

Dendrogram Tools plugin

PLUGINS
VERSION 7.6



Contents

1	Starting and setting up BioNumerics	3
1.1	Startup program	3
1.2	Installing the Dendrogram Tools plugin	4
2	Dendrogram tools functionality	5
2.1	Auto collapse branches	5
2.2	Auto cutoff branches	5
2.3	Export a dendrogram in Newick format	6
2.4	Fill field with cluster number	6
2.5	Find best approximating branch	7
2.6	Rearrange dendrogram to match other	8
2.7	Find cluster in other dendrogram	9
2.8	Similarity – Cluster size plot	9
2.9	Copy group names to information field	10

NOTES

SUPPORT BY APPLIED MATHS

While the best efforts have been made in preparing this manuscript, no liability is assumed by the authors with respect to the use of the information provided.

Applied Maths will provide support to research laboratories in developing new and highly specialized applications, as well as to diagnostic laboratories where speed, efficiency and continuity are of primary importance. Our software thanks its current status for a part to the response of many customers worldwide. Please contact us if you have any problems or questions concerning the use of BioNumerics[®], or suggestions for improvement, refinement or extension of the software to your specific applications:

Applied Maths NV

Keistraat 120
9830 Sint-Martens-Latem
Belgium
PHONE: +32 9 2222 100
FAX: +32 9 2222 102
E-MAIL: info@applied-maths.com
URL: <http://www.applied-maths.com>

Applied Maths, Inc.

11940 Jollyville Road, Suite 115N
Austin, Texas 78759
U.S.A.
PHONE: +1 512-482-9700
FAX: +1 512-482-9708
E-MAIL: info-US@applied-maths.com

LIMITATIONS ON USE

The BioNumerics[®] software, its plugin tools and their accompanying guides are subject to the terms and conditions outlined in the License Agreement. The support, entitlement to upgrades and the right to use the software automatically terminate if the user fails to comply with any of the statements of the License Agreement. No part of this guide may be reproduced by any means without prior written permission of the authors.

Copyright ©1998, 2018, Applied Maths NV. All rights reserved.

BioNumerics[®] is a registered trademark of Applied Maths NV. All other product names or trademarks are the property of their respective owners.

BioNumerics[®] uses following third-party software tools and libraries:

- The Python[®] 2.7.4 release from the Python Software Foundation (<http://www.python.org/>).
- A library for XML input and output from the Apache Software Foundation (<http://www.apache.org>).
- NCBI toolkit version 2.2.10 (<http://www.ncbi.nlm.nih.gov/BLAST/>).
- The Boost c++ libraries (<http://www.boost.org/>).
- Samtools for interacting with SAM / BAM files (<http://www.htslib.org/download/>)
- The 7-Zip command line version (7za.exe) from 7-Zip, copyright 1999-2010 Igor Pavlov. <http://www.7-zip.org/>
- Velvet for Windows, source code can be downloaded from <http://www.applied-maths.com/download/open-source>.
- Ray for Windows, source code can be downloaded from <http://www.applied-maths.com/download/open-source>.
- Mothur for Windows, source code can be downloaded from <http://www.applied-maths.com/download/open-source>.
- Cairo 2D graphics library version 1.12.14 (<http://cairographics.org/>).
- Crypto++ Library version 5.5.2 (<http://www.cryptopp.com/>).
- libSVM library for Support Vector Machines (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>).
- SQLite version 3.7.17 (<http://www.sqlite.org/>).
- Gecko engine version 21 (<https://developer.mozilla.org/en-US/docs/Mozilla/Gecko>).
- pymzML Python[®] module for high throughput bioinformatics on mass spectrometry data (<https://github.com/pymzml/pymzML>).
- Numpy Python[®] library version 1.8.1 (<http://www.numpy.org/>).
- BioPython Python[®] library version 1.64 (<http://www.biopython.org/>).
- PIL Python library[®] version 1.1.7 (<http://www.pythonware.com/products/pil/>).
- The SPAdes genome assembler version 3.7.1 (<http://bioinf.spbau.ru/spades>).

