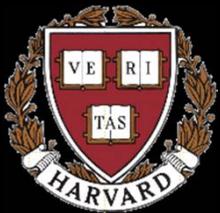




Population epidemiology of the gut microbiome and human health outcomes

Curtis Huttenhower



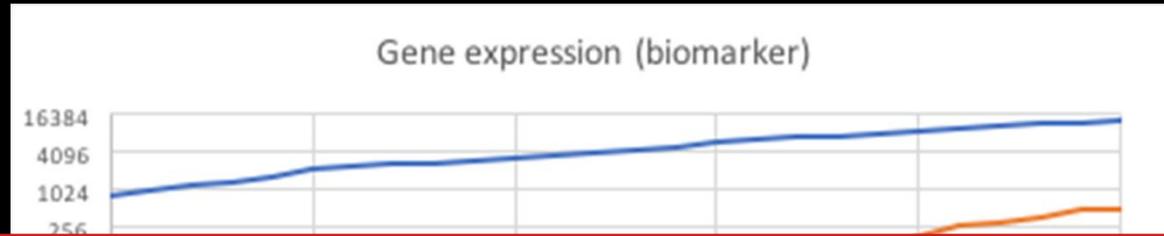
Harvard T.H. Chan School of Public Health
Department of Biostatistics

05-30-17





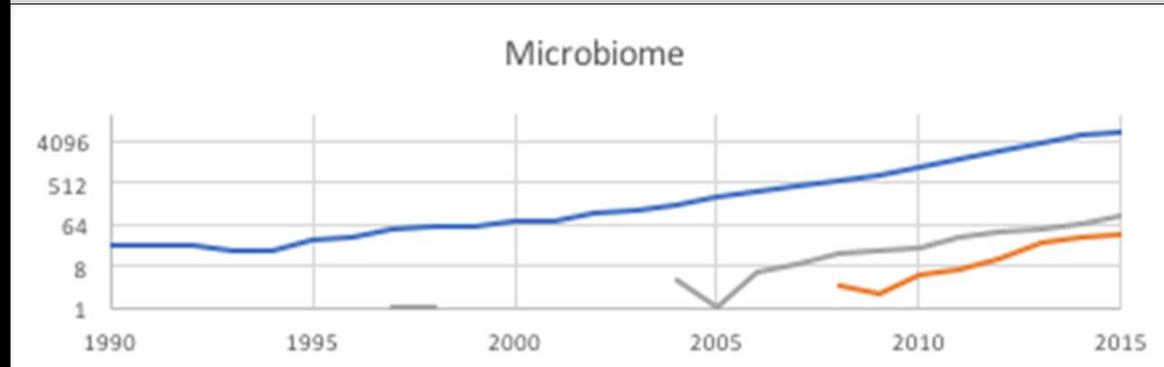
Population studies of the human microbiome



Meta-analysis

More publications contained “microbiome” than “GWAS” in 2015, but multi-population, translatable studies are lagging.

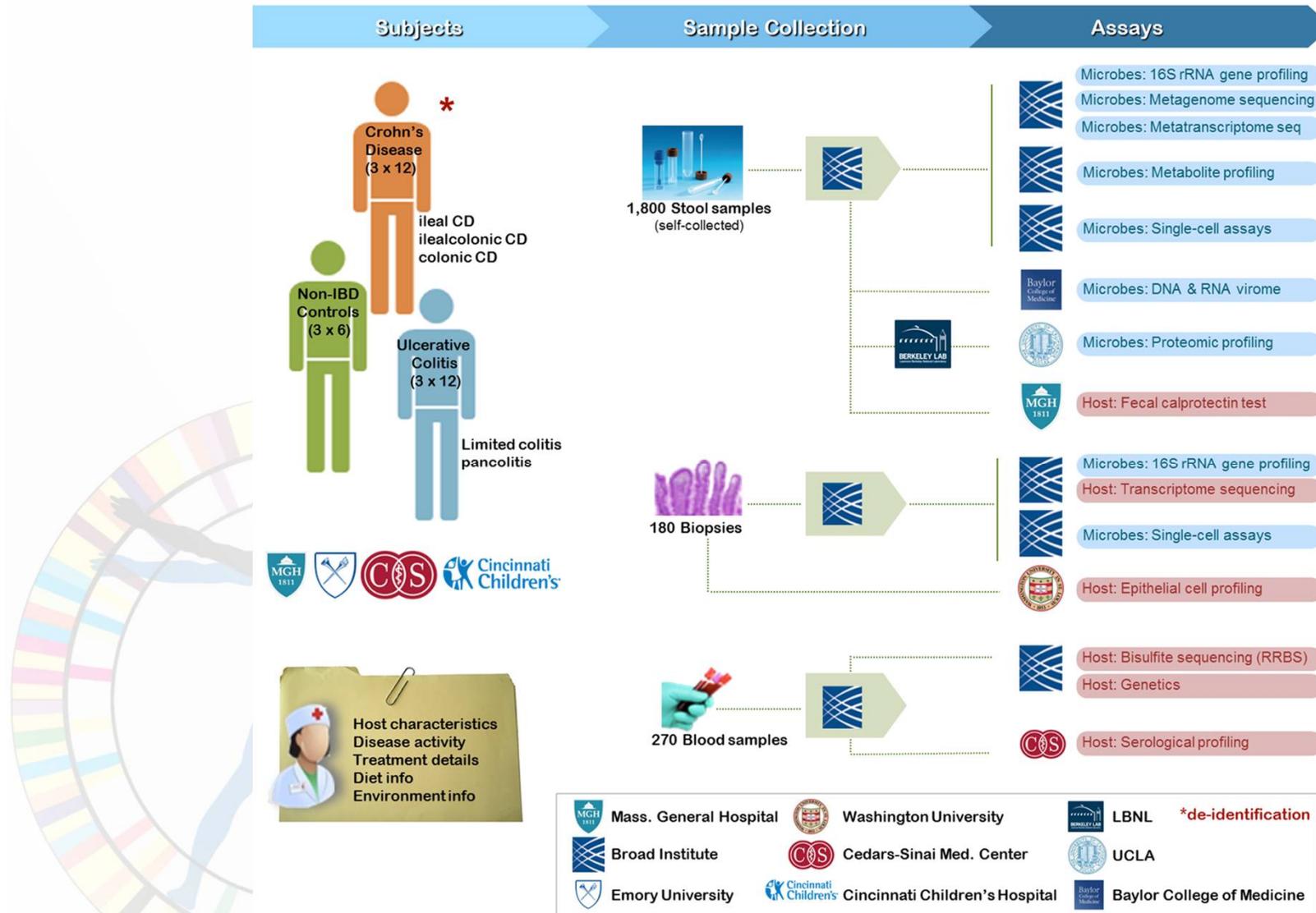
translation



The "HMP2" IBD Multi'omics Data resource

<http://ibdmdb.org>

With Ramnik Xavier



Funded by National Institutes of Health, Dept. of Health and Human Services

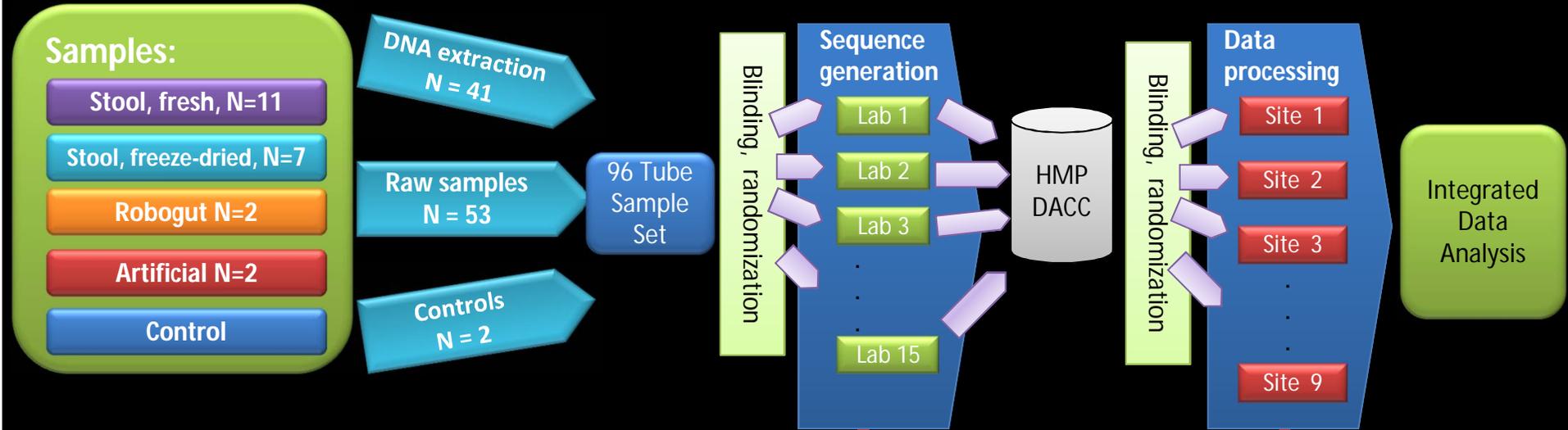




The MicroBiome Quality Control Project (MBQC)

<http://mbqc.org>

With Rashmi Sinha, Christian Abnet, Rob Knight, Owen White



Lab	Location	Lab	Location
Robbie Burk	Einstein	David Mills	UC Davis
Rick Bushman	UPenn	Joe Petrosino	Baylor
Roberto Flores	NCI	Jacques Ravel	UMD
Dirk Gevers	Broad	Pat Schloss	UMich
Greg Gloor	UWO	Orin Shanks	EPA
Andy Goodman	Yale	Rashmi Sinha	NCI
Rob Knight	UCSD	Peter Turnbaugh	UCSF
Dan Littman	NYU		

Lab	Location
Robbie Burk	Einstein
Greg Caporaso	NAU
Nick Chia	Mayo
Greg Gloor	UWO
Curtis Huttenhower	HSPH
Rob Knight	UCSD
Joe Petrosino	Baylor
Jacques Ravel	UMD
Rashmi Sinha	NCI

2,238 samples seq'd, 16,555 samples analyzed:
~3x HMP data, ~6x HMP seqs.!

