

Running AMR bioinformatic tools using Google colab



**Public Health Alliance for
Genomic Epidemiology**

Antimicrobial Resistance Sub-Grant Team Malaysia

Dr. Su Datt Lam, Universiti Kebangsaan Malaysia

Dr. Sabrina Di Gregorio, University of Buenos Aires

Mr. Mia Yang Ang, University of Tokyo

Dr. Tengku Zetty Maztura Tengku Jamaluddin, Universiti Putra Malaysia

Prof. Dr. Sheila Nathan, Universiti Kebangsaan Malaysia

Dr. Hui-min Neoh, Universiti Kebangsaan Malaysia

Running AMR bioinformatic tools using Google colab

¿Why Google colab?

Google colab is free, requires no configuration and works entirely in the cloud (write, run and share code). It allows the creation of different workflows using different bioinformatic tools that can be shared and allow a reproducible analysis of the data.

These features will allow users in public health laboratories without bioinformatic knowledge and/or dedicated computers, to run on the cloud different bioinformatic software on bacterial WGS datasets to detect antimicrobial resistance (AMR).

AMRfinder plus

AMRfinder plus is a tool designed to identify AMR genes and some point mutations from assembled genome sequences or proteins.

For more details regarding the software, checkout the AMRfinder plus github page:
<https://github.com/ncbi/amr/wiki>

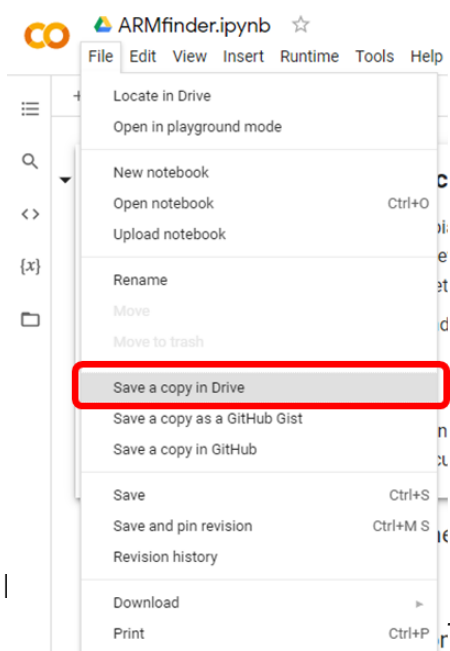
1) Access WGS data

Download assembled genome sequence/s from shared folder on drive.

2) Start Google colab for AMRfinder plus from the following link:

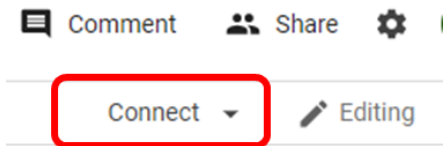
<https://bit.ly/3uFX6Vo>

a) On the upper left panel, click con “file” and “Save a copy in drive”.



button found on the upper right of the page to
re.


Running AMR bioinformatic tools using Google colab



After it's done you will see the allocated RAM and space on disk appear.

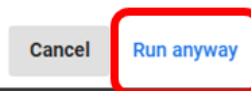


3) Run the pipeline

- a) The pipeline consists of 5 steps. To start running each step, click on the  icon found to the left.
- b) The first time you run google colab, the following message will appear, click on "Run anyway"

Warning: This notebook was not authored by Google.

This notebook was authored by sudatt.lam@gmail.com. It may request access to your data stored with Google, or read data and credentials from other sessions. Please review the source code before executing this notebook. Please contact the creator of this notebook at sudatt.lam@gmail.com with any additional questions.



- c) The currently running step is indicated by a circle with a stop sign next to it. 


When its finished a green tick will appear to the left 

- d) **STEP1: Install AMRFinderPlus and other prerequisites**

Running AMR bioinformatic tools using Google colab

Click on 

e) STEP2: Test AMRFinderPlus installation

Click on . Test is finished when a message similar to the following appears:

```
AMRFinder took 7 seconds to complete
```

f) STEP3: Upload your file

Click on . Then, the upload file button  will then appeared.

When the upload is finished (100% done) proceed with the next step.

g) STEP4: Provide your parameters and run AMRFinderPlus

Complete all boxes marked in red as appropriate.

filename:

- the name of the file you uploaded. Please include the extension name (for example, .txt).

sequence_type:

- n - translated-nucleotide
- p - protein sequence with no genomic coordinates

organism:

- organism.

output_filename:

- Filename of output file.