Chapter 1

Starting and setting up BioNumerics

1.1 Startup program

When BioNumerics is launched from the Windows start panel or when the BioNumerics shortcut () on your computer's desktop is double-clicked, the **Startup program** is run. This program shows the *BioNumerics Startup* window (see Figure 1.1).

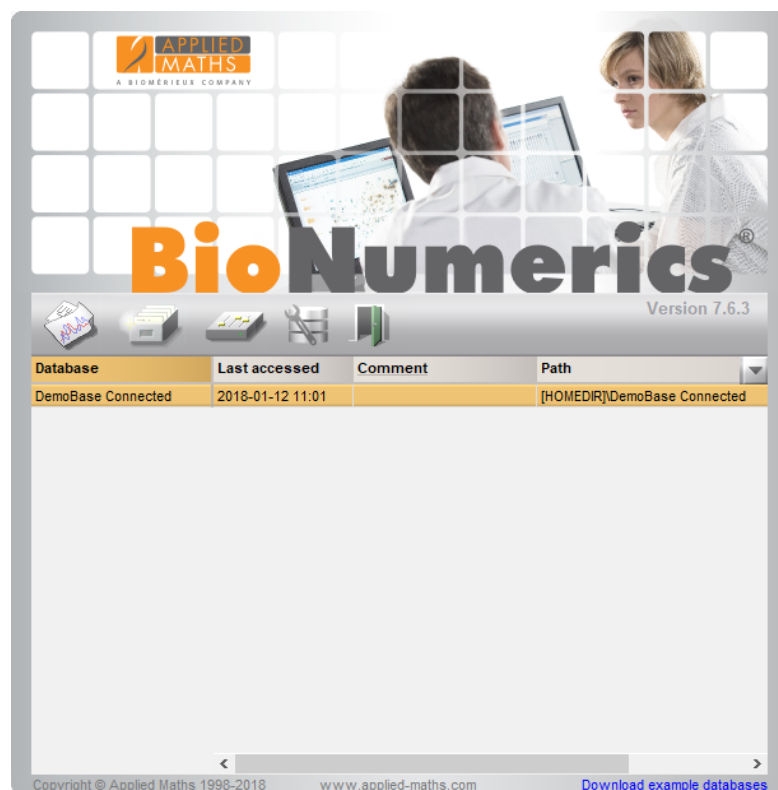


Figure 1.1: The *BioNumerics Startup* window.

A new BioNumerics database is created from the Startup program by pressing the  button.

An existing database is opened in BioNumerics with  or by simply double-clicking on a database name in the list.

1.2 Installing the Dendrogram Tools plugin

If a database is opened for the first time, the *Plugins* dialog box will appear by default (see Figure 1.2).

If the database has already been opened previously, the *Plugins* dialog box can be called from the *Main* window by selecting **File > Install / remove plugins...** (🔧).

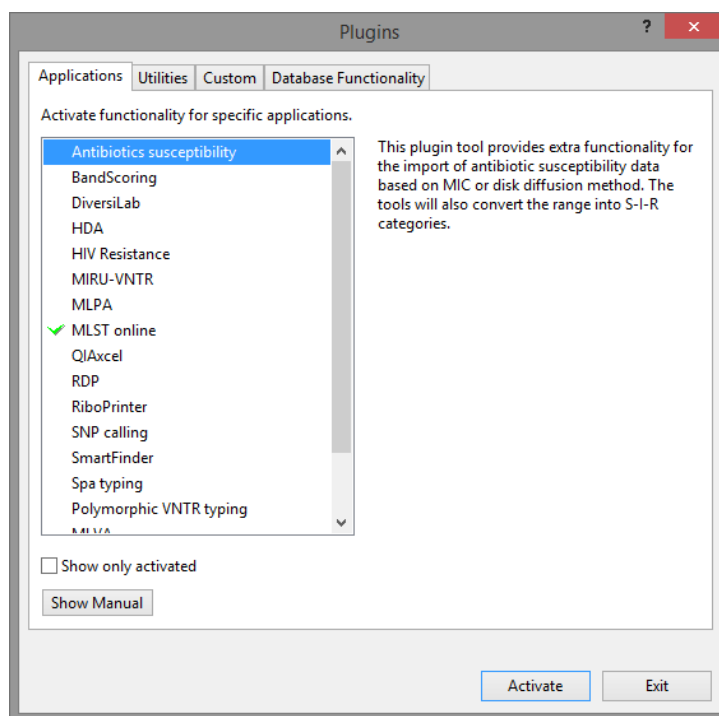


Figure 1.2: The *Plugins* dialog box.

