

BIOL550- Bioinformatics
Fall 2023 Long Assignment Answers
Names- Utsav Pathak, Harshini Balaga
A IDs- A20454823, A20517086

Read Processing

i) Working Directory- /media/Data_2/BIOL550/F23/TeamAssignments/T02

ran fastqc GUI to check the read quality of all the three files.

This can also be done as using:-

`fastqc /media/Data_2/BIOL550/F23/LongAssignment/illumina.R1.fastq.gz`

`fastqc /media/Data_2/BIOL550/F23/LongAssignment/illumina.R2.fastq.gz`

`fastqc /media/Data_2/BIOL550/F23/LongAssignment/nanopore.fastq.gz`

The read quality is good for the Illumina because the sequences flagged as poor are 0 and the adapter content, which is important for the Illumina data, is also satisfied.

The image contains two side-by-side screenshots of the FastQC software interface. Both screenshots show the 'Basic sequence stats' table for two different fastq files: 'illumina.R1.fastq.gz' on the left and 'illumina.R2.fastq.gz' on the right. The table provides detailed information about the sequence quality and content.

| Measure | Value |
|-----------------------------------|-------------------------|
| Filename | illumina.R1.fastq.gz |
| File type | Conventional base calls |
| Encoding | Sanger / Illumina 1.9 |
| Total Sequences | 570450 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 35-301 |
| %GC | 39 |

On the left side of each screenshot, there is a list of analysis modules with their status icons:

- Basic Statistics (Green checkmark)
- Per base sequence quality (Orange exclamation mark)
- Per sequence quality scores (Green checkmark)
- Per base sequence content (Red X)
- Per sequence GC content (Orange exclamation mark)
- Per base N content (Green checkmark)
- Sequence Length Distribution (Orange exclamation mark)
- Sequence Duplication Levels (Green checkmark)
- Overrepresented sequences (Orange exclamation mark)
- Adapter Content (Green checkmark)

ii) wget https://raw.githubusercontent.com/PombertLab/Misc/main/read_len_plot.py
Copied the nanopore file to working directory for ease

```
python read_len_plot.py -f nanopore.fastq.gz
```

The above command shows the metrics as below.

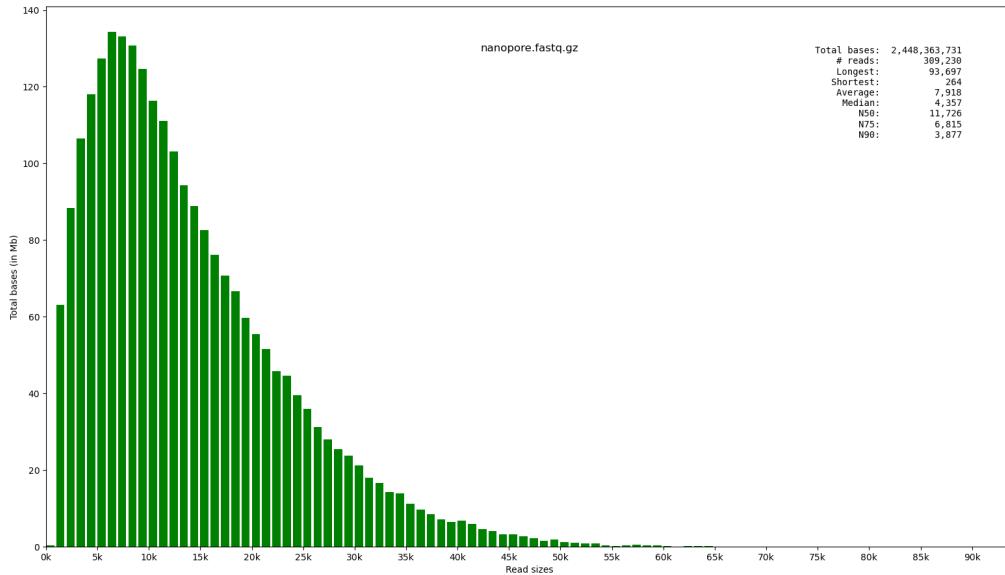
```
Working on nanopore.fastq.gz...

Metrics for nanopore.fastq.gz:

Total bases:      2,448,363,731
# reads:          309,230
Longest:          93,697
Shortest:         264
Average:          7,918
Median:           4,357
N50:              11,726
N75:              6,815
N90:              3,877
```

```
python read_len_plot.py -f nanopore.fastq.gz -o output.png
```

