Basic statistics

Basic Statistics

Measure	Value
Filename	sample_trimmed.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	197343
Sequences flagged as poor quality	0
Sequence length	18-85
%GC	63

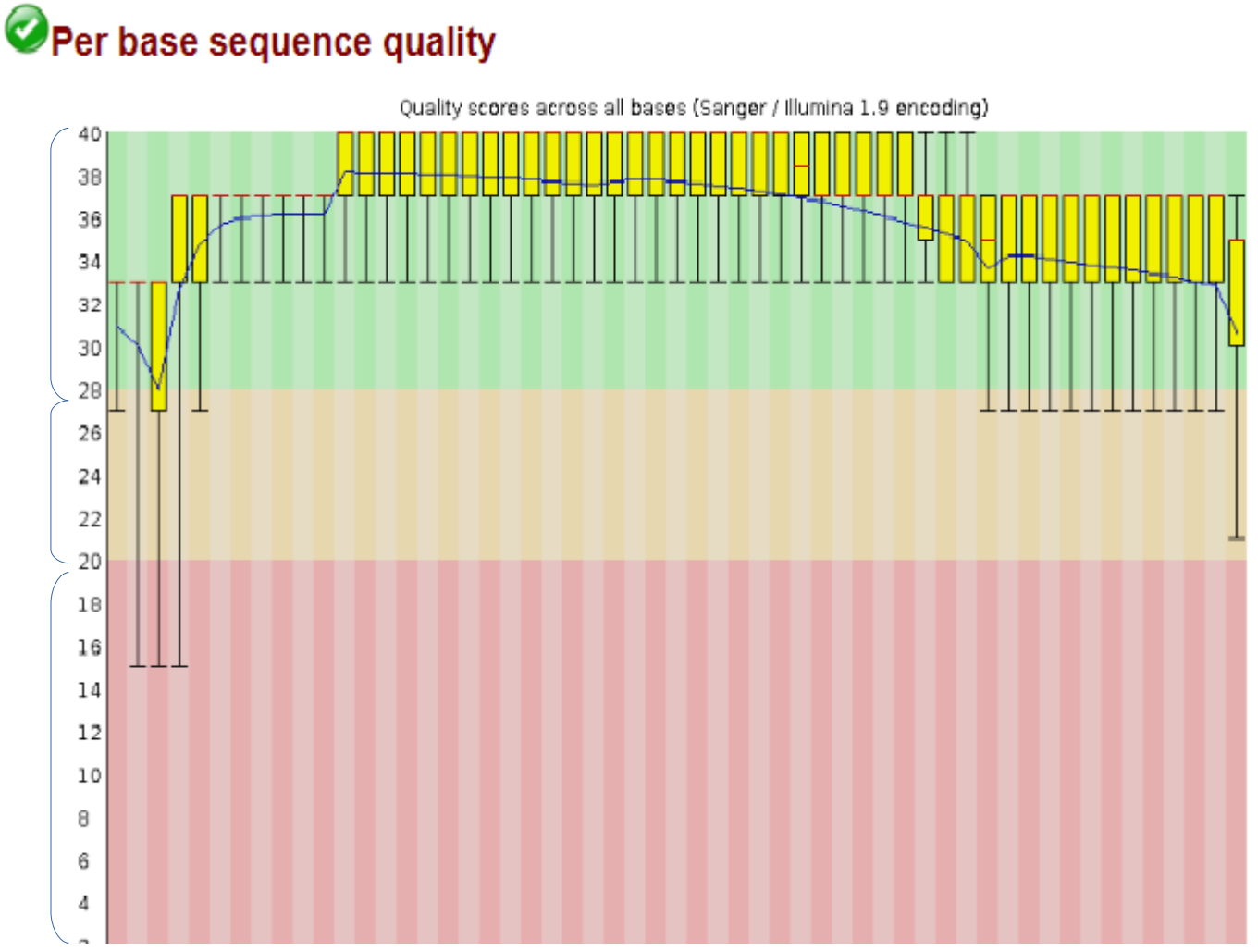
Summary of the fastq file: total reads; fastq version; total number of reads; length; GC content

