



PHA4GE Newsletter

November 2021

Editorial

As we wrap up the year that had its own fair challenges with the COVID-19 pandemic, we take pride in the tremendous work done by PHA4GE members. PHA4GE's first **open meeting** held earlier in the year showcased the consortium's Working Groups' activities towards strengthening bioinformatics for public health. The development and increased uptake among public health institutes globally of the **PHA4GE SARS-CoV-2 metadata specification**, is worth mentioning.

Through seed funding from the Bill & Melinda Gates Foundation, 14 teams across Africa and southeast Asia are either implementing data standards for antimicrobial resistance (AMR) and SARS-CoV-2, or implementing

projects in genomics ethics and data sharing.

PHA4GE also engages the wider genomics and public health communities on genomics work through monthly **Global Genomics Webinars** that are free and open for all to participate. In this current quarter, our members have been actively involved in meetings with partners such as the Africa CDC **Pathogen Genomics Initiative**, Bill & Melinda Gates Foundation **Grand Challenges Annual Meeting** and Global Alliance for Genomic Epidemiology (**GA4GH**) **9th Plenary Meeting** all with the aim of strengthening public health bioinformatics in low and middle income contexts. PHA4GE also

partnered with [CLIMB-BIG-DATA](#) and the [Joint Programming Initiative on Antimicrobial Resistance \(JPIAMR\)](#) in running a successful [workshop on antimicrobial resistance \(AMR\)](#).

In this edition of the newsletter, an announcement is given on the projects in genomics ethics and data sharing. The Bioinformatics Pipelines and Data Visualizations Working Group give an update on their current activities. Dr. Rosaline Macharia shares her views on bioinformatics training for different personnel in public health. The map in this edition shows our footprint.

We continue to aspire towards One Health goals and strengthen our global partnerships for bioinformatics for public health.

To all our members and partners, enjoy the rest of 2021!

Rangarirai Matima

PHA4GE Ethics and Data Sharing sub-awards Announcement



Various applications were received for the [call for small grants to fund ethics and data sharing projects](#). The Ethics and Data Sharing Working Group approved four projects for funding. These successful projects are available on the [PHA4GE Ethics and Data Sharing Platform](#) and are listed below.

1. **Ethical oversight of genomic data sharing in research for human health: A case of the Kenya Medical Research Institute (KEMRI) Scientific Ethics Review Unit (SERU)** - *“There is little information about involvement of [Research Ethics Committees] RECs in sub-Saharan Africa with genomic data sharing. It will be interesting to know how the RECs conduct oversight on genomic data sharing.”*

TEAM KENYA 



Mr. Geoffrey K. Ngasura Sang Mr. Enock Kebenei Ms. Lilian Achacha Mr. Kisiinya Odhiambo Cyprian



Ms. Daisy Cheruiyot Ms. Victoria Chepkorir Soi Mr. Gideon Cornel Msee

Kenya Medical Research Institute (KEMRI)

2. **Networking for Ethics Committee biomedical research in**

Central Africa – *“The aim of the project is to strengthen the Ethics review capacity of the institutional ethics committee of [Republic of Congo] RoC, the national ethics committees of [Democratic Republic of Congo] DRC and the Gabon for the evaluation of proposals during outbreaks of infectious diseases. As a secondary objective, the investigators including scientists, health workers, students) in the Brazzaville (RoC) will be strengthened in Good Clinical Practices and on ethics on health research using online tools and materials.”*

TEAM CONGO 



Prof. Francine Ntoumi Dr. Herman Ghislain N'dinga Ms. Carine Kades

Fondation Congolaise pour la Recherche Médicale (FCRM), Congo



Prof. Félicien Munday Mulopo Dr. Pierre-Blaise Matsiegui

Comité National d'Ethique de la santé (CNES) - DRC Comité National d'Ethique pour la Recherche du Gabon

3. **A survey to capture perspectives on the COVID-19 pandemic and vaccine hesitancy, at slums and market areas in Plateau State, Nigeria** – “We aim to capture the knowledge levels and perspectives that focus on the COVID-19 epidemic and the effects on personal choices of protection.”

