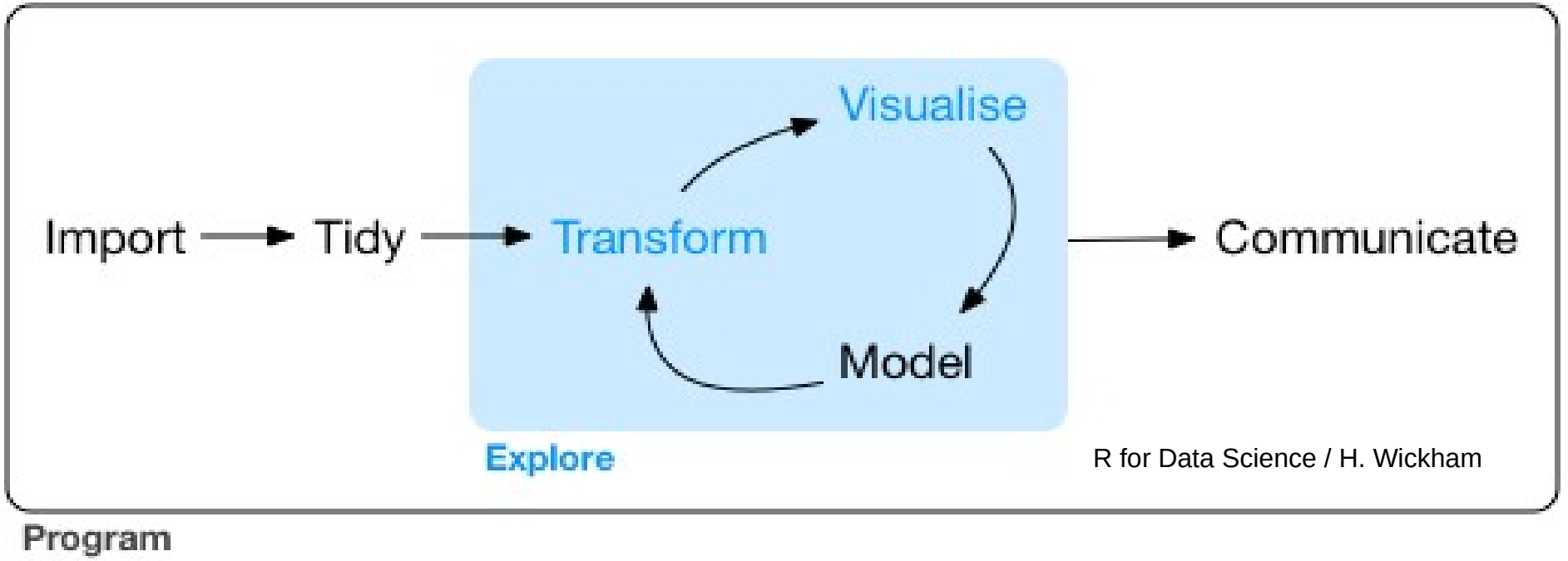

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Data science workflow



R for Data Science / H. Wickham

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

Ben J. Callahan[1], Kris Sankaran[1], Julia A. Fukuyama[1], Paul J. McMurdie[2], Susan P. Holmes

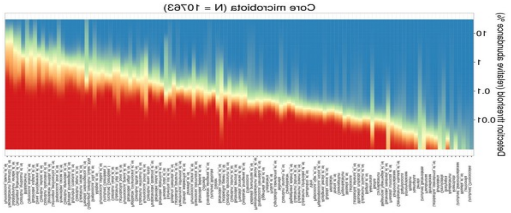
This article is included in the Bioconductor gateway.

microbiome R package

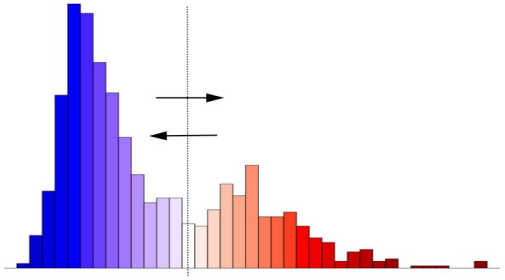
Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

chat on gitter | build passing | codecov 24% | PRs welcome

Core & prevalence
prevalence(x)
core(x)
core_members(x)

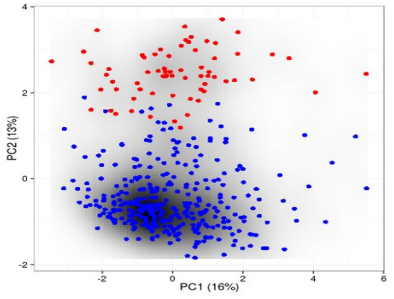


Stability & resilience



Alpha & beta diversity

alpha(x)
diversity(x)
evenness(x)
dominance(x)
rarity(x)
readcount(x)



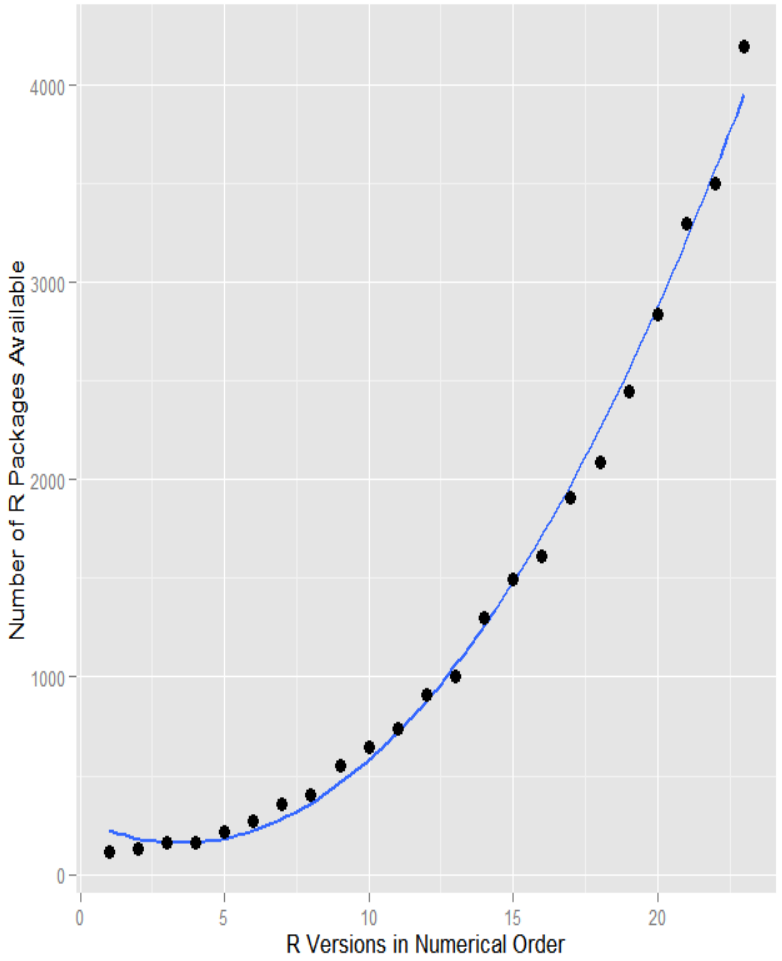
Transformations

transform(x, “compositional”)
transform(x, “clr”)
transform(x, “log10p”)
transform(x, “hellinger”)
transform(x, “identity”)

