



LAB 1.1

TIACOS

March 24, 2017

In this laboratory, we will use the standard Shell commands to get information about a biological sequence stored in a FastQ format (https://en.wikipedia.org/wiki/FASTQ_format).

- 1) Understand what the scripts in files ***FastQ-Histogram.sh*** and ***FastQ-Histogram-pipes.sh*** do by reading the manual (man). You have a copy also in your virtual machine (labs/1.1/Scripts)

FastQ-Histogram.sh

```
#!/bin/bash  
  
awk '0 == (NR) % 4' $1 > out1.txt  
  
#the line above prints each 4th line  
  
fold -w1 out1.txt >out2.txt  
  
sort out2.txt >out3.txt  
  
uniq -c out3.txt
```

FastQ-Histogram-pipes.sh

```
#!/bin/bash  
  
awk '0 == (NR) % 4' | fold -w1 | sort | uniq -c
```

If you are interested in visualizing this information, you can use an asci version of gnuplot with commands underneath, whe “file” ids the output of your scripts.

```
$gntuplot  
  
gnuplot> set terminal dumb  
  
gnuplot> plot “file” using 1:xtic(2) with boxes
```

- 2) Assuming you have a FastQ file that stores a sequence from the Illumina software, what commands would you use to get the number of sequences and the identifier of the instrument?
- The following commands may be useful: cat, grep, wc, cut, head
- a) Implement redirecting to files (one step per action)
- b) Implement using pipes whenever possible (reducing the number of temporary files to the minimum)
- 3) Write a very simple script that implements the workflow done in 0
- 4) Compile file kseq_text.c with library z (in labs/1.1/C you will find a makefile that may help you in steps a) and b))
- a) Using the flags -static (this will link the library statically)
- b) Without using the flags -static (this will link the library dynamically)
- c) Execute both versions (./kseq_test test.seq). Do you get the same result? Why?
- d) Check the size of the both files. Do they have different sizes? Why?
- e) Execute nm on both executable files. Do we get the same result? Why?

Report: To submit the work done in this lab, please send a document (it can be a simple text file) **answering all questions 3, 4.d, and 4.e**, and submit it via moodle.