MODULE 1

Foundations for Bioinformatics:

Python, Linux, R Programming for Biological Data

Build a strong computational foundation by learning how to work with real biological datasets in a research-oriented environment.

Module Duration

17 July 2026 – 30 July 2026

Class Timings: 10:00 AM – 03:00 PM

Week 1 Classes (Python & Linux/Bash)

17 July 2026

18 July 2026

19 July 2026

Independent Assignment Period

20 July 2026 – 23 July 2026

Participants will independently complete a Linux and Python-based data handling exercise using real biological files and submit a short report.

Week 2 Classes (R Programming & Data Visualization) 24

July 2026

25 July 2026

26 July 2026

Capstone Project Period

27 July 2026 – 30 July 2026

Participants will integrate Python, Linux, and R skills to process, analyze, and visualize a biological dataset and present their findings.

Total Learning Commitment

a) 6 Live Interactive Classes
b) 18 Hours of Instructor-Led Training
c) 2 Independent Research Projects
d) Real Biological Datasets
e) Research-Oriented Assignments
f) Project Presentation & Evaluation

Week 1: Python & Linux Foundations

Theme:

From the Command Line to Your First Biological Script

Participants will learn to navigate Linux environments, write Bash and Python scripts, and handle common biological file formats.

Class 1

Introduction to Linux & the Command Line

Topics

- Introduction to Linux operating systems
- File system structure and navigation
- Basic Linux commands (file/directory operations, permissions)
- Working with text files (cat, head, tail, grep, sed, awk)
- Pipes, redirection, and wildcards
- Introduction to Bash scripting
- Writing and executing simple shell scripts

Hands-on

- Navigating a Linux file system
- Practicing core commands on biological data files
- Writing a basic Bash script to automate a file-handling task

Class 2

Python Programming Fundamentals

Topics

- Introduction to Python for biological data
- Variables, data types, and operators

- Control flow (loops, conditionals)
- Functions and modules
- Working with strings, lists, dictionaries
- File handling in Python
- Introduction to Biopython

Hands-on

- Writing scripts to parse biological file formats
- Automating simple sequence-handling tasks
- Practicing with Biopython basics

Class 3

Biological File Formats & Data Handling

Topics

- Overview of common biological file formats (FASTA, FASTQ, BED, GFF/GTF, VCF, SAM/BAM)
- Reading, parsing, and manipulating biological files
- Combining Bash and Python for data processing pipelines
- Introduction to data cleaning and organization
- Best practices for reproducible research workflows

Hands-on

- Parsing and extracting information from FASTA/FASTQ files
- Writing a combined Bash + Python workflow
- Organizing a sample project directory structure

Week 1 Assignment & Project

Independent Linux & Python Data Handling Project

Each participant will receive a set of biological data files and a defined set of processing tasks.

Tasks

1. Navigate and organize the provided dataset using Linux commands
2. Write Bash scripts to automate file operations
3. Write Python scripts to parse and summarize biological file formats
4. Extract specific information (e.g., sequence counts, read lengths, feature counts)
5. Document the workflow and outputs

Deliverables

- Bash scripts used for file handling
- Python scripts with parsing/summary outputs
- Short written report describing the workflow and results

Expected Outcome

Participants will independently navigate a Linux environment and use Python and Bash scripting to process and summarize real biological data files.

Week 2: R Programming & Data Visualization

Theme:

From Data Tables to Biological Insight

Participants will learn R programming fundamentals and apply visualization techniques to explore and interpret biological data.

Class 4

R Programming Fundamentals

Topics

- Introduction to R and RStudio
- Data types and data structures (vectors, data frames, lists)
- Importing and exporting biological data in R
- Data manipulation basics
- Introduction to the tidyverse (dplyr, tidyr)
- Writing simple functions in R

Hands-on

- Loading and exploring a biological dataset in R
- Performing basic data manipulation with dplyr
- Writing simple custom functions

Class 5

Data Wrangling & Exploratory Analysis

Topics

- Data cleaning and reshaping

- Filtering, grouping, and summarizing data
- Merging and joining datasets
- Handling missing data
- Descriptive statistics for biological data
- Introduction to exploratory data analysis (EDA)

Hands-on

- Cleaning and reshaping a real biological dataset
- Performing grouped summaries and descriptive statistics
- Conducting exploratory analysis on sample data

Class 6

Data Visualization & Interpretation

Topics

- Principles of effective data visualization
- Introduction to ggplot2
- Creating bar plots, scatter plots, histograms, and boxplots
- Visualizing distributions and group comparisons
- Customizing plots for publication quality
- Interpreting visual patterns in biological data

Hands-on

- Building a variety of plots using ggplot2
- Customizing themes, labels, and aesthetics
- Interpreting visualizations to draw biological conclusions

Week 2 Assignment & Project

Integrated Computational Foundations Project

Each participant will receive a biological dataset requiring end-to-end processing.

Tasks

1. Use Linux/Bash to organize and pre-process raw data files
2. Use Python to parse and extract relevant information
3. Import the processed data into R
4. Perform data wrangling and exploratory analysis in R
5. Generate visualizations to highlight key patterns
6. Interpret findings in biological context
7. Prepare a short presentation summarizing the workflow and results

Deliverables

- Combined Bash/Python processing scripts
- R script/notebook with wrangling and visualization code
- Set of generated plots
- Final presentation summarizing the workflow and findings

Expected Outcome

Participants will independently complete an end-to-end data handling and visualization workflow spanning Linux, Python, and R.

Module Outcomes

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

a) Work confidently in Linux-based environments using the command line and Bash scripting
b) Write basic Python scripts to handle and parse biological file formats
c) Use R and the tidyverse to wrangle and explore biological datasets
d) Create publication-quality visualizations using ggplot2
e) Build integrated Bash/Python/R workflows for biological data processing
f) Interpret visualized biological data to identify meaningful patterns
g) Present computational workflows and findings in a research-ready format