ecodive: Parallel and Memory-Efficient R Package for Ecological Diversity Analysis

15 August 2025



ecodive package logo

# Summary

In ecology, diversity measures the composition of communities and is the first step toward understanding the role communities play within their environment. The most common measures of diversity in microbiome research are alpha-diversity and beta-diversity. While alpha-diversity aims to describe the richness and evenness of features within a single sample, beta-diversity assesses the dissimilarities between two or more communities. Diversity calculations may include the number of species or other features present, relative abundances, evolutionary relationships, or a combination thereof.

Applying these metrics to large collections of communities, such as thousands of gut microbiome samples, offers insights into predicting or diagnosing disease states through ecological “fingerprints,” a computationally intensive task.

# Statement of Need

Processing diversity metrics for thousands of communities is computationally intensive. The speed and memory footprint of these calculations often become a bottleneck for analysis, limiting the scope and scale of research studies. This is especially true for Faith’s PD (Faith 1992) and UniFrac (C. Lozupone and Knight 2005), which require complex integration of species counts with evolutionary distances by traversing phylogenetic trees. A faster and more memory-efficient implementation enables researchers to analyze a greater number of samples, leading to more robust and comprehensive insights. The ecodive R package addresses these challenges by employing a compiled C library with pthreads parallelization to efficiently compute these metrics, offering significant performance gains.

## Related Works

There are currently nine other R packages that can calculate alpha and beta diversity metrics: abdiv (Bittinger 2020), adiv (Pavoine 2020), ampvis2 (Andersen et al. 2018), entropart (Marcon and Herault 2015), GUniFrac (Chen et al. 2023), phyloregion (Daru, Karunarathne, and Schliep 2020), phyloseq (McMurdie and Holmes 2013), picante (Kembel et al. 2010), and vegan (Oksanen et al. 2025). While several R packages offer diversity metric calculations, ecodive distinguishes itself by providing an implementation that is both significantly faster and more memory efficient. This superior performance, across various diversity metrics, is rigorously demonstrated in Figures 1-3 through comprehensive benchmarking.

The bench R package (Hester and Vaughan 2025) was used to compare abdiv, adiv, ampvis2, entropart, ecodive, GUniFrac, phyloregion, phyloseq, picante, and vegan. The benchmarking runs are detailed in the benchmark vignette, which is available from within R with vignette('benchmark') and online at <https://cmmr.github.io/ecodive/articles/benchmark.html>. Note that not all R packages offer all diversity metrics.

Figure 1: UniFrac benchmarks. ecodive demonstrates substantial performance gains for UniFrac, being 2 to 3,900x faster and using 50 - 32,000x less memory, which helps overcome computational bottlenecks in large-scale analyses.

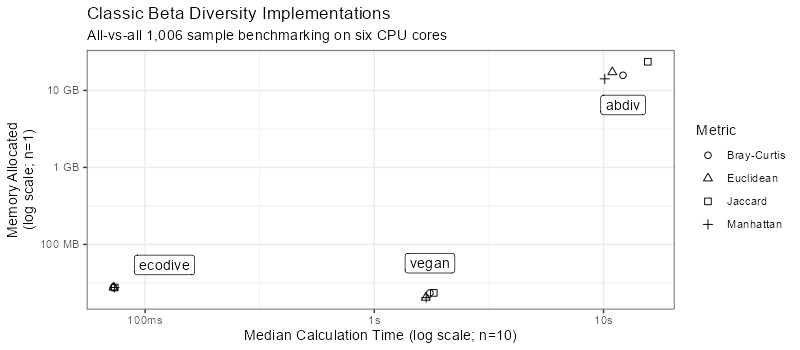


Figure 2: Classic beta diversity benchmarks. ecodive is 23 to 210x faster and uses 1 to 850x less memory, enabling more efficient analysis of community dissimilarities.

