

BIONUMERICS® version 8 - PLUGINS



Spa typing plugin

Contents

1	5 51	5
2	Getting started12.1Browsing Spa repeats or types12.2Updating Spa types and Spa repeats1	
3	Importing and assembling trace files13.1Importing and assembling trace files in batch13.2Reports1	
4	Checking assemblies in Assembler24.1Introduction24.2Showing Spa repeats on the consensus24.3Showing the repeat succession plot24.4Changing the status of error (and warning) messages24.4.1Principles24.4.2Option1: Changing the status in Assembler24.4.3Option2: Changing the status in the Detailed report window2	22 24 26 26 27
5	Spa-Typing in BIONUMERICS 2 5.1 Selections in the main window 2 5.2 Assigning Spa types 2 5.2.1 Principles 2 5.2.2 Step 1: The assembly is screened for repeats 3 5.2.3 Step 2: Repeat type (if available) is assigned to each selected entry 3	29 29 30
6	6.1Introduction36.2The Comparison window36.3Creating a cost matrix36.4Cluster analysis settings36.5Minimum spanning tree3	33 34 36
7	7.1 Selections in the main window	41 41 41
8	· · · · · · · · · · · · · · · · · · ·	15 15

8.2	SpaSe	erver synchronization settings 47
	8.2.1	Introduction
	8.2.2	Add SpaServer users
	8.2.3	Link BIONUMERICS information fields to SpaServer fields
	8.2.4	Store SpaServer results in BIONUMERICS database fields
8.3	Synch	ronizing with SpaServer (batch mode)
8.4	Synch	ronizing with SpaServer (entry mode)

NOTES

SUPPORT BY APPLIED MATHS

While the best efforts have been made in preparing this manuscript, no liability is assumed by the authors with respect to the use of the information provided.

Applied Maths will provide support to research laboratories in developing new and highly specialized applications, as well as to diagnostic laboratories where speed, efficiency and continuity are of primary importance. Our software thanks its current status for a part to the response of many customers worldwide. Please contact us if you have any problems or questions concerning the use of BIONUMERICS[®], or suggestions for improvement, refinement or extension of the software to your specific applications:

Applied Maths NV Keistraat 120 9830 Sint-Martens-Latem Belgium PHONE: +32 9 2222 100 FAX: +32 9 2222 102 E-MAIL: BE-DAU-INFO@biomerieux.com URL: https://www.applied-maths.com

Applied Maths, Inc.

11940 Jollyville Road, Suite 115N Austin, Texas 78759 U.S.A. PHONE: +1 512-482-9700 FAX: +1 512-482-9708 E-MAIL: US-DAU-INFO@biomerieux.com

LIMITATIONS ON USE

The BIONUMERICS[®] software, its plugin tools and their accompanying guides are subject to the terms and conditions outlined in the License Agreement. The support, entitlement to upgrades and the right to use the software automatically terminate if the user fails to comply with any of the statements of the License Agreement. No part of this guide may be reproduced by any means without prior written permission of the authors.

Copyright ©1998-2020, Applied Maths NV. All rights reserved.

BIONUMERICS[®] is a registered trademark of Applied Maths NV. All other product names or trademarks are the property of their respective owners.

BIONUMERICS^(R) uses following third-party software tools and libraries:

- Python 2.7.4 release from the Python Software Foundation, http://www.python.org/
- Xerces library for XML input and output from the Apache Software Foundation, https://xerces.apache.org/
- NCBI toolkit version 2.2.28, http://www.ncbi.nlm.nih.gov/BLAST/
- SRA Toolkit, https://ncbi.github.io/sra-tools/
- Boost c++ libraries, http://www.boost.org/
- Samtools for interacting with SAM / BAM files, http://www.htslib.org/download/
- 7-Zip (7za.exe), http://www.7-zip.org/
- Zlib library, https://zlib.net/
- Pigz for parallel gzip compression, https://zlib.net/pigz/
- Cairo 2D graphics library version 1.12.14, http://cairographics.org/
- Crypto++ library version 5.5.2, http://www.cryptopp.com/
- OpenSSL library, https://www.openssl.org/
- libSVM library for Support Vector Machines, http://www.csie.ntu.edu.tw/~cjlin/libsvm/
- SQLite version 3.7.17, http://www.sqlite.org/
- pymzML Python module for high throughput bioinformatics on mass spectrometry data, https://github.com/pymzml/pymzML
- Numpy Python library version 1.8.1, http://www.numpy.org/
- BioPython Python library version 1.64, http://www.biopython.org/
- PIL Python library version 1.1.7, http://www.pythonware.com/products/pil/
- Chromium Embedded Framework, https://bitbucket.org/chromiumembedded/cef/wiki/ Home
- SPAdes genome assembler version 3.13.1, http://bioinf.spbau.ru/spades *
- SKESA version 2.3.0, https://github.com/ncbi/SKESA/releases
- Unicycler version 0.4.8, https://github.com/rrwick/Unicycler/releases *
- Velvet for Windows, source code can be downloaded from https://www.applied-maths.com/download/open-source
- Ray for Windows, source code can be downloaded from https://www.applied-maths.com/download/open-source
- Bowtie2 version 2.2.5 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml)*
- SNAP version 1.0.18, http://snap.cs.berkeley.edu/
- RAxML version 8.2.11, https://github.com/stamatak/standard-RAxML/releases
- FastTree version 2.1.10, http://www.microbesonline.org/fasttree/

- CFSAN SNP pipeline version 0.8.2, https://github.com/CFSAN-Biostatistics/snp-pipeline
- Prokka version 1.12, https://github.com/tseemann/prokka *
- *: On Calculation Engine only

Chapter 1

Starting and setting up BIONUMERICS

1.1 Introduction

This guide is designed as a tutorial for the *Spa typing plugin*. This plugin offers extra functionality to BIONUMERICS to do Spa typing for *Staphylococcus aureus*. Sequences in the database can be screened for known spa repeats and types downloaded from the SpaServer, data can be submitted to the SpaServer via a synchronization process, and entries can be clustered based on the spa types.

The features of the plugin will be illustrated using data available on the Applied Maths website (https://www.applied-maths.com/download/sample-data, click on "SPA typing data files"). The *Spa typing plugin* is supported in the **BIONUMERICS-SEQ** and **BIONUMERICS-SUITE**.

1.2 Startup program

Make sure the latest version of BIONUMERICS is installed (https://www.applied-maths.com/ download/software). The installation manual can be downloaded from https://www.applied-maths. com/download/manuals.

When BIONUMERICS is launched from the Windows start panel or when the BIONUMERICS

shortcut (**L**) on your computer's desktop is double-clicked, the **Startup program** is run. This program shows the *BIONUMERICS Startup* window (see Figure 1.1).

A new BIONUMERICS database is created from the Startup program by pressing the 🗟 button.

An existing database is opened in BIONUMERICS with 🗐 or by simply double-clicking on a database name in the list.

1.3 Creating a new database

3.1 Press the BIONUMERICS *BIONUMERICS Startup* window to enter the *New database* wizard.

3.2 Enter a name for the database, and press < *Next*>.

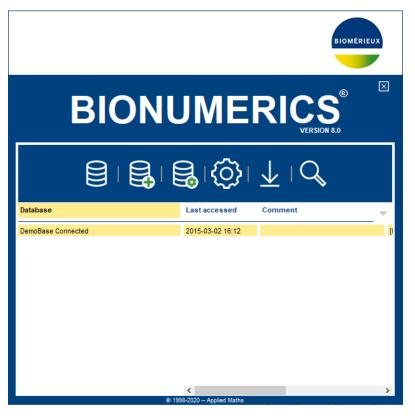


Figure 1.1: The BIONUMERICS Startup window.

A new dialog box pops up, prompting for the type of database (see Figure 1.2).

3.3 Leave the default option selected and press < Next>.

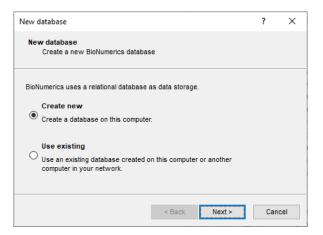


Figure 1.2: The New database wizard page.

A new dialog box pops up, prompting for the database engine (see Figure 1.3).

3.4 Leave the default option selected and press $<\!\!\textit{Finish}\!>$ to complete the setup of the new database.

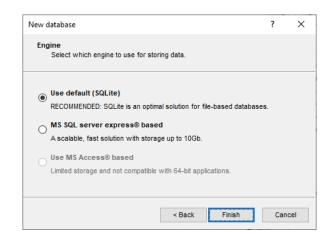


Figure 1.3: The Database engine wizard page.

1.4 Installing the Spa typing plugin

The *Plugins* dialog box can be called from the *Main* window by selecting *File* > *Install* / *remove plugins...* (,) (see Figure 1.4).

Activate fun	ctionality f	or specific	applications.				
Antibiotics BandScorir E. coli func Listeria fun MTBC func Salmonella S. aureus f HIV Resiste MIRU-VNTF MLST onlin QIAxcel RDP RiboPrinter SNP calling Sopa typing Polymorphi MLVA WGS tools	ig tional genc ctional genc tional genc functional unctional g ince e	otyping typing typing genotyping genotyping	-	the import of a on MIC or disk	l provides extra fu ntibitic susceptib diffusion method. te range into S-LR	ility data based The tools will	
	h, a stivate						
Show or Show Mar	-	ed					

Figure 1.4: The Plugins dialog box.

When a particular plugin is selected from the list of plugins, a short description appears in the right panel.

A selected plugin can be installed with the *Activate*> button. The software will ask for confirmation before installation. Some plugins are only supported in specific BIONUMERICS configurations. If the plugin is not supported by your BIONUMERICS configuration, it cannot be installed and an error message will be generated.

Once a plugin is installed, it is marked with a green V-sign. It can be removed again with the *<Deactivate*> button.

If the selected plugin is documented, pressing < Show Manual > will open its manual in the Help

window.

- 4.1 Select the *Spa typing plugin* from the list in the *Applications tab* and press the *Activate*> button.
- 4.2 The program will ask to confirm the installation of the plugin. Press <*OK*> twice to confirm the installation.

The Spa typing settings dialog box pops up (see Figure 1.5).

Spa typing settings		?	×			
General Spa typing plugin settings. Normally NO changes are required!						
Experiment Settings						
Repeat succession:	Spa-repsuc					
Sequence	Spa-typing					
Start target:	RCAMCAAAA					
Stop target:	TAYATGTCGT					
	Advanced assembly settings					
Type Detection Set	tings					
Allow IUPAC						
Allow gaps						
Max # of mismatches	(2-4): 2					
Information Fields						
Spa type:	SpaType 🗸					
Repeats succession:	RepeatSuccession ~					
Kreiswirth succ.:	<none> ~</none>					
Clonal complex:	<none> ~</none>					
Update URL	Update URL					
Repeats: http://spa	.ridom.de/dynamic/sparepeats.fasta					
Types: http://spa	Types: http://spa.ridom.de/dynamic/spatypes.txt OK					
Update automatically when database is opened Cancel						

Figure 1.5: The Spa typing settings dialog box.

Experiment Settings:

- The sequence type **Spa-typing** is automatically created upon installation of the *Spa typing plugin*, and will be used for the storage of the imported sequences (*Sequence*).
- The Spa repeat successions are stored in the character type experiment **Spa-repsuc** (*Repeat succession*).
- The start and stop trim patterns are automatically filled out in the *Start* and *Stop target* boxes respectively, but can be changed if desired.



If you want to submit sequences to the online SpaServer (see 8), the sequences must contain the following signatures: 5' signature: RCAMCAAAA, 3' signature: TAYATGTCGT.

When pressing the <*Advanced Assembly settings*> button, the *Assembly settings* dialog box pops up (see Figure 1.6).

The Assembly settings are grouped per settings dialog box in Assembler: *Quality Assignment*, *Calculate Assembly*, and *Consensus Determination*. For a detailed description of the Assembler program settings, see the reference manual.

Assembly settings	? ×
Quality assignment Curve sliding window (bp) 5 Min. good/bad peak ratio 1.30 Min. short/long peak distance ratio 0.50 Base calling sliding window (bp) 41	Calculate assembly Min. match word size (bp) 7 Min. score (matches) 30.0 Unit penalty per gap (x match) 12.0 Unit penalty per mismatch (x match) 3.0
Min. resolved positions (bp) 25	Maximum number of gaps 45
Min. consecutive good bases (bp) 6	Consensus determination
Min. length of usable sequence (bp) 50 Min. fraction of good bases (%) 20	Req. bases to include (%) 51 Consensus for unique base (%) 50
Note: The Spa default 'Quality assignment' settings are required for submission of new types to the Ridom/SeqNet SpaServer.	Consensus for 2-fold degen. (%) 70 Consensus for 3-fold degen. (%) 90 Spa defaults OK Cancel

Figure 1.6: The Assembly settings dialog box for Spa Typing.



The Assembly settings can still be changed after installation of the plugin with *Spa-Typing* > *Settings*.



The default Spa *Quality Assignment* settings are required for submission of new types to the SpaServer (see 8). Pressing the *<Spa defaults>* button will reset all settings in to their defaults.

Type Detection Settings:

- **Allow IUPAC**: When this option is enabled, the tool will consider the different possibilities for the ambiguous positions for the repeat calling in Assembler. (see 4.2). This option is enabled by default.
- *Allow gaps*: When this option is checked, gaps are allowed when searching for possible repeats in the consensus sequence.
- *Maximum number of mismatches*: In the *Spa typing plugin*, a visualization tool is available (see 4.3) with editing suggestions for the unknown repeat(s). With this option, you can specify the maximum number of mismatches you want to consider between the source sequence and the repeat sequence. The maximum value is 4. Entering a higher number will cause the value to be set to 4.

Information Fields:

In the *Information fields panel*, you can choose the names of the database information fields that will contain the *Spa type*, the *Repeat succession*, the *Kreiswirth succession* string, and *Clonal complex* information for the entries in the database (see Figure 1.5). You can choose the default suggested names, select an existing field, enter a new field name or set the box to "None".



The storage of a repeat succession in an information field is used for illustration purposes only. Long repeat successions may be truncated due to size limitations of the information field. The repeat information stored in the associated character type will be used when using the matching and clustering tools.



If you want to change the name of one of the information fields selected in the *Information fields panel*, you need to rename the information fields in the database **and** in the *Information fields panel* in order to run the plugin tool properly.



A new information field can not start with a space.

