Forschungszentrum Borstel
Leibniz Lungenzentrum

ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

Tutorial on genomic analysis of non-tuberculous mycobacteria

Analysis of a skin infection outbreak of *M. abscessus*

Trainer
Dr. ir. Margo Diricks (Research Center Borstel, Germany)

mdiricks@fz-borstel.de

Date:
05.05.2026

Contents

1 Background 3
1.1 Introduction to *Mycobacterium abscessus* 3
1.2 Case presentation 4
1.3 Tasks 4
2 Gather required data 5
3 Bioinformatic analysis 6
3.1 Web-based and desktop GUI Tools 6
3.1.1 Species identification and typing using Pathogenwatch 6
3.1.2 Typing using PubMLST 8
3.1.3 Species identification and resistance prediction with NTM-Profiler (web-server) 15
3.1.4 CgMLST-based cluster analysis using Bruker MBioSEQ Ridom Typer 17
3.2 Command-line Tools 29
3.2.1 NTM-Profiler 29
3.2.2 NTMseq 29
4 References 30

1 Background

1.1 Introduction to *Mycobacterium abscessus*

In this tutorial we will investigate a real outbreak of skin infections caused by *Mycobacterium abscessus* which was described in the publication "Investigation of Two Mycobacterium abscessus Outbreaks in Quebec Using Whole Genome Sequencing" by Wuzinski and coworkers[1].

M. abscessus is a rapidly growing non-tuberculous mycobacterium with high intrinsic resistance. It is mainly affecting individuals with pre-existing lung diseases such as cystic fibrosis, bronchiectasis or previous TB. It can also cause extra-pulmonary disease such as skin and soft tissue infections via wounds, surgery or cosmetic procedures.

Phylogenetic analyses have shown that *M. abscessus* can be divided into three subspecies: subsp. *abscessus*, *massiliense* and *bolletii*. In addition to these subspecies, several clusters of closely related isolates have been identified within both *subsp. abscessus* and *subsp. massiliense.* These so called dominant circulating clones (DCC) have been isolated from both CF and non-CF patients across the whole globe and have been associated with increased virulence, higher rates of resistance, and worse clinical outcomes compared to unclustered isolates. Seven such DCCs have been described by Bryant and coworkers but DCC1, DCC2 and DCC3 are most commonly isolated from patients[2–4].

At the Research Center Borstel, we have developed a core genome multi-locus sequence typing scheme (cgMLST) that can be used for population structure analysis as well as transmission and outbreak analysis of *M. abscessus* strains (Figure 1)[5].

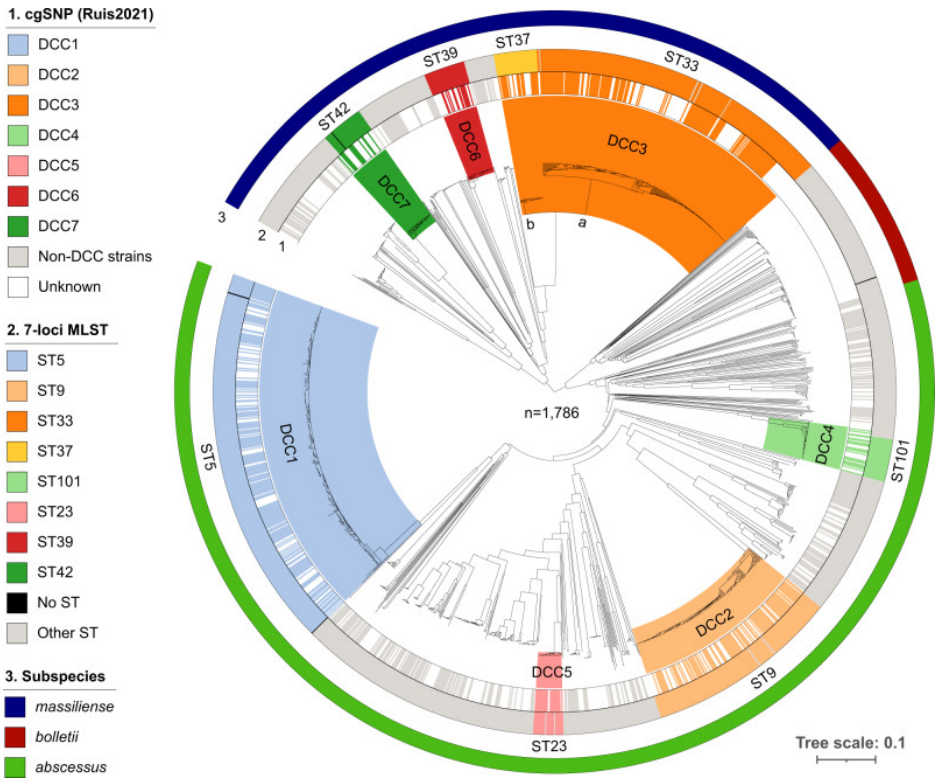


Figure 1. Core genome multi-locus sequence typing (cgMLST)-based neigbor-joining tree comprising 1,786 M. abscessus strains. 1. Strains that were previously classified into one of seven dominant circulating clones (DCC) based on core genome single nucleotide polymorphism (cgSNP)/FastBAPS analysis. DCC clades (inner circle) are delineated and colored based on the most recent common ancestral node comprising all strains with known DCC classification. 2. Sequence types (ST) according

to the nomenclature of the 7-loci MLST scheme hosted on pubMLST. 3. Mash distance-based subspecies classification. (Diricks et al. 2022)[5].

The cgMLST scheme consists of 2,904 core loci (core genes) representing 59% of the gene set of *M. abscessus* type strain ATCC19977. These core genes are present in nearly all *M. abscessus* strains, and comparing their allelic profiles allows determination of the genetic relatedness between strains. Using this scheme, we confirmed and amended the nomenclature of the main dominant circulating clones previously proposed by Bryant and coworker and found that these also correlate well with traditional 7-loci multi-locus sequence typing (MLST)[5]. Moreover, we found that DCC strains could be linked to a corresponding DCC reference genome with less than 250 alleles while 99% of pairwise genetic comparisons between epidemiologically linked (epi-linked) isolates were below 25 alleles and 90% below 10 alleles (i.e. most epi-linked isolates had a pairwise distance of <10 alleles). Epi-linked isolates are isolates with a known epidemiological link, such as a shared host (e.g. isolates from a chronically infected patient) or a common transmission pathway, exposure, or outbreak context (e.g. isolates from individuals exposed to the same contaminated water source). Clustering of isolates using thresholds of 25 or 10 alleles could therefore be used to guide further epidemiological investigations, as isolates differing by fewer than 25 alleles (and more stringently fewer than 10 alleles) may be epidemiologically linked. In general, smaller allelic distances indicate a higher likelihood of recent transmission or a common source.

Note: An allele is a specific gene variant, represented by an allele number according to a fixed nomenclature (i.e. each unique sequence is assigned a specific number). CgMLST analysis compares the allele profiles of core genes (i.e. the numbers for each of these genes) and calculates pairwise genetic distances. These distances (e.g. 10 alleles) refer to the number of core genes that differ between two isolates, providing a measure of their genetic relatedness: the higher the number, the more distantly related the isolates are.

1.2 Case presentation

Within a short time period of 6 weeks, three females and one male patient presented to the doctor with similar skin lesions around their new tattoo. Biopsies were taken and tested positive for mycobacteria. The isolates were identified as *M. abscessus subsp. abscessus* by MLST and PCR sequencing of the 16S and hsp65 genes. Patient interviews revealed a common exposure at a specific tattoo studio. The Public Health Laboratory in Quebec suspected an outbreak and sent the samples to the National Microbiology Laboratory to be investigated through whole genome sequencing. Subsequent core genome single nucleotide polymorphism (cgSNP) analysis revealed close genetic similarity with SNP distances ranging between 3 and 11 SNPs, indicative of a transmission cluster[1].

1.3 Tasks

We want to trace back the source of the outbreak. As person-to-person transmission is considered rare for *M. abscessus*, we start a hypothetical environmental investigation and take samples within the tattoo shop: 500 mL of tap water, 50 mL of ink, 500 mL of shower water and 500 mL of water from the aquarium. Samples are sent to the National Reference Laboratory in Borstel, Germany. After standard pre-processing procedures, including filtration and decontamination, samples were inoculated into liquid medium (mycobacterial growth indicator tubes, MGIT) as well as onto solid media (7H10 and Löwenstein-Jensen), incubated for 8 weeks at 30 and 37 °C, and monitored closely. If colonies grew, they were examined microscopically after Ziehl–Neelsen staining to identify acid-fast bacteria. Two different looking acid-fast bacteria with different growth rates were identified from the shower sample (isolate L and N). Also the tap (sample M) and aquarium water (sample O) tested positive for mycobacteria but not the ink. After DNA extraction and library preperation, the isolates

were sequenced on a Illumina machine. The raw data (fastQ) were assembled using the skesa[6] assembler and resulting assemblies (fastA) have been shared with you on the ECDC platform.

Table 1 Overview of isolates to be studied

Isolate identifier	Reference	Source
H	Wuzinski 2020	Patient 1
I	Wuzinski 2020	Patient 2
J	Wuzinski 2020	Patient 3
K	Wuzinski 2020	Patient 4
L	Hypothetical investigation	Shower
M	Hypothetical investigation	Tap water
N	Hypothetical investigation	Shower
O	Hypothetical investigation	Aquarium

The ultimate goal of this tutorial is to identify the source of the skin infection outbreak. To achieve this, several subtasks must be completed:

- Task 1: Determine which NTM (sub)species were isolated from the environmental samples (L-O) using Pathogenwatch and/or NTM-Profiler
- Task 2: Determine the 7-loci MLST Sequence Type (ST) of the *M. abscessus isolates* using Pathogenwatch and/or pubMLST
- Task 3: Determine whether the *M. abscessus* isolates belong to known dominant circulating clones (DCCs) using pubMLST and/or Ridom Typer (for existing users only)
- Task 4: Determine the genetic distance (expressed as allele differences) between outbreak isolates using pubMLST and/or Ridom Typer (for existing users only)
- Task 5: Determine the genetic resistance profiles of the isolates using NTM-Profiler

Complete all results in **Task_Exercise2026_NTM.xlsx.**

- Task 6: Determine the source of the outbreak: i.e. which strains isolated from the environmental samples are genetically similar to the ones from the patients → formulate a hypothesis regarding the transmission route

2 Gather required data

Make sure you have downloaded all required files for today's tutorial from the ECDC learning platform to a folder you can access:

1. Folder *FastA* → containing draft genomes (fastA files) of the 8 isolates (clinical and environmental)

2. *Task_Exercise2026_NTM.xlsx* → to store results

3. *110_Ref_Mabscessus_pubMLST.xlsx* → to determine DCCs

The file *110_Ref_Mabscessus_Ridom.xlsx* is optional and intended only for users with a valid Ridom Typer (SeqSphere+) license. FastQ files are also optional and can be downloaded and used by those with command-line experience.

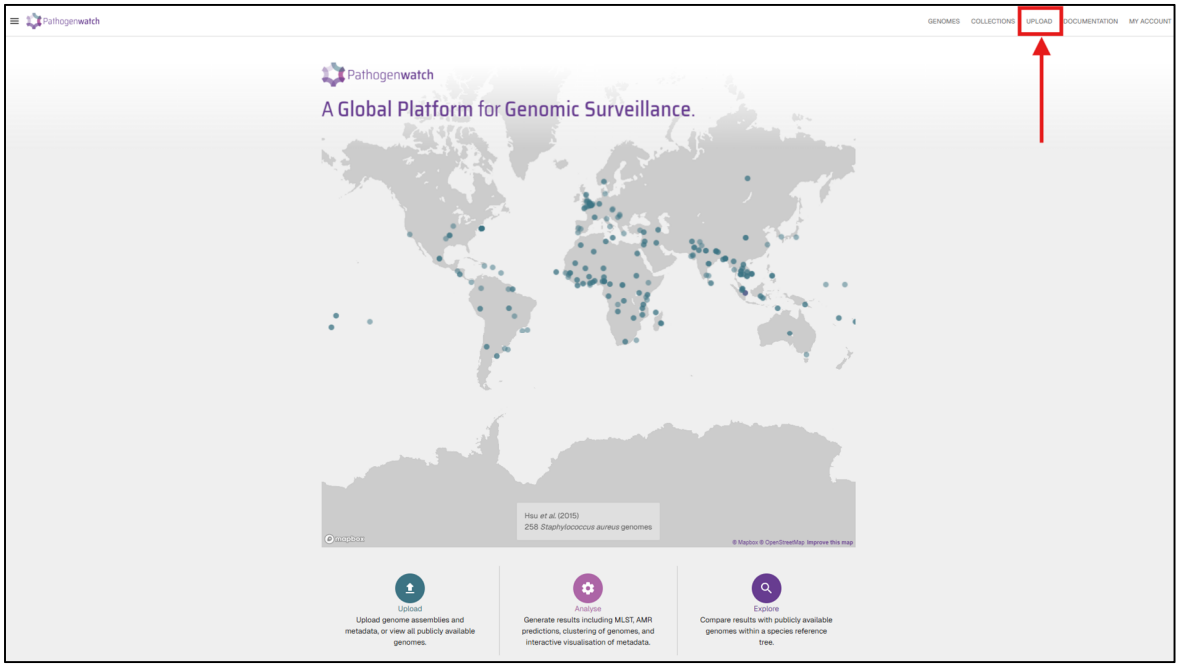
3 Bioinformatic analysis

Note: Software tools often change over time, so some screenshots might not fully reflect the current interface. If you notice any mistakes or need assistance, please feel free to contact me at mdiricks@fz-borstel.de.

