Day 1 Session 2: Practical approaches - specimen, genomics and data infrastructure

Aquillah Kanzi, Kareemah Suleiman, Shavanthi Rajatileka, Gerald Mboowa, Dr Dawit Wolday

Course roadmap

Sun 7 May
Introduction Day

Mon 8 May
Day 1
Capacity Building

Introduction
Overview of basic concepts
Case studies

Tue 9 May
Day 2
Specimen and Sequencing

Wed 10 May
Day 3
Data Tools and Pipelines

Thu 11 May
Day 4
Frameworks, Guidelines, and Decision-making

Fri 12 May
Day 5
Projects Review and Action Planning

Next steps and Beyond

Session objectives

Showcase case studies about setting up infrastructure and processes - completed projects on a piece of infrastructure that has been set up for pathogen genomics
Panel showcase - 5 case studies - 8 minutes each

- Biobank at IHV - Kareemah
- Sequence infrastructure set up at a specific place or project network - Shavanthi
- Data infrastructure for surveillance (or network) (Africa CDC) - Gerald
- Genomics infrastructure and pipelines for routine diagnostics (ASLM) - Aquillah

Institute of Human Virology Nigeria (IHVN) Biobank

Kareemah Suleiman

The Biobank : Institute of Human Virology Nigeria H3Africa Biorepository (I-HAB)

Goal:

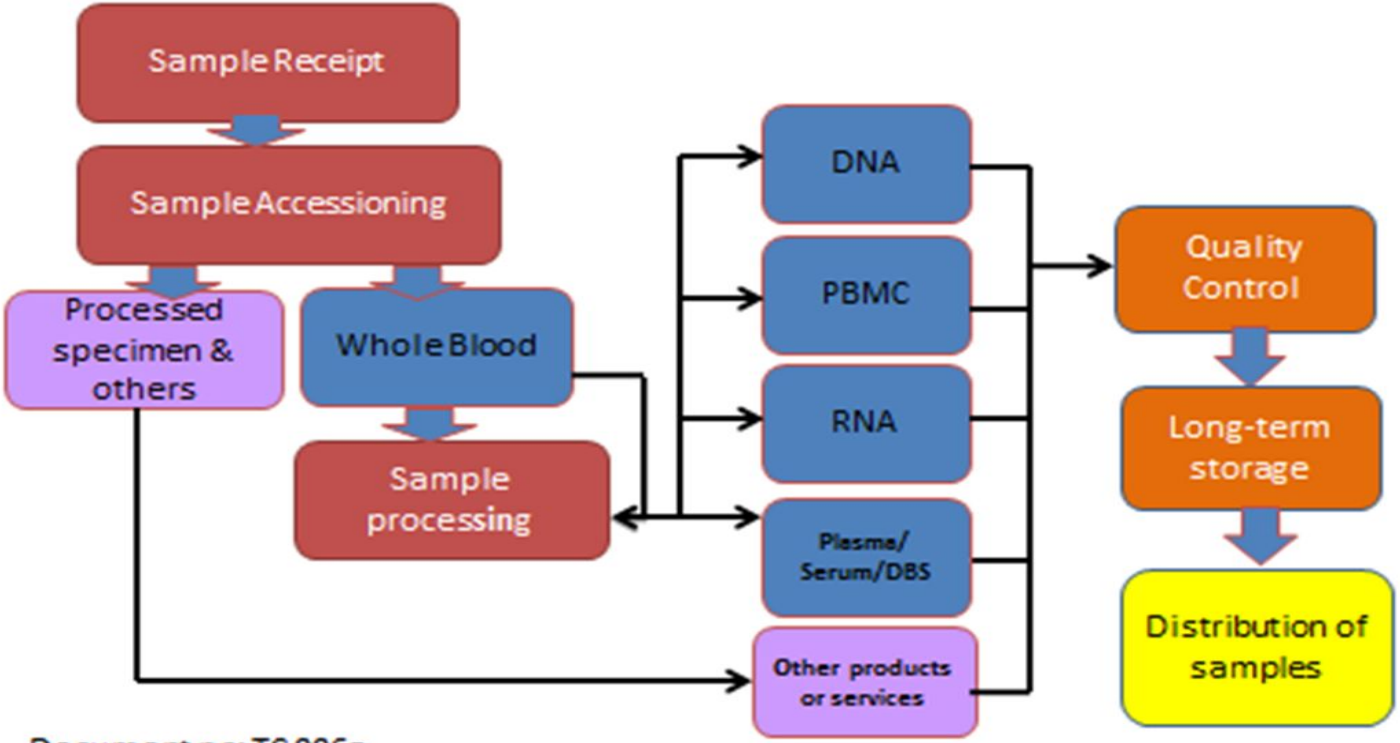
To promote population and personal Health, by facilitating cutting edge research and collaborations among African communities and beyond by providing high quality, affordable

biobanking services.

