GC%

- GC content for most eukaryotes: 40% - 50%
- GC content for bacterial genome: <15% - >75%

Per Base Sequence Quality

- X-axis: position indexes
- Y-axis: Phred quality scores
- Blue line represents the mean quality scores
- The quality scores on the:
  - ✔ green region (Q > 28) are very good
  - ✔ orange region (20 < Q <28) are reason-able
  - ✔ red region (Q < 20) are poor

Per Base Sequence Content

- X-axis: position indexes
- Y-axis: base percentage
- Higher percentage of some bases at the beginning of the x-axis may indicate contaminating remnants of adaptor sequences or other contaminating sequences.
- If the difference between any of the four bases in any position:
  - ✔ greater than 10%: a warning message
  - ✔ greater than 20%: failure of this metric

Per Sequence GC Content

- X-axis: mean GC percentage per read
- Y-axis: number of reads
- A warning sign is displayed if the observed distribution deviates from normal distribution by a sum of more than 15% of the reads.
- A failure sign will be displayed if the distribution deviates by a sum of more than 30% of reads.

Per Base N Content

- X-axis: position indexes
- Y-axis: N base percentages
- A warning is issued if any position shows an N content of greater than 5% and a failure sign if any position shows an N content of greater than 20%.

Sequence Length Distribution

- X-axis: sequence length
- Y-axis: frequency
- A warning is displayed if the reads do not have the same length.

Overrepresented Sequences

- A warning will be issued if a sequence is overrepresented more than 0.1% of the total
- A failure will occur if the overrepresentation is more than 1% of the total.

Adapter Content

- X-axis: position indexes
- Y-axis: frequency
- A warning is raised if any sequence is present in more than 5% of all reads
- A failure occurs if any sequence is present in more than 10% of all reads.

FastQC Command

- fastqc filename.fastq.gz -o output_directory

e.g., fastqc SRR19551358.fastq.gz -o output/

Trimmomatic command

- TrimmomaticSE -threads 4 SRR19551358.fastq.gz output/trimming/SRR19551358_trimmed.fastq.gz ILLUMINACLIP:adapters.fa:2:30:10 LEADING:28 TRAILING:28 HEADCROP:15 CROP:172 SLIDINGWINDOW:4:15 MINLEN:36

Hisat2 command

- hisat2-build SRR19551358.fasta output/mapping/genome_index
- hisat2 -p 4 -x output/mapping/genome_index -U output/trimming/SRR19551358_trimmed.fastq.gz -S output/mapping/SRR19551358_aligned.sam --summary-file output/mapping/alignment_summary.txt