ACEMFS FUT Minna Bioinformatics Workshop

Sequence Retrieval & Quality Control using Galaxy

Itunuoluwa Isewon PhD
Covenant University

Workshop Outline

By the end of this session, participants will be able to:

● Retrieve sequences of mycotoxins & fungal enzymes from databases.

● Upload & organize datasets in **Galaxy**.

● Perform QC using Galaxy tools (**FastQC**)

Introduction to Galaxy

- Web-based platform for bioinformaticians.

- No coding required.

- Supports reproducible research.

- Widely used for genomics, proteomics and transcriptomics analysis.

Sequence Retrieval

NCBI Exploration:

1. Go to **NCBI**

2. In the search bar, paste the **lcc9 gene**.

3. Click on the dropdown arrow next to the search bar and select **Nucleotide**.

4. Click the search button.

Sequence Retrieval

Sequence Retrieval

5. On the results page click the **first hit**.

6. What is the name of the **organism**?

7. How many **base pairs** does it have?

8. What is its **accession** and **version numbers**?

Sequence Retrieval

Sequence Retrieval

9. On the right, click on the dropdown arrow next to Send to:

10. Select file under, Choose **Destination**.

11. Change the format to **FASTA**

12. Click on **Create File** to download the FASTA File.

13. **Repeat** this process for a gene interesting to you.

Sequence Retrieval

Quality Control in Galaxy

Obtain **FASTQ** files:

1. Go to **Galaxy**

2. In the tool Bar, click on **Get Data**

3. Choose "**Faster Download and Extract Reads in FASTA/Q format from NCBI SRA**".

4. In the Accession tab, write the accession number of the fastq file: **SRR18453616**.
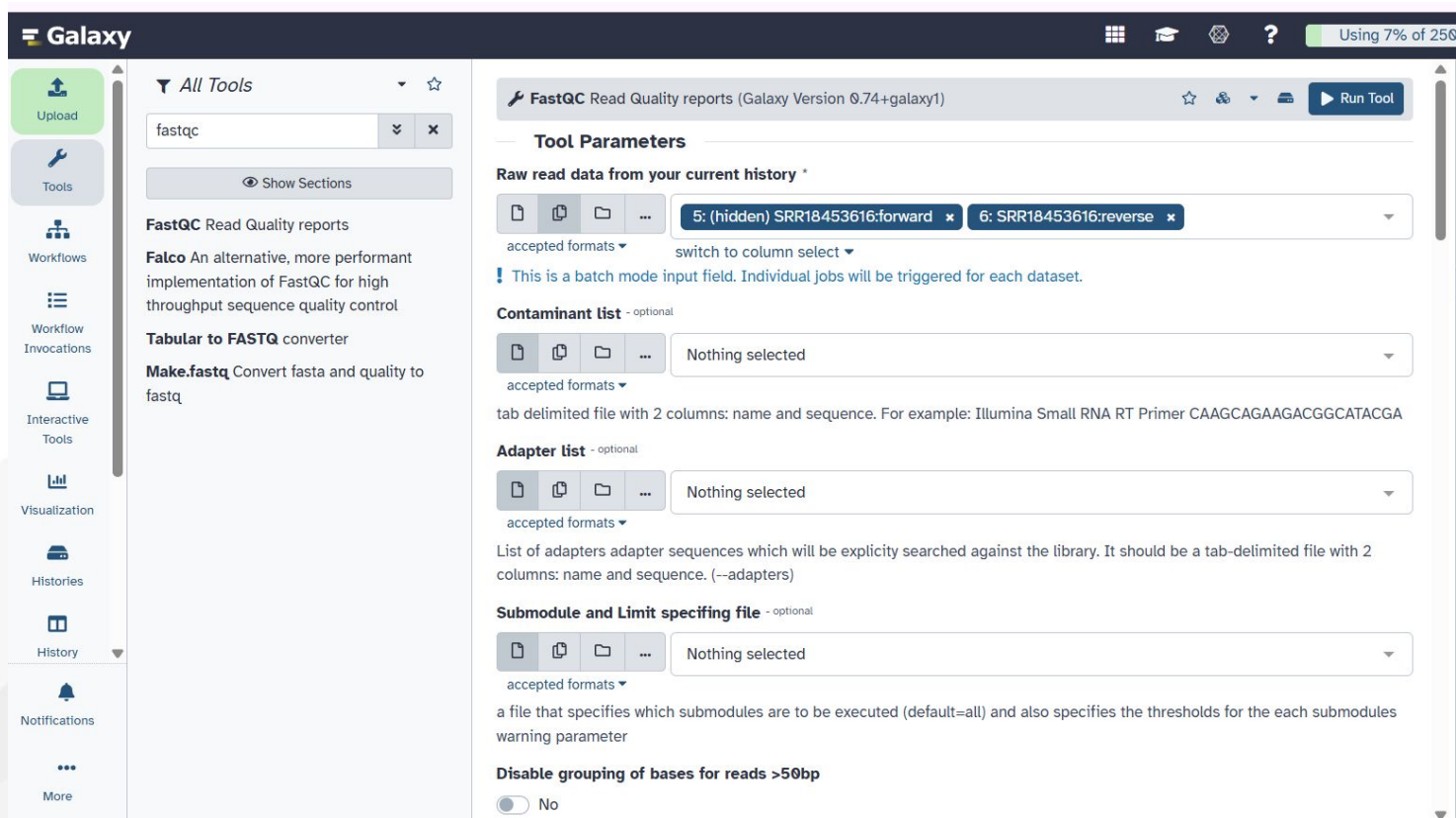
Quality Control in Galaxy

Quality Control in Galaxy

Perform **QC**:

1. On the search bar, type **fastqc**.

2. **Choose** the desired fastq file (paired end) in the raw read tab.

