

Microbiome data science

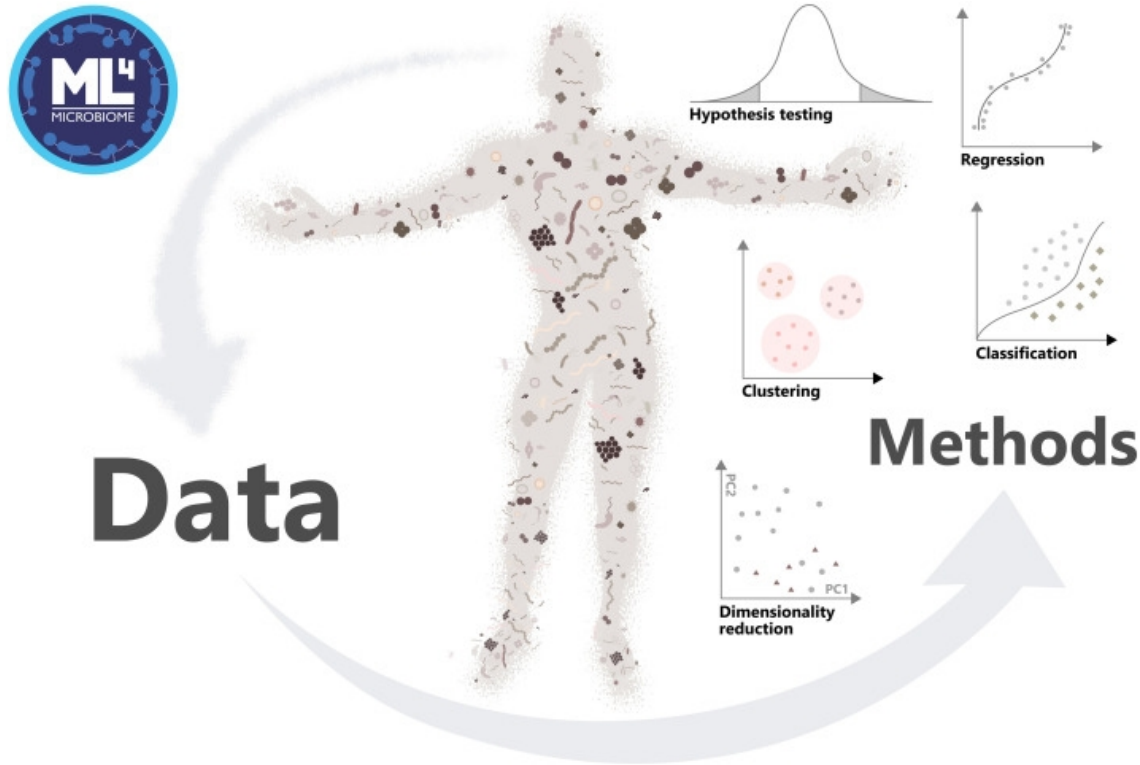


Figure source: Moreno-Indias et al. (2021)
Statistical and Machine Learning Techniques
in Human Microbiome Studies: Contemporary
Challenges and Solutions. URL:
<https://doi.org/10.3389/fmicb.2021.635781>.
Frontiers in Microbiology 12:11.

Associate Prof. Leo Lahti | datascience.utu.fi

Department of Computing, University of Turku, Finland

 @antagomir

Acknowledgments

Course organizers

- Mirjam Bloemendaal

Course assistants

- Henrik Eckermann
- Tuomas Borman
- Chouaib Benchraoui

Miaverse collective

- Sudarshan Shetty
- Felix Ernst

Learning goals

Key concepts in microbiome bioinformatics

Common data science tools & best practices

Overview of the week

Key data analytical concepts

- Intro to microbiome data science
- Alpha diversity
- Beta diversity
- Differential abundance analysis (Henrik)

Conclusion

- Course work presentations

Monday: from raw sequences to ecological data analysis

Lectures (by Tom)

Microbiota analysis: association studies vs. causality; microbiota sequencing methods (16S, shotgun, metagenomics)

DNA isolation and 16S rRNA gene sequencing: from raw sequences to OTU table in a biom file

Getting started

Prerequisites

Data generation and import

Open science methods

Practical

Importing microbiome data to R for interactive data analysis

Initializing reproducible report with Rmarkdown

Getting started

- Checklist
- R, R studio, R packages installation ok?
- additional reading
- online chat (Gitter & Slack)
- course assignment

Course assignment – case study options

Option 1 ADHD-associated changes in gut microbiota and brain in a mouse model

Option 2 Western diet intervention study in serotonin transporter modified mouse model

Mice received three weeks of Western (high fat, high sugar) or Control diet in three Genotype groups, a wild-type group and 2 groups with either partial or complete knock-out of the serotonin transporter (SERT), leading to an excess of serotonin in the gut, brain and blood circulation. Mice were measured before and after the intervention. In total there are three factors in this design, Time (pre-,post-intervention), Diet (Control, Western diet) and Genotype group (WT, HET, KO). There are about 6-8 mice per group.

