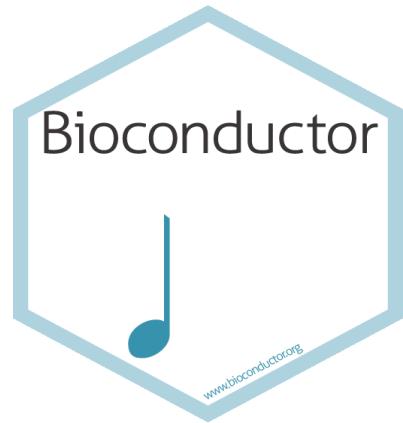


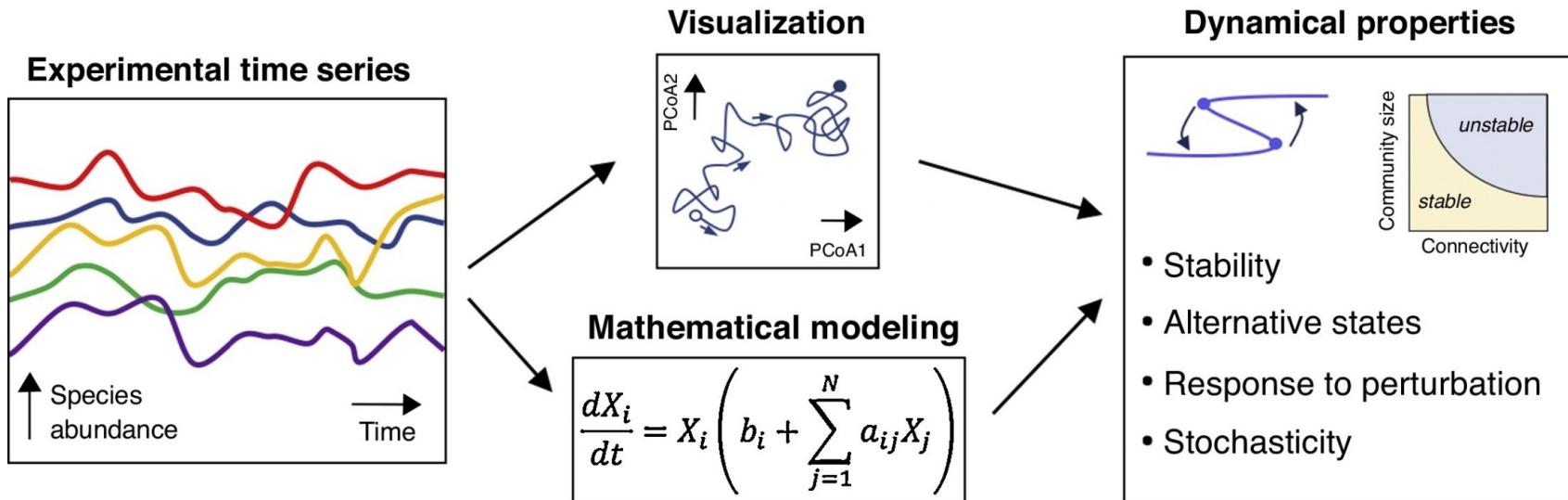
Microbiome package ecosystem

Leo Lahti, CSC course, Nov 28 – Dec 2, 2022



Microbial communities as dynamical systems

Didier Gonze ^{1, 2✉}, Katharine Z Coyte ^{3, 4}, Leo Lahti ^{5, 6, 7}, Karoline Faust ^{5✉}

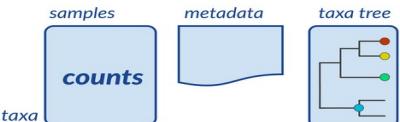


Workflow with *mia* & *miaViz* R packages

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

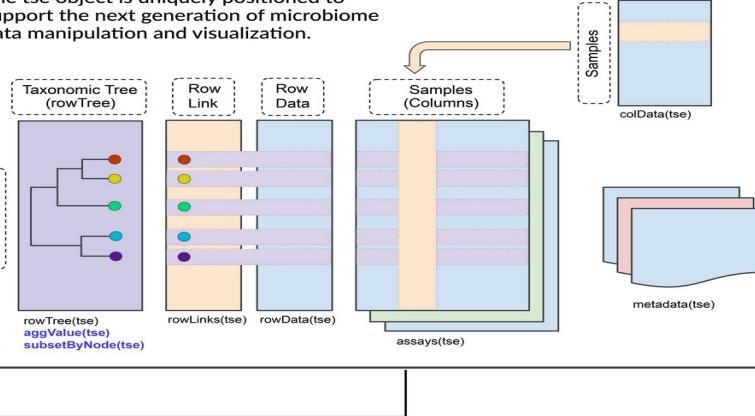


Amplicon Sequencing

DA²

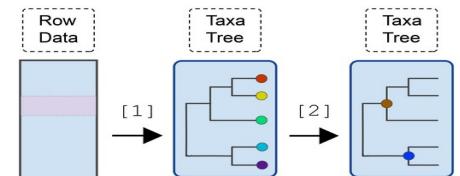
The TreeSE object

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

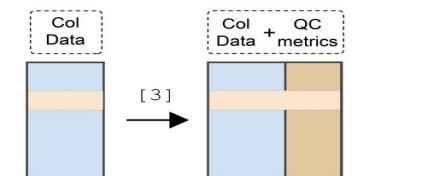


The mia Pipeline

Accessing Taxonomic Info.

