

Upgrading microbiome research in R/Bioc: MicrobiomeExperiment

Leo Lahti @antagomir
University of Turku, Finland

with
Felix G.M. Ernst, Sudarshan A. Shetty,
Tuomas Borman, Ruizhu Huang,
Domenick J. Braccia, Héctor Corrada Bravo



phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data

Paul J. McMurdie, Susan Holmes 

RESEARCH ARTICLE

REVISED Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses [version 2; referees: 3 approved]

Ben J. Callahan¹, Kris Sankaran¹, Julia A. Fukuyama¹, Paul J. McMurdie²,  Susan P. Holmes¹

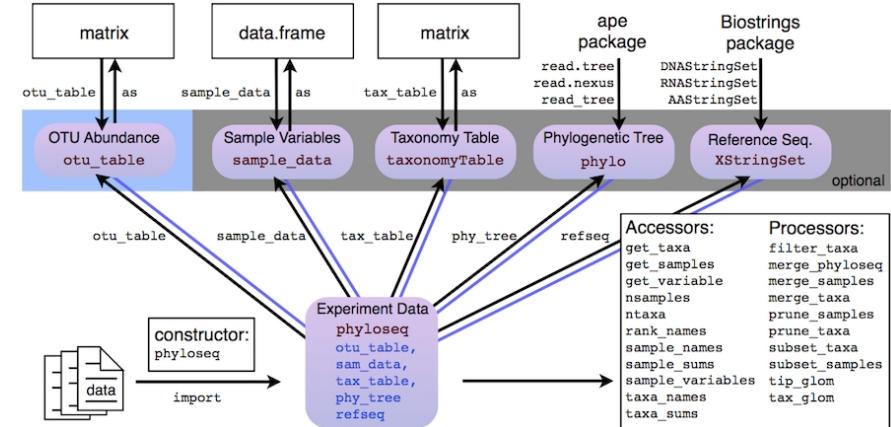


Table 1 Overview of contemporary online resources for microbiome data science in R. The indicated groupings are approximations as many packages span over multiple categories

From: [Microbiome data science](#)

Pre-processing of raw reads to ASVs/OTUs	BioC: dada2 (Callahan et al. 2016b)
Taxonomic classification and analysis	BioC: rRDP (Hahsler and Nagar 2014), DECIPHER (IDTAXA algorithm) (Murali et al. 2018); CRAN: taxize (Chamberlain et al. 2014), microclass (Liland et al. 2017)
General data manipulation and visualisation	BioC: Phyloseq (McMurdie and Holmes 2013), microbiome (Lahti and Shetty 2017); CRAN: vegan (Oksanen et al. 2011); theseus (Price et al. 2018), metacoder (Foster et al. 2017); Github: mare (Korpela 2016), ampvis2 (Andersen et al. 2018), microbiomeutilities (Lahti and Shetty 2017), microbiomeSeq, yingtools2
Diversity analysis	CRAN: picante (Kembel et al. 2010), GUniFrac (Chen 2012), labdsv (Roberts 2007), breakaway (Willis and Bunge 2016), ape (Paradis et al. 2004), RAM (Chen et al. 2016); Github: DivNet (Willis and Martin 2018)
Community types	BioC: DirichletMultinomial (Morgan 2017)
Network analysis	BioC: CCREPE (Schwager et al. 2014); CRAN: igraph (Csardi and Nepusz 2006); Github: SPIEC-EASI (Kurtz et al. 2015)
Group-wise comparisons and association analysis	BioC: structSSI, edgeR, DESeq2, metagenomeSeq; CRAN: mixOmics (Rohart et al. 2017), mixDIABLO (Singh et al. 2019), mixMC (Le Cao et al. 2016), Sigtree (Stevens et al. 2017), ALDEX2 (Fernandes et al. 2014)
Time series analysis	Github: Seqtime (Faust et al. 2018), bootLong (Jeganathan and Holmes 2018), treelapse (Sankaran and Holmes 2018b)
Pipelines/GUIs	BioC: Pathostat (Manimaran et al. 2018), shiny-phyloseq (McMurdie and Holmes 2015), metavizr (Bravo et al. 2017); Github: Rhea (Lagkouvardos et al. 2017), DAME (Piccolo et al. 2018)
Interoperability	CRAN: qiime (Bittinger 2014), BIOM format (McMurdie and Paulson 2016)
Workflows and Tutorials	Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses (Callahan et al. 2016a)
The Riffomonas Reproducible Research Tutorial Series	(Schloss 2018)
Happy belly bioinformatics	(https://astrobioinformatic.github.io/); Microbiome package tutorial series (http://microbiome.github.io/microbiome/); Open & Reproducible Microbiome Data Analysis (https://goo.gl/CPChhd); Random Forest Modelling of the Lake Erie microbial community (https://tinyurl.com/ycz4rgfv)

