

BIONUMERICS Tutorial:

Extracting subsequences from (whole genome) sequences

1 Introduction

With the sequence extraction tools in BIONUMERICS, subsequences can be extracted from (whole genome) sequences and stored in separate sequence experiments. Any subsequence can be searched for (resistance gene sequences, virulence gene sequences, etc) and used for more in-depth study.

In this tutorial we will screen the whole genome sequences of some *Staphylococcus aureus* samples for the *mecA* sequence. 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

The whole genome demonstration database of *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).

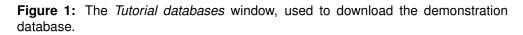
2.2 Option 1: Download demo database from the Startup Screen

1. 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).

- 2. Select the WGS_demo_database_for_Staphylococcus_aureus from the list and select Database > Download (2).
- 3. Confirm the installation of the database and press < OK > after successful installation of the database.
- 4. Close the *Tutorial databases* window with *File* > *Exit*.

🖆 Tutorial databases					- 0
File Database Tutorial Window					
Tutorial databases					
<u></u>					
⊥- Name	Size (MB)	Min I	Max	-	
BandScoring demo database	15	7	8	-	The DemoBase Connected demonstration
DemoBase Connected	4	7.5	8		database in BioNumerics version 8.0 format.
Neisseria MLST demo database	1	7.5	8		This database is used in several tutorials and
SNP demonstration database	0	7.5	8		comes with the BioNumerics 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		
				× .	
	<				1
Tutorials					
آگ ا					
Name				-	
Band matching and polymorphism analysis				~	This tutorial illustrates how to calculate a
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
					fingerprint data set and how to change the
Combined analysis of character data					layout of the obtained plots.
Combined analysis of character data Combined analysis of fingerprint data					
Combined analysis of fingerprint data					



The **WGS_demo_database_for_Staphylococcus_aureus** appears in the *BIONUMERICS Startup* window.

5. 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.

6. Download the file wgMLST_SAUR.bnbk file from https://www.applied-maths.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.

- 7. In the *BIONUMERICS Startup* window, press the button. From the menu that appears, select **Restore database...**.
- 8. Browse for the downloaded file and select *Create copy*. Note that, if *Overwrite* is selected, an existing database will be overwritten.

- 9. Specify a new name for this demonstration database, e.g. "Whole genome Staphylococcus aureus demobase".
- 10. Click < OK > to start restoring the database from the backup file (see Figure 2).

C:11	Jsers\Downloads\w	MIST SAUP1	bobk		Browse
0.1	Set 3 Do Willouds Wi	gincor_okorri	DIDK		browse.
Dest	ination				
	Overwrite				
0	Overwrite the curr coli with a previous since the backup w recovered.	version of the o	data as stored in a	backup. An	y new data
	Create copy				
۲	Create a new data have access to bo copy a database fr	in old and currer	it versions of an e	kup. Use thi: xisting datal	s option to base or to
New	/ database name:	Whole genom	e Staphyloco		

