

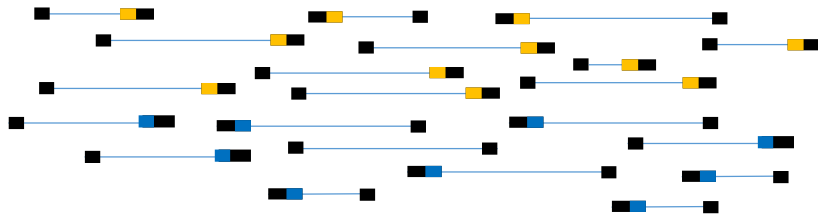
Whole-genome short-read sequencing

DNA

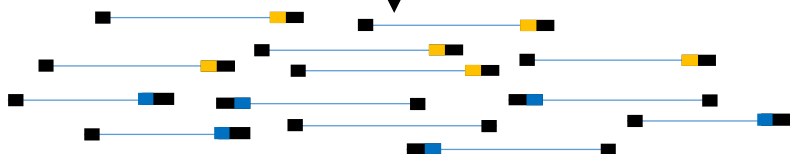
↓ Random shearing



↓ Illumina adapter ligation
incl. individual index

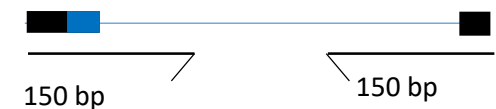


↓ Size selection



paired-end sequencing

Up to 20 billion read pairs

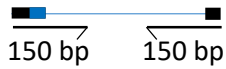


1. Quality check and trimming raw reads

paired-end Illumina sequencing



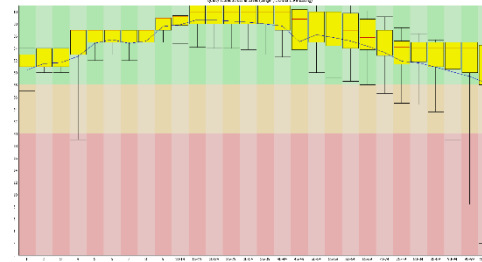
Up to 20 billion
read pairs



Forward
reads

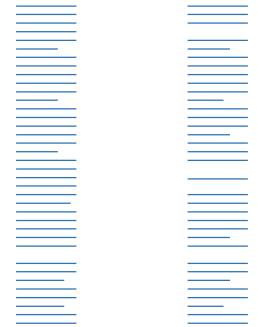
Reverse
reads

Fastqc quality plots



fastp

Trimmed and filtered
reads of good quality



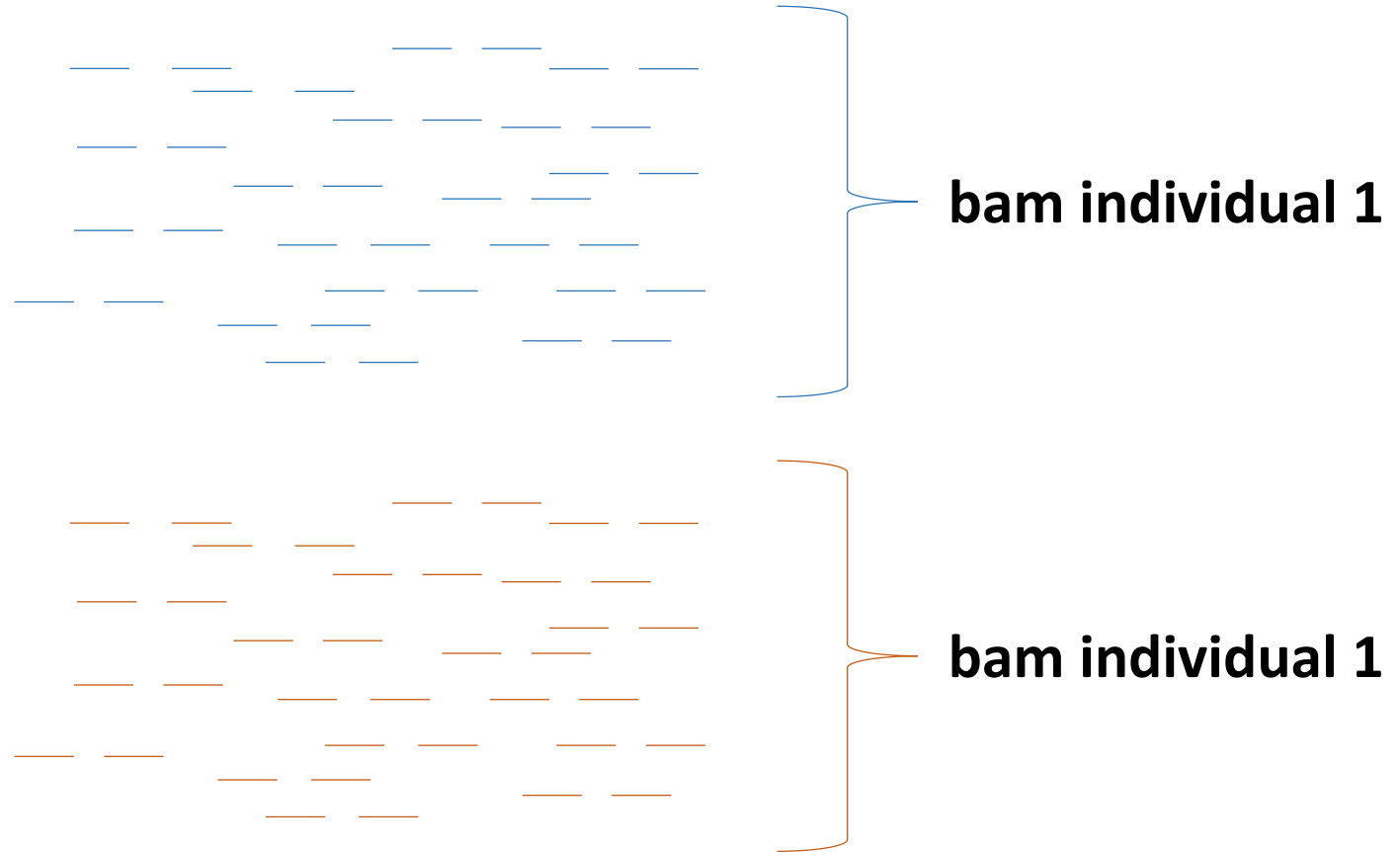
Fastq file of reads

```
@HWUSI-EAS611:34:6669YAAXX:1:1:5069:1159 1:N:0:
TCGATAATACCGTTTTTTTCCGTTTGATGTTGATACCAT
+
IIHHIIHHIIIIIIIIIIIIIIIIIIIIIIIIIIHHIIIIHHIIII
```

