# Response of the gut microbiome and metabolome to dietary fiber in healthy dogs

Amrisha Bhosle[1,2,3], Matthew I. Jackson[4], Aaron Walsh[2], Eric A. Franzosa[1,2,3], Curtis Huttenhower[1,2,3,5], Dayakar V. Badri[4]

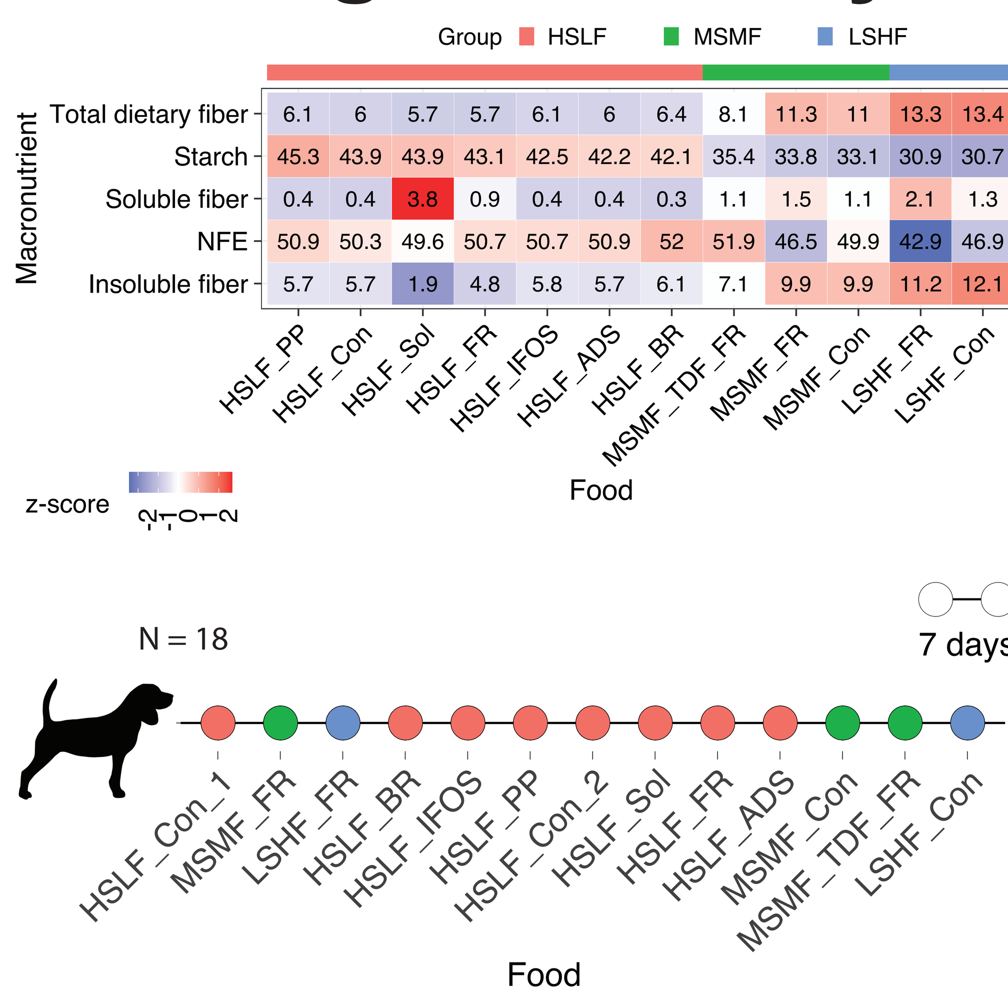
[1]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, [2]Department of Biostatistics, Harvard T. H. Chan School of Public Health,
[3]Harvard Chan Microbiome in Public Health Center, Harvard T. H. Chan School of Public Health
[4]Pet Nutrition Center, Hill's Pet Nutrition Inc.,[5]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health

BROAD INSTITUTE · THE HARVARD CHAN MICROBIOME IN PUBLIC HEALTH CENTER · Hill's · VERITAS · ONE HEALTH MICROBIOME RESOURCE

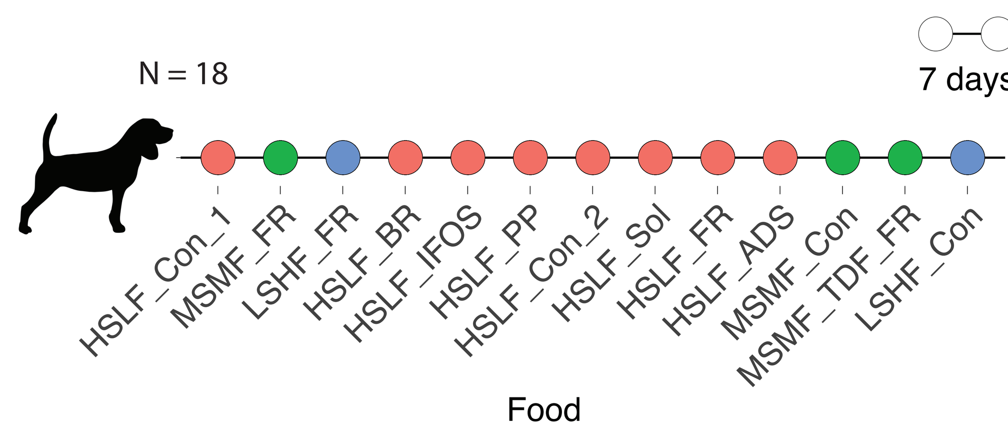
## Dietary fiber and the microbiome

Nutrients and compounds from diet can directly influence the gut microbiome and microbial metabolism of these compounds can in turn influence the host. Metabolism of dietary fiber by the microbiome provides several health-relevant metabolites such as short chain fatty acids (SCFAs) which participate in intestinal homeostasis and immune regulation, and fiber-released compounds that affect gastrointestinal physiology. Dietary fiber interventions in both humans and dogs have shown alterations to microbiome structure and metabolism. However, differentiating the effects of individual fibers in humans with complex diets and heterogeneous lifestyles is challenging. Companion animals provide a particularly relevant context to study diet-microbiome interactions due to more consistent foods and environments. In this work, we investigated the gut microbial and metabolomic responses to various dietary fiber sources and quantities using a canine colony population. This design allowed us to study the association of specific microbial and metabolic responses with different carbohydrates including fiber and starch as well as the consistency of these associations across subjects.

## Design of dietary fiber study

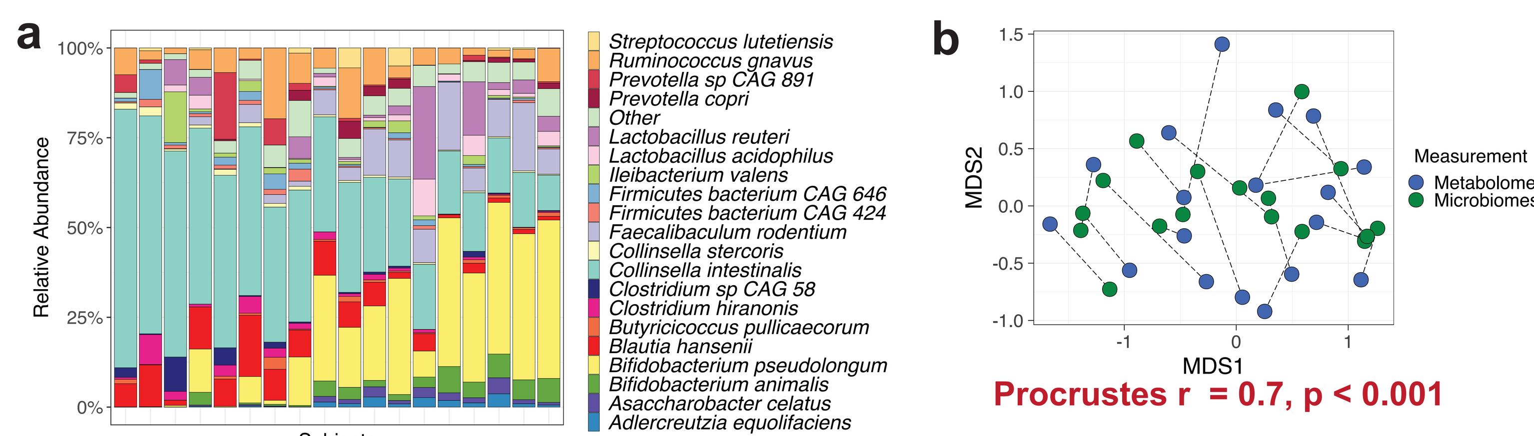



12 foods containing different sources and amounts of fiber were tested. They were classfied into 3 groups - high starch low fiber (HSLF) (n = 7), medium starch medium fiber (MSMF) (n = 3), and low starch high fiber (LSHF) (n = 2).




18 dogs were fed the 12 foods in a random order for 7 days each. HSLF_Con (control) was fed twice. Fecal samples collected on the last day of each treatment were used for metagenomic and metabolomic analyses.

## Baseline microbiomes and metabolomes are diverse

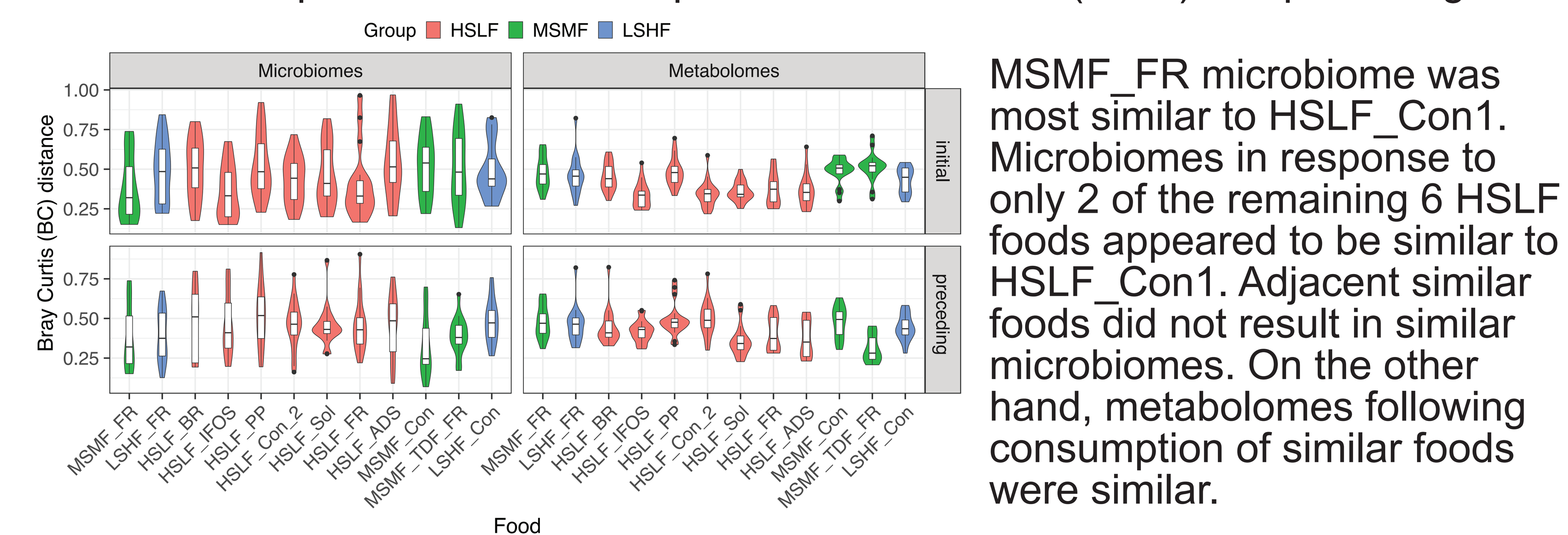



(a) *Collinsella* and *Bifidobacterium* sp were the most abundant species in baseline gut microbiomes of dogs followed by Firmicutes(36.98%) and Bacteroidetes(3.23%), Proteobacteria(0.44%), and Fusobacteria (0.005%). (b) Abundances of 818 metabolites were assayed from the same samples. Inter-individual variation in metabolomes was concordant with variation in microbiomes.

## Fiber affects the metabolome more than microbiome composition

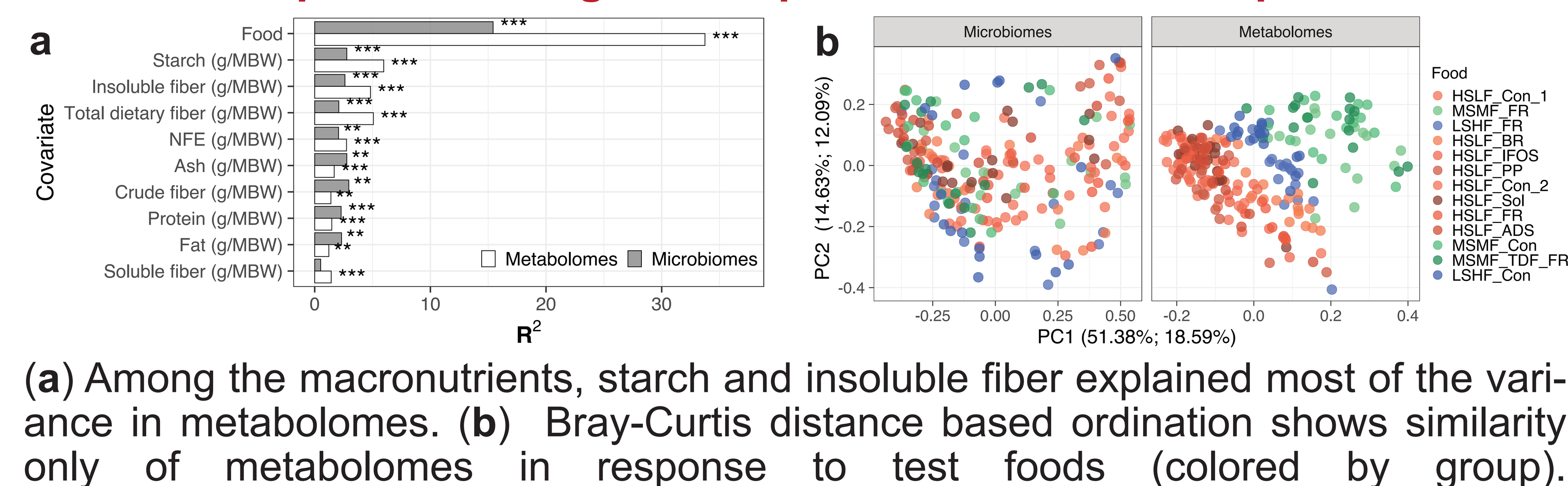
### Similar foods give rise to similar metabolomes but not microbiomes

We examined the extent to which microbiomes and metabolomes changed in response to food. Microbiomes/metabolomes following consumption of a particular food were compared to those in response to control food (initial) and preceding food.



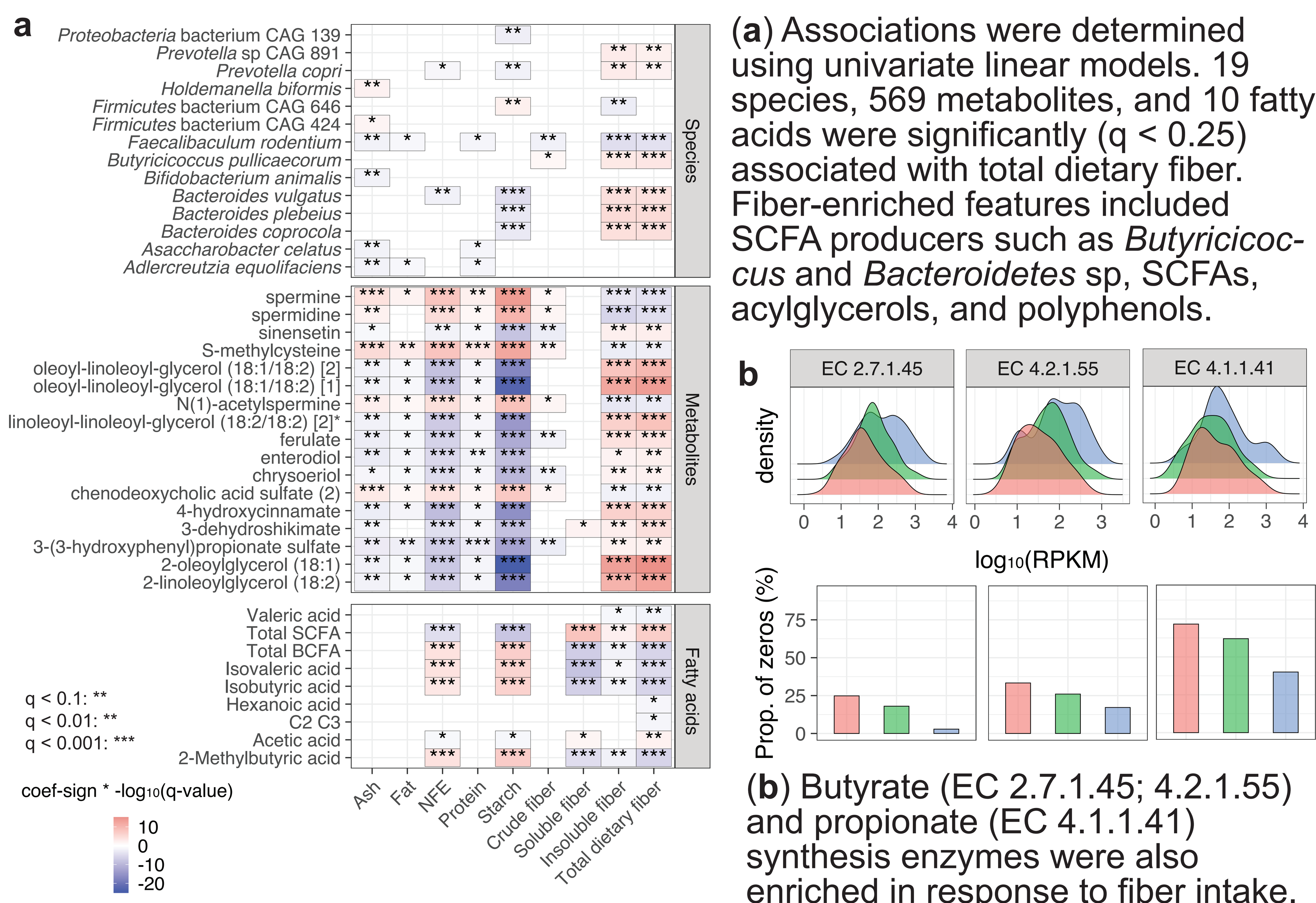

MSMF_FR microbiome was most similar to HSLF_Con1. Microbiomes in response to only 2 of the remaining 6 HSLF foods appeared to be similar to HSLF_Con1. Adjacent similar foods did not result in similar microbiomes. On the other hand, metabolomes following consumption of similar foods were similar.

### Food explains the largest component of variation in profiles




(a) Among the macronutrients, starch and insoluble fiber explained most of the variance in metabolomes. (b) Bray-Curtis distance based ordination shows similarity only of metabolomes in response to test foods (colored by group).

### Macronutrient-metabolite associations are stronger and more numerous than macronutrient-microbial feature associations

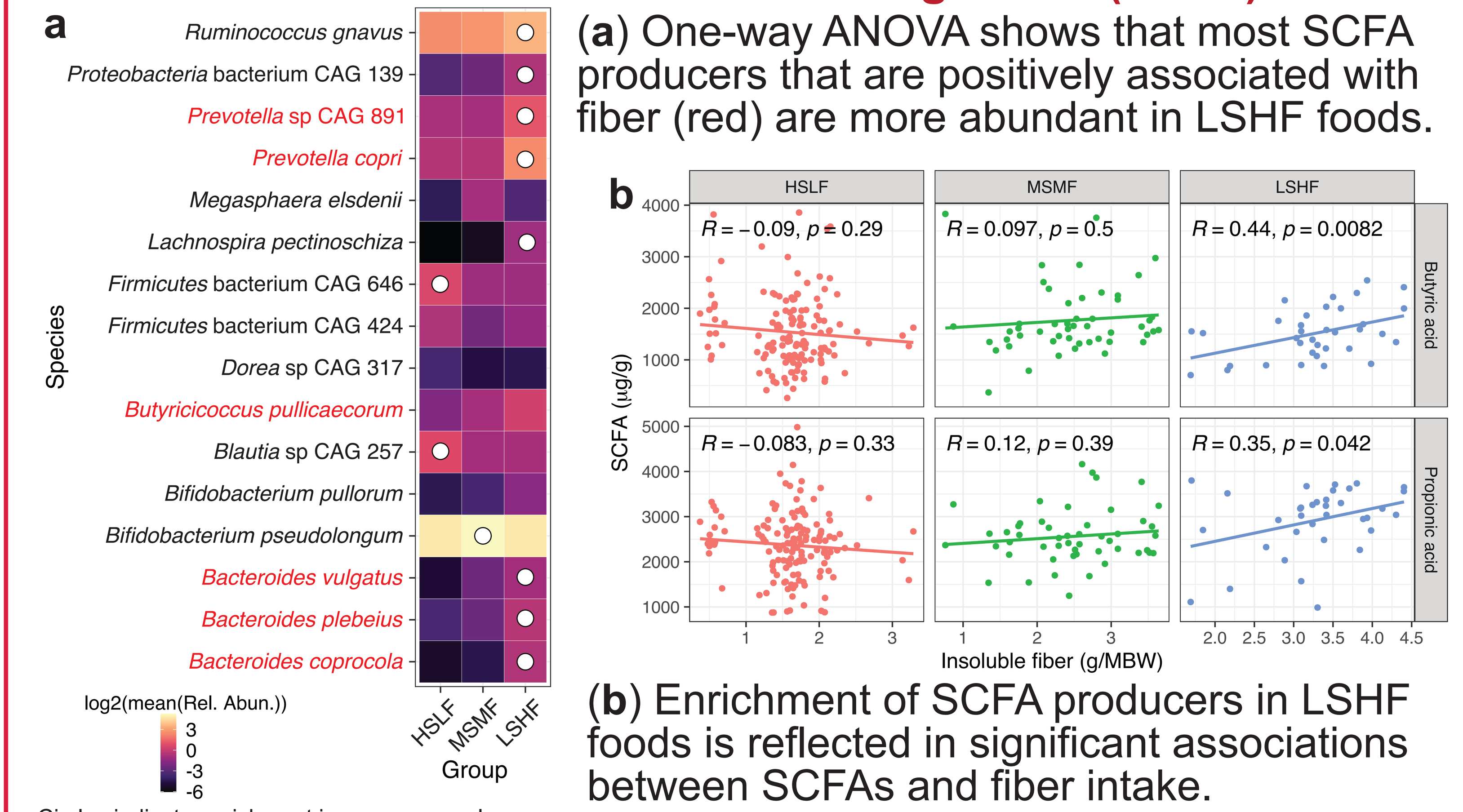



(a) Associations were determined using univariate linear models. 19 species, 569 metabolites, and 10 fatty acids were significantly (q < 0.25) associated with total dietary fiber. Fiber-enriched features included SCFA producers such as *Butyricicoccus* and *Bacteroidetes* sp, SCFAs, acylglycerols, and polyphenols.

(b) Butyrate (EC 2.7.1.45; 4.2.1.55) and propionate (EC 4.1.1.41) synthesis enzymes were also enriched in response to fiber intake.

## Strength of associations with fiber varies by food group and subject

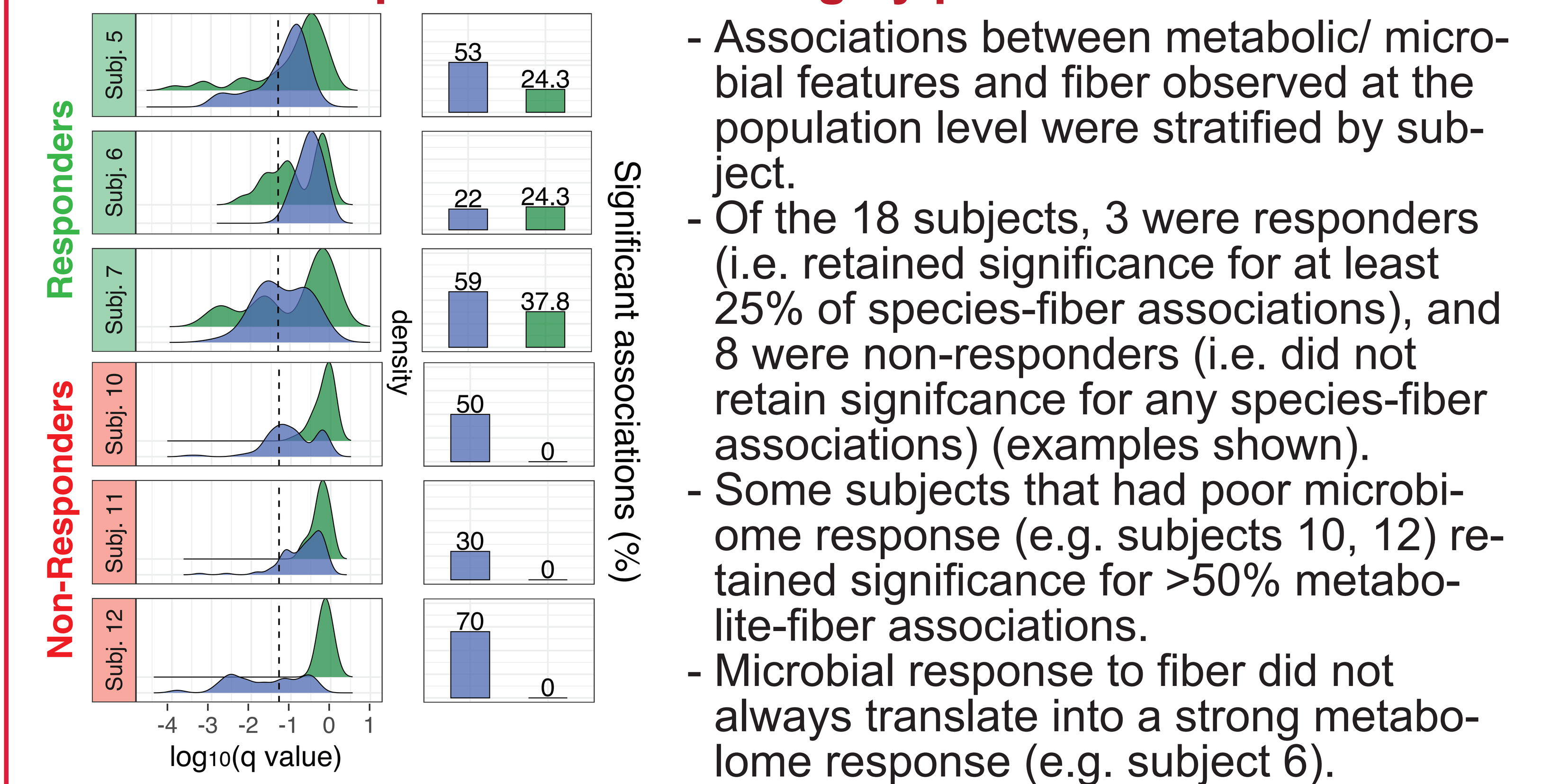
### Associations of fiber-responsive species and metabolites are more pronounced in low starch high fiber (LSHF) foods




(a) One-way ANOVA shows that most SCFA producers that are positively associated with fiber (red) are more abundant in LSHF foods.

(b) Enrichment of SCFA producers in LSHF foods is reflected in significant associations between SCFAs and fiber intake.

### Response to fiber is highly personalized




- Associations between metabolic/ microbial features and fiber observed at the population level were stratified by subject.
- Of the 18 subjects, 3 were responders (i.e. retained significance for at least 25% of species-fiber associations), and 8 were non-responders (i.e. did not retain signifcance for any species-fiber associations) (examples shown).
- Some subjects that had poor microbiome response (e.g. subjects 10, 12) retained significance for >50% metabolite-fiber associations.
- Microbial response to fiber did not always translate into a strong metabolome response (e.g. subject 6).

## Conclusions

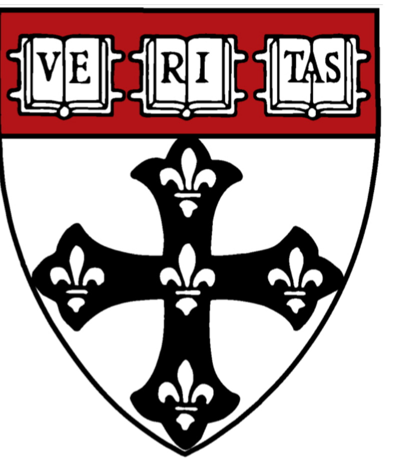
(1) Canine gut microbiomes and metabolomes change in response to diet. Similar foods are more likely to elicit similar metabolomic than microbiome responses. This suggests that different microbiomes can provide convergent metabolic potential to yield similar metabolomes from similar foods.
(2) Features are associated with fiber intake include SCFA producing species, SCFAs, and metabolites that are released upon fiber degradation such as acylglycerols and polyphenols.The strength of associations varies by both the type and quantity of fiber.
(3) Responses to fiber are subject-specific and cannot be predicted from intake or microbiome composition.

## Acknowledgements

We thank Hill's Pet Nutrition Inc. and the National Institutes of Health for funding this study and all of the pet partners and those who care for them.

http://huttenhower.sph.harvard.edu

# Host-Childcare Microbiome Interactions Highlighted by Using Long Read Sequencing




Marina Chen[1,2], Ya Wang[3,4,5], Kelsey N. Thompson[3,4,5], Jeremy E. Wilkinson[6], Jacob Nearing[3,4,5], David D. Christiani[1,7], Eric A. Franzosa[3,4,5], John D. Spengler[1], Curtis Huttenhower[3,4,5,8]

[1]Department of Environmental Health, Harvard T. H, Chan School of Public Health [2]Harvard Graduate School of Arts and Sciences [3]Harvard Chan Microbiome in Public Health Center
[4]Department of Biostatistics, Harvard T.H. Chan School of Public Health [5]Broad Institute of MIT and Harvard [6]Pacific Biosciences [7]Massachussetts General Hospital
[8]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health

Early-life exposure to microorganisms plays a crucial role in shaping children's health by instructing immune maturation and modulating the risk of disease development. Most preschool-aged children spend seven to ten hours a day in childcare centers, yet the microbial communities in childcare settings and how they interact with human and human-associated microbiomes have yet to be fully elucidated. A limited number of previous studies have primarily focused on bacterial communities, using amplicon-based profiling methods, largely due to the low biomass of these communities in this built-environment setting. These studies thus omit functional or genetic information and non-bacterial community members. Additionally, none of the prior research has incorporated host-associated phenotypes and prior infections to investigate transmission. Expanding upon these efforts, we collected a variety of indoor and outdoor environmental samples from two childcare centers, as well as nasal and oral swabs from 34 participating children aged two to four. We profiled all samples using PacBio full-length 16S rRNA gene and internal transcribed spacer (ITS) sequencing, respectively, for bacterial and fungal community members, and a subset of pooled samples using shotgun metagenomic sequencing including both short- and long-read metagenomics. This approach not only offered enhanced resolution to identify previously unknown aspects of the microbial communities in childcare environments, but also provided additional insight regarding the functional potential in the ecosystem. Current results revealed distinct microbial profiles associated with different host-associated and environmental communities. Most of shared species between host and environmental communities were identified to be specific host-associated ones, including those associated with host food consumption, such as Lactococcus lactis and Streptococcus thermophilus. These results suggest potential microbial transmission and interactions via host shedding. Our work thus expands the understanding of microbial ecology in childcare environments and these communities' relevance to childhood health. The knowledge gained from this study can allow us to identify potential environmental reservoirs of pathogens, track microbial transmission routes, and develop targeted interventions to benefit early-life human health.

