qdiv: A Python package for microbial ecology analysis using the Hill number framework

Downloaded from: https://research.chalmers.se, 2026-09-12 14:31 UTC

Citation for the original published paper (version of record):

Modin, O. (2026). qdiv: A Python package for microbial ecology analysis using the Hill number framework. The Journal of Open Source Software, 11(121). http://dx.doi.org/10.21105/joss.10089

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

The Journal of Open Source Software

qdiv: A Python package for microbial ecology analysis using the Hill number framework

Oskar Modin [1]

1 Chalmers University of Technology, Department of Architecture and Civil Engineering, Division of Water Environment Technology, Gothenburg, Sweden

DOI: 10.21105/joss.10089

Software

- Review
- Repository
- Archive

Editor: Richard Liu

Reviewers:

- @erinyoung
- @Ruslan-Kurmashev
- @Shaun-Jiang

Submitted: 05 January 2026
Published: 05 May 2026

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

Summary

Microorganisms are ubiquitous and play essential roles in biogeochemical cycles, human health, and engineered systems. Microbial ecology research seeks to understand the structure and function of microbial communities. Common approaches, such as amplicon sequencing and metagenomics, generate tabular data that track the relative abundances of hundreds or thousands of species or genes across time or space. To explore how communities assemble and respond to environmental factors, microbial diversity is quantified using various metrics, some incorporating phylogenetic and functional relationships.

The Hill number framework, also known as effective numbers, provides a unified and intuitive methodology for assessing diversity and changes in community composition. However, few tools systematically implement this framework across alpha and beta diversity, phylogenetic metrics, multivariate statistics, and null models. The Python package `qdiv` fills this gap by applying the Hill number framework to a broad range of ecological metrics, offering a streamlined and versatile tool for investigating patterns of community assembly.

Statement of need

`qdiv` is a Python package designed for microbial ecology. It was developed to facilitate the management and analysis of data generated from amplicon sequencing and metagenomics within a Python programming environment. Specifically, `qdiv` implements the Hill number framework (Chao et al., 2014) for assessing diversity across a broad range of ecological metrics and analyses. This framework enables systematic evaluation of how relative abundances influence diversity measures.

Traditionally, Hill numbers (Hill, 1973) are applied to quantify alpha diversity, which is the diversity of taxa within a single community. However, the framework can also be applied to beta diversity (Jost, 2007), which measures differences in composition between communities, and extended to metrics that incorporate phylogenetic and functional relationships among taxa (Chiu et al., 2014; Chiu & Chao, 2014).

Many ecological statistical tests and null models, such as Raup-Crick (Raup & Crick, 1979), the Net Relatedness Index, and the Nearest-Taxon Index (Webb et al., 2002), have used presence-absence data or relied on a single relative abundance-based index (e.g. Bray-Curtis). `qdiv` fills a critical gap by enabling the implementation of the Hill number framework for analyzing diversity and microbial community assembly with a range of statistical methods, all within Python.

Statement of the field

A variety of software tools support microbial community analysis, particularly within the R programming environment. Packages such as vegan (Oksanen et al., 2026), phyloseq (McMurdie & Holmes, 2013), microbiome (Lahti et al., 2017), and ampvis2 (Andersen et al., 2018) provide comprehensive workflows for diversity calculations, ordination, and phylogenetic analysis, but they primarily rely on classical indices rather than a unified diversity framework. Several R packages implement the Hill number framework, including hilldiv (Alberdi & Gilbert, 2019), hillR (Li, 2018), vegetarian (Charney, 2019), and iNEXT (Hsieh et al., 2024). However, to our knowledge there is no Python package that implements the Hill number framework in a unified way for diversity metrics and statistical workflows. The `qdiv` package fills this gap by offering a systematic Python implementation of Hill-based alpha and beta diversity, phylogenetic diversity, and associated statistical methods.

Other commonly used tools include QIIME (Bolyen et al., 2019), an end-to-end microbiome analysis platform typically used through a command-line interface, which limits flexible, script-based workflows in Python. The Python library scikit-bio (Aton et al., 2025) provides data structures and algorithms for omics analyses, including sequence handling, phylogenetic tree operations, and ecological statistics, but it does not center around, or generalize, the Hill number framework.

Software design

`qdiv` has a modular architecture centered around the core MicrobiomeData class. This class serves as a container for all essential components, including relative abundance tables, taxonomic annotations, sequences, phylogenetic trees, and associated metadata. The motivation for this design is to provide users with a single, coherent object that keeps data, metadata, and analysis context together, reducing the risk of mismatched inputs.

A key design decision concerns how much functionality should reside inside the MicrobiomeData class versus in separate subpackages. Methods related to data wrangling, such as subsetting, merging, and saving or loading data, are implemented as class methods because they directly modify or manage the internal state of the dataset. Keeping these operations inside the class provides users with a simple interface.

In contrast, more specialized analyses such as diversity calculations, statistical tests, and visualizations, are implemented in dedicated subpackages that operate on MicrobiomeData instances but do not alter them. This separation follows the principle of keeping data management and analytical logic independent.

Another important trade-off relates to how much general-purpose data manipulation the package should include. The primary goal of qdiv is to implement the Hill number framework and extend it to statistical methods for which it has not previously been applied, such as null models. For convenience, qdiv includes a limited set of helper functions for tasks such as simple phylogenetic tree manipulations. However, for advanced tree operations, users are encouraged to rely on specialized libraries such as ete3 (Huerta-Cepas et al., 2016).

