Searching for a consensus among inflammatory bowel disease studies: a systematic meta-analysis

Abdel Rahman L[1], Morgan XC[1]

1.Department of Microbiology and Immunology, University of Otago, Dunedin, Otago, New Zealand

Project Overview

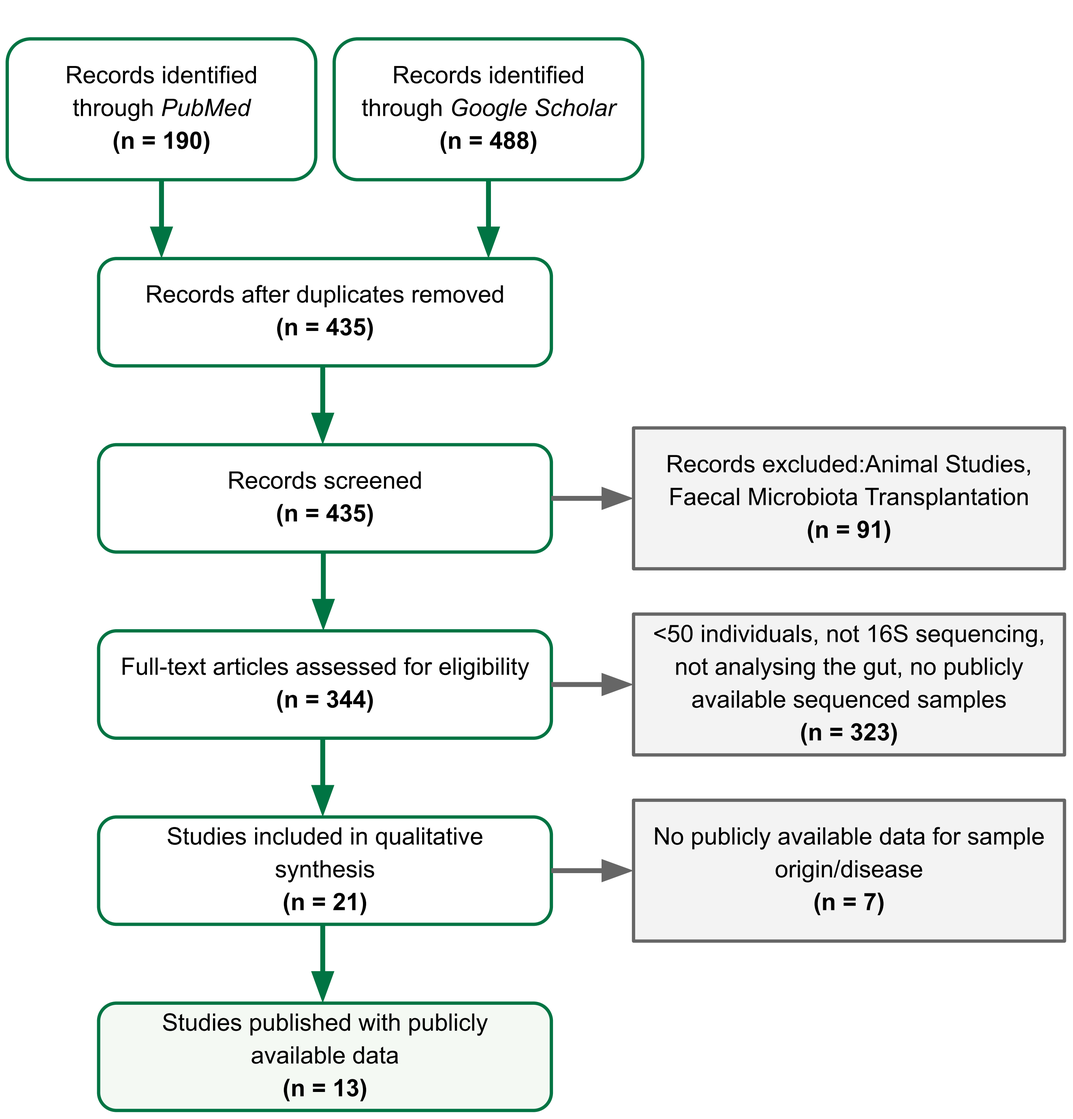
Numerous studies have examined the gut microbial ecology of patients with Crohn's disease (CD) and ulcerative colitis (UC), but IBD-associated taxa and ecological effect sizes are not consistent between studies. We aimed to find consensus on ecological effect sizes and direction associated with IBD.

We performed a meta-analysis of 13 studies to analyse how variables such as sample type (stool, biopsy, and lavage) affect results in IBD gut microbiome studies.

To the best of our knowledge, this project represents the most up-to-date and largest systematic meta-analysis on the IBD gut microbiome using 16S rRNA sequencing data.

Approach

We systematically searched PubMed and Google Scholar for IBD studies that met our criterion



Uniform bioinformatics methods were used throughout this project for analysis. All datasets were processed through DADA2.

Conclusions

IBD has a consistent effect on taxa richness and microbiome community compostion

There is variation in microbiome results associated with IBD with different sample types.

Our results suggest that stool type may be superior to biopsy due to decreased heterogeneity.

Results

IBD patients are significantly more likely to have fewer bacterial taxa in their gut microbiome than controls; however, heterogeneity is high

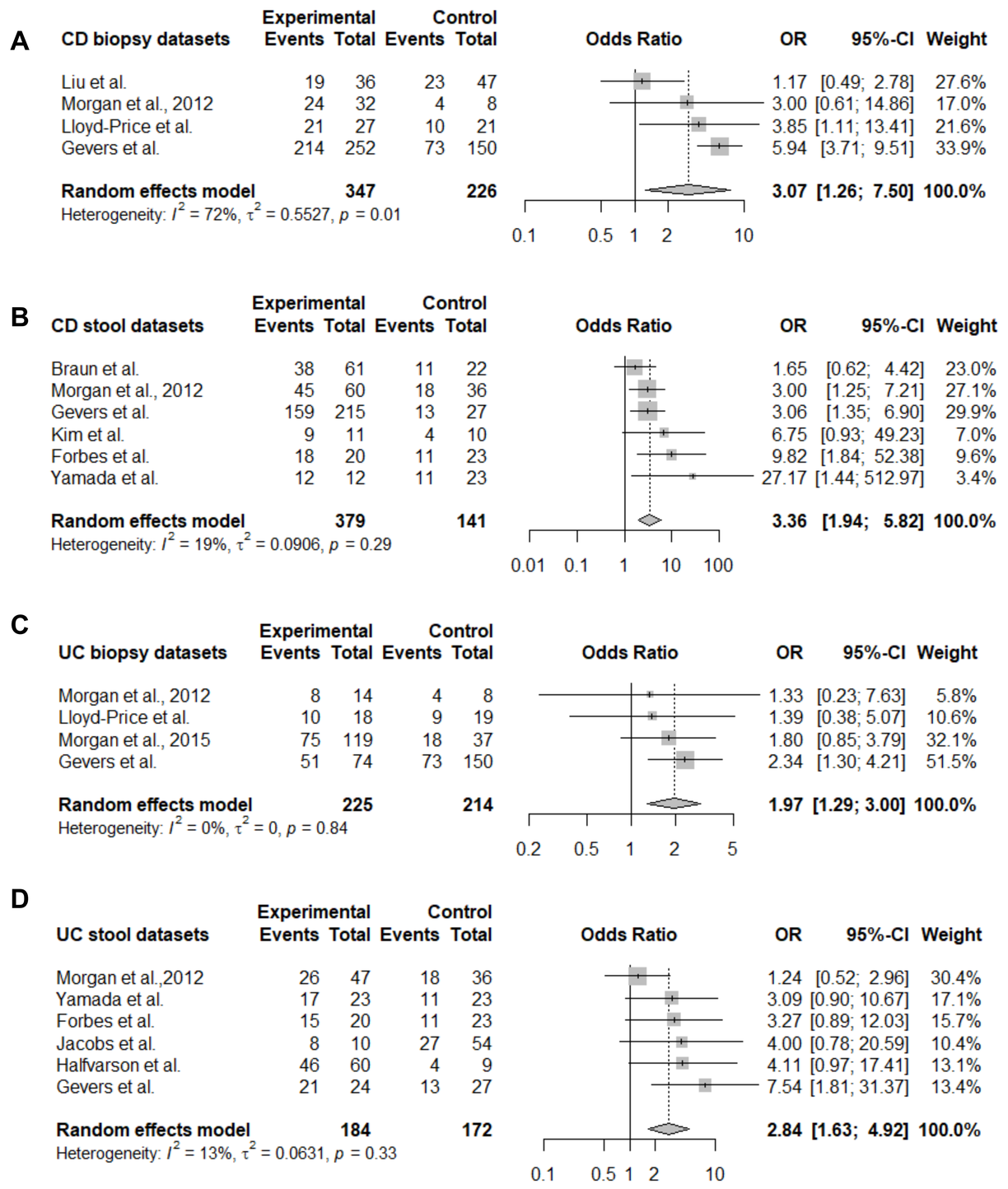


Figure 1. UC and CD patients are more likely to have less bacterial taxa in their gut microbiome than controls p<0.05 (**A:** CD biopsy, **B:** CD stool, **C:** UC biopsy, **D:** UC stool). P values on each figure is that of heterogeneity (I2). I2>50% indicates high heterogeneity in results across studies. Odds ratios were used to compare the proportion of cases and controls with observed richness greater than the median of controls.

Both disease and sample type significantly contribute to variation in microbiome community composition

Sample type	Dataset	Disease		Sample type	
		R^2	P value	R^2	P value
Lavage	Mottawea *et al*	3.6%	n.s.		
Stool and biopsy combination	Kim *et al*	3.9%	*	8.4%	***
	Morgan *et al* 2012	2.9%	***	7.9%	***
	Schirmer *et al*			8.3%	***
Stool	Forbes *et al*	10.1%	***		
	Yamada *et al*	12.9%	***		
	Jacobs *et al*	5.4%	***		
	Halfvarson *et al*	5.9%	***		
	Braun *et al*	1.9%	**		
	Gevers *et al* - stool	0.6%	n.s.		
Biopsy	Lloyd-Price *et al*	2.7%	n.s.		
	Liu *et al*	3.1%	**		
	Gevers *et al* -biopsy	1.5%	***		
	Morgan *et al* 2015	2.3%	**		

Table 1. Effects of disease and sample type on community structure. PERMANOVA analysis is based on Bray-Curtis distance. R2 = percentage of variation. *, **, ***, n.s. denote P<0.05, P<0.01, P<0.001 and not significant. Empty cells represent unavailable metadata.

Discovery of new bioactive microbial metabolites in Inflammatory Bowel Disease

Sena Bae[a], Amrisha Bhosle[b,c], Eunyoung Chun[a], Yancong Zhang[b,c], Grace Cao[a], Julian Avila-Pacheco [c], Jessica K. Lang[a], Kathryn G. Rosinski[a], Clary Clish[c], Ramnik Xavier[c], Hera Vlamakis[c] , Eric A. Franzosa[b,c] , Curtis Huttenhower[a,b,c*] and Wendy Garrett[a,c*]

[a]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[b]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston MA, 02115, USA

[c]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

Abstract

The gut microbiota and associated bioactive compounds have been implicated as causal and as protective factors in gastrointestinal disorders, including the inflammatory bowel diseases (IBD). Both host immune interactions with gut microbes and microbial small molecule products are likely responsible for these bioactivities. Several gut microbial metabolites, e.g. short-chain fatty acids and a subset of omega-3 fatty acids depleted in GI inflammation, have demonstrated therapeutic potential in IBD by attenuating gut inflammation. However, discovery of new bioactive compounds from the gut microbiome relevant to IBD or inflammation is challenging due to the vast numbers of uncharacterized metabolites produced by the microbiome.

To address this challenge, we investigated two IBD cohorts with integrated metagenomic and metabolomic profiles of the gut microbiome: PRISM, the Prospective Registry in IBD Study at MGH, and the Integrative Human Microbiome Project (HMP2). Putrescine and a potentially novel family of metabolites microbially derived from it were among the ~10,000 metabolites differentially abundant (PRISM n=8,792 and HMP2 n=9,444) during gut inflammation, of which only ~100 were characterized (PRISM n=157 and HMP2 n=99). We validated the dependence of these putrescine derivatives on the gut microbiome and their bioactivity *in vivo* by treating germ-free, gnotobiotic and conventional mice with dietary putrescine, which induced changes in immune system activity in a microbial community-dependent manner. This included that putrescine selectively affects host colonic and ileum M2 macrophage cell populations only in conventional mice. These results underscore the power of combined computational and experimental approaches for identifying microbially derived metabolites with general immunomodulatory activity and specific relevance for IBD patient care.

Introduction

Although there are highly effective IBD therapies that directly target the immune system, many IBD patients do not achieve durable remission, lose responsiveness to treatment over time, or suffer from the broad immuno-suppressive effects of such treatments. Despite the strong association of gut microbiome configurations with IBD and advances in taxonomical profiling of the gut microbiome, the effective translation of specific mechanisms of host-microbiota signaling and microbial metabolites for IBD clinical care remains largely elusive.

	Non-IBD Control (control)	Crohn's Disease (CD)	Ulcerative Colitis (UC)
PRISM	34	68	53
HMP2	27	67	38

Nontargeted Metabolites	Identified met
48,000	628
81,000	597

Table 1. IBD metabolomic datasets used for this project and IBD cohorts. Number of identified and unidentified metabolites in each data set.

Human fecal metabolomics, using untargeted high-resolution liquid chromatography-mass spectrometry (LC-MS), can provide comprehensive functional readouts of gut microbial activity and host-microbial interactions. Untargeted LC-MS techniques profile tens of thousands of metabolites in individual human stool samples; however, our understanding of their bioactivity is limited to ~<1% (Table 1). Thus, an *in silico* technique to prioritize these metabolites is a critical unmet medical need for realizing the potential of microbial metabolites for IBD treatment. We identified new IBD-associated uncharacterized metabolites using two publicly available IBD metabolomic datasets, PRISM and IBD, by MACARRoN and tested its biological function *in vivo*.

Validation of MACARRoN

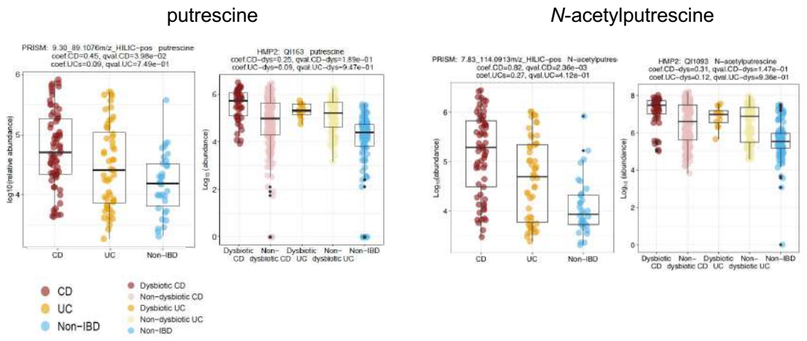


Figure 1. Putrescine and N-acetylputrescine were significantly enriched in patients with CD relative to control individuals, non-IBD, in the PRISM cohort, and in dysbiotic CD samples relative to non-dysbiotic control samples in the HMP2 cohort. Box plots show median and bottom and top quartiles, with outliers outside of box plot whiskers.

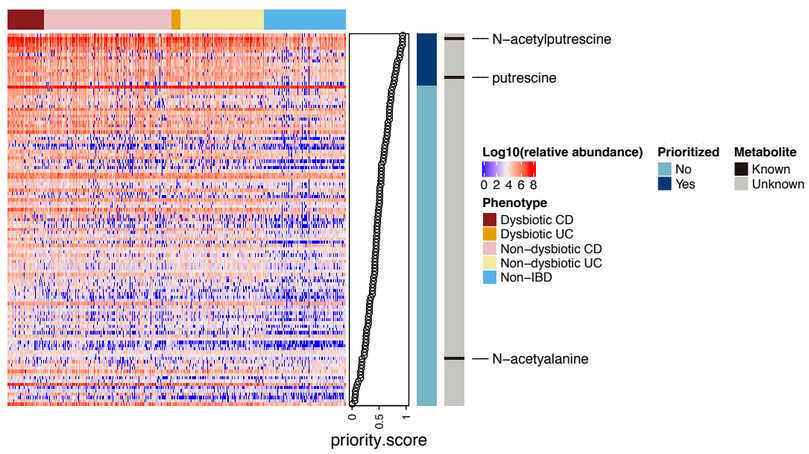


Figure 2. Heatmap representing putrescine/N-acetylputrescine module from HMP2 fecal metabolomics. This module was clustered after regressing out diagnosis, dysbiosis, age and medication use. The module contains 115 metabolites features and 16 metabolite features are significantly enriched and highly prioritized in dysbiotic CD samples.

Screen high-priority metabolites *in vivo*

Objective : (1) to demonstrate that the uncharacterized metabolites could be generated *in vivo* from a chemical precursor in a gut microbiota dependent process and (2) to evaluate a change in the host immune system in response to the precursor in a gut microbiota dependent manner.

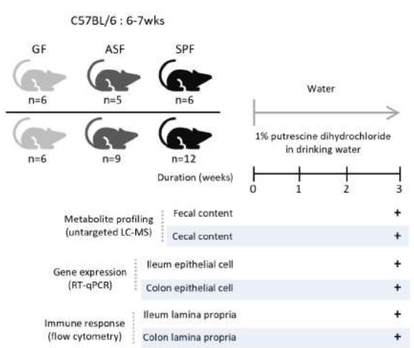


Figure 3. Schematic of the experimental design.

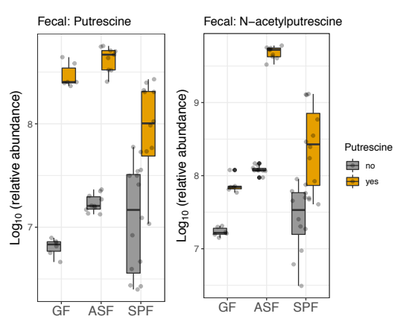


Figure 4. Abundance plots of putrescine and N-acetylputrescine. Putrscine acetylation is in creased in the presence of gut microbiota

We employed mice with distinct gut microbiota communities, germ free (GF), Altered Schaedler Flora (ASF, a minimal microbiota of 8 species), and SPF C57BL/6J mice, in the presence or absence of putrescine and profile their microbial activities, host gut barrier function, and immune cell phenotypes.

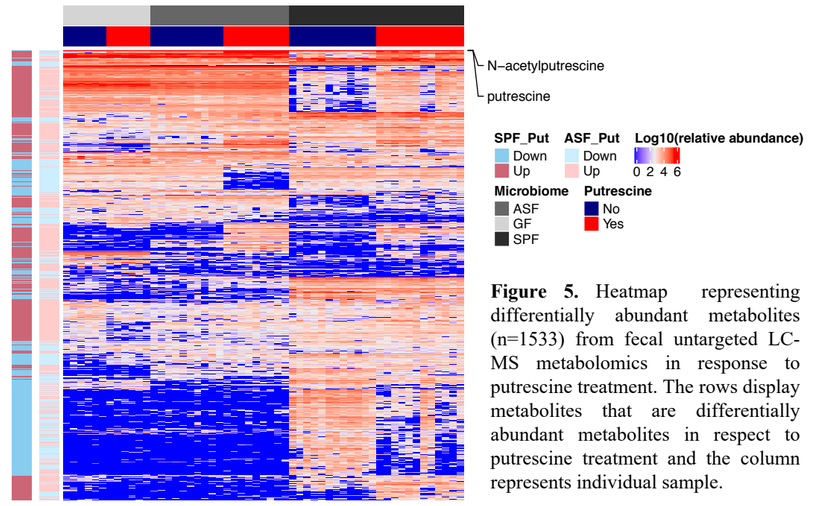
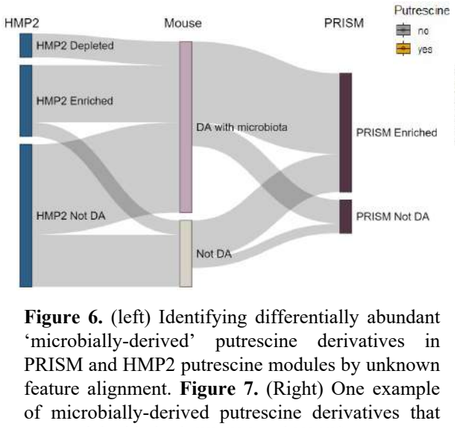


Figure 5. Heatmap representing differentially abundant metabolites (n=1533) from fecal untargeted LC-MS metabolomics in response to putrescine treatment. The rows display metabolites that are differentially abundant metabolites in respect to putrescine treatment and the column represents individual sample.



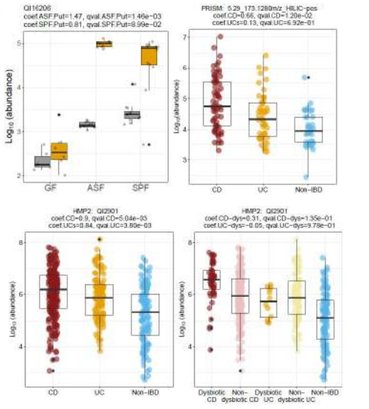


Figure 6. (left) Identifying differentially abundant 'microbially-derived' putrescine derivatives in PRISM and HMP2 putrescine modules by unknown feature alignment. **Figure 7.** (Right) One example of microbially-derived putrescine derivatives that was enriched in both PRISM and HMP2.

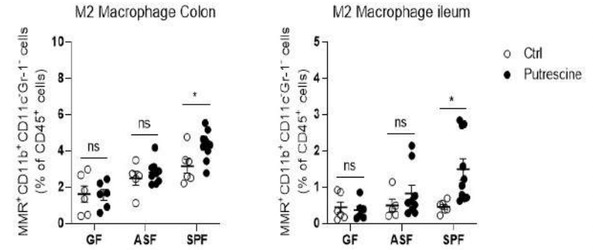


Figure 8. Lamina propria flow cytometry of colonic and ileal M2 macrophage cells, $MMR^+CD11b^+CD11c^-Gr\text{-}1^-$ cells out of $CD45^+$ cells, from GF, ASF, and SPF mice fed putrescine or control. Data shown as the mean ± SEM *p <0.05, two-tailed t-test.

Future Directions

- Characterize the chemical structure of the microbially-associated new bioactive metabolites followed by metabolite synthesis.
- Determine the efficacy of the bioactive metabolites in IBD preclinical mouse models.

Conclusion

- Putrescine selectively affects host colonic and ileum M2 macrophage cell populations in a gut microbiota- dependent manner.
- Putrescine level in gut is regulated by gut microbiota and microbially-derived putrescine derivatives are strongly associated with IBD phenotype.

References & Funding

1. Franzosa. *et al*. Gut Microbiome Structure and Metabolic Activity in Inflammatory Bowel Disease, *Nature Microbiology* (2019)
2. Lloyd-Price. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* (2019)
3. Quinn, R.A. *et al.* Global chemical effects of the microbiome include new bile-acid conjugations. *Nature*, (2020)
4. NIH NIDDK grant R24DK110499

Pathogenomic studies on dysbiosis in skin microbiome of leprosy patients undergoing drug regimen treatment from India

Nitin Bayal[1], Shekhar C. Mande[2], Milind S. Patole[1], Sharmila S. Mande[3], Sunil Nagpal[3], Monzoorul M Haque[3]

[1]National Centre for Cell Science, Pune

[2]Council of Scientific & Industrial Research, Anusandhan Bhavan, 2 Rafi Marg, New Delhi

[3]Bio-Sciences R&D Division, TRDDC, TCS, Pune

Introduction

Skin harbors a large repertoire of microbiota having symbiotic, saprophytic, commensal and opportunistic characteristics.

Skin is extremely affected in leprosy and the changes are irreversible.

Leprosy is caused by *Mycobacterium leprae* and *Mycobacterium lepromatosis*. Leprosy is treated by multi-drug treatment (MDT) therapy primarily includes Rifampicin, Dapsone and Clofazimine.

We have investigated and reported community structure of skin microbiota from lesional and non-lesional skin of Indian leprosy patients and healthy individuals as controls.

Aims and Objectives

Skin microbiota profiling and metagenomic predictions from skin of Indian leprosy patients (lesional and non-lesional skin sites) and healthy individuals.

Study Design

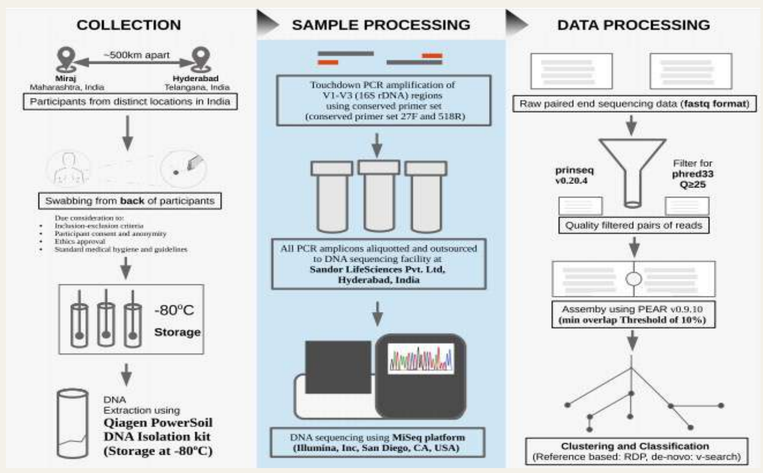


Figure 1 Schematic workflow depicting overall study design and methodology followed by sample collection, processing and data processing. (Bayal et. al., 2019)

Figure 2 Consent images of leprosy patients taken at leprosy rehabilitation clinics. All skin swab samples were collected from back having BI>3

Inclusion and Exclusion Criteria

Participant Type	Hyderabad		Miraj		Total Samples
Healthy (Control)	15		15		30
Patients	Affected (Lesions)	Unaffected	Affected (Lesions)	Unaffected	60
	14	14	16	16	

Table 1 An overview of number of study participants and samples collected from two sampling locations viz. Hyderabad and Miraj, India

Inclusion criteria

- Age between 21 to 70 years.
- Patients having a high (3+) bacillary index (MB/BL/LL).
- Active patches present on the patients back (scapular and lumbar regions).

Exclusion criteria

- Pregnant women.
- Patients co-infected with Leprosy-TB or Leprosy-HIV.
- Individuals with a different illness or those who have taken any medications in previous 10 days for any other symptoms.

Results

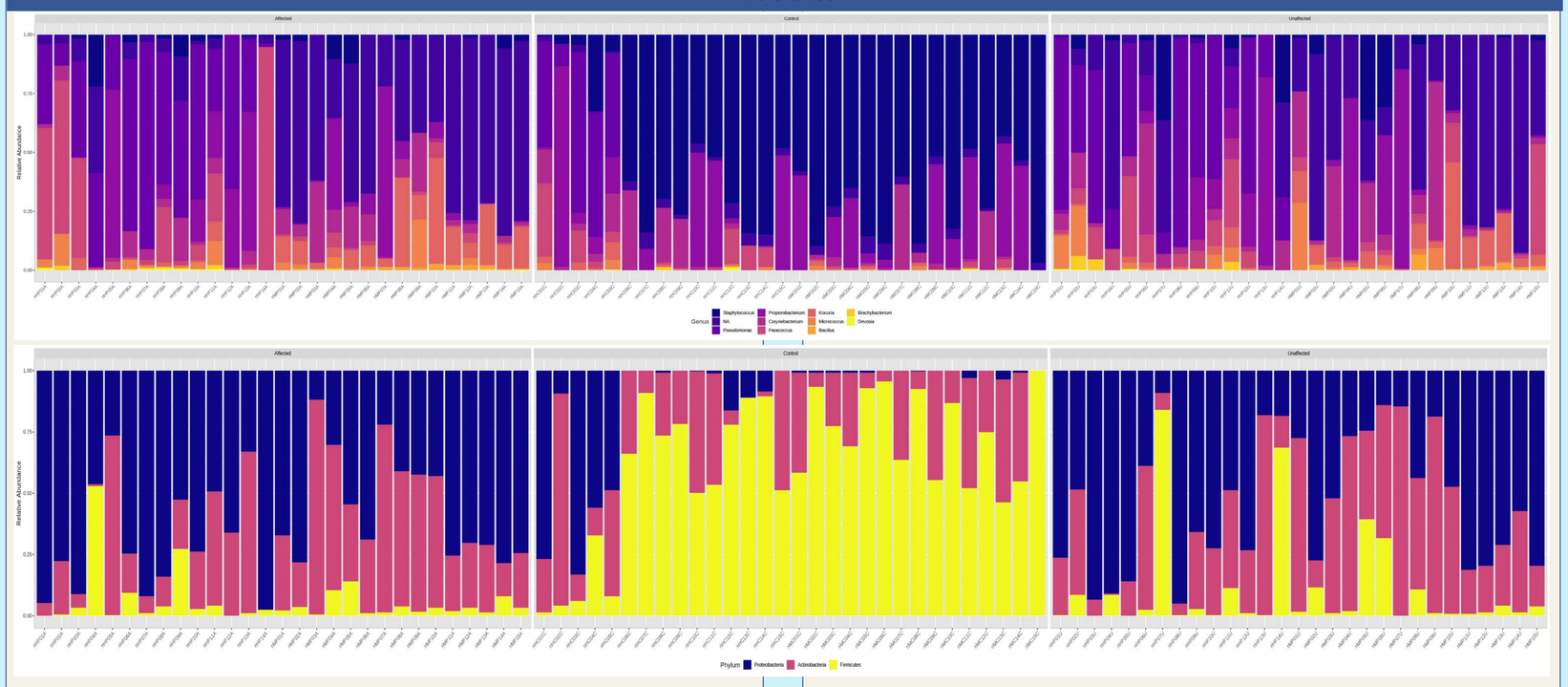


Figure 3 and 4 Taxonomic abundance profiles of lesional, non-lesional and control skin microbiota at the genera and phylum level.

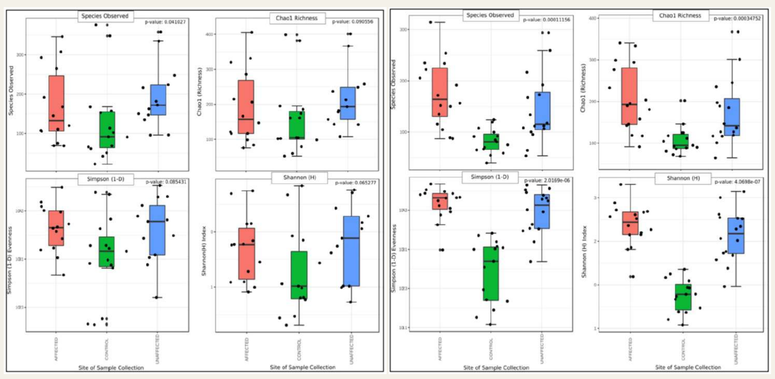


Figure 5 Box plots for the alpha diversity indices Chao1, Shannon, Simpson and observed species based on OTUs for Hyderabad and Miraj sampling locations separately.

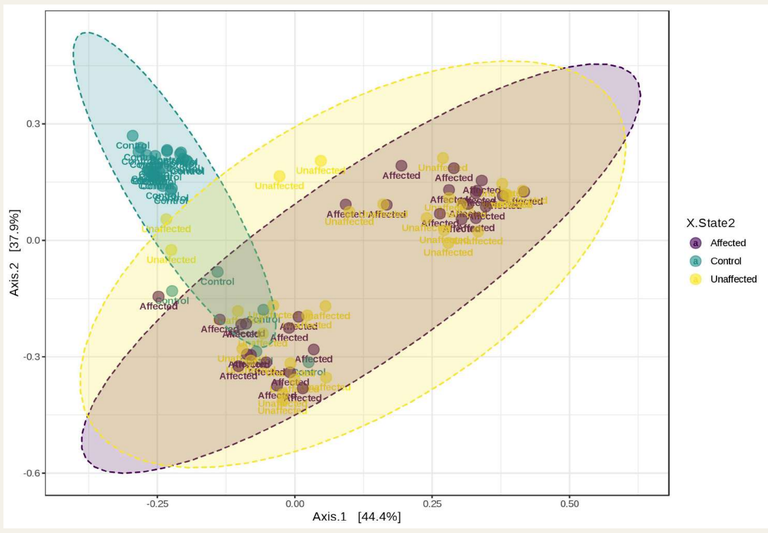


Figure 6 PCoA plots using using Jensen Shannon divergence distance metric and ANOSIM statistics for different study groups.

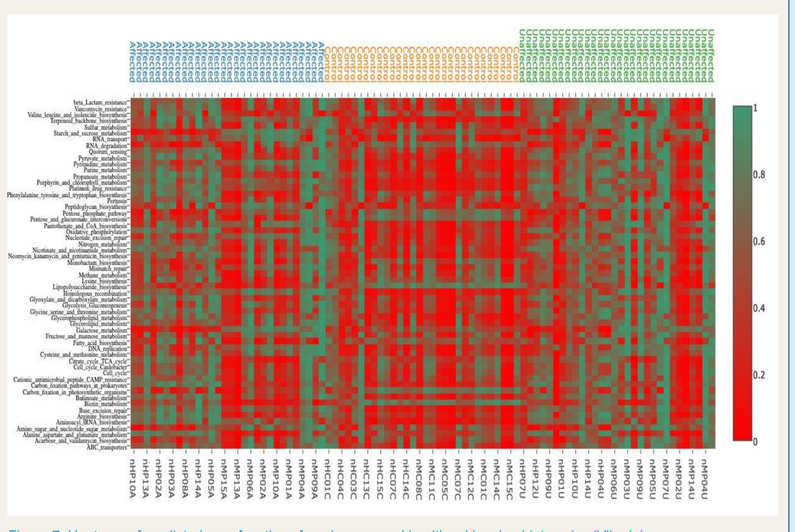


Figure 7 Heatmap of predicted core functions from leprous and healthy skin microbiota using iVikodak.

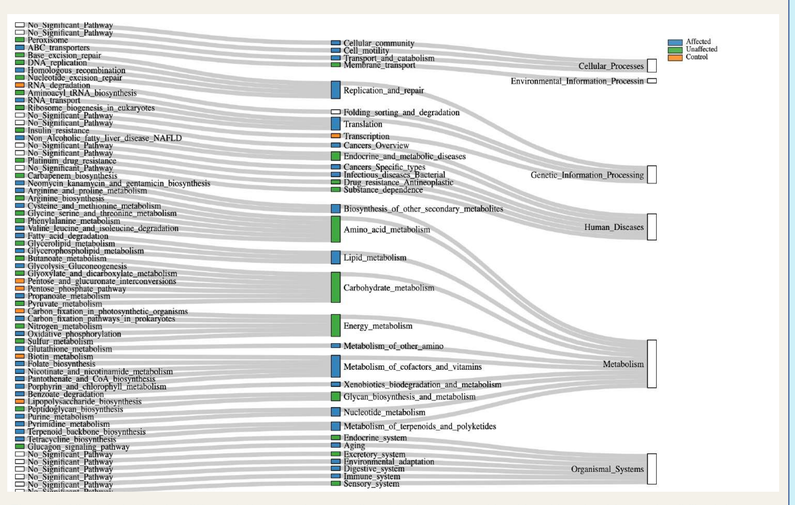


Figure 8 Alluvial plot for the differentiating functions predicted for leprous and healthy skin microbiota using iVikodak.

Conclusion

1. In both sets of samples (Hyderabad and Miraj), a stark difference is observed between the taxonomic profiles of skin microbiome samples of healthy controls as compared to that from participants affected with leprosy. Data Analysis indicates a distinct depletion of Staphylococcus in samples taken from leprosy affected participants.

2. There appears to be uniformity in skin microbiome profiles of healthy controls irrespective of geographical location (Hyderabad and Miraj). In contrast, the skin microbiome profiles of samples taken from leprosy affected participants from Hyderabad appear to have significant differences from that taken from affected participants from Miraj.

3. Absence of a distinct difference between samples taken directly from the leprosy lesions and those taken from adjoining non-lesional regions.

References

Bayal, N., Nagpal, S. et al. Structural aspects of lesional and non-lesional skin microbiota reveal key community changes in leprosy patients from India. Sci Rep 11, 3294 (2021).

Bayal N, Nagpal S, Haque MM, Patole MS, Valluri V, Suryavanshi R, Mande SS, Mande SC et al., 16S rDNA based skin microbiome data of healthy individuals and leprosy patients from India. *Nature Sc. Data* (2019) 6(1):225.

Rob Knight, Alison Vrbanac, Bryn C. Taylor et al., Best practices for analysing microbiomes. *Nature Microbiome* (2018) 16, 410-422.

Jacquelyn S. Meisel, Geoffrey D. Hannigan, Amanda S. Tyldsley et al., Skin Microbiome Surveys Are Strongly Influenced by Experimental Design. *JID* (2016) 136, 947-956.

Funding

This study was funded by Department of Biotechnology, Government of India under the SysTB project (DBT/PR3260/BRB/10/967/2011) and the Centre of Excellence grant (DBT/PR15450/COE/34/46/2016).

Acknowledgements

- **Doctoral Advisor - Dr. Shekhar C. Mande**
- **Co-Supervisor - Dr. Milind S. Patole**
- Collaborators - Dr. Sharmila S. Mande (TRDDC, TCS, Pune), Dr. Vijaya Lakshmi Valluri (LEPRA, Hyderabad) & Dr. Rohini Suryawanshi (TLM, Miraj)

Contact Information

NITIN BAYAL

Laboratory #2, New Building,

National Centre for Cell Science, Pune, INDIA

Phone: +91 9923 653 535

Email: nitin.bayal@protonmail.com

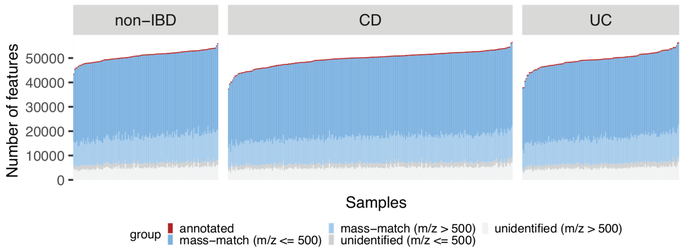
Identifying Novel Bioactive Metabolites in Inflammatory Bowel Disease

Amrisha Bhosle[1,2,3], Sena Bae[4], Yancong Zhang[1,2,3], Eunyoung Chun[4], Julian Avila-Pacheco[5], Ludwig Geistlinger[7,8], Jonathan Glickman[6,9], Monia Michaud[4], Levi Waldron[7], Clary Clish[5], Ramnik J. Xavier[1,10,11], Hera Vlamakis[1], Eric A. Franzosa[1,2,3], Wendy S. Garrett[1,3,4,12], Curtis Huttenhower[1,2,3,4*]

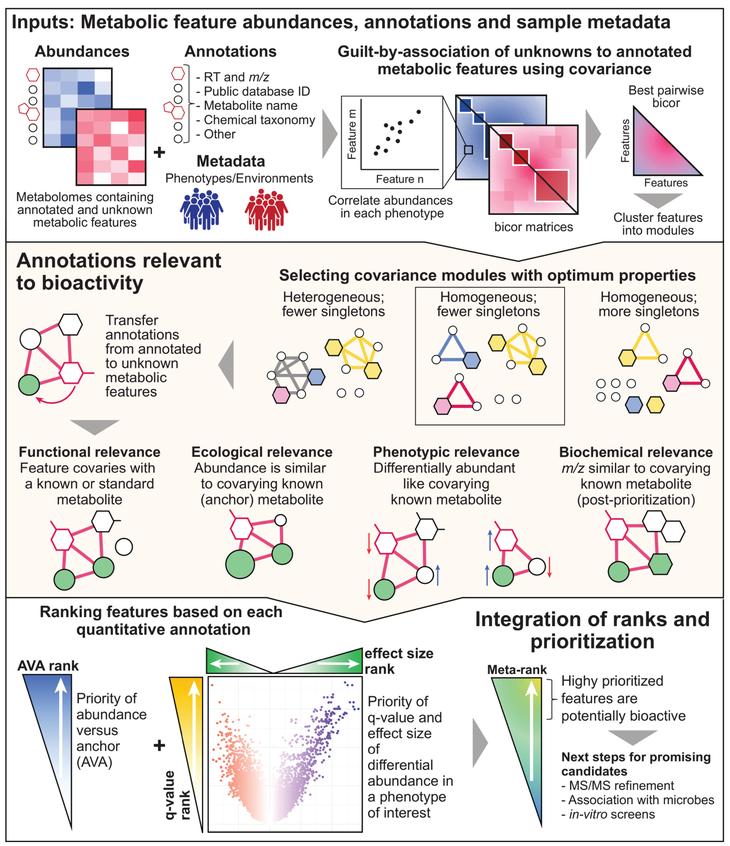
[1]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, [2]Department of Biostatistics, Harvard T. H. Chan School of Public Health, [3]Harvard Chan Microbiome in Public Health Center, Harvard T. H. Chan School of Public Health, [4]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, [5]Metabolomics Platform, Broad Institute of MIT and Harvard, [6]Department of Pathology, Harvard Medical School, [7]Department of Epidemiology and Biostatistics, Graduate School of Public Health and Health Policy, City University of New York, [8]Center for Computational Biomedicine, Harvard Medical School, [9]Beth Israel Deaconess Medical Center, [10]Gastrointestinal Unit and Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital and Harvard Medical School, [11]Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, [12]Department of Medical Oncology, Dana-Farber Cancer Institute

Chemical dark matter in IBD

Thousands of metabolites have been assayed from microbial communities, the gut microbiome in particular, but as yet with minimal biochemical characterization or knowledge of their therapeutic potential. Here, we developed a new approach, MACARRoN (Metabolome Analysis and Combined Annotation Ranks to pRioritize Novel bioactives), for identifying potential bioactives by integrating knowledge of annotated/standard compounds with phenotypic or environmental indicators of bioactivity to annotate and prioritize small molecules from microbiomes. We have applied this approach to identify novel bioactives from the inflammatory bowel disease (IBD) metabolomes in the Integrative Human Microbiome Project (iHMP/HMP2).



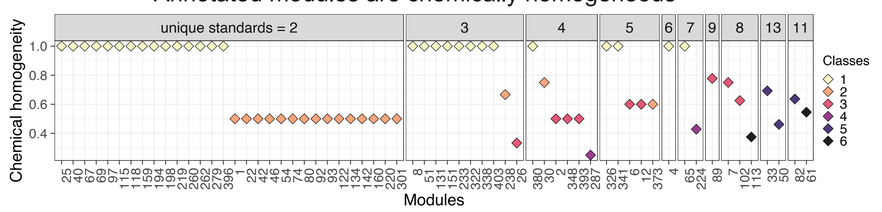
MACARRoN methodology

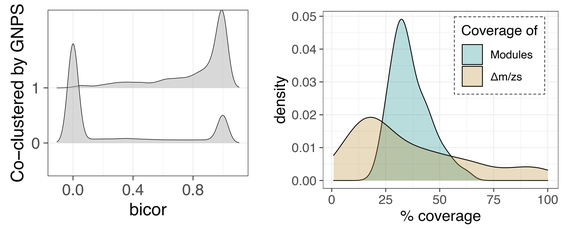


Quantitative metabolite annotations

To associate unannotated features with annotated metabolites i.e., standards, we clustered features based on covarying abundances in the different phenotypes i.e., CD, UC, and non-IBD. Of the 44,757 features that were prevalent in at least one phenotype, 43,043 features were clustered into 410 modules. Modules varied in size, ranging from 36 to 1,423 features; 138 modules had at least one standard in them and were called as “annotated modules”.