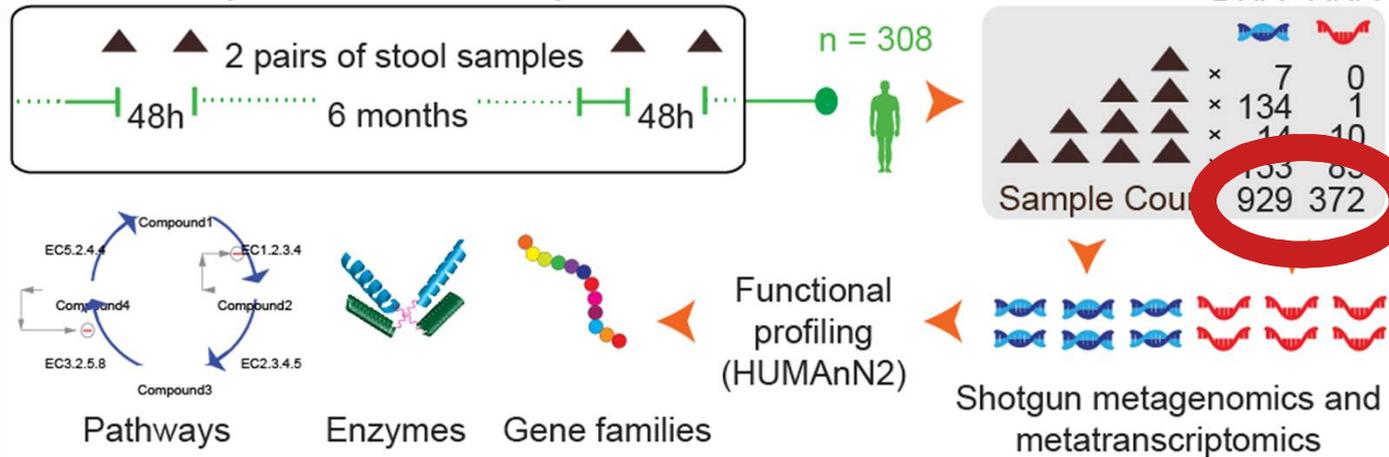


Microbiome functional epidemiology in dietary cancer risk and population studies

With Andy Chan, Zsofia Stadler, Wendy Garrett, Jacques Izard

A Stool self-collection, data generation and profiling

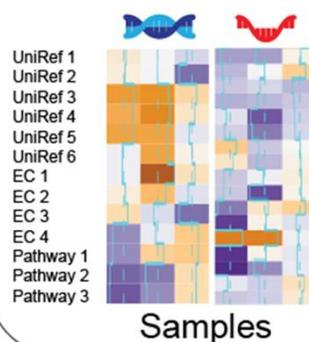
Men's Lifestyle Validation Study



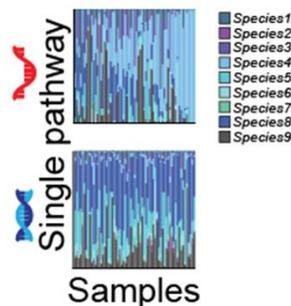
Galeb Abu-Ali

B Quantification and ecology of the metatranscriptome

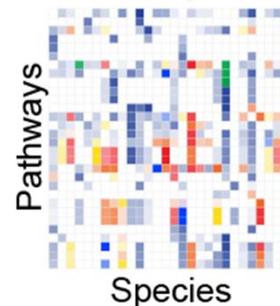
Up/down regulation



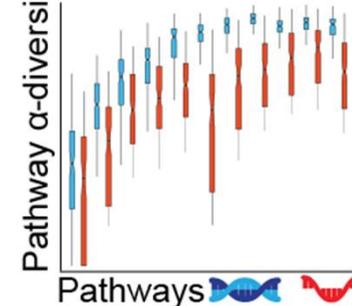
Species stratification



Species-specific transcription

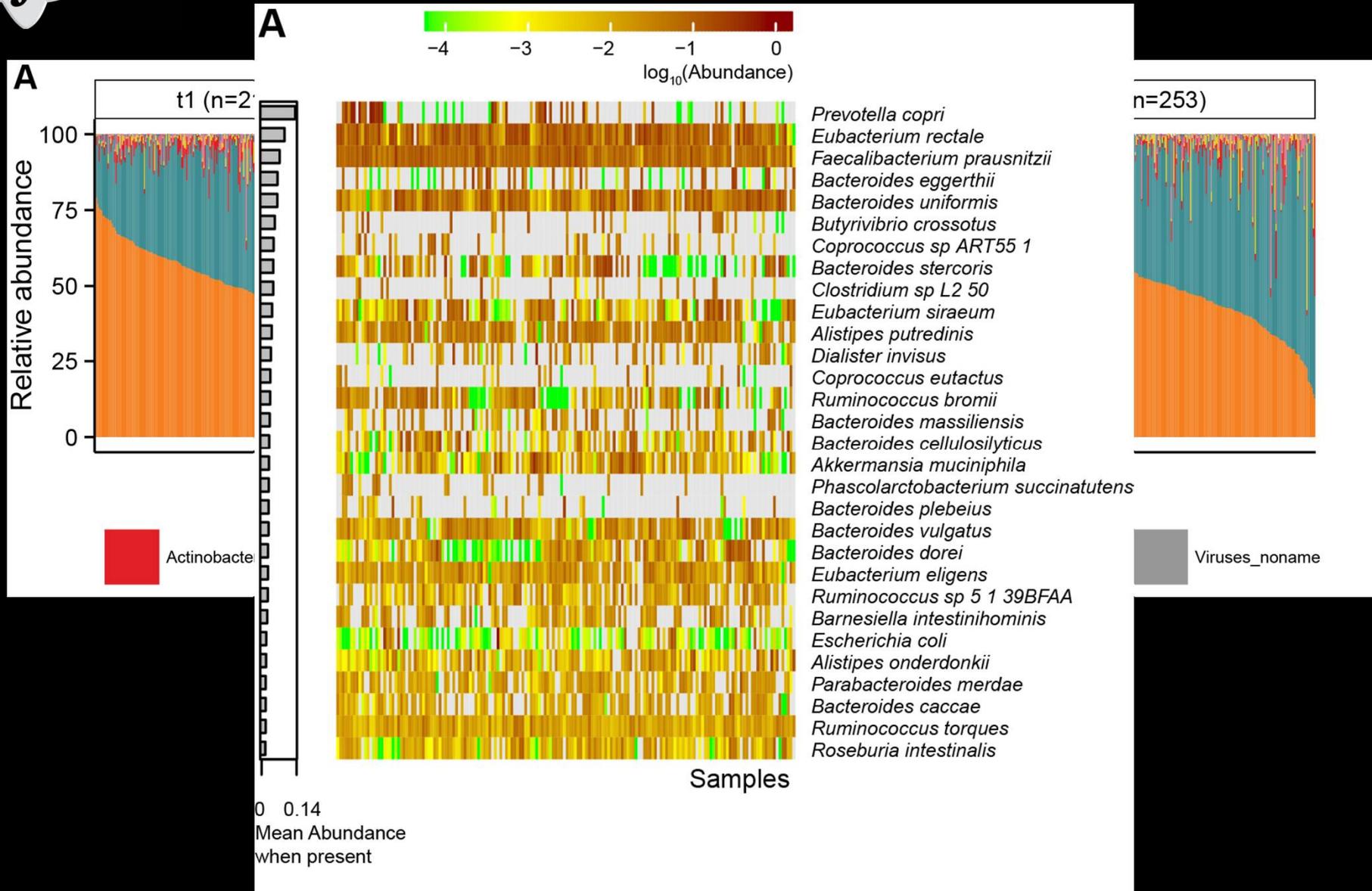


Distribution and utilization



Taxonomic profiles as expected, down to the strain level

<http://huttenhower.sph.harvard.edu/metaphlan2>

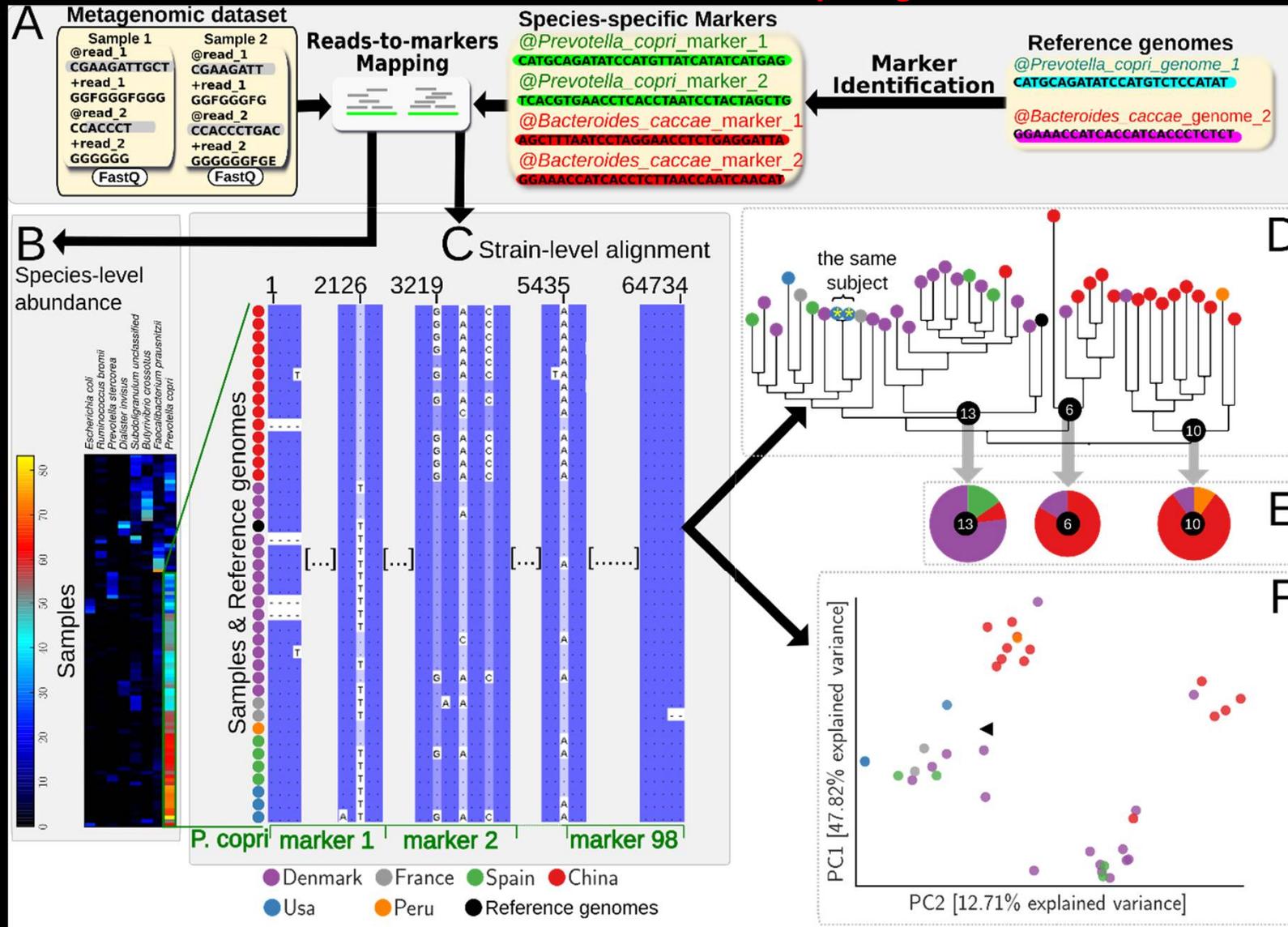


StrainPhlan: metagenomic strain identification and tracking

<http://segatalab.cibio.unitn.it/tools/strainphlan>



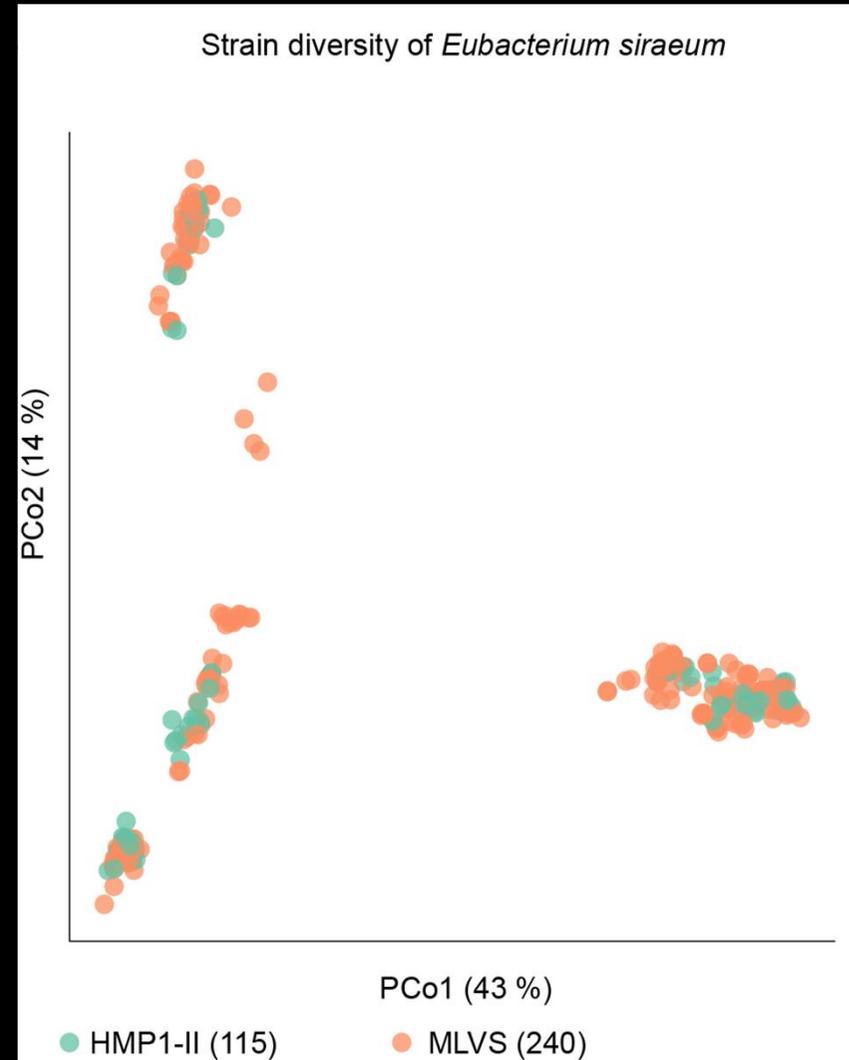
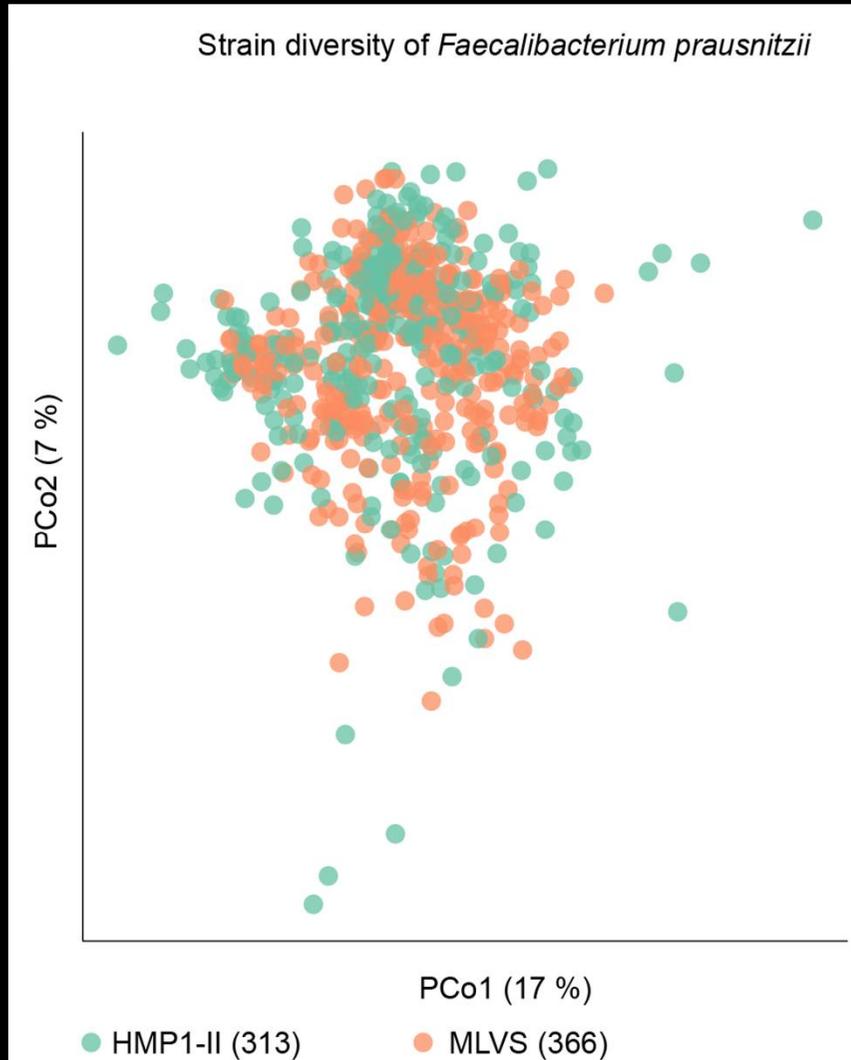
Nicola Segata





