

BIONUMERICS Tutorial:

Mycobacterium tuberculosis complex functional genotyping: predicting phenotypic traits from whole genome sequences

1 Aim

In this tutorial we will screen whole genome sequences of *Mycobacterium tuberculosis* complex samples for phenotypic traits such as spoligotype, lineage and antibiotic resistance using the *Mycobacterium tuberculosis complex functional genotyping plugin*. The plugin also allows you to perform species confirmation.

The different steps are illustrated using the whole genome demonstration database of the *My*cobacterium tuberculosis complex. This database is available for download on our website (see 2) and contains 29 publicly available sequence read sets of the *Mycobacterium tuberculosis* complex with already calculated de novo assemblies.

2 Preparing the database

2.1 Introduction to the demonstration database

We provide a **WGS demo database** for the *Mycobacterium tuberculosis* complex containing sequence read set data links for 29 samples, calculated de novo assemblies and wgMLST results (allele calls and quality information).



The wgMLST workflow and results will not be discussed in this tutorial.

The **WGS demo database** for the *Mycobacterium tuberculosis* complex can be downloaded directly from the *BIONUMERICS Startup* window (see 2.2), or restored from the back-up file available on our website (see 2.3).

Installation of the *Mycobacterium tuberculosis complex functional genotyping plugin* is only possible when no spaces are present in the BIONUMERICS home directory and in the name of the database. Before downloading or restoring the **WGS demo database** for the *Mycobacterium tuberculosis* complex, please check if your BIONUMERICS home directory does not contain any spaces:

1. Click the 😟 button, located in the toolbar in the *BIONUMERICS Startup* window and select *Change home directory...* to call the *Home directory* dialog box.

2. In case the currently specified home directory contains spaces, update the path to a path containing no spaces and close the *Home directory* dialog box.

2.2 Option 1: Download demo database from the Startup Screen

3. Click the *button*, located in the toolbar in the *BIONUMERICS Startup* window.

This calls the *Tutorial databases* window (see Figure 1).

🖆 Tutorial databases					- D >	<
File Database Tutorial Window						
r Tutorial databases						
1 T						
Name	Size (MB)	Min	Max	•		
BandScoring demo database	15	7	8	^	The DemoBase Connected demonstration	\sim
DemoBase Connected	4	7.5	8		database in BioNumerics version 8.0 format. This database is used in several tutorials and	
Neisseria MLST demo database	1	7.5	8		comes with the BioNumerics installation.	
SNP demonstration database	0	7.5	8		comes with the Diorumencs installation.	
WGS demo database for Listeria monocytogenes	311	7.5	8			
WGS_demo_database_for_Brucella_spp	214	7.6	8			
WGS_demo_database_for_Escherichia_coli	601	7.5	8			
WGS_demo_database_for_Staphylococcus_aureus	624	7.5	8			
	_			~		\sim
	<			>		
Tutorials						
Ē.						
Name				Ŧ		
Band matching and polymorphism analysis				^	This tutorial illustrates how to calculate a	\sim
Building automated decision and action workflows					Principal Components Analysis (PCA) and a	
Calculating a PCA and an MDS on a fingerprint data set					Multi Dimensional Scaling (MDS) (sometimes also called Principal	
Clustering a phenotypic test assay					Coordinates Analysis (PCoA)) on a	
Combined analysis of character data					fingerprint data set and how to change the	
Combined analysis of fingerprint data					layout of the obtained plots.	
Configuring the database layout						
Dendrogram layout options						\sim
Entry information fields and their properties				~		

Figure 1: The *Tutorial databases* window, used to download the demonstration database.

- 4. Select WGS_demo_database_for_MTBC from the list and select *Database* > *Download* ().
- 5. Confirm the installation of the database and press < OK > after successful installation of the database.
- 6. Close the *Tutorial databases* window with *File* > *Exit*.

The WGS_demo_database_for_MTBC appears in the BIONUMERICS Startup window.

7. Double-click the WGS_demo_database_for_MTBC in the *BIONUMERICS Startup* window to open the database.

2.3 Option 2: Restore demo database from back-up file

A BIONUMERICS back-up file of the demo database for the *Mycobacterium tuberculosis* complex is also available on our website. This backup can be restored to a functional database in

BIONUMERICS.

8. Download the file WGS_MTBC.bnbk file from https://www.applied-maths.com/download/ sample-data, under 'WGS_demo_database_for_MTBC'.



In contrast to other browsers, some versions of Internet Explorer rename the WGS_MTBC.bnbk database backup file into WGS_MTBC.zip. If this happens, you should manually remove the .zip file extension and replace with .bnbk. A warning will appear ("If you change a file name extension, the file might become unusable."), but you can safely confirm this action. Keep in mind that Windows might not display the .zip file extension if the option "Hide extensions for known file types" is checked in your Windows folder options.

- 9. In the *BIONUMERICS Startup* window, press the button. From the menu that appears, select **Restore database...**.
- 10. Browse for the downloaded file and select *Create copy*. Note that, if *Overwrite* is selected, an existing database will be overwritten.
- 11. Specify a new name for this demonstration database, e.g. "WGS_MTBC_demobase".
- 12. Click < OK > to start restoring the database from the backup file.
- 13. Once the process is complete, click < Yes > to open the database.

The *Main* window is displayed (see Figure 2).