Annotated modules are chemically homogeneous

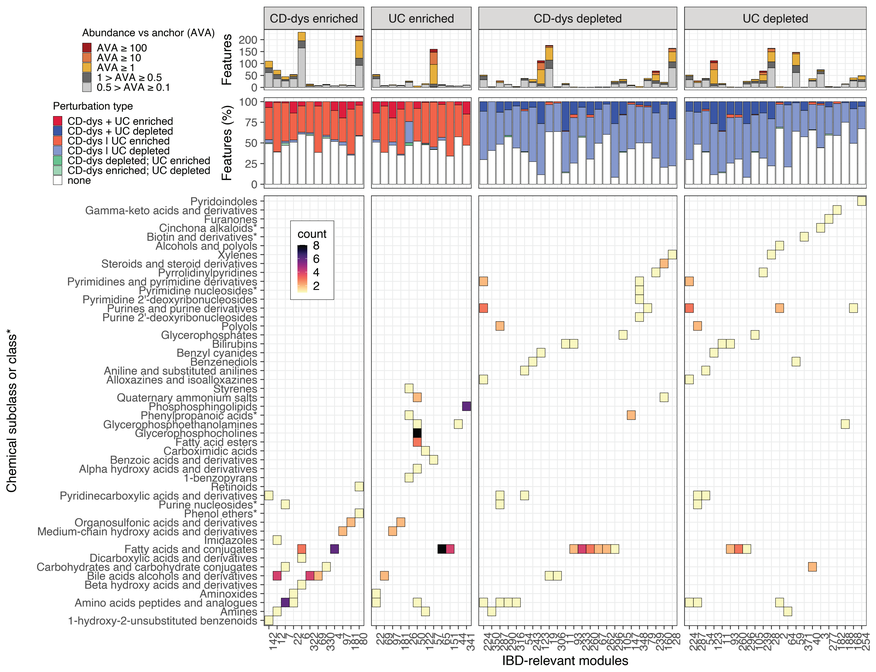




(I) Covarying metabolites tend to be chemically similar.
(r) Modules capture distinct $\Delta m/z$s whereas $\Delta m/z$s associated with common biochemical transformations are universally more prevalent.

Evidence of chemically diverse bioactives in IBD

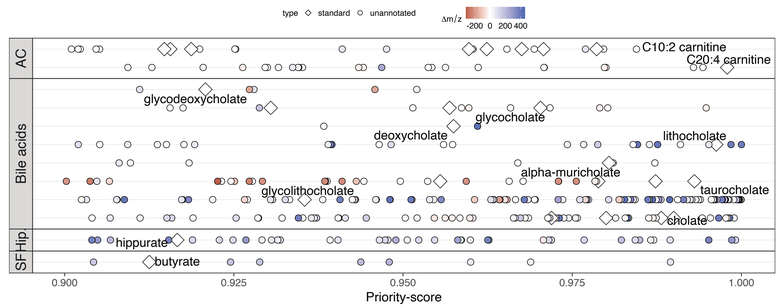
Unannotated features that are as abundant as a co-clustered standard (AVA ≥ 1) and perturbed in IBD covary with standards belonging to different chemical classes.



Prioritized bioactives

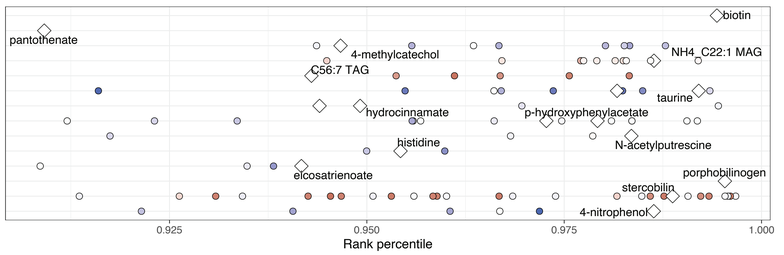
MACARRoN prioritizes known IBD-linked bioactives

Bile acids, acylcarnitines, and butyrate and several unannotated features that covary with them are highly-prioritized.



Metabolites with lesser-known roles in IBD are also highly-prioritized

Microbial metabolites including hydrocinnamate, stercobilin, and 4-methylcatechol are depleted whereas p-hydroxyphenylacetate and N-acetylputrescine are enriched in IBD.



Conclusions and future work

(1) MACARRoN integrates ecological, epidemiological, and biochemical annotations to prioritize metabolites.
(2) Novel highly-prioritized compounds covary with gut- or IBD-relevant metabolites, are abundant and significantly differentially abundant in IBD and dysbiosis.
(3) MACARRoN prioritizes known IBD-linked classes such as bile acids, acylcarnitines, and SCFAs which serves to validate the workflow.
(4) It also prioritized lesser-understood classes such as bilirubins, polyamines, and vitamins whose roles in IBD require further study.
(5) MACARRoN is generalizable to microbial community metabolomes and is available as a Bioconductor package.
https://bioconductor.org/packages/release/bioc/html/Macarron.html

Acknowledgements

This work has been funded by NIH NIDDK grant R24K110499.

@hutlab

https://huttenhower.sph.harvard.edu
https://huttenhower.sph.harvard.edu/MACARRoN

The aged mouse microbiome has obesogenic characteristics

(Dana Binyamin et al., Genome Medicine 2020)

Background

During aged, there is a physiological decline, increase of morbidity and mortality, and changes in the gut microbiome. As human lifespan is increasing worldwide, the morbidity associated with aged is becoming a serious public health concern.

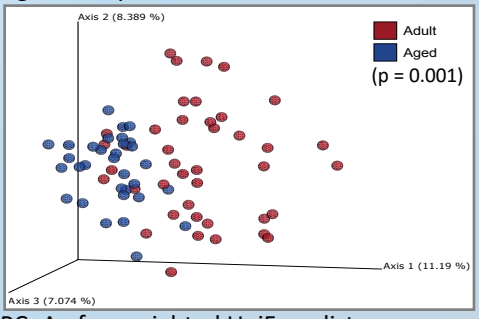
In this study, we investigated the influence of the gut microbiome on different metabolic parameters in adult and aged mice.

Methods

Fecal and blood samples from adult and aged mice were collected. Microbiome analysis was done using QIIME2. Weight and body composition were measured and Insulin and leptin levels in the blood were quantified. Fecal microbiota transplantation experiments from adult and aged mice into young germ-free mice were carried out in order to examine the effect of the gut microbiome on adult and aged mice

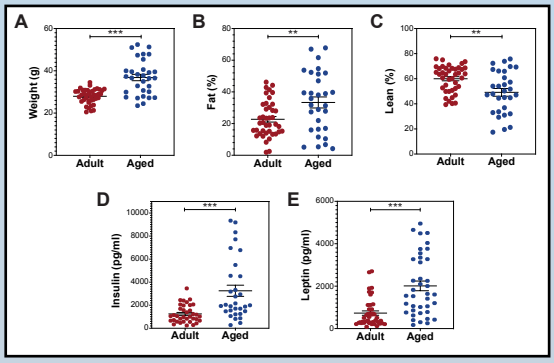
Result

The microbiomes of aged mice are significantly different then adults.

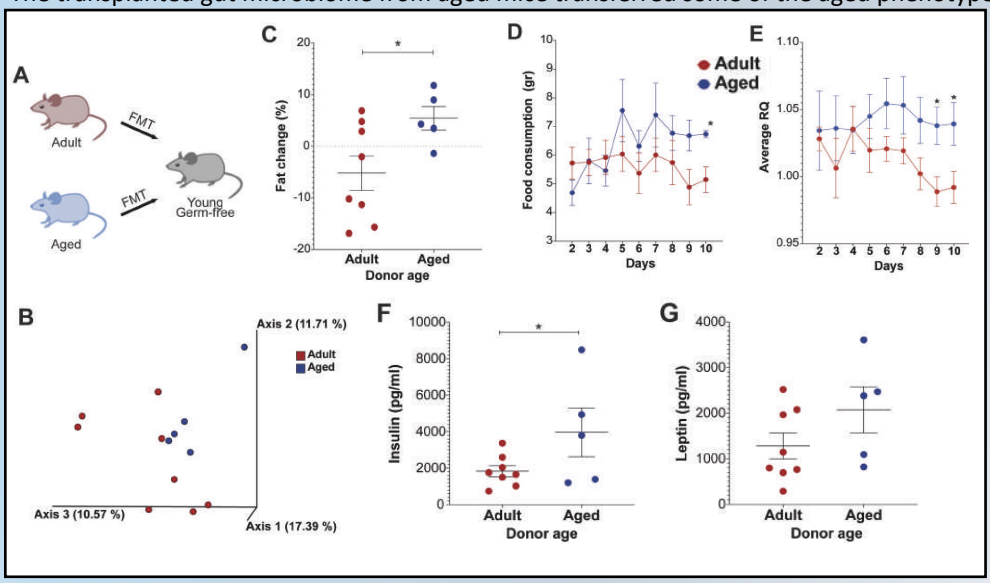


PCoA of unweighted UniFrac distances

Aged mice had significantly higher weight (A), fat mass (B) and low lean mass (C) than adult mice along with higher insulin (D) and leptin (E) levels in the blood.



The transplanted gut microbiome from aged mice transferred some of the aged phenotypes. The recipents microbiome was different (B). Fat body mass increased in the mice who received feces from aged mice (C). In addition, they consumed more food (D), had a higher respiratory quotient (E) and high insulin (F) and leptin (G) levels compared to mice receiving adult fece.



Conclusions

The gut microbiota of aged mice has obesogenic characteristics.

The gut bacterial population itself is sufficient to induce some of the manifestations of age-related obesity.

KORENLAB
The Azrieli Faculty of Medicine Bar-Ilan University
ADAMS Fellowships מלגות אדאמס
Dana Binyamin: dsimoni925@gmail.com

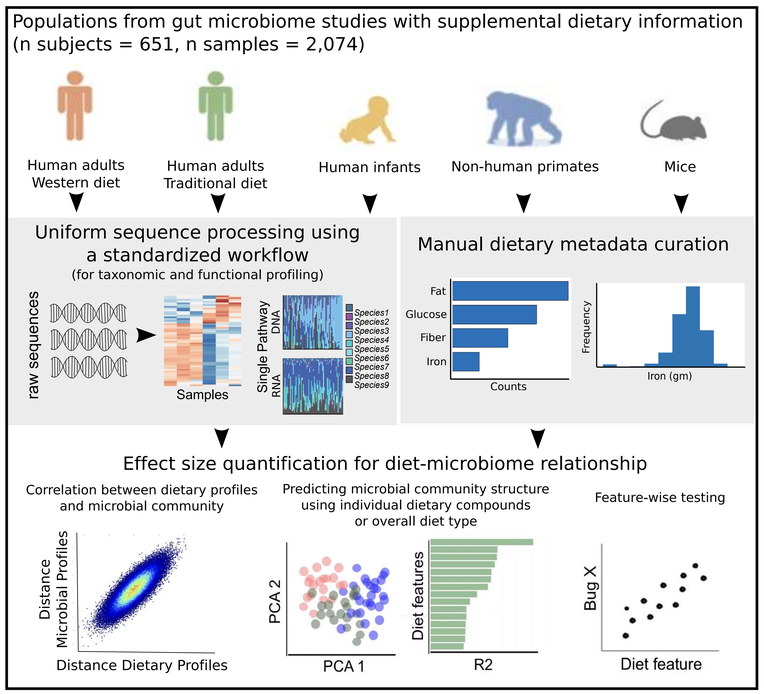
Quantifying the direct and indirect effects of diet and the microbiome

Tobyn Branck[1,2,3], Jason Lloyd-Price[3,7], Kelsey N. Thompson[2,3,7], George Weingart[2,3], Long H. Nguyen[2,3,4,5], Raaj S. Metha[2,3,4,5], Dong D. Wang[2,3,6], Wenjie Ma[2,3,4,5], Yan Yan[2,3], Meghan I. Short[2,3,7], Cesar Arze[3], Galeb Abu-Ali[3], Himel Mallick[3,7], Gholamali Rahnavard[3,7], Amit D. Joshi[4,5], Kerry L. Ivey[6], Jacques Izard[8], Wendy S. Garrett[2,7,10,11], Eric B. Rimm[2,6], Andrew T. Chan[2,4,5,7,11], Curtis Huttenhower[1,2,3,7]

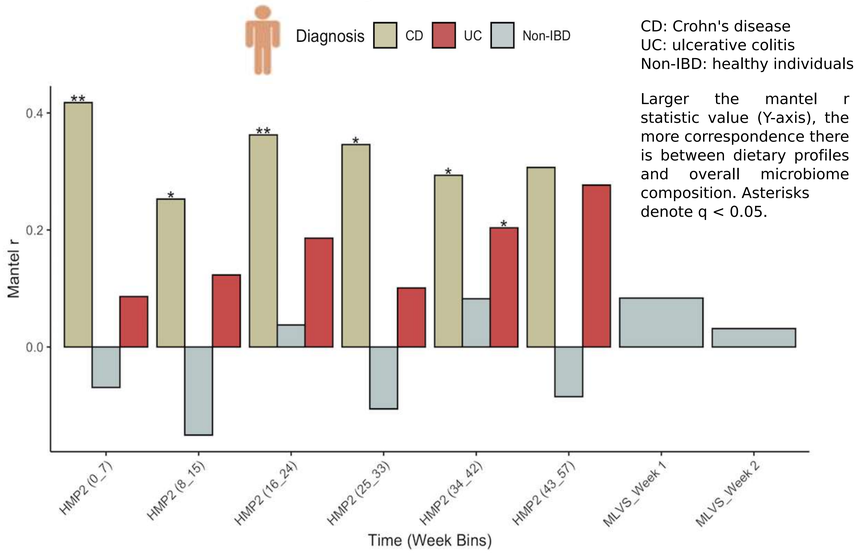
[1]Department of Systems Biology, Harvard Medical School, Boston, MA, [2]Harvard Chan Microbiome in Public Health Center, Boston, MA, [3]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, [4]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, [5]Clinical & Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, [6]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, [7]Broad Institute of MIT and Harvard, Cambridge, MA, [8]Department of Food Sciences & Technology, University of Nebraska, Lincoln, NE, [9]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, [10]Department of Medicine, Dana-Farber Cancer Institute, Boston, MA, [11]Department of Immunology and Infectious Disease, Harvard T.H. Chan School of Public Health, Boston, MA

Host dietary effects are far more context-dependent than is generally acknowledged. For instance, dietary changes drastically affect the microbiome in animal models, and in human infants during a shift to solid foods. In contrast, day-to-day dietary variation in healthy adults typically elicit only minor compositional shifts, although differences in pre-existing resident microbes among individuals can result in distinct chemical and metabolic responses to the same dietary intake. Here, we quantify these previously undifferentiated effects across different populations, life stages, environments, and dietary metadata. Seven publicly available metagenomic datasets spanning human adult, human infant, non-human primate, and mouse populations were reanalyzed through a standard bioinformatic workflow (total n=2,074 samples). The metagenomic profiles were accompanied by cohort-specific dietary information ranging from general diet types to resolved profiles of dietary components. To measure the effects of diet and the microbiome, we uniformly applied a set of models across studies. We assessed the relationship between overall dietary patterns or individual dietary compounds and microbial profiles, in addition to the specific interactions between dietary compounds and microbes. We found that, in a typical Western diet, the effect of day-to-day diet variation is small but significant, as expected. Instead, diet affects the microbiome indirectly via alterations in microbial transcription (but not, generally, organismal abundances), which in turn can have a mediating effect on host responses to diet. Applying the same models across populations, we also found that when the microbiome has not fully stabilized, as in human infants and laboratory animals, diet changes have a substantially larger effect on the gut community structure. Direct diet-driven variation thus depends largely on the resilience of the microbiome and on the extent of the dietary perturbation, while indirect interactions can be highly chemically and microbiologically specific.

Cohort microbiome sampling and meta'omics



Day-to-day diet variation has small effect on gut communities in healthy Western adults

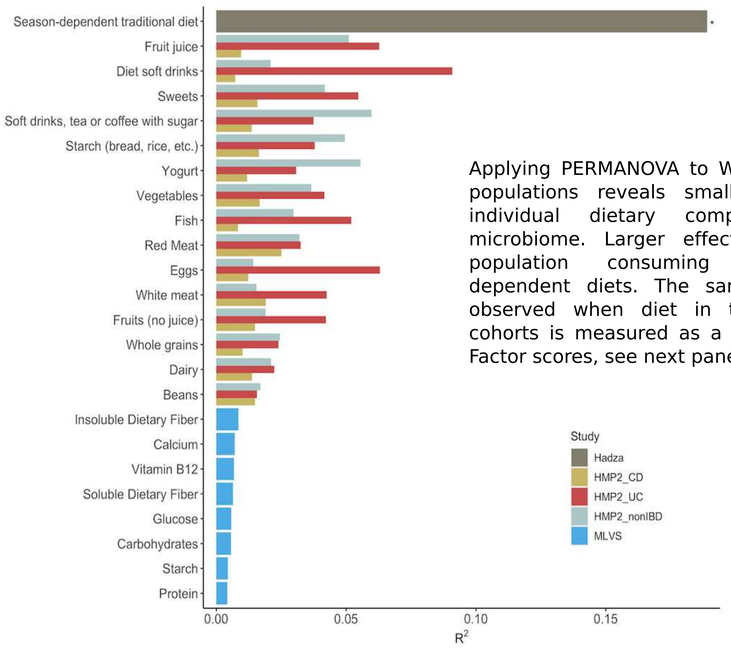


CD: Crohn's disease
UC: ulcerative colitis
Non-IBD: healthy individuals

Larger the mantel r statistic value (Y-axis), the more correspondence there is between dietary profiles and overall microbiome composition. Asterisks denote q < 0.05.

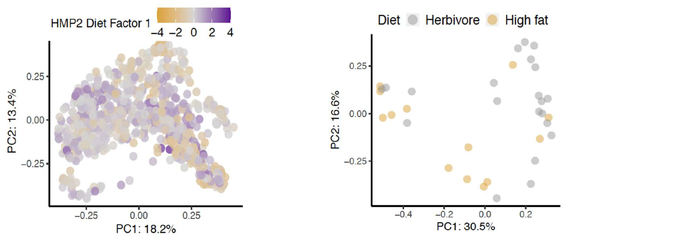
In healthy human adults, overall diet minimally agreed with the structure of the microbiome. In other words, the already perturbed microbiomes of individuals with disease seem to respond more to diet than the microbiomes of healthy individuals. HMP2: Human Microbiome Project II (Western diet), MLVS: Men's Lifestyle Validation Study cohort (Western diet)

Typical diet variation has small effect relative to consumption of seasonably variable diets



Applying PERMANOVA to Western adult populations reveals small effects of individual dietary components on microbiome. Larger effects seen for population consuming seasonally-dependent diets. The same trend is observed when diet in the Western cohorts is measured as a pattern (i.e., Factor scores, see next panel).

Diet effects on microbiome are context-dependent



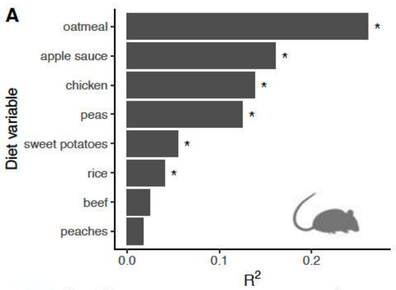
Relative to subtle differences in food profiles, when diet differences are large, we see greater differences in composition. Left: Human adult microbiome taxonomic profiles (Human Microbiome Project II), colored by dietary profile pattern scores. Right: Non-human primate microbiome taxonomic profies colored by diet type.

Diet variation drives gut microbiome variation in the developing gut microbiomes of human infants

Diet has larger effect in Western human infants compared to adults.

Differences in infant gut microbiomes observed across those soley breastfed, solely formul fed, or fed both breast milk and formula.

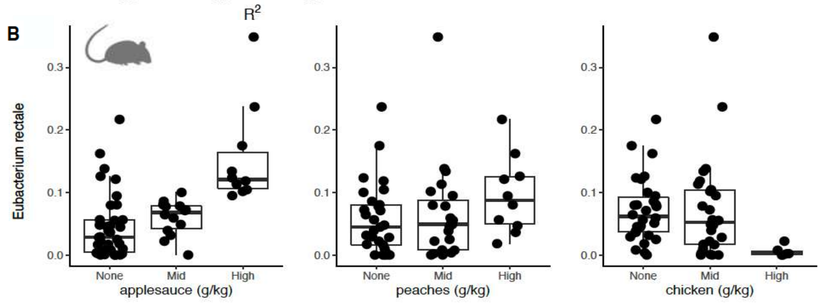
Drastic dietary changes alter gut community in non-human hosts

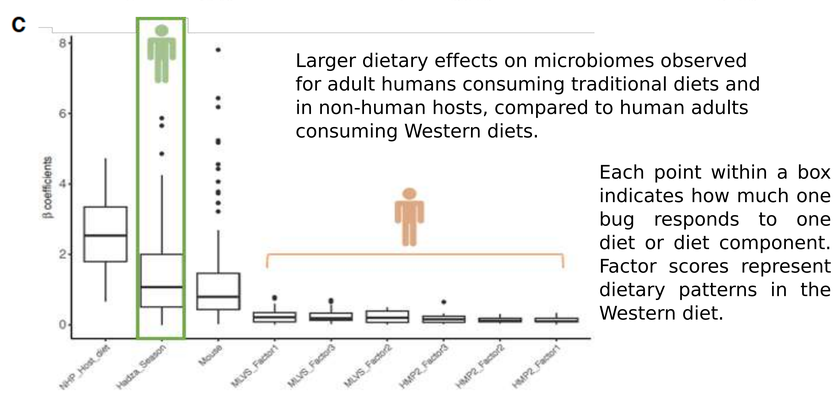


A) PERMANOVA model applied to individual dietary components fed to mice. Asterisks denote q < 0.05.

B) Diet~microbiome interactions where we see direct changes to the relative abundance of a specific taxon

C) Relative comparison of effect sizes measured for diet and taxa associations across host species





Larger dietary effects on microbiomes observed for adult humans consuming traditional diets and in non-human hosts, compared to human adults consuming Western diets.

Each point within a box indicates how much one bug responds to one diet or diet component. Factor scores represent dietary patterns in the Western diet.

Acknowledgements

This work has been supported by funding from STARR Cancer Consortium Award #17-A714, NIH grants P01DK046763, U01DK062413, U54DK102557, UL1TR001881, P30DK043351, R24DK110499, R01HG005969, U54DE023798, NSF grant DBI-1053486, and Army Research Office grant W911NF-11-1-0473.

http://huttenhower.sph.harvard.edu

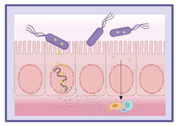
Human commensal bacteria produce novel genotoxic metabolites and exacerbate colorectal cancer

Yiyun Cao, *Joonseok Oh, Mengzhao Xue, Anjelica L. Martin, Deguang Song, Jason M. Crawford, Seth B. Herzon,* ***Noah W. Palm****

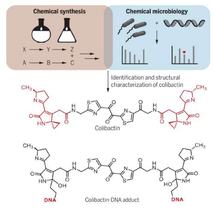
Department of Immunobiology, Yale University School of Medicine, New Haven, CT 06519, USA.

Introduction

Microbiota dysbiosis is associated with colorectal cancer (CRC). However, the gut commensals produce thousands of metabolites that mostly lack functional annotations or structural identification, so the causative mechanisms mediated by the gut microbiota through metabolites are still to be illustrated.



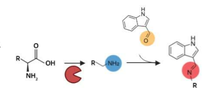
Escherichia coli strains produce the reactive small molecule genotoxin colibactin, which directly alkylates and crosslinks DNA, triggering double-stranded DNA breaks (DSBs) and facilitating intestinal tumorigenesis [1]. Human CRC tumors also contain mutational signatures consistent with colibactin-induced DNA damage [2]. However, aside from colibactin, the role of microbiota-derived small molecule genotoxins in CRC initiation or progression remains mostly unexplored.



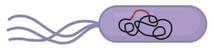
We hypothesized that ***diverse taxa from the human gut microbiome may produce previously undiscovered small molecules that cause DNA damage in intestinal epithelial cells and contribute to the development of CRC***.

Key findings

1) Established a forward-screening pipeline to evaluate the genotoxicity of microbial small molecule metabolites causing DNA damage.

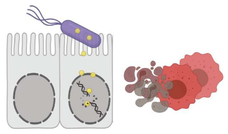


2) Discovered a previously undescribed family of genotoxic metabolites—**indolimines**—produced by the CRC-associated species ***Morganella morganii***.



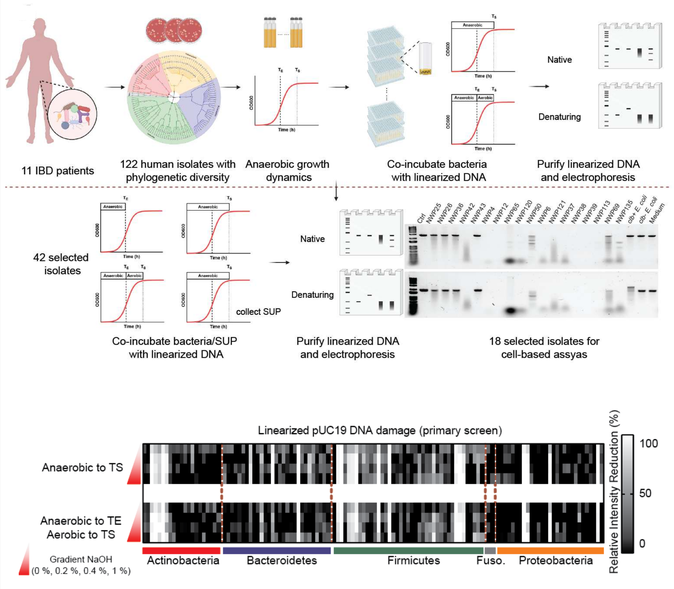
3) Identified the gene ***aat*** **(encoding AAT_I protein, aspartate aminotransferase fold type I)** necessary for the synthesis of indolimines and genotoxicity of *M. morganii*.

4) Evaluated the tumorigenesis effect of indolimine-producing *M. morganii* with AOM/DSS CRC model in gnotobiotic mice.

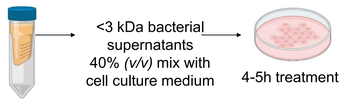


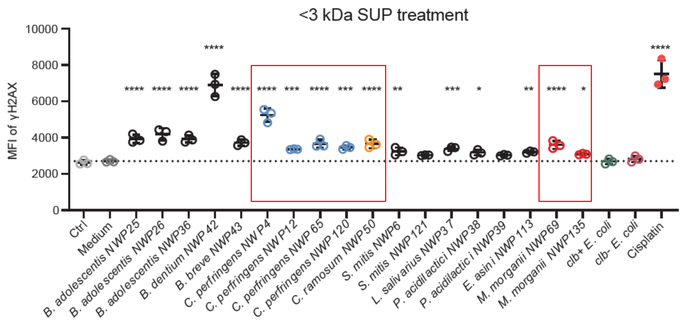
Screening pipeline

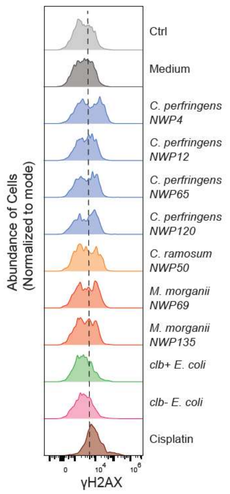
Figure 1. DNA damaging activity was widespread in diverse taxa of the human gut microbiota



Genotoxic species







Human Microbiome Project
(~1700 samples)

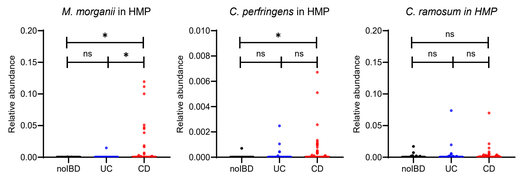
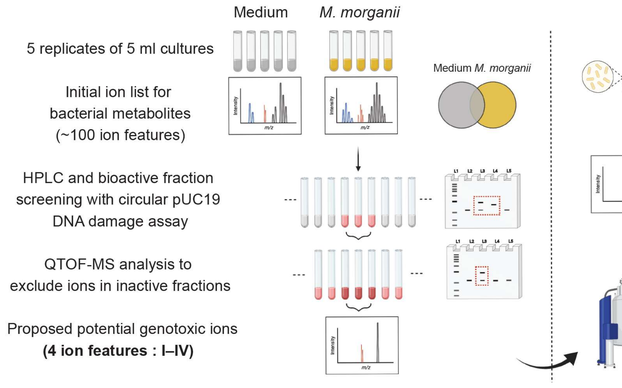


Figure 2. Small molecule metabolites produced by human gut microbes induced DNA damage.

Genotoxin(s)

QTOF-MS-bsased untargeted metabolomics and bioactivity-guided fractionation

Large-scale cultivation for chemical isolation and identification



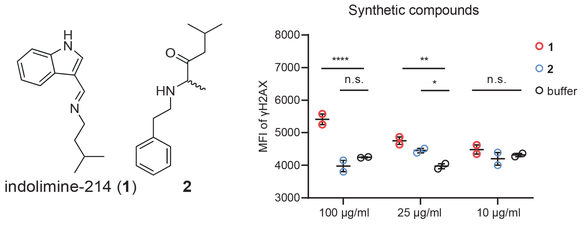
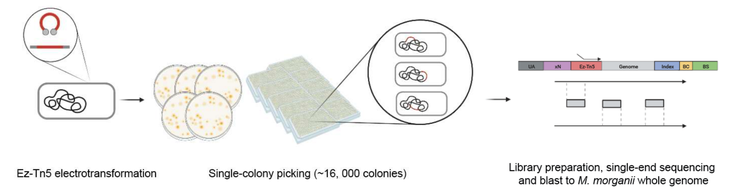
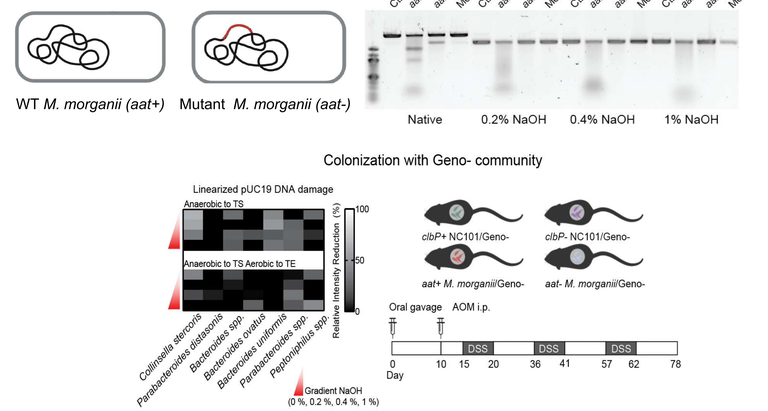


Figure 3. Isolation and identification of a previously undescribed genotoxic metabolite derived from *M. morganii*.

AAT for genotoxicity





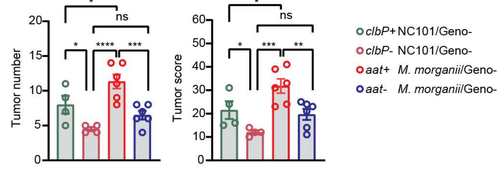
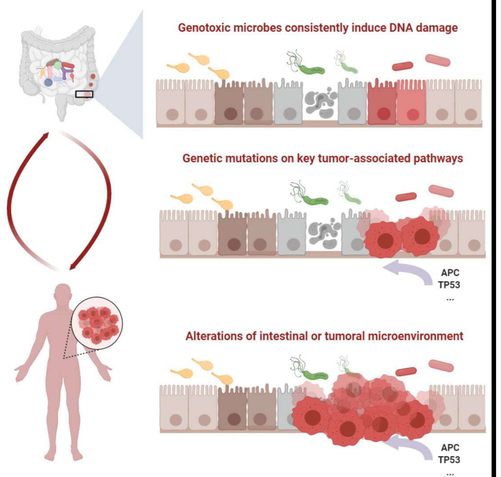


Figure 4. *aat* (aspartate aminotransferase fold type I) gene regulated indolimine synthesis and genotoxicity of *M. morganii* both in vitro and in vivo.

Model



References

[1] Xue, M., et al., *Structure elucidation of colibactin and its DNA cross-links.* Science, 2019. **365**(6457).

[2] Pleguezuelos-Manzano, C., et al., *Mutational signature in colorectal cancer caused by genotoxic pks(+) E. coli.* Nature, 2020.

[3] Anzai, I.A., et al., *Rapid curation of gene disruption collections using Knockout Sudoku.* Nat Protoc, 2017. **12**(10): p. 2110-2137.

Yale Cancer Center Team Challenge award; Common Fund of the National Institutes of Health (DP2DK125119); Leona M. and Henry B. Helmsley Charitable Trust (3083); Chan Zuckerberg Initiative; Michael J. Fox Foundation for Parkinson's Research; Ludwig Family; Mathers Foundation; Pew Charitable Trust; NIA and NIGMS (R01AG068863 and RM1GM141649); F. Hoffmann-La Roche Ltd.

Characterizing bacterial and fungal communities in childcare microbiome: a study overview

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

THE HARVARD CHAN MICROBIOME IN PUBLIC HEALTH CENTER

[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 6Pacific Biosciences [7]Massachussetts General Hospital
[8]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health

Indoor environments harbor diverse microbes to which occupants are constantly exposed. Exposure to environmental microbes has been shown to have both negative and positive healthimpacts, particularly relating to children's immune maturation during early development. Many preschool-aged children spend 7 to 10 hours per day in childcares, almost as much as at home, yet the environmental microbes associated with childcares have yet to be fully elucidated. Although several studies have explored the taxonomic composition of childcare microbiomes using DNA sequencing, this method on its own suffers from an inability to discern viability, which hinders the interpretation of health implications. Additionally, exposure to fungi can lead to adverse health effects, but indoor fungal communities are much less studied as compared to bacterial communities and the extent to which fungal communities affect children in childcares remains unclear. In this study, first, we will use paired metagenomics and metatranscriptomics to characterize viable bacterial communities, including functional molecular mechanisms responsible for microbial persistence and antimicrobial resistance burden, in childcare environments. Second, we will target the full-length internal transcribed spacer (ITS) region and curate a full-length ITS reference database for identifying indoor fungal communities, to overcome the drawbacks associated with traditional short-read sequencing, which usually results in lower taxonomic resolution, higher proportion of unidentified taxa and greater extent of underclustering.

Study design and sample collection

Study Participants

RCCC

N = 50
Toddlers/Preschoolers
Age 3-5

SFPCC

Exclude if:
- Used oral or intravenous ABX or chemotherapy within 9 months
- Self report acute infectious disease (cold/flu, 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)

Sample and Data Collection

Swab samples

Environment	# samples	Source
Doorknob	8	2 per classroom x 2 rooms x 2 centers
Desk	4	1 per classroom
Chair	8	2 per classroom
Toy	8	2 per classroom
Drain	4	1 per classroom
Sink	5	1 per classroom

Dust samples

Carpet	4	1 per classroom
Air	4	2 per center
HVAC	4	1 per classroom
Infiltration	8	2 per classroom
Soil	8	4 different locations per center

Oral swabs Nasal swabs

Questionnaires

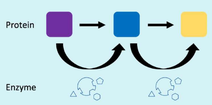
- Age
- Sex
- Race
- Ethnicity
- Breastfeeding
- Delivery mode
- Siblings
- Pets
- ABX use
- Diet (probiotics)
- Respiratory illness
- Oral health

Analysis

bioBakery v3.0

Taxonomic profiling

Functional profiling



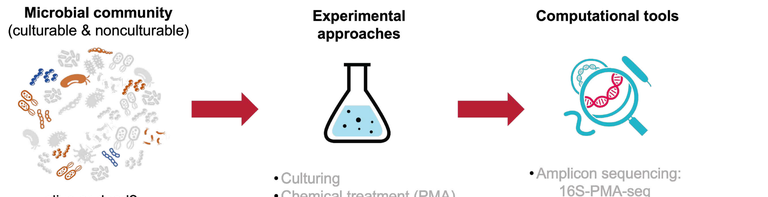
Strain profiling

Statistics: PERMANOVA (community composition), metatranscriptomics-based differential expression analysis, omnibus and per-feature test (MaAsLin 3)

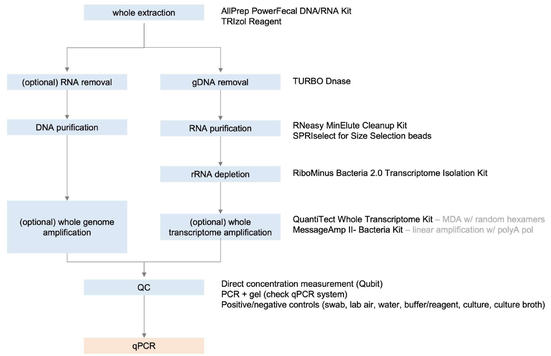
RCCC: Radcliffe Child Care Center
SFPCC: Soldiers Field Park Children's Cente
ABX: antibiotics

Metagenomes and metatranscriptomes for viability assessment

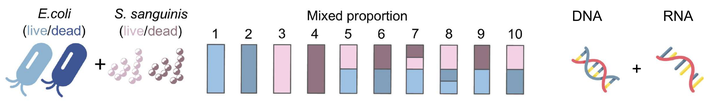
In anthropogenic, biochemically unnatural settings such as the built environment (BE), viable microbes are often outnumbered by their "dead" counterparts. Critically, the functions of a microbial community are defined by these viable microbes, and despite their low abundances, these viable microbes can still pose health risks. Previous evaluations of "PMA-seq" (PMA treatment + 16S rRNA seq) and "16S-RNA-seq" (16S rRNA transcript-based amplicon sequencing) proved to function well in low-complexity communities, but poorly quantified viable microbes in realistically complex community samples.



Protocol evaluation and validation



We first evaluate in ten synthetic communities comprising live and/or heat-killed *E.coli* and *S.sanguinis* mixed at different ratios.
Total DNA and RNA are extracted simultaneously, with RNA samples immediately reverse transcribed into cDNA followed by shotgun metagenomic sequencing for all cDNA and their respective DNA samples in parallel.



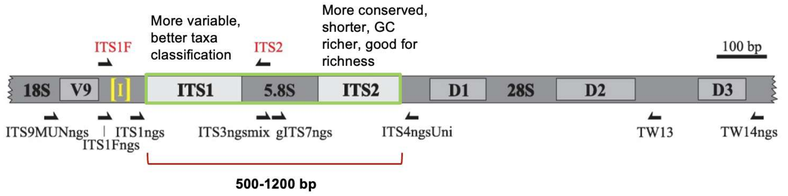
To evaluate its performance under variable biomass, biochemical background and diversity conditions, we next apply the protocol to swabbings of natural microbial communities from computer screens and keyboards, soil, and human saliva spiked with known concentrations of living and heat-killed *E. coli* controls.

Environmental samples from childcare centers

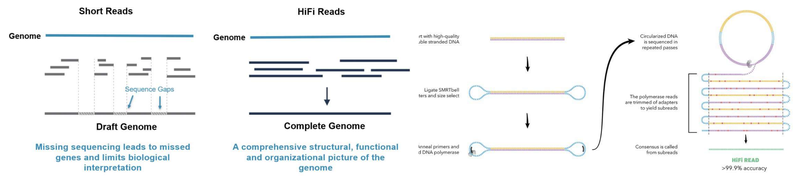
We further validate this protocol in built environment microbiome samples collected from childcare centers. In total, we will have a total of 250 samples comprising of those from synthetic cultures, *E. coli* spike-in experiments, childcare environmental microbiome samples and several experimental and technical controls for shotgun sequencing.

qPCR is performed targeting 16S V4 region and species-specific protein-coding genes on synthetic communities to determine the viable (mRNA/cDNA) and overall (DNA) bacterial mass.

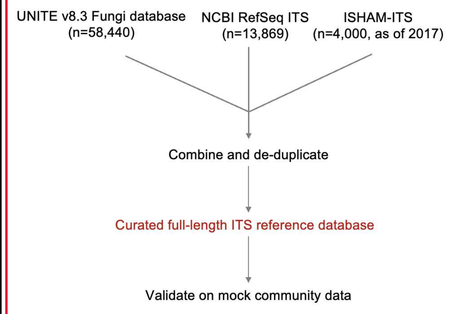
Full-length ITS amplicon-based fungal identification



The full fungal ITS region ranges from 500-1200 bp, which cannot be sequenced as a single DNA fragment using short-read sequencing (typically 50-500 bp fragment size). Because of this, fungal community studies mostly use only ITS subregions, either ITS1 or ITS2. There have been controversies in the selection of markers, and using either alone can lead to sequence gaps and thus provide limited taxonomic and phylogenetic information for species discrimination and taxonomic assignment.



Here we use third-generation circular consensus sequencing from Pacific Biosciences to amplify the full-length ITS marker gene and several flanking rRNA subunit regions simultaneously. This enables to focus on variable regions for species-level identification and conserved regions for phylogeny and higher level identification. The primer pair determined to be best suitable for pan-eukaryote identification is ITS9MUNngs+ITS4ngsUni, because of the avoidance of an intron site just upstream, terminal mismatch to all bacteria and coverage of the SSU V9 variable region.



Two major challenges remain in downstream community analyses:
- Scarcity of automated pipelines to analyze long amplicon data. Most software developed for sequencing-error correction and sequence assembly have been optimised for short reads and have not been evaluated for long reads in community analyses;
- Difficulty in fungal identification largely because of the incompleteness of reference databases. These reference databases play a critical role in the sequence-based species identification, cultured strains and fungal communities from environmental samples.

Several databases, including UNITE, NCBI ITS RefSeq, and ISHAM-ITS, contain high-quality ITS sequences generated from reliably identified and preserved specimens. We combine and de-duplicate into our curated full-length ITS reference database and validate it on mock community data consisting of 10-20 mixed fungi and bacteria.

Acknowledgments

We appreciate the support and assistace 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, with this research, particularly for involving in the recruitment and sample collection. This work has been supproted by the Harvard Chan-NIEHS Center for Environmental Health (NIH; #P30ES000002).

http://huttenhower.sph.harvard.edu

Developing omics pipelines to wrangle strain diversity in staphylococci contextualizes skin microbiomes.

INTRODUCTION

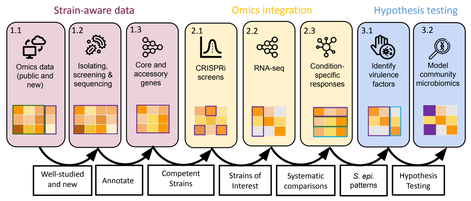
Of the trillions of microbial cells associated with a human body at any given point in time, about half are identifiable at the genus level, a quarter are culturable, and only a handful have been isolated and extensively studied in the laboratory over decades. Analogous to how model organisms, such as mice, have been used to study human biology, “transfers” of the great depth of knowledge accumulated for model microbial species, such as the pathogen *Staphylococcus aureus*, to related but less-studied microbes such as the commensal and opportunistic pathogen *Staphylococcus epidermidis*, will greatly facilitate understanding microbial diversity and microbial communities.

To date, transfer learning has been successful in the fields of image, video and natural language processing and has been applied in genomics to bridge different mammalian cell types and is starting to be used to connect different species, including plant and insect model and non-model organisms. The genetic diversity and transcriptional plasticity of *S. epidermidis* is sparsely annotated, and we aim to apply transfer learning to genetic, transcriptomic and functional data to integrate the wealth of *S. aureus* data collected over previous decades and drive hypothesis generation around *S. epidermidis* pathogenicity.

METHODS

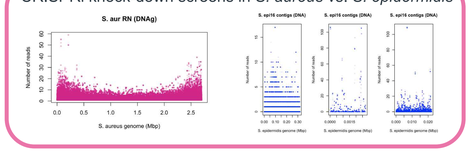
Computational pipelines that analyze large datasets from well-studied strains and new data from diverse strains sets the stage for cross-strain comparisons and transfer learning approaches that leverage similarities and identify differences.

Genome-wide CRISPR-interference knock-down screens and tRNA-seq transcriptomic profiling, along with annotated reference genomes, promote hypothesis generation in interpretable pipelines.

