

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 *BION-UMERICS 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 Obstrate 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).

🖆 Tutorial databases					— 🗆	×
File Database Tutorial Window						
Tutorial databases						
<u>1</u>						
Name	Size (MB)	Min	Max	-		
BandScoring demo database	15	7	8.1	^	The DemoBase Connected demonstration	^
DemoBase Connected	4	7.5	8.1		database. This database is used in several	
Neisseria MLST demo database	1	7.5	8.1		tutorials.	
SNP demonstration database	0	7.5	8.1			
WGS_demo_database_for_Brucella_spp	214	7.6	8.1			
WGS_demo_database_for_Burkholderia_cepacia	142	8	8.1			
WGS_demo_database_for_Escherichia_coli	601	7.5	8.1			
WGS_demo_database_for_Listeria_monocytogenes	288	7.5	8.1			
WGS_demo_database_for_MTBC	279	8	8.1			
WGS_demo_database_for_Salmonella_enterica	618	8	8.1			
WGS_demo_database_for_Staphylococcus_aureus	624	7.6	8.1			
						\sim
				~	I	-
Tutorials						
<u>г</u> а.						
¹						
Name				•		
Band matching and polymorphism analysis				~	This tutorial illustrates how to calculate a	^
Entry information fields and their properties					Principal Components Analysis (PCA) and a	
Configuring the database layout					Multi Dimensional Scaling (MDS) (sometimes also called Principal	
Selecting entries in a database					Coordinates Analysis (PCoA)) on a	
Combined analysis of fingerprint data					fingerprint data set and how to change the	
Calculating a PCA and an MDS on a fingerprint data set					layout of the obtained plots.	~
Clustering a phenotypic test assay					-	

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 (B).
- 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 BIONU-MERICS.

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).

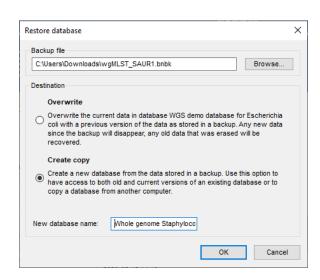


Figure 2: Restoring the whole genome demonstration database from the BioNumerics backup file wgMLST_SAUR.bnbk.

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).

