

Tuesday: Key concepts & alpha diversity

Demo

Microbiome data exploration

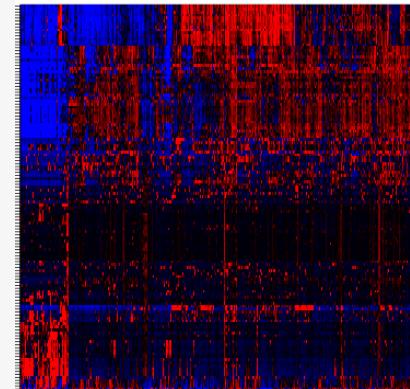
Lecture

Key concepts in microbiome data science

Practical

Alpha diversity: estimation, analysis, and visualization

OTU Abundances



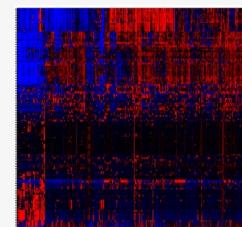
OTUs

Samples

Side information on samples

Metadata
(age, bmi, sex..)

OTU
Abundances



OTUs

Samples

Side information on samples

Side information on
taxonomic units

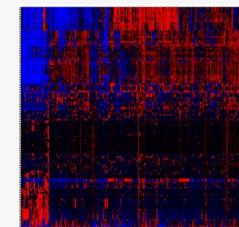
Metadata
(age, bmi, sex..)

Ref
Seq

Phylo
tree

Tax
table

OTU
Abundances



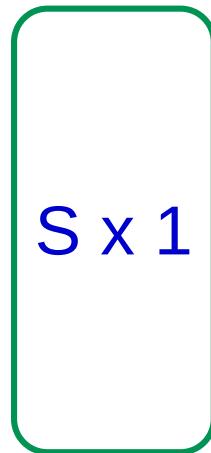
OTUs

Samples

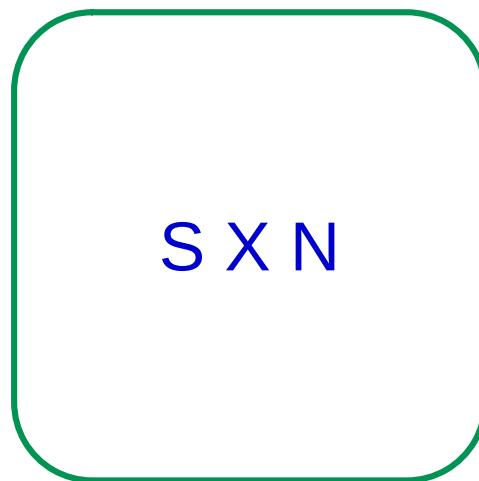
From individuals to populations, follow-ups, and multimodal data

Sequence
Variants /
OTUs

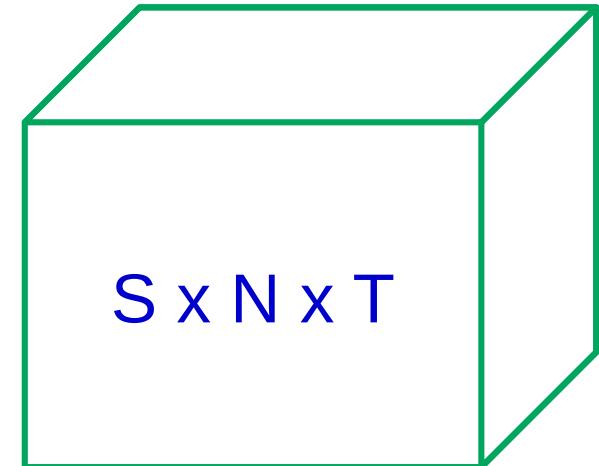
Individual



Population

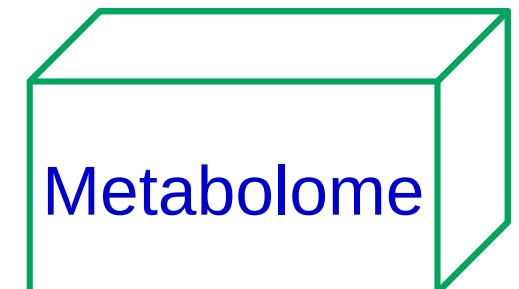
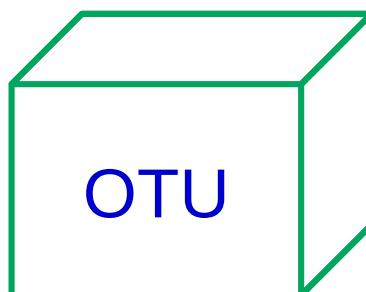


