BIONUMERICS Tutorial:

wgMLST typing in the Listeria monocytogenes 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 of *Listeria monocytogenes*.

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

2 Preparing the database

The **WGS demo database** for *Listeria monocytogenes* 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).

2. Select the **WGS_demo_database_for_Listeria_monocytogenes** from the list and select ***Database*** > ***Download*** ( ).

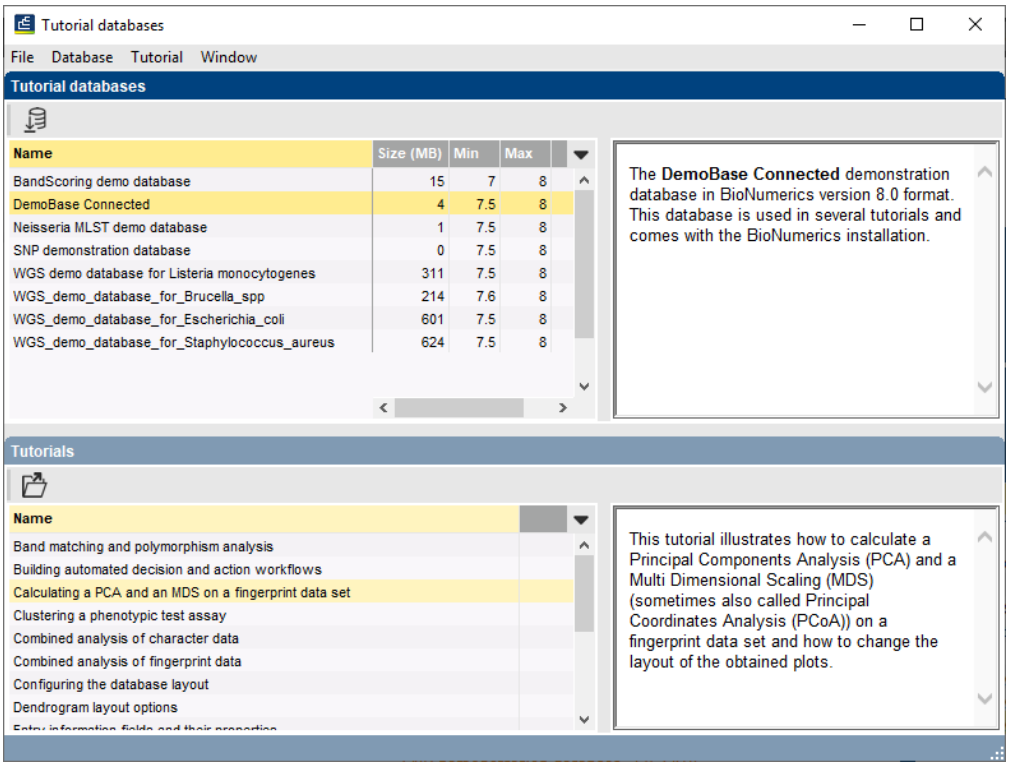


Figure 1: The *Tutorial databases* window, used to download the demonstration database.

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_Listeria_monocytogenes** appears in the *BIONUMERICS Startup* window.

5. Double-click the **WGS_demo_database_for_Listeria_monocytogenes** 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 *Listeria monocytogenes* is also available on our website. This backup can be restored to a functional database in BIONUMERICS.

6. Download the file `WGS_LMO1.bnbk` file from https://www.applied-maths.com/download/sample-data, under 'WGS_demo_database_for_Listeria_monocytogenes'.

In contrast to other browsers, some versions of Internet Explorer rename the `WGS_LMO1.bnbk` database backup file into `WGS_LMO1.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 Listeria demobase”.

10. Click <***OK***> to start restoring the database from the backup file (see Figure 2).

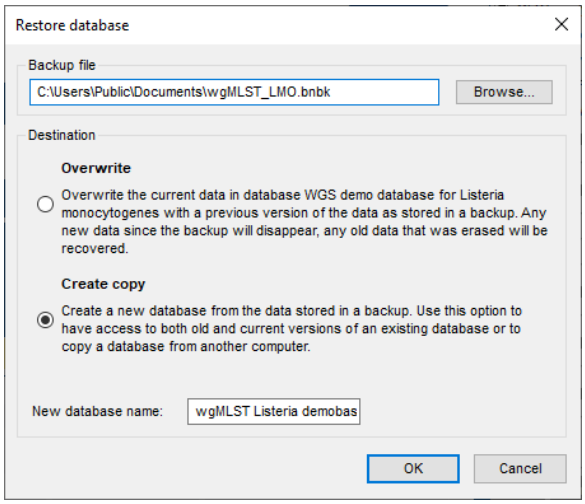


Figure 2: Restoring the WGS demonstration database from the BN backup file `WGS_LMO1.bnbk`.

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

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

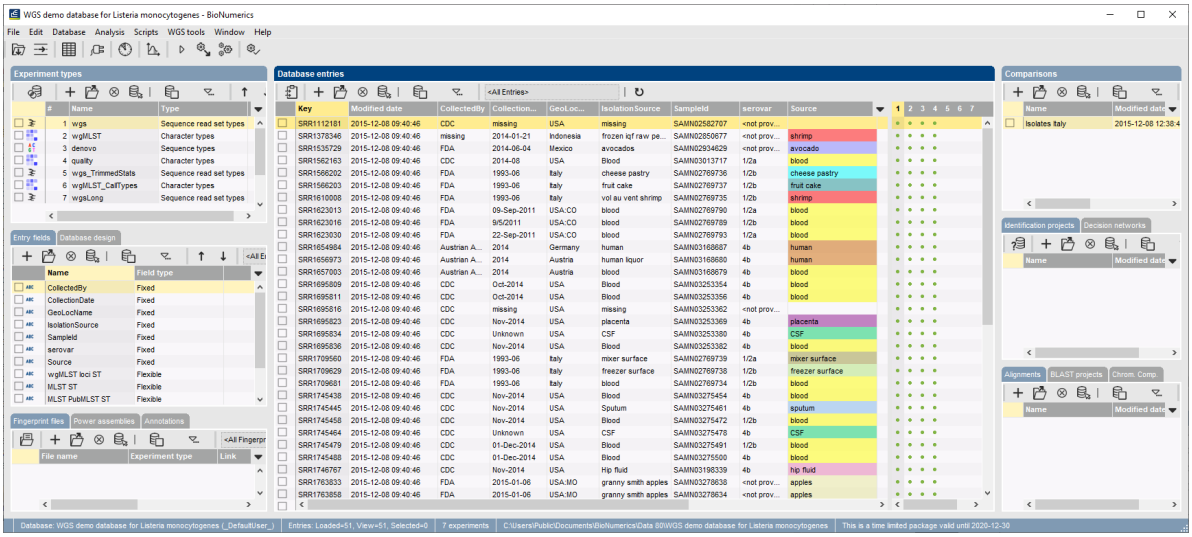


Figure 3: The *Listeria monocytogenes* demonstration database: the *Main* window.

3 About the demonstration database

The WGS Listeria demobase (see 2) contains links to sequence read set data on NCBI's sequence read archive (SRA) for 51 publicly available sequencing runs. Sequence read set experiment type **wgs** contains the link to the sequence read set data on NCBI (SRA) with some raw data statistics.

The full wgMLST analysis (de novo assembly, assembly-based calls and assembly-free calls) was performed on this set of samples using default settings and the *L. monocytogenes* wgMLST scheme on the Applied Maths Calculation Engine.

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

The calculation engine project is linked to the *Listeria monocytogenes* 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).

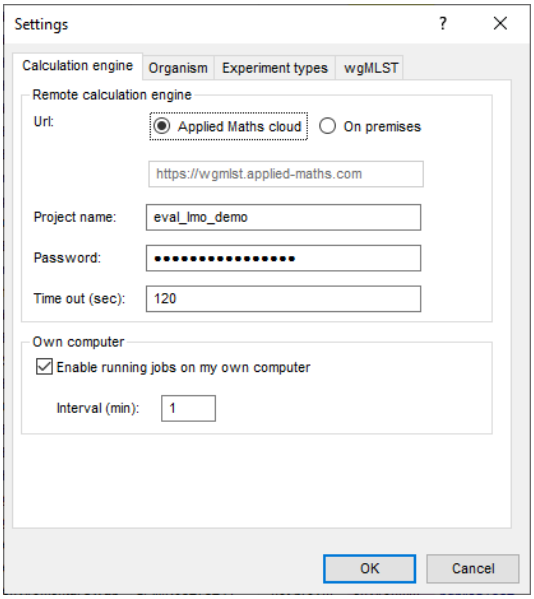


Figure 4: The *Calculation engine* tab of the *Calculation engine settings* dialog box.

2. Click on the *wgMLST tab* (see Figure 5) and press the <***Auto submission criteria***> button (see Figure 6).

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.

3. Click <***Cancel***> twice to close the *Calculation engine settings* dialog box.

Experiment types linked to wgMLST analysis are present in the database for each of the entries and are displayed in the *Experiment types* panel (see Figure 7):

- Character experiment type **wgMLST** contains the allele calls for detected loci in each sample, where the consensus from assembly-based and assembly-free calling resulted in a single allele ID.
- Sequence experiment type **denovo** contains the results from the de novo assembly algorithm, i.e. concatenated de novo contig sequences.

