Manual (under construction)

Input files

In this tab, you can supply files to *massoc*. Moreover, you can save or clear the current settings, and load saved settings from a .xyz settings file. If you supplied the correct inputs, the dialog box on the right will summarize some properties of your files.

Set default directory

This directory contains the filepath to a location where you would like output files to be stored.

Open BIOM files

Specify the complete filepaths of your BIOM files. You can select multiple BIOM files at once, although you can only specify settings once; hence, if these settings are incompatible with some of your BIOM files, this may cause *massoc* to crash.

Open count tables

Specify the complete filepaths of your tab-delimited count tables. You can select multiple files at once, but make sure these are selected in the same order as your taxonomy files. The count tables should be in a format that can be accept by the BIOM-format parser.

Preprocessing

In this tab, you can specify preprocessing steps that *massoc* should carry out before network inference.

Prefix for saved files

Helps identify intermediate files stored by *massoc*.

Remove taxa with mean count below:

Before a prevalence filter or rarefaction, you can remove taxa that have a count number below the supplied number. This can help remove singletons or low-abundance taxa. The removed counts are binned.

Rarefy:

Rarefy to even depth or to a specific count number. If the supplied count number is higher than the sample count number, the sample is removed from the dataset.

Set prevalence filter in %

Filter taxa so only those remain that are present in the specified percentage of samples. The removed counts are binned. Prevalence filters can help remove spurious associations, and the default is therefore set to 20%.

Split samples by metadata variable

Select a metadata column to split the datasets by. For example, half of your datasets were taken from lakes, and the other half from rivers. If this is in the BIOM file, you can use it to split the BIOM file and run network inference on the samples separately (as well as on the original network).

Taxonomic levels to analyze

Select taxonomic levels to perform network inference on. The BIOM file is collapsed to different taxonomic levels, and the BIOM format *collapse* function is adapted to preserve the taxonomy up to the collapsed level. The identifiers of the collapsed taxa are also preserved in these BIOM files.

Clustering algorithm

If there are no metadata available that explain microbiome structure, there may still be a structure present that could induce spurious associations (e.g. pH may explain microbiome clustering, but pH was not measured). In that case, *massoc* can run clustering algorithms: it first transforms the data using the centered log-ratio, and then ordinates them in PCA space. The PCA coordinates are used as input for the clustering algorithms. The cluster identities can either be added to the metadata, or used to split the BIOM file in multiple files.

Network inference

In this tab, you can specify which tools should be used to run network inference. For some of these tools, you can also adjust the settings. Finally, an overview of the settings is given, as well as an option to copy these to a command line call.

Select network inference tools to run

Currently, SparCC, CoNet and SPIEC-EASI are supported. When selected, these tools will run network inference using the default settings.
For more details on these tools, please check out the original publications:
SparCC: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002687
SPIEC-EASI: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004226
CoNet: https://f1000research.com/articles/5-1519/v1

Change settings for SparCC

SparCC bootstraps correlations, and uses the bootstraps to compute pseudo p-values. Lower p-values can result in sparser networks.

Change settings for SPIEC-EASI

SPIEC-EASI can use two different algorithms: the Meinshausen-Buhlmann algorithm and the Graphical Lasso algorithm. For both of these algorithms, the sparsity of the final network is controlled by the number of StARS repetitions.

Change settings for CoNet

As CoNet has a multitude of settings, there is no option to adjust them inside *massoc.* However, a future update will contain a utility to supply alternative CoNet.sh scripts to *massoc.*

Number of processes

After all settings have been entered, *massoc* will specify the number of unique networks that need to be inferred. Network inference can be sped up significantly if networks are inferred simultaneously. More cores can be added if a particularly large number of networks needs to be inferred.

Export as command line call

While there is an option to generate a command line call, there is no compiled version of the command line utility available.

Run network inference

Network inference will start, and a loading bar keeps track of progress. When finished, one or more output GraphML files will be written to the previously specified filepath.