A guide to the MLTreeMap imagemaker

MLTreeMap is a software framework, designed for phylogenetic and functional analysis of metagenomic data. It searches for instances of marker genes on nucleotide sequences and deduces their most likely origin in a set of reference phylogenies. The MLTreeMap imagemaker has been designed to visualize the placement results within the reference phylogenies. It is not part of the MLTreeMap stand-alone package and has to be installed individually.

This guide can be downloaded from http://mltreemap.org/treemap_cgi/show_download_page.pl.

A) Installation of the MLTreeMap imagemaker

The MLTreeMap imagemaker consists of Perl scripts and has the following dependencies:

SVG

Math::Trig

If these dependencies are satisfied, the MLTreeMap imagemaker is ready to use. As a viewer for the SVG images we recommend Inkscape.

B) Usage of the MLTreeMap imagemaker

The MLTreeMap imagemaker has to be accessed on the command line and needs MLTreeMap output files as input.

An example for a valid input command is:
./mltreemap_imagemaker.pl –i example_input

This will concatenate all files of the same analysis type into one file and then generate the pictures (they will be written to the output directory
MLTreeMap_imagemaker_2_06/output/).

If you want to enter a single file as input, this is also possible:
./mltreemap_imagemaker.pl –i example_input/p_E_coli_RAxML_parsed.txt

Optional input parameters are:

-b   parameter for the size of the placement bubbles (default: 0.9).

-d   use different colors for different datasets (0: don't use this mode (default), 1: use this mode, 11: use this mode but concatenate the projected placement weights across the datasets). See below for a more detailed description.

-h   high quality placement bubbles (1, default) or low quality placement bubbles (0). Adobe Illustrator CS4 runs into an error if there are too many placement bubbles. If Adobe Illustrator CS4 is your sole viewer, you can use low quality bubbles that do not provoke this error.

-o   use this option if you want to write the output to another directory than the default.

-r   display 16S and 18S rRNA hits in different trees (default: 2) or in one tree-of-life (1). Note: MLTreeMap treats the 16S and 18S rRNA reference data as two different trees. For displaying reasons it might make sense to print all results into a single tree-of-life.

-t   display trees inclusive text and labels (1, default) or without text (0)

C) Detailed information about the –d input option.

The –d option is designed to print the concatenated results from different datasets into one single picture, while each dataset is represented by its own color. For the principle see Figure 0.1.

This mode works only if you enter a directory as input e.g. –i example_input (if you enter only one file as input, the –d option is irrelevant).

Currently the imagemaker supports up to 9 datasets. If your input directory contains more than 9 files of the same type of analysis, the program will die with an error message. You can expand the number of supported datasets by adding more colors in RGB format to the file tree_data/available_dataset_colors.txt.
The input files are alphanumerically assigned to the content of this file.
E.g. Let us assume that the color list contains blue, green and red as follows:
0 0 255
0 255 0
255 0 0
Let us further assume that your input files are the same as in Figure 0.1. If that is the case, then 'h_file_1...' will get the blue, 'h_file_2...' the green and 'h_file_3...' the red color. This information will also be written into two legend files in the output directory. One file is in text format, while the other is an SVG file with colored labels.

If you use the '-d 1' option, the projected placement weights will be given for every dataset. Unfortunately the current version of the MLTreeMap imagemaker does not check for the length of the resulting text string. Thus if you have many datasets, the percentages may end up being drawn on top of the leaf labels. You can solve this issue by manually adjusting the picture in a program like inkscape. Another solution can be to use the '-d 11' option. In this case, the percentages are concatenated and given as a total. In publications it usually does not make sense to show all the labels (in pictures of ordinary sizes, they will be to small to be legible). If you use the '-t 0' option, the text will not be displayed and the pictures optimized to look good without it.
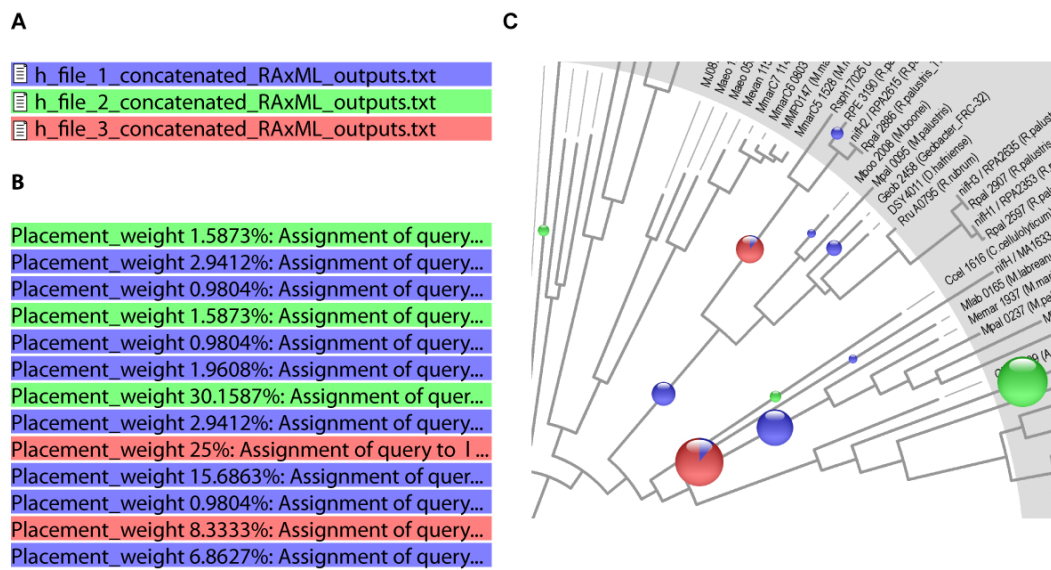
2

A

B

C

Figure 0.1: The –d input option. Usage: './mltreemap_imagemaker.pl –i your_input –d 1'. A: If you use the –d option, each file in 'your_input' will be treated as an individual dataset (as indicated by the colored bars). B: Before concatenating the files, each line is labelled with a color code, representing its origin. C: The output picture is drawn according to this color code. If several datasets have placements at the same nodes, pie chart bubbles are drawn to represent the fractional weights.