BIONUMERICS Tutorial:

CFSAN SNP pipeline

1 Introduction

The ***CFSAN SNP pipeline*** is a SNP pipeline created by the FDA Center for Food Safety and Applied Nutrition (CFSAN) [1] which can be launched on sequence read set experiments in BIONUMERICS.

A ***CFSAN SNP pipeline*** analysis in BIONUMERICS consists of the following steps:

1. Choose an appropriate reference sequence for your samples of interest.
2. Create a comparison of your samples of interest.
3. Launch a wgSNP analysis with the ***CFSAN SNP pipeline*** on the Calculation Engine.
4. Import the results in BIONUMERICS and analyse the wgSNP clustering in the *Comparison* window.

The CFSAN SNP pipeline was developed for closely related organisms and is therefore not suited for the analysis of organisms for which no single appropriate reference sequence is available. For more details on the CFSAN SNP pipeline and the performed mapping, variant calling, SNP filtering and clustering steps, we refer to Davis *et al.* 2015 [1].

2 Preparing the database

The CFSAN SNP pipeline can only be performed in BIONUMERICS after installation of the *WGS tools plugin* in the BIONUMERICS database (***File*** > ***Install / remove plugins...*** (🔌)).

As the CFSAN SNP pipeline is only available on the Cloud Calculation engine make sure to select the options ***Use default Cloud Calculation Engine*** and ***Enable running jobs on Cloud Calculation Engine*** during installation of the *WGS tools plugin*. The Calculation engine option requires credits for running jobs on the Applied Maths cloud calculation engine. Credits are linked to credentials that you need to enter when installing the *WGS tools plugin*.

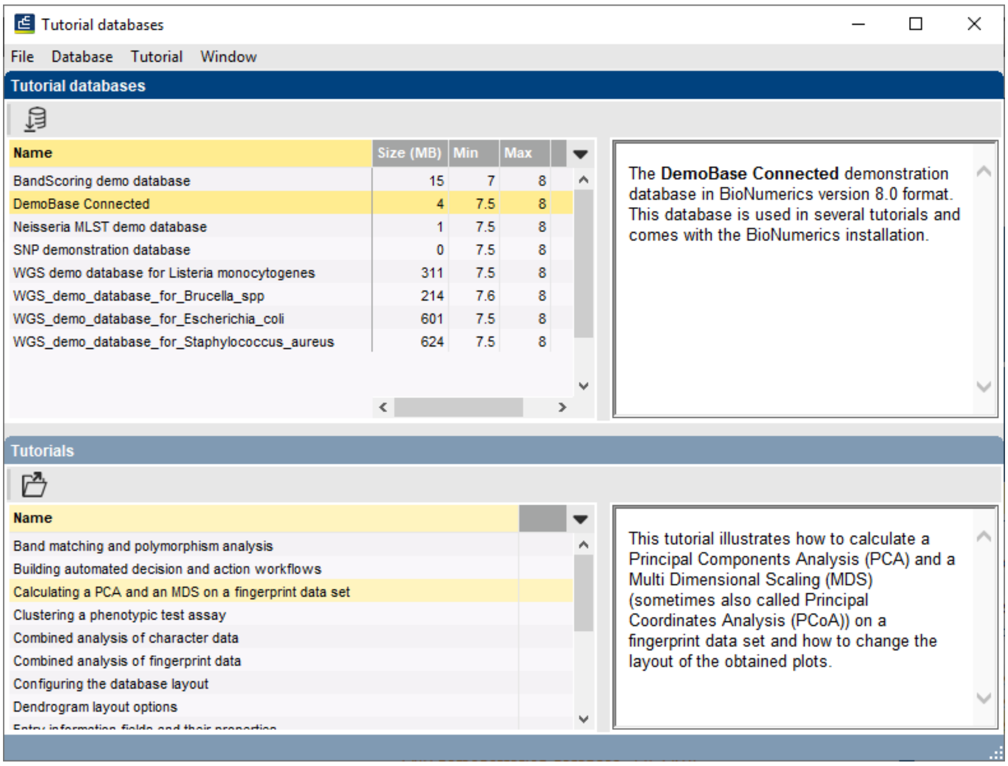
In this tutorial the **WGS demo database for *Salmonella*** will be used in which the *WGS tools plugin* is already installed. No credits are assigned to the demo project so no CFSAN SNP pipeline jobs can be launched on the external calculation engine. Please contact Applied Maths to obtain more information.

The **WGS demo database for *Salmonella*** 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).



Name	Size (MB)	Min	Max
BandScoring demo database	15	7	8
DemoBase Connected	4	7.5	8
Neisseria MLST demo database	1	7.5	8
SNP demonstration database	0	7.5	8
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

The **DemoBase Connected** demonstration database in BioNumerics version 8.0 format. This database is used in several tutorials and comes with the BioNumerics installation.

Name
Band matching and polymorphism analysis
Building automated decision and action workflows
Calculating a PCA and an MDS on a fingerprint data set
Clustering a phenotypic test assay
Combined analysis of character data
Combined analysis of fingerprint data
Configuring the database layout
Dendrogram layout options
Entry information fields and their properties

This tutorial illustrates how to calculate a Principal Components Analysis (PCA) and a Multi Dimensional Scaling (MDS) (sometimes also called Principal Coordinates Analysis (PCoA)) on a fingerprint data set and how to change the layout of the obtained plots.

Figure 1: The *Tutorial databases* window, used to download the demonstration database.

2. Select **WGS_demo_database_for_Salmonella_enterica** 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_Salmonella_enterica** appears in the *BIONUMERICS Startup* window.

5. Double-click the **WGS_demo_database_for_Salmonella_enterica** 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 demo database for *Salmonella enterica* is also available on our website. This backup can be restored to a functional database in BIONUMERICS.

