

BIONUMERICS Tutorial:

wgMLST typing in the *Brucella* demonstration database

1 Introduction

This guide is designed for users to explore the wgMLST functionality present in BIONUMERICS without having to post calculation jobs on their own computer or on the external calculation engine. The whole genome demonstration database used in this tutorial contains the results obtained from the full wgMLST analysis in BIONUMERICS on publicly available sequence read sets and genome sequences of *Brucella* spp.

Although this guide provides the necessary information to start working with the wgMLST functionality present in BIONUMERICS, it is recommended to read the following documentation available for download on the tutorial page on our website:

- Tutorial "wgMLST typing: routine workflow starting from sequence read sets"
- Tutorial "wgMLST typing: routine workflow starting from imported genomes"
- Tutorial "wgMLST typing: detailed exploration of results"
- WGS tools plugin manual

Furthermore, a leaflet on our website (https://www.applied-maths.com/sites/default/files/ extra/Brucella-spp-how-to-make-the-most.pdf) explains how to use the trunk and branch wgMLST schema for *Brucella* spp. efficiently.

2 Preparing the demonstration database

The **WGS demo database** for *Brucella* can be downloaded directly from the *BIONUMERICS Startup* window (see 2.1), or restored from the back-up file available on our website (see 2.2).

2.1 Option 1: Download demo database from the Startup Screen

1. Click the *button*, located in the toolbar in the *BIONUMERICS Startup* window.

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

🖆 Tutorial databases					×
File Database Tutorial Window					
Tutorial databases					
<u>D</u>					
Name	Size (MB)	Min	Max	•	
BandScoring demo database	15	7	8	^	The DemoBase Connected demonstration
DemoBase Connected	4	7.5	8		database in BioNumerics version 8.0 format.
Neisseria MLST demo database	1	7.5	8		This database is used in several tutorials and
SNP demonstration database	0	7.5	8		comes with the biowumencs installation.
WGS demo database for Listeria monocytogenes	311	7.5	8		
WGS_demo_database_for_Brucella_spp	214	7.6	8		
WGS_demo_database_for_Escherichia_coli	601	7.5	8		
WGS_demo_database_for_Staphylococcus_aureus	624	7.5	8		
				~	×
	<			>	
Tutorials					
P ^A					
Name				•	This is the field Westerlage from the anti-state later of the
Band matching and polymorphism analysis				^	Principal Components Applyois (PCA) and a
Building automated decision and action workflows					Multi Dimensional Scaling (MDS)
Calculating a PCA and an MDS on a fingerprint data set					(sometimes also called Principal
Clustering a phenotypic test assay					Coordinates Analysis (PCoA)) on a
Combined analysis of character data					fingerprint data set and how to change the
Combined analysis of fingerprint data					layout of the obtained plots.
Configuring the database layout					
Dendrogram layout options					~
Entry information fields and their properties	_			~	



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

The WGS_demo_database_for_Brucella_spp appears in the BIONUMERICS Startup window.

5. Double-click the **WGS_demo_database_for_Brucella_spp** in the *BIONUMERICS Startup* window to open the database.

2.2 Option 2: Restore demo database from back-up file

A BIONUMERICS back-up file of the WGS demo database for *Brucella* is also available on our website. This backup can be restored to a functional database in BIONUMERICS.

6. Download the file WGS_BRU.bnbk file from https://www.applied-maths.com/download/ sample-data, under 'WGS_demo_database_for_Brucella_spp'.



In contrast to other browsers, some versions of Internet Explorer rename the WGS_BRU.bnbk database backup file into WGS_BRU.zip. If this happens, you should manually remove the .zip file extension and replace with .bnbk. A warning will appear ("If you change a file name extension, the file might become unusable."), but you can safely confirm this action. Keep in mind that Windows might not display the .zip file extension if the option "Hide extensions for known file types" is checked in your Windows folder options. 7. In the *BIONUMERICS Startup* window, press the button.

- 8. From the menu that appears, select *Restore database...*.
- 9. Browse for the downloaded file and select *Create copy*. Note that, if *Overwrite* is selected, an existing database will be overwritten.
- 10. Specify a new name for this demonstration database, e.g. "WGS Brucella demobase" (see Figure 2).
- 11. Click < OK > to start restoring the database from the backup file.

C:\Users\Public\WGS_E	RU.bnbk	Browse
Destination		
Overwrite		
 Overwrite the curr with a previous ve the backup will disa 	ent data in database WGS demo data rsion of the data as stored in a back appear, any old data that was erase	abase for Brucella spp. up. Any new data since d will be recovered.
Create copy		
 Create a new data have access to bo copy a database fr 	base from the data stored in a backu h old and current versions of an exi om another computer.	up. Use this option to isting database or to
New database name:	WGS Brucella demobase	

Figure 2: Restore Brucella demonstration database from backup file.

