National Institute of Technology Andhra Pradesh
Tadepalligudem, Andhra Pradesh – 534101
Department of Biotechnology

Workshop Report: Next-Generation Sequencing (NGS) Hands-on Training

Date: September 12th – 13th, 2025

Resource Person: Dr. Srinivas Bandaru (Associate Professor, Department of Biotechnology, KL University)

Platform Used: Galaxy (web-based bioinformatics workbench)

Organized by : Training and placement cell in association with BEA

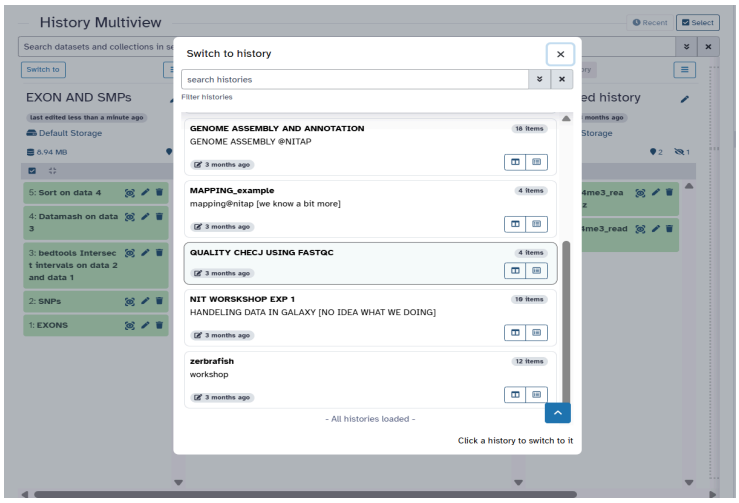
Executive Summary

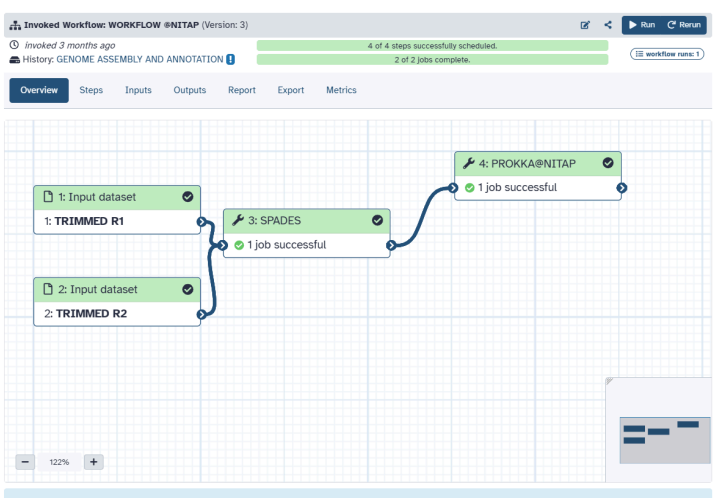
A comprehensive two-day hands-on workshop on **Next-Generation Sequencing (NGS)** data analysis was conducted by Dr. Srinivas Bandaru. The training focused on bridging the gap between raw biological data and biological insight using the **Galaxy platform**. Participants were guided through various bioinformatics pipelines, including genome assembly, functional annotation, and exploratory data analysis.

Introduction to the Galaxy Platform

The workshop began with an introduction to **Galaxy**, an open-source, web-based platform designed for accessible, reproducible, and transparent computational biomedical research.

- **Accessibility:** It allows researchers without extensive programming or command-line experience to perform complex bioinformatics tasks.
- **Reproducibility:** Every step of the analysis is recorded in a "History", ensuring that experiments can be shared and replicated exactly.
- **Workflow Integration:** It enables the linking of multiple tools into a single automated pipeline.