WGS_demo_database_for_MTBC	- BioNumerics							 	- 0
e Edit Database Analysis So	ripts WGS tools Window Help								
Შ 글 ▦ ▱ ♥	<u>⊳</u> ⊳ ®, ®⊗ ®,								
xperiment types		Database entries							Comparisons
			0.0						
ଡ଼ + 🗗 ⊗ 🗞	; Ē₁ ײַ ↑ ↓	\$1 + 🗗 ⊗		<all entries=""></all>	ני ט				+ 🖻 ⊗ 🔩 🛍
# Name	Туре 💌	Key	Modified date			Sample 🔫	1		Name Modified date
] ≹ 1 wgs	Sequence read set types	ERR067612	2020-08-12 16:10:00	ERR067612	Mycobacterium tuberculosis	0344A	•	 •	
2 wgsLong	Sequence read set types	ERR551067	2020-08-12 15:25:14	ERR551067	Mycobacterium tuberculosis complex sp.	10349-03		 •	
3 wgMLST	Character types	ERR551511	2020-08-12 15:25:14	ERR551511	Mycobacterium tuberculosis complex sp.	10348-03		 •	
4 denovo	Sequence types	ERR551847	2020-08-12 16:10:00	ERR551847	Mycobacterium tuberculosis complex sp.	3861-03		 •	
5 quality	Character types	ERR553130	2020-08-12 16:10:00	ERR553130	Mycobacterium tuberculosis complex sp.	1322-04		 •	
3 ≥ 6 wgs_TrimmedStats	Sequence read set types	ERR551192	2020-08-12 16:10:00	ERR551192	Mycobacterium tuberculosis complex sp.	3859-03		 •	
7 wgMLST_CalTypes	Character types	ERR551568	2020-08-12 16:10:00	ERR551568	Mycobacterium tuberculosis complex sp.	3736-04	•		
		ERR553211	2020-08-12 16:10:00	ERR553211	Mycobacterium tuberculosis complex sp.	10517-03	•	•	
	~	ERR552921	2020-08-12 16:10:00	ERR552921	Mycobacterium tuberculosis complex sp.	4779-04	•		<
<	>	ERR552281	2020-08-12 16:10:00	ERR552281	Mycobacterium tuberculosis complex sp.	7538-03		 •	Identification projects Decision networks
		ERR553386	2020-08-12 16:10:01	ERR553386	Mycobacterium tuberculosis complex sp.	5472-03	•	•	
y fields Database design		ERR551118	2020-08-12 16:10:00	ERR551118	Mycobacterium tuberculosis complex sp.	8867-03		•	29 + 🗗 ⊗ 🖳 । 🖣
- P3 ⊗ B. B.		SRR671775	2020-08-12 16:10:01	SRR671775	Mycobacterium tuberculosis	BGI_10		•	Name Modified date
		ERR553228	2020-08-12 16:10:01	ERR553228	Mycobacterium tuberculosis complex sp.	2509-04		 •	
	ield type 🔍 🔻	ERR551998	2020-08-12 16:10:01	ERR551998	Mycobacterium tuberculosis complex sp.	5248-04	•		
	ixed	ERR551807	2020-08-12 16:10:01	ERR551807	Mycobacterium tuberculosis complex sp.	10350-03	•		
	ixed	SRR671874	2020-08-12 16:10:01	SRR671874	Mycobacterium tuberculosis	BGI_XJ06090	•		
	ixed	SRR671757	2020-08-12 16:10:01	SRR671757	Mycobacterium tuberculosis	BGL_FJ05189	•		
AMC Core (v2) ST F	lexible	SRR671751	2020-08-12 16:10:01	SRR671751	Mycobacterium tuberculosis	BGI_FJ05132	•		
		SRR671740	2020-08-12 16:10:01	SRR671740	Mycobacterium tuberculosis	BGI_0-xz06008	•		
		SRR671750	2020-08-12 16:10:01	SRR671750	Mycobacterium tuberculosis	BGI_FJ05121	•	•	
		SRR671720	2020-08-12 16:10:00	SRR671720	Mycobacterium tuberculosis	BG[_322			<
erprint files Power assemblies		SRR671719	2020-08-12 16:10:01	SRR671719	Mycobacterium tuberculosis	BGL_60		•	Alignments BLAST projects Chrom. Co.
	Annotatons	SRR671762 SRR671741	2020-08-12 16:10:01 2020-08-12 16:10:01	SRR671762 SRR671741	Mycobacterium tuberculosis	BG[_FJ05060	1		
∃ + 🗗 ⊗ 🗟 ।	All Fingerprint f	ERR552935	2020-08-12 16:10:01 2020-08-12 16:10:01	SRR671741 ERR552935	Mycobacterium tuberculosis	BGLP-xz09119 5685-04			+ 12 ⊗ 8, 61
File name Exc	periment type Link 🗸	ERR550665	2020-08-12 16:10:01 2020-08-12 16:10:00	ERR552935 ERR550665	Mycobacterium tuberculosis complex sp.	1599-04			Name Modified date
	•	ERR551070	2020-08-12 16:10:00	ERR550665	Mycobacterium tuberculosis complex sp. Mycobacterium tuberculosis complex sp.	4139-04			
		ERR552454	2020-08-12 15:25:14	ERR552454	Mycobacterium tuberculosis complex sp. Mycobacterium tuberculosis complex sp.	1597-04	1.1		
		ERN552454	2020-08-12 15:25:14	ERR552454	Mycobacterium tuberculosis complex sp.	1597-04			
		-							
<	>	<				>			<

Figure 2: The MTBC demonstration database: the Main window.

3 About the demonstration database

The WGS demo database contains links to sequence read set data on NCBI's sequence read archive (SRA) for 29 publicly available sequencing runs. Additional information (in entry info fields Sample and Organism) was collected from NCBI and added to the demonstration database.

Seven experiments are present in the demo database and are listed in the *Experiment types* panel (see Figure 3).

Experim	ent ty	pes		
କ୍ଷ	-	+ 🗗 ⊗ 🕵	Call Experiment types>	
		Name	Туре	•
≣ ≨	1	wgs	Sequence read set types	
≣	2	wgsLong	Sequence read set types	
	3	wgMLST	Character types	
🗆 🛟	4	denovo	Sequence types	
	5	quality	Character types	
≣	6	wgs_TrimmedStats	Sequence read set types	
	7	wgMLST_CallTypes	Character types	~

Figure 3: The Experiment types panel in the Main window.

1. Click on the green colored dot for one of the entries in the first column in the *Experiment presence* panel. Column 1 corresponds to the first experiment type listed in the *Experiment types* panel, which is **wgs** in the default configuration.

In the *Sequence read set experiment* window, the link to the sequence read set data on NCBI (SRA) with a summary of the characteristics of the sequence read set is displayed: *Read set size*, *Sequence length statistics*, *Quality statistics*, *Base statistics* (see Figure 4).

- 2. Close the Sequence read set experiment window.
- 3. Click on the green colored dot for one of the entries in the fourth column in the *Experiment presence* panel. Column 4 corresponds to the fourth experiment type listed in the *Experiment types* panel, which is **denovo** in the default configuration.

The *Sequence editor* window opens, containing the results from the de novo assembly algorithm, i.e. concatenated de novo contig sequences (see Figure 5).

