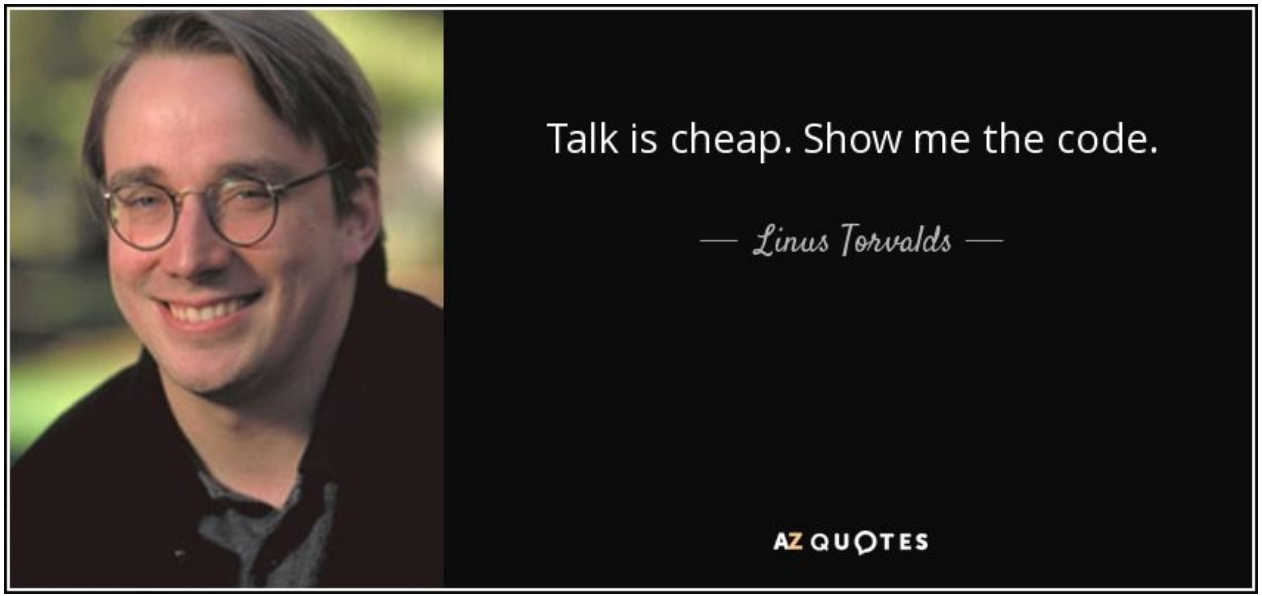
WRAP-UP LINUX COMMAND LINE:

An example for real-world (bioinformatics) application

Duy Dao

- 2024.05.30 -

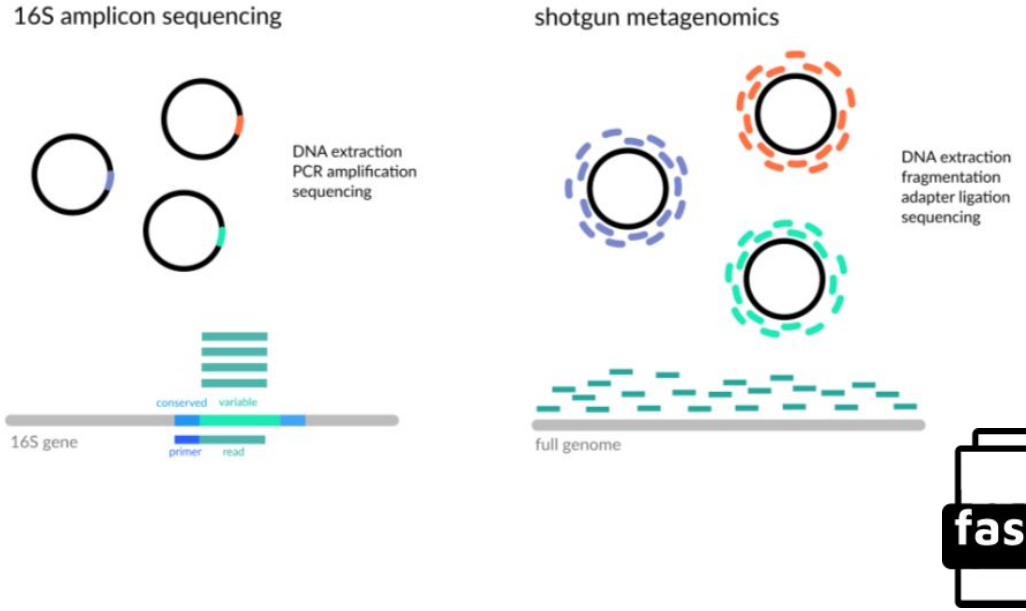

Talk is cheap. Show me the code.
— Linus Torvalds —
AZ QUOTES

Exploring the FASTQ file

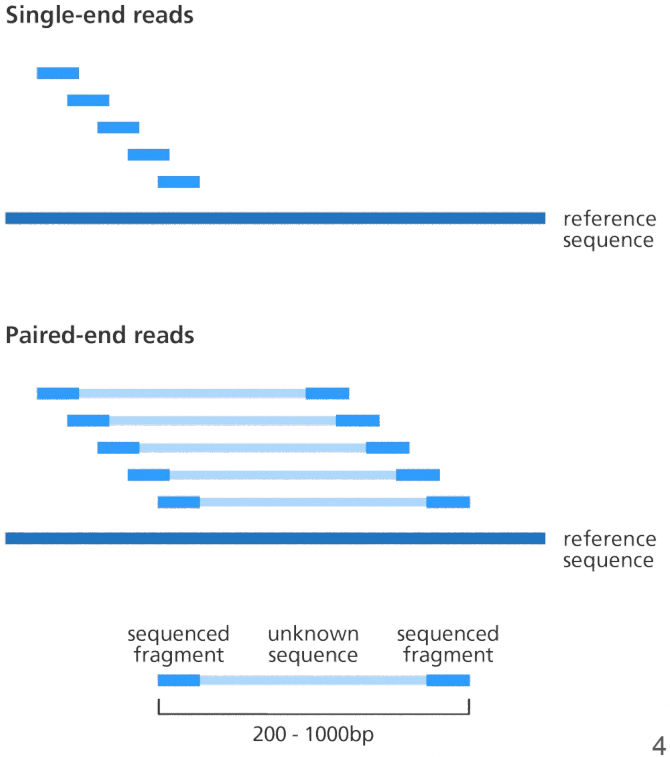
- Read checking (length, number of reads)
- Find the sequencing depth per sample

Short introduction of FASTQ file

The read (sequence)