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

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

7. In the *BIONUMERICS Startup* window, press the button. From the menu that appears, select ***Restore database...***.

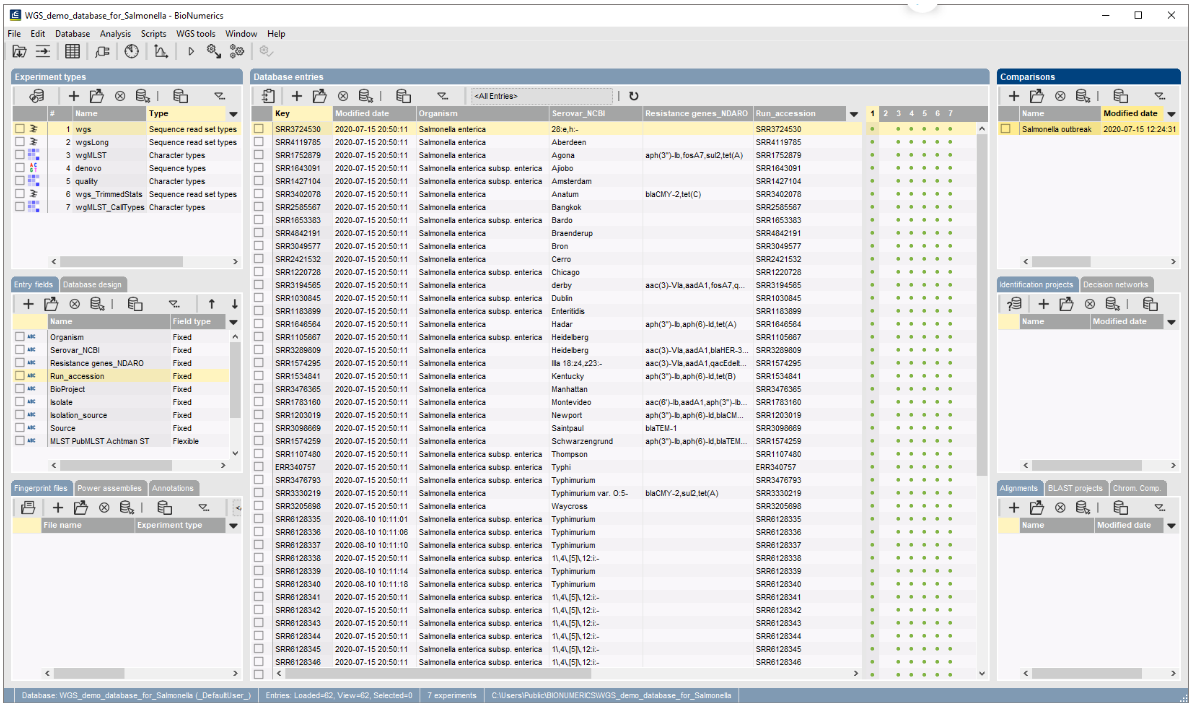
8. Browse for the downloaded file and select ***Create copy***. Note that, if ***Overwrite*** is selected, an existing database will be overwritten.

9. Specify a new name for this demonstration database, e.g. "WGS_Salmonella_demobase".

10. Click <***OK***> to start restoring the database from the backup file.

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

The *Main* window is displayed (see Figure 2).



#	Name	Type
1	wgs	Sequence read set types
2	wgsLong	Sequence read set types
3	wgMLST	Character types
4	denovo	Sequence types
5	quality	Character types
6	wgs_TrimmedStats	Sequence read set types
7	wgMLST_CallTypes	Character types

Name	Field type
Organism	Fixed
Serovar_NCBI	Fixed
Resistance genes_NDARO	Fixed
Run_accession	Fixed
BioProject	Fixed
Isolate	Fixed
Isolation_source	Fixed
Source	Fixed
MLST PubMLST Achtman ST	Flexible

File name	Experiment type

Key	Modified date	Organism	Serovar_NCBI	Resistance genes_NDARO	Run_accession
SRR3724530	2020-07-15 20:50:11	Salmonella enterica	28:e,h:-		SRR3724530
SRR4119785	2020-07-15 20:50:11	Salmonella enterica	Aberdeen		SRR4119785
SRR1752879	2020-07-15 20:50:11	Salmonella enterica	Agona	aph(3")-Ib,fosA7,sul2,tet(A)	SRR1752879
SRR1643091	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Ajiobo		SRR1643091
SRR1427104	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Amsterdam		SRR1427104
SRR3402078	2020-07-15 20:50:11	Salmonella enterica	Anatum	blaCMY-2,tet(C)	SRR3402078
SRR2585567	2020-07-15 20:50:11	Salmonella enterica	Bangkok		SRR2585567
SRR1653383	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Bardo		SRR1653383
SRR4842191	2020-07-15 20:50:11	Salmonella enterica	Braenderup		SRR4842191
SRR3049577	2020-07-15 20:50:11	Salmonella enterica	Bron		SRR3049577
SRR2421532	2020-07-15 20:50:11	Salmonella enterica	Cerro		SRR2421532
SRR1220728	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Chicago		SRR1220728
SRR3194565	2020-07-15 20:50:11	Salmonella enterica	derby	aac(3)-VIa,aadA1,fosA7,q...	SRR3194565
SRR1030845	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Dublin		SRR1030845
SRR1183899	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Enteritidis		SRR1183899
SRR1646564	2020-07-15 20:50:11	Salmonella enterica	Hadar	aph(3")-Ib,aph(6)-Id,tet(A)	SRR1646564
SRR1105667	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Heidelberg		SRR1105667
SRR3289809	2020-07-15 20:50:11	Salmonella enterica	Heidelberg	aac(3)-VIa,aadA1,blaHER-3...	SRR3289809
SRR1574295	2020-07-15 20:50:11	Salmonella enterica	IIIa 18:z4,z23:-	aac(3)-VIa,aadA1,qacEdelt...	SRR1574295
SRR1534841	2020-07-15 20:50:11	Salmonella enterica	Kentucky	aph(3")-Ib,aph(6)-Id,tet(B)	SRR1534841
SRR3476365	2020-07-15 20:50:11	Salmonella enterica	Manhattan		SRR3476365
SRR1783160	2020-07-15 20:50:11	Salmonella enterica	Montevideo	aac(6')-Ib,aadA1,aph(3")-Ib...	SRR1783160
SRR1203019	2020-07-15 20:50:11	Salmonella enterica	Newport	aph(3")-Ib,aph(6)-Id,blaCM...	SRR1203019
SRR3098669	2020-07-15 20:50:11	Salmonella enterica	Saintpaul	blaTEM-1	SRR3098669
SRR1574259	2020-07-15 20:50:11	Salmonella enterica	Schwarzengrund	aph(3")-Ib,aph(6)-Id,blaTEM...	SRR1574259
SRR1107480	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Thompson		SRR1107480
ERR340757	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Typhi		ERR340757
SRR3476793	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Typhimurium		SRR3476793
SRR3330219	2020-07-15 20:50:11	Salmonella enterica	Typhimurium var. O:5-	blaCMY-2,sul2,tet(A)	SRR3330219
SRR3205698	2020-07-15 20:50:11	Salmonella enterica	Waycross		SRR3205698
SRR6128335	2020-08-10 10:11:01	Salmonella enterica subsp. enterica	Typhimurium		SRR6128335
SRR6128336	2020-08-10 10:11:06	Salmonella enterica subsp. enterica	Typhimurium		SRR6128336
SRR6128337	2020-08-10 10:11:10	Salmonella enterica subsp. enterica	Typhimurium		SRR6128337
SRR6128338	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128338
SRR6128339	2020-08-10 10:11:14	Salmonella enterica subsp. enterica	Typhimurium		SRR6128339
SRR6128340	2020-08-10 10:11:18	Salmonella enterica subsp. enterica	Typhimurium		SRR6128340
SRR6128341	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128341
SRR6128342	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128342
SRR6128343	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128343
SRR6128344	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128344
SRR6128345	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128345
SRR6128346	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128346

