Public Health Alliance for Genomic Epidemiology

Robert Sobukwe Rd, Bellville, Cape Town, 7530, South Africa

SUB-GRANTS PROMOTING SUSTAINABLE DEVELOPMENT IN BIOINFORMATICS TO SUPPORT PUBLIC HEALTH (REFERENCE: PHA4GE/2021-01)

Public Health Alliance for Genomic Epidemiology (PHA4GE) (www.pha4ge.org) is a global consortium established in 2019, to ensure a rapid global genomic-driven public health response to disease outbreaks. Our community includes partners and stakeholders such as funders, public health organizations at international, regional and national levels; and non-specialist individuals across the world. The consortium has Working Groups, that are made up of representative individuals, from the growing list of over 30 partner organizations. These Working Groups drive various activities to fulfil the consortium's mission. Our mission is to establish global consensus on data standards, document and share best practices, improve the availability of critical bioinformatic tools and resources and advocate for greater openness, interoperability, accessibility and reproducibility in public health bioinformatics.

SECTION A: APPLICATION PROCESS

1. Overview

Public Health Alliance for Genomic Epidemiology (PHA4GE) is calling for proposals for six (6) sub-grants with a total value of USD150,000 to assist work in bioinformatics within low-to-middle-income countries (LMICs) in Africa and Asia. This encompasses development of open and interoperable public health bioinformatic software, development of standardized bioinformatic tools and infrastructure, usage of bioinformatic software or tools, implementation of standards developed by PHA4GE, maintenance of bioinformatic software or infrastructure, capacity development in bioinformatics, and any other related activities to bioinformatics.

1.1 The themes to these sub-grants are as follows:

Theme 1: Implementation of standardized bioinformatics practices, pipelines, and data structures in AMR sequencing laboratories in LMICs (3 X 20 000USD)

Antimicrobial resistance (AMR) is a global health problem that contributes to tens of thousands of deaths per year around the globe. The World Health Organization has declared that antimicrobial resistance is one of the top ten global public health threats facing humanity, and has identified large gaps in existing surveillance. Genomic surveillance of AMR is a powerful tool that is increasingly being used to understand the evolution of resistance, as well as perform source attribution and track the spread of AMR determinants (genes). A number of tools exist that can be used to detect the presence of AMR genes in public health microbial sequence data. These widely-used AMR detection tools differ in terms of their inputs, functionality (including parameters and reference databases), and outputs. Differences in the meaning, structure and range of values in the different outputs of these tools can make comparing and interpreting results difficult for public health laboratories. To address these issues, PHA4GE has developed a standardized AMR gene detection output specification (https://github.com/pha4ge/hAMRonization/blob/master/docs/hAMRonization_specification_details.csv) to better harmonize AMR analysis results, resulting in improved interoperability and information exchange across tools and laboratories. The goal of this award is to put data standards into practice to better facilitate the exchange of AMR genomic surveillance data between public health laboratories. Projects should describe how teams will collaborate to use the specification and associated resources to harmonize and exchange AMR surveillance data, where data sharing did not previously exist, or was difficult.

Theme 2: Implementation of standardized bioinformatics practices, pipelines, and data structures in SARS-CoV-2 sequencing laboratories in LMICs (3 X 30 000USD)

Genome sequencing of the SARS-CoV-2 virus has been a key tool for understanding the spread of the disease at global, national and local scales, developing diagnostic tests and vaccines, and refuting misinformation during the pandemic. With the emergence of viral variants of concern, sequencing and sequence analysis have become essential activities for national public health institutes (NHPIs). While routinely used pipelines for SARS-CoV-2 analyses are available in the scientific community (https://github.com/CDCgov/SARS-CoV-2_Sequencing), laboratories experience challenges associated with SARS-CoV-2 analyses include difficulty in submitting raw sequences to international databases such as GISAID and INSDC in an interoperable and standardized manner; lack of access to reliable compute resources to execute analyses and training of public health professionals in bioinformatics skills for SARS-CoV-2 analysis.

PHA4GE and its partners have identified best practices for infrastructure, analysis pipelines and metadata associated with SARS-CoV-2 sequencing. The goal of the award is for NPHIs' to work with PHA4GE to adopt these best practices identified as they build capacity for SARS-CoV2 sequence data analysis and data sharing. Applications are sought that would implement interoperable and reproducible SARS-CoV-2 analysis and sharing of data via INSDC, GISAID and any other international database.

1.2 The scope to these sub-grants is detailed in ***Appendix 1: Scope for PHA4GE subgrants***; that can be found in **SECTION B**.

2. Eligibility Criteria

2.1 This call is open to the following types of sequencing laboratories:

- national public health institutes (NPHIs) in low to middle income countries in Africa and Asia
- academic bioinformatics laboratories in Africa that partner with any NPHIs in Africa and Asia
- public health laboratories in Africa and Asia

2.2 Only one award may be awarded to one organization.

2.3 The organization should not have any cases levelled against it by any Professional Councils and/or public courts relating to professional misconduct.

2.4 None of the desired funds will be diverted to support activities that are not related to the submitted proposal. Allowable expenses include salaries and travel costs.

2.5 None of the individuals to be involved in the work, such as staff and volunteers, have potential conflicts of interest.

3. Timelines

Spending of the grant may cater for deliverables that are within a time-frame of five (5) months.

