# User Manual for MEGAN V5.11.3

Daniel H. Huson

January 21, 2016

## Contents

Contents 1

1 Introduction 3

2 Getting Started 5

3 Obtaining and Installing the Program 5

4 Licensing 6

5 Program Overview 7

6 Importing, Reading and Writing Files 7

7 Taxonomic Binning 9

8 Functional Analysis 11

9 Comparison of samples 12

10 Main Window 12

11 SEED Window 24

12 COG Window 24

13 KEGG Window 24

14 Sample Viewer 25

15 Cluster Viewer 27

16 Import Dialog 29

17 Inspector Window 31

18 Microbial Attributes Window 32

19 Rarefaction Window 32

20 Chart Window 32

21 Alignment Viewer 33

22 Group Viewer 35

23 Find Toolbar 35

24 Format Dialog 36

25 Message Window 36

26 Parameters Dialog 36

27 Compare Dialog 37

28 Extractor Dialog 37

29 Export Image Dialog 37

30 About Window 38

31 Server Browser Window 38

32 File Formats 38

33 Command-Line Options 45

34 Command-Line Commands 47

35 Examples 51

36 Using More Memory 51

37 MEGAN Tools 52

38 Acknowledgments 52

**References** **52**

**Index** **55**

# 1 Introduction

**Disclaimer**: This software is provided "AS IS" without warranty of any kind. This is developmental code, and we make no pretension as to it being bug-free and totally reliable. Use at your own risk. We will accept no liability for any damages incurred through the use of this software. Use of the MEGAN is free for academic usage, however the program is not open source.

**Type-setting conventions**: In this manual we use e.g. `Edit`→Find... to indicate the `Find...` menu item in the `Edit` menu.

**How to cite**: If you publish results obtained in part by using MEGAN , then we require that you acknowledge this by citing the program as follows:

- D.H. Huson *et al*, Integrative analysis of environmental sequences using MEGAN 4, Genome Res. 2011. 21:1552-1560.

The initial idea to develop MEGAN was born in 2005, when Daniel Huson and Stephan C. Schuster (then at Penn State) discussed that a metagenomic analysis tool was required to analyze reads produced using 454 sequencing for a mammoth sample, the goal being to try to identify the origin of any non-mammoth reads. The resulting paper was the first publication to use next-generation sequencing for metagenomics [16].

The first public version of MEGAN was written by Daniel Huson during a sabbatical at the University of Canterbury in 2006, hosted by Mike Steel at the Biomathematics Research Centre. The result of this work is presented in [7]. A second sabbatical was invested in 2011 to significantly extend the scope of the program and to rewrite major parts of the code. The result of this work (version 4) is presented in [8].

MEGAN consists of $\approx 275,000$ lines of Java code, generated over a period of nine years.

Stephan C. Schuster, Suparna Mitra, Daniel C. Richter, Paul Rupek, Hans Ruscheweyh and Nico Weber contributed ideas and some supporting code.

The term *metagenomics* has been defined as "The study of DNA from uncultured organisms" (Jo Handelsman), and an approximately 99% of all microbes are believed to be unculturable. A *genome* is the entire genetic information of one organism, whereas a*metagenome* is the entire genetic information of an *ensemble* of organisms. Metagenome projects can be as complex as large-scale vertebrate projects in terms of sequencing, assembly and analysis.

The aim of MEGAN is to provide a tool for studying the taxonomic content of a set of DNA reads, typically collected in a metagenomics project. In a preprocessing step, a sequence comparison of all reads with a suitable database of reference DNA or protein sequences must be performed to produce an input file for the program. MEGAN is suitable for DNA reads (metagenome data), RNA reads (metatranscriptome data), peptide sequences (metaproteomics data) and, using a suitable synonyms file that maps SILVA ids to taxon ids, on 16S rRNA data (amplicon sequencing).

At start-up, MEGAN first reads in the current NCBI taxonomy (consisting of over one million taxa). A first application of the program is that it facilitates interactive exploration of the NCBI taxonomy.

However, the main application of the program is to parse and analyze a the result of a BLAST comparison of a set of reads against one or more reference databases, typically using BLASTN, BLASTX or BLASTP to compare against NCBI-NT, NCBI-NR or genome specific databases. The result of a such an analysis is an estimation of the taxonomical content ("species profile") of the sample from which the reads were collected. The program uses a number of different algorithms to "place" reads into the taxonomy by assigning each read to a taxon at some level in the NCBI hierarchy, based on their hits to known sequences, as recorded in the BLAST file.

Alternatively, MEGAN can also parse files generated by the RDP website [5] or the Silva website [18]. Moreover, MEGAN can parse files in SAM format [10]. Note that DIAMOND [4] and MALT produces SAM format for BLASTX-like alignments and MEGAN is able to parse this, see also sam2rma.

As of version 4, MEGAN provides functional analysis using both the SEED classification [15] and also using KEGG pathways [9].

An early version of this software (called GenomeTaxonomyBrowser) was used to analyze metagenomic reads in the very first paper to use second-generation sequencing for metagenomics [16].

This document provides both an introduction and a reference manual for MEGAN .

## 2 Getting Started

This section describes how to get started.

First, download an installer for the program from www-ab.informatik.uni-tuebingen.de/software/megan5, see Section 3 for details.

Upon startup, the program will automatically load its own version of the NCBI-taxonomy and will then display the first three levels of the taxonomy. To explore the NCBI taxonomy further, leaves of this overview tree can be uncollapsed. To do so, first click on a node to select it. Then, use the `Tree→Uncollapse` item to show all nodes on the next level of the taxonomy, and use the `Tree→Uncollapse Subtree` item to show all nodes in the complete subtree below the selected node (or nodes).

To analyze a data set of reads, first BLAST the reads against a database of reference sequences, such as NCBI-NR [3] using BLASTX [1] or BLASTP, NCBI-NT [3] using BLASTN [1]. In addition, the output of a number of other programs can also be parsed, for example, RAPSearch2 [22].

Then import the BLAST file into MEGAN using the `File→Import From BLAST...` menu item. The Files tab allows you to enter the name of the BLAST file, a reads file containing all the read sequences in multi-FastA format (if available), and the name of the new output RMA file. As of version 4, you can also specify more than one BLAST file and one more than one reads file.

Alternatively, instead of supplying a BLAST file, one can also specify a file obtained from the RDP website or from the Silva website. In additional, MEGAN can also parse files in SAM format.

Some implementations or output formats of BLAST suppress those reads for which no alignments were found. In this case, use the `Options→Set Number Of Reads...` menu item to set the total number of reads in the analysis.

Clicking on a node will cause the program to display the exact number of hits of any given node, and the number of hits in the subtree rooted at the node. Right-clicking on a node will show a popup-menu and selecting the first item there, `Inspect` , will open the `Inspector` window which is used to explore the hits associated with any given taxon. A node is selected by clicking on it. Double-clicking on a node will select the node and the whole subtree below it. Double-clicking on the label of a node will open the node in the `Inspector` window.

Example files are provided with the program. They are contained in the `examples` subdirectory of the installation directory. The precise location of the installation directory depends upon your operating system.

## 3 Obtaining and Installing the Program

MEGAN is written in Java and requires a Java runtime environment version 1.7 or newer, freely available from www.java.org. Note that the Windows and MacOS X installers both contain a bundled JRE and so separate installation of Java should not be necessary for these two operating systems.

MEGAN is installed using an installer program that is freely available from www-ab.informatik.uni-tuebingen.de/software/megan. There are four different installers, targeting different operating systems:

- `MEGAN_windows_5.11.3.exe` provides an installer for a 32-bit version of MEGAN for Windows.
- `MEGAN_windows-64x_5.11.3.exe` provides an installer for a 64-bit version of MEGAN for Windows.
- `MEGAN_macos_5.11.3.dmg` provides an installer for MacOS X.
- `MEGAN_unix_5.11.3.sh` provides an installer for Linux and Unix.

The 32-bit Windows version of MEGAN is configured to use 1.1 GB of memory. For all other versions of the software, the installer will allow you to set the maximal amount of memory during the installation process. By default, the program will suggest to use 2 GB. If your computer has more memory available, then it is a good idea to set this limit higher. For example, if you have 4 GB of main memory, then set the limit for MEGAN to 3 GB. This is because the program runs faster, the more memory it is given. To change the maximum amount of memory used after installation of the program, see Section 36.

To install MEGAN using a command-line dialog, launch the installer from the command line and pass the command-line option `-c`. For example, under MacOS X, type the following:

```
/Volumes/MEGAN/MEGAN\ Installer.app/Contents/MacOS/JavaApplicationStub -c
```

## 4 Licensing

Any usage of MEGAN requires a license. The following license types are available:

- *Academic license* This license permits use of the software exclusively for academic research (publications in academic journals and papers at academic conferences) and instruction. This type of license is available free of charge from the program website.
- *Single user license* This license permits a single user to use the program. This type of license is granted for a charge from the University of Tübingen, see the program website.
- *Site license* This license permits use of the program at a single physical location, within a single organization. This type of license is granted for a charge from the University of Tübingen, see the program website.
- *Enterprise license* This license permits use of the program anywhere within a single organization. This type of license is granted for a charge from the University of Tübingen, see the program website.
- *Evaluation license* This type of license is granted for 45 days and is for evaluation purposes only. It is available free of charge from the University of Tübingen, see the program website.

# 5 Program Overview

In this section, we give an overview over the main design goals and features of this program. Basic knowledge of the underlying design of the program should make it easier to use the program.

MEGAN is written in the programming language Java. The advantages of this is that we can provide versions that run under the Linux, MacOS, Windows and Unix operating systems.

Typically, after generating a RMA file (read-match archive) from a BLAST file, the user will then interact with the program, using the Find toolbar to determine the presence of key species, collapsing or un-collapsing nodes to produce summary statistics and using the `Inspector` window to look at the details of the matches that are the basis of the assignment of reads to taxa. The assignment of reads to taxa is computed using the LCA-assignment algorithm, see [7] for details.

The program is designed to operate in two different modes: in a GUI mode, the program provides a GUI for the user to interact with the program. In command-line mode, the program reads commands from a file or from standard input and writes output to files or to standard output.

The linux and MacOS X distributions of MEGAN each provide two executables. The executable `MEGAN` launches the program in interactive GUI mode. The executable `megan-commandline` runs the program in non-interactive command-line mode.

# 6 Importing, Reading and Writing Files

To open an existing RMA file or MEGAN text file, select the `File→Open...` menu item and then browse to the desired file. Alternatively, if the file was recently opened by the program, then it may be contained in the `File→Open Recent` submenu.

By default, when parsing an input file, for each read, taxon and RefSeq id, only one best-scoring match is kept. For example, if read $R$ has two equally high-scoring matches $M_1$ and $M_2$ to two sequences from *E. coli*, say, then MEGAN will discard one of the two matches, unless they have different RefSeq accession numbers, or unless exactly one of the two matches does not have a RefSeq accession number. To turn this filter off, use the `Window→Command Input...` menu item to enter the following command `setProp oneMatchPerTaxon=false`.

## 6.1 Blast Files

New input to the program is usually provided as a BLAST file obtained from a BLAST comparison of the given set of reads to a database such as NCBI-NR or NCBI-NT, see Section 32 for details of the file formats used. MEGAN supports BLASTN, BLASTX and BLASTP standard text-format, and BLAST XML format.

The BLAST files and reads files may also be parsed in *gzip format* or *zip format* . In this case, the file names should have concatenated suffixes ending on `.gz` or `.zip`, such as `.blastn.gz` or `.fna.zip`. However, this is incompatible selected the `Don't save` button in the Advanced Tab.

MEGAN can also parse tabular BLAST output (generated using BLAST option `-m 8`, however as this form of output does not contain the subject line for sequences matched, it is unsuitable for MEGAN because MEGAN cannot determine the taxon or gene associated with the database sequence. However, if you add an additional column to this format containing the associated taxon name or numerical NCBI taxon-id for each line then MEGAN will parse these and use them as input. For unknown taxa, write either `unknown` or `-1` in the column.

Note that, in all cases, the reads file should be given to use the full potential of the program.

The BLAST file and reads file are supplied to MEGAN when setting up a new *MEGAN project* . Both files are parsed and all information is stored in the project file. The input data is then analyzed and can be interactively explored. All reads and BLAST matches are contained in the project file and MEGAN provides different mechanisms for extracting them again. A MEGAN project file contains all reads and all significant BLAST matches (by default, up to 100 matches per read) in a binary and incrementally compressed format. The size of such a project file is around 20% of the size of the original input files and is thus usually smaller than the file that one obtains by simply compressing the BLAST file. As of version 4, MEGAN provides more control over whether and how BLAST matches and reads are stored, see the discussion of the `Import Blast` window.

MEGAN5 uses a new algorithm for determining the taxon associated with a given reference sequence. In previous versions, the program looked in the header line of a reference sequence for the longest substring that matches some valid taxon name (or synonym) in the NCBI taxonomy. This determined which taxon to assign to the match. However, because many entries in the NR database mention multiple different species for a given match, the program now determines only maximal matching names in the header line and assigns the match to the LCA of the taxa mentioned. (So, in particular, the LCA algorithm is used twice in MEGAN, namely once to figure out which taxon to assign to a match and then, based on this, again to determine which taxon to assign to a given read.)

Thus, it is important that alignment programs used in conjunction with MEGAN must list all taxa associated with a given reference sequence.

## 6.2 SAM Files

MEGAN can now parse files in *SAM* format [10]. Note, however, that SAM files usually do not contain the names of the taxa associated with the reference sequences and so one must supply suitable mapping files that map identifiers used for the reference sequences to NCBI taxa, KEGG, COG and/or SEED identifiers, see below.

## 6.3 RDP Files

In addition, MEGAN can import rRNA analysis files downloaded from the *RDP* website at `http://rdp.cme.msu.edu/` [5]. Go to the website and upload your rRNA sequences and then let the website process them for you. Please note that the RDP website allows one to download two types of files, namely a *hierarchy as text* file from its `Classifier :: Hierarchy View` window and a *text* file obtained from its `Classifier :: Assignment Detail` window. Input to MEGAN must be of the latter type. The RDP website recommends using a `Min Score` setting of 80. MEGAN calls this the *RDP-Assignment-Detail* format.

If you use the standalone RDP classifier, then the output has a different format. MEGAN calls this the `RDP-standalone` format. In this case, MEGAN expects the format to be a tab-separated file in which each line corresponds to one read:

```
read-name [-] [taxon-name rank-name score] [taxon-name rank-name score] ...
```

In more detail, the first token is a string that identifies the read. The next token is either empty, or a minus, in the latter case indicating that the read is reverse complemented. Then all further tokens come in groups of three, indicating the name of a taxon, the name of the rank of the taxon and a score between 0 and 1 (which MEGAN will multiple by 100). MEGAN reports each such taxon as a separate hit for the read.

### 6.4 Silva Files

Similarly, MEGAN can import rRNA analysis files downloaded from the *Silva* website at `http://www.arb-silva.de` [18]. To create a file using the Silva website that can be imported into MEGAN, go to the Aligner tab of the Silva website and upload your sequences and then press the *align sequences* button. Once the Silva website has computed an alignment, you will be able to download two files, an *arb file* and a *log file*. MEGAN requires the log file as input, *not* the arb file. When importing such a file into MEGAN, one must specify that MEGAN uses the synonyms file called `silva2ncbi.map` to map Silva accession numbers to NCBI taxa. This file is available from the MEGAN download page.

### 6.5 CSV Files

MEGAN supports import of data from other programs in a delimiter-separated format (using comma's or tabs) from a CSV file.

### 6.6 BIOME Format Files

*BIOME* is a new file format for exchanging data between different metagenome analysis tools. MEGAN can import and export data in BIOME format, see `http://biom-format.org/`. For example, can import OTU classification data generated by the *QIIME* package [**?**], or taxonomic and functional classifications generated by *MG-RAST* [6], using the `Import→BIOM1 Format...` menu item. To export data in BIOME format, open the viewer for the type of data that you would like to export. For example, if you want to export a SEED classification, then open the `SEED` Analyzer window and select those nodes that you want to export. Then use the `Export→BIOM1 Format...` to save the data to a file. The suffix of a BIOME file is `.biom`.

## 7 Taxonomic Binning

By default, MEGAN uses the NCBI taxonomy for taxonomic binning.

## 7.1 The NCBI Taxonomy

The *NCBI taxonomy* provides unique names and IDs for over 660,000 taxa, including approximately 25,000 prokaryotes, 84,000 animals, 65,000 plants, and 17,000 viruses. The individual species are hierarchically grouped into clades at the levels of: Superkingdom, Kingdom, Phylum, Class, Order, Family, Genus, and Species (and some unofficial clades in between).

At startup, MEGAN automatically loads a copy of the complete NCBI and then displays the taxonomy as a rooted tree. The taxonomy is stored in an NCBI tree file and an NCBI mapping file, which are supplied with the program.

## 7.2 The NCBI-NR and NCBI-NT Databases

The *NCBI-NR* ("non-redundant") protein sequence database is available from the NCBI website. It contains entries from GenPept, Swissprot, PIR, PDF, PDB and RefSeq. It is non-redundant in the sense that identical sequences are merged into a single entry.

