Tutorial

Antibiotic Resistance Analysis

July 9, 2021

Sample to Insight

QIAGEN Aarhus · Silkeborgvej 2 · Prismet · 8000 Aarhus C · Denmark
digitalinsights.qiagen.com · ts-bioinformatics@qiagen.com

Tutorial

Antibiotic Resistance Analysis

This tutorial will give an introduction to the analysis of metagenomics samples with respect to the presence and abundance of antimicrobial resistance (AMR) with the Drug Resistance Analysis tools and accompanying databases of CLC Microbial Genomics Module.

Prerequisites For this tutorial, you will need CLC Genomics Workbench 12.0 or later with CLC Microbial Genomics Module 4.8 installed. How to install modules and plugins is described here: http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Install.html.

Basic Resistance Mechanisms

According to the definition of WHO, Antimicrobial resistance (AMR) is the ability of a microorganism to stop an antimicrobial agent (such as antibiotics, antivirals and antimalarials) from working against it (https://www.who.int/antimicrobial-resistance/en/).

The NCBI - NIH (https://www.ncbi.nlm.nih.gov/m/pubmed/2025137) lists three fundamental mechanisms of antimicrobial resistance:

- Alteration of bacterial proteins that are antimicrobial targets (1)
- Enzymatic degradation of antibacterial drugs (2)
- Changes in membrane permeability to antibiotics (3)

Although resistance to an antimicrobial compound (or class) can be attained through different mechanisms and pathways, we can distinguish two major genetic adoption strategies of bacteria (Munita and Arias, 2016 and Dever and Dermody, 1991). These approaches are reflected in the different database types and detection algorithms available in CLC Microbial Genomics Module.

- Resistance conferred by mutations in gene(s) that are targeted by an antibiotic compound (corresponding to mechanism 1 above). Generally speaking, such mutations alter the structure of the protein (drug-target) in a way that inhibits the compound from binding specifically. Examples are mutations in bacterial DNA gyrases and topoisomerases targeted by fluoroquinolones or ribosomal mutations conferring resistance to macrolides. In order to detect the presence of resistance-conferring mutations versus susceptible alleles of the same gene(s) for a set of NGS reads, CLC Microbial Genomics Module offers the **Find Resistance with PointFinder** tool and corresponding database.
- Acquisition of resistance determinants through horizontal gene transfer (HGT) is the second major adoption strategy (corresponding to mechanisms 2 and 3 above). Common examples are beta-lactamases which break down penicillins and cephalosporins by hydrolysis, or efflux pumps which actively expel antibiotic compounds from the bacteria and hence keeping the concentrations at levels tolerable by the microbes. Here, the presence of a gene or members of a gene-family are sufficient indicators of resistance. The **Find Resistance with Nucleotide DB** tool can be used to identify AMR genes in a set of assembled contigs and the **Find Resistance with ShortBRED** tool can be used to find AMR gene families in a set of NGS reads.

Introduction to QMI-AR

The QIAGEN Microbial Insight - Antimicrobial Resistance database (QMI-AR) addresses a number of challenges faced by users of Antibiotic Resistance Gene Data Resources (see Xavier et al., 2016 for an in-depth overview). Successful AR analysis relies on the quality of both bioinformatics analysis software and reference data. On the software-side, CLC Microbial Genomics Module enables AMR-analyses using different databases while striving to provide consistency between the query tools and reports. On the data-side, the QMI-AR database provides integrated up-to-date reference information from a number of different resources, namely ARGAnnot [Gupta et al., 2014], NCBI (https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/), ResFinder [Zankari et al., 2012] and CARD [Jia et al., 2017], making it probably the most comprehensive resource of antibiotic resistance currently available (figure 1).

