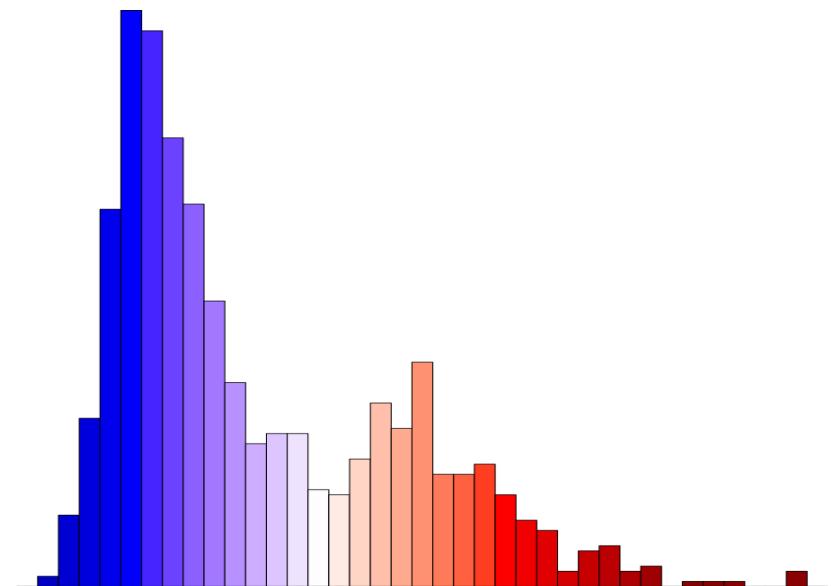


# A roadmap for microbiome bioinformatics in R

Leo Lahti

University of Turku, Finland

[leo.lahti@iki.fi](mailto:leo.lahti@iki.fi) | @antagomir



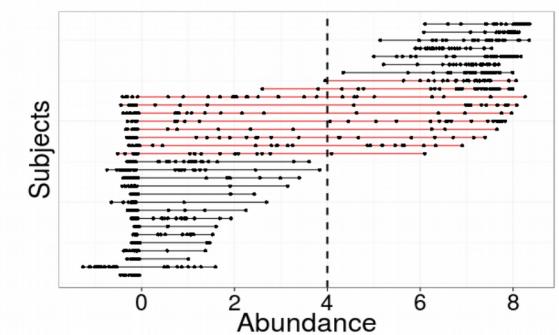
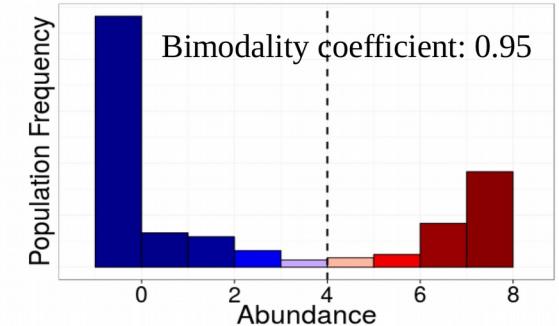
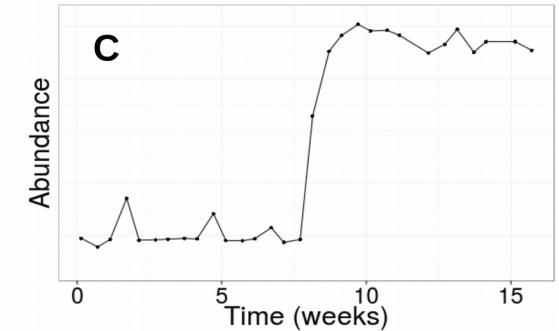
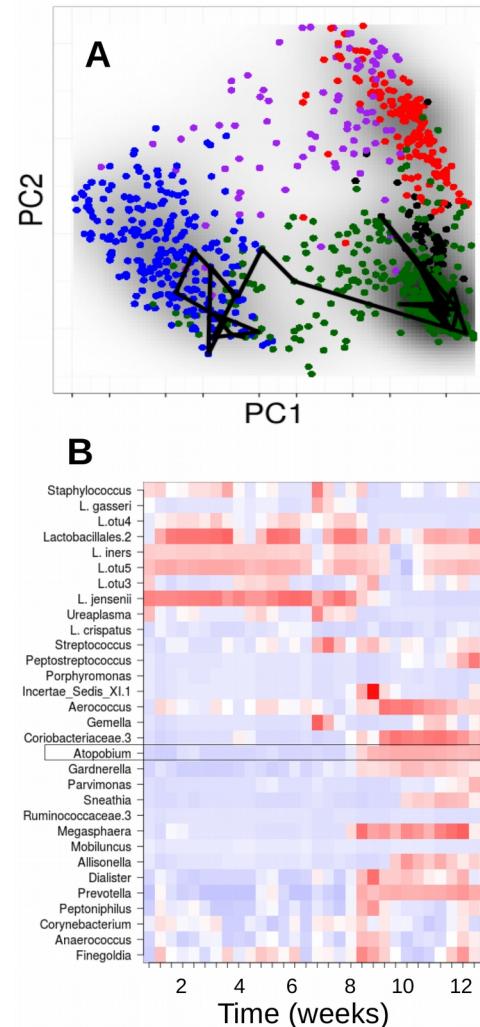
# Visualizing & modeling vaginal microbiome dynamics

Variation:

- cross-sectional
- spatial
- temporal

Levels of analysis:

- ecosystem(s)
- metagenome
- function (metabolome)
- host interactions



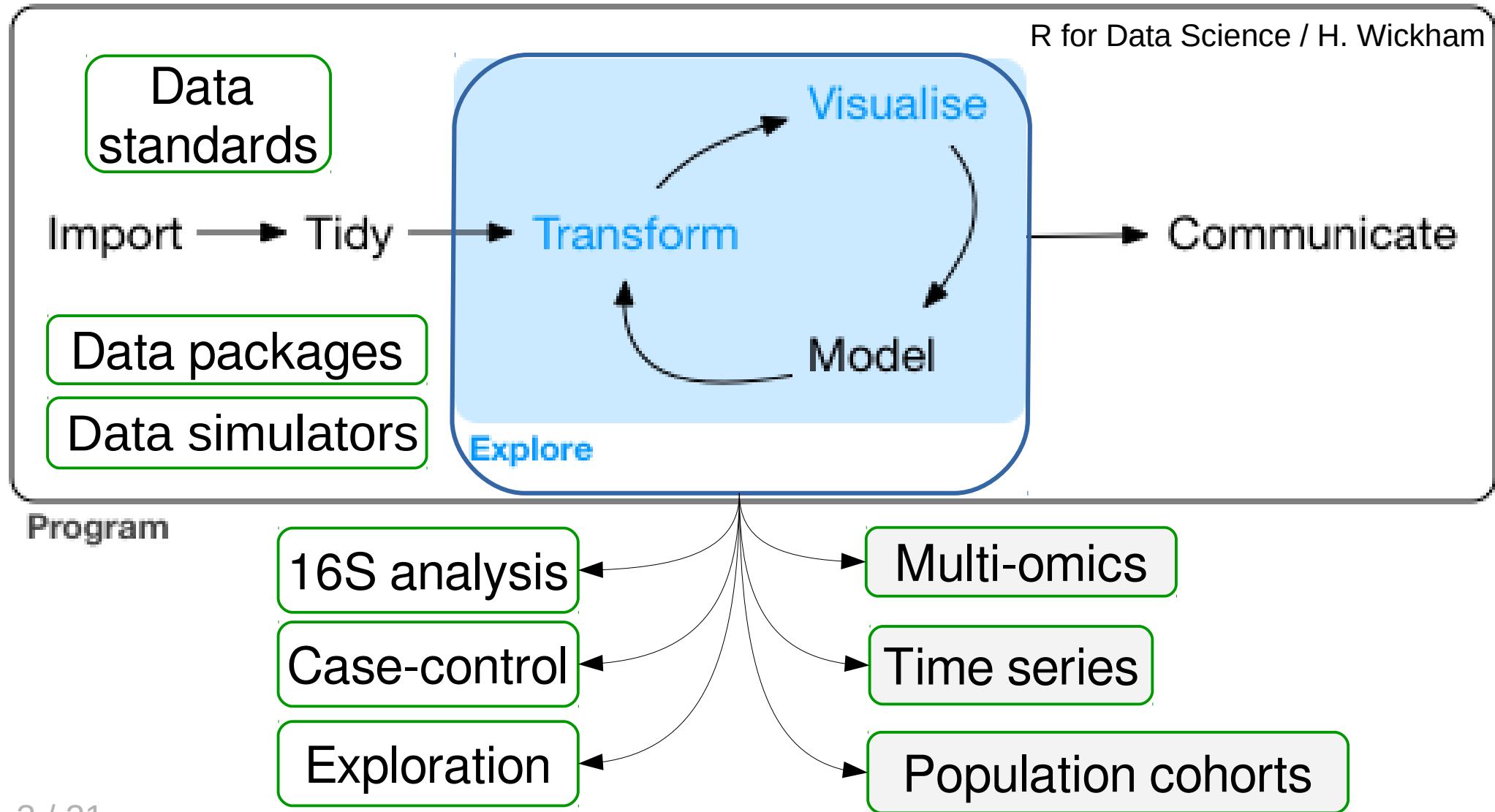
Metagenomics meets time series analysis: unraveling microbial community dynamics

Karoline Faust<sup>1,2,3,9</sup>, Leo Lahti<sup>4,5,9</sup>, Didier Gonze<sup>6,7</sup>,  
Willem M de Vos<sup>4,5,8</sup> and Jeroen Raes<sup>1,2,3</sup>