Current status of I-HAB

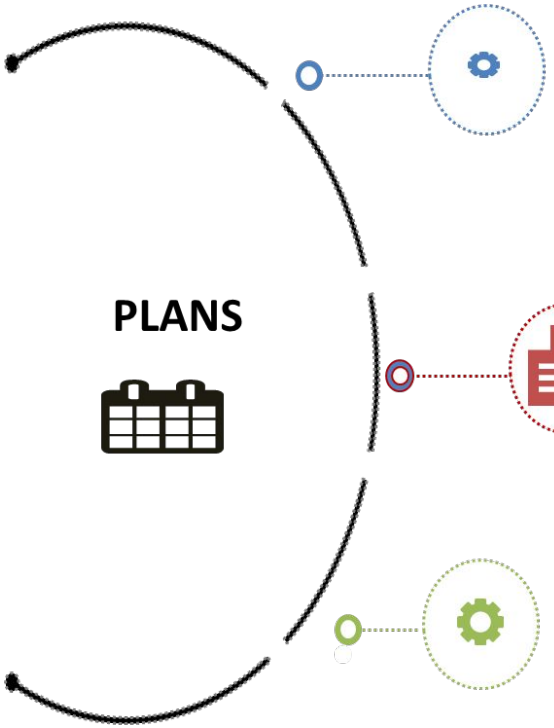
- 2009: IHVN with University of Maryland Baltimore established a biorepository network in Abuja, Zaria and Jos to support IHVN clinical and research activities.
- 2012:Abuja biorepository received funding through the H3 Africa grant supported by NIH to support African investigators in genomic research. The Abuja Biorepository became I-HAB
- 2013:I-HAB partnered with Coriell to bring the biorepository to international standards.
- 2015:H_3Africa biospecimen shipment pilot study between H_3Africa biorepository network
- Sample deposit began from 4 west African Countries-Nigeria, Benin, Ghana and Mali
    - Total biospecimen deposited at I-HAB- 32,687
- Trained 176 research staff in over 30 topics related to biobanking

How the IHVN Biobank is set up



Document no: TC 006a

Ongoing initiatives and future plans for IHV Biobank



- Expanded Fee for Service
- Strengthened Partnership and collaboration with diverse groups
- Expand capacity for Integrated Disease Diagnostics and Surveillance

Sequencing set-up

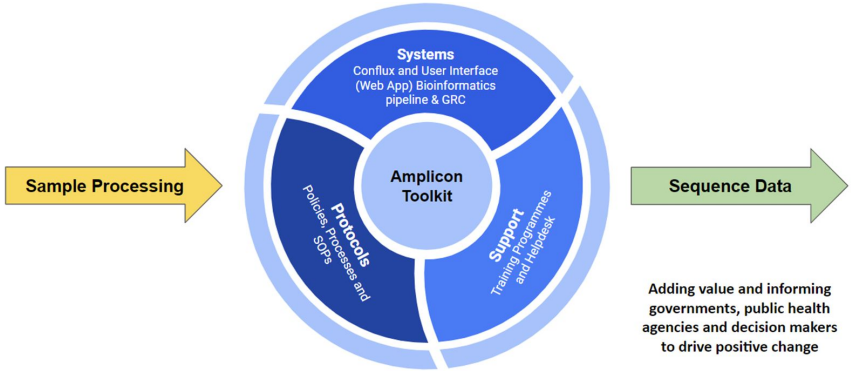
Dr Shavanthi Rajatileka

FUNDED BY NIHR National Institute for Health and Care Research
UKaid from the British people
MRC Unit The Gambia
COG-TRAIN COVID-19 GENOMICS GLOBAL TRAINING

GENOMIC SURVEILLANCE OF MALARIA IN WEST AFRICA

PROJECT OVERVIEW

- Funded by the National Institute for Health Research
- Project based at the West African Centre for Cell Biology of Infectious Pathogens (Ghana) and MRC Unit, The Gambia (The Gambia) in collaboration with the Wellcome Sanger Institute (UK).



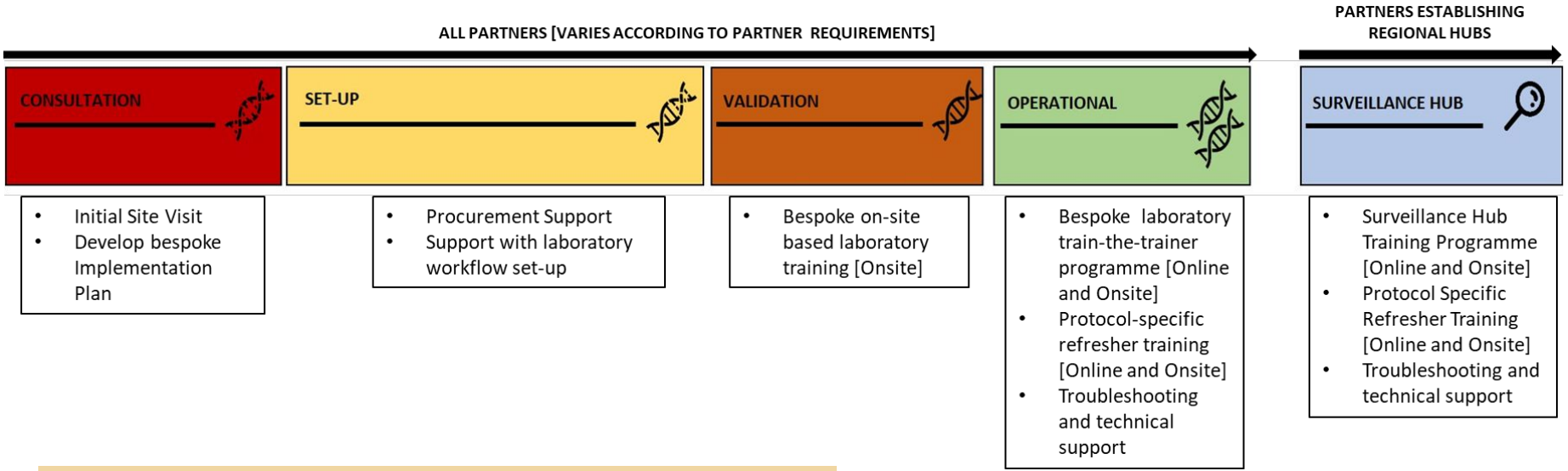
PROJECT AIMS

- To build infrastructure (laboratory & some data systems) for genomic surveillance using amplicon sequencing in at the University of Ghana and MRC Unit, The Gambia.
- To establish a working proof of concept for genomic surveillance of malaria parasites and vectors in Ghana and The Gambia.
- To work with National Malaria Control Programmes (NMCPs) to learn to translate genomic data into actionable outputs that can be integrated into their operations.