Update URL:

- The URL for the update of the *Repeats* and the *Types* can be changed in the *Update URL panel*.
- When the option **Update automatically when database is opened** is checked, the software will automatically update all online repeat and type information each time the database is opened.
 - 4.3 Leave all settings unaltered and press < OK >.

The software updates the types and repeats from the Ridom/Seqnet SpaServer.

- 4.4 When the *Spa typing plugin* is successfully installed, a confirmation message pops up. Press < OK > twice.
- 4.5 Press < *Proceed*> (or <*Exit*>) to close the *Plugins* dialog box and to continue to the *Main* window.
- 4.6 Close and reopen the database to activate the features of the Spa typing plugin.

The *Spa typing plugin* installs itself in a menu of the BIONUMERICS software. In the *Main* window, some initialization has been done to the database with the installation of the *Spa Typing plugin* (see Figure 1.7).

		×
Spa typing plugin - BioNumerics		- 🗆 ×
File Edit Database Analysis Spa-Typing Scripts Window Help		
🕼 主 🏢 🔎 🕚 Assign Spa types		
Experiment types Match Spa types		Comparisons
Settings		
Browse repeats/types	ି 🗞 । 🔓 🗠	+ 💾 🛇 🗟 🔂 🖓 🖓 🖓 + 🖓
* Raine	Level Modified date SpaType RepeatSuccession 👻 1 2	Name Modified date Level 🔻
1 Spa-typing Update repeats/types (SpaServer) 2 Spa-repsuc		
Synchronize with SpaServer		
SpaServer synchronization settings		
	_	
Entry fields Database design		Identification projects Decision networks
+ 🗗 ⊗ 🛱, 🛍 🖳 ↑ ↓ «Alle		? 🕄 + 🗁 ⊗ 🔩 🛍 🗠 ≪Allide
Name Field type 🗨		Name Modified date 🗨
All SpaType Fixed		
ANC RepeatSuccession Fixed		
v		
Fingerprint files Power assembles Annotations		Alignments BLAST projects Chromosome comparisons
円 + P ⊗ B F マ. <al fingerpi<="" td=""><td></td><td>+ 🖓 🖉 🗞 🗗 🖾 🖾 All Alignments:</td></al>		+ 🖓 🖉 🗞 🗗 🖾 🖾 All Alignments:
File name Experiment type Link 👻		Name Modified date 🗸
< > <	>	
Database: Spa typing plugin (_DefaultUser_) Entries: Loaded=0, View=0, Selected=0 2 ex	xperiments C.VUsers\Public\Documents\BioNumerics\Data 80\Spa typing plugin This is a time finited package valid until 2020-12-30	

Figure 1.7: The Main window after installation of the Spa Typing plugin.

- The information fields specified in the *Information fields panel* of the *Spa typing settings* dialog box are present in the *Database entries* panel.
- The *Experiment types* panel lists the experiment types. BIONUMERICS has automatically created a sequence type called **Spa-typing** and a character type called **Spa-repsuc** upon installation of the plugin.
 - 4.7 To call the *Spa typing settings* dialog box from the *Main* window, select *Spa-Typing* > *Settings*.
 - 4.8 Press the <*Advanced Assembly settings*> button to call the *Assembly settings* dialog box for the **Spa-typing** sequence experiment.

Chapter 2

Getting started

2.1 Browsing Spa repeats or types

The lists of Spa repeats and Spa types, downloaded from the SpaServer, can be queried by the user.

1.1 Select *Spa-Typing* > *Browse repeats/types* in the *Main* window.

This action calls the *Browse types/repeats* dialog box (see Figure 2.1).

rowse types	/repeats	?	×	
Repeats/T	ypes			
Browse: I repeats				
⊖ types				
Browse				
Repeat	Sequence	KW	^	
r01	GAGGAAGACAACAAGACTAGC			
r02	AAAGAAGACAACAAAAAACCTGGC	A1		
r03	GAGGAAGACAATAACAAACCTGGT	D2		
r04	GAGGAAGACAATAACAAGCCTGGT	Z1		
r05	AAAGAAGACAACAAAAAGCCTGGC	C1		
r06	AAAGAAGACGGCAAAAAACCTGGC	G2		
r07	GAGGAAGACAACAACAACCTGGT	U1		
r08	GAGGAAGACAACAACAAGCCTGGT	X1		
r09	GAGGAAGACGGCAACAAACCTGGT	A2		
r10	AAAGAAGACAATAACAAGCCTGGT	C2	\checkmark	
Add	Edit	Find	1	
		Clos	e	

Figure 2.1: The Browse types/repeats dialog box.

In the Repeats/Types panel specify which list you want to browse: the repeats list or types list.

In the Browse panel, all repeats/types are listed that were downloaded from the online SpaServer.

Use the scroll bar to browse through the repeats/types.

Select repeats/types in the *Repeats/Types panel* and press the *<Find>* button to look for a repeat/type.

When looking for a repeat, enter the sequence in the *Find type* dialog box and press the *<Find>* button. If the repeat is present in the list of repeats, the *Repeat ID* of the sequence is displayed.



The *Kreiswirth* information box is only shown in the *Find type* dialog box if its corresponding information field is present in the database.

Find repeat	?	×
Repeat ID:	r07	
Kreiswirth:	U1	
Sequence:	GAGGAAGACAACAACAAAC	стс
	Find Close	•

Figure 2.2: The Find type dialog box: Find repeat.

Find type	? ×
Type ID:	t007
CC:	
Succession:	15-12-16-16-16-02-25-17
	Find Close

Figure 2.3: The Find type dialog box: Find type.

When looking for a type, enter the succession string in the *Find type* dialog box and press the *<Find>* button. If the succession string is present in the list of types, the *Type ID* of the succession string is displayed.



The *Clonal complex* information box is shown in the *Find type* dialog box if its corresponding information field is present in the database.

With the < Add> button, a new type can be added to the list of existing repeats/types.

Add Spa type	?	×
Type ID:		
Succession:		
CC:		
	ОК	
	Canc	el

Figure 2.4: The Add Spa type dialog box.

Enter a *Succession* string, a *Type ID* and optional a clonal complex notation in the *Add Spa type* dialog box.

When pressing the $\langle OK \rangle$ button the type is added to the list of types.



When updating the spa repeat and spa type information from the SpaServer, types that were added manually to the database might be overwritten if the *Type ID* entered in the *Add Spa type* dialog box corresponds to a Type ID of a new online spa type.

Select a repeat/type in the *Browse panel* and press the *<Edit>* button to add/edit the Kreiswirth or clonal complex information.

If an information field name for the *Kreiswirth* information is specified in the *Information fields* panel of the *Spa typing settings* dialog box, Kreiswirth information can be linked in the *Edit Spa repeat* dialog box to the spa repeats that are present in the database.

If an information field name for the *Clonal complex* information is specified in the *Information*

Edit Spa repe	at	?	×
Repeat ID:	r04		
Sequence:	GAGGAAGA	CAATAACAA	GCCTG
Kreiswirth:	Z1		
		OK	
		Cano	el :

Figure 2.5: The Edit Spa repeat dialog box.

Edit Spa type	? ×
Type ID:	t002
Succession:	26-23-17-34-17-20-17-12-17-16
CC:	
	ОК
	Cancel

Figure 2.6: The Edit Spa type dialog box.

fields panel of the *Spa typing settings dialog box*, clonal complex information can be linked in the *Edit Spa type* dialog box to the spa types that are present in the database.

Clonal complex information that is stored outside BIONUMERICS (for example in an Excel or text file) can be imported into BIONUMERICS and linked to the spa types with a special script. Please contact Applied Maths to obtain this script.

It is also possible to view all repeats and types stored in the database with an object query.

- 1.2 In the *Main* window, select *Database* > *Object queries...* (III) and select "<Create new>" from the drop-down menu that appears. Press < OK >.
- 1.3 As *Object to report*, select "Spa repeats" or "Spa types" and press *OK* (see Figure 2.7).

For more information on object queries, we refer to the reference manual.

2.2 Updating Spa types and Spa repeats

When installing the *Spa typing plugin*, the known repeats and types are downloaded from the SpaServer.

2.1 The list of known repeats and types can be updated with the command *Spa-Typing* > *Update repeats/types (SpaServer)*.



You can perform analyses without being connected to the internet, but you will be unable to update the list of known repeats and types.

When the option **Update automatically when database is opened** is checked in the Spa typing settings dialog box, BIONUMERICS automatically updates all online repeat and type information each time the database is opened.

🖆 ОЬ	oject	t query			-		\times
File Q	Quer	ry Ob					
6	I		$\otimes \blacksquare \psi \gg arsigma$				
Objec	ct q	uery			Pare	nt obje	
		ра гере	eats	^			
+					+		
\otimes		rtopour			\otimes		
		<u>Repeat</u>	sequence				
		Kreiswi	irth notation		AND		
OR OR		0	ram		DR .		
O NOT		Owner	[Al]		O NOT		
		Shared		~			
	<			>		<	>
Objec	ct li	st					
전	Ć	96	2				
		eat ID	Repeat sequence Kreiswirth notation Owner			Shared	-
l r	01		GAGGAAGACAACAACAAGCCTAGCDefaul			Yes	^
	02		AAAGAAGACAACAAAAAACCTGGC A1 _Defaul	tUser_		Yes	
	03		GAGGAAGACAATAACAAACCTGGT D2 _Defaul	_		Yes	
	04		GAGGAAGACAATAACAAGCCTGGT Z1 _Defaul	_		Yes	
	05 06		AAAGAAGACAACAAAAAAGCCTGGC C1Defaul AAAGAAGACGGCAAAAAACCTGGC G2Defaul	_		Yes	
	07		GAGGAAGACAACAACAACCTGGT U1Defaul	_		Yes	
n 🗆 r	08		GAGGAAGACAACAACAAGCCTGGT X1Defaul	tUser_		Yes	
	09		GAGGAAGACGGCAACAAACCTGGT A2 _Defaul	tUser_		Yes	~
	<					>	•
802 obje	ects						_
002 00j0							
_	_	t querv	······		_		: ×
🖆 Ob	oject				-		 ×
E Ob	oject Quer	ry Obj	ject Window Help		-		.:: ×
🖆 Ob	oject Quer				-		.: ×
E Ob	oject Quer	ry Obj	ject Window Help		- Parei	nt obje	: ×
File Q Dbjed	oject Quer ct q	ry Obj	iject Window Help ⊗ Ⅲ ψ> ∳> ▷	^			×
File Q Cobject Cobject +	oject Quer ct q 	ny Ob uery pa type	iject Window Help ⊗ Ⅲ ψ> ∳> ▷	^	+		×
File Q Dbjed	ject Quer Ct q S	ny Ob uery pa type Type ID	iject Window Help ⊗ I III v V V V es [Alī] ✓	^			×
ie Objec File Q Objec + ⊗	ject Quer Ct q S	ny Ob uery pa type Type ID	iject Window Help ⊗ I III v V V PS	^	+		×
File Object)ject Quer Ct q S	ry Objuery uery pa type Type ID Repeat	iject Window Help ⊗ I III v V V V es [Alī] ✓	^	+ ⊗ "×0		×
E Objec File O Objec + &	Quer Quer Ct q S	ry Ob uery pa type Type ID Repeat	iject Window Help	^	+		×
File Object	Quer Quer Ct q S	ry Ob uery pa type Type ID Repeat	iject Window Help	^	+ ⊗ "×0		×
E Objec File O Objec + &	ject Quer Ct q S	ry Ob uery pa type Type ID Repeat Clonal C	iject Window Help	^	+		×
E Objec File O Objec + &	ject Quer Ct q S	ry Ob uery pa type Type ID Repeat Clonal C Owner Shared	iject Window Help	^	+		×
Constant of the second	oject Quer S	ry Ob Juery pa type Type ID Repeat Clonal C Owner Shared	iject Window Help		+	nt obje	
E Objec File O + ⊗ @ a a objec	oject Quer S S Ct li	ny Ob uuery pa type Type ID Repeat Clonal C Owner Shared	ject Window Help ⊗ I I I I I I I I I I I I I I I I I I I		+	nt obje	
e objec File 0 + ⊗ @ a s objec P	oject Quer s s ct q	ry Obj i i i i i i i i i i i i i i i i i i i	ject Window Help ⊗ I	>	+	nt obje	
Cobject Cob	oject Quer Ct q S S Ct li Ct li Ct li Ct li	ry Obj pa type pa type ID Repeat Clonal C Owner Shared St St E D	yiect Window Help	> 0v	+ () () () () () () () () () ()	<	
E Objec File O Cobjec + ⊗ Objec P Cobjec P T T T t	<pre>ct q s s ct q s f s f f f f f f f f f f f f f f f f</pre>	y Obj y Obj pa type pa type pa type pa type pa type Clonal C Owner Shared Shared St e ID 6	yiet Window Help	> 0v	+ \bigotimes \bigotimes_{AS0} \bigotimes_{DS} \bigotimes_{DS} \bigotimes_{DS} \bigotimes_{DS}	nt obje	
E Objec File O Cobjec + ⊗ Objec Pile Cobjec Pile Cobjec Pile Cobjec Pile Cobjec	ct q s ct q s ct li ct li f f y e 167	y Ob y Ob pa type type D Repeat Clonal C Owner Shared St C C C C C C C C C C C C C	yiect Window Help	> 0w _Du	+ $\Delta X 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 $	nt obje	
Cbject	<pre>ct q s s ct q s f s f f f f f f f f f f f f f f f f</pre>	y Ob y Ob pa type pa type D Repeat Clonal C Owner Shared st t e ID 6 7 8	ight Window Help ⊗ Image: Set in the set	> 0w _Du _Du _Du	+ \bigotimes \bigotimes_{AS0} \bigotimes_{DS} \bigotimes_{DS} \bigotimes_{DS} \bigotimes_{DS}	nt obje <	
	ct q s ct q s s ct li ct	y Ob y Ob pa type pa type D Repeat Clonal C Owner Shared Stared 6 7 8 9	ight Window Help ⊗ Image: Set in the set	> 0 0 0 0 0 0 0 0 0 0 0 0 0	+	nt obje <	
Cobject Cob	ct q s ct q s ct li ct li f f f f f f f f f f f f f f f f f f f	y Ob y Ob uery uery Repeat Clonal C Owner Shared St C C C Shared Shared C Shared C Shared C Shared C Shared C Shared C Shared C Shared C Shared C Shared C Shared C Shared Shared C Shared C Shared Sha	Image: Second State Image: Second State Second State [AII] Succession [AII] Successio	> Ow _Du _Du _Du _Du _Du _Du _Du _Du	+ S S S S S S S S S S S S S	er_	
Cbject	ct q s ct q s ct li f f f f f f f f f f f f f f f f f f f	y Ob y Ob pa type pa type Ppa type Repeat Clonal C Owner Shared St Shared 6 7 8 9 9 6 6 7 8 9 9	Image: Set Window Help Image: Set	> 0w _0v _0v _0v _0v _0v _0v _0v _0v _0v	+ S S S S S S S S S S S S S	st obje	
Image: Second system Image: Second system Image: Second	Ct q S S Ct q S S Ct lis (Ct lis (Ct lis (Ct lis (Ct lis (Ct lis (Ct lis (Ct lis)) (Ct lis (Ct lis)) (Ct lis (Ct lis)) (Ct lis)) (y Ob y Ob pa type type D Repeat Clonal C Owner Shared St C C C C C C C C C C C C C	Image: Second State Image: Second State 28 [All] succession [All] Succession [All] Complex [All] [All] [All]	> 0w _0v _0v _0v _0v _0v _0v _0v _0v _0v _0v	+ S S S S S S S S S S S S S	ererererererere	>
Image: Second system Object Image: Second system Image: Second system Image: Second system	 ct q ct q s s ct li ct li f f	y Ob y Ob pa type type D Repeat Clonal C Owner Shared St C C C C C C C C C C C C C	Image: Set Window Help Image: Set	> 0w _0v _0v _0v _0v _0v _0v _0v _0v _0v _0v	+ S S S S S S S S S S S S S	st obje	>
Image: Second system Object Image: Second system Image: Second system Image: Second system	<pre>ct q ct q s ct q ct li ct</pre>	y Ob y Ob uery uery Repeat Clonal C Owner Shared St Clonal C Owner Shared F Shared St St St St St St St St St St	Image: Second State Image: Second State 28 [All] succession [All] Succession [All] Complex [All] [All] [All]	> 0w _0v _0v _0v _0v _0v _0v _0v _0v _0v _0v	+ S S S S S S S S S S S S S	ererererererere	>

Figure 2.7: Object queries: Spa Repeats and Types.

