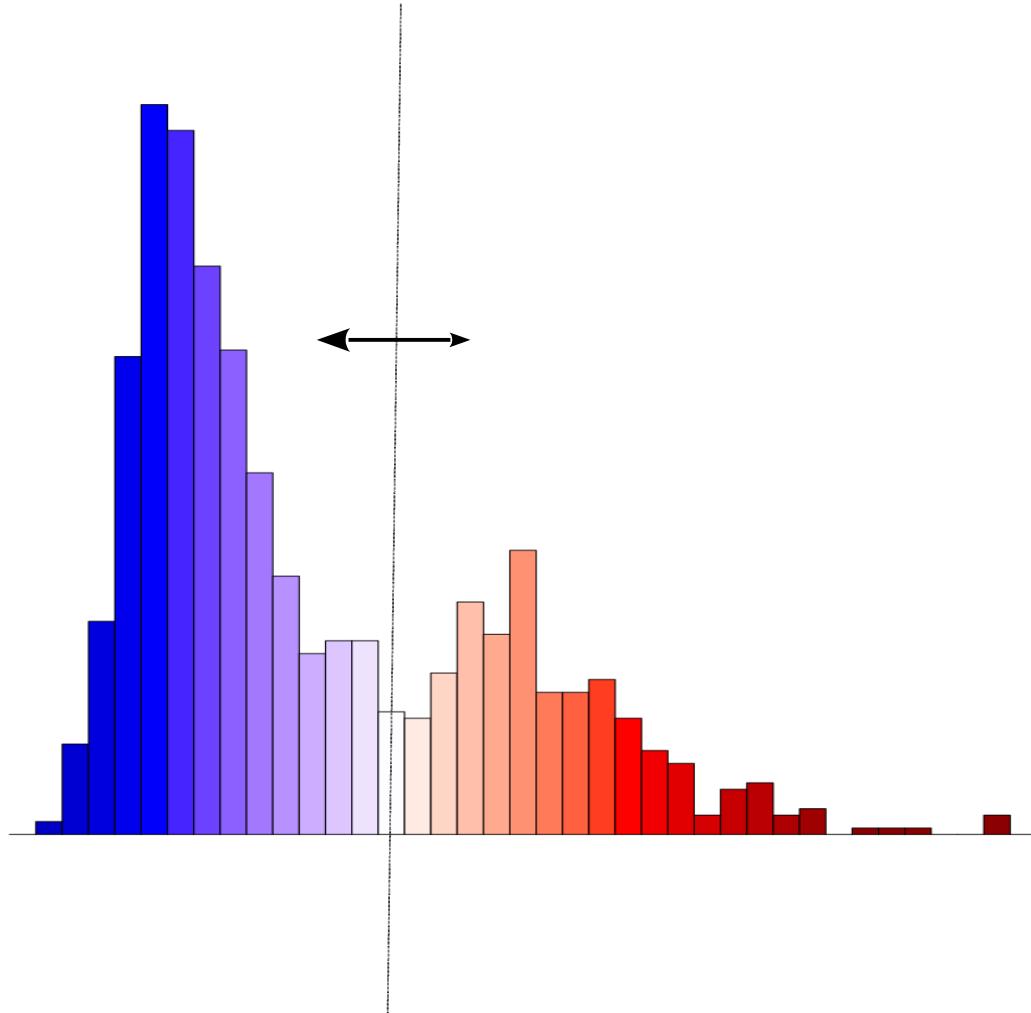


Quantitative models of microbiome variation



Leo Lahti

Dpt Mathematics and Statistics
University of Turku, Finland
leo.lahti@iki.fi | @antagomir

Cross-sectional
population (cohort) studies

Prospective
long-term follow-ups

Longitudinal
ecosystem dynamics

Microbiome
data science

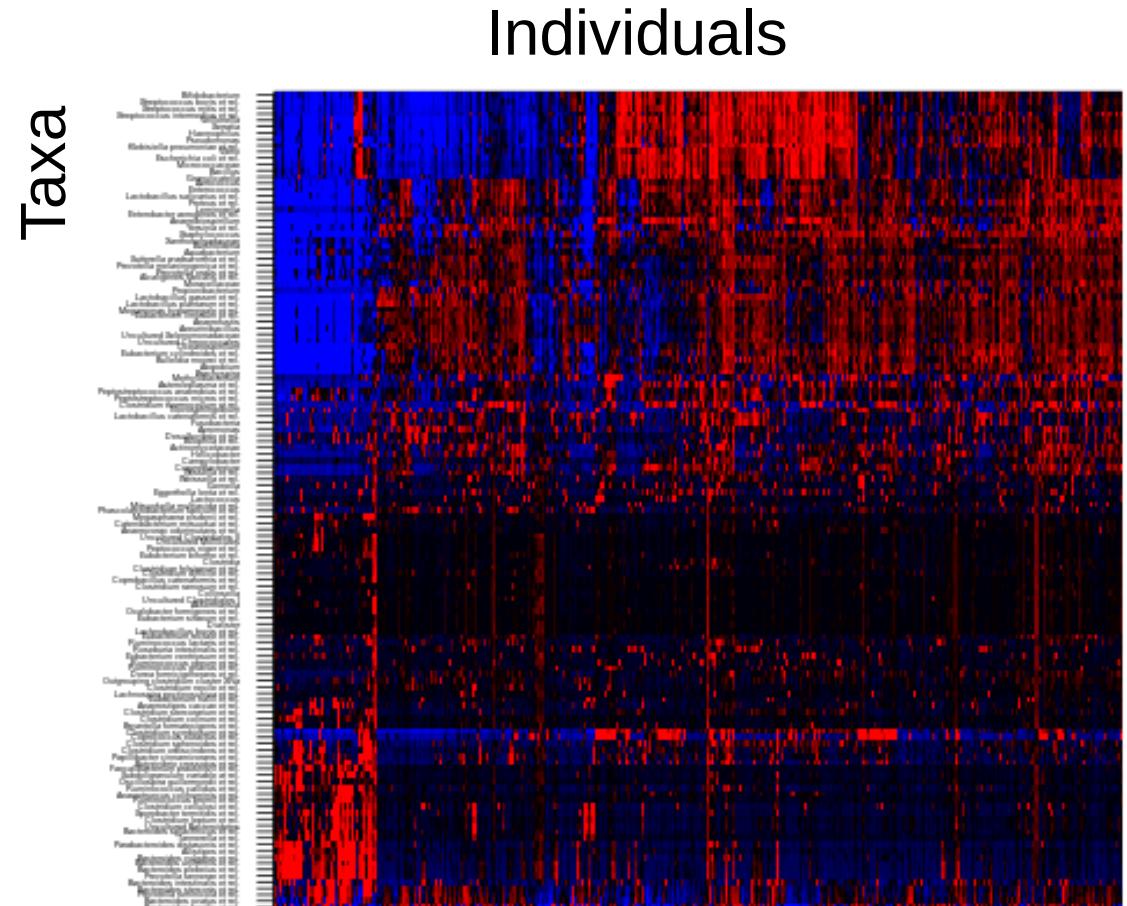
Human Intestinal Tract (HIT)Chip Atlas: 100+ genera ~ 10,000+ samples



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

Background data for

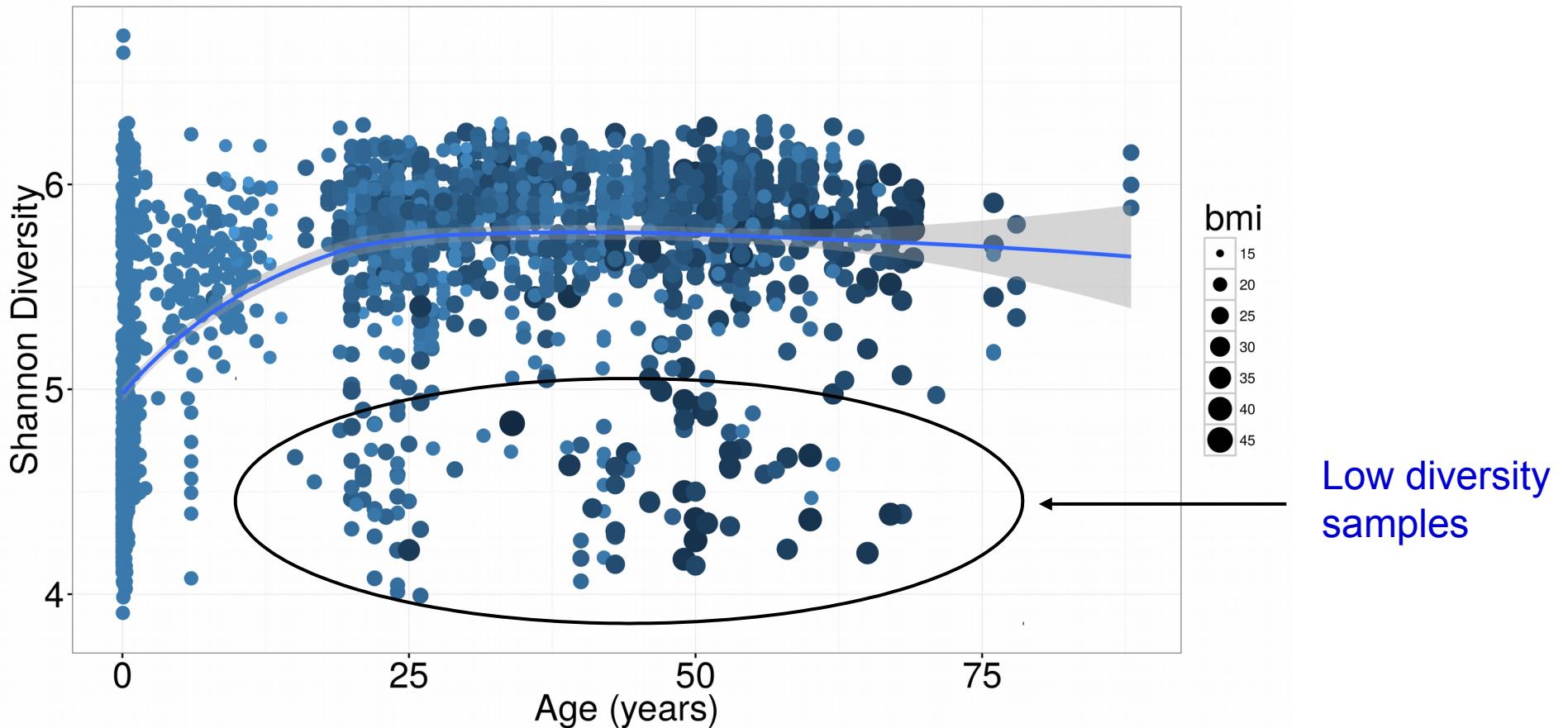
- Time series
- Interventions
- Multi-omics



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



15 year prospective view (Finland / FINRISK2002)

2002

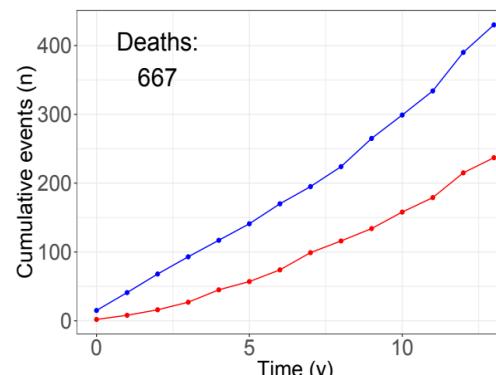
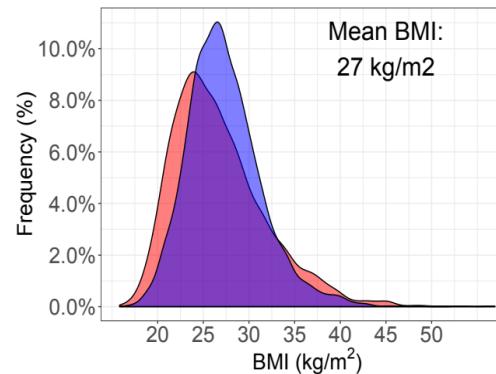
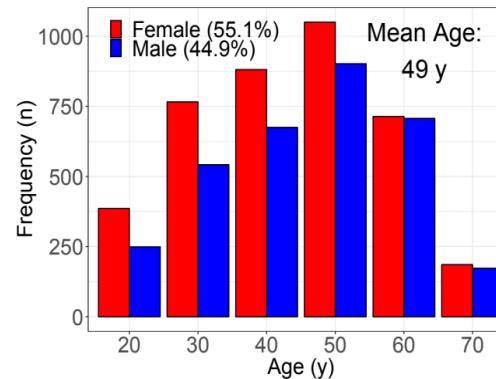
~7000+ adult stool samples:
omics and clinical
measurements.

