BioNumerics Tutorial:

Dendrogram layout options

1 Introduction

A range of dendrogram display options are available in BioNumerics facilitating the interpretation of a tree. In this tutorial some of these display options will be illustrated in the *Comparison* window and *Advanced cluster analysis* window.

2 Preparing the database

The **DemoBase Connected** will be used in this tutorial and can be downloaded directly from the *BioNumerics Startup* window or restored from the back-up file available on our website:

1. To download the database directly from the *BioNumerics Startup* window, click the ***Download example databases*** link, located in the lower right corner of the *BioNumerics Startup* window. Select **DemoBase Connected** from the list and select ***Database*** > ***Download***. Confirm the download action.

2. To restore the database from the back-up file, first download the file DemoBase_Connected.bnbk from http://www.applied-maths.com/download/sample-data, under 'DemoBase Connected'.

   In the *BioNumerics Startup* window, press the ![button] button, select ***Restore database***, browse for the downloaded file and select ***Create copy***. Specify a name and click <***OK***>.

> In contrast to other browsers, some versions of Internet Explorer rename the DemoBase_Connected.bnbk database backup file into DemoBase_Connected.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.

3 Working in the database

1. In the *BioNumerics Startup* window, double-click on the **DemoBase Connected** database to open it.

2. Right-click on the **Species** information field in the *Main* window and choose ***Field properties*** from the floating menu (see Figure 1).

The *Database field properties* dialog box appears (see Figure 2).

3. Press <***Add all***> to create all existing states for the **Species** field. Confirm the action.

4. Check ***Use colors*** to display a specific color code for each field state (see Figure 2).

5. Press <***OK***> to accept the new settings.

The *Database entries* panel is updated.
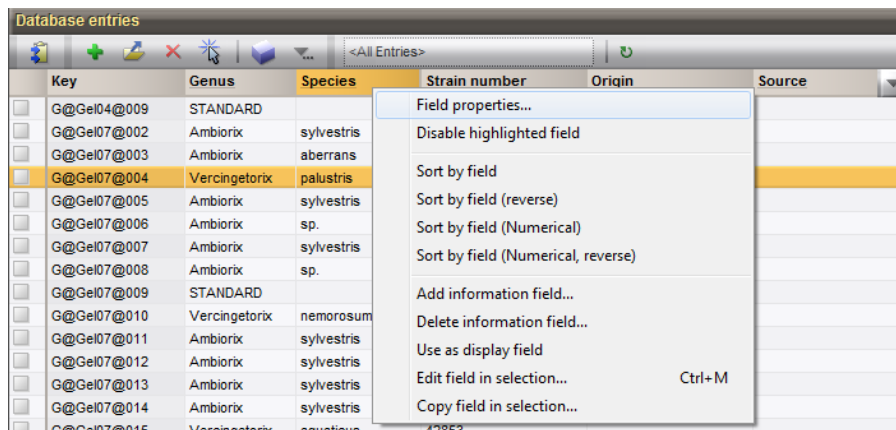
Figure 1: Floating menu.
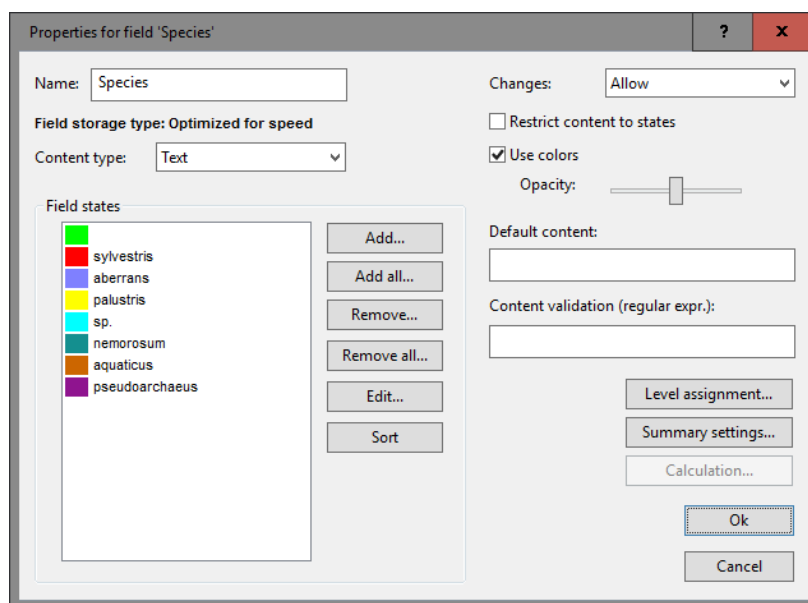

Figure 2: Species states.

4 Comparison window

1. In the *Database entries* panel of the *Main* window, select all entries using ***Edit*** > ***Select all*** (**Ctrl+A**) and unselect the STANDARDs.

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

3. Click on the next to the experiment name **FAME** in the *Experiments* panel.

Default, the character values are displayed in the *Experiment data* panel as colors according to the color scale defined for each character (see Figure 3).