Base quality distributions per position

Green: good/ great quality

Orange: Acceptable quality

Red: Poor quality



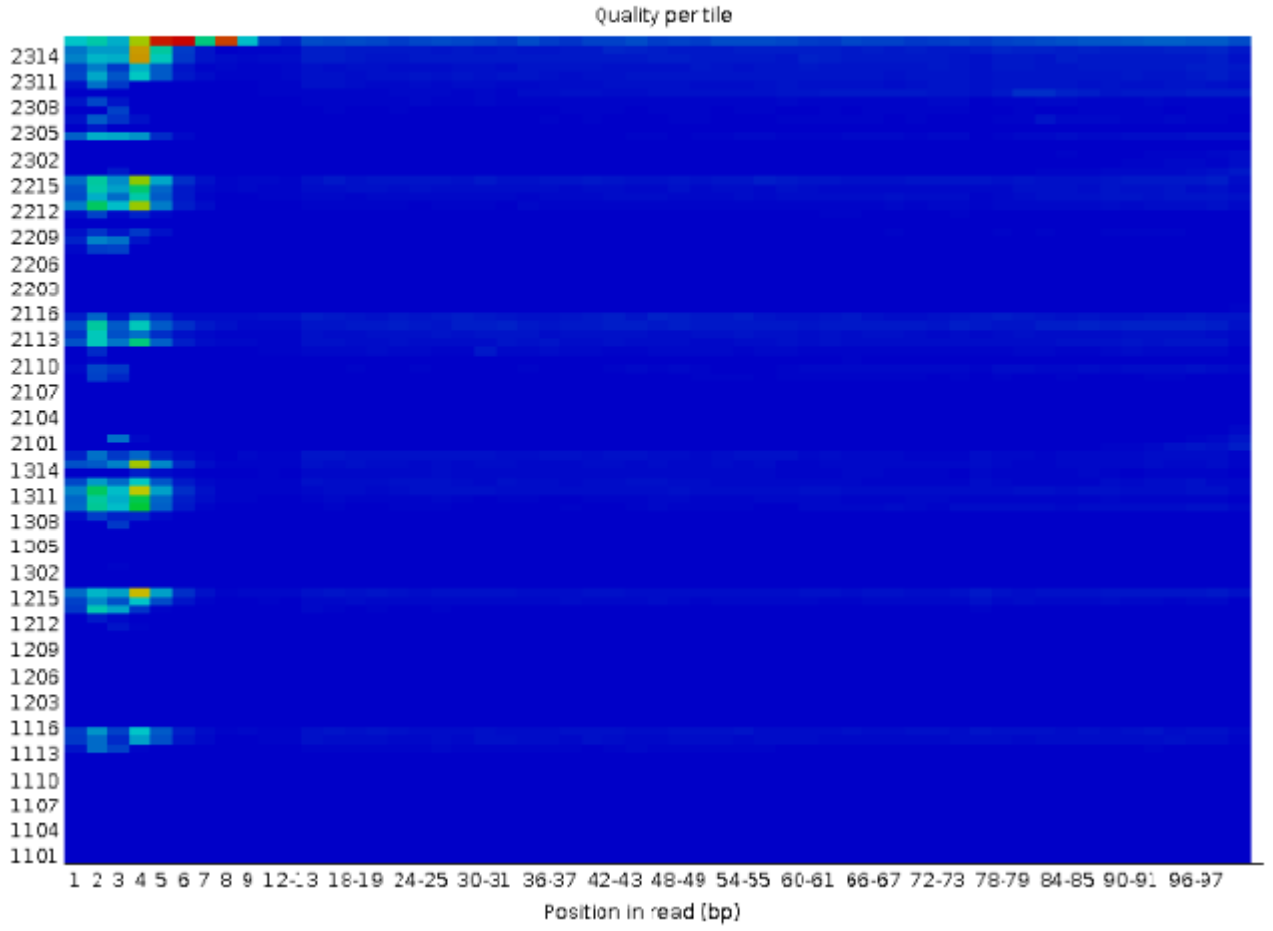
Qual = 30 – 1/1000 error probablity

Qual = 20 – 1/100

Qual = 10 - 1/10

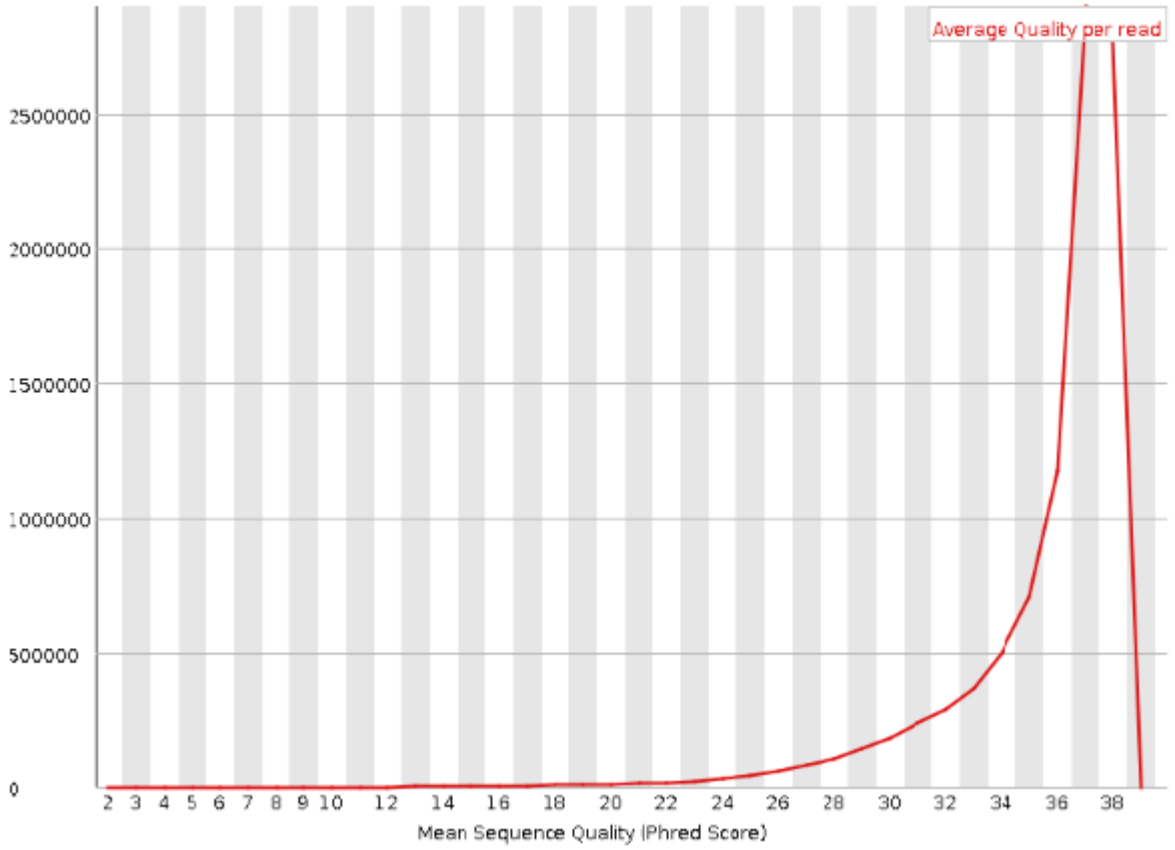
Base qualities per tile

Per tile sequence quality



A number of tiles were under-performing, but to the extent that seriously affect quality of the run.

Distribution of average read qualities (sum of base qualities divided by length of the read)