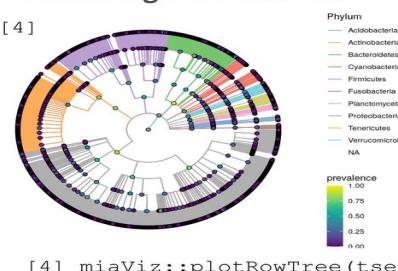


Quality Control



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

Visualizing with miaViz



European Bioconductor Meeting 2020

- Where: Virtual Conference
- When: 14-18 December 2020
- On twitter: #EuroBioC2020



Special thanks

Felix G.M. Ernst
Sudarshan A. Shetty

Tuomas Borman

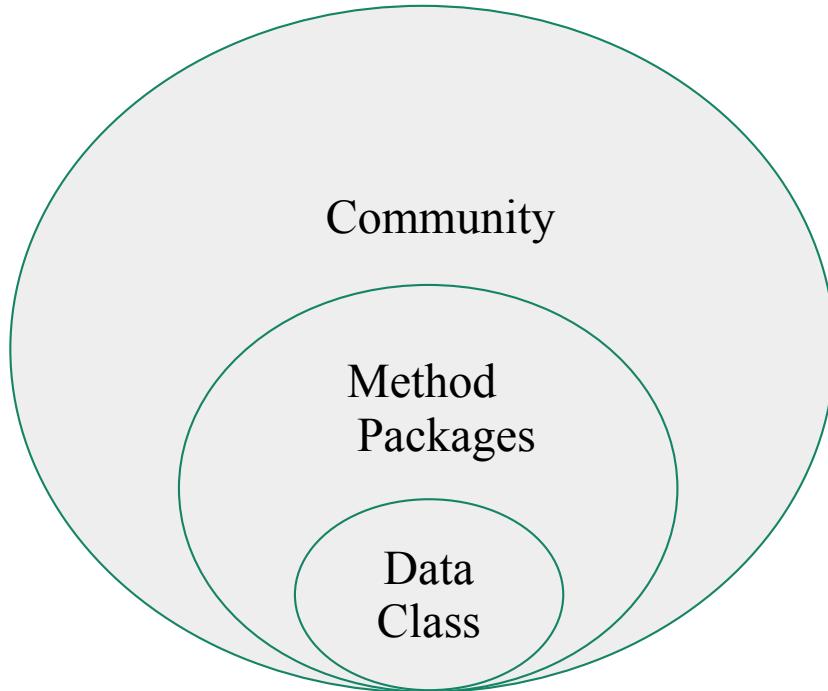
Ruihu Huang

Domenick J. Braccia

Héctor Corrada Bravo

Leo Lahti

Reduce overlapping efforts, improve interoperability, ensure sustainability.



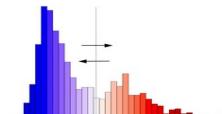
Data packages

ExperimentHub

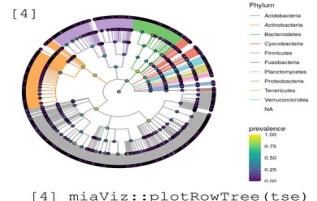
platforms all rank 76 / 1974 posts 2 / 1 / 2e+01 / 1 in Bioc 4 years
build ok updated before release dependencies 72

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

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



miaViz - Visualization



Package ecosystem

1. Ampvis2 Tools for visualising amplicon sequencing data
2. CCREEPE 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. MMint Methodology for the large-scale assessment of microbial metabolic interactions (MMint) 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. qilmer R tools compliment qlime
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](https://github.com/microsud/Tools-Microbiome-Analysis)

 Springer Link

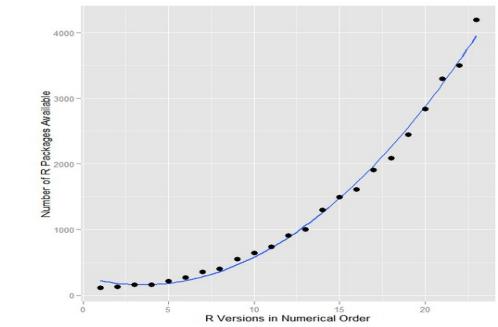
[Journal of Biosciences](#)
October 2019, 44:115 | [Cite as](#)



Microbiome data science

Authors [Authors and affiliations](#)

Sudarshan A Shetty, Leo Lahti 



microbiome R package

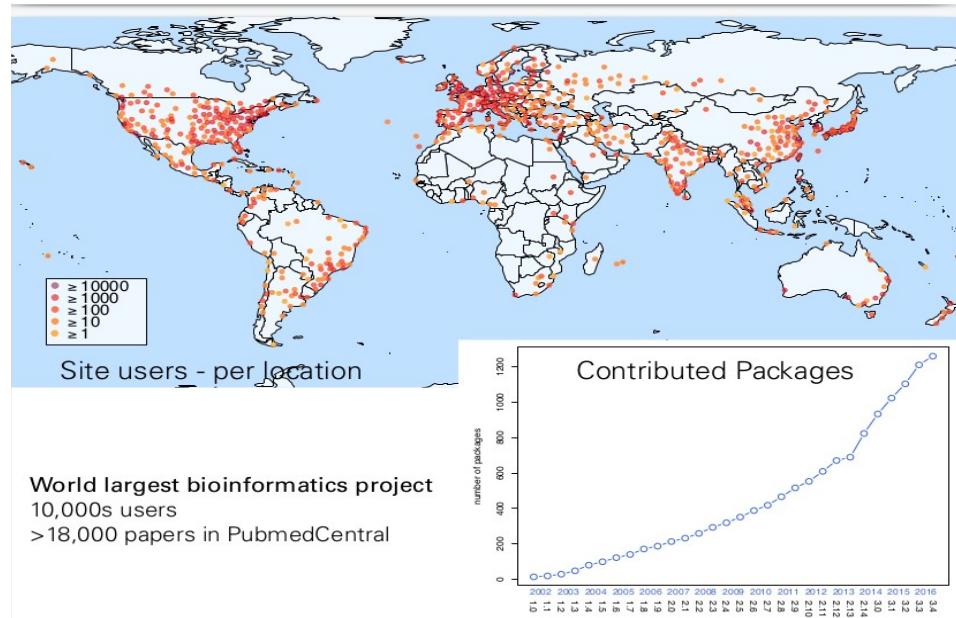
[chat](#) [on gitter](#) [build](#) [passing](#) [codecov](#) [24%](#) [PRs](#) [welcome](#)



