Installing and Running FastQC and MultiQC

# (Using the Linux Command Line)

About this Protocol

This protocol is for users who are interested in checking the quality of their raw next-generation sequence files (DNA or RNA). FastQC takes FASTQ, BAM, or SAM files. It outputs a HTML report that provides commonly used quality scores. MultiQC is a convenient tool to merge all single result files from FastQC into a single report. This makes it easier to compare quality statistics between multiple samples. If you only have one sample, running MultiQC is not necessary.

Challenge Level: *For beginners in bioinformatics. Need to understand some basic Linux shell scripting. To learn more about FastQC and MultiQC before beginning, check out these links: FastQC: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/*

MultiQC: https://multiqc.info/ and https://github.com/ewels/MultiQC

Context: *For this project, I am working on a Linux subsystem. I have raw RNA-seq files in FASTQ format, and I would like to check the quality of the reads.*
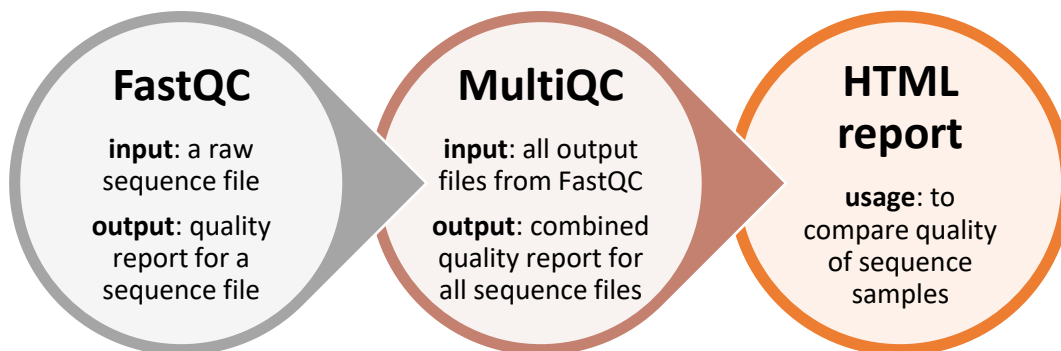
Let's get started

Install FastQC:

1. You can install FastQC using the built in Advanced Packaging Tool (apt):

> ```
> > sudo apt install fastqc
> ```

> ```
> #It will ask for your password. This is your Linux account
> password.
> ```

Running FastQC:

2. Navigate into the folder that contains the files you would like to assess the quality for. My files are unzipped FASTQ files. You can also use BAM or SAM files.

> If you only want to assess the contents of one file:

> ```
> > fastqc sample1.fastq
> ```

> If you only want to assess a couple of files, you can write the name of each file out with a space between each file name:

> ```
> > fastqc sample1.fastq sample2.fastq
> ```

> If you want to assess all FASTQ files in your folder, you can use a wildcard character (*):

> ```
> > fastqc *.fastq
> ```

Other options:

3. FastQC comes with additional features, such as where to redirect the output files (-o) (if you do not want the files in the working directory), if you want to increase the number of threads working on the analysis (-t), or if you want to specify adapter sequences. Check all the options in the built-in manual:

> ```
> > fastqc --help
> ```

Output:

4. Click on each of your .html files. They should open separate webpages. These are your quality scores for each sample. The interpretation of results will vary depending on the sequencing strategies used, the type of sequence data, and the requirements of your experiment. Here are good resources to help understand the output:

https://dnacore.missouri.edu/PDF/FastQC_Manual.pdf

https://rtsf.natsci.msu.edu/genomics/tech-notes/fastqc-tutorial-and-faq/

Install MultiQC:

1. We will install MultiQC through the Python Package Index (PyPI). If you do not already have the PyPI (or do not know if it is installed), run this command to install it first:

> **`> sudo apt install python3-pip`**

Type your password when prompted, then type 'Y' when asked "Do you want to continue?" It will take a few minutes to install (~10-20 minutes).

2. Now we can use PyPI to install MultiQC.

> **`> pip install multiqc`**

*If this installation method does not work, see *Troubleshooting* section ("Alternative method to install MultiQC").

Running MultiQC:

3. Navigate into the folder that contains all the output files from FastQC. This will include .html and .zip files. From the previous example, we would have: sample1_fastqc.html, sample1_fastqc.zip, sample2_fastqc.html, sample2_fastqc.zip.

> **`> multiqc .`**

Other options:

4. MultiQC also comes with additional features, and all the options can be reviewed in the built-in manual:

> **`> multiqc --help`**

Output:

5. The interpretation of results will be like your FastQC results; however, it will combine all your sample reports into a single report file. This makes it easier to compare between multiple samples. Click on your ***multiqc_report.html*** file. It should also open a webpage and this page should include all your samples into a single report file. There are several video tutorials that explain each of the statistics, such as this one: Using MultiQC Reports - YouTube .

Troubleshooting

Alternative method to install MultiQC:

1. If the Python Package Index (PyPI) installation method did not work then fear not, it also did not work for me. Instead, we are going to try to install MultiQC using Bioconda. To do this lets first install Bioconda.

> **`> curl -O https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh`**

> **`> sh Miniconda3-latest-Linux-x86_64.sh`**

> This might ask you several questions, accept the defaults.

2. Close your terminal window, and re-open. You should see "(base)" before your Linux username. Then we will set up channels. Channels gives Conda the location where the packages we want to install are stored.

> **`> conda config --add channels defaults`**

> **`> conda config --add channels bioconda`**

> **`> conda config --add channels conda-forge`**

3. Now we will install two packages (bwa and multiqc):

> **`> conda install bwa`**

> **`> conda install -c bioconda multiqc`**

4. You should be able to run MultiQC now through the conda environment.

Tips

1. If you don't want Conda to activate by default (i.e., if every time you open a new terminal window, you do not want to be in the Conda environment- when you see "(base)" before your Linux username):

> **`> conda config --set auto_activate_base false`**

2. Then, whenever you want to use Conda or multiqc, you can manually activate or deactivate:

> **`> conda activate`**

> **`> multiqc --help`**

> **`> conda deactivate`**