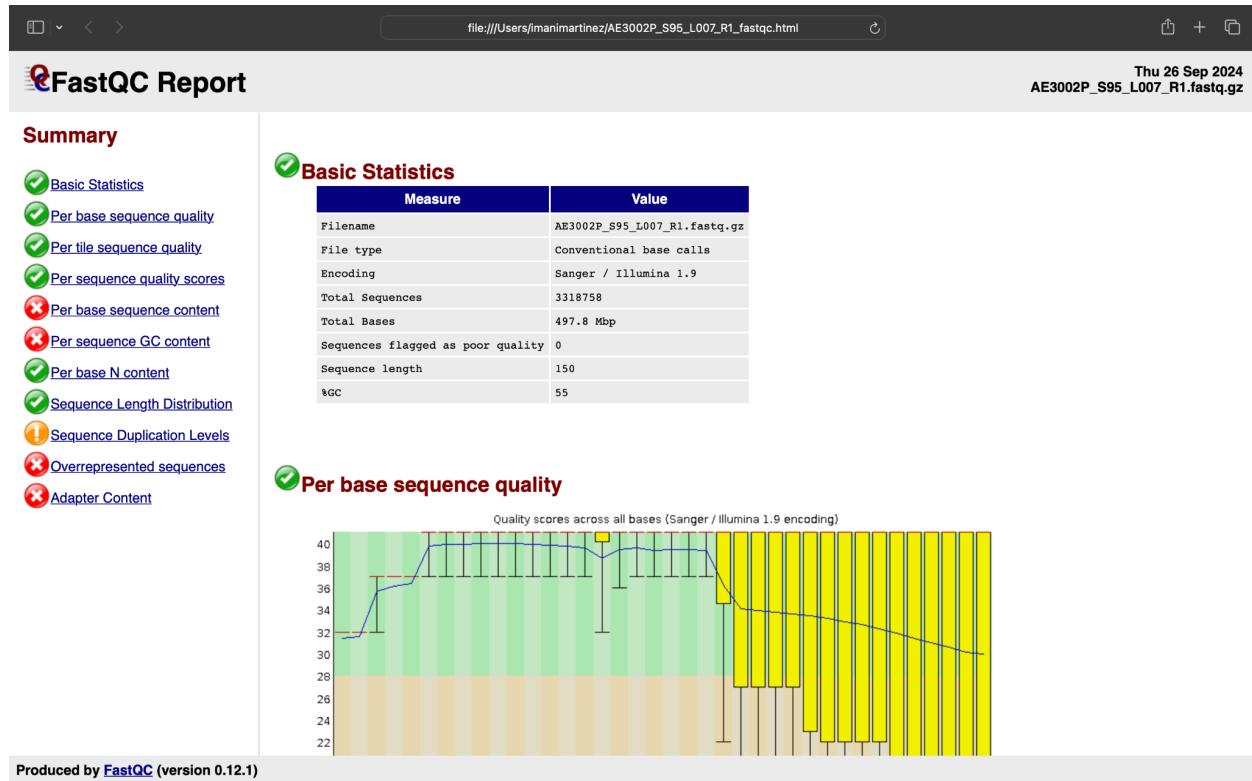


## Lab Assignment 6

### 1. CHECK QUALITY OF FASTQ SEQUENCE FILE



```

bamtools/2.4.1          i-tasser/5.1           pangolin/4          wrf/4.6.0
bamtools/2.5.2          i-tasser/5.2           parallel/20240422    wrf/4.6.0-mpi
bayesase/21.1.13        interprocсан/5.60-92.0  paraview/5.12.1-mpi
beagle/1.1              interprocсан/5.67-99.0  paraview/5.12.1-osmesa-mpi
bedtools/2.29.0          iqtree/1.6.12         parcels/2.1.5
bedtools/2.31.1          iqtree/2.1.2          parcels/2.1.6
blast/2.11.0+           iqtree/2.3.4          pcangs/1.2.9
blast/2.15.0+           jellyfish/2.3.0       plumed/2.2.9.1
blatsuite/36            jellyfish/2.3.1       moy/5.1.2
blatsuite/37            kraken2/2.1.4       prokka/1.14.6
blatsuite/38            kraken2/2.1.5       purge_haplotype/1.1.3
bowtie2/2.5.1           kraken2/2.1.6       pytorch/2.1.0-cudal1.7
bowtie2/2.5.3           lammps/02Aug23-cuda  DPLR/2.3.0-cudal1.7
bracken/2.9              lammps/02Aug23-mpi   oe/7.1-intel-mpi
braker/3.0.7             lammps/23Jun22-cuda  oe/7.3-intel-mpi
busco/3.0.2              lammps/23Jun22-mpi   qime2/2024.2
busco/4.1.4              liggghts/3.8.0       quast/5.0.2
busco/4.7.1              longstitch/1.0.6     R/4.2.2
bwameth/2.7              maf/1.0.457w000      R/4.2.2-mpi
canu/2.1.1               mafft/2.1.9w000      R/4.3.0
canu/2.2                 maker/3.0.0.03      R/4.3.3-mpi
cd-hit/4.8.1              masurca/4.1.0       ragtag/1.1.1
checkm/1.2.2             masurca/4.1.1       ragtag/2.1.0
clustal-omega/1.2.4      mathematica/12.3.1  raxml/8.2.4
consol/6.2                mathematica/13.3.0  raxml/8.2.4-mpi
cufflinks/2.2.1          mcrl/22018a       raxml/8.2.4-mpi
ddocent/2.9.8             mcrl/22019          raxml/8.2.4-mpi
decon/2.1                 mcrl/22019b       repeatmasker/9.1.0
degenprime/0.1.2          mega/19.2.6       repeatmasker/4.1.2
diamond/2.0.9              mega/11.0.13       repeatmasker/4.1.6
diamond/2.1.9              megadock/4.1.1     repeatmodeler/2.0.2
dm_control/1.0.16         megadock/4.1.1-mpi  repeatmodeler/2.0.5
dram/1.5.0                megalodon/2.5.0     repeatscout/1.0.6
emboss/6.0                 mega/19.2.6       rmblast/2.11.0