wellcome connecting science

COG-TRAIN COVID-19 GENOMICS GLOBAL TRAINING

DEPLOYMENT AND IMPLEMENTATION STRUCTURE



BCL uploader & Samplesheet Checker

Bioinformatics Pipeline

GRC returned to partners

wellcome connecting science

COG-TRAIN COVID-19 GENOMICS GLOBAL TRAINING

DEPLOYMENT AND IMPLEMENTATION STRUCTURE

Pillar	Consultation	Laboratory Set-up	Sample Collection	Laboratory Protocol validation	Sequence Generation (Operational)	Genomic analysis	Dissemination
Major activities	• Site visits • Resource review • Develop bespoke implementation plan	• SOP review • Establish workflow • Procurement of equipment, reagents, consumables • Compliance, collaboration agreements, MTAs, DTAs	• Collect sample • Collect metadata • Microscopy confirmation of species • Laboratory confirmation of species	• Optimise protocols • Process test samples • Sequence test samples • Validation samples processed • Laboratory Training • Protocol QC checks	• Process samples • Sequence samples • Refresher training • Laboratory train-the-trainer programme • Scaling up operations and improve QC	• Bioinformatics and computational analysis • Metadata linkage • Generate genetic report card • Visualize data	• Report information to NMCPs/ Public Health officials/Policy makers
Implementing partner	Laboratory team Project leads Administration staff (Finance, Grants, Legal, Procurement)	Laboratory team Project leads Administration staff (Finance, Grants, Legal, Procurement)	Fieldworkers National malaria control programmes Clinic staff Laboratory staff	Laboratory staff Procurement teams	Laboratory staff Procurement teams	Bioinformatics team Data Scientists	Project PI Project Leads NMCPs Public Health Officials Policy Makers

LESSONS LEARNED

ADMINISTRATIVE

- GFGP and due diligence understand requirements
- Put research collaboration agreements, material transfer agreements and data transfer agreements in place at project inception
- Dedicated research managers based at each site
- Dedicated project website to track project progress and showcase work
- Factor in costs for compliance, shipping and dissemination

SUPPLY CHAINS

- Identify the correct routes for procurement
- Standardized protocols and list of reagents and consumables and equipment
- Working with other national/regional institutions to combine orders

LABORATORY TRAINING

- Develop protocols and training materials in collaboration with partner laboratories
- Bespoke training for each implementing lab
- Facilitate discussions between regional labs
- Online training materials
- Incorporate a laboratory train-the-trainer programme at earliest
- Buddy system with key members of lab teams

DATA PROCESSING

- Work with partners to build required tools for data upload and processing.
- Understand data requirements for National Malaria Control Programmes

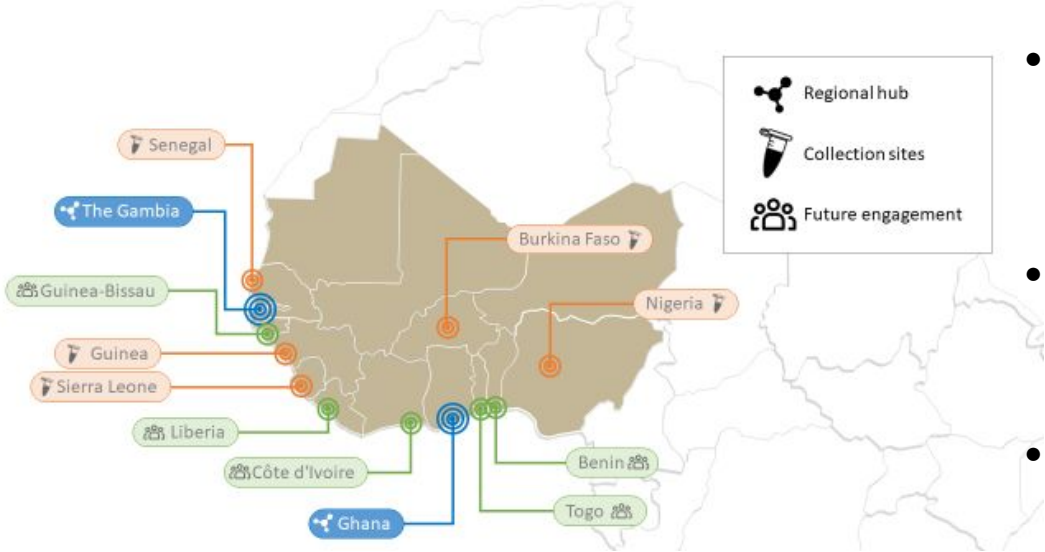
QUALITY CHECKS

- Develop a monitoring and evaluation programme in in collaboration with partners
- Work with partners to improve processes and quality checks
- Support partners with obtaining accreditation

COMMUNITY

- Including NMCPs and public health laboratories in the project from the start
- Ensure the teams know they are valued and the importance of their roles
- Partner-led community and public engagement activities