Community
- Online tutorials
- Mailing list
- Gitter chat
- Example data
- Workshops

Quality control
- continous integration
- unit tests

Number of open analysis tools has grown exponentially



github
SOCIAL CODING

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

R

ROpenSci

1. Ampvis2 Tools for visualising amplicon sequencing data
2. CCREPE Compositionality Corrected by PErmutation and REnormalization
3. DADA2 Divisive Amplicon Denoising Algorithm
4. DESeq2 Differential expression analysis for sequence count data
5. edgeR empirical analysis of DGE in R
6. mare Microbiota Analysis in R Easily
7. Metacoder An R package for visualization and manipulation of community taxonomic diversity data
8. metagenomeSeq Differential abundance analysis for microbial marker-gene surveys
9. microbiome R package Tools for microbiome analysis in R
10. MINT Multivariate INTegrative method
11. mixDIABLO Data Integration Analysis for Biomarker discovery using Latent variable approaches for 'Omics studies
12. mixMC Multivariate Statistical Framework to Gain Insight into Microbial Communities
13. MMinte Methodology for the large-scale assessment of microbial metabolic interactions (MMinte) from 16S rDNA data
14. pathostat Statistical Microbiome Analysis on metagenomics results from sequencing data samples
15. phylofactor Phylogenetic factorization of compositional data
16. phylogeo Geographic analysis and visualization of microbiome data
17. Phyloseq Import, share, and analyze microbiome census data using R
18. qiimer R tools compliment qiime
19. RAM R for Amplicon-Sequencing-Based Microbial-Ecology
20. ShinyPhyloseq Web-tool with user interface for Phyloseq
21. SigTree Identify and Visualize Significantly Responsive Branches in a Phylogenetic Tree
22. SPIEC-EASI Sparse and Compositionally Robust Inference of Microbial Ecological Networks
23. structSSI Simultaneous and Selective Inference for Grouped or Hierarchically Structured Data
24. Tax4Fun Predicting functional profiles from metagenomic 16S rRNA gene data
25. taxize Taxonomic Information from Around the Web
26. labdsv Ordination and Multivariate Analysis for Ecology
27. Vegan R package for community ecologists
28. igraph Network Analysis and Visualization in R
29. MicrobiomeHD A standardized database of human gut microbiome studies in health and disease *Case-Control*
30. Rhea A pipeline with modular R scripts
31. microbiomeutilities Extending and supporting package based on microbiome and phyloseq R package
32. breakaway Species Richness Estimation and Modeling

[A survey for 16S Github.com/microsud/Tools-Microbiome-Analysis]

Springer Link

Journal of Biosciences

October 2019, 44:115 | Cite as

Microbiome data science

Authors | Authors and affiliations

Sudarshan A Shetty, Leo Lahti

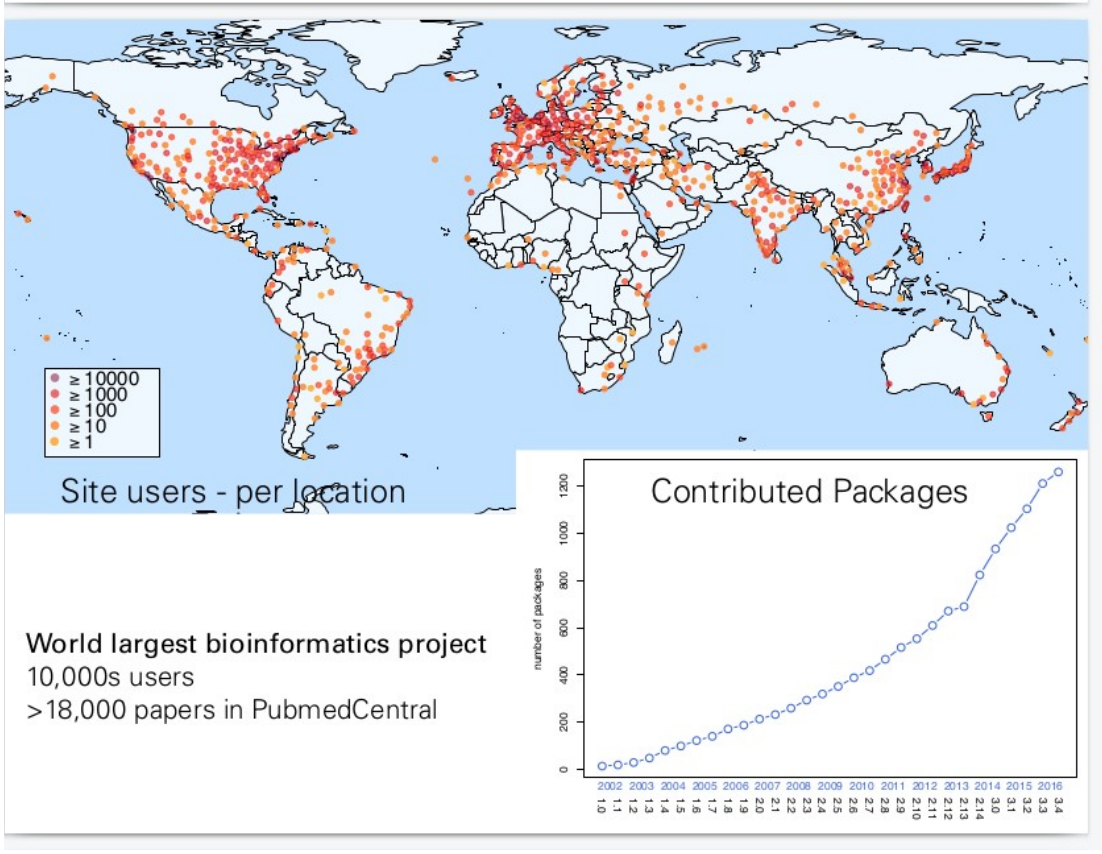
Bioconductor

OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Started in 2001 as a platform for microarray data analysis

Now 2140+ packages:

- Sequencing (DNA, RNA, ChIP, SC)
- Microarrays
- Flow cytometry
- Proteomics
- Multi-omics data integration



What is Bioconductor ?