3.1 Web-based and desktop GUI Tools

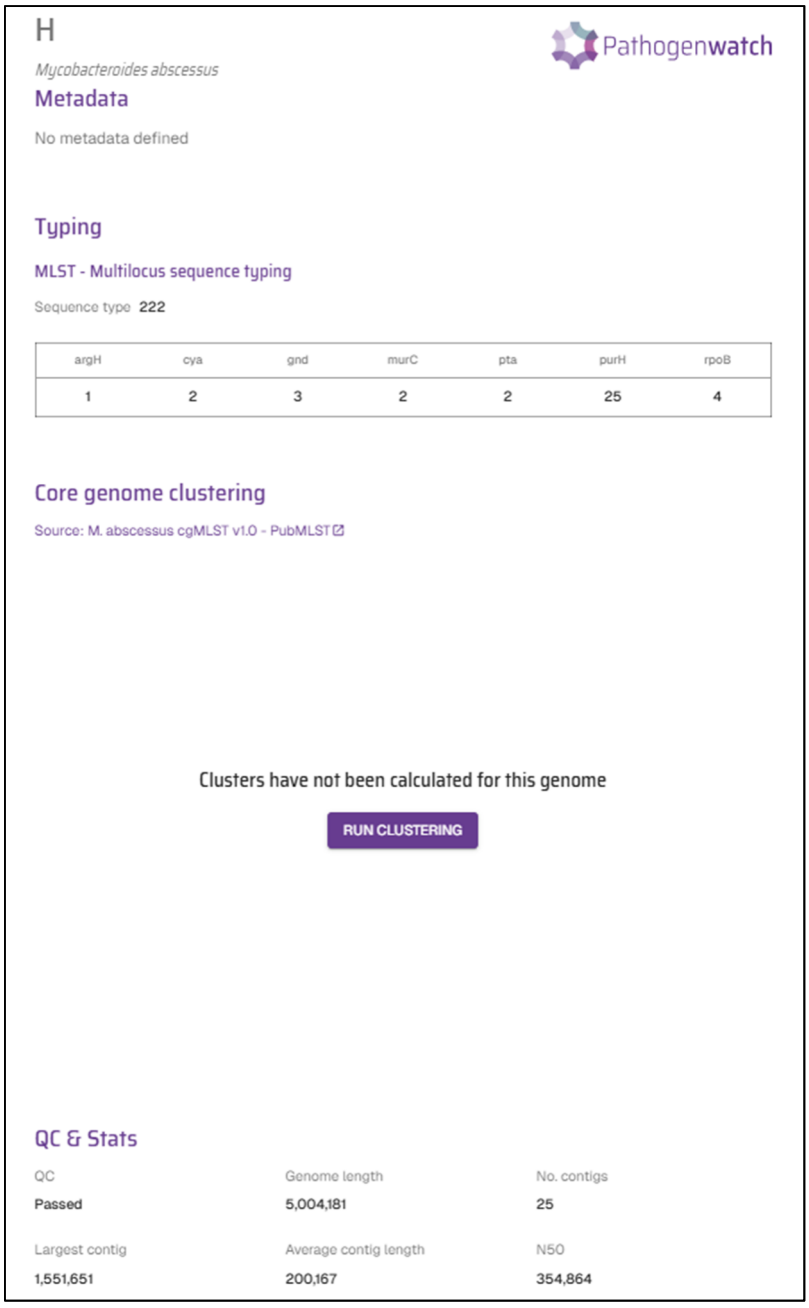
3.1.1 Species identification and typing using Pathogenwatch

1. Browse to https://pathogen.watch/



2. Click on "**Upload**" at the **top** of the website and sign in. If it is the first time you sign in, you will need to click on the activation link that will be sent to you.

3. Open Windows Explorer and navigate to the downloaded *FastA* folder. Drag and drop the 8 fastA files into the middle of the Pathogenwatch website. Alternatively, click on the "**+**" sign at the bottom right, browse to the *FastA* folder, select all 8 files, and click "**Open**".
4. Once all samples are analysed, click on "**View Genomes**" in the center of the circle. You will see the species annotation of the sequences you have just uploaded. Complete these results in **Task_Exercise2026_NTM.xlsx** (Column species)**.**
5. Click on a sample (e.g. H) in the column "Name" to open a **sample-specific report**. Here you can see the full MLST ST profile, assembly stats and organism prediction. Check these reports for all samples and complete the results in **Task_Exercise2026_NTM.xlsx** (Column MLST ST)**.**

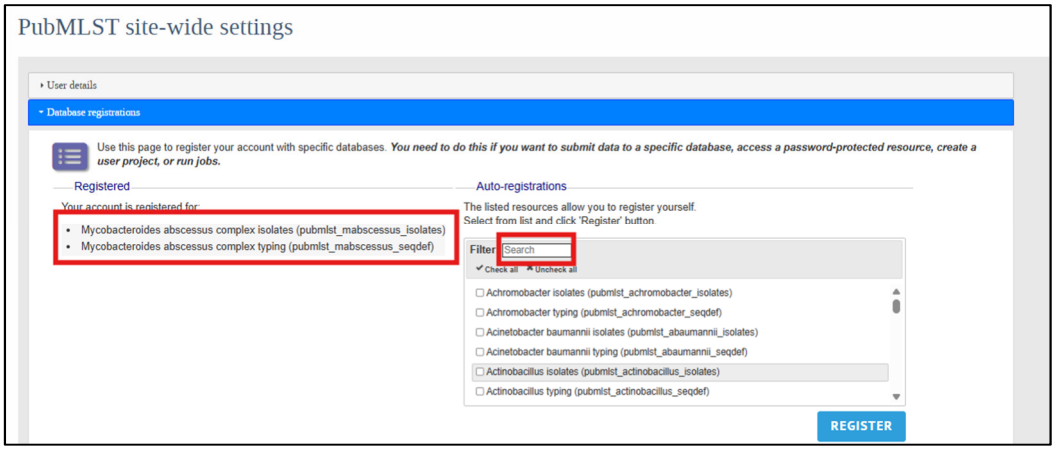


6. Click in the middle of the report on „**Run Clustering**" and click on "**View Cluster**". Are the patient samples related? Which environmental isolate clusters with the patient isolates?

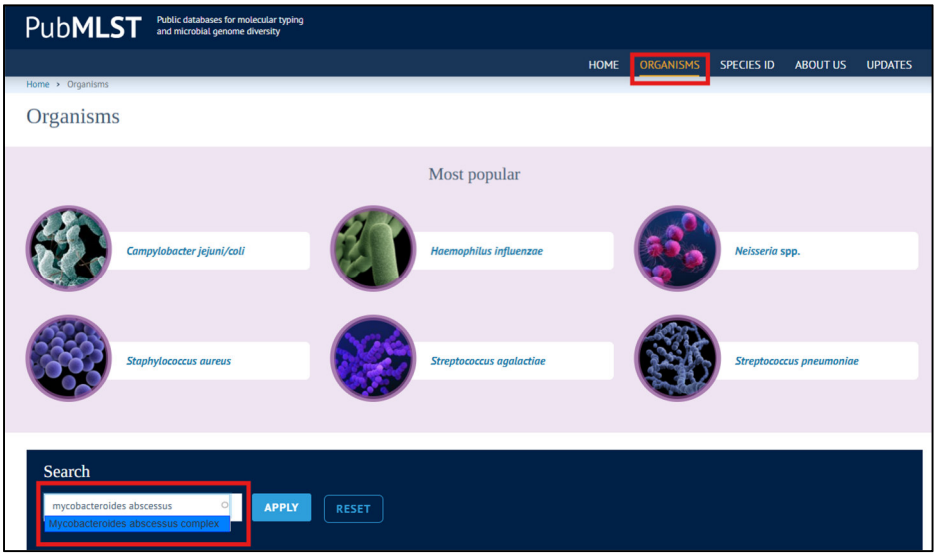
3.1.2 Typing using PubMLST

We will make use of the web-based typing tool PubMLST to determine the multi-locus sequence type (MLST ST) and genetic relatedness of the 6 *M. abscessus* strains.

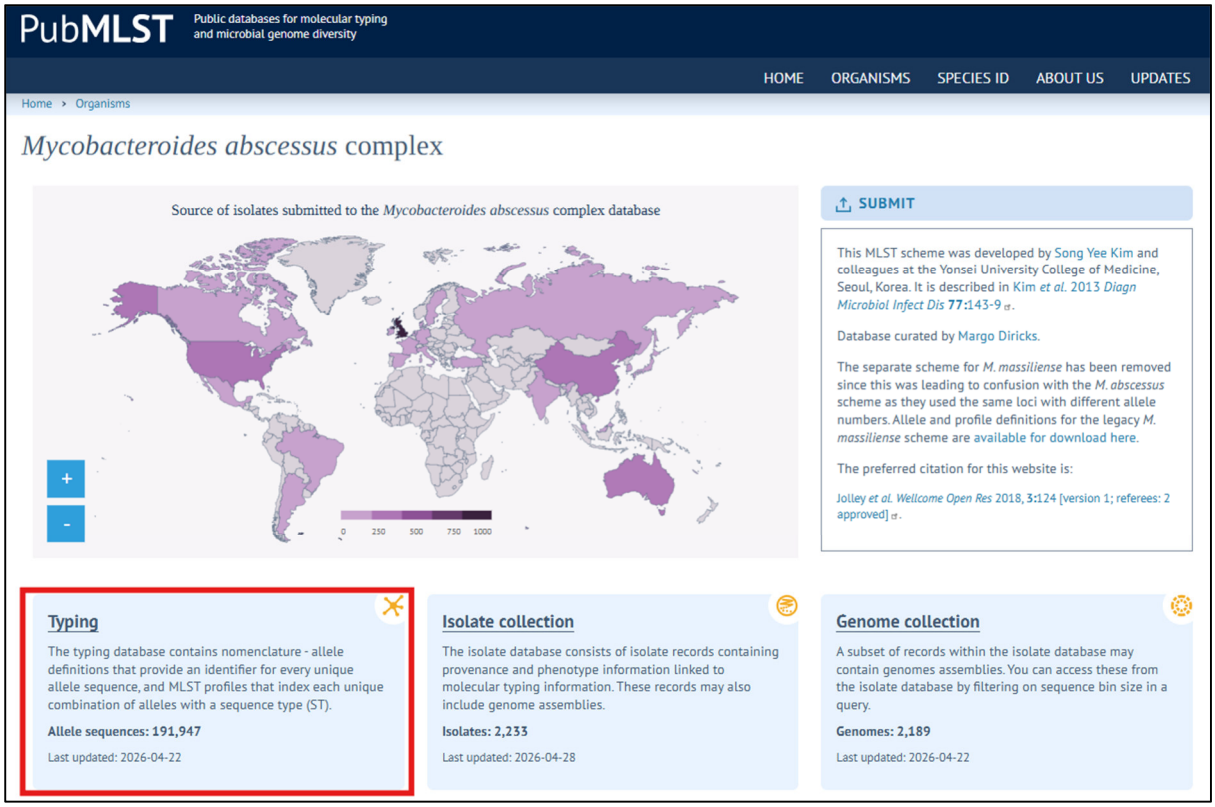
- Before you can start:
  - Make a pubMLST account: follow instructions at https://pubmlst.org/site-accounts
  - Log into your account. Click at the upper right on My account and look for the *Mycobacteroides abscessus* databases via the Search function in the tab "Database registrations". Click on Register, they now should appear on the left under "Registered".



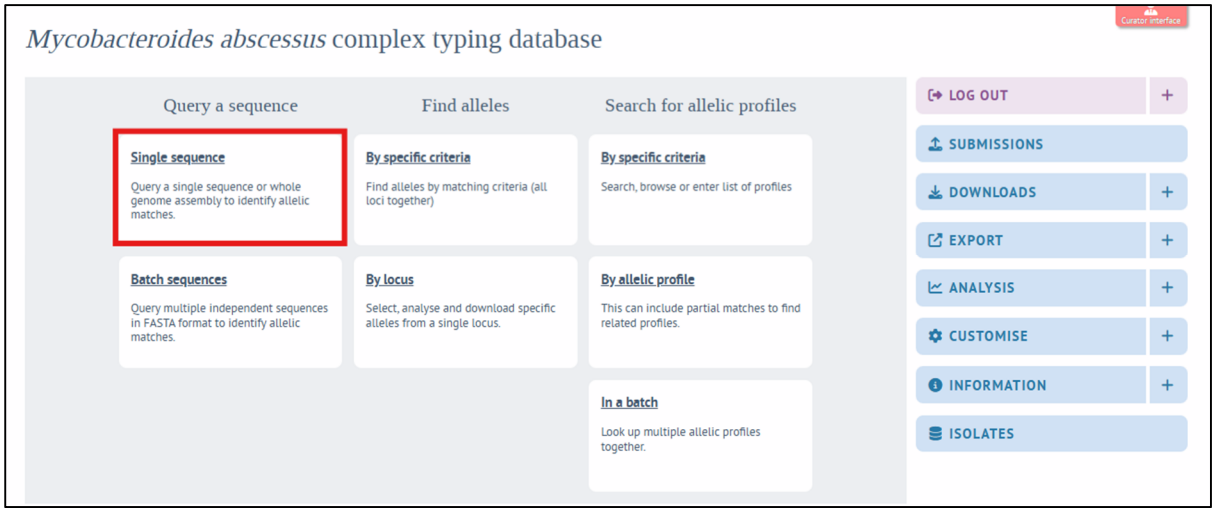
7. Browse to the website https://pubmlst.org/, click at the top on "**ORGANISMS**" and search for *Mycobacteroides abscessus*



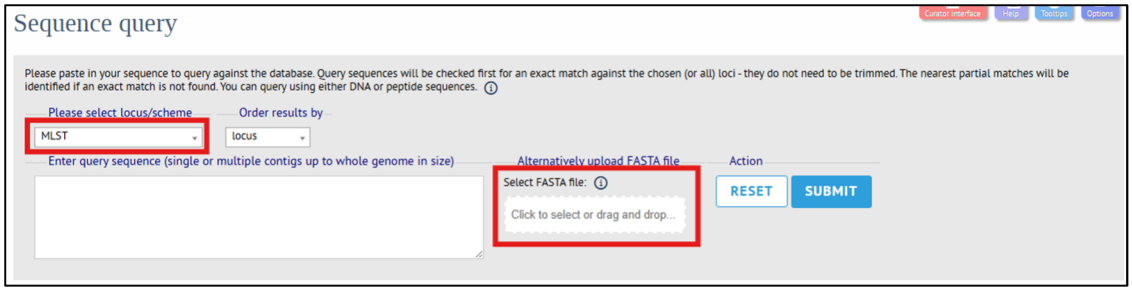
8. Click on **Mycobacteroides abscessus complex** and click on "**Typing**" (box at the bottom left).