4. Select ***Characters*** > ***Show values+colors*** ( ) to display the corresponding character values and colors in overlay.

5. Select ***Characters*** > ***Show colors*** ( ) to return to the original color display.

Cluster analysis is a two-step process. First, all pairwise similarity values are calculated with a **similarity**
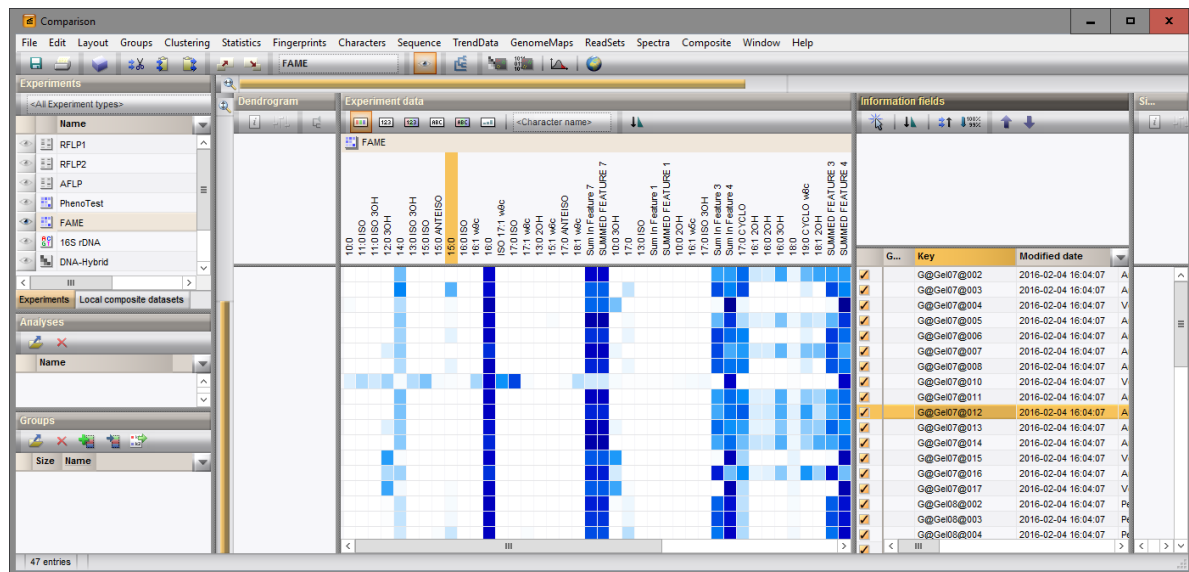
Figure 3: The *Comparison* window.

coefficient. Then, the resulting similarity matrix is converted into a dendrogram using a **clustering algorithm**. Although in practice these steps are performed together, they each require their own comparison settings.

6. Make sure **FAME** is selected in the *Experiments* panel and select *Clustering* > *Calculate* > *Cluster analysis (similarity matrix)...*.

The first step deals with the similarity coefficient for the calculation of the similarity matrix.

7. Select *Euclidean distance* from the list and press <*Next*>.

In step two, the options related to the clustering algorithms are grouped. Under *Method*, the clustering algorithm to be applied on the similarity matrix can be selected. A *Dendrogram name* can be entered in the corresponding text box. By default, the name of the experiment type appended with the aspect (here: "FAME(<All characters>)") will be used.

8. Select *UPGMA* and <*Finish*> to start the cluster analysis.

During the calculations, the program shows the progress in the *Comparison* window's caption (as a percentage), and there is a green progress bar in the bottom of the window.

When finished, the dendrogram and the similarity matrix are displayed in their corresponding panels. The cluster analysis is listed by its defined name in the *Analyses* panel of the *Comparison* window.

9. Press the **F4** key to clear any selection in the database.

10. Left-click on the dendrogram to place the cursor on any node or tip (where a branch ends in an individual entry).

11. To select entries in a cluster, click on the node of the cluster while holding the **Ctrl**-key.

12. Press *Edit* > *Cut selection* (, **Ctrl+X**) to remove the selected entries from the cluster analysis. Confirm the action. The dendrogram is automatically updated.

13. Select *Edit* > *Paste selection* (, **Ctrl+V**). The cluster analysis is recalculated automatically, and the selected entries are placed back in the dendrogram.

A branch can be moved up or down to improve the layout of a dendrogram:

14. Click the branch which you want to move up in the dendrogram and select **Clustering** > **Move branch up** (⬆).

15. Click the branch which you want to move down in the dendrogram and select **Clustering** > **Move branch down** (⬇).

To simplify the representation of large and complex dendrograms, it is possible to simplify branches by abridging them as a triangle.

16. Select a cluster of closely related entries and select **Clustering** > **Collapse/expand branch** (◁). Repeat this action to undo the abridge operation.

Comparison groups can be defined from clusters, from database fields, or just from any selection you want. As an example, we will let BioNumerics create groups based on the **Genus** names.

17. In the *Comparison* window, right-click on the field name **Genus** in the *Information fields* panel, and select **Create groups from database field** from the floating menu.

