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

Version 2.0

2023

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

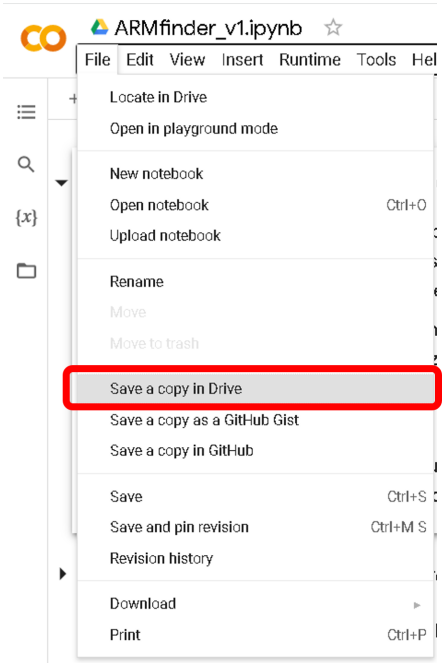
1) **Access WGS data**

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

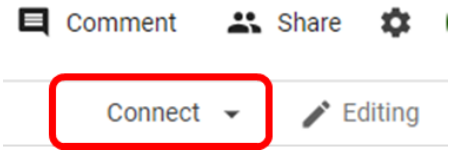
2) **Start Google colab for AMRfinder plus from the following link:**

   https://bit.ly/3mce0ee

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



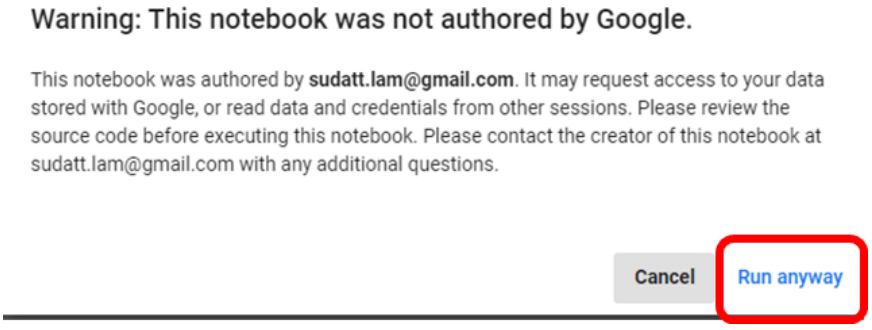
b) Please press the "Connect" button found on the upper right of the page to connect to a Google machine.



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

3) Run the pipeline

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 may appear, click on “Run anyway”



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

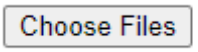
Click on the ▶ button on the left.

e) STEP2: Test AMRFinderPlus installation

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

```
AMRFinder took 8 seconds to complete
Success!
```

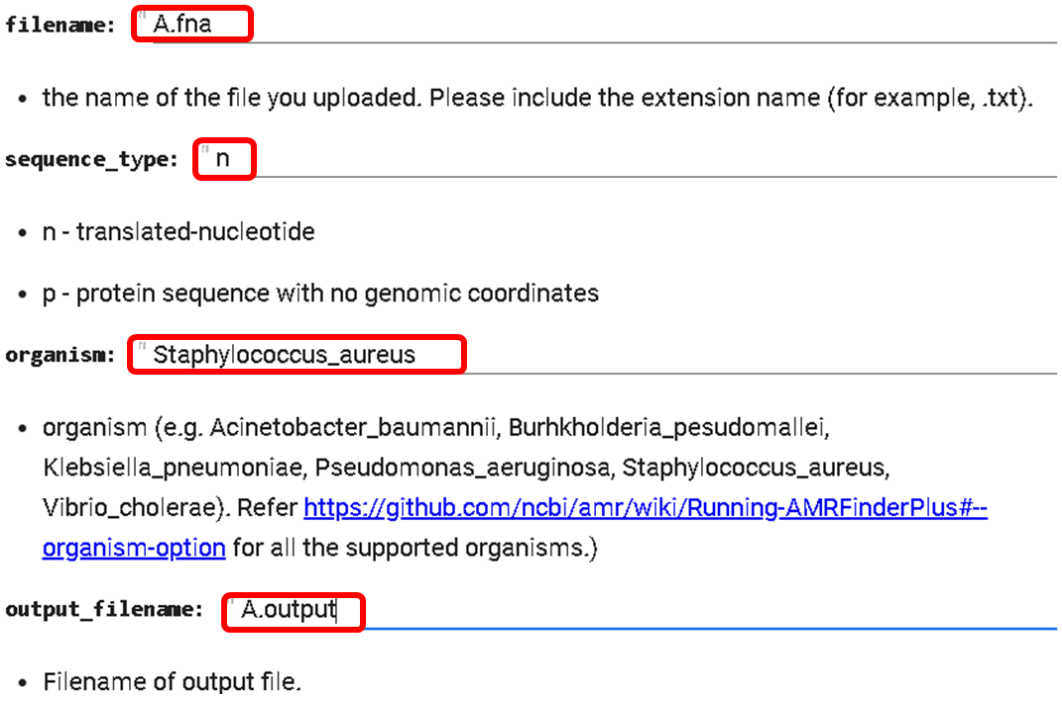
f) STEP3: Upload your file

Click on ▶ . Then, the upload file button  will appear.

Click on “Choose Files” and select your input file. 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.



Click on ▶ .

