

Microbiome data containers in R/Bioconductor

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



Overview

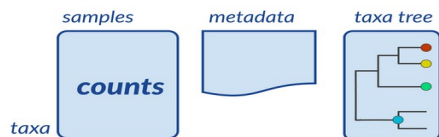
- Rationale for data containers
- Microbiome data containers

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.

RAW DATA

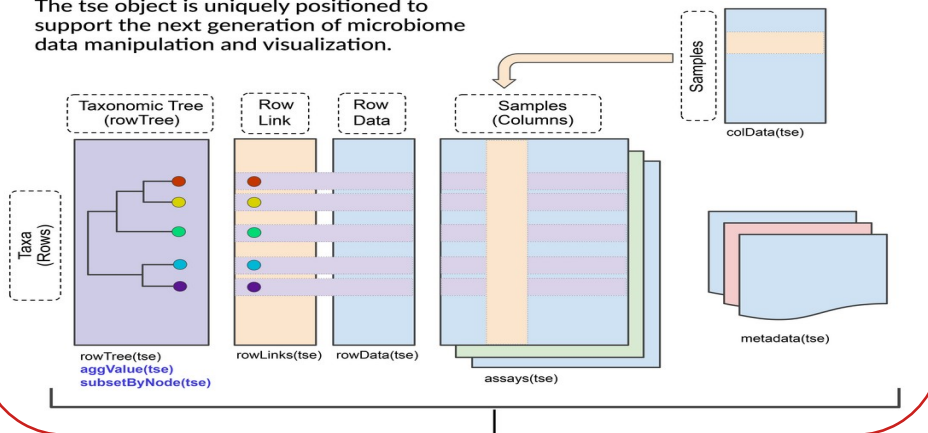


EXISTING DATA



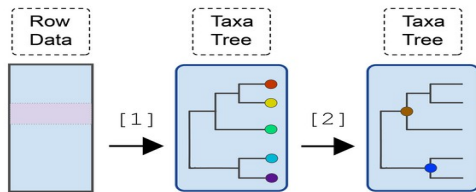
The TreeSE object

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



The mia Pipeline

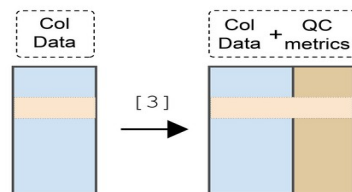
Accessing Taxonomic Info.



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

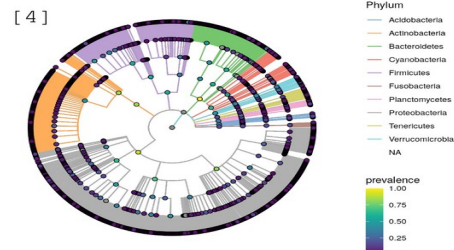
```
[2] TreeSE::aggValue(tse)
```

Quality Control



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

Visualizing with miaViz

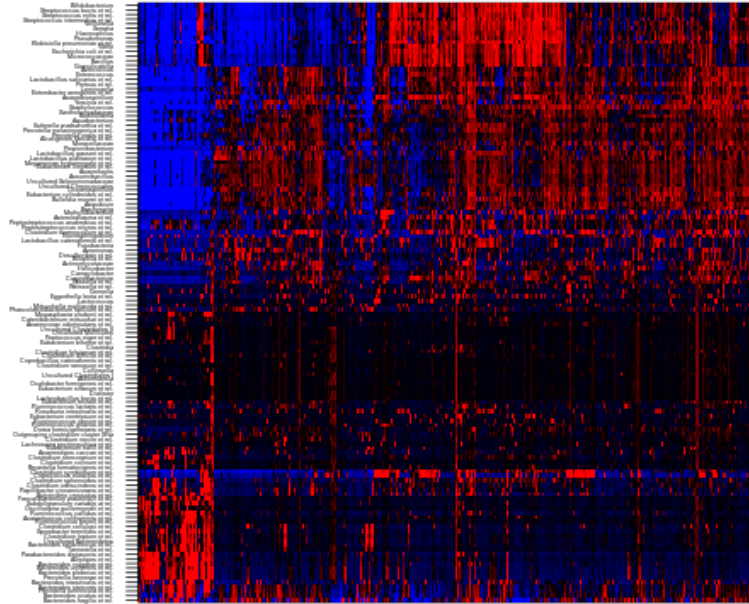


```
[4] miaViz::plotRowTree(tse)
```

Fig: Domenick Braccia, EuroBioC 2020

"Omics" data

taxonomic abundance table
(features x samples)