4. Close the Sequence editor window.

The sequence read set experiment type **wgs_TrimmedStats** contains some data statistics about the reads retained after trimming, used for the de novo assembly.

The sequence read set experiment type **wgsLong** contains the links to long read sequence read data (typically PacBio or MinION datasets). In this demo database, no links are defined for this experiment.

The other three experiments contain data related to the wgMLST analysis performed on the samples:

- Character experiment type wgMLST contains the allele calls for detected loci in each sample, where the consensus from assembly-based and assembly-free calling resulted in a single allele ID.
- Character experiment type **quality** contains quality statistics for the raw data, the de novo assembly and the different allele identification algorithms.
- Character experiment type wgMLST_CallTypes: contains details on the call types.

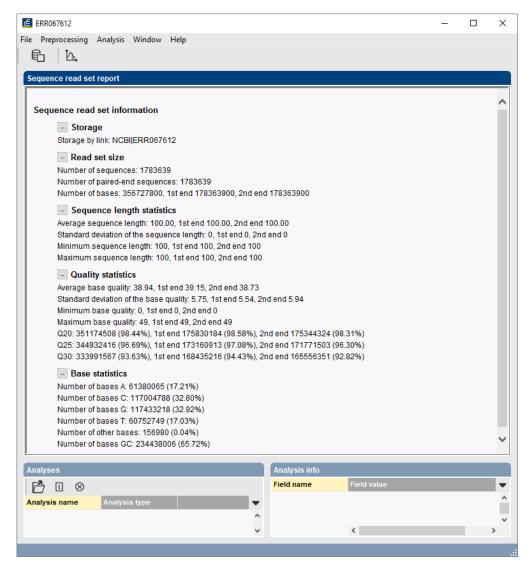


Figure 4: The sequence read set experiment card for an entry.

4 Installing the MTBC functional genotyping plugin

To store the results of your genotyping jobs, you will have to create some new information fields and character/sequence experiments.

- 1. Make sure the *Database entries* panel is the active panel in the *Main* window and select *Edit* > *Information fields* > *Add information field...*.
- Enter a name, e.g. Species confirmation and choose Space next to Optimize storage for (see Figure 6). The latter is important because otherwise results can get truncated if they contain too much characters. All other settings can be left default.
- 3. Repeat previous step for following information fields. Make sure you optimize the storage for *Space* each time:
- Unknown genotype
- Resistance summary
- · Spoligotype (octal)

	2	
	-	

			ce Viewer)											-		
Se	quence	Heade	er Annotati	on View	v Tool	Wind	low Hel	0								
11	[]	fi	00	× X	ទា (3 ⊗	01			-2 -2						
	<u>~~</u>	ЕП	<u>v / (</u>	<u>*</u> %	ĽL	<u> </u>	\simeq			~~ ~	~ ~					
que	ence Edi	itor														
cc	GCCG	CA T	GGAACA	CT G	TGAG	CCAG	GGAGI	ATTCG	AGCCCA	AGGC	AACGCCGGCC	CAGACGCAGG	CATTGGAACG	11760		
											CIGCGGCGCT					
											GCCGAACAGA					
									TGTGGG.							
											ACGACCCGGC					
											GACCAGCAGG					
											GTGGCACCGT					
											CGACGGTGCC					
											TGGCCGCGCT					
											GACGGTGACG					
											GGATTCCCGG					
	ence Vie 단민 (5000 B	MTBC_15	10 37 MT	1000 	'1€ -мтв			0000 '25	000 '30		000 MTBMTBC_1664	40000	•
				'5000' ⊐ ₿∳Ç	MTBC_15	110 	1000	ЧЕ Д МТВ	NTEC 1	· · · /2	0000 22				40000	
D			MTBM	₽₩	MTBC_15	37 MT	. мтви	/1€ ∦ <mark></mark> ∕р		'''''''''			‱ '35(⊐CX ≰		40000 * *	
D	ation ure list	€	MTB M		MTBC_15	37 MT	I.	₩TB	8000	··· '2		077212190	‱ 'ss: ⊐CX ≰⊐		40000	
D	ation ure list	€			tart	⊗ Enc	I I	Length	NTEC.1		+ & /allel="1	077212190	000 '350 ₽CX ¥X (188)		40000	
D	ation ure list	e	ature key ♦ CDS ♦ CDS			37 MT	<u>Т</u> 94	∜ <u></u> ∕	¢.		the set of	077212190 "TBC_1539"	000 '35(+		40000 ****	
D	ation ure list	€ Fea ¢	♦ CDS ♦ CDS		tart 672	⊗ Enc 106	<mark>Т</mark> 94 90	Length 1023	мтро_1 (мт. %GC 64.29		the last the	077212190 "TBC_1539" 00.0		MTBMTBC_1584	40000 ****	
nota ieatu	ation ure list		🔶 CDS	B	tart 672 0772	⊗ Enc 106 121		Length 1023 1419	*GC 64.29 67.28		the second secon	077212190 " " " " " " " " " " " " " " " " " " "	000 '350	MTBMTBC_1564		
nota Featu	ation ure list		 CDS CDS CDS CDS 	B	tart 672 0772 2196	⊗ Enc 106 121 130	. MTB	Length 1023 1419 900	*GC 64.29 67.28 67.07		the set of	077212190 "ITBC_1539" 00.0 Wd=1,start=1077 "STGEVMGDLPFGFSS	1:stop=12190;cid gdpPpeDPSGRKRGK	MTBMTBC_1664	AFGIGG SIGFIA	
nota	ation ure list		 CDS CDS CDS CDS CDS CDS 	S S S S S S S S S S S S S S S S S S S	tart 672 0772 2196 3104 3432	⊗ Enc 106 121 130 133	<mark>Т</mark> 94 90 95 04 75	Length 1023 1419 900 201	%GC 64.29 67.28 67.07 60.50		the second secon	077212190 " " " " " " " " " " " " " " " " " " "	1; stop=12190; cid GDDPPEDPSGRDKRGK MFGGVGTAMAAGKTSG ADTMLDGATSLPAGAT	ArtBWTBC_1564	AFGIGG SIGFIA ILATWK	
	ation ure list # 11 \$ 12 13 @= 14 15 6 16		 CDS CDS CDS CDS CDS 	S S S S S S S S S S S S S S S S S S S	tart 672 0772 2196 3104	⊗ Enc 106 121 130 133 147	<mark>Т</mark> 94 90 95 04 75 87	Length 1023 1419 900 201 1344	*GC 64.29 67.28 67.07 60.50 64.18		<pre>the control of the control of t</pre>	077212190 "TBC_1539" 00.0 Wd=1;start=1077 "STGEVMGDLFGFSS FINMADLGQ1FTRLGG "IPAATNSAIADAVHI LCDPMAQQ1STVWAS	1;stop=12190;cid GDDPPEDPSGRDKRGK MFGGVGTAMAAGKTSG ADTWLDGATSLPSGAL	MTBMTBC_1564	AFGIGG SIGFIA LLATWK SQALGR	
nota	ation ure list		 CDS CDS CDS CDS CDS CDS CDS CDS CDS 	S S S S S S S S S S S S S S S S S S S	tart 672 0772 2196 3104 3432 4809	⊗ Enc 106 121 130 133 147 150	Ц мтв. м. 94 90 95 004 75 87 19	Length 1023 1419 900 201 1344 279	%GC 64.29 67.28 67.07 60.50 64.18 62.59 67.46		the search of th	077212190 "IBC_1539" 00.0 wd=1;start=1077 STGEVMGDLFPGF55 STRNADLG01FTRLOE "IPAATNSATADAVHL LCDPMAQ01STVWASS SREVLTSTDIGLPLG	1; stop=12190; cid GDDPPEDPSGRDKRGK MFGGVGTAMAAGKTSG ADTMLDGATSLPAGAT	HEL.MTBC_1564	AFGIGG SIGFIA FLATKK SQALGR LATREA	
nota	ation ure list		 CDS CDS CDS CDS CDS CDS CDS 	S S S S S S S S S S S S S S S S S S S	tart 672 0772 2196 3104 3432 4809 5517	⊗ Enc 106 121 130 133 147 150 176	Ц мтв. м. 94 90 95 004 75 87 19	Length 1023 1419 900 201 1344 279 2103	%GC 6429 67.28 67.07 60.50 64.18 62.59 67.46 58.27	•	tec 1445 Nec 146 /allel="1 /locus_tag="M /evidence=1 /hote="f /translation="V R L A	077212190 " " " " " " " " " " " " " " " " " " "	1; stop=12190; cid GDDPPEDPSGRDKRGK ADTWLDGATSLPAGAT SLPEEAKSMAGPLLSI FRGVATLPGAVESPA	HEdenovo_0" DGADSGSGANPLGJ FVNTDLARVASSF FVNTDLARVASSF LAVGASFTOMVINI LISQMGGIAFGSOLG LAGLEQPRSEILTFI	AFGIGG SIGFIA LLATWK SQALGR LATREA LADPAA	