The *NCBI-NT* nucleotide sequence database is available from the NCBI website. It contains entries from GenBank and is not non-redundant. It contains untranslated gene coding sequences and also mRNA sequences.

## 7.3 Assigning Reads to Taxa

The main problem addressed by MEGAN is to compute a "species profile" by assigning the reads from a metagenomics sequencing experiment to appropriate taxa in the NCBI taxonomy. At present, this program implements the following naive approach to this problem:

1. Compare a given set of DNA reads to a database of known sequences, such as NCBI-NR or NCBI-NT [3], using a sequence comparison tool such as BLAST [1].
2. Process this data to determine all hits of taxa by reads.
3. For each read $r$, let $H$ be the set of all taxa that $r$ hits.
4. Find the lowest node $v$ in the NCBI taxonomy that encompasses the set of hit taxa $H$ and assign the read $r$ to the taxon represented by $v$.

We call this the naive LCA-assignment algorithm (LCA = "lowest common ancestor"). In this approach, every read is assigned to some taxon. If the read aligns very specifically only to a single taxon, then it is assigned to that taxon. The less specifically a read hits taxa, the higher up in the taxonomy it is placed. Reads that hit ubiquitously may even be assigned to the root node of the NCBI taxonomy.

If a read has significant matches to two different taxa $a$ and $b$, where $a$ is an ancestor of $b$ in the NCBI taxonomy, then the match to the ancestor $a$ is discarded and only the more specific match to $b$ is used.

The program provides a threshold for the bit score of hits. Any hit that falls below the threshold is discarded. Secondly, a threshold can be set to discard any hit whose score falls below a given percentage of the best hit. Finally, a third threshold is used to report only taxa that are hit by a minimal number of reads or minimal percent of all assigned reads. By default, the program requires at least 0.1% of all assigned reads to hit a taxon, before that taxon is deemed present. All reads that are initially assigned to a taxon that is not deemed present are pushed up the taxonomy until a node is reached that has enough reads. This is set using the `Min Support Percent` or `Min Support` button.

Taxa in the NCBI taxonomy can be excluded from the analysis. For example, taxa listed under `root - unclassified sequences - metagenomes` may give rise to matches that force the algorithm to place reads on the `root` node of the taxonomy. This feature is controlled by `Options→Taxon Disabling` menu. At present, the set of disabled taxa is saved as a program property and not as part of a MEGAN document.

Note that the *LCA-assignment algorithm* is already used on a smaller scale when parsing individual blast matches. This is because an entry in a reference database may have more than one taxon associated with it. For example, in the NCBI-NR database, an entry may be associated with up to 1000 different taxa. This implies, in particular, that a read that may be assigned to a high level node (even the root node), even though it only has one significant hit, if the corresponding reference sequence is associated with a number of very different species.

Note that the list of disabled taxa is also taken into consideration when parsing a BLAST file. Any taxa that are disabled are ignored when attempting to determine the taxon associated with a match, unless all recognized names are disabled, in which case the disabled names are used.

MEGAN contains two enhancements to the naive LCA algorithm:

The *minimum coverage heuristic* greedily identifies a subset of all taxa such that each read that has at least one significant alignment to the complete set of taxa will also have at least one significant match to a taxon in the identified subset. The LCA algorithm is then run using only matches to taxa in the selected subset. A greedy heuristic is used to select a subset of minimum size. This is activated using the `Use Use Minimum Coverage Heuristic` button.

The *LCA of a fixed percent* places each read on the lowest node that covers the given percentage $t$ (between 50 and 100) of taxa associated with its alignments. By default, $t = 100$ and the algorithm is identical to the simple naive LCA. This is set using the `LCA Percent` button.

# 8 Functional Analysis

MEGAN allows functional analysis using SEED, COG and KEGG.

## 8.1 Identification of SEED Functional Classes

The *SEED* classification of gene function consists of a collection of biologically defined *subsystems* [15]. The SEED classification can be displayed as a tree containing about 10,000 nodes and edges. Genes are mapped onto *functional roles* and these are present in one or more subsystems. The program will attempt to map each read onto a gene that has an known functional role and then into one or more subsystems.

To perform this analysis, MEGAN uses a mapping of *RefSeq* ids to SEED functional roles. Hence, if a SEED-based analysis is desired, then the database that is used in the BLAST comparison must contain RefSeq-ids. This is the case for the NCBI-NR database. The `Import Blast` dialog provides other options for identifying SEED functional roles.

### 8.2 Identification of COGs

The *COG* classification of gene function consists of a collection of biologically defined *clusters of orthologous groups* [19, 17]. The COG classification can be displayed as a tree containing many nodes and edges. Genes are mapped onto *COGs* and *NOGs* . The program will attempt to map each read onto a gene that has an known COG or NOG.

To perform this analysis, MEGAN uses a mapping of *RefSeq* ids to COGs. Hence, if a COG-based analysis is desired, then the database that is used in the BLAST comparison must contain RefSeq-ids. This is the case for the NCBI-NR database. The `Import Blast` dialog provides other options for identifying COGs.

### 8.3 Mapping of Reads to KEGG groups and pathways

The *KEGG* database provides a collection of *metabolic pathways* and other pathways [9]. The KEGG classification can be displayed as a tree, which we refer to as the *Kegg tree* . Genes are mapped onto so-called *KO* groups and these are present in one or more pathways. The program will attempt to map each read onto a gene that has a valid KO identifier and thus to one or more pathways.

To perform this analysis, MEGAN uses a mapping of RefSeq-ids to KO groups. Hence, if a KEGG-based analysis is desired, then the database that is used in the BLAST comparison must contain RefSeq-ids. This is the case for the NCBI-NR database. The `Import Blast` dialog provides other options for identifying KO groups.

Due to a change of KEGG licensing in 2011, MEGAN ships with a old version of KEGG. If you have a license for KEGG, then you can update your installation of MEGAN to use the latest versions of KEGG pathways as follows. Using your paid access to the KEGG ftp site, download the `pathway/map` directory from the KEGG ftp site and copy all `map*.png` and `map*.conf` to the directory called `class/resources/kegg` in the MEGAN installation directory. MEGAN will then use those files rather than the old file set.

## 9 Comparison of samples

Multiple samples can be opened simultaneously and then displayed together in a comparison view.

## 10 Main Window

The `Main` window is used to display the taxonomy and to control the program via the main menus. Initially, at startup, before reopening or creating a new RMA file, the `Main` window displays the NCBI taxonomy. By default, the taxonomy is only drawn to its second level. Parts of the taxonomy, or the full taxonomy, can be explored using the menu items of the window.

Once a data set has been read in, the full NCBI taxonomy is replaced by the taxonomy that is induced by the data set. The size of nodes indicates the number of reads that have been assigned to the nodes using the algorithm described in Section 7.3.

Double-clicking on a node will produce a textual report stating how many reads have been assigned to the corresponding taxon and how many reads have been assigned in total to the taxon and to any of the taxa below the given node in summary.

Subtrees can be collapsed and expanded, as described below. Most windows in MEGAN provide access to their functionality through menus, a tool bar that contains a selection of the menu items, and popup menus that also provide contextual access to menu items.

We now discuss all menus of the `Main` window.

## 10.1 The File menu

The `File` menu contains the following items:

- The `File→New...` item: Open a new empty document.
- The `File→Open...` item: Open a MEGAN file (ending on .rma, .meg or .megan).
- The `File→Open Recent` submenu.
- The `File→Open From Server...` item: Open the `Server Browser` window to allow access to files hosted on an instance of MeganServer.
- The `File→Import From BLAST...` item: Show the `Import Blast` dialog.
- The `File→Import` submenu.
- The `File→Save As...` item: Save current data set.
- The `File→Export Image...` item: Export content of window to an image file.
- The `File→Export Legend...` item: Export content of legend window.
- The `File→Export` submenu.
- The `File→Page Setup...` item: Setup the page for printing.
- The `File→Print...` item: Print the main panel.
- The `File→Extract To New Document...` item: Extract all reads and matches on or below selected node(s) to a new document.
- The `File→Extract Reads...` item: Extract reads for the selected nodes.
- The `File→Properties...` item: Show document properties.

- The `File→Close` item: Close the window.
- The `File→Quit` item: Quit the program (Windows and Linux only).

## 10.2 The Open Recent menu

The `Open Recent` menu allows one to reopen recently opened files.

## 10.3 The Import menu

The `Import` menu contains the following items:

- The `Import→CSV Format...` item: Load data in delimiter-separated-values (CSV) format: READ-NAME,CLASS-NAME,SCORE or CLASS,COUNT(,COUNT...).
- The `Import→BIOM1 Format...` item: Import data from a table in BIOME format (see `http://biom-format.org/documentation/format-versions/biom-1.0.html`).
- The `Import→Metadata...` item: Import a metadata mapping file (as defined in `http://qiime.org/documentation/file_formats.html`).

## 10.4 The Export menu

The `Export` menu contains the following items:

- The `Export→CSV Format...` item: Export assignments of reads to nodes to a CSV (tab or comma-separated) file.
- The `Export→BIOM1 Format...` item: Export assignments of reads to nodes in BIOME format.
- The `Export→Metadata...` item: Export a metadata mapping file (as defined in `http://qiime.org/documentation/file_formats.html`).
- The `Export→Taxonomic Paths...` item: Export assignments of reads weighted taxonomic paths.
- The `Export→Tree...` item: Export induced taxonomic tree (in Newick format).
- The `Export→Reads...` item: Export all reads to a text file (or only those for selected nodes, if any selected).
- The `Export→Matches...` item: Export all matches to a text file (or only those for selected nodes, if any selected).
- The `Export→Alignments...` item: Calculate and export alignments (or assembled contigs) for all selected leaves.

- The `Export→Coverage vs CG Content...` item: Export coverage vs GC content for reference sequences of all selected nodes.
- The `Export→MEGAN5 File...` item: Export an RMA computed using version 4 of MEGAN as a new RMA file that is fully compatible with MEGAN5. In MEGAN4, the refSeq identifier for each match was stored and this was used to identify the SEED or KEGG accession for the match during analysis. In MEGAN5, the SEED, KEGG and COG identifier for each match is stored In MEGAN4, the refSeq identifier for each match was stored and this was used to identify the SEED or KEGG accession for the match during analysis. In MEGAN5, the SEED, KEGG and COG identifier for each match is stored. When exporting a MEGAN4 as a MEGAN5 file, all matches are reparsed and by default, RefSeq identifiers are are used to identify SEED, COG and KEGG identifiers. To use other identifiers for this purposes, load the corresponding mapping files before running this command, either entering the appropriate commands using the `Window→Command Input...` menu iem or by using the `Import Blast` dialog.
- The `Export→MEGAN Summary File...` item: Export as summary file.

## 10.5 The Edit menu

The `Edit` menu contains the following items:

- The `Edit→Cut` item: Cut.
- The `Edit→Copy` item: Copy.
- The `Edit→Copy Image` item: Copy image to clipboard.
- The `Edit→Copy Legend` item: Copy legend image to clipboard.
- The `Edit→Paste` item: Paste.
- The `Edit→Edit Node Label` item: Edit the node label.
- The `Edit→Edit Edge Label` item: Edit the edge label.
- The `Edit→Description...` item: Add a short description to the sample.
- The `Edit→Format...` item: Format nodes and edges.
- The `Edit→Find...` item: Open the find toolbar.
- The `Edit→Find Again` item: Find the next occurrence.
- The `Edit→Preferences` submenu.

### 10.6 The Preferences menu

The `Preferences` menu contains the following items:

- The `Preferences→Show Legend` item: Show horizontal or vertical legend, or hide.
- The `Preferences→Fix Taxon Mapping` submenu.
- The `Preferences→Use Alternative Taxonomy...` item: Open alternative taxonomy.tre and taxonomy.map files.
- The `Preferences→Use Default NCBI Taxonomy` item: Open default NCBI taxonomy.
- The `Preferences→Use Magnitudes` item: Parse *magnitude* or *read weights* in read FastA header lines. Use this when reads have weights. Read weights must be specified in one of the following formats: `magnitude=`*number* or `magnitude|`*number*. If neither of these expressions if found, then the program will scan for `weight=`*number* or `weight|`*number*. In all cases, *number* must be a positive integer.

### 10.7 The Fix Taxon Mapping menu

The `Fix Taxon Mapping` menu contains the following items:

- The `Fix Taxon Mapping→Add A Change...` item: Change the taxon name to taxon id mapping for a given taxon.
- The `Fix Taxon Mapping→List All Changes...` item: List all changes.
- The `Fix Taxon Mapping→Clear All Changes...` item: Clear all changes.

### 10.8 The Select menu

The `Select` menu contains the following items:

- The `Select→All Nodes` item: Select nodes.
- The `Select→None` item: Deselect all nodes.
- The `Select→From Previous Window` item: Select from previous window.
- The `Select→All Leaves` item: Select all leaves (except Not Assigned, No Hits and Low Complexity).
- The `Select→All Internal Nodes` item: Select all internal nodes.
- The `Select→All Intermediate Nodes` item: Select all intermediate nodes.
- The `Select→Subtree` item: Select subtree.

- The `Select→Leaves Below` item: Select allow leaves below.
- The `Select→Invert` item: Invert selection.
- The `Select→Level` submenu.

### 10.9 The Level menu

The `Level` menu contains the following items:

- The `Level→Super Kingdom` item: Select Super Kingdom.
- The `Level→Kingdom` item: Select Kingdom.
- The `Level→Phylum` item: Select Phylum.
- The `Level→Class` item: Select Class.
- The `Level→Order` item: Select Order.
- The `Level→Family` item: Select Family.
- The `Level→Genus` item: Select Genus.
- The `Level→Species` item: Select Species.

### 10.10 The Options menu

The `Options` menu contains the following items:

- The `Options→Change LCA Parameters...` item: Rerun the LCA analysis with different parameters.
- The `Options→Set Number Of Reads...` item: Set the total number of reads in the analysis (will initiate recalculation of all classifications).
- The `Options→Taxon Disabling` submenu.
- The `Options→List Summary...` item: List summary of hits for selected nodes of tree.
- The `Options→List Path...` item: List path from root to selected node(s).
- The `Options→List Microbial Attributes...` item: List NCBI microbial attributes for selected microbes.
- The `Options→Shannon-Weaver Index...` item: Compute the Shannon-Weaver diversity index.
- The `Options→Simpson-Reciprocal Index...` item: Compute the Simpson-Reciprocal diversity index.

- The `Options→Compare...` item: Open compare dialog to produce a comparison of multiple samples.
- The `Options→Compute Taxonomic Profile...` menu item opens a dialog that offers algorithms for computing a taxonomic profile at a given taxonomic rank from the taxonomic binning displayed by MEGAN . This is work in progress (together with Vincent Moulton at UEA) and should be considered an experimental feature.
- The `Options→Open NCBI Web Page...` item: Open NCBI Taxonomy web site in browser.
- The `Options→Inspect...` item: Inspect the read-to-taxon assignments.

## 10.11 The Taxon Disabling menu

The `Taxon Disabling` menu contains the following items:

- The `Taxon Disabling→Enable All` item: Enable all taxa.
- The `Taxon Disabling→Disable...` item: Disable all selected taxa or all named ones.
- The `Taxon Disabling→Enable...` item: Enable all selected taxa or all named ones.
- The `Taxon Disabling→List Disabled...` item: List all disabled taxa.

## 10.12 The Layout menu

The `Layout` menu contains the following items:

- The `Layout→Show Legend` item: Show horizontal or vertical legend, or hide.
- The `Layout→Increase Font Size` item: Set the font size.
- The `Layout→Decrease Font Size` item: Decrease the font size.
- The `Layout→Expand/Contract` submenu.
- The `Layout→Layout Labels` item: Layout labels.
- The `Layout→Scale Nodes By Assigned` item: Scale nodes by number of reads assigned.
- The `Layout→Scale Nodes By Summarized` item: Scale nodes by number of reads summarized.
- The `Layout→Set Max Node Height...` item: Set the maximum node height in pixels.
- The `Layout→Zoom To Selection` item: Zoom to the selection.
- The `Layout→Fully Contract` item: Contract tree vertically.
- The `Layout→Fully Expand` item: Expand tree vertically.

- The `Layout→Draw Circles` item: Draw data as circles.
- The `Layout→Draw Pies` item: Draw data as pie charts.
- The `Layout→Draw Coxcombs` item: Draw data as coxcombs.
- The `Layout→Draw Bars` item: Draw nodes as bars.
- The `Layout→Draw Heatmaps` item: Draw data as heat maps.
- The `Layout→Linear Scale` item: Show values on a linear scale.
- The `Layout→Sqrt Scale` item: Show values on square-root scale.
- The `Layout→Log Scale` item: Show values on log scale.
- The `Layout→Cladogram` item: Draw tree as cladogram with all leaves aligned right.
- The `Layout→Phylogram` item: Draw tree as phylogram with all leaves positioned as left as possible.
- The `Layout→Use Magnifier` item: Turn the magnifier on or off.
- The `Layout→Draw Leaves Only` item: Only draw leaves.
- The `Layout→Highlight Differences` submenu.

