

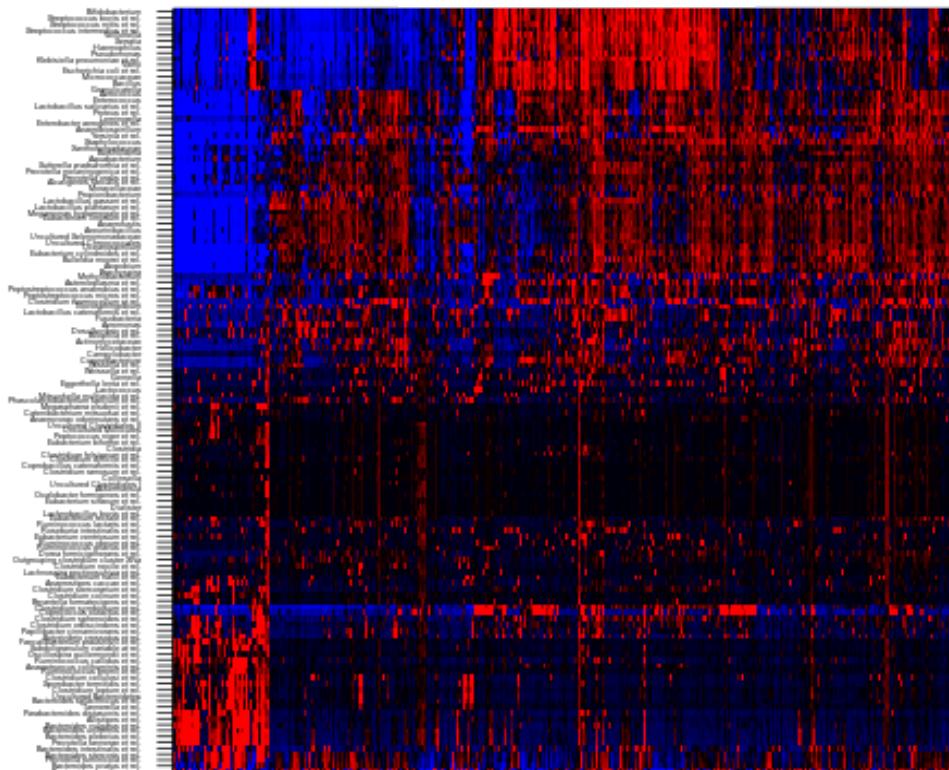
# Opportunities and challenges in large-scale microbiome profiling studies

Leo Lahti

<http://www.iki.fi/Leo.Lahti>

University of Turku, Finland & VIB/KU Leuven, Belgium

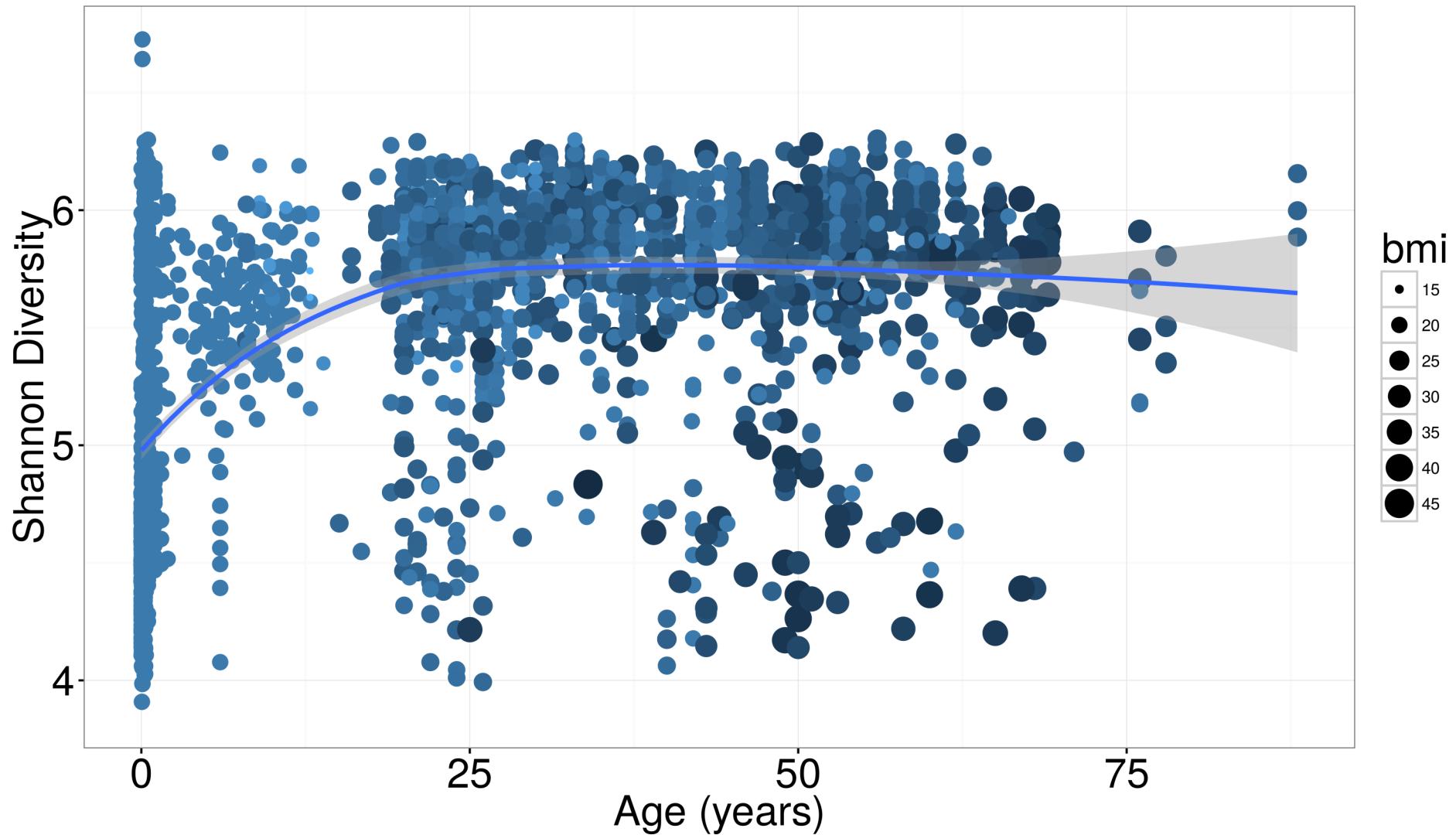
1000 bacterial photypes



N ~ 10,000+ samples

# Microbiome diversity and age: healthy & normal obese subjects

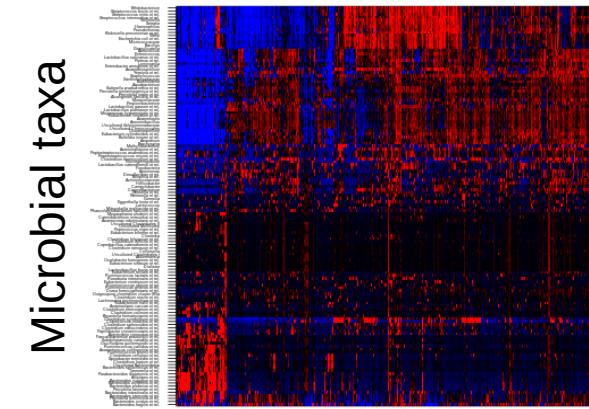
N = 2363



Data: HITChip Atlas;  
healthy fecal RBB samples

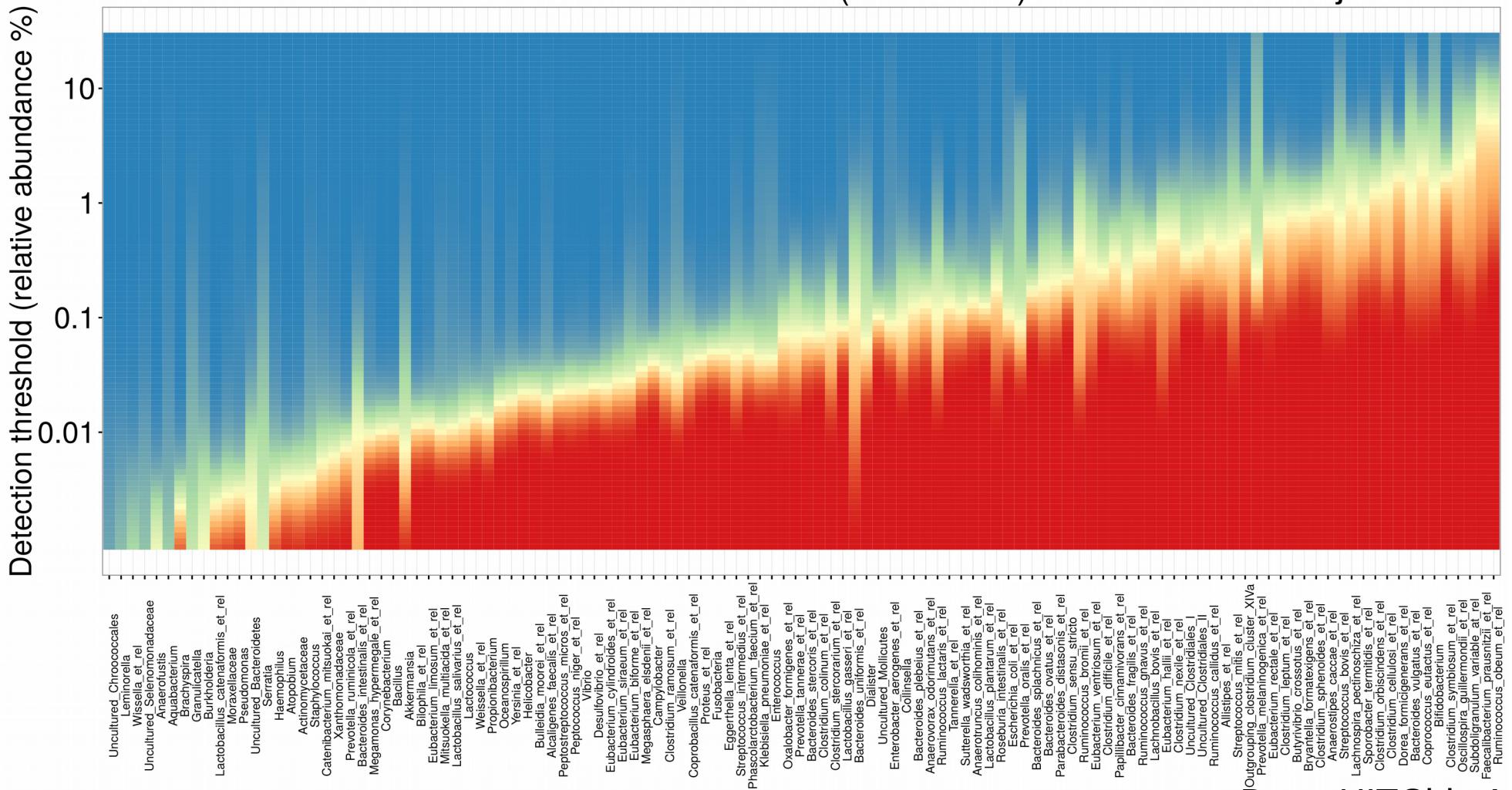
# Core microbiota

only few species are prevalent (shared)  
in population at a high abundance



Microbial taxa

Subjects



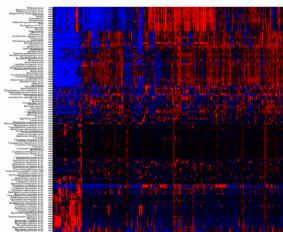
Uncultured Chroococcales  
Lemniscella  
Uncultured Selenomonadaceae  
Anaerofustis  
Aquadacterium  
Brachyspira  
Granulicella  
Burkholderia  
Lactobacillus\_cateniformis\_et\_rei  
Moraxellaceae  
Pseudomonas  
Uncultured Bacteroidetes  
Serrata  
Haenophilus  
Altoplumbum  
Actinomytaceae  
Staphylococcus  
Catenibacterium\_mitsuokai\_et\_rei  
Xanthomonadaceae  
Prevotella\_ruminicola\_et\_rei  
Bacteroides\_intestinalis\_et\_rei  
Megamona\_hypermegale\_et\_rei  
Corynebacterium  
Bacillus  
Akermansia  
Bifidobacteria\_et\_rei  
Eubacterium\_limosum\_et\_rei  
Mitsoukella\_multicida\_et\_rei  
Lactobacillus\_salivarius\_et\_rei  
Lactococcus  
Weissella\_et\_rei  
Propionibacterium  
Oceanospirillum  
Oscina\_et\_rei  
Helicobacter  
Bullidida\_moorei\_et\_rei  
Alcaligenes\_faecalis\_et\_rei  
Peptococcus\_niger\_et\_rei  
Vibrio  
Desulfobulbrio\_et\_rei  
Eubacterium\_cylindroides\_et\_rei  
Eubacterium\_siraenum\_et\_rei  
Eubacterium\_bifidum\_et\_rei  
Megaplaera\_lesdenii\_et\_rei  
Campylobacter  
Clostridium\_ramosum\_et\_rei  
Coprobacter  
Coprobacter\_cateniformis\_et\_rei  
Proteus\_et\_rei  
Fusobacteria  
Eggerthella\_lenta\_et\_rei  
Streptococcus\_intermedius\_et\_rei  
Phascolarctobacterium\_faecium\_et\_rei  
Klebsiella\_pneumoniae\_et\_rei  
Enterococcus  
Oxalobacter\_formigenes\_et\_rei  
Prevotella\_tannerae\_et\_rei  
Bacteroides\_stercoris\_et\_rei  
Clostridium\_colinum\_et\_rei  
Clostridium\_stercorarium\_et\_rei  
Lactobacillus\_gasseri\_et\_rei  
Bacteroides\_uniformis\_et\_rei  
Dialister  
Untutured\_Molluscues  
Enterobacter\_aerogenes\_et\_rei  
Collinsella  
Bacteroides plebeius\_et\_rei  
Anaerovorax odorifaciens\_et\_rei  
Ruminococcus\_lactaris\_et\_rei  
Tannerella\_et\_rei  
Sutterella\_wadsworthia\_et\_rei  
Anaerotruncus\_colloformis\_et\_rei  
Lactobacillus\_planatum\_et\_rei  
Roseburia\_intestinis\_et\_rei  
Escherichia\_coli\_et\_rei  
Prevotella\_oralis\_et\_rei  
Bacteroides\_spalchicus\_et\_rei  
Bacteroides\_ovatus\_et\_rei  
Parabacteroides\_distasonis\_et\_rei  
Clostridium\_sensu\_stricio  
Ruminococcus\_bromii\_et\_rei  
Eubacterium\_vestricosum\_et\_rei  
Clostridium\_difficile\_et\_rei  
Papillibacter\_citramavorans\_et\_rei  
Bacteroides\_fragilis\_et\_rei  
Ruminococcus\_gravus\_et\_rei  
Lachnobacillus\_bovis\_et\_rei  
Eubacterium\_hallii\_et\_rei  
Clostridium\_nexile\_et\_rei  
Uncultured\_Clostridiates  
Ruminococcus\_calidus\_et\_rei  
Allistipes\_et\_rei  
Streptococcus\_mitis\_et\_rei  
Prevotella\_melanogena\_et\_rei  
Eubacterium\_rectale\_et\_rei  
Clostridium\_leptum\_et\_rei  
Butyrivibrio\_crossotus\_et\_rei  
Bryantella\_formalexesiensis\_et\_rei  
Clostridium\_sphenoides\_et\_rei  
Anaerostipes\_accaee\_et\_rei  
Streptococcus\_dovii\_et\_rei  
Lachnospira\_pepticuschia\_et\_rei  
Sporobacter\_termidis\_et\_rei  
Clostridium\_tribeculindens\_et\_rei  
Clostridium\_cellulosi\_et\_rei  
Dorea\_formicigenans\_et\_rei  
Bacteroides\_vulgaris\_et\_rei  
Coprooccus\_eutacticus\_et\_rei  
BlidoBacterium  
Clostridium\_symbosum\_et\_rei  
Oschroma\_guillermondi\_et\_rei  
Subodoligranulum\_varabile\_et\_rei  
Faecalibacterium\_prausnitzii\_et\_rei  
Ruminococcus\_oebeum\_et\_rei