2016

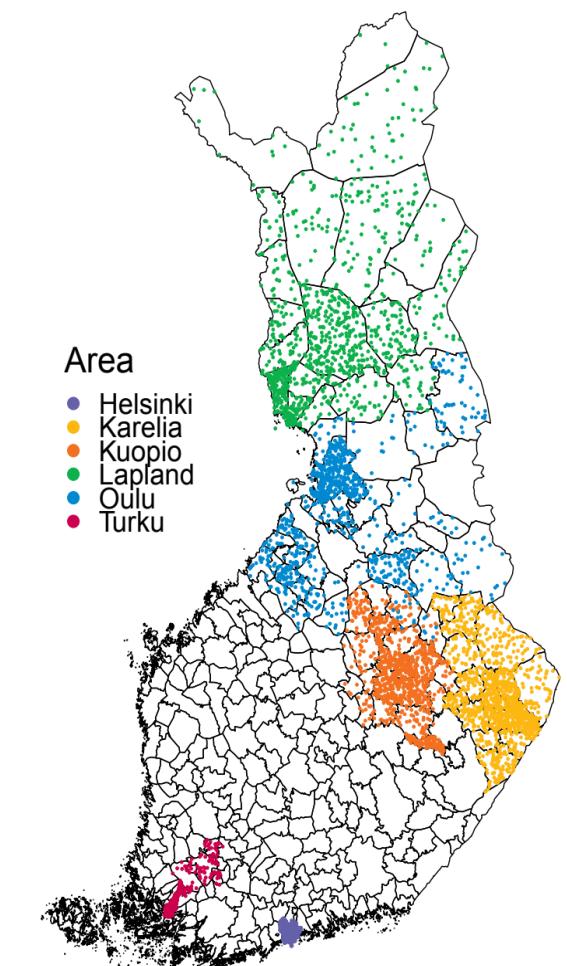
comprehensive health
information from Finnish
health registers

15 year follow up

16S, metagenome (1M
reads/sample), metabolome



N=7231



With: Teemu Niiranen, Veikko Salomaa, Rob Knight, Mohit Jain, Mike Inouye, Aaro Saloenssaari, Ville Laitinen, and others



NATIONAL INSTITUTE
FOR HEALTH AND WELFARE

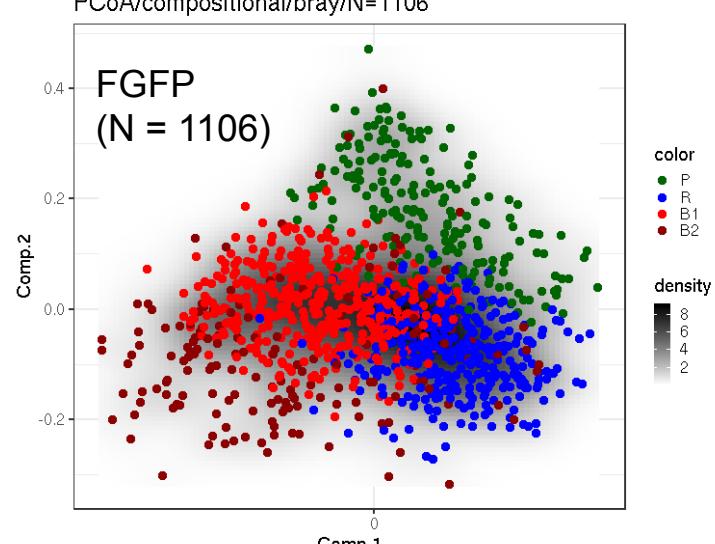
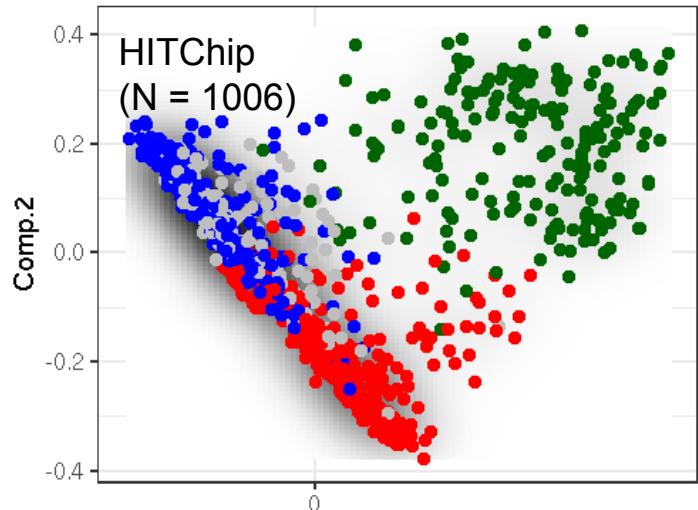
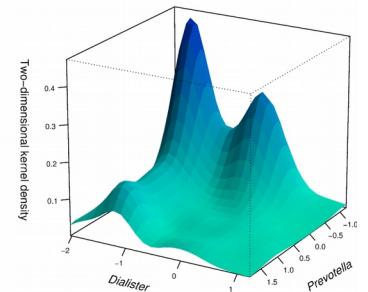
Intestinal microbiome landscaping: insight in community assemblage and implications for microbial modulation strategies

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)



Fundamental considerations in beta diversity analysis

Feature selection

(all/core taxa; genus/strain level..?)

Transformation

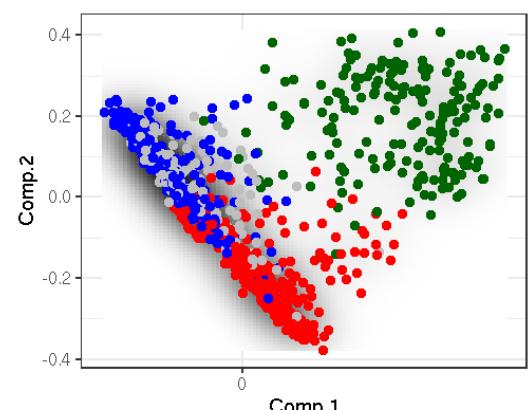
(absolute, compositional, CLR, Hellinger..?)

Dissimilarity measure

(Euclidean/L2, Bray-Curtis, Unifrac..?)

Analysis method

(PCA, PCoA, NMDS, t-SNE, UMAP..)



Why (nonlinear) PCoA and not (linear) PCA?

PCoA Principal Coordinates Analysis (a.k.a MDS)

Transformation: compositional

Dissimilarity: Bray-Curtis

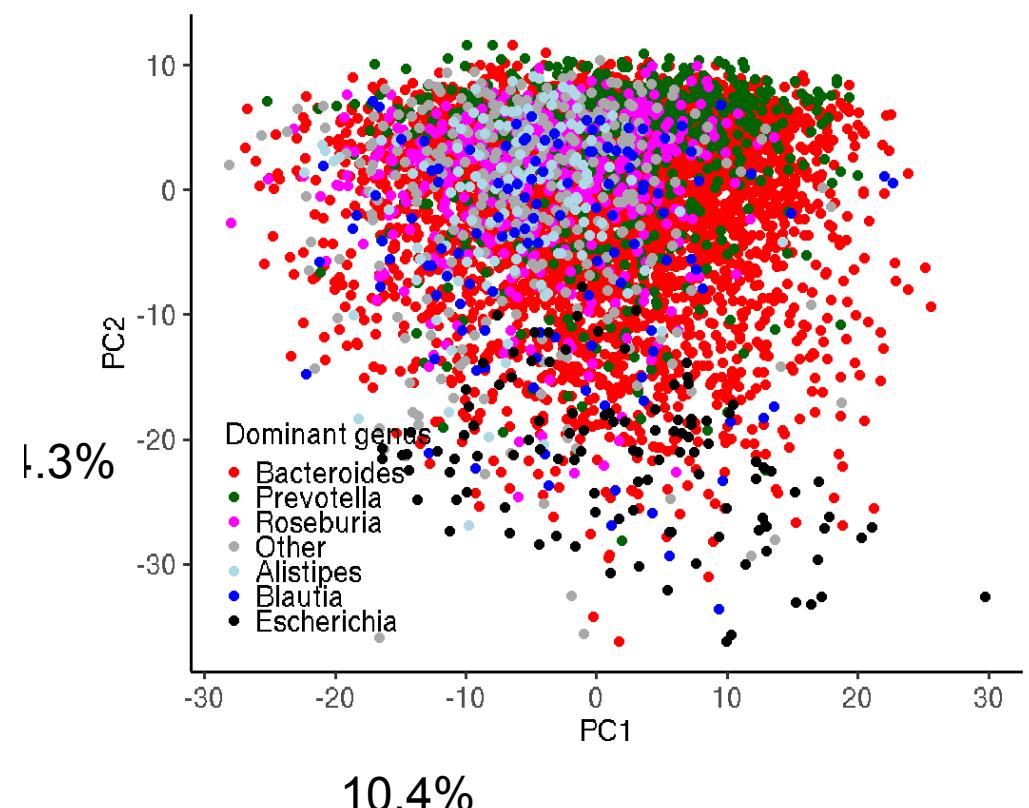
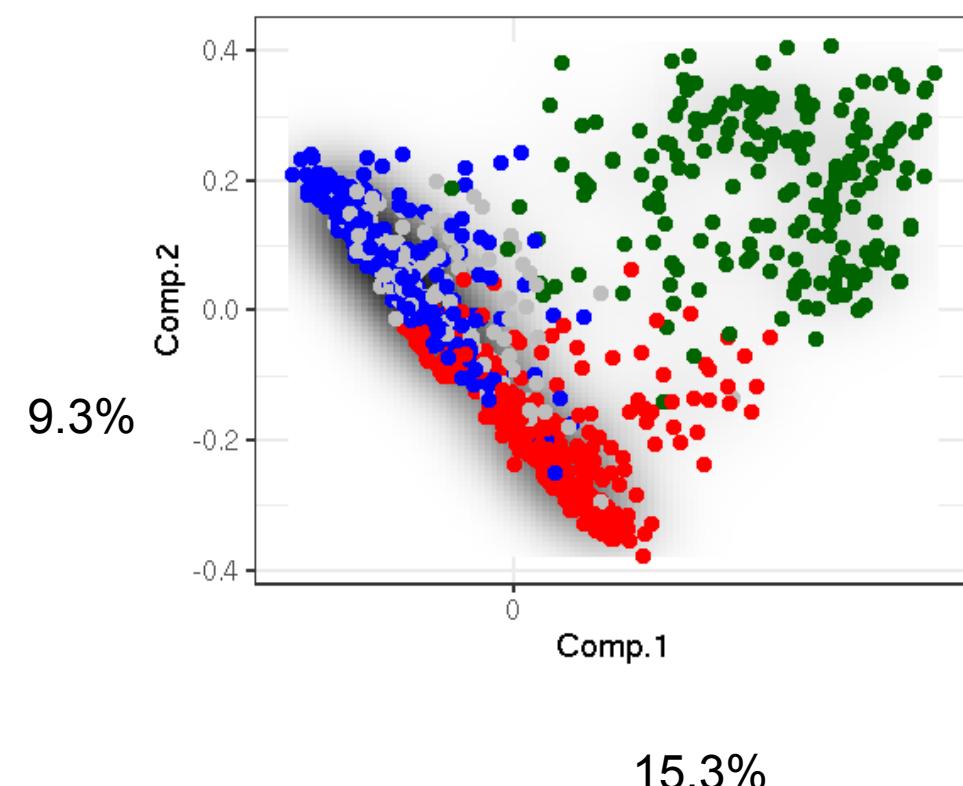
Method: **Preserves distances**

PCA Principal Component Analysis

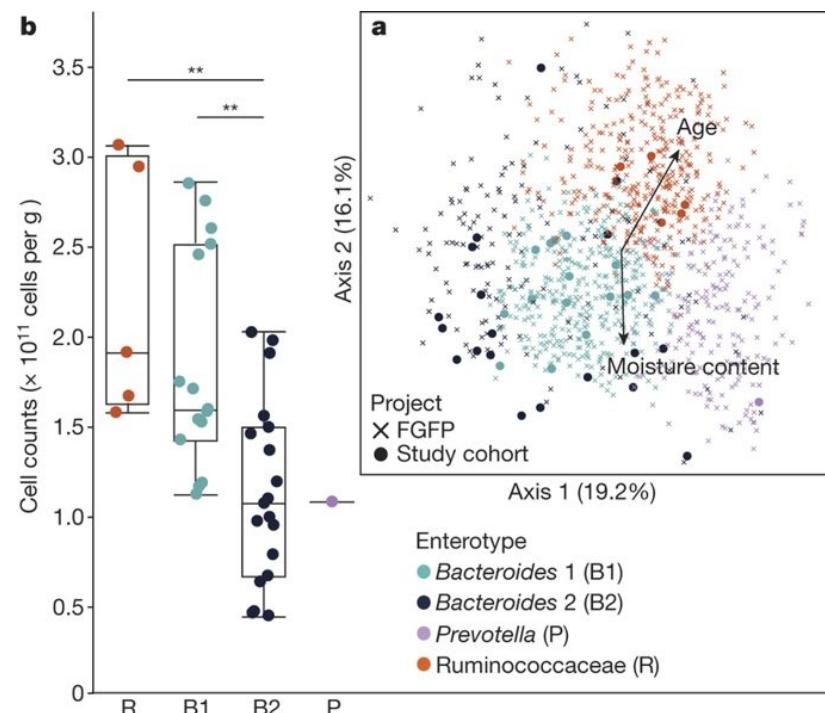
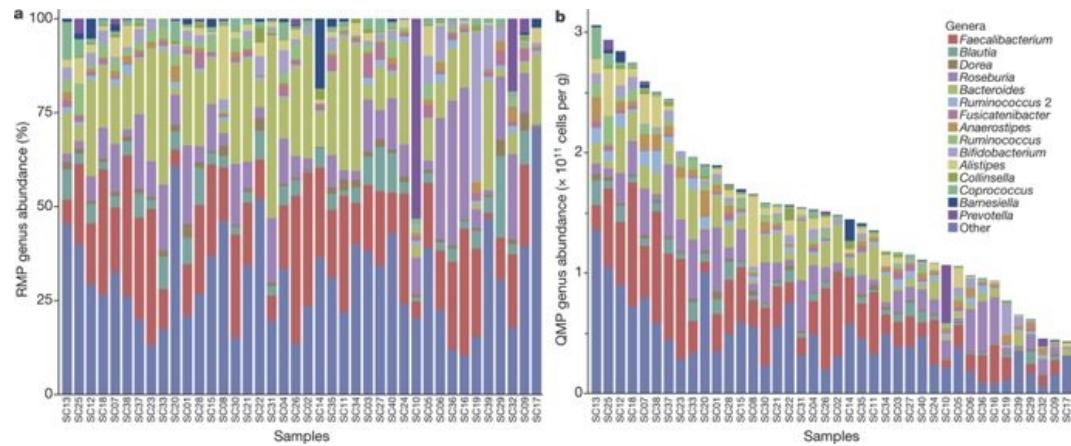
Transformation: CLR

