



# Marine Microbial Biodiversity, Bioinformatics & Biotechnology



Grant agreement n°287589

Acronym : Micro B3

Start date of project: 01/01/2012, funded for 48 month

## Deliverable 5.9

# Ecological Analysis Tools for Microbial Ecology (EATME)

Version: 1.0

Circulated to: WP5 leader (10.02.2014)

Approved by: Frank Oliver Glöckner

Expected Submission Date: 31.12.2013

Actual submission Date: 28.02.2014

Lead Party for Deliverable: ALFRED-WEGENER-INSTITUT FUER POLAR- UND MEERESFORSCHUNG (PIC: 999497507)

Mail: [pbuttigi@mpi-bremen.de](mailto:pbuttigi@mpi-bremen.de), [aramette@mpi-bremen.de](mailto:aramette@mpi-bremen.de) Tel.: +49 421 2028 984

### Dissemination level:

Public (PU)	✓
Restricted to other programme participants (including the Commission Services) (PP)	
Restricted to a group specified by the consortium (including the Commission Services) (RE)	
Confidential, only for members of the consortium (including the Commission Services) (CO)	



The Micro B3 project is funded from the European Union's Seventh Framework Programme ( Joint Call OCEAN.2011-2: Marine microbial diversity – new insights into marine ecosystems functioning and its biotechnological potential) under the grant agreement no 287589. The Micro B3 project is solely responsible for this publication. It does not represent the opinion of the EU. The EU is not responsible for any use that might be made of data appearing herein.





## Summary

This deliverable comprises the Ecological Analysis Tools for Microbial Ecology (EATME) software package. A range of analysis functions have been consolidated and implemented as easy-to-use web applications integrated into the Micro B3-IS. These functions have been selected due to their applicability to the high dimensional and complex data sets generated by modern microbial ecology research. Relatively new techniques as well as canonical methods have been included in the initial release and new functions may be readily added in a modular fashion. Each function can be used through an interactive web-page with simple, point-and-click usability. Ancillary functions allowing, for example, data transformation and normalisation augment EATME's core functions and are seamlessly integrated into the relevant web-pages. Together with the Interactive Ecological Analysis Guide (IEAG; see Deliverable 5.4), EATME provides an approachable solution to apply multivariate statistical techniques to microbial ecology data sets.

## Access

- EATME is available online at: <http://mb3is.megx.net/eatme/>
- The source code is available in a subversion repository: <https://colab.mpi-bremen.de/micro-b3/svn/eatme/trunk>
- An issue tracker is viewable at: [https://colab.mpi-bremen.de/its/browse/MB3\\_IS/component/10112](https://colab.mpi-bremen.de/its/browse/MB3_IS/component/10112)
- To post issues, register via the invitation at: <https://colab.mpi-bremen.de/wiki/signup.action?token=19ca1e66caf77208>

