



Strollur: an R package for working with amplicon sequence data in R

Sarah L. Westcott,¹ Gregory Johnson, Jr.,¹ Patrick D. Schloss¹

AUTHOR AFFILIATION See affiliation list on p. 2.

ABSTRACT Microbiologists are increasingly relying on R to analyze and visualize amplicon sequence data from microbiome studies. We present strollur, which is an open-source package that facilitates the import, export, and handling of amplicon sequence data within the R programming language.

KEYWORDS microbiome, bioinformatics, R, amplicon sequence data, 16S rRNA gene sequences

Amplicon sequencing projects, including 16S rRNA gene sequence collections, are widely used by microbiologists to analyze diverse microbiomes. These data sets are often large and complicated, making it difficult for researchers to keep their data organized and record the provenance of derivative data products. A long-standing difficulty for many has been the ability to move data between different analytical tools. This often requires reformatting data, which is challenging for researchers with limited programming skills. Several efforts have been made to facilitate the management of amplicon sequencing data and their associated metadata, including the biom file format (1), mothur output files (2), the qiime2 archive (3), and phyloseq R objects (4). Each of these formats and tools has strengths but also limitations, which make their use difficult for researchers new to microbiome data analysis or who are more accustomed to using R's tidyverse collection of packages.

To overcome these challenges, we developed the strollur R package. Our goal was to develop a framework to store amplicon sequence data that could be integrated within other packages without the end user necessarily knowing they were using strollur. This is meant to be analogous to how many users of the tidyverse use the tibble package without knowing they are using the package. By analogy, strollur strives to provide a consistent interface to amplicon sequence data, which also integrates with analysis and visualization tools from the tidyverse and other tools used by microbiome researchers.

Strollur's data structures allow users to keep track of sequence data, abundance data, classification data, phylogenetic trees, operational taxonomic unit (OTU), and amplicon sequence variant assignments, reports from different processing steps, and sample metadata (Fig. 1). Furthermore, the data structures can accommodate information describing the commands that generated the data products and information about the databases that were used. The package includes user-facing functions for getting, assigning, and summarizing various forms of data. The package also provides functions to read data from or write data to other popular formats, including mothur (2), QIIME 2 (3), DADA2 (5), phyloseq (4), and BIOM (1). Strollur can be used on any operating system with no additional dependencies.

Strollur also includes functions designed for developers to integrate strollur objects into their own R packages. These developer-facing functions provide additional access to the back-end data. The ability to import and export data as a strollur object has already been integrated into the clustur (6) and phylotypyr (7) packages. The package leverages

Editor Leighton Pritchard, University of Strathclyde, Glasgow, United Kingdom

Address correspondence to Patrick D. Schloss, pschloss@umich.edu.

The authors declare no conflict of interest.

Received 6 July 2026

Accepted 9 September 2026

Published 29 September 2026

Copyright © 2026 Westcott et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