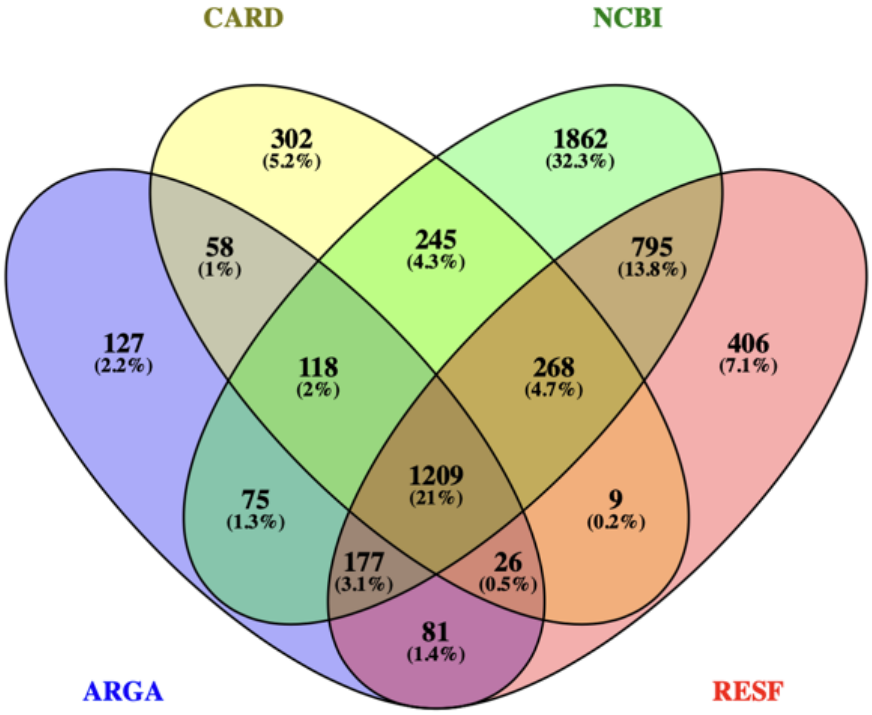


Figure 1: *Illustration of the overlap between some of the databases available within QMI-AR.*

As illustrated in the figure above, each database covers different aspects of antimicrobial resistance. Only about 1/5 of the sequences are common to all of them. While the Comprehensive Antibiotic Resistance Database CARD contains data with high quality ontology annotations, the NCBI database currently has the most comprehensive collection of AMR related genes despite lacking systematic semantic annotation. Similarly, but to a lesser extend, ResFinder (RESF) and ARG-Annot (ARGA) contain AMR genes that are not available in any of the other resources.

Not only does the integration of the aforementioned resources enable the use of the same algorithms for an analysis which is consistent for all the genes of the different resources, it also opens up for systematically applying consistent annotations to the genes using the Antibiotic Resistance Ontology (http://www.obofoundry.org/ontology/aro.html). This ontology describes antibiotic resistance genes and mutations, their products, mechanisms, and associated phenotypes, as well as antibiotics and their molecular targets.

Annotating the genes with terms from an open and curated ontology helps both with knowledge sharing and gene consolidation across resources. For example, the sulfonamide-resistance conferring dihydropteroate-synthase gene is denoted as 'sul3', 'SUL3', 'Sul-3' and even with roman numbering as 'SulIII' in different databases. By assigning the label 'ARO:3000413' it becomes unambiguous which gene is actually meant. Also, the public CARD knowledge database contains valuable additional information such as definitions, descriptions and links to the primary literature accessible for each of the ARO identifiers. While the ontology is constantly updated [Tsang et al., 2019], the assigned ARO-identifiers should remain stable and serve as a link to lookup the latest information via the Ontology Lookup Service provided by EMBL-EBI at https://www.ebi.ac.uk/ols/ontologies/aro. For example looking up the entry ARO:3000413 we get the description "Sul3 is a sulfonamide resistant dihydropteroate synthase (...)" and obtain a link to the original publication http://www.ncbi.nlm.nih.gov/pubmed/12604565. The nucleotide version of QMI-AR contains ontology annotations down to the gene level, and thus enables the precise identification of resistance genes while linking to the public ARO database entries.

The ontology can furthermore be used to identify the family of a gene, e.g., sul3 (ARO:3000413) is straightforwardly identified to be a member of the sulfonamide resistant sul family (ARO:3004238). This is important when the type of resistance is of interest rather than identifying the exact gene that confers resistance. In CLC Microbial Genomics Module, the Find Resistance with ShortBRED tool has been designed to rapidly detect peptide markers specific to families (or groups) of genes directly from the NGS reads. Using the CARD ontology, (ShortBRED) peptide markers have been created for the genes in QMI-AR and annotated with the ARO identifier which captures the most precise information about the family (or group) for which the marker is specific.

Downloading and importing the tutorial data

This tutorial illustrates the analysis of antimicrobial resistance (AMR) determinants in both a metagenomic sample and several clinical isolates. In this section, we describe the download of the Resistance Databases and NGS tutorial data sets before analysis.

Downloading Resistance Databases Find the Download Resistance Database tool in the Toolbox here:

> **Databases** ( ) | **Drug Resistance Analysis** ( ) | **Download Resistance Database** ( )

1. For the purpose of this tutorial and to save storage space, we only download the nucleotide and ShortBRED marker versions of the QMI-AR, and the PointFinder point-mutation database for the detection of single-point mutation that confer resistance (figure 2).
2. After clicking **Next**, you will need to agree to the license terms for each of the Databases.
3. Finally **Save** the results into a new dedicated folder you can name 'AMR databases'.

After the download is completed, you should see in the Navigation Area both the QMI-AR peptide Marker and Nucleotide Databases, as well as the PointFinder databases for a number of organisms (figure 3). We only need the Salmonella PointFinder database for this tutorial, so if you are pressed for storage space you can delete the other PointFinder databases.

