Genomics training in the EMRO region

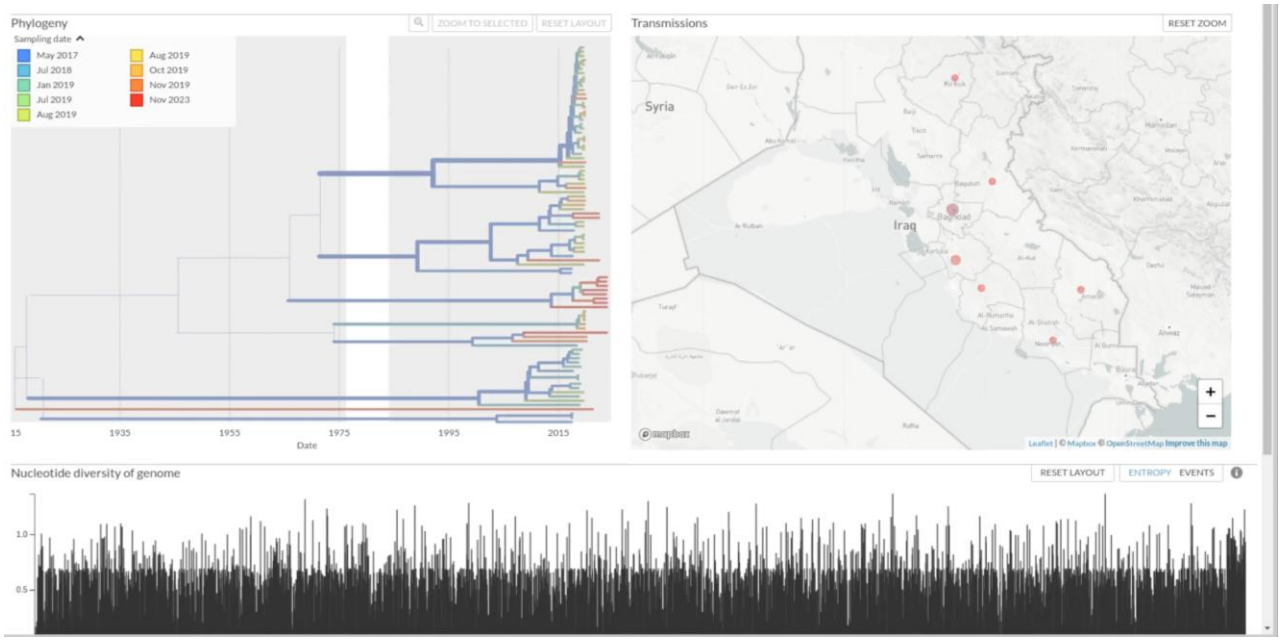
Dr Luke W. Meredith

WHO EMRO WHE/IHP Lab Team

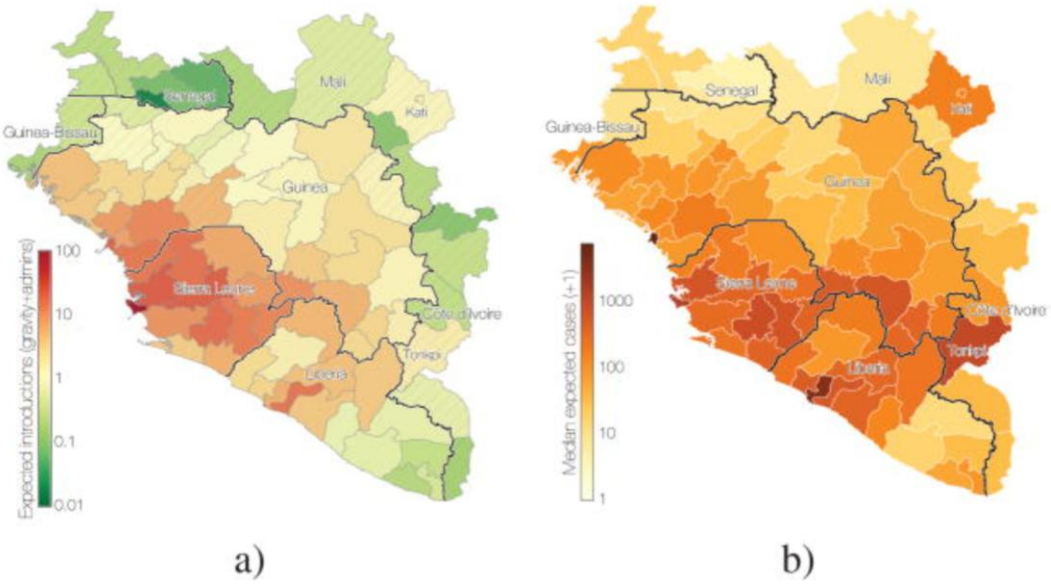
What can we do with sequence data?