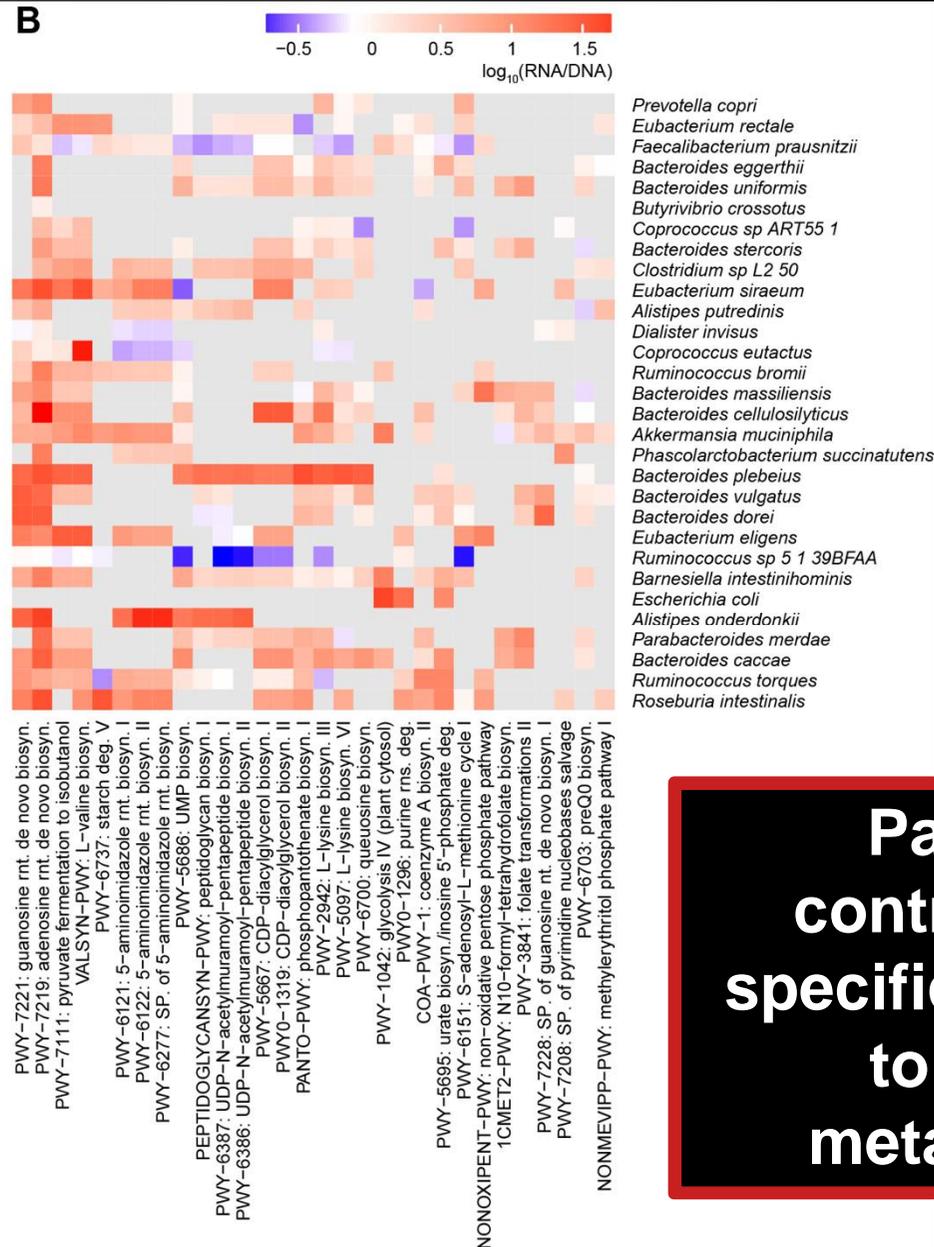
Taxonomic profiles as expected, down to the strain level

<http://segatalab.cibio.unitn.it/tools/strainplan>





But important to link taxonomy to metagenomic and metatranscriptomic functional profiles



**Pathways
contributed by
specific organisms
to the gut
metagenome.**



HUMAN2 for taxon-specific metagenome and metatranscriptome functional profiling

Input

Taxonomic prescreen

Pangenome mapping

Translated search

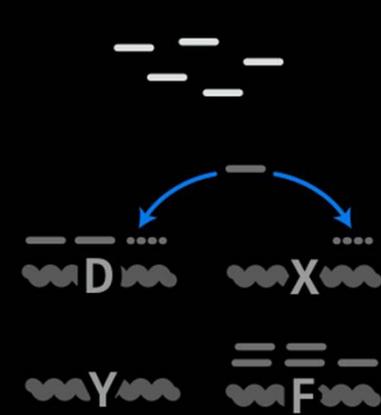
Gene Family	Abd.
UniRef50_A6L0N6	67
UniRef50_A6L0N6 s__Bacteroides_fragilis	8
UniRef50_A6L0N6 s__Bacteroides_finegoldii	5
UniRef50_A6L0N6 s__Bacteroides_stercoris	4
UniRef50_A6L0N6 unclassified	1
UniRef50_G9S1V7	60
UniRef50_G9S1V7 s__Bacteroides_vulgatus	31
UniRef50_G9S1V7 s__Bacteroides_thetaiotaomicron	22
UniRef50_G9S1V7 s__Bacteroides_stercoris	7

- Species1
- Species2
- Ambiguous
- Novel

Reads mapped to
clade-specific marker
genes to rapidly identify
community species



Reads mapped to
pangenomes of



Unclassified reads
multi-hit mapped to

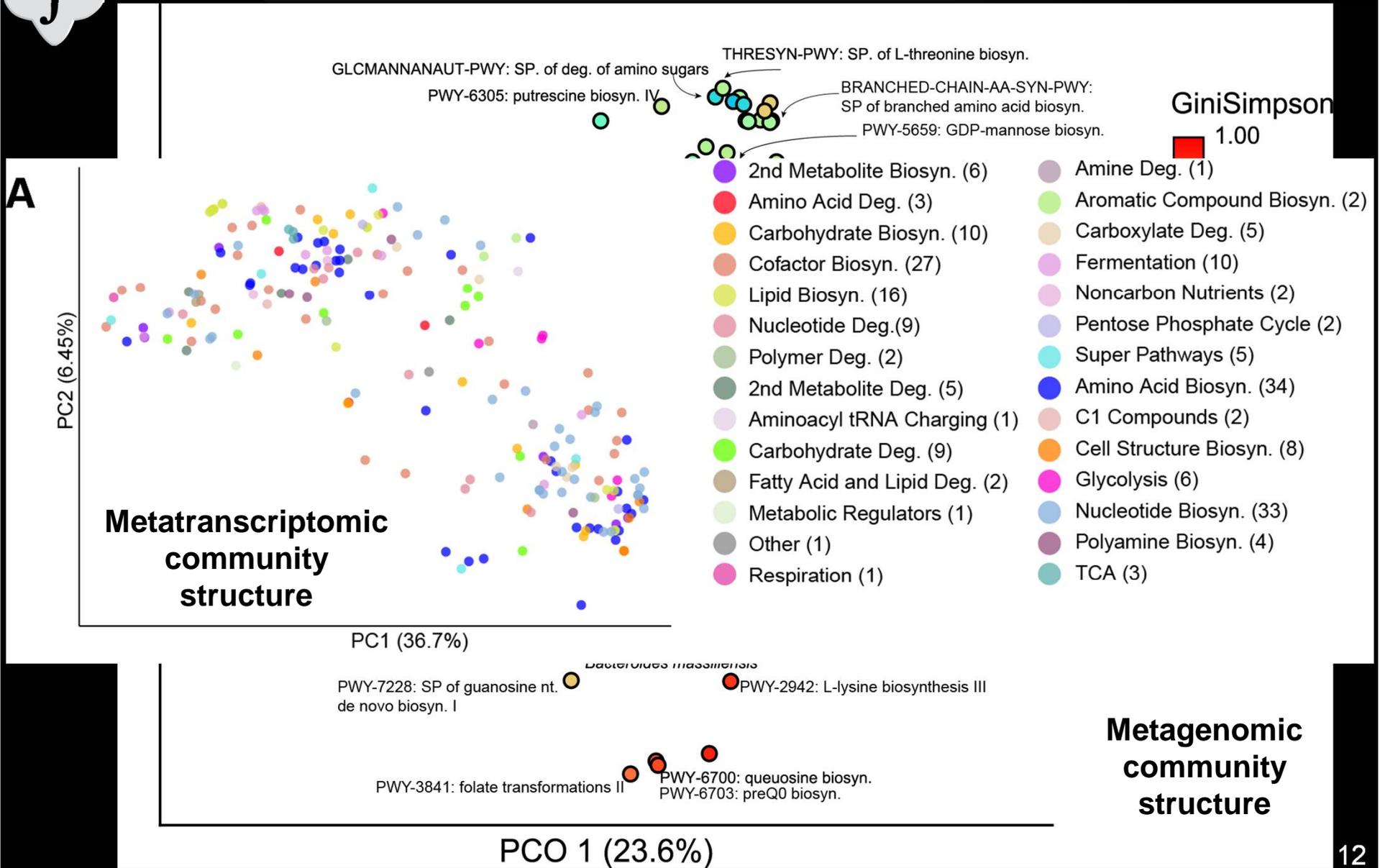
Pathway	Abd.
indole-3-acetate activation	57.01
indole-3-acetate activation unclassified	32.26
indole-3-acetate activation s__Bacteroides_ovatus	4.5
indole-3-acetate activation s__Alistipes_putredinis	3
indole-3-acetate activation s__Bacteroides_caccae	2.25
melibiose degradation	55
melibiose degradation unclassified	17
melibiose degradation s__Parabacteroides_merdae	8
melibiose degradation s__Bacteroides_caccae	6

(external to HUMAN2)





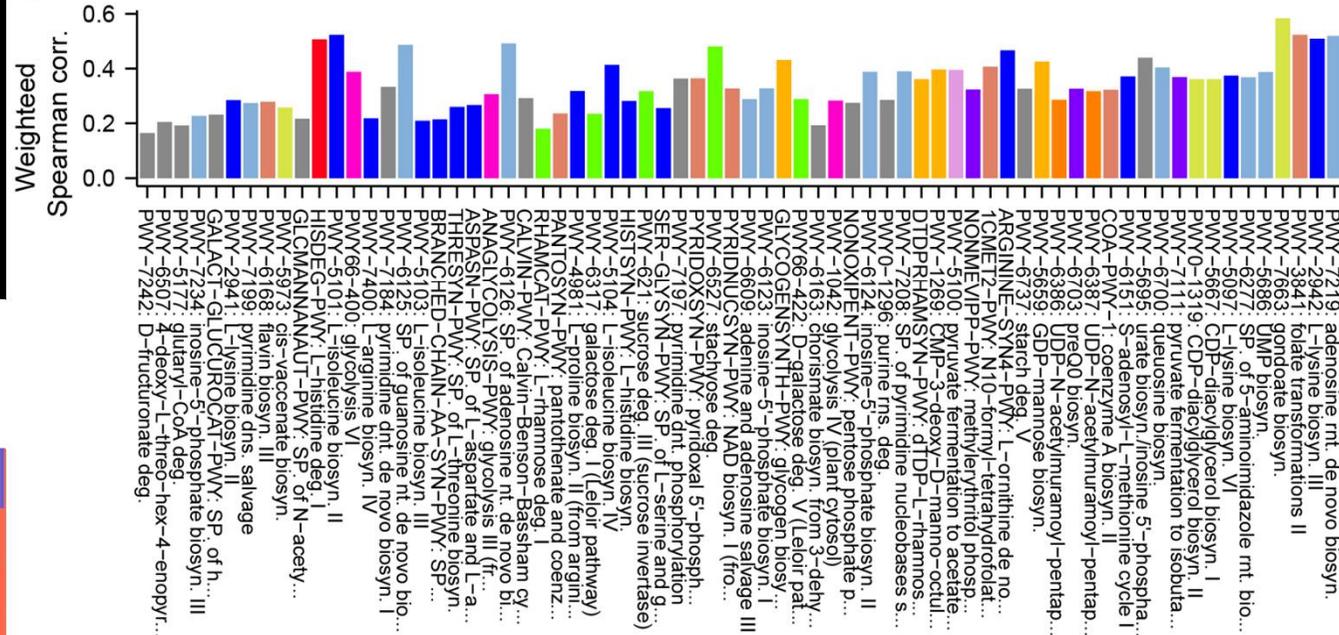
Functional potential and expression are correlated, but with very different contributing taxa





Functional potential and expression are correlated, but with very different contributing taxa

B

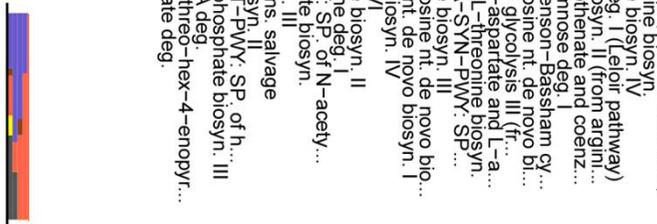


Functional Category

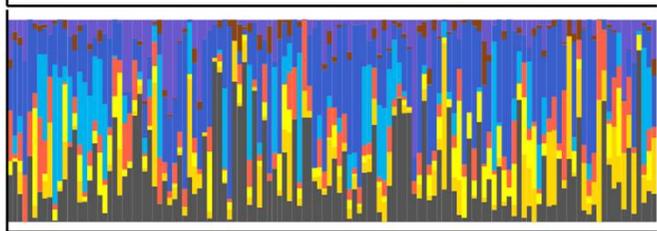
- Amino Acid Biosyn.
- Amino Acid Deg.
- Carbohydrate Biosyn.
- Carbohydrate Deg.
- Cell Structure Biosyn.
- Cofactor Biosyn.
- Fermentation
- Glycolysis
- Lipid Biosyn.
- Nucleotide Biosyn.
- Other
- 2nd Metabolite Biosyn.

