

Overview of the week

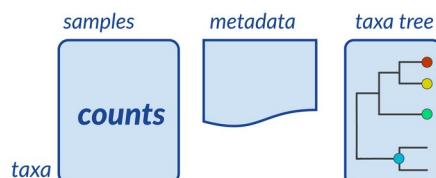
- Day 1: data science framework (*data containers*)
- Day 2: single-assay data analysis (*tabular data*)
- Day 3: multi-assay data analysis (*multi-table data*)
- Day 4: advanced topics, summary & conclusion

Summary & conclusions

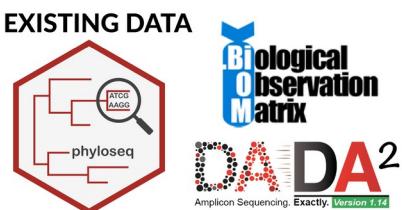
Import Data

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

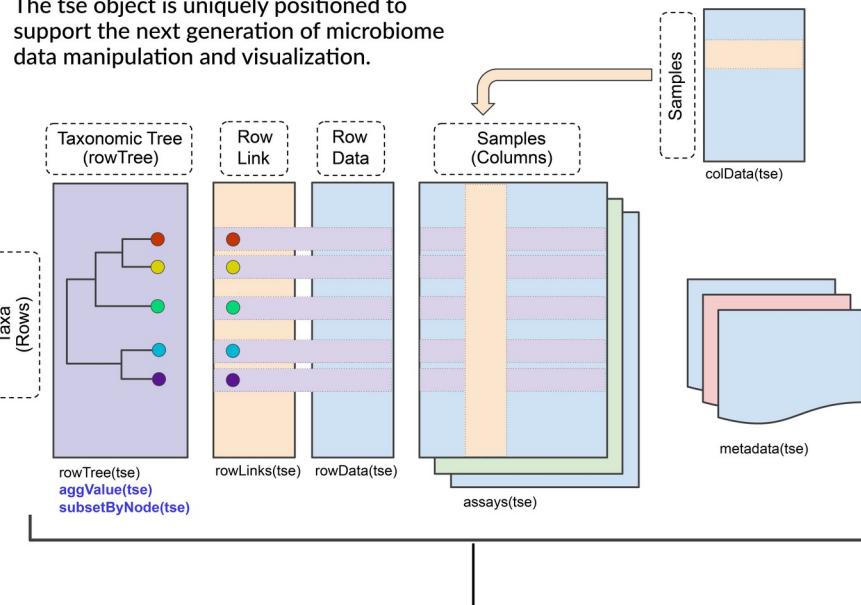


EXISTING DATA



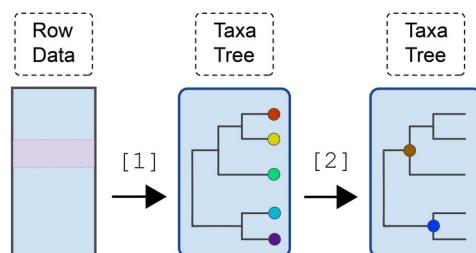
The TreeSE object

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

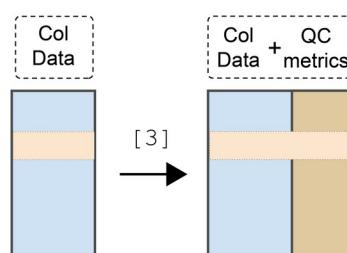


The mia Pipeline

Accessing Taxonomic Info.



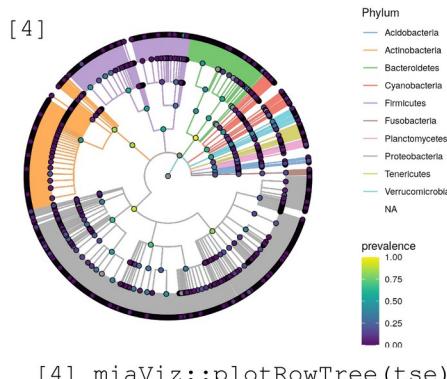
Quality Control



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

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

Visualizing with miaViz



Support & contributions

Online support:

- Gitter

Join R/Bioconductor developer community:

- Github issues & pull requests (package github page)
- R/Bioc Slack
- Email

FINRISK population cohort

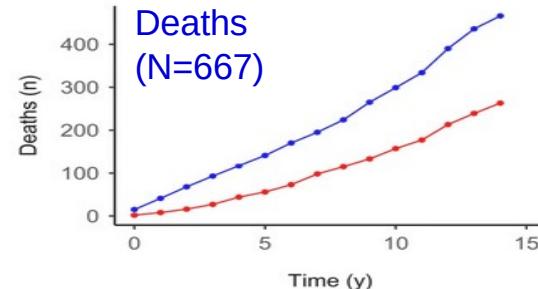
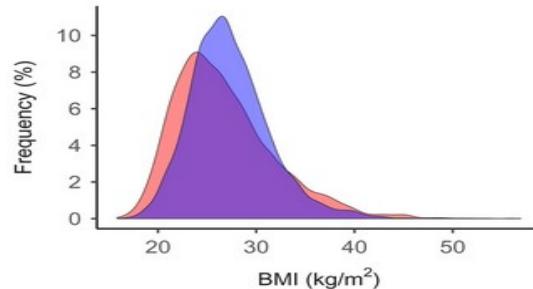
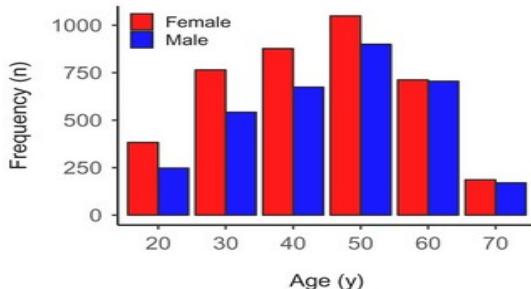
N=7211, 2002

Taxonomic signatures of cause-specific mortality risk in human gut microbiome

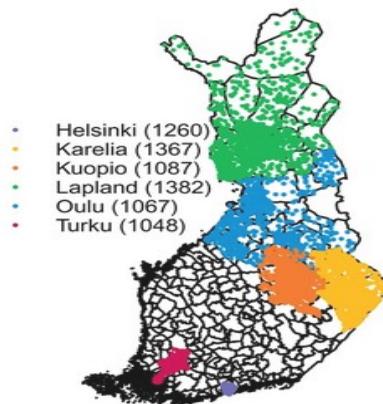
Aaro Saloenssaari, Ville Laitinen, Aki S. Havulinna, Guillaume Meric, Susan Cheng, Markus Perola, Liisa Valsta, Georg Alfrhan, Michael Inouye, Jeremie D. Watrous, Tao Long, Rodolfo A. Salido, Karenina Sanders, Caitriona Brennan, Gregory C. Humphrey, Jon G. Sanders, Mohit Jain, Pekka Jousilahti, Veikko Salomaa, Rob Knight, Leo Lahti & Teemu Nilanen

Nature Communications 12, Article number: 2671 (2021) | Cite this article

a

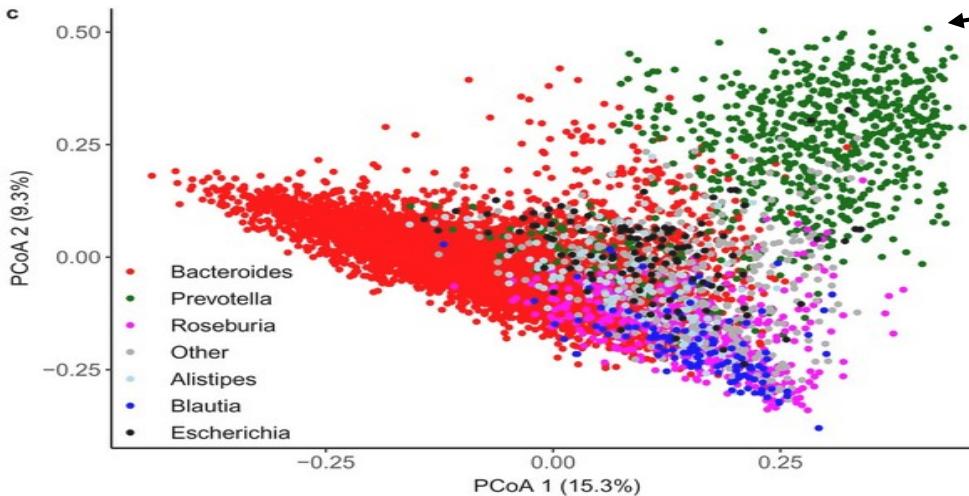


b

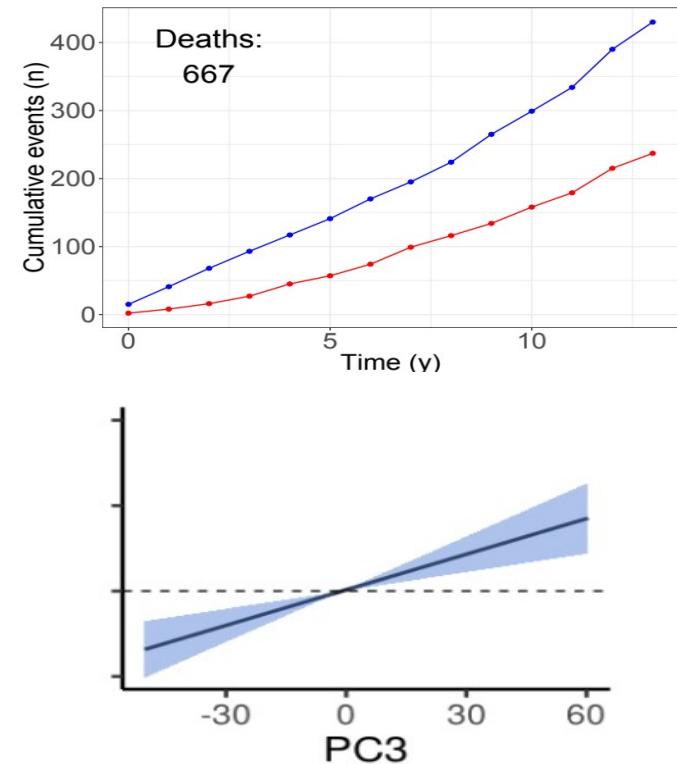
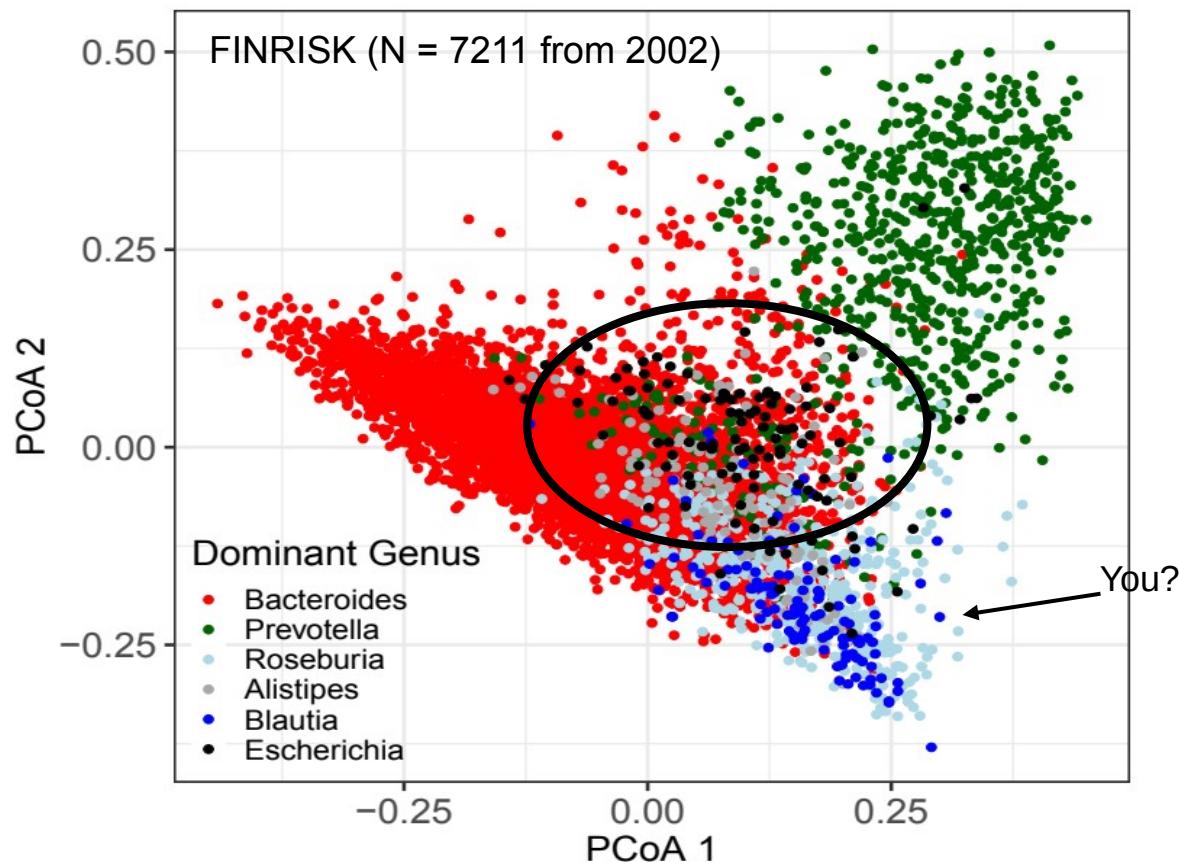


NATIONAL INSTITUTE
FOR HEALTH AND WELFARE

c



Predicting overall mortality risk in Finnish adult population



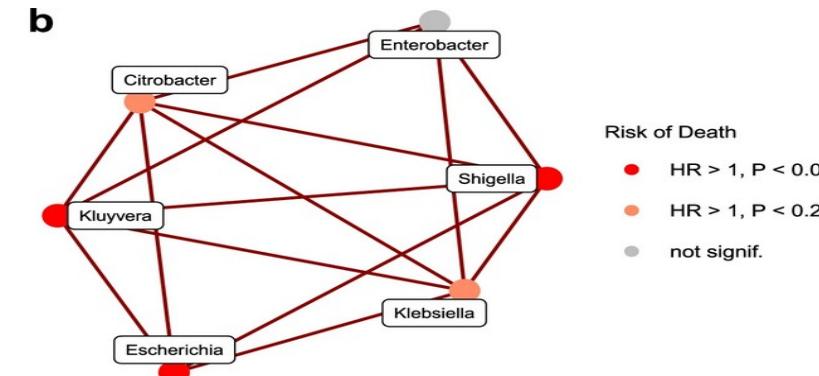
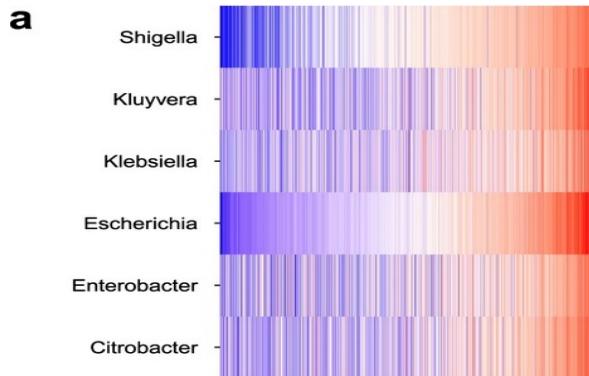
Taxonomic signatures of cause-specific mortality risk in human gut microbiome

Aaro Saloenssaari, Ville Laitinen, Aki S. Havulinna, Guillaume Meric, Susan Cheng, Markus Perola, Liisa Valsta, Georg Alftman, Michael Inouye, Jeramie D. Watrous, Tao Long, Rodolfo A. Salido, Karenina Sanders, Caitriona Brennan, Gregory C. Humphrey, Jon G. Sanders, Mohit Jain, Pekka Jousilahti, Veikko Salomaa, Rob Knight, Leo Lahti & Teemu Niiranen

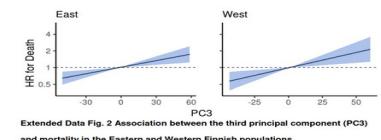
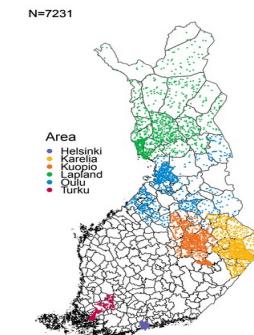
Nature Communications 12, Article number: 2671 (2021) | Cite this article

Fig. 4: Taxonomic subnetwork associated with increased mortality risk.

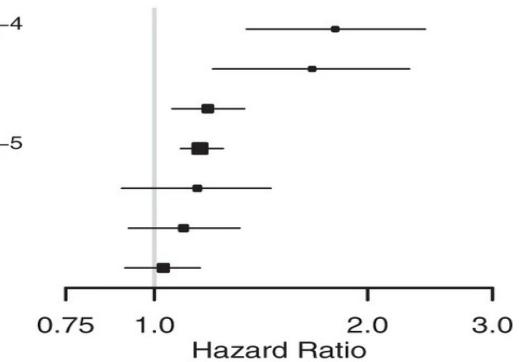
From: [Taxonomic signatures of cause-specific mortality risk in human gut microbiome](#)



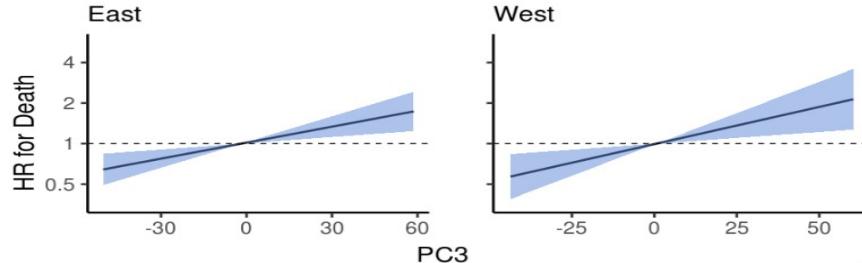
a Abundance variation across the study population for the subnetwork that exhibits the strongest mortality associations (CLR-transformed abundances centred at zero and scaled to unit variance). The samples are ordered by the total relative abundance of the subnetwork. **b** The observed subnetwork structure and mortality risk. The total subnetwork abundance was associated with elevated mortality with a hazard ratio of 1.155 (95% confidence interval [CI], 1.08–1.24; $P = 0.0002$, Wald two-tailed test statistic for Cox regression, 4.07) The respective hazard ratios were 1.17 (95% CI, 1.07–1.27; $P = 0.001$, Wald statistic 3.66) in the Eastern and 1.14 (95% CI, 1.001–1.31; $P = 0.15$, Wald statistic 2.02) in the Western Finnish populations. The analyses are conducted after excluding rare taxa and adjusted for age, body mass index, sex, smoking, diabetes, use of antineoplastic and immunomodulating agents, systolic blood pressure and self-reported antihypertensive medication; P values are FDR-adjusted.



Cause of Death	Deaths	HR	FDR
Gastrointestinal	36	1.80 (1.35–2.41)	2.2×10^{-4}
Respiratory	31	1.67 (1.21–2.29)	0.004
Cancer	238	1.19 (1.06–1.34)	0.007
All	729	1.16 (1.09–1.25)	9.5×10^{-5}
Neurological	60	1.15 (0.90–1.46)	0.333
Other	110	1.10 (0.92–1.32)	0.333
Cardiovascular	254	1.03 (0.91–1.16)	0.633

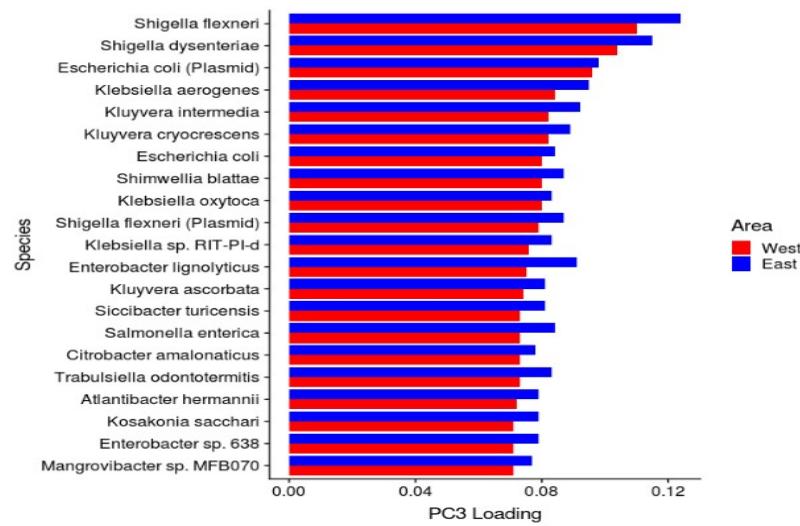


Prospective analysis with FINRISK2002 cohort: PC axes



Extended Data Fig. 2 Association between the third principal component (PC3)

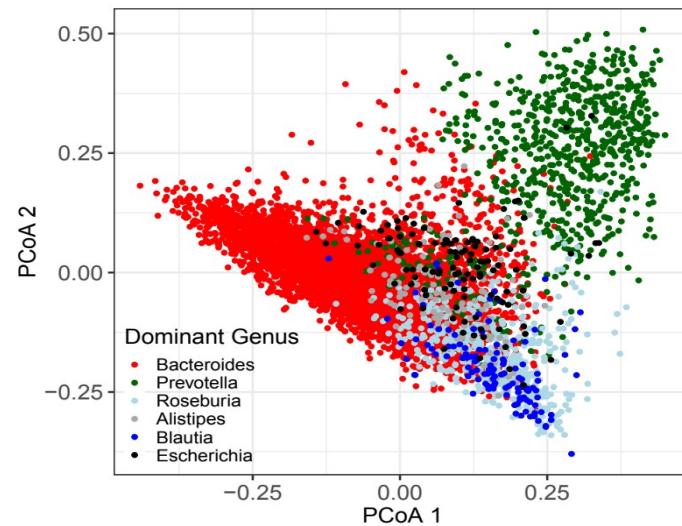
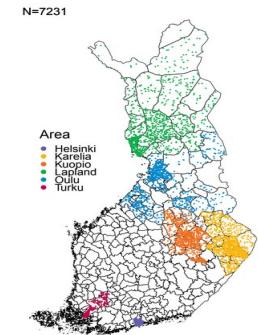
and mortality in the Eastern and Western Finnish populations.



