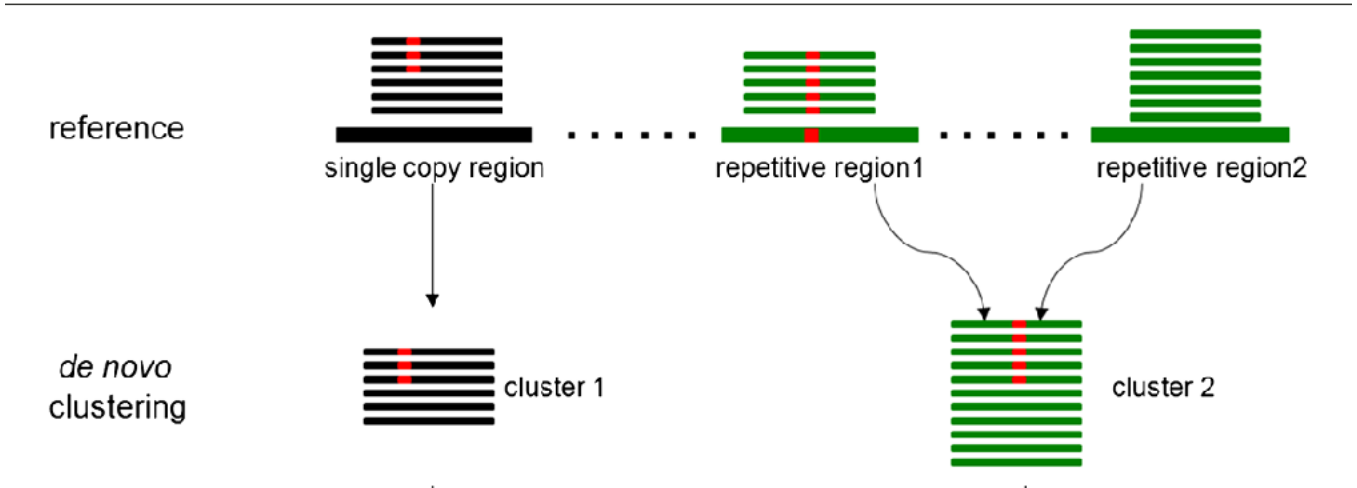
SNP Calling

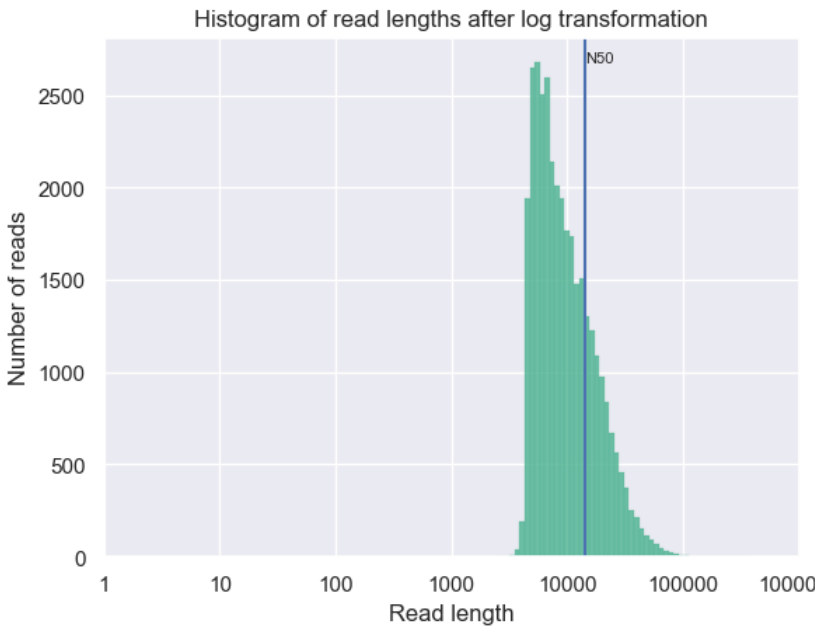
Population genetic studies

- Many options and different requirements for SNP calling
- RAD-Seq
  - Stacks, both reference guided or *de novo*
- Hyb-Seq, RNA-Seq or Whole Genome Sequencing
  - GATK, Freebayes, mpileup
  - Must have a reference genome or at the very least something to map reads to



