**Comprehensive Bioinformatics Training Program**

*Duration: 3-4 Months | Weekly Sessions: 2 Hours | Format: Hands-on Learning*

**Program Overview**

This beginner-friendly training program is specifically designed for microbiologists with no prior bioinformatics or computer science background. The curriculum focuses on practical, hands-on learning with weekly 2-hour sessions over 3-4 months, covering essential bioinformatics skills from basic computing concepts to advanced genomic analysis.

**Program Structure**

**Month 1: Foundation and Environment Setup**

**Week 1: Computing Infrastructure Basics**

**Duration:** 2 hours  
**Focus:** Understanding the computational backbone of bioinformatics

**Theoretical Content (30 minutes):**

* Introduction to bioinformatics and its role in microbiology
* Operating systems: Linux/Unix vs Windows vs macOS
* Understanding servers, clusters, and cloud computing
* Introduction to github with practical examples
* File systems and directory structures

**Hands-on Activities (90 minutes):**

* Setting up a Linux environment (VirtualBox or WSL)
* Basic navigation: pwd, ls, cd, mkdir
* File operations: cp, mv, rm, touch
* Text viewing: cat, less, head, tail
* Introduction to github console tour

**Assignment:**  
Create a directory structure for a mock project and practice basic file operations. GitHub account creation and push to code to repo.

**Week 2: Environment Management and Software Installation**

**Duration:** 2 hours  
**Focus:** Tool management and reproducible environments

**Theoretical Content (30 minutes):**

* Package managers: conda vs pip vs apt
* Virtual environments and their importance
* Introduction to Docker containers
* Best practices for software installation
* Version control basics

**Hands-on Activities (90 minutes):**

* Installing Miniconda/Anaconda
* Creating conda environments
* Installing bioinformatics tools (FastQC, BWA, SAMtools)
* Basic Docker commands and pulling containers
* Environment export and sharing

**Assignment:**  
Create three different conda environments for different analyses (QC, assembly, annotation) and document the process.

**Week 3: Programming Languages Overview**

**Duration:** 2 hours  
**Focus:** Introduction to essential programming languages

**Theoretical Content (30 minutes):**

* Bash scripting fundamentals
* Python vs R for bioinformatics
* Data types and variables
* When to use which language

**Hands-on Activities (90 minutes):**

* Writing simple bash scripts
* Basic Python syntax and data types
* Introduction to R and RStudio
* File input/output operations
* Simple data manipulation

**Assignment:**  
Write a bash script to organize FASTQ files, a Python script to count sequences and simple plot.

**Week 4: File Types and Best Practices**

**Duration:** 2 hours  
**Focus:** Understanding bioinformatics data formats

**Theoretical Content (30 minutes):**

* Common file formats: FASTA, FASTQ, SAM, BAM, VCF, GFF
* File naming conventions
* Data organization strategies
* Metadata management

**Hands-on Activities (90 minutes):**

* Examining different file formats
* Converting between formats
* File compression and decompression
* Creating standardized naming schemes
* Building a project directory template

**Assignment:**  
Organize a provided dataset using proper naming conventions and document the file structure with metadata descriptions.

**Month 2: NGS Technologies and Quality Control**

**Week 5: NGS Technologies Deep Dive**

**Duration:** 2 hours  
**Focus:** Understanding sequencing technologies and their applications

**Theoretical Content (45 minutes):**

* Illumina short-read sequencing
* Oxford Nanopore long-read sequencing
* PacBio sequencing technology
* Whole genome sequencing (WGS) vs targeted sequencing
* Metagenomics applications and considerations

**Hands-on Activities (75 minutes):**

* Downloading data from NCBI SRA
* Examining read quality and characteristics
* Comparing short vs long read data
* Understanding read headers and quality scores

**Assignment:**  
Download and examine bacterial WGS data and metagenomic samples, comparing their characteristics and preparing a summary report.