C

RNA

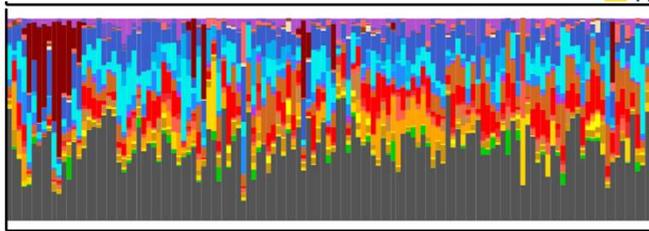


DNA



Subject

DNA



Subject

- *Eubacterium rectale*
- *Eubacterium siraeum*
- *Faecalibacterium prausnitzii*
- *Methanobrevibacter smithii*
- *Odoribacter splanchnicus*
- *Parabacteroides distasonis*
- *Prevotella copri*
- *Roseburia intestinalis*
- *Ruminococcus bromii*
- *Ruminococcus torques*
- other

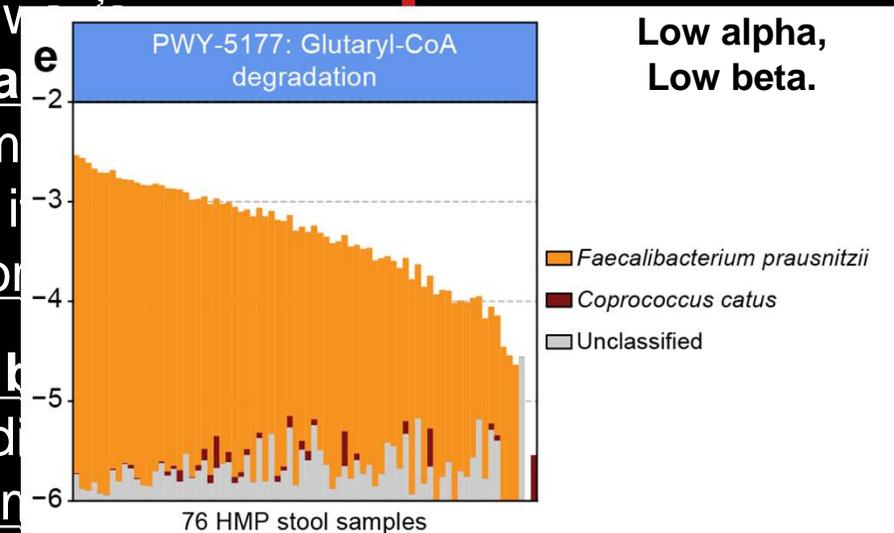
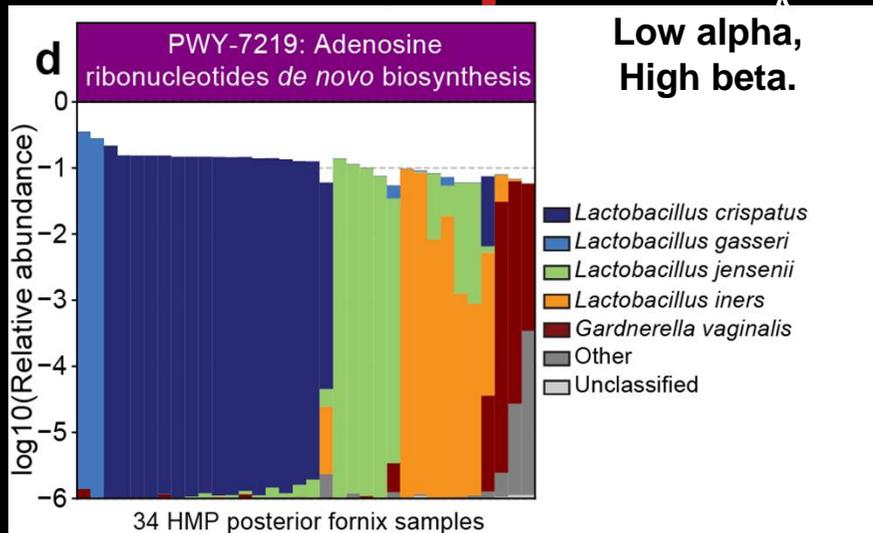
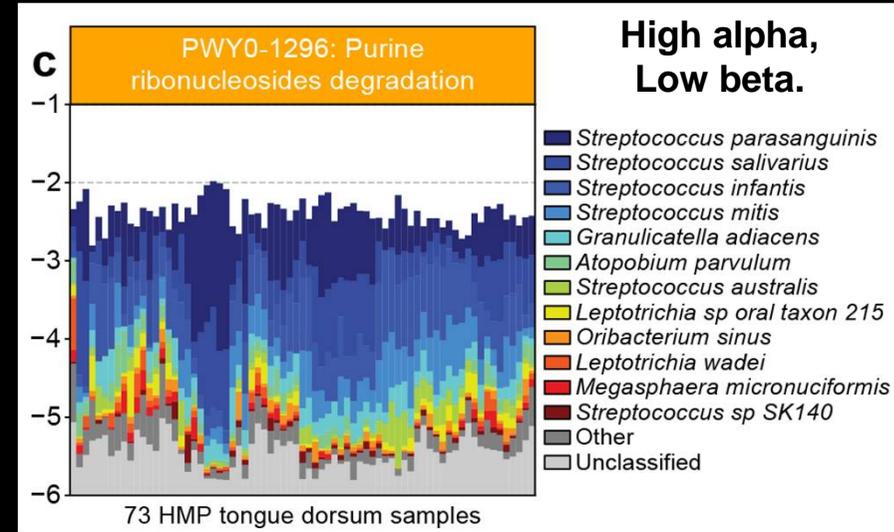
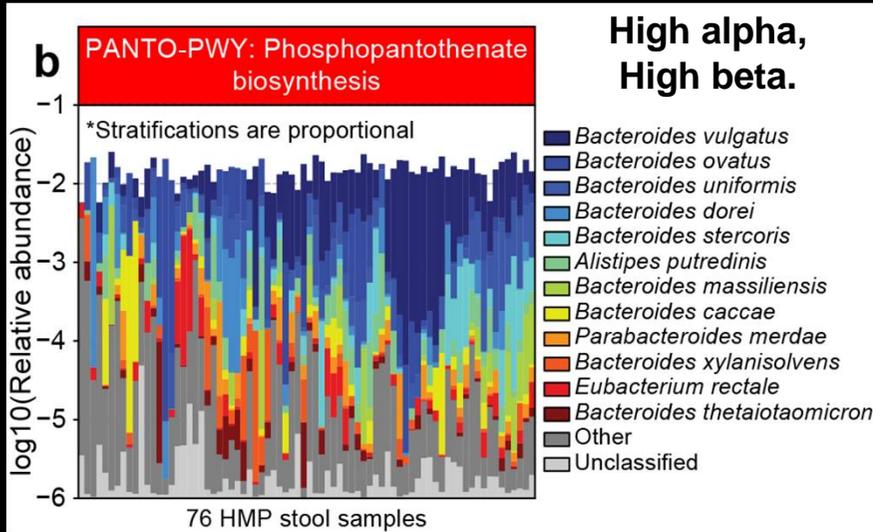
phila
i

nsis

aomicron
is
s
;



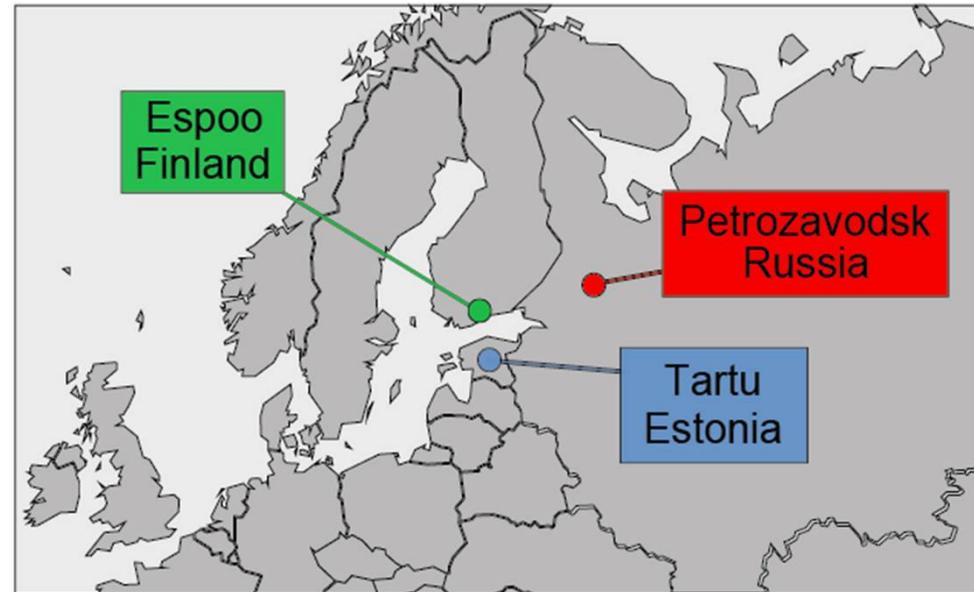
Contributional diversity: linking ecological structure and function



Type 1 diabetes, the gut microbiome, and the hygiene hypothesis in infancy



Tommi Vatanen

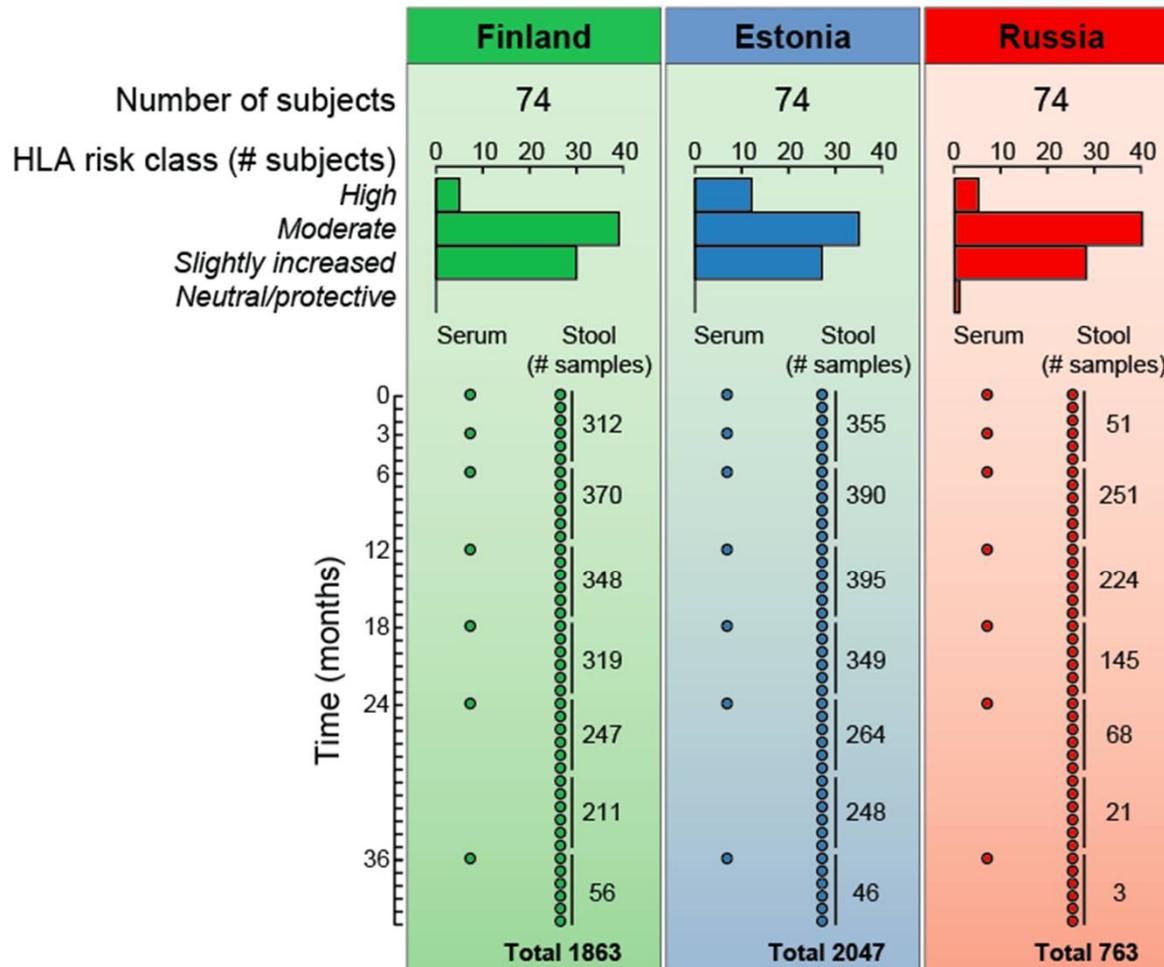


In Finland, 6-fold higher incidence of T1D compared to Russia

In Estonia, rapid transition towards Finnish incidence after Soviet exit

~1000 newborn infants were screened for increased HLA-conferred susceptibility to autoimmunity