Data: HITChip Atlas

# Data availability: large & standardized population cohorts

## HITChip Atlas (microarrays):

- > 10,000 samples
- > 5,000 subjects
- > 1,000 phylotypes (~0.1%)
- versatile geography
- highly reproducible (>98%)
- long follow-ups (10+ y)
- open data: Lahti et al. 2014



## Flemish Gut Flora (NGS):

- > 5,000 samples
- > 1,000 subjects
- > 7,000 OTUs
- normal Flemish population
- dense time series
- open data: Falony et al. 2016



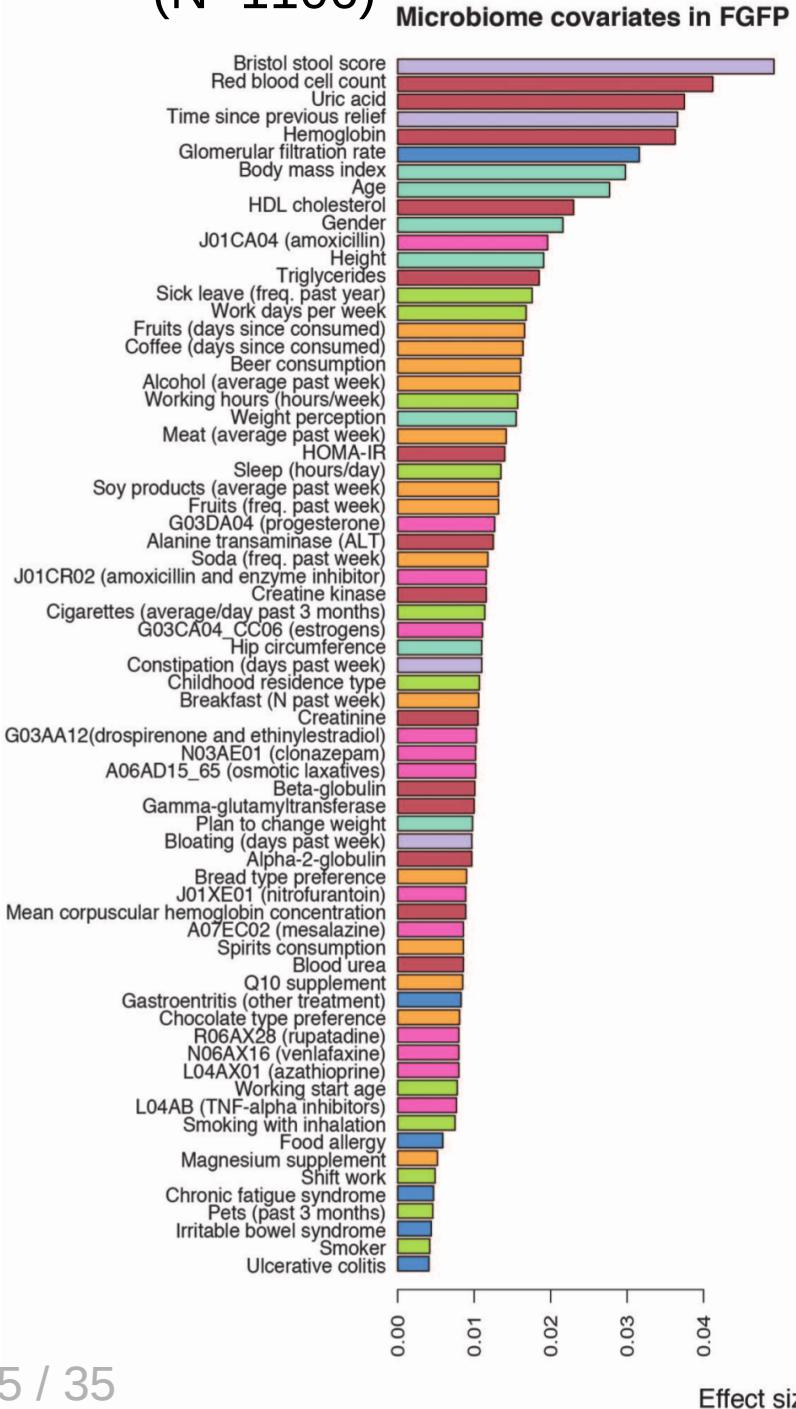
## Challenges for data availability:

- Protocol & metadata standardization
- Custom tools & reproducible workflows
- Ethical considerations

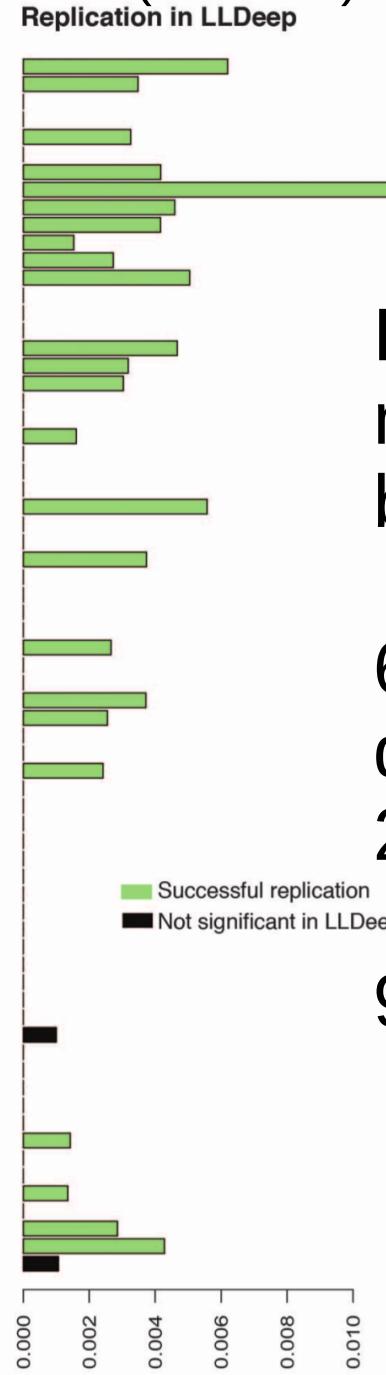
## Analysis & visualization tools: <http://microbiome.github.io>

Other collections: MetaHIT, Human Microbiome Project, American Gut...

## Flemish Gut Flora (N=1106)



## Dutch LifeLines- DEEP (N=1135)



Determinants of  
microbiota  
composition ?

Diet, health, lifestyle,  
medication, host variables,  
blood, bowel habits..

69 covariates (clinical &  
questionnaire) in FGFP;  
26 shared with DEEP

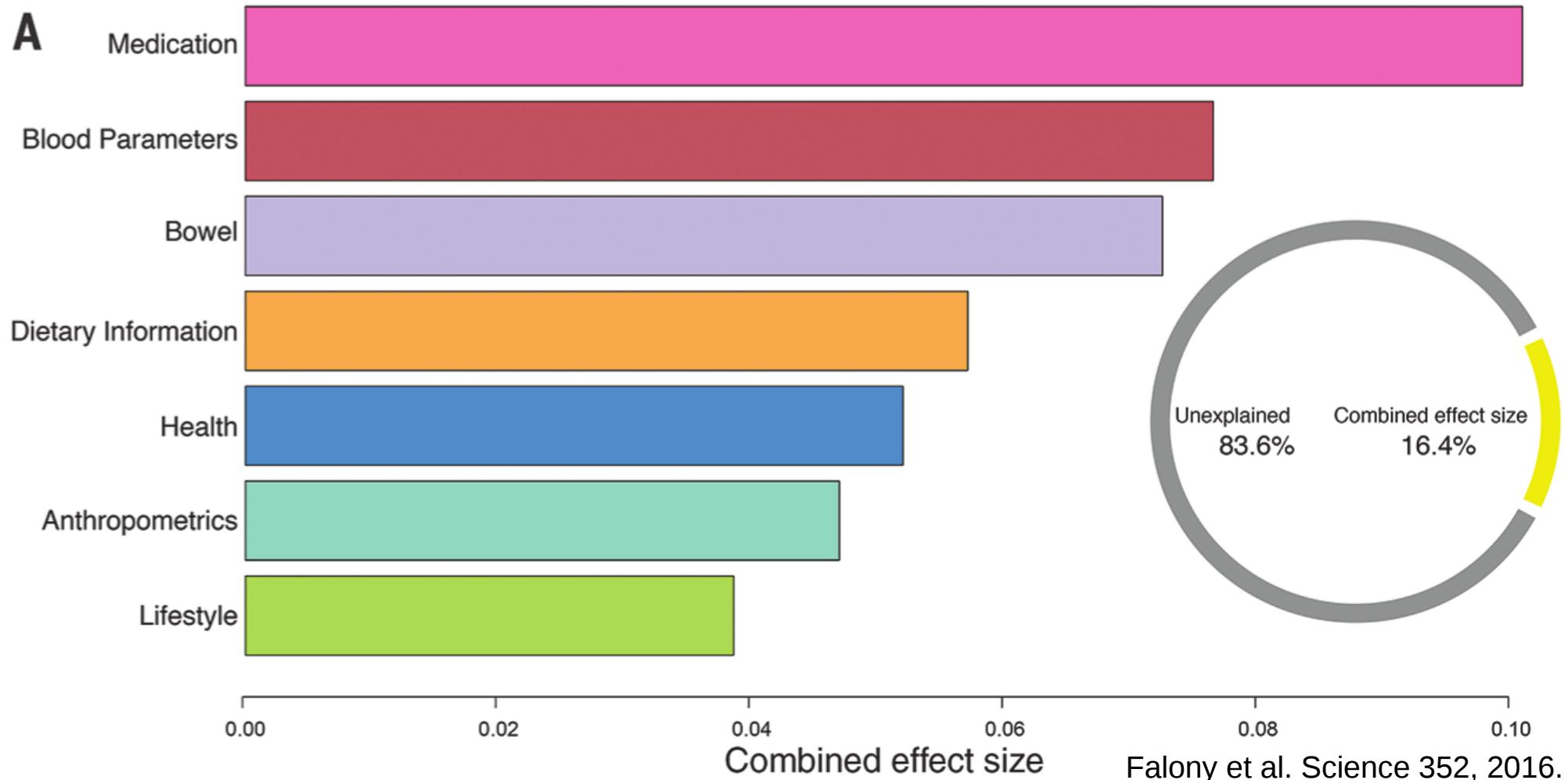
92% replication rate !

Falony et al. Science 352, 2016.

# Total explained variation: 16.4%

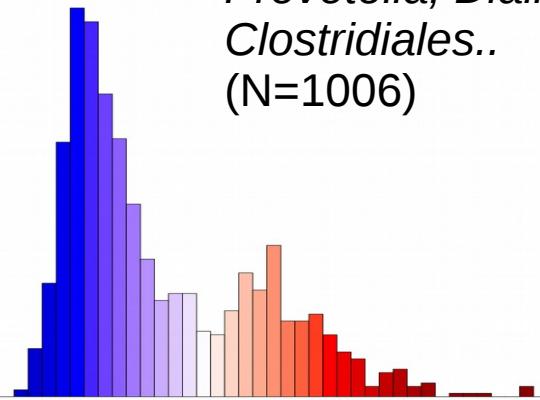
## (Flemish Gut Flora Project)

Proposed disease marker genera associated to host covariates and medication - inclusion in study design is essential !