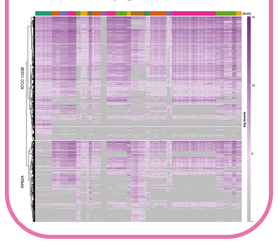


EXAMPLES

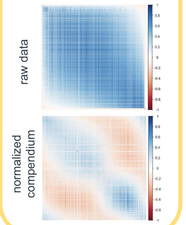
CRISPRi knock-down screens in *S. aureus* vs. *S. epidermidis*



Computationally-efficient mapping of RNA-seq data to many genomes



Exploring gene-gene correlation networks



RNA-seq compendium preprint

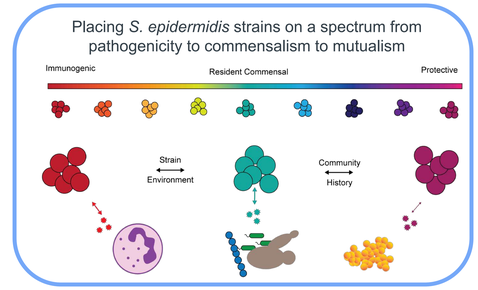
The Jackson Laboratory

Leading the search for tomorrow’s cures

SIGNIFICANCE

- Different strains of *S. epidermidis* have different immunomodulatory activities and microbial interactions.
- Different strains of *S. epidermidis* respond to environmental cues and stressors differently.
- Mapping out strain by environmental effects requires high throughput tools that can be applied to clinical and environmental isolates.
- Spoto et al CRISPRi & RNA-seq preprint describe conditional essentiality and expression in strain Tu3298.

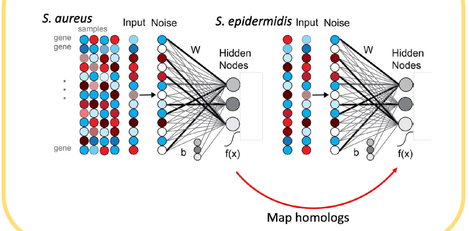
Placing *S. epidermidis* strains on a spectrum from pathogenicity to commensalism to mutualism



FUTURE DIRECTIONS

- Gene essentiality and expression profiling of *S. epidermidis* strains when grown in skin model microbial communities.
- Interpretable deep learning approaches to compare data from diverse *S. epidermidis* strains to the large amounts of public data for *S. aureus* and other skin pathogens and commensals.
- Paired genomics, metabolomics and transcriptomics to determine functional consequences of strain diversity.
- Characterization of essential and conditionally essential genes in diverse strains of *S. epidermidis.*
- Identifying similarities and differences between isolates and well-studied strains can help place individual isolates on the commensal-pathogen spectrum, isolating defining factors.
- Transfer learning to aid in the generation of new hypotheses can be applied to genomic and transcriptomic data via homolog mapping.

Transfer learning from well-studied to diverse strains



Georgia Doing, Johanna Riera, Michelle Spoto, Julia Oh
The Jackson Laboratory for Genomic Medicine
Farmington, CT 06032 USA

The Oh Lab

SYNTROPHIC INTERACTIONS AMONG DIFFERENT MEMBERS OF THE HUMAN GUT MICROBIOTA IN THE METABOLISM OF β-GLUCAN FROM *FUSARIUM VENENATUM*

Pedro Jesús Fernández Juliá, Jose Luis Muñoz Muñoz

Department of Applied Sciences, Northumbria University, Newcastle Upon Tyne NE1 8ST, Tyne & Wear, England, United Kingdom

BACKGROUND

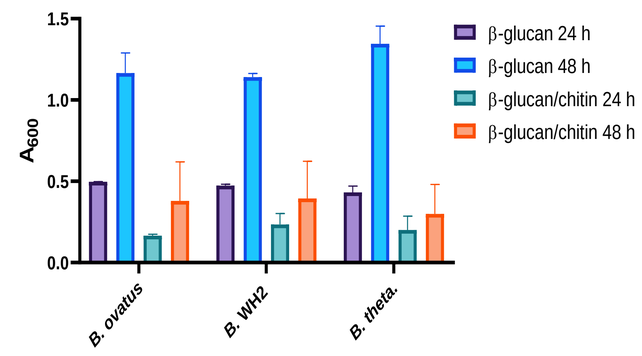
β-glucans are polysaccharides which have been described as potential prebiotics[1]. The mechanism of action underpinning these health effects related to β-glucans are still nuclear. *Bacteroides* spp. are described as glycan degraders capable of using a wide range of substrates[2] whereas other bacteria such as *Bifidobacterium* spp. commonly metabolize smaller glycans[3], in particular oligosaccharides, sometimes through syntrophic interactions with *Bacteroides* spp.

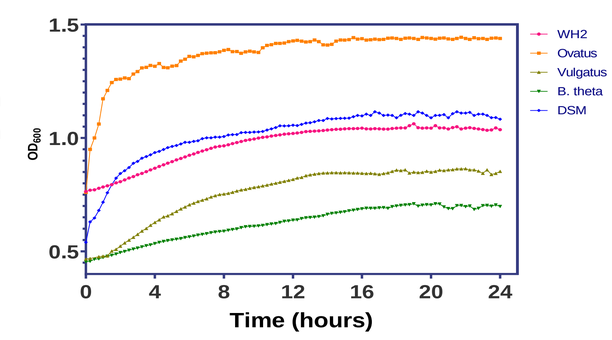
PROJECT AIMS

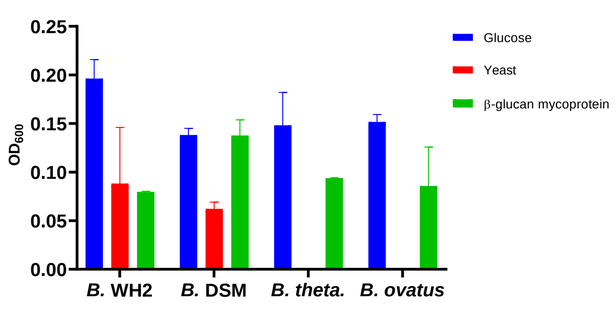
This project is going to explore how human gut *Bacteroides* spp. and *Bifidobacterium* spp. can use β-glucan from the microfungus *Fusarium venenatum,* which is used to elaborate mycoprotein of QUORN® products. The study is focused on the search of the metabolic pathways for the β-glucan degradation together with the presumable cross-feeding relations established between different members of the Human Gut Microbiota (HGM) involved in the process.

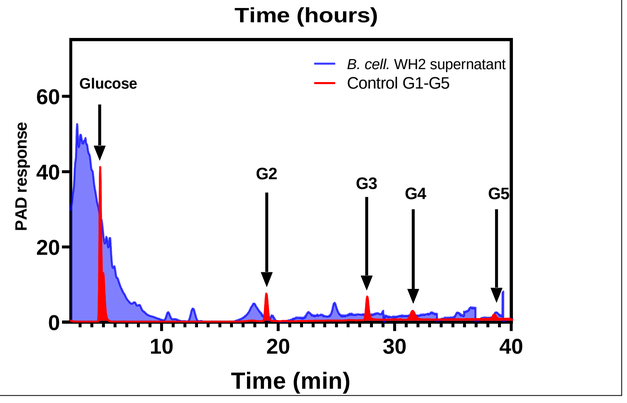
UTILIZATION OF β-GLUCAN BY *BACTEROIDES*

The utilization of β-glucan as carbon source is present in different *Bacteroides* spp. Data shows the β-glucan from mycoprotein leads to higher growth rates than yeast β-glucan and similar values compared with glucose. Moreover, the utilization of fungal β-glucan is spread within *Bacteroides* spp., but the growth rates are different between them, probably due to the presence of different Polysaccharide Utilization Loci (PULs).









UP-REGULATION OF THE GENES INVOLVED

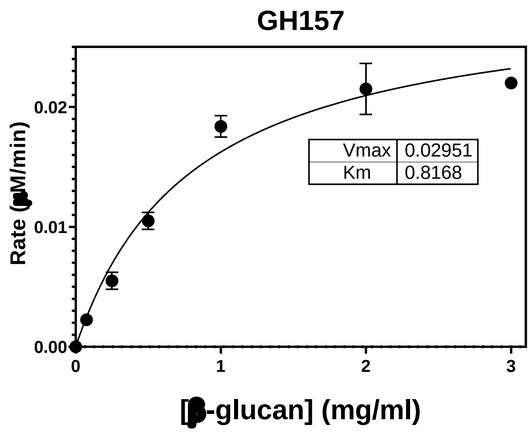
The metabolism of β-glucan of fungi is related to the up-regulation of different genes in *Bacteroides* spp. and, therefore, the overexpression of the encoded proteins in those genes. The degradation of β-glucan involves different PULs which are composed of a wide range of proteins such as SGBPs, glycoside hydrolases, transporters, etc.

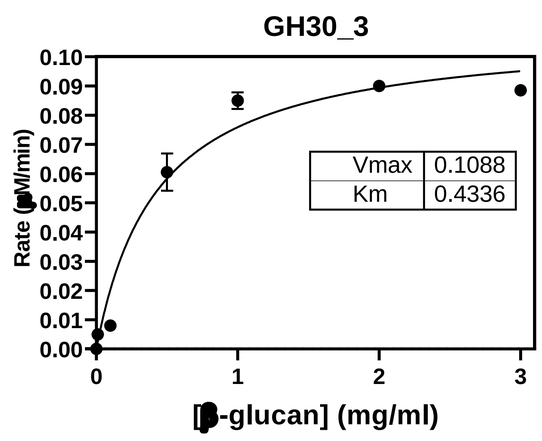
Locus tag	Family	Abundance
BcellWH2_02126	GH3	54
BcellWH2_02127	GH3	41
BcellWH2_02128	SusD	55
BcellWH2_02129	SusC	57
BcellWH2_02131	GH157	89
BcellWH2_02132	HTCS	41
Locus tag	**Family**	**Abundance**
BcellWH2_01837	GH30_3	100
BcellWH2_01838	Unknown	89
BcellWH2_01839	SusC	80
BcellWH2_01840	SusD	65
BcellWH2_01841	GH2/CBM57	40
BcellWH2_01842	PL38/GH88	2

PUL24		
BVU_1039	HTCS	23
BVU_1040	Unknown	25
BVU_1041	SusC	80
BVU_1042	SusD	91
BVU_1043	GH30_4	67
BVU_01044	GH30_4	59
PUL33		
BVU_1250	HTCS	31
BVU_1251	GH2	48
BVU_1252	SusC	88
BVU_1253	SusD	102

KINETICS ASSAYS OF THE ENZYMES INVOLVED

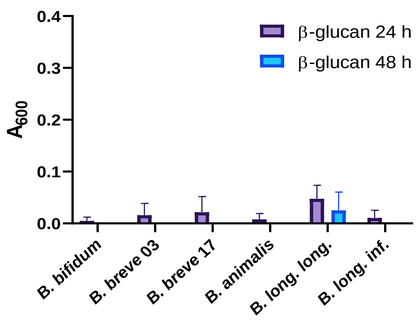
The kinetic assays reveal how different enzymes take part in the metabolization of β-glucan obtained from mycoprotein. The analysis of the kinetic parameters shows how this enzymes possess different affinities and specificities for the substrate, depending on its catalytic role.

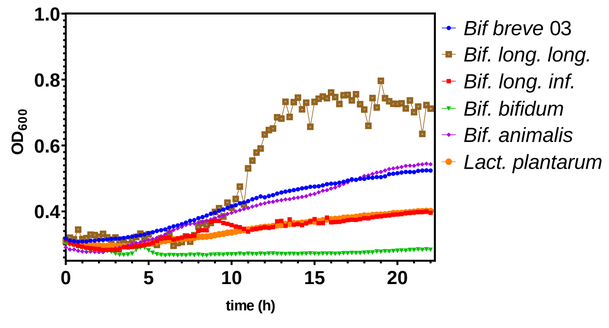


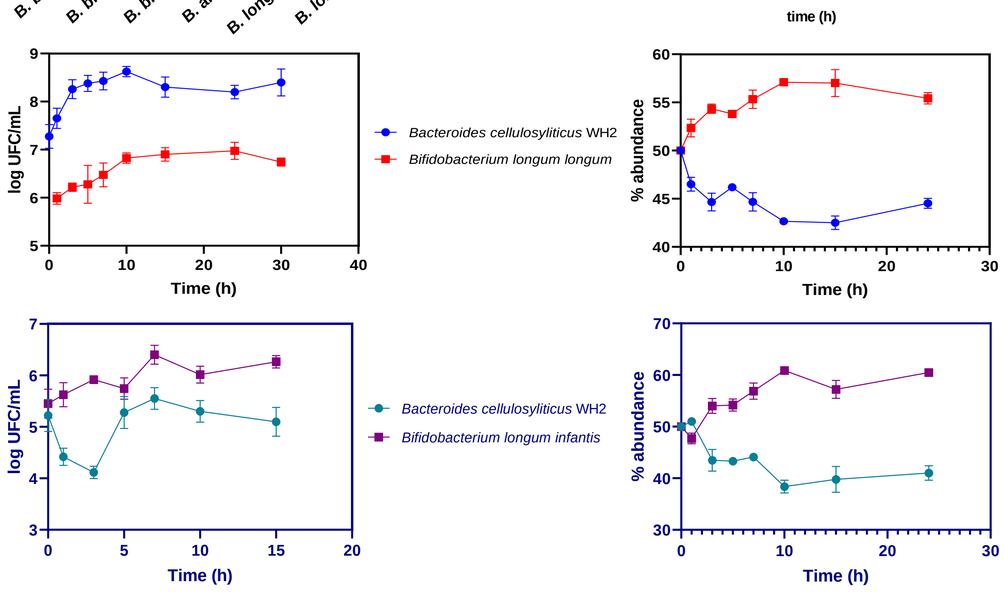


CO-CULTURES OF *BACTEROIDES* AND *BIFIDOBACTERIUM*

Bifidobacterium* spp. show no ability to grow with the crude polysaccharide but they show this ability when the supernatants obtained in *Bacteroides* spp. cultures are use as substrate. Moreover, coculture experiments using both *Bacteroides spp.* and *Bifidobacterium spp.* together reflect the crossfeeding interaction between them. Because of this feeding connection among bacteria, the secondary degrader *Bifidobacterium spp.* is able to persist in the media taking advantage of the metabolites released by the primary degrader *Bacteroides spp.







CONCLUSION

1. **Fungals β-glucan obtained from mycoprotein is metabolized for several species within *Bacteroides* genus.**
2. **The degradation of mycoprotein by *Bacteroides spp.* allows the growth of benefitial bacteria such as *Bifidobacterium spp.* and *Lactobacillus spp.* because of the production of oligosaccharides and SCFAs.**

FUTURE PERSPECTIVES

1. **Identify the structure and abundance of oligosaccharides and SCFAs produced as metabolites during β-glucan degradation.**
2. **Amplify the croosfeeding experiments to other bacteria species present in the Human Gut Microbiota.**

REFERENCES

[1] Zhu, F., et al., A critical review on production and industrial applications of betaglucans, *Food Hydrocoll.* 52 (2016) 275-288. doi: 10.1016/j.foodhyd.2015.07.003. [2] Grondin, J., et al., Polysaccharide Utilization Loci: Fueling Microbial Communities, *J. Bacteriol.* 199(15):e00860-16. doi: 10.1128/JB.00860-16 (2017). [3] Seth, E., et al., Nutrient cross-feeding in the microbial world, *Front. Microbiol.* 5:350. doi: 10.3389/fmicb.2014.00350 (2014).

Northumbria University NEWCASTLE

Quorn™ Healthy Protein. Healthy Planet.

Statistical approaches for differential expression analysis in metatranscriptomics

Yancong Zhang[1,2,3], Kelsey N. Thompson[1,2,3], Curtis Huttenhower[1,2,3,4], Eric A. Franzosa[1,2,3]

[1]Harvard Chan Microbiome in Public Health Center; [2]Dept. of Biostatistics, Harvard T. H. Chan School of Public Health
[3]Broad Institute of MIT and Harvard, [4]Dept. of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health

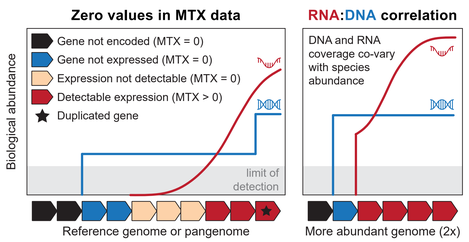
THE HARVARD CHAN MICROBIOME IN PUBLIC HEALTH CENTER

BROAD INSTITUTE

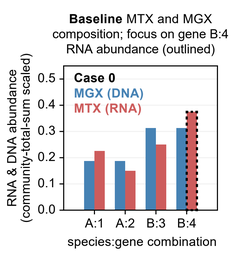
DE analysis challenges in MTX data

Metatranscriptomics (MTX) has become an increasingly practical way to profile the functional activity of microbial communities *in situ*. However, MTX remains underutilized due to experimental and computational challenges. The latter include non-independent changes in RNA transcript levels and their underlying genomic DNA copies, genetic plasticity, measurement compositionality, and zero-inflation.

Here, we present a systematic evaluation of and recommendations for differential expression (DE) analysis in MTX in the presence or absence of paired metagenomics (MGX) data.

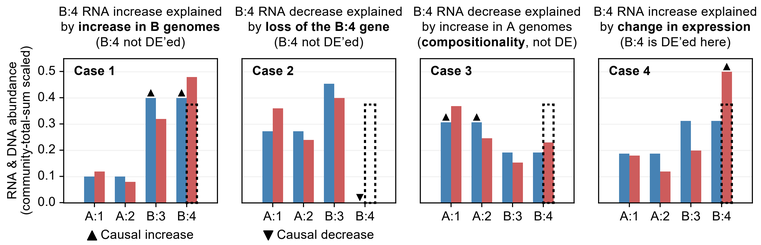


Gene copy number variation induces spurious DE in MTX



Left: *The reference state of a simple community (two species, A and B, each contributing two genes) profiled by MTX and MGX.*

Below: *Cases 1-4 correspond to perturbations of the community. While the RNA abundance of gene B:4 (dashed outline) differs under each perturbation, only in Case 4 is the difference attributable to DE rather than gene copy number variation.*



We evaluated six statistical models for DE discovery in MTX

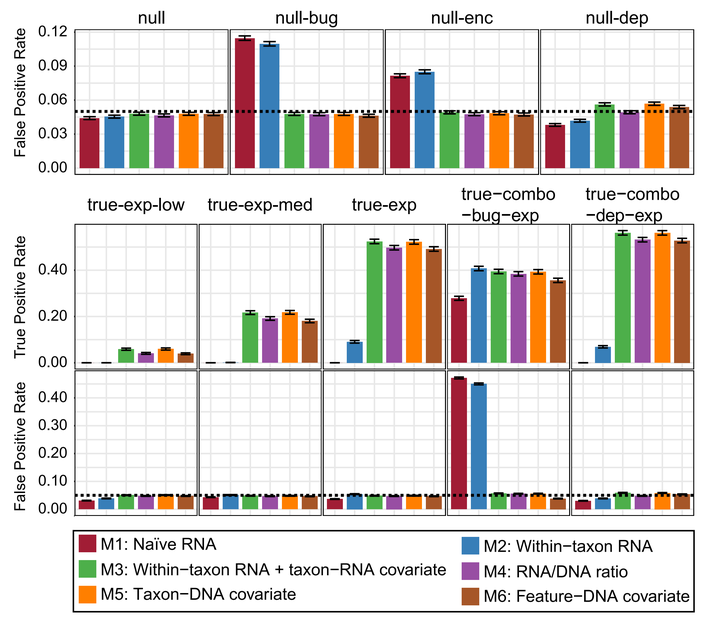
Models incorporate different combinations of DNA and RNA normalization and assumptions about the underlying changes of gene copies or species abundance within communities.

	Model number: name	Model formulation	Model description
	M1: Naïve RNA	$C(f_{RNA}) \sim p$	The community total-sum-scaled abundance $C(x)$ of a feature f's RNA count (f_{RNA}) is modeled as a function of a sample phenotype/property (p).
×	M2: Within-taxon RNA	$T(f_{RNA}) \sim p$	f_{RNA} is subjected to within-taxon sum-scaling, $T(x)$, i.e. normalizing against the total RNA pool of f's source taxon.
×	M3: Taxon-RNA covariate	$T(f_{RNA}) \sim Tax_{RNA}(f) + p$	f_{RNA} is scaled within-taxon as in M2; an RNA-level estimate of f's source taxon abundance, $Tax_{RNA}(f)$, is included as a covariate.
+	M4: RNA/DNA ratio	$C(f_{RNA})/C(f_{DNA}) \sim p$	The ratio of f's total-sum scaled RNA and DNA abundances (the "relative expression ratio") is modeled as a function of p.
✱	M5: Taxon-DNA covariate	$C(f_{RNA}) \sim Tax_{DNA}(f) + p$	f_{RNA} is total-sum scaled as in M1; a DNA-level estimate of f's source taxon abundance, $Tax_{DNA}(f)$, is included as a covariate.
+	M6: Feature-DNA covariate	$C(f_{RNA}) \sim C(f_{DNA}) + p$	f_{RNA} is total-sum scaled as in M1; a total-sum scaled estimate of f's DNA-level abundance is included as a covariate.

× = *The model requires a mapping of features to taxa* + = *the model requires paired MGX (DNA) data* ✱ = *both are required*

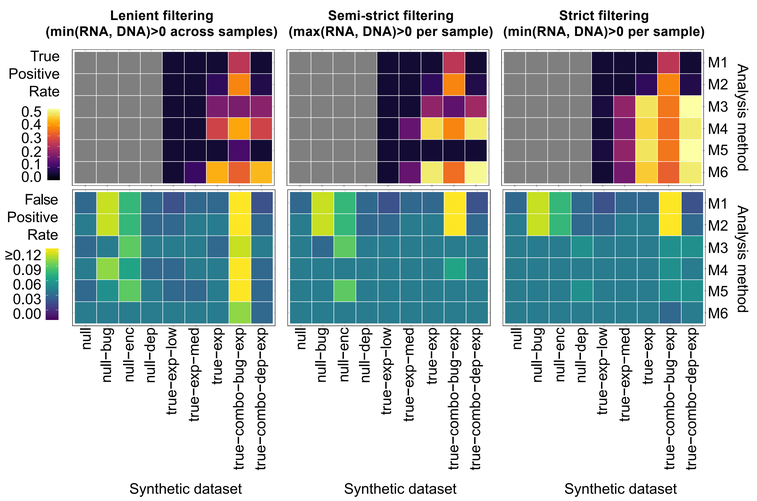
Controlling for gene copies is critical

We evaluated models M1-M6 across a range synthetic datasets (see right column). M1 and M2, which fail to control for gene copy number (GCN), were prone to mistaking changes in species and gene abundance for DE. M3 achieved high specificity using species-total RNA as a proxy for GCN, though models using DNA-based estimates of GCN (M4 and M6) were slightly better. M3-M6 were similarly sensitive to positive DE signals.



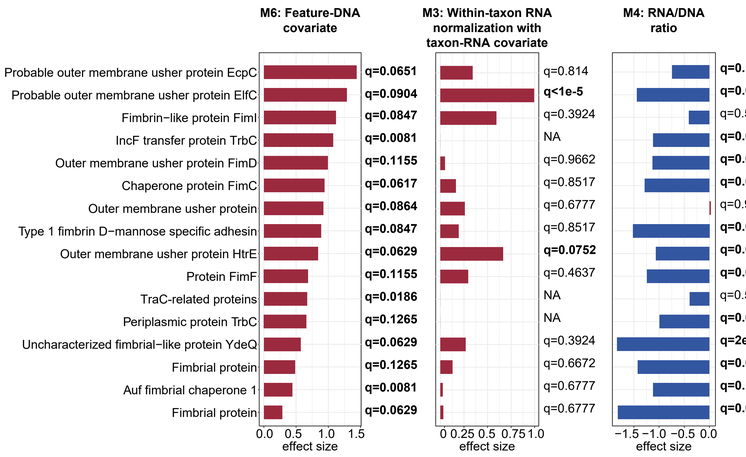
Paired MGX data add robustness to RNA zero values

The above analyses employed strict filtering of RNA zero values. Under less-strict filtering, models lacking DNA-based control of GCN were prone to mistaking gene absence/deletion for down-regulation. The gene-DNA covariate model (M6) was especially robust to RNA zero values.



MTX finds *E. coli* pilins DE'ed in IBD

We tested 113 *E. coli* pilin-family proteins for DE in IBD-associated dysbiosis using three well-performing models (M3, M4, and M6). M6 identified 16 genes with FDR-significant elevated expression in dysbiotic IBD samples. The significance and sign of these trends differed in M3 and M4 (respectively), possibly due to those models' greater sensitivity to real-world gene and genome copy-number variation.



Description of synthetic datasets and guide to abbreviations

We simulated synthetic MTX and MGX data for model evaluation based on the human gut microbiome. Species abundances were modeled as zero-inflated log-normal (LN) distributions. Strain-specific genes were sampled randomly from simulated pangenomes. Gene expression values were modeled as LN distributions within strains and across samples.

Species, gene, and transcript features were permuted to associate with simulated phenotypes.

	null	null-bug	null-enc	null-dep	true-exp	true-exp-med	true-exp-low	true-combo-bug-exp	true-combo-dep-exp	group-null-enc	group-true-exp
% of transcripts associated (DE'ed) with phenotype	–	–	–	–	10%	10%	10%	10%	10%	–	10%
Strength of DE relationships	–	–	–	–	High	Med	Low	High	High	–	High
% of taxa strongly confounded with phenotype	–	10%	–	–	–	–	–	50%	–	–	–
% of gene gain/loss strongly confounded with phenotype	–	–	10%	–	–	–	–	–	–	10%	–
Sequencing depth confounded with phenotype	–	–	–	Yes	–	–	–	–	Yes	–	–
Gene taxonomy is unknown; trends at orthogroup level	–	–	–	–	–	–	–	–	–	Yes	Yes

← 11 synthetic MTX + MGX datasets →

Acknowledgments

The authors gratefully acknowledge the use of the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University. This work was supported by a research grant from Takeda Pharmaceuticals and by NIH NIDDK grant R24DK110499.

@hutlab

Learn more (and download data & code) via:

`http://huttenhower.sph.harvard.edu/mtx2021`

Discover bioBakery software and tutorials via:

`http://huttenhower.sph.harvard.edu/biobakery`

Gut microbiome diversity and biochemical traits associated with metabolic diseases in urban and semiurban adults in Mexico City, 2021

Miguel A. Garcia-Salcido, Eira E. Huerta-Avila, Angélica Martínez-Hernández,
Francisco M. Barajas-Olmos, Humberto García-Ortiz, Lorena Orozco.
Immunogenomics and Metabolic Diseases Laboratory, Instituto Nacional de Medicina Genómica, Secretaría de Salud, México City, México

Background. One of the main determinants of gut microbiome (GM) is diet, which is also determined by the physical, economic, sociocultural and political context (food enviroment) (2). Replacement of traditional food enviroment by Western industrialized lifestyles, characterized by high consumption of saturated fat, animal protein, and food additives, as well as low fiber and vegetable protein intake, leads to changes in the GM diversity and composition, which could be related with an increased prevalence of metabolic diseases (3). In Mexico there is a great diversity of food enviroment (urban, semiurban and traditional), however, there are a few studies about the relationship between these with GM diversity and composition and metabolic diseases.

Aim. To determine the relationship between food environments with the GM diversity and biochemical and anthropometric measures associated with metabolic diseases in the Mexican population.

Methods. We conducted an observational, descriptive, cross-sectional study, based on the GM analysis of metagenomic data. In a population composed of 42 participants from urban and semiurban environments in Mexico City.

Figure 1. Gut microbiome α- diversity indices by food environment, boxplot

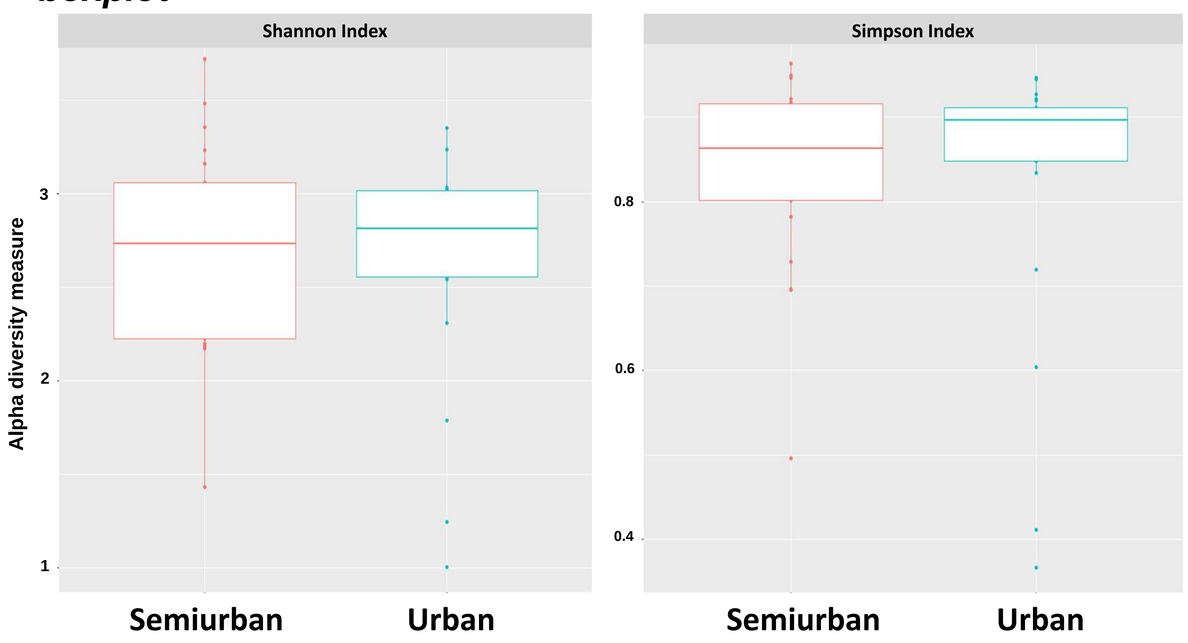


Figure 2. Gut microbiome β- diversity by food environment, principal coordinates analysis

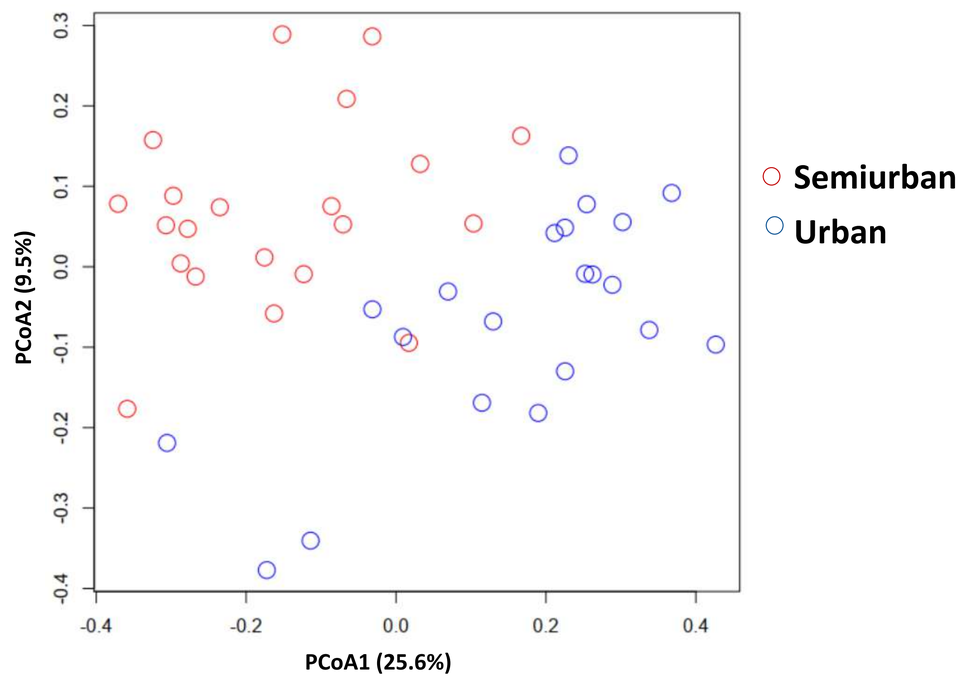
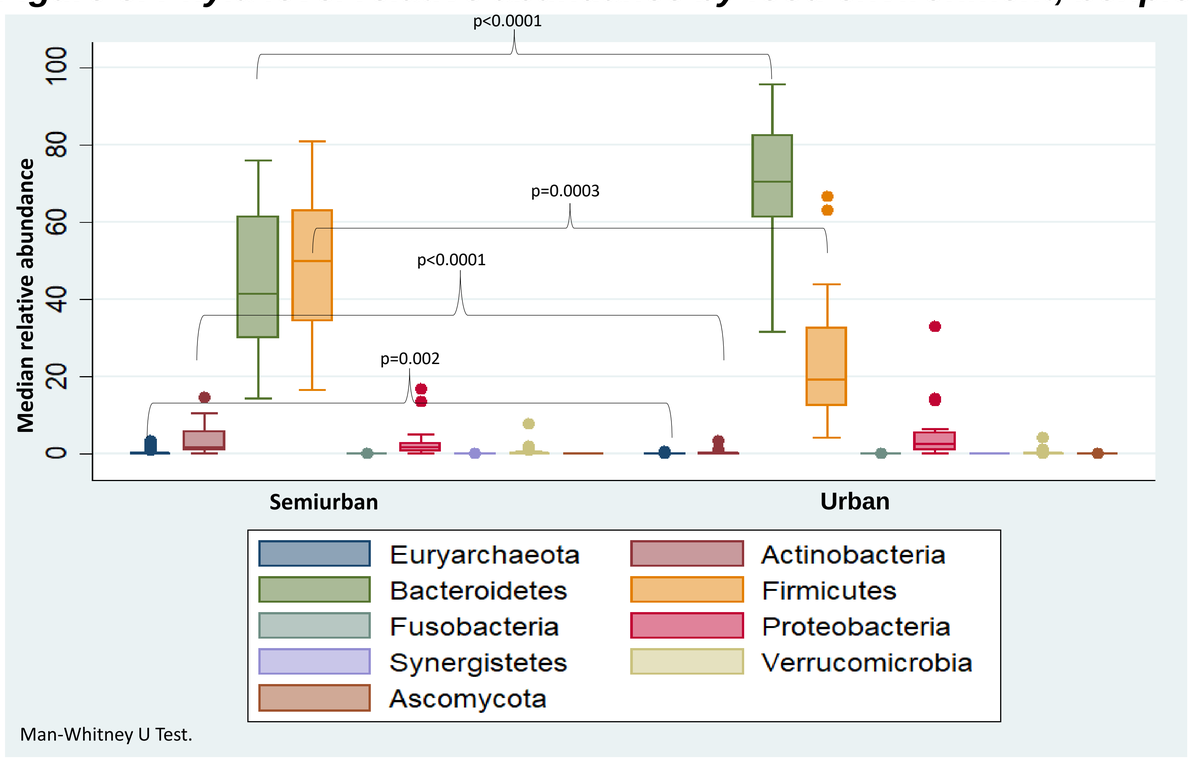
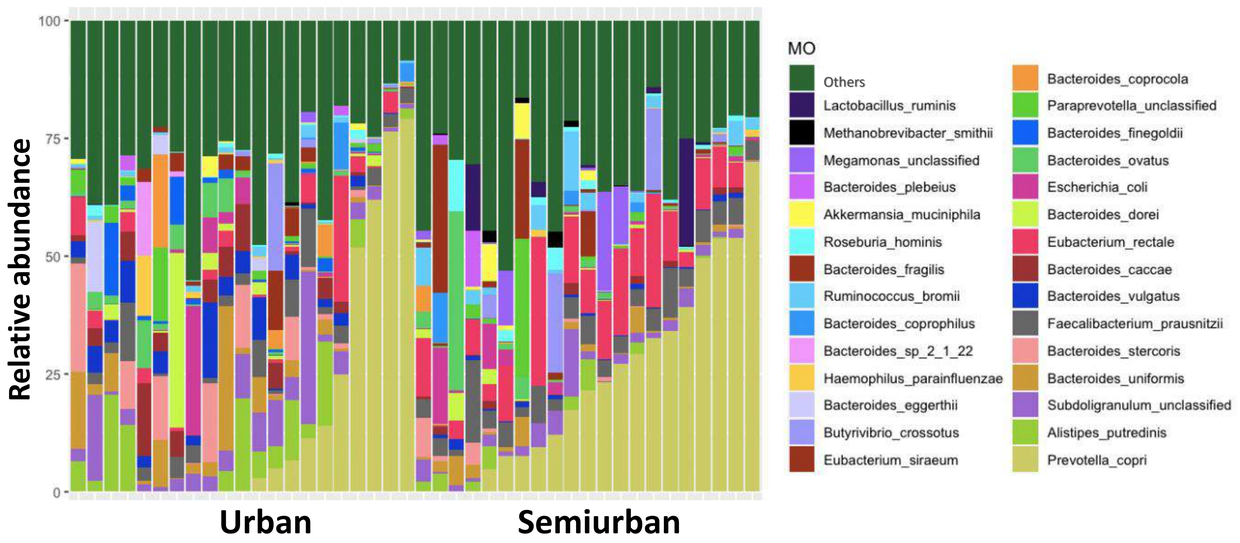


Figure 3. Phyla level relative abundance by food environment, boxplot



Man-Whitney U Test.

Figure 4. Species level relative abundance by food environment, stacked barplot



Results. We observed that 68.5% of the participants were women and 31.43% men, with a mean age of 41.8 (±11.3) years. There were no statistically significant differences for age (p=0.560) or sex (p=0.413) by environment. Semiurban participants showed higher daily consumption frequencies of ultra-processed cereals (p=0.023), soft drinks (p<0.0001) and corn-based traditional meals (p=0.018). Statistically significant differences between food environments and HDL (p=0.001) as well as fast glucose serum levels (0.004), were also observed.

Conclusions. The GM composition shows differences between individuals from urban and semiurban food environments. There are also differences in beta diversity but not in alpha diversity, between both groups. Further studies are necessary to deepen in the relationship among food enviroment, GM and metabolic diseases.

Gratefulness. CONACyT PN 2016-3251.

Bibliography.

1. Munckhof. Role of gut microbiota in chronic low-grade inflammation aspotential driver for atherosclerotic cardiovascular disease: a systematicreview of human studies. Obesity Reviews 19, 1719–1734, December 2018.
2. Swinburn B, Egger G, Raza F, et al. Dissecting Obesogenic Environments:The Development and Application of a Framework for Identifying andPrioritizing Environmental Interventions for Obesity. Preventive Medicine.1999; 29 (1): 563–570.
3. Kolodziejczyk, A.A., Zheng, D. & Elinav, E. Diet–microbiota interactions andpersonalized nutrition. Nat Rev Microbiol 17, 742–753 (2019).

Inferring the effect of microbial strains on host health outcomes with anpan

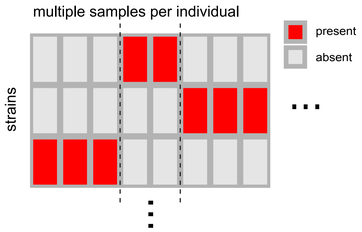
Andrew R. Ghazi[1,2] Yan Yan[1], Eric A Franzosa[1,2], Curtis Huttenhower[1,2]

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

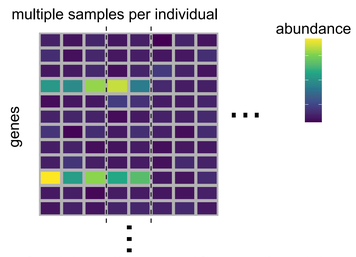
Microbial strain variation can strongly influence the impact of microbes on host health, though methods for quantitatively understanding these important differences have been lacking. Strain data have several features that make traditional statistical methods challenging to use, including high dimensionality, person-specific strain carriage, and complex phylogenetic relatedness. We present anpan, an R package that consolidates methods for strain statistics. Combining modern hierarchical modeling strategies with novel adaptive filtering methods specifically designed to interrogate microbial strain profiles, anpan facilitates the identification of strain-specific genetic elements associated with host health outcomes. Additionally, we use regularized phylogenetic generalized linear mixed models to characterize the effect of strain-level community structure. We validate our methods by simulation, as well as application to a dataset of 1262 colorectal cancer patients, showing that we achieve more accurate effect size estimation and a lower false positive rate compared to current methodologies. The open source repository with help documentation and a tutorial vignette are available at https://github.com/biobakery/anpan

Strain analysis challenges

Strains by sample:

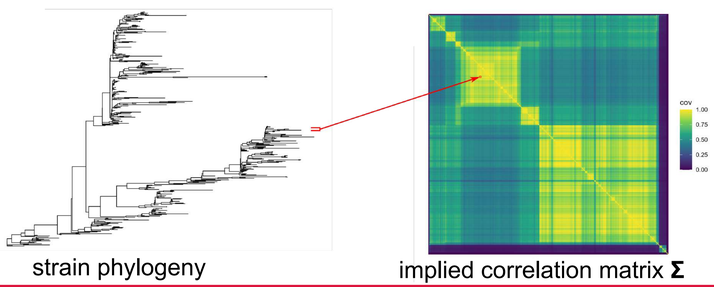


Genetic element by sample:



Microbial strains can be defined in terms of arbitrarily specific nucleotide identity thresholds. However, as a result of the finely-resolved nature of the data, unique strains rarely recur across individuals. This non-recurrence means that individual strains usually cannot be associated with clinical covariates. Analyzing the data in terms of individual microbial genes alleviates the non-recurrence at the cost of increasing the multiple testing burden by multiple orders of magnitude. Furthermore, gene-level analysis inherits error from upstream profiling steps in the form of reads being misallocated to the incorrect gene in a given microbe.

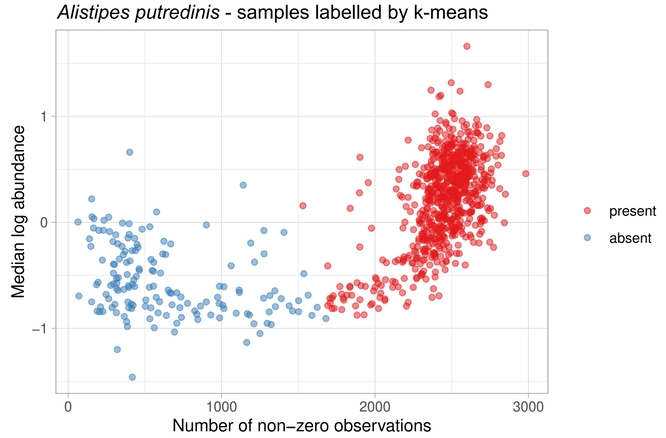
Alternatively, the phylogenetic structure may be utilized as a way of inferring the effect of strain-level variation on outcomes. However methods for probabilistically evaluating complex phylogenetic structure are lacking. Notably, methods like PERMANOVA do not accurately handle the covariance structure implied by phylogenetic trees.



strain phylogeny implied correlation matrix **Σ**

Adaptive filtering of gene profiles

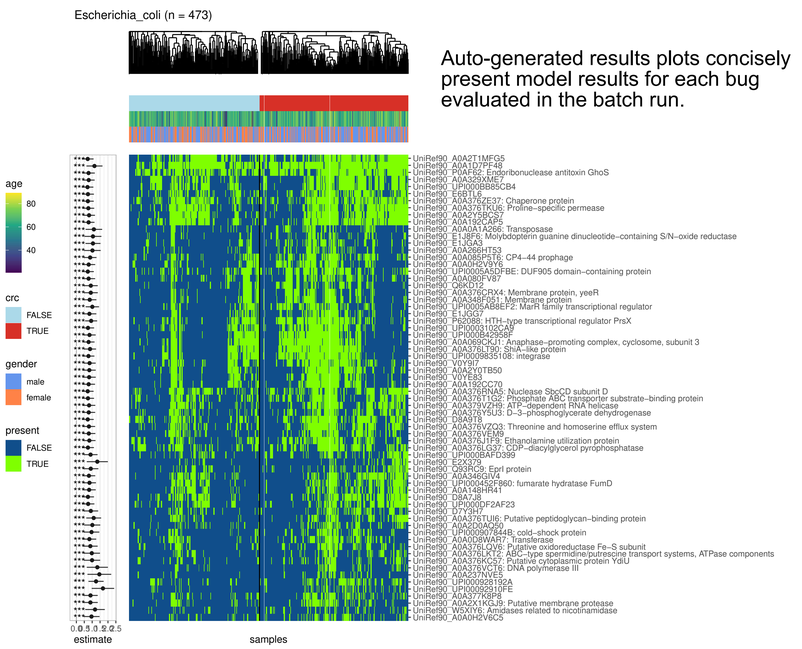
Samples with gene reads mis-allocated to the incorrect microbe must be discarded to avoid bias. Applying k-means clustering to simple summary statistics of each microbe in each samples allows assigning a clear labelling of "bug present" or "bug absent". Samples where the bug is absent are discarded be fore proceeding to the modeling step.