Data: Gajer *et al.* 2012

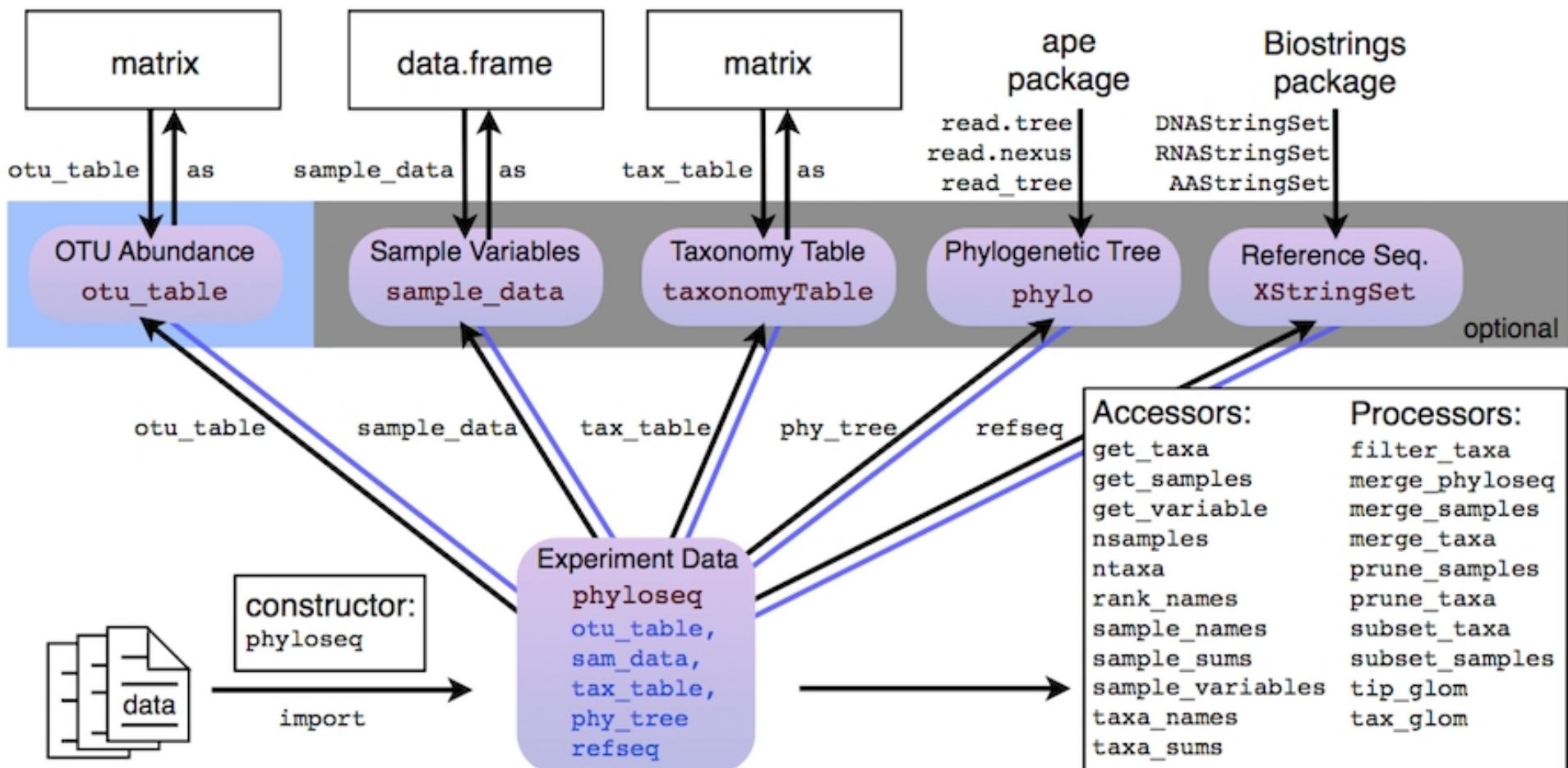
# Microbiome data science

Sudarshan Shetty<sup>1</sup>, Leo Lahti<sup>2\*</sup>



# Data standard: phyloseq

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





RESEARCH ARTICLE

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

Ben J. Callahan<sup>1</sup>, Kris Sankaran<sup>1</sup>, Julia A. Fukuyama<sup>1</sup>, Paul J. McMurdie<sup>2</sup>, Susan P. Holmes<sup>1</sup>

Author details



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



METRICS

38052

VIEWs

5288

DOWNLOADS

Get PDF

Get XML

Cite

## Normalization and microbial differential abundance strategies depend upon data characteristics

Sophie Weiss, Zhenjiang Zech Xu, Shyamal Peddada, Amnon Amir, Kyle Bittinger, Antonio Gonzalez, Catherine Lozupone, Jesse R. Zaneveld, Yoshiaki Vázquez-Baeza, Amanda Birmingham, Embriette R. Hyde and Rob Knight

*Microbiome* 2017 5:27

<https://doi.org/10.1186/s40168-017-0237-y> | © The Author(s). 2017

Received: 9 October 2015 | Accepted: 27 January 2017 | Published: 3 March 2017



# microbiome R package

chat on gitter

build passing

codecov 24%

PRs welcome



## Complementing *phyloseq*

### Data

- Population cohorts
- Time series
- Interventions
- Multi-omics

### Utilities

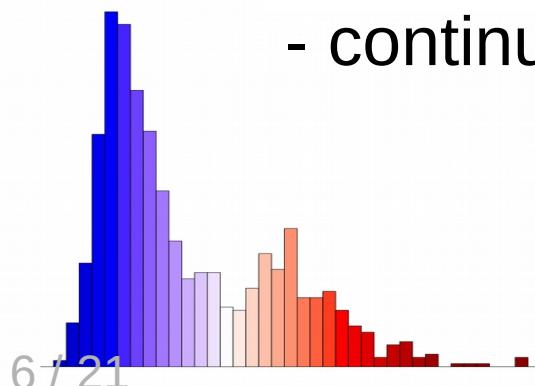
- transformations
- alpha & beta diversity
- core microbiota
- visualizations

### Analysis & modeling

- stability analysis
- tipping elements

### Quality control:

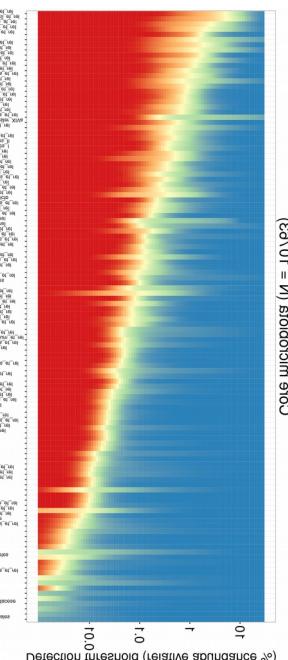
- unit tests
- continuous integration..



### Community:

- mailing list
- gitter
- workshops

<http://microbiome.github.io>



# Example Data



**Population cohorts** Gut microbiome profiles of 1006 normal western adults. HITChip Atlas / Wageningen. (Lahti *et al.* *Nat. Comm.* 2014)

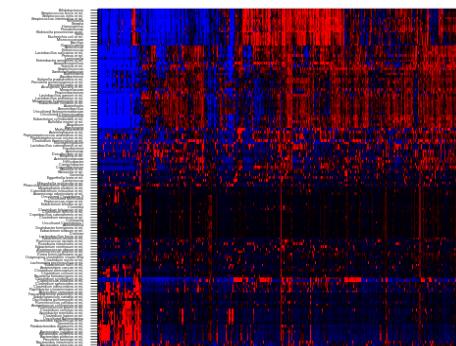
**Time series** Vaginal microbiome (Ravel *et al.* 2013); Dense gut microbiome (David *et al.* 2014)

**Interventions** African-American diet swap study. (O'Keefe et al. *Nat. Comm.* 2015)

# **Multi-omics Gut microbiome – serum lipid association study.**

**Lahti *et al.* PeerJ 2013**

**Data simulators** with varying assumptions on ecosystem composition and dynamics



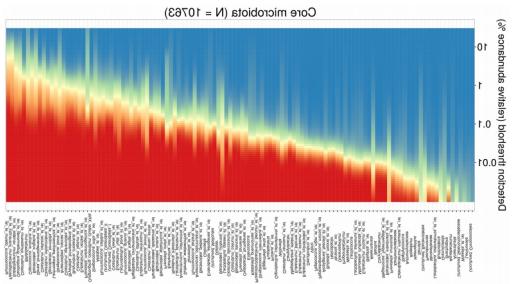
# Some utilities

## Core & prevalence

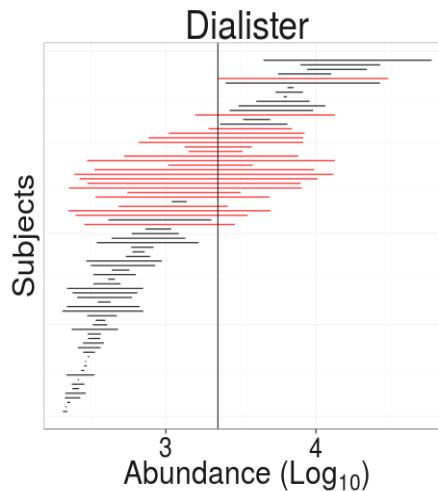
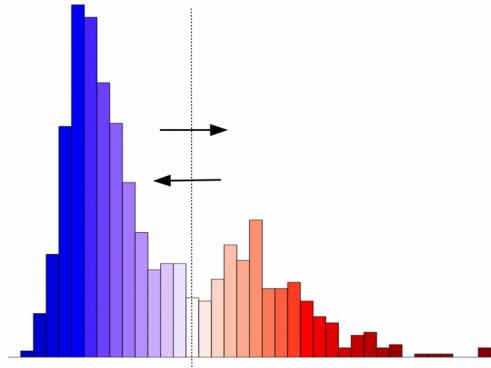
`prevalence(x)`