Longitudinal cohort

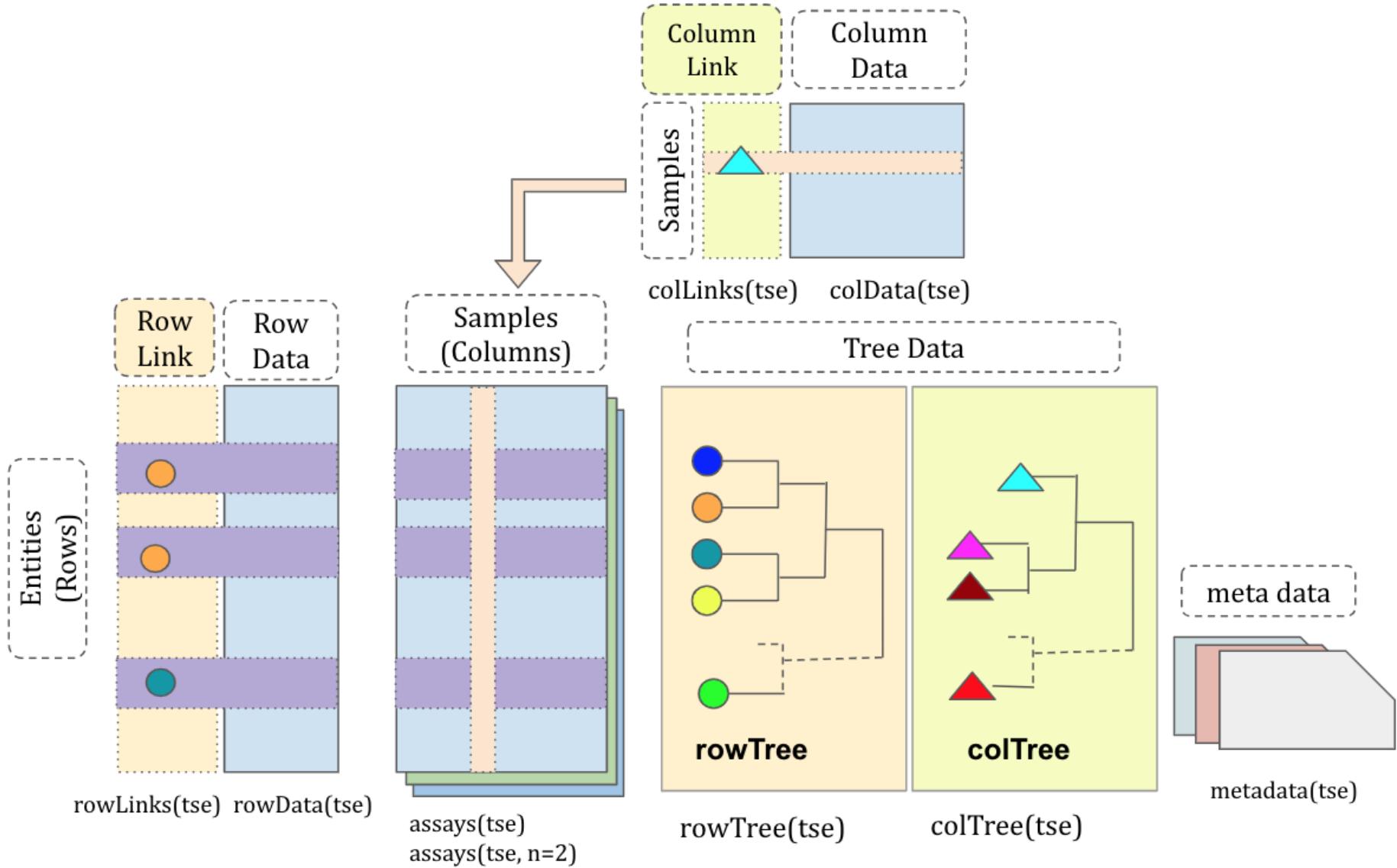


$S \times N \times T \times K$

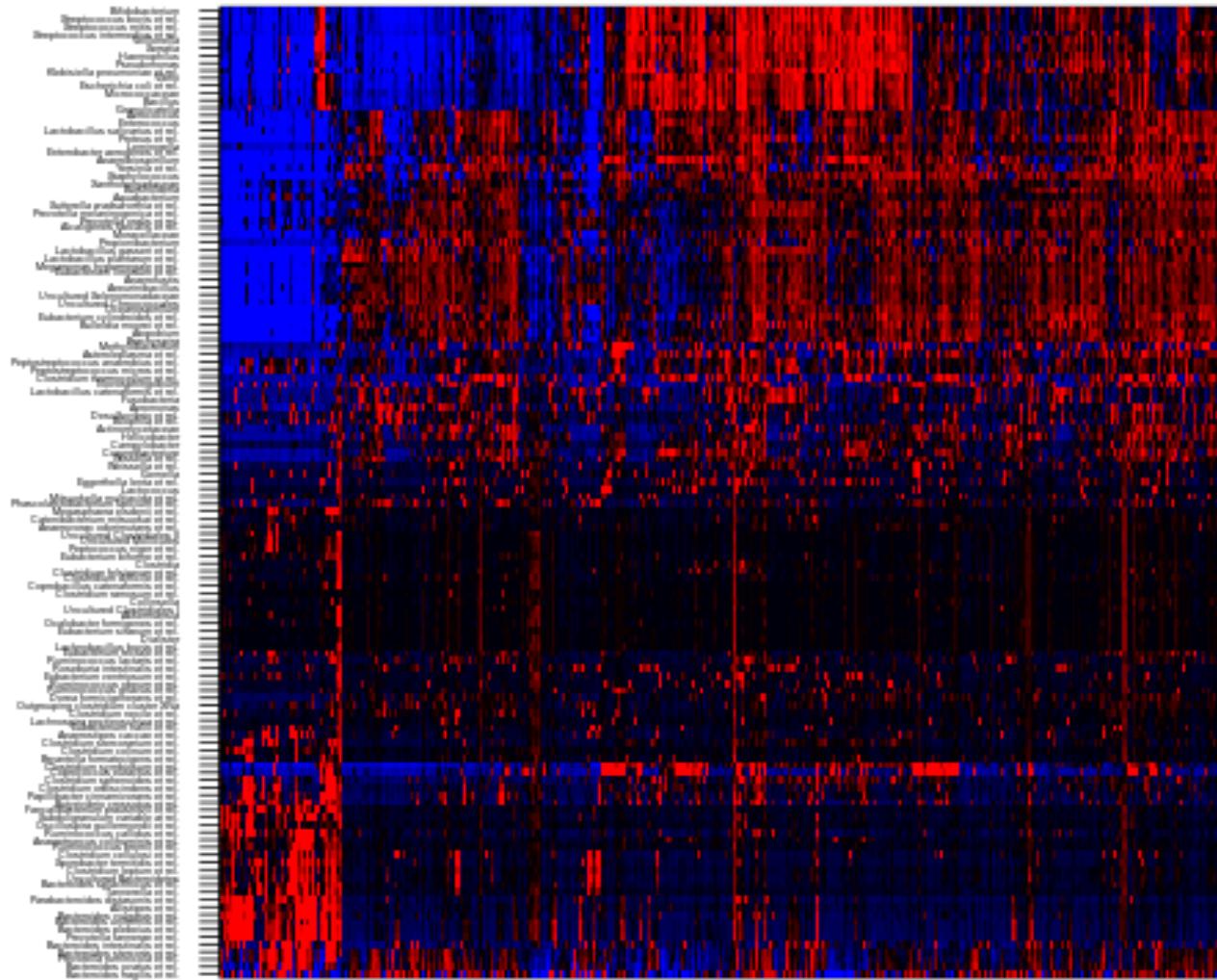
“Multi-modal” longitudinal cohort



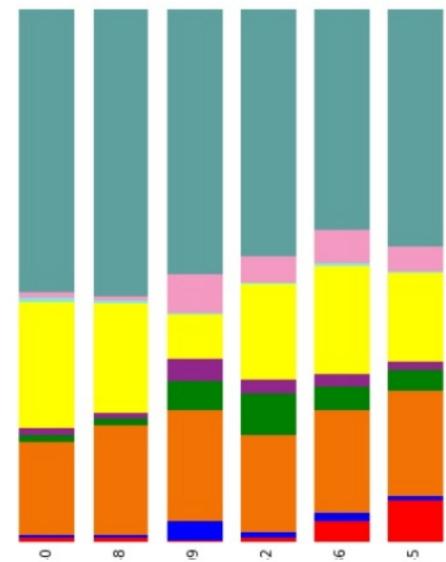
Anatomy of TreeSummarizedExperiment



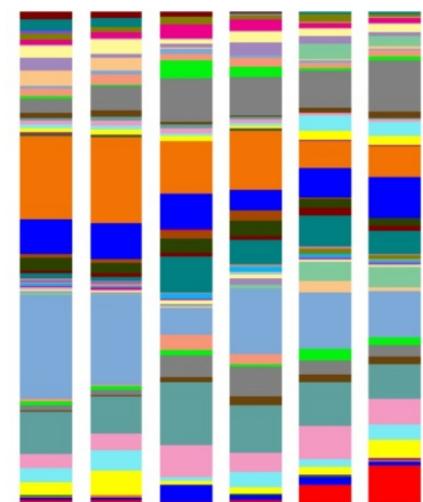
Aggregation & transformations