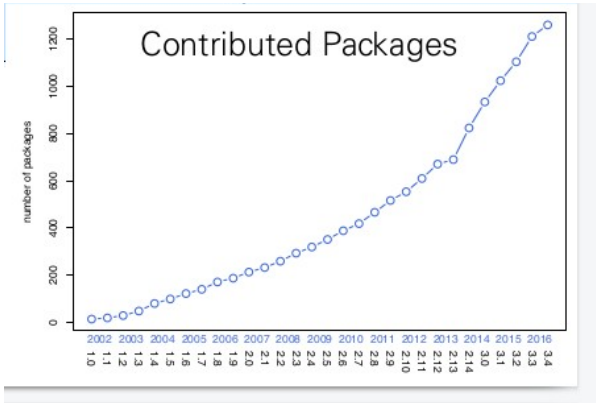
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Collaborative research software development

- software & data repository

- tutorials & documentation

- bioinformatics peer support



50,000+ papers in PubMed

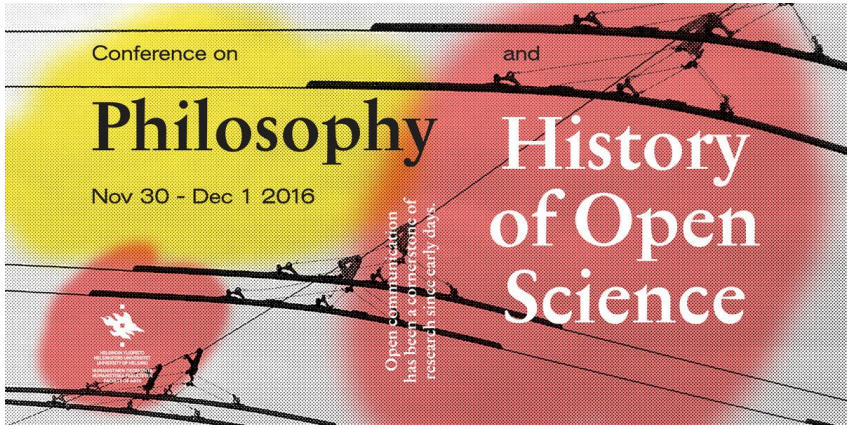
Management:
- Core team
- SAB & TAB & CAB
- Global contributor base

What marks out modern science is not the conduct of experiments (alchemists conducted plenty of experiments), but the formation of a *critical community capable of assessing discoveries and replicating results*.

The Invention of Science: A New History of the Scientific Revolution, by David Wootton

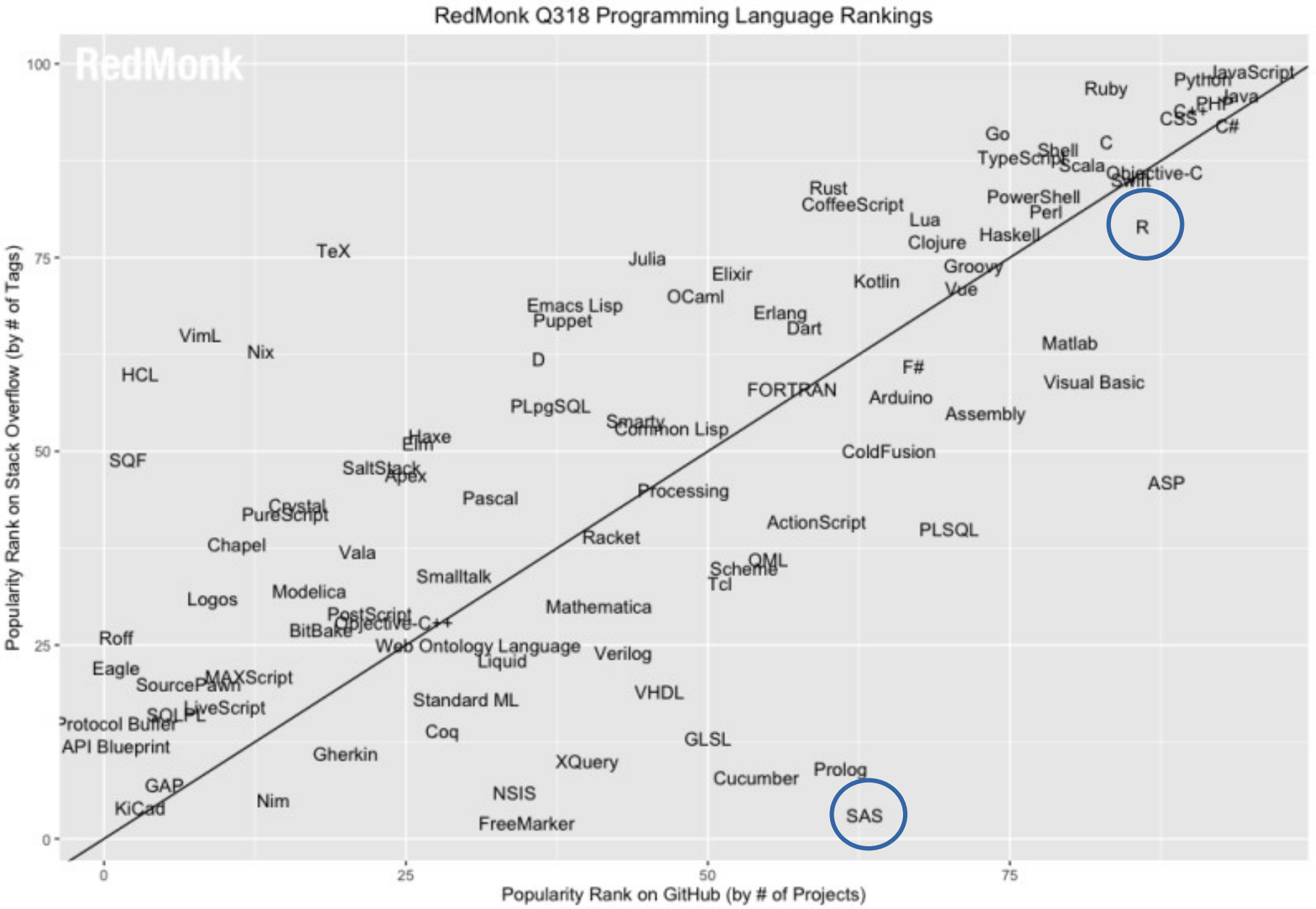
Alchemy & algorithms: perspectives on the philosophy and history of open science

Leo Lahti, Filipe da Silva, Markus Petteri Laine, Viivi Lähteenoja, Mikko Tolonen

