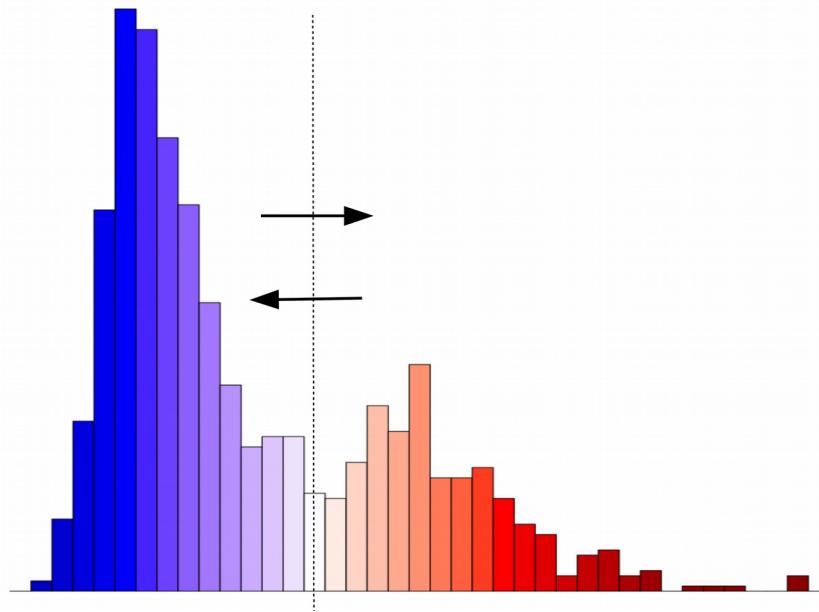


Microbiome bioinformatics: challenges and solutions



Turun yliopisto
University of Turku

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

An Aura of Our Microbes Surrounds Us...

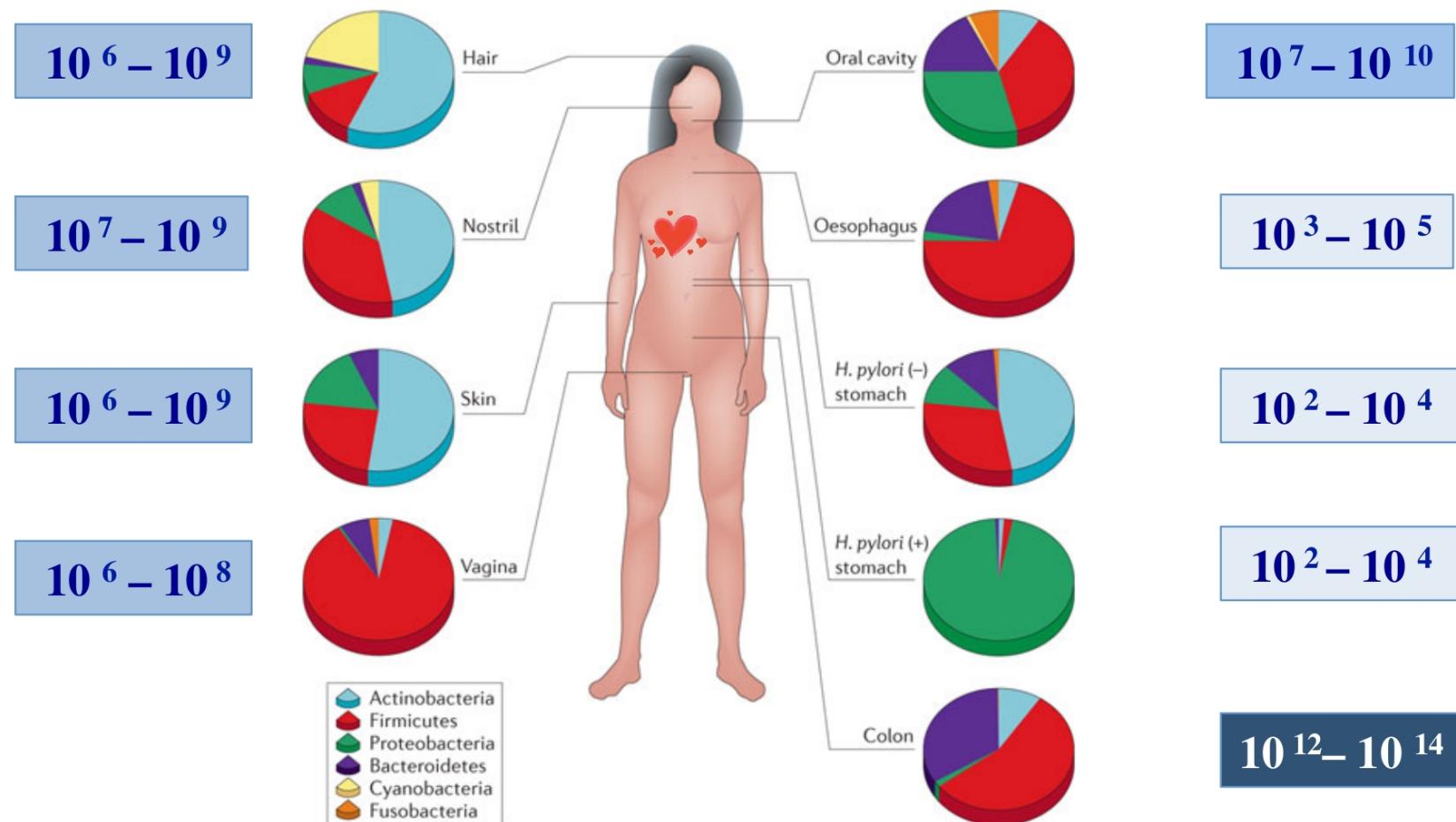
Everything Is Everywhere But The Environment Selects - *Baas-Bekking & Beijerinck 1934*



Human Emit Over 10 Million Biological Particles per Hour - Personal Cloud
From Cradle to Grave

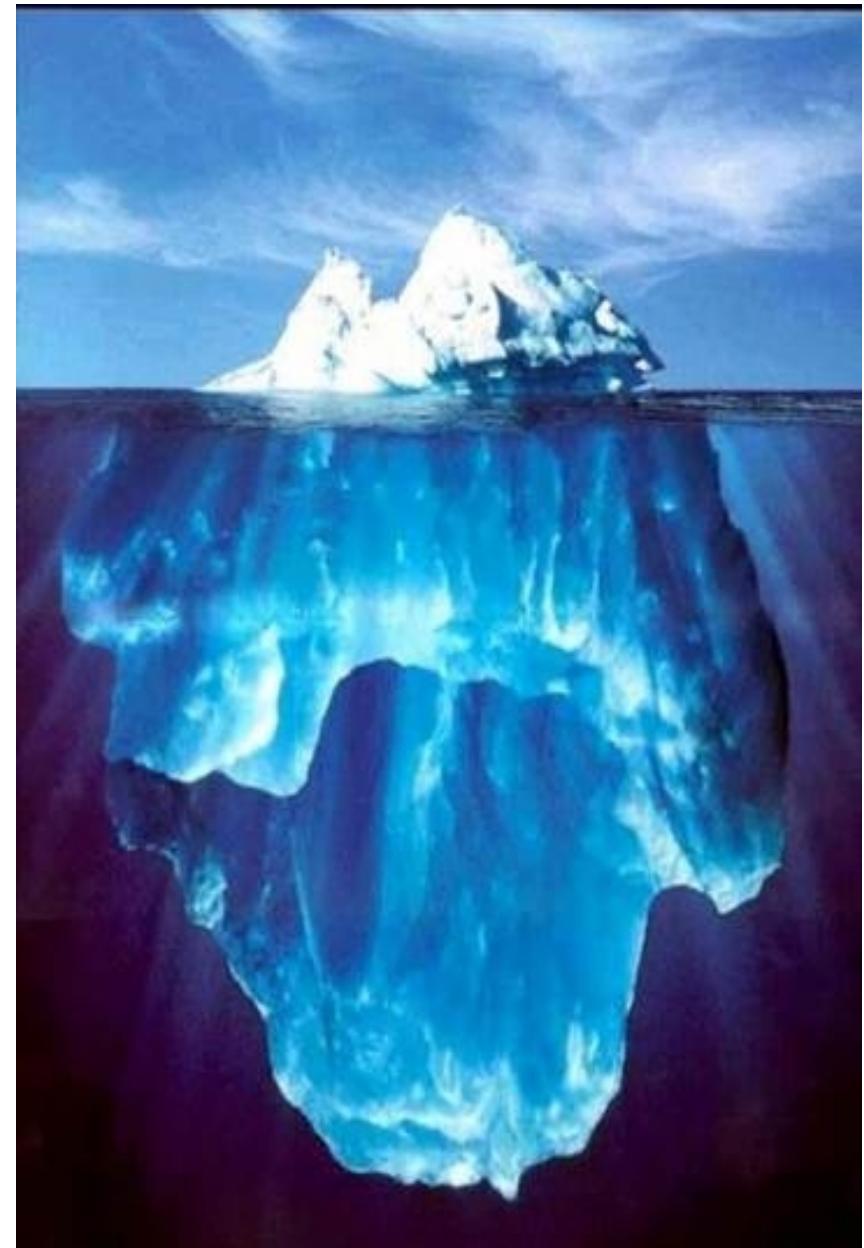
Meadows et al Peer J 2015 - Metcalf et al Science 2015

...And We Are Colonized By Many More !

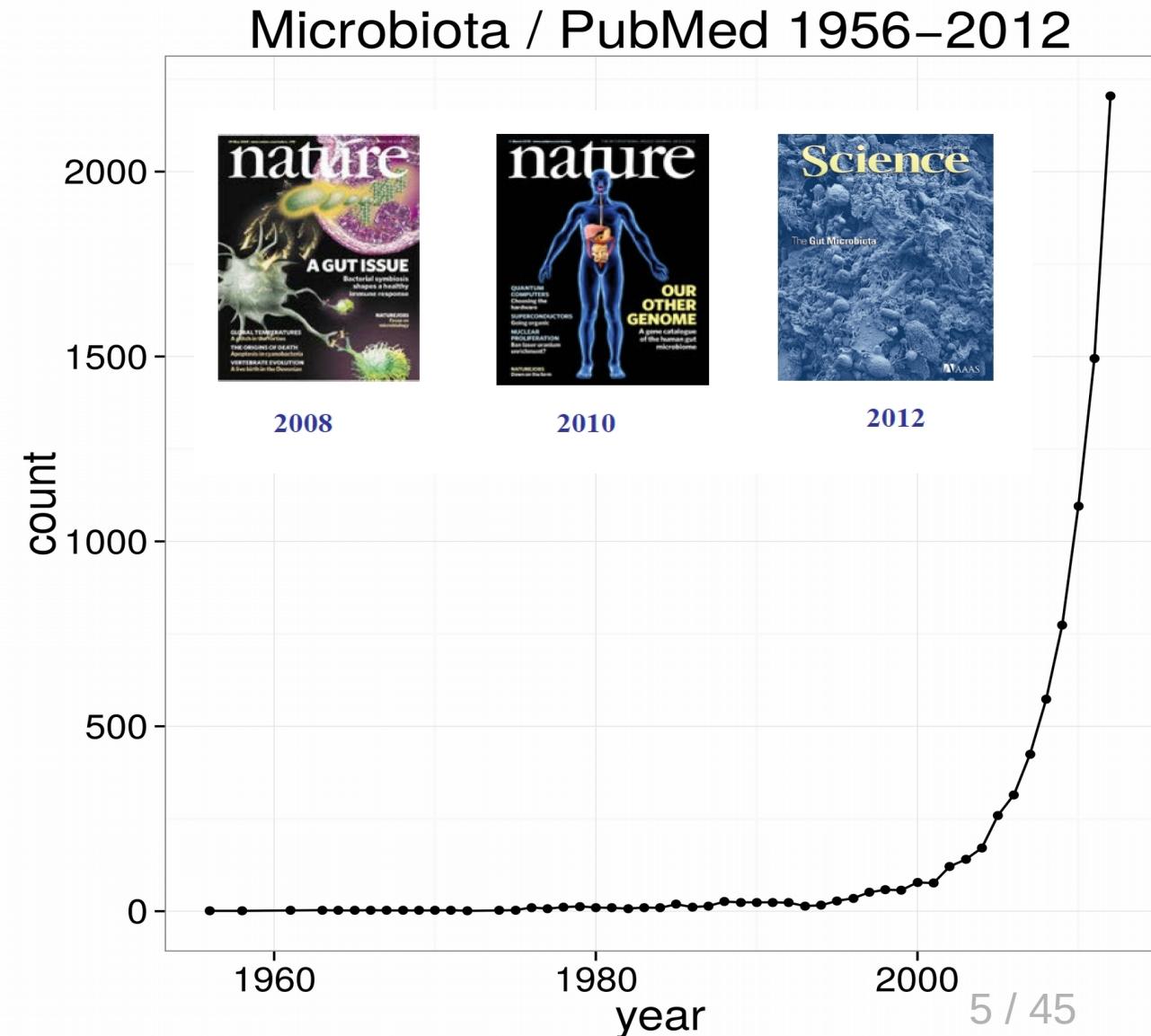


Nature Reviews | Genetics

Cho & Blaser Nature Rev Genetics 2012
Douillard & De Vos Micr Cell Fact 2014



Research on human microbiota increasing at an accelerating pace



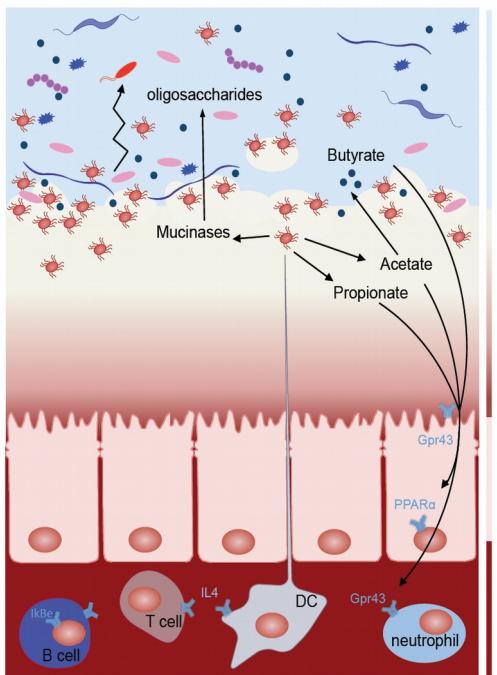
Outline

- Human microbiome research
- Challenges in statistical ecology
- Open data science



