




Tidytacos: An R package for analyses on taxonomic composition of microbial communities



Stijn Wittouck ¹, Tim Van Rillaer ¹, Wenke Smets ¹, and Sarah Lebeer ¹

¹ Department of Bioscience Engineering, University of Antwerp, Belgium  Corresponding author

DOI: [10.21105/joss.06313](https://doi.org/10.21105/joss.06313)

Software

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

Editor: [Renata Diaz](#)  

Reviewers:

- [@kelly-sovacool](#)
- [@david-barnett](#)

Submitted: 13 November 2023

Published: 31 January 2025

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary



The tidytacos (tidy TAXonomic COMpositionS) package is an R package for the exploration of microbial community data. Such community data consists of read counts generated by amplicon sequencing (e.g. a region of the 16S rRNA gene) or metagenome (shotgun) sequencing. tidytacos is based on the tidy principle introduced by Hadley Wickham where the data are stored in a consistent format ([Wickham et al., 2023](#)). Specifically, tidytacos uses the tidy format and grammar to select, transform, and prepare microbial community data for visualization and analyses. In addition, it offers a range of functions for popular and less known analyses and visualizations of microbial community data. tidytacos is designed for researchers across various levels of expertise, to offer both an improved accessibility to microbial community data and to allow easy transformation of data for novel approaches of visualization and analysis.

Statement of need

In the past two decades, DNA sequencing for the determination of microbial community composition has emerged as a powerful and transformative tool in the field of microbiology. Advancements in sequencing technologies have made it possible to explore and understand the vast diversity of microbial communities present in various environments (including the human microbiome) with unprecedented accuracy and depth at increasingly lower costs. This has led to more and more researchers, many with no prior experience in this field, attempting to determine the microbial community composition of samples using amplicon and metagenome sequencing technologies. Unsurprisingly, there is a high need for accessible tools both to process raw sequencing reads to taxonomically annotated read counts and subsequently interpret these taxonomic compositions. Researchers having to handle this kind of data for the first time often struggle with the gathering, transformation, and preparation of data for visualisation and analyses, which is a crucial step in the data analysis process. These steps lay the foundation of accurate and reliable insights. tidytacos is a toolkit for data wrangling of DNA-sequencing-based microbial community data. It increases accessibility of microbial community data, but also allows researchers who are more experienced with this type of data to address novel questions. The tool includes a range of simple functions for common exploratory analyses, visualizations, and statistical analyses. tidytacos offers similar functionalities as packages

such as `microViz` (Barnett et al., 2021), `mia` (Borman et al., 2025), and `phyloseq` (McMurdie & Holmes, 2013), but was designed specifically to be very accessible to researchers with limited R and general programming experience. To achieve this, `tidytacos` was embedded in the `tidyverse` framework, which provides a style of programming that is very intuitive for new (but also experienced) programmers and is often the starting point for many people learning R. `Tidytacos` is also very easy to install, since it has only the main `tidyverse` packages as dependencies for the vast majority of its functionality. Finally, `tidytacos` also serves the needs of researchers who are looking to implement some rarely used or new visualizations and methods of analysis. `tidytacos` has already been used in a number of scientific publications by different researchers (Ahannach et al., 2021, 2024; Allonsius et al., 2019; De Boeck et al., 2017, 2019, 2020, 2021, 2022; Jörissen et al., 2021; Lebeer et al., 2022, 2023; Legein et al., 2022; E. Oerlemans et al., 2022; E. F. Oerlemans et al., 2020; Papini et al., 2023; Peng et al., 2020; Smets et al., 2022, 2023; Spacova et al., 2022; M. Van Tendeloo, Bundervoet, et al., 2021; M. Van Tendeloo, Xie, et al., 2021; Weyns et al., 2024) and has also been used by several undergraduate, graduate, and PhD students during their research projects (Dillen, 2024; Eilers, 2023; Legein, 2022; I. M. Van Tendeloo, 2022), demonstrating its versatility in topics and user experience.

