



BIONUMERICS Tutorial:

Staphylococcus aureus functional genotyping: predicting phenotypic traits from whole genome sequences

1 Aim

In this tutorial we will screen whole genome sequences of *S. aureus* samples for phenotypic traits such as virulence and antibiotic resistance using the *Staphylococcus aureus functional genotyping plugin*. The plugin also allows you to detect phages and to extract *spa* repeat sequences.

The different steps are illustrated using the whole genome demonstration database of *Staphylococcus aureus*. This database is available for download on our website (see [2](#)) and contains 97 publicly available sequence read sets of *Staphylococcus aureus* with already calculated de novo assemblies.

2 Preparing the database

2.1 Introduction to the demonstration database

We provide a **WGS demo database** for *Staphylococcus aureus* containing sequence read set data links for 97 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 *Staphylococcus aureus* 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 *Staphylococcus aureus 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 *Staphylococcus aureus*, 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. To download the database directly from the *BIONUMERICs* Startup window, click the  button, located in the toolbar in the *BIONUMERICs* Startup window.

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

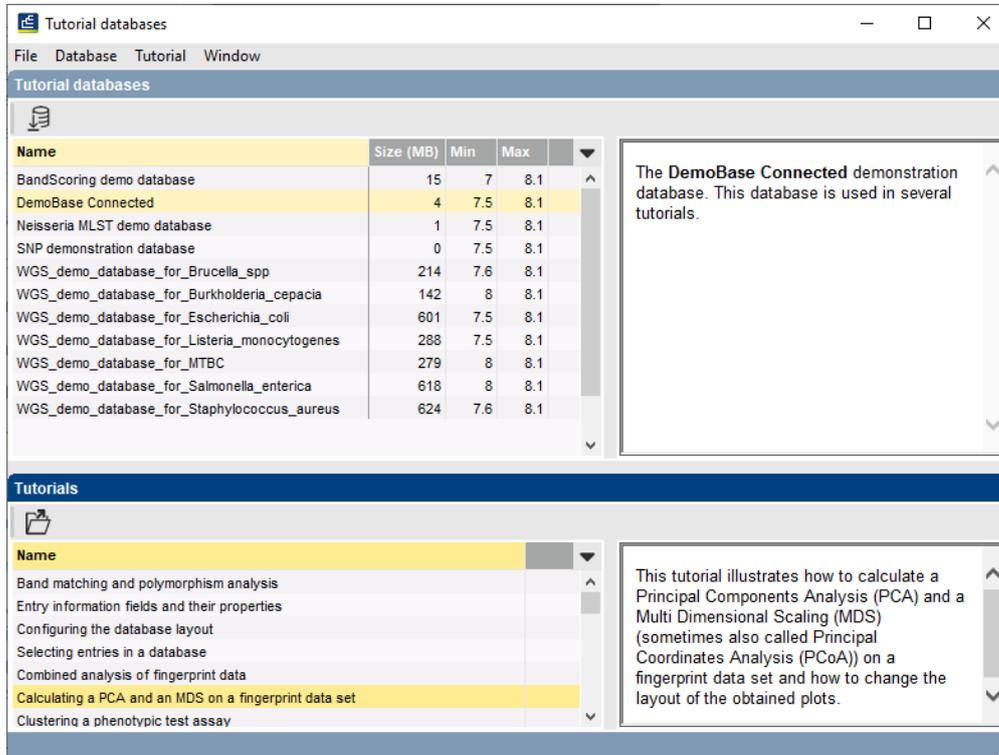


Figure 1: The *Tutorial databases* window, used to download the demonstration database.

4. Select the **WGS_demo_database_for_Staphylococcus_aureus** 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_Staphylococcus_aureus** appears in the *BIONUMERICs* Startup window.

7. Double-click the **WGS_demo_database_for_Staphylococcus_aureus** 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 whole genome demo database for *Staphylococcus aureus* is also available on our website. This backup can be restored to a functional database in BIONUMERICs.

8. Download the file `wgMLST_SAUR.bnbk` file from <https://www.bionumerics.com/download/sample-data>, under 'WGS_demo_database_for_Staphylococcus_aureus'.



In contrast to other browsers, some versions of Internet Explorer rename the wgMLST_SAUR.bnbk database backup file into wgMLST_SAUR.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 and make sure the name does not contain any spaces to ensure the successful installation of the *S. aureus functional genotyping plugin*. Specify for example: "WGS_Saur_demobase".
12. Click <**OK**> to start restoring the database from the backup file (see Figure 2).

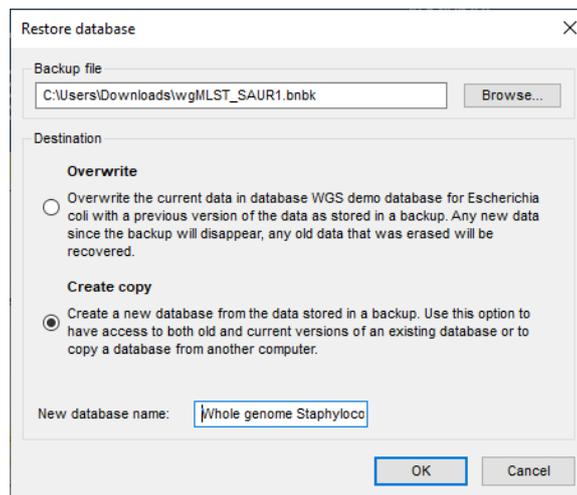


Figure 2: Restoring the whole genome demonstration database from the BioNumerics backup file wgMLST_SAUR.bnbk.

