

User Manual for Hybroscale

by Benjamin Albrecht

November 6, 2015

Contents

1	Introduction	2
2	Program Overview	2
3	Obtaining and Installing the Program	2
4	Getting started	2
5	Main Window	3
5.1	File Menu	3
5.2	Edit Menu	4
5.3	Algorithms Menu	4
5.4	View Menu	5
6	File Formats	6
6.1	Extended Newick Format	6
6.2	Hybroscale Files	6
7	Additional Windows	6
7.1	Computing Hybridization Networks	6
7.2	Showing the Embedding of Trees	7
8	More Memory	7
9	Command Line	7
10	Short Example	8

1 Introduction

This write-up gives a description of how to use the software package Hybroscale. If you have any questions or ideas how to improve the program, please do not hesitate and write a mail to albrecht@bio.ifi.lmu.de.

Type-setting convention: Throughout this manual we use $A \mapsto B$ to indicate the location of menu item B in menu A. For example by $FILE \mapsto LOAD$ we refer to the menu item LOAD being contained in the FILE menu, which can be found at the top of the main window.

How to cite: When using Hybroscale please cite the following paper:

Benjamin Albrecht. Computing all hybridization networks for multiple binary phylogenetic input trees. BMC Bioinformatics, vol. 16, no. 236, pp. 439-441, July 2015.

2 Program Overview

Hybroscale is developed specifically for the research of certain kinds of parsimonious phylogenetic networks, i.e. minimum hybridization networks, including its computation and visualization. The software is freely available (see Sec. 3) and runs on all three major operating systems. Its handling is similar to the software package *Dendroscope* [1], a well known software package for visualizing phylogenetic trees and rooted networks. In the following we briefly describe its main features, which are discussed again in more detail in Section 5:

- The program enables loading and visualizing rooted trees given in *Extended Newick* format (cf. Sec.6.1).
- Given a set of rooted input trees sharing an overlapping set of taxa, the program enables an efficient computation of **all** networks with minimum hybridization number displaying each input tree (cf. Sec.5.3). Until now, this algorithm is unique regarding its functionality and efficiency.
- Having computed a set of minimum hybridization networks, the program offers a special visualization mode (see $VIEW \mapsto HYBRIDVIEW$) showing the embedding of all initial input trees by introducing edges of particular colors (cf. Sec.7.2).
- A network filter, selecting networks fulfilling specific topological constraints, can be applied in order to test well-directed hypothesis.

3 Obtaining and Installing the Program

Hybroscale is offered as runnable jar file requiring Java 1.7 or higher and runs on all three major operating systems. It is freely available from

<http://www.bio.ifi.lmu.de/software/services/hybroscale>

4 Getting started

You can run the graphical user interface of the program by just typing

```
java -jar hybroscale.jar
```

without any extra parameters. To run the program via command line please see Section 9. Note that for complex input trees, it helps to increase the available amount of working memory, which is briefly described in Section 8.

5 Main Window

The main window of Hybroscale contains several tab-separated *Viewer*, which are again separated as follows: On the left hand side two data tables show all trees and networks that have been added or computed so far. On the right hand side you can find a grid panel for visualizing trees and networks which can be done intuitively by first selecting a panel and then by choosing one of the entries of the two data tables. At the top of the main window you can find all menus that are now discussed in detail in the upcoming part of this section.