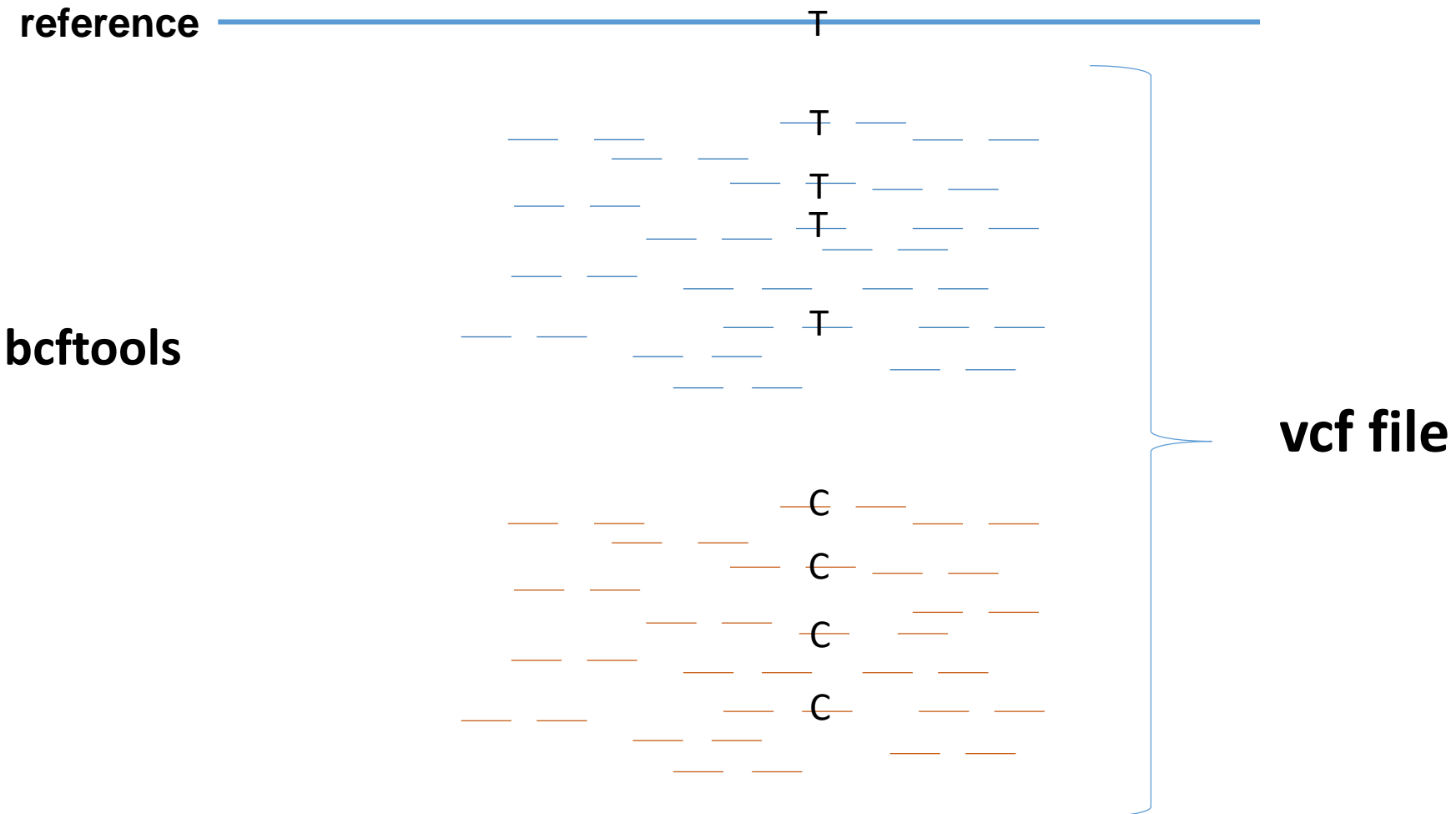
2. Alignment to the reference genome with bwa

