

## **BioNumerics** Tutorial:

# Creating a minimum spanning tree based on MLST data

## 1 Aim

In this tutorial we will create a minimum spanning tree based on MLST data. We will also see how we can alter the layout of the minimum spanning tree and how to export the picture to use it in a publication, presentation, etc.

## 2 Preparing the database

#### 2.1 Introduction to the MLST demo database

The **MLST demo database** contains for 500 *Neisseria meningitidis* isolates following information: a unique identifier ("Key"), a strain number, an MLST sequence type that was deduced from the analysis ("ST"), the clonal complex information ("CC"), the serogroup, the country where the strains originate from, the year of isolation, the species and the disease in which the strains were involved (see Figure 1).

The allele number is reported for each of the seven loci sequenced (sequence types **abcZ**, **adk**, **aroE**, **fumC**, **gdh**, **pdhC** and **pgm**) for all 500 strains and is stored in the **MLST** character type experiment.

The **MLST demo database** can be downloaded directly from the *BioNumerics Startup* window (see 2.2), or the data can be imported from a file available on our website, in a new, empty BioNumerics database (see 2.3), or the database can be restored from a back-up file available on our website (see 2.4).

File	Edit	Database Ana		Window	Help													
Data	ibase e	entries											1	xperin	ient t	ypes		
ź		► 🏄 🗙 →	š 🖌 🧉 👻	<aii< th=""><th>Entries&gt;</th><th>10</th><th></th><th>_</th><th></th><th></th><th></th><th></th><th></th><th>225</th><th>•</th><th>🌽 🗙 🐇</th><th></th><th><b>T</b></th></aii<>	Entries>	10		_						225	•	🌽 🗙 🐇		<b>T</b>
-	Key	Modified date	Strain	ST	CC	Serogroup	Country	Year	Species	Disease	Genus 🐨	1	T		#	Name	1	Гуре
1	1	2013-11-25 1	A4/M1027	4	ST-4 complex/subgrou	A	USA	1937	meningitidis	invasive(unspecified	Neisseria	•	A [		1	MLST	(	Character typ
1	10	2013-11-25 1		1	ST-1 complex/subgroup	A	Canada	1971	meningitidis	invasive(unspecified	Neisseria		E					
1	1015	2013-11-25 1	MDU 9720319	158	ST-41/44 complex/Lin	в	Australia	1997	meningitidis	invasive(unspecified	Neisseria							
/	1016	2013-11-25 1	MDU 9802543	159	ST-41/44 complex/Lin	В	Australia	1998	meningitidis	invasive(unspecified	Neisseria	•						
1	1017	2013-11-25 1	BM33a	160	ST-35 complex	в	Greece		meningitidis	carrier	Neisseria				•			•
1	1018	2013-11-25 1	BM34	161	ST-103 complex	в	Greece		meningitidis	invasive(unspecified	Neisseria	•	E	xperime	nt type	Entry fields		
1	1019	2013-11-25 1	BM35	162	ST-162 complex	в	Greece		meningitidis	invasive(unspecified	Neisseria	•		ingerp	in (- 6)	00		
1	1020	2013-11-25 1	BM55	163		В	Greece		meningitidis	invasive(unspecified	Neisseria			ingerp	inite in	es		
1	1021	2013-11-25 1	BM65	164		в	Greece		meningitidis	invasive(unspecified	Neisseria	•	- 14	۵.		🚽 🗡 🖏		<ai< td=""></ai<>
/	1027	2013-11-25 1		170	ST-41/44 complex/Lin	в	Netherlands	1982	meningitidis	unspecified	Neisseria	•		F	ile na	me	Experim	ent type
1	1028	2013-11-25 1		170	ST-41/44 complex/Lin	в	Netherlands	1986	meningitidis	unspecified	Neisseria	•			_			
/	1029	2013-11-25 1		170	ST-41/44 complex/Lin	в	Netherlands	1990	meningitidis	unspecified	Neisseria	•						
1	105	2013-11-25 1	Z2720	80	ST-4 complex/subgrou	A	Gambia	1983	meningitidis	unspecified	Neisseria	•						
/	1057	2013-11-25 1	BM95	187		в	Greece		meningitidis	invasive(unspecified	Neisseria	•						
1	1059	2013-11-25 1	BM95b	187		в	Greece		meningitidis	carrier	Neisseria	•						
1	1065	2013-11-25 1		191	ST-41/44 complex/Lin	В	Netherlands	1981	meningitidis	unspecified	Neisseria	•	F	ingerpri	nt files	Power assembl	es MTG	Ident. proj.
/	107	2013-11-25 1	Mrs 98082	1151	ST-11 complex/ET-37	В	Saudi Arabia	1998	meningitidis	meningitis	Neisseria	•					_	
/	1071	2013-11-25 1	M3388	5	ST-5 complex/subgroup II	A	Ghana	1997	meningitidis	unspecified	Neisseria	•	<u>د</u>	Compar	isons		-	
1	1072	2013-11-25 1	M970	7	ST-5 complex/subgroup III	A	Mongolia	1994	meningitidis	unspecified	Neisseria	•	- 64	+ (	۵.	× 🕆 🔰	<b>V</b>	<all compari<="" td=""></all>
/	1073	2013-11-25 1	M960	7	ST-5 complex/subgroup II	A	Mongolia	1994	meningitidis	unspecified	Neisseria	•		Nai	ne	1	Aodified d	ate
1	1080	2013-11-25 1		15	ST-364 complex	в	Netherlands	1983	meningitidis	unspecified	Neisseria	•						
1	109	2013-11-25 1	Z3204	4	ST-4 complex/subgrou	A	Gambia	1983	meningitidis	unspecified	Neisseria	•						
1	1095	2013-11-25 1		43	ST-41/44 complex/Lin	в	Netherlands	1973	meningitidis	unspecified	Neisseria	•						
1	1096	2013-11-25 1		43	ST-41/44 complex/Lin	в	Netherlands	1986	meningitidis	unspecified	Neisseria	•						
1	1					-						•	-	4				