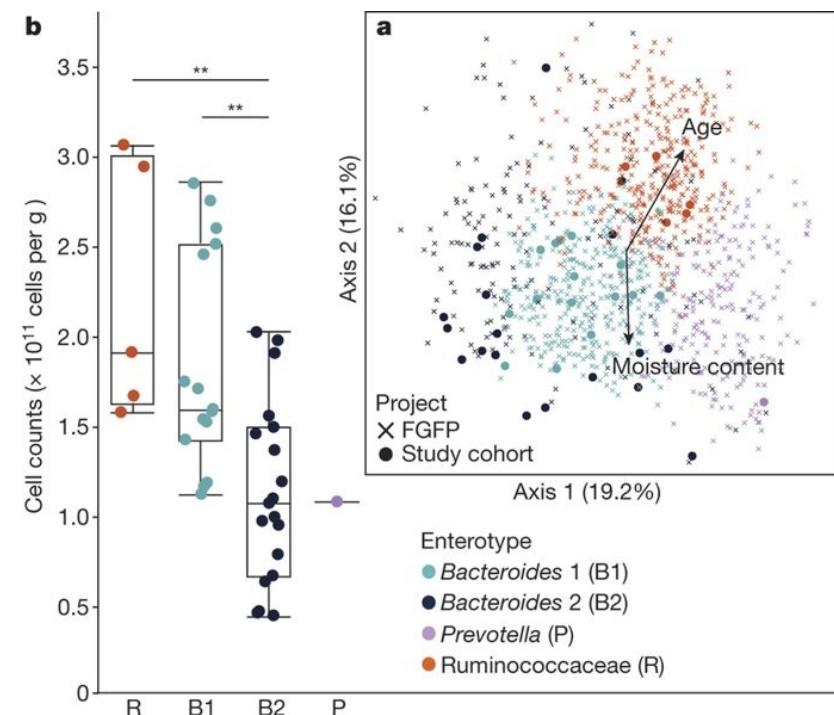
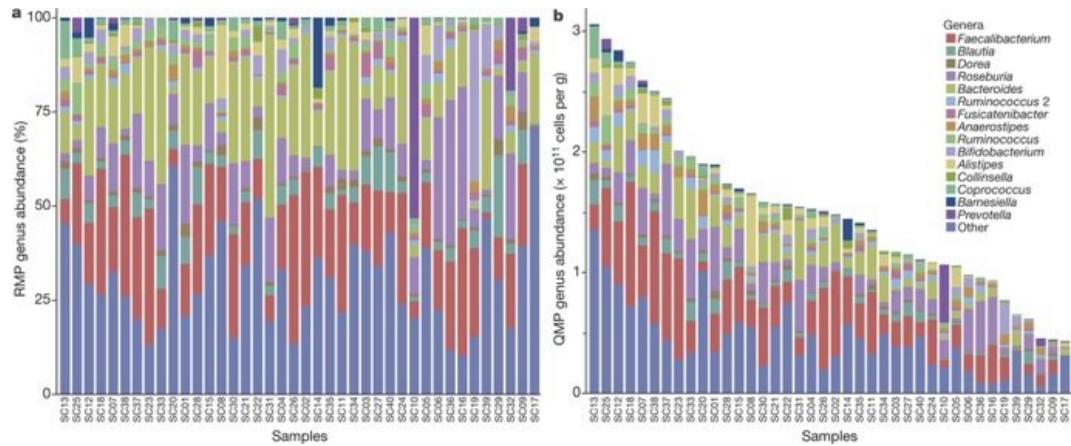
Phylum level



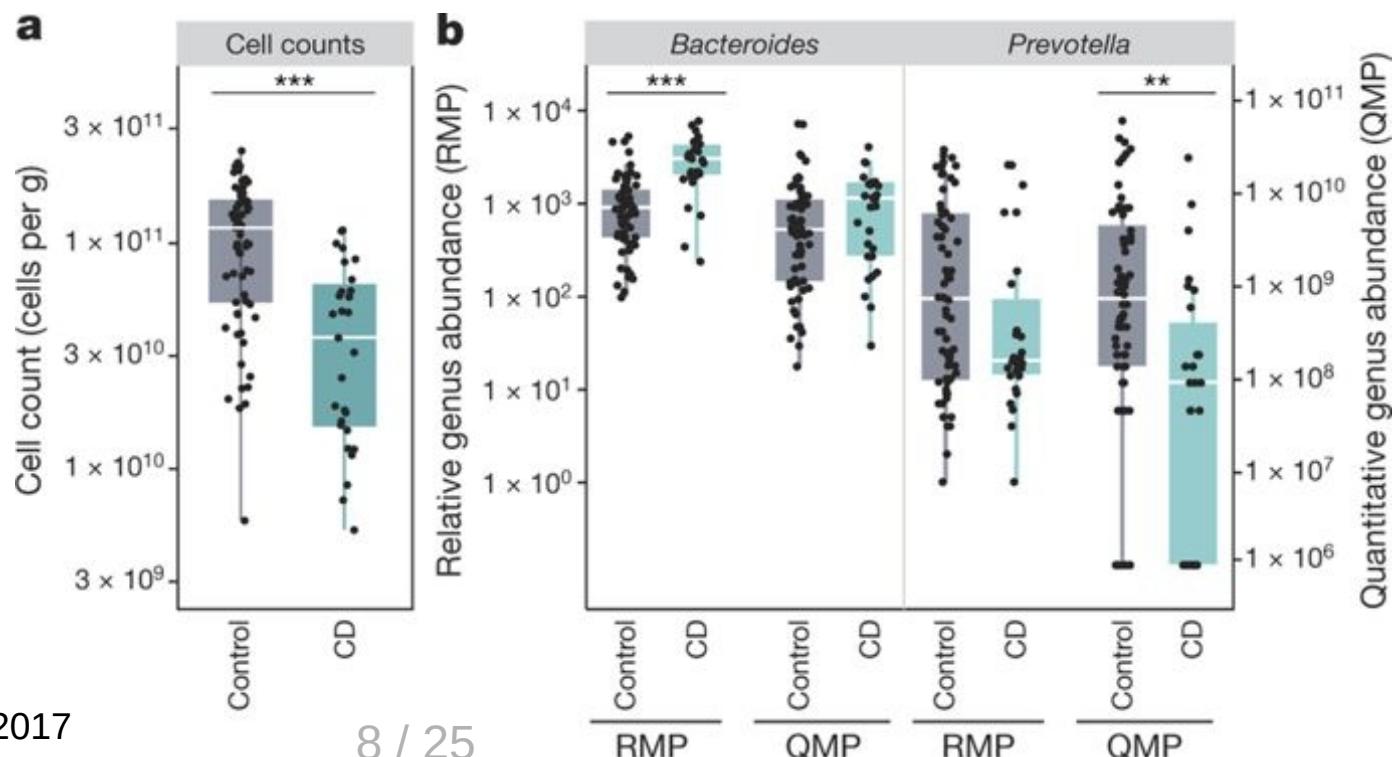
Genus level



Compositional (relative) data vs. quantitative microbiome profiling

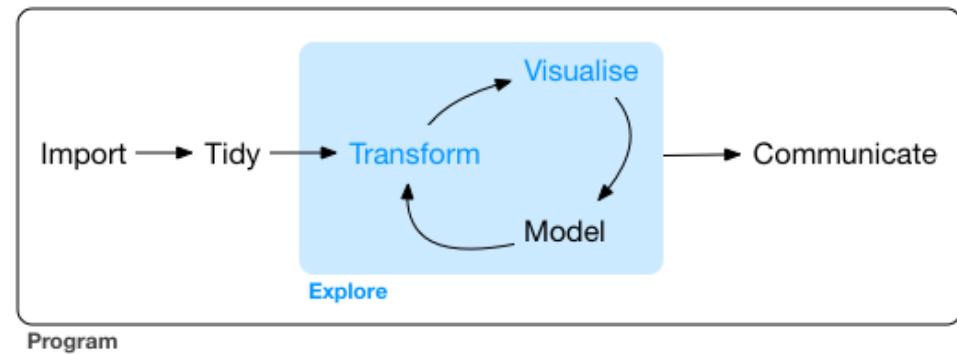


RMP vs. QMP: drastic effect on conclusions!