Figure 5: The Sequence editor window.

Add information	n field			?	×
Name : Spe	cies confirmatio	n ID			
Origin					
Store field	d content in data	ibase			
 Calculate 	field content fro	om other fields			
Ontinen					
Options Create a	Text field				
create a	Text field	~			
Optimize stor	Spee	d	character	rs)	
	Spac	е		Canc	el

Figure 6: Adding a new information field with the *Space* option.

- Lineage number
- Lineage name

Proceed as follows to install the *MTBC genotyping plugin*:

- 4. Call the *Plugins* dialog box from the *Main* window with *File* > *Install / remove plugins...* (,).
- 5. Select *MTBC functional genotyping* from the list in the *Applications tab* and press the <*Activate*> button.
- 6. Confirm the installation of the plugin.

7. In the *General tab* of the wizard (see Figure 7) choose wgs as WGS experiment type and the *Info fields* that will appear in the report. The Key and the Resistance/Lineage results are default shown in the report. Check for example the Species confirmation and Spoligotype (octal) information field or any other field that was added. Click <Next>.

ttings			?	>
General Set general settings.				
Export				
WGS experiment type	wgs	\sim		
Info fields to show in report	Core (v2) ST Species confirmation Unknown genotype Resistance summary Spoligotype (octal) Lineage number Lineage number <lineage <loaded="" entry="" fields="" name=""> <all entry="" fields=""> <all entry="" fields=""> <all entry="" fields=""> <all entry="" fields=""></all></all></all></all></lineage>	~		

Figure 7: The General tab.

- In the Species tab of the wizard choose the Species field for the storage of the species confirmation results, change the detection parameters if wanted and click < Next> (see Figure 8).
- In the *Lineage tab* choose the fields for the storage of the *Lineage number* and *Lineage name*, change the detection parameters if wanted and click < *Next*> (see Figure 9).
- In the Spoligotyping tab select Create New as Spoligo presence/absence experiment type. Choose the field for the storage of the Octal spoligotype. Modify the detection parameters if wanted and click <Next> (see Figure 10).
- 11. If *Create new* was selected in the previous step, a dialog box appears prompting for the *Spolig-otype absence/present* experiment type name. Enter e.g. *Spoligotype* and click <*OK*>.
- 12. In the *Resistance tab* (see Figure 11), select the *Resistance database*. Currently, only one database is available: *Curated database (version 6.0)*.
- 13. Select *Create new* next to the three *Resistance results* experiment types (see Figure 11).
- 14. Choose the *Resistance summary* information field for the storage of the antibiotics for which known resistance-related mutations (or indels) were found.
- 15. Choose the **Unknown genotype** information field for the storage of the antibiotics for which unknown mutations or indels (which are not included in the resistance database) were found in known resistance-related genes.

Species Set species determination specific settings. Detection parameters Minimum coverage 5 Output fields
Minimum coverage 5
Output fields
Species field Species confirmation V
Species rold Species commation V
< Back Next > Cancel

Figure 8: The Species tab.

- 16. Check Detect genomic variants if you want to store all the mutations in resistance-related genes in an experiment type. Select Create new or select an existing experiment from the Nucleotide variants and Amino acid variants experiment type lists.
- 17. Click <*Finish*>.
- 18. If *Create new* was selected in the previous step, a dialog box appears prompting for the experiment type name(s). Enter the name(s) and click <*OK*>.

When the *MTBC functional genotyping* installation is complete, you will be prompted to restart the database. The *Plugins* dialog box can be closed by pressing the $\langle Exit \rangle$ button and the database via *File* > *Exit*.

Open the database again from the *BIONUMERICS Startup* window. A *MTBC* menu item is now available in the *Main* window (see Figure 12).