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

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

🛃 wgs	Brucella demobase - BioNun	nerics												-	• ×
File Edit	t Database Analysis Scr	ipts WGS tools Wine	dow	Help											
₩ Ξ	± ⊞ æ ©	<u>⊳</u> , ⊳ ©, °€		l l											
Experin	ment types		Dat	abase entries									Comparisons		
୶	+ [→ ⊗ 🗞		f] + P ⊗ (<all entries=""></all>		U					+ 🖻 🛛 🗟	, Ē. v.	<all comp<="" td=""></all>
	# Name	Туре 🔻		Key	Species	Strain	Run accession	Sample	Accession	Origin 🗸	1234567		Name	Modified date	
∏≩	1 wos	Sequence read set		BRUCELLA DEMO001	meltensis	16M			AE008917			^	_		
	2 woMLST	Character types		BRUCELLA DEMO002	suis	1330			CP002997			_			
	3 denovo	Sequence types	n	BRUCELLA DEMO003	ovis	ATCC 25840			CP000709						
	4 quality	Character types	n	BRUCELLA DEMO004	abortus	S19			CP000887						
⊒≩	5 wgs TrimmedStats	Sequence read set		BRUCELLA DEMO005	microti	CCM 4915			CP001578						
	6 wgMLST CallTypes	Character types		BRUCELLA DEMO006	ceti	TE10759-12			CP006897						
⊒≹	7 wasLong	Sequence read set		BRUCELLA DEMO007	canis	SVA13			CP007630						
	/			BRUCELLA DEMO008	neotomae	5K33			JMSC01000001				/	_	
		,		BRUCELLA DEMO009	abortus	870			CP007709						,
Entry fie	Ids Database design			BRUCELLA DEMO010	inopinata	802			NZ ADFA00				Identification projects	lecision networks	
				BRUCELLA DEMO011	vulpis	F60			LN997863						- 101
+		∠ T ↓		BRUCELLA DEMO012	canis	ATCC 23365			CP000872				29 + 🗁		♥
		Field type 🛛 🗨		BRUCELLA_DEMO013	meltensis	63/9			CP007789				Name		•
ARC	Species	Fixed A		BRUCELLA_DEMO014	meltensis	ATCC 23457			CP001488						
AK	Strain	Fixed		BRUCELLA_DEMO015	abortus	86/8/59			CP007765						
AK	Run accession number	Fixed		BRUCELLA_DEMO016	inopinata	B01	SRR4038982								
AIC	Sample Name	Fixed		BRUCELLA_DEMO017	pinnipedialis	B2-94	SRR4038996								
AK	Accession	Fixed		SRR2058928	suis	B00-0729	SRR2058928	Swine		TX					
AK	Origin	Fixed		SRR2058929	suis	B00-0771	SRR2058929	Swine		AL					
AK	MLST PubMLST 9 loci ST	Flexible y		SRR2058942	suis	B11-0001	SRR2058942	Cow		TX					
AK	د	>		SRR2058946	suis	B11-0193	SRR2058946	Cow		TX					
_				SRR2058961	suis	B11-0566	SRR2058961	Cow		TX					
Fingerpr	int files Power assemblies	Annotations		SRR2058968	suis	B13-0195	SRR2058968	Feralpig		TX			Alignments BLAST pro		omparisons
r =				SRR2058969	suis	B13-0198	SRR2058969	Feralpig		TX					<al aligner<="" td=""></al>
	T [] O est	EL		SRR2058971	suis	B13-0219	SRR2058971	Human		OR			+000	3 1 1 1	- All Alight
	File name Expe	eriment type 📃 🔻		SRR2058977	suis	B94-0444	SRR2058977	Swine		TX			Name	Modified dat	•
				SRR2058979	suis	B95-0322	SRR2058979	Swine		AR					
				SRR2058982	suis	B13-0257	SRR2058982	Swine		TX					
				SRR2058984	suis	B13-0234	SRR2058984	Human		Tonga					
				SRR2058985	suis	B13-0235	SRR2058985	Human		Tonga					
				SRR2058986	suis	B13-0236	SRR2058986	Human		Tonga					
				SRR2058987	suis	B13-0237	SRR2058987	Human		Tonga					
				SRR2058988	suis	B13-0238	SRR2058988	Human		Tonga					
	<	>		<						>		~			
Datab	ase: WGS Brucella demobase (DefaultUser) Entries	: Loade	ed=35. View=35. Selected=	0 7 experiments C.N.	Isers\Public\Docum	ents\BioNumerics\Data	80/WGS Bru	cella demobase	This is a time limited package	valid until 2020-12-30				

Figure 3: The Brucella spp. demonstration database: the Main window.

3 About the demonstration database

The *Brucella* spp. demonstration database contains 35 *Brucella* entries with linked data. The *WGS tools plugin* has already been installed in the demo database.

1. Select *WGS tools* > *Settings...* to access the settings of the plugin.

The calculation engine project is linked to the *Brucella* spp. allele database. No credits are assigned to this project so no jobs can be submitted to the external calculation engine, however since the option *Enable running jobs on my own computer* is checked in the *Calculation engine* tab, it is possible to run jobs on your own computer (see Figure 4).

alculation engine	Organism	Experiment hyp	T2 IMow		
Demote colouistic	Organism	Experiment type	es wymicor		
Remote calculatio	n engine —				
Uri:	Applie	ed Maths cloud	On premise	es	
	https://w	amist applied_mat	the com		
	mponti	gmot applied ma			
Project name:	eval_bru_	demo			
Password:					
Time out (sec):	30				
Own computer					
Enable runnin	g jobs on m	y own computer			
Interval (min):	1]			

Figure 4: The Calculation engine tab of the Calculation engine settings dialog box.

- 2. In the *Experiment types* tab, the experiment types that are created during installation of the plugin are listed. These experiments are automatically linked to the datasets used for wgMLST analysis:
- Character experiment type **wgMLST** contains the allele calls for detected loci in each sample, where the consensus from assembly-based and assembly-free calling (if performed) resulted in a single allele ID.
- Sequence read set type wgs contains the link to the sequence read files on NCBI.
- Sequence experiment type **denovo** contains (1) the imported (assembled) whole genome sequences, or (2) the concatenated de novo contig sequences resulting from the de novo assembly performed on the sequence read sets.
- Character experiment type **quality** contains quality statistics for the raw data (if available), the de novo assembly (if calculated) and the allele identification algorithm(s).



No data is available for the sequence read set type **wgsLong** in the demo database. This sequence read set is used to store links to long read sequence read data (e.g. PacBio or MinION datasets).

3. Click on the *wgMLST tab* and press the *<Auto submission criteria>* button.

By default, the *Use nomenclature acceptance criteria* option will be checked, meaning that the automatic submission settings are defined by the curator of the allele database.

4. Click <*Cancel*> in both dialog boxes.

Additional information (in entry info fields Strain, Sample name, Origin, etc.) was collected from the corresponding publications and added to the demonstration database.

The 35 *Brucella* entries can be divided in three categories:

- Published genome assemblies are imported for 15 database entries (Key BRUCELLA_ DEMO001 to BRUCELLA_DEMO015) and 2 database entries (Key BRUCELLA_DEMO016 and BRUCELLA_DEMO017) contain the link to sequence read set data NCBI's sequence read archive (SRA). These 17 entries represent the covered diversity of the *Brucella* spp. schema.
- 17 *Brucella suis* samples contain links to sequence read set data on NCBI's sequence read archive (SRA).
- 1 sample is present that has not yet been allocated to a species.

By clicking on one of the green dots next to an entry in the database, the corresponding results can be viewed, either in a separate window or in an experiment card for the character data types:

5. Click on a green colored dot for one of the entries in the first column in the *Experiment presence* panel. Column 1 corresponds to the first experiment type listed in the *Experiment types* panel, which is **wgs** in the default configuration.