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

Each data type has its own special properties

- Genomics
- Epigenomics
- Microbiomics
- Lipidomics
- Proteomics
- Glycomics
- Foodomics
- Transcriptomics
- Metabolomics
- Culturomics

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



[+ Author & Article Information](#)

Cancer Res (2017) 77 (21): e39–e42.

<https://doi.org/10.1158/0008-5472.CAN-17-0344> [Article history](#)

- (Tree)SummarizedExperiment
- RangedSummarizedExperiment
- MultiAssayExperiment
- SingleCellExperiment
- SpatialExperiment

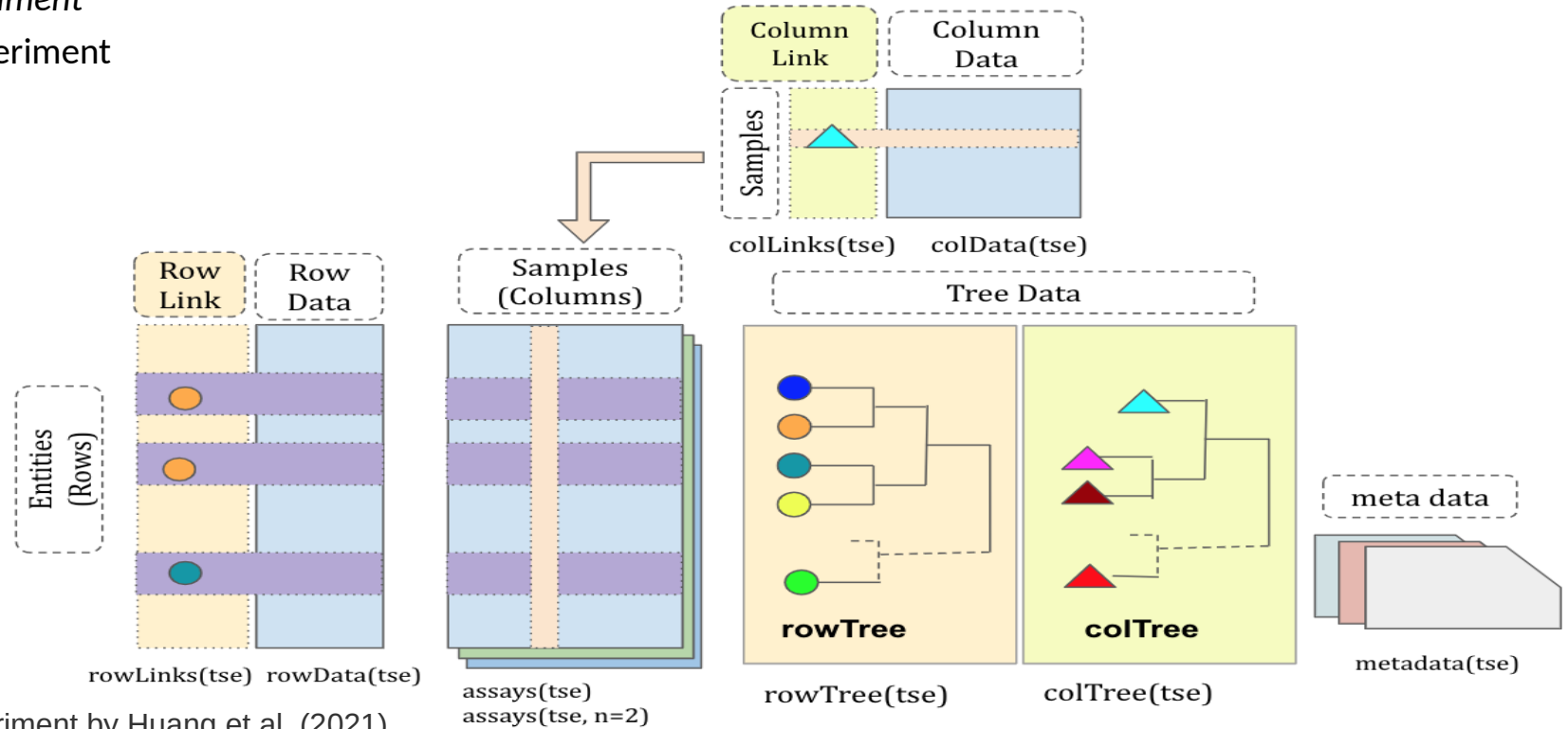


Fig: TreeSummarizedExperiment by Huang et al. (2021)

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

F1000Research

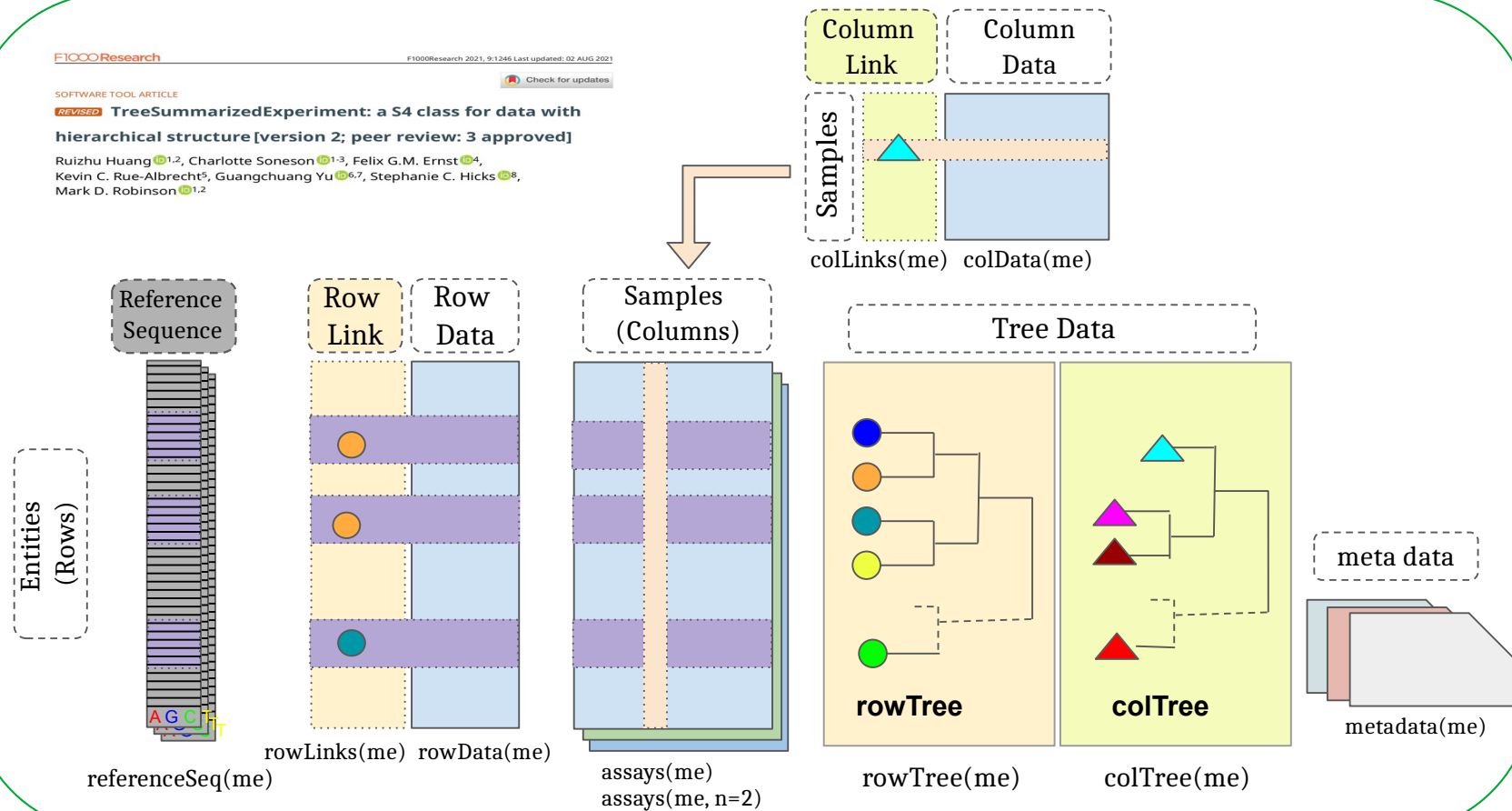
F1000Research 2021, 9:1246 Last updated: 02 AUG 2021

