DTU

CGELabs: Portable Genomic Analysis for Disease Surveillance

Frederik Duus Møller

Introduction - CGE Labs

- **How do we bring genomic diagnostics to remote regions with limited infrastructure?**
  - CGE Labs: “real-time” sequencing analysis in the field
  - Developed for GREAT LIFE project
  - Pathogen detection without the need for specialized bioinformatics experts, HPCs or internet connection

CGE LABS

Oxford NANOPORE Technologies

Image: https://nanoporetech.com/software/devices/minion-mk1b

GREAT-LIFE

- Establishing capacity for mobile ONT sequencing, laptop analysis and real-time data sharing for disease detection
- Focused around the African Great Lakes region
- Partners: institutes in Tanzania, Rwanda, Burundi, Uganda, Kenya, plus partners from UK, Netherlands

Image: https://www.globalsurveillance.eu/projects/great-life

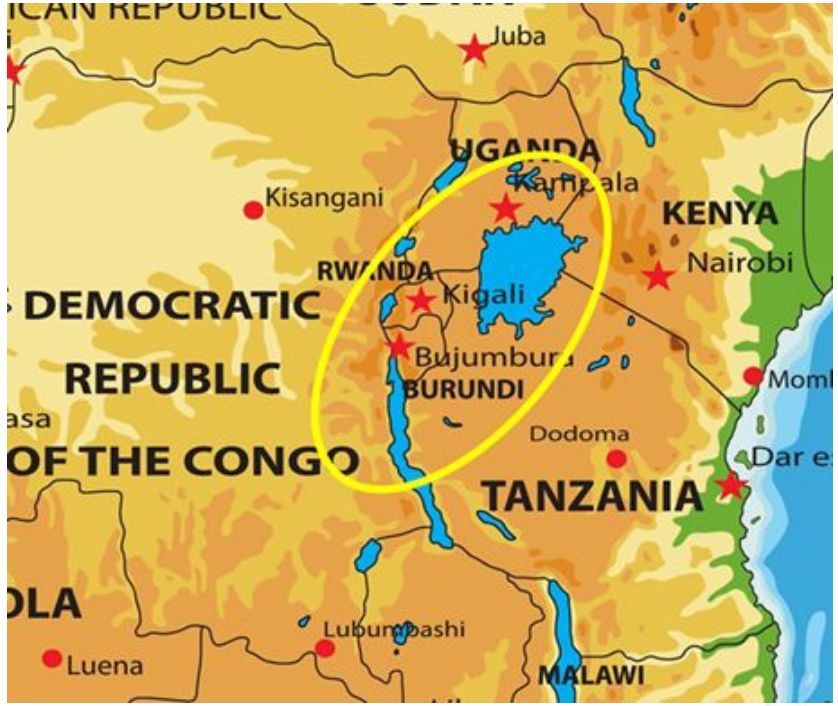
GREAT-LIFE: The challenge

- Great Lakes region:
  Dense population and high biodiversity

- Limited infrastructure for sequencing, sample transport

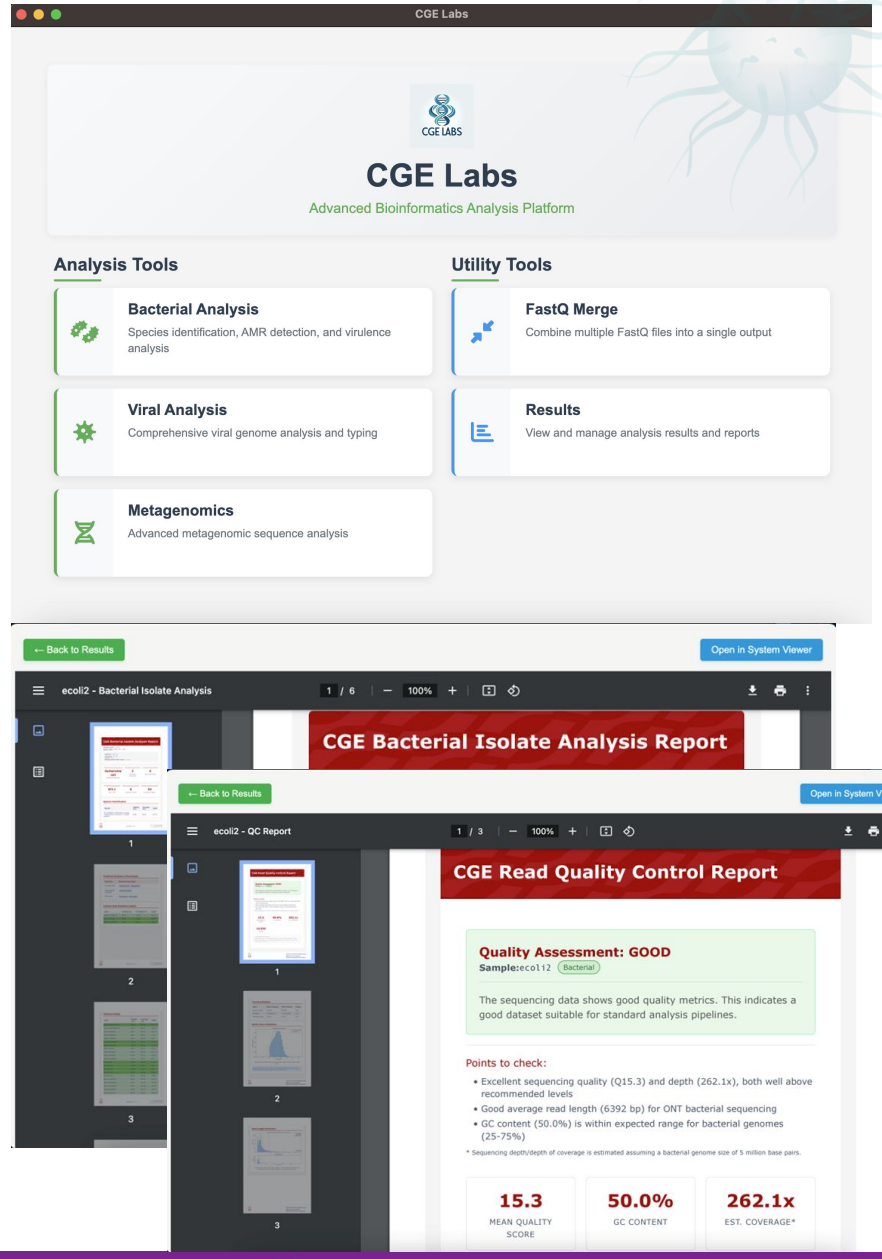
- Impossible to transfer large sequencing data over mobile networks, limited access to high-performance computing