exav/1.8.21                mega/11.0.13       rmblast/2.14.1
examv/1.8.22              metacerberus/1.3.1  rnaquast/2.3.0
exonerate/2.4.0            metaphlan/4.1.0     rseqc/5.0.3
famsr/2.2                 microbeannotator/2.0.5  rstudio/2024.06
fastqc/8.11.9              minilign/0.6.0     samtools/1.11
FastQC/0.12.1             mira/5rc2         samtools/1.19
ffmpeg/4.2.1              mitobim/1.9.1      seqtk/1.3
:

^Z^Z^C
[imarti18@gal-19 BINF2111]$ ^C
[imarti18@gal-19 BINF2111]$ 
[1]+ Stopped                  sbatch
[imarti18@gal-19 BINF2111]$ sbatch qc.sh
Submitted batch job 101
[imarti18@gal-19 BINF2111]$ queue --me
JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)
1849 Centaurus myjob imarti18 R 0:12 1 gal-c1
[imarti18@gal-19 BINF2111]$ queue --me
JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)
[imarti18@gal-19 BINF2111]$ ls -lh
total 25M
-rw-r--r-- 1 imarti18 imarti18 623K Sep 26 14:48 AE3802P_S9P_L007_R1_fastqc.html
-rw-r--r-- 1 imarti18 imarti18 250K Sep 26 14:48 AE3802P_S9P_L007_R1_fastqc.zip
-rw-r--r-- 1 imarti18 imarti18 463 Sep 26 14:49 AE3802P_S9P_L007_R1_fastq.gz
-rw-r--r-- 1 imarti18 imarti18 497 Sep 26 14:42 qc.sh
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 14:44 slurm-1840.out
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 14:48 slurm-1849.out
[imarti18@gal-19 BINF2111]$ 

```

## 2. FIND AMOUNT OF PHIX IN FASTQ FILE

```

bucket 8: 88%
bucket 8: 98%
bucket 8: 100%
Sorting block of length 516 for bucket 8
(Using difference cover)
Sorting block time: 00:00:00
Returning block of 517 for bucket 8
Exited Ebtw loop
fchr[A]: 0
fchr[C]: 1292
fchr[G]: 2447
fchr[T]: 5386
fchr[$]: 5386
Exiting Ebwt::buildToDisk()
Returning from initFromVector
Wrote 4196384 bytes to primary EBWT file: phixDB.rev.1.bt2.tmp
Wrote 1352 bytes to secondary EBWT file: phixDB.rev.2.bt2.tmp
Re-opening _in1 and _in2 as input streams
Returning from Ebwt constructor
Headers:
len: 5386
bwtLen: 5387
sz: 1347
bwtSz: 1347
lineRate: 6
offRate: 4
offMask: 0xfffffffff0
ftabChars: 10
ftabLen: 20
ftabSz: 80
ftablen: 1048577
ftabsz: 4194308
offslen: 337
offsSz: 348
lineSz: 64
sideSz: 44
sideBwtSz: 48
sideBwtLen: 192
numSides: 29
numLines: 29
ebwtTotLen: 1856