Name	Modified date
Salmonella outbreak	2020-07-15 12:24:31

Database: WGS_demo_database_for_Salmonella (_DefaultUser_) | Entries: Loaded=62, View=62, Selected=0 | 7 experiments | C:\Users\Public\BIONUMERICS\WGS_demo_database_for_Salmonella

Figure 2: The *Salmonella* demonstration database: the *Main* window.

3 Select a reference sequence

To demonstrate the CFSAN SNP pipeline analysis in BIONUMERICS a publicly available benchmark dataset (NCBI BioProject PRJNA412988) for which epidemiological data is available will be used. The sequence read set data of this BioProject was already imported in the demo database and used as input for a denovo assembly job and wgMLST analysis.

We will first import the closed genome sequence of *S. typhimurium* LT2 which is most referred to in literature and is also used in Saltykova *et al.* (2018) [2] for the same benchmark dataset. An appropriate reference sequence for wgSNP analysis should be selected with care and is preferably a high quality and complete genome sequence of an organism which is closely related to all samples in the dataset. It is recommended to remove all contigs which are smaller than 1000 bp from the reference sequence.

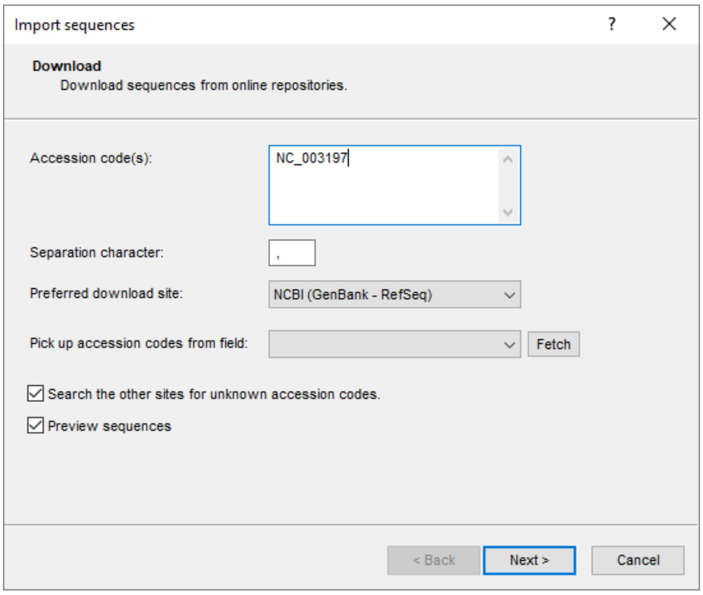
1. In the *Main* window, select ***File*** > ***Import...*** (📂, **Ctrl+I**) to open the *Import* dialog box.

2. Choose the option ***Download sequences from internet*** under the ***Sequence type data*** item in the tree and click <***Import***> (see Figure 3).

3. Enter the accession code **NC_003197** in the ***Accession codes*** input field.

4. Choose one of the available download sites from the list, e.g. ***NCBI***.

5. With the option ***Preview sequences*** checked, press <***Next***>.



Download

Download sequences from online repositories.

Accession code(s): NC_003197

Separation character: ,

Preferred download site: NCBI (GenBank - RefSeq)

Pick up accession codes from field:

☑ Search the other sites for unknown accession codes.

☑ Preview sequences

Figure 3: Import sequences dialog box.

The import routine fetches the sequence from the selected database and shows detailed information in the next step.