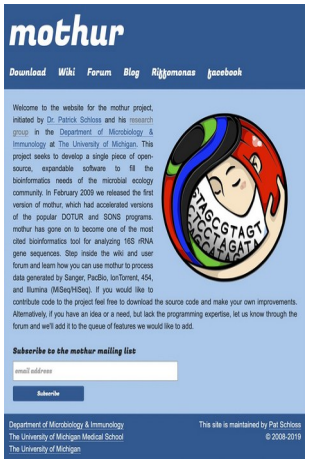


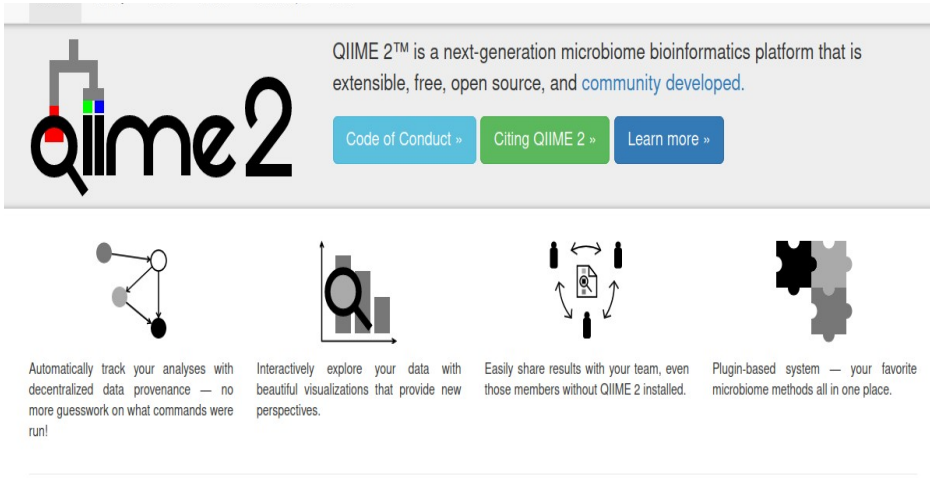
A family of alchemists at work, an engraving by Philip Galle, after a painting by Pieter Bruegel the Elder, published by Hieronymus Cock, *c.*1558.

Varying cultures of open collaboration



open data science ecosystems





Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

PeerJ ›

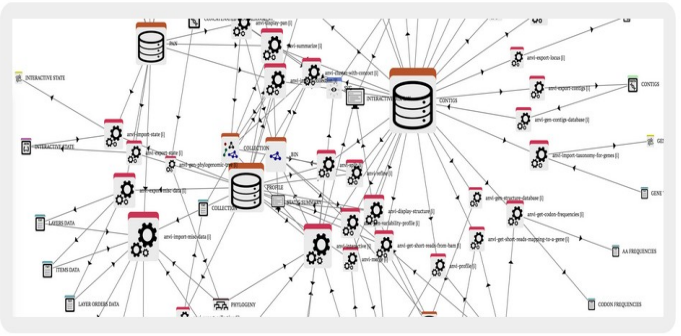
Anvi'o: an advanced analysis and visualization platform for 'omics data

Research article | Bioinformatics | Biotechnology | Computational Biology | Genomics | Microbiology

A. Murat Eren[1,2], Özcan C. Esen[1], Christopher Quince[3], Joseph H. Vineis[1], Hilary G. Morrison[1], Mitchell L. Sogin[1], Tom O. Delmont[1]

blished October 8, 2015

Anvi'o in a nutshell



Anvi'o is an open-source, community-driven analysis and visualization platform for 'omics data.

FOCUS ON COMPUTER RESOURCES | OCTOBER 31 2017

Software for the Integration of Multiomics Experiments in Bioconductor FREE

Marcel Ramos; Lucas Schiffer; Angela Re; Rimsha Azhar; Azfar Basunia; Carmen Rodriguez; Tiffany Chan; Phil Chapman; Sean R. Davis; David Gomez-Cabrero; Aedin C. Culhane; Benjamin Haibe-Kains; Kasper D. Hansen; Hanish Kodali; Marie S. Louis; Arvind S. Mer; Markus Riester; Martin Morgan; Vince Carey; Levi Waldron

Cancer Res (2017) 77 (21): e39–e42.

https://doi.org/10.1158/0008-5472.CAN-17-0344 **Article history**

Important themes:

- Reproducible research
- Interoperable packages
- Usability & documentation

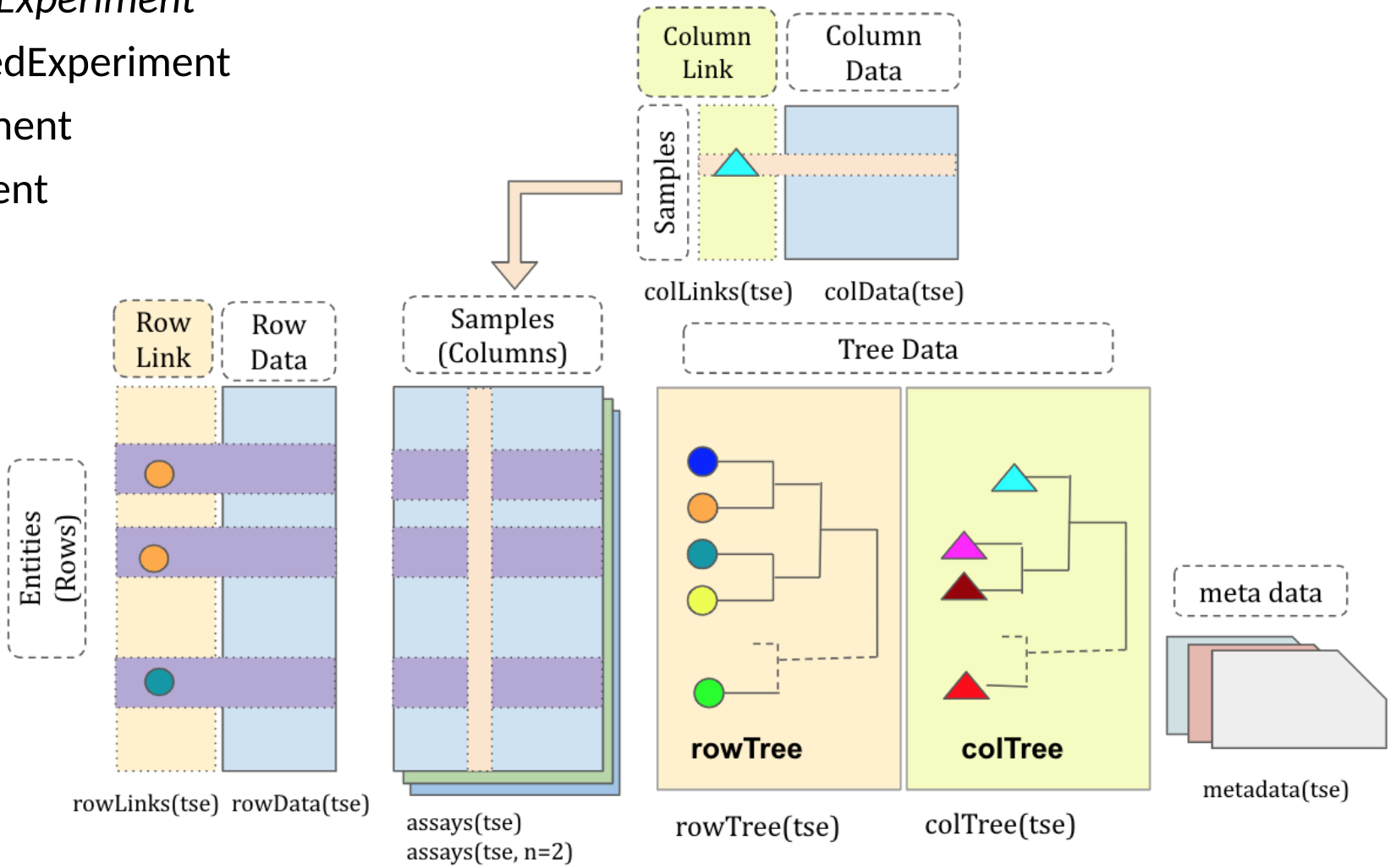
(Tree)SummarizedExperiment

RangedSummarizedExperiment

MultiAssayExperiment

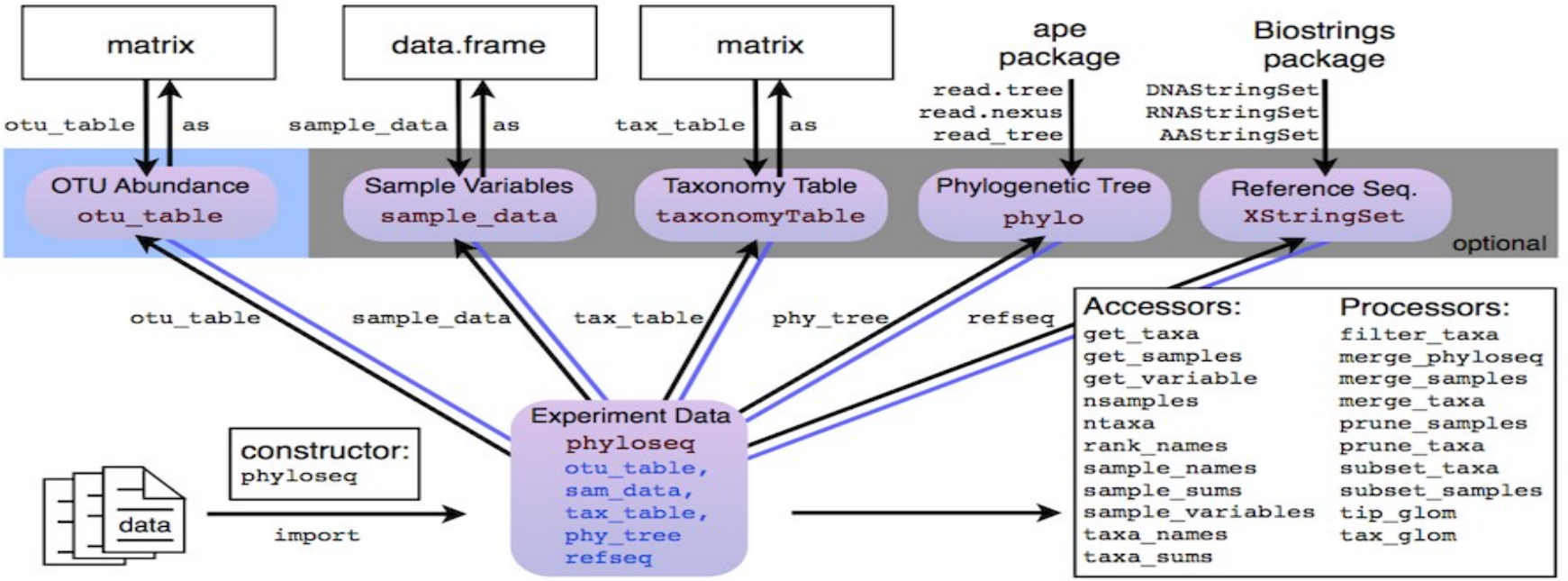
SingleCellExperiment

SpatialExperiment



Alternative data container: *phyloseq*

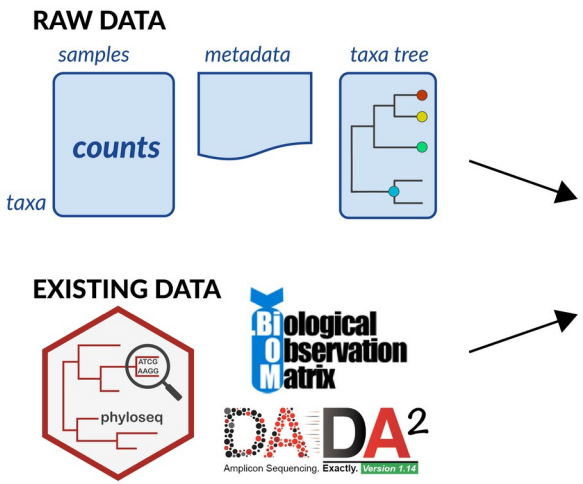
Standard for (16S) microbiome bioinformatics in R (J McMurdie, S Holmes *et al.*)



Data containers support collaborative development of analysis methods & workflows

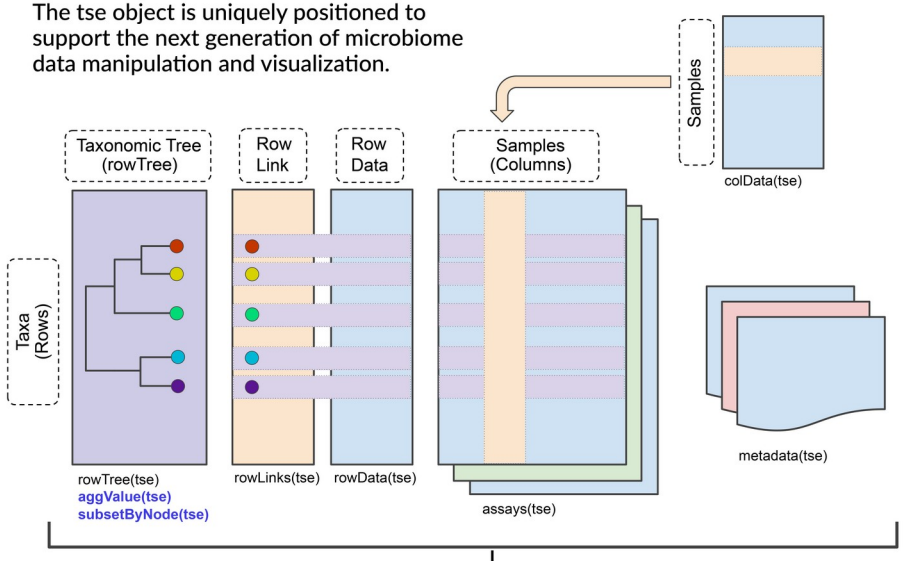
Import Data

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



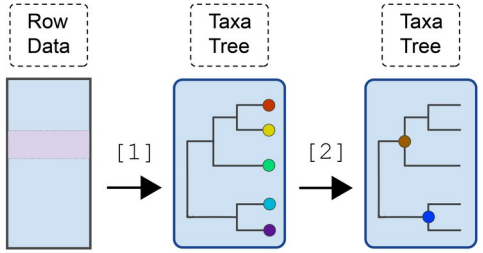
The `TreeSE` object

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

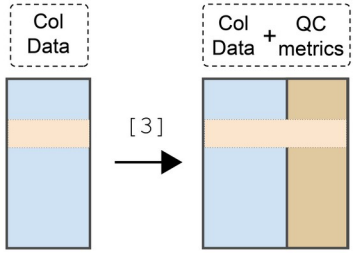


The `mia` Pipeline

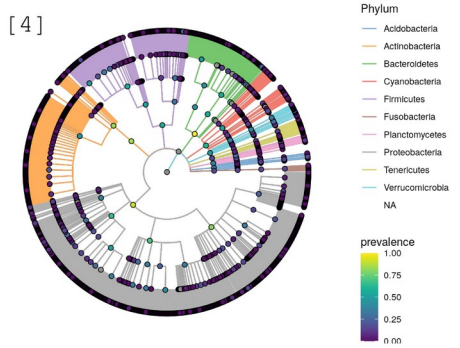
Accessing Taxonomic Info.