18. Keep the first option selected and confirm.

In our example three groups are created. The groups are listed in the *Groups* panel. The group color is displayed next to each entry in the *Information fields* panel (see Figure 4). The **Group** column is also updated with the name of the group, in this case the **Genus** name.

Figure 4: The *Comparison* window with groups defined.
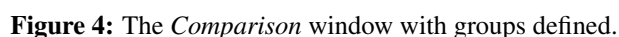
19. Select **Clustering** > **Dendrogram display settings...** (╫) to call the *Dendrogram display settings* dialog box (see Figure 5).

20. Enable **Show group colors** and press <**OK**>.

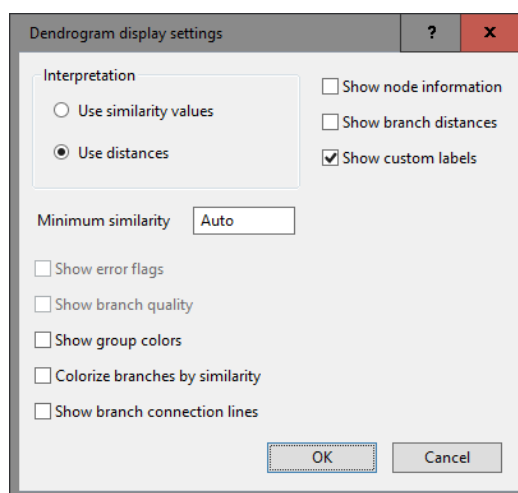The dendrogram branches are now colored according to the group colors (see Figure 6).

Figure 5: Dendrogram display settings.

The similarity values (here: distances) in the *Similarities* panel are represented by colors, based on the color scale displayed on top of the matrix.

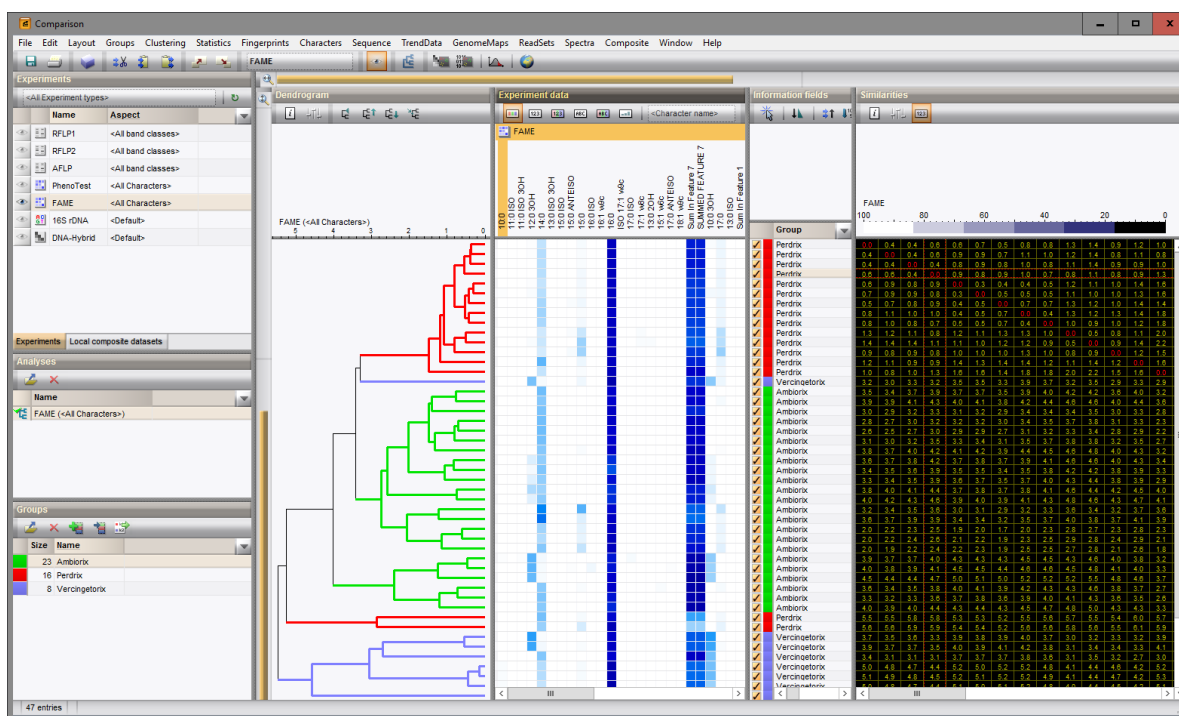21. To show the values in the matrix, select ***Clustering*** > ***Similarity matrix*** > ***Show values*** ([123]).

Figure 6: The dendrogram with group colors.

22. Select ***Clustering*** > ***Dendrogram display settings...*** ([⊞]) to call the *Dendrogram display settings* dialog box again.

