



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

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

All Databases ▾ lcc9 gene Search

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

NCBI
The National Library of Medicine advances science and health by providing access to genomic information.
[Home](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit
Submit manuscripts and databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Build and code applications

Analyze
Identify an NCBI tool for your data analysis task

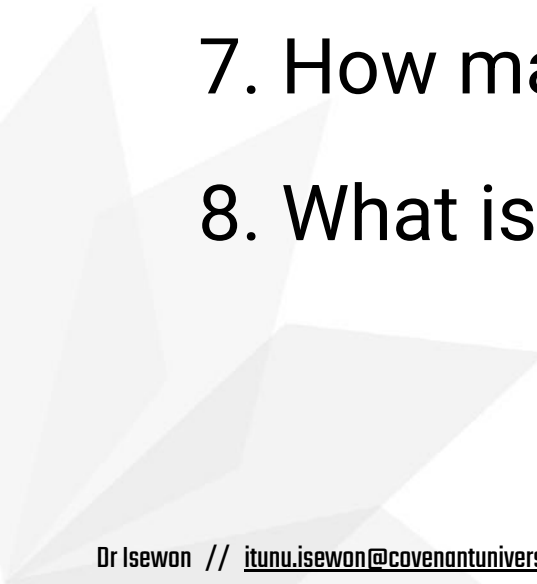
Research
Explore NCBI research and collaborative projects

Popular Resources
[PubMed](#)
[Bookshelf](#)
[PubMed Central](#)
[BLAST](#)
[Nucleotide](#)
[Genome](#)
[SNP](#)
[Gene](#)
[Protein](#)
[PubChem](#)


NCBI News & Blog
[Upcoming Changes to GenBank Project Lists and Symlinks](#)
20 Aug 2025
The volume of GenBank WGS, TSA, and T1 S data sets is large and continues to grow.
[PubMed Central \(PMC\) Beta Search to Replace Current Search](#)
19 Aug 2025
Coming the week of September 7th The

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

 **National Library of Medicine**
National Center for Biotechnology Information

Log in

Nucleotide

Nucleotide

lcc9

Search

Create alert Advanced

Help

Species
Animals (4)
Fungi (21)
Bacteria (1)
Viruses (1)
Customize ...

Molecule types
genomic DNA/RNA (19)
mRNA (8)
Customize ...

Source databases
INSDC (GenBank) (20)
RefSeq (6)
Customize ...

Sequence Type
Nucleotide (27)

Genetic compartments
Mitochondrion (4)
Custom range...

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

Summary 20 per page Sort by Default order Send to: Filters: [Manage Filters](#)

See [lcc9 laccase-9 precursor](#) in the Gene database
[lcc9](#) reference sequences [Protein](#) (1)

Items: 1 to 20 of 27

<< First < Prev Page 1 of 2 Next > Last >>

☐ [Fusarium oxysporum strain 4287 Lcc9 \(lcc9\) gene, complete cds](#)
1. 3,669 bp linear DNA
Accession: EF990898.1 GI: 152013647
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Neurospora crassa OR74A Lcc9 \(NCU04593\), partial mRNA](#)
2. 1,635 bp linear mRNA
Accession: XM_952834.3 GI: 758995457
[BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Neurospora crassa OR74A linkage group V, whole genome shotgun sequence](#)
3. 6,436,246 bp linear DNA
Accession: NC_026505.1 GI: 759001908
[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

Results by taxon
Top Organisms [\[Tree\]](#)
[Laccaria bicolor](#) (4)
[Neurospora crassa OR74A](#) (3)
[Lenitula edodes](#) (3)
[Agaricus bisporus var. bisporus H97](#) (3)
[Triphophysa robusta](#) (3)
All other taxa (11)
[More...](#)

Find related data
Database: [Select](#)
[Find items](#)

Search details
lcc9[All Fields]
[Search](#)


Activate Windows
Go to Settings to activate Windows.
[See more...](#)

