Задание 2. чекпоинт 1.

Дубинская ЕК.

Репозиторий со всем - всем - всем: https://github.com/villanellex2/NSU\_Bioinformatics

1. Ссылка на загруженные прочтения:   
   <https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR6140826>
2. Скрипт на bash:

fastqc sra\_data.fastq

mv sra\_data\_fastqc.html $1\_1.html

rm sra\_data\_fastqc.zip

minimap2 -a GCF\_000005845.2\_ASM584v2\_genomic.fna sra\_data.fastq.gz > kotiki.sam

samtools view kotiki.sam   
samtools flagstat kotiki.sam

1. Результат команды

7267 + 0 in total (QC-passed reads + QC-failed reads)

1324 + 0 secondary

944 + 0 supplementary

0 + 0 duplicates

7145 + 0 mapped (98.32% : N/A)

0 + 0 paired in sequencing

0 + 0 read1

0 + 0 read2

0 + 0 properly paired (N/A : N/A)

0 + 0 with itself and mate mapped

0 + 0 singletons (N/A : N/A)

0 + 0 with mate mapped to a different chr

0 + 0 with mate mapped to a different chr (mapQ>=5)

1. Скрипт разбора файлов с результатами:

import sys

if \_\_name\_\_ == '\_\_main\_\_':

file\_name = sys.argv[1]

file1 = open(file\_name, 'r')

Lines = file1.readlines()

mylist = Lines[4].split("(")

k = 0

a = mylist[1][k]

res = ''

while a != '%':

res += a

k += 1

a = mylist[1][k]

if float(res) > 90.0:

print("OK")

else:

print("Not OK")