Option 3 Open data set of your own choice, see e.g. Bioconductor ExperimentHub

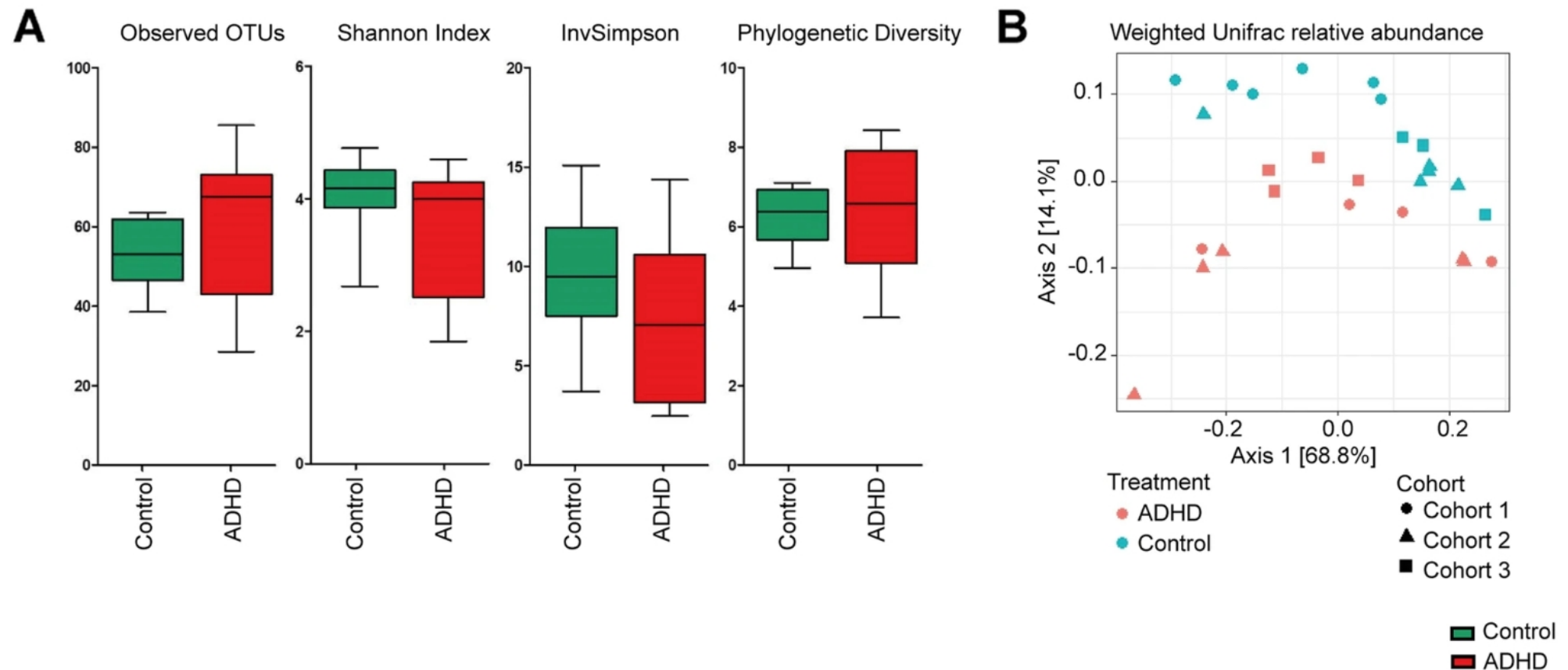
- microbiomeDataSets
- curatedMetagenomicData

Gut microbiota from persons with attention-deficit/hyperactivity disorder affects the brain in mice

Anouk C. Tengeler, Sarita A. Dam, Maximilian Wiesmann, Jilly Naaijen, Miranda van Bodegom, Clara Belzer, Pieter J. Dederen, Vivienne Verweij, Barbara Franke, Tamas Kozicz, Alejandro Arias Vasquez & Amanda J. Kiliaan ✉

Microbiome 8, Article number: 44 (2020) | [Cite this article](#)

From: [Gut microbiota from persons with attention-deficit/hyperactivity disorder affects the brain in mice](#)



Initializing reproducible report

Task: create Rmarkdown document that imports the data

1 Welcome!

1.1 Installation Instructions

2 Amazing Resources

3 Dynamic Documents

3.1 Reproducible Research

4 Markdown

5 RMarkdown

5.1 Why R Markdown?

5.2 Simple Workflow

5.3 Creating a .Rmd File

5.4 YAML Headers

5.5 Markdown Basics

5.6 Embed Code

5.6.1 Inline R Code

5.6.2 R Code Chunks

6 Knitr

6.1 Chunk Labels

6.2 Chunk Options

6.3 Global Options

Creating Dynamic Documents with RMarkdown and Knitr

Code ▾

By: Marian L. Schmidt, @micro_marian, marschmi at umich.edu

May 11th, 2016

1 Welcome!

This workshop was hosted on **May 11th, 2016** and has 2 supplemental materials available:

1. **Class notes** can be found on the workshop [etherpad](#).
2. This workshop was **recorded** live and is available on [YouTube](#). Please be welcome to tune in on YouTube!