TEAM NIGERIA 






Dr. Rejoice Helma Abimiku Ms. Sophia Osawe Ms. Felicia Okolo Ms. Olajumoke Felicia Obishakin






Ms. Simi Patience Wilson-Dindam Mr. Hammed Abdullahi Abidemi Ms. Nenbammun Daniel Prof. Alash'le Abimiku

Plateau State Human Virology Research Centre (PLASVIREC), Nigeria

Institute of Human Virology, Nigeria

in the use of molecular-based techniques and microscopy for infectious disease diagnostics... learning modules consisting of video media of experimental scientific protocols, (will be) translated into the main local languages - Shona and Ndebele.”

TEAM ZIMBABWE 





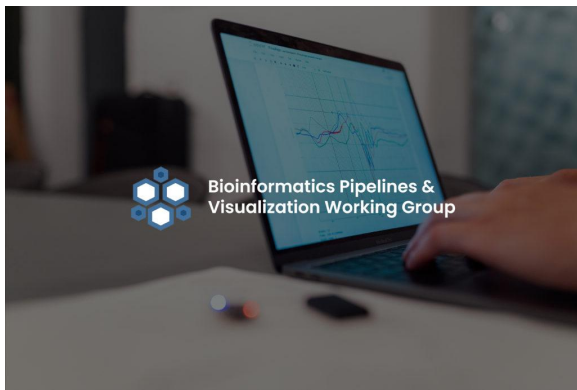
Mr. Peter Chimkupete Dr. Alice Matimba Dr. Joice Ndlovu

De Montfort University, United Kingdom Wellcome Genome Campus, United Kingdom Chinhoyi University, Zimbabwe

4. **Development of educational scientific tools and resources in local languages – ”**

The goal of the project is to pilot development of tools and resources to enhance learning and improve understanding of basic scientific concepts

Bioinformatics Pipelines and Visualization Working Group update



The Bioinformatics Pipeline and Data Visualization Working Group of the Public Health Alliance for Genomic Epidemiology (PHA4GE) has continued to host bi-monthly meetings to discuss critical challenges in the field of public health bioinformatics, highlight valuable

resources and open-source software that address such challenges, and freely discuss working-group members' current projects and scientific pursuits to foster open communication and collaboration amongst members.

In the past few months, meeting highlights have included

- Ongoing conversations regarding the challenges of SARS-CoV-2 data submission to the various internationally-accessible databases such as GISAID, NCBI, and ENA,
- Presentations by the US Centers for Disease Control and Prevention's Technical Outreach and Assistance for States Team (CDC TOAST) outlining their approach to providing SARS-CoV-2 bioinformatics support for US states and territories,

- A presentation by UCSC USHER developers detailing the utility of rapid phylogenetic placement in the context of SARS-CoV-2 outbreak and surveillance investigations,
- Various discussions and presentations about the dynamic nature of SARS-CoV-2 lineage nomenclature, and
- Multiple Happy Hour discussions during which working group members have had opportunities to discuss their current projects and general scientific pursuits to seek feedback from leaders in the field and encourage collaborative work

Outside of the bi-monthly meetings, the working group has continued work in creating, maintaining, and distributing guidance documents and resources that address common challenges regarding

the integration and application of bioinformatics solutions that support global COVID-19 response efforts. These resources have been made publicly available through a PHA4GE hosted Github repository:

<https://github.com/pha4ge/pipeline-resources>.

~ by Kevin Libuit, on behalf of the Bioinformatics Pipelines and Data Visualization Working Group

Community Profile - Dr. Rosaline Macharia

In this quarter's edition of the PHA4GE newsletter, we shine a spotlight on Dr. Rosaline Macharia from the University of Nairobi, Kenya.



Dr. Rosaline Macharia

Please tell us about yourself.