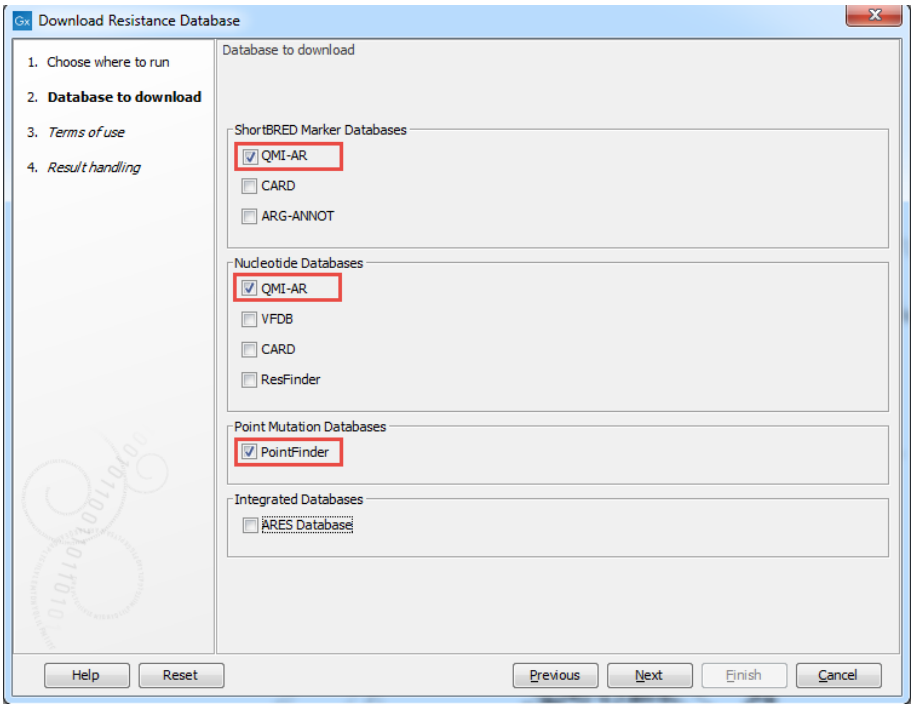


Figure 2: *Download Resistance Database with only the databases selected which are required for this tutorial.*

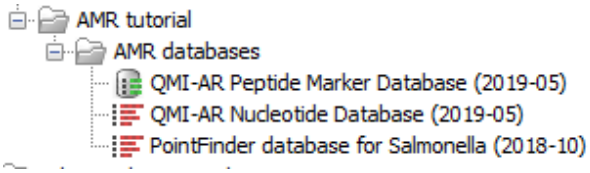


Figure 3: *Result folder of the Download Resistance Database tool containing only the databases required for this tutorial.*

Download of the tutorial datasets For demonstrating the range of possible applications, the data for this tutorial comes from two different studies, a complex environmental metagenome and clinical isolates.

1. Click on the following link (or paste it into your web browser) to download the tutorial data: `http://resources.qiagenbioinformatics.com/testdata/AMR_data.zip`.
2. Use the Import tool here:

   File | **Import (**  **)** | **Standard Import (**  **)**
3. Choose the file `AMR_data` you just downloaded on your computer. Ensure the import type under **Options** is set to **Automatic import** (figure 4) and click **Next**.
4. Select the location where you want to store the imported sequences (a folder you can call 'AMR data') and click **Finish**.

Afterwards, you should have a directory in the Navigation area called 'AMR data' with the following two sub-directories (figure 5):