Started in 2001 as a platform for microarray data analysis

Now 2140+ packages:

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

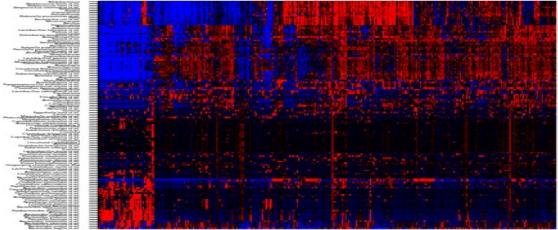


Microbiome data container

TreeSummarizedExperiment (Huang et al. 2021)

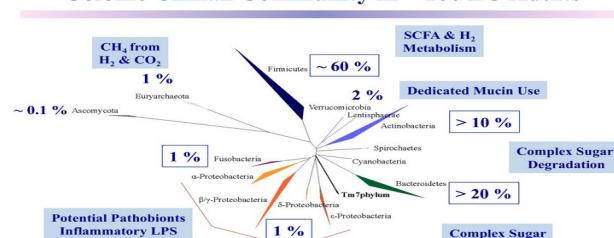
Taxonomic features

Individuals

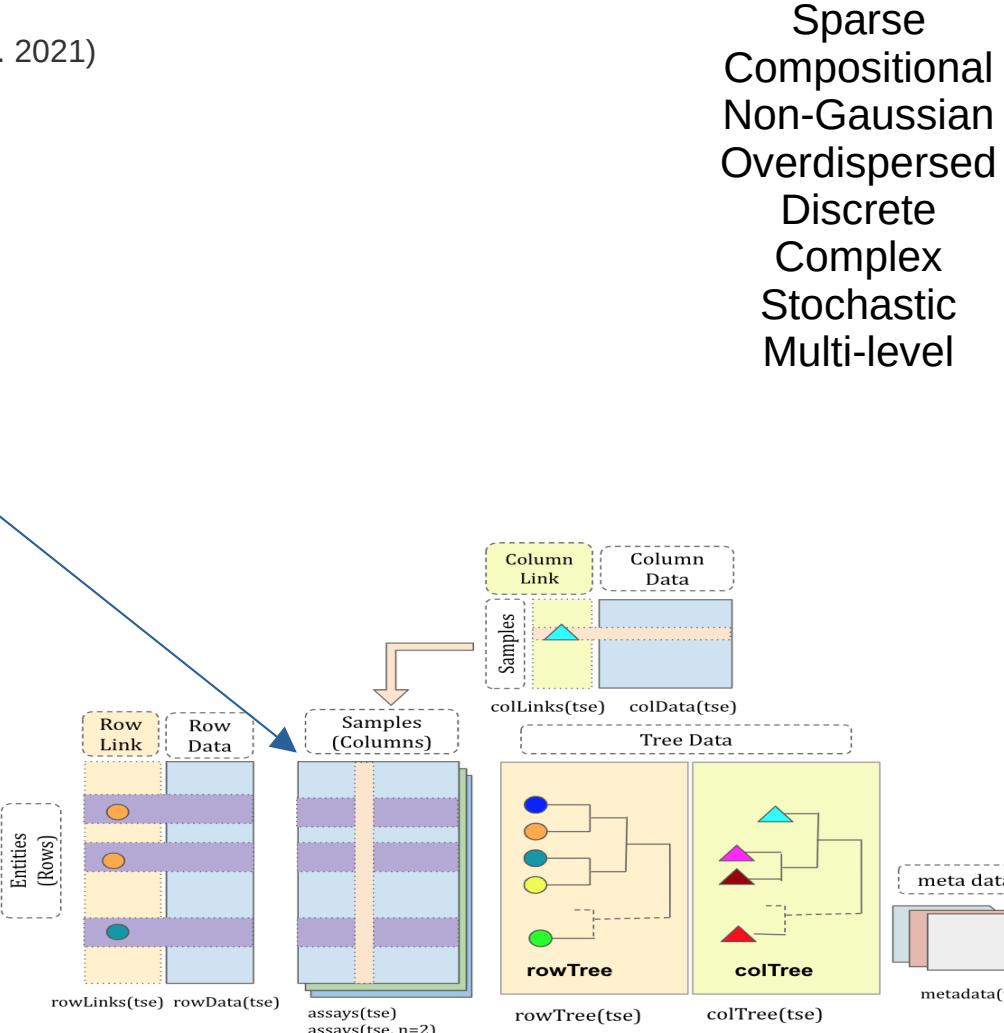


Gut microbiota: 1000
western adults
(Lahti et al. Nature
Comm. 2014)