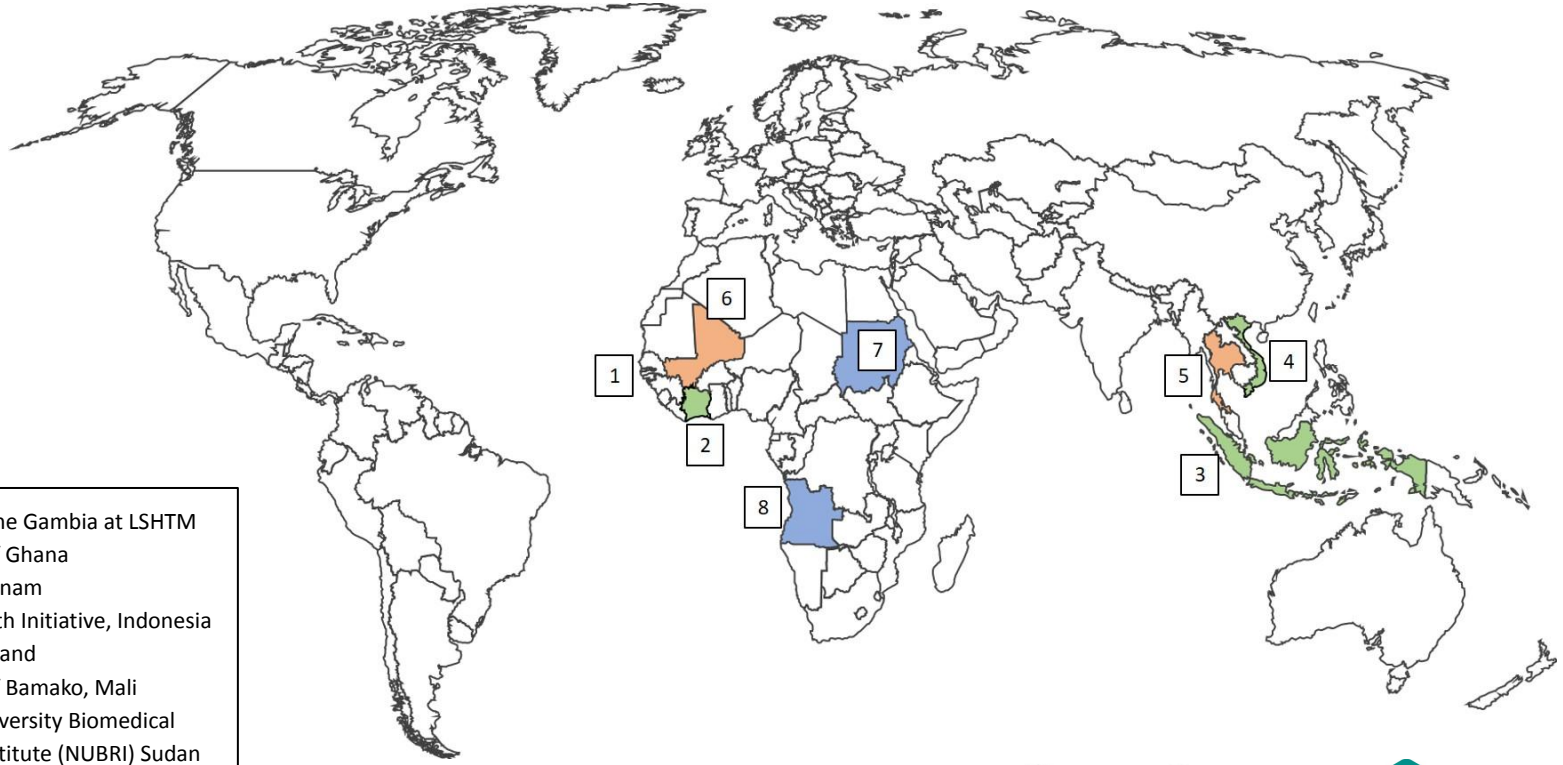
NEXT STEPS: PHASE 2



OBJECTIVES

- **Expand capacity for genomic surveillance** of malaria parasites and mosquito vectors in Africa and Asia, and to generate essential information required for National Malaria Control Programmes (NMCPs) to plan sustainable interventions
- **Establish regional sequencing hubs** capable of processing samples from neighbouring countries and providing NMCPs with timely, actionable genomic surveillance data
- **Integrate genomic data into the routine working practices of NMCPs** and to provide a working example of how such end to end genomic surveillance systems could be deployed at other locations in the world.

NEXT STEPS: EXPANSION OF PARTNER SUPPORT



1. MRC Unit, The Gambia at LSHTM
2. University of Ghana
3. OUCRU, Vietnam
4. EXEINS Health Initiative, Indonesia
5. MORU, Thailand
6. University of Bamako, Mali
7. National University Biomedical Research Institute (NUBRI) Sudan
8. INIS - National Health Research Institute, Angola

wellcome connecting science

COG-TRAIN COVID-19 GENOMICS GLOBAL TRAINING

AFRICA PGI DATA MANAGEMENT & EXCHANGE PLATFORM

Gerald Mboowa

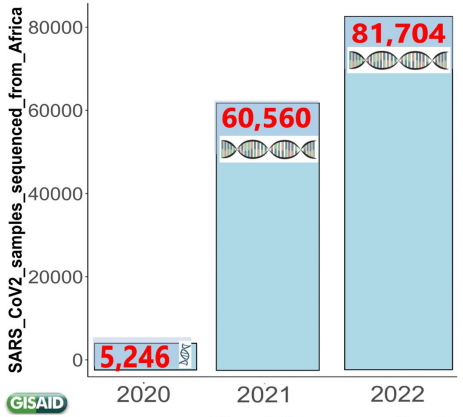
wellcome connecting science
COG-TRAIN
COVID-19 GENOMICS GLOBAL TRAINING