Check for updates

SOFTWARE TOOL ARTICLE

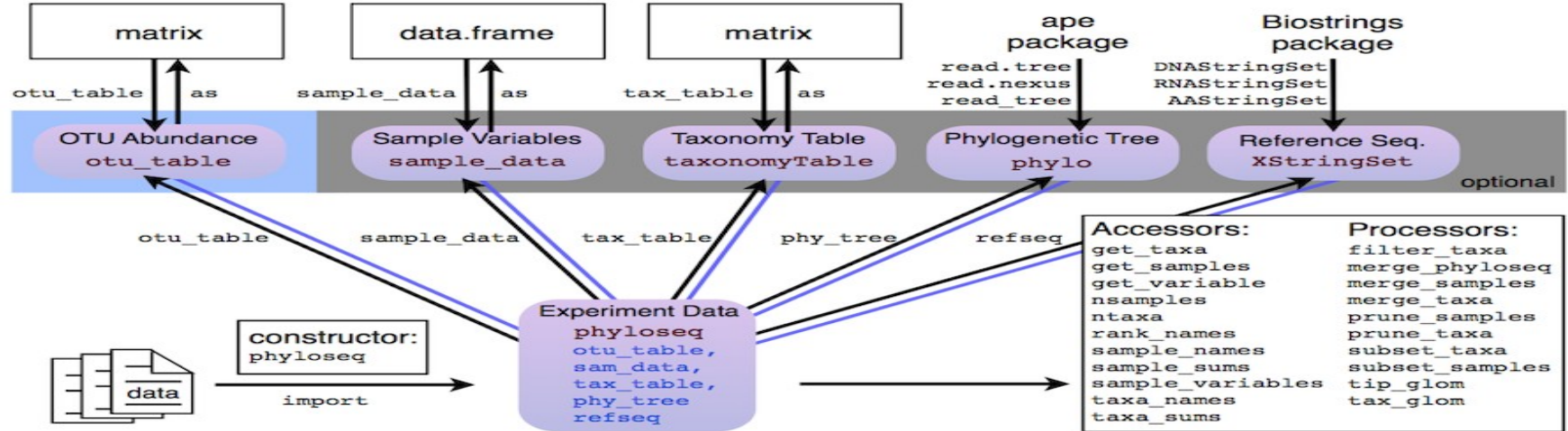
REVISED TreeSummarizedExperiment: a S4 class for data with hierarchical structure [version 2; peer review: 3 approved]

Ruizhu Huang^{1,2}, Charlotte Soneson^{1,3}, Felix G.M. Ernst⁴, Kevin C. Rue-Albrecht⁵, Guangchuang Yu^{6,7}, Stephanie C. Hicks⁸, Mark D. Robinson^{1,2}