13. Once the process is complete, click <**Yes**> to open the database.

The *Main* window is displayed (see Figure 3).

3 About the demonstration database

The WGS demo database contains links to sequence read set data on NCBI's sequence read archive (SRA) for 97 publicly available sequencing runs of three *Staphylococcus aureus* whole genome sequencing studies ([?] [?] [?]). Additional information (in entry info fields Organism name, Instrument, Study accession, etc.) was collected from the corresponding publications and added to the demonstration database.

Eight experiments are present in the demo database and are listed in the *Experiment types* panel (see Figure 4).

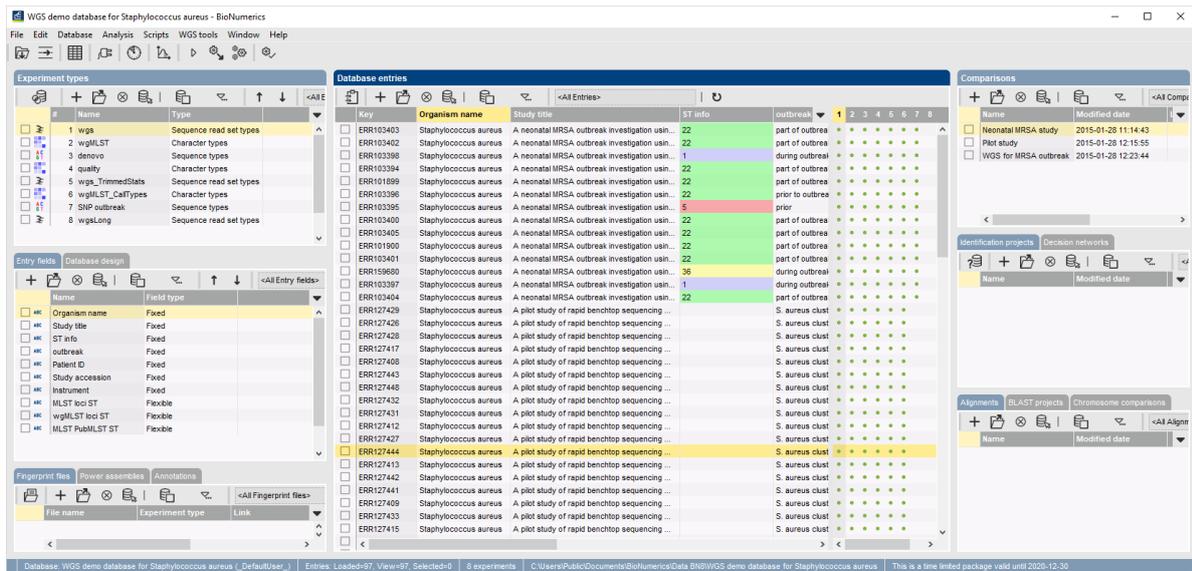


Figure 3: The *Staphylococcus aureus* demonstration database: the *Main* window.

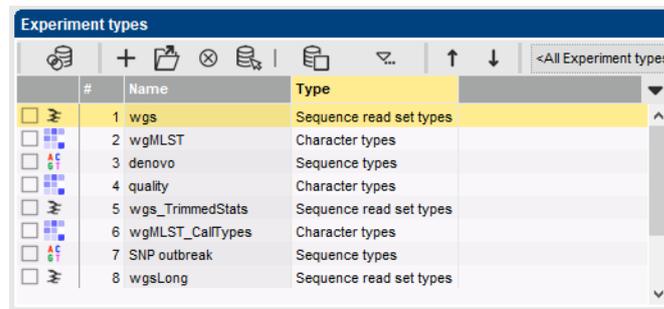


Figure 4: The *Experiment types* panel in the *Main* window.

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 *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 5).

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

The *Sequence editor* window opens, containing the results from the de novo assembly algorithm, i.e. concatenated de novo contig sequences (see Figure 6).