6. Press <***Next***>.

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

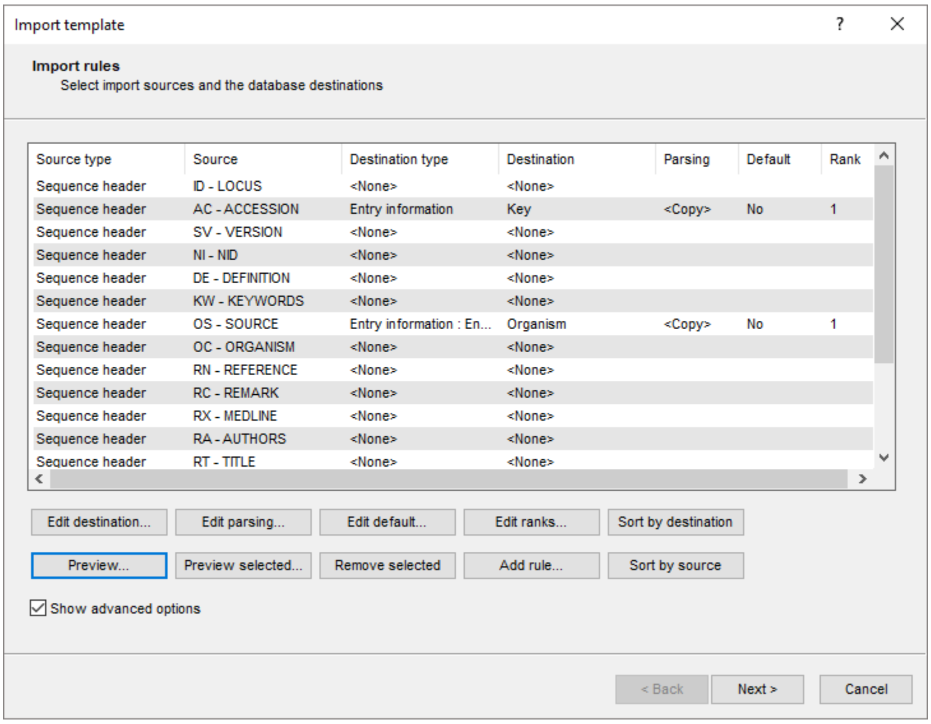
7. Click <***Create new***> to create a new import template.

Each header tag (e.g. ID, AC, . . .) corresponds to a row in the grid panel.

8. Select ***AC - ACCESSION*** in the list and click <***Edit destination***> or double-click on ***AC - ACCESSION***. Select ***Key***, and press <***OK***>.

9. Select ***OS - SOURCE*** in the list and click <***Edit destination***> or double-click on ***OS - SOURCE***. Under ***Entry info field*** select ***Organism***, and press <***OK***>.

The grid is updated (see Figure 4).



Import template

Import rules
Select import sources and the database destinations

Source type	Source	Destination type	Destination	Parsing	Default	Rank
Sequence header	ID - LOCUS	<None>	<None>			
Sequence header	AC - ACCESSION	Entry information	Key	<Copy>	No	1
Sequence header	SV - VERSION	<None>	<None>			
Sequence header	NI - NID	<None>	<None>			
Sequence header	DE - DEFINITION	<None>	<None>			
Sequence header	KW - KEYWORDS	<None>	<None>			
Sequence header	OS - SOURCE	Entry information : En...	Organism	<Copy>	No	1
Sequence header	OC - ORGANISM	<None>	<None>			
Sequence header	RN - REFERENCE	<None>	<None>			
Sequence header	RC - REMARK	<None>	<None>			
Sequence header	RX - MEDLINE	<None>	<None>			
Sequence header	RA - AUTHORS	<None>	<None>			
Sequence header	RT - TITLE	<None>	<None>			

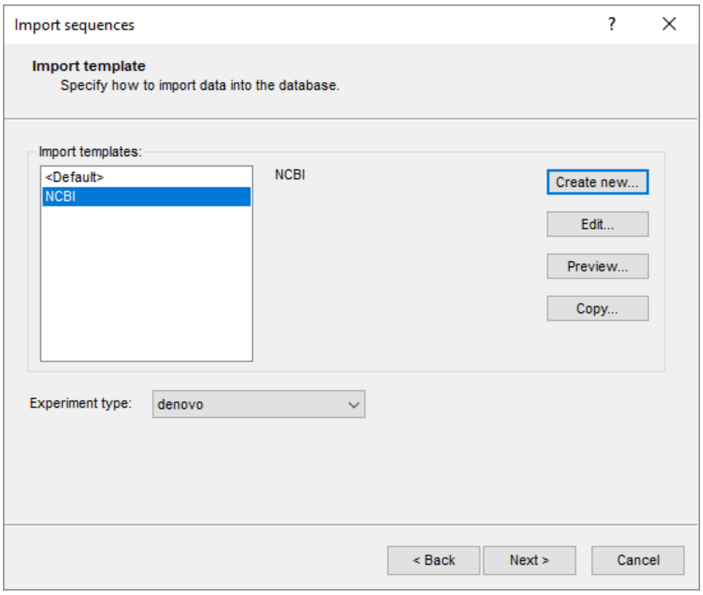
☑ Show advanced options

Figure 4: Import template.

10. Click <***Next***> and press <***Finish***>.

11. Specify a template name (e.g. **NCBI**) and optionally enter a description. Press <***OK***>.

12. Highlight the newly created template and select ***denovo*** as ***Experiment type*** (see Figure 5).



Import sequences

Import template
Specify how to import data into the database.

Import templates:
<Default>
NCBI

NCBI

Experiment type: denovo

Figure 5: Import sequences dialog box.

13. Press <***Next***>.

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

14. Press <***Finish***>.

The sequence is imported in the database and is automatically selected.

15. Click on the green colored dot of the newly created entry in the *Experiment presence* panel to open the *Sequence editor* window.

The sequence is displayed in the upper panel and a graphical representation of the sequence is displayed in the panel below. The *Annotation* panel holds the GenBank features, and the header information is stored in the *Header* panel.

16. Close the *Sequence editor* window.

4 Create a comparison

We will now create a comparison containing the entries of the benchmark dataset (i.e. the samples of BioProject PRJNA412988).

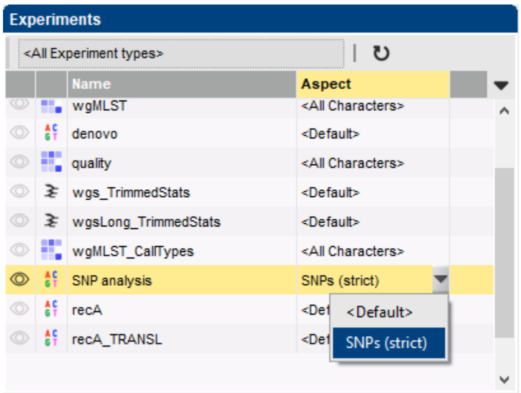
1. In the *Database entries* panel of the *Main* window make sure no entries are selected. All entries can be deselected at once by pressing ***Edit*** > ***Clear selection*** (**Ctrl+Shift+A**).

