Pre-processing the raw sequencing data

System requirement:

* macOS X Mojave Version 10.14 with python 2.7
* galaxy local version: 17.09
* FASTQ Trimmer version 1.1.1
* FASTQ Groomer version 1.1.1
* FastQC Hight Throughtput Sequence QC Report version 0.11.8

Installations:

* Galaxy
  + https://galaxyproject.org/admin/get-galaxy/
  + configure the local serve according to the instruction
* FASTQ Trimmer, FASTQ Groomer
  + Install Galaxy tools according to the instructions
  + https://galaxyproject.org/admin/tools/add-tool-from-toolshed-tutorial/
* FastQC
  + https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

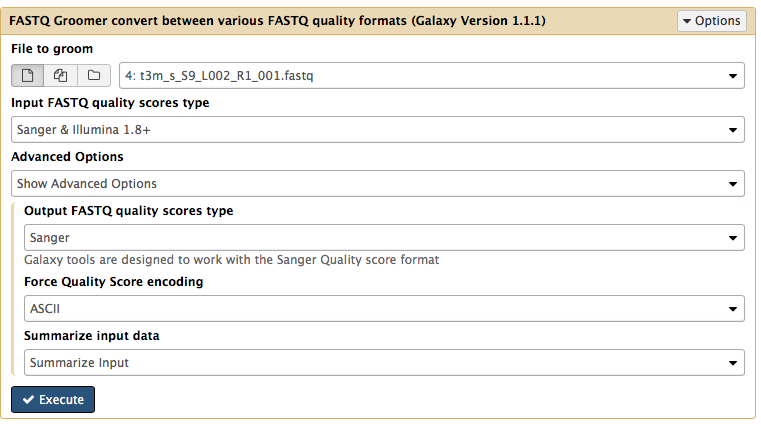
Instruction:

* open terminal
* cd galaxy (change directory to the local galaxy folder)
* sh run.sh (launch the local galaxy server)
* open a web browser, copy and paste the local server address:

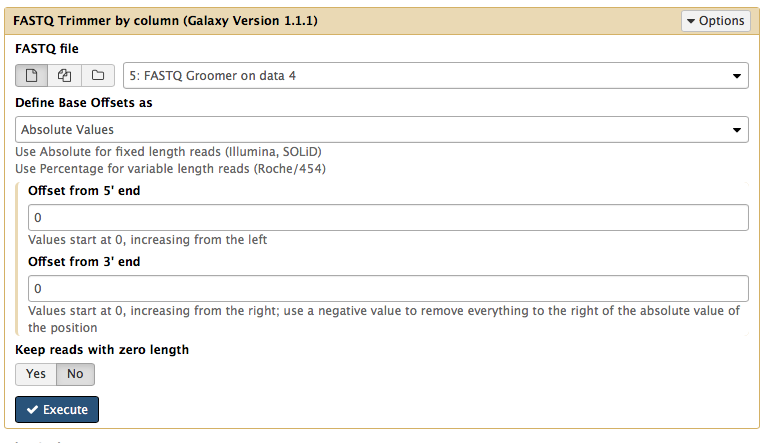
http://127.0.0.1:8080

* Upload the raw data into the Galaxy local server
* Run FASTQ Groomer (Galaxy)

1. File to groom: select the .fastq file
2. Input FASTQ quality scores type: Sanger & Illumina 1.8+
3. Advanced Options: Show Advanced Options
4. Output FASTQ quality scores type: Sanger
5. Force Quality Score encoding: ASCII
6. Summarized input data: Summarize input
7. Start by pressing "Execute"



* Run FASTQ Trimmer (Galaxy)
  1. Select the groomed FASTQ file
  2. Define Base Offsets as: Absolute Values
  3. Offset from 5' end: 3
  4. Offset from 3' end: 0
  5. Keep reads with zero length: No
  6. Start by pressing "Execute"



* Save the processed sequence files and check the quality by FastQC (command line)
  + fastqc $DIR/$file
  + $DIR: directory where the processed sequence located
  + $file: the name of the sequence file