`core(x)`

`core_members(x)`



## Stability & resilience



## Alpha & beta diversity

`alpha(x)`

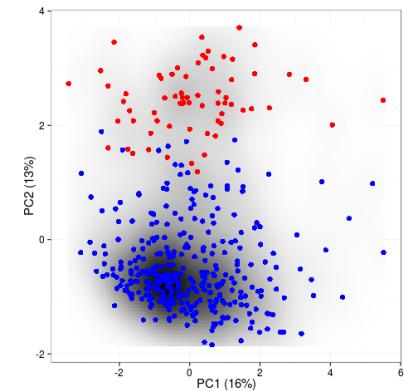
`diversity(x)`

`evenness(x)`

`dominance(x)`

`rarity(x)`

`readcount(x)`



## Transformations

`transform(x, "compositional")`

`transform(x, "clr")`

`transform(x, "log10p")`

`transform(x, "hellinger")`

`transform(x, "identity")`

## Package website

- Tutorials & support
- Data & code sharing

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. qimper 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

**Other available R tools?**  
a survey for 16S  
Github.com/microsud/  
Tools-Microbiome-Analysis

**Challenges:**

- sparse
- non-gaussian
- count data
- complex
- multi-level

# Packages

dplyr

tidyr

ggplot

# Package collections

Data science:  
tidyverse



microbiome

earlywarnings

seqtime

# Distribution ecosystems

CRAN



Microbiomics:  
[microbiome.github.io](http://microbiome.github.io)

# Synergies in package development



Sequence  
variants

Samples

$N \sim 1000$

Scaling up  
from individuals  
to populations

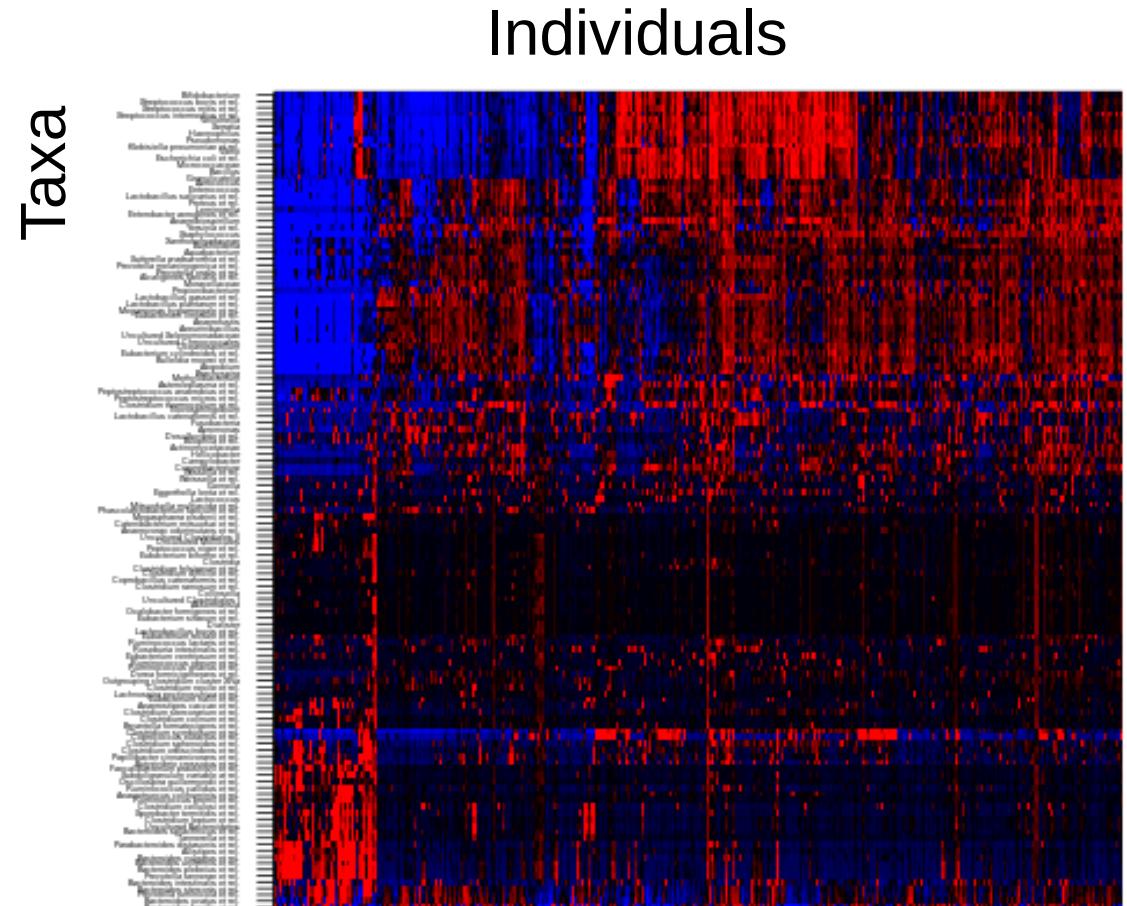
# Data: Human Intestinal Tract (HIT)Chip Atlas: 100 genera ~ 1000 samples