2. Select ***Edit*** > ***Find object in list...*** (, **Ctrl+Shift+F**) and in the *Find* dialog box type "PRJNA412988" and press <***Select all***>.

The 32 entries of the benchmark dataset are now selected.

3. Click on ***Edit*** > ***Create new object...*** (+) in the *Comparisons* panel to open the comparison window for the 32 selected entries.

A CFSAN SNP pipeline analysis can be launched on the calculation engine for the entries in the created comparison. Following the same principles as cluster analyses in comparisons, comparison jobs apply on the whole comparison and the data stored in the *active* experiment and (where applicable) the active aspect in the *Experiments* panel (see Figure 6 for an example).



Name	Aspect
wgMLST	<All Characters>
denovo	<Default>
quality	<All Characters>
wgs_TrimmedStats	<Default>
wgsLong_TrimmedStats	<Default>
wgMLST_CallTypes	<All Characters>
SNP analysis	SNPs (strict)
recA	<Def
recA_TRANSL	<Def

Figure 6: The *Experiments* panel in the *Comparison* window. The experiment type highlighted in yellow (**SNP analysis** is the active experiment, the **SNPs (strict)** is the active aspect.

4. Save the comparison by selecting ***File*** > ***Save*** (💾, **Ctrl+S**). Specify a name for the comparison for example "CFSAN SNP analysis" and press <***OK***>.

5 Launch the CFSAN SNP pipeline

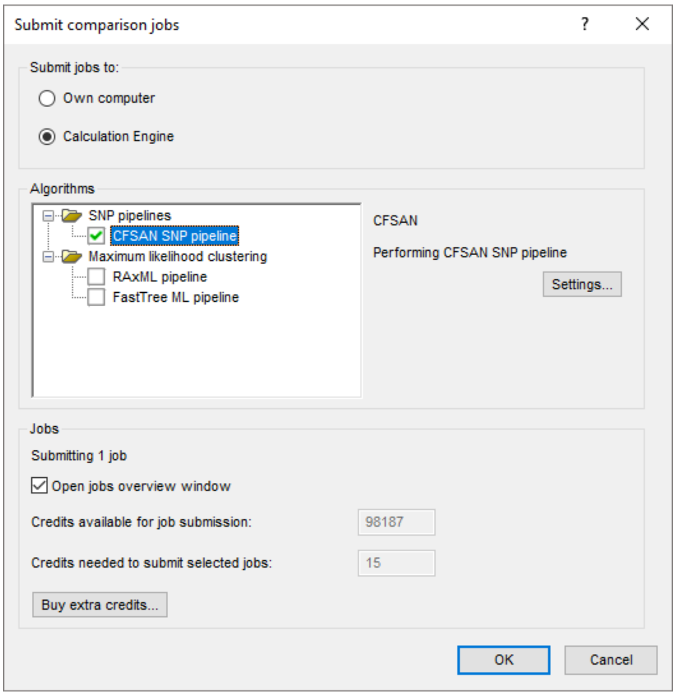
1. In the *Experiments* panel select the ***wgs*** experiment which contains the sequence read set data of the entries in the *Comparison* window.

2. To launch the CFSAN SNP pipeline comparison job, select ***File*** > ***Launch comparison jobs...*** (▷).

In case the comparison was not saved to the database yet, you will be prompted to save first. Comparisons should be saved before any jobs can be launched.

Subsequently, the *Submit comparison jobs* dialog box will open (see Figure 7).

If the active experiment does not support any comparison jobs, an error message is shown and the *Submit comparison jobs* dialog box will not appear.



Submit comparison jobs ? ✕

Submit jobs to:

○ Own computer

● Calculation Engine

Algorithms

- SNP pipelines
  - ☑ CFSAN SNP pipeline
- Maximum likelihood clustering
  - □ RAxML pipeline
  - □ FastTree ML pipeline

CFSAN

Performing CFSAN SNP pipeline

Settings...

Jobs

Submitting 1 job

☑ Open jobs overview window

Credits available for job submission: 98187

Credits needed to submit selected jobs: 15

Buy extra credits...

OK Cancel

Figure 7: The *Submit comparison jobs* dialog box.

The CFSAN SNP pipeline is not available on the local calculation engine.

3. In the ***Submit jobs to*** panel select ***Calculation engine*** and in the ***Algorithms*** panel select the ***CFSAN SNP pipeline*** algorithm (see Figure 7).

4. With the ***CFSAN SNP pipeline*** algorithm highlighted, press the <***Settings...***> to open the *CFSAN SNP pipeline settings* dialog box in which the reference sequence(s) and other settings for the CFSAN SNP pipeline job can be defined (see Figure 8).

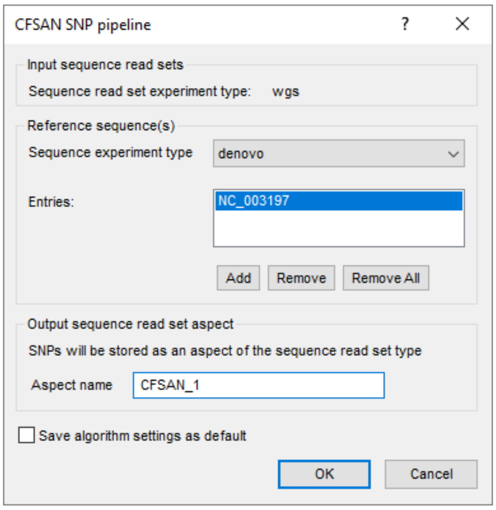
The ***Sequence read set experiment type*** selected as ***Input sequence reads*** is read-only because it was set by choosing the active experiment prior to calling the *Submit comparison jobs* dialog box.

Under ***Reference sequence(s)***, the reference genome that should be used for the mapping needs to be specified.

5. Select the ***denovo*** sequence experiment type from the sequence experiment type drop-down list.

6. Press the <***Add***> button and specify the entry that contains the reference genome i.e. the entry with Key NC_003197 (see Figure 8).

If needed, multiple reference genomes can be chosen by repeating this step.