## Description of EATME

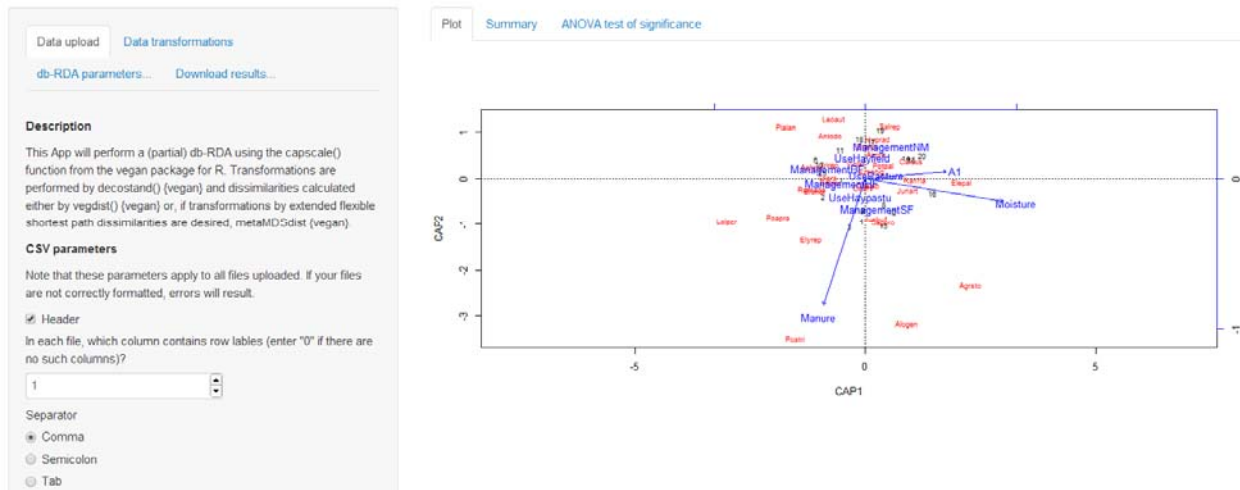
### Development and implementation

In developing EATME, as outlined in Deliverable 5.1, functions from selected R (R Core Team 2013) packages such as the numerical ecology package `vegan` (Oksanen et al. 2013), `Hmisc` (Harrell & with contributions from many other users 2009), and `mvoutlier` (Filzmoser & Gschwandtner 2013; Filzmoser et al. 2008) were rendered as interactive web-based applications using the `shiny` R package (RStudio Inc. 2013). Each application comprises a server script and user interface (UI) script specifying its execution and interactivity with the user, respectively. Rather than a command line interface (CLI), users interact with familiar web-based input elements such as drop-down menus, radio buttons, and sliders (**Figure 1**, **Figure 2**). These elements transmit user input to the server script where conditional flow control ensures that appropriate functions and parameters are defined given the input harvested through the UI. Results, in the form of graphical and textual output, are immediately returned to the user on the same web-page. An extended

range of results, which are not comfortably viewable on a web-page, are made available as downloads.

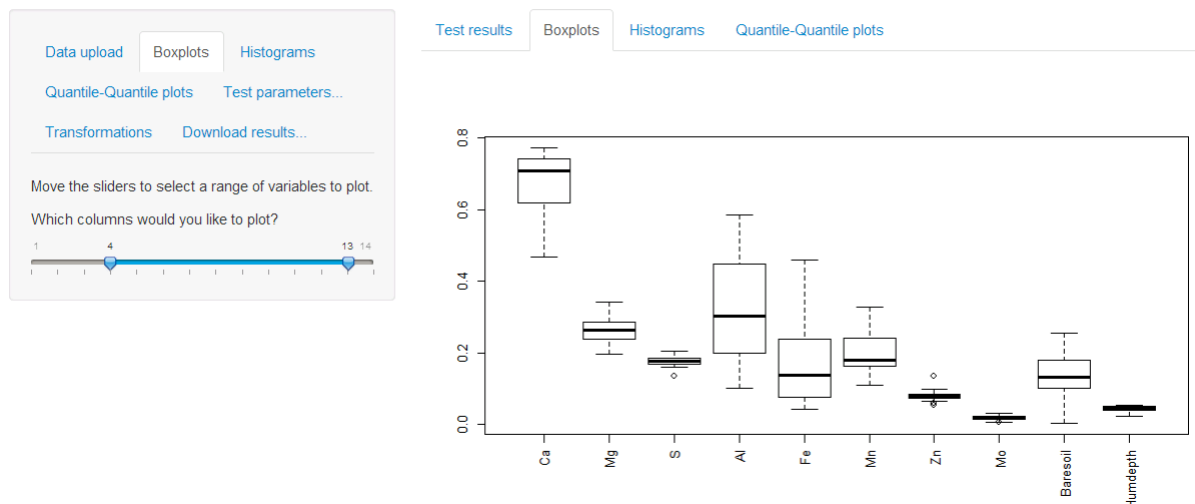
Errors generated due to, for example, incorrectly formatted data or unsuitable data types are forwarded from the R session running on the server to the user via the UI. Several custom error messages, which are more easily understandable to users unfamiliar with R, have been implemented to ease usage and direct corrective action. These can readily be extended as required.