23. Check the option ***Use similarity values*** to interpret the dendrogram in terms of similarity values.

24. Check the option ***Show node information*** and press <***OK***> to apply the new display settings.

The similarity values are now displayed both in the matrix in the *Similarities* panel and on each dendrogram node in the *Dendrogram* panel (see Figure 7).
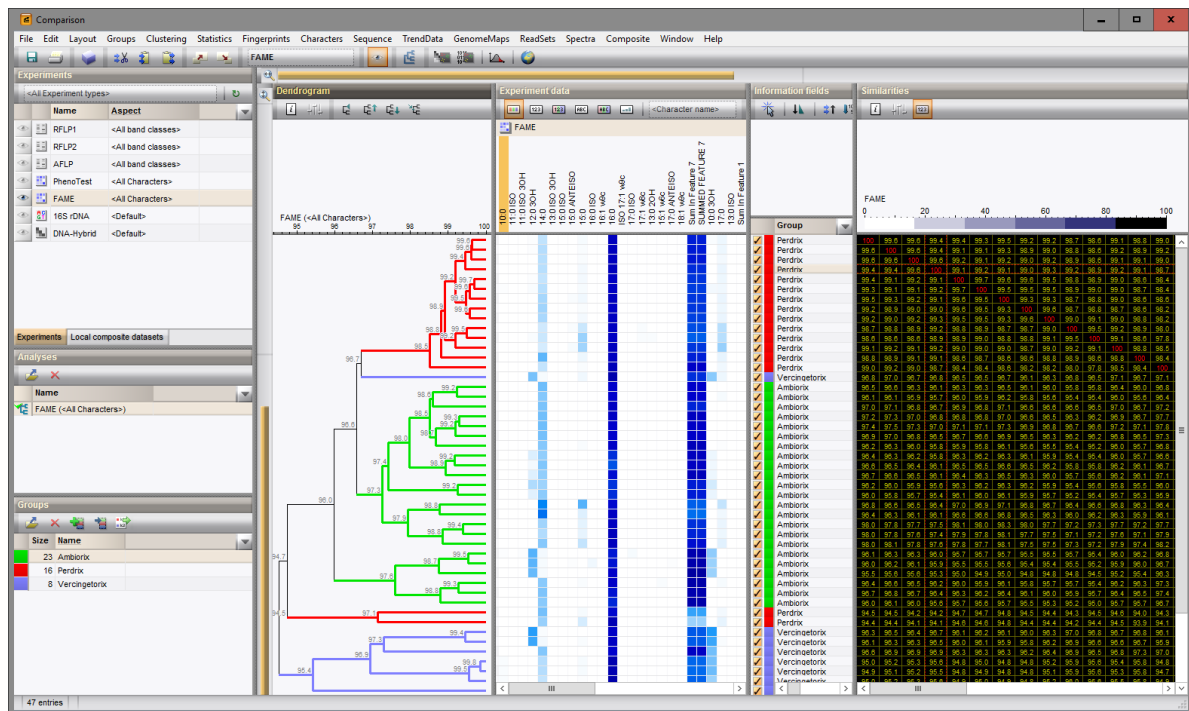
Figure 7: Similarity values.

25. Select ***Clustering*** > ***Dendrogram display settings...*** (⊞) to call the *Dendrogram display settings* dialog box again.

26. Uncheck the option ***Show node information*** and check the options ***Show branch distances*** and ***Colorize branches by similarity***. Press <***OK***> to apply the changes.

The branch lengths (***Show branch distances***) are displayed underneath each branch, and the branches are shaded according to their similarity value (***Colorize branches by similarity***) (see Figure 8).
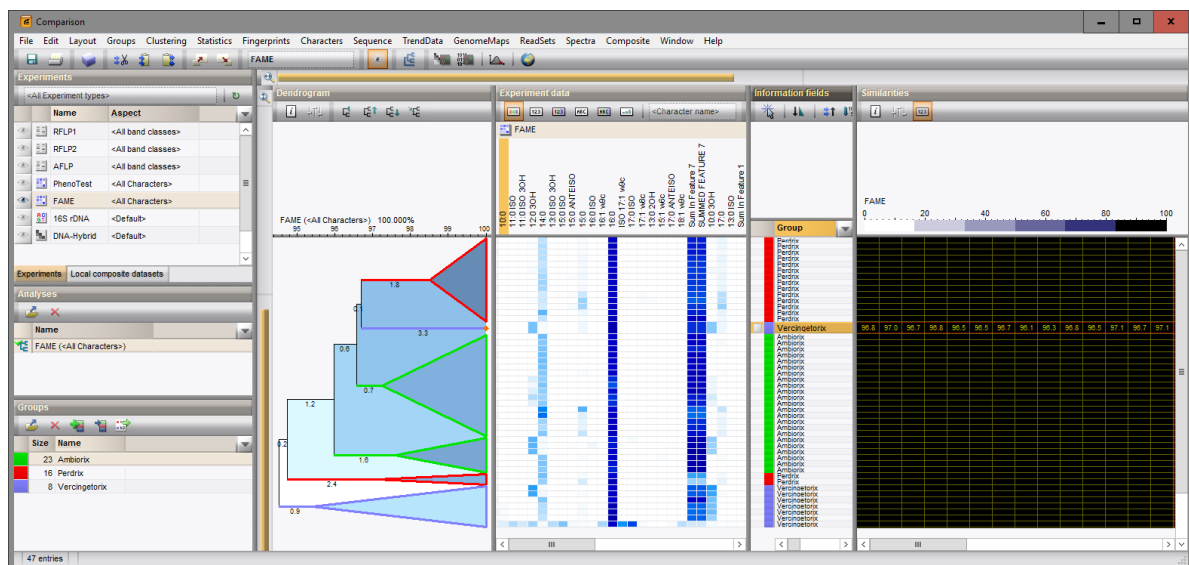
Figure 8: The *Comparison* window with similarity shades and abridged branches.