Alternative data container: *phyloseq*

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



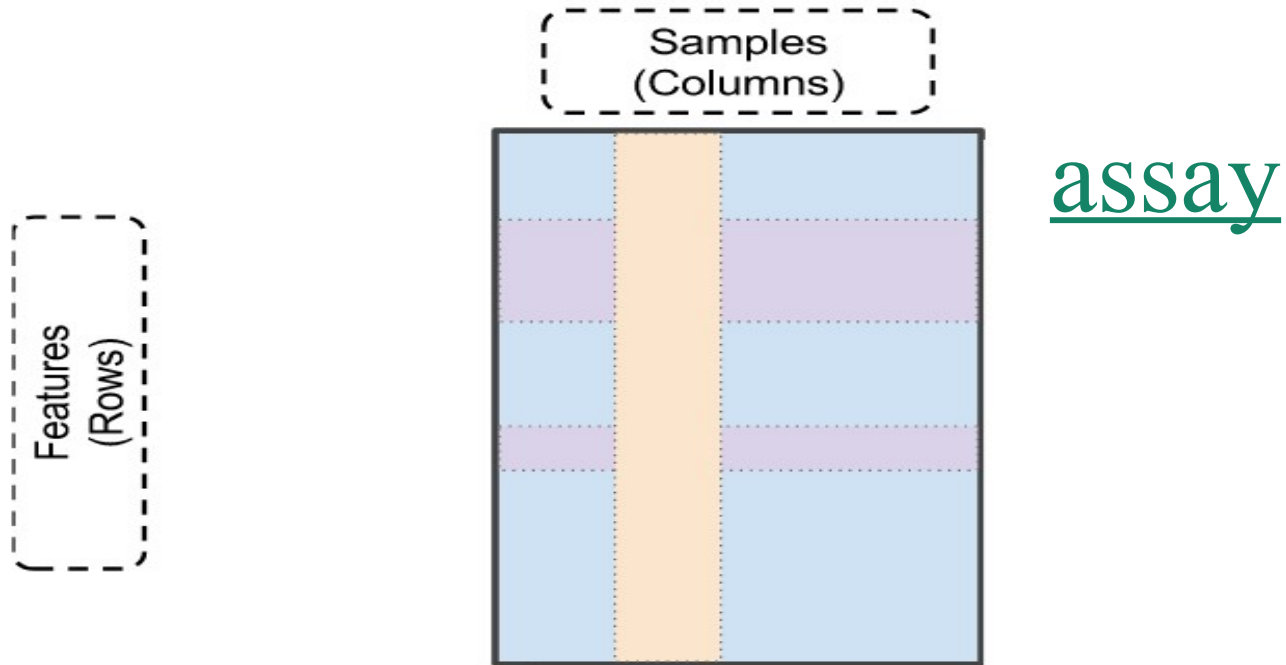
REVISED Bioconductor Workflow for Microbiome Data
Analysis: from raw reads to community analyses
[version 2; peer review: 3 approved]