I graduated from SANBI, University of Western Cape with a PhD in Bioinformatics in the year 2016. Soon after I joined the University of Nairobi, Kenya as a full-time lecturer at the Centre for Biotechnology and Bioinformatics (CEBIB). I got interested in the field of Bioinformatics because I found it to be aligned to the tasks I enjoyed as a young girl: playing with computers, data analysis and training. Through my current position, I have an opportunity to train and mentor other

upcoming scientists and to use data analysis to improve public health. Prior to taking up my current role I served as a Bioinformatics consultant at the International Center for Insect Physiology and Ecology (icipe) and as a visiting lecturer in Pan African University of Science and Technology. I am a member of the PHA4GE and the African Biogenome Project and I serve as an external reviewer for the DAAD-in country scholarship programme.

How does your work at University of Nairobi relate with that of PHA4GE?

My current role as a lecturer involves training different cadres of people ranging from students to established scientists. While training a cohort of doctors on Bioinformatics back in the year 2019, I realized that there was a grim need for basic training in Bioinformatics for the personnel in public health. I chose to join the Training and Workforce Development

Working Group under PHA4GE to participate in development of standardized training materials for different groups of people in Public health. I am keen to witness the transformation in the public health sector in our country as a result of such structured training

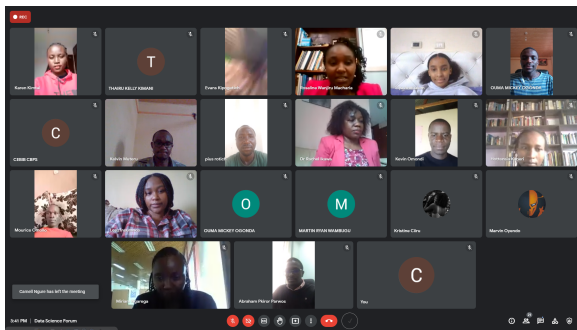
What do you view as one of your greatest achievements in being a scientist?

I would say through science I have developed passion for training and supporting learners to achieve their goals. Good training is key for future development in science.

What is your guidance to anyone who wishes to be in the field of genomics and bioinformatics?

I would encourage one to have curiosity and open-mindedness beyond classwork. It is also important for one to identify the competency they wish to acquire before beginning the formal training.

For more information about the activities of the Training and Workforce Development Working Group, please [click this link](#).



CEBIB Data Science Forum

**Standardization, not perspiration:
#TeamDataStructures partners with organizations around the world to implement data standards to build interoperability**

“Data harmonization should be 10% perspiration, 90% standardization.”

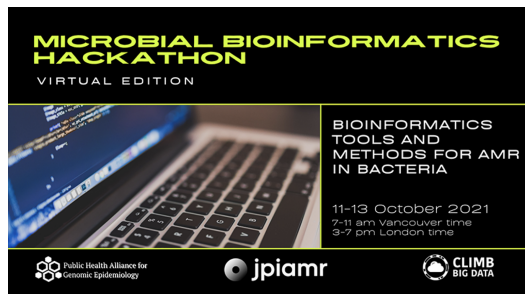
~Data Structures Working Group

A “data standard” is a technical specification that describes how data should be stored or exchanged for the consistent collection and

interoperability of that data across different systems, sources, and users. The Data Structures working group (DS WG) has been very busy this quarter working with the community to improve and implement different PHA4GE data standards, and to train users around the world.

At the beginning of October, the PHA4GE DS WG was pleased to partner with the UK’s Bioinformatics Cloud Infrastructure platform, CLIMB BIG DATA, as well as Europe’s JPIAMR (Joint Programming Initiative on Antimicrobial Resistance) to host the 7th Microbial Bioinformatics Virtual Hackathon. The goal of the hackathon was to bring international bioinformatics researchers, scientists and clinicians together to collaborate and improve/build/extend bioinformatics tools and methods for the AMR community. The three day hackathon attracted over 78 participants from 32 different countries, and projects

included testing/expanding PHA4GE's SNV variant detection standard for AMR (with special focus on its application to harmonizing Mycobacterium tuberculosis (Mtb) mutation-based resistance information), the alignment of AMR databases (programmatic merging and deduplicating), the creation of standardised benchmarking datasets (genomic, metagenomic, assembled, unassembled), the use of uncorrected long read data for AMR analyses, and GPU enabled AMR calling and analysis. Learn more about some of the tools and resources created during the hackathon by visiting <https://github.com/AMR-Hackathon-2021> and <https://github.com/pha4ge/hAMRoni-zation>.