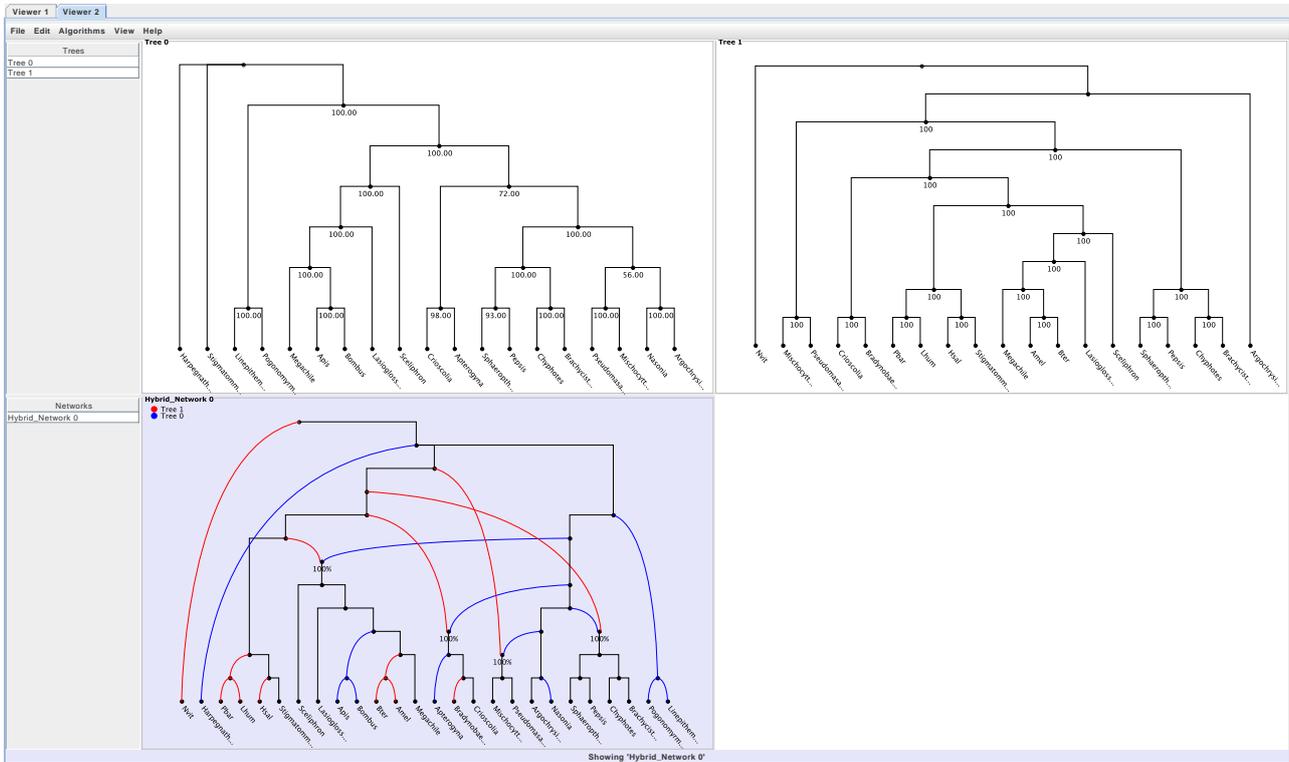


Figure 1: Main window of Hybroscale: On the left hand side two data tables show all trees and networks that have been loaded or calculated so far. On the right hand side you can find a 2x2 grid containing the visualization of tow trees and one network.

5.1 File Menu

The FILE menu provides several items for saving and loading trees and networks:

- FILE \mapsto NEW VIEWER
Creates a new Viewer providing new empty data tables and an empty grid panel.
- FILE \mapsto ENTER TREES/NETWORKS
Opens a window where you can enter trees or networks given in *Extended Newick* format (cf. Sec. 6.1). Please note that only one string per line is accepted.
- FILE \mapsto LOAD TREES/NETWORKS
Opens a window and automatically pastes in the content of the file that has been selected. Again, please note that only one string in *Extended Newick* format (cf. Sec. 6.1) per line is accepted.
- FILE \mapsto LOAD RECENT TREES/NETWORKS
Opens a window and automatically pastes in the content of the respective recently loaded file.
- FILE \mapsto LOAD VIEWER
Restores the content of the saved Viewer including all trees and networks as well as the mapping of a set of trees back to its computed networks what is necessary for showing its embedding via VIEW \mapsto HYBRID LAYOUT.

- **FILE ↦ LOAD RECENT VIEWER**
Restores the content of a recently saved Viewer.
- **FILE ↦ SAVE TREES/NETWORKS**
Saves certain trees or networks to an arbitrary file. You can choose between ...
 - saving all trees or networks (↦ SAVE ALL),
 - just all trees (↦ SAVE ALL TREES),
 - just all networks (↦ SAVE ALL NETWORKS),
 - and just those trees or networks whose panel is selected (↦ SAVE ALL SELECTED PANELS).
- **FILE ↦ SAVE VIEWER**
Saves the whole content of the Viewer including all trees and networks as well as the mapping of a set of trees back to its calculated networks which is necessary for showing its embedding via VIEW ↦ HYBRID LAYOUT.
- **FILE ↦ EXPORT PANEL**
Exports each selected panel into a single file. You can choose between the three well known image formats *pdf*, *png*, and *jpg*.
- **FILE ↦ EXPORT NEWICK-STRING**
Opens a window displaying each *Extended Newick* string (cf. Sec. 6.1) of all selected trees or networks.
- **FILE ↦ CLOSE VIEWER**
Closes the current Viewer.
- **FILE ↦ CLOSE OTHER VIEWERS**
Closes all viewers except the one which is displayed.
- **FILE ↦ QUIT**
Quits the program. Please get sure that all important trees and networks have been saved.