Extended Data Fig. 4. Third principal component drivers in Eastern and Western Finland.

Most important PC3 drivers in Eastern and Western Finland.

**Verification:
two distinct sub-populations**



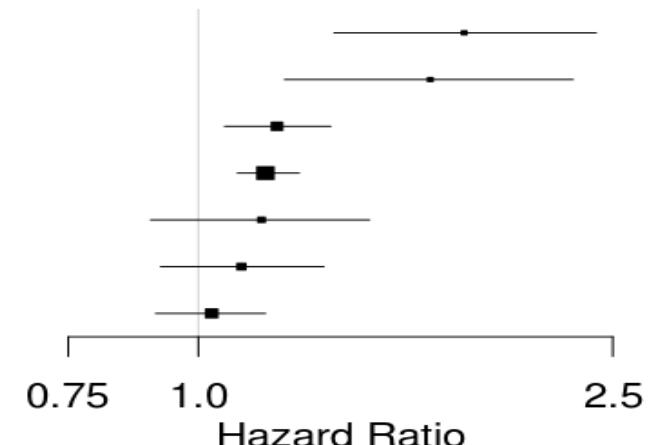
Taxonomic signature associated & specific causes of mortality

Cause of Death

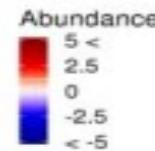
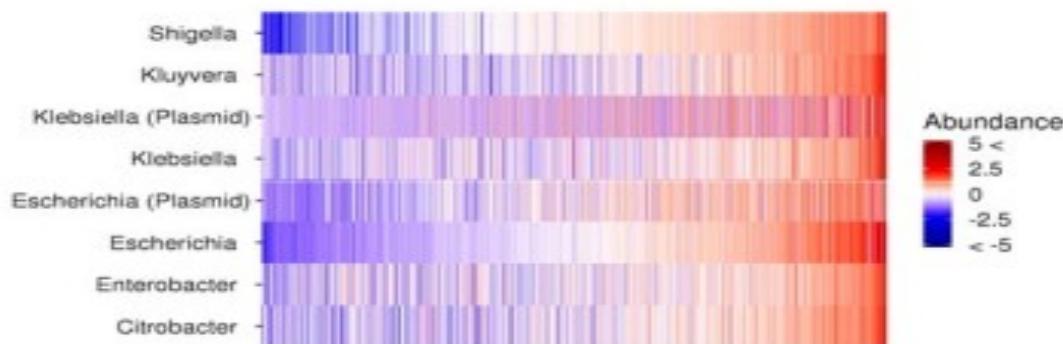
	Deaths	HR
Gastrointestinal	36	1.80 (1.35-2.41)
Respiratory	31	1.67 (1.21-2.29)
Cancer	238	1.19 (1.06-1.34)
All	729	1.16 (1.09-1.25)
Neurological	60	1.15 (0.90-1.46)
Other	110	1.10 (0.92-1.32)
Cardiovascular	254	1.03 (0.91-1.16)

FDR

2.2×10^{-4}
0.004
0.007
9.5×10^{-5}
0.333
0.333
0.633



A



B

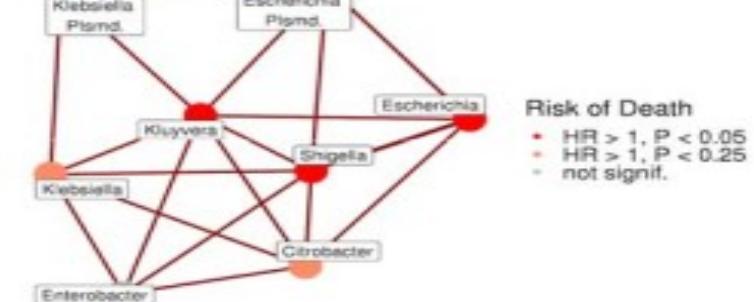
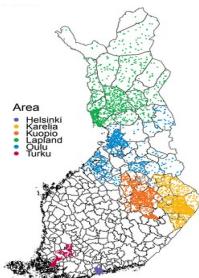
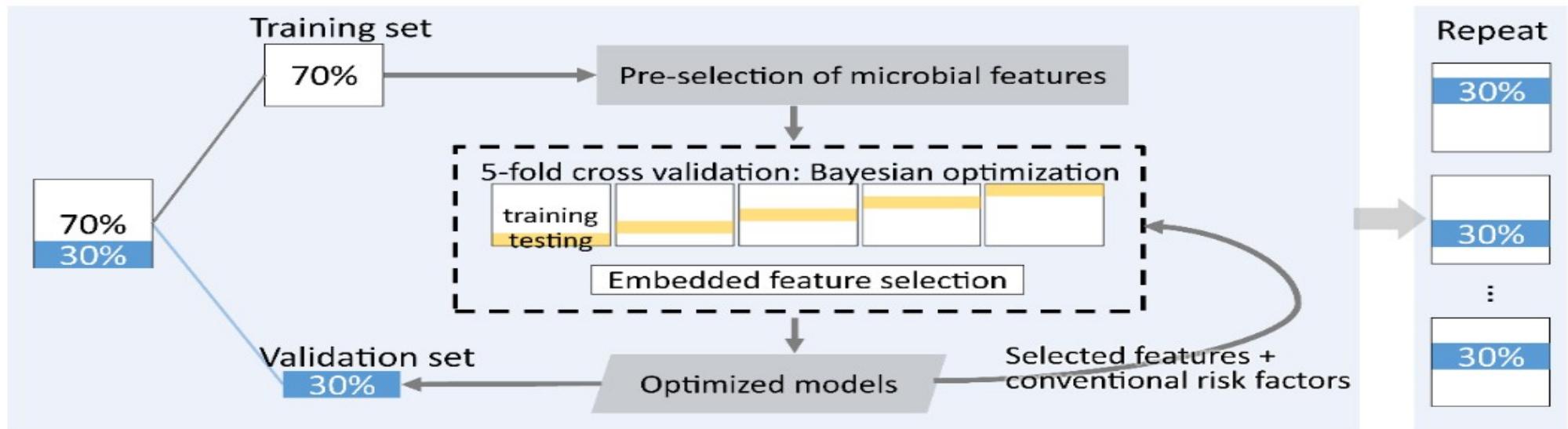


Figure 1. Machine learning framework for predicting incident liver disease



Early prediction of liver disease using conventional risk factors and gut microbiome-augmented gradient boosting

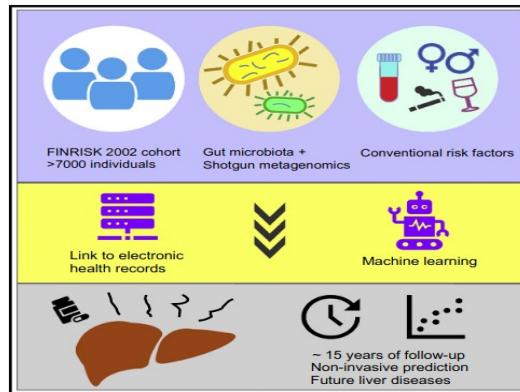
[Comment on this paper](#)

Yang Liu, Guillaume Meric, Aki S Havulinna, Shu Mei Teo, Matti Ruuskanen, Jon Sanders, Qiyun Zhu, Anupriya Tripathi, Karin Verspoor, Susan Cheng, Mo Jain, Pekka Jousilahti, Yoshiki Vazquez-Baeza, Rohit Loomba, Leo Lahti, Teemu Niiranen, Veikko Salomaa, Rob Knight, Michael Inouye

doi: <https://doi.org/10.1101/2020.06.24.20138933>

Early prediction of incident liver disease using conventional risk factors and gut-microbiome-augmented gradient boosting

Graphical abstract



Authors

Yang Liu, Guillaume Méric,
Aki S. Havulinna, ..., Veikko Salomaa,
Rob Knight, Michael Inouye

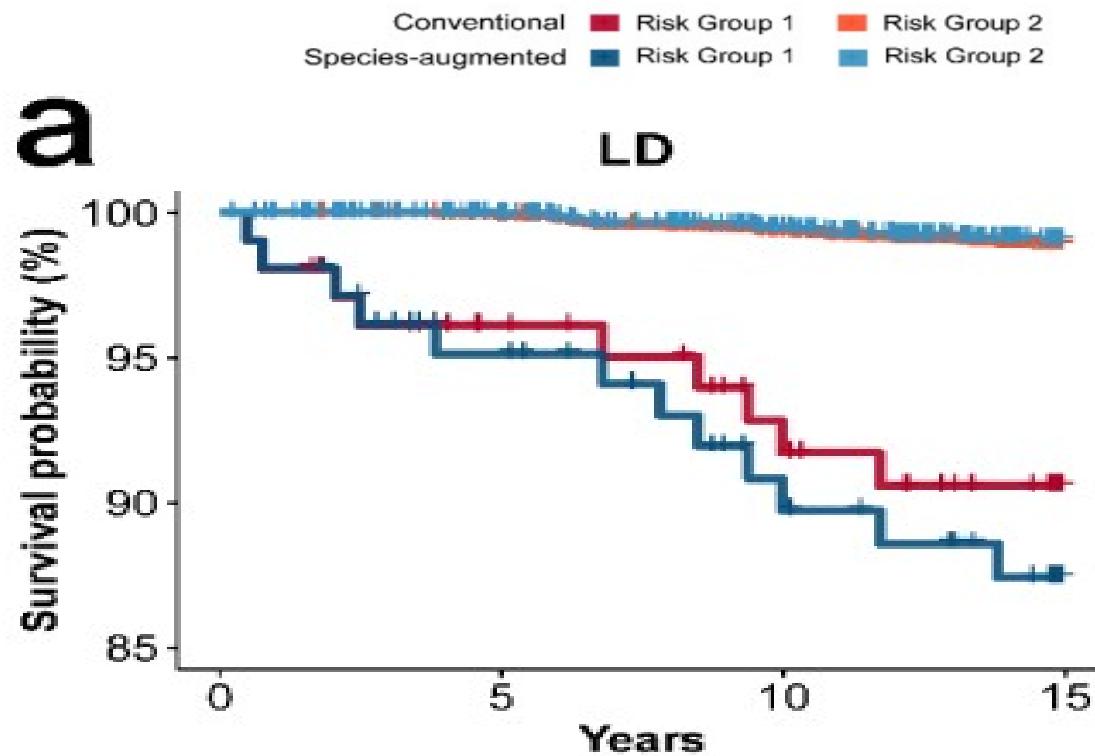
Correspondence

yang.liu2@baker.edu.au (Y.L.),
minouye@baker.edu.au (M.I.)

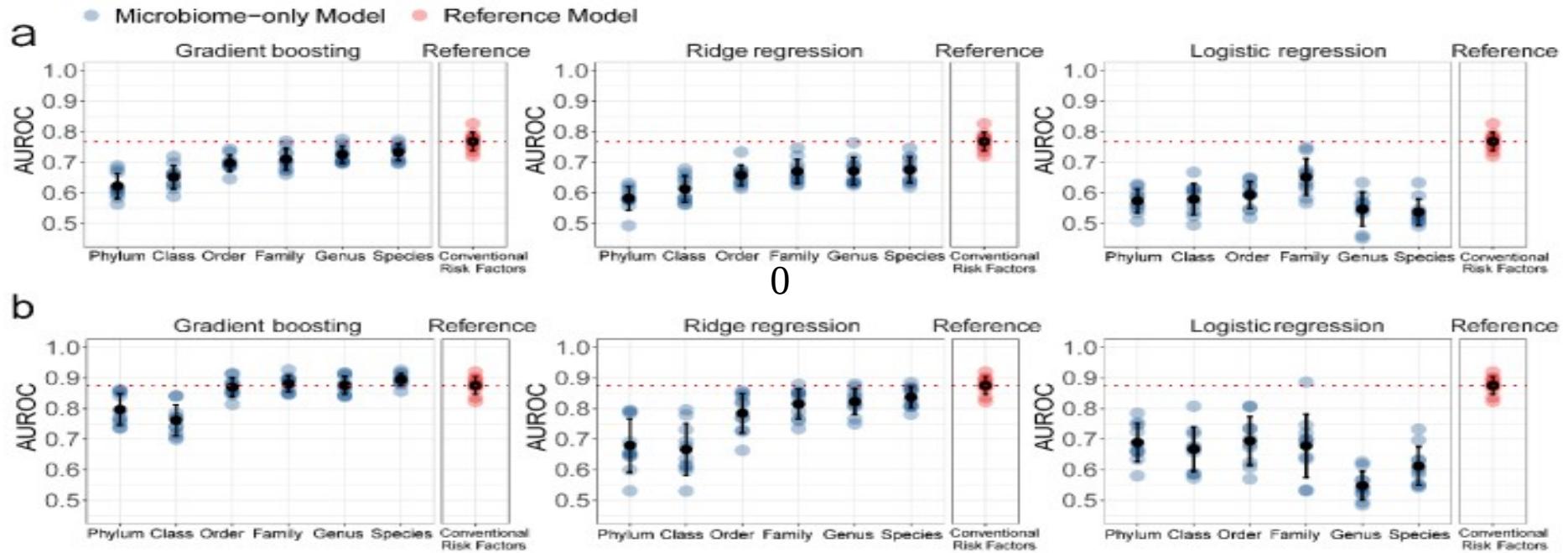
In brief

The predictive capacity of the gut microbiome for future liver diseases is investigated using machine learning in a large prospective cohort.

Survival curves for predicted risk groups of incident liver disease: microbiome signatures improve predictions



Taxonomic resolution and machine learning methodology have notable effect on prediction performance

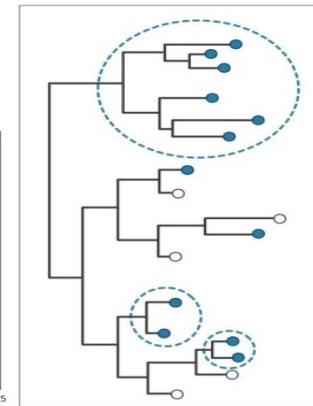
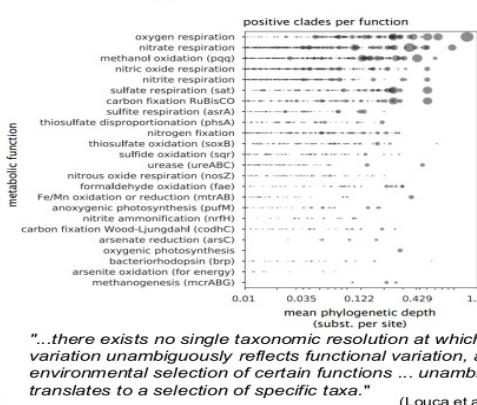
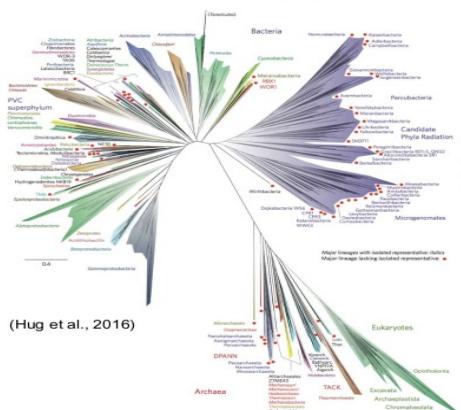


Early prediction of liver disease using conventional risk factors and gut microbiome-augmented gradient boosting

[Comment on this paper](#)

Yang Liu, Guillaume Meric, Aki S Havulinna, Shu Mei Teo, Matti Ruuskanen, Jon Sanders, Qiyun Zhu, Anupriya Tripathi, Karin Verspoor, Susan Cheng, Mo Jain, Pekka Jousilahti, Yoshiki Vazquez-Baeza, Rohit Loomba, Leo Lahli, Teemu Niiranen, Veikko Salomaa, Rob Knight, Michael Inouye
doi: <https://doi.org/10.1101/2020.06.24.20138933>

The use of phylogenetic information in metagenomics



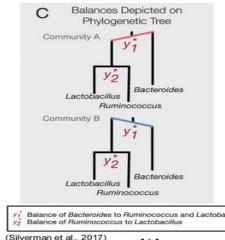
A phylogenetic transform enhances analysis of compositional microbiota data

Justin D Silverman, Alex D Washburne, Sayan Mukherjee, Lawrence A David
Duke University, United States; University of Colorado, United States

Details for FLI and PhILR transform

$$\text{FLI} = \left(e^{0.953 \times \log_e(TG)} + 0.139 \times \text{BMI} + 0.718 \times \log_e(GGT) \right) \\ / \left(1 + e^{0.953 \times \log_e(TG)} + 0.139 \times \text{BMI} + 0.718 \times \log_e(GGT) \right) \times 100$$

(Bedogni et al., 2006)



TURUN YLIOPISTO

Research Paper

Links between gut microbiome composition and fatty liver disease in a large population sample

Matti O. Ruuskanen , Fredrik Åberg , Ville Männistö , Aki S. Havulinna , Guillaume Méric , Yang Liu, ...show all

Pages 1-22 | Received 17 Aug 2020, Accepted 28 Jan 2021, Published online: 02 Mar 2021

Download citation

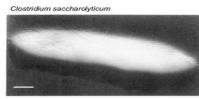
<https://doi.org/10.1080/19490976.2021.1888673>

Check for updates



Conclusions

Gut microbiome composition is connected with FLI



Previously NAFLD-affiliated Clostridia XIVa group bacteria were detected

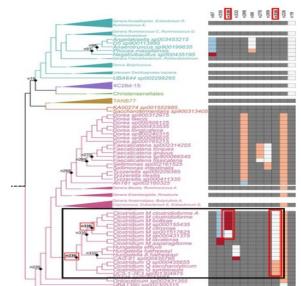
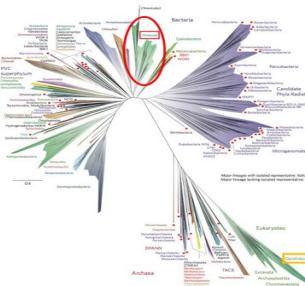
- Two species in Clostridia IV group (*Phocae massiliensis* and *Negativibacillus* sp.) have not been affiliated previously with fatty liver disease

Bacteria in clades connected with FLI likely ferment polysaccharides into SCFA and ethanol, and modify bile acids

- C. bolteae* has also been affiliated with disruption of the intestinal barrier

Our results can be generalized at the population level in Finland and should be followed by studies in other populations, and targeted studies of the specific clades

TURUN YLIOPISTO



Specific clades within Clostridia IV and XIVa subclusters are likely connected with fatty liver pathophysiology

After model selection and optimization:

FLI $\geq 60 \sim \text{age} + \text{gender} + 11 \text{ balances}$
AUC = 0.75

Genera:

- Negativibacillus*
- Phocae*
- Dorea*
- Faecalicitena*
- Sellimonas*
- Clostridium*
- (Lachnospiraceae)*

TURUN YLIOPISTO

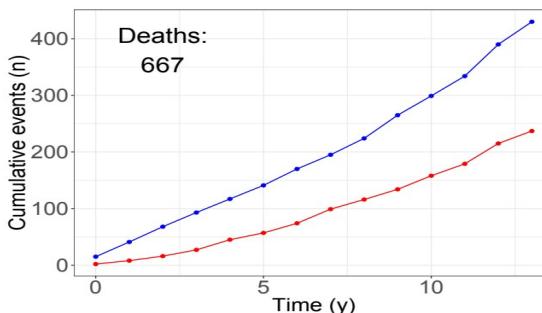
Population cohort studies: FINRISK2002 gut microbiome

2002 ~7000+ adult stool samples:
omics and clinical measurements

2017: 16+ year follow up

comprehensive health information
from Finnish health registers

Shallow shotgun metagenome (~1M
reads/sample), also other info: 16S,
metabolome, host genome & exome,
clinical parameters..