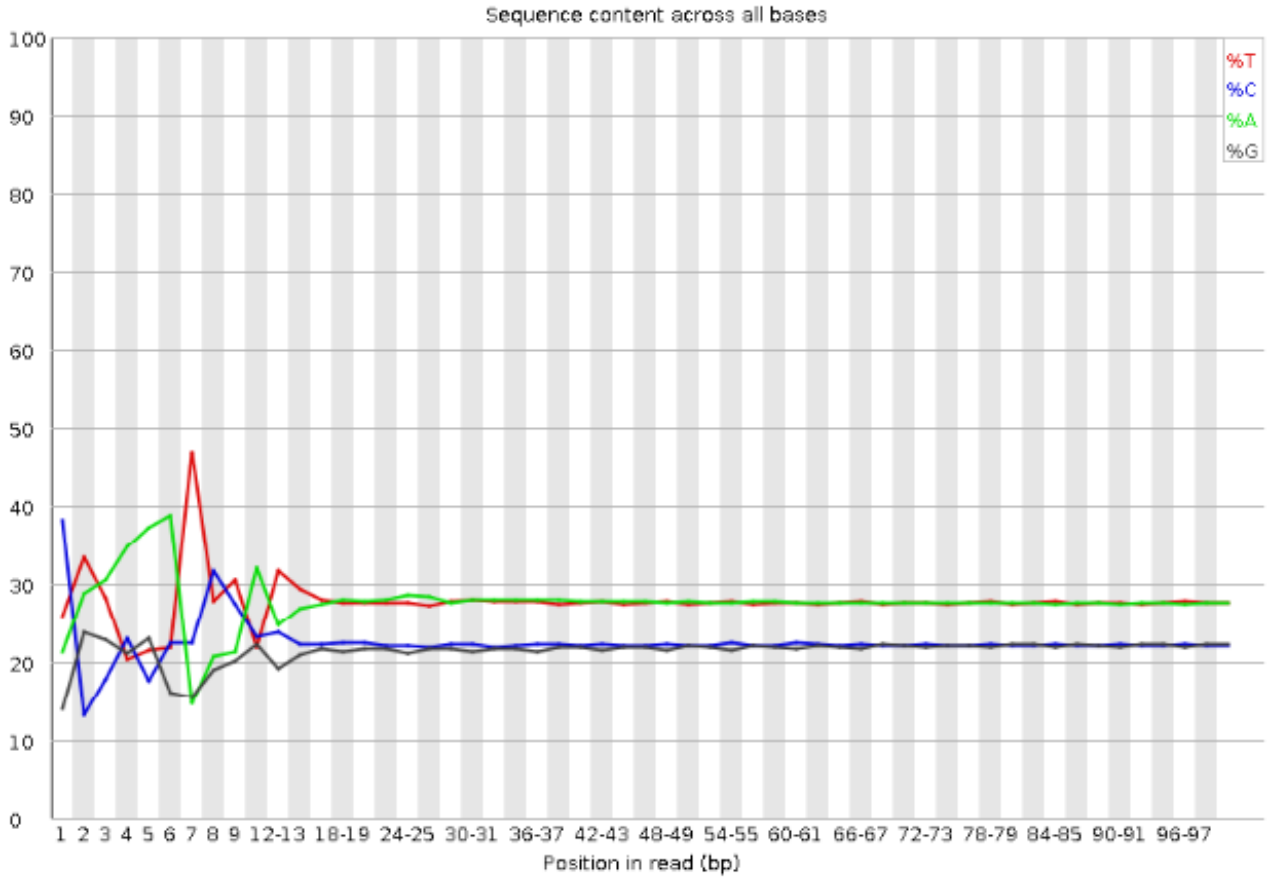


The majority of the reads have mean quality between 32 – 35

Good base qualities

Base proportions per position across all reads

Per base sequence content

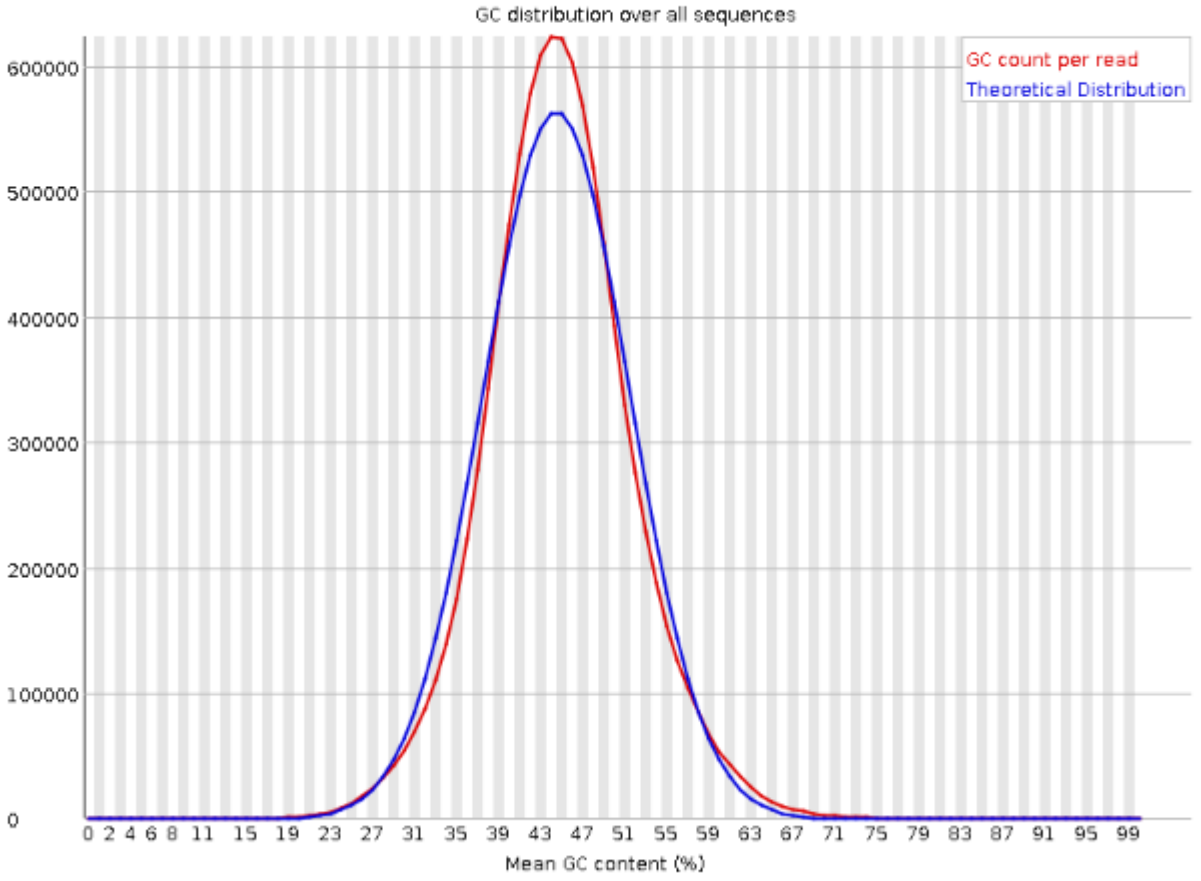


Expected to be about 25% for each base in the case of human genome.

Deviates from the expected values in other genomes, mRNAs, smallRNAs, bisulfite, etc.

The sequence content bias at 5’ end is the result of ligation and amplification.