Modeling associations of microbial genes with clinical outcomes.

The filtered data are analyzed alongside relevant covariates using either:

- GLMs one gene at a time followed by FDR correction
- all genes and coviarates at once with a horseshoe prior on gene effects



Auto-generated results plots concisely present model results for each bug evaluated in the batch run.

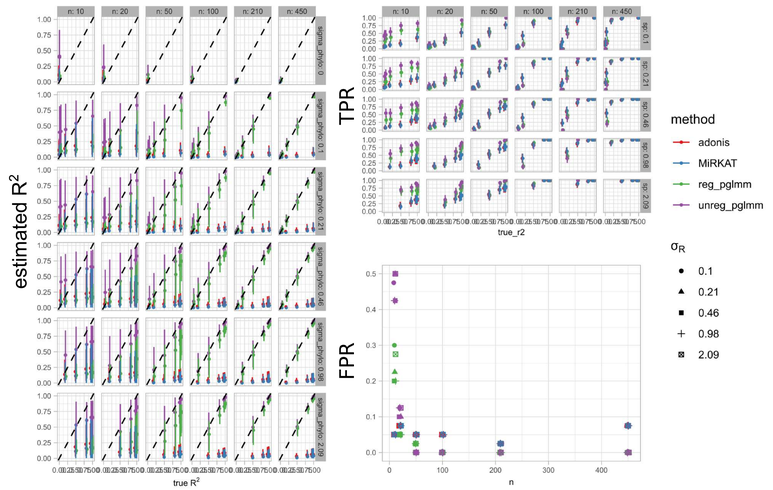
Phylogenetic generalized linear mixed models accurately quantify the effect of strains on outcomes

Mixed models can be used to analyzed data with a phylogenetic structure by incorporating a "random effect" term for each leaf. The random effects are correlated among leaves according to the phylogenetic correlation matrix **Σ** implied by the inter-leaf distances of the tree. Other model components follow familiar linear model conventions.

$$y = X\beta + (1|\text{phylogeny}) + \varepsilon$$
$$\textbf{(1|phylogeny)} \sim \text{MVNormal}(0, \sigma_P^2\boldsymbol{\Sigma})$$
$$\varepsilon \sim \text{Normal}(0, \sigma_R^2)$$

Model comparison against a "base" fit without the (1|phylogeny) term assesses the impact of phylogeny on the outcome. Integrated leave-one-out expected log pointwise predictive density is used to quantify predictive performance of the two models. Bugs where the predictive performance of the phylogenetic model substantially exceeds that of the base model indicates that within-species phylogeny impacts the distribution of the outcome variable in at least a subset of leaves.

Evaluating the performance of regularized and unregularized phylogenetic models against "permutational" linear models (e.g. PERMANOVA and MiRKAT) showed that the permutational models dramatically and consistently under-estimated effect size, while also yielding inferior classification metrics for typical sample sizes.



Acknowledgments

We appreciate the help of Dr. Aki Vehtari, the Stan Developer team, and all the other users of the Stan help forum who helped the development of this work. This work was supported by NIH NIDDK R24DK110499.

https://huttenhower.sph.harvard.edu/anpan

Maternal Stress During Pregnancy and the Neonatal Gut Microbiome

Maryam Hamidi, PhD, RN, Sandra J. Weiss, PhD, RN, FAAN

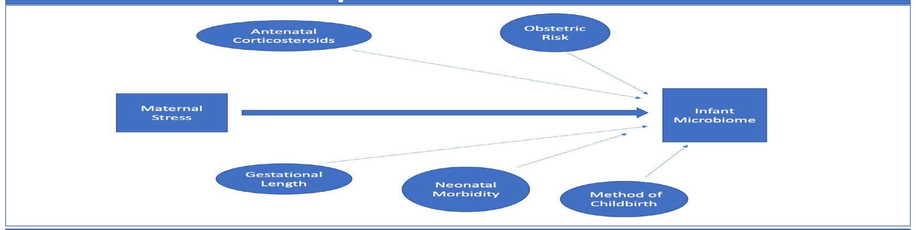
UCSF School of Nursing

Abstract

Purpose: Preliminary research suggests that prenatal stress may alter programming of the fetal microbiome. The purpose of this study was to determine relationships of pregnancy stress to diversity and composition of the neonate's gut microbiome.
Design: This cross-sectional analysis is part of a longitudinal, cohort study.
Methods: Fifty-one women and their future newborns were recruited during the third trimester of pregnancy. Women completed demographics and Cohen's Perceived Stress Scale at recruitment. A stool sample was collected from the neonate at one month of age. Data on potential confounds (e.g. gestational age) were extracted from the medical record to control for their effects. 16s rRNA gene sequencing and DESeq were used to identify diversity and abundance of species and test for differential expression of various taxa. We employed multiple linear regression to examine the aims.
Results: Greater pregnancy stress was associated with greater diversity of the neonate's gut microbiome (β = .30, p=.025). However, the abundance of certain species appeared perturbed in neonate's exposed to greater stress in utero. For instance, they had a significantly lower abundance of potentially beneficial bacteria such as Lactobacillus, Lactococcus, and Bifidobacterium.
Conclusions: A more complex, multi-species gut microbiome may lead to less stability and greater susceptibility to functional perturbations during development. Lower levels of beneficial bacteria can lead to less ability to ward off pathogenic organisms, and related Infections or intestinal disorders.
Relevance: Research could eventually yield microbial markers and microbial gene pathways that are bio-signatures of risk and inform targets for probiotic therapies.

Key words: microbiome, maternal stress, neonate, pregnancy

Conceptual Framework



Findings

Pearson Correlations between Covariates and Diversity Measures

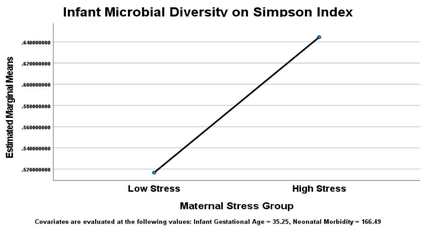
	Obstetric Risk	Corticosteroid Exposure	Method of Birth	Gestational Length	Neonatal Morbidity
Richness	-.09	-.26	.13	.33**	-.23
Evenness	-.11	-.21	.01	.31*	-.29*
Shannon Diversity	-.11	-.24	.03	.33**	-.29*
Simpson Diversity	-.11	-.23	.03	37***	-.29*

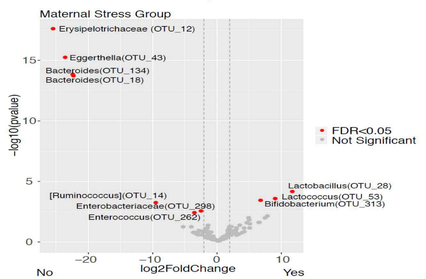
Multiple Linear Regression for Effects of Pregnancy Stress on Infant Microbial Diversity [Shannon Index]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	-.927	1.612		-.575	.568	-4.168	2.314
	Infant Gestational Age	.069	.044	.246	1.563	.125	-.020	.159
	Neonatal Morbidity	-.001	.000	-.164	-1.042	.302	-.002	.000
2	(Constant)	-.603	1.581		-.382	.705	-3.784	2.577
	Infant Gestational Age	.051	.044	.180	1.144	.258	-.039	.140
	Neonatal Morbidity	-.001	.000	-.211	-1.355	.182	-.002	.000
	Maternal Pregnancy Stress	.026	.014	.350	1.869	.068	-.002	.054

Multiple Linear Regression for Effects of Pregnancy Stress on Infant Microbial Diversity [Simpson Index]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	-.495	.606		-.817	.418	-1.714	.724
	Infant Gestational Age	.032	.017	.294	1.891	.065	-.002	.065
	Neonatal Morbidity	.000	.000	-.143	-.920	.362	-.001	.000
2	(Constant)	-.347	.584		-.594	.555	-1.521	.828
	Infant Gestational Age	.023	.016	.215	1.407	.166	-.010	.056
	Neonatal Morbidity	.000	.000	-.199	-1.320	.193	-.001	.000
	Maternal Pregnancy Stress	.012	.005	.303	2.315	.025	.002	.022





Conclusion

- Findings suggest that greater perceived stress during pregnancy is associated with greater diversity of the infant gut microbiome, not less diversity
- Complex, multispecies communities may be inherently vulnerable to destabilization - The equilibrium of an unstable microbiome can be more easily disrupted by perturbations
- Greater maternal stress may lead to reduced levels of commensal bacteria in the infant gut microbiome

Discussion and Future Direction

➢Lactobacillus, Lactococcus, and Bifidobacterium are considered commensal or beneficial bacteria with a role in regulating and enhancing immune function
➢Their lower levels among infants whose mothers had greater stress could lead to less ability to ward off pathogenic organisms, along with greater vulnerability to infections and intestinal disorders
➢Identification of distinct microbial genes and gene pathways associated with pregnancy stress is needed, along with assessment of their persistence over time
➢A longitudinal, case-control study is recommended that controls for even more covariates and involves a larger sample size

Acknowledgment

This research was funded by NICHD, RO1 HD081188 - 05 and the Robert and Delphine Wentland Eschbach Endowment (PI: Weiss, Sandra)
Dr. Hamidi's involvement was supported by NINR, T32 NR016920 - 05

Polygenic risk: gut microbiome and genetic risk on preprandial and postprandial metabolic in apparently healthy individuals.

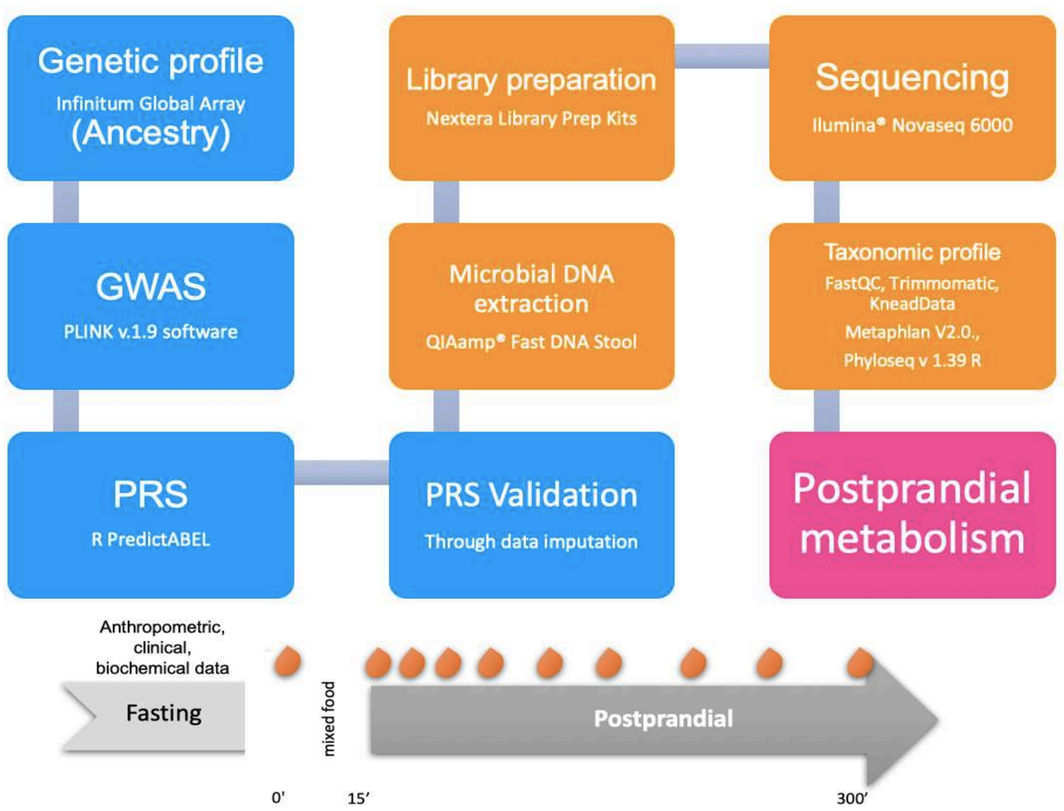
Eira E. Huerta-Avila[1], Angélica Martínez-Hernández[2], Fabiola Escalante-Araiza[2], Francisco M. Barajas-Olmos[2], Humberto García-Ortiz[2], Raul A. Bastarrachea[3], Lorena Orozco[2]

1 Immunogenomics and Metabolic Diseases Laboratory, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City/Universidad Nacional Autónoma de México, México City. 2 Immunogenomics and Metabolic Diseases Laboratory, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City. 3 Sansum Diabetes Research Institute, Santa Barbara CA, USA.

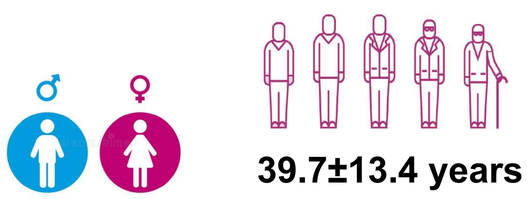
Background. The prevalence of the different metabolic syndrome phenotypes (MetSP) is increasing (1). It has been documented that, even before the onset of MetSP, altered postprandial metabolism may occur (2). A large number of genetic variants associated with the risk of developing MetSP have been identified and in recent years, it has been documented that the gut microbiota (GM) is strongly involved in the development of these entities. A widely used methodology is the polygenic risk score (PRS) that allows identifying people at risk by combining the environmental and genetic part (3-5).

Aim: To determine if there is a correlation between the GM composition, the pre and postprandial metabolic state of the individual and the PRS to develop MetSP.

Methods:

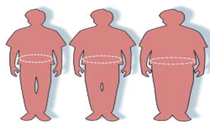


Results:

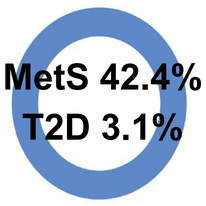


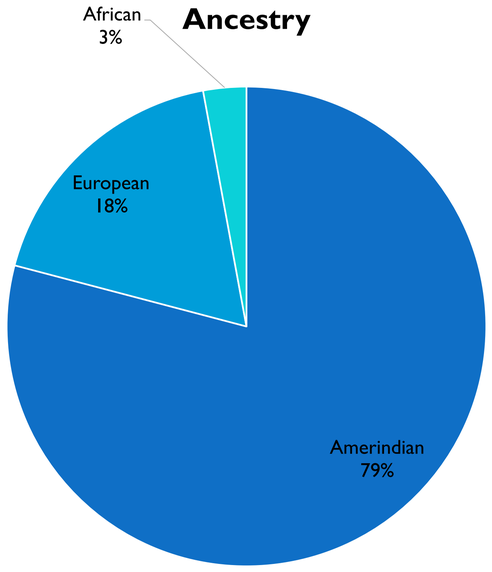
24% 76%

39.7±13.4 years



29.7±5.1 kg/m²





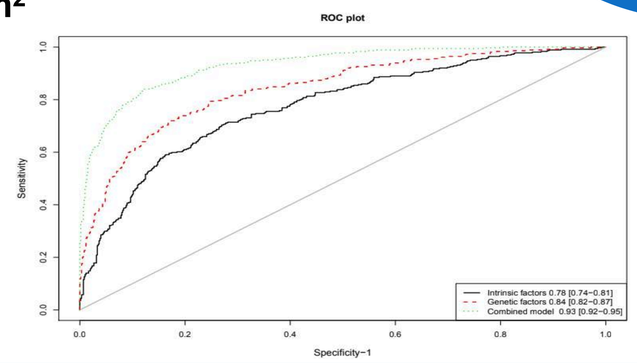
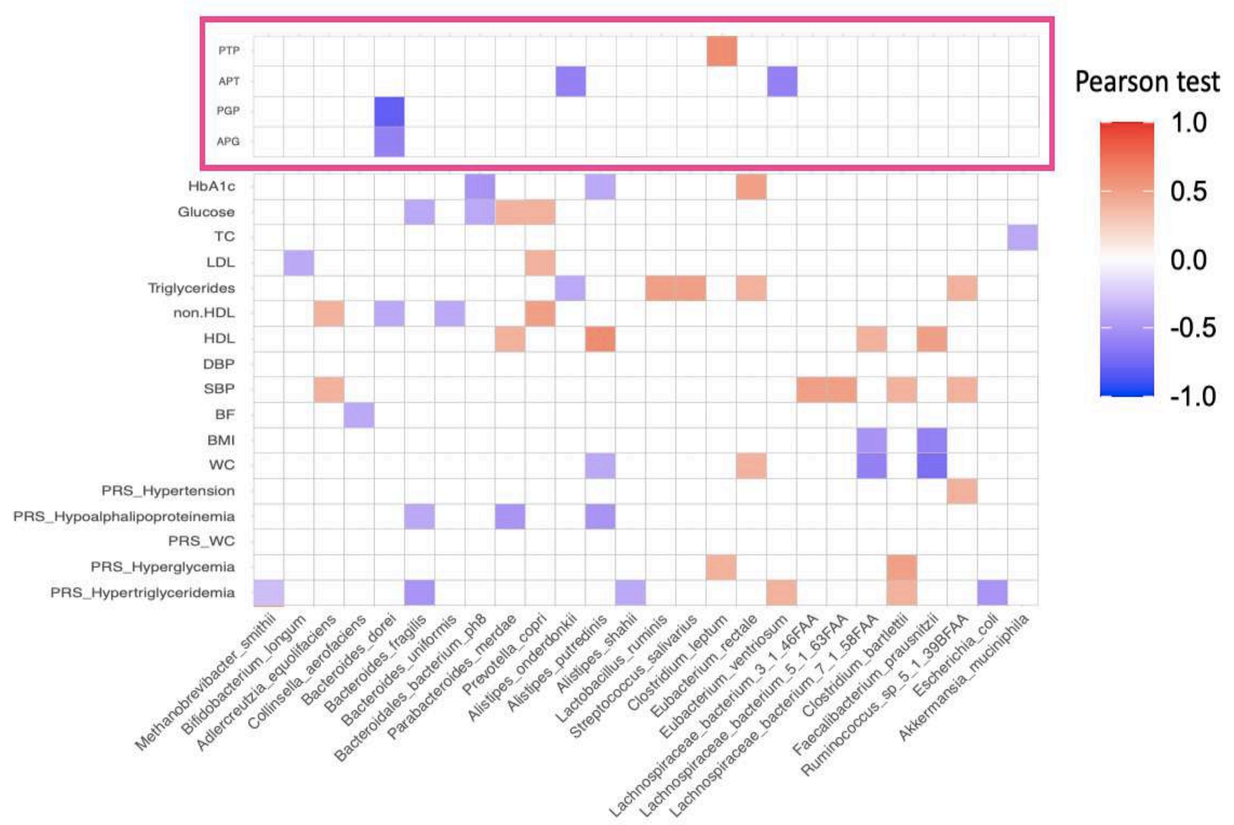
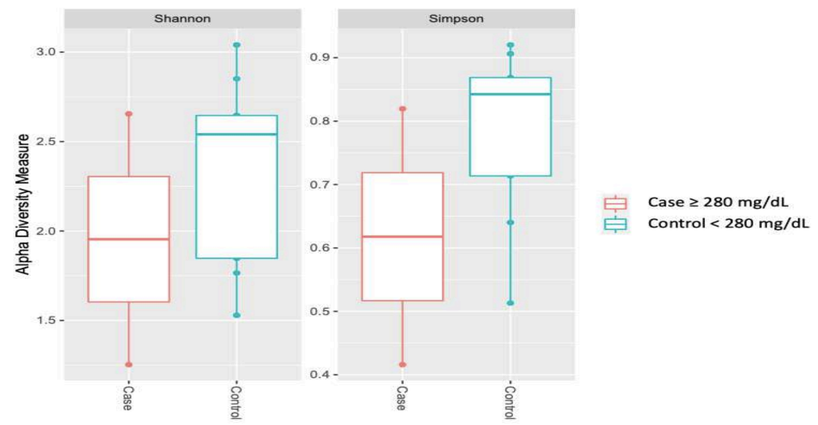


Figure 1. Correlation matrix of PRS, MetSP, postprandial state with microbiome at the species level.



Square with color = p-value <0.05 by Pearson test.
Hba1c: glycated hemoglobin, TC: total cholesterol, LDL: low density lipoprotein, non-HDL: non-high density lipoprotein, HDL: high density lipoprotein, DBP: diastolic blood pressure, SBP: systolic blood pressure, BF: body Fat, BMI: body Mass Index, WC: waist circunference, PRS: polygenic risk score, MetSP: metabolic syndrome phenotypes, PGP: postprandial glucose peak, PTP: postprandial triglyceride peak, APG: average postprandial glucose, APT: average postprandial triglycerides.

Figure 2. Alpha diversity indices according to postprandial triglyceride response.



Conclusions:

1. The decrease in gut microbiome diversity could occur in the stages before the onset of manifestation MetSP.
2. Some PRS, such as for hypertriglyceridemia, correlate negatively with several species of microorganisms, such as the controversial archaea Methanobrevibacter smithii. Meanwhile, Alistipes onderdonkii correlated negatively with triglyceride levels both fasting and postprandial.

It is necessary to increase the sample size to corroborate these data.

Gratefulness:CONACyT: PN 2016-3251. CONACYT (scholarship number 704009) for the financial support for my PhD degree in biomedical sciences.

Bibliography

1. Huang P. A comprehensive definition for metabolic syndrome Dis Model Mech. 2009. 2(5-6): 231–237.
2. Sottero B, et al. Postprandial Dysmetabolism and Oxidative Stress in Type 2 Diabetes: Pathogenetic Mechanisms and Therapeutic Strategies. *Medicinal Research Reviews*. 2015. 35(5):968-1031.
3. Sanghera D, *et al.* Obesity genetics and cardiometabolic health: Potential for risk prediction. Diabetes Obes Metab. 2019; 21(5).
4. Kruttika D, et al. The gut microbiome and metabolic syndrome. J Clin Invest. 2019.129(10).
5. Martin AR, *et a*l. Clinical use of current polygenic risk scores may exacerbate health disparities. Nature. 2019. 51(1):584

Evaluating assembly- and reference-based methods for virome analysis with and without enrichment

Jordan Jensen[1,2], Lea Wang[2,3,4], Moreno Zolfo[5], Nicola Segata[5], Eric A. Franzosa[2,3,4], Curtis Huttenhower[1,2,3,4]

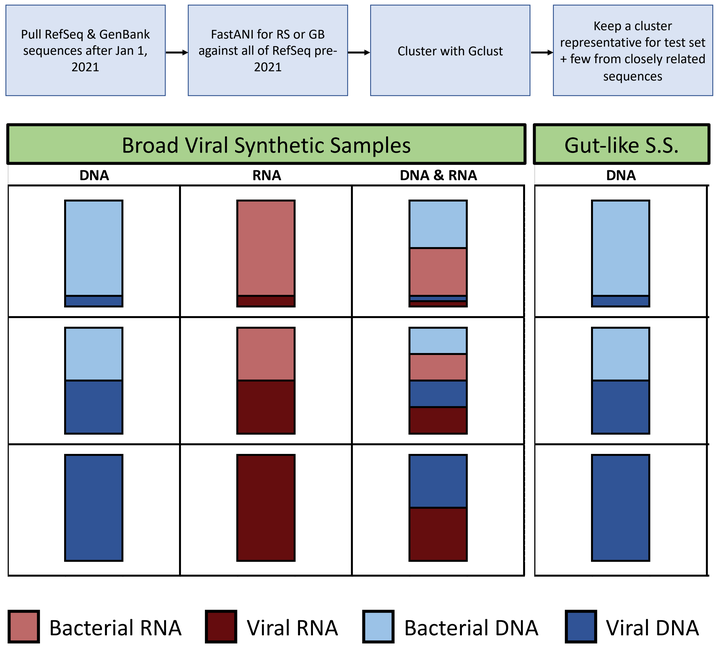
THE HARVARD CHAN MICROBIOME IN PUBLIC HEALTH CENTER

[1]: Department of Immunology and Infectious Diseases, Harvard TH Chan School of Public Health, Harvard University, Boston, MA, USA; [2]:Harvard Chan Microbiome in Public Health Center, Harvard TH Chan School of Public Health, Harvard University, Boston, MA, USA; [3]: Department of Biostatistics, Harvard TH Chan School of Public Health, Harvard University, Boston, MA, USA; [4]: Broad Institute of MIT and Harvard, Boston, MA, USA; [5]: Centre for Integrative Biology, University of Trento, Italy.

Abstract

Capturing an accurate representation of the viral members of a microbial community presents significant experimental and computational challenges. We systematically evaluated a series of assembly- and reference-based approaches to viral profiling, using synthetic sample sets along with shotgun metagenomes, metatranscriptomes, and virus-like particle (VLP)-enriched viromes from the IBDMDB cohort of the Integrative Human Microbiome Project. We found that mapping to well-characterized reference sets such as RefSeq maintained high specificity across a wide range of bacterial contamination, but failed to capture highly novel viral content. Viral metagenome assembled genome (vMAG) reference sets varied in mapping rates of viral reads, but were able to expand mapping of synthetic novel microbial sequences.

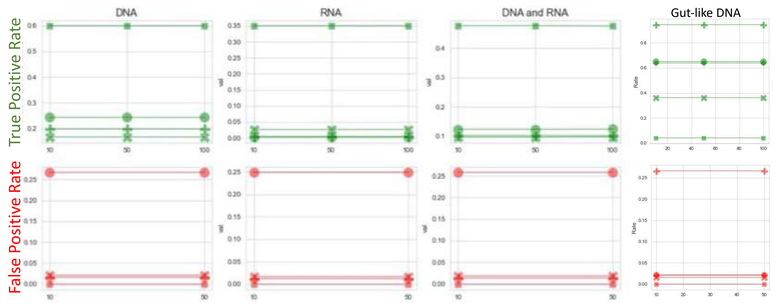
Synthetic Viral Samples



Discover Huttenhower Lab software & tutorials via
http://huttenhower.sph.harvard.edu/biobakery

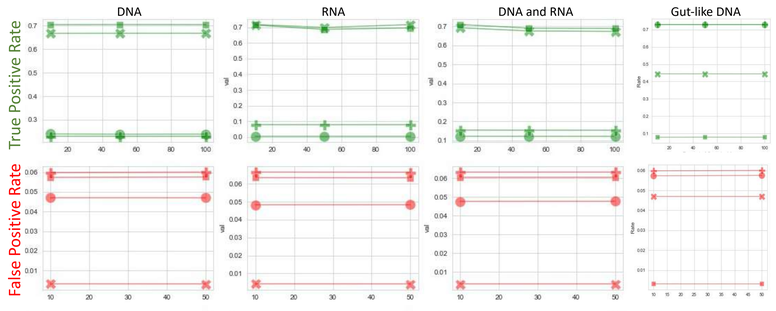
Nucleotide and translated mapping of synthetic samples

Synthetic samples were mapped to either vMAG sets: Gut Phage Database (GPD), Gut Virome Database (GVD), Viral Sequence Clusters (VSC), or gold-standard RefSeq (RS). Nucleotide mapping was carried out on each database with bowtie2, then databases were translated into the 6 possible translation frames and translated mapping was performed with DIAMOND BLASTX.



Nucleotide Mapping DNA, RNA, and DNA/RNA: RS □, GPD O, GVD ×, VSC +. **Gut-like DNA:** RS □, GPD +, GVD ◊, VSC ×, RS+GVD O

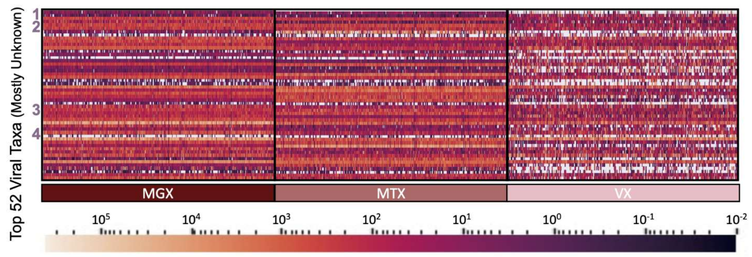
- RefSeq has the highest true positive rate (TPR) and lowest false positive rate (FPR) for broad viral synthetic samples
- Gut Phage Database has the highest TPR as a consequence of high inaccuracy- demonstrating the highest FPR
- Viral MAG sets perform better against novel microbial sequences
- Combining RefSeq and a vMAG set (GVD) expands TPR without a subsequent large raise in FPR



Translated Mapping DNA, RNA, and DNA/RNA: RS ×, GVD +, VSC O , RS+GVD □. **Gut-like DNA:** RS O, GVD +, VSC ×, RS+GVD O

- Translated mapping increases TPR for RNA viruses significantly, while DNA viral gains are smaller, comparatively
- Translated mapping raises FPR only minimally

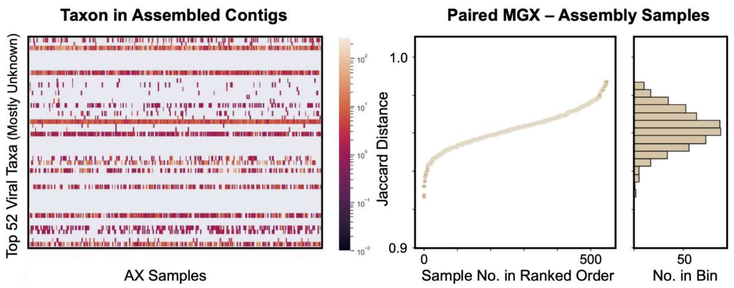
Expanded viral reference sets greatly enriches viral profiling



Heatmap of joined top 25 taxa found in MGX, MTX, and MVX datasets (rows) across all samples (columns) for MGX, MTX, and MVX datasets.

- Most taxonomic observations (M-groups) did not have previous or known taxonomic assignment. Those associated with previously identified taxa are marked in purple: **1 & 4:** Lactococcus Phage **2:** Uncultured CrAssphage **3:** Enterobacteria Phage
- MGX and MTX samples are species-rich and show similar trends in taxonomic observations.
- MVX taxon observations are more sparse and exhibit more disagreement with MTX and MGX taxon observations.

Assembly generates shallow but accurate species observations



Left Heatmap of taxa found in AX set for the top 25 taxa from MGX, MTX, and MVX datasets (rows) across all samples (columns). **Right** Jaccard distances between assembled (AX) and non-assembled MGX reads mapped to VSCs with Bowtie2 (MGX) or BLAST (AX).

- Top taxa identified among MGX, MTX, and MVX are sparse or absent from many samples in the AX set.
- High Jaccard distances between AX and MGX data from the same samples reflects low richness in species observations from AX data alone, in contrast to high MGX richness.

The Long-term Association of Post-Traumatic Stress Disorder with Dietary Pattern and Gut Microbiome in a Cohort of Women

Shanlin Ke[1], Xu-Wen Wang[1], Andrew D. Ratanatharathorn[2], Tianyi Huang[1,3], Andrea Roberts[4], Francine Grodstein[5], Laura D. Kubzansky[7], Karestan C. Koenen[6,#], Yang-Yu Liu[1,#]

[1] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School. [2] Department of Epidemiology, Mailman School of Public Health, Columbia University. [3] Department of Nutrition, Harvard T.H. Chan School of Public Health. [4] Department of Environmental Health, Harvard T.H. Chan School of Public Health. [5] Rush Alzheimer's Disease Center, Rush University Medical Center. [6] Department of Epidemiology, Harvard T.H. Chan School of Public Health. [7] Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health.

Abstract

- Post-Traumatic Stress Disorder (PTSD) is a psychiatric condition that may occur in people who have experienced or witnessed traumatic or horrifying events. The gut microbiome plays a critical role in modulating the immune, metabolic, psychological and cognitive activities of the host. Understanding PTSD's long-term associations on dietary pattern and gut microbiome may improve the physical and mental health of people with PTSD, but remains unexplored. Here we analyzed information on trauma exposure and PTSD symptoms with microbiome data and dietary information collected about 5 years later, in 191 individuals enrolled in the Mind-Body Study (MBS). We found that inter-individual differences in gut microbiome appear to be stable over time intervals as long as six months, and thus a limited number of measurements may be adequate to reliably investigate associations with long-term health. Notably, we demonstrated that PTSD has a long-term inverse association with host dietary habits, especially a healthy Mediterranean-style dietary pattern. Moreover, three (i.e., *Bacteroides ovatus*, *Roseburia inulinivorans*, and *Dorea longicatena*) and four (i.e., *Eubacterium siraeum*, *Bacteroides massiliensis*, *Ruminococcus gnavus*, and *Oscillibacter unclassified*) differentially abundant species were identified in No-trauma vs. Trauma-no-PTSD and Trauma-no-PTSD vs. PTSD comparisons, respectively. Several functional pathways were found significantly enriched in PTSD group, including pyrimidine deoxyribonucleotide and L-ornithine de novo biosynthesis. Overall, these findings suggest that PTSD is associated with long-term changes in dietary pattern and gut microbiome, highlighting the critical importance of incorporating the human gut microbiome and diet in our understanding of PTSD and their association with physical health.

Introduction

PTSD is a leading contributor to the global disease burden and is estimated to affect almost 4% of the world's population. General population studies have shown that a large proportion of people in developed countries have been exposed to at least one traumatic event in their lifetime. Numerous studies have shown that PTSD is associated with medical comorbidities, including asthma6, cardiovascular disease, chronic pain and inflammation, obesity, type 2 diabetes, and gastrointestinal disorders. However, our understanding of how PTSD influences chronic disease development remains limited.

We recently found that PTSD was associated with lower improvement in overall diet quality over 20 years. Food and nutrition is not only essential to human health, but also modulates the human gut microbiome. Comprising trillions of microbes including bacteria, archaea, fungi, and viruses, the human gut microbiome plays a vital role in our physiology, metabolism, homeostasis, and immunity. In recent years, the gut microbiome has been linked to the development and function of the central nervous system.

We therefore hypothesize that PTSD may link with host health through the associations with diet and the gut microbiome. To test this hypothesis, we using data from a large cohort of female registered nurses in the United States (the Nurses' Health Study II: NHS-II) to systematically examined the associations of trauma exposure and PTSD status in 2008 with dietary data and whole-genome shotgun sequencing from stool samples collected in 2013. The primary goal of this study is to understand the longitudinal association of PTSD with dietary intake and the gut microbiome, which may have important implications for human health (**Fig.1**).

Materials and methods

- All participants completed the baseline questionnaire, including basic demographic characteristics, and the cohort has been followed by biennially mailed questionnaires to update information on a variety of lifestyle and health-related factors and ascertain incident diseases.
- Among the 191 participants, 160, 21, 10 and 0 participants provided four, three, two and one stool samples, respectively.
- Trauma exposure was measured with a 16-item modified version of the Brief Trauma Questionnaire (BTQ).
- Microbial taxonomic profiling and functional profiling were performed using MetaPhlAn2 and HUMAnN2, respectively.

Fig. 1. Conceptual framework of study.

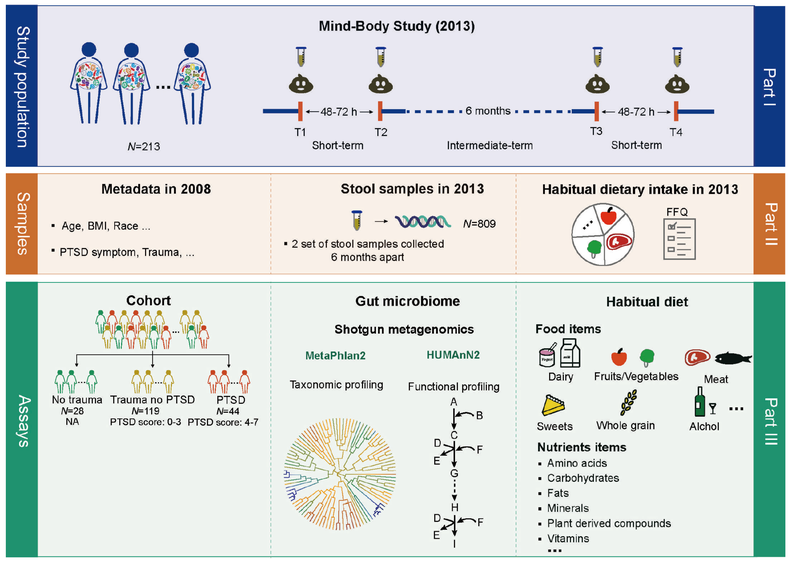
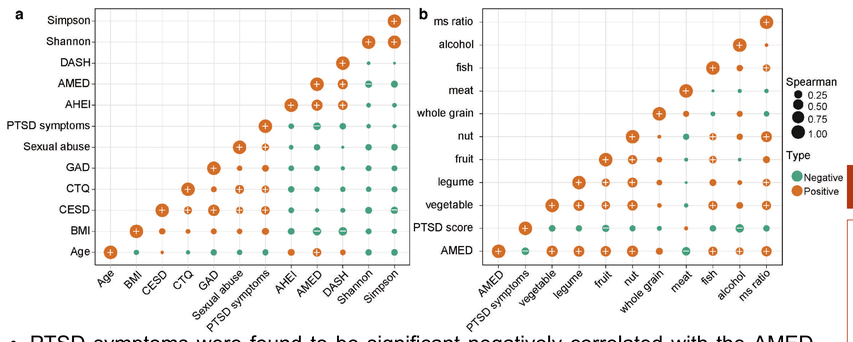
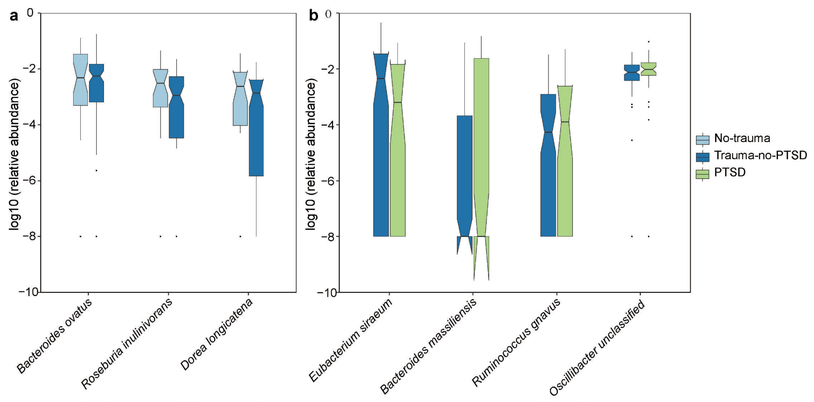


Fig. 2. Correlation between PTSD score, host factors, and microbiome diversity.



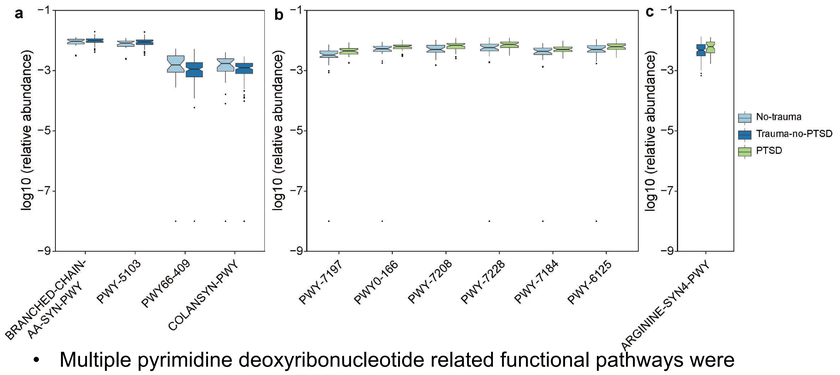
- PTSD symptoms were found to be significant negatively correlated with the AMED score (**Fig.2a**). Specifically, the consumption of plant-based foods was negatively correlated with the PTSD score, while the consumption of red/processed meat was slightly positively correlated with PTSD symptoms (**Fig.2b**).

Fig. 3. Relative abundance of PTSD related microbial features.



- A total of three differential abundant species were identified from the comparison between No-trauma and Trauma-no- PTSD, including *Bacteroides ovatus*, *Roseburia inulinivorans*, and *Dorea longicatena* (**Fig.3a**).
- *Eubacterium siraeum*, *Bacteroides massiliensis*, *Ruminococcus gnavus*, and *Oscillibacter unclassified* were significantly different between Trauma-no-PTSD and PTSD groups (**Fig.3b**).

Fig. 4. Relative abundance of PTSD related microbial features.



- Multiple pyrimidine deoxyribonucleotide related functional pathways were significantly enriched in PTSD group when compared to No-trauma group. (**Fig.4a**).
- L-ornithine de novo biosynthesis (ARGININE-SYN4-PWY) was significantly over-represented in PTSD group when compared to Trauma-no-PTSD group. (**Fig.4b**).

Conclusions

In this study, we comprehensively evaluated the long-term associations of PTSD with dietary patterns and gut microbiome in 191 women. In our analysis, several species and functional pathways were shown to be significantly correlated with the PTSD symptoms. Our study also linked severe PTSD symptoms to the low MedDiet adherence, particularly in association with low intake of plant based foods. Overall, these findings highlighting the critical importance of incorporating the human gut microbiome and diet in our understanding of PTSD and their association with physical health.

Acknowledgements

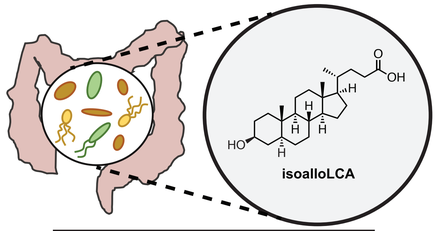
Y-Y Liu is supported by grants R01AI141529, R01HD093761, R01AG067744, UH3OD023268, U19AI095219, and U01HL089856 from National Institutes of Health, USA. K.C.K., A.D.R., L.K. and A.R. are supported by R01MH101269 from National Institutes of Health, USA.

Human gut bacteria produce T_{reg}-modulating bile acid metabolite

Wei Li[1,§], Saiyu Hang[1,§], Yuan Fang[1,2], Sena Bae[3], Yancong Zhang[3,4], Minghao Zhang[5], Gang Wang[1], Megan D. McCurry[1], Munhyung Bae[1], Donggi Paik[1], Eric A. Franzosa[3,4], Fraydoon Rastinejad[5], Curtis Huttenhower[3,4], Lina Yao[1,*], A. Sloan Devlin[1,*], Jun R. Huh[1,6,*]

[1]Harvard Medical School; [2]Harvard University; [3]T. H. Chan School of Public Health; [4]Broad Institute of MIT and Harvard; [5]Target discovery Institute; [6]Harvard Medical School and Brigham and Women's Hospital. [§]These authors contributed equally.

Abstract



- Bacteria influence the development and function of host immune cells, including T_{reg} cells.
- In a previous study, we showed that the bile acid metabolites isoallolithocholic acid (isoalloLCA) modulate T_{reg} cell differentiation [1].
- Here, we identified human gut bacteria that produce isoalloLCA from a known bile acid metabolite, 3-oxoLCA, as well as the enzymes responsible.

Figure 1. A previous study shows that isoalloLCA, a bacterial metabolite affect T_{reg} cell differentiation.

Screen for bacterial producers in human stool

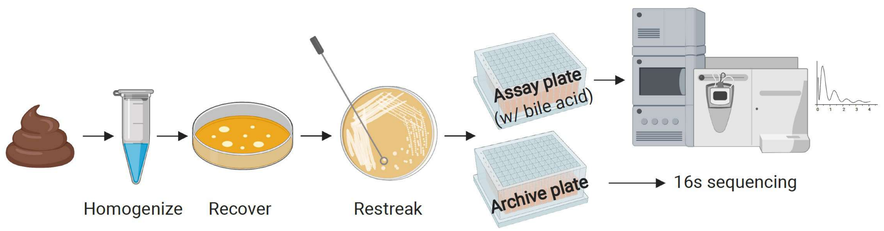


Figure 2. Schematic of the screen for bacterial producers of the bile acid metabolites isoalloLCA from two human stool samples (p#3 and p#27). In total, 990 cultivable colonies were screened.