Quality Control



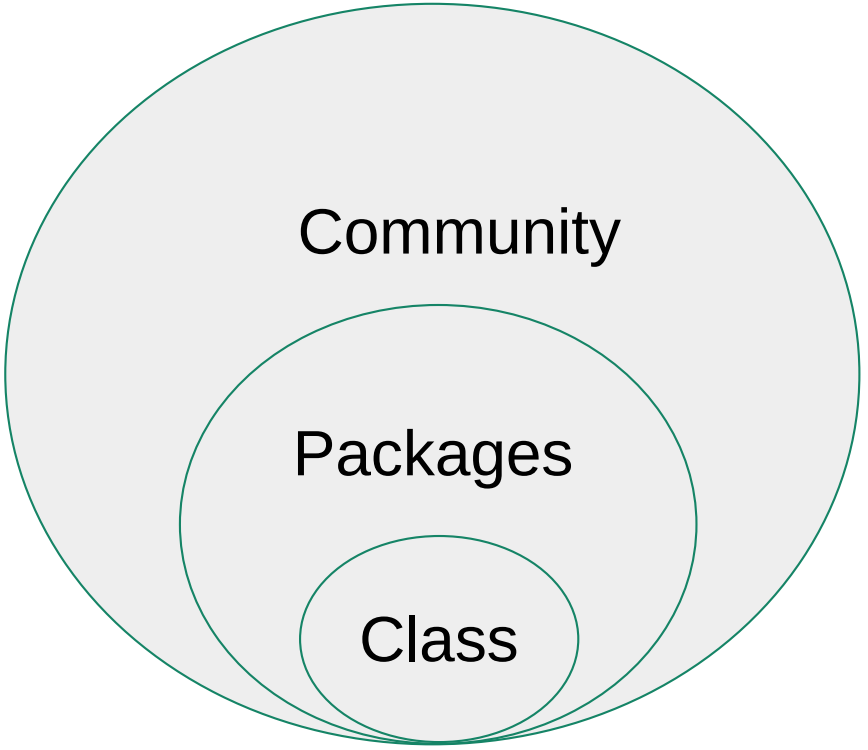
Visualizing with `miaViz`



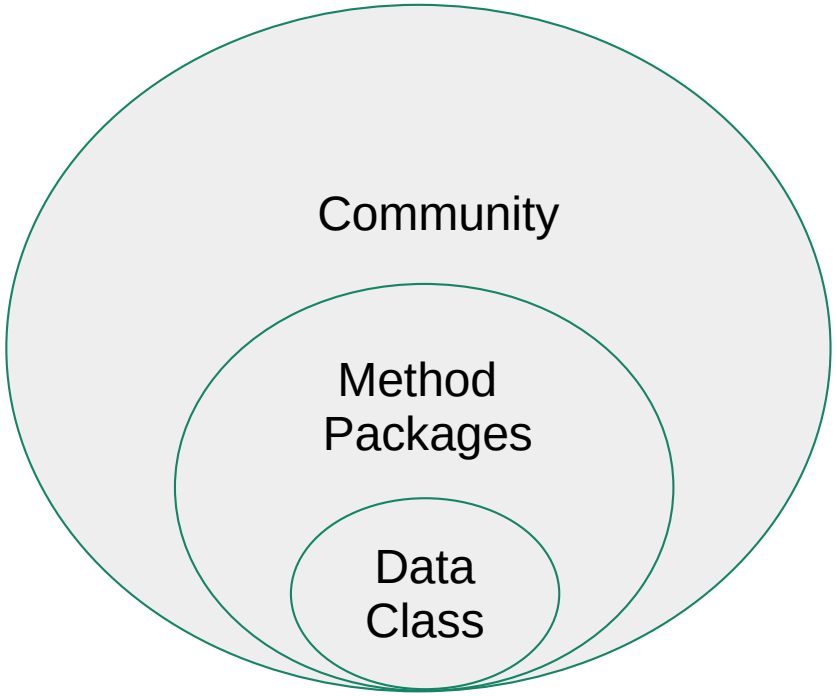

```
[1] mia::addTaxonomyTree(tse)
[2] TreeSE::aggValue(tse)
[3] scatter::addPerCellQC(tse)
[4] miaViz::plotRowTree(tse)
```

microbiome.github.io

Standardized data containers are central for the R/Bioc ecosystem



Reduce overlapping efforts, improve interoperability, ensure sustainability.



Data packages

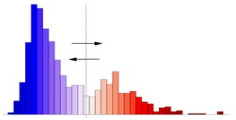
ExperimentHub

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

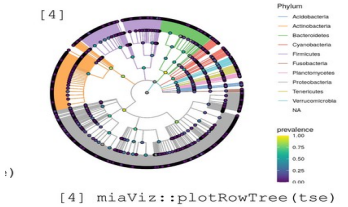
DOI: 10.18129/B9.bioc.ExperimentHub

mia – microbiome analysis

getDiversity(x)
calculateDMM(x)