The above command plots and saves the plot in the working directory with the name output.png, this can be seen as below.



iii) gunzip -c nanopore.fastq.gz | NanoFilt -q 10 -l 500 --headcrop 50 | gzip > trimmedreads.fastq.gz

The command above filters and trims the head size for the nanopore datasets to the specified values as above.

iv)

screen -S SPAdes

spades.py -k 21,33,55,77 --careful --pe1-1 illumina.R1.fastq.gz --pe1-2 illumina.R2.fastq.gz -t 8 -o spades

screen -S ShastaRun

gunzip nanopore.fastq.gz

shasta --input nanopore.fastq --config Nanopore-May2022

screen -S Flye

flye -t 8 --nano-raw nanopore.fastq.gz -o flye

screen -S Unicycler

unicycler --spades_path /opt/SPAdes-3.13.0-Linux/bin/spades.py -t 8 -1 illumina.R1.fastq.gz -2 illumina.R2.fastq.gz --pilon_path /opt/pilon/pilon-1.24.jar -l nanopore.fastq.gz -o hybrid_assembly

v)

FLYE:-

```
runTaxonomizedBLAST.pl -t 8 -p blastn -a megablast -db ref_prok_rep_genomes -q  
draft_assembly.fasta -e 1e-30 -c 1 -o BLAST_RESULTS
```

```
parseTaxonomizedBLAST.pl -v -b BLAST_RESULTS/contigs.blastn.6 -f contigs.fasta -n  
"Listeria monocytogenes" Listeria Monocytogenes -c sscinames -o S_listeria.fasta
```

```
less BLAST_RESULTS/contigs.blastn.6
```

SPADES:-

```
runTaxonomizedBLAST.pl -t 8 -p blastn -a megablast -db ref_prok_rep_genomes -q  
contigs.fasta -e 1e-30 -c 1 -o BLAST_RESULTS
```

```
parseTaxonomizedBLAST.pl -v -b BLAST_RESULTS/contigs.blastn.6 -f contigs.fasta -n  
"Listeria monocytogenes" Listeria Monocytogenes -c sscinames -o S_listeria.fasta
```

```
less BLAST_RESULTS/contigs.blastn.6
```

```
Best hit for NODE_983_length_569_cov_1.441057 = Moraxella osloensis  
Best hit for NODE_984_length_569_cov_1.428732 = Moraxella osloensis  
Best hit for NODE_985_length_569_cov_1.195122 = Moraxella osloensis  
Best hit for NODE_986_length_568_cov_2.057026 = Moraxella osloensis  
Best hit for NODE_987_length_568_cov_1.890020 = Moraxella osloensis  
Best hit for NODE_988_length_568_cov_1.802444 = Moraxella osloensis  
Best hit for NODE_989_length_568_cov_1.737271 = Moraxella osloensis  
Best hit for NODE_98_length_1462_cov_5.418830 = Moraxella osloensis  
Best hit for NODE_990_length_568_cov_1.588448 = Moraxella osloensis  
Best hit for NODE_991_length_567_cov_2.116327 = Moraxella osloensis  
Best hit for NODE_992_length_567_cov_1.891837 = Moraxella osloensis  
Best hit for NODE_993_length_567_cov_1.265306 = Moraxella osloensis  
Best hit for NODE_994_length_566_cov_2.728016 = Moraxella osloensis  
Best hit for NODE_995_length_566_cov_1.319018 = Moraxella osloensis  
Best hit for NODE_996_length_566_cov_1.122699 = Moraxella osloensis  
Best hit for NODE_997_length_565_cov_2.266393 = Moraxella osloensis  
Best hit for NODE_998_length_565_cov_1.852459 = Moraxella osloensis  
Best hit for NODE_999_length_565_cov_1.295082 = Moraxella osloensis  
Best hit for NODE_99_length_1441_cov_1.849707 = Moraxella osloensis  
Best hit for NODE_9_length_151126_cov_55.931817 = Listeria monocytogenes EGD-e  
Match found for Monocytogenes: NODE_9_length_151126_cov_55.931817 gi|16802048|ref|NC_003210.1| 179 101479 93.926 101354 1.529e+05 0.0 169963 L  
isteria monocytogenes EGD-e Bacteria firmicutes  
[biolF23_23@Hozart spades]$
```