### 10.13 The Expand/Contract menu

The `Expand/Contract` menu contains the following items:

- The `Expand/Contract→Expand Horizontal` item: Expand view horizontally.
- The `Expand/Contract→Contract Horizontal` item: Contract view horizontally.
- The `Expand/Contract→Expand Vertical` item: Expand view vertically.
- The `Expand/Contract→Contract Vertical` item: Contract view vertically.

### 10.14 The Highlight Differences menu

The `Highlight Differences` menu contains the following items:

- The `Highlight Differences→Uncorrected` item: In a comparison of exactly two samples, highlight statistically significant differences, using no correction.
- The `Highlight Differences→Holm-Bonferroni Corrected` item: In a comparison of exactly two samples, highlight statistically significant differences, using Holm-Bonferroni correction.

- The `Highlight Differences→Bonferroni Corrected` item: In a comparison of exactly two samples, highlight statistically significant differences, using Bonferroni correction.
- The `Highlight Differences→Set Highlight Color...` item: Set the pairwise comparison highlight color.

## 10.15 The Tree menu

The `Tree` menu contains the following items:

- The `Tree→Collapse` item: Collapse nodes.
- The `Tree→Collapse at Level...` item: Collapse all nodes at given depth in tree.
- The `Tree→Collapse At Taxonomic Rank` submenu.
- The `Tree→Collapse All Others` item: Collapse all parts of tree that are not above or below the selected nodes.
- The `Tree→Uncollapse` item: Uncollapse selected nodes.
- The `Tree→Uncollapse Subtree` item: Uncollapse whole subtree beneath selected nodes.
- The `Tree→Uncollapse All` item: Uncollapse all nodes.
- The `Tree→Show Names` item: Display the full names of taxa.
- The `Tree→Show IDs` item: Display the NCBI ids of taxa.
- The `Tree→Show Number of Reads Assigned` item: Display the number of reads assigned to a taxon.
- The `Tree→Show Number of Reads Summarized` item: Display the total number of hits to a taxon and its descendants.
- The `Tree→Node Labels On` item: Show labels for selected nodes.
- The `Tree→Node Labels Off` item: Hide labels for selected nodes.
- The `Tree→Show Intermediate Labels` item: Show intermediate labels at nodes of degree 2.

## 10.16 The Collapse At Taxonomic Rank menu

The `Collapse At Taxonomic Rank` menu contains the following items:

- The `Collapse At Taxonomic Rank→Super Kingdom` item: Collapse Super Kingdom.
- The `Collapse At Taxonomic Rank→Kingdom` item: Collapse Kingdom.

- The `Collapse At Taxonomic Rank→Phylum` item: Collapse Phylum.
- The `Collapse At Taxonomic Rank→Class` item: Collapse Class.
- The `Collapse At Taxonomic Rank→Order` item: Collapse Order.
- The `Collapse At Taxonomic Rank→Family` item: Collapse Family.
- The `Collapse At Taxonomic Rank→Genus` item: Collapse Genus.
- The `Collapse At Taxonomic Rank→Species` item: Collapse Species.

## 10.17 The Window menu

The `Window` menu contains the following items:

- The `Window→About...` item: About MEGAN and the authors (Windows and Linux only).
- The `Window→How to Cite...` item: Show how to cite the program.
- The `Window→Website...` item: Go to the program website.
- The `Window→License...` item: Show license window.
- The `Window→Change License...` item: Change license.
- The `Window→Message Window...` item: Open the message window.
- The `Window→Set Window Size...` item: Set the window size.
- The `Window→Inspector Window...` item: Open inspector window.
- The `Window→Show Alignment...` item: Show alignment of reads to a specified reference sequence.
- The `Window→Main Viewer...` item: Brings the main viewer to the front.
- The `Window→SEED Analyzer...` item: Opens the SEED Analyzer.
- The `Window→COG Analyzer...` item: Opens the COG Analyzer.
- The `Window→KEGG Analyzer...` item: Opens the KEGG Analyzer.
- The `Window→Sample Viewer...` item: Opens the Sample Viewer.
- The `Window→Chart...` item: Show chart of assigned reads.
- The `Window→Radial Chart...` item: Show radial space filling chart. If labels are very long and do not fit completely into the window, then enter the following command to scale the chart: `setprop RadialChartScalingFactor=`⟨*factor*⟩, where *factor* is a scaling factor between 0 and 1.

- The `Window→Word Cloud...` item: Show word cloud.
- The `Window→Comparison Plot...` item: Plot pairwise comparison of assignments to classes.
- The `Window→Chart Microbial Attributes...` item: Chart microbial attributes.
- The `Window→Cluster Analysis...` item: Open a cluster analysis window.
- The `Window→Rarefaction Analysis...` item: Compute and chart a rarefaction curve based on the leaves of the tree shown in the viewer.
- The `Window→Command-Line Syntax...` item: Shows syntax help for commands.
- The `Window→Command Input...` item: Execute a command.

## 10.18 Popup Menus

Many of the menu items listed above are also available through context-specific popup menus, which are activated by a right mouse click.

## 10.19 The Toolbar

The toolbar contains the following items:

- The `Open...` item: Open a MEGAN file (ending on .rma, .meg or .megan).
- The `Print...` item: Print the main panel.
- The `Export Image...` item: Export content of window to an image file.
- The `Find...` item: Open the find toolbar.
- The `Expand Vertical` item: Expand view vertically.
- The `Contract Vertical` item: Contract view vertically.
- The `Expand Horizontal` item: Expand view horizontally.
- The `Contract Horizontal` item: Contract view horizontally.
- The `Fully Contract` item: Contract tree vertically.
- The `Fully Expand` item: Expand tree vertically.
- The `Cladogram` item: Draw tree as cladogram with all leaves aligned right.
- The `Phylogram` item: Draw tree as phylogram with all leaves positioned as left as possible.
- The `Collapse` item: Collapse nodes.

- The `Collapse at Level...` item: Collapse all nodes at given depth in tree.
- The `Uncollapse` item: Uncollapse selected nodes.
- The `Uncollapse Subtree` item: Uncollapse whole subtree beneath selected nodes.
- The `Draw Circles` item: Draw data as circles.
- The `Draw Pies` item: Draw data as pie charts.
- The `Draw Coxcombs` item: Draw data as coxcombs.
- The `Draw Bars` item: Draw nodes as bars.
- The `Draw Heatmaps` item: Draw data as heat maps.
- The `Linear Scale` item: Show values on a linear scale.
- The `Sqrt Scale` item: Show values on square-root scale.
- The `Log Scale` item: Show values on log scale.
- The `Chart...` item: Show chart of assigned reads.
- The `Radial Chart...` item: Show radial space filling chart.
- The `Word Cloud...` item: Show word cloud.
- The `Comparison Plot...` item: Plot pairwise comparison of assignments to classes.
- The `Inspect...` item: Inspect the read-to-taxon assignments.
- The `Show Alignment...` item: Show alignment of reads to a specified reference sequence.
- The `Extract Reads...` item: Extract reads for the selected nodes.
- The `Rarefaction Analysis...` item: Compute and chart a rarefaction curve based on the leaves of the tree shown in the viewer.
- The `Chart Microbial Attributes...` item: Chart microbial attributes.
- The `Main Viewer...` item: Brings the main viewer to the front.
- The `SEED Analyzer...` item: Opens the SEED Analyzer.
- The `COG Analyzer...` item: Opens the COG Analyzer.
- The `KEGG Analyzer...` item: Opens the KEGG Analyzer.
- The `Sample Viewer...` item: Opens the Sample Viewer.
- The `Show Legend` item: Show horizontal or vertical legend, or hide.
- The `Cluster Analysis...` item: Open a cluster analysis window.

### 10.20 The MEGAN Menu

Under MacOS, there is an additional, standard menu associated with the program, called the `MEGAN` menu. As usual, this contains the `Window→About...` and `File→Quit` menu items.

### 10.21 Wheel Mouse and Special Keys

Use of a wheel mouse is recommended for zooming of graphics displayed in different windows. The default is *vertical zoom*. For *horizontal zoom*, additionally press the shift key.

To scroll the graph, either press and drag the mouse (using the right mouse button), or use the arrow keys. To zoom the graph in verticial or horizontal direct, press the shift-key while using the arrow keys. To increase the zoom factor, additionally press the alt key or the control key.

To select a region of nodes using the mouse, click, hold for a second until the cursor changes to an arrow and then drag the mouse to capture the nodes to be selected.

## 11 SEED Window

The `SEED` window is used to display a SEED analysis of gene function, based on [15]. The SEED classification is displayed as a tree. Genes are mapped onto functional roles and these are present in one or more subsystems. Modes of interaction and available menu items are similar to those of the main window.

The window is split into two panes. The right pane contains a tree-based display of the result of the SEED classification. The left pane contains two tabs, one containing a textual tree-based view and the other using a heat-map style listing of the current leaf nodes of the tree displayed in the right pane.

## 12 COG Window

The `COG` window is used to display a COG analysis of gene function, based on [15]. The SEED COG is displayed as a tree. Genes are mapped onto COGs and these are present in one or more subsystems. Modes of interaction and available menu items are similar to those of the main window.

The window is split into two panes. The right pane contains a tree-based display of the result of the COG classification. The left pane contains two tabs, one containing a textual tree-based view and the other using a heat-map style listing of the current leaf nodes of the tree displayed in the right pane.

## 13 KEGG Window

The `KEGG` window is used to display a KEGG analysis of gene function, based on [9]. The KEGG classification is displayed as a tree. Genes are mapped onto *enzymes* and these are present in one or more pathways. Modes of interaction and available menu items are similar to those of the main window.

The window is split into two panes. The right pane contains a tree-based display of the result of the KEGG classification. The left pane contains two tabs, one containing a textual tree-based view and the other using a heat-map style listing of the current leaf nodes of the tree displayed in the right pane.

Additionally, the right pane of the window is tabbed. Initially, only the tree-based display of the KEGG classification is visible. However, by double-clicking on any item in the left pane for which a KEGG-pathway diagram exists, a new*pathway tab* is opened containing the corresponding pathway. Different shades of green are used to indicate how many reads were assigned to any given enzyme of gene-product in the pathway.

Another way to open a pathway tab is to use the following menu item, which is available in the `Options` menu and from context menus associated with nodes:

- The `Options→Show KEGG Pathway...` item: Show the specified KEGG pathway.

# 14 Sample Viewer

The `Sample Viewer` provides a tabular view of all samples present in a document. The samples can have multiple attributes and these attributes can be modified. They can also be used to color the samples. Samples can be extracted or merged into *biomes* in a number of different ways.

The Sample Viewer has a number of specific menus:

## 14.1 The Attributes menu

The `Attributes` menu contains the following items:

- The `Attributes→New...` item: Create a new attribute (column) in the data table.
- The `Attributes→Import From File...` item: Import one or more attributes from a file into the data table.
- The `Attributes→Set Color...` item: Set the color for all selected item.
- The `Attributes→Set Value...` item: Set value for all selected items. item The `Attributes→Duplicate...` item: Duplicate an existing attribute (column).
- The `Attributes→Rename...` item: Rename an existing attribute (column).
- The `Attributes→Delete...` item: Delete an existing attribute (column).
- The `Attributes→Select All Same` item: Select all cells that have the same attribute and value.

## 14.2 The Samples menu

The `Samples` menu contains the following items:

- The `Samples→Node Shape` item: Set the shape used to represent the selected samples.
- The `Samples→Group Nodes` item: All selected samples will be displayed as grouped (using their convex hull) in the PCoA plot.
- The `Samples→Ungroup All` item: Ungroup all sample.
- The `Samples→Add...` item: Add samples from open document.
- The `Samples→Add From File...` item: Add samples from files.
- The `Samples→Show All` item: Show all samples.
- The `Samples→Show Selected` item: Show selected samples.
- The `Samples→Hide Selected` item: Hide selected samples.
- The `Samples→Hide Unselected` item: Hide samples.
- The `Samples→Duplicate...` item: Duplicate selected samples (rows).
- The `Samples→Rename...` item: Rename selected samples (rows).
- The `Samples→Delete...` item: Delete an existing sample (row).
- The `Samples→Move Up` item: Move Up.
- The `Samples→Move Down` item: Move samples down.
- The `Samples→Apply Reordering To Viewers` item: Reorder samples in all viewers as currently listed in table.
- The `Samples→Set Color...` item: Set the color for all selected samples.
- The `Samples→Color By Attribute` item: Color samples by attribute states.

## 14.3 The Algorithms menu

The `Algorithms` menu contains the following items:

- The `Algorithms→Extract Samples...` item: Extract selected samples to a new document.
- The `Algorithms→Compute Core Biome...` item: Determine taxa and functions that appear in a majority of the samples.
- The `Algorithms→Compute Total Biome...` item: Determine total (union) taxonomic and functional content.

- The `Algorithms→Compute Rare Biome...` item: Determine taxa and functions that appear in a minority of samples.
- The `Algorithms→Compute Shared Biome...` item: Determine shared taxonomic and functional content of samples (i.e. intersection).
- The `Algorithms→Resample...` item: Resample selected samples to a new document.

# 15 Cluster Viewer

The `Cluster` viewer provides methods for comparing multiple samples. It can be opened for any comparison document containing at least four samples. The Cluster viewer allows one to compute a distance matrix on the set of samples, based either on their taxon profiles, or based on their SEED, COG or KEGG profiles. The viewer provides a number of different ecological indices to compute the distances. By default the distances are based on the leaves of the corresponding tree, however, if some nodes of the tree are selected, then only those nodes are used in the calculation.

The calculated distances are displayed as a PCoA plot (principle coordinates analysis), a hierarchical clustering (UPGMA tree), an unrooted tree (Neighbor-Joining tree) or an unrooted split network (Neighbor-net), see [12] for details.

MEGAN also computes and displays biplot vectors in a *PCoA plot*. These vectors indicate which taxa or functional groups have the largest influence in PCoA plot. MEGAN is able to show both a two dimensional PCoA and a three dimensional PCoA plot.

When viewing a three dimensional PCoA plot, click and drag the mouse to roll the plot around. To rotate it, in addition, press the control key. To slide it, in addition, press the alt key.

The cluster viewer has a number of specific menus:

## 15.1 The Cluster Viewer Layout menu

The `Cluster Viewer Edit` menu contains the following additional items:

- The `Edit→Node Shape` sub menu is used to set the shape of the selected nodes.
- The `Edit→Group Nodes` item group all selected node; this is indicated by drawing their convex hull in the PCoA plot.
- The `Edit→Ungroup All` ungroups all nodes.
- The `Edit→Show Groups` is to used to turn the representation of groups on or off.

The `Cluster Viewer Layout` menu contains the following items:

- The `Layout→Show Legend` item: Show horizontal or vertical legend, or hide.
- The `Layout→Increase Font Size` item: Set the font size.

- The `Layout→Decrease Font Size` item: Decrease the font size.
- The `Layout→Expand/Contract` submenu.
- The `Layout→Zoom to Fit` item: Zoom to fit.
- The `Layout→Flip Horizontally` item: Flip horizontally.
- The `Layout→Flip Vertically` item: Flip vertically.
- The `Layout→Use Colors` item: Use colors.
- The `Layout→Show Labels` item: Show node labels.
- The `Layout→Set Node Radius...` item: Set node radius.
- The `Layout→PC1 vs PC2` item: Set principle components to use.
- The `Layout→PC1 vs PC3` item: Set principle components to use.
- The `Layout→PC2 vs PC3` item: Set principle components to use.
- The `Layout→PCi vs PCj...` item: Set principle components to use.
- The `Layout→PC1 PC2 PC3` item: Display three-dimensional PCoA plot using first three components.
- The `Layout→PCi PCj PCk` item: Display three-dimensional PCoA plot using three selected components.
- The `Layout→Show BiPlot` item: Turn the display of biplot vectors on or off.
- The `Layout→BiPlot Size...` item: Determine the number of biplot vectors to display.

## 15.2 The Cluster Viewer Options menu

The `Cluster Viewer Options` menu contains the following items:

- The `Options→Use Normalized Goodall` item: Use normalized Goodall's ecological index.
- The `Options→Use Goodall` item: Use Goodall's ecological index.
- The `Options→Use Chi-Square` item: Use ChiSquare ecological index.
- The `Options→Use Kulczynski` item: Use Kulczynski ecological index.
- The `Options→Use Bray-Curtis` item: Use Bray-Curtis ecological index.
- The `Options→Use Hellinger` item: Use Hellinger ecological index.
- The `Options→Use Euclidean` item: Use Euclidean ecological index.
- The `Options→Use Pearson` item: Use Pearson's correlation distance.
- The `Options→Use JSD` item: Use square root of Jensen-Shannon divergence [2].
- The `Options→Sync` item: Sync view of data.