- Need for immediate actionable results at frontline for timely intervention

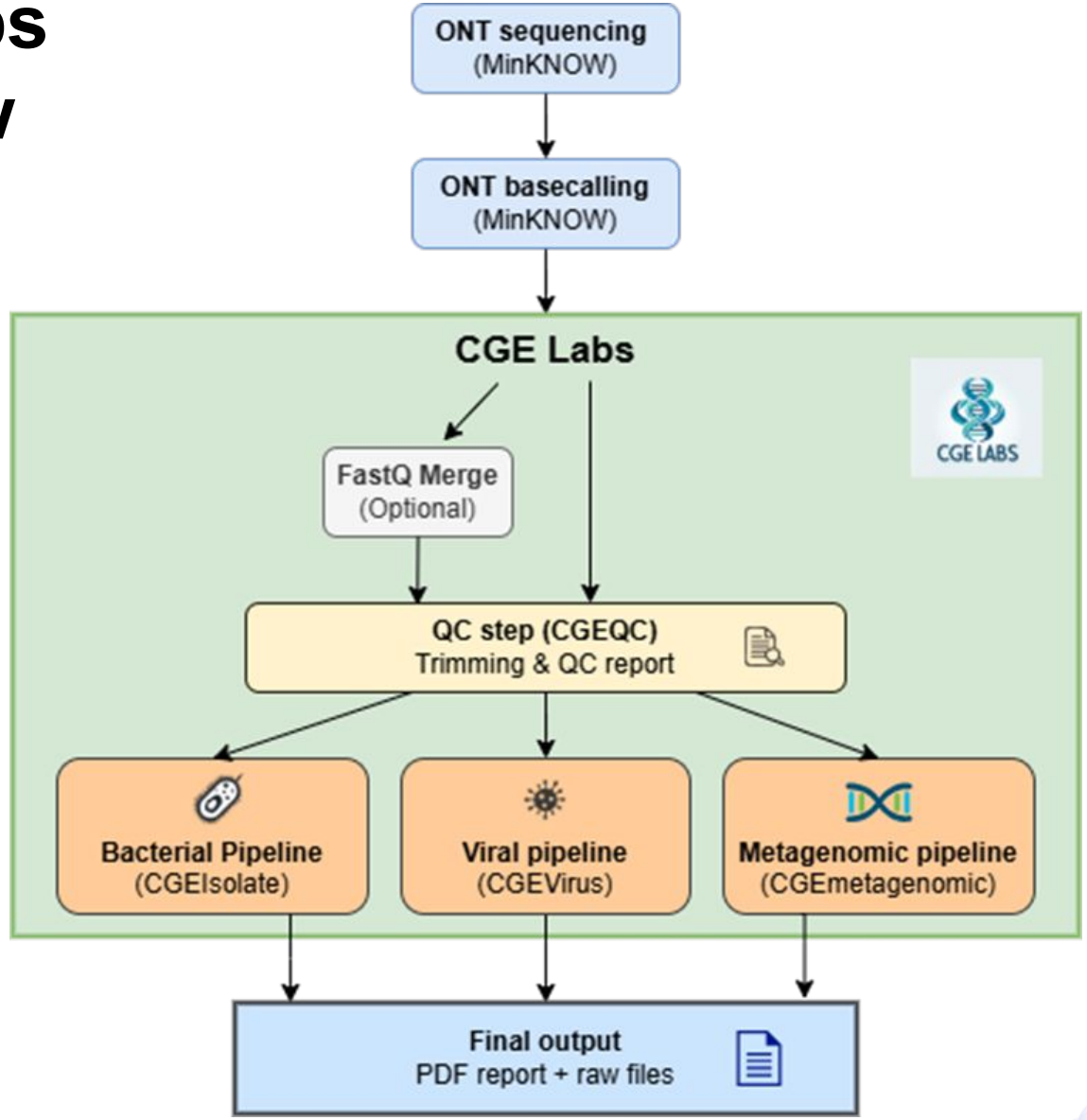


The CGE Labs solution

- Laptop-based tool for Nanopore sequence analysis
- Works offline
- User-friendly GUI, no command line required
- Analyzes bacteria, viruses, metagenomic samples
- Creates PDF reports with visualizations



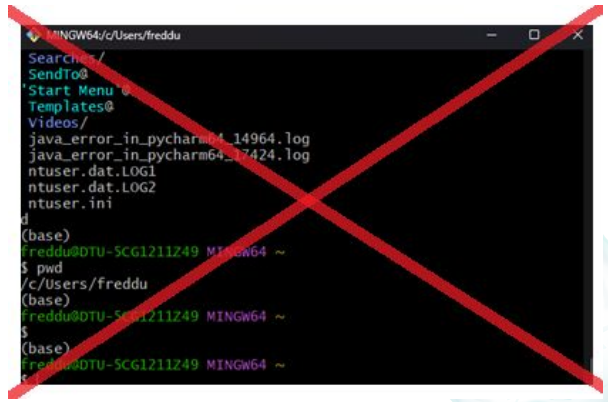
CGE Labs workflow



Demonstration

Application architecture

- Electron Application Framework:
    - Cross-platform desktop app (Linux/MacOS, later Windows)
    - Built on Chromium → Runs like a webpage inside a desktop app, same behavior across different systems
    - Uses web technologies (JavaScript, HTML, CSS) → Fast development & easy updates
    - Modular design → Easy to expand and add features
    - Modern, interactive UI → No command line required
    - Uses Conda for dependency management

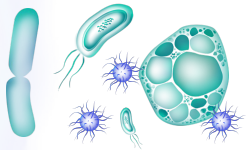


Analysis modules

- Four specialized modules and a utility tool:

- **cgeisolate**:
  species identification, AMR profiling, plasmids, virulence factors

- **cgevirus**:
  viral identification, gene annotation



- **cgemetagenomics**:
  species identification, pathogen detection, AMR profiling

