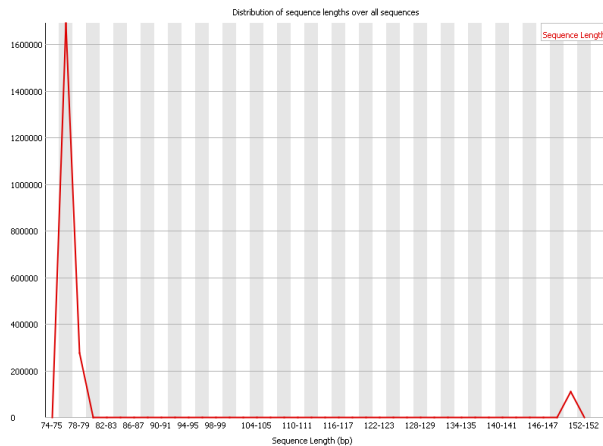


# Bioinformatika projekat 2

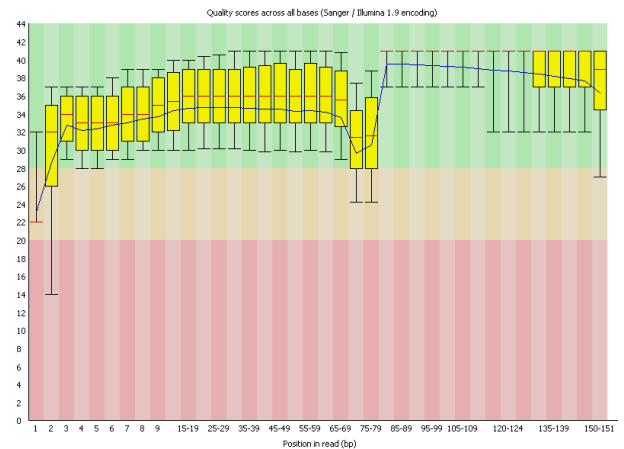
## 1. Upotrebom FastQC alata, dolazimo do sledećih problematičnih kriterijuma (warning):

Pe1:

### Sequence Length Distribution

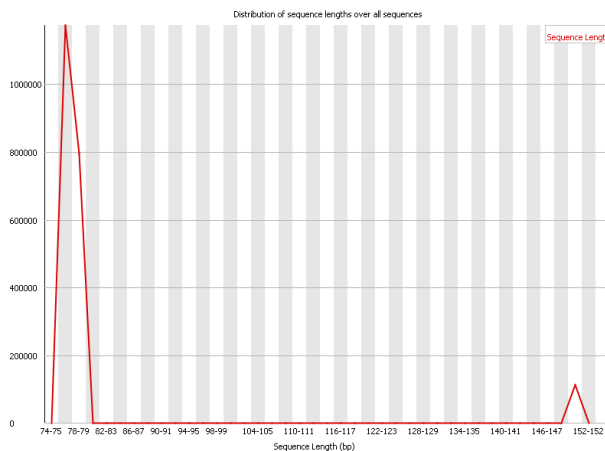


### Per base sequence quality

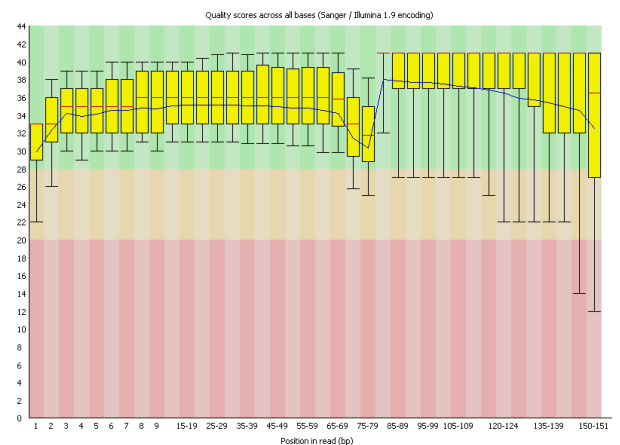


Pe2:

### Sequence Length Distribution

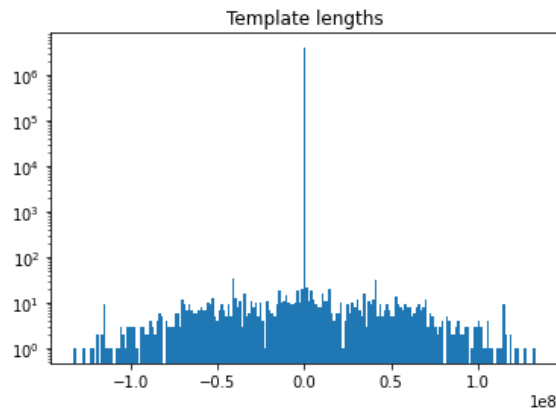


### Per base sequence quality



- Per base sequence quality – Pri početku, prve baze u oba slučaja kreću sa nižom vrednošću kvaliteta, kod prvog pravi problem i drop kod baza 70-79.
  - Sequence Length Distribution – kod baza 74-80 imamo različite dužine nizova, zbog toga imamo upozorenje
2. Nakon mapiranja uzoraka na referentni genom dobijamo 3 913 079 mapiranih i 25695 nemapiranih read-ova. Od toga 3 887 368 readova su uparena (1 943 684 para).

Histogram dužina sekvenciranih fragmenata:



3. Nakon markiranja duplikata korišćenjem Picard alata, dobijamo sledeći izveštaj:

| LIBRARY           | UNPAIRED_READS_EXAMINED | READ_PAIRS_EXAMINED | SECONDARY_OR_SUPPLEMENTARY_RDS | UNMAPPED_READS | UNPAIRED_READ_DUPLICATES |
|-------------------|-------------------------|---------------------|--------------------------------|----------------|--------------------------|
| 0 Unknown Library | 25695                   | 1942489             | 2406                           | 250893         | 15009                    |

| READ_PAIR_DUPLICATES | READ_PAIR_OPTICAL_DUPLICATES | PERCENT_DUPLICATION | ESTIMATED_LIBRARY_SIZE |
|----------------------|------------------------------|---------------------|------------------------|
| 263630               | 0                            | 0.138664            | 6493052                |

Pronađeno je ukupno 542269 duplikata, od toga 263630 parova duplikata i 15009 duplikata bez para. Optičkih duplikata nema, što znači da su 100% PCR duplikati.

4. Upotrebom alata Haplotype Caller identifikovano je 5767 mutacija, od toga 5235 SNP-ova i 532 INDEL-a. Kriterijume filtriranja prolaze 5577 mutacija, 190 ne prolazi. Ti/Tv odnos pre filtriranja je približno 2,2227692 (Ti : 3612, Tv : 1625), a posle filtriranja približno 2,2265475 (Ti : 3489, Tv : 1567).

5. Alatom Funcotator nakon anotiranja podataka, dobijaju se sledeće anotacije, i broj njihovog ponavljanja:

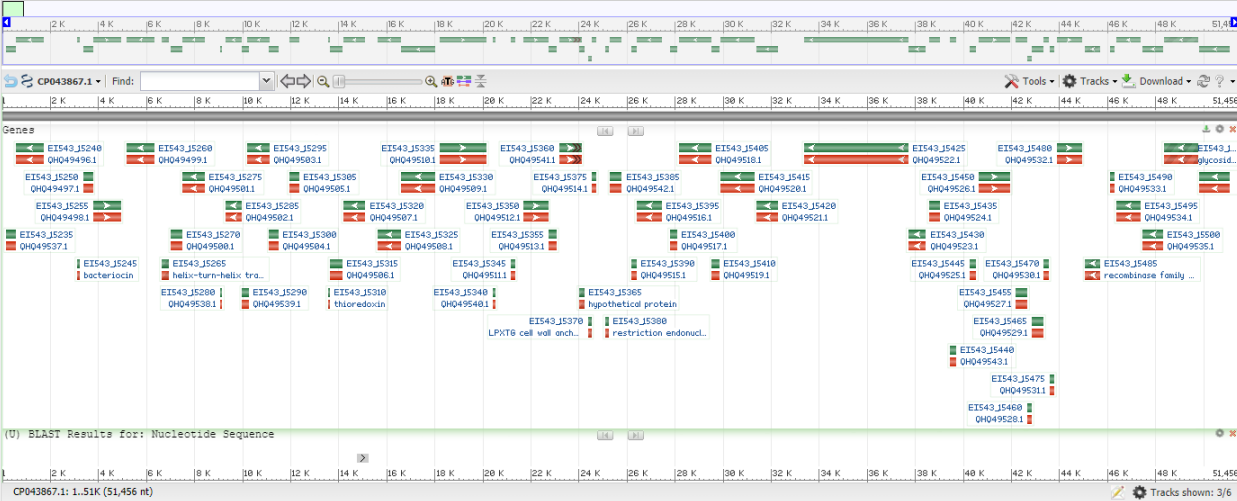
```
Benign 235
Benign/Likely_benign 47
not_provided 4
Likely_benign 23
Uncertain_significance 1
drug_response 1
```

