

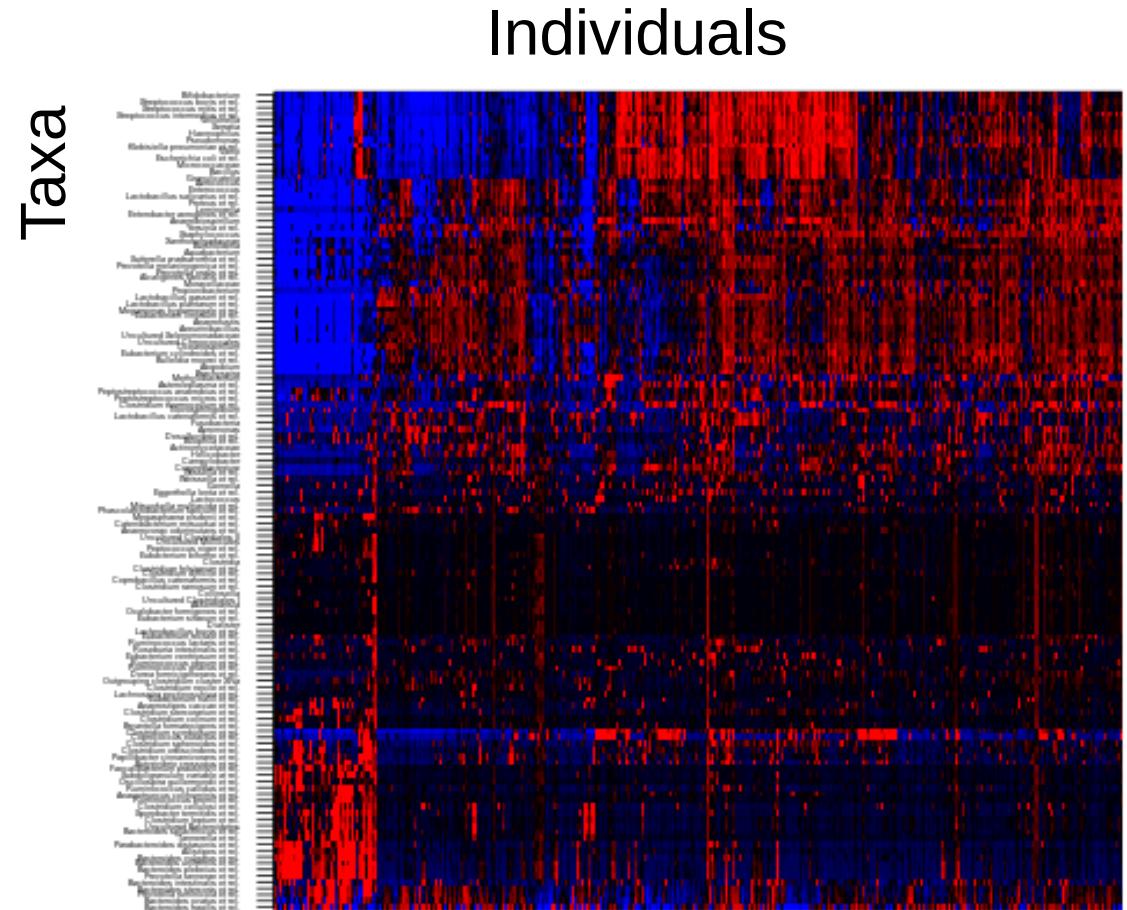
Lecture: Key concepts

- 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



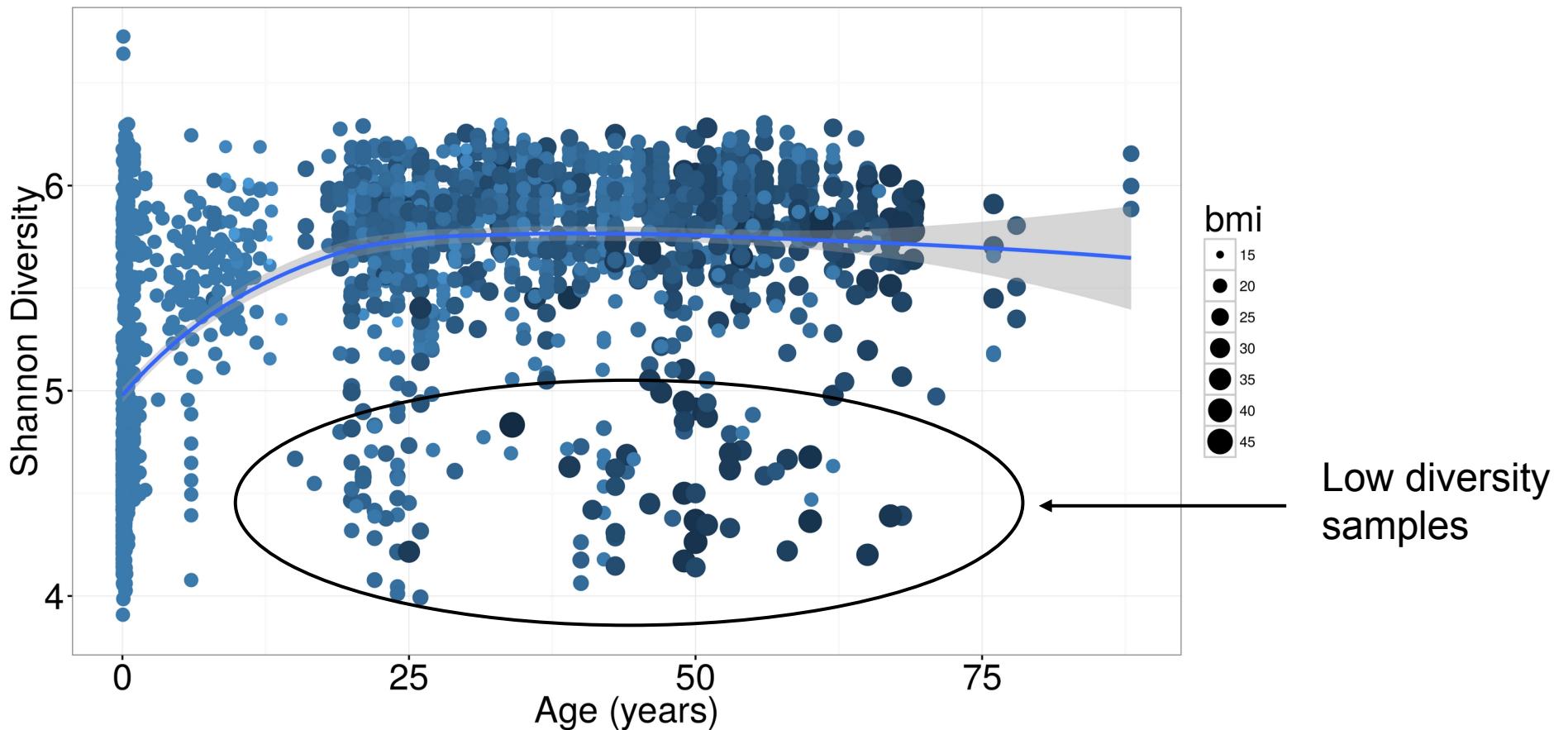
Gut microbiota: 1000 western adults (Lahti *et al.* Nature Comm. 2014)



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

Aging, microbiome diversity & tipping elements: healthy & normal obese subjects

N = 2363



Key sources of microbial ecosystem variation

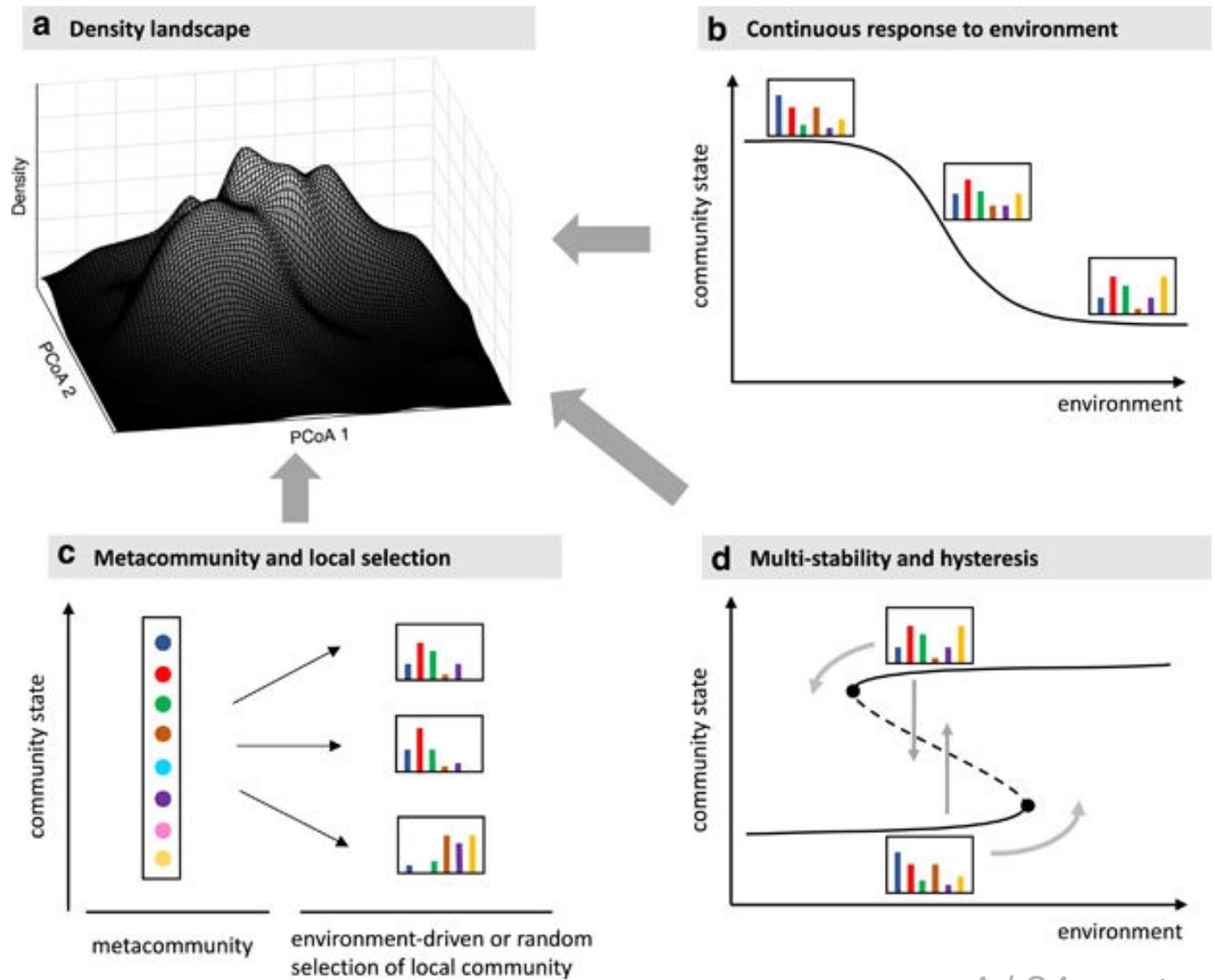
External perturbations (push & pulse)