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

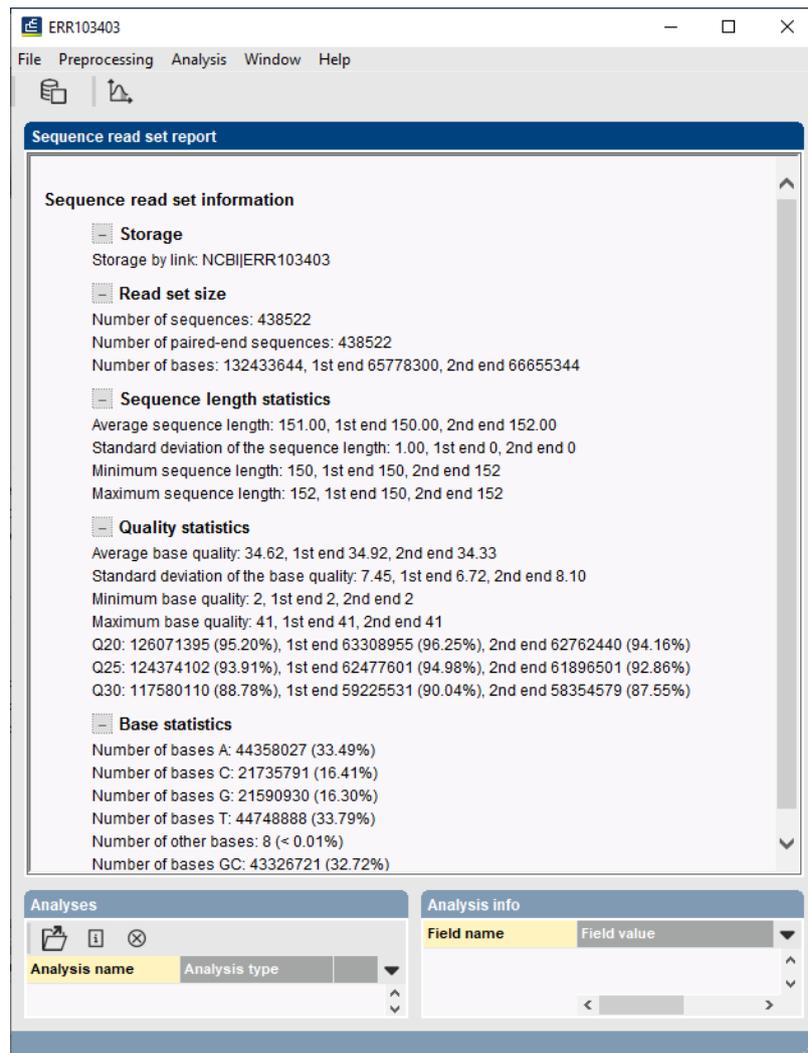


Figure 5: The sequence read set experiment card for an entry.

experiment.

A reference mapping has been calculated for all entries from the Neonatal MRSA study and the resulting sequences are stored in the **SNP outbreak** sequence type. These sequences are used in the wgSNP tutorials to illustrate the wgSNP functionality present in BIONUMERICs.

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.

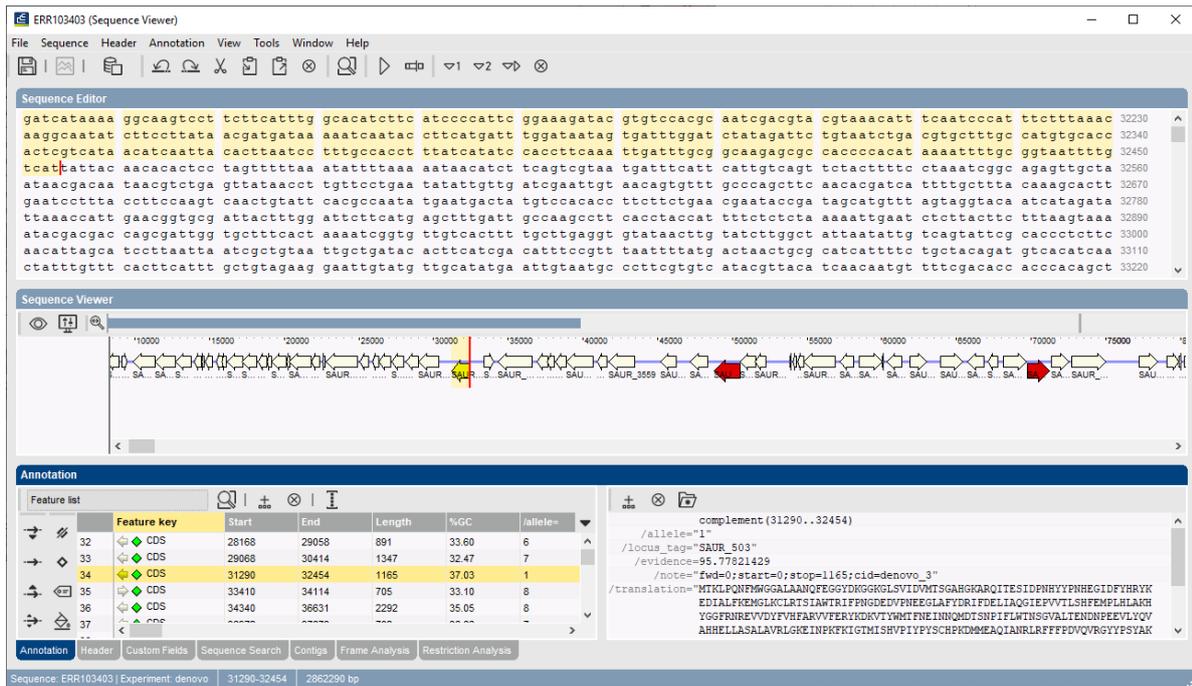


Figure 6: The Sequence editor window.

