

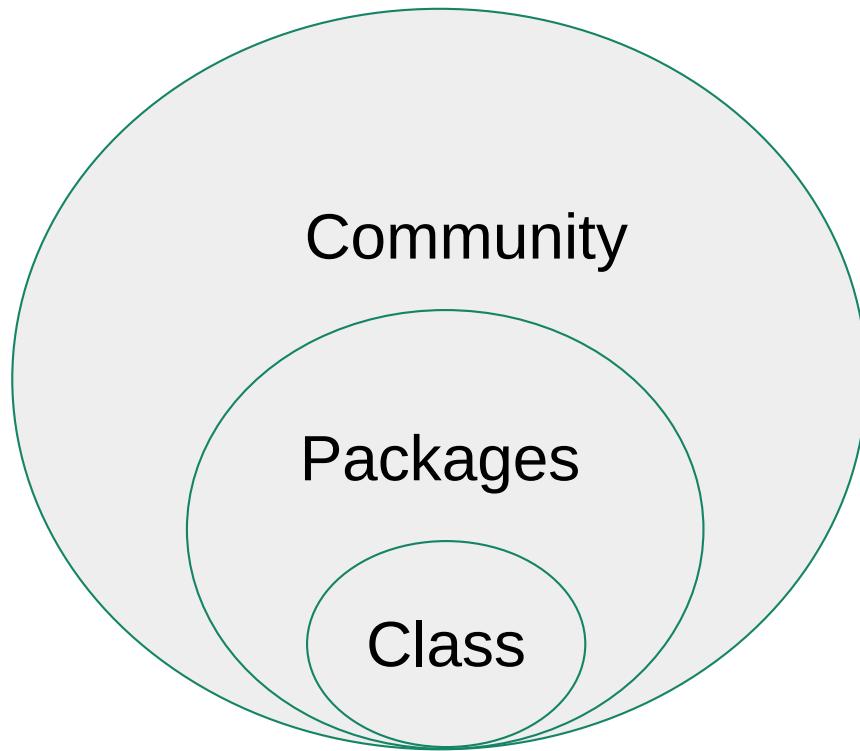
# Microbiome data science in the *SummarizedExperiment* universe

