1. Mapiranje readova

Izvor podataka: results/03_dedup/dedup_flagstat.txt

Komanda: cat results/03_dedup/dedup_flagstat.txt

- Ukupno readova: 6,063,198
- Mapiranih readova: 5,466,994 (≈ 90.17%)
- Properly paired: 5,398,440 (≈ 89.08%)

```
6063198 + 0 in total (QC-passed reads + QC-failed reads)
6060318 + 0 primary
0 + 0 secondary
2880 + 0 supplementary
193987 + 0 duplicates
193987 + 0 primary duplicates
5466994 + 0 mapped (90.17% : N/A)
5464114 + 0 primary mapped (90.16% : N/A)
6060318 + 0 paired in sequencing
3030159 + 0 read1
3030159 + 0 read2
5398440 + 0 properly paired (89.08% : N/A)
5434414 + 0 with itself and mate mapped
29700 + 0 singletons (0.49% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

2. Duplikati (PCR i optički)

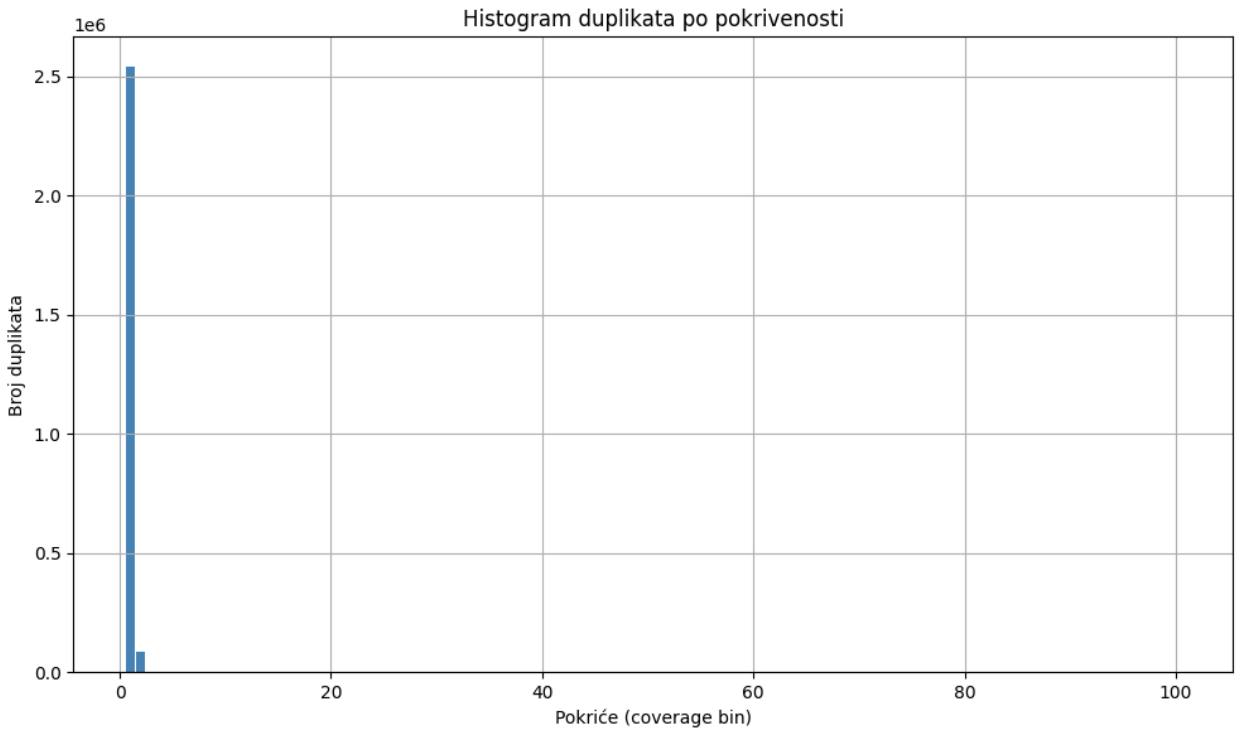
Izvor podataka : results/03_dedup/dup_metrics.txt

Komanda : cat ~/projekat2/results/03_dedup/dup_metrics.txt

PCR duplikata: 90,029
Optičkih duplikata: 0
Ukupan procenat duplikata: 3.55%

Na histogramu duplikata po pokrivenosti vidi se da se većina duplikata javlja pri nižim vrednostima pokrivenosti (bin = 1), što je očekivano za PCR duplikate. Histogram takođe potvrđuje da nema optičkih duplikata, jer su svi duplikati svrstani u opšti PCR region.

Vizualizacija histogram duplikata je urađena pomoću matplotlib u Colab-u (Histogrami.ipynb).



3. Broj mutacija (VCF analiza)

Izvor: filtered_variants.vcf.gz

a) Broj ukupno detektovanih mutacija

Komanda : bcftools view filtered_variants.vcf.gz | grep -v '^#' | wc -l

```
15452
```

b) Koliko je SNP-ova

Komanda : bcftools view -v snps filtered_variants.vcf.gz | grep -v '^#' | wc -l

```
13533
```

c) Koliko je INDEL-a

Komanda : bcftools view -v indels filtered_variants.vcf.gz | grep -v '^#' | wc -l