4 Installing the *Staphylococcus aureus* functional genotyping plugin

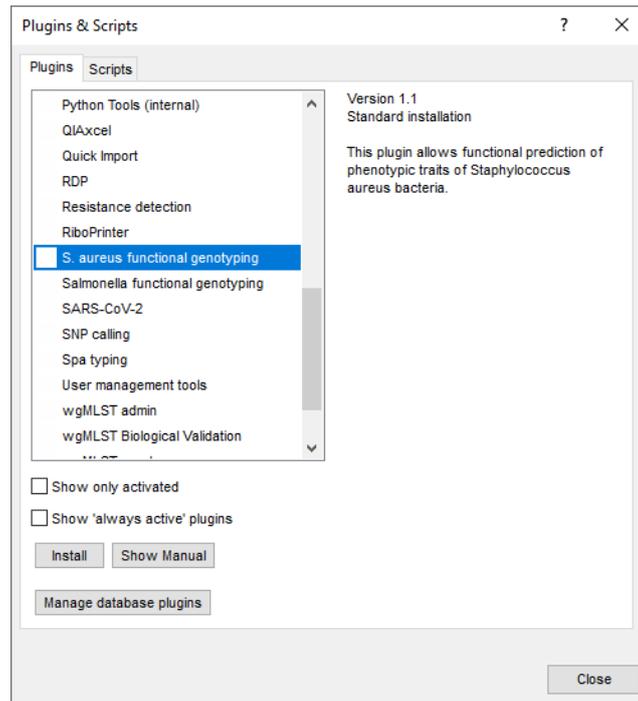
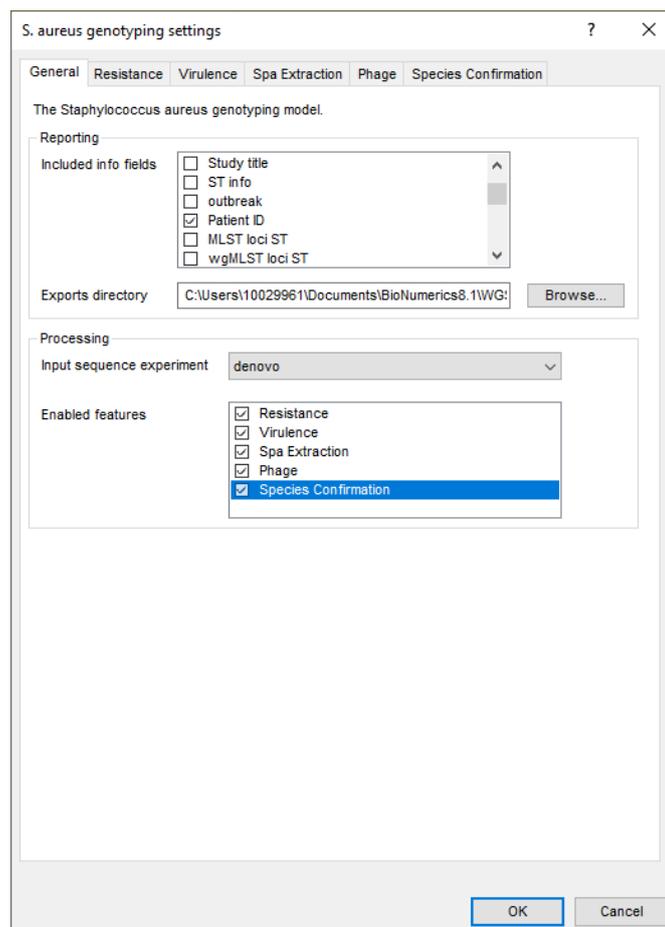
1. Call the *Plugins and Scripts* dialog box from the *Main* window by selecting **File > Install / remove plugins...** ().
2. Select the *S. aureus* genotyping plugin and press the **<Install>** button (see Figure 7).
3. Confirm the installation of the plugin.

During installation, the plugin downloads online knowledge bases from <https://www.bionumerics.com>, which requires a connection to the internet.

4. Click on **<Yes>** to review the settings.

The *S. aureus* genotyping settings dialog box pops up, consisting of 6 tabs (see Figure 8). In the *General* tab the following general settings need to be specified:

- **Included info fields:** In this list the entry information fields that will be displayed in the genotyping report can be specified.
- **Exports directory:** With **<Browse...>** you can specify an export directory to store all exports from the genotyping reports.
- **Input Sequence experiment:** From the drop-down list you can specify the sequence experiment that holds the (whole) genome sequences that will be screened.
- **Enabled features:** This list contains all offered features of the genotyping plugin. Features which are not required can be disabled in this list to save on processing time and omit the corresponding sections from the report. By default, all features are enabled.

**Figure 7:** Install plugin.**Figure 8:** The *Settings dialog: General* tab.

5. In our demonstration database, the assembled sequences are stored in the **denovo** sequence experiment. Make sure this experiment is selected from the drop-down list and optionally check the **Patient ID** to include in the report.

The other tabs group the settings for each possible search: Resistance (Acquired and mutational resistance), Virulence (Acquired virulence and Virulence islands), Spa extraction, Phage and Species confirmation.

Except for the **Spa extraction** and **Species confirmation** tab, all tabs contain a *Knowledgebase*, *BLAST* and *Results* panel:

1. *Knowledgebase*: in this panel the **Version** and **Name** of the knowledge base that is being used for this feature is shown. A different knowledge base version can be selected by pressing the **<Change...>** button. With **Check for updates on startup** checked, BIONUMERICS will check if a newer knowledge base version is available online for this feature each time the database is opened.
2. *BLAST*: in this panel two settings for the BLAST algorithm can be specified; the **Minimum percent identity (%)** and the **Minimum coverage (%)** of your query sequence against the knowledge base's reference sequences. If the option **Combine fragments** is checked, genes that occur fragmented in the genome (i.e. split over two contigs) can still be detected.
3. *Results*: in this panel the output database information fields and experiments to which the screening results will be written can be dictated. Use the drop-down list to choose an existing experiment type or field, or the **<Create>** option to create new experiments and fields. A default name for the experiment or information field is suggested, but you can adjust this if you want to. Check **Annotate sequence experiment** to annotate the input sequence with the detected genotyping features.