## Study design and sample collection

### Study Participants

N = 50
Toddlers/Preschoolers
Age 3-5

Exclude if:
- Used oral or intravenous ABX or chemotherapy within 9 months
- Self report acute infectious disease (cold, gastroenteritis, etc.)
- Used topical ABX or antifungal applications nose and mouth within 1 week
- Have open wound in the nose or mouth
- Had infections within 1 week
- Had surgery involving the nasal or oral cavity within 9 months (cavity filling and routine dental cleaning are OK)

**Host samples (n = 68)**
34 children of age 2-4
- Oral swab n = 34
- Nasal swab n = 34
- Questionnaires

**Environmental samples (n = 109)**
8 classrooms, 2 childcare centers
- Doorknob n = 16
- Infiltration n = 6
- HVAC n = 2
- Desk n = 16
- Chair n = 16
- Toy n = 16
- Sink n = 8
- Drain n = 8
- Carpet n = 16
- Soil n = 5

*pooled subset*

PacBio full-length 16S and ITS amplicon sequencing

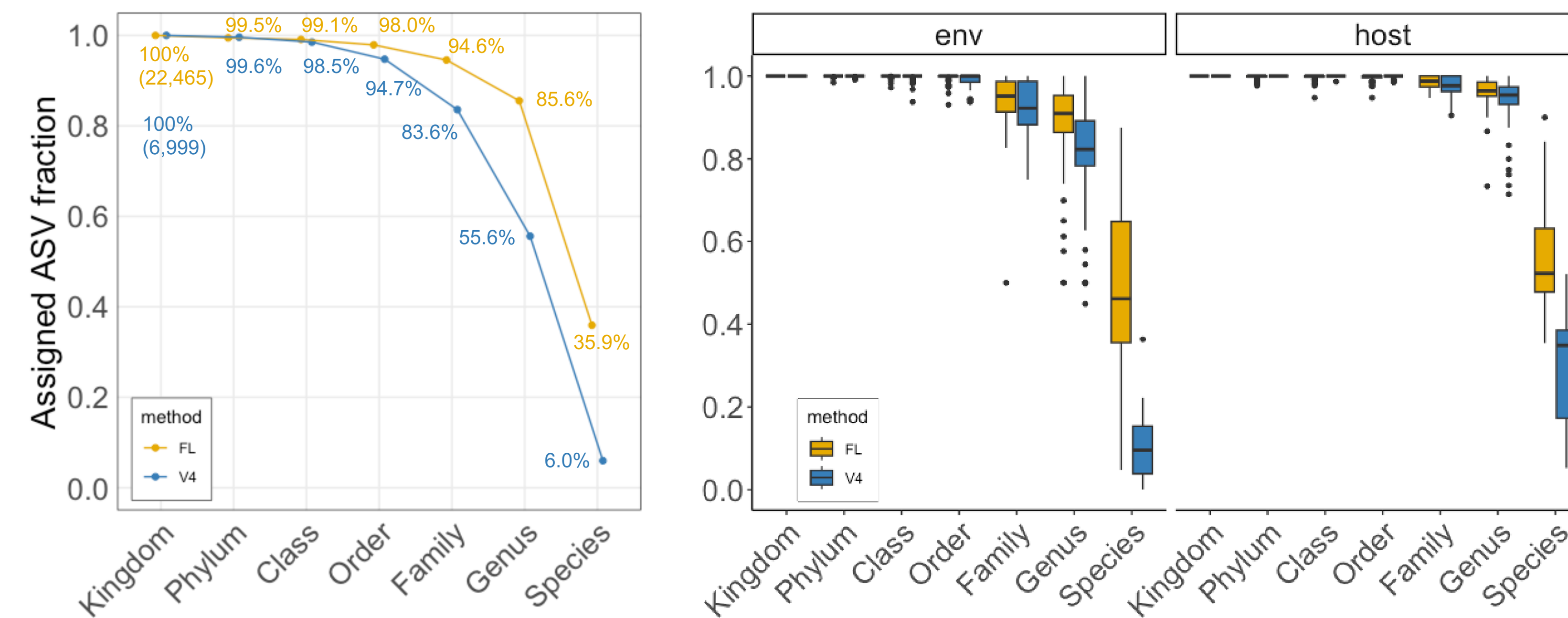
Illumina and PacBio shotgun metagenomic sequencing

**Amplicon analysis**
- Filter low depth samples and putative contam ASVs (*Decontam*)
- DADA2 profiling

**Metagenomic assembly**
- Short read: *bioBakery workflow*
- Long read: *hifiasm*
- Hybrid assembly: *hybridSPAdes*

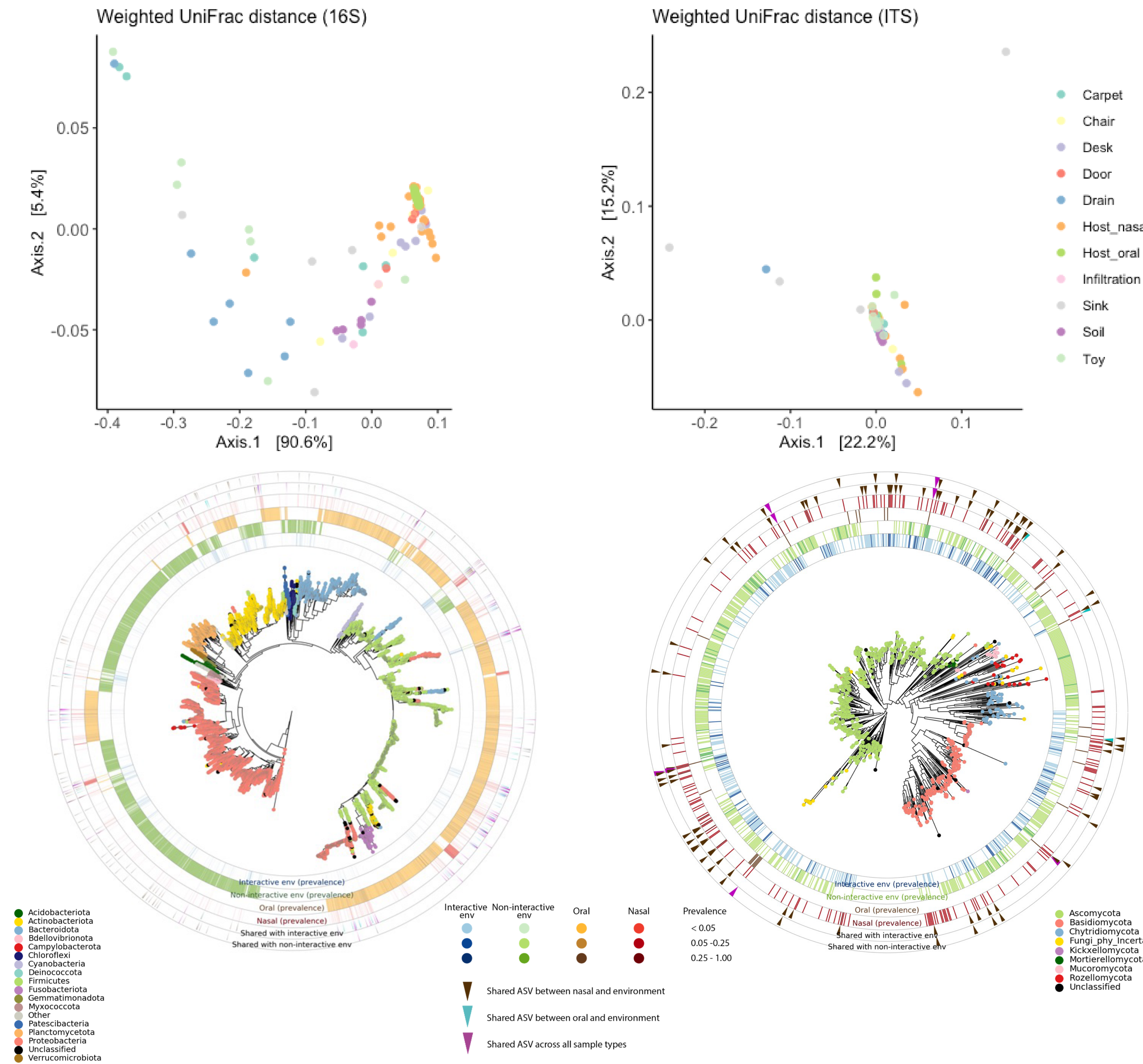
In total, we collected177 samples and 7 controls from two Harvard-affiliated childcare centers. Following sample and data filtering, 109 FL-16S and 61 FL-ITS samples were selected for subsequent amplicon-based analyses, respectively.

## Full-length 16S amplicons greatly improves species-level resolution




- FL-16S data yields ASVs which are approximately 3 times that from *in silico* V4 amplicons.
- Both types of amplicons are comparable in ASV taxonomic assignment at higer taxonomic levels, but diverge down to species level; FL-16 is found able to assign 30% more compared to V4 amplicons.

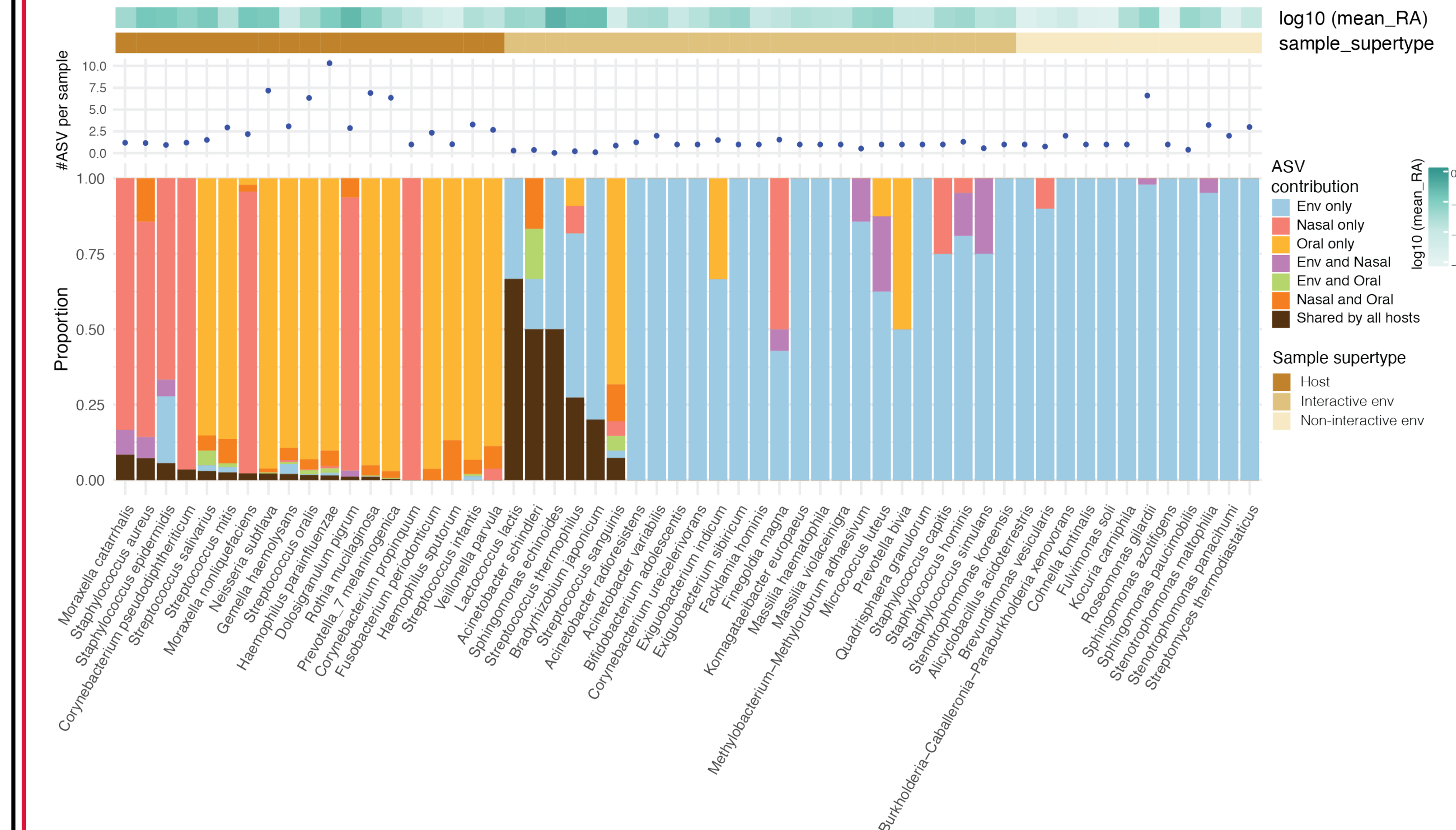
## Full-length amplicons reconstructs bacterial and fungal phylogeny




*Non-interactive environments include Drain, Sink, and Soil. Interactive environments include the rest of sample types.

- Environmrntal microbial communities exhibit greater phylogenetic diversity across different environmental sample types compared to host-associated communities.
- Full-length 16S and ITS amplicons generally reconstruct consistent phylogenetic relationships.
- Bacterial ASVs are found commonly shared between both oral/nasal communities and environments, whereas shared fungal ASVs are mostly restricted to nasal communities.

## Microbial interactions mainly driven by host shedding




Among the top 5 abundant and/or prevalent species, approximately one-third comprise ASVs universally shared between hosts and environments. The majority of these are host-associated, while those lacking shared ASVs predominantly consist of environmental ASVs. This suggests the shedding of host-associated microbial community members into the environment, with minimal uptake from environmental communities into the host microbiome.

Several food-associated species exhibit high levels of sharing, with ASVs (strains) likely originating from dairy products or other food sources. This is evidenced by their closer phylogenetic proximity to (non)dairy-associated reference sequences compared to the non-food outgroup.

## Acknowledgments

We gratefully acknowledge the support and assistance from Katy Donovan, Executive Director of Harvard's Campus Child Care, Inc., Trecia Mayo, Director of Soldiers Field Park Children's Center, and Barbara Carlson, Director of Radcliffe Child Care Center, in conducting this research, as well as all participating children and parents. This work has been supported by the Harvard Chan-NIEHS Center for Environmental Health (NIH; #P30ES000002) and Pacific Biosciences of California, Inc.

**http://huttenhower.sph.harvard.edu**

# Acetoacetate Regulates Anti-Tumor Immunity

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Slater L. Clay[1], Sena Bae[1], Geicho Nakatsu[1], Diogo Fonseca-Pereira[1], Monia Michaud[1], Meghan MacDonald[1], Jonathan N. Glickman[2], and Wendy S. Garrett[1]

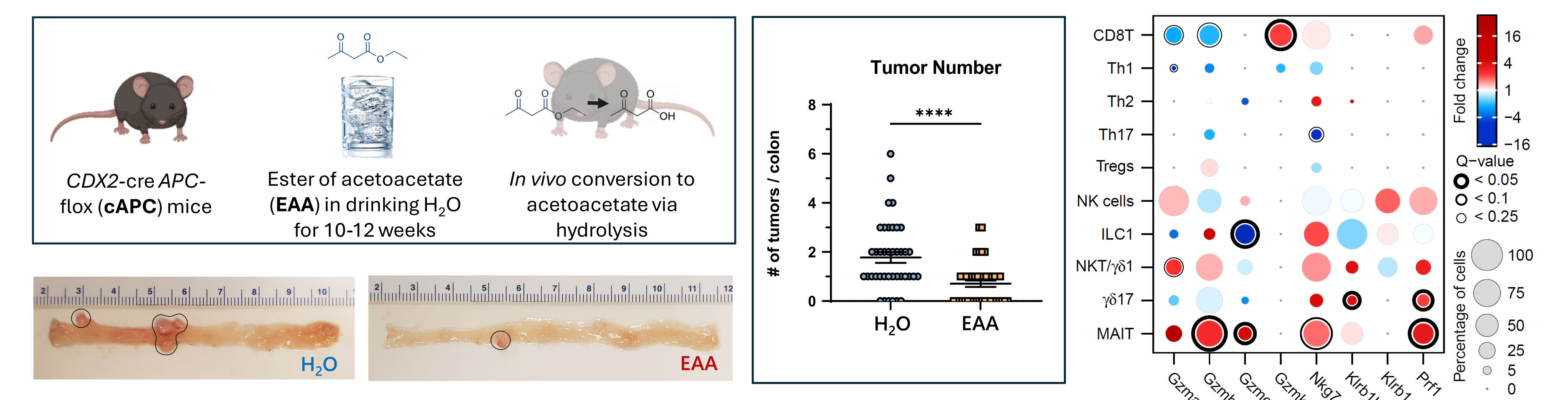
[1]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA

[2]Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA

## Background

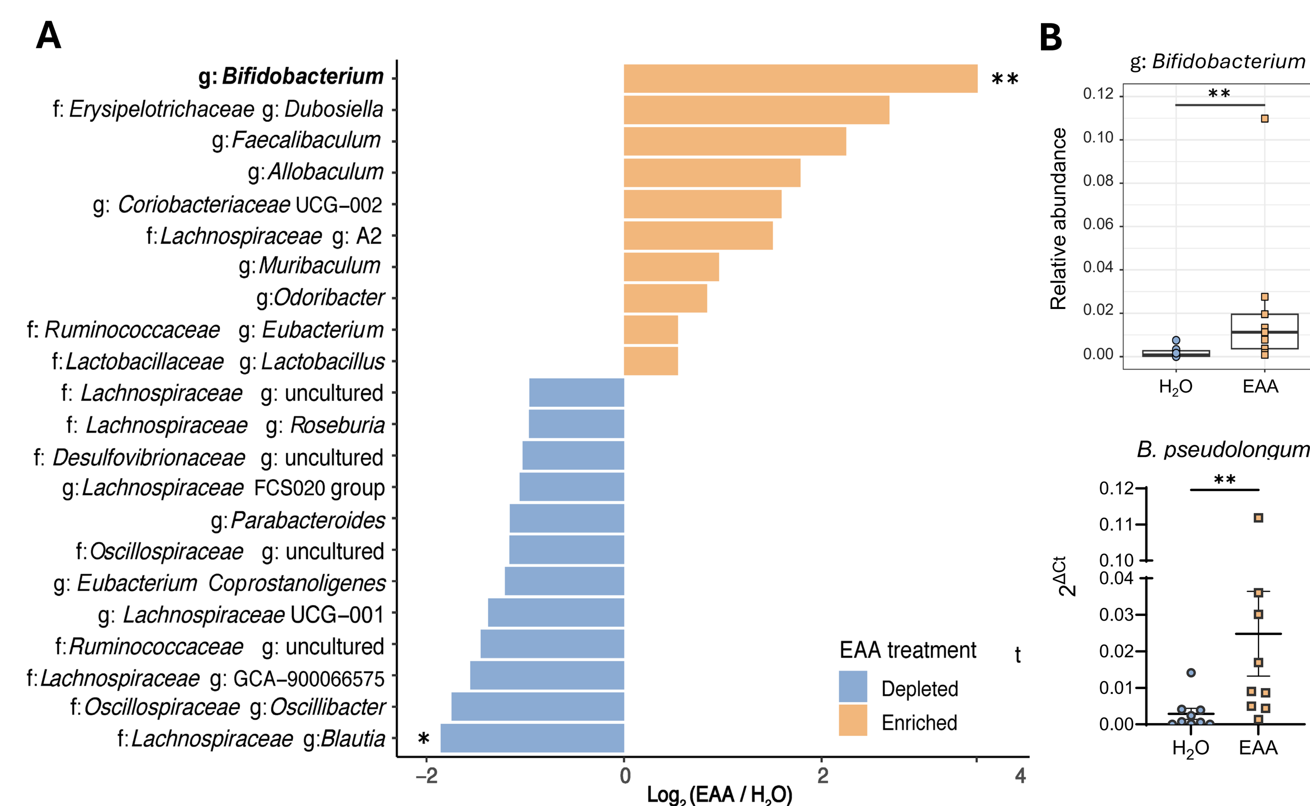
- Ketogenic diets and β-hydroxybutyrate (βHB) have been studied extensively as potential cancer therapy, but the effects of acetoacetate (AcAc) on tumor growth and anti-tumor immunity remains unclear.
- *In vivo* administration of an esterized acetoacetate (EAA) increases serum AcAc and reduces tumor growth in a genetic model of colorectal cancer (see panel right).
- EAA promotes expansion of mucosal associated invariant T (MAIT) cells and increased cytotoxic effectors (right).
- We investigated alterations to the fecal microbiota and tumor, colon, fecal and serum metabolites in response to AcAc. We also performed *in vivo* stable isotope tracing with $^{13}$C EAA to assess how AcAc is utilized in each tissue.
- Overall, we seek to identify altered microbial and host metabolic pathways affecting AcAc-mediated regulation of tumorigenesis

## Colorectal Cancer Model and Immune Regulation

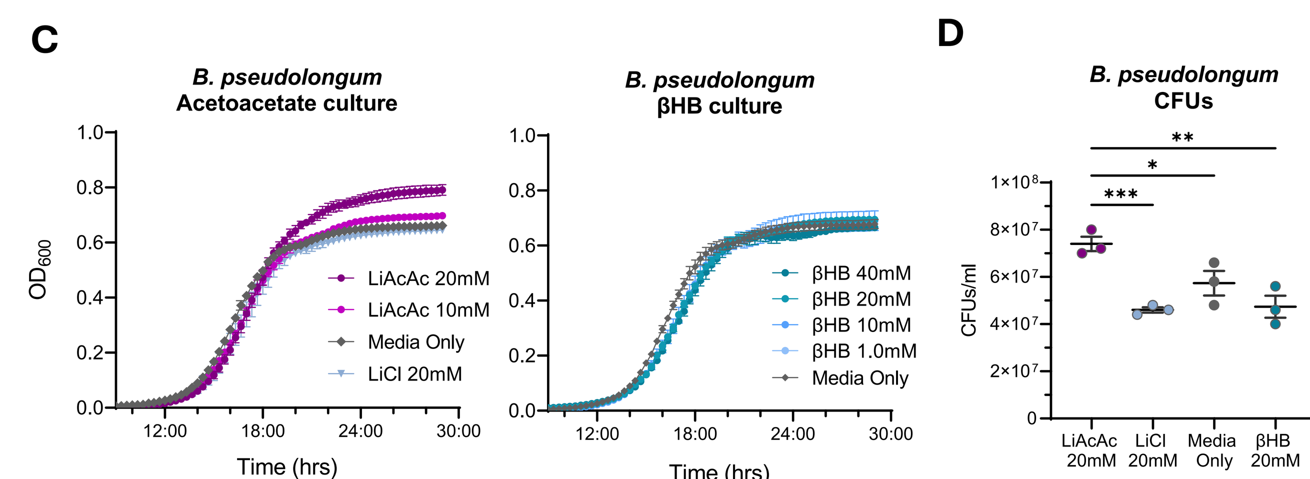



## Results

### EAA treatment alters the fecal microbiota of cAPC mice

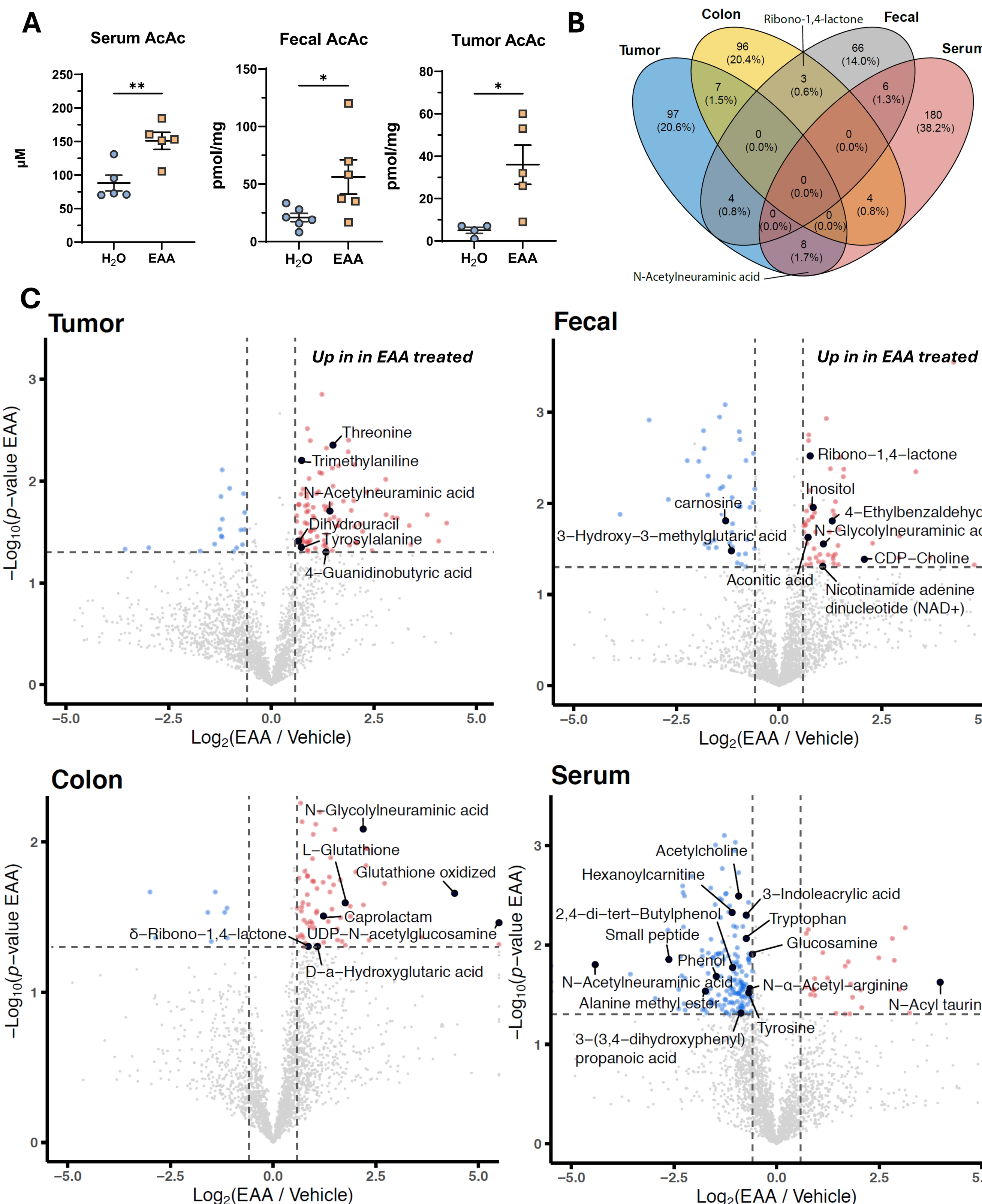



### Acetoacetate increases *Bifidobacteria* growth in culture



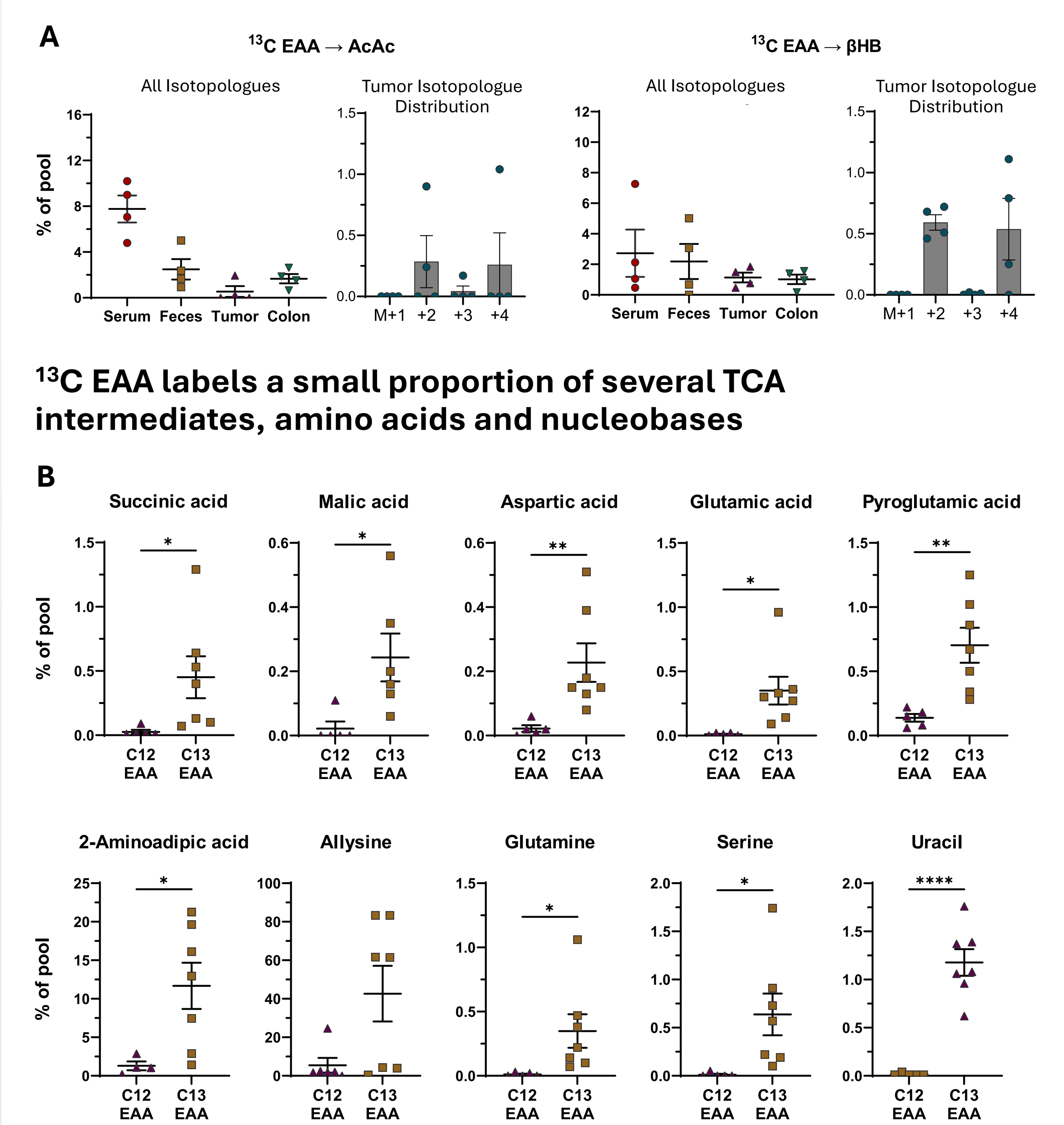

**A.** EAA treatment alters the composition of the fecal microbiota in cAPC mice. 16S rRNA amplicon sequencing was carried out on fecal samples. *Bifidobacteria* was the most increased taxa. **B.** Enrichment of *B. pseudolongum* was determined by BLAST and validated by qPCR. **C.** AcAc, but not βHB, increased the growth of *B. pseudolongum* but was not affected by. *B. pseudolongum* was cultured in MRS media in an anaerobic chamber with LiAcAc, LiCl control, NaβHB or media only for 30hrs. CFU was quantified by plating culture media on MRS plates.

### EAA alters the metabolome in a site-specific manner




**A.** EAA treatment increases AcAc in the serum, feces and tumor. **B.** Untargeted LC-MS/MS analysis identified 346 differentially abundant metabolites between EAA and Vehicle treated animals across four tissues. Only 32 of these are shared between tissues, and two are identified. **C.** Each tissue reveals distinct profiles of differentially abundant metabolites.

### Oral $^{13}$C EAA administration yields detectable signal in AcAc and βHB




**A.** Oral administration of $^{13}$C EAA allows tracing of $^{13}$C in AcAc and βHB. This is consistent with interconversion via *Bdh1* (see inset). The highest labelling is detected in serum, while lowest labelling is detected in the tumor and adjacent colon. **B.** Stable isotope tracing shows $^{13}$C incorporation into a range of metabolites including TCA cycle intermediates, amino acids, and uracil, although at a low proportion of the total pool for each metabolite.