Gut microbiome: 300-400 m² surface area

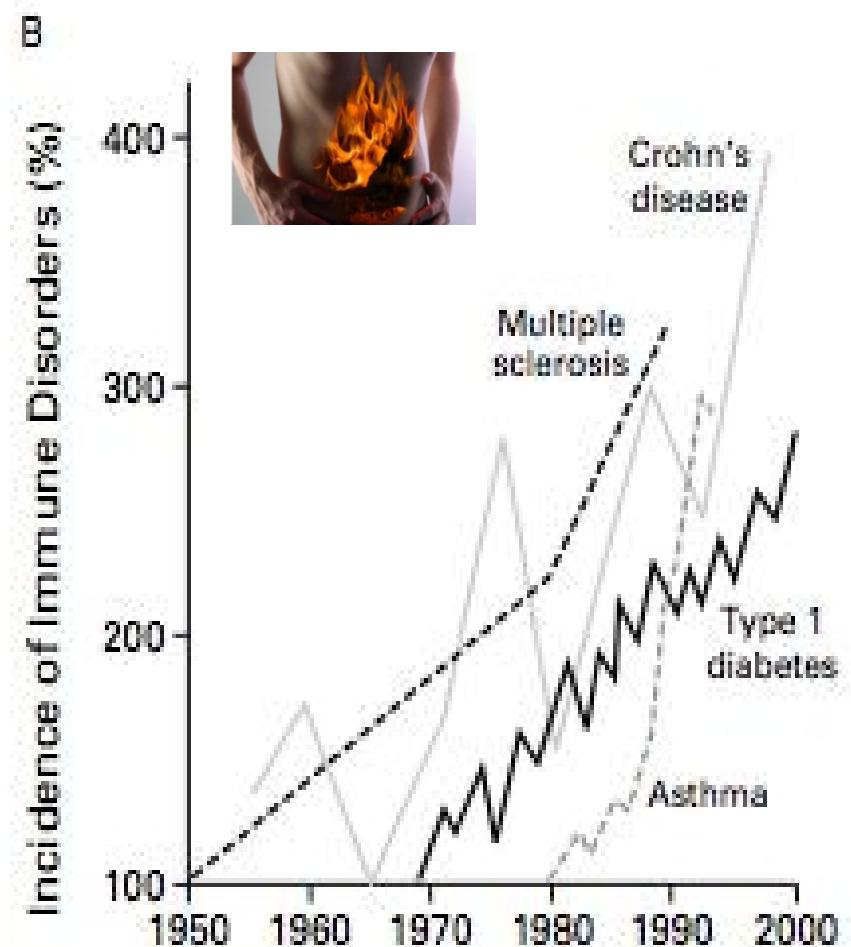
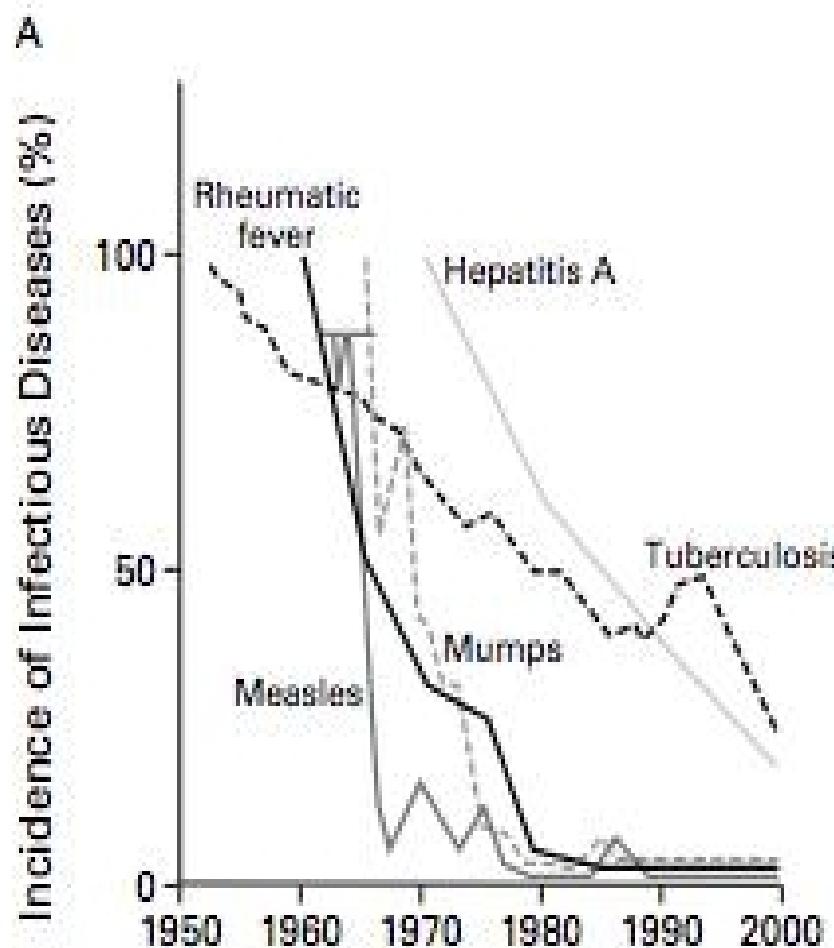
- ~ bacterial cells outnumber our own
- ~ 100 billion bacteria / gram in the gut
- ~ 10,000 gut bacterial species
- ~ 10M uniq genes (500x human genome)
- ~ Plastic, varies in time



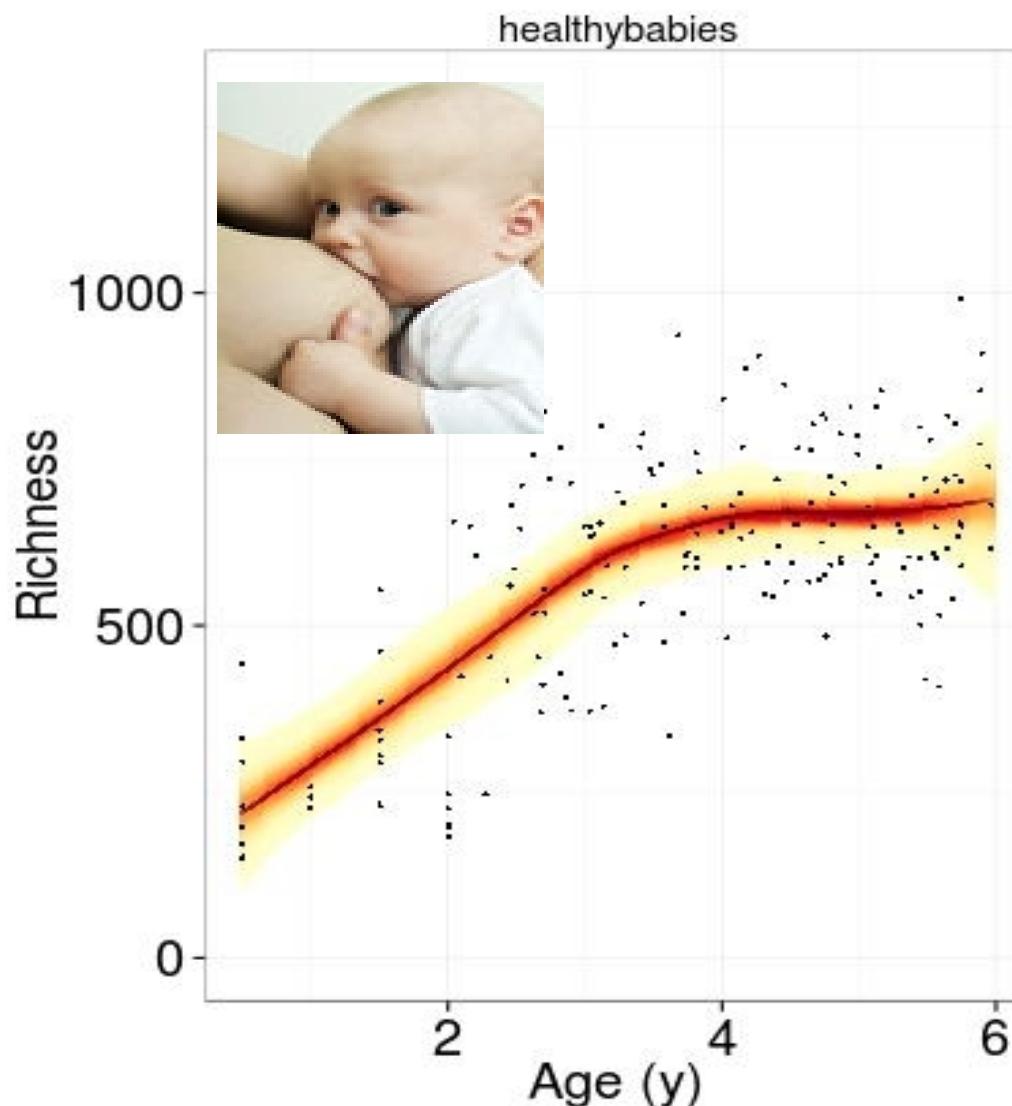
Brussel central square flower carpet



Many diseases linked to gut microbiota have become more prevalent in the past decades !



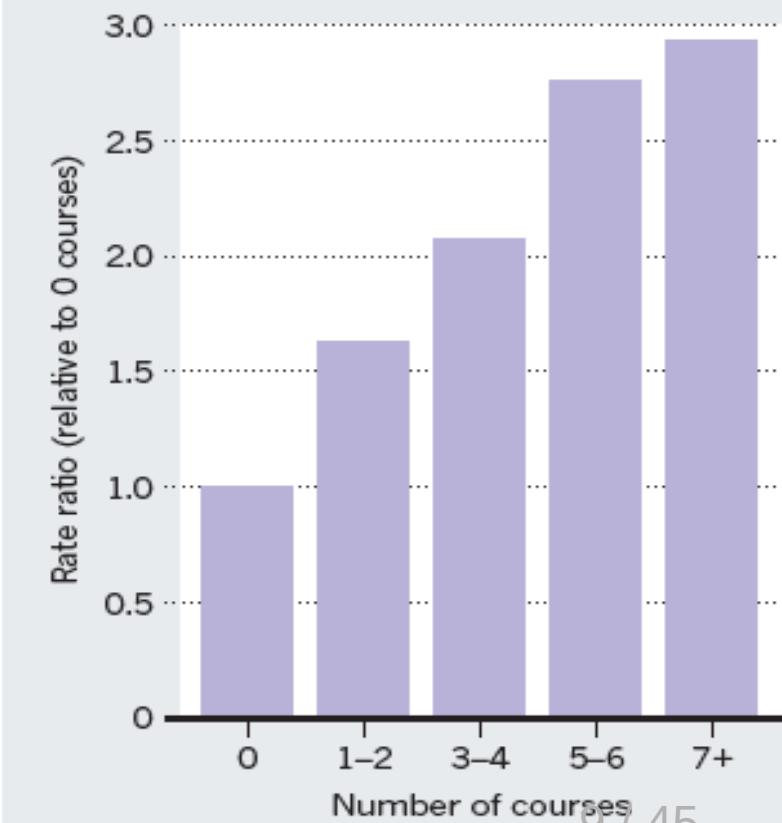
Microbiota develops in early life



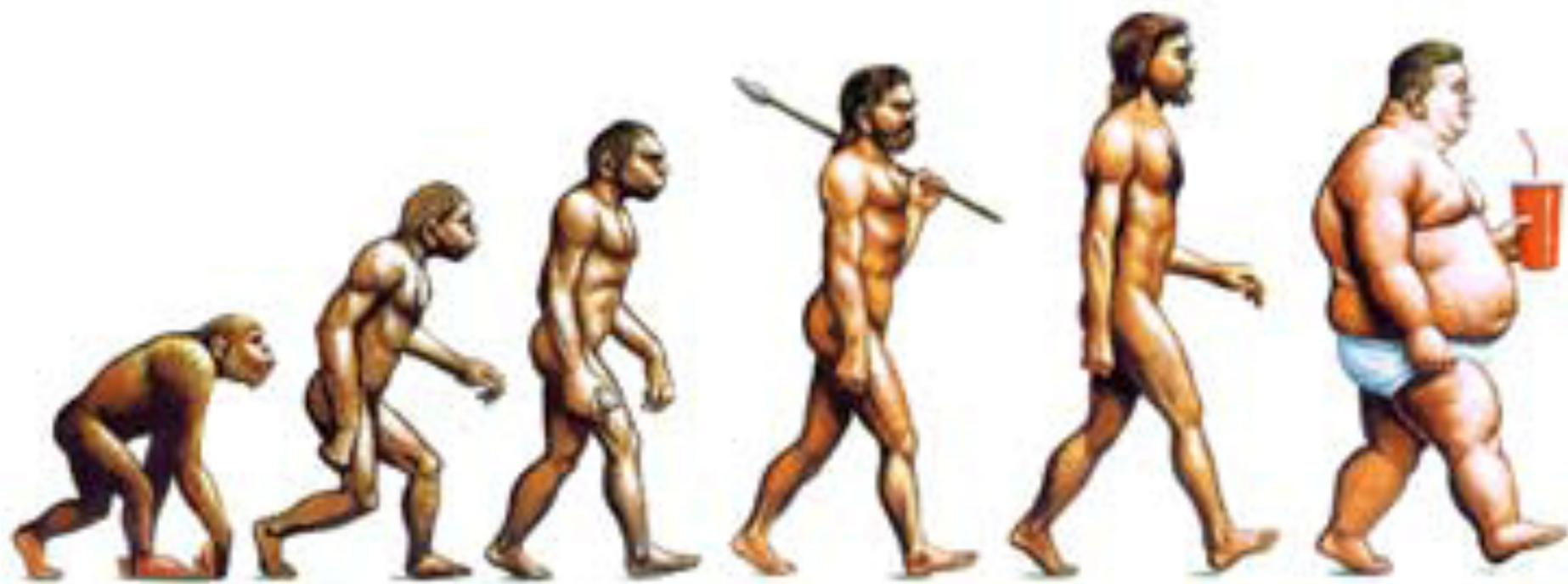
Frequent antibiotics in childhood increases the risk of intestinal disorders

TROUBLING CORRELATION

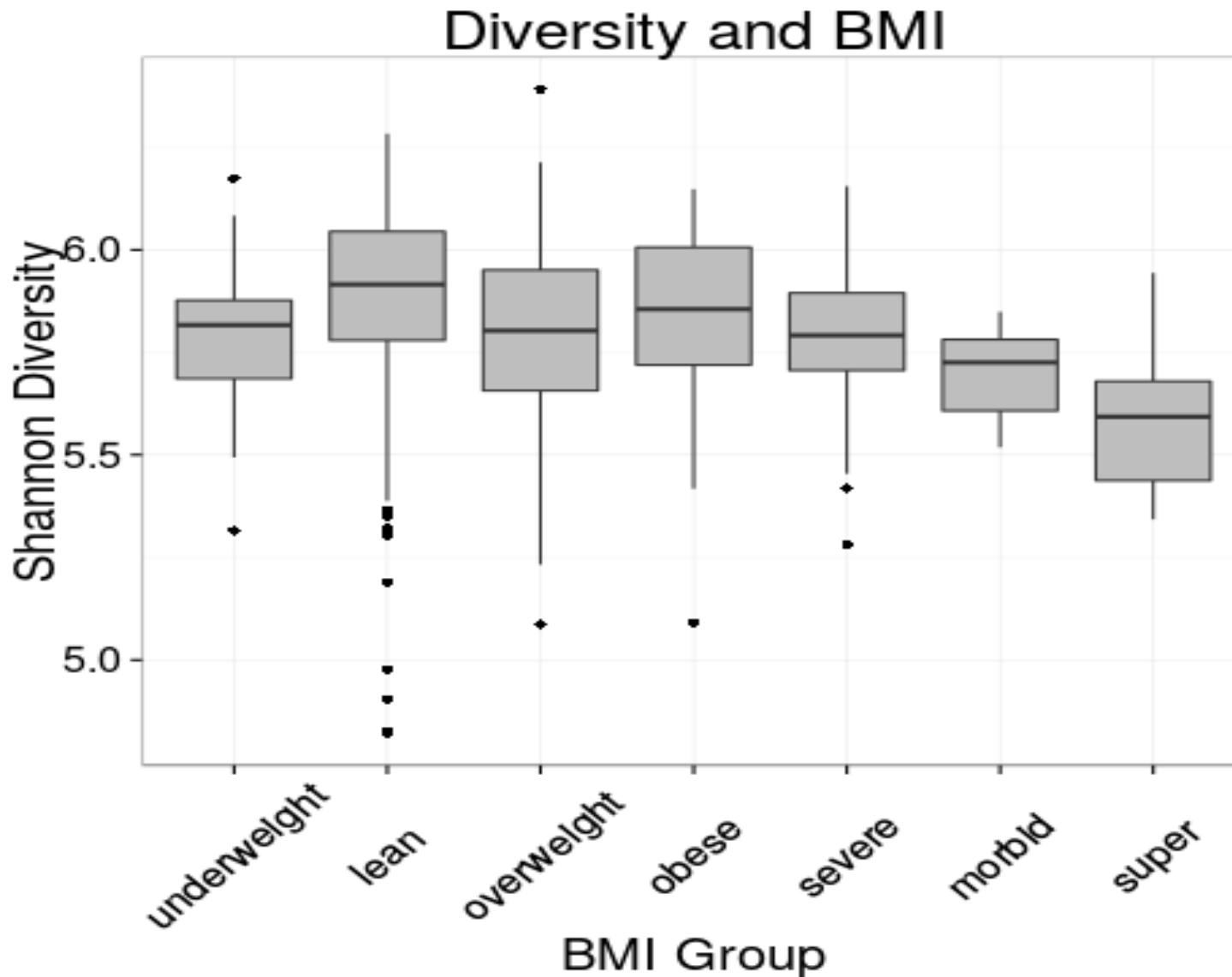
The risk of inflammatory bowel diseases in children rises with the number of courses of antibiotics taken.