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

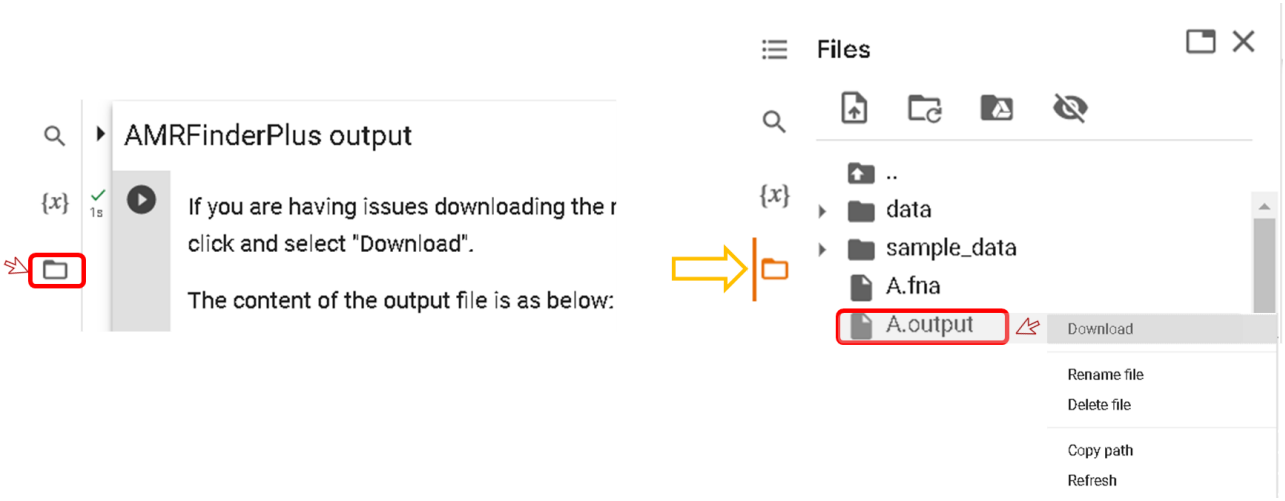
```
AMRFinder took 66 seconds to complete
```

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 output file: A.output



Right-click on the file and select "Download". The content of the output file will be displayed below.

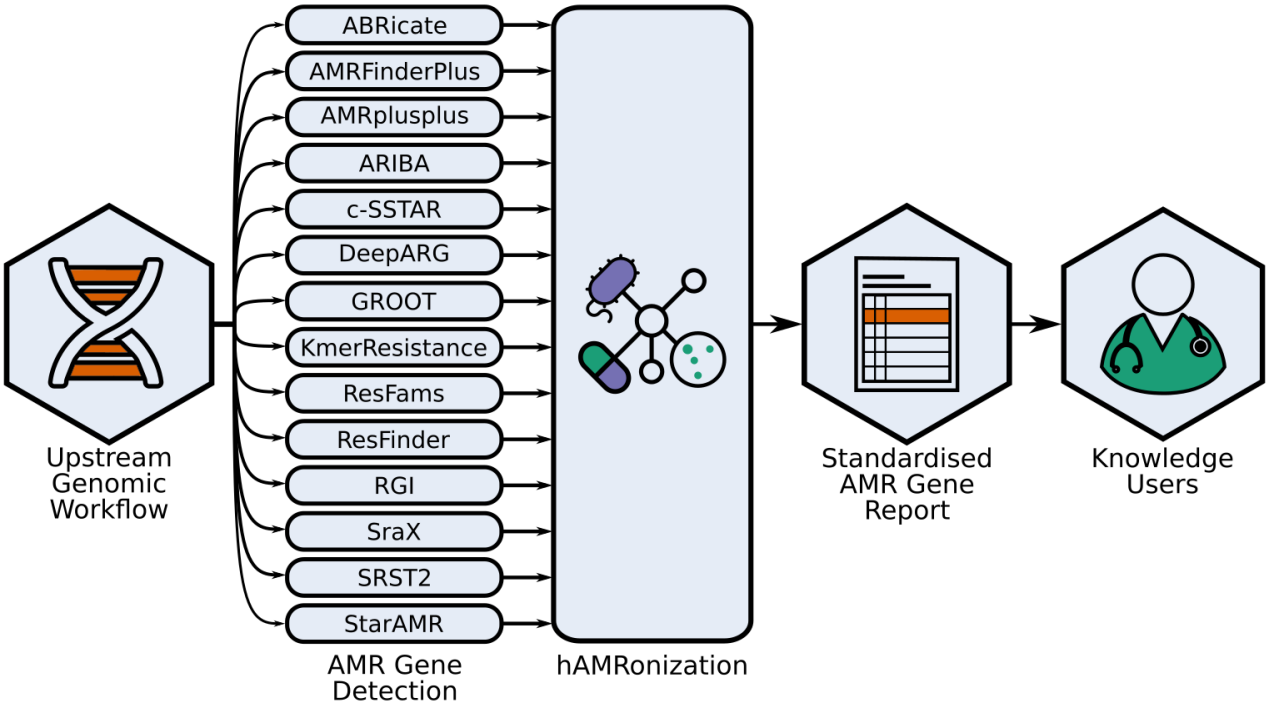
If you wish to run this tool in batch to multiple genomes, follow this link: https://bit.ly/41ewgSX and run the pipeline following the same steps as described above in this manual.

