1. Fasta saugoja seką ir sekos ID. FASTQ saugoja sekos ID, prasideda su ženklu @. Po to eina pati seka. Po sekos eina nauja eilutė kuri prasideda + ženklu, čia gali būti pakartota sekos ID ir aprašymas. Ir galiausiai eina sekos kokybės ivertinimo eilutės. Taigi, pagrindinis skirtumas tarp FASTA ir FASTQ – FASTA neturi informacijos apie sekos kokybę.

2. Pirmi 32 ASCII simboliai neturi atitinkamų ženklų, tai visokie line feed, carriage return, back space ir t.t. Todėl jie netinka.

3. Gimiau gegužės 30. 30+33= 63. 63 atitinka ? ženklą.

4.

a) Sekos kokybė sanger+phred 33. Skaičiavau taip : radau kiekvieno reado mažiausia ir didžiausia kokybės reikšmę. Tada naudodamas algoritmo poslinkius ieškojau, kurio algoritmo yra mažiausias skirtumas tarp readuose rastų min ir max reiškšių ir tame algoritme galiojančių min ir max reiksmių. Pasirinktas tas algoritmas, kurio rėžiai labiausiai sutapo su readuose rastuose skirtumu tarp min ir max reikšmių.

Naudoti šie režiai (rasti vikipedijoje)

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)

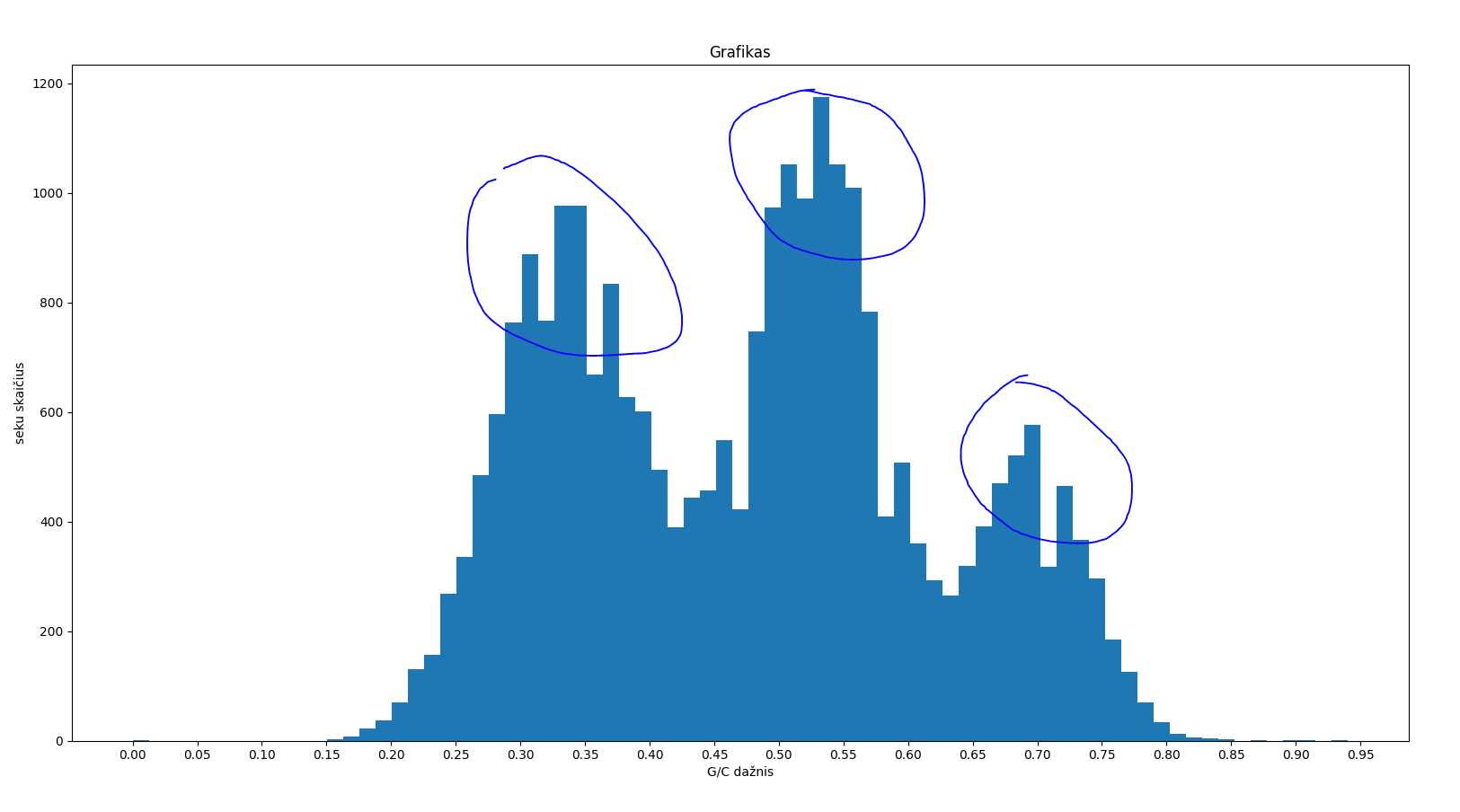
J - Illumina 1.5+ Phred+64, raw reads typically (3, 41)