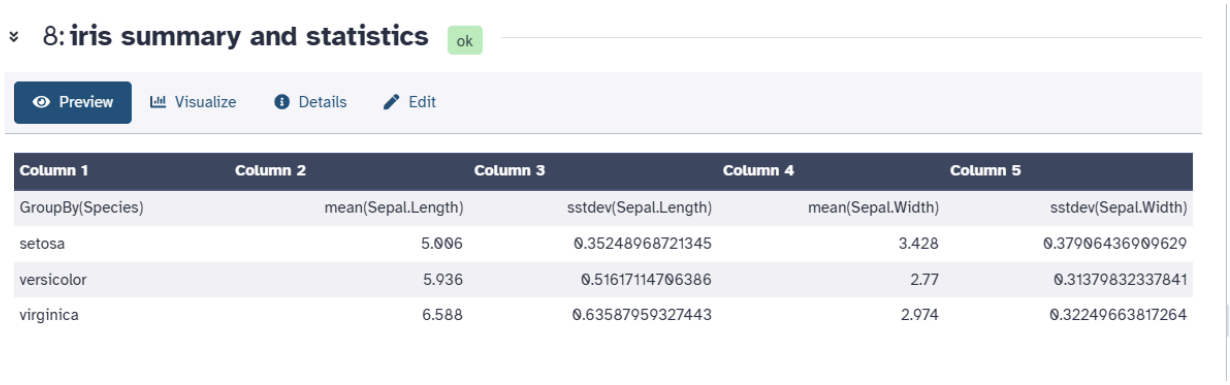
Day 1: Data Pre-processing and Exploratory Analysis

The initial sessions focused on data handling and quality control.

1. Handling Data in Galaxy (EXP 1)

We utilized the classic **Iris dataset** to learn the fundamentals of the Galaxy interface:

- **Data Cleaning:** Managing tabular data and filtering specific columns.
- **Statistical Analysis:** Generating summary statistics to understand data distribution.
- **Visualization:** Creating scatterplots using the **ggplot2** library to visualize relationships between biological features.



8: iris summary and statistics ok

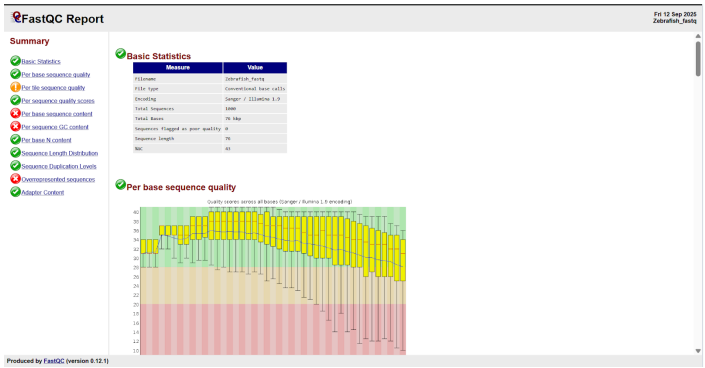
Preview | Visualize | Details | Edit

Column 1	Column 2	Column 3	Column 4	Column 5
GroupBy(Species)	mean(Sepal.Length)	sstdev(Sepal.Length)	mean(Sepal.Width)	sstdev(Sepal.Width)
setosa	5.006	0.35248968721345	3.428	0.37906436909629
versicolor	5.936	0.51617114706386	2.77	0.31379832337841
virginica	6.588	0.63587959327443	2.974	0.32249663817264

2. Quality Control (QC) and Trimming

Analyzing raw NGS data requires verifying the quality of the sequences before assembly.

- **Tool Used: FastQC**.
- **Process:** We processed multiple datasets (e.g., Zebrafish.fastq, mutant_R1.fastq) to assess per-base quality scores and GC content.
- **Outcome:** Generation of Webpage reports for each dataset to identify adapter contamination or low-quality reads.