Functionality

`tidytacos` is a package for R, which is the most widely used programming language for microbial community analyses. It adheres to the tidy principles of coding and of structuring data and builds on the `tidyverse` collection of packages (Wickham et al., 2019). The package stores microbial community data in one “`tidytacos` object” which is a list of three tables: 1) The counts: these are the counts of reads for each taxon (OTU/ASV/phyloTYPE) in each sample. Each row contains exactly one read count, representing a single observation. This is the typical long format in the `tidyverse`. 2) The samples: this table contains the sample metadata. Each row represents a sample. 3) The taxa: this table contains the taxonomy and other metadata for the taxa. Each row represents a taxon. The package is called `tidytacos` because each of the tables is tidy: each row represents an observation and each column a variable. Additionally, it adopts the grammar for data manipulation introduced by the `dplyr` package of the `tidyverse`, allowing, for example, the filtering and selection of samples and taxa and their associated information (e.g. `filter_samples`, `filter_taxa`, `select_samples`, `select_taxa`). Also, aggregating samples or taxa (`aggregate_samples`, `aggregate_taxa`) and conversion of the taxon counts (`add_rel_abundance/add_clr_abundance`) are supported, including a function to determine absolute abundances (`add_absolute_abundance`) based on the spike-in method (Smets et al., 2016; Wang et al., 2021). There are quick functions for visualizing community composition of samples using stacked bar plots (`tacoplot_stack`) or visualizing a two-dimensional ordination of samples (`tacoplot_ord`) based on the `ggplot2` package of the `tidyverse`. Additionally, functions to study the effect of sample variables on community variation are available (e.g. `perform_adonis`) and differentially abundant taxa can be determined with a method based on the logratio differences introduced by Aitchison (1982). The full overview of functions, code, a quick start tutorial, and more advanced tutorials are available on the [Github](https://github.com/LebeerLab/tidytacos) page, <https://github.com/LebeerLab/tidytacos>.

Acknowledgements

We would like to acknowledge the work of the R-core team who maintain the R language and the contributors of the `tidyverse`. We also want to thank other members and alumni of the Lebeer Lab who contributed code and/or documentation: Wannes Van Beeck, Sander Wuyts, and Tom Eilers. We furthermore acknowledge the financial support of the FWO (fellowship 11A0620N of SW and fellowship 12ZJ821N of WS) and ERC (starting grant Lacto-Be, 852600, of SL).