Figure 1: The Main window of the MLST demo database.

#### 2.2 Option 1: Download the demo database from the Startup Screen

1. Click the *Download example databases* link, located in the lower right corner of the *BioNumerics Startup* window.

This calls the *Tutorial databases* window (see Figure 2).

File Database Tutorial Window         Tutorial databases         Name       Size (MB)       Min       Max         BandScoring demo database       19       7       7.6         DemoBase Connected       3       7.5       7.6         Nesseria MLST demo database       1       7.5       7.6         SIP demonstration database       0       7.5       7.6         SIP demonstration database       0       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         Winde genome SNPA analysis with mapping performed on the extern	🖪 Tutorial databases						x
Name       Size (MB)       Min       Max       Max         BandScoring demo database       19       7       7.6         DemoBase Connected       3       7.5       7.6         DemoBase Connected       3       7.5       7.6         Nesseria MLST demo database       1       7.5       7.6         SNP demonstration database       0       7.5       7.6         VGS demo database for Listeria monocytogenes       311       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         VGS demo database for Staphylococcus aureus       1086       7.5       7.6         VGS demo database for Staphylococcus aureus       1086       7.5       7.6         VGS demo database for Staphylococcus aureus       1086       7.5       7.6         VGS demo database for Staphylococcus aureus       1086       7.5       7.6         VGS demo database for Staphylococcus aureus       1086       7.5       7.6         VGS demo database for Staphylococcus aureus       1086       7.5       7.6         VGS demo database for Staphylococcus aureus       1086       7.5       7.6         VGS demo database for Staphylococcus aureus       1086       7.5       7.6	File Database Tutorial Window						
Name       Size (MB)       Min       Max         BandScoring demo database       19       7       7.6         DemoBase Connected       3       7.5       7.6         Neisseria MLST demo database       1       7.5       7.6         Neisseria MLST demo database       0       7.5       7.6         SNP demonstration database       0       7.5       7.6         WGS demo database       0       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo	Tutorial databases						
BandScoring demo database       19       7       7.6         DemoBase Connected       3       7.5       7.6         DemoBase Connected       3       7.5       7.6         Neisseria MLST demo database       1       7.5       7.6         SNP demonstration database       0       7.5       7.6         SNP demonstration database       0       7.5       7.6         WGS demo database for Listeria monocytogenes       311       7.5       7.6         WGS demo database for Staphylococccus aureus       1086       7.5       7.6         WGS demo database for Staphylococccus aureus       1086       7.5       7.6         WGS demo database for Listeria monocytogenes       311       7.5       7.6         WGS demo database for Staphylococccus aureus       1086       7.5       7.6         WGS demo database for Staphylococccus aureus       1086       7.5       7.6         WGS demo database for Staphylococccus aureus       1086       7.5       7.6         WGS demo database for Staphylococccus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6 <td>2</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	2						
Dambase       1 </td <td>Name</td> <td>Size (MB)</td> <td>Min</td> <td>Max</td> <td>-</td> <td></td> <td></td>	Name	Size (MB)	Min	Max	-		