5 Screening of entries

5.1 Submitting jobs

MTBC jobs need to be submitted to the calculation engine. The Calculation engine option requires credits for running jobs on the Applied Maths cloud calculation engine. Credits are linked to credentials that you need to enter when installing the WGS tools plugin. In our demo database, the WGS tools plugin is installed but no credits are assigned to the demo project so no MTBC jobs

Lineage Set lineage determi	nation specific setting	s.		
Detection parameters				
Minimum coverage	5			
Minimum relative cover	age 10.0	%		
Output fields				
Lineage number field	Lineage number		\sim	
Lineage name field	Lineage name		~	

Figure 9: The Lineage tab.

can be performed on the external calculation engine. Please contact Applied Maths to obtain more information.

Once the *MTBC genotyping plugin* is installed and the settings have been specified, an MTBC genotyping job (spoligotyping, resistance/lineage prediction and species prediction) can be submitted.

- 1. Select a single entry in the *Database entries* panel by holding the **Ctrl**-key and left-clicking on the entry. Alternatively, use the **space bar** to select a highlighted entry or click the ballot box next to the entry.
- 2. In order to select a group of entries, hold the Shift-key and click on another entry.
- 3. Select *WGS tools* > *Submit jobs...* () to open the *Submit jobs* dialog box.

From the *Submit jobs* dialog box, one can define which algorithms need to be run on the selected samples and as such, define and launch the related jobs on the calculation engine.

- 4. Select Calculation Engine.
- 5. Check the box next to *MTBC genotyping* (see Figure 13). Note that this option is only available after successful installation of the plugin and closing and reopening of the database after installation.

At this stage, you can also still change the settings by clicking the *Settings* button next to *MTBC genotyping*.

6. Uncheck all other boxes if you do not want to perform any additional analyses (e.g. wgMLST).

Settings				?	×
Spoligotyping Set spoligotyping specific	settings.				
Set spoligotyping specific Detection parameters Minimum total coverage Minimum forward coverage Minimum reverse coverage Spoligotyping results Spoligo presence/absence ex Octal code field	4 2 2	Spoligotype Spoligotype (octal)		y
		< Back	Next >	Car	ncel

Figure 10: The Spoligotyping tab.

Jobs that already have been submitted and have been imported successfully, will not be relaunched for analysis, unless the check box in front of *Re-submit already processed data* in the **Jobs** part is checked.

By default, the *Job overview* window will be opened after submission of the jobs. However, this can be changed by unchecking the option *Open jobs overview window*.

To analyze one sample with the MTBC genotyping tool, you need 2 credits. The number of credits required to run the selected jobs for the selected entries can be consulted at the bottom of the *Submit jobs* dialog box.

7. Click <*OK*> to launch the MTBC genotyping jobs and open the *Job overview* window for the calculation engine (see Figure 14).

In the *Job overview* window you can see the status of the submitted jobs. The *Job overview* window can be opened from the *Main* window with *File* > *Jobs overview...* ().

Finished jobs can be imported with a manual action (*Jobs* > *Get results* ([®])) or through an automatic update: select *File* > *Settings*, check both options and specify an interval (e.g. 10 min).

The job results can also be imported starting from the entry selection in the *Main* window:

9. Make an entry selection in the *Database entries* panel and select *WGS tools* > *Get results* (%).



The job log files are saved in the *Job log* panel of the *Entry* window. Double-click on an entry in the *Database entries* panel to open the *Entry* window and to consult this information.

Set resistance specific settings.			
Resistance database			
Resistance database Curated datab	base (version 6.0) V		
Detection parameters			
Minimum coverage 5			
Minimum relative coverage 10.0	%		
Resistance results			
Mapping experiment type	<create new=""></create>	~	
Antibiotic resistance experiment type	<create new=""></create>	\sim	
Resistance mutations experiment type	<create new=""></create>	~	
Resistance summary field	Resistance summary	~	
Unknown genotype field	Unknown genotype	\sim	
Genomic variants			
Detect genomic variants			
Nucleotide variants experiment type	<create new=""></create>	\sim	
Amino acid variants experiment type	<create new=""></create>	\sim	
Amino acid variants experiment type	<create new=""></create>	~	

Figure 11: The Resistance tab.

🖆 MTBC genotyping - BioNumerics	
File Edit Database Analysis Scripts	s WGS tools MTBC Window Help
☞ 글 ▦ ☞ ♥ ፟፟	
Experiment types	Show report
😔 🕂 🗗 😣 🗟 I	E Settings 🛞
# Name	Туре 🗾 Кеу
□ ≵ 1 wgs	Sequence read set types 🔨

Figure 12: New menu items after installation of the plugin.

Once the results are imported, the corresponding jobs and their underlying data sets are automatically deleted from the calculation engine and as such, from the *Job overview* window.

5.2 Local screening

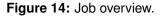
If you want to redo the lineage or resistance prediction of your samples with other settings, you do not need to submit a new job to the CE. This means that no additional credits will be charged:

- 1. Make new information fields to store the new lineage or resistance results (optional step).
- 2. Open the Settings dialog box with MTBC > Settings... in the Main window.
- 3. Select the new information fields and/or create new character experiments to store the new resistance results (optional step).

Submit jobs	? ×
Submit jobs to: Own computer Calculation Engine	
Algorithms	MTBC genotyping Performs lineage, resistance,
MTBC genotyping	spoligotype and species prediction starting from the sequence read sets Settings
Jobs Submitting 1 job for 1 entry.	
Re-submit already processed data	
Open jobs overview window	
Credits available for job submission:	0
Credits needed to submit selected jobs:	2
Buy extra credits	
	OK Cancel

Figure 13: Submitting jobs to the calculation engine.

đ (Overview									- 0	×
File	Jobs Vi	iew Window	/ Help								
×	- 1	°, 61	ບ ↓ໄ Alljobs		8						
Ov	erview of	submitted jo	bs								
	Туре	Name	Submitted time (UTC)	Status	Message	Progress	Job type	Description	User	JobID	
1	Entry	ERR067612	2020-05-13 13:20:36	Running	2020-05-13 13	0%	MTBC genotyping	Performs lineage, resistance,	_DefaultUser_	e0f5e114-1e5a-41b8-9b7e-3b	a 🔨
											~
	<										>
		Leven and in here (
) All running jobs: 1								



- 4. Select other settings.
- Resistance and lineage prediction can be re-analyzed locally by clicking in the *Main* window on *MTBC* > *Analyze* > *Resistance* or *MTBC* > *Analyze* > *Lineage* respectively. If you want to redo both analysis, select *MTBC* > *Analyze* > *All enabled*.