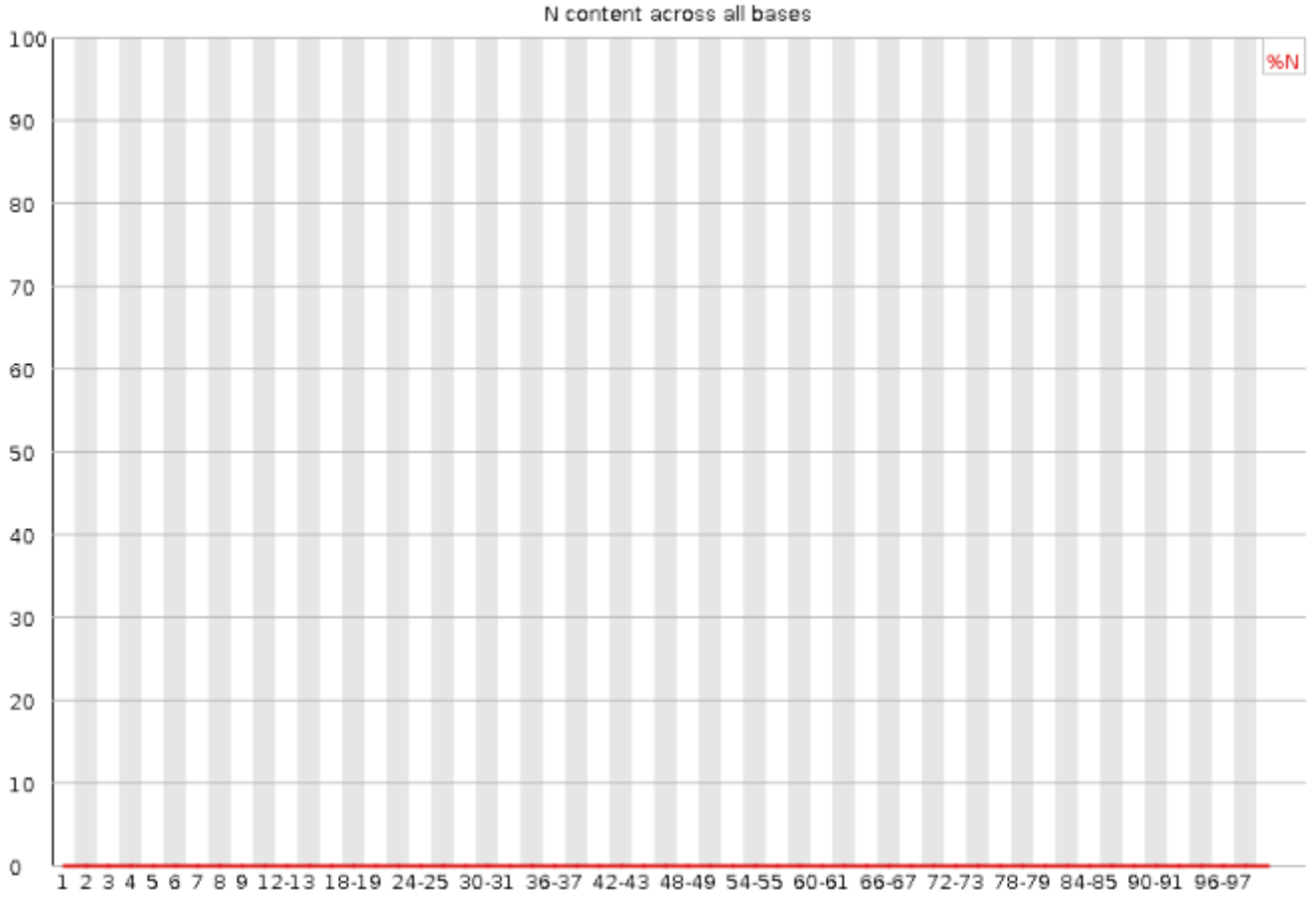
Per sequence GC content



Slightly different from theoretical, which was calculated based on Human genome.

There are many cases when this parameter will deviate from theoretical (different organism, mRNA, smallRNA, adapter contamination, etc.).

