Computational Genomics

B V SOMA ADITHYA

NGS DATA ANALYSIS

Galaxy Details:

URL: https://usegalaxy.eu/

(European)

Step 1:

SRA ID: DRR319902 (SRA Paired)

Step 2:

spots read   : 12,001,829

reads read   : 24,003,658

reads written   : 24,003,658

Step 3:

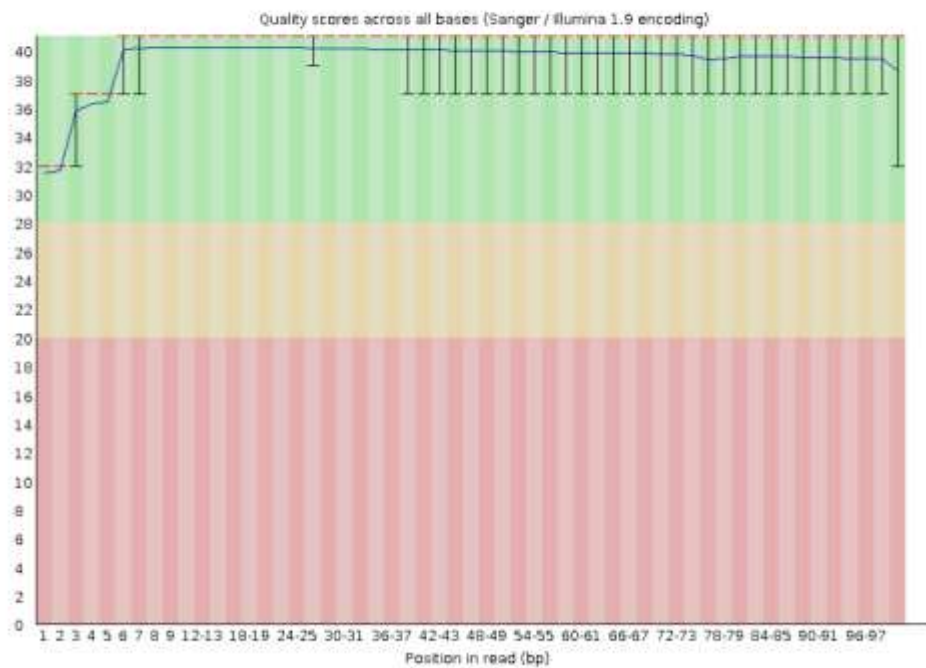Quality Control (Before Trimming)

FORWARD

- Basic Statistics

Measure	Value
Filename	DRR319902_forward.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	12001829
Total Bases	1.2 Gbp
Sequences flagged as poor quality	0
Sequence length	100
%GC	43

- Per base sequence quality

- Overrepresented sequences

- Adapter Content

Adapter Content

REVERSE

- Basic Statistics

Basic Statistics

Measure	Value
Filename	DRR319902_reverse.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	12001829
Total Bases	1.2 Gbp
Sequences flagged as poor quality	0
Sequence length	100
%GC	43

- Per base sequence quality

- Overrepresented sequences

- Adapter Content

Step 4:

I trimmed the raw reads accordingly. This can be seen in my galaxy account.