In the *Results* panel of the *Spa Extraction* tab you can click on **<Change...>** next to *Sequence extraction* to open the *Change sequence experiment* dialog box. For each PCR target in the list a new or existing sequence experiment type can be selected from the drop-down list.

6. In this tutorial, specify the experiment types and information fields in all tabs by selecting the **<Create>** option in the drop-down lists and accepting the default names. Leave the other settings unaltered.
7. In the *Results* panel of the *Spa Extraction* tab click on **<Change...>** next to *Sequence extraction* to open the *Change sequence experiment* dialog box. In the drop-down list of each PCR target select the **<Create>** option and accept the default name. Click on **<OK>** to close the *Change sequence experiment* dialog box (see Figure 9).
8. Click on **<OK>** in the *S. aureus genotyping settings* dialog box.
9. When the *S. aureus functional genotyping plugin* is successfully installed, a confirmation message pops up. Press **<OK>**.
10. Press **<Close>** to close the *Plugins and Scripts* dialog box.
11. Close and reopen the database to activate the features of the *S. aureus functional genotyping plugin*.

The *S. aureus functional genotyping plugin* installs menu items in the main menu of the software under **S. aureus** (see Figure 10).



The settings specified during installation of the plugin can be called again at any time with **S. aureus > Settings...**

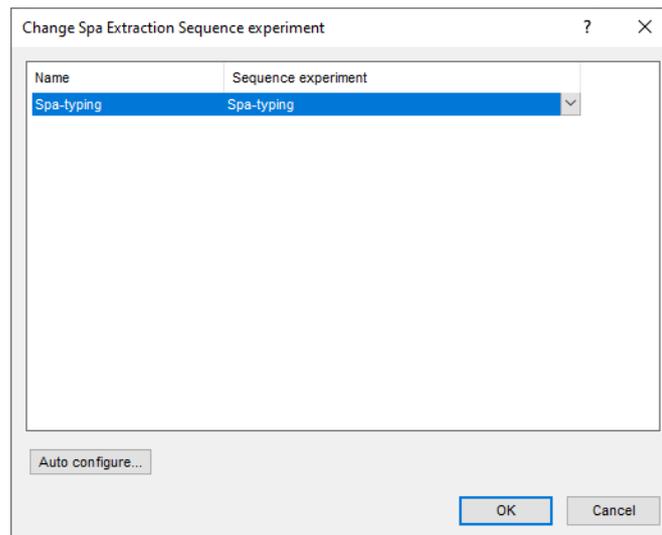


Figure 9: The *Change sequence experiment* dialog box.

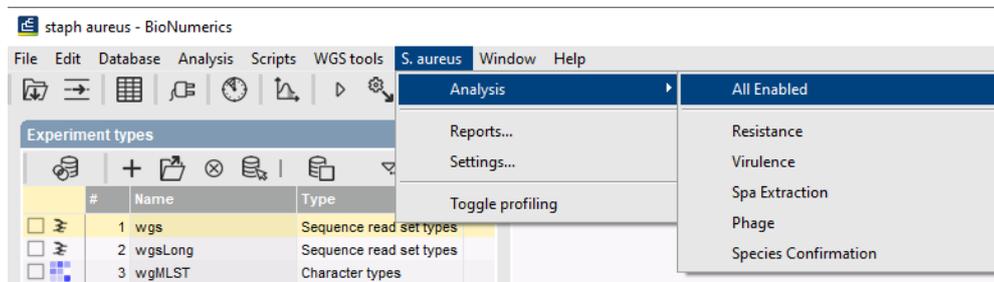


Figure 10: New menu-items after installation of the *S. aureus functional genotyping plugin*.

5 Screening of entries

The screening can be done on any selection of entries in the database.

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.

Selected entries are marked by a checked ballot box (☑) and can be unselected in the same way.

2. In order to select a group of entries, hold the **Shift**-key and click on another entry.

A group of entries can be unselected the same way.

3. Make sure a few entries are selected in the *Database entries* panel of the demonstration database.

Screening for the phenotypic traits can be done for all tools checked in the *S. aureus genotyping settings* dialog box (**S. aureus > Analysis > All Enabled**) or for each tool separately (**S. aureus > Analyze > ...**).

4. Select **S. aureus > Analysis > All Enabled** to screen the selected entries for all enabled traits.

A progress bar appears. The analysis time depends on the number of selected entries. When the

analysis is finished, the progress bar disappears. The detected traits for the screened entries are stored in the database.

The resistance type prediction (**resistance type**), Virulence islands results (**Total Islands**) and Species confirmation **Species confirmation** are written to the information fields in the *Database entries* panel (see Figure 11). Please note that the shown names of the information fields are those created per default, but can be different in your case depending on whether you have chosen an alternative name during installation.



Key	resistance type	Total islands	Species confirmation
<input checked="" type="checkbox"/> ERR101900	methicillin-resistant S. aureus (MRSA)	0	Staphylococcus aureus
<input checked="" type="checkbox"/> ERR103397	methicillin-resistant S. aureus (MRSA)	1	Staphylococcus aureus

Figure 11: Example output of the **Resistance type**, **Total Islands** and **Species confirmation** information fields.

The character experiment types for **Resistance**, **Virulence**, **Spa extraction** and **Phage** and the sequence experiment type for **Spa extraction** are created and updated with the predicted traits. Please note that the shown names of the experiment types are those created per default, but can be different in your case depending on whether you have chosen an alternative name during installation.