Million years of human evolution

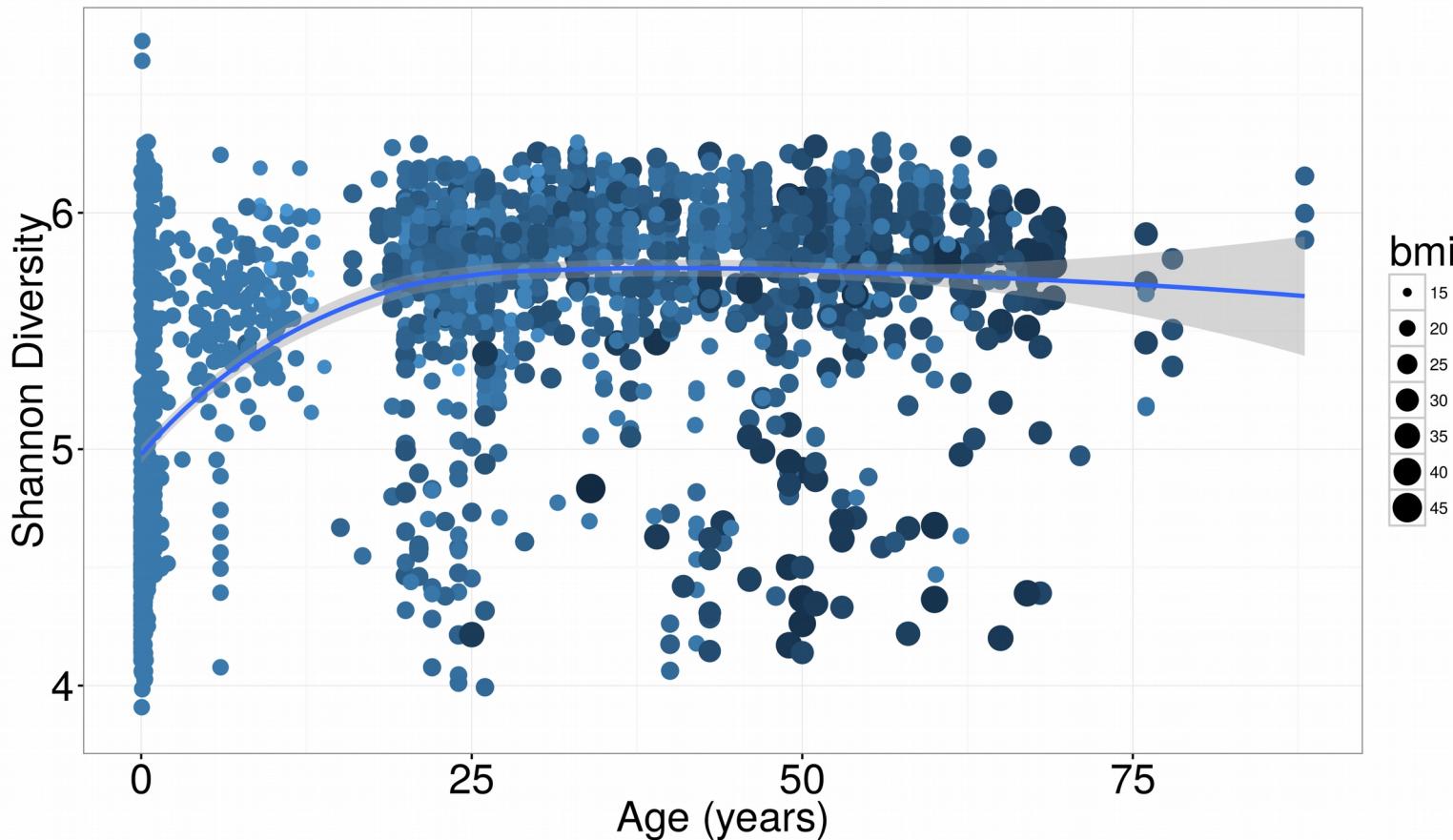


Microbial diversity decreases with obesity



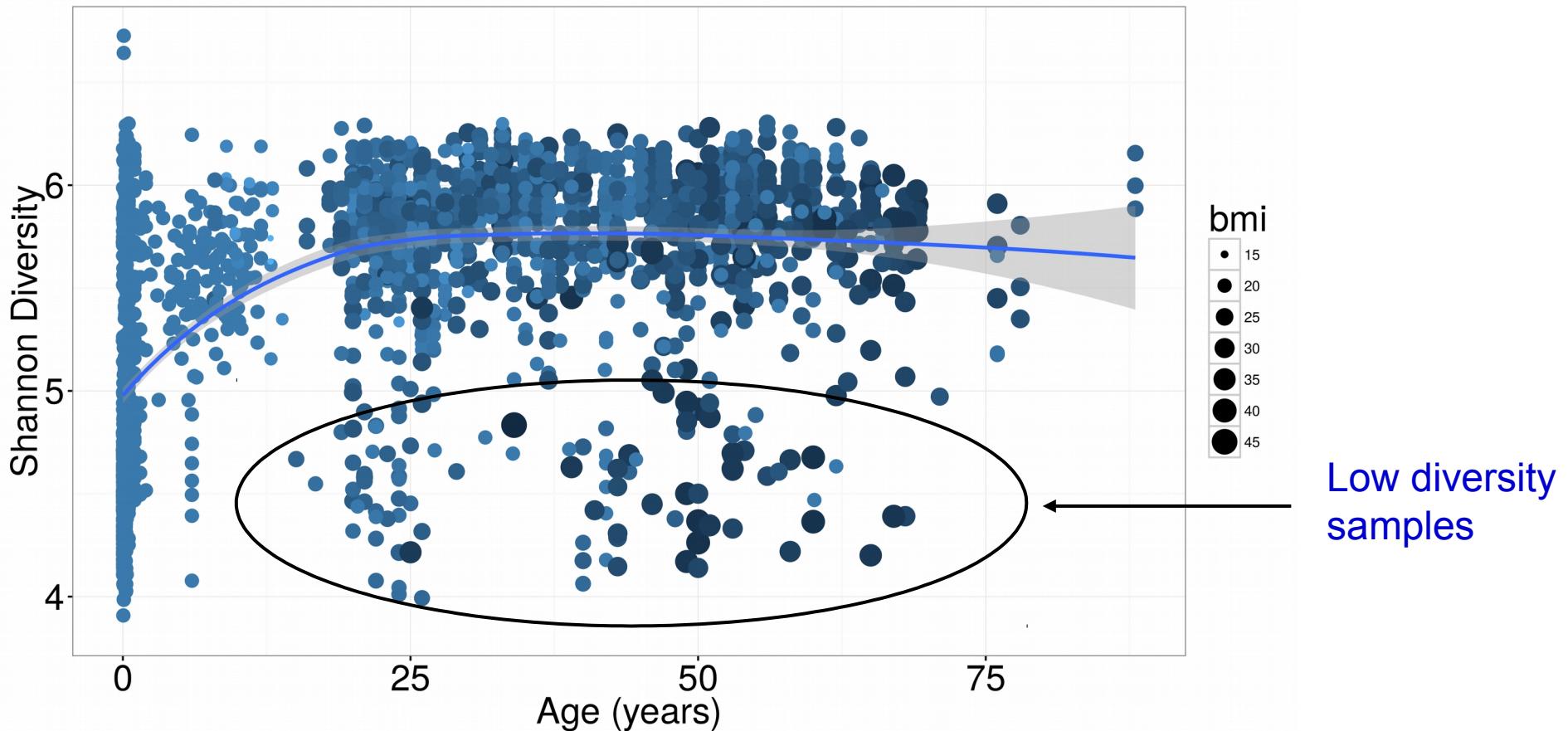
Microbiome diversity and age: healthy & normal obese subjects

N = 2363



Microbiome diversity and age: healthy & normal obese subjects

N = 2363



The promise of the microbiome field: diagnostics & cure for all major diseases on the planet





A scanning electron micrograph of bacteria in human faeces, in which 50% of species originate from the gut.

Microbiome science needs a healthy dose of scepticism

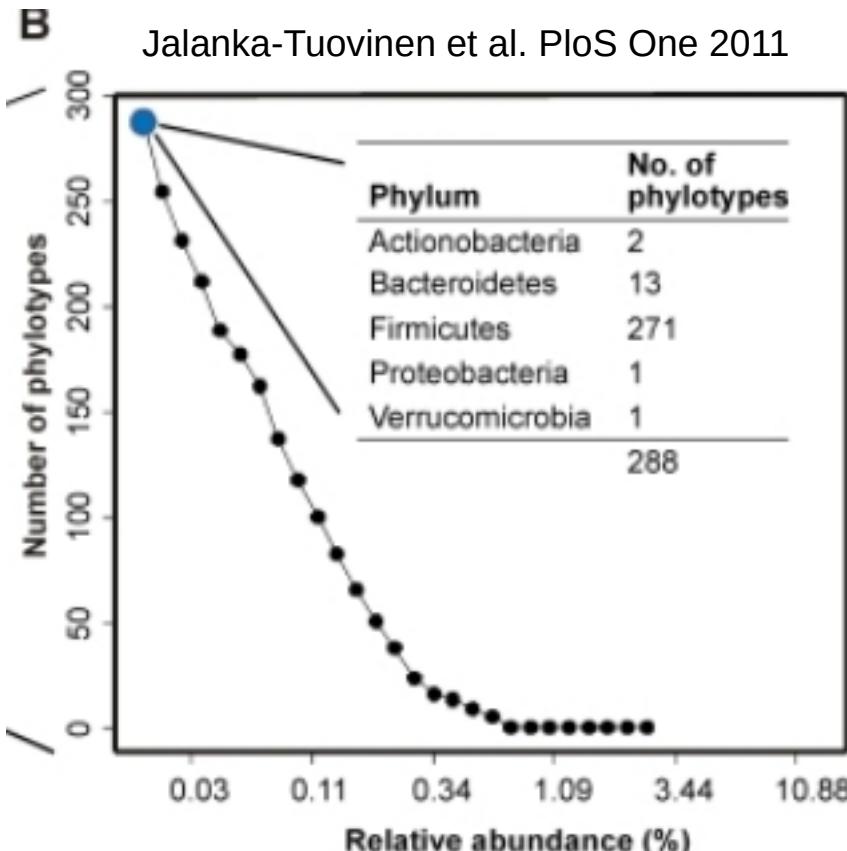
To guard against hype, those interpreting research on the body's microscopic communities should ask five questions, says **William P. Hanage**.

Comment August 2014 Nature

(Alpha) Diversity

True diversity, or the effective number of types, refers to the number of equally abundant types needed for the average proportional abundance of the types to equal that observed in the dataset of interest.

$$qD = \left(\sum_{i=1}^R p_i^q \right)^{1/(1-q)}$$



R: richness (number of distinct types)
p_i: proportion of type I

Order of diversity:

q = 0 : Species Richness

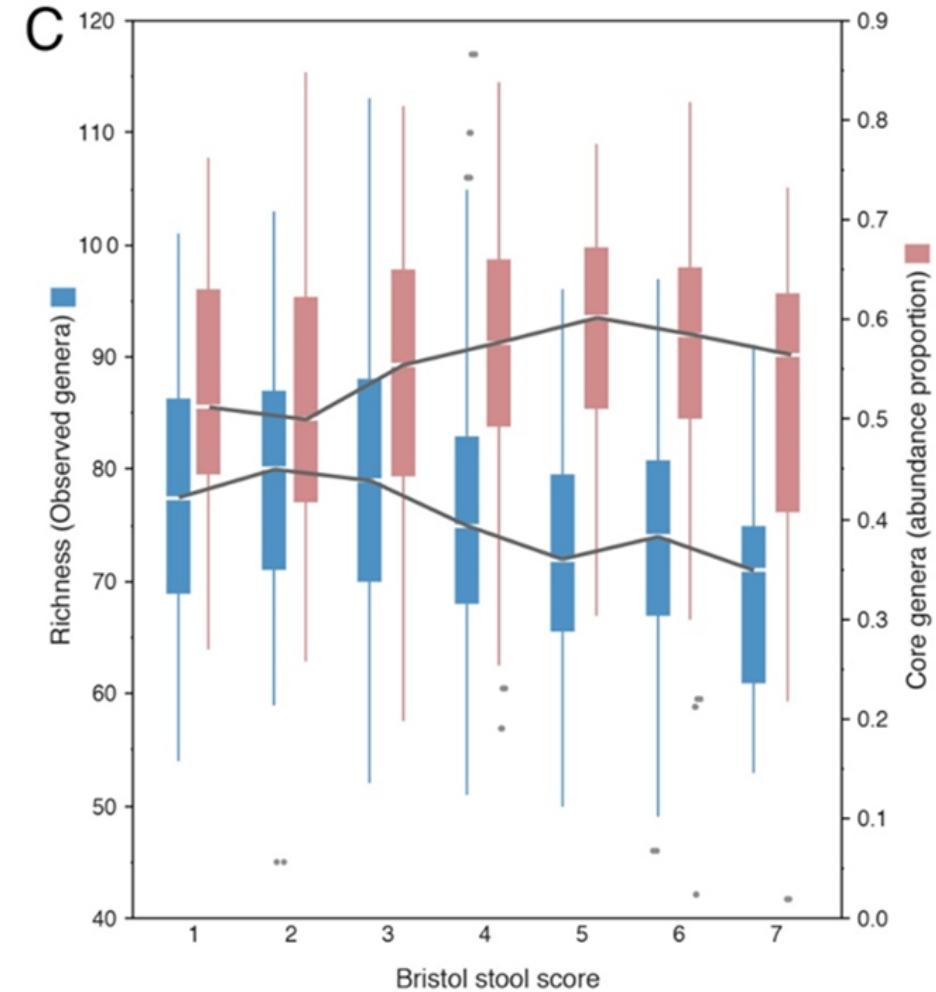
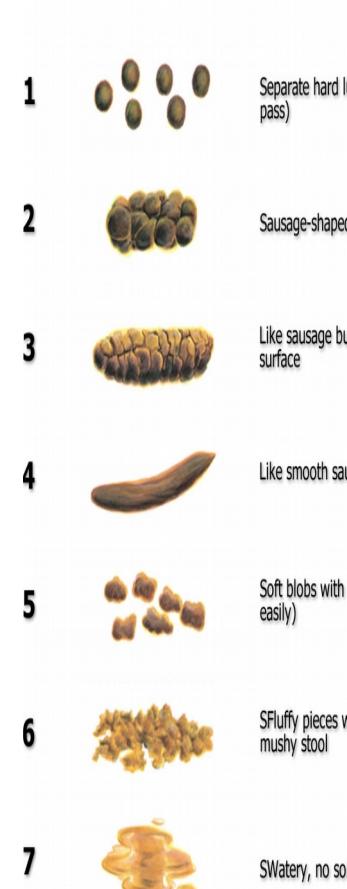
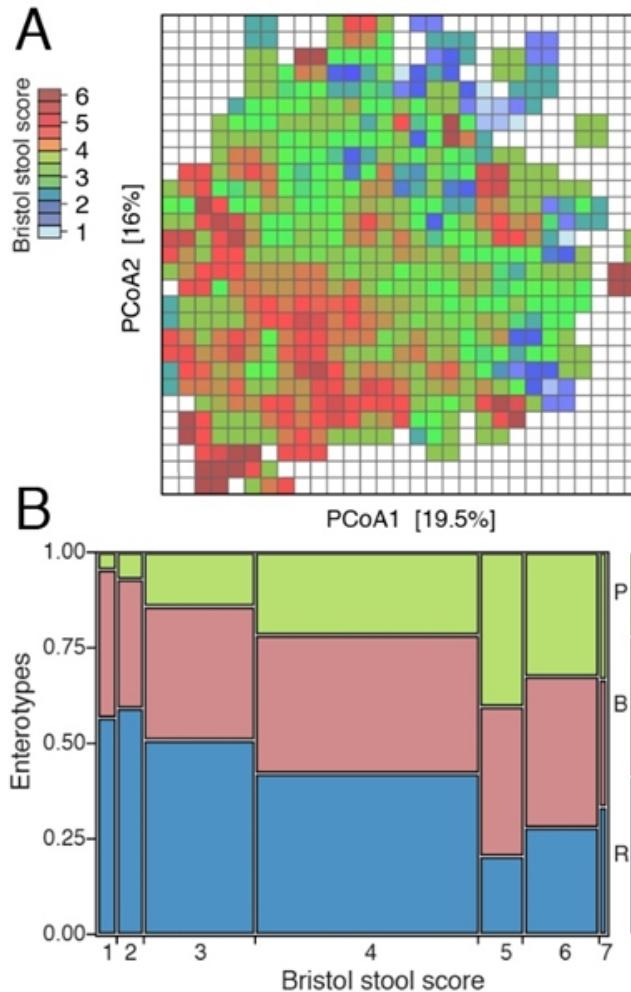
q = 1 : Shannon diversity