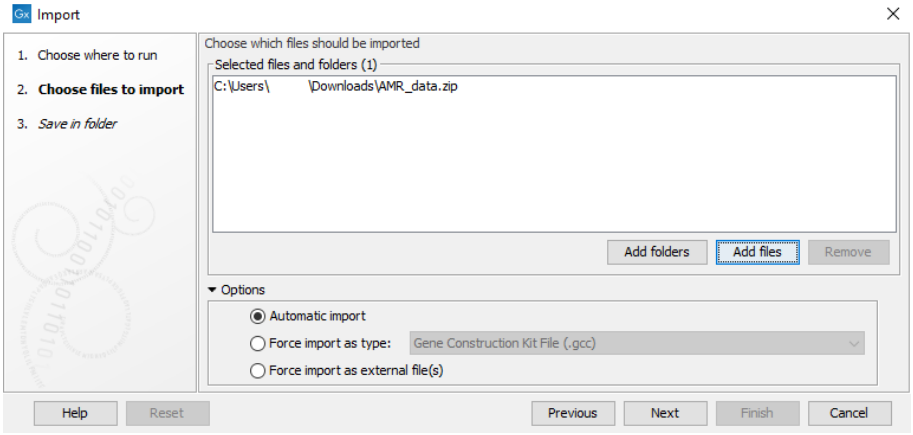


Figure 4: *Automatic Standard Import of a zip file.*

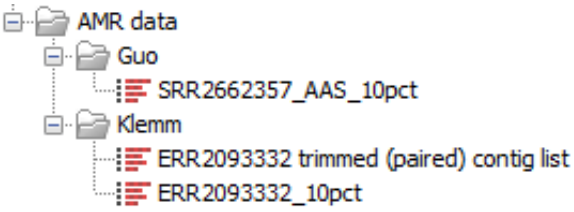


Figure 5: *Data after import.*

- The folder **Guo** contains data from the study by Guo et al., 2017. The samples present a broad-spectrum of antibiotic resistance genes obtained from a waste-water treatment plant. The file `SRR2662357_AAS_10pct` contains a 10% down-sampled fractions of the aerobic sample SRR2662357 and has approx. 4.78 million paired reads obtained on the Illumina HiSeq 2000 platform. We will use the Find Resistance with ShortBRED tool to get an overview of the broad antimicrobial classes and particular resistance genes present.
- The folder **Klemm** includes multidrug-resistant (MDR) isolates of Salmonella typhi strains from an outbreak in Pakistan [Klemm et al., 2018]. While the entire study contains whole genome sequencing of over 80 samples, we use data from run ERR2093332 down-sampled at 10% to approximately 200.000 paired reads each obtained from Illumina HiSeq 2500. We provide `ERR2093332 trimmed (paired) contig list`, a contig-list assembled from the full ERR2093332 set of reads for finding resistance genes with a nucleotide database and `ERR2093332_10pct` the down-sampled set of reads to find point-mutations.

Find Resistance with ShortBRED: a metagenomic sample

For performing AMR profiling in shotgun metagenomic sequencing data we will be using the pre-computed set of peptide Markers derived from QMI-AR.

1. Start the Find Resistance with ShortBRED tool from the Toolbox:

   Drug Resistance Analysis ( ) | **Find Resistance with ShortBRED** ( )

2. Specify the file from the Guo folder as seen in figure 6 and click **Next**.
3. Provide the QMI-AR Marker Database as reference (figure 7) and leave the other parameters as they are set by default before clicking **Next**.

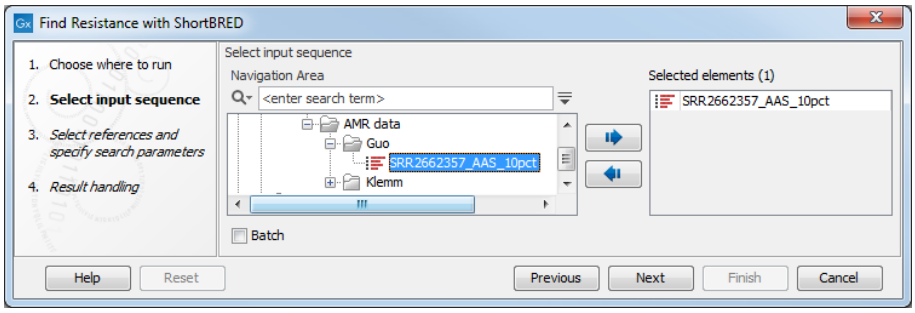


Figure 6: *Specify the input from Find Resistance with ShortBRED.*

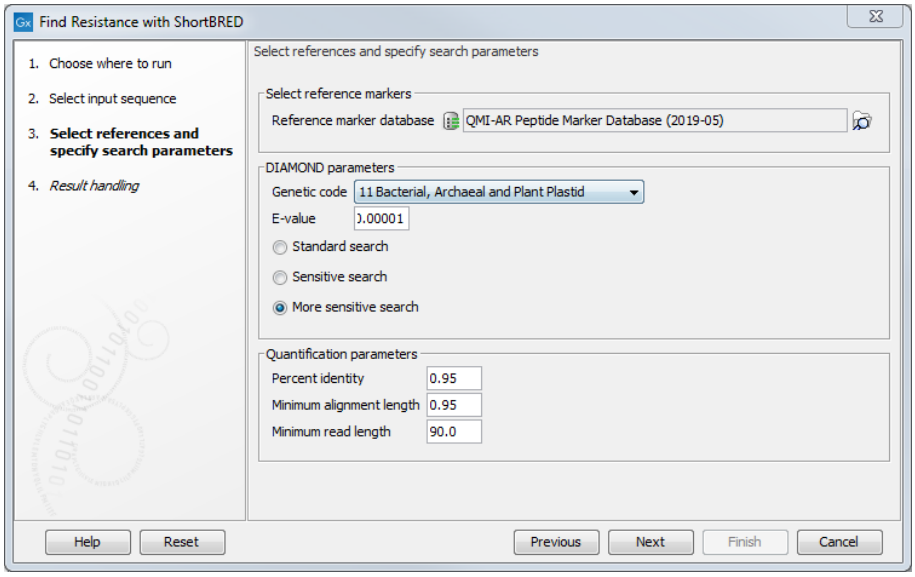


Figure 7: *Specify the reference from Find Resistance with ShortBRED.*

4. Choose to **Save** the results to a new subfolder called 'AAS ShortBRED'.

The ShortBRED report gives a brief overview of the input data and a summary of the AR profiles and genes found. The ShortBRED table offers an aggregated summary at the phenotype level. You can sort the table by Number of reads by clicking on the header of column - this provides an overview which resistances are present, how many genes were detected, and how many markers were hit (figure 8).

Note that the table also contains rows with zero counts to indicate Antimicrobial resistance profiles that were investigated but for which no indication of resistance has been found.

Open the ShortBRED abundance table, and switch it to the Sunburst graphical overview (figure 9) by clicking on the Sunburst icon ( ) at the bottom of the View.