Demodate connected       1       3       7.3       1.3       uses publicly available next-generation sequence reads from the Sequence         Neisseria MLST demo database       0       7.5       7.6       Image: Sequence reads from the Sequence         SNP demonstration database       0       7.5       7.6       Image: Sequence reads from the Sequence         WGS demo database for Listeria monocytogenes       311       7.5       7.6       Image: Sequence reads from the Sequence         WGS demo database for Staphylococcus aureus       1086       7.5       7.6       Image: Sequence reads from the Sequence         WGS demo database for Staphylococcus aureus       1086       7.5       7.6       Image: Sequence reads from the Sequence         Tutorialis       Image: Sequence from (whole genome) sequences       Image: Sequence from (whole genome) sequences of some Staphylococcus aureus are used in this turing we will screen a set of whole genome sequences of some Staphylococcus aureus are used in the extern       Image: Sequence from (whole genome Staphylococcus aureus are used in the start         Name       Image: Sequence from (whole genome) sequences       Image: Sequence from from from from do not not extern       Image: Sequence from from from from from from from from	BandScoring demo database	19	7	7.6	^		$\sim$
Nessena MLS1 demo database       1       7.5       7.5       Image: Constraint of the second sec	DemoBase Connected	3	7.5	7.6			
SNP demonstration database       0       7.5       7.6       into genome sequences. wgMLST alleles were called using the assembly-based and assembly-free method.         WGS demo database for Listeria monocytogenes       311       7.5       7.6       into genome sequences. wgMLST alleles were called using the assembly-based and assembly-free method.         WGS demo database for Staphylococcus aureus       1086       7.5       7.6       into genome sequences.         Tutorials       Image: Staphylococcus aureus       Image: Staphylococcus aureus       Image: Staphylococcus aureus       Image: Staphylococcus aureus         Name       Image: Staphylococcus aureus       Image: Staphylococcus       Image: Staphylococcus <td>Neisseria MLST demo database</td> <td>1</td> <td>7.5</td> <td>7.6</td> <td>=</td> <td></td> <td></td>	Neisseria MLST demo database	1	7.5	7.6	=		
WGS demo database for Listeria monocytogenes       311       7.5       7.6         WGS demo database for Staphylococcus aureus       1086       7.5       7.6         Tutorialis       Image: Comparison of the presence of the mechange of the staphylococcus aureus are used in this whole genome demonstration database of Staphylococcus aureus are used in this whole genome demonstration database of Staphylococcus aureus are used in this	SNP demonstration database	0	7.5	7.6			
Tutorials         Image: State of the sequences from (whole genome) sequences         Extracting subsequences from (whole genome) sequences         Image: State of the sequence of the sequence of the sequence of the sequence. The assembled sequences present in the whole genome data in this whole genome data in the whole genome data in the whole genome data in the sequence. The sequence of sequences are used in this whole genome demonstration database of Staphylococcus aureus are used in this database of	WGS demo database for Listeria monocytogenes	311	7.5	7.6			
Tutorials  Tutorials  Name  Extracting subsequences from (whole genome) sequences  Performing whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming whole genome SNP analysis with mapping performed on the extern  Ferforming wh	WGS demo database for Staphylococcus aureus	1086	7.5	7.6			$\sim$
Name       Image: Standard					×		
Name         Image: Constraint of the second se	Tutorials						
Extracting subsequences from (whole genome) sequences  Extracting subsequences from (whole genome) sequences  Performing whole genome SNP analysis with mapping performed on the extern  In this tutorial we will screen a set of whole genome sequences of some Staphylococcus  aureus samples for the presence of the mecA sequence. The assembled sequences present in the whole genome demonstration database of Staphylococcus aureus are used in this	4						
Performing whole genome SNP analysis with mapping performed on the extern	Name			-			
Performing whole genome site analysis with mapping performed on the extern.	Extracting subsequences from (whole genome) sequences			^			~
	Performing whole genome SNP analysis with mapping perfor	med on the e	extern	=			
	Whole genome MLST typing in the Staphylococcus aureus d	emonstration	data			nome demonstration database of Staphylococcus aureus are used in this	
· ·					tutonal.		$\sim$
				~	<u> </u>		