Standard workflow in microbiome data science

- Retrieve data
- Preprocess
- Explore
- Analyse
- Model
- Report



microbiome R package

chat on [gitter](#)

build [passing](#)

codecov

24%

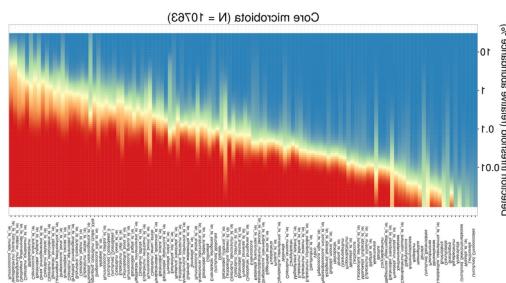
PRs [welcome](#)

Core & prevalence

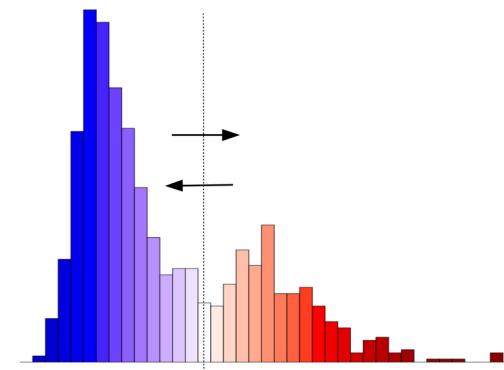
`prevalence(x)`

`core(x)`

`core_members(x)`



Stability & resilience



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

Alpha & beta diversity

`alpha(x)`

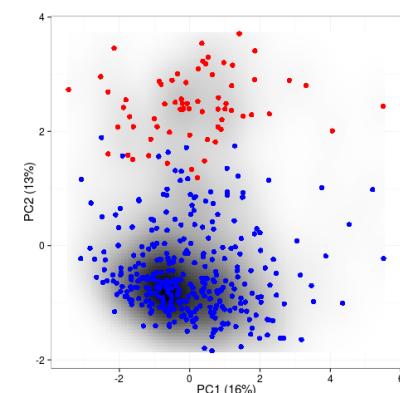
`diversity(x)`

`evenness(x)`

`dominance(x)`

`rarity(x)`

`readcount(x)`



Quality control

- continuous integration
- unit tests



Figure: Wikipedia

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. 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)



[Journal of Biosciences](#)

... October 2019, 44:115 | [Cite as](#)



Microbiome data science

Authors

Authors and affiliations

Sudarshan A Shetty, Leo Lahti

RESEARCH ARTICLE

RESEARCH ARTICLE

**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,

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

Susan P. Holmes¹

¹Department of Statistics, Stanford University, Stanford, CA, 94305, USA

²Whole Biome Inc., San Francisco, CA, 94107, USA

¹Statistics Department, Stanford University, Stanford, CA, 94305, USA

²Whole Biome Inc., San Francisco, CA, 94107, USA

First published: 04 Jun 2016 | F1000

F1000Research 2016, 5:1492 Last updated: 02 AUG 2016

Check for updates

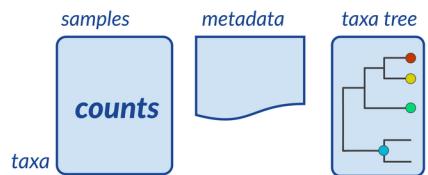
Check for updates

Example workflow

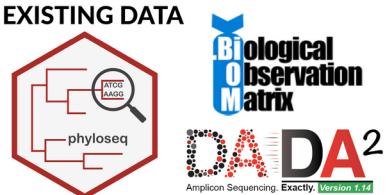
Import Data

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

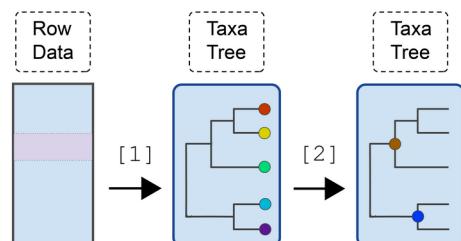


EXISTING DATA

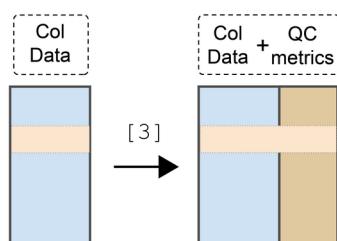


The mia Pipeline

Accessing Taxonomic Info.



Quality Control

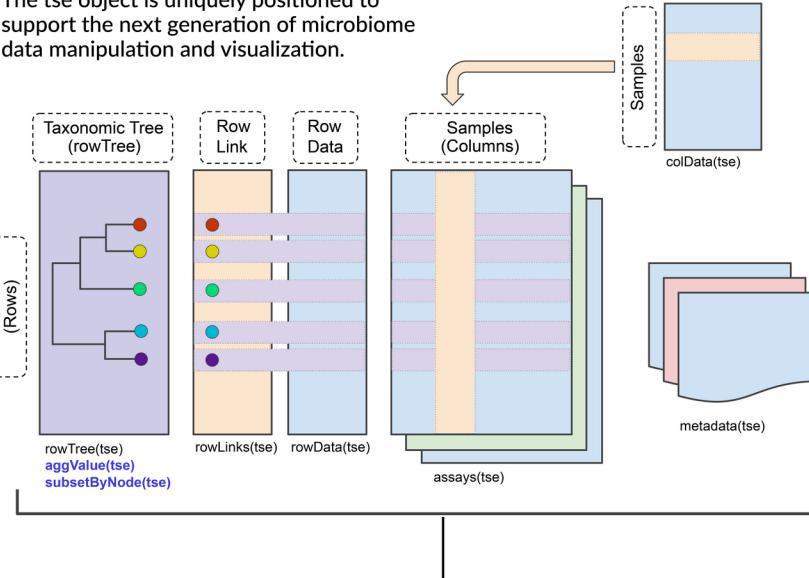


```
[1] mia::addTaxonomyTree(tse)  
[2] TreeSE:::aggValue(tse)
```

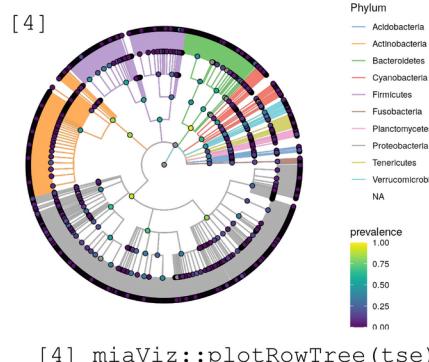
```
[3] scatter:::addPerCellQC(tse)
```

The TreeSE object

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



Visualizing with miaViz



Check the poster
F1000 / EuroBioC!