GS demo database for Staphy	lococcus aureus - BioNu	imerics									- 0
Edit Database Analysis S	cripts WGS tools Wi	ndow Help									
<u>→</u> III / C ()	h. D .	0 0,									
	LT	• •v									
riment types				Database entries						Comparisons	
3 + 13 8 8	. h ~	t t	<aii e<="" th=""><th>£] + 🖻</th><th>⊗ & ₽</th><th>All Entries></th><th>3</th><th></th><th></th><th>+ ₽→ ⊗</th><th></th></aii>	£] + 🖻	⊗ & ₽	All Entries>	3			+ ₽→ ⊗	
# Name	Туре			Key Key	Organism name	Study title	ST info	outbreak 🛨 1	234567		Modified date
			•	ERR103403						Neonatal MRS	
	Sequence read set	types	^		Staphylococcus aureus	A neonatal MRSA outbreak investigation usin.		part of outbrea		Plot study	
	Character types			ERR103402 ERR103398	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin.	22	part of outbrea ·			2015-01-28 12:15:55 A outbreak 2015-01-28 12:23:44
	Sequence types		- 11		Staphylococcus aureus	A neonatal MRSA outbreak investigation usin.	1	during outbreak ·		U WGS TOP MRS	A outpreak 2015-01-28 12:23:44
	Character types				Staphylococcus aureus	A neonatal MRSA outbreak investigation usin		part of outbrea •			
E 5 wgs_TrimmedStats	Sequence read set	types		ERR101899	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin	22	part of outbreal •			
6 wgMLST_CalTypes	Character types		_	ERR103396	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin	22	prior to outbrea •	• • • • • •		
7 SNP outbreak	Sequence types			ERR103395	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin.		prior •			
8 wgsLong	Sequence read set	types		ERR103400	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin.		part of outbrea •		<	
			~	ERR103405	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin.	22	part of outbreal			
				ERR101900	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin.	22	part of outbrea		Identification projects	Decision networks
elds Database desion				ERR103401	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin	22	part of outbrea		29 + P	9 ⊗ & ि
				ERR159680	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin	36	during outbreal •			/
	∽ † ↓	<all entr<="" td=""><td>y fields></td><td>ERR103397</td><td>Staphylococcus aureus</td><td>A neonatal MRSA outbreak investigation usin</td><td>1</td><td>during outbreal</td><td></td><td>Name</td><td>Modified date</td></all>	y fields>	ERR103397	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin	1	during outbreal		Name	Modified date
Name				ERR103404	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin.	22	part of outbrea			
Organism name	Fixed		0	ERR127429	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust •			
	Fixed			ERR127426	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust			
	Fixed		- 11	ERR127428	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust •			
	Fixed			ERR127417	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust			
				ERR127408	Staphylococcus aureus	A pilot study of rapid benchtop sequencing					
	Fixed			ERR127443	Staphylococcus aureus	A pilot study of rapid benchtop sequencing					
	Fixed			ERR127448	Staphylococcus aureus	A plot study of rapid benchtop sequencing		S. aureus clust			
	Fixed										
	Flexible				Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust		Alignments BLAST	Eprojects Chromosome comparise
	Flexible				Staphylococcus aureus	A pliot study of rapid benchtop sequencing		S. aureus clust		+ 🖻 😣	
MLST PubMLST ST	Flexible			ERR127412	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust ·			
				ERR127427	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust ·		Name	Modified date
			~	ERR127444	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust ·			
				ERR127413	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust •			
print files Power assemblies	Annotations			ERR127442	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust •	• • • • •		
+ 13 8 8.	6 2	<all fingerprint<="" td=""><td>fies></td><td>ERR127441</td><td>Staphylococcus aureus</td><td>A pilot study of rapid benchtop sequencing</td><td></td><td>S. aureus clust •</td><td></td><td></td><td></td></all>	fies>	ERR127441	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust •			
	-0			ERR127409	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust ·			
File name Ex	periment type	ink	•	ERR127433	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust ·			
			0	ERR127415	Staphylococcus aureus	A pliot study of rapid benchtop sequencing		S. aureus clust ·		~	
4			> [×]					> <	_	>	

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

Experim	ent ty	pes	
କ୍ର	-	+ 🗗 🛛 🖳	All Experiment types
		Name	Type 🗸 🗸
≣ ≨ []	1	wgs	Sequence read set types
	2	wgMLST	Character types
	3	denovo	Sequence types
	4	quality	Character types
≣ ≨	5	wgs_TrimmedStats	Sequence read set types
	6	wgMLST_CallTypes	Character types
🗆 🛟	7	SNP outbreak	Sequence types
∃≹	8	wgsLong	Sequence read set types

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

🖆 ERR103403	-	×
ile Preprocessing Analysis Window Help		
Sequence read set report		
Sequence read set information		
- Storage		
Storage by link: NCBIJERR103403		
- Read set size		
Number of sequences: 438522 Number of paired-end sequences: 438522		
Number of bases: 132433644, 1st end 65778300, 2nd end 66655344		
- Sequence length statistics		
Average sequence length: 151.00, 1st end 150.00, 2nd end 152.00		
Standard deviation of the sequence length: 1.00, 1st end 0, 2nd end 0 Minimum sequence length: 150, 1st end 150, 2nd end 152		
Maximum sequence length: 152, 1st end 150, 2nd end 152		
- Quality statistics		
Average base quality: 34.62, 1st end 34.92, 2nd end 34.33		
Standard deviation of the base quality: 7.45, 1st end 6.72, 2nd end 8.10 Minimum base quality: 2, 1st end 2, 2nd end 2		
Maximum base quality: 41, 1st end 41, 2nd end 41		
Q20: 126071395 (95.20%), 1st end 63308955 (96.25%), 2nd end 62762440 Q25: 124374102 (93.91%), 1st end 62477601 (94.98%), 2nd end 61896501	· · · ·	
Q30: 117580110 (88.78%), 1st end 59225531 (90.04%), 2nd end 58354579		
- Base statistics		
Number of bases A: 44358027 (33.49%)		
Number of bases C: 21735791 (16.41%) Number of bases G: 21590930 (16.30%)		
Number of bases T: 44748888 (33.79%)		
Number of other bases: 8 (< 0.01%) Number of bases GC: 43326721 (32.72%)		~
Analyses Analysis info		
Field name Field	value	
Analysis name Analysis type		^
		×
v		

