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

18 August 2025

# Summary

Characterizing the composition of biological communities is a fundamental task in ecology, but the calculations involved can be computationally prohibitive. ecodive is an R package that addresses this challenge by providing a highly optimized implementation of common ecological diversity metrics, including alpha-diversity (within-sample richness and evenness) and beta-diversity (between-sample dissimilarity). These metrics can incorporate species counts, relative abundances, and evolutionary relationships, providing a multi-faceted view of ecological structure. By leveraging a compiled C library with pthreads for parallelization, ecodive delivers substantial performance gains in both speed and memory usage, enabling researchers to analyze larger datasets more efficiently.

# Statement of Need

The analysis of ecological diversity in large-scale studies is often hampered by the computational demands of calculating metrics across thousands of communities, a common requirement in modern microbiome research. This is particularly true for phylogenetic metrics like Faith’s PD (Faith 1992) and the UniFrac distance family (C. Lozupone and Knight 2005), which integrate species abundance with evolutionary data from phylogenetic trees. The resulting high demand on processing time and memory can limit the scope and scale of scientific inquiry.

ecodive overcomes these limitations by offering a significantly faster and more memory-efficient solution. This allows researchers to analyze more samples, explore more complex questions, and obtain more robust insights from their data. By providing a high-performance, parallelized engine for these calculations, ecodive empowers researchers to push the boundaries of large-scale ecological analysis.

## Comparison to Existing Packages

While numerous R packages can calculate diversity metrics, our comparison focuses on those that provide their own implementations: abdiv (Bittinger 2020), adiv (Pavoine 2020), ampvis2 (Andersen et al. 2018), ecodist (Goslee and Urban 2007), 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). For every metric implemented, ecodive offers superior performance, setting it apart from these packages. Furthermore, ecodive has zero external R dependencies. This makes it a lightweight, stable, and secure computational backend, minimizing installation conflicts and simplifying long-term maintenance for developers who build upon it.

Comprehensive benchmarks, conducted using the bench package (Hester and Vaughan 2025), demonstrate these advantages across a range of metrics (Figures 1-3). The benchmark suite also confirms that ecodive produces numerically identical output to other packages for every metric. The sole exception was the Jaccard metric, for which abdiv’s output differed from the concordant results of ecodive, vegan, and ecodist. The complete benchmark code and results are available in the package vignette (vignette('benchmark')) and online.

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 6 to 2,300x faster and uses 1 to 1,800x 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.

# Implemented Metrics

ecodive provides a comprehensive suite of alpha and beta diversity metrics. The current implementation includes:

### Alpha Diversity

* Classic: Shannon Index (Shannon 1948), Simpson Index (Simpson 1949; Gini 1912), Inverse Simpson Index (Simpson 1949), and Chao1 (Chao 1984).
* Phylogenetic: Faith’s Phylogenetic Diversity (Faith 1992).

### Beta Diversity

* Classic: Bray-Curtis (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), and Manhattan (Kaufman and Rousseeuw 1990).
* Phylogenetic: 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), and Variance Adjusted Weighted UniFrac (Chang, Luan, and Sun 2011).

For the most up-to-date list and detailed descriptions, please refer to the official ecodive documentation at <https://cmmr.github.io/ecodive/reference/index.html>.

# Example Usage

ecodive is designed for ease of use and integrates seamlessly with existing bioinformatics workflows, such as those using phyloseq objects. For example, calculating weighted UniFrac distances is straightforward:

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 AI144297), and Baylor College of Medicine and Alkek Foundation Seed. The authors also acknowledge the use of Google’s Gemini for 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.

Goslee, Sarah C., and Dean L. Urban. 2007. “The Ecodist Package for Dissimilarity-Based Analysis of Ecological Data.” *Journal of Statistical Software* 22: 1–19. <https://doi.org/10.18637/jss.v022.i07>.

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. <https://doi.org/10.1093/bioinformatics/btq166>.

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.