---

hAMRonization

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

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



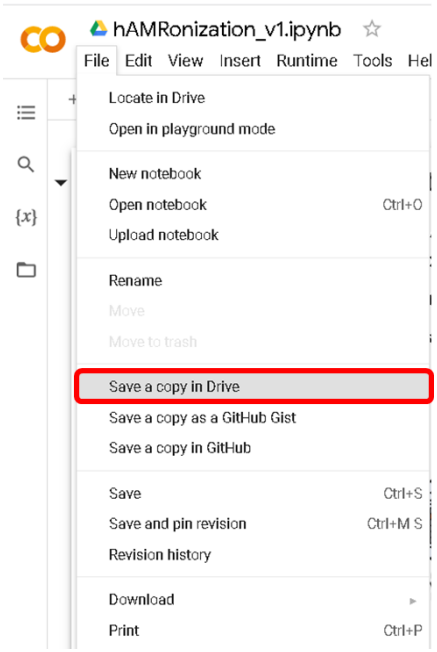
1) Access AMR output data

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

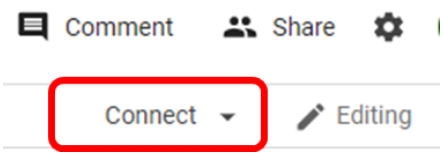
2) Start Google colab for hAMRonization from the following link:

https://bit.ly/3zyluLs

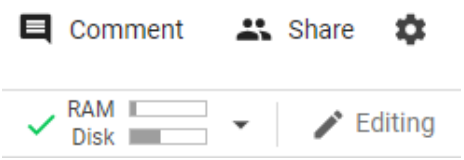
a) On the upper left panel, click con “file” and “Save a copy in drive”.



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



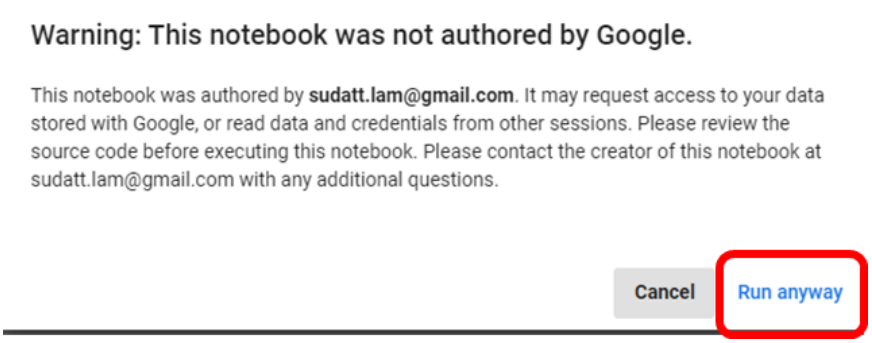
After it's done you will see the allocated RAM and space on disk appear.



3) Run the pipeline

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

d) The first time you run google colab, the following message may appear, click on "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 the hAMRonization software

Click on the ▶ button on the left.

g) STEP2: Test the hAMRonization installation

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

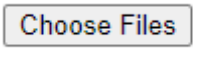
```
usage: hamronize.py abricate <options>

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
```

h) STEP3: Upload your file

Click on ▶ . Then, the upload file button 

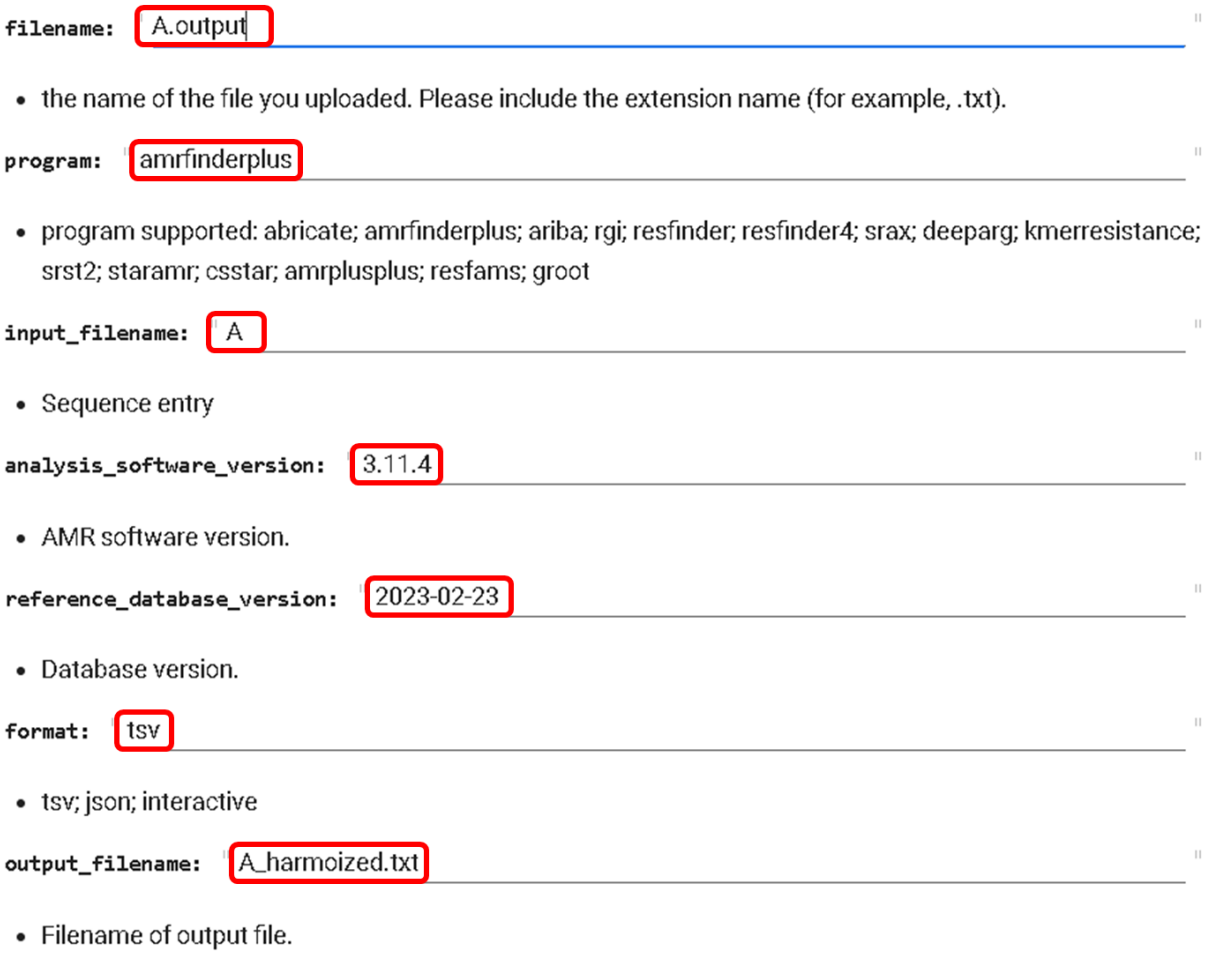
will appear.