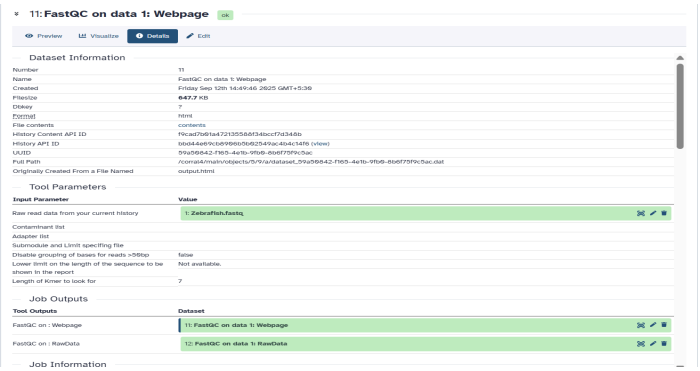


FastQC Report

Summary

Basic Statistics

Per base sequence quality



11: FastQC on data 1: Webpage ok

Preview | Visualize | Details | Edit

Dataset Information

Tool Parameters

Job Outputs

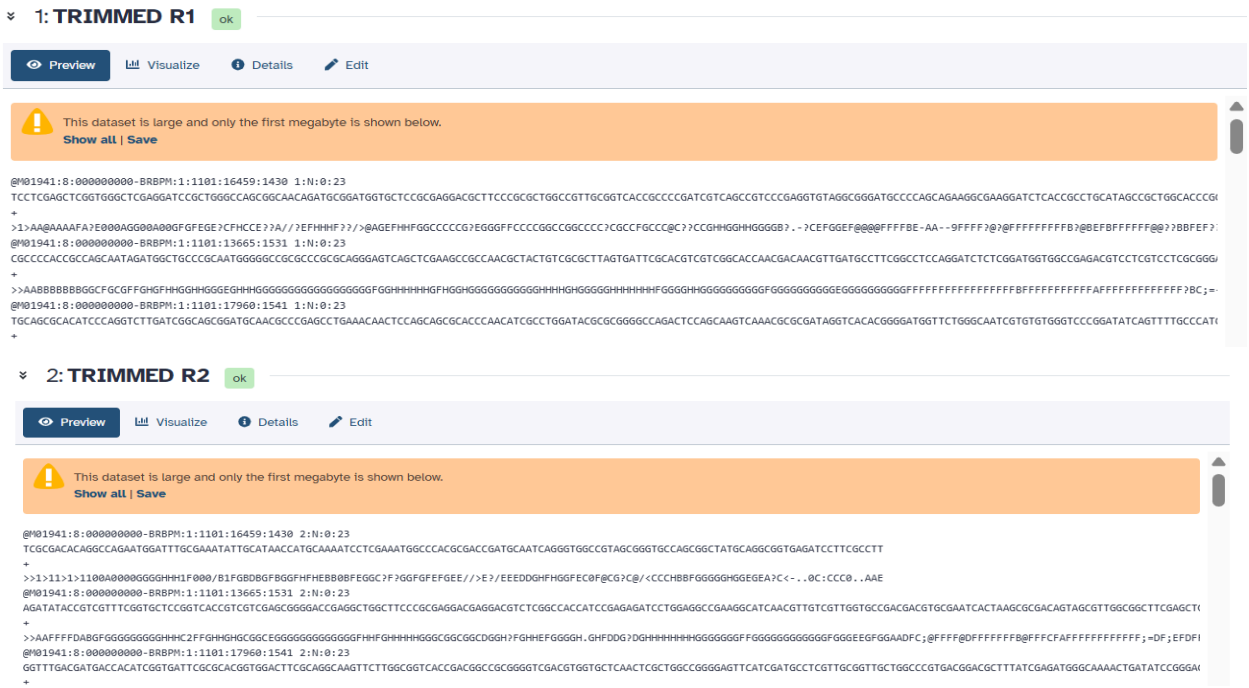
Day 2: Genome Assembly, Annotation, and SNP Analysis

The second day transitioned into advanced genomic workflows.

1. De Novo Genome Assembly

We performed assembly using the **SPAdes** (St. Petersburg Genome Assembler) tool.

Input: Paired-end reads (Trimmed R1 and R2).

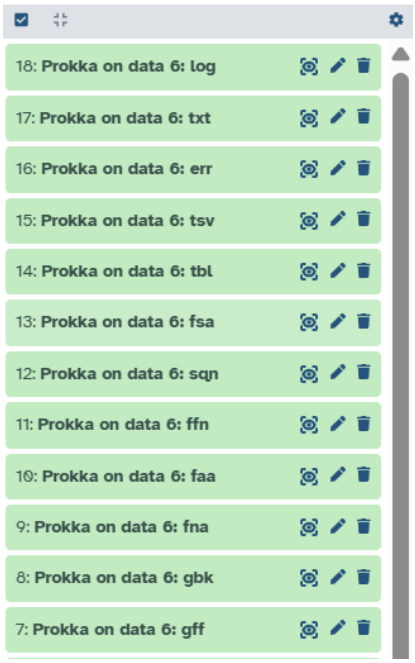


Outputs: The process generated assembly graphs, **Contigs**, and **Scaffolds**, representing the reconstructed genome structure.

2. Functional Genome Annotation

Once assembled, the genome needs to be "read" to identify genes and functional elements.

- **Tool Used: Prokka**.
- **Outputs:** Prokka provided a suite of files including:
  - .fna: Nucleotide FASTA file.
  - .faa: Protein FASTA file (for amino acid sequences).
  - .gff: General Feature Format (locations of genes).
  - .txt: Statistics and summary of the annotation.



