Public Health Alliance for Genomic Epidemiology

PHA4GE Newsletter

March 2022

Editorial

2022 kicked off with PHA4GE Working Groups continuing with closing-off activities from yesteryear, which all have an aim to strengthen microbial bioinformatics for public health.

In this edition of our quarterly newsletter, the Data Structures Working Group gives an update on upgrading their SARS-CoV-2 and AMR tools.

From some of the sub-awardees, we hear of their success stories on implementing standardized SARS-CoV-2 and AMR tools within national public health laboratories.

The Ethics and Data Sharing Working Group gives an account on their new publication that interrogates benefit sharing in health research and proposes a framework that can be applied for benefit sharing in health research.

Do enjoy the read!

Rangarirai Matima

Responding to community needs: The Data Structures Working Group is updating its tools for AMR and SARS-CoV-2

The ability to compare how SARS-CoV-2 lineages are evolving around the world in different contexts depends on harmonizable contextual data across labs and datasets. Contextual data is the sample metadata, methods information as well as the lab, clinical and epidemiological data that enables the interpretation of sequence data. In

August 2020, the Data Structures Working Group (DSWG) responded to the need for a contextual data standard designed for public health genomic surveillance and pandemic response releasing its first contextual data specification. The specification contains standardized fields and terms for critical information such as sampling strategies, information about samples and hosts, as well as software and sequencing tools. Over the past two years, the specification has continued to evolve according to user data needs and requests. In December 2021, the DSWG released a major update to the PHA4GE SARS-CoV-2 contextual data specification. Version 3.0 contains updated vocabulary, improved mappings to public repositories as well as contextual data recommendations made by the World Health Organization, and includes terms and identifiers from 24 different OBO Foundry Ontologies to improve interoperability and to implement FAIR principles (Findable, Interoperable, Accessible, Reusable) for scientific data management. The specification was also published in February 2022 in GigaScience, and is supported by a number of tools, reference materials and protocols for data curation and submission. Read more about the specification package in GigaScience here.

In 2021, members of the DSWG worked with 10 teams across Africa and southeast Asia to implement data standards for antimicrobial resistance (AMR) and SARS-CoV-2, through seed funding from the Bill & Melinda Gates Foundation. The goals of these partnerships included piloting the standards and resources developed by PHA4GE in real-world settings, learning from partners in a wide variety of contexts about how they should be improved, and building lasting relationships with public health bioinformatics practitioners in the community. One such partnership included a team led by researchers at the National University of Malaysia (UKM). The team piloted PHA4GE's "hAMRonization" - a specification and command-line parsing tool used to harmonize the outputs of widely used gene and mutation detection software in a standardized report - for sharing data about clinically relevant methicillin resistant Staphylococcus aureus isolates between labs in Malaysia and Argentina.

In the past month, the DSWG met with the team, which included researchers Dr. Hui-min Neoh, Dr. Su Datt Lam, Dr. Sabrina Di Gregorio, Mr. Mia Yang Ang, Dr. Tengku Zetty Maztura Tengku Jamaluddin and Prof. Dr. Sheila Nathan to discuss further improvements to the tool and its supporting materials. These discussions included ways the Malaysia team could increase the usability of the tool for non-bioinformatician colleagues working in hospitals. The team created a “Google Collaboratory'' (known as a Google Colab) which enables users to execute python code through a browser without any software installations. Google Colabs are Jupyter notebooks that run in the cloud and are highly integrated with Google Drive, making them easy to set up, access, and share. The team hopes that the simplicity of the Google Colab version of hAMRonization will better enable their colleagues (which include clinicians and microbiologists less familiar with command-line) to quickly compare antimicrobial resistance in hospital settings. The DSWG is now working with the Malaysia team to make the Google Colab publicly available in GitHub. The Malaysia team will be hosting a workshop in April that will include training for using hAMRonization via the Google Colab.

PHA4GE sub-awards success stories

Thanks to the Bill and Melinda Gates Foundation, PHA4GE received funds that were utilized to implement standardized bioinformatics practices, pipelines, and data structures in either antimicrobial resistance (AMR) or SARS-CoV-2 sequencing within national public health laboratories. We look at six teams, from five different low to middle income countries (LMICs), that successfully completed these projects. They share what worked well, what could have been done differently and what they envision as next steps in being responsive to disease outbreaks.