The options *MTBC* > *Analyze* > *Resistance/lineage/all enabled* can only be used when there is already a mapping present for the selected entries (this will only be the case if you have already submitted an MTBC genotyping job for this sample to the CE). Please note that only the mapping generated by the MTBC genotyping plugin can be used and not the standard mapping.



If you did not create/select new experiment types in the settings, results in the current experiment types will be overwritten when you re-analyze your samples.



If you click on MTBC > Analyze > Resistance, and you also selected new, empty fields for lineage results, the lineage results obtained with the original settings will be used. If you also want to analyze the lineage with new settings, you have to click on MTBC > Analyze > Lineage or MTBC > Analyze > All enabled.

5.3 Results

The *Species*, *Lineage*, *Spoligotyping* and *Resistance* results are written to the information fields in the *Database entries* panel (see Figure 15). Please note that the shown names of the information fields can be different in your case depending on whether you choose an alternative name during installation.

An additional information field called *MDR_or_XDR* is generated automatically during the first import of results (see Figure 15). MDR: multidrug resistance is defined as predicted resistance to INH and RMP. XDR: extensively drug resistance is defined as predicted resistance to INH, RMP, (CAP, KAN or AMK) and 1 of the FQ (Moxifloxacin, Ofloxacin, Levofloxacin or fluoroquinolone in general). No: not MDR or XDR.

Dat	abase entries								
] + 🖻		∽ <a∥< th=""><th>Entries></th><th>ម ខ</th><th></th><th></th><th></th><th></th></a∥<>	Entries>	ម ខ				
	Key	Species confirmation	Lineage number	Lineage name	Spoligotype (octal)	Resistance summary	Unknown genotype	MDR_or_XDR	-
	ERR067612	MTBC complex	4/4.3/4.3.3	Euro-American	777777663760771	EMB / PZA / RMP / SM	CAP / FQ / INH / LFX /	No	
	ERR551067	MTBC complex	4 / 4.1	Euro-American	700377777760771	EMB / INH / SM	ETH / FQ / LFX / MFL /	No	

Figure 15: Example output of the Species, Lineage, Spoligotyping and Resistance information fields.

The character experiment types for **Spoligotype**, **Resistance** and **Genomic variants** are created and updated with the predicted traits. Please note that the shown names of the experiment types can be different in your case depending on whether you choose an alternative name during installation.

10. Open a character card for one of the analyzed entries by clicking on the corresponding green colored dot in the *Experiment presence* panel.

Below, the interpretation of the results gathered in the character experiment types is given.

Spoligo presence/*absence experiment type*: character experiment in which the binary code (for each of the 43 spacers: absence or presence) is stored (see Figure 16 and Figure 17).

Antibiotic resistance experiment type: character experiment containing all antibiotics from the selected resistance KB (see Figure 18).

The summary resistance call for an antibiotic can be (see Figure 19):

- **Resistant (R; -5)**: at least 1 known resistance-related mutation or indel from the selected resistance KB was found with sufficient coverage.
- Unknown (U; -3): there are no known resistance-related mutations from the selected resistance KB found with sufficient coverage but there is at least 1 unknown mutation or indel found in a resistance-related gene. For positions already included in the resistance KB, every (non-synonymous) mutation or indel with sufficient coverage (as defined in the settings) leads to the 'unknown' status. For positions in resistance-related genes not yet included in the resistance KB, only majority/consensus mutations with more than 75% relative coverage lead to the 'unknown' status. Lastly, if no bases are covered in resistance related genes for a particular antibiotic, the outcome is also 'unknown'.
- *Failed (F; -2)*: no known or unknown mutations with sufficient coverage were found but also not all positions in the resistance-related genes were covered.
- **Susceptible** (S; +1): there are no known or unknown mutations found and all bases of the resistance-related genes were covered sufficiently.

Character Spacer Seq Enabled Min. Max. Color scale 1 ATAGAGGGTCGCCGGTTCTGGATCA ✓ 0 1 2 CCTCATAATTGGGCGACAGCTTTTG ✓ 0 1 3 CCGTGCTTCCAGTGATCGCCTTCTA ✓ 0 1 4 ACGTCATACGCCGACCAATCATCAG ✓ 0 1 5 TITTCTGACCACTTGTGCGGGATTA ✓ 0 1 6 CGTCGTCTTTTCCGGCGGCTTCC ✓ 0 1 7 GAGGAGAGCGAGTACTCGGGGGTGC ✓ 0 1 8 CGTGAAACCGCCCCCAGCCTGCCG ✓ 0 1 9 ACTCGGAATCCCATGTGCTGACAGC ✓ 0 1	
2 CCTCATAATTGGGCGACAGCTITTG 0 1 3 CCGTGCTTCCAGTGATCGCCTTCTA 0 1 4 ACGTCATACGCCGACCAATCATCAG 0 1 5 TITTCTGACCACTTGTGCGGGGATTA 0 1 6 CGTCGTCATTTCCGGCTTCAATTTC 0 1 7 GAGGAGAGCGAGTATCATCGGGGGCTGC 0 1 8 CGTGAAACCGCCCCCAGCCTGCCG 0 1 9 ACTCGGAATCCCATGTGCTGACAGC 0 1	
2 CCGTGCTICCAGTGATCGCCTTCTA 0 1 3 CCGTGCTTCCAGTGATCGCCTTCTA 0 1 4 ACGTCATACGCCGACCAATCATCAG 0 1 5 TTTTCTGACCACTTGTGCGGGATTA 0 1 6 CGTCGTCATTTCCGGCTTCAATTTC 0 1 7 GAGGAGAGCGAGTACTCGGGGGCTGC 0 1 8 CGTGAAACCGCCCCCAGCCTGCCG 0 1 9 ACTCGGAATCCCATGTGCTGACAGC 0 1	
4 ACGTCATACGCCGACCAATCATCAG 0 1 5 TITTCTGACCACTTGTGCGGGGATTA 0 1 6 CGTCGTCATTTCCGGCGGCTTCAATTTC 0 1 7 GAGGAGAGCGAGTACTCGGGGGCTGC 0 1 8 CGTGAAACCGCCCCAGCCTGGCGG 0 1 9 ACTCGGAATCCCATGTGCTGACAGC 0 1	
4 ACGICALACGGCGACAALAALAG 0 1 5 THITCTGACCACTTGTGCGGGGATTA 0 1 6 CGTCGATTTCCGGCGTCAATTC 0 1 7 GAGGAGAGCGAGTACTCGGGGCTGC 0 1 8 CGTGAAACCGCCCCAGCCTGCCG 0 1 9 ACTCGGAATCCCATGTGCTGACAGC 0 1	
6 CGTCGTCATTTCCAGTTTC 0 1 7 GAGGAGAGCGAGTACTCGGGGCTGC 0 1 8 CGTGAAACCGCCCCCAGCCTGCCG 0 1 9 ACTCGGAATCCCATGTGCTGACAGC 0 1	
7 GAGGAGAGCGAGTACTCGGGGGCTGC ✓ 0 1 8 CGTGAAACCGCCCCCAGCCTGCCG ✓ 0 1 9 ACTCGGAATCCCATGTGCTGACAGC ✓ 0 1	1 1 1
8 CGTGAAACCGCCCCCAGCCTCGCCG ✓ 0 1 9 ACTCGGAATCCCATGTGCTGACAGC ✓ 0 1	1
9 ACTCGGAATCCCATGTGCTGACAGC 🗸 0 1	1
10 TCGACACCCGCTCTAGTTGACTTCC V 0 1	
mparison settings	_