DIABIMMUNE infants have similar genetic risk for autoimmunity

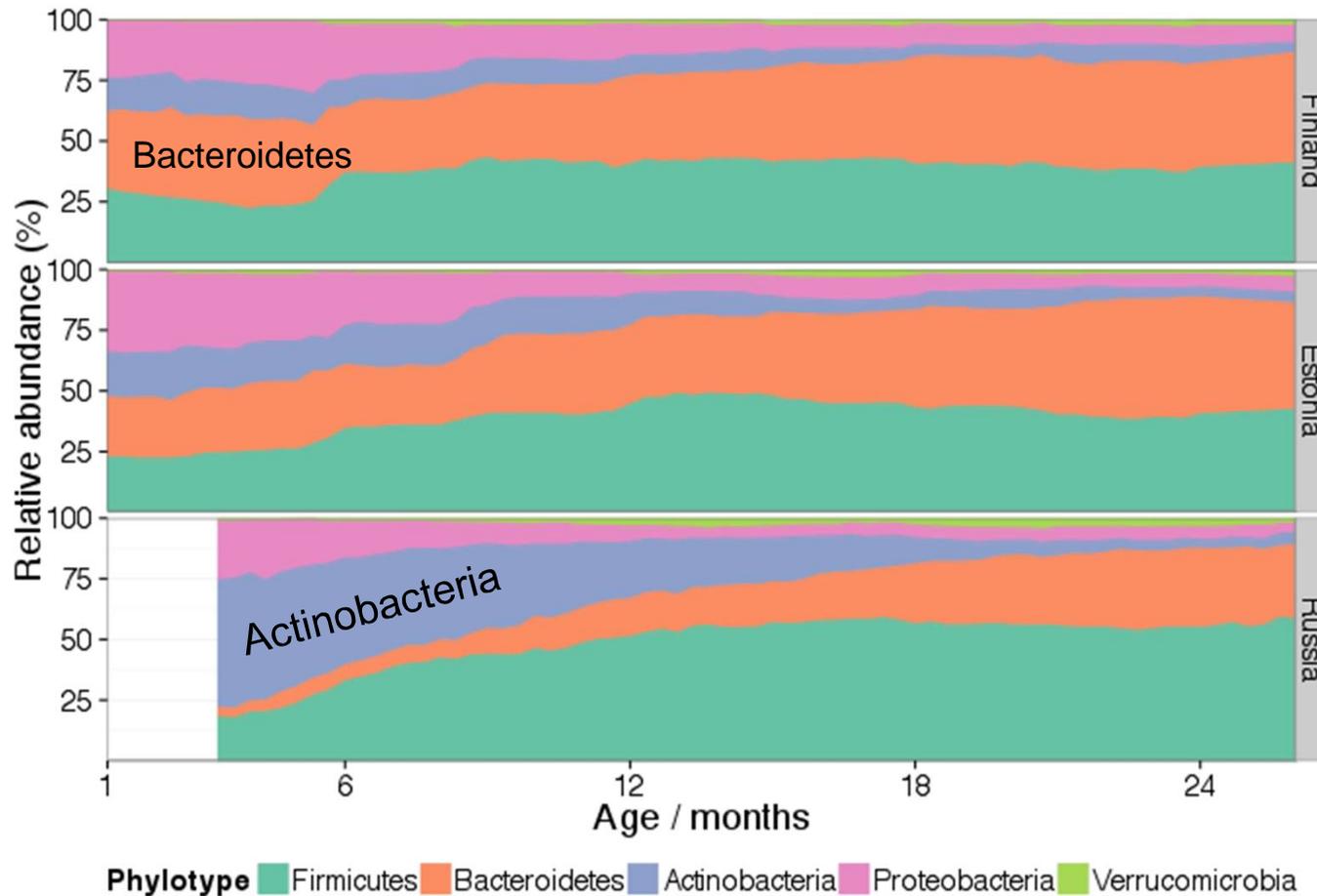


Number of stool samples sequenced:

1585 (16S sequencing)

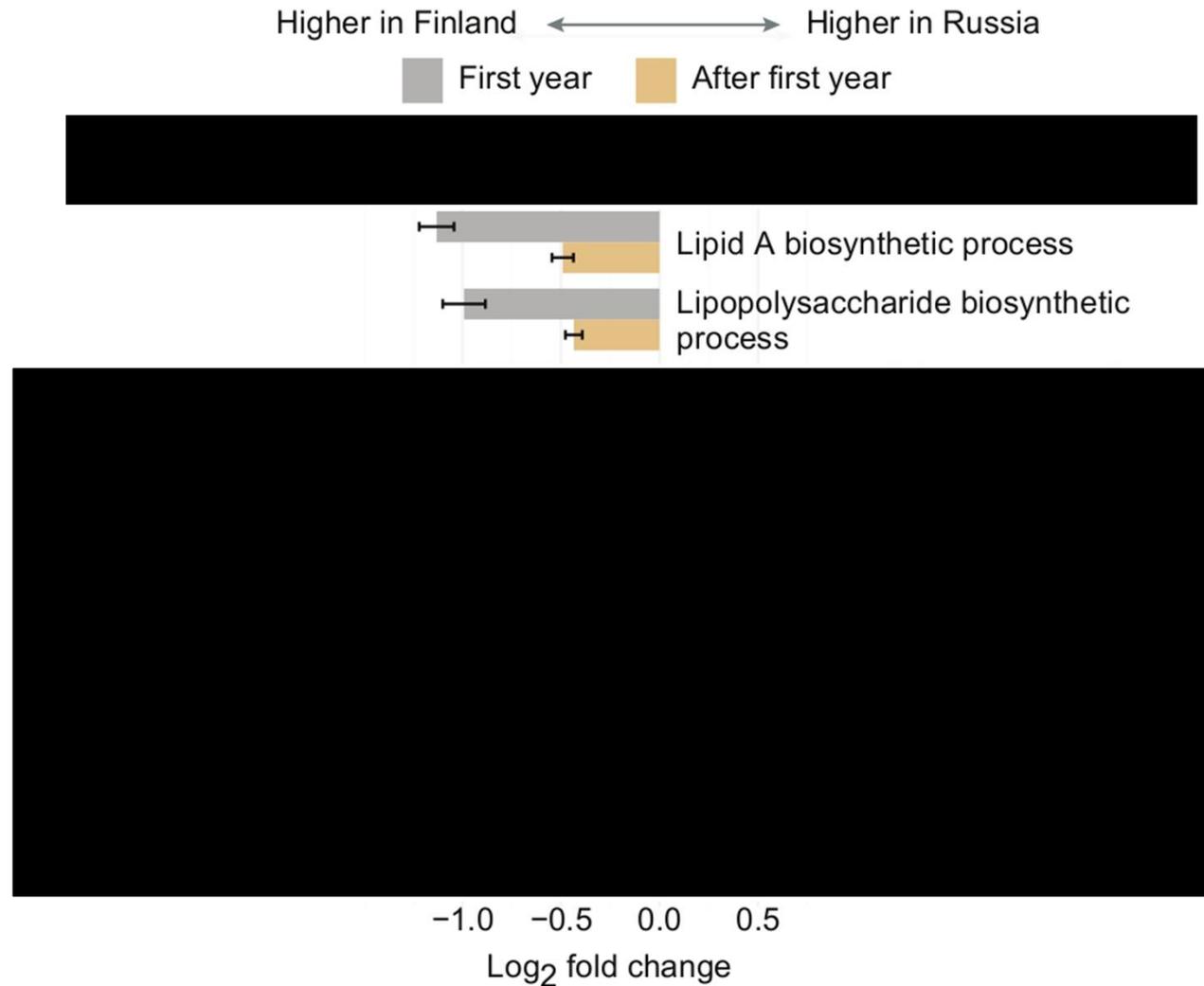
785 (WGS sequencing)

HMO metabolism is driven by different species in Finland and Russia

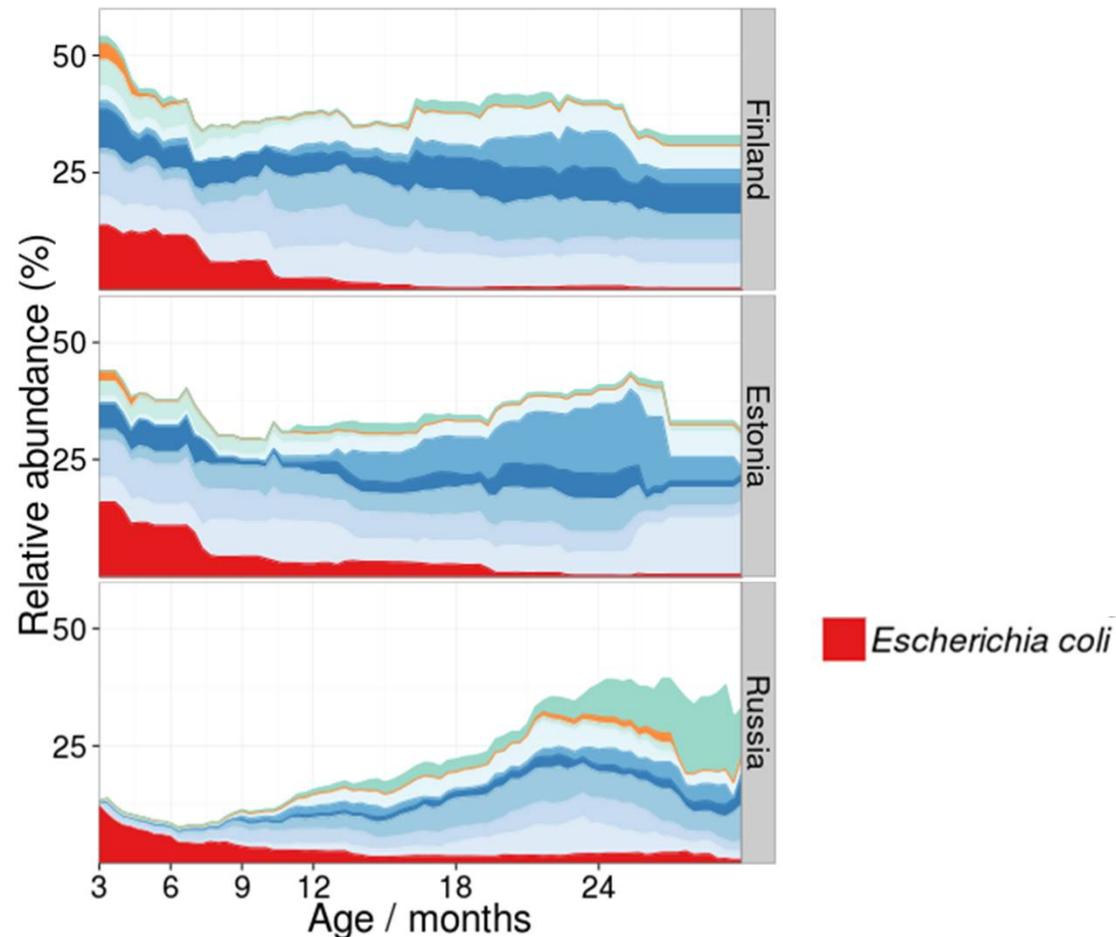


Most HMO utilization genes are conferred by *Bifidobacterium* in Russians and *Bacteroides* in Finns and Estonians

Lipopolysaccharide (LPS) biosynthesis pathways are differentially abundant between Finland and Russia