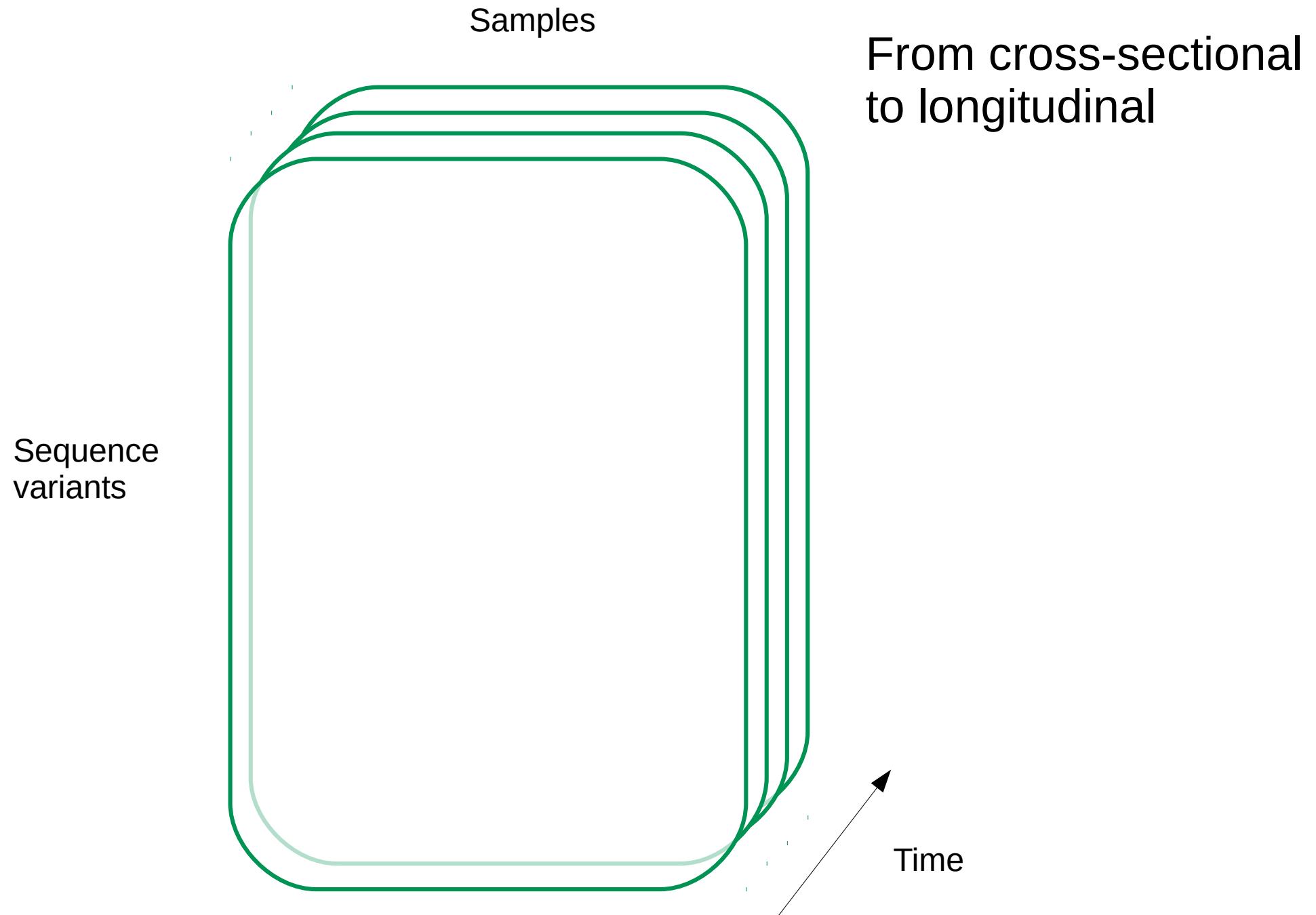
Gut microbiota: 1000 western adults (Lahti *et al.* Nature Comm. 2014)

## Other example data

- Population cohorts
- Time series
- Interventions
- Multi-omics



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



# Early warning signals to predict state shifts ?

Early warning signals for a critical transition in a time series generated by a model of a harvested population<sup>77</sup> driven slowly across a bifurcation.

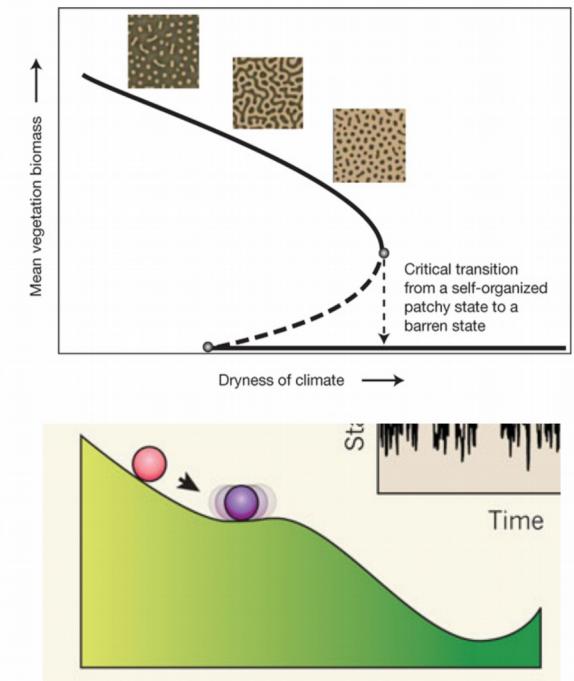
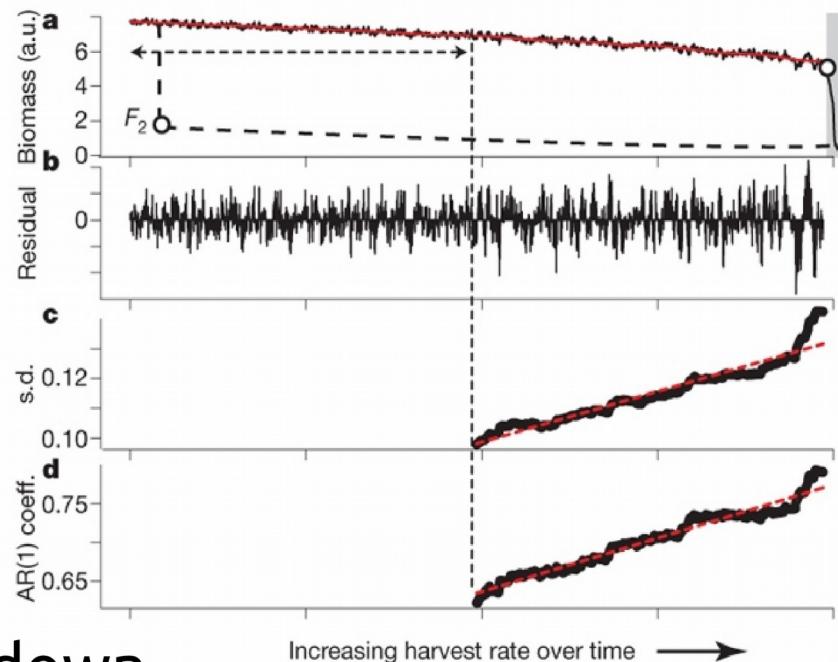
Autocorrelation

