

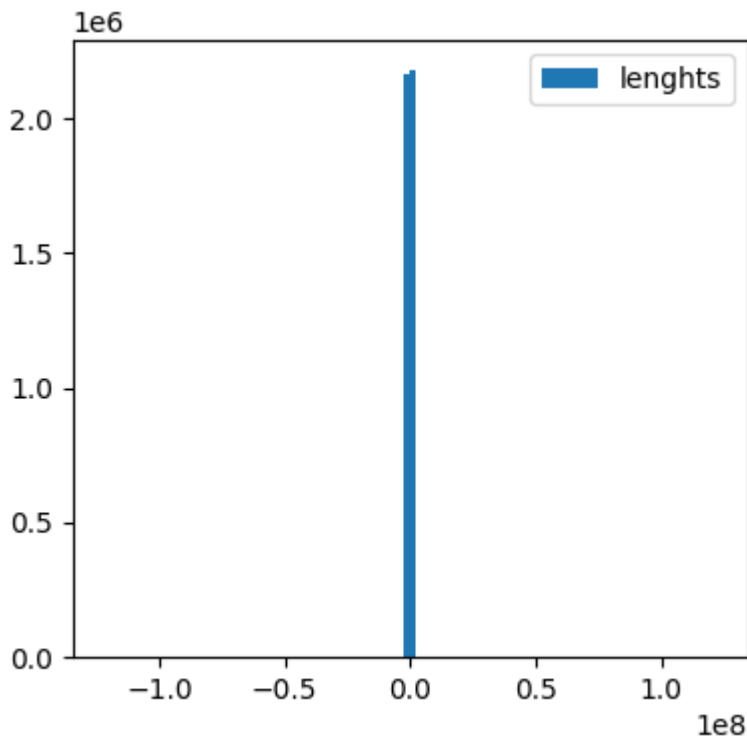
# Izveštaj za drugi projekat iz Uvoda u bioinformatiku

1. Mapirati sekvencirane readove na referentni genom hg38 upotrebom alata BWA Mem. Upotrebom python biblioteke pysam odrediti sledeće:

- 1a. Koliko je readova uspešno mapirano?
- 1b. Koliko je parova readova mapirano tako da su oba para mapirana?
- 1c. Nacrtati histogram dužina sekvenciranih fragmenata (template\_length).

**1a.** Uspešno je mapirano 4367188 readova.

**1b.** Ukupno 4294503 ima parova readova tako da su oba mapirana



**1c.**

2. Izvršiti obradu dobijenog BAM fajla prema GATK protokolu (Markiranje Duplikata, rekalkulacija kvaliteta baza)

- 2a. Koliki su procenati PCR i optičkih duplikata?

**2a.** Procenti PCR i optičkih duplikata su 6,7953%

3. Identifikovati mutacije upotrebom alata Haplotype Caller i filtrirati mutacije predefinisanim filterima (hard filtering) prema Broad preporukama (kao što je rađeno na vežbama).

- 3a. Koliko je ukupno mutacija identifikovano, koliko od njih su SNP-ovi, a koliko INDEL-i?
- 3b. Koliko mutacija prolazi, a koliko ne prolazi kriterijume filtriranja.
- 3c. Izračunati Ti/Tv odnos pre i posle filtriranja.

**3a.** Ukupno je identifikovano 191 mutacija, od čega su 187 SNP-ovi a 4 INDEL-i

**3b.** 191 mutacija prolazi filtriranje, a 13645 ne prolazi

**3c.** Ti/Tv odnos pre filtriranja je 2.1755610972568578, a posle filtriranja je 2.1692346552159636

4. Anotirati mutacije alatom Funcotator

- 4a. Izbrojati različite vrednosti ClinVar značajnosti (anotacija ClinVar\_VCF\_CLNSIG, koliko mutacija je označeno kao Benign, Likely\_benign, itd.).

**4a.** Anotacija ClinVar\_VCF\_CLNSIG:

Broj mutacija koje su označene kao Benign je: 265

Broj mutacija koje su označene kao Likely\_Benign je: 25

5. Svi uzorici sadrže određenu količinu kontaminacije DNK materijalom bakterijskog ili virusnog porekla. Većina ovakvih readova se neće mapirati na ljudski genom. Izvući readove koji nisu mapirani u procesu mapiranja, asemblovati ih alatom abyss, i identifikovati organizam od kojeg potiče najduži skafold upotrebom alata Blast.

**5.** Organizam od kog potiče najduži skafold je: *Bradyrhizobium diazoefficiens* strain 172S4 chromosome