- **When used properly**, sequencing provides invaluable **support** for epidemiological, diagnostic and therapeutic interventions.
- Sequencing provides support for epidemiological tracing across and within borders.
- Monitoring AMR to guide drug distribution for clinical support.
- Monitor efficacy of public health interventions.
- At this stage sequencing is **NOT** a primary diagnostic tool approved for use for anything other than COVID-19 unless under extraordinary circumstances.

Transmission chains

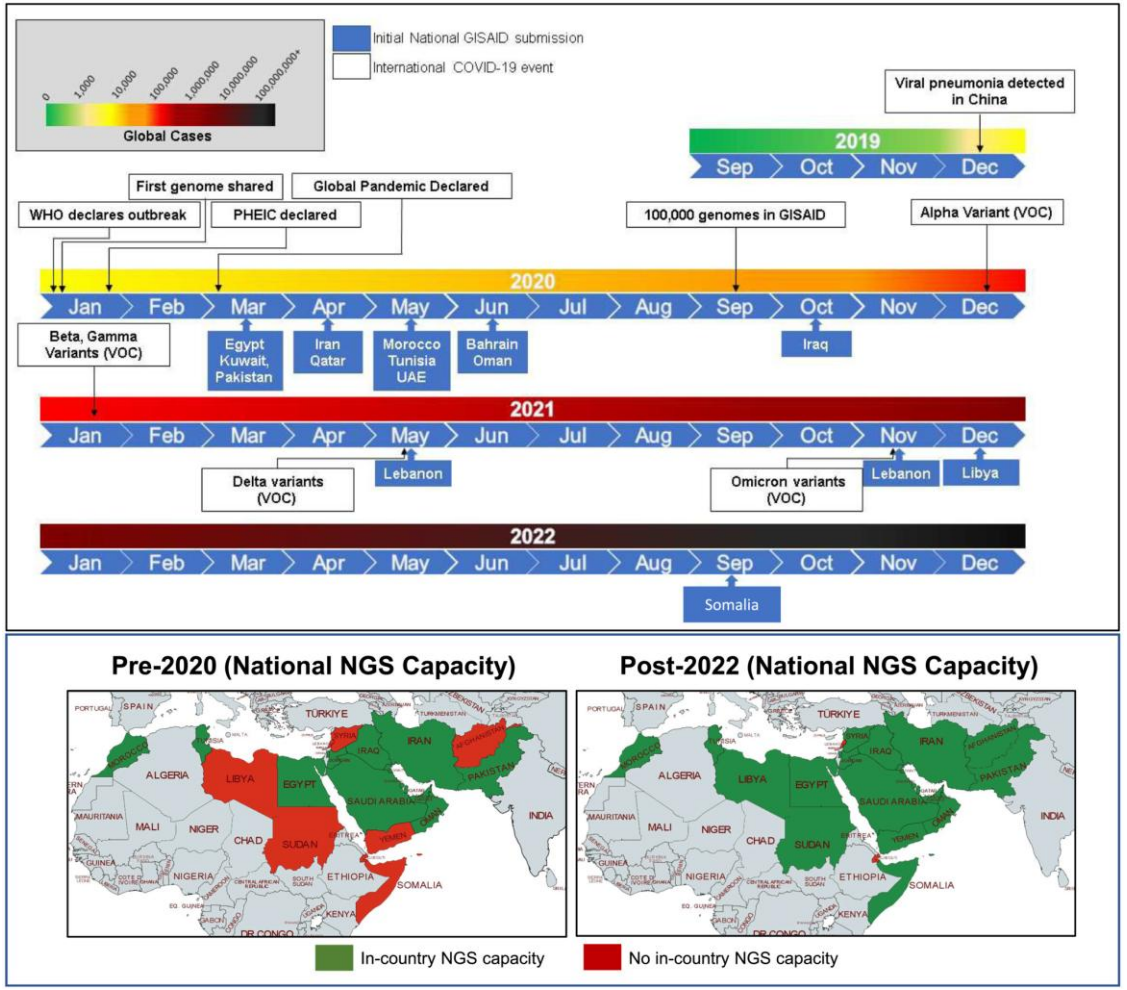


Effect of interventions



Genomic goals

- Build on and strengthen genomics capacity in the Region.
- Integrate genomic surveillance with existing surveillance networks to enhance awareness and understanding of disease spread in the region to prevent outbreaks becoming epidemics or pandemics.
- Build a sustainable regional network of genomic expertise at both the laboratory and surveillance level
- Support the development of sustainable genomic surveillance at the National policy level.



21/22 countries now have sequencing capacity at the national level for COVID-19, expand this to other viral and bacterial pathogens.