Taxonomic Signatures of Long-Term Mortality Risk in Human Gut Microbiota

Aaro Saloenssaari, Ville Laitinen, Aki Havulinna, Guillaume Meric, Susan Cheng, Markus Perola, Liisa Valsta, Georg Alftan, Michael Inouye, Jeramie D. Watrous, Tao Long, Rodolfo Salido, Karenina Sanders, Caitriona Brennan, Gregory C. Humphrey, Jon G. Sanders, Mohit Jain, Pekka Jousilahti, Veikko Salomaa, Rob Knight, Leo Lahti, Teemu Niiranen

doi: <https://doi.org/10.1101/2019.12.30.19015842>



Research Paper

Links between gut microbiome composition and fatty liver disease in a large population sample

Matti O. Ruuskanen, Fredrik Åberg, Ville Männistö, Aki S. Havulinna, Guillaume Méric, Yang Liu, ...show all

Early prediction of liver disease using conventional risk factors and gut microbiome-augmented gradient boosting

Yang Liu, Guillaume Meric, Aki S. Havulinna, Shu Mei Teo, Matti Ruuskanen, Jon Sanders, Qiyun Zhu, Anupriya Tripathi, Karin Verspoor, Susan Cheng, Mo Jain, Pekka Jousilahti, Yoshiki Vazquez-Baeza, Rohit Loomba, Leo Lahti, Teemu Niiranen, Veikko Salomaa, Rob Knight, Michael Inouye

doi: <https://doi.org/10.1101/2020.06.24.20138933>

Comment on this paper

Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort

Youwen Qin, Aki S. Havulinna, Yang Liu, Pekka Jousilahti, Scott C. Ritchie, Alex Tokolyi, Jon G. Sanders, Liisa Valsta, Marta Brožíkova, Qiyun Zhu, Anupriya Tripathi, Yoshiki Vazquez-Baeza, Rohit Loomba, Susan Cheng, Mohit Jain, Teemu Niiranen, Leo Lahti, Rob Knight, Veikko Salomaa, Michael Inouye, Guillaume Méric

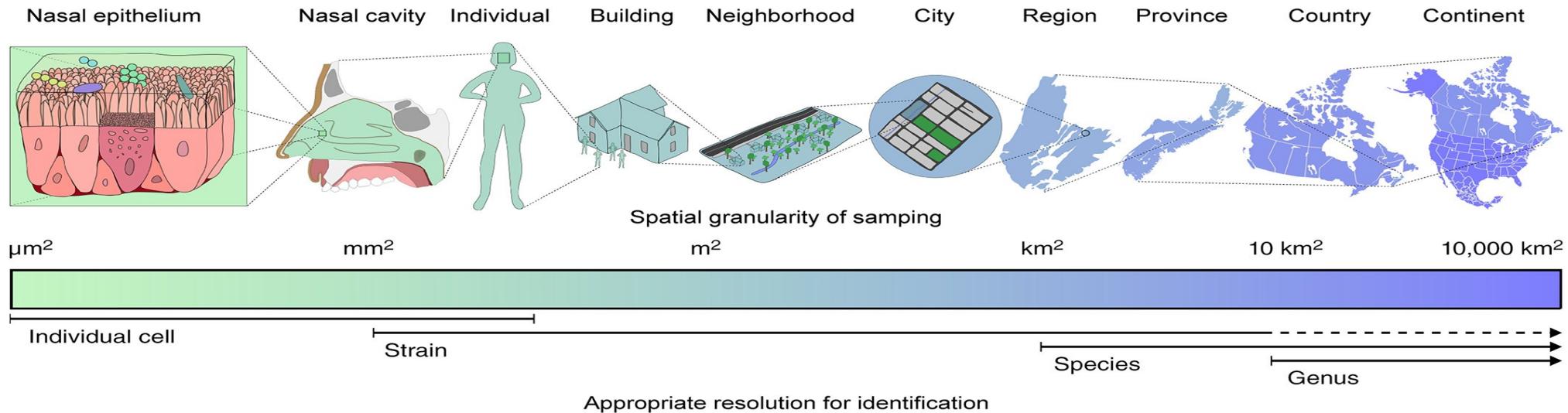
doi: <https://doi.org/10.1101/2020.09.12.20193045>

Comment on this paper

OGUs enable effective, phylogeny-aware analysis of even shallow metagenome community structures

Qiyun Zhu, Shi Huang, Antonio Gonzalez, Imran McGrath, Daniel McDonald, Niina Haiminen, George Armstrong, Yoshiki Vázquez-Baeza, Julian Yu, Justin Kuczynski, Gregory D. Sepich-Poore, Austin D. Swafford, Promi Das, Justin P. Shaffer, Franck Lejzerowicz, Pedro Belda-Ferre, Aki S. Havulinna, Guillaume Méric, Teemu Niiranen, Leo Lahti, Veikko Salomaa, Ho-Cheol Kim, Mohit Jain, Michael Inouye, Jack A. Gilbert, Rob Knight

doi: <https://doi.org/10.1101/2021.04.04.438427>



Environmental Microbiology, Volume: 23, Issue: 5, Pages: 2374-2388, First published: 18 March 2021, DOI: (10.1111/1462-2920.15462)

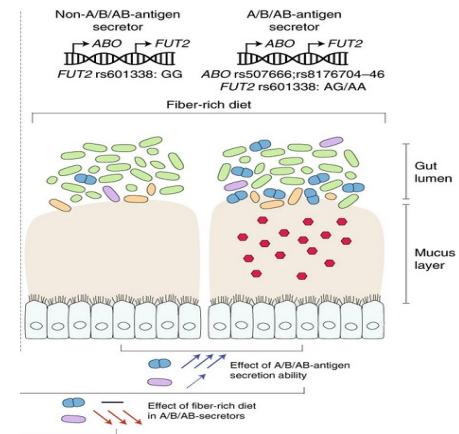
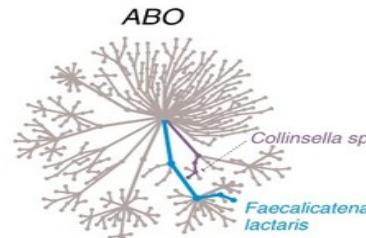
Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort

Youwen Qin, Aki S. Havulinna, Yang Liu, Pekka Jousilahti, Scott C. Ritchie, Alex Tokolyi, Jon G. Sanders, Liisa Valsta, Marta Brožínska, Qijun Zhu, Anupriya Tripathi, Yoshiaki Vázquez-Baeza, Rohit Loomba, Susan Cheng, Mohit Jain, Teemu Niiranen, Leo Lahti, Rob Knight, Veikko Salomaa, Michael Inouye & Guillaume Méric

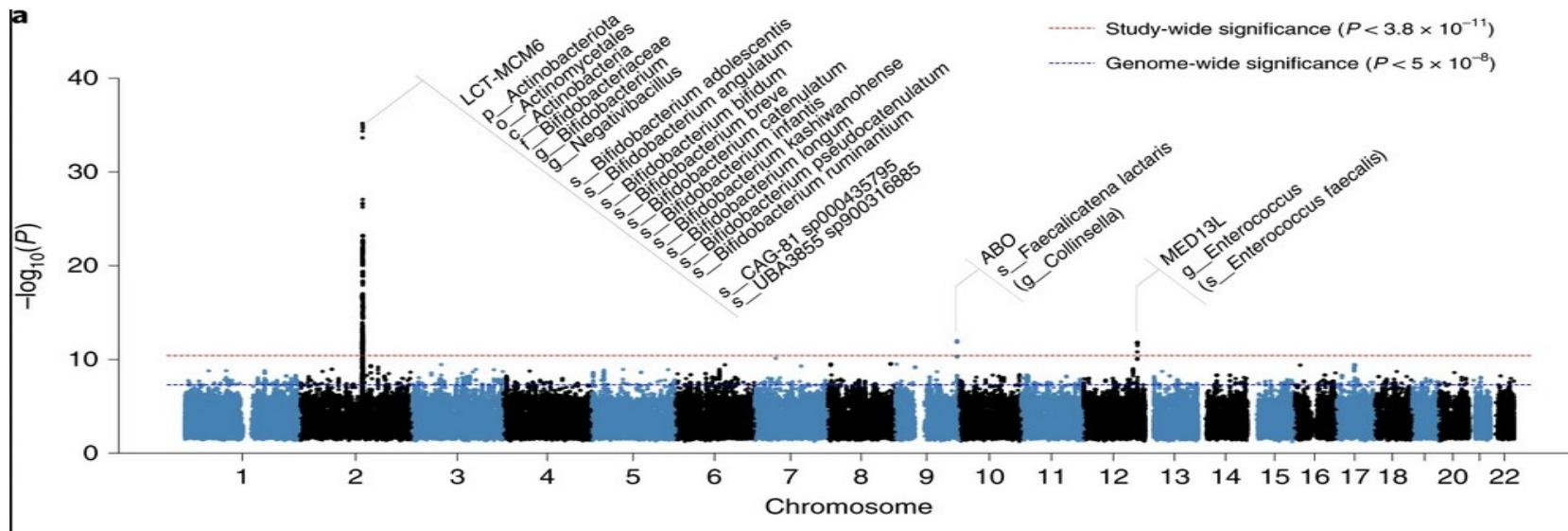
Nature Genetics 54, 134–142 (2022) | Cite this article

22k Accesses | 6 Citations | 241 Altmetric | Metrics

Fig. 4: Effects of host genetics and dietary fiber intake on gut abundance variation of two bacterial taxa associated with the *ABO* locus.



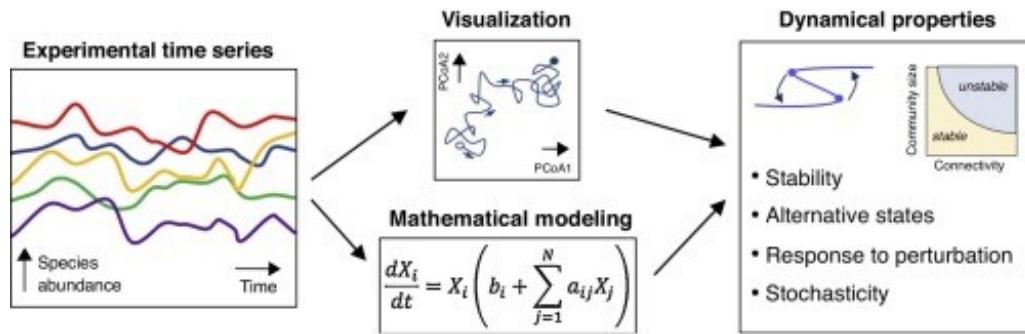
Top SNP locus -microbe associations



Temporal dynamics & critical transitions in microbial communities

Microbial communities as dynamical systems

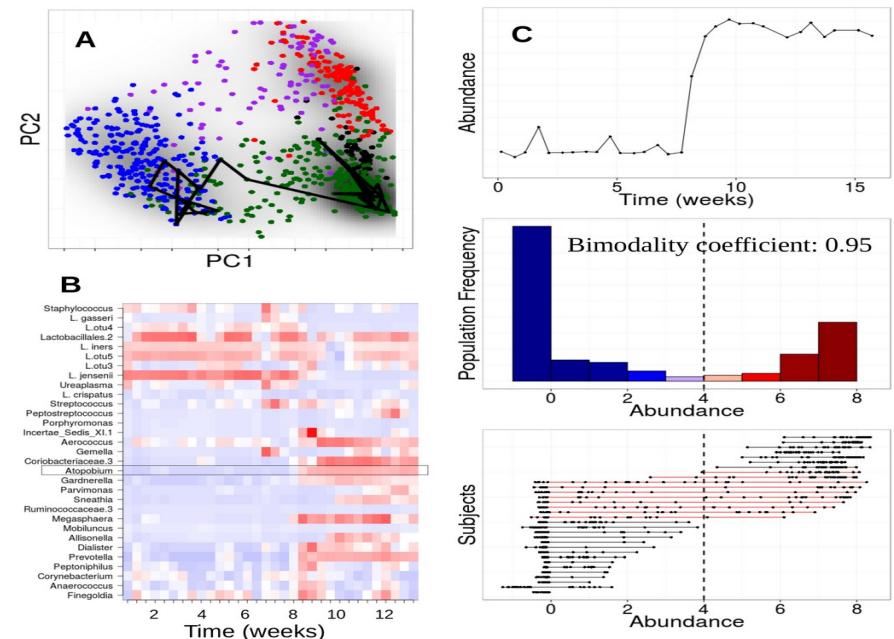
Didier Gonze^{1, 2}✉, Katharine Z Coyte^{3, 4}, Leo Lahti^{5, 6, 7}, Karoline Faust⁵✉



Neutral models, stochasticity, interaction models (gLV), tipping points, memory effects, individuality, eco-evo

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}



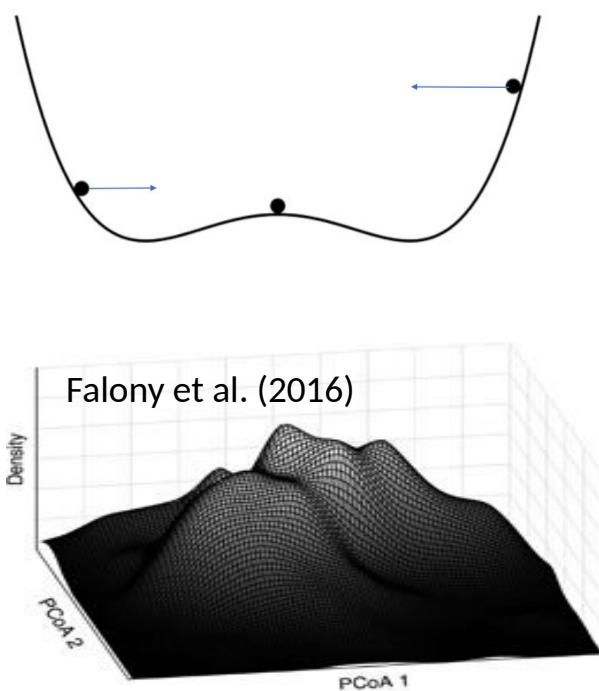
Data: Vaginal microbiota time series from 32 women. Gajer et al. 2012

Three-Species Lotka-Volterra Model with Respect to Caputo and Caputo-Fabrizio Fractional Operators

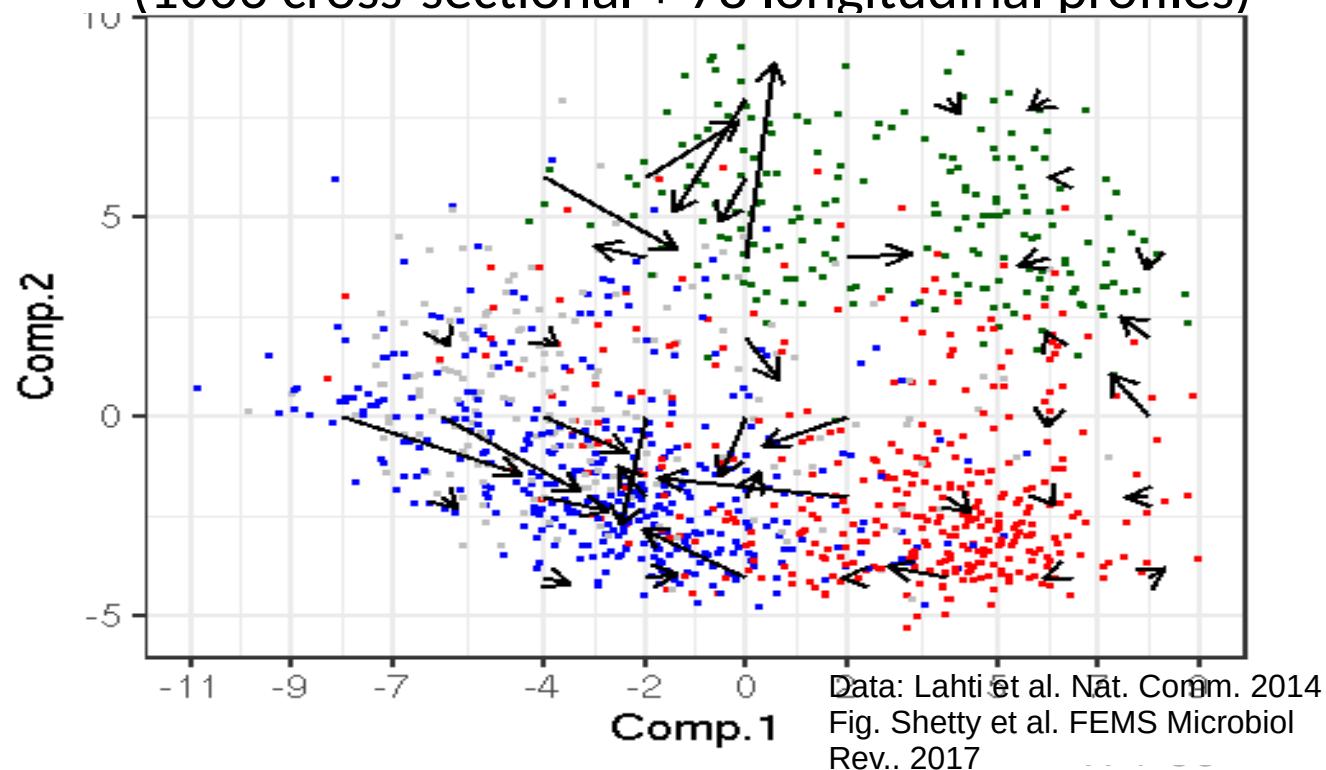
by Moein Khalighi^{1,*}✉, Leila Eftekhari²✉, Soleiman Hosseinpour³✉ and Leo Lahti¹✉

Towards a dynamic landscape model of the gut microbiome

Prior info from
background cohorts &
pooling evidence across
individuals

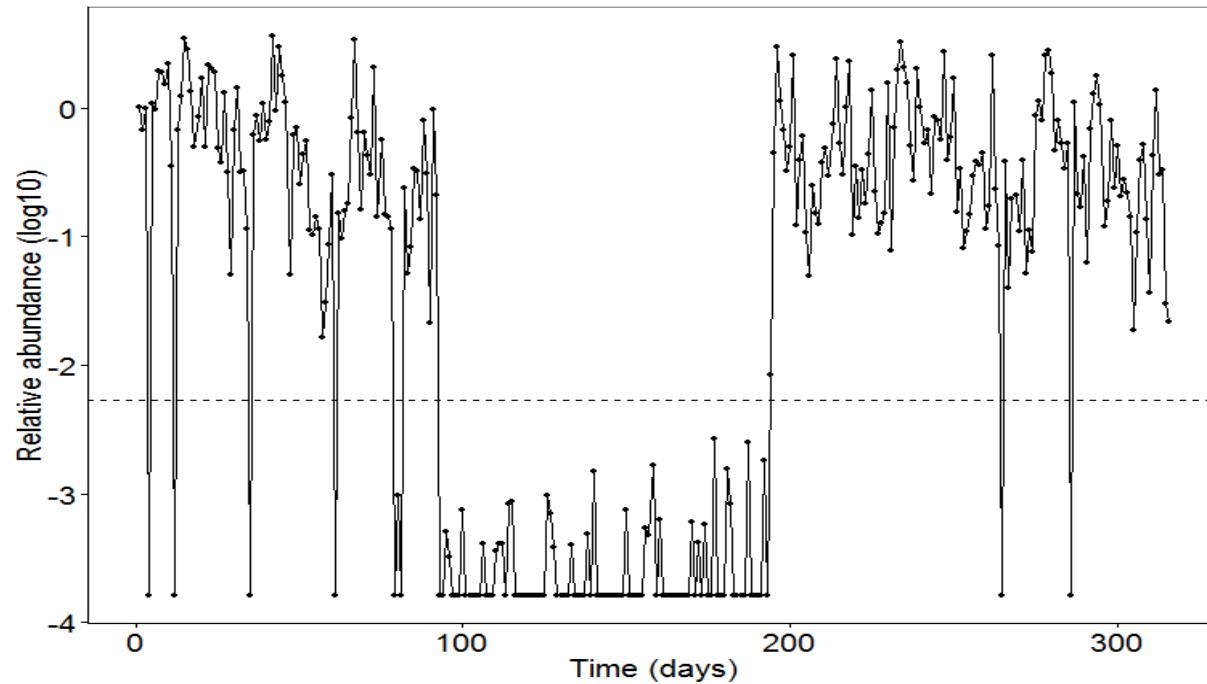


Gradients of change on HITChip PCA landscape
(1006 cross-sectional + 78 longitudinal profiles)