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)

b) skaičiavau G/C dalis sekose ir gavau diagramą :



Čia matomi 3 pikai.

c)

1 0.327

TGCGTTTATTACATTACCTTCTATTATATTTGCTGACGAACCAACAGGTGCACTGGATTCTAAAAGTACTCAAGATTTATTAAAACGATTAACAAGAATGAATGAAGCATTTAAGTCTACAATTATTATGGTAACGCATGATCCTGTTGC

Staphylococcus aureus strain GDY8P96A chromosome, complete genome

2973 0.338

TCATTATTTCACTTCACCATTAAATTCATTCATAACATGGTCAAATCGTGATGTTTCAACAAACTTTTCAATTGCGCGTGCTGCGTGTTCGATAACTTTTTCCATCGTTACCATTTCATCATTTGAAAAGCGTTGTAAAACATAATCAGGT

Staphylococcus aureus subsp. aureus strain CC1153-MRSA chromosome, complete genome

3047 0.325

GTTGGAATTGACGCTCTATTTCCTTTAACTGAATATCACGCAAGCGACGTTCTGTATGCTCAATAGAATGGATCAGTTGCTCATTTGATTCAAATAAATCGTAAATATAATTATTAATTTCTTCTAGTTCACTGTTGTTTTTTAAAGGCGT

Staphylococcus aureus strain UP\_644 chromosome, complete genome

7267 0.322

GAATGCTGAACCTTTACCAGGCTTTACATGTTGGAAGTCTATAACTTTCCAAATAGCGTTATCAACAGAAATTGTTAAACCTGTTTTAAAATCATTAACCGAAATCATTCAGTTTCCTCCTCATTTTACACGCTTATGTTAAAACTATA

Staphylococcus aureus subsp. aureus strain CC1153-MRSA chromosome, complete genome

25740 0.325

ATAAGGTATTCCAGGATTTTTAACAATTATTGGATTATTATCAAGCAACGTTAATGGATGACTTCCACTTACAACAGAAATGCCCATAGATTCTAAATCTTTTGCATGAGCATCTTGTGATAAGTCTTTTCCATCATTGACAGTTACATTC

Staphylococcus aureus strain GDY8P96A chromosome, complete genome

iš pirmo piko

969 0.51

CACCCTCAATTTTCACCGCTTCACCAAAAACGATCTTCGGCACGATGACACCGGCCCGTTCCAGTTCTTTAATTACCGCAACCTCACGGACAATCGTTGGTCGGCCGAACGGATAACGTACGGAATGAAACAGATGATGCGTCATGCGCTT

