Practical: Antimicrobial resistance genes and point mutations detection in Salmonella

## Overview

General

|  |  |  |
| --- | --- | --- |
| **Time** | **Activity description** | **ILOs** |
| 14.30-16.00 | Practical: “Antimicrobial resistance genes and point mutations detection in *Salmonella*”* ResFinder (incl. PointFinder)
* AMRFinderPlus
 | * Utilize command-line tools to identify AMR genes and point mutations (PMs).
* Explain why different bioinformatic tools may give different results.
* Discuss the difference between a tool, a database, and a method.
* Explain the difference between genotype and phenotype.
 |

Specific

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Time** | **Duration** | **Activity description** | **Duration** | **Activity** |
| 14:30-14:50 | 20 min | Introduction to AMR  | 20min |  Lecture #1 |
| 14:50-15:20 | 30 min | ResFinder + PointFinder | 5 min | Interactive lecture #1 |
| 15 min | Practice #1 |
| 10 min | Recap #1 |
| 15:20-15:50 | 30 min | AMRFinderPlus | 5 min | Interactive lecture #2 |
| 15 min | Practice #2 |
| 10 min | Recap #2 |
| 15:50-16:00 | 10min | Discussion | 10min |  Discussion #1 |

How to use this document

**Conda environment with required software**: CGEfinders and NCBItypers

**Path to working folder**: ~/BTG\_2024

**Path to assemblies**: ~/BTG\_2024/precomputed\_data/day7/amr

**Path to databases**: ~/BTG\_2024/data/databases

**Path to output**: ~/BTG\_2024/amr/output

**Path to precomputed output:** ~/BTG\_2024/precomputed\_data/day7/amr

Software commands are highlighted in grey. Arguments must be filled in manually (so don’t copy-paste the commands!); all arguments are indicated by square brackets. For example, if you need to use the following command:

bgzip -k [input vcf-file]

on a vcf-file named “exciting.vcf”, you would type the following in your terminal:

bgzip -k exciting.vcf

## ResFinder + PointFinder

“*ResFinder identifies acquired antimicrobial resistance genes in total or partial sequenced isolates of bacteria.*

Online server: <http://genepi.food.dtu.dk/resfinder>

Tool: <https://bitbucket.org/genomicepidemiology/resfinder/src/master/>

Databases:
<https://bitbucket.org/genomicepidemiology/resfinder_db/>

<https://bitbucket.org/genomicepidemiology/pointfinder_db/>

**Tutorial**

In this tutorial we will use ResFinder to identify AMR genes and point mutations in assembled contigs

To start, go to your working folder “/path/to/working/folder/”

cd [path\_to\_working\_folder]

Look for the environment that you will use in this session

mamba env list

Activate the environment “CGEfinders”

mamba activate [environment\_name]

To get started with ResFinder, you can use the help option and review the command options you get.

resfinder.py -h

python3 -m resfinder -h

But before running a sample through ResFinder, you will want to create folders to save your output. Use the following command with “/path/to/output/” to create a folder called “amr” and a folder called “output” inside it

mkdir -p [path\_to\_output]

Now, this is how you run a sample through ResFinder. Run the sample “**SRR27241771**”, use the path to databases /path/to/databases/” and remember the genes database name “resfinder\_db” and the point mutations database name “pointfinder\_db”

resfinder.py -o [path\_to\_output]/[sample1\_name] -l 0.6 -t 0.9 -acq -ifa [path\_to\_assembly\_from\_sample1] -db\_res [path\_to\_databases]/[database\_name] -c -db\_point [path\_to\_databases]/[database\_name] --species Salmonella

python3 -m resfinder -o [path\_to\_output]/[sample1\_name] -l 0.6 -t 0.9 -acq -ifa [path\_to\_assembly\_from\_sample1] -db\_res [path\_to\_databases]/[database\_name] -c -db\_point [path\_to\_databases]/[database\_name] --species Salmonella

See the results for sample “**SRR27241771**”

less -S [path\_to\_output]/[sample1\_name]/ResFinder\_results\_tab.txt

**Extra:**

Try to do the same for sample “**SRR27241772**”

Once you are done running both samples, you can concatenate the results to get a better overview.

awk 'FNR==1 {if (!header\_printed) {print "File:", $0; header\_printed=1}; next} {print FILENAME, $0}' [path\_to\_output]/ResFinder\_results\_tab.txt > [path\_to\_output]/resfinder\_all\_salm.txt

An important part of explaining why different tools give different results consists in doing database queries.