Click on .

When AMRFinderPlus its done it will display a message similar to this one:

Running AMR bioinformatic tools using Google colab

```
AMRFinder took 66 seconds to complete
```

h) SETP5: AMRFinderPlus Output

Click on  . The output file will be downloaded to your computer.

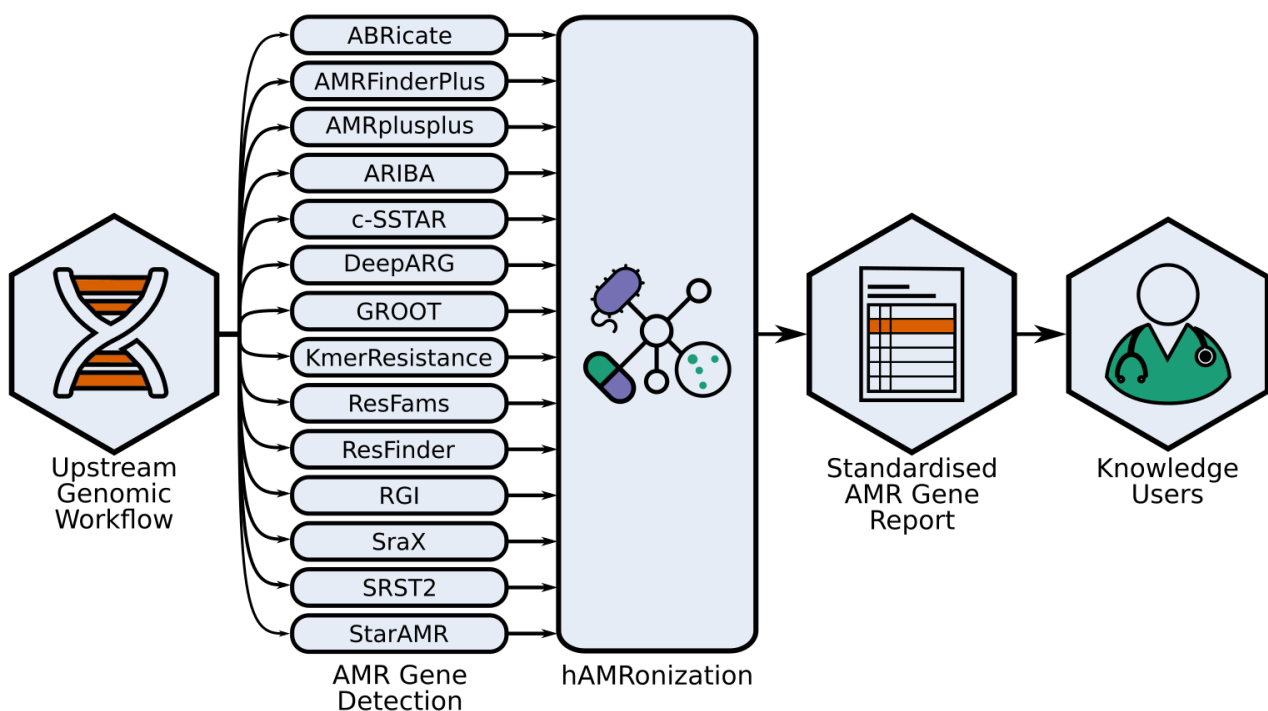
If you are having issues downloading the result archive, try disabling your ad blocker and run this cell again. If that fails click on the little folder icon to the left, navigate to file: output_filename, right-click and select "Download". The content of the output file will be displayed below.

Running AMR bioinformatic tools using Google colab

hAMRonzation

hAMRonzation is a software which converts AMR gene detection tool output (such as ariba, resfinder4, amrfinderplus) to a standardized AMR gene report (i.e. hAMRonzation specification format).