- **cgeqc**:
  trims raw reads and produces a QC report

- **cgemerge**:
  utility tool for merging fastq files from Nanopore sequencings

Application architecture

- Core Analysis:
  - Uses KMA for mapping/aligning reads to reference databases
  - Custom reference databases for the three pipelines (bacterial, viral, metagenomic)

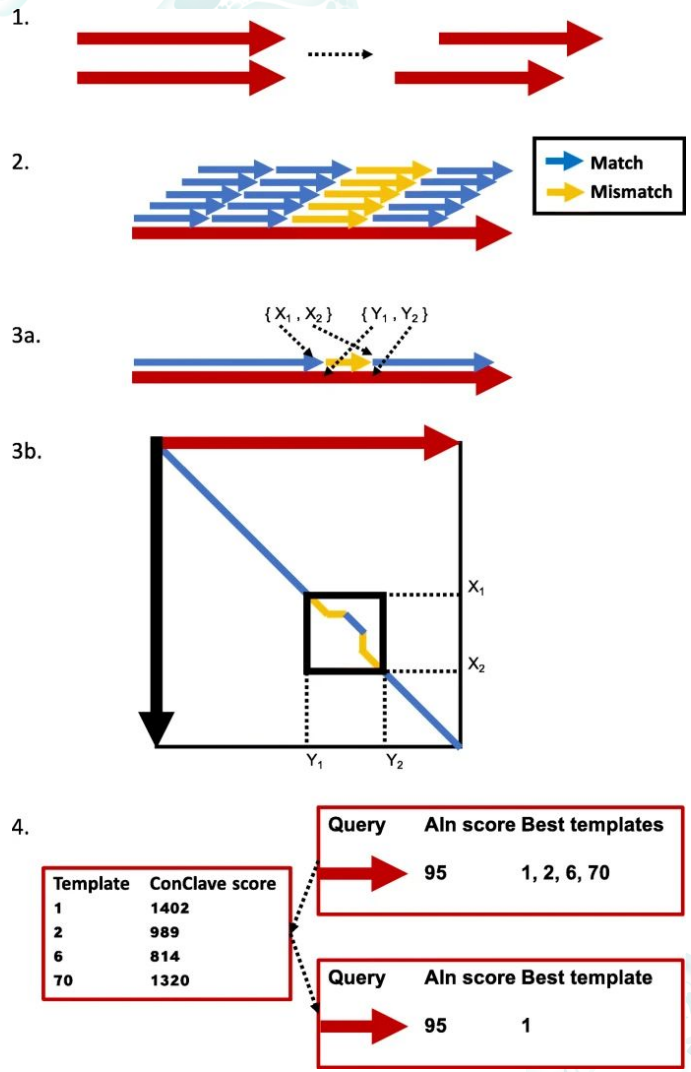


Image: Clausen, P.T.L.C., Aarestrup, F.M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 19, 307 (2018). https://doi.org/10.1186/s12859-018-2336-6

Databases

- Three custom databases being developed for the tool
  - bacteria: complete bacterial genomes from NCBI (<30 GB)
  - virus: complete viral genomes from NCBI
  - metagenomic: bacteria + virus + human genome
    - with planned expansion to fungi/parasites

Analysis modules

- All CGELabs tools are hosted on our BitBucket, where source code, updates, and documentation are maintained:

- CGE Labs:
  https://bitbucket.org/genomicepidemiology/cgelabs/wiki/Home

- CGE isolate:
  https://bitbucket.org/genomicepidemiology/cgeisolate

- CGE Virus:
  https://bitbucket.org/genomicepidemiology/cgevirus/src/main/

- CGE metagenomics:
  https://bitbucket.org/genomicepidemiology/cgemetagenomics/src/main/

- CGE QC:
  https://bitbucket.org/genomicepidemiology/cgeqc/src/main/

Current status

- Currently works on Linux and MacOS (hopefully Windows support in the future)
- Not yet deployed to African partners in its current form with updated databases
- Future development:
    - Enable data sharing → Allow partners to exchange results
    - Phylogenetic tree generation → Compare relatedness of isolates across different samples
    - Centralized strain database → Continuously update with new data to improve reference data and track pathogens in the region

Thank you!