Figure 16: Spoligo presence/absence experiment type. This experiment contains 43 spacer sequences.

Figure 17: Example output of spoligotyping (spoligo presence/absence experiment type) for a sample with spoligotype 0000000003771. White = spacer is absent, black = spacer is present.

	」│↓႞┼╶∔ racters · ⊗ €₅			↑ ·	1	₽	<all chara<="" th=""><th>cters></th><th>U</th><th></th><th></th><th></th></all>	cters>	U			
	Character			Enable	d	Min.		Max.	C	olor scale		
	Rifampicin (RM	P)		~				5	1			
	Moxifloxacin (1	IFL)		~			-	5	1			
	Pyrazinamide (PZA)		×			-	5	1			
	Streptomycin (SM)		 ✓ 			-	5	1			
	Para-aminosali	cylic ac	id (PAS)	 			-	5	1			
	Clofazimine (Cl	_0)		 * 			-	5	1			
	Ofloxacin (OF)	()		 				5	1			
	Ethambutol (EN			×				5	1			
	Fluoroquinolon			 * 				5	1			
	Capreomycin (CAP)		 ✓ 			-	5	1			1
_	acters Mappin parison settir											
	tibiotic resis		•			data a	at (17 abora	etere)				
Ant	ibiotic resistai	ice: nu	merical v	alues, (open	data si	et (17 chara	cters)				

Figure 18: Antibiotic resistance experiment type.

Character	Value	Mapping	-
Rifampicin (RMP)	-5	R	^
Moxifloxacin (MFL)	-3	U	
Pyrazinamide (PZA)	-5	R	
Streptomycin (SM)	-5	R	
Para-aminosalicylic	-3	U	
Clofazimine (CLO)	1	S	
Ofloxacin (OFX)	-3	U	
Ethambutol (EMB)	-5	R	
Fluoroquinolone (FQ)	-3	U	
Capreomycin (CAP)	-3	U	
Levofloxacin (LFX)	-3	U	
Bedaquiline (BED)	1	S	
lsoniazid (INH)	-3	U	
Amikacin (AMK)	1	S	
Kanamycin (KAN)	1	S	
Ethionamide (ETH)	1	S	
Linezolid (LZD)	1	S	
	<	,	

Figure 19: Example output of antibiotic resistance experiment type for sample ERR067612. U = Unknown, R = Resistant, F = Failed, S = Susceptible.

Nucleotide variants experiment type: Character experiment that contains all nucleotide mutations in resistance-related genes that have sufficient coverage. These include (new) mutations which are not in the selected resistance KB (see Figure 20).

+		Ø√ Ø×	↑↓	17 <	All Characters>	ບ		
	Character	Enabled	Min.	Max.	Color scale	LocusName	Gene	•
	2288850	×	0	100		Rv2043c	pncA	
	2155802	×	0	100		Rv1908c	katG	
	775639	×	0	100		Rv0676c	mmpL5	
	759615	×	0	100		(rpoB_promoter)	rpoB_promoter	
	4247730	 	0	100		Rv3795	embB	
	4242643	×	0	100		Rv3793	embC	
	1917972	 	0	100		Rv1694	tlyA	
	2518919	×	0	100		Rv2245	kasA	
	3073868	 Image: A second s	0	100		Rv2764c	thyA	
	764995	 Image: A set of the set of the	0	100		Rv0668	rpoC	
	761101	 Image: A second s	0	100		Rv0667	rpoB	
	762310	 Image: A set of the set of the	0	100		Rv0667	rpoB	
	576744	 Image: A second s	0	100		Rv0486	mshA	
	1834836	 Image: A set of the set of the	0	100		Rv1630	rpsA	
	<							>

Figure 20: Nucleotide variants experiment type. Character: genomic position (H37Rv numbering).

The resistance call for a resistance-related position can be (Figure 21):

- A/C/T/G: nucleotide substitution.
- Multiple variants: there are multiple bases with sufficient coverage.

- Insertion: an insertion is present at this position.
- - : a deletion is present at this position.

Character	Value	Ma	pping	•
2288850	200	0 Ins	ertion	^
2155802		5 T		
775639		4 G		
759615		3 C		
4247730		2 A		
4242643		5 T		
1917972		4 G		
2518919		2 A		
3073868		4 G		
764995		4 G		
761101		3 C		
762310		4 G		
576744		4 G		
1834836		3 C		
4408156		4 G		
4408102		3 C		
4407720		4 G		Y
2726350	<		>	

Figure 21: Example output of nucleotide variants experiment type for sample ERR067612.

Amino acid variants experiment type: Character experiment that contains all amino acid mutations in resistance-related genes that have sufficient coverage. These include (new) mutations which are not in the selected resistance KB.