Clicking on a peripheral section will zoom into that part, clicking into the center will zoom out to the higher level. Figure 9 shows that the 25% of hits to 'antibiotic target replacement' is comprised of hits to sul1 and sul2.

Even more details can be found in the (ShortBRED) Sequence list of all the reads with their classification. The reads are linked up with additional information derived from the ontology such as phenotype, confer-resistance-to links to ARO and additional pubmed-ids for detailed inspection of the results.

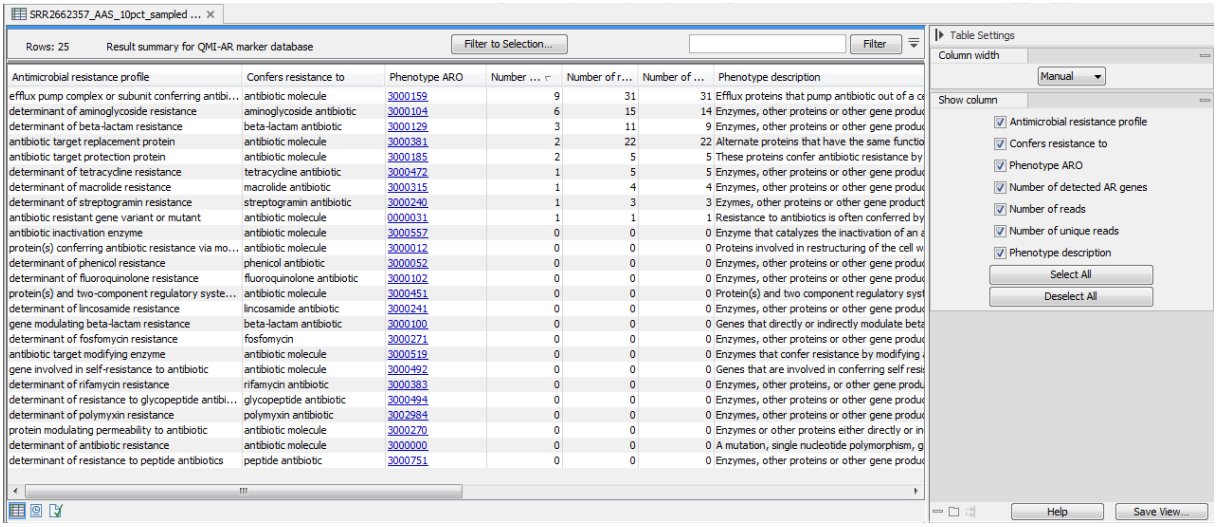


Figure 8: *The table showing the result summary for QMI-AR marker database.*

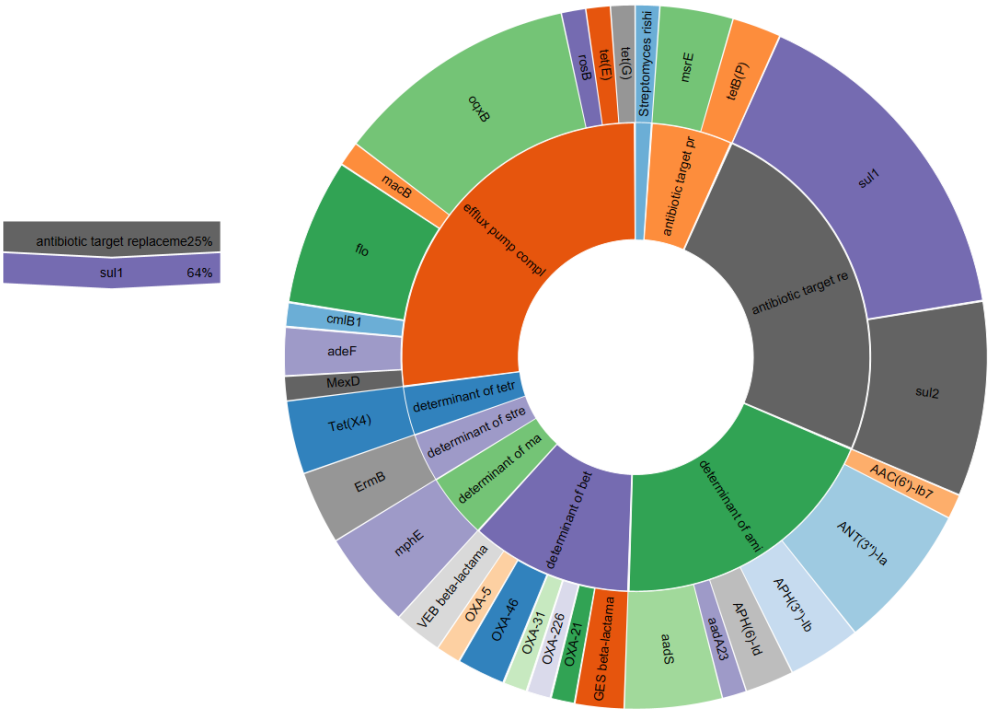


Figure 9: *ShortBRED abundance table seen as a zoomable Sunburst.*

Find Resistance with Nucleotide DB

The QMI-AR nucleotide-database contains the AR gene information which is used for sequence searching of whole genes in a set of assembled contigs. For this tutorial, we provide the assembled contig list from sample ERR2093332.

