

FASTQ-Analiza podataka

BASIC STATISTICS

Encoding: Sanger/Illumina 1.9

Total sequences: 4504104

Sequence length: 75-151

Per base sequence quality

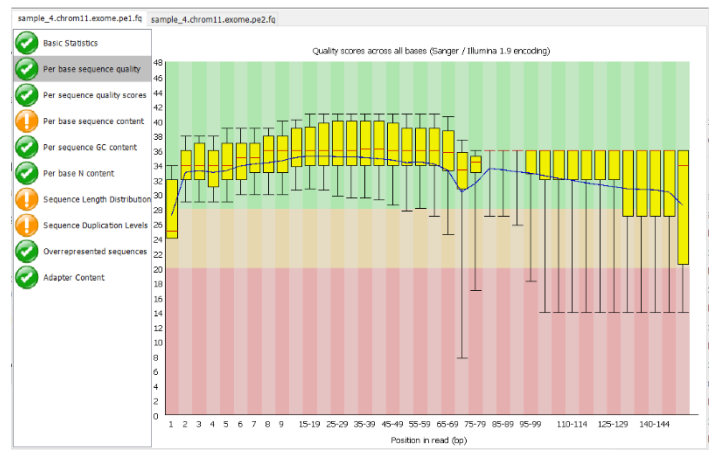
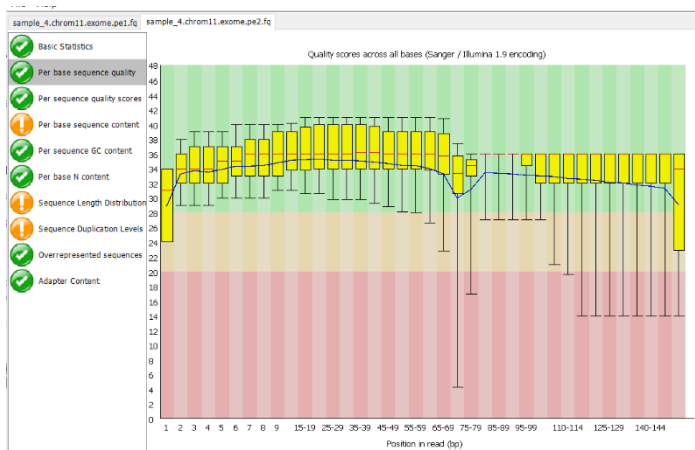
x-osa: 150 baznih parova

zuta granica je od 25. Do 75. Percentile

crvena linija: medium value

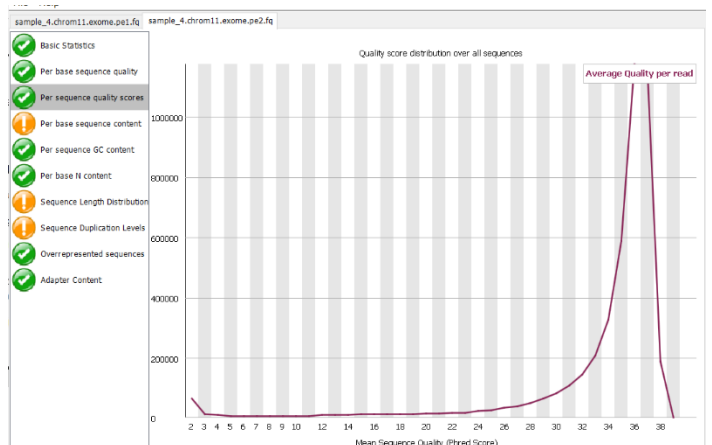
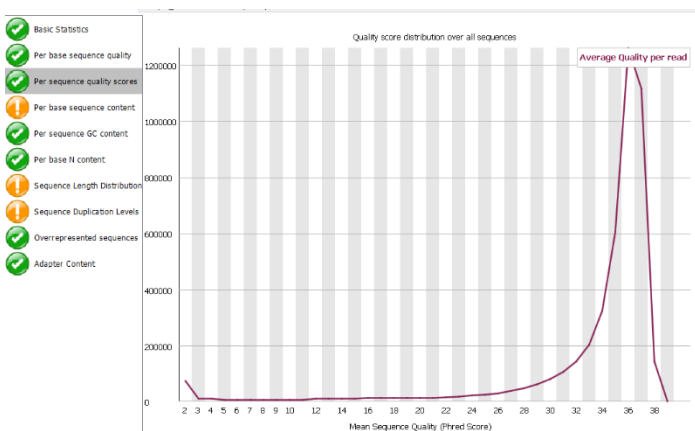
crna: 10-90 percentila