3. **Leave** all other tabs unchanged.

Quality Control in Galaxy

Quality Control in Galaxy

4. Once it runs, two files are generated, a raw data file and a Web Page file.

5. **View** the result by clicking on the web page file produced.

Quality Control Result

Quality Control (MultiQC)

Why: It helps us to obtain a more intuitive
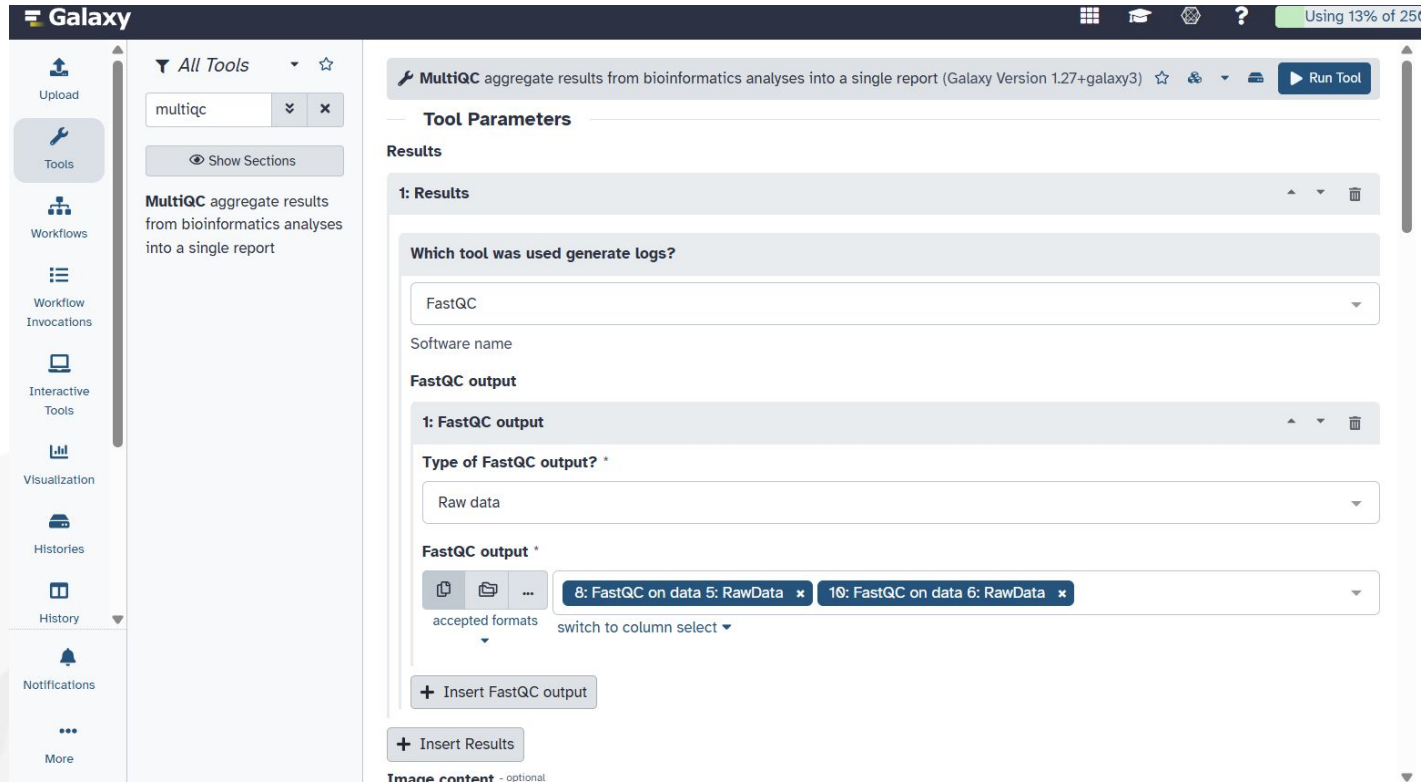
comparison

1. On the search bar, type **multiqc**

2. On the "Which tool was used generate logs?" tab,

    choose **Fastqc**

3. Then click on "**Insert FastQC output**"

Quality Control (MultiQC)

4. Type of output is raw data

5. Add the raw data files generated earlier

6. **Leave** all other parameters at default

7. **Run** tool

8. **View** the result by clicking on the web page file produced

Quality Control (MultiQC)

Quality Control Result

Why QC is Important

- Ensures data integrity before downstream analysis.

- Detects contamination, errors, or poor-quality sequences.

- Prevents misleading results.