

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

Stijn Wittouck<sup>1</sup>, Tim Van Rillaer<sup>1</sup>, Wenke Smets<sup>1</sup>, and Sarah Lebeer<sup>1</sup>

<sup>1</sup> Department of Bioscience Engineering, University of Antwerp, Belgium ¶ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

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

Editor: [Renata Diaz](#)

## Reviewers:

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

Submitted: 13 November 2023

Published: unpublished

## License

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

## Summary



7

8 The tidytacos (tidy TAXonomic COMpositionS) package is an R package for the exploration  
9 of microbial community data. Such community data consists of read counts generated  
10 by amplicon sequencing (e.g. a region of the 16S rRNA gene) or metagenome (shotgun)

sequencing. tidy\_tacos is based on the tidy principle introduced by Hadley Wickham where the data are stored in a consistent format (Wickham et al., 2023). Specifically, tidy\_tacos 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. tidy\_tacos 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, some 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. tidy\_tacos 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. tidy\_tacos was designed with two types of users in mind. A first target audience are newcomers in the field of microbial ecology with little programming experience, who need an intuitive and user-friendly way of coding their analyses. The second group are more advanced users who are looking to implement some rarely used or new visualizations and methods of analysis. tidy\_tacos has already been used in a number of scientific publications by different researchers (Ahannach et al., 2021; Allonsius et al., 2019; De Boeck et al., 2017, 2019, 2020, 2021, 2022; Jörissen et al., 2021; Lebeer et al., 2022, 2023; Leguin 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) and has also been used by several undergraduate, graduate, and PhD students during their research projects (Leguin, 2022; I. M. Van Tendeloo, 2022), demonstrating its versatility in topics and user experience.

## Functionality

tidy\_tacos 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 “tidy\_tacos object” which is a list of three tables: 1) The counts: these are the counts of reads for each taxon (OTU/ASV/phylo type) in each sample. Each row represents such a read count. 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 tidy\_tacos 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 `ggplot` package of the tidyverse. Additionally, also functions to study the effect of sample variables on community variation are available (e.g. `perform_adonis`). The full overview of functions, code, a quick start tutorial, and more advanced tutorials are available on the [Github](https://github.com/LebeerLab/tidyacos) page, <https://github.com/LebeerLab/tidyacos>.

## 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).
- 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.
- De Boeck, I., Broek, M. F. van den, Allonsius, C. N., Spacova, I., Wittouck, S., Martens, K., Wuyts, S., Cauwenberghs, E., Jokicevic, K., Vandenhuevel, D., & others. (2020). Lactobacilli have a niche in the human nose. *Cell Reports*, 31(8).
- 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.
- 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.
- De Boeck, I., Wittouck, S., Martens, K., Spacova, I., Cauwenberghs, E., Allonsius, C. N., Jorissen, J., Wuyts, S., Van Beeck, W., Dillen, J., & others. (2021). The nasal mutualist *Dolosigranulum pigrum* AMBR11 supports homeostasis via multiple mechanisms. *Iscience*, 24(9).
- De Boeck, I., Wittouck, S., Wuyts, S., Oerlemans, E. F., Van den Broek, M. F., Vandenhuevel, 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.
- Jorissen, 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).