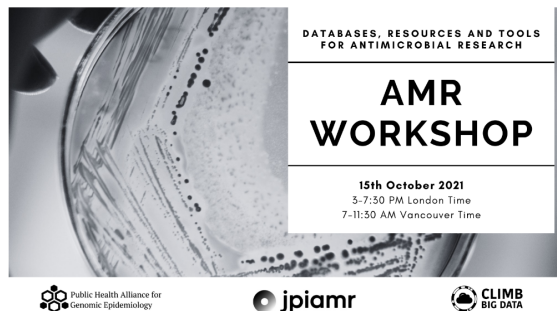


This community effort yielded many impressive results. Most notable for the DS WG were the development of parsers for harmonizing the outputs of different Mtb tools (e.g. TBProfiler, Mykrobe) for easier data sharing and comparisons of analytical results, a layman's "translator" for the widely used (and sometimes tricky to interpret) sequence variant nomenclature HGVS, and the integration of ChamrDb (<https://gitlab.com/antunderwood/chamredb/-/tree/master>) into PHA4GE's hAMRonization package (<https://github.com/pha4ge/hAMRoni-zation>). ChamrDb is a tool for matching predicted genes across different widely used databases to better enable harmonization of AMR gene/variant prediction tool outputs. Many thanks to the hackathon participants who contributed to these achievements!

Applications for the hackathon suggested that many people were interested in AMR bioinformatics training. To accommodate this need, a virtual Joint PHA4GE-CLIMB BIG DATA-JPIAMR AMR Bioinformatics Workshop was held on October 15, which attracted over 200 participants from around the globe. The half day webinar-style event featured talks by Dr. Kara Tsang (London School of Hygiene & Tropical Medicine, UK), who provided an overview of existing AMR-related databases and resources including the Comprehensive Antimicrobial Resistance Database (CARD), NCBI's National Database of Antibiotic Resistant Organisms (NDARO), the Center for Genomic Epidemiology's ResFinder database, and the Pathosystems Resource Integration Center (PATRIC). Dr. Michael Feldgarden (National Center for Biotechnology Information, USA) provided an overview of the theory and use of bioinformatics

tools to detect AMR genes from genomes (e.g., AMRFinderPlus). Ines Mendes (Instituto de Medicina Molecular, PT) described PHA4GE's AMR gene/mutation detection data standard, as well as the hAMRonization program designed to help put standards into practice in public health settings. Dr. Finlay Maguire (Dalhousie University, CA) demonstrated how databases and tools can work together in workflows to efficiently generate AMR gene reports from bacterial genomic reads (https://github.com/fmaguire/amr_training_workshop_practical). Those who missed the event need not worry as all of the talks are freely available online for streaming (<https://www.youtube.com/watch?v=DNsF8U4EsIY&list=PLwflvG-RsluTp6zDaBhcDVC7OthvF5Hpx>). Slide decks are also available for reference and training purposes on the PHA4GE website

(<https://pha4ge.org/amr-workshop-2021/>). Special thanks to Drs. Finlay Maguire, Andrew Page, and Lisa Marchioretto, and the rest of the Steering Committee for all their hard work organizing the hackathon and workshop.



Also in October, members of PHA4GE (Drs. Duncan MacCannell, Danny Park, and Emma Griffiths) met with the GA4GH community to present our activities and explore opportunities for collaboration at the GA4GH Connect meeting (Oct 12). The Global Alliance for Genomics and Health (GA4GH) is a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a human rights

framework (<https://www.ga4gh.org/>). GA4GH and PHA4GE will continue to align efforts and work together to implement data standards in the future.