Typical study designs

Case-control studies

Interventions

Cross-sectional population cohorts

Prospective follow-ups

Longitudinal time series

Multi-omics

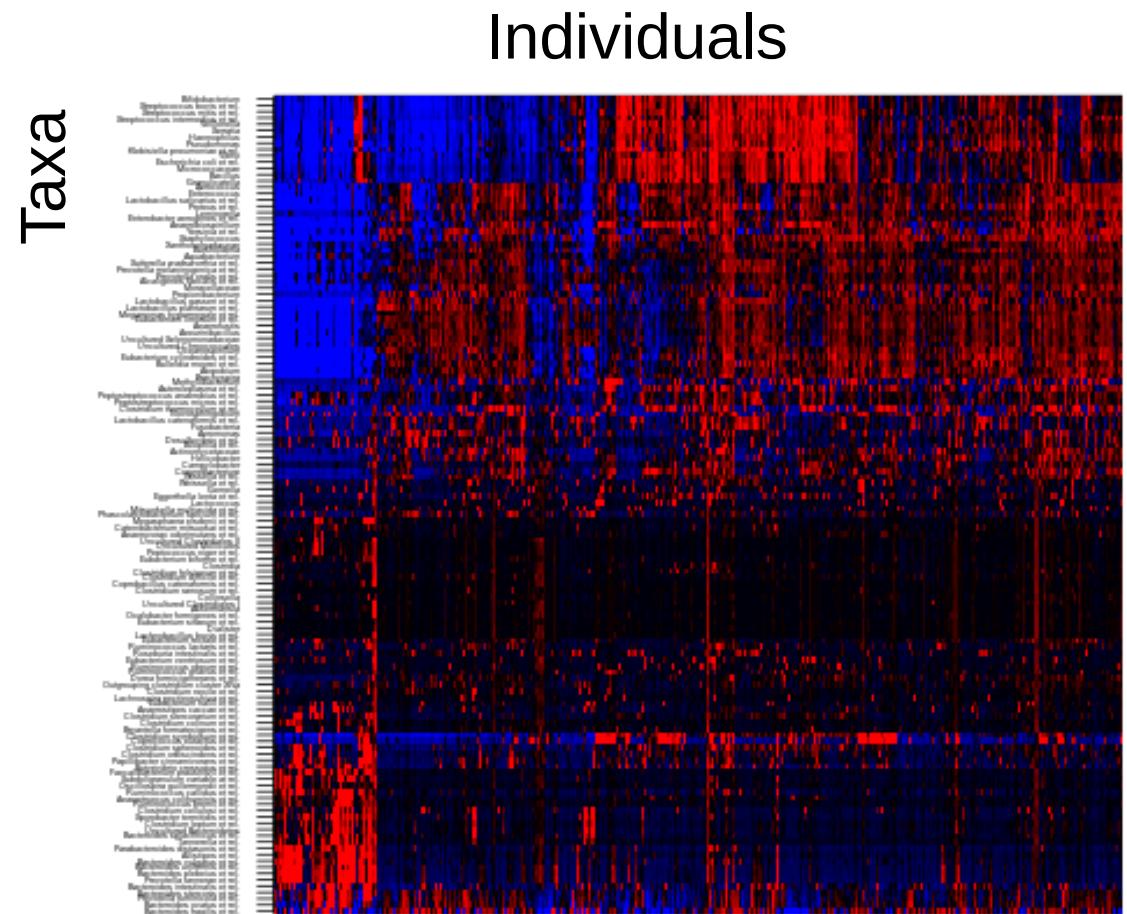
Lecture: Key concepts

- special properties of microbiome data
 - data science workflows
-
- alpha diversity
 - beta diversity
 - differential abundance

Human Intestinal Tract (HIT)Chip Atlas: 100+ genera ~ 10,000+ samples



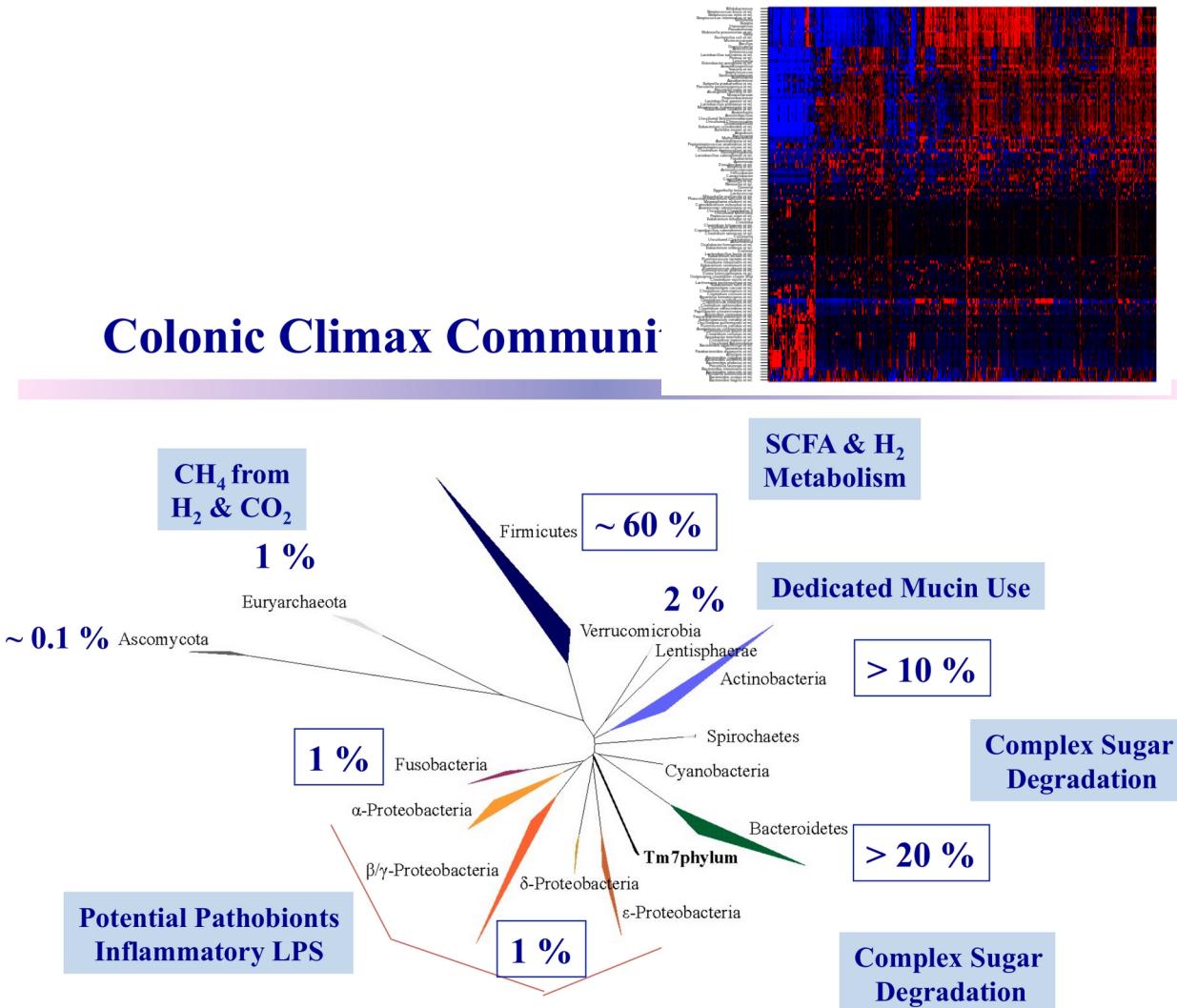
Gut microbiota: 1000 western adults (Lahti *et al.* Nature Comm. 2014)