BioNumerics can export the cluster analysis as it appears in the *Comparison* window.

27. Select ***File*** > ***Print preview...*** (🖶, **Ctrl+P**) to call the *Comparison print preview* window.

More information about the *Comparison print preview* window can be found in the BioNumerics reference manual, accessible through **Help** > **Help on window**.

    28. Close the *Comparison print preview* window.

5   Advanced clustering window

The *Advanced cluster analysis* window can be launched from the *Comparison* window.

1. Make sure **FAME** is selected in the *Experiments* panel of the *Comparison* window.

2. Select **Clustering** > **Calculate** > **Advanced cluster analysis...** in the *Comparison* window or press the button and select **Advanced cluster analysis** to launch the *Create network* wizard.

3. Enter "FAME+UPGMA" as **Name** of the cluster analysis. Make sure **FAME (<All characters>)** is selected under **Choose an experiment**.

4. Under **Analysis template**, choose **No template** and press **<Next>**.

5. Select **Similarity matrix** and press **<Next>**.

6. Select **UPGMA** and press **<Next>** twice.

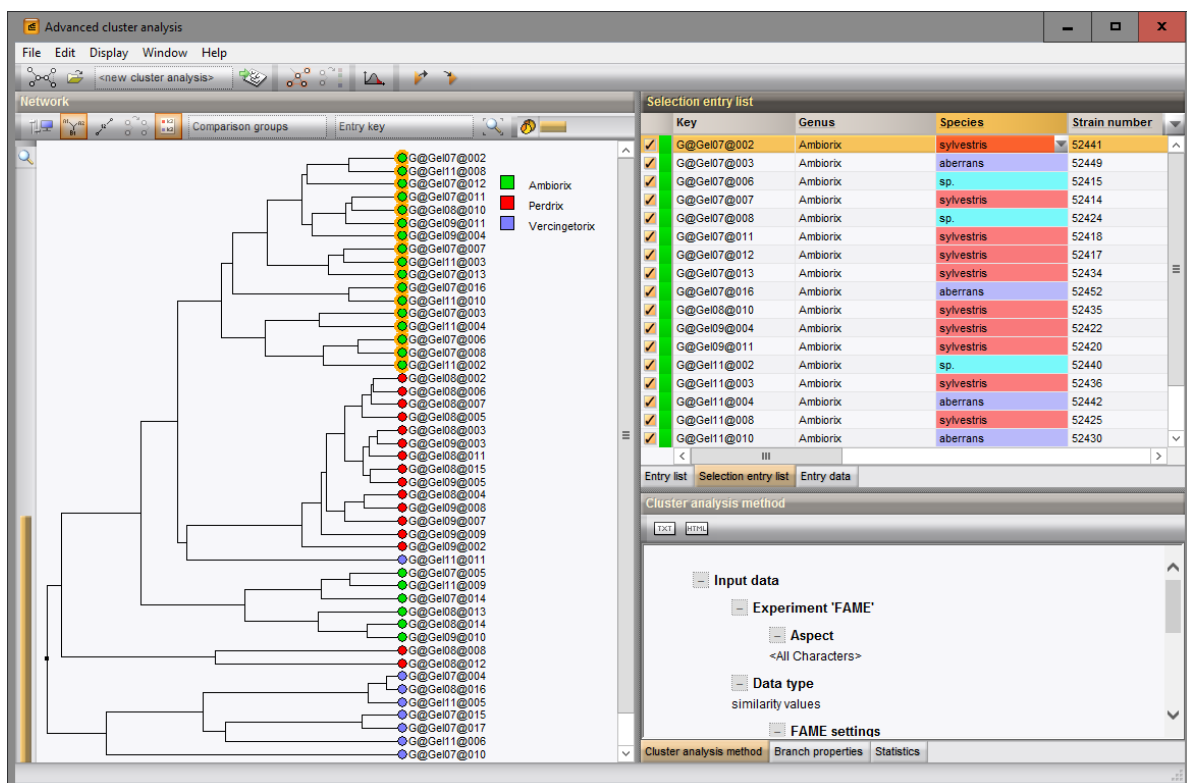The *Advanced cluster analysis* window pops up (see Figure 9).

Figure 9: The *Advanced cluster analysis* window.