In the *Sequence read set experiment* window, the link to the sequence read set data on NCBI (SRA) with a summary of the characteristics of the sequence read set is displayed: *Read set size*, *Sequence length statistics*, *Quality statistics*, *Base statistics* (see Figure 5).

- 6. Close the Sequence read set experiment window.
- 7. Click on the green colored dot for one of the entries in the second column in the *Experiment presence* panel. Column 2 corresponds to the second experiment type listed in the *Experiment types* panel, which is **wgMLST** in the default configuration.

Character experiment type **wgMLST** contains the allele calls for detected loci in each sample, where the consensus from assembly-based and assembly-free calling (if performed) resulted in a single allele ID (see Figure 6).

- 8. Close the character experiment card by clicking on the triangle in the top left corner.
- 9. Click on the green colored dot for one of the entries in the third column in the *Experiment presence* panel. Column 3 corresponds to the third experiment type listed in the *Experiment types* panel, which is **denovo** in the default configuration.

Depending on the entry, the *Sequence editor* window contains (1) the imported (assembled) whole genome sequence, or (2) the concatenated de novo contig sequences resulting from the de novo assembly performed on the sequence read set.

- 10. Close the *Sequence editor* window.
- 11. Click on the green colored dot in column 4 to open the **quality** character card (default configuration) for an entry in the database.

The **quality** character card contains quality statistics for the raw data (if present), the de novo assembly (if calculated) and the different allele identification algorithm(s).

12. Close the character experiment card by clicking on the triangle in the top left corner.

C SRR2058929	_									
File Preprocessing Analysis Window Help										
Sequence read set report										
Sequence read set information										
- Storage										
Storage by link: NCBI SRR2058929										
- Read set size										
Number of sequences: 750000										
Number of bases: 330199710, 1st end 164260568, 2nd end 165939141										
- Sequence length statistics										
Average sequence length: 220.13, 1st end 219.01, 2nd end 221.25										
Standard deviation of the sequence length: 49.38, 1st end 49.66, 2nd end 49.07 Minimum sequence length: 35, 1st end 35, 2nd end 35										
Maximum sequence length: 251, 1st end 251, 2nd end 251										
- Quality statistics										
Average base quality: 32.82, 1st end 35.10, 2nd end 30.57 Standard deviation of the base quality 9.40, 1st and 6.49, 2nd and 0.59										
Minimum base quality: 2, 1st end 2, 2nd end 2										
Maximum base quality: 40, 1st end 40, 2nd end 40										
Q20: 282915019 (85.68%), 1st end 153065998 (93.18%), 2nd end 129849021 (78 Q25: 275767530 (83.52%). 1st end 150981392 (91.92%). 2nd end 124786138 (75	.25%) .20%)									
Q30: 262994821 (79.65%), 1st end 147212909 (89.62%), 2nd end 115781912 (69	.77%)									
- Base statistics										
Number of bases A: 70881220 (21.47%)										
Number of bases C: 94465289 (28.61%) Number of bases G: 94126993 (28.51%)										
Number of bases T: 70544637 (21.36%)		~								
Analyses Analysis info										
Field name Field value		-								
Analysis name Analysis type		^								
~		~								
γ <		>								

Figure 5: The Sequence read set experiment window.

SRR2058929									
Character	Value	Mapping	•						
BRUC_1	4	<+>	^						
BRUC_2	9	<+>							
BRUC_3	1	<+>							
BRUC_4	4	<+>							
BRUC_5	5	<+>							
BRUC_6	1	<+>							
BRUC_7	1	<+>							
BRUC_8	2	<+>							
BRUC_9	3	<+>							
BRUC_10	7	<+>							
BRUC_11	1	<+>							
BRUC_12	6	<+>	~						
Press Insert to add char	acter								

Figure 6: Character card.

4 Subschemes

During installation of the plugin, the **wgMLST** character experiment is created and synchronized with the *Brucella* spp. specific locus scheme. All detected loci and subschemes are added to this

experiment.

- 1. In the *Main* window double-click the character experiment type **wgMLST** in the *Experiment types* panel to call the *Character type* window.
- 2. Click on the drop-down bar in the toolbar (see Figure 7 for an example).

The views that have been defined at the curator level are synchronized upon installation and are listed. In the *Brucella* spp. database following views are defined by the curator (see Figure 7): the default view **All loci**, the genus-wide **TRUNK** loci view, 11 species specific loci views, the MLST 9 loci view (**MLST PubMLST 9 loci**) and the MLST 21 loci view (**MLST PubMLST 21 loci**).