Internal dynamics and multi-stability

Immigration

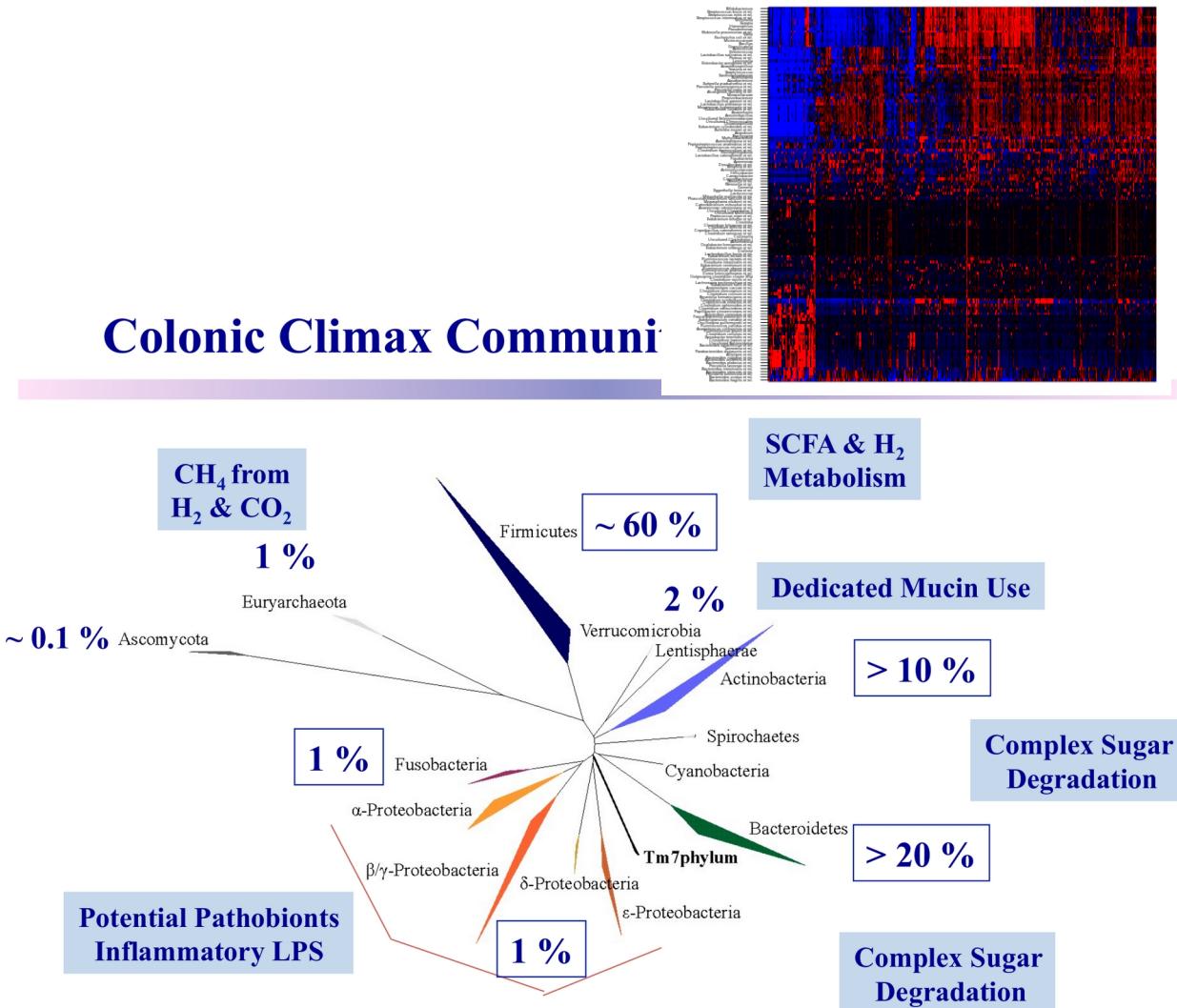
Stochasticity

Memory



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

State diagnosis & manipulation: from specific targets to the overall ecosystem

Diet

Life style

Antibiotics

Probiotics

Prebiotics

Fecal transplants

Figure 3: Spectrum of microbiome-derived modulators being pursued by biotech companies, ranging from ecosystem-level interventions to single-target approaches.

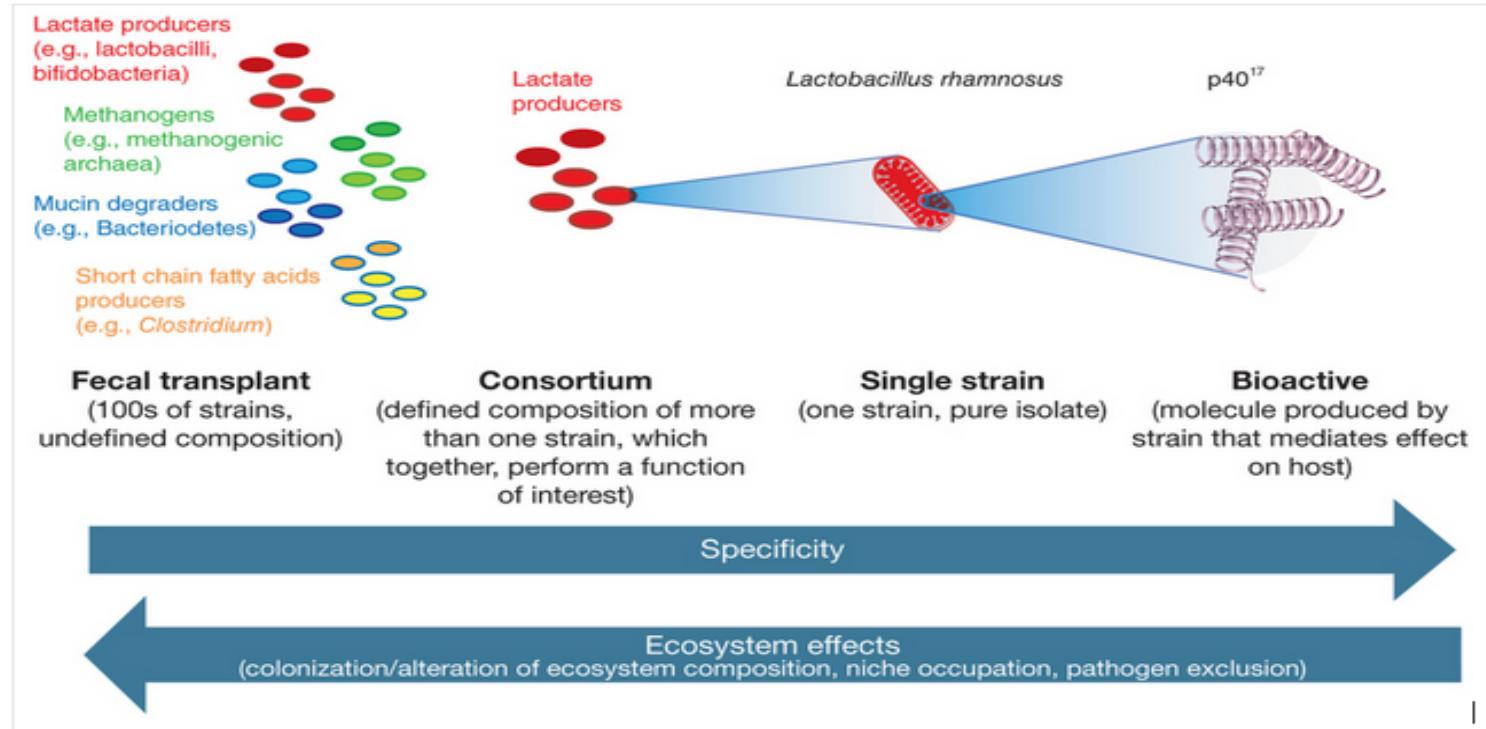
From

Medicines from microbiota

Bernat Ollé

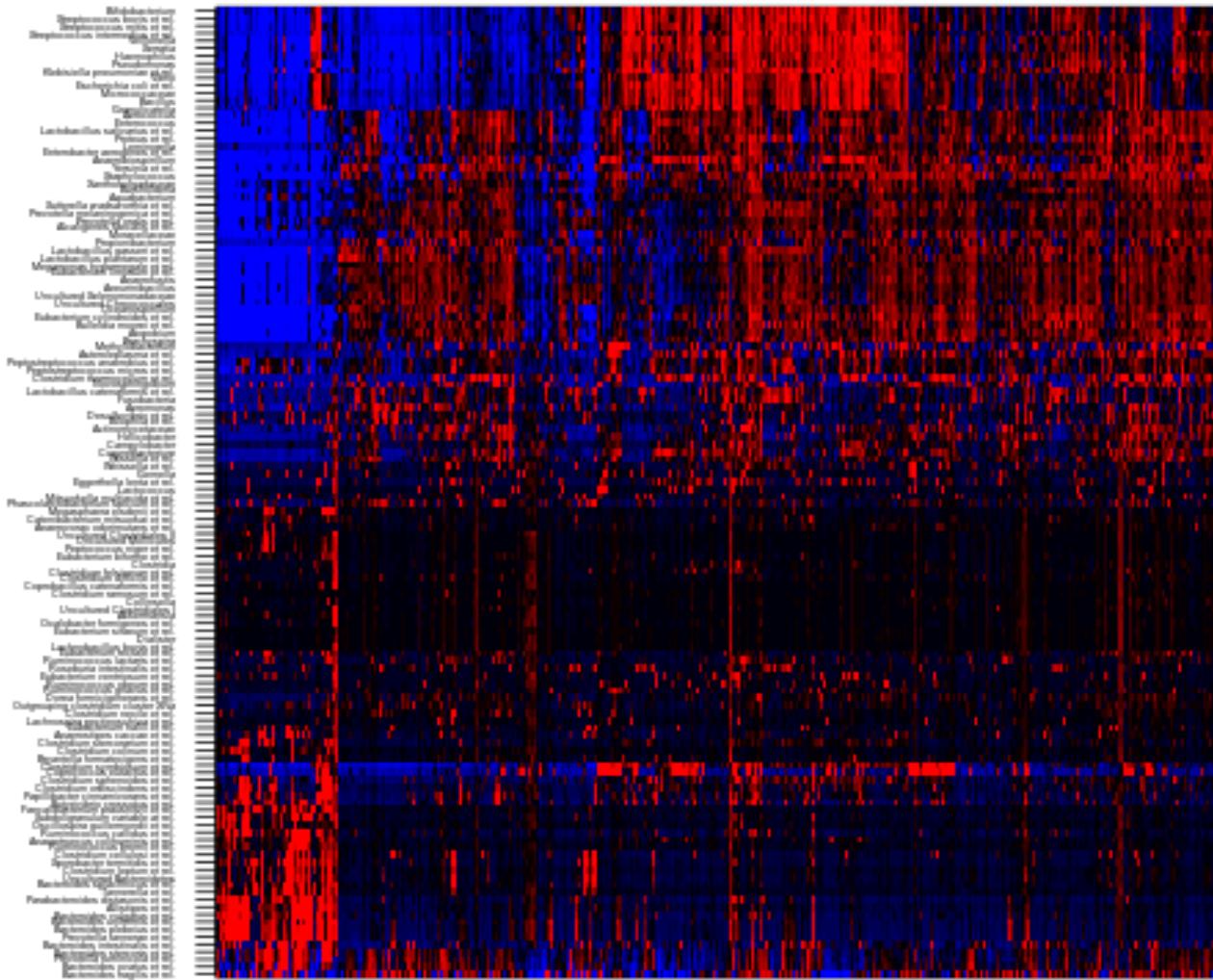
Nature Biotechnology 31, 309–315 (2013) doi:10.1038/nbt.2548

Figure 3: Spectrum of microbiome-derived modulators being pursued by biotech companies, ranging from ecosystem-level interventions to single-target approaches.