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

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

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

	is Scripts WGS tools	Window Help ②③ ④	>								
eriment types				Database	entries					Comparisons	
a + Pੈ ⊗		1 † 1	<aii e<="" th=""><th>£1</th><th>+ 🖻</th><th>⊗ €. €.</th><th>All Entries></th><th>3</th><th></th><th>+ 13 ⊗ €. 6 </th><th><ai< th=""></ai<></th></aii>	£1	+ 🖻	⊗ €. €.	All Entries>	3		+ 13 ⊗ €. 6	<ai< th=""></ai<>
a Name	Туре		-	Key		Organism name	Study title	ST info	outbreak 🔻 1 2 3 4 5 6 7 8	Name Modified date	
			•		00400						
	Sequence read	set types	Ŷ		103403	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin		part of outbrea	Neonatal MRSA study 2015-01-28 11:1 Pliot study 2015-01-28 12:1	
	Character types				103402	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin		part of outprod	Plot study 2015-01-28 12:1 WGS for MRSA outbreak 2015-01-28 12:2	
	Sequence types					Staphylococcus aureus	A neonatal MRSA outbreak investigation usin			WGS for MRSA outbreak 2015-01-28 12:2	23:44
	Character types				103394	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin		part of outbrea		
		set types			101899	Staphylococcus aureus		22	partorodorea		
			_			Staphylococcus aureus		22	prior to outbrea • • • • • •		
	Sequence types				103395	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin	5	prior • • • • • •		
8 wgsLong	Sequence read	set types			103400	Staphylococcus aureus		22	part of outbreal • • • • • • •	<	
			~		103405	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin		part of outbreal • • • • • • •	Identification projects Decision networks	
						Staphylococcus aureus	A neonatal MRSA outbreak investigation usin		part of outbrea • • • • • • •		
ields Database design						Staphylococcus aureus	A neonatal MRSA outbreak investigation usin		part of outbreal • • • • • • •	2 + 13 ⊗ 6. 6.	∽
13 ⊗ €.	6 v. 1	1	try fields>		159680	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin		during outbreat • • • • • • •	Name Modified date	
	E ~ T	4 KAIIEM	try heids>	ERR1	103397	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin	1	during outbreat • • • • • • •	Name Modified date	
Name				ERR1	103404	Staphylococcus aureus	A neonatal MRSA outbreak investigation usin	22	part of outbrea • • • • • •		
Crganism name	Fixed		^	ERR1	127429	Staphylococcus aureus	A plot study of rapid benchtop sequencing		S. aureus clust • • • • • •		
K Study title	Fixed			ERR1	127426	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust • • • • • •		
K ST info	Fixed			ERR1	127428	Staphylococcus aureus	A plot study of rapid benchtop sequencing		S. aureus clust • • • • • •		
e outbreak	Fixed			ERR1	127417	Staphylococcus aureus	A plot study of rapid benchtop sequencing		S. aureus clust • • • • • •		
R Patient ID	Fixed			ERR1	127408	Staphylococcus aureus	A pliot study of rapid benchtop sequencing		S. aureus clust • • • • • •		
K Study accession	Fixed			ERR1	127443	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S, aureus clust · · · · · ·		
K Instrument	Fixed			ERR1	127448	Staphylococcus aureus	A plot study of rapid benchtop sequencing		S. aureus clust • • • • • •		
MLST loci ST	Flexible				127432	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust • • • • • •	Algnments BLAST projects Chromosome con	maninar
woMLST loci ST	Flexible		_		127431	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust • • • • • •		Tyan sea
MLST PubMLST ST					127412	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust • • • • • •	+1030€,161 ₹	<4
MLST PUBMEST ST	Flexible				127427	Staphylococcus aureus	A plot study of rapid benchtop sequencing		S. aureus clust • • • • • •	Name Modified date	
					127444	Staphylococcus aureus	A plot study of rapid benchtop sequencing		S. aureus clust • • • • • •	mouned date	
			Ŷ		127413	Staphylococcus aureus	A plot study of rapid benchtop sequencing		S aureus clust		
print files Power assem	bies Assolutions					Staphylococcus aureus	A plot study of rapid benchtop sequencing A plot study of rapid benchtop sequencing		S. aureus clust		
					127442						
+ 🖻 🛛 🗎	3. 6 ~	<all fingerprin<="" td=""><td>nt files></td><td></td><td></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></all>	nt files>			Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust		
File name	Experiment type	Link			127409	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust		
	in provide the second sec		^		127433	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust • • • • • •		
			Ç		127415	Staphylococcus aureus	A pilot study of rapid benchtop sequencing		S. aureus clust • • • • • • • • •		
<			>	- <					> < >		

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

3 About the demonstration database

The whole genome demonstration database of *Staphylococcus aureus* contains links to sequence read set data on NCBI's sequence read archive (SRA) for 97 publicly available sequencing runs.

The sequence read set experiment type wgs contains the link with some raw data statistics.

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).

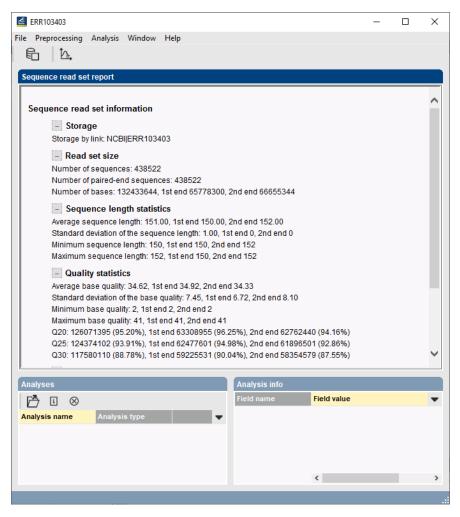


Figure 4: The sequence read set experiment card for an entry.