Character type WgMLST						
File Settings Characters Fields Mapping Window Help Image: Settings Characters Image: Settings Image: Setings Image	🖆 Character type 'wgMLST'				-	\times
If	File Settings Characters Fields	Mapping Window	/ Help			
Characters + * * * * - Characters - BRUC_1 - BRUC_2 - BRUC_3 - BRUC_3 - MI Characters> - Mino - Selected Characters> - Mino - All Characters> - BRUC_3 - Mino - Mino <t< td=""><td>局此症症</td><td></td><td></td><td></td><td></td><td></td></t<>	局此症症					
Characters Characters Couns Tay Gene BRUC_1 Characters> Colong VBBruAbo93 saX1.joD BRUC_2 Characters> Colong VBBruAbo93 saX1.joD BRUC_3 Characters> Colong VBBruAbo93 saX1.joD BRUC_4 Characters> Colong VBBruAbo93 saX1.joD BRUC_5 Colong VBBruAbo93 colong VBBruAbo93 saX1.joD BRUC_6 Colong VBBruAbo93 saX1.joD colong VBBruAbo93 saX1.joD BRUC_5 Colong VBBruAbo93 saX1.joD colong VBBruAbo93 saX1.joD BRUC_6 Colong VBBruAbo93 saX1.joD colong VBBruAbo93 saX1.joD BRUC_10 Colong VBBruAbo93 suB11 colong VBBruAbo93 suB11 BRUC_11 Vulpis ceti suis solong VBBruAbo93 purL1.jou BRUC_13 BRUC_14 Solong VBBruAbo93 purL ceti suis abortus colong VBBruAbo93 purL c						
Image: Second sector in the second sector is the second second sector is the second second second second	Characters					
Character Enabled Min. < <all characters=""> BRUC_1 - <</all>	+ 🛛 🗟 🛛 🗠 🗠	↑ ↓ ↓\	<all characters=""> U</all>			
BRUC_1 Selected Characters> 0000_VBBruAb093 salX1,bD ^ BRUC_3 0000_VBBruAb093 fmaE2,dmaE-2 BRUC_4 0000_VBBruAb093 fmaE2,dmaE-2 BRUC_5 All loci 0000_VBBruAb093 0000_VBBruA	Character Enabled	d Min.	✓ <all characters=""></all>	LocusTag	Gene	•
BRUC_2 • • BRUC_3 • • BRUC_3 • • BRUC_4 • • BRUC_5 • • BRUC_7 • • BRUC_8 • • BRUC_7 • • BRUC_9 • • BRUC_10 • • BRUC_11 • • BRUC_12 • • BRUC_13 • • BRUC_14 • • BRUC_14 • • BRUC_15 • • BRUC_14 • • BRUC_15 • • BRUC_14 • • BRUC_15 • • BRUC_16 • •	BRUC_1	o	<selected characters=""></selected>	0000_VBIBruAbo93	salX1,IoID	^
BRUC_3 0000_VBIBruAb093 pu000_VBIBruAb093 pu111 pu000_VBIBruAb093 pu12_purQ pu000_VBIBruAb093 pu12_purQ pu000_VBIBruAb093 pu12_purQ pu000_VBIBruAb093 pu12_purQ pu000_VBIBruAb093 purC purC pu11 pu000_VBIBruAb093 purC purC purC pu11 pu000_VBIBruAb093 purC pu11 pu000_VBIBruAb093 pu11 pu000_VBIBruAb093 purC <td>BRUC_2</td> <td>o</td> <td></td> <td>0000_VBIBruAbo93</td> <td>dnaE2,dnaE-2</td> <td></td>	BRUC_2	o		0000_VBIBruAbo93	dnaE2,dnaE-2	
BRUC_4 Image: Constraints Image: Constraints <td>BRUC_3</td> <td>o</td> <td><manage defined="" user="" views=""></manage></td> <td>0000_VBIBruAbo93</td> <td></td> <td></td>	BRUC_3	o	<manage defined="" user="" views=""></manage>	0000_VBIBruAbo93		
BRUC_5 All foci BRUC_5 All foci BRUC_6 TRUNK BRUC_7 inopinata BRUC_8 inopinata BRUC_9 inopinata BRUC_10 microti BRUC_11 upinipedialis Vulpis vulpis BRUC_13 ceti BRUC_14 upins BRUC_15 upins Comparison settings ovis MLST PubMLST 1 loci MLST PubMLST 2 loci Comparison settings Comparison settings Compari	BRUC_4	o		0000_VBlBruAbo93		
BRUC_6 IRUNK 0000_VBBruAb033 mesl1 BRUC_7 inopinata 0000_VBBruAb033 suhB1 BRUC_8 inopinata 0000_VBBruAb033 bull BRUC_9 inicroti 0000_VBBruAb033 bull BRUC_10 inicroti 0000_VBBruAb033 bull BRUC_11 inicroti 0000_VBBruAb033 bull BRUC_12 isis outpits 0000_VBBruAb033 purl1 0000_VBBruAb033 purl1 0000_VBBruAb033 purl1 0000_VBBruAb033 purl1 0000_VBBruAb033 purl1 0000_VBBruAb033 purl1 0000_VBBruAb033 purl1 0000_VBBruAb033 purl2.purl2 0000_VBBruAb033 purl2.purl2 0000_VBBruAb033 purl2.purl2 0000_VBBruAb033 purl2.purl2 0000_VBBruAb033 purl2.purl2 0000_VBBruAb033 purl2.purl2 0000_VBBruAb033 purl2.purl2 0000_VBBruAb033 purl2.purl2 0000_VBBruAb033 purl2.purl2 ovis MLST PubMLST 9 loci MLST PubMLST 21 loci Comparison settings C	BRUC_5	o	All loci	0000_VBlBruAbo93	rpsD	
BRUC_7 Implication BRUC_8 Implication BRUC_9 Implication BRUC_10 Implication BRUC_11 Implication BRUC_12 Implication BRUC_13 Implication BRUC_14 Implication BRUC_15 Implication Implication Implication Implicatio	BRUC_6	0	TRUNK	0000_VBlBruAbo93	mesJ1	
BRUC_8 Implified BRUC_9 microti BRUC_10 microti BRUC_11 microti BRUC_12 microti BRUC_13 microti BRUC_14 microti BRUC 15 microti Characters Mapping melitensis ovis multistrinucial values, open data set (5325 char Comparison settings Mist PubMLST 9 loci MLST PubMLST 10 coi Mist PubMLST 21 loci	BRUC_7	0	inopinata	0000_VBIBruAbo93	suhB1	
BRUC_9 microti BRUC_10 microti BRUC_11 microti ulpis 0000_VBIBruAbe93 bruL1 0000_VBIBruAbe93 bruL11 ulpis ceti 0000_VBIBruAbe93 bruL212 ceti bruL14 ceti suis abortus neotomae canis melitensis ovis MLST settings wgMLST: numerical values, open data set (5325 char Similarity coefficient Categorical (values) comparison settings Character type wgMLST	BRUC_8	0	inopinata	0000_VBIBruAbo93		
BRUC_10 pinnipedialis BRUC_11 pinnipedialis Ulpis ulpis Ulpis<	BRUC_9	0	microti	0000_VBIBruAbo93	bolA1	
BRUC_11 Vulpis BRUC_12 Ceti BRUC_13 Ceti BRUC_14 Ceti BRUC_15 Suis Characters Mapping Mapping NullsT Comparison settings ovis Similarity coefficient ovis Categorical (values) Comparison settings Comparison settings Comparison settings Comparison settings Comparison Similarity coefficient Categorical (values) Comparison settings Characters in view=5325	BRUC_10	0	pinnipedialis	0000_VBIBruAbo93	purL1	
BRUC_12 Vulpis BRUC_13 Ceti BRUC_15 000_VBIBruAbo93 BRUC_15 000_VBIBruAbo93 Doug_VBIBruAbo93 pur2 Doug_VBIBruAbo93 pur2 Ocharacters Mapping Mapping neotomae Comparison Similarity coefficient Categorical (values) ovis Comparison settings Attachments	BRUC_11	0		0000_VBIBruAbo93		
BRUC_13 ceti BRUC_14 suis BRUC_15 suis BRUC 15 od Characters Mapping Comparison settings ceti wgMLST settings melitensis wgMLST: numerical values, open data set (5325 char Similarity coefficient Categorical (values) Comparison settings	BRUC_12	0	vuipis	0000_VBIBruAbo93	gst3	
BRUC_14 0 BRUC 15 suis abortus abortus neotomae canis wgMLST settings ovis WgMLST: numerical values, open data set (5325 char MLST PubMLST 9 loci Similarity coefficient Categorical (values) Comparison settings Crosslinks Attachments Attachments	BRUC_13	0	ceti	0000_VBIBruAbo93	purL2,purQ	
BRUC 15 0 Characters Mapping Comparison settings neotomae canis melitensis wgMLST settings ovis MLST numerical values, open data set (5325 char MLST PubMLST 9 loci Comparison Similarity coefficient Categorical (values) Comparison settings Comparison settings Crosslinks Attachments Attachments	BRUC_14	0	suis	0000_VBIBruAbo93	purS	
Characters Mapping Comparison settings neotomae wgMLST settings canis wgMLST: numerical values, open data set (5325 char melitensis Orus Orus Similarity coefficient MLST PubMLST 21 loci Categorical (values) MLST PubMLST 21 loci Comparison settings Crosslinks Attachments Characters in view=5325	BRUC 15	0	shadus	0000 VBIBruAbo93	purC	~
Comparison settings neotomae wgMLST settings canis wgMLST: numerical values, open data set (5325 char melitensis Comparison Similarity coefficient Categorical (values) MLST PubMLST 21 loci Comparison settings Crosslinks Attachments Characters in view=5325	Characters Mapping		abortus			
comparison canis wgMLST: numerical values, open data set (5325 char ovis Comparison Similarity coefficient Categorical (values) MLST PubMLST 21 loci Comparison settings Crosslinks Attachments Character type wgMLST Character type wgMLST Characters in view=5325	Comparison sottings		neotomae			
wgMLST settings wgMLST: numerical values, open data set (5325 char Comparison Similarity coefficient Categorical (values) Comparison settings Crosslinks Attachments Character type wgMLST Characters in view=5325 Characters in view=5325	comparison settings		canis			
wgMLST settings ovis wgMLST: numerical values, open data set (5325 char ovis Comparison Similarity coefficient Categorical (values) MLST PubMLST 21 loci Comparison settings Crosslinks Attachments Character type wgMLST Character type wgMLST Characters in view=5325			melitensis			~
wgMLST: numerical values, open data set (5325 char Comparison Similarity coefficient Categorical (values) Comparison settings Crosslinks Attachments Character type wgMLST Characters in view=5325 Characters in view=5325	wgMLST settings					
Comparison MLST PubMLST 9 loci MLST PubMLST 21 loci MLST PubMLST 21 loci MLST PubMLST 21 loci MLST PubMLST 21 loci Comparison settings Crosslinks Attachments Character type wgMLST Characters in view=5325 .	wgMLST: numerical values, open	ı data set (5325 char	ovis			
Comparison settings Crosslinks Attachments Character type wgMLST Characters in view=5325	Comparison		MLST PubMLST 9 loci			
Comparison settings Crosslinks Attachments Character type wgMLST Characters in view=5325	Comparison		MLST PubMLST 21 loci			
Categorical (values) Comparison settings Crosslinks Attachments Character type wgMLST Characters in view=5325 .	- Similarity coe	efficient				
Comparison settings Crosslinks Attachments Character type wgMLST Characters in view=5325	Categorical (values	3)				~
Character type wgMLST Characters in view=5325	Comparison settings Crosslinks Att	tachments				
	Character type wgMLST Characters i	in view=5325				