EMRO NGS Trainings 2023/4

Regional/Sub-regional trainings

- ***AWD Genomics and Bioinformatics, AUB Lebanon, 2023***
- Arboviral Genomics and Bioinformatics, CPHL Oman, 2023
- Bioinformatics, Online, CLC Workbench, 2023

National (in-country) trainings

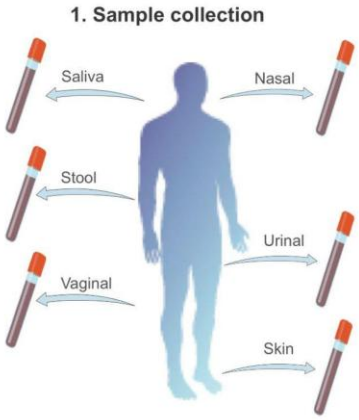
- ***AWD Genomics***
  - ***Sudan, Yemen, Iraq, UAE, Oman***
- Arboviral genomics
  - Iraq, UAE, Oman, Afghanistan*
- Respiratory genomics
  - UAE, Oman, Afghanistan*, Lebanon*

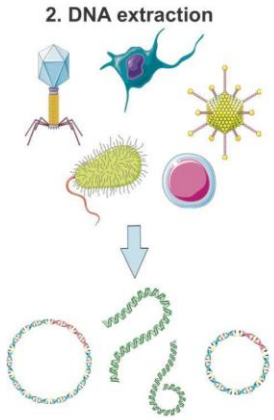
Remote Trainings / Data analysis

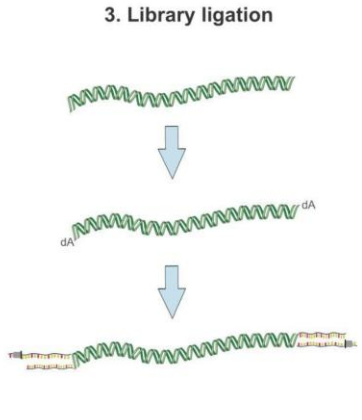
- AWD genomics
  - Iraq, UAE, Sudan, Oman

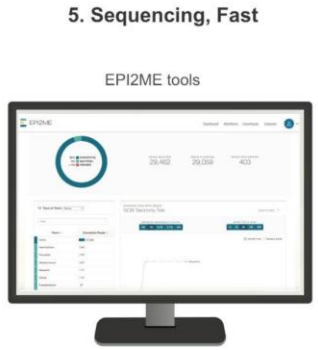
**Pending in July*

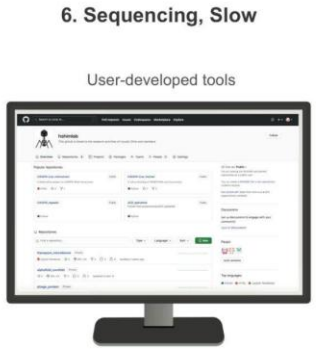
Cholera Sequencing: Nanopore rapid sequencing











- Identify lineage, AMR, strain and toxigenicity in one assay.
- One plate or broth for culture.
- 24 hours from sample receipt to sequence using ONT rapid barcoding.
- Requires an extraction kit, flowcells and a barcoding kit, plus some ethanol.
- Simple, reproducible protocol
- ~$90USD/sample (if barcoding).

Why start with Nanopore?

By 2022, 21/22 countries in the region had either a MK1C or a GridION to sequencing COVID-19.

The Midnight protocol was the most widely used protocol in the region.