## Summary

- EAA alters colonic microbial composition and increases *B. pseudolongum* abundance
- AcAc but not βHB increases growth of *B. pseudolongum*
- Treatment increases AcAc in serum, feces and tumors, and alters the metabolome of specific sites with distinct profiles of differentially abundant metabolites
- Oral administration of $^{13}$C EAA allows tracing of several TCA intermediates, amino acids and nucleobases- although at low proportions of the total pool
- Questions remain about how distinct cell populations utilize AcAc, and how the spatial distribution of metabolites relates to the location of immune populations

## Ongoing and Future Work

- *In vitro* co-cultures of immune and cancer cells, with ketones
- Metabolomics and $^{13}$C tracing on purified cell populations and culture supernatant
- Analysis of intra-tumoral microbiota using shotgun sequencing
- *In vitro* and *in vivo* gnotobiotic studies of intra-tumoral bacterial strains
- Spatial transcriptomics and mass spectrometry imaging of tumors to identify co-localization of cells and metabolites of interest.

## Acknowledgements

Thanks to the Harvard Center for Mass Spectrometry for metabolomics analysis. Thanks to the Dana Farber Cancer Institute and Cancer Grand Challenges for research funding. Thanks to the British Society of Immunology for travel funding.

Dana-Farber Cancer Institute | Department of Cancer Immunology & Virology
Cancer Grand Challenges
British Society for Immunology
Cancer Research UK | NIH National Cancer Institute

# Enhanced Feature Selection for Microbiome Data using FLORAL

**Teng Fei[1]*, Tyler Funnell[2], Nicholas Waters[2], Sandeep Raj[3], Marcel van den Brink[4]**

[1]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center (MSKCC) * **feit1[AT]mskcc[DOT]org**

[2]Department of Immunology, Sloan Kettering Institute, MSKCC [3]Department of Medicine, MSKCC [4]City of Hope Los Angeles and City of Hope National Medical Center

## Motivating Data Example

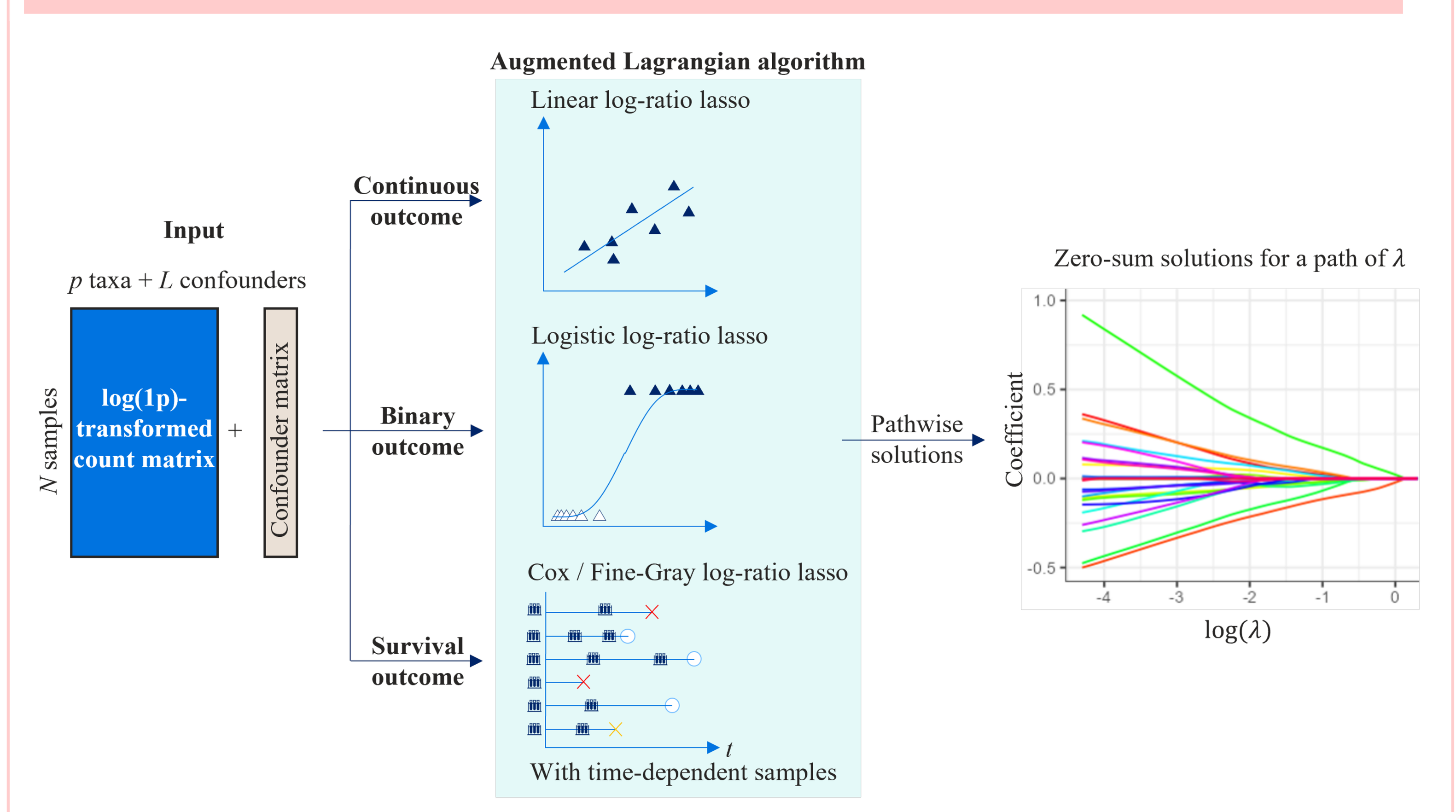
**MSK allogeneic hematopoietic cell transplant (allo-HCT) 16S rRNA data**

- 8,967 samples collected from 1,415 patients between transplant day [-30,730]
- Survival endpoints: overall survival (OS), transplant-related mortality (TRM), GvHD-related mortality (GRM); **TRM/GRM formulated as competing risks**

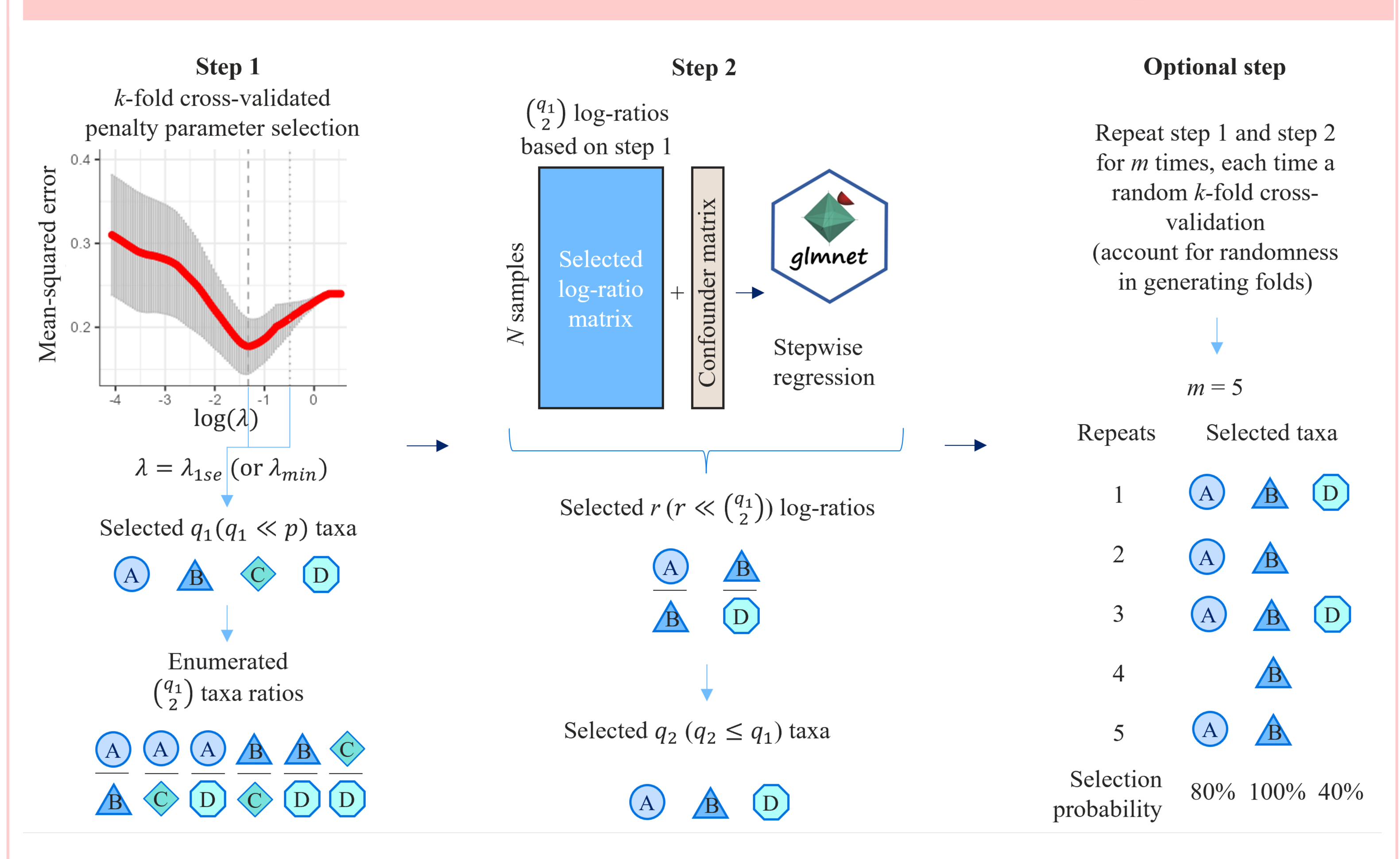
**Research question:** Whether and how the longitudinal microbial taxa features are associated with survival endpoints?

## FLORAL: Fit a LOg-RAtio Lasso

### Part 1: Zero-Sum Lasso Model Fitting with Compositional Taxa Features




### Part 2: Feature Selection Based on Cross-Validation and a Step 2 Procedure




## Variable Selection Performance by Simulation

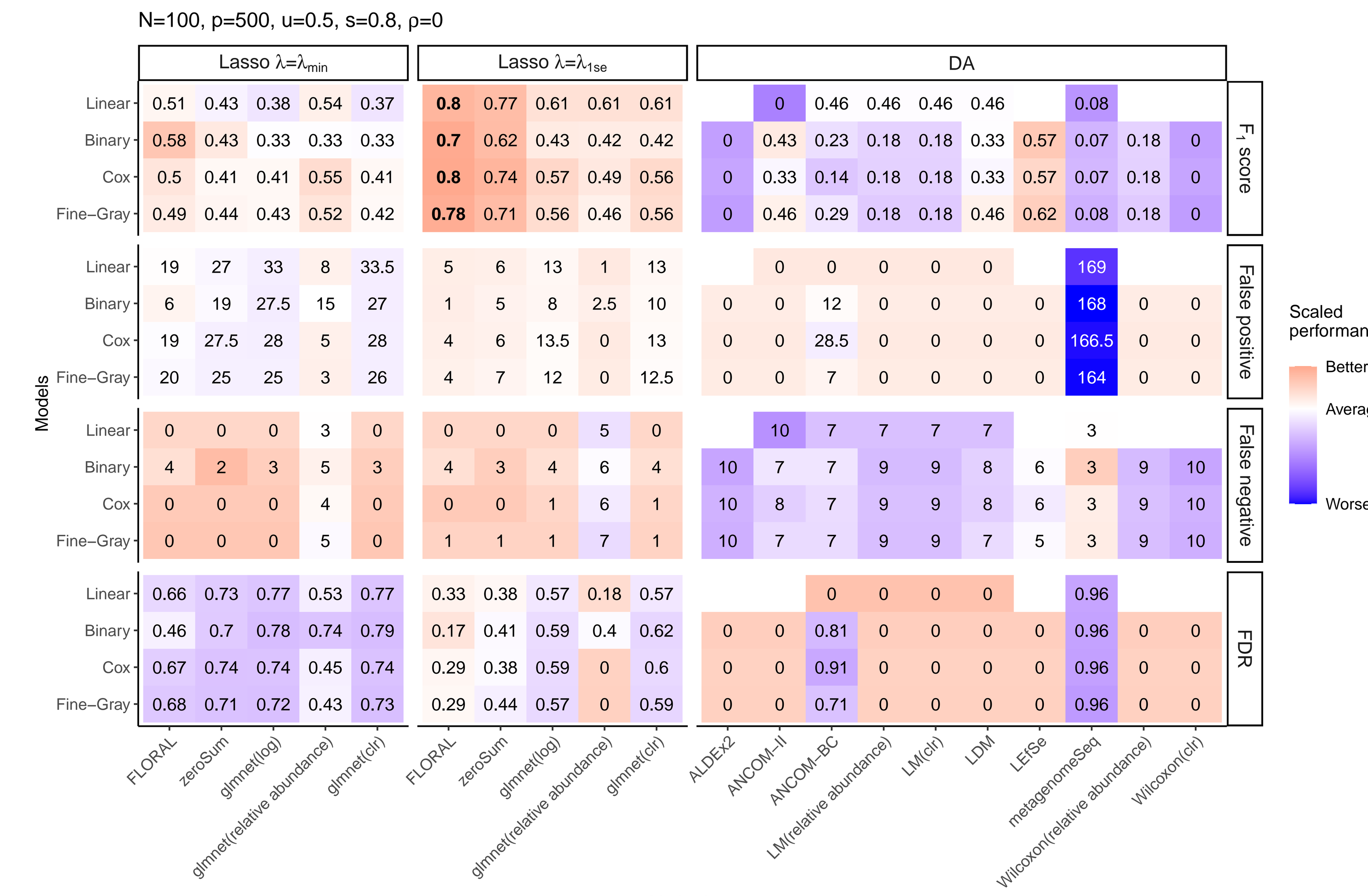



**Figure 1:** Median $F_1$ score, median number of false positive features, median number of false negative features, and median false discovery rate (FDR) obtained by lasso and differential abundance (DA) methods for linear, binary, survival, and competing risk models out of 100 simulations, where there were 10 true features out of $p = 500$ features in each simulation run. For each type of regression model, metrics across all methods were scaled for color visualization. For the DA methods, the censoring indicator of the survival outcomes were used to define patient groups except for LDM.

## Taxa Selection for MSK Allo-HCT Cohort

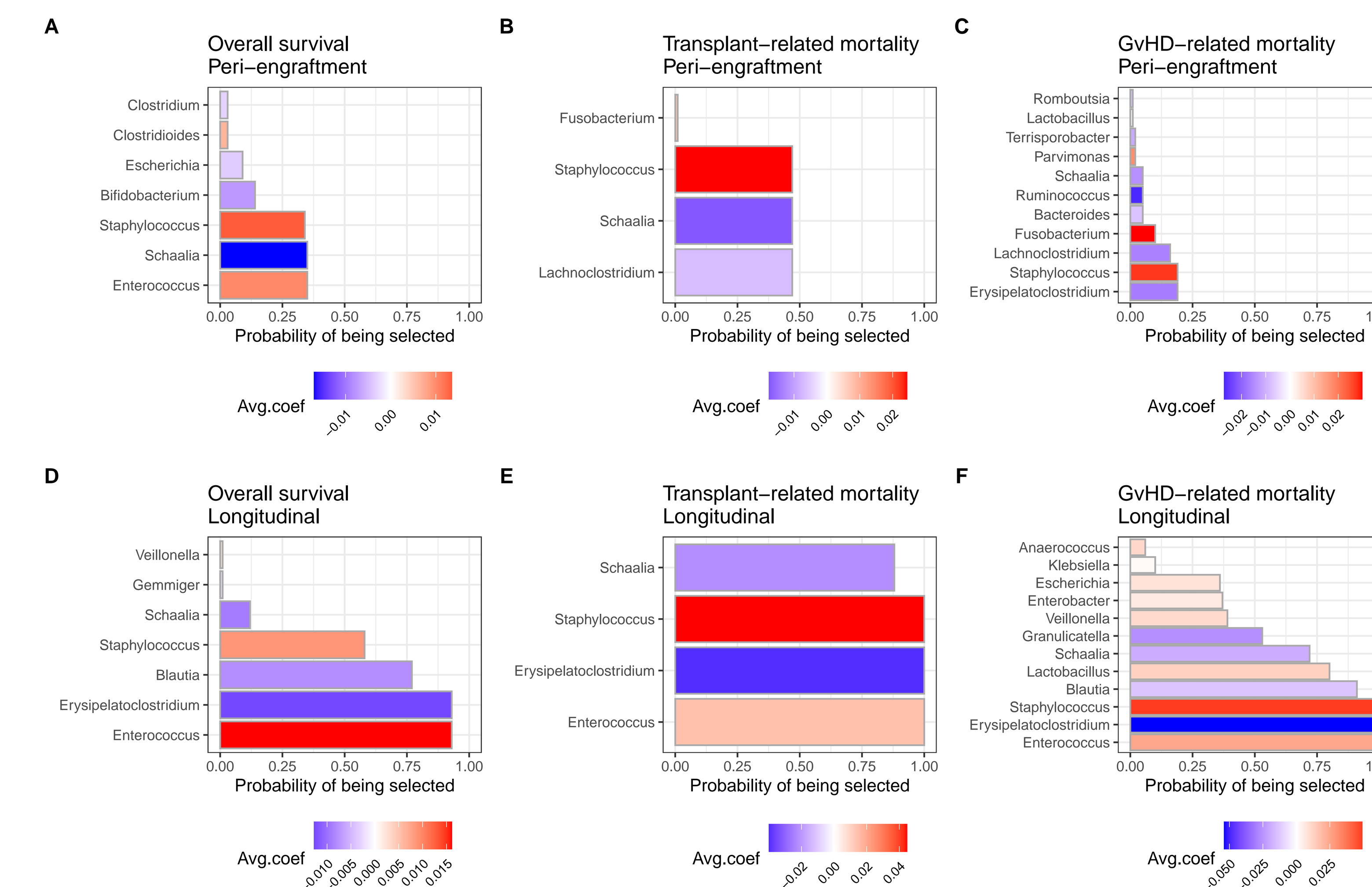



**Figure 2:** Probabilities of genera being selected by FLORAL with $\lambda = \lambda_{1se}$ and two-stage variable selection from the MSKCC allo-HCT cohort data, based on 100 repeats of 5-fold cross-validation with random fold split for Cox model of overall survival with **A.** peri-engraftment samples and **D.** longitudinal samples; Fine-Gray model of transplant-related mortality with **B.** peri-engraftment samples and **E.** longitudinal samples; Fine-Gray model of GvHD-related mortality with **C.** peri-engraftment samples and **F.** longitudinal samples. The color scheme represents the average lasso coefficient estimates of the corresponding genus at $\lambda^{(i)} = \lambda_{1se}$ over $i = 1, \ldots, 100$ repeats.

## False Positive Rates on 39 Publicly Available Datasets

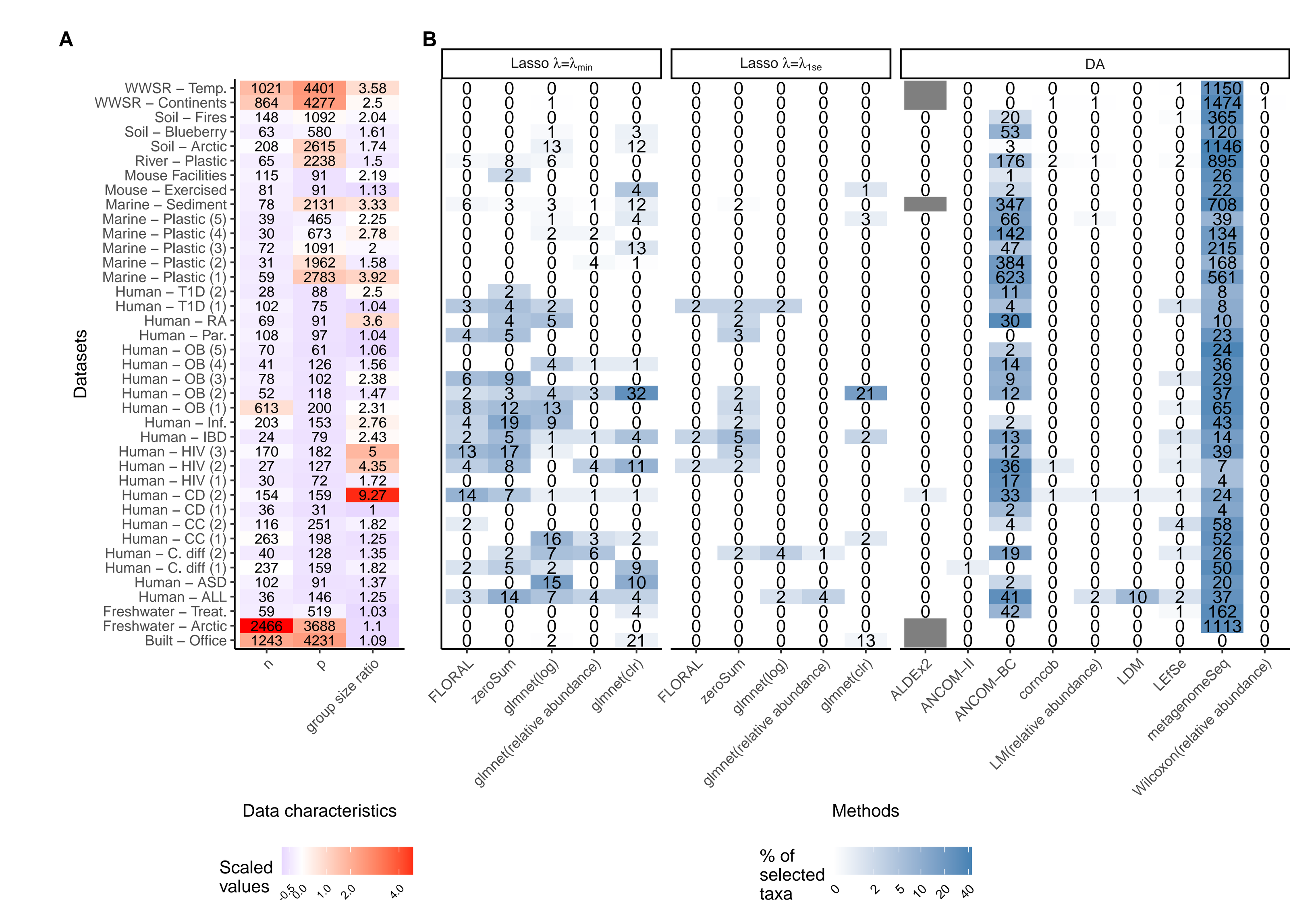



**Figure 3: A.** Data characteristics of the 39 publicly available 16S microbiome datasets, including sample size (n), number of genera (p), and ratio between the sizes of comparison groups. The color scheme represents scaled characteristics across all datasets. **B.** Number of selected taxa from the 39 publicly available 16S microbiome datasets by feature selection methods, with comparison group labels randomly shuffled. Part of data were unavailable for ALDEx2 due to memory overflow. The color scheme represents the percentage of selected taxa out of all taxa in a certain data set.

## Software Availability: https://github.com/vdblab/FLORAL

## Acknowledgements

This research was supported by National Cancer Institute award numbers, R01-CA228358, R01-CA228308, P30 CA008748 MSK Cancer Center Support Grant/Core Grant and P01-CA023766; National Heart, Lung, and Blood Institute (NHLBI) award number R01-HL123340 and R01-HL147584; National Institute on Aging award number P01-AG052359, and Tri-Institutional Stem Cell Initiative.

# Investigating the Effects of Probiotic Treatment on the Preterm Infant Gut Microbiome


Isabella M. Goodchild-Michelman [1,2,3], Shirin Moossavi [4], Emily Mercer [4], Marie-Claire Arrieta[4], Ali R. Zomorrodi [2,3]

[1] Department of Molecular and Cellular Biology, Harvard Faculty of Arts and Sciences, Boston, MA. [2] Harvard Medical School, Boston, MA. [3] Mucosal Immunology and Biology Research Center, Mass General Hospital for Children, Boston, MA.
[4] Department of Physiology and Pharmacology, University of Calgary , Alberta, CA [4]

MassGeneral Hospital for Children
Mucosal Immunology and Biology Research Center
VE RI TAS
MGH 1811


## Background

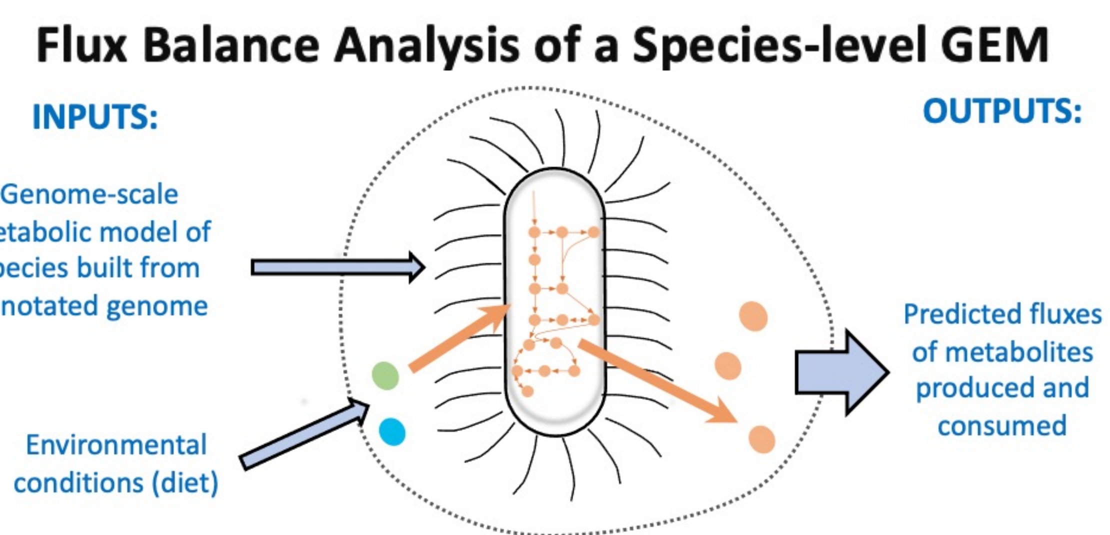
- **Infants born prematurely have an abnormal set of birth conditions that lead to a sparse, low-diversity population of microbes** initially colonizing their guts.
- **Prior studies have shown the ability of probiotic treatments to shift the preterm microbiome** to resemble that of a healthy, term infant, however, there is still little known about the functional mechanisms that underlie probiotic's therapeutic effects.
- All data used in this project is from the **BLOOM study, a longitudinal study on preterm infants run by the University of Calgary**

**Objective:** Use metagenomic and metadata from the BLOOM study to construct metabolic models of the preterm gut community to better understand the probiotic treatment's impact on the maturation of the preterm infant gut microbiota.

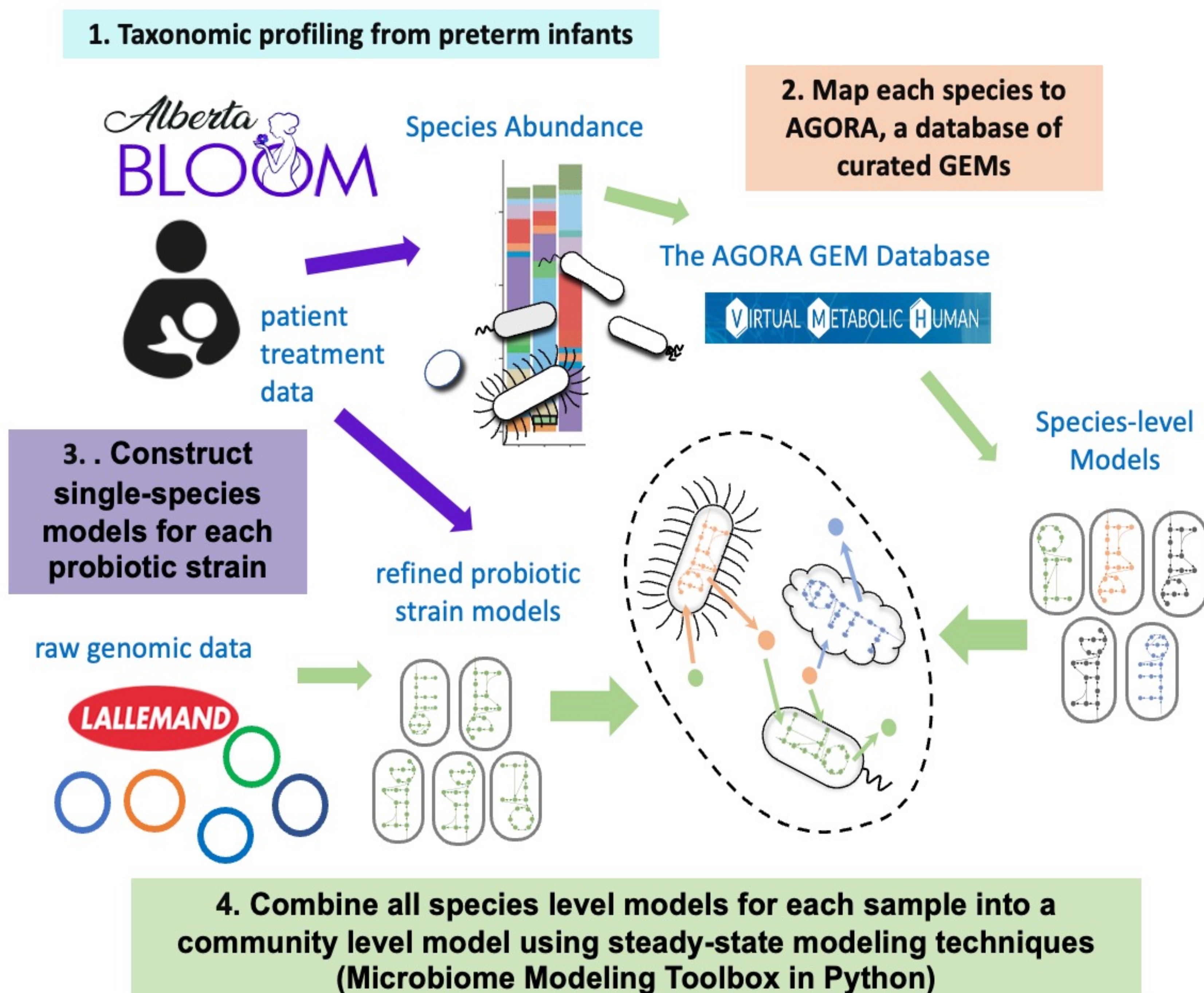
## Genome-Scale Metabolic Models

- Assume a cell can be approximated by the network of its metabolic pathways and can be analyzed to trace a metabolite's production back to a specific microbial species in the gut.