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.

													_		
C		uence Viewer)	ion View Tools	107-stars 11	t-t-									_	
		~					-								
] 🖂		1 <u> </u>	7 % 🖞 [3 ⊗ Q		▽1 ▽2	$\neg \rangle \otimes$								
quence	Editor														
atcat	taaa	a ggcaagt	cct tcttca	tttg gca	catcttc a	tccccatt	c ggaaa	gatac	gtgtccac	gc aatcgacgt	a cgtaaacat	t tcaatcccat	ttctttaaad	32230	
aggca	aata	t cttcctt	ata acgatg	ataa aaa	tcaatac c	ttcatgat	t tggat	aatag	tgatttgg	at ctatagatt	c tgtaatctg	a cgtgctttgc	catgtgcaco	32340	
tcgt	tcat	a acatcaa	tta cactta	atcc ttt	gccacct t	tatcatat	c cacct	tcaaa	ttgatttg	cg gcaagagcg	c caccccaca	t aaaattttgc	ggtaatttt	32450	
atta	atta	c aacacac	tcc tagttt	ttaa ata	ttttaaa a	ataacato	t tcagt	cgtaa	tgatttca	tt cattgtcag	t tctactttt	c ctaaatcggc	agagttgcta	a 32560	
aaco	gaca	a taacgtc	tga gttata	acct tgt	tcctgaa t	atattgtt	g atoga	attgt	aacagtgt	tt gcccagctt	c aacacgatc	a ttttgcttta	a caaagcactt	32670	
atco	ttt	a ccttcca	agt caactg	tatt cac	gccaata t	gaatgact	a tgtcc	acacc	ttcttctg	aa cgaataccg	a tagcatgtt	t agtaggtaca	atcatagata	a 32780	
aaac	ccat	t gaacggt	gcg attact	ttgg att	cttcatg a	gctttgat	t gccaa	gcctt	cacctacc	at tttctctct	a aaaattgaa	t ctcttacttc	: tttaagtaaa	32890	
acga	acga	c cagcgat	tgg tgcttt	cact aaa	atcggtg t	tgtcactt	t tgett	gaggt	gtataact	tg tatcttggc	t attaatatt	g tcagtattcg	g caccetette	33000	
catt	tage	a tccttaa	tta atcgct	gtaa ttg	ctgatac a	cttcatcg	ja cattt	ccgtt	taatttta	tg actaactgo	g catcatttt	c tgctacagat	gtcacatcaa	33110	
attt	tgtt	t cacttca	ttt gctgta	gaag gaat	ttgtatg t	tgcatatg	ja attgt	aatgc	ccttcgtg	tc atacgttac	a tcaacaatg	t tttcgacacc	: acccacaget	33220	
_															
		'10000 ⊐I}-{=-}=; sAsAs	'15000 IQIQIQIS_S_IQIQ S_S_S_IQIQ	'20000 X K SASAUR.	'25000 ⊐≮ो⊀КҚЪҚ⊐Қ	'30000 	'35000	40000 ' '	45000	50000	'55000 '60000 SAUR SA SA SA		70000	75000 SAU	, ,
] @	10000	14044247444	SA SAUR.	25000 コインイベスティント S SA	'30000 			ÂUR_3559 SÂU S					75000 SAU	, , ,
> [↑] / ₁) @,		2011 ±	SASAUR. ⊗ <u>I</u>		UR <mark>Sau</mark> r S			AUR_3559 SAU S	L Star	SAUR SA SA SA			75000	- J
> [↑] / ₁) @,	<	14044247444	SA SAUR.	'25000 试试试试试	'30000 			AUR_3559 SAU S.	complement (31290	SAUR SA SA SA			75000	, , ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
Determine the second secon) @,	<pre>SASASAS.</pre>	2011 ±	SASAUR. ⊗ <u>I</u>		UR <mark>Sau</mark> r S			AUR_3559 SAU S	complement (31290	SAUR SA SA SA			75000 SAL	