Variance

Skewness

Flickering

Critical slowing down



M Scheffer Nature 2010

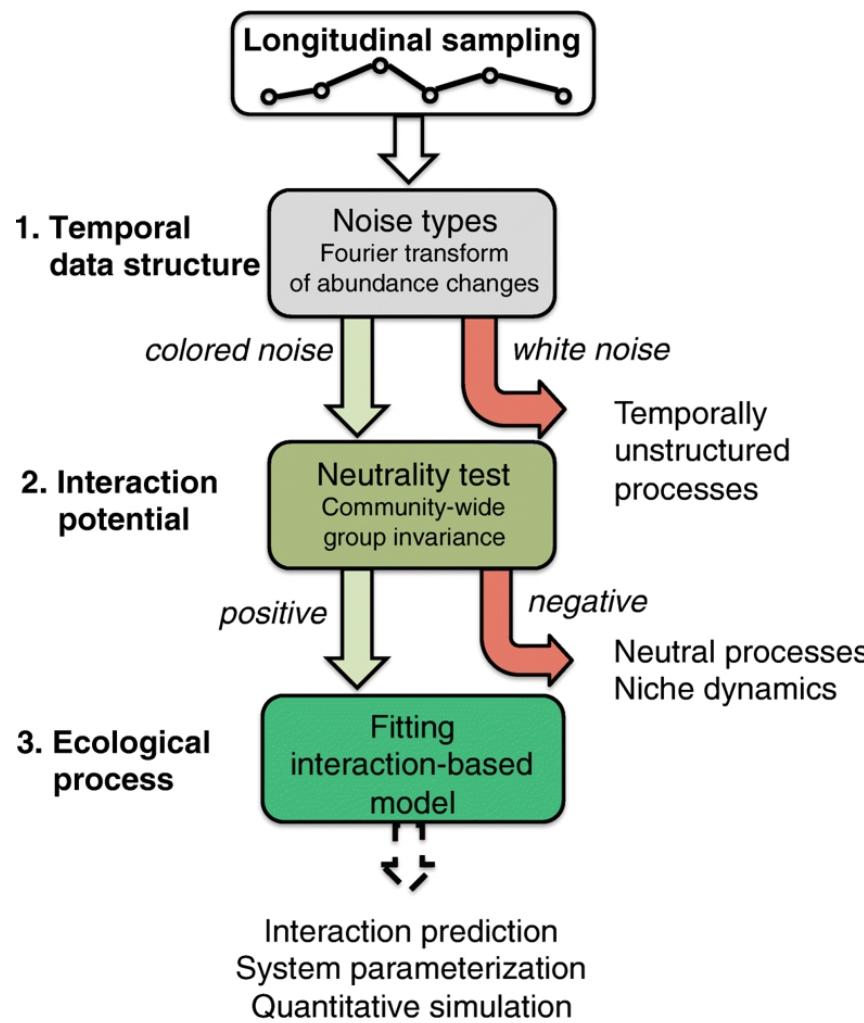
M Scheffer et al. *Nature* **461**, 53-59 (2009) doi:10.1038/nature08227

CRAN: earlywarnings R package  
V. Dakos & L. Lahti

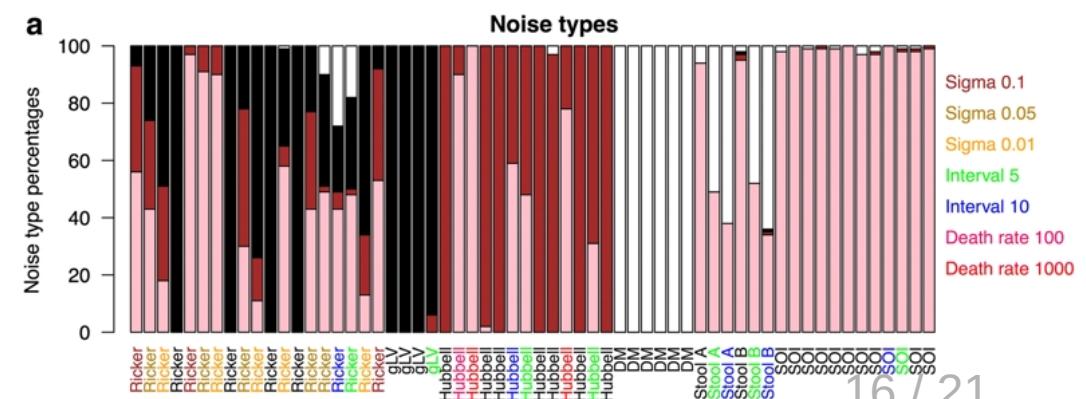
nature

# Signatures of ecological processes in microbial community time series

Karoline Faust<sup>1\*</sup> , Franziska Bauchinger<sup>2</sup>, Béatrice Laroche<sup>3</sup>, Sophie de Buyl<sup>4,5</sup>, Leo Lahti<sup>1,6,7</sup>, Alex D. Washburne<sup>8,9</sup>, Didier Gonze<sup>5,10</sup> and Stefanie Widder<sup>11,12,13\*</sup>