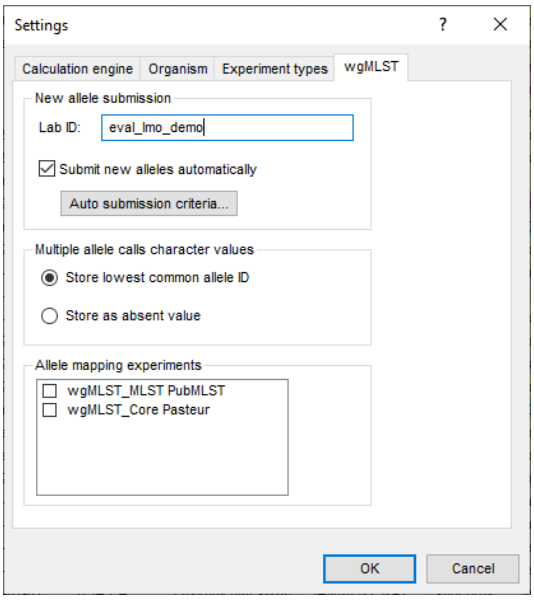


Figure 5: The *wgMLST tab* of the *Calculation engine settings* dialog box.

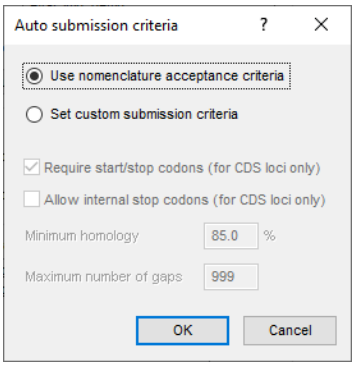


Figure 6: The *Auto submission criteria* dialog box.

- Character experiment type **quality** contains quality statistics for the raw data, the de novo assembly and the different allele identification algorithms.
- Sequence read set experiment type **wgs_TrimmedStats**: contains some data statistics about the reads retained after trimming.
- Character experiment type **wgMLST_CallTypes**: contains details on the call types.

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).

Additional information, stored in entry info fields (CollectionDate, CollectedBy, serovar, etc.) was collected from the corresponding publications and added to the demonstration database.

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:

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

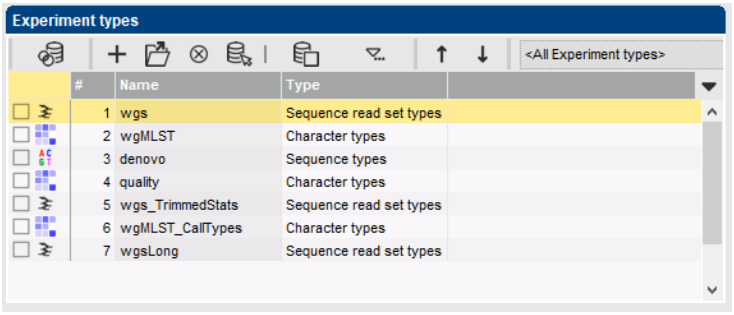


Figure 7: The *Experiment types* panel of the *Main* window.

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 8).

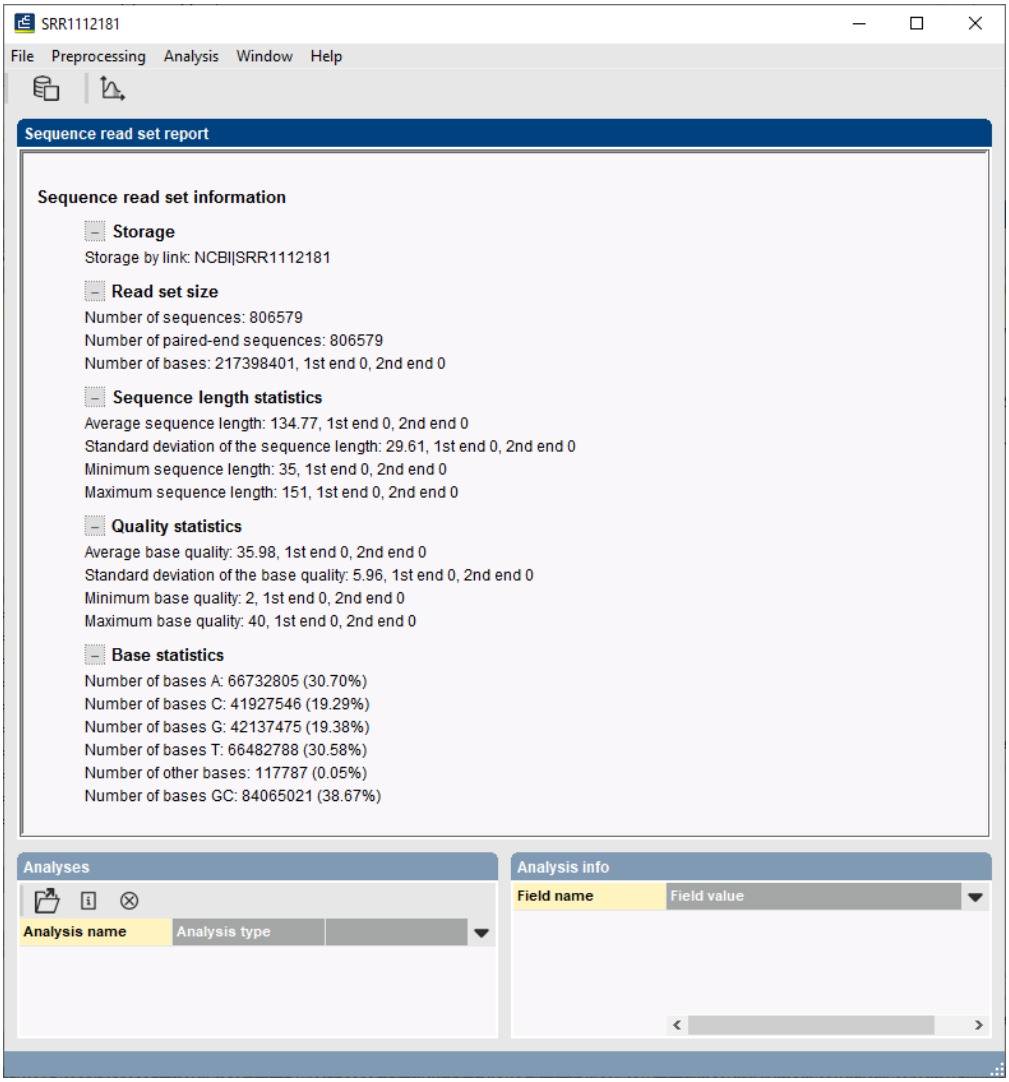


Figure 8: The sequence read set experiment card for an entry.

5. Close the *Sequence read set experiment* window.

6. 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 resulted in a single allele ID (see Figure 9).

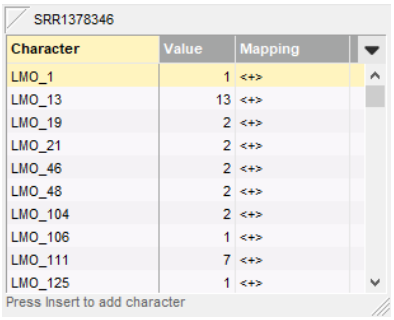


Figure 9: The character experiment card for an entry.

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

8. Click on the green colored dot for one of the entries in the third column in the *Experiment presence* panel. Column 3 corresponds to the third experiment type listed in the *Experiment types* panel, which is **denovo** in the default configuration.

The *Sequence editor* window opens, containing the results from the de novo assembly algorithm, i.e. concatenated de novo contig sequences (see Figure 10).

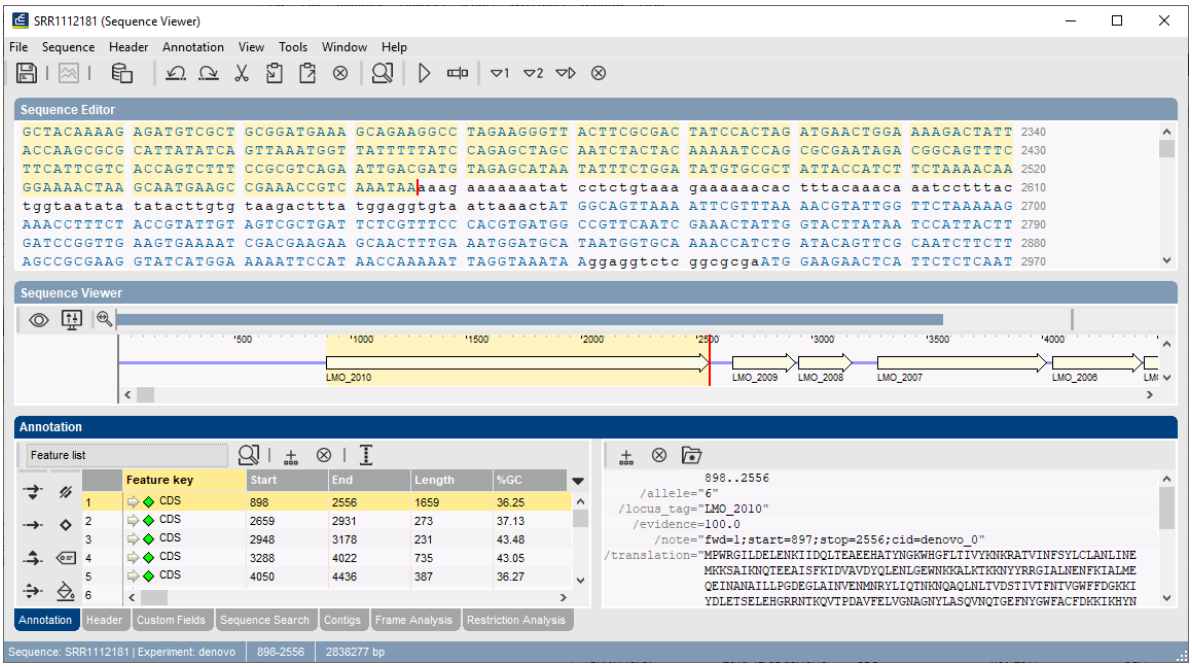


Figure 10: The *Sequence editor* window.