Determine the second secon	n st	Feature key		SA SAUR	Length	URSAURS.	AUR_			complement (31290	SAUR SA SA SA			75000 SAL	
Dotation eature list	n st 32	<pre>SASASAS.</pre>	2014 10 4 2 - 2 - 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 +	⊗ <u>]</u> End 29058	Length 891	ия <mark>Sau</mark> Rs.s %GC 33.60	/allele= 6		AUR_3559 SAU S	complement (31290 1" SAUR_503" 55.7721429 fwd=0:stat=0;st	32454)	€ <u>, , , , , , , , , , , , , , , , , , , </u>	A. SALR.	SAU	>
notation eature list	n st 32 33 34	Feature key	Q	★ ↓ ↓ SA SAUR. End 29058 30414	Length 891 1347	ия <mark>S₄u</mark> яs s ₄ uяs ss 33.60 32.47	/allele= 6 7		AUR_3559 SAU S	complement (31290 1" SAUR_503" 55.77621429 Ywd-0;start-0;st	32454)	VO_3"	ESIDPNHYYPNHEGI	DFYHRYK	
■ attion eature list	n st 32 33 34	 ✓ ✓ ✓ ✓ ✓ ✓ <	Q] ± Sart 2868 31290	Image: Same Same Same Same Same Same Same Same	Length 891 1347 1165 2292	33.60 32.47 37.03 33.10 35.05	AUR		AUR_3559 SAU S AUR_3559 SAU S /allele=" /locus_tag=" /evidence= /note=" translation="	complement (31290 1" 55.77621429 MTKLPDURMEGALAR MTKLPDURMEGALAR	32454)	vo_3" IUVATSGARGKARQIT IUVATSGARGKARQIT IUVATSGARGKARQIT	ESIDPNHYYPNHEGI AQGIEFVVTLSHEM	DFYHRYK	>
■ 10tation eature list %	n st 32 33 34 35	Feature key ⇔ CDS ⇔ CDS ⇔ CDS	CALC ALC AND A CONTRACT AND A CONTRA	Image: SA SA	Length 891 1347 1165 705	<mark>жарана (1996) 33.60 32.47 37.03 33.10</mark>	AUR AUR 6 7 1 8 8 8		AUR 3009 SAU S AUR 3009 SAU S /allele=' /locus tag=' /evidence- /note=' translation='	complement (31290 1" SAUR_503" 53,77021429 fwd=0;start=0;st Rel=0;start=0;st Rel=0;start=0;st		vo_3" IDMISGAHGKARQITI IDMISGAHGKARQITI INEGLAYDRITPELI INEDIMONSIPIT	ESIDPNHYYPNHEGI AQGIEVVTJSHEM	DFYHRYK PLHLAKH EEVLYQV	
notation eature lis - # - \$	n st 32 33 34 35 36	Feature key	Child (K, T) - Child	BÀ BÀUR BÀ BÀUR	Length 891 1347 1165 2292	33.60 32.47 37.03 33.10 35.05	Aur Charles Contraction Contra		AUR 3009 SAU S AUR 3009 SAU S /allele=' /locus tag=' /evidence- /note=' translation='	complement (31290 1" SAUR_503" 53,77021429 fwd=0;start=0;st Rel=0;start=0;st Rel=0;start=0;st		vo_3" IUVATSGARGKARQIT IUVATSGARGKARQIT IUVATSGARGKARQIT	ESIDPNHYYPNHEGI AQGIEVVTJSHEM	DFYHRYK PLHLAKH EEVLYQV	

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 functional 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.