Figure 2: The *Tutorial databases* window, used to download the Neisseria MLST demonstration database.

2. Select the Neisseria MLST demo database from the list and select Database > Download ().

- 3. Confirm the installation of the database and press *<Yes* > after successful installation of the database.
- 4. Close the *Tutorial databases* window with *File* > *Exit*.

The Neisseria demo database appears in the BioNumerics Startup window.

5. Double-click the Neisseria demo database in the BioNumerics Startup window to open the database.

The *Main* window should look like Figure 1.

#### 2.3 Option 2: Import the data from an Excel file in a new database

- 6. Create a new database or open an existing database.
- 7. Import the MLST dataset from the example Excel file Neisseria MLST.xlsx as described in the tutorial: "Importing MLST data from an Excel file". The Excel file contains preprocessed MLST information for about 500 *Neisseria meningitidis* strains.

After import the *Main* window should look like Figure 1.

#### 2.4 Option 3: Restore demo database from back-up file

A BioNumerics back-up file of the **Neisseria MLST demo database** is also available on our website. This backup can be restored to a functional database in BioNumerics.

8. Download the file Neisseria.bnbk from http://www.applied-maths.com/download/ sample-data, under 'Neisseria MLST demo database'. In contrast to other browsers, some versions of Internet Explorer rename the Neisseria.bnbk database backup file into Neisseria.zip. If this happens, you should manually remove the .zip file extension and replace with .bnbk. A warning will appear ("If you change a file name extension, the file might become unusable."), but you can safely confirm this action. Keep in mind that Windows might not display the .zip file extension if the option "Hide extensions for known file types" is checked in your Windows folder options.

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

The *Main* window should look like Figure 1.

#### 3 Working in the database

The character data is stored in the character type MLST.

- 1. To view the values in a list, double-click on the experiment **MLST** in the *Experiment types* panel, select *Settings* > *General settings...* (*III*), select the *Experiment card tab* and change the representation to *List*. Close the two windows.
- 2. Click on a green colored dot in the *Experiment presence* panel to open the experiment card for an entry.

The imported allele numbers are displayed in the experiment card next to the corresponding housekeeping gene names.

Character	Value	Mapping	
abcZ	1	<+>	
adk	3	<+>	
aroE	3	<+>	
fumC	1	<+>	
gdh	4	<+>	
pdhC	2	<+>	
pgm	3	<+>	

Figure 3: The experiment card.

- 3. Close the experiment card by clicking in the left upper corner of the card.
- 4. Right-click on the **Serogroup** information field in the *Main* window and choose *Field properties* from the floating menu.
- 5. Press <*Add all*> to create all existing states for the **Serogroup** field. Confirm the action.
- 6. Check *Use colors* to display a specific color code for each field state (see Figure 5).

