










Easy16S: a user-friendly Shiny web-service for exploration and visualization of microbiome data.

Cédric Midoux ^{1,2,3}, Olivier Rué ^{2,3}, Olivier Chapleur ¹, Ariane Bize ¹,
Valentin Loux ^{2,3}, and Mahendra Mariadassou ^{2,3}

¹ Université Paris-Saclay, INRAE, PROSE, 92761, Antony, France ² Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France ³ Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

DOI: [10.21105/joss.06704](https://doi.org/10.21105/joss.06704)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Frederick Boehm](#)  

Reviewers:

- [@adriantaudiere](#)
- [@tmaklin](#)

Submitted: 27 March 2024

Published: 15 November 2024

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

The analysis of microbiome data has become a major asset for investigating microbial diversity and dynamics, in diverse fields, like health ([Chatelier et al., 2013](#)), environmental studies ([Karimi et al., 2020](#)), food-processing ([Chaillou et al., 2014](#)), or environmental biotechnologies ([Poirier et al., 2016](#)). Due to sequencing advances, microbiome studies now require exploration, analysis and interpretation of large and high-dimensional datasets. Metabarcoding approaches, in particular, are based on a two-step process. First, a bioinformatics pipeline processes raw amplicon sequencing reads into Operational Taxonomic Unit (OTU) or Amplicon Sequence Variant (ASV), generating counts and taxonomic affiliations for each of them. Second, these tables are enriched with sample metadata to investigate relevant biological questions using statistical analyses. The affordability of amplicon sequencing has led to its widespread use in microbial ecology. Therefore, there is a growing demand for user-friendly interactive tools, enabling researchers to analyze their data autonomously, alleviating the dependence on bioinformaticians, biostatisticians or the need to acquire skills in R programming.

Regarding the bioinformatics part, many solutions are available to produce count tables from reads ([Hakimzadeh et al., 2023](#)), relying either on command-line tools provided through a Galaxy interface (ie: QIIME ([Caporaso et al., 2010](#)), FROGS ([Escudié et al., 2017](#))), or on R pipelines (ie: DADA2 ([Callahan et al., 2016](#))). For the second step, several packages dedicated to the analysis and visualization of microbiomes are available, such as phyloseq ([Paul J. McMurdie & Holmes, 2013](#)), microbiome ([Lahti & Shetty, 2012-2019](#)), metacoder ([Foster et al., 2017](#)) and many others. Their use can however sometimes be complex. There are relatively few tools available for rapid, interactive analysis of microbiome data (shiny-phyloseq ([P. J. McMurdie & Holmes, 2014](#)) is no longer supported ; ranacapa ([Kandlikar et al., 2018](#)) lacks flexibility, especially during import step ; animalcules ([Zhao et al., 2021](#)) requires local installation ; shaman ([Volant et al., 2020](#)) is integrated into a global workflow and cannot be decoupled from the bioinformatics components).

Statement of need

Here, we introduce Easy16S, an R-package and an interactive Shiny application ([Chang et al., 2012](#)), aiming to facilitate exploratory microbiome data analysis, data visualization, and statistical analysis. This tool is specifically designed for biologists eager to swiftly explore their data and generate figures interactively. It is easy-to-use and especially focused on the mapping of covariates of interest to microbiome structure.

This application is built upon phyloseq-class objects powered by ([Paul J. McMurdie & Holmes, 2013](#)). These objects integrate a matrix of OTU/ASV abundances per sample, a data.frame of

metadata and covariates associated to the samples, a matrix of taxonomic affiliations for each OTU/ASV, and optionally, a phylogenetic tree and reference sequences for the OTU/ASV.

App Features

For data loading, users have three options: they can use one of the demo dataset, upload flat files to enable the application to construct a phyloseq object, or directly upload a phyloseq object.

Before analysis, uploaded data can be preprocessed by the user to help refine and clean raw data. This includes options such as sample/taxa filtering, modification of the taxonomy table, rarefaction of data and application of mathematical functions to transform the count matrix.

The various exploration and analysis sections are :

- Key tables constituting the phyloseq object;
- Metadata visualization using *esquisse* (Meyer & Perrier, 2018);
- Taxonomic composition barplot;
- Rarefaction curves;
- Abundance heatmap;
- Richness within a sample (α -diversity): table, scatterplot and ANOVA;
- Dissimilarity between samples (β -diversity): table, sample heatmap, sample clustering, MultiDimensional Scaling and Multivariate ANOVA;
- Principal Component Analysis;
- Differential abundance analysis with DESeq2 (Love et al., 2014) with sensible defaults.