Unicycler:-

```
runTaxonomizedBLAST.pl -t 8 -p blastn -a megablast -db ref_prok_rep_genomes -q  
assembly.fasta -e 1e-30 -c 1 -o BLAST_RESULTS
```

```
parseTaxonomizedBLAST.pl -v -b BLAST_RESULTS/assembly.blastn.6 -f assembly.fasta -n  
"Listeria monocytogenes" Listeria Monocytogenes -c sscinames -o S_listeria.fasta
```

```
less BLAST_RESULTS/assembly.blastn.6
```

```
[biolF23_23@Mozart hybrid_assembly]$ parseTaxonomizedBLAST.pl \v \b BLAST_RESULTS/assembly.blastn.6 \f assembly.fasta \n "Listeria monocytogenes" Listeria Monocytogenes \c sscinames \o S_listeria.fasta
Best hit for 1 = Listeria monocytogenes EGD-e
Match found for Monocytogenes: 1 gi|16802048|ref|NC_003210.1| 788775 980271 94.909 191557 2.996e+05 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 2 = Listeria monocytogenes EGD-e
Match found for Monocytogenes: 2 gi|16802048|ref|NC_003210.1| 126616 256152 96.144 129585 2.115e+05 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 3 = Listeria monocytogenes EGD-e
Match found for Monocytogenes: 3 gi|16802048|ref|NC_003210.1| 36137 49649 96.510 13525 22336 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 4 = Listeria innocua
Match found for Listeria: 4 gi|1930446050|ref|NZ_JACTKR010000006.1| 19689 31862 85.087 12224 12386 0.0 1642 Listeria innocua Bacteria firmicutes
Best hit for 5 = Listeria monocytogenes EGD-e
Match found for Monocytogenes: 5 gi|16802048|ref|NC_003210.1| 223 1279 82.846 1061 942 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
[biolF23_23@Mozart hybrid_assembly]$
```

ShastaRun:-

```
runTaxonomizedBLAST.pl \t 8 \p blastn \a megablast \db ref_prok_rep_genomes \q
Assembly.fasta \e 1e-30 \c 1 \o BLAST_RESULTS
```

```
parseTaxonomizedBLAST.pl \v \b BLAST_RESULTS/Assembly.blastn.6 \f Assembly.fasta \n
"Listeria monocytogenes" Listeria Monocytogenes \c sscinames \o S_listeria.fasta
```

less BLAST_RESULTS/Assembly.blastn.6

```
ia firmicutes
Best hit for 828 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 828 gi|16802048|ref|NC_003210.1| 1 9302 94.598 9348 14397 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 830 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 830 gi|16802048|ref|NC_003210.1| 1 561 93.250 563 822 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 832 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 832 gi|16802048|ref|NC_003210.1| 1 2396 95.102 2409 3775 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 834 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 834 gi|16802048|ref|NC_003210.1| 1 4712 91.687 4812 6514 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 836 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 836 gi|16802048|ref|NC_003210.1| 1 3600 93.009 3619 5260 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 838 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 838 gi|16802048|ref|NC_003210.1| 6 2680 95.433 2693 4268 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 84 = Listeria innocua
Match found for Listeria: 84 gi|1930445938|ref|NZ_JACTKR010000001.1| 8875 11306 95.238 2436 3843 0.0 1642 Listeria innocua Bacteria firmicutes
Best hit for 840 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 840 gi|16802048|ref|NC_003210.1| 1 5524 95.059 5566 8700 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 842 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 842 gi|16802048|ref|NC_003210.1| 2 3460 96.317 3475 5681 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 86 = Listeria innocua
Match found for Listeria: 86 gi|1930445938|ref|NZ_JACTKR010000001.1| 1 22221 97.385 22291 37796 0.0 1642 Listeria innocua Bacteria firmicutes
Best hit for 88 = Listeria marthii
Match found for Listeria: 88 gi|1893642213|ref|NZ_JAARXK010000003.1| 1423 12043 90.087 10663 13740 0.0 529731 Listeria marthii Bacteria firmicutes
Best hit for 90 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 90 gi|16802048|ref|NC_003210.1| 1 4757 94.207 4782 7258 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 92 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 92 gi|16802048|ref|NC_003210.1| 1 18042 96.409 18155 29673 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 94 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 94 gi|16802048|ref|NC_003210.1| 1 15124 93.023 15192 22066 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 96 = Listeria monocytogenes EGD-e
Match found for Listeria monocytogenes: 96 gi|16802048|ref|NC_003210.1| 1 23831 94.254 23914 36387 0.0 169963 Listeria monocytogenes EGD-e Bacteria firmicutes
Best hit for 98 = Listeria innocua
Match found for Listeria: 98 gi|1930445992|ref|NZ_JACTKR010000003.1| 6349 10056 88.685 3756 4490 0.0 1642 Listeria innocua Bacteria firmicutes
[biolF23_23@Mozart ShastaRun]$
```