Click on “Choose Files” and select your input file. When the upload is finished (100% done) proceed with the next step.

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.



Click on  .

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

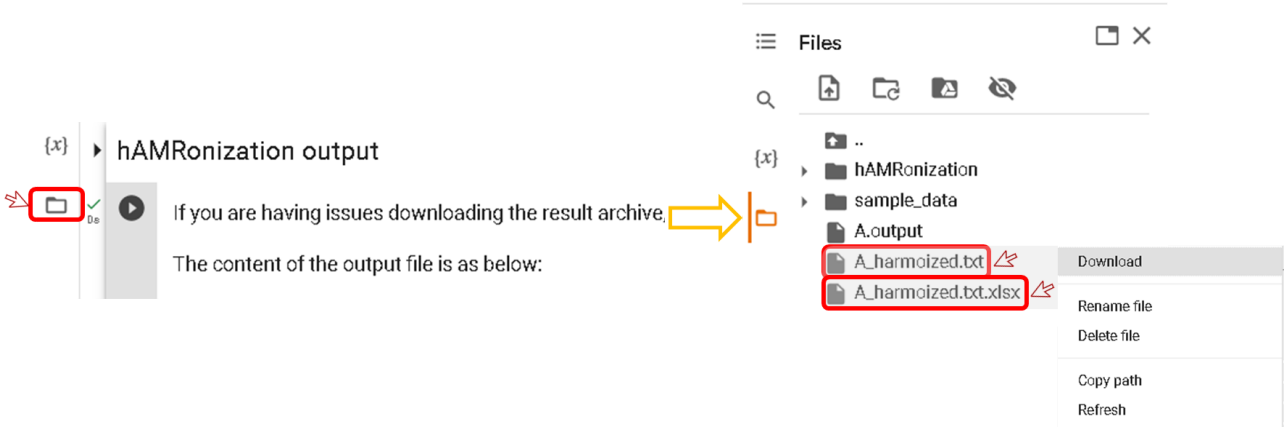
```
Harmonization complete for A_harmoized.txt.
```

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 the output filenames: A_harmonized.txt.xlsx and A_harmonized.txt.



Right-click and select "Download". The content of the output file will be displayed below.

To explore the functionality of hAMRonization software, we recommend to download both the Excel (.xlsx) and plain text (.txt) files. The Excel file can be used to easily explore and navigate the results. Meanwhile, the text file can be used for downstream analysis (if there is any).

If you wish to run this tool in batch to multiple hAMRonization reports, follow this link: https://bit.ly/3zOFtpm and run the pipeline following the same steps as described above in this manual.

Convert hAMRonization output to Microreact ready format

Microreact (https://microreact.org/) is a web-based application developed by the Centre for Genomic Pathogen Surveillance (CGPS) that allows you to upload, visualise and interactively explore any combination of clustering (trees), geographic (map), temporal (timeline) and other metadata variables (such as the presence of AMR genes and point mutations).

This google colab takes all the standardized AMR genes and point mutations reports (hAMRonization outputs, .txt or .xlsx files) and summarizes them into one single .csv file that can be used in microreact.

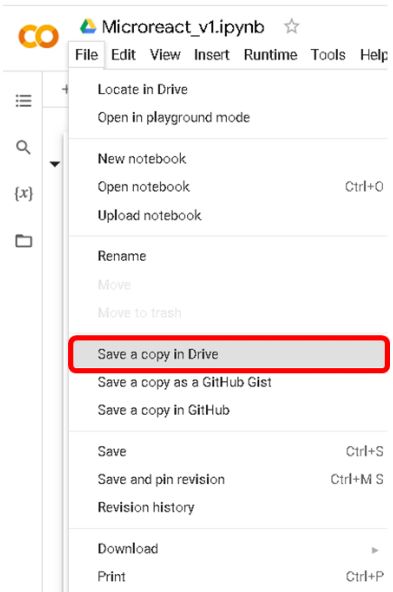
1) **Access hAMRonized data**

   Download all your hAMRonization outputs (.txt files) to your computer.

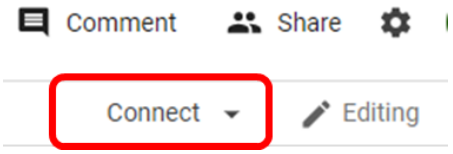
2) **Start Google colab from the following link:**

   https://bit.ly/3KDXjBW

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



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

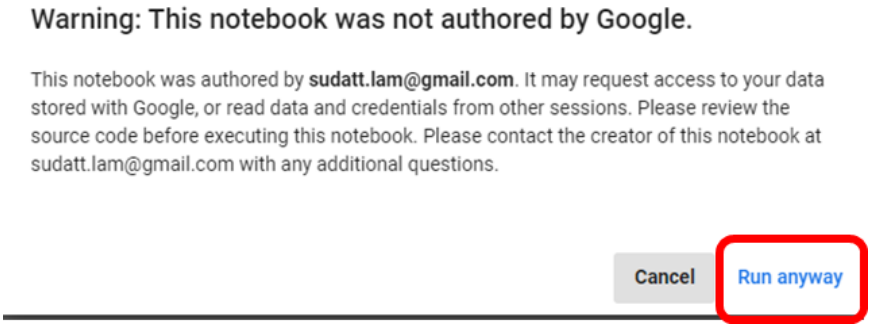


After it's done you will see the allocated RAM and space on disk appear.