reference

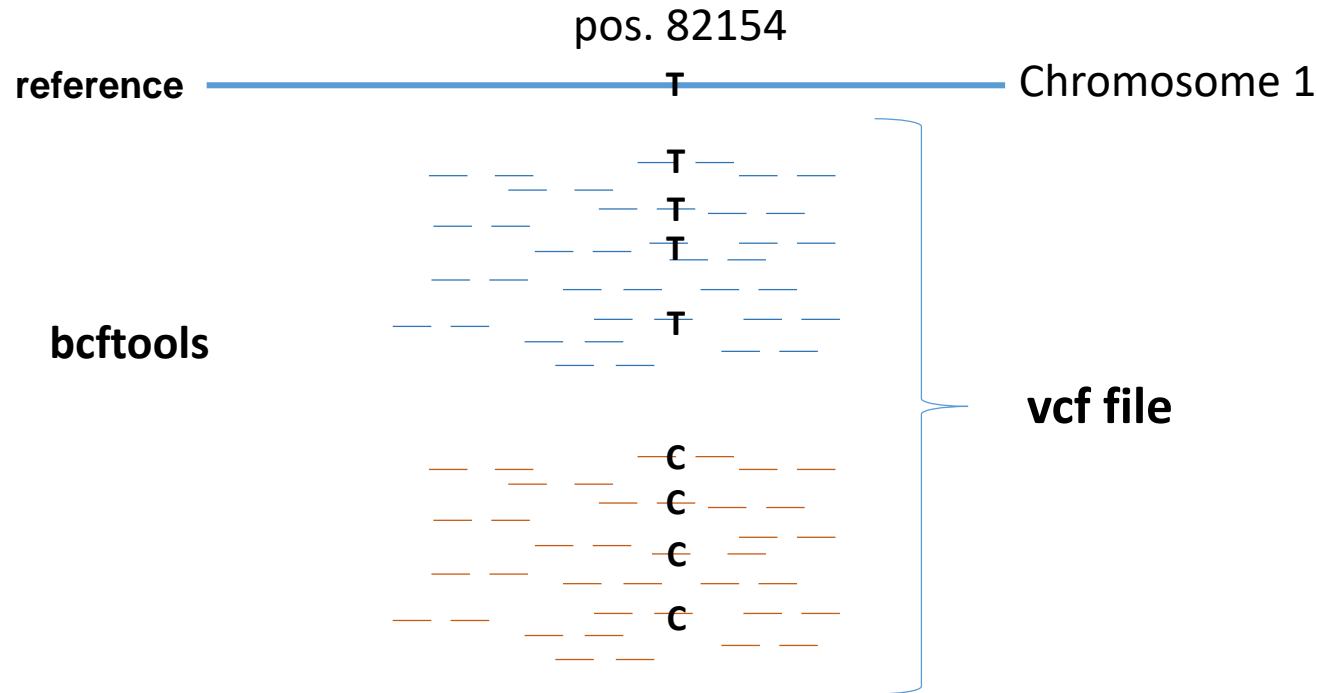
bwa



3. Variant and genotype calling with bcftools



3. Variant and genotype calling with bcftools



vcf file: Genotypes for each individual at genomic sites

```
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype"  
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT GEN  
chr1 82154 . T C . GT 0/0 1/1 0/0 1/1 0/0 1/1  
chr1 752566 . T . . GT 0/0 0/0 0/0 0/0 0/0 0/0  
chr1 752721 . T C . GT 1/1 1/1 1/1 1/1 1/1 1/1  
chr1 752721 . A . . GT ./.
```

