# 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

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, 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. `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` 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. `tidytacos` 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; 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) and has also been used by several undergraduate, graduate, and PhD students during their research projects (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 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 `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

60 and taxa and their associated information (e.g. filter_samples, filter_taxa, select_samples,
61 select_taxa). Also, aggregating samples or taxa (aggregate_samples, aggregate_taxa) and
62 conversion of the taxon counts (add_rel_abundance/add_clr_abundance) are supported,
63 including a function to determine absolute abundances (add_absolute_abundance) based
64 on the spike-in method (Smets et al., 2016; Wang et al., 2021). There are quick functions
65 for visualizing community composition of samples using stacked bar plots (tacoplot_stack)
66 or visualizing a two-dimensional ordination of samples (tacoplot_ord) based on the ggplot
67 package of the tidyverse. Additionally, also functions to study the effect of sample variables
68 on community variation are available (e.g. perform_adonis). The full overview of functions,
69 code, a quick start tutorial, and more advanced tutorials are available on the Github page,
70 https://github.com/LebeerLab/tidytacos.

Acknowledgements

References

79 Ahannach, S., Delanghe, L., Spacova, I., Wittouck, S., Van Beeck, W., De Boeck, I., & Lebeer,
80   S. (2021). Microbial enrichment and storage for metagenomics of vaginal, skin, and saliva
81   samples. *Iscience*, *24*(11).

82 Allonsius, C. N., Van Beeck, W., De Boeck, I., Wittouck, S., & Lebeer, S. (2019). The
83   microbiome of the invertebrate model host galleria mellonella is dominated by enterococcus.
84   *Animal Microbiome*, *1*, 1–7.

85 De Boeck, I., Broek, M. F. van den, Allonsius, C. N., Spacova, I., Wittouck, S., Martens,
86   K., Wuyts, S., Cauwenberghs, E., Jokicevic, K., Vandenheuvel, D., & others. (2020).
87   Lactobacilli have a niche in the human nose. *Cell Reports*, *31*(8).

88 De Boeck, I., Cauwenberghs, E., Spacova, I., Gehrmann, T., Eilers, T., Delanghe, L., Wittouck,
89   S., Bron, P. A., Henkens, T., Gamgami, I., & others. (2022). Randomized, double-blind,
90   placebo-controlled trial of a throat spray with selected lactobacilli in COVID-19 outpatients.
91   *Microbiology Spectrum*, *10*(5), e01682–22.

92 De Boeck, I., Wittouck, S., Martens, K., Claes, J., Jorissen, M., Steelant, B., Broek, M. F. van
93   den, Seys, S. F., Hellings, P. W., Vanderveken, O. M., & others. (2019). Anterior nares
94   diversity and pathobionts represent sinus microbiome in chronic rhinosinusitis. *MSphere*,
95   *4*(6), 10–1128.

96 De Boeck, I., Wittouck, S., Martens, K., Spacova, I., Cauwenberghs, E., Allonsius, C. N.,
97   Jörissen, J., Wuyts, S., Van Beeck, W., Dillen, J., & others. (2021). The nasal mutualist
98   dolosigranulum pigrum AMBR11 supports homeostasis via multiple mechanisms. *Iscience*,
99   *24*(9).

100 De Boeck, I., Wittouck, S., Wuyts, S., Oerlemans, E. F., Van den Broek, M. F., Vandenheuvel,
101   D., Vanderveken, O., & Lebeer, S. (2017). Comparing the healthy nose and nasopharynx
102   microbiota reveals continuity as well as niche-specificity. *Frontiers in Microbiology*, *8*, 2372.

103 Jörissen, J., Broek, M. F. van den, De Boeck, I., Van Beeck, W., Wittouck, S., Boudewyns,
104   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.

Lebeer, S., Ahannach, S., Gehrmann, 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.

Lebeer, S., Oerlemans, E. F., Claes, I., Henkens, T., Delanghe, L., Wuyts, S., Spacova, I., Broek, M. F. van den, Tuyaerts, I., Wittouck, S., & others. (2022). Selective targeting of skin pathobionts and inflammation with topically applied lactobacilli. *Cell Reports Medicine*, *3*(2).

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.

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.

Oerlemans, E., Ahannach, S., Wittouck, S., Dehay, E., De Boeck, I., Ballet, N., Rodriguez, B., Tuyaerts, I., & Lebeer, S. (2022). Impacts of menstruation, community type, and an oral yeast probiotic on the vaginal microbiome. *Msphere*, *7*(5), e00239–22.

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.

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 nitritation/anammox. *Environmental Science & Technology*, *54*(9), 5822–5831.

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.

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.

Smets, W., Spada, L. M., Gandolfi, I., Wuyts, K., Legein, M., Muyshondt, 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.

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.

Van Tendeloo, I. M. (2022). *Resource-efficient nitrogen removal from sewage: Kinetic, physical and chemical tools for mainstream partial nitritation/anammox* [PhD thesis]. University of Antwerp.

Van Tendeloo, M., Bundervoet, B., Carlier, N., Van Beeck, W., Mollen, H., Lebeer, S., 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 nitritation/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., Grolemund,
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., & Grolemund, G. (2023). *R for data science*. " O'Reilly
165 Media, Inc.".