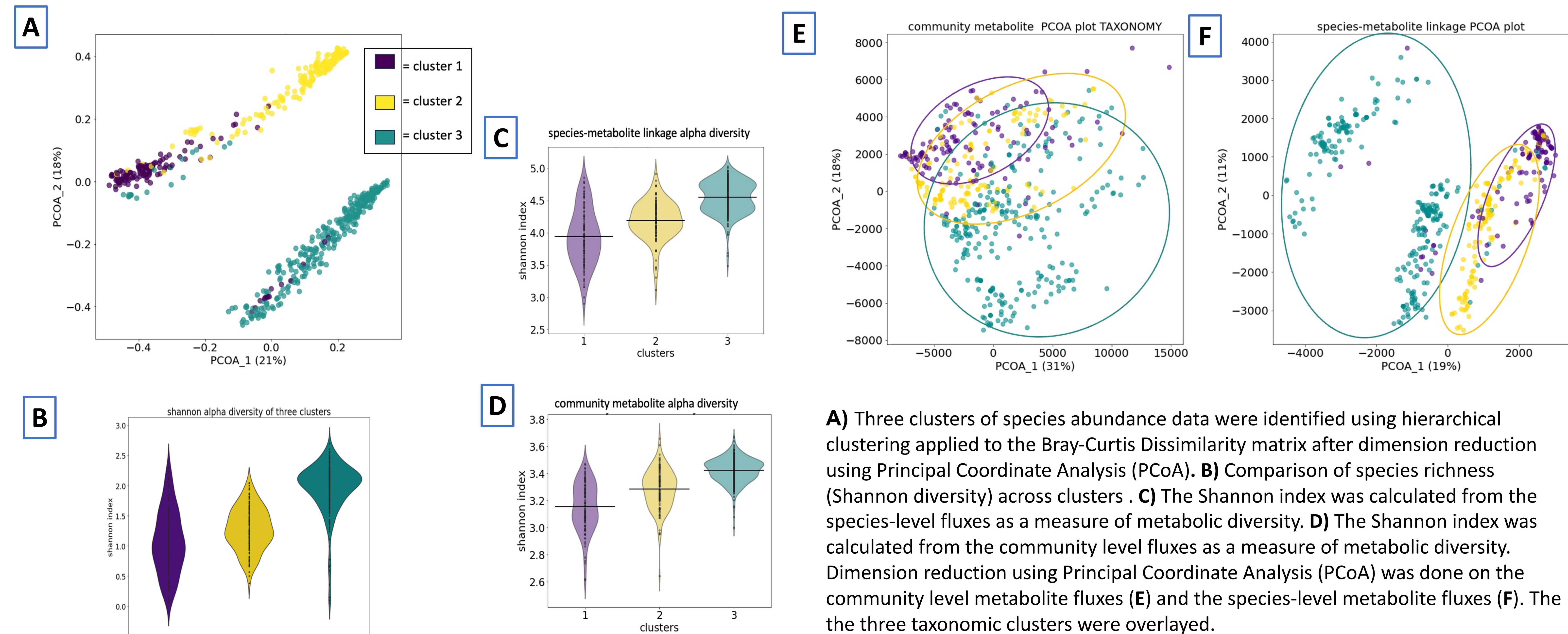
### Flux Balance Analysis of a Species-level GEM




## Methods

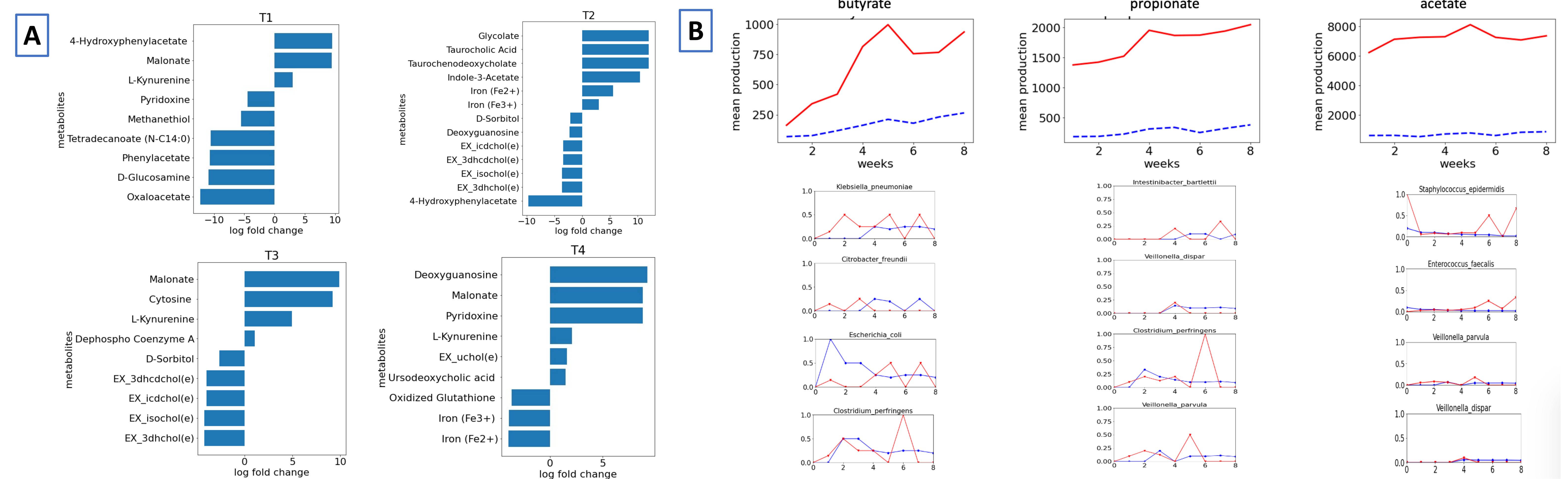



## Taxonomic and Metabolic Maturity Signatures




**A)** Three clusters of species abundance data were identified using hierarchical clustering applied to the Bray-Curtis Dissimilarity matrix after dimension reduction using Principal Coordinate Analysis (PCoA). **B)** Comparison of species richness (Shannon diversity) across clusters . **C)** The Shannon index was calculated from the species-level fluxes as a measure of metabolic diversity. **D)** The Shannon index was calculated from the community level fluxes as a measure of metabolic diversity. Dimension reduction using Principal Coordinate Analysis (PCoA) was done on the community level metabolite fluxes (**E**) and the species-level metabolite fluxes (**F**). The the three taxonomic clusters were overlayed.

## Species-Metabolite Linkages of Differentially Produced Metabolites




**A)** Fold change of metabolites only produced in probiotic or untreated group (a positive value indicates production in probiotic group) across four time bins. **B)** The three SCFAs were tracked at the community level and the species that contribute to their production at each time point are listed below

## Conclusion

**Our preliminary result support the feasibility of this study, and they will serve as a platform for large-scale computational studies of the host-microbiota interactions**

**Next Steps:**
- Large-scale simulations: Analyze all control vs probiotic microbiomes from BLOOM
- Test breastmilk vs formula diet in gut microbiome flux simulations

HONORHEALTH.

# Comparative Analysis of Skin Microbiome in Pancreatic Cancer Patients, Individuals with Other Cancers, and Cancer-Free Controls: A Pilot Study


Taylor Davis[1], Katherine Decker[2], Dana Hosseini[2], Gayle Jameson[1], Erkut Borazanci[1]

1: HonorHealth Research Institute; 2: ProdermIQ, Inc.



Supportive Fundations: HonorHealth Foundation


PRO DERM IQ

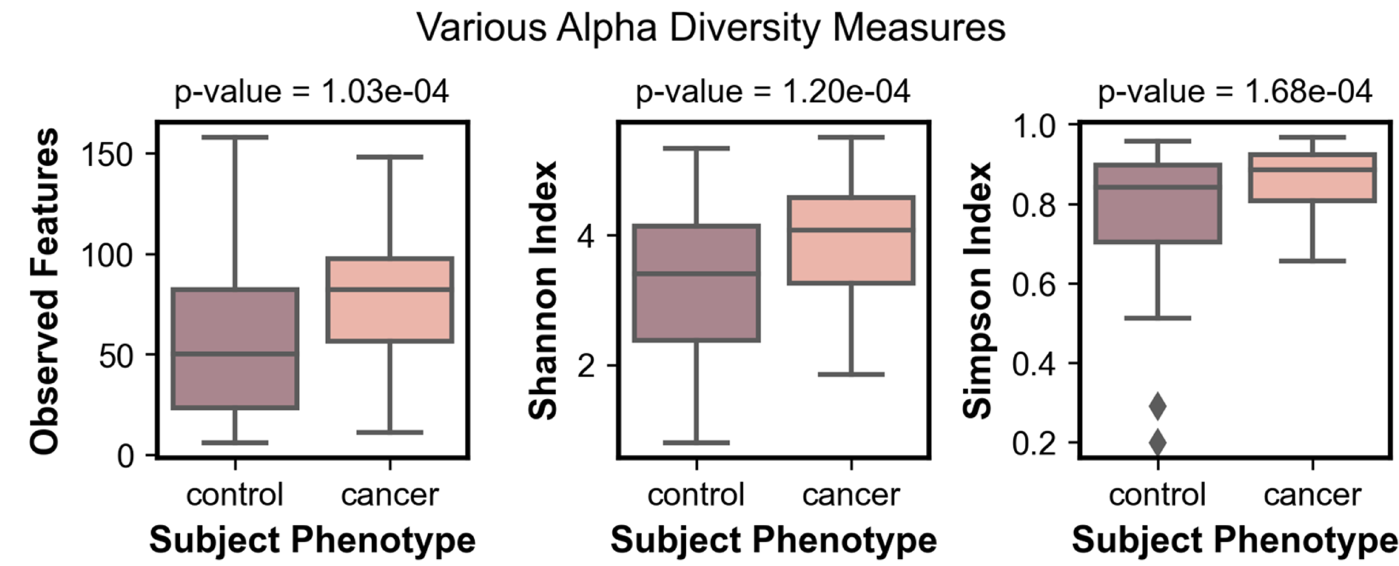
## Background

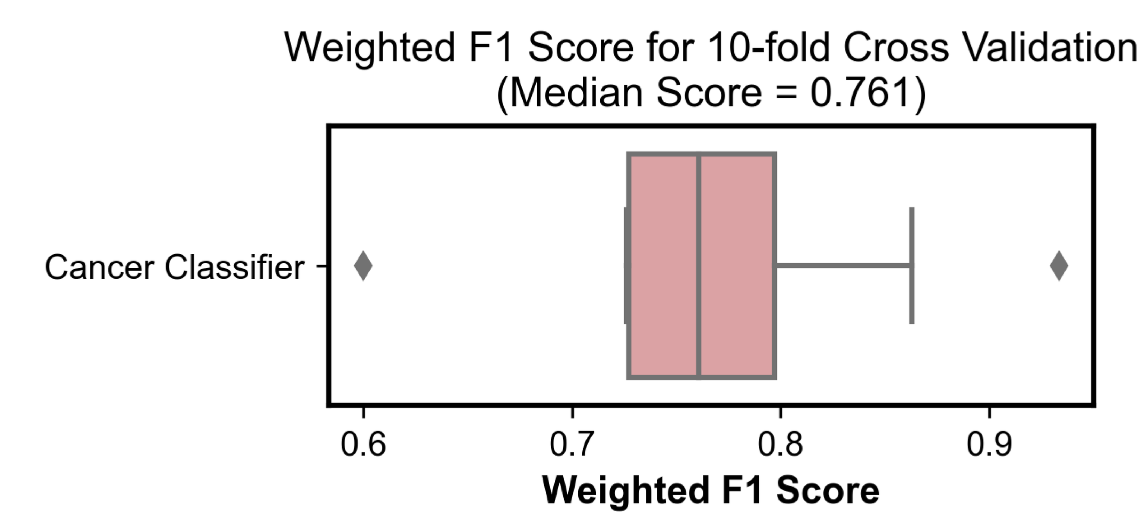
Several studies have reported the importance of the human microbiome in the overall health of its host. While recent studies have explored the microbiome's role in various types of cancer compared to healthy patients, this study narrows the focus to pancreatic cancer. This study aims to characterize the skin microbiomes on the forehead and cheek of individuals from three groups: 1) patients with pancreatic cancer, 2) patients with other forms of cancer, and 3) patients without any form of cancer. The goal is to determine if the results from this trial could provide insight on associations of microbial flora with the state or severity of cancer, status of host immune system, or progress of an ongoing therapy, which could have therapeutic applications.

## Methods

A total of 58 participants were enrolled in the study. Participants were given a questionnaire that prompted them to provide information including age, gender, ethnicity, race, weight, height, and status of skin health. An additional 60 control samples were drawn from an existing broader database of healthy skin samples at ProdermIQ to supplement the analysis. The participants were enrolled from three groups: cancer patients with pancreatic cancer, cancer patients with other types of cancer, and individuals without cancer. Skin microbiome samples from the forehead and cheek collection sites were processed and then analyzed by incorporating both statistical methods and machine learning techniques.

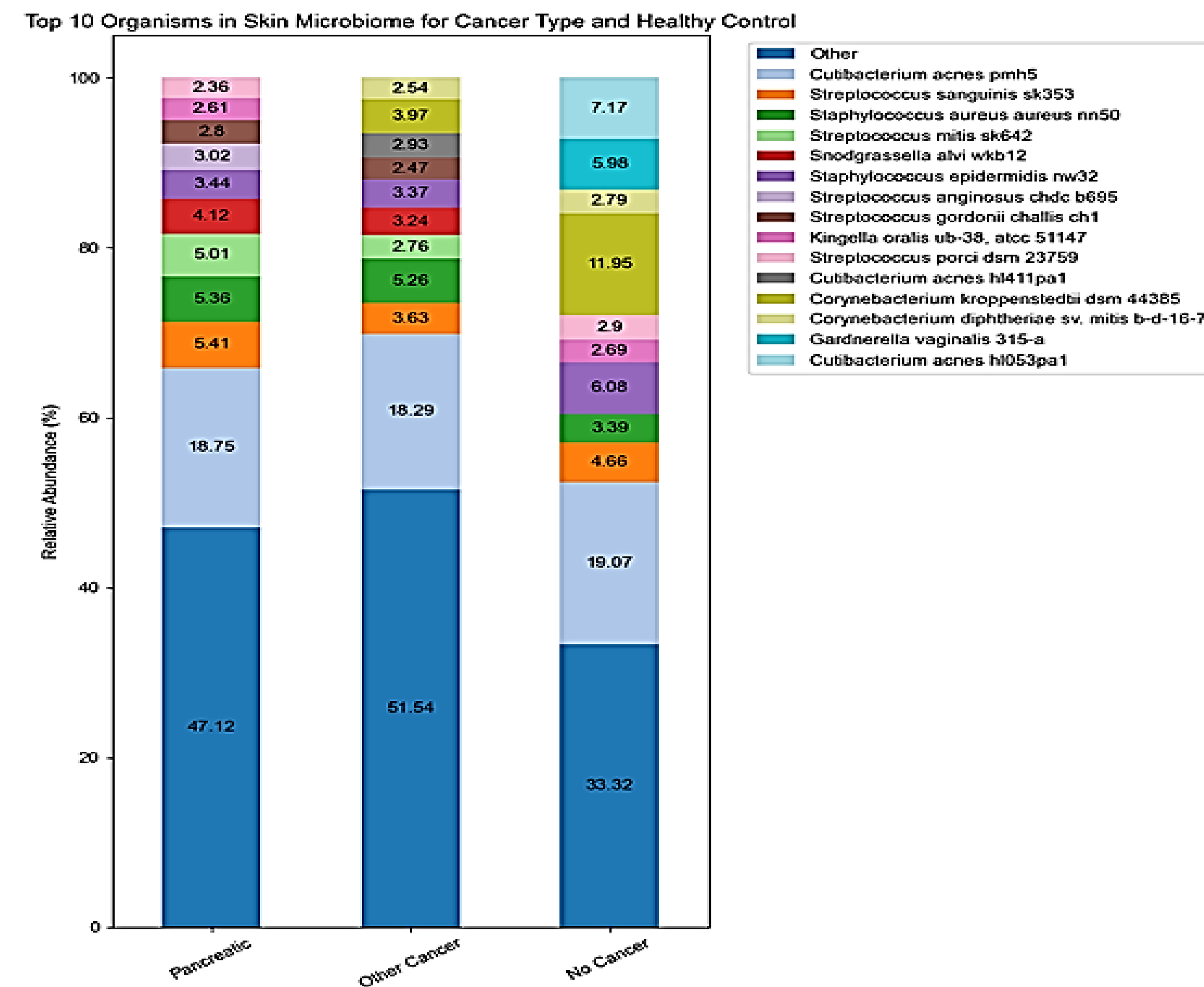
The skin microbiome of cancer subjects is more diverse than that of subjects without cancer, as indicated by various alpha diversity measures.




A machine learning classification model trained on the skin microbiome data achieved a good F1 Score (≥0.7) for predicting the status of cancer.

There are differences in microbe relative abundances observed for the three study groups: pancreatic cancer patients, patients with other cancer types, and healthy controls.




## Subject Demographics

| | Pancreatic Cancer (n=23) | Other Cancer (n=21) | No Cancer (n=14) |
|---|---|---|---|
| **Average Age (yr.)** | 65.8 | 59.8 | 39.6 |
| **Gender (M/F/Unlisted)** | 12/11/0 | 5/16/0 | 3/7/4 |
| **Smoker (Y/N/Unlisted)** | 1/22/0 | 0/21/0 | 0/10/4 |
| **Ethnicity** (Caucasian/Hispanic or Latino/Asian/African American/Native American/Unlisted) | 20/1/0/1/0/0 | 13/5/3/0/0/0 | 8/0/0/1/1/4 |

## Results

A total of 150 samples were analyzed, including 79 samples from subjects with cancer and 71 samples from control subjects. The mean age of the control group was 60 years, and the mean age of the cancer group was 63 years. Characterization of the two analysis groups was further refined using observed features and alpha diversity metrics. The cancer group displayed a significantly higher mean alpha diversity compared to the control group. Our analysis showed 15 dominant organisms as seen in the bottom figure. Organisms such as *Streptococcus mitis SK*642, *Snograssella alvi* wkB12, and *Streptococcus gordonii* Challis CH1 were seen in abundance within the pancreatic and other cancer groups but not within the no cancer group. *Streptococcus porci* DSM 23759 and *Kingella oralis* UB-38 were seen significantly within the pancreatic and no cancer group but not within the other cancer group. Additionally, a machine learning classification model built on the microbiome data demonstrated a median F1 Score of 0.761 for accurately classifying the cancer (all types) versus control samples. Given that F1 scores above 0.70 are generally regarded as satisfactory, this result indicates the skin microbiome can be predictive of cancer status.

## Conclusion

This analysis showed that there were significant differences in the skin microbiome of cancer patients versus patients without cancer. The cancer groups showed an increase in alpha diversity versus the no cancer group, and the machine learning model achieved a satisfactory F1 Score for differentiating the control and cancer samples . This could indicate the presence of dysbiosis in cancer subjects' skin microbiomes due to their clear differentiation from the healthy skin microbiomes. Additional research could provide potential opportunities to develop biomarkers that can identify pancreatic and other types of cancer.

## References


1. Deng T, Zheng H, Zhu Y, Liu M, He G, Li Y, Liu Y, Wu J, Cheng H. Emerging Trends and Focus in Human Skin Microbiome Over the Last Decade: A Bibliometric Analysis and Literature Review. Clin Cosmet Investig Dermatol. 2023 Aug 10;16:2153-2173. doi: 10.2147/CCID.S420386. PMID: 37583484; PMCID: PMC10424697.
2. Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. Nat Rev Microbiol. 2018 Mar;16(3):143-155. doi: 10.1038/nrmicro.2017.157. Epub 2018 Jan 15. PMID: 29332945.
3. Lee HJ, Kim M. Skin Barrier Function and the Microbiome. Int J Mol Sci. 2022 Oct 28;23(21):13071. doi: 10.3390/ijms232113071. PMID: 36361857; PMCID: PMC9654002.
4. Grice EA, Segre JA. The skin microbiome. Nat Rev Microbiol. 2011 Apr;9(4):244-53. doi: 10.1038/nrmicro2537. Erratum in: Nat Rev Microbiol. 2011 Aug;9(8):626. PMID: 21407241; PMCID: PMC3535073.
5. Eloe-Fadrosh EA, Rasko DA. The human microbiome: from symbiosis to pathogenesis. Annu Rev Med. 2013;64:145-63. doi: 10.1146/annurev-med-010312-133513. PMID: 23327521; PMCID: PMC3731629.
6. Wang Y, Yang G, You L, Yang J, Feng M, Qiu J, Zhao F, Liu Y, Cao Z, Zheng L, Zhang T, Zhao Y. Role of the microbiome in occurrence, development and treatment of pancreatic cancer. Mol Cancer. 2019 Dec 1;18(1):173. doi: 10.1186/s12943-019-1103-2. PMID: 31785619; PMCID: PMC6885316.
7. Bouferraa Y, Fares C, Bou Zerdan M, Boyce Kennedy L. Microbial Influences on Immune Checkpoint Inhibitor Response in Melanoma: The Interplay between Skin and Gut Microbiota. Int J Mol Sci. 2023 Jun 2;24(11):9702. doi: 10.3390/ijms24119702. PMID: 37298653; PMCID: PMC10253513.
8. Saud Hussein A, Ibraheem Salih N, Hashim Saadoon I. Effect of Microbiota in the Development of Breast Cancer. Arch Razi Inst. 2021 Oct 31;76(4):761-768. doi: 10.22092/ari.2021.355961.1750. PMID: 35096312; PMCID: PMC8790982.
9. Li, N. et al. The Relationship Between Gut Microbiome Features and Chemotherapy Response in Gastrointestinal Cancer. Front. Oncol. 11, 781697 (2021).
10. Chen, X.-H., Wang, A., Chu, A.-N., Gong, Y.-H. & Yuan, Y. Mucosa-Associated Microbiota in Gastric Cancer Tissues Compared With Non-cancer Tissues. Front. Microbiol. 10, 1261 (2019).
11. Kubinski, R. et al. Benchmark of Data Processing Methods and Machine Learning Models for Gut Microbiome-Based Diagnosis of Inflammatory Bowel Disease. Front. Genet. 13, 784397 (2022).
12. Tan-Torres, A. L., Jr, Brooks, J. P., Singh, B. & Seashols-Williams, S. Machine learning clustering and classification of human microbiome source body sites. Forensic Sci. Int. 328, 111008 (2021).
13. Kang, B. E. et al. Machine learning-derived gut microbiome signature predicts fatty liver disease in the presence of insulin resistance. Sci. Rep. 12, 21842 (2022).
14. Tiffany CR, Bäumler AJ. Dysbiosis: from fiction to function. Am J Physiol Gastrointest Liver Physiol. 2019 Nov 1;317(5):G602-G608. doi: 10.1152/ajpgi.00230.2019. Epub 2019 Sep 11. PMID: 31509433; PMCID: PMC6879887.
15. Ahluwalia J, Borok J, Haddock ES, Ahluwalia RS, Schwartz EW, Hosseini D, Amini S, Eichenfield LF. The microbiome in preadolescent acne: Assessment and prospective analysis of the influence of benzoyl peroxide. Pediatr Dermatol. 2019 Mar;36(2):200-206. doi: 10.1111/pde.13741. Epub 2019 Jan 18. PMID: 30656737.

# The Harvard T.H. Chan School of Public Health Microbiome Collection Core


Steven Medina[1], Curtis Huttenhower[1,2]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health

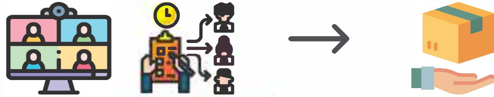
[2]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health


BROAD INSTITUTE

HCMPH MICROBIOME CENTER

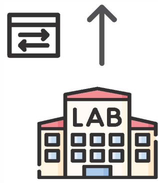
The Microbiome Collection Core at the Harvard T.H. Chan School of Public Health (HCMCC) was established in response to a strong demand among the research community for validated microbiome sample collection kit configurations and easy usability for in-home sampling. Under the umbrella of the Harvard Chan Microbiome in Public Health Center (HCMPH), HCMCC aims to support population-scale microbiome sample collection and expand our understanding of the microbiome to improve population health. The HCMCC has developed a multi carrier-compatible home stool and oral sample collection kit that permits cost-effective multi'omic microbiome studies, leveraging the intellectual and infrastructure foundation laid by the HMP2 (the 2nd phase of the NIH Human Microbiome Project) and the MLSC (Massachusetts Life Sciences Center)-funded MICRO-N (MICRObiome Among Nurses) collection. By providing this customizable microbiome collection kit, we enable researchers to perform multiple different molecular assays and tailor collection plan to study-specific needs.

## HCMCC services



**Microbiome sample collection plan development** - Collection kit configuration - Kit distribution & logistics - Sample transport plan- Sample handling & storage plan

**Kit ordering & shipment** - Kit customization & implementation - Ambient temperature shipping - to selected clinical sites - direct to participants



**Streamlined post-collection assistance** - Automated aliquoting - Barcode tracking - -80˚C storage in the BiOS Freezer - Fast sample retrieval - Sample shipment to sequencing labs for meta'omics & metabolomic profiling

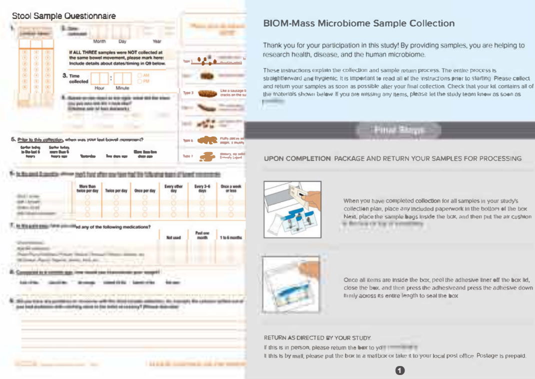
**At-home sample collection**

**Pre-paid return shipment**

The Microbiome Collection Core is a part of the Harvard Chan Microbiome in Public Health Center (HCMPH). Want to learn more? Visit **https://hcmph.sph.harvard.edu**

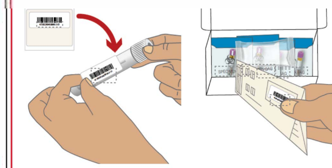
## A scalable gut and oral microbiome sample collection platform

This customizable microbiome sample collection kit avoids the need for expensive, bulky, and inconvenient ice packs by providing several different room temperature storage media that are also compatible with multiple different molecular assays including **any combination of amplicon (16S), metagenomic, metatranscriptomic sequencing, metabolomics, and other molecular assays**. This kit further includes a collection method that uses anaerobic transport media that **yields live microbes for culture or gnotobiotic research.**



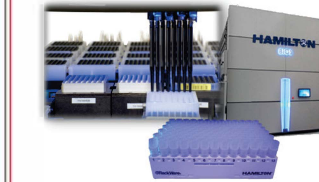

In addition to storage media, this sample collection kit includes **user-friendly instructions** and toilet accessories to maximumly facilitate and smooth the in-home stool sample collection experience. **Standardized questionnaires**, as companions to collected samples, are included to capture **recent medications, diet, anthropometric measurements, and gastrointestinal health status measured by the Bristol Stool Scale**. The modularity of this kit allows researchers to tailor kit components to study-specific needs and conduct cost-effective microbiome research ranging from **pilot studies to large-scale studies involving 10,000s of participants.**

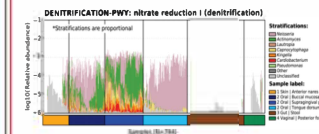
## HCMCC-supported study activities within the BIOM-Mass platform



**Pre-collection** - Participant enrollment - Kit ordering - Kit distribution

**Collection** - Self-collection - Sample return through pre paid shipment




**Post-collection** - Sample aliquoting via Hamilton STAR automated liquid handler - Long-term -80˚C storage via the BiOS Freezer Core - Data generation - Data analysis via the Microbiome Analysis Core

## Microbiome population health research opportunities

- Accessible microbiome population studies' data on the BIOM-Mass Data Portal **https://biom-mass.org**
- Integrative microbiome informatics and analysis via the Harvard Chan Microbiome Analysis Core **https://hcmph.sph.harvard.edu/hcmac/**
- Long-term sample storage via the Harvard Chan BiOS Freezer Core
- Gnotobiotic mice experiments via the Harvard Chan Gnotobiotic Center for Mechanistic Microbiome Studies
- Course offerings on microbial communities and human microbiome research via the Harvard Chan Microbiome in Public Health Center


Special thanks to the Massachusetts Life Sciences Center (MLSC), the Harvard Chan Microbiome Platform Steering Committee, the Harvard Chan Freezer Core Director John Obrycki, and the BWH/Harvard Cohorts Biorepository Laboratory Manager Christine Everett.


## Contact us:


hcmcc@hsph.harvard.edu
Manger: Steven Medina
Scientific Director: Curtis Huttenhower

**https://hcmph.sph.harvard.edu/hcmcc (i)**

**Follow us on Twitter @hutlab (ii)**

HARVARD MEDICAL SCHOOL | BRIGHAM HEALTH BRIGHAM AND WOMEN'S HOSPITAL | BROAD INSTITUTE

# A Cross-Cohort Multi-Omics Study of Diet, Gut Microbiome, and Metabolomics in Type 2 Diabetes in MicroCardio Consortium

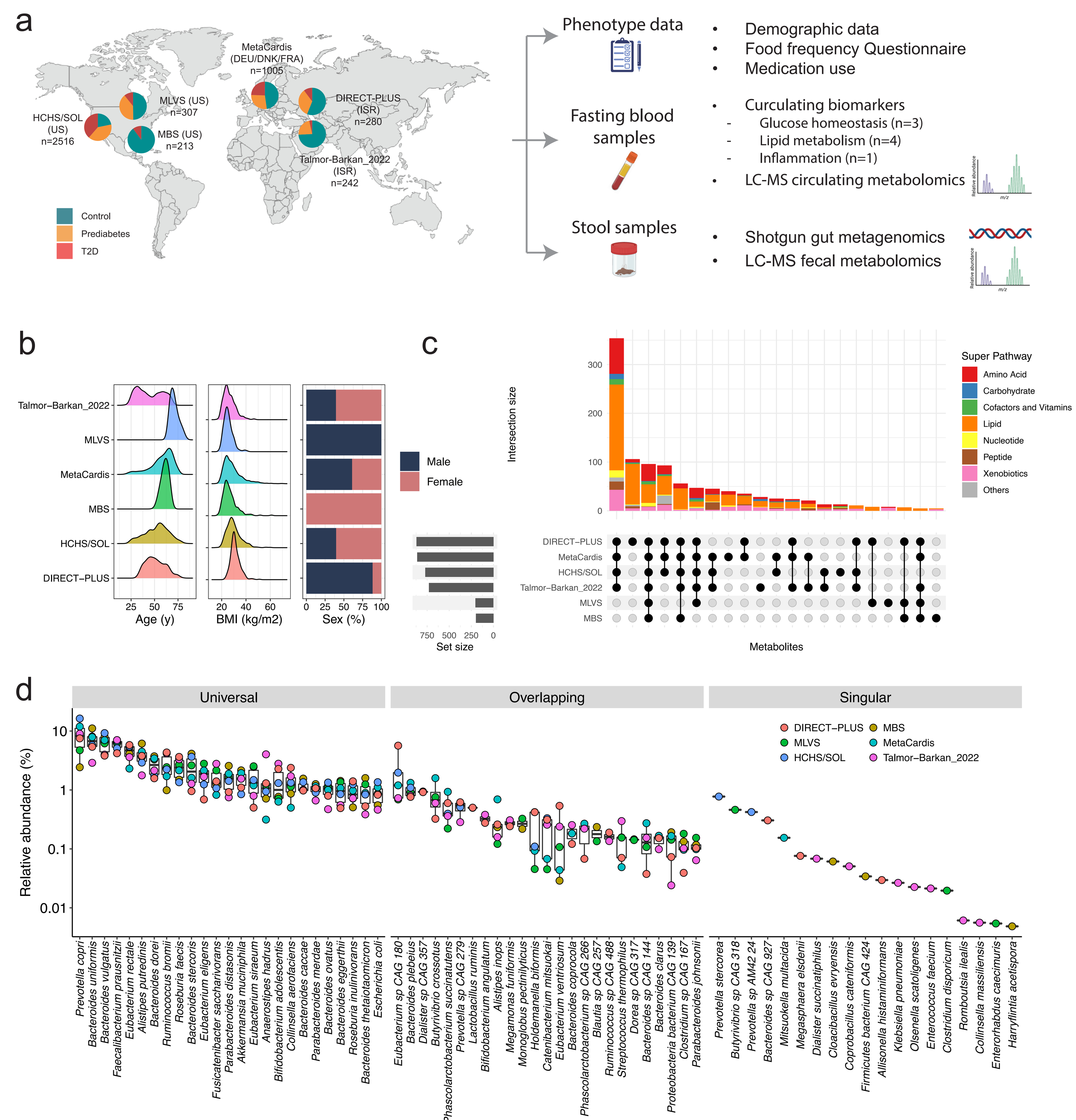

**Zhendong Mei[1,2], Amrisha Bhosle[2,3], Fenglei Wang[2,4], Danyue Dong[1,2], Jacob T. Nearing[2,3], Qibin Qi[4,5], Iris Shai[4,6], Robert D. Burk[5,7,8,9], Robert C. Kaplan[5,10], Curtis Huttenhower[2,3,11,12], Dong D. Wang[1,2,4]**

[1]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; [2]Broad Institute of MIT and Harvard, Cambridge, MA; Departments of [3]Biostatistics, [4]Nutrition, and [11]Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA; Departments of [5]Epidemiology and Population Health, [7]Pediatrics, [8]Microbiology & Immunology, and [9]Obstetrics, Gynecology & Women's Health, Albert Einstein College of Medicine, Bronx, NY; [6]Faculty of Health Sciences, The Health & Nutrition Innovative International Research Center, Ben-Gurion University of the Negev, Be'er Sheva, Israel; [10]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; [12]Harvard Chan Microbiome in Public Health Center, Harvard T.H. Chan School of Public Health, Boston, MA



The gut microbiome, interacting with dietary intake, modulates host metabolism and contributes to the pathogenesis of type 2 diabetes (T2D). Yet, large-scale multi-omics studies to examine these complex interactions are limited. Applying a validated data harmonization pipeline, we conducted a comprehensive study that integrates data on long-term habitual diet, gut microbiome, and circulating metabolomes from six studies of 4,563 participants with T2D, prediabetes, and normoglycemic status in the US, Europe, and Israel. Our analysis identified diet- and host-derived metabolites (e.g., quinate) and microbial-host co-metabolites (e.g., cinnamoylglycine and indole propionate) associated with T2D, independent of major risk factors. We also identified interactions between microbe and metabolite implicated in T2D risk. In addition, the inter-individual difference in the association of species (such as *Roseburia inulinivorans*) with T2D risk could potentially be explained by the strain-specific processing of metabolite implicated in the pathogenesis of T2D. Our study offers robust insights into the intricate interplay of diet, gut microbes, and their metabolites underlying the development of T2D in diverse populations.