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

 **National Library of Medicine**
National Center for Biotechnology Information

Log in

Nucleotide Nucleotide Search

Advanced Help

GenBank ▼

Fusarium oxysporum strain 4287 Lcc9 (lcc9) gene, complete cds

GenBank: EF990898.1

[FASTA](#) [Graphics](#)

Go to: ▼

LOCUS	EF990898	3669 bp	DNA	linear	PLN 22-JUL-2007
DEFINITION	Fusarium oxysporum strain 4287 Lcc9 (lcc9) gene, complete cds.				
ACCESSION	EF990898				
VERSION	EF990898.1				
KEYWORDS	.				
SOURCE	Fusarium oxysporum				
ORGANISM	Fusarium oxysporum Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Fusarium; Fusarium oxysporum species complex.				
REFERENCE	1 (bases 1 to 3669)				
AUTHORS	Cordoba C.D. and Roncero,M.I.G.				
TITLE	Isolation and characterization of laccases genes from the plant pathogenic fungus Fusarium oxysporum				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 3669)				
AUTHORS	Cordoba C.D. and Roncero,M.I.G.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-JUN-2007) Genetic, University of Cordoba, Campus de Rabanales, Cordoba E-14071, Spain				
FEATURES	Location/Qualifiers				
source	1..3669				

☒ Complete Record
☐ Coding Sequences
☐ Gene Features

Choose Destination
☒ File
☐ Clipboard
☐ Collections
☐ Analysis Tool

Download 1 item.
Format
FASTA
Show GI ☐
Create File

Send to: ▼

Annotations shown

Annotations

Sequence

Gene Features

Gene

Protein

Taxonomy

Recent activity

Fusarium oxysporum strain 4287 Lcc9 (lcc9) gene, complete cds

Activated

lcc9 (27)

Go to Settings

Nucleotide

Dr Isewon // itunu.isewon@covenantuniversity.edu.ng

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

The screenshot displays the Galaxy web interface. On the left is a vertical sidebar with navigation icons and labels: Upload, Tools, Workflows, Workflow Invocations, Interactive Tools, Visualization, Histories, History, Notifications, and More. The main content area is titled 'All Tools' and contains a search bar. Below the search bar, a list of tools is shown under the heading 'Get Data'. The tool 'Faster Download and Extract Reads in FASTQ format from NCBI SRA' is selected and expanded. The tool's configuration page shows the following details:

- Tool Parameters:**
 - select input type:** A dropdown menu with 'SRR accession' selected.
 - Accession *:** A text input field containing 'SRR18453616'. Below it, a note states: 'Must start with SRR, DRR or ERR, e.g. SRR925743, ERR343809'.
- Advanced Options:** A section with a downward arrow.
- Additional Options:**
 - Email notification:** A toggle switch set to 'No'. Below it, text reads: 'Send an email notification when the job completes.'
 - Attempt to re-use jobs with identical parameters?:** A toggle switch set to 'No'. Below it, text reads: 'This may skip executing jobs that you have already run.'

At the bottom of the tool configuration area, there is a 'Run Tool' button and a 'Help' link.

