

Part 1 Demultiplexing | BIO 622

Jared Galloway

Part 1

1.

Filename	Read
1294 _S 1 _L 008 _R 1 ₀ 01. <i>fastq.gz</i>	Read 1
1294 _S 1 _L 008 _R 2 ₀ 01. <i>fastq.gz</i>	Index 1
1294 _S 1 _L 008 _R 3 ₀ 01. <i>fastq.gz</i>	Index 2
1294 _S 1 _L 008 _R 4 ₀ 01. <i>fastq.gz</i>	Read 2

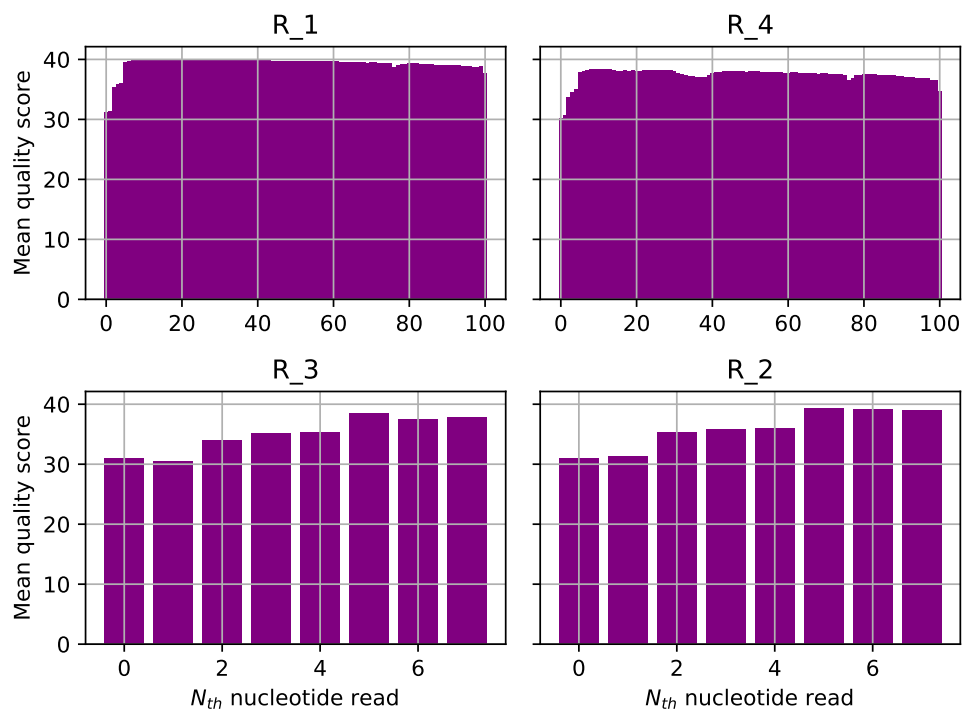


Figure 1: Distribution of average quality score R1

Due to the first two biological reads – for both R1 and R2 – are quite low $\approx < 36$, I would set the cutoff for 36, to make sure we are getting high quality biological reads.

To find the total number of N's within the two index reads, I used the following bash command

```
zcat emp_files/1294_S1_L008_R[2-3]* | awk 'NR%4==2' | grep N | wc -l > countN_Index.txt
```

Part 2