q = 2 : (Inverse) Simpson diversity

q ≠ 1 : Renyi entropy

Stool consistency (transit time proxy) linked to Prevotella enterotype, richness and proportion of core genera

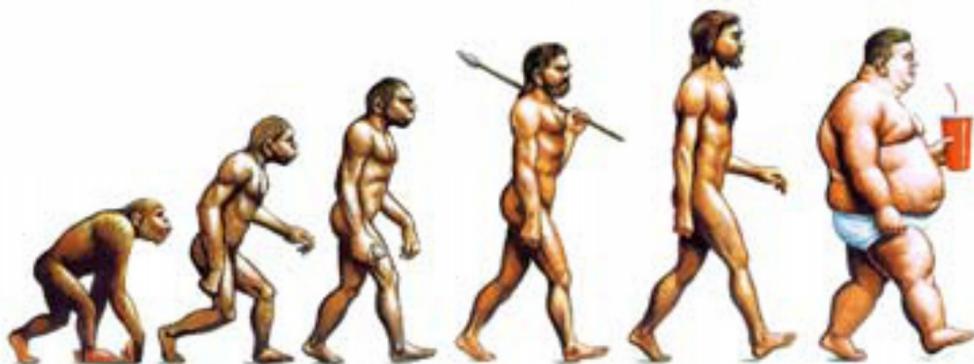
(cfr. Vandepitte et al Gut 2015 paper [n=54])



Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol* 1997

Bowel cancer risk may be reduced by rural African diet, study finds

Tests on subjects who swapped a fatty, meat-heavy diet for foods rich in beans and vegetables found a drop in biological markers for cancer in just two weeks



Diet swap – Two weeks ?



[Home](#) [About the journal](#) [Authors and referees](#) [Browse archive](#) [Search](#)

[nature.com](#) > [Journal home](#) > [Current month](#) > [Abstract](#)

ARTICLE PREVIEW

[View full access options](#) >

NATURE COMMUNICATIONS | ARTICLE



Fat, fibre and cancer risk in African Americans and rural Africans

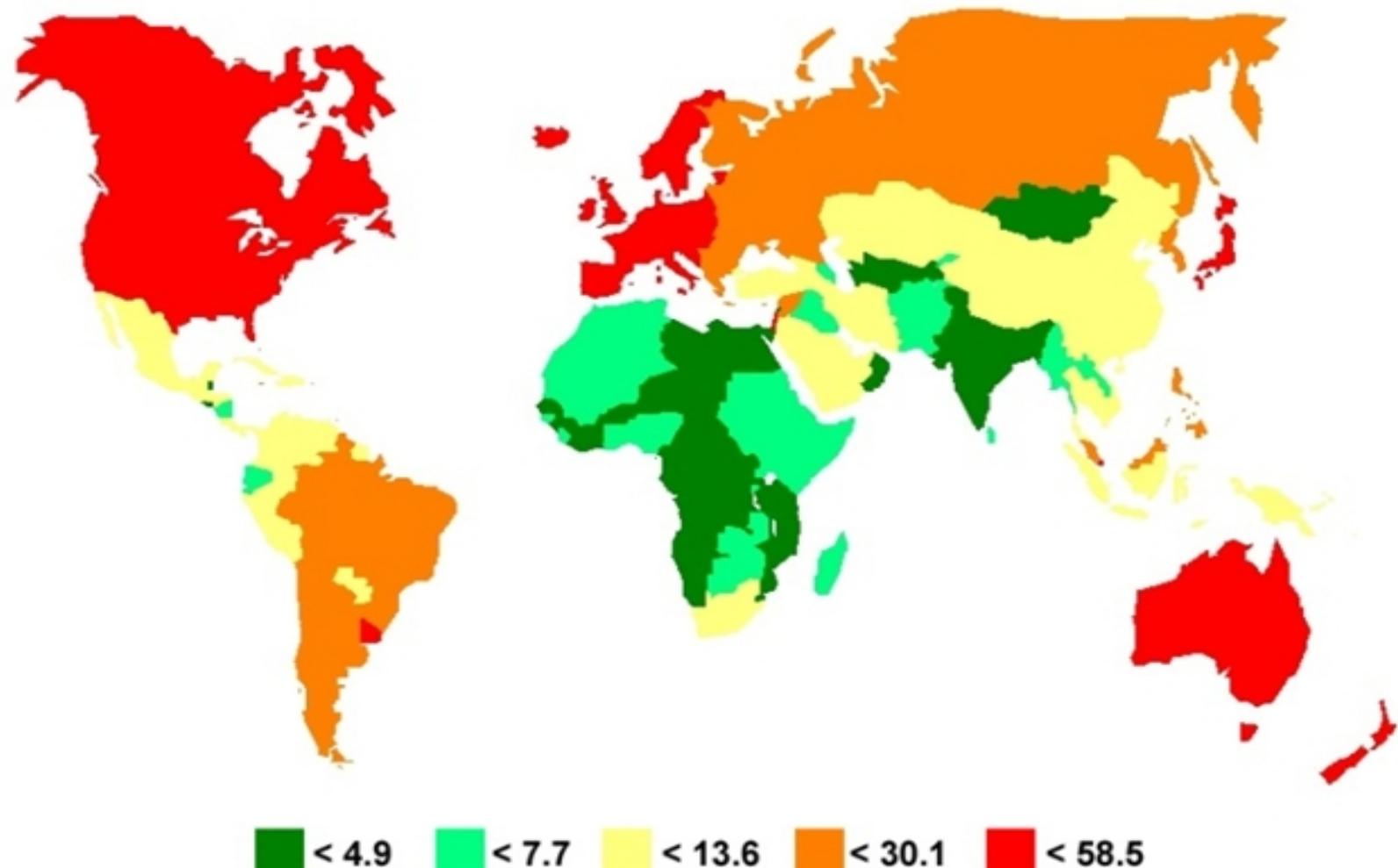
Stephen J. D. O'Keefe, Jia V. Li, Leo Lahti, Junhai Ou, Franck Carbonero, Khaled Mohammed, Joram M. Posma, James Kinross, Elaine Wahl, Elizabeth Ruder, Kishore Vipperla, Vasudevan Naidoo, Lungile Mtshali, Sebastian Tims, Philippe G. B. Puylaert, James DeLany, Alyssa Krasinskas, Ann C. Benefiel, Hatem O. Kaseb, Keith Newton [+ et al.](#)

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

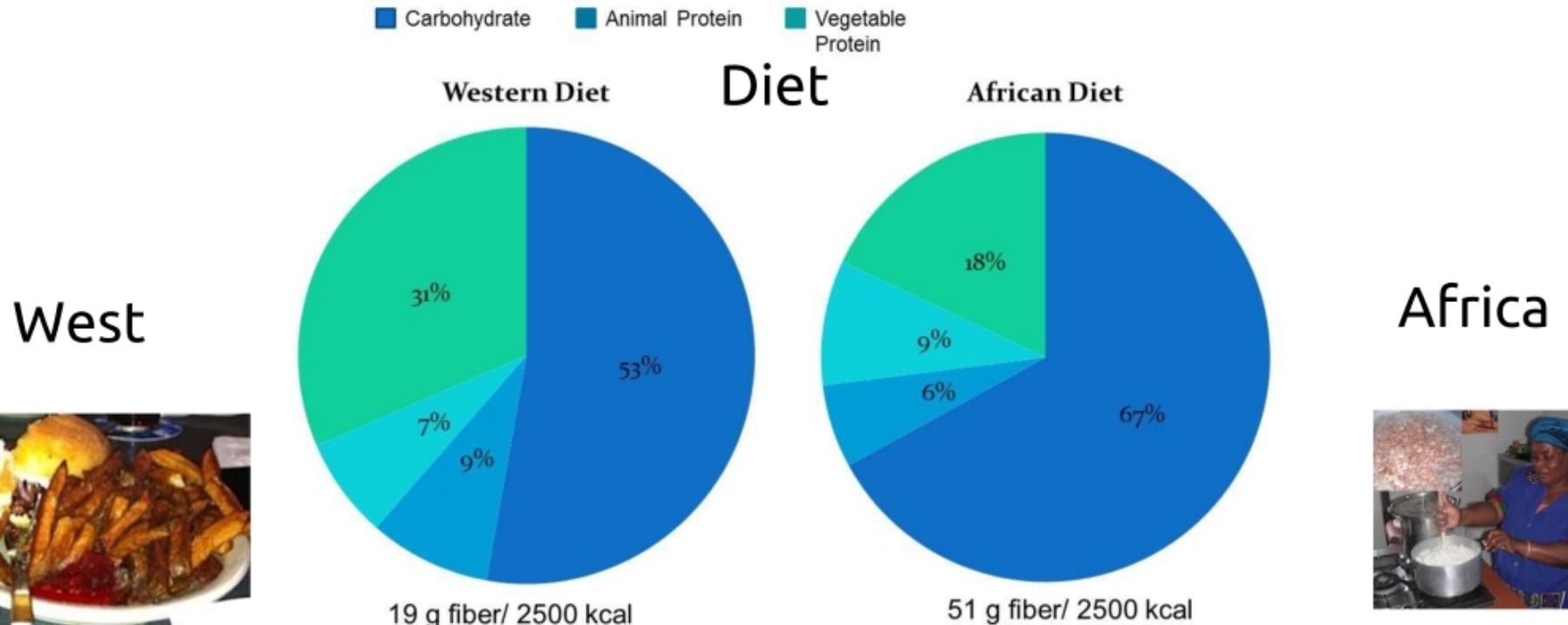
Nature Communications 6, Article number: 6342 | doi:10.1038/ncomms7342

Received 23 May 2014 | Accepted 20 January 2015 | Published 28 April 2015

Colon cancer prevalence



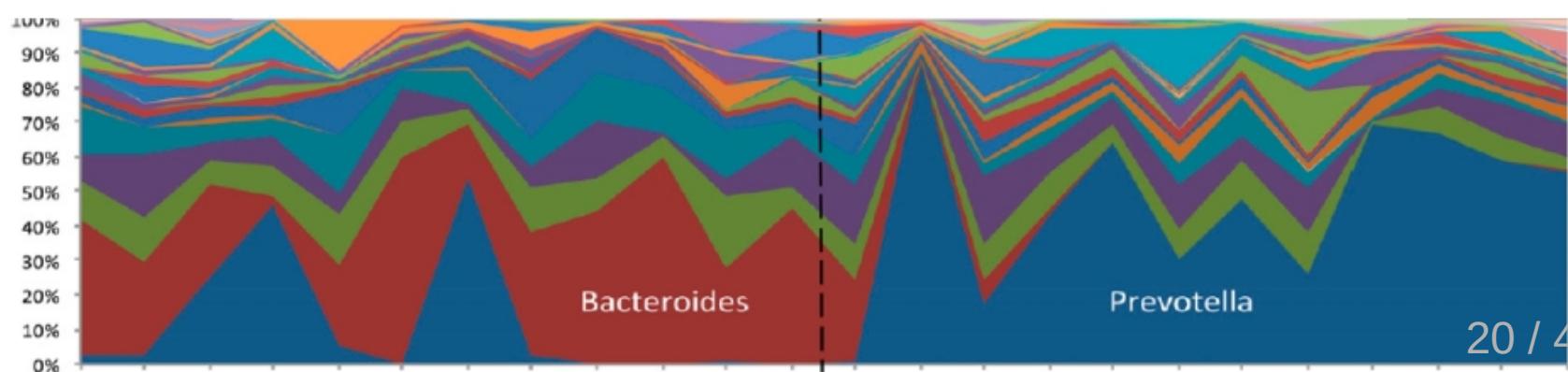
Colon cancer rates per country



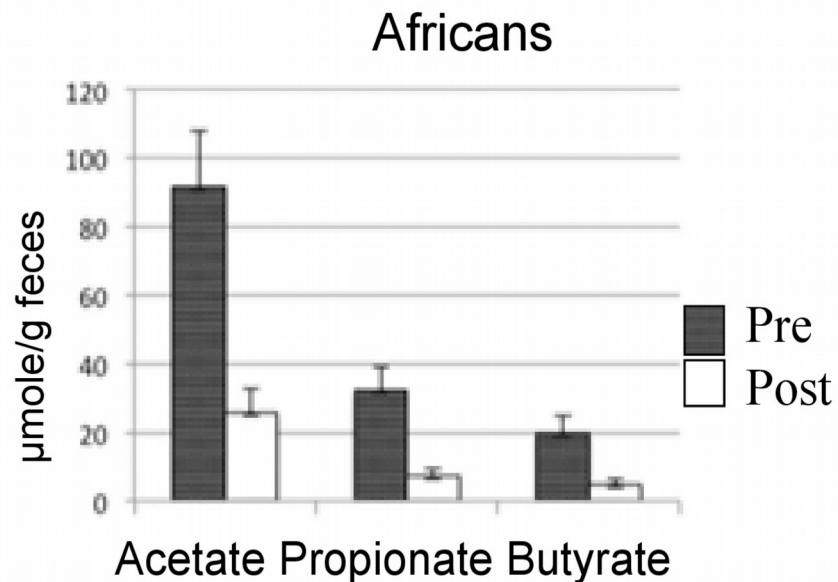
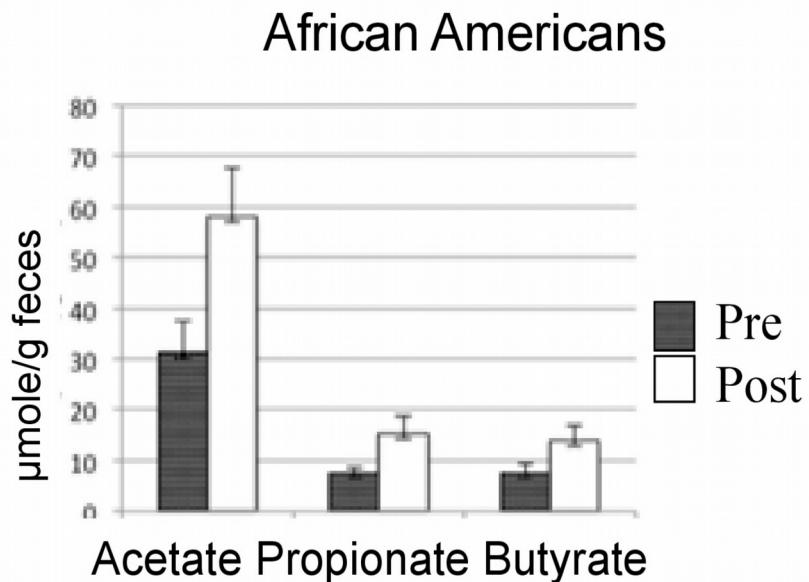
Microbiota composition

