

microbiome vignette

Leo Lahti and Jarkko Salojarvi

2015-02-04

Contents

microbiome R package 1

microbiome R package

The microbiome package contains general-purpose tools for microarray-based analysis of microbiome profiling data sets.

Example workflows

- [Minimal example](#)
- [Atlas](#)

Installation, example data sets and preprocessing

- [Installation](#)
- [Data](#)
- [RPA](#)
- [Preprocessing](#)
- [Phylogeny](#)

Visualization and related tools

- [Barplots](#)
- [Boxplots](#)
- [Heatmaps](#)
- [Matrix visualization](#)
- [Motion charts](#)
- [Ordination](#)
- [Oligo heatmap](#)
- [Cross hybridization](#)

Clustering

- [Bimodality](#)
- [Clustering](#)
- [Distance metrics](#)

Microbiota composition

- [Core microbiota](#)
- [Diversity](#)
- [Probe level studies](#)
- [Stability](#)

Linear models, comparisons, and association studies

- [Linear models](#)
- [Pairwise comparisons](#)
- [Cross correlations](#)

Other statistical analysis

- [ROC curves](#)
- [RDA](#)

Miscellaneous

- [leaveout](#)
- [misc](#)

Licensing and Citations

This work can be freely used, modified and distributed under the [Two-clause FreeBSD license](#).

Kindly cite the work as ‘Leo Lahti and Jarkko Salojarvi (2014). microbiome R package. URL: <http://microbiome.github.com>’.

References

You can embed citations, for example2: Lahti et al. (2014)

You can embed citations, for example: (Lahti et al. 2014)

Cite with DOI: (Abrams et al. 2012)

Cite URL (Lord 2012)

Packages: (Xie 2014a; Xie 2013; Xie 2014b)

For automated markdown citations, check [this](#), [this](#), and [this](#). The package utilizes tools from a number of other CRAN and Bioconductor extensions, including ade4, df2json, rjson, fastcluster, ggplot2, MASS, methods, minet, mixOmics, plyr, qvalue, RCurl, reshape2, RPA, vegan, and WGCNA. We thank all authors for these contributions:

- N. Caballero (2013). [df2json: Convert a dataframe to JSON](#)
- A. Couture-Beil (2013). [rjson: JSON for R](#)
- A. Dabney, John D. Storey and with assistance from Gregory R. Warnes. qvalue: Q-value estimation for false discovery rate control.

- S. Dray and A. B. Dufour, (2007): The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*. 22(4): 1-20.
- S. Dejean et al. (2013). [mixOmics: Omics Data Integration Project](#)
- L. Lahti et al. A fully scalable online-preprocessing algorithm for short oligonucleotide microarray atlases. *NAR* 41(10):e110, 2013
- L. Lahti et al. Analysis of Probe Reliability in Differential Gene Expression Studies with Short Oligonucleotide Arrays. *TCBB/IEEE* 8(1):217-225, 2011
- L. Lahti et al. Associations between the human intestinal microbiota, *Lactobacillus rhamnosus* GG and serum lipids indicated by integrated analysis of high-throughput profiling data. *PeerJ* 1:e32, 2013.
- D. T. Lang (2013). [RCurl: General network \(HTTP/FTP/...\) client interface for R](#)
- P. Langfelder and S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008, 9:559
- P. Langfelder, S. Horvath (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of Statistical Software*, 46(11), 1-17
- P. E. Meyer, Frederic Lafitte and Gianluca Bontempi (2008). MINET: An open source R/Bioconductor Package for Mutual Information based Network Inference. [BMC Bioinformatics](#)
- D. Mullner (2013). fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. *Journal of Statistical Software*, 53(9), 1-18
- Jari Oksanen 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. [ISBN 3-900051-07-0](#)
- W. N. Venables and B. D. Ripley (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- H. Wickham (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1-20
- H. Wickham. ggplot2: elegant graphics for data analysis. Springer New York, 2009.
- H. Wickham (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1), 1-29