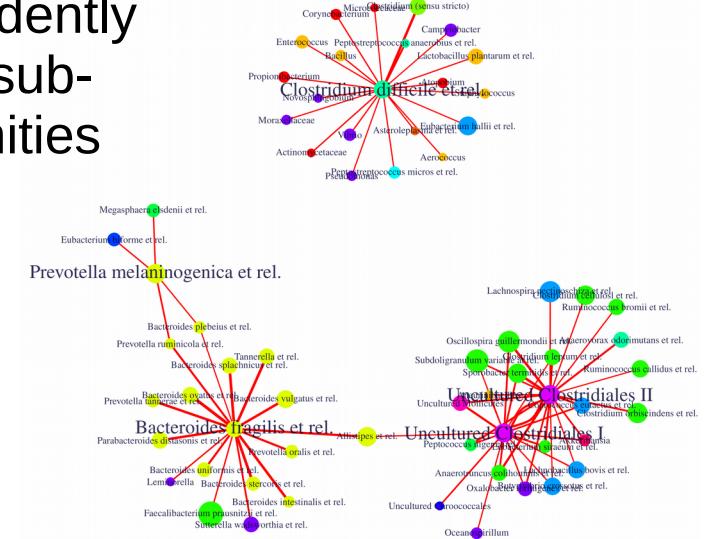


# Subject stratification into community types could reduce complexity

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

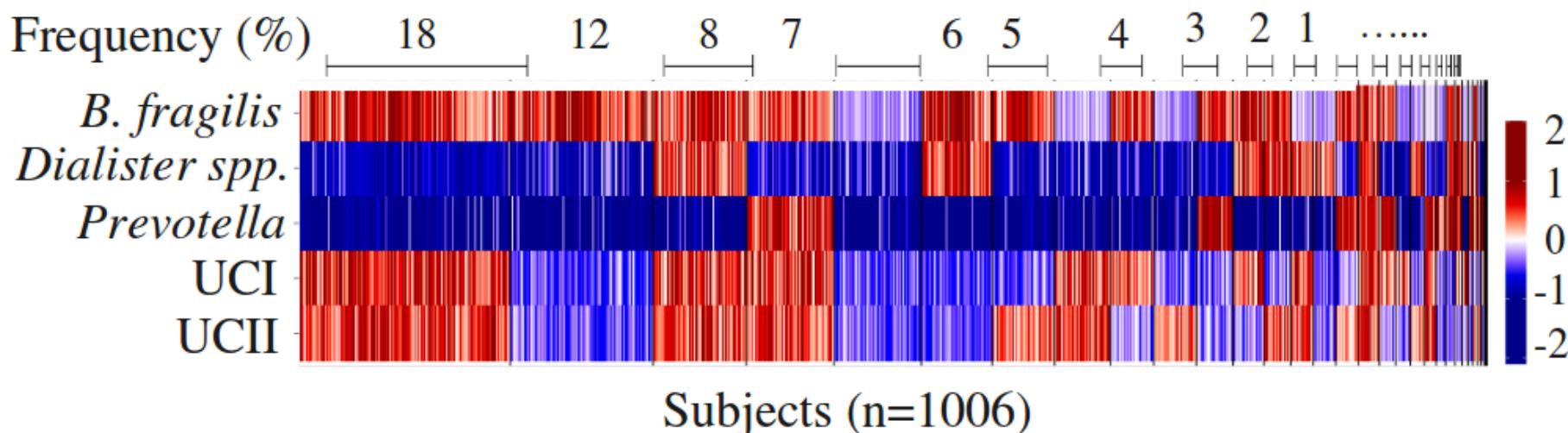


Independently varying sub-communities

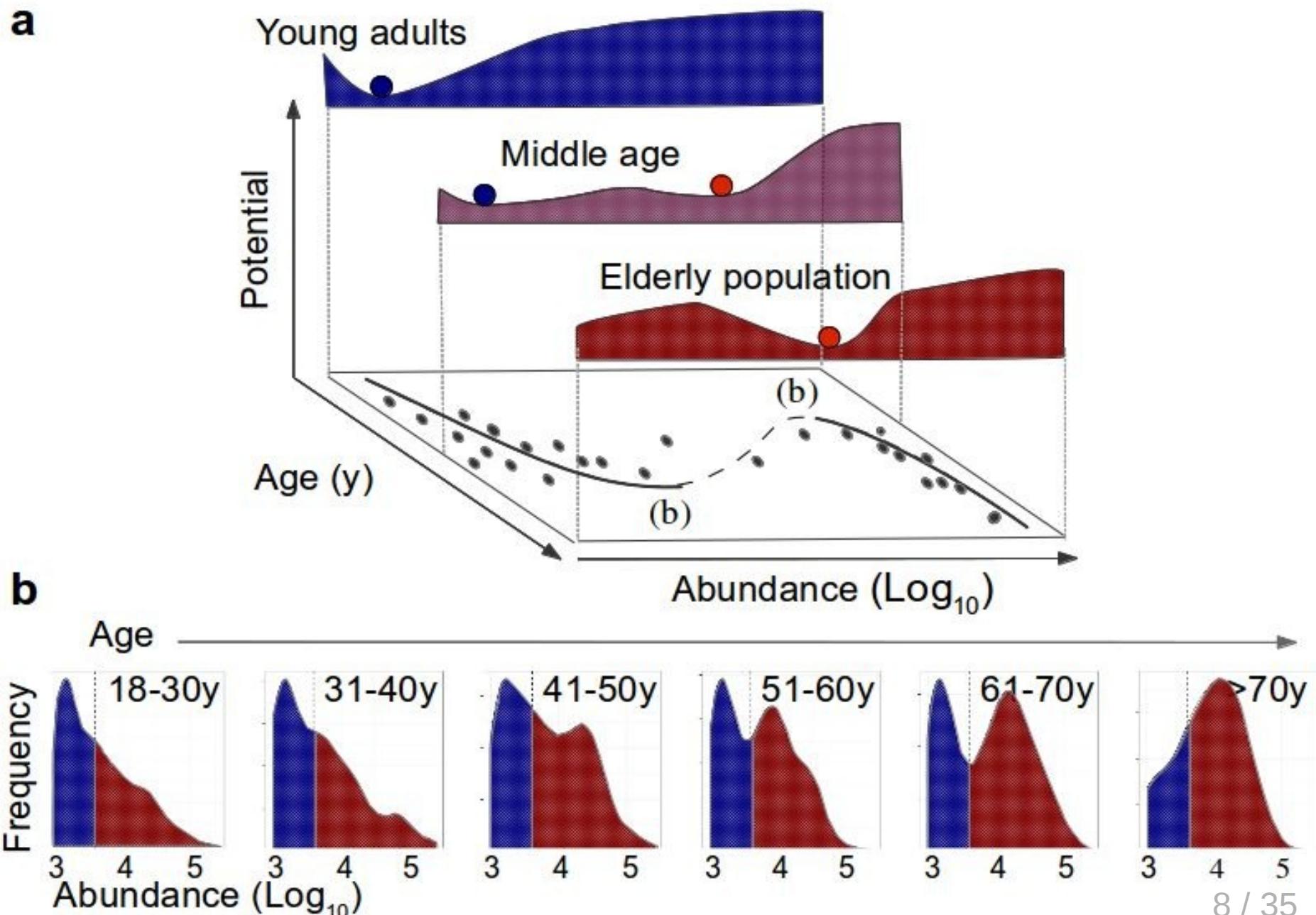


Ecosystem states are rich combinations

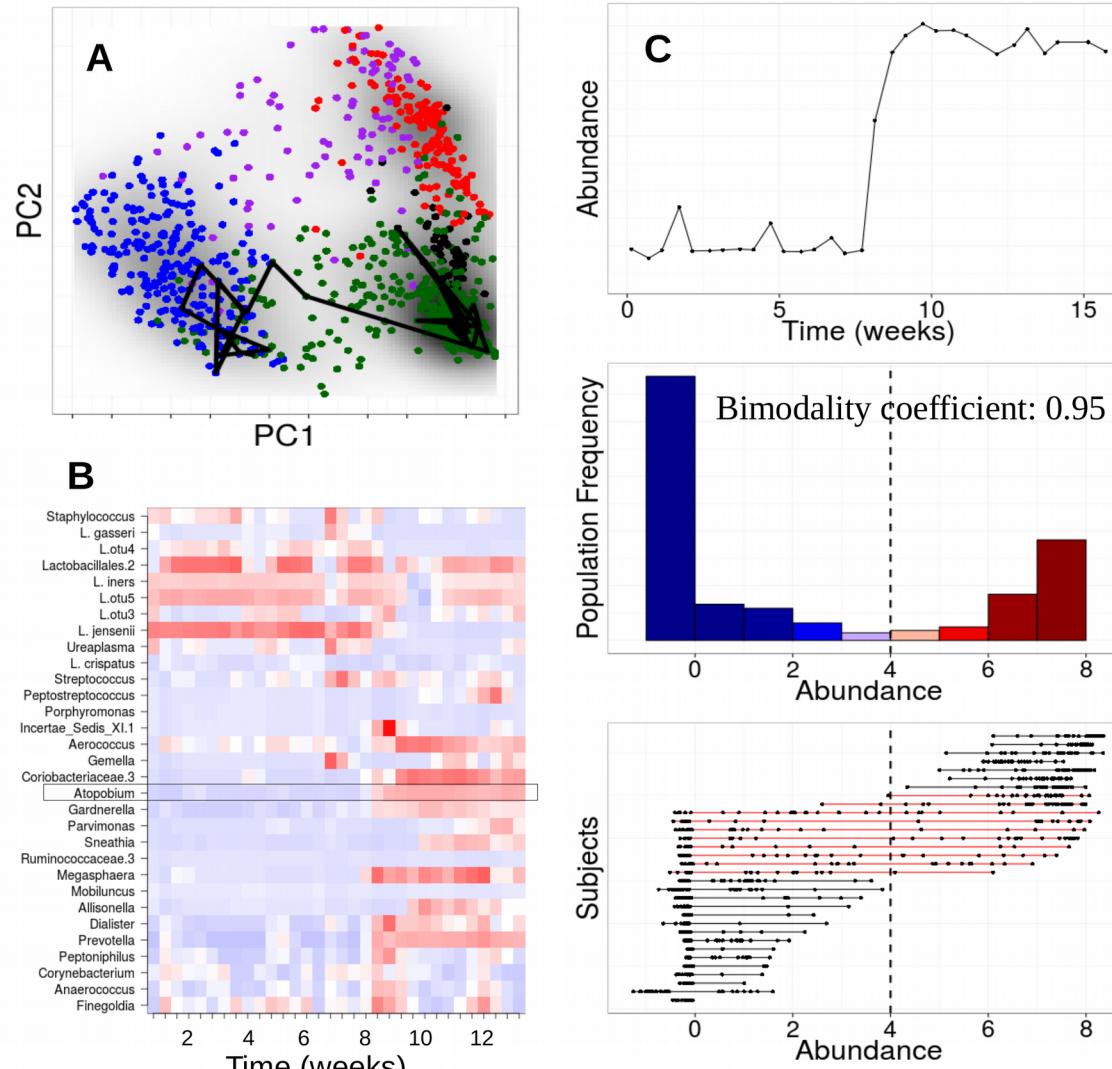
of independent tipping elements ? Lahti et al. Nat. Comm. 2014



# Cross-sectional analysis suggests a lifetime trend: state shift in Uncultured Clostridiales



# Cross-sectional vs. longitudinal analyses: community-level dynamics in vaginal microbiota is supported by pooling limited longitudinal data from multiple individuals

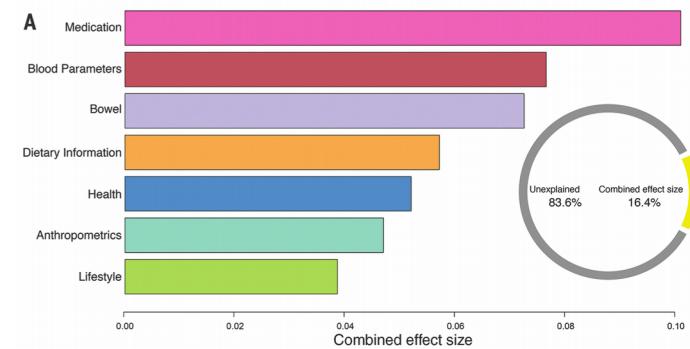
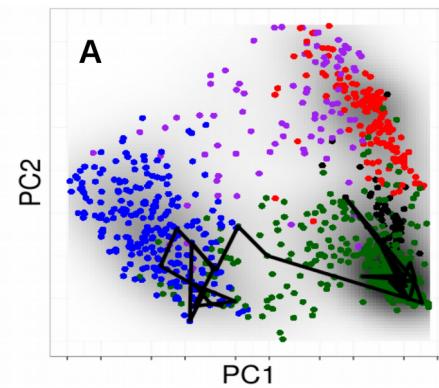
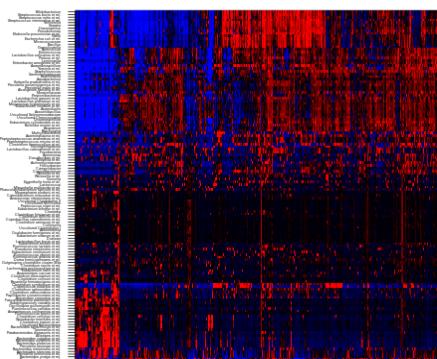


Data: Gajer et al. 2012

Figure: Faust et al. Curr. Op. Microbiol. 2015

# Summary

- Individuality and temporal dynamics better understood in the context of larger populations
- Majority (>80%) of total variation still to be explained
- Standardization, study design and confounder analysis remain essential yet easily overlooked
- Open data & algorithms can advance research



# Thank You !

Anne Salonen  
Jarkko Salojärvi  
Karoline Faust  
Christopher Quince  
Jarmo Ritari  
Jonna Jalanka  
Philippe Puylaert  
Stephen O'Keefe  
Erwin Zoetendal  
Marten Scheffer  
Jeroen Raes  
Willem M de Vos

Univ. Turku, Dpt.  
Mathematics & Statistics

Univ. Helsinki, Dpt.  
Veterinary Biosciences

Wageningen University, Lab.  
Of Microbiology, Netherlands

VIB/KU Leuven, Center for  
Biol. Of Disease, Belgium

## Welcome to poster #46 !

<http://microbiome.github.io>

Population-level analysis of gut microbiome variation. Falony et al. *Science* 352(6285):560-4, 2016

Metagenomics meets time series analysis: unraveling microbial community dynamics. Faust, Lahti et al. *Current Opinion in Microbiology* 15:56-66 2015

Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database Ritari et al. *BMC Genomics* 16:1056, 2015.

Fat, Fiber and Cancer Risk in African, Americans and Rural Africans. O'Keefe et al. *Nat. Comm.* 6:6342, 2015

Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. Salonen et al. *ISME Journal* 8:2218-30, 2014

Tipping elements in the human intestinal ecosystem. Lahti et al. *Nat. Comm.* 5:4344, 2014



# Data processing & interpretation

- Dedicated HITdb for gut 16S rRNA outperforms generic ones (Silva, Greengenes)
- Improved sensitivity and coverage in gut profiling
- Other body sites to be done
- Also in light-weight OTU clustering pipeline Lotus:  
<http://psbweb05.psb.ugent.be/lotus/>

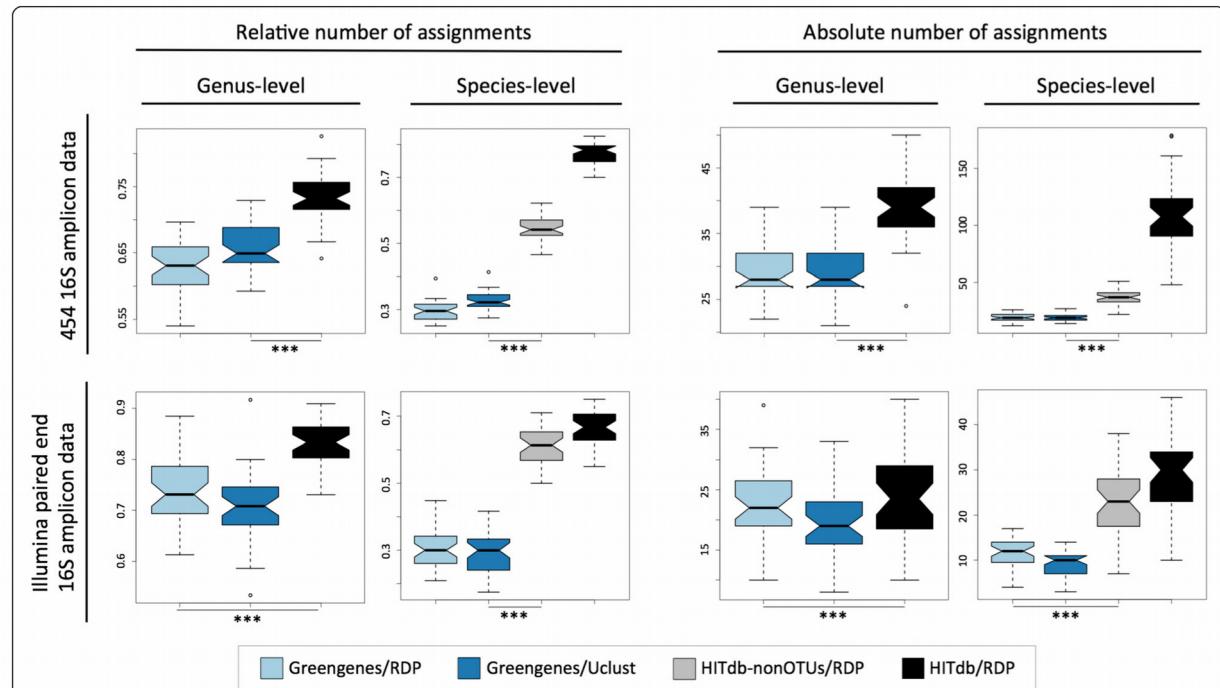
## RESEARCH ARTICLE

## Open Access



# Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database

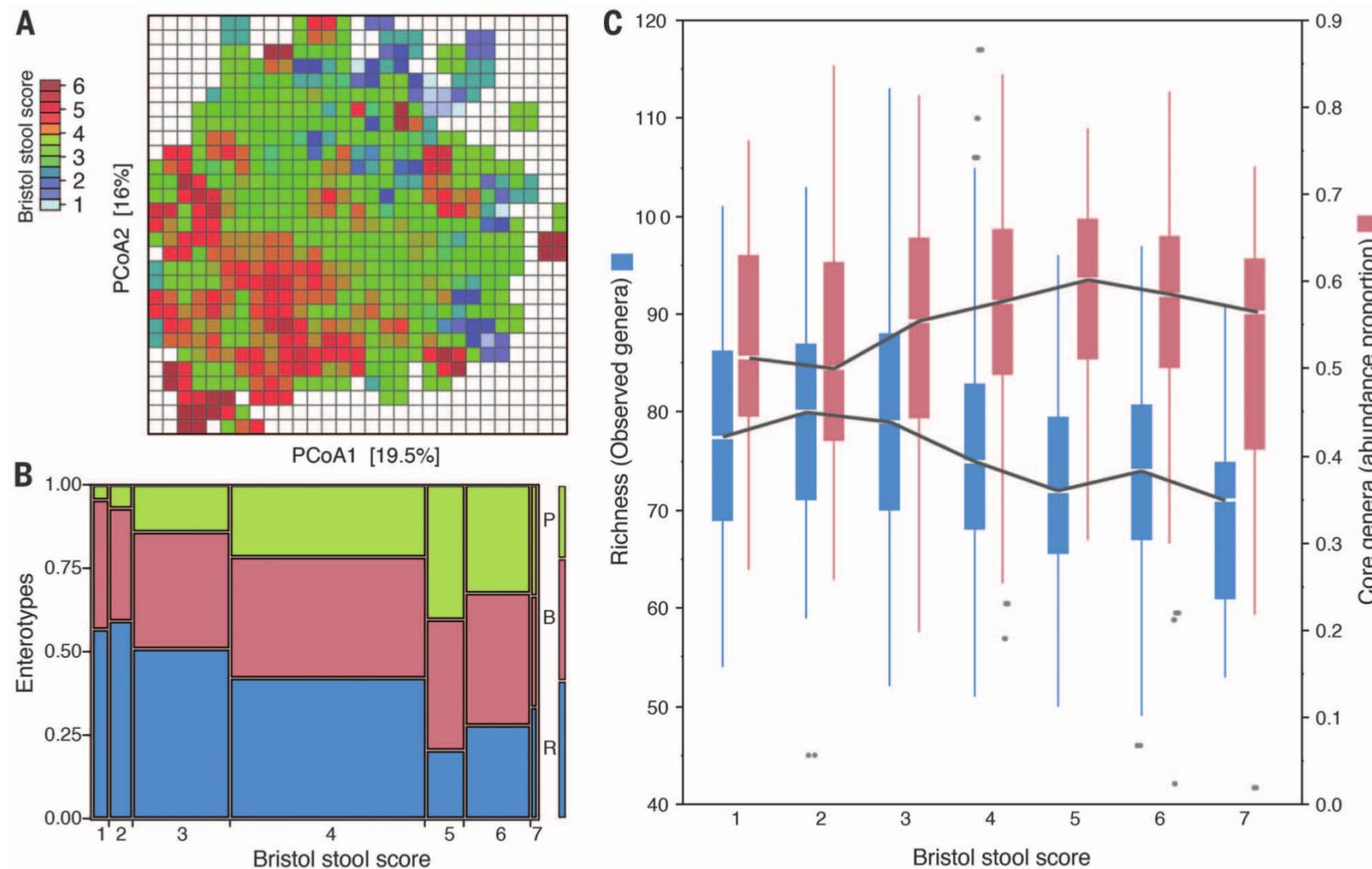
Jarmo Ritari<sup>1\*</sup>, Jarkko Salojärvi<sup>1</sup>, Leo Lahti<sup>1,2</sup> and Willem M. de Vos<sup>1,2,3</sup>