3) Run the pipeline

c) The pipeline consists of 3 steps. To start running each step, click on the icon found to the left.

d) The first time you run google colab, the following message may appear, click on "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: Upload all your hAMRonization outputs

Click on the ▶ button on the left. Then, the upload file button [Choose Files] will appear.

Click on “Choose Files” and select ALL your input files (genomeA_hAMRonized.txt, genomeB_hAMRonized.txt, etc) at once .

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

g) STEP2: Convert hAMRonization output to microreact format (part 1)

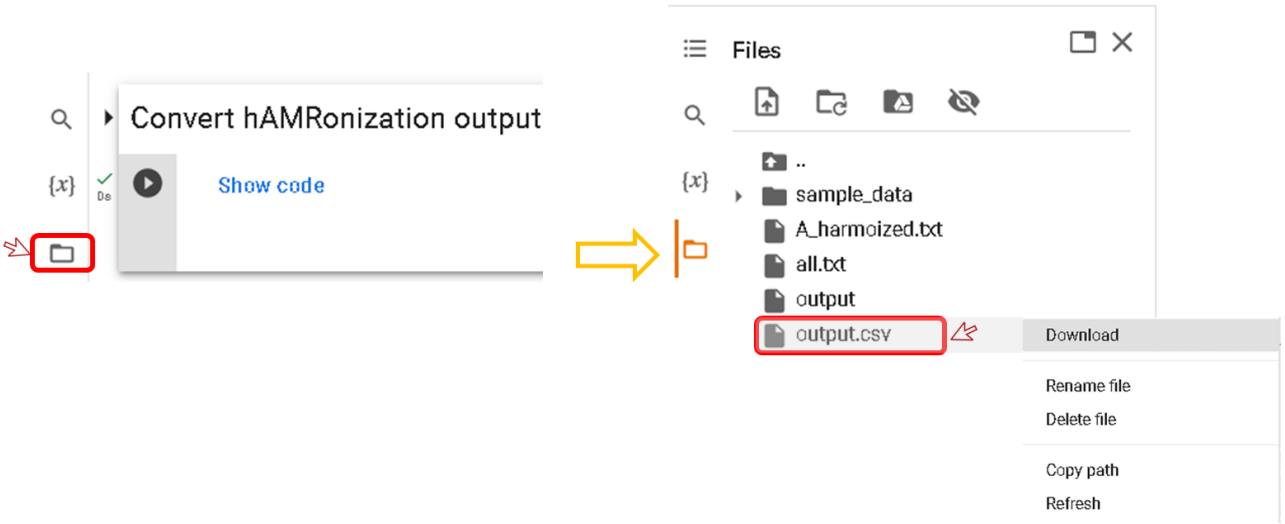
Click on ▶ .

h) STEP3: Convert hAMRonization output to microreact format (part 2)

Click on ▶ . The output file (output.csv) 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 the output filename: output.csv



Right-click and select "Download". The content of the output file will be displayed below.

4) Create a phylogenetic tree using Pathogenwatch

Pathogenwatch

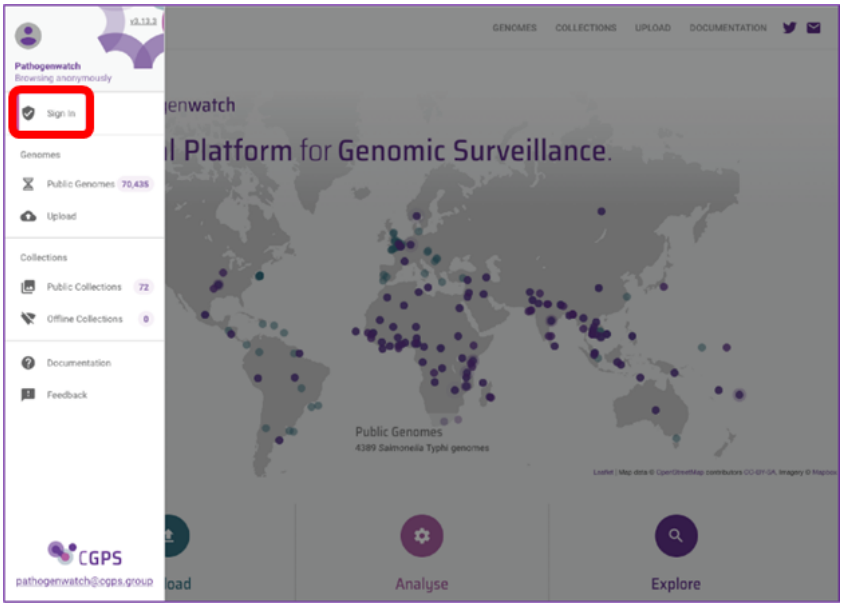
Pathogenwatch (https://pathogen.watch/) is a web-based application developed by the Centre for Genomic Pathogen Surveillance (CGPS) that allows you to upload, analyse and and compare genomes from a wide variety of microbial species.

The platform allows you to browse or download public sequence collections. In addition, each user can upload their own genomes privately (reads or *de novo* assemblies). The platform performs automatic analysis using pipelines and public or internal databases. The results of these analyzes can later be downloaded, or also viewed on the platform.