Colon cancer rates:
 -Africans:<10:100,000
 -African Americans:>65:100,000

Ou et al. Am J Clin Nutr.
 Jul 2013; 98(1): 111–120



Impact diet exchange on SCFA

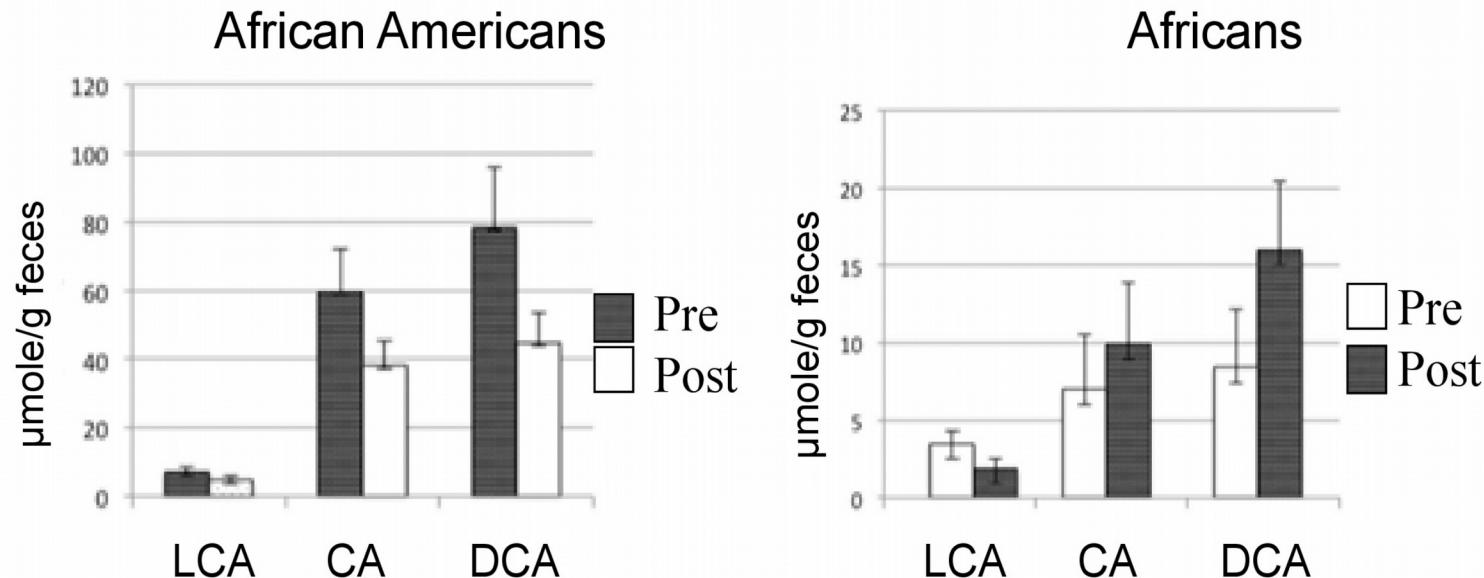


Reciprocally increased SCFA production with African diet: reported anti-inflammatory and anti-carcinogenic properties !

See also
Louis et al. Nat. Rev.
Microbiol Sept 2014

O'Keefe et al. Nat. Comm. 6:6342, 2015

Impact diet exchange on bile acids



LCA: Lithocholic Acid, CA: Cholic Acid, DCA: Deoxycholic Acid

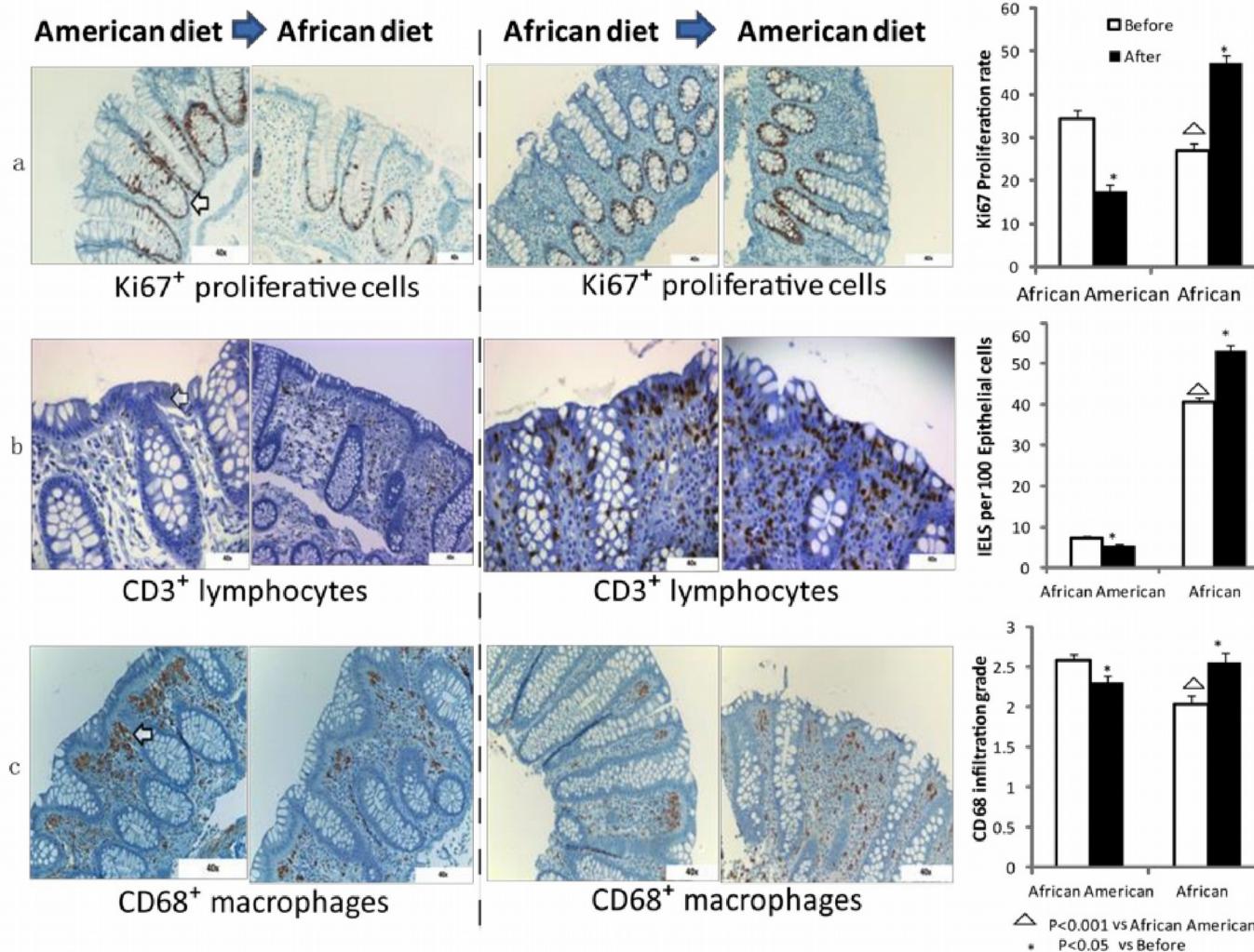
See also
Louis et al. Nat. Rev.
Microbiol Sept 2014

Primary and secondary bile acids
reciprocally increased with American diet:
reported pro-inflammatory and
carcinogenic properties !

22 / 45

Diet swap (2 weeks) and colon cancer risk

Reciprocal impact on epithelial biomarkers for colon cancer risk seen already after two weeks !



Fiber rich vegetarian food



Finnish Savo people?



African Zulu
(< 5:100,000)



African
americans
(65:100,000)

Meat and fat



Alaska Yupiks
>100:100,000



Stephen O'Keefe

“African” vegetarian diet – Two weeks

- + Inflammation-reducing microbes
- + Short-chain fatty acids
- Inflammation-promoting microbes
- Bile acids
- Cancer biomarkers (inflammation, cell proliferation)

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

Diet

Life style

Antibiotics

Probiotics

Prebiotics

Fecal transplants

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

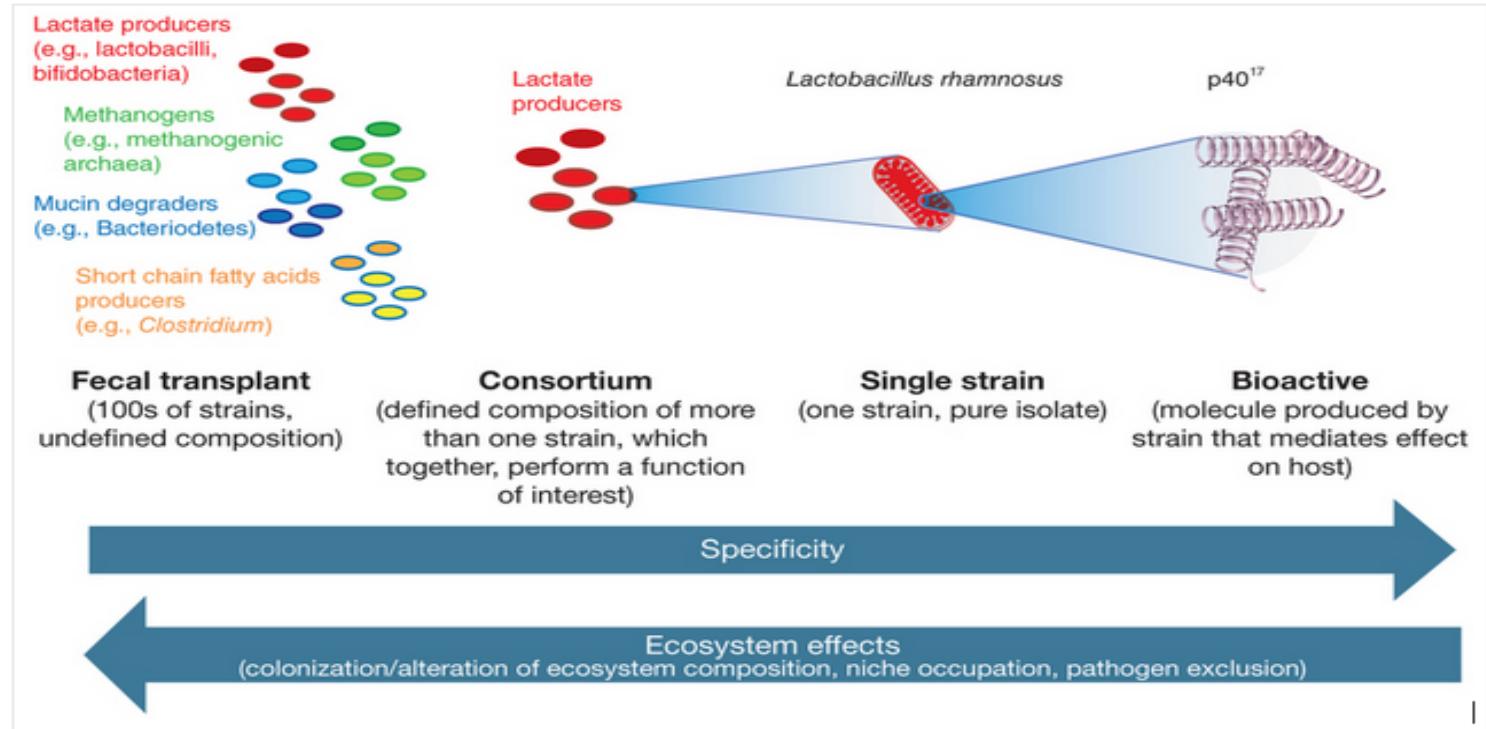
From

Medicines from microbiota

Bernat Ollé

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

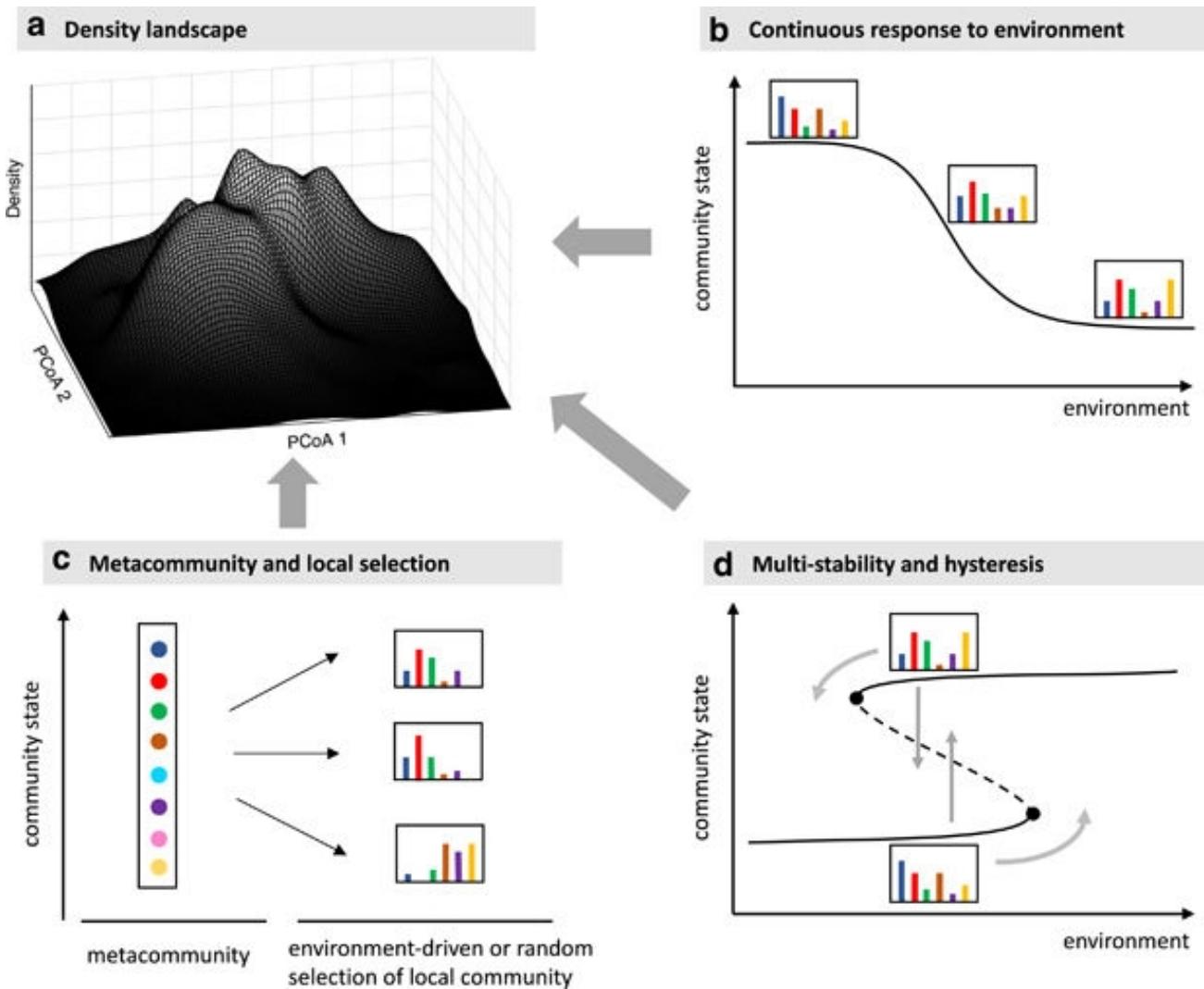
Figure 3: Spectrum of microbiome-derived modulators being pursued by biotech companies, ranging from ecosystem-level interventions to single-target approaches.