Long and dense time series

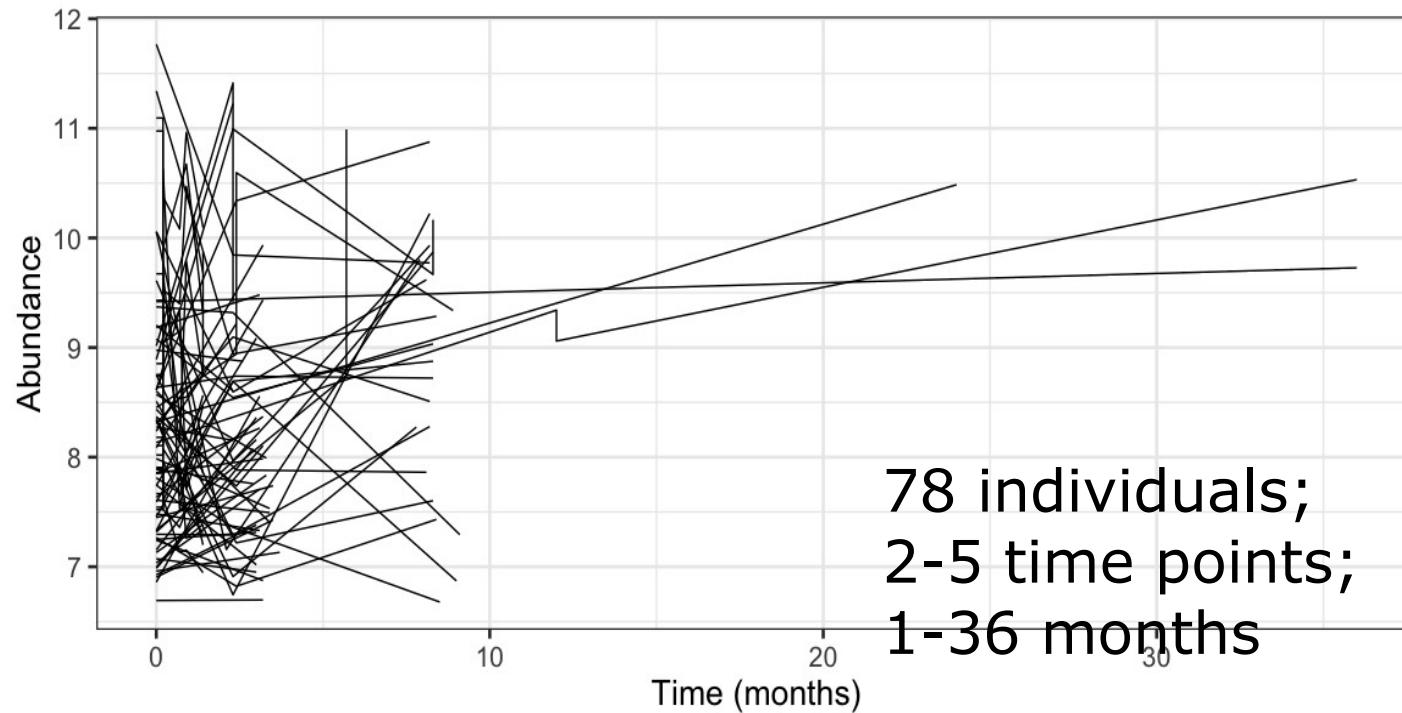
1 individual & 300+ days



Data: David et al (2012)

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

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

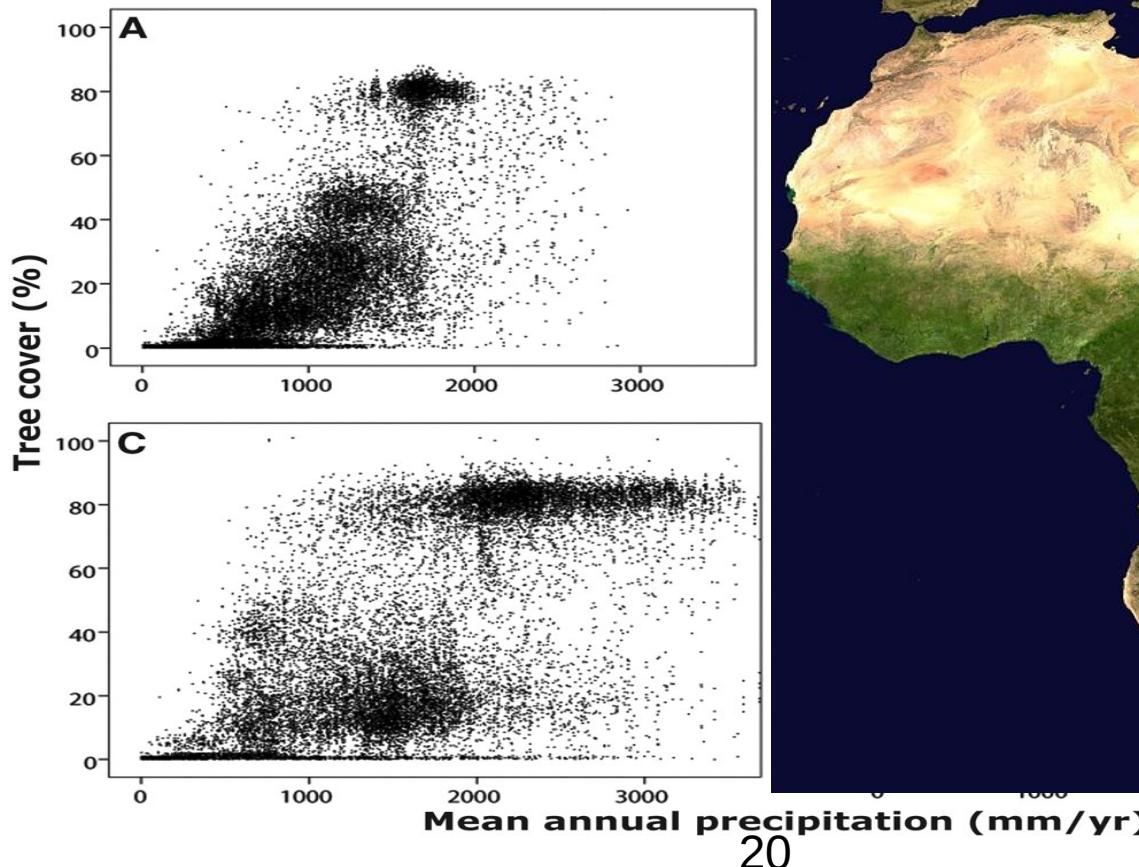


Data:
HITChip Atlas
Lahti et al (2014)

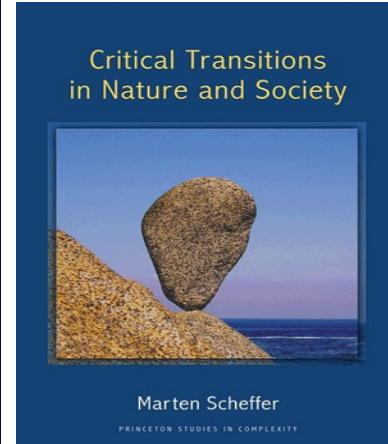
REPORT

Global Resilience of Tropical Forest and Savanna to Critical Transitions

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



20



20 / 53

Fecal transplant induced ecosystem level state shift

Efficient treatment in C. difficile infected patients.
Shift from Bacilli- and Proteobacteria-dominated state to a healthy state dominated by Bacteroides
and many butyrate producing taxa

Better than
antibiotics !

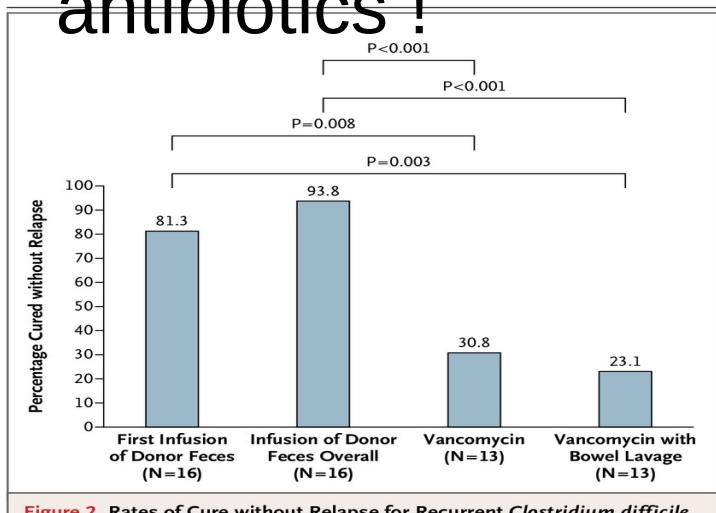
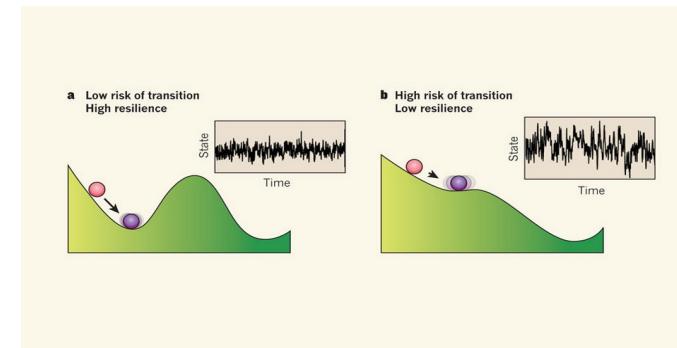


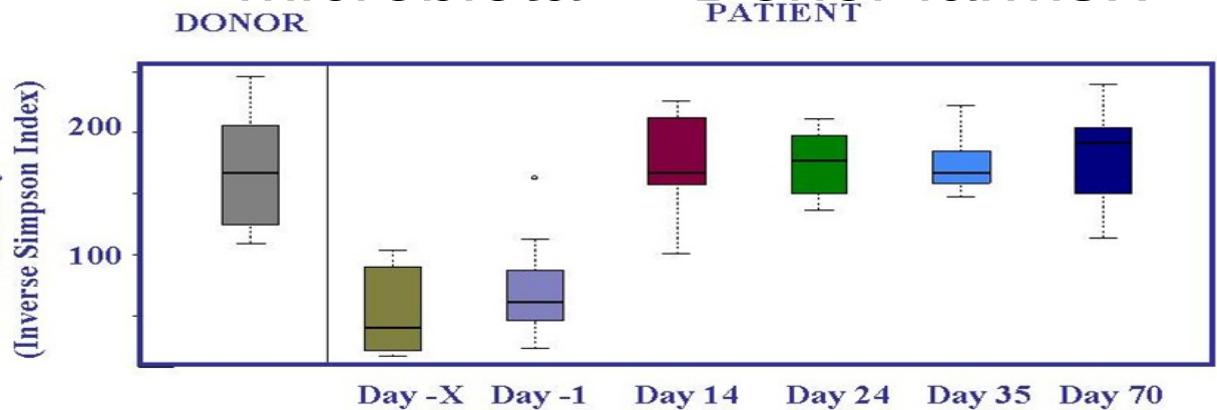
Figure 2. Rates of Cure without Relapse for Recurrent *Clostridium difficile* Infection.

Shown are the proportions of patients who were cured by the infusion of donor feces (first infusion and overall results), by standard vancomycin therapy, and by standard vancomycin therapy plus bowel lavage.

21 / 53



Patient gets donor
microbiota -> Donor farms!?

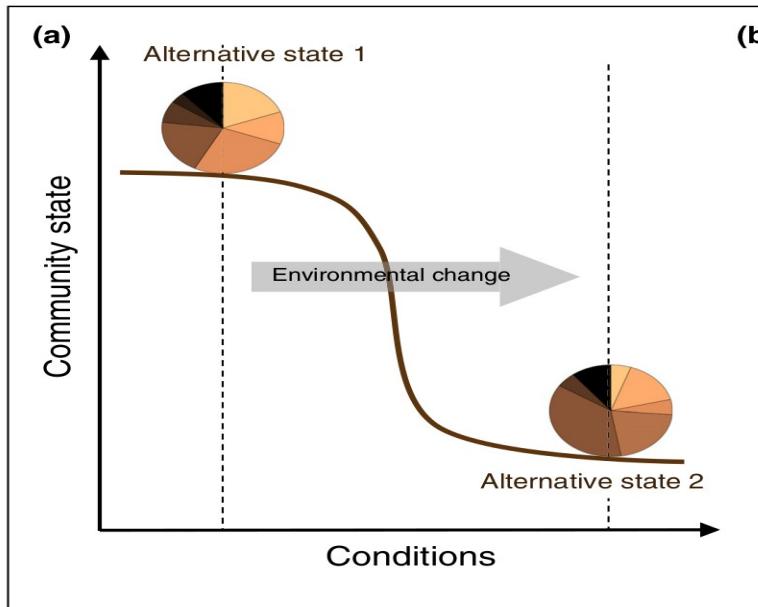


Els van Nood et al., NEJM 368(5) 2013
Fuentes et al. ISME J, 8:1621-33, 2014

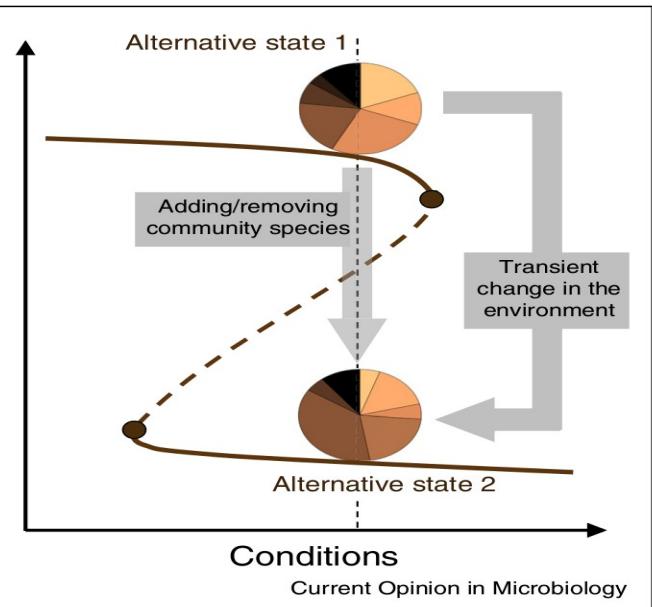
Alternative (stable) states

Diversit
y?

State induced by
external factors



Intrinsic stability:
robust to external factors



Environmental or host factor (e.g. body-mass index)

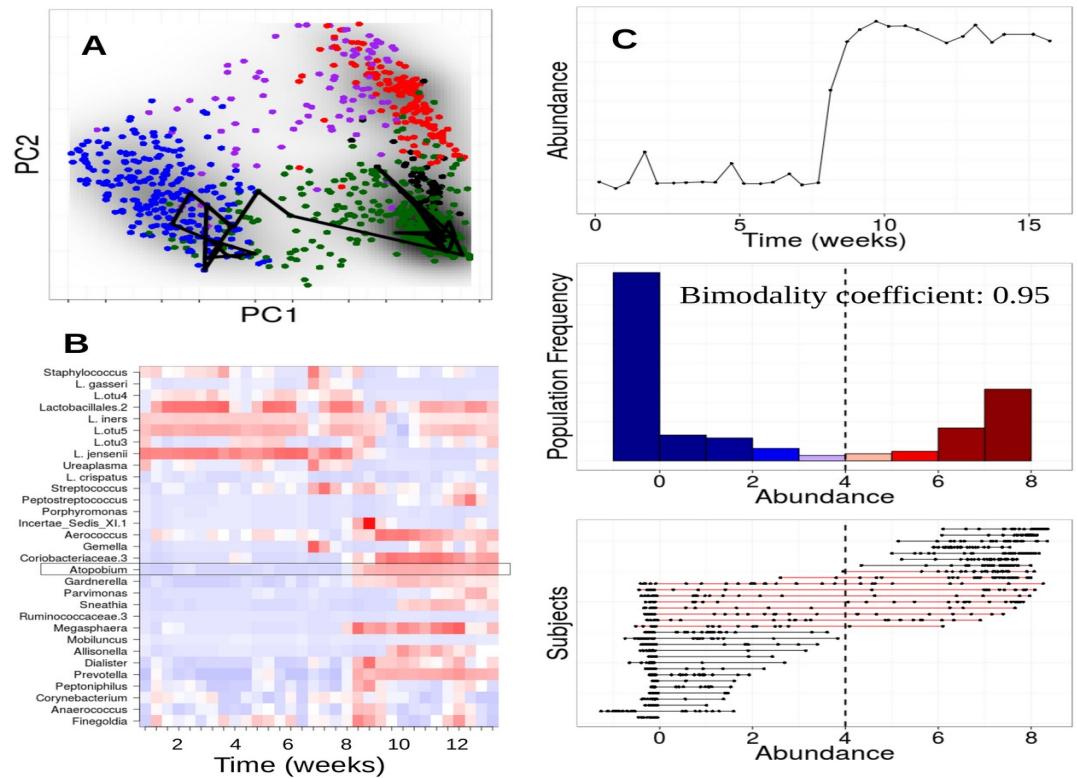
Faust et al. Curr. Op. Microbiol. 2015

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:

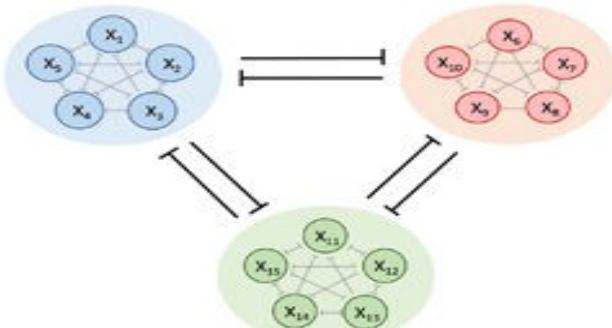
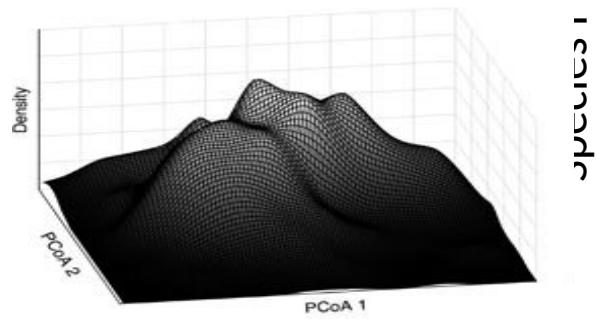
- cross-sectional
- spatial
- temporal



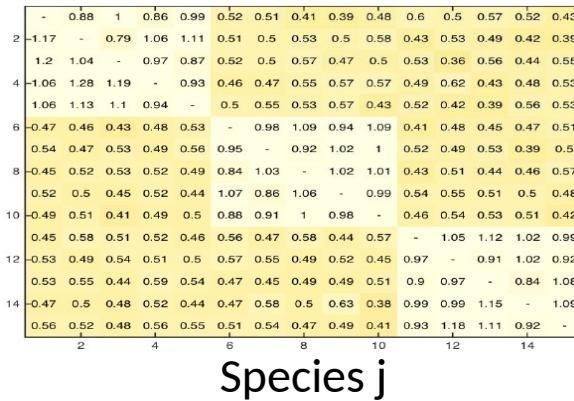
Data: Vaginal microbiota time series from 32 women. Gajer et al. 2012

23 / 53

Interaction models yield more complex dynamics and predict alternative stable states



Legend



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

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

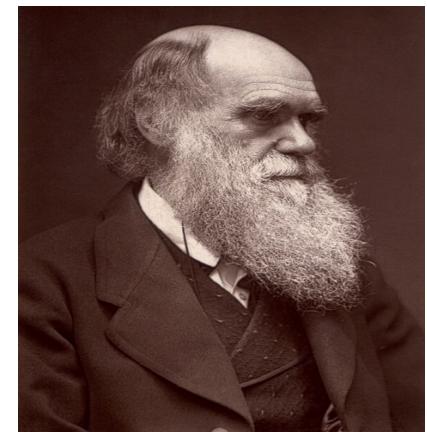
$$f_i(\{X_k\}) = \prod_{\substack{k=1 \\ k \neq i}}^N \frac{K_{ik}^n}{K_{ik}^n + X_k^n}$$

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

How to explain the observed variation –classical 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!"

– Charles Darwin, The Origin of Species.



Neutral vs. niche models in ecology ?

After >25 years on the Barro Colorado Island tropical forests,
Hubbell controversially proposed that..

random chance may in fact be the best explanation of the
observed biodiversity (Hubbell 2001).