Chapter 3

Importing and assembling trace files

3.1 Importing and assembling trace files in batch

A set of sequences run on an Applied Biosystems Genetic Analyzer can be downloaded from the Applied Maths website (https://www.applied-maths.com/download/sample-data, click on "SPA typing data files") and are used in this guide to explain the work flow of the *Spa-Typing plugin*.

- 1.1 Select *File* > *Import...* (, Ctrl+I) to call the *Import* dialog box.
- 1.2 Select *Import and assemble trace files* under *Sequence type data* and press <*Import*>.
- 1.3 Select the *<Browse>* button, navigate to the correct path, select all the sequence trace files and press *<Open>*.

The dialog is updated (see Figure 3.1).

Import sequence trace files		?	\times
Import sequence traces Select the data to import.			
Select file(s): C:\Users\10023342\Docume\Strain1_f.ab1 C:\Users\10023342\Docume\Strain1_r.ab1 C:\Users\10023342\Docume\Strain2_f.ab1 C:\Users\10023342\Docume\Strain3_f.ab1 C:\Users\10023342\Docume\Strain3_r.ab1 C:\Users\10023342\Docume\Strain3_r.ab1 C:\Users\10023342\Docume\Strain4_f.ab1 C:\Users\10023342\Docume\Strain4_r.ab1 C:\Users\10023342\Docume\Strain5_r.ab1 C:\Users\10023342\Docume\Strain5_r.ab1 C:\Users\10023342\Docume\Strain5_r.ab1 C:\Users\10023342\Docume\Strain5_r.ab1 C:\Users\10023342\Docume\Strain5_r.ab1 C:\Users\10023342\Docume\Strain5_r.ab1	Browse Delete Delete All		
Template file:	Browse		
< Back	Next >	Cance	el

Figure 3.1: Select trace files.

1.4 Press < *Next* > to go the next step.

The way the information should be imported in the database can be specified with an import template. In the example data set, the *Key* is provided in the trace file name.

1.5 Make sure the *Example import 1* template is selected and press the *<Preview>* button.

The *Example import 1* template will parse the *Key* from the file names.

- 1.6 Close the preview.
- 1.7 Make sure the *Example import 1* template is selected, and select the *Spa-typing* from the *Experiment type* list (see Figure 3.2).

mport sequence trace files		? >
Import template Specify how to import	lata into the database.	
Import templates: Example import 1 Example import 2	Entry keys parsed from file names.	Create new
		Edit Preview
		Сору
Experiment type: Spa-typ	ing ~	
	< Back N	ext > Cancel

Figure 3.2: Import sequence trace files.

- 1.8 Press <*Next*>.
- 1.9 Press < *Next* > once more to confirm the creation of 13 new entries (see Figure 3.3).

Import sequence trace files	?	\times
Database links Link the imported records to database entries. Double click on a cell to get an overview.		
Overview In 'All levels' Create 13 entries and update 0 entries		
Select modified entries Overwrite existing assemblies Update existing assemblies		
< Back Next >	Can	cel

Figure 3.3: Database links.

The *Processing* wizard page opens (see Figure 3.4).

In the *Reports panel*, the *Maximum# of unresolved bases reported* can be specified (default value 20). Likewise, the *Maximum # of align inconsistencies reported* can be entered (default value 20). Align inconsistencies are positions where the consensus is resolved, but where one or

Import sequence trace files			?	×
Processing Further processing of the trace files.				
Report Max. number of unresolved bases to report. 20 Report align inconsistencies Max. number of inconsistencies to report: 20 Open assembly overview report Settings Assembly settings Trimming settings				
	< Back	Finish	Can	cel

Figure 3.4: The Processing wizard page.

more sequences are different from the consensus.

1.10 Press < *Trimming settings*> to pop up the *Assembly trimming settings* dialog box (see Figure 3.5).

Assemb	ly trimming settings for 'Spa-ty	ping'	?	×
Minimur	n number of sequences contributir	ig to consensi	us: [1	
	Trim pattern	Tolerance	Offset Search range	
Start:	RCAMCAAAA	0	0 -	
Stop:	TAYATGTCGT	0	0 -	
			ОК	Cancel

Figure 3.5: The Assembly trimming settings dialog box.

Following settings can be specified:

- *Minimum # of sequences* specifies the minimum number of trace sequences that should contribute to the subsequence on the consensus that matches the trimming targets. For example, if "2" is entered, a trimming target will only be set if the matching region on the consensus is *fully* defined by at least 2 sequences.
- For both the *Start position* and *Stop position*, a *Trim pattern* is displayed. The use of IUPAC code for ambiguous positions is supported. The *Tolerance* defines the number of mismatches allowed for a sequence to be recognized as a trim pattern. With the *Offset*, one can specify that the consensus is trimmed at a certain offset from the start and end trimming target positions. If no offset is specified (zero), the trimming targets are included in the trimmed consensus. With the *Search range* one can restrict the search to certain regions on the consensus, e.g. to prevent incidental matches inside the targeted consensus sequence.

The entered trim patterns will be searched on the consensus sequence in both directions, i.e. on the consensus as it appears as well as on its complementary strand. In case the trim patterns

match the complementary strand of the consensus, it will be automatically invert-complemented. If the *Trim pattern* text boxes are left empty, no preference sense is available.

The trimming patterns specified in the *Spa typing settings* dialog box (see Figure 1.5) are shown in the *Start pattern* and *Stop pattern* text boxes.

- 1.11 Leave the predefined settings unaltered and press < OK > to close the trimming dialog box.
- 1.12 Press the <**Assembly settings**> button to call the **Assembly settings** dialog box (see Figure 3.6).

Assembl	y settings for 'Spa-typing'		?	×
Quality	Assembly Consensus Cop	oy settings		
Curve s	liding window:	5 b	p	
Minimum	i good/bad peak ratio:	1.30		
Minimum	short/long peak distance ratio:	0.50		
Base ca	Iling sliding window:	41 b	p	
Minimum	resolved positions:	25 b	р	
Minimum	consecutive good bases:	6 b	p	
Minimum	length of usable sequence:	50 b	p	
Minimum	fraction of good bases:	20 9	6	
	Reset to default	ОК	(Cancel

Figure 3.6: The Assembly settings dialog box.

The Assembly settings are grouped in tabs per settings dialog box in *Assembler*: **Quality** assignment, **Assembly** and **Consensus** determination. For a detailed description of the Assembler program settings, we refer to the reference manual. In the last tab the Assembly settings can be copied from or to another sequence type experiment.



The default Spa *Quality* assignment settings are required for submission of new types to the SpaServer (see 8).

- 1.13 For this exercise, do not change the settings and press < OK >.
- 1.14 Make sure the option *Open assembly overview report* is checked and press <*Finish*> to assemble the selected trace files from the example dataset into separate contig projects.

3.2 Reports

When the assemblies are processed, an interactive report window appears (see Figure 3.7). This window can also be displayed from the *Main* window with *Analysis* > *Sequence types* > *Batch assembly reports...*.

The *Overview* panel displays the entries (keys) as rows and the experiments as columns. Each cell, corresponding to a key/experiment pair, provides information about the current status of the contig project. This information can be:

- N/A: No such experiment exists with this key.
- **N/B**: An experiment with this key exists, but (a) the assembly was not created from this batch; or (b) no batch sequence assembly is present for this sequence.

æ	Assembly rep	oort: Batch - 2020-0	6-04 17:19:01					 _	×
File	Overview	Details Window	Help						
Ove	rview				Details				
ŧ					\approx				
	Кеу	Spa-typing	Message	-	Code	Message	Status		
~	Strain1	warning	Align inconsistencies		warning	Inconsistency in alignment at position 398 in sequence(s) 1, 2	new		
~	Strain10	warning	Align inconsistencies		warning	Inconsistency in alignment at position 384 in sequence(s) 1, 2	new		
~	Strain2	ок			info	Created new assembly			
~	Strain22	ок				Report for Strain1 / Spa-typing			
~	Strain23	warning	Align inconsistencies						
~	Strain24	ок							
~	Strain3	ок							
~	Strain4	ок							
~	Strain5	warning	Align inconsistencies						
~	Strain6	warning	Align inconsistencies						
~	Strain7	ок							
~	Strain8	warning	Align inconsistencies						
~	Strain9	warning	Align inconsistencies						
	<			>					

Figure 3.7: The Batch sequence assembly report window.

- **OK** (green): A contig was assembled without any problems.
- Warning (orange): Align inconsistencies occurred that were resolved under the applied consensus determination settings.
- Error (red): At least one of several possible assembly errors occurred, e.g. a trace sequence did not meet the quality criteria, more than one contig was created, the trimming positions were not found or unresolved bases are present in the consensus.
- Solved (green): A warning or error that was solved by the user.

2.1 Click a cell, e.g. *Strain23/Spa-typing* to update the *Details* panel on the right-hand side.

The *Details* panel is organized in message rows with four columns.

- The first column displays a message Code, which can be either "info", "warning" or "error".
- The second column shows the actual **Message**. Double-clicking on this cell opens the *Contig* assembly window (if not already open), with the corresponding position highlighted.
- The third column displays the **Status** of the message, which can be "new", "read" or "solved". The status can be changed by the user.
- The fourth column is a **Comment** field. A comment can be entered by the user.

Chapter 4

Checking assemblies in Assembler

4.1 Introduction

The *Contig assembly* window can be launched from the *Batch sequence assembly report* window or from the *Main* window:

- Double-click on a message cell in the *Details* panel*Details* panel of the *Batch sequence assembly report* window of an key/experiment combination to launch Assembler.
- As soon as an experiment is linked to a database entry, the *Experiment presence* panel shows a colored dot for the experiment of this entry. Click on the colored dot in the *Experiment presence* panel while holding the **Shift**-key to open the *Experiment card* window for an entry. In the *Experiment card* window, click on the Motor to launch Assembler.
 - 1.1 Open the *Contig assembly* window for the entry with key **Strain23** by double-clicking on the first message in the *Details* panel of the *Batch sequence assembly report* window.

The *Alignment* panel in the *Contig assembly* window shows the consensus sequence (upper line) and the individual trace sequences that contribute to the displayed consensus. The upper panel (*Alignment overview* panel) displays the aligned trace sequences. If the arrow points to the left, the program has invert-complemented the sequence to obtain the correct alignment. The upper left panel displays the selected consensus with its length and the number of sequences that are part of it.

1.2 Select the *Aligned traces* panel.

The bottom panel now displays the chromatogram files for both trace sequences (see Figure 4.1).

1.3 To obtain an optimal view of the curves, use the zoom sliders in the *Traces* panel or use the zoom buttons.

The parameter **Req. bases to include** in the Assembly settings dialog box is by default set to 51% (see Figure 1.6). This means that a gap in one sequence and a nucleotide in the other will insert a gap in the consensus sequence. If you take a closer look at the alignment inconsistencies of this assembly, two gaps are present in the forward sequence (at positions 160 and 218), resulting in two gaps in the consensus sequence. These positions will be further investigated in the next steps.



Figure 4.1: The Aligned traces panel.

4.2 Showing Spa repeats on the consensus

2.1 In the *Contig assembly* window, select *Spa-Typing* > *Show repeats* or use the shortcut **Shift+F5**.

Assembler screens the consensus sequence for repeats.

- Known repeats are shown in *green* (see Figure 4.2) and the name of the repeat is shown on top of the known repeat sequence.
- Bases in the repeat succession string that are not assigned to a known repeat are shown in *red*.
- The 5' and 3' signatures are displayed in *yellow*.

If the option *Allow IUPAC code* is checked in the *Spa typing settings* dialog box (see Figure 1.5) and *one of the bases* of a IUPAC code in the consensus results in a match with a known repeat, the repeat is shown in green and the name of the repeat is shown on top of the corresponding repeat sequence in the *Alignment* panel.

If the option **Allow IUPAC code** is checked in the *Spa typing settings* dialog box (see Figure 1.5) and *more than one* of the bases of a IUPAC code in the consensus results in a match with a known repeat, the *Multiple repeat successions* dialog box displays the different repeat succession options.

The repeat of the selected match is shown in *orange* (see Figure 4.4) and the name of the matched repeat is shown on top of the corresponding repeat sequence followed by a question mark (e.g. "r12?").