References

- Ahannach, S., Delanghe, L., Spacova, I., Wittouck, S., Van Beeck, W., De Boeck, I., & Lebeer, S. (2021). Microbial enrichment and storage for metagenomics of vaginal, skin, and saliva samples. *Iscience*, 24(11). <https://doi.org/10.1016/j.isci.2021.103306>
- Ahannach, S., Gehrman, T., Spacova, I., Wittouck, S., Hiers, J., Bron, P., Donck, L. V., Cromphout, M., Swayambhu, M., Arora, N., Schuh, L., Tournoy, I., Smeers, I., Wuestenbergs, J., Bekaert, B., Decorte, R., Jehaes, E., & Lebeer, S. (2024). *Microbial and seminal traces of sexual intercourse and forensic implications*. <https://doi.org/10.21203/rs.3.rs-4302243/v1>
- Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2), 139–160. <https://doi.org/10.1111/j.2517-6161.1982.tb01195.x>
- Allonsius, C. N., Van Beeck, W., De Boeck, I., Wittouck, S., & Lebeer, S. (2019). The microbiome of the invertebrate model host *Galleria mellonella* is dominated by *Enterococcus*. *Animal Microbiome*, 1, 1–7. <https://doi.org/10.1186/s42523-019-0010-6>
- Barnett, D. J., Arts, I. C., & Penders, J. (2021). microViz: An R package for microbiome data visualization and statistics. *Journal of Open Source Software*, 6(63), 3201. <https://doi.org/10.21105/joss.03201>
- Borman, T., Ernst, F. G. M., Shetty, S. A., & Lahti, L. (2025). *Mia: Microbiome analysis*. <https://github.com/microbiome/mia>
- De Boeck, I., Broek, M. F. van den, Allonsius, C. N., Spacova, I., Wittouck, S., Martens, K., Wuyts, S., Cauwenberghs, E., Jokicevic, K., Vandenneuvel, D., & others. (2020). Lactobacilli have a niche in the human nose. *Cell Reports*, 31(8). <https://doi.org/10.2139/ssrn.3422481>
- De Boeck, I., Cauwenberghs, E., Spacova, I., Gehrman, T., Eilers, T., Delanghe, L., Wittouck, S., Bron, P. A., Henkens, T., Gamgami, I., & others. (2022). Randomized, double-blind, placebo-controlled trial of a throat spray with selected lactobacilli in COVID-19 outpatients. *Microbiology Spectrum*, 10(5), e01682–22. <https://doi.org/10.1128/spectrum.01682-22>
- De Boeck, I., Wittouck, S., Martens, K., Claes, J., Jorissen, M., Steelant, B., Broek, M. F. van den, Seys, S. F., Hellings, P. W., Vanderveken, O. M., & others. (2019). Anterior nares diversity and pathobionts represent sinus microbiome in chronic rhinosinusitis. *MSphere*, 4(6), 10–1128. <https://doi.org/10.1128/msphere.00532-19>
- De Boeck, I., Wittouck, S., Martens, K., Spacova, I., Cauwenberghs, E., Allonsius, C. N., Jörissen, J., Wuyts, S., Van Beeck, W., Dillen, J., & others. (2021). The nasal mutualist *Dolosigranulum pigrum* AMBR11 supports homeostasis via multiple mechanisms. *Iscience*, 24(9). <https://doi.org/10.1016/j.isci.2021.102978>
- De Boeck, I., Wittouck, S., Wuyts, S., Oerlemans, E. F., Van den Broek, M. F., Vandenneuvel, D., Vanderveken, O., & Lebeer, S. (2017). Comparing the healthy nose and nasopharynx microbiota reveals continuity as well as niche-specificity. *Frontiers in Microbiology*, 8, 2372. <https://doi.org/10.3389/fmicb.2017.02372>
- Dillen, J. (2024). *Investigating the role of antimicrobial specialized metabolites and bacteriocins in the vaginal microbial community* [PhD thesis, University of Antwerp]. <https://doi.org/10.63028/10067/2067410151162165141>
- Eilers, T. (2023). *Functional potential and ecology of Lactobacillaceae in vegetable fermentations* [PhD thesis]. University of Antwerp.
- Jörissen, J., Broek, M. F. van den, De Boeck, I., Van Beeck, W., Wittouck, S., Boudewyns, A., Van de Heyning, P., Topsakal, V., Van Rompaey, V., Wouters, I., & others. (2021).