2. Close the Sequence read set experiment window.

The sequence experiment type **denovo** contains the results from the de novo assembly algorithm, i.e. concatenated de novo contig sequences.

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** 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. If not all panels are in place select *Window* > *Restore default configuration* to restore the default configuration of the *Sequence editor* window.

In this tutorial we will extract subsequences in batch from the sequences stored in the denovo

ERR127	7414 (Se	equence	Viewer))														_		×
e Seque	ence	Header	Anno	tation	View	Tools	Wind	low H	elp											
8) 🖂	ā I	ß	0	0	XS	ר ז ו	\otimes	Q	Þ			⊽1	▽2	⊲D	\otimes					
					00 <u>-</u>		0								0					
Sequenc	e Edito																			
agaaa	ctt	tt tt	acct	ttat	atg	ttcc	ggt	aaaa	ccaa	ac	at	gtt	tcg	ca	ctg	tati	tgaa	2280		^
ctgtt	cca	ca to	atcc	aaat	aag	tttc	tgc	aatg	faact	ta	get	tct	taa	cg	gat	ctc	ctgg	2340		
caata	gaad	cc gt	ttct	gcaa	ttt	cgac	gtc	atto	attg	gt	tta	aat	gtg	tg	gtg	ttga	attt	2400		
cataa	aata	at to	ctcc	caat	att	tgta	tag	tgat	ttgt	ct	ati	ttg	tct	аa	cca	tgta	atat	2460		
aataa	tati	tt tt	taaa	agcg	ctt	tcaa	gca	ttac	attt	tg	aca	aat	att	аa	gaa	cat	ttta	2520		
cagtt	acaq	gc ta	ctaa	cctt	gta	taaa	act	aatt	ttca	aa	gco	cta	tct	at	gtt	taat	tttt	2580		
tataa	ttt	tt co	aatt	cgag	taa	aaat	tat	tgta	tgat	ga	ga	gtg	att	аa	gta	act	ttAT	2640		
GGAGG	TTG	IC TA	CGTG	ACGC	AAA	CACA	ACC	AAAA	TTTT	ΤA	AC	GAT	ΤΤΑ	ТΑ	ATA	CGT	FATA	2700		
TAAAG	AAA	TT CA	ATAA	GGAA	AAT	ATCC	GAG	TGGC	CATG	CA	TT	ACC	TAC	ΤG	AAA	AGGI	AACT	2760		~
			1200		'1400 `			AUR_319		'20	000 .		'2200		24	100	'28	00	2800	
		<	' '1200		'1400					'20	000 .		' '2200		24	100				
Annotatio	on		' '1200		'1400 '					'2(000 .		' '2200		24	100				
Annotatio			'1200		'1400 ·			AUR_319		'20	000 .		'2200	+						3
Feature	list	<	'1200			±	s	AUR_319			000	%4	'2200				compl	SAUR_38		
	list	< Feat			<u>Q</u>]	±	× s	AUR_319		igth	000	%(+	×	ele='	compl '1"	SAUR_35	5	.24 /
Feature	list	Feat	ure key		Q] Start	±	S S S End	AUR_319	Len	igth	000 .	% • 32.) 37.)		+	/all	ele=' tag='	compl '1" 'SAUR_	ement(5	
Feature	list	Feat ¢¢	ure key → CDS		Q Start	±	⊗ End 168	AUR_319	Len 135	igth	000 .			+	/all ccus_ /evid	ele=' tag=' lence=	compl '1" 'SAUR_ =100.0	ement(319"	5	.24 /
	list 1 2	Feat	ure key CDS		Start 337 1696	±	⊗ End 168 240	AUR_319	Len 1353 708	igth	000 .	37.)	~ ^ ~	/1	/all ocus_ /evid /r	ele=' tag=' lence=	compl '1" 'SAUR_ =100.0 'fwd=0 ;cid=	ement(319" ;start denovo	3 (1696 = 1695; p_0"	.24 /
Feature	list 1 2	Feat	ure key > CDS > CDS > CDS > CDS		Start 337 1696 2639	±	End 168 240 339	AUR_319	Len 135: 708 756	igth	000 * 000	37.) 35.)	~ ^ ~	/1	/all ocus_ /evid /r	ele=' tag=' lence=	compl '1" 'SAUR_ =100.0 'fwd=0 ;cid=	ement(319" ;start denovo	5 (1696 ;=1695;	.24 /
Feature	list 1 2 3 4 5	< <tr> Feat \$\overline\$ \$\overline\$ \$\overline\$ \$\overline\$ \$\overline\$ \$\overline\$ \$\overline\$ \$\overline\$ \$\overline\$ \$\overline\$</tr>	ure key > CDS > CDS > CDS > CDS		Start 337 1696 2639	+	S End 168 240 339 431	AUR_319	Len 135: 708 756	ıgth 3		37.) 35.) 35.)	~	/1.	/all ocus_ /evid /r	ele=' tag=' lence=	compl '1" 'SAUR_ =100.0 'fwd=0 ;cid=	ement(319" ;start denovo	3 (1696 = 1695; p_0"	.24 /