1. Start the Find Resistance with Nucleotide DB tool from the Toolbox:

   Drug Resistance Analysis ( ) | **Find Resistance with Nucleotide DB** ( )

2. Specify the file from the Klemm folder as seen in figure 10 and click **Next**.

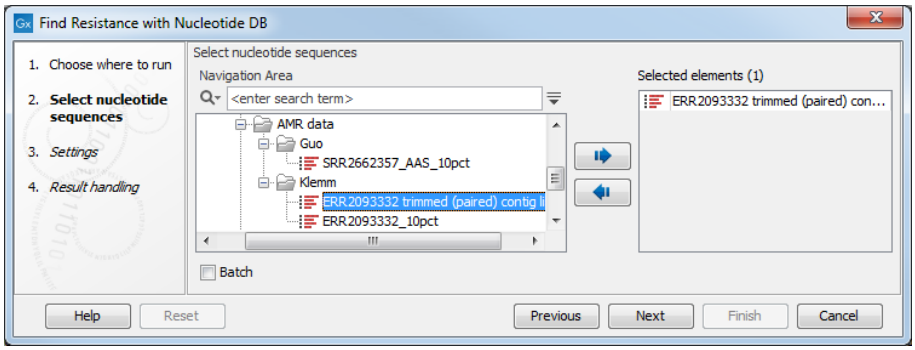


Figure 10: *Specify the input from Find Resistance with Nucleotide DB.*

3. Then provide the QMI-AR Nucleotide Database (figure 11) and leave the other parameters as they are set by default before clicking **Next**.

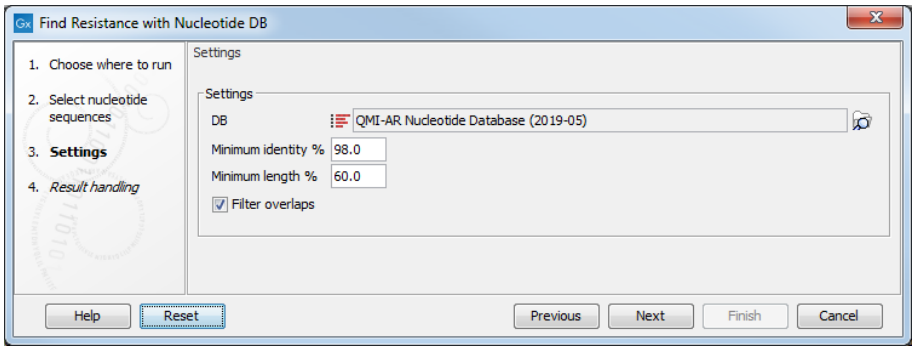


Figure 11: *Specify the reference from Find Resistance with Nucleotide DB.*

4. Save into a new sub-folder you can call 'ResFinder'

When the tool has completed, open the Find Resistance table, and de-select several columns such that we only see the Gene, %Identity, Predicted Phenotype and Gene ARO (figure 12).

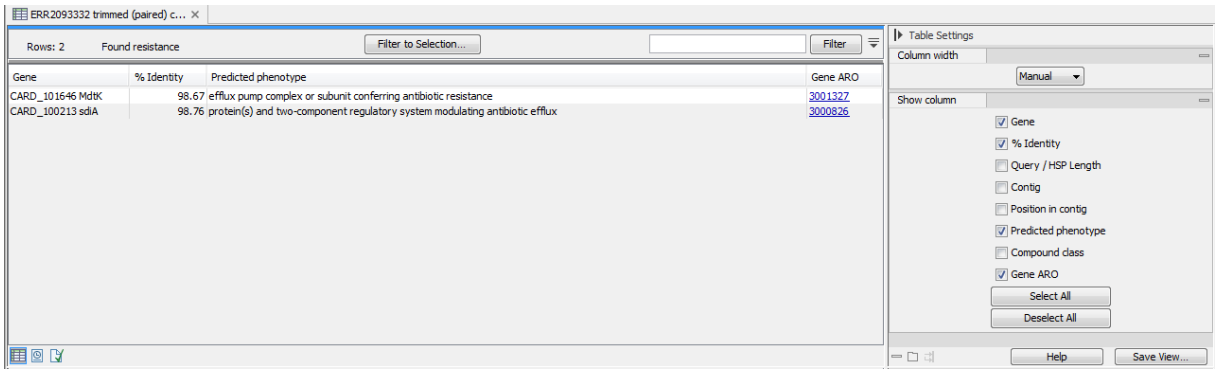


Figure 12: *The Find Resistance table.*

The table shows two genes which both (directly and indirectly) relate to the category of efflux pump complexes: MdtK is a 'multidrug and toxic compound extrusions (MATE) transporter' conferring resistance to several **fluoroquinolone** antibiotics such as ciprofloxacin, norfloxacin, doxorubicin and acriflavine; SdiA is a cell division regulator. Clicking on the Gene ARO link reveals that SdiA is also 'a positive regulator of AcrAB when expressed from a plasmid'. Similarly, the Gene ARO links

describes AcrAB-TolC as 'a tripartite RND efflux system that confers resistance to tetracycline, **chloramphenicol**, **ampicillin**, nalidixic acid, and rifampin'.