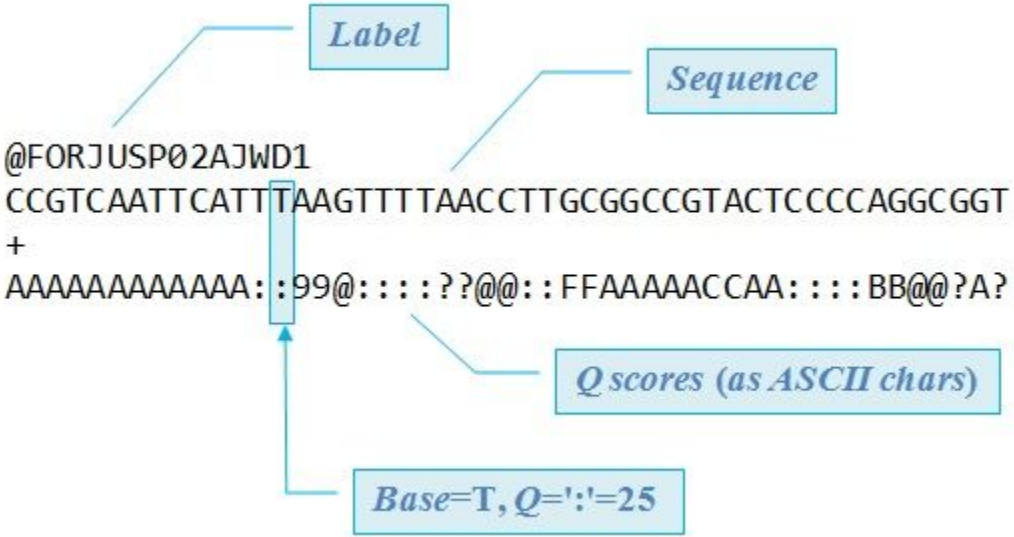


Short reads



https://gibbons-lab.github.io/isb_course_2020/16S/

Short introduction of FASTQ file



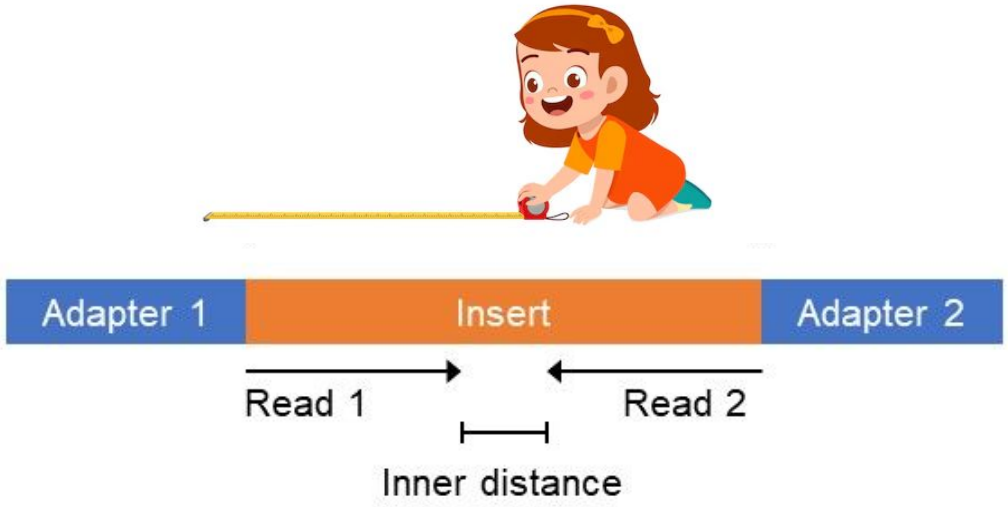
File Format

★ *What is a fastq file (Illumina)?*

```
@<title and optional description>
<sequence line>
+<optional repeat of title line>
<quality line>
```

Given a fastq file, what should we do?

Question 1: What is the length of our read?



zcat

length

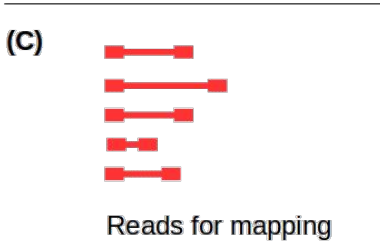
awk

uniq

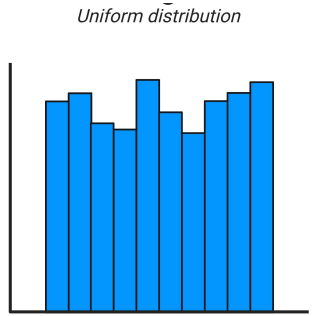
sort

```
zcat M6D364_S191_L001_R2_001.fastq.gz | awk
'(NR%4==2) {print length($0)}' | sort | uniq
```

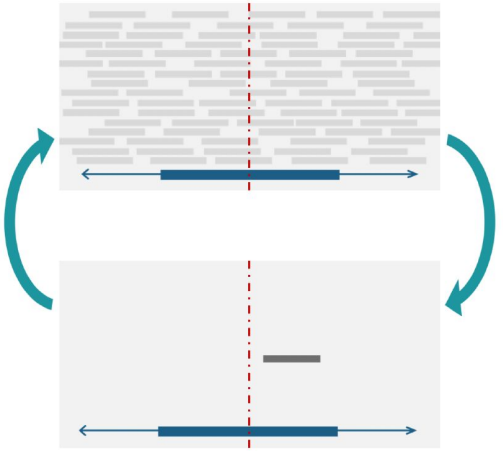
Question 2: Do our samples have good sequencing depths?



Sequencing depth = total read number



What we hope…



But…

→ Count the number of reads within a fastq file

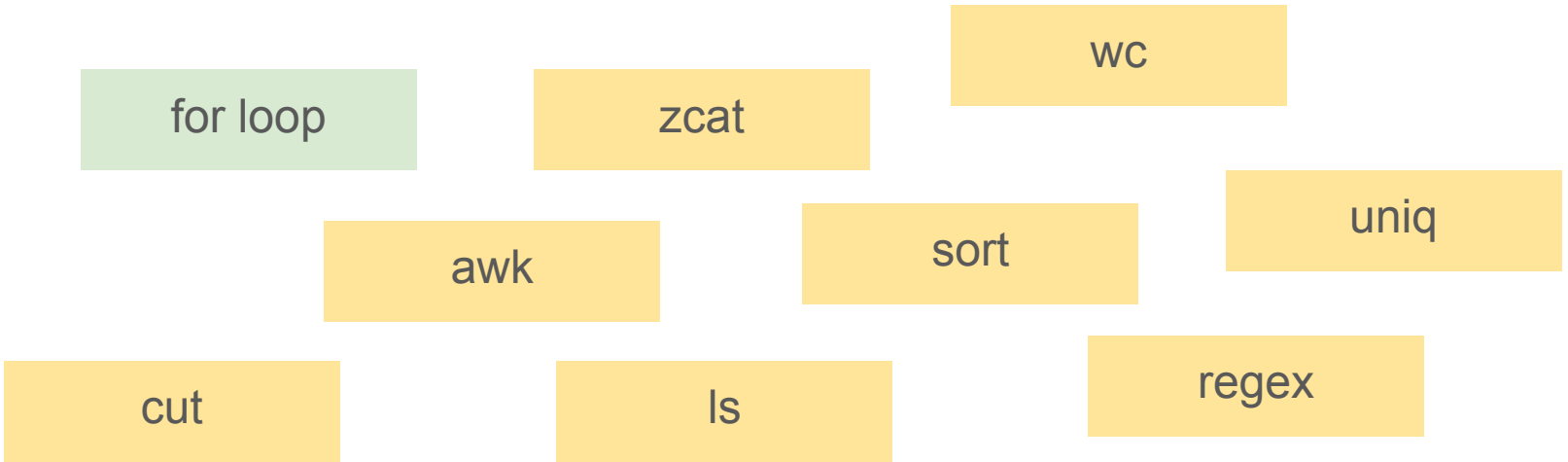
https://www.biorender.com/template/histogram-uniform-distribution