9. Close the *Sequence editor* window.

10. 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, the de novo assembly and the different allele identification algorithms (see Figure 11).

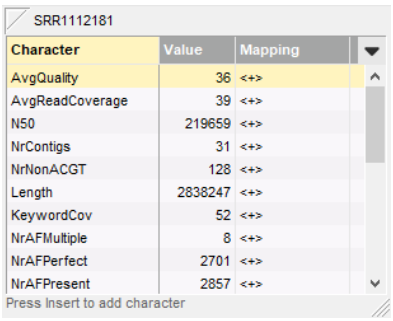


Figure 11: The character experiment card for an entry.

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

4 Subschemes

1. In the *Main* window double-click the character experiment type **wgMLST** in the *Experiment types* panel to call the *Character type* window (see Figure 12).

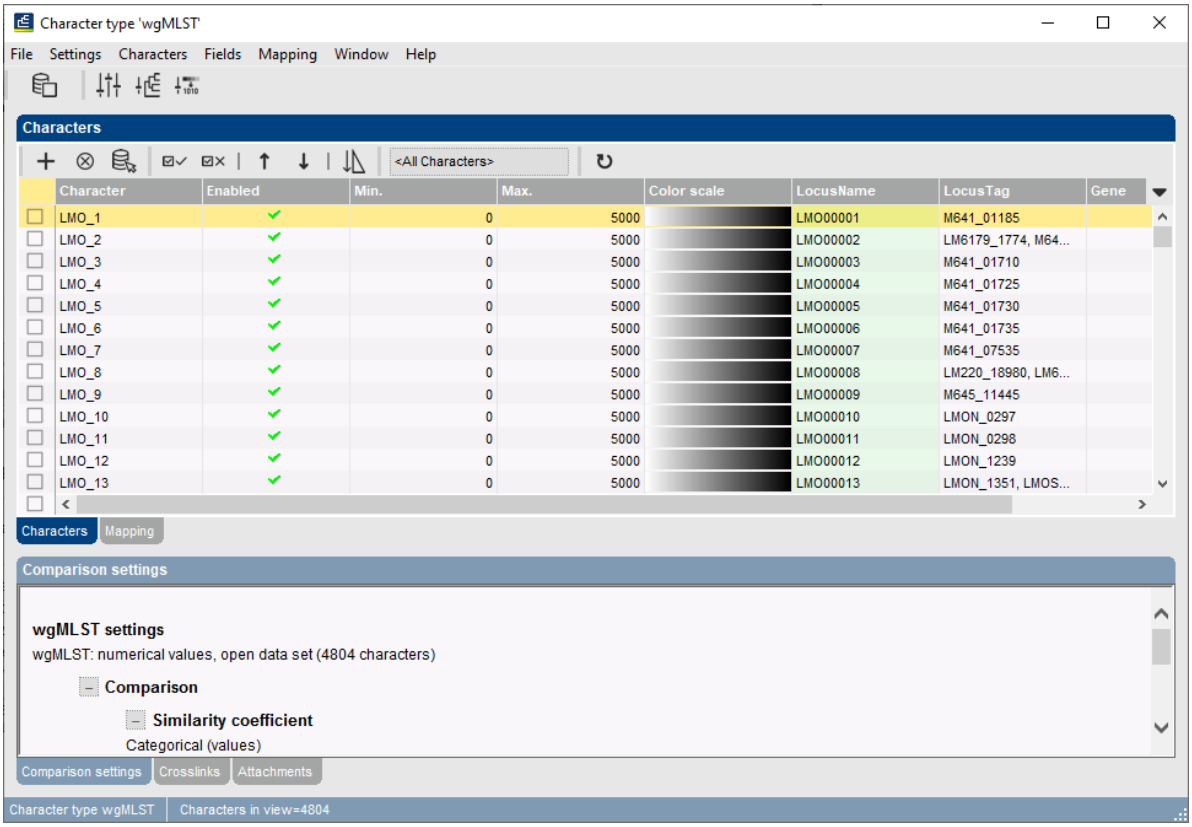


Figure 12: The *Character type* window.

Within a character experiment type, a character view can be defined that specifies a particular subset of characters.

2. Click on the drop-down bar in the toolbar (see Figure 13).

In this database following views have been defined at the curator level and are synchronized upon installation (see Figure 13): the default view **All loci**, the **MLST PubMLST** view for the traditional seven housekeeping loci, the **Core Pasteur** view, and the **wgMLST loci** view containing all loci except the ones present in the **MLST PubMLST** view.

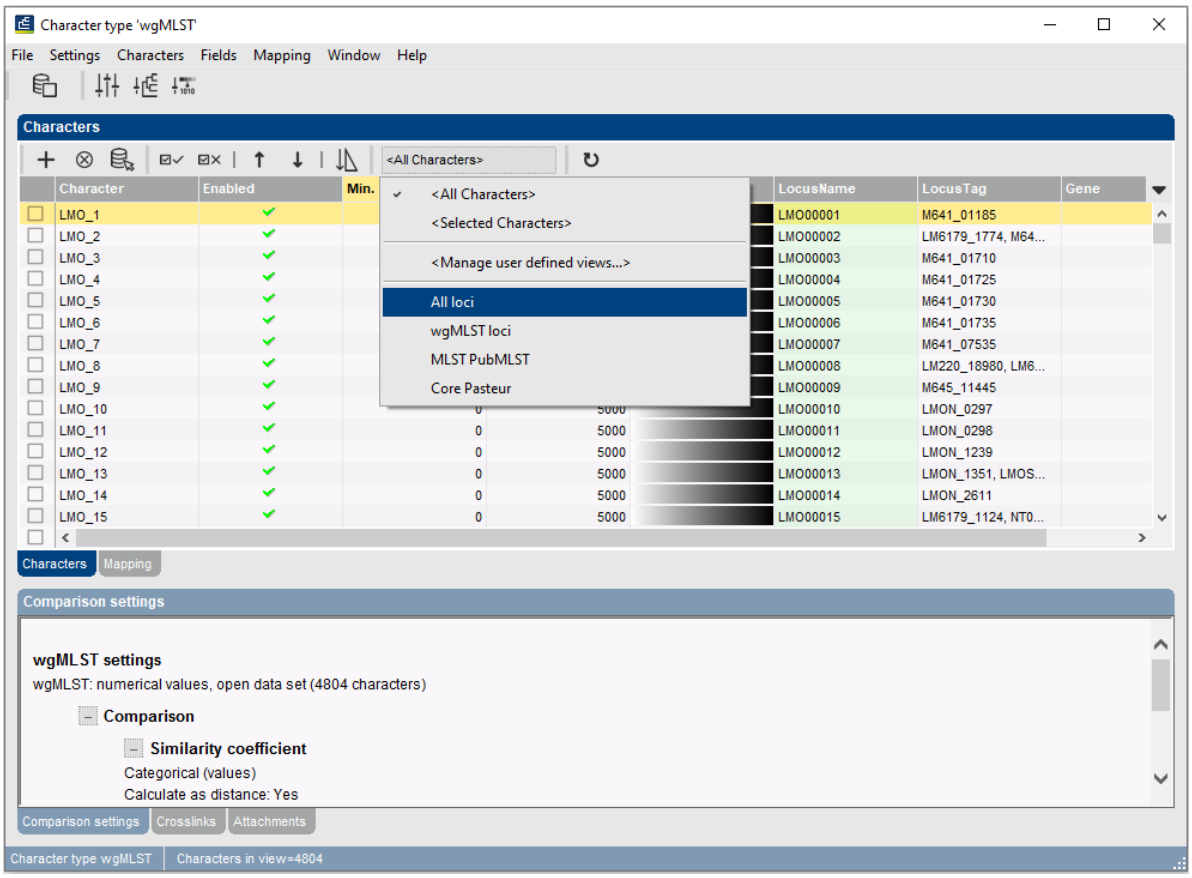


Figure 13: Views defined at the curator side.

3. Select the **MLST PubMLST** view from the list.

After selecting a character view, the window is updated (see Figure 14), and the number of characters in 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. Creating a character view can be done in two ways:

- The first method is based on a character *selection*.
- The second method is based on a *dynamic query* using the character information fields.

5. Select a few characters by selecting the characters directly in the *Character type* window (**Ctrl+click** or **Shift+click**).

The selection is synchronized with the database: any selection of characters made in the *Character type* window is reflected in other windows, e.g. the *Comparison* window, and vice versa.

6. Click on the drop-down bar in the toolbar and choose ***Manage user defined views*** (see Figure 13), alternatively select ***Characters*** > ***Character Views*** > ***Manage user defined views...*** (<All Characters>).