When a particular plugin is selected from the list of plugins, a short description appears in the right panel.

A selected plugin can be installed with the **<Activate>** button. The software will ask for confirmation before installation. Some plugins depend on functionality offered by specific BioNumerics modules. If a required module is missing, the plugin cannot be installed and an error message will be generated.

Once a plugin is installed, it is marked with a green V-sign. It can be removed again with the **<Deactivate>** button.

If the selected plugin is documented, pressing **<Show Manual>** will open its manual in the *Help* window.

- 2.1 To install the *Dendrogram tools plugin* in your database, select the *Utilities tab* and select the *Dendrogram tools plugin* from the list of plugins.
- 2.2 Press the **<Activate>** button, confirm the installation of the plugin and close the *Plugins* dialog box.
- 2.3 Close and reopen the database to activate the features of the *Dendrogram tools plugin*.

The *Dendrogram tools plugin* installs menu items in the *Comparison* window under the **File** and **Clustering > Advanced tools** menus.

Chapter 2

Dendrogram tools functionality

2.1 Auto collapse branches

The **Auto collapse branches tool** can be used to have the branches of a tree automatically displayed as *abridged*.

- 1.1 If a dendrogram does not yet exist in the *Comparison* window, perform a cluster analysis on an experiment for which data is linked to the entries in the comparison.
- 1.2 The *Auto collapse branches* dialog box is called with **Clustering > Advanced tools > Auto collapse branches** (see Figure 2.1).

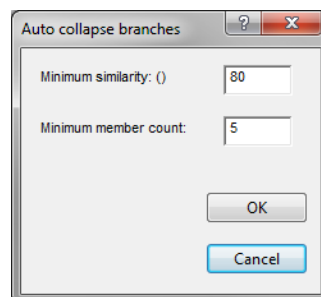


Figure 2.1: The *Auto collapse branches* dialog box.

Specify the **Minimum similarity** and enter the minimum number of entries a branch should contain (**Minimum member count**).

When pressing the **<OK>** button, the branches that meet the specified criteria are abridged and are displayed as a triangle in the *Dendrogram* panel.

2.2 Auto cutoff branches

The **Auto cutoff branches tool** can be used to display the clusters that appear below a certain similarity level in a light gray color. This function can be useful to distinguish insignificant clusters from significant clusters.

- 2.1 If a dendrogram does not yet exist in the *Comparison* window, perform a cluster analysis on an experiment for which data is linked to the entries in the comparison.
- 2.2 The *Cut off clusters* dialog box is called with **Clustering > Advanced tools > Auto cut off branches** (Figure 2.2).

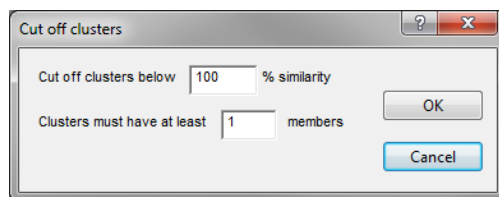


Figure 2.2: The *Cut off clusters* dialog box.

Specify a *Cut off clusters similarity value* at which level the clusters will be cut off.

Enter a *Minimum number of members* a cluster should have in order to be cut off.

When pressing the <OK> button, the clusters that meet the cut off criteria are displayed in a light gray color.

2.3 Export a dendrogram in Newick format

With the **Export Dendrogram (Newick) tool** a dendrogram can be exported as a text file in Newick format. The Newick format contains both the structure of the dendrogram and the lengths of the branches. This common format can be imported into many tree-processing applications.