The *Network panel* displays the UPGMA 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 the displayed tree.

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

8. The zoom slider on the left allows further zooming in or out on the network.

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

The rooted tree can be changed to an unrooted or circular tree via the *Change layout* dialog box.

10. Choose ***Display*** > ***Layout*** to call the *Change layout* dialog box (see Figure 10).
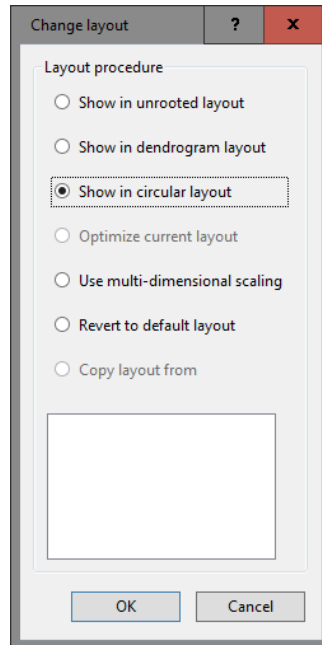

Figure 10: Change tree layout.

11. In this example, choose ***Show in circular layout*** and press <***OK***>.

The circular tree is now shown in the *Network panel*.

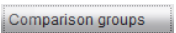12. Press ⊞ or choose ***Display*** > ***Display settings*** to call the *Display settings* dialog box.

The colors of the comparison groups are automatically shown as node colors and the entry keys are default used as node labels. These settings can be changed in the *Display settings* dialog box.

13. In the *Node labels and sizes* tab, choose ***Strain number*** from the ***Use label from*** drop-down list. Under ***Use color from*** choose ***Nothing*** for the moment (see Figure 11).

In the *Node colors* tab, the node coloring can be defined (see Figure 12). Default the colors from the ***Comparison groups*** are taken but this can be changed to a field state coloring, no coloring, or based on the number of entries.

14. In the *Node colors* tab, leave the ***Comparison groups*** selected as node coloring and press <***OK***> to apply the new settings.

The circular tree is updated accordingly (see Figure 13).

Changing the node labels and the node coloring is also possible using the drop-down lists in the toolbar of the *Network panel* (Entry key and Comparison groups respectively).

15. Use the horizontal zoom slider in the toolbar of the *Network panel* to adjust the node sizes.

Default, the node labels are shown next to the nodes.

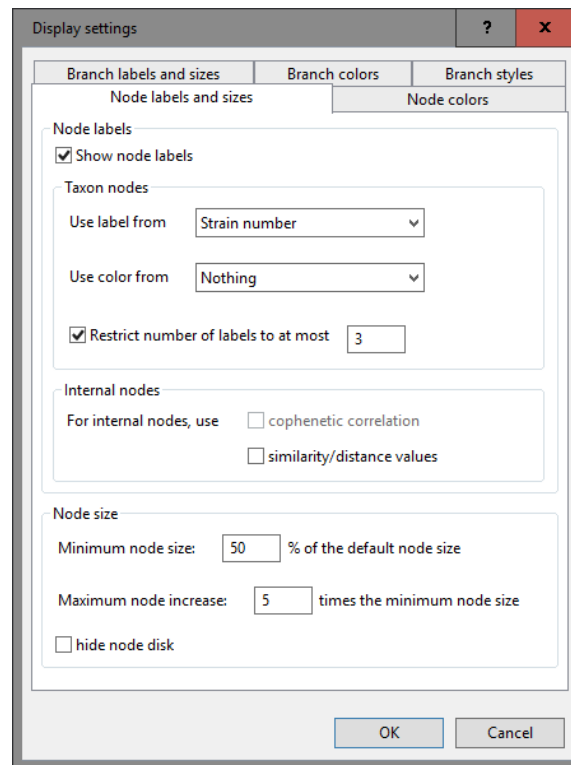16. Select ***Display*** > ***Show node labels on top of the nodes***.
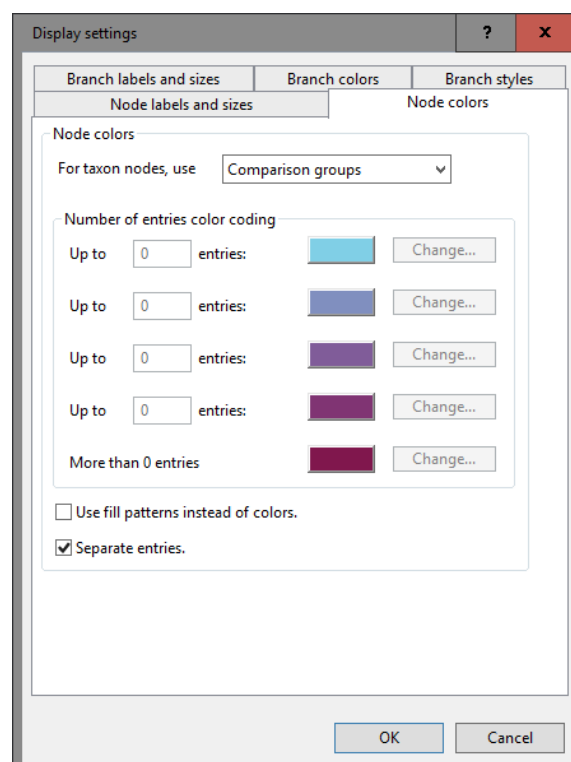
Figure 11: Node labels.