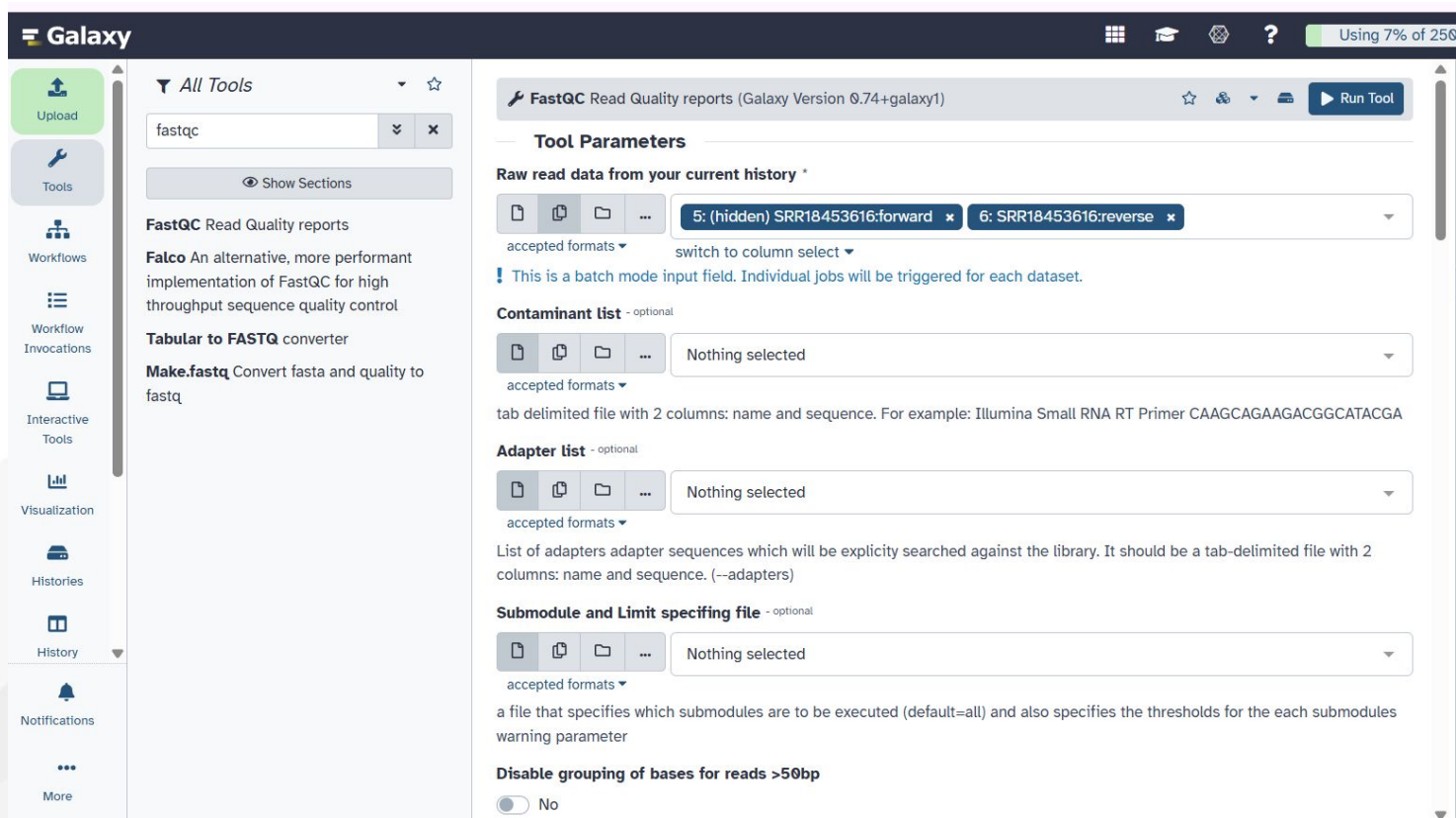
Quality Control in Galaxy



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



The screenshot displays the Galaxy web interface for configuring the FastQC tool. The left sidebar contains navigation links: Upload, Tools, Workflows, Workflow Invocations, Interactive Tools, Visualization, Histories, History, and Notifications. The main panel shows the 'FastQC Read Quality reports (Galaxy Version 0.74+galaxy1)' tool configuration. The 'Tool Parameters' section includes:

- Raw read data from your current history ***: A dropdown menu showing selected datasets: '5: (hidden) SRR18453616:forward' and '6: SRR18453616:reverse'. Below the dropdown are links for 'accepted formats' and 'switch to column select'. A warning message states: 'This is a batch mode input field. Individual jobs will be triggered for each dataset.'
- Contaminant list - optional**: A dropdown menu currently set to 'Nothing selected'. Below it are links for 'accepted formats' and a description: 'tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA'.
- Adapter list - optional**: A dropdown menu currently set to 'Nothing selected'. Below it are links for 'accepted formats' and a description: 'List of adapters adapter sequences which will be explicitly searched against the library. It should be a tab-delimited file with 2 columns: name and sequence. (--adapters)'.
- Submodule and Limit specifying file - optional**: A dropdown menu currently set to 'Nothing selected'. Below it are links for 'accepted formats' and a description: 'a file that specifies which submodules are to be executed (default=all) and also specifies the thresholds for the each submodules warning parameter'.
- Disable grouping of bases for reads >50bp**: A toggle switch currently set to 'No'.