Align	iment																																														
			5'-signat	ure	-			r26							20)							30		r	23				4	0							- 50)					r	13	60	
I→I	СТС	ΑΑ	G C A	C (СΑ	A.	A A	G	AG	G	A A	١G	А	C A	ΑA	С	A A	ΑA	А	A /	A C	C	ΤG	G	Т	ΑĂ	ĂΑ	G	А	A (ÂΑ	C C	G	ςĊ	А	A	сA	Ā	A	C (ст	Ġ	G	c /	A A	A	G
1	стс										A A									Α.		С	ΤØ	G								۱C															
←	CYC	: A A	GC	A C (CA	A	АА	G	AG	G	AA	A G	A	C /	A A	C	A	A A	A	A	A C	с	то	G	т	AA	A A	G	A	A	GΑ	C C	G	GC	: A	A	СA	A A	A	C	сı	G	G	C /	A A	AA	G
	<																																														>
Align	nment																																														
	80	r??				90							1	00				31			1								120							r??1	30								40		
l→l	GGT	A A	AG/	A A I	GΑ	С	GG	i C	- 4	\C	Α.	λA	C	C 1	ΓG	G	C /	A A	A	G.	A A	G	ΑT	ΓG	G	C /	A A	ιC	А	A /	۹ C	: C	Т	GG	i C	A	A A	۱ G	iΑ	A	G /	۱C	A	A (C /	A A	A
	GGT	A A	AG	AΑ	GΑ	С	GG	i C	A A	٩C	A	ΑA	С	C 1	ΤG	G	C /	A A	A	G	ΑA	G	ΑT	ΓG	G	C /	ΑA	۱C	А	A	A C	: C	Т	G	i C	А	A A	4 6	iΑ	А	G /	A C	Α	A	C /	A A	Α
l←l	GGC	: A A	AG	A A	GΑ	С	GG	C	- /	A C	A	A M	С	C	ΓG	G	C	A A	A	G	AA	G	AI	ΓG	G	C,	AA	C	A	A	4 C	c	т	G	i C	A	A A	4 6	iΑ	A	G /	۹ C	A	A	C /	A A	A
																																															-

Figure 4.2: Showing the repeats on the consensus sequence.

Multiple repeat successions		?	×
There are multiple possible repeat successions, probably due to IUPAC code in the sequence. Select the repeat succession to use.			
RepeatSuccession	Туре		
15-174-16-02-16-02-25-17-24-24-24			
15-12-16-02-16-02-25-17-24-24-24	t018		
		ОК	
		Canc	el

Figure 4.3: The Multiple repeat successions dialog box.

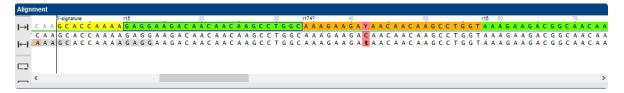


Figure 4.4: IUPAC code resulting in more than one known spa repeats.

The repeat succession string and the corresponding Spa type (if known) are displayed in the caption of the *Contig assembly* window (see Figure 4.5).

Strain23_f.ab1 (1) Assembly: 183(raw) - Sequence: 143 (158) RepeatSuccession: 26-23-13-??-31-29-17-25-17-25-16-28 Spatype: ???

Figure 4.5: Repeat succession string.

When importing and assembling spa sequences, BIONUMERICS uses the parameters defined in the *Assembly settings* dialog box (see Figure 3.6).

2.2 Select *File* > *Show report* (1) to view all parameters.

After import, these parameters can still be changed for each individual assembly.

1. Select the Trimming panel and select File > Quality assignment... (1) to change the

quality assignment settings. This action can only be used if the alignment is removed.

- 2. Select the *Assembly* panel and choose *Assembly* > *Assemble sequences...* (⅔) to change the assembly settings.
- 3. If you want to change the Consensus determination parameters, select the *Assembly* panel and select *Assembly* > *Consensus determination...*.

Detailed information on each of these parameters can be found in the reference manual.

4.3 Showing the repeat succession plot

3.1 Select *Spa-Typing* > *Show repeats plot* or use the shortcut *Shift+F6*.

The repeats are displayed in the Spa repeat plot window (see Figure 4.6).

at plot Strain	23/Spa-typi	ng						-	-		Х
Spa-Typing	Window	Help									
											^
		Found	15 Spa	-repeats with	correspon	ding Spa-Typ	e: ???				
5' r	26	r23	r13	r??	r31	r??	r17	r31	r	29	
		Four				COCTAC					
		3001	r29								
			115			111111C					
						IIIIAC					
					1111111						
			190								>
											-
	Spa-Typing	Spa-Typing Window	Spa-Typing Window Help Found 5' r26 r23 Sour r r r r	Spa-Typing Window Help Found 15 Spa 5' <u>r26</u> <u>r23</u> r13 Source:	Spa-Typing Window Help 5' r26 r23 r13 r?? Source: AAAGAAGACA r29 r115 r115 r116 r116 r116 r116 r467 r471 r05 r100 r100	Spa-Typing Window Help Found 15 Spa-repeats with correspondence 5' r26 r23 r13 r?? r31 Source: AAAGAAGACAACAACAAAAA r29 r115 r116 r1	Spa-Typing Window Help Found 15 Spa-repeats with corresponding Spa-Typ 5' r26 r23 r13 r?? r31 r?? Source: AAAGAAGACAACAAAAAGCCTAC- r29 r115 r467 r467 r471 GAC r05 r06	Spa-Typing Window Help Found 15 Spa-repeats with corresponding Spa-Type: ??? 5' r26 r23 r13 r?? r31 r?? r17 Source: AAAGAAGACAACAAAAAGCCTAC- r29 r115 r467 r471 GAC r467 r471 GAC r05 r100 r10 GC r100 GC r100 GC r100 GC r100 GC	Spa-Typing Window Help Found 15 Spa-repeats with corresponding Spa-Type: ??? 5' r26 r23 r13 r?? r31 r?? r17 r31 Source: AAAGAAGACAACAAAAAGCCTAC- r29 r115 r467 r467 r471 GAC r05 r06 GGC r100 A GAC	Spa-Typing Window Help Found 15 Spa-repeats with corresponding Spa-Type: ??? 5' r26 r23 r13 r?? r31 r?? r17 r31 r Source: AAAGAAGACAACAAAAAGCCTAC- r29 r115 r467 r467 r471 GAC r05 r06 GGC r100 r06 r06 GGC	Spa-Typing Window Help Found 15 Spa-repeats with corresponding Spa-Type: ??? 5' r26 r23 r13 r?? r31 r?? r17 r31 r29 Source: AAAGAAGACAACAAAAAAGCCTAC- r29 GC r115 GC r467 AC r471 GAC r471 GAC GC r156 GC GC

Figure 4.6: The repeat plot with editing suggestions for the second unknown repeat.

- 3.2 Click on the second unknown red "r??" repeat. A table is displayed with suggestions to edit the sequence. In the left column, the repeat is shown.
- 3.3 Use the zoom functions \square and \square (*View* > *Zoom in* and *View* > *Zoom out*) to obtain the best view of the plot.

Editing the sequence as suggested by the first row will give repeat "r29" (see Figure 4.6). Looking at this position in the *Contig assembly* window gives additional information about the missing base: in the chromatogram of the forward sequence, there is a missing "G".

- 3.4 Place the cursor on the gap in the trace sequence and type "G". The consensus sequence is automatically updated (see Figure 4.7 (b)).
- 3.5 Select the sequence you have edited (click on any position on the sequence, in the chromatogram or on the overview) and call *File* > *Sequence information...* (I), Ctrl+I). This brings up the *Sequence information* dialog box, listing all base corrections that are made to the sequence. Press <*Exit*>.
- 3.6 Select *Spa-Typing* > *Show repeats*.

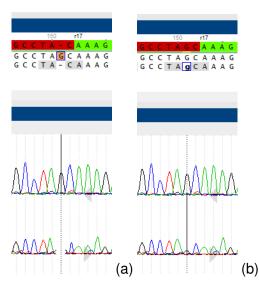


Figure 4.7: Missing peak in the chromatogram (a), Editing the trace sequence (b).

The repeat assignment in the consensus sequence is updated. The "r29" repeat is displayed in green in the *Assembly view*.

3.7 Select *Spa-Typing* > *Refresh* in the repeat plot to update the information.

The corrected repeat is displayed in green.

3.8 Click on the remaining unknown repeat in the repeat plot.

🖆 S	pa repe	at plot Strain	23/Spa-typi	ing					-	- 🗆	\times
File	View	Spa-Typing	Window	Help							
R	Ę,										
5											
_											
Q.	Plot										
				Found 15 Sp	a-repeats with c	orrespond	ling Spa-Typ	e. 222			Â
						oncopone	ing opu iyp				_
		5' r	26	r23 r13	r??	r31	r29	r17	r31	r29	
				Source:	AAAGAAGACG	BCACAAA	CTGGC				
				r23			CTIGC	t022			
1 1				r432	TTTTTTTTT	TC /	CTIGC				
				r458		GC /	CTIGC				
				r103			CTIGG				
				r06	TITTTTTTT	A A	CTIGC				
				r16	TITTTTTTT		CTIGT	t1616			
				r22	TITTTT		CTIGC				~
	<										>
		_	_			_	_	_	_	_	

Figure 4.8: The repeat plot with editing suggestions for the remaining "unknown" repeat.

The table with suggestions is displayed. In the left column, the repeat is shown. In the right column, the associated spa type is displayed (Figure 4.8). Editing the sequence as suggested by the first row will give repeat "r23" and type "t022". Looking at this position in the *Contig assembly* window gives additional information: in the chromatogram of the forward sequence, there is a missing "A" and based on the default *Consensus determination* parameters, this leads to a gap in the consensus sequence (see Figure 4.9 (a)).

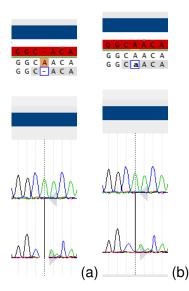


Figure 4.9: Missing peak in the chromatogram (a), Editing the trace sequence (b).

3.9 To insert the base in the trace sequence, place the cursor on the gap in the trace sequence and type "A".

The consensus sequence is automatically updated (see Figure 4.9 (b)).

3.10 Select *Spa-Typing* > *Show repeats*.

The repeat assignment in the consensus sequence is updated. All repeats are now displayed in green in the *Assembly view*.

3.11 Select *Spa-Typing* > *Refresh* in the repeat plot to update the information.

3.12 To copy the repeat plot to the clipboard, select *File* > *Copy to clipboard*.

3.13 The plot can be printed with File > Print.

3.14 Close the *Repeat plot window* with *File* > *Exit*.

The two warning messages (*Inconsistency in alignment at position 218* and *Inconsistency in alignment at position 160*) are checked and corrected for Strain23).



The plugin will not take into account unresolved bases in the consensus sequence when looking for spa repeats. Make sure no unresolved bases are present in the consensus sequence when looking for repeats.



The status of a contig project is set to **ERROR** if unresolved bases are detected in the consensus sequence.

4.4 Changing the status of error (and warning) messages

4.4.1 Principles

Only for those entries that have a green (= **OK** or **Solved**) or orange (= **Warning**) status, Spa types can be assigned.

- It is recommended to check the *warning* messages and solve them if needed. Since Spa types can be assigned to entries that have a Warning status, it is not required to change the status to "Solved".
- *Errors* need to be checked in the *Contig assembly* window and solved. Since Spa types cannot be assigned to entries that have an Error status, it is required to change the status to "Solved" after having solved all errors in Assembler.

4.4.2 Option1: Changing the status in Assembler

4.1 Select *Batch sequence assembly* > *Set report to solved, save and close* (Ctrl+Shift+S) in the *Contig assembly* window.

The corresponding key/experiment cell in the overview *Batch sequence assembly report* window is updated and displayed in green. The status "Solved" is displayed in the key/experiment field.

4.4.3 Option2: Changing the status in the Detailed report window

4.2 Open the *Contig assembly* window for the entry with key **Strain5** by double-clicking on one of the two warning messages in the *Details* panel of the *Batch sequence assembly report* window reporting an *Inconsistency in alignment*.

The contig is shown in the *Contig assembly* window, with the corresponding position in focus.

4.3 Select *Spa-Typing* > *Show repeats*.

The start and stop positions and 10 known repeats are detected.

4.4 Make sure the *Aligned traces* panel is selected and use the zoom sliders or the zoom buttons to obtain an optimal view of the curves.

If you look at the chromatograms at position 395 and 397, false peaks introduce a "G" at position 395 and an "A" at position 397 in the reverse sequence. Gaps in the forward sequence at these positions result in two gaps in the consensus sequence (see Figure 4.10). These gaps are the reason why two warning messages are reported for this contig project. We could delete these false base callings in the forward sequence, resulting in the removal of the gaps in the consensus sequence, but since we have allowed gaps to be present in the consensus sequence when searching for repeats and signatures in the source sequence, these gaps do not interfere with the analysis, and so we do not need to edit the sequence.

- 4.5 Select *File* > *Save* (B, Ctrl+S) and *File* > *Exit* to close the *Contig assembly* window.
- 4.6 In the *Batch sequence assembly report* window, select *Details* > *Set all messages to solved* (Ctrl+S).

The corresponding key/experiment cell in the *Overview* panel is updated and displayed in green. The status "solved" is displayed in the cell and in the *Status* column of the *Details* panel (see Figure 4.11).

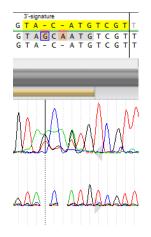


Figure 4.10: False peaks.

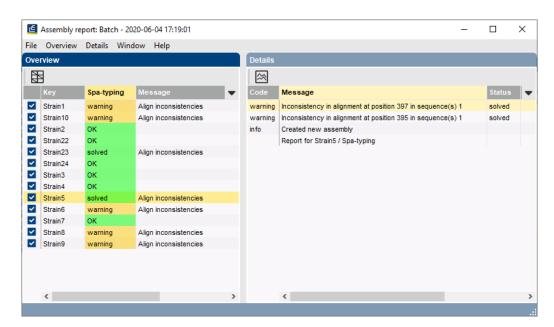


Figure 4.11: Solve errors/warnings.

Chapter 5

Spa-Typing in BIONUMERICS

5.1 Selections in the main window

In the *Main* window, a **Spa-typing** experiment is present for each contig project (see colored dot in the **Spa-typing** column in the *Experiment presence* panel).

Screening for spa repeats and types can be done for all entries present in the database, or for any selection of entries in database.

1.1 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 (\mathbf{z}) and can be unselected in the same way.

1.2 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.

- 1.3 All entries can be selected at once with *Edit* > *Select all* (Ctrl+A).
- 1.4 Clear all selected entries with *Database* > *Entries* > *Unselect all entries (all levels)* (F4).

5.2 Assigning Spa types

5.2.1 Principles

- 2.1 Make a selection in the *Main* window.
- 2.2 Select *Spa-Typing* > *Assign Spa types* in the *Main* window.