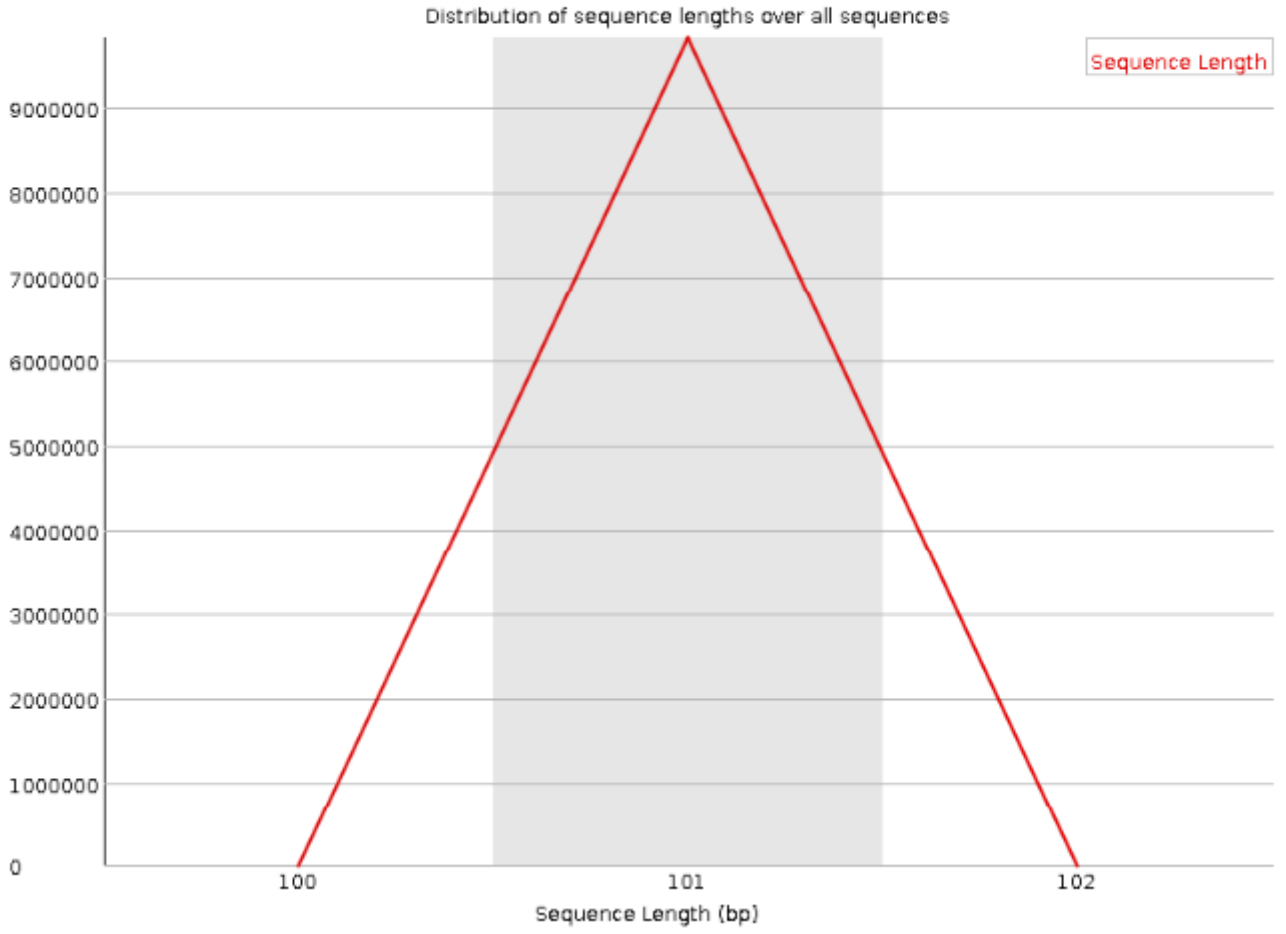
Per base N content



When the base cannot be identified it is replaced with **N**

In this case no N bases were found

