



The
Fleming Fund
Regional Grants

Whole Genome sequencing and Bioinformatics Training

4-7 March 2025

Lilongwe, Malawi

Day 1 – Introduction to WGS and Online Bioinformatics Basics

Focus: Setting the stage, understanding sequencing technologies, performing BLAST and exploring different online tools.

Morning Session 10:00-12:30: Welcome and Overview

- **Introductions & Goals:**
 - o Welcome and participant introductions.
 - o Outline training goals.
- **WGS Overview:**
 - o Overview of whole genome sequencing (WGS) technologies and their laboratory roles.
 - o Basic bioinformatics concepts: file formats and sequence data types
 - o Basic bioinformatics operations: assembly, mapping, alignment, variant calling

Lunch 12:30-13:30

Afternoon Session 13:30-16:00: Introduction to Online Tools

- **Demonstrations**
 - o Overview of browser-based platforms such as BLAST, Galaxy, Epi2me, PathogenWatch, CGE, Assembly Server, Illumina Dragen, Terra Bio, etc.
 - o Walk-through of each platform's interface and navigation tips.
 - o Optional: brief demo of command-line workflow (e.g., BAP)
- **Hands-On Exercise:**
 - o BLAST Mystery Species Identification, and troubleshoot what went wrong between wet and dry lab
 - o Identify bacterial species using on-line platforms (BLAST, KmerFinder, Pathogen Watch) and learn to detect a common quality issue

Day 2 – Quality Control (QC) and Assembly

Focus: Understanding data quality, performing QC on raw reads, and exploring assembly theory and practice.

Morning Session 9:30-12:30: Quality Control (QC) of Raw Reads

- **Presentation:**
 - o Concepts of data quality and its impact on downstream analysis.
 - o Common issues: adapter contamination and low-quality reads.
 - o Illumina Reads QC (and 3 puzzles)
- **Demonstration:**
 - o Review examples of good and poor run reports from Oxford Nanopore Technologies (ONT)
- **Hands-On Exercises:**
 - o (Optional) install WSL and Ubuntu command-line environment
 - o Use off-line FastQC and on-line ONT QC tool to assess sample data.
 - o Review pre-loaded sample datasets and interpret QC reports.

Lunch 12:30-13:30

Afternoon Session 13:30-16:00: Assembly Theory, Exercises, and QC

- **Presentation:**
 - o Assembly Approaches:
 - Reference-based vs. de-novo assembly (including de Bruijn graphs)
 - Issues such as contamination and high contig counts.
 - o Assembly Types:
 - Long, short, and hybrid assemblies.
 - o Tools like BV-BRC for assembly.
 - o Quality Assessment:
 - QC tools such as QUAST, MultiQC, CheckM, and BUSCO.
- **Hands-On Exercises:**
 - o Generate assemblies (ideally using participants' own data) via the BV-BRC Assembly Service.
 - o Evaluate assembly quality using QC reports, using NICD report for CG samples as example.

Day 3 – Typing and AMR Analysis Using Online Platforms

Focus: Understanding antimicrobial resistance (AMR) fundamentals and using online tools for bacterial typing and AMR gene detection.

Morning Session 9:30-12:30: Typing & AMR Fundamentals

- **Presentation:**
 - Compare traditional lab-based typing methods with bioinformatics approaches.
- **Demonstration:**
 - KmerFinder, MLST, serotyping.
 - ResFinder, , PlasmidFinder, VirulenceFinder
- **Hands-On Exercise:**
 - Recap on how ResFinder/PlasmidFinder/etc work.
 - Build your own AnythingFinder with MyDBFinder, MyKmerFinder, MyKMAFinder

Lunch 12:30-13:30

Afternoon Session 13:30-16:00: Hands-On AMR Exercise

- **Review & Discussion:**
 - Revisit CGE results and discuss findings.
- **Hands-On Exercises:**
 - Use web-based tools for AMR gene detection (ResFinder, RGI-CARD, PathogenWatch, VirulenceFinder, PlasmidFinder, MEFinder)
 - Demo / Integrate data with hAMRonization
 - Discuss interpretation of results and the importance of consensus among tools (e.g., comparing ResFinder, AMRFinderPlus, CARD)

Day 4 – Bacterial Typing and Phylogenetic Analysis via Online Tools

Focus: Leveraging online platforms for bacterial typing and exploring phylogenetic relationships.

Morning Session 9:30-12:30: Bacterial Typing Overview

- **Presentation - Overview of Typing Methods:**
 - Introduction to methods such as MLST and cgMLST and their applications in tracking AMR epidemiologically.
 - Discuss SNP and cgMLST analysis: kmer-distance vs. SNPs vs genes/alleles.
- **Additional Topics:**
 - Use of multiple sequence alignment (MSA) and tree-building tools.
 - Overview of tools such as CSI Phylogeny, MinTyper, and BEAST (including rooted trees, bootstrapping, and clock/time analysis).
- **Demonstration:**
 - Live demo using Enterobase.

Lunch 12:30-13:30

Afternoon Session 13:30-16:00 Hands-On Phylogenetic Analysis & Simulated Outbreak Scenario

- **Hands-On Exercise:**
 - Participants work with sample datasets to generate and visualize phylogenetic trees using off-line GrapeTree and online tools such as Microreact, and iTOL.
- **Theory & Discussion:**
 - Interpreting tree topology and understanding outbreak clusters.
 - Discuss practical implications for AMR surveillance.
- **Simulated Outbreak Scenario**
 - **Tasks:**
 - Identify AMR genes using a web-based tool.
 - Perform bacterial typing and generate a phylogenetic tree.
 - Interpret the integrated results to propose potential clinical and epidemiological actions.
- **Wrap-Up:**
 - Q&A session and final discussion.