### 15.3 The Cluster Viewer View menu

The `Cluster Viewer View` menu contains the following items:

- The `View→PCoA` item: Open the PCoA tab.
- The `View→UPGMA Tree` item: Open the UPGMA tree tab.
- The `View→NJ Tree` item: Open the NJ tree tab.
- The `View→Network` item: Open the network tab.
- The `View→Matrix` item: Open the matrix tab.

## 16 Import Dialog

The `Import Blast` dialog is used to import new data from BLAST (or a similar tool) and to create a new RMA file. The dialog has a number of tabbed panes.

### 16.1 Files Tab

The *Files* tab is used to setup the location of the input and output files. The program allows one to open more than one BLAST file or reads files, for the case that reads and matches are spread across multiple files. The first item is used to specify the location of the BLAST file or similar comparison file. Once one the file has been specified, MEGAN will attempt to detect the type of the file provided. If the program is unsuccessful at this, then use the dropdown menu to set the file type. The second item is used to specify the location of the reads file. If the reads are from a paired-read project, then selecting the `Paired reads` check box will request MEGAN to perform a paired-read analysis (see [13]). The third item is used to specify the location of the new RMA file. The *Max number of matches per read* item specifies how many matches per read to save in the RMA file. A small value will reduce the size of the RMA file, but may exclude some important matches. By default, the 100 highest scoring matches per read are saved.

Once this information has been collected, the user should review the other panels, as described below, before pressing the *Apply* button to import the data.

### 16.2 Taxonomy Tab

The *Taxonomy Tab* is used to specify how MEGAN identifies taxa in a BLAST or similar file. By default, the program attempts to parse taxon names. Additionally, a mapping file that maps RefSeq ids to taxon ids can be used, or a file that maps GI accession numbers to taxon ids, or a file that maps arbitrary strings (“synonyms”) to taxon ids. The format of the mapping files is as follows: each line contains two items, separated by a tab. The files should end on `.map`, `.txt` or `.gz`. In additional, MEGAN5 can process indexed GI mapping files that are available from the MEGAN5 website. These files end on `.bin` and contain a binary index that is not read into memory.

## 16.3 SEED Tab

The *SEED Tab* is used to specify how MEGAN identifies SEED functional roles in a BLAST or similar file. By default, the program uses a built-in RefSeq to SEED mapping. Other options are an external RefSeq ids to SEED id mapping, or a file that maps GI accession numbers to SEED ids, or a file that maps arbitrary strings ("synonyms") to SEED ids. Note that the SEED ids used by MEGAN are not official ids. See Taxonomy Tab for a brief description of file formats.

## 16.4 COG Tab

The *COG Tab* is used to specify how MEGAN identifies COG classes in a BLAST or similar file. By default, the program uses a built-in RefSeq to COG mapping. Other options are an external RefSeq ids to COG id mapping, or a file that maps GI accession numbers to COG ids, or a file that maps arbitrary strings ("synonyms") to COG ids. See Taxonomy Tab for a brief description of file formats.

## 16.5 KEGG Tab

The *KEGG Tab* is used to specify how MEGAN identifies KEGG KO groups in a BLAST or similar file. By default, the program uses a built-in RefSeq to KO mapping. Other options are an external RefSeq ids to KO mapping, or a file that maps GI accession numbers to KOs, or a file that maps arbitrary strings ("synonyms") to KOs ids. See Taxonomy Tab for a brief description of file formats.

## 16.6 LCA Parameters Tab

The *LCA Parameters Tab* contains all items of the `Parameters` dialog which allows one to set the parameters used by the LCA algorithm. Because re-computation of an analysis at a later stage can take quite long on a very large sample, it is recommended to set these values appropriately before starting the import process.

## 16.7 Notes Tab

The *Notes Tab* allows the user to enter a one-line description of the sample. This will appear as a `Description` attribute in the samples `Sample Viewer`.

## 16.8 Advanced Tab

The *Advanced Tab* controls how MEGAN parses files how much data MEGAN saves to a file.

MEGAN supports three different *text storage policies*. Select `Save in main file` to have all reads and BLAST matches embedded in the computed RMA file. This provides best portability of files. If the `Save in separate file` button is selected, then all reads and matches are stored in a separate *RMAZ* file. In this case, the RMA file will be much smaller and can be used independently of the RMAZ file, unless one wants to access the reads or matches, in which case the RMAZ file will be asked for. Finally, if the `Don't save` button is selected, reads and matches are not stored explicitly. If they are requested by the user, then the program will obtain them from the original files. This mode leads to the smallest RMA files and shortest computation time, but is less portable. This mode cannot be used when either the BLAST files or reads files are compressed.

By default, MEGAN5 parses through BLAST files and the corresponding DNA sequence files simultaneously so as to associate read sequences and BLAST matches with each other on the fly. This is faster and less memory intensive that is implemented in MEGAN4. This processing requires that the sequences and matches associated with reads appear in the same order in the FastA and BLAST files. If this requirement is not fulfilled, then one can select the `Load All Reads Into Memory` checkbox to make MEGAN5 load all DNA reads into memory and then to lookup the appropriate sequences when parsing the BLAST files (which is how MEGAN4 handled parsing).

Please note that in either case, all reads must have unique names (where the name of a read is the first word of hits header line) and all BLAST matches associated with a given read must appear consecutively within a BLAST file. If the matches associated with a given read are located in different parts of a file or even in different files, then MEGAN5 will be able to handle this gracefully and a messy error may occur.

# 17 Inspector Window

The `Inspector` Window can be used to inspect the alignments that are the basis of the assignment of reads to taxa. It can be opened either using the `Window→Inspector Window...` menu item or by right-clicking on a taxon and then selecting the `Inspect` popup item. This window displays data hierarchically using a data tree. The root node of this tree represents the current input file. This window can only be opened when data has been loaded into the program.

Any taxon added to the window, either by right-clicking a taxon and then selecting the `Inspect` popup item in the main viewer, or by using the `Options→Show Taxon` item, is shown at a second level below the root. Clicking on such a *taxon node* will open a new level of nodes, each *read node* representing a read that has been assigned to the named taxon. Clicking on a read node will then open a new level of nodes, each such *read hit node* representing an alignment of the given read to a sequence associated with some taxon. Finally, double-clicking on a read hit node will display the actual BLAST alignment provided to deduce the relationship.

## 17.1 Inspector Menus

Here we describe those menu items that are different from the main window.

## 17.2 The Inspector Edit Menu

The `Edit` menu contains the following item:

- The `Edit→Show Taxon...` item: Show the named taxon and all reads assigned to it.

### 17.3 The Inspector Options Menu

The `Inspector Options` menu contains the following items:

- The `Options→Collapse` item: Collapse the selected nodes, or all, if none selected.
- The `Options→Expand` item: Expand the selected nodes, or all, if none selected.
- The `Options→Use Hit` item: Use all selected hits.
- The `Options→Use All Hits` item: Use all hits.

## 18 Microbial Attributes Window

The `Chart Microbial Attributes...` is used to analyze the composition of microbial taxa (Bacteria and Archeae) and their various physiological features. Taxa have to be assigned with at least one read to be considered. The classification and its nomenclature is based on the NCBI's prokaryotic attribute table (derived from: http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi ).

## 19 Rarefaction Window

The `Rarefaction` Window be used to compute and draw a species rarefaction plot. This operates by repeatedly sampling subsets from a set of reads and computing the number of leaves to which taxa have been assigned. This analysis uses the current leaves of the taxonomy, in other words collapsing or uncollapsing nodes will lead to a different result. There are two properties that can be set: Use `setprop NumberRareFactionDataPoints=`⟨*number*⟩ to set the number of datapoints to be plotted for each sample. Use `setprop NumberRareFactionReplicates=`⟨*number*⟩ to set the number of replicates to be used for each datapoint.

## 20 Chart Window

There is one chart window per document per type of analysis.

### 20.1 Taxon Chart Window

The `Taxon Chart Window` is used to visualize the abundance distribution of the taxa as a *pie chart* , *bar chart*, *graph*, *radial space filling tree*, *heat map* or *co-occurrence network*.

It can be opened using the `Window→Chart...` menu item. To present a specific set of taxa in the chart, select them in the main window and then press the `sync` button in the chart window tool bar.

The co-occurrence network is controlled by a number of parameters. The *threshold* sets the minimum count required for a taxon to be considered present in a sample. The *min prevalence* and *max prevalence* parameters are used to set the minimum and maximum percentage of samples in which a taxon can occur, respectively, so as to have the taxon to be presented by a node in the network. The *probablity* sets parameter defines the minimum probability that a co-occurrence between two taxa $A$ and $B$ must attain so as to be represented by an edge in the graph. The *probability of coocurrence* is computed as the number of samples that contain both $A$ and $B$, divided by the number number of samples that contain either $A$ or $B$.

### 20.2 SEED Chart Window

The `SEED Chart Window` is used to visualize the abundance distribution of the SEED classes, as in the `Taxon Chart Window`.

### 20.3 KEGG Chart Window

The `KEGG Chart Window` is used to visualize the abundance distribution of the KEGG classes, as in the `Taxon Chart Window`.

### 20.4 COG Chart Window

The `COG Chart Window` is used to visualize the abundance distribution of the COG classes, as in the `Taxon Chart Window`.

## 21 Alignment Viewer

The `Alignment Viewer` is used to compute and visualize a multiple sequence alignment of all reads that have significant matches to a reference sequences associated with a given taxon, SEED class or KEGG class. It can be opened using the `Window→Show Alignment...` menu item. Here is an overview of the menus available in this viewer and those menu items that do not appear in the main viewer.

The File menu contains the following items:

- The `File→Export Alignment...` item: Save alignment to a file.
- The `File→Export Contigs...` item: Assemble reads using exact overlaps (only) and save the contigs to a file.
- The `File→Export Reference...` item: Save reference sequence to a file.
- The `File→Export Consensus...` item: Save consensus to a file.

The minimum overlap required to assemble two reads in to an alignment is 20 bp. At present, this is a program property (i.e., not document specific). It can be changed to 40, for example, by entering the command `setprop alignment-min-overlap=40` using the `Window→Command Input...` window.

The Edit menu contains the following items:

- The `Edit→Copy Alignment` item: Copy selected part of the alignment.
- The `Edit→Copy Consensus` item: Copy selected consensus sequence to clipboard.
- The `Edit→Copy Reference` item: Copy selected reference sequence to clipboard.
- The `Edit→Translate...` item: Translate DNA or cDNA sequence.

The Options menu contains the following items:

- The `Options→Show As Alignment` item: Show as alignment.
- The `Options→Show As Mapping` item: Show alignment as mapping.
- The `Options→Unsorted` item: Do not sort sequences.
- The `Options→Sort By Names` item: Sort sequences by names.
- The `Options→Sort By Start` item: Sort sequences by start positions.
- The `Options→Sort By Contigs` item: Assemble reads and then sort by contigs.
- The `Options→Move Up` item: Move selected sequences up.
- The `Options→Move Down` item: Move selected sequences down.
- The `Options→Sort By Similarity` item: Sort sequences by pairwise similarity.
- The `Options→Set Amino Acid Colors...` item: Set the color scheme for amino acids.
- The `Options→Color Matches` item: Color letters that match the reference sequence.
- The `Options→Color Mismatches` item: Color letters that do not match the reference sequence.
- The `Options→Chart Diversity` item: Opens a chart showing a “diversity analysis” that aims at estimating the number of distinct genomic sequences corresponding to a given gene. Using a sliding window (default length 25) along the reference sequence, the program records the total number $n$ of reads that cover the window and the number $k$ of such reads that have distinct sequences over the window. These are then depicted in a scatter plot. Using a simple function based on Michaelis-Menten kinetics [20], the program plots a curve for the data that is used to estimate the number of different genomes involved.

The Layout menu contains the following items:

- The `Layout→Show Insertions` item: Show insertions in reads.
- The `Layout→Contract Gaps` item: Contract all columns consisting only of gaps.

- The `Layout→Show Nucleotides` item: Show nucleotides in alignment.
- The `Layout→Show Amino Acids` item: Show amino-acids in alignment.
- The `Layout→Show Reference` item: Show reference sequence.
- The `Layout→Show Consensus` item: Show consensus sequence.
- The `Layout→Show Unaligned` item: Show the unaligned prefix and suffix of reads.

By default, the alignment viewer only shows as many positions as covered by alignments. To show additional positions at the end of the alignment, enter the following command `setprop alignmentViewerAdditionalPositions=`⟨number⟩ using the `Window→Command Input...` menu item.

# 22 Group Viewer

The `Group Viewer` is used to define or display groups of samples. Currently, groups are of limited use in MEGAN . This will change in future releases.

# 23 Find Toolbar

The `Find` toolbar can be opened using the `Edit→Find...` item. Its purpose is to find taxa, genes or other strings in a window. Use the following check boxes to parameterize the search:

- If the `Whole words only` item is selected, then only taxa or reads matching the complete query string will be returned.
- If the `Case sensitive` item is selected, then the case of letters is distinguished in comparisons.
- If the `Regular Expression` item is selected, then the query is interpreted as a Java regular expression.
- Use the `From File...` item to read a list of search terms from a text file, one per line.
- In the taxonomy viewer, use the `Uncollapse` item to allow MEGAN to search for taxa among currently collapsed nodes.

Press the `Close`, `Find First` or `Find Next` buttons to close the toolbar, or find the first, or next occurrence of the query, respectively. Press the `Find All` button to find all occurrences of the query.

Press the `From File` button to load a set of queries, one per line, from a file.

## 24 Format Dialog

The `Format` dialog is opened using the `Edit→Format...` item. This is used to change the font, color, size and line width of all selected nodes and edges. Also, it is used to turn labels on and off.

## 25 Message Window

The `Message` window is opened using the `Window→Message Window...` item. The program writes all messages to this window. The window contains the usual File and Edit menu items.

## 26 Parameters Dialog

The `Parameters` dialog is used to control the parameters of the LCA-assignment algorithm. It can be invoked by selecting `Options→Change LCA Parameters...` . The dialog options are:

- The `Min Score` item is used to set a minimum threshold for the bit score of hits. Any hit in the input data that scores less than the given threshold is ignored.
- The `Max Expected` item is used to set a maximum threshold for the expected value of hits. Any hit in the input data whose *E-value* exceeds this value is ignored.
- The `Top Percentage` item is used to set a threshold for the maximum percentage by which the score of a hit may fall below the best score achieved for a given read. Any hit that falls below this threshold is discarded.
- The `Min Support Percent` item is used to set a threshold for the minimum support that a taxon requires, as a percentage of assigned reads. This feature is turned off by setting the value to 0. If a value greater than 0 (and at most 100) is given, then the program will set the `Min Support` threshold appropriately (see next item).
- The `Min Support` item is used to set a threshold for the minimum support that a taxon requires, that is, the number of reads that must be assigned to it so that it appears in the result. Any read that is assigned to a taxon that does not have the required support is pushed up the taxonomy until a node is found that has sufficient support (version 3.10 onward, earlier versions counted the read as *unassigned*).
- The `LCA Percent` item is used to set the percent of matches that the LCA of a read must cover, in the range 50-100. When a value of less than 100 is specifiedm then the LCA of a fixed percent is used. The `Min Complexity` item is used to identify low complexity reads [21, 14]. These are placed on a special *Low Complexity* node. To turn this filter off, set the value to 0. A value of 0.3 catches most low complexity short reads.
- The `Paired Reads` item is used to turn paired-read awareness of MEGAN on and off. In paired-read mode, MEGAN utilities read-pairing information to enhance the taxonomic assignment of reads.

- The `Use Use Minimum Coverage Heuristic` item turns on the minimum coverage heuristic.
- The `Use 16S Percent Identity Filter` item is used to turn on an additional filter for assigning reads to a specific taxonomic level. When this is active, the percent identity of a match must exceed the given value of percent identity to be assigned at the given rank: Species 99%, Genus 97%, Family 95%, Order 90%, Class 85%, Phylum 80%. This should only be used when analyzing 16S rRNA sequences.

## 27 Compare Dialog

The `Compare` dialog is opened using the `Options→Compare...` item. This dialog provides a list of currently open samples. To construct a comparison, select at least two different samples and then press "ok". Select `Use Absolute Counts` , if you want the comparison the original counts of reads for each sample. Select `Use Normalized Counts` , if you want all counts to be normalized to the smallest number of reads of any of the selected samples. Select `Use Square Root Normalization`, if you want to normalize counts by taking their square-root (recommended by statistician Susan Holmes). Select `Use Sub-Sampled Counts` , if you want all counts to be randomly subsampled download to the smallest count of any of the given samples. Each sample is subsampled 1000 times and then the rounded average counts are used. To globally change the number of replicates used, enter the following command `setprop subSampleReplicates=`⟨number⟩ using the `Window→Command Input...` menu item. Select `Ignore all unassigned reads` , if you want all reads assigned to the three special nodes labeled 'Not Assigned', 'No Hits' and 'Low Complexity' (if present) to be ignored.

## 28 Extractor Dialog

This provides an alternative to the `Export→Reads...` item which allows to save reads from different taxa to files whose names contain the taxon name.

