



Vilniaus Universitetas
Matematikos ir informatikos fakultetas
Programų sistemų katedra

Programų sistemų studijų programa
Bioinformatikos III-ojo laboratorinio darbo ataskaita

Parengė: Laura Tumaitė

Darbo vadovas: Gediminas Alzbutas, Partn. Prof., Dr.

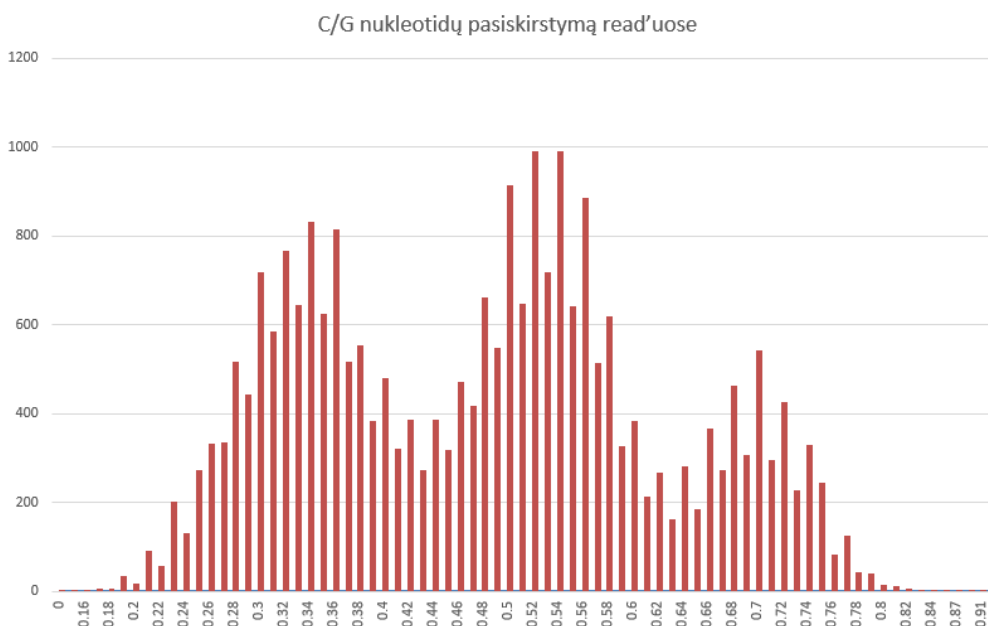
Vilnius
2022

1. ATASKAITA

1. FASTQ formatas tai yra tekstinių duomenų formatas ir FASTA formato plėtinys, naudojamas biologinei sekai (dažniausiai nukleotidų sekai) ir kiekvieno sekos elemento kokybės balams pateikti. Sekos elementai ir jų kokybės rodikliai yra užkoduoti pavieniais ASCII simboliais. Lyginant FASTQ su FASTA formatu, FASTQ formatas papildomai pateikia kiekvienos sekos elemento kokybės rodiklį.
2. Gimiau 12 dieną, tai $33+12=45$. 45 ASCII lentelėje atitinka – simbolį.
3. Pirmi 32 ASCII kodai negali būti naudojami sekos kokybei koduoti, nes tie kodai nėra skirti spausdintinai informacijai atvaizduoti, t.y. tie kodai yra skirti apibūdinti tam tikrą funkcinę klaviatūros įvestį, pvz. space, backspace, cancel.
4. Skriptas yra Lab3.ipynb faile
 - a. Pateikto failo kokybės kodavimas yra Sanger Phred+33. Tam, kad nustatyti formatą buvo naudotas Bioinfokit biblioteka. Taip pat galima atskirti pagal naudojamus ASCII simbolius (imta iš vikipedijos):

```
S - Sanger Phred+33, raw reads typically (0, 40)
X - Solexa Solexa+64, raw reads typically (-5, 40)
I - Illumina 1.3+ Phred+64, raw reads typically (0, 40)
J - Illumina 1.5+ Phred+64, raw reads typically (3, 41) with 0=unused,
  1=unused, 2=Read Segment Quality Control Indicator (bold) (Note: See
  discussion above).
L - Illumina 1.8+ Phred+33, raw reads typically (0, 41)
P - PacBio Phred+33, HiFi reads typically (0, 93)
```

- b. Gautame grafike matosi, kad yra 3 stambūs pikai:



c. Buvo naudojama iš šio įrankio:

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=MicrobialGenomes tam kad padaryti šį uždavinį.

READ ID	RASTO MIKROORGANIZMO RŪŠIS
PIRMO PIKO	
M00827:12:000000000-AEUNW:1:1101:12983:1610	Staphylococcus aureus subsp. aureus NCTC 8325 chromosome, complete genome
M00827:12:000000000-AEUNW:1:1101:13901:2032	Staphylococcus aureus subsp. aureus NCTC 8325 chromosome, complete genome
M00827:12:000000000-AEUNW:1:1101:17616:2339	Escherichia coli str. K-12 substr. MG1655, complete genome
M00827:12:000000000-AEUNW:1:1101:21543:2685	Staphylococcus aureus subsp. aureus NCTC 8325 chromosome, complete genome
M00827:12:000000000-AEUNW:1:1101:20192:2848	Staphylococcus aureus subsp. aureus NCTC 8325 chromosome, complete genome
ANTRO PIKO	
M00827:12:000000000-AEUNW:1:1101:16324:1777	Escherichia coli str. K-12 substr. MG1655, complete genome
M00827:12:000000000-AEUNW:1:1101:18840:1877	Escherichia coli str. K-12 substr. MG1655, complete genome
M00827:12:000000000-AEUNW:1:1101:18967:1954	Escherichia coli str. K-12 substr. MG1655, complete genome
M00827:12:000000000-AEUNW:1:1101:18037:2122	Staphylococcus aureus subsp. aureus NCTC 8325 chromosome, complete genome
M00827:12:000000000-AEUNW:1:1101:11654:2202	Escherichia coli str. K-12 substr. MG1655, complete genome
TREČIO PIKO	
M00827:12:000000000-AEUNW:1:1101:20958:2801	Thermus thermophilus HB8 chromosome 1, complete sequence
M00827:12:000000000-AEUNW:1:1101:22220:3113	Thermus thermophilus HB8 chromosome 1, complete sequence
M00827:12:000000000-AEUNW:1:1101:18070:3392	Thermus thermophilus HB8 chromosome 1, complete sequence
M00827:12:000000000-AEUNW:1:1101:14615:4214	Thermus thermophilus HB8 chromosome 1, complete sequence
M00827:12:000000000-AEUNW:1:1101:23350:4251	Thermus thermophilus HB8 chromosome 1, complete sequence

5. Buvo rasti šios bakterijos rūšys:

- Staphylococcus aureus
- Escherichia coli
- Thermus thermophilus