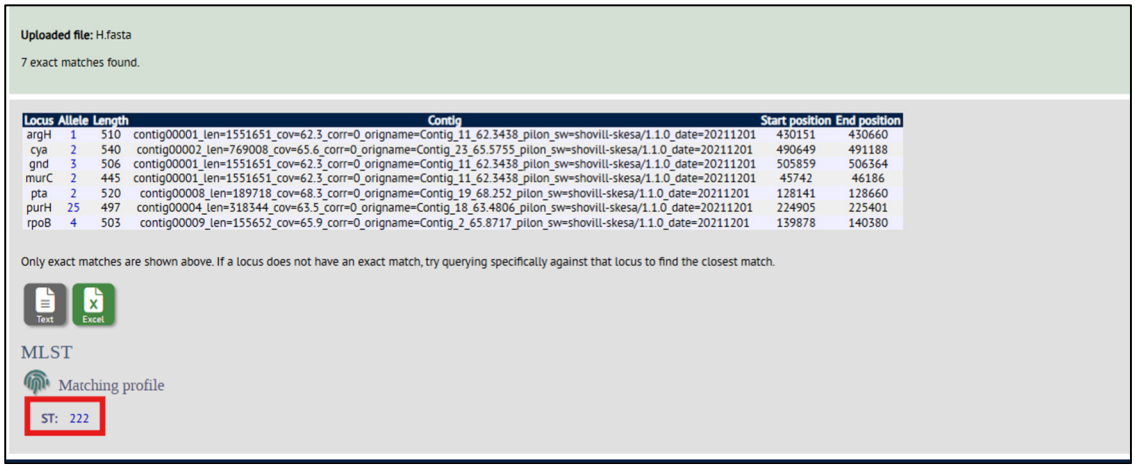


9. At the right, click on **+** sign next to "**Information**" and click on **"database status"**. You will see that there are two typing schemes available: MLST (based on 7 loci) and cgMLST (based on 2,904 loci defined in Diricks et al. 2022[5]). These can both be used to type and compare *M. abscessus* isolates. The cgMLST scheme can also be used for outbreak analysis.
10. Go back to the previous screen and click on the box "**Single Sequence**".

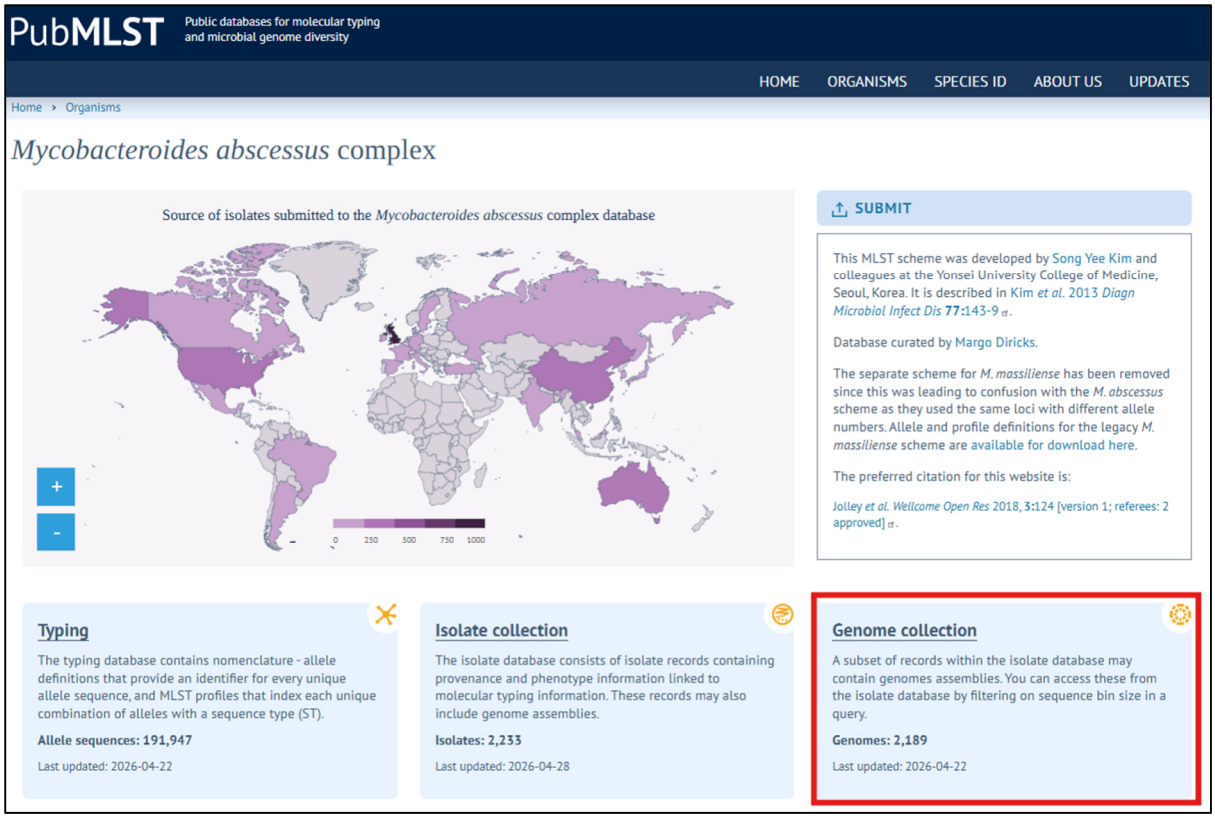


11. Select "**MLST**" in the first box (Please select locus/scheme). In the middle, click on "**select fasta file**" and navigate to the folder FastA and choose H.fasta. Click on "**Submit**".

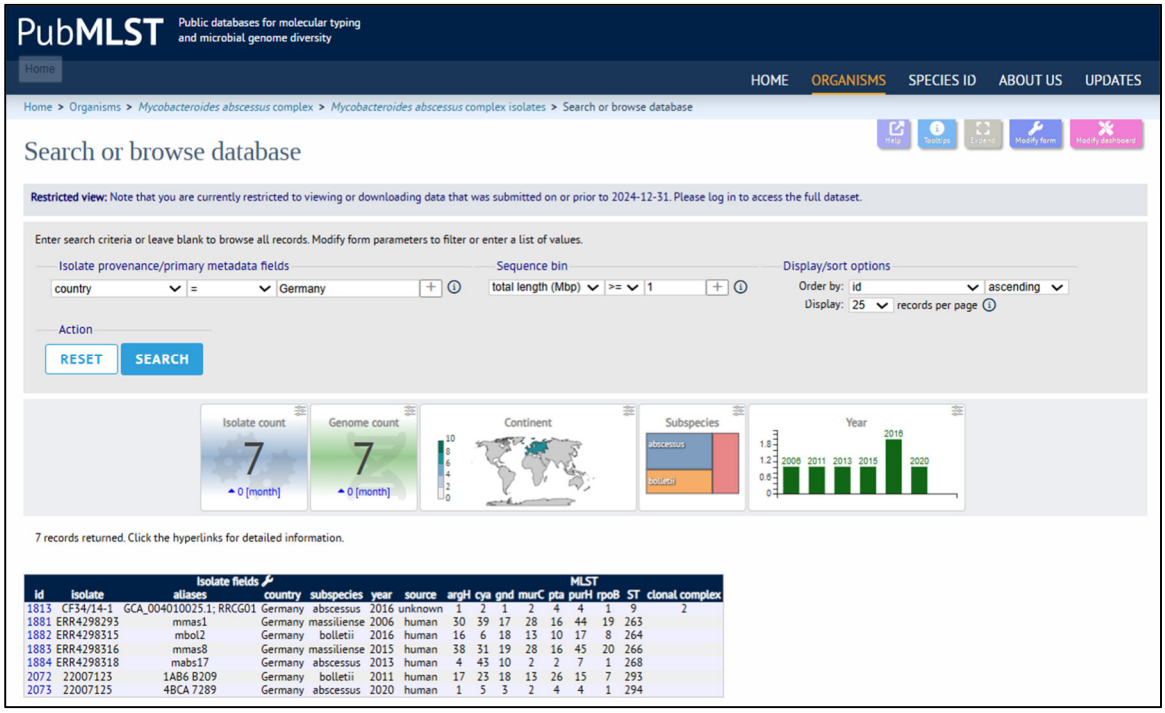




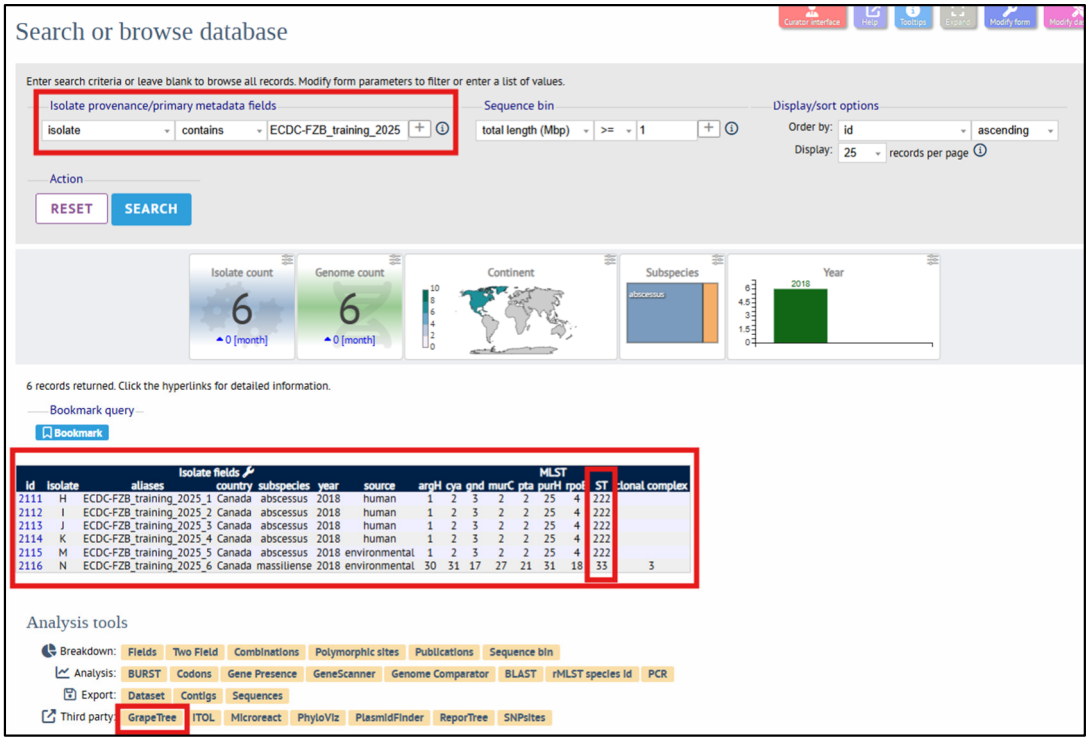
12. Do the same for the other 7 fastA files. Enter the MLST sequence type results in **Task_Exercise2026_NTM.xlsx** (column MLST ST). Compare the MLST ST from pubMLST with that from pathogenwatch. It should be the same as they are using the same nomenclature!
13. Go back to https://pubmlst.org/organisms/mycobacteroides-abscessus-complex and click on „**Genome collection**"



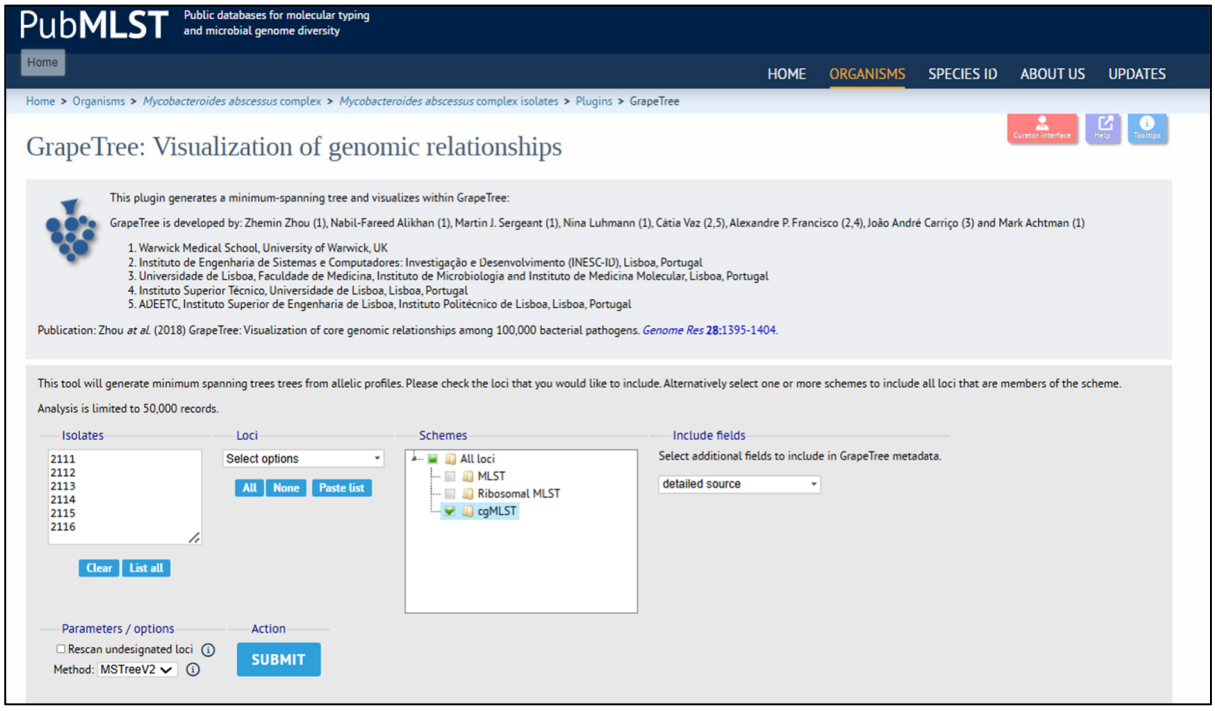
14. Look if there are *M. abscessus* genomes of your country included in this database: choose "**Country**" in first field of "**Isolate provenance/primary metadata fields**" and your country name in the third field.



