

Bioinformatikos Lab. Nr. 2

Rytis Kaplūnas PS6 2018-11-28

1. FASTQ formatas - tekstinis formatas, skirtas laikyti sekas ir jų kokybes viename faile. Sekos ir jų kokybės yra užkoduotos su ASCII simboliais.

FASTA nuo FASTQ skiriasi tuo, kad FASTQ failai savyje laiko informaciją apie sekos kokybę, kuri pasako kaip tiksliai buvo parinkta bazė.

2. Gimiau 7 mėnesio dieną: $33 + 7 = 40$
Atitinkantis ASCII simbolis: ‘(‘
3. Pirmieji 32 ASCII simboliai yra nespausdinami simboliai, jie naudojami teksto atvaizdavimui apibrėžti, todėl jų negalima naudoti sekos kokybei koduoti.

4.
 - a. Su parašytu script'u buvo nustatyta, kad koduotė yra Sanger/Illumina-1.8. Abi koduotės naudoja Phred+33 kokybės vertinimą. Nustatymas yra paremtas tuo, kokie ASCII simboliai sudaro kiekvieno read'o kokybės seką.

Koduočių intervalai:

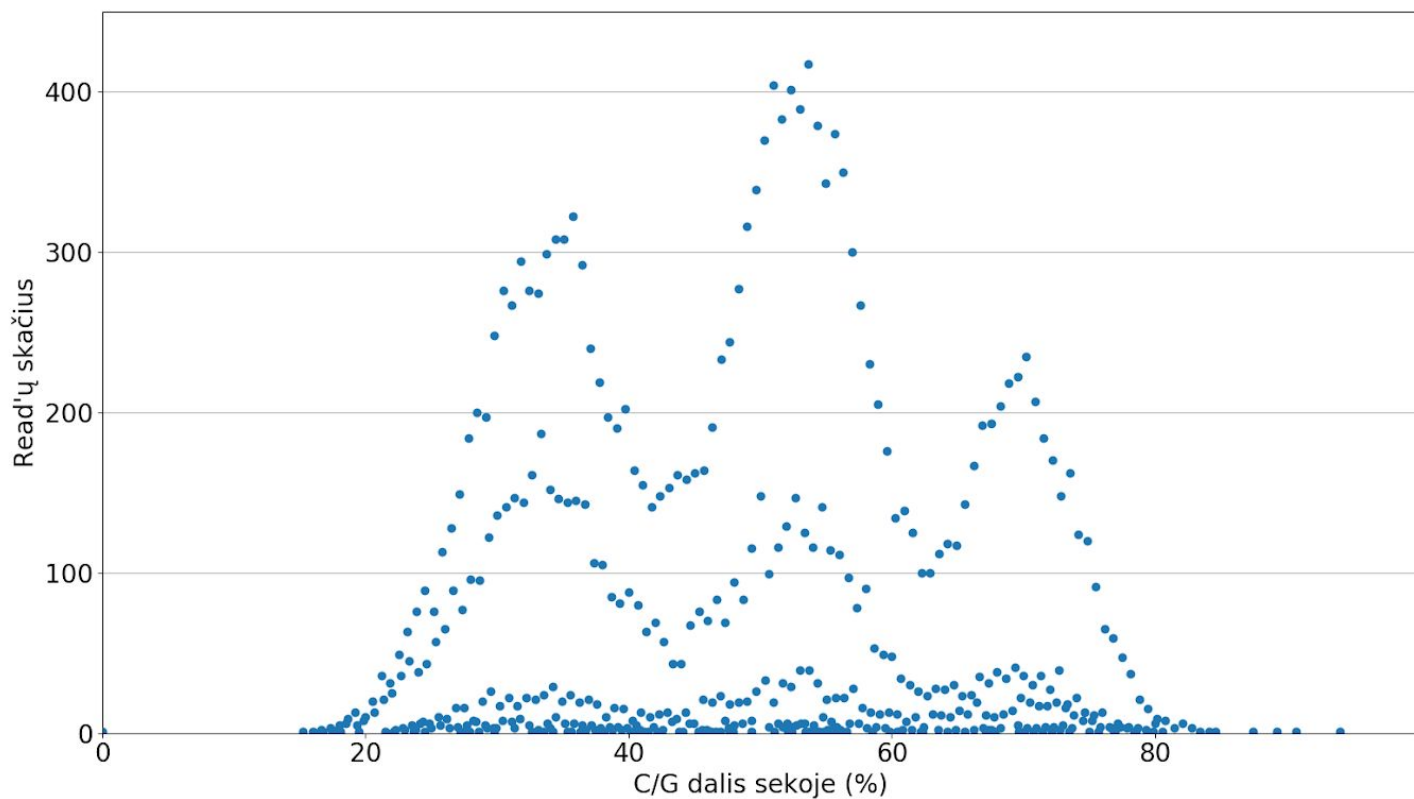
- Sanger: [33, 73],
- Illumina-1.8: [33, 74]
- Solexa: [59, 104]
- Illumina-1.3: [64, 104]
- Illumina-1.5: [66, 105]

Einant per visus read'us ieškomos min ir max reikšmės. Ir pagal jas sudarome intervalą ir nustatome koduotę.

Gautas intervalas buvo [35, 73], todėl koduotė yra Sanger/Illumina-1.8.

- b. Išanalizavus C/G nukleotidų pasiskirstymą read'uose buvo nubrėžtas grafikas. Iš jo parinkti trys pikai:
 - 35.7616% - 322 read'ų.
 - 53.6424% - 417 read'ų.
 - 70.1987% - 235 read'ų.

Gautas grafikas:



c. Išsirinkus po penkis read'us iš kiekvieno piko, buvo atlikta BLAST paieška. Gauti rezultatai pavaizduoti lentelėje:

M00827:12:000000000-AEUNW:1:1101:19585:8123	<i>Shigella boydii</i>
M00827:12:000000000-AEUNW:1:1101:20008:11235	<i>Escherichia coli</i>
M00827:12:000000000-AEUNW:1:1101:7922:8647	<i>Thermus thermophilus</i>
M00827:12:000000000-AEUNW:1:1101:11761:10309	<i>Staphylococcus aureus</i>
M00827:12:000000000-AEUNW:1:1101:11742:3950	<i>Shigella dysenteriae</i>
M00827:12:000000000-AEUNW:1:1101:23128:3788	<i>Shigella flexneri</i>
M00827:12:000000000-AEUNW:1:1101:11245:8780	<i>Thermus thermophilus</i>
M00827:12:000000000-AEUNW:1:1101:19588:16449	<i>Staphylococcus aureus</i>
M00827:12:000000000-AEUNW:1:1101:7248:14390	<i>Escherichia coli</i>
M00827:12:000000000-AEUNW:1:1101:5538:14656	<i>Escherichia coli</i>
M00827:12:000000000-AEUNW:1:1101:26090:9604	<i>Escherichia coli</i>
M00827:12:000000000-AEUNW:1:1101:11998:18470	<i>Thermus thermophilus</i>
M00827:12:000000000-AEUNW:1:1101:28070:13077	<i>Thermus thermophilus</i>
M00827:12:000000000-AEUNW:1:1101:17868:6469	<i>Staphylococcus aureus</i>
M00827:12:000000000-AEUNW:1:1101:23294:5998	<i>Thermus thermophilus</i>

5. Taigi, atlikus sekų, pateiktų FASTQ faile, analizę, buvo nustatyta, kad mėginyje buvo:

- *Shigella boydii*
- *Escherichia coli*
- *Thermus thermophilus*
- *Staphylococcus aureus*
- *Shigella dysenteriae*
- *Shigella flexneri*