**Fig. 4** Comparison between Greengenes and HITdb using data from biological samples. The used data set is indicated in rows, and relative and absolute numbers of assignments at genus and species levels in columns. At species level, the results for HITdb additionally show the biological species only (i.e. without OTUs) for easier comparison with Greengenes. \*\*\* $p < 0.001$ .  $n = 119$  and  $n = 40$  samples for 454 and Illumina data sets, respectively

# Confounding variables: stool consistency showed the largest effect size on microbiota variation

RESEARCH | RESEARCH ARTICLES



**Fig. 4. BSS score association to microbiota variation.** (A) BSS score variation across the FGFP cohort, as represented on the genus-level PCoA ordination (Bray-Curtis dissimilarity). Each cell is colored according to median BSS score of individual samples allocated to the cell coordinates. (B) Enterotype distribution over BSS scores [JSD enterotyping (18)] showing an increase in *Prevotella* individuals with looser stool consistency. (C) Median differences in abundance of the core microbiota (FGFP genus-level core at 99%) and in observed genus richness across BSS score.

## Opportunities

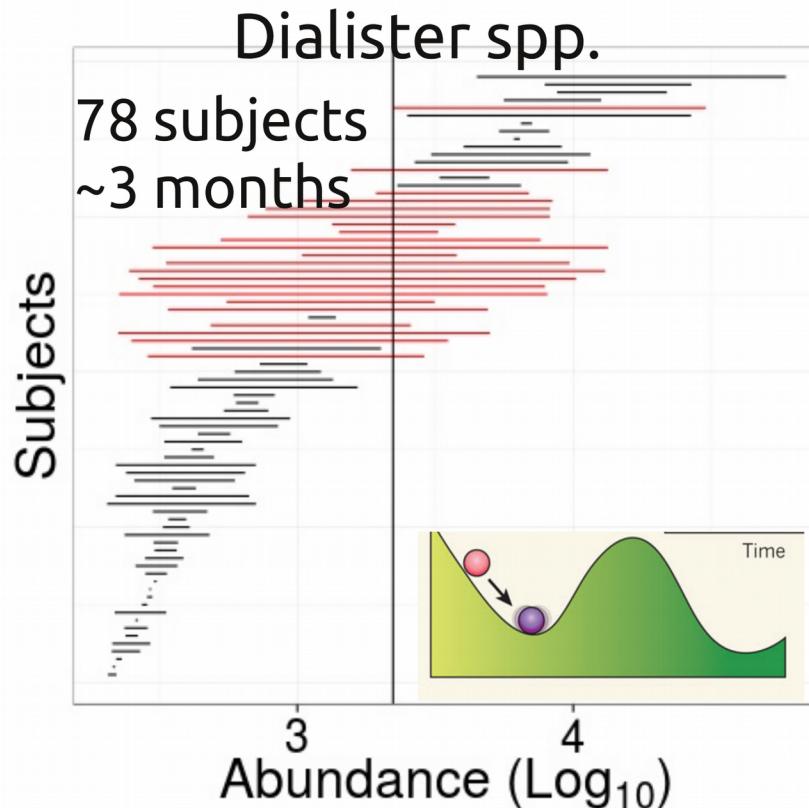
- Increasing statistical power and sensitivity
- Comprehensive population coverage & subject stratification
- Understanding individuality in the larger context

## Challenges

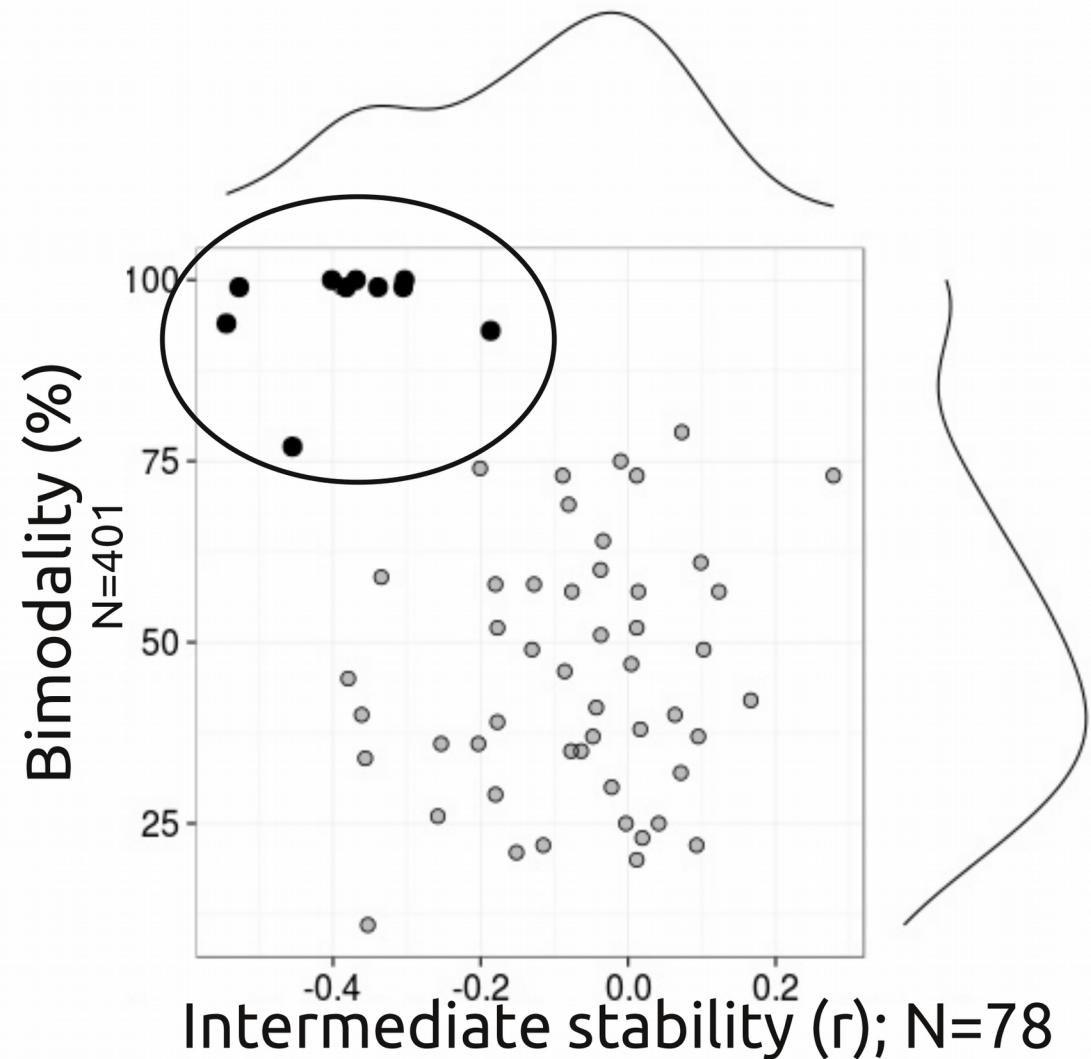
- High individuality and variation
- Data availability
- Lack of standardization & custom tools

# Confirming bi-stability with temporal analysis

-> alternative attractors divided by a critical unstable tipping point



State transitions indicate instability in intermediate abundance range

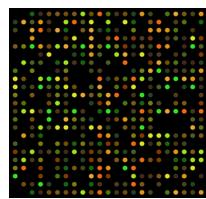


# Microbiome profiling

## Nucleic Acid – Based Approaches

### Metagenome, Metatranscriptome & Taxonomic composition with 16S rRNA Gene Analysis

#### Microarrays (HITChip Atlas)



- Known targets
- Cost-efficient
- Standardized & reproducible
- Semi-quantitative ('relative')
- Deep phylotype level profiling

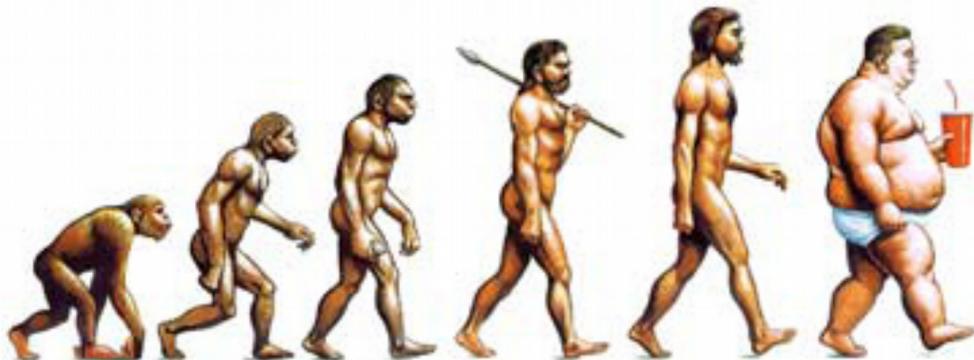
#### Sequencing (Flemish Gut Flora)



- De novo discovery
- More expensive
- Many platforms and protocols
- Quantitative ('absolute')
- Sequencing depth determines resolution

## 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 ?

**nature**  
**COMMUNICATIONS**

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

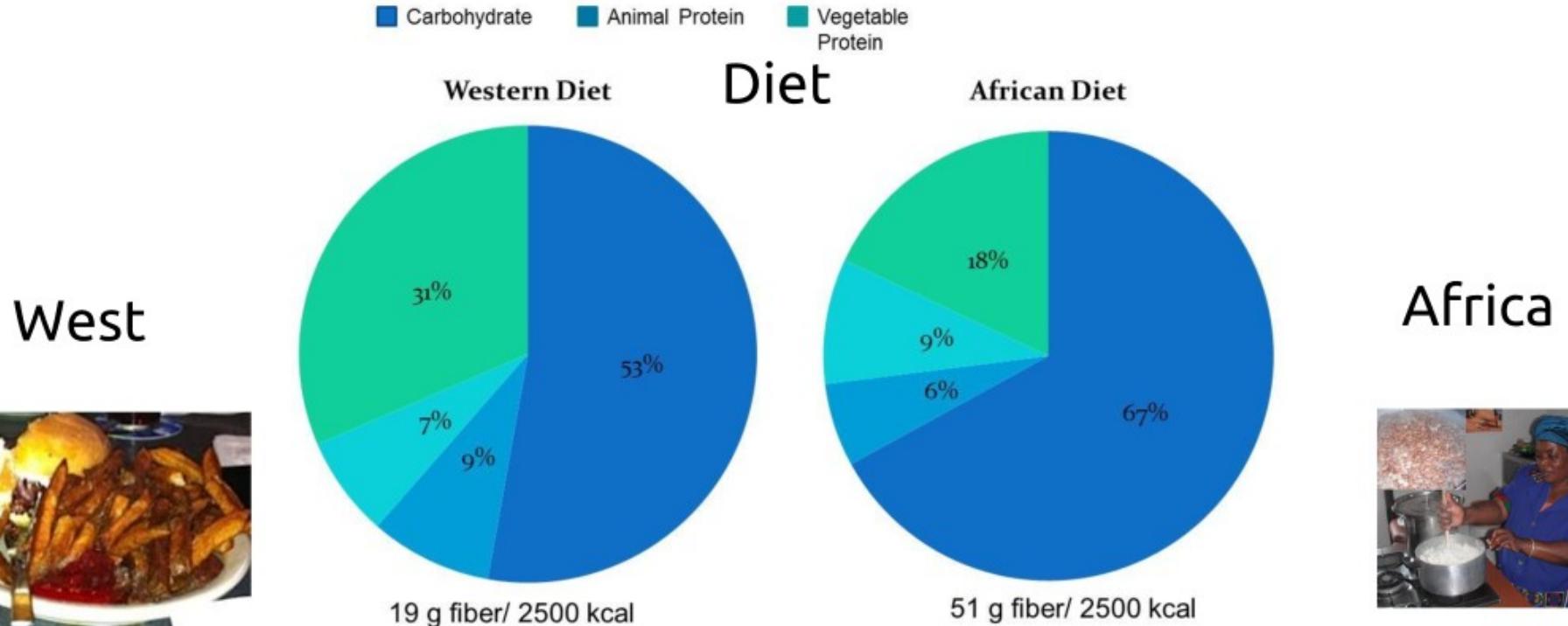


### Fat, fibre and cancer risk in African Americans and rural Africans

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 <sup>✉</sup> et al.