**Week 6: Quality Control and Preprocessing**

**Duration:** 2 hours  
**Focus:** Ensuring data quality for downstream analysis

**Theoretical Content (30 minutes):**

* Quality metrics: Phred scores, GC content, adapter contamination
* Coverage calculations and requirements
* Contamination detection principles

**Hands-on Activities (90 minutes):**

* Running FastQC on various datasets
* Interpreting quality reports
* Trimming adapters with Trimmomatic
* Quality filtering and read processing
* Calculating coverage statistics

**Assignment:**  
Process a low-quality dataset through the complete QC pipeline and document improvements at each step.

**Week 7: Contamination Assessment**

**Duration:** 2 hours  
**Focus:** Identifying and addressing contamination issues

**Theoretical Content (30 minutes):**

* Types of contamination in sequencing data
* Host contamination removal
* Cross-contamination between samples

**Hands-on Activities (90 minutes):**

* Using Kraken2 for taxonomic classification
* Generating Krona charts for visualization
* Filtering reads by taxonomy ID
* Assessing GC content distributions
* Host read removal strategies

**Assignment:**  
Analyze a contaminated sample, identify contamination sources, and clean the dataset for further analysis.

**Week 8: Biological Databases**

**Duration:** 2 hours  
**Focus:** Navigating and utilizing major biological databases

**Theoretical Content (30 minutes):**

* NCBI databases: GenBank, SRA, RefSeq
* European databases: EBI, EMBL
* Specialized databases: CARD, ResFinder
* API access and programmatic data retrieval

**Hands-on Activities (90 minutes):**

* Searching and downloading from NCBI
* Using BLAST for sequence similarity
* Accessing antimicrobial resistance databases
* Building custom databases
* Batch downloading techniques

**Assignment:**  
Create a local database of antimicrobial resistance genes and document the curation process.

**Month 3: Genomic Analysis and Specialized Tools**

**Week 9: Pathogen Genomic Analysis**

**Duration:** 2 hours  
**Focus:** Tools and workflows for pathogen characterization

**Theoretical Content (30 minutes):**

* Genome assembly principles
* Annotation strategies
* Comparative genomics approaches
* Phylogenetic analysis basics

**Hands-on Activities (90 minutes):**

* Assembling bacterial genomes with SPAdes
* Genome annotation with Prokka
* Quality assessment with QUAST
* Basic comparative analysis
* Visualization with genome browsers

**Assignment:**  
Complete assembly and annotation of a bacterial pathogen, comparing results with reference genomes.

**Week 10: Antimicrobial Resistance Analysis**

**Duration:** 2 hours  
**Focus:** Identifying and characterizing AMR genes

**Theoretical Content (30 minutes):**

* AMR mechanisms and gene families
* Databases: CARD, ResFinder, ARG-ANNOT
* Phenotype prediction from genotype

**Hands-on Activities (90 minutes):**

* Running AMRFinderPlus on assemblies
* Using ResFinder for AMR gene detection
* Analyzing resistance gene networks
* Correlating genotype with phenotype
* Creating AMR summary reports

**Assignment:**  
Analyze AMR profiles of multiple bacterial isolates and prepare a surveillance report.

**Week 11: Advanced Taxonomy and Environmental Analysis**

**Duration:** 2 hours  
**Focus:** Comprehensive taxonomic analysis and environmental microbiology

**Theoretical Content (30 minutes):**

* Kraken2 database construction
* Taxonomic classification algorithms
* Environmental vs clinical sample considerations

**Hands-on Activities (90 minutes):**

* Building custom Kraken2 databases
* Running classification on metagenomic data
* Advanced Krona visualization
* Filtering by specific taxonomic IDs
* Comparative taxonomy between samples

**Assignment:**  
Compare microbial communities from different environmental sources using taxonomic profiling.

**Week 12: Introduction to Workflows**

**Duration:** 2 hours  
**Focus:** Automated pipeline development with Nextflow

**Theoretical Content (30 minutes):**

* Workflow management systems
* Nextflow DSL2 basics
* Reproducibility and scalability
* nf-core community pipelines

**Hands-on Activities (90 minutes):**

* Installing Nextflow
* Running nf-core/bacass for bacterial assembly
* Running nf-core/mag for metagenomics
* Customizing pipeline parameters
* Understanding workflow outputs

**Assignment:**  
Run both bacterial and metagenomic pipelines on provided datasets and compare results with manual analysis.