Standardized – cost efficient – accurate at 0.1% relative abundance
Rajilic-Stojanovic et al. Env. Microbiol. 2009

Special properties of microbiome data

- Sparse
- Compositional
- Non-Gaussian
- Overdispersed
- Discrete
- Complex
- Stochastic
- Multi-level

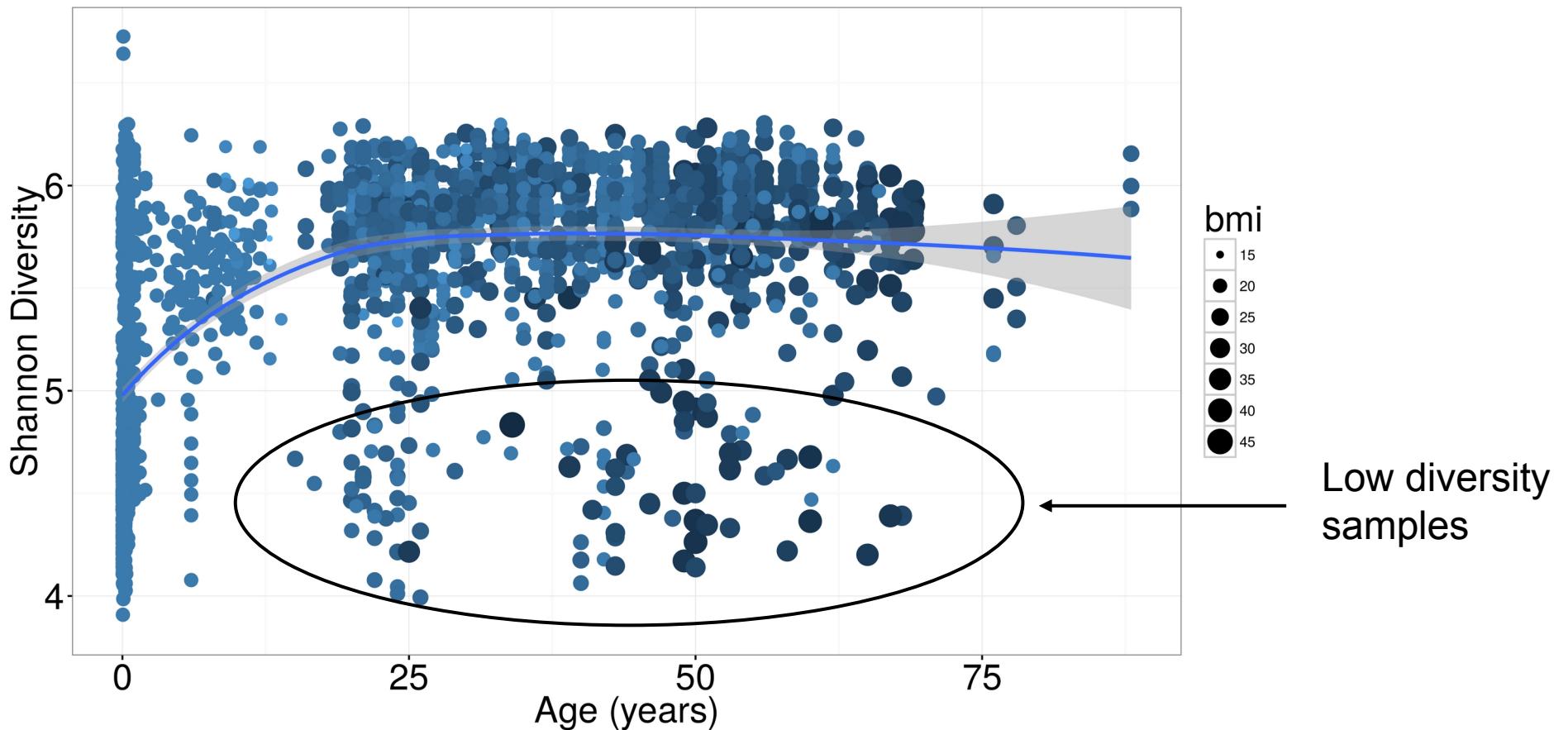


Zoetendal EG, EE Vaughan & WM de Vos (2006) Mol Microbiol 59: 1639

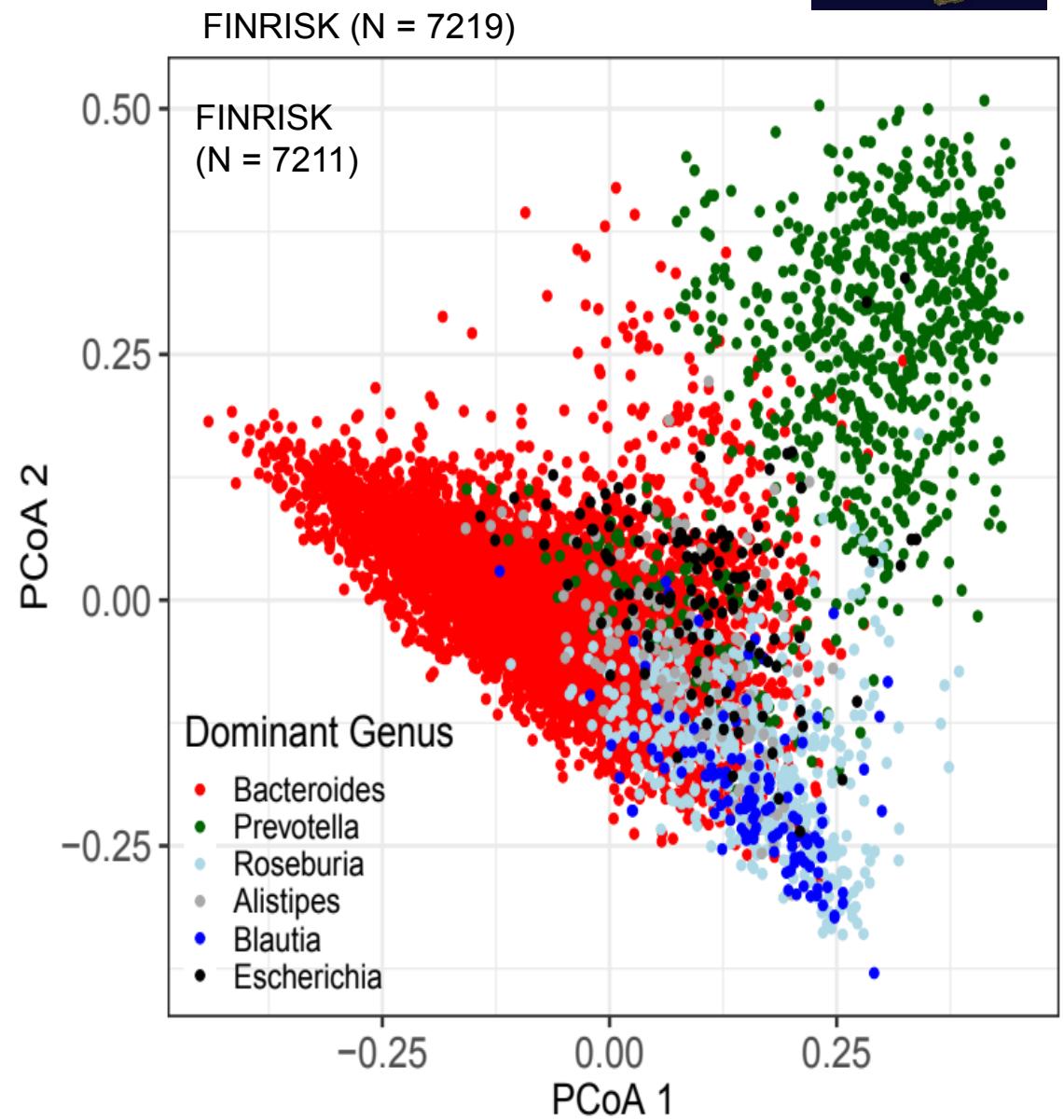
