[5E] Basic Quality control of raw reads

Tutorial

In this session you will learn to import, view and check the quality of a sequenced data using FastQC

- Install FastQC in your computer:

    1. FastQC has a graphical interface and can be downloaded and run on a Windows /Linux/Mac computer. It is available here

- Download the Fastq files or the raw data to your computer:
    - The dataset you will be working with is an E.coli isolate sequenced by an Illumina platform. The dataset contain millions of reads and are therefore quite big. We are only going to use a subset (EC505_subset_R1.fastq.gz and EC505_subset_R2.fastq.gz) of the original dataset for this tutorial which are available via Jottacloud (~100MB, 2 files). https://www.jottacloud.com/s/259ce4585cc6cd14401984f9add3ba548db

- Import EC505_subset_R1.fastq.gz and EC505_subset_R2.fastq.gz to FastQC
    - To know how to run FastQC please refer to FastQC documentation here

- Run FastQC, look what it producing, and wait for the outputs
    - For each fastq file, a .zip archive containing all the plots, and a .html report will be produced
    - Open the .html files with your favourite web browser

- Study both files and FastQC results and answer the questions in the following survey: https://www.surveymonkey.com/r/5EBasicQC