- 3.1 If a dendrogram does not yet exist in the *Comparison* window, perform a cluster analysis on an experiment for which data is linked to the entries in the comparison.
- 3.2 The export dialog is called with **Clustering > Advanced tools > Export dendrogram (Newick)** (see Figure 2.3).

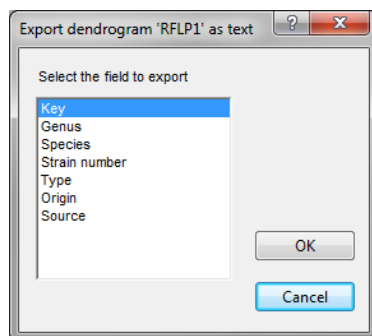


Figure 2.3: The *Export dendrogram as text* dialog box.

The dialog asks to choose the *Field to export*.

2.4 Fill field with cluster number

The **Fill field with cluster number tool** can be used to assign numbers to clusters that appear above a certain similarity level.

- 4.1 If a dendrogram does not yet exist in the *Comparison* window, perform a cluster analysis on an experiment for which data is linked to the entries in the comparison.
- 4.2 The *Fill field with cluster number* dialog box is called with **Clustering > Advanced tools > Fill field with cluster number** (see Figure 2.4).

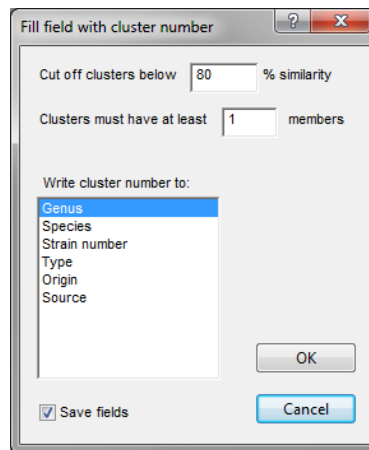


Figure 2.4: The *Fill field with cluster number* dialog box.

Specify a similarity value at which level the dendrogram will be split up into clusters (*Cut off similarity value*).

Enter a *Minimum number of members* a cluster should have in order to be counted.

Select a field to *Write the cluster number to*.

By checking *Save fields*, the field containing the cluster number is permanently saved.

When pressing the <OK> button, the obtained clusters are numbered, and the cluster numbers are filled into the field that was specified.

2.5 Find best approximating branch

The **Find best approximating branch tool** can be used to compare two dendrograms in a comparison.

5.1 Perform at least two cluster analyses in the *Comparison* window.

5.2 Select a branch on the dendrogram to start from and to make interpretation of the result easier, select the entries of that branch (**Ctrl+click** on the branch).

5.3 The *Find best matching branch* dialog box is called with **Clustering > Advanced tools > Find best approximating branch** (see Figure 2.5).

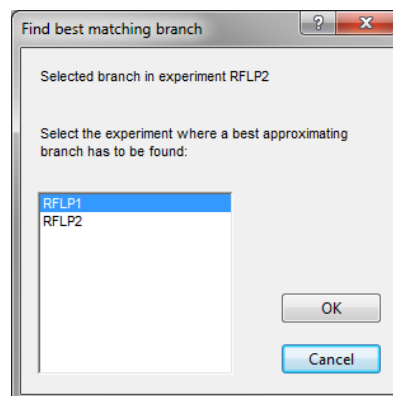


Figure 2.5: The *Find best matching branch* dialog box.

Select the dendrogram where a best approximating branch has to be found and press **<OK>** to call the *Set penalties* dialog box.

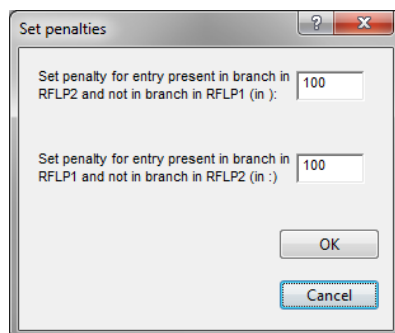


Figure 2.6: The *Set penalties* dialog box.

In the *Set penalties* dialog box, non-exactly matching branches can be dealt with by allowing relative **Penalties** (in percentage) to be set for entries present in the branch of the first tree but not in the branch of the second tree, and vice versa (see Figure 2.6).