Figure 5: The Sequence editor window.

sequence experiment. We will search for the *mecA* sequence.

5. Close the Sequence editor window.

Additional information, stored in entry info fields (Organism name, Instrument, Study accession, etc.) was collected from the corresponding publications and added to the demonstration database.



The wgMLST analysis settings and results (assembly-based calls and assembly-free calls) performed on the Applied Maths Calculation Engine are in depth discussed in the tutorial "wgMLST typing in the Staphylococcus aureus demonstration database" available on our website.

4 Extracting subsequences

4.1 Principles

The sequence extraction tools in BIONUMERICS use a BLAST approach to extract subsequences in batch from sequences stored in an *Origin* experiment type and saves the retrieved subsequences in a *Destination* experiment type. The BLAST search is based on a single *query sequence* per destination experiment type.

Before we can extract subsequences from the sequences stored in the *denovo* experiment, we first need to specify a query sequence in our demonstration database (see 4.2), and specify the sequence extraction settings (see 4.3).

4.2 Provide query sequence

A FASTA formatted text file can be found on our website, containing the *mecA* sequence extracted from the reference sequence with accession number **BX571856.1** (see Figure 6). The example file can be found on the download page on our website (https://www.applied-maths.com/ download/sample-data, "mecA reference sequence").



Figure 6: mecA reference sequence.

In order to use this sequence as query sequence we first need to import this sequence in our demonstration database.

- 1. Select File > Import... (, Ctrl+I) to open the Import dialog box.
- 2. Choose the option *Import FASTA sequences from text files* under the *Sequence type data* item in the tree and click <*Import*>.
- 3. Press < Browse>, navigate to the downloaded file, select the Reference.txt file and press < Open>.
- 4. With the option *Preview sequences* checked, press <*Next*>.

The import wizard now displays a preview of the sequence data in the FASTA file. From this preview, it is clear that the first (and only) FASTA field contains the name of gene (see Figure 7). We will use this name for our sequence experiment in our database.

5. Press <*Next*>.

The next step of the import wizard lists the templates that are present to import sequence information in the database. As this is the first time we import FASTA formatted sequences in the database, we need to create a new import template by specifying *Import rules*.

- 6. Click < *Create new*> to create a new import template.
- Select *Field 1* in the list and click < *Edit destination*> or simply double-click on "Field 1". Select *Sequence type* from the list and press < *OK*>.
- 8. Scroll down the list in the grid using the scroll bar on the right and select the last row in the grid, *File Name*, and press <*Edit destination*>. Choose *Key* and press <*OK*>.
- 9. Press < *Preview* > to obtain a preview of the data you are about to import (see Figure 8).
- 10. Close the preview and click <*Next*> and <*Finish*>.
- 11. Specify a template name, e.g. **FASTA**, and optionally enter a description. Press < **OK** >.

nport se	quences			?	×
	nces preview uble-click on a seque	ence to open in the	Sequence Viewer.		
Nr.		Length	Header		
⊻ 1	Reference	2007	mecA		

Figure 7: Preview.

revie	w		?	×
Nr.	Sequence type	Кеу		
1	mecA	Reference		
			Clos	se



- 12. Highlight the newly created template (see Figure 9) and press < *Next*>.
- Press < Yes > twice to confirm the creation of the mecA sequence experiment in the database (see Figure 10).