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

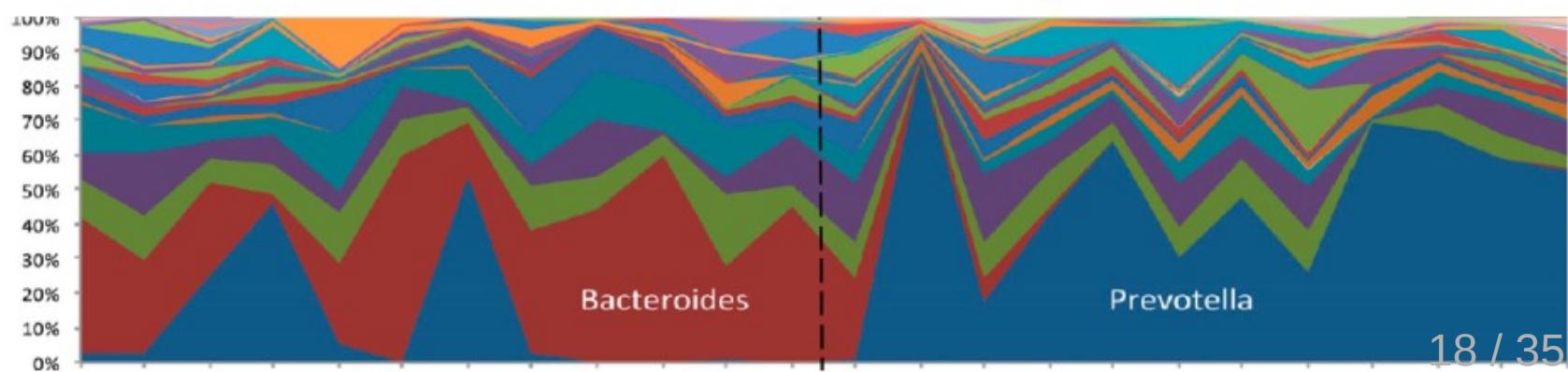
*Nature Communications* 6, Article number: 6342 | doi:10.1038/ncomms7342  
Received 23 May 2014 | Accepted 20 January 2015 | Published 28 April 2015



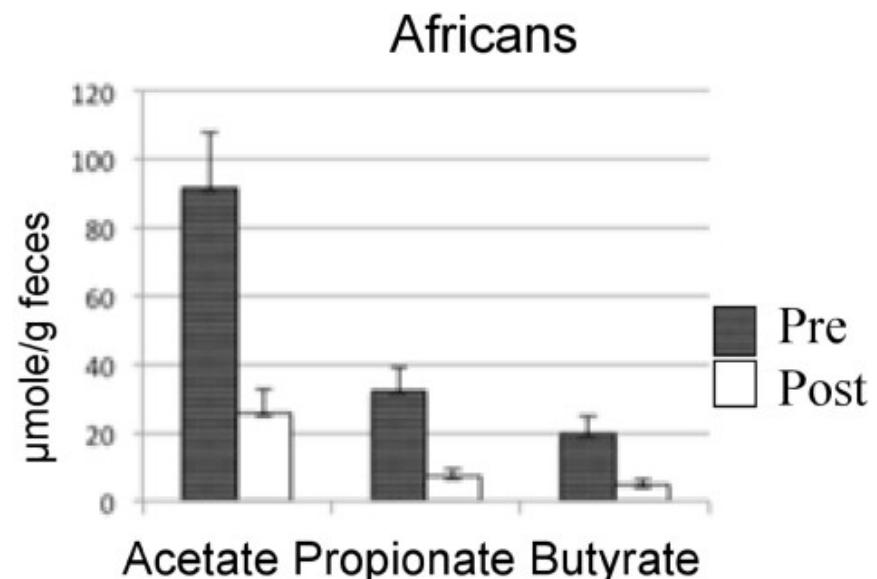
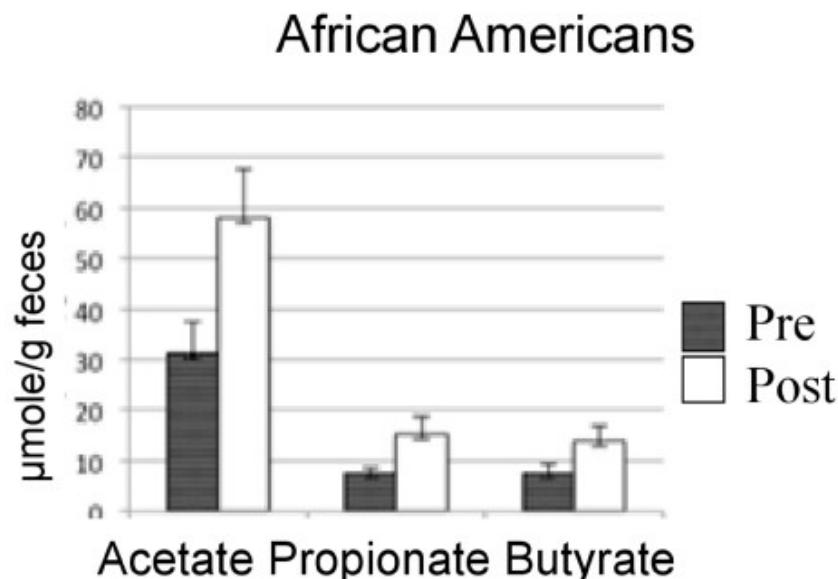
## Microbiota composition

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

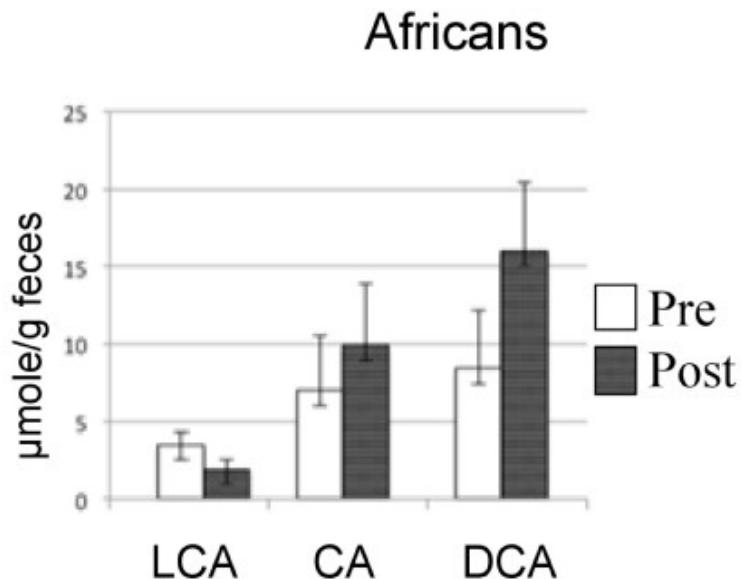
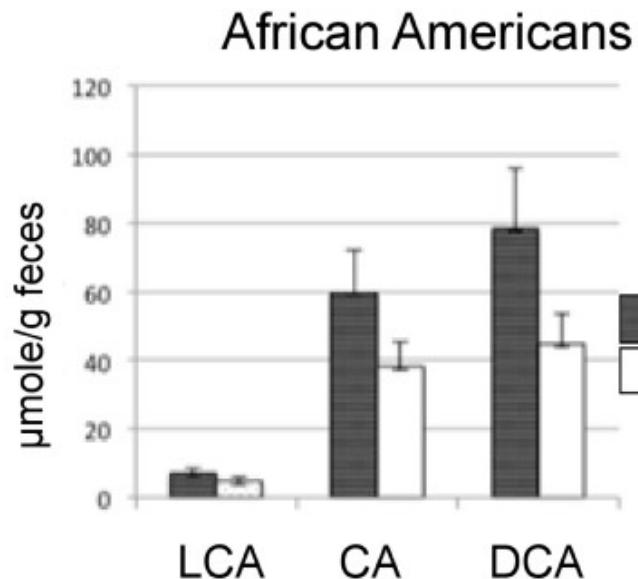


Reciprocally increased SCFA production with African diet: reported anti-inflammatory and anti-carcinogenic properties !

O'Keefe et al.  
Nat. Comm. 2015

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

# Impact diet exchange on bile acids



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

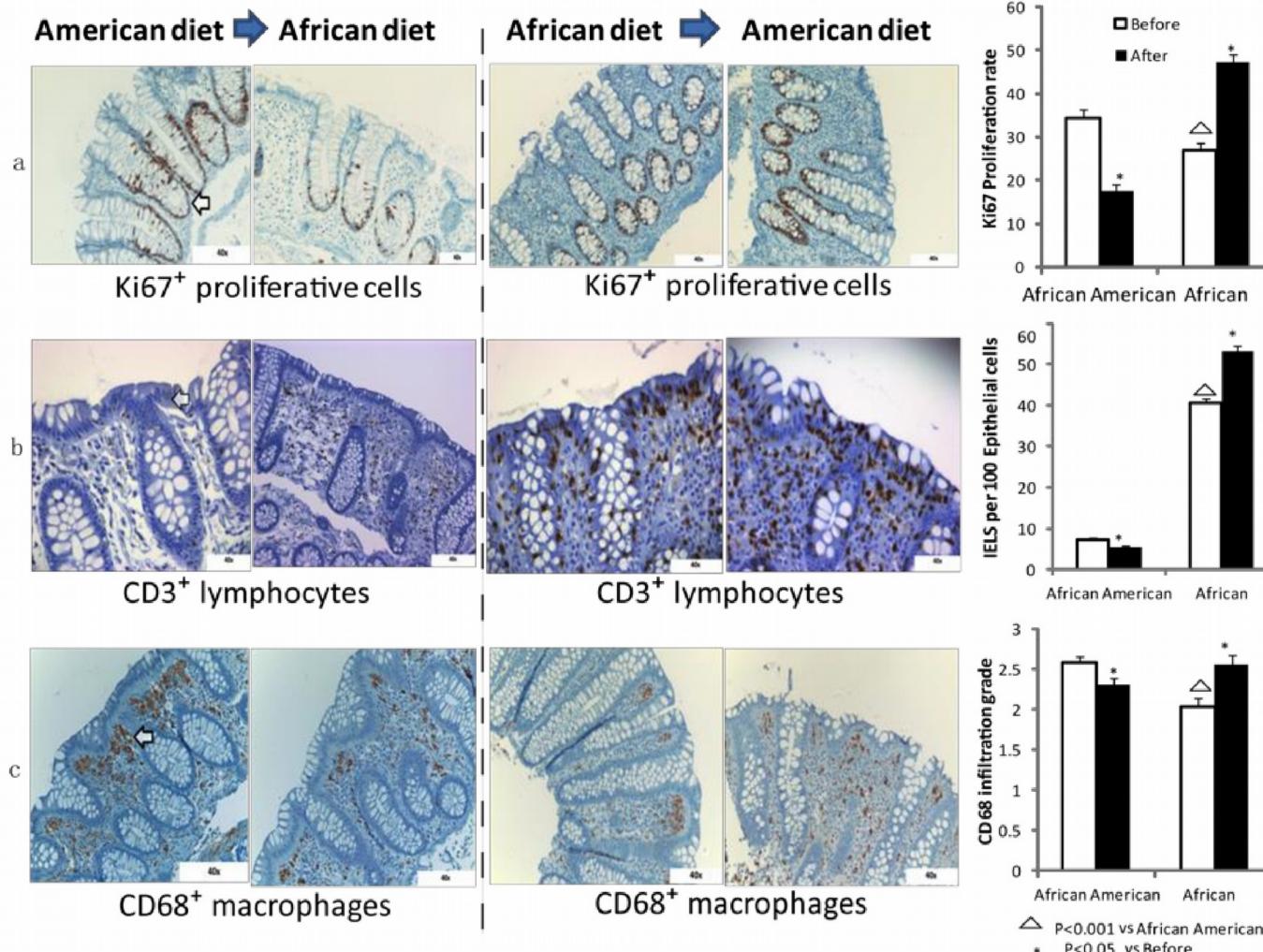
Primary and secondary bile acids reciprocally increased with American diet: reported pro-inflammatory and carcinogenic properties !

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

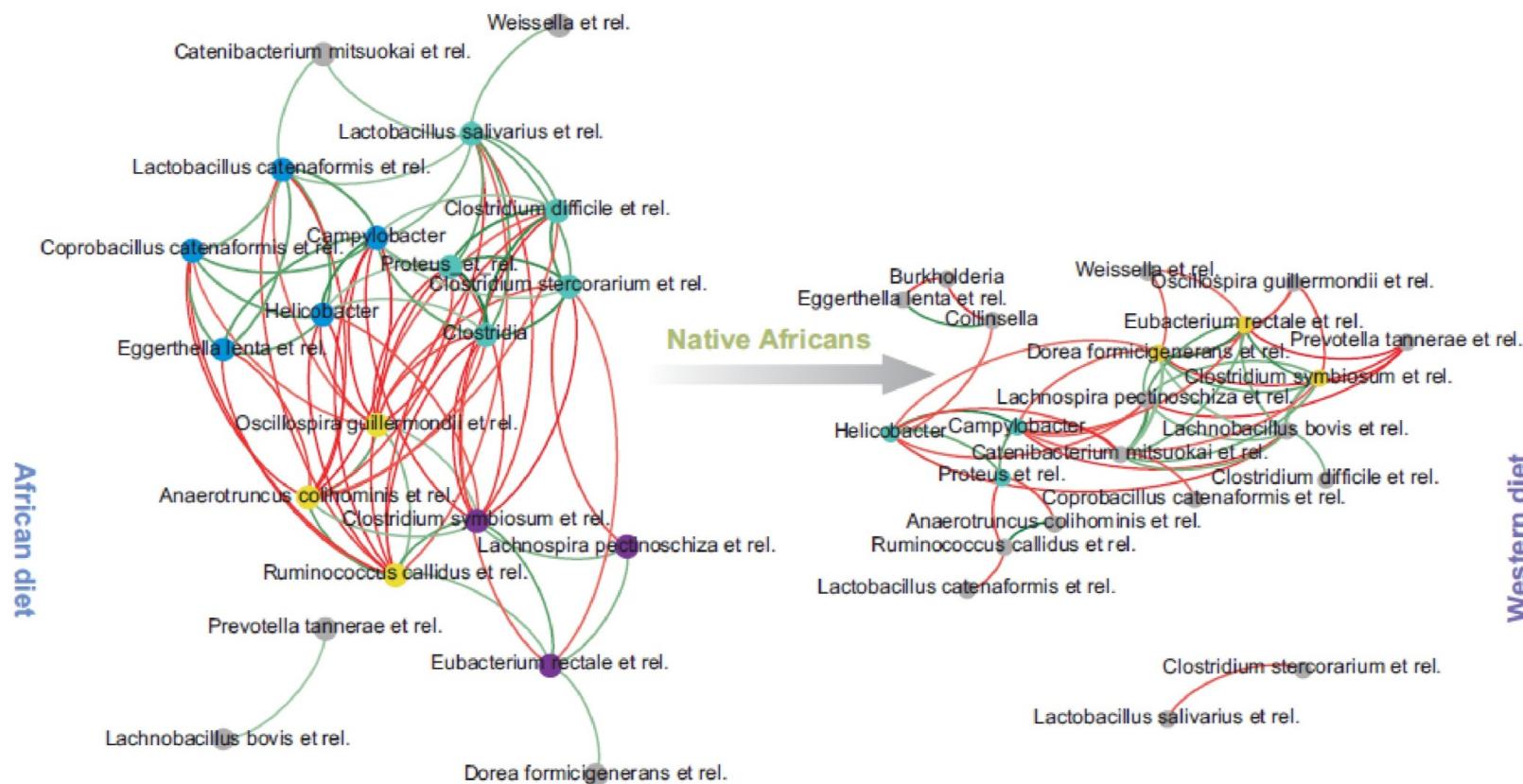
O'Keefe et al.  
Nat. Comm. 2015

# Diet swap (2 weeks) and colon cancer risk

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



# Westernization of African communities?



## Short-term diet shift

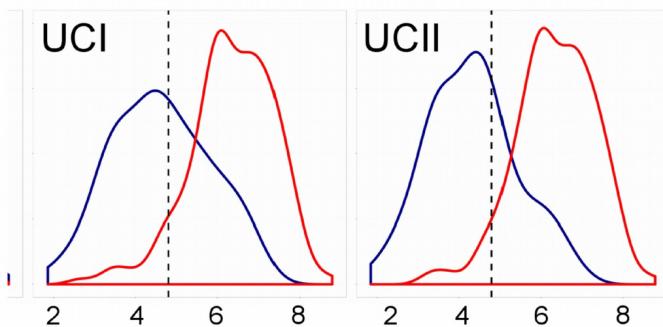
- > Limited impact on microbiota composition
- > Large impact on microbiota activity
- > Can microbiota interactions explain the differences ?