**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



*Standardized data containers*  
are central for the R/Bioc ecosystem



# Optimal container for microbiome data?

**Multiple assays**  
seamless interlinking

**Hierarchical data**  
supporting samples & features

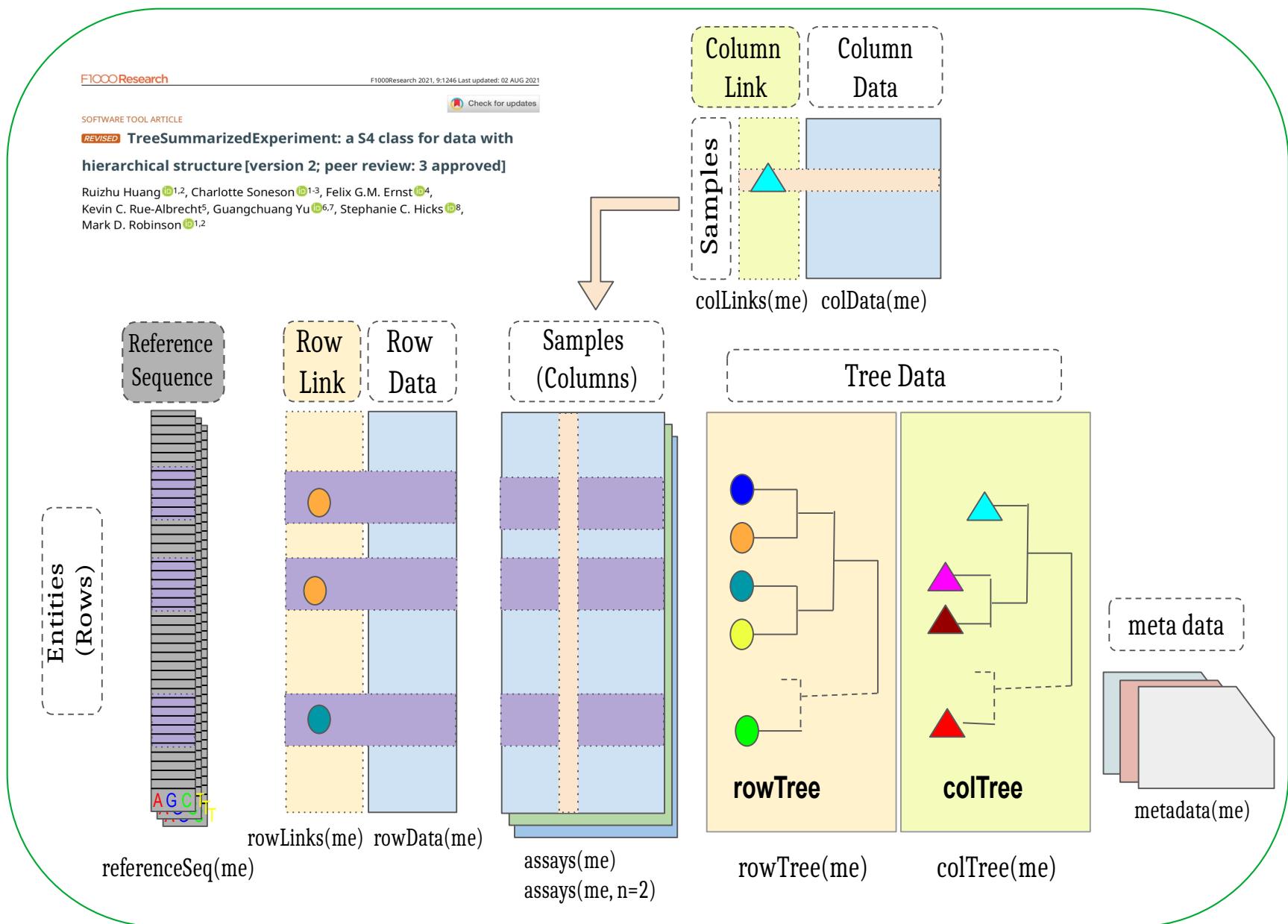
**Side information**  
extended capabilities & data types

**Optimized**  
for speed & memory

**Integrated**  
with other applications & frameworks

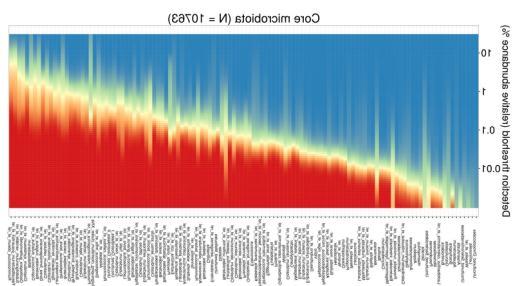
Reduce overlapping efforts, improve interoperability, ensure sustainability.

# (Tree)SummarizedExperiment

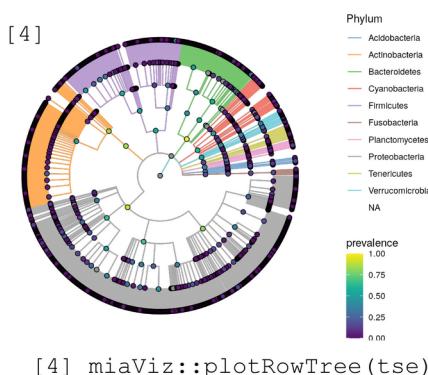


# Current package ecosystem: *miaverse*

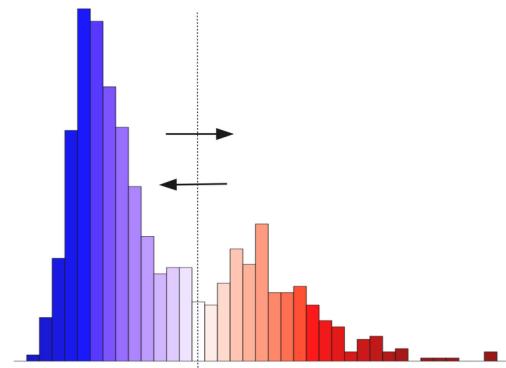
**mia – microbiome analysis**  
getPrevalentTaxa(x)  
getDiversity(x)  
calculateDMM(x)



**miaViz - visualization**



**miaTime - time series**



**Data packages - *microbiomeDataSets***

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.