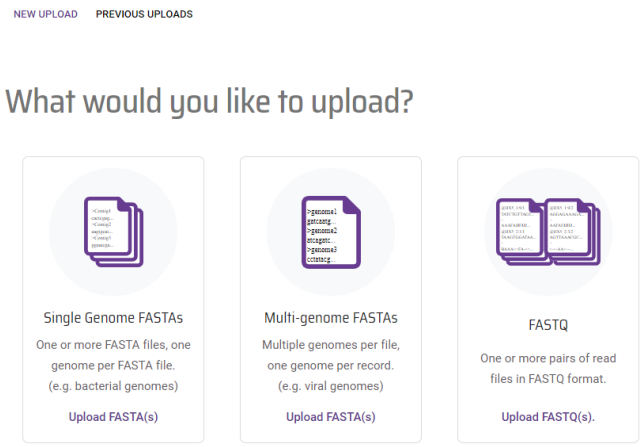
For a curated set of species (detailed in the website) Pathogenwatch provides the ability to construct a phylogenetic tree based on a core genome distance.

a) SETP1: Upload your genomes into Pathogenwatch

Enter https://pathogen.watch/. Click on the ☰ menu at the top left and on Sign In to create an account with your email address or social media account.



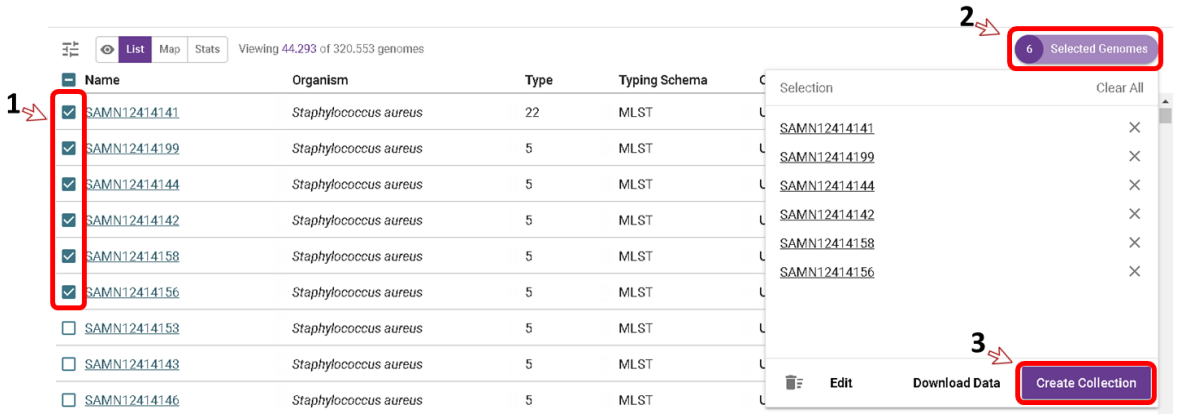
Go to Upload (https://pathogen.watch/upload) on the upper right corner. Select "Single Genome FASTAs" if your genomes are individual assembly files. Select "Multi-Genome FASTAs" if your genomes have been concatenated into a big file.



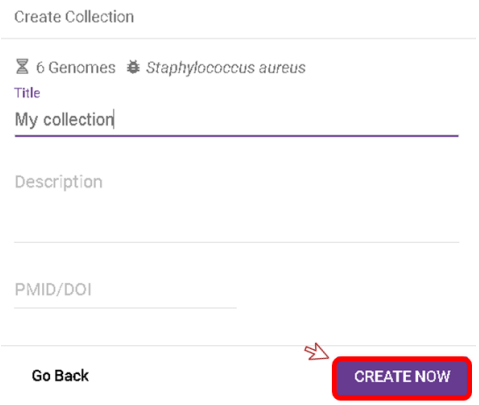
Upload the file: Drag and drop your genomes into the website. After the genomes have been uploaded, wait for a while while the system performs the analyses.

b) SETP2: Create a collection:

Select all the genomes uploaded and place them in the same collection. At the genome page, select all genomes interested by checking the respective box. After that, select "Selected Genomes" followed by "Create Collection".



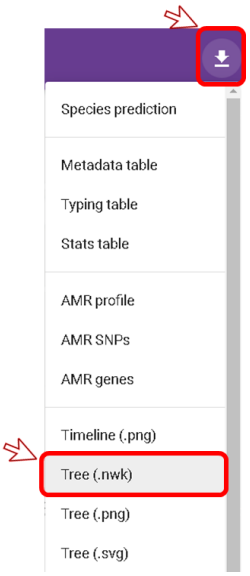
Add a name for the collection and click “Create now”.



c) SETP3: Open the collection and download the generated tree

Click on the collection created. The system will try to create a phylogenetic tree using all the sequences inside the collection. If a tree is not available, wait for a while until the system creates one. Once it is ready, it will be displayed in the system by clicking on “View Tree”.

Click the Download button on Top Right followed by "Tree (.nwk)".

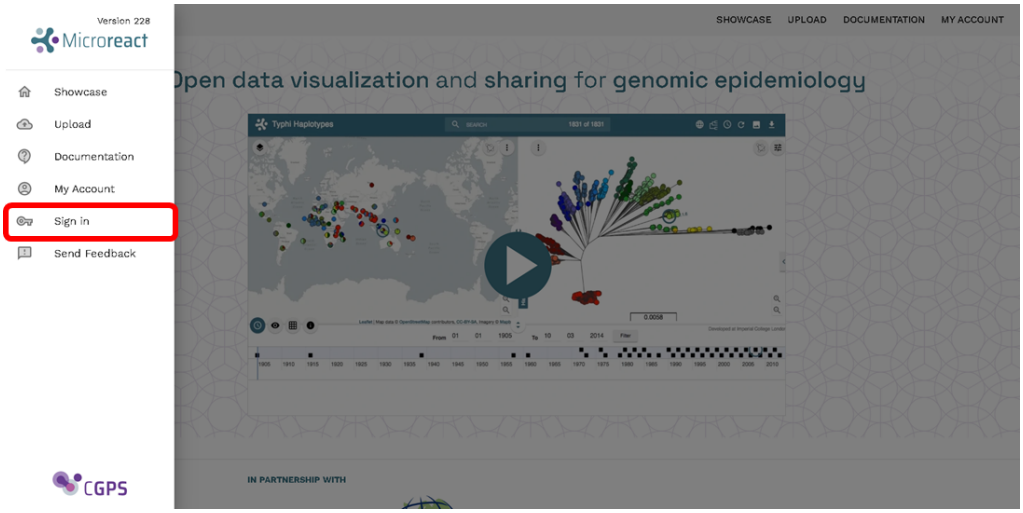


Congratulations you managed to generate a tree using Pathogenwatch!!