The top right of the interface shows a status bar with 'Using 7% of 250'.

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

Summary

- ✓ [Basic Statistics](#)
- ✓ [Per base sequence quality](#)
- ✓ [Per sequence quality scores](#)
- ✗ [Per base sequence content](#)
- ✓ [Per sequence GC content](#)
- ✓ [Per base N content](#)
- ✓ [Sequence Length Distribution](#)
- ✗ [Sequence Duplication Levels](#)

✓ Per base sequence quality



Quality Control (MultiQC)



Why: It helps us to obtain a more intuitive comparison

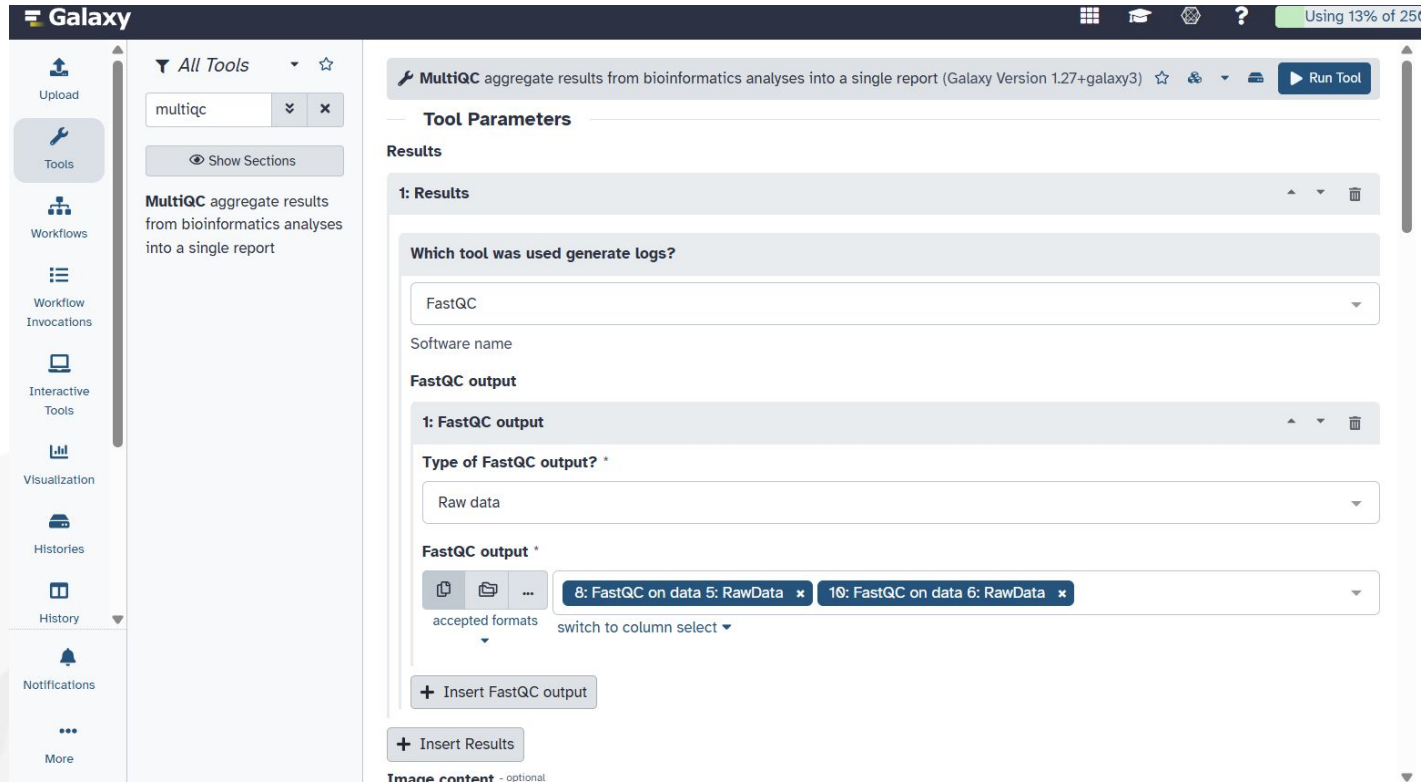
1. On the search bar, type **multiqc**
2. On the “Which tool was used generate logs?” tab, choose **Fastqc**
3. Then click on “**Insert FastQC output**”