Colonic Climax Community in ~ 100 EU Adults



Zoetewij EG, EE Vaughan & WM de Vos (2006) *Int Microbiol* 9: 1639
Lay C, I Rigolet-Guis, K Holtenius, M Rajile, EE Vaughan, WM de Vos, MD Collins, R Thiel, P Nusseleck, M Blaut & J Dore (2005) *AEM* 71: 4183



Example / demonstration data sets

in *TreeSummarized* data container format

- R package data (*mia*, *miaViz*, *miaTime*)
- *curatedMetagenomicData*
- *microbiomeDataSets*

curatedMetagenomicData

platforms all rank 30 / 408 support 1 / 1 build ok
updated < 1 month dependencies 155

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

Curated Metagenomic Data of the Human Microbiome

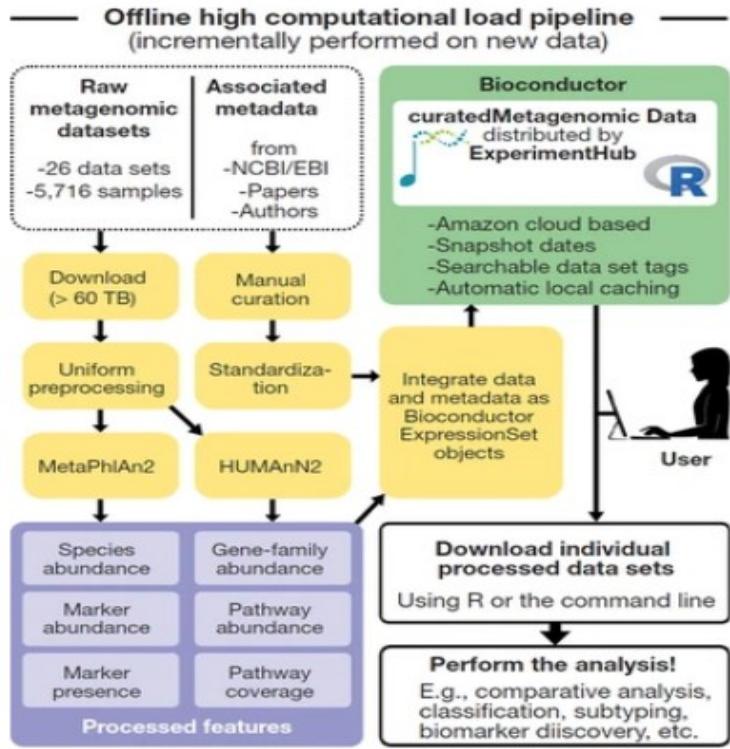
microbiomeDataSets

platforms all rank 99 / 408 support 0 / 0 build ok
updated before release dependencies 113

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

Experiment Hub based microbiome datasets

curatedMetagenomicData provides standardized, public data sets for microbiome analysis in R/Bioc TreeSummarizedExperiment

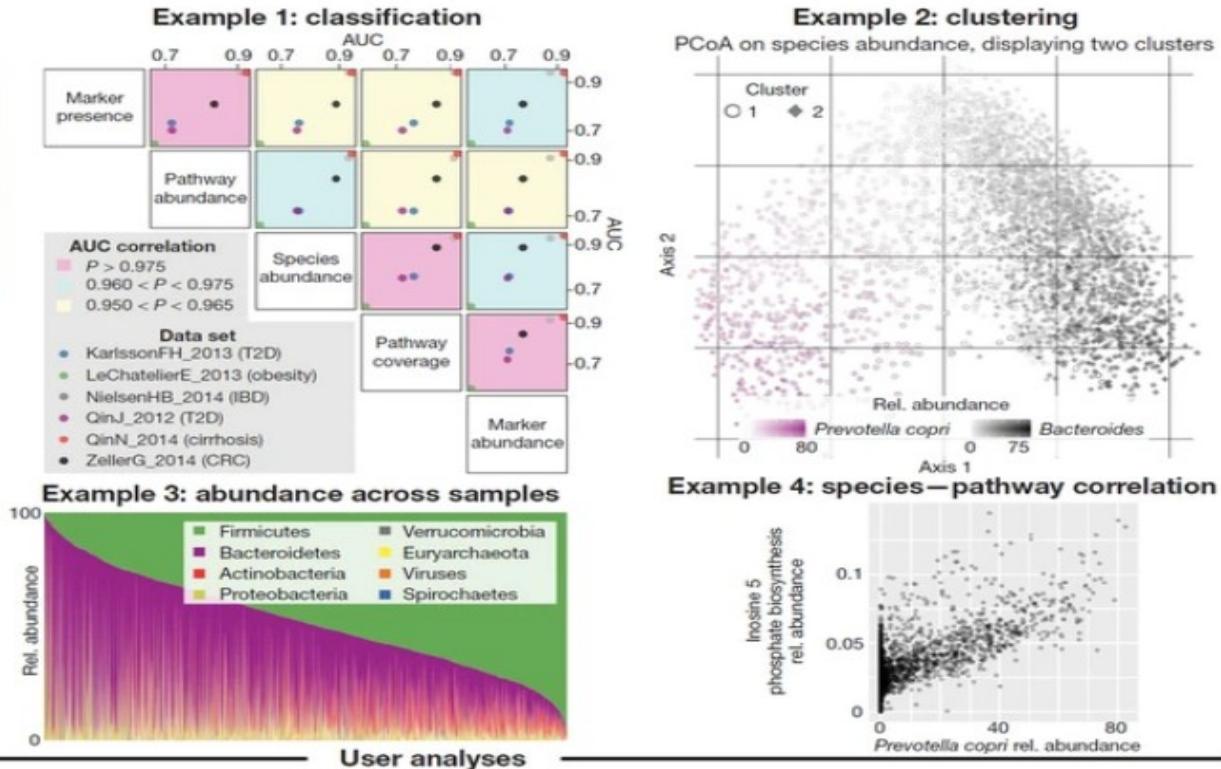


Accessible, curated metagenomic data through ExperimentHub

[Edoardo Pasolli](#), [Lucas Schiffer](#), [Paolo Manghi](#), [Audrey Renson](#), [Valerie Obenchain](#), [Duy Tin Truong](#), [Francesco Beghini](#), [Faizan Malik](#), [Marcel Ramos](#), [Jennifer B Dowd](#), [Curtis Huttenhower](#), [Martin Morgan](#), [Nicola Segata](#) & [Levi Waldron](#)