The Database links wizard page will indicate that 1 new entry will be created during import.

14. Press < *Finish*> to start the import into the database.

An entry with key *Reference* is created in the database and the sequence from the text file is linked to the *mecA* sequence experiment.

Import template	IS:					
<default> FASTA</default>		FASTA			Create new	
				[Edit	
				[Preview	
				[Сору	
kperiment type:	< From imp	ort template>	~			

Figure 9: Import template.

New experiment types	?	×
These 1 new experiment types	s will be cre	ated:
mecA		
Do you want to continue?		
-		
Yes	N	o

Figure 10: Add new sequence type to the database.

4.3 Specify sequence extraction settings

15. Select *Analysis* > *Sequence types* > *Extract sequences* > *Settings...* in the *Main* window to call the *Sequence extraction settings* dialog box.

This dialog box gives access to *Sequence extraction* settings per sequence experiment type and the general *Reports* settings. Initially, the tree control on the left will be empty.

- 16. Press the <*Add*> button to call the *Extract gene from genome* dialog box.
- 17. Select *denovo* as *Origin experiment type* (see Figure 11). This is the sequence experiment, containing the whole genome sequences, that will be screened and from which a subsequence will be copied from.
- 18. Check the *mecA* experiment as *Destination experiment type* (see Figure 11).
- 19. Press < *Next*> to call the second step of the wizard.

The *Search sequence* is what the BLAST algorithm will use to screen the origin experiment type (here *denovo*) for. In our demonstration database, entry with key *Reference* contains the query sequence, stored in the *Destination experiment type*.

20. Press < *Pick* > to open the *Select entry* dialog box.

Similarity-based sequence e			
	ence from an origin experime	nt type (usually a	
genomic sequence) into a d	estination experiment type.		
Origin experiment type:	denovo	\sim	
Destination experiment type(s):	denovo		
	SNP outbreak		
	M meca		

Figure 11: Specify the origin and destination sequence types.

21. Scroll down the list, highlight *Reference* and press *<OK>*. Alternatively, start typing the text *Reference* in the *Search for* text box, highlight the reference entry and press *<OK>*.

The **BLAST settings** include two thresholds that a BLAST hit should fulfill in order to be considered:

- A *Minimum sequence identity (%)* between the search sequence and the matched subsequence in the origin sequence experiment, expressed as a percentage.
- A *Minimum length for coverage (%)*, i.e. a minimum overlap between the search sequence and the matched subsequence.

In case more than one BLAST result is found that fulfills both criteria, the best match will be copied to the destination experiment.

Optionally, the length of the extracted sequence can be corrected (see *Extracted sequence correction* options).

22. For this exercise, make sure the *Reference* entry is specified as query entry, leave the other settings at their defaults and press <*Next*> (see Figure 12).

The tree in the *Extract gene from genome* dialog box is updated (see Figure 13).

Default report settings will be applied when running a report (see 4.4), but can be modified by highlighting *Reports* in the tree and pressing *< Edit >*.

23. Press < OK > to close the *Extract gene from genome* dialog box.

4.4 Sequence extraction analysis

Now that we have specified the sequence extraction settings (see 4.3), we can now start the actual sequence extraction process.

24. In the *Database entries* panel of the *Main* window select all entries with *Edit* > *Select all* (Ctrl+A) and unselect entry with key *Reference*.

Settings		
Please enter the BLAST settings and determine how to handle the extra sequence.	cted	
Search sequence		
Pick entry with search sequence:		
Reference Pick		
BLAST settings		
Minimum sequence identity (%) 80		
Minimum length for coverage (%) 95		
Extracted sequence correction		
Take best BLAST hit		
O Correct BLAST hit based on CDS		
O Correct BLAST hit based on PCR primers		

Figure 12: Sequence extraction settings.

Sequence extraction settings	? >	<
Gequence extraction mecA	Extract sequences using a similarity based method.	
Add Edit Remove	OK Cancel	

Figure 13: Sequence extraction settings.

The status bar, displayed at the bottom of the *Main* window will indicate that **97** entries are selected.