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)



The screenshot displays the MultiQC tool interface within the Galaxy web portal. The top navigation bar includes the Galaxy logo, a search bar, and a status indicator showing 'Using 13% of 250 MB'. The left sidebar contains navigation links for Upload, Tools, Workflows, Workflow Invocations, Interactive Tools, Visualization, Histories, History, and Notifications. The main content area is titled 'MultiQC' and includes a search bar with 'multiqc' entered. Below the search bar, a description states: 'MultiQC aggregate results from bioinformatics analyses into a single report'. The 'Tool Parameters' section is currently empty. The 'Results' section shows a list of results, with the first result being '1: Results'. This result is expanded, showing a dropdown menu for 'Which tool was used generate logs?' set to 'FastQC'. Below this, the 'FastQC output' section shows a dropdown menu for 'Type of FastQC output?' set to 'Raw data'. The 'FastQC output' section also displays a list of outputs: '8: FastQC on data 5: RawData' and '10: FastQC on data 6: RawData'. At the bottom of the results section, there are buttons for '+ Insert FastQC output' and '+ Insert Results'.

Quality Control Result

The screenshot displays the Galaxy web interface. On the left is a sidebar with navigation icons for Upload, Tools, Workflows, Workflow Invocations, Interactive Tools, Visualization, Histories, History, Notifications, and More. The main area shows the 'MultiQC' tool results for a dataset named '11: MultiQC on data 8 and data 10: Webpage'. The tool version is v1.27. The report is generated on 2025-08-25, 03:54 UTC based on data in: /corral4/main/jobs/070/149/70149916/working/multiqc_WDir. The report includes sections for General Stats, FastQC, Sequence Counts, Sequence Quality Histograms, Per Sequence Quality Scores, Per Base Sequence Content, Per Sequence GC Content, Per Base N Content, Sequence Length Distribution, Sequence Duplication Levels, Overrepresented sequences by sample, and Top overrepresented sequences. A 'General Statistics' section is also visible, showing a table with columns: Sample Name, Dups, GC, and Seqs. The table is currently showing 2/2 rows and 3/6 columns.

Galaxy Using 13% of 25

11: MultiQC on data 8 and data 10: Webpage ok

4.7 MB
format **html** database ? size **4.7 MB**

MultiQC aggregate results from bioinformatics analyses into a single report

multiqc v1.27

General Stats

FastQC

Sequence Counts

Sequence Quality Histograms

Per Sequence Quality Scores

Per Base Sequence Content

Per Sequence GC Content

Per Base N Content

Sequence Length Distribution

Sequence Duplication Levels

Overrepresented sequences by sample

Top overrepresented sequences

multiqc v1.27

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2025-08-25, 03:54 UTC based on data in: /corral4/main/jobs/070/149/70149916/working/multiqc_WDir

Welcome! Not sure where to start? don't show again

Watch a tutorial video (6:06)

General Statistics

Copy table Configure columns Scatter plot Violin plot Export as CSV...

Showing 2/2 rows and 3/6 columns.

Summarize table

Sample Name	Dups	GC	Seqs
-------------	------	----	------

https://usegalaxy.org/datasets/f9cad7b01a47213585b73fee9118e9d4

Why QC is Important

- Ensures data integrity before downstream analysis.
- Detects contamination, errors, or poor-quality sequences.
- Prevents misleading results.