[Nature Methods](#) 14, 1023–1024 (2017) | [Cite this article](#)

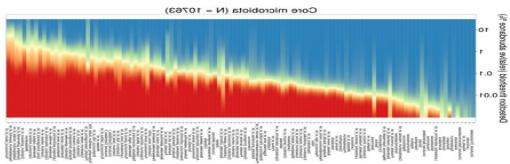
5710 Accesses | 103 Citations | 29 Altmetric | [Metrics](#)



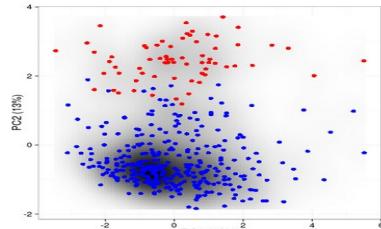
Interoperable packages for microbial community analysis

Standard manipulation: mia

```
transformSamples(x, "compositional")
getPrevalentFeatures(x)
splitByRanks(x)
estimateDiversity(x)
```



Visualization: miaViz



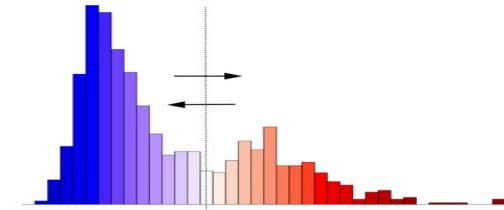
Community

- Online tutorials
- Gitter chat
- Workshops

Broader Bioconductor ecosystem:

- RNA-seq (e.g. DESeq2)
- Single cell (e.g. scater)
- etc.

Time series: miaTime



Community simulation: miaSim

Quality control

- continuous integration
- unit tests
- open source

Microbiome data import from standard formats (*mia*)

tse <- loadFromMetaphlan(file)

```
makeTreeSEFromPhyloseq()  
makeTreeSummarizedExperimentFromPhyloseq()
```

Coerce a `phyloseq` object to a `TreeSummarizedExperiment`

```
makeTreeSEFromDADA2()  
makeTreeSummarizedExperimentFromDADA2()
```

Coerce ‘DADA2’ results to `TreeSummarizedExperiment`

```
loadFromBiom() makeTreeSEFromBiom()  
makeTreeSummarizedExperimentFromBiom()
```

Loading a biom file

```
loadFromQIIME2() readQZA()
```

Import QIIME2 results to `TreeSummarizedExperiment`

```
loadFromMothur()
```

Import Mothur results as a `TreeSummarizedExperiment`

```
loadFromMetaphlan
```

Import Metaphlan results to `TreeSummarizedExperiment`

```
makePhyloseqFromTreeSE()  
makePhyloseqFromTreeSummarizedExperiment()
```

Create a `phyloseq` object from a `TreeSummarizedExperiment` object

Data manipulation: *mia*

mia 1.5.17 Get started Reference

mia - Microbiome analysis

This project provides functions and workflows examples for analyses of microbiome data. The main class for working with microbiome data in this package is `TreeSummarizedExperiment`.

Currently following things are implemented:

- data wrangling functions (`agglomerate*`, `merge*`, and more)
- CCA analysis via `vegan` package
- Bray-Curtis dissimilarity via `vegan` package
- JSD and UniFrac distance calculation ported from `phyloseq` to work with `TreeSummarizedExperiment` objects
- MDS via the `scater` package for any other distance objects
- import functions for `biom` data, `DADA2` objects, `phyloseq` objects and more

Contribution

Contributions are welcome in the form of feedback, issues, pull requests etc, see [contributor guidelines](#).

Technical aspects

Let's use a git flow kind of approach. Development version should be done against the `master` branch and then merged to `release` for release. (<https://guides.github.com/introduction/flow/>)

Installation

Bioc-release

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("mia")
```

Agglomeration

```
agglomerateByRank(<SummarizedExperiment>)
agglomerateByRank(<SingleCellExperiment>)
agglomerateByRank(<TreeSummarizedExperiment>)

mergeRows() mergeCols()
splitByRanks() unsplitByRanks()

splitOn() unsplitOn()

mergeSEs() full_join() inner_join()
left_join() right_join()
```

Agglomerate data using taxonomic information

Merge a subset of the rows or columns of a `SummarizedExperiment`

Split/Unsplit a `SingleCellExperiment` by taxonomic ranks

Split `TreeSummarizedExperiment` column-wise or row-wise based on grouping variable

Merge SE objects into single SE object.

Links

[View on Bioconductor](#)

[Browse source code](#)

[Report a bug](#)

License

[Artistic-2.0](#) | file [LICENSE](#)

Community

[Contributing guide](#)

Citation

[Citing mia](#)

Developers

Felix G.M. Ernst

Author 

Sudarshan A. Shetty

Author 

Tuomas Borman

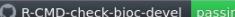
Author, maintainer 

Leo Lahti

Author 

[More about authors...](#)

Dev status

 passing

 76%

Transformation