Figure 7: Character views defined by the curator.

3. Select another view from the list to update the set of loci in the *Characters* panel.

The number of loci in the selected view is displayed in the status bar at the bottom of the window.

4. To view all characters again, select <**All loci**> again from the drop-down list.

Besides these curator views, the user can create as many additional local character views as needed and use them as subscheme e.g. for clustering or when inspecting the allele calls for a subset of loci via *Characters* > *Character Views* > *Manage user defined views...* (<a character views).

5. Close the *Character type* window.

5 Sequence type assignment

Sequence types can be assigned for selected entries, based on a specific wgMLST subscheme. Note that only some of the curator-defined subschemes have associated sequence types.

- 1. Select the entries for which you would like to assign sequence types. For this example, select all entries with *Edit* > *Select all* (Ctrl+A).
- 2. Select WGS tools > Assign wgMLST sequence types....

This opens the *Assign sequence types* dialog box, where all available typing schemes can be checked to be included in the assignment of the sequence types (see Figure 8).

Assign sequence types	?	×
Select one or more typing schem	as:	
MLST PubMLST 21 loci		
MLST PubMLST 9 loci		
ОК	Car	ncel

Figure 8: Sequence type assignment.

Only check the schema *MLST PubMLST 9 loci* to assign sequence types based on the 9 loci used for traditional MLST analysis and press < *OK* >.

The sequence type results are saved to an entry information field (one information field for each typing scheme). In our example, a sequence type number is added in the field **MLST PubMLST 9 loci ST** for all entries (see Figure 9).