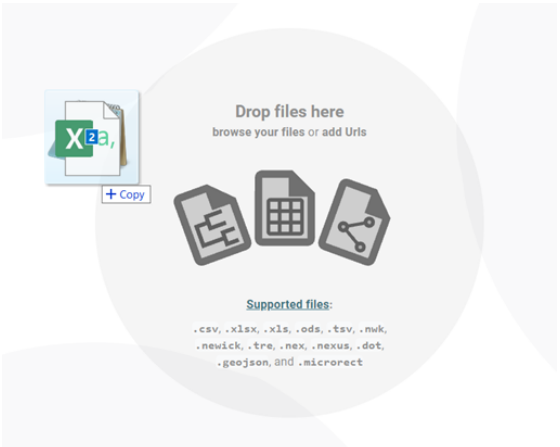
5) Display the tree and the hAMRonization output using Microreact

a) SETP1: Upload your files into Microreact

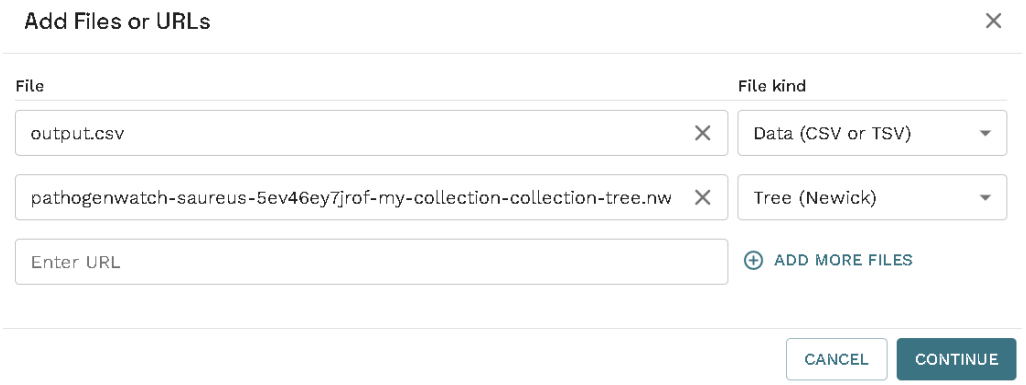
Log into Microreact website (https://microreact.org/). Click on the menu at the top left and on Sign In to create an account with your email address or social media account.



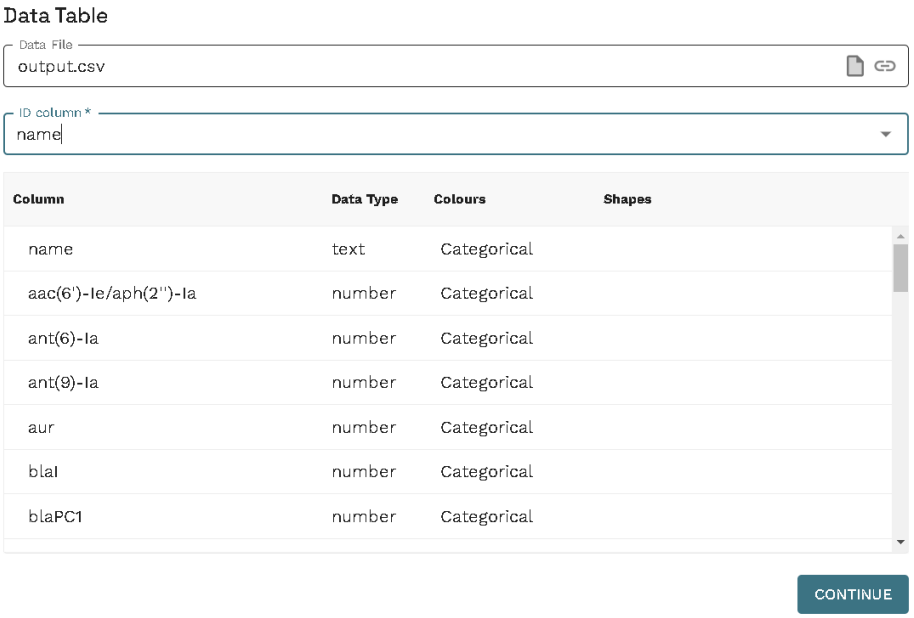
Go to Upload (https://pathogen.watch/upload) on the upper right corner. Upload the files: Drag and drop the microreact ready hAMRonization file (i.e. output.csv) and the phylogenetic tree (i.e. pathogenwatch-XXX-XXX-test-collection-tree.nwk).



Make sure the file type for microreact ready hAMRonization file is "Data" and phylogenetic tree is "Tree (Newick)".



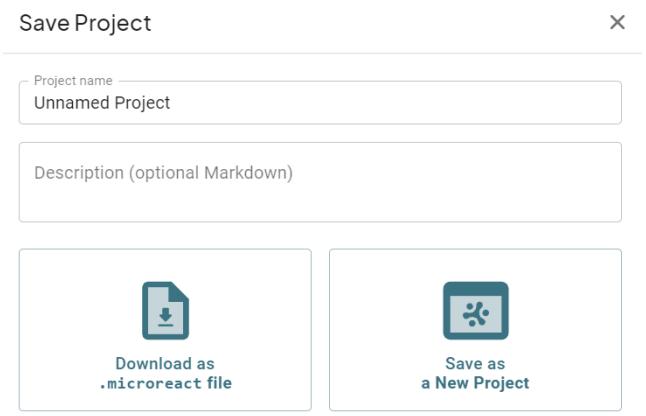
Make sure column "name" is the column label. Click on “continue” two times.