## Overview of metagenomic, metabolomic, and phenotypic data in 1259 T2D patients, 1577 prediabetic and 1727 normoglycemic individuals

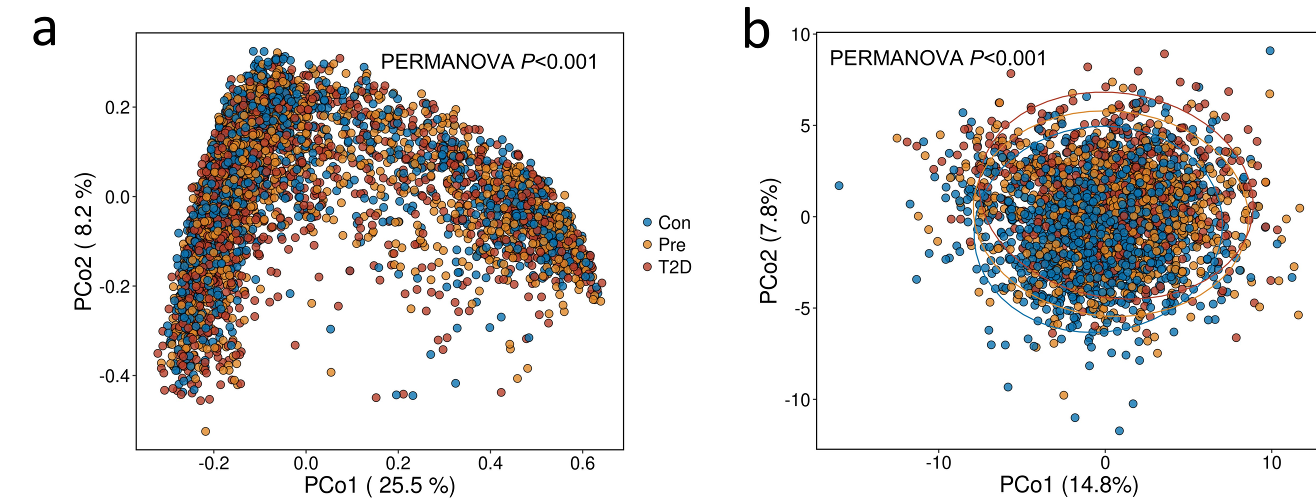



## Substantial proportions of variation in plasma metabolome are explained by the gut microbiome and habitual diet

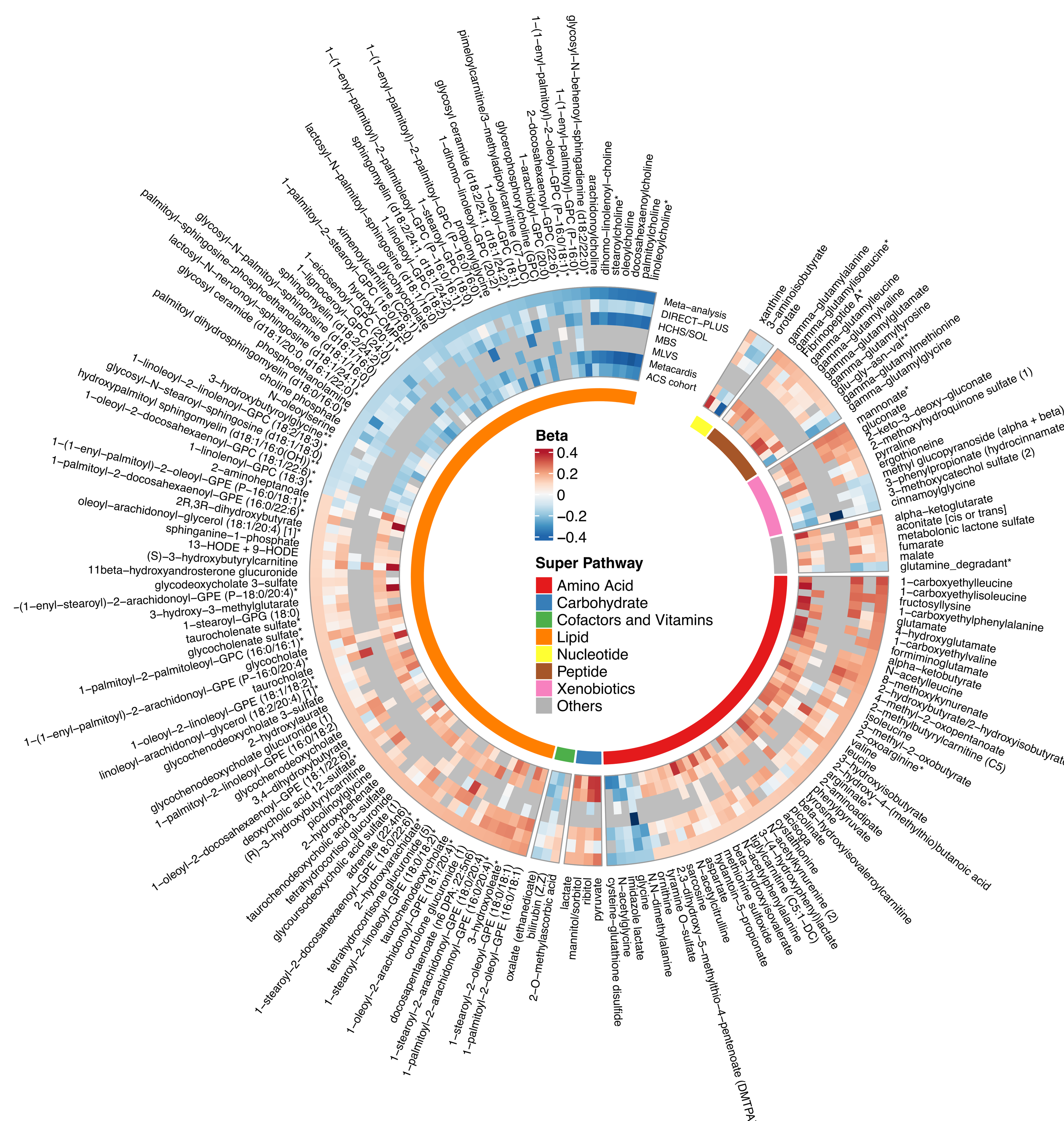
| | All | Species | Diet | T2D | Metformin | BMI | Sex | Age |
|---|---|---|---|---|---|---|---|---|
| Meta-analysis | 18.53 | 9.82 | 11.07 | 0.99 | 0.84 | 1.31 | 2.20 | 1.19 |
| DIRECT-PLUS | 19.93 | 8.60 | 9.32 | 0.64 | 0.56 | 0.84 | 1.63 | 2.02 |
| HCHS/SOL | 15.78 | 8.01 | 9.02 | 1.53 | 0.99 | 0.98 | 2.23 | 2.07 |
| MBS | 26.19 | 11.49 | 13.62 | 0.82 | 1.12 | 3.05 | | 0.64 |
| MLVS | 22.47 | 10.74 | 12.71 | 0.76 | 0.42 | 2.34 | | 0.81 |
| MetaCardis | 13.85 | 9.82 | | 2.48 | 1.91 | 2.50 | 3.28 | 2.82 |
| Talmor-Barkan_2022 | 30.43 | 16.41 | 17.97 | 1.80 | | 2.60 | 3.64 | 4.18 |

PERMANOVA R2

## T2D is significantly associated with overall gut microbiome (a) and plasma metabolome (b) although not a major driver

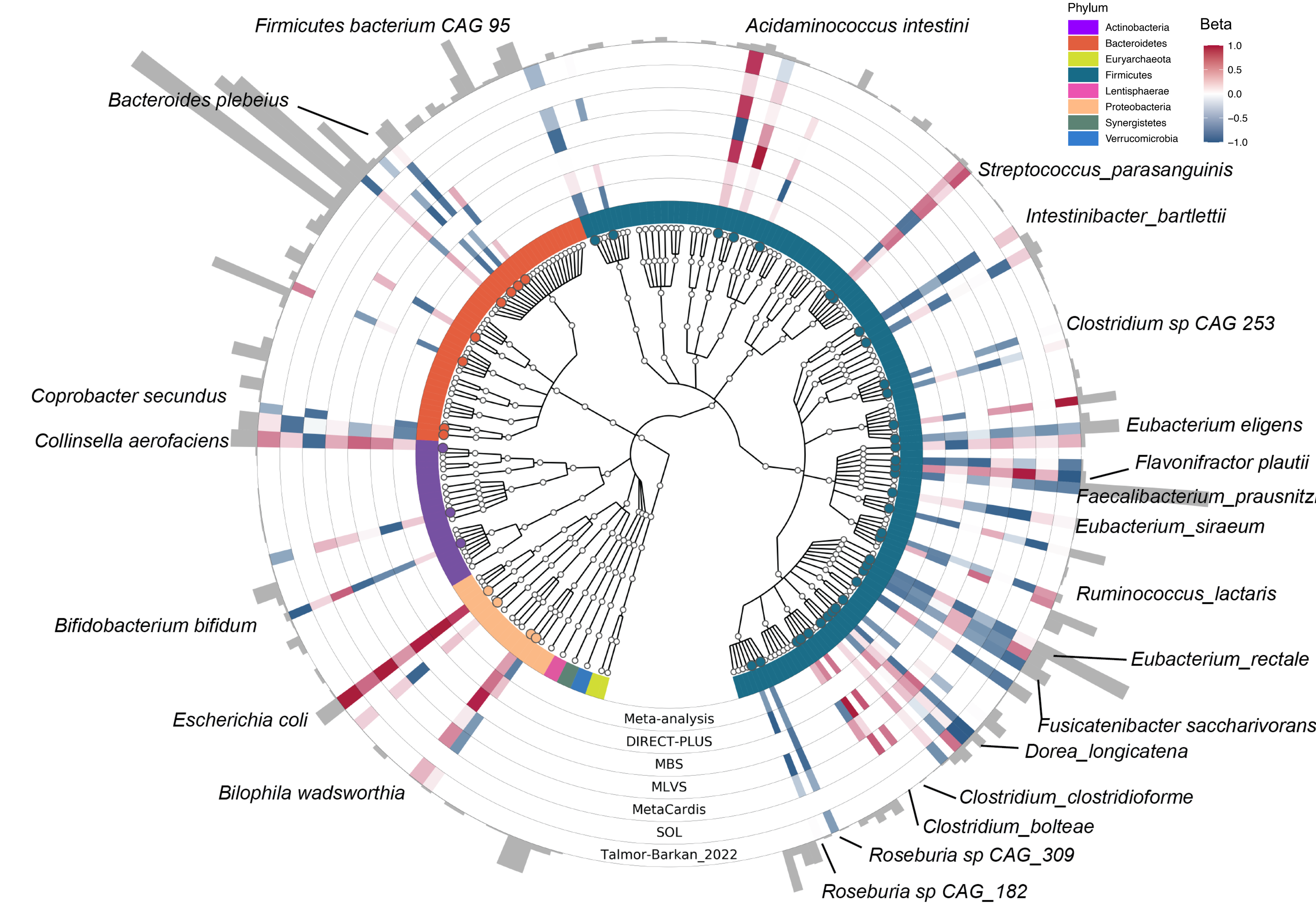



## Metabolites from various pathways are constantly linked with T2D across studies, highlighting the systematic nature of T2D pathogenesis



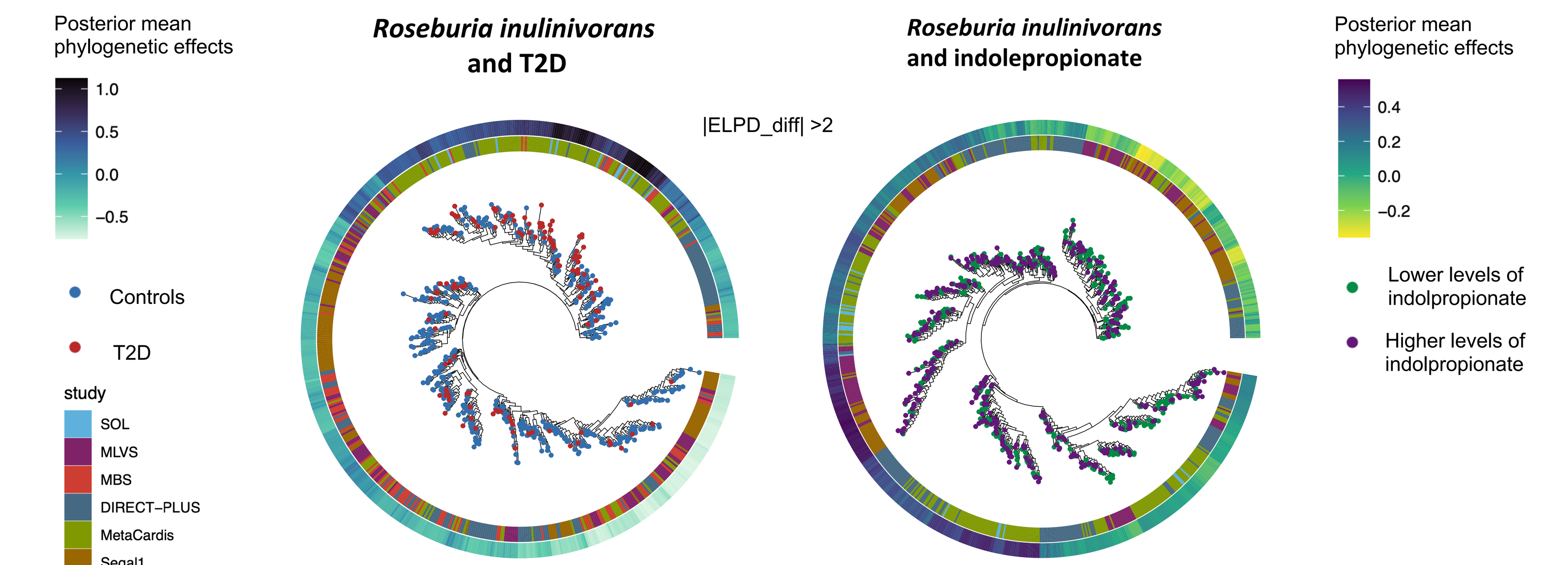

The beta coefficients were estimated using the linear regression model with adjustment for age, sex, and BMI in MaAsLin2. Metabolites with FDR *q* value<0.05 were shown.

## Phylogenetically and functionally diverse species linked to T2D




The beta coefficients were estimated using the linear regression model with adjustment for age, sex, and BMI in MaAsLin2. Species with FDR *q* value<0.25 were highlighted with colors.

## Phylogeny within *Roseburia inulinivorans* associated with T2D and indolepropionate




## ACKNOWLEDGEMENT


- This research was funded by the National Institutes of Health (R01NR01999, R00DK119412, P30DK046200, U01CA167552, R01HL035464, and U01CA176726) and the National Cancer Institute (P01CA055075). HCHS/SOL was funded by NHLB (HHSN268201300001I/N01-HC-65233, HHSN268201300004I/N01-HC-65234, HHSN268201300002I/N01-HC-65235, HHSN268201300003I/N01-HC-65236, HHSN268201300005I/N01-HC-65237). DIRECT-PLUS was funded by the Deutsche Forschungsgemeinschaft (209933838).
- We sincerely thank all participants and staff of MBS, MLVS, HCHS/SOL, and DIRECT-PLUS for their great contributions to this research.

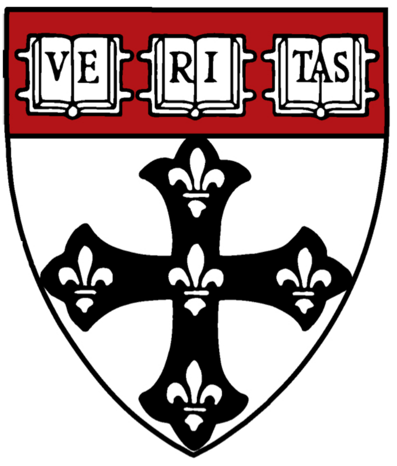


For any questions, please contact
Dr. Zhendong Mei: nhzme@channing.harvard.edu
Dr. Dong D. Wang: dow471@mail.harvard.edu

# Harvard T.H. Chan School of Public Health Microbiome Analysis Core

Xochitl C. Morgan[1], Lauren J. McIver[1], Thomas Kuntz[1], Curtis Huttenhower[1,2]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health [2]Broad Institute of MIT and Harvard



The Microbiome Analysis Core at the Harvard T.H. Chan School of Public Health was established in response to the rapidly emerging field of microbiome research and its potential to affect studies across the biomedical sciences. The Core's goal is to aid researchers with microbiome study design and interpretation, reducing the gap between primary data and translatable biology. The Microbiome Analysis Core provides end-to-end support for microbial community and human microbiome research, from experimental design through data generation, bioinformatics, and statistics. This includes general consulting, power calculations, selection of data generation options, and analysis of data from amplicon (16S/18S/ITS), shotgun metagenomic sequencing, metatranscriptomics, metabolomics, and other molecular assays. The Microbiome Analysis Core has extensive experience with microbiome profiles in diverse populations, including taxonomic and functional profiles from large cohorts, qualitative ecology, multi'omics and meta-analysis, and microbial systems and human epidemiological analysis. By integrating microbial community profiles with host clinical and environmental information, we enable researchers to interpret molecular activities of the microbiota and assess its impact on human health.

## Core services

**Consultation for microbiome project development.**
We provide consultation on experimental design, sample collection and sequencing, grant proposal development, study power estimation, bioinformatics, and statistical data analysis.

**Validated end-to-end meta'omic analysis of microbial community data.**
Using open-source analytical methods developed in the Huttenhower laboratory and by other leaders in the field, we provide cutting-edge microbiome informatics and analysis.

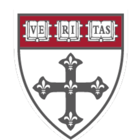
**Fully-collaborative support for all stages of funded investigations**

From preliminary data development to hypothesis formulation, grant narrative development, data analysis and inference, custom software development, and co-authored dissemination of findings.

### Study Design
- Consultation
- Grant assistance
- Power analysis
- Collection methods
- Wet lab
- Dry lab

### Analysis
- Bioinformatics (raw data processing, taxonomic and functional profiling)
- Downstream analysis and statistics

### Interpretation
- Results
- Discussion
- Manuscript writing/editing
- Response to reviewers

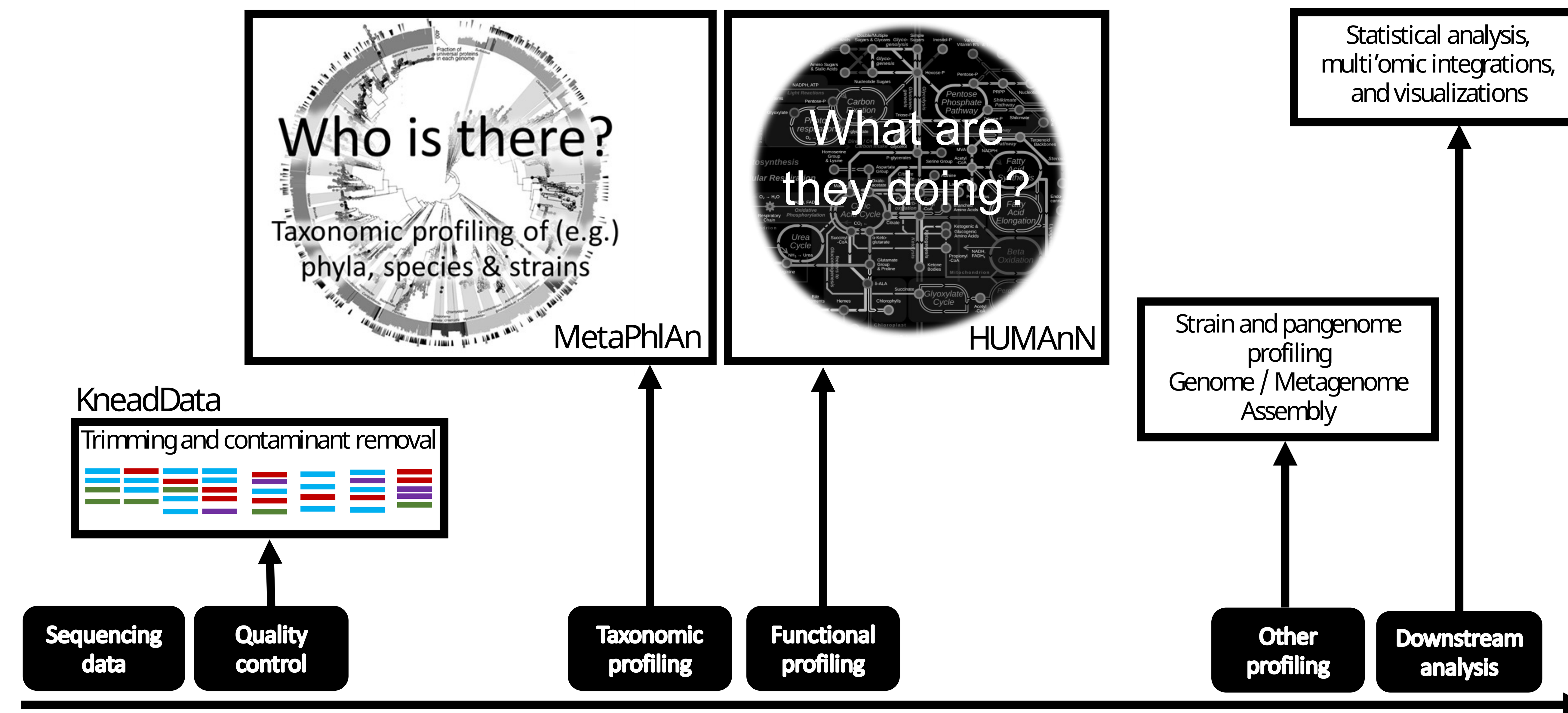


HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH
Microbiome Analysis Core

Director: Xochitl Morgan
Senior Software Developer: Lauren McIver
Data Analyst: Thomas Kuntz
Scientific Director: Curtis Huttenhower

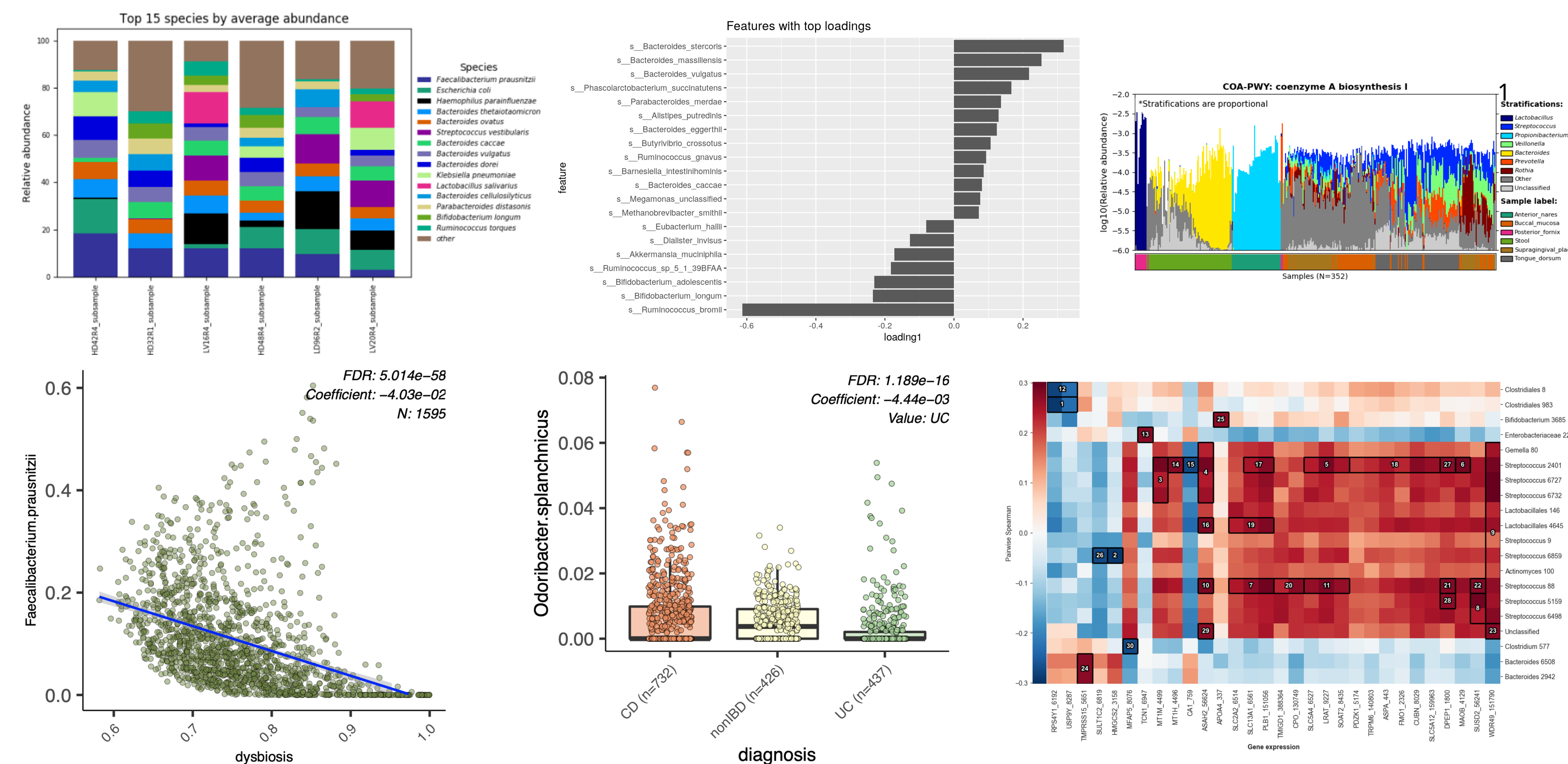
Contact us at: hcmac@hsph.harvard.edu

## Microbial community profiling




The first step in microbiome molecular data analysis is quality control (KneadData) and profiling to transform raw data into biologically interpretable features using a reproducible workflow (AnADAMA/bioBakery). This includes identifying microbial species (MetaPhlAn) and strains (PanPhlAn/StrainPhlAn), characterizing their functional potential or activity (HUMAnN), and integrating metagenomics with other data types.

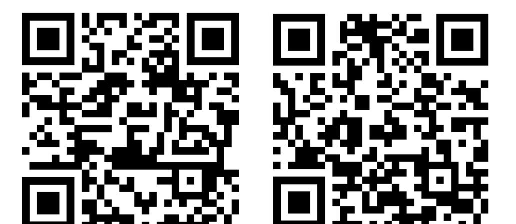
## Downstream analysis and statistics




Once profiled, microbial communities are amenable to downstream statistics and visualization much like other molecular epidemiology data such as human genetic or transcriptional profiles. Like these other data types, microbial communities often require tailored statistics for environmental, exposure, or phenotype association (MaAsLin 2.0, MMUPHIN) or for ecological interaction discovery (BAnOCC). The Harvard Chan Microbiome Analysis Core provides a variety of analyses for researchers working in the microbiome space.