🛃 WGS B	Irucella demobase - BioNun	merics													-	• ×	-
File Edit	Database Analysis Scri	ipts WGS tools Win	dow	Help													
同→	ല രില	h D Q 20		,													
Experim			Dat	abase entries										Comparisons			1
A	+ 🖻 🛛 🖻		t,	1 + 🖻 🛛 🖻		<all entries=""></all>	1	<u>ت</u>								Z sAl Comp	Í.
609			선			Of the local data	1	0		0.1-1-	MI OT DU ANN OT O IN						
	• Nallie	туре		Rey	opecies	strain	Run accession number	sample name	Accession	origin	MEST PUDMEST STO	4		Name	mounieu u		
	1 wgs	Sequence read set		BRUCELLA_DEMODU1	meikensis	16M			AE008917		publics 17		· · · ·				
	2 WgML51	Character types		BRUCELLA_DEMODU2	suis	1330			CP002997		public5114						
	4 qualty	Character types		BRUCELLA_DEMODOS	abortus	C10			CP000703		publicSTTS						
	6 was TrimmedState	Sequence read set		BRUCELLA DEMODOS	microti	CCM 4915			CP000007		publicST63						
	6 wold ST Callfornee	Character types		BRUCELLA DEMONDS	cati	TE10759.12			CP008897		publicST28						
	7 wosl ong	Sequence read set		BRUCELLA DEMODOZ	canie	SVA13			CP007630		publicST20						
	1 highling	Codecilion Long Col		BRUCELLA DEMODOR	nentomae	5K33			IMSC01000001		publicST22						
		, v		BRUCELLA DEMOND	abortus	870			CP007709		publicST4				_		
	`			BRUCELLA DEMOQ10	inopinata	802			NZ ADEA00		N/A						
Entry fields	s Database design		V	BRUCELLA DEMO011	vulpis	F60			LN997863		N/A			Identification projects	acision networks		
	* • A · A			BRUCELLA DEMO012	canis	ATCC 23365			CP000872		publicST20						
+ 6		⊻ T ↓		BRUCELLA_DEMO013	melitensis	63/9			CP007789		publicST8			29 + 10 0	9 6, 1 6	⊻	1
		Field type 🛛 🔻	V	BRUCELLA_DEMO014	melitensis	ATCC 23457			CP001488		publicST8			Name		ate 🔍 🔫	
ARC	Species	Fixed A	~	BRUCELLA_DEMO015	abortus	86/8/59			CP007765		publicST1						
- AK	Strain	Fixed	V	BRUCELLA_DEMO016	inopinata	B01	SRR4038982				publicST52						
- AK	Run accession number	Fixed	Z	BRUCELLA_DEMO017	pinnipedialis	B2-94	SRR4038996				publicST25		• •				
- AK	Sample Name	Fixed	V	SRR2058928	suis	B00-0729	SRR2058928	Swine		TX	publicST14		• •				
ARC .	Accession	Fixed	Z	SRR2058929	suis	B00-0771	SRR2058929	Swine		AL	publicST14		• •				
- AK	Origin	Fixed	~	SRR2058942	suis	B11-0001	SRR2058942	Cow		TX	publicST14		• •				
- AK	MLST PubMLST 9 loci ST	Flexible	~	SRR2058946	suis	B11-0193	SRR2058946	Cow		TX	publicST14		• •				
ARC	MLST PubMLST 21 lbci ST	Flexible Y	~	SRR2058961	suis	B11-0566	SRR2058961	Cow		TX	publicST14		• •				
	<	>	≤	SRR2058968	suis	B13-0195	SRR2058968	Feralpig		TX	publicST14		• •				
Eingernrint	Har Dowar serambiar	Anostations		SRR2058969	suis	B13-0198	SRR2058969	Feralpig		TX	publicST14		•••	Alignmente BLAST proj		comparizone	
r ingerprint	rites retret assertates in			SRR2058971	suis	B13-0219	SRR2058971	Human		OR	publicST14		•••	Augumenta (Dexor proj		computitiona	
巴	+ [2] ⊗ [4]		× <u>×</u>	SRR2058977	suis	B94-0444	SRR2058977	Swine		TX	publicST14	•••	•••	+ 12 ⊗ €		Z. <all aligne<="" td=""><td>۴.</td></all>	۴.
Fi	ile name Expe	eriment type 🔍 🔻		SRR2058979	suis	B95-0322	SRR2058979	Swine		AR	publicST14		•••	Name	Modified da	ate 💌 🔻	
			¥.	SRR2058982	suis	813-0257	SRR2058982	Swine		TX	publicST14						
			Ľ	SRR2058984	SUIS	813-0234	SRR2058984	Human		Tonga	publicS114						
			×	SRR2058985	SUIS	813-0235	SRR2058985	Human		Tonga	publicS114	111					
				SRR2058986	suis	B13-0236	SRR2058966	Human		Tonga	publicS114	111					
				CDD2050000	ouis	B13-0237	CDD2050000/	Human		Tonga	public T14	111					
				CDD2068080	euis	813 0220	CDD2050000	Numan		Tonga	publicST14	1.1					
				SRR3096458	Unknown	0.00000	SRR3096458			. onga	publicST1						
													~				
<		>		<							>	< C	>				
Databas	se: WGS Brucella demobase (_DefaultUser_) Entries	s: Loade	d=35, View=35, Selected=	35 7 experiments C	\Users\Public\Docu	ments\BioNumerics\Data 80\V	VGS Brucella den	obase This is	a time limited ;	ackage valid until 2020-12-30						į,

Figure 9: The *Main* window after sequence type assignment.

4. Press <*F*4> to clear the entry selection.

6 Follow-up analysis

A cluster analysis on the **wgMLST** character experiment (or a subscheme thereof) is created in the *Comparison* window or the *Advanced cluster analysis* window. The steps to create a new comparison and to perform cluster analysis on wgMLST data are explained in the next sections. The trunk-and-branch structure of the *Brucella* spp. schema allows flexibility in how to analyze different types of samples. Three possible scenarios are outlined here to illustrate this:

6.1 Compare multiple Brucella species

When you have multiple species of the genus *Brucella* in the same database and you want to compare them to each other, you will want to use the genus-wide subschema **TRUNK**.

- In the Database entries panel of the Main window, select the entries with Key BRUCELLA_ DEMO001 to BRUCELLA_ DEMO017. Selected entries are marked by a checked ballot box (<).
- 2. Highlight the *Comparisons* panel in the *Main* window and select *Edit* > *Create new object...* (+) to create a new comparison for the 17 selected entries.
- 3. Select the **wgMLST** character experiment in the *Experiments* panel of the *Comparison* window.

Since we are dealing with multiple species, we will use the genus-wide **TRUNK** subschema.

- 4. Make sure the **TRUNK** aspect is selected for the **wgMLST** experiment (see Figure 10).
- 5. In the *Experiments* panel click on the eye icon () that proceeds **wgMLST** to display the values of the selected aspect.