+	Character	⊡√ ⊠X Enabled	↑ Min.	†	JL <a⊪c Max.</a⊪c 	haracters>	ひ LocusName	Gene	
_	2288850		winti.	0	100	color scale	Rv2043c		
-	2288850			0	100		Rv2043c Rv1908c	pncA katG	
_	2155802 775639			0	100		RV1908C	mmpL5	
_	759615	- ÷		0	100		(rpoB_promoter)	rpoB promoter	
_	4247730	- ÷		0	100		Rv3795	embB	
	4242643	×		0	100		Rv3793	embC	
_	1917972	~		0	100		Rv1694	tlyA	
	2518919	×		0	100		Rv2245	kasA	
_	3073868	×		ő	100		Rv2764c	thyA	
_	764995	×		ŏ	100		Rv0668	rpoC	
_	761101	×		0	100		Rv0667	rpoB	
_	762310	×		0	100		Rv0667	rpoB	
	576744	×		0	100		Rv0486	mshA	
	1834836	×		0	100		Rv1630	rpsA	
	<								>

Figure 22: Aminoacid variants experiment type. Character: genomic position (H37Rv numbering).

The resistance call for a resistance-related position can be:

- An amino acid (one letter abbreviation).
- Multiple variants: there are multiple bases with sufficient coverage.
- · Insertion: an insertion is present at this position.
- - : a deletion is present at this position.

Character	Value	Map	oping	•
2288850	20	00 Inse	rtion	^
2155802	1	18 W		
775639	1	20 V		
4247730	1	04 D		
4242643	1	02 R		
1917972	1	11 L		
2518919	1	16 S		
3073868	1	01 A		
764995	1	01 A		
761101	1	15 P		
762310	1	02 R		
576744	1	08 G		
1834836	1	17 T		
4408156	1	02 R		
4408102	1	01 A		
4407720	1	01 A		
2726350	1	11 L		Y
7362	<		>	

Figure 23: Example output of amino acid variants experiment type for sample ERR067612.

11. Close the character card(s) and click on the green colored dot corresponding to the Mapping experiment.

Mapping experiment type: Sequence experiment in which the mapped sequence of your sample will be stored. This sequence has the same length as the H37Rv reference genome (4411532 bp) (see Figure 24).

ERR067612 (Sequence Viewer)	_	×
File Sequence Header Annotation View Tools Window Help		
Sequence Editor		
G KDCCGATG ACCCCGGTTC AGGCTTCACC ACAGTGTGGA ACGCGGTCGT CTCCGAACTT	60	^
AACGGCGACC CTAAGGTTGA CGACGGACCC AGCAGTGATG CTAATCTCAG CGCTCCGCTG	120	
ACCCCTCAGC AAAGGGCTTG GCTCAATCTC GTCCAGCCAT TGACCATCGT CGAGGGGTTT	180	
GCTCTGTTAT CCGTGCCGAG CAGCTTTGTC CAAAACGAAA TCGAGCGCCA TCTGCGGGCC	240	
CCGATTACCG ACGCTCTCAG CCGCCGACTC GGACATCAGA TCCAACTCGG GGTCCGCATC	300	
GCTCCGCCGG CGACCGACGA AGCCGACGAC ACTACCGTGC CGCCTTCCGA AAATCCTGCT	360	
ACCACATCGC CAGACACCAC AACCGACAAC GACGAGATTG ATGACAGCGC TGCGGCACGG	420	
GGCGATAACC AGCACAGTTG GCCAAGTTAC TTCACCGAGC GCCCGCACAA TACCGATTCC	480	
GCTACCGCTG GCGTAACCAG CCTTAACCGT CGCTACACCT TTGATACGTT CGTTATCGGC	540	
GCCTCCAACC GGTTCGCGCA CGCCGCCGCC TTGGCGATCG CAGAAGCACC CGCCCGCGCT	600	
TACAACCCCC TGTTCATCTG GGGCGAGTCC GGTCTCGGCA AGACACACCT GCTACACGCG	660	~

Figure 24: Example output of mapping experiment type for sample ERR067612.

12. Close the Mapping experiment type.

6 Reports

1. To generate a summary report, select the entries of interest and click on *MTBC* > *Show report*.

The *Report* window contains a genotype report for each of the selected entries (see Figure 25).

2. Select another entry in the *Entries* panel to update the results in the *Genotype report* panel.

The creation date of the report (Date), the Key (Name), and information fields checked in the Settings are displayed in the Genotype report panel.

🖆 Report Wind	low			- 0	×
File Entries R	eport Window Help				
Entries	ERR067612 - MTBC functional genotypin	g report - RUO			
Key 🔻	<default report=""></default>	ย			
ERR067612	· · · · · · · · · · · · · · · · · · ·				
ERR551067					
	ERR067612				
	MTBC functional genotyping repo	ort - RUO			
	The contents of this report are fo making. Disclaimer.	r research purpo	ses only and not ir	ntended for clinical decision	
		08/13/20 15:28	:37		
	Name:	ERR067612			
	Spoligotype (octal):	777777663760	771		
	Species confirmation:	MTBC complex			
	Predicted lineage:			-	
	Freutited inteage.	4	Euro- American		
		4.3	Euro-		
			American		
			(LAM)		
		4.3.3	Euro- American		
			(LAM)		
				-	
	Predicted resistance :				
		Ethambutol (EN Streptomycin (S			
		Pyrazinamide (· · · · · · · · · · · · · · · · · · ·		
	Predicted susceptibility:	Amikacin (AMK			
		Bedaquiline (Bl	*		
		Kanamycin (KA Clofazimine (Cl			
		Ethionamide (E			
		Linezolid (LZD)	· · · · · · · · · · · · · · · · · · ·		
	Unknown:		· · ·		
		Para-aminosali Ofloxacin (OFX	cylic acid (PAS)		
		Capreomycin (
		Moxifloxacin (N	IFL)		
		Isoniazid (INH)			-
< >	4	Levofloxacin (L	FAI) F

Figure 25: Example of a functional genotyping report.

- 3. To rerun an analysis from the Report window (with current MTBC settings) for one entry, click on *Report* > *Reanalyze current report* or click on at the top of the report.
- 4. To rerun the analysis for all entries displayed in the *Report window*, click on *Entries* > *Reana*lyze all.
- 5. Click on a hyperlink of one of the predicted traits to display the detailed results in the Genotype report panel.
- 6. Select *File* > *Exit* to close the *Report* window.