Plugins & Scripts			?	×
Plugins Scripts				
Python Tools (internal) QIAxcel Quick Import RDP Resistance detection RiboPrinter Saureus functional genotyping Salmonella functional genotyping SARS-CoV-2 SNP calling Spa typing User management tools wgMLST admin wgMLST Biological Validation USE Show only activated Show 'always active' plugins Install Show Manual Manage database plugins	~	Version 1.1 Standard installation This plugin allows functional prophenotypic traits of Staphyloco aureus bacteria.		
			Close	e

Figure 7: Install plugin.

General	Resistance	Virulence	Spa Extraction	Phage	Species Confirma	ition		
					0,000,000,000,000,000			
Reporti	-	iureus gen	otyping model.					
	d info fields	Stud	lv title		^			
		ST ir	nfo					
		outb Patie						
			T loci ST					
			ILST loci ST		*			
Exports	s directory	C:\Users	s\10029961\Docum	ents\Biol	Numerics8.1\WG	Bro	owse	
Proces	sing							
Input se	equence expe	riment	denovo			~		
		-						
Enable	d features		 Resistance Virulence 					
			Spa Extraction					
			 Phage Species Confir 	rmation				

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:

- 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.
- 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.

- 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...*.

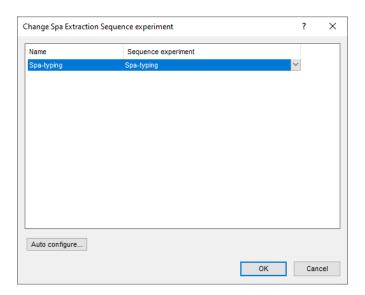


Figure 9: The Change sequence experiment dialog box.

🖆 staph aureus - BioNumerics		
File Edit Database Analysis Scripts WG	S tools S. aureus Window Help	
☞ 글 🖩 🕫 🕚 📐	🗅 🔍 Analysis 🔸	All Enabled
Experiment types	Reports	Resistance
ଡ଼ + 🗗 ⊗ 🗞 ଘ	Settings	Virulence
# Name Type	Toggle profiling	Spa Extraction
□ ≩ 1 wgs Seque	ence read set types	Phage
	ence read set types	Species Confirmation
3 wgMLST Chara	acter types	

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.

Data	abase entries			
Lt 1] + 🖻 🛛 🗎	🕄 🗗 🖳 <selected er<="" th=""><th>ntries></th><th>ן ט ו</th></selected>	ntries>	ן ט ו
		resistance type		Species confirmation
~	ERR101900	methicillin-resistant S. aureus (MRSA)	0	Staphylococcus aureus
~	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...* (\downarrow).

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):

ERR101900				ERR101900			
Character	Value	Mapping	•	Character	Value	Mapping	
Minocycline	0	<->	^	tet(A)	() <->	1
Neomycin	0	<->		tet(B)	() <->	
pristinamycin IIA	0	<->		tet(C)	() <->	
Amikacin	1	<+>		tet(D)) <->	
Quinupristin+Dalfopri	0	<->		tet(E)) <->	
Streptomycin	1	<+>		tet(G)	() <->	
Cephalothin	0	<->		tet(H)) <->	
Virginiamycin M	0	<->		tet(J)) <->	
Tylosin	0	<->		tet(K)) <->	
Lincomycin	0	<->		tet(L)) <->	
Clindamycin	1	<+>		tetA(P)) <->	
Oleandomycin	0	<->		tet(V)) <->	
Aztreonam	0	<->		tet(Y)) <->	
Unknown Tetracycline	0	<->		tet(Z)) <->	
Butirosin	0	<->		tet(30)) <->	
quinupristin	1	<+>		tet(31)) <->	
Pristinamycin IA	1	<+>		tet(33)) <->	
Dalfopristin	0	<->		tet(35)) <->	
Butiromycin	0	<->		tet(39)) <->	
Ceftaroline	0	<->		tet(40)) <->	
Teicoplanin	0	<->		tet(41)) <->	
Fusidic acid	0	<->		tet(42)) <->	
penicillin	1	<+>		tet(43)) <->	
Spiramycin	0	<->		tet(45)	() <->	
Cefotaxime	1	<+>		tetA(46)	() <->	
Linezolid	0	<->	~	tetB(46)	() <->	
Press Insert to add chara	acter		1	Press Insert to add	character		