For more details regarding the software, checkout the hAMRonzation github page:
<https://github.com/pha4ge/hAMRonzation>



1) Access AMR output data

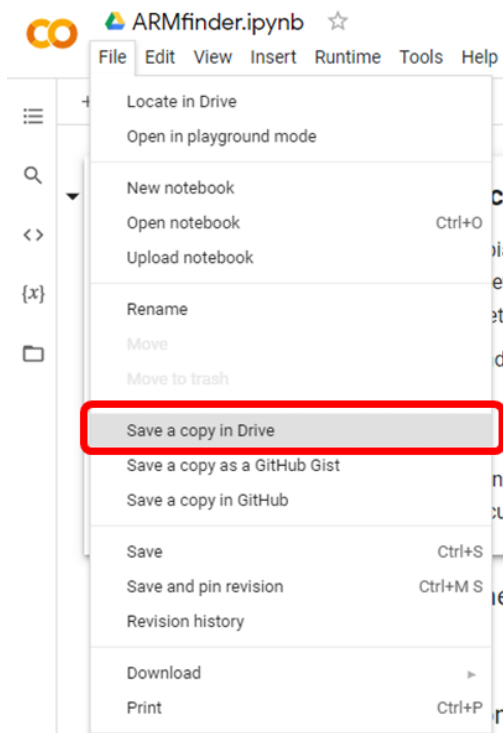
Download AMR output data from genome sequence/s from to your computer.

2) Start Google colab for hAMRonzation from the following link:

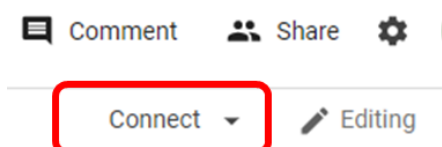
<https://bit.ly/3iOx10Q>

a) On the upper left panel, click con “file” and “Save a copy in drive”.

Running AMR bioinformatic tools using Google colab



- b) Please press the "Connect" button found on the upper right of the page to connect to a Google machine.



After it's done you will see the allocated RAM and space on disk appear.



3) Run the pipeline

- c) The pipeline consists of 5 steps. To start running each step, click on the icon found to the left.



Running AMR bioinformatic tools using Google colab

- d) The first time you run google colab, the following message will appear, click on "Run anyway"

Warning: This notebook was not authored by Google.

This notebook was authored by **sudatt.lam@gmail.com**. It may request access to your data stored with Google, or read data and credentials from other sessions. Please review the source code before executing this notebook. Please contact the creator of this notebook at **sudatt.lam@gmail.com** with any additional questions.

Cancel

Run anyway

- e) The currently running step is indicated by a circle with a stop sign next to it. 

When its finished a green tick will appear to the left 

f) STEP1: Install hAMRonization

Click on 

g) STEP2: Test hAMRonization installation

Click on  . Test is finished when a message similar to the following appears:

```
usage: hamronize.py abricate <options>

Applies hAMRonization specification to output(s) from abricate (OUTPUT.tsv)

positional arguments:
  report                Path to report(s)

optional arguments:
  -h, --help            show this help message and exit
  --format FORMAT       Output format (tsv or json)
  --output OUTPUT       Output location
  --analysis_software_version ANALYSIS_SOFTWARE_VERSION
                        Input string containing the analysis_software_version
                        for abricate
  --reference_database_version REFERENCE_DATABASE_VERSION
                        Input string containing the reference_database_version
                        for abricate
```


Running AMR bioinformatic tools using Google colab

h) STEP3: Upload your file

Click on  . Then, the upload file button  will then appeared.

When the upload is finished (100% done) proceed with the next step.

i) STEP4: Provide your parameters and run hAMRonization

Complete all boxes marked in red as appropriate.

filename:

- the name of the file you uploaded. Please include the extension name (for example, .txt).

program:

- program supported: abricate; amrfinderplus; ariba; rgi; resfinder; resfinder4; srax; deeparg; km

input_filename:

- Sequence entry

analysis_software_version:

- AMR software version.

reference_database_version:


- Database version.

format:

- tsv; json; interactive

output_filename:

- Filename of output file.


Click on  .

When AMRFinderPlus its done it will display a message similar to this one:

```
Harmonization complete for PPUKM-261-2009_amrfinderplus_harmoized.txt.
```

Running AMR bioinformatic tools using Google colab

j) SETP5: AMRFinderPlus Output

Click on  . The output file will be downloaded to your computer.

If you are having issues downloading the result archive, try disabling your ad blocker and run this cell again. If that fails click on the little folder icon to the left, navigate to file: `output_filename`, right-click and select "Download". The content of the output file will be displayed below.