blue: mean quality score



Per sequence quality scores

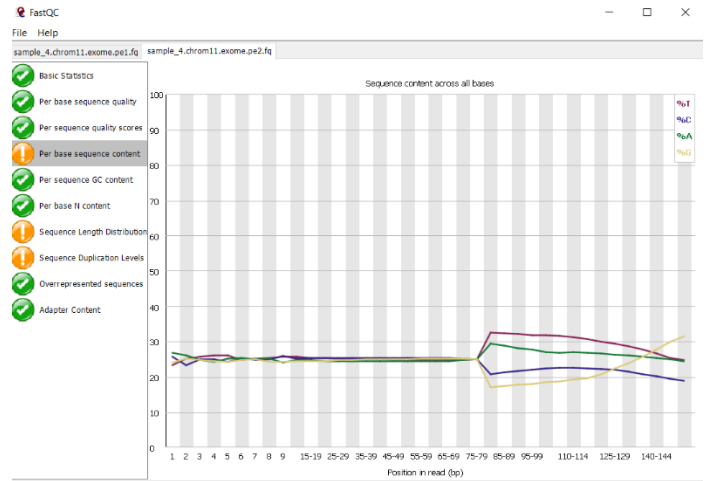
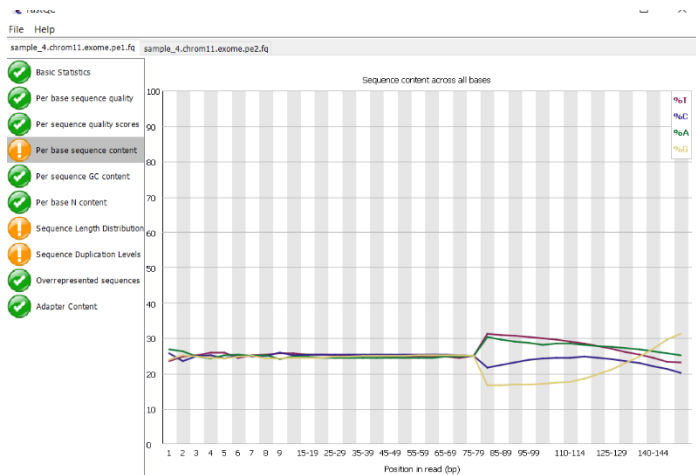
Mean score across all the basis in that particular sequence- ovaj je dobar



Per base sequence content

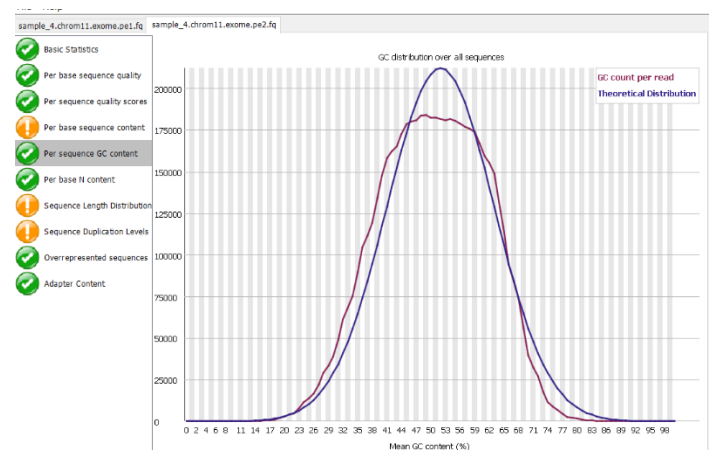
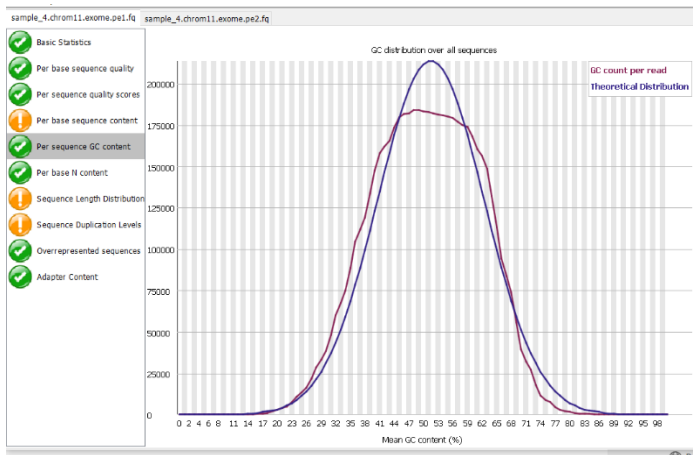
Ako se oslonimo I na sequence length distribution I na basic statistics sequence length

U 25% se “pojavljuju” u base pair koje ni nemamo, dok u 75bp nisu uopste jednako prisutne sve ATCG baze.