Session info

This vignette was created with

```
sessionInfo()
```

```
## R version 3.1.2 (2014-10-31)
## Platform: x86_64-pc-linux-gnu (64-bit)
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
```

```

## [9] LC_ADDRESS=C LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] grid parallel stats graphics grDevices utils datasets
## [8] methods base
##
## other attached packages:
## [1] knitr_1.18.1 microbio_0.99.34 devtools_1.17.1
## [4] googleVis_0.5.6 limma_3.20.9 netresponse_1.17.13
## [7] mclust_4.4 minet_3.20.2 infotheo_1.2.0
## [10] Rgraphviz_2.8.1 graph_1.42.0 ggplot2_1.0.0
## [13] sorvi_0.7.13 dplyr_0.3.0.2 rddryad_0.1.1
## [16] knitr_1.8 gdata_2.13.3 AnnotationDbi_1.26.1
## [19] GenomeInfoDb_1.0.2 Biobase_2.24.0 BiocGenerics_0.10.0
## [22] RSQlite_1.0.0 DBI_0.3.1 reshape_0.8.5
## [25] vegan_2.2-1 lattice_0.20-29 permute_0.8-3
## [28] e1071_1.6-4 rmarkdown_0.3.10
##
## loaded via a namespace (and not attached):
## [1] acepack_1.3-3.3 ape_3.1-4 assertthat_0.1
## [4] bibtex_0.4.0 class_7.3-11 cluster_1.15.3
## [7] codetools_0.2-9 colorspace_1.2-4 df2json_0.0.2
## [10] digest_0.6.4 dmt_0.8.20 doParallel_1.0.8
## [13] dynamicTreeCut_1.6.2 evaluate_0.5.5 fastcluster_1.1.15
## [16] foreach_1.4.2 foreign_0.8-61 formatR_1.0
## [19] Formula_1.1-2 G0.db_2.14.0 gtable_0.1.2
## [22] gtools_3.4.1 Hmisc_3.14-5 htmltools_0.2.6
## [25] httr_0.5 igraph_0.7.1 impute_1.38.1
## [28] IRanges_1.22.10 iterators_1.0.7 labeling_0.3
## [31] latticeExtra_0.6-26 lazyeval_0.1.9 lubridate_1.3.3
## [34] magrittr_1.0.1 MASS_7.3-37 Matrix_1.1-4
## [37] matrixStats_0.10.3 memoise_0.2.1 mgcv_1.8-3
## [40] mixOmics_5.0-3 munsell_0.4.2 mvtnorm_1.0-0
## [43] nlme_3.1-118 nnet_7.3-8 OAIHarvester_0.1-7
## [46] pheatmap_0.7.7 plyr_1.8.1 preprocessCore_1.26.1
## [49] proto_0.3-10 qvalue_1.38.0 RColorBrewer_1.0-5
## [52] Rcpp_0.11.3 RCurl_1.95-4.3 RefManageR_0.8.45
## [55] reshape2_1.4.1 RGCCA_2.0 rgl_0.95.1158
## [58] rjson_0.2.15 RJSONIO_1.3-0 R.methodsS3_1.6.1
## [61] rpart_4.1-8 scales_0.2.4 splines_3.1.2
## [64] stats4_3.1.2 stringr_0.6.2 survival_2.37-7
## [67] tcltk_3.1.2 tools_3.1.2 WGCNA_1.43
## [70] XML_3.98-1.1 yaml_2.1.13

```

Abrams, Peter A., Lasse Ruokolainen, Brian J. Shuter, and Kevin S. McCann. 2012. "Harvesting Creates Ecological Traps: consequences of Invisible Mortality Risks in Predatorprey Metacommunities." *Ecology* 93 (2) (February): 281–293. doi:[10.1890/11-0011.1](https://doi.org/10.1890/11-0011.1). <http://dx.doi.org/10.1890/11-0011.1>.

Lahti, Leo, Jarkko Salojärvi, Anne Salonen, Marten Scheffer, and Willem M. de Vos. 2014. "Tipping Elements in the Human Intestinal Ecosystem." *Nature Communications* 5 (July): 4344. doi:[10.1038/ncomms5344](https://doi.org/10.1038/ncomms5344).

Lord, Phillip. 2012. "Greycite." *Knowledge Blog*. <http://knowledgeblog.org/greycite>. <http://knowledgeblog.org/greycite>.

- Xie, Yihui. 2013. *Dynamic Documents with R and Knitr*. Boca Raton, Florida: Chapman; Hall/CRC. <http://yihui.name/knitr/>.
- . 2014a. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. <http://yihui.name/knitr/>.
- . 2014b. “knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.