The `Extractor` dialog is opened using the `File→Extract Reads...` item. The dialog is used to extract all reads assigned to selected nodes. For any selected nodes, all reads assigned to it, or to *any node below* it in the hierarchy, are saved to a file.

Use the Browse button to specify the output directory. As the MacOS X dialog does not support the selection of a dialog, select any file inside the desired target directory. Specify the file name for output in the `File name` field. If the name contains `%t`, then the program will produce one output file per node, and the name of the file is generated by replacing `%t` by the node name. Otherwise, all reads are written to one file.

## 29 Export Image Dialog

The`Export Image` dialog is opened using the `File→Export Image...` item. This dialog is used to save a picture of the current tree in a number of different formats, see Section 32.6.

The format is chosen from a menu. There are two radio buttons `Save whole image` to save the whole image, and `Save visible image` to save only the part of the image that is currently visible in the main viewer. If the chosen format is EPS, then selecting the `Convert text to graphics` check box will request the program to render all text as graphics, rather than fonts.

Pressing the apply button will open a standard file save dialog to determine where to save the graphics file.

## 30 About Window

The `About` Window is opened using the `Window→About...` item. It shows the program's splash screen.

## 31 Server Browser Window

The `Server Browser` window is opened using the `File→Open From Server...` menu item.

The window is used to connect MEGAN to one or more *MeganServer* instances, allowing the user to open MEGAN files that are stored remotely on a server.

MeganServer is a light-weight server program, written by Hans-Joachim Ruscheweyh, that runs on a server and provides access to selected RMA files on the server via a REST interface. (See the documentation for MeganServer for details).

Use the `Server Browser:Open remote server` tab to connect to a remote server.

For example, set the server to `http://megan-db.org/Permafrost`, and the user and password both to `guest` to access 12 MEGAN RMA files computed for the samples described in [11].

Once an instance of MeganServer has been successfully contacted, a directory tree will be displayed that mirrors the directory structure of the server. Double-clicking on a file name will open that file. To open a comparison document for multiple files, select the files of interest and then press the `Compare` button.

MEGAN will remember previously contacted MeganServer instances and thus previously opened files can be directly reopened using the `Open Recent` menu.

## 32 File Formats

MEGAN uses its own file formats to store the data describing the result of a sequence comparison computation between a file of DNA reads and a database of reference sequences, such as computed by BLASTX, BLASTP or BLASTN [1].

### 32.1 RMA Files

Files ending in `.rma` are in a compressed binary format called RMA (read-match archive), which is a new open format that we will describe in a separate document. MEGAN 1 used a text format (files ending on `.megan` or `.meg`), which are now deprecated and will not be supported by futher versions of the program. By convention, we use the suffix `.megan` for MEGAN text files and `.rma` for binary read-match archive files.

With MEGAN 4, we have introduced a new version of the RMA format, internally known as RMA 2. This format is more flexible, as it does not necessarily need to contain all reads and matches. Moreover, it has better locality and thus updating it is much faster. MEGAN 4 can read RMA files generated by earlier versions of MEGAN, but not vice-versa.

A *RMA file* is generated using the `File→Import From BLAST...` menu item from a BLAST file and a *read file* (or from multiple files, if the reads are spread across multiple files). Depending on which of the three possible text storage policies is chosen, the RMA file may contain all reads and matches in a compressed form, or these may be stored in a separate RMAZ file, or otherwise only links to there locations in the original reads and match files are stored.

In the first case, the size of such a file is around 10-20% of the size of the original input files and is thus usually smaller than the file that one obtains by simply compressing the BLAST file. The file is indexed and thus provides MEGAN with fast access to data stored in it. The reads and matches can be extracted from the file and so the MEGAN file provides a means of keeping all reads, BLAST matches and analysis in one document.

To accommodate the every growing size of datasets, MEGAN now also uses a new version 3 of the RMA format, when using MEGAN together with DIAMOND or MALT.

### 32.2 The Text File Summary Format

As of version 4, the *MEGAN text file* format has been completely rewritten to accommodate SEED and KEGG classifications.

A MEGAN file starts with a number of header lines, each starting with a . These lines can occur in any order. This is best illustrated by an example:

```
 1 @Creator        MEGAN (version 4.0alpha20, built 14 Oct 2010)
 2 @CreationDate   Wed Oct 27 17:10:52 CEST 2010
 3 @ContentType    Summary4
 4 @Names  155_PE_1_fixed-paired   ecoli-testrun-2000-nr
 5 @Uids   1288068180866   1288190195887
 6 @Sizes  51246   2000
 7 @TotalReads     200000
 8 @Collapse       SEED    2000041
 9 @Algorithm      Taxonomy        tree-from-summary
10 @Parameters     normalizedTo=100000
11 @NodeStyle      KEGG    piechart
```

The first and second lines are optional descriptions of who generated the file when. The third line identifies the format as Summary4, indicating that this is a summary file in the format introduced with MEGAN 4. The fourth line lists the names of all samples that are represented by this file. In this case there are two. Line 5 of this example lists the unique identifier numbers associated with the samples, if any. Line 6 lists the original sizes of the samples. Line 7 lists the total number of reads. This is not necessary the sum of the original sizes. Line 8 specifies, for the SEED classification, which nodes are to be collapsed in the visual representation of the classification. The keyword SEED can be replaced by TAXONOMY or KEGG for the other classifications. Line 9 contains the name of the algorithm used to compute a classification. The second word here is a keyword to identify which classification is meant. Line 10 lists parameters of the computation used to generate the file. Line 11 specifies the style used to draw nodes in a given classification, in this case KEGG.

The main body of a MEGAN text file contains multiple lines as follows:, in any order:

```
TAX     199310  0       1250
TAX     1       271     100
TAX     28216   35
TAX     32523   8
TAX     2       8336    1350
KEGG    7716    12
KEGG    3859    2
KEGG    7714    2       100
SEED    54      6       50
```

The general format is *classification*, *count-1*, *count-2*, ..., *count-n*. Here, *classification* is either `TAX` for taxonomy, `SEED` or `KEGG`. This is followed by a number indicating a class in the given classification. In the case of taxonomy, this is the NCBI taxonId. This is followed by up to $n$ numbers, where $n$ is the number of samples mentioned in the header, indicating how many reads in the 1-st, 2-nd etc sample were assigned to the given class.

Because this format is also embedded in RMA files to provide a summary of the data, when opening an RMA file generated by an earlier version of MEGAN, the program automatically updates the summary to the new format. As a consequence, any RMA file that has been opened by MEGAN 4 cannot later be opened by an earlier version of the program.

## 32.3 Required Syntax of BLAST Files

MEGAN imports data from a *BLAST file*. MEGAN can parse BLAST files in standard or XML format obtained using the BLAST output option `-m 0` or `-m 7`, respectively. MEGAN can also parse tabular format (BLAST output option `-m 8`). For this to work, the subject field must either contain taxon names or GI accession numbers. In the latter case, please use the `Load GI-Lookup File` button to load a GI lookup file. Alternatively, a nine column may be supplied that contains either the taxon name or taxon Id associated with the database sequence. The program also scans the subject field for RefSeq identifiers to determine the associated gene.

MEGAN can read *gzipped BLAST files* and *zipped BLAST files*.

For human readable format, any *BLASTX file* or *BLASTP file* is expected to adhere to the format shown in Figure 1. Any *BLASTN file* is expected to adhere to the format shown in Figure 2.

`BLASTX` text
text...

followed by 0 or more blocks of the following type:

> `Query=` ⟨*query-id*⟩ text `length=`⟨*length*⟩ text
> *or*
> `Query=` ⟨*query-id*⟩ text`|length|`⟨*length*⟩`|`text
> text...
>
> followed by 0 or more blocks of the following type:
>
> > `>` text `[`⟨*NCBI-taxon-name*⟩`]` text
> > (line breaks ok)
> >
> > `Score = `⟨*score*⟩ `bits` `(`⟨*bits*⟩ `)` `Expect = `⟨*e-value*⟩
> > `Identities = ` text `(`⟨*percent-identities*⟩`%)`
> > `Positives = ` text `(`⟨*percent-positives*⟩`%)`, `Gaps = ` text `(`⟨*percent-gaps*⟩`%)`
> > `Frame = ` ⟨*frame*⟩
> > followed by 0 or more blocks of the following type:
> >
> > > `Query` text
> > > text
> > > `Sbjct` text

Figure 1: The required structure of a BLASTX file. Labels shown as `label` are tokens that must occur verbatim in the file. Labels shown as ⟨*label*⟩ are values that are read into the program. The first word in the file must be `BLASTX`. The header line starting with `Query =`, which is taken from the Fasta header of the query sequence (a read), must start with a one word unique identifier for the read and must also contain a statement containing the length of the read, in the format `length=`⟨*length*⟩, or as `|length|`⟨*length*⟩`|`. Another important feature is that the comment line of the database sequence must contain a NCBI-taxon name. If names are not contained in the comment lines, then the accession lookup support must be used. Finally, the `Gaps=` statement is optimal.

> `BLASTN` text
> text...
>
> followed by 0 or more blocks of the following type:
>
> > `Query=` ⟨*query-id*⟩ text `length=`⟨*length*⟩ text
> > *or*
> > `Query=` ⟨*query-id*⟩ text`|length|`⟨*length*⟩`|`text
> > text...
> >
> > followed by 0 or more blocks of the following type:
> >
> > > `>` text ⟨*NCBI-taxon-name*⟩ text
> > > (line breaks ok)
> > >
> > > `Score =`⟨*score*⟩ `bits` `(`⟨*bits*⟩ `)` `Expect =`⟨*e-value*⟩
> > > `Identities =` text `(`⟨*percent-identities*⟩`%)` `Gaps =` text `(`⟨*percent-gaps*⟩`%)`
> > > `Strand=` ⟨*strand*⟩ `/` ⟨*strand*⟩
> > >
> > > followed by 0 or more blocks of the following type:
> > >
> > > > `Query` text
> > > > text
> > > > `Sbjct` text

Figure 2: The required structure of a BLASTN file. Labels shown as `label` are tokens that must occur verbatim in the file. Labels shown as ⟨*label*⟩ are values that are read into the program. The first word in the file must be `BLASTN`. The header line starting with `Query=`, which is taken from the Fasta header of the query sequence (a read), must start with a one word unique identifier for the read and must also contain a statement containing the length of the read, in the format `length=`⟨*length*⟩. Another important feature is that the comment line of the database sequence must contain a NCBI-taxon name. If names are not contained in the comment lines, then the accession lookup support must be used.

## 32.4 How MEGAN Parses Taxon Names

MEGAN uses the following algorithm to determine the taxon from the header line of a reference sequence. If the string consists only of an integer, then this is interpreted as a taxon id. Otherwise, if `Use Synonyms` is turned on, then MEGAN attempts to match an entry in the given synonyms file. A *synonyms file* defines a mapping between words that appear in the match header and taxa or other classes. The longest matching synonym is used to determine the taxon. Otherwise, if `Use GI Lookup` is turned on, then MEGAN searches for an occurrence of the string `gi|` followed by a number and tries to use the number as a *GI accession* to determine the taxon.

Otherwise, if the header line contains a semi-colon, then MEGAN assumes that a list of taxon names is given, e.g. `Bacteria;Proteobacteria; Alpha proteobacteria`, as present, for example, in the Silva database. In this case, MEGAN uses the right-most name to determine the taxon id.

Otherwise, if the header line contains the text `/TAXON_ID=`, then MEGAN will attempt to read a taxon id following the text. This syntax is used in BLAST files obtained from the CAMERA website.

Otherwise, MEGAN searches for all pairs of disjoint square brackets and attempts to parse the strings between such brackets to obtain a set of taxon ids. The taxon id for the match is then set to the LCA of the ids. (In the NCBI-NR database, names of taxa are placed between square brackets.)

Otherwise, MEGAN searches for maximal and non-overlapping substrings that can be mapped onto an NCBI taxon id. Again, the taxon id of the match is set to the LCA.

Otherwise, the taxon is set to `unknown''`.

## 32.5 Required Format of Read Files

Reads from sequencing are assume to be provided in *fasta format* or *fastq format* in a *reads file* . The file may be gzipped. The first word of a FastA header is assumed to be the read-id.

## 32.6 Graphics Formats

The following graphics formats are supported:

- BMP, “Bitmap”.
- EPS, “Encapsulated PostScript”, vector format.
- GIF, “Graphics Interchange Format”.
- JPEG, “Joint Photographic Experts Group”.
- PDF, “Portable Document Format”, vector format.
- PNG, “Portable Network Graphics”.
- SVG, “Scalable Vector Graphics”, vector format.

## 32.7 CSV File Format

MEGAN supports importing data from other programs in a comma-separated format from a *CSV file*, using the `Import→CSV Format...` menu item. The input file must be a text file in which either all lines each contain two strings that are separated by a comma. or all lines each contain three strings separated by commas.

**Importing read assignments** If each line of the CSV file contains two strings separated by a comma, then the first string will be intepreted as a taxon name or taxon id and the second string will be intepreted as an integer specifying the number of reads assigned to the named taxon. MEGAN will assume that this is the result of some analysis and thus will produce a summary file from it and will simply display it on the NCBI taxonomy with no further analysis.

For example, assume that you have performed a metagenome analysis using some other method and have obtained the following result:

- Gammaproteobacteria: 55 reads
- Mollicutes: 400 reads
- Escherichia coli K-12: 42 reads
- Unknown: 100 reads

To import this data into MEGAN so as to visualize the taxonomical assignments, produce the following CSV file:

```
Gammaproteobacteria, 55
Mollicutes, 400
Escherichia coli K-12, 42
Not assigned, 100
```

MEGAN will draw a tree with four nodes, one for each of the named taxa.

**Importing read matches** Otherwise, if each line of the CSV file contains three strings separated by a comma, the first string will be interpreted as a read id, the second one as a taxon name or id and the third one will be interpreted as a bit score for this assignment. MEGAN will assume that this data describes a collection of reads and their matches. This data will be analysed using the LCA algorithm and the result will be displayed on the NCBI taxonomy.

For example, assume that you have done a database search using some other method than BLAST and have obtained the following result:

- The read r01 matches *Escherichia coli CFT073* with a bitscore of 100,
- The read r01 matches *Escherichia coli K-12* with a bitscore of 110, and
- The read r01 matches *Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67* with a bitscore of 120.

- The read r02 matches *Caldicellulosiruptor saccharolyticus DSM 8903* with a bitscore of 90.

To import this data into MEGAN so as to analyze is using the LCA algorithm, produce the following CSV file:

```
r01, Escherichia coli CFT073, 100
r01, Escherichia coli K-12, 110
r01,  Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67,120
r02, Caldicellulosiruptor saccharolyticus DSM 8903, 90
```

MEGAN can also import SEED or KEGG counts. In addition, MEGAN is able to map entries consisting of a RefSeq Id and counts to KEGG or SEED.

### 32.8 Tree and Map Format

The NCBI taxonomy is loaded by MEGAN at startup. It is contained in a *NCBI tree file* in the standard Newick tree format. The mapping from taxon-IDs to taxon names is loaded by MEGAN at startup. It is contained contained in a *NCBI mapping file* in a line based format in which each has three entries: taxon-ID, taxon name and then a number indicating the size of the genome, or -1, if the size is unknown.

## 33 Command-Line Options

MEGAN has the following *command-line* options:

```
 Mode:
-g, --commandLineMode            Run MEGAN in command-line mode. (Default value: false)
 Input:
-f, --files [string(s)]          MEGAN file(s) to open.
 Commands:
-x, --execute [string]           Command to execute at startup (do not use for multiple command
-c, --commandFile [string]       File of commands to execute in command-line mode.
 Configuration:
-E, --quitOnException            Quit if exception thrown in command-line mode. (Default value:
-p, --propertiesFile [string]    Alternate properties file.
+w, --hideMessages               Hide message window. (Default value: false)
-L, --licenseFile [string]       Specify alternate license file.
-V, --version                    Show version string. (Default value: false)
-S, --silentMode                 Silent mode. (Default value: false)
-d, --debug                      Debug mode. (Default value: false)
+s, --hideSplash                 Hide startup splash screen. (Default value: true)
-MC, --maxCores [number]         Max number of cores (usually ignored). (Default value: 4)
-B, --bufferSize [number]        Write buffer size in bytes. (Default value: 10240000)
 Other:
```

```
-v, --verbose                    Be verbose. (Default value: false)
-h, --help                       Show program usage. (Default value: false)
```

The linux and MacOS X distributions of MEGAN each provide two executables. The executable `MEGAN` launches the program in interactive *GUI mode*. The executable `megan-commandline` runs the program in *command-line mode*.

When running in command-line mode, the program will first executing any command given with the `-x` option and then will read commands from the file specified using the `-c` command. If no such file is given, additional commands are read from standard input.

Please note that windows will still open when in command-line mode, but should not be used interactively. (This is necessary for the program to fully implement all graphical commands.) To prevent windows from opening, or to use the command-line mode on a server, please use the linux *virtual frame buffer* command `xvfb-run`, as shown here:

```
xvfb-run --auto-servernum --server-num=1 MEGAN +g
```

One way to set the maximum number of cores used in any one computation is to enter the following command `setprop MaxNumberCores=`⟨*number*⟩. This value is currently used in when running the LCA algorithm and also when computing a comparison file using sub-sampling. In future releases, it will also apply to other multi-threaded tasks.