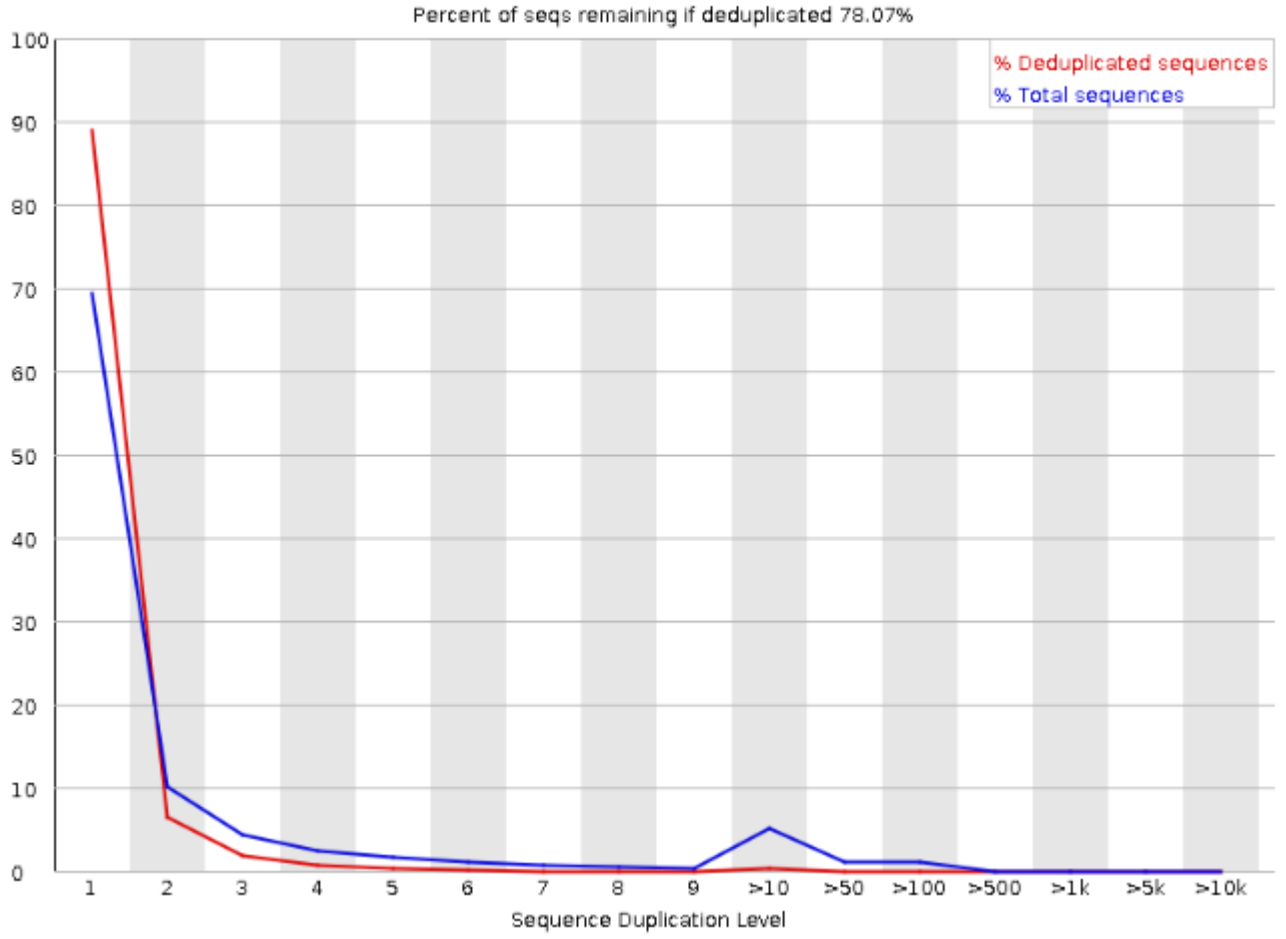
Sequence Length Distribution



All of the sequences are the same length, since they were not manipulated in any way.