AMR Theme

Team Cambodia

Dr Vandelannoote *, what worked well in implementing the AMR project?

The Medical Laboratory of the Institut Pasteur du Cambodge (IPC) has been working with partnering institutions to build genomic AMR surveillance capacity for Cambodia. We have been exploring various bioinformatic pipelines capable of analyzing genomic AMR data in an interoperable manner. An awarded PHA4GE sub-grant supported us in exploring the PHA4GE AMR gene detection output standard for exchanging AMR genomic surveillance results between IPC and the Microbiological Diagnostic Unit Public Health Laboratory (MDU PHL) at the Peter Doherty Institute, who have implemented pathogen genomics for public health in the state of Victoria (Australia).

In your own opinion, what did not work well and how could this have been done differently?

We didn't experience any issues during the installation / implementation of hAMRonization and successfully performed a series of exchange exercises with MDU-PHL. During these exercises we didn't identify any discrepant results. We did notify the developers of a number of minor points that could be better described in the hAMRonization documentation.

What are the next steps for you and your team?

We plan to continue using hAMRonization when comparing output from different AMR gene detection tools during exchanges with other institutes.

**Dr Koen Vandelannoote worked with Rutaiwan Dusadeepong, Gauthier Delvallez, Kristy Horan, Tuyet Hoang and Tim Stinear to complete this project.*

Team Malaysia

Dr. Neoh*, what worked well in implementing the AMR project?

Our team members are from laboratories in different locations (i.e. Malaysia, Argentina and Tokyo). The open access format with clear and easy installation instructions from the

hAMRonization GitHub made the platform very user-friendly. hAMRonization helped us obtain standardized analysis output from various AMR software used in the different laboratories of our team members; this eased the comparison of AMR genes from pathogens of interest that we were working on. We found both html (GUI) and excel output formats of hAMRonization useful and we highly recommend the platform for our colleagues working on AMR.

In your own opinion, what did not work well and how could this have been done differently?

Due to inherent differences in AMR software algorithms, there will be some differences in the AMR analysis output between laboratories which use different AMR software for analysis. This could be solved with collaborating laboratories using the same software for analysis prior to plugging the output into hAMRonization. Inclusion of epidemiological and clinical information (besides AMR gene information) into the tool will be useful for public health analysis.

What are the next steps for you and your team?

We are working to establish a Google Colab suite to ease hAMRonization installation in laboratories without bioinformaticians. We also plan to create a manual of the tool for future users.

**Dr Hui-min Neoh (bottom right corner) worked with (from left to right) Sabrina Di Gregorio, Su Datt Lam, Sheila Nathan, Mia Yang Ang and Tengku Zetty Maztura Binti Tengku Jamaluddin to complete this project.*

Team Nigeria

Dr. Okeke and Dr. Afolayan, what worked well in implementing the AMR project?

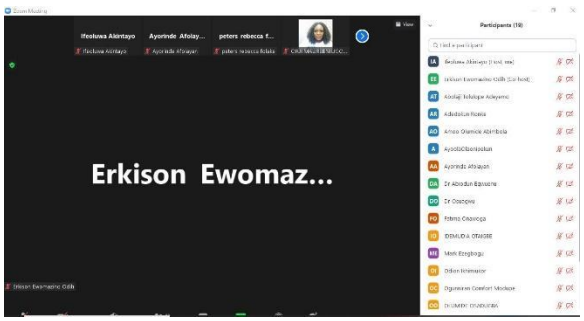
The AMR Project was aimed at providing an entry point for hospital laboratory scientists and field epidemiologists in the understanding of genomic science, the analysis of sequence data, and interpretation of sequence data. We were able to successfully organize two modules of bioinformatics virtual workshops which introduced participants to the use of web- and command line-based tools in analyzing and interpreting genomic sequence data. Participants were also able to utilize our successfully-launched Nextflow Tower platform that deploys our AMR gene prediction pipeline together with a crucial Hamronization step to predict AMR genes in silico, as well as generate a report for the exchange of AMR genomic surveillance data.

In your own opinion, what did not work well and how could this have been done differently?