Review | Published: 20 September 2019

Microbiome data science

Sudarshan A Shetty & Leo Lahti 

[Journal of Biosciences](#) 44, Article number: 115 (2019) | [Cite this article](#)

616 Accesses | 1 Citations | 14 Altmetric | [Metrics](#)

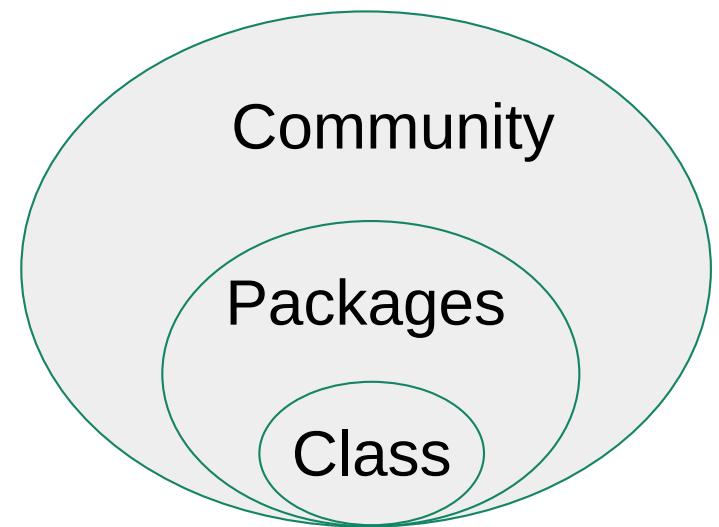
Updated tool listing:
[github.com/microsud/
Tools-Microbiome-Analysis](https://github.com/microsud/Tools-Microbiome-Analysis)



Time for an upgrade?

Tighter integration with related R/Bioc classes

- General performance optimization
- Access to a wider array of existing methods



Support for richer data structures

- Enhancements for hierarchical, multi-assay, and other data types

Reduce overlapping efforts, improve interoperability, ensure sustainability.

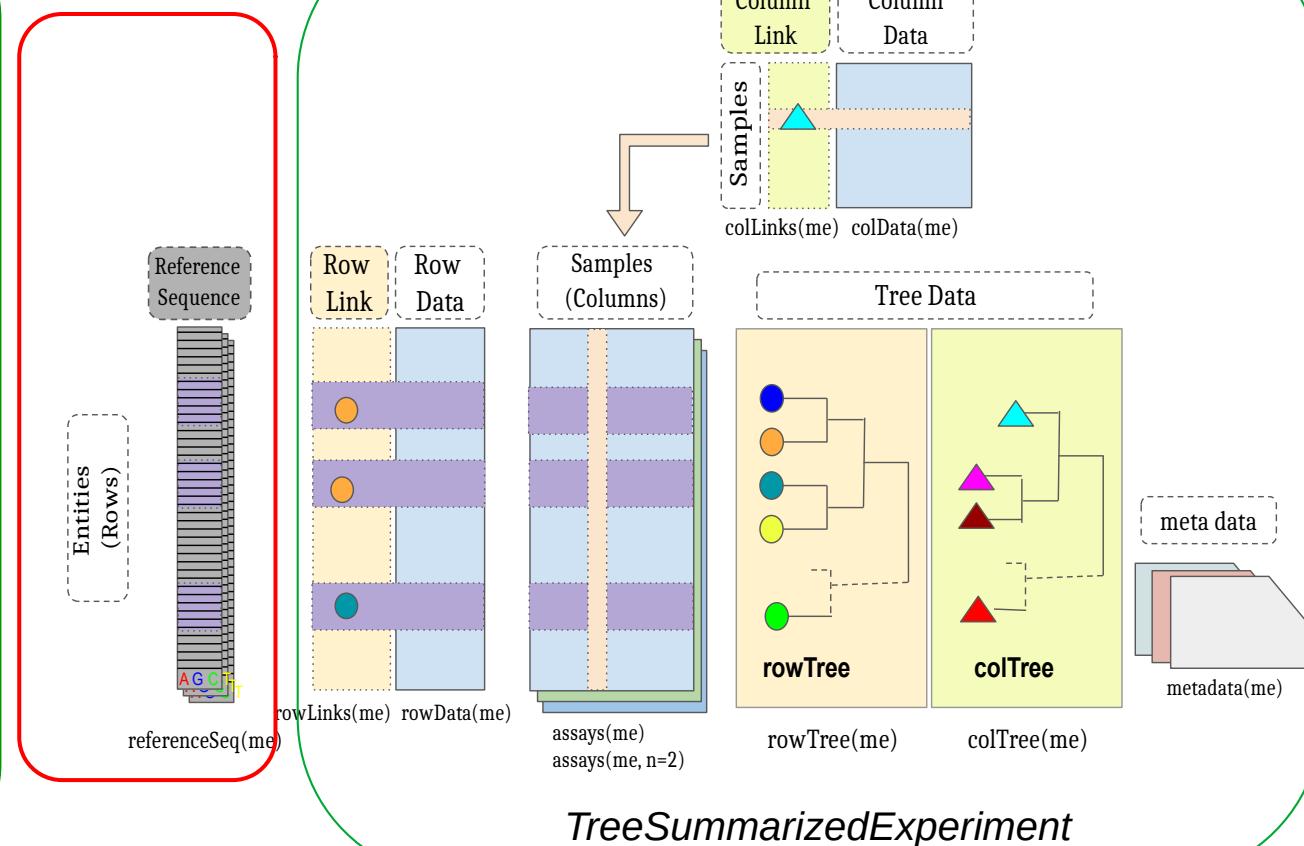
MicrobiomeExperiment class?