Exp	erim	ients			
</th <th>All Ex</th> <th>periment types></th> <th>1 ប</th> <th></th> <th></th>	All Ex	periment types>	1 ប		
		Name	Aspect		•
	₹	wgs	<default></default>		^
		wgMLST	TRUNK	-	
	AC 61	denovo	<default></default>		
	٠.	quality	<all characters=""></all>		
	₹	wgs_TrimmedStats	<default></default>		
	Π.	wgMLST_CallTypes	<all characters=""></all>		
	₹	wgsLong	<default></default>		
					¥

Figure 10: The TRUNK subschema.

- 6. Select *Clustering* > *Calculate* > *Cluster analysis (similarity matrix)...* and choose the *Categorical (values)* coefficient from the list.
- 7. Press <*Next*>, choose *Complete Linkage* in the last step and press <*Finish*>.

The resulting dendrogram is displayed in the *Dendrogram* panel and the analysis is stored in the *Analyses* panel. The subscheme that was used is indicated between brackets: e.g. **wgMLST** (**TRUNK**).

- 8. The settings used to calculate the dendrogram that is displayed in the *Dendrogram* panel can be called with *Clustering* > *Show information* (1).
- 9. To view the similarity values on the nodes, select *Clustering* > *Dendrogram display settings...* (11), and tick the option *Show node information*. Press <*OK*>.

- Right-click on the column header of Species in the *Information fields* panel and select *Create* groups from database field. In the *Group creation preferences* dialog box, leave the settings at their defaults and press < OK >.
- 11. Select *Clustering* > *Dendrogram display settings...* (11) again, and tick the option *Show group colors*. Press <*OK*>.

The group colors are now displayed on the dendrogram. The *Comparison* window should now look like Figure 11.



Figure 11: The *Comparison* window.

Another analysis tool that can be applied on wgMLST data is the calculation of a Minimum spanning tree (MST). A minimum spanning tree is calculated in the *Advanced cluster analysis* window which is launched from the *Comparison* window.

12. Select *Clustering* > *Calculate* > *Advanced cluster analysis...* in the *Comparison* window to launch the *Create network* wizard.

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.

13. Specify an analysis name, make sure the **TRUNK** subscheme is selected, select *MST for categorical data*, and press <*Next*> (see Figure 12).



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

A MST is now computed in the *Advanced cluster analysis* window. 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. The analysis is also added to the *Analyses* panel in the *Comparison* window.

- 14. Press 🖽 or choose *Display* > *Display settings* to open the *Display settings* dialog box.
- 15. In the Branch labels and sizes panel, you can specify that you want to see the distances between the nodes (i.e. the number of allele differences): check Show branch labels and set Number of digits to "0".

the analysis template,	or create a custom analysis.	•
Analysis name		
Name of the cluster analy	sis: MST TRUNK	
Experiment		
Choose an experiment:	wgMLST (TRUNK) V Settings	
Analysis template		
No template		
🖻 🦢 Predefined temp	plates	
MST for cat	egorical data arsimony tree	
O Maximum lik	elihood tree	
 Topscore U Majority UD(PGMA	
······· ······························	5//	
Standard MST with single	e and double locus variance priority rules.	
	ns for new analysis	

Figure 12: Settings for the calculation of the MST.

- 16. In the *Node labels and sizes panel*, check *Show node labels* and choose *MLST PubMLST 9 loci ST* as *Label*.
- 17. Click <*OK*> to close the *Display settings* dialog box. The MST is now displayed with branch and node labels (see Figure 13).
- 18. Close the Advanced cluster analysis window.
- 19. Save the comparison with *File* > *Save as...*. Specify a name (e.g. **Genus-wide comparison**) and close the comparison with *File* > *Exit*.

6.2 Compare single Brucella species

If you have a set of samples for which you determined the species via different methods, you can use the corresponding species-specific subschema to analyze them.

- 20. In the *Main* window, select all *Brucella suis* entries: press <*F*4> to unselect all entries, select *Edit* > *Find object in list...* (R, Ctrl+Shift+F), search for "suis" and select <*Select all*>.
- 21. Highlight the *Comparisons* panel in the *Main* window and select *Edit* > *Create new object...* (+) to create a new comparison for the selected entries.
- 22. Select the wgMLST character experiment in the *Experiments* panel of the *Comparison* window.

Since we are dealing with a single species, we can use the species-specific suis subschema.

- 23. Make sure the suis aspect is selected for the wgMLST experiment (see Figure 14).
- 24. In the *Experiments* panel click on the eye icon () that proceeds **wgMLST** to display the values of the selected aspect.
- 25. Select *Clustering* > *Calculate* > *Cluster analysis (similarity matrix)....*



Figure 13: The minimum spanning tree based on the TRUNK schema.

Experiments											
</th <th>All Ex</th> <th>periment types></th> <th>ט </th> <th colspan="6">ט </th>	All Ex	periment types>	ט	ט							
		Name	Aspect		•						
	₹	wgs	<default></default>		^						
	Η.	wgMLST	suis	-							
	AC 6 T	denovo	<default></default>								
\odot	Π.	quality	<all characters=""></all>								
	₹	wgs_TrimmedStats	<default></default>								
		wgMLST_CallTypes	<all characters=""></all>								
	₹	wgsLong	<default></default>								
					~						

Figure 14: The suis subschema.

26. Since we want to compare closely related isolates choose the *Categorical (differences)* coefficient from the list. Specify a *Scaling factor* of 1.

The *Categorical (differences)* coefficient treats each different value as a different state, and results in a distance matrix. With the *Scaling factor* one can deal with the hard-coded maximum of 200 that can be calculated for a distance value. Values that make sense are 1, 10 and 100, allowing the correct visualization of maximally 200, 2000 and 20000 different character values, respectively, in a cluster analysis.

27. Press < *Next*>, choose *Complete Linkage* in the last step and press < *Finish*>.

The resulting dendrogram is displayed in the Dendrogram panel and the analysis is stored in the

Analyses panel. The subscheme that was used is indicated between brackets: e.g. **wgMLST** (suis).

28. To view the number of allele differences on the branches, select *Clustering* > *Dendrogram display settings...* (11), and tick the option *Show node information*. Press <*OK*>.

To trace back the number of different loci from the branches or distance matrix, the displayed values needs to be multiplied with the *Scaling factor* used (1 in this example).

- 29. Right-click on the column header of **Sample name** in the *Information fields* panel and select *Create groups from database field*. In the *Group creation preferences* dialog box, leave the settings at their defaults and press <*OK*>.
- 30. Select *Clustering* > *Dendrogram display settings...* (11) again, and tick the option *Show group colors*. Press <*OK*>.