Functional equivalence hypothesis

Differences between individuals are neutral in terms of fitness, regardless of their species -> Identical *per capita* demographic rates (birth, death) - At least within specific taxonomic guild.

-> No competition but stochastic fluctuations !

How much niche differences affect community assembly, stability and resilience ?

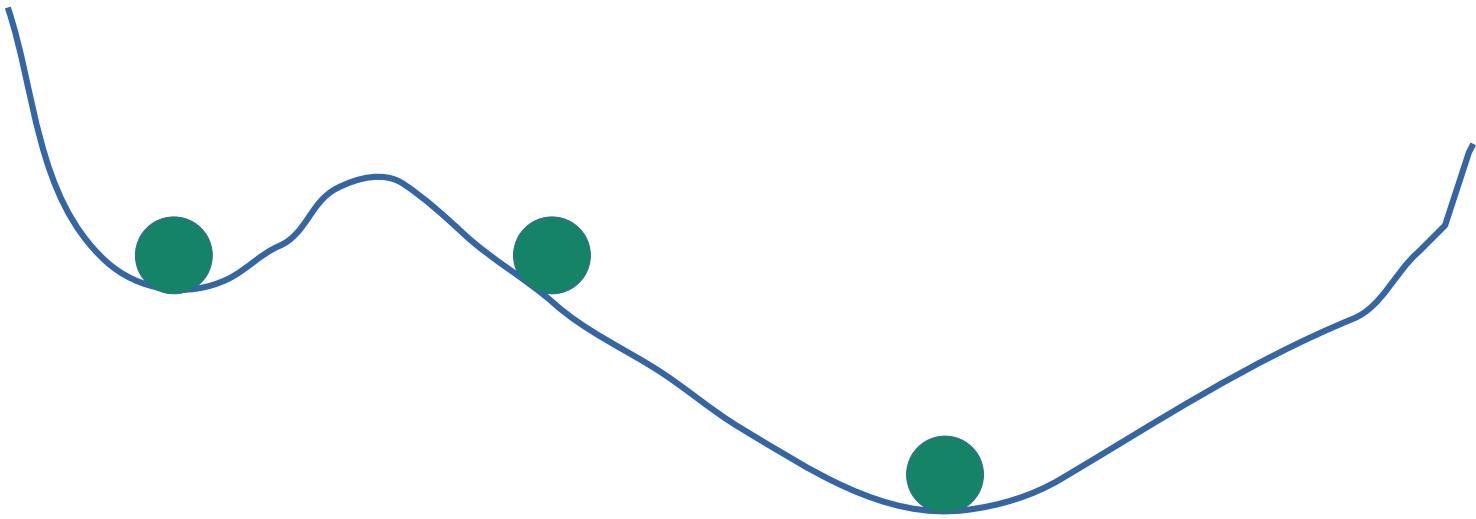


Hubbell:

Is there a limiting niche similarity for species in functional groups? -> No evidence in the investigated tree communities

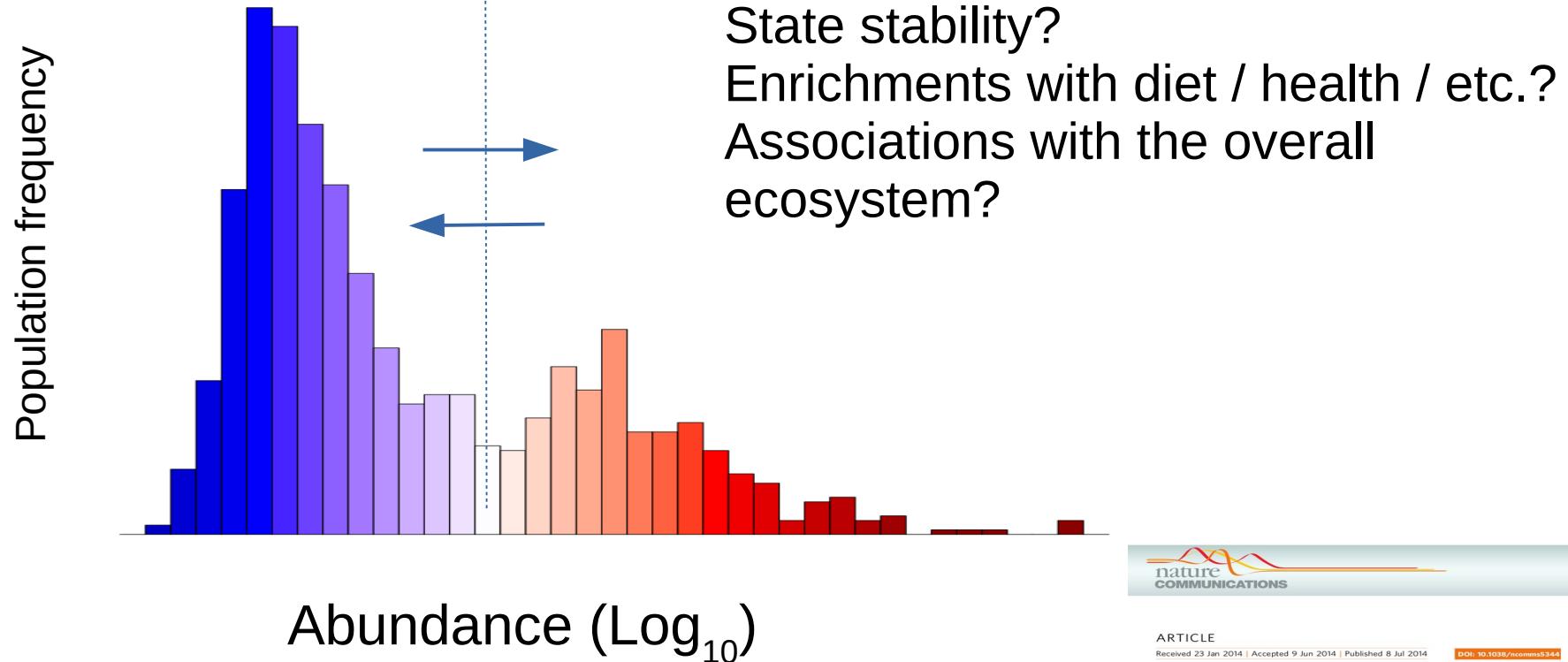
How many coexisting tree species can be classified in the same functional group? -> Arbitrarily many

Resilience



Alternative states in specific bacterial groups ? → Bimodal population distribution is one indicator

Dialister spp. (N = 1006)



ARTICLE
Received 23 Jan 2014 | Accepted 9 Jun 2014 | Published 8 Jul 2014 | DOI: 10.1038/ncomms5344 | OPEN
Tipping elements in the human intestinal ecosystem
Leo Lahti^{1,2}, Jarkko Salojärvi^{1,*}, Anne Salonen^{3,*}, Marten Scheffer⁴ & Willem M. de Vos^{1,2,3}

Mean exit time as a measure of “life expectancy” estimated from fluctuations.

Science

Current Issue First release papers Archive About ▾

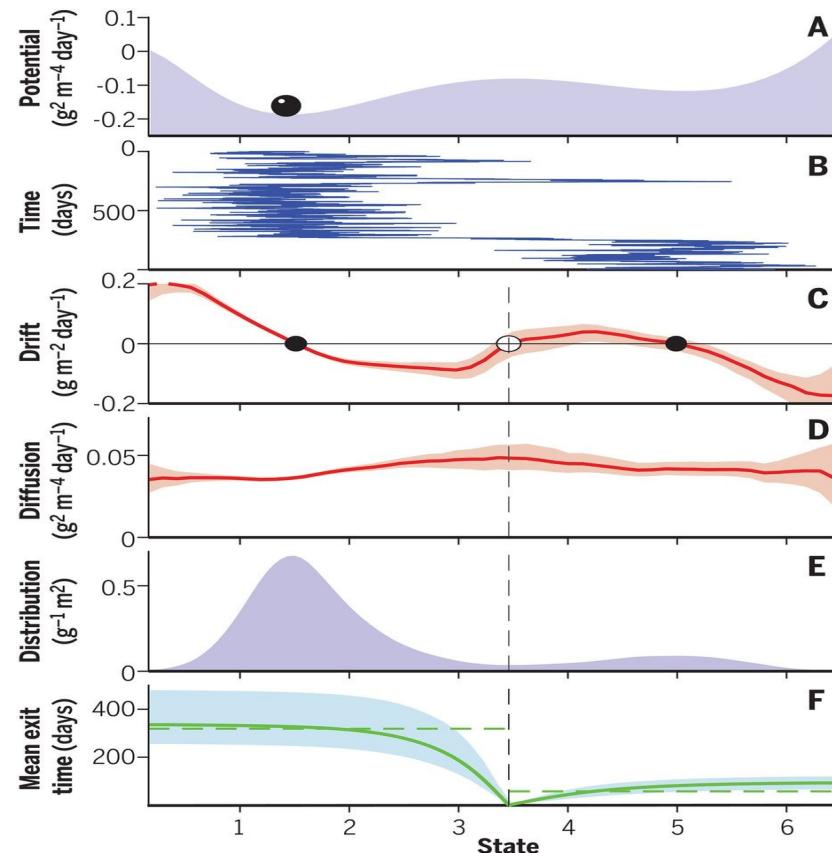
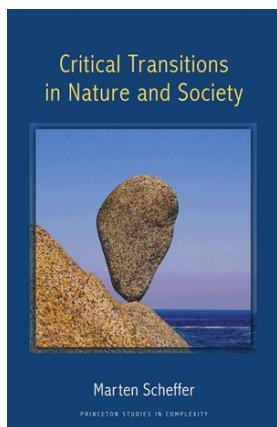
HOME > SCIENCE > VOL. 372, NO. 6547 > EXIT TIME AS A MEASURE OF ECOLOGICAL RESILIENCE

RESEARCH ARTICLE

Exit time as a measure of ecological resilience

BABAK M. S. ARANI , STEPHEN R. CARPENTER , LEO LAHTI , EGBERT H. VAN NES , AND MARTEN SCHEFFER  [Authors Info & Affiliations](#)

SCIENCE • 11 Jun 2021 • Vol 372, Issue 6547 • DOI:10.1126/science.aaay4895

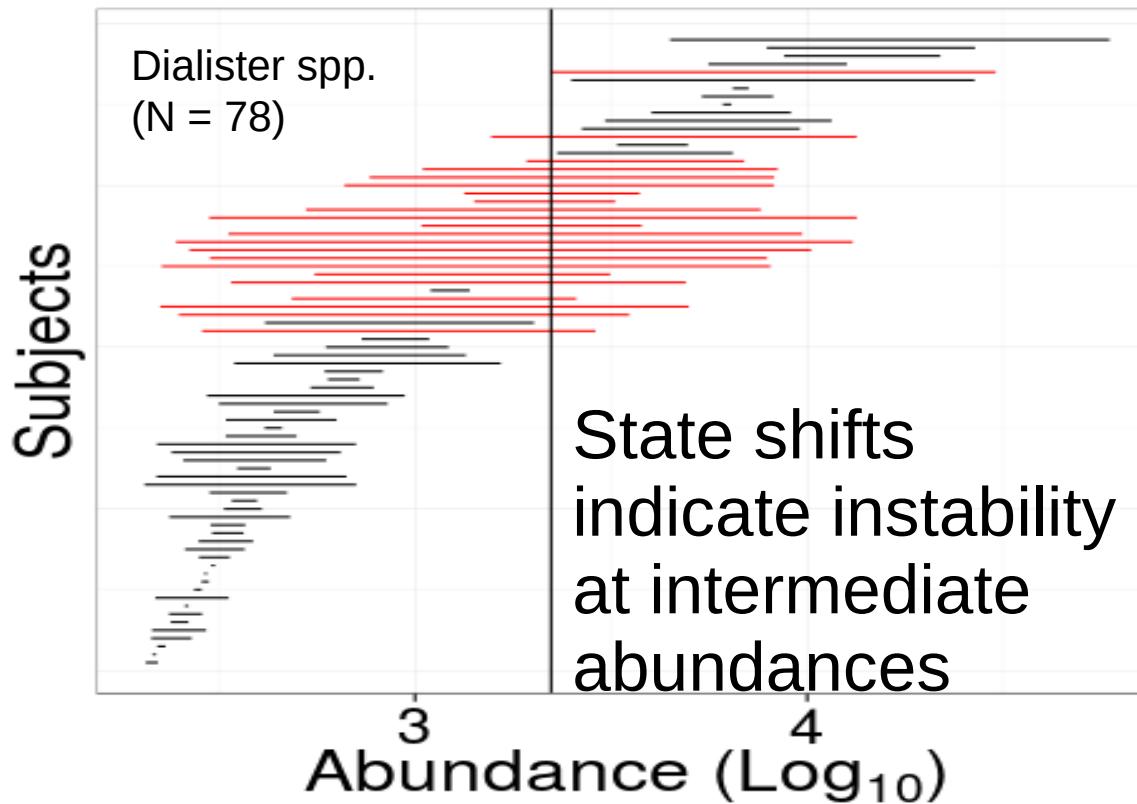


Babak M. S. Arani et al. Science 2021;372:eaay4895

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Bi-stability analysis with many short time series

Non-parametric potential analysis & Fokker-Planck dynamics



Lahti et al. Nat. Comm 2014

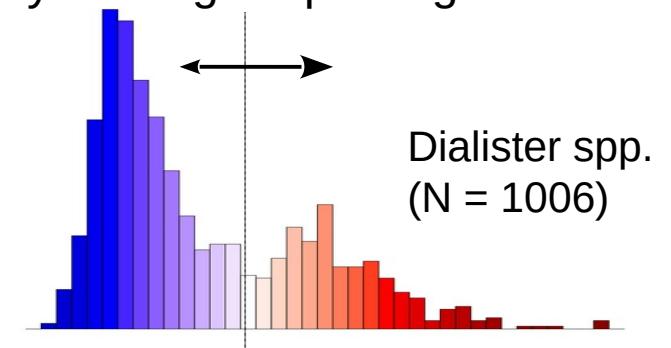
Hirota et al. Science 2011

Livina et al. Clim Past. 2010

System state described by a stochastic process:
 $dz = -U(z)dt + \sigma dW$

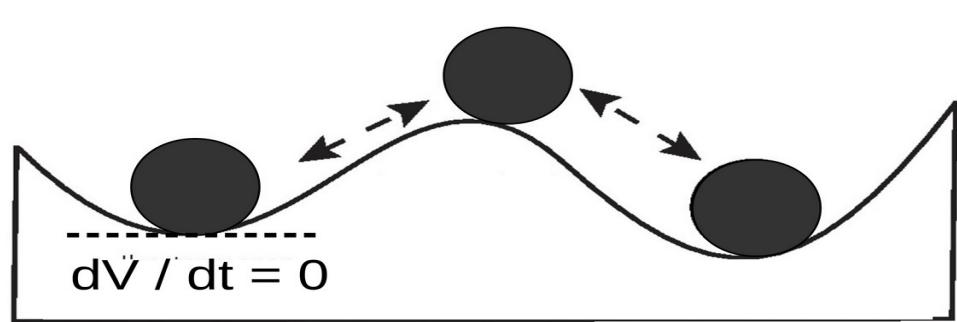
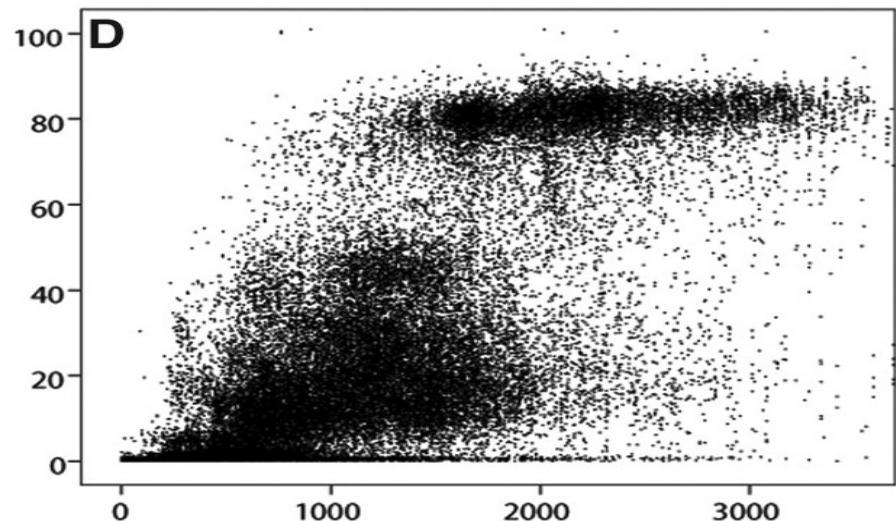
$$U(z) = -\frac{\sigma^2}{2} \log P(z)$$

Used in climatology and ecology.
earlywarnings R package

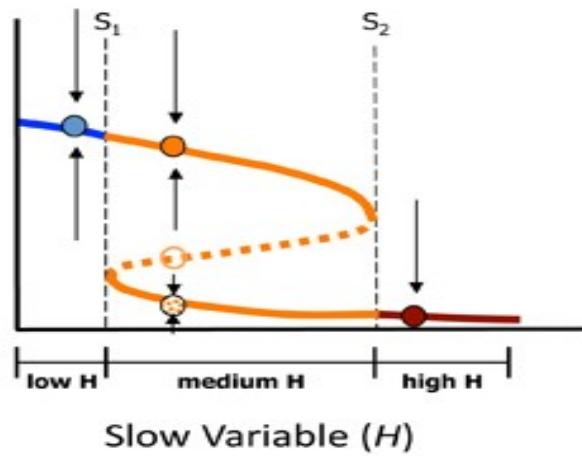


Tipping point

Kasvillisuus (V)

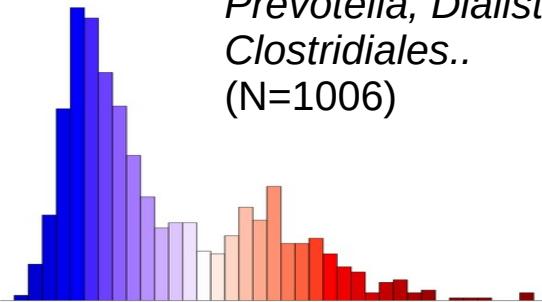


Kasvillisuus (V)



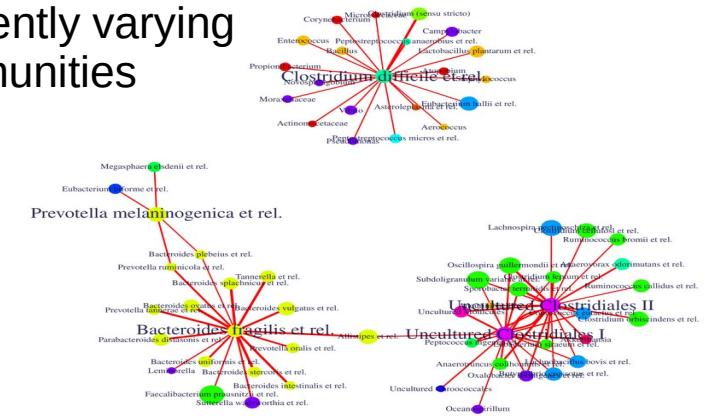
Subject stratification into community types could reduce complexity

Bistable taxa:
Prevotella, *Dialister*,
Clostridiales..
(N=1006)

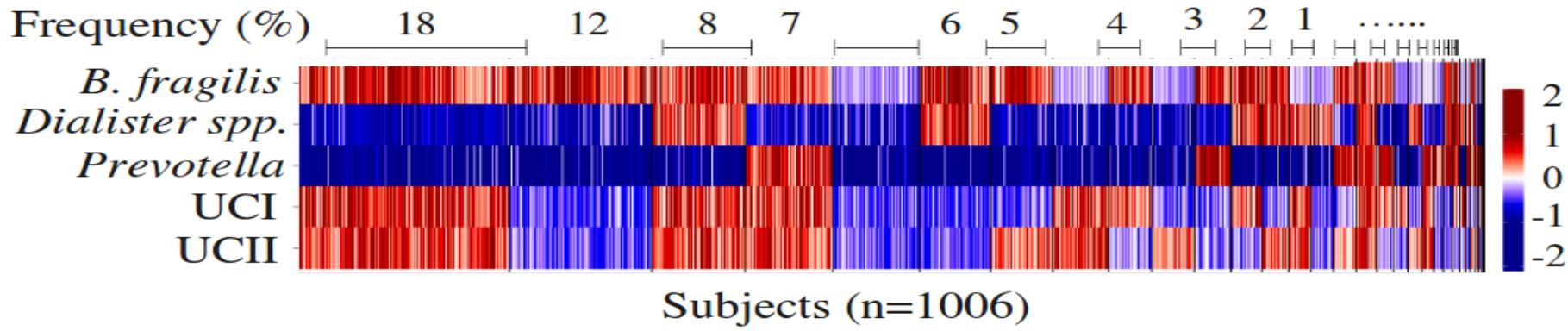


Ecosystem states are rich combinations
of independent tipping elements ?