The `qdiv` package leverages the widely used Python libraries pandas, numpy, and matplotlib, to ensure efficiency, interoperability, and familiarity for users already working in the Python data-science ecosystem. A deliberate effort was made to keep dependencies to a minimum, to reduce installation friction and decrease the risk of version conflicts.

Research impact statement

The `qdiv` package was first introduced by Modin et al. (2020) to enable Hill-based beta diversity calculations and extend the Raup-Crick null model to such metrics. Since then, it

has been adopted by other researchers, e.g., (Alberdi et al., 2021; Nikolova et al., 2021), and used by our own research group, e.g., (Abadikhah et al., 2024). It has been downloaded 835 times from PyPI (excluding mirrors) from 2025-09-13 to 2026-03-12. Version 4 introduces the MicrobiomeData class for streamlined data management, along with new statistical and visualization capabilities.

AI usage disclosure

Microsoft 365 Copilot powered by the GPT-5 chat model was used during the development of `qdiv` version 4.0.0 and higher. Copilot was used to revise existing functions, draft a new function for distance-based redundancy analysis, and revise sections of this manuscript. All design decisions, code implementations, and the content of the manuscript remain the sole responsibility of the author.

Acknowledgements

Research projects that have supported the continued development of `qdiv` were funded by the Swedish Research Council VR (2023-03908), the Swedish Research Council FORMAS (2018-00622, 2024-01814), and the NovoNordisk Foundation (NNF24OC0093678).

References

Abadikhah, M., Persson, F., Farewell, A., Wilen, B.-M., & Modin, O. (2024). Viral diversity and host associations in microbial electrolysis cells. *ISME Communications*, *4*. https://doi.org/10.1093/ismeco/ycae143

Alberdi, A., Bideguren, G. M., & Aizpurua, O. (2021). Diversity and compositional changes in the gut microbiota of wild and captive vertebrates: A meta-analysis. *Scientific Reports*, *11*. https://doi.org/10.1038/s41598-021-02015-6

Alberdi, A., & Gilbert, M. T. P. (2019). Hilldiv: An r package for the integral analysis of diversity based on hill numbers. *bioRxiv*. https://doi.org/10.1101/545665

Andersen, K. S., Kirkegaard, R. H., Karst, S. M., & Albertsen, M. (2018). ampvis2: An r package to analyse and visualise 16S rRNA amplicon data. *bioRxiv*. https://doi.org/10.1101/299537

Aton, M., McDonald, D., Cañardo Alastuey, J., Azom, R., Batra, P., Bezshapkin, V., Bolyen, E., Cagle, A., Caporaso, J. G., Debelius, J. W., Gorlick, K., Hamsanipally, N., Hunger, L., Keluskar, A., Liao, D., Lu, Y. Y., Navas-Molina, J. A., Pitman, A., Rideout, J. R., … Morton, J. T. (2025). Scikit-bio: A fundamental python library for biological omic data analysis. *Nature Methods*. https://doi.org/10.1038/s41592-025-02981-z

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*, 852–857. https://doi.org/10.1038/s41587-019-0209-9

Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, *45*. https://doi.org/10.1146/annurev-ecolsys-120213-091540

Charney, N. (2019). *Vegetarian (version 1.2): Jost diversity measures for community data*.

https://www.rdocumentation.org/packages/vegetarian/versions/1.2

Chiu, C.-H., & Chao, A. (2014). Distance-based functional diversity measures and their decomposition: A framework based on hill numbers. *PLoS ONE*, *9*(7), e100014. https://doi.org/10.1371/journal.pone.0100014

Chiu, C.-H., Jost, L., & Chao, A. (2014). Phylogenetic diversity measures and their decomposition: A framework based on hill numbers. *Ecological Monographs*, *84*(1), 21–44. https://doi.org/10.1890/12-0960.1

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, *54*(2), 427–432. https://doi.org/10.2307/1934352

Hsieh, T. C., Ma, K. H., & Chao, A. (2024). *iNEXT: Interpolation and extrapolation for species diversity*. https://doi.org/10.32614/CRAN.package.iNEXT

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*, *33*(6), 1635–1638. https://doi.org/10.1093/molbev/msw046

Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, *88*(10), 2427–2439. https://doi.org/10.1890/06-1736.1

Lahti, L., Shetty, S., & others. (2017). *Tools for microbiome analysis in r*. Bioconductor. https://doi.org/10.18129/B9.bioc.microbiome

Li, D. (2018). hillR: Taxonomic, functional, and phylogenetic diversity and similarity through hill numbers. *Journal of Open Source Software*, *3*(31), 1041. https://doi.org/10.21105/joss.01041

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

Modin, O., Liebana, R., Saheb-Alam, S., Wilen, B.-M., Suarez, C., Hermansson, M., & Persson, F. (2020). Hill-based dissimilarity indices and null models for analysis of microbial community assembly. *Microbiome*, *8*. https://doi.org/10.1186/s40168-020-00909-7

Nikolova, C., Ijaz, U. Z., & Gutierrez, T. (2021). Exploration of marine bacterioplankton community assembly mechanisms during chemical dispersant and surfactant-assisted oil biodegradation. *Ecology and Evolution*, *11*. https://doi.org/10.1002/ece3.8091

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., … Weedon, J. (2026). *Vegan: Community ecology package*. https://doi.org/10.32614/CRAN.package.vegan

Raup, D. M., & Crick, R. E. (1979). Measurement of faunal similarity in paleontology. *Journal of Paleontology*, *53*. https://www.jstor.org/stable/1304099

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology, Evolution, and Systematics*, *33*. https://doi.org/10.1146/annurev.ecolsys.33.010802.150448