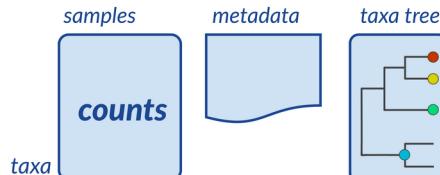
# Example workflow

Figure by Domenick Braccia (EuroBioC 2020)

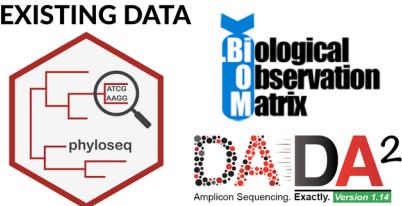
## Import Data

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

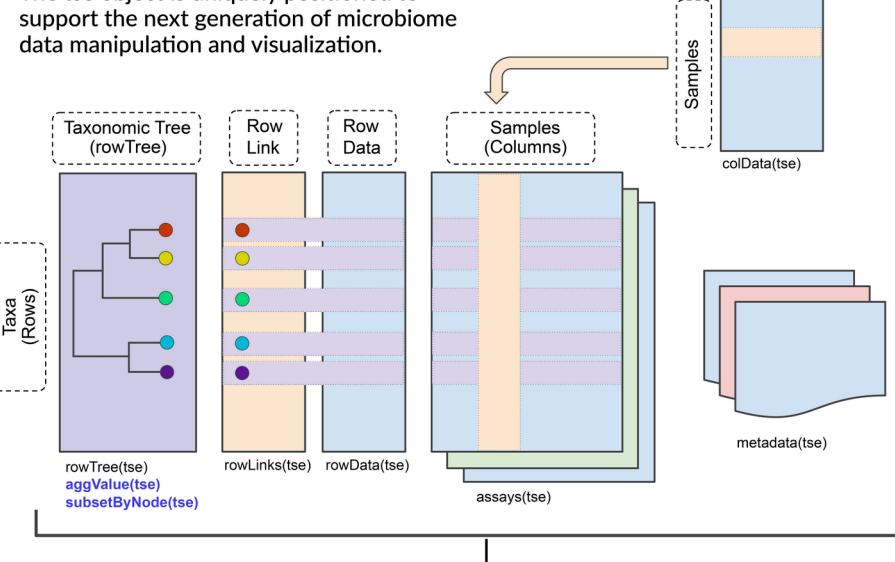


### EXISTING DATA



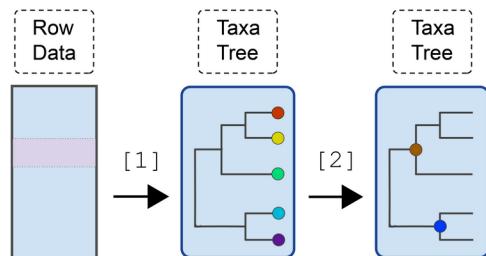
## The TreeSE object

The tse object is uniquely positioned to support the next generation of microbiome data manipulation and visualization.



## The mia Pipeline

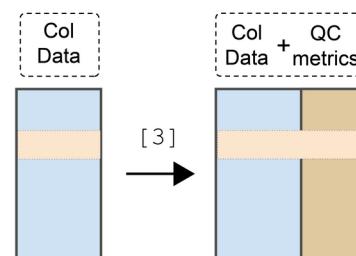
### Accessing Taxonomic Info.



[1] `mia::addTaxonomyTree(tse)`

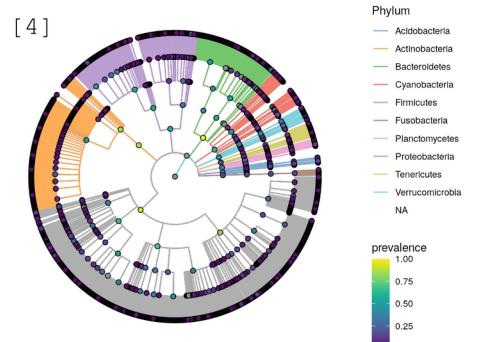
[2] `TreeSE::aggValue(tse)`

### Quality Control



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

### Visualizing with miaViz



[4] `miaViz::plotRowTree(tse)`

# 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. Callahan<sup>1</sup>, Kris Sankaran<sup>1</sup>, Julia A. Fukuyama<sup>1</sup>, Paul J. McMurdie<sup>2</sup>,  Susan P. Holmes<sup>1</sup>

