

GGG 201B: Assignment 1

Gitanshu Munjal

2/22/2015

1. What quality type are these reads, phred+33 or phred+64? How can you tell? (be specific)

These reads are phred+33 type. I found #s at the ends of reads as an indicator and used the python program at <https://github.com/brentp/bio-playground/blob/master/reads-utils/guess-encoding.py> to confirm.

2. Use a unix command to get only the first 250,000 fastq reads and write them to a file. What command would this be?

```
head -n 1000000 Rice1.fq > Rice1reduced.fq
```

We use 1,000,000(= 250,000 * 4) because each read is made up of 4 lines.

3. How many reads were rejected?

15733

5. How many well-mapped reads were found in gene LOC_Os01g64330?

15