When pressing the **<OK>** button, the second tree is displayed in the *Dendrogram* panel and the best approximating branch is selected on the tree.

2.6 Rearrange dendrogram to match other

The **Rearrange dendrogram to match other tool** facilitates the interpretation and comparison of two dendrograms, by rearranging the branches of the current dendrogram so that the order of entries approaches the order of entries in the other dendrogram as closely as possible.

6.1 Perform at least two cluster analyses in the *Comparison* window.

6.2 The *Re-arrange dendrogram* dialog box is called with **Clustering > Advanced tools > Rearrange dendrogram to match other** (see Figure 2.7).

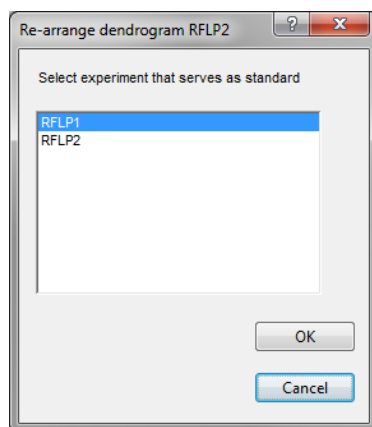


Figure 2.7: Rearrange dendrogram based on another dendrogram.

Select the cluster analysis that will serve as a **standard** to rearrange the tree. Obviously, the rearrangement can only be done when at least two clusterings are present.

When pressing the **<OK>** button, the branches of the dendrogram are rearranged to resemble to other dendrogram as closely as possible.

2.7 Find cluster in other dendrogram

The **Find cluster in other dendrogram tool** facilitates the interpretation and comparison of clusters in two dendrograms.

7.1 Perform at least two cluster analyses in the *Comparison* window.

7.2 Select a cluster on the dendrogram to start from.

7.3 A new dialog is called with **Clustering > Advanced tools > Find cluster in other dendrogram** (see Figure 2.8).

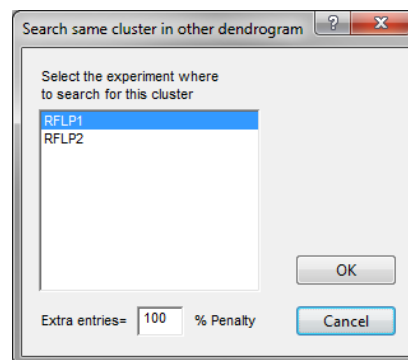


Figure 2.8: Search for same cluster in other dendrogram.

A list of all available dendrograms is shown in a new dialog box.

A cost (**Penalty**) can be specified for extra entries that are found in the matching cluster. When a cost of 100% is specified, an extra entry is penalized as heavily as a missing entry. When 0% is specified, no cost is given to extra entries, and the root of the other dendrogram will always be selected.

When pressing the **<OK>** button, the entries of the chosen branch are selected and the dendrogram of the cluster analysis is displayed with the closest matching branch selected.



Any other selection of entries you made in the comparison will get lost when running this tool.

2.8 Similarity – Cluster size plot

The **Similarity - cluster size plot tool** can be used to get an idea about the range of similarity most clusters fall in.

8.1 If a dendrogram does not yet exist in the *Comparison* window, perform a cluster analysis on an experiment for which data is linked to the entries in the comparison.

8.2 The *Charts and statistics* window is called with **Clustering > Advanced tools > Similarity - cluster size plot** (see Figure 2.9).

A plot is created with the cumulative number of the clusters plotted against the similarity scale.

More information about the *Charts and statistics* window can be found in the Reference manual, Part Charts.