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



The characters in the characters experiments are displayed in the same order they are listed in their knowledge base. However, it might be more convenient for interpretation to have them displayed alphabetically. This can be done in the *Character type* window with the option **Characters > Arrange characters by field...** (⌵).

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

Acquired Resistance (see Figure 12):

- **Resistance traits:** contains the results for each antibiotic: 0 = not detected (sensitive), 1 = detected (resistant).
- **Resistance loci:** contains the results for each resistance gene: 0 = not detected (sensitive), when detected (resistant) the % identity of the best hit is shown.

Mutational Resistance (see Figure 13):

- **Resistance mutations:** contains the results for each resistance mutation: -2 = partially indecisive, -1 = fully indecisive, 0 = not detected (sensitive), 1 = detected (resistant).

Acquired Virulence (see Figure 14):

- **Virulence loci:** contains the results for each virulence gene: 0 = not detected, when detected the % identity of the best hit is shown.
- **Virulence traits:** contains the results for each virulence type: 0 = not detected, 1 = detected.

Virulence Islands (see Figure 15):

Character	Value	Mapping
Minocycline	0	<->
Neomycin	0	<->
pristinamycin IIA	0	<->
Amikacin	1	<->
Quinupristin+Dalfofri...	0	<->
Streptomycin	1	<->
Cephalothin	0	<->
Virginiamycin M	0	<->
Tylosin	0	<->
Lincomycin	0	<->
Clindamycin	1	<->
Oleandomycin	0	<->
Aztreonam	0	<->
Unknown Tetracycline	0	<->
Butirosin	0	<->
quinupristin	1	<->
Pristinamycin IA	1	<->
Dalfofristin	0	<->
Butiromycin	0	<->
Ceftaroline	0	<->
Teicoplanin	0	<->
Fusidic acid	0	<->
penicillin	1	<->
Spiramycin	0	<->
Cefotaxime	1	<->
Linezolid	0	<->

Character	Value	Mapping
tet(A)	0	<->
tet(B)	0	<->
tet(C)	0	<->
tet(D)	0	<->
tet(E)	0	<->
tet(G)	0	<->
tet(H)	0	<->
tet(J)	0	<->
tet(K)	0	<->
tet(L)	0	<->
tetA(P)	0	<->
tet(V)	0	<->
tet(Y)	0	<->
tet(Z)	0	<->
tet(30)	0	<->
tet(31)	0	<->
tet(33)	0	<->
tet(35)	0	<->
tet(39)	0	<->
tet(40)	0	<->
tet(41)	0	<->
tet(42)	0	<->
tet(43)	0	<->
tet(45)	0	<->
tetA(46)	0	<->
tetB(46)	0	<->

Figure 12: Example output of the *Resistance_traits* and the *Resistance_loci* experiment types for sample ERR101900.

Character	Value	Mapping
23S_dC2447T	0	<->
23S_dT2500A	0	<->
23S_dG2576T	0	<->
cls2_pA23V	0	<->
cls2_pL52F	0	<->
cls2_pF60S	0	<->
dfrB_pL21V	0	<->
dfrB_pH31N	0	<->
dfrB_pL41F	0	<->
dfrB_pN60I	0	<->
dfrB_pF99Y	0	<->
dfrB_pF99S	0	<->
dfrB_pF99I	0	<->
dfrB_pH150R	0	<->
dhps_pT51M	0	<->
fusA_pM16I	0	<->
fusA_pA67T	0	<->
fusA_pA70V	0	<->
fusA_pA71V	0	<->
fusA_pR76C	0	<->
fusA_pV90I	0	<->
fusA_pP114H	0	<->
fusA_pQ115L	0	<->
fusA_pD189G	0	<->
fusA_pE233Q	0	<->
fusA_pT326I	0	<->

Figure 13: Example output of the *Resistance_mutations* experiment type for sample ERR101900.

- **island_counts:** contains the number of detected loci associated to a pathogenicity island.
- **island_percentages:** contains the percentage of detected loci associated to a pathogenicity island.

Phage detection (see Figure 16):

Character	Value	Mapping
toxin	1	<=>
exoenzyme	1	<=>
immune evasion	1	<=>
adherence	1	<=>
plasminogen activator	1	<=>
antiphagocytosis	1	<=>
secretion system	1	<=>
iron uptake	1	<=>

Character	Value	Mapping
sely	0	<=>
sel27	0	<=>
selx	95	<=>
set	0	<=>
selz	0	<=>
ses	0	<=>
selv	0	<=>
sel26	0	<=>
nuc	99	<=>
edinC	0	<=>
edinB	0	<=>
edinA	0	<=>
seu	100	<=>
ser	0	<=>
seo	100	<=>
sen		

Figure 14: Example output of the *Virulence_traits* and the *Virulence_loci* experiment types for sample ERR101900.