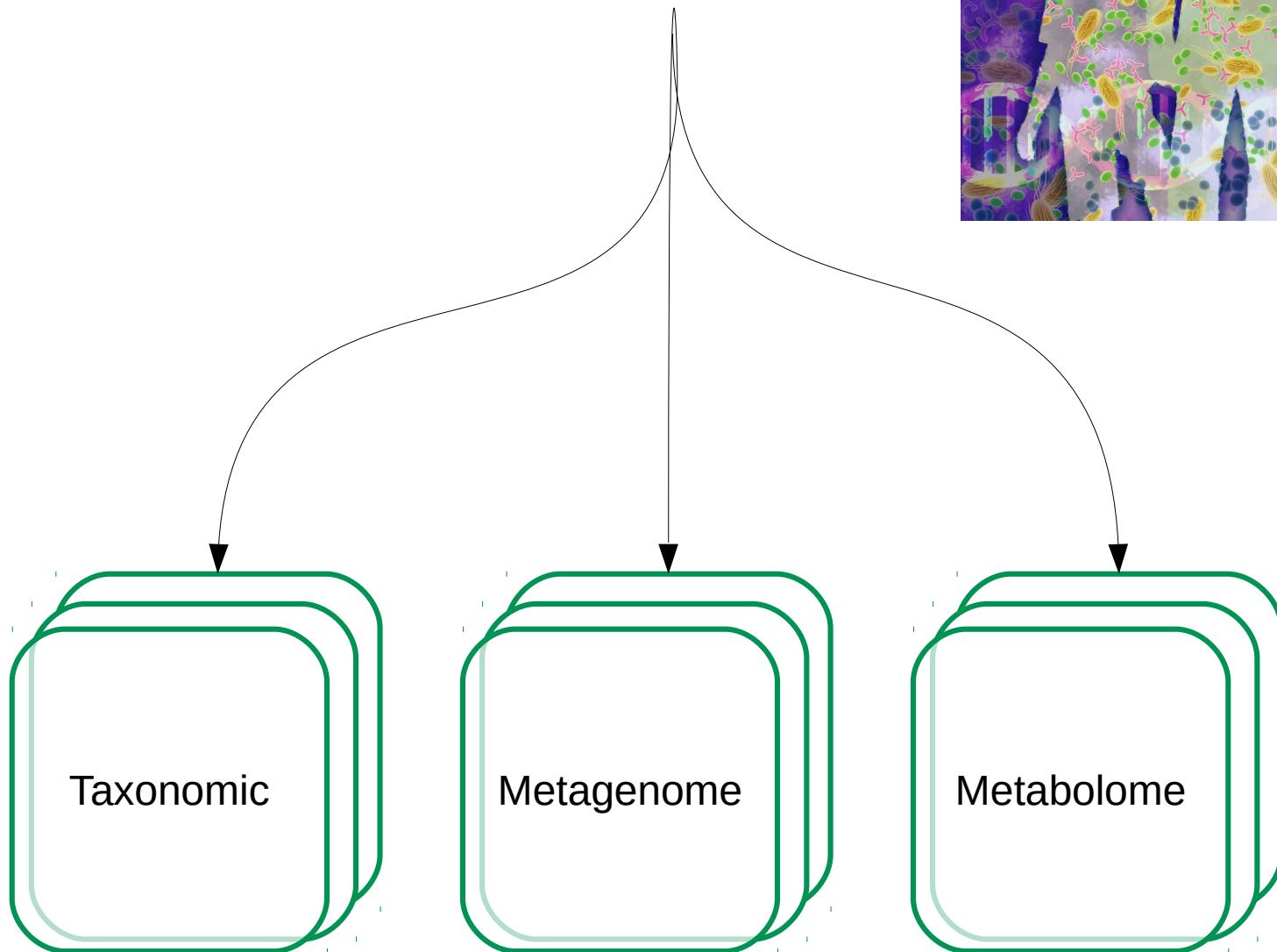
White noise → Unstructured  
 Pink noise → Temporal structure  
 Brown noise → Neutrality  
 Black noise → Interactions



Towards multi-assay experiments,  
biogeography,  
longitudinal population cohorts..



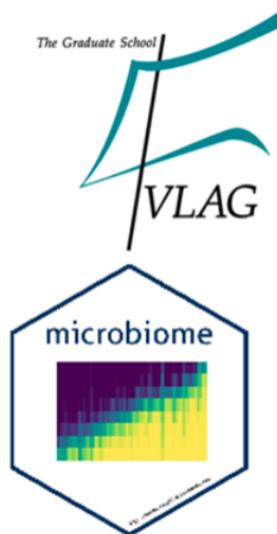
Fig: NHGRI



# Tutorials / Education

OPEN & REPRODUCIBLE MICROBIOME DATA ANALYSIS  
SPRING SCHOOL

Wageningen, The Netherlands, May 28-30, 2018



Turun yliopisto  
University of Turku

Three day spring school on basics of high throughput 16S rRNA gene sequencing data analysis. This spring school is organised in

## Getting started

- Overview (vignette)
- Installation
- Example data
- Data manipulation

## Microbiome analysis

- Alpha diversity
- Beta diversity / Community heterogeneity
- Community composition
- Core microbiota
- Landscapes (population density analysis)
- Stability and tipping elements

### Tutorials

- Examples
- Workflows
- Literature
- Educational resource

### Blogs

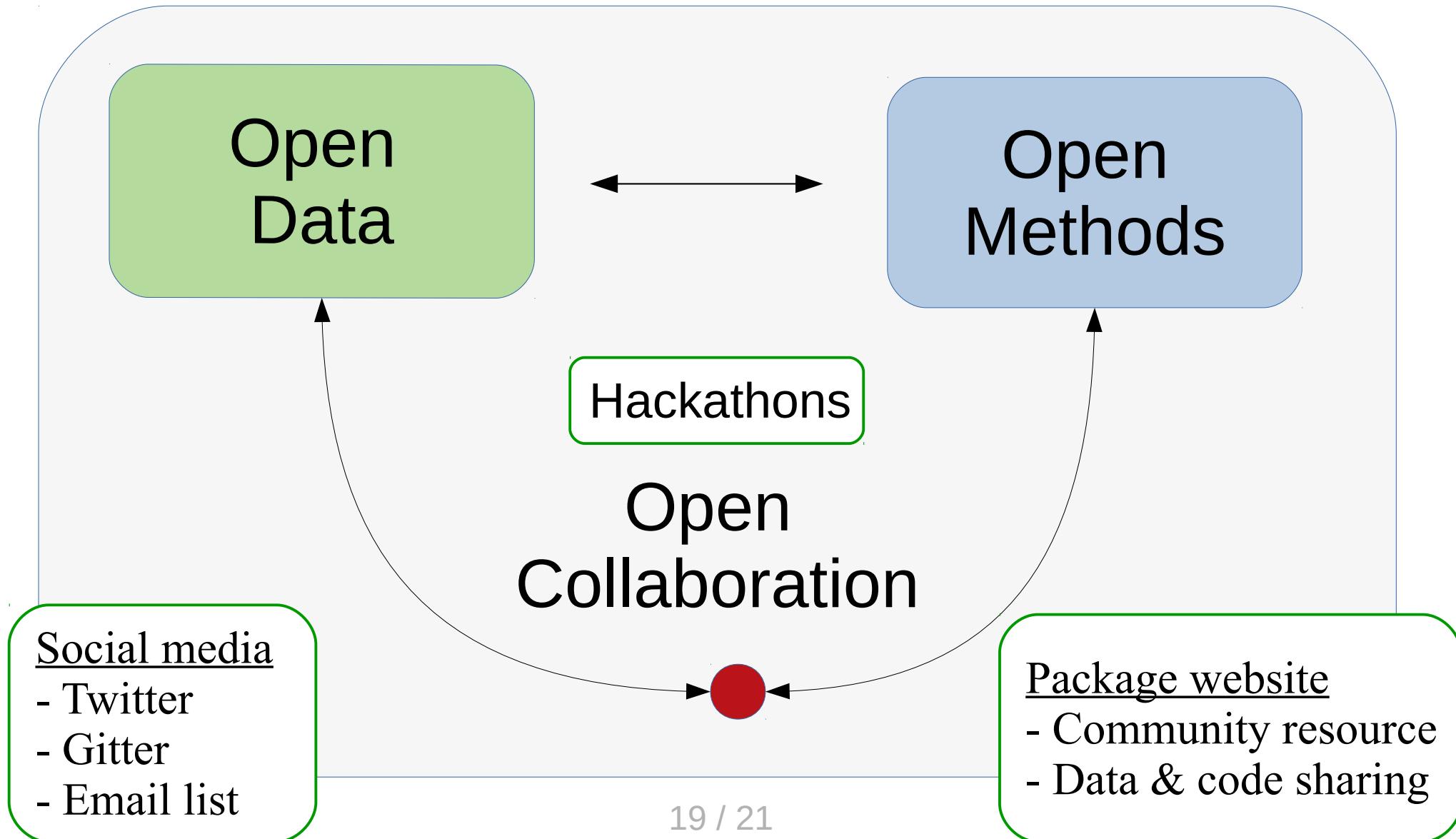
- Data & code sharing
- Support for users
- PR