miaViz - Visualization



Package ecosystem

Transparency

Reproducibility

Collaboration

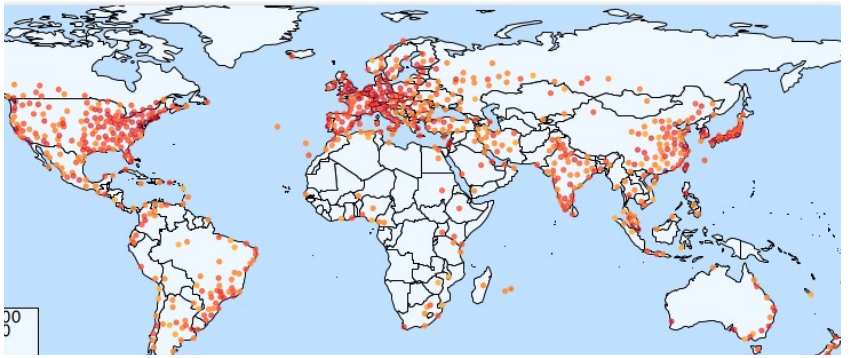
R/Bioc community

(Global) Bioconductor conference

Asian Bioconductor conference

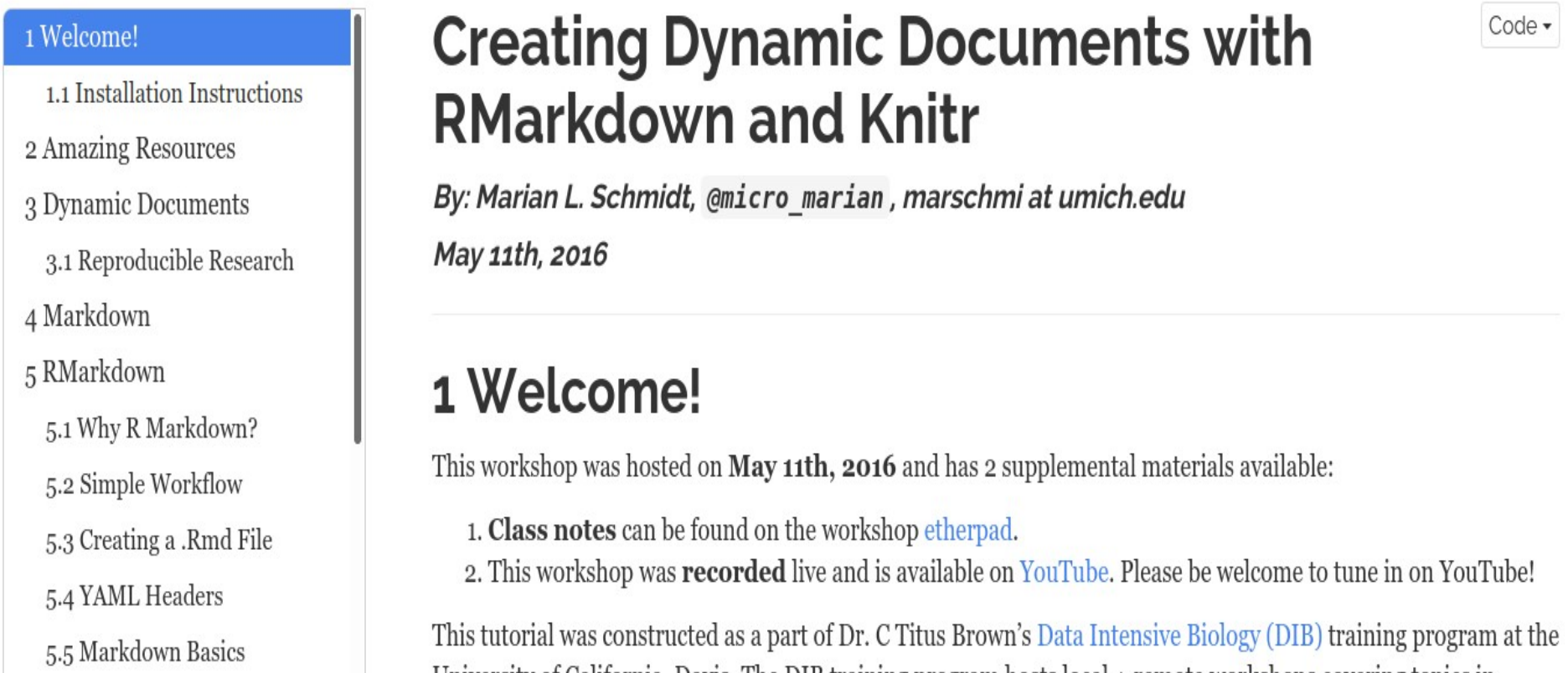
European Bioconductor conference

- *Heidelberg*, Sep 14-16, 2022



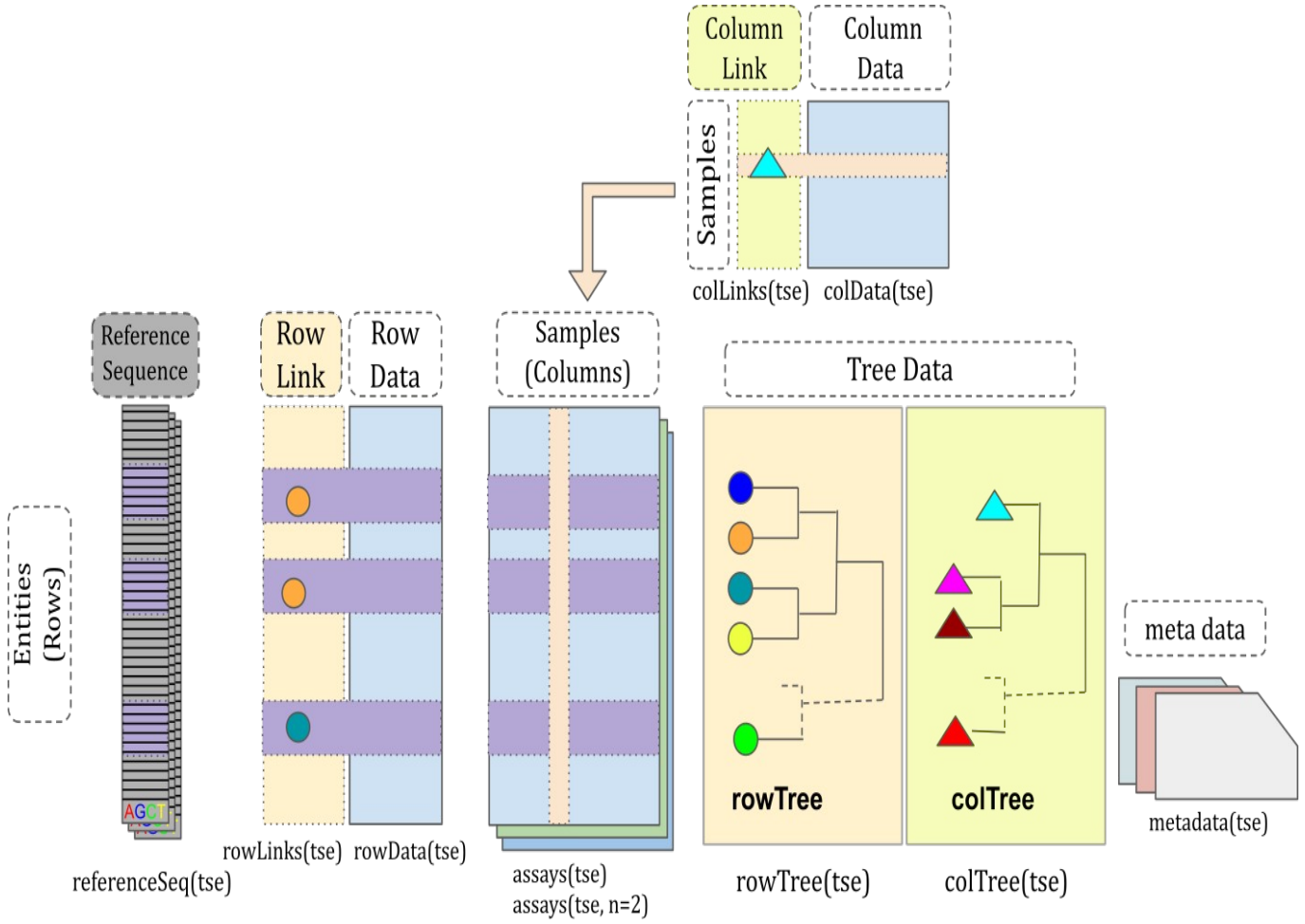
Initializing reproducible report

Task: create Rmarkdown document that imports the data



Importing microbiome data to R

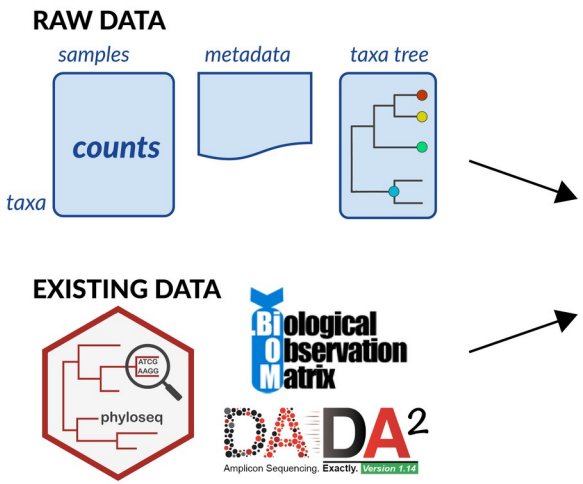
Task: import biom data files for your chosen case study into a specific data container (structure) in R, TreeSummarizedExperiment (TSE) Huang et al. (2020). This provides the basis for downstream data analysis.



Data containers support collaborative development of analysis methods & workflows

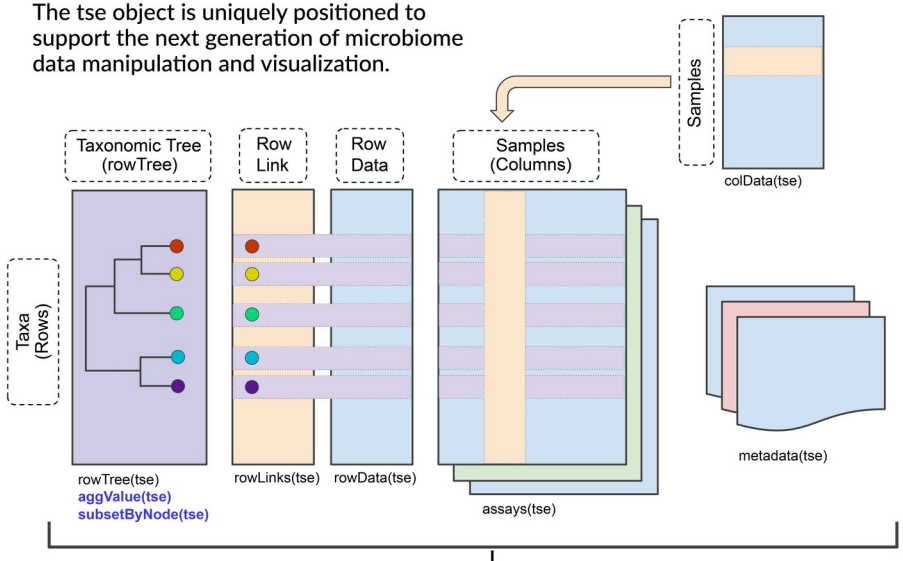
Import Data