Character	Value	Mapping
vSa_gamma	0	<=>
vSa_beta	0	<=>
SaPIbov2	0	<=>
vSa4	0	<=>
SaPIAv	0	<=>
SaPI4	0	<=>
SaPI68111	0	<=>
vSa1	0	<=>
SaPI3	0	<=>
vSa3	0	<=>
SaPI5	0	<=>
SaPIbov	0	<=>
SaPI2	0	<=>
SaPI1	0	<=>
vSa_alpha	0	<=>
SAPI-S0385	0	<=>
etd_PI	0	<=>

Character	Value	Mapping
vSa_gamma	0	<=>
vSa_beta	0	<=>
SaPIbov2	0	<=>
vSa4	0	<=>
SaPIAv	0	<=>
SaPI4	0	<=>
SaPI68111	0	<=>
vSa1	0	<=>
SaPI3	0	<=>
vSa3	0	<=>
SaPI5	0	<=>
SaPIbov	0	<=>
SaPI2	0	<=>
SaPI1	0	<=>
vSa_alpha	0	<=>
SAPI-S0385	0	<=>
etd_PI	0	<=>

Figure 15: Example output of the *island_counts* and the *island_percentages* experiment types for sample ERR101900.

- **Phage_seq_ids:** contains the results of the phages detection by sequence IDs: 0 = not detected, when detected the % of the detected full phage is shown.
- **Phage_categories:** contains the results of the phages detection by phage categories: 0 = not detected, when detected the % of the detected full phage is shown.

Spa extraction:

- **Spa Extraction amplicons:** contains the result for the in-silico PCR: 0 = no amplicon, 1 = amplicon generated.
- **Spa-typing:** contains the *spa*-repeat region sequence.

6. Close the character and sequence card(s).

Character	Value	Mapping
Staphylococcus virus 44AHJD (44AHJD)	0	<->
Staphylococcus virus 80alpha	40	<+>
Staphylococcus virus 3a (3A)	92	<->
Staphylococcus phage vB_SauS_phI2	81	<->
Staphylococcus phage tp310-3	52	<->
Staphylococcus phage tp310-2	93	<->
Staphylococcus phage tp310-1	49	<->
Staphylococcus phage phiJB	56	<->
Staphylococcus phage phi2958PVL (ph...)	83	<->
Staphylococcus phage Stau2	0	<->
Staphylococcus phage SPbeta-like	0	<->
Staphylococcus phage SLPW	0	<->
Staphylococcus phage SA13	72	<->
Staphylococcus phage S25-4	0	<->
Staphylococcus phage S25-3	0	<->
Staphylococcus phage P108	0	<->
Staphylococcus phage CNPx	0	<->
Staphylococcus phage BP39	0	<->
Staphylococcus phage 812	0	<->
Staphylococcus phage 3MRA	50	<->
Staphylococcus phage 23MRA		

Character	Value	Mapping
Spbetavirus	0	<->
Unclassified Siphovi...	0	<->
Phietavirus I	0	<->
Unclassified Siphovi...	0	<->
Silviavirus	0	<->
Unclassified Myoviri...	0	<->
Unclassified phage	0	<->
Biseptimavirus I	0	<->
Sepunavirus	0	<->
Kayvirus	0	<->
Biseptimavirus II	0	<->
Phietavirus V	0	<->
Rosenblumvirus	0	<->
Triavirus	93	<->
Phietavirus II	0	<->
Phietavirus III	76	<->
Biseptimavirus III	81	<->
Phietavirus IV	44	<->
Sextaevirus	0	<->

Figure 16: Example output of the *Phage_seq_ids* and *Phage_categories* experiment types for sample ERR101900.

6 Reports

1. Open the genotype report for the selected entries with **S. aureus** > **Reports....**

The *Report* window contains a genotype report for each of the selected entries (see Figure 17).

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 *S. aureus genotyping settings* dialog box are displayed in the *Genotype report* panel.

3. Select **Report** > **Report styles** in the *Report* window and make sure the option **Summary** is selected (see Figure 18).

A summary of the results of all analyzed traits is displayed in the *Report* window.

4. Select **Report** > **Report styles** in the *Report* window and select the option **Complete**.

In the **Complete** view, the summarized results as well as all available details are shown. All hits that passed the settings for **Resistance**, **Virulence**, **Spa Extraction**, **Phage** and **Species confirmation** screening are listed and described.

5. Click on a hyperlink of one of the predicted traits to display the detailed results in the *Genotype report* panel (see Figure 19).
6. Select **File** > **Exit** to close the *Report* window.

For more detailed information on the genotyping analyses and interpretation of the reported results, please check the genotyping plugin manual.