https://www.cancer.gov/ccg/blog/2019/low-coverage-seq

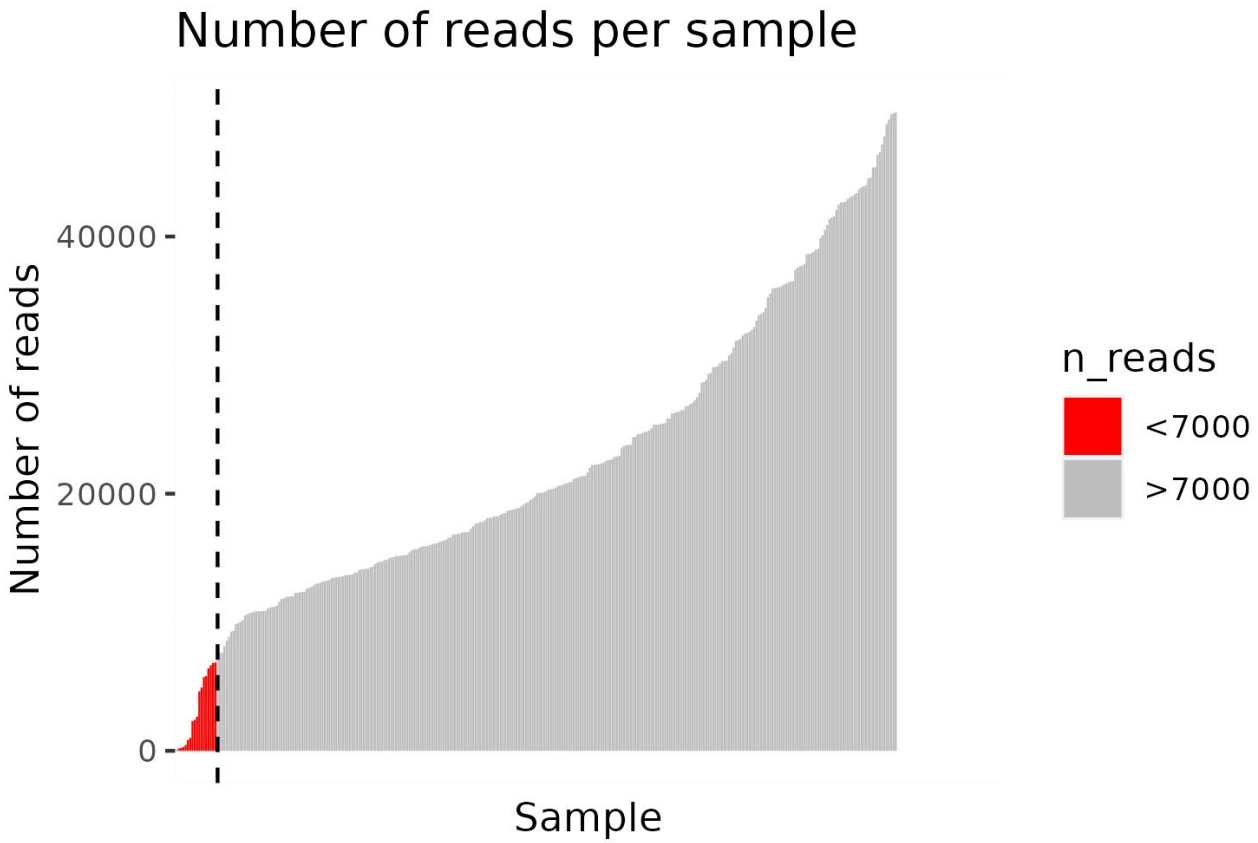
Question 2: Do our samples have good sequencing depths?