- Case-control microbiome study of chronic otitis media with effusion in children points at streptococcus salivarius as a pathobiont-inhibiting species. *MSystems*, 6(2), 10–1128. <https://doi.org/10.1128/msystems.00056-21>
- Lebeer, S., Ahannach, S., Gehrman, T., Wittouck, S., Eilers, T., Oerlemans, E., Condori, S., Dillen, J., Spacova, I., Vander Donck, L., & others. (2023). A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. *Nature Microbiology*, 1–13. <https://doi.org/10.1038/s41564-023-01500-0>
- Lebeer, S., Oerlemans, E. F., Claes, I., Henkens, T., Delanghe, L., Wuyts, S., Spacova, I., Broek, M. F. van den, Tuytaerts, I., Wittouck, S., & others. (2022). Selective targeting of skin pathobionts and inflammation with topically applied lactobacilli. *Cell Reports Medicine*, 3(2). <https://doi.org/10.1016/j.xcrm.2022.100521>
- Legein, M. (2022). *Phyllosphere bacteria in greenhouses and their use in biocontrol* [PhD thesis]. University of Antwerp.
- Legein, M., Smets, W., Wuyts, K., Bosmans, L., Samson, R., & Lebeer, S. (2022). The greenhouse phyllosphere microbiome and associations with introduced bumblebees and predatory mites. *Microbiology Spectrum*, 10(4), e01755–22. <https://doi.org/10.1128/spectrum.01755-22>
- McMurdie, P. 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>
- Oerlemans, E. F., Bellen, G., Claes, I., Henkens, T., Allonsius, C. N., Wittouck, S., Broek, M. F. van den, Wuyts, S., Kiekens, F., Donders, G. G., & others. (2020). Impact of a lactobacilli-containing gel on vulvovaginal candidosis and the vaginal microbiome. *Scientific Reports*, 10(1), 7976. <https://doi.org/10.1038/s41598-020-64705-x>
- Oerlemans, E., Ahannach, S., Wittouck, S., Dehay, E., De Boeck, I., Ballet, N., Rodriguez, B., Tuytaerts, I., & Lebeer, S. (2022). Impacts of menstruation, community type, and an oral yeast probiotic on the vaginal microbiome. *Msphere*, 7(5), e00239–22. <https://doi.org/10.1128/msphere.00239-22>
- Papini, G., Muys, M., Van Winckel, T., Meerburg, F. A., Van Beeck, W., Vermeir, P., & Vlaeminck, S. E. (2023). Boosting aerobic microbial protein productivity and quality on brewery wastewater: Impact of anaerobic acidification, high-rate process and biomass age. *Bioresource Technology*, 368, 128285. <https://doi.org/10.1016/j.biortech.2022.128285>
- Peng, L., Xie, Y., Van Beeck, W., Zhu, W., Van Tendeloo, M., Tytgat, T., Lebeer, S., & Vlaeminck, S. E. (2020). Return-sludge treatment with endogenous free nitrous acid limits nitrate production and n₂o emission for mainstream partial nitrification/anammox. *Environmental Science & Technology*, 54(9), 5822–5831. <https://doi.org/10.1021/acs.est.9b06404>
- Smets, W., Chock, M. K., Walsh, C. M., Vanderburgh, C. Q., Kau, E., Lindow, S. E., Fierer, N., & Koskella, B. (2023). Leaf side determines the relative importance of dispersal versus host filtering in the phyllosphere microbiome. *mBio*, e01111–23. <https://doi.org/10.1128/mbio.01111-23>
- Smets, W., Leff, J. W., Bradford, M. A., McCulley, R. L., Lebeer, S., & Fierer, N. (2016). A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biology and Biochemistry*, 96, 145–151. <https://doi.org/10.1016/j.soilbio.2016.02.003>
- Smets, W., Spada, L. M., Gandolfi, I., Wuyts, K., Legein, M., Muyschondt, B., Samson, R., Franzetti, A., & Lebeer, S. (2022). Bacterial succession and community dynamics of the emerging leaf phyllosphere in spring. *Microbiology Spectrum*, 10(2), e02420–21. <https://doi.org/10.1128/spectrum.02420-21>

- Spacova, I., Ahannach, S., Breynaert, A., Erreygers, I., Wittouck, S., Bron, P. A., Van Beeck, W., Eilers, T., Alloul, A., Blansaer, N., & others. (2022). Spontaneous riboflavin-overproducing *Limosilactobacillus reuteri* for biofortification of fermented foods. *Frontiers in Nutrition*, 9, 1231. <https://doi.org/10.3389/fnut.2022.916607>
- Van Tendeloo, I. M. (2022). *Resource-efficient nitrogen removal from sewage: Kinetic, physical and chemical tools for mainstream partial nitrification/anammox* [PhD thesis]. University of Antwerp.
- Van Tendeloo, M., Bundervoet, B., Carlier, N., Van Beeck, W., Mollen, H., Lebeer, S., Colson, J., & Vlaeminck, S. E. (2021). Piloting carbon-lean nitrogen removal for energy-autonomous sewage treatment. *Environmental Science: Water Research & Technology*, 7(12), 2268–2281. <https://doi.org/10.1039/d1ew00525a>
- Van Tendeloo, M., Xie, Y., Van Beeck, W., Zhu, W., Lebeer, S., & Vlaeminck, S. E. (2021). Oxygen control and stressor treatments for complete and long-term suppression of nitrite-oxidizing bacteria in biofilm-based partial nitrification/anammox. *Bioresour. Technol.*, 342, 125996. <https://doi.org/10.1016/j.biortech.2021.125996>
- Wang, X., Howe, S., Deng, F., & Zhao, J. (2021). Current applications of absolute bacterial quantification in microbiome studies and decision-making regarding different biological questions. *Microorganisms*, 9(9), 1797. <https://doi.org/10.3390/microorganisms9091797>
- Weyns, A.-S., Ahannach, S., Van Rillaer, T., De Bruyne, T., Lebeer, S., & Hermans, N. (2024). Enhancing pediatric attention-deficit hyperactivity disorder treatment: Exploring the gut microbiota effects of french maritime pine bark extract and methylphenidate intervention. *Frontiers in Nutrition*, 11. <https://doi.org/10.3389/fnut.2024.1422253>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., & others. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. " O'Reilly Media, Inc."