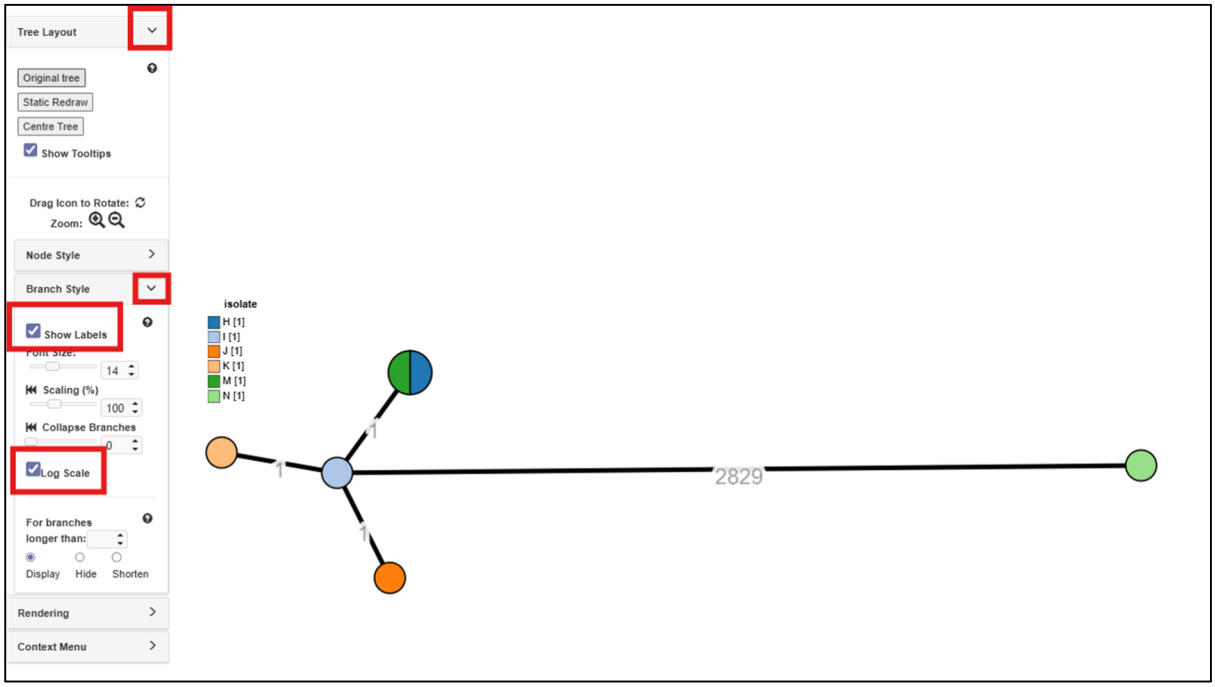
15. Choose "**isolate**" in the "**Isolate provenance/primary metadata fields**", choose "**contains**" in the second field and enter "**ECDC-FZB_training_2025**" in the third field and click on Search (only visible if you have signed in!). You will see the 6 *M. abscessus* isolates from our exercise which I have uploaded before to this database. Note that the MLST ST types are also indicated here.



16. Click at the bottom on “**GrapeTree**”.
17. Check the box next to cgMLST in the middle box (schemes). And choose „**Detailed source**“ in the „**Include fields**“ at the right. Click on “**Submit**”.



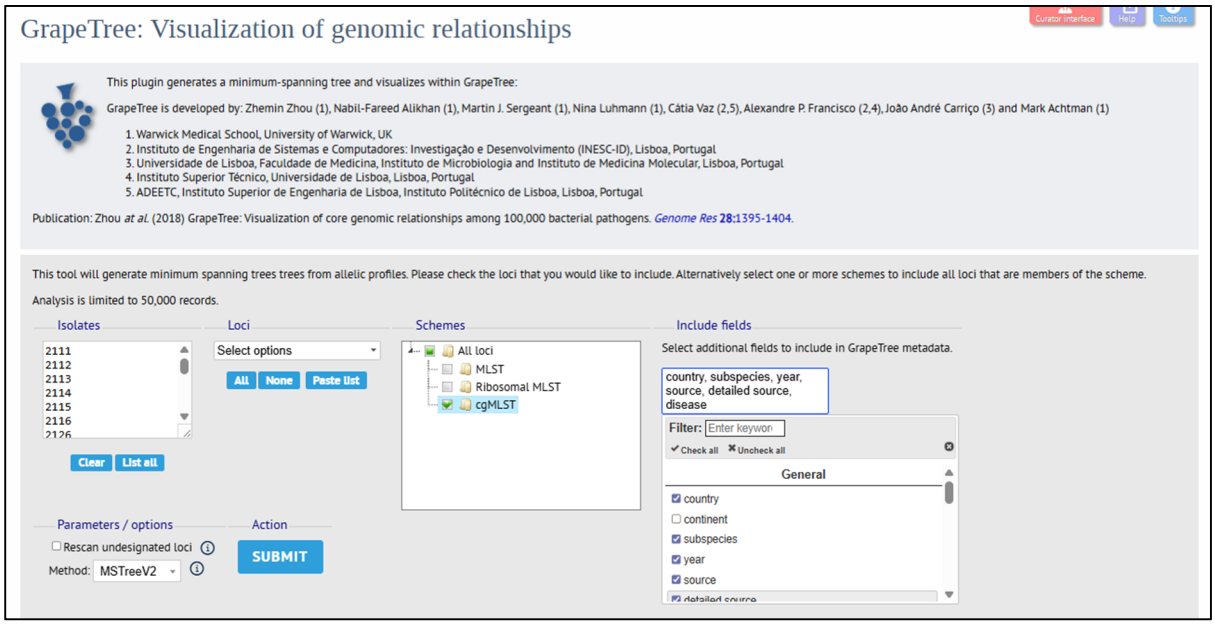
18. After calculations are done, click on „**Launch GrapeTree**“. Click on the “>” next to **“Tree Layout**” and on “>” next to **“Branch style**”. Check box next to “**Show labels**“ on branches to display the allele distance between samples. Also check the box next to “**Log scale**” to better see the small distances.



What do you think is the source of the outbreak? Which environmental isolate is closely related to the clinical isolates from the skin infection?

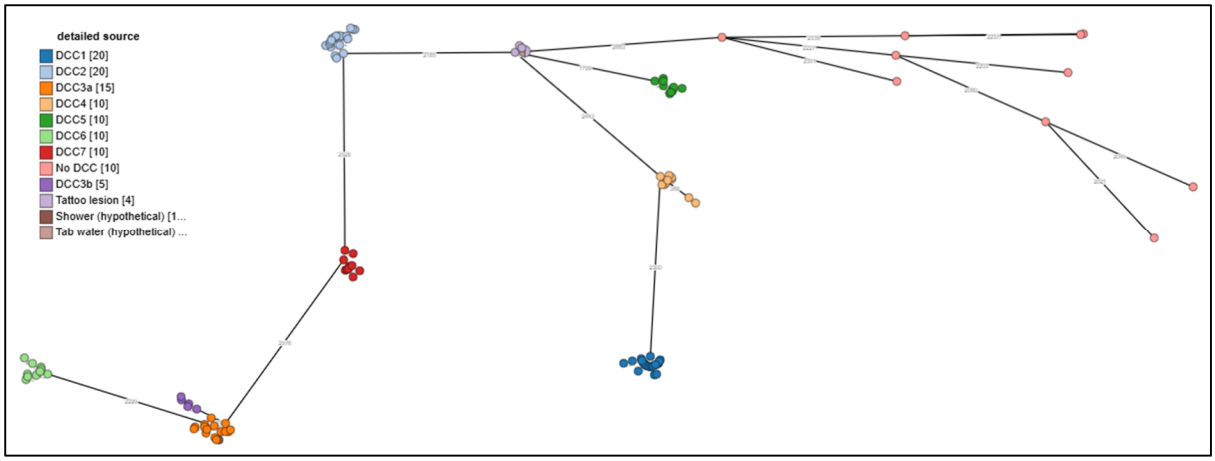
Do our *M. abscessus* isolates belong to one of the known seven dominant circulating clones (DCCs)?

1. To address this question, multiple approaches can be used. As a first step, a preliminary check can be performed. DCC strains typically share the same 7-loci MLST sequence type (see Diricks et al., 2022[5]). Please open the file "**110_Ref_Mabscessus_pubMLST.xlsx**" for the most common ST types associated with each DCC (e.g. DCC1 → ST5). Do the isolates H, I, J, K, M, and N have the same ST type as any of the known DCCs?
2. A more accurate approach is to compare your isolates with known DCC isolates at the whole-genome level, for example through phylogenetic positioning. This can be done, for example, using pubMLST.
3. Go back to the **genome collection**: https://pubmlst.org/bigsdb?db=pubmlst_mabscessus_isolates&page=query&genomes=1
4. Click on "**Grapetree**" at the Bottom
5. Click on "**Clear**" To remove all current Isolate IDs and past the list of identifiers from column A in the "**110_Ref_Mabscessus_pubMLST.xlsx**" file in the "**Isolates field**"
6. Check the box next to "**cgMLST**" in the "**Schemes**" Panel and under "**Include fields**" select country, subspecies, year, source, detailed source and disease.

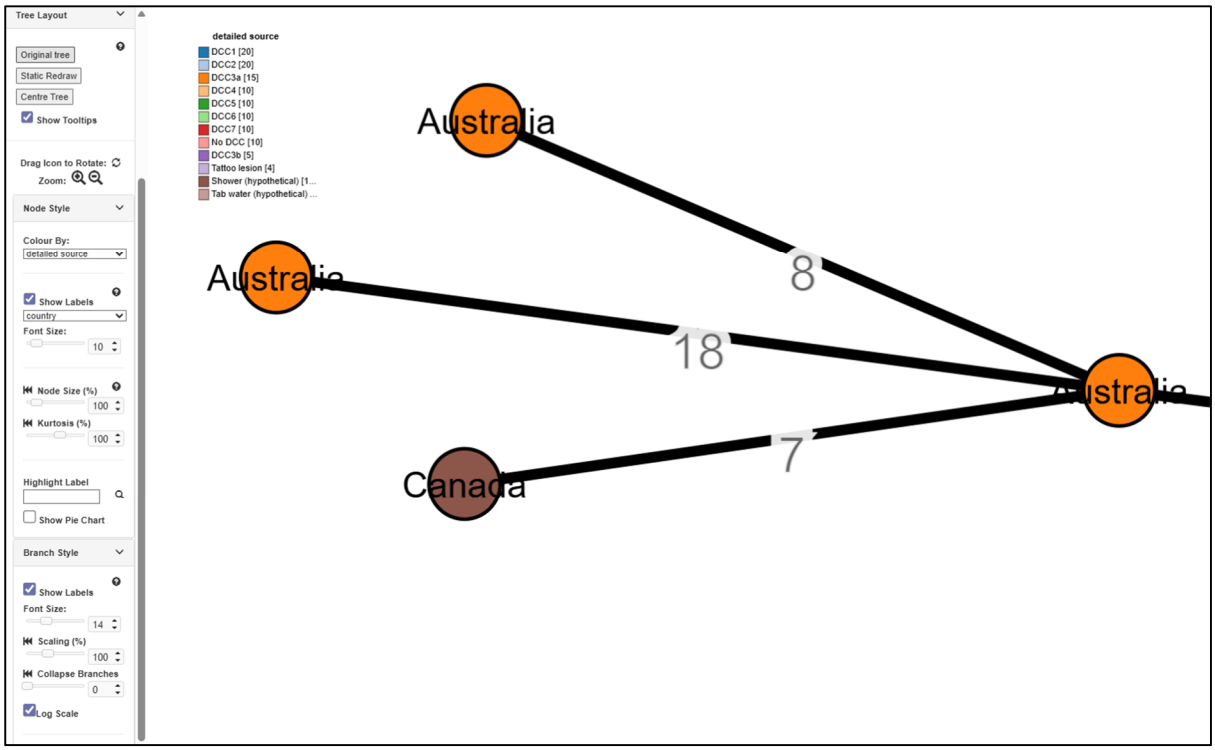


7. Click on submit and next on "**Launch Grapetree**" as soon as the job is done.
8. In the left panel, click on the > arrow next to "**Tree Layout**" and choose the option "**Colour** by: "**detailed source**"
9. In the left panel, click on the > arrow next to "**Branch style**" and choose the option "**Show labels**".
10. Tip: you can turn a tree by clicking on "**Drag Icon to Rotate**" in the left panel. You can also move the branches by clicking on the branches and dragging them in a specific direction.

Do the outbreak isolates belong to one of the seven DCCs?



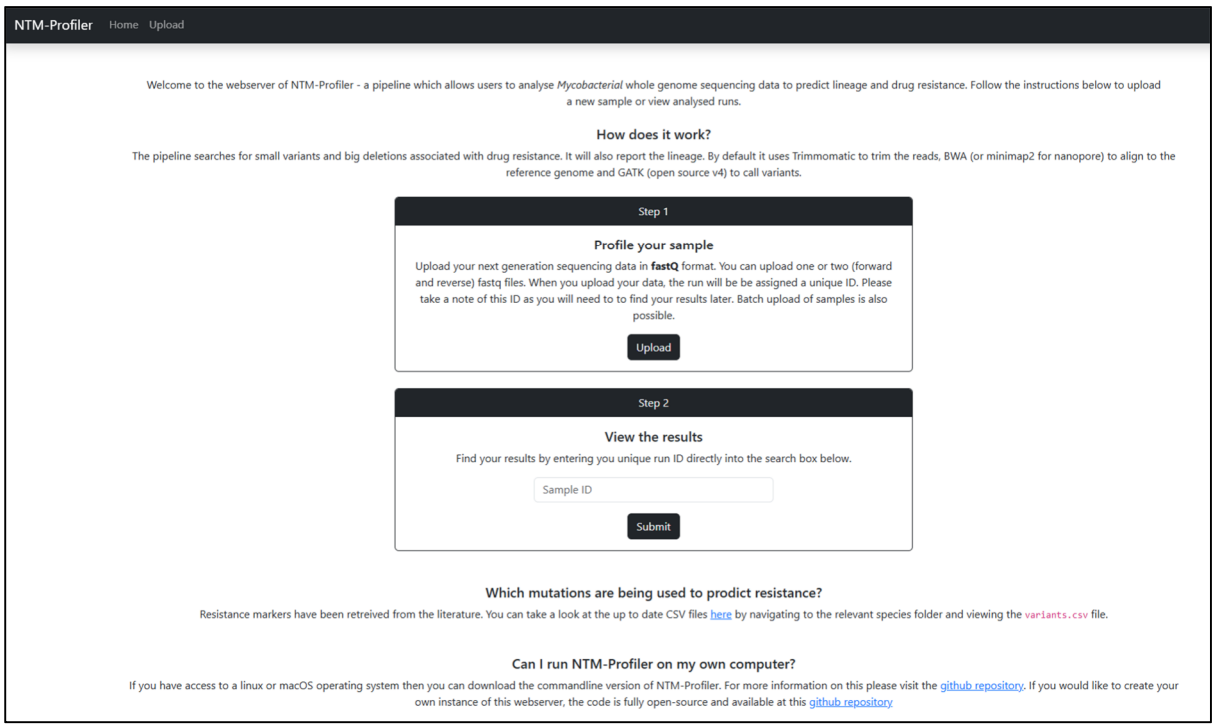
11. Put on the "**log scale"** (in the left panel under "**Branch style**") to better see the genetic distance (number on branches) between isolates. Under "**node Style**" select the option "**Show labels**" and choose "**country**" as label. The shower isolate N from Canada (brown disc) seems to be very closely related to a DCC3a isolate from Australia with a genetic distance of only 7 alleles. As isolate N is closely related (<250 allelic differences) to other DCC3a isolates (i.e., it falls within the same cluster), it can be classified as belonging to DCC3a. Please record this result in **Task_Exercise2026_NTM.xlsx** (column DCC).