As stated in the abstract from the paper Klemm et al., 2018: 'Reduced susceptibility to fluoroquinolones is also widespread, and sporadic cases of resistance to third-generation cephalosporins or azithromycin have also been reported. (...) Here, we report the first large-scale emergence of resistance to three first-line drugs (chloramphenicol, ampicillin, and trimethoprim-sulfamethoxazole) as well as fluoroquinolones and third-generation cephalosporins'. In this single isolate, we could already identify several of the drugs (highlighted in bold above) from the study.

Find Resistance-Conferring Point Mutations

The Find Resistance with PointFinder tool finds known antimicrobial resistance conferring mutations in NGS reads (from isolates or shotgun metagenome samples after binning) and summarizes them in a concise and easily understandable report. In contrast to the Find Resistance with Nucleotide DB tool that identifies the occurrence of entire resistance conferring genes, the aim here is to detect the presence of resistance conferring mutations in antibiotic targets.

1. Start the Find Resistance with Nucleotide DB tool from the Toolbox:

   Drug Resistance Analysis ( ) | **Find Resistance with PointFinder** ( )

2. Specify the file from the Klemm folder as seen in figure 13 and click **Next**.

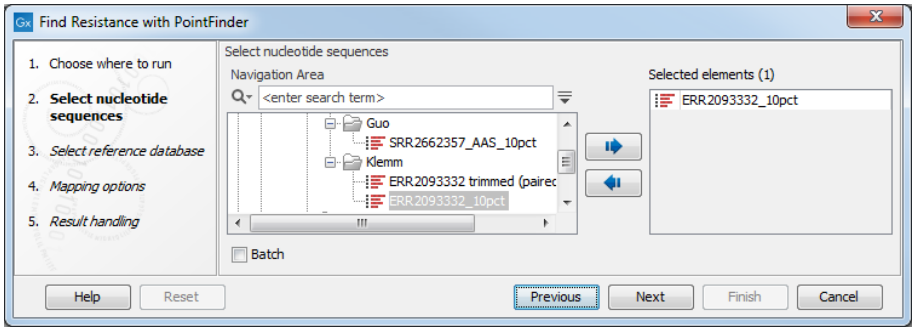


Figure 13: *Specify the input from Find Resistance with PointFinder.*

3. Then provide the Pointfinder database for Salmonella (figure 14) and click **Next**.

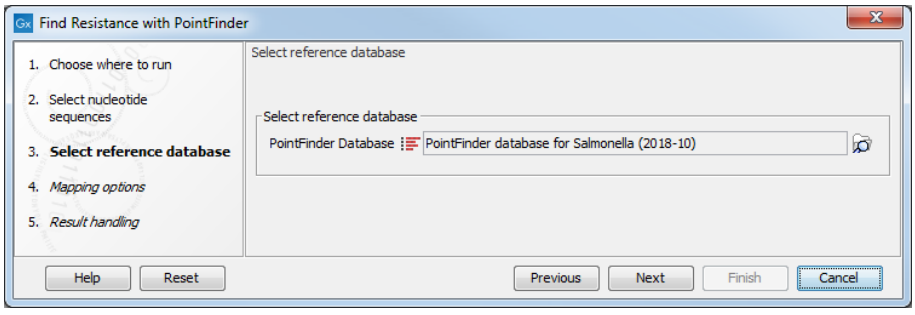


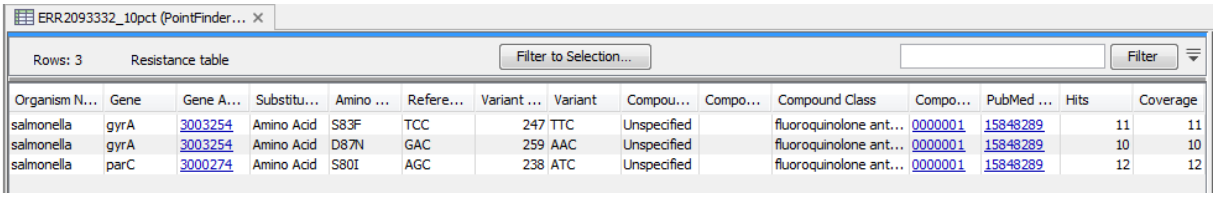
Figure 14: *Specify the reference from Find Resistance with PointFinder.*

4. Leave the Mapping options as they are set by default and click **Next**.

5. Save all possible outputs into a new subfolder you can call 'Pointfinder'.

The report gives a high-level overview of the number of sequences in the database and the reference database that was used. It also shows the number of hits and which compound / class they are corresponding to - in this case that would be the two fluoroquinolones ciprofloxacin and nalidixic acid. The 'Output annotated reads' provides the most level of detail in the form of an annotated sequence list, detailing precisely the reads that match an entry of the database.