**Month 4: Advanced Topics and Integration**

**Week 13: AI and Machine Learning in Genomics**

**Duration:** 2 hours  
**Focus:** Understanding AI applications in pathogen surveillance

**Theoretical Content (45 minutes):**

* Machine learning types: supervised, unsupervised
* Applications in genomics: classification, prediction
* AI tools for pathogen identification
* Protein language models and genomic surveillance

**Hands-on Activities (75 minutes):**

* Using AI-powered tools (AmrProfiler, MetaPhlAn)
* Basic Python machine learning with scikit-learn
* Feature extraction from genomic data
* Simple classification examples
* Interpreting AI model outputs

**Assignment:**  
Apply machine learning tools to classify bacterial isolates based on their genomic features.

**Week 14: Integrated Analysis Project**

**Duration:** 2 hours  
**Focus:** Combining all learned skills in a comprehensive project

**Hands-on Activities (120 minutes):**

* Complete analysis of a real-world dataset
* From raw reads to final report
* Integration of multiple analysis types
* Documentation and reproducibility
* Presentation preparation

**Assignment:**  
Complete a full analysis report suitable for presentation.

**Week 15: Advanced Topics and Career Development**

**Duration:** 2 hours  
**Focus:** Staying current and advancing skills

**Theoretical Content (45 minutes):**

* Keeping up with bioinformatics developments
* Building a bioinformatics portfolio
* Contributing to open-source projects
* Career paths in bioinformatics

**Hands-on Activities (75 minutes):**

* Setting up GitHub portfolios
* Contributing to documentation
* Joining bioinformatics communities
* Planning continued learning paths

**Assignment:**  
Create a professional portfolio showcasing skills developed throughout the program.

**Assessment and Evaluation**

**Weekly Assessments:**

* Practical assignments with real datasets
* Problem-solving exercises
* Code documentation and sharing
* Peer review activities

**Monthly Evaluations:**

* Comprehensive projects integrating multiple skills
* Presentation of findings
* Troubleshooting challenges
* Portfolio development

**Final Capstone Project:**

Students will complete an end-to-end analysis of a complex dataset, incorporating:

* Quality control and preprocessing
* Multiple analysis approaches
* Visualization and interpretation
* Written report and presentation

**Resources and Materials**

**Required Software:**

* Linux environment (WSL or native)
* Miniconda/Anaconda
* Docker
* Python 3.8+
* Git

**Datasets:**

* Curated bacterial isolate collections
* Metagenomic samples from various environments
* AMR surveillance datasets
* Quality control challenge datasets

**Learning Outcomes**

Upon completion of this program, participants will be able to:

1. **Set up and manage** bioinformatics computing environments
2. **Quality control and preprocess** NGS data effectively
3. **Identify and characterize** antimicrobial resistance genes
4. **Perform taxonomic classification** and contamination detection
5. **Build and execute** automated workflows
6. **Apply basic AI/ML tools** to genomic data
7. **Create reproducible analyses** with proper documentation
8. **Troubleshoot common issues** in bioinformatics pipelines
9. **Communicate findings** effectively to diverse audiences
10. **Continue learning** independently in the rapidly evolving field

**Prerequisites**

* Basic understanding of molecular biology and microbiology
* Familiarity with bacterial genetics concepts
* Access to a computer with internet connection
* Willingness to learn command-line interfaces
* Commitment to hands-on practice between sessions

**Certification**

Participants who complete all assignments and the capstone project will receive a certificate of completion, documenting their acquired skills in bioinformatics for microbiology applications.