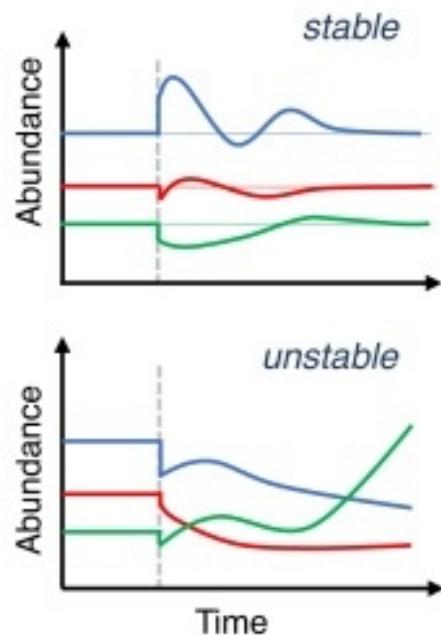
"Lactate producer" is used here as a functional attribute descriptive of a community. Species belonging to the "lactate producers" community (e.g., *L. rhamnosus*) may also belong to other communities. A community may be described by a metabolic function (e.g., lactate production) or by any other functional attribute (e.g., regulatory T-cell induction or vitamin K production). p40 is a bioactive, soluble protein expressed by *L. rhamnosus*, which mediates intestinal epithelial homeostasis¹⁷.

Microbial communities as dynamical systems

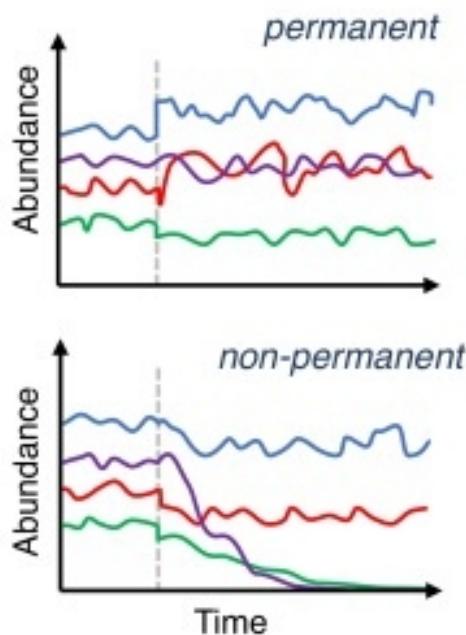
Didier Gonze ^{1, 2✉}, Katharine Z Coyte ^{3, 4}, Leo Lahti ^{5, 6, 7}, Karoline Faust ^{5✉}



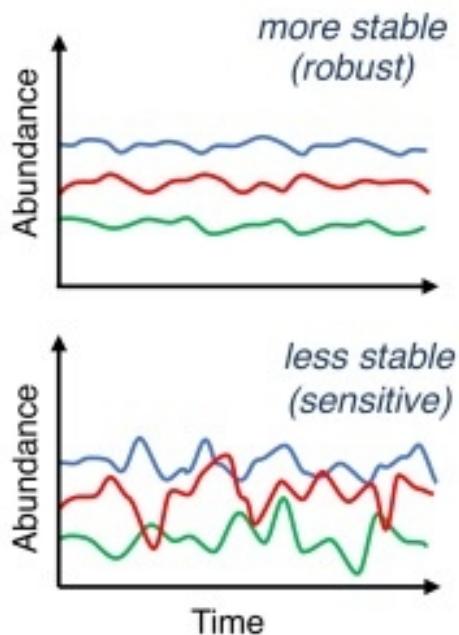
**(a) Steady state stability
(nonlinear theory)**



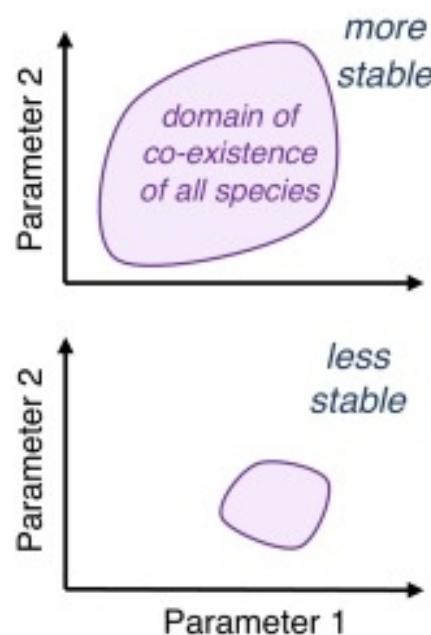
**Permanence
(long-term survival)**



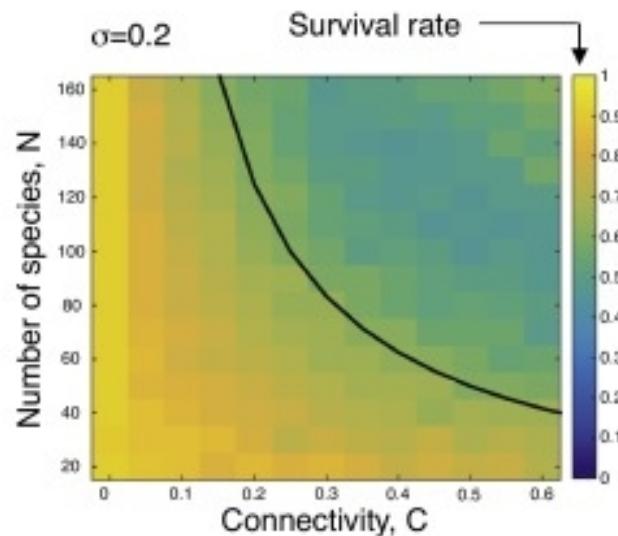
**Temporal stability
(robustness/sensitivity)**



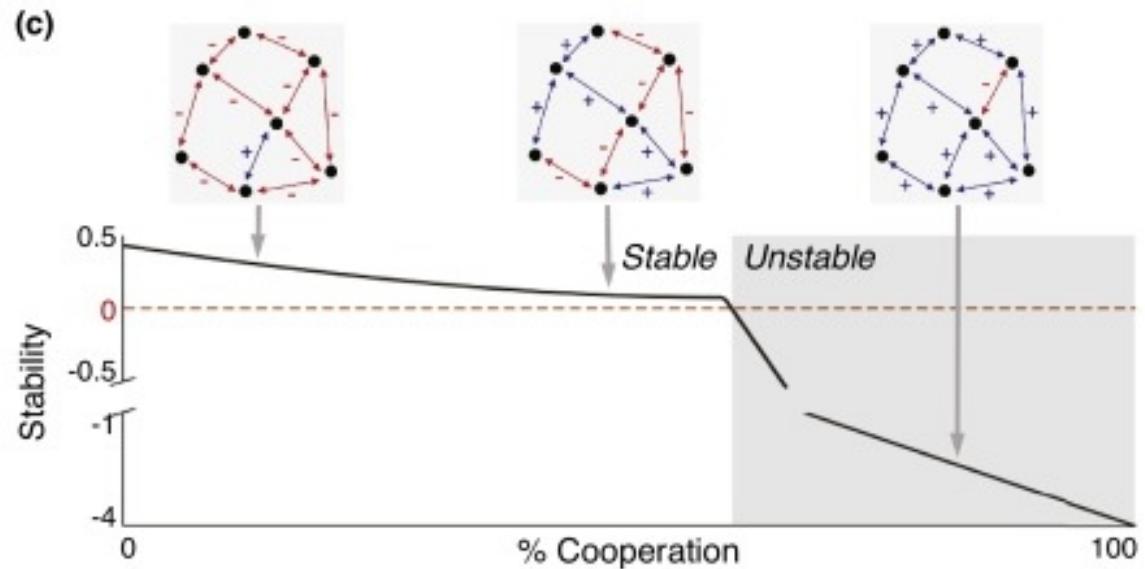
**Structural stability
(parameter space analysis)**



(b)



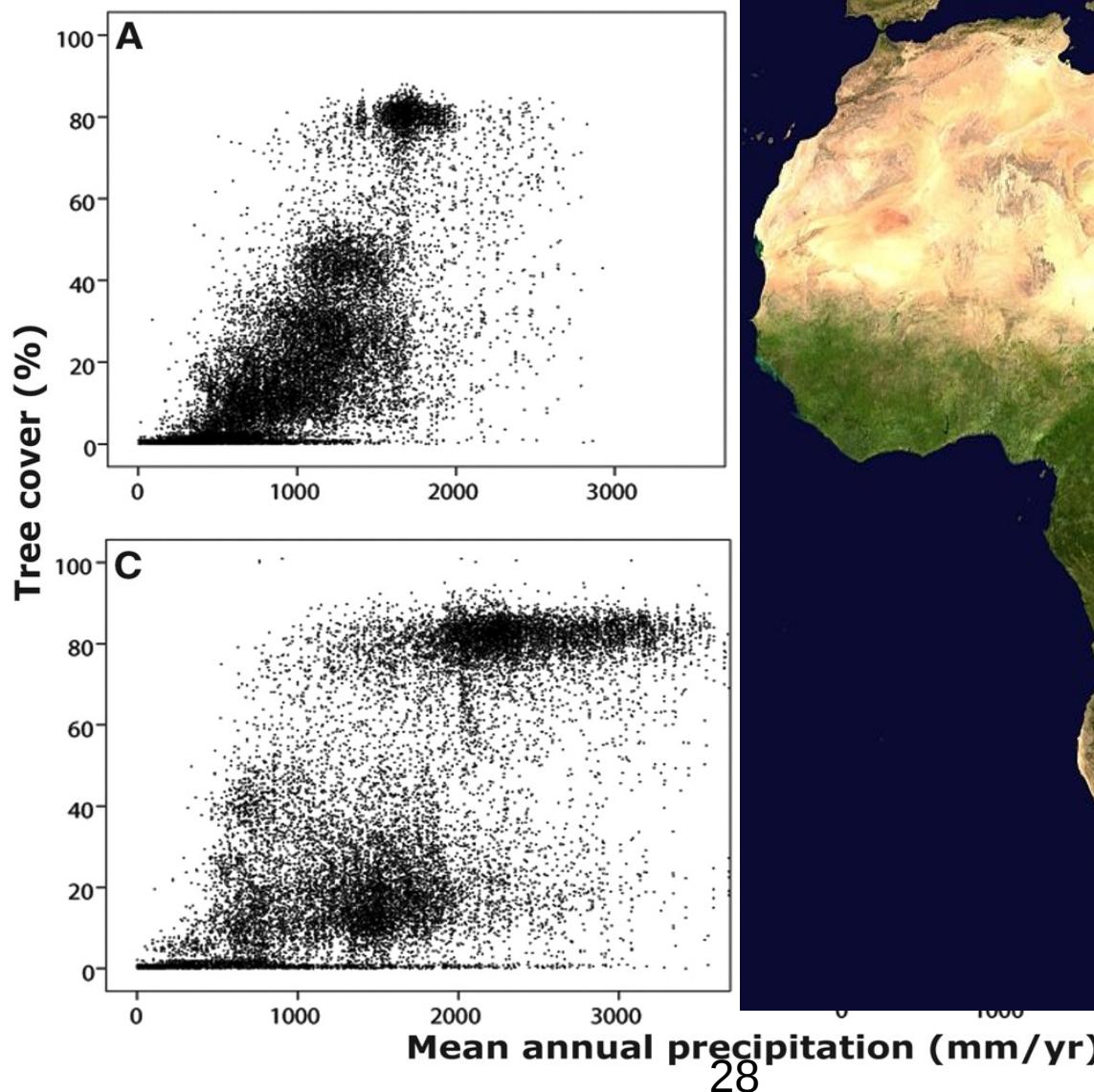
(c)



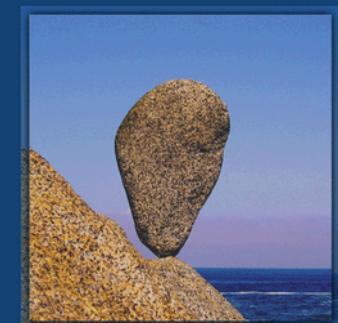
REPORT

Global Resilience of Tropical Forest and Savanna to Critical Transitions

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



Critical Transitions
in Nature and Society

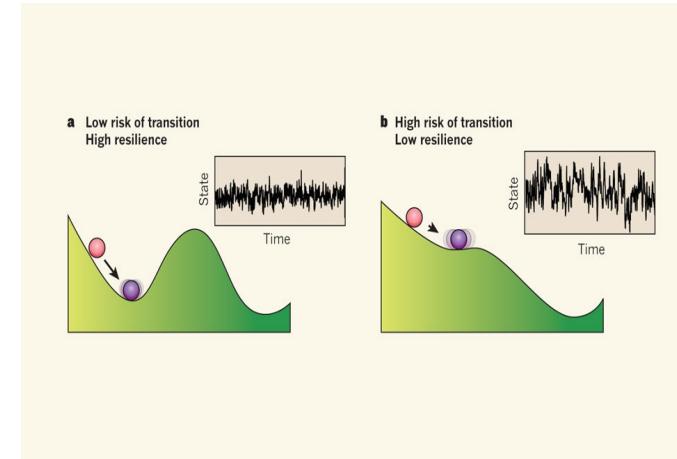


Marten Scheffer

PRINCETON STUDIES IN COMPLEXITY

Manipulation: 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 !

Patient gets donor microbiota -> Donor farms!?