Dissimilarity: Euclidean

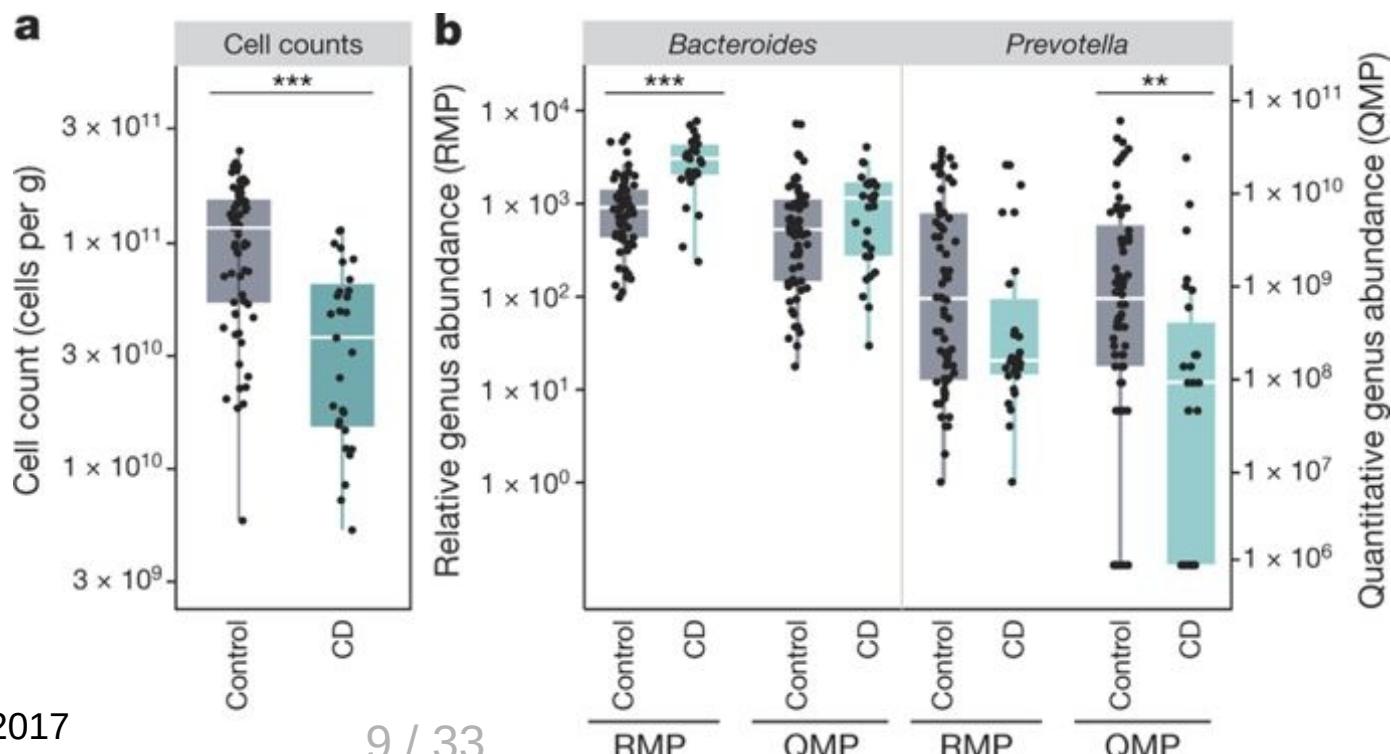
Method: **Captures largest variation**



Relative versus absolute abundance: quantitative microbiome profiling



RMP vs. QMP:
drastic effect
on conclusions!



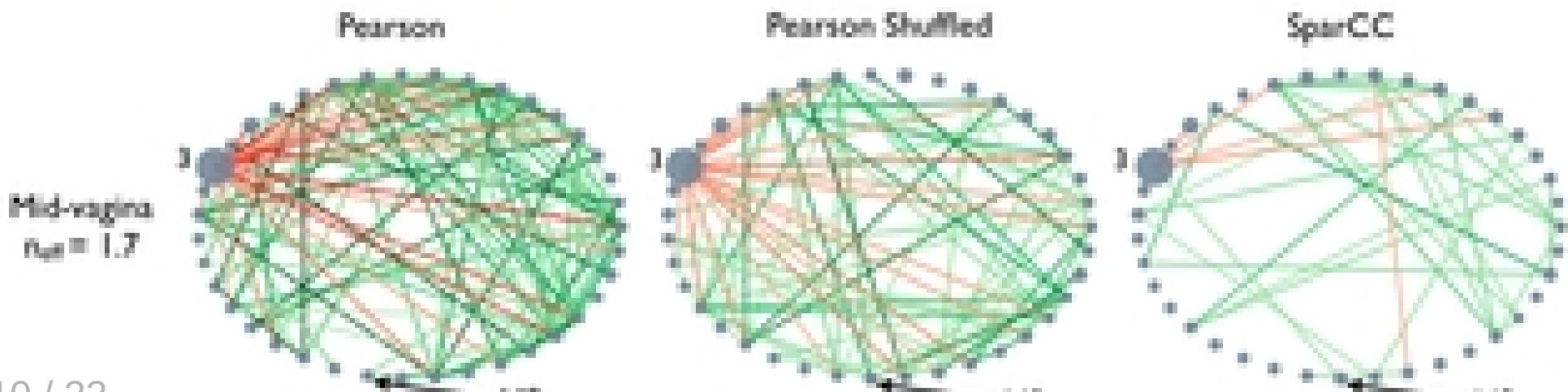
Aitchison transformations remove compositionality

Centered log-ratio transformation (CLR). Ratio stays constant when transforming from QMP to RMP and can therefore be used for sample comparisons that are free from compositionality bias.

$g(x)$: geometric mean

$$\text{clr}(x) = \left[\log \frac{x_1}{g(x)} \dots \log \frac{x_{D-1}}{g(x)} \right]$$

Drastic effects on co-occurrence network inference



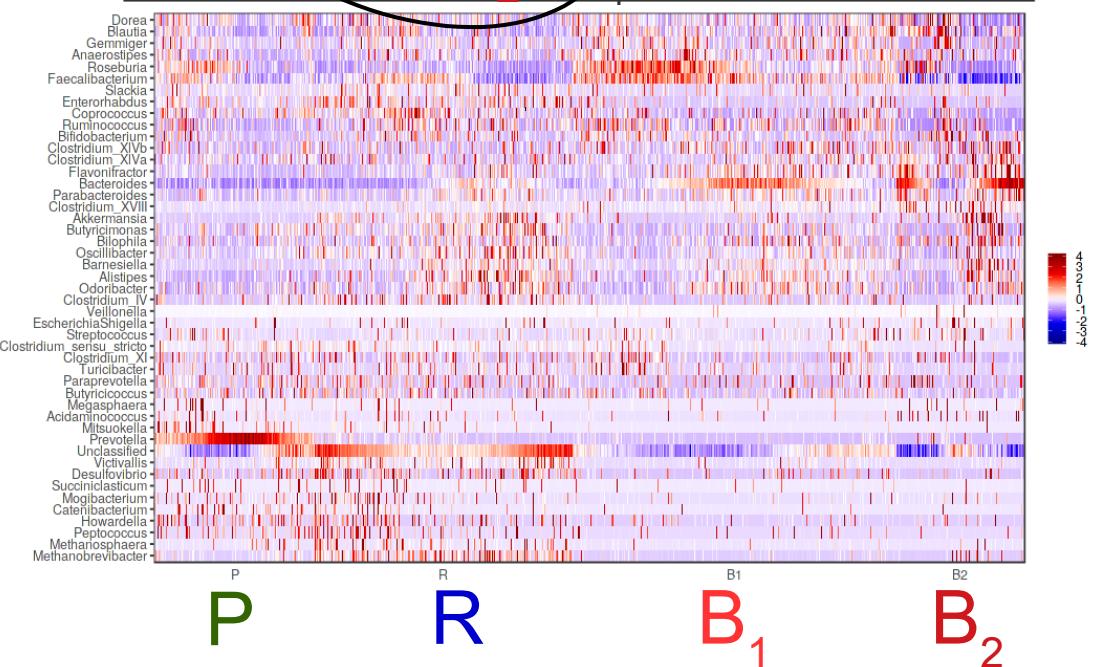
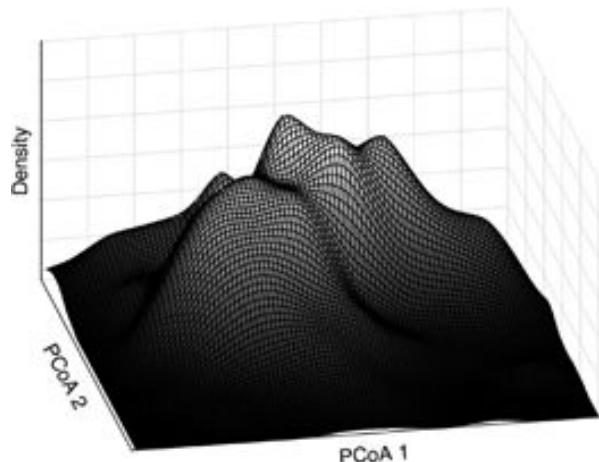
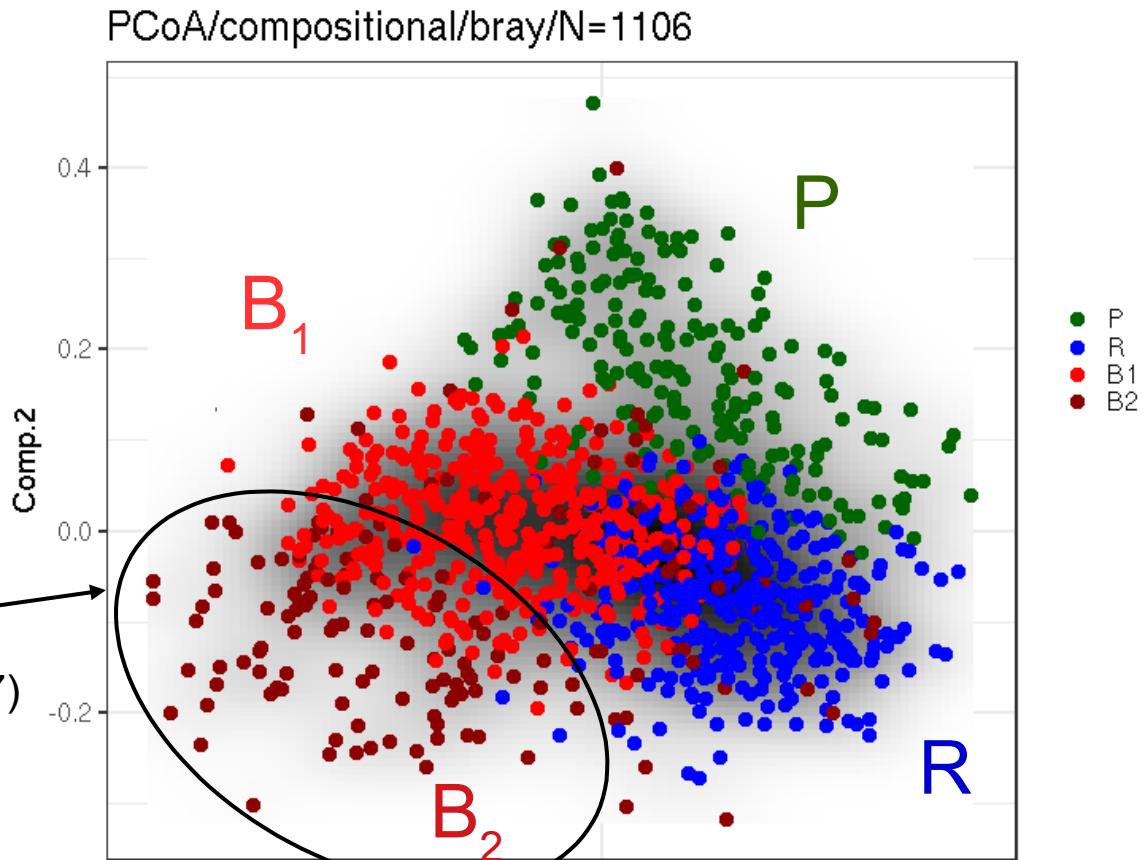
Broad community types (Flemish Gut Flora)

Bacteroides has two subtypes

B₂ subtype associated with

IBD & low cell count

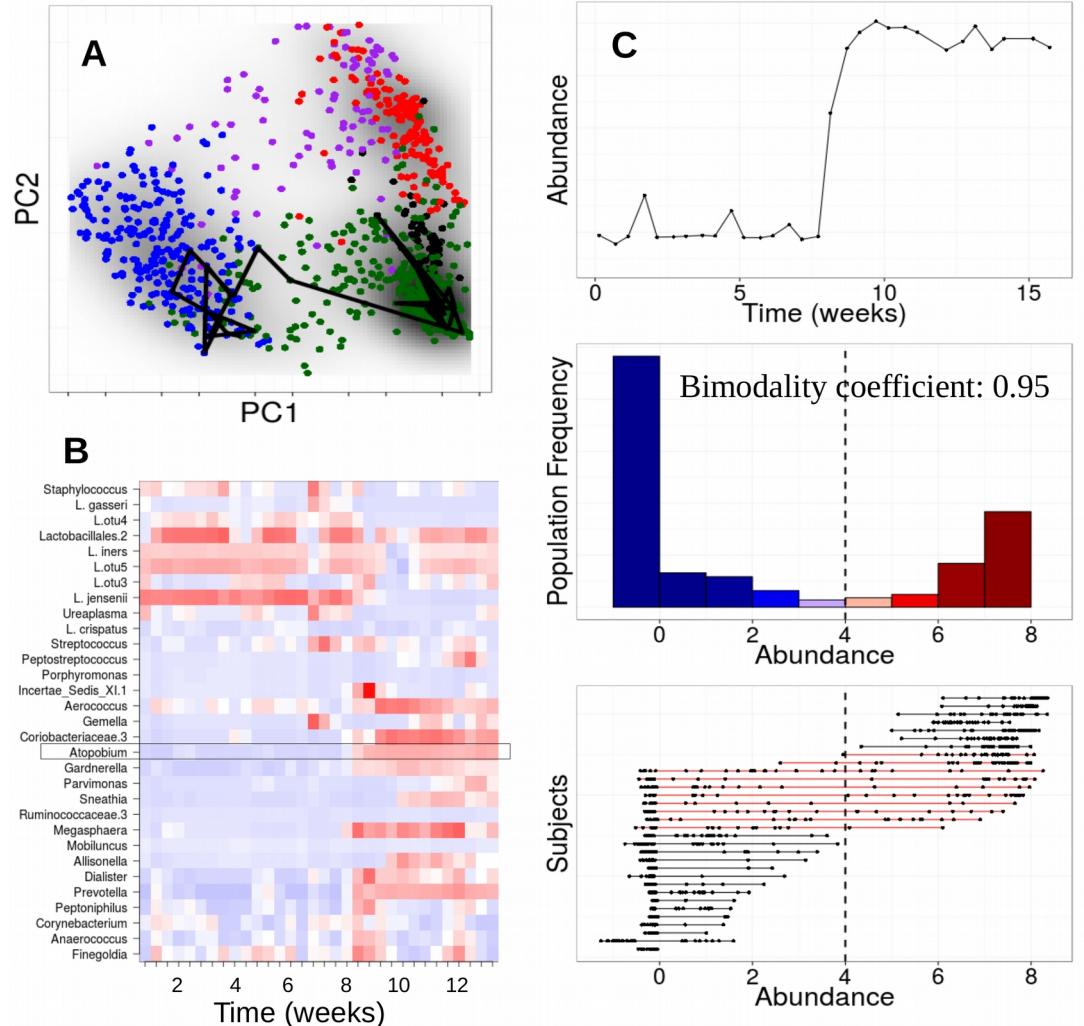
(Raes Lab / Vandeputte et al. Nature 2017)