CFSAN SNP pipeline

Input sequence read sets

Sequence read set experiment type: wgs

Reference sequence(s)

Sequence experiment type: denovo

Entries: NC_003197

Output sequence read set aspect

SNPs will be stored as an aspect of the sequence read set type

Aspect name: CFSAN_1

□ Save algorithm settings as default

Figure 8: The *CFSAN SNP pipeline settings* dialog box.

The resulting SNP matrix (i.e. the output from the CFSAN SNP pipeline) will be stored as an aspect of the input sequence read set experiment type. An ***Aspect name*** can be entered manually or the default name can be accepted.

To avoid having to re-enter the above settings for a subsequent analysis, one can save them as defaults to the database with ***Save algorithm settings as default***.

7. Select <***OK***> to confirm the settings and close the *CFSAN SNP pipeline settings* dialog box.

8. In the *Submit comparison jobs* dialog box leave the option ***Open jobs overview window*** selected and select <***OK***> to launch the CFSAN SNP pipeline job.

The job is submitted to the Calculation Engine and the *Job overview* window opens. In the *Job overview* window, the job type, job name, time of submission, job status, a description of the job, its progress and much more can be monitored.

6 Import and analyse the CFSAN SNP pipeline results

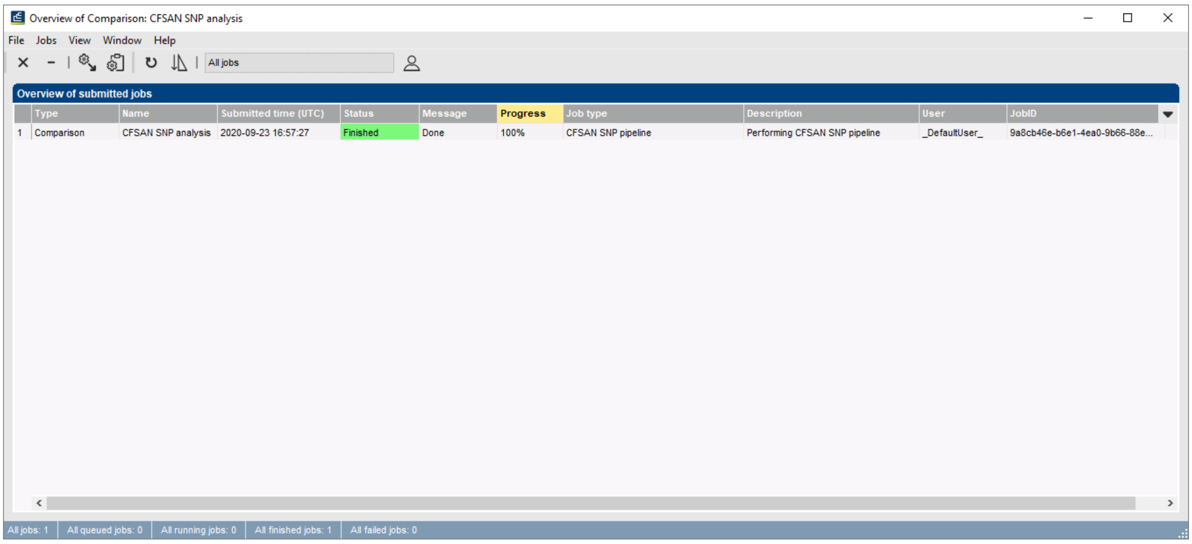
Once the job has been finished (see Figure 9), the results can be imported in the database by selecting ***Jobs*** > ***Get results*** ( ) from the *Job overview* window.

1. When the job is finished, highlight the job and select ***Jobs*** > ***Get results*** ( ) to import the results in the comparison window.

2. Close the *Job overview* window.

The *Comparison* window now looks like Figure 10.

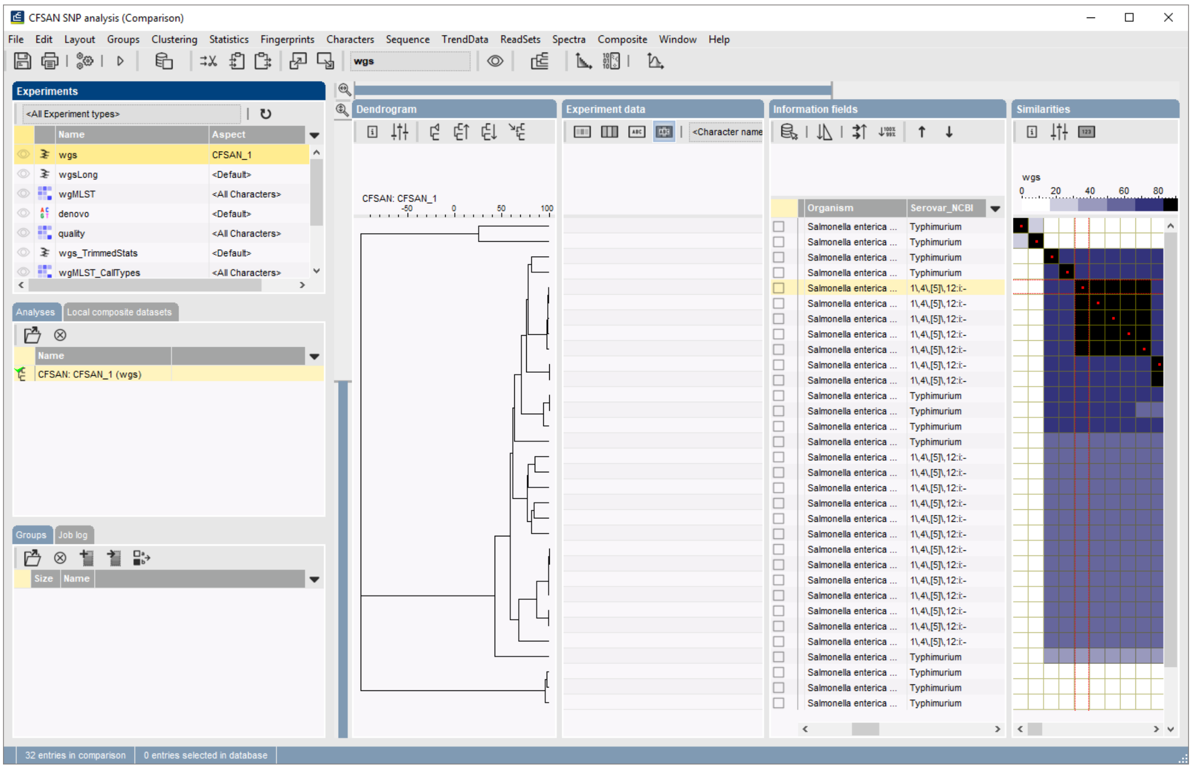
In the *Experiments* panel, the aspect containing the CFSAN SNP pipeline results is selected. The default name of the first CFSAN aspect is ”CFSAN_1”). The *Analyses* panel lists the performed analyses (for a CFSAN analysis, the analysis type (CFSAN), the aspect (CFSAN_1) and the ex-