Data	ibase e	ntries						Exper
\$		• 🎿 🗙 붟   🥪 -	<a>All Ent</a>	tries>	<u> </u> ଅ			- 299
	Key	CC	Country	Disease	Genus	Serogroup	Species ST 💌 1	
	1	ST-4 complex/subgroup IV	USA	invasive(unspecified/other)	Neisseria	A	Field properties	
	10	ST-1 complex/subgroup VII	Canada	invasive(unspecified/other)	Neisseria	A	Disable highlighted field	
	1015	ST-41/44 complex/Lineage 3	Australia	invasive(unspecified/other)	Neisseria	В		
	1016	ST-41/44 complex/Lineage 3	Australia	invasive(unspecified/other)	Neisseria	В	Sort by field	perir
	1017	ST-35 complex	Greece	carrier	Neisseria	В	Sort by field (reverse)	ige
	1018	ST-103 complex	Greece	invasive(unspecified/other)	Neisseria	В	Sort by field (Numerical)	
	1019	ST-162 complex	Greece	invasive(unspecified/other)	Neisseria	В		
	1020		Greece	invasive(unspecified/other)	Neisseria	В	Sort by field (Numerical, reverse)	
	1021		Greece	invasive(unspecified/other)	Neisseria	в	Add information field	
	1027	ST-41/44 complex/Lineage 3	Netherlands	unspecified	Neisseria	в	Delete information field	
	1028	ST-41/44 complex/Lineage 3	Netherlands	unspecified	Neisseria	в		iger
	1029	ST-41/44 complex/Lineage 3	Netherlands	unspecified	Neisseria	в	Use as display field	
	105	ST-4 complex/subgroup IV	Gambia	unspecified	Neisseria	A	Edit field in selection Ctrl+	M pmp
	1057		Greece	invasive(unspecified/other)	Neisseria	в	Copy field in selection	+
	1059		Greece	carrier	Neisseria	в —		

Figure 4: Field properties.

Properties for field 'SEROGROUP'			? ×
Name: Serogroup Field storage type: Optimized for speed Content type: Text Field states A B C	Add Add all Remove	Changes: Restrict content Use colors Opacity: Default content: Content validati	Allow
	emove all Edit Sort		Level assignment Summary settings Calculation Ok Cancel

Figure 5: The *Database field properties* dialog box.

7. Press  $\langle OK \rangle$  to accept the new settings.

The *Database entries* panel is updated (see Figure 6).



Since it is also possible to create groups based on the field content in the *Comparison* window, we will use the content of the **CC** column as an example there (see 4).

### 4 Comparison window

- 1. In the Database entries panel of the Main window, select all entries using Edit > Select all (Ctrl+A).
- 2. Highlight the *Comparisons* panel in the *Main* window and select *Edit* > *Create new object...* (\*) to create a new comparison for the selected entries.
- 3. Click on the select *Characters* > *Show values* (2) to display the allele numbers in the *Experiment data* panel (see Figure 7).