ebwtTotSz: 1856
color: 0
reverse: 1
Total time for backward call to driver() for mirror index: 00:00:00
Renaming phixDB.3.bt2.tmp to phixDB.3.bt2
Renaming phixDB.4.bt2.tmp to phixDB.4.bt2
Renaming phixDB.1.bt2.tmp to phixDB.1.bt2
Renaming phixDB.2.bt2.tmp to phixDB.2.bt2
Renaming phixDB.rev.1.bt2.tmp to phixDB.rev.1.bt2
Renaming phixDB.rev.2.bt2.tmp to phixDB.rev.2.bt2
3318758 reads; of these:
3318758 (100.00%) were unpaired; of these:
3315706 (99.91%) aligned 0 times
3035 (0.0%) aligned exactly 1 time
17 (0.0%) aligned >1 times
0.0% overall alignment rate
=====
End Time : Thu Sep 26 03:07:49 PM EDT 2024
=====

[iamt118@gal-i9 BINF2111]$ 

```

```

Headers:
len: 5386
bwtLen: 5387
sz: 1347
bwtSz: 1347
lineRate: 6
offRate: 4
offMask: 0xfffffffff0
ftabChars: 10
ftabLen: 20
ftabSz: 80
ftablen: 1048577
ftabsz: 4194308
offslen: 337
offsSz: 1348
lineSz: 64
sideSz: 64
sideBwtLen: 192
numSides: 29
numLines: 29
ebwtTotLen: 1856
ebwtTotSz: 1856
color: 0
reverse: 1
Total time for backward call to driver() for mirror index: 00:00:00
Renaming phixDB.3.bt2.tmp to phixDB.3.bt2
Renaming phixDB.4.bt2.tmp to phixDB.4.bt2
Renaming phixDB.1.bt2.tmp to phixDB.1.bt2
Renaming phixDB.2.bt2.tmp to phixDB.2.bt2
Renaming phixDB.rev.1.bt2.tmp to phixDB.rev.1.bt2
Renaming phixDB.rev.2.bt2.tmp to phixDB.rev.2.bt2
3318758 reads; of these:
3318758 (100.00%) were unpaired; of these:
3315706 (99.91%) aligned 0 times
3035 (0.0%) aligned exactly 1 time
17 (0.0%) aligned >1 times
0.0% overall alignment rate
=====
End Time : Thu Sep 26 03:07:49 PM EDT 2024
=====

[iamt118@gal-i9 BINF2111]$ ls -lh
total 1.4G
-rw-r--r-- 1 iamt118 imart118 1.2G Sep 26 15:07 AE3002P_local.sam
-rw-r--r-- 1 iamt118 imart118 623K Sep 26 15:07 AE3002P_598_L007_R1.fastqc.html
-rw-r--r-- 1 iamt118 imart118 1.1G Sep 26 14:49 AE3002P_598_L007_R1.fastqc.zip
-rw-r--r-- 1 iamt118 imart118 258M Sep 26 14:49 AE3002P_598_L007_R1.fasta.gz
-rw-r--r-- 1 iamt118 imart118 636 Sep 26 14:59 bowtie2.sh
-rw-r--r-- 1 iamt118 imart118 463 Sep 26 14:37 example.sh
-rw-r--r-- 1 iamt118 imart118 5.4K Sep 26 14:58 phix174.fa
-rw-r--r-- 1 iamt118 imart118 4.1M Sep 26 15:00 phixDB.1.bt2
-rw-r--r-- 1 iamt118 imart118 1.4K Sep 26 15:00 phixDB.2.bt2