Users can export preprocessed data for further analysis within R or for use in future Easy16S sessions. Additionally, tables and plots can be easily exported, enhancing the usability and accessibility of both data and results. This flexibility enables users to seamlessly integrate Easy16S with their preferred analysis tools and workflows.

Use-case

Three major use cases have been identified for Easy16S:

- Easy16S empowers beginner users to autonomously conduct their analyses without any technical skills.
- More advanced users can use Easy16S to swiftly explore data, identify patterns, before developing their R code for a more in-depth analysis.
- During training sessions, Easy16S serves as a valuable tool, allowing users to concentrate on mastering biological concepts without being encumbered by programming challenges.

Nevertheless, the simplicity of analysis should not mask the complexity of interpretation and conclusion. Some parameters and guardrails have been established to decrease the risk of misuse of statistics. However, the outcomes must be analyzed with caution and must not be over-interpreted.

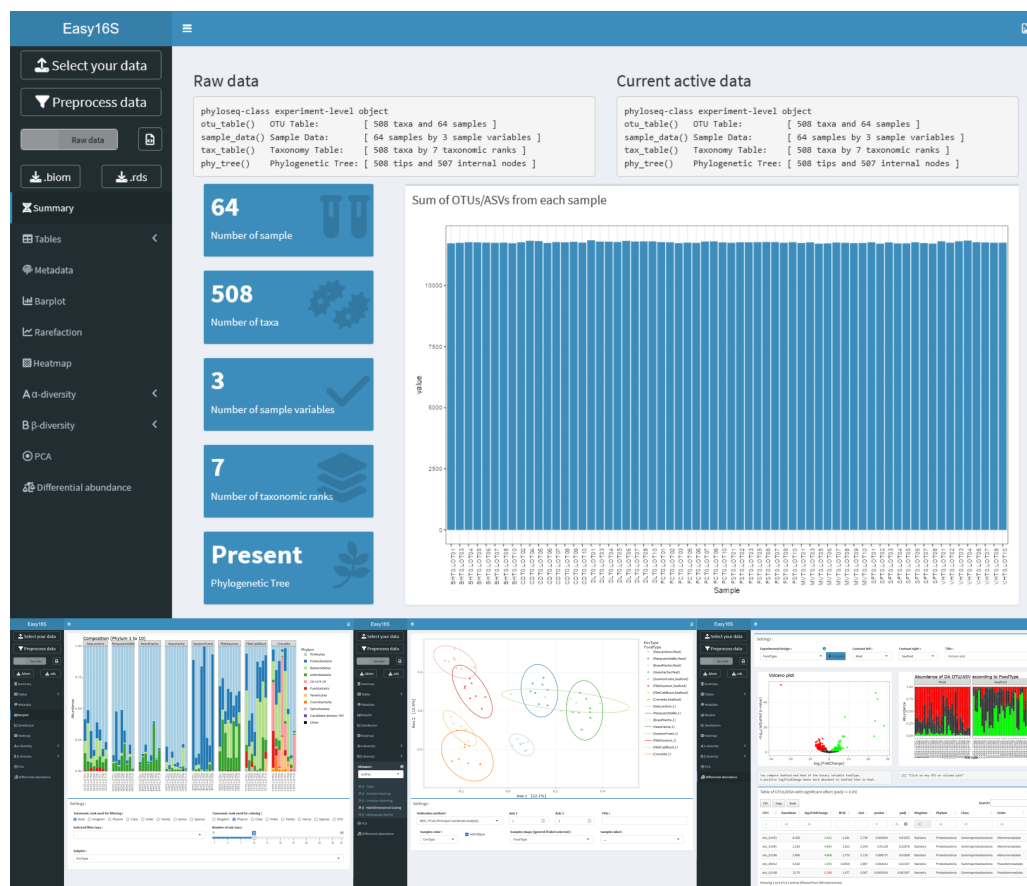


Figure 1: Summary in Easy16S and three examples of data visualization

Acknowledgements

We are grateful to the INRAE MIGALE (2018) bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi: 10.15454/1.5572390655343293E12) for providing help, computing and storage resources. Special thanks to Chrystelle Bureau, Patrick Dabert, and all testers for their invaluable tests and suggestions during the development of Easy16S. We also extend our thanks to the INRAE SK8 Team for their advice and recommendations throughout the Easy16S development and deployment.

References

- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., ... Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5), 335–336. <https://doi.org/10.1038/nmeth.f.303>
- Chaillou, S., Chaulot-Talmon, A., Caekebeke, H., Cardinal, M., Christieans, S., Denis, C., Desmonts, M. H., Dousset, X., Feurer, C., Hamon, E., Joffraud, J.-J., Carbona, S. L., Leroi, F., Leroy, S., Lorre, S., Macé, S., Pilet, M.-F., Prévost, H., Rivollier, M., ...