```
transformSamples() transformCounts()
transformFeatures() ZTransform()
relAbundanceCounts()
```

Transform Counts

Data loading

```
makeTreeSEFromPhyloseq()
makeTreeSummarizedExperimentFromPhyloseq()

makeTreeSEFromDADA2()
makeTreeSummarizedExperimentFromDADA2()

loadFromBiom() makeTreeSEFromBiom()
makeTreeSummarizedExperimentFromBiom()

loadFromQIIME2() readQZA()
loadFromMothur()
loadFromMetaphlan
makePhyloseqFromTreeSE()
makePhyloseqFromTreeSummarizedExperiment()
```

Coerce a `phyloseq` object to a `TreeSummarizedExperiment`

Coerce 'DADA2' results to `TreeSummarizedExperiment`

Loading a biom file

Import QIIME2 results to `TreeSummarizedExperiment`

Import Mothur results as a `TreeSummarizedExperiment`

Import Metaphlan results to `TreeSummarizedExperiment`

Create a `phyloseq` object from a `TreeSummarizedExperiment` object

Taxonomic data

```
TAXONOMY_RANKS taxonomyRanks()
taxonyRankEmpty() checkTaxonomy()
getTaxonomyLabels() taxonomyTree()
addTaxonomyTree() mapTaxonomy()
IdTaxaToDataFrame()
```

Functions for accessing taxonomic data stored in `rowData`.

Diversity indeces

Alpha Diversity

```
estimateDiversity() estimateFaith()
estimateDivergence()
```

Estimate (alpha) diversity measures

Estimate divergence

Beta Diversity

```
calculateUnifrac() runUnifrac()

calculateJSD(<ANY>)
calculateJSD(<SummarizedExperiment>)
runJSD()

calculateNMDS() runNMDS() plotNMDS()
```

Calculate weighted or unweighted (Fast) Unifrac distance

Calculate the Jensen-Shannon Divergence

Perform non-metric MDS on sample-level data

mia: example methods

Compositional (Aitchison) transformations:

```
x <- transformSamples(tse,  
                      method = "clr",  
                      pseudocount = 1)
```

Takes advantage on *vegan* transformations.

Phylogeny-aware distance metrics:

```
calculateUnifrac(tse, weighted = TRUE)
```

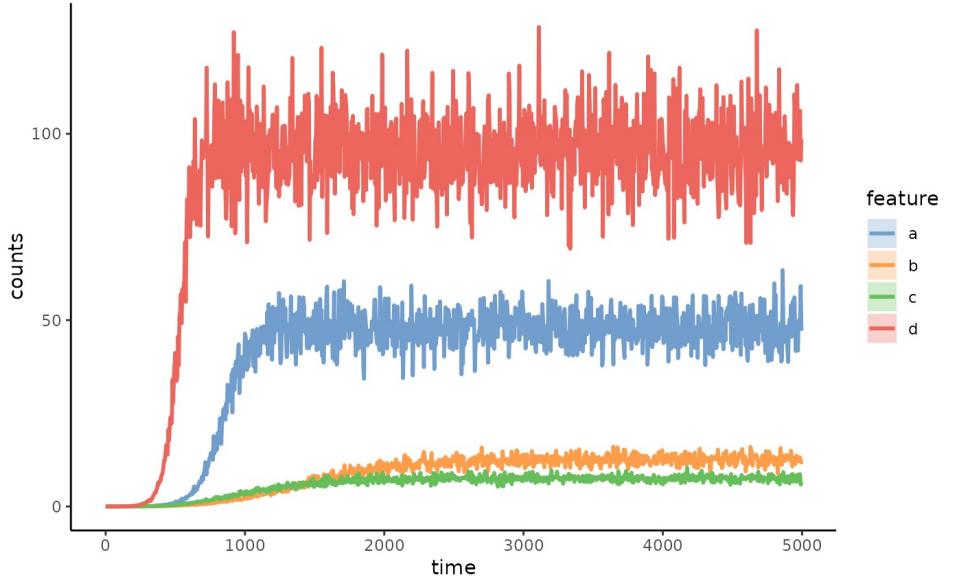
Unifrac method by Rob Knight. Initial R/Bioc design for *phyloseq* Paul J. McMurdie. Adapted for *mia/TreeSummarizedExperiment* by Felix G.M. Ernst

Microbiome data simulation: *miaSim*

Ongoing: /w Daniel Garza, Yagmur Simsek, Yu Gao, Emma Gheysen, Karoline Faust

- Hubbell's neutral model
- Generalized Lotka-Volterra
- Self-organised instability