File	Edit	Database Analysis Scrip	ts Window	Help														
-4¥ iete	ibase e		<b>L</b> .										Experi	montt	whore			_
Jala	ibase e					_		_		_	_	_		nenti	ypes	**		
3		🕨 🖌 🗶 👘		ntries>	<u>ଅ</u>	_	_	_	_	_			- 29	. *	-	×		in   1
	Key	CC	Country	Disease	Genus	Serogroup	Species	ST	Strain	Year 💌	1			#	Name		Ту	pe
1	1	ST-4 complex/subgroup IV	USA	invasive(unspecified/other)	Neisseria	A	meningitidis	4	A4/M1027	1937	•	•		1	1 MLST		Ch	aracter
1	10	ST-1 complex/subgroup //I	Canada	invasive(unspecified/other)	Neisseria	A	meningitidis	1		1971	•			4				. F
1	1015	ST-41/44 complex/Lineage 3	Australia	invasive(unspecified/other)	Neisseria	В	meningitidis	158	MDU 9720319	1997	•		Evenerim	anthon	es Entry			
1	1016	ST-41/44 complex/Lineage 3	Australia	invasive(unspecified/other)	Neisseria	В	meningitidis	159	MDU 9802543	1998	•		Experim	ent typ	es chuy	TIERUS		_
1	1017	ST-35 complex	Greece	carrier	Neisseria	В	meningitidis	160	BM33a		•		Finger	orint fi	les			
1	1018	ST-103 complex	Greece	invasive(unspecified/other)	Neisseria	В	meningitidis	161	BM34		•				4	× *		, [
1	1019	ST-162 complex	Greece	invasive(unspecified/other)	Neisseria	В	meningitidis	162	BM35		•			• File na		· · · · · · · · · · · · · · · · · · ·		
1	1020		Greece	invasive(unspecified/other)	Neisseria	В	meningitidis	163	BM55		•			riie na	me		Experimen	t type
1	1021		Greece	invasive(unspecified/other)	Neisseria	В	meningitidis	164	BM65		•	_						
1	1027	ST-41/44 complex/Lineage 3	Netherlands	unspecified	Neisseria	В	meningitidis	170		1982	•	_		٠ 📃				Þ
1	1028	ST-41/44 complex/Lineage 3	Netherlands	unspecified	Neisseria	В	meningitidis	170		1986	•		Fingerp	int files	Power	assemblies	s MTG k	Jent. pr
1	1029	ST-41/44 complex/Lineage 3	Netherlands	unspecified	Neisseria	В	meningitidis	170		1990	•	_	Compo	rác o na				-
1	105	ST-4 complex/subgroup IV	Gambia	unspecified	Neisseria	A	meningitidis	80	Z2720	1983	•	_	Compa	IIISOII	5			
1	1057		Greece	invasive(unspecified/other)	Neisseria	В	meningitidis	187	BM95		•	_	+	<i>2</i>	× 72		V </td <td>All Com</td>	All Com
1	1059		Greece	carrier	Neisseria	В	meningitidis	187	BM95b		•		N	me		Mo	odified dat	e
1	1065	ST-41/44 complex/Lineage 3	Netherlands	unspecified	Neisseria	В	meningitidis	191		1981	•							
1	107	ST-11 complex/ET-37 com	Saudi Arabia	meningitis	Neisseria	В	meningitidis	1151	Mrs 98082	1998	•	-		_				
	٠ 📖			III						•	٠.	•	•					

Figure 6: The Main window.

- 4. In the *Comparison* window, right-click in the header of the "CC" field and select *Create groups from database field* from the floating menu. Alternatively select *Groups* > *Create groups from database field*.
- 5. In the *Group creation preferences* dialog box, make sure *Create largest group first* is selected, select *Skip empty content*, specify a maximum count of *20* and press *<OK>* twice.

Every clonal complex with at least three members is now assigned to a unique group. The 20 groups appear in the *Groups* panel along with their color, size and name (see Figure 7).