Open the PointFinder resistance table to find detailed information on the Genes and the identified aminoacid-substitutions, the links to the Gene ARO, details of the substitution, reference, description of the variant (figure 15).



Organism N...	Gene	Gene A...	Substitu...	Amino ...	Refere...	Variant ...	Variant	Compou...	Compo...	Compound Class	Compo...	PubMed ...	Hits	Coverage
salmonella	gyrA	3003254	Amino Acid	S83F	TCC	247	TTC	Unspecified		fluoroquinolone ant...	0000001	15848289	11	11
salmonella	gyrA	3003254	Amino Acid	D87N	GAC	259	AAC	Unspecified		fluoroquinolone ant...	0000001	15848289	10	10
salmonella	parC	3000274	Amino Acid	S80I	AGC	238	ATC	Unspecified		fluoroquinolone ant...	0000001	15848289	12	12

Figure 15: *The PointFinder resistance table.*

By using Find Resistance with PointFinder we have now identified three point-mutations which are very much in agreement with the findings by Klemm et al., 2018: The concurrent substitutions in gyrA (S83F and D87N) and parC (S80I) are known to increase the minimum inhibitory concentration (MIC) for fluoroquinolones (see also Parry et al., 2010).

References

[Dever and Dermody, 1991] Dever, L. and Dermody, T. (1991). Mechanisms of bacterial resistance to antibiotics. *Arch Intern Med*, 151(5):886–895. doi:10.1001/archinte.1991.00400050040010.

[Guo et al., 2017] Guo, J., Li, J., Chen, H., Bond, P., and Yuan, Z. (2017). Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Research*. doi: 10.1016/j.watres.2017.07.002.

[Gupta et al., 2014] Gupta, S., Padmanabhan, B., Diene, S., Lopez-Rojas, R., Kempf, M., Landraud, L., and Rolain, J.-M. (2014). Arg-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrobial Agents and Chemotherapy*, 58(1):212–220. doi:10.1128/AAC.01310-13.

[Jia et al., 2017] Jia, B., Raphenya, A., Alcock, B., Waglechner, N., Guo, P., Tsang, K., Lago, B., Dave, B., Pereira, S., Sharma, A., Doshi, S., Courtot, M., Lo, R., Williams, L., Frye, J., Elsayegh, T., Sardar, D., Westman, E., Pawlowski, A., Johnson, T., Brinkman, F., Wright, G., and McArthur, A. (2017). Card 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res*, 4(45):566–573. doi: 10.1093/nar/gkw1004.

[Klemm et al., 2018] Klemm, E., Shakoor, S., Page, A., Qamar, F. N., Judge, K., Saeed, D., Wong, V., Dallman, T., Nair, S., Baker, S., Shaheen, G., Qureshi, S., Yousafzai, M., Saleem, M., Hasan, Z., Dougan, G., and Hasan, R. (2018). Emergence of an extensively drug-resistant salmonella enterica serovar typhi clone harboring a promiscuous plasmid encoding resistance to fluoroquinolones and third-generation cephalosporins. *mBio*, 9(1):1–10. doi: 10.1128/mBio.00105-18.

[Munita and Arias, 2016] Munita, J. and Arias, C. (2016). Mechanisms of antibiotic resistance. *Microbiology Spectrum*, 4(2). doi:10.1128/microbiolspec.VMBF-0016-2015.

[Parry et al., 2010] Parry, C., Thuy, C., Dongol, S., Karkey, A., Vinh, H., Chinh, N., Duy, P., Thieu Nga, T., Campbell, J., Minh Hoang, N., Arjyal, A., Bhutta, Z., Bhattacharya, S., Agtini, M., Dong, B., Canh, D., Naheed, A., Wain, J., Hien, T., Basnyat, B., Ochiai, L., Clemens, J., Farrar, J., Dolecek, C., and S., B. (2010). Suitable disk antimicrobial susceptibility breakpoints defining salmonella enterica serovar typhi isolates with reduced susceptibility to fluoroquinolones. *Antimicrobial Agents and Chemotherapy*, 54(12):5201–5208. doi 10.1128/AAC.00963-10.

[Tsang et al., 2019] Tsang, K., Speicher, D., and McArthur, A. (2019). Pathogen taxonomy updates at the comprehensive antibiotic resistance database: Implications for molecular epidemiology. *Preprints*. doi:10.20944/preprints201907.0222.v1.

[Xavier et al., 2016] Xavier, B., Das, A., Cochrane, G., De Ganck, S., Kumar-Singh, S.and Mï¿½ller Aarestrup, F., Goossens, H., and Malhotra-Kumar, S. (2016). Consolidating and exploring antibiotic resistance gene data resources. *Journal of Clinical Microbiology*, 54(4):851--859. doi: 10.1128/JCM.02717-15.

[Zankari et al., 2012] Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F., and Voldby Larsen, M. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial*, 67(11):2640–2644. doi: 10.1093/jac/dks261.