6. Uz pomoć alata Blast, identifikovan je organizam od kog potiče najduži skafold. Detaljni podaci o organizmu su prikazani na sledećoj strani.

Enterococcus faecium strain ME3 plasmid unnamed2

GenBank: CP043867.1  
GenBank FASTA

[Link To This View](#) | [Feedback](#)



Enterococcus faecium strain ME3 plasmid unnamed2

Sequence ID: [CP043867.1](#) Length: 51456 Number of Matches: 1

Range 1: 14762 to 15227 [GenBank](#) [Graphics](#)

[▼ Next Match](#) [▲ Pre](#)

| Score         | Expect  | Identities    | Gaps      | Strand    |
|---------------|---|---------------|-----------|-----------|
| 861 bits(466) | 0.0   | 466/466(100%) | 0/466(0%) | Plus/Plus |
| Query 1       | TAAAAAGTCCTTTAAATTCTGGATTAACTGAAATTATAACAAGTTGTTTATCTCGTATAA    | 60            |           |           |
| Sbjct 14762   | TAAAAAGTCCTTTAAATTCTGGATTAACTGAAATTATAACAAGTTGTTTATCTCGTATAA    | 14821         |           |           |
| Query 61      | TGCTATATTTTGTGAAAGCAACAAATTTAGTTAAAGTATTTCCATCATCATTTCCAAGCAT   | 120           |           |           |
| Sbjct 14822   | TGCTATATTTTGTGAAAGCAACAAATTTAGTTAAAGTATTTCCATCATCATTTCCAAGCAT   | 14881         |           |           |
| Query 121     | TAATACTTAATAATTTGTCATACGTGCTACTAAGATCTTTTATAAAATTGTTCCCCACGTT   | 180           |           |           |
| Sbjct 14882   | TAATACTTAATAATTTGTCATACGTGCTACTAAGATCTTTTATAAAATTGTTCCCCACGTT   | 14941         |           |           |
| Query 181     | CTGTATAACCGCTTAGTTTTTTCAAATCGTGAAATGAAAAAATATCTCATCTACACCTT     | 240           |           |           |
| Sbjct 14942   | CTGTATAACCGCTTAGTTTTTTCAAATCGTGAAATGAAAAAATATCTCATCTACACCTT     | 15001         |           |           |
| Query 241     | TTTCACGTACACGTGAAGCAATTGAAAAAATAAGTTCATTTACGACTATTTAAATTTTC     | 300           |           |           |
| Sbjct 15002   | TTTCACGTACACGTGAAGCAATTGAAAAAATAAGTTCATTTACGACTATTTAAATTTTC     | 15061         |           |           |
| Query 301     | TCAGAGGAATATGATTTAATCTATTTTCATATTTTACAATTTTCATTTGCCATAAGCTTCT   | 360           |           |           |
| Sbjct 15062   | TCAGAGGAATATGATTTAATCTATTTTCATATTTTACAATTTTCATTTGCCATAAGCTTCT   | 15121         |           |           |
| Query 361     | CCTACTTTTATAAAGTATGAAATAAAAGTAAACATACCTTTATAAAAAATACTCAATAATGTC | 420           |           |           |
| Sbjct 15122   | CCTACTTTTATAAAGTATGAAATAAAAGTAAACATACCTTTATAAAAAATACTCAATAATGTC | 15181         |           |           |
| Query 421     | GGAAAACCTACTTGAAAAGCGTAGAAAACCTACTTGAAAAGCGTAG                  | 466           |           |           |
| Sbjct 15182   | GGAAAACCTACTTGAAAAGCGTAGAAAACCTACTTGAAAAGCGTAG                  | 15227         |           |           |