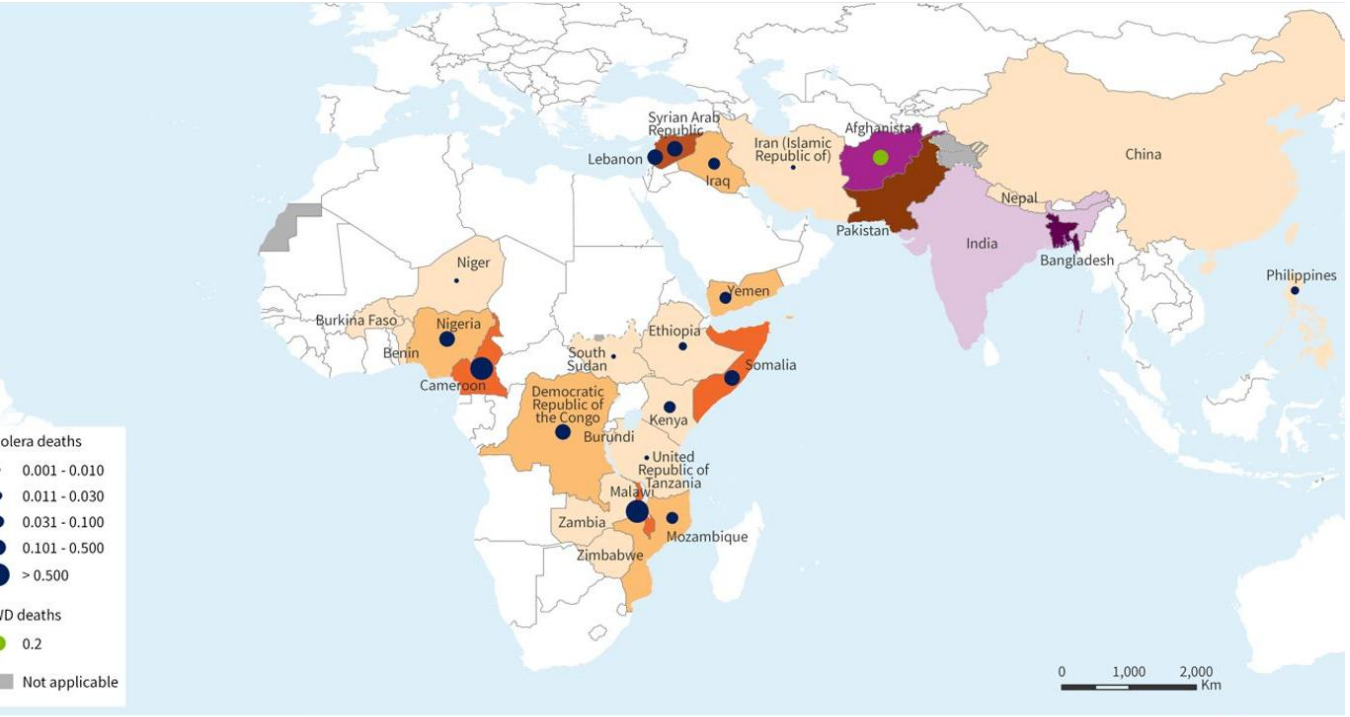
Logistically speaking, the nanopore offers the smallest footprint and ease of use in complex settings.

Adapting existing protocols to use Artic-style approaches is possible as is shotgun sequencing

https://onlinelibrary.wiley.com/doi/full/10.1111/irv.13205

What do we hope to achieve: **Cholera**

- Cholera outbreaks declared across almost the entire region.
- Vaccine resistant strains?
  Increased virulence?
  Drug resistance?
  Novel transmission chains?
  Effect of WASH/interventions?
- **NOT primary diagnosis** outside specific circumstances.



d the presentation of the material in this publication do not imply the expression of part of WHO concerning the legal status of any country, territory, city or area or of its delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent hich there may not yet be full agreement.

Data Source: World Health Organization, United Nations Population Division (population prospect 2021)
Map Production: WHO Health Emergencies Programme
Map Date: 9 December 2022

World Organi

AWD training, AUB Lebanon

Why AUB?

- WHO Collaborating Center for Bacterial Pathogens.
- Good facilities and experience with sequencing.

Trainers (10)

- WHO EMRO
- AUB (5 laboratory staff)
- UK HAS NVAP (1 laboratory, 1 bioinformatician)
- University of Cambridge (4 bioinformaticians)

AWD Training Structure (Nanopore)

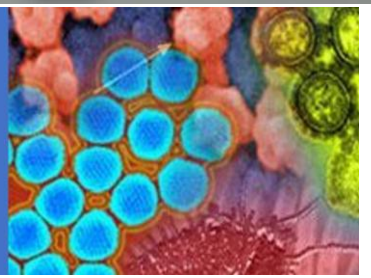
- Theory (4-10 hours)
  - Bacteriology basics.
  - NGS technology.
  - Surveillance and interpretation of sequencing data.
  - Biosafety and sample handling.
  - Data management (the importance of metadata)
  - Quality control.

Sampling and Testing Strategies:
What are they and why do we need them?

Next-Generation Sequencing:
What is it and what can we do with it?

Targeted vs Non-targeted sequencing

Shotgun Metagenomic Sequencing



Metagenomic Analysis:
Bioinformatics

On the Importance of Data Management and Quality Control

AWD Training Structure (Nanopore)

- Hands-on Laboratory (3-5 days)
  - Nanopore focused, end-to-end starting from sample selection, handling through to loading and starting a sequencing run.
  - Detailed, step-by-step protocols.
  - Focus on quality and prevention of cross-contamination
  - Importance of negative controls
  - Unidirectional workflows etc

Appendix 1: Bench protocol – Cholera shotgun metagenomic sequencing protocol with rapid barcoding (SQK-RBK110.96)

Appendix 1. Bench Protocol

Shotgun metagenomic sequencing of Cholera using Oxford Nanopore Technology and Rapid Barcoding