Independently varying
sub-communities

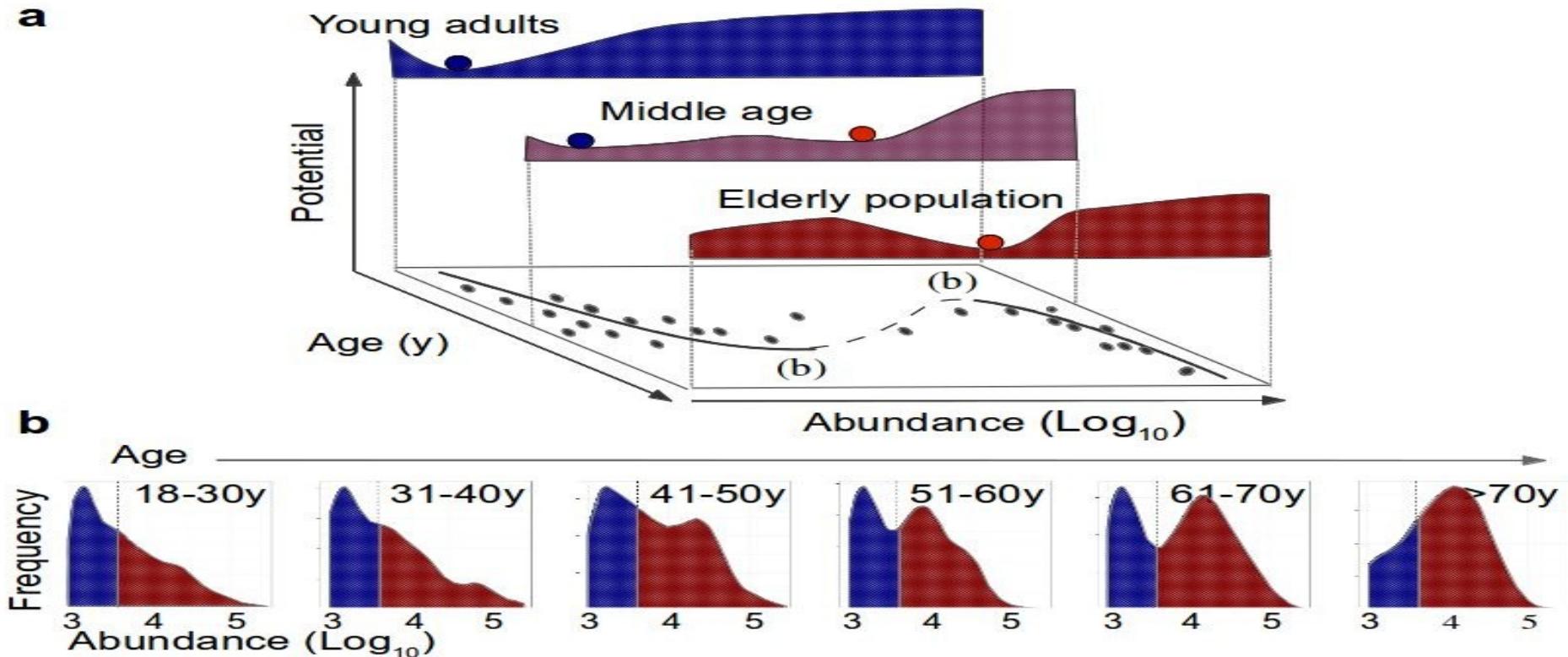


Lahti et al. Nat. Comm. 2014



Tipping elements of the human gut microbiota: Uncultured Clostridiales I

Lahti et al. Nat. Comm. 5:4344, 2014



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 population⁷⁷ driven slowly across a bifurcation.

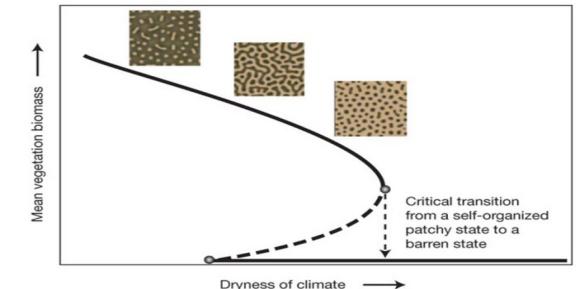
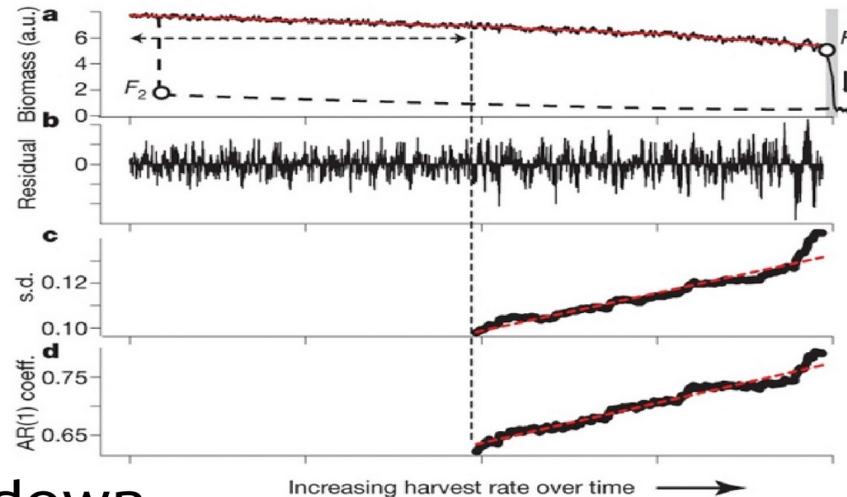
Autocorrelation

Variance

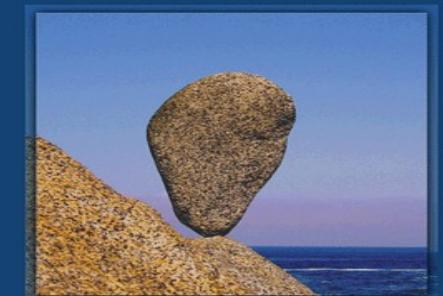
Skewness

Flickering

Critical slowing down



Critical Transitions
in Nature and Society



CRAN: earlywarnings R package

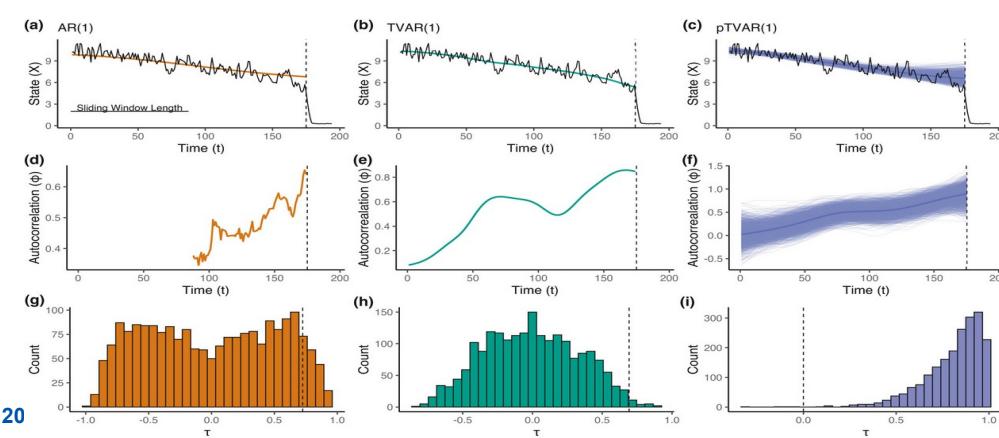
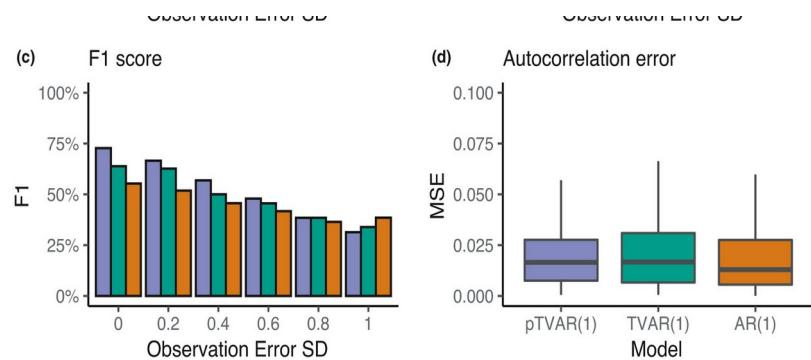
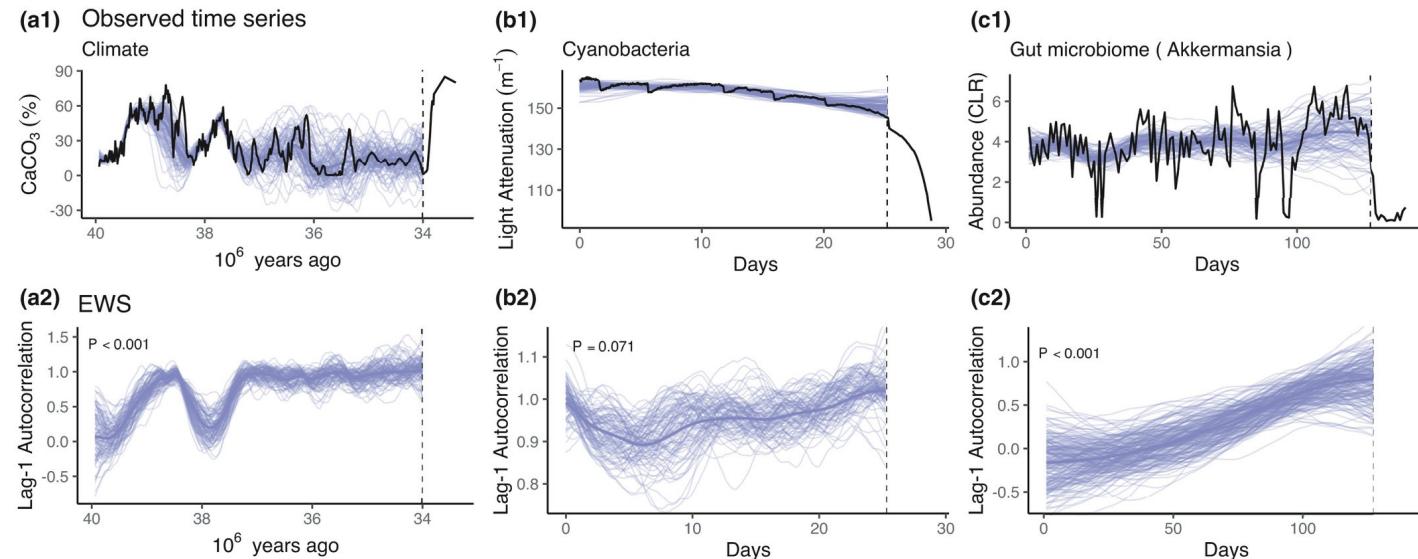
V. Dakos & L. Lahti



Probabilistic early warning signals

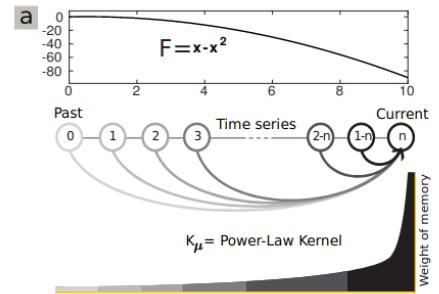
Ville Laitinen , Vasilis Dakos, Leo Lahti,

First published: 26 September 2021

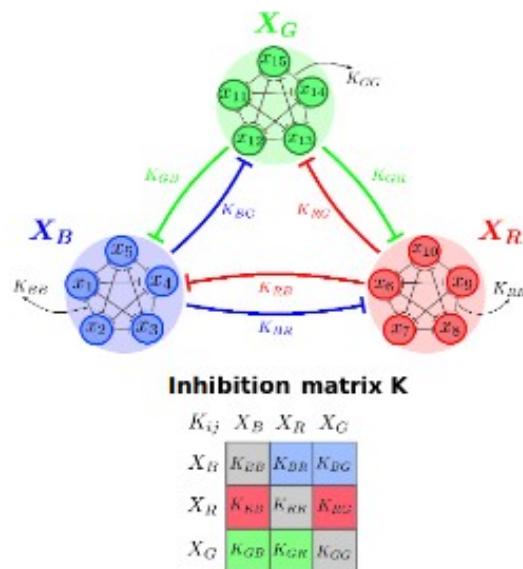
<https://doi.org/10.1002/ece3.8123>

Quantifying the impact of ecological memory on the dynamics of interacting communities

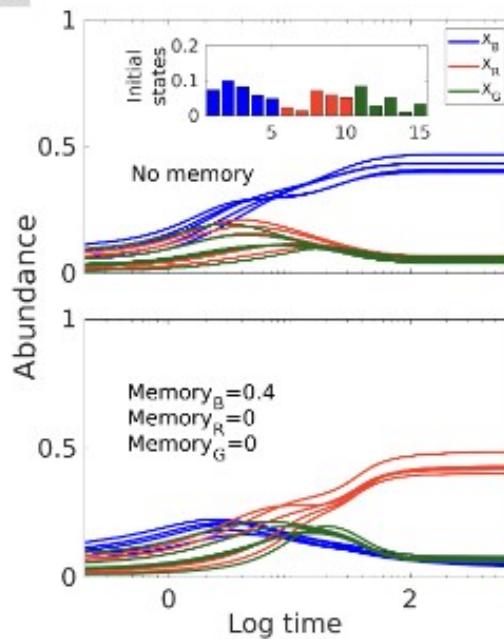
Moein Khalighi^{1*}, Guilhem Sommeria-Klein^{1¶}, Didier Gonze², Karoline Faust³, Leo Lahti^{1*}



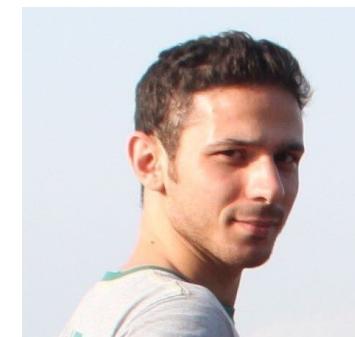
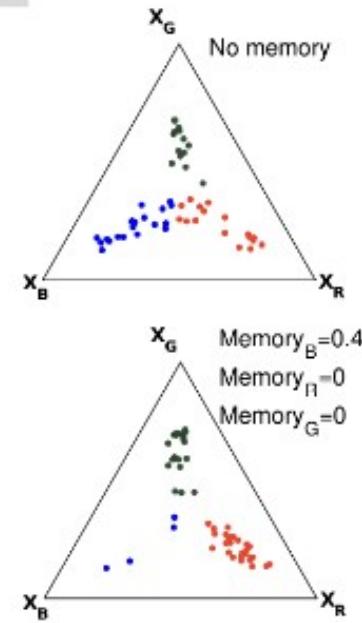
a
15 Species in 3 groups



b



c



d

1

Key sources of microbial ecosystem variation

External perturbations
(push & pulse)

Internal dynamics
and multi-stability

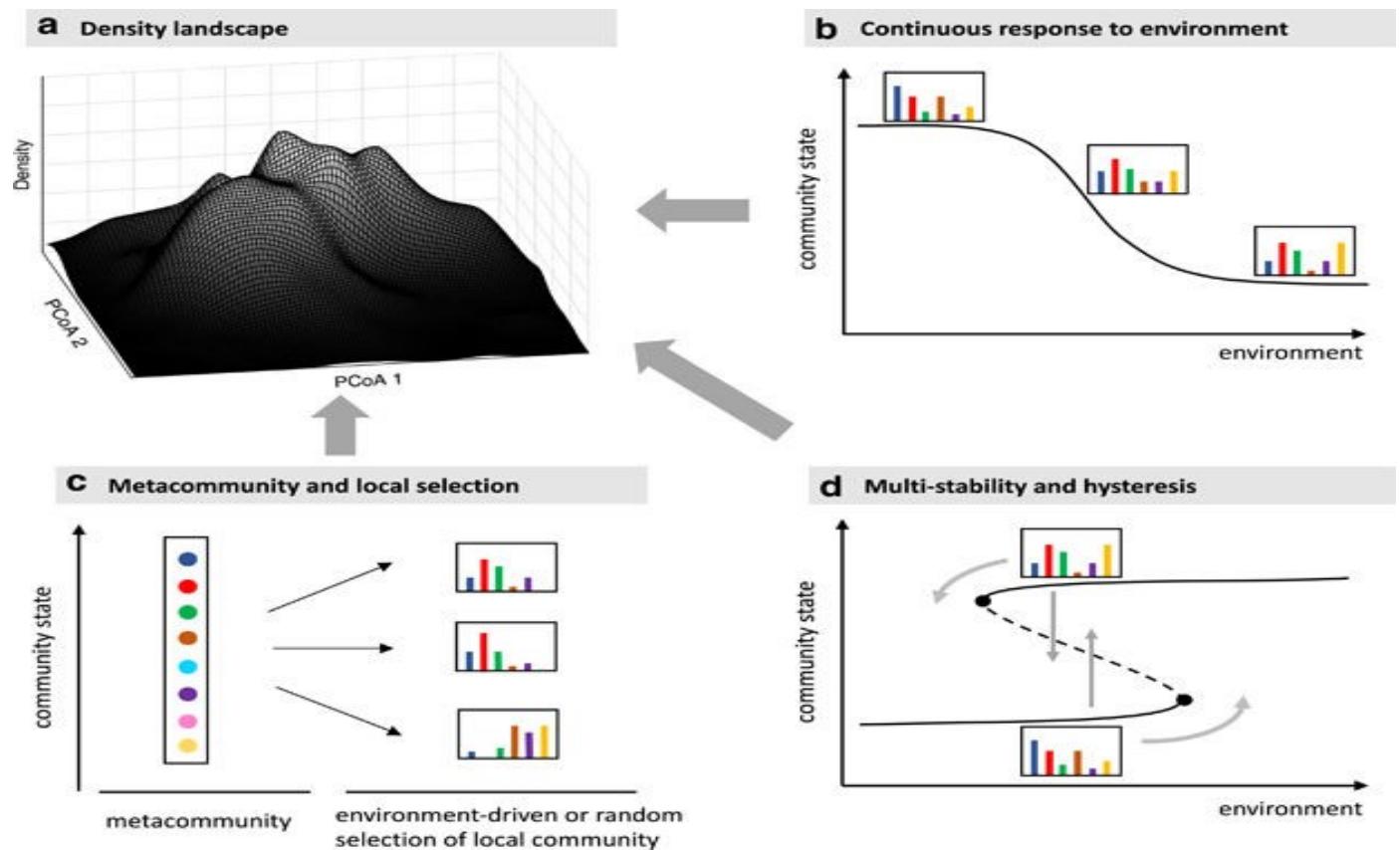
Immigration

Stochasticity

Memory

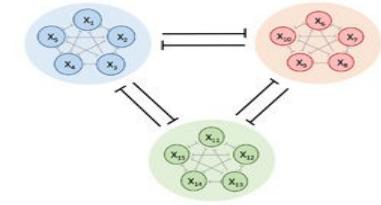
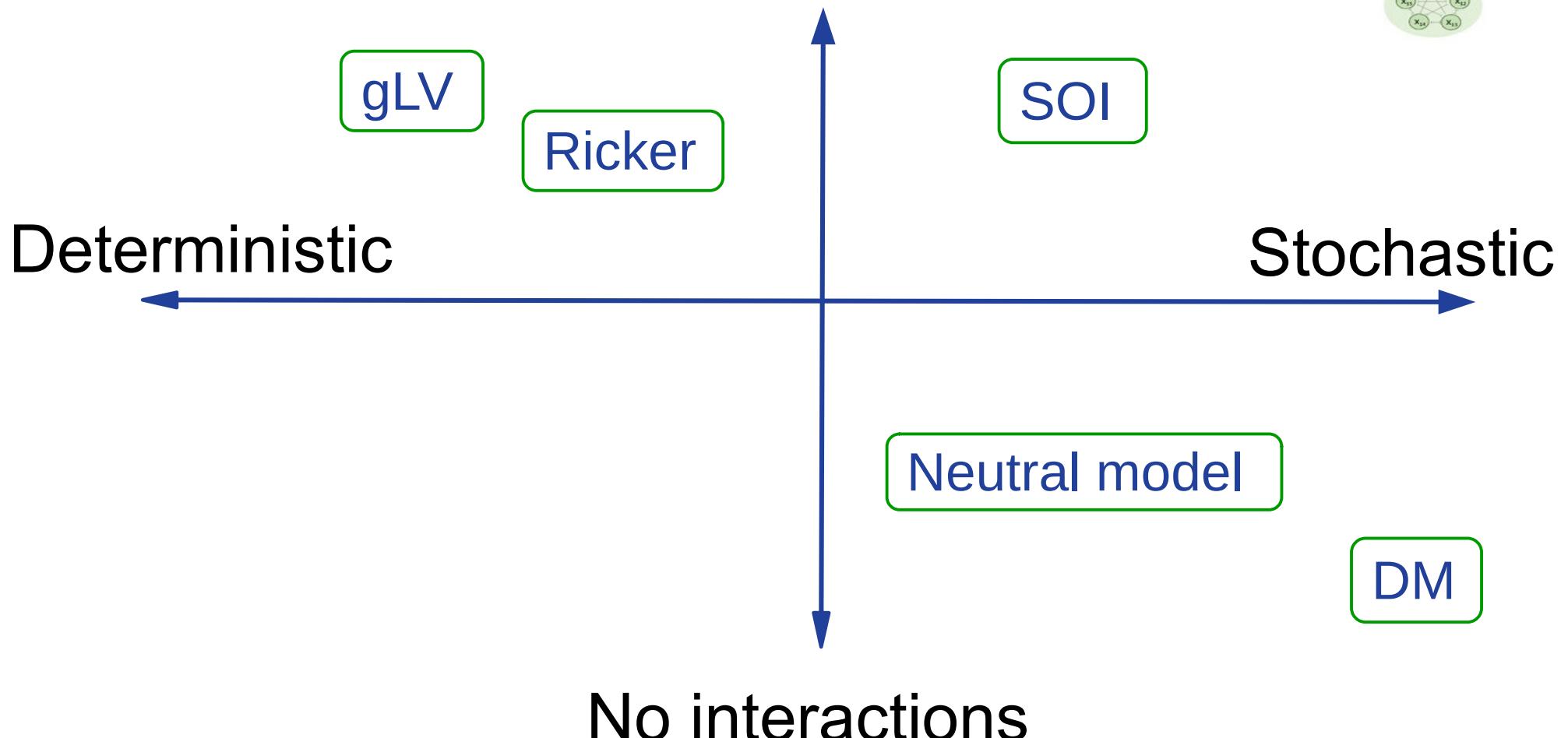
Microbial communities as dynamical systems