→ Count the number of reads within a fastq file

```
for sample in `ls *.gz | cut -d "_" -f1 | uniq`; do
   n_read=`zcat ${sample}_*.gz | grep "^@M" | wc -l`;
   echo -e $sample '\t' $n_read;
done > n_read_per_sample.tsv
```



Question 2: Do our samples have good sequencing depths?



Low sequencing depth. Why?

- Library preparation issues
- Sample quality issues
- Sequencing issues
- …

SUMMARY

Advantage of Bash scripting:

1. Automating Workflows

2. Lightweight

3. Working with command-line tools

4. Scripting for System Tasks (system administration, file manipulation, process automation)

Disadvantages:

1. Limited Data Structures

2. Error Handling

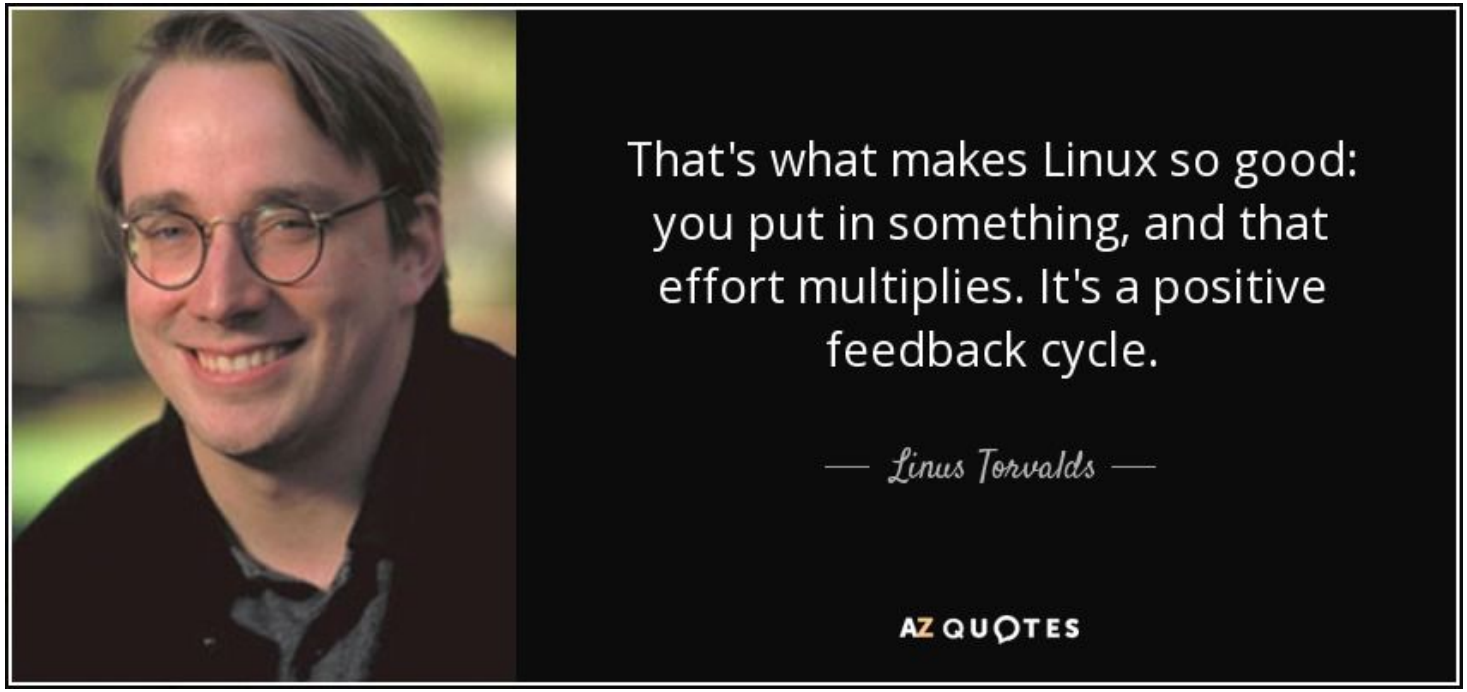
3. Less Versatile for Application Development

SUMMARY

Get our hands dirty with…

https://rosalind.info/problems/list-view/

SUMMARY


That's what makes Linux so good: you put in something, and that effort multiplies. It's a positive feedback cycle.
— Linus Torvalds —
AZ QUOTES