(base) MacBook-Airpan4da:raw\_data pan4da$ wc -l amp\_res\_2.fastq

1823504 amp\_res\_2.fastq (the same number in file\_1, so we have 455876 reads)

fastqc -o . /Users/pan4da/project2/raw\_data/amp\_res\_1.fastq /Users/pan4da/project2/raw\_data/amp\_res\_2.fastq

conda create --env trimmomatic

conda install trimmomatic -c bioconda

(command for installation and then you have to activate it)

trimmomatic PE -phred33 /Users/pan4da/project2/raw\_data/amp\_res\_1.fastq /Users/pan4da/project2/raw\_data/amp\_res\_2.fastq /Users/pan4da/project2/raw\_data/pairedPE\_1.fq /Users/pan4da/project2/raw\_data/singlePE\_1.fq /Users/pan4da/project2/raw\_data/pairedPE\_2.fq /Users/pan4da/project2/raw\_data/singlePE\_2.fq LEADING:20 TRAILING:20 MINLEN:20 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10

(base) MacBook-Airpan4da:raw\_data pan4da$ wc -l pairedPE\_3.fq

15400 pairedPE\_3.fq

(base) MacBook-Airpan4da:raw\_data pan4da$ wc -l pairedPE\_4.fq

15541 pairedPE\_4.fq

bwa index amp\_res\_1.fastq it’s working, create 5 new files, but we need to use reference file!

7703 + 0 in total (QC-passed reads + QC-failed reads)

7702 + 0 primary

0 + 0 secondary

1 + 0 supplementary

0 + 0 duplicates

0 + 0 primary duplicates

7699 + 0 mapped (99.95% : N/A)

7698 + 0 primary mapped (99.95% : N/A)

7702 + 0 paired in sequencing

3851 + 0 read1

3851 + 0 read2

7682 + 0 properly paired (99.74% : N/A)

7694 + 0 with itself and mate mapped

4 + 0 singletons (0.05% : N/A)

0 + 0 with mate mapped to a different chr

0 + 0 with mate mapped to a different chr (mapQ>=5)

Only SNPs will be reported

Warning: No p-value threshold provided, so p-values will not be calculated

Min coverage: 8

Min reads2: 2

Min var freq: 0.2

Min avg qual: 15

P-value thresh: 0.01

Reading input from /Users/pan4da/project2/result/my.mpileup

4641524 bases in pileup file

9 variant positions (6 SNP, 3 indel)

0 were failed by the strand-filter

6 variant positions reported (6 SNP, 0 indel)

snpEff

wget https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF\_000005845.2\_ASM584v2/GCF\_000005845.2\_ASM584v2\_genomic.gbff.gz

snpeff.config (after making file add in it k12.genome : ecoli\_K12)

mkdir -p data/k12

conda activate snpEff

gunzip GCF\_000005845.2\_ASM584v2\_genomic.gbff.gz

cp GCF\_000005845.2\_ASM584v2\_genomic.gbff data/k12/genes.gbk

snpEff build -genbank -v k12

snpEff ann k12 ../VarScan\_results.vcf