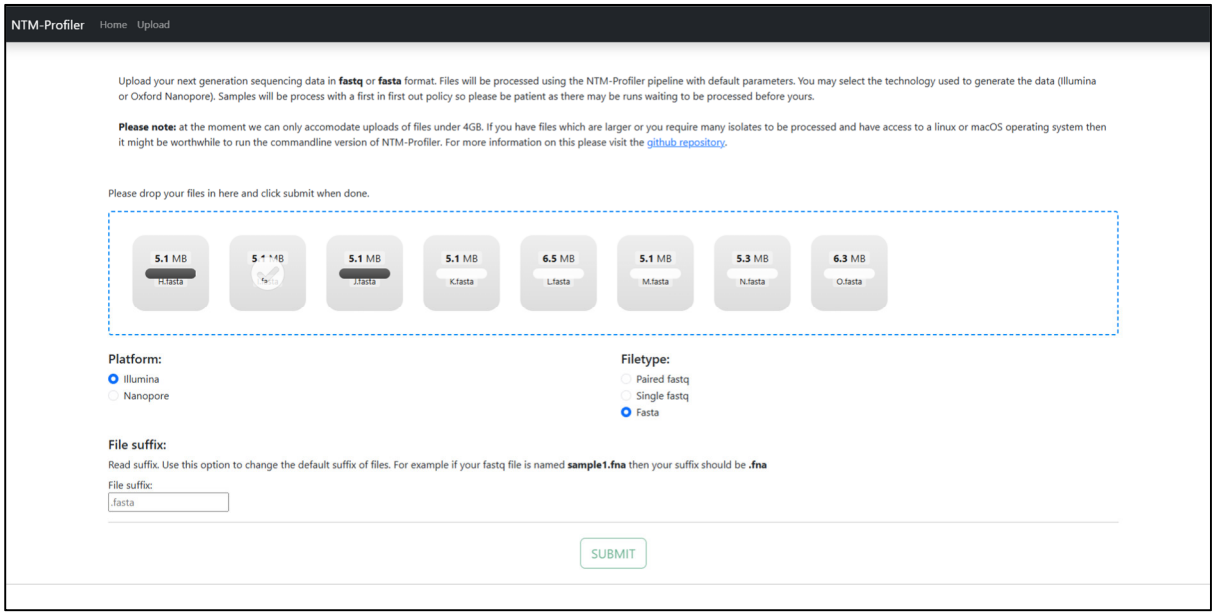
12. Note: you can also submit and analyse your own genomes in pubMLST: https://pubmlst.org/submit-data. After submission, you will need to wait for the curator (which is me for *M. abscessus* 😊) to accept your genomes.

3.1.3 Species identification and resistance prediction with NTM-Profiler (web-server)

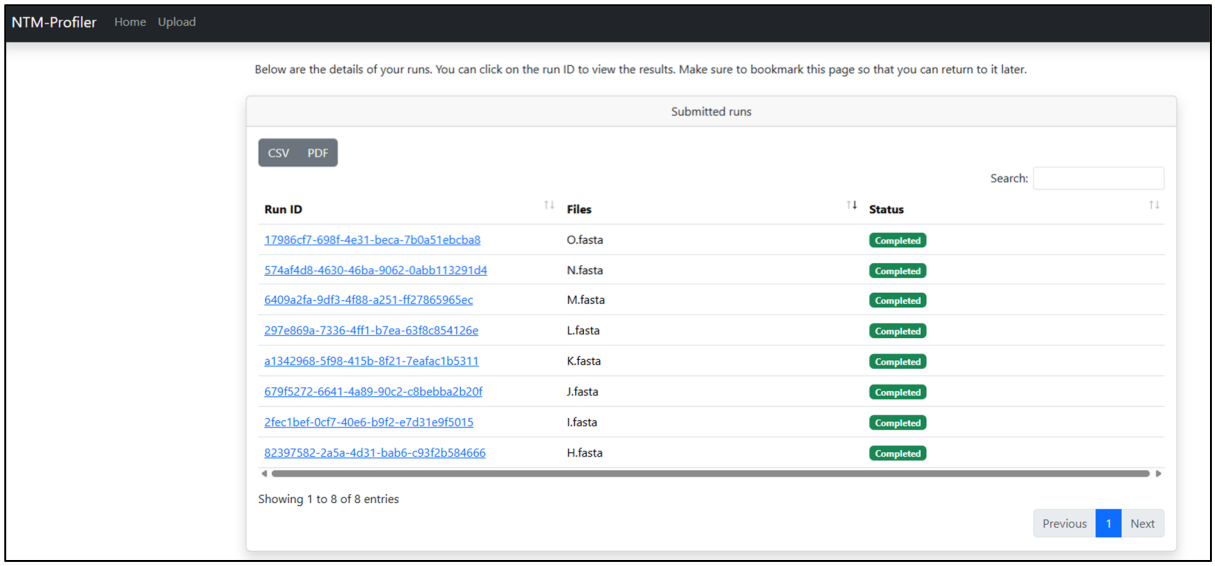
1. Browse to https://bioinformatics.lshtm.ac.uk/ntm-profiler/
2. Click on “Upload”



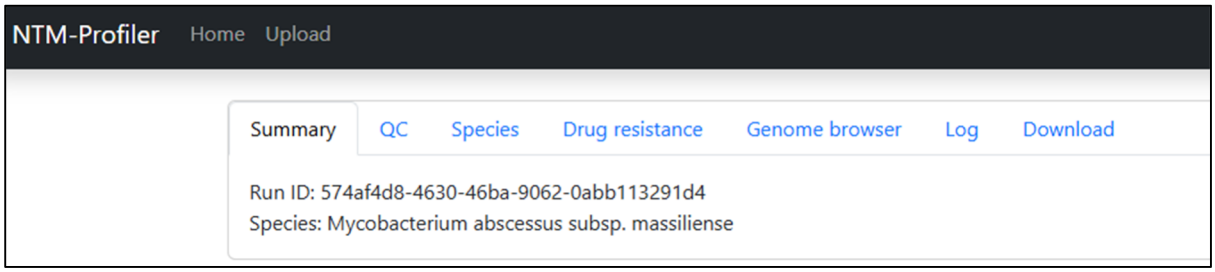
3. Select “**Fasta**” as filetype and drop the 8 fastA files from windows explorer in the central part.

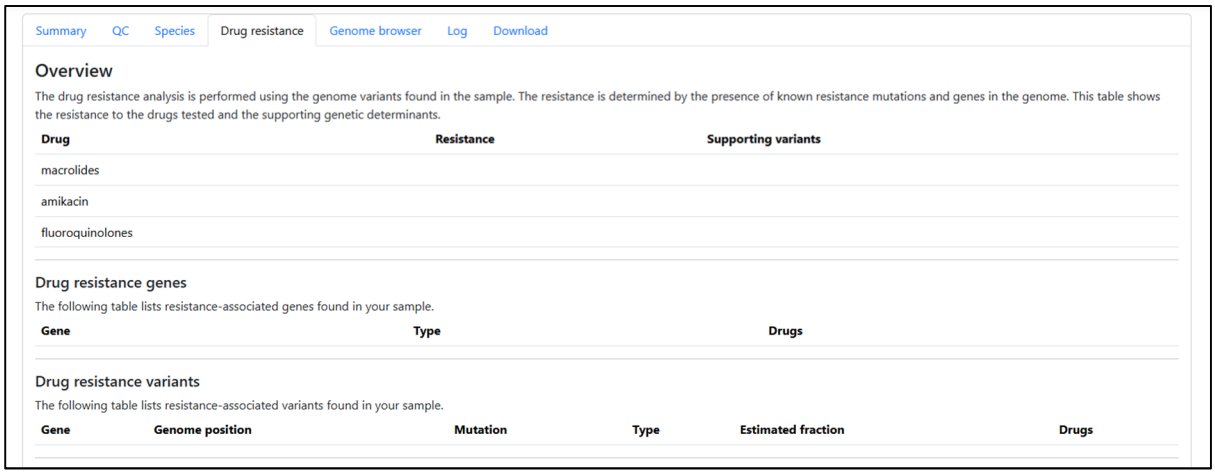


4. Click on “**Submit**”. Processing will take around 2 min.
5. After the processing finished, you can click on every single isolate to look at the results. Do you get the same results as Pathogenwatch? Complete the results in **Task_Exercise2026_NTM.xlsx** (especially column **subspecies**).



6. Have a look at the drug resistance tabs of all samples. Do you see any known resistance related mutations? What would you have expected for *M. abscessus* subsp. *abscessus*; can you explain the macrolide susceptibility of the outbreak isolates?





Optional: Only for current users of Ridom Typer (SeqSphere+) with an active license

3.1.4 CgMLST-based cluster analysis using Bruker MBioSEQ Ridom Typer

This is commercial software: https://www.ridom.de/ridom-typer/

Note: Bruker MBioSEQ Ridom Typer was formerly known as SeqSphere+. The name of this software application changed after the company Ridom was acquired by Bruker in 2025.

3.1.4.1 Preliminaries

Installation: This tutorial requires Ridom Typer (client and server) to be downloaded and installed (https://www.ridom.de/ridom-typer/update/)

- **License:** To be able to use Ridom Typer you need to have a valid license → e.g a paid license from your institute or a free demo license (only for in-person trainings in Borstel, not available for this online training)

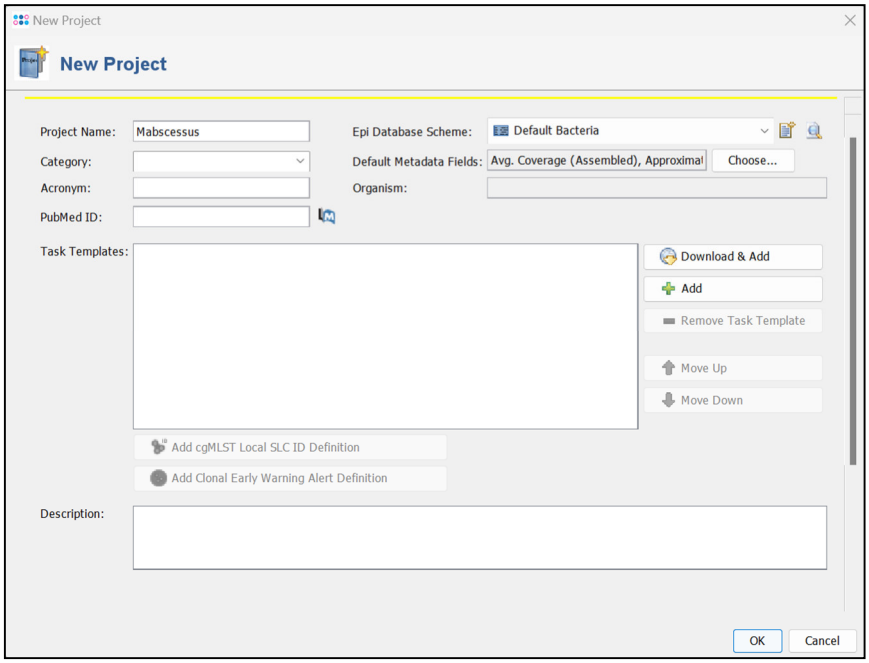
3.1.4.2 Start Ridom Typer

1. Double click on the icon Ridom Typer on your desktop
2. Step 3: Login with your credentials

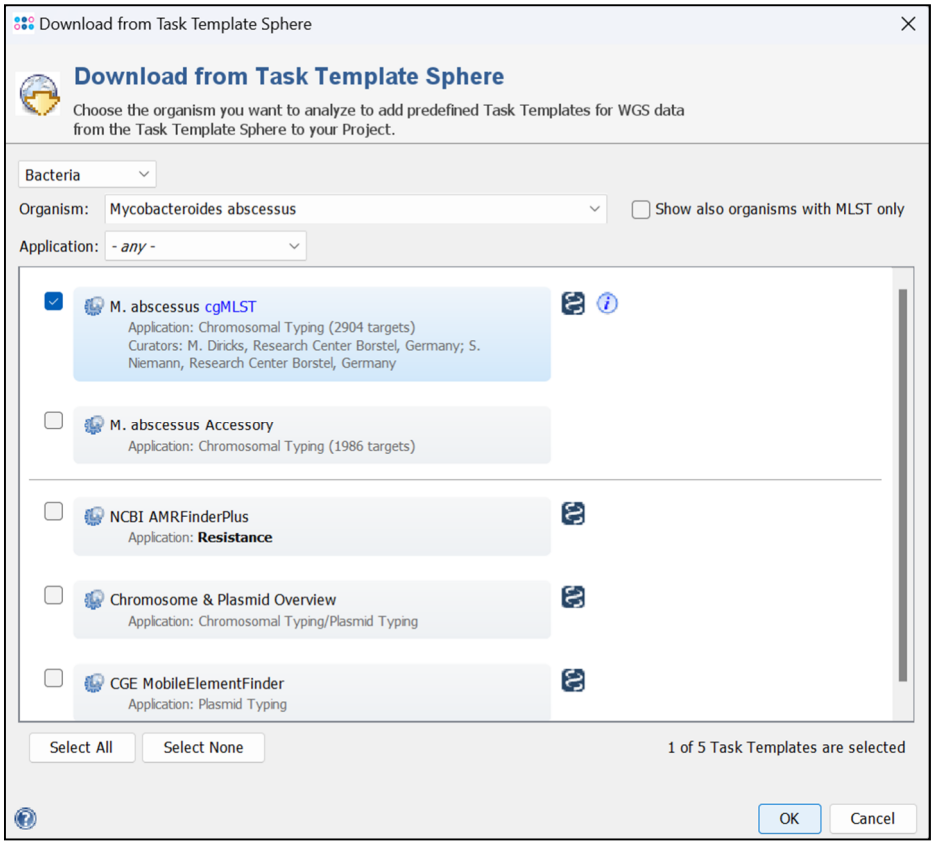
3.1.4.3 Define Project, Database Scheme, and Tasks