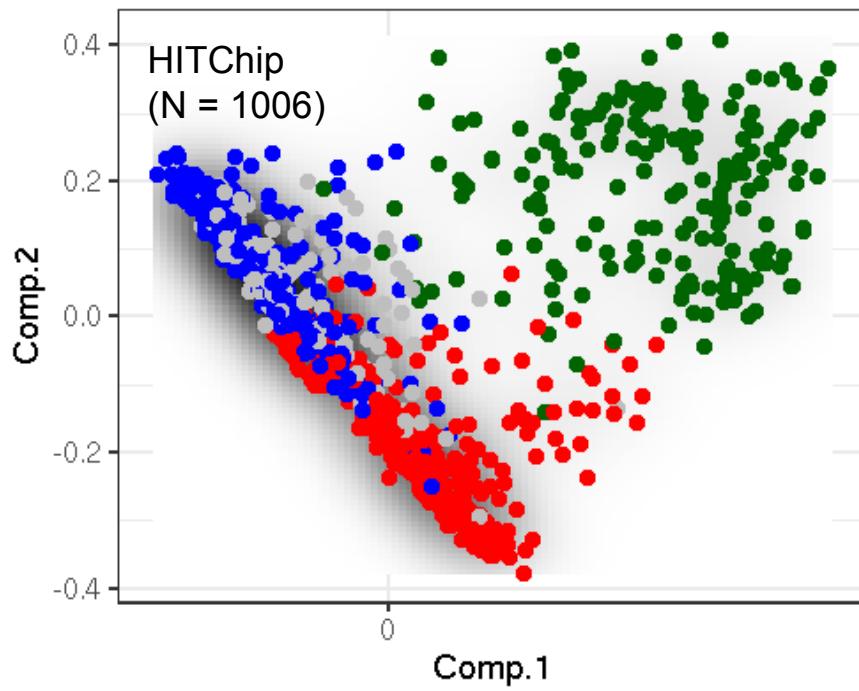
Lay C, L Rigottier-Gois, K Holmstrom, M Rajilic, EE Vaughan, WM de Vos, MD Collins, R Their, P Namsolleck, M Blaut & J Dore (2005) AEM 71: 4153

Alpha diversity & aging healthy & normal obese subjects

N = 2363



Beta diversity & population landscape

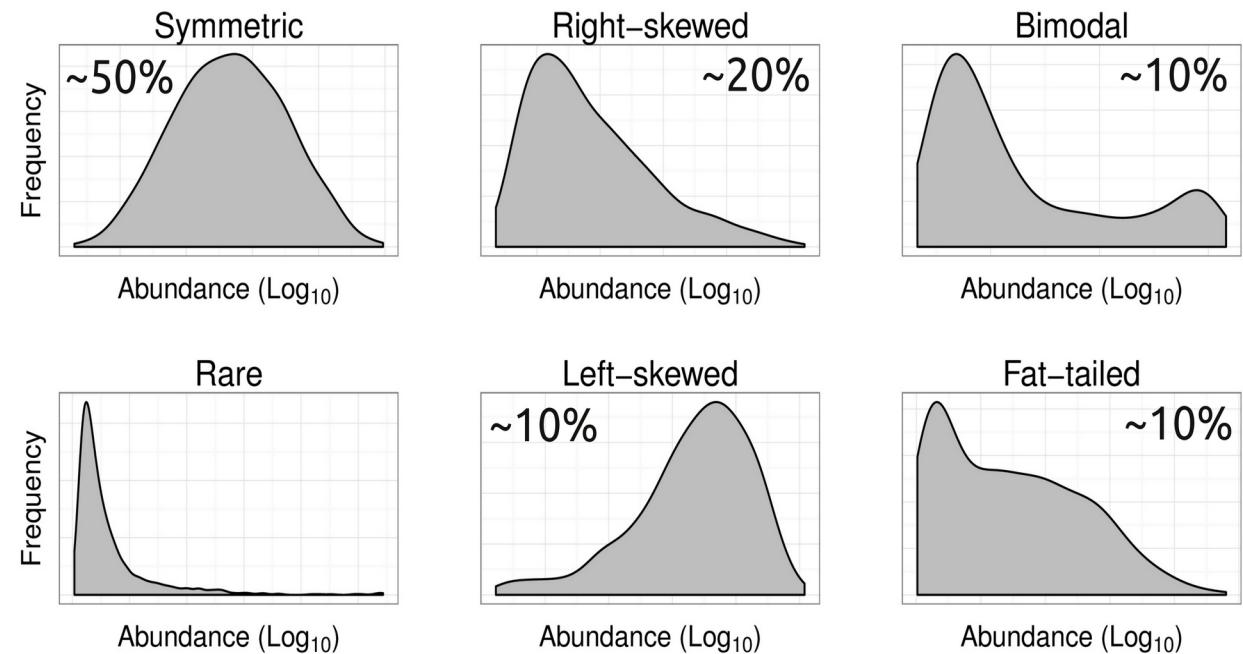


Differential abundance

Standard t-test for two-group comparison?