```
ExampleCR <- simulateConsumerResource(
  n_species = n_species,
  n_resources = n_resources, names_species = letters[1:n_species],
  names_resources = paste0("res", LETTERS[1:n_resources]), E = ExampleE,
  x0 = rep(0.001, n_species), resources = ExampleResources,
  growth_rates = runif(n_species),
  error_variance = 0.01,
  t_end = 5000,
  t_step = 1
)
ExampleCR_SE <- TreeSummarizedExperiment(
  assays = list(counts = t(ExampleCR$matrix[, 1:n_species])),
  colData = DataFrame(time = ExampleCR$matrix[, "time"]),
  metadata = ExampleCR[-which(names(ExampleCR) == "matrix")]
)
miaViz::plotSeries(ExampleCR_SE, x = "time")
```



Microbiome time series analysis: miaTime

Minimal gut bioreactor example by Sudarshan Shetty

- Individual time series
- Community divergence

Community dynamics

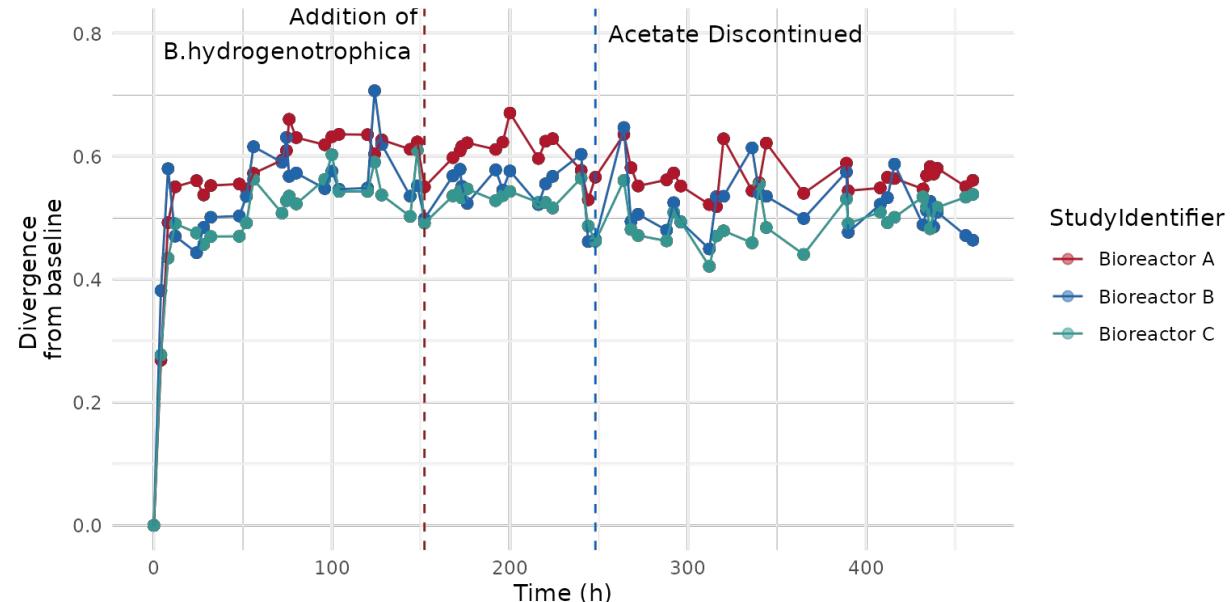
The `minimalgut` dataset, mucus-diet based minimal microbiome (MDbMM-16), consists of 16 species assembled in three bioreactors. We can investigate the succession of mdbMM16 from the start of experiment here hour zero until the end of the experiment.

```
## Divergence from baseline i.e from hour zero.
tse <- mia::relAbundanceCounts(minimalgut) # get relative abundance
tse <- getBaselineDivergence(tse,
  group = "StudyIdentifier",
  time_field = "time.hr",
  name_divergence = "divergence_from_baseline",
  name_timedifference = "time_from_baseline",
  assay_name="rrelabundance",
  FUN = vegan::vegdist,
  method="bray")
```

Visualize the divergence

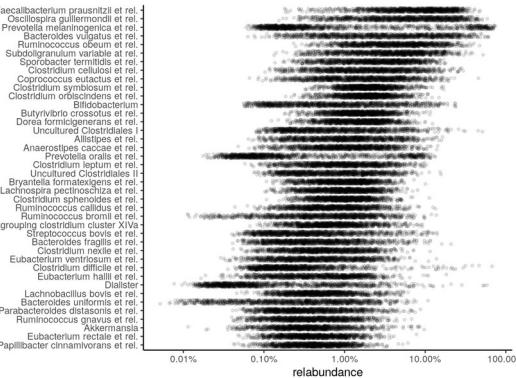
```
# First define nice colors for bioreactors
bioreac_cols <- c(`Bioreactor A'= "#b2182b",
  `Bioreactor B'= "#2166ac",
  `Bioreactor C' = "#35978f")