The length profile will change after the adapter trimming

Sequence Duplication Levels

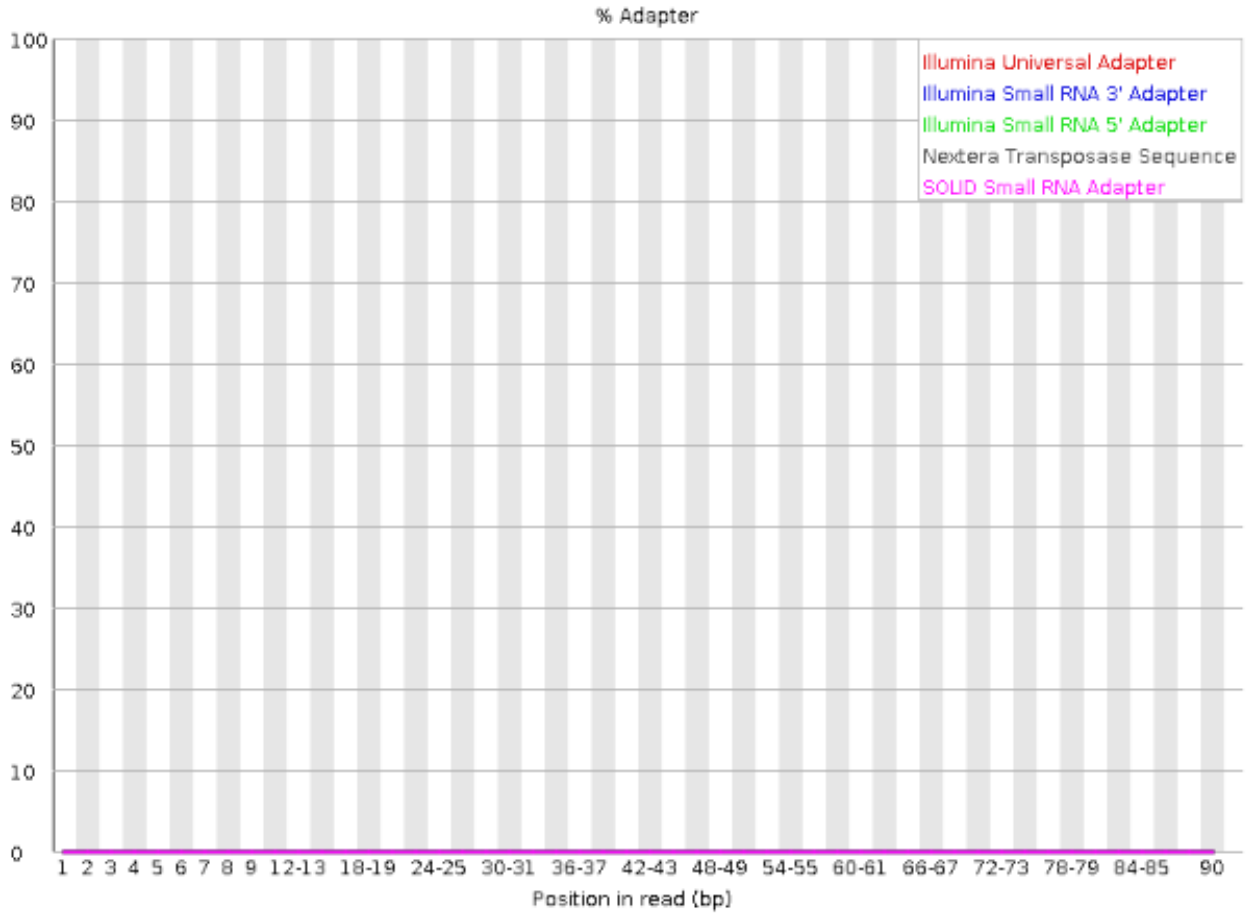


The frequency of duplicated sequences.

Sequence duplicates may arise as PCR and optical artifacts.

Typically, we do not remove duplicate sequences in RNA-seq projects since they may represent genuine transcripts from short or highly expressed genes.

Adapter Content



FastQC detects Illumina adapters automatically.

No adapters are found in this case.