# Effect on the overall ecosystem

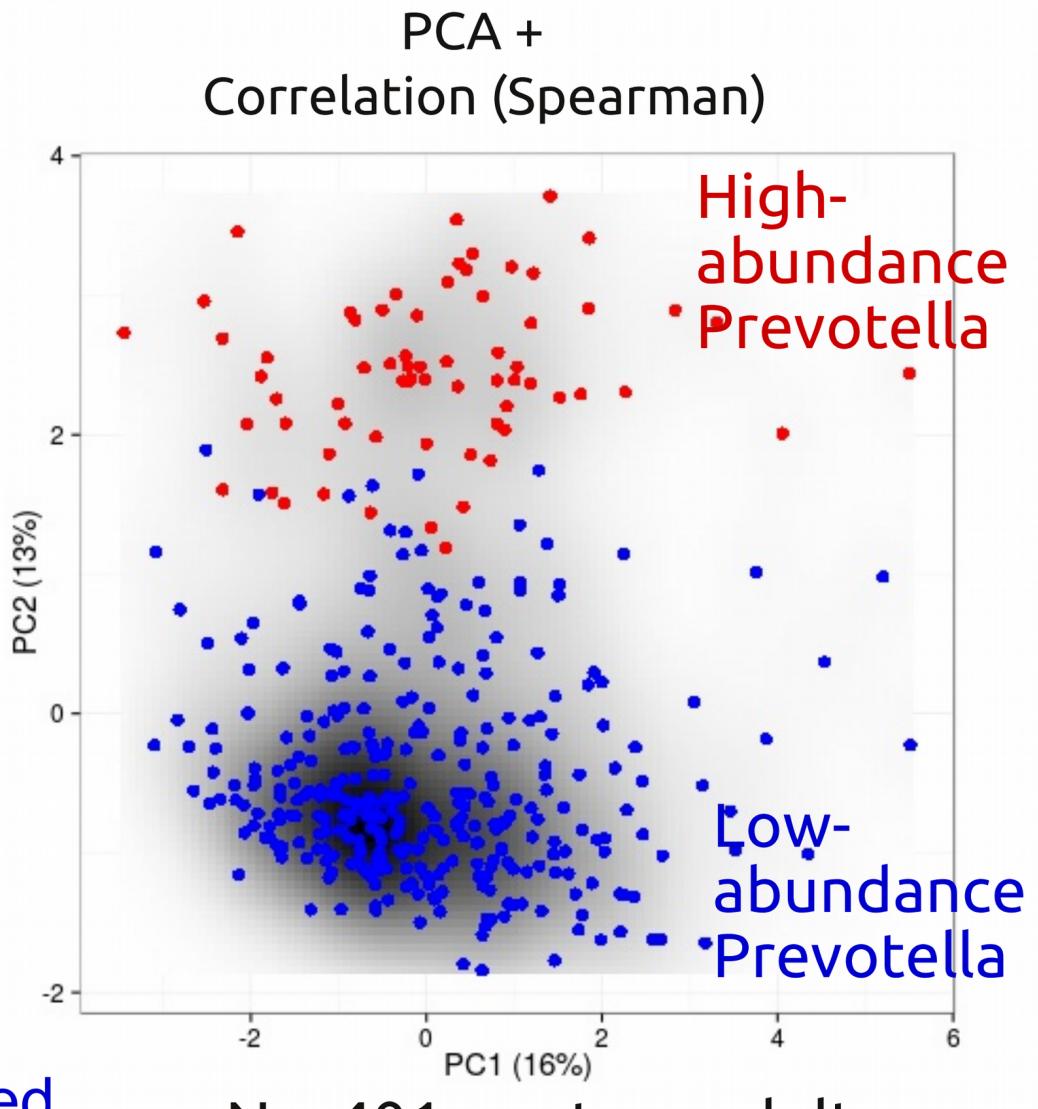
Prevotella states seen at ecosystem level

High ~10% relative abundance !

Other bi-stable taxa:  
<2% relative abundance,  
no ecosystem-level switches are visible



Uncultured Clostridiales associated with overall metagenomic richness



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



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)
- Role of bistable bacteria ?

O'Keefe et al.  
Nat. Comm. 2015

# Early warning signals to predict state shifts ?

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

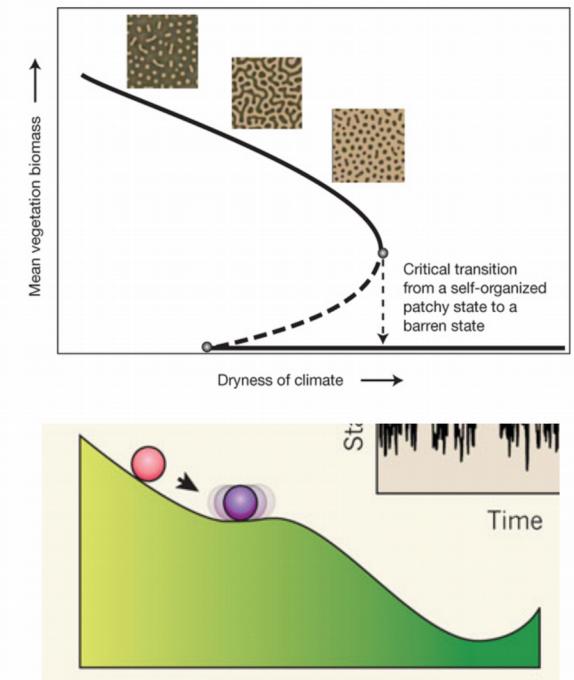
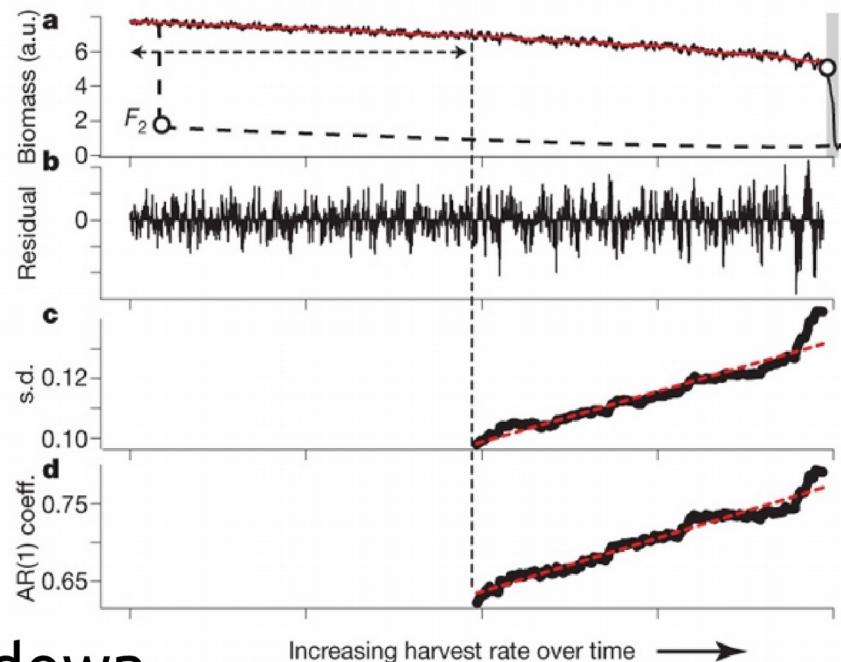
Autocorrelation

Variance

Skewness

Flickering

Critical slowing down



M Scheffer Nature 2010

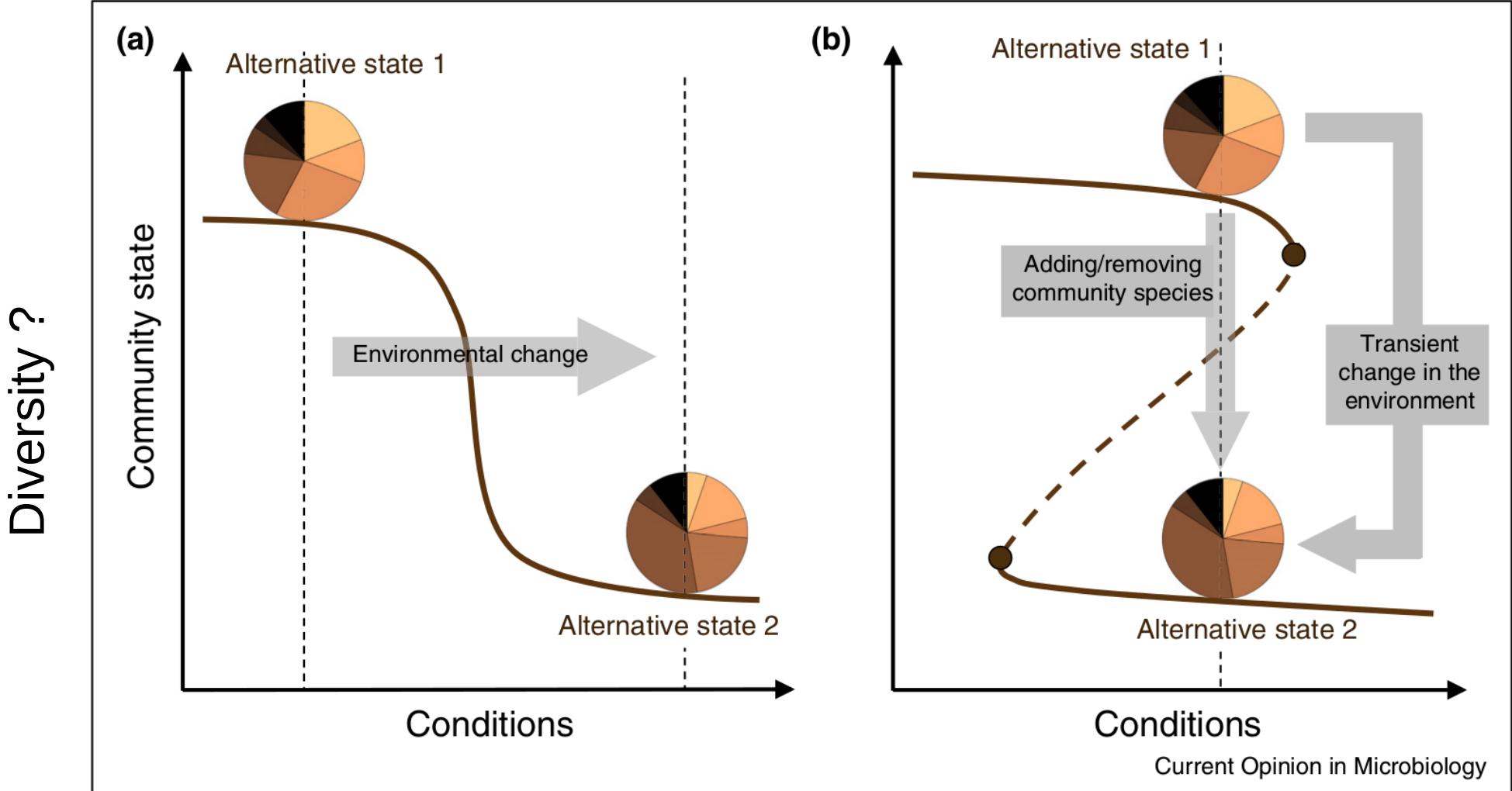
M Scheffer et al. *Nature* **461**, 53-59 (2009) doi:10.1038/nature08227

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

nature

## State induced by external factors

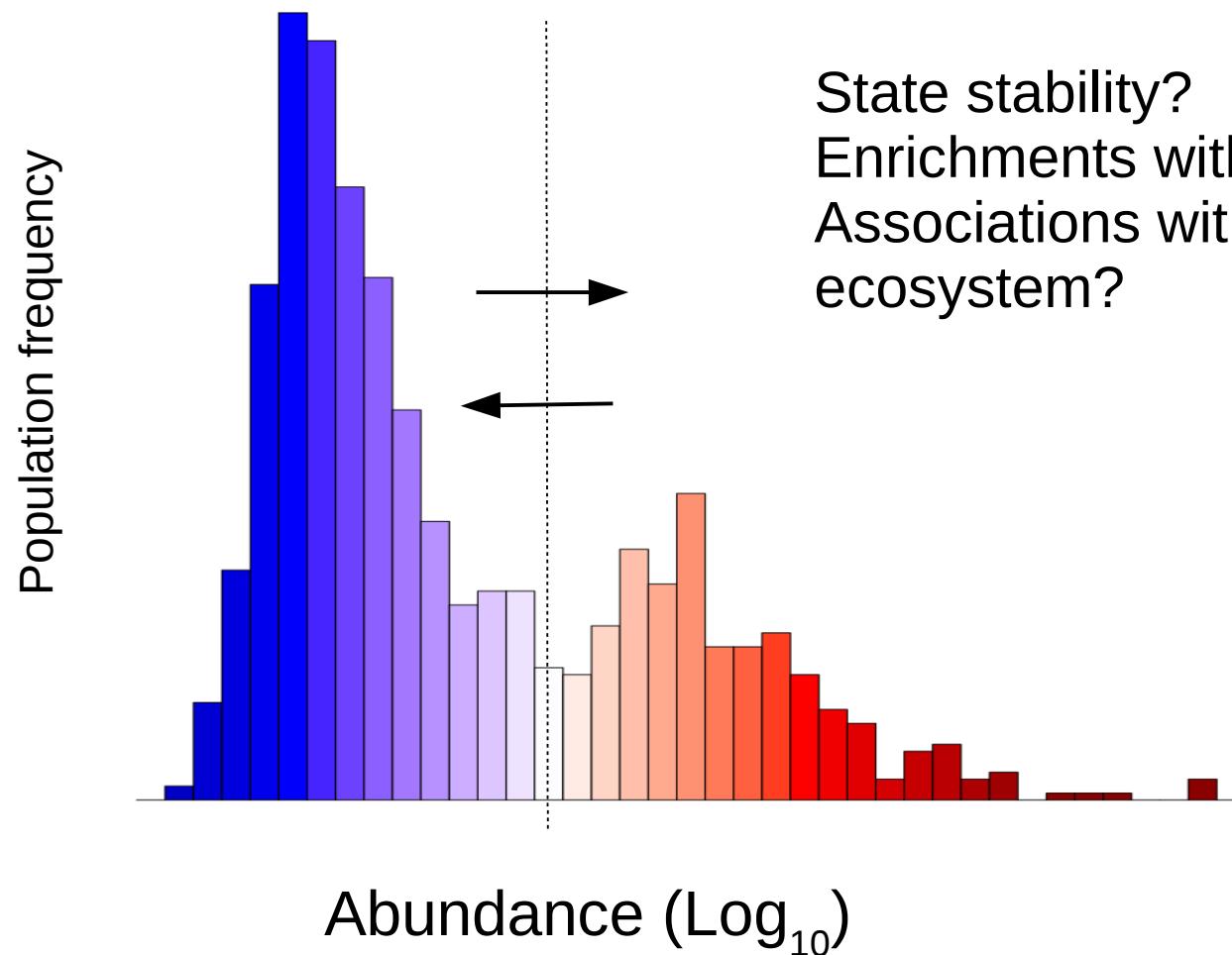
## Intrinsic stability: robust to external factors



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

# Alternative states in bacterial abundance ?

*Dialister* spp. (N = 1006)