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

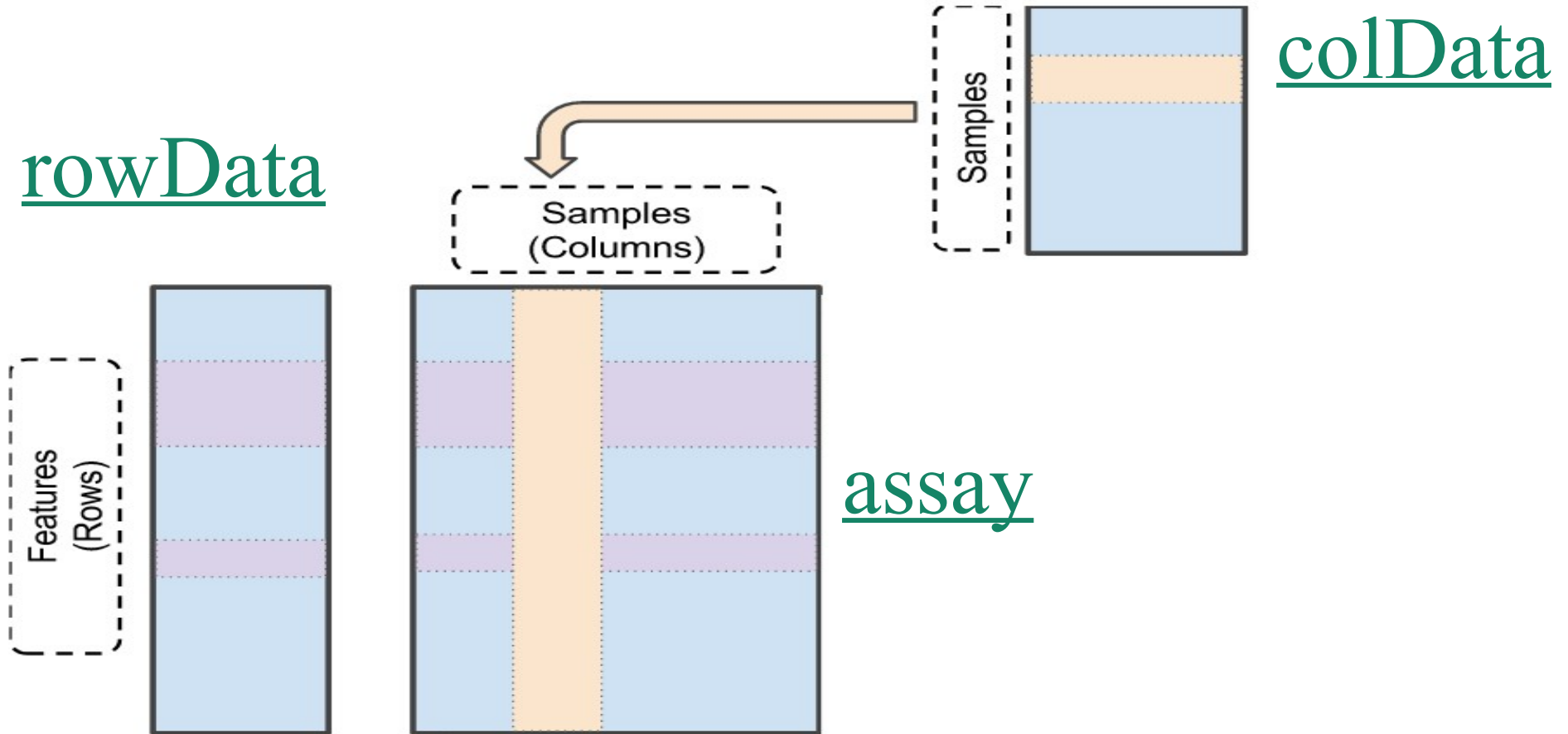


This article is included in the [Bioconductor](#) gateway.