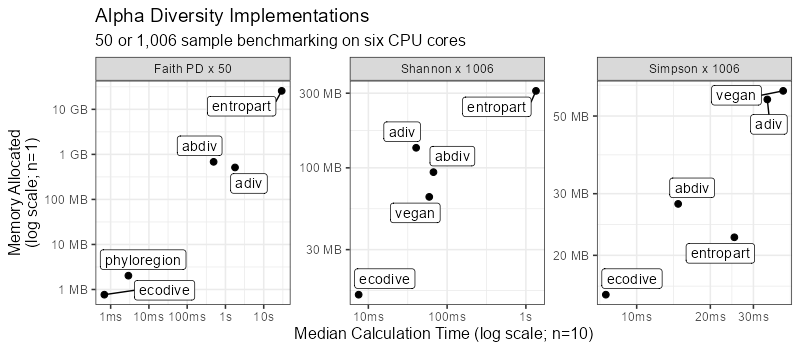


Figure 3: Alpha diversity benchmarks. ecodive is 2 to 43,000x faster and uses 1 to 33,000x less memory, significantly accelerating the analysis of diversity within single samples.

# Algorithms

The full list of alpha and beta diversity metrics currently implemented by ecodive is provided below. This set of metrics is subject to expansion as new functionalities are developed. Refer to ecodive’s official documentation at <https://cmmr.github.io/ecodive/reference/index.html> for the most up-to-date list and detailed descriptions.

### Classic Alpha Diversity

* Shannon Index (Shannon 1948)
* Simpson Index (Simpson 1949; Gini 1912)
* Inverse Simpson Index (Simpson 1949)
* Chao1 (Chao 1984)

### Phylogenetic Alpha Diversity

* Faith’s Phylogenetic Diversity (Faith 1992)

### Classic Beta Diversity

* Bray-Curtis Index (Bray and Curtis 1957; Sorenson 1948)
* Canberra (Lance and Williams 1967)
* Euclidean (Gower and Legendre 1986; Legendre and Caceres 2013)
* Gower (Gower 1971; Gower and Legendre 1986)
* Jaccard (Jaccard 1908)
* Kulczynski (Kulczynski 1927)
* Manhattan (Kaufman and Rousseeuw 1990)

### Phylogenetic Beta Diversity

* Unweighted UniFrac (C. Lozupone and Knight 2005)
* Weighted UniFrac (C. A. Lozupone et al. 2007)
* Normalized Weighted UniFrac (C. A. Lozupone et al. 2007)
* Generalized UniFrac (Chen et al. 2012)
* Variance Adjusted Weighted UniFrac (Chang, Luan, and Sun 2011)

# Usage

Users can easily compute alpha and beta diversity metrics using ecodive. For example, to calculate weighted UniFrac distances with a phyloseq object:

library(phyloseq)  
data(esophagus)  
  
ecodive::weighted\_unifrac(esophagus)  
#> B C  
#> C 0.1050480   
#> D 0.1401124 0.1422409

# Acknowledgements

This study was supported by NIH/NIAD (Grant number U19 AI44297), and Baylor College of Medicine and Alkek Foundation Seed.

The authors would like to thank Gemini for its assistance in refining this manuscript.

# References

Andersen, Kasper S., Rasmus H. Kirkegaard, Søren M. Karst, and Mads Albertsen. 2018. “Ampvis2: An r Package to Analyse and Visualise 16S rRNA Amplicon Data.” *bioRxiv*. <https://doi.org/10.1101/299537>.

Bittinger, Kyle. 2020. *Abdiv: Alpha and Beta Diversity Measures*. <https://doi.org/10.32614/CRAN.package.abdiv>.

Bray, J. Roger, and John T. Curtis. 1957. “An Ordination of the Upland Forest Communities of Southern Wisconsin.” *Ecological Monographs* 27 (4): 325–49. <https://doi.org/10.2307/1942268>.

Chang, Qin, Yihui Luan, and Fengzhu Sun. 2011. “Variance Adjusted Weighted UniFrac: A Powerful Beta Diversity Measure for Comparing Communities Based on Phylogeny.” *BMC Bioinformatics* 12 (1). <https://doi.org/10.1186/1471-2105-12-118>.

Chao, Anne. 1984. “Non-Parametric Estimation of the Number of Classes in a Population.” *Scandinavian Journal of Statistics* 11: 265–70. <https://doi.org/10.2307/4616294>.

Chen, Jun, Kyle Bittinger, Emily S. Charlson, Christian Hoffmann, James Lewis, Gary D. Wu, Ronald G. Collman, Frederic D. Bushman, and Hongzhe Li. 2012. “Associating Microbiome Composition with Environmental Covariates Using Generalized UniFrac Distances.” *Bioinformatics* 28 (16): 2106–13. <https://doi.org/10.1093/bioinformatics/bts342>.