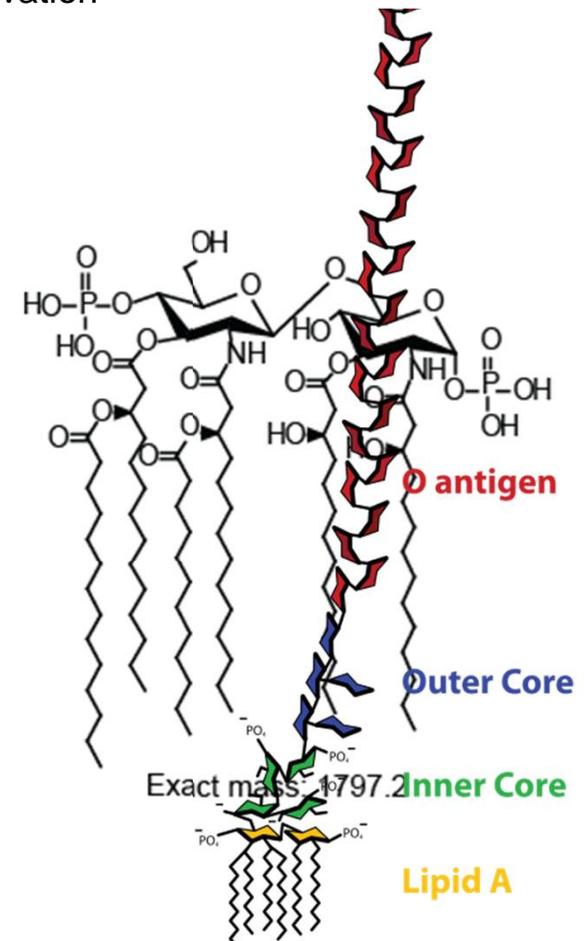
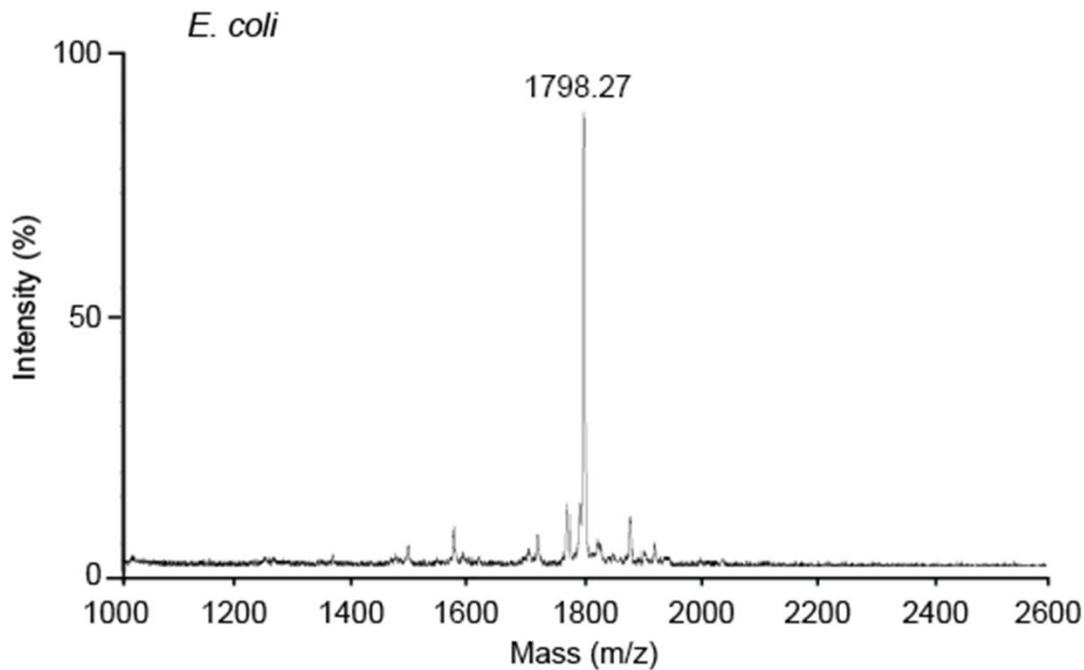
Finnish and Estonian microbiome has LPS from many Bacteroidetes species



***Bacteroides dorei* has been previously associated with T1D pathogenesis (Davis-Richardson et al. 2014)**

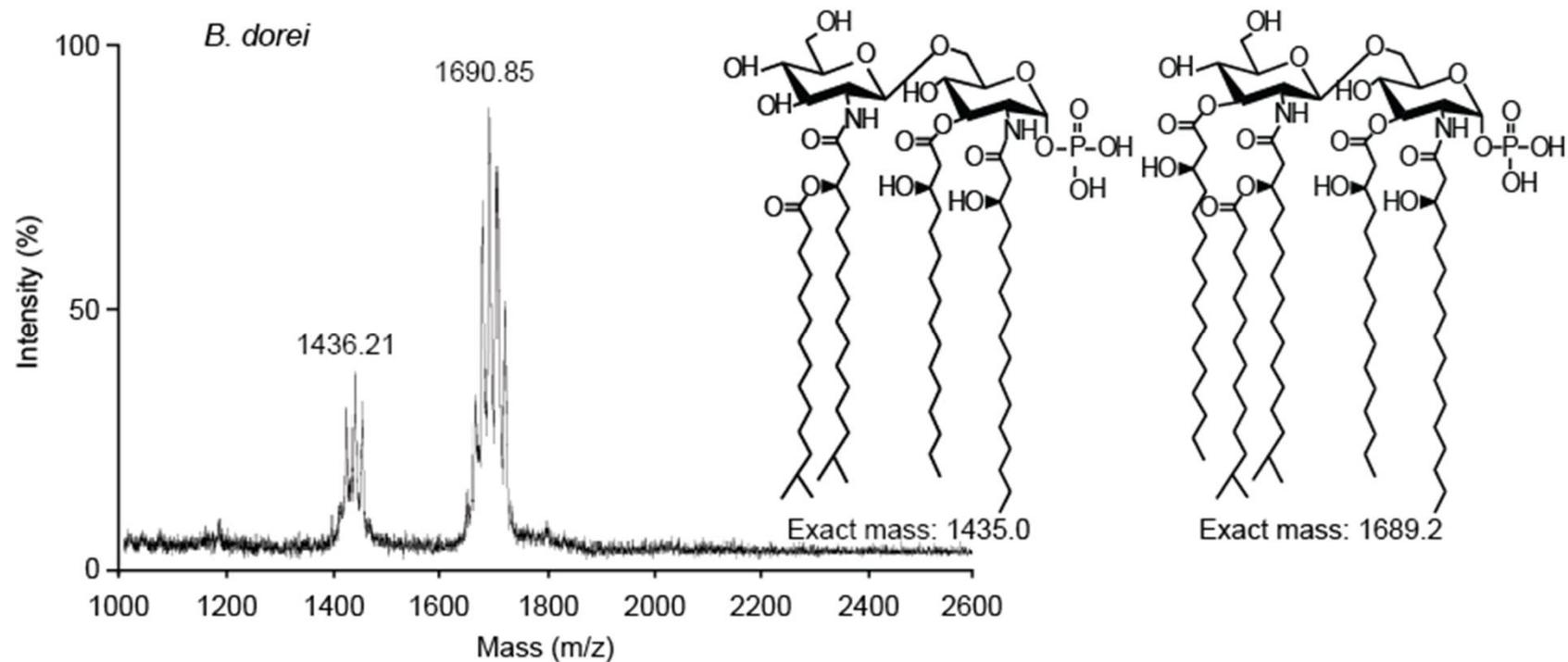
Lipid A of *E. coli* is hexa-acylated

- Lipid A is a lipid component of LPS responsible for its immunostimulatory properties
- The number of acyl chains is known to affect the immune activation
- *E. coli* has six acyl chains



Lipid A in *B. dorei* is tetra/penta acylated

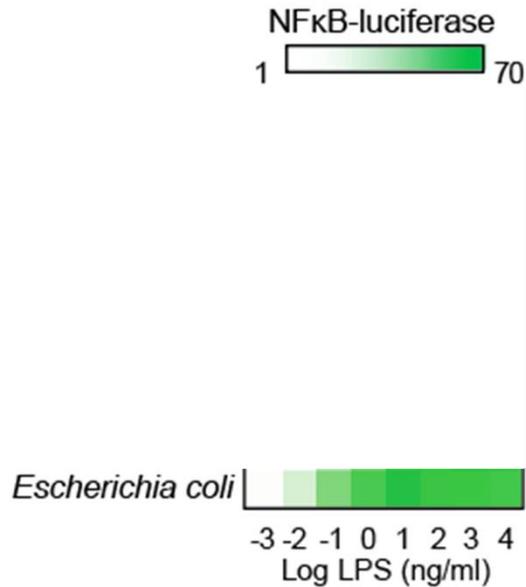
- Repeating the same experiment with *B. dorei* LPS shows two peaks



The structure of *B.dorei* lipid A
may be less immunostimulatory

LPS from *Bacteroides* species is less immunostimulatory compared to *E. coli* LPS

HEK293 NFκB reporter
downstream TLR4 signaling



***B. dorei* and *B. vulgatus* harbor immunologically silent form of LPS**



Population studies of the human microbiome

Predicting and Manipulating Cardiac Drug Inactivation by the Human Gut Bacterium *Eggerthella lenta*

Henry J. Haider,¹ David B. Gootenberg,¹ Kelly Chatman,¹ Gopal Sirasani,² Emily P. Balskus,² Peter J. Turnbaugh^{1*}

A conserved bacterial protein induces pancreatic beta cell expansion during zebrafish development

Jennifer Hampton Hill¹, Eric A Franzosa^{2,3}, Curtis Huttenhower^{2,3}, Karen Guillemin^{1,4*}

Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease

James D. Lewis,^{6,7,9} Eric Z. Chen,^{6,9} Robert N. Baldassano,¹ Anthony R. Otley,³ Anne M. Griffiths,⁴ Dale Lee,^{1,2} Kyle Bittinger,⁸ Aubrey Bailey,⁸ Elliot S. Friedman,⁷ Christian Hoffmann,⁸ Lindsey Albenberg,¹ Rohini Sinha,⁸ Charlene Compher,³ Erin Gilroy,³ Lisa Nessel,⁸ Amy Grant,³ Christel Chehoud,⁸ Hongzhe Li,⁹ Gary D. Wu,^{7,*} and Frederic D. Bushman^{6,*}

MUCOSAL IMMUNITY

Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut

Michael R. Howitt,¹ Sydney Lavoie,¹ Monia Michaud,¹ Arthur M. Blum,² Sara V. Tran,³ Joel V. Weinstock,² Carey Ann Gallini,¹ Kevin Redding,³ Robert F. Margolskee,³ Lisa C. Osborne,^{4,*} David Artis,⁴ Wendy S. Garrett^{1,5,6†}

Propionibacterium-Produced Coproporphyrin III Induces *Staphylococcus aureus* Aggregation and Biofilm Formation

Michael S. Wollenberg,^{a,b*} Jan Claesen,^c Isabel F. Escapa,^{a,b} Kelly L. Aldridge,^{a*} Michael A. Fischbach,^c Katherine P. Lemon^{a,d}

Translating the microbiome *starts* with population studies...

- Not all mechanisms are taxon-linked – multi'omics.
- Dynamic on an intermediate time scale – longitudinal samples.
- Microbial physiology is strain-specific – culture isolates.
- Causality is unclear in humans – model systems.



bioBakery
A virtual environment for
meta'omic analysis

<http://huttenhower.sph.harvard.edu/biobakery>
First hit when you google "biobakery"

Composition Analysis

These tools can determine the composition in terms of (i) microbial species and their associated abundances (MetaPhlAn) or (ii) genes and associated pathways (HUMAnN) in the dataset. Please click on the links below for detailed tutorials:

HUMAnN

- Microbial species and associated genes and pathways

MetaPhlAn

- Microbial species and abundances

PhyloPhlAn

- Reconstruction of phylogenetic trees

PICRUSt

- Predict metagenome functional content from marker gene

ShortBRED

- Abundance of proteins of interest in genetic data

PPANINI

- Prioritize microbial genes based on their metagenomic properties

Statistical Analysis

These tools can determine the associations from the provided metadata information and microbial composition tables. Please click on the links below for detailed tutorials:

HAIIA

- Hierarchical All-against-All association testing

ARepA

- Extract 'omics data from repositories

CCREPE

- Assess the significance of general similarity measures in compositional datasets

LEfSe

- Association between metadata (max. 2) and microbial species and abundances

MaAsLin

- Association between metadata (no restriction) and microbial species and abundances

microPITA

- Sample selection in two stage-tiered studies

SparseDOSSA

- A hierarchical model of microbial ecological population structure

Infrastructure and Utilities

The following tools may be used for additive utility and framework for your projects:

GraPhlAn

- Generating cladograms

KneadData

- Removing 'contaminant' reads

AnADAMA

- Automating data analysis

- Upstream microbial genome curation
- Taxonomic & functional profiling of meta'omes (metagenomes + metatranscriptomes)
- Microbial & clinical association discovery
- Reproducible workflows & infrastructure

**Source code,
documentation,
& tutorials**



Thanks!

<http://huttenhower.sph.harvard.edu>



Brantley Hall



Yan Yan



Cesar Arze



Casey DuLong



Daniela Boernigen



Lauren McIver



Ramnik Xavier



Human Microbiome Project 2

Lita Procter

Bruce Birren

Jon Braun

Chad Nusbaum

Dermot McGovern

Clary Clish

Subra Kugathasan

Joe Petrosino

Ted Denson

Thad Stappenbeck

Janet Jansson

Owen White



George Weingart



Emma Schwager



Eric Franzosa



Boyu Ren



Tiffany Hsu



Ali Rahnavard



Hera Vlamakis



Mikael Knip



Zsafia Stadler
Annie Lincoln



Jason Lloyd-Price



Chrystal Chan



Tommi Vatanen



Koji Yasuda



Siyuan Ma



Galeb Abu-Ali



Wendy Garrett



Andy Chan
Mingyang Song
Raaj Mehta



Owen White
Victor Felix
Jonathan Crabtree



Bahar Sayoldin



Toby Branch



Melanie Schirmer



Himel Mallick



Moran Yassour



Long Nguyen



Nicola Segata



Interested? We're looking for a bioinformatics center lead!