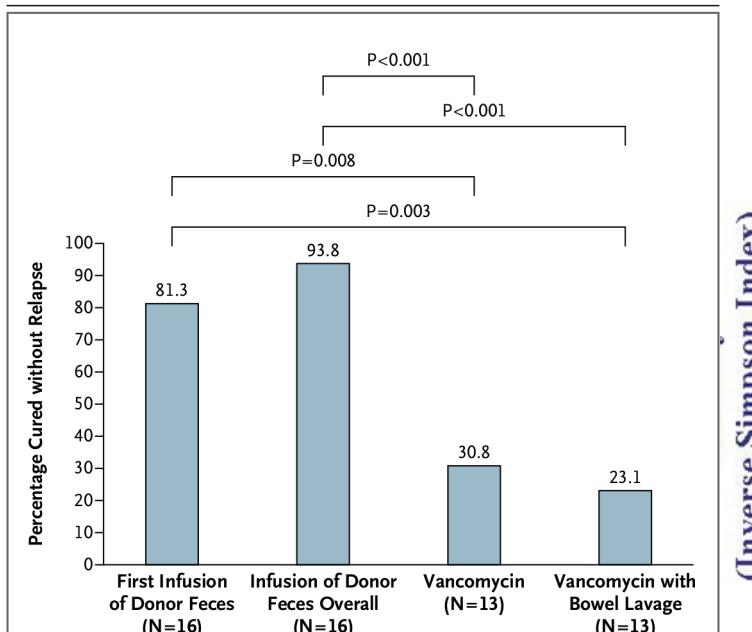
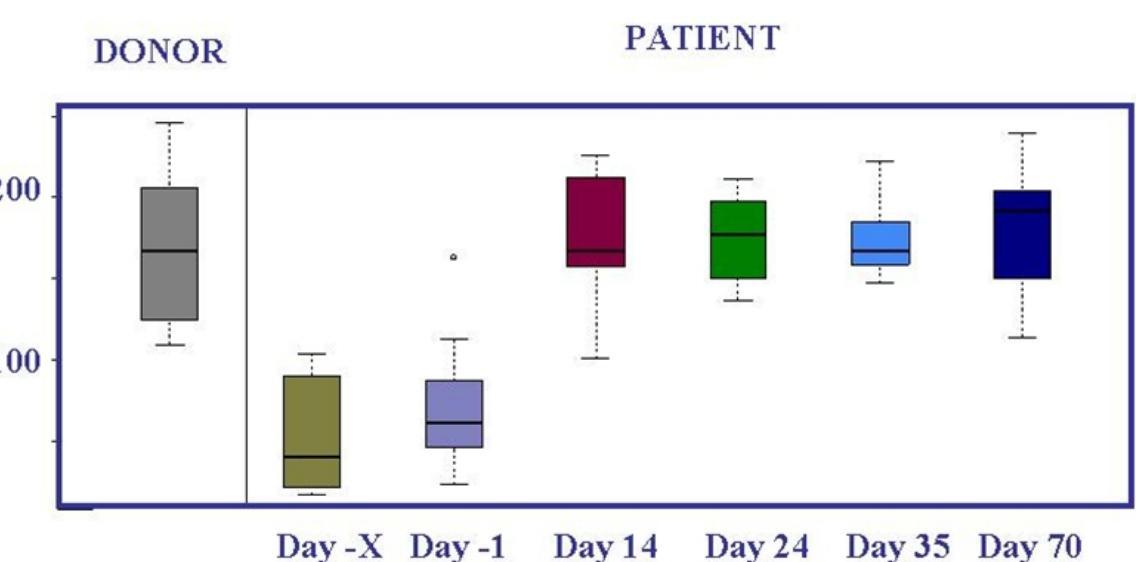


Figure 2. Rates of Cure without Relapse for Recurrent *Clostridium difficile* Infection.

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.



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

- dogs, camels and some other animals eat feces when sick; horses sometimes treated with fecal water

- 300AD Ge Hong (China) fecal water for food poisoning and sever diarrhea

- 1500 herbal doctors (Ming Dynasty) fermented fecal water for intestinal issues (“yellow soup”)

- 1958 first scientific publication on fecal transplants in humans

- 2000 onwards more research



Visualizing & modeling vaginal microbiome dynamics

Variation:

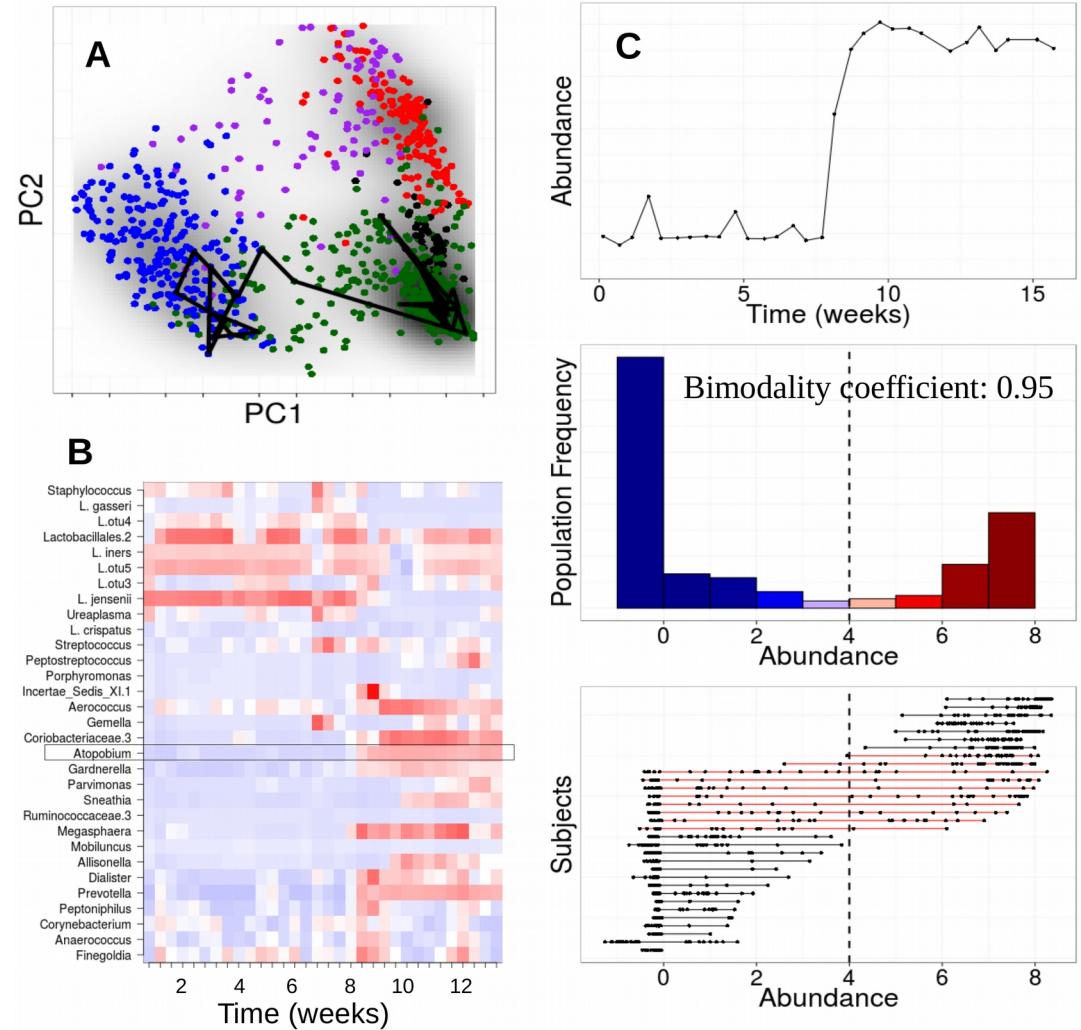
- cross-sectional
- spatial
- temporal

Levels of analysis:

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

Metagenomics meets time series analysis: unraveling microbial community dynamics

Karoline Faust^{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: Gajer *et al.* 2012

31

31 / 45

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 

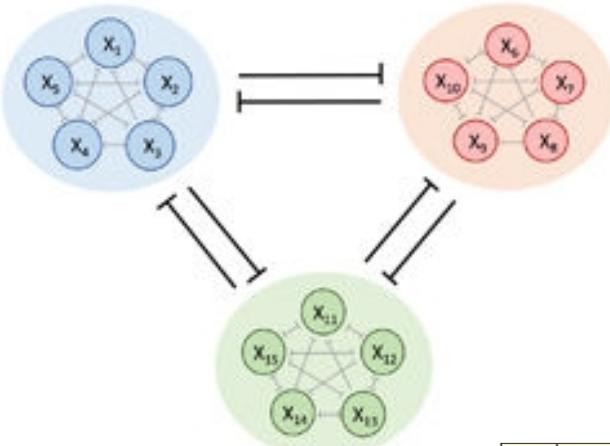
The ISME Journal (2017) **11**, 2159–2166 (2017)

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)

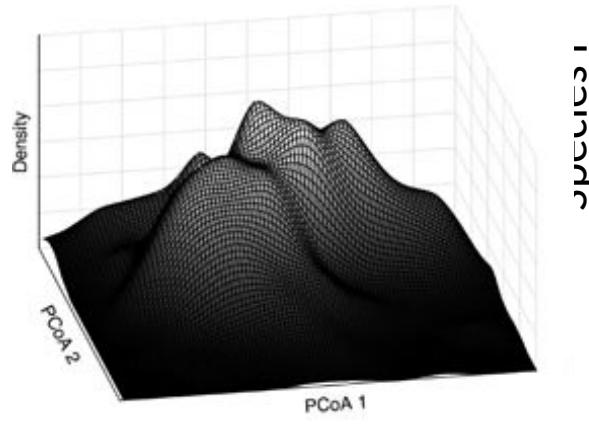


b

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

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

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



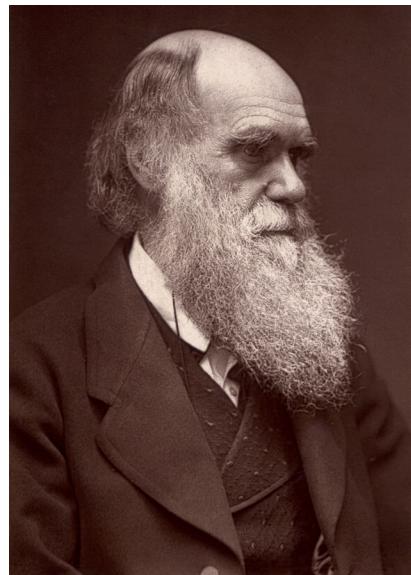
	2	4	6	8	10	12	14								
-	0.88	1	0.86	0.99	0.52	0.51	0.41	0.39	0.48	0.6	0.5	0.57	0.52	0.43	
2	1.17	-	0.79	1.06	1.11	0.51	0.5	0.53	0.5	0.58	0.43	0.53	0.49	0.42	0.39
4	1.2	1.04	-	0.97	0.87	0.52	0.5	0.57	0.47	0.5	0.53	0.36	0.56	0.44	0.55
6	1.06	1.28	1.19	-	0.93	0.46	0.47	0.55	0.57	0.57	0.49	0.62	0.43	0.48	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: could random chance explain variation in human gut microbiomes ?

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*)