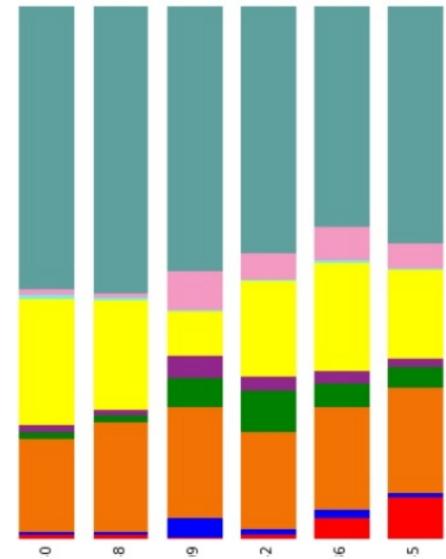


'Lactate producer' is used here as a functional attribute descriptive of a community. Species belonging to the 'lactate producers' community (e.g., *L. rhamnosus*) may also belong to other communities. A community may be described by a metabolic function (e.g., lactate production) or by any other functional attribute (e.g., regulatory T-cell induction or vitamin K production). p40 is a bioactive, soluble protein expressed by *L. rhamnosus*, which mediates intestinal epithelial homeostasis¹⁷.

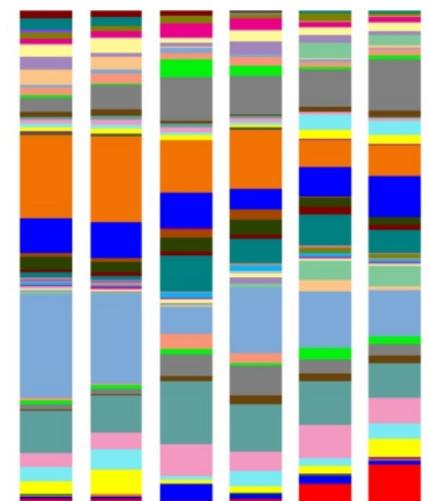
Feature selection?



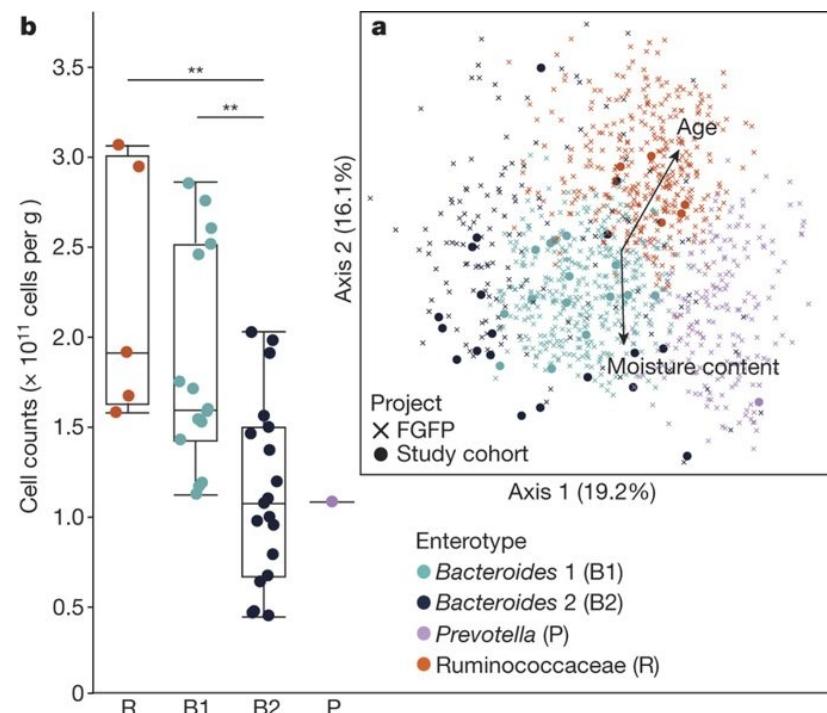
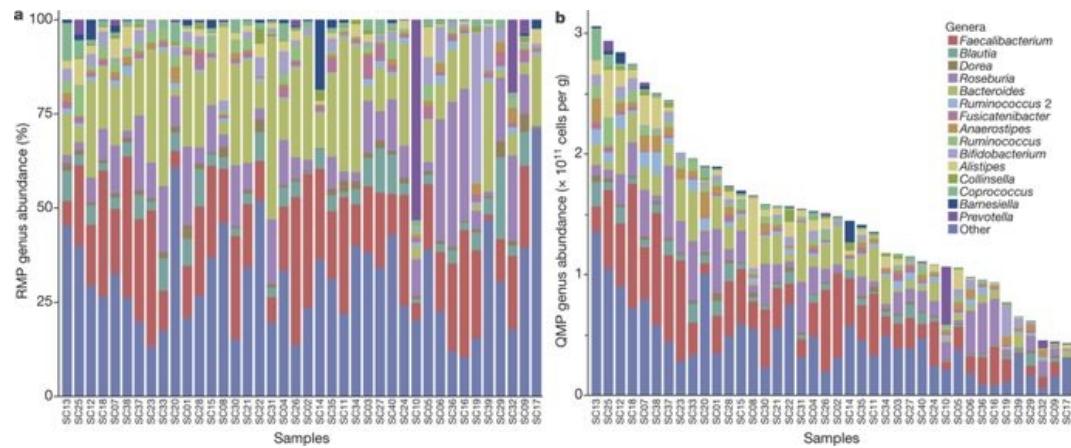
Phylum level