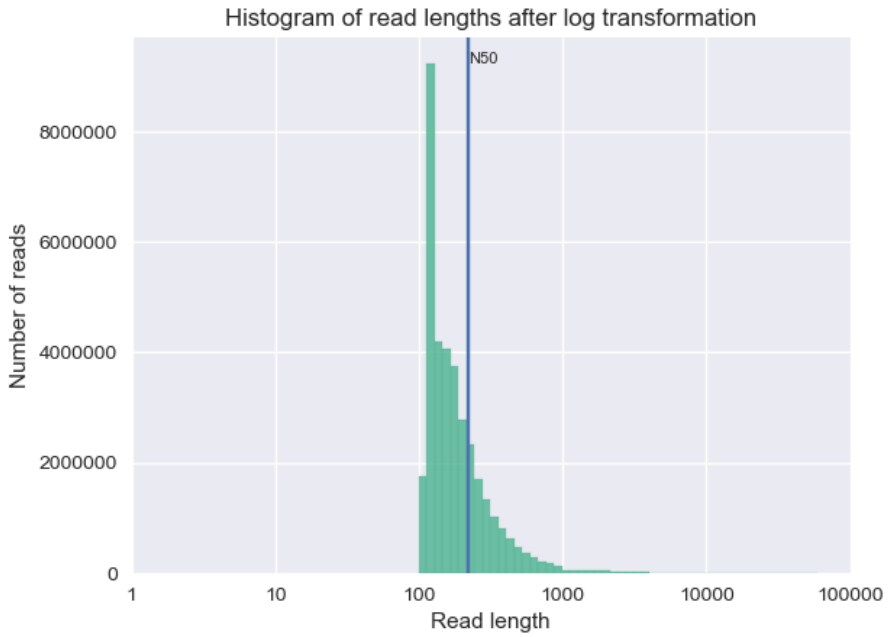
Having a draft genome helps with SNPs

Nanopore only



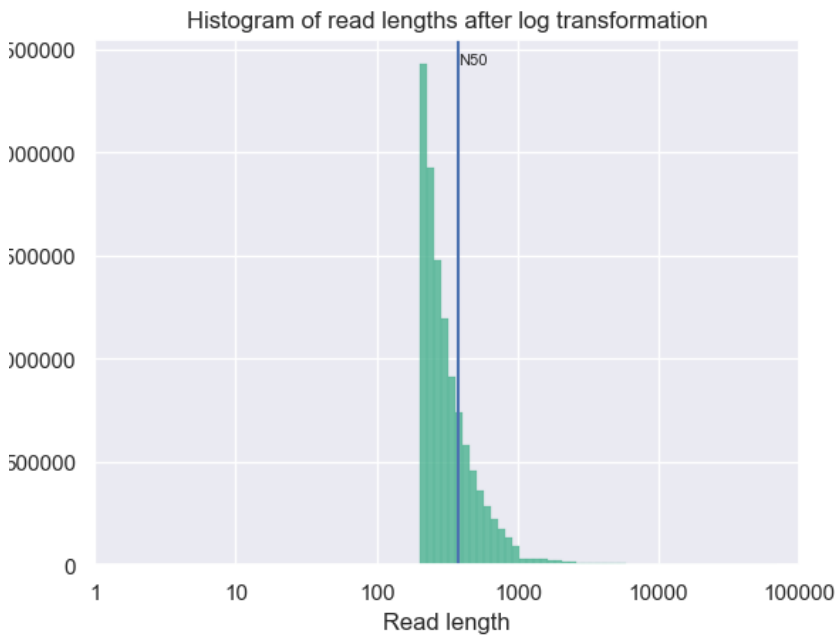
Mean read length:	11,771.3
Median read length:	8,595.0
Number of reads:	33,757.0
Read length N50:	14,352.0
Total bases:	397,362,391.0

Illumina only



Mean read length:	218.2
Median read length:	159.0
Number of reads:	35,527,267.0
Read length N50:	220.0
Total bases:	3,051,257,368.0

Hybrid Assembly



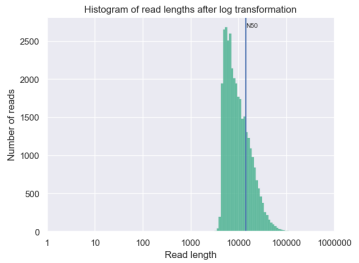
Mean read length:	383.3
Median read length:	278.0
Number of reads:	11,239,396.0
Read length N50:	377.0
Total bases:	4,307,778,281.0

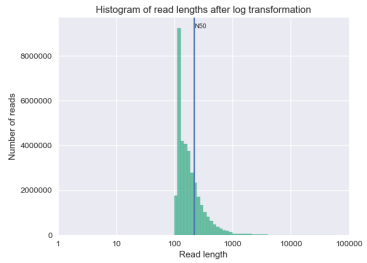
Preliminary RAD-Seq

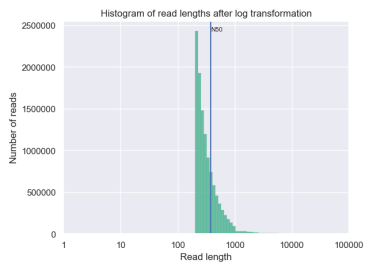
- RAD-seq data of 25 test individuals
- Used Stacks to call SNPs, after using BWA MEM for mapping to reference
- Non-stringent filtering, max missing 70% loci

SNPs

	de novo	Nanopore only	Illumina	Hybrid
Raw	5,903	50,723	203,143	258,958
Filtered	2,188	4,976	8,660	15,533







Steps in module

- Map transcriptome reads to reference genome
  - BWA MEM
- Call SNPs with Stacks
  - Perl wrapper ref_map.pl
- Filter with VCFtools
  - Manual filtering to remove poor individuals and poor loci
- PCA using filtered SNP data
  - SNPRelate package in R