Trimmomatic, Trim Galore, Cutadapt tools used for trimming. Parameters are set accordingly.
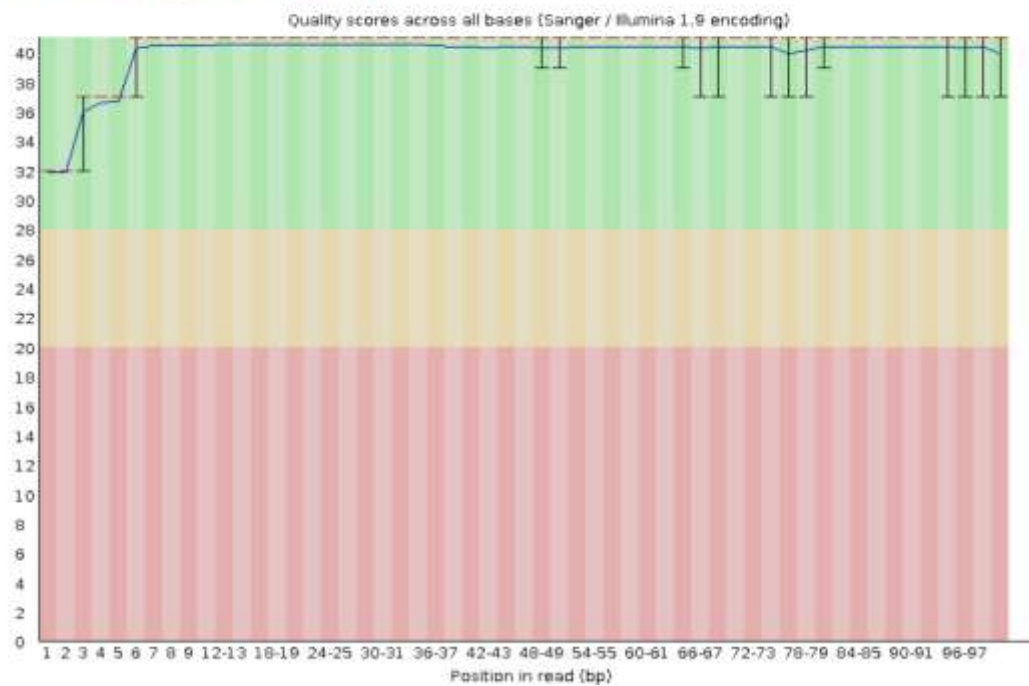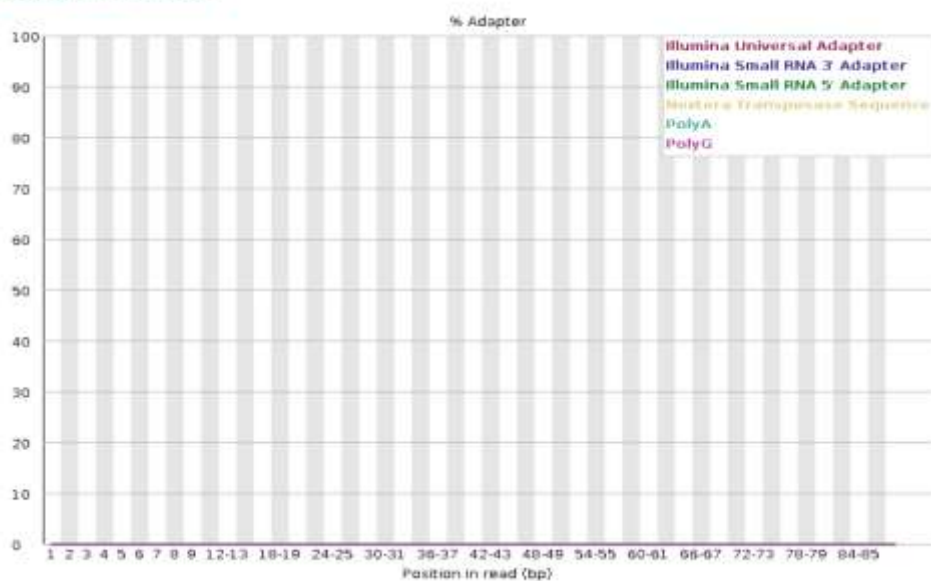
Step 5:

Quality Control (After trimming)

1st

- Per base sequence quality

- Adapter Content

2 nd

- Per base sequence quality

- Adapter Content

Step 5:

Before Trimming:-

Total Sequences : 12001829

Total Bases   : 1.2 Gbp

Sequence Length : 100

After Trimming:

Total Sequences : 10485186

Total Bases   : 912.3 Mbp

Sequence Length : 3-100

Step 6:

Generated BAM File.

Step 7:

Samtools Flagstat tool used for summary.

```
20970372 + 0 in total (QC-passed reads + QC-failed reads)
20970372 + 0 primary
0 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
20557168 + 0 mapped (98.03% : N/A)
20557168 + 0 primary mapped (98.03% : N/A)
20970372 + 0 paired in sequencing
10485186 + 0 read1
10485186 + 0 read2
19900090 + 0 properly paired (94.90% : N/A)
20454318 + 0 with itself and mate mapped
102850 + 0 singletons (0.49% : N/A)
149824 + 0 with mate mapped to a different chr
55158 + 0 with mate mapped to a different chr (mapQ>=5)
```

Mapped Reads = 20557168

Unmapped Reads = 413204

Tools used throughout this assignment :

Faster Download and extract reads, FastQC, Trimmomatic, Trim Galore, Cutadapt, Bowtie2, samtools Flagstat.