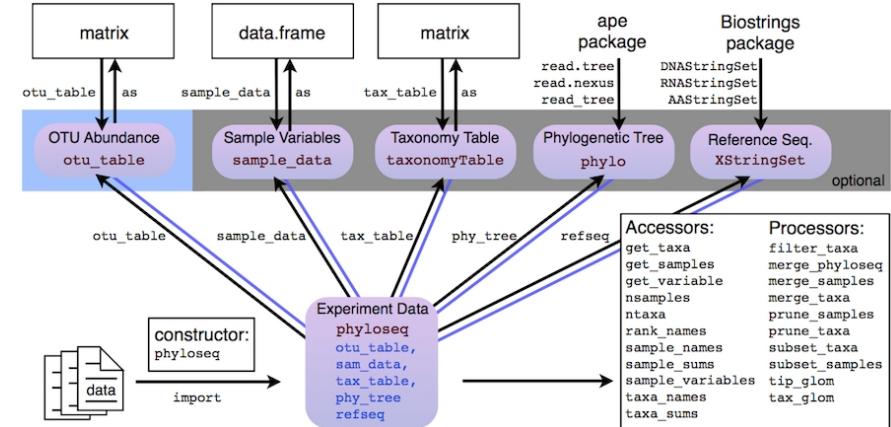


Table 1 Overview of contemporary online resources for microbiome data science in R. The indicated groupings are approximations as many packages span over multiple categories

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), microbiomewutilities (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)



# Orchestrating Microbiome Analysis with R/Bioc

## Preface

## I Introduction

### 1 Data Infrastructure

#### 1.1 Installation

#### 1.2 Background

#### 1.3 Loading experimental microbi...

#### 1.4 Metadata

#### 1.5 Microbiome and tree data speci...

#### 1.6 Data conversion

#### 1.7 Conclusion

## Session Info

## Orchestrating Microbiome Analysis

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

**Version:** 0.98.9

**Modified:** 2021-04-10

**Compiled:** 2021-07-29

**Environment:** R version 4.1.0 (2021-05-18), Bioconductor 3.14

**License:** CC BY-NC-SA 3.0 US

**Copyright:**

**Source:** <https://github.com/microbiome/OMA>

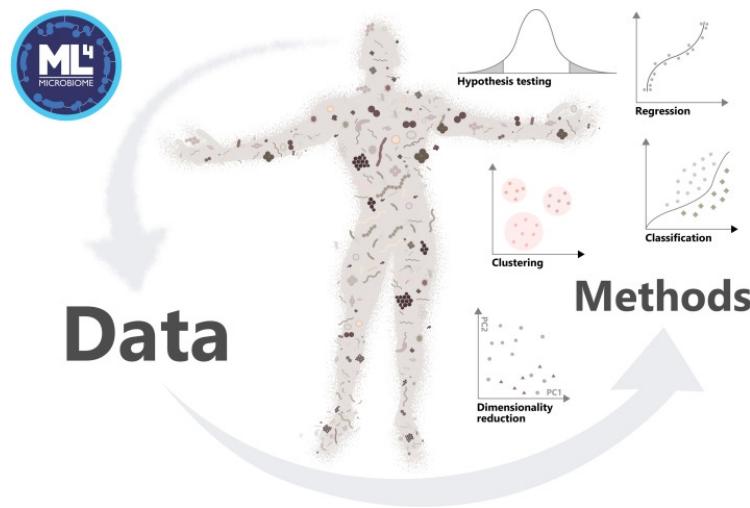


Figure source: Moreno-Indias et al. (2021) Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. *Frontiers in Microbiology* 12:11.

# Community

*Special thanks:*

Felix G.M. Ernst, Sudarshan A. Shetty, Tuomas Borman, Henrik Eckermann, Chandler Ross, Chouaib Benchraka, Shigdel Rajesh, Ruizhu Huang, Domenick J. Braccia , Héctor Corrada Bravo, Leo Lahti

**Slack:**  
**#miaverse**

**Project website:**  
[microbiome.github.io](https://microbiome.github.io)



MICROBIOME



Turun yliopisto  
University of Turku



SUOMEN AKATEMIA  
FINLANDS AKADEMI • ACADEMY OF FINLAND