Metagenomics meets time series analysis: unraveling microbial community dynamics

Karoline Faust^{1,2,3,9}, Leo Lahti^{4,5,9}, Didier Gonze^{6,7},
Willem M de Vos^{4,5,8} and Jeroen Raes^{1,2,3}

Variation:
- cross-sectional
- spatial
- temporal



Data: Vaginal microbiota time series from 32 women. Gajer et al. 2012

The maternal vaginal microbiome partially mediates the effects of prenatal stress on offspring gut and hypothalamus

Eldin Jašarević^{1,2,3,4}, Christopher D. Howard^{1,2}, Kathleen Morrison^{1,2,3,4}, Ana Misic^{1,5},
Tiffany Weinkopff³, Phillip Scott^{1,3}, Christopher Hunter^{1,3}, Daniel Beiting^{1,3} and Tracy L. Bale^{1,2,3,4*}



OPEN

Stress during pregnancy alters temporal and spatial dynamics of the maternal and offspring microbiome in a sex-specific manner

Received: 12 October 2016

Accepted: 06 February 2017

Published: 07 March 2017

Eldin Jašarević^{1,2}, Christopher D. Howard^{1,2}, Ana M. Misic^{1,3}, Daniel P. Beiting^{1,3} & Tracy L. Bale^{1,2}

Psychoneuroendocrinology (2015) 60, 58–74



Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/psyneuen

Prenatal stress-induced alterations in major physiological systems correlate with gut microbiota composition in adulthood

Anna V. Golubeva^a, Sean Crampton^b, Lieve Desbonnet^{a,1},
Deirdre Edge^c, Orla O'Sullivan^d, Kevin W. Lomasney^{a,e,2},
Alexander V. Zhdanov^f, Fiona Crispie^{a,d}, Rachel D. Moloney^{a,3},
Yuliya E. Borre^a, Paul D. Cotter^{a,d}, Niall P. Hyland^{a,e},
Ken D. O'Halloran^c, Timothy G. Dinan^{a,g},
Gerard W. O'Keeffe^{a,b,h,**}, John F. Cryan^{a,b,*}



Early life & Gut-Brain axis: Finnbrain birth cohort Maternal Prenatal Distress, Infant Gut Microbiota, and Cognitive Development

Aatsinki Anna-Katariina
Anniina Keskitalo, Ville Laitinen, Anastasia Karavaeva,
Eveliina Munukka, Paula Mustonen, Henna-Maria
Jusitupa, Leo Lahti, Erkki Eerola, Hasse Karlsson, Linnea
Karlsson

FinnBrain Birth Cohort Study,
University of Turku, Finland

13 / 33

Funded by: Academy of Finland, Gyllenberg Foundation, Yrjö Jansson Foundation, Finnish State Grants for Clinical Research

Microbial communities as dynamical systems

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

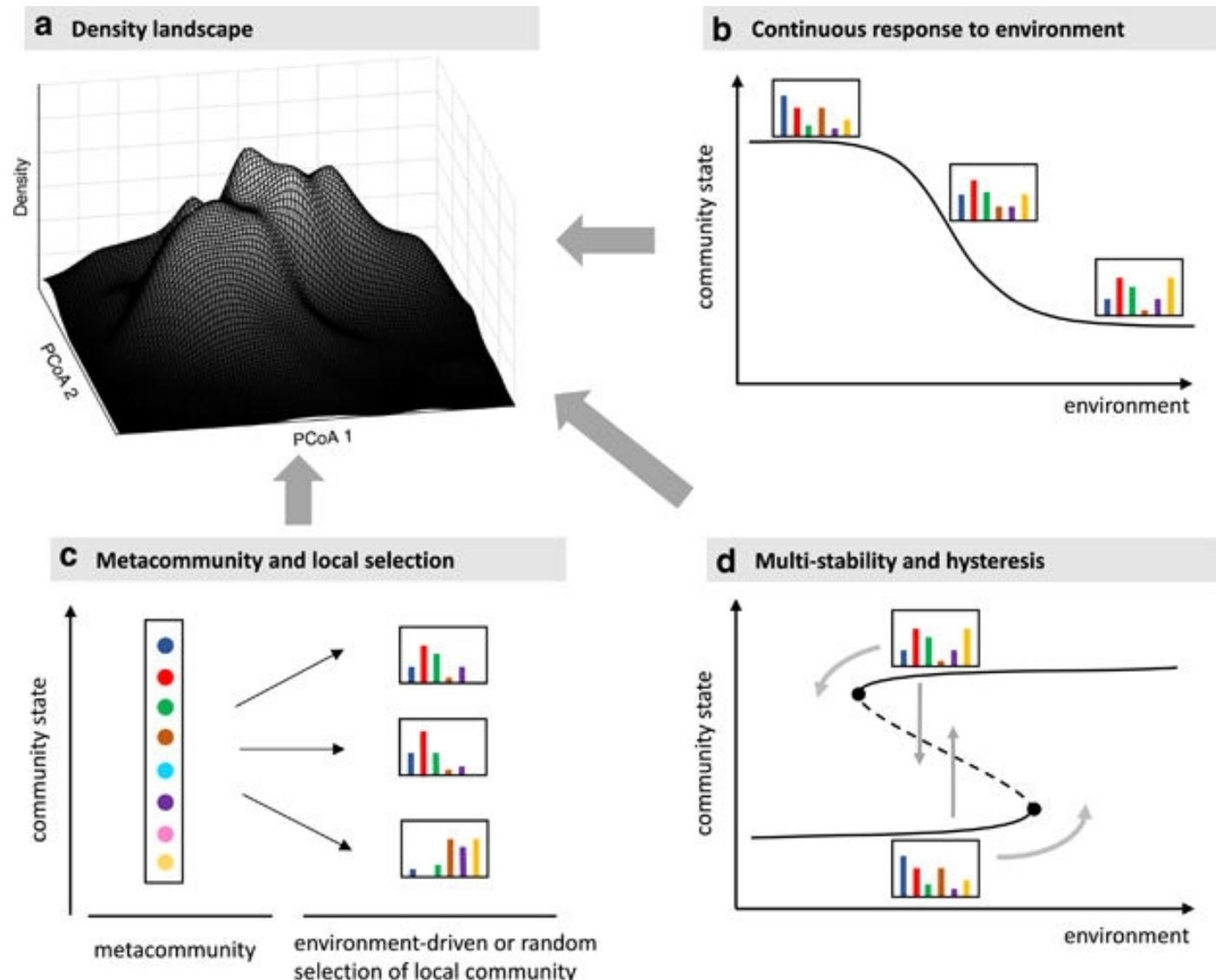
External perturbations
(push & pulse)

Internal dynamics and
multi-stability

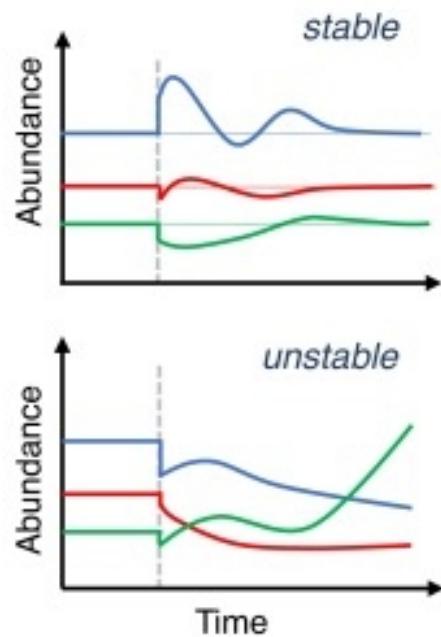
Immigration

Stochasticity

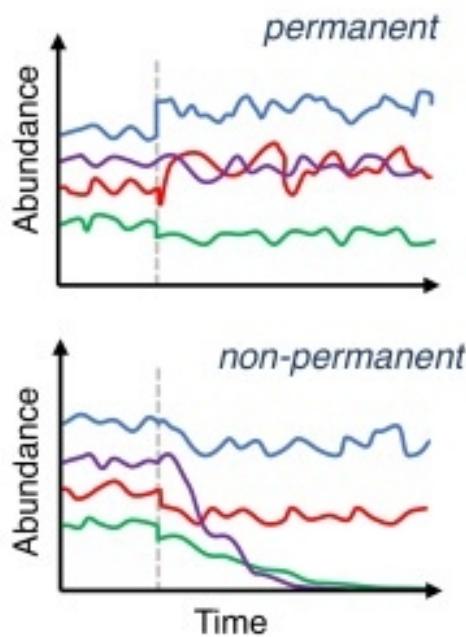
Memory



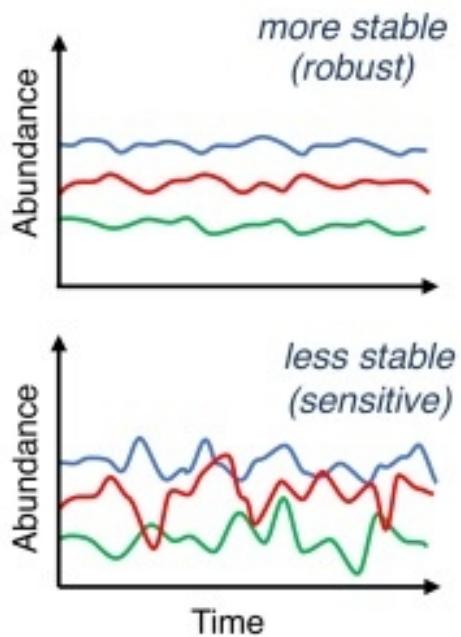
**(a) Steady state stability
(nonlinear theory)**



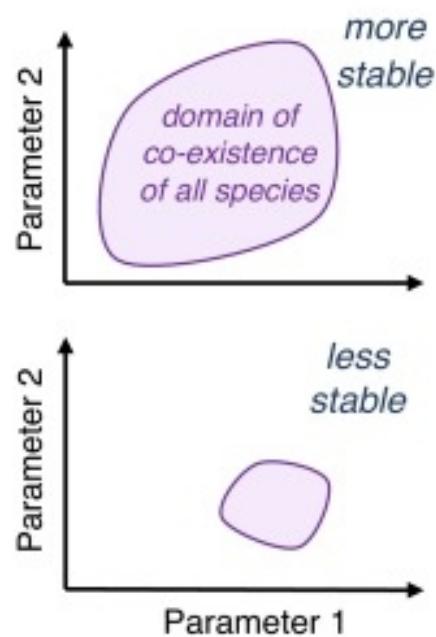
**Permanence
(long-term survival)**



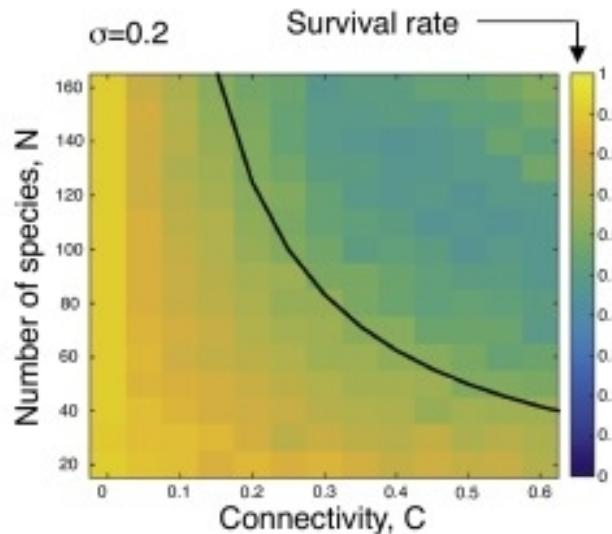
**Temporal stability
(robustness/sensitivity)**



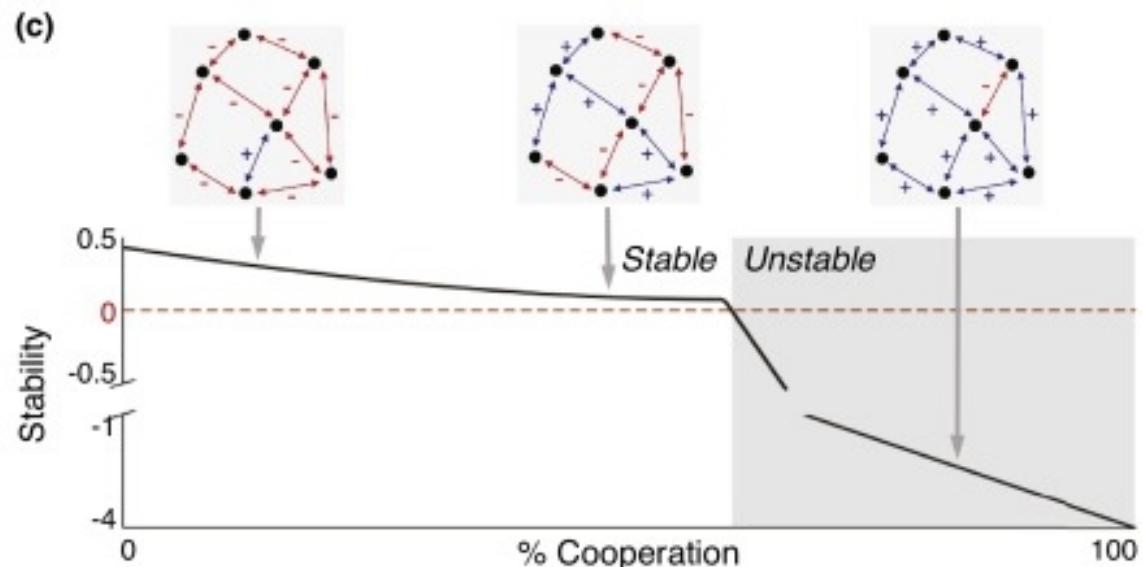
**Structural stability
(parameter space analysis)**