-rw-r--r-- 1 iamt118 imart118 1.3K Sep 26 15:00 phixDB.3.bt2
-rw-r--r-- 1 iamt118 imart118 1.4K Sep 26 15:00 phixDB.4.bt2
-rw-r--r-- 1 iamt118 imart118 4.1M Sep 26 15:00 phixDB.rev.1.bt2
-rw-r--r-- 1 iamt118 imart118 1.4K Sep 26 15:00 phixDB.rev.2.bt2
-rw-r--r-- 1 iamt118 imart118 499 Sep 26 14:42 qc.sh
-rw-r--r-- 1 iamt118 imart118 1.4K Sep 26 14:44 slurm-1840.out
-rw-r--r-- 1 iamt118 imart118 1.4K Sep 26 14:48 slurm-1849.out
-rw-r--r-- 1 iamt118 imart118 14K Sep 26 15:07 slurm-1868.out
[iamt118@gal-i9 BINF2111]$ 

```

### 3. REMOVE PHIX AND CONVERT SAM TO FASTQ

```
-rw-r--r-- 1 imarti18 imarti18 5.4K Sep 26 14:58 phix174.fa
-rw-r--r-- 1 imarti18 imarti18 4.1K Sep 26 15:00 phixDB.2.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.2.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.3.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.4.bt2
-rw-r--r-- 1 imarti18 imarti18 4.1M Sep 26 15:00 phixDB.rev.1.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4M Sep 26 15:00 phixDB.rev.2.bt2
-rw-r--r-- 1 imarti18 imarti18 499 Sep 26 14:42 qc.sh
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 14:44 slurm-1840.out
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 14:48 slurm-1849.out
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:07 slurm-1886.out
[imarti18@gal-i9 BINF2111]$ cp example.sh samtools.sh
[imarti18@gal-i9 BINF2111]$ nano samtools.sh
[imarti18@gal-i9 BINF2111]$ sbatch samtools.sh
Submitted batch job 1886
[imarti18@gal-i9 BINF2111]$ squeue --me
JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)
1886_centaurus myjob imarti18 R 0:07 1 gal-c1
[imarti18@gal-i9 BINF2111]$ squeue -m
JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)
[imarti18@gal-i9 BINF2111]$ cat samtools.sh
#!/bin/bash

#SBATCH --job-name=myjob
#SBATCH --partition=Centaurus
#SBATCH --time=00:10:00

echo "=====
echo "Start Time : $(date)"
echo "=====
echo "# ===== PUT YOUR COMMANDS HERE ====="
#load the module
module load samtools

#convert sam to bam
samtools view -B AE3002P_local.sam > AE3002P_local.bam

#convert bam to unmapped bam
samtools view -b -f 4 AE3002P_local.bam > AE3002P_unmapped-local.bam

#convert unmapped bam to fastq
samtools fastq AE3002P_unmapped-local.bam > AE3002P_unmapped-local.fastq