HOW THE DATA MANAGEMENT & EXCHANGE PLATFORM WAS SET UP

Data portal for surveillance of pathogens & antimicrobial resistance in near real-time

Goal of the project to develop infrastructure

- Africa PGI has equipped over 40 national public health institutions (NPHIs) with NGS platforms
- Many NPHIs lack capacity to analyse sequence data
- Federated data analysis, management, sharing and archiving



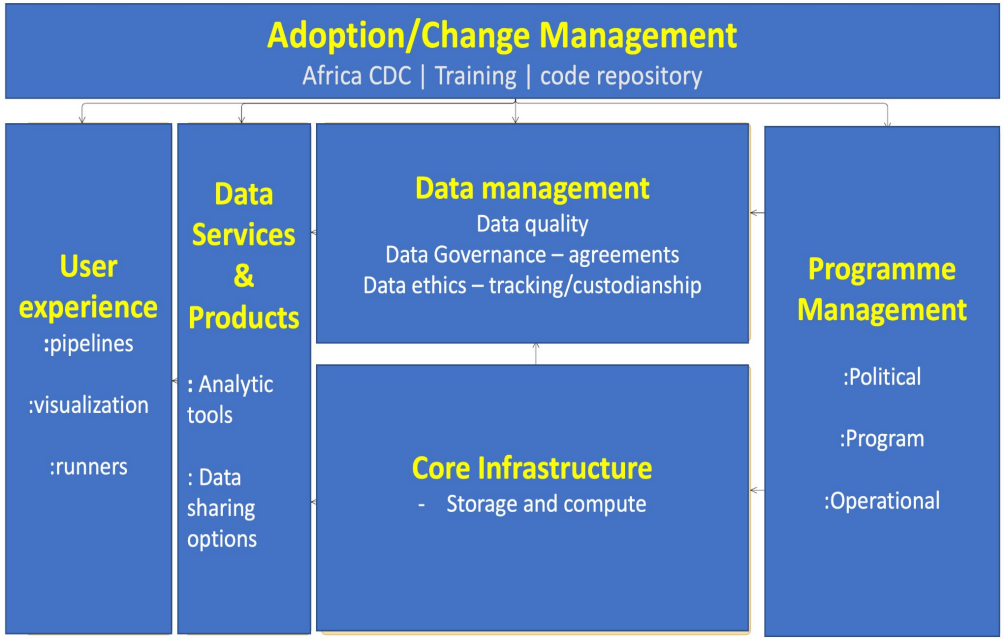
Need for genomic data infrastructure

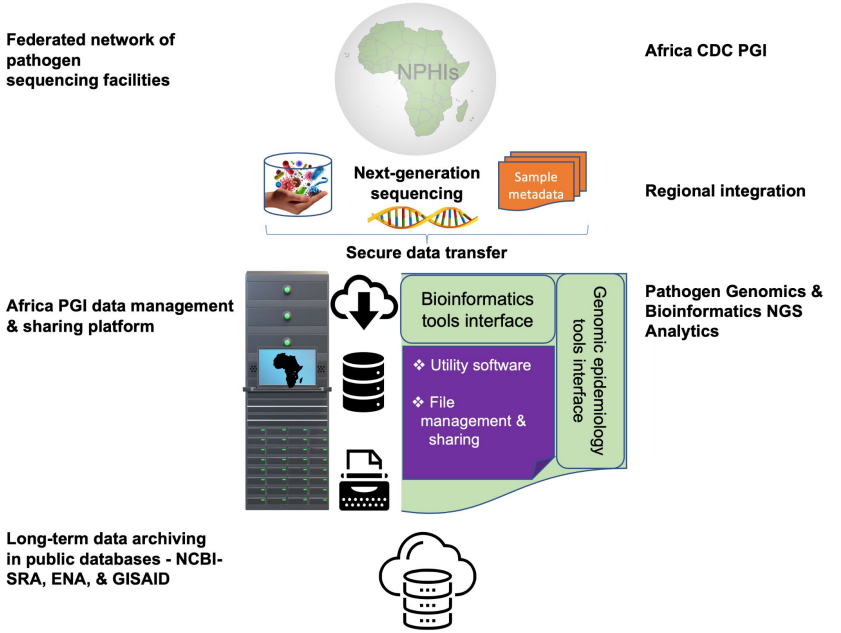
Africa has generated at least 10 TB of Pathogen Genomic Data deposited in NCBI-SRA

DATA DELUGE

Estimate period	Estimated output at full capacity
Weekly	15 TB
Monthly	60 TB
Annually	720 TB
5-Year	3.6 Petabytes