5.2 Edit Menu

The EDIT menu enables the selection of panels within the grid:

- **EDIT ↦ SELECT ALL PANELS**
Selects all panel within in the grid.
- **EDIT ↦ SELECT NO PANELS**
Deselects all panels within in the grid.
- **VIEW ↦ ADD**
Enables the adding of all trees and networks, just all trees, or just all networks to the grid panel. Please note that before the adding all other trees and networks within the grid are removed.
- **VIEW ↦ EXPLORE**
Opens a new window showing all trees and networks, just all trees, or just all networks. You can scroll through all loaded trees or networks by pressing the arrow buttons on your keyboard. Note that the new window again contains a subset of all menu items allowing to change the grid dimension and the visualization of the trees and networks.

5.3 Algorithms Menu

The ALGORITHM menu contains the main features of Hybroscale. It enables running a sophisticated algorithm calculating all minimum hybridization networks for an arbitrary number (at least two) of rooted non-binary trees sharing an overlapping set of taxa. For further details please see Section 7.1.

- **ALGORITHMS ↦ HYBRIDIZATION NETWORKS**
Computes all hybridization networks of minimum reticulation number. As input an arbitrary number (at least two) of multifurcating trees sharing an overlapping set of taxa is required.

- **ALGORITHMS ↦ HYBRIDIZATION NUMBER**
Computes just the hybridization number of an arbitrary number (at least two) of multifurcating trees sharing an overlapping set of taxa.
- **ALGORITHMS ↦ SHOW GRAPH PROPERTIES**
Computes topological features of the chosen trees and networks like the time-consistency degree (TC-Degree), the add-taxa degree (AT-Degree), and its level. The features are written in the top left corner of the respective panel.
- **ALGORITHMS ↦ HIDE GRAPH PROPERTIES**
Hides all topological features of the chosen trees and networks.

5.4 View Menu

The VIEW menu contains several items for changing the visualization of trees and networks:

- **VIEW ↦ SET GRID**
Enables the setting of grid dimension.
- **VIEW ↦ FILTER ALL NETWORKS**
Allows to explore all networks fulfilling certain constraints as follows: In a first step, a window for creating certain constraints is shown. Secondly, a window is opened showing all networks fulfilling these constraints. You can scroll through all these networks by either using your mouse wheel or by pressing the arrow buttons on your keyboard. Please note that this window contains a subset of all menu items allowing to change the grid dimension and the visualization of the networks.
- **VIEW ↦ SHOW RETICULATIONS STATISTICS**
Shows the frequency of taxa sets occurring as direct descendants of reticulation nodes.
- **VIEW ↦ RECOMPUTE LAYOUT**
Recomputes the layout of each selected tree and network.
- **VIEW ↦ HYBRID LAYOUT**
Opens a window that allows to visualize the embedding of certain trees within the selected network (cf. Sec. 7.2). Please note that this only works for those networks that have been computed by the program before (cf. Sec. 5.3). Otherwise, you will receive a message pointing out that there is no *hybrid information* available.
- **VIEW ↦ SHOW NODE LABELS**
Shows all node labels of inner nodes.
- **VIEW ↦ SHOW EDGE LABELS**
Enables the showing of certain edge labels.
- **VIEW ↦ HIDE NODE LABELS**
Hides all node labels of inner nodes.
- **VIEW ↦ HIDE EDGE LABELS**
Enables the hiding of certain edge labels.
- **VIEW ↦ INCREASE FONT SIZE**
Increases the font size of each label within the selected graphs.
- **VIEW ↦ DECREASE FONT SIZE**
Decreases the font size of each label within the selected graphs.

6 File Formats

Hybroscale can handle the following two file formats:

6.1 Extended Newick Format

Hybroscale uses the *Extended Newick* format for loading and saving trees and networks initially proposed by [2]. This format describes a tree with the same bracket notation as used in the standard Newick format. In this case, however, a node can contain three different kinds of information: The first part always refers to the node label. The following part either is surrounded by squared brackets, what refers to the label of its incoming edge, or to a character beginning with #H followed by a number denoting a part that is either the source or the target node of a reticulate edge. An illustration of an *Extended Newick* string is given in Figure 2.