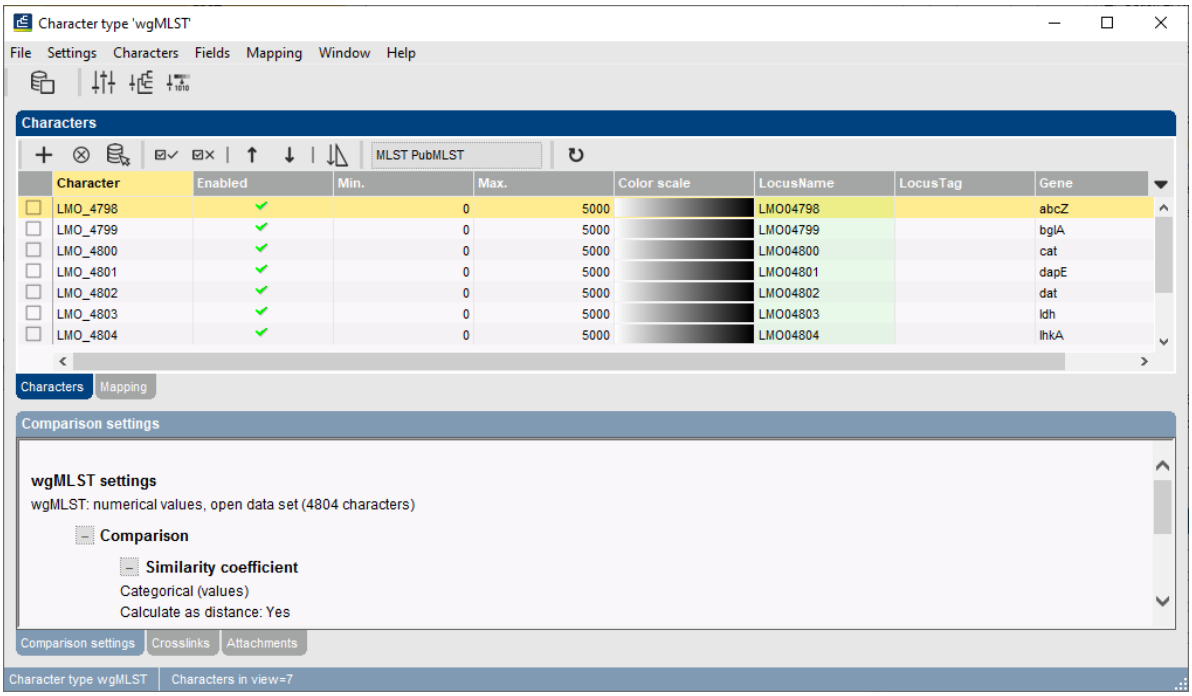


Figure 14: MLST PubMLST view.

7. Press <***Add...***>, specify a name, e.g. **MySubsetExample**, make sure **Subset based** is selected, and press <***OK***> and <***Exit***>.

The new view is added to the database and is automatically selected in the *Character type* window. The new view is available for use e.g. in the *Character type* window, *wgMLST quality assessment* window or *Comparison* window.

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

As a second example we will create a query-based view of all loci encoding a ribosomal protein. Because all those loci have a gene name starting with ”rpl” (ribosomal proteins of the large subunit) or ”rps” (ribosomal proteins of the small subunit), this subset can be easily defined with a query-based view.

9. Click on the drop-down bar in the toolbar and choose ***Manage user defined views*** (see Figure 13), alternatively select ***Characters*** > ***Character Views*** > ***Manage user defined views...*** (<All Characters>).

10. Select <***Add...***>, specify a name, e.g. “ribosomal proteins”, make sure ***Query based*** is selected and click <***OK***>.

11. Select the ***Gene*** field, change the ***Equals*** condition to ***Contains*** and type “rpl” in the white box.

12. Press <***Add new***> in the ***Statements*** panel and edit it to ***Gene Contains*** “rps”.

13. Press <***Remove all unused***>.

14. Finally, select both remaining rules (use **Ctrl+click**) and press <***OR***> in the ***Group by*** panel.

The query should now look like in Figure 15.

15. Press <***OK***> to validate the query and <***Yes***> to confirm and press <***Exit***>.

The new query-based view is created with the 47 characters that fulfill the specified criteria (see

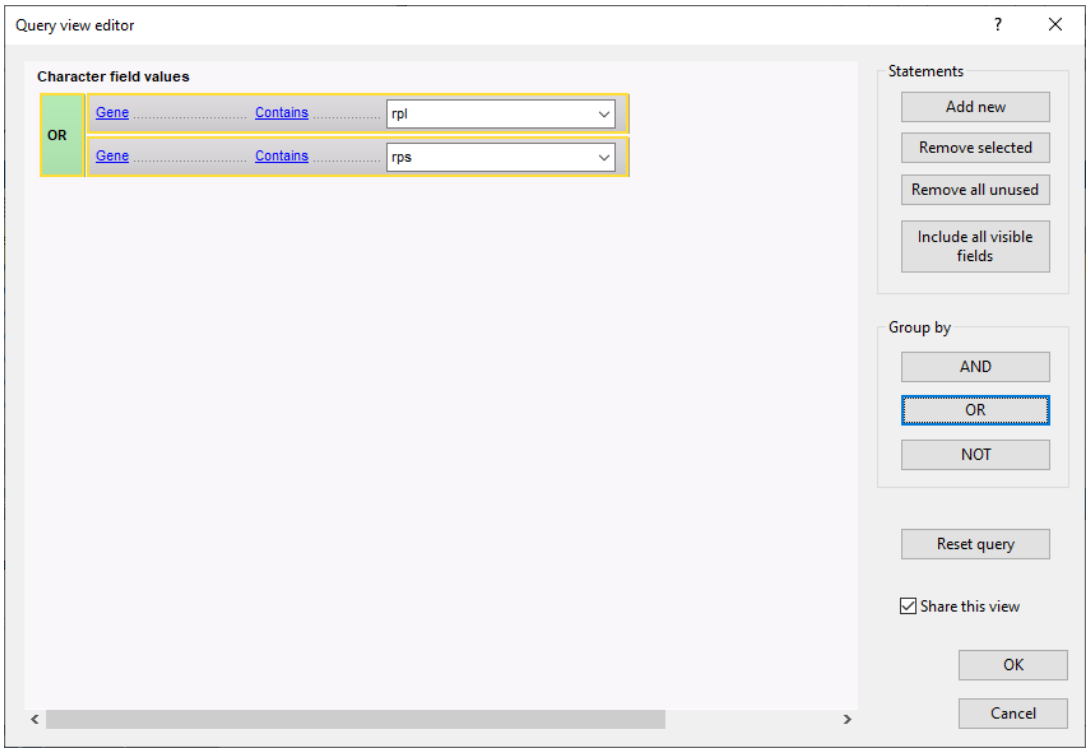


Figure 15: Query based view.

Figure 16). The new view is available for use e.g. in the *Character type* window, *wgMLST quality assessment* window or *Comparison* window.

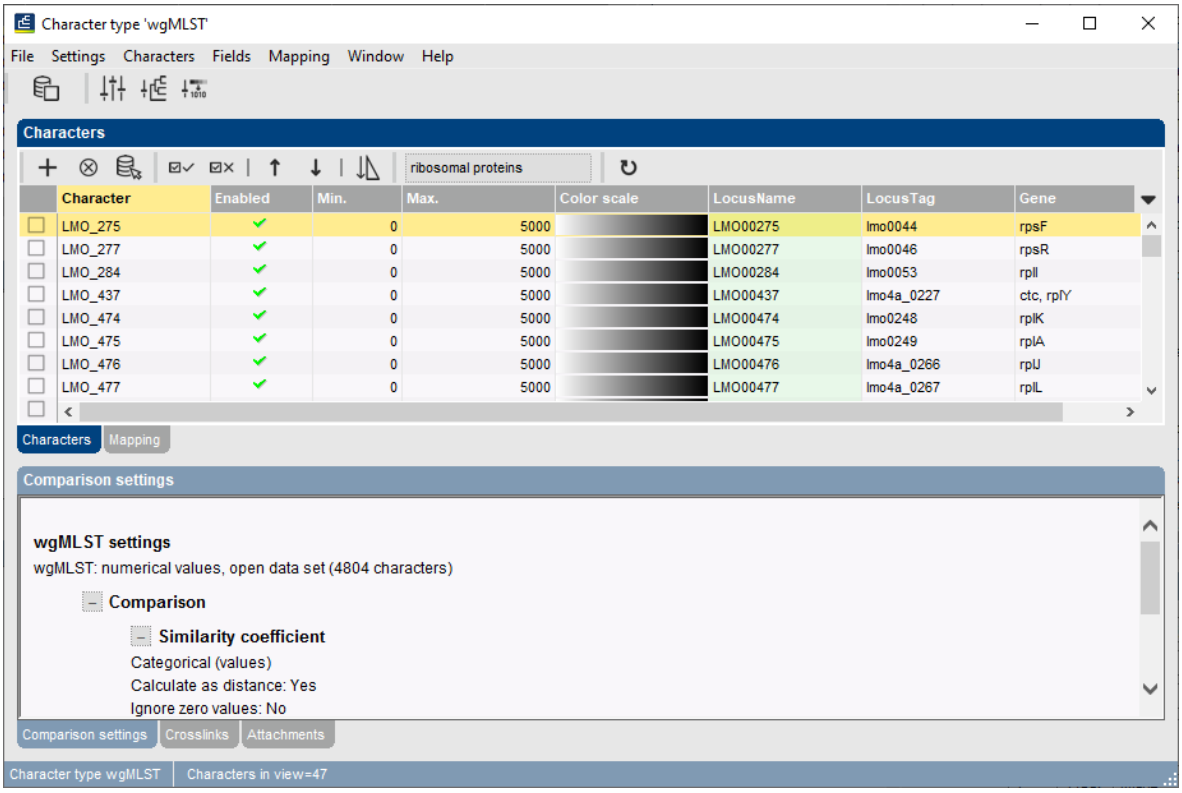


Figure 16: Query based view.