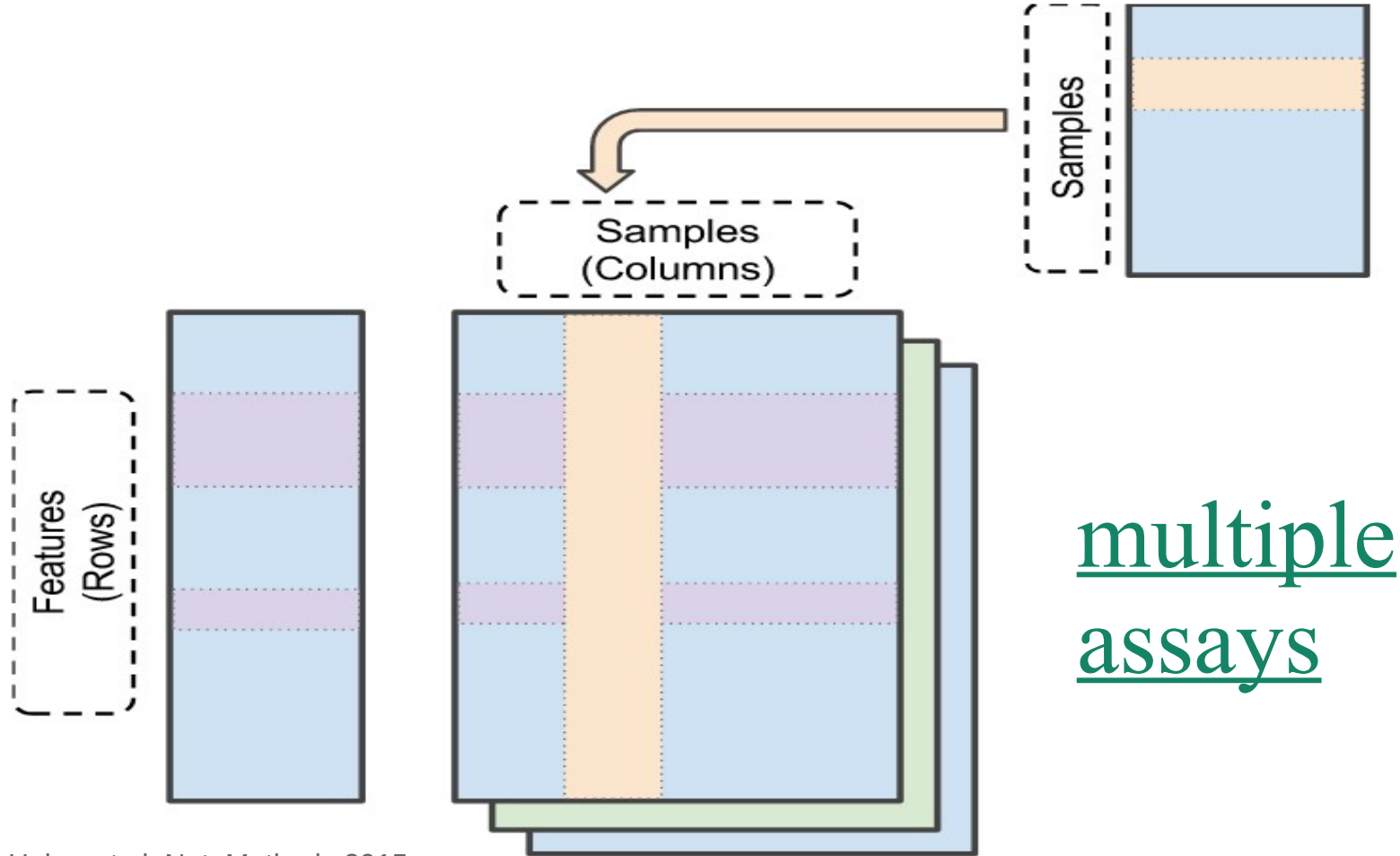
SummarizedExperiment



SummarizedExperiment



SummarizedExperiment



Integrating parallel data sources

Option	Rows (features)		Cols (samples)	Recommended
assays	match		match	Data transformations
altExp	free		match	Alternative experiments
MultiAssay	free		free (mapping)	Multi-omic experiments

Data operations

Abundance data: `assay(tse, "counts")`

Feature information: `rowData(tse)`

Sample information: `colData(tse)`

Alternative experiments: `altExp(tse, "Genus")`

TreeSummarizedExperiment

F1000Research

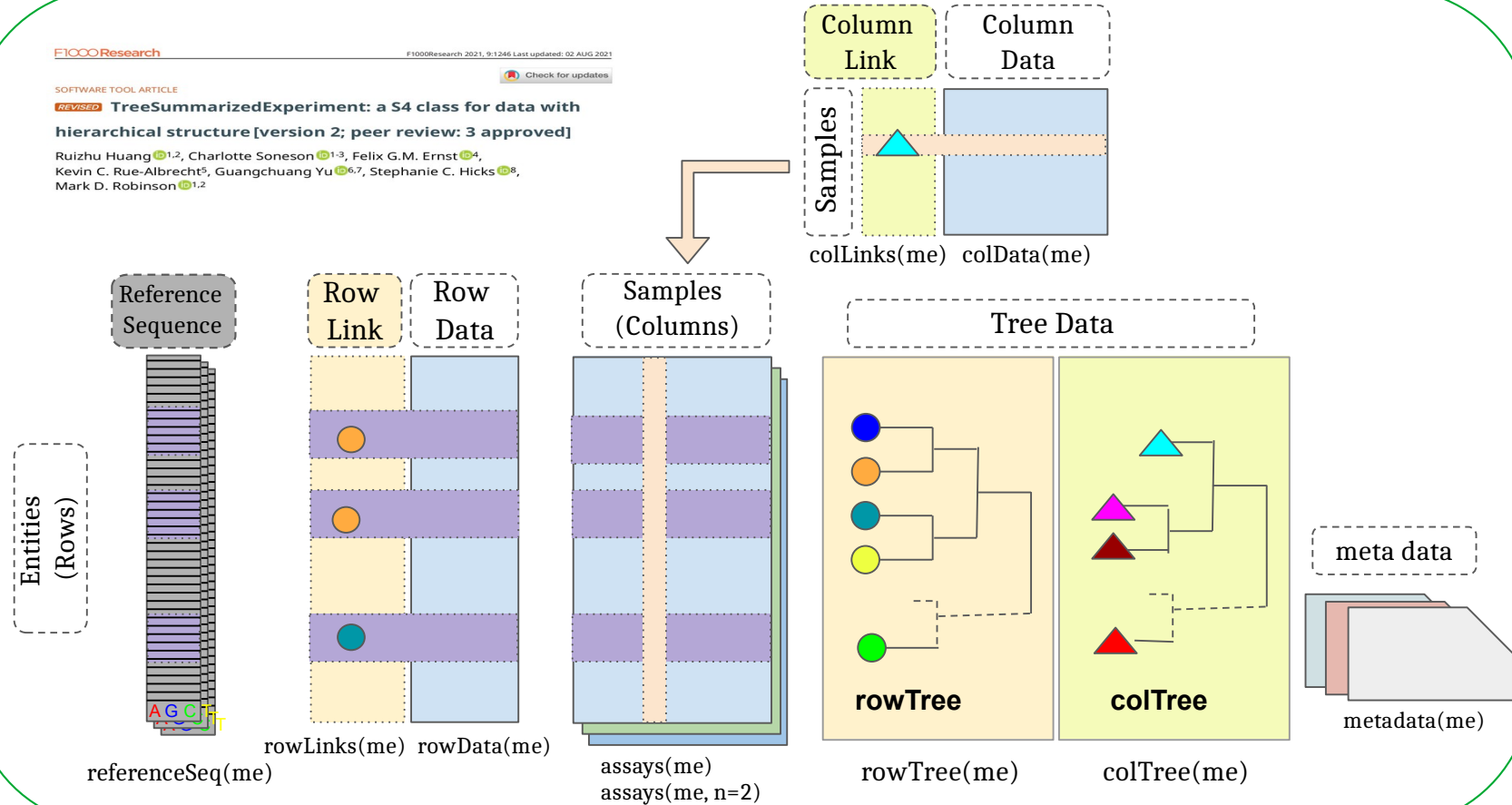
F1000Research 2021, 9:1246 Last updated: 02 AUG 2021