Please be aware that the command-line version of the program uses the same *properties file* as the interactive version. So, any *preferences* set using the interactive version of the program will also apply to the command-line version of the program. It this is not desired, then please use the `-p` option to supply a different properties file.

Another important thing to note is that the command-parser operates in a line-by-line fashion. When processing commands in a given line, the parser makes note of required updates to the taxonomy and data-structures. These updates are not executed until all commands in the current input line have been processed. For example, if you want to open and MEGAN file and then to save a picture of the taxonomical analysis in a PDF file, then the two commands should be entered on separate lines because otherwise the taxonomy will be drawn before the data from the MEGAN file has been processed. Here is an example of the correct way to produce a picture of a taxonomic analysis:

```
open file='/Users/huson/data/megan/x.rma'
exportimage file='/Users/huson/data/megan/x.pdf'format=PDF replace=true
quit
```

Alternatively, the `update` command is used to explicitly force MEGAN to update all data-structures, in this case the commands show appear together on one line, e.g.:

```
open file='x.rma';update;exportimage file='x.pdf' format=PDF replace=true;
```

As described below, the `update` command takes a number of different parameters that is used to determine exactly what type of update is required.

Please use the `-x` option only to specify a single command, as updating may otherwise not work correctly.

To save an image of a SEED analysis, after loading a file, we must then open the `SEED` viewer, change the command context to the seed viewer, then configure the size of the window and how much of the tree to uncollapse. Then we save the image file. Here is an example:

```
open file='/Users/huson/data/megan/x/x.rma'
show window=seedviewer
set context=seedviewer
set windowSize=1000 x 1000
select nodes=all
uncollapse subtrees
exportimage file='/Users/huson/data/megan/x/x.pdf'format=PDF replace=true
quit
```

## 34 Command-Line Commands

Command processing has been completely rewritten for MEGAN 4. Each type of window that can be opened by MEGAN has its own command interpreter. Initially, on startup the program will open a `Main` window and all commands piped to the program will be executed using the command interpreter associated with the main window. The main window provides a number of commands for opening other windows. For example, the command `show window=seedviewer;` will open the `SEED` classification viewer. To pipe commands to the SEED viewer, the *command context* has to be set to the SEED viewer, by entering `set context=seedviewer;`. After entering this command, all subsequent commands are handled by the interpreter associated with the `SEED` viewer. To obtain a list of all commands available for the current interpreter, enter `help;`. In obtain help on a particular command, for example on *export*, enter `help export;`. All command description lines that contain the word "export" (case insensitive) will be listed.

In the following we list all commands available in the `Main` viewer. Other viewers support many of these commands, too, but also other, viewer-specific ones. To determine which commands are available for a given window, run MEGAN in GUI mode, open the window of interest and then select the `Window→Command-Line Syntax...` item to obtain a listing of all commands available for the given window. Here are the commands that are available in the `Main` viewer:

```
Available commands (context=MainViewer):

File menu:
new; - Open a new empty document
open file=<filename> [readonly={false|true}]; - Open a MEGAN file (ending on .rma, .meg or .megan)
show window=ImportBlast; - Show the 'Import from Blast' dialog
save file=<filename> [summary={true|false}]; - Save current data set
exportImage file=<filename> [descriptionFile=<filename>] [format={bmp|eps|gif|jpg|pdf|png|svg}] [replace={false|true}]
[visibleOnly]={false|true}] [textAsShapes={false|true}] [title=<string>];
- Export content of window to an image file
exportLegend file=<filename> [format={bmp|eps|gif|jpg|pdf|png|svg}] [replace={false|true}] [textAsShapes={false|true}];
- Export content of legend window
show window=pagesetup; - Setup the page for printing
show window=print; - Print the main panel
extract what=document file=<megan-filename> [sparseFile={false|true}] [data={Taxonomy,SEED,KEGG,COG,REFSEQ|readNames}]
[ids=<numbers...>] [names=<names...>] [namesFile=<file>] [allBelow={false|true}];
- Extract all reads and matches on or below selected node(s) to a new document
show window=ExtractReads; - Extract reads for the selected nodes
show window=properties; - Show document properties
```

```
close; - Close the window

Import sub-menu:
import csv={reads|summary} separator={comma|tab} file=<fileName> [topPercent=<num>] [taxonomy={true|false}] [seed={false|true}]
[cog={false|true}] [kegg={false|true}] [minScore=<num>] [minSupportPercent=<number>] [minSupport=<num>] [multiplier=<number>] [mapper=<mapperConfiguration>];
- Load data in delimiter-separated-values (CSV) format: READ_NAME,CLASS-NAME,SCORE or CLASS,COUNT(,COUNT...)
import format=biome file=<fileName>;
- Import data from a table in BIOME format (see http://biom-format.org/documentation/format_versions/biom-1.0.html)
import metaData=<file> [format={metaDataMapping}];
- Import a metadata mapping file (as defined in http://qiime.org/documentation/file_formats.html)

Export sub-menu:
export what=CSV format={readname_taxonname|readname_taxonid|readname_taxonpath|readname_matches|taxonname_count|taxonpath_count|taxonid_count|
    taxonname_readname|taxonpath_readname|taxonid_readname|taxonname_length|taxonpath_length|taxonid_length|taxonname_normalized_count|
    taxonpath_normalized_count|taxonid_normalized_count|readname_refseqid|reference_readname|readname_seedname|readname_seedpath|seedname_count|
    seedpath_count|seedname_length|seedpath_length|seedname_readname|seedpath_readname|cogname_count|cogpath_count|cogname_length|cogpath_length|
    cogname_readname|cogpath_readname|readname_keggname|readname_keggpath|keggname_count|keggpath_count|keggname_length|keggpath_length|
    keggname_readname|keggpath_readname|ko_taxa}
separator={comma|tab} file=<filename>;
- Export assignments of reads to nodes to a CSV (delimiter-separated values) file
export what=biome data={Taxonomy,SEED,KEGG,COG,REFSEQ file=<filename>;
- Export assignments of reads to nodes in BIOME format
export metaData=<file> [format={metaDataMapping}];
- Export a metadata mapping file (as defined in http://qiime.org/documentation/file_formats.html)
export what=paths file=<filename>; - Export assignments of reads weighted taxonomic paths
export what=tree file=<filename> [simplify={true|false}] [showInternalLabels={false|true}] [showUassigned={false|true}];
- Export induced taxonomic tree (in Newick format)
export what=reads [data={Taxonomy|SEED|KEGG}] file=<filename>;
- Export all reads to a text file (or only those for selected nodes, if any selected)
export what=matches [data={Taxonomy|SEED|KEGG}] file=<filename>;
- Export all matches to a text file (or only those for selected nodes, if any selected)
export what=alignment file=<filename> data={Taxonomy|SEED|KEGG} classId={number[,number...]|selected} [asConsensus={false|true}]
[useEachReadOnlyOnce={true|false}] [useEachReferenceOnlyOnce={true|false}] [includeInsertions={true|false}]
[refSeqOnly={false|true}] [contractGaps={false|true}] [translateCDNA={false|true}] [saveDiversityExtrapolation={false|true}];
- Calculate and export alignments for all selected leaves
export what=cg-vs-coverage file=<filename> data={Taxonomy|SEED|KEGG} classId={number[,number...]|selected};
- Export coverage vs GC content for reference sequences of all selected nodes
upgrade file=<filename>; - Upgrade a MEGAN4 file to MEGAN5. Use this if your RMA file was generated by MEGAN4 and you want to make it fully compatible with MEGAN5.

Edit menu:
copyLegend; - Copy legend image to clipboard
show window=formatter; - Format nodes and edges
show findToolbar={true|false}; - Open the find toolbar

Preferences sub-menu:
show legend={horizontal|vertical|none}; - Show horizontal or vertical legend, or hide
load treeFile=<filename> [mapFile=<filename>]; - Open alternative taxonomy.tre and taxonomy.map files
set useMagnitude={true|false}; - Use reads magnitude values to weight reads, if present in FastA header lines of reads

Fix Taxon Mapping sub-menu:
changeMapping taxName=<taxon-name> taxId=<taxon-id>;
- Change the taxon name to taxon id mapping for a given taxon
changeMapping list; - List all changes
changeMapping clear; - Clear all changes

Select menu:
select nodes={all|none|leaves|internal|previous|subtree|subleaves|intermediate|invert} - Select nodes
select nodes=previous; - Select from previous window

Level sub-menu:
select rank={SuperKingdom|Kingdom|Phylum|Class|Order|Family|Varietas|Genus|Species_group|Subspecies|Species}
- select rank

Options menu:
recompute [minSupportPercent=<number>] [minSupport=<number>] [minScore=<number>] [maxExpected=<number>] [topPercent=<number>] [minComplexity=<number>]
[useMinimalCoverageHeuristic={false|true}] [pairedReads={false|true}] [useIdentityFilter={false|true}]
[useSeed={false|true}] [useCog={false|true}] [useKegg={false|true}];
- Rerun the LCA analysis with different parameters
set totalReads=<num>;
- Set the total number of reads in the analysis (will initiate recalculation of all classifications)
list summary={all|selected}; - List summary of hits for selected nodes of tree
list path=selected; - List path from root to selected node(s)
compute index={Shannon|SimpsonReciprocal} [data={Taxonomy|SEED|KEGG}];
- Compute the Shannon-Weaver diversity index
compare mode={absolute|relative|sqrt|subsample} [ignoreUnassigned={false|true}] [pid=<number> ...] [meganFile=<filename> ...];
- Open compare dialog to produce a comparison of multiple samples
compute projection; - Assign all reads to the current leaves of the taxonomic classification
show webpage taxon=<name|id>; - Open NCBI Taxonomy web site in browser
inspector taxa=selected; - Inspect the read-to-taxon assignments

Taxon Disabling sub-menu:
enable taxa=all; - Enable all taxa
disable taxa={selected|<name,...>}; - Disable all selected taxa or all named ones
enable taxa={selected|<name,...>}; - Enable all selected taxa or all named ones
list taxa=disabled; - List all disabled taxa

Layout menu:
show legend={horizontal|vertical|none}; - Show horizontal or vertical legend, or hide
```

```
set fontSize={<number>|increase|decrease}; - Set the font size
set autoLayoutLabels={true|false}; - Layout labels
set scaleBy={Summarized|Assigned|None}; - Scale nodes by number of reads assigned
set maxNodeHeight=<number>; - Set the maximum node height in pixels
zoom selected; - Zoom to the selection
zoom fit; - Contract tree vertically
zoom full; - Expand tree vertically
set nodeDrawer={Summarized|Assigned|None}; - Draw data as pie charts
set scale={linear|percent|log}; - Show values on a linear scale
set drawer={RectangularCladogram,RectangularPhylogram};
- Draw tree as cladogram with all leaves aligned right
set magnifier={true|false}; - Turn the magnifier on or off
set drawLeavesOnly={true|false}; - Only draw leaves

Expand/Contract sub-menu:
expand direction=horizontal; - Expand view horizontally
contract direction=horizontal; - Contract view horizontally
expand direction=vertical; - Expand view vertically
contract direction=vertical; - Contract view vertically

Highlight Differences sub-menu:
set highlightDifferences={true|false} [correction={none|bonferroni|holm_bonferroni}];
- In a comparison of exactly two samples, highlight statistically significant differences, using no correction
set comparisonHighlightColor=<number>; - Set the pairwise comparison highlight color

Tree menu:
collapse nodes={SELECTED|name [name name ...]}; - Collapse nodes
collapse level=<num>; - Collapse all nodes at given depth in tree
collapse except={name}; - Collapse all parts of tree that are not above or below the selected nodes
uncollapse nodes={all|selected|<name ...>} [subtree={false|true}]; - Uncollapse selected nodes
nodeLabels names={true|false}; - Display the full names of taxa
nodeLabels ids={true|false}; - Display the NCBI ids of taxa
nodeLabels assigned={true|false}; - Display the number of reads assigned to a taxon
nodeLabels summarized={true|false}; - Display the total number of hits to a taxon and its descendants
show labels=selected; - Show labels for selected nodes
hide labels=selected; - Hide labels for selected nodes
show intermediate=<bool>; - Show intermediate labels at nodes of degree 2

Collapse At Taxonomic Rank sub-menu:
collapse rank={SuperKingdom|Kingdom|Phylum|Class|Order|Family|Varietas|Genus|Species_group|Subspecies|Species}

Window menu:
show window=howToCite; - Show how to cite the program
show window=website; - Go to the program website
show window=license; - Show license window
show window=changeLicense; - Load a new license
show window=message; - Open the message window
set windowSize=<width> x <height>; - Set the window size
show window=inspector; - Open inspector window
show window=aligner; - Show alignment of reads to a specified reference sequence
show window=mainViewer; - Brings the main viewer to the front
show window=seedViewer; - Opens the SEED Analyzer
show window=cogViewer; - Opens the COG Analyzer
show window=keggViewer; - Opens the KEGG Analyzer
show window=sampleViewer; - Opens the Sample Viewer
show chart data={taxonomy|seed|kegg|cog|attributes|taxaVsSeed}; - Show chart of assigned reads
show radialChart data={taxonomy|seed|kegg|cog}; - Show radial chart
show wordCloud data={taxonomy|seed|kegg|cog}; - Show word cloud
show comparisonPlot data={Taxonomy|SEED|COG|KEGG};
- Plot pairwise comparison of assignments to classes
show window=clusterViewer; - Open a cluster analysis window
show rarefaction data={Taxonomy|SEED|COG|KEGG};
- Compute and chart a rarefaction curve based on the leaves of the tree shown in the viewer
help [keyword(s)]; - Shows syntax help for commands

Additional commands:
addSample [sample=<name>] source=<filename|pid> ... [overwrite={false|true}]; - Add samples from other documents
extract what=reads outDir=<directory> outFile=<filename-template> [data={Taxonomy|SEED|KEGG|COG|REFSEQ}][ids=<SELECTED|numbers...>]
[names=<names...>] [allBelow={false|true}]; - Extract reads for the selected nodes
fixLinks [old=<filename> new=<filename>] [...]; - Fix links to source files
import blastFile=<name> [,<name>...] [fastaFile=<name> [,<name>...]] meganFile=<name> [maxMatches=<num>] [minScore=<num>] [maxExpected=<num>]
 [topPercent=<num>] [minSupportPercent=<number>] [minSupport=<num>] [minComplexity=<num>] [useMinimalCoverageHeuristic={false|true}]
[useSeed={true|false}] [useCOG={true|false}] [useKegg={true|false}] [paired={false|true} [suffix1=<string> suffix2=<string>]]
[useIdentityFilter={false|true}] [textStoragePolicy={inRMA|inOriginal|inRMAZ}]
[blastFormat={GUESS|BlastX|BlastN|BlastP|BlastXML|BlastTAB|RAPSearch2|RDPAssignmentDetails|RDPStandalone|Mothur|SilvaLogFile|SAM|NR_as_FastA}]
[mapping=<mapperConfiguration>];
- Import BLAST (or RDP or Silva or SAM) and reads files to create a new MEGAN file
list assignments; - List the number of reads assigned to each level of the taxonomy
load cogGIFile=<filename>; - Load a file mapping GI accession numbers to COG-ids
load cogRefSeqFile=<filename>; - Load a file mapping RefSeq-ids to COG-ids
load colorFile=<filename>; - Load a palette of colors from a file (one RGB color per line)
load keggGIFile=<filename>; - Load a file mapping GI accession numbers to KEGG-ids
load keggRefSeqFile=<filename>; - Load a file mapping RefSeq-ids to KEGG-ids
load seedGIFile=<filename>; - Load a file mapping GI accession numbers to SEED-ids
load seedRefSeqFile=<filename>; - Load a file mapping RefSeq-ids to SEED-ids
load synonymsFileCog=<filename>; - Load a file mapping synonyms to COG-ids
load synonymsFileKegg=<filename>; - Load a file mapping synonyms to KEGG-ids
load synonymsFileSeed=<filename>; - Load a file mapping synonyms to SEED-ids
load synonymsFileTaxonomy=<filename>; - Load a file mapping synonyms to taxon-ids
load taxGIFile=<filename>; - Load a file mapping GI accession numbers to taxon-ids
load taxRefSeqFile=<filename>; - Load a file mapping RefSeq-ids to taxon-ids
load taxonomyFile=<filename>; - Load the ncbi.tre and ncbi.map files
mpAnalyzer what={lca-ranks|compare} infile=<filename> outfile=<filename>;
- Compute the rank at which the LCA is found for each mate-pair, or preprocess comparison
quit; - Quit the program
scrollTo node=<name>; - Scroll to a specific node
select id=<id> ...; - Select node(s) for the given id(s)
select name=<name> ... [state={true|false}]; - Select node(s) for the given name(s)
select nodes=all; - Select all nodes
select nodes=invert; - Invert selection
select nodes=leaves; - Select all leaves (except Not Assigned, No Hits and Low Complexity)
select nodes=none; - Deselect all nodes
select nodes=previous; - Select from previous window
select nodes=subTree; - Select subtree below currently selected nodes
select nodes=leavesBelow; - Select all nodes below currently selected nodes
select nodes=above; - Select all nodes above currently selected nodes
set color={<color>|null}; - Set the color of selected nodes and edges
set context={<window-type>|?};
- Choose command context, i.e. the window that will parse subsequent commands. Use ? to list current context and all available contexts.
set dir=<directory> - Set the current directory
set edgeShape={angular|straight|curved}; - Set the shape of selected edges
set edgeWidth=<integer>; - Set the width of selected edges
set fillColor={<color>|null} - Set the fill color of selected nodes
set font=<name-style-size>; - Set font nodes or edges, e.g. arial-italic-12
set highlightContrasts={true|false} [alpha={number}] [bonferroni={false|true}] [ignoreUnassigned={true|false}];
- In a comparison of exactly two samples, highlights nodes that show a significant difference
set labelColor={<color>|null}; - Set the label color of selected nodes and edges
set labelFillColor={<color>|null}; - Set the label color of selected nodes and edges
set loadAllReadsIntoMemory={false|true};
- Determine whether to load all reads into memory when importing BLAST. Should only be used when order of reads differs between BLAST files and FastA files.
set magnifyAllMode={true|false}; - Magnify the whole tree
set margin [left=<number>] [right=<number>] [bottom=<number>] [top=<number>];
- Set margins used in tree visualization
set nodeShape={none|circle|square|triangle|diamond}; - Node shape
set nodeShape={oval|rectangle|none}; - Set the shape of selected nodes
set nodeSize=<integer>; - Set the size of selected nodes
set order=<number> <number>...; - Change the order of samples in a comparison view
set useKegg={true|false}; - Turn KEGG analysis on or off
set usePercentIdentity={false|true};
- Adjust assignment based on best percent identity of matches, using the following minimum requirements:
Species 99%, Genus 97%, Family 95%, Order 90%, Class 85%, Phylum 80%
set useSeed={true|false}; - Turn SEED analysis on or off
setProp <name>=<value>; - Set a property
show histogram taxonId=<num>; - Shows the distribution of matches for a given taxon
show webpage classification=<name> id=<id>; - Search for this in web browser
show window=about; - About MEGAN and the authors
show window=attributes; - Open Microbial Attributes window
show window=checkForUpdate; - Check for an update of the program
show window=colorPalette; - Edit the color palette used in comparison views
show window=comparisonStats; - Open dialog to produce a statistical comparison of two samples
toFront; - Bring window to front
update [reprocess={false|true}] [reset={false|true}] [reinduce={false|true}]; - Update data
version; - Show version info
```