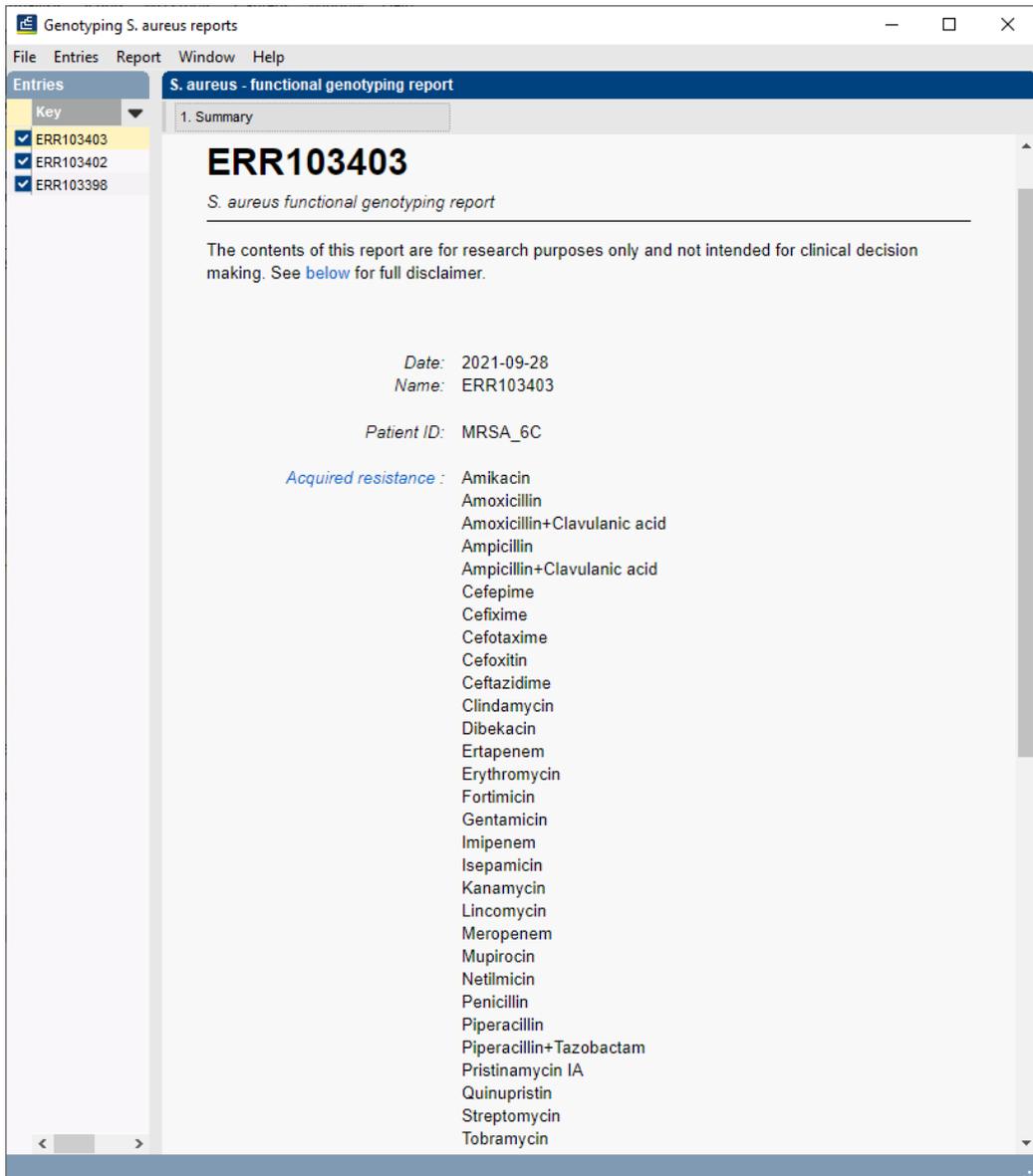


Figure 17: Example of a functional genotyping report.

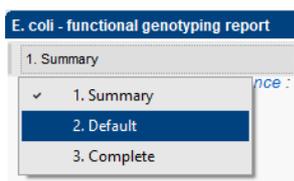


Figure 18: Report templates in the Report window.

The screenshot shows a software window titled 'Genotyping S. aureus reports'. The main content area displays a report for 'S. aureus - functional genotyping report' at step '3. Complete'. The report is titled 'Resistance' and shows 'Acquired resistance results' in a table. The table has six columns: Trait, Locus, Coverage (%), Identity (%), Position, and Accession. Each row includes a 'Publication' link. The table lists various antibiotics and their corresponding genetic markers, all with 100% coverage and identity.

Trait	Locus	Coverage (%)	Identity (%)	Position	Accession
Erythromycin	erm(C)	100.00	100.00	4145..3411	M13761
<i>Publication: 3087948</i>					
Lincomycin	erm(C)	100.00	100.00	4145..3411	M13761
<i>Publication: 3087948</i>					
Clindamycin	erm(C)	100.00	100.00	4145..3411	M13761
<i>Publication: 3087948</i>					
Quinupristin	erm(C)	100.00	100.00	4145..3411	M13761
<i>Publication: 3087948</i>					
Pristinamycin IA	erm(C)	100.00	100.00	4145..3411	M13761
<i>Publication: 3087948</i>					
Virginiamycin S	erm(C)	100.00	100.00	4145..3411	M13761
<i>Publication: 3087948</i>					
Amikacin	aadD	100.00	100.00	1000570..999809	M19465
<i>Publication: 3010356</i>					
Tobramycin	aadD	100.00	100.00	1000570..999809	M19465
<i>Publication: 3010356</i>					
Amoxicillin	blaZ	100.00	100.00	1599504..1598617	CP003979
<i>Publication: 23977354</i>					
Ampicillin	blaZ	100.00	100.00	1599504..1598617	CP003979
<i>Publication: 23977354</i>					
Penicillin	blaZ	100.00	100.00	1599504..1598617	CP003979
<i>Publication: 23977354</i>					
Piperacillin	blaZ	100.00	100.00	1599504..1598617	CP003979
<i>Publication: 23977354</i>					
Gentamicin	aac(6')-aph(2'')	100.00	100.00	2744304..2742865	M13771
<i>Publication: 3015884</i>					
Tobramycin	aac(6')-aph(2'')	100.00	100.00	2744304..2742865	M13771
<i>Publication: 3015884</i>					
Netilmicin	aac(6')-aph(2'')	100.00	100.00	2744304..2742865	M13771

Figure 19: Report details.