tse |>
  ggplot(aes(x=Time.hr, y=divergence_from_baseline)) +
  geom_point(aes(color=StudyIdentifier), size=2, alpha=0.5) +
  geom_line(aes(color=StudyIdentifier)) +
  theme_minimal() +
  scale_color_manual(values = bioreac_cols) +
  labs(x="Time (h)", y="Divergence \nfrom baseline") +
  # highlight specific timepoints
  geom_vline(xintercept = 152, lty=2, color="#991720") +
  geom_vline(xintercept = 248, lty=2, color="#0963bd") +
  annotate("text",x=c(152, 248),y=c(0.8, 0.8),
    label=c("Addition of\nB.hydrogenotrophica","Acetate Discontinued"),
    hjust=c(1.05,-0.05))
```

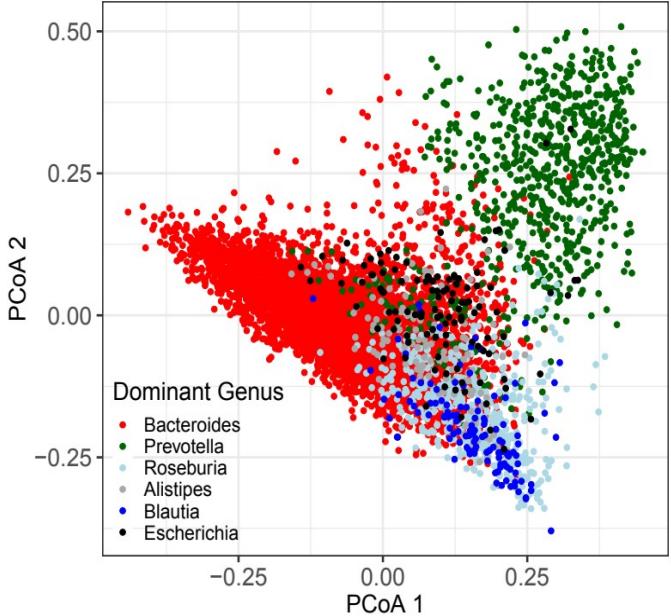
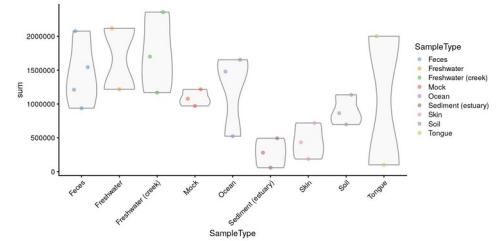
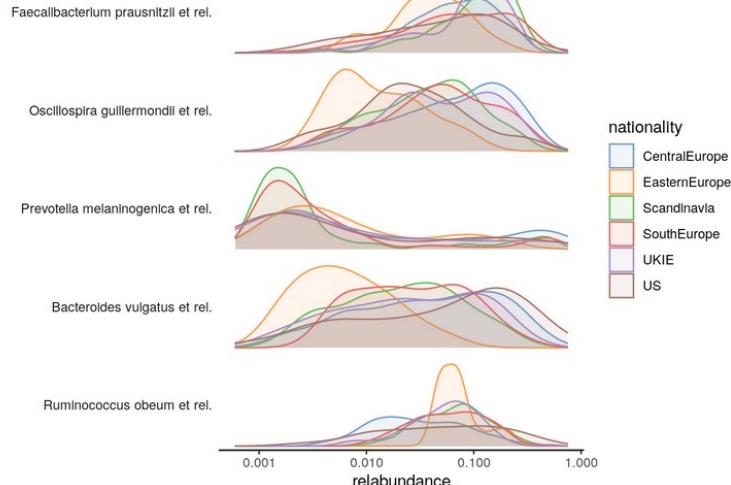


Microbiome data visualization: *miaViz*

```
x <- altExp(GlobalPatterns,"Genus")
plotRowTree(x[rownames(x) %in% top_genus],
tip_colour_by = "log_mean",
tip_size_by = "detected")
```



```
plotAbundanceDensity(tse, layout = "density", abund_values = "relabundance",
n = 5, colour_by="nationality", point_alpha=1/10) +
scale_x_log10()
```



Taxonomic Signatures of Long-Term Mortality Risk in Human Gut Microbiota

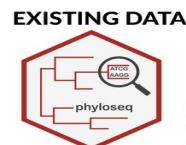
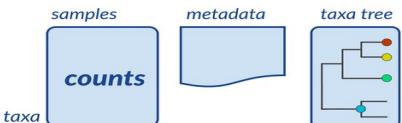
● Aaro Saloenssaari, ● Ville Laitinen, ● Aki Havulinna, Guillaume Meric, ● Susan Cheng, ● Markus Perola, Liisa Valsta, ● Georg Althan, ● Michael Inouye, Jeramie D. Watrous, Tao Long, ● Rodolfo Salido, Karenina Sanders, Caitriona Brennan, Gregory C. Humphrey, Jon G. Sanders, ● Mohit Jain, Pekka Jousilahti, ● Veikko Salomaa, ● Rob Knight, ● Leo Lahti, ● Teemu Niiranen
doi: <https://doi.org/10.1101/2019.12.30.19015842>

Open, interoperable data & tools for microbiome analysis workflows

Import Data

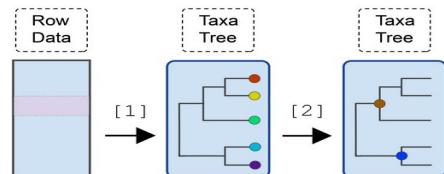
This workflow starts with either raw data directly from relative abundance estimation or taxonomic classification OR pre-existing data objects from widely used software.

RAW DATA

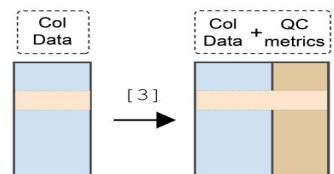


The mia Pipeline

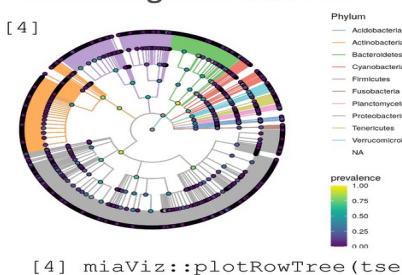
Accessing Taxonomic Info.



Quality Control



Visualizing with miaViz



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

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

European Bioconductor Meeting 2020

- Where: Virtual Conference
- When: 14-18 December 2020
- On twitter: #EuroBioC2020



Special thanks

Felix G.M. Ernst
Sudarshan A. Shetty

Tuomas Borman

RuiZhu Huang

Domenick J. Braccia

Héctor Corrada Bravo

Leo Lahti

Acknowledgments

Course organizers:

- Finnish IT Center for Science (CSC)
- Department of Computing, University of Turku, Finland

Material preparation supported by:



Turun yliopisto
University of Turku



BIOCITY TURKU
BIOCITY TURKU
BIOCITY TURKU



SUOMEN AKATEMIA
FINLANDS AKADEMI • ACADEMY OF FINLAND



MICROBIOME