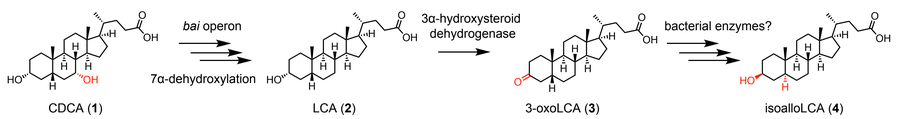


Figure 3. The known pathway from human bacteria converting primary bile acid chenodeoxycholic acid (CDCA) (1) to 3-oxolithocholic acid (3-oxoLCA) (3), a putative substrate for isoalloLCA (4). The transformation of (1) to (2) and (2) to (3) is known and characterized.

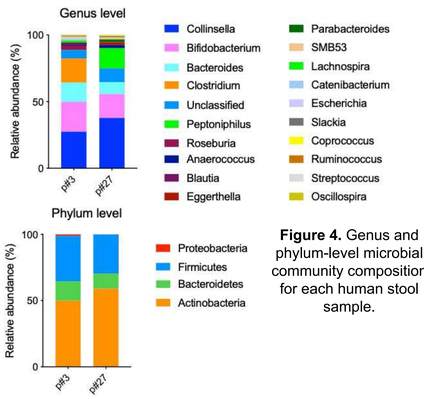


Figure 4. Genus and phylum-level microbial community composition for each human stool sample.

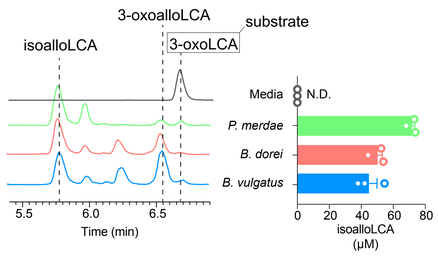


Figure 5. Identified gut bacteria in the Bacteriodetes phylum that can convert 3-oxolithocholic acid (3-oxoLCA) (3) to for isoalloLCA (4)

Elucidate the genes and pathways for isoalloLCA

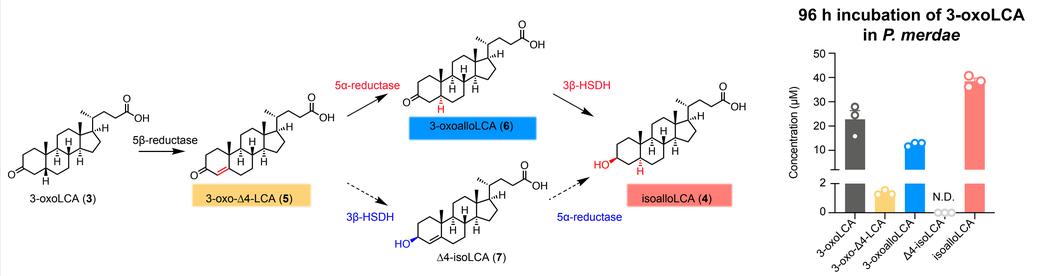


Figure 6. This study identified novel biosynthetic genes from gut bacteria that convert 3-oxoLCA (3) to isoalloLCA (4). Two pathways were proposed, with the top red pathway being the identified pathway for the isoalloLCA-producing bacteria.

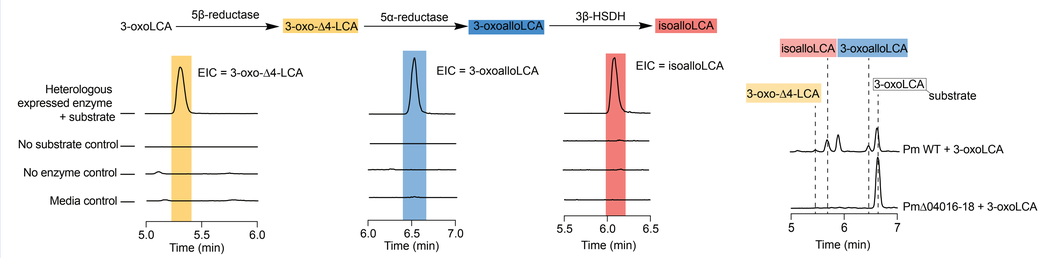


Figure 7. The pathway is confirmed by characterizing the activity of each enzyme and by deleting the genes in the native bacteria.

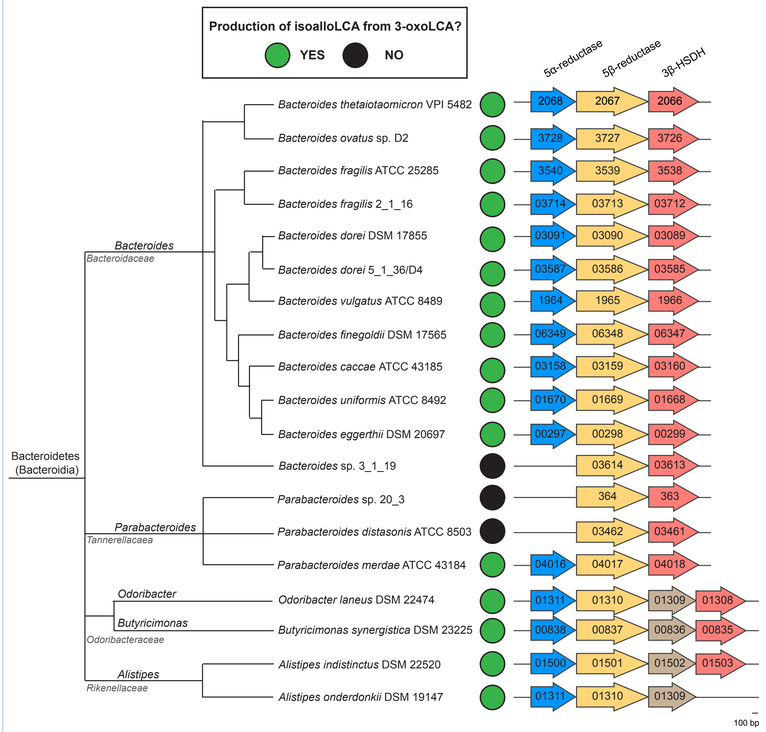


Figure 8. IsoalloLCA production from Bacteroidetes containing complete or partial gene clusters homologous to the recombinantly verified gene cluster.

Conclusion

① Discovered bacterial producers of immunomodulatory isoalloLCA (**Fig. 2-5**)

② Uncovered a novel bacterial biosynthetic gene cluster (**Fig. 6-8**)

③ Gene cluster is required to produce isoalloLCA in vivo (**Fig. 9**)

④ Human disease relevance: IBD (**Fig 10**)

In vivo production of isoalloLCA

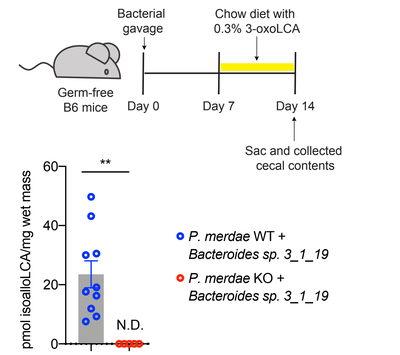


Figure 9. IsoalloLCA is produced in vivo, but not when the isoalloLCA-producing gene cluster is deleted.

Human Disease Correlation

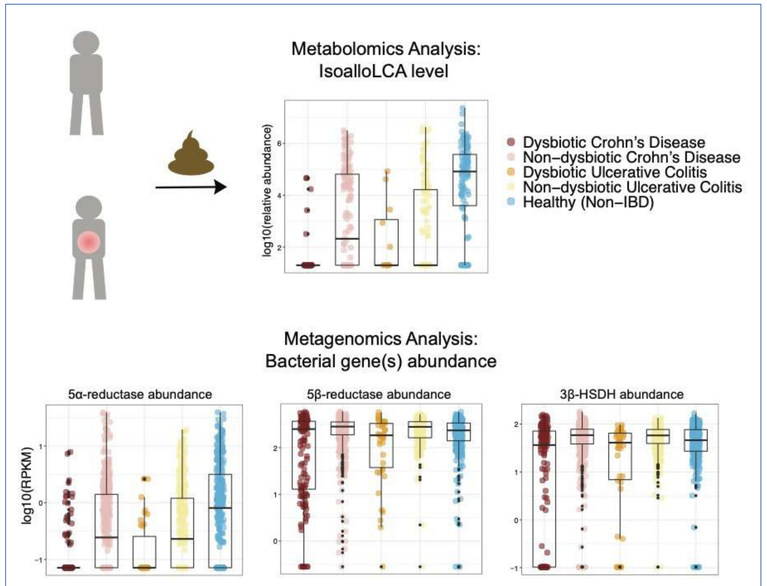


Figure 10. Both metabolite levels and the corresponding bacterial gene(s) were depleted in human inflammatory bowel disease (IBD), Crohn's disease (CD) or/ and ulcerative colitis (UC) patients.

Acknowledgement

The Devlin, Huh, and Clardy labs (HMS) for helpful discussions. We are grateful to the human patients who participated in the human stool screen, PRISM and HMP2 studies. Fig. 2 was created using BioRender.

References

1. Hang, S.[§], Paik, D.[§], ... Fischbach, M.[*], Littman, D.[*], **Huh, J.**[*]. Bile acid metabolites control T_H17 and T_{reg} cell differentiation. Nature 576, 143–148 (2019).
2. Paik, D.[§], Yao, L.[§], ... Huh, J.[*], **Devlin, A.S.**[*]. Human gut bacteria produce T_H17-modulating bile acid metabolites. bioRxiv, under revision (2021).
3. Li, W.[§], Hang, S.[§], ... Yao, L.[*], Devlin, A.S.[*], **Huh, J.**[*]. A Bacterial Bile Acid Metabolite Modulates Treg Activity through the Nuclear Hormone Receptor NR4A1. Cell Host Microbe 29(9), 1366–1377 (2021).

Engineered *E. coli* for the targeted deposition of therapeutic payloads to sites of disease

Jason P. Lynch[1,2], Coral González-Prieto[1,2]; Analise Z. Reeves[1,2]; Neha Godbole[1,2]; Nadia Sahli[1,2]; Charles B. Shoemaker[3]; Wendy S. Garrett[4,5]; Cammie F. Lesser[1,2,5]

[1]Center for Bacterial Pathogenesis, Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA 02114; [2]Department of Microbiology, Harvard Medical School, Boston, MA 02115; [3]Cummings School of Veterinary Medicine, Tufts University, North Grafton, MA 01536; [4]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA 02115; [5]Broad Institute of Harvard and MIT, Cambridge, MA 02142;

INTRODUCTION

- New drug platforms are needed which enable the directed delivery of therapeutics to sites of disease to maximize efficacy and limit off-target effects.
- Here we developed PROT$_3$EcT (probiotic type 3 *E. coli* therapeutic), commensal *Escherichia coli* engineered for the direct secretion of proteins into their surroundings
- PROT$_3$EcT has four modular components: an *E. coli* chassis, a modified bacterial protein secretion system, a regulatable transcriptional activator, and a secretable therapeutic payload (Figure 1).

Goals

- Develop designer probiotics for targeted delivery of therapeutic payloads into the gut lumen.
- Provide proof of principle for this new therapeutic strategy for the treatment of gut inflammation

PROT$_3$EcT: a programmable therapeutic delivery device

Interchangeable and Optimizable Components:

1) E. coli chassis
2) Regulator (VirB)
3) T3SS, Delivery Apparatus
4) Heterologous substrate, SS-target protein fusion

Fig.1 Schematic overview of PROT$_3$EcT.

Development of nanobodies (aka VHH) as type III secreted substrates

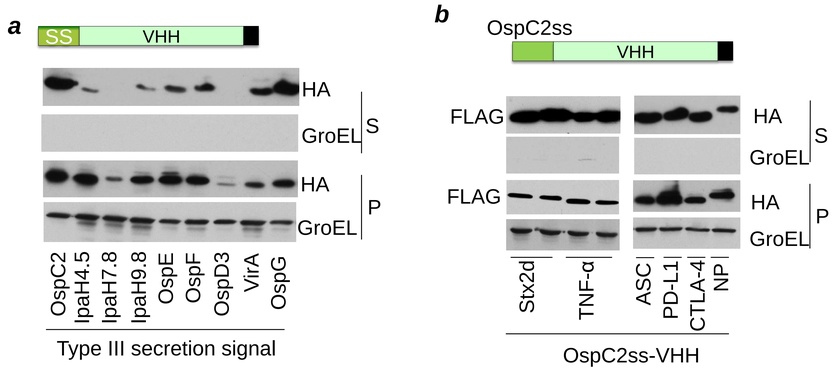


Fig. 2. Liquid secretion assays were performed for 6 h. IPTG 1 mM was used to induce protein expression. Pellets (P) and supernatants (S) were harvested and FLAG or HA western blots performed (**d**) GelCode Blue staining.

PROT$_3$EcT secreted VHH are functional, stable and are the predominant secreted protein

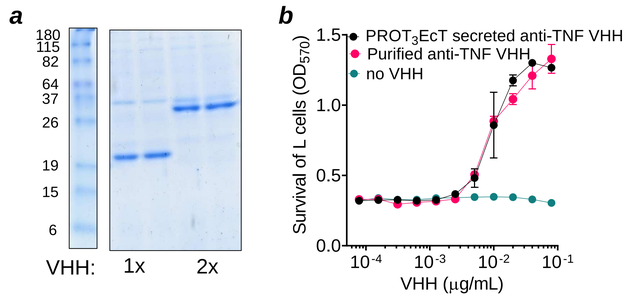


Fig. 3 (**a**) FLAG western blot of a 6 h liquid secretion assay. For VHHs, 1x=monomer; 2x=dimer. (**b**) Inhibition of TNF-induced killing of L929 cells as measured by an MTT assay.

Summary of circuit optimization

- T3SS activator (virB) integrated into the EcN chromosome.
- virB driven by a synthetic constitutive promoter.
- VHH on a plasmid that complements an auxotrophic variant of T3EcN.

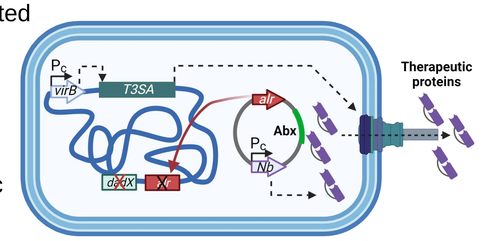


Fig. 4 Circuit optimization in PROT$_3$EcT

PROT$_3$EcT colonizes and remains secretion competent in the GI tract

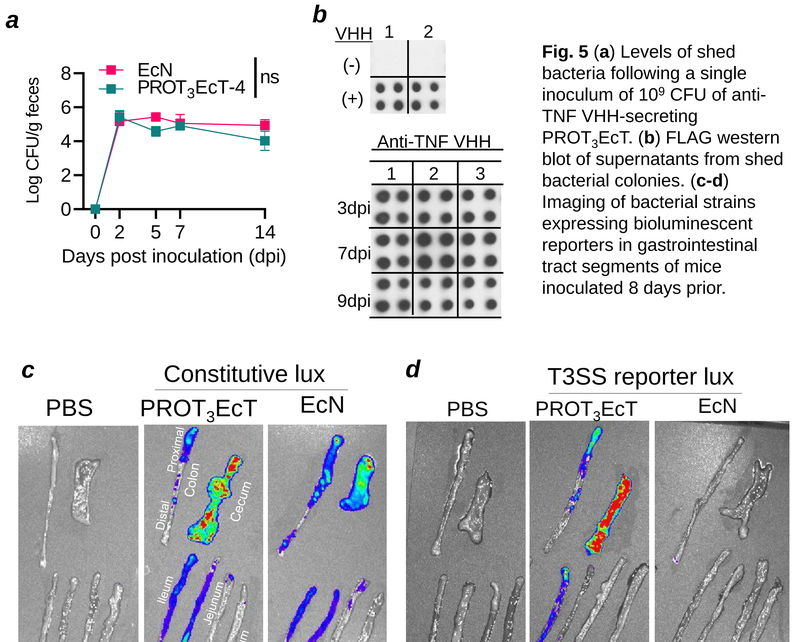


Fig. 5 (**a**) Levels of shed bacteria following a single inoculum of 10^9 CFU of anti-TNF VHH-secreting PROT$_3$EcT. (**b**) FLAG western blot of supernatants from shed bacterial colonies. (**c-d**) Imaging of bacterial strains expressing bioluminescent reporters in gastrointestinal tract segments of mice inoculated 8 days prior.

Anti-TNF VHH-secreting PROT$_3$EcT ameliorates inflammation in the TNBS colitis model

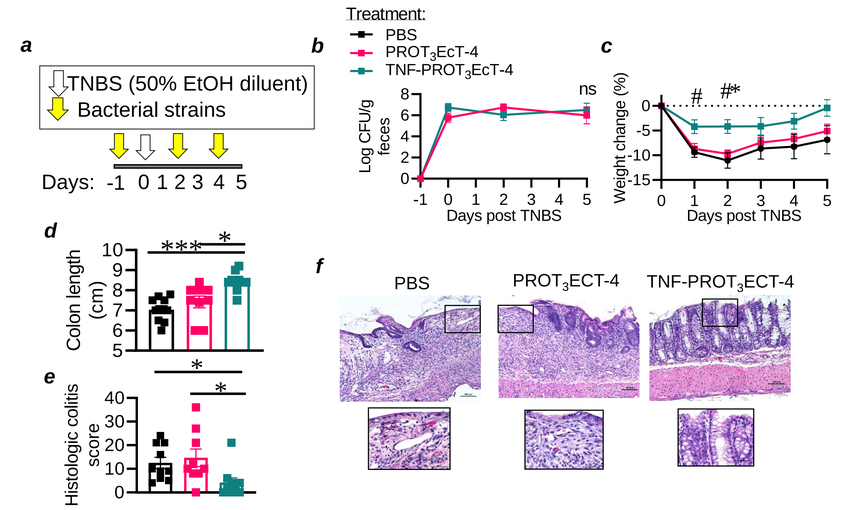


Fig. 7 (**a**) BALBc mice (Jax) received TNBS (2 mg, enema) plus either PBS or inoculum of 10^8 CFU of PROT$_3$EcT or TNF-PROT$_3$EcT. (**b**) Shed bacteria. (**c**) Weight change. (**d**) Colon lengths. (**e**) Histologic colitis score. (**f**) Representative histology of H&E stained colon sections.

Single dose of TNF-PROT$_3$EcT ameliorates TNBS colitis model

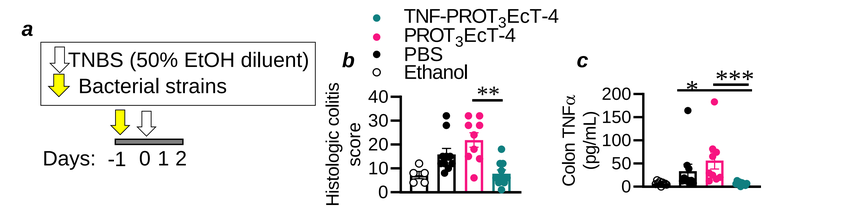


Fig. 8 (**a**) BALBc mice (Jax) received TNBS (2 mg, enema) plus either PBS or inoculum of 10^8 CFU of PROT$_3$EcT or TNF-PROT$_3$EcT. (**b**) Histologic colitis score. (**c**) TNF-α was measured by ELISA.

Conclusions and Ongoing Work

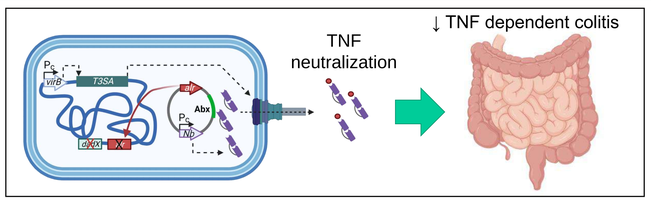


Fig. 9 Working model.

- PROT$_3$EcT-secreted VHH retain activity following type III secretion.
- PROT$_3$EcT stably colonizes the GI tract and its T3SS is active in the gut.
- Anti-TNF VHH-secreting PROT$_3$EcT ameliorate gut inflammation in the TNBS colitis model.
- Ongoing work is focused on developing variants that target other cytokines (e.g., IL-12/IL-23) and other diseases (e.g., cancer).

Acknowledgements

This work was funded by a Transformative Research Award (R01DK113599) to CFL and WSG; a Rainin Award to CFL; an Endeavour Australia Fellowship and a Crohn's & Colitis Foundation Fellowship to JPL. The authors thank Jonathan N. Glickman (HMS) for evaluating colitis scores and Karen Inouye (HSPH) for assistance with IVIS imaging.

Association of Bowel Movement Frequency with Cognitive Function in Women and Men

Chaoran Ma[1], Kristine K. Dennis[1], Raaj Mehta[2], Changzheng Yuan[3], Jae Hee Kang[1], Francine Grodstein[4], Alberto Ascherio[1,5], Walter C. Willett[1,5], Andrew T. Chan[2,5], Curtis Huttenhower[5,6], Meir J. Stampfer[1,5], Dong D. Wang[1,5]

[1]Brigham and Women's Hospital/Harvard Medical School, Boston, MA; [2]Massachusetts General Hospital/Harvard Medical School, Boston, MA; [3]Zhejiang University, Hangzhou, China; [4]Rush Medical College, Chicago, Illinois; [5]Harvard T.H. Chan School of Public Health, Boston, MA; [6]Broad Institute of Harvard and MIT, Boston, MA

Background

- Abnormal gastrointestinal motility and perturbations in the gut microbiome may contribute to cognitive impairment and dementia
- Few studies have investigated this potential link through the gut microbiome in humans using high-resolution microbial profiling technology and examined species-level microbial features

Objectives

- To examine the associations of bowel movement frequency with objective and subjective cognitive function and explore the mediating role of the gut microbiome in a subsample

Methods

- 139,978 women and men from Nurses' Health Study (NHS), Nurses' Health Study II (NHSII) and Health Professionals Follow-Up Study (HPFS)
- Bowel movement frequency reported in 2012/2013
- Self-reported subjective cognitive dysfunction collected in 2014-2017
- A subset (n=14,586) of participants completed the CogState neuropsychological battery for objective cognitive assessment in 2014-2018
- The gut microbiome using shotgun metagenomics in fecal samples collected from a subpopulation of 515 women and men

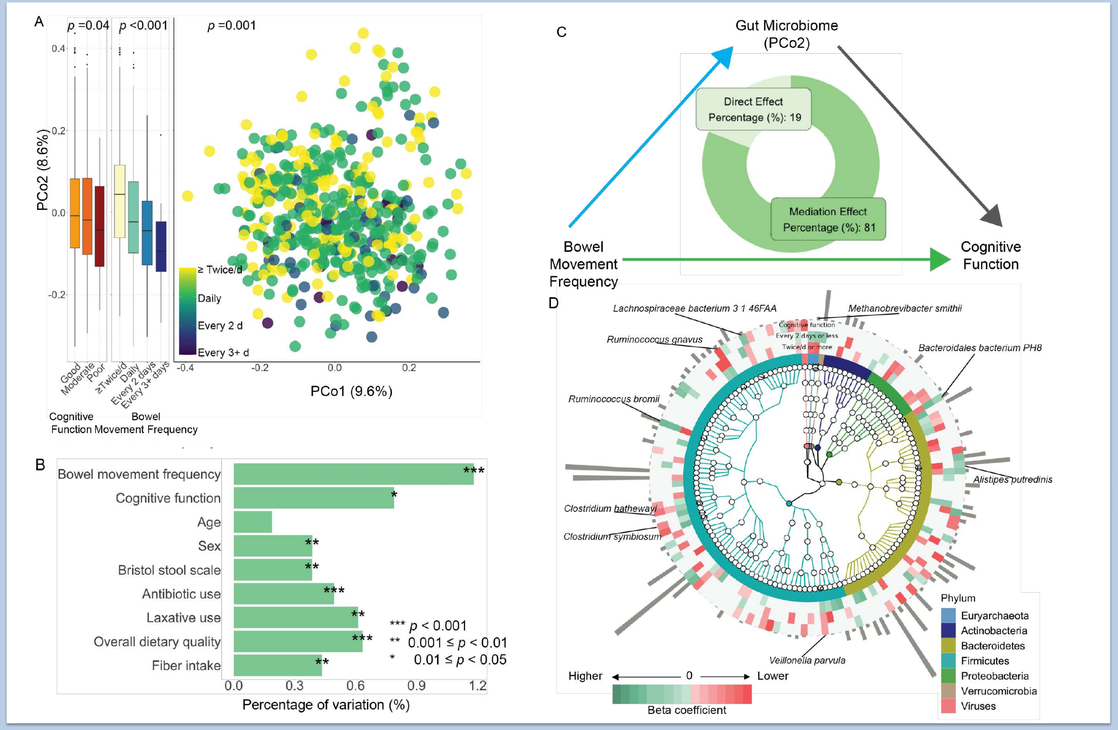
Results

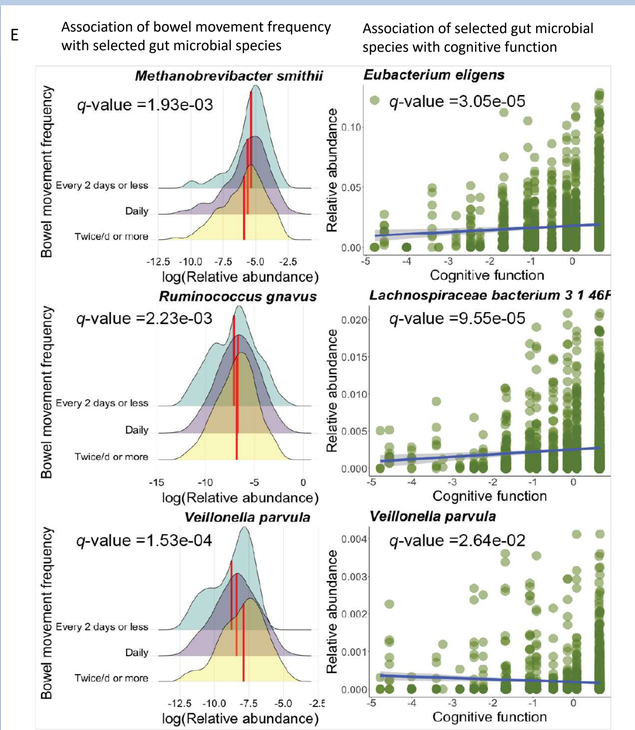
Table 1. Bowel movements and objective cognitive function (NHS2)

	Frequency of Bowel Movements					
	Every 3+ day	Every 2 days	Daily	Twice/day	>Twice/day	P quadratic
Global Cognitive Function	-0.06 (-0.10, -0.02)	0.02 (-0.01, 0.04)	Ref.	0.01 (-0.02, 0.03)	-0.03 (-0.07, 0.01)	0.02
Learning and Working Memory	-0.07 (-0.12, -0.02)	-0.00 (-0.04, 0.03)	Ref.	-0.01 (-0.05, 0.02)	-0.05 (-0.09, -0.01)	0.007
Psychomotor Speed and Attention	-0.05 (-0.11, 0.01)	0.04 (-0.01, 0.08)	Ref.	0.03 (-0.02, 0.07)	-0.01 (-0.07, 0.05)	0.38

Table 2. Bowel movements and subjective cognitive function (NHS, NHS2, HPFS)

	Frequency of Bowel Movement				
	Every 3+ days	Every 2 days	Daily	≥Twice/day	P quadratic
Difference in global cognitive function	-0.17 (-0.19, -0.14)	-0.06 (-0.07, -0.04)	Ref.	-0.08 (-0.09, -0.06)	<0.001
Odds ratio of cognitive decline	1.18 (1.08, 1.30)	1.08 (1.01, 1.15)	Ref.	1.10 (1.05, 1.15)	<0.001





(A) Associations of bowel movement frequency and cognitive function with overall variation of the gut microbiome based on the principal coordinate analysis using species-level Bray-Curtis dissimilarity. (B) The proportion of variation in taxonomy explained by bowel movement frequency, cognitive function, and covariables based on permutational multivariate analysis of variance. (C) The gut microbiome mediated the association between bowel movement frequency and subjective cognitive function. (D) Phylogenetic associations of bowel movement frequency and subjective cognitive function with microbial species. (E) Associations of bowel movement frequency and subjective cognitive function with select gut microbial species.

Conclusions

- An abnormal intestinal motility pattern was associated with worse cognitive function.
- The association may be mediated through changes in the gut microbiome composition

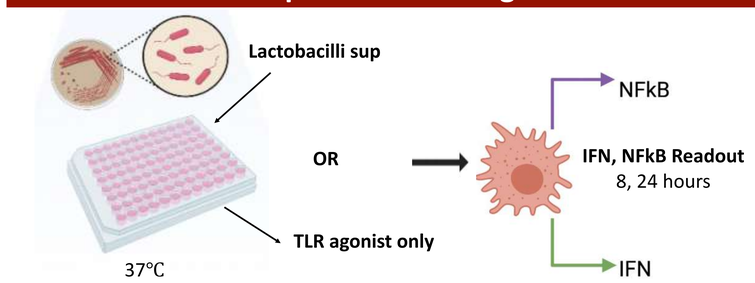
Vaginal Lactobacilli-mediated Anti-inflammatory Effects on Host Immune Response

Morgan Martin[1]; Cecilia Webber[1]; Ki Hyun Kim[2]; Sunghee Bang[2]; Jon Clardy[2]; Smita Gopinath[1]
[1]Harvard T.H. Chan School of Public Health, [2]Harvard Medical School

Abstract

Vaginal bacterial communities of people across the world are dominated by a few lactobacilli species. Loss of lactobacilli dominance is linked to increased vaginal inflammation with higher levels of pro-inflammatory cytokines. Our work aims to understand how the presence of dominant Lactobacilli species may modulate host immune response to reduce inflammation. We used a human monocyte-derived macrophage reporter cell line, THP1, to screen cell-free supernatants from vaginal lactobacilli for immunomodulatory effects on type I Interferon and NFκB activation and downstream signaling. Addition of cell-free supernatant from vaginal but not intestinal lactobacilli strains suppressed activation of multiple Toll-Like Receptors including TLR2, 3 and 4. Prior activation of cells with TLR agonists induced Interferon (IFN) and NFκB activation which was suppressed upon addition of cell-free supernatant from vaginal but not intestinal lactobacilli species. Our results suggest that vaginal lactobacilli secrete compounds that suppress inflammatory signaling in human macrophages. With our collaborators, we have screened for and identified active fractions from the supernatant of *Lactobacillus crispatus* strain MV1A. This work aims to help elucidate the bacterial metabolite(s) responsible for suppression of host TLR signaling pathways. The findings of this study will help increase our understanding of how the lactobacilli-dominated vaginal microbial communities influence host immunity to reduce inflammation.

Experimental Design



Screen of Lactobacilli supernatants:
Human monocyte-derived macrophage reporter cells, THP1s, were stimulated for 2 hours with TLR 2, 3, or 4, to induce IFN and NFκB activation. Next, the THP1s were incubated with cell-free supernatant from various vaginal and intestinal lactobacilli species. Supernatant was collected from the cells 8- and 24-hours post-incubation, and IFN and NFκB induction were measured.

Fractionation of *Lactobacillus crispatus*:
Lactobacillus crispatus strain MV1A was grown in MRS media, and the supernatant was separated from the cell pellet by centrifuge prior to a standard fractionation process. The obtained supernatant was applied to solvent-partitioning with acetate (EtOAc) to obtain EtOAc fraction. High-performance liquid chromatography (HPLC) using a reverse-phase HPLC column eluting a gradient solvent system of water and methanol was performed on the EtOAc fraction to obtain the individual fractions tested in our screening.

Results

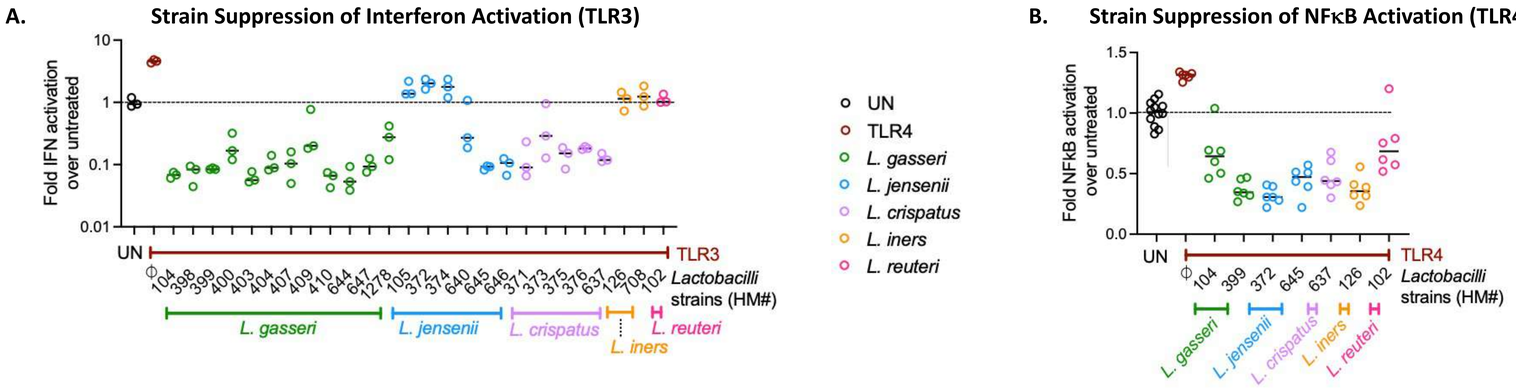


Figure A-B. Vaginal Lactobacilli Cell-Free Supernatants Suppress IFN and NFkB Induction upon TLR3 and TLR4 Activation: THP1s were stimulated with 1mg/ml Kasugamycin (A) or 500ng/ml LPS (B) prior to being incubated with 5% v/v cell-free supernatant from vaginal Lactobacilli strains including *L. gasseri, L. jensenii, L. crispatus,* and *L. iners* and an intestinal Lactobacilli strain *L. reuteri.* Control wells received media only (Φ). Interferon induction (A) and NFkB induction (B) were measured 24- and 8-hours post-incubation respectively.

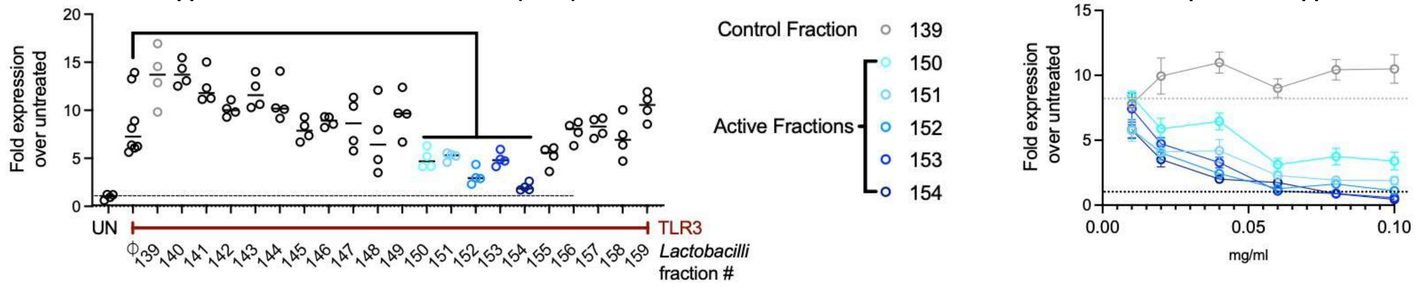


Figure C-D. *L. crispatus* Fractions (#150-154) Suppress IFN upon TLR3 Activation in a Dose-Dependent Manner: THP1s were stimulated with 1mg/ml Kasugamycin prior to being incubated with the indicated *L. crispatus* MV1A fraction at the concentration of 0.1mg/ml (C) or a serial dilution ranging from 0.01 - 0.1mg/ml (D). Interferon induction was measured via luminescence 24 hours post-incubation. The grey dotted line (D) represents the average fold induction of interferon activity by the TLR3 agonist control.

Summary and Future Directions

- Vaginal lactobacilli cell-free supernatants suppress IFN while intestinal lactobacilli (*L. reuteri*) cell-free supernatant does not. Both vaginal and intestinal lactobacilli cell-free supernatants suppress NFκB induction.
- Complete suppression to untreated levels is observed in both basal IFN and post-TLR agonist activation conditions, using multiple TLR agonists.
- Active fractions of *L. crispatus* MV1A were successfully identified and confirmed to be the same active fractions from three independent bacterial cultures. Their suppressive activity is dose-dependent.
- The future directions of this project include expanding the screen to epithelial cells for increased understanding of the bacterial supernatants' effect on immune response and potential identification of active compound(s) in *L. crispatus* responsible for the anti-inflammatory effect observed.

Acknowledgements

This work is supported by Harvard University's Biological Sciences in Public Health (BPH) PhD program training grant. In the Gopinath lab, we are grateful for Maryam Ahmad's help in maintaining the cell line of interest.

References

1. Ravel, J. et al. (2011). Vaginal microbiome of reproductive-age women. Proceedings of the National Academy of Sciences, 108 (1), 4680-4687. DOI: 10.1073/pnas.1002611107
2. Gosmann, C. et al. (2017). Lactobacillus-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young South African women. *Immunity, 46*(1), 29-37. DOI: 10.1016/j.immuni.2016.12.013

Linking individual microbial and metabolite biomarkers of Colorectal Cancer using computational models of gut microbiota metabolism

Matthew S. Miyasaka[1,2], Ali R. Zomorrodi[2,3]

[1]Depaetment of Economics, Harvard Faculty of Arts and Sciences, [2]Mucosal Immunity and Biology Research Center, Mass General Hospital for Children, [3]Harvard Medical School,

1. Abstract

Colorectal cancer (CRC) is the second most deadly cancer in the world, affecting almost 150,000 Americans each year and leading to over 50,000 deaths. Numerous microbial species and the metabolites they produce in the gut have been associated with the development and spread of CRC. However, it remains unknown to what extent the gut microbes, individually or collectively, contribute to the gut metabolome and affect CRC carcinogenesis. To better understand the functional role of individual microbial species in CRC carcinogenesis, we used a bottom-up systems approach based on GEnome-scale Models (GEMs) of metabolism to functionally profile the gut microbiota at species and molecular level resolution in CRC. To this end, we used publicly available fecal metagenomic data from 30 subjects with CRC and 30 non-CRC controls to construct GEMs of the gut microbiota metabolism at species-level resolution (spanning 127 microbial species). By computationally simulating these models, we could infer the metabolic activity of each microbial species in the gut, which allowed us to trace back individual microbial species producing several secreted metabolites. This analysis identified 338 metabolites with differential production levels by the gut microbiota as well as 656 linkages between specific microbial species and metabolites that were significantly different between CRC subjects and non-CRC controls (Wilcoxon, adjusted $p < 0.05$).Several of these identified metabolites and species have been previously implicated in CRC, examples of which include chicory insulin, deoxycytidine, and acetate that are produced by *Faecalibacterium prausnitzi*, *Eubacterium eligens*, and *Bacteroides vulgatus,* respectively, amongst other species, according to our models. Overall, our study provides a roadmap for mechanistically linking microbial and metabolite biomarkers of CRC.

2. The gut microbiome and CRC

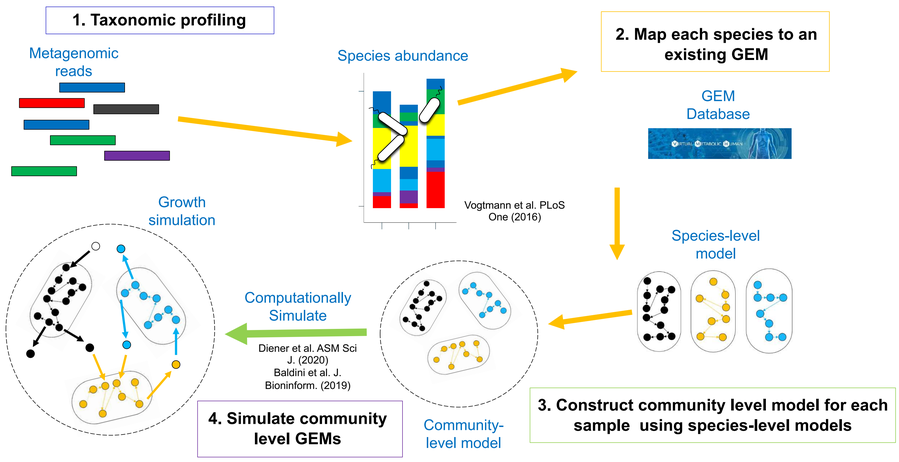
- ❑ The colon contains the highest microbial density in the gut
- ❑ Diseases like CRC may cause or be caused by changes to the microbial environment
- ❑ Analyzing differences in the microbiome between CRC patients and controls helps us identify relevant species and metabolites

3. Individual microbes' effect

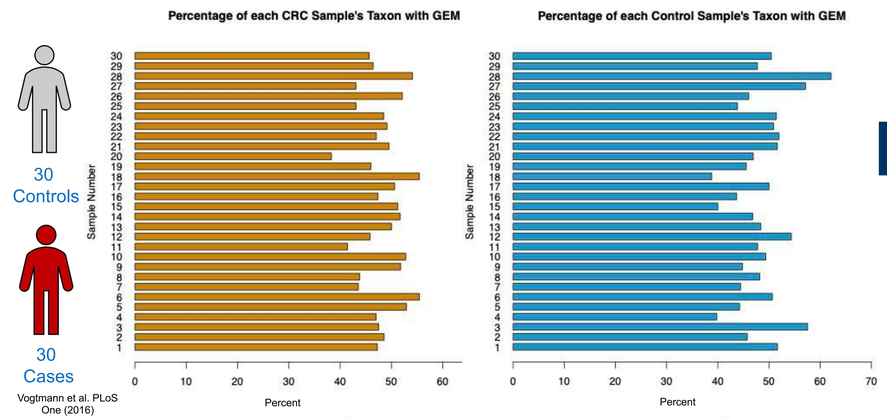
- ❑ While many species and metabolites have been associated with CRC, individual species' functional role in its carcinogenesis is less understood
- ❑ The goal of our research is to identify and analyze Specific species-metabolite links to instances of CRC

4. GEnome-scale Models (GEMs)

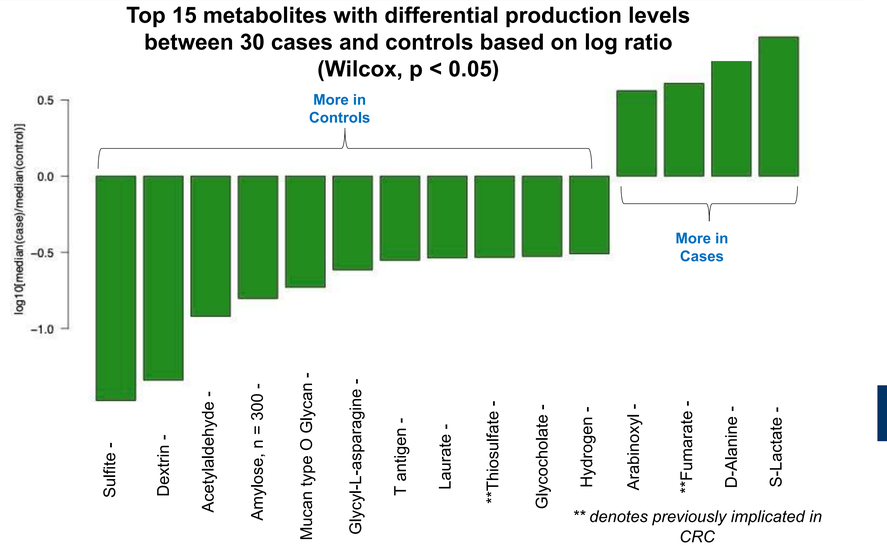
5. Methods



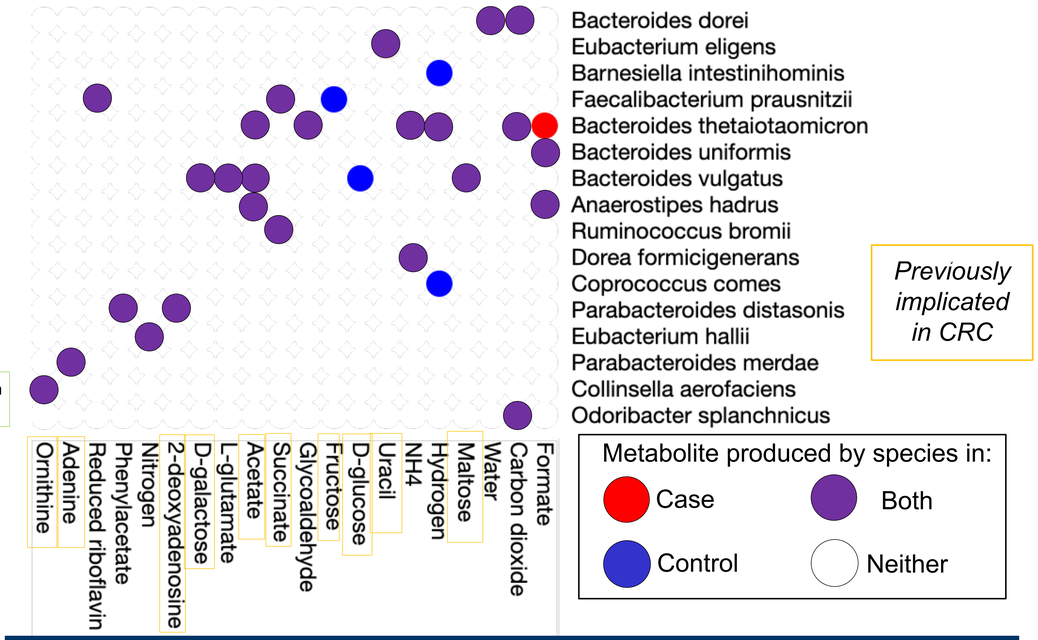
6. Study populations and community GEM reconstruction



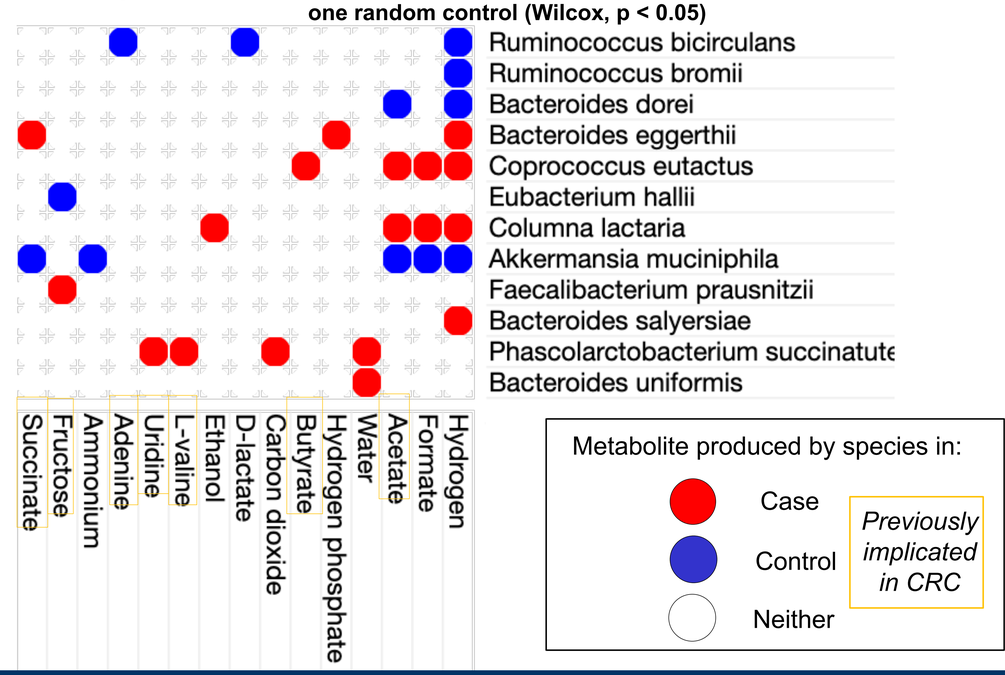
7. Metabolites differentially produced by cases and controls



8. Metabolite-species link for 30 cases and 30 controls



9. Metabolite-species links for one CRC case vs one control



10. Conclusions

- ❑ Our analysis successfully linked previously identified metabolites relevant to CRC to specific species
- ❑ Moving forward, we will conduct this analysis on a larger scale, also considering samples from other cohorts to consider the effects of different diets on CRC carcinogenesis

EFFECTS OF WHEY PROTEIN SUPPLEMENTATION ON GUT MICROBIOTA IN COLITIS MOUSE MODEL

Montibeller, Maria Jara[1]; Ferrocino, Ilario[2]; Rodrigues Cardoso, Daniel[3]; Cardoso Umbelino Cavallini, Daniela[1]

[1] School of Pharmaceutical, São Paulo State University, Araraquara, Brazil, [2] DISAFA - Department of Forestry, Agriculture and Food Sciences, University of Torino, Italy, [3] São Carlos Institute of Chemistry, University of Sao Paulo, São Carlos, Brazil.