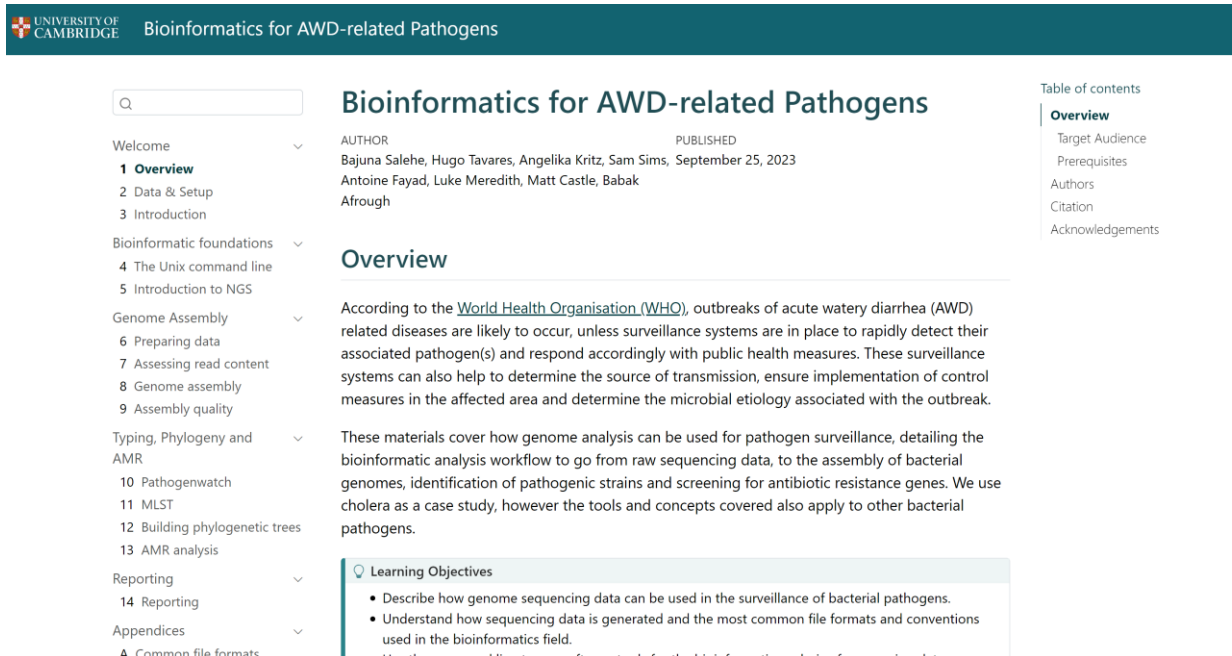
Date: _____

No. of Samples: _____

Operator:

AWD Training Structure (Nanopore)

- Key aspect of training.
- Data analysis (2-4 days)
  - Basecalling and assembling a genome
  - Quality control
  - Basic phylogenetics (where appropriate)
  - All protocols available online
  - Support given remotely on request



UNIVERSITY OF CAMBRIDGE Bioinformatics for AWD-related Pathogens

Welcome
1 Overview
2 Data & Setup
3 Introduction
Bioinformatic foundations
4 The Unix command line
5 Introduction to NGS
Genome Assembly
6 Preparing data
7 Assessing read content
8 Genome assembly
9 Assembly quality
Typing, Phylogeny and AMR
10 Pathogenwatch
11 MLST
12 Building phylogenetic trees
13 AMR analysis
Reporting
14 Reporting
Appendices
A Common file formats

Bioinformatics for AWD-related Pathogens

AUTHOR
Bajuna Salehe, Hugo Tavares, Angelika Kritz, Sam Sims, Antoine Fayad, Luke Meredith, Matt Castle, Babak Afrough

PUBLISHED
September 25, 2023

Table of contents
Overview
Target Audience
Prerequisites
Authors
Citation
Acknowledgements

Overview

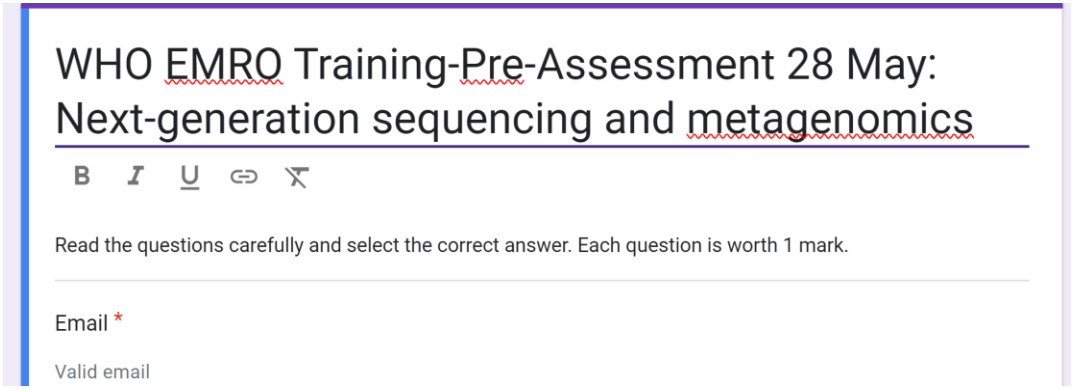
According to the World Health Organisation (WHO), outbreaks of acute watery diarrhea (AWD) related diseases are likely to occur, unless surveillance systems are in place to rapidly detect their associated pathogen(s) and respond accordingly with public health measures. These surveillance systems can also help to determine the source of transmission, ensure implementation of control measures in the affected area and determine the microbial etiology associated with the outbreak.