Overview of Comparison: CFSAN SNP analysis
File Jobs View Window Help
All jobs
Overview of submitted jobs

	Type	Name	Submitted time (UTC)	Status	Message	Progress	Job type	Description	User	JobID
1	Comparison	CFSAN SNP analysis	2020-09-23 16:57:27	Finished	Done	100%	CFSAN SNP pipeline	Performing CFSAN SNP pipeline	_DefaultUser_	9a8cb46e-b6e1-4ea0-9b66-88e...

All jobs: 1 | All queued jobs: 0 | All running jobs: 0 | All finished jobs: 1 | All failed jobs: 0

Figure 9: The *Job overview* window listing a finished CFSAN SNP pipeline job.



Name	Aspect
wgs	CFSAN_1
wgsLong	<Default>
wgMLST	<All Characters>
denovo	<Default>
quality	<All Characters>
wgs_TrimmedStats	<Default>
wgMLST_CallTypes	<All Characters>

Name	
CFSAN: CFSAN_1 (wgs)	

Organism	Serovar_NCBI
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	1\,4\,[5]\,12:i:-
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	Typhimurium
Salmonella enterica ...	Typhimurium

32 entries in comparison | 0 entries selected in database

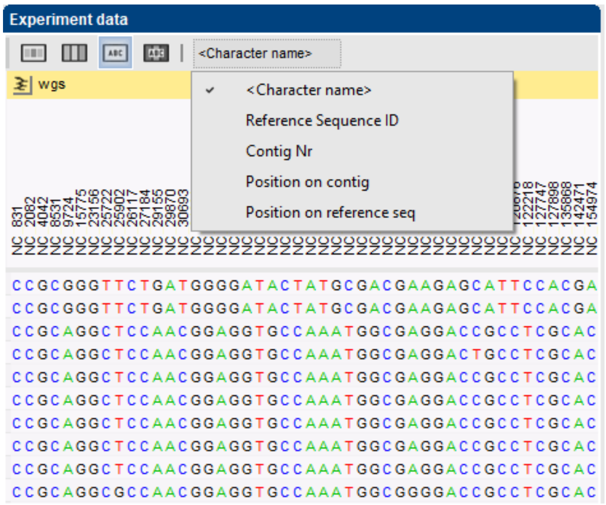
Figure 10: The *Comparison* window after import of the CFSAN SNP pipeline job results.

periment type on which the analysis was performed (wgs)is included in the name of the analysis).

3. Click on the ◎ next to the experiment name **wgs** in the *Experiments* panel to display the SNP matrix in the *Experiment data* panel.

The SNP matrix is now visible in the *Experiment data* panel. By default, the character name is present above the SNP positions but this can be changed to the reference sequence ID, the contig number, the position on the contig or the position on the reference sequence (see Figure 11).

The default dendrogram display settings in BIONUMERICS are set to similarity values. However,



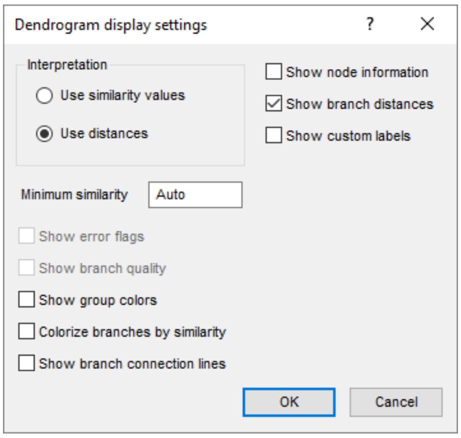
```
CCGCGGGTTCTGATGGGGATACTATGCGACGAAGAGCATTCCACGA
CCGCGGGTTCTGATGGGGATACTATGCGACGAAGAGCATTCCACGA
CCGCAGGCTCCAACGGAGGTGCCAAATGGCGAGGACCGCCTCGCAC
CCGCAGGCTCCAACGGAGGTGCCAAATGGCGAGGACTGCCTCGCAC
CCGCAGGCTCCAACGGAGGTGCCAAATGGCGAGGACCGCCTCGCAC
CCGCAGGCTCCAACGGAGGTGCCAAATGGCGAGGACCGCCTCGCAC
CCGCAGGCTCCAACGGAGGTGCCAAATGGCGAGGACCGCCTCGCAC
CCGCAGGCTCCAACGGAGGTGCCAAATGGCGAGGACCGCCTCGCAC
CCGCAGGCTCCAACGGAGGTGCCAAATGGCGAGGACCGCCTCGCAC
CCGCAGGCGCCAACGGAGGTGCCAAATGGCGGGGACCGCCTCGCAC
```

Figure 11: The *Experiment data* panel.

the CFSAN SNP pipeline calculates pairwise SNP distances for clustering. The dendrogram display settings should therefore be adjusted.

4. Select ***Clustering*** > ***Dendrogram display settings...*** (⫶) to open the *Dendrogram display settings* dialog box.