(b)



(c)



Interaction models yield more complex dynamics and predict alternative stable states

Multi-stability and the origin of microbial community types

Didier Gonze, Leo Lahti, Jeroen Raes & Karoline Faust 

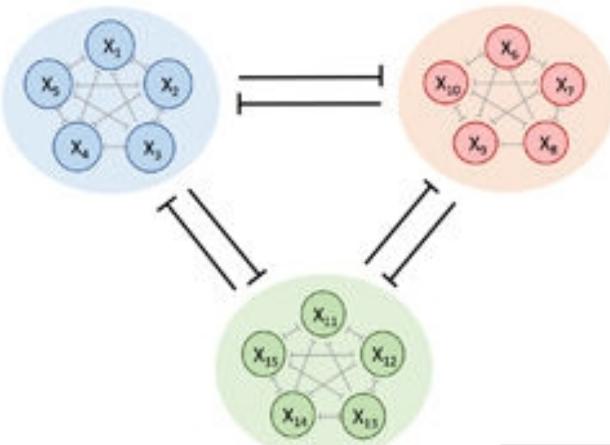
The ISME Journal (2017) **11**, 2159–2166 (2017)

doi:10.1038/ismej.2017.60

Received: 06 December 2016

Revised: 28 February 2017

Inter-group inhibition stronger than intra-group inhibition (the smaller the stronger)

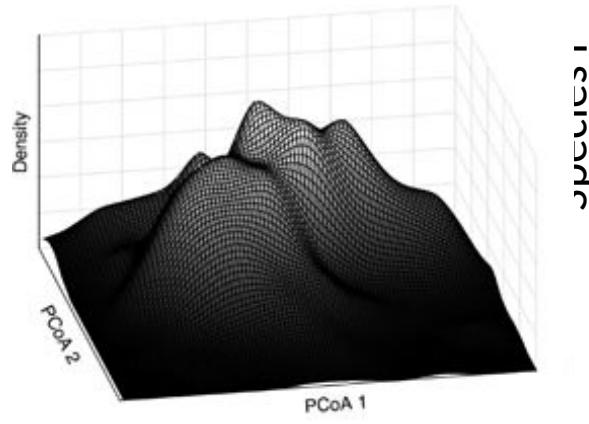


b

$$\frac{dX_i}{dt} = X_i(b_i f_i(\{X_k\}) - k_i X_i)$$

$$f_i(\{X_k\}) = \prod_{\substack{k=1 \\ k \neq i}}^N \frac{K_{ik}^n}{K_{ik}^n + X_k^n}$$

K_{ij}	X_1	X_2	X_3	b_i	k_i
X_1	-	0.1	0.1	1	1
X_2	0.1	-	0.1	0.95	1
X_3	0.1	0.1	-	1.05	1



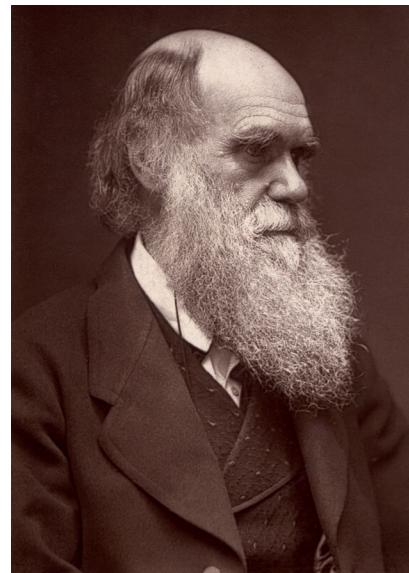
	2	4	6	8	10	12	14								
-	0.88	1	0.86	0.99	0.52	0.51	0.41	0.39	0.48	0.6	0.5	0.57	0.52	0.43	
2	1.17	-	0.79	1.06	1.11	0.51	0.5	0.53	0.5	0.58	0.43	0.53	0.49	0.42	0.39
4	1.2	1.04	-	0.97	0.87	0.52	0.5	0.57	0.47	0.5	0.53	0.36	0.56	0.44	0.55
6	1.06	1.28	1.19	-	0.93	0.46	0.47	0.55	0.57	0.57	0.49	0.62	0.43	0.48	0.53
8	1.06	1.13	1.1	0.94	-	0.5	0.55	0.53	0.57	0.43	0.52	0.42	0.39	0.56	0.53
10	0.47	0.46	0.43	0.48	0.53	-	0.98	1.09	0.94	1.09	0.41	0.48	0.45	0.47	0.51
12	0.54	0.47	0.53	0.49	0.56	0.95	-	0.92	1.02	1	0.52	0.49	0.53	0.39	0.5
14	0.45	0.52	0.53	0.52	0.49	0.84	1.03	-	1.02	1.01	0.43	0.51	0.44	0.46	0.57
16	0.52	0.5	0.45	0.52	0.44	1.07	0.86	1.06	-	0.99	0.54	0.55	0.51	0.5	0.48
18	0.49	0.51	0.41	0.49	0.5	0.88	0.91	1	0.98	-	0.46	0.54	0.53	0.51	0.42
20	0.45	0.58	0.51	0.52	0.46	0.56	0.47	0.58	0.44	0.57	-	1.05	1.12	1.02	0.99
22	0.53	0.49	0.54	0.51	0.5	0.57	0.55	0.49	0.52	0.45	0.97	-	0.91	1.02	0.92
24	0.53	0.55	0.44	0.59	0.54	0.47	0.45	0.49	0.49	0.51	0.9	0.97	-	0.84	1.08
26	0.47	0.5	0.48	0.52	0.44	0.47	0.58	0.5	0.63	0.38	0.99	0.99	1.15	-	1.09
28	0.56	0.52	0.48	0.56	0.55	0.51	0.54	0.47	0.49	0.41	0.93	1.18	1.11	0.92	-

Species j

Neutral model of biodiversity suggests no interaction: could random chance explain microbiome variation?

Niche model

"When we look at the plants and bushes clothing an entangled bank, we are tempted to attribute their proportional numbers and kinds to what we call chance. But how false a view is this!" (**Darwin**, *The Origin of Species*)

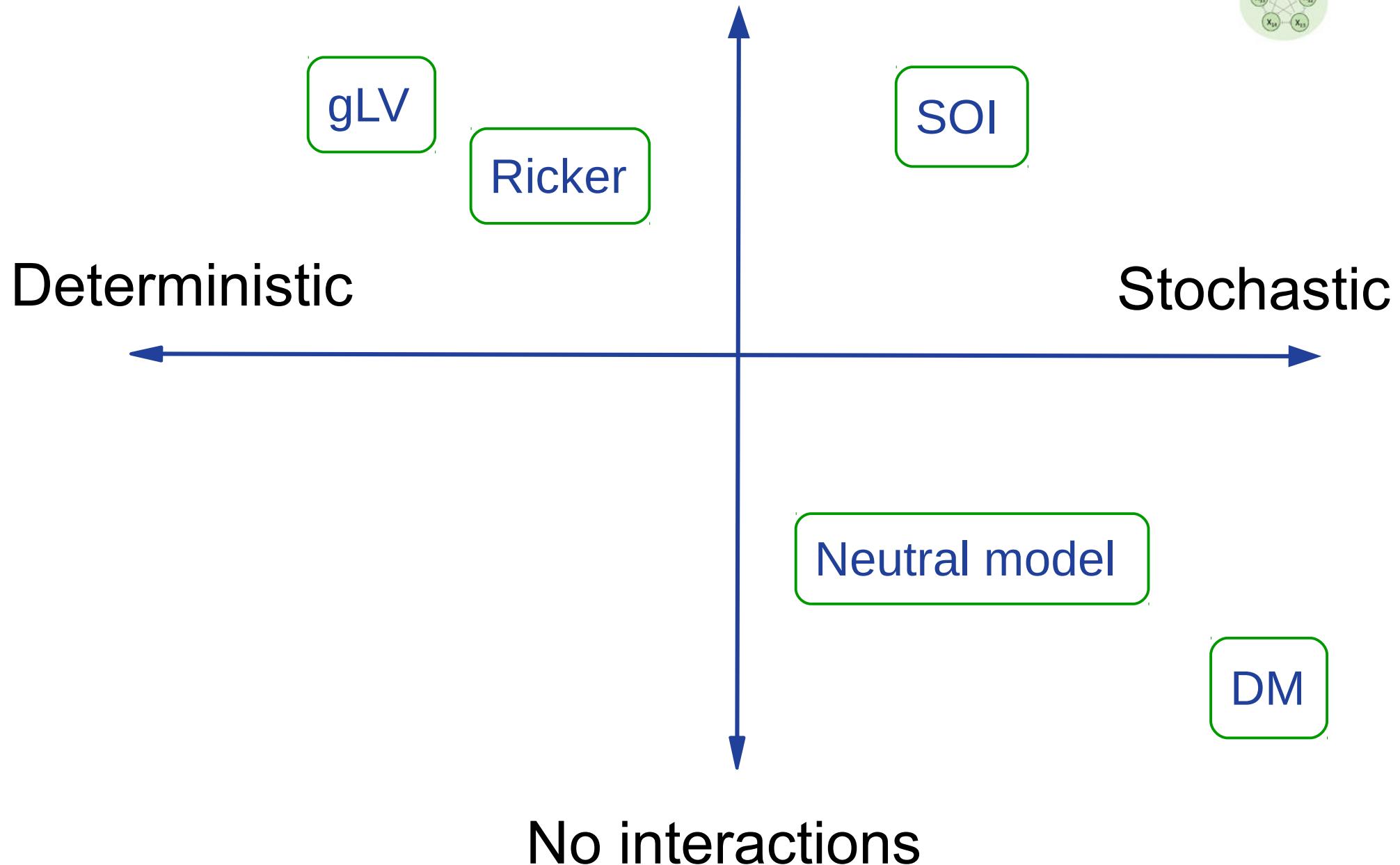


Neutral model

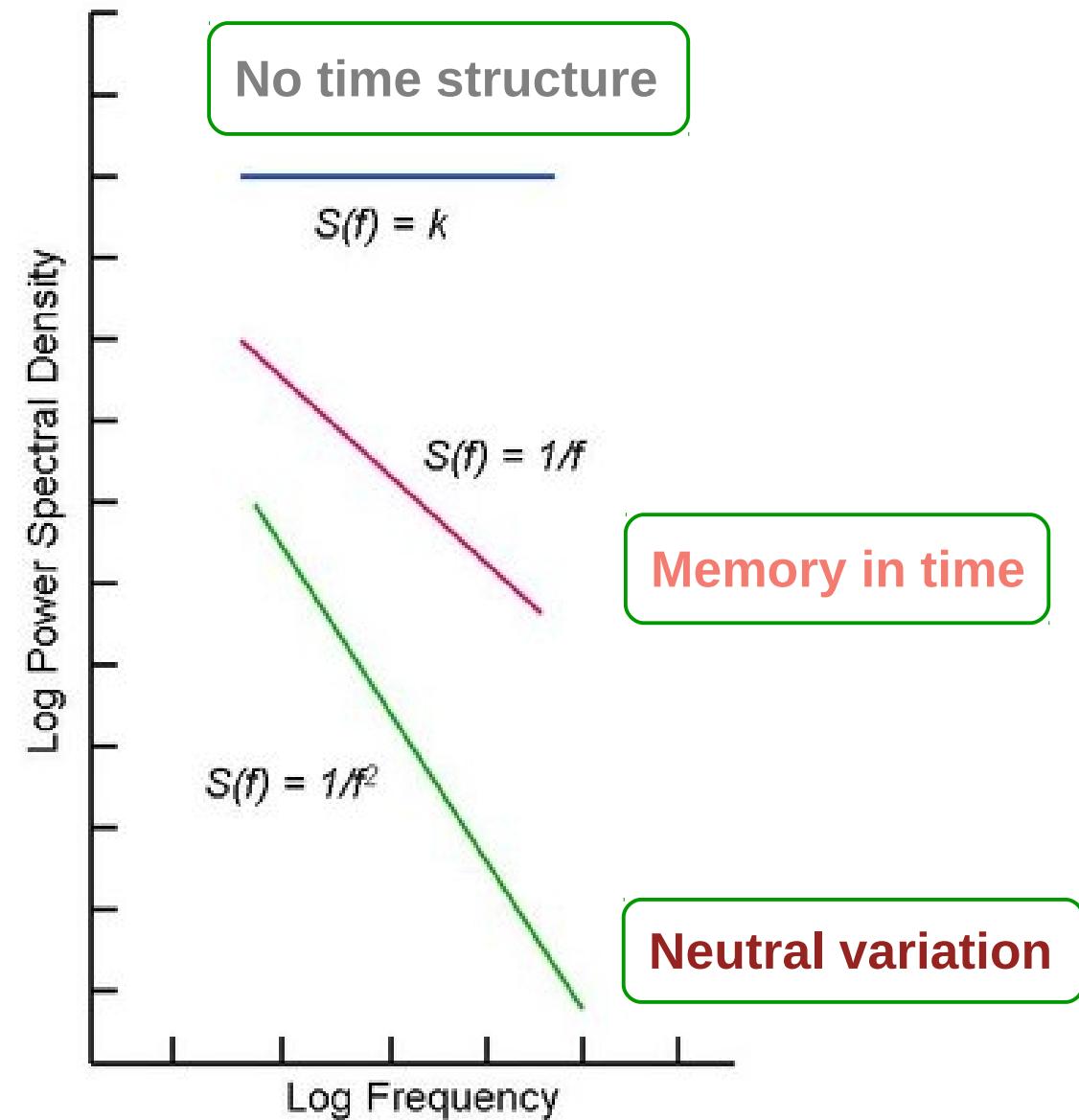
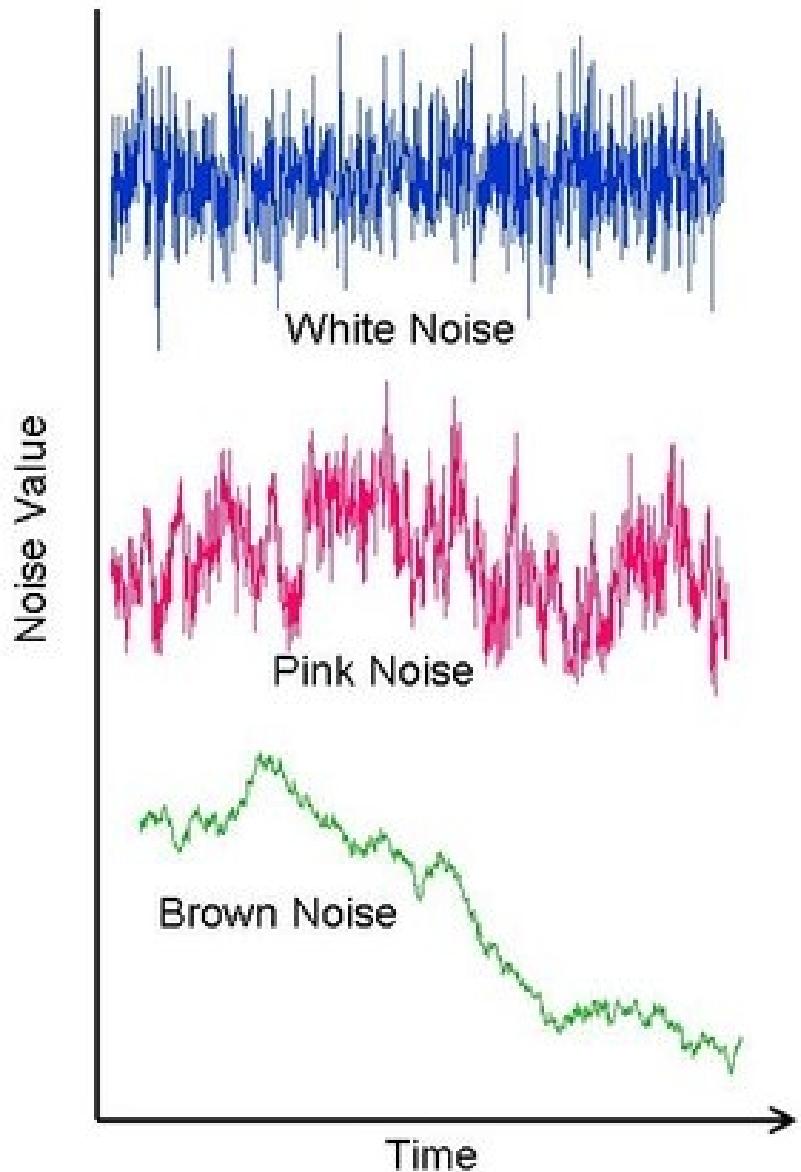
After >25 years on the Barro Colorado Island tropical forests, **Hubbell** proposed that.. random variation could in fact best explain observed biodiversity (Hubbell 2001).