## 34.1 Writing scripts

The best way to run scripts with MEGAN is to prepare a file of commands and then pipe these to MEGAN in command-line mode. Use of the -x option to supply commands is not encouraged because of update issues. MEGAN uses updates all windows etc after a line of commands has been entered and all commands provided using the -x option are considered to be contained in a single line.

Here is an example of how one would use MEGAN in command-line mode on a Mac to save some information on KEGG assignments:

```
 /Applications/Megan/MEGAN.app/Contents/MacOS/JavaApplicationStub -g -E < commands.txt
```

where the file `commands.txt` contains the following lines:

```
open file='/Users/huson/data/megan/microbiome.rma';
```

```
show window=KeggViewer;
set context=KeggViewer;
update;
uncollapse nodes=all;
select nodes=leaves;
export what=CSV format=keggpath_count separator=tab file='/Users/huson/data/megan/kegg.txt';
quit;
```

The first line is used to open a MEGAN file. Please surround the file name with single quotes as shown here.
The second line opens the `KEGG` window (or KeggViewer, as it is referred to here).
The third line sets the command context to the KeggViewer (so that subsequent commands are interpreted by the KeggViewer). The argument of this command is case sensitive. Please use `KeggViewer` and not `keggviewer`.
The fourth line ensures that the `KEGG` window is uptodate.
The fifth line uncollapses the whole Kegg tree.
The sixth line selects all leaves of the Kegg tree.
The seventh line exports all Kegg paths and read counts to a file in "Delimiter separated format".
The eight line quits the program.

## 35 Examples

Example files can be downloaded from the MEGAN website.

## 36 Using More Memory

The MEGAN installer allows you to specify the amount of MEGAN that the program can use. We recommend at least 2 GB on a 64-bit machine and recommend 8 GB on a desktop.

MEGAN is a memory-hungry application. When importing BLAST files, we recommend that you use a machine that allows you to run MEGAN with at least 4 GB of main memory. Using less memory will work, but Java will be forced to perform frequent garbage collection, which will slow the program down. Also, because the program is i/o intensive, it is best to have all files on local disks, as this will increase the speed of the program.

To run MEGAN with more than 2GB under MacOS X on an intel Mac, edit the file
`/Applications/MEGAN/MEGAN.app/Contents/Info.plist` as follows: Find the lines

```
<key>VMOptions</key>
<string>-server -Xms2000M -Xmx2000M </string><!-- I4J_INSERT_VMOPTIONS -->
```

and replace them by:

```
<key>VMOptions</key>
<string>-server -Xms2000M -Xmx8000M </string><!-- I4J_INSERT_VMOPTIONS -->
```

to run using 8 gigabytes, for example.

To run MEGAN with more than 2GB on a 64-bit unix/linux system, open the file ⟨*installation-dir*⟩`/MEGAN.vmoptions` in a text editor. Find the current memory specification (e.g. `-Xmx1600M`) and replace it by `-Xmx8G` to run with 8 gigabytes of memory, say.

## 37 MEGAN Tools

The Linux and MacOS distributions of MEGAN have an additional *tools directory* that contains a number of standalone utility programs.

The program *sam2rma* can be used on a server to generate a RMA file on a server. The input must be one or more SAM files generated by either DIAMOND or MALT.

The program *blast2lca* provides a standalone implementation of the LCA algorithm and can also perform KEGG classification. Input is a BLAST file or similar.

More tools will follow.

## 38 Acknowledgments

This program uses a number of external Java libaries. The JARs and their licenses are contained in the jars directory. MEGAN is obfuscated using yguard.

## References

[1] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25:3389–3402, 1997.

[2] Manimozhiyan Arumugam, Jeroen Raes, Eric Pelletier, Denis Le Paslier, Takuji Yamada, D R. Mende, Gabriel R. Fernandes, Julien Tap, Thomas Bruls, Jean-Michel Batto, Marcelo Bertalan, Natalia Borruel, Francesc Casellas, Leyden Fernandez, Laurent Gautier, Torben Hansen, Masahira Hattori, Tetsuya Hayashi, Michiel Kleerebezem, Ken Kurokawa, Marion Leclerc, Florence Levenez, Chaysavanh Manichanh, H. Bjorn Nielsen, Trine Nielsen, Nicolas Pons, Julie Poulain, Junjie Qin, Thomas Sicheritz-Ponten, Sebastian Tims, David Torrents, Edgardo Ugarte, Erwin G. Zoetendal, Jun Wang, Francisco Guarner, Oluf Pedersen, Willem M. de Vos, Soren Brunak, Joel Dore, Jean Weissenbach, S. Dusko Ehrlich, and Peer Bork. Enterotypes of the human gut microbiome. *Nature*, 473(7346):174–180, 05 2011.

[3] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, and D.L. Wheeler. Genbank. *Nucleic Acids Res*, 1(33):D34–38, 2005.

[4] B. Buchfink, C. Xie, and D.H. Huson. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12:59–60, 2015.

[5] J. R. Cole, Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje. The ribosomal database project: improved alignments and new tools for rrna analysis. *Nucleic Acids Research*, 37(suppl 1):D141–D145, January 2009.

[6] Elizabeth M. Glass, Jared Wilkening, Andreas Wilke, Dionysios Antonopoulos, and Folker Meyer. Using the metagenomics rast server (mg-rast) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc*, 2010(1):pdb.prot5368+, January 2010.

[7] D. H. Huson, A. F. Auch, J. Qi, and S. C. Schuster. MEGAN analysis of metagenomic data. *Genome Res*, 17(3):377–386, March 2007.

[8] D. H. Huson, S. Mitra, N. Weber, H.-J. Ruscheweyh, and S. C. Schuster. Integrative analysis of environmental sequences using MEGAN 4. *Genome Research*, 21:1552–1560, 2011.

[9] M. Kanehisa and S. Goto. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1):27–30, Jan 2000.

[10] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The sequence alignment/map (SAM) format and SAMtool. *Bioinformatics*, 25:2078–9, 2009.

[11] Rachel Mackelprang, Mark Waldrop, Kristen DeAngelis, Maude David, Krystle Chavarria, Steven Blazewicz, Edward Rubin, and Janet Jansson. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature*, 480(7377):368–371, 2011.

[12] S. Mitra, J.A. Gilbert, D. Field, and D.H. Huson. Comparison of multiple metagenomes using phylogenetic networks based on ecological indices. *ISME J*, 2010. doi:10.1038/ismej.2010.51.

[13] Suparna Mitra, Max Schubach, and D H Huson. Short clones or long clones? A simulation study on the use of paired reads in metagenomics. *BMC Bioinformatics*, 11(Suppl 1):S12+, 2010.

[14] David W. Mount. Using the basic local alignment search tool (BLAST). *Cold Spring Harbor Protocols*, 2007(7):pdb.top17, 2007.

[15] Ross Overbeek, Tadhg Begley, Ralph M Butler, Jomuna V Choudhuri, Han-Yu Chuang, Matthew Cohoon, Valérie de Crécy-Lagard, Naryttza Diaz, Terry Disz, Robert Edwards, Michael Fonstein, Ed D Frank, Svetlana Gerdes, Elizabeth M Glass, Alexander Goesmann, Andrew Hanson, Dirk Iwata-Reuyl, Roy Jensen, Neema Jamshidi, Lutz Krause, Michael Kubal, Niels Larsen, Burkhard Linke, Alice C McHardy, Folker Meyer, Heiko Neuweger, Gary Olsen, Robert Olson, Andrei Osterman, Vasiliy Portnoy, Gordon D Pusch, Dmitry A Rodionov, Christian Rückert, Jason Steiner, Rick Stevens, Ines Thiele, Olga Vassieva, Yuzhen Ye, Olga Zagnitko, and Veronika Vonstein. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res*, 33(17):5691–5702, 2005.

[16] Hendrik N Poinar, Carsten Schwarz, Ji Qi, Beth Shapiro, Ross D E Macphee, Bernard Buigues, Alexei Tikhonov, D H Huson, Lynn P Tomsho, Alexander Auch, Markus Rampp, Webb Miller, and Stephan C Schuster. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science*, 311(5759):392–394, Jan 2006.

[17] Sean Powell, Damian Szklarczyk, Kalliopi Trachana, Alexander Roth, Michael Kuhn, Jean Muller, Roland Arnold, Thomas Rattei, Ivica Letunic, Tobias Doerks, Lars Juhl Jensen, Christian von Mering, and Peer Bork. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Research*, 40(Database-Issue):284–289, 2012.

[18] E. Pruesse, C. Quast, K. Knittel, B. Fuchs, W. Ludwig, J. Peplies, and F.O. Glöckner. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nuc. Acids Res.*, 35(21):7188–7196, 2007.

[19] R. L. Tatusov, E. V. Koonin, and D. J. Lipman. A genomic perspective on protein families. *Science*, 278(5338):631–637, Oct 1997.

[20] Wikipedia. Michaelis-Menten kinetics. `http://en.wikipedia.org/wiki/Michaelis--Menten_kinetics`, 2012.

[21] John Wootton and Scott Federhen. Statistics of local complexity in amino acid sequences and sequence databases. *Computers & Chemistry*, 17(2):149–163, 1993.

[22] Yongan Zhao, Haixu Tang, and Yuzhen Ye. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics*, 28(1):125–126, 2012.

# Index

-m 0, 40
-m 7, 40
-m 8, 40
.bin, 29
.biom, 9
.gz, 29
.map, 29
.meg, 39
.megan, 39
.rma, 39
.txt, 29

.blastn.gz, 7
.fna.zip, 7
.gz, 7
.zip, 7

About, 38
About..., 21, 24, 38
Academic license, 6
Add A Change..., 16
Add From File..., 26
Add..., 26
Advanced Tab, 30
Algorithms, 26
Algorithms→Compute Core Biome..., 26
Algorithms→Compute Rare Biome..., 27
Algorithms→Compute Shared Biome..., 27
Algorithms→Compute Total Biome..., 26
Algorithms→Extract Samples..., 26
Algorithms→Resample..., 27
Alignment Viewer, 33
Alignments..., 14
All Intermediate Nodes, 16
All Internal Nodes, 16
All Leaves, 16
All Nodes, 16
Apply Reordering To Viewers, 26
arb file, 9
assemble, 33
Attributes, 25
Attributes→Delete..., 25
Attributes→Duplicate..., 25
Attributes→Import From File..., 25
Attributes→New..., 25
Attributes→Rename..., 25
Attributes→Select All Same, 25
Attributes→Set Color..., 25
Attributes→Set Value..., 25

bar chart, 32
BIOM1 Format..., 9, 14
BIOME, 9
BiPlot Size..., 28
BLAST file, 40
blast2lca, 52
BLASTN file, 40
BLASTP file, 40
BLASTX file, 40
BMP, 43
Bonferroni Corrected, 20

Case sensitive, 35
Change LCA Parameters..., 17, 36
Change License..., 21
Chart Diversity, 34
Chart Microbial Attributes..., 22, 23, 32
Chart..., 21, 23, 32
Cladogram, 19, 22
Class, 17, 21
class/resources/kegg, 12
Classifier :: Assignment Detail, 8
Classifier :: Hierarchy View, 8
Clear All Changes..., 16
Close, 14, 35
Cluster, 27
Cluster Analysis..., 22, 23
Cluster Viewer Edit, 27
Cluster Viewer Layout, 27
Cluster Viewer Options, 28
Cluster Viewer View, 29
clusters of orthologous groups, 12
co-occurrence network, 32
COG, 12, 24
COG Analyzer..., 21, 23
COG Chart Window, 33
COG Tab, 30
COGs, 12
Collapse, 20, 22, 32
Collapse All Others, 20

Collapse at Level..., 20, 23
Collapse At Taxonomic Rank, 20
Collapse At Taxonomic Rank→Class, 21
Collapse At Taxonomic Rank→Family, 21
Collapse At Taxonomic Rank→Genus, 21
Collapse At Taxonomic Rank→Kingdom, 20
Collapse At Taxonomic Rank→Order, 21
Collapse At Taxonomic Rank→Phylum, 21
Collapse At Taxonomic Rank→Species, 21
Collapse At Taxonomic Rank→Super Kingdom, 20
Color By Attribute, 26
Color Matches, 34
Color Mismatches, 34
color, change, 36
command context, 47
Command Input..., 7, 15, 22, 33, 35, 37
command-line, 45
command-line installation, 6
command-line mode, 46
Command-Line Syntax..., 22, 47
commands.txt, 50
Compare, 37, 38
Compare..., 18, 37
Comparison Plot..., 22, 23
compressed, 31
Compute Core Biome..., 26
Compute Rare Biome..., 27
Compute Shared Biome..., 27
Compute Taxonomic Profile..., 18
Compute Total Biome..., 26
Contract Gaps, 34
Contract Horizontal, 19, 22
Contract Vertical, 19, 22
Convert text to graphics, 38
Copy, 15
Copy Alignment, 34
Copy Consensus, 34
Copy Image, 15
Copy Legend, 15
Copy Reference, 34
Coverage vs CG Content..., 15
CSV file, 44
CSV Format..., 14, 44
Cut, 15
Decrease Font Size, 18, 28
Delete..., 25, 26
Description, 30
Description..., 15
Disable..., 18
Disclaimer, 3
Don't save, 31
Draw Bars, 19, 23
Draw Circles, 19, 23
Draw Coxcombs, 19, 23
Draw Heatmaps, 19, 23
Draw Leaves Only, 19
Draw Pies, 19, 23
Duplicate..., 25, 26