vi)

The best approach for this is the Hybrid approach as we have both short reads and long reads in the data.

Flye and Unicycler are good for Oxford Nanopore only.

GC% (Garbage Collection) is least for hybrid as well with 38.06%

```
quast.py \--min-contig 500 \Assembly.fasta \o flye_quast  
firefox flye_quast/report.html
```

```
quast.py \--min-contig 500 \Assembly.fasta \o Shasta_quast  
firefox Shasta_quast/report.html
```

```
quast.py \--min-contig 500 \Assembly.fasta \o spades_quast
```

```
firefox spades_quast/report.html
```

```
quast.py \--min-contig 500 \assembly.fasta \o hybrid_quast
```

```
firefox hybrid_quast/report.html
```

Report

| | Assembly |
|----------------------------|----------|
| # contigs (>= 0 bp) | 422 |
| # contigs (>= 1000 bp) | 412 |
| # contigs (>= 5000 bp) | 345 |
| # contigs (>= 10000 bp) | 220 |
| # contigs (>= 25000 bp) | 52 |
| # contigs (>= 50000 bp) | 16 |
| Total length (>= 0 bp) | 5862116 |
| Total length (>= 1000 bp) | 5856173 |
| Total length (>= 5000 bp) | 5642514 |
| Total length (>= 10000 bp) | 4703092 |
| Total length (>= 25000 bp) | 2125816 |
| Total length (>= 50000 bp) | 1012040 |
| # contigs | 418 |
| Largest contig | 90984 |
| Total length | 5860991 |
| GC (%) | 38.10 |
| N50 | 18899 |
| N90 | 7490 |
| L50 | 90 |
| L90 | 286 |
| # N's per 100 kbp | 0.00 |

All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

Report

| | contigs |
|----------------------------|---------|
| # contigs (>= 0 bp) | 1990 |
| # contigs (>= 1000 bp) | 265 |
| # contigs (>= 5000 bp) | 29 |
| # contigs (>= 10000 bp) | 27 |
| # contigs (>= 25000 bp) | 25 |
| # contigs (>= 50000 bp) | 18 |
| Total length (>= 0 bp) | 4260154 |
| Total length (>= 1000 bp) | 3276592 |
| Total length (>= 5000 bp) | 2945221 |
| Total length (>= 10000 bp) | 2932066 |
| Total length (>= 25000 bp) | 2900569 |
| Total length (>= 50000 bp) | 2661684 |
| # contigs | 1250 |
| Largest contig | 340237 |
| Total length | 3950861 |
| GC (%) | 39.67 |
| N50 | 144632 |
| N90 | 730 |
| L50 | 10 |
| L90 | 592 |
| # N's per 100 kbp | 0.00 |

All statistics are based on contigs of size >= 500 bp, unless otherwise noted
(e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs).