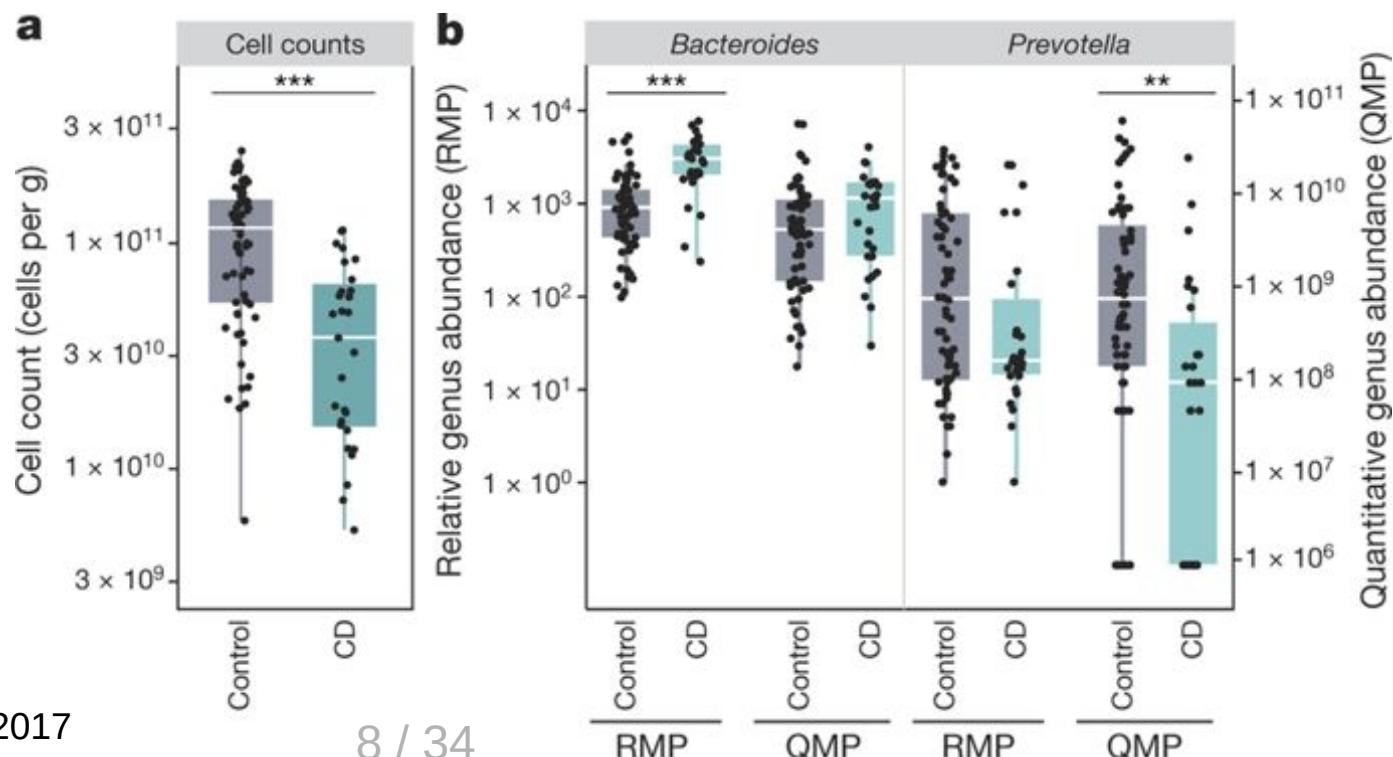
Genus level



Relative versus absolute abundance: quantitative microbiome profiling

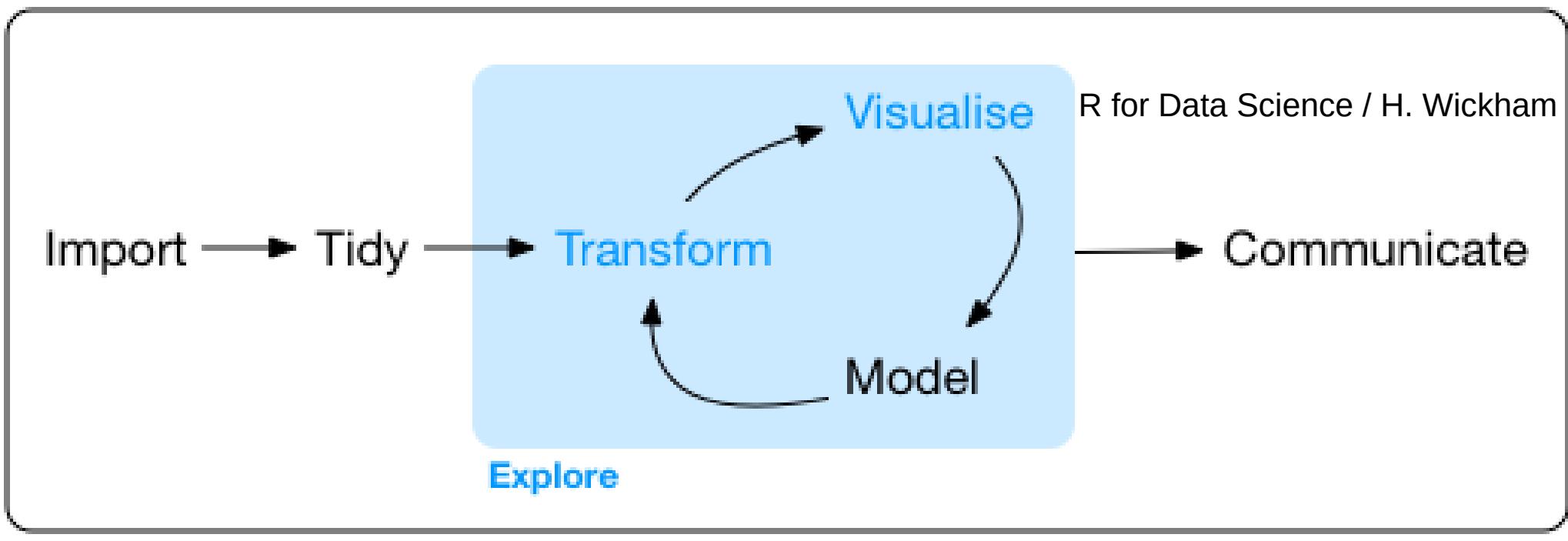


RMP vs. QMP:
drastic effect
on conclusions!



Demo: Data exploration

Task: see the online material



Reproducible reporting and best practices

Typical study designs

Case-control studies

Interventions

Cross-sectional population cohorts

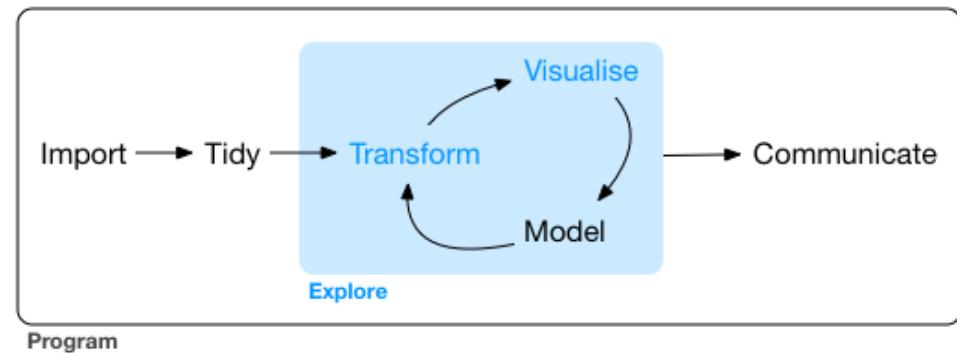
Prospective follow-ups

Longitudinal time series

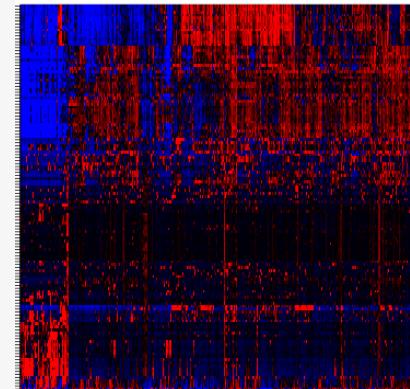
Multi-omics

Standard workflow in microbiome data science

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



OTU Abundances



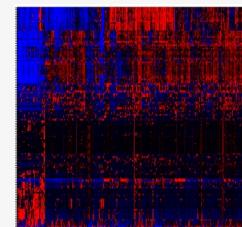
OTUs

Samples

Side information on samples

Metadata
(age, bmi, sex..)

OTU
Abundances



OTUs

Samples

Side information on samples

Side information on
taxonomic units

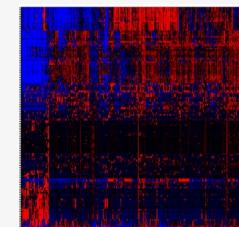
Ref
Seq

Phylo
tree

Tax
table

Metadata
(age, bmi, sex..)

OTU
Abundances



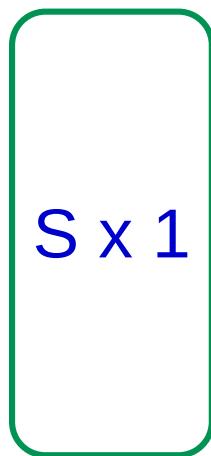
OTUs

Samples

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

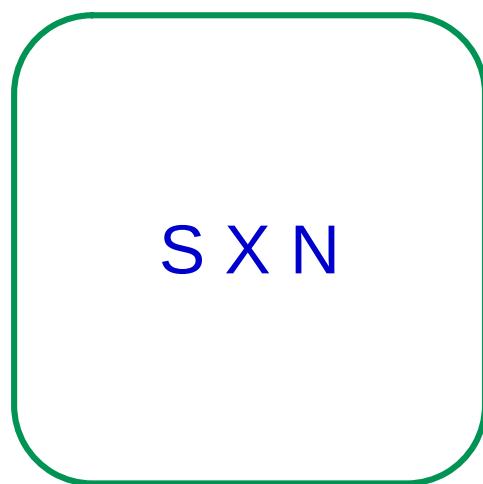
16 / 34

Individual

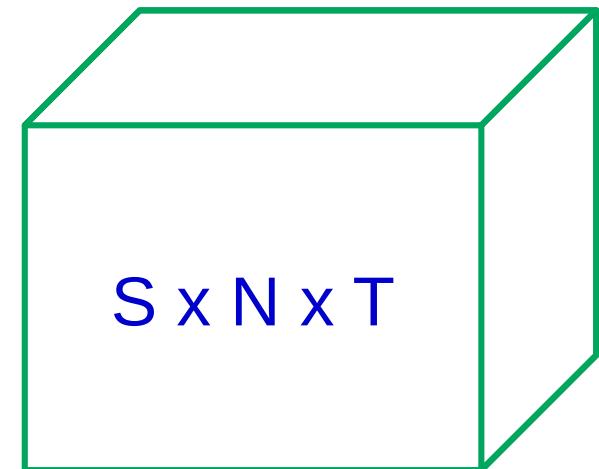


Sequence
Variants /
OTUs

Population

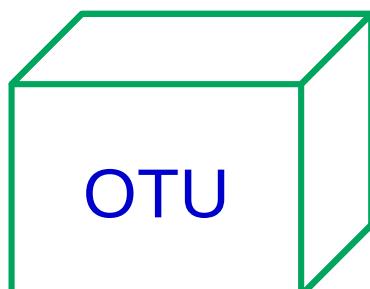


Longitudinal cohort

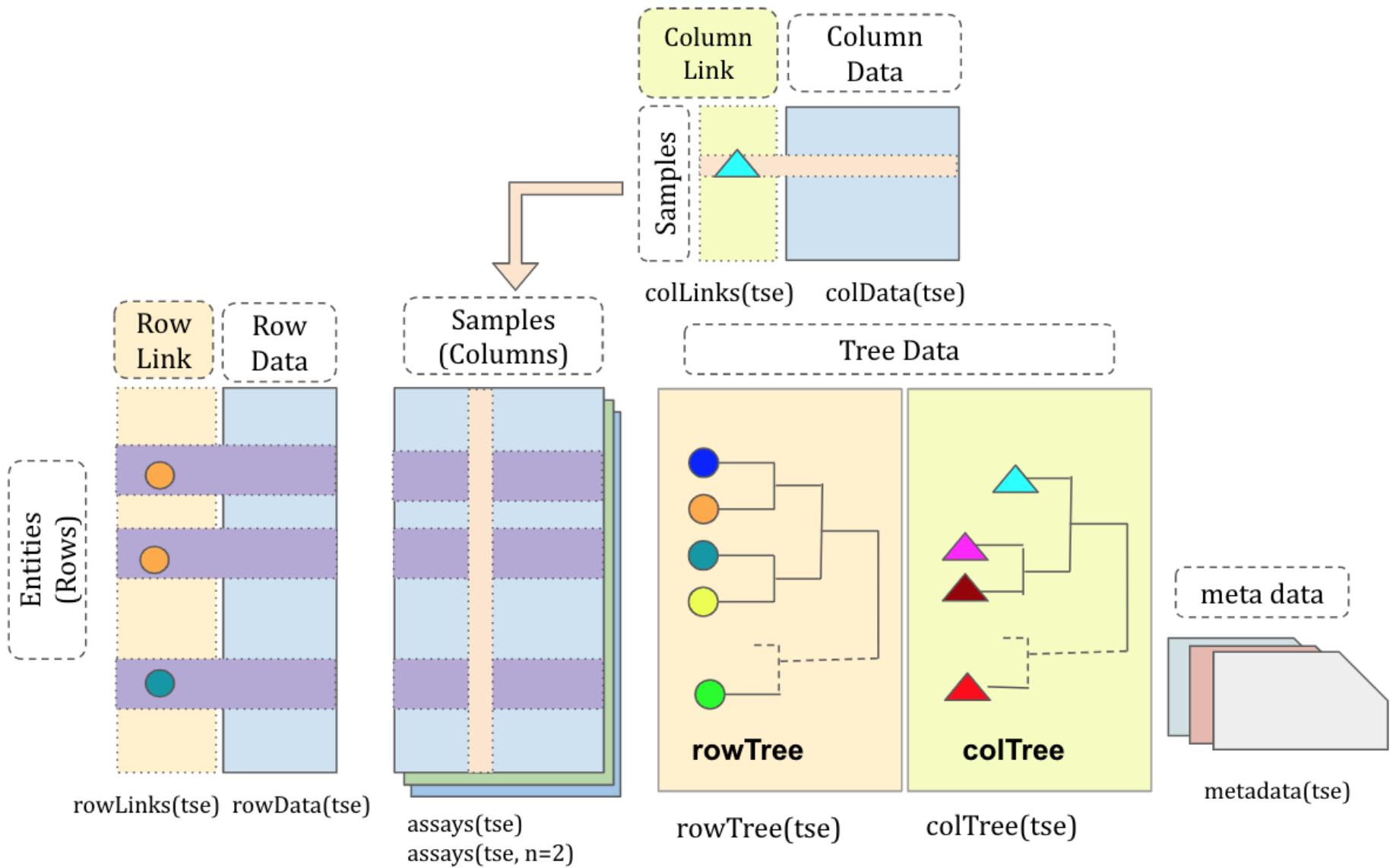


$S \times N \times T \times K$

“Multi-modal” longitudinal cohort



Anatomy of TreeSummarizedExperiment

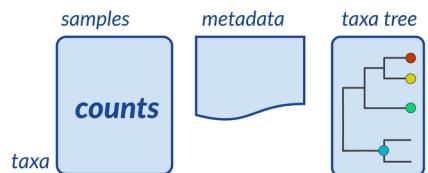


Example workflow

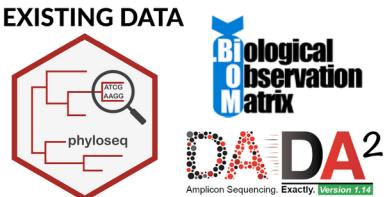
Import Data

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

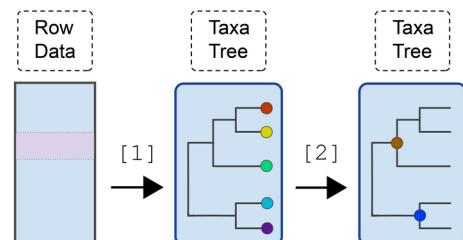


EXISTING DATA

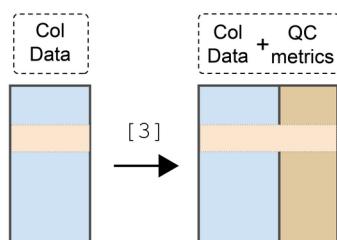


The mia Pipeline

Accessing Taxonomic Info.