derivative of *TreeSummarizedExperiment* by Ruizhu @fiona Huang;
initially proposed by Hector Bravo & Domenick Braccia

Tested tools for hierarchical data,
spanning both sample and
feature space + other methods for
SummarizedExperiment
derivatives

Inherit support for sparse
matrices & multiple assays while
providing improvements in speed
and memory compared to the
current solutions for microbiome
data.

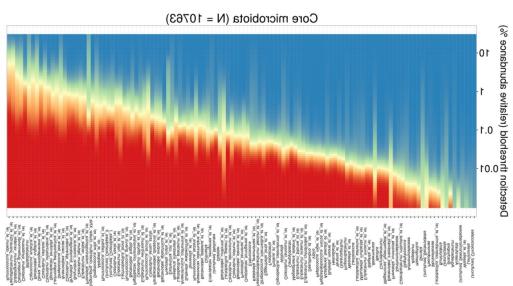
Improved support for microbiome
features: more detailed sequence
information based on classes
such as the *DNAStringSet(List)*.



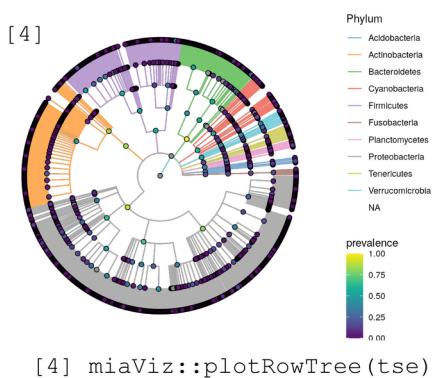
Seamless conversion from *phyloseq* & other raw data types

Package ecosystem

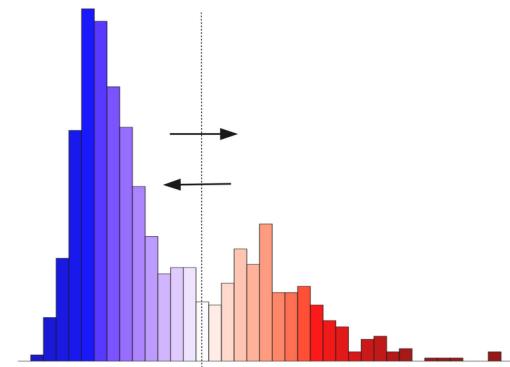
mia – microbiome analysis
getPrevalentTaxa(x)
getDiversity(x)
calculateDMM(x)



miaViz - visualization



miaTime - time series



Data packages

ExperimentHub

platforms all rank 76 / 1974 posts 2 / 1 / 2e+01 / 1 in Bioc 4 years
build ok updated before release dependencies 72

DOI: [10.18129/B9.bioc.ExperimentHub](https://doi.org/10.18129/B9.bioc.ExperimentHub)

Other tools?