Figure 12: Nodes and node colors.

The node labels are now shown inside the nodes (see Figure 14).

17. Press ⊞ or choose **Display** > **Display settings** to open the *Display settings* dialog box again.
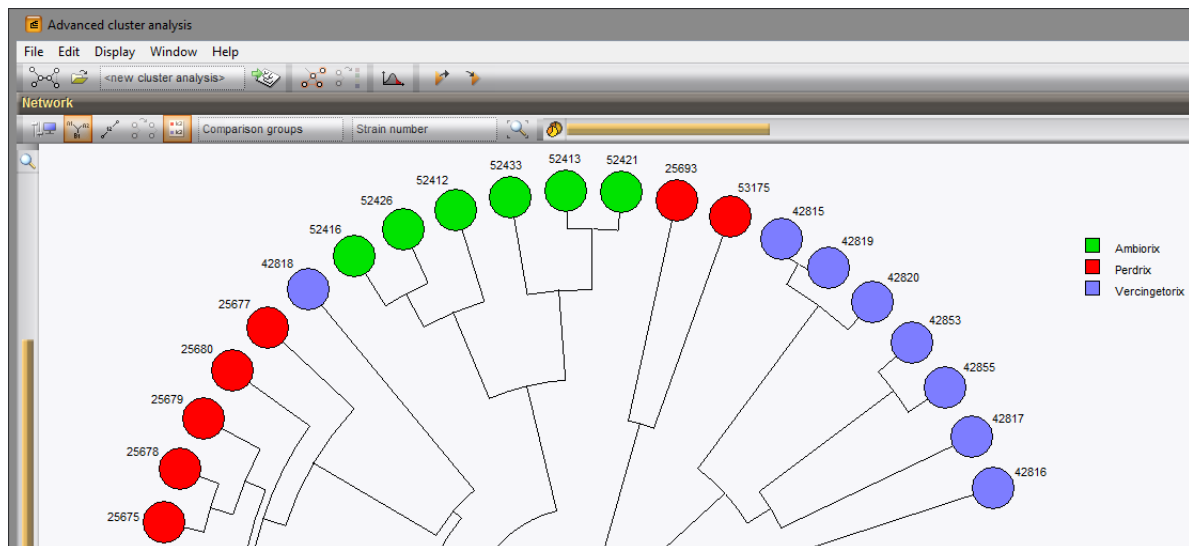
Figure 13: Drop-down lists in the *Network panel* toolbar.

18. In the *Branch labels and sizes* tab, check **Show branch labels** and **branch length** (checked by default) and press <**OK**>. Alternatively select **Display** > **Show branch labels**.

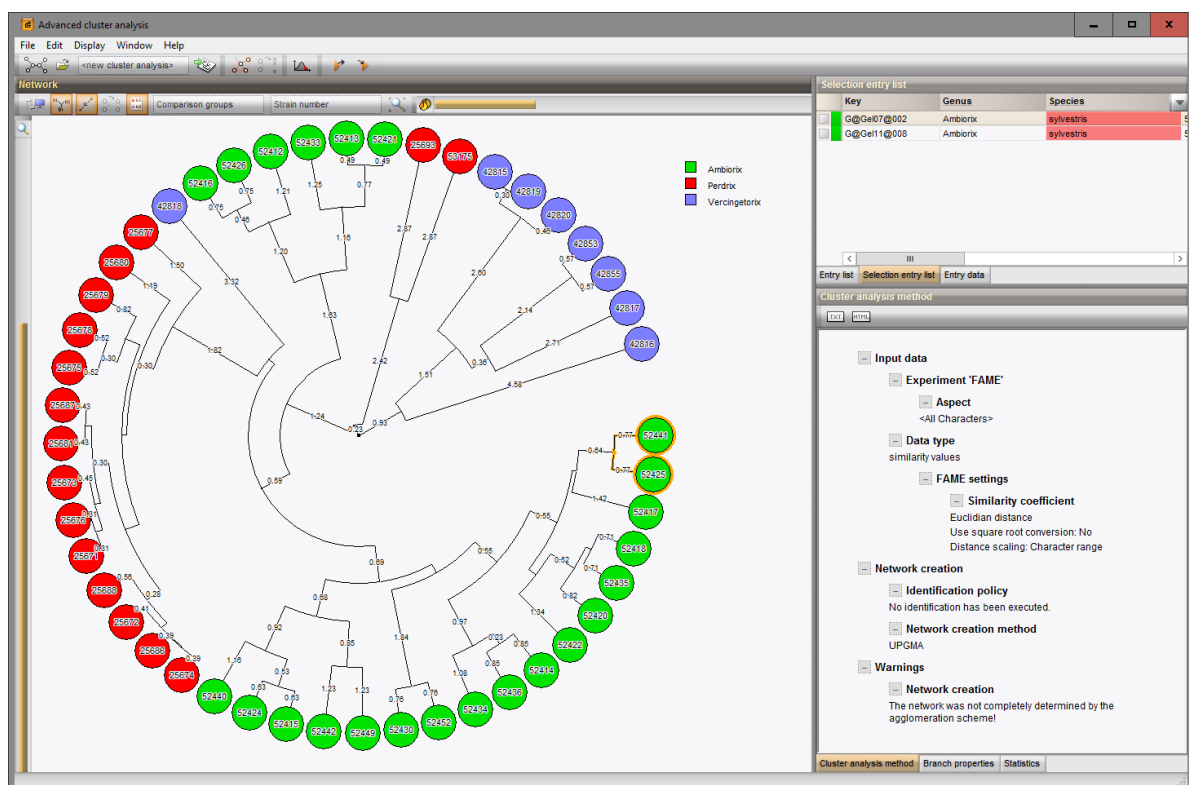The branch lengths are now displayed next to the branches (see Figure 14).