5. Select ***Use distances*** and ***Show branch distances*** (see Figure 12). Press <***OK***>.


Dendrogram display settings

Interpretation
○ Use similarity values
● Use distances

Minimum similarity: Auto

□ Show error flags
□ Show branch quality
□ Show group colors
□ Colorize branches by similarity
□ Show branch connection lines

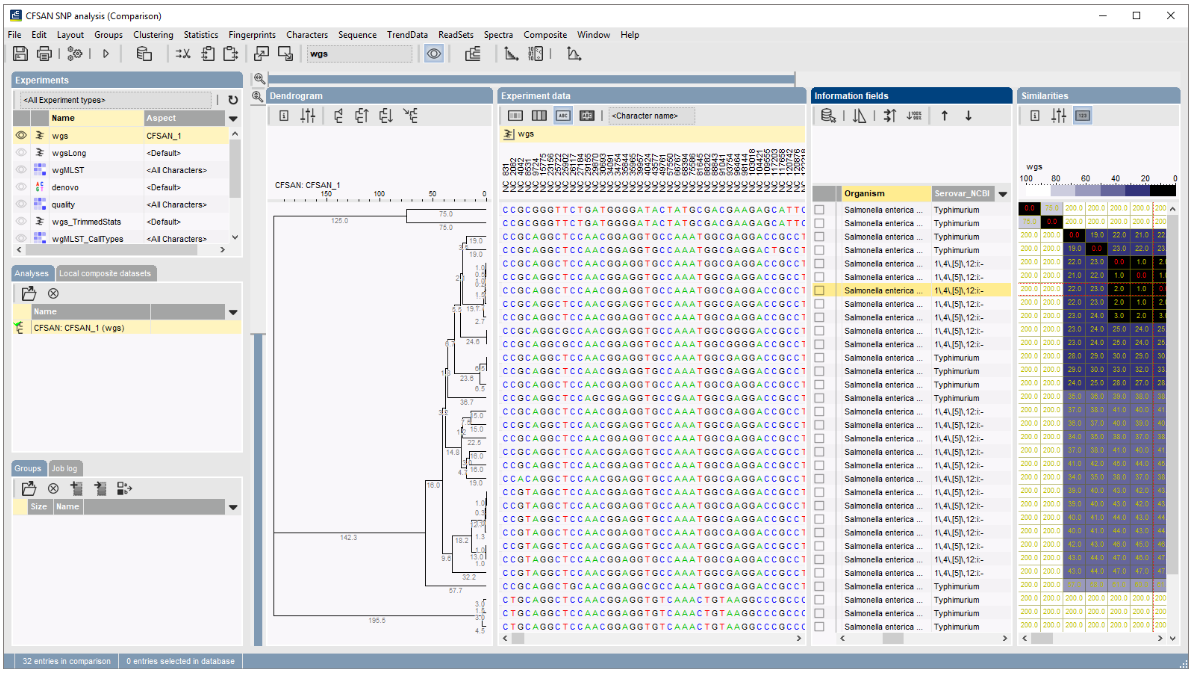
□ Show node information
☑ Show branch distances
□ Show custom labels

Figure 12: The dendrogram display settings.

6. Select ***Clustering*** > ***Similarity matrix*** > ***Show values*** (▤) to visualise the pairwise SNP distances in the *Similarities* panel.

The *Comparison* window now looks like Figure 13.

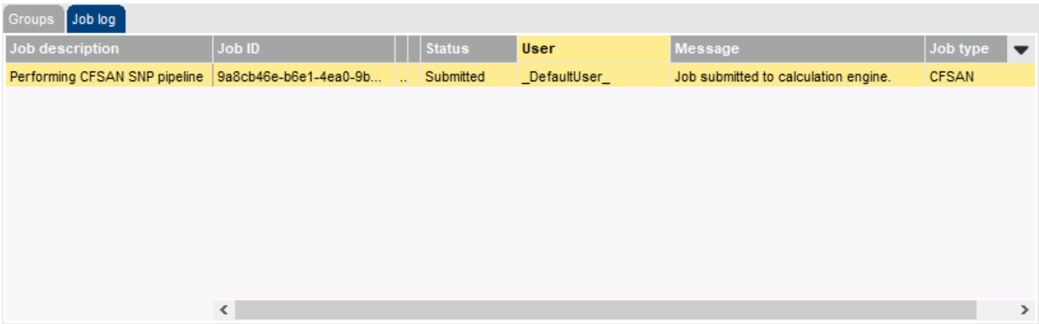
The log file of the CFSAN SNP pipeline job can be consulted in the Job log panel (see Figure 14).


CFSAN SNP analysis (Comparison)
File Edit Layout Groups Clustering Statistics Fingerprints Characters Sequence TrendData ReadSets Spectra Composite Window Help
Experiments
<All Experiment types>
Dendrogram
Experiment data
<Character name>
Information fields
Similarities
Analyses
Local composite datasets
Groups
Job log
32 entries in comparison
0 entries selected in database

Name	Aspect
wgs	CFSAN_1
wgsLong	<Default>
wgMLST	<All Characters>
denovo	<Default>
quality	<All Characters>
wgs_TrimmedStats	<Default>
wgMLST_CallTypes	<All Characters>

Name
CFSAN: CFSAN_1 (wgs)

Figure 13: The *Comparison* window after visualising the SNP matrix, adjusting the dendrogram display settings and visualising the pairwise SNP distances.



Job description	Job ID		Status	User	Message	Job type
Performing CFSAN SNP pipeline	9a8cb46e-b6e1-4ea0-9b...	..	Submitted	_DefaultUser_	Job submitted to calculation engine.	CFSAN

Figure 14: The Job log panel.

Bibliography

[1] Steve Davis, James B Pettengill, Yan Luo, Justin Payne, Al Shpuntoff, Hugh Rand, and Errol Strain. Cfsan snp pipeline: an automated method for constructing snp matrices from next-generation sequence data. *PeerJ Computer Science*, 1:e20, 2015.

[2] A. Saltykova, V. Wuyts, W. Mattheus, S. Bertrand, NHC Roosens, and K. Marchal. Comparison of snp-based subtyping workflows for bacterial isolates using wgs data, applied to salmonella enterica serotype typhimurium and serotype 1,4,[5],12:i:-. *PLoS ONE*, 13(2), 2018.