echo "=====
echo "End Time : $(date)"
echo "=====
echo "[imarti18@gal-i9 BINF2111]$ cat slurm-1886.out
=====
Start Time : Thu Sep 26 03:12:44 PM EDT 2024
=====

[M::bam2fq_mainloop] discarded 0 singletons
[M::bam2fq_mainloop] processed 3315786 reads
=====
End Time : Thu Sep 26 03:14:16 PM EDT 2024
=====

[imarti18@gal-i9 BINF2111]$ "
```

```
#SBATCH --partition=Centaurus
#SBATCH --time=00:10:00
echo "=====
echo "Start Time : $(date)"
echo "=====
echo
# ----- PUT YOUR COMMANDS HERE -----
#load the module
module load samtools

#convert sam to bam
samtools view -bS AE3002P_local.sam > AE3002P_local.bam

#convert bam to unmapped bam
samtools view -b -f 4 AE3002P_local.bam > AE3002P_unmapped-local.bam

#convert unmapped bam to fastq
samtools fastq AE3002P_unmapped-local.bam > AE3002P_unmapped-local.fastq

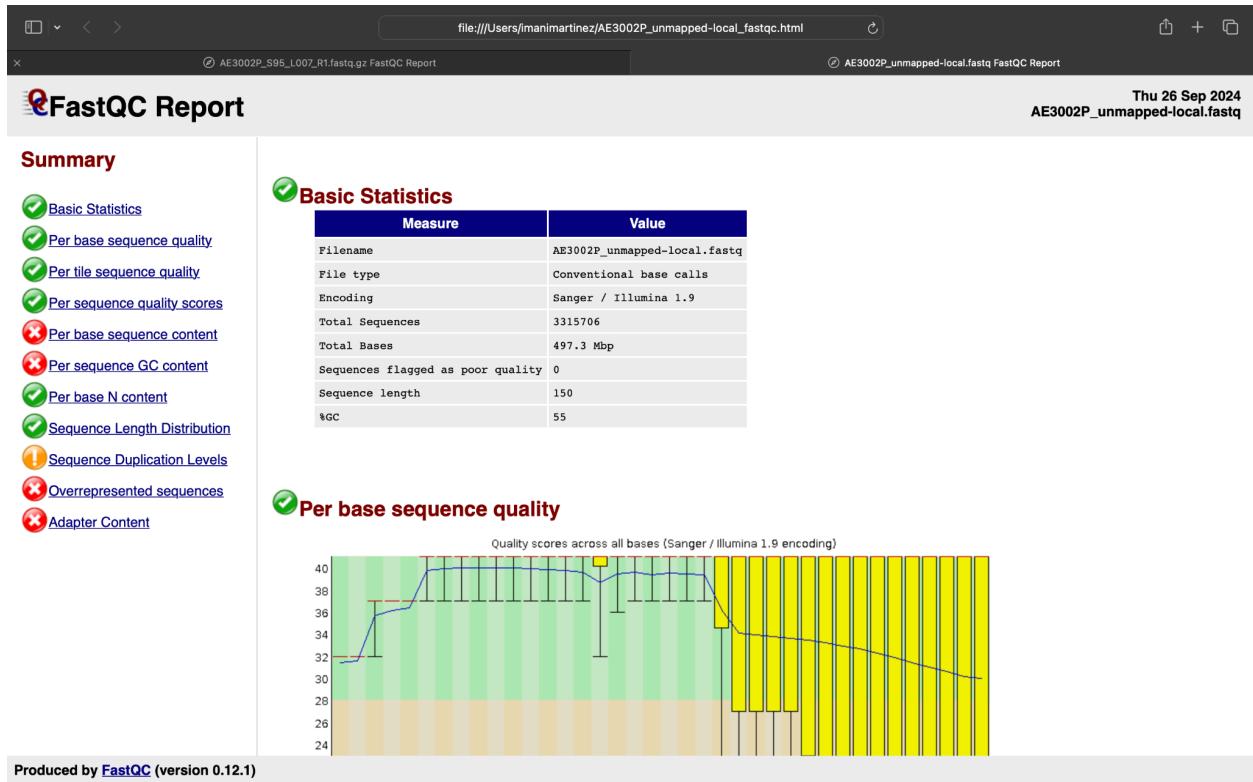
echo
echo "=====
echo "End Time : $(date)"
echo "=====
echo
[imarti18@gal-19 BINF2111]$ cat slurm-1886.out
=====
Start Time : Thu Sep 26 03:12:44 PM EDT 2024
=====

[M::bam2fq_mainloop] discarded 0 singletons
[M::bam2fq_mainloop] processed 3315786 reads

=====
End Time : Thu Sep 26 03:14:16 PM EDT 2024
=====

[imarti18@gal-19 BINF2111]$ ls -lh
total 3.0G
-rw-r--r-- 1 imarti18 imarti18 232M Sep 26 15:13 AE3002P_local.bam
-rw-r--r-- 1 imarti18 imarti18 1.2G Sep 26 15:07 AE3002P_local.sam
-rw-r--r-- 1 imarti18 imarti18 623K Sep 26 14:48 AE3002P_S95_L007_R1_fastqc.html
-rw-r--r-- 1 imarti18 imarti18 498K Sep 26 14:48 AE3002P_S95_L007_R1.fastqc.zip
-rw-r--r-- 1 imarti18 imarti18 258M Sep 26 14:48 AE3002P_S95_L007_R1.fastq.gz
-rw-r--r-- 1 imarti18 imarti18 232M Sep 26 15:14 AE3002P_unmapped-local.bam
-rw-r--r-- 1 imarti18 imarti18 1.1G Sep 26 15:14 AE3002P_unmapped-local.fastq
-rw-r--r-- 1 imarti18 imarti18 636 Sep 26 14:51 bowtie1.sh
-rw-r--r-- 1 imarti18 imarti18 465K Sep 26 14:57 bowtie2.sh
-rw-r--r-- 1 imarti18 imarti18 5.1K Sep 26 14:58 bowtie2_14_fa
-rw-r--r-- 1 imarti18 imarti18 4.1M Sep 26 15:00 phixDB.1.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.2.bt2
-rw-r--r-- 1 imarti18 imarti18 17 Sep 26 15:00 phixDB.3.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.4.bt2