Quality Control

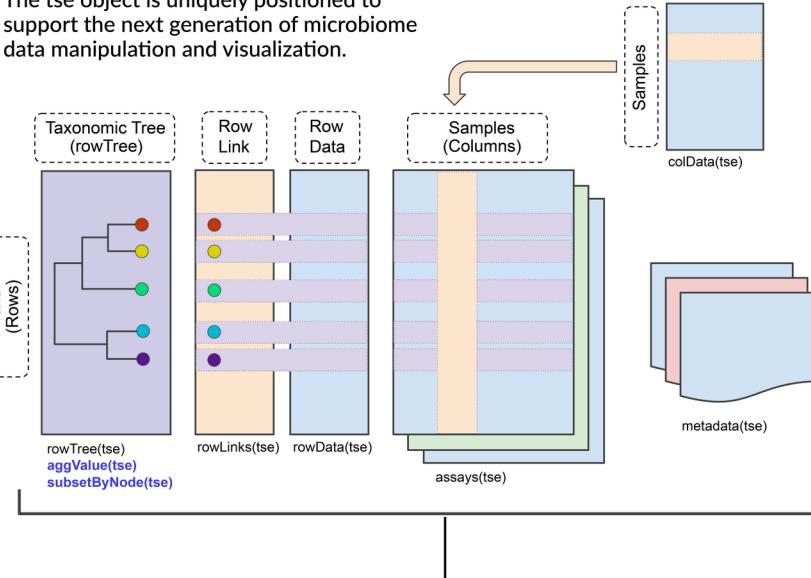


```
[1] mia::addTaxonomyTree(tse)  
[2] TreeSE:::aggValue(tse)
```

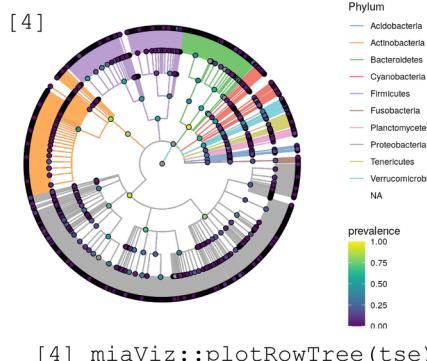
```
[3] scatter:::addPerCellQC(tse)
```

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!



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

microbiome R package

chat on [gitter](#)

build [passing](#)

codecov

24%

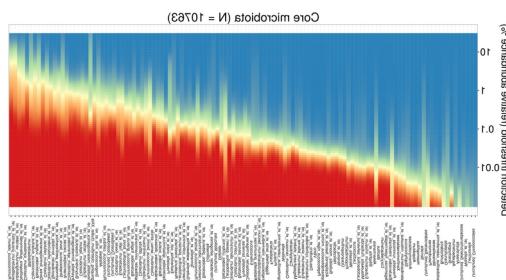
PRs [welcome](#)

Core & prevalence

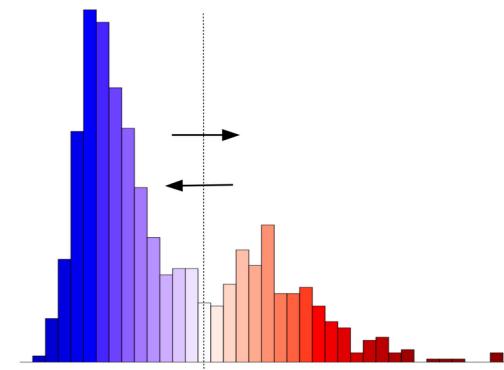
`prevalence(x)`

`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)`

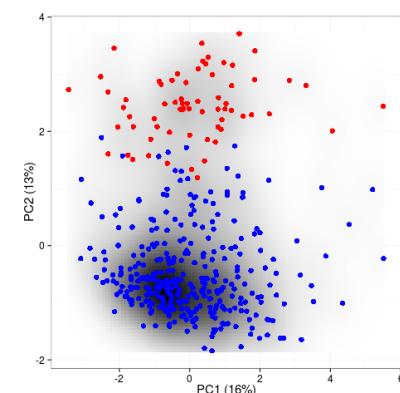
`diversity(x)`

`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. qimer 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