Report

| | assembly |
|----------------------------|----------|
| # contigs (>= 0 bp) | 6 |
| # contigs (>= 1000 bp) | 5 |
| # contigs (>= 5000 bp) | 4 |
| # contigs (>= 10000 bp) | 4 |
| # contigs (>= 25000 bp) | 4 |
| # contigs (>= 50000 bp) | 2 |
| Total length (>= 0 bp) | 3031848 |
| Total length (>= 1000 bp) | 3031646 |
| Total length (>= 5000 bp) | 3029963 |
| Total length (>= 10000 bp) | 3029963 |
| Total length (>= 25000 bp) | 3029963 |
| Total length (>= 50000 bp) | 2945138 |
| # contigs | 5 |
| Largest contig | 1903610 |
| Total length | 3031646 |
| GC (%) | 38.06 |
| N50 | 1903610 |
| N90 | 1041528 |
| L50 | 1 |
| L90 | 2 |
| # N's per 100 kbp | 0.00 |

```
vii) prokka --outdir prokka_annotation \--genus Listeria --species monocytogenes \--locustag F23 --increment 10 \--compliant \S_listeria.fasta
```

```
dfast -g ./S_listeria.fasta -o DFAST --strain Listeria Monocytogenes --cpu 8
```

```
organism: Listeria monocytogenes strain
contigs: 5
bases: 3031646
CDS: 2970
CRISPR: 1
rRNA: 18
tRNA: 67
tmRNA: 1
```

```
GNU nano 5.8
Total Sequence Length (bp)      3031646
Number of Sequences      5
Longest Sequences (bp) 1903610
N50 (bp)          1903610
Gap Ratio (%)    0.000000
GCcontent (%)    38.1
Number of CDSs   2990
Average Protein Length 300.9
Coding Ratio (%) 89.0
Number of rRNAs  18
Number of tRNAs 67
Number of CRISPRs 2
```

viii)

```
grep -c '/locus_tag=' Listeriamonocytogenes.gb prokka_annotation/PROKKA_12082023.gbk
\DFAST/genome.gbk
```

```
prokka_annotation/PROKKA_12082023.gbk:3056
DFAST/genome.gbk:3076
[biolF23_23@Mozart hybrid_assembly]$
```

The results are not congruent between PROKKA and DFAST, this is because the default reference database for PROKKA is 18276 Sequences while DFAST has a larger database of 417922 Sequences.