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):

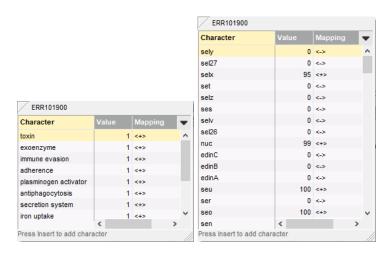


Figure 14: Example output of the *Virulence_traits* and the *Virulence_loci* experiment types for sample ERR101900.

·			_	-	_	1	
Character	Value	Mapping	•	Character	Value	Mapping	•
vSa_gamma	0	<->	^	vSa_gamma	0	<->	1
vSa_beta	0	<->		vSa_beta	0	<->	
SaPlbov2	0	<->		SaPlbov2	0	<->	
vSa4	0	<->		vSa4	0	<->	
SaPIAv	0	<->		SaPIAv	0	<->	
SaPl4	0	<->		SaPl4	0	<->	
SaPI68111	0	<->		SaPI68111	0	<->	
vSa1	0	<->		vSa1	0	<->	
SaPI3	0	<->		SaPI3	0	<->	
vSa3	0	<->		vSa3	0	<->	
SaPI5	0	<->		SaPI5	0	<->	
SaPlbov	0	<->		SaPlbov	0	<->	
SaPI2	0	<->		SaPl2	0	<->	
SaPI1	0	<->		SaPI1	0	<->	
vSa_alpha	0	<->		vSa_alpha	0	<->	
SAPI-S0385	0	<->		SAPI-S0385	0	<->	
etd_PI	0	<->		etd_PI	0	<->	
			×			_	`
Press Insert to add	<		>	Press Insert to add cha	<		>

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	-	Character	Value	Mapping	
Staphylococcus virus 44AHJD (44AHJD)		<->	^	Spbetavirus	-	<->	
Staphylococcus virus 80alpha		<+>		Unclassified Siphovi		<->	
Staphylococcus virus 3a (3A)		<+>		Phietavirus I	-	<->	
Staphylococcus phage vB_SauS_phi2	81	<+>		Unclassified Siphovi	0	<->	
Staphylococcus phage tp310-3	52	<+>		Silviavirus	0	<->	
Staphylococcus phage tp310-2	93	<+>		Unclassified Myoviri	0	<->	
Staphylococcus phage tp310-1	49	<+>		Unclassified phage	0	<->	
Staphylococcus phage phiJB	56	<+>		Biseptimavirus I	0	<->	
Staphylococcus phage phi2958PVL (ph	83	<+>		Sepunavirus	0	<->	
Staphylococcus phage Stau2	0	<->		Kayvirus	0	<->	
Staphylococcus phage SPbeta-like	0	<->		Biseptimavirus II	0	<->	
Staphylococcus phage SLPW	0	<->		Phietavirus V	0	<->	
Staphylococcus phage SA13	72	<+>		Rosenblumvirus	0	<->	
Staphylococcus phage S25-4	0	<->		Triavirus	93	<+>	
Staphylococcus phage S25-3	0	<->		Phietavirus II	0	<->	
Staphylococcus phage P108	0	<->		Phietavirus III	76	<+>	
Staphylococcus phage CNPx	0	<->		Biseptimavirus III	81	<+>	
Staphylococcus phage BP39	0	<->		Phietavirus IV	44	<+>	
Staphylococcus phage 812	0	<->		Sextaecvirus	0	<->	
Staphylococcus phage 3MRA	50	<+>	\checkmark				
Staphylococcus phage 23MRA	<	>					

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.