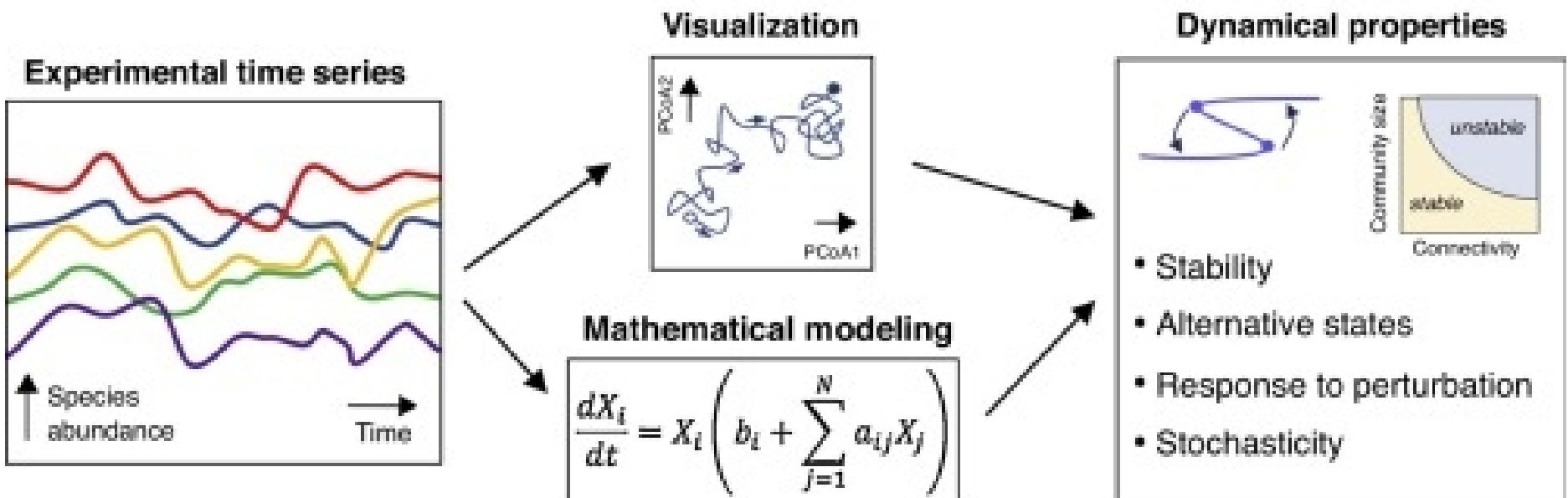


Neutral model

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

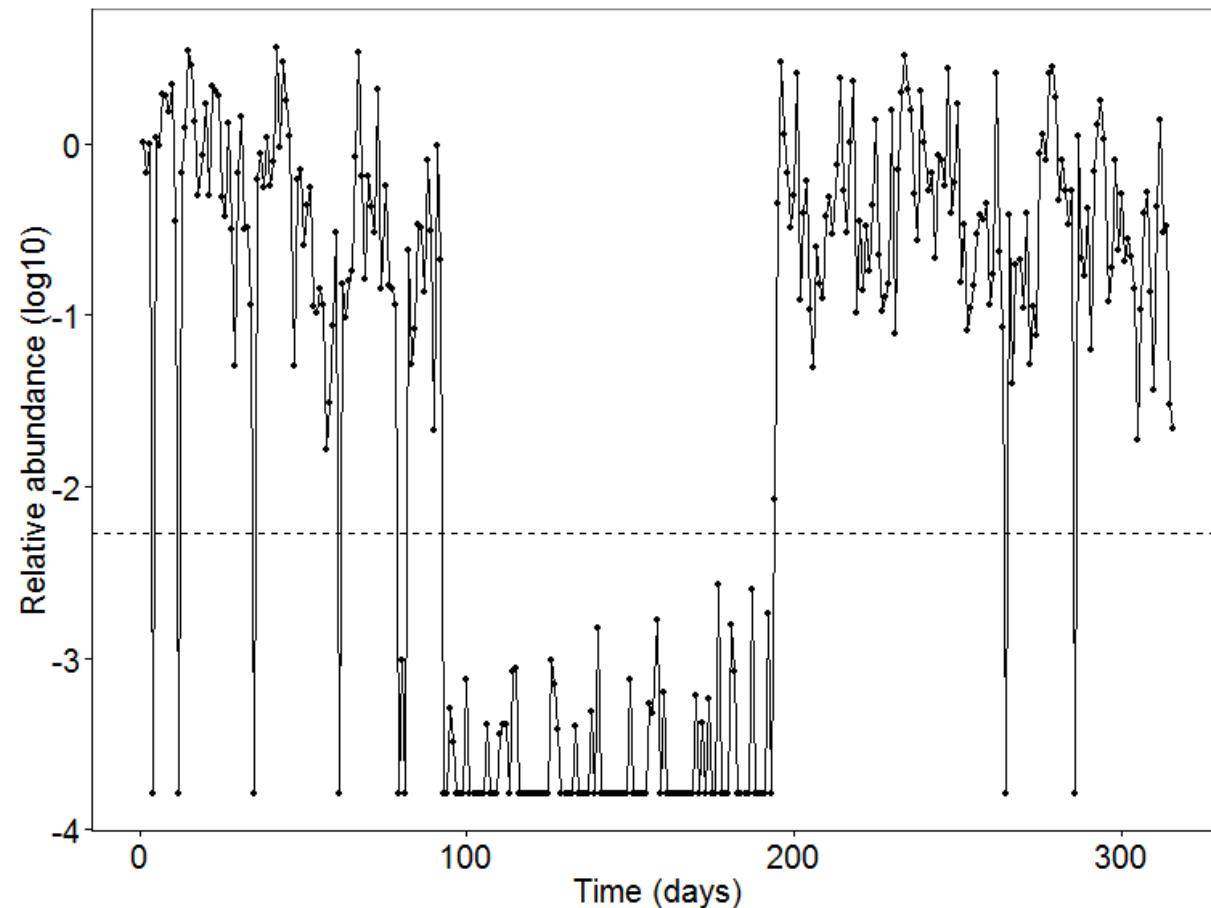


Modeling community assembly



Long and dense time series

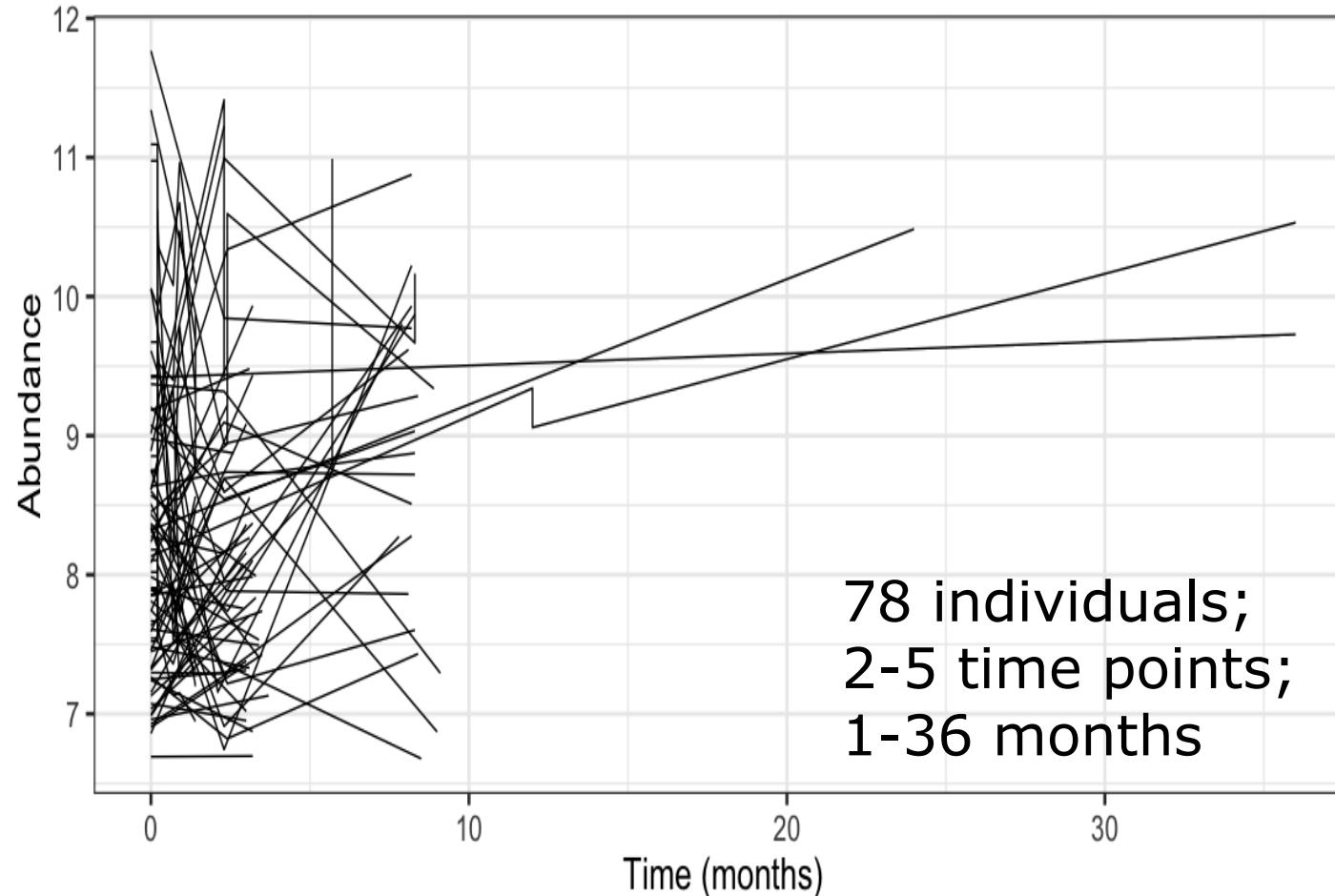
1 individual & 300+ days



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)

Modeling challenges

- Ecological and biochemical processes unknown
- Scarce data
- Lack of algorithms
- Lack of multi-disciplinary expertise

OPEN ACCESS

ESSAY

898,944

1,119

4,143

7,007

VIEWS

CITATIONS

SAVES

SHARES

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: Aug 30, 2005 • DOI: 10.1371/journal.pmed.0020124



Abstract

Modeling the Framework for False Positive Findings

Bias

Testing by Several Independent Teams

Corollaries

Most Research Findings Are False for Most Research Designs and for Most Fields

Claimed Research Findings May Often Be Simply Accurate Measures of the

Abstract

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate

Related PLOS Articles

[When Should Potentially False Research Findings Be Considered Acceptable?](#)

[Most Published Research Findings Are False—But a Little Replication Goes a Long Way](#)

[Minimizing Mistakes and Embracing Uncertainty](#)

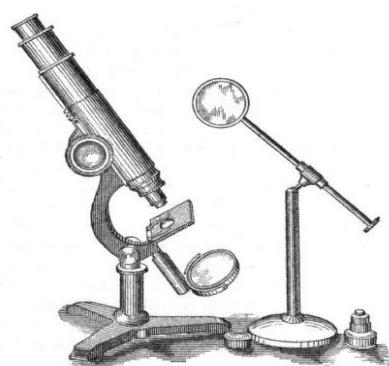
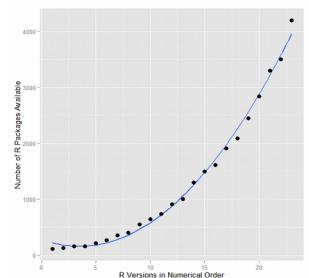
Subject Areas



How to Make More Published Research True

John P. A. Ioannidis 

Published: October 21, 2014 • DOI: 10.1371/journal.pmed.1001747



Science 13 April 2012:
Vol. 336 no. 6078 pp. 159-160
DOI: 10.1126/science.1218263

POLICY FORUM

RESEARCH PRIORITIES

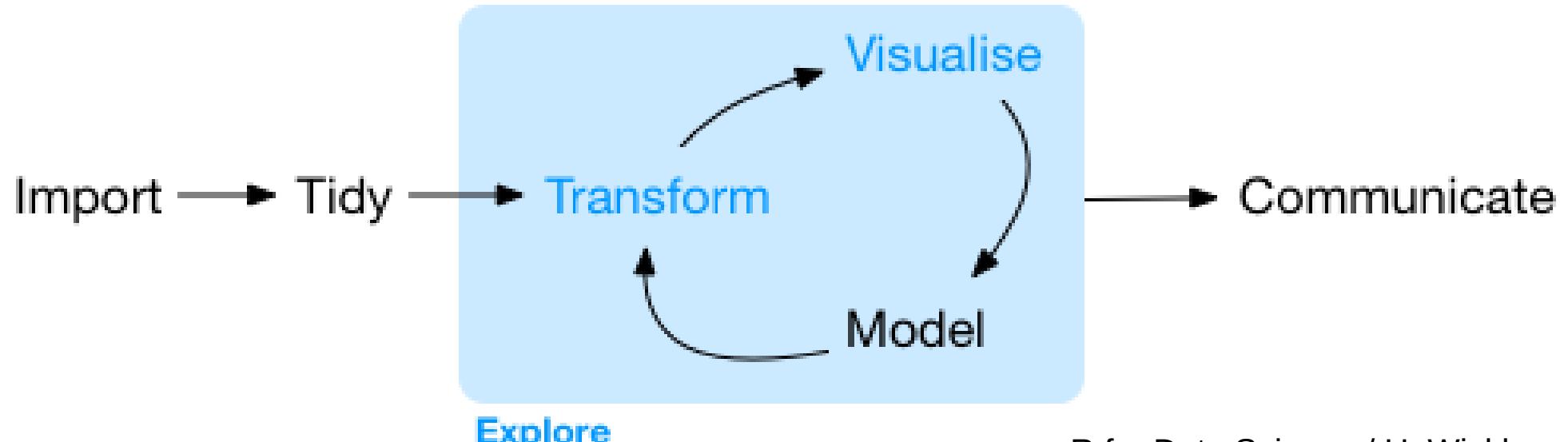
Shining Light into Black Boxes

A. Morin¹, J. Urban², P. D. Adams³, I. Foster⁴, A. Sali⁵, D. Baker⁶, P. Sliz^{1,*}

¹

Data science workflow

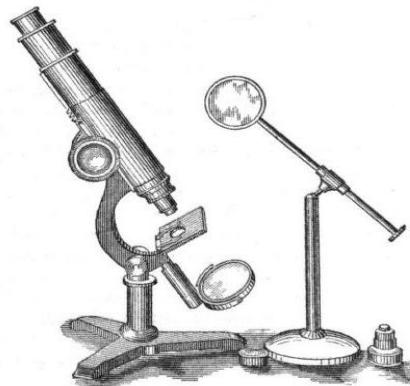
<http://microbiome.github.io>



Program

R for Data Science / H. Wickham

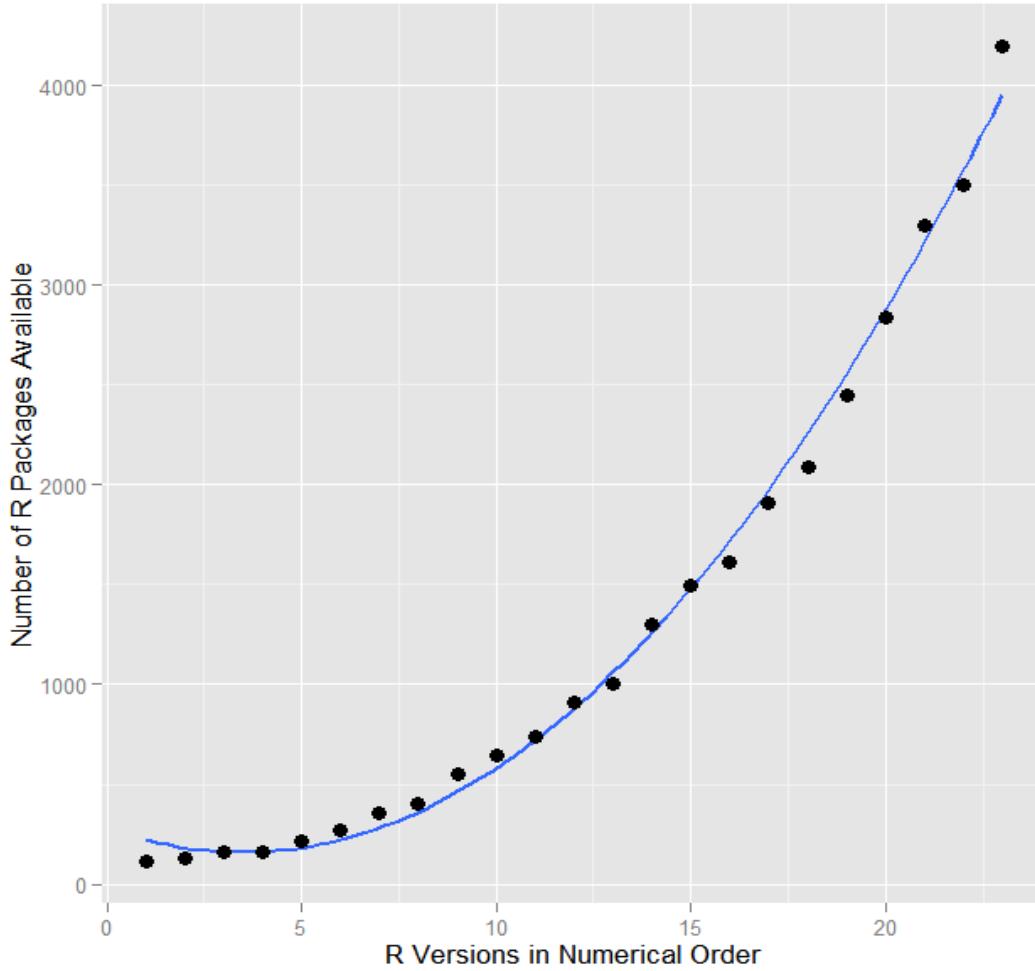
Computational workflows have an increasingly central role in data-intensive research fields



(bioinformatics, climatology, economics, ecology, physics)

similar data formats, related analytical challenges !

Number of analysis tools for R now growing exponentially



microbiome R package

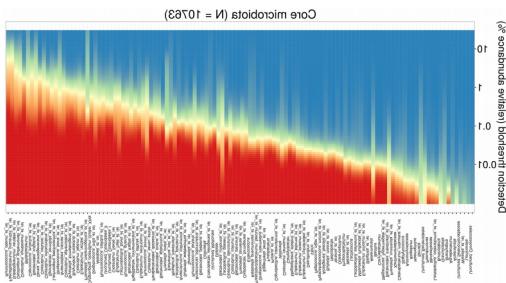


Core & prevalence

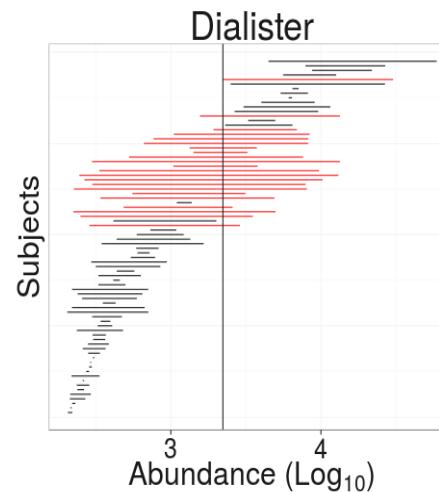
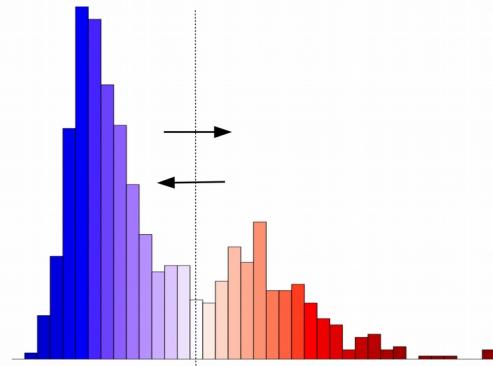
`prevalence(x)`

`core(x)`

`core_members(x)`



Stability & resilience



Transformations

`transform(x, "compositional")`

`transform(x, "clr")`

`transform(x, "log10p")`

`transform(x, "hellinger")`

`transform(x, "identity")`

Alpha & beta diversity

`alpha(x)`

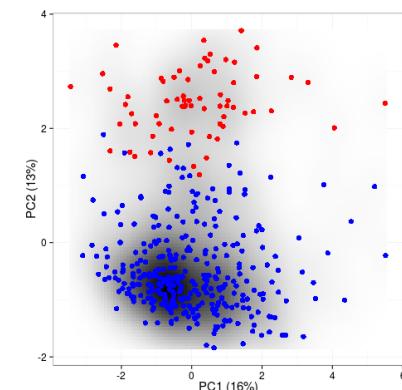
`diversity(x)`

`evenness(x)`

`dominance(x)`

`rarity(x)`

`readcount(x)`



Package website

- Tutorials & support
- Data & code sharing

Conclusion

- Human microbiomics: emerging multi-disciplinary field
- Majority of the variation still is poorly understood
- Population studies provide context for understanding individuality and dynamics
- Scientific computing is increasingly central for data-intensive research
- Open science can improve research quality and accelerate scientific progress

openresearchlabs.github.io

Acknowledgments

openresearchlabs.github.io

National Institute for Health and Welfare (THL), Finland
de Vos Lab, Wageningen University, The Netherlands
Raes Lab, VIB / KU Leuven, Belgium