This tutorial was constructed as a part of Dr. C Titus Brown's [Data Intensive Biology \(DIB\)](#) training program at the University of California, Davis. The DIB training program hosts local + remote workshops covering topics in bioinformatic tools and data analysis.

There were three remote classrooms that tuned in:

1. University of California, Davis
2. Simon Fraser University
3. Ontario Institute for Cancer Research

The **Github repository** for this lesson can [be found here](#).

Orchestrating Microbiome Analysis with R and Bioconductor – online book: in progress

microbiome.github.io

Preface

I Introduction

1 Data Infrastructure

1.1 Installation

1.2 Background

1.3 Loading experimental microbio...

1.4 Metadata

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](#).

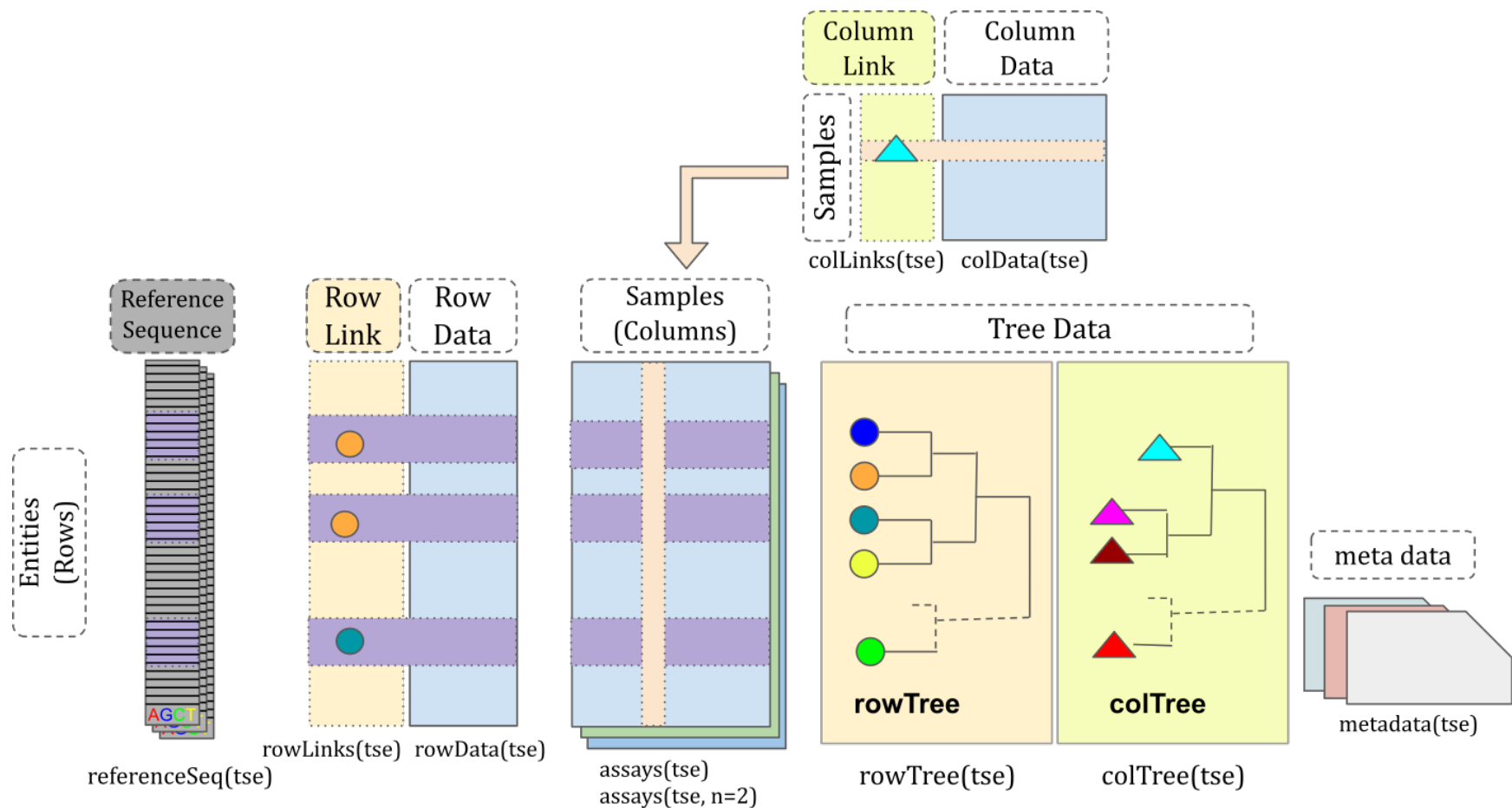
We focus on microbiome analysis tools, new, updated and established methods. In the *Introduction* section, we show how to work with the key data infrastructure `TreeSummarizedExperiment` and related classes, how this framework relates to other infrastructure and how to load microbiome analysis data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

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 exploration

Task: see the online material