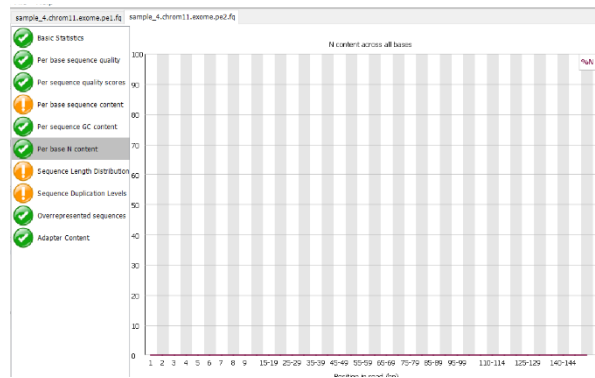
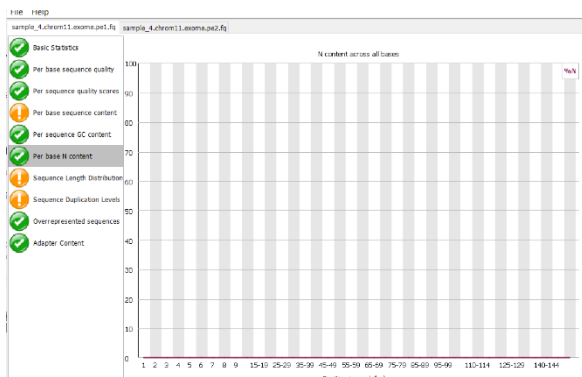
Per sequence GC content

The Per sequence GC content graph gives an idea of contamination in present. Optimalno 50%-55%



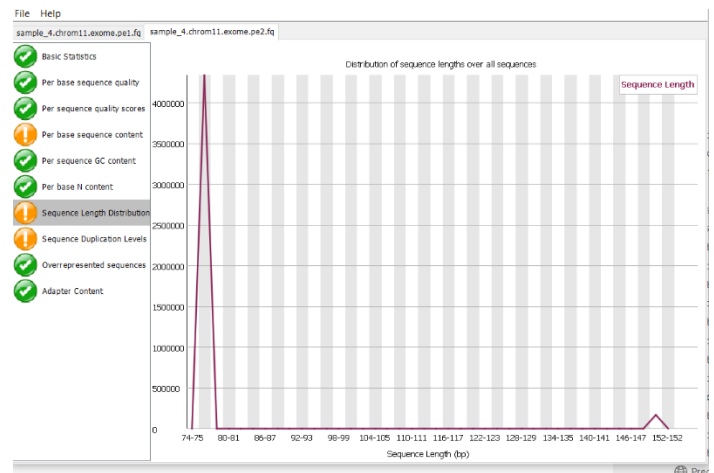
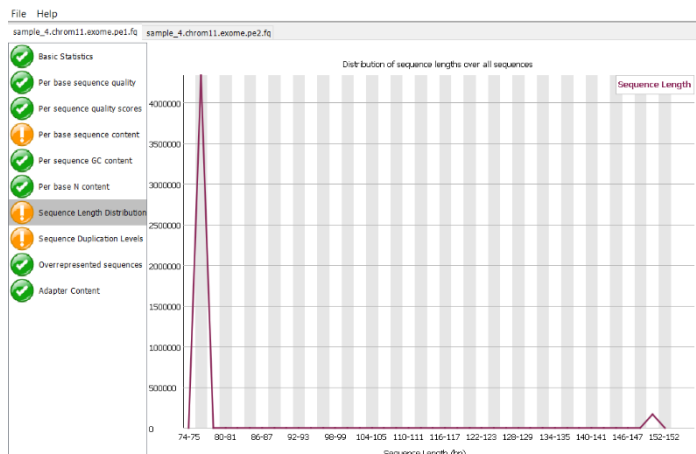
Per base N content

Ako ne moze da nadje bazu zameni je sa N



Sequence length distribution

Nema base pair pre74-75 base pair, njih ima najvise, zatim 152, dakle nije uniformna raspodela



Sequence duplication levels

Izmedju 35%-45% su predstavljeni jednom. 20%-35% su predstavljeni 2x. 10%-20% su predstavljeni 3x

5-10% 4x. 66% je procentat sequences koji ostaje kada se izbace duplikati za pe1. 65% je za pe2.