## Perform a (partial) distance-based redundancy analysis



**Figure 1:** An example of an R function implemented as an interactive web-page in EATME. A solution is computed immediately upon the successful upload of a data set to the server. Results, such as plots and textual/numeric output, are displayed on the web-page and are available for download.

## Test a dataset for normality...



**Figure 2:** An example of a data screening EATME application. Simple user interface elements, such as “sliders” are included to allow users to quickly examine their data sets. Ancillary functions such as data transformations may be applied with immediate effect.



## Core functions

The first release version of EATME includes a set of core functions that are relevant to – and used in – microbial ecology (Ramette 2007; Thioulouse et al. 2012). These methods are described at length in the Interactive Ecological Analysis Guide (IEAG; Deliverable 5.4), EATME’s sister project, and are listed below.

### *Diagnostic functions*

- Multivariate and univariate normality detection
- Multivariate and univariate outlier detection
- Multicollinearity detection

### *Unconstrained approaches*

- Non-metric dimensional scaling (NMDS)
- Metric dimensional scaling or principal coordinates analysis (PCoA)
- Principal components analysis (PCA)
- Hierarchical cluster analysis

### *Constrained approaches*

- Redundancy analysis (RDA)
- Distance-based redundancy analysis (db-RDA)
- Canonical correspondence analysis (CCA)

### *Hypothesis testing approaches*

- Analysis of similarity (ANOSIM)
- Permutational multivariate analysis of variance using distance matrices (PERMANOVA)

### *Spatial analysis*

- Principal coordinates of neighbour matrices (PCNM)

### *Ancillary functions*

- Data transformations
- Data and result export to comma separated value (CSV) formats
- Plot and graphical export to high resolution (300 dpi) tagged image file format (TIFF).
- Dissimilarity matrix calculation using a range of measures

## Maintenance and extensibility

By virtue of implementing EATME exclusively in R via shiny, the software can be maintained and extended entirely with R. New applications and features may be added in a straightforward and modular manner and updating existing applications does not require sweeping re-writes of the software. All code pertaining to EATME is under version control using subversion.



### Additional components and notes

- A set of example data sets have been included for demonstrative purposes and to guide users in the proper formatting of their own data sets.
- Currently, routines which would estimate the duration of a calculation task and prompt users that they will be notified of lengthy task completion via e-mail have not yet been implemented. At present, it is unclear if this functionality will be needed. Further testing with a wider range of data sets (a likely outcome of using EATME in teaching activities; see below) will clarify this need.

### Interaction with the Interactive Ecological Analysis Guide (IEAG)

The EATME web applications do not provide explanations of the methods implemented. These are described at length in the Interactive Ecological Analysis Guide (IEAG; Deliverable 5.4). EATME web applications will be paired with their corresponding IEAG pages, combining reference and application to assist dissemination and well-informed usage of these key analytical approaches through the consortium and the wider microbial ecology community.

### Use in teaching and dissemination

EATME will be demonstrated and used as a teaching aid in the upcoming "Micro B3 Marine Metagenomics Bioinformatics" course at the European Bioinformatics Institute (Hinxton, UK; March 24-28, 2014). This will provide a valuable opportunity to beta-test the software and enhance its functionality and usability in accordance with user behaviour.

### Reference list

- Filzmoser P & Gschwandtner M (2013) mvoutlier: Multivariate outlier detection based on robust methods.
- Filzmoser P, Maronna R & Werner M (2008) Outlier identification in high dimensions. *Comput Stat Data Anal* 52: 1694–1711.
- Harrell FE & with contributions from many other users (2009) Hmisc: Harrell Miscellaneous.
- Oksanen J et al. (2013) vegan: Community Ecology Package.
- R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramette A (2007) Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* 62: 142–160.
- RStudio Inc. (2013) shiny: Web Application Framework for R.
- Thioulouse J, Prin Y & Duponnois R (2012) Multivariate analyses in soil microbial ecology: a new paradigm. *Environ Ecol Stat* 19: 499–520.