HOW THE DATA INFRASTRUCTURE WAS SET UP





wellcome connecting science

COG-TRAIN COVID-19 GENOMICS GLOBAL TRAINING

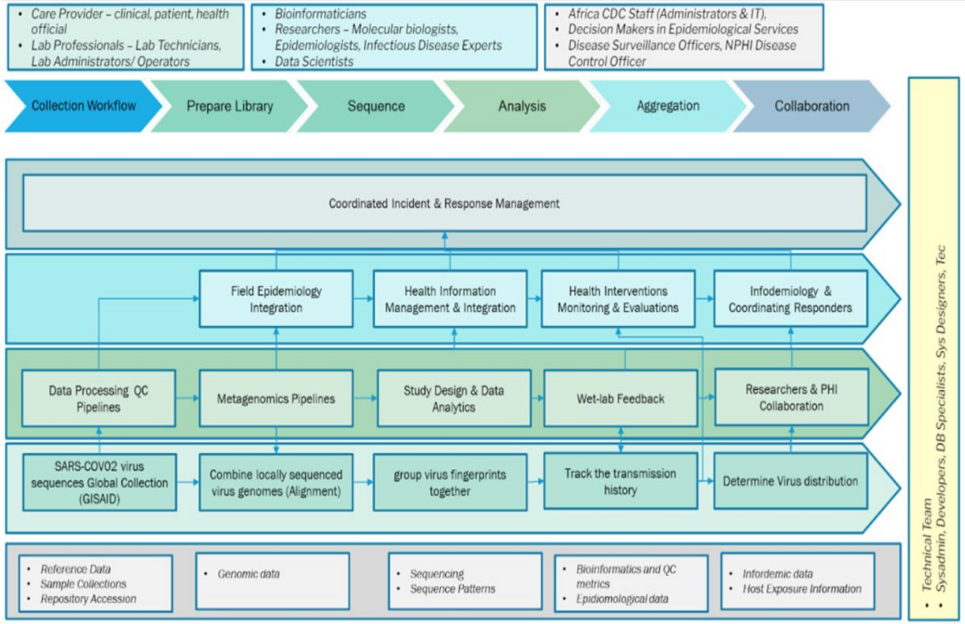
CURRENT STATUS OF THE DATA INFRASTRUCTURE CAPACITY PROJECT

Phase 1: SARS-CoV-2 shared platform	October 2021/2022
Hosting infrastructure	
Genomic Data platform setup	
VirusSeq pilot setup	
VirusSeq available to Africa CDC NGS network	
Phase 2: NGS analytics platform	October 2022/2023
Visualising Variants	
Paas: Workflow management	
Daas: Pathogen data integration	
Daas: Analytics on demand	
Phase 3: Fully operational data platform	October 2023/2024
Workspace/Application management	
Regional hybrid integration	
Use experience & data portal	
Accessible African pathogen genomics shared data platform	

INV-036413 ($0.5M)

Through an RFP process, select partners to deliver a pilot of the initial technical implementation (Phase 1: INV-036413), including solution configuration & enhancement, data curation services, & ongoing maintenance & support services.

$7M 55% operational costs



ONGOING INITIATIVES FOR DATA INFRASTRUCTURE PROJECT

- **Pathogen genomics use-cases projects**

  1. *Vibrio cholerae* genomic surveillance
  2. *Klebsiella pneumoniae* genomic surveillance
  3. Respiratory pathogen panel (RPIP) pilot project

- **Developing national multi-pathogen genomic surveillance strategies**

African Union

AFRICA CDC
Centres for Disease Control and Prevention
Safeguarding Africa's Health

TECHNICAL SUPPORT FOR
DEVELOPMENT OF NATIONAL MULTI-PATHOGEN GENOMIC SURVEILLANCE STRATEGY

WWW.AFRICACDC.ORG AFRICACDC

ADDIS ABABA, ETHIOPIA
2-3 DECEMBER, 2022

wellcome connecting science

COG-TRAIN COVID-19 GENOMICS GLOBAL TRAINING

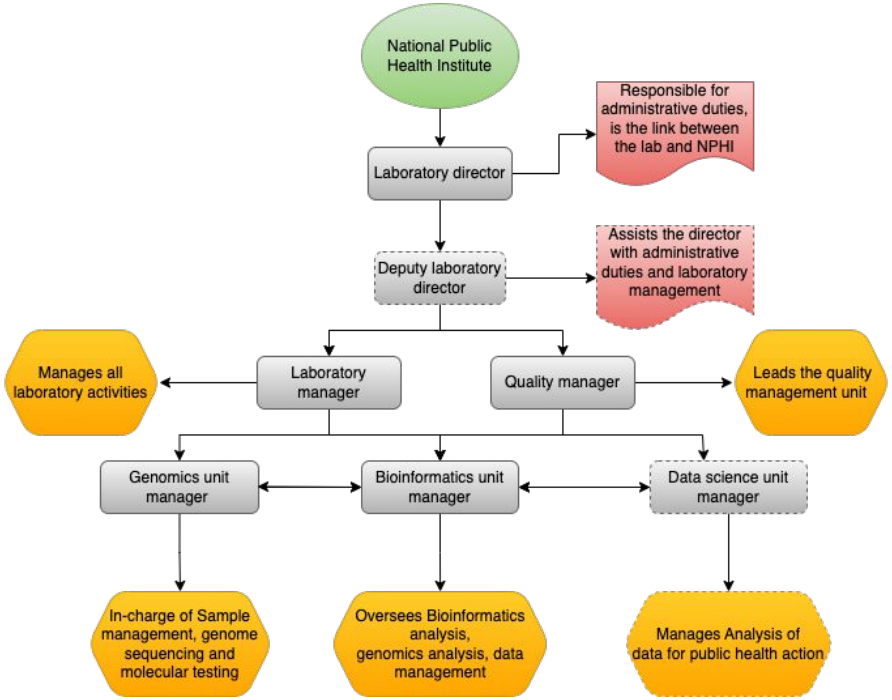
Genomics infrastructure and pipelines for routine diagnostics

Dr Aquillah Kanzi

Setting up genomics infrastructure, quality management and capacity development: CARES Pathogen genomics

- ASLM CARES Pathogen Genomics project is funded by the US CDC CARES laboratory strengthening program
- Supports Africa CDC Africa PGI activities through the following:
  - Equipment and infrastructure upgrades
  - Sample collection and referral
  - Staffing and workforce development
  - Quality assurance - quality management systems (QMS) and EQA
- CARES Pathogen Genomics supports nine African member states in the Africa PGI laboratory network