Background

The Ulcerative Colitis (UC) is characterized by chronic relapsing intestinal inflammation. Considering the impact of UC on the quality of life, dietary interventions have been proposed to reduce the associated symptoms. Supplementation with food proteins as a source of bioactive peptides like whey protein (WP) may reduce symptoms of UC by modulation of gut microbiota and immune system due the production of mucins and metabolites. In this way, the short-wave ultraviolet light (UV-C) procedures have been used to improve the digestibility and release of bioactive peptides during WP digestion. In the present study induced UC mice were fed a standard diet plus WP with or without UV-C light treatment.

Methodology

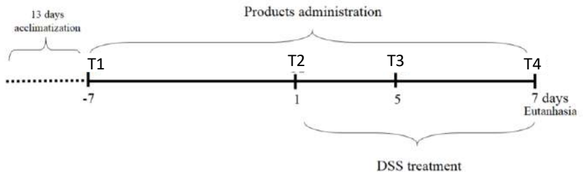


Figure 1. Colitis induction and products administration during the experimental protocol.

We induced UC in mice by a treatment with dextran sulfate sodium (DSS). Male C57BL/6J mice (n = 10 per group) colitis induction and were divided in: **H group** - healthy mice; **C group** - DSS mice that did not receive WP; **W group** - DSS mice fed with WP without UV-C treatment and **WU group** - DSS mice fed with WP functionalized by UV-C treatment. WP administration started seven days before continued through the seven days of DSS induction. The severity of colitis was daily determined by disease activity index (DAI) and the fecal microbiota was analyzed in the days -7, 1, 5 and 7.

Results

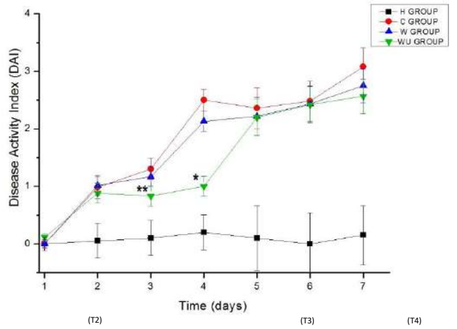


Figure 2. Disease activity index (DAI) of each group during DSS induction period.

** no statistical difference with day 1

* no statistical difference with healthy group

In all DSS-induced mice, a reduction in the alpha-diversity was observed across the experimental time. Moreover, all groups display a similar Shannon index through the experiment ($p > 0.05$).

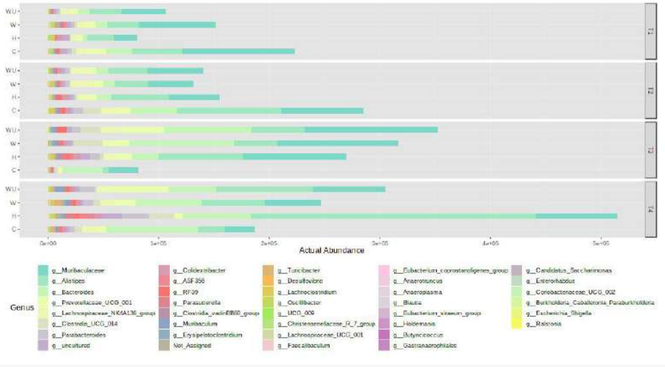


Figure 3. Frequency of microbiota ASVs between mice groups during timeline of experiment.

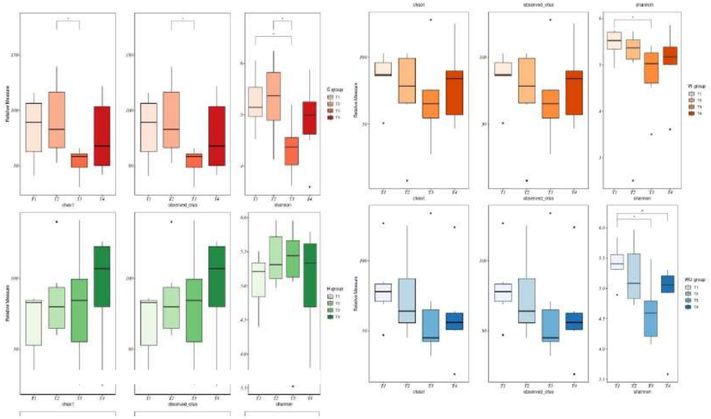


Figure 4. Box lots showing the Alpha-diversity measures in each group during timeline of experiment.

Microbiota composition was significanty affected by the treatment. A decrease in *Lachnospiraceae* both in W and WU groups ($p < 0.05$) was observed. However, at the study end (day 14) a restore in *Lachnospiraceae* was observed ($p < 0.05$). Finally, *Oscillospiraceae* (associated with H group) and *Turicibacter* (W group), increased during the experiment. Mice in WU group showed a reduction in the characteristic symptoms of colitis (diarrhea, weight loss, and presence of blood in the feces), without differing from the H group on day 11 of the protocol ($p<0.05$).

Conclusions

The results indicate that the different WP production could result in a restoration in beneficial taxa and slower symptom progression in colitis mouse model. Although future studies are needed to confirm these beneficial effects.

Acknowledgment: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Finance Code 001 and FAPESP Grant number 2017/01189-0.

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

Xochitl C. Morgan[1], Lauren J. McIver[1], Thomas M. 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

HCMPH MICROBIOME CENTER

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.
This includes 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.

Support fully-collaborative grant-funded investigations.
Includes preliminary data development, 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

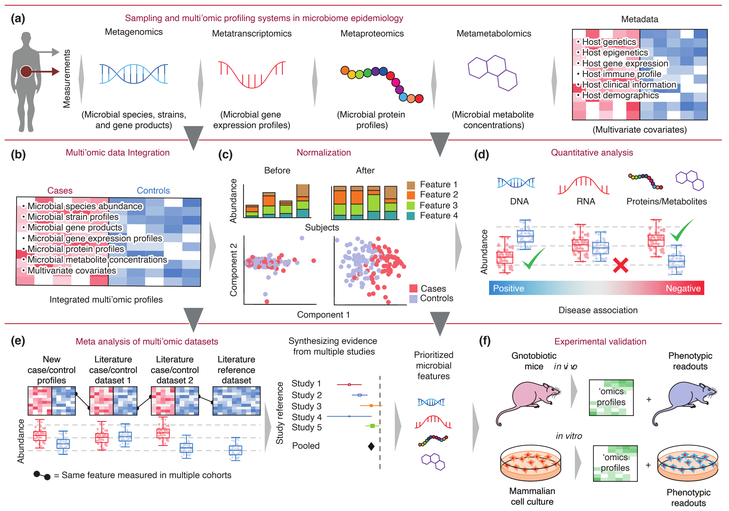
Analysis
- Bioinformatics (raw data processing, taxonomic and functional profiling)
- Downstream analysis and statistics

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

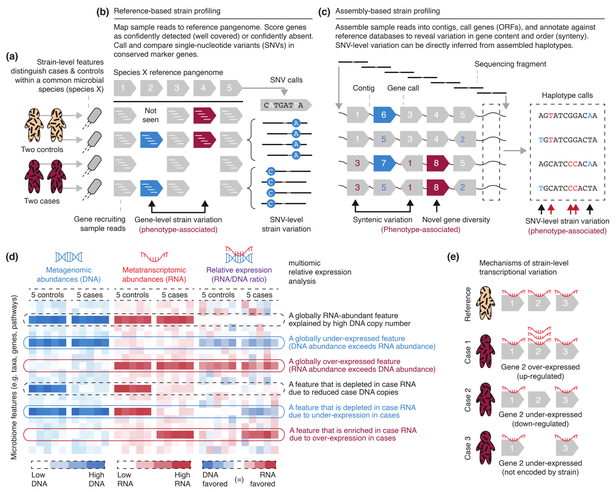
HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH
Microbiome Analysis Core

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**

Microbial multi'omics



The Harvard Chan Microbiome Analysis Core supports microbiome analysis for a variety of molecular data types in human populations or in model systems. Typical analysis workflow steps include **a)** molecular data generation of a variety of types, including but not limited to sequencing, which are **b)** bioinformatically processed into biologically interpretable features and **c)** quality controlled per dataset. This permits **d)** microbiome-tailored statistical methods to associate molecular features with covariates and outcomes, and optionally **e)** meta-analysis of multiple data types per project or across multiple projects. Finally, **f)** the Core can assist with study design for downstream evaluation of statistical associations in in vivo or in vitro model systems.



Shotgun metagenomic and metatranscriptomic sequence data are particularly amenable to detailed computational analysis, including multiple complementary methods for **a)** strain tracking or differential microbial expression. **b)** Reference-based methods can identify strains using either single nucleotide or structural (genomic) variants, and **c)** can be used in tandem with assembly-based methods for novel microbial discovery. **d)** Whole-community microbial differential expression can additionally be detected either in tandem with or in addition to metagenomic copy number changes, and **e)** analyzed per gene, pathway, microbe, or human individual.

Mallick, H. et al. Experimental design and quantitative analysis of microbial community multiomics. Genome Biology. 18:228 (2017).

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 (AnADAMA2). This includes identifying microbial species (MetaPhlAn2) and strains (PanPhlAn/StrainPhlAn), characterizing their functional potential or activity (HUMAnN3, ShortBRED), and priotization of bioactive genes (MetaWIBELE) or metabolites (MACARRoN).

KneadData — Perform quality control of meta'omic reads, including host read removal

MetaPhlAn — Profile microbial clades and their abundances

StrainPhlAn — Identify and track microbial strains using SNPs in conserved, unique marker genes

HUMAnN — Profile microbial genes and pathways; stratify by contributing species

PICRUSt — Predict metagenome functional content from marker genes

ShortBRED — Profile proteins of interest with very high specificity

MetaWIBELE — Identify and prioritize bioactive gene products

MelonnPan — Predict metabolome content from microbial enzyme abundances

PhyloPhlAn — Reconstruct microbial phylogenetic trees

MACARRoN — Identify and prioritize bioactive metabolites

AnADAMA — Perform and document automated scientific workflows

bioBakery — A meta'omic analysis environment

Downstream analysis and statistics

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

SparseDOSSA — A hierarchical model of microbial ecological population structure

ARepA — Extract and normalize 'omics data from online repositories

MMUPHin — Correct batch effects, meta-analyze microbes, genes, and pathways across multiple studies

GraPhlAn — Generate cladograms and decorate with metadata

LEfSe — Associate up to two metadata with microbiome features

MaAsLin — Associate arbitrarily complex metadata with microbiome features

CCREPE — Assess the significance of similarity measures in compositional data

HAllA — Perform well-powered comparisons of paired, high-dimensional datasets

BAnOCC — Bayesian assessment of association in compositional data

MicroPITA — Select samples for follow-up analysis in two-stage tiered studies

McIver, L. J. et al. bioBakery: a meta'omic analysis environment. Bioinformatics, 34:7, 1235-1237 (2018).

Director: Xochitl C. Morgan
Senior Software Developer: Lauren J. McIver
Postdoctoral Fellow and Data Analyst: Thomas M. Kuntz
Scientific Director: Curtis Huttenhower

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

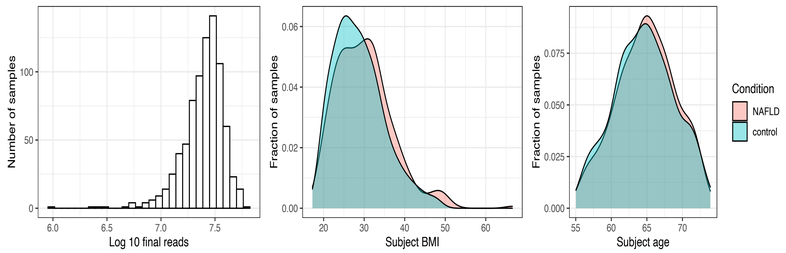
Changes in taxonomic and metabolic composition among NAFLD patients

Paul Nelson[1], Hanseul Kim[1,2,3], Kelsey N. Thompson[1], Long H. Nguyen[1,2,3], Curtis Huttenhower[1]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health [2]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School [3]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School

Non-alcoholic fatty liver disease (NAFLD), which is strongly associated with obesity and type 2 diabetes, affects up to 25% of the adult population in the US. Human and mouse studies have suggested gut microbiome as a causal factor in the pathogenesis of NAFLD. To investigate the role of the gut microbiome in NAFLD, we examined metagenomics and metabolomics from 211 subjects with NAFLD and 457 healthy controls from the Nurses' Health Study II. We found that NAFLD explains a significant, but relatively small (>1%), amount of the taxonomic and metabolic variability in the microbiome. Despite this relatively weak signal, we identified several metabolites and species associated with the disease phenotype. Specifically, we highlight the role of the microbially produced bile acid, isoallolithocholic acid (isoalloLCA) in NAFLD. Our results link together previous research on microbial manipulation of Treg and Th17 immune cells, and the role of the immune system in NAFLD. All together this study represents the largest multi-omics study to date of the microbiome in NAFLD.

Participants characteristics:



We adjusted for the most recent diabetes diagnosis, cumulative average of body mass index (BMI), cumulative average of Alternative Healthy Eating Index (AHEI), and cumulative average of physical activity.

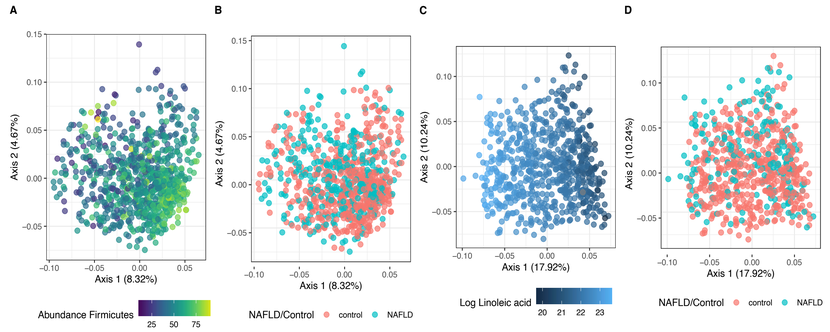
	NAFLD		
	Control	Case	P-value
n	457	211	
Age (mean (SD))	62.2 (4.2)	62.2 (3.9)	0.90
Hypertension (%)	118 (25.8)	77 (36.5)	0.01
Hypercholesterolemia (%)	137 (30.0)	86 (40.8)	0.01
Diabetes Mellitus (%)	24 (5.3)	46 (21.8)	<0.001
Pack-Years (mean (SD))	12.8 (10.1)	14.7 (11.1)	0.21
Physical activity (mean (SD))	24.0 (19.3)	17.1 (15.8)	<0.001
BMI (mean (SD))	27.3 (5.5)	30.7 (6.1)	<0.001
Postmenopausal Hormone Use (%)			0.60
Current User	100 (21.9)	40 (19.0)	
Never Used	197 (43.1)	83 (39.3)	
Past User	122 (26.7)	64 (30.3)	
Aspirin Use (%)			0.07
Current user	158 (34.6)	89 (42.2)	
Nonuser	299 (65.4)	122 (57.8)	
AHEI (mean (SD))	58.5 (10.0)	56.1 (8.9)	0.004
Total Calories (mean (SD))	1815.6 (452.8)	1876.1 (469.0)	0.11

Data generation:

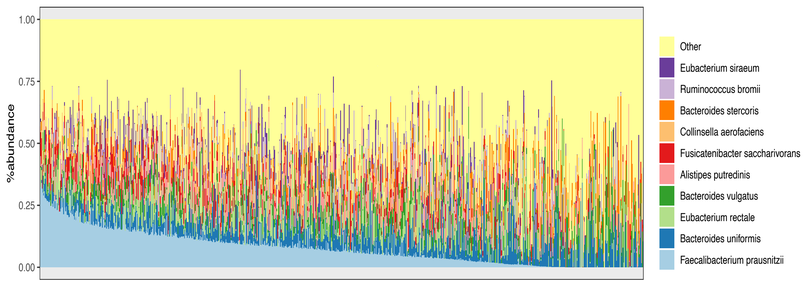
Metagenomic data was generated by the Broad Institute were using High-Output Microbial WGS, which yields 12M 2 x 150bp reads. Sequences were processed using Biobakery Workflows v3 in Nov 2021 using default parameters. Sequences were trimmed and decontaminated using KneadData, yielding and average of microbial 7.5M reads per sample (Fig 1C). Taxonomic composition was then determined with MetaPhlAn 3.0 and functional profiles were generated in HUMAnN 3.0.Metabolites were profiled using hydrophilic interaction chromatography (HILIC) and C18 HPLC columns, in both positive and negative ion mode each. Nontargeted spectra were processed using Progenesis CoMet software (v 2.0, Nonlinear Dynamics) to detect and de-isotope peaks, perform chromatographic retention time alignment, and integrate peak areas. Peaks of unknown ID were tracked by method, m/z and retention time (RT). Metabolite identification was conducted by matching measured mass and RT to reference metabolites matching an internal database of >600 characterized compounds. 10 samples were chosen as internal technical replicates and processed twice each; metabolite abundances for these 10 samples were averaged across replicates.

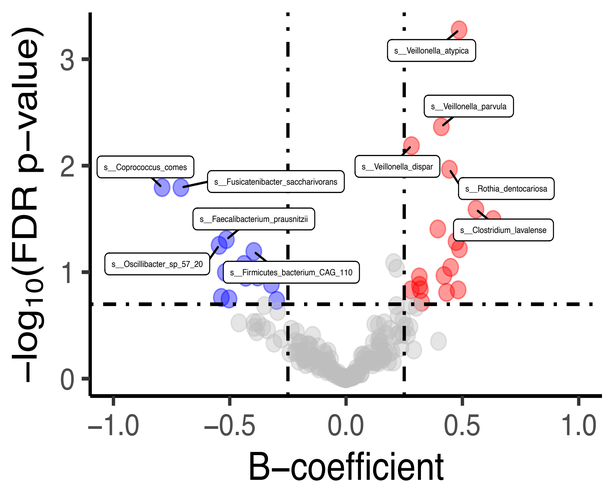
Ordination analyses:

PCoA shows patterns characteristic of human stool microbiome, such as a Firmicutes gradient along the first principal component of species composition (A). However, we see very little differentiation between NAFLD and healthy subjects for both species and metabolites (B,C)



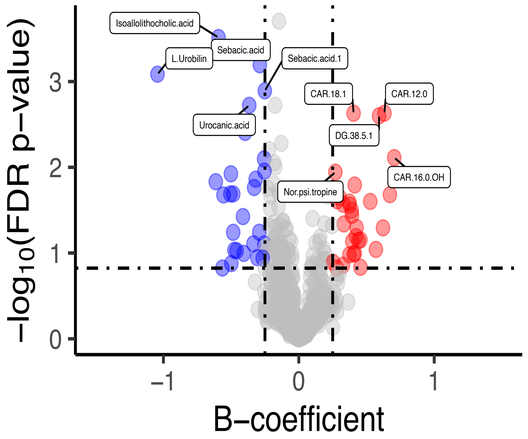
Taxonomic changes in NAFLD:





33 species differed between NAFLD and healthy controls. A permanova of Bray-Curtis similarity revealed that NAFLD explained a small (R^2 = 0.005), but significant ($p < 0.001$) amount of variation in microbial species abundances.

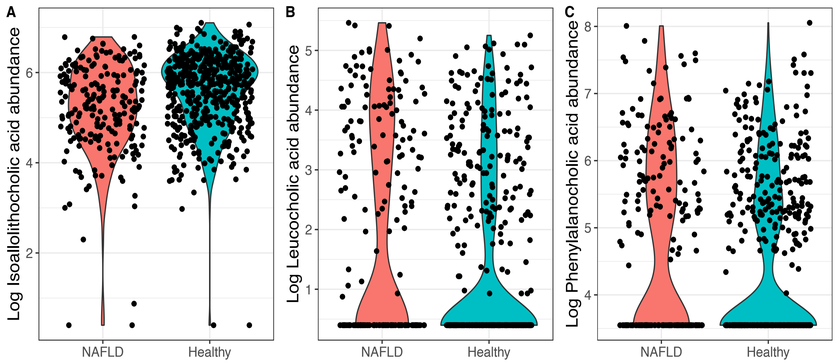
Metabolic changes in NAFLD:



We observed changes in several bile acids, fatty acids, and increases in acylcarnitines in NAFLD. Because bile acids are known to be involved in NAFLD, we pay special attention to this group.

Acylcarnitines make up the largest class of the compounds higher in subjects with NAFLD. Increases in acylcarnitines in response to steatosis and liver fibrosis has been noted previously in plasma.

Violin plots of the abundance of the secondary bile acids isalloLCA (A), phenylalanocholic acid (PheCA, B), and leucocholic acid (LeuCA, C) stratified by NAFLD.



References:

Enooku, K., Nakagawa, H., Fujiwara, N., Kondo, M., Minami, T., Hoshida, Y., Shibahara, J., Tateishi, R. and Koike, K., 2019. Altered serum acylcarnitine profile is associated with the status of nonalcoholic fatty liver disease (NAFLD) and NAFLD-related hepatocellular carcinoma. Scientific reports, 9(1), pp.1-9.

Gottlieb, A. and Canbay, A., 2019. Why bile acids are so important in non-alcoholic fatty liver disease (NAFLD) progression. Cells, 8(11), p.1358.

Fei N, Bruneau A, Zhang X, et al. Endotoxin Producers Overgrowing in Human Gut Microbiota as the Causative Agents for Nonalcoholic Fatty Liver Disease. mBio 2020; 11.

Burz SD, Monnoye M, Philippe C, et al. Fecal Microbiota Transplant from Human to Mice Gives Insights into the Role of the Gut Microbiota in Non-Alcoholic Fatty Liver Disease (NAFLD). Microorganisms 2021; 9.

Alexander M, Ang QY, Nayak RR, et al. Human gut bacterial metabolism drives Th17 activation and colitis. Cell Host Microbe 2022; 30: 17-30.e9.

Świderska M, Jaroszewicz J, Stawicka A, Parfieniuk-Kowerda A, Chabowski A, Flisiak R. The interplay between Th17 and T-regulatory responses as well as adipokines in the progression of non-alcoholic fatty liver disease. Clin Exp Hepatol 2017; 3: 127-34.

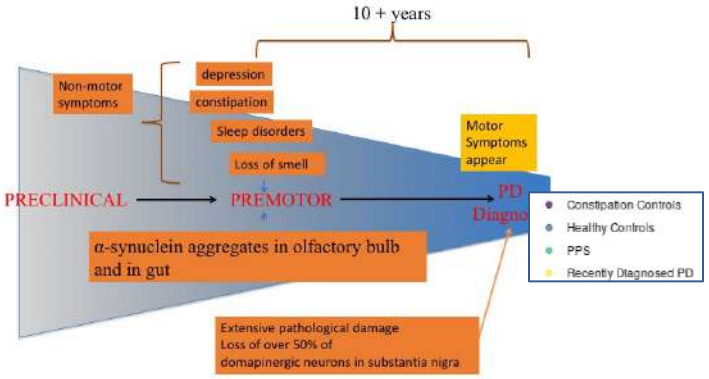
Acknowledgements:

We would like to thank the Massachusetts Life Sciences Center for funding this project as well as the Huttenhower lab and support from Empress therapeutics, especially the computational team and Boris Hayete.

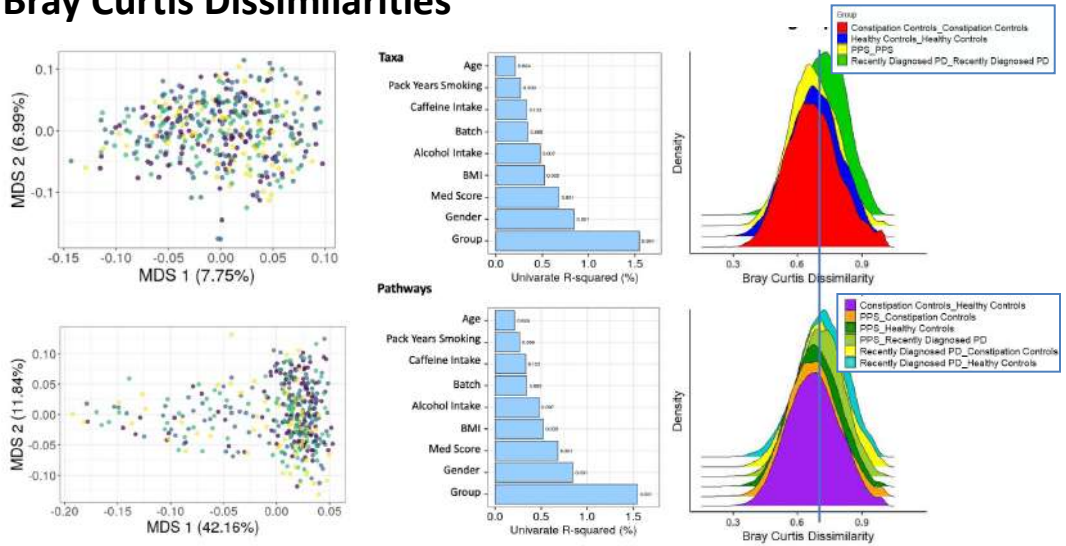
The Gut Microbiome and Parkinson's Disease

Natalia Palacios, Jeremy Wilkinson, Kjetil Bjornevik, Michael Schwarzschild, Lauren McIver, Alberto Ascherio, Curtis Huttenhower

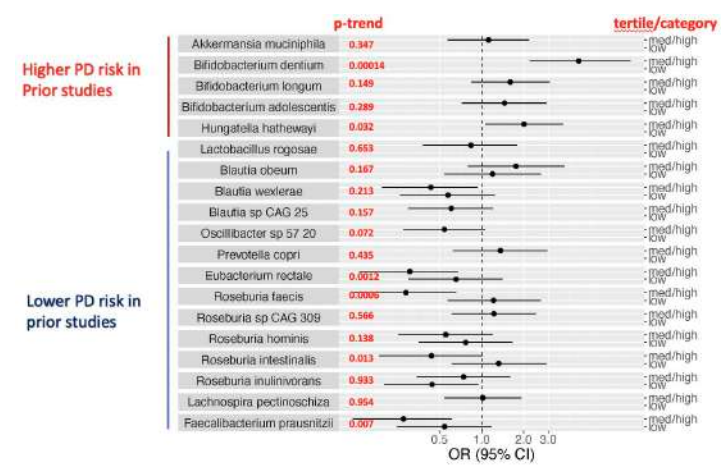
Prodromal Parkinson Disease



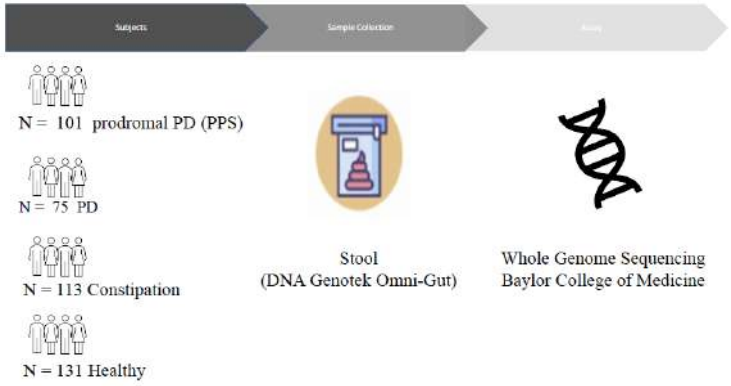
Bray Curtis Dissimilarities



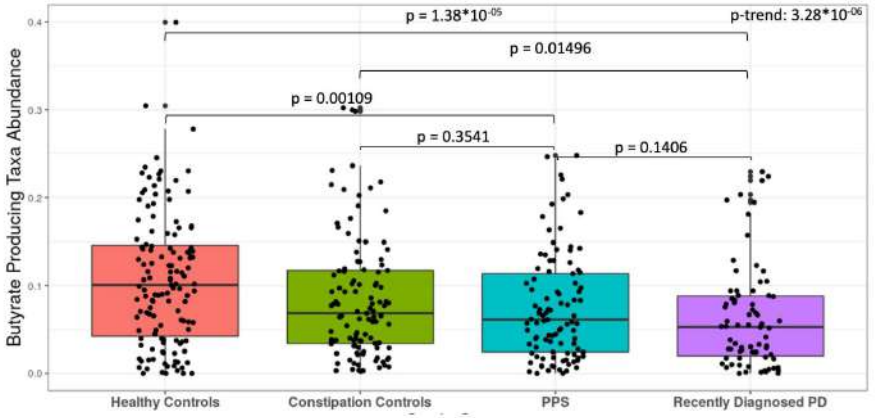
OR, PD vs. Control for a-prior species



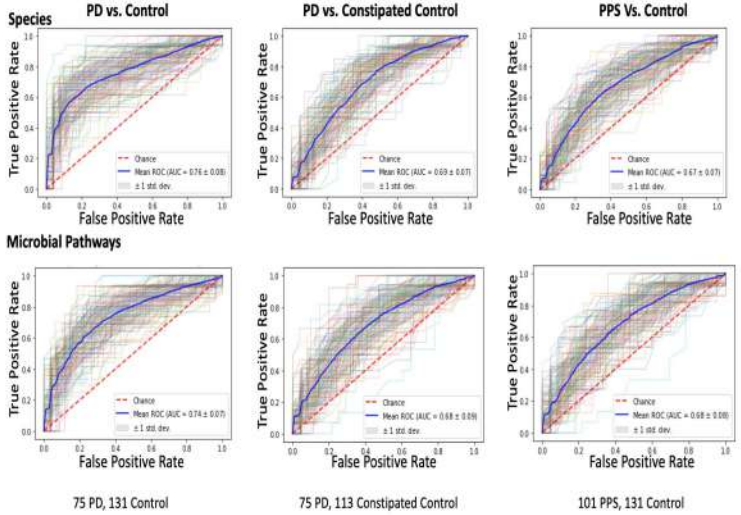
Study Design



Butyrate-Producing Taxa



Random Forest Classifier



Funding: NINDS R01NS097723 PI: Natalia Palacios

natalia_palacios@uml.edu

Introduction & Methods

Hypothesis
Limiting the consumption of three specific simple sugar groups (fructose, oligo-saccharides, polyols) to a cumulative total of less than 40 grams per day per individual will improve the symptoms of *Acne Vulgaris* (AV)

Objectives
1. To observe the effect of an exclusion diet on cases with *Acne Vulgaris* (AV) over a period of 15 days
2. To compare differences in stool microbiota of cases before and after a dietary intervention with symptom relief

Limiting fructose, oligosaccharides & polyol sugar groups
Malabsorption of three simple sugar groups were thought to trigger a rapid inflammatory effect appearing as AV but potentially moderated by limiting simple sugars in the diet.

While the microbiome can be modified within 3-5 days (David 2013) a 15-day elimination diet was applied to a small adult AV cohort in order to allow sufficient time both to adjust the gut microbiome and to observe the effect on acne severity. A comparison of DNA sequenced stool samples at baseline and endpoint was used to identify common trends in this cohort.

The usual diet for Lao people is already low in the target sugars under review suggesting better adherence, while adult AV candidates were chosen for their reliability and expected adherence rather than a teenage cohort. Skin microbiome or specific blood bio-markers were not examined and no other intervention (e.g. pharmaceutical / OTC preparation) formed part of the study.

A local restaurant was contracted to deliver a meal plan that conformed with the study requirements, supplying three meals/snacks per day/per case. Cases collected meals in-person every two days from the restaurant where for each visit the Co-PI observed their skin and asked each case a standard checklist of questions in relation to dietary adherence and bathroom frequency.

picture booklet designed for project
Foods to enjoy vs. Foods to avoid

AV severity evaluation
Observations of acne severity at baseline (Day 1) and at endpoint (Day 15) were agreed by two independent dermatologists using the *Investigator Global Assessment* (IGA scale 0-5) for a total of 27 AV cases (Dreno 2011). Photographs were taken at baseline and endpoint of each case by a professional photographer.

While the microbiome can be expected to be modified within 3-5 days (David 2013) the 15-day timeframe allows for both for AV to be observed, as well as allowing an expectation for high dietary adherence. DNA sequencing of the microbiome from stool samples at baseline and endpoint were retrieved to establish if a statistical significant change in gut microbes could be matched across the cohort with AV improvement.

Microbiome analysis
DNA sequencing of stool microbiome was completed at baseline and endpoint for every case by Diversigen Laboratories MN USA (https://www.diversigen.com)

Intestinal parasites
Considerable effort was made to ensure cases were clear of intestinal parasites as prior (unpublished) work by co-PI indicated parasites appeared to increase malabsorption of target sugars in AV cases.

Infection or re-infection of intestinal parasites may be due to non-adherence, however to limit the risk of false-negative results, parasite exams were cross-checked by using two separate laboratories.

Included
summary

polished rice
rice noodles
meat; fish; poultry
eggs; dairy products
low-fat UHT milk

fiber was increased
(oats/nuts/vegetables)

no restrictions on
glycemic load

Excluded
summary

fruit; fruit juice
fruit products
honey
added fructose
HFCS / FOS
wheat products
soy-bean oil
onions; garlic
mushrooms
beans/lentils
sorbitol, xylitol

Pilot Study

An exclusion diet to examine the effect on Acne Vulgaris (AV)

limiting fructose, oligosaccharides & polyol sugar groups for 15 days

Marie Ryan M.Med.Sc (Epi) M.Man (Health) & Dr. Laurent Ferradini M.D (Derm) PhD (Immun)
in cooperation with the Lao National Center for Dermatology & Venereology (NDC)
Director, Dr. Ammala Philavanh

Methods

Study Type Non-randomized, prospective observational dietary study

Inclusion Criteria
- Adults >18 years with persistent moderate-severe facial acne for the previous five years
- Resident in Vientiane within the radius of supplied map for 6-8 weeks
- Must be contactable by smart phone, and be registered users of WhatsApp
- Must have completed their food diaries from pre-inclusion interview to Day 1

Exclusion Criteria
- Use of oral isotretinoin, or antibiotics in the past three months
- Persistent intestinal parasite infection
- Diagnosed food allergy or intolerance, celiac, Crohn's disease, ulcerative colitis
- Veganism, vegetarianism or other restrictive diet
- Current consumption of soft drinks containing fructose that exceeds 600ml per day
- Intended continued use of complimentary medicines, supplements or anabolic steroids
- *Other:* Diabetes; PCOS; pregnancy; current diagnosis of dengue fever, chikunguyna fever, malaria; alcohol intake more than three standard drinks per day; illicit drug use; physical or mental illness

Study population A total of three small groups comprising 27 adult cases were recruited in Vientiane, Laos (October 2019-June 2020) via advertising on social media and word-of-mouth. Interested candidates forwarded a 'selfie' of their skin condition via WhatsApp messenger and a follow-up phone call was made with screening questions according to the inclusion/exclusion criteria.

Pre-enrolment evaluation Candidates were requested to record their dietary habits leading up to the pre-inclusion group information session, where the *Informed Consent form was signed* at the Lao *National Center for Dermatology and Venereology* (NDC). At pre-inclusion, candidates provided a blood sample for general FBC testing, and stool for a parasite exam. Weight, height and blood pressure were also recorded.
Cases were confirmed but were not given specific details regarding foods included/excluded in the study.
Day 1 a second stool sample was obtained both for parasite exam and also for DNA sequencing. Cases were asked not to change their usual skin regimen and to supply photographs of all products used on their skin prior to/during the intervention.

Exclusion diet implementation
Day 1 cases were photographed and received the baseline skin assessment (IGA) by two dermatologists. Detailed information about the study was provided Day 1 with Q&A. Any additional information throughout the intervention was sent to all cases via group chat using WhatsApp messenger so as not to bias or influence individual behaviors. However, all cases were expected to refer to detailed printed information and food picture book.

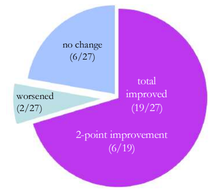
Exclusion diet The same meals/snacks were collected by cases every two days. The co-PI was present for each collection to monitor cases. A standard set of questions was asked of each case regards their dietary adherence and bathroom frequency. A review was made of the completeness of individual diet diaries along with a record of skin condition. Self-reported adherence was around ~90% although infection by intestinal parasites by five cases during the intervention would suggest reported adherence was likely lower at least for these cases.

Parasites diagnosis and treatment A parasite exam was applied in the week prior to Day 1, again on Day 1 and lastly on Day 15, with appropriate medication as necessary. Administration of *Albendazole* (for example) has a one day wash out and was not expected to alter the gut microbiome[1]

Stool sampling and DNA sequencing Stool samples on Day 1 and Day 15 were transferred using Diversigen test kits and stored according to manufacturers specifications before being shipped all together by courier on Day 15 to Diversigen Laboratories (MN USA)

Results

AV clinical score evolution at Day 15
Investigator Global Assessment (IGA) scores (Dreno 2011) of the AV enrolled patients (n=27) made in agreement by two dermatologists at baseline and endpoint at Day 15 of elimination diet.



A clinically meaningful outcome[2] of at least a two-point (IGA) improvement from baseline (USA-FDA guidelines) **was achieved for 6/27 cases with 19/27 cases showed any improvement while 6/27 showed no change in IGA score and 2/27 worsened .**

While the time for AV lesions to resolve completely can be in the range of 4-6 weeks, this short 15-day intervention had a positive effect on the IGA scores improving both skin quality and self-confidence for 19/27 cases.

Other outcomes of the diet
- Weight loss was observed for 20/27 cases (minus 1-4kg) while weight remained stable for 7/27 cases.
- Cases reported that within few days of the intervention the texture of their skin had improved with reduced oiliness.
- Cases also reported improved intestinal transit with more frequent and easier bathroom visits.
- Constipation is known to be a symptom of simple sugar malabsorption and can be further exacerbated by parasitic infection like ascariasis, taeniasis or *Opisthorchis* species.

Indeed, parasitic infection was found to be clearly associated to AV lesion worsening in one case (below) with parasitic re-infection during the diet (suggesting diet non-adherence) since anti-parasitic treatment at day 15 and strict self-compliance with the diet led to considerable improvement of AV at 6 weeks post-intervention.

Parasite infection/re-infection was suspected to be from consumption of snacks or low-cost lunches (often papaya salad or fermented fish paste) purchased from street vendors.

DNA analysis
- DNA was isolated from fecal samples for 26 cases (24F, 2M) and 4 controls by *Shallow Shotgun Metagenomic Sequencing* (SSMS).
- Comparison of the relative abundances of operational taxonomic units (OTU) for 26 cases at baseline and endpoint (15-day exclusion diet) is underway

Case characteristics

cases	gender	average age	weight loss (1-4 kg)
n=27	2M 25F	29 (20-48)	20/27

Self-reported
- no new acne if fully adherent to diet
- lesion(s) within two hours when non-adherent
- smoother skin
- less oily skin
- reduced pore size
- easier/more regular bathroom visits

Conclusions

- This exclusion diet limiting simple sugars (<40grams total/day) had a positive clinical impact on AV evolution
- This exclusion diet also resulted in weight loss and improved frequency and ease of bowel movements
- Intestinal parasitic infection appeared to worsen AV lesions probably through further malabsorption of simple sugars.
- The exclusion diet also induced weight loss (minus 1-4 kilograms) during the intervention

The mechanism of action of the exclusion diet is unknown but was able to modify the intestinal microbiome. For this reason, a systematic analysis of bacterial DNA at baseline and after the diet intervention is underway to understand to which extend the intestinal microbiome could have been modified by the diet and how this correlates with AV improvement.