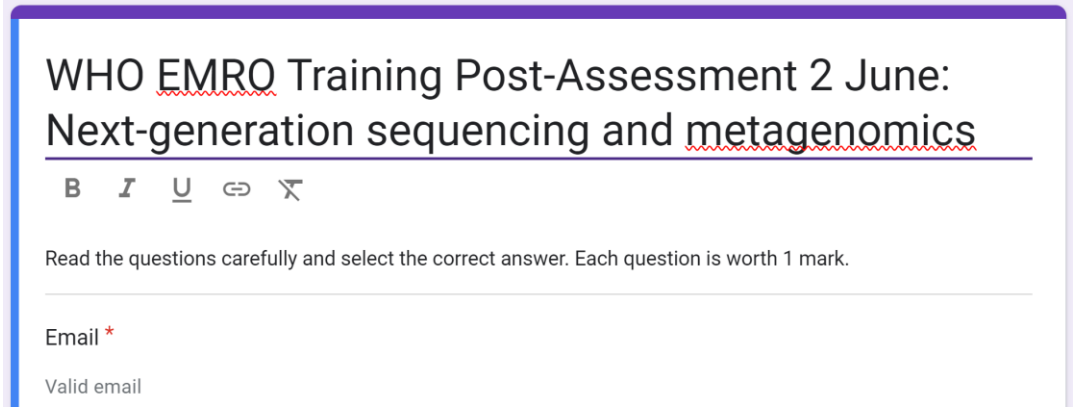
These materials cover how genome analysis can be used for pathogen surveillance, detailing the bioinformatic analysis workflow to go from raw sequencing data, to the assembly of bacterial genomes, identification of pathogenic strains and screening for antibiotic resistance genes. We use cholera as a case study, however the tools and concepts covered also apply to other bacterial pathogens.

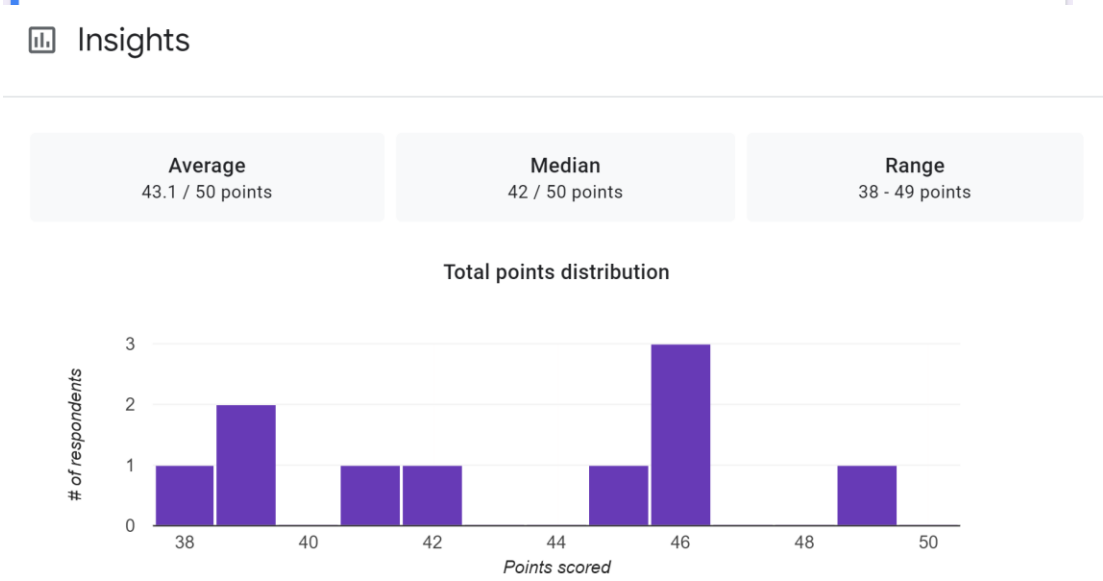
Learning Objectives
- Describe how genome sequencing data can be used in the surveillance of bacterial pathogens.
- Understand how sequencing data is generated and the most common file formats and conventions used in the bioinformatics field.

AWD Training Structure (Nanopore)

- Assessment
  - Pre- and post-course assessment of learning.
  - Look for an improvement in knowledge.
  - Scores confidential and shared only with participants.
  - Provide feedback on the key areas that need focusing on.
  - No “pass/fail” purely informative.