The *Find Spa types* dialog box pops up (see Figure 5.1).

The *Include Kreiswirth notation* and *Include clonal complexes* check boxes are only shown if their corresponding information fields are present in the database (see Figure 1.5).

2.3 Press <*OK*>.



If no selection is present in the database, the software will display a message asking you if you wish to run the tool on the complete database.

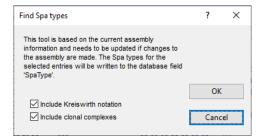


Figure 5.1: The Find Spa types dialog box.

Errors occurred	?	×
Errors occurred; entries with problems are selected in the dat Select entries below to open in Assembler:	abase.	
Strain1: unknown repeats Strain6: unknown repeats Strain8: unknown repeats Strain1: unknown repeats Strain10: unknown repeats Strain23: unknown repeats	OK	el

Figure 5.2: The Errors occurred dialog box.

If entries are detected with sequence assembly problems or unknown repeats, the *Errors occurred* dialog box pops up, listing all these entries with one of the following error messages:

- Unknown repeats: One or more unknown repeats are detected in the consensus sequence.
- Problems with assembly: The status box in the Overview report window reports an error message (= red status box). Spa types can only be assigned to entries that have a green (= OK or Solved) or orange (= Warning) status.

Entries can be selected and their assemblies can be opened in Assembler.

All entries with sequence assembly problems or unknown repeats are selected.

The *Spa typing plugin* uses a 2-step approach when the command *Spa-Typing* > *Assign Spa types* is selected:

5.2.2 Step 1: The assembly is screened for repeats

The repeat succession is displayed in the database information field that holds the repeat succession information (default name: **RepeatSuccession**, see Figure 1.5). The repeat succession is stored in the database information field that holds the repeat succession information *and* in the character type **Spa-repsuc**. If the *Include Kreiswirth notation* is checked in the *Find Spa types* dialog box, the Kreiswirth information is stored in the database information field that holds the Kreiswirth notation field that holds the Kreiswirth notation field that holds the Kreiswirth notation if this information field is present in the database.

- 2.4 Click on the colored dot in the **Spa-repsuc** column of the *Experiment presence* panel to open the character *Experiment card* window for an entry (see Figure 5.3).
- 2.5 Close the experiment card by clicking in the small triangle-shaped button in the left upper corner.

Character	Value	Mapping
rs_001	9	r08
rs_002	17	r16
rs_003	3	r02
rs_004	17	r16
rs_005	3	r02
rs_006	26	r25
rs_007	18	r17
rs_008	25	r24

Figure 5.3: The Spa-repsuc character card, displaying the repeat succession in the *Mapping* column.

When a repeat sequence does not match one of the repeats in the database, or when a IUPAC code is present in the consensus sequence, a "??" is displayed in the **RepeatSuccession** information field and in the **Mapping** column of the character card. In the Kreiswirth information field - if present in the database - the text "NA" (Not Available) is displayed.

When a sequence is found that is too short or too long to be considered as a repeat sequence, an asterisk (*) is displayed in the **RepeatSuccession** information field and in the **Mapping** column of the character card. In the Kreiswirth information field - if present in the database - the text "NA" (Not Available) is displayed.

When no repeats are found, no information is written in the repeat succession and Kreiswirth information fields.

5.2.3 Step 2: Repeat type (if available) is assigned to each selected entry

The Spa type is displayed in the information field that holds the Spa Type information (default name: **SpaType**, see Figure 1.5).

The Spa type is denoted as "???" if the repeat succession is incomplete. When the repeat information is currently not linked to a Spa type in the database, "Unknown" is displayed in the spa type information field. If no repeats are found, "NA"(Not Available) is displayed.

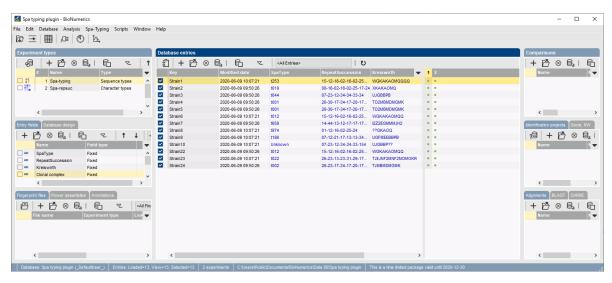


Figure 5.4: The Main window after repeat and type assignment.

Chapter 6

Cluster analysis of Spa types

6.1 Introduction

In this chapter, we are going to take a look at the evolutionary relationship between the Spa sequences by means of the construction of a dendrogram and a minimum spanning tree.

The Spa typing plugin uses a multi-step approach for this cluster analysis.

- The plugin uses an algorithm based on a DSI model [1] for the pairwise alignment of the Spa repeats. This model considers three mutational events: Duplication of tandem repeats, Substitutions and Indels.
- Next, the cost matrix is used to correct for the evolutionary distances between the repeats.

Taking these costs into account, the output of the DSI model is a similarity matrix. From this similarity matrix a dendrogram and/or a minimum spanning tree can be constructed.

6.2 The Comparison window

- 2.1 For this exercise, make sure all entries are selected in the Main window (Ctrl+A).
- 2.2 Highlight the *Comparisons* panel in the *Main* window and select *Edit* > *Create new object...* (+) to create a new comparison for the selected entries.
- 2.3 Drag the separator lines between the panels to the left or to the right, in order to divide the space among the panels optimally.
- 2.4 Move the panels by clicking in the header of a panel and while keeping the mouse button pressed dragging it to another location in the *Comparison* window.

The character type **Spa-repsuc** is created upon installation of the *Spa typing plugin* and displayed in the *Experiments* panel. The repeat information stored in the associated character type will be used when using the clustering tools. The repeat succession stored in the associated repeat information field is only used when no repeat information is present in the associated character type.

2.5 Click on the eye button (()) of the character type **Spa-repsuc** in the *Experiments* panel.

X Y	Comparison																						- 0	1
Construction Experiment data Enderforgram Experiment data Information fields Image: State of the st	e Edit Layout Groups Clust	ering Statistics Fingerprints	Charao	cters	Sequ	ence	Spa cl	usterin	ng Tr	endDat	a Re	adSet	s Spe	ctra (Compos	te Win	dow	Help						
Consistent Experiment data Information fields Information fields Image: Sparteneous Image: Sparteneous <t< th=""><th>36 6 ≭x £</th><th>다. Spa-rep:</th><th>suc</th><th></th><th></th><th>\bigcirc</th><th>đ</th><th></th><th>Ъ,</th><th>1017 o</th><th>Ľ</th><th>ŀ,</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></t<>	36 6 ≭x £	다. Spa-rep:	suc			\bigcirc	đ		Ъ,	1017 o	Ľ	ŀ,												
Normal	xperiments	e																						
Bit datasets Spar-regular Bit datasets Bit data Bit data Bit data Bit datasets Fig. diff. Bit datasets Fig. diff. Bit datasets Fig. diff.	<all experiment="" types=""></all>		Expe	erime	nt data	1										Infor	nation	fields						
Big Big <td>Name 🔻</td> <td>이 나는 것 같</td> <td>1V</td> <td>, I 🛛</td> <td></td> <td>12</td> <td>2</td> <td>10</td> <td>000</td> <td> <0</td> <td>haract</td> <td>ter nan</td> <td>ie></td> <td></td> <td></td> <td>e,</td> <td>ιtΓ</td> <td> \$↑ ↓</td> <td>∰ ↑ ↓</td> <td></td> <td></td> <td></td> <td></td> <td></td>	Name 🔻	이 나는 것 같	1V	, I 🛛		12	2	10	000	<0	haract	ter nan	ie>			e,	ιtΓ	\$↑ ↓	∰ ↑ ↓					
No No<			. s	Spa-re	psuc																			
Bit Shareds Control Contro Control Control	a opa-repsuc		_	~	~							_	~	m +	10									
Construction Construction<			8	8	8	8	8, 8	8, 8	8	8	ē,	8	ē,	δ δ	δ									
n08 n16 n12 r16 r12 r16 r12 r17 r14 C Strand 2202.06-80 95:02.05 1019 06.16-02.05-10.24 XXAAADU 07 r23 r23 r14 r14 <td< td=""><td>,</td><td></td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2 2</td><td>2 2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2 2</td><td>2</td><td></td><td></td><td>Key</td><td>Modified date</td><td></td><td></td><td></td><td>Kreis 🔻</td><td></td></td<>	,		2	2	2	2	2	2 2	2 2	2	2	2	2	2 2	2			Key	Modified date				Kreis 🔻	
n07 c3 c12 c4 c4 c3 c4 c3 c4 c4 c3 c3 c4 c4 c4 c3 c3 c4 c4 c4 c3 c3 c4 c	vses Local composite datasets									r24	r24	r24	r24					Strain1	2020-06-09 10:	7:21 t2	53	15-12-16-02-16-02-25-17-24-24-24-24	WGKAKAG	
C8 60 17 74 17 72 17 71 72 17 71 72 17 74 72 74 72 74 72 74 72 74 72 74<	9 ⊗		r08	r16	r02	r16 ri	02 r2	5 r17	r24									Strain2	2020-06-09 09:	0:26 t0	19	08-16-02-16-02-25-17-24	XKAKAOM	
AB 100 17 74 17 101 17 1			r07	r23	r12	r34 r.	34 r3	3 r34								~		Strain3	2020-06-09 09:	0:26 t0	44	07-23-12-34-34-33-34	UJGBBPB	
11 11 11 11 11 12 12 11 12 <	Name 🔫	1 m	r26	r30	r17	r34 r	17 r2	0 r17	r12	r17	r16					~		Strain4	2020-06-09 09:	0:26 t0	01	26-30-17-34-17-20-17-12-17-16	TO2MBMDI	
14 rt4 rt3 rt2 rt7 rt7 rt2 rt8 22 Strain 2020-66.09 65.90 14.44.13-12.17.17.17.23.18 2222EGMM r0 rt2 rt6 rt2 rt4 rt3 rt2 rt4 rt3 rt2 rt4 rt4 rt3 rt2 rt4 rt4 rt3 rt2 rt4 rt4 rt3 rt2 rt4 rt3 rt4 rt4 rt3 rt4 rt7 rt7 rt4 rt4 rt3 rt4			r26	r30	r17	r34 r	17 r2	0 r17	r12	r17	r16					~		Strain5	2020-06-09 09:	0:26 t0	01	26-30-17-34-17-20-17-12-17-16	TO2MBMDI	
O rt1 rt6 rt2 rt7 rt2 rt7 rt7 rt2 rt6 rt2 rt7 rt2 rt7 rt2			r15	r12	r16	r02 r	16 r0	2 r25	i r17	r24	r24					~		Strain6	2020-06-09 10:	7:21 t0	12	15-12-16-02-16-02-25-17-24-24	WGKAKAC	
NO NO<			r14	r44	r13	r12 r	17 r1	7 11	r23	r18						V		Strain7	2020-06-09 09:	0:26 t6	59	14-44-13-12-17-17-17-23-18	I2Z2EGMM	
NO 12 14 13 154 2 Struit 2020-66-90 100721 Unknown 07-23-12-34-33-31-94 UNGB8977 NS 112 16 02 57 117 124 2 Strait 2020-66-90 00721 Unknown 07-23-12-34-33-31-94 UNGB8977 NS 62 37 112 16 02 57 117 124 2 Strait 2020-66-90 00721 Unknown 07-23-12-34-33-31-94 WIGKB4AC NS 62 37 112 121 16 02 17 125 1	< >		r01	r12	r16	r02 r3	25 r2	4								V		Strain8	2020-06-09 10:	7:21 t9	74	01-12-16-02-25-24	??GKAOQ	
Image: Note of the transmission of the transmission of the transmission of the transmission of transmis			r07	r12	r21	r17 r	13 r1	3 r34	r34	r33	r34					V		Strain9	2020-06-09 10:	7:21 t1	86	07-12-21-17-13-13-34-34-33-34	UGFMEEBE	
📭 7 126 123 113 123 131 129 117 131 129 117 125 117 125 116 128 🗹 Stran23 2020-06-09 10:07:21 1022 28-23-13-23-31-29-17-31-29-17-25-1 TJEJNF2M			r07	r23	r12	r34 r.	34 r3	3 r15	54							V		Strain10	2020-06-09 10:	7:21 Ur	nknown	07-23-12-34-34-33-154	UJGBBP??	
r26 r23 r13 r23 r31 r29 r17 r31 r29 r17 r25 r17 r25 r16 r28 🗹 Strain23 2020-06-09 10:07:21 t022 26-23-13-29-17-31-29-17-31-29-17-25-1 TJEJNF2M			r15	r12	r16	r02 r	16 r0	2 125	r17	r24	r24					V		Strain22	2020-06-09 09:	0:26 t0	12	15-12-16-02-16-02-25-17-24-24	WGKAKAC	
▼ 126 123 117 134 117 120 117 112 117 116 🔽 Strain24 2020-08-09 09:50:26 1002 26:23-17:34-17:20-17:12-17:16 ТЈИВИОМК	/ - Not Not -		r26	r23	r13	r23 r.	31 r2	9 r17	r31	r29	r17	r25	17 13	5 r16	r28	V		Strain23	2020-06-09 10:	7:21 t0	22	26-23-13-23-31-29-17-31-29-17-25-1	TJEJNF2M	
	Size Name 🗨		r26	r23	r17	r34 r	17 r2	n 17	r12	r17	r16					V		Strain24	2020-06-09 09:	0:26 t0	02	26-23-17-34-17-20-17-12-17-16	TJMBMDMK	
	Groups		r07 r15 r26	r23 r12 r23	r12 r16 r13	r34 ri r02 r r23 ri	34 r3 16 r0 31 r2	3 r15 2 r25 9 r17	54 5 r17 7 r31	r24 r29	r24 r17	r25	r17 r3	!5 r16	r28	y y y		Strain10 Strain22 Strain23	2020-06-09 10: 2020-06-09 09: 2020-06-09 10:	7:21 Ur 0:26 t0 7:21 t0	nknown 12 22	07-23-12-34-34-33-154 15-12-16-02-16-02-25-17-24-24 26-23-13-23-31-29-17-31-29-17-25-1		UJGBBP?? WGKAKAC TJEJNF2MI
			<												,		<						>	$\langle \rangle$

Figure 6.1: The Comparison window.

The pattern images are displayed in the *Experiment data* panel. Initially, the character values are displayed as colors according to the color scale defined for each character (see the reference manual for more information).