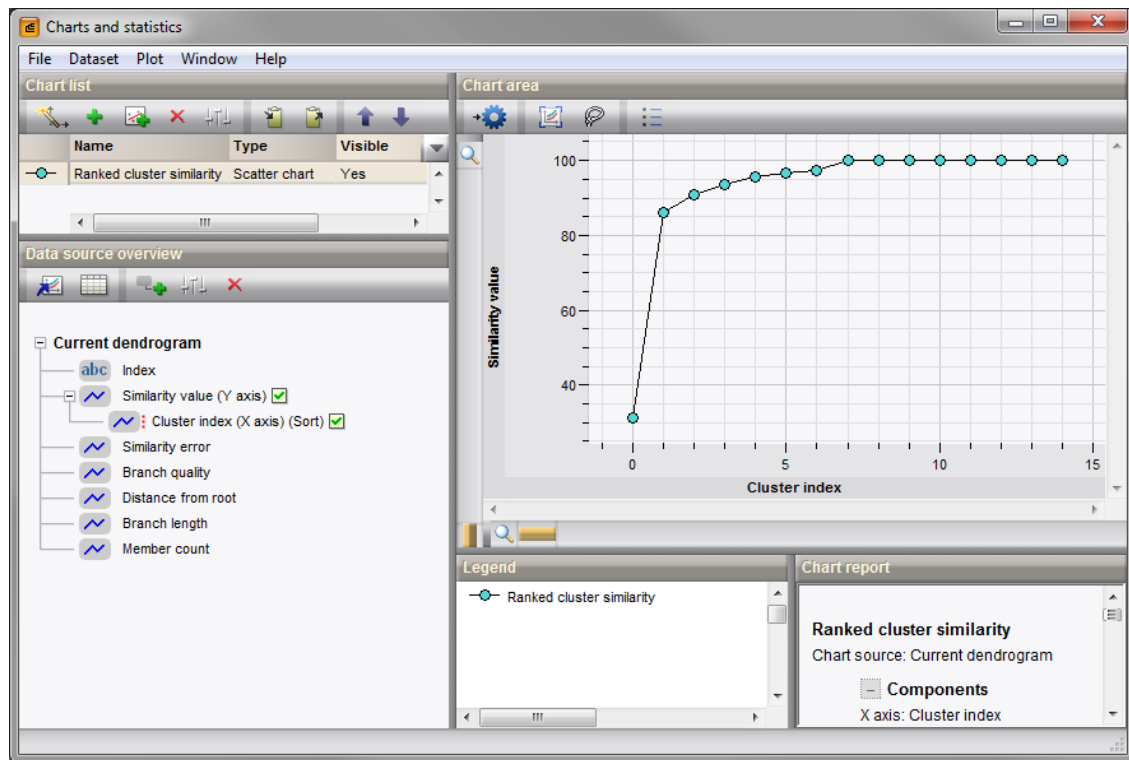


Figure 2.9: The *Charts and statistics* window.

2.9 Copy group names to information field

The **Copy group names to information field tool** can be used to fill the group names into an information field appointed by the user.

- 9.1 If groups do not yet exist in the *Comparison* window, assign groups to the entries as specified in the Reference manual, Chapter General comparison functions.
- 9.2 The *Copy group names to info field* dialog box is called with **Groups > Copy group names to information field** (see Figure 2.10).

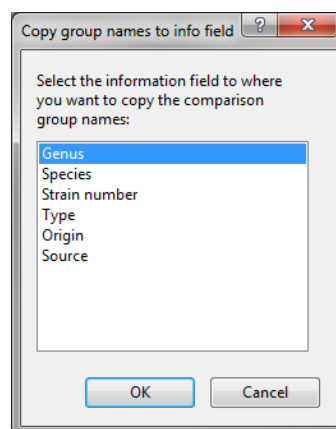


Figure 2.10: The *Copy group names to info field* dialog box.

Select a field to **Write the group names to**.

When pressing the **<OK>** button, the group numbers are filled into the field that was specified.



A B I O M É R I E U X C O M P A N Y

Copyright 1998-2018, Applied Maths NV. All rights reserved.

Please contact us for any additional information you might require, we will gladly help you!

Headquarters

📍 Keistraat 120 • 9830 Sint-Martens-Latem • Belgium
☎ +32 922 22 100 ✉ info@applied-maths.com

USA and Canada

📍 11940 Jollyville Rd., Suite 115N • Austin, TX 78750 USA
☎ +1 512 482 9700 ✉ info-us@applied-maths.com