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.

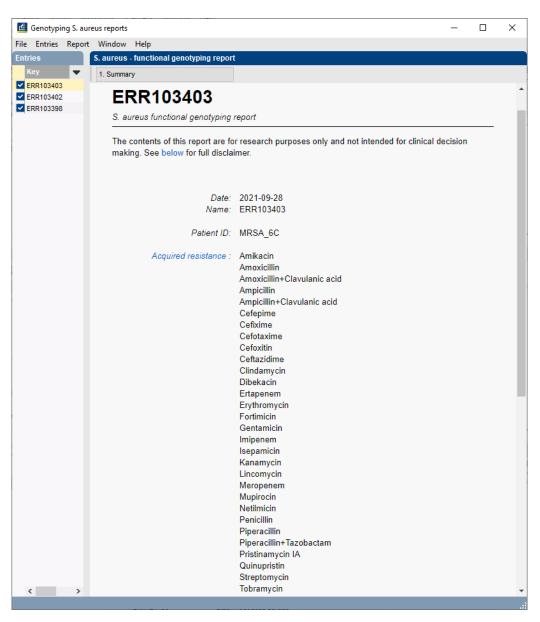


Figure 17: Example of a functional genotyping report.

E. co	E. coli - functional genotyping report				
	ummary				
~	1. Summary	nce :			
	2. Default				
	3. Complete				

Figure 18: Report templates in the *Report window*.

Genotyping S. au	reus reports					- 🗆
Entries Report	· · · · · · · · · · · · · · · · · · ·					
ries	S. aureus - functional genotypin	ig report				
Key 🔻	3. Complete					
ERR103403 ERR103402	Resistance					
ERR103398	Acquired resistance resu	ilts				
	Trait	Locus	Coverage (%)	Identity (%)	Position	Accession
	Erythromycin	erm(C)	100.00	100.00	41453411	M13761
	Publication: 30879	948				
	Lincomycin	erm(C)	100.00	100.00	41453411	M13761
	Publication: 30879	948				
	Clindamycin	erm(C)	100.00	100.00	41453411	M13761
	Publication: 30879	948				
	Quinupristin	erm(C)	100.00	100.00	41453411	M13761
	Publication: 30879	948				
	Pristinamycin IA	erm(C)	100.00	100.00	41453411	M13761
	Publication: 30879					
	Virginiamycin S	erm(C)	100.00	100.00	41453411	M13761
	Publication: 30879			100.00	4000570 000000	
	Amikacin	aadD	100.00	100.00	1000570999809	M19465
	Publication: 30103		400.00	400.00	4000570 000000	140.405
	Tobramycin	aadD	100.00	100.00	1000570999809	M19465
	Publication: 30103 Amoxicillin	blaZ	100.00	100.00	1500504 1509617	CP003979
			100.00	100.00	15995041598617	CP003979
	Publication: 2397 Ampicillin	blaZ	100.00	100.00	15995041598617	CP003979
	Publication: 2397		100.00	100.00	13333041330017	0F003919
	Penicillin	blaZ	100.00	100.00	15995041598617	CP003979
	Publication: 2397		100.00	.00.00	10000011000011	0100010
	Piperacillin	blaZ	100.00	100.00	15995041598617	CP003979
	Publication: 2397					
	Gentamicin	aac(6')-	100.00	100.00	27443042742865	M13771
		aph(2")				
	Publication: 30158	384				
	Tobramycin	aac(6')- aph(2'')	100.00	100.00	27443042742865	M13771
	Publication: 30158					
	Netilmicin	aac(6')- aph(2'')	100.00	100.00	27443042742865	M13771

Figure 19: Report details.