The Harvard Chan Microbiome Analysis Core is a part of the Harvard Chan Microbiome in Public Health Center (HCMPH). Want to learn more? Visit **https://hcmph.sph.harvard.edu**

**https://hcmph.sph.harvard.edu/hcmac** **http://huttenhower.sph.harvard.edu**



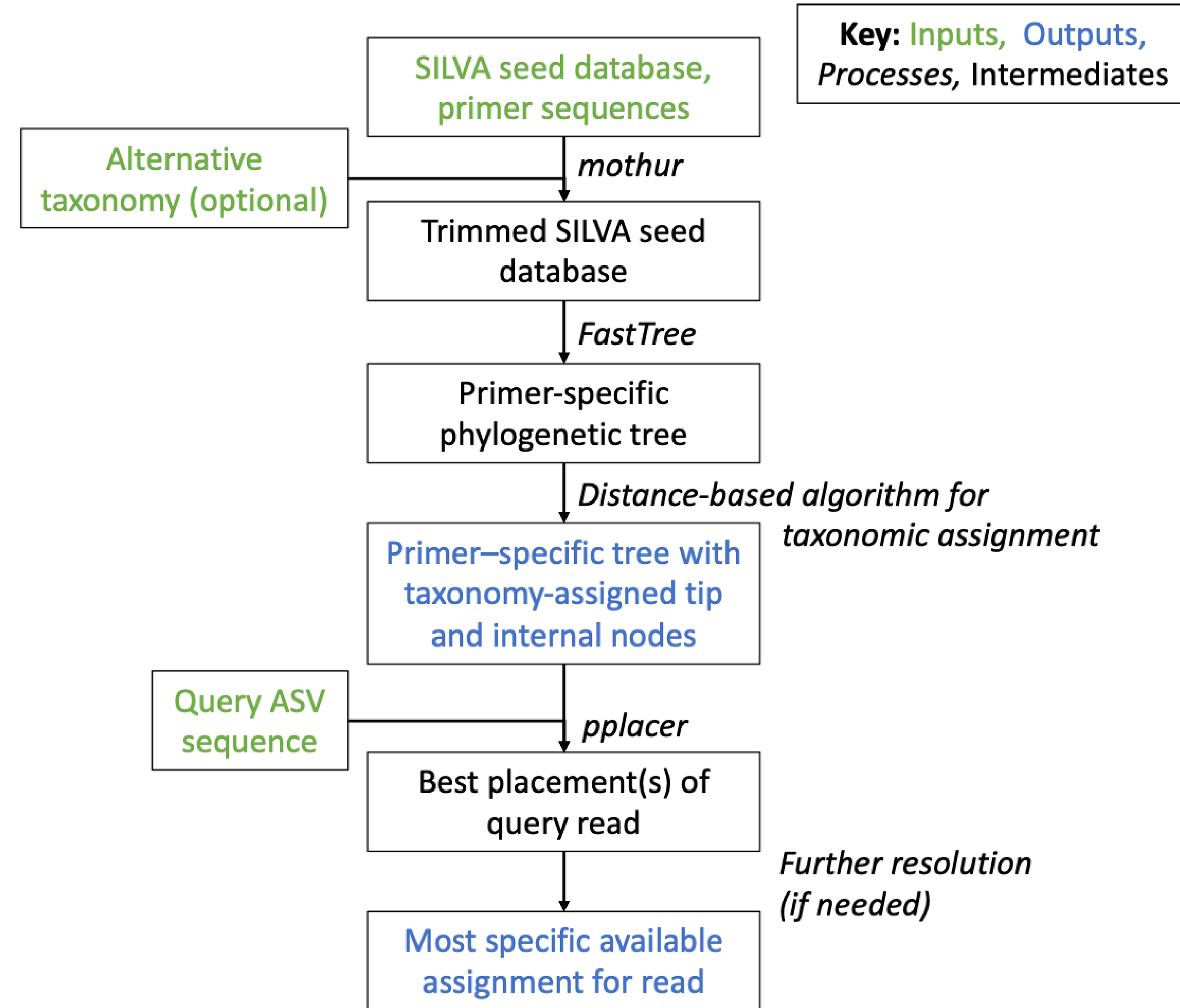
# Preserving and Assimilating Region-specific Ambiguities in Taxonomic Hierarchical Assignments for Amplicons - PARATHAA

Jacob T. Nearing[1,2,3], Kelsey N. Thompson[1,2,3], Amrisha Bhosle[1,2,3], Tobyn Branck[1,3], Eric A. Franzosa[1,2,3], Dayakar Badri[4], Christoph Brockel[4], Curtis Huttenhower[1,2,3], Meghan I. Short[5]

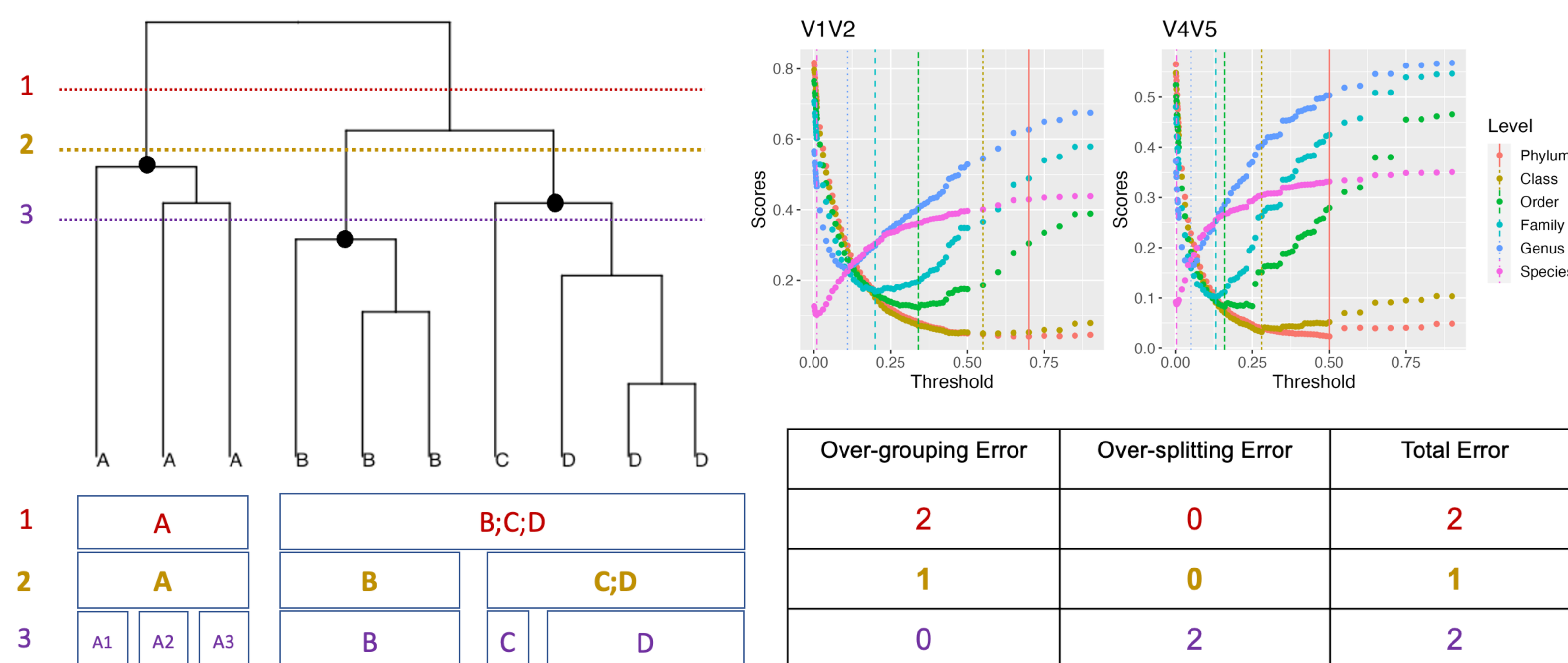
[1] Broad Institute of Harvard and MIT - [2] Harvard T.H. Chan School of Public Health - [3] Harvard Chan Microbiome in Public Health Center - [4] Science & Technology Center, Hill's Pet Nutrition
[5] Institute for Clinical Research and Health Policy Studies, Tufts Medical Center

Short read amplicon sequencing is a common strategy to taxonomically profile microbial communities. This involves the PCR amplification of specific variable regions on markers such as the 16S rRNA gene. While this is relatively low cost and high-throughput, it does suffer from drawbacks such as bias associated with usage of differing variable regions. Additionally, short variable regions do not always differ among taxa, introducing further ambiguity.

## PARATHAA Workflow



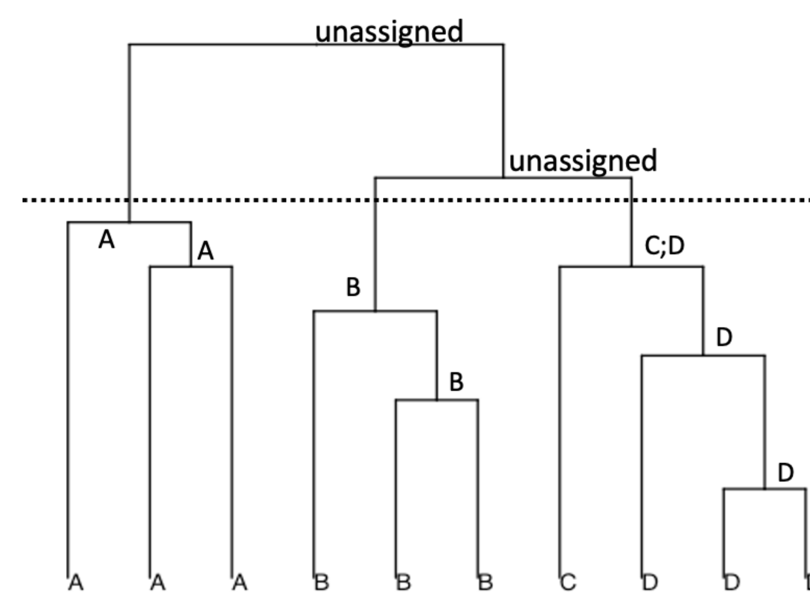

## Identification of Optimal Taxonomic Distance Thresholds




| | Over-grouping Error | Over-splitting Error | Total Error |
|---|---|---|---|
| 1 | 2 | 0 | 2 |
| 2 | 1 | 0 | 1 |
| 3 | 0 | 2 | 2 |

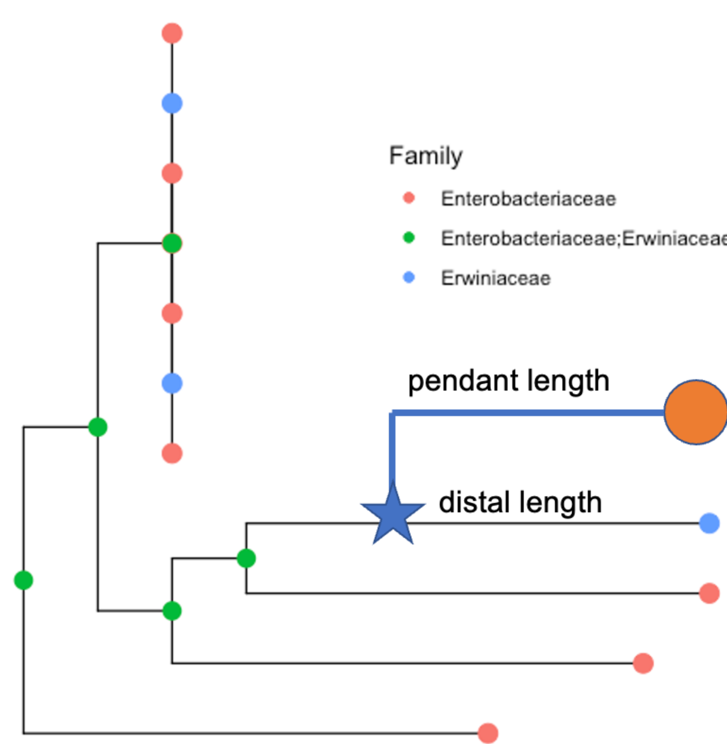
- For each taxonomic level PARATHAA selects multiple distance thresholds to test (3 dotted lines)
- For each distance the over-grouping and over-splitting error is summed to a total error
  - Over-grouping: areas in the tree where multiple different labels are under the chosen distance
  - Over-splitting: areas in the tree where the same label is split-up under the chosen distance
- The total error is then normalized to a score between 0 and 1. The distance with the lowest score is chosen as the optimal distance threshold for that variable region (top right hand plots)

## Taxonomic Assignment of Internal Tree Nodes




- We next want to assign taxonomy to the internal nodes of the primer-specific phylogenetic tree
- For each taxonomic level:
  - Nodes that are above the optimal distance cutoff for that taxonomic level are left unassigned
  - Nodes that are below the threshold are assigned taxonomy based on the underlying dominate tip labels
    - When multiple differing taxonomies are present the dominant label is chosen based on a binomial error model that allows for multiple assignments (see node C;D).

## Taxonomic Assignment of New Query Seqeunces




- Placed sequences (using pplacer) are then assigned taxonomy using the following criteria:
- If the maximum distance to child nodes is smaller than the species threshold:
  - Taxonomy is assigned to the placed nodes species
  - A neighborhood search is then used to eliminate assignments when conflicting labels (from nearby tips) are within a small radius of the placed tip
- If the maximum distance is larger:
  - Taxonomy is assigned at the level where the maximum distance is below that taxonomic level's optimal threshold

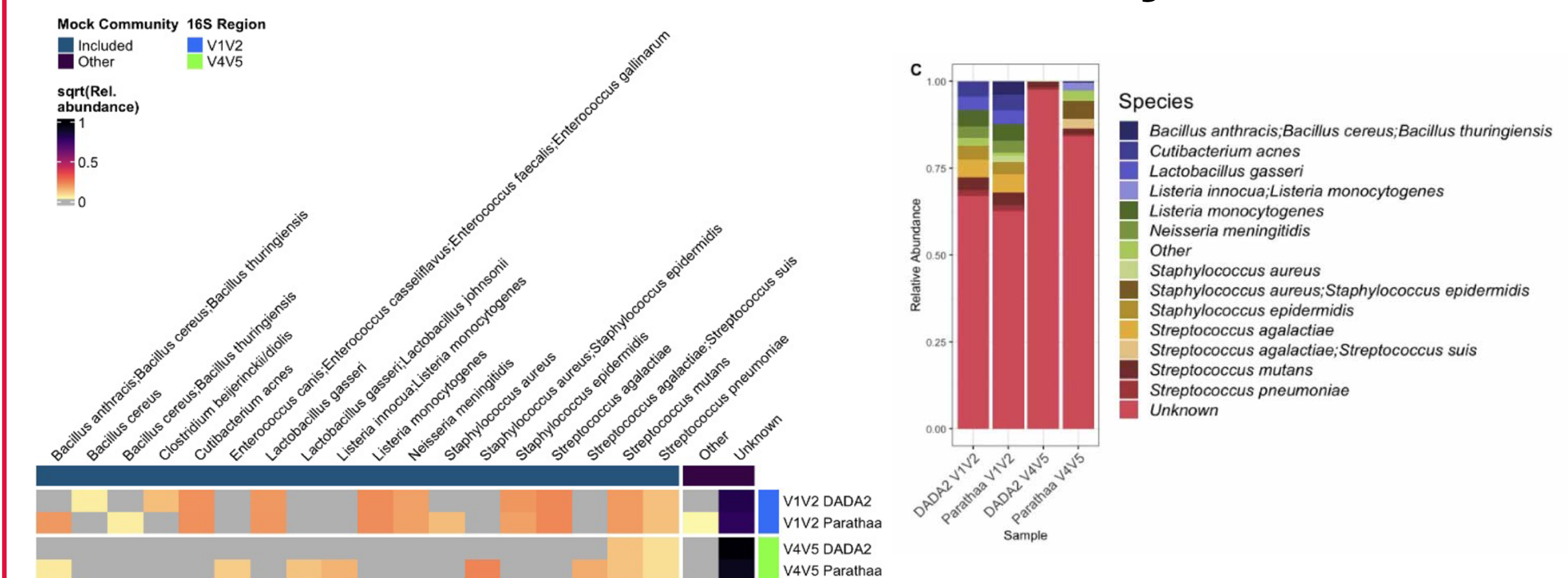
## PARATHAA Maintains Precision While Being More Sensitive Than State-of-the-art Assignment Methods

| Metric | V1V2 | | | V4V5 | | |
|---|---|---|---|---|---|---|
| | Parathaa | DADA2 (default) | DADA2 Multi | Parathaa | DADA2 (default) | DADA2 Multi |
| Accuracy | 0.781 | 0.689 | 0.697 | 0.820 | 0.757 | 0.809 |
| Precision | 0.926 | 0.986 | 0.982 | 0.930 | 0.950 | 0.937 |
| Recall | 0.806 | 0.696 | 0.706 | 0.875 | 0.788 | 0.856 |
| F1 Score | 0.877 | 0.816 | 0.822 | 0.901 | 0.862 | 0.894 |
| Uniquely Correct | 0.123 | 0.039 | 0.039 | 0.146 | 0.128 | 0.128 |
| One-to-many Correct | 0.024 | 0 | 0.011 | 0.065 | 0 | 0.065 |
| Incorrect | 0.031 | 0.010 | 0.013 | 0.062 | 0.040 | 0.054 |
| Unassigned Correct | 0.633 | 0.651 | 0.648 | 0.609 | 0.629 | 0.616 |
| Unassigned Incorrect | 0.188 | 0.301 | 0.290 | 0.117 | 0.203 | 0.137 |

- Synthetic sequences that are not represented in the seed database that PARATHAA uses were chosen from SILVA to be assigned taxonomy by PARATHAA and DADA2

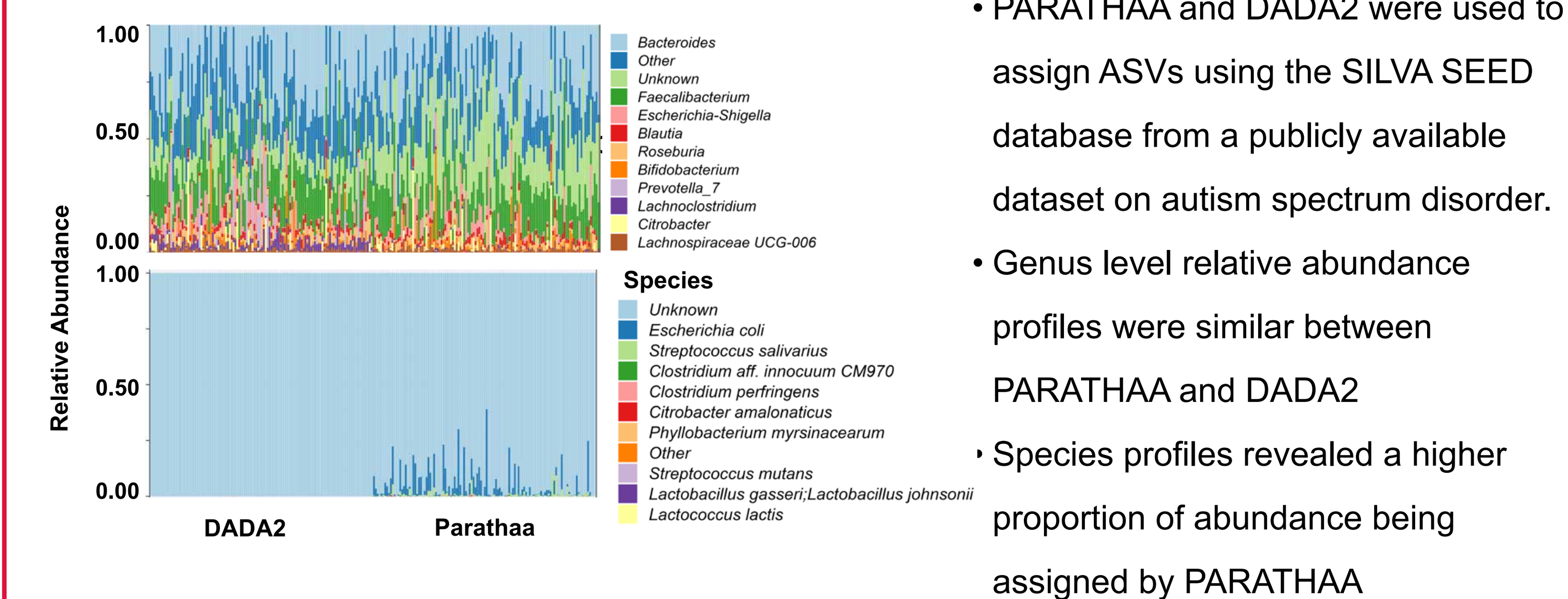
- Sequences were chosen from genera but not necessarily species that were represented within the seed database
- For each given genera at least 20 sequences were chosen (all sequences not in the seed database were chosen when there was less than 20)
- Query sequences were then trimmed to the V1V2 region
- DADA2 assignments were completed using the same seed database that PARATHAA uses

## PARATHAA Performs Comparably to Other Methods on Mock Commuty Data



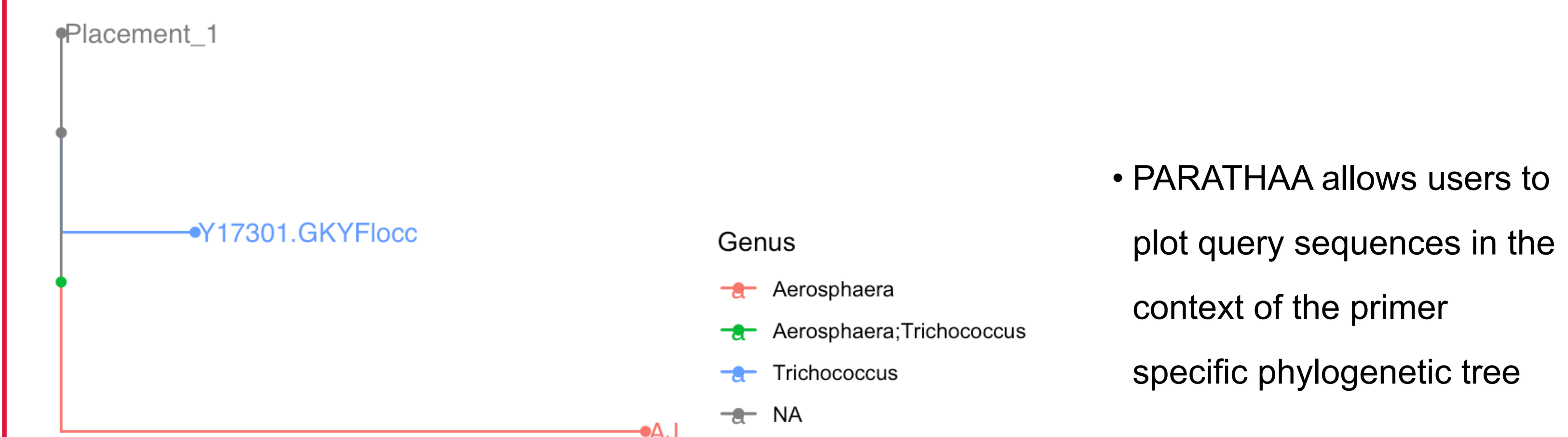

- PARATHAA assigns reads to a similar number of species (V1V2) or more (V4V5) compared to DADA2 on a published mock community.
- The mock community is composed of DNA from 20 species in equal quantities (5% relative abundance).
- Species with less than 0.1% relative abundance are listed as "Other".

## PARATHAA Assigns Similar Genera but more Species on Real Microbiome Datasets




- PARATHAA and DADA2 were used to assign ASVs using the SILVA SEED database from a publicly available dataset on autism spectrum disorder.
- Genus level relative abundance profiles were similar between PARATHAA and DADA2
- Species profiles revealed a higher proportion of abundance being assigned by PARATHAA

## PARATHAA Allows for Better Interpretation of Seqeunce Assignment




- PARATHAA allows users to plot query sequences in the context of the primer specific phylogenetic tree

- Plotting allows users to determine why their sequence receive the assignment given allowing for better interpretation then other black box machine learning methods

**http://huttenhower.sph.harvard.edu**

# Concurrent and Habitual Diets Differentially Associate with Microbial Multi-Omic Profiles in Inflammatory Bowel Disease (IBD)

NGUYEN LAB

Jiaxian Shen[1,2], Etienne Nzabarushimana[1,2], Hanseul Kim[1,2], Hannah VanEvery, Yiqing Wang[2], Kelsey N. Thompson[1,3], Andrew T. Chan[1,2], Curtis Huttenhower[1,3], Long H. Nguyen[1,2]

[1]Harvard T.H. Chan School of Public Health [2]Massachusetts General Hospital and Harvard Medical School [3]Broad Institute of MIT and Harvard

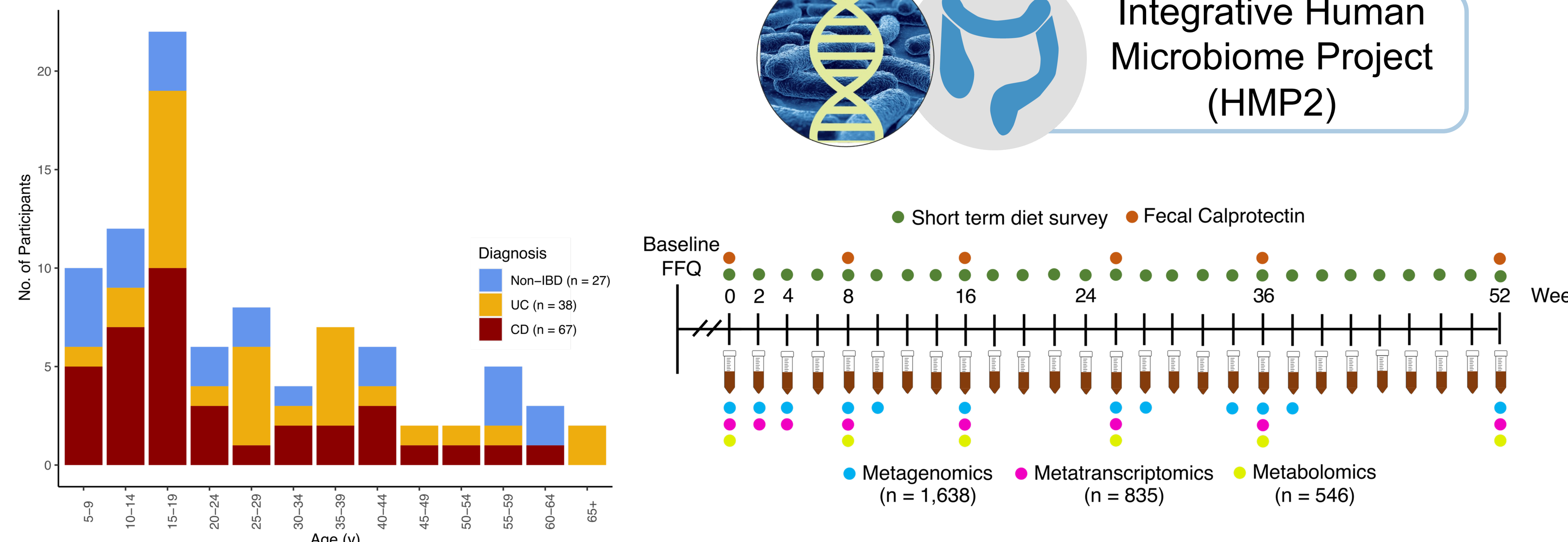
MGH 1811

## Background

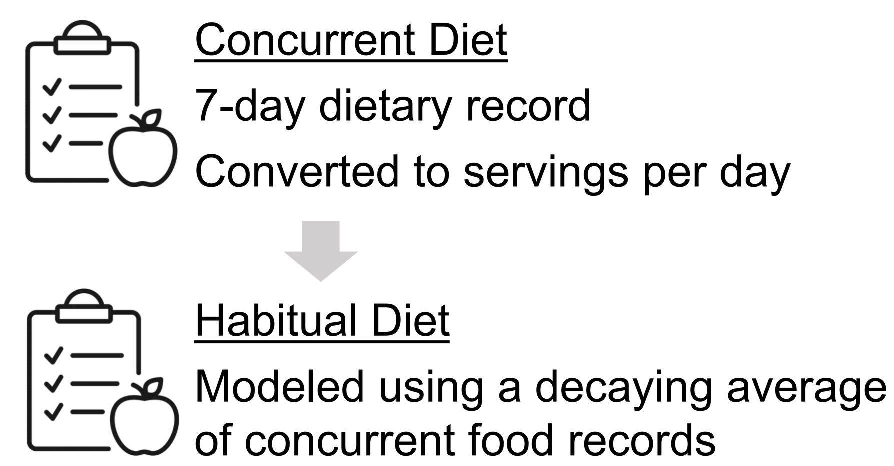
- Studies have linked diet to the risk and severity of IBD and its subtypes, Crohn's disease (CD) and ulcerative colitis (UC).
- Similarly robust evidence has associated disease activity to characteristic alterations in gut microbial taxonomy (metagenomics, MGX), community functions (metatranscriptomics, MTX), and microbial metabolites (metabolomics, MBX).
- However, in IBD, explorations into how these multi-omic readouts are affected by concurrent/short-term vs. habitual/long-term diets are limited.

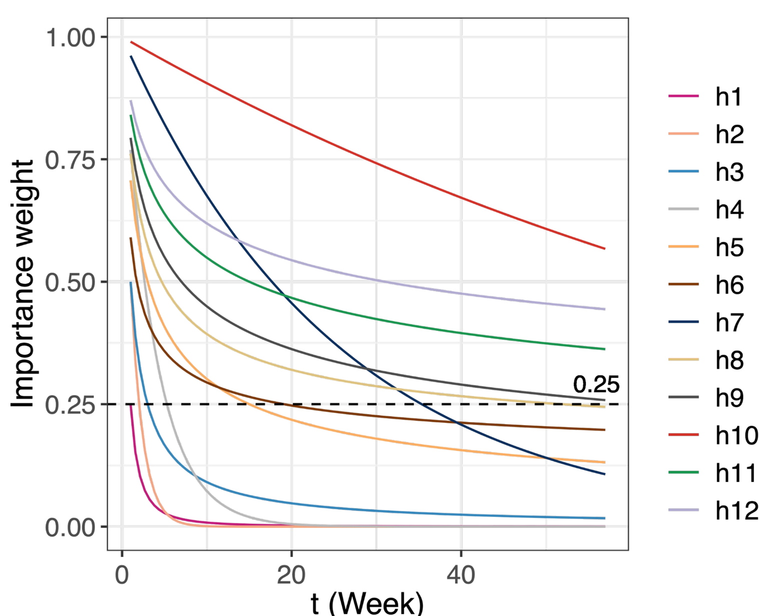
## Study population and analysis design

### A. Data acquisition




### B. Construction of dietary profiles



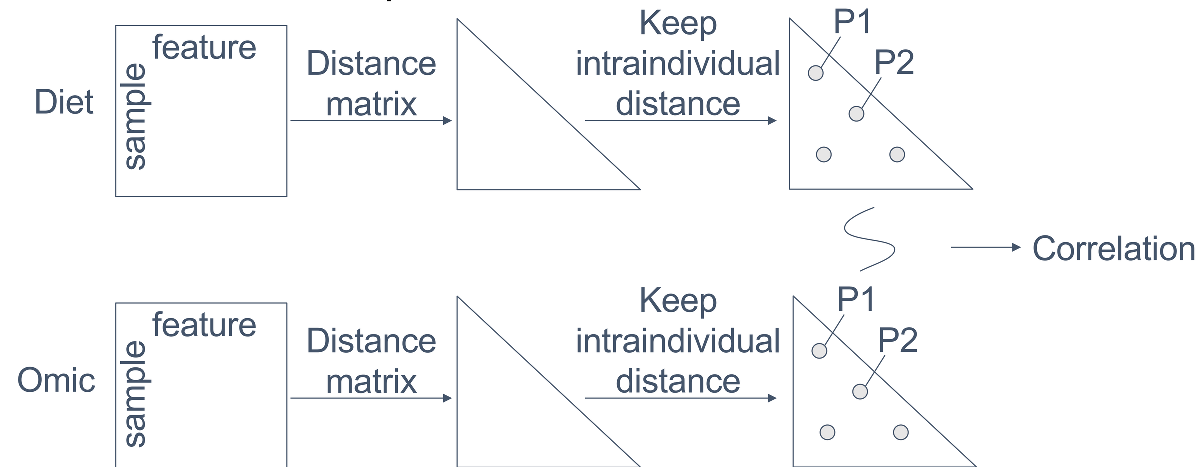




- We explored different decay formulas, which vary the duration and extent of the influence from a past diet.
- h1→h12: Decay became more gradual generally.
- For example, with "importance weight= $2^{-t}$", habitual diet of a sample was calculated as a decaying average weighted by $2^{-n}$ for the $n^{th}$ prior week of food records, accounting for all available data points.

### C. Quantification of associations

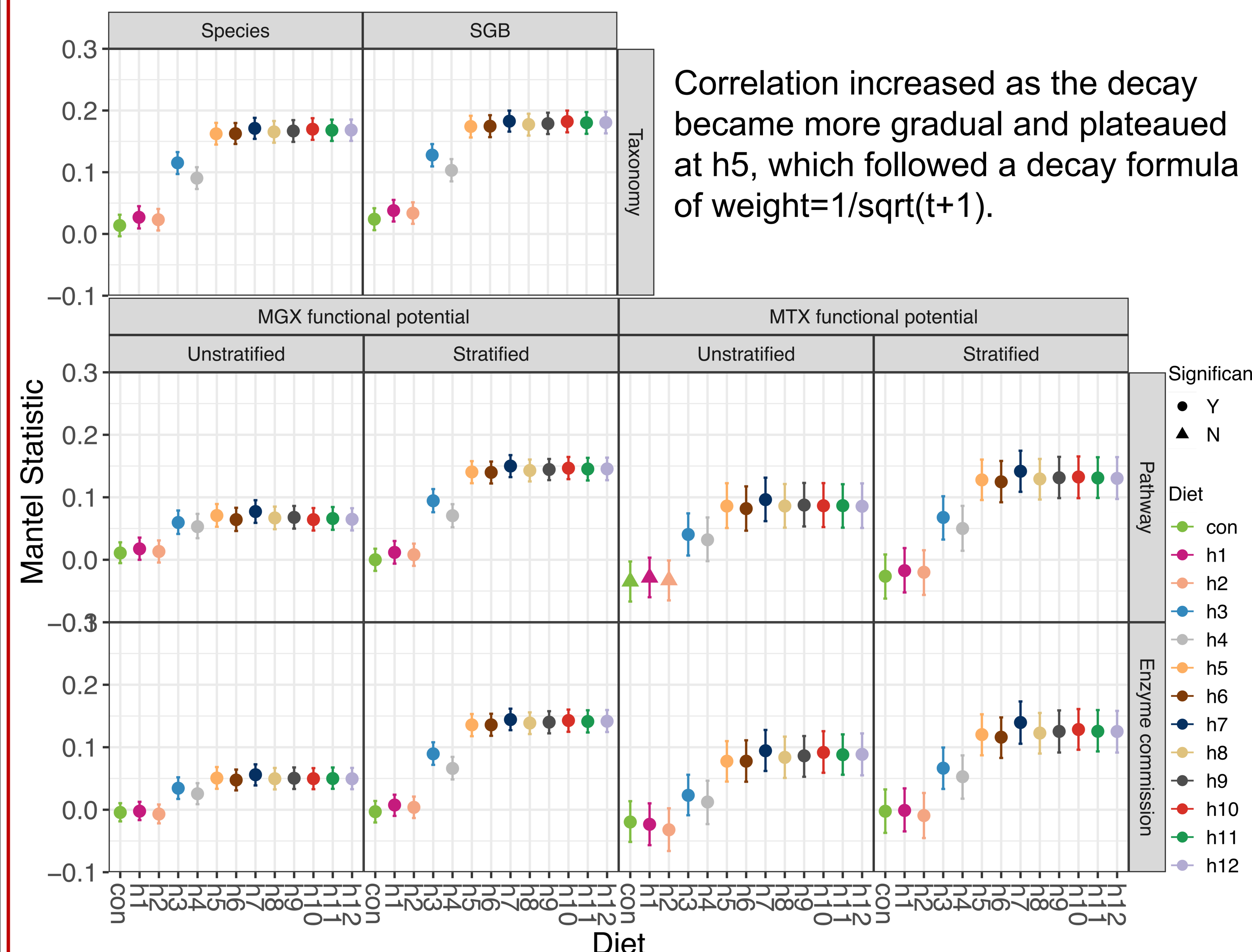
We linked diet and microbiome matrices via intra-individual Mantel tests, quantifying associations within participants only.