AWD Training Structure (Nanopore)

- Post-course feedback
  - Survey on the quality of the training and the implementation of genomics when the participants get back to their own lab.

Course Feedback - Overview

- 12 participants provided feedback in the post-course survey
- The overall feedback regarding the training was **extremely positive**

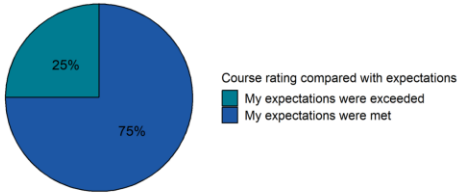


Figure 1. Participant responses to the "How did this course compare to your expectations?" question in the post-course survey. Most respondents noted that their expectations were either met or exceeded. No respondents noted that their expectations were not met.

Knowledge Increase

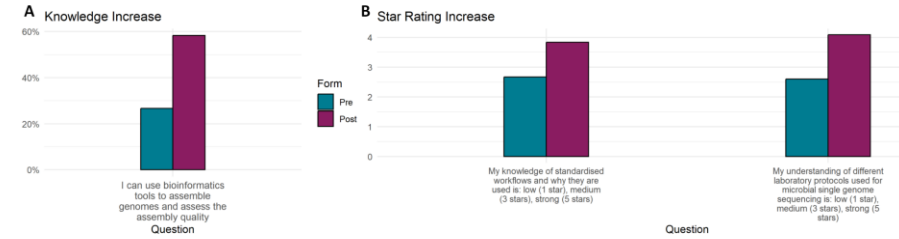


Figure 2. **(A)** Percentage of participants agreeing with statements and **(B)** mean star rating in the pre- (n = 15) and post-course (n = 12) surveys.

ens of Interest

Figure 4. When asked "What other pathogens would your laboratory be interested in receiving bioinformatics training for in the future?", the most common response was polio.

Challenges

- Availability of sequencing reagents
- Availability of sequencing training and troubleshooting (outside workshop)
- Access to Unix environment (either dedicated PC or VM)
- Sequencing errors with ONT (older chemistries)
- Steep learning curve/no prior exposure to command line

Course Delivery

- Most participants noted that the course was **too short**
  - Of these respondents, most would like to spend more time on the bioinformatics portion of the course (9/10)
  - The remainder wanted to spend more time on both the lab and bioinformatics portions of the course

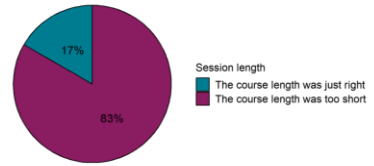


Figure 3. When asked "Do you think that the length of the course was suitable?", most participants noted that the course was too short.

Regional and National Training courses

- Series of trainings supported by the Rockefeller Foundation and WHO country offices.
- Follow-up regional trainings with National training in active outbreaks.
  - Sudan, Yemen, Oman, UAE and Iraq, Somalia (pending).
- “Train the trainers” approach, Central labs followed by sub-national.
- Focus on high-quality, reproducibility and utility of the data.

Successes

Sudan:

- Submission of 31 genomes to SRA (in progress).
- Lab work done by national staff (analysis by EMRO and NVAP).
- AMR analysis with the intended goal of supporting therapeutics.
- Toxigenicity and AMR profiles from regions.

Yemen, Oman

- Analysis of ~30 samples from across the countries (analysis ongoing) to understand where they came from.

Lebanon

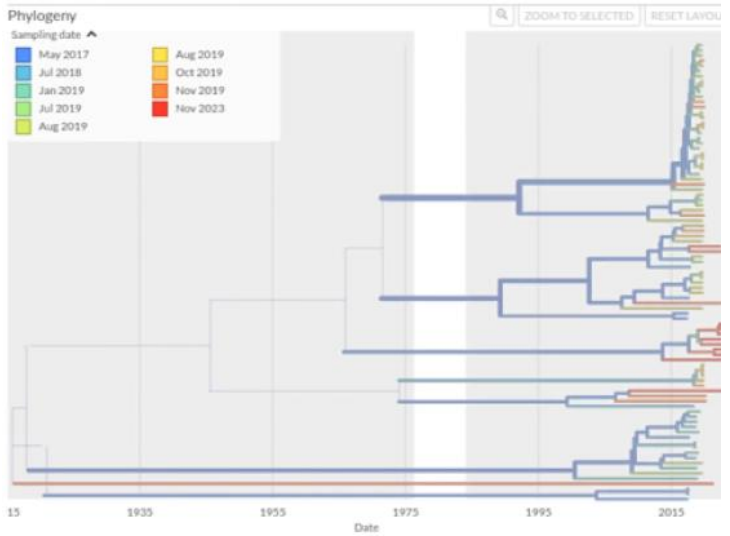
- Experience as trainers for genomics.
- Preparation of training course for Somalia.

