Lopper: A tool to prune phylogenetic datasets to a common list of taxa.

A common requirement of comparative phylogenetic analysis tools is that the objects to compare contain exactly the same taxa. Often this is not the case: for example a phylogenetic tree might contain taxa for which data has not been collected, or different studies of a family might contain slightly different species.

Lopper allows the user to take two files that contain taxonomic information and compare them to each other. The result is two new files, in the same format as the input ones, that contain only those taxa that are present in both inputs. These files can be a list, a phylogenetic tree and a trait data table in any combination.

Optionally, the tool can return the lists of matched and mismatched taxa.

Usage: Rscript lopper.R -f FORMATS [OPTIONS] FILE1 FILE2

Reduces a pair of files which contain taxa names (tree, trait matrix, list) to the common set of taxa.

-h --help Prints this help.

-v --version Prints the program's version.

-f FORMAT --format FORMAT Format of FILE1 and FILE2. If only one format is provided it is used for both files.

Currently supported formats are:

csv: table of traits in comma separated value format

csvx: table of traits in comma separated value without header format

nwk: phylogenetic tree in newick format

nex: phylogenetic tree in nexus format

lst: list of taxa names in plain text.

-m NUM --mismatches NUM: lists of mismatched taxa to save:

0 = none (default)

1 = FILE1

2 = FILE2

3 = both FILES.

-a AUTHORITY --auth AUTHORITY: Authority for resolution of taxa name. Ignored if TNRS is not used

-t NUM --tnrs NUM Use Taxonomic Resolution Service for:

0 = none (default)

1 = FILE1

2 = FILE2

3 = both FILES.

Currently supported authorities

"tropicos" (default): Tropicos, botanical information system at the Missouri Botanical Garden

-j --json The output is sent as a JSON string to STDOUT and no files are generated (options -m and -o are ignored).