E-value, 36
Edit, 15
Edit Edge Label, 15
Edit Node Label, 15
Edit→Copy, 15
Edit→Copy Alignment, 34
Edit→Copy Consensus, 34
Edit→Copy Image, 15
Edit→Copy Legend, 15
Edit→Copy Reference, 34
Edit→Cut, 15
Edit→Description..., 15
Edit→Edit Edge Label, 15
Edit→Edit Node Label, 15
Edit→Find Again, 15
Edit→Find..., 15, 35
Edit→Format..., 15, 36
Edit→Group Nodes, 27
Edit→Node Shape, 27
Edit→Paste, 15
Edit→Preferences, 15
Edit→Show Groups, 27
Edit→Show Taxon..., 31
Edit→Translate..., 34
Edit→Ungroup All, 27
Enable All, 18
Enable..., 18
Enterprise license, 6
enzymes, 24
EPS, 43
Evaluation license, 6

examples, 5
Expand, 32
Expand Horizontal, 19, 22
Expand Vertical, 19, 22
Expand/Contract, 18, 19, 28
Expand/Contract→Contract Horizontal, 19
Expand/Contract→Contract Vertical, 19
Expand/Contract→Expand Horizontal, 19
Expand/Contract→Expand Vertical, 19
Export, 13, 14
Export Alignment..., 33
Export Consensus..., 33
Export Contigs..., 33
Export Image, 37
Export Image..., 13, 22, 37
Export Legend..., 13
Export Reference..., 33
Export→Alignments..., 14
Export→BIOM1 Format..., 9, 14
Export→Coverage vs CG Content..., 15
Export→CSV Format..., 14
Export→Matches..., 14
Export→MEGAN Summary File..., 15
Export→MEGAN5 File..., 15
Export→Metadata..., 14
Export→Reads..., 14, 37
Export→Taxonomic Paths..., 14
Export→Tree..., 14
Extract Reads..., 13, 23, 37
Extract Samples..., 26
Extract To New Document..., 13
Extractor, 37

Family, 17, 21
fasta format, 43
fastq format, 43
File, 13
File→Close, 14
File→Export, 13
File→Export Alignment..., 33
File→Export Consensus..., 33
File→Export Contigs..., 33
File→Export Image..., 13, 37
File→Export Legend..., 13
File→Export Reference..., 33
File→Extract Reads..., 13, 37
File→Extract To New Document..., 13
File→Import, 13
File→Import From BLAST..., 5, 13, 39
File→New..., 13
File→Open From Server..., 13, 38
File→Open Recent, 7, 13
File→Open..., 7, 13
File→Page Setup..., 13
File→Print..., 13
File→Properties..., 13
File→Quit, 14, 24
File→Save As..., 13
Files, 29
Find, 35
Find Again, 15
Find All, 35
Find First, 35
Find Next, 35
Find..., 15, 22, 35
Fix Taxon Mapping, 16
Fix Taxon Mapping→Add A Change..., 16
Fix Taxon Mapping→Clear All Changes..., 16
Fix Taxon Mapping→List All Changes..., 16
Flip Horizontally, 28
Flip Vertically, 28
font, change, 36
Format, 36
Format..., 15, 36
From File, 35
From File..., 35
From Previous Window, 16
Fully Contract, 18, 22
Fully Expand, 18, 22
functional roles, 11

genome, 3
Genus, 17, 21
GI accession, 43
GIF, 43
graph, 32
Group Nodes, 26, 27
Group Viewer, 35
guest, 38
GUI mode, 46
gzip format, 7
gzipped BLAST files, 40

heat map, 32
Hide Selected, 26
Hide Unselected, 26
Highlight Differences, 19
Highlight Differences→Bonferroni Corrected, 20
Highlight Differences→Holm-Bonferroni Corrected, 19
Highlight Differences→Set Highlight Color..., 20
Highlight Differences→Uncorrected, 19
Holm-Bonferroni Corrected, 19
horizontal zoom, 24
How to cite, 3
How to Cite..., 21
http://megan-db.org/Permafrost, 38

Ignore all unassigned reads, 37
Import, 13, 14
Import Blast, 29
Import From BLAST..., 5, 13, 39
Import From File..., 25
Import→BIOM1 Format..., 9, 14
Import→CSV Format..., 14, 44
Import→Metadata..., 14
Increase Font Size, 18, 27
Inspect, 5, 31
Inspect..., 18, 23
Inspector, 31
Inspector Options, 32
Inspector Window..., 21, 31
Invert, 17

JPEG, 43

KEGG, 12, 24
KEGG Analyzer..., 21, 23
KEGG Chart Window, 33
KEGG Tab, 30
Kegg tree, 12
Kingdom, 17, 20
KO, 12

Layout, 18
Layout Labels, 18
Layout→BiPlot Size..., 28
Layout→Cladogram, 19
Layout→Contract Gaps, 34
Layout→Decrease Font Size, 18, 28
Layout→Draw Bars, 19
Layout→Draw Circles, 19
Layout→Draw Coxcombs, 19
Layout→Draw Heatmaps, 19
Layout→Draw Leaves Only, 19
Layout→Draw Pies, 19
Layout→Expand/Contract, 18, 28
Layout→Flip Horizontally, 28
Layout→Flip Vertically, 28
Layout→Fully Contract, 18
Layout→Fully Expand, 18
Layout→Highlight Differences, 19
Layout→Increase Font Size, 18, 27
Layout→Layout Labels, 18
Layout→Linear Scale, 19
Layout→Log Scale, 19
Layout→PC1 PC2 PC3, 28
Layout→PC1 vs PC2, 28
Layout→PC1 vs PC3, 28
Layout→PC2 vs PC3, 28
Layout→PCi PCj PCk, 28
Layout→PCi vs PCj..., 28
Layout→Phylogram, 19
Layout→Scale Nodes By Assigned, 18
Layout→Scale Nodes By Summarized, 18
Layout→Set Max Node Height..., 18
Layout→Set Node Radius..., 28
Layout→Show Amino Acids, 35
Layout→Show BiPlot, 28
Layout→Show Consensus, 35
Layout→Show Insertions, 34
Layout→Show Labels, 28
Layout→Show Legend, 18, 27
Layout→Show Nucleotides, 35
Layout→Show Reference, 35
Layout→Show Unaligned, 35
Layout→Sqrt Scale, 19
Layout→Use Colors, 28
Layout→Use Magnifier, 19
Layout→Zoom to Fit, 28
Layout→Zoom To Selection, 18
LCA, 8
LCA of a fixed percent, 11
LCA Parameters Tab, 30
LCA Percent, 36

LCA-assignment algorithm, 11
Leaves Below, 17
Level, 17
Level→Class, 17
Level→Family, 17
Level→Genus, 17
Level→Kingdom, 17
Level→Order, 17
Level→Phylum, 17
Level→Species, 17
Level→Super Kingdom, 17
License..., 21
line width, change, 36
Linear Scale, 19, 23
Linux, 6, 7
List All Changes..., 16
List Disabled..., 18
List Microbial Attributes..., 17
List Path..., 17
List Summary..., 17
Load All Reads Into Memory, 31
Load GI-Lookup File, 40
log file, 9
Log Scale, 19, 23
Low Complexity, 36

MacOS, 7
MacOS X, 6
magnitude, 16
Main, 12
Main Viewer..., 21, 23
map*.conf, 12
map*.png, 12
Matches..., 14
Matrix, 29
Max Expected, 36
Max number of matches per read, 29
max prevalence, 33
MEGAN, 7, 24, 46
MEGAN project, 8
MEGAN Summary File..., 15
MEGAN text file, 39
megan-commandline, 7, 46
MEGAN5 File..., 15
MEGAN_macos_5.11.3.dmg, 6
MEGAN_unix_5.11.3.sh, 6
MEGAN_windows-64x_5.11.3.exe, 6
MEGAN_windows_5.11.3.exe, 6
MeganServer, 38
Message, 36
Message Window..., 21, 36
metabolic pathways, 12
Metadata..., 14
metagenome, 3
metagenomics, 3
MG-RAST, 9
Min Complexity, 36
min prevalence, 32
Min Score, 8, 36
Min Support, 36
Min Support Percent, 36
minimum coverage heuristic, 11
minimum overlap, 33
Move Down, 26, 34
Move Up, 26, 34

NCBI mapping file, 45
NCBI taxonomy, 10
NCBI tree file, 45
NCBI-NR, 10
NCBI-NT, 10
Network, 29
New..., 13, 25
NJ Tree, 29
Node Labels Off, 20
Node Labels On, 20
Node Shape, 26, 27
node size, change, 36
Node→Inspect, 5, 31
NOGs, 12
None, 16
Notes Tab, 30

Open From Server..., 13, 38
Open NCBI Web Page..., 18
Open Recent, 7, 13, 14
Open..., 7, 13, 22
Options, 17
Options→Change LCA Parameters..., 17, 36
Options→Chart Diversity, 34
Options→Collapse, 32
Options→Color Matches, 34

Options→Color Mismatches, 34
Options→Compare..., 18, 37
Options→Compute Taxonomic Profile..., 18
Options→Expand, 32
Options→Inspect..., 18
Options→List Microbial Attributes..., 17
Options→List Path..., 17
Options→List Summary..., 17
Options→Move Down, 34
Options→Move Up, 34
Options→Open NCBI Web Page..., 18
Options→Set Amino Acid Colors..., 34
Options→Set Number Of Reads..., 5, 17
Options→Shannon-Weaver Index..., 17
Options→Show As Alignment, 34
Options→Show As Mapping, 34
Options→Show KEGG Pathway..., 25
Options→Show Taxon, 31
Options→Simpson-Reciprocal Index..., 17
Options→Sort By Contigs, 34
Options→Sort By Names, 34
Options→Sort By Similarity, 34
Options→Sort By Start, 34
Options→Sync, 28
Options→Taxon Disabling, 11, 17
Options→Unsorted, 34
Options→Use All Hits, 32
Options→Use Bray-Curtis, 28
Options→Use Chi-Square, 28
Options→Use Euclidean, 28
Options→Use Goodall, 28
Options→Use Hellinger, 28
Options→Use Hit, 32
Options→Use JSD, 28
Options→Use Kulczynski, 28
Options→Use Normalized Goodall, 28
Options→Use Pearson, 28
Order, 17, 21

Page Setup..., 13
Paired Reads, 36
Paired reads, 29
Parameters, 36
Paste, 15
pathway tab, 25
pathway/map, 12
PC1 PC2 PC3, 28
PC1 vs PC2, 28
PC1 vs PC3, 28
PC2 vs PC3, 28
PCi PCj PCk, 28
PCi vs PCj..., 28
PCoA, 29
PCoA plot, 27
PDF, 43
Phylogram, 19, 22
Phylum, 17, 21
pie chart, 32
PNG, 43
Preferences, 15, 16
preferences, 46
Preferences→Fix Taxon Mapping, 16
Preferences→Show Legend, 16
Preferences→Use Alternative Taxonomy..., 16
Preferences→Use Default NCBI Taxonomy, 16
Preferences→Use Magnitudes, 16
Print..., 13, 22
probability of coocurrence, 33
probablity, 33
properties file, 46
Properties..., 13

QIIME, 9
Quit, 14, 24

Radial Chart..., 21, 23
radial space filling tree, 32
RAPSearch2, 5
Rarefaction, 32
Rarefaction Analysis..., 22, 23
rarefaction plot, 32
RDP, 8
RDP-Assignment-Detail, 8
RDP-standalone, 9
read file, 39
read hit node, 31
read node, 31
read weights, 16
reads file, 43
Reads..., 14, 37
RefSeq, 12
Regular Expression, 35

Rename..., 25, 26
Resample..., 27
RMA, 39
RMA file, 39
RMAZ, 30

SAM, 8
sam2rma, 52
Sample Viewer, 25
Sample Viewer..., 21, 23
Samples, 26
Samples→Add From File..., 26
Samples→Add..., 26
Samples→Apply Reordering To Viewers, 26
Samples→Color By Attribute, 26
Samples→Delete..., 26
Samples→Duplicate..., 26
Samples→Group Nodes, 26
Samples→Hide Selected, 26
Samples→Hide Unselected, 26
Samples→Move Down, 26
Samples→Move Up, 26
Samples→Node Shape, 26
Samples→Rename..., 26
Samples→Set Color..., 26
Samples→Show All, 26
Samples→Show Selected, 26
Samples→Ungroup All, 26
Save As..., 13
Save in main file, 30
Save in separate file, 30
Save visible image, 38
Save whole image, 38
Scale Nodes By Assigned, 18
Scale Nodes By Summarized, 18
SEED, 11, 24
SEED Analyzer..., 21, 23
SEED Chart Window, 33
SEED Tab, 30
Select, 16
Select All Same, 25
Select→All Intermediate Nodes, 16
Select→All Internal Nodes, 16
Select→All Leaves, 16
Select→All Nodes, 16
Select→From Previous Window, 16
Select→Invert, 17
Select→Leaves Below, 17
Select→Level, 17
Select→None, 16
Select→Subtree, 16
Server Browser, 38
Server Browser:Open remote server, 38
Set Amino Acid Colors..., 34
Set Color..., 25, 26
Set Highlight Color..., 20
Set Max Node Height..., 18
Set Node Radius..., 28
Set Number Of Reads..., 5, 17
Set Value..., 25
Set Window Size..., 21
Shannon-Weaver Index..., 17
Show Alignment..., 21, 23, 33
Show All, 26
Show Amino Acids, 35
Show As Alignment, 34
Show As Mapping, 34
Show BiPlot, 28
Show Consensus, 35
Show Groups, 27
Show IDs, 20
Show Insertions, 34
Show Intermediate Labels, 20
Show KEGG Pathway..., 25
Show Labels, 28
Show Legend, 16, 18, 23, 27
Show Names, 20
Show Nucleotides, 35
Show Number of Reads Assigned, 20
Show Number of Reads Summarized, 20
Show Reference, 35
Show Selected, 26
Show Taxon, 31
Show Taxon..., 31
Show Unaligned, 35
Silva, 9
silva2ncbi.map, 9
Simpson-Reciprocal Index..., 17
Single user license, 6
Site license, 6
Sort By Contigs, 34

Sort By Names, 34
Sort By Similarity, 34
Sort By Start, 34
Species, 17, 21
Sqrt Scale, 19, 23
subsystems, 11
Subtree, 16
Super Kingdom, 17, 20
SVG, 43
Sync, 28
sync, 32
synonyms file, 43

taxon assignment to matches, 8
Taxon Chart Window, 32, 33
Taxon Disabling, 11, 17, 18
Taxon Disabling→Disable..., 18
Taxon Disabling→Enable All, 18
Taxon Disabling→Enable..., 18
Taxon Disabling→List Disabled..., 18
taxon node, 31
Taxonomic Paths..., 14
Taxonomy Tab, 29
text storage policies, 30
three dimensional PCoA, 27
threshold, 32
tools directory, 52
Top Percentage, 36
Translate..., 34
Tree, 20
Tree→Collapse, 20
Tree→Collapse All Others, 20
Tree→Collapse at Level..., 20
Tree→Collapse At Taxonomic Rank, 20
Tree→Node Labels Off, 20
Tree→Node Labels On, 20
Tree→Show IDs, 20
Tree→Show Intermediate Labels, 20
Tree→Show Names, 20
Tree→Show Number of Reads Assigned, 20
Tree→Show Number of Reads Summarized, 20
Tree→Uncollapse, 5, 20
Tree→Uncollapse All, 20
Tree→Uncollapse Subtree, 5, 20
Tree..., 14
two dimensional PCoA, 27
Type-setting conventions, 3

unassigned, 36
Uncollapse, 5, 20, 23, 35
Uncollapse All, 20
Uncollapse Subtree, 5, 20, 23
Uncorrected, 19
Ungroup All, 26, 27
Unix, 6, 7
Unsorted, 34
update, 46
UPGMA Tree, 29
Use 16S Percent Identity Filter, 37
Use Absolute Counts, 37
Use All Hits, 32
Use Alternative Taxonomy..., 16
Use Bray-Curtis, 28
Use Chi-Square, 28
Use Colors, 28
Use Default NCBI Taxonomy, 16
Use Euclidean, 28
Use GI Lookup, 43
Use Goodall, 28
Use Hellinger, 28
Use Hit, 32
Use JSD, 28
Use Kulczynski, 28
Use Magnifier, 19
Use Magnitudes, 16
Use Normalized Counts, 37
Use Normalized Goodall, 28
Use Pearson, 28
Use Square Root Normalization, 37
Use Sub-Sampled Counts, 37
Use Synonyms, 43
Use Use Minimum Coverage Heuristic, 37

vertical zoom, 24
View→Matrix, 29
View→Network, 29
View→NJ Tree, 29
View→PCoA, 29
View→UPGMA Tree, 29
virtual frame buffer, 46

Website..., 21

Whole words only, 35
Window, 21
Window→About..., 21, 24, 38
Window→Change License..., 21
Window→Chart Microbial Attributes..., 22
Window→Chart..., 21, 32
Window→Cluster Analysis..., 22
Window→COG Analyzer..., 21
Window→Command Input..., 7, 15, 22, 33, 35, 37
Window→Command-Line Syntax..., 22, 47
Window→Comparison Plot..., 22
Window→How to Cite..., 21
Window→Inspector Window..., 21, 31
Window→KEGG Analyzer..., 21
Window→License..., 21
Window→Main Viewer..., 21
Window→Message Window..., 21, 36
Window→Radial Chart..., 21
Window→Rarefaction Analysis..., 22
Window→Sample Viewer..., 21
Window→SEED Analyzer..., 21
Window→Set Window Size..., 21
Window→Show Alignment..., 21, 33
Window→Website..., 21
Window→Word Cloud..., 22
Windows, 6, 7
Word Cloud..., 22, 23

xvfb-run, 46

zip format, 7
zipped BLAST files, 40
Zoom to Fit, 28
Zoom To Selection, 18