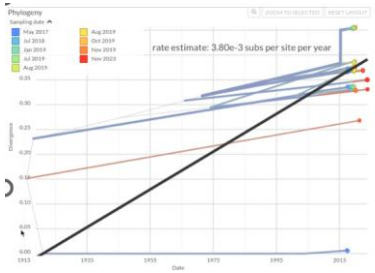
Regional

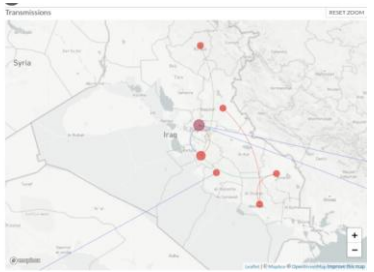
- Increase expertise in bacterial whole genome sequencing.
- Local Nextstrain monitoring for cholera transmission (in development).
- Fantastic resource for training
- https://cambiotraining.github.io/awd-pathogen-bioinformatics/

Challenges: Bioinformatics and data analysis

- The pipeline is straightforward but requires working knowledge of linux, which is a challenge.
- Understanding what the data means and how to use it.
- Phylogenetics require significant knowledge and computational power.
- Reporting in a useful manner to surveillance and public health teams is challenging.
- Minimal metadata requirements need to be coordinated from response to laboratory.







Quality Control

- While the sequencing works great, it will only correctly identify cholera if cholera samples are added.
- Sample selection is critical to making sequencing worthwhile.
- Bacteriology skills need to be excellent.
- Confirmed diagnosis is critical to cost-effectiveness and utility.

Training	Genome Coverage	Cholera Sequence Positive	Pathogens Detected
Regional (1)*	~93%	1/30	*Aeromonas, Shigella, Brucellosis, E.coli, Enterobacter, Klebsiella, Kluyveria* and more
National (1)	~90%	12/15	
National (2)	~88%	2/30	
National (3)	~94%	2/31	

Sharing and Reporting of Data post-training.

- Reporting results and sharing data with national, cross-border, regional or international partners to allow for public health discussions to be had.
- Doing this is such a way that it’s non-punitive.

Frustration

What does genomic surveillance tangibly need to function?

Outcome	Team
Clearly defined goals for sequencing (priority pathogens etc).	Regional strategy (10 priority pathogens for genomics)
Strong logistics, infrastructure and funding.	Dubai Logistics Hub and partners
A functional surveillance and epidemiology network.	Epidemiology and RRT network
Sample collection and identification kits (and know how to use them).	Logistics, labs and training
A strong laboratory network.	Logistics, labs and training
Clearly defined diagnostic algorithms for in vitro diagnosis.	Labs and training.
Capacity to produce and analyse high quality, meaningful sequencing data.	Labs, EQA, bioinformatics, genetic epidemiology, training teams
Ability to interpret and assess genomic data in relation to public health (risk assessment).	Bioinformatics and epidemiology, public health teams.
Strong links between the laboratory, epi and public health teams (preferably OneHealth).	Communications, data and information sharing teams
Ability to communicate and share that data with partners to mount a response (or not).	Communications, data and information sharing, national and regional policy.

Jordan, 2023

Influenza and other respiratory viruses Open Access isirv

ORIGINAL ARTICLE | Open Access

Key aspects defining the development and implementation of a regional genomic surveillance strategy for the Eastern Mediterranean Region

Luke W. Meredith, Mustafa Aboualy, Rachel Ochola, Patrick Okwarah, Mehmet Ozel, Abdinasir Abubakar, Amal Barakat

First published: 18 October 2023 | https://doi.org/10.1111/irv.13205

Acknowledgements

- **WHO EMRO WHE/IHP Lab Team**
  Amal Barakat
  Mustafa Aboualy
  Dr. Abdinasir Abubakar
  Dr. Thomas Mollet
- **WHO EMRO WHE/IHP Surveillance and Response Teams**
- **WHO Dubai Hub**
  Robert Blanchard
  Nevien Attala
- **US CDC Viral Special Pathogens Branch**
  Shannon Whitmer
  Jessica Rowland
  John Klena
  Frank Chiang
- **WHO Bacterial Collaborative Centre, AUB**
  Antoine Abou-Fayad and the team
- **UKHSA New Variant Assessment Program**
  Babak Afrough
  Leena Inamdar
- **University of Cambridge Training Team**
- **Rockefeller Foundation**
- **Oman Central Public Health Laboratory**
  Hanan al-Kindi
  Yathrib al-Zakwari
  Samiha Kharusi and the team
- **SKMC Virus Reference Laboratory**
  Francis Armithraj Selvaraj and team
- **Nanopore Abu Dhabi Team**
- **Central Public Health Laboratories and health care workers from across the region who are working tirelessly under incredibly tough conditions.**
- **All funding and technical supporting partners from WHO, US-CDC**