An important part of putting standards into practice is training public health personnel to use them. The PHA4GE DS WG was pleased to partner with the The Africa Pathogen Genomics Initiative (Africa PGI)'s NGS Academy

(<https://africacdc.org/institutes/ipg/>) to offer a workshop on SARS-CoV-2 contextual data curation and stewardship. Contextual data for public health microbial genomics consists of all the clinical, epidemiological, lab data as well as sample metadata, methods information and quality control metrics that enable the interpretation of sequence data for decision making. This information is often encoded in different systems and spreadsheets using different fields, terms, and formats. Different public repositories have different submission

requirements, and different public health agencies have different reporting requirements. Participants learned about the PHA4GE contextual data specification

[\(<https://github.com/pha4ge/SARS-CoV-2-Contextual-Data-Specification>\)](https://github.com/pha4ge/SARS-CoV-2-Contextual-Data-Specification)

and how this information can be standardized to improve interoperability and analyses, as well as different privacy and practical considerations for data sharing.

The workshop was very timely, as the variability in global data standards for SARS-CoV-2 contextual data can be difficult to navigate. As this topic is of broad interest, the PHA4GE DS WG will also be participating in SARS-CoV-2 data standards workshop hosted by ELIXIR, an intergovernmental organisation that coordinates European compute resources, including databases, software tools, training materials, cloud storage and more

[\(<https://elixir-europe.org/about-us>\)](https://elixir-europe.org/about-us).

The PHA4GE DS WG will present different data standards (e.g. GISAID, ENA, WHO) and how to map between them. We will also describe how the PHA4GE SARS-CoV-2 contextual data specification and associated tools and protocols can provide data structures to help better standardize the information going into different repositories. The workshop is scheduled for Nov 16 (3pmCET), and those interested can register for the workshop [here](#).

To learn more about our activities, and/or how to join, [check out our website](#).

Publications

Corner

Take a read on some of the bioinformatics for public health papers published in the last three months

Whole genome sequencing of clinical samples reveals extensively drug resistant tuberculosis (XDR TB) strains from the Beijing lineage in Nigeria, West Africa

- Highlights the importance of whole genome sequencing (WGS) in public health and diagnosis of MDR and XDR tuberculosis in Nigeria. The major discovery in this study is that the authors were able to identify the Beijing lineage in Nigeria for the first time using WGS from direct clinical samples which were also extensively drug resistant.

MicrobeTrace: Retooling molecular epidemiology for rapid public health response

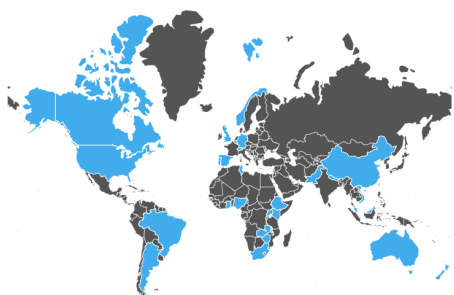
- A useful browser-based tool which combines contact tracing and/or microbial genomic data with demographic or behavioural information to investigate disease outbreak data

A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa

- An important paper with over a 100 African groups collaborating on describing genomic surveillance analysis of SARS-CoV-2 variants and lineages across Africa.

PHA4GE

Membership



■ PHA4GE Members

As of November 2021, PHA4GE has 173 members based at over 25 institutions around the globe

Interested in being featured in our newsletter?

Send an email to communications@pha4ge.org

Events

We'd like to share some of the other upcoming events in our community.

	ASLM2021 Conference 15 - 18 November 2021 Learn More
	Genomics in Health Implementation Forum (GHIF) 16 - 17 November 2021 Learn More
	ASTMH 2021 Annual Meeting 17 - 21 November 2021 Learn More

Web edition: bit.ly/3bbnxZm

Website: pha4ge.org

Twitter: [@pha4ge](https://twitter.com/pha4ge)

Facebook: facebook.com/pha4ge