Comparison																		2
ile Edit Layout Groups Clust	ering	Statistics	Finger	prints	C	harac	ters	Seq	uenc	e Tren	dData	GenomeMaps ReadS	ets Spectra	Composite Window H	elp			
🖶 👝 🛛 🥪 🔹 🐒		2 3	ML	.ST					- 2	6								
kperiments		Ð	-													_		Π
<all experiment="" types=""></all>		Ð	Expe		nt di	ata				Infor	mation f	elds			_		Similaritie	s
		~				-	nc]	000		NK		<b>\$↑ ↓</b> <sup>100%</sup>					i 47	
Name		i -				23	1BC	ABC		13	1	i i i i i i i i i i i i i i i i i i i	+				1	ř.
MLST	<u> </u>		N 122	ILST														
	E																	
	-		13		щ	è	-	é	F									
			ą	adk	aroE	func	đđ	pdho	mgq		Key	CC	Country	Disease	Genus	-		
beriments Local composite datasets			1	3	3	1	4	2	3		1	ST-4 complex/subg	USA	invasive(unspecified/other)	Neisseria	A		Ĩ
alyses			1	3	1	1	1	1	3		10	ST-1 complex/subg	Canada	invasive(unspecified/other)	Neisseria	A		
-			10	6	12	5	9	6	9		1015	ST-41/44 complex/L	Australia	invasive(unspecified/other)	Neisseria	в		
🌽 🗙 👘			3	10	9	5	9	6	9		1016	ST-41/44 complex/L	Australia	invasive(unspecified/other)	Neisseria	в		
Name	-		4	3	11	18	6	10	12		1017	ST-35 complex	Greece	carrier	Neisseria	в		
	-		8	7	6	17	26	18	2		1018	ST-103 complex	Greece	invasive(unspecified/other)	Neisseria	в		
	-		1	5	13	53	26	41	3		1019	ST-162 complex	Greece	invasive(unspecified/other)	Neisseria	в		
oups			12	2	9	9	9	54	10	<b>/</b>	1020		Greece	invasive(unspecified/other)	Neisseria	в		
🌽 🗙 📲 🍟			39	16	9	9	5	11	44	<b>/</b>	1021		Greece	invasive(unspecified/other)	Neisseria	в		
	-		3	6	9	60	9	6	9	<b>Z</b>	1027	ST-41/44 complex/L	Netherlands	unspecified	Neisseria	в		
Size Name			3	6	9	60	9	6	9		1028	ST-41/44 complex/L	Netherlands	unspecified	Neisseria	в		
117 ST-11 complex/ET-37 complex	<b>^</b>		3	6	9	60	9	6	9		1029	ST-41/44 complex/L	Netherlands	unspecified	Neisseria	в		
76 ST-41/44 complex/Lineage 3			1	3	15	31	4	2	3		105	ST-4 complex/subg	Gambia	unspecified	Neisseria	A		
37 ST-32 complex/ET-5 complex	-		1	6	50	38	35	6	9	1	1057		Greece	invasive(unspecified/other)	Neisseria	в		
32 ST-1 complex/subgroup VII			1	6	50	38	35	6	9	<ul> <li>Image: A set of the set of the</li></ul>	1059		Greece	carrier	Neisseria	в		
30 ST-5 complex/subgroup III			28	6	9	24	9	6	9	<b>Z</b>	1065	ST-41/44 complex/L	Netherlands	unspecified	Neisseria	в		
24 ST-4 complex/subgroup IV			2	3	15	3	8	110	20		x 107	ST-11 complex/ET-3	Saudi Arabia	meningitis	Neisseria	в		
17 ST-269 complex			1	1	2	1	3	2	3	<b>Z</b>	1071	ST-5 complex/subg	Ghana	unspecified	Neisseria	A		
13 ST-35 complex			1	1	2	1	3	2	19	<b>Z</b>	1072	ST-5 complex/subg	Mongolia	unspecified	Neisseria	A		
10 ST-376 complex			1	1	2	1	3	2	19	<b>Z</b>	1073	ST-5 complex/subg	Mongolia	unspecified	Neisseria	A		
< III )			•													•	٠ 📄	

Figure 7: The Comparison window with groups defined.

## 5 Advanced clustering window

A minimum spanning tree in BioNumerics is calculated in the *Advanced cluster analysis* window. This window can be launched from the *Comparison* window.

1. Select *Clustering* > *Calculate* > *Advanced cluster analysis...* or press the *i* button and select *Advanced cluster analysis* to launch the *Create network* wizard.

Due to the arbirariness of the allele numbers, the similarity coefficient for clustering MLST data is the categorical coefficient. The categorical coefficient compares the allele numbers to see if they are the same or different but does not quantify the difference. The predefined template *MST for categorical data* uses the categorical coefficient for the calculation of the similarity matrix, and will calculate a standard minimum spanning tree with single and double locus variance priority rules.

2. Specify an analysis name (for example MLST1), make sure *MLST* is selected, select *MST for categorical data*, and press *<Next>*.



To view and modify the settings of a selected template check the option *Modify template settings for new analysis*.

The Advanced cluster analysis window pops up. The Network panel displays the minimum spanning tree, the upper right panel (*Entry list*) displays the entries that are present in the tree. The *Cluster analysis method panel* displays the settings used, in this example the priority rules that result in the displayed network.