Problems:

- Few replicates
- Non-gaussian, discrete, positive, skewed..
- Multiple testing



Lecture: Key concepts

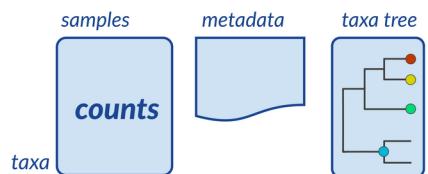
- special properties of microbiome data
 - data science workflows
-
- alpha diversity (within sample)
 - beta diversity (between samples)
 - differential abundance (individual taxa)

Example workflow

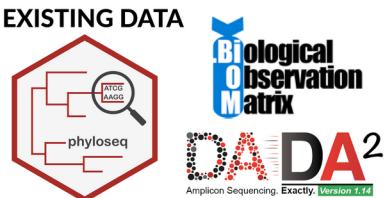
Import Data

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

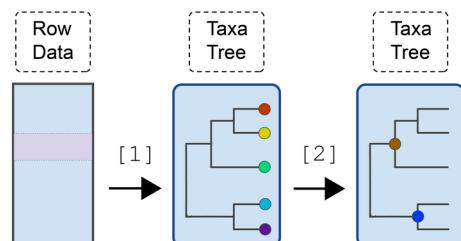


EXISTING DATA

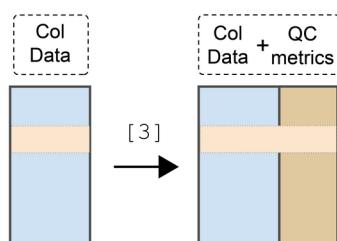


The mia Pipeline

Accessing Taxonomic Info.



Quality Control

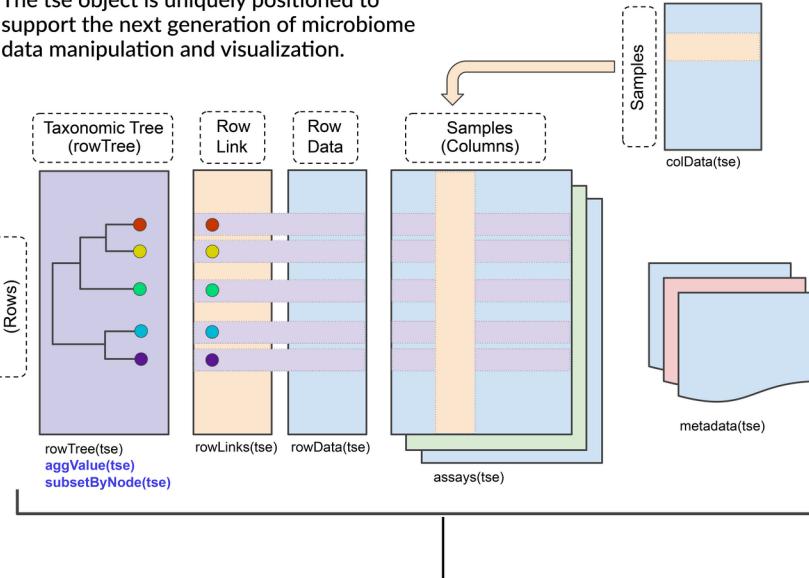


```
[1] mia::addTaxonomyTree(tse)  
[2] TreeSE:::aggValue(tse)
```

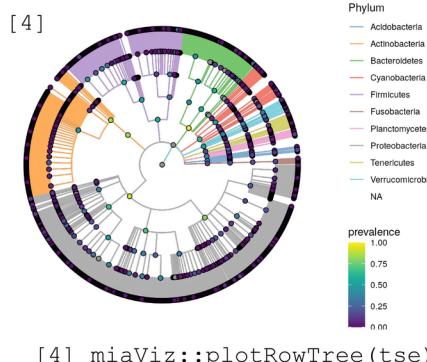
```
[3] scatter:::addPerCellQC(tse)
```

The TreeSE object

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



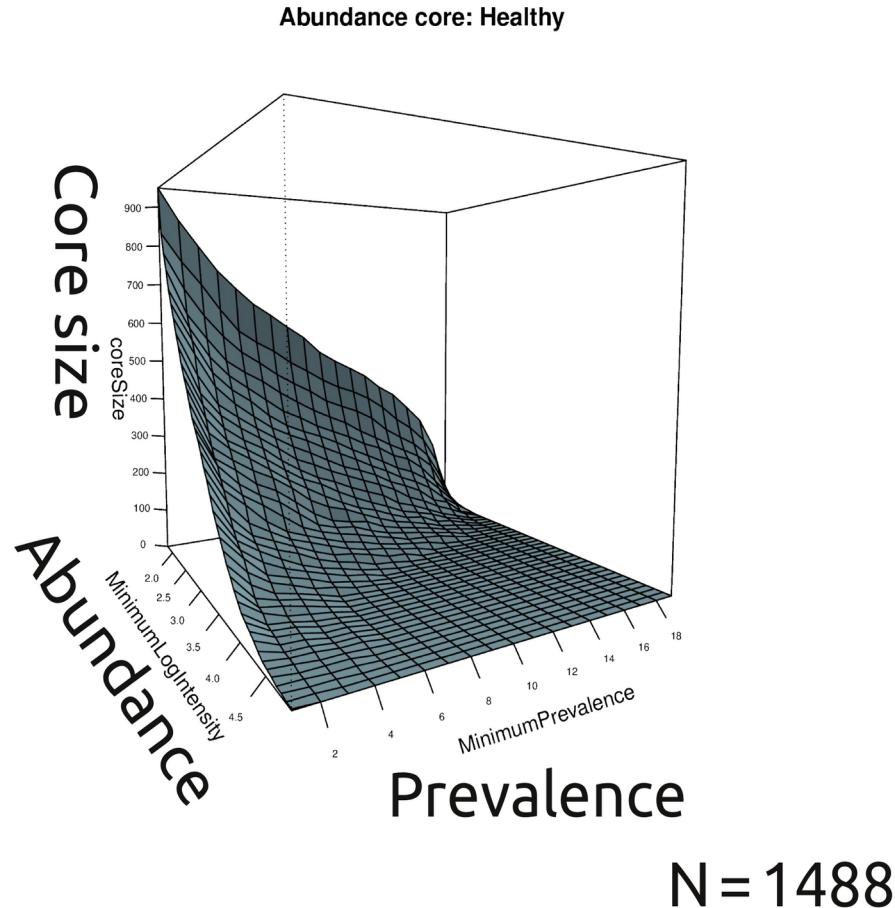
Visualizing with miaViz



Check the poster
F1000 / EuroBioC!



Shared core microbiota in healthy adults depends on analysis depth and prevalence



"Blanket analysis"
github.com/microbiome

Estimate frequency in the core for each phylotype & bootstrap for confidence intervals

Jalanka-Tuovinen J et al. (2011) Intestinal microbiota in healthy adults: Temporal analysis reveals individual and common core and relation to digestive symptoms. PLoS One 6:e23035

Salonen A et al. (2012) The adult intestinal core microbiota is determined by analysis depth and health status. Clinical Microbiology and Infection 18:16–20.

Microbial communities as dynamical systems

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