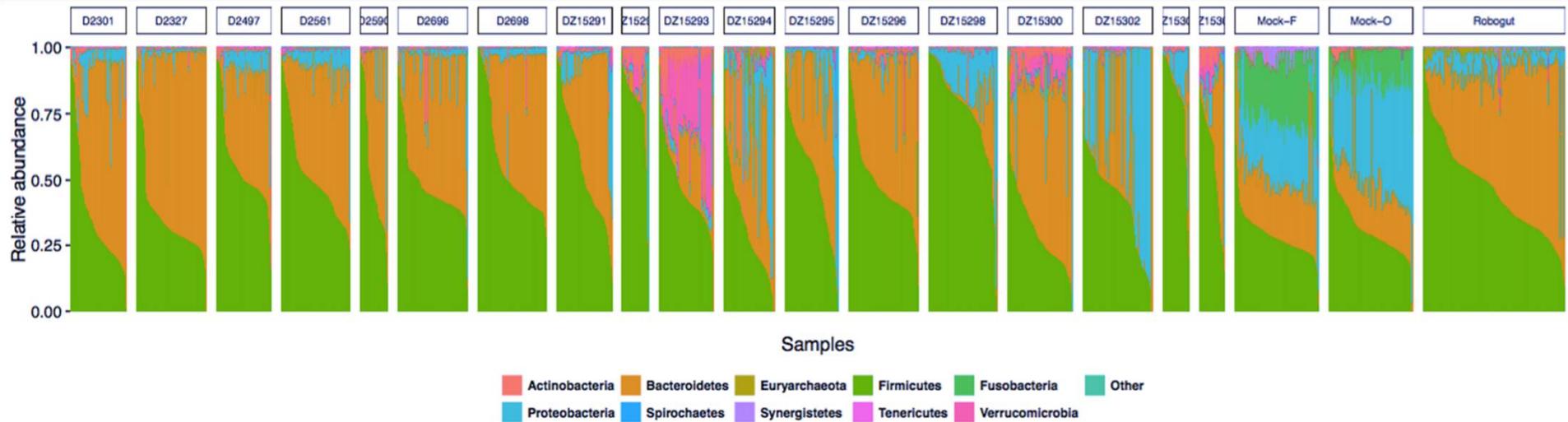




High-level taxonomy is as expected, but with tremendous diversity

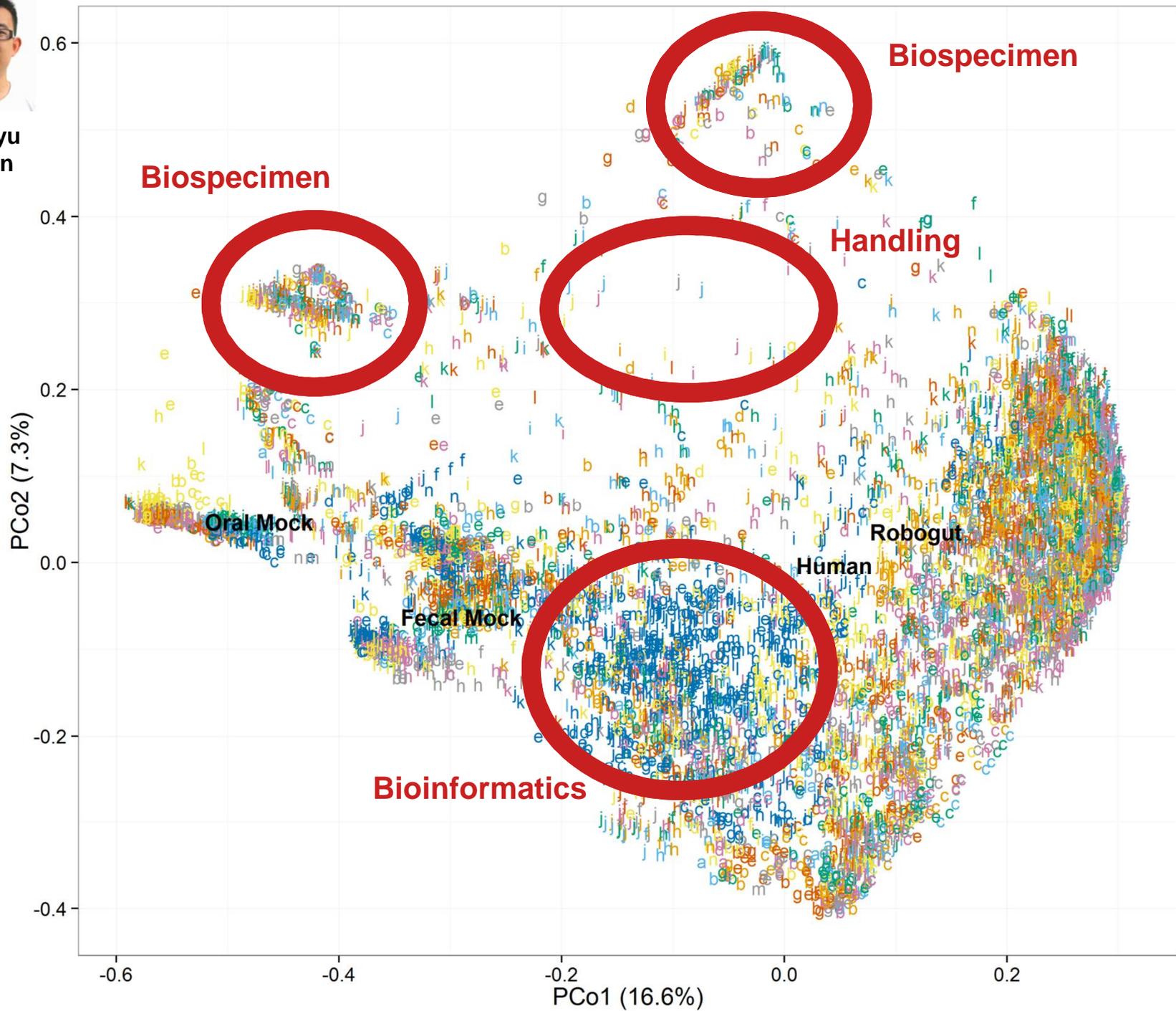


Galeb
Abu-Ali

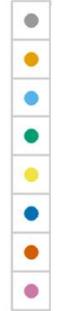




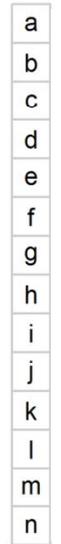
Boyu Ren



Dry lab



Wet lab



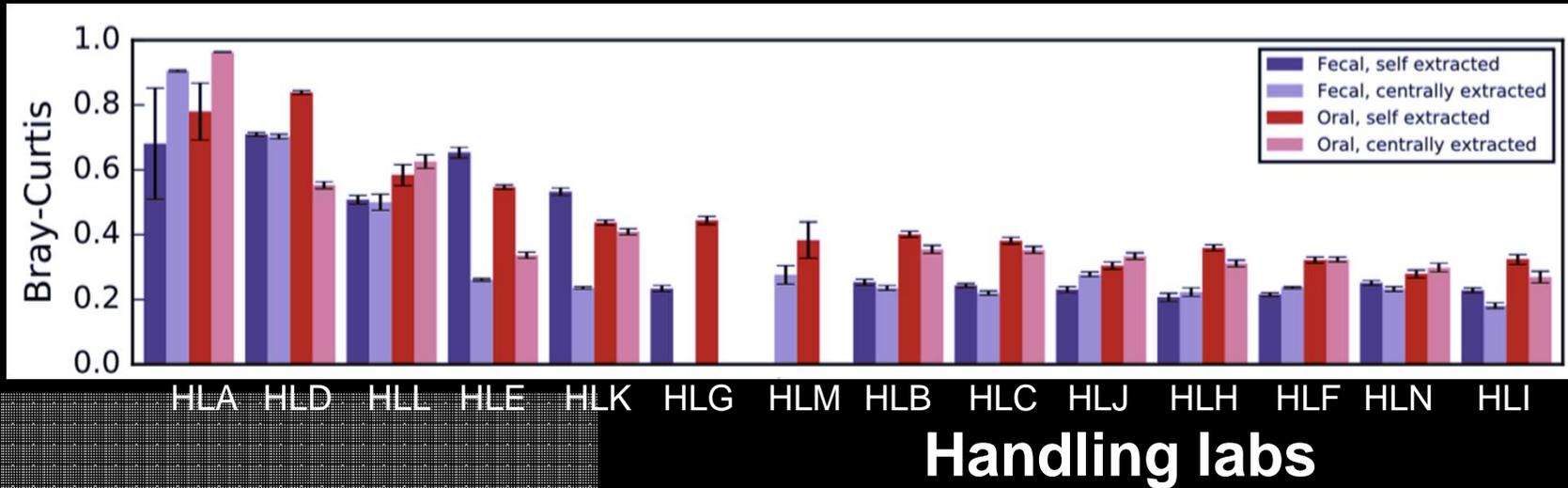


Positive controls: fecal and oral defined communities

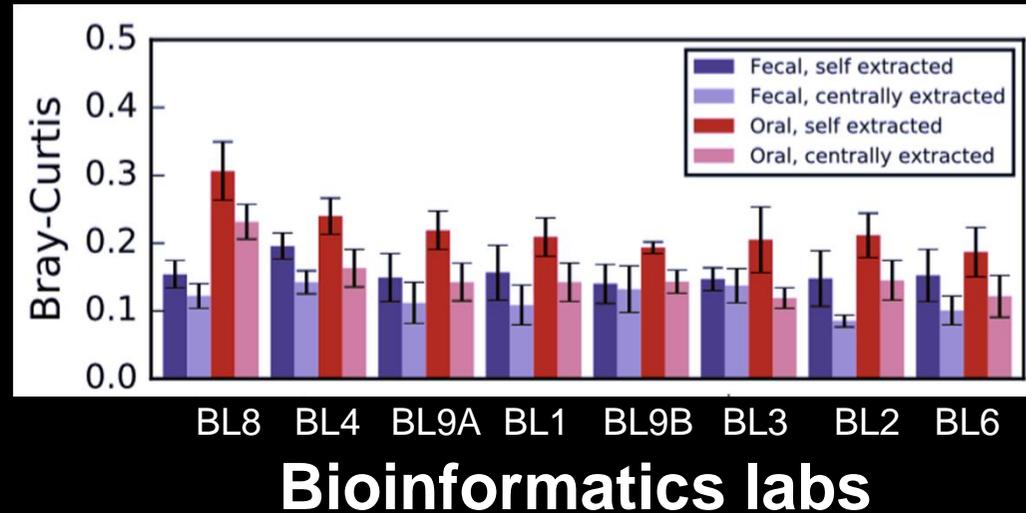


Emma Allen-Vercoe

Distance from truth



Distance from handler





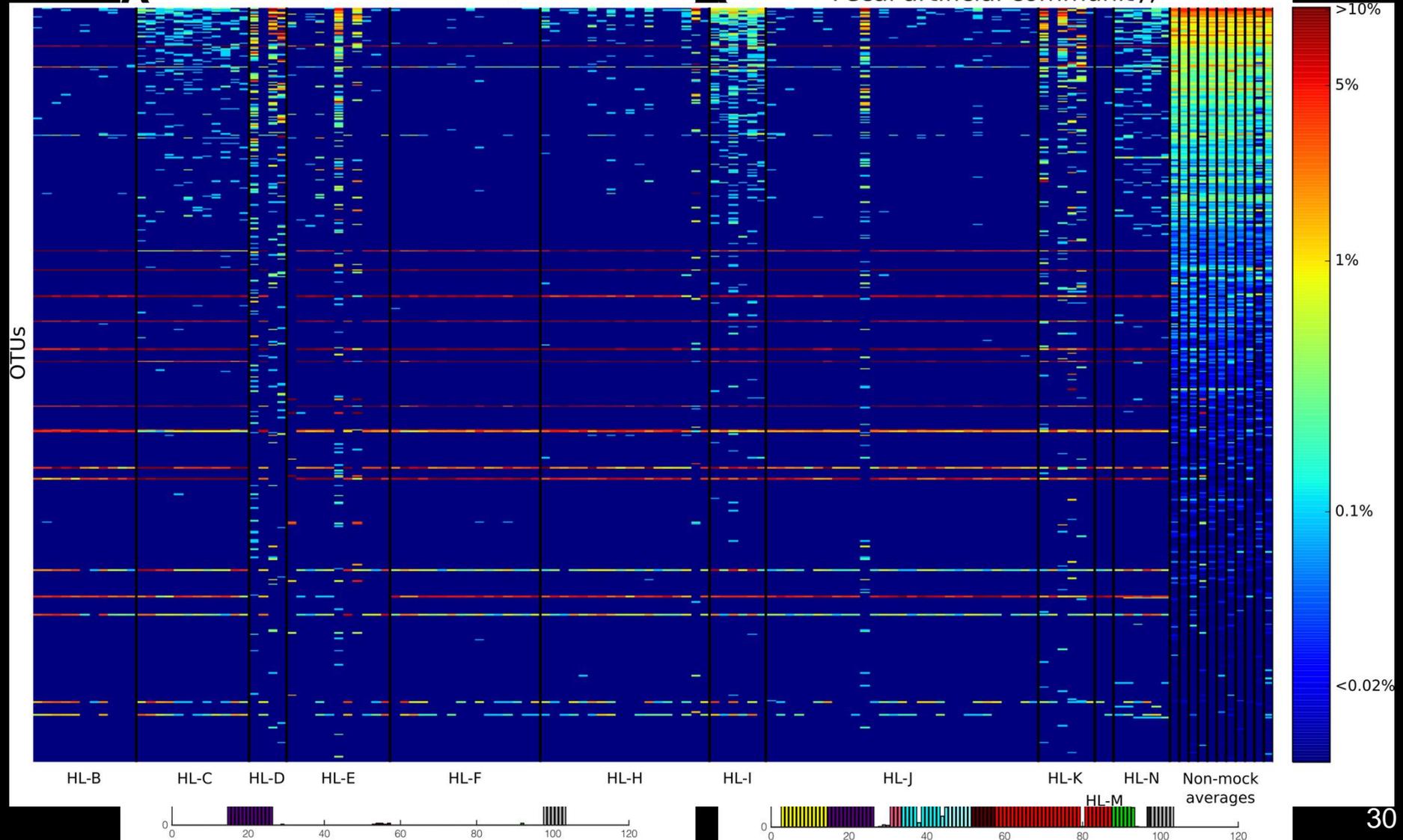
Anything that can go wrong, will go wrong, sometimes