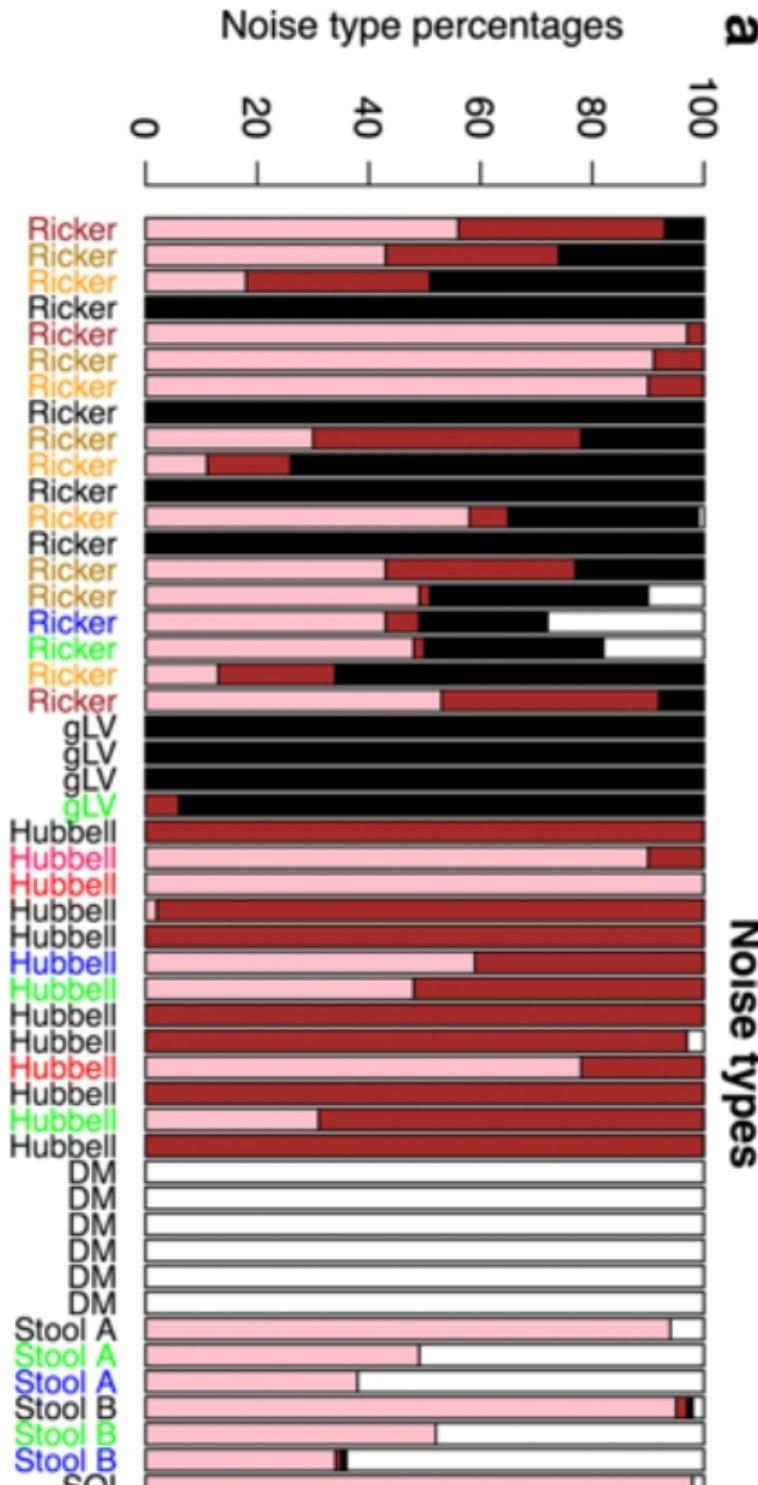


Species Interactions



“Colors of noise” characterize time series, providing an alternative testing scheme





Distinct noise profiles of ecological processes confirmed in simulations

White noise → Unstructured

Pink noise → Time dependence

Brown noise → Neutrality

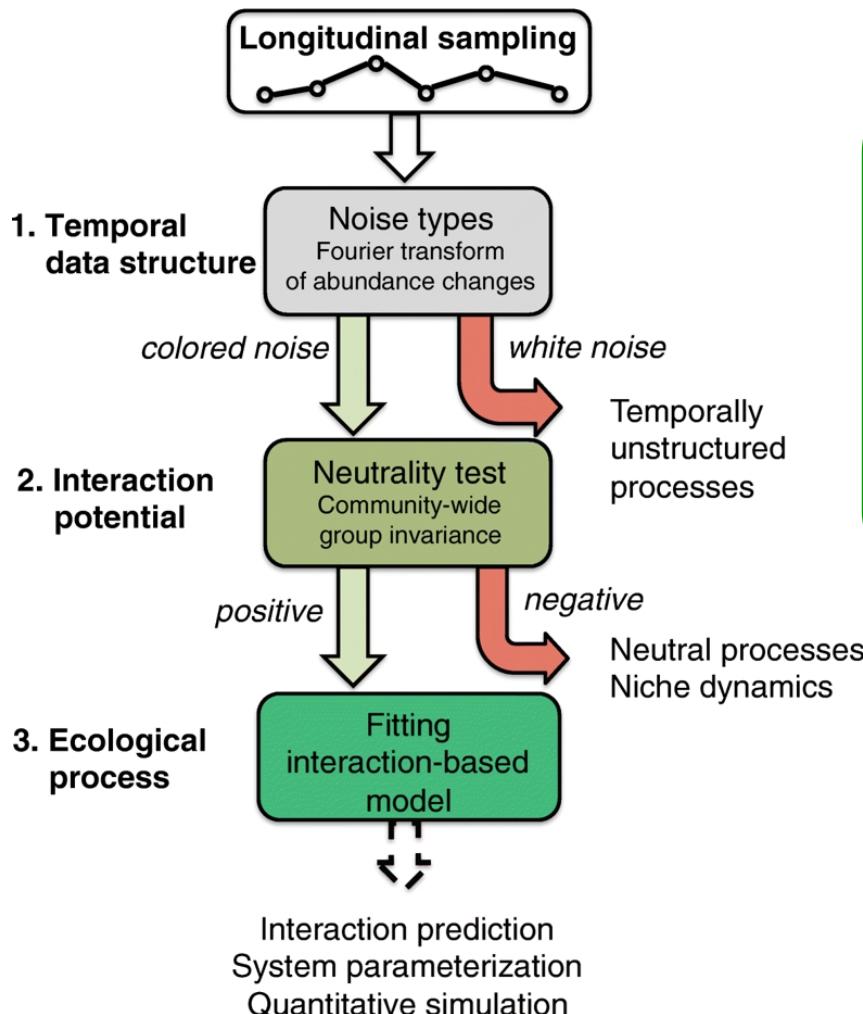
Black noise → Interactions

“Colors of noise” hint at ecological mechanisms

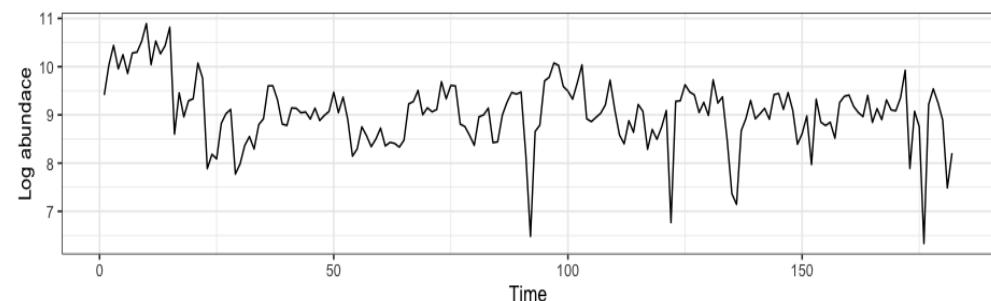


Signatures of ecological processes in microbial community time series

Karoline Faust^{1*} , Franziska Bauchinger², Béatrice Laroche³, Sophie de Buyl^{4,5}, Leo Lahti^{1,6,7}, Alex D. Washburne^{8,9}, Didier Gonze^{5,10} and Stefanie Widder^{11,12,13*}

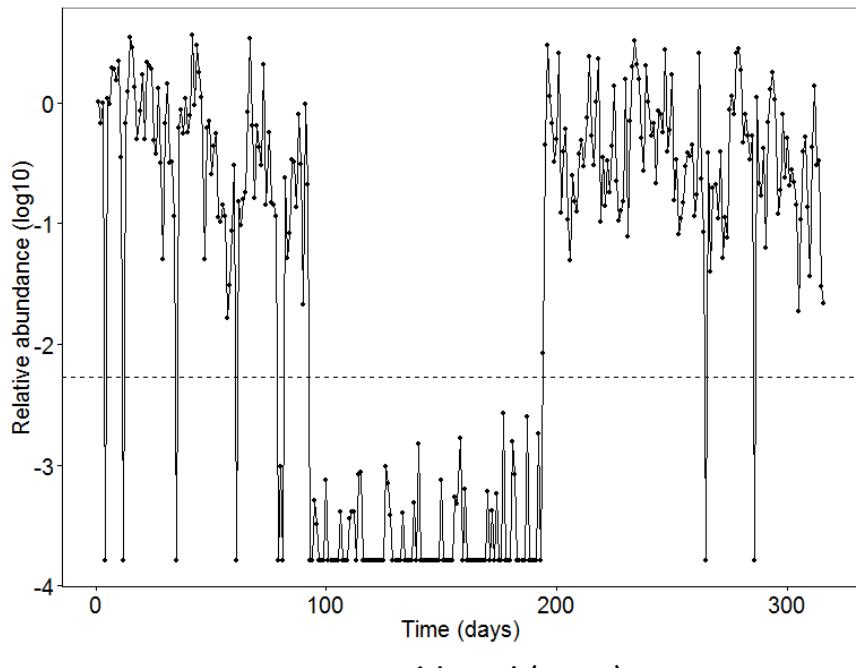


White noise → Unstructured
Pink noise → Time dependence
Brown noise → Neutrality
Black noise → Interactions

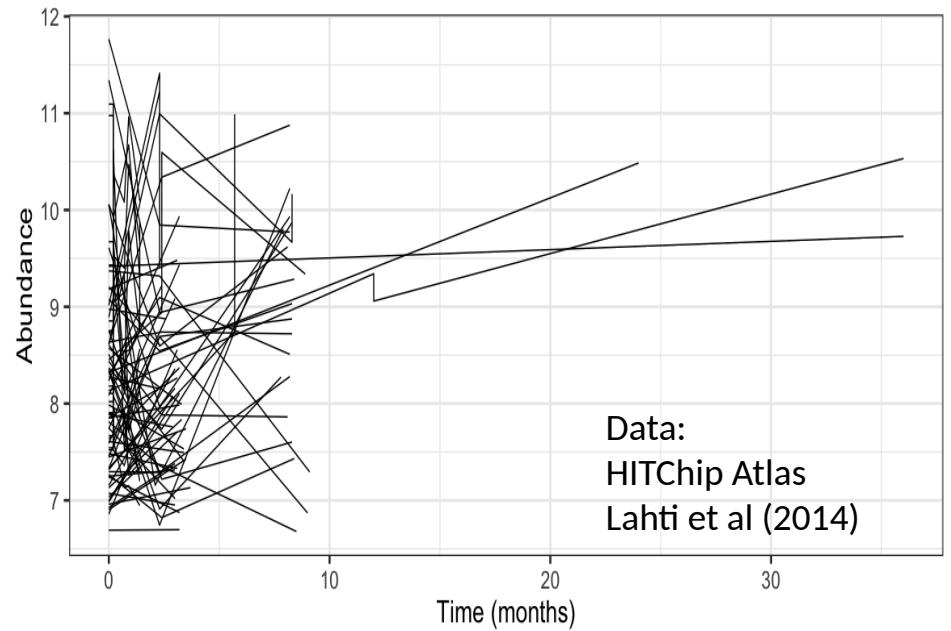


Typical microbiome time series: short, sparse, noisy
→ challenge for fitting parametric models