Escherichia coli strain FDAARGOS\_943 chromosome, complete genome

1585 0.523

CTGCTCGGTATTCAAAGCGGCAAGAATCTCATTGCCCGCTTCATGATCTTTCGCCGAGCCAAACAGAACCACCTGATAACCTTCATCAATCAGCTGCTTTGCCAGCTCCGCATAGTGGTAGTGTGGCCAGCGTTTTGCCGGACCAAACTCC

Escherichia coli strain FDAARGOS\_943 chromosome, complete genome

2032 0.53

CCATTCGATATTAAAACCATCAAGGTGCGCGAGATAATCCTGACAGGCGTTATCCCGCACGCTAACCAGCGCACTAACTTGCGGTTTAACGGCAAGTTTGCTCAACGTTTCCAGGCGGCGGCGCAATGCAGGCATTCCGGATTCATCGC

Escherichia coli strain NEB5-alpha\_F'Iq chromosome, complete genome

7280 0.51

GAACCTGATAATGAAGATGCGCCTTATTATGAATACGCGCAACGCGTCTATTGTGGAGAAGGCGAAAGCCTGTTGGCAGAACTTTGTGACTACTGGCGCGAGTATCCCTCCACCCAGGCGGATGCTTTAATGTTGCAATGGTGCCGTCAGC

Escherichia coli TA8571 DNA, complete genome

7661 0.51

TATCCTGCAATTCACAGCAGTGAACGGTACGACCAATAATACGGCCCAGCGTGGAGTGCAGCTGTGCCACTTCCAGCGTGTTGCCAGTCAGTTTCTGATAAATCGCAACGATTTCATTCAGTTTGTTCTGGGTAGATTCGCGACCTGCCGC

Escherichia coli TA8571 DNA, complete genome

iš antro piko

3936 0.682

GGACCGCCTGCCTGACCCCCTCCCCGAGGAAAGGCTCCTCGCCCTTTACCGGGAGGACCCCAGGCCCAAGAGCCTGGACGATTTTCCCGAGCTCATCGCCCACCTGAGGCGGCTCCACCCCATCCGCCACCGCTACCTCTTTGACGGCCTG

Thermus thermophilus HB8 genomic DNA, complete genome

4040 0.695

GAGGGCCTTTTGCAGGGCGTGGCCGGAAAGCTTGGGGGGGAGGGTATAGCCCAGGCGGGCCAGGGCCTGGCGGAGCTTCTCGTAGGCGTCCTGGACCGGCTCCTCGTGGACGCGGAAGAGGCCGGGGAGGCCCTTTTGCACCAGGTACTCC

Thermus thermophilus AA2-29 DNA, complete genome

249 0.678

GGTGGTCATGGACGAGGAGGCGCGGCTCGGCTACACCGAGGTGAAGATCGGCTTCGTGGCCGCCTTGGTCTCGGTGATCCTGGTGCGGGCCGTGGGGGAAAAGGCCGCCAAGGACCTCCTCCTCACGGGAAGGCTCGTGGAGGCAAGGG

Thermus thermophilus AA2-29 DNA, complete genome

3319 0.675

CCATCCGCACCCCCGTGGCCGAGATCCGGCAGTACTCCCGGGCCACCGCCGGCGTCAGGGTGATGAACCTCCCGGAGGACGACGAGGTGGCGAGCGCCTTCGTGGTGGAGGAGGAAAAGTGATGGAAGAGGTCGTGCTCATCACGGTGCCG

Thermus thermophilus HC11 DNA, complete genome

15188 0.675

GAGGGCGGTGGCCAGGGCCCTTGCCTCCTCCTTGCGGTGGAGACCGAAAAGGACCTGCAAAAACCCCGCCCGGGTGCGGGCGTAGGTGGCCAGGGCCGCGTTCTCCAGGAGCTTCATCTTGACCCTCCTTTTCTCCGCACCGGGCTTGCCC

Thermus thermophilus HB8 genomic DNA, complete genome

iš trečio piko

5.

Buvo rastos šios bakterijos – thermus thermopilus, E Coli, Staphylococcus aureus.