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

17. Close the *Character type* window.

5 Obtaining MLST profiles and sequence types

Using the *WGS tools plugin*, MLST profiles with public allele numbers can be obtained, i.e. using the same allele numbering as PubMLST. Additionally, the plugin allows the retrieval of public sequence types.

First, we need to activate the corresponding allele mapping experiment in the wgMLST settings:

1. Select ***WGS tools*** > ***Settings...*** to open the *Calculation engine settings* dialog box.

2. Click on the *wgMLST tab* to bring the wgMLST settings into focus.

3. Under ***Allele mapping experiments***, check ***wgMLST_MLST PubMLST*** and press <***OK***>.

A character experiment type called **wgMLST_MLST PubMLST** is created in the database in case it did not exist yet. Now, MLST profiles with exactly the same allele IDs as used on PubMLST can be obtained for all entries with a **wgMLST** experiment:

4. In the *Experiment types* panel, highlight the **wgMLST** experiment type and select ***Database*** > ***Entries*** > ***Select entries with experiment*** to make the entry selection.

5. Select ***WGS tools*** > ***Get alleles mapping***.

The allele numbers from the **wgMLST** experiments are translated into public nomenclature. The public allele numbers are then retrieved and stored in the **wgMLST_MLST PubMLST** experiments. Optionally, this can be verified in the *Comparison* window:

6. Highlight the *Comparisons* panel and select ***Edit*** > ***Create new object...*** (+) to open a comparison with the selected entries.

7. In the *Experiments* panel, click on the icon next to **wgMLST_MLST PubMLST** to visualize the MLST profiles in the *Experiment data* panel. Select ***Characters*** > ***Show values*** (123) to display the values (see Figure 17).

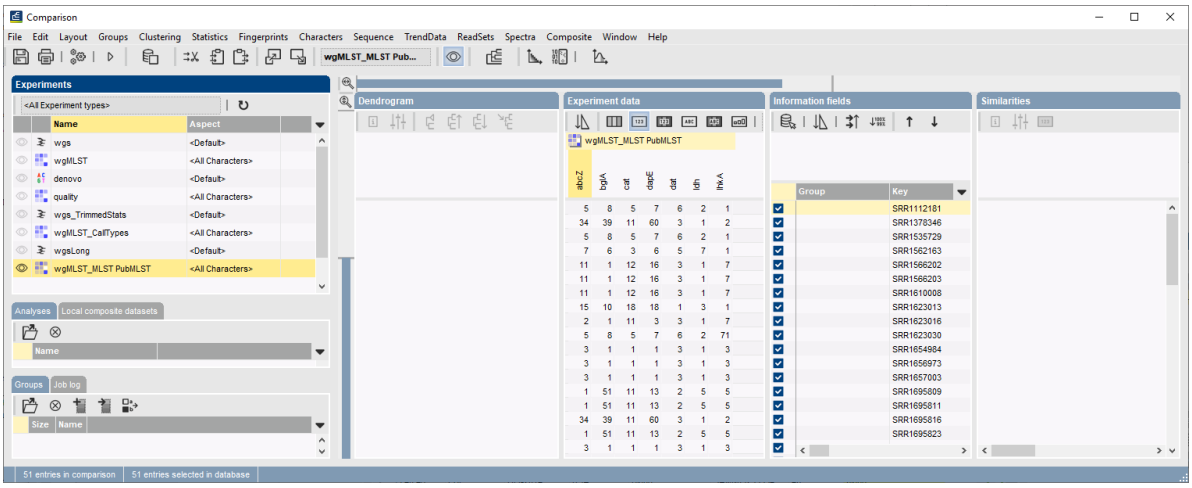


Figure 17: The *Comparison* window.

8. Close the *Comparison* window.

Next, sequence types can be assigned for the selected entries, based on the **MLST PubMLST** subscheme.

9. In the *Main* window, select ***WGS tools*** > ***Assign wgMLST sequence types...***.

This opens the *Assign sequence types* dialog box, where available typing schemes can be checked to be included in the assignment of the sequence types (see Figure 18).

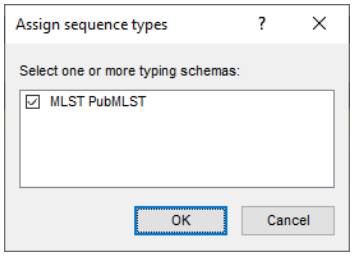


Figure 18: The *Assign sequence types* dialog box, with a single typing scheme listed.

10. Leave the subscheme **MLST PubMLST** checked and press <***OK***> to assign a sequence typing based on the 7 loci used for traditional MLST analysis.

Per entry and typing scheme, a list of allele identifications is sent to the allele database and sequence type information is returned. The sequence types are then saved to a dedicated entry information field.

In our example database, a sequence type is added in the field ***MLST PubMLST ST*** for the selected entries (see Figure 19).

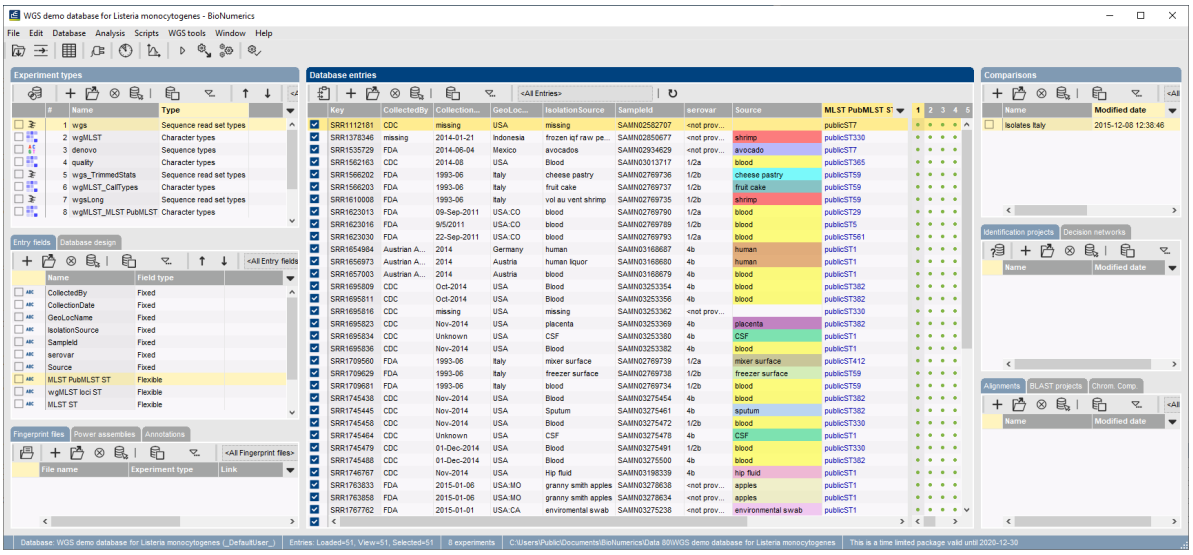


Figure 19: MLST PubMLST ST numbers.

In case an entry has an incomplete profile for the **MLST PubMLST** subscheme, no sequence type can be assigned and an error message will be generated for that entry.

6 Import of sample-specific allele sequences into the database

Once the wgMLST allele results have been imported in the database, it is possible to import the actual allele sequences for a specific wgMLST locus or a combination of loci, as defined in a subscheme.

As an example, we will import the allele sequences for the seven MLST loci from PubMLST into the database, using sequence type names that can be recognized by the *MLST online plugin*.

1. Double-click the character experiment type **wgMLST** in the *Experiment types* panel of the *Main* window).

A character information field should be present with the exact locus names as defined in the online MLST scheme. The names of the seven MLST loci as they are defined in the MLST scheme on http://bigsdb.web.pasteur.fr/listeria/ are: **abcZ**, **bglA**, **cat**, **dapE**, **dat**, **ldh**, **lhkA**.

2. In the character views drop down menu, select the ***MLST PubMLST*** view from the list.

The **Gene** character information field contains the loci names as they are defined in the online MLST scheme (see Figure 20).