- Significance: 4,999 permutations
- Robustness: 4,999 bootstraps




## Habitual diets had a significantly stronger correlation with taxonomy and functional potential

Compared to concurrent, habitual diets had a significantly stronger correlation with MGX taxonomy, MGX functional potential, and MTX functional potential.

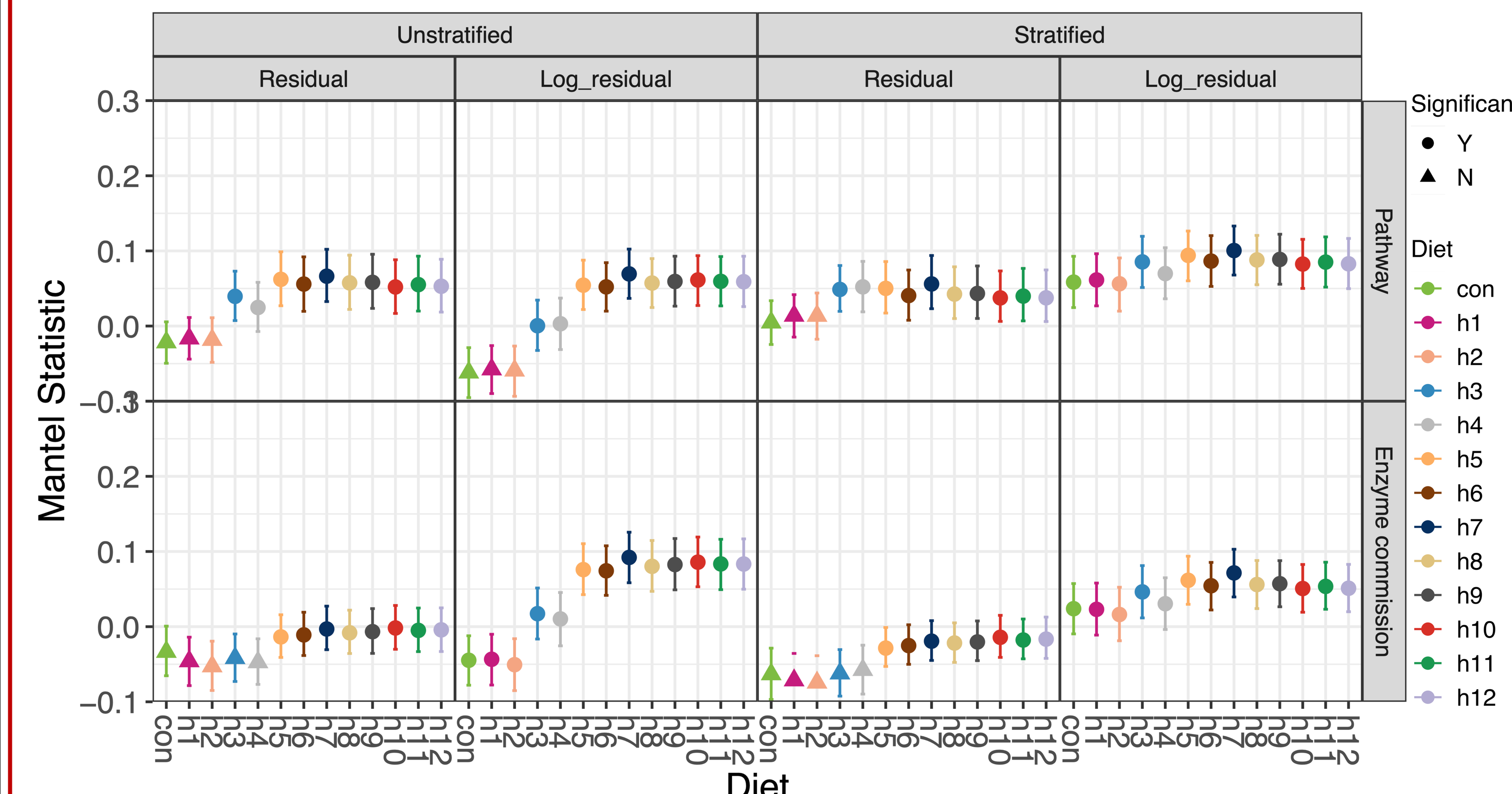



Correlation increased as the decay became more gradual and plateaued at h5, which followed a decay formula of weight=1/sqrt(t+1).

* Distance: Bray-Curtis; Abbreviation: con, concurrent diet; h1~h12, habitual diets.

## MTX functional activities exhibited patterns consistent with MGX only at the community-level

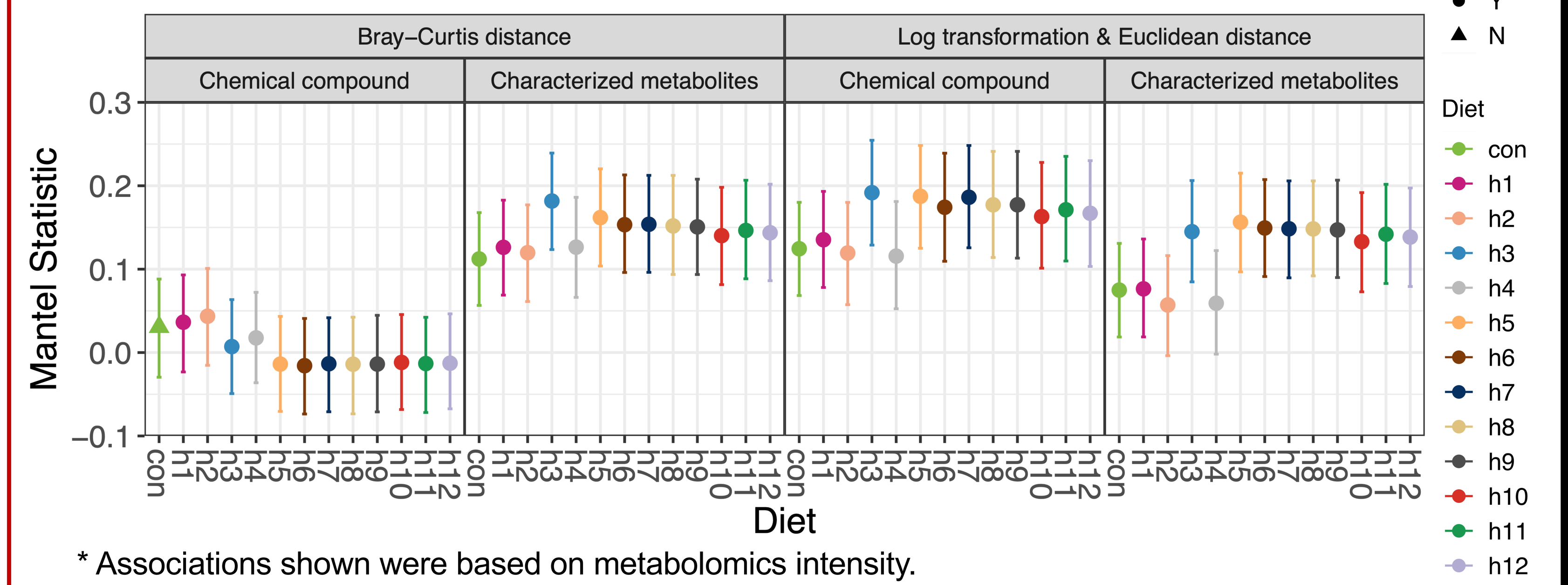
- The pattern was more prominent for pathways than enzyme profiles.
- When functions were stratified by contributing microorganisms, no differences emerged between concurrent and habitual diets.
- This suggests that although habitual diet more significantly shapes community-level functions than concurrent diet, it does not predict which species contribute to specific functions.




* MTX functional activities were assess by normalizing MTX against DNA gene copy number; Distance: Euclidean; Log_residual: log transformation was conducted before residual calculation.

## Neither habitual nor concurrent diets had additive effects on shaping MBX

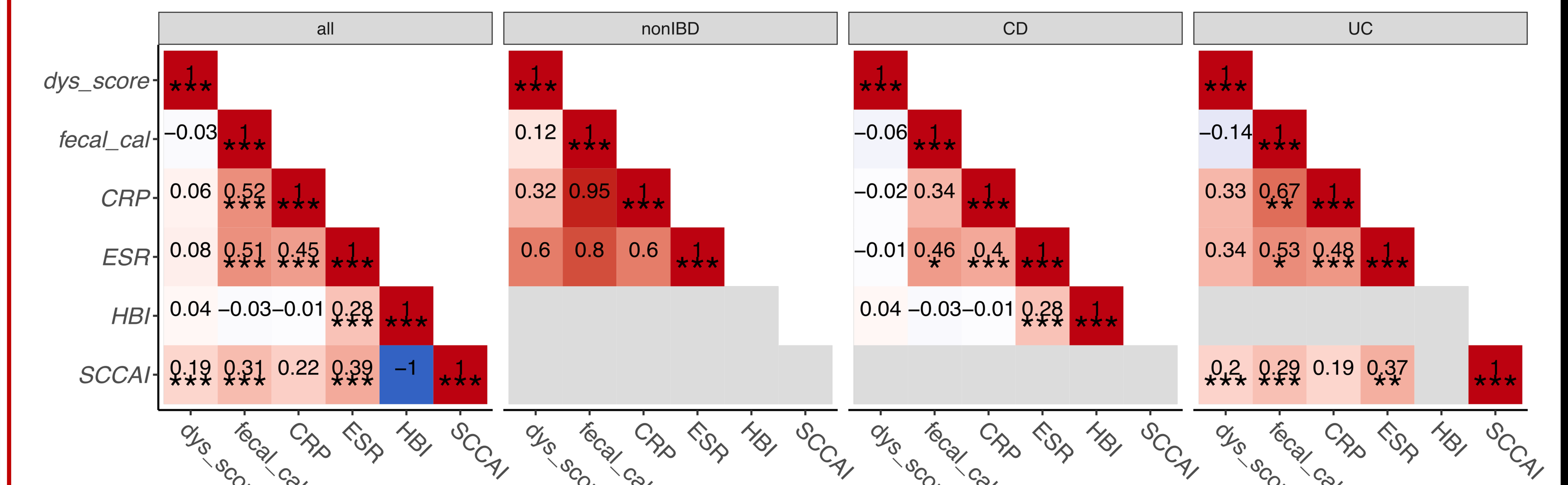
- This might be an actual biological pattern or potentially technical noise in MBX data, as the association tests were less stable during bootstrapping.
- The coupling between diet and MBX became tighter at a log scale and was more pronounced for characterized metabolites, as opposed to all chemical compounds.



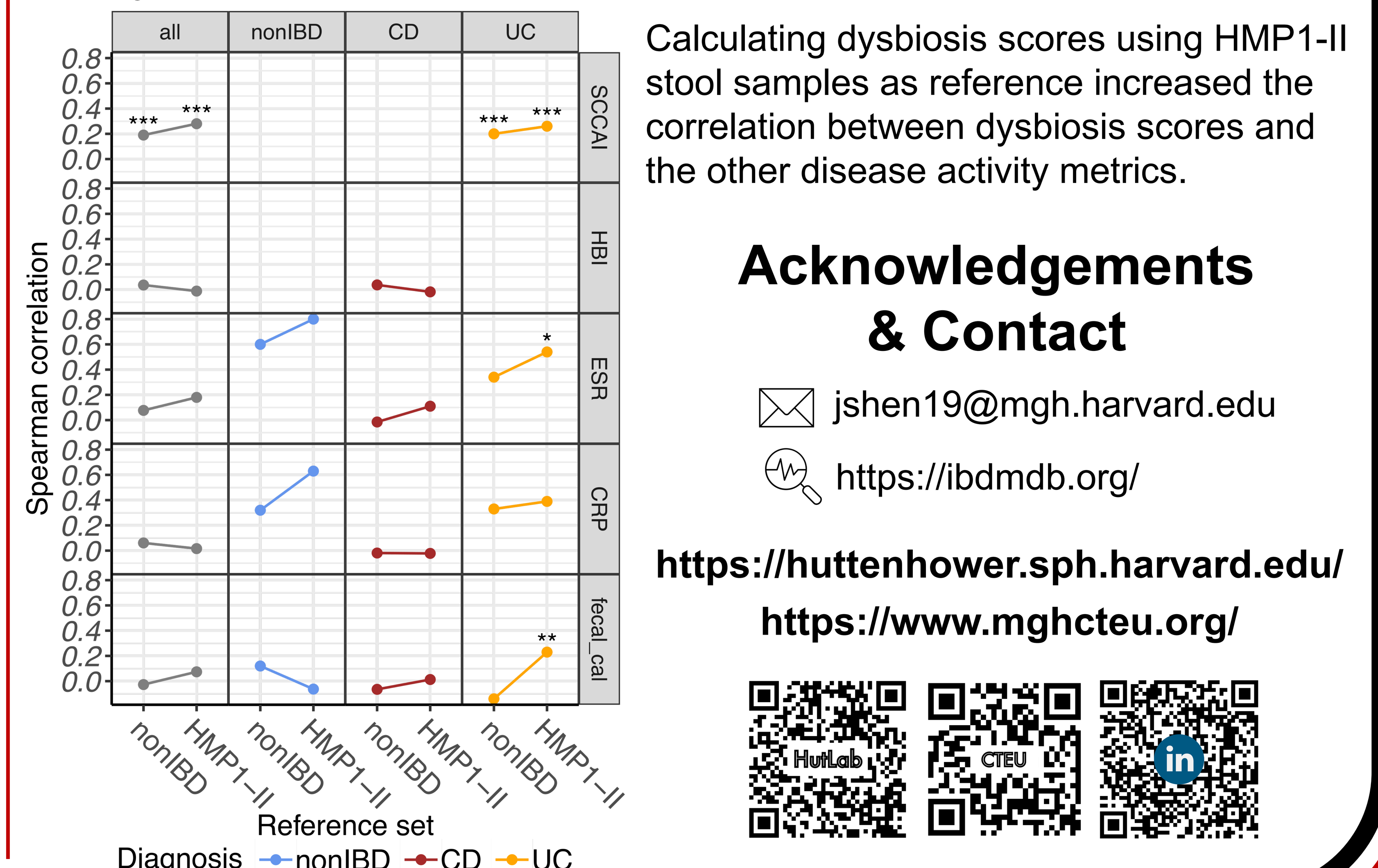

* Associations shown were based on metabolomics intensity.

## Disease activity metrics were modestly correlated

- Correlation: CD < UC < nonIBD
- CSR, ERP, and fecal calprotectin correlated the best.




* Abbreviation: dys_score, dysbiosis score; fecal_cal, fecal calprotectin; CRP, C-Reactive Protein; ESR, Erythrocyte Sedimentation Rate; HBI, Harvey-Bradshaw Index; SCCAI, Simple Clinical Colitis Activity Index. Dysbiosis score was calculated using nonIBD sub-cohort as reference.

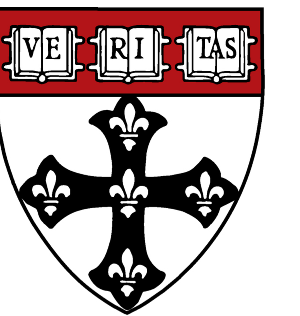



Calculating dysbiosis scores using HMP1-II stool samples as reference increased the correlation between dysbiosis scores and the other disease activity metrics.

## Acknowledgements & Contact

jshen19@mgh.harvard.edu
https://ibdmdb.org/
https://huttenhower.sph.harvard.edu/
https://www.mghcteu.org/

# Using housekeeping gene *cpn60* as a marker for microbial community profiling and viability assessment



Ya Wang[1,2,3], Kelsey Thompson[1,2,3], Sagun Maharjan[1,3], Marina Chen[1], Meghan I. Short[4]
Yancong Zhang[1,2,3], Jacob Nearing[1,2,3], Eric A. Franzosa[1,2,3], Curtis Huttenhower[1,2,3]

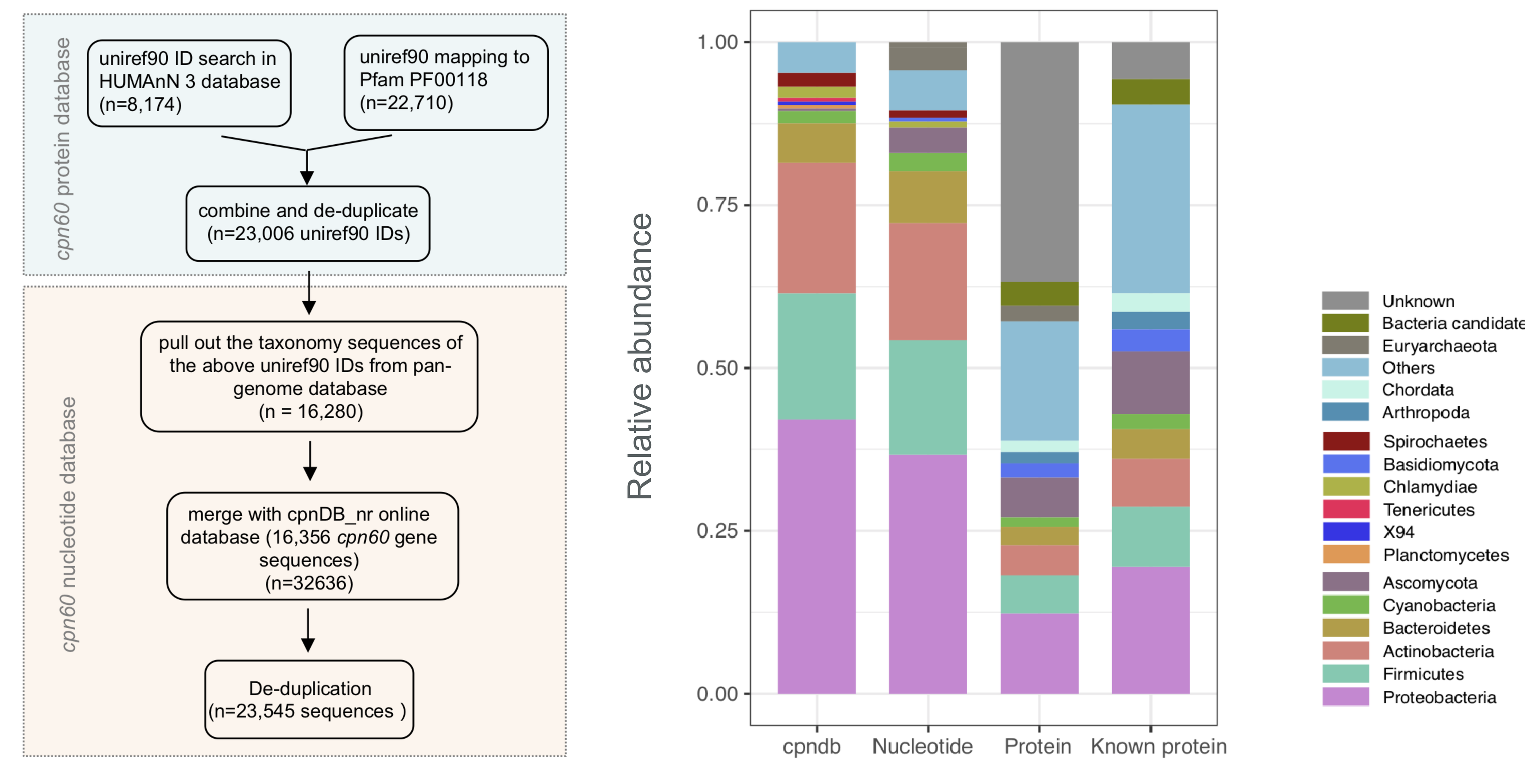
[1]Harvard T.H. Chan School of Public Health [2]Broad Institute of MIT and Harvard [3]Harvard Chan Microbiome in Public Health Center [4]Tufts University, Boston

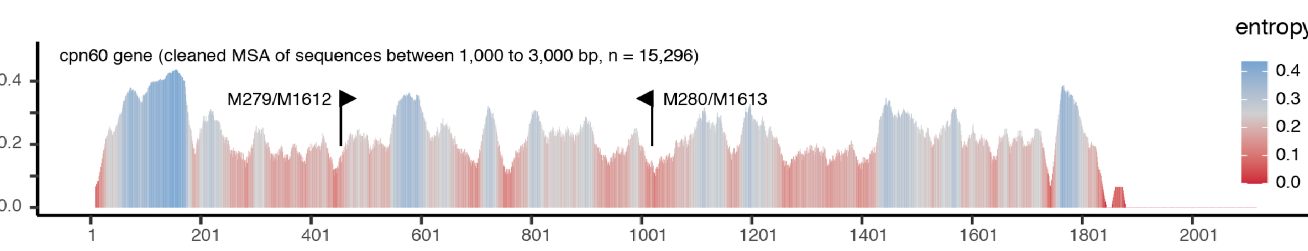
While high-throughput metagenomic sequencing has transformed microbial community studies, it remains challenging to discern between "alive" and "dead" microbes. This limits our understanding of microbial community functions and their interactions with their surroundings and human health. To address this, we introduced a novel protein-coding marker gene approach using the *cpn60* gene for comprehensive microbial community profiling and functional activity assessment via metagenomic and metatranscriptomic sequencing. We first constructed an extensive database integrating *cpn60* protein and enriched *cpn60* nucleotide sequences. Using *cpn60* protein IDs for protein-based taxonomy inference in the Human Microbiome Project II dataset, we found strong agreement between *cpn60*-protein-based taxonomy and shotgun metagenomic results. This would suggest *cpn60* being a discriminative marker for taxonomy profiling. Additionally, we explored *cpn60* protein expression in metatranscriptomic data as an indicator of bacterial species' activity. The *cpn60*-protein-based analysis correlated positively with growth rates estimated using the bPTR method, suggesting *cpn60* proteins as robust markers indicating the activity of microbial community. These findings suggest *cpn60*'s potential as a discriminative marker for microbial community taxonomy profiling and viability characterization, offering insights into microbial community dynamics. This study underscores the promise of marker gene approaches in advancing microbial viability assessment and functional activity profiling.

## A comprehensive database with *cpn60* protein and nucleotide sequences

We first constructed a comprehensive database that inculdes 23,006 unique *cpn60* protein IDs and 20,556 nucleotide sequences.

- The updated database exhibits a taxonomic composition at the phylum level similar to cpnDB, with dominant phyla including Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes.
- The protein database reveals a diverse array of taxonomies, including unclassified groups, bacteria, fungi, animals, plants, and other eukaryotic organisms.
- Multiple sequence alignment revealed multiple variable regions and conserved regions in *cpn60* gene.

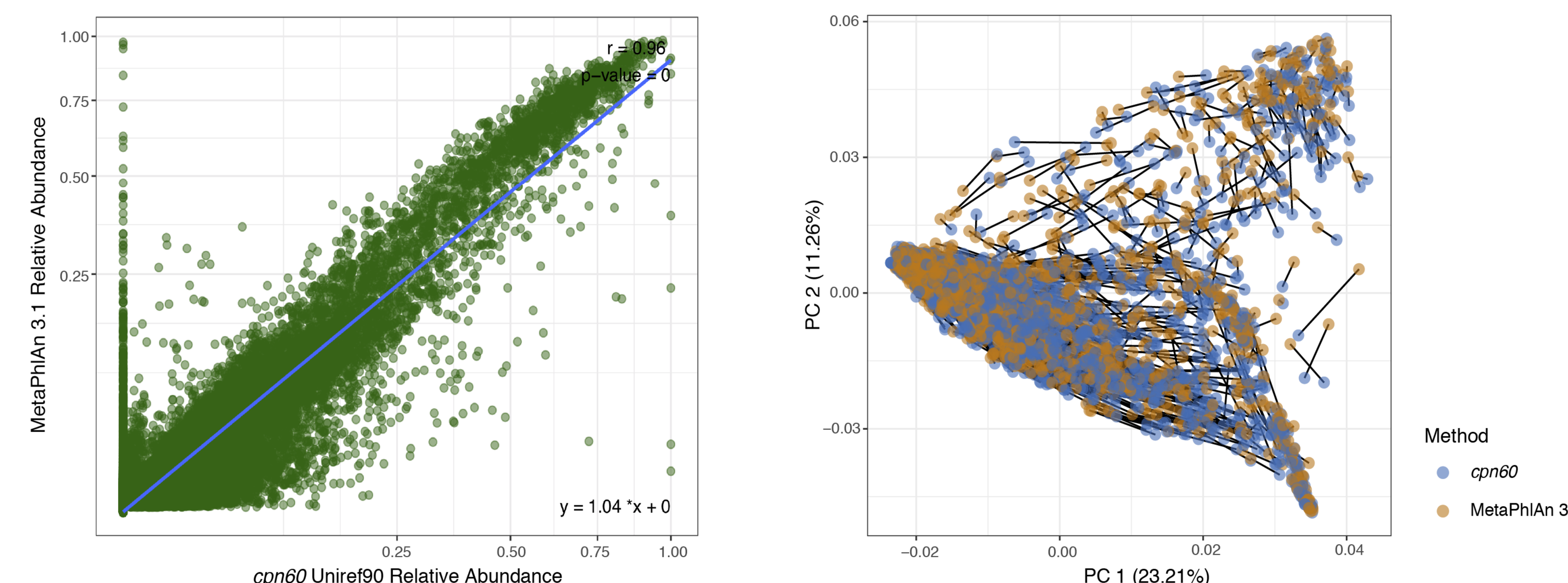



## *Cpn60*-amplicon-sequencing can be used as a supplement of 16S rRNA sequencing

- Taxonomy profiles differed notably between 16S-seq and *cpn60*-seq at the phylum level.
- *cpn60*-based method identifying fewer Firmicutes, Bacteroidetes, Cyanobacteria, and Fusobacteria, but more Proteobacteria and Saccharibacteria.
- Higher within-phylum identities compared to between-phyla identities for both methods.
- Identities from 16S rRNA V4 region were generally higher than thosw from *cpn60* UT region.

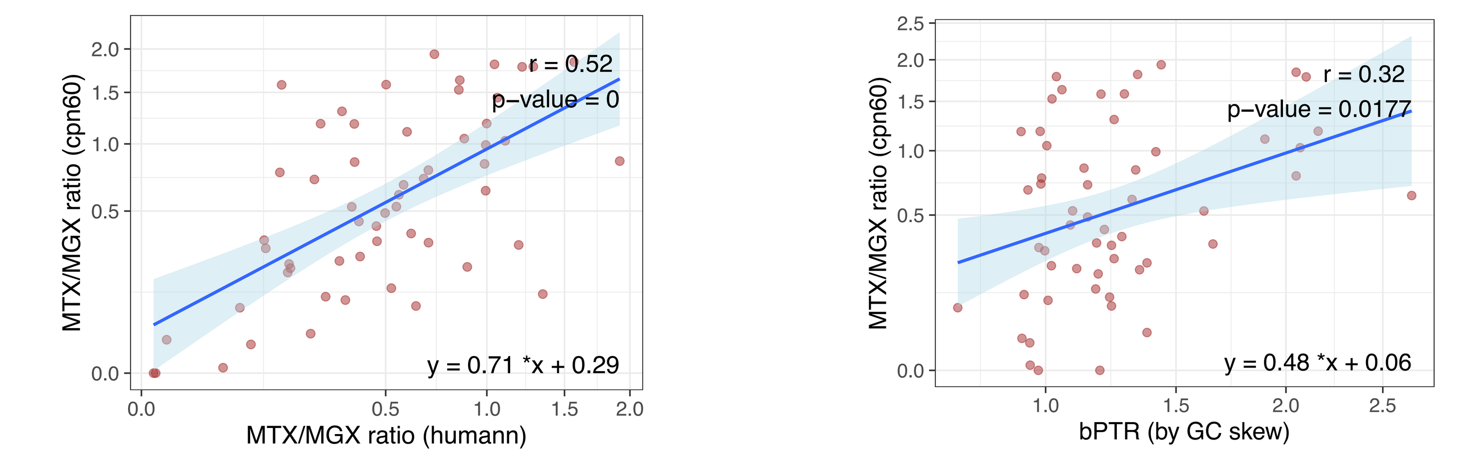
## *Cpn60* is a reliable marker for taxonomy profiling in human stool microbiome



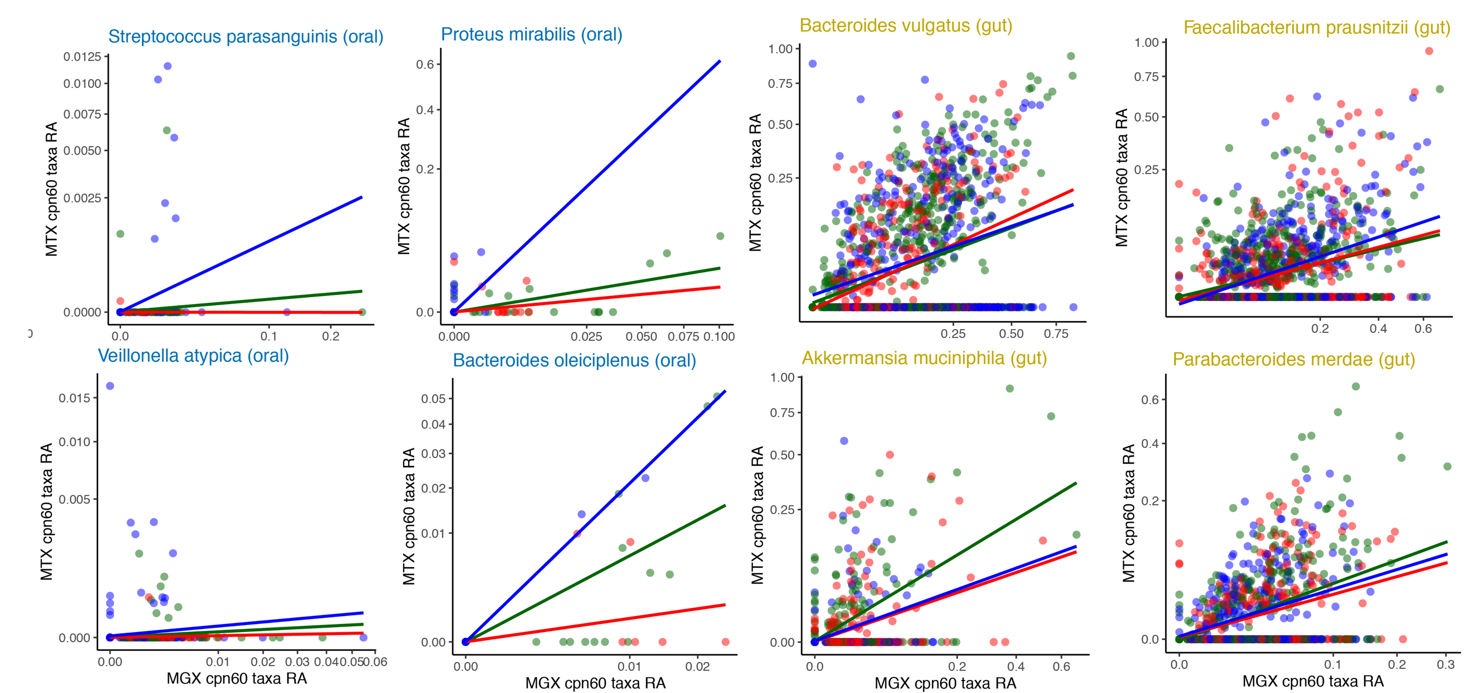

- In HMP2 dataset, both MetaPhlAn 3.1 and *cpn60* profiling showed high agreement at various taxonomic levels.
- Genus-level relative abundances revealed strong agreement between the two methods (correlation coefficient = 0.96, *p*-value = 0.00).
- Comparable compositional dissimilarity confirmed by Principal Coordinate Analysis, highlighting the compatibility of both approaches.