1. Use the "**New Project"** Shortcut from the Ridom Typer Welcome screen.

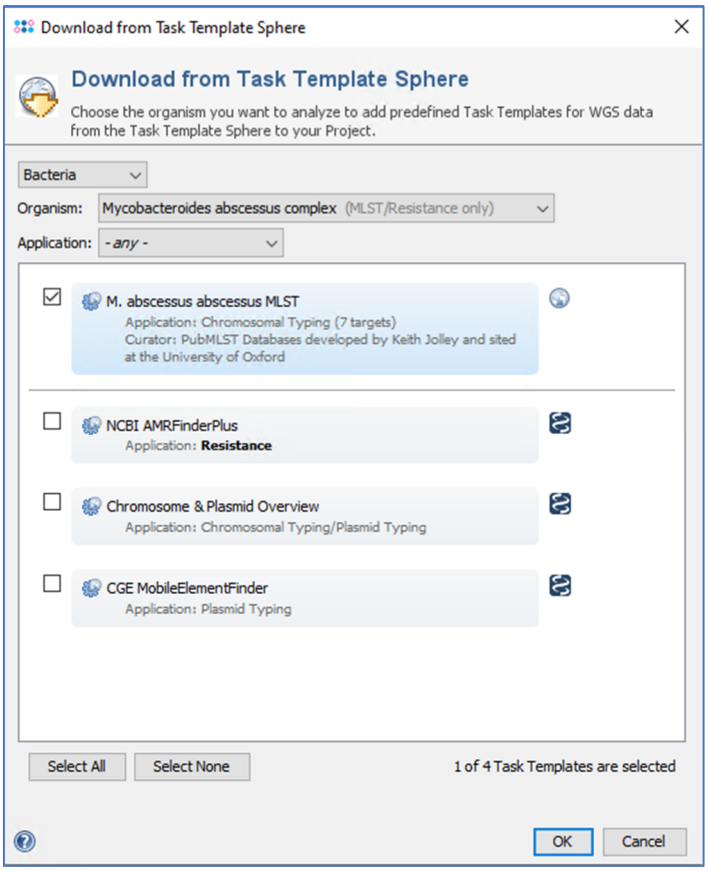
2. The New Project window opens. Enter a "**Project Name"**, e.g., Mabscessus.



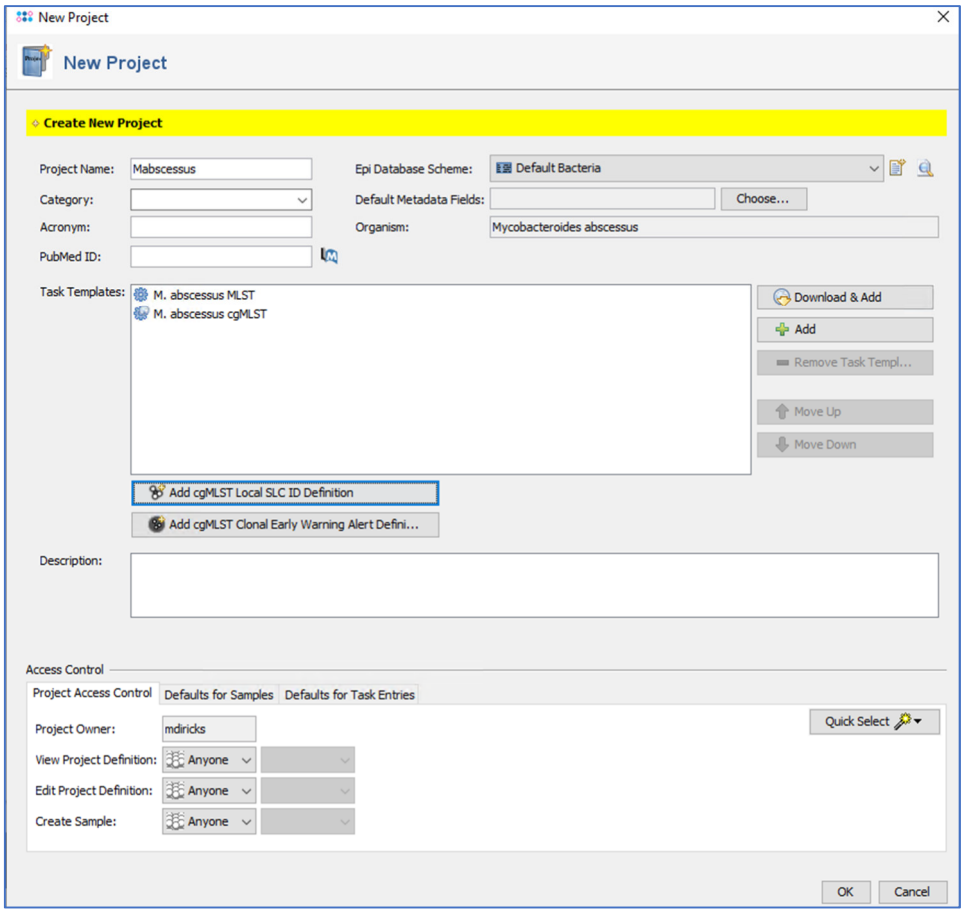
3. Click the button "**Download & Add**" to select Task Templates. Select **Mycobacteroides abscessus** as organism and **uncheck "M. abscessus Accessory**". Click "**OK**".



4. Click again on **Download & Add to select Task Templates**. Select **Mycobacteroides abscessus** complex (MLST/Resistance only) as organism and click on OK.



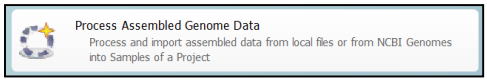
The project window should now look like this:



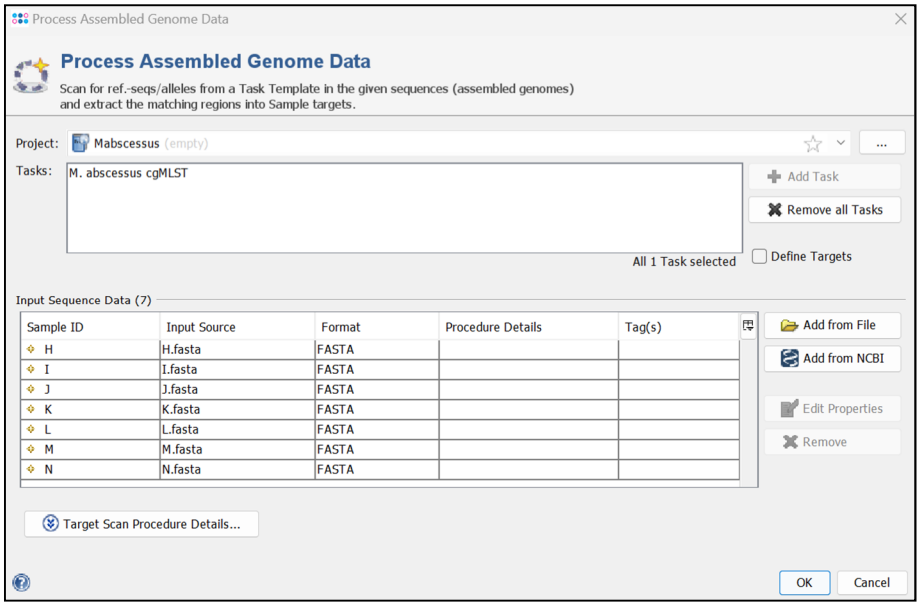
5. **Close** the New project window by clicking on „**Ok**" at the bottom.

3.1.4.4 Process already assembled genome data

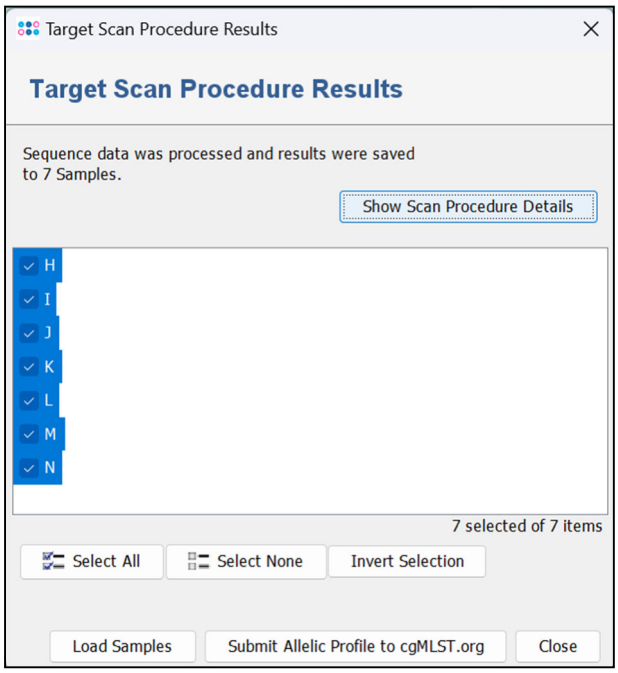
1. Click on **Process Assembled Genome Data** button from the Ridom Typer Welcome screen.



2. Make sure your previously defined project (Mabscessus) is selected and click at "**Add from File**" at the right. Browse to your training folder **FastA** and select sample H to N (7 in total) and click on "**OK**". Click again on "**OK**" to start processing the samples. Ridom Typer will now scan the genomes for the 2,904 genes included in the *M. abscessus* cgMLST scheme and determine the allele numbers. This will only take about 7 min.



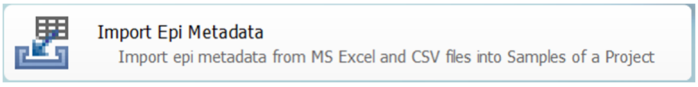
Once the samples are processed you should see following screen:



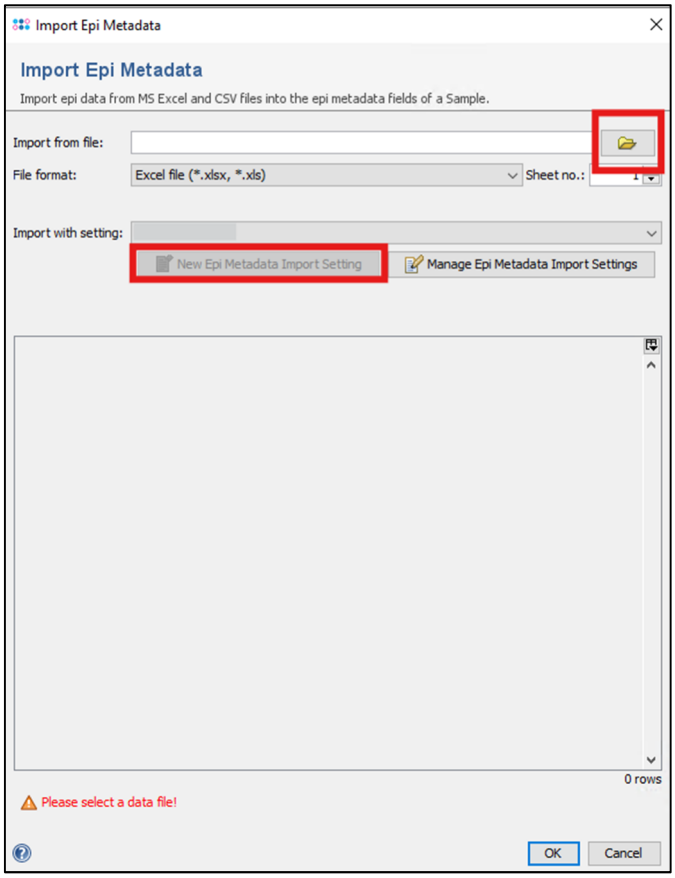
3. After the samples are processed. Click on "**Close**".

3.1.4.5 Import epidata of M. abscessus strains under investigation

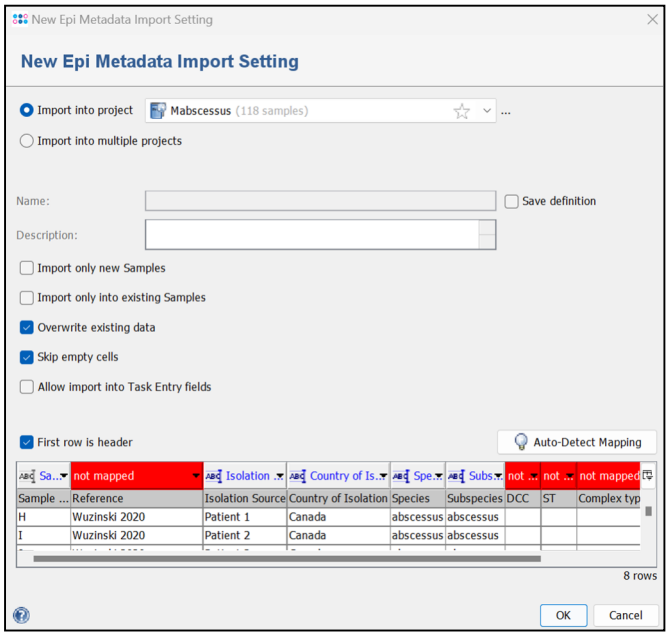
1. Click on the "**Import Epi Metadata"** button in the main window of Ridom Typer.



2. In the upcoming "**Import Epi Metadata"** dialog click on the folder icon and select the Excel file **Task_Exercise2026_NTM.xlsx.** Click on "**OK**".



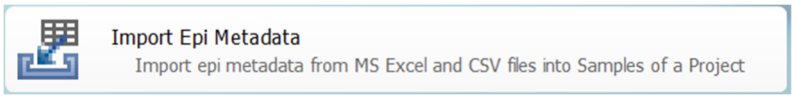
3. Click on the "**New Epi Metadata Import Setting button**" (will be come available once you have selected a file, see previous step). In the upcoming "**New Epi Metadata Import Setting"** select your previously defined **project** (Mabscessus) in the first row. Check the box next to "**Overwrite existing data**". Check the box next to "**First row is header**". Click on the "**Auto-Detect Mapping**" function on the right. Click on "**OK"**. Click once again on "**OK**". The data will now be imported. After the data has been imported, click on „**Close**".



