



Whole Genome sequencing and Bioinformatics Training

2025

In-person training for Fleming Fund Country Grants

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

Lunch 12:30-13:30

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

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

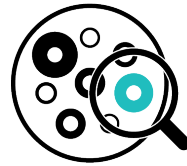
Morning Session 9:30-12:30: Quality Control (QC) of Raw Reads

- **Presentation:**
 - Concepts of data quality and its impact on downstream analysis.
 - Common issues: adapter contamination and low-quality reads.
 - Illumina Reads QC (and 3 puzzles).
- **Demonstration:**
 - Review examples of good and poor run reports from Oxford Nanopore Technologies (ONT).
- **Hands-On Exercises:**
 - (Optional) install WSL and Ubuntu command line environment.
 - Use off-line FastQC and online ONT QC tools to assess sample data.
 - 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:**
 - Assembly Approaches:
 - Reference-based vs. de-novo assembly (including de Bruijn graphs).
 - Issues such as contamination and high contig counts.
 - Assembly Types:
 - Long, short, and hybrid assemblies.
 - Tools like BV-BRC for assembly.
 - Quality Assessment:
 - QC tools such as QUAST, MultiQC, CheckM, and BUCO.
- **Hands-On Exercises:**
 - Generate assemblies (ideally using participants' own data) via the BV-BRC Assembly Service.
 - Evaluate assembly quality using QC reports.



The
Fleming Fund
Regional Grants

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/PlasminFinder/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 (e.g., 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/alles.
- **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: In-depth analysis of participants' own data using typing, AMR detection, and visualization tools.

- Data Preparation & Tool Selection
 - Review isolate metadata and perform QC/assembly if needed.
 - Select tools: EnteroBase, EPI2ME, Pathogenwatch, CGE webtools (e.g., ResFinder, MLST), and R for visualisation.
- Typing and AMR Profiling
 - Perform MLST/cgMLST and AMR gene detection (e.g., using EnteroBase, CGE tools, or local workflows).
 - Extract relevant typing and resistance results and metadata for downstream analysis.
- Phylogenetic & Clustering Analysis
 - Construct trees using EnteroBase or GrapeTree and explore them in visualisation tools such as iTOL or Microreact.
 - Annotate trees with isolates source, location, resistance profile, etc.
- Visualisation & Figure Creation
 - Use R packages (e.g., ggplot2, pheatmap, ComplexHeatmap, ggtree) to generate resistance heatmaps, barplots, and phylogenetic trees.
 - Begin drafting figures suitable for use in reports or publications.
- Wrap-Up & Discussion
 - Share results and visualisations with the group for discussion.