The colors of the comparison groups (see 4) are automatically shown as node colors, but this can very easily be changed to a field state grouping defined in the *Main* window (see 2):

- 3. Press 414 or choose *Display* > *Display settings* to open the *Display settings* dialog box.
- 4. In the *Node colors tab* select the **Serogroup** from the list and press *<OK>*.

The node colors are updated according to the serogroups.

- 5. A node or branch can be selected by clicking on them. To select several nodes/branches hold the **Shift**-key.
- 6. The zoom slider on the left always further zooming in or out on the network. The zoom slider on top adjusts the size of the nodes.
- 7. Select *Display* > *Zoom to fit* or press 🔀 to optimize the view of the tree.
- 8. Press 44 or choose *Display* > *Display settings* to open the *Display settings* dialog box again.
- 9. In the Branch labels and sizes tab, check Use logarithmic scaling.
- 10. In the *Node colors tab* select the **Comparison groups** option again from the list and make sure the option *Separate entries* is unchecked.
- 11. Press  $\langle OK \rangle$  to apply the new settings.

The Advanced cluster analysis window should now look like Figure 8.

In the *Advanced cluster analysis* window it is possible to create *partitions*. In case of an MST, the partitioning algorithm will group nodes in partitions (complexes) when the distance between the connected nodes is less than or equal to a distance entered by the user. As soon as a connection has a longer distance, the partition ends.

12. A partitioning can be created with *Edit* > *Create partitioning* or using the *button*. This calls the *Partitioning* dialog box.

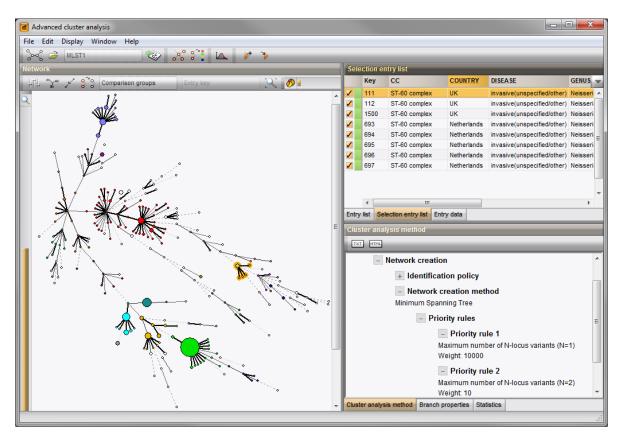


Figure 8: The Advanced cluster analysis window.

## 13. For the current example, enter a *Maximum distance between nodes in the same partition* of 2 and a *Minimum number of entries in a partition* of 2. Choose *Color from majority* and press *<OK>*.

The result looks as in Figure 9. The color of the partitions is adopted from the node colors. In case the nodes have different colors, the color from the majority is taken.

From this picture it is clear that the definition of a partitioning in an MST corresponds to the clonal complexes as defined for MLST and similar allele-based typing techniques.

14. The image can be exported with *File* > *Export image*.

15. Close the Advanced cluster analysis window and Comparison window with File > Exit.

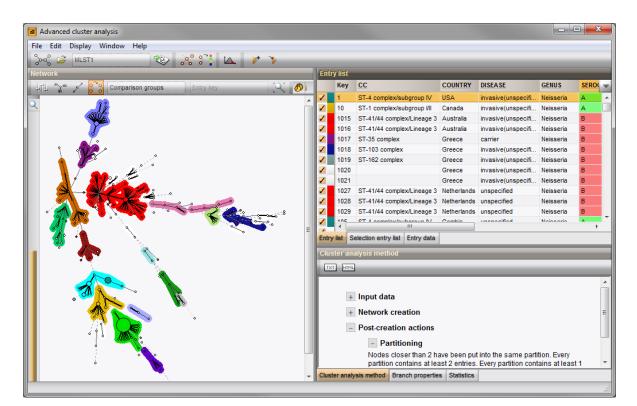


Figure 9: Partitions in the Advanced cluster analysis window.