-rw-r--r-- 1 imarti18 imarti18 4.1M Sep 26 15:00 phixDB.rev.1.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.rev.2.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 14:42 phixDB.sh
-rw-r--r-- 1 imarti18 imarti18 781 Sep 26 15:12 samtools.sh
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 14:44 slurm-1848.out
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 14:48 slurm-1849.out
-rw-r--r-- 1 imarti18 imarti18 14K Sep 26 15:07 slurm-1868.out
-rw-r--r-- 1 imarti18 imarti18 399 Sep 26 15:14 slurm-1886.out
[imarti18@gal-19 BINF2111]$
```

## 4. RERUN QC



```

-rw-r--r-- 1 imarti18 imarti18 1.20 Sep 26 15:07 AE3802P_local.sam
-rw-r--r-- 1 imarti18 imarti18 623K Sep 26 14:48 AE3802P_S98_L007_R1_fastqc.html
-rw-r--r-- 1 imarti18 imarti18 490K Sep 26 14:48 AE3802P_S98_L007_R1_fastqc.zip
-rw-r--r-- 1 imarti18 imarti18 258K Sep 26 15:14 AE3802P_unmapped-local.bam
-rw-r--r-- 1 imarti18 imarti18 230M Sep 26 15:14 AE3802P_unmapped-local.fastq
-rw-r--r-- 1 imarti18 imarti18 1.10 Sep 26 15:14 AE3802P_unmapped-local.fastq.gz
-rw-r--r-- 1 imarti18 imarti18 636 Sep 26 14:59 bowtie.sh
-rw-r--r-- 1 imarti18 imarti18 463 Sep 26 14:37 example.sh
-rw-r--r-- 1 imarti18 imarti18 5.4K Sep 26 14:58 phixDB.1.bt2
-rw-r--r-- 1 imarti18 imarti18 4.1M Sep 26 15:00 phixDB.1.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.2.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.3.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.4.bt2
-rw-r--r-- 1 imarti18 imarti18 4.1M Sep 26 15:00 phixDB.rev.1.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.rev.2.bt2
-rw-r--r-- 1 imarti18 imarti18 499 Sep 26 14:42 qc.sh
-rw-r--r-- 1 imarti18 imarti18 781 Sep 26 15:12 samtools.sh
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 14:44 slurm-1840.out
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:07 slurm-1840.out
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:07 slurm-1868.out
-rw-r--r-- 1 imarti18 imarti18 399 Sep 26 15:14 slurm-1886.out
[imarti18@gal-19 BINF2111]$ cp qc.sh qc-post.sh
[imarti18@gal-19 BINF2111]$ nano qc-post.sh
[imarti18@gal-19 BINF2111]$ sbatch qc-post.sh
Submitted batch job 1901
[imarti18@gal-19 BINF2111]$ squeue --me
              JOBID PARTITION   NAME   USER ST      TIME NODES NODELIST(REASON)