1 individual & 300+ days



78 individuals;
2-5 time points;
1-36 months



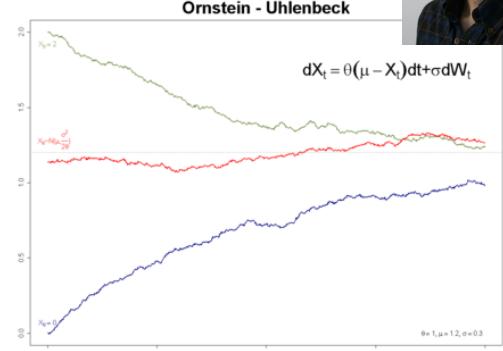
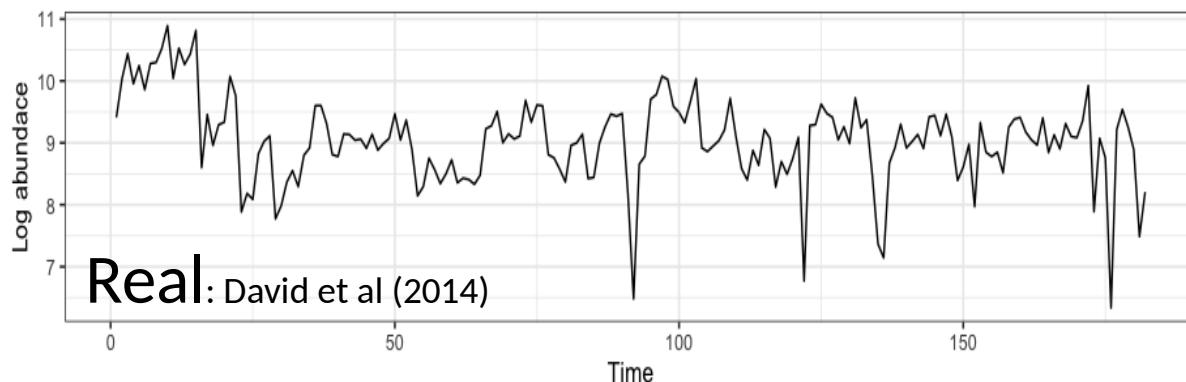
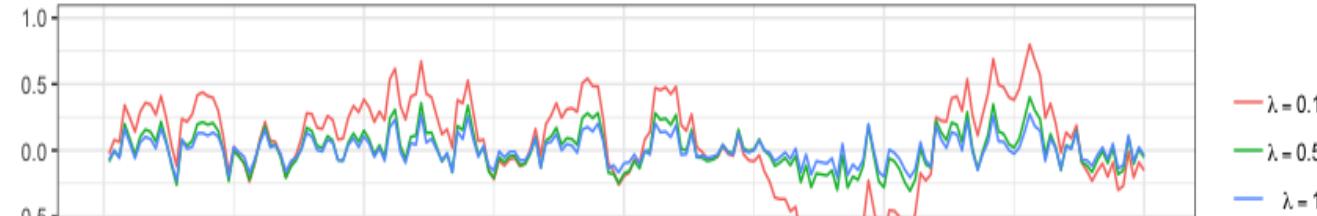
Raes Lab / Flemish Gut Flora project now collecting
dense time series for hundreds of individuals 150+ days!

Quantifying resilience & stochasticity

Ornstein-Uhlenbeck Process (OUP)
quantifies key properties of a stable state:

$$dX = \lambda(\mu - X)dt + \sigma dS$$

$\mu = 0; \sigma = 0.2$

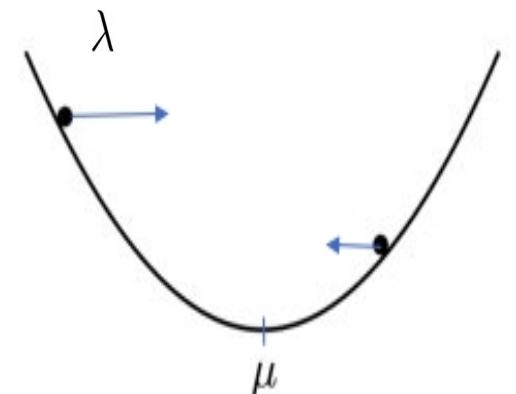


λ = mean reversion rate

μ = long-term mean

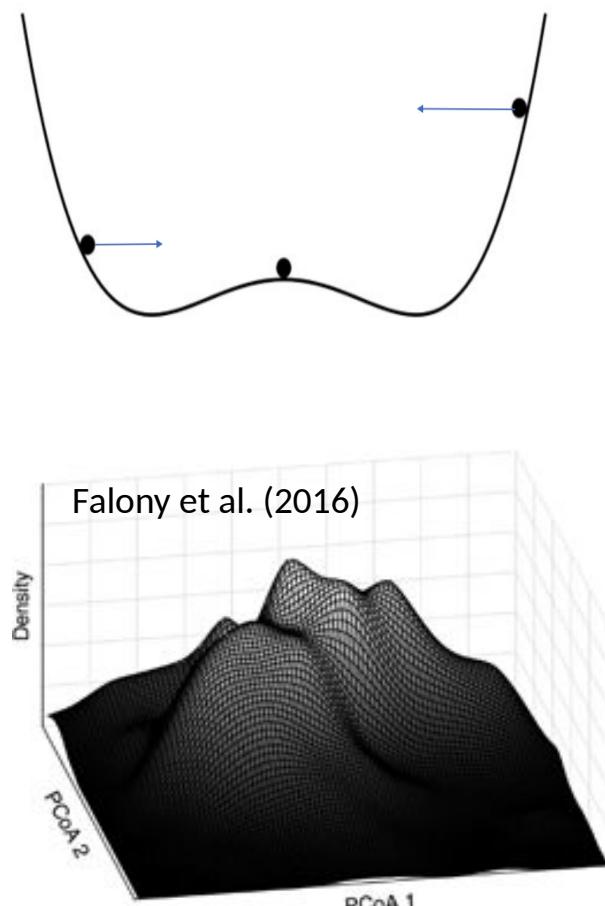
σ = stochasticity level

S = stochastic process

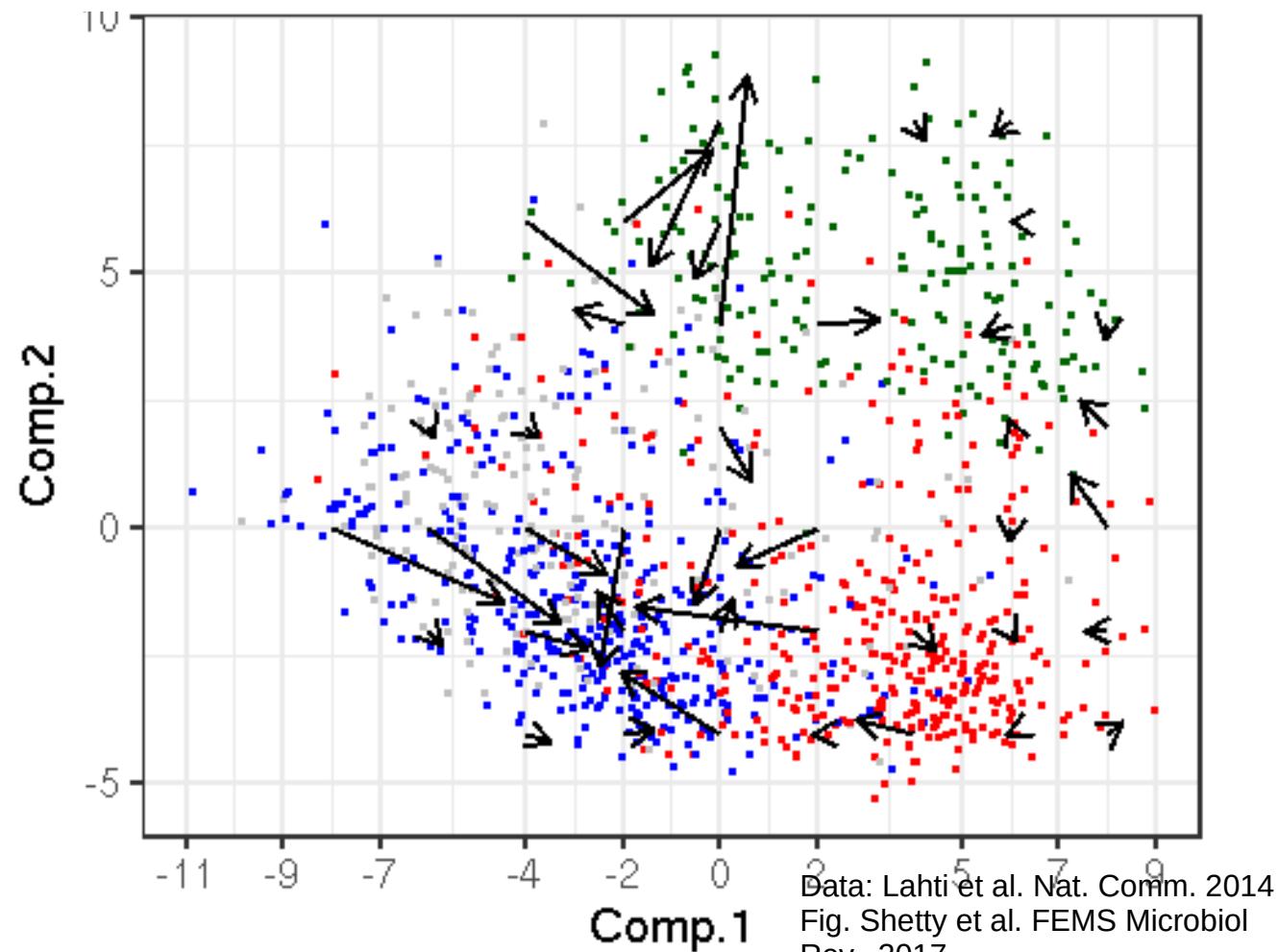


Towards a dynamic landscape model of the gut microbiome

Prior info from background cohorts & pooling evidence across individuals



Gradients of change on HITChip PCA landscape (1006 cross-sectional + 78 longitudinal profiles)



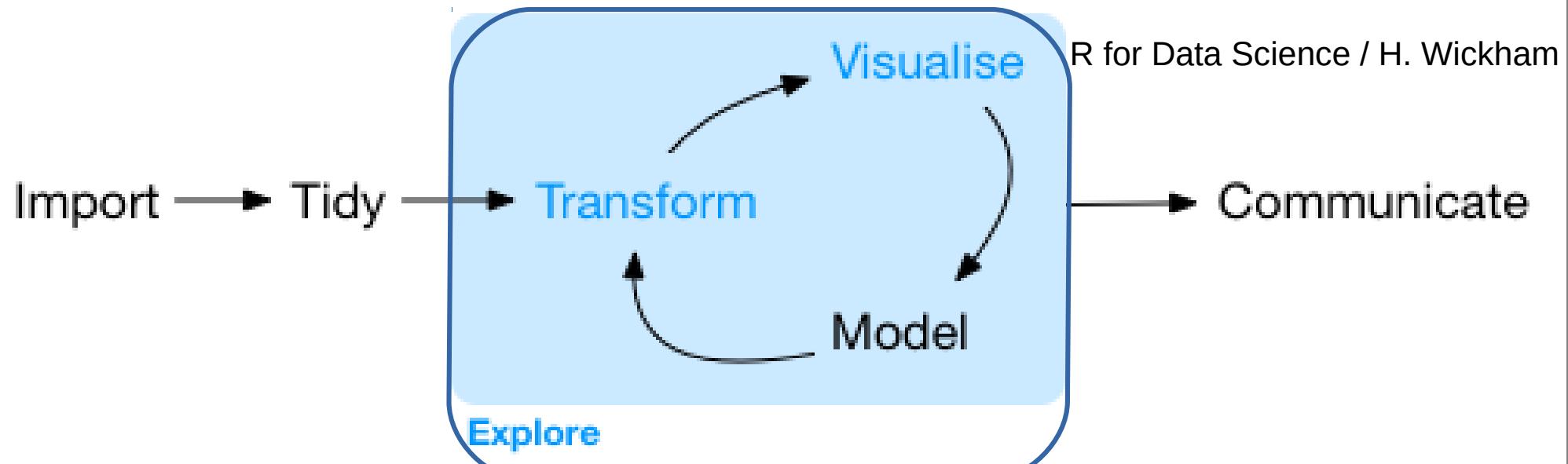
Data: Lahti et al. Nat. Comm. 2014
Fig. Shetty et al. FEMS Microbiol Rev.. 2017

“I have begun to think that no one ought to publish biometric results, without lodging a well arranged and well bound manuscript copy of all his data, in some place where it should be accessible, under reasonable restrictions, to those who desire to verify his work.”

Francis Galton (1901), *Biometrika* 1:1, pp. 7-10.