Limitations
- Study group was a pilot study on a limited size cohort, a larger cohort study would be needed to confirm such preliminary data
- SARS-CoV-2 disrupted recruitment from July 2021

Financing is being sought
to resume & expand this important study in 2022/23
please contact AsiaSkinProject.com for partnership opportunities

Asia Skin Project ©
www.AsiaSkinProject.com
Vientiane, Laos
May 2022

This study was undertaken in cooperation with *Lao National Center for Dermatology & Venereology* (NDC).

Ethics approval was obtained from the
Lao National Ethics Committee for Health Research (No.075/NECHR)

Acknowledgements
Dr. Monivong Sisavath fmr. Dir NDC

Bibliography

1956 Chambers R.A & Pratt R.T.C Idiosyncrasy to Fructose. The Lancet Aug 1956: DOI:10.1016/S0140-6736(56)92196-1
2009 Raj et al: *Acne prevention* 2009 Indian jour derm & veren Vol. 75, No. S1, 2009, pp. 61-61
2010 Gibson P.R & Shepherd S.J: Evidence-based dietary management of functional gastrointestinal symptoms: The FODMAP approach. Journal of Gastroenterology and Hepatology 25 (2010) 252–258: DOI:10.1111/j.1440-1746.2009.06149.x
2011 Dreno et al: Development and evaluation of a Global Acne Severity Scale (GEA Scale) suitable for France and Europe Journal of the European Academy of Dermatology and Venereology (JEADV) Vol.25: Pg.45. DOI: 10.1111/j.1468-3083.2010.03685.x
2012 Zarchi K. & Jemec G.: Severity Assessment and Outcome Measures in Acne Vulgaris. Curr Derm Rep (2012) 1:131–136 DOI:10.1007/s13671-012-0016-8
2012 Reynolds .R et al: Effect of the Glycemic Index of Carbohydrates on Acne vulgaris. Nutrients 2010, 2, 1060-1072; DOI:10.3390/nu2101060
2013 David L. et al: Diet rapidly and reproducibly alters the human gut microbiome Nature: Letter|Research (2013) DOI:10.1038/nature12820
2018 Informed Consent form adapted from WHO template. *Informed Consent Form Template for Clinical Studies*. WHO Research Ethics Review Committee (WHO-ERC) http://www.who.int/rpc/research_ethics/informed_consent/en/ (downloaded 25 May 2018)
1 https://pubchem.ncbi.nlm.nih.gov/compound/albendazole#section=Metabolism-Metabolites
2 https://www.fda.gov/media/71152

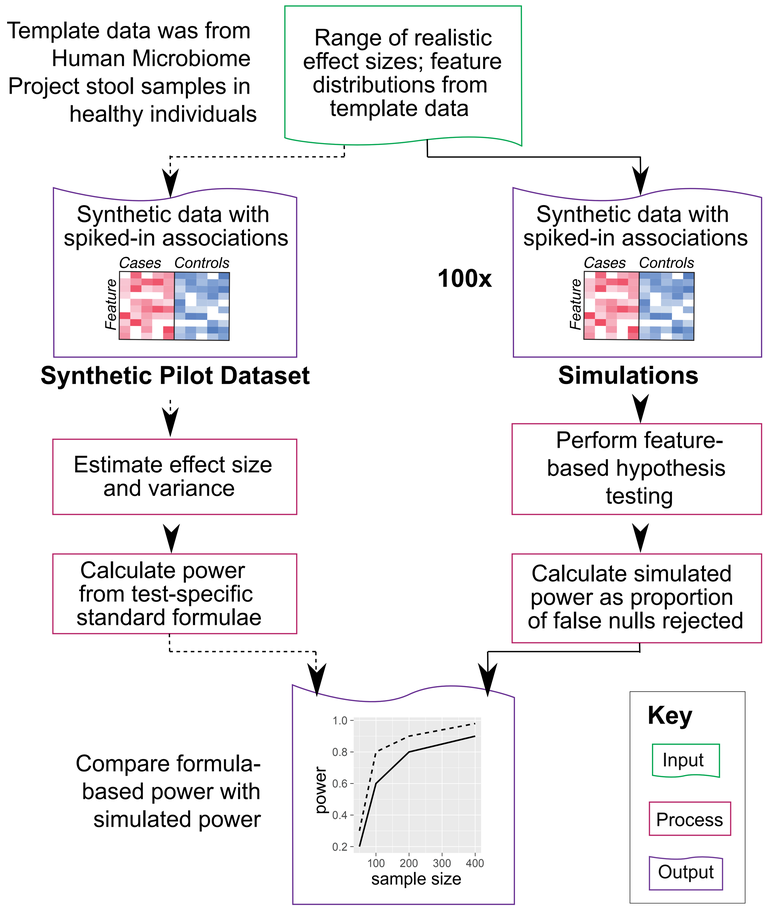
Sample size calculation for differential abundance tests in microbiome epidemiology

Meghan I. Short[1,2,3], Emma Schwager[1,2,3], Siyuan Ma[1,2,3], Paulo Manrique[1], Lauren McIver[1,2,3], Jeremy E. Wilkinson[1,3], Eric A. Franzosa[1,2,3], Curtis Huttenhower[1,2,3]

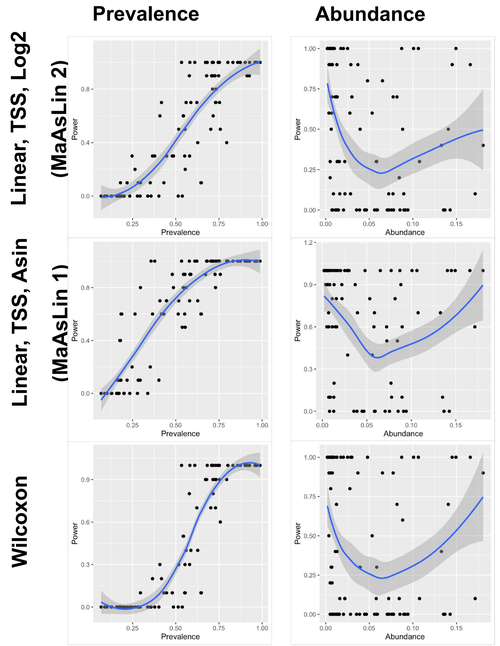
[1]Harvard T.H. Chan School of Public Health [2]Broad Institute of Harvard and MIT [3]The Harvard Chan Microbiome in Public Health Center

Tests of feature differential abundance are a cornerstone of microbiome analysis, and a wide variety of such tests is available. Reliable methods for power and sample size estimation for such tests, however, are lacking. Traditional parametric or rank-based sample size formulae do not account for the unique challenges posed by microbial feature data, including an abundance of biological and technical zero values, compositionality, and the potential for associations of clinical or environmental variables with feature abundance and/or prevalence. To benchmark existing power formulae, we use a rich simulation framework previously implemented in SparseDOSSA2 to fit zero-inflated log-linear models to microbial read counts and generate realistic synthetic feature tables. By simulating many feature tables with the same underlying distributions, we estimate power for various scenarios. We identified strong relationships between power and feature prevalence, which is unaccounted for in standard formulae for parametric tests. Use of an "effective" sample size accounting for feature prevalence improved power calculation accuracy (i.e., similarity to simulated power) in these cases. We plan to streamline best approaches in a new software, Sample Sizes for Microbiome Research (SSMoRe), which we will validate using resampling techniques in previously published data from 16S and metagenomic studies.

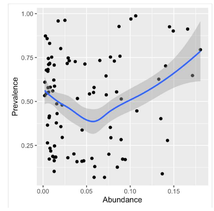
Benchmarking power formulae against simulated datasets



Power correlates with feature prevalence, less with abundance



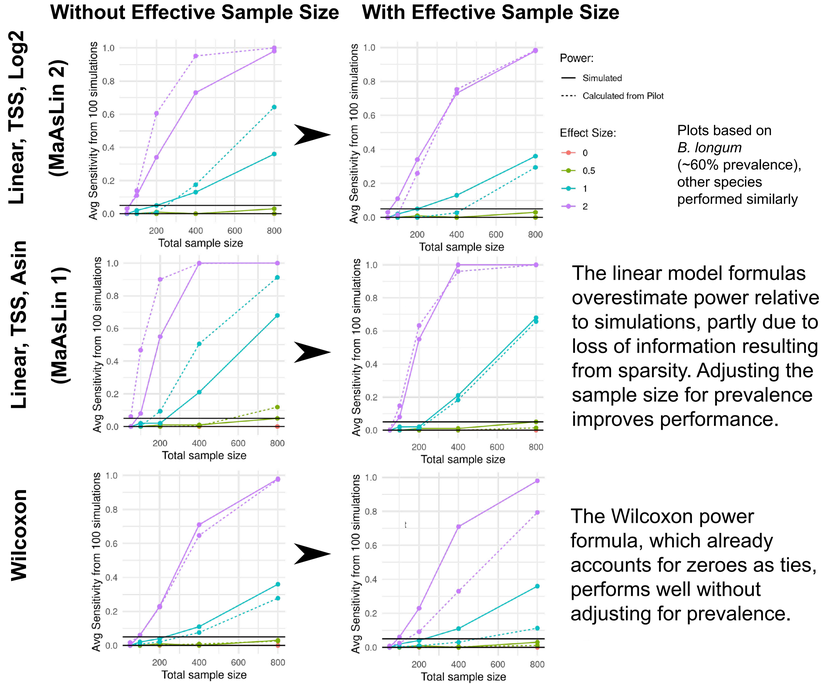
Abundance-Power relationships resemble Abundance-Prevalence relationships:



Therefore, we focused on prevalence as a target to improve power estimation.

We considered a prevalence-adjusted "Effective Sample Size":
$N_{eff} = N \times Prevalence$

Prevalence-adjusted sample size (N_{eff}) improves parametric power estimation



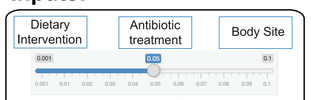
Plots based on *B. longum* (~60% prevalence), other species performed similarly

The linear model formulas overestimate power relative to simulations, partly due to loss of information resulting from sparsity. Adjusting the sample size for prevalence improves performance.

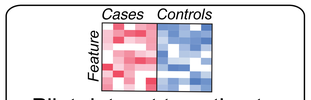
The Wilcoxon power formula, which already accounts for zeroes as ties, performs well without adjusting for prevalence.

Sample Sizes for Microbiome Research (SSMoRe) software will allow simple study planning

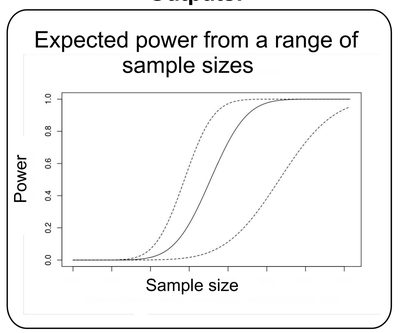
Inputs:



OR



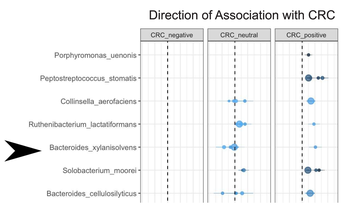
Outputs:



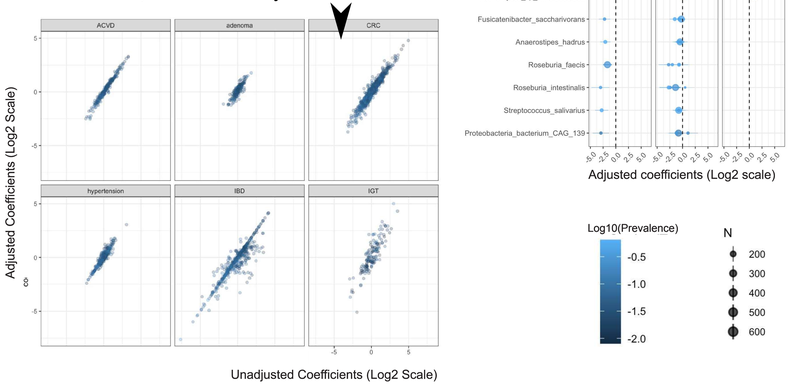
Cataloging effect sizes from published data aids study design

curatedMetagenomicData is an R package with uniformly processed metagenomic frequency data with harmonized metadata.

Effect size estimates between microbial features and health conditions such as colorectal cancer (CRC) will be incorporated into SSMoRe as references.



Information about spread and variability of effect sizes across conditions, with optional adjustment for covariates, offers flexibility to SSMoRe users.



Acknowledgments

This work has been supported by sponsored research from Hill's Pet Nutrition and NIH NIDDK R24DK110499

http://huttenhower.sph.harvard.edu

Communication Breakdown: sensing of host mucin regulates a symbiont's biogeography and inflammatory potential in the intestine


T. Jarrod Smith[1]
@tjarrods
jarrods@uoregon.edu


Deepika Sundarraman[2]
deepikas@uoregon.edu


Raghuveer Parthasarathy[2]
@RParthasarathy7


Karen Guillemin[1]
@GuilleminLab

[1]Institute of Molecular Biology, University of Oregon, Eugene, Oregon, USA
[2]Department of Physics and Materials Science Institute, University of Oregon, Eugene, Oregon, USA

National Institute of Diabetes and Digestive and Kidney Diseases

UNIVERSITY OF OREGON

META

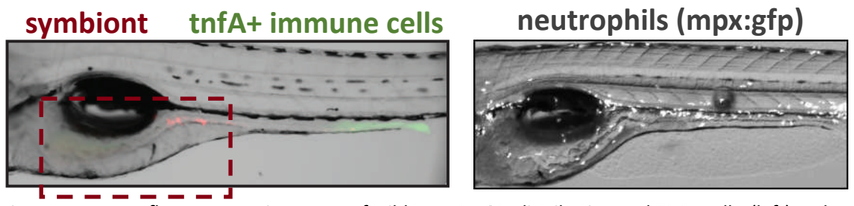
Introduction

Intestinal mucus is thought to promote host health by entangling and distancing potentially inflammatory symbiotic microbes from the gut epithelium. Here we report new evidence that beneficial microbes can self-limit their intestinal distribution and inflammatory capacity in response to intestinal mucus to actively promote host health. Using experimental evolution, live imaging of bacterial dynamics and host immune cell populations across the entire intestine, we probed the distribution, physiology, and inflammatory potential of a beneficial bacterium *Aeromonas* sp. ZOR0001 (Aer01) in its native host, the larval zebrafish. Aer01 typically forms large aggregates in a distinct intestinal region and aggregates in response to commercially available mucin in culture. Phenotypic and subsequent genomic analysis of evolved Aer01 that do not respond to mucin in culture identified a putative mucin-sensing two-component system and surface-associated mucin-binding adhesin with analogs in human gut microbiota such as *Oxalobacter*, *Akkermansia*, and *Ruminococcus* that are crucial for driving Aer01 aggregation and localization in the host intestinal environment. Disruptions in either pathway dramatically transformed Aer01 intestinal distribution and aggregative clustering, and enhanced Aer01 inflammation potential. We also found Aer01 mucin-sensing disruptions altered the community composition in defined 2- and 5- member communities. Together, our work highlights the important but largely unexplored role of microbial mucin sensory pathways in promoting host health and contributing to community composition.

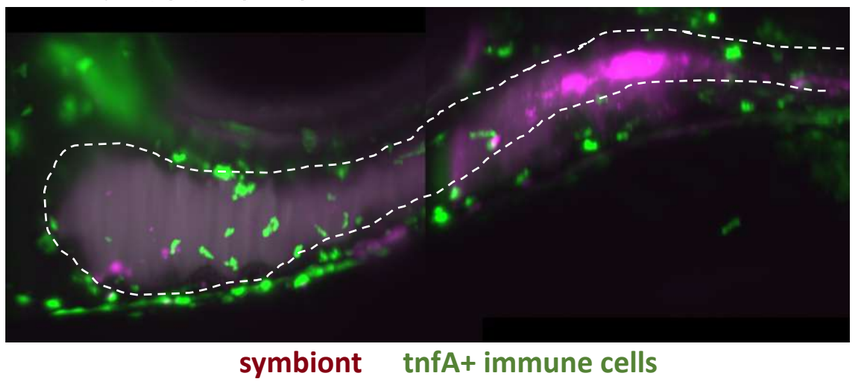
Leveraging the larval zebrafish to uncover mechanistic principles of symbiotic host-microbe interactions

live imaging of microbial populations and host immune cells

symbiont **tnfA+ immune cells** **neutrophils (mpx:gfp)**



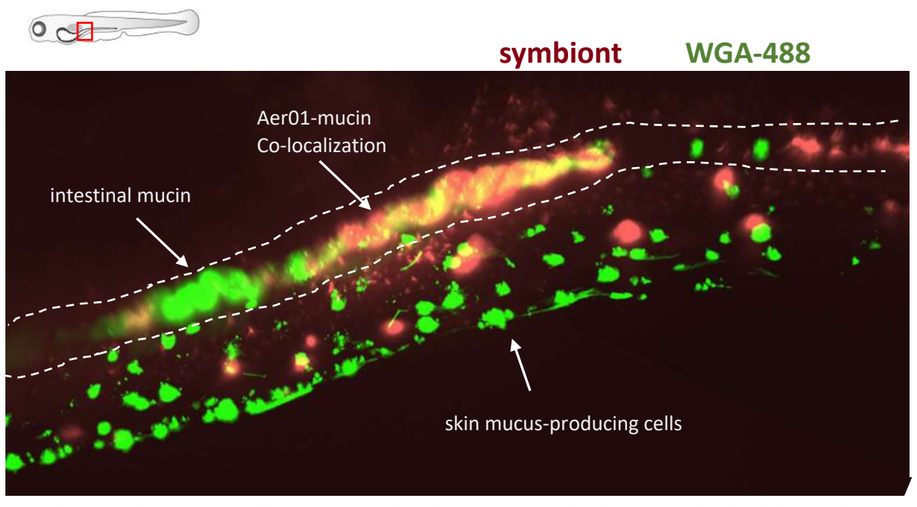
Live stereoscope fluorescent microscopy of wild type Aer01 distribution and TNF+ cells (left) and host neutrophils (right) using transgenic bacteria and zebrafish.



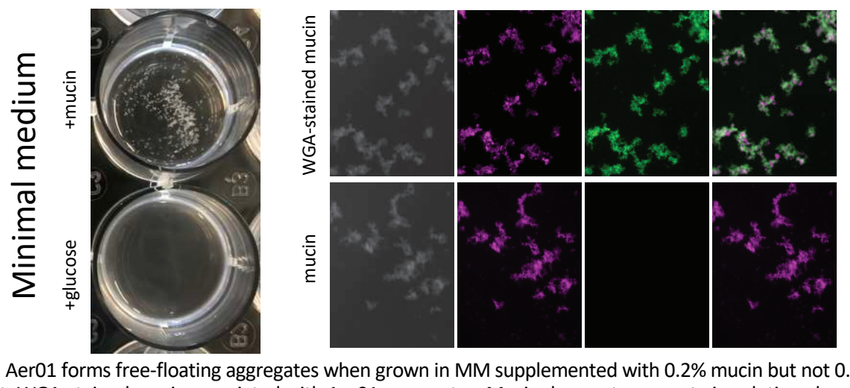
symbiont **tnfA+ immune cells**

Live light sheet fluorescent microscopy of wild type Aer01 distribution and TNF+ cells in the proximal and distal foregut highlight the typical Aer01 distribution and luminal aggregate clustering.

Live imaging reveals Aer01 co-localizes with mucin in the intestine and in culture

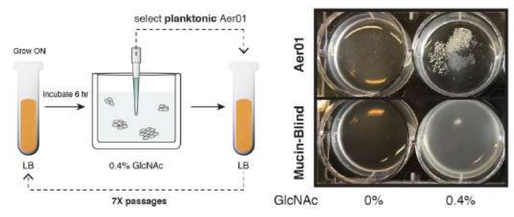


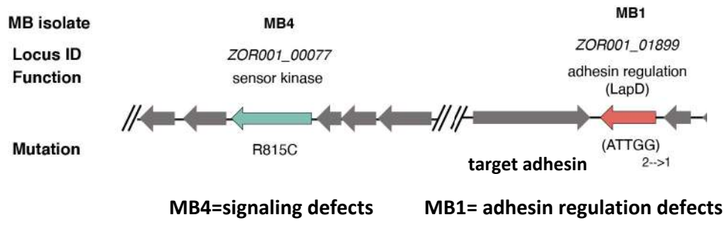
Simultaneous live imaging of host mucus and symbiont Aer01 co-localization in the distal foregut lumen using light sheet fluorescent microscopy. Image is maximum intensity projection of 98 z-stack images.



Left: Aer01 forms free-floating aggregates when grown in MM supplemented with 0.2% mucin but not 0.2% glucose.
Right: WGA-stained mucin associated with Aer01 aggregates. Mucin does not aggregate in solution alone (not shown).

Mucin-blind experimental evolution selects for mutants in environmental sensing and adhesin localization pathways

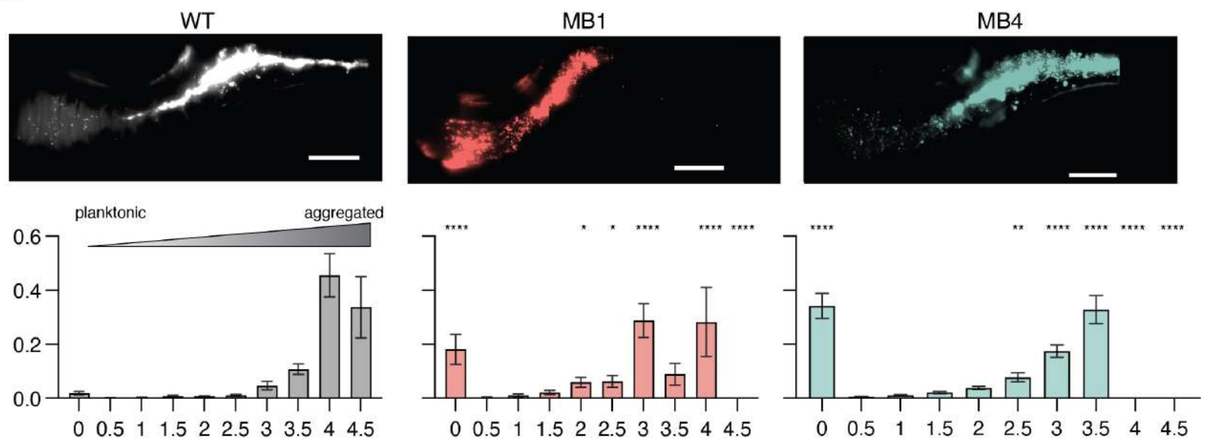




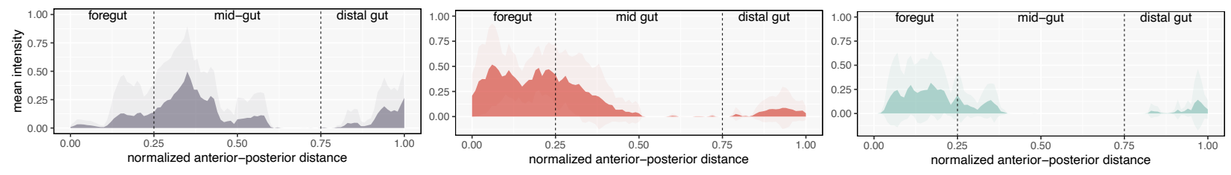
MB4=signaling defects **MB1= adhesin regulation defects**

Mucin-blind isolates exhibit altered luminal aggregation and localization in a wild-type host

MB1 and MB4 isolates are more planktonic and poorly aggregated

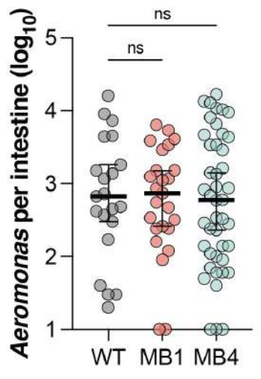


MB1 and MB4 isolates are displaced into the foregut region of the intestine

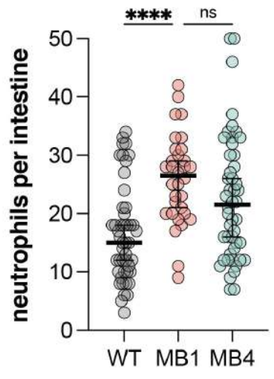


Mucin sensing determines symbiont inflammation potential

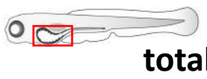
colonization



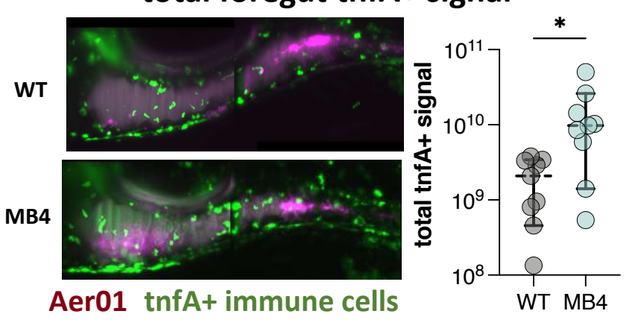
intestinal neutrophils



Aeromonas CFUs and host intestinal neutrophil abundance 24 hr after mono-association of the indicated strains



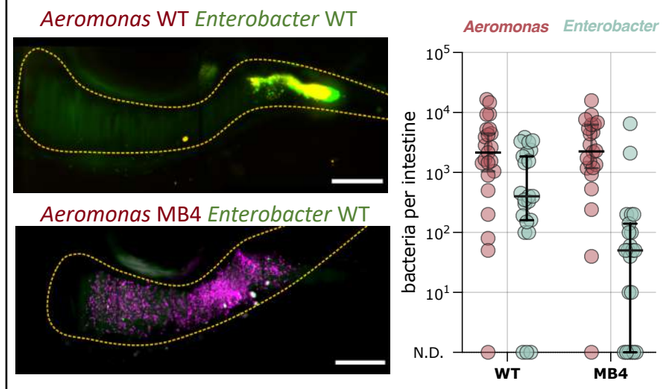
total foregut tnfA+ signal



Aeromonas CFUs and host intestinal neutrophil abundance 24 hr following mono-association of the indicated strains. Images are maximum intensity projections;

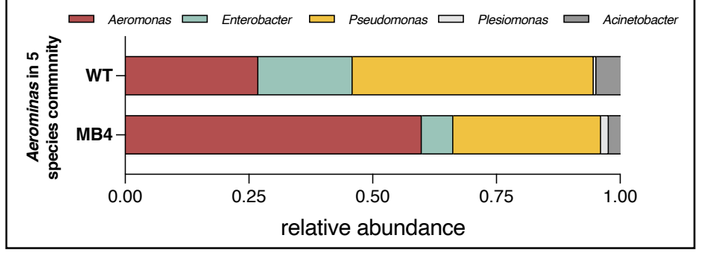
Mucin-blind isolates disrupt community composition and architecture

Two-member communities



Right: *Aeromonas* and *Enterobacter* typically co-localize in the distal foregut. MB4 antagonizes *Enterobacter* aggregation, decreasing Enterobacter abundance through yet known mechanisms. Images are maximum intensity projections.

Five-member communities



Funding

This research was supported by the National Institutes of Health under Awards 1P01GM125576 and F32DK124033.

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

Xochitl Morgan[1], Magnus Stefansson[1], Curtis Huttenhower[1,2,3]

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

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

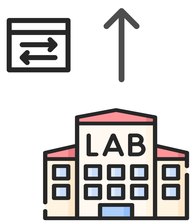
HCMPH MICROBIOME CENTER

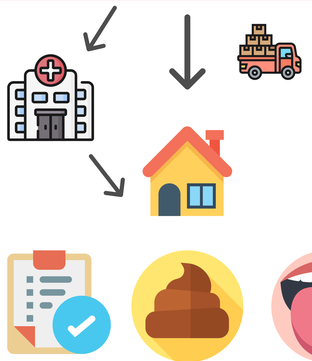
The Microbiome Collection Core at the Harvard T.H. Chan School of Public Health (HCMCC) was established in response to a strong demand among the research community for validated microbiome sample collection kit configurations and easy usability for in-home sampling. Under the umbrella of the Harvard 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





At-home sample collection

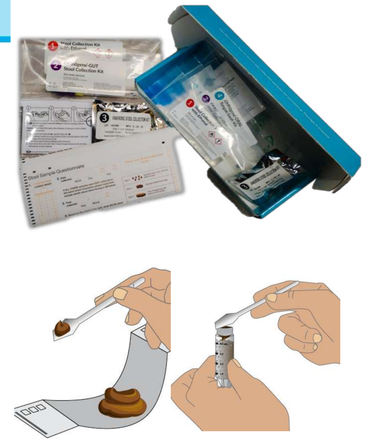
Streamlined post-collection assistance - Automated aliquoting - Barcode tracking - -80˚C storage in the BiOS Freezer - Fast sample retrieval - Sample shipment to sequencing labs for meta'omics & metabolomic profiling

Pre-paid return shipment

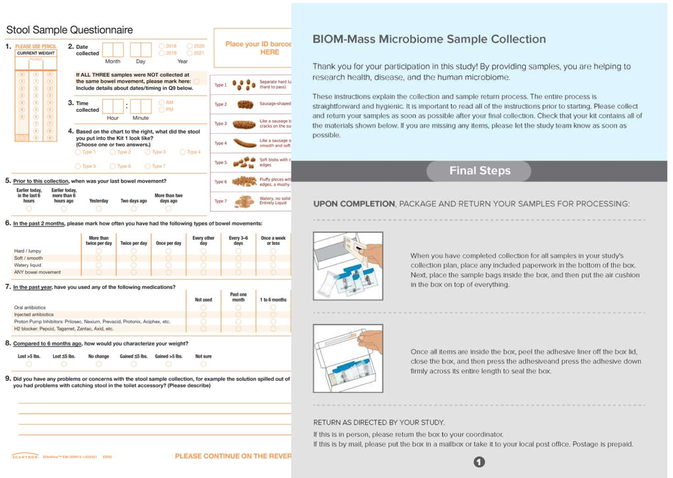
THE HARVARD CHAN MICROBIOME COLLECTION CORE

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

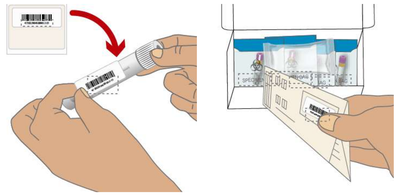


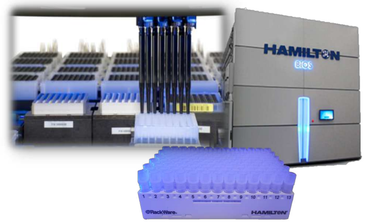
This customizable microbiome sample collection kit avoids the need for expensive, bulky, and inconvenient ice packs by providing several different room temperature storage media that are also compatible with multiple different molecular assays including **any combination of amplicon (16S), metagenomic, 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**.

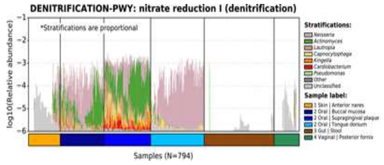


In addition to storage media, this sample collection kit includes **user-friendly instructions** and toilet accessories to maximumly facilitate and smooth the in-home stool sample collection experience. **Standardized questionnaires**, as companions to collected samples, are included to capture **recent 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 the Massachusetts Life Sciences Center (MLSC), the Harvard Chan Microbiome Platform Steering Committee, the Harvard Chan BiOS Freezer Director Eric Rimm, the BWH/Harvard Cohorts Biorepository Laboratory Manager Christine Everett.

Contact us: hcmcc@hsph.harvard.edu

Interim Managers: Xochitl Morgan & Magnus Stefansson

Microbiome in Public Health Center Strategic Manager: Magnus Stefansson

Microbiome Analysis Core Director: Xochitl Morgan

Scientific Director: Curtis Huttenhower

https://hcmph.sph.harvard.edu/hcmcc
https://huttenhower.sph.harvard.edu

Evaluation of methods to isolate human gut viral communities for high-throughput sequencing

Ya Wang[1,2], Jordan Jensen[1], Eric Franzosa[1,2], Seth Rakoff-Nahoum[3], Curtis Huttenhower[1,2]

[1]Harvard T.H. Chan School of Public Health [2]Broad Institute of MIT and Harvard [3]Boston Children's Hospital

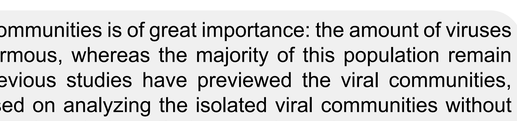
Characterizating gut viral communities is of great importance: the amount of viruses living in human gut is enormous, whereas the majority of this population remain uncharacterized. While previous studies have previewed the viral communities, these studies mostly focused on analyzing the isolated viral communities without systematic evaluating the isolation strategy. Particularly, it remains unclear whether the selected protocol could efficiently remove bacterial-originated nucleic acids while leaving the viral potion unaffected. Here, we present our work benchmarkng an experimental workflow to isolate virus-like particles (VLP) from gut microbial communities. Different experimental parameters were first evaluated in stool samples spiked with simple synthetic communities to come up with an optimized protocol, which was further validated in complicated mock communities consisting of five viruses representing common gut viral families (P1, T1, T4, phiX174 and MS2 phages) and then in stool samples spiked with the above mock communities. The optimized VLP isolation protocol efficiently reduced bacterial signals (16S rRNA gene copies) from 10^5 copies/ml to undetectable in mock viral communities and brought only minimal variations in virus copies between whole microbial vs. VLP-isolated communities. In spiked stool samples, the protocol depleted bacterial signals by approximately 100-fold. P1, T1 and T4 phages were slightly depleted in VLP-concentrated communities, while phiX174 and MS2 phages were enriched. Overall, these results show that the VLP isolated from gut microbial communities were affected by different experimental parameters. We thus come up with a standardized and relatively more optimized protocol for gut VLP isolation. This protocol will be further validated using using shotgun metagenomic and metatranscriptomic sequencing in stool samples collected from preemie babies.

Study design

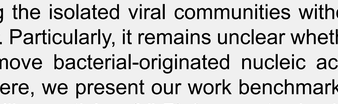
We first evaluated different experimental parameters known to potentially affect viral extraction, including storage buffer (SM buffer vs. 95% ethanol), filtration combinations (no filter vs. 0.45-μm-pore-size filter vs. 0.22-μm-por (e-size filter), methods for nucleic acid concentration after filtration (ultrafiltration with 100-KDa centrifugal filter vs. chemical-precipitation using PEG8000), and various enzymatic treatments (10U, 20U, 40U and 100U DNase; DNase alone vs. DNase plus RNase).

VLP protocol optimization using simple synthetic communities

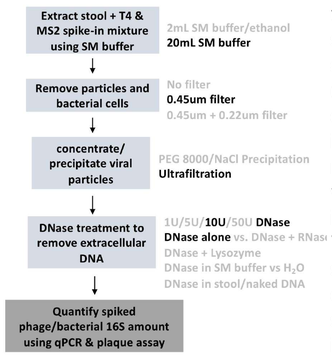
VLP protocol evaluation in mock viral communities and spiked stool samples



VLP protocol application in preemie stool sampleset



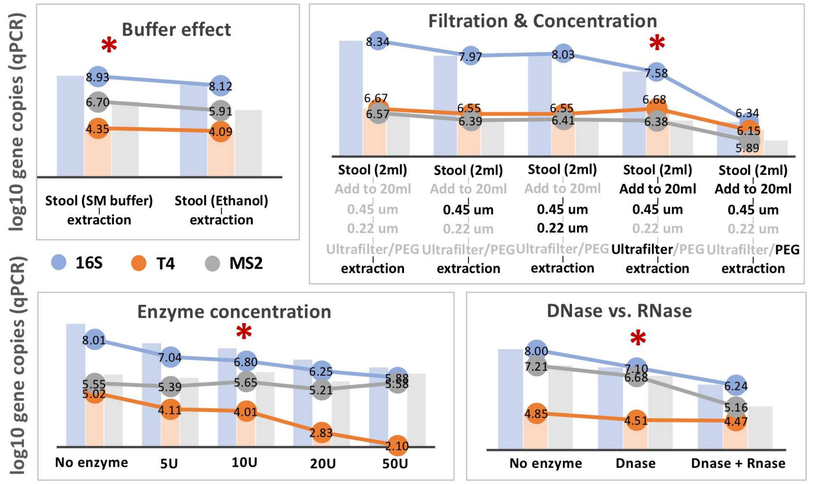
Optimized VLP isolation protocol



Different combinations of the parameters were tested by qPCR in stool samples spiked with simple synthetic viral communities comprising an sRNA phage (MS2) and a dsDNA phage (T4). Conditions with the highest purification efficiency (largest depletion of bacterial signals, i.e. lowest 16S rRNA gene copies) and the minimum impact on the spiked viruses (highest viral gene copies) were selected and combined into an optimized protocol.

The optimized protocol was then evaluated in a slighly more complex viral community comprising five viruses and in spiked stool samples. We further applied this VLP protocol in stool samples from preemie babies.

Evaluating different experimental parameters for gut VLP isolation



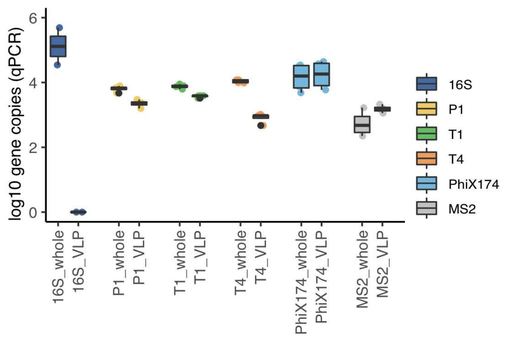
Purification efficiency of a gut VLP isolation protocol could be affected by various experimental parameters:

--Ethanol did not differ greatly in preserving bacterial/viral signals compared to the SM buffer;
--Stool samples filted with 0.45 μm filter and then concentrated using 100kDa centrifugal filter depletd the most bacterial signals while preserving the most viruses;
--Enzymatic treatment using 10 Unit DNase alone effectively remove bacterial signals while causing relatively small depletion on DNA virus. The application of RNase caused negative effect to RNA viruses.

VLP protocol depleted bacterial signals effectively in mock communities and in spiked stool samples

The optimized VLP protocol was applied in more complex mock viral communities consisting of five viruses representing the common viral families in the gut (i.e. T4, T1, P1, phiX174 and MS2 phages) and then in stool samples spiked with the above mock communities, to see if the optimized VLP protocol caused any differential bias to different families of viruses

VLP protocol in mock viral communities



--The optimized VLP isolation protocol effectively reduced bacterial signals (16S rRNA gene copies) from 10^5 copies/ml to undetectable in mock viral communities;

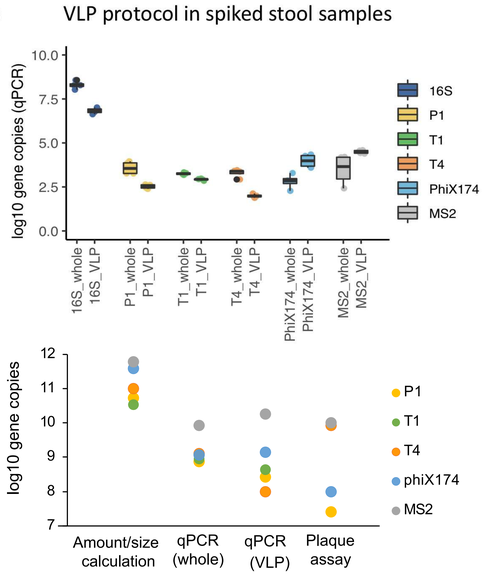
--Compared to the obvious depletion in bacterial signals, only minor differentiations were observed in different viruses between whole-community samples vs. VLP-concentrated samples.

in stool samples spiked with mock viral communities, the VLP protocol depleted bacterial signals by approximately 100-fold.

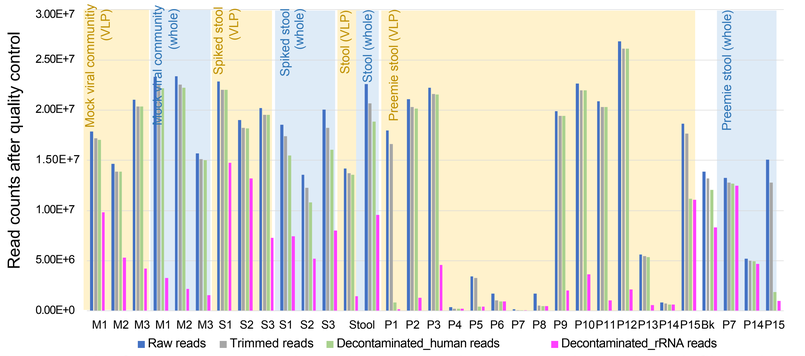
Similar as in mock viral communities, minor viral differentiations were observed in whole-community samples vs. VLP-enriched samples: communities:

--P1, T1 and T4 phages were slightly depleted in samples treated with VLP protocol, while phiX174 and MS2 phages had little increases.

--The differentiations in viruses were possibly introduced by the variations in detection methods (i.e. qPCR) rather than the VLP-protocol itself.



VLP protocol applied in stool samples for shot gun sequencing



Raw reads from metatranscriptomic sequencing were trimmed and filtered to remove low quality reads and sequencing adaptors, and then decontaminated by removed human- and 16S rRNA-originated reads. In mock viral communities and spiked stool samples, VLP-enriched samples (VLP) got higher read coverage after the above QC steps compared to the original ones (whole). This would likely suggest that the VLP protocol works effectively in enriching sequencing signals from virus-like particles.

Ongoing works

We are currently continuing the analysis of shotgun metagenomic and metatranscriptomic sequencing results generated from the VLP protocol, as well as to develop computational approaches for the taxonomic and functional profiling of gut viral communities. These can circumvent some of the current limitations of the VLP isolation protocol, providing a complementary view of viral communities, and directly observing functional activities involved in the gut-virome interaction.

Acknowledgments

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

http://huttenhower.sph.harvard.edu

Diet and gut microbial interactions in irritable bowel syndrome subtypes

Yiqing Wang,[1,2] Wenjie Ma,[1,2] Raaj Mheta,[1,2] Long H. Nguyen,[1,2,3] Mingyang Song,[1,2,4,5] David A. Drew,[1,2] Francesco Asnicar,[7] Curtis Huttenhower,[3,6] Nicola Segata,[7,8] Jonathan Wolf,[9] Tim Spector,[10] Sarah E. Berry,[11] Kyle D. Staller,[1,2] Andrew T. Chan.[1,2,5,12]

[1] Clinical and Translational Epidemiology Unit, Massachusetts General Hospital (MGH) & Harvard Medical School (HMS); [2] Division of Gastroenterology, MGH & HMS;
[3] Department of Biostatistics, Harvard T.H. Chan School of Public Health (HSPH); [4] Department of Epidemiology, HSPH; [5] Department of Nutrition, HSPH; [6] Broad Institute of MIT and Harvard;
[7] Department of Cellular, Computational and Integrative Biology, University of Trento; [8] European Institute of Oncology Scientific Institute for Research, Hospitalization and Healthcare; [9] Zoe Global Ltd;
[10] Department of Twin Research, King's College London; [11] Department of Nutritional Sciences, King's College London; [12] Channing Division of Network Medicine, Brigham and Women's Hospital & HMS

Irritable bowel syndrome (IBS) is a common functional gastrointestinal disorder, yet the role of diet and gut microbial communities in the pathophysiology of IBS is not fully understood. Thus, we investigated the interplay between dietary risk factors and specific taxonomic and functional groups in IBS subtype (IBS-C, constipation; IBS-D, diarrhea; IBS-M, mixed). We found that participants with IBS were predominantly female, younger, and attained higher levels of education than those without IBS. Participants with IBS-D more frequently consumed healthy plant-based foods compared to other participants, as well as higher levels of animal-based foods but lower levels of lactose than those without IBS. Gut microbiome composition differed slightly by IBS subtype, as reflected by nominally lower microbial diversity in IBS-D. Using linear regression adjusted for a wide range of host factors, we identified several taxa and functional pathways associated with specific IBS subtypes. Although limited by the available population size, *Faecalibacterium prausnitzii* showed intriguing evidence of interaction with dietary risk factors in association with IBS subtypes. Our findings add further evidence to the alterations in the gut microbial composition, function, and diet-microbiota interactions specific to IBS subtypes.