          1901 Centaurus myjob imarti18 R      0:07      1 gal-c1
[imarti18@gal-19 BINF2111]$ squeue --me
              JOBID PARTITION   NAME   USER ST      TIME NODES NODELIST(REASON)
          1901 Centaurus myjob imarti18 R      0:39      1 gal-c1
[imarti18@gal-19 BINF2111]$ squeue --me
              JOBID PARTITION   NAME   USER ST      TIME NODES NODELIST(REASON)
[imarti18@gal-19 BINF2111]$ ls -lh
total 3.8G
-rw-r--r-- 1 imarti18 imarti18 232M Sep 26 15:13 AE3802P_local.bam
-rw-r--r-- 1 imarti18 imarti18 1.20 Sep 26 15:07 AE3802P_local.sam
-rw-r--r-- 1 imarti18 imarti18 623K Sep 26 15:20 AE3802P_S98_L007_R1_fastqc.html
-rw-r--r-- 1 imarti18 imarti18 490K Sep 26 15:20 AE3802P_S98_L007_R1_fastqc.zip
-rw-r--r-- 1 imarti18 imarti18 258K Sep 26 15:14 AE3802P_unmapped-local.bam
-rw-r--r-- 1 imarti18 imarti18 230M Sep 26 15:14 AE3802P_unmapped-local.fastq
-rw-r--r-- 1 imarti18 imarti18 623K Sep 26 15:21 AE3802P_unmapped-local.fastqc.html
-rw-r--r-- 1 imarti18 imarti18 490K Sep 26 15:21 AE3802P_unmapped-local_fastqc.zip
-rw-r--r-- 1 imarti18 imarti18 636 Sep 26 14:59 bowtie.sh
-rw-r--r-- 1 imarti18 imarti18 463 Sep 26 14:37 example.sh
-rw-r--r-- 1 imarti18 imarti18 5.4K Sep 26 14:58 phixDB.1.bt2
-rw-r--r-- 1 imarti18 imarti18 4.1M Sep 26 15:00 phixDB.1.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.2.bt2
-rw-r--r-- 1 imarti18 imarti18 1.7 Sep 26 15:00 phixDB.3.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.4.bt2
-rw-r--r-- 1 imarti18 imarti18 4.1M Sep 26 15:00 phixDB.rev.1.bt2
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:00 phixDB.rev.2.bt2
-rw-r--r-- 1 imarti18 imarti18 535 Sep 26 15:19 qc-post.sh
-rw-r--r-- 1 imarti18 imarti18 499 Sep 26 15:12 samtools.sh
-rw-r--r-- 1 imarti18 imarti18 781 Sep 26 15:12 samtools.sh
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 14:44 slurm-1840.out
-rw-r--r-- 1 imarti18 imarti18 1.4K Sep 26 15:07 slurm-1868.out
-rw-r--r-- 1 imarti18 imarti18 399 Sep 26 15:14 slurm-1886.out
-rw-r--r-- 1 imarti18 imarti18 2.5K Sep 26 15:21 slurm-1901.out
[imarti18@gal-19 BINF2111]$ 

```

**How many sequences were removed? (HINT: Subtract the Total Sequences from the fastqc HTML outputs)**

3318758 - 3315706 = 3052 sequences removed