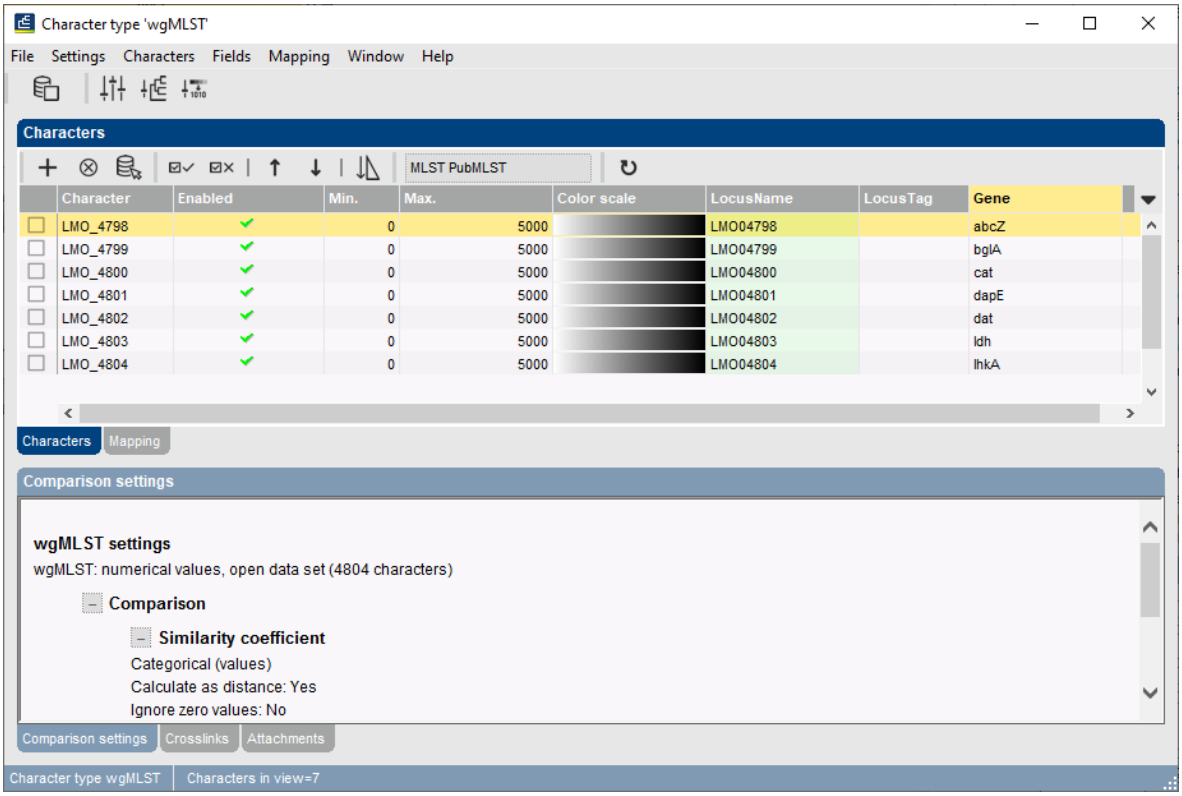


Figure 20: The *Character type* window for **wgMLST**, with locus names for the 7 MLST loci, as known in the online MLST scheme, filled in in the **Gene** character info field.

3. Close the *Character type* window.

Now the allele sequences can be imported into sequence type experiments that have the correct name for analysis by the *MLST online plugin*.

4. Make sure the *Database entries* panel is the active panel and select ***Edit*** > ***Select all*** (**Ctrl+A**) to select all entries at once.

5. Select ***WGS tools*** > ***Store wgMLST locus sequences...***, specify **MLST PubMLST** as the ***Subschema*** and select **Gene** for the ***Sequence experiment type*** (see Figure 21).

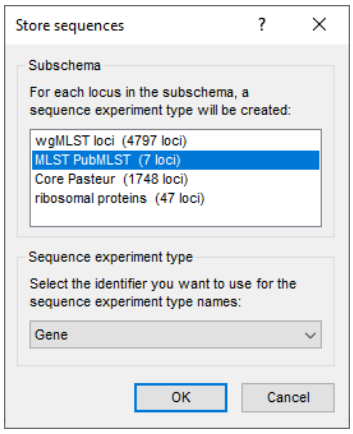


Figure 21: Store sequences.

6. Click <***OK***> to start importing the allele sequences and <***Yes***> to confirm the creation of new experiment types.

The database now contains the allele sequences for the 7 MLST loci, stored in 7 sequence experiment types that can be accessed by the *MLST online plugin*.

This can be illustrated as follows:

7. Select ***File*** > ***Install / remove plugins...*** ( ), select ***MLST online*** from the list, press <***Activate***> and confirm.

8. Choose ***Select organism from on-line list***, press <***Next***> and select **Listeria monocytogenes** from the list. Click <***Next***> three times.

9. Specify **MLST ST Bigs** next to ***Sequence types*** (see Figure 22) and press <***Next***> and <***Finish***>, press <***OK***> three times and close the *Plugins* dialog box.

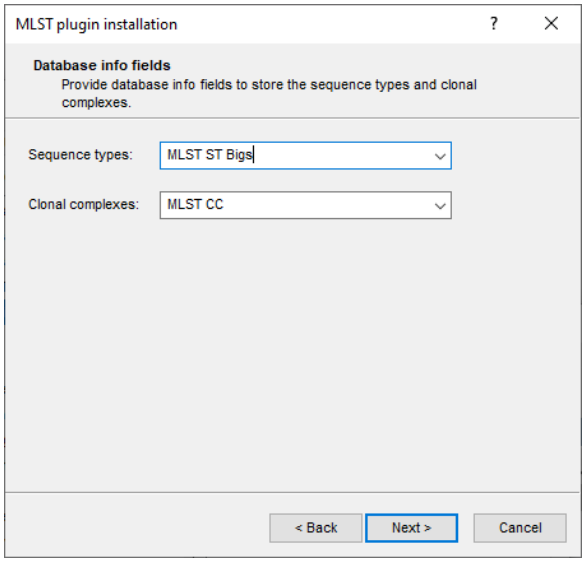


Figure 22: Sequence type information field.

10. In the *Main* window, make sure the *Database entries* panel is the active panel and select ***Edit*** > ***Select all*** (**Ctrl+A**) to select all entries at once and choose ***MLST*** > ***Identify alleles and profiles***.

The character type **MLST** now contains the allele numbers for the 7 loci as they are known in the online MLST scheme, the public sequence types are written to the entry field **MLST ST Bigs** (see Figure 23).

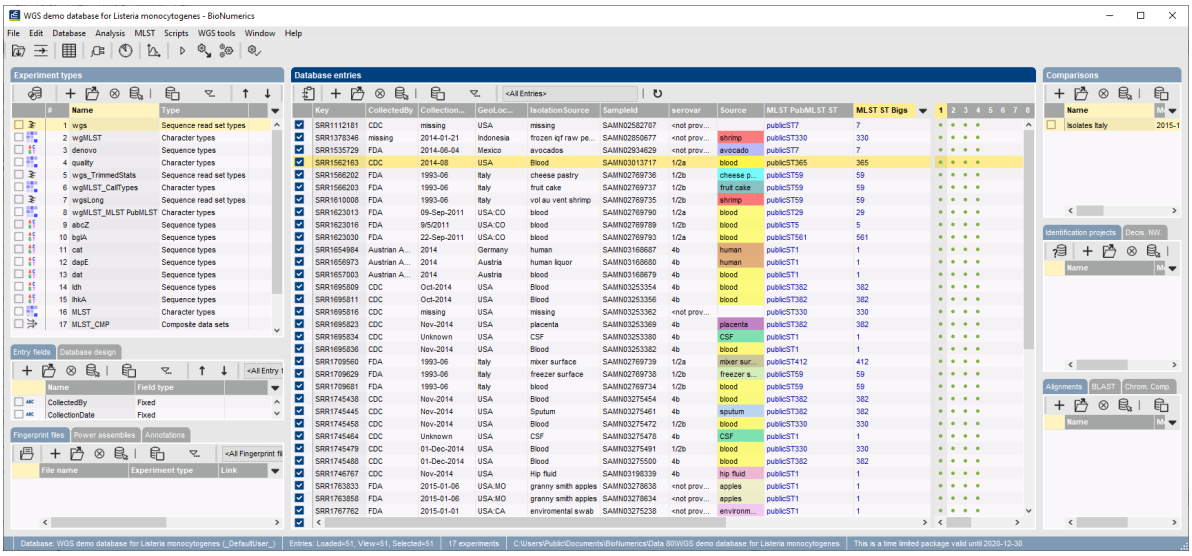
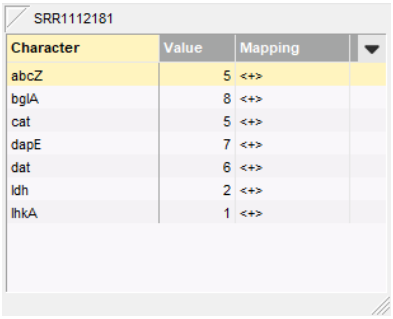


Figure 23: The *Main* window.

11. Click on the green colored dot for one of the entries in the **MLST** column in the *Experiment presence* panel.



SRR1112181

Character	Value	Mapping
abcZ	5	<+>
bglA	8	<+>
cat	5	<+>
dapE	7	<+>
dat	6	<+>
ldh	2	<+>
lhkA	1	<+>

Figure 24: The character card experiment for an entry.

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

Please consult the *MLST online plugin* manual for detailed instructions.

7 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.

7.1 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. Select the **wgMLST** character experiment in the *Experiments* panel of the *Comparison* window.

A valuable addition in the analysis of wgMLST data is the use of character views, i.e. wgMLST subschemes consisting of a subset of loci for a specific research question. Default **All characters** are included in the analysis. Another character view can be selected from the drop-down list in the **Aspect** column (see Figure 25).