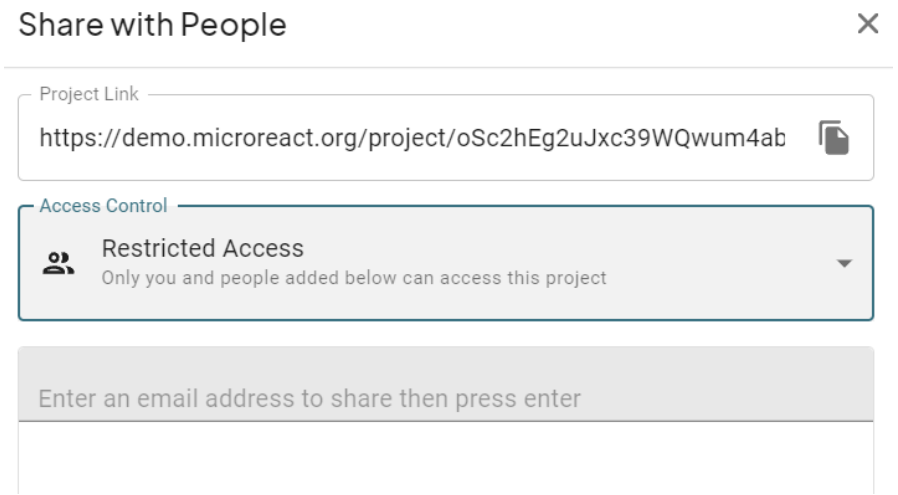
b) SETP2: Create a project

Projects are not saved on the upload page unless the user requests it. Save the project by clicking Save  .

Then, to name the project, write its name under “Project name”, and click on “Save as New Project”. In this way we will be saving the project on the microreact.org platform. Alternatively you can download the project to analyze it at another time.



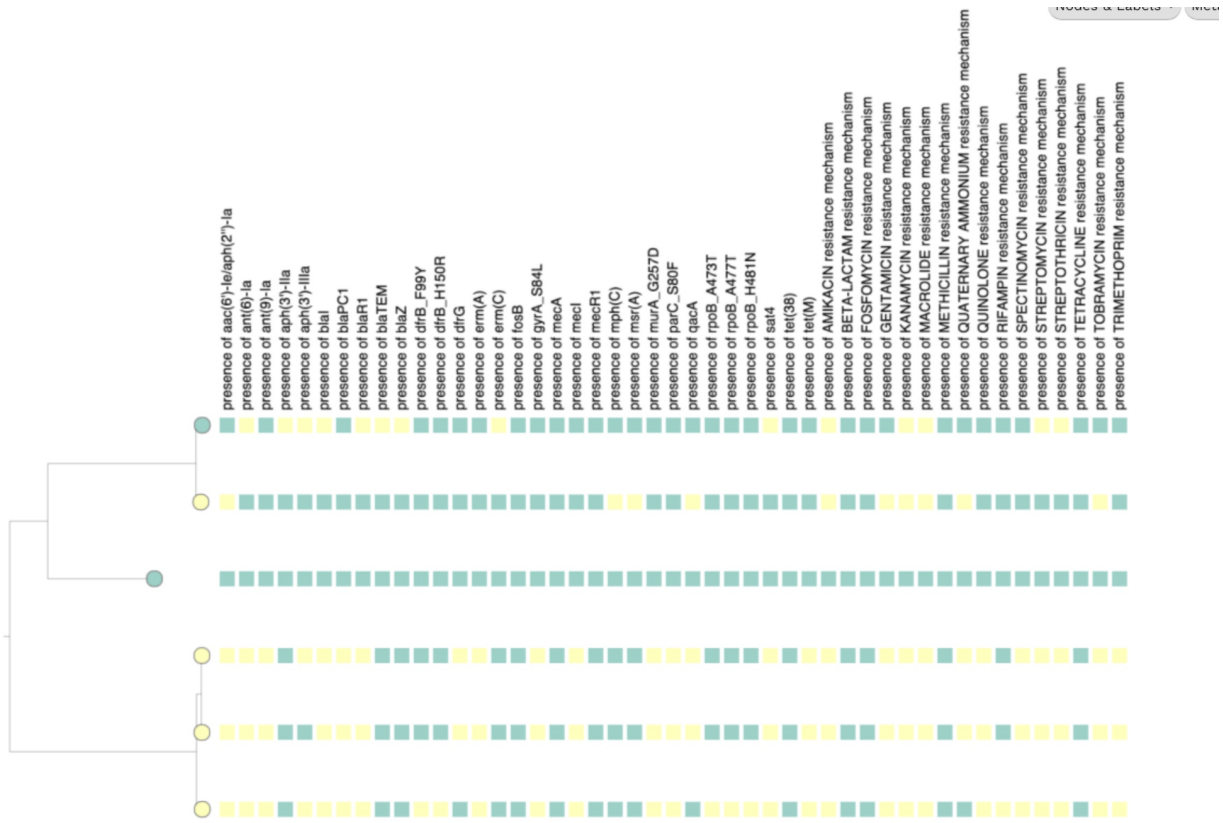
Next you will be able to obtain the url of the project, and choose if it is of restricted or public access using the drop-down menu for that purpose. Select “Restricted Access”. Close the window by clicking on the “X”.



c) SETP3: Analyse your data

In the tree window, click on the menu then on Metadata Blocks and activate Select all.

With this the phylogenetic tree will be displayed with the presence of AMR gene and also the presence of Antibiotic Resistance.



Feel free to include additional Metadata in the .csv file (e.g. latitude, longtitude, anatomical material, body product, year) to make the visualization even more informative.