- Champomier-Vergès, M.-C. (2014). Origin and ecological selection of core and food-specific bacterial communities associated with meat and seafood spoilage. *The ISME Journal*, 9(5), 1105–1118. <https://doi.org/10.1038/ismej.2014.202>
- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2012). Shiny: Web application framework for r. In *CRAN: Contributed Packages*. The R Foundation. <https://doi.org/10.32614/CRAN.package.shiny>
- Chatelier, E. L., Trine Nielsen, and, Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Garup, N., Jørgensen, T., Brandslund, I., Nielsen, H. B., Juncker, A. S., Bertalan, M., ... Pedersen, O. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature*, 500(7464), 541–546. <https://doi.org/10.1038/nature12506>
- Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S., Hernandez-Raquet, G., Combes, S., & Pascal, G. (2017). FROGS: Find, rapidly, OTUs with galaxy solution. *Bioinformatics*, 34(8), 1287–1294. <https://doi.org/10.1093/bioinformatics/btx791>
- Foster, Z. S. L., Sharpton, T. J., & Grünwald, N. J. (2017). Metacoder: An r package for visualization and manipulation of community taxonomic diversity data. *PLOS Computational Biology*, 13(2), e1005404. <https://doi.org/10.1371/journal.pcbi.1005404>
- Hakimzadeh, A., Abdala Asbun, A., Albanese, D., Bernard, M., Buchner, D., Callahan, B., Caporaso, J. G., Curd, E., Djemiel, C., Brandström Durling, M., Elbrecht, V., Gold, Z., Gweon, H. S., Hajibabaei, M., Hildebrand, F., Mikryukov, V., Normandeau, E., Özkurt, E., M. Palmer, J., ... Anslan, S. (2023). A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses. *Molecular Ecology Resources*. <https://doi.org/10.1111/1755-0998.13847>
- Kandlikar, G. S., Gold, Z. J., Cowen, M. C., Meyer, R. S., Freise, A. C., Kraft, N. J. B., Moberg-Parker, J., Sprague, J., Kushner, D. J., & Curd, E. E. (2018). ranacapa: An r package and shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations. *F1000Research*, 7, 1734. <https://doi.org/10.12688/f1000research.16680.1>
- Karimi, B., Villerd, J., Dequiedt, S., Terrat, S., Bouré, N. C.-P., Djemiel, C., Lelièvre, M., Tripied, J., Nowak, V., Saby, N. P. A., Bispo, A., Jolivet, C., Arrouays, D., Wincker, P., Cruaud, C., & Ranjard, L. (2020). Biogeography of soil microbial habitats across france. *Global Ecology and Biogeography*, 29(8), 1399–1411. <https://doi.org/10.1111/geb.13118>
- Lahti, L., & Shetty, S. (2012-2019). *Microbiome r package*. <https://doi.org/10.18129/b9.bioc.microbiome>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12). <https://doi.org/10.1186/s13059-014-0550-8>
- McMurdie, Paul J., & Holmes, S. (2013). Phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8(4), e61217. <https://doi.org/10.1371/journal.pone.0061217>
- McMurdie, P. J., & Holmes, S. (2014). Shiny-phyloseq: Web application for interactive microbiome analysis with provenance tracking. *Bioinformatics*, 31(2), 282–283. <https://doi.org/10.1093/bioinformatics/btu616>
- Meyer, F., & Perrier, V. (2018). Esquisse: Explore and visualize your data interactively. In *CRAN: Contributed Packages*. The R Foundation. <https://doi.org/10.32614/CRAN.package.esquisse>
- MIGALE. (2018). *Migale bioinformatics facility*. INRAE. <https://doi.org/10.15454/1.5572390655343293E12>

- Poirier, S., Bize, A., Bureau, C., Bouchez, T., & Chapleur, O. (2016). Community shifts within anaerobic digestion microbiota facing phenol inhibition: Towards early warning microbial indicators? *Water Research*, *100*, 296–305. <https://doi.org/10.1016/j.watres.2016.05.041>
- Volant, S., Lechat, P., Woringer, P., Motreff, L., Campagne, P., Malabat, C., Kennedy, S., & Ghoulane, A. (2020). SHAMAN: A user-friendly website for metataxonomic analysis from raw reads to statistical analysis. *BMC Bioinformatics*, *21*(1). <https://doi.org/10.1186/s12859-020-03666-4>
- Zhao, Y., Federico, A., Faits, T., Manimaran, S., Segrè, D., Monti, S., & Johnson, W. E. (2021). Animalcules: Interactive microbiome analytics and visualization in r. *Microbiome*, *9*(1). <https://doi.org/10.1186/s40168-021-01013-0>