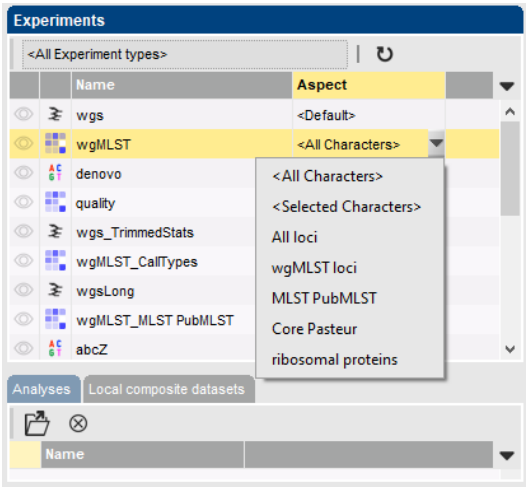


Figure 25: Character views.

7.2 Similarity based clustering

4. Make sure the correct subscheme of the **wgMLST** character experiment that you want to use for your analysis is selected in the *Experiments* panel. As an example select the **MLST PubMLST** aspect for **wgMLST** (see Figure 26).

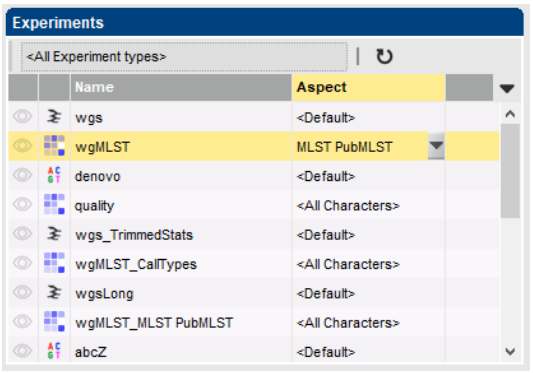


Figure 26: The **MLST PubMLST** aspect.

5. In the *Experiments* panel click on the eye icon ( ) that proceeds **wgMLST**. Select ***Characters*** > ***Show values*** ( ) to display the values of the selected aspect.

6. Select ***Clustering*** > ***Calculate*** > ***Cluster analysis (similarity matrix)...***, select ***Categorical (values)***, make sure ***Calculate as distance*** is unchecked, press <***Next***>, choose ***UPGMA*** 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 (MLST PubMLST)**.

7. Right-click on the column header of **MLST PubMLST ST** 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***>.

The *Comparison* window should now look like Figure 27.

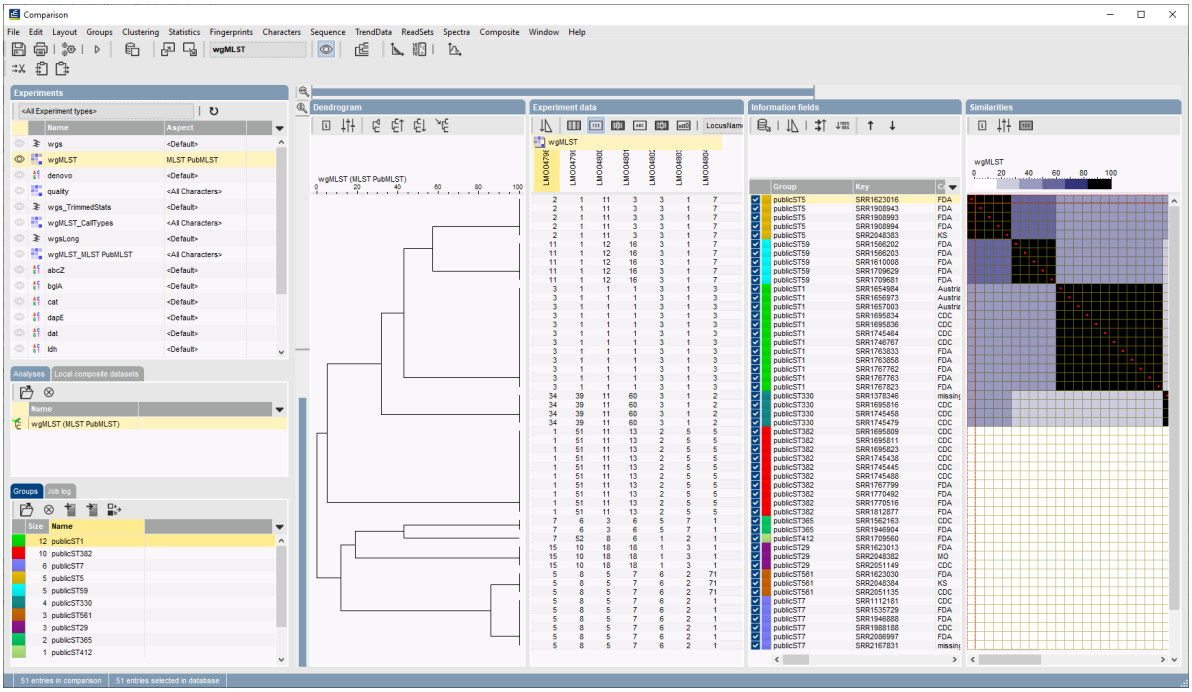


Figure 27: The *Comparison* window: dendrogram based on the MLST allele numbers.

8. Now, select the **wgMLST loci** aspect for **wgMLST** (see Figure 28).

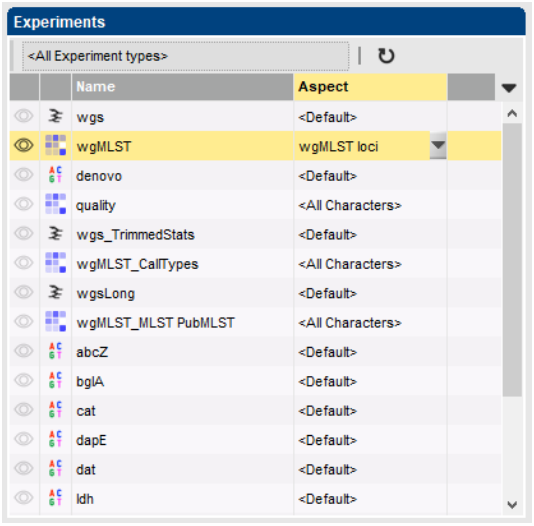


Figure 28: The **wgMLST loci** aspect.

9. Select ***Clustering*** > ***Calculate*** > ***Cluster analysis (similarity matrix)...***, select ***Categorical (values)***, make sure ***Calculate as distance*** is unchecked, press <***Next***>, choose ***UPGMA*** in the last step and press <***Finish***>.

The resulting dendrogram is displayed in the *Dendrogram* panel and the analysis is added to the *Analyses* panel (see Figure 29).

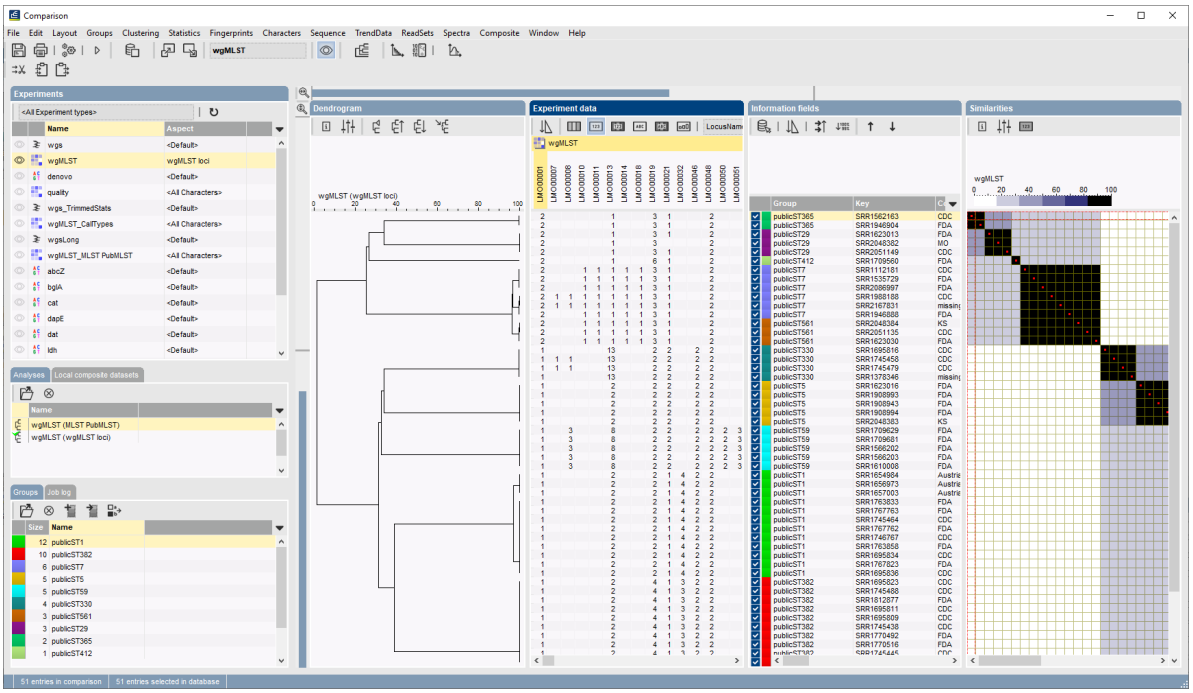


Figure 29: The *Comparison* window: dendrogram based on the wgMLST allele numbers.