As undeniable as our successes were, nothing beats a physical bioinformatics workshop. One reason is that some participants faced challenges in the installation and execution of a bioinformatics tool on their local PCs during the second module of the workshop, and it would have helped if a bioinformatics tutor was on ground to help out quickly enough. Although we were able to solve almost all installation challenges remotely, the process was much slower than one would have loved it to be, and this sometimes extended the workshop time for the day. Knowing that this pandemic will not be over any time soon, a third module of the bioinformatics workshop slated to take place later this year will most likely be conducted virtually. Luckily, all the bioinformatics tools needed for the workshop are already installed on our server, thereby nullifying any potential installation issues.

A few of the participants did not have the foundation for the first module, which made learning a little difficult for them, and for the rest of the cohort. We are now developing and piloting a pre-assessment and a 'Module Zero' of basic foundational concepts that could prepare participants with little molecular biology background to succeed in the first module.

Odion Ikhimiukor



Erkison Ewomazino Odih's Presentation

Gather

What are the next steps for you and your team?

We intend to conduct a third bioinformatics module so that participants who have attended the first and second module of the workshop will have a more grounded understanding and interpretation of sequence data, which is a necessary skill for the enhancement of Nigeria's growing genomic AMR surveillance

efforts. Afterwards, we will conduct a Train-the-Trainer workshop for a subset of participants so that they can, in turn, conduct similar bioinformatics training sessions in their hospitals or research centers around Nigeria.

Our modules one and two are templates for ongoing and future training programs. For example, we are preparing a cohort of African typhoid researchers to enter an adaptation of Module 1 next month.

We also intend to make our analysis pipelines more accessible to the public using the web-based Nextflow Tower platform, so that public health practitioners with a working knowledge of genomic science but little or no experience in the use of the command-line can analyze and interpret data easily. Funding for the development and maintenance of these pipelines, as well as funding for the storage and computation costs on Amazon Web Services (where all submitted analyses "jobs" via the Nextflow Tower Platform are run), will ensure continued development and sustenance of sequence analyses pipelines.

**Dr Iruka N. Okeke and Dr. Ayorinde O. Afolayan worked Erkison Ewomazino Odih, Odion Ikhimiukor, Rotimi Dada, Ifeoluwa Akintayo and Faith Popoola Oni to complete this project.*

SARS-CoV-2 Theme

Team Kenya

Dr. Oyola and Dr. Entfellner*, what worked well in implementing the SARS-CoV-2 project?

The practical implementation of technical sample and data analysis worked well. This started with a very intensive training of research assistants on the general lab molecular procedures in a laboratory where varied genomic projects have been running. As a result of a worldwide effort to control SARS-CoV-2, technical information of developing a streamlined workflow on genomic surveillance was readily available and regularly updated by

different research groups. Our team was vigilant in acquiring the latest technical development and implementing efficiently under the supervision of the principal investigator. In a relatively short period of time, technical aspects of SARS-CoV-2 genomic analysis and near real-time reporting of results to the Ministry of Health were established and scaled to high throughput levels. Access to seed funding, presence of a lead genomic expert within the Institute, steady sample collection and referral and good collaboration with the government and her public health teams were the major drivers of this success.

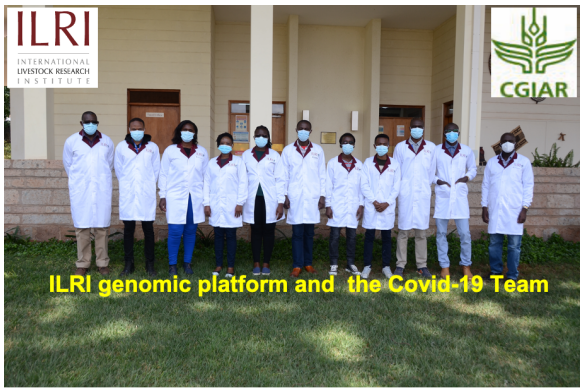
In your own opinion, what did not work well and how could this have been done differently?

Although great success was achieved in the end, a steady supply of reagents was a big challenge. With global demand for similar reagents, laboratories in the developing world received a raw deal. There was a massive delay in procuring and delivery of critical reagents. Coupled with a dynamic workflow where new and optimized reagents were continuously produced, it was difficult to keep up-to-date with the improvement in sample analysis. This challenge could have been subverted by donors' involvement to ensure direct access to critical reagents from manufactures by the funded labs in the developing world.