```
1915
```

4. Histogram dužina sekvenciranih fragmenata (template_length)

U ovom koraku analizirane su dužine fragmenata iz BAM fajla posle deduplikacije. Histogram prikazuje koliko puta se koja dužina pojavljuje.

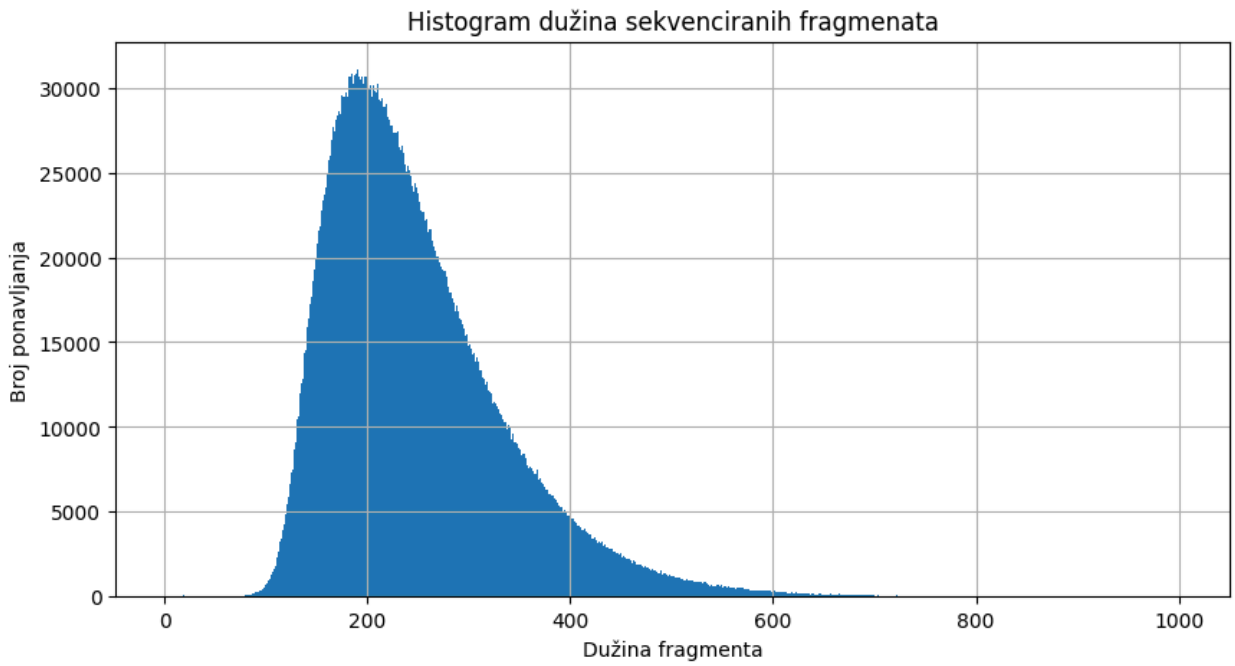
Komanda: head template_lengths.txt

```
179
179
258
258
247
247
313
190
190
313
```

Histogram prikazuje distribuciju dužina sekvenciranih fragmenata (template_length) dobijenih iz BAM fajla nakon deduplikacije.

Najveći broj fragmenata ima dužinu između 150 i 400 baza (bp), sa najvišim vrhom oko 180–200 bp. Distribucija ima asimetričan oblik sa blagim desnim repom, što je karakteristično za exome sekvenciranje gde se cilja određeni skup regiona.
Ovakav obrazac dužina fragmenata potvrđuje uspešnu biblioteku za sekvenciranje i validnu pripremu uzorka.

Vizualizacija histogram dužina je urađena pomoću matplotlib u Colab-u (Histogrami.ipynb).



5. JSON fajl sa putanjama

Priložen fajl results.json sadrži strukturisane putanje do svih ključnih rezultata po koracima obrade.

Komanda za kreiranje JSON fajla :

cat <<EOF > results.json

{

"fastqc": {

"read1_html": "results/01_fastqc/sample_0.chrom11.exome.pe1_fastqc.html",

"read2_html": "results/01_fastqc/sample_0.chrom11.exome.pe2_fastqc.html"

},

"mapping": {

"sam": "results/02_mapping/aligned.sam",

"bam": "results/02_mapping/aligned.bam",

```
    "sorted_bam": "results/02_mapping/sorted.bam",
    "index": "results/02_mapping/sorted.bam.bai"
  },
  "deduplication": {
    "bam": "results/03_dedup/dedup.bam",
    "index": "results/03_dedup/dedup.bai",
    "metrics": "results/03_dedup/dup_metrics.txt"
  },
  "bqsr": {
    "recal_data": "results/04_bqsr/recal_data.table",
    "bam": "results/04_bqsr/recalibrated.bam",
    "index": "results/04_bqsr/recalibrated.bai"
  },
  "variants": {
    "raw_vcf": "results/05_haplotypecaller/raw_variants.vcf",
    "filtered_vcf": "results/06_filtering/filtered_variants.vcf.gz"
  }
}
EOF
```

Komanda za ispis : cat results.json

```
{
  "fastqc": {
    "read1_html": "results/01_fastqc/sample_0.chrom11.exome.pe1_fastqc.html",
```

```
    "read2_html": "results/01_fastqc/sample_0.chrom11.exome.pe2_fastqc.html"
  },
  "mapping": {
    "sam": "results/02_mapping/aligned.sam",
    "bam": "results/02_mapping/aligned.bam",
    "sorted_bam": "results/02_mapping/sorted.bam",
    "index": "results/02_mapping/sorted.bam.bai"
  },
  "deduplication": {
    "bam": "results/03_dedup/dedup.bam",
    "index": "results/03_dedup/dedup.bai",
    "metrics": "results/03_dedup/dup_metrics.txt"
  },
  "bqsr": {
    "recal_data": "results/04_bqsr/recal_data.table",
    "bam": "results/04_bqsr/recalibrated.bam",
    "index": "results/04_bqsr/recalibrated.bai"
  },
  "variants": {
    "raw_vcf": "results/05_haplotypecaller/raw_variants.vcf",
    "filtered_vcf": "results/06_filtering/filtered_variants.vcf.gz"
  }
}
```