The Personalized Responses to Dietary Composition Trial (PREDICT) 1 study

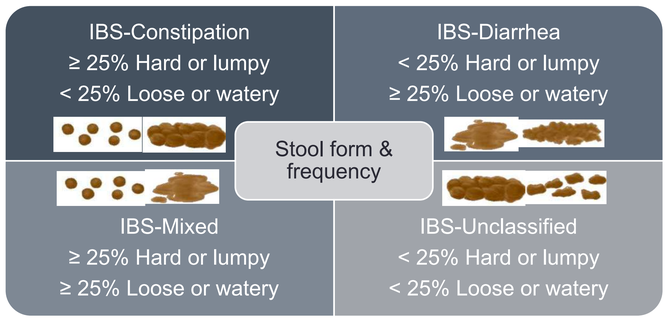
- 131-item European Prospective Investigation into Cancer and Nutrition (EPIC) food frequency questionnaire (FFQ) capturing average intake in the past year
- Other health and socioeconomic information

A single-arm, single-blinded intervention study using genetic, metabolomic, metagenomic and meal-context information to predict individual's postprandial responses to food.

ROME II: Diagnosis criteria for IBS

At least three months, with onset at least six months previously, of recurrent abdominal pain or discomfort and associated with two or more of the following:

- Improvement with defecation;
- Onset associated with a change in frequency of stool;
- Onset associated with a change in form (appearance) of stool.

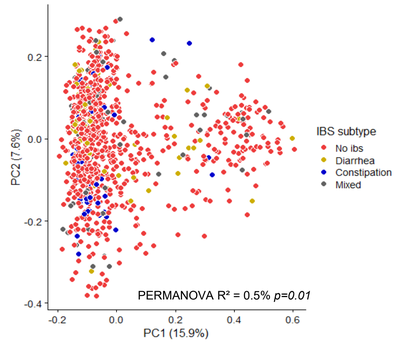


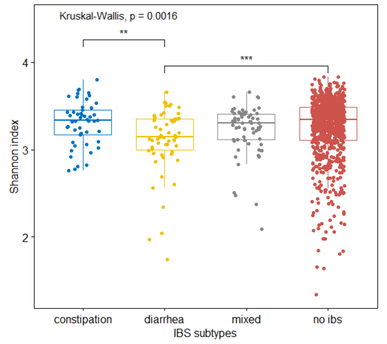
Participant characteristics

		IBS subtype			
	No IBS	Constipation	Diarrhea	Mixed	P-value
N	797	49	59	64	
Age (y)	46.1 (11.9)	44.5 (13.7)	42.9 (12.3)	41.3 (11.3)	0.005
Female, N (%)	547 (68.6%)	45 (91.8%)	55 (93.2%)	54 (84.4%)	<0.001
BMI (kg/m²)	25.5 (4.82)	25.8 (6.82)	26.9 (6.08)	26.3 (5.47)	0.31
Education, N (%)					0.29
Low (level 0-2)	97 (12.2%)	6 (12.2%)	5 (8.5%)	8 (12.5%)	
Medium (Level 3-4)	225 (28.2%)	8 (16.3%)	10 (16.9%)	15 (23.4%)	
High (college and above)	466 (58.5%)	34 (69.4%)	44 (74.6%)	41 (64.1%)	
Depression, N (%)	73 (9.2%)	3 (6.1%)	10 (16.9%)	10 (15.6%)	0.08
Physical activity, N (%)					0.10
Low (<1/wk)	189 (23.7%)	24 (49.0%)	24 (40.7%)	29 (45.3%)	
Medium (1-2/wk)	283 (35.5%)	16 (32.7%)	11 (18.6%)	15 (23.4%)	
High (>=3/wk)	324 (40.7%)	9 (18.4%)	23 (39.0%)	20 (31.3%)	
Energy intake (kcal/d)	1740 (516)	1690 (524)	1600 (459)	1630 (475)	0.08
Fiber intake (g/day)	17.1 (6.0)	17.3 (6.1)	17.4 (6.6)	16.8 (6.3)	0.91
Lactose intake (g/day)	7.60 (5.31)	7.89 (5.91)	5.54 (4.40)	6.33 (5.09)	0.01
Animal-based food score	28.9 (4.49)	30.2 (4.26)	30.4 (4.00)	29.1 (4.29)	0.03

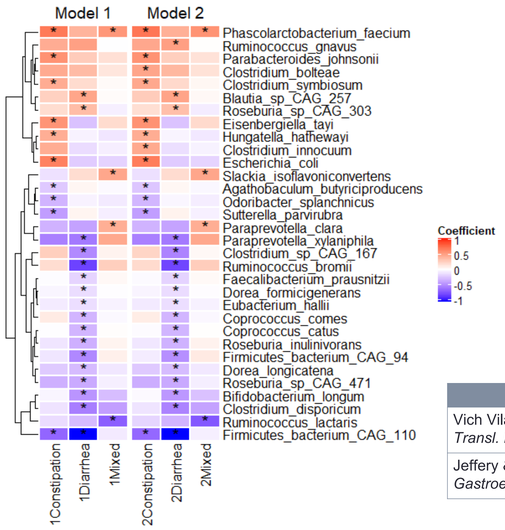
Differences among numeric variables categorical variables were tested using Kruskal-Wallis test and Chi-square test.

Gut microbiome composition differed slightly by IBS subtype





Individual taxa associated with specific IBS subtypes

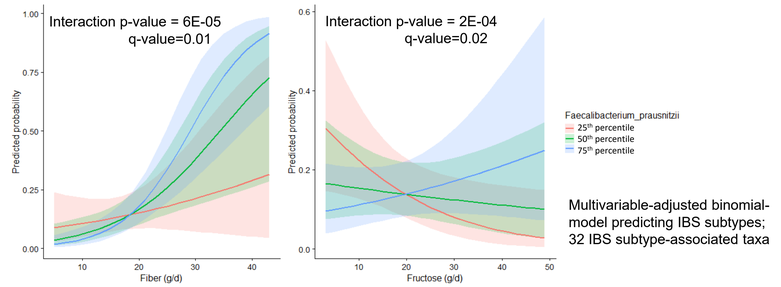


- A slight increases in typically pro-inflammatory taxa during IBS-C (e.g. *Ruminococcus gnavus, Escherichia coli*).
- A loss of typical gut strict anaerobes during IBS-D (e.g *Faecalibacterium prausnitzii*), explaining the overall lower diversity.
- Our results were modestly correlated with prior studies.

	IBS-C	IBS-D	IBS-M
Vich Vila et al., *Sci. Transl. Med.* (2018)	r = 0.15 p = 0.19	r = 0.43 p < 0.001	r = 0.02 p = 0.86
Jeffery & Das et al., *Gastroenterology*. (2020)	r = 0.45 p < 0.001	r = 0.34 p < 0.001	r = 0.03 p = 0.75

Taxa ≤0.01% relative abundance with ≤10% prevalence were filtered, leaving 170 taxa;
Model 1 covariates: sex, age, education, smoking, menopausal status, hormone therapy, antibiotic use, probiotic use, BMI, Bristol;
Model 2 covariates: Model 1 + diet quality (AHEI), total energy intake;
* FDR < 0.25.

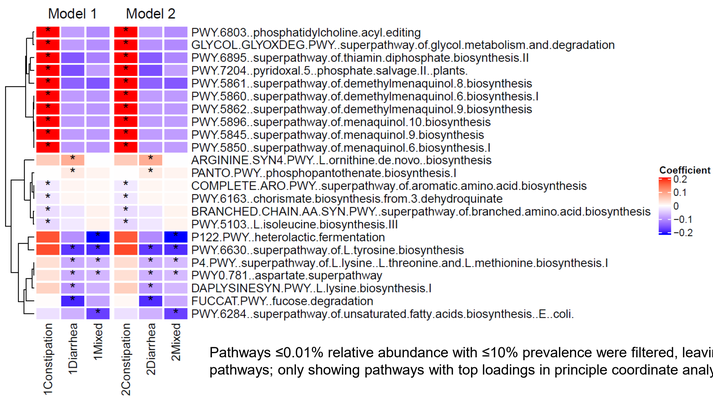
Interaction between diet and taxa in IBS-D



Multivariable-adjusted binomial-model predicting IBS subtypes; 32 IBS subtype-associated taxa

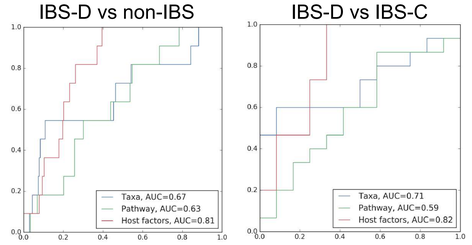
While the predicted probability for was lower at higher levels of *Faecalibacterium prausnitzii* when fiber and fructose intake was low, this association was reversed as fiber and fructose intake increased.

Individual pathways associated with specific IBS subtypes



Pathways ≤0.01% relative abundance with ≤10% prevalence were filtered, leaving 301 pathways; only showing pathways with top loadings in principle coordinate analysis.

Random forest model distinguishing IBS subtype



Although outperformed by other host factors, as expected for the notoriously multifactorial etiology of IBS, gut microbial taxa and functional pathways were significant independent machine learning predictors for distinguishing participants with IBS-D from those with IBS-C or no IBS.

5-fold Cross validation with 500 trees and weighted by inverse proportion of the respective frequencies of IBS subtypes; 769 taxa, 445 pathways, & 12 host factors (Model 2 covariates).

Conclusions

- There are IBS subtype-specific variations in gut microbial composition, function, and diet-microbiota interactions.
- Our findings may provide insights into microbiome-aware dietary interventions for IBS treatment.
- Further longitudinal studies are needed to confirm our results.

Acknowledgement

This work was supported by R35 CA253185, and R01 CA202704, Zoe Ltd, Biotechnology and Biological Sciences Research Council, MGH, American Gastroenterological Association, Crohn's and Colitis Foundation, American Cancer Society, Wellcome Trust, MRC, European Union, Chronic Disease Research Foundation, and the NIHR-funded BioResource, Clinical Research Facility, and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

Identifying strain-specific functional genes in colorectal cancer

Yan Yan[1,2], Kelsey N. Thompson[1,2], Andrew Ghazi[1,2], Andrew M. Thoma[3], Long H. Nguyen[4,5], Paolo Manghi[3], Lauren J. Mciver[1,2], Emma Accorsi[1,2], Eric A. Franzosa[1,2], Nicola Segata[3], Andrew T. Chan[4,5], Wendy S. Garrett[2,6], Curtis Huttenhower[1,2,6*]

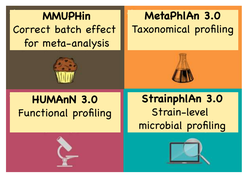
[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, [2]Broad Institute of MIT and Harvard, [3]CIBIO Department, University of Trento, Trento, Italy
[4]Clinical and Translational Epidemiology Unit, [5]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School,
[6]Department of Immunology & Infectious Diseases, Harvard T.H. Chan School of Public Health

Changes in the gut microbiota have been associated with colorectal cancer (CRC), but neither the causal mechanisms nor corresponding microbial strains and small molecule products have been elucidated for CRC. We have developed a new strain-level meta-analysis using stool metagenomic profiles of 600 CRC patients, 143 with pre-cancerous adenomas, and 662 healthy controls from nine recently published CRC microbiome studies. We created the MMUPHin framework to jointly normalise these datasets and identify potential consistently significant links between CRC neoplasia, severity, and microbial species and strains. We identified several species as novel CRC biomarkers including several typical oral species. Interestingly, we identified instances where the gene carriage was significantly different among strains of the same species in CRC patients compared to control individuals, indicating strain-specific functional distinctions in the CRC microbiome that would be invisible in taxonomically-focused analyses. We also found a group of genes unique to subsets of *E. coli* pan-genome associated with CRC phenotypes, comprising encoded pathways for lysine degradation, adhesion, and flagellar motility. This study adds further evidence to the hypothesis that strain-level genomic variation in gut microbes may be a major driver in the initiation or development of CRC.

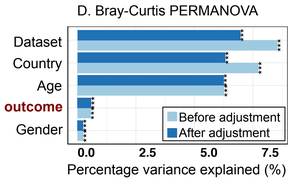
Expanded metagenomes and methods for meta-analyzing the CRC microbiome

A. Comparison of size of previous and expanded CRC datasets

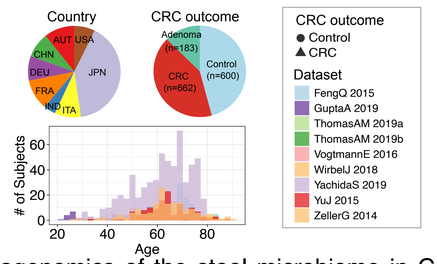
C. Microbiome profiling (bioBakery 3.0)



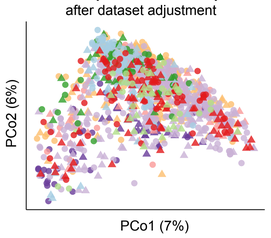
D. Bray-Curtis PERMANOVA



B. Size and characteristics of the expanded CRC dataset

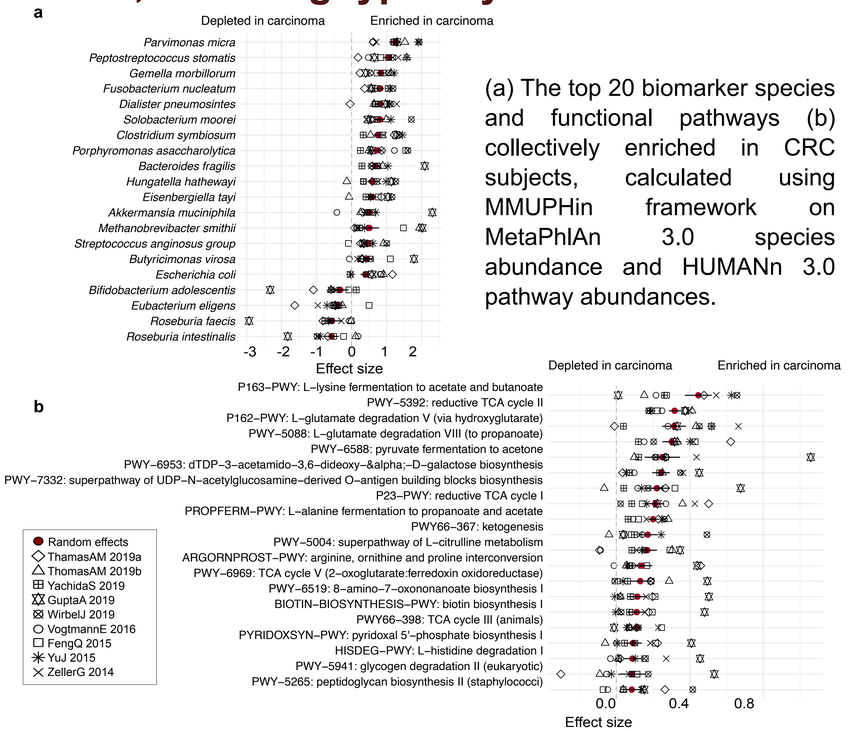


E. Bray-Curtis dissimilarity after dataset adjustment

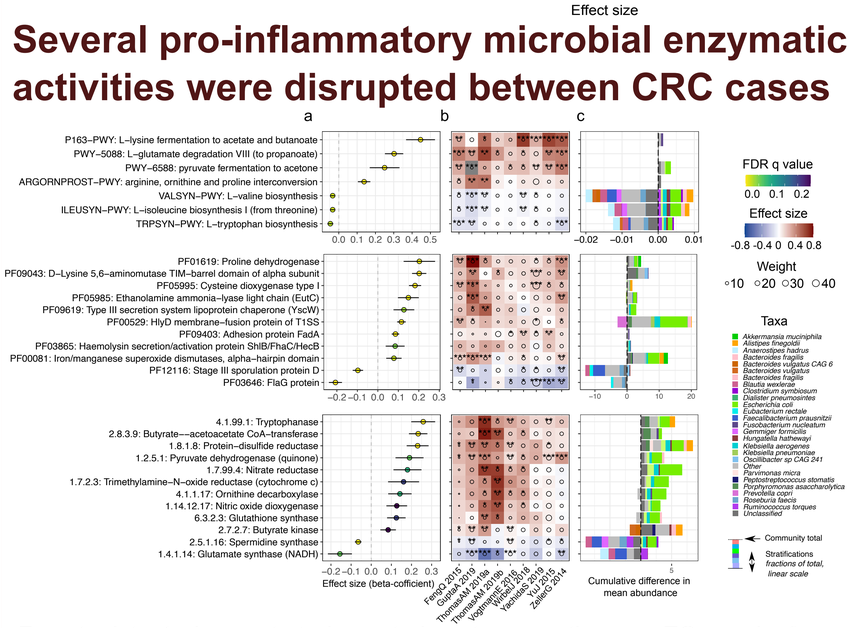


Metagenomics of the stool microbiome in CRC populations. a) size and characteristics of the large scale CRC metagenomic datasets. b) Performing batch (study) effect adjustment in CRC microbial features. c) Principal corrdinate analysis (PCoA) of stool metagenomic species. d) Typical of western populations, gradients of Bacteriodetes and Firmicutes dominance are seen across populations.

Known and novel taxa and pathway enriched in CRC, including typically oral microbes

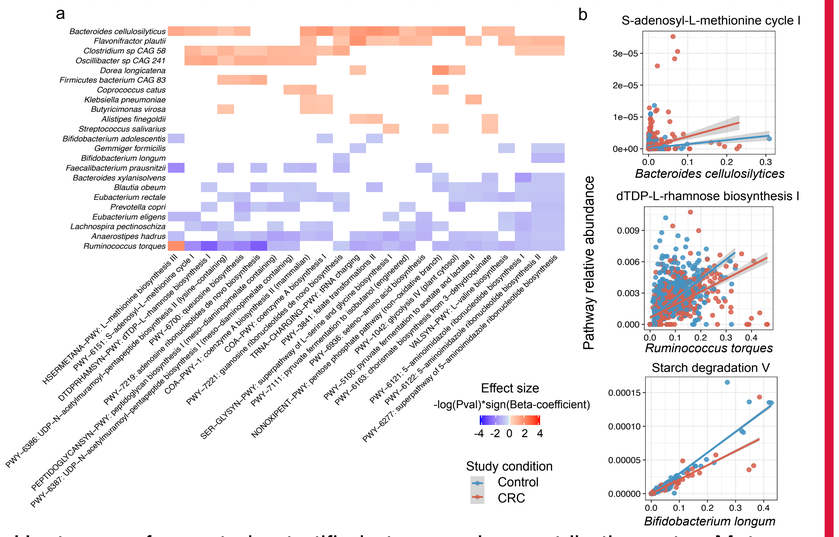


Several pro-inflammatory microbial enzymatic activities were disrupted between CRC cases



Forest plot a) demonstrated curated metacyc pathways, ECs and pfams statistics summary from MMUPhin models. Colored points estimated by FDR corrected P-value between carcinoma and control group. b) Heatmap illustrates broad agreement in the effect size per study, but does not always reach statistical significance per study. c) Stacked barplots indicated the curated 27 taxonomic contributors to related functions.

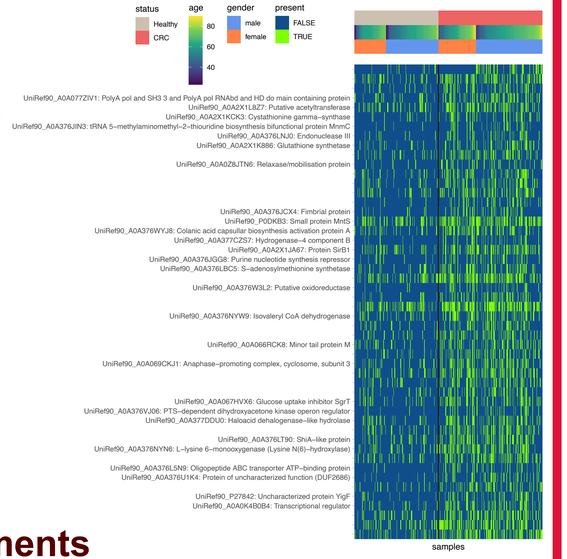
Sub-clades and their gene products are linked to strain fitness in the CRC niche



Heatmap of curated stratified taxonomic contributions to Metacyc pathways (a). Selected instances of scatter plots (b) of species by stratified pathway abundance in carcinoma and control, with more strain effects in the pathways than expected based on metagenomic abundance.

E. coli strain-specific gene carriage in CRC cases

A group of genes unique to subsets of *E. coli* pan-genome associated with CRC phenotypes, comprising encoded pathways for lysine degradation, flaggear motility. Each column indicates a strain from individuals (n=414). Rows represent top 50 CRC associated genes (FDR q < 0.25).



Acknowledgements

This work was funded in part by Charles A. King Trust Postdoctoral Research Fellowship Program (YY), by Cancer Research UK Grand ChallengeInitiative C10674/A27140 (WG), and by NIH NIDDK R24DK110499 (CH). Methods used for analysis are available in the bioBakery at:

http://huttenhower.sph.harvard.edu

Pharmacological inhibition of gut microbial β-glucuronidase enzymes blocks DSS-induced colon inflammation

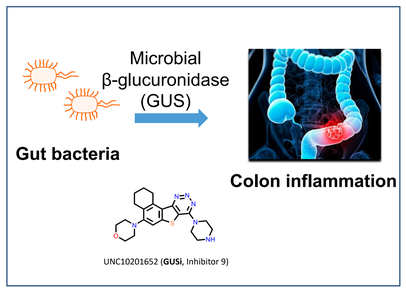
SCHOOL OF MEDICINE
Biochemistry and Biophysics

Jianan Zhang [1,2], Shan Sun [3], Ermin Zhao [2], Anthony Fodor [3], Eric A. Decker [2], Guodong Zhang [4,*], Matthew R. Redinbo [1,*]

1. Departments of Chemistry, Biochemistry, Microbiology and Genomics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
2. Department of Food Science, University of Massachusetts Amherst, Amherst, MA, USA
3. Departments of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, USA.
4. Department of Food Science and Technology, National University of Singapore, Singapore, Singapore.

The roles of gut bacteria on colon inflammation are little known

Colon inflammation and colon cancer are serious health problems. A gradually increasing incidence of colitis and colon cancer is reported in people with age less than 50. Due to the complicated genetic and environmental factors involved, however, there are currently no targeted therapeutic approaches available to reduce colon inflammation. Here we report that microbial β-glucuronidase (GUS) enzymes act as a therapeutic target to reduce colon inflammation in mice.



Inhibition of β-glucuronidase (GUS) enzymes attenuated DSS-induced colon inflammation

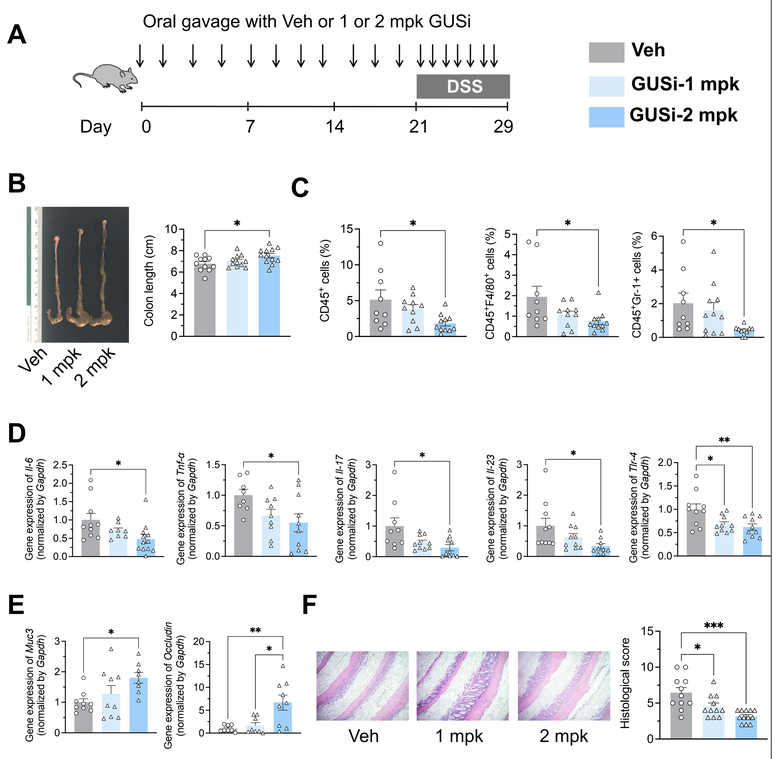


Fig. 1. Targeted inhibition of gut microbial GUS enzymes attenuated DSS-induced colon inflammation. A. Experiment scheme: C57BL/6 mice were treated with 1mpk or 2 mpk dose of GUSi or vehicle via oral gavage, then stimulated with DSS to induce colitis. **B.** Colon length. **C.** Immune cells in colon tissue: CD45+ cells, CD45+F4/80+ cells and CD45+ Gr1+ cells. **D.** Gene expression of inflammatory cytokines in colon tissue: *Tnf- α, Il-6, Il-17, Il-23 and Tlr-4* . **E.** Gene expression of tight junction protein in colon tissue: *Muc3* and *Occludin*. **F.** H&E histological staining of colon tissue. Left: representative H&E histological images of colon. Right: quantification of histology score. n = 10-12/ group, the data are mean ± SEM. *P < 0.05, ** P < 0.01, ***P < 0.001, **** P < 0.0001.

Inhibition of GUS enzymes altered beta diversity and the composition of gut bacteria

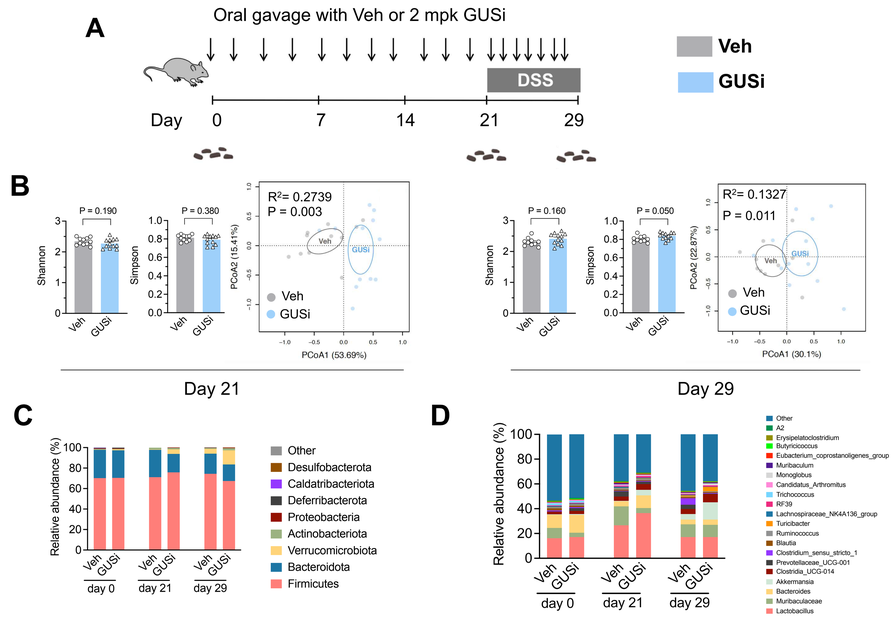


Fig.2 Inhibition of gut microbial GUS enzymes altered the alpha, beta diversity and composition of gut microbiota in DSS-treated mice. A. Experimental scheme. **B.** Alpha diversity and beta diversity on day 21 (before DSS- treatment) and on day 29 (after DSS treatment). **C-D.** GUSi altered the composition of the microbiota at phylum level and at genus level. The data are mean ± SEM. P < 0.05. n=10 or 11 in veh and n=12 in GUSi group.

The specific altered gut bacteria by GUS inhibitor at phylum and genus levels

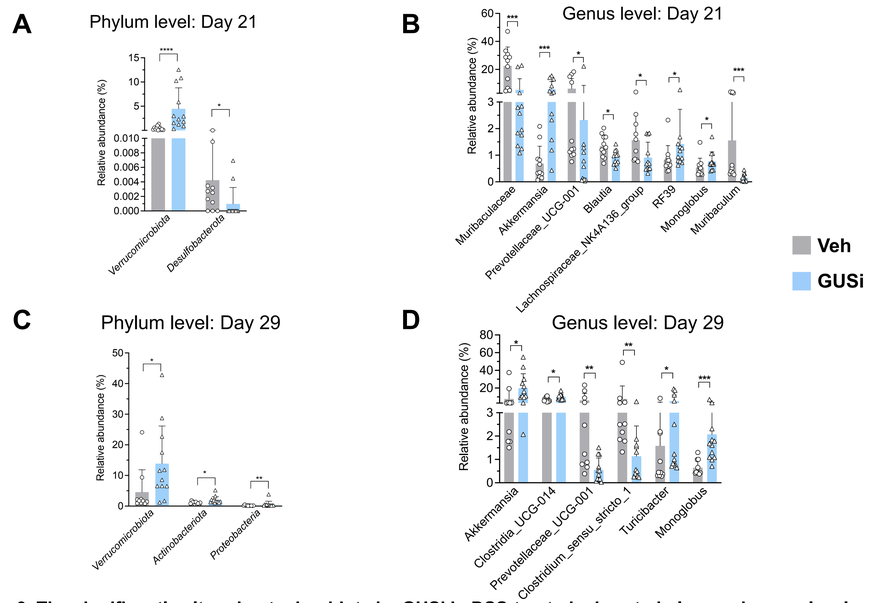


Fig.3 The significantly altered gut microbiota by GUSi in DSS-treated mice at phylum and genus level. The data are mean ± SEM. P < 0.05, FDR < 0.1. *P < 0.05, ** P < 0.01, ***P < 0.001, **** P < 0.0001.

Inhibition of gut microbial GUS altered gut bacteria composition at all taxonomic levels

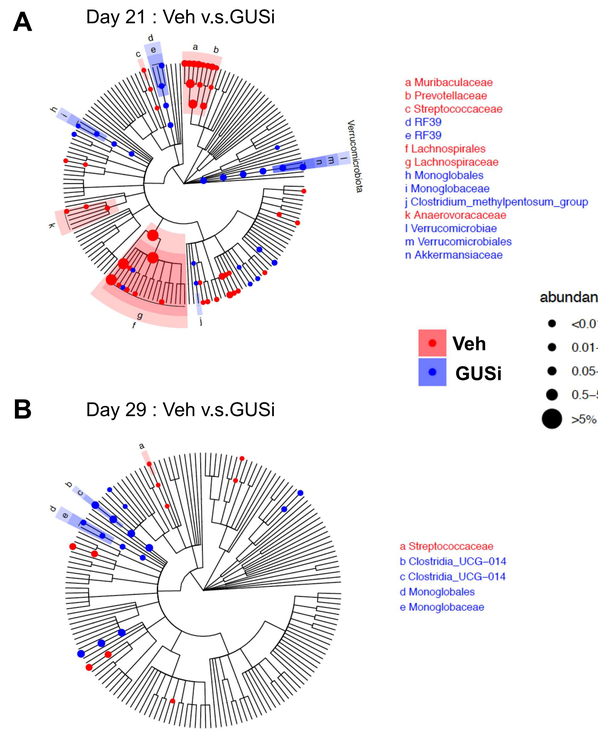
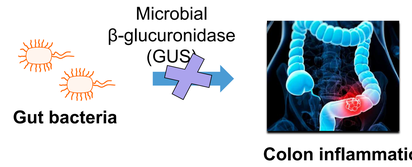


Fig.4 Inhibition of gut microbial GUS enzymes altered the composition of gut microbiota in DSS-treated mice. A. GUSi altered the composition of the microbiota at all taxonomic levels on day 21. **B.** GUSi altered the composition of the microbiota at all taxonomic levels on day 29. The data are mean ± SEM. P < 0.05, FDR < 0.1. n=10 or 11 in veh and n=12 in GUSi group.

Conclusion and future directions



Targeted inhibition of gut microbial GUS alleviates DSS induced colon inflammation and microbial GUS enzymes could be a potential therapeutic target for the prevention and/or treatment of colon inflammation.

Acknowledgement

1. Funded by NIH grants GM135218 and GM137286 (MRR).
2. Funded by interdisciplinary faculty research award from the University of Massachusetts Amherst, USDA NIFA 2019-67017-29248 and 2020-67017-30844, and USDA/ Hatch MAS00556 (G.Z.)

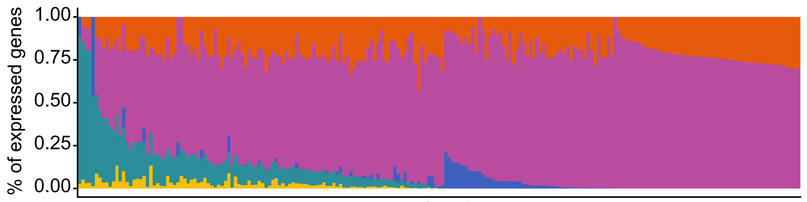
Predicting functions of uncharacterized gene products from microbial communities

Yancong Zhang[1,2,3], Xueying (Sonia) Huang[2,3], Amrisha Bhosle[1,2,3], Sena Bae[2,3], Andy Krueger[4], Wendy S. Garrett[2,3], Eric A. Franzosa[1,2,3], Curtis Huttenhower[1,2,3]

[1]Broad Institute, [2]Harvard T. H. Chan School of Public Health,
[3]Harvard Chan Microbiome in Public Health Center, [4]Takeda Pharmaceutical Company Limited

Functions of microbial gene products

Microbial communities are rich reservoirs for molecular functions that influence environmental and host-associated chemistry, with numerous roles in ecosystem maintenance, health, and disease. However, our knowledge of these molecular mechanisms is limited, due to the massive range of microbial genetic material in comparison to the limited throughput available for experimental characterization. Here, we assessed a novel method to systematically predict functional capacity of uncharacterized proteins in microbiome by integrating high-dimensional community-wide data from the Integrative Human Microbiome Project (HMP2).



Protein classes (and abbreviations)

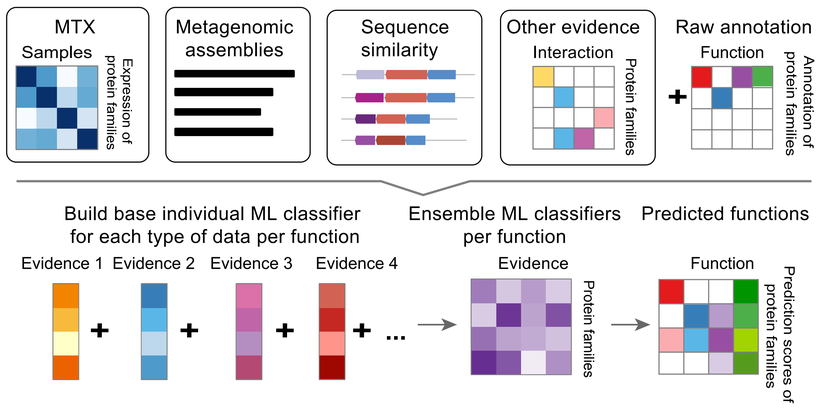
- SC = **S**trong homologs to **C**haracterized UniProtKB proteins
- SU = **S**trong homologs to **U**ncharacterized UniProtKB proteins
- UPI = Strong homologs to uncharacterized **UniP**arc proteins
- RH = **R**emote **H**omologs to UniProt proteins
- NH = **N**o **H**omologs to UniProt proteins

We enumerated expression profiles of five groups of proteins from HMP2 based on homology and functional annotation (abbreviated SC, etc. and defined above):
- Metatranscriptomes (MTX) capture expression profiles of community proteins;
- Expressed proteins without characterization are dominant in the community.

Here, "characterized" proteins are defined as those annotated with "informative" biological process (BP) terms, i.e. each BP term contains >1% of annotated genes without any child term passing the criteria.

FUGAsseM for function prediction in microbiome

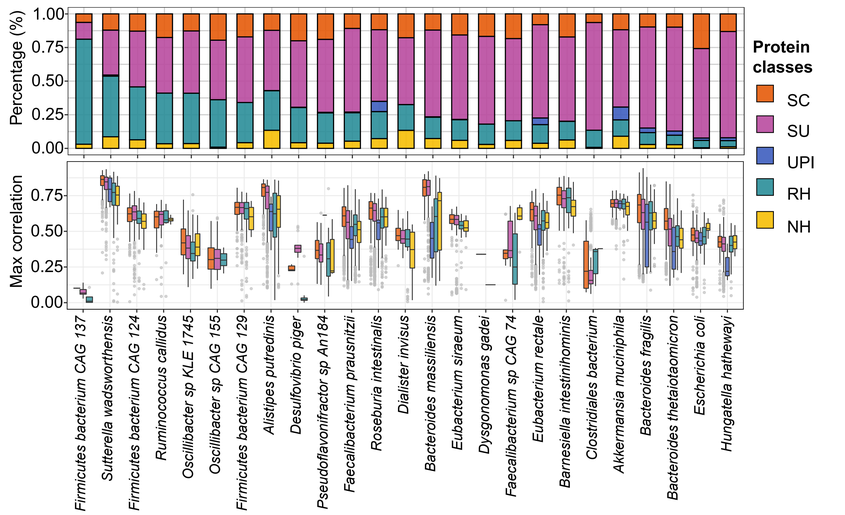
FUGAsseM (a **F**unction predictor of **U**ncharacterized **G**ene products by **Asse**ssing high-dimensional community data in **M**icrobiome) is generalizable to any types of microbial communities, providing a new approach to predict microbial protein functions.



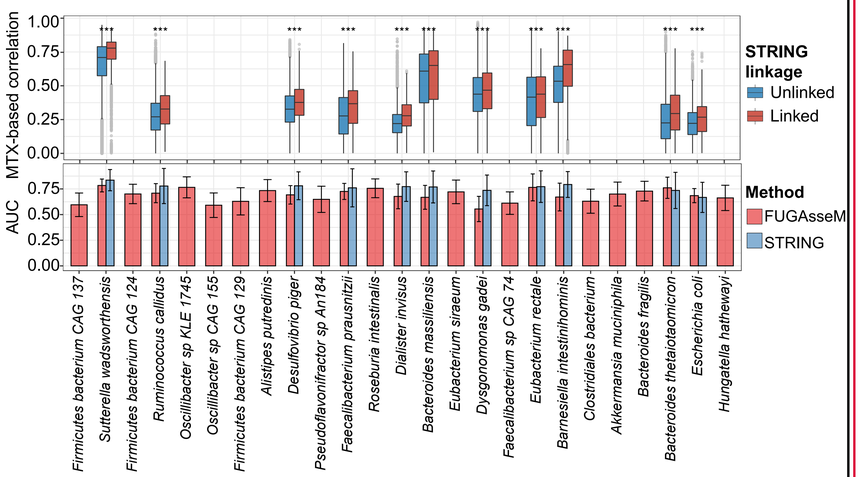
Novel proteins coexpressed with known proteins in stratified species

Coexpression patterns encoded by MTX

Among the top 25 species with the largest number of new proteins (lacking strong homologs to UniProt proteins), uncharacterized proteins, regardless of annotation status, are highly correlated with characterized proteins in the community, enabling "guilt-by-association" characterization.



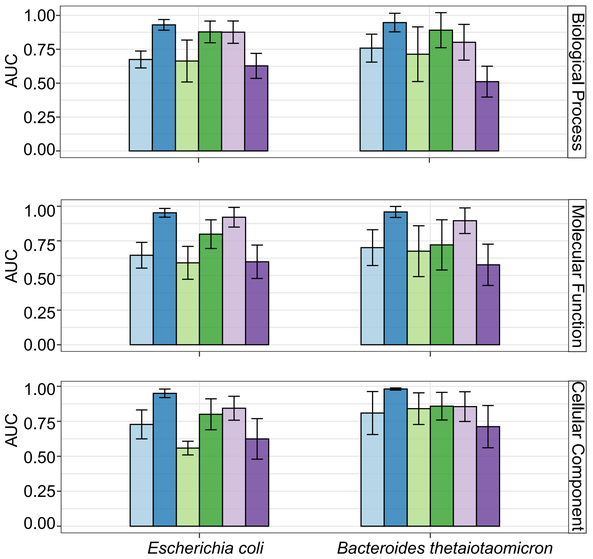
Coexpression patterns used for function prediction



We evaluated MTX-based coexpression by comparing with isolate-based coexpression:
- Among the overlapped species, proteins linked in STRING networks tend to have higher correlation among MTX networks;
- FUGAsseM's predictions based on MTX coexpression are comparable to STRING where they overlap, but FUGAsseM applies to many more species from communities.

FUGAsseM's performance is comparable to single-organism tools

We compared FUGAsseM with other single-organism methods in the task of predicting functions of proteins from stratified species in the community, and assessed performances of the same set of GO terms from different methods.



- FUGAsseM's performance is improved by aggregating other community-wide data;
- FUGAsseM shows comparable predictions to state-of-art single-organism tools;
- FUGAsseM works better for predicting BP terms than MF terms.

Conclusion and next steps

Our current results show that:
- MTX coexpression is comparable to isolate-based coexpression;
- MTX-based coexpression patterns are informative for function prediction;
- FUGAsseM predicts functions with high accuracy for microbial communities.

Further work including, but not limited to:
- Assess the performance for more species in the community;
- Examine applicability of FUGAsseM in other microbial communities;
- Validate predicted functions of novel proteins from communities.

Acknowledgements

This work has been supported in part by a grant from Takeda Pharmaceuticals (CH) and NIH NIDDK grant R24DK110499 (CH).

https://huttenhower.sph.harvard.edu/fugassem

@hutlab

Quantifying Shared and Unique Gene Content Across 17 Different Microbial Ecosystems

Samuel Zimmerman*, Braden T Tierney*, Chirag J Patel+, Aleksander D Kostic+

Joslin Diabetes Center · HARVARD MEDICAL SCHOOL · BLAVATNIK INSTITUTE BIOMEDICAL INFORMATICS

Introduction

The inhabitants of the human microbiome are not isolated to a single body site. Many microbes are transient, while others evolved from ancestors in other hosts or environments. As a result, we must contextualize the human microbiome across different ecologies to fully comprehend its diversity and evolutionary history. Here we built a resource for ecologically contextualizing the gene universe of metagenomic data.

Methods

We built a non-redundant gene catalog of 14, 183 samples across 17 different ecological contexts. We compared gene content of each ecology, identifying which are most similar. We then found genes shared between different ecologies and based on prevalence, we identified genes present in all 17 types of metagenomes.

Results

- We found 117 million genes from ~14,00 samples in 17 different ecosystems, almost 66% of which were singletons (fig 1A)
- Gut and environmental metagenomes have distinct genetic fingerprints (fig1B)
- Thousands of genes are shared between different ecologies (Fig 2A) and have functions representative of their environment (Fig 2B)
- There are 1,864 genes prevalent in all 17 ecologies (Fig 3)

Discussion

- The large number of singletons illustrates that our gene catalogue is far from complete.
- Genes enriched in environmental ecologies are enriched in energy production while those in the gut are enriched in carbohydrate metabolism and recombination
- The 1,800 genes found in all ecologies represent a new class of genes, those found in every metagenome, but not every bacterial genome

Citations

- Turnbaugh, Peter J., Ruth E. Ley, Micah Hamady, Claire M. Fraser-Liggett, Rob Knight, and Jeffrey I. Gordon. 2007. "The Human Microbiome Project." *Nature* 449 (7164): 804–10.
- S. L. Schnorr, K. Sankaranarayanan, C. M. Lewis Jr, C. Warinner, Insights into human evolution from ancient and contemporary microbiome studies. Curr. Opin. Genet. Dev. 41, 14–26 (2016).

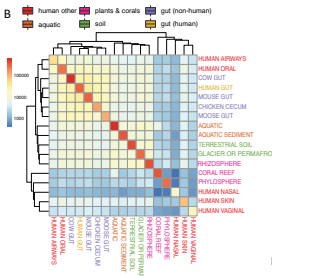
A

	Sample Size	Predicted ORFs	ORFs present at 30% identity
Oral	1,863	165,909,772	9,291,964
Nasