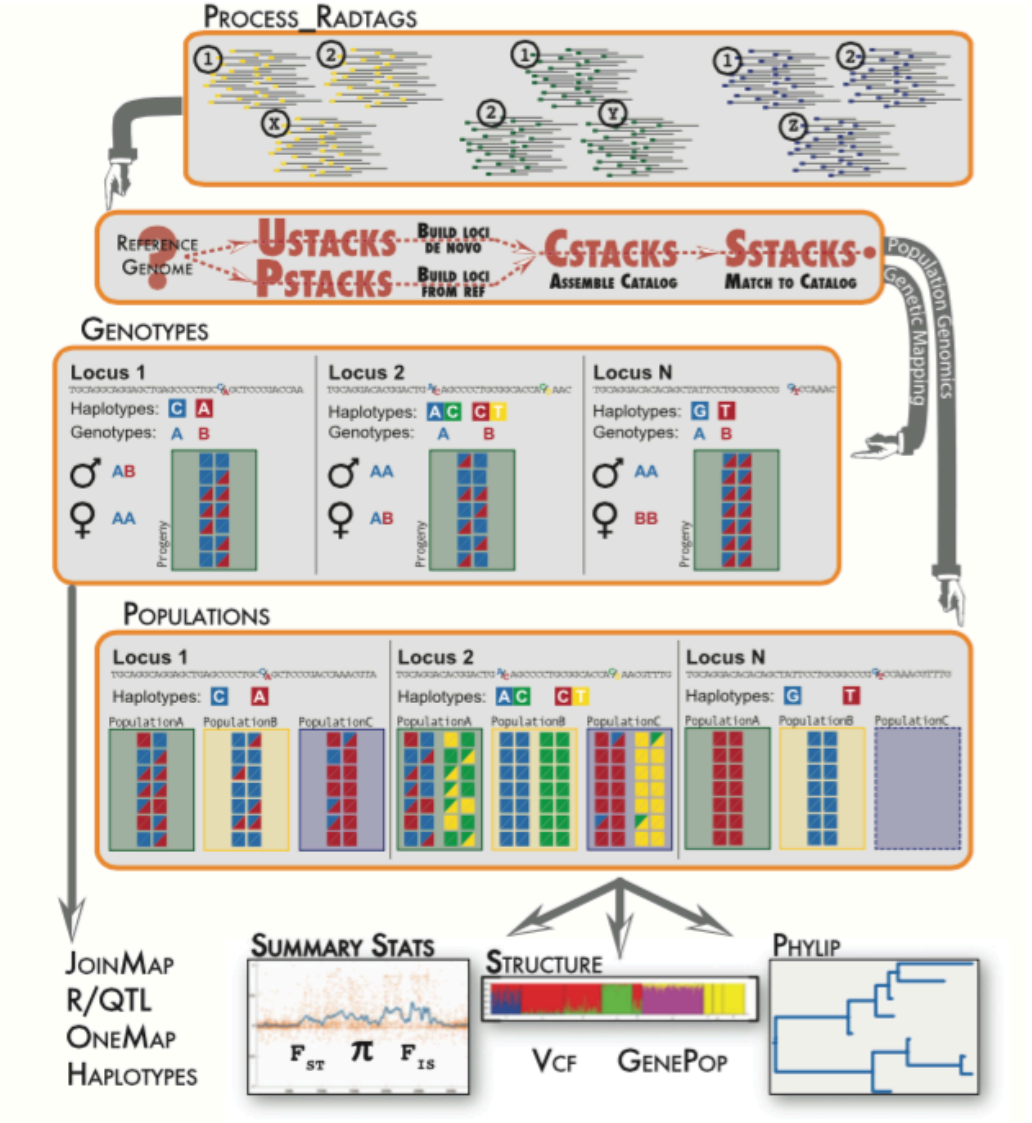
1. Stacks workflow tutorial
2. Install Stacks_workflow
3. Download your raw datafiles (Illumina or prepared IonProton lanes)
4. Extract individual data with `process_radtags`
5. Rename samples
6. Align reads to a reference genome (optional)
7. STACKS pipeline
8. Filtering the results
9. Conclusion

STACKs software



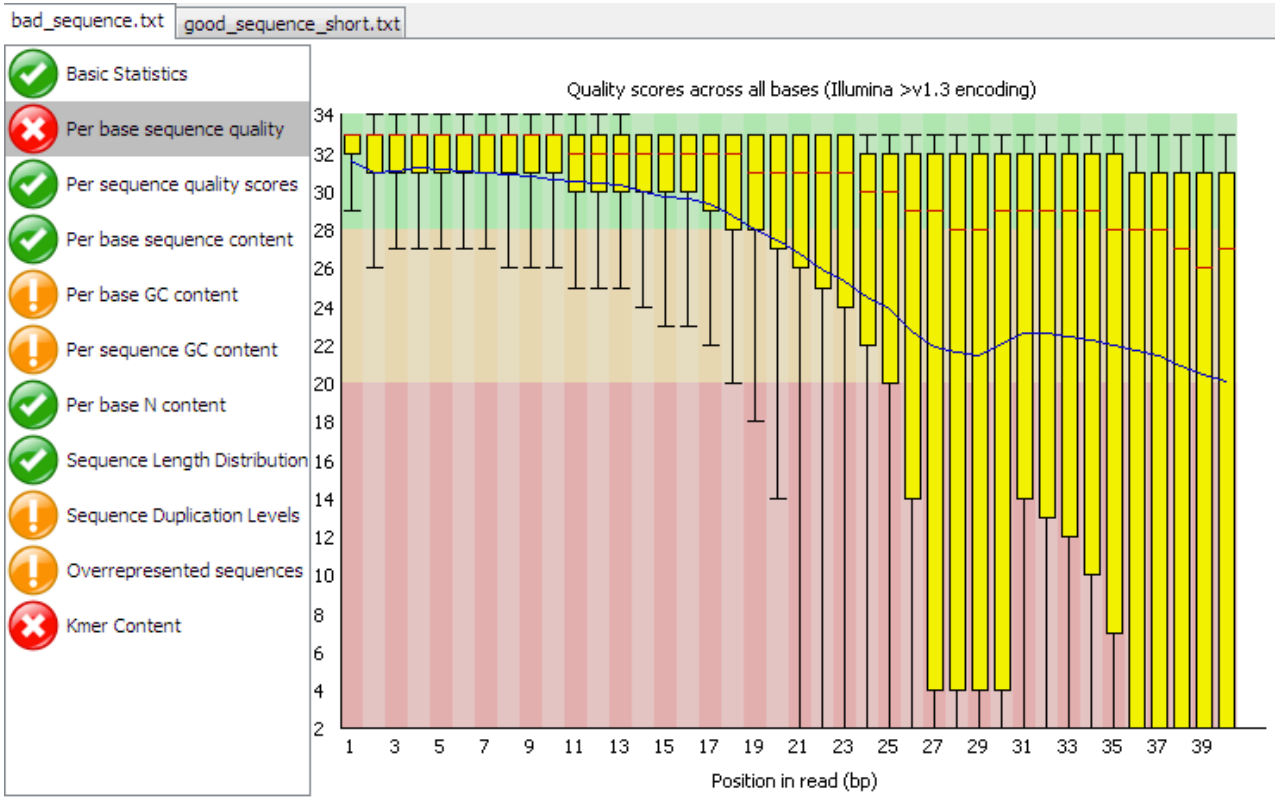


STACKs parameters



FASTQC

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/



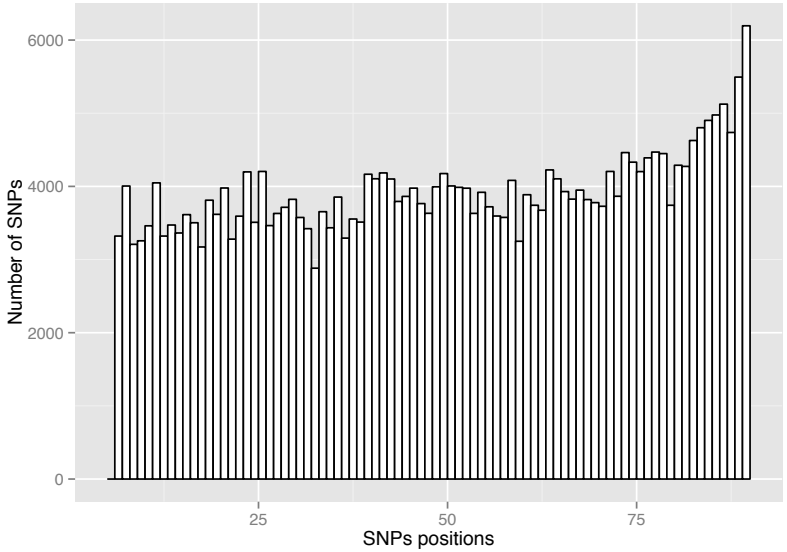
Cut the adaptaters with CUTADAPT

https://code.google.com/p/cutadapt/

1. Process_Radtags

-> cut-off number of bp

-> Enzyme used

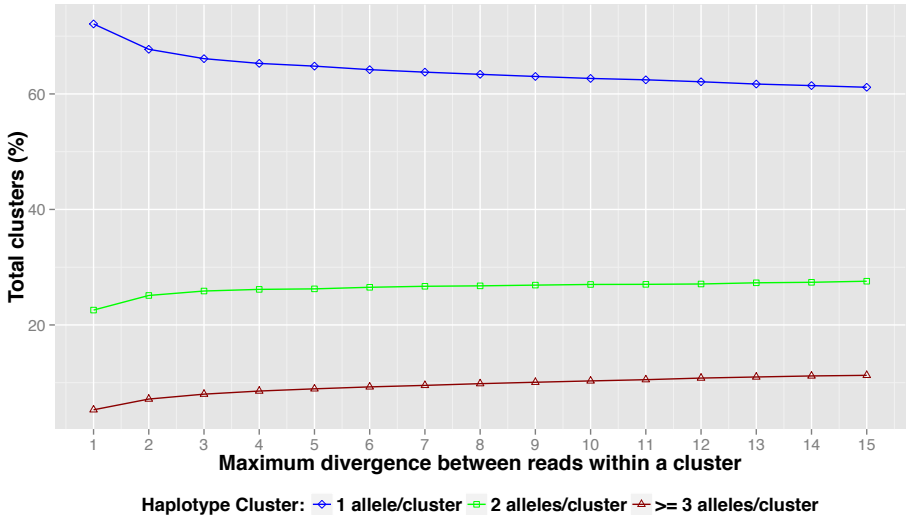


2. USTACKS

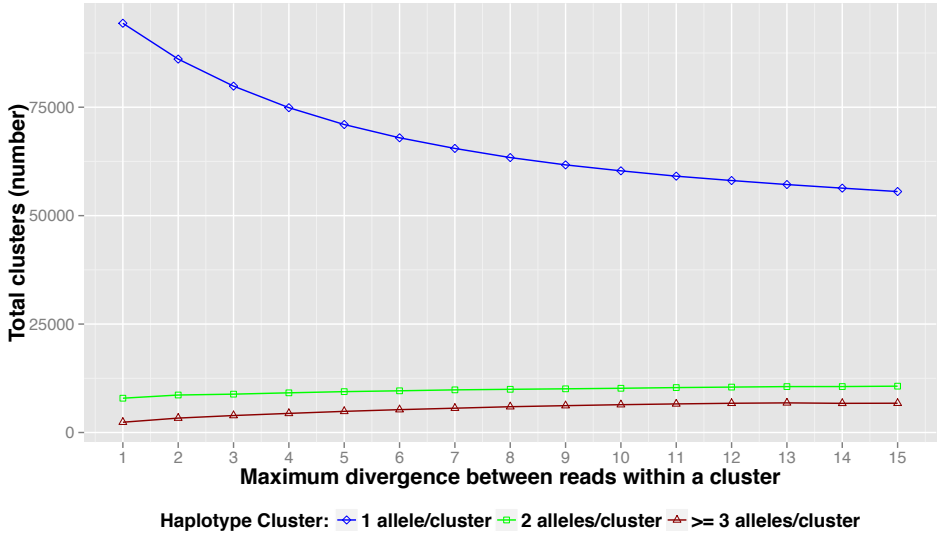
-> m = minimum depth of coverage

-> M= Maximum de distance allowed between stacks (Ilut et al. 2014)

Thinking about your species



Lobster



Lake trout

3. CSTACKS (default setting)

4. SSTACKS (default setting)

5. POPULATION

-> m : minimum depth required for individuals

-> r : minimum percentage of individuals in a population

"The ideal molecular approach for population genomics should uncover hundreds of polymorphic markers that cover the entire genome in a single, simple and reliable experiment. Unfortunately, at present there is no such approach."

Luikart et al, 2003

Entering in the genomics world

1. VCFTOOLS

http://vcftools.sourceforge.net

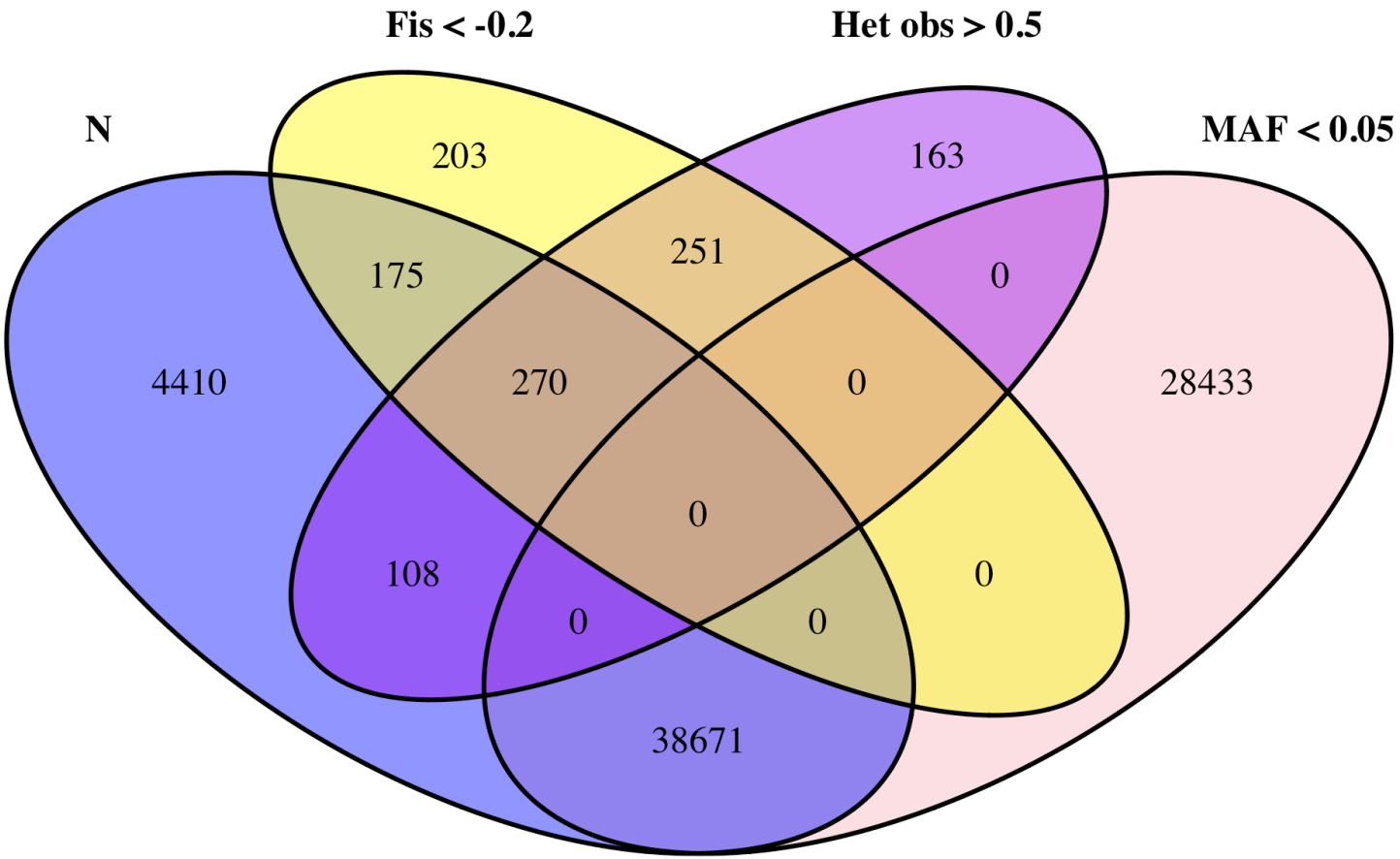
Go to the **full manual** page

2. PLINK

http://pngu.mgh.harvard.edu/~purcell/plink/

3. R software

Exploring data



From STACKS to SNPs

FROM READS TO SNPS	**SNP count**
STACKS CATALOG	200,293
POPULATION FILTERS	
Genotyped	
> 70% of the samples	74,229
> 70% of the populations	
MAF FILTERS	
Global MAF > 0.05	15,552
Local MAF > 0.1	
COVERAGE FILTER	
From 10 to 100x	15,505
HWE FILTERS	
F_{IS} > -0.3	10,324
Hardy-Weinberg equilibrium	
(P-value 0.05)	
H_{OBS} < 0.5	10,156
GENOME SCAN FILTER	
Putatively neutral	**8144**
Putatively under divergent selection	32

Benestan et al. 2015 (accepted)

PGDspider is your best friend

PGDSpider version 2.0.8.1 (April 2015)

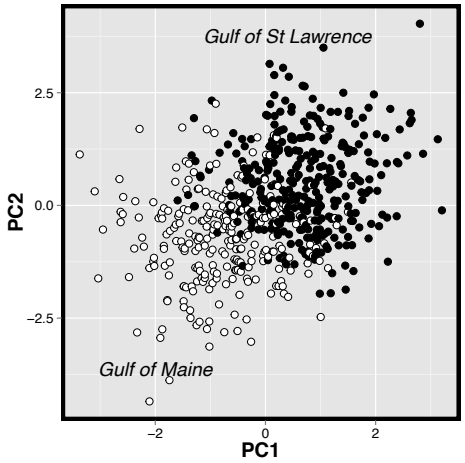
http://www.cmpg.unibe.ch/software/PGDSpider/

Finding the number of clusters

1. DPCA (library adegenet)
http://adegenet.r-forge.r-project.org

2. FASTSTRUCTURE
http://rajanil.github.io/fastStructure/

3. ADMIXTURE
https://www.genetics.ucla.edu/software/admixture/



F-statistics

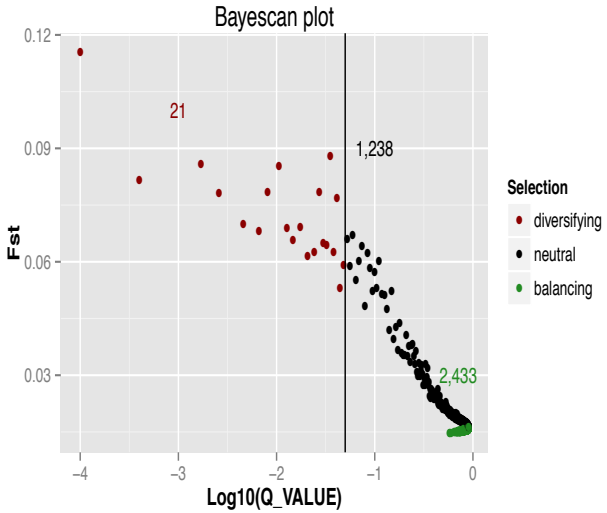
1. GENODIVE
http://www.bentleydrummer.nl/software/software/GenoDive.html

2. ARLEQUIN
http://cmpg.unibe.ch/software/arlequin35/

Genome scan

1. BAYESCAN

http://cmpg.unibe.ch/software/BayeScan/



2. ARLEQUIN

http://cmpg.unibe.ch/software/arlequin35/

Learning is an ongoing process

Thanks for your attention