4. Format for proposals

All proposals are to be submitted, in English, on the **PHA4GE Sub-grants Proposal Form** (*Appendix 2)*, that is attached to this call**.**

5. Submission of proposals

5.1 All submissions are to be submitted to **PHA4GE Secretariat** as **a single .pdf file** on the following email address: subawards@pha4ge.org. Please use the following subject line: **Application for Bioinformatics sub-grant (PHA4GE/2021-01)**

5.2 Regrettably, incomplete proposals and/or submissions will not be reviewed.

5.3 Only successful applicants will be informed within thirty (30) days after the closing date.

Closing date for applications: 10 March 2021

For any assistance or queries relating to finalising and submission of this application, please email us at help@pha4ge.org

SECTION B: APPENDICES

Appendix 1: Scope for PHA4GE subgrants

Theme 1: Implementation of standardized bioinformatics practices, pipelines, and data structures in AMR sequencing laboratories in LMICs

1.1 Background

This sub-grant is intended for public health and research labs in LMICs that are performing genomic surveillance of antimicrobial resistance and are seeking guidance and resources to improve data exchange. PHA4GE has developed a standardized AMR gene detection output specification to better harmonize the AMR detection results across tools and resources and improve interoperability (https://github.com/pha4ge/hAMRonization/blob/master/docs/hAMRonization_specification_details.csv). PHA4GE has also developed different tools to enable labs to benchmark different AMR gene detection software, and parse outputs according to the data standard (https://github.com/pha4ge/hAMRonization).
The goal of the project is to implement the AMR specification package to facilitate data exchange between laboratories where interoperability currently prevents sharing, integration, comparison, and interpretation of AMR genomic surveillance data.

1.2 Funding Opportunity Description

Funding scope: The aim of this award is to put data standards into practice to better facilitate the exchange of AMR genomic surveillance data between public health laboratories. Teams will work with PHA4GE to utilize the AMR specification package (standard and tools) to demonstrate successful data exchange between laboratories where interoperability currently prevents sharing, integration, comparison, and interpretation. While we hope that these awards catalyze lasting benefits for AMR genomic surveillance data exchange, the purpose of the award is to demonstrate that harmonization and interoperability can be enhanced through implementation of data standards and specifications.

As such, steps to performing this AMR genomic surveillance data exchange project are:

1. Teams of two or more laboratories will identify as a “network” that aims to exchange AMR genomic surveillance data.
2. Teams will describe how currently, the lack of interoperability between analysis workflows and/or standardization in reporting formats creates barriers to data sharing/integration/interpretation in their network.
3. Teams will describe how interoperable data and harmonized reporting would benefit their AMR surveillance programs.
4. Teams will work with PHA4GE to use the AMR gene detection output standard and accompanying tools to produce harmonized reports.
5. Teams will use the harmonized reports to exchange AMR gene detection results.

1.3 Objectives

Awardees will work with PHA4GE to:

1. To facilitate exchange of AMR gene detection information between laboratory networks to improve public health genomic surveillance of AMR.
2. To demonstrate that harmonization and interoperability can be achieved through implementation of the PHA4GE AMR gene detection output specification package.
3. To provide feedback to PHA4GE for improving their tools and standards.

1.4 Deliverables

Awardees will:

1. Develop a harmonized report based on the PHA4GE AMR gene detection output standard for exchanging AMR genomic surveillance results.
2. Perform an exercise using the report to exchange AMR genomic surveillance results.
3. Write a report outlining their successes, the methods used to create the report and use the AMR specification package, describe their experience using the standard and tools, and provide suggestions for implementation on a broader scale.

Theme 2: Implementation of standardized bioinformatics practices, pipelines, and data structures in SARS-CoV-2 sequencing laboratories in LMICs

2.1 Background

This sub-grant is intended for public health and research labs in LMICs that are beginning to sequence SARS-CoV-2 genomes from human patients and are seeking guidance with data analysis and data sharing best practices. This award intends to closely partner such labs with working groups in PHA4GE that can equip labs with the tools needed to routinely analyze, report on, and release SARS-CoV-2 genomic data generated within their lab, including the contextual data (epidemiological, clinical, laboratory and methods information) critical for interpreting different kinds of analyses based on public health sequence data.

The goal of this project is to develop the processes and practices required to continually produce, analyze, and share high quality sequence and contextual data with public repositories.

2.2 Funding Opportunity Description

Funding scope: awardees should have the knowledge and capability to process SARS-CoV-2 patient samples, produce libraries, and generate sequencing data, using a commonly accepted Illumina or Oxford Nanopore based laboratory protocol. Awardees should already have access to sequencing equipment and viral samples. Although consumables may be budgeted for with this award, this award is intended primarily to fund the analysis, data management, and project costs.

During the course of this project, awardees will learn and routinely utilize containerized bioinformatics pipelines to analyze SARS-CoV-2 sequencing data, set up or adopt an appropriate compute infrastructure to execute these pipelines, and publicly release their data using the PHA4GE contextual data specification. PHA4GE will provide guidance, recommendations, and feedback on each of these elements as needed.

2.3 Objectives

Awardees will work with PHA4GE to:

1. Establish processes to generate FASTQ and FASTA files with appropriate quality controls
2. Establish processes to create (PANGO and nextclade) lineage prevalence reports to facilitate surveillance
3. Establish a compute environment capable of running containerized workflows to perform the above analyses.
4. Establish processes to capture standardized contextual data, including bioinformatics methods

2.4 Deliverables

Awardees will:

1. Generate SARS-CoV-2 viral sequence data on at least two small batches of samples (approximately 12 samples each).
2. Submit sequence and contextual data to INSDC repositories (e.g. SRA+GenBank (NCBI) or ENA (EMBL-EBI)), GISAID and any other international database.
3. Document their implemented bioinformatics protocols (i.e. on protocols.io or GitHub using a GitBook (https://www.gitbook.com/).
4. Write a report outlining their successes and challenges, and suggestions for other labs implementing PHA4GE recommendations and resources for SARS-CoV-2 analyses.