What are the next steps for you and your team?

One experience gained from the ongoing global pandemic is the powerful role of genomics in controlling a disease outbreak. We have strengthened our genomic platform with updated equipment including sequencing machines and high-performance computing infrastructure. We have also established a platform for genomic training and prepared materials for a continuous capacity development program. Our goal is to create and establish a

genomic response program to pandemics and disease outbreaks. This platform will continue to run research programs on detection, discovery and genomic characterization of emerging and potential pandemic-causing pathogens. We will focus on creating pathogen genomic analysis tools and pipelines with the ability to analyze different kinds of pathogens. Additionally, we are expanding our network for sample collection and referral systems to ensure that our surveillance logistics are swift, timely and efficient.



**Dr Samuel O. Oyola and Dr Jean-Baka Domelevo Entfellner worked with Collins Muli, Gilbert Kibet, Daniel Ouso, Shebbar Osiany, Edward Kiritu, Paul Dobi and Dr Sonal Henson to complete this project.*

Team Malawi

1.Dr. Kamng'ona*, what worked well in implementing the SARS-CoV-2 project?

We had dedicated time working together and were able to learn modern bioinformatics techniques. During the Delta wave, for example, we were analyzing data that had been generated only a few days before as part of the

pandemic response, which helped to keep things "fresh". We were quite excited to be able to train and transfer the SARS-CoV-2 sequence data analysis skills to emerging scientists working on the project. It was also quite satisfying to be able to set up containers on servers in the UK, and then deploy those pipelines on Kamuzu University of Health Sciences (KUHeS) infrastructure by the end of the project. As a team, we were able to engage the Ministry of Health through the Public Health Institute of Malawi (PHIM) and use their platform to highlight the goals of the PHA4GE project. Another notable achievement is the purchase of the MinION Sequencing machine, and we are already in discussion with PHIM and other relevant stakeholders to determine how best we can collaborate.

2. In your own opinion, what did not work well and how could this have been done differently?

Despite continuing our collaboration with PHIM, there was an 'anti-climax' after the project, in which team members returned to their regular projects, creating a sense of lack of direction.

How this could have been done differently may be a difficult question to answer but suffice it to say that we should have paid more attention to continuity after the project.

However, with the acquisition of the MinION, we hope to be able to strengthen our existing collaborations and put the learning into practice a little more.

3. What are the next steps for you and your team?

I think it will be great to sequence some SARS-CoV-2 at KUHeS using the Mk1C and then analyze it there with the pipelines deployed as part of the PHA4GE project.

The MinION Mk1C sequencer and reagents finally delivered!

**Dr Arox Kamng'ona worked with Philip Koen, Khuzwayo Jere and Benjamin Kumwenda to complete this project.*

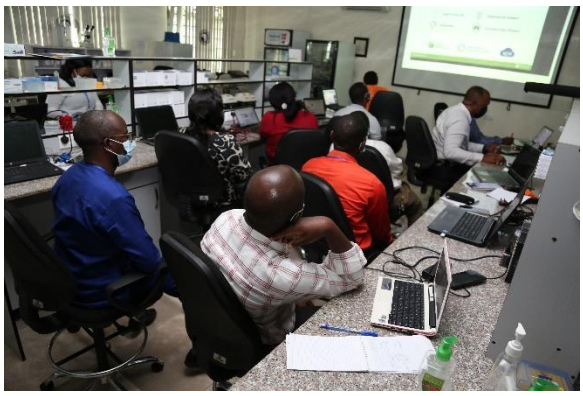
Team Nigeria

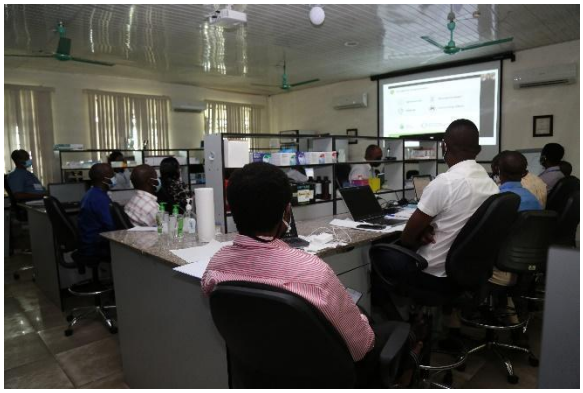
Dr Olawoye and Dr Oluniyi, what worked well in implementing the SARS-CoV-2 project?