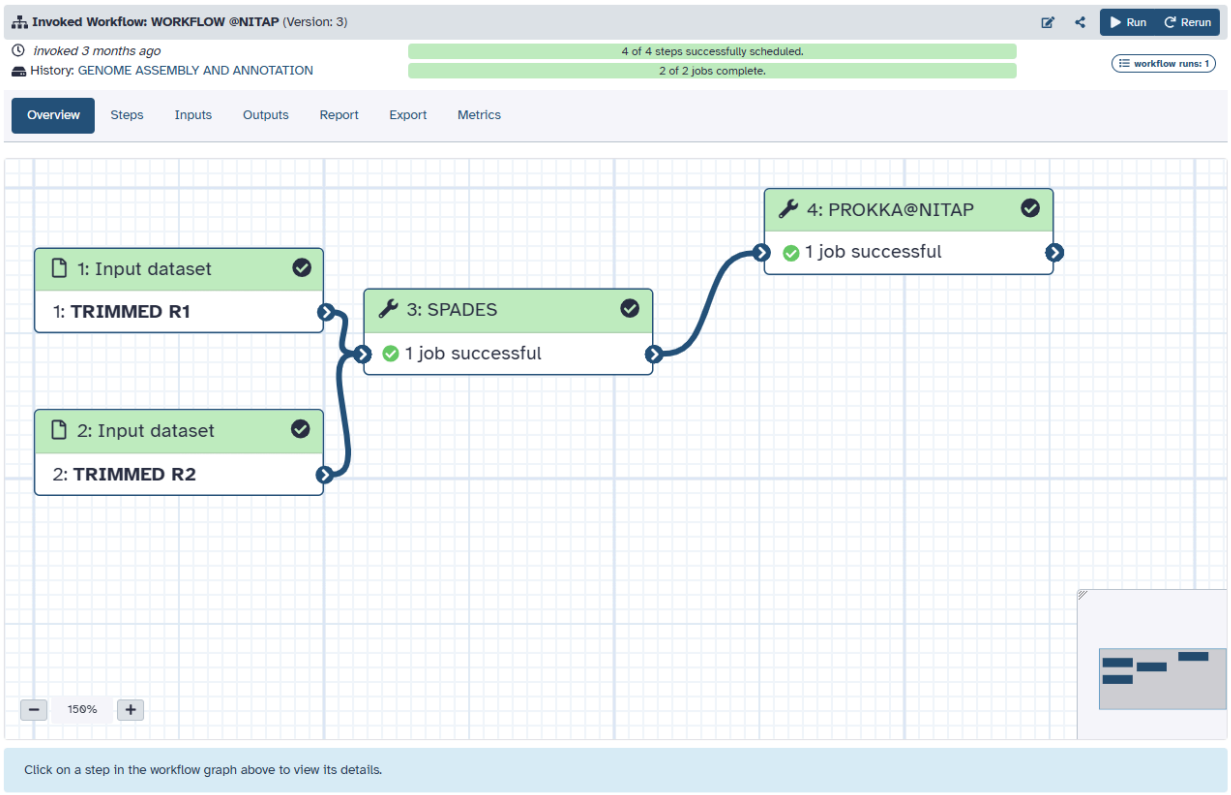
3. SNP Analysis and Mapping

The final technical module covered identifying genetic variations.

- **Objective:** Mapping reads to a reference genome to identify Single Nucleotide Polymorphisms (SNPs).
- **Significance:** Crucial for understanding mutations in mutant strains compared to wild-type (WT) samples.

Key Takeaways and Learning Outcomes

- **Workflow Mastery:** Successfully invoked complex workflows (e.g., WORKFLOW @NITAP) that automated the transition from Trimmed Reads →SPAdes→ Prokka.



- **Data Interpretation:** Learned to distinguish between raw data (FASTQ) and processed genomic outputs (FASTA, GFF)
- **Tool Proficiency:** Gained hands-on experience with industry-standard tools like FastQC, SPAdes, and Prokka without needing to write code.

Images of day 2

Conclusion

The workshop provided a robust foundation in NGS data logistics. Under the guidance of Dr. Srinivas Bandaru, we successfully navigated the complexities of genomic assembly and annotation, transforming raw sequencing reads into annotated files ready for downstream biological interpretation.