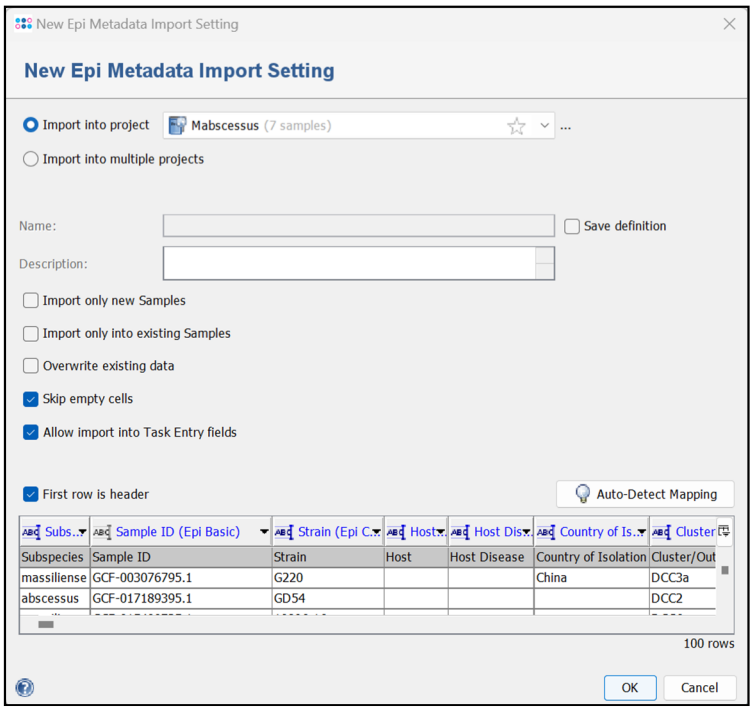
3.1.4.6 Import results and epidata of M. abscessus reference strains

To put our strains in a global context, to determine the subspecies and to be able to determine whether our strains belong to one of the known seven dominant circulating clones (DCCs), we will compare the cgMLST profiles of our 6 *M. abscessus* strains with a set of 110 *M. abscessus* reference strains. Accession numbers of these 110 references strains can be found here: https://github.com/ngs-fzb/NTMtools/tree/main/References. To safe time, we have already processed these genomes. The resulting cgMLST profiles as well as epidata can be found in file 110_Ref_Mabscessus_Ridom.xlsx. We can import this excel file directly into Ridom Typer.

4. Click on the "**Import Epi Metadata"** button in the main window of Ridom Typer.



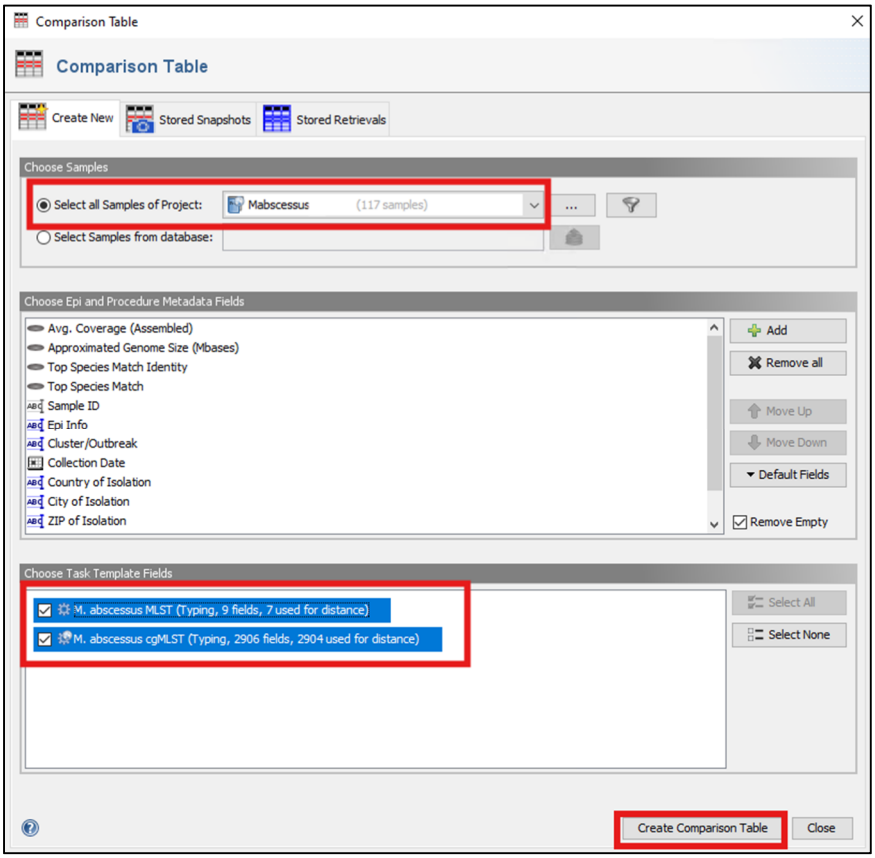
5. In the upcoming "**Import Epi Metadata**" dialog click on the folder icon and select the file **110_Ref_Mabscessus_Ridom.xlsx**. Click on "**OK"**.
6. Click on the button "**New Epi Metadata Import Setting**". In the upcoming "**New Epi Metadata Import Setting"** select your previously defined **project** (Mabscessus) in the first row. Check the box next to "**Allow import into Task Entry fields**". Check the box next to "**First row is header**". Click on the "**Auto-Detect Mapping**" button on the right. Click on "**OK**". Click once again on "**OK**". The data will now be imported (will take a couple of minutes).



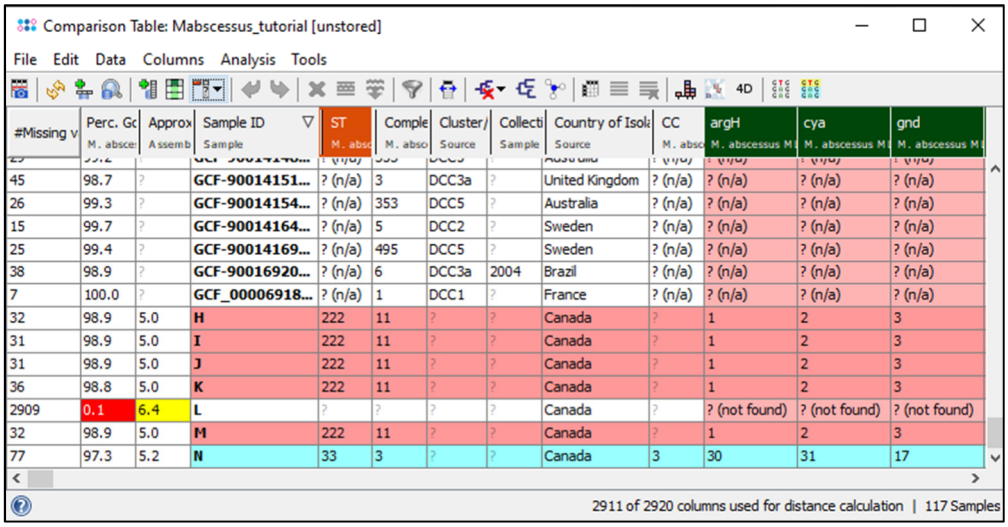
3.1.4.7 Cluster analysis

1. Click on the **Comparison Table** button in the main window of Ridom Typer.

2. Make sure "**Select all Samples**" of Project Mabscessus is selected, which contains data from 117 samples (our seven outbreak samples and 110 reference *M. abscessus* strains)
3. In addition, make sure "**M. abscessus MLST**" is also checked
4. Click on "**Create Comparison table**" at the bottom right.

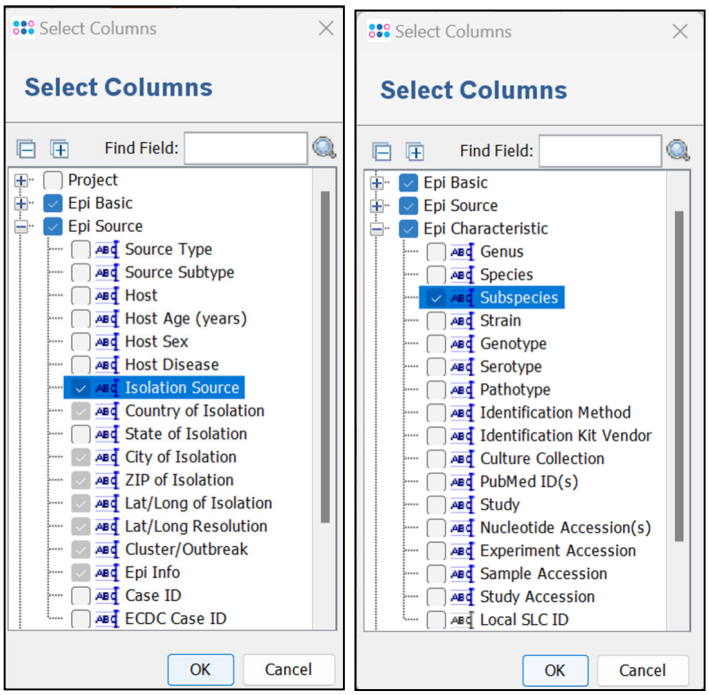


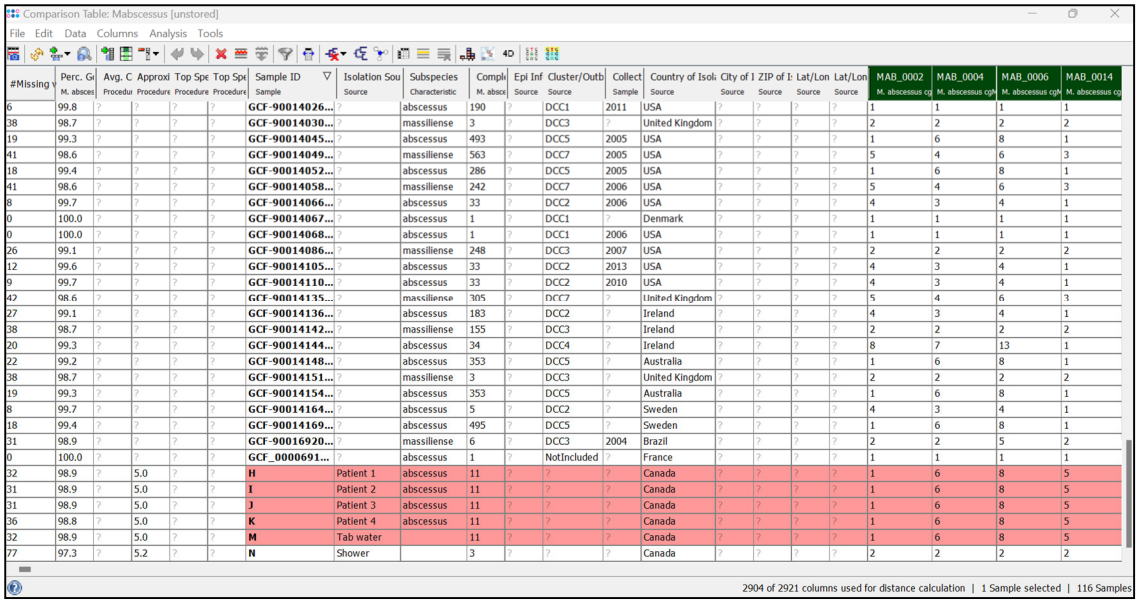
5. Left **click** on the header of column "**Sample ID**" and scroll to the bottom.
6. You can see that sample L only has 0.1% good cgMLST targets (2nd column Perc. Good Targets). This is indeed not a *M. abscessus* strain but a *M. intracellulare* strain, for which the scheme is not suitable. All other isolates have a percentage of good cgMLST targets way above 95% indicating that the data are of sufficient quality and can be used for further analysis.



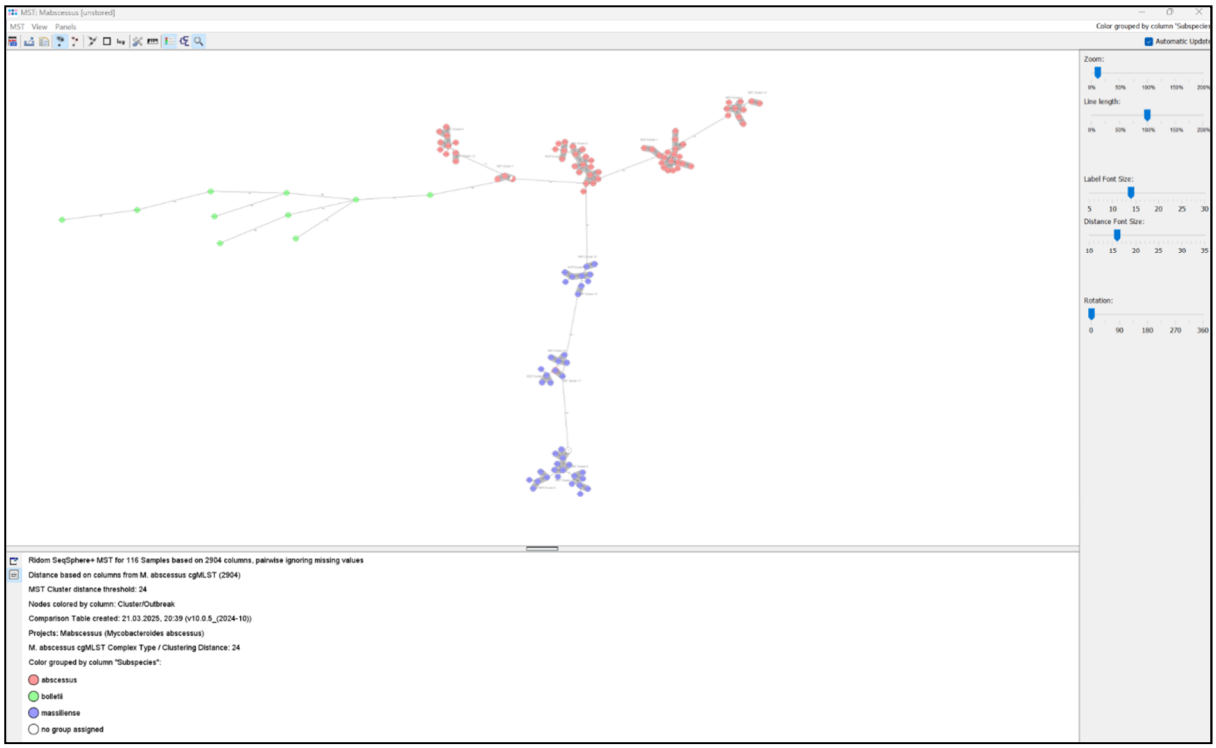
#Missing v	Perc. Gc	Approx	Sample ID	ST	Comple	Cluster/	Collecti	Country of Isol	CC	argH	cya	gnd
	M. absce	Assemb	Sample	M. absc	M. absc	Source	Sample	Source	M. absc	M. abscessus M	M. abscessus M	M. abscessus M
45	98.7	?	GCF-90014151...	? (n/a)	3	DCC3a	?	United Kingdom	? (n/a)	? (n/a)	? (n/a)	? (n/a)
26	99.3	?	GCF-90014154...	? (n/a)	353	DCC5	?	Australia	? (n/a)	? (n/a)	? (n/a)	? (n/a)
15	99.7	?	GCF-90014164...	? (n/a)	5	DCC2	?	Sweden	? (n/a)	? (n/a)	? (n/a)	? (n/a)
25	99.4	?	GCF-90014169...	? (n/a)	495	DCC5	?	Sweden	? (n/a)	? (n/a)	? (n/a)	? (n/a)
38	98.9	?	GCF-90016920...	? (n/a)	6	DCC3a	2004	Brazil	? (n/a)	? (n/a)	? (n/a)	? (n/a)
7	100.0	?	GCF_00006918...	? (n/a)	1	DCC1	?	France	? (n/a)	? (n/a)	? (n/a)	? (n/a)
32	98.9	5.0	H	222	11	?	?	Canada	?	1	2	3
31	98.9	5.0	I	222	11	?	?	Canada	?	1	2	3
31	98.9	5.0	J	222	11	?	?	Canada	?	1	2	3
36	98.8	5.0	K	222	11	?	?	Canada	?	1	2	3
2909	0.1	6.4	L	?	?	?	?	Canada	?	? (not found)	? (not found)	? (not found)
32	98.9	5.0	M	222	11	?	?	Canada	?	1	2	3
77	97.3	5.2	N	33	3	?	?	Canada	3	30	31	17