## *Cpn60* transcripts can be used for community activity assessment

- Evaluate protein-based activity assessment using *cpn60* Uniref90 and all Uniref90 IDs from HUMAnN 3.6, alongside the peak-to-trough (PTR) coverage ratio method in the iRep tool.
- As sequencing depth exceeded 100, *cpn60*-based activity assessment increasingly correlated with bPTR values.
- *cpn60* transcript is a potential viability marker for discerning activity in complex microbial communities.




## *Cpn60*-based analysis reveals different activity of gut vs. oral microbes in IBD




Gut species showed stable relative abundance profiles across Crohn's Disease (CD), ulcerative colitis (UC), and healthy control conditions, whereas oral bugs exhibited elevated transcriptional activity in CD. Conversely, among oral species, *K. pneumonia* displayed increased transcription in CD but not in UC compared to healthy controls, while *S. parasanguinis* and *P. mirabilis* were increasingly transcribed in UC, but not CD, relative to healthy controls.

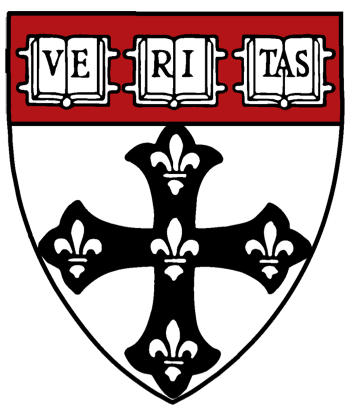
## Ongoing works

We are currently integrating additional house-keeping genes into marker gene-based microbial community characterization. This would enhance accuracy and provide a more comprehensive view of microbial community physiology, leading to refined sequencing workflows for more accurate and cost-effective community profiling and viability assessment processes,

## Acknowledgments

The works has been supported by Grant U19AI110820 and the DFSA Incubation Award from the Harvard Chan Dean's Fund for Scientific Advancement.

http://huttenhower.sph.harvard.edu



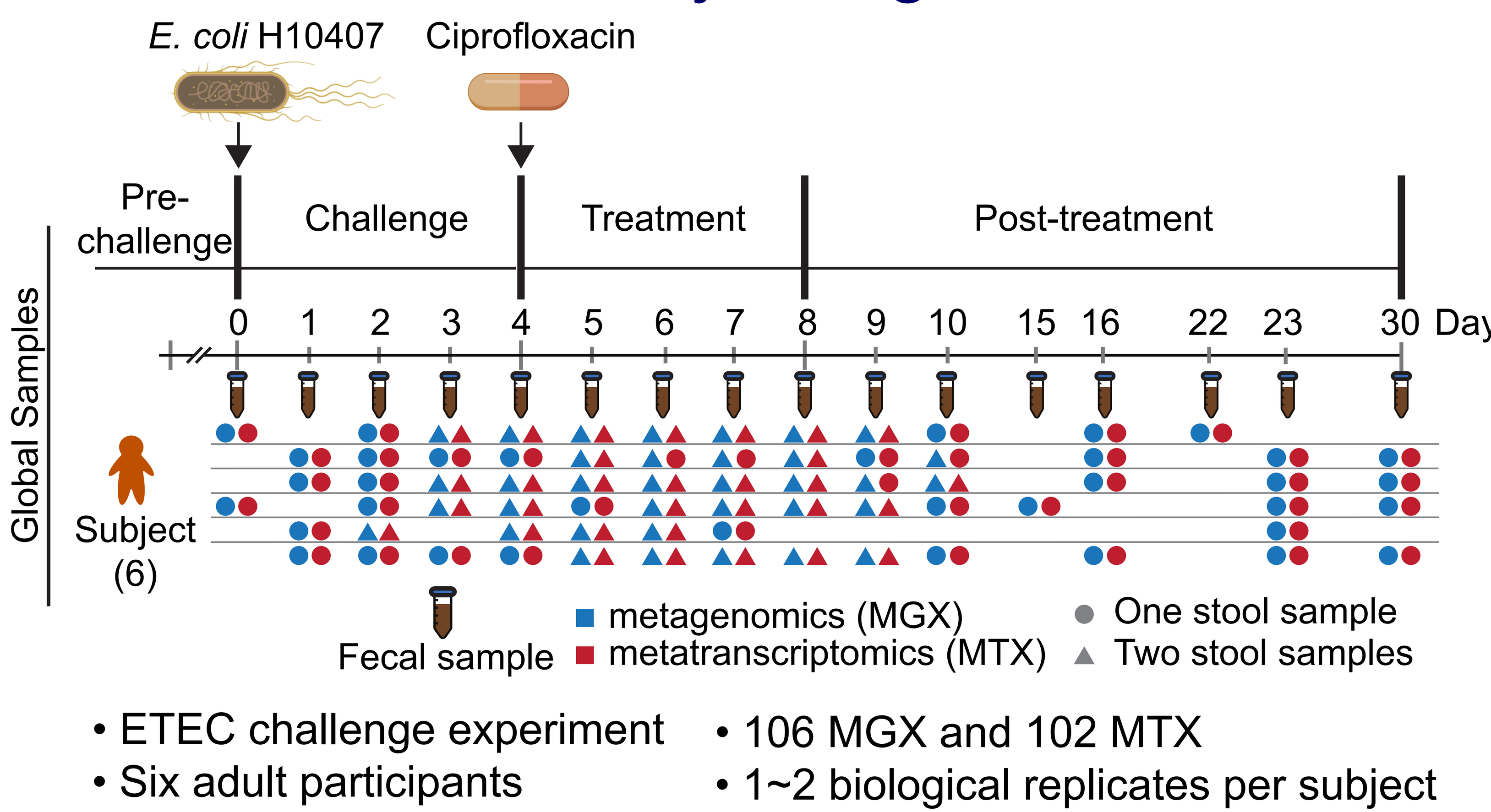
# Response of the gut microbiome to acute enteric pathogen infection and antibiotic treatment

Yancong Zhang[1,2,3], Sagun Maharjan[2,3], Anup Mahurkar[4], Jane M. Michalski[4,5], Taylor K.S. Richter[4,5], Wilbur H. Chen[6], Eric A. Franzosa[1,2,3], Curtis Huttenhower[1,2,3], David A. Rasko[4,5]

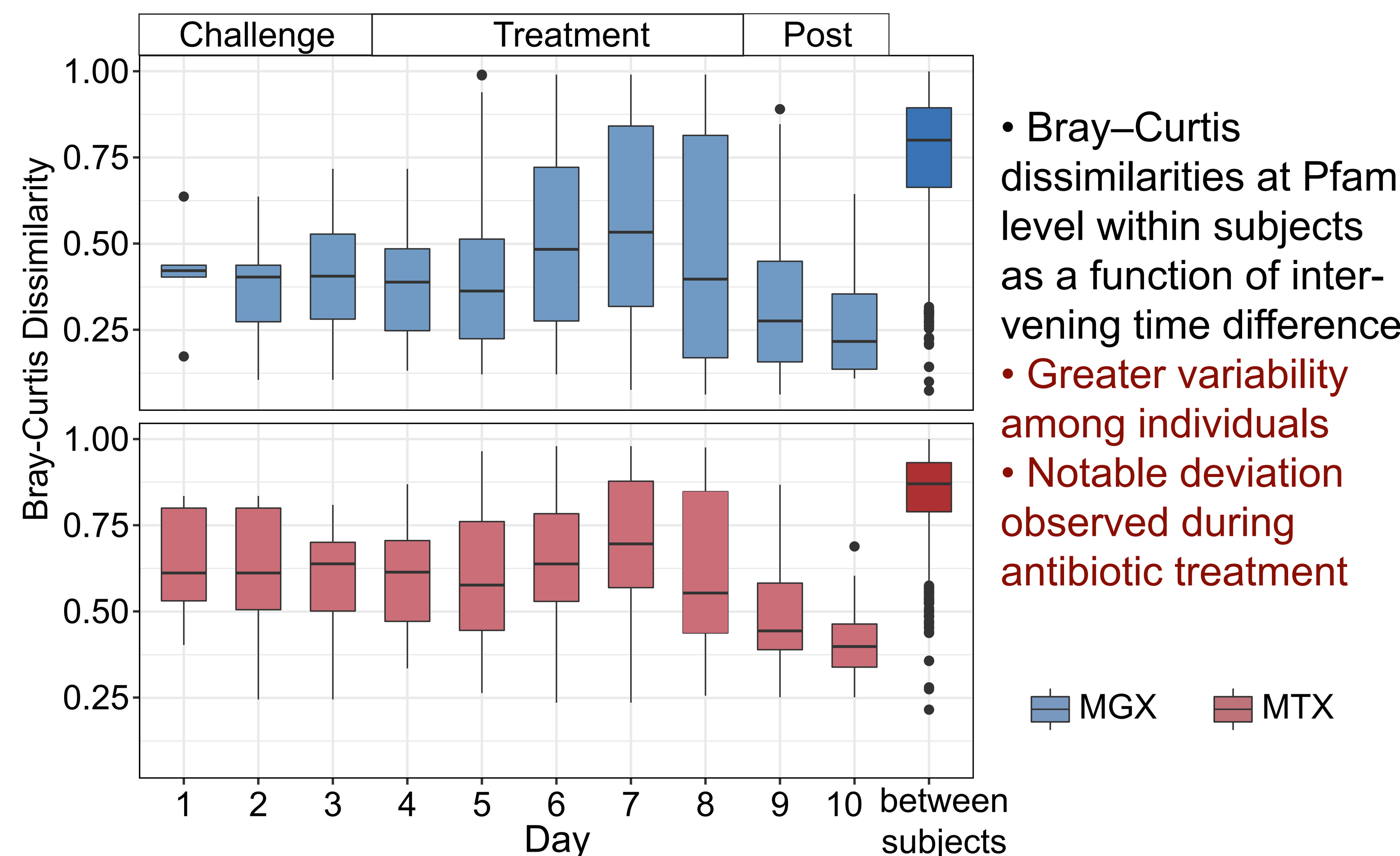
[1]Broad Institute; [2]Harvard T. H. Chan School of Public Health; [3]Harvard Chan Microbiome in Public Health Center; [4]Institute for Genome Sciences, [5]Department of Microbiology and Immunology, [6]Department of Medicine, University of Maryland School of Medicine

Enterotoxigenic *Escherichia coli* (ETEC) is a major global cause of diarrheal illnesses, particularly in low- and middle-income countries. Despite its significance, our knowledge of how the gut ecosystem responds to ETEC infection and antibiotic treatment is limited. We conducted a comprehensive study on six adults, analyzing their gut microbiome before, during, and after controlled exposure to ETEC and antibiotic recovery. The gut microbiome's response varied widely among individuals. Some species, such as *Ruminococcus bromii* and *Lachnospira eligens*, were more abundant in severe cases, likely due to microbial growth dysfunction during diarrhea. Transcriptional changes revealed insights into how the gut microbiome reacts to pathogenic challenge and antibiotic treatment, with species-specific metabolic processes identified. This study enhances our understanding of the gut microbiome's role in responding to enteric pathogens and antibiotics, offering valuable insights for mitigating effects in both adults and vulnerable infant populations.


## Study design



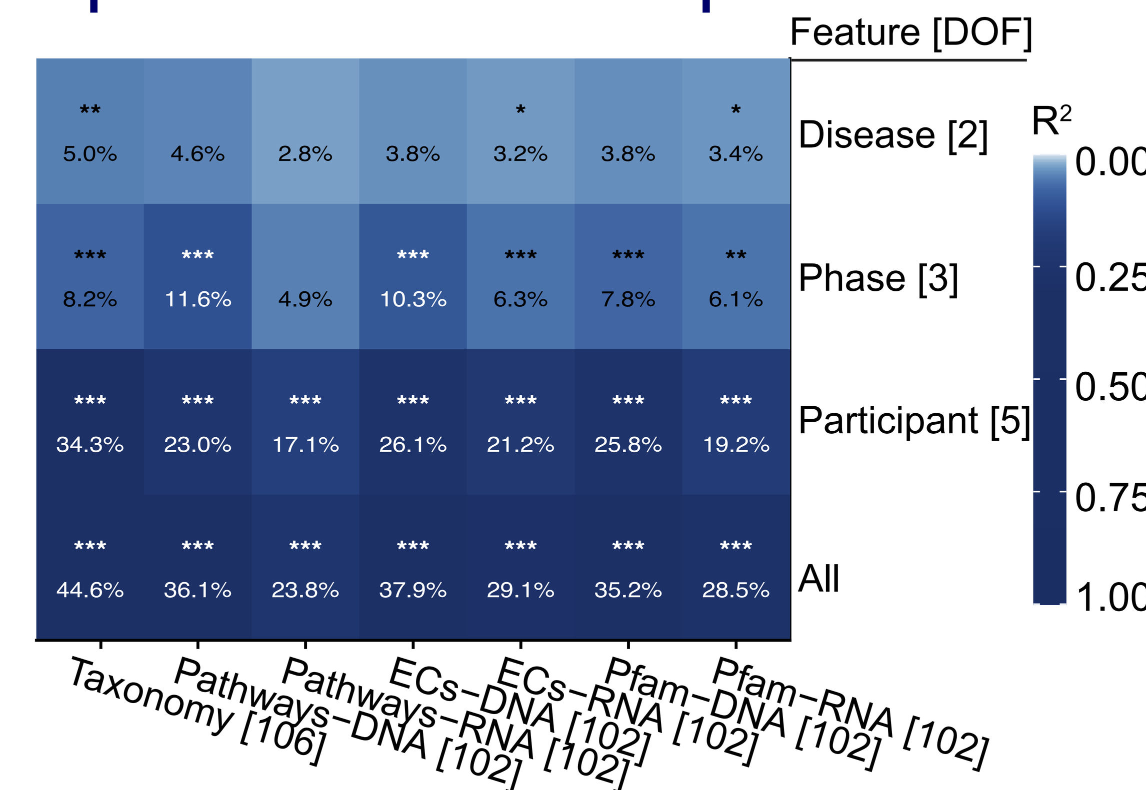

## Microbial divergence changes over time




- Bray–Curtis dissimilarities at Pfam level within subjects as a function of intervening time difference
- Greater variability among individuals
- Notable deviation observed during antibiotic treatment

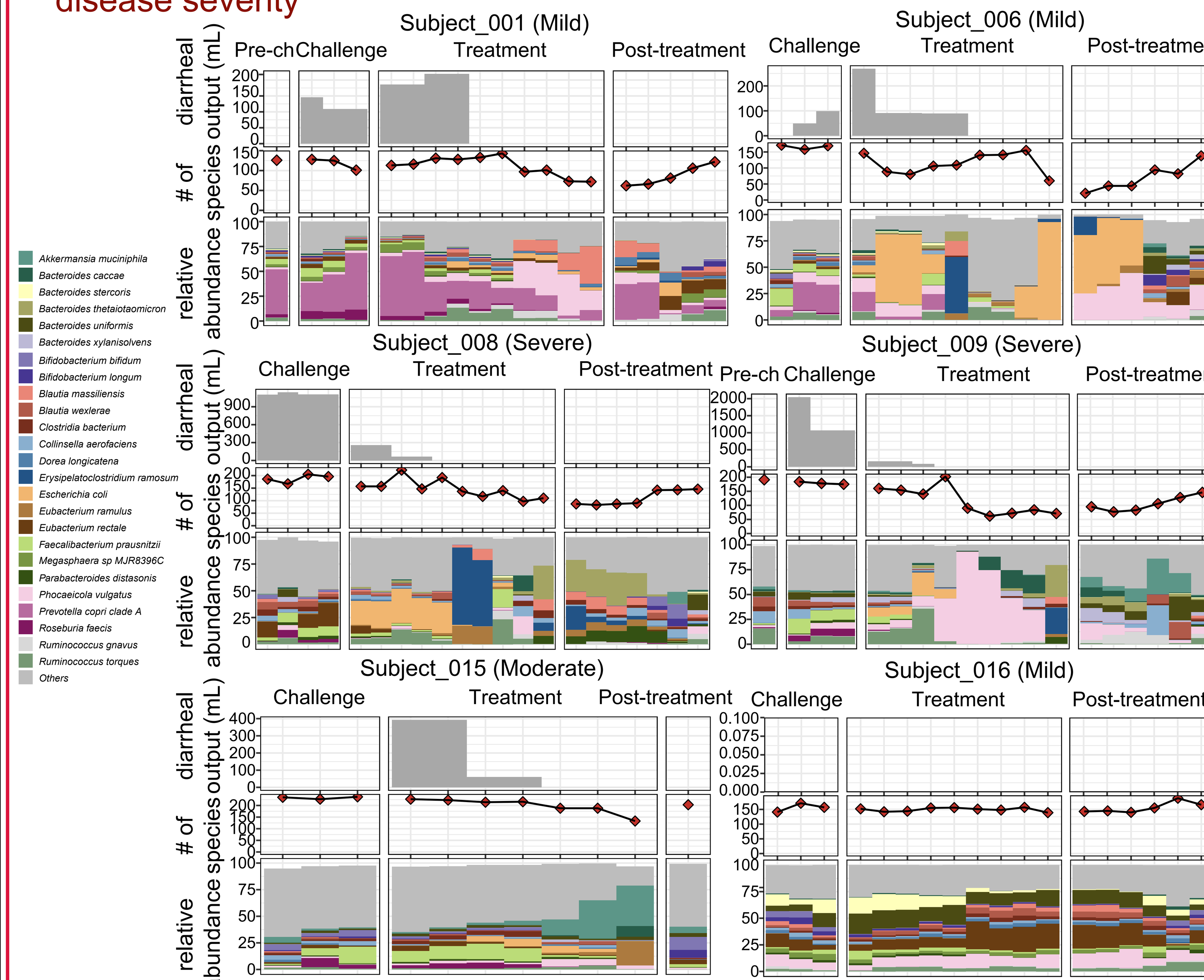
## Individual-specific microbial profiles

- Cohort characteristics are linked to microbial measurements by PERMANOVA
- Individual diversity explained the largest variation across measurement types
- Disease severity accounted for a smaller portion of variability




### Dynamics of species abundances in the human gut

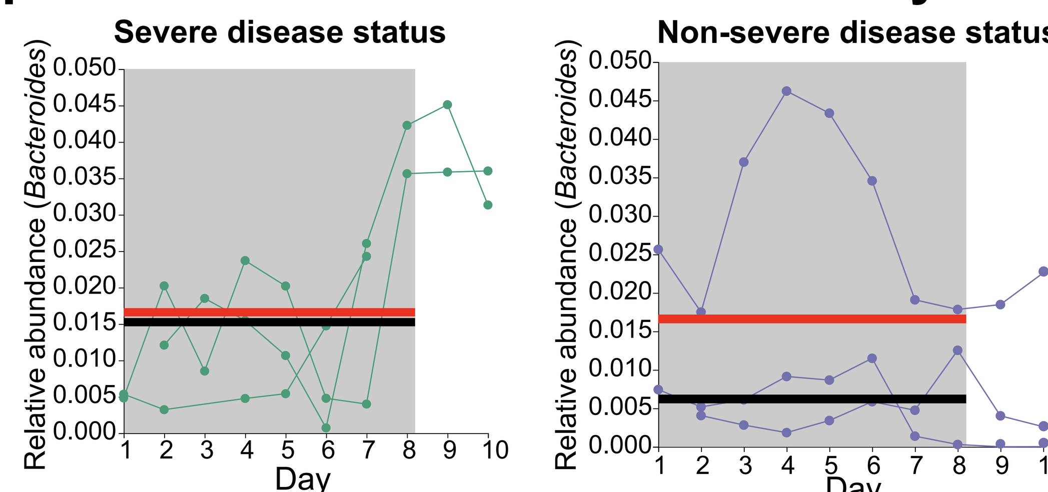
- Distinct dominant taxa in each participant during challenge
- Compositions rebounded post-antibiotic, showing variable stability linked to disease severity




### High-confidence prediction for host disease severity

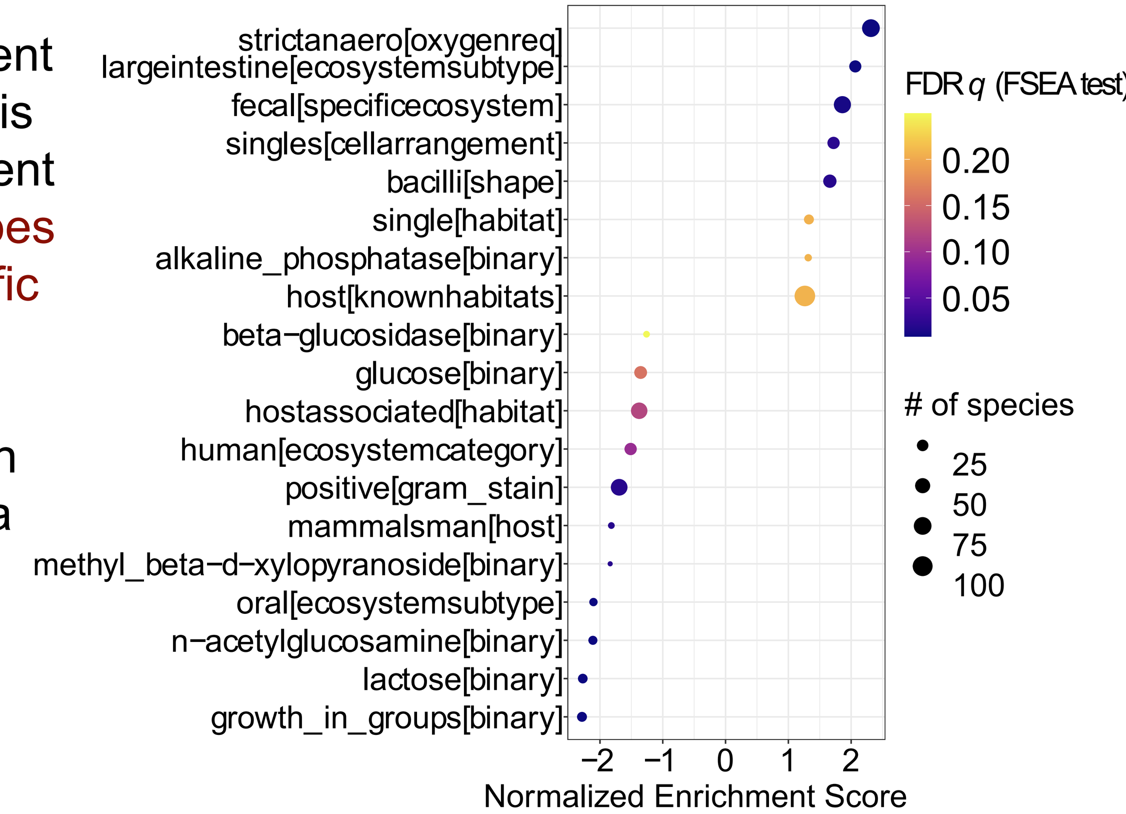
- Predict participant disease status from longitudinal microbiota data
- MITRE distinguished host disease severity with high confidence

- Red lines: threshold slopes/abundances
- Black lines: median slopes/abundances

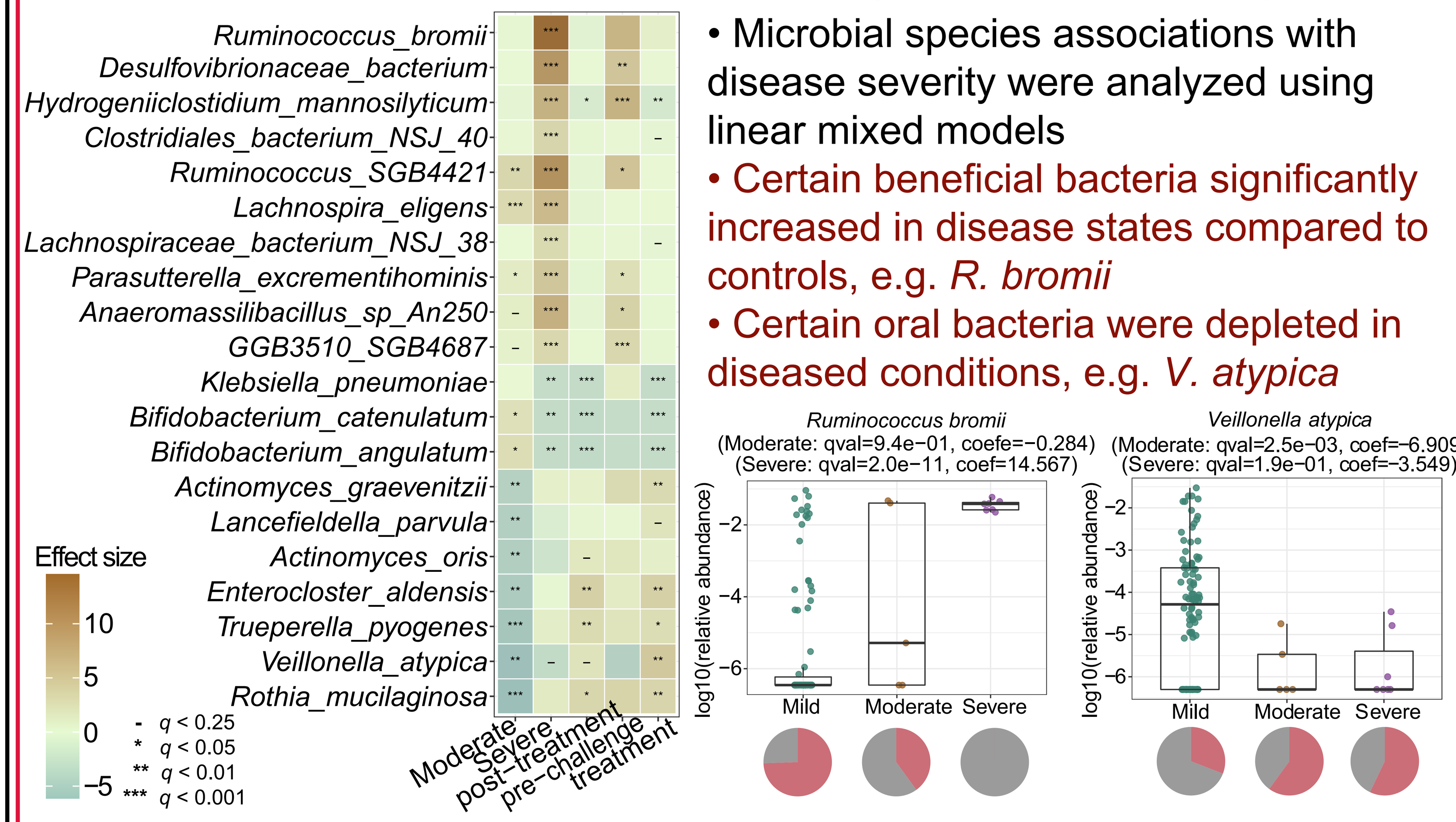



## Disruption of microbial profiles

- A feature-set enrichment analysis (FSEA) analysis for phenotypic enrichment
- Disease-linked microbes showed enriched specific phenotypes
- Gut microbes showed significant enrichment in strict anaerobic bacteria
- Oral bacteria showed significant negative enrichment

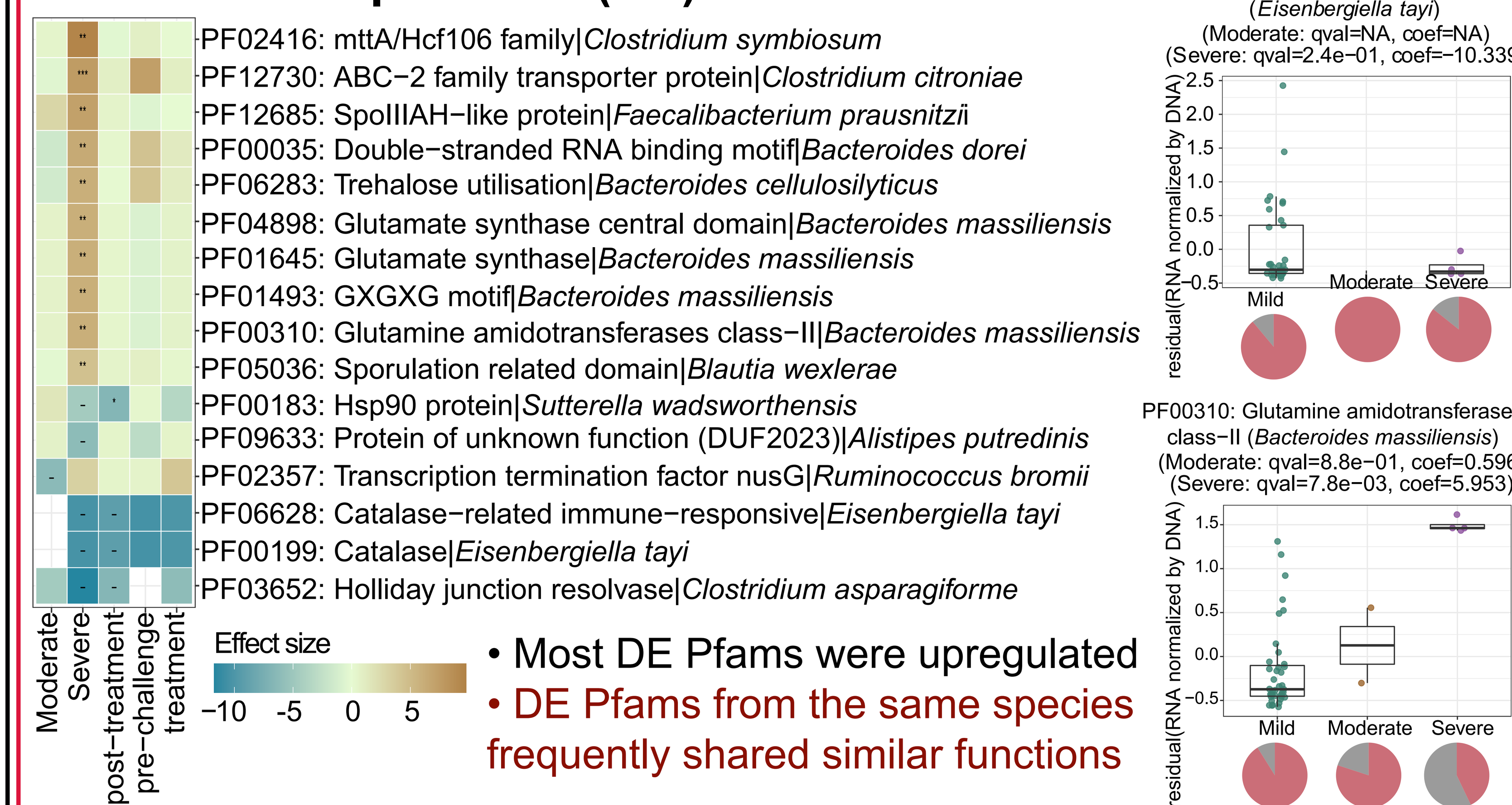



### Species abundance shifts among disease phenotypes




- Microbial species associations with disease severity were analyzed using linear mixed models
- Certain beneficial bacteria significantly increased in disease states compared to controls, e.g. *R. bromii*
- Certain oral bacteria were depleted in diseased conditions, e.g. *V. atypica*

### Differential expression (DE) of microbial functions




- Most DE Pfams were upregulated
- DE Pfams from the same species frequently shared similar functions

## Acknowledgements


This work has been supported in part by NIH HHS grant U19AI110820 (D.A.R) and NIDDK grant R24DK110499 (C.H.).