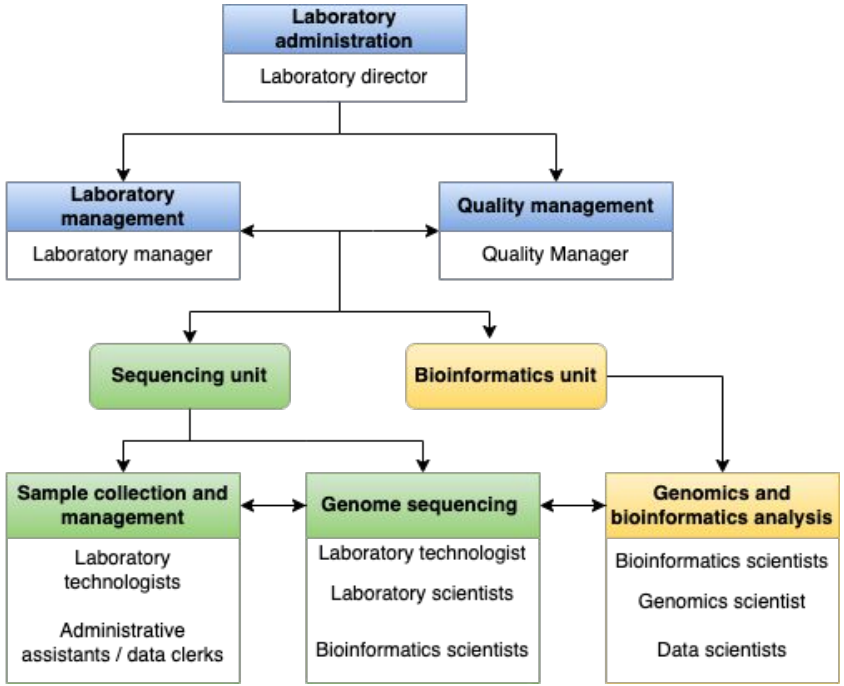
General organization of CARES supported genomics laboratories



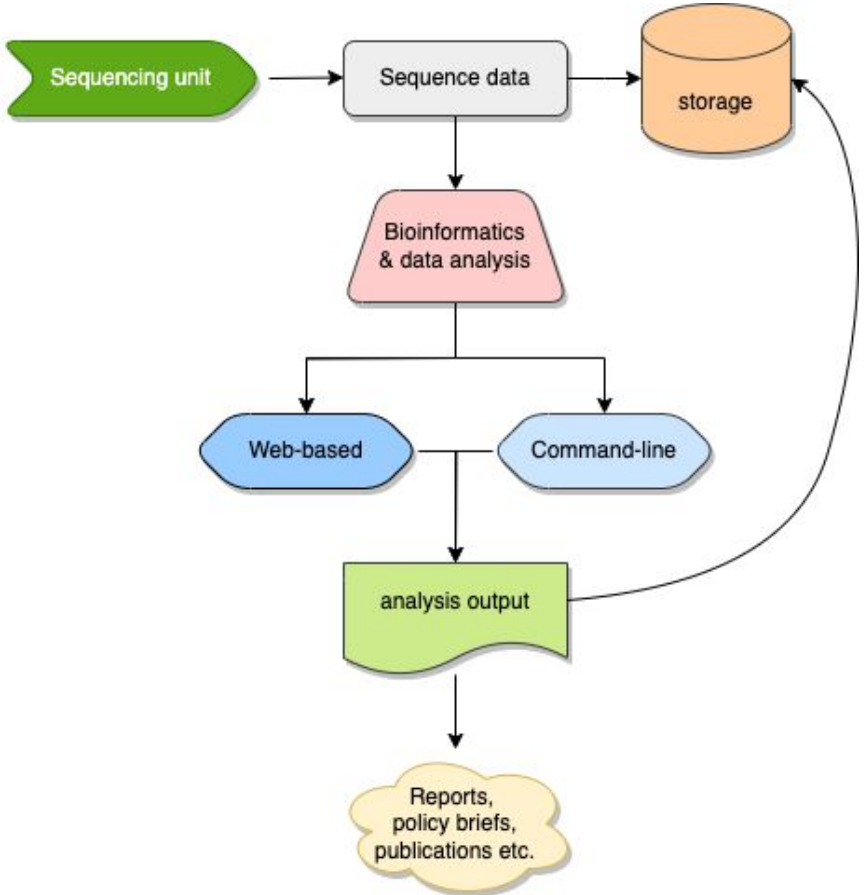
General organization of CARES supported genomics laboratories

Management

Sequencing and bioinformatics analysis



Implementation of bioinformatics analysis workflows



Africa PGI is working on a data architecture framework for Africa

Current status of the genomics infrastructure capacity of CARES supported laboratories