Tuesday: Key concepts & alpha diversity

Demo

Microbiome data exploration

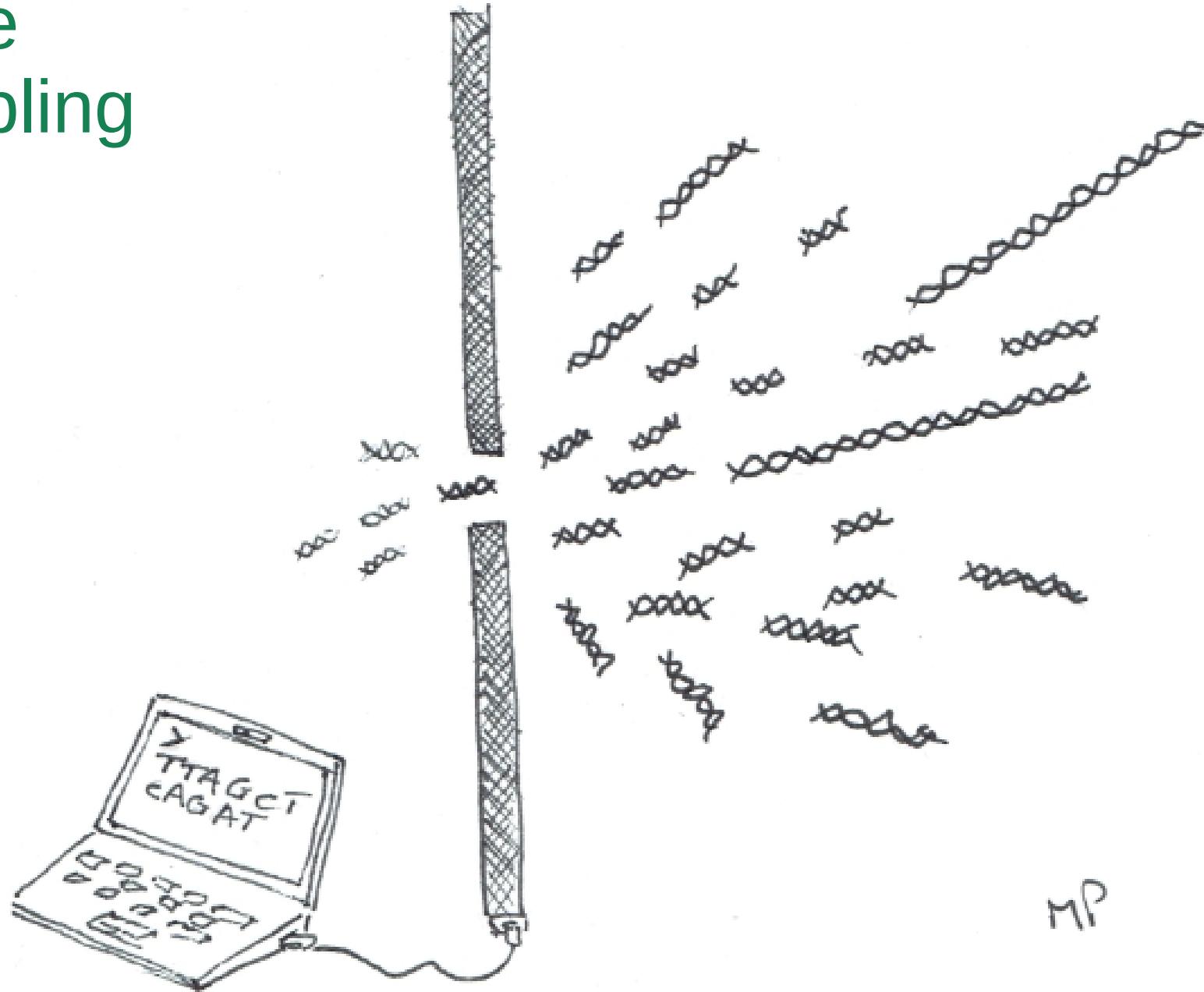
Lecture

Key concepts in microbiome data science

Practical

Alpha diversity: estimation, analysis, and visualization

Finite sampling



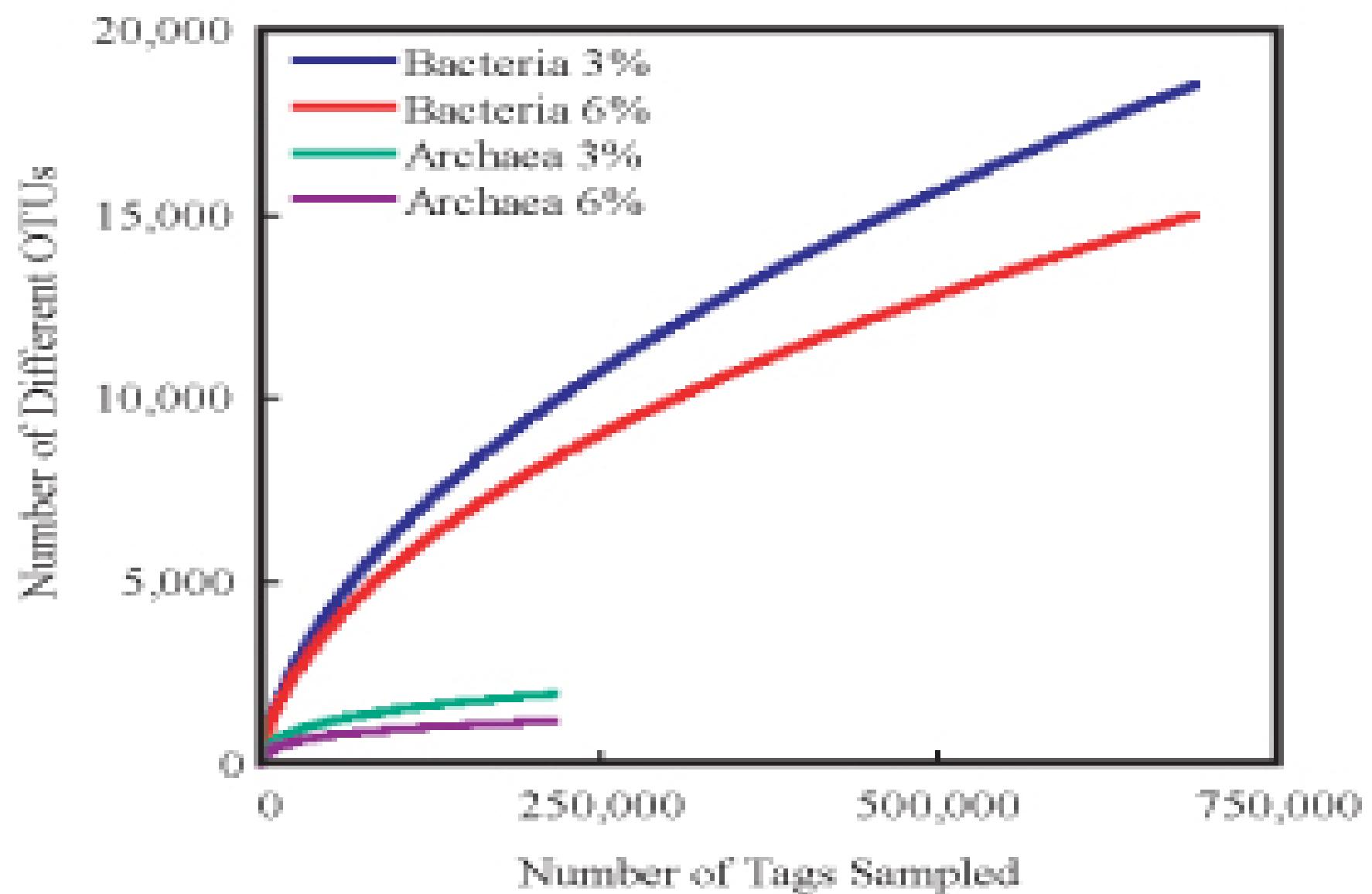


Figure 2

Wednesday: Beta diversity

Demonstration

Community similarity

Practical

Beta diversity: estimation, analysis, and visualization

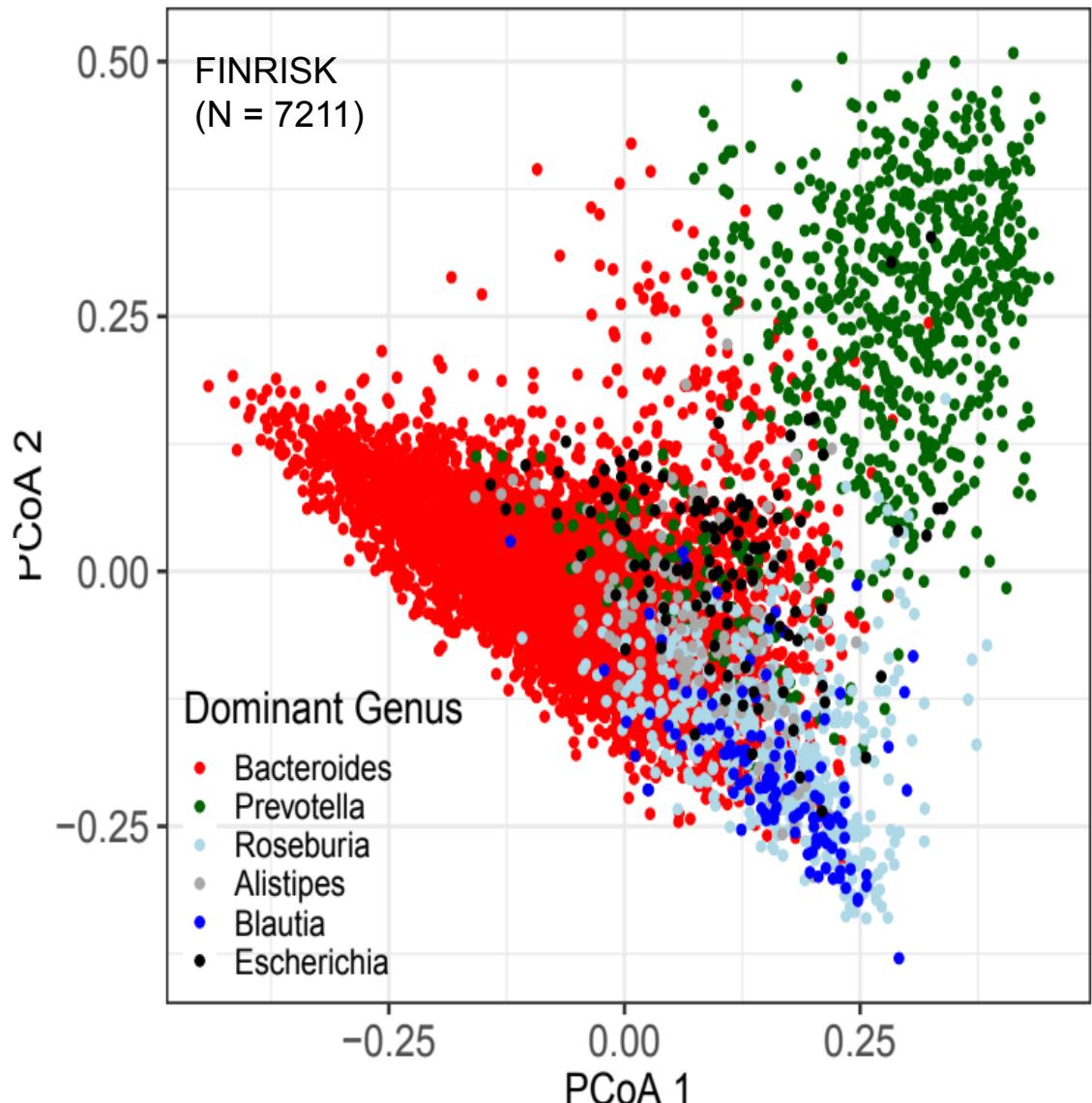
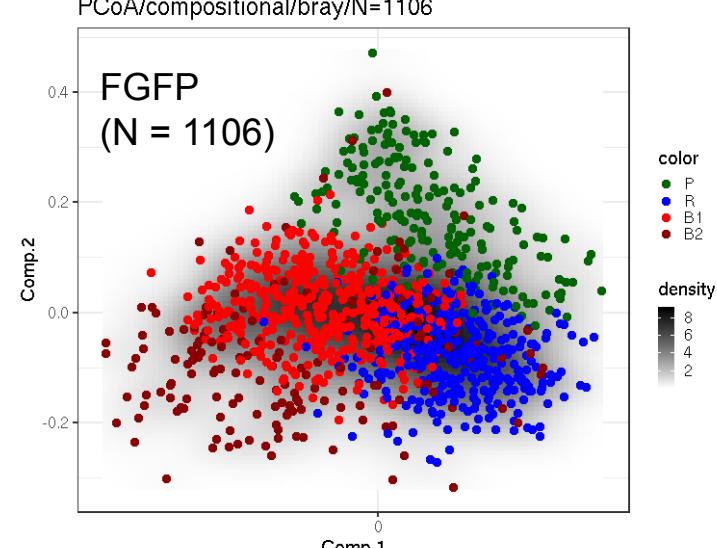
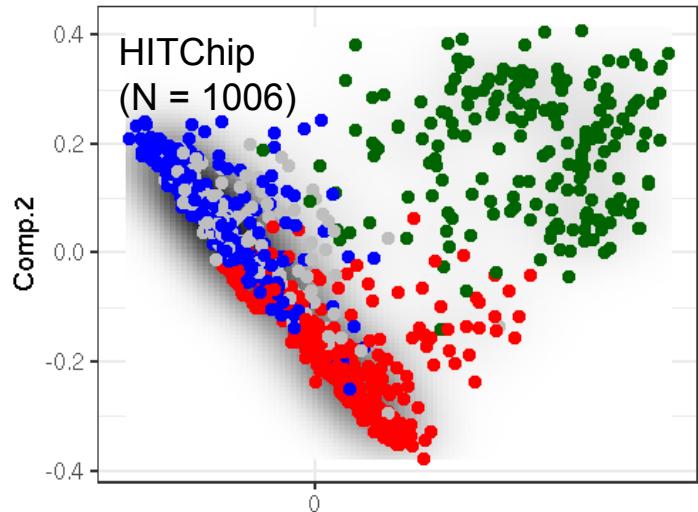
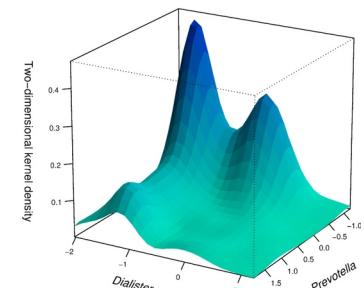
Intestinal microbiome landscaping: insight in community assemblage and implications for microbial modulation strategies FREE

Sudarshan A. Shetty ✉, Floor Hugenholtz, Leo Lahti, Hauke Smidt, Willem M. de Vos

FEMS Microbiology Reviews, Volume 41, Issue 2, 1 March 2017, Pages 182–199, <https://doi.org/10.1093/femsre/fuw045>

