rbiom: An R package for Analyzing Ecological Data

13 August 2025

Figure 1: Rbiom package logo

Summary

rbiom is a software toolkit for the R programming language that simplifies the analysis of complex ecological data. Its main purpose is to help researchers uncover hidden trends and patterns by connecting what they observe in a biological sample (like the types and numbers of bacteria) with information they know about that sample (like where it was collected or the health of the individual). rbiom makes this process easy by focusing on creating clear, publication-quality figures that include statistical test results right on the chart. This allows scientists to quickly visualize and understand the important relationships within their data.

Statement of Need

A core challenge in microbiome analysis is identifying meaningful associations between a community's composition and its associated metadata. While established packages like phyloseq exist for handling the large and complex datasets often stored in Biological Observation Matrix (BIOM) files, there is a persistent need for a tool that streamlines the entire workflow from data ingestion to the creation of publication-quality figures with integrated statistical results. rbiom addresses this need by providing a unified interface that connects the analysis of taxonomic abundance and diversity with powerful statistical methods and visualizations. Its design prioritizes speed and reproducibility, making it particularly suitable for both routine data exploration and rigorous statistical reporting. Functionality

The rbiom package offers the following key functionalities:

- Data Management and Manipulation: rbiom seamlessly imports various data formats, including BIOM files, QIIME2 and mothur outputs, and objects from other popular R packages using the phyloseq or SummarizedExperiment classes. The package includes a robust set of tools for rarefaction, filtering, and summarization, enabling users to prepare their data for downstream analysis.

- Diversity and Composition Analysis: The package focuses on three key community features: **alpha diversity**, beta diversity, and taxa abundance. For each, it can compute statistics against categorical or continuous metadata using appropriate methods such as Kruskal-Wallis, Mann-Whitney, PERMANOVA, and estimated marginal means.
- Statistical Visualization: A core feature of rbiom is its ability to directly overlay statistical test results onto ggplot2-based figures. Functions like adiv_boxplot() and stats_corrplot() generate visualizations with automated annotations for p-values, trend lines, confidence intervals, and methodology. Customization is supported through parameters like p.label to display only significant results, ensuring clean, publication-ready graphics.
- High-Level Plotting: The package provides a diverse set of highly customizable plot types, including boxplots, heatmaps, stacked bar charts, and ordination plots. Users have extensive control over aesthetics, with options for specifying metadata variables for the x-axis, statistical groups, and plot faceting.

Related Works

The ecosystem of R packages for microbiome and ecological data analysis is rich and active. The most notable and widely used package is phyloseq, which introduced a foundational data structure (the phyloseq object) that has been widely adopted. phyloseq provides extensive functionality for data import, normalization, and classic ecological analyses. rbiom complements this ecosystem by offering an alternative, streamlined workflow that prioritizes speed and reproducibility. A key differentiator is rbiom's tight integration of statistical testing with visualization, allowing users to generate figures with embedded statistical results, such as p-values and trend lines, directly from a single function call.

A recent review by Wen et al. (2023) highlighted the vast number of R packages available for microbiome analysis. However, few make a concerted effort to display statistical results directly on generated figures. microeco, a comprehensive tool known for its microtable class and extensive integration with other packages, is one of the exceptions. While microeco offers a broad range of capabilities, rbiom provides a more focused approach, simplifying common analyses and offering a completely automated method for overlaying statistics on figures. Uniquely, rbiom also provides the underlying R code for its generated figures and statistics. This transparency is crucial for a deeper level of reproducible research and allows users to easily customize or extend the analysis. This specialization makes rbiom an excellent choice for researchers who need to quickly and reliably generate publication-ready figures with clear statistical evidence.

Acknowledgements

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

References