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



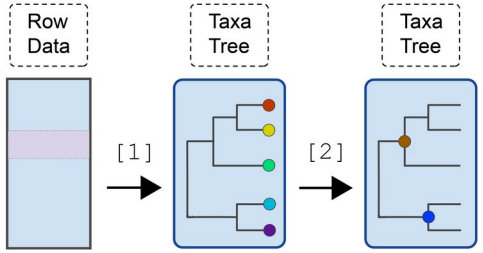
The `TreeSE` object

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

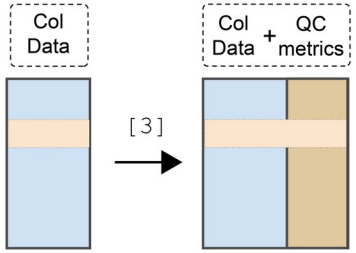


The `mia` Pipeline

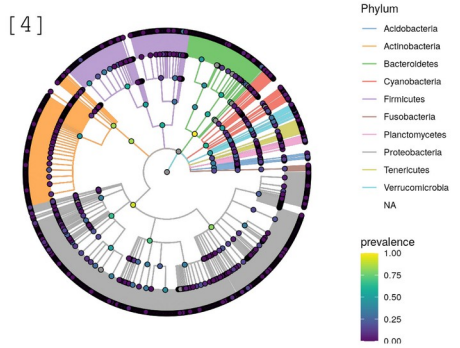
Accessing Taxonomic Info.



Quality Control



Visualizing with `miaViz`




```
[1] mia::addTaxonomyTree(tse)
[2] TreeSE::aggValue(tse)
[3] scatter::addPerCellQC(tse)
[4] miaViz::plotRowTree(tse)
```

microbiome.github.io

Orchestrating Microbiome Analysis with R and Bioconductor – online book: *beta version*

microbiome.github.io

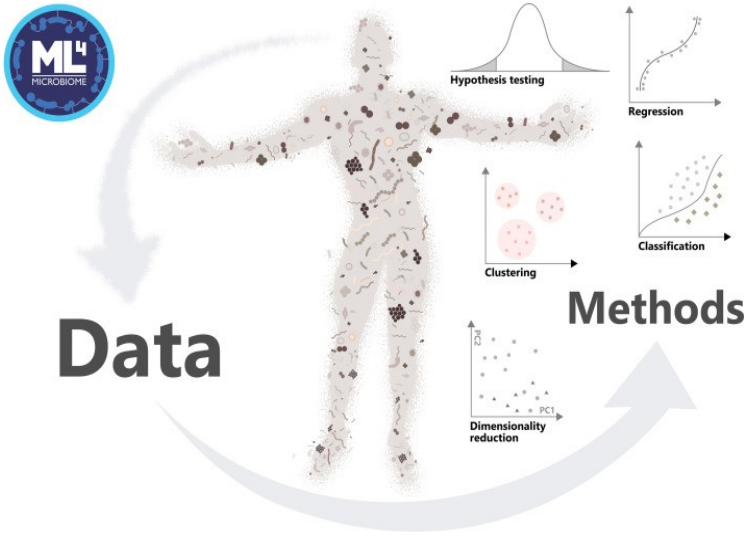


Figure source: Moreno-Indias et al. (2021) Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. Frontiers in Microbiology 12:11.

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.

Figure 5: The online version provides the text in HTML, data files and up-to-date code.