We were able to train 20 public health scientists in Nigeria on SARS-CoV-2 genome analysis, data sharing and metadata standards. We were also able to improve our bioinformatics capacity and processes by procuring equipment for data analysis and internet plans for data processing.

In your own opinion, what did not work well and how could this have been done differently?

We conducted a follow-up zoom meeting/training with the participants and we experienced a drastic reduction in participation numbers. However, we recorded the sessions which were shared with them but we envisaged that the concentration levels of the participants were not on par with the hands-on training that was initially held. We believe that a hands-on approach would yield better results than the online training sessions.





What are the next steps for you and /or with your team?

We plan to build on the already established networks and also expand our training to more public health labs within Nigeria and the Africa region as many labs require skills in bioinformatics, standardizing workflows and contextual data.

Ethics and Data Sharing Manuscript

...a chat with Nicki Tiffin on the recent publication of an ethics and data sharing paper....

Congratulations on the publications of your Ethics and Data Sharing paper! Tell us, what prompted you to write this paper?

Thank you, we are very pleased to have published the benefit sharing framework in BMJ Global Health. There is increasing recognition of the need to implement benefit sharing as part of all research programmes, alongside the increased need to ensure equitable research practices. I do think, however, that many research stakeholders are not sure how to practically implement or describe benefit sharing efforts even though their projects may already contain some elements of benefit sharing. We hope that the framework that we have developed will help researchers, funders and other stakeholders in health research to identify ways that they can include benefit-sharing in their projects, and also describe benefits that may already be shared during their programmes. The framework is designed to make it much easier to design and implement benefit sharing from the earliest stages of programme or project design.

Please share with us what the paper is about.

In the paper, we have defined two dimensions to benefit sharing: The first describes the stakeholders who might benefit from a programme, recognising that there are many different types of stakeholders ranging from individuals such as research participants and researchers, all the way through to global or regional organizations. The second dimension identifies types of benefit sharing that are possible – they might be financial, but we also describe many other types of benefits, for example benefits might impact health and well-being, research careers or economic activity of stakeholders. Using these two dimensions in a matrix makes it possible to identify and plan ways to

ensure specific benefits reach relevant stakeholders.

What is the take home message?

Our take home message is that this systematic approach to identifying benefit sharing opportunities can break benefit sharing down into manageable and practical pieces, and the framework can be used by a wide variety of health research stakeholders to plan and undertake intentional benefit sharing in their diverse programmes.

So, what projects are you working on next?

We are in the process of writing up our risk framework which partners the benefit sharing framework: the two frameworks are designed to complement each other in supporting the design of ethical and equitable research. Undertaking ethical research requires balancing risks and benefits for stakeholders, and the two frameworks can be used synergistically to identify the risks and benefits of a programme and to ensure that they are distributed fairly between stakeholders.

The full title for the paper is "A framework for the promotion of ethical benefit sharing in health research" and is **available** in the BMJ Global Health Journal.

Publications Corner

"...highlighting some of the bioinformatics for public health papers published in the last three months..."

A framework for the promotion of ethical benefit sharing in health research

- A PHA4GE Ethics and Data Sharing Working Group paper on benefit that proposes a framework that can be applied for benefit sharing in health research.

Future-proofing and maximizing the utility of metadata: The PHA4GE SARS-CoV-2 contextual data specification package

- A PHA4GE Data Structures Working Group paper describing the PHA4GE SARS-CoV-2 contextual data standard

Events

13th International Meeting on Microbial Epidemiological Markers (IMMEM XIII)

Bath, United Kingdom

14–17 September 2022

Interested in being featured in our newsletter?

Send an email to communications@pha4ge.org

Web edition: bit.ly/3vInosq

Website: pha4ge.org

Twitter: @pha4ge

Facebook: facebook.com/pha4ge