10. Save the comparison with ***File*** > ***Save as...***. Specify a name (e.g. **All**) and close the comparison with ***File*** > ***Exit***.

We will now look at the data of the entries belonging to the **publicST7** MLST group:

11. Press <***F4***> to clear any selection and select the six entries belonging to the **publicST7** MLST group.

12. Highlight the *Comparisons* panel in the *Main* window and select ***Edit*** > ***Create new object...*** (+) to create a new comparison for the six selected entries.

13. Select the **wgMLST loci** aspect for **wgMLST** in the *Experiments* panel.

14. Select ***Clustering*** > ***Calculate*** > ***Cluster analysis (similarity matrix)...***.

A disadvantage of the ***Categorical (values)*** similarity coefficient is that the number of different loci cannot easily be deduced from the dendrogram or similarity matrix. The ***Categorical (differences)*** coefficient is more suitable for this purpose.

15. Choose the ***Categorical (differences)*** coefficient from the list.

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.

16. In this example, choose a ***Scaling factor*** of **1**.

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

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

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 (in this example: 1).

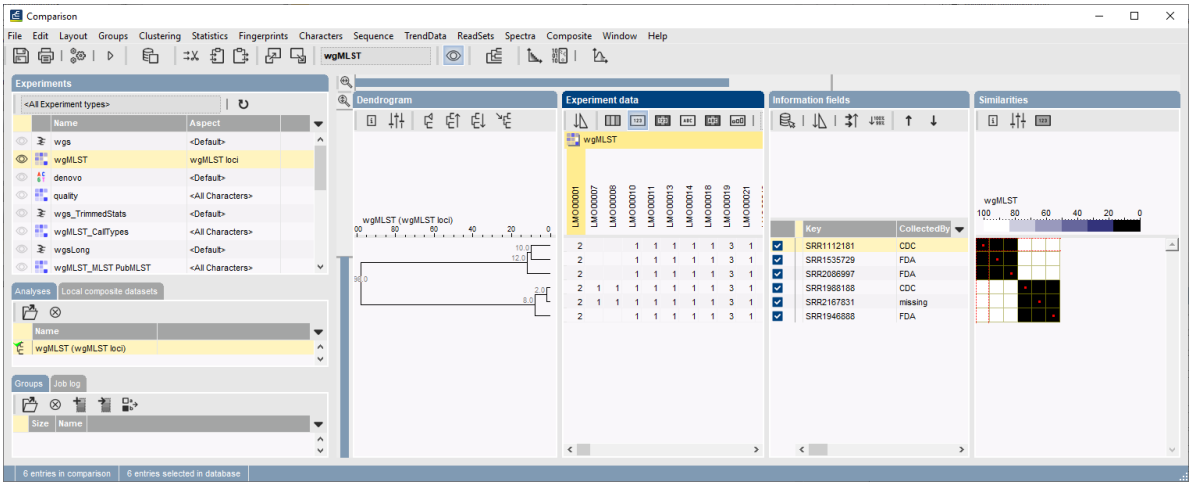


Figure 30: Complete linkage tree based on categorical differences.

19. The polymorphic loci for the set of samples in the selected scheme can be displayed with ***Characters*** > ***Filter characters*** > ***Select polymorphic characters...***.

20. 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.

21. To export the cluster analysis as it appears in the *Comparison* window select ***File*** > ***Print preview...*** (, **Ctrl+P**). The *Comparison print preview* window appears.

22. Close and optionally save the comparison.

7.3 Minimum spanning tree

A minimum spanning tree is calculated in the *Advanced cluster analysis* window which is launched from the *Comparison* window.

23. Open the saved comparison **All** or create a new comparison containing all entries in the database.

24. Create comparison groups based on the **MLST PubMLST ST** (if not already present): right-click on the column header of **MLST PubMLST ST** in the *Information fields* panel and select ***Create groups from database field***. Press <***OK***>.

25. 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 with single and double locus variance priority rules.

26. Specify an analysis name (for example **wgMLST MST**), make sure ***wgMLST (wgMLST loci)*** 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***.

A MST is now computed in the *Advanced cluster analysis* window (see Figure 31). 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.

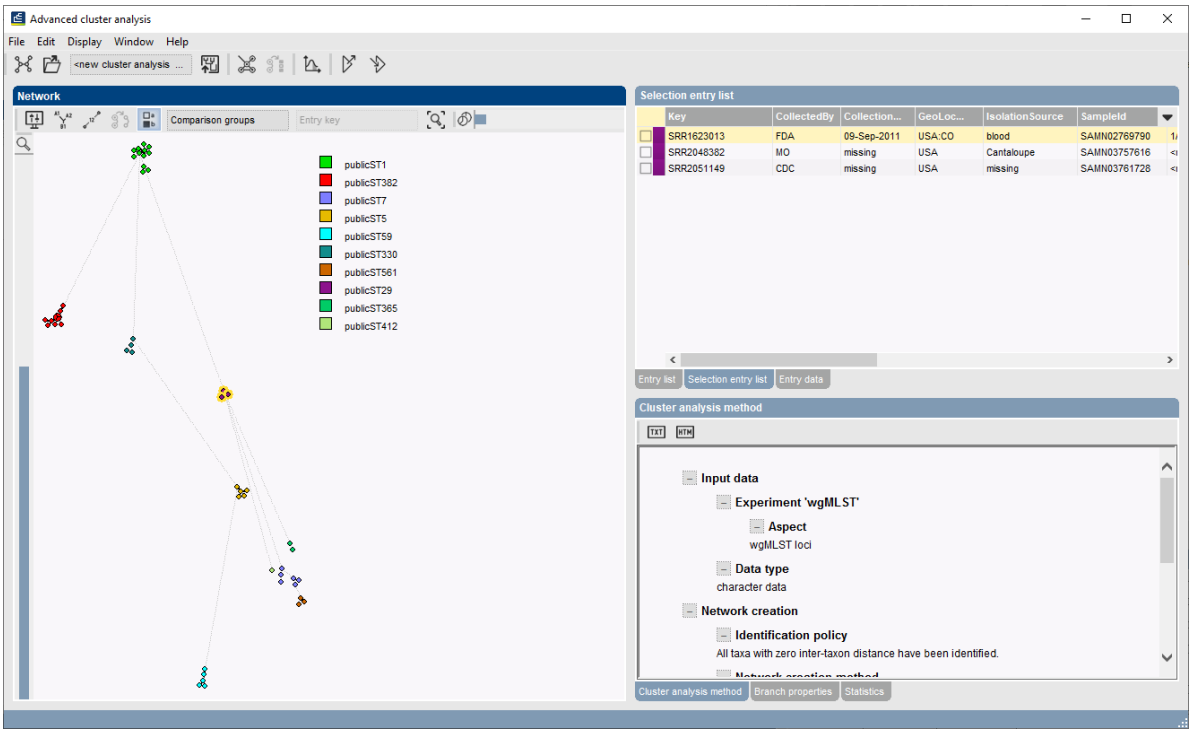


Figure 31: The *Advanced cluster analysis* window.

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

27. Press or choose ***Display*** > ***Display settings*** to open the *Display settings* dialog box.

28. In the *Node colors tab* select the **Source** from the list and press <***OK***>.

The node colors are updated according to the isolation source.

29. To go back to the comparison group coloring, repeat the previous action, or select the ***Comparison groups*** option from the toolbar (see Figure 32).

30. A node or branch can be selected by clicking on them. To select several nodes/branches hold the **Shift**-key.

31. 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.

32. Select ***Display*** > ***Zoom to fit*** or press to optimize the view of the tree.

33. Press or choose ***Display*** > ***Display settings*** to open the *Display settings* dialog box again.

34. To add more information to the MST, go to ***Display*** > ***Display settings***. In the *Branch labels and sizes panel* of the *Display settings* dialog box, we can specify that we 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”.

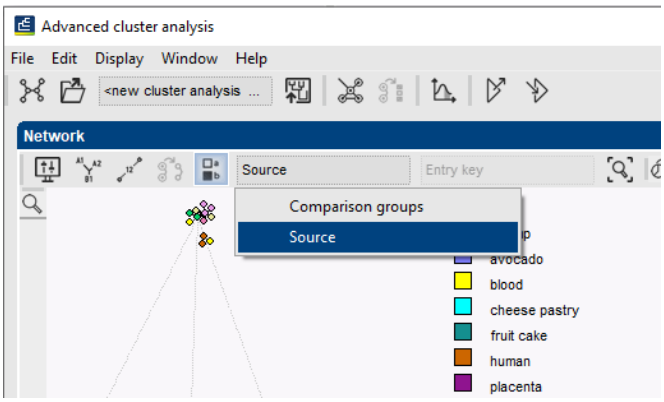


Figure 32: Groupings available in the *Advanced cluster analysis* window.

35. Click <***OK***> to close the *Display settings* dialog box. The MST is now displayed with branch labels.

36. Zooming can be done with the zoom slider on the left side of the image, and the size of the nodes can be adjusted with the zoom slider at the top. By holding the **Ctrl**-key and dragging a node with the mouse, the node can be repositioned in any direction.

37. Export the image via ***File*** > ***Export image...*** and save in the format of your choice.

38. Close the *Advanced cluster analysis* window.

39. Close the *Comparison* window.