Microbiome data science

Sudarshan Shetty¹, Leo Lahti^{2*}



Program

Challenges:

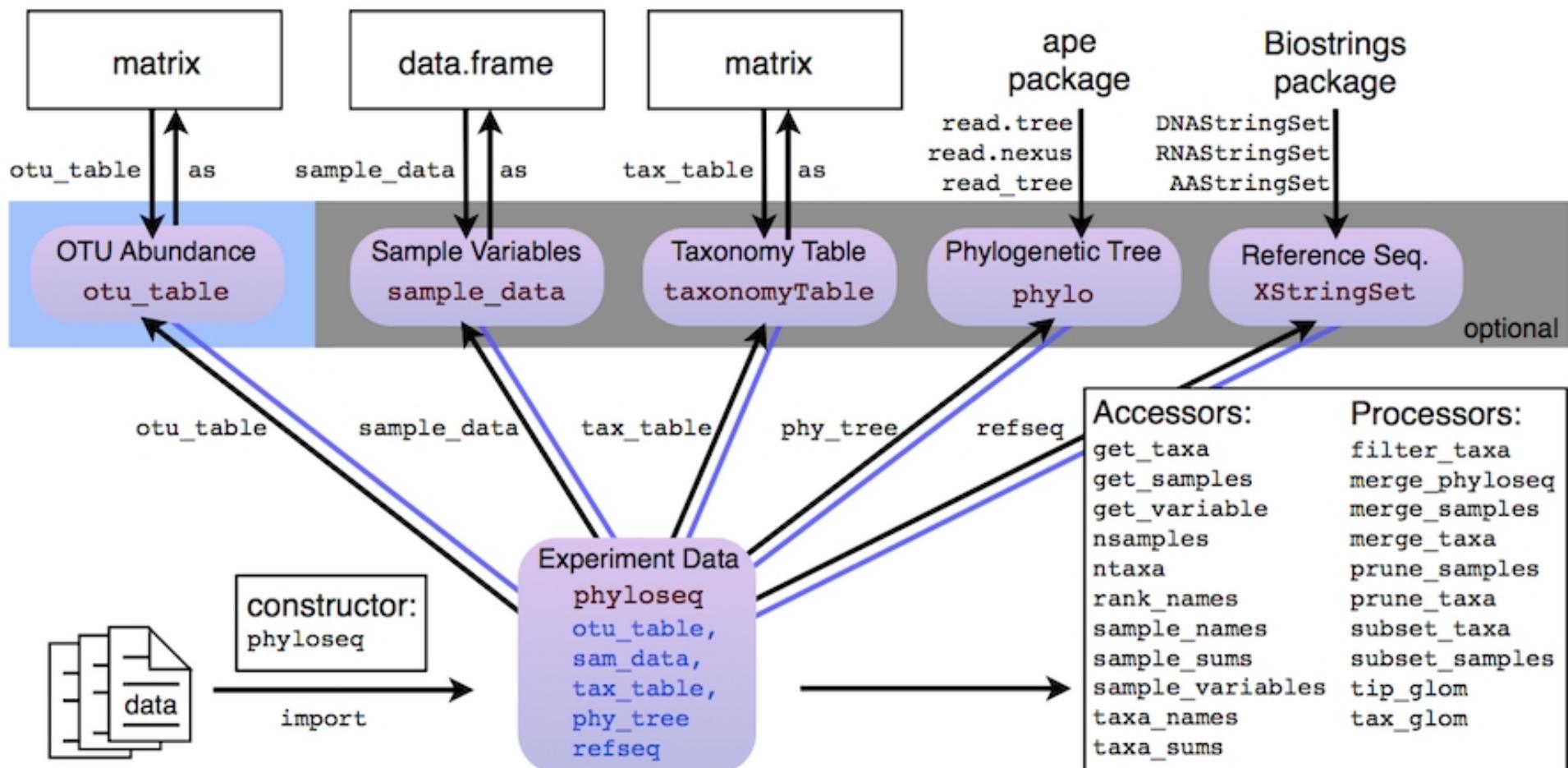
- sparse
- non-gaussian
- count data
- complex
- multi-level

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. qimic 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

Data standard: phyloseq

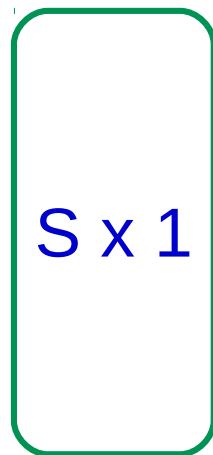
Standard for (16S) microbiome bioinformatics in R
(J McMurdie, S Holmes et al.)



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

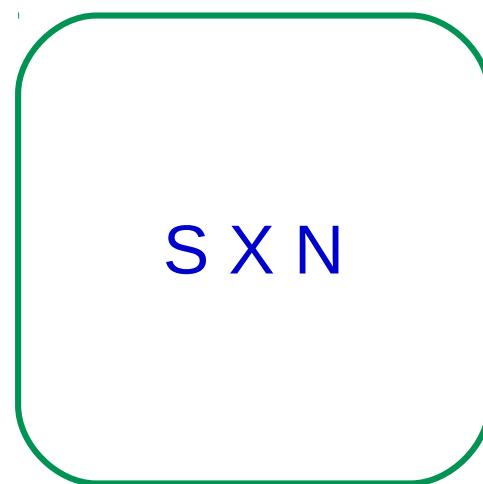
QUESTION

Individual

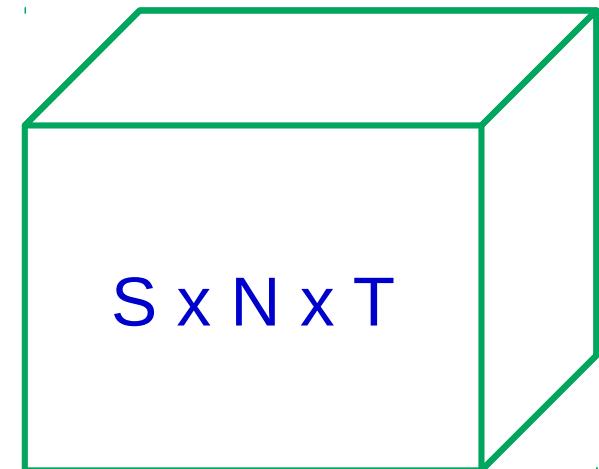


Sequence
Variants /
OTUs

Population

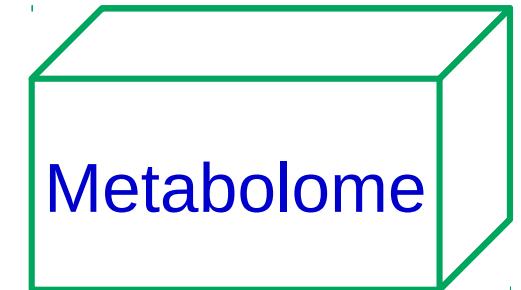
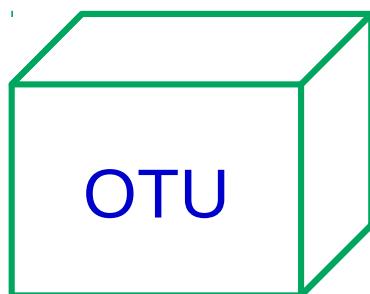


Longitudinal cohort



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

“Multi-modal” longitudinal cohort



Synergies & networking in package development



microbiome R package

chat on [gitter](#)

build [passing](#)

codecov

24%

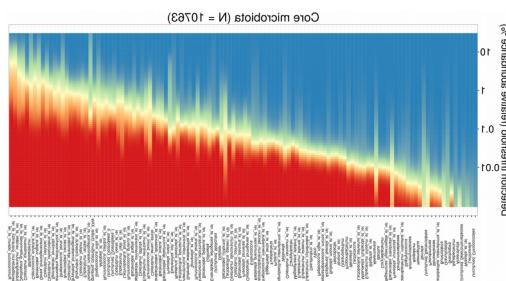
PRs [welcome](#)

Core & prevalence

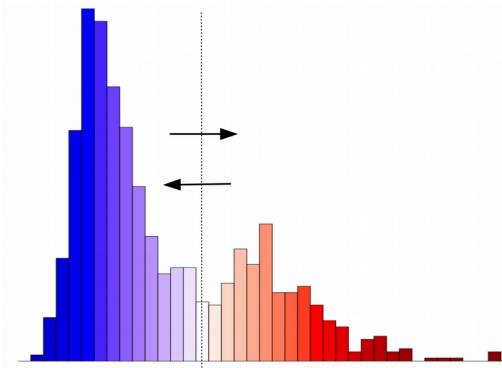
`prevalence(x)`

`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)`

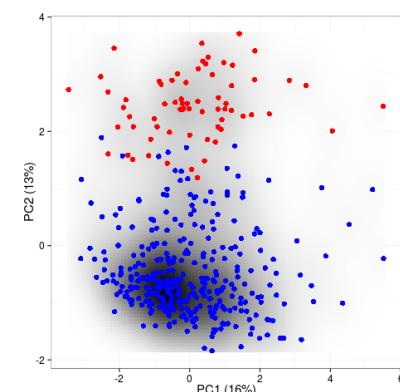
`diversity(x)`

`evenness(x)`

`dominance(x)`

`rarity(x)`

`readcount(x)`

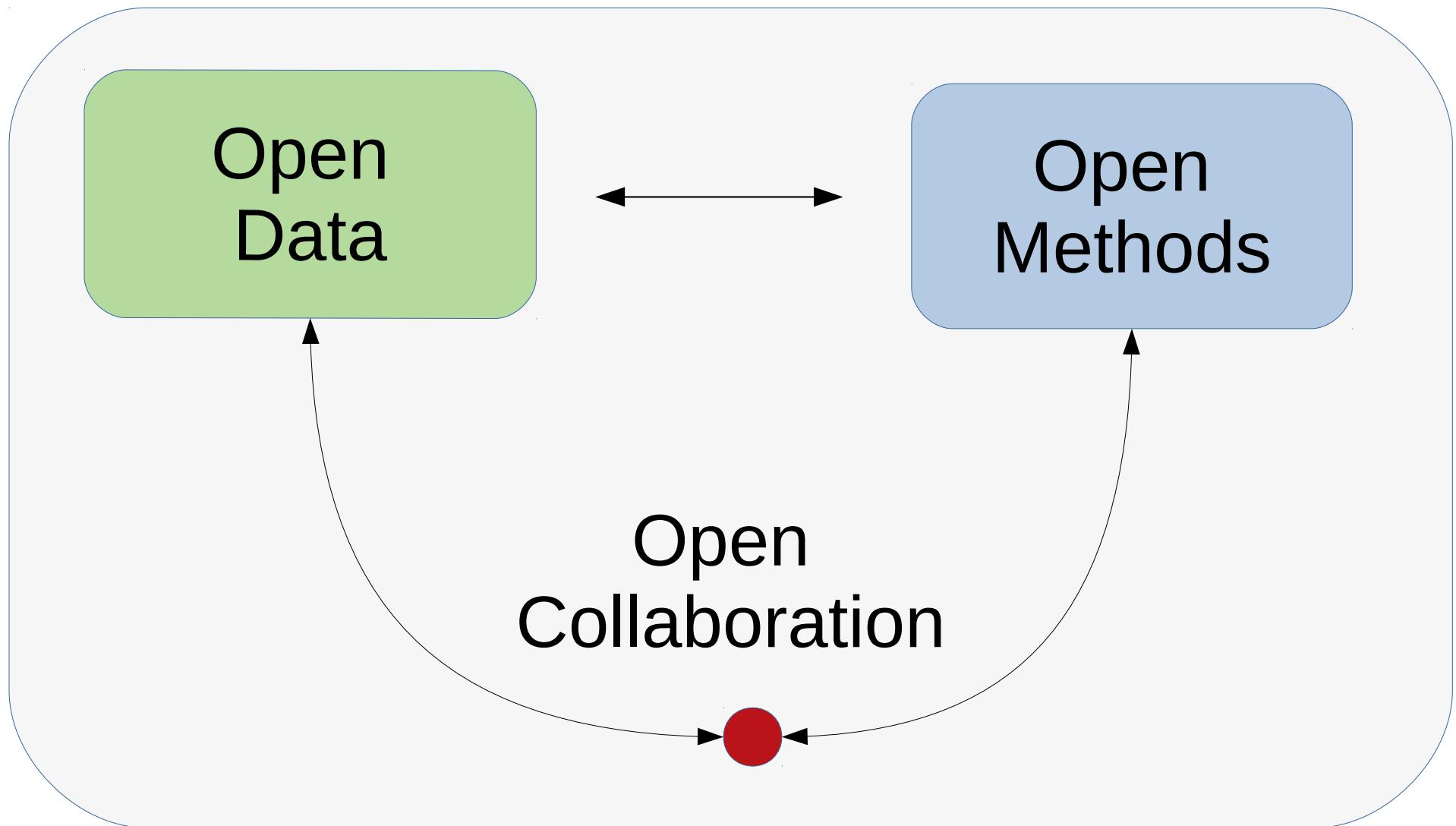


Quality control

- continuous integration
- unit tests

Open data science

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





Ville Laitinen
PhD student
email: ville.laitinen@utu.fi



Aaro Salosensaari
PhD student (co-supervised with Doc Teemu Niiranen).
[Research homepage](#).
email: aaro.salosensaari@utu.fi



Joonatan Palmu
PhD student (co-supervised with Doc Teemu Niiranen)
email: joonatan.palmu@utu.fi



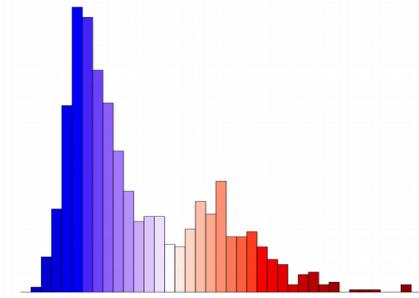
Anna Sorjamaa
PhD student (Oulu; co-supervised with Doc Justus Reunanen)
email: anna.sorjamaa@oulu.fi



Anna Aatsinki
PhD student (Turku; co-supervised with Doc Linnea Karlsson)
email: anna.sorjamaa@oulu.fi

openresearchlabs.github.io

Teemu Niiranen



Collaboration

Helsinki / Wageningen
Willem de Vos
Anne Salonen
Sudarshan Shetty

Raes Lab
Jeroen Raes
Karoline Faust

THL
Veikko Salomaa
Teemu Niiranen
Aki Havulinna

Rob Knight
Susan Cheng
Mohit Jain
Michael Inouye
Scott Ritchie

Turku
Turku Microbiome Biobank
Finnbrain
Turku PET Center
STRIP