- 105 Case-control microbiome study of chronic otitis media with effusion in children points at  
106 streptococcus salivarius as a pathobiont-inhibiting species. *MSystems*, 6(2), 10–1128.
- 107 Lebeer, S., Ahannach, S., Gehrman, T., Wittouck, S., Eilers, T., Oerlemans, E., Condori,  
108 S., Dillen, J., Spacova, I., Vander Donck, L., & others. (2023). A citizen-science-enabled  
109 catalogue of the vaginal microbiome and associated factors. *Nature Microbiology*, 1–13.
- 110 Lebeer, S., Oerlemans, E. F., Claes, I., Henkens, T., Delanghe, L., Wuyts, S., Spacova, I.,  
111 Broek, M. F. van den, Tuyaerts, I., Wittouck, S., & others. (2022). Selective targeting of  
112 skin pathobionts and inflammation with topically applied lactobacilli. *Cell Reports Medicine*,  
113 3(2).
- 114 Legein, M. (2022). *Phyllosphere bacteria in greenhouses and their use in biocontrol* [PhD  
115 thesis]. University of Antwerp.
- 116 Legein, M., Smets, W., Wuyts, K., Bosmans, L., Samson, R., & Lebeer, S. (2022). The  
117 greenhouse phyllosphere microbiome and associations with introduced bumblebees and  
118 predatory mites. *Microbiology Spectrum*, 10(4), e01755–22.
- 119 Oerlemans, E. F., Bellen, G., Claes, I., Henkens, T., Allonsius, C. N., Wittouck, S., Broek,  
120 M. F. van den, Wuyts, S., Kiekens, F., Donders, G. G., & others. (2020). Impact of a  
121 lactobacilli-containing gel on vulvovaginal candidosis and the vaginal microbiome. *Scientific  
122 Reports*, 10(1), 7976.
- 123 Oerlemans, E., Ahannach, S., Wittouck, S., Dehay, E., De Boeck, I., Ballet, N., Rodriguez, B.,  
124 Tuyaerts, I., & Lebeer, S. (2022). Impacts of menstruation, community type, and an oral  
125 yeast probiotic on the vaginal microbiome. *Msphere*, 7(5), e00239–22.
- 126 Papini, G., Muys, M., Van Winckel, T., Meerburg, F. A., Van Beeck, W., Vermeir, P., &  
127 Vlaeminck, S. E. (2023). Boosting aerobic microbial protein productivity and quality on  
128 brewery wastewater: Impact of anaerobic acidification, high-rate process and biomass age.  
129 *Bioresource Technology*, 368, 128285.
- 130 Peng, L., Xie, Y., Van Beeck, W., Zhu, W., Van Tendeloo, M., Tytgat, T., Lebeer, S., &  
131 Vlaeminck, S. E. (2020). Return-sludge treatment with endogenous free nitrous acid  
132 limits nitrate production and N<sub>2</sub>O emission for mainstream partial nitrification/anammox.  
133 *Environmental Science & Technology*, 54(9), 5822–5831.
- 134 Smets, W., Chock, M. K., Walsh, C. M., Vanderburgh, C. Q., Kau, E., Lindow, S. E., Fierer,  
135 N., & Koskella, B. (2023). Leaf side determines the relative importance of dispersal versus  
136 host filtering in the phyllosphere microbiome. *Mbio*, e01111–23.
- 137 Smets, W., Leff, J. W., Bradford, M. A., McCulley, R. L., Lebeer, S., & Fierer, N. (2016).  
138 A method for simultaneous measurement of soil bacterial abundances and community  
139 composition via 16S rRNA gene sequencing. *Soil Biology and Biochemistry*, 96, 145–151.
- 140 Smets, W., Spada, L. M., Gandolfi, I., Wuyts, K., Legein, M., Muysshondt, B., Samson, R.,  
141 Franzetti, A., & Lebeer, S. (2022). Bacterial succession and community dynamics of the  
142 emerging leaf phyllosphere in spring. *Microbiology Spectrum*, 10(2), e02420–21.
- 143 Spacova, I., Ahannach, S., Breynaert, A., Erreygers, I., Wittouck, S., Bron, P. A., Van  
144 Beeck, W., Eilers, T., Alloul, A., Blansaer, N., & others. (2022). Spontaneous riboflavin-  
145 overproducing limosilactobacillus reuteri for biofortification of fermented foods. *Frontiers  
146 in Nutrition*, 9, 1231.
- 147 Van Tendeloo, I. M. (2022). *Resource-efficient nitrogen removal from sewage: Kinetic, physical  
148 and chemical tools for mainstream partial nitrification/anammox* [PhD thesis]. University of  
149 Antwerp.
- 150 Van Tendeloo, M., Bundervoet, B., Carlier, N., Van Beeck, W., Mollen, H., Lebeer, S.,  
151 Colsen, J., & Vlaeminck, S. E. (2021). Piloting carbon-lean nitrogen removal for energy-

- 152 autonomous sewage treatment. *Environmental Science: Water Research & Technology*,  
153 7(12), 2268–2281.
- 154 Van Tendeloo, M., Xie, Y., Van Beeck, W., Zhu, W., Lebeer, S., & Vlaeminck, S. E. (2021).  
155 Oxygen control and stressor treatments for complete and long-term suppression of nitrite-  
156 oxidizing bacteria in biofilm-based partial nitrification/anammox. *Bioresource Technology*,  
157 342, 125996.
- 158 Wang, X., Howe, S., Deng, F., & Zhao, J. (2021). Current applications of absolute bacterial  
159 quantification in microbiome studies and decision-making regarding different biological  
160 questions. *Microorganisms*, 9(9), 1797.
- 161 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Golemund,  
162 G., Hayes, A., Henry, L., Hester, J., & others. (2019). Welcome to the tidyverse. *Journal*  
163 *of Open Source Software*, 4(43), 1686.
- 164 Wickham, H., Çetinkaya-Rundel, M., & Golemund, G. (2023). *R for data science*. " O'Reilly  
165 Media, Inc."

DRAFT