7. Remove sample L by **right clicking** on the row and selecting "**Remove selected Samples From Table**". Click OK.
8. Note that H, I, J, and K all belong to Complex Type (CT) 11, while sample N belongs to CT 3. Note that complex types group together isolates with very similar cgMLST profiles based on a species-specific threshold (more info: https://www.ridom.de/u/Core_Genome_MLST_Complex_Type.html). In case of *M. abscessus*, isolates are classified into a CT if the distance is less than 25 alleles compared to the CT founder profile. If strains belong to the same CT, they might be epidemiologically related. You could copy the CT´s of the *6 M. abscessus* strains in Task_Exercise2026_NTM.xlsx. In addition, you can compare the ST types (222 and 3) with the ones from pubMLST. These should be the same as Ridom Typer is using the same nomenclature as pubMLST.
9. Click at the top bar on "**Columns**" > **Add additional database fields as columns**. Click on + sign next to "**Epi Source**" and check the box next to "**Isolation Source**". Also click on the + sign next to "**Epi characteristic**" and select "**Subspecies**". Click two times on "**OK**".

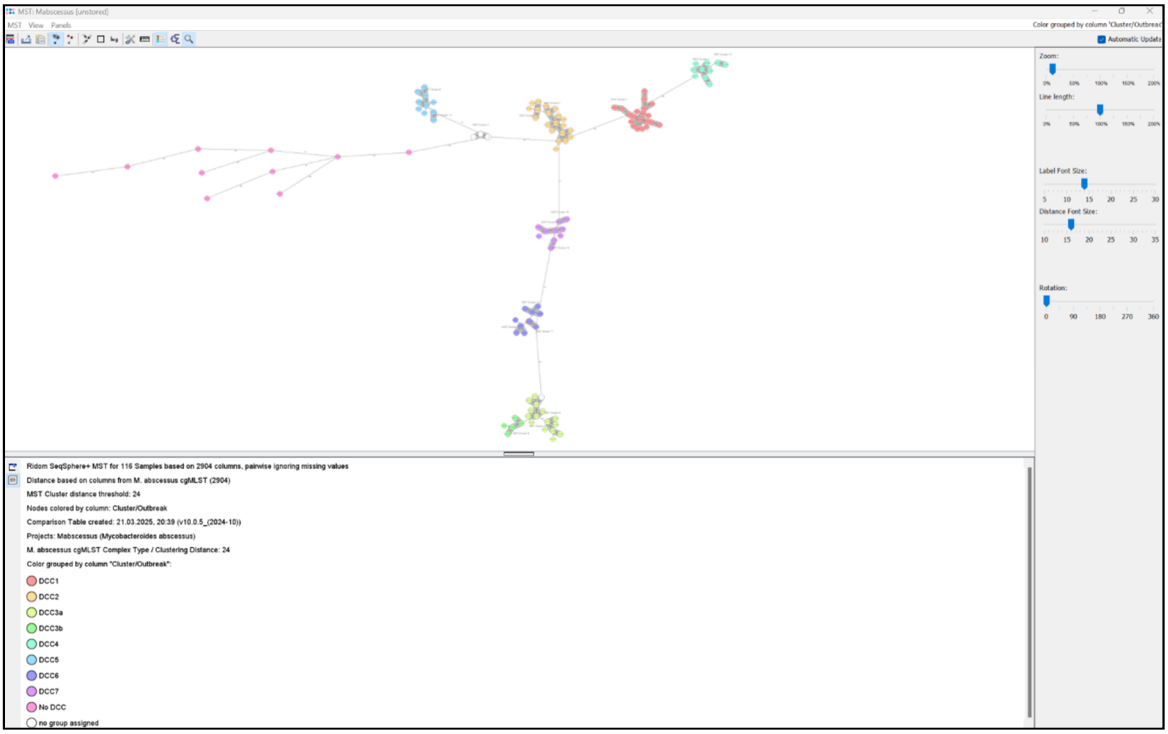




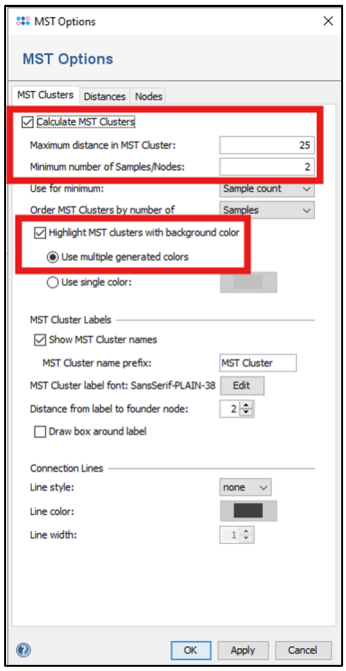
10. Right click on the "**Subspecies**" header and choose "**set color groups by column values**". Click on "**OK**".
11. Click on "**Analysis**" in top bar of the comparison window and click "**Minimum spanning tree**" (MST). Click "**OK**".
12. Play around with the parameters at the right of the window (Zoom, line length, rotation,…)
13. Based on the phylogenetic positioning, to which subspecies does sample M and N belong? I.e. to which of the known subspecies is your sample most closely related. Complete this info in **Task_Exercise2026_NTM.xlsx** if you have not already done so**.**



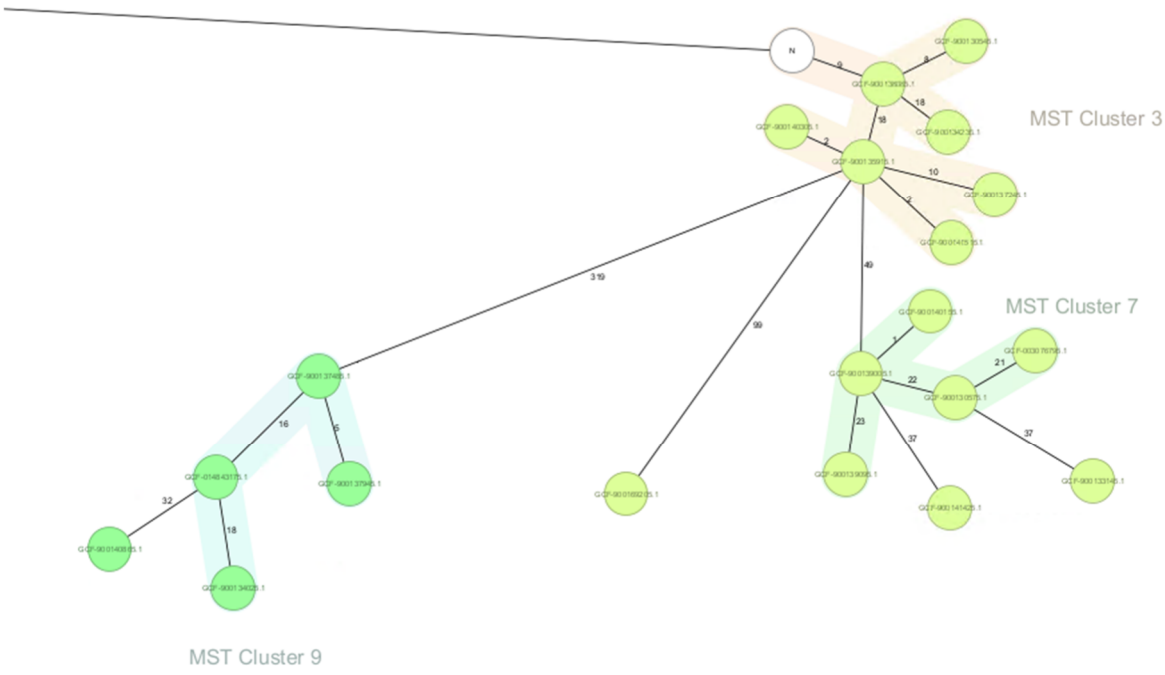
14. Go back to the **Comparison Table** window. Right click on the "**Cluster/Outbreak**" header and choose "**set color groups by column values**". Click on **"OK**". Have another look at the tree.



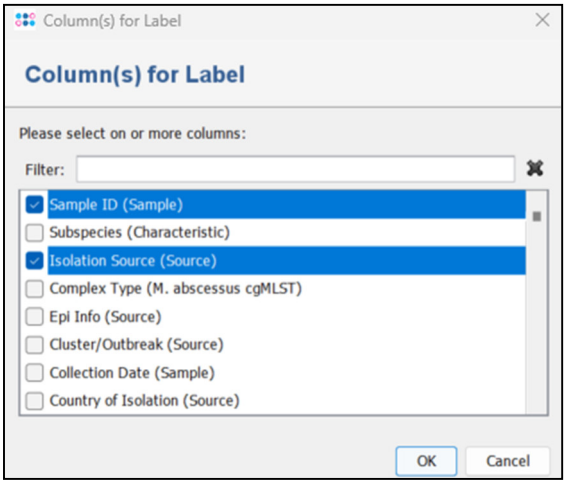
15. To which DCC do the 7 *M. abscessus* strains belong? Note that DCCs are defined with a threshold of 250 alleles (Diricks et al. 2022)[5]. Are there any isolates with a distance less than 250 alleles to a known DCC strain? Complete this info in **Task_Exercise2026_NTM.xlsx** if you have not already done so.
16. In the MST window, click on "**View**" > **MST options**. Check box next to "**Calculate MST clusters**" and choose 25 as "**Maximum distance in MST cluster**". Choose "**Use multiple generated colors**". Click on "Apply" and then on "**OK**". You will see now that the different clusters are indicated. Each of the isolates in a cluster can be connected to another cluster member with less than 25 alleles. As mentioned before, also a more strict threshold can be used here to define potential transmission clusters, such as 10 alleles[5].

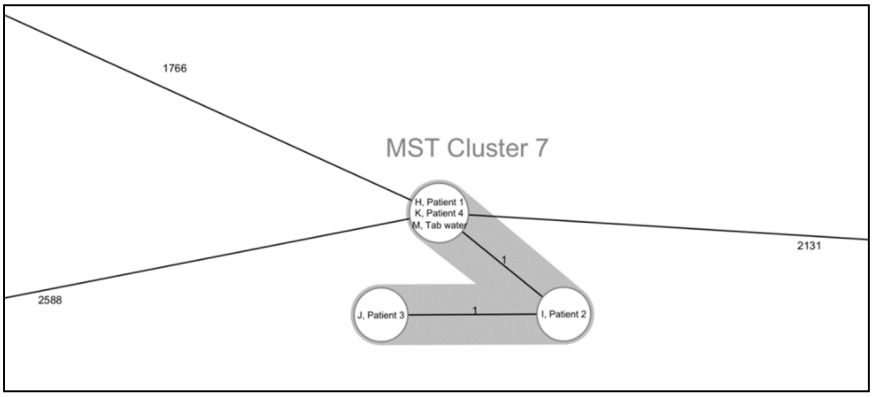


17. Have a **closer look** at sample N (zoom in). It is closely related (9 alleles distance) to one of the reference samples (DCC3a strain). Note that within a DCC, you can have multiple subclusters of more closely related isolates.



18. Click on "**View**" > **Show columns** for Label. Check box next to "**isolation source**". Click 2 times on „**OK**".





19. Based on this analysis, are the patient samples and water samples part of the outbreak? complete this info in **Task_Exercise2026_NTM.xlsx**. Compare the genetic distances with those from pubMLST (grapetree analysis).
20. What is the source of the outbreak? Describe in detail how transmission could have occured.

Optional: Only for participants with command-line experience

3.2 Command-line Tools

3.2.1 NTM-Profiler

1. Install NTM-profiler using the instructions at the github website: https://github.com/jodyphelan/NTM-Profiler
2. Run NTM-Profiler using the instructions at the github website

3.2.2 NTMseq

Note: a demonstration of this pipeline will be given next session (11th of May 2026)

1. Install NTMseq using the instructions at the github website: https://github.com/ngs-fzb/NTMseq
2. Run NTMseq using the instructions at the github website

4 References

1. Wuzinski, M. *et al.* Investigation of Two Mycobacterium abscessus Outbreaks in Quebec Using Whole Genome Sequencing. *BioMed Research International* **2020**, (2020).
2. Bryant, J. M. *et al.* Emergence and spread of a humantransmissible multidrug-resistant nontuberculous mycobacterium. *Science* **354**, 751–757 (2016).
3. Bryant, J. M. *et al.* Whole-genome sequencing to identify transmission of Mycobacterium abscessus between patients with cystic fibrosis: A retrospective cohort study. *The Lancet* **381**, 1551–1560 (2013).
4. Ruis, C. *et al.* Dissemination of Mycobacterium abscessus via global transmission networks. *Nature Microbiology* **6**, 1279–1288 (2021).
5. Diricks, M. *et al.* Delineating Mycobacterium abscessus population structure and transmission employing high-resolution core genome multilocus sequence typing. **13**, 4936 (2022).
6. Souvorov, A., Agarwala, R. & Lipman, D. J. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biology 2018 19:1* **19**, 1–13 (2018).