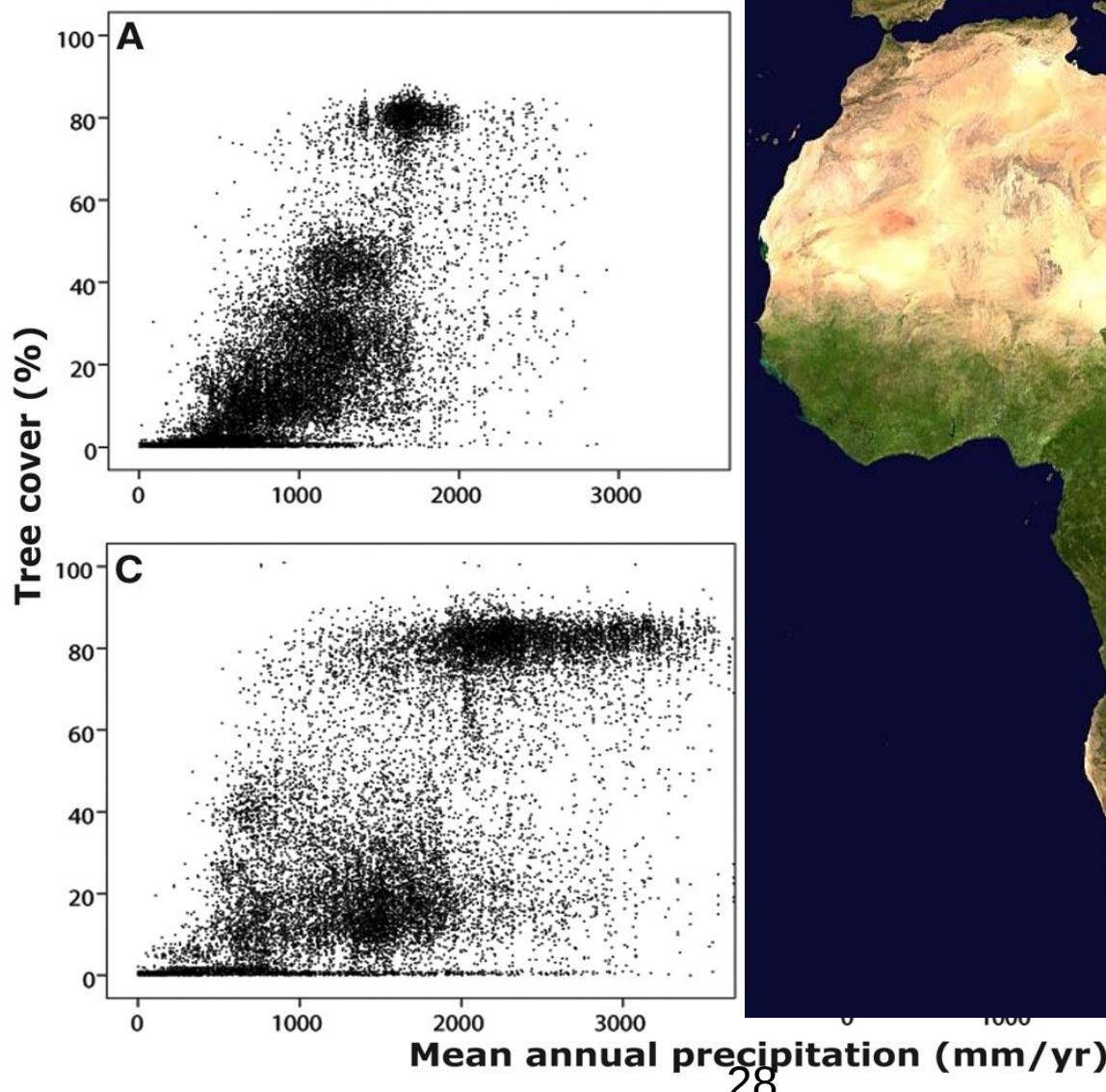


Hirota et al. Science 2011  
Livina et al. Clim Past. 2010  
Lahti et al. Nat. Comm 2014

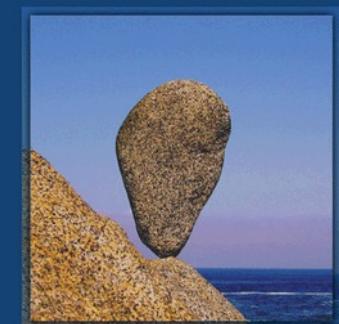
REPORT

## Global Resilience of Tropical Forest and Savanna to Critical Transitions

Marina Hirota<sup>1</sup>, Milena Holmgren<sup>2,\*</sup>, Egbert H. Van Nes<sup>1</sup>, Marten Scheffer<sup>1</sup>



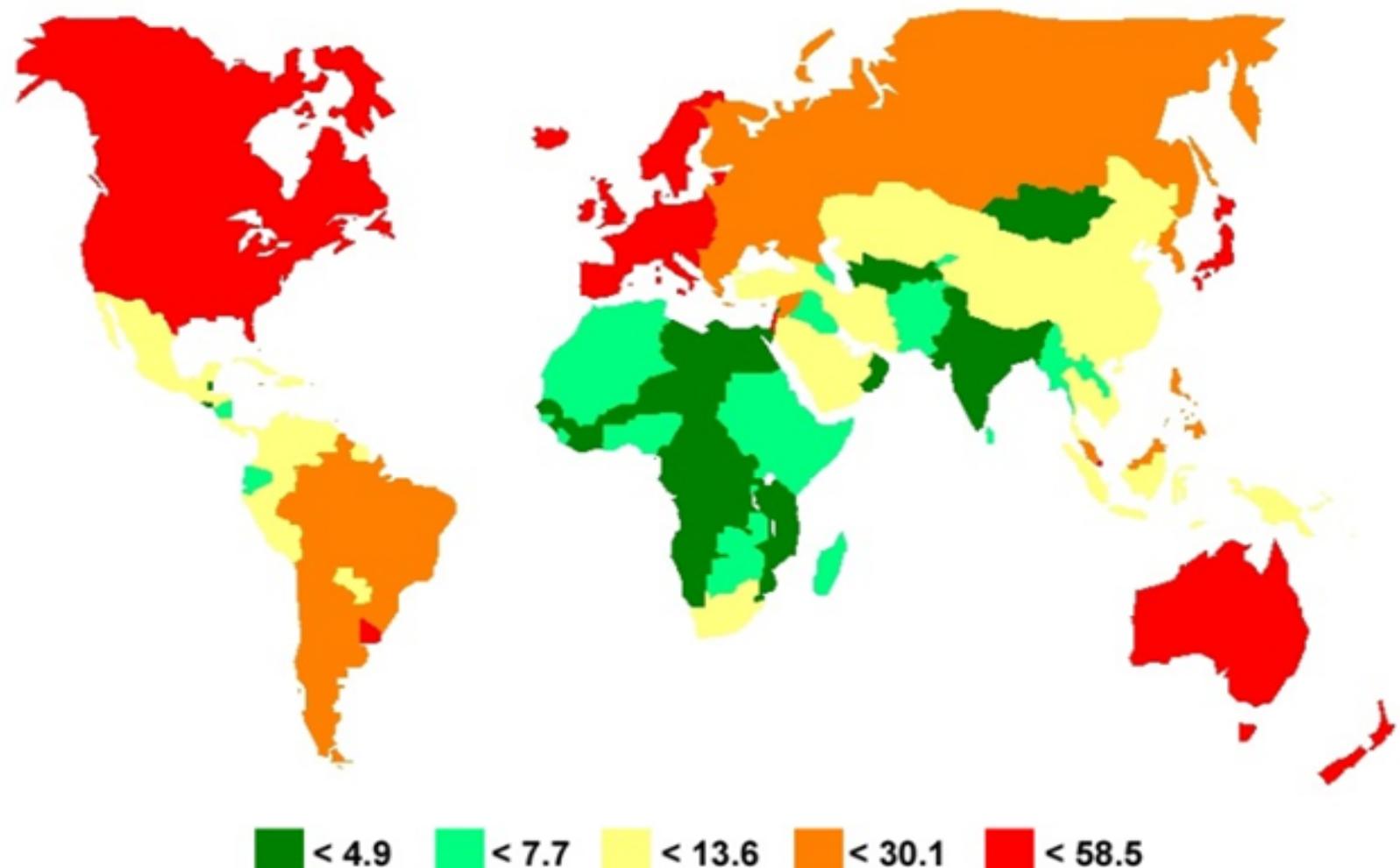
Critical Transitions  
in Nature and Society



Marten Scheffer

PRINCETON STUDIES IN COMPLEXITY

# Colon cancer prevalence



Colon cancer rates per country

## Health associations of bi-stable tipping elements

Health status	Bimodal group	Enriched state	Compromised (%)	Controls (%)	FDR (%)
Severe obesity (n=136)	UCI	Low abundance	29	55	<0.1
Severe obesity	UCII	Low abundance	38	61	<0.1
IBS (n=106)	UCII	Low abundance	50	61	1
MetS (n=66)	<i>B.fragilis</i> group	High abundance	89	78	<0.1
MetS	<i>Prevotella</i> group	Low abundance	11	22	11
MetS	<i>Dialister</i>	High abundance	36	28	13

- Causality: drivers or passengers?
- Diagnostic & prognostic indicators ?
- Targets for manipulation?
- Early warnings signals of state shift ?

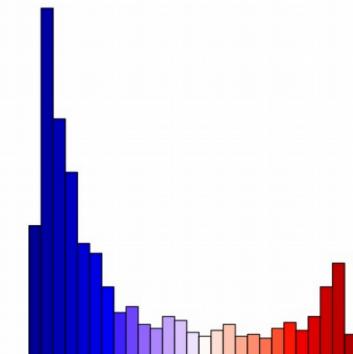
Received 23 Jan 2014 | Accepted 9 Jun 2014 | Published 8 Jul 2014

DOI: 10.1038/ncomms5344

OPEN

### Tipping elements in the human intestinal ecosystem

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



# Core microbiota, healthy western adults (n=456)

## Top-10 genus-level groups

Faecalibacterium prausnitzii et rel.  
Ruminococcus obeum et rel.  
Subdoligranulum variable at rel.  
Dorea formicigenerans et rel.  
Oscillospira guillermondi et rel.  
Lachnospira pectinoschiza et rel.  
Bifidobacterium  
Clostridium symbiosum et rel.  
Clostridium orbiscindens et rel.  
Clostridium sphenoides et rel.

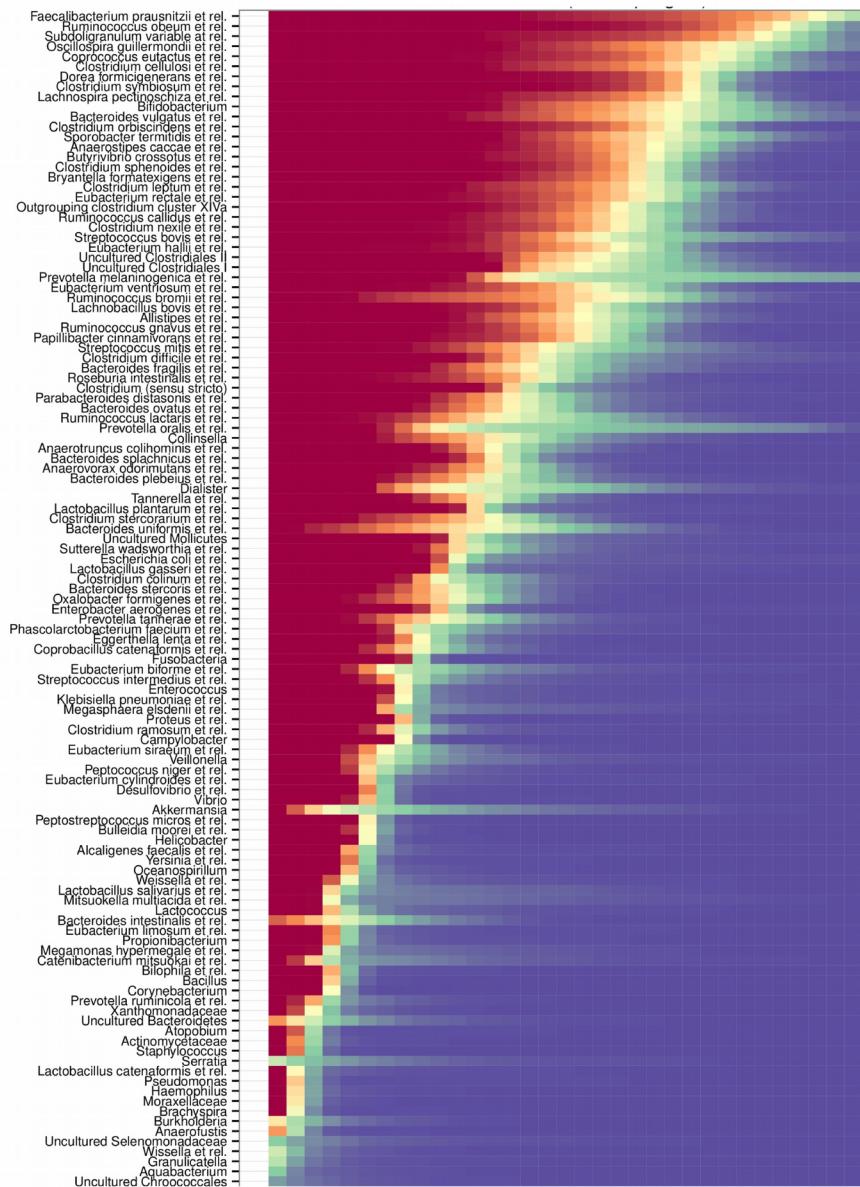
Jalanka-Tuovinen et al. (2011) PLoS One

6:e23035

Salonen et al. (2012) Clinical microbiology and  
infection 18:16–20.

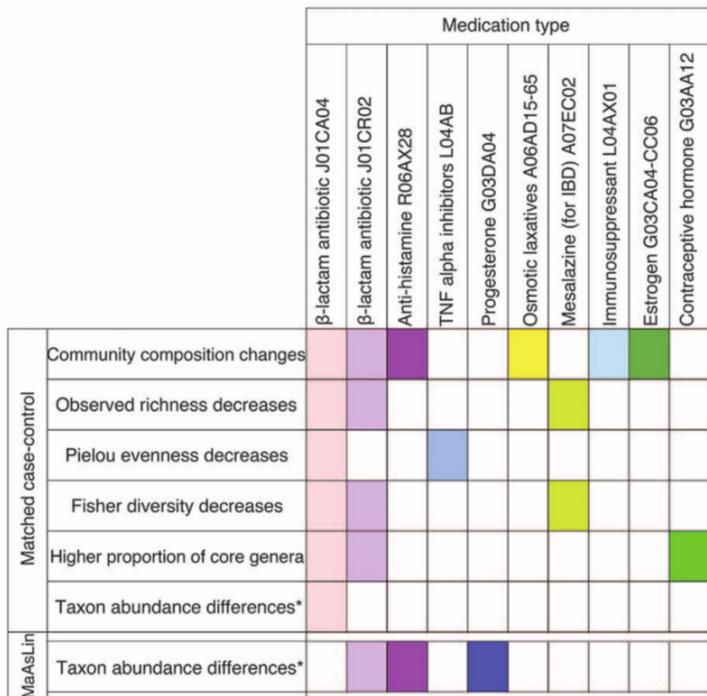


Code: [microbiome.github.com](https://github.com/microbiome)