Check for updates

SOFTWARE TOOL ARTICLE

TreeSummarizedExperiment: a S4 class for data with hierarchical structure [version 2; peer review: 3 approved]

Ruizhu Huang^{1,2}, Charlotte Soneson^{1,3}, Felix G.M. Ernst⁴, Kevin C. Rue-Albrecht⁵, Guangchuang Yu^{6,7}, Stephanie C. Hicks⁸, Mark D. Robinson^{1,2}



Summary

Data containers:

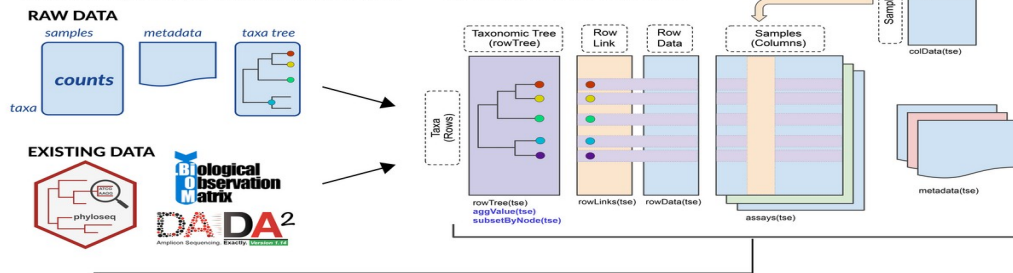
- Standardized data presentations
- Optimized techniques
- Key part of the data science workflow

They facilitate:

- Transparent analysis
- Reproducible research
- Interoperability of tools
- Usability & documentation
- Collaboration

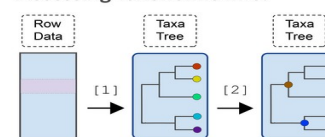
Import Data

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 mia Pipeline

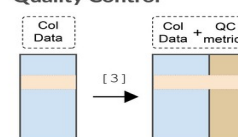
Accessing Taxonomic Info.



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

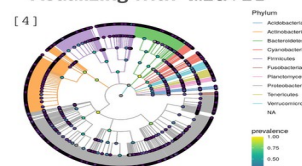
[2] `TreeSE::aggValue(tse)`

Quality Control



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

Visualizing with miaViz



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