ix)

```
get_SNPs.pl \-fa hybrid_assembly/assembly.fasta \-pe1 *R1.fastq.gz \-pe2 *R2.fastq.gz \-mapper  
minimap2 \-preset sr \-rmo \-bam
```

```
samtools bam2fq \-f 1 \-F 12 \-1 map_R1.fastq \-2 map_R2.fastq  
\minimap2.BAM/illumina.R1.fastq.gz.assembly.fasta.minimap2.bam
```

```
samtools bam2fq \-f 12 \-F 256 \-1 unmap_R1.fastq \-2 unmap_R2.fastq  
\minimap2.BAM/illumina.R1.fastq.gz.assembly.fasta.minimap2.bam
```

```
get_SNPs.pl \-fa hybrid_assembly/assembly.fasta -pe1 *R1.fastq.gz -pe2 *R2.fastq.gz -mapper  
minimap2 \-preset sr \-rmo \-bam \-t 8
```

```
[biolF23_23@Mozart T02]$ samtools bam2fq \-f 1 \-F 12 \-1 map_R1.fastq \-2 map_R2.fastq \minimap2.BAM/illumina.R1.fastq.gz.assembly.fasta.minim  
ap2.bam  
[M:\bam2fq_mainloop] discarded 0 singlettons  
[M:\bam2fq_mainloop] processed 1077985 reads  
[biolF23_23@Mozart T02]$ samtools bam2fq \-f 12 \-F 256 \-1 unmap_R1.fastq \-2 unmap_R2.fastq \minimap2.BAM/illumina.R1.fastq.gz.assembly.fasta.  
.minimap2.bam  
[M:\bam2fq_mainloop] discarded 0 singlettons  
[M:\bam2fq_mainloop] processed 58332 reads  
[biolF23_23@Mozart T02]$ get_SNPs.pl \-fa hybrid_assembly/assembly.fasta -pe1 *R1.fastq.gz -pe2 *R2.fastq.gz -mapper minimap2 \-preset sr \-rmo \-b  
am \-t 8  
Option t is ambiguous (threads, type)  
  
## FASTQ information:  
R1 FASTQ parsed as: illumina.R1.fastq.gz  
R2 FASTQ parsed as: illumina.R2.fastq.gz  
FASTQ input DIR parsed as: ./  
  
## FASTA information:  
FASTA parsed as: assembly.fasta  
FASTA input DIR parsed as: hybrid_assembly/  
  
Mapping illumina.R1.fastq.gz and illumina.R2.fastq.gz on hybrid_assembly/assembly.fasta with minimap2...  
Running samtools on ./illumina.R1.fastq.gz.assembly.fasta.minimap2.sam...  
Using 1624 Mb per thread for samtools  
[bam_sort_core] merging from 0 files and 16 in-memory blocks...  
  
Calculating stats...  
Time to calculate stats: 41 seconds  
  
Mapping/SNP calling started on: Fri Dec 8 14:40:47 2023  
Mapping/SNP calling ended on: Fri Dec 8 14:41:37 2023  
Time elapsed: 50 seconds  
[biolF23_23@Mozart T02]$ █
```

```
cat minimap2.rmo.stats/*
```

```
[biolF23_23@Mozart T02]$ cat minimap2.rmo.stats/*
FASTQ file(s) used: illumina.R1.fastq.gz (and mate, if PE)
Reference fasta file used: hybrid_assembly/assembly.fasta

Total number of bases in the reference genome 3031848 bp
Number of bases covered by at least one read 3017967
Number of bases without coverage 13881
Maximum sequencing depth 292X
Average sequencing depth 88.94X
Sequencing breadth (percentage of bases covered by at least one read) 99.54%

## SAMTOOLS flagstat metrics
1141808 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
908 + 0 supplementary
0 + 0 duplicates
1080733 + 0 mapped (94.65% : N/A)
1140900 + 0 paired in sequencing
570450 + 0 read1
570450 + 0 read2
1067320 + 0 properly paired (93.55% : N/A)
1077082 + 0 with itself and mate mapped
2743 + 0 singletons (0.24% : N/A)
264 + 0 with mate mapped to a different chr
149 + 0 with mate mapped to a different chr (mapQ>=5)

[biolF23_23@Mozart T02]$
```

```
gzip -k *.fastq
```

```
[biolF23_23@Mozart T02]$ ls
flye           mapping.minimap2.log  minimap2.rmo.coverage  output.png      spades          unmap_R2.fastq
hybrid_assembly map_R1.fastq.gz   minimap2.rmo.depth   read_len_plot.py  time.minimap2.rmo.log
illumina.R1.fastq.gz map_R2.fastq.gz minimap2.rmo.stats  read_metrics.txt  trimmedreads.fastq.gz
illumina.R2.fastq.gz minimap2.BAM    nanopore.fastq.gz ShastaRun       unmap_R1.fastq
[biolF23_23@Mozart T02]$ gzip -k *.fastq
[biolF23_23@Mozart T02]$ ls
flye           mapping.minimap2.log  minimap2.rmo.coverage  output.png      spades          unmap_R1.fastq.gz
hybrid_assembly map_R1.fastq.gz   minimap2.rmo.depth   read_len_plot.py  time.minimap2.rmo.log
illumina.R1.fastq.gz map_R2.fastq.gz minimap2.rmo.stats  read_metrics.txt  trimmedreads.fastq.gz
illumina.R2.fastq.gz minimap2.BAM    nanopore.fastq.gz ShastaRun       unmap_R2.fastq
[biolF23_23@Mozart T02]$
```

```
du -sh illumina.R1.fastq.gz unmap_R1.fastq.gz
```

```
du -sh illumina.R2.fastq.gz unmap_R2.fastq.gz
```

```
[biolF23_23@Mozart T02]$ du -sh illumina.R1.fastq.gz unmap_R1.fastq.gz
80M    illumina.R1.fastq.gz
1.7M   unmap_R1.fastq.gz
[biolF23_23@Mozart T02]$ du -sh illumina.R2.fastq.gz unmap_R2.fastq.gz
98M    illumina.R2.fastq.gz
1.8M   unmap_R2.fastq.gz
[biolF23_23@Mozart T02]$
```