The group colors are now displayed on the dendrogram, emphasizing the clustering of the samples of human origin (see *Comparison* window).



Figure 15: The Comparison window: clustering of species specific entries.

- 31. The polymorphic loci for the set of samples in the selected scheme can be displayed with *Characters* > *Filter characters* > *Select polymorphic characters...*
- 32. The information displayed in the *Experiment data* panel can be exported with *Characters* > *Export character table*. The character table will open as a export.csv file in MS Excel.
- 33. To export the cluster analysis as it appears in the *Comparison* window select *File* > *Print preview...* (, Ctrl+P). The *Comparison print preview* window appears.
- 34. Select *Clustering* > *Calculate* > *Advanced cluster analysis...* in the *Comparison* window to launch the *Create network* wizard.
- 35. Specify an analysis name, make sure the **suis** subscheme is selected, select *MST for categor-ical data*, and press <*Next*>.
- 36. Press 🖽 or choose *Display* > *Display settings* to open the *Display settings* dialog box.
- 37. In the *Branch labels and sizes panel* check **Show branch labels** and set **Number of digits** to "0".
- Click < OK > to close the Display settings dialog box. The MST is now displayed with branch labels (see Figure 16).



Figure 16: The minimum spanning tree based on the suis subschema.

- 39. Close the Advanced cluster analysis window.
- 40. Save the comparison with *File* > *Save as...*. Specify a name (e.g. **Suis samples comparison**) and close the comparison with *File* > *Exit*.

6.3 Identify unknown Brucella species

When you have a sample, from which you want to determine the type of species, you can analyze it with a species-specific subschema in the following way:

- 41. In the Database entries panel of the Main window press < OK > to clear any previous selection. Select the entries with Key BRUCELLA_ DEMO001 to BRUCELLA_ DEMO017 and include the Unknown entry.
- 42. Highlight the *Comparisons* panel in the *Main* window and select *Edit* > *Create new object...* (+) to create a new comparison for the 18 selected entries.
- 43. Select the wgMLST character experiment in the *Experiments* panel of the *Comparison* window.
- 44. To determine to which species this unknown sample relates the most, choose the **TRUNK** aspect for the **wgMLST** experiment (see Figure 10).
- 45. In the *Experiments* panel click on the eye icon () that proceeds **wgMLST** to display the values of the selected aspect.
- 46. Select *Clustering* > *Calculate* > *Cluster analysis (similarity matrix)...* and choose the *Categorical (values)* coefficient from the list.

47. Press <*Next*>, choose *Complete Linkage* in the last step and press <*Finish*>.

The resulting dendrogram is displayed in the Dendrogram panel.

The unknown sample falls within the *Brucella abortus* cluster (see Figure 17).



Figure 17: The Brucella abortus cluster.

This allows us to use the species-specific subschema.

- 48. Unselect all entries with *<F*4*>*. Select the three *B. abortus* samples and the *Unknown* sample.
- 49. Highlight the *Comparisons* panel in the *Main* window and select *Edit* > *Create new object...* (+) to create a new comparison for the 4 selected entries.
- 50. Select the wgMLST character experiment in the *Experiments* panel of the *Comparison* window.
- 51. Make sure the **abortus** aspect is selected for the **wgMLST** experiment (see Figure 18).

Experiments						
</th <th>All Ex</th> <th>periment types></th> <th>ט </th> <th></th>	All Ex	periment types>	ט			
		Name	Aspect 🗨	,		
	₹	wgs	<default></default>	•		
0	85	wgMLST	abortus			
	A C G T	denovo	<default></default>			
\odot		quality	<all characters=""></all>			
	₹	wgs_TrimmedStats	<default></default>			
		wgMLST_CallTypes	<all characters=""></all>			
	₹	wgsLong	<default></default>			
<			>			

Figure 18: The abortus aspect.

- 52. Select *Clustering* > *Calculate* > *Cluster analysis (similarity matrix)...*, choose the *Categorical (differences)* coefficient from the list. Specify a *Scaling factor* of 10.
- 53. Press <*Next*>, choose *Complete Linkage* in the last step and press <*Finish*>.
- 54. To view the number of allele differences on the branches, select *Clustering* > *Dendrogram display settings...* (11), and tick the option *Show node information*. Press < *OK* >.

To trace back the number of different loci from the branches or distance matrix, the displayed values needs to be multiplied with the *Scaling factor* used (10 in this example).

Comparison							
File Edit Layout Groups Clustering Statistics Fingerprints Characters Sequence TrendData ReadSets Spectra Composite Window Help							
B Gel‰l⊳ El ≭X £ C‡ {Pa G₂ wymust OO dÉ l⊾, 101 l Å,							
Experiments (9,							
<all experiment="" types=""> 10 (2) Dendrogram</all>	Experiment data	Information fields	Similarities				
Name Aspect 🗸 🖬 나타 면 단다 탄	C III III III III III III III III III I	₿, ↓ \$↑↓₩ ↑↓	i † 120				
⊘ ≩ wgs <defaut> ^</defaut>	wgMLST						
WgMLST abortus							
O AC denovo <default></default>							
quality <all characters=""></all>							
wgMLST_CallTypes <all characters=""></all>							
⊘ ≱ wgsLong <defaut></defaut>	<mark>8</mark> 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9		WgMLST 100 80 60 40				
< wgMLST (abortus) 60 40	20 0 8 8 8 8 8 8 8 8 8 8 8 8 8 8	Key Species Strain 🗸	· · · · · · · · · · · · · · · · · · ·				
		BRUCELLA DEMO015 abortus 86/8/59	• • • • •				
Analyses Local composite datasets	12.9 1 20 1 1 1 1 1 1 1 1 1 1 1 1 1	SRR3096458 Unknown :					
P ^{70,8}	1 20 1 9 1 1 1 1 1 1 1 1 1	BRUCELLA_DEMO004 abortus S19					
Name	1 21 1 1 1 1 1 1 3 1 1 1 1	BRUCELLA_DEMO009 abortus 870					
YE wgMLST (abortus)							
×							
Groups Jub bg							
	< >	< >	v .				
4 entries in comparison 4 entries selected in database							

Figure 19: Clustering based on the abortus aspect.

55. Close the *Comparison* window.