Published: 09 February 2017 Article history ▾

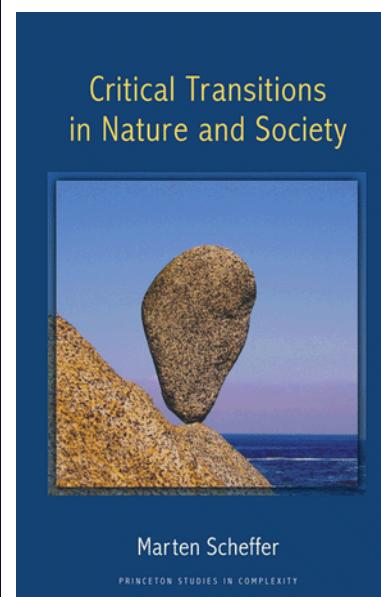
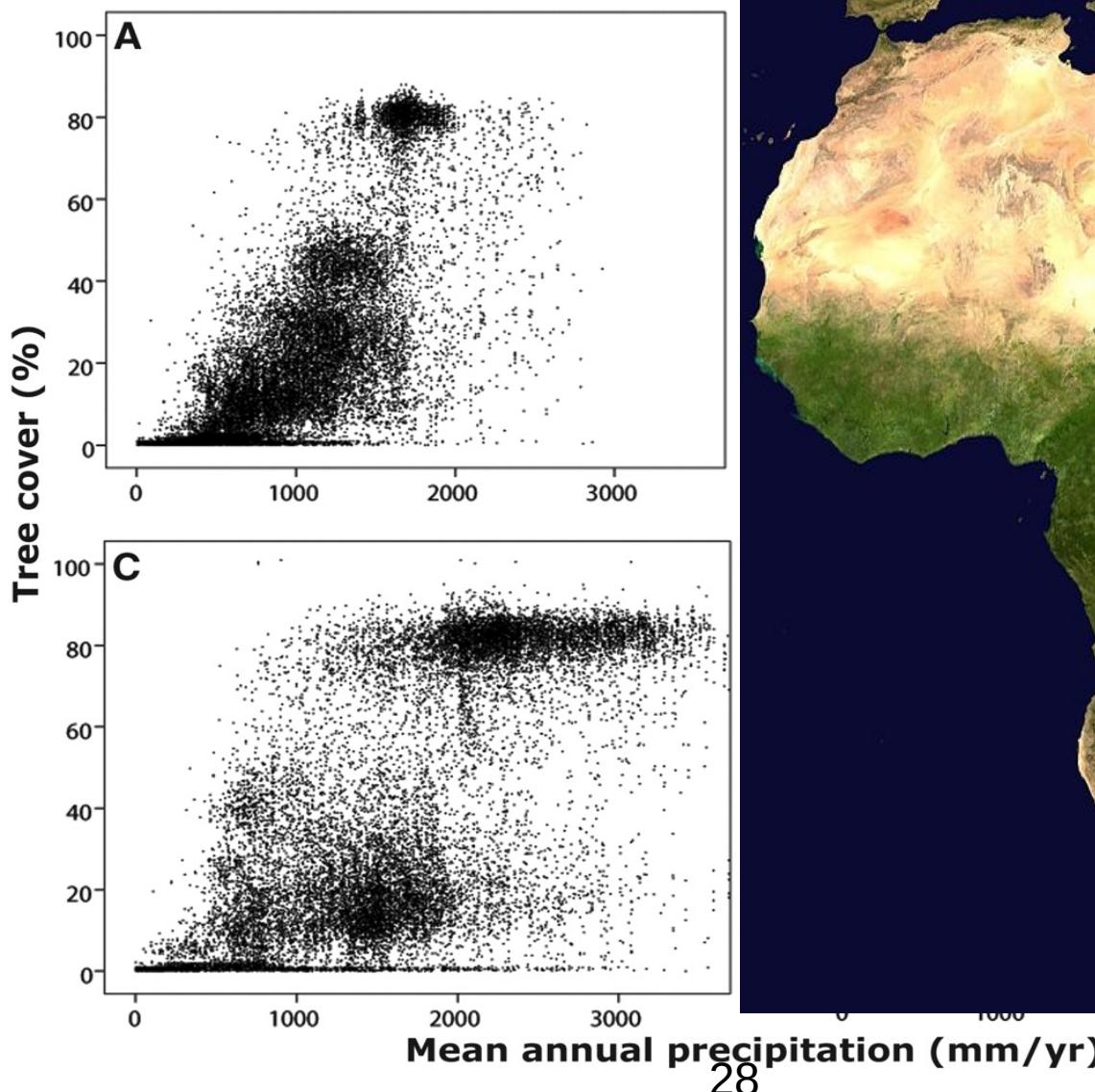
(B)



REPORT

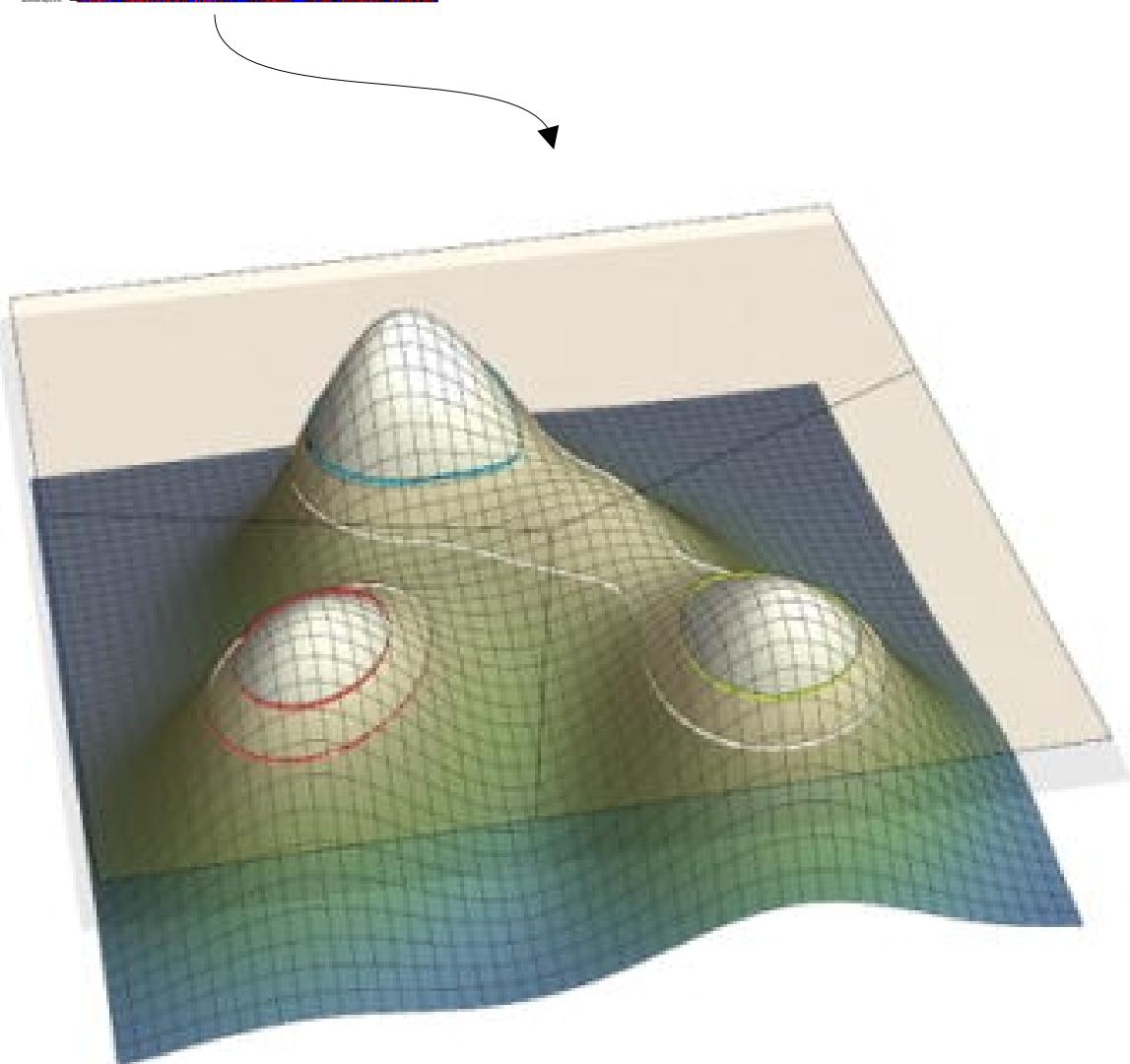
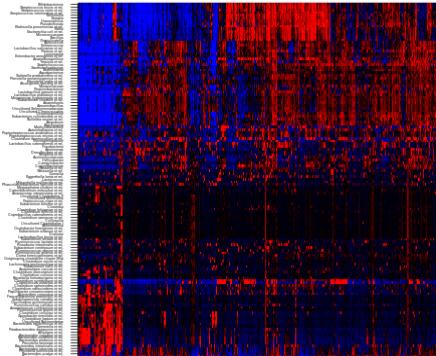
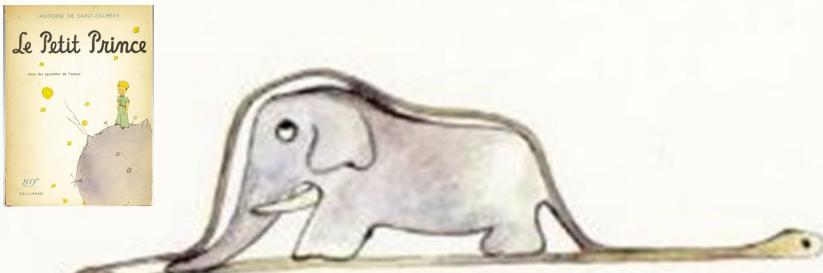
Global Resilience of Tropical Forest and Savanna to Critical Transitions

Marina Hirota¹, Milena Holmgren^{2*}, Egbert H. Van Nes¹, Marten Scheffer¹





Mon dessin ne représentait pas un chapeau. Il représentait un serpent boa qui digérait un éléphant