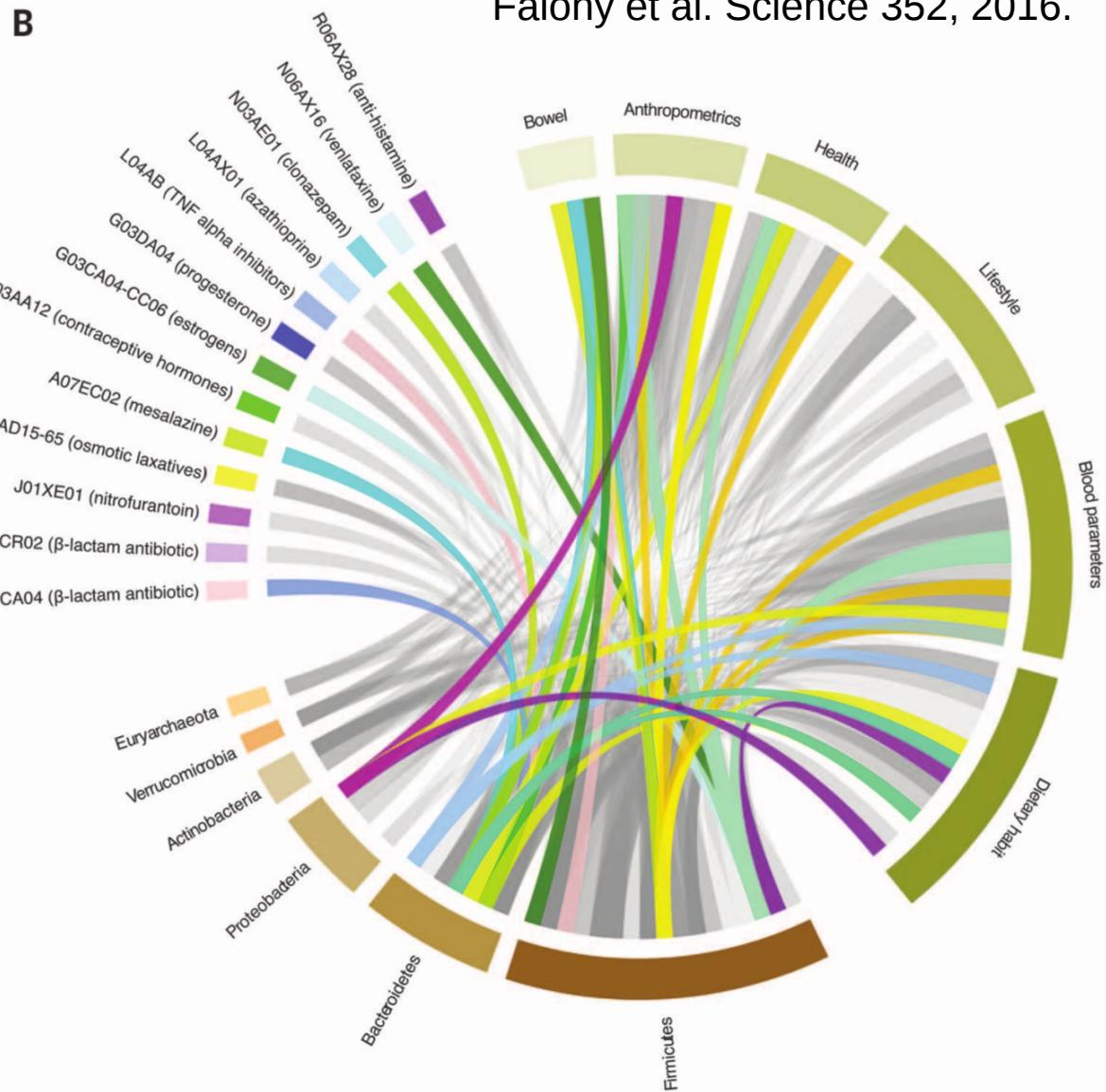
Detection threshold (abundance)

A



\*Decreased in medicated \*Increased in medicated

Butyrivibrio	Anaerostipes
Coprococcus	Bacteroides
Prevotella	Collinsella
uncl. Bdellovibrionaceae	Flavonifractor
uncl. Clostridiaceae	Parabacteroides
uncl. Clostridiales	
uncl. Desulfovibrionaceae	
uncl. Gracilibacteraceae	
uncl. Lachnospiraceae	
uncl. Prevotellaceae	
uncl. Ruminococcaceae	
uncl. Veillonellaceae	
	uncl. Enterobacteriaceae
	Eggerthella
	Flavonifractor
	Parasutterella
	uncl. Eubacteriaceae

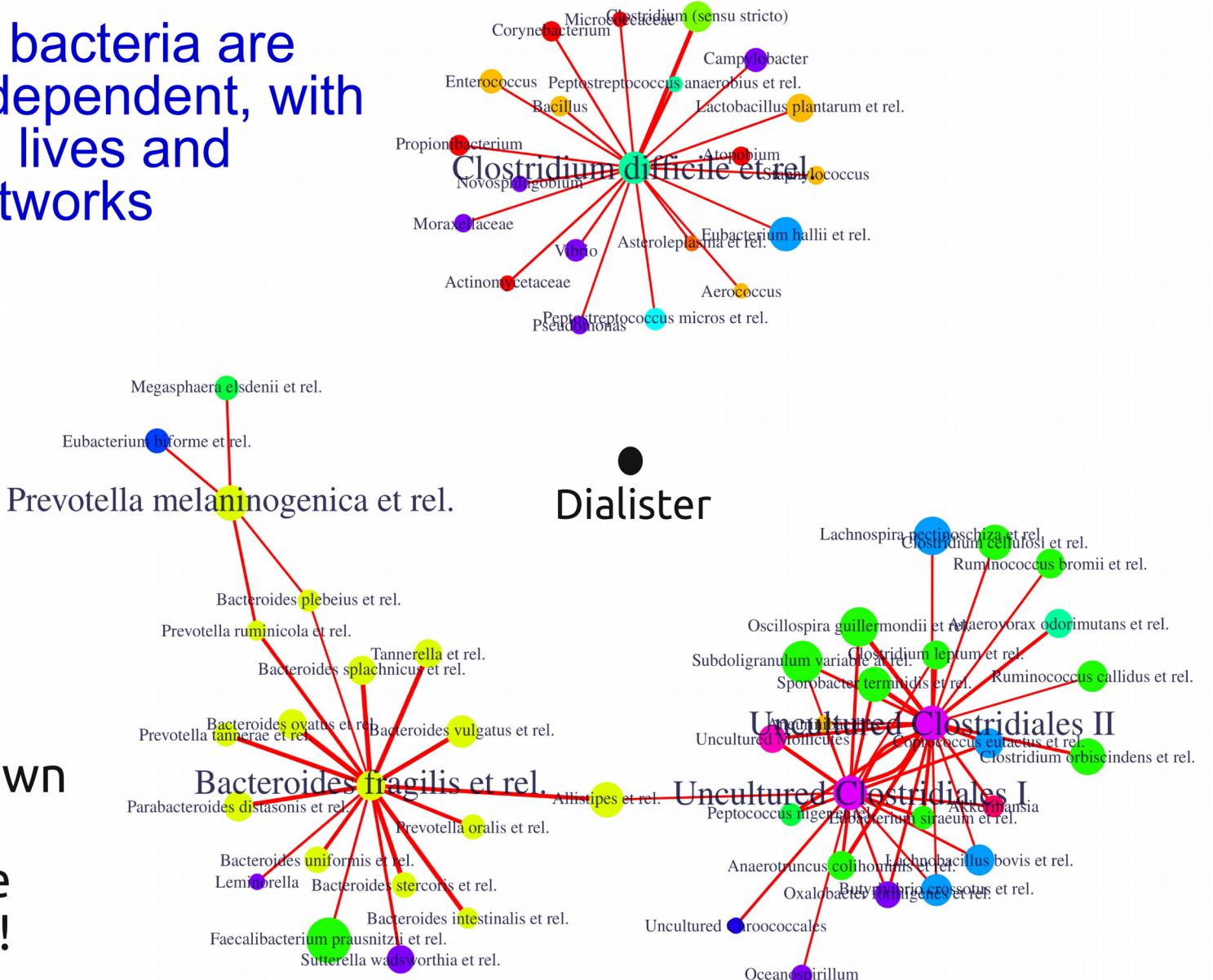


**Fig. 5. Drug interactions in the FGFP.** (A) Overview of the association between different types of medication and microbiome composition. Colored boxes (color coding according to medication) represent a significant result in the matched case-control (FDR<5%) or boosted additive general linear modeling (FDR<10%, table S11) analyses. The effect (decrease/increase) of medication on genera abundances is specified. (B) Circos plot showing correlations between covariates and genus abundances (FDR<10%) interacting with drugs. Genera are grouped at phylum level; ribbons represent genus-phenotype associations and are colored according to the confounding medication (gray indicates nonconfounded).

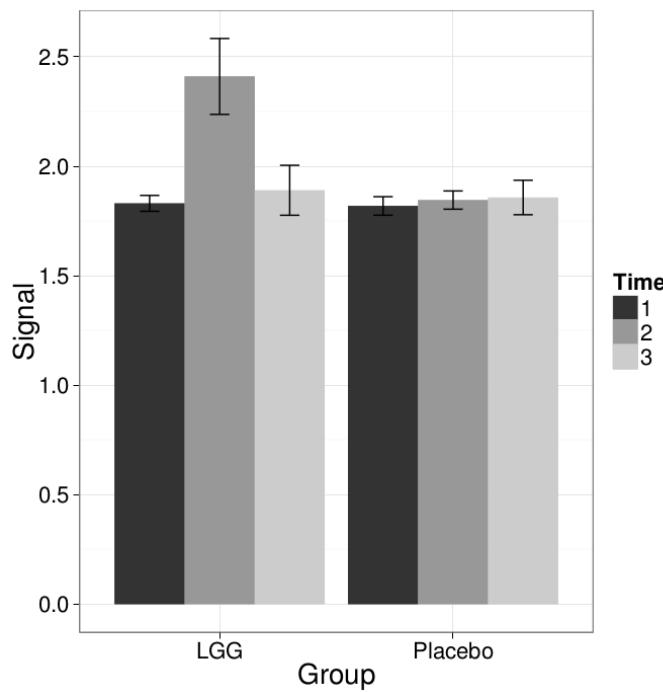
Bi-stable bacteria are rather independent, with their own lives and social networks

$|r| > 0.33$  shown

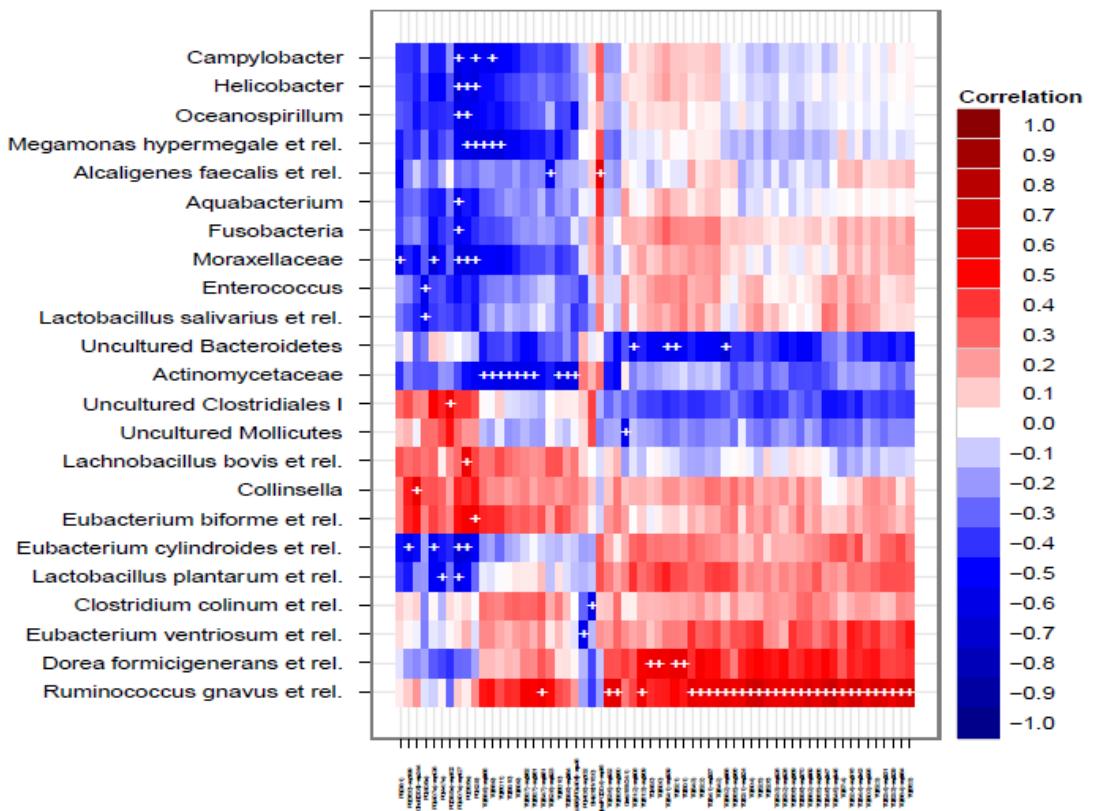
Only positive correlations !



# Probiotic intervention (*L. rhamnosus* GG) 3-week follow-up → No global effect on microbiota composition



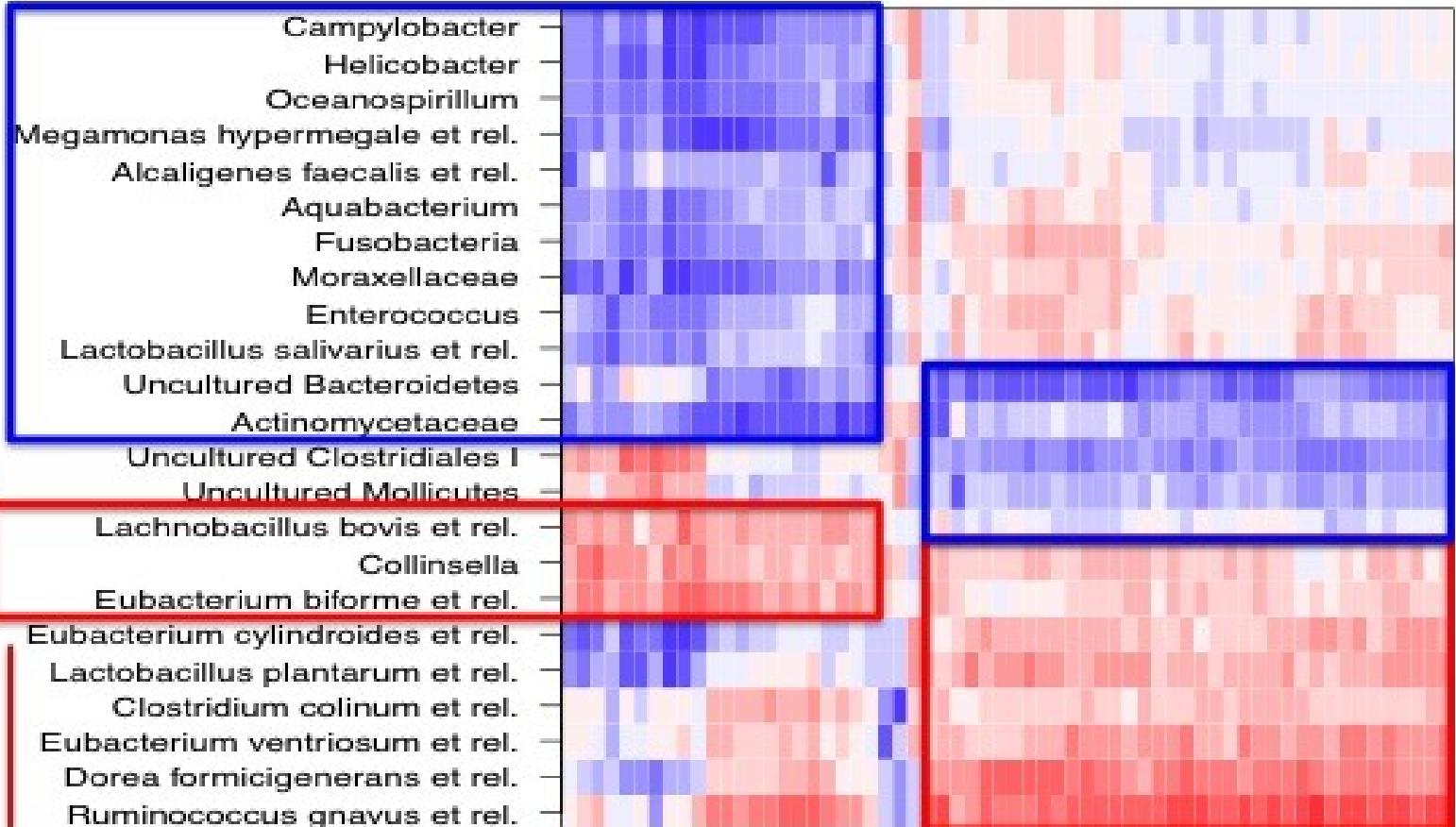
## Correlations microbiota & serum lipids



# Gut microbiota and blood serum lipid links

Metabolic rather than phylogenetic clusters which are likely to indicate functional links

Potential pathogens



Lahti et al. PeerJ 32, 2013

$Q < 0.05$

Red = positive correlation Blue = negative correlation