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



38 / 53

Species Interactions

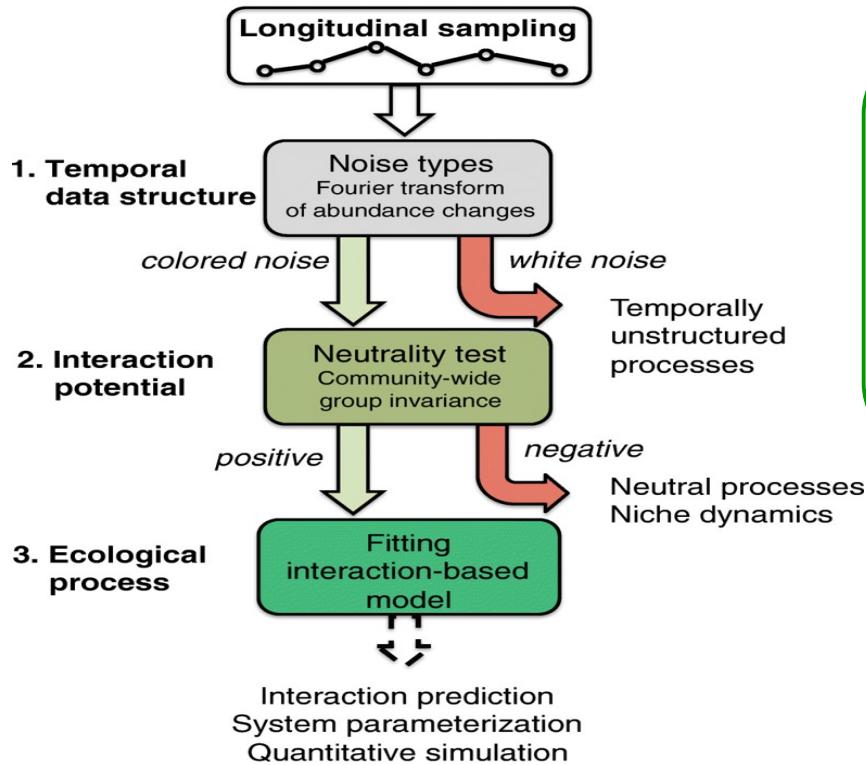


“Colors of noise” hint at ecological mechanisms



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*}

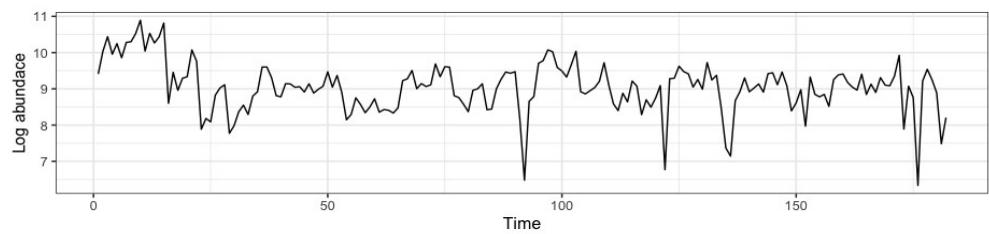


White noise → Unstructured

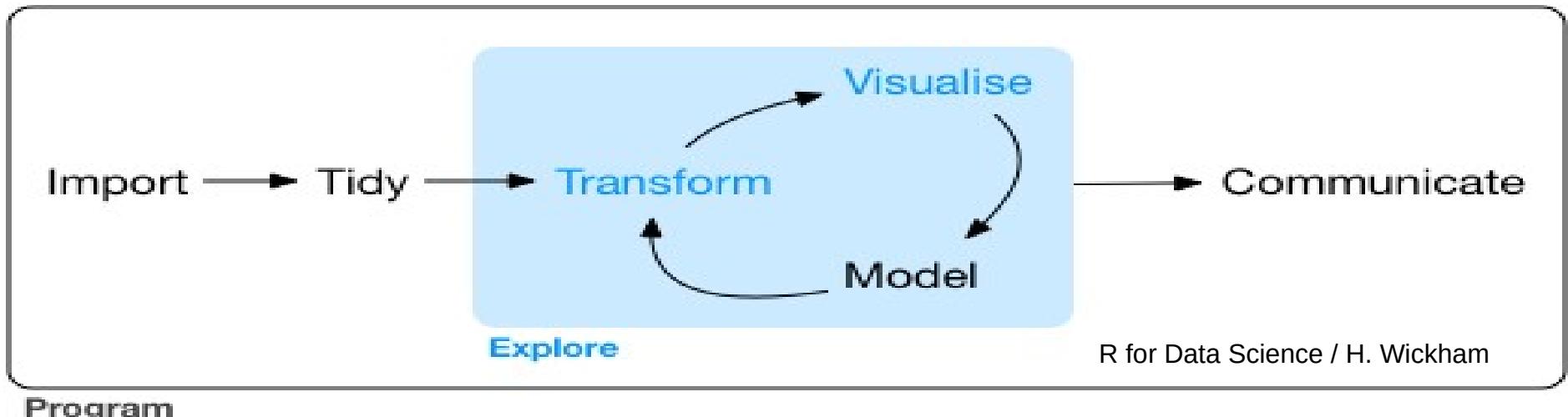
Pink noise → Time dependence

Brown noise → Neutrality

Black noise → Interactions



Data science workflow



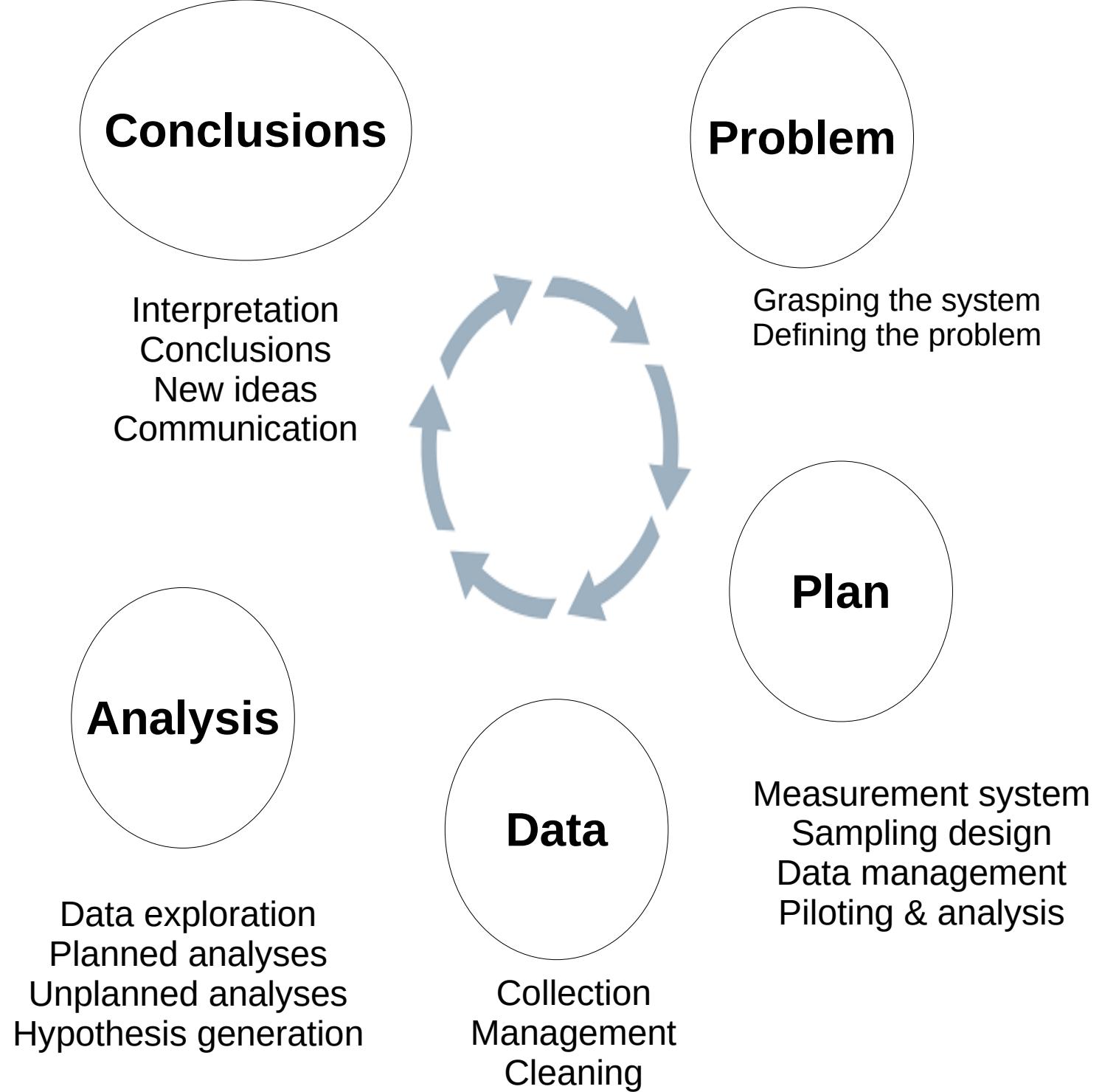
REVISED Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses [version 2; peer review: 3 approved]

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



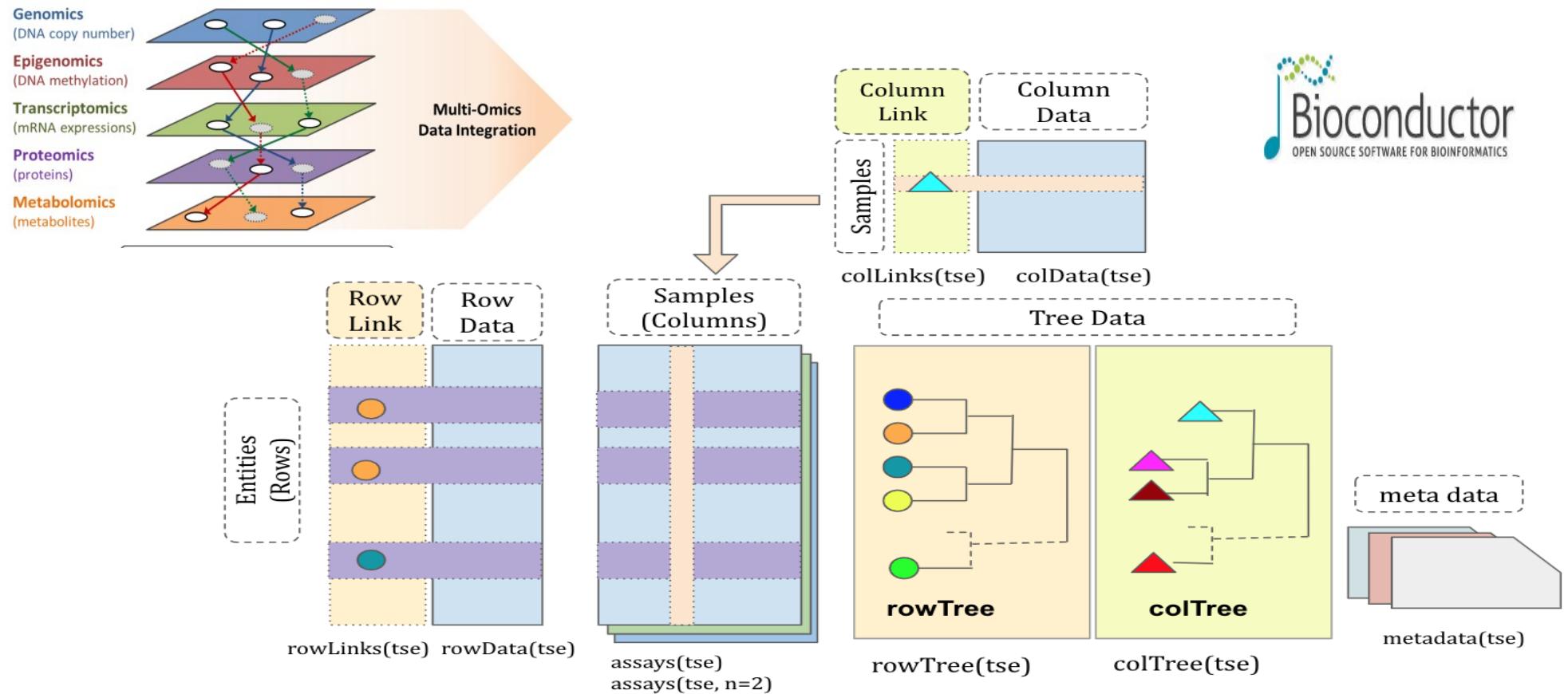
This article is included in the [Bioconductor](#) gateway.

PPDAC cycle



TreeSummarizedExperiment data container

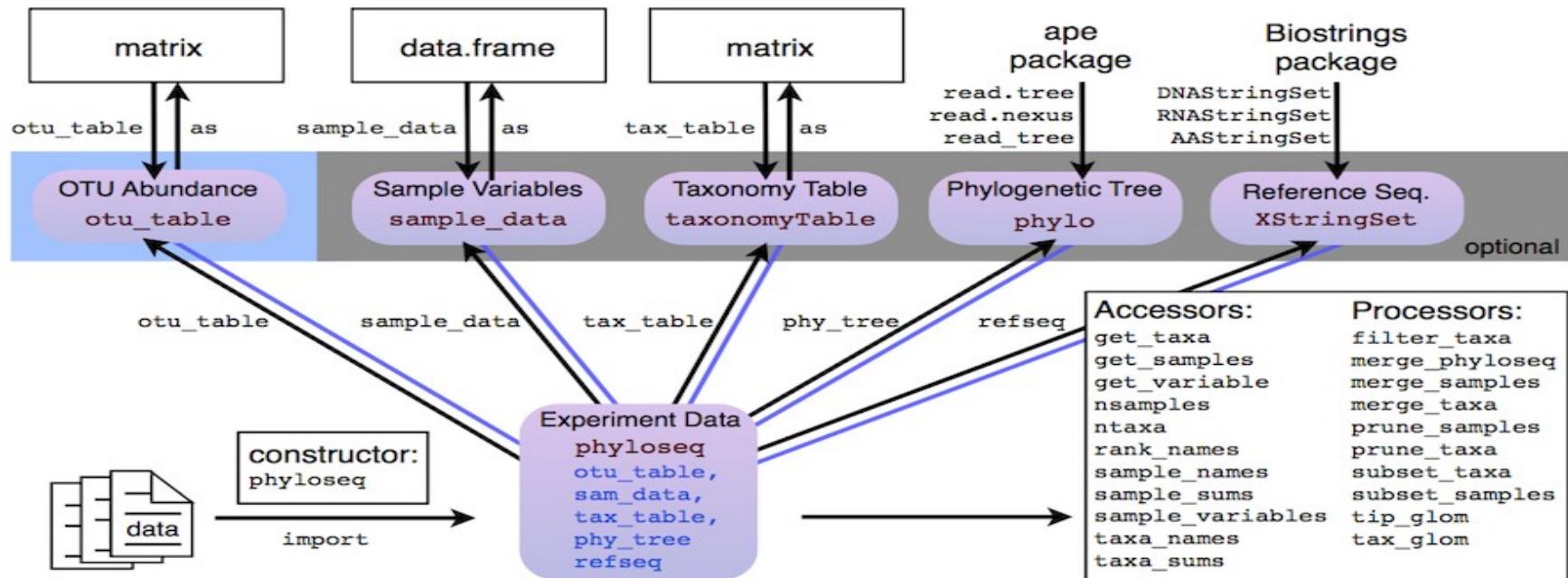
by Ruizhu @fiona Huang; initially proposed for microbiome research by Hector Bravo & Domenick Braccia



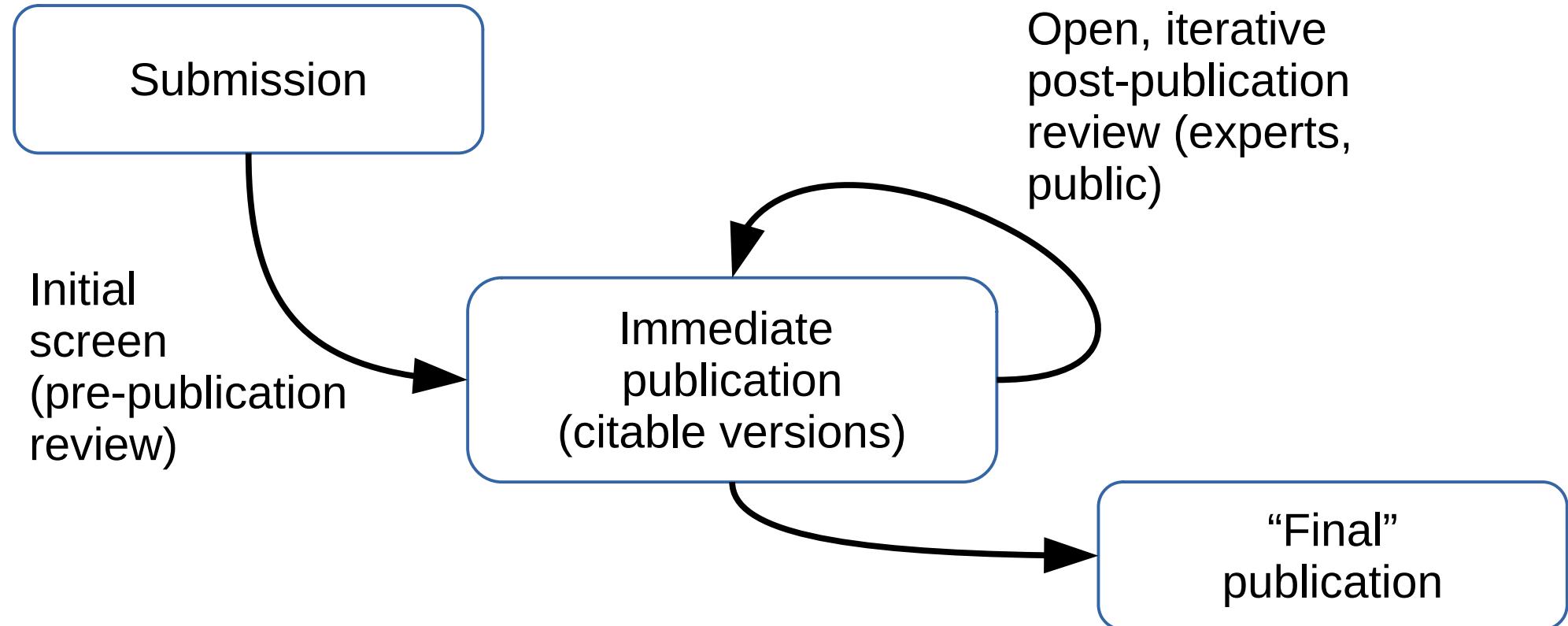
Seamless conversion from *phyloseq* & other raw data types

Alternative data container: *phyloseq*

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



Transparency: pre- vs. post-publication review

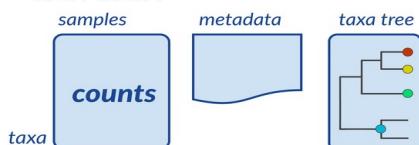


Example workflow

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

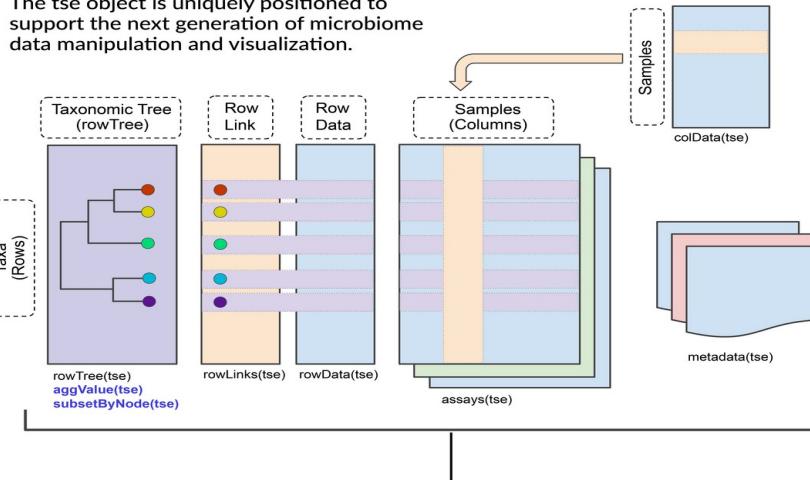


EXISTING DATA



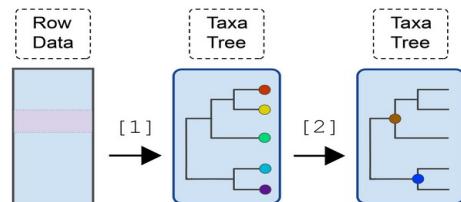
The TreeSE object

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



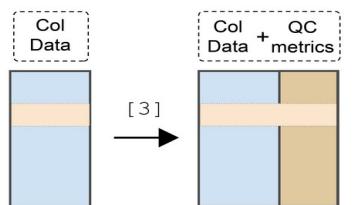
The mia Pipeline

Accessing Taxonomic Info.