### Function help

- Inputs
- Outputs
- Explanation
- Examples

# Open data science

Leo Lahti. In: Advances in Intelligent Data Analysis XVII.  
Lecture Notes in Computer Science 11191., Springer Nature, 2018.



Special thanks:

Sudarshan Shetty

Ville Laitinen

Aaro Salosensaari

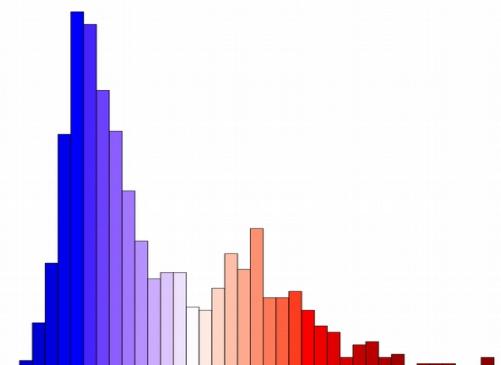
Karoline Faust

Umer Ijaz

Willem de Vos

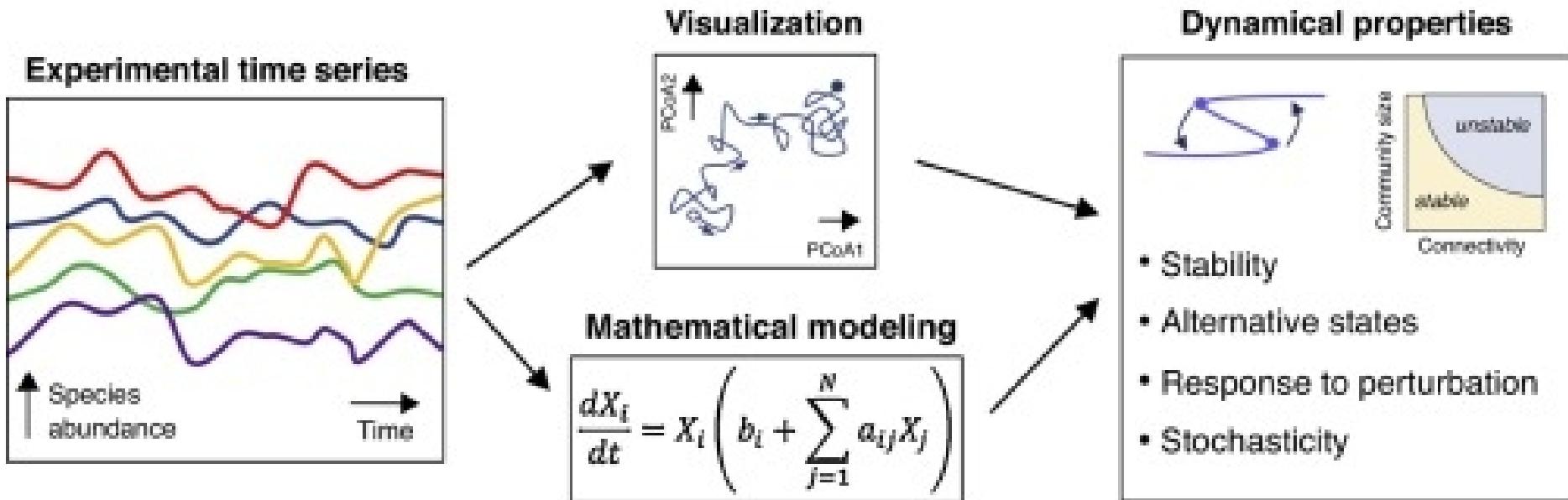
R package:  
<http://microbiome.github.io>

Twitter:  
@openreslabs



# Microbial communities as dynamical systems

Didier Gonze <sup>1, 2</sup>✉, Katharine Z Coyte <sup>3, 4</sup>, Leo Lahti <sup>5, 6, 7</sup>, Karoline Faust <sup>5</sup>✉



## Metagenomics meets time series analysis: unraveling microbial community dynamics

Karoline Faust<sup>1,2,3,9</sup>, Leo Lahti<sup>4,5,9</sup>, Didier Gonze<sup>6,7</sup>,  
Willem M de Vos<sup>4,5,8</sup> and Jeroen Raes<sup>1,2,3</sup>

## A fully scalable online pre-processing algorithm for short oligonucleotide microarray atlases

[Leo Lahti](#), [1,2,\\* Aurora Torrente](#), [3,4 Laura L. Elo](#), [5,6 Alvis Brazma](#), [3](#) and [Johan Rung](#)<sup>3</sup>

## 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

## Tipping elements in the human intestinal ecosystem

Leo Lahti<sup>1,2</sup>, Jarkko Salojärvi<sup>1,\*</sup>, Anne Salonen<sup>3,\*</sup>, Marten Scheffer<sup>4</sup> & Willem M. de Vos<sup>1,2,3</sup>

Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical Dirichlet process 1  
Keith Harris<sup>1</sup>, Todd L Parsons<sup>2</sup>, Umer Z Ijaz<sup>3</sup>, Leo Lahti<sup>4</sup>, Ian Holmes<sup>5</sup>, Christopher Quince<sup>6,\*</sup>