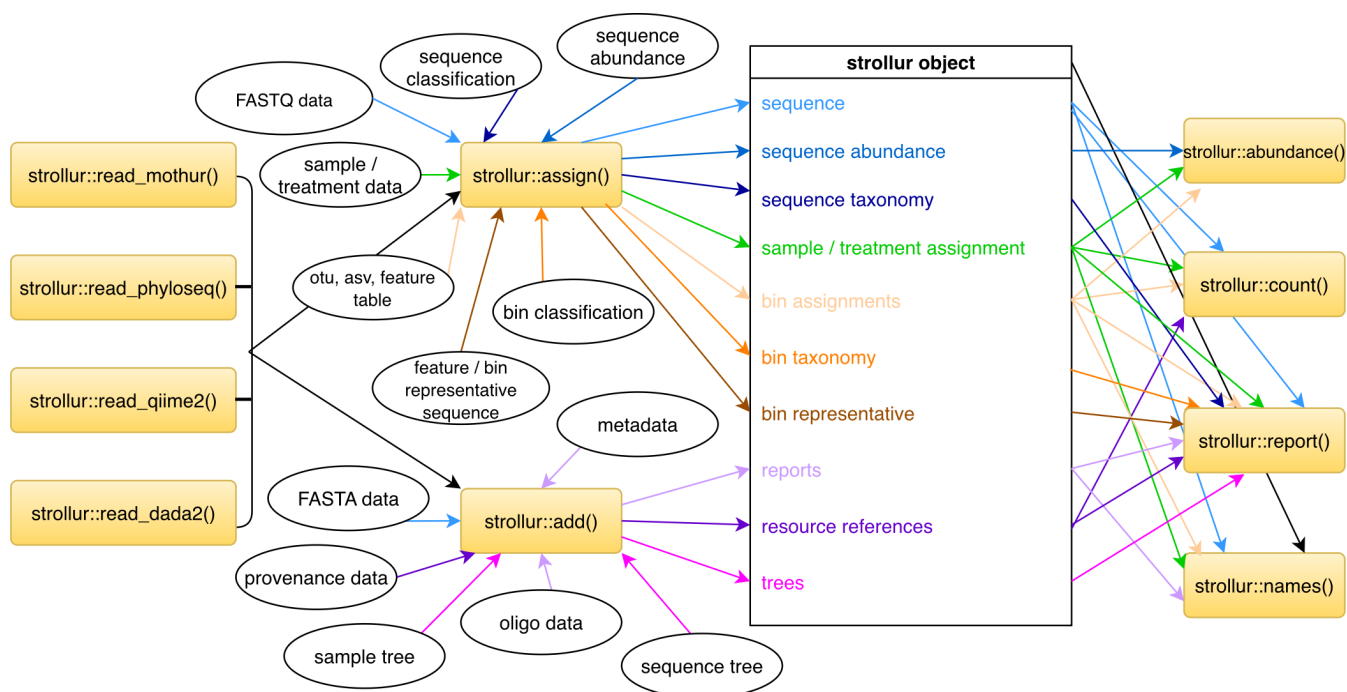


FIG 1 The strollur R package facilitates the import of microbiome data generated by popular software packages into R and allows users to export the data in a variety of formats.

the R6 object-oriented programming framework and the Rcpp package to accelerate data processing. By using these features of R, strollur minimizes the need to create deep copies of large data objects.

AUTHOR AFFILIATION

¹Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan, USA

AUTHOR ORCID*s*

Gregory Johnson, Jr. <http://orcid.org/0009-0008-3890-0297>

Patrick D. Schloss <http://orcid.org/0000-0002-6935-4275>

AUTHOR CONTRIBUTIONS

Sarah L. Westcott, Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – review and editing | Gregory Johnson, Jr., Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – review and editing | Patrick D. Schloss, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Strollur can be accessed through the Comprehensive R Archive Network (CRAN; DOI: [10.32614/CRAN.package.strollur](https://doi.org/10.32614/CRAN.package.strollur)) and developmental versions are available through the project's GitHub website (<https://github.com/mothur/strollur>). The package's GitHub repository includes the package's source code, the website's source code, automated tests written using the testthat R package, and an issue tracker where users can post questions, bug reports, and suggestions for future features. A pkgdown version of

the documentation including multiple articles describing its use is hosted at (<https://mothur.org/strollur>). The strollur package is available under the GNU General Public License (v3).

REFERENCES

1. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, Knight R, Caporaso JG. 2012. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *Gigascience* 1:7. <https://doi.org/10.1186/2047-217X-1-7>
2. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>
3. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852–857. <https://doi.org/10.1038/s41587-019-0209-9>
4. McMurdie PJ, Holmes S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>
5. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>
6. Johnson G, Westcott SL, Schloss PD. 2025. Clustur: an R package for clustering features using sparse distance matrices. *Microbiol Resour Announc* 14:e01238-24. <https://doi.org/10.1128/mra.01238-24>
7. Schloss PD. 2025. Phylotypr: an R package for classifying DNA sequences. *Microbiol Resour Announc* 14:e01144-24. <https://doi.org/10.1128/mra.01144-24>