Improving AMR genomic surveillance and knowledge translation in public health settings using PHA4GE's hAMRonization specification and tool

Background

Antimicrobial resistance (AMR), the ability of microbes to proliferate despite exposure to antimicrobial drugs, represents a present and growing public health crisis with a global scope. Improving surveillance of AMR in clinical, animal and environmental settings to better understand AMR evolution and transmission is a core component of national and international action plans to mitigate this crisis. The detection of AMR determinants directly from genomic data has become a standard procedure in public health surveillance, with a large number of different bioinformatic tools currently available to perform this task (e.g., Abricate, AMRFinderPlus, RGI, ResFinder etc). These tools, although implementing similar principles, differ in supported inputs, search algorithms, parameterisation, and underlying reference databases.

The Challenge

These many AMR gene detection tools each generate output reports of detected AMR genes in distinct, non-standard formats. This is a huge barrier to the comparison of results across tools, modularity of tools within workflows, and communication of results between national surveillance efforts and to downstream knowledge users.

The Public Health Alliance for Genomic Epidemiology (PHA4GE) AMR Gene Detection Output Specification Package

The Public Health Alliance for Genomic Epidemiology (https://pha4ge.org) is a global coalition that is actively working to establish consensus 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 microbial bioinformatics.

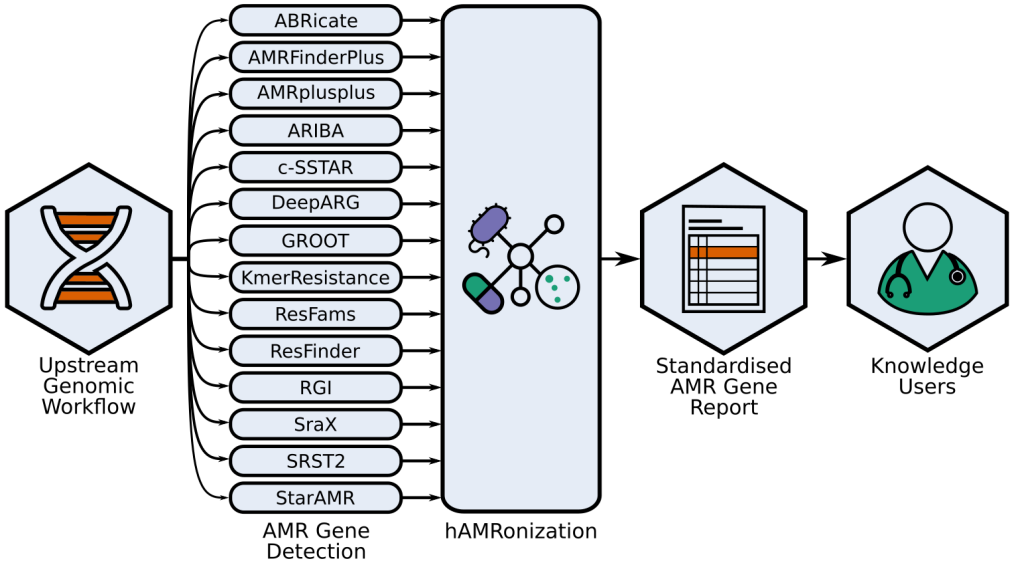

Figure 1: PHA4GE's hAMRonization package within a typical AMR genomic analysis workflow. Without hAMRonization each AMR gene detection tool would generate differently formatted outputs and thus would require tool-specific downstream workflows with tailored communication to knowledge users (e.g., clinicians, public health experts). hAMRonization enables changing (and comparison) of the AMR gene detection software without impacting the rest of the workflow or resulting information dissemination.

To facilitate public health AMR surveillance, PHA4GE has developed a data standard for the bioinformatic detection of AMR genes from genomic data through iterative testing and expert consultation.

This specification provides standardized field names for reporting AMR gene detection results along with definitions, guidance, and ontology identifiers to validate supplied values across fields. The specification has been implemented in the "hAMRonization" which automatically converts the varied outputs of the 14 most-common species-agnostic AMR gene detection tools to a standardized unified report following this PHA4GE standard. This tool also supports creation of summary reports across tools and data-sets in a variety of formats (tabular, JSON, or an interactive HTML file viewable in a browser). The specification and hAMRonization are open-source and freely available at https://github.com/pha4ge/hAMRonization.

Sub-Grant hAMRonization Evaluations

A test of the PHA4GE AMR specification package within the context of the AMR sub-grants is being requested to better explore its utility in improving genomics-based AMR surveillance communication and data sharing among academic and public health groups. In this test, participants are asked to perform the following:

1. Install hAMRonization (see instructions below).
2. Run your dataset(s) of interest using the tool, and examine the tabular and html summary outputs which have been harmonized to the PHA4GE AMR data standard.
3. Answer the following questions below and submit your responses to Emma Griffiths (emma_griffiths@sfu.ca) and Finlay Maguire (finlay.maguire@dal.ca).

Optional Feedback questions (point form answers are preferred):

1. What is your role in analyzing AMR data (e.g. bioinformatician and/or data scientist, lab administrator, epidemiologist, lab technician, other)?
2. How easy was it to install the hAMRonization tool using the instructions on GitHub? If you had any difficulties, please briefly describe them. If someone else installed it, put Not Applicable.
3. Which fields in the harmonized output were the most useful for your work?
4. Were you able to find all the information you needed to interpret the results? If not, what kinds of information would help you to do this?
5. Did you prefer the tabular or html output summary, and why did that work best for you?
6. Did you find anything surprising or strange in the harmonized results? If yes, please briefly describe what you were expecting vs what you observed.

Instructions for Implementation

For full instructions, alternative installation options (e.g., Galaxy, docker, and conda), and details of all options available see the hAMRonization documentation (https://github.com/pha4ge/hAMRonization)

- *System requirements*: python installation (at least version 3.7)
- *Installation*: `python -m pip install hAMRonization`
- *Basic Usage:* `hamronize <tool_name> <tool_output_files> <required_metadata>`
- *Example*: `hamronize abricate abricate_genome1.tsv abricate_genome2.tsv --reference_database_version v3.1.1 --analysis_software_version v1.0.1 > abricate_hamronized_output.tsv`

The executable `hamronize` will be installed in your path and can be used on the output files of any of the 14 AMR gene detection tools listed to automatically apply the hAMRonization specification and generate a standardised report. Each tool has different required additional metadata that must be provided on the command-line (see the help message for each tool for details e.g., `hamronize abricate -h`)

A collated single report of hAMRonized outputs from multiple tools can be generated using `hamronize summarize`.

- *Example:* `hamronize summarize abricate_hamronized_output.tsv rgi_hamronized_output.tsv > combined_report.tsv`