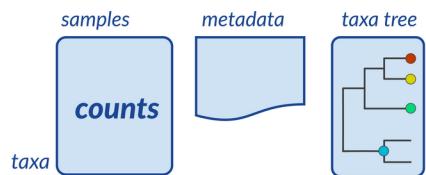
- Utilities
- SingleCellExperiment adaptations
- Independent contributions
- etc.

Example workflow

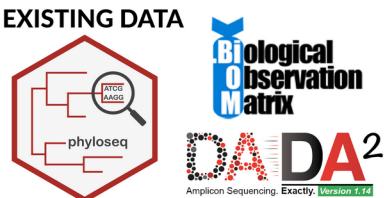
Import Data

This workflow starts with either raw data directly from relative abundance estimation or taxonomic classification OR pre-existing data objects from widely used software.

RAW DATA

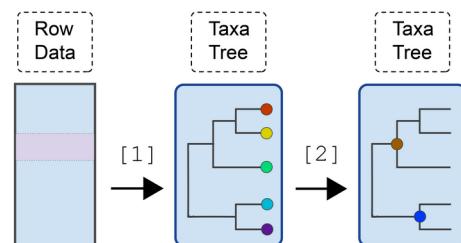


EXISTING DATA

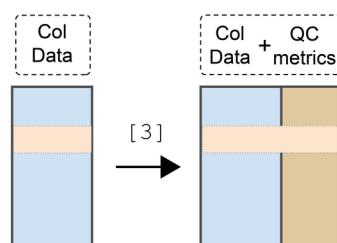


The mia Pipeline

Accessing Taxonomic Info.



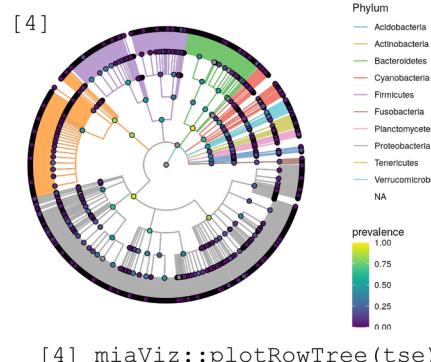
Quality Control



```
[1] mia::addTaxonomyTree(tse)  
[2] TreeSE:::aggValue(tse)
```

```
[3] scatter:::addPerCellQC(tse)
```

Visualizing with miaViz



Check the poster
F1000 / EuroBioC!



MiaBook - Manual & Education

Preface

I Introduction

1 Data Infrastructure

1.1 Installation

1.2 Background

1.3 Loading experimental microbio...

1.4 Metadata

1.5 Microbiome and tree data spec...

1.6 Data conversion

1.7 Conclusion

Session Info

Microbiome Analysis

Authors: Leo Lahti [aut], Sudarshan Shetty [aut], Felix GM Ernst [aut, cre]

Version: 0.98.0003

Modified: 2020-12-06

Compiled: 2020-12-13

Environment: R version 4.0.0 (2020-04-24), Bioconductor 3.11

License: CC BY-NC-SA 3.0 US

Copyright:

Source: <https://github.com/microbiome/MiaBook>

Preface

This website is a book on microbiome analysis in the Bioconductor universe and is showing common principles and workflows of performing microbiome analysis.

The book was borne out of necessity, while updating tools for microbiome analysis to work with common classes of the Bioconductor project handling count data of various sorts. It is heavily influenced by similar resources, such as the [Orchestrating Single-Cell Analysis with Bioconductor](#) book, [phyloseq tutorials](#) and [microbiome tutorials](#).

We focus on microbiome analysis tools, new, updated and established methods. In the *Introduction* section, we show how to work with the key data infrastructure `TreeSummarizedExperiment` and related classes, how this framework relates to other infrastructure and how to load microbiome analysis data to work with in the context of this framework.

The second section, *Focus Topics*, is all about the steps for analyzing microbiome data, beginning with the most common steps and progressing to more specialized methods in subsequent sections.

The third section, *Appendix*, contains the rest of things we didn't find another place for, yet.

Community

Slack:
#MicrobiomeExperiment

Workshop:
Friday, 17:05-18:05 CET

Project website:
microbiome.github.io

microbiome.github.io



Special thanks

Felix G.M. Ernst
Sudarshan A. Shetty
Tuomas Borman
Ruizhu Huang
Domenick J. Braccia
Héctor Corrada Bravo
Leo Lahti