Figure 14: Branch lengths.

19. Select **Display** > **Show node labels next to the nodes** to display the node labels again next to the nodes.

20. Choose **Display** > **Display settings** to call the *Display settings* dialog box again.

21. In the *Node labels and sizes* tab, choose **Genus** from the **Use label from** drop-down list. Under **Use color from** choose **Comparison groups**. Press <**OK**> to apply the new settings (see Figure 15).

22. Select **Display** > **Show color legend** to hide or show the information about the node colors in the *Network panel*.
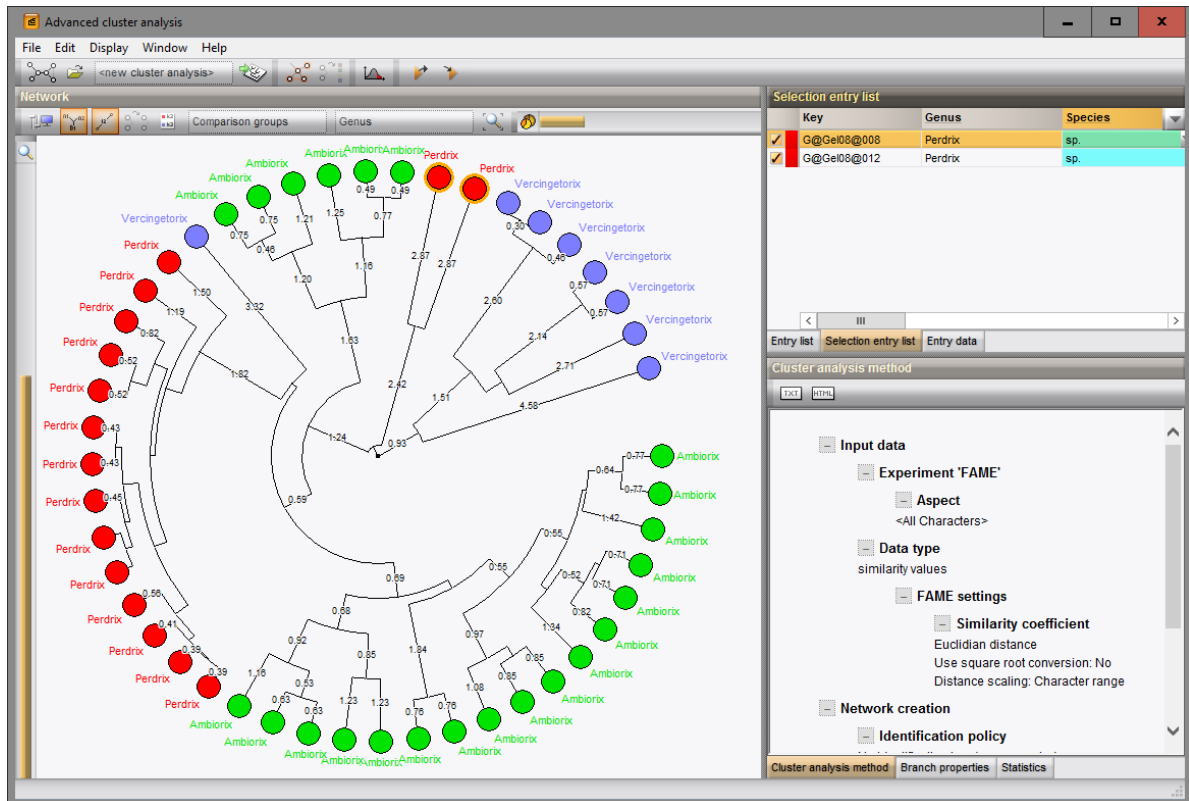

Figure 15: The *Advanced cluster analysis* window.

23. The image as it appears in the *Network panel* can be exported with **File** > **Export image**.

More information about the different export options can be found in the BioNumerics reference manual.

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