25. Select Analysis > Sequence types > Extract sequences > Analyze or use the Process data dialog box: select File > Process... (→), highlight Extract sequences under Sequence type and press < OK >.

A progress bar appears. The complete analysis may take up to several minutes. When the analysis is finished, the question "Do you want to open the reports?" pops up.

26. Press < **Yes**> to open the *Report* window. Alternatively, a sequence extraction report can be opened for the selected entries with *Analysis* > *Sequence types* > *Extract sequences* > *Show reports*.

The *Report* window displays a summary of the extraction results (see Figure 14).

The *Report* window contains a gene extraction report for each of the selected entries. For each destination experiment type ('Locus') that has sequence extraction settings (here: only one), it is indicated whether or not a BLAST hit was found, its position on the origin sequence ('Start' and

🖆 Report Window	w					-					-		×
File Entries Rep	oort Wind	dow Help											
Entries		ERR103402 - S	equence extraction	i report									
Key	-	<default report<="" td=""><td>></td><td>U</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></default>	>	U									
ERR103403	^	1 E											
ERR103402													- 11
ERR103398													- 1
ERR103394													- 1
ERR101899		FRR1	03402										- 1
ERR103396			03402										- 1
ERR103395		Sequence ex	traction report										- 1
ERR103400													- 1
ERR103405													- 1
ERR101900													- 1
ERR103401			Date	05/18/20 14:55:16									- 1
ERR159680			Name										- 1
ERR103397				Entributue									- 1
ERR103404													- 1
ERR127429													- 1
ERR127426		Similar	itv-based o	ene extractio	on								- 1
ERR127428		ommu	ity-busou g										- 1
ERR127417			Result			BLAST							
ERR127408										-	-		_
ERR127443		Locus	Found	Start	End	Identity (%)	Length (%)	Ref. Length	Mismatches	Open gaps	Start		En
ERR127448		mecA	Yes	2034299	2036305	99.95	100.00	2007	1	0	2034299	203	3630
ERR127432													
ERR127431	~												-
 ✓ < 	>	•											÷.

Figure 14: Summary of the sequence extraction results.

'Stop'), sequence identity ('Identity (%)') and sequence overlap ('Length (%)'). Furthermore, the length of the retrieved subsequence is reported ('Ref length'), the number of mismatches with the query sequence ('Mismatches'), number of gaps ('Open gaps') and length correction applied.

27. Close the *Report* window.

The extracted sequences - if found - are stored in the *mecA* sequence experiment type.

28. Clicking on a green colored dot in the *Experiment presence* panel for a *mecA* experiment will open the *Sequence editor* window, containing the extracted sequence (see Figure 15).

C ERR103403 (Sequence Viewer)	-		×
File Sequence Header Annotation View Tools Window Help			
□ ∞ ြ ାଠାତା ୪ ଥିଥି⊗ ଠା଼ ▷ === ▽1 ▽2 ▽> ⊗			
Sequence Editor			
ttattcatct atatcgtatt ttttattacc gttctcatat agctcatcat acactttacc	: 60		^
tgagattttg gcattgtagc tagccattcc tttatcttgt acatctttaa cattaatago	: 120		
catcatcatg tttggattat ctttatcata tgatataaac cacccaattt gtctgccag	: 180		
ttctccttgt ttcattttga gttctgcagt accggatttg ccaattaagt ttgcataaga	a 240		
totataaata tottotttat gtgttttatt tacgaottgt tgoatacoat cagttaatag	j 300		
attgatattt totttggaaa taatattttt ottocaaact ttgtttttog tgtottttaa			
taagtgaggt gcgttaatat tgccattatt ttctaatgcg ctatagattg aaaggatctg	420		~
		_	-
Sequence Viewer			
<pre></pre>			>
Annotation			
Feature list			
→ W Feature key Start End →			
\rightarrow \diamond			
Annotation Header Custom Fields Sequence Search Contigs Frame Analysis Restriction Analysis			
Sequence: ERR103403 Experiment: mecA 1 2007 bp			

Figure 15: The Sequence editor window.

29. Close the Sequence editor window.



In the tutorial "MLST analysis starting from whole genome sequences" available on our website, the sequence extraction tools are used to extract the seven MLST house-keeping genes from a set of whole genome sequences of *Listeria monocytogenes*. Follow-up analysis is illustrated in combination with the *MLST online plugin*.