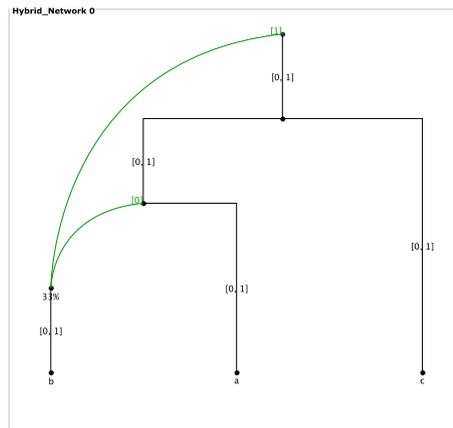


Figure 2: Visualization of the network referring to the following *Extended Newick* string:

```
((c[[0, 1]],(a[[0, 1]],(b[[0, 1]])33%#H0[[0]])[[0, 1]])[[0, 1]],33%#H0[[1]]);
```

6.2 Hybroscale Files

A file ending on `.hybroscale` provides all information to restore a recent *Viewer* including its trees and networks as well as the mapping of tree sets to certain networks which is necessary for generating the Hybrid Layout (cf. Sec. 5.4).

7 Additional Windows

7.1 Computing Hybridization Networks

This additional window pops up for starting the computation of all hybridization networks (cf. Sec. 5.3). It provides several functions depending on the status of the algorithm:

Before pressing the Run Button you can set several parameters as follows:

- **Add Constraints Button**
Allows to set certain constraints each reported network has to fulfill. Note that setting these constraints can limit the search space and, thus, enables a faster execution of the algorithm.
- **Set Cores Button**
Sets the number of cores that can be used by Hybroscale to run exhaustive parts of the algorithm in parallel.
- **Close Button**
Closes the window.

- **Run Button**
Starts the execution of the algorithm.

After pressing the **Run Button** each row of the table shows the progress of a specific thread. Please note that the larger the progress the more complex the algorithm gets and, thus, more runtime for gaining progress is necessary. However, a thread can finish at an arbitrary value meaning that it can suddenly jump to 100 percent. During the execution of a thread the user has the following options:

- **Abort Button**
The execution of the corresponding thread is aborted and, if no sub-network has been computed so far, the corresponding part of each resulting network is replaced by a cluster network.
- **Details Button**
Details about the status of its execution is reported. For example you can see the number of networks that has been computed so far.
- **Cancel Button**
If no final network has been computed so far, the computation is aborted without reporting any results. Otherwise, all so far computing networks are reported.

7.2 Showing the Embedding of Trees

If a network provides any *Hybrid Information* you can generate a *Hybrid Layout* (cf. Sec. 5.4) which can be handled by a certain window. The window contains one row per embedded tree that enables the setting of its edge color outlining the embedding. Moreover, you can hide all edges that are necessary for displaying a certain tree to highlight the embedding of other trees.

8 More Memory

The algorithm for computing all hybridization networks of minimum reticulation number might needs a lot of memory. Thus, to enable its execution, you might have to increase the heap size of the java virtual machine (JVM) by setting the known arguments, e.g. `-Xmx` or `-Xms`, when running the program. As example, the following command runs the graphical user interface of Hybroscale by setting the maximum heap size of the JVM to 8 GB:

```
java -jar -Xmx8G hybroscale.jar
```

9 Command Line

You can run Hybroscale via command-line by using the following syntax:

Syntax:

```
java -jar Hybroscale.jar -i <inputfile> [-o <outputfile>] [-mode <hNum|hNet|reroot>] [--upperBound]
                                     [--v] [-cores <#cores>] [-format <norm|ext|none>]
```

Options:

```
-i <inputfile>          path to input file containing ONE newick string per line
-o <outputfile>        path to output file (default is StdOut)
-mode <hNet|hNum|reroot> mode of the algorithm (default is hNet)
  hNet:  compute all hybridization networks
  hNum:  compute hybridization number
  reroot: minimize hybridization number by re-rooting trees
--v          activate verbose mode reporting details of the running algorithm
-cores <#cores> number of available cores (default is #cores minus one)
-format <norm|ext|none> format of reported newick strings (default is normal)
  norm:  normal newick format
  ext:   extended newick format including edge labels
  none:  just the number of networks/tree-sets is reported
```

Please note that all options in squared brackets are optional. The syntax as well as all options, listed above, can be displayed as follows:

```
java -jar hybroscale.jar -h
```

As example, the following command computes all hybridization networks referring to all trees within the input file `trees.txt`:

```
java -jar hybroscale.jar -i trees.txt
```

Please note that the input file must contain **only one tree per line** following the *Extended Newick* format (cf. Sec. 6.1)

10 Short Example



Figure 3: Main window of Hybroscale showing the embedding of three trees within the network at the bottom right.

References

- [1] Huson, D.H. and Scornavacca, C. (2012) Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.*, accepted, doi:10.1093/sysbio/sys062.
- [2] Gabriel Cardona, Francesc Rossello, and Gabriel Valiente. Extended Newick: It is time for a standard representation of phylogenetic networks. *BMC Bioinformatics*, 9:532, 2008.