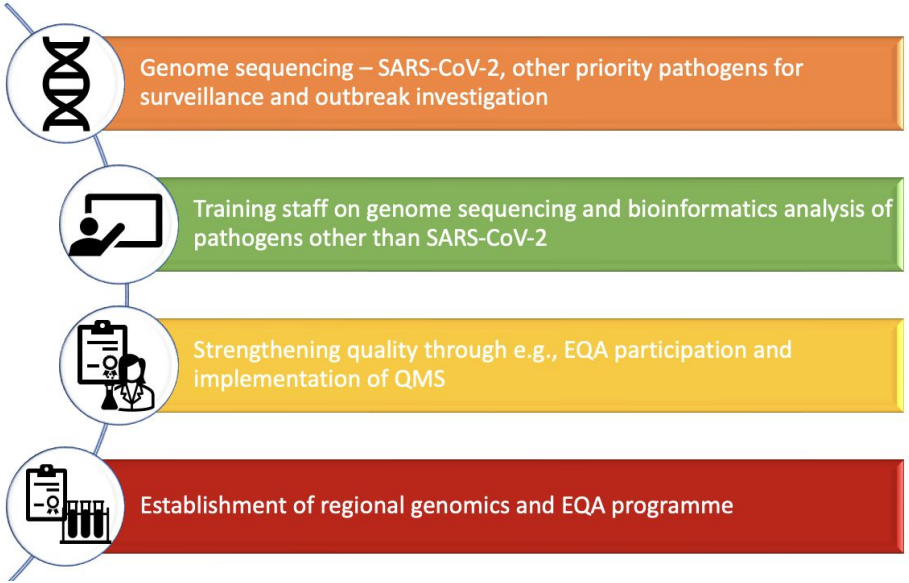
Genome sequencing

- The laboratories have varying sequencing capacity depending on scope and needs
- Examples of sequencing equipment available in these laboratories include
  - Miseq
  - MiniSeq
  - Nextseq 2000
  - Oxford Nanopore - MinION, GridION
  - Accessory equipment e.g, liquid handlers, DNA quantitators and quality analyzers, DNA size selection equipment, DNA shearing equipment etc.

Bioinformatics analysis

- Varying capacity of bioinformatics analysis infrastructure including,
  - desktop computers, laptops, and servers

Ongoing initiatives and future plans for strengthening capacity for CARES supported laboratories



Establishing Pathogen Genomics Excellence at Ethiopian Public Health Institute

Dr Dawit Wolday

EPHI Pathogen Genomics Initiative

Overall aim:

- Building capacity for integrated pathogen genomic surveillance for informed public health decision process

Overarching ***specific objectives*** include:

- Strengthen collection and analysis of **clinical and epidemiological data** and clinical samples; perform **translational clinical research** demonstrating the application of genomic epidemiology to **inform public health decision-making**
- **Enhance capacity** pathogen genomic sequencing Ethiopia, including strengthening lab infrastructure, human work force, pathogen genomic data analysis, integration with metadata
- Develop and implement **innovative digital diagnostic platforms** and create **semi-real time mobile phone applications** for policy decisions
- Promote **communities of practice** and **knowledge exchange** through fostering African collaboration and networking on pathogen genomic surveillance

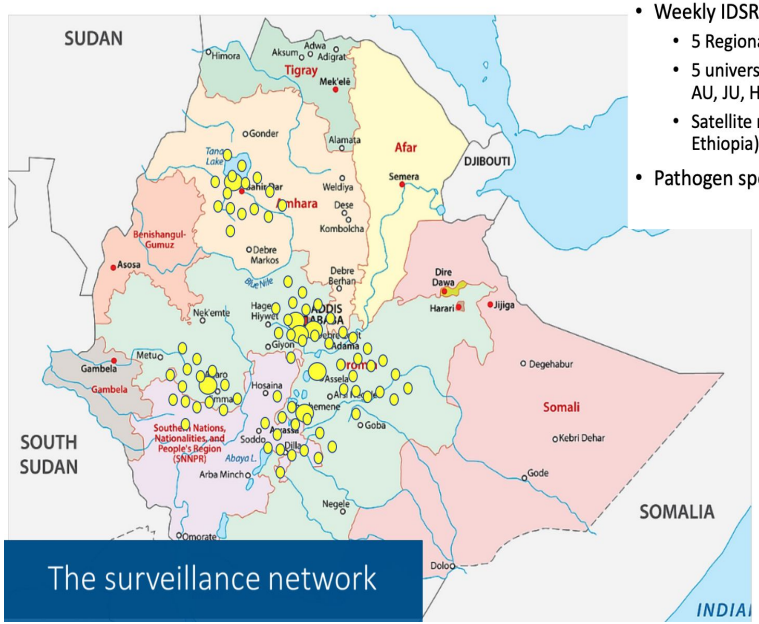
Pathogen Genomics Centre of Excellence @EPHI

Overview of the project

Building Scalable Pathogen Genomic Epidemiology in Ethiopia

'EpiGen-Ethiopia'

Funded by EDCTP3/EU



- Weekly IDSR
  - 5 Regional Public Health Laboratories
  - 5 university-based hospitals (SPHMMC, AU, JU, HU, BU)
  - Satellite networks (400 sites throughout Ethiopia)
- Pathogen specific surveillance in 30 sites

aighd AMSTERDAM INSTITUTE FOR GLOBAL HEALTH & DEVELOPMENT

hhu Heinrich Heine Universität Düsseldorf

UNIVERSITAT DE BARCELONA

wellcome connecting science

COG-TRAIN COVID-19 GENOMICS GLOBAL TRAINING

Current status of projects

- Current active Pathogen Genomic studies at EPHI:
    - EpiGen (recently funded by EDCTP3/EU - BMGF)
    - DESTINE project on Hepatitis-C molecular epidemiology (fund: NIHR-UK)#
    - SUPER (Africa CDC/BMGF)
    - FUO-MetaGenomics (NIH-funded)
- Infrastructure already available at EPHI will be strengthened

Future plans

- **EpiGen** and other projects in collaboration with Africa CDC to serve as a platform

- Establishing '**Center of Excellence**' for Pathogen Genomics at the EPHI

- Foster collaborations with international and regional bodies

Small group activity

You will be split into small groups.

What are the key focus capacity development areas needed in your country?

- On your worksheet, there are areas which are important for sequencing. You need to identify what exists in your country now (as much as you can) and then discuss and decide what would need to be improved/put in place to run successful sequencing.