Perspective | Published: 18 December 2017

Enterotypes in the landscape of gut microbial community composition

Paul I. Costea, Falk Hildebrand, [...] Peer Bork

Nature Microbiology 3, 8–16(2018) | Cite this article

6840 Accesses | 253 Citations | 100 Altmetric | Metrics

Thursday: Differential abundance

Lecture

Differential abundance analysis methods

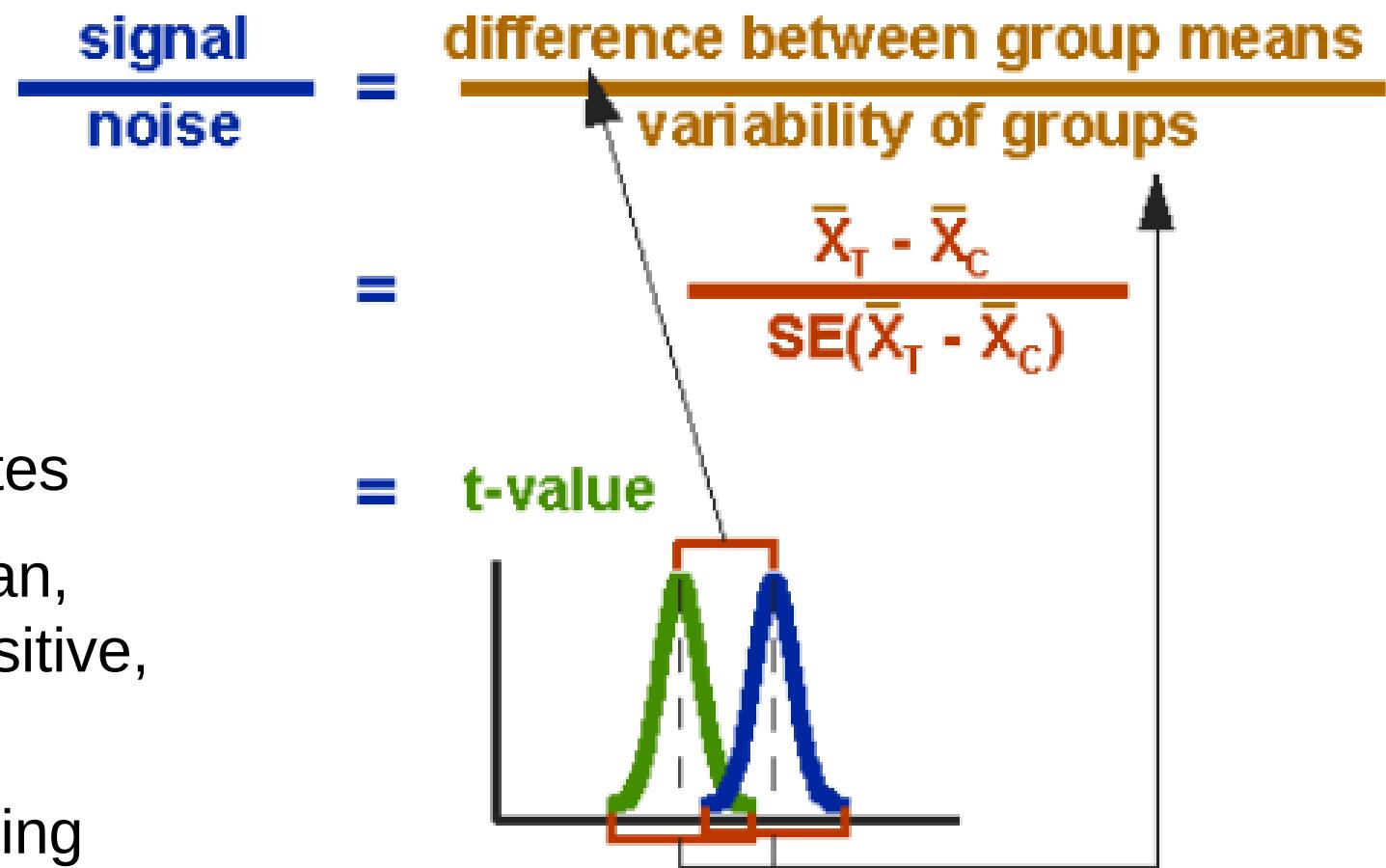
Practical

Differential abundance in practice

Lecture

Overview of microbiota data science methods & concepts

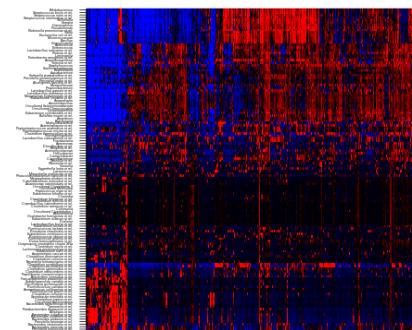
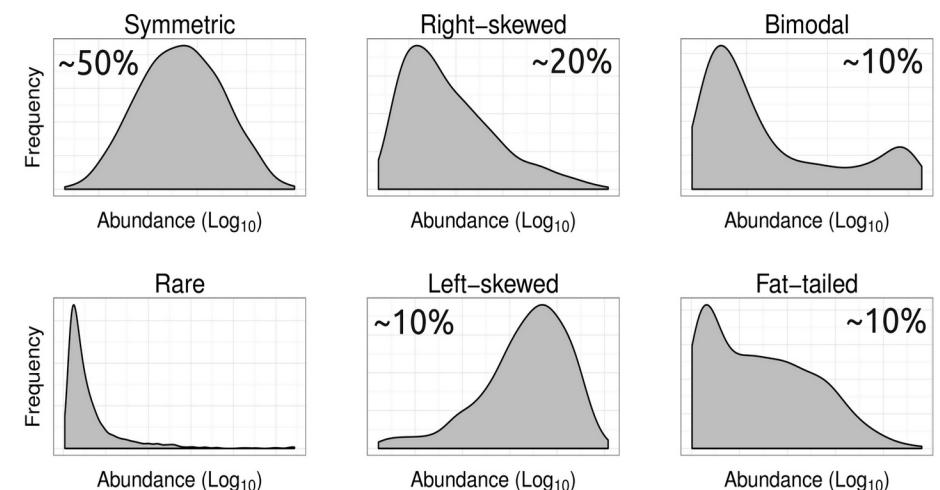
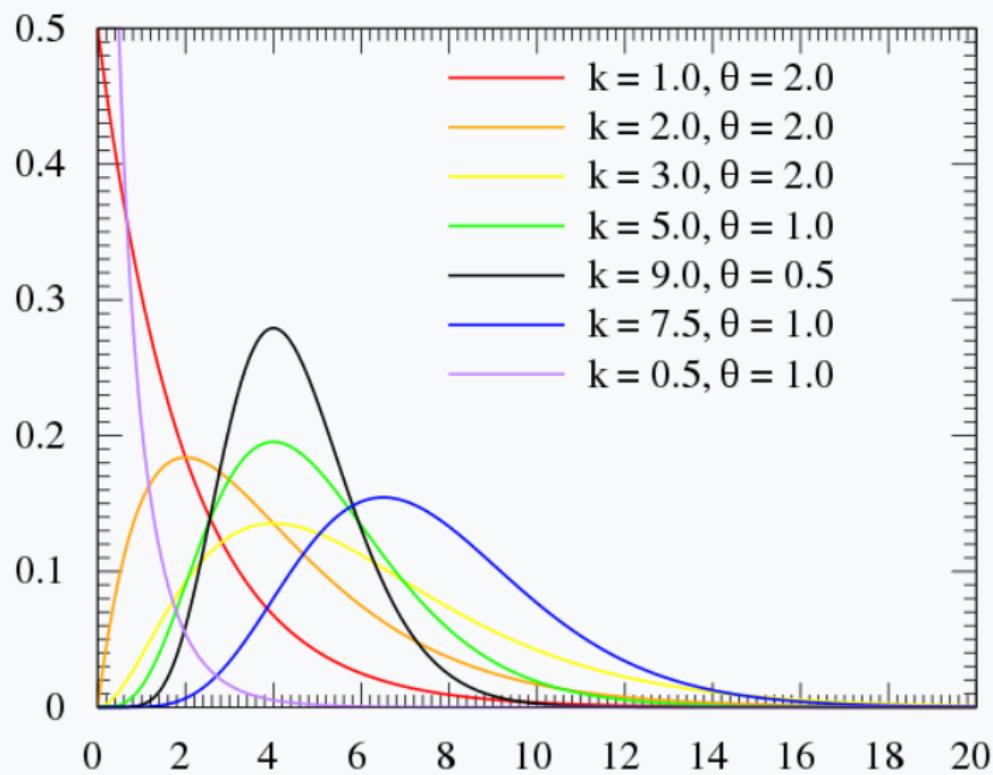
Standard t-test for two-group comparison?



Problems:

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

Gamma distribution gives a flexible model for biological abundance variation



Taylor's law (in HITChip Atlas)

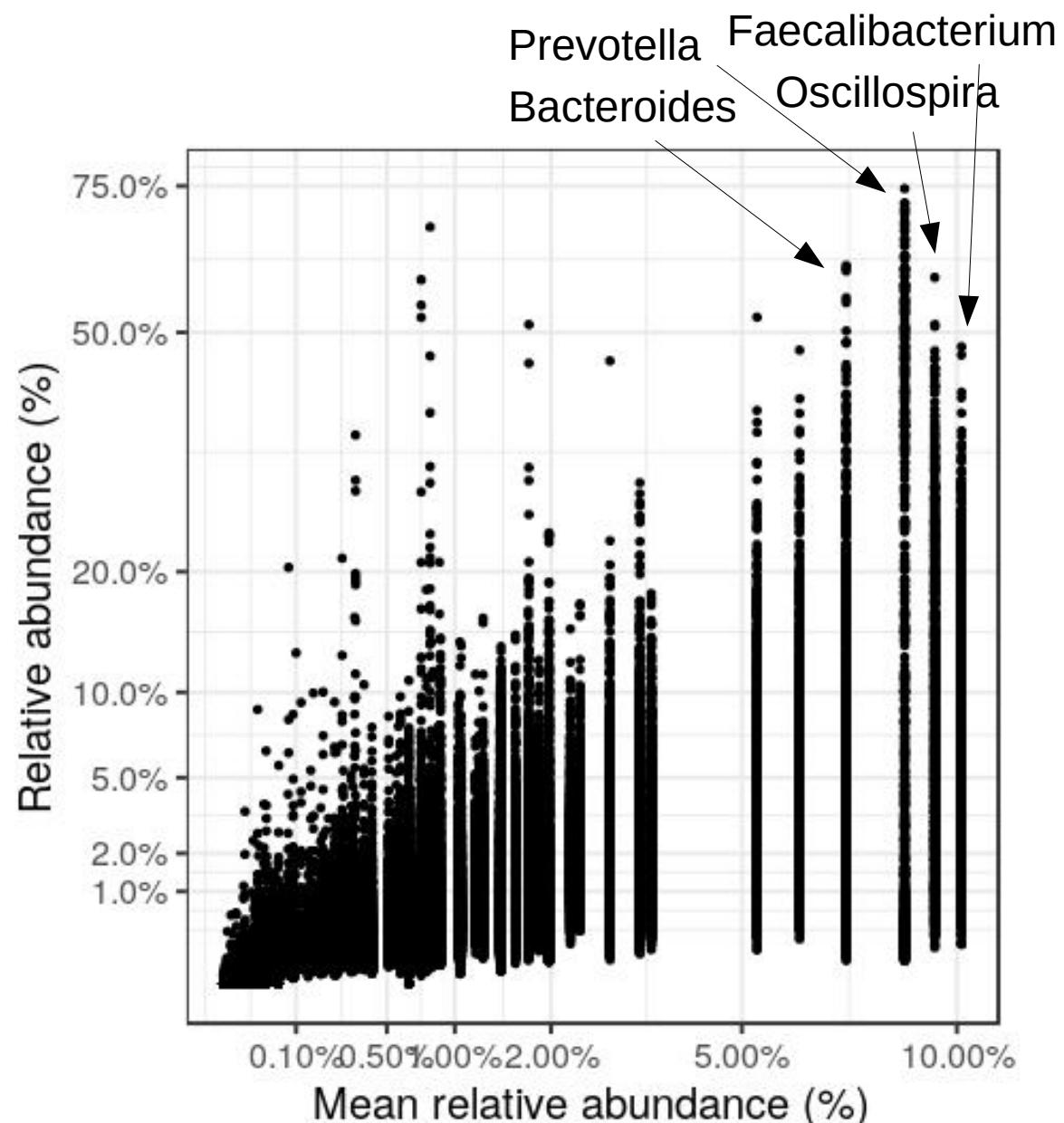
Heteroschedasticity:

Variance increases with the mean

Overdispersion:

Variance increases faster than proposed by the model

Data: HITChip Atlas



Lecture: Key concepts

- special properties of microbiome data
- data science workflows

- alpha diversity
- beta diversity
- differential abundance