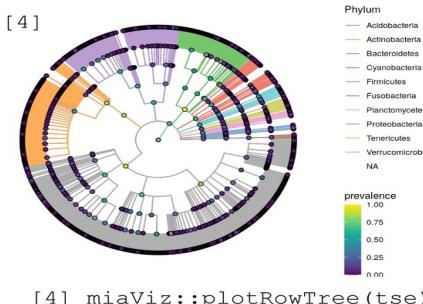
```
[1] mia::addTaxonomyTree(tse)  
[2] TreeSE::aggValue(tse)
```

Quality Control



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

Visualizing with miaViz

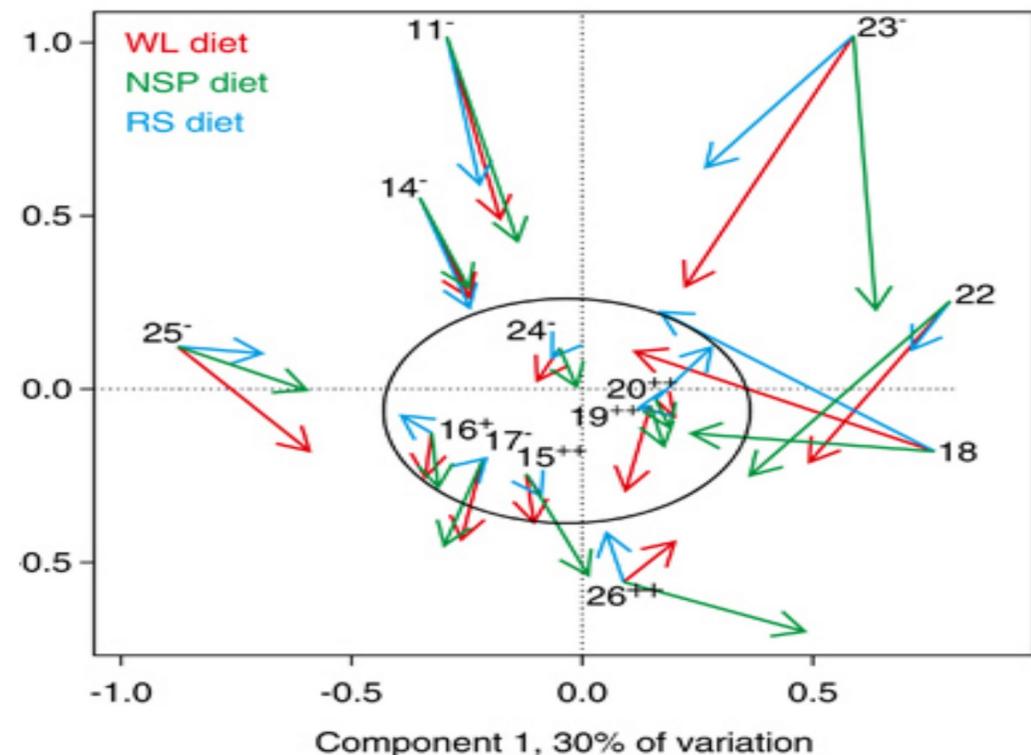
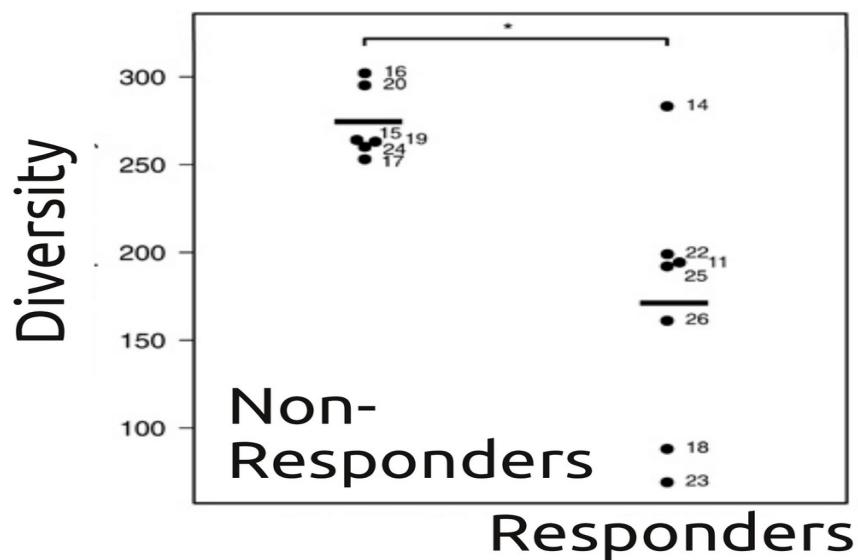


Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men

Anne Salonen, Leo Lahti, Jarkko Salojärvi, Grietje Holtrop, Katri Korpela, Sylvia H Duncan, Priya Date, Freda Farquharson, Alexandra M Johnstone, Gerald E Lobley, Petra Louis, Harry J Flint and Willem M de Vos

Lahti et al. Nat. Comm. 5:4344, 2014

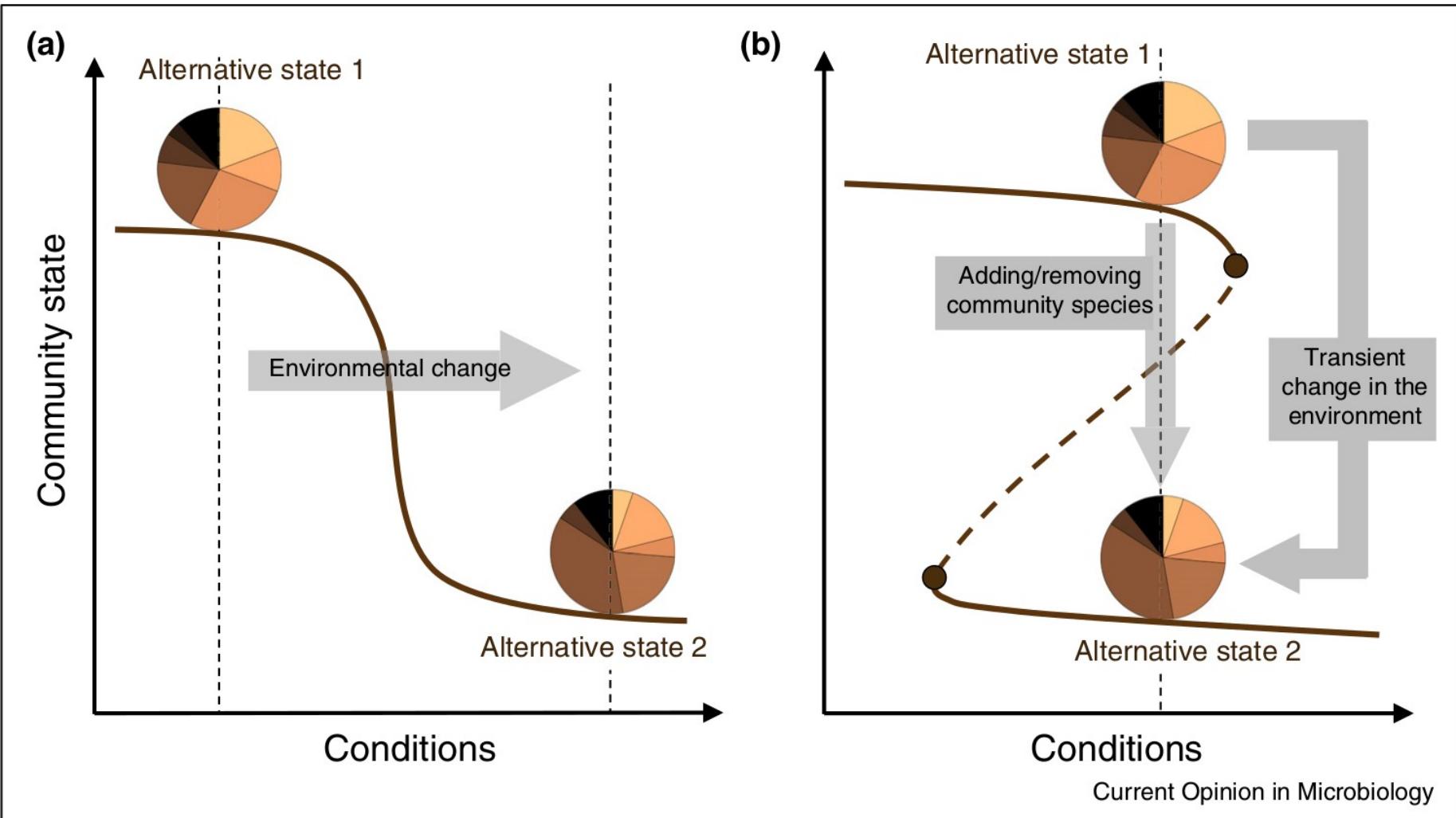
Korpela et al. PLoS ONE 9(3): e90702, 2014



State induced by external factors

Intrinsic stability: robust to external factors

Diversity ?

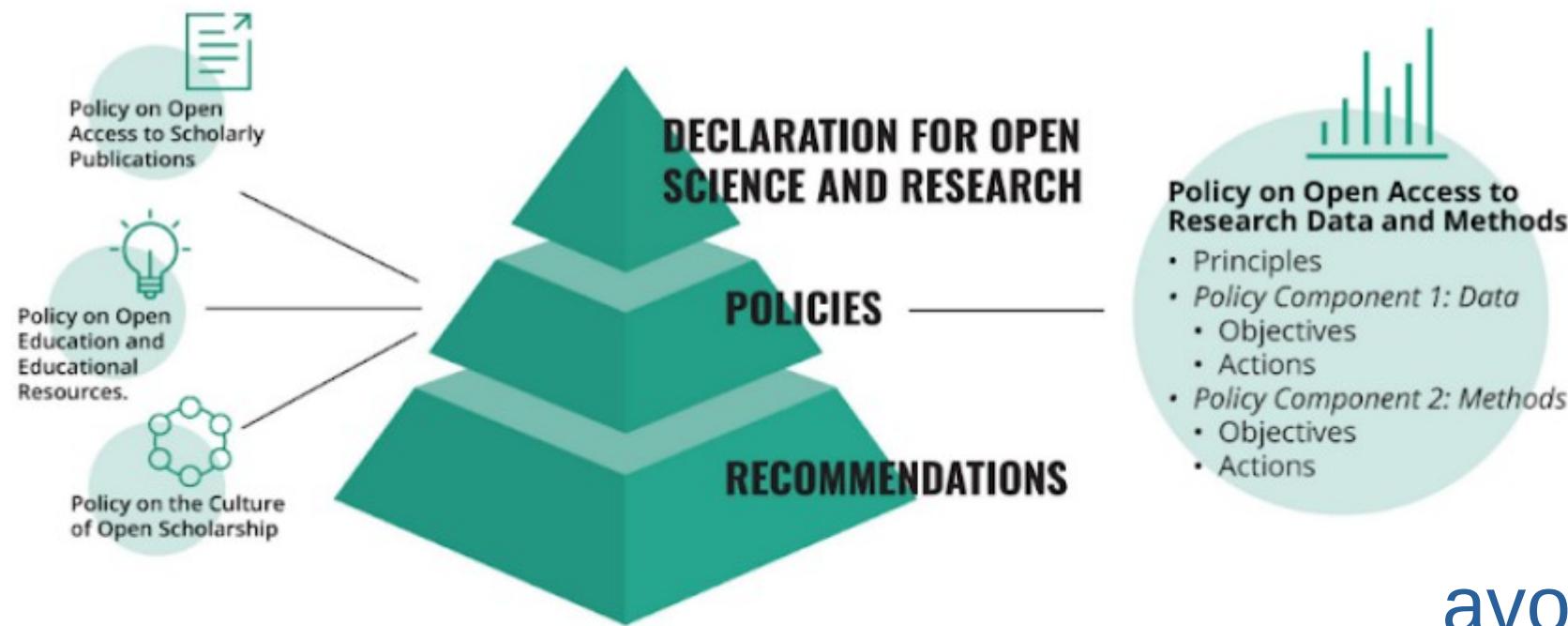


Environmental or host factor (e.g. body-mass index)



Picture 1. The policy in relation to other national open science documents.

National Policy on Open Science National policy and executive plan by the higher education and research community for 2021-2025



avointiede.fi

Reusability

PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS

EDUCATION

A Quick Guide to Software Licensing for the Scientist-Programmer

Andrew Morin, Jennifer Urban, Piotr Sliz 

Published: July 26, 2012 • <https://doi.org/10.1371/journal.pcbi.1002598>



Software citation principles

Arfon M. Smith^{1,*}, Daniel S. Katz^{2,*}, Kyle E. Niemeyer^{3,*}
FORCE11 Software Citation Working Group

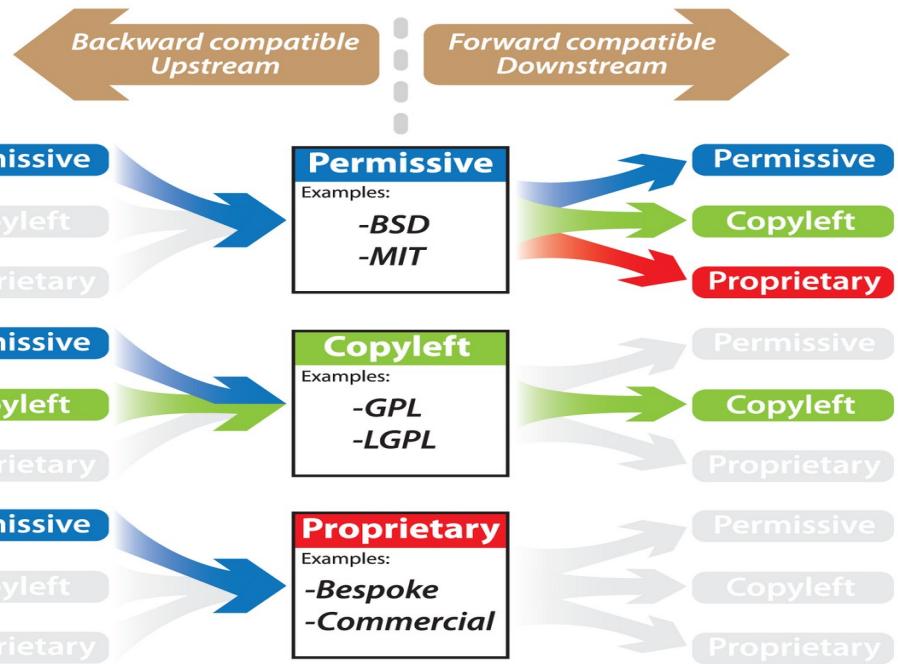
¹ GitHub, Inc., San Francisco, California, United States

² National Center for Supercomputing Applications & Electrical and Computer Department & School of Information Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States

³ School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State Corvallis, Oregon, United States

* These authors contributed equally to this work.

MIT License



Copyright (c) <year> <copyright holders>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Learning goals

reproducible workflows

advanced R/Bioconductor tools

multi-omics methods

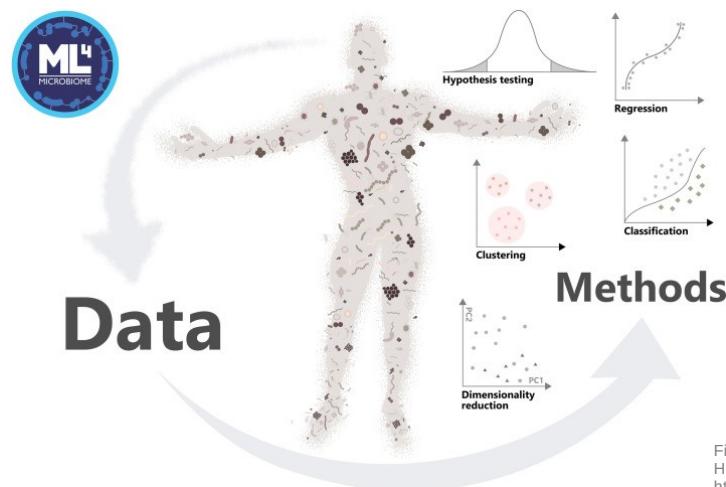


Figure source: Moreno-Indias et al. (2021) Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. URL: <https://doi.org/10.3389/fmicb.2021.635781>. Frontiers in Microbiology 12:11.

Support & contributions

Online support:

- Gitter, Slack
- Email

Join R/Bioconductor developer community:

- Github issues & pull requests (package github page)
- R/Bioc Slack

microbiome.github.io

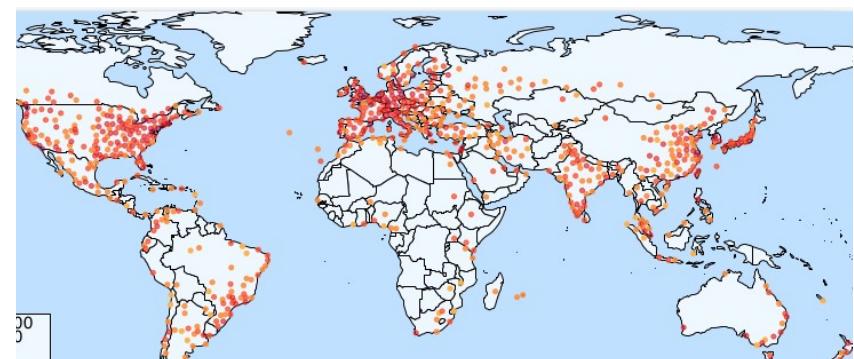
R/Bioc community

(Global) Bioconductor conference

Asian Bioconductor conference

European Bioconductor conference

- *Heidelberg, Sep 14-16, 2022*



A screenshot of the EuroBioC 2022 conference website. At the top, there's a navigation bar with links for HOME, SUBMISSIONS, REGISTRATION, SCHEDULE, SPONSORS, and ABOUT. The main title "EUROBIOC 2022 CONFERENCE, HEIDELBERG, BIOQUANT, SEPTEMBER 14-16, 2022." is prominently displayed in large white text. Below it, the tagline "WHERE SOFTWARE AND BIOLOGY CONNECT" is visible. In the bottom right corner, there's a decorative hexagonal graphic featuring a landscape with a castle and a musical note, with the text "EuroBioC2022 Heidelberg" and the website "www.bioconductor.org". The background of the page has a repeating hexagonal pattern with text from previous conferences like "BioC Asia 2016 Brisbane", "BioC Asia 2017 Adelaide", and "BioC Asia 2018 Melbourne".