Amnon
Amir,
Knight lab

A Fecal artificial community,

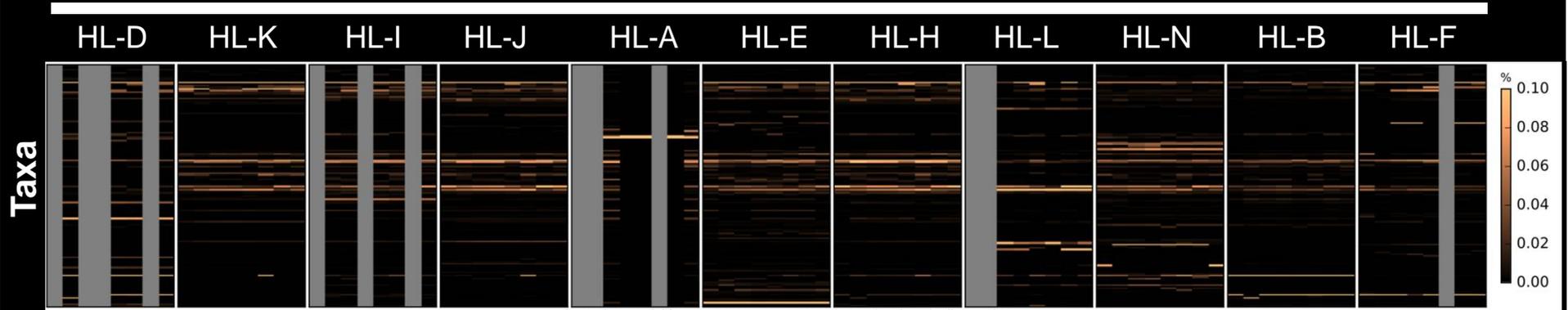
C Fecal artificial community,



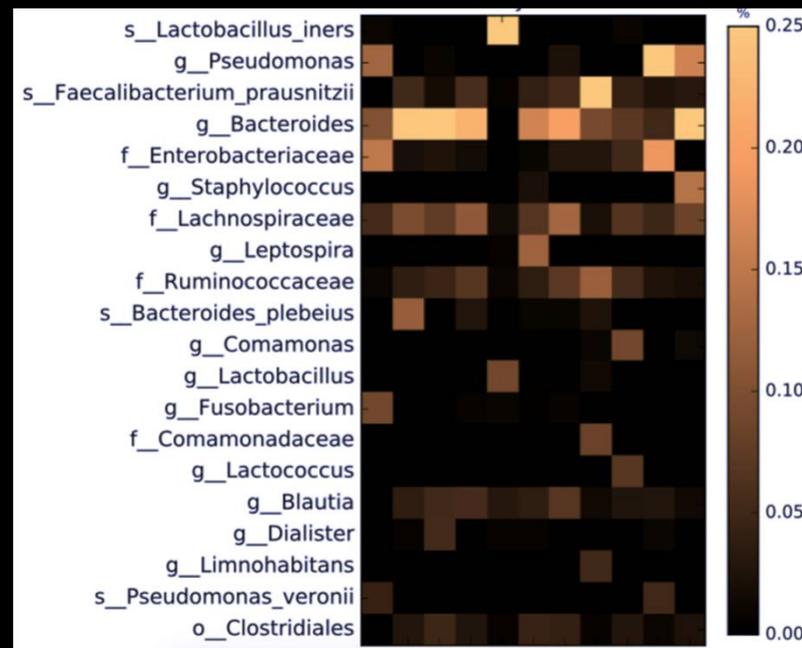


Negative controls: buffer blanks

Handling labs



Bioinformatics labs



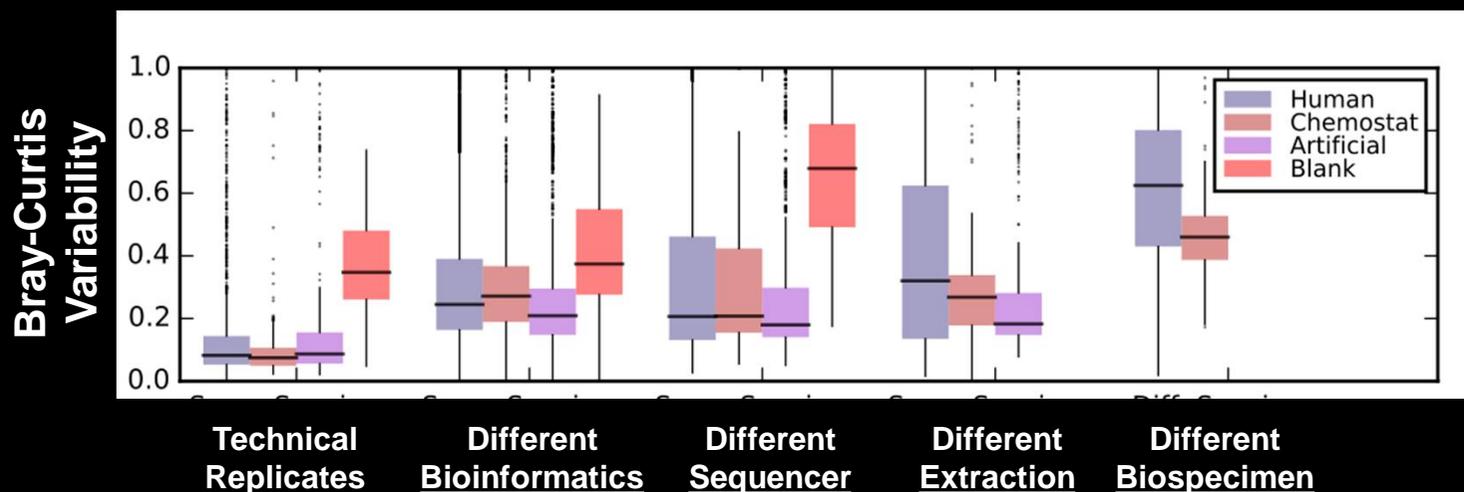
Emma Schwager

Handling labs



Outcomes from the MBQC-baseline

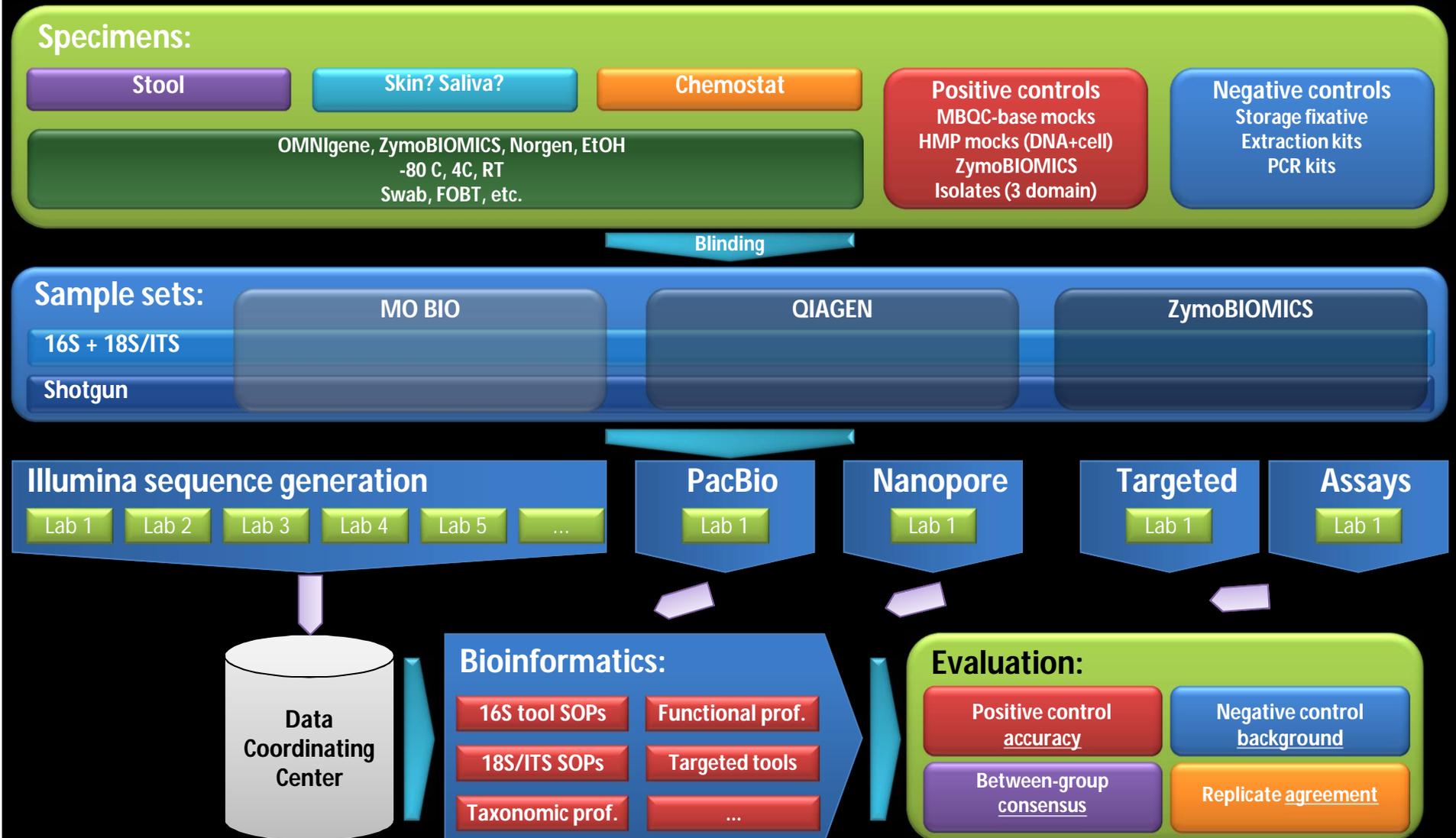
- Explored sequencing and bioinformatics protocol space
 - Labs chose very broadly from among nearly all possible protocol variables
 - The HMP explored microbiome variation; MBQC-base outlined protocol variation
- All of that possible technical variation...
 - ...was no bigger than, and often smaller than, biological phenotypes





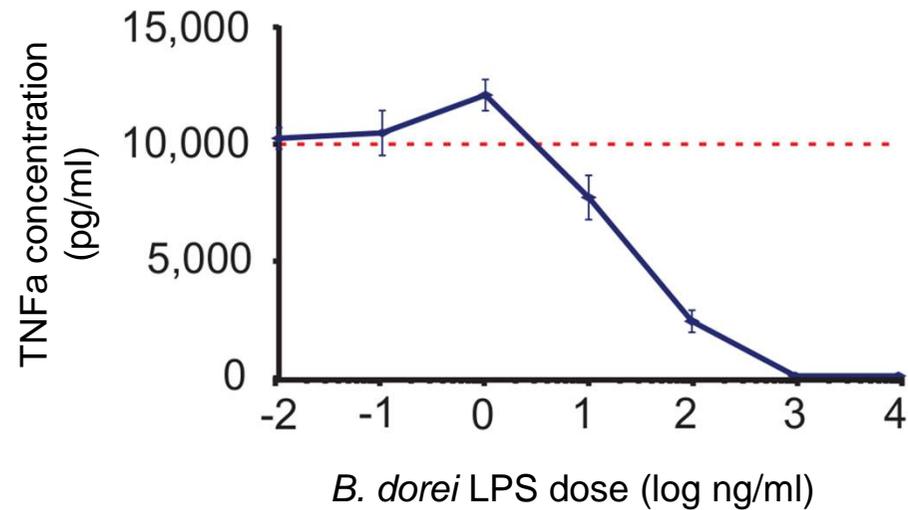
What's next for MBQC-II?

Modified from MBQC-base figure from Owen White



B. dorei LPS antagonizes immune stimulation by *E. coli* LPS

Stimulate human PBMCs:



Microbe-specific LPS stimulation may be responsible for a component of the hygiene hypothesis