Chen, Jun, Xianyang Zhang, Lu Yang, and Lujun Zhang. 2023. *GUniFrac: Generalized UniFrac Distances, Distance-Based Multivariate Methods and Feature-Based Univariate Methods for Microbiome Data Analysis*. <https://doi.org/10.32614/CRAN.package.GUniFrac>.

Daru, Barnabas H., Piyal Karunarathne, and Klaus Schliep. 2020. “Phyloregion: R Package for Biogeographic Regionalization and Macroecology.” *Methods in Ecology and Evolution* 11: 1483–91. <https://doi.org/10.1111/2041-210X.13478>.

Faith, Daniel P. 1992. “Conservation Evaluation and Phylogenetic Diversity.” *Biological Conservation* 61: 1–10. <https://doi.org/10.1016/0006-3207(92)91201-3>.

Gini, Corrado. 1912. *Variabilita e Mutabilita*. Bologna: Tipografia di Paolo Cuppini.

Gower, JC. 1971. “A General Coefficient of Similarity and Some of Its Properties.” *Biometrics* 27 (4): 857–71. <https://doi.org/10.2307/2528823>.

Gower, JC, and P Legendre. 1986. “Metric and Euclidean Properties of Dissimilarity Coefficients.” *Journal of Classification* 3: 5–48. <https://doi.org/10.1007/BF01896809>.

Hester, Jim, and Davis Vaughan. 2025. *Bench: High Precision Timing of r Expressions*. <https://doi.org/10.32614/CRAN.package.bench>.

Jaccard, Paul. 1908. “Nouvelles Recherches Sur La Distribution Florale.” *Bulletin de La Societe Vaudoise Des Sciences Naturelles* 44 (163): 223–70. <https://doi.org/10.5169/seals-268384>.

Kaufman, Leonard, and Peter J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons. <https://doi.org/10.1002/9780470316801>.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. “Picante: R Tools for Integrating Phylogenies and Ecology.” *Bioinformatics* 26: 1463–64.

Kulczynski, S. 1927. “Die Pflanzenassoziationen Der Pieninen.” *Bulletin International de l’Académie Polonaise Des Sciences Et Des Lettres, Classe Des Sciences Mathématiques Et Naturelles, Série B: Sciences Naturelles*, 57–203.

Lance, Godfrey N, and William T Williams. 1967. “A General Theory of Classificatory Sorting Strategies II. Clustering Systems.” *The Computer Journal* 10 (3). <https://doi.org/10.1093/comjnl/10.3.271>.

Legendre, P, and M Caceres. 2013. “Beta Diversity as the Variance of Community Data: Dissimilarity Coefficients and Partitioning.” *Ecology Letters* 16 (8). <https://doi.org/10.1111/ele.12141>.

Lozupone, Catherine A., Micah Hamady, Scott T. Kelley, and Rob Knight. 2007. “Quantitative and Qualitative Beta Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities.” *Applied and Environmental Microbiology* 73 (5): 1576–85. <https://doi.org/10.1128/aem.01996-06>.

Lozupone, Catherine, and Rob Knight. 2005. “UniFrac: A New Phylogenetic Method for Comparing Microbial Communities.” *Applied and Environmental Microbiology* 71 (12): 8228–35. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>.

Marcon, Eric, and Bruno Herault. 2015. “Entropart: An r Package to Measure and Partition Diversity.” *Journal of Statistical Software* 67 (8): 1–26. <https://doi.org/10.18637/jss.v067.i08>.

McMurdie, Paul J., and Susan Holmes. 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>.

Oksanen, Jari, Gavin L. Simpson, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O’Hara, et al. 2025. *Vegan: Community Ecology Package*. <https://doi.org/10.32614/CRAN.package.vegan>.

Pavoine, Sandrine. 2020. “Adiv: An r Package to Analyse Biodiversity in Ecology.” *Methods in Ecology and Evolution* 11: 1106–12. <https://doi.org/10.1111/2041-210X.13430>.

Shannon, Claude E. 1948. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27 (3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.

Simpson, Edward H. 1949. “Measurement of Diversity.” *Nature* 163 (4148): 688–88. <https://doi.org/10.1038/163688a0>.

Sorenson, Thorvald. 1948. “A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content.” *Kongelige Danske Videnskabernes Selskab* 5: 1–34.