

microbiome vignette

Leo Lahti and Jarkko Salojarvi

2015-02-05

microbiome R package

The microbiome package contains general-purpose tools for microarray-based analysis of microbiome profiling data sets in R (R Core Team, 2013); also relevant (Venables and Ripley, 2002)

Example workflows

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

Installation, example data sets and preprocessing

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

The peerj32 example data set is from (Lahti, Salonen, Kekkonen, et al., 2013).

Preprocessing of HITChip data is based on RPA [(Lahti, Torrente, Elo, et al., 2013); (Lahti, Elo, Aittokallio, et al., 2011)]

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 Salojärvi (2014). microbiome R package. URL: <http://microbiome.github.com>’.

Dependencies

The package utilizes tools from a number of other CRAN and Bioconductor extensions, including:

- [df2json](#) (Caballero, 2013)
- [rjson](#) (Couture-Beil, 2014)
- [ade4](#) (Dray and Dufour, 2007; Chessel, Dufour, and Thioulouse, 2004; Dray, Dufour, and Chessel, 2007)
- [mixOmics](#) (Dejean, Gonzalez, Monget, et al., 2014)
- [RCurl](#) (Temple Lang, 2014)
- [vegan](#) (Oksanen, Blanchet, Kindt, et al., 2015)
- [reshape](#) (Wickham and Hadley, 2007)
- [WGCNA](#) (Langfelder and Horvath, 2008; Langfelder and Horvath, 2012)
- [ggplot2](#) (Wickham, 2009)
- [RPA](#) (Lahti, Torrente, Elo, et al., 2013)
- [minet](#) (Meyer, Lafitte, and Bontempi, 2008)
- [fastcluster](#) (Müllner, 2013)
- [plyr](#) (Wickham, 2011)

References

- [1] N. Caballero. *df2json: Convert a dataframe to JSON*. R package version 0.0.2. 2013. .
- [2] D. Chessel, A. Dufour and J. Thioulouse. “The ade4 package-I: One-table methods”. In: *R News* 4 (2004), pp. 5-10.
- [3] A. Couture-Beil. *rjson: JSON for R*. R package version 0.2.15. 2014. .
- [4] S. Dejean, I. Gonzalez, K. L. C. w. c. f. P. Monget, et al. *mixOmics: Omics Data Integration Project*. R package version 5.0-3. 2014. .
- [5] S. Dray and A. Dufour. “The ade4 package: implementing the duality diagram for ecologists”. In: *Journal of Statistical Software* 22.4 (2007), pp. 1-20.
- [6] S. Dray, A. Dufour and D. Chessel. “The ade4 package-II: Two-table and K-table methods.” In: *R News* 7.2 (2007), pp. 47-52.
- [7] L. Lahti, L. L. Elo, T. Aittokallio, et al. “Probabilistic Analysis of Probe Reliability in Differential Gene Expression Studies with Short Oligonucleotide Arrays”. In: *IEEE/ACM Trans. Comput. Biol. and Bioinf.* 8.1 (Jan. 2011), pp. 217-225. DOI: 10.1109/tcbb.2009.38. .
- [8] L. Lahti, A. Salonen, R. A. Kekkonen, et al. “Associations between the human intestinal microbiota, Lactobacillus rhamnosus GG and serum lipids indicated by integrated analysis of high-throughput profiling data “. In: *PeerJ* 1 (2013), p. e32. DOI: 10.7717/peerj.32. .
- [9] L. Lahti, A. Torrente, L. L. Elo, et al. “A fully scalable online-preprocessing algorithm for short oligonucleotide microarray atlases”. In: *Nucleic Acids Research* 41 (10 2013). R/BioC: <http://bioconductor.org/packages/release/bioc/html/RPA.html>, p. e110. .
- [10] P. Langfelder and S. Horvath. “Fast R Functions for Robust Correlations and Hierarchical Clustering”. In: *Journal of Statistical Software* 46.11 (2012), pp. 1-17. .
- [11] P. Langfelder and S. Horvath. “WGCNA: an R package for weighted correlation network analysis”. In: *BMC Bioinformatics* (2008), p. 559.
- [12] P. E. Meyer, F. Lafitte and G. Bontempi. “MINET: An open source R/Bioconductor Package for Mutual Information based Network Inference”. In: *BMC Bioinformatics* 9 (2008). .
- [13] D. Müllner. “fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python”. In: *Journal of Statistical Software* 53.9 (2013), pp. 1-18. .
- [14] J. Oksanen, F. G. Blanchet, R. Kindt, et al. *vegan: Community Ecology Package*. R package version 2.2-1. 2015. .
- [15] R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013. ISBN: ISBN 3-900051-07-0. .
- [16] D. Temple Lang. *RCurl: General network (HTTP/FTP/...) client interface for R*. R package version 1.95-4.3. 2014. .
- [17] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. fourth. New York: Springer, 2002. ISBN: ISBN .
- [18] H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN: 978-0-387-98140-6. .
- [19] H. Wickham. “The Split-Apply-Combine Strategy for Data Analysis”. In: *Journal of Statistical Software* 40.1 (2011), pp. 1-29. .
- [20] Wickham and Hadley. “Reshaping data with the reshape package”. In: *Journal of Statistical Software* 21.12 (2007). .

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] parallel stats      graphics  grDevices utils      datasets  methods
## [8] base
##
## other attached packages:
##  [1] microbiome_0.99.34  ggplot2_1.0.0      sorvi_0.7.13
##  [4] dplyr_0.3.0.2       rdryad_0.1.1       AnnotationDbi_1.26.1
##  [7] GenomeInfoDb_1.0.2  Biobase_2.24.0     BiocGenerics_0.10.0
## [10] RSQlite_1.0.0       DBI_0.3.1          reshape_0.8.5
## [13] vegan_2.2-1         lattice_0.20-29    permute_0.8-3
## [16] e1071_1.6-4         devtools_1.7.0     knitr_1.15.3
## [19] 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        doParallel_1.0.8   dynamicTreeCut_1.62
## [13] evaluate_0.5.5      fastcluster_1.1.15 foreach_1.4.2
## [16] foreign_0.8-61      formatR_1.0        Formula_1.1-2
## [19] gdata_2.13.3        GO.db_2.14.0       grid_3.1.2
## [22] gtable_0.1.2        gtools_3.4.1       Hmisc_3.14-5
## [25] htmltools_0.2.6     httr_0.5           igraph_0.7.1
## [28] impute_1.38.1       IRanges_1.22.10    iterators_1.0.7
## [31] knitr_1.8           labeling_0.3        latticeExtra_0.6-26
## [34] lazyeval_0.1.9      lubridate_1.3.3    magrittr_1.0.1
## [37] MASS_7.3-37         Matrix_1.1-4       matrixStats_0.10.3
## [40] memoise_0.2.1       mgcv_1.8-3         mixOmics_5.0-3
## [43] munsell_0.4.2       nlme_3.1-118       nnet_7.3-8
## [46] OAIHarvester_0.1-7  pheatmap_0.7.7     plyr_1.8.1
## [49] preprocessCore_1.26.1 proto_0.3-10       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] tools_3.1.2         WGCNA_1.43         XML_3.98-1.1
```

```
## [70] yaml_2.1.13
```