Write a query to find out how many antimicrobial resistance classes are in the resfinder database, remember the database\_name “resfinder\_db”, and the file extension “.fsa”

cd [path\_to\_databases]/[database\_name]

ls \*[file\_ext]

ls \*[file\_ext] | wc

Write a query to retrieve the total number of amr genes in the resfinder database, you need to use the filename “all.fsa”

cd [path\_to\_databases]/[database\_name]

grep '>' [filename] | wc

Write a query to retrieve the number of aminoglycoside genes in the resfinder database, you need to use the filename “aminoglycoside.fsa”

grep '>' [filename] | wc

Write a query to find how many genes have alternate names in resfinder database, you need to use the file “notes.txt”

grep 'Alternate name' [filename] | wc

## AMRFinderPlus

 “*This software and the accompanying database are designed to find acquired antimicrobial resistance genes and point mutations in protein and/or assembled nucleotide sequences.”*

Tool: <https://github.com/ncbi/amr>

Database: <https://ftp.ncbi.nlm.nih.gov/pathogen/Antimicrobial_resistance/AMRFinderPlus/database/3.11/2023-11-15.1/>

**Tutorial**

In this tutorial, we will use AMRFinderPlus to identify AMR genes and point mutations in assembled contigs

To start, go to your working folder “/path/to/working/folder”

cd [path\_to\_working\_folder]

Look for the environment that you will use in this session

mamba env list

Activate the environment “NCBI\_typers”

mamba activate [environment\_name]

To get started with AMRFinderPlus, you can use the help option and review the command options you get

amrfinder -h

Then, download the most up-to-date AMRFinderPlus database

 amrfinder -u

But before running a sample through AMRFinderPlus, you will want to create folders to save your output. Use the following command with “/path/to/output” to create a folder called “amr” and a folder called “output” inside it

mkdir -p [path\_to\_output]

Now this is how you run a sample through AMRFinderPlus. Run the sample “**SRR27241771**”

amrfinder --ident\_min 0.9 --coverage\_min 0.5 --organism Salmonella --nucleotide [path\_to\_assembly\_from\_sample1] --output [path\_to\_output]/[sample1\_name]/amrfinderplus\_output.tsv

See the results for sample “**SRR27241771**”

less -S [path\_to\_output]/[sample1\_name]/amrfinderplus\_output.tsv

**Extra:**

Try to do the same for sample “**SRR27241772**”

Once you are done running both samples, you can concatenate the results to a better overview, you can use the following command:

awk 'FNR==1 {if (!header\_printed) {print "File:", $0; header\_printed=1}; next} {print FILENAME, $0}' [output]/amrfinderplus\_output.tsv > [path\_to\_output]/amrfinderplus\_all\_salm.txt

An important part of explaining why different tools give different results consist in doing database queries.

Write a query to find out how many antimicrobial resistance classes are in the amrfinderplus database, you will need the filename “ReferenceGeneCatalog.txt”

cd [path\_to\_databases]/amrfinderplus/data/2023-11-15.1

wget <https://ftp.ncbi.nlm.nih.gov/pathogen/Antimicrobial_resistance/AMRFinderPlus/database/3.11/2023-11-15.1/ReferenceGeneCatalog.txt>

cut -f 6,7,8 [filename] | grep 'AMR'| sort -u

cut -f 6,7,8 [filename] | grep 'AMR' | cut -f 3 | sort -u | wc

Write a query to retrieve the antimicrobial resistance classes conferred by point mutations in the amrfinderplus database, you will need the filename “ReferenceGeneCatalog.txt”

cut -f 6,7,8 [filename] | grep 'AMR'| grep 'POINT' |sort -u

cut -f 6,7,8 [filename] | grep 'AMR'| grep 'POINT' |sort -u | wc

Write a query to retrieve the total number of amr genes in the amrfinderplus database, you will need the filename “AMR\_CDS”

grep '>' [filename]| wc

Write a query to retrieve the total number of amr proteins in the amrfinderplus database, you will need the filename “AMRProt”

grep '>' [filename] | wc

Write a query to find why the genename aac(6')-Iaa not reported by AMRFinderPlus?, you will need the filename “changes.txt”

grep "[genename]" [filename] -B 2