2.6 Select *Characters* > *Show mappings* (Im) or *Characters* > *Show mappings+colors* (Im) to display the mapped name for each character value (see Figure 6.1).

6.3 Creating a cost matrix

In the *Spa typing plugin*, there is a default binary cost matrix available for the calculation of the dendrogram, consisting of two states: a match between the repeats and no match.

3.1 Select *Spa clustering* > *Cost matrices* in the *Comparison* window for the creation of your own cost matrix.

The *Cost matrices* dialog box appears (see Figure 6.2).

A weight

Figure 6.2: The Cost matrices dialog box.

The Cost matrices dialog box displays all cost matrices defined by the user (initially empty).

Selecting < *Create new* > displays the *Create Spa cost matrix* dialog box (see Figure 6.3).

You can define a *Name* for the cost matrix and set the costs for nucleotides and amino acids.

• *Maximum number of differences*: defines the maximum number of differences in nucleotides/amino acids between two repeats. The default is 5 and there is a gradual cost

Create Spa cost mat	rix	? ×
Cost matrix name:	My own matrix	
Nucleotide cost		
Maximum number of differences:	5	
Relative weight:	1	
Amino acid cost	t	
Maximum number of differences:	5	ОК
Relative weight:	2	Cancel

Figure 6.3: The Create Spa cost matrix dialog box.

between 0 and 5 mismatches. Differences larger than 5 will get 100% of the cost as well.

- **Relative weight**: defines the relative weight between the nucleotides and the amino acids. The settings in Figure 6.3 penalize a change in an amino acid twice as much as a change in a nucleotide.
 - 3.2 Select < *Create new*>, specify a *Cost matrix name*, leave the settings unaltered and press < *OK*>.

This calls the Cost matrix dialog box (see Figure 6.4).

Repeat No1 No2 No3 No4 No5 No6 NO7 No6 NO9 NO N11 N12 N13 N14 N15 N16 N01 0 73 40 33 67 67 33 27 60 60 27 53 60 27 20 73 N02 73 40 60 0 7 60 73 7 13 33 33 13 40 60 27 33 53 73 73 73 73 73 33 53 77 73 73 73 73 53 60 73 60 53 73 73 53 60 73 60 73 73 73 53 60 73 60 73 60 73 60 73 60 73 73 73 73 73 73 73 73 73 73 73	Repeat	r01	r02	r03	r04	r05	r06	r07	r08	r09	r10	r11	r12	r13	r14	r15	r16
r02 73 0 60 60 7 27 53 60 73 40 60 33 27 47 53 53 r03 40 60 0 7 60 73 7 13 33 33 13 40 33 13 20 60 r04 33 60 7 0 60 73 13 7 40 27 7 33 40 20 13 60 r05 67 7 60 60 0 33 60 73 73 33 53 27 33 53 47 60 r06 87 27 73 73 73 73 73 73 73 33 13 27 33 13 7 13 53 73																	
r03 40 60 0 7 60 73 7 13 33 13 40 33 13 20 60 r04 33 60 7 0 60 73 13 7 40 27 7 33 40 20 13 60 r05 67 7 60 60 0 33 60 53 73 33 53 27 33 53 47 60 r06 87 27 73 73 73 73 60 73 60 73 73 73 53 60 73 60 53 7		-															
r04 33 60 7 0 60 73 13 7 40 27 7 33 40 20 13 60 r05 67 7 60 60 0 33 60 53 73 33 53 27 33 53 47 60 r06 87 27 73 73 73 53 60 73 60 53 73 73 53 60 73 60 53 73 73 53 60 73 60 73 60 73 60 73 60 73 60 73 60 73 7 <th7< td=""><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></th7<>			-														
r05 67 7 60 60 0 33 60 53 73 33 53 27 33 53 47 60 r06 87 27 73 73 33 0 73 73 53 60 73 60 53 73 73 53 60 73 60 53 73 73 73 53 60 73 60 53 73 73 73 53 60 73 60 53 73 73 73 73 73 73 60 73 <th< td=""><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></th<>				-													
106 87 27 73 73 33 0 73 73 73 73 73 73 73 73 73 73 60 73 60 73 60 73 70 13 60 73 <th< td=""><td></td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></th<>					-												
r07 33 53 7 13 60 73 0 7 27 40 20 33 27 7 13 53 r08 27 60 13 7 53 73 7 0 33 33 13 27 33 13 7 60 r09 60 73 33 40 73 53 27 33 0 60 47 60 53 33 40 27 r10 60 40 33 27 33 60 40 33 60 47 60 53 33 40 27 r11 27 60 13 7 53 73 20 13 47 40 40 40 40 40 40 33 33 33 33 33 33 33 33 33 33 33 34 40 33 33 33 33 33 33 33 33 33 33 33 33						-											
r08 27 60 13 7 53 73 7 0 33 33 13 27 33 13 7 60 r09 60 73 33 40 73 53 27 33 0 60 47 60 53 33 40 27 r10 60 40 33 27 33 60 40 33 60 0 33 7 13 47 40 40 r11 27 60 13 7 53 73 20 13 47 33 0 40 47 13 7 60 r12 53 33 40 33 53 27 60 33 47 7 0 33 40 23 33 33 33 33 33 33 34 40 33 40 27 40 33 40 7 0 33 40 7 60 7 7 13 33 47 13							-										
r09 60 73 33 40 73 53 27 33 0 60 47 60 53 33 40 27 r10 60 40 33 27 33 60 40 33 60 0 33 7 13 47 40 40 r11 27 60 13 7 53 73 20 13 47 33 0 40 47 13 7 60 r12 53 33 40 33 53 27 60 7 40 0 7 40 33 40 33 33 40 27 7 60 7 40 0 7 40 33 40 7 60 7 40 33 40 7 60 60 60 7 60 60 60 60 60 60 60 60 60 60 7 7 60 60 7 73 60 60 73 73								-									
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$																	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$										-							
$\begin{array}{cccccccccccccccccccccccccccccccccccc$											-						
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$												-					
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$													-				
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$																	
r16 73 53 60 60 60 27 53 60 27 40 60 33 27 60 60 0 r17 73 60 60 60 53 33 60 53 33 33 60 27 33 60 27 33 60 60 7 r18 60 73 73 73 73 60 73 73 60 60 73 73 53 60 73 73 53 60 73 73 53 60 73 73 60 60 73 73 53 60 73 73 73 73 60 60 73 73 73 73 73 60 60 73 73 73 73 73 60 60 73 73 73 73 73 73 73 73 73 73 73 73 73 73 74 74 74 74 74 74 73 73															-		
17 73 60 60 60 53 33 60 57 33 60 27 33 60 60 73 73 73 73 73 73 73 73 73 73 73 73 73 73 60 60 73 73 53 60 73 73 53 60 73 73 53 60 73 73 53 60 73 73 53 60 73 73 33 71 73 73 60 60 73 73 53 60 73 73 33 73 73 73 73 60 60 73 74 73 73 74 <th< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>-</td><td></td></th<>																-	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$																	-
$\begin{array}{cccccccccccccccccccccccccccccccccccc$																	
r21 47 27 47 40 20 53 40 33 60 13 33 7 13 33 27 40 r22 73 53 60 60 47 27 60 60 40 40 60 33 40 60 53 13 r23 73 47 60 60 53 20 60 60 33 47 60 60 53 13 r24 73 60 60 60 53 20 60 60 33 47 60 40 53 60 7 r24 73 60 60 60 40 60 40 60 33 60 60 13 r25 73 60 60 60 33 60 60 33 60 60 13 r24 73 60 60 60 33 60 60 33 60 60 13 r25 73 60 </td <td></td> <td>53</td> <td>33</td> <td>40</td> <td>33</td> <td>27</td> <td>60</td> <td>47</td> <td>40</td> <td>60</td> <td>7</td> <td>27</td> <td>13</td> <td>20</td> <td>40</td> <td>33</td> <td></td>		53	33	40	33	27	60	47	40	60	7	27	13	20	40	33	
r21 47 27 47 40 20 53 40 33 60 13 33 7 13 33 27 40 r22 73 53 60 60 47 27 60 60 40 40 60 33 40 60 53 13 r23 73 47 60 60 53 20 60 60 33 47 60 60 53 13 r24 73 60 60 60 53 20 60 60 33 47 60 40 53 60 7 r24 73 60 60 60 40 60 40 60 33 60 60 13 r25 73 60 60 60 33 60 60 33 60 60 13 r24 73 60 60 60 33 60 60 33 60 60 13 r25 73 60 </td <td>r20</td> <td>53</td> <td>20</td> <td>40</td> <td>47</td> <td>27</td> <td>47</td> <td>33</td> <td>40</td> <td>60</td> <td>20</td> <td>40</td> <td>13</td> <td>7</td> <td>27</td> <td>33</td> <td>33</td>	r20	53	20	40	47	27	47	33	40	60	20	40	13	7	27	33	33
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		47		47	40		53		33	60		33	7				
r24 73 60 60 60 60 40 60 60 40 60 33 40 60 60 60 13 r25 73 60 60 60 60 33 60 60 33 60 60 13 r26 53 32 37 32 40 60 27 47 60 40 52 47 37 32 7	r22	73	53	60	60	47	27	60	60	40	40	60	33	40	60	53	13
r25 73 60 60 60 33 60 60 33 47 60 40 33 60 60 60 70 r26 53 37 32 40 60 20 27 47 60 40 52 47 27 23 72	r23	73	47	60	60	53	20	60	60	33	47	60	40	33	53	60	7
-ne E2 22 27 22 40 60 20 27 47 60 40 E2 47 27 22 73	r24	73	60	60	60	60	40	60	60	40	40	60	33	40	60	60	13
	r25	73	60	60	60	60	33	60	60	33	47	60	40	33	60	60	7
< >>		60	00	77	00	40	60	20	77	47	en	40	60	47	77	00	
	<																>

Figure 6.4: The Cost matrix dialog box.

The cost matrix is calculated and shown. The higher the costs, the more distantly related the repeats are. Press < *Close*> to close the *Cost matrix* dialog box.

A selected cost matrix is removed from the list in the *Cost matrices* dialog box with <*Delete*>. The cost matrix is shown in the *Cost matrix* dialog box when pressing the <*Show*> button. The *Cost matrices* dialog box can be closed with <*Close*>.

6.4 Cluster analysis settings

4.1 Select *Spa clustering* > *Cluster Spa types* in the *Comparison* window.

The Spa Clustering dialog box appears (see Figure 6.5).

Spa Clustering	? ×
Alignment Gap creation cost: 250 % Gap extension cost: 50 % Duplicate creation cost: 25 % Duplicate extension 25 % Maximum duplication 3 reps	Cluster Method UPGMA Neighbor Joining Single Linkage Complete Linkage Minimum Spanning Tree
Matrix Cost matrix: Default ~	MST Bin grouping distance: 1.00 %

Figure 6.5: The Spa Clustering dialog box.

Following settings can be specified in the Spa Clustering dialog box:

Alignment settings:

- *Gap creation cost*: specifies the cost for the introduction of a single gap in one of the repeats (in %).
- Gap extension cost: defines the cost for the extension of a created gap (in %).
- Duplicate creation cost: gives the cost for the duplication of a repeat (in %).
- Duplicate extension: defines the cost for the extension of a duplicated repeat (in %).
- *Maximum duplication length*: defines the maximum number of neighboring repeats that are taking into account to create a duplicate from.

Matrix:

In the *Matrix panel*, the default cost matrix or a custom cost matrix can be selected from the drop-down menu (see 6.3 for the creation of a cost matrix).

Cluster Method:

In the upper right box, five cluster methods are listed: *Minimum spanning tree*, *UPGMA*, *Neighbor Joining*, *Single Linkage*, and *Complete Linkage*.

An additional setting called **Distance bin size** is displayed in the **MST panel** when the **Minimum spanning tree** option is checked. Based on this setting, the software creates bins of certain distance intervals, that are converted into distance units. When for example the distance bin size is set to 1%, two entries having a similarity of 99.6% will have a distance of 0 (interval 100%-99% =

distance 0). Two entries that have a similarity of 98.7% will have a distance of 1 (interval 99%-98% = distance 1). The default setting is 1%.

In this example, we will create a minimum spanning tree (see 6.5) and a UPGMA dendrogram (see 6.6).

6.5 Minimum spanning tree

Minimum spanning trees are trees calculated from a distance matrix and possess the property of having a total branch length that is as small as possible. A MST chooses the sample with the highest number of related samples as the root node, and derives the other samples from this node. This may result in trees with star-like branches and allows for a correct classification of population systems that have a strong mutational or recombinational rate.

5.1 Select *Spa Clustering* > *Cluster Spa types* in the *Comparison* window and select *Minimum Spanning Tree* in the *Cluster Method panel* (see Figure 6.5).

An additional setting called **Distance bin size** is displayed in the **MST panel**. Based on this setting, the software creates bins of certain distance intervals, that are converted into distance units. When for example the distance bin size is set to 1%, two entries having a similarity of 99.6% will have a distance of 0 (interval 100%-99% = distance 0). Two entries that have a similarity of 98.7% will have a distance of 1 (interval 99%-98% = distance 1). The default setting is 1%.

5.2 Leave the settings unaltered and press < OK >.

The *Advanced cluster analysis* window pops up (see Figure 6.6). The *Network panel* displays the minimum spanning tree, the upper right panel (*Entry list*) displays the entries that are present in the tree. The *Selection entry list* lists the entries that are present in the selected node(s).

5.3 Select a node or branch by clicking on them, or several nodes/branches by holding the **Shift**-key while clicking.

As an exercise we will change some display settings. More detailed information about the *Ad*-vanced cluster analysis window can be found in the reference manual.

5.4 Choose Display > Display settings to open the Display settings dialog box.

5.5 In the Node labels and sizes tab, select Show node labels and select SpaType from the list.

- 5.6 In the Node colors tab, select Number of entries from the drop-down list.
- 5.7 In the Branch styles tab, select branch length from the drop-down list.
- 5.8 In the Branch labels and sizes tab, select Show branch labels and branch length.
- 5.9 Press *<OK>* to apply the new settings.

The Advanced cluster analysis window should now look like Figure 6.6.

- 5.10 In the *Advanced cluster analysis* window, select *Display* > *Zoom to fit* to optimize the view of the tree in the current window.
- 5.11 Close the Advanced cluster analysis window.

Advanced cluster analysis Edit Display Window Help C 🏠 Spa typing MST	X 31 V V V						- 1	
letwork			Entr	y list				
🖽 🏆 🖍 🗊 🔛	SpaType	S. 0		Key	SpaType	RepeatSuccession	Kreiswirth	
				Strain1	t253	15-12-16-02-16-02-25-17-24-24-24-24	WGKAKAOMQC	100
t659			~	Strain2	t019	08-16-02-16-02-25-17-24	XKAKAOMQ	
			~	Strain3	t044	07-23-12-34-34-33-34	UJGBBPB	
			~	Strain4	t001	26-30-17-34-17-20-17-12-17-16	TO2MBMDMGMH	<
9:00		1	~	Strain5	t001	26-30-17-34-17-20-17-12-17-16	TO2MBMDMGMH	< C
	t012 t974	Unknown 2-5	~	Strain6	t012	15-12-16-02-16-02-25-17-24-24	WGKAKAOMQC	2
			~	Strain7	t659	14-44-13-12-17-17-17-23-18	I2Z2EGMMMJH2	
	1253	t044 11 - 20	~	Strain8	t974	01-12-16-02-25-24	??GKAOQ	
u /	ha	800 2 1	~	Strain9	t186	07-12-21-17-13-13-34-34-33-34	UGFMEEBBPB	
9,00		6:00 21	~	Strain10	Unknown	07-23-12-34-34-33-154	UJGBBP??	
		<u>}</u>	~	Strain22	t012	15-12-16-02-16-02-25-17-24-24	WGKAKAOMQC	2
t001		U t186	~	Strain23	t022	26-23-13-23-31-29-17-31-29-17-25-17-25-16-28	TJEJNF2MNF2M	OMOKR
1002 10,00				ter analysis	t002 on entry list Entr s method	28-23-17-34-17-20-17-12-17-16	TJMBMDMGMK	>
1022					out data	t 'Spa-repsuc'		
					- Aspe	ct		
					Data type similarity values Renc ethod Branch pr	urt		

Figure 6.6: The Advanced cluster analysis window.

6.6 Cluster analysis sensu stricto

Cluster analysis *sensu stricto* is based upon the similarity matrix and a subsequent algorithm for calculating bifurcating dendrograms to cluster the entries. In the *Spa typing plugin* you can choose between the following four methods: Unweighted Pair Group Method using Arithmetic averages (*UPGMA*), the *Neighbor Joining* method and two variants of UPGMA: *Single linkage* and *Complete linkage* (see Figure 6.5).

- 6.1 In the *Comparison* window, choose *Spa clustering* > *Cluster Spa types*.
- 6.2 Select UPGMA, use the default alignment settings and default cost matrix and press < OK >.

The dendrogram is shown in the *Comparison* window (see Figure 6.7).

Comparison						- 0	×
e Edit Layout Groups Clustering Statistics Fingerprints Charact	cters Sequence Spa clustering Tren	dData ReadSets Spectra Composi	Window Help				
B = E ≭X £ C - Spa-repsuc	o 🖻 ⊾ !	S I 12.					
Experiments							
<all experiment="" types=""></all>			Information fields				
Name 🗨 🗉 👫 🖞 🗄	123 123 ANC ALC 000	<character name=""></character>	le_ ↓ ;↓ ↓∰ ↑	T			í
	Spa-repsuc		G ¹ 1 1 1 1 1 1 1 1	•			
	spa-repsuc						
	8 4 8 8 8 8	2 7 7 7 9 9					Spa-
< Spa-repsuc of	s_003 s_004 s_005 s_006 s_006	s_009 s_010 s_011 s_013 s_013 s_013 s_015				_	9
90 95 100			oup Key Modifie	date SpaType Rep	eatSuccession	Kreis 🔻	_
	r30 r17 r34 r17 r20 r17 r12					TO2MBMDI	
	r30 r17 r34 r17 r20 r17 r12					TO2MBMDI	
120	r23 r17 r34 r17 r20 r17 r12					TJMBMDM	-
		-18				2Z2EGMM	-
	r23 r12 r34 r34 r33 r34 r23 r12 r34 r34 r33 r154					UJGBBPB UJGBBP??	-
		22 - 24				UGFMEEBE	
	r12 r16 r02 r16 r02 r25 r17					WGKAKAC	-
	r12 r16 r02 r16 r02 r25 r17					WGKAKAC	-
	r12 r16 r02 r16 r02 r25 r17					WGKAKAC	
PA 🛛 🕇 🔭 🕞	r16 r02 r16 r02 r25 r17 r24			09 09:50:26 t019 08-	16-02-16-02-25-17-24	XKAKAOM	
r01	r12 r16 r02 r25 r24		Strain8 2020-06-	09 10:07:21 t974 01-	12-16-02-25-24	??GKAOQ	
Size Name 🔽 126	r23 r13 r23 r31 r29 r17 r31	29 r17 r25 r17 r25 r16 r28	Strain23 2020-06-	09 10:07:21 t022 26-3	23-13-23-31-29-17-31-29-17-25-1	TJEJNF2MI	
<		>	<			>	<>
13 entries in comparison 13 entries selected in database							

Figure 6.7: The *Comparison* window with a dendrogram and a similarity matrix.

6.3 Click on the dendrogram to place a cursor on any node or tip (where a branch ends in an individual entry). The average similarity at the cursor's place is shown in the upper part of the *Experiment data* panel. You can move the cursor with the arrow keys.

More detailed information about the dendrogram display settings can be found in the reference manual.

6.4 Save and close the *Comparison* window.

Chapter 7

Matching Spa types

7.1 Selections in the main window

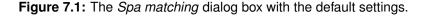
One or more selected Spa types can be matched (identified) against the complete database, all Spa types, or a selection in the database.

1.1 As an exercise, select a few entries in the *Main* window (e.g. **Strain22**, **Strain23**, and **Strain 24**).

7.2 Matching Spa types

2.1 Call the *Spa matching* dialog box with *Spa-Typing* > *Match Spa types*.

Spa matching			? ×
Alignment	250 %	Matrix & Results	
Gap creation cost: Gap extension cost:	250 %		Default V
Duplicate creation cost:	25 %	Match against: Result set size:	<all entries=""></all>
Duplicate extension	25 %	Result set size.	
Maximum duplication	3 reps		OK Cancel



In the Spa matching dialog box, following settings can be specified:

Alignment settings:

- Gap creation cost: specifies the cost for the introduction of a single gap in one of the repeats (in %).
- Gap extension cost: defines the cost for the extension of a created gap (in %).
- Duplicate creation cost: gives the cost for the duplication of a repeat (in %).
- Duplicate extension: defines the cost for the extension of a duplicated repeat (in %).
- *Maximum duplication length*: defines the maximum number of neighboring repeats that are taking into account to create a duplicate from.

Matrix & Results:

- **Cost matrix**: The drop-down menu lists the default cost matrix and the user-defined cost matrices (if created).
- Match against: The selection can be matched against all entries in the database (<All Entries>), all entries of which the currently logged-in user is the owner (<My Entries>), all entries currently loaded into memory (<Loaded Entries>), all selected entries (<Selected Entries>), or all known types (<All Spa types>).
- Result set size: Defines the number of best matches that are shown in the detailed report.
 - 2.2 For this exercise, choose <*All Entries*> from the *Match against* menu, leave all other settings at their defaults and press <*OK*>.

The program tries to find the best matches for the selected entries based on their repeats. The *Spa matching window* appears (see Figure 7.2).

🖆 Spa matching					– 🗆 X
File Window Help	0				
Table					
Key (unknown)	Match distance	Repeats (unknown vs. match)	Key (best match)	ЅраТуре	Kreiswirth
Strain22	0	15-12-16-02-16-02-25-17-24-24 15-12-16-02-16-02-25-17-24-24	Strain6	t012	WGKAKAOMQQ
Strain23	975	26-23-13-23-31-29-17-31-29-17-25-17-25-16-28 26-23-17-34-17-20-17-12-17-16	Strain24	t022	TJEJNF2MNF2MOMOKR
Strain24	100	26-23-17-34-17-20-17-12-17-16 26-30-17-34-17-20-17-12-17-16	Strain4	t002	ТЈМВМДМGМК
				·	
<					

Figure 7.2: The Spa matching window.



the repeat information stored in the associated character type will be used when matching entries. The repeat succession stored in the associated repeat information field is only used when no repeat information is present in the associated character type.

- In the first column, the keys of the selected "unknown" entries are shown.
- The distance between the selected entry and its match is displayed in the second column. The smaller the value, the better the match with "0" being an exact match.
- The repeats of the selected entries and their matches are shown in the third column.
- The fourth column displays the best matching entry.
- In the last column(s), the entry field content of the unknown entry is listed.

2.3 Double-click on an entry in the Spa matching window (e.g. entry with key Strain22).

A detailed report pops up (see Figure 7.3). The best matching entries are shown in descending order.



In both report windows, you can select or unselect entries by pressing the **Ctrl-** or **Shift**-key while holding the left mouse button.

File Window H	Help			
Table				
Key (match)	Match distance	Repeat	SpaType	Kreiswirth
Strain22		15-12-16-02-16-02-25-17-24-24	t012	WGKAKAOMQQ
Strain6	0	15-12-16-02-16-02-25-17-24-24	t012	WGKAKAOMQQ
Strain1	50	15-12-16-02-16-02-25-17-24-24-24-24	1253	WGKAKAOMQQQQ
Strain2	300	08-16-02-16-02-25-17-24	t019	ΧΚΑΚΑΟΜQ
Strain8	325	01-12-16-02-25-24	t974	??GKAOQ
Strain9	900	07-12-21-17-13-13-34-34-33-34	t186	UGFMEEBBPB
Strain7	950	14-44-13-12-17-17-17-23-18	t659	I2Z2EGMMMJH2
Strain4	1000	26-30-17-34-17-20-17-12-17-16	t001	TO2MBMDMGMK
Strain5	1000	26-30-17-34-17-20-17-12-17-16	t001	TO2MBMDMGMK
Strain24	1000	26-23-17-34-17-20-17-12-17-16	t002	TJMBMDMGMK
Strain3	1050	07-23-12-34-33-34	t044	UJGBBPB
Strain10	1050	07-23-12-34-34-33-154	Unknown	UJGBBP??
Strain23	1325	26-23-13-23-31-29-17-31-29-17-25-17-25-16-28	t022	TJEJNF2MNF2MOMOKR

Figure 7.3: Detailed report of the Spa matching window.

Chapter 8

Synchronizing with SpaServer

8.1 SpaServer information fields

In BIONUMERICS it is possible to submit new Spa types to the online SpaServer via a synchronization process. Spa data can only be submitted to the online SpaServer, if all mandatory SpaServer strain information is provided when uploading the information to the SpaServer.

Mandatory SpaServer strain information includes: **isolation date** (YYYY-MM-DD), **country**, **MR-SA/MSSA** (MRSA, MSSA), and **origin** (person, animal, environment, unknown). More information can be found on the *Submission* page of the SpaServer website.

A number of information fields are automatically created when a new database is created and after installation of the *Spa typing plugin* (see Figure 1.7).

In addition, extra information fields can be added to the *Database entries* panel with *Edit* > *Information fields* > *Add information field...*. This command can also be accessed by rightclicking in the information toolbar of the *Database entries* panel.

- 1.1 Add information fields to the database for the storage of all mandatory SpaServer strain information: isolation date, country, MRSA/MSSA, and origin (see Figure 8.3 for an example).
- 1.2 Optionally, add information fields to the database for the storage of additional (not mandatory) strain information (e.g. City, ZIP,...).

Strain information can be entered in the database in several ways:

- Importing information stored outside BIONUMERICS (e.g. in a text file or an ODBC-compatible source) with the import routines (select *File > Import...* (, Ctrl+I) to call the *Import* dialog box).
- Entering information using the *Entry* window (double-click on an information field to call this window).
- Editing information directly by clicking twice on an information field.

Detailed information on each of these options can be found in the reference manual. In this section only the second option will be illustrated.

1.3 Double-click on a database entry to open the *Entry* window. Right-clicking on the entry, and selecting *Open highlighted entry* also opens this window.

In default configuration, the upper left panel of the *Entry* window shows the information fields. The upper right panel shows the available experiments for the entry (see Figure 8.1).

🖆 Entry edit				_	×
	Edit Sequence Experiment Window He	elp			
√ок 🗙 🚰	Ď ⊗ 🛛 Ĺ, Ē				
Database fields			Experiments		
Кеу	lso1		<all experiment="" types=""></all>		
SpaType	Unkown	-	12 ∞ • 6		
RepeatSuccession	15-12-17-02-16-02-25-17-24-24-24-24	-	Name		•
Kreiswirth		-	💍 🛟 Spa-typing		
Clonal complex		-	💍 📑 Spa-repsuc		
Date	2008-10-22	-			
Country	Austria	-			
Origin	person	-			
MRSA/MSSA	MRSA	*			
Attachments Depe	ndencies Crosslinks		Comparisons		
17100	\otimes		Þ		
Content type	Name	Descrip 🔷	Name		•
<		>			
Level:					.:

Figure 8.1: The Entry window.

1.4 Enter the information in the fields.

In the *SpaServer synchronization settings* dialog box (see 8.2), the BIONUMERICS information fields containing mandatory (and optional) strain information can be linked to the SpaServer information fields (see 8.2.3). When there is a link present between the BIONUMERICS information fields and the **MRSA/MSSA** and the **Origin** SpaServer information fields (see 8.2.3), the history lists for these BIONUMERICS information fields contain all possible online options (**Origin**: unknown, person, animal, environment; **MRSA/MSSA**: MRSA, MSSA). These history lists can be used to save time and work and to avoid typographical errors.

1.5 The history lists can be accessed by clicking the button on the right hand from the information field in the *Entry* window. A floating menu appears from which the correct information string can be selected (see Figure 8.2).

Database fields		
Кеу	lso2	*
SpaType	Unkown	-
RepeatSuccession	08-23-12-34-34-33-34	-
Kreiswirth		-
Clonal complex		-
Date	2008-09-11	-
Country	Belgium	-
Origin	animal	•
MRSA/MSSA	MRSA	animal
		unknown
Attachments Depe	ndencies Crosslinks	environment
[] [] ◎	\otimes	person

Figure 8.2: History list in the *Entry* window.

1.6 Press the **Enter**-key or select < **OK**> to close the *Entry* window. The information is stored in the database.

Spa typing plugin - BioNumerics				- 0 X
File Edit Database Analysis Spa-Typing Scripts Window	Help			
	http			
Experiment types	Database entries			Comparisons
@ + [∄ ⊗ 🖳 Ē 🤜 ↑	🕄 🕂 💾 🛞 🕄 🛛 🔂 🗸 - <selected entr<="" td=""><td>ries> し</td><td></td><td>+ 🖒 ⊗ 🗞 🛍</td></selected>	ries> し		+ 🖒 ⊗ 🗞 🛍
# Name Type 🔻	Key SpaType RepeatSuccession	Date Country Origin MRS	SA/MSSA 🔻 1 2	Name N 🔫
AC 1 Spa-typing Sequence types	Iso1 Unkown 15-12-17-02-16-02-25-17-24-24-24-24	2008-10-22 Austria person MRS	iA • •	
2 Spa-repsuc Character types	Iso2 Unkown 08-23-12-34-34-33-34	2008-09-11 Belgium animal MRS	A •	
< >				
Entry fields Database design				
				< >
+ [2] ⊗ 🗟 🗗 🤄 ↑ ↓ 🧃				Identification projects Decis. NW.
Name Field type 🗨				29 + 🗗 ⊗ 民 I
✓ MC SpaType Fixed ^				
V MC RepeatSuccession Fixed				Name N 🔻
Alt Date Fixed Alt Country Fixed				
V MK Origin Fixed				
V MK MRSA/MSSA Fixed				
< >				< >
Fingerprint files Power assemblies Annotations				Alignments BLAST CHRMC
				+ 🖻 🛛 🗟 🔓
File name Experiment type Link 🔻				Modified date 🗨
< >	<		>	< >
Database: Spa typing plugin (_DefaultUser_) Entries: Loaded=15, '	View=2, Selected=2 2 experiments C:\Users\Public\Documents\BioNumeri	cs\Data 80\Spa typing plugin This is a time limited package valid i	until 2020-12-30	

Figure 8.3: The *Main* window with information fields containing date, country, MR-SA/MSSA and origin information.

8.2 SpaServer synchronization settings

8.2.1 Introduction

Before new Spa types can be submitted to the SpaServer, some synchronization settings need to be specified in BIONUMERICS. These settings can be accessed via the menu command *Spa*-*Typing* > *SpaServer synchronization settings*.

2.1 Select *Spa-Typing* > *SpaServer synchronization settings* in the *Main* window.

This pops up the SpaServer synchronization settings dialog box (see Figure 8.4).

8.2.2 Add SpaServer users

Synchronization with the SpaServer is only possible if at least one registered SpaServer user is defined in the BIONUMERICS database. To edit or view the user settings of the user selected in the list *SpaServer users panel* press the <*Edit*> button. All user information of the selected user is deleted with the <*Delete button*>. To add the contact details of a registered SpaServer user to the database, select the <*Add*> button in the *SpaServer users panel*.

2.2 Press the <*Add*> button to call the *Add user* dialog box (see Figure 8.5).

The Add user dialog box prompts for the user ID of the new user.

2.3 Enter a user ID in the *Add user* dialog box and press <*OK*> to call the *Edit SpaServer user* dialog box (see Figure 8.6).

In the *Edit SpaServer user* dialog box, all information fields marked with an asterisk are mandatory fields.



To obtain a *SeqNet.org release code*, please contact SeqNet.org.

A BIONUMERICS script is available that generates XML files of the certification trial data processed in BIONUMERICS. Please contact Applied Maths to obtain this script.

SpaServer u	sers		SpaServer Info			
Jack		Add	Strain info to be synd	chronized with the Spa	Server:	
			SpaServer Info	Database Info field		^
		Edit	Isolation date (*)	DATE		
			Isolate city			
		Delete	Isolate ZIP			
			Isolate state			
			Isolate country (*)	COUNTRY		
SpaServer re	sults		Origin (*)	ORIGIN		
Assign databa	ase info field to	store	Acquisition			U
SpaServer re	sults		Accoriation			
Spa Type:	<new field=""></new>	~	Assign info field Assign database info	field to selected SpaS	erver ir	ıfo
Repeat succession:	<new field=""></new>	~	DATE		\sim	
NOTE					C	к
		rains is published or	 00			

Figure 8.4: The SpaServer synchronization settings dialog box.

Add user	?	×
New user ID:	John	
	, OK	
[Cance	el

Figure 8.5: The Add user dialog box.

Edit SpaServer user	? ×
Contact info Salutation: Mr. Title: First name: Last name*: Email*:	Lab info Organization* Department: Street: Nr: City*: ZIP:
Allow email contact from other users Fields marked with a "*" are mandatory For information on how to obtain a release code, please contact SeqNet.org.	State: Country*: SeqNet.org release code*: Import lab info OK Cancel

Figure 8.6: The Edit SpaServer user dialog box.

When there is at least one user defined in the list of SpaServer users, the option <*Import lab info*> becomes available when a new SpaServer user is added to the database. Pressing this button calls the *Import lab info* dialog box.

The Lab information of the selected user is copied to the *Lab info panel* of the new user when pressing < OK >.

2.4 Enter the user information in the *Edit SpaServer user* dialog box and press <*OK*> to add the SpaServer user to the database.



Figure 8.7: The Import lab info dialog box.

8.2.3 Link BIONUMERICS information fields to SpaServer fields

In the *SpaServer Info panel*, the SpaServer fields are listed in the *SpaServer info column*. SpaServer fields marked with an asterisk are mandatory fields, and need to be linked to one of the BIONU-MERICS information fields. Fields without an asterisk are optional fields. The database information field selected in the *Assign info field panel* is displayed in the right column of the *SpaServer info panel*. All mandatory SpaServer fields need to be linked to their corresponding database information fields. If one or more mandatory SpaServer fields are not linked to a BIONUMERICS information field, the synchronization will fail.



The entry key is automatically linked to the SpaServer ID.

- 2.5 In the *SpaServer info panel*, select a SpaServer info field in the left column and select the corresponding database information field from the drop-down menu in the *Assign info field panel*.
- 2.6 Repeat the previous action for at least all mandatory SpaServer fields.

8.2.4 Store SpaServer results in BIONUMERICS database fields

BIONUMERICS information fields to store the SpaServer **SpaType** and **Repeat succession** results can be selected from two drop-down menus in the *SpaServer results panel*. An existing or new field can be selected from the list.

2.7 After having specified all settings in the *SpaServer synchronization settings* dialog box press <*OK*>.

New info field	?	×	New info field	?	×
New information field for 'Spa Type': Spa Type SpaServer	ОК		New information field for 'Repeat succession': Repeat succession SpaServer	ОК	
	Cance	el		Cance	el

Figure 8.8: The New info field dialog box.

If the online results are assigned to new database information fields the *New info field* dialog box is displayed prompting for the information field name(s).

2.8 Specify a new information field name for the storage of the spa type and/or the repeat succession, or keep the default suggested name(s). Press < OK > to add the fields to the database.

8.3 Synchronizing with SpaServer (batch mode)

Spa data present in the BIONUMERICS database can be synchronized with the SpaServer in *batch*.

- 3.1 Select the entries you wish to synchronize with the SpaServer. To select entries use the **Ctrl**and **Shift**-keys. Check boxes for selected entries are indicated as **v**.
- 3.2 Select *Spa-Typing* > *Synchronize with SpaServer* in the *Main* window.

Select SpaServer submitter	?	×
Select submitter:		
Jack		
John		
	OK	
	Cance	el

Figure 8.9: The Select SpaServer submitter dialog box.

If more than one SpaServer user is defined in the database, the *Select SpaServer submitter* dialog box pops up, listing all users defined in the *SpaServer synchronization settings* dialog box. Select the correct user from the list and press < OK >.

When there are no users defined in the database, an error message pops up (see Figure 8.10 (a)). Select *Spa-Typing* > *SpaServer synchronization settings* in the *Main* window, and press the <*Add*> button to add a spa user to the database (see 8.2).

When one or more mandatory SpaServer fields are not linked to a BIONUMERICS information field, an error message pops up (see Figure 8.10 (b)). When pressing < OK >, the *SpaServer synchronization settings* dialog box automatically pops up.



Figure 8.10: An error message pops up when (a) no SpaServer users are defined in the database,(b) when not all mandatory SpaServer fields are linked to a BIONU-MERICS information field.

3.3 When all fields are correctly linked BIONUMERICS tries to submit the data to the Ridom/Seqnet SpaServer. The *SpaServer synchronization* dialog box is displayed (see Figure 8.11).

The *SpaServer synchronization* dialog box lists all selected entries. In the left column, the keys of the entries are shown. Depending on the information that BIONUMERICS has tried to submit to the online SpaServer, different messages are displayed in the **SpaServer response** and **BN** error columns in the *Report window*.

BN errors:

• The experiment "Spa-Typing" is not present: no Spa-Typing experiment is defined for the

Server synchro	onization			?
lesults:				
Entry	Spa Type	SpaServer response	BN error	
			Export to Notepad	Close

Figure 8.11: The SpaServer synchronization dialog box.

selected entry.

- Sequence is empty: a Spa-Typing experiment is present for the selected entry, but this experiment does not contain an assembly.
- Assembly quality problem: the assembly quality settings differ from the default settings that are required for the submission of new types to the SpaServer, and/or the number of active traces in the contig is not equal to 2, and/or one or both trimming patterns are not detected on the consensus sequence.
- Problems with assembly: the status of the assembly is not set to solved.
- Invalid input in field "Name information field": the information field in the database does not contain the correct information.

SpaServer responses:

- *New Spa type; SpaServer ID*: a new Spa type is submitted to the SpaServer. The new Spa type is shown in the **Spa Type** column (see Figure 8.11).
- *Existing Spa type*: this message is displayed when an existing spa type is submitted to the SpaServer. The Spa type is shown in the second column.



Only the strain info of NEW Spa types are stored. Updating information fields of already existing types is not possible.

- The LabCode ****** is not valid!: the SeqNet.org release code of SpaServer submitter is invalid.
- Server in maintenance
- ...

To export the SpaServer synchronization results to notepad use the *<Export to Notepad>* button. The *SpaServer synchronization* dialog box can be closed with the *<Close>* button.

8.4 Synchronizing with SpaServer (entry mode)

- 4.1 Double-click on a database entry to open the *Entry* window (see Figure 8.1). Right-clicking on the entry, and selecting *Open entry* also works.
- 4.2 In the *Entry* window select *Spa-Typing* > *Synchronize with SpaServer*.

When no users are defined in the database, the error message 'No SpaServer users defined yet' is generated. In that event, select *Spa-Typing* > *SpaServer synchronization settings* in the *Main* window, and press the <*Add*> button to add a spa user to the database (see 8.2).

Prior to the submission of the data to the SpaServer, BIONUMERICS checks the presence of **BIONUMERICS errors** for this entry (see 8.3 for all possible BIONUMERICS errors):

- As a first check, BIONUMERICS checks if the experiment *Spa-typing* is present for the entry.
- In a second step, BIONUMERICS checks if the Spa-typing experiment contains an assembly.

If BIONUMERICS detects the presence of an assembly, the program checks if problems are present in the assembly and reports this:

- The assembly quality settings differ from the default settings that are required for the submission of new types to the SpaServer.
- The number of active traces in the contig is not equal to 2.
- One or both trimming patterns are not detected on the consensus sequence.
- The status of the Assembly reports an error (= red status box).

If none of the above described errors are present for the entry, BIONUMERICS checks in a next step if the information fields that are linked to one of the mandatory SpaServer information fields, contain (the correct) information.

Error ? × Some SpaServer info field contents are not in the proper format. Please provide the information requested below: (fields marked with a (*) are mandatory) Isolation date (*) V Tuesday , June 9,2020 V NOTE: information fields in the database will also be changed. OK Cancel			
Please provide the information requested below: (fields marked with a (*) are mandatory) Isolation date (*) I Tuesday , June 9,2020 V NOTE: information fields in the database will also be changed.	Error	?	×
(fields marked with a (*) are mandatory) Isolation date (*)	Some SpaServer info field contents are not in the	proper fo	ormat.
NOTE: information fields in the database will also be changed.	the second se		
	Isolation date (*) Tuesday , June	9, 2020	~
OK Cancel	NOTE: information fields in the database will also	be chang	ed.
	ОК	Cancel	I

Figure 8.12: The Error dialog box.

When BIONUMERICS detects one or more information fields that do not contain the correct information, the *Error* dialog box pops ups, listing all the information fields that are not in the proper format. Select the correct information from the drop-down list(s) and press < OK >.

4.3 If no BIONUMERICS errors are detected, and if all fields are properly filled in, the *Submit to Ridom/SeqNet SpaServer* dialog box pops up (see Figure 8.13).

The first time spa information for an entry is submitted to the SpaServer, the *Submit to Ridom/SeqNet SpaServer* dialog box pops up, listing all spa information that BIONUMERICS will try to

Select a Spat	Server user: Jack ~	Strain info	Content		J
		Fwd. strain Rev. strain	C:\Users\10023342\Docum C:\Users\10023342\Docum		
		Edits	2	ients\Data	p
User info	Content	Quality	2 97.9		
Name	Mr. Jack Jones	Quality	51.5		
Department		Isolate ID	Iso2		
Organization	My demo organization	Isolation date (*)			
Street + Nr		Isolate city			
ZIP + City	Gent	Isolate ZIP			
State		Isolate state			
Country	BELGIUM	Isolate country (*)	Belgium		
Email	Jack.jones@biomerieux.com	Orinin (*)	nerson		
		<		>	

Figure 8.13: The Submit to Ridom/SeqNet SpaServer dialog box.

submit to the SpaServer (see Figure 8.13). The SpaServer submitter can be selected from the drop-down list in the left upper panel. Optionally, a comment can be entered in the *Comment* box.

The **Quality** score displayed in the list in the right panel is calculated based on the **Assembly Quality** settings. The quality of a strain is the percentage of bases in the consensus that has an average sequence base quality greater than or equal to 100. Spa data can only be submitted to the online SpaServer if the quality of the strain is greater than 70.

Pressing the <*Submit*> button submits the spa data to the online SpaServer. To cancel the submission, press the <*Cancel*> button.

4.4 Press the <*Submit*> button.

BIONUMERICS submits the spa data to the online SpaServer. Depending on the information that is submitted, **SpaServer response** errors may pop up (see 8.3 for all possible **SpaServer response** errors).

If (corrected) spa information is **re-submitted** to the SpaServer (e.g. when the server returned an error that was subsequently corrected), the *Submitted SpaServer data* dialog box pops up, displaying all previously submitted information. The *SpaServer response* of the submission is displayed in the *Response* field.

Pressing the *<Submit again*> button will pop up the *Submit to Ridom/SeqNet SpaServer* dialog box, listing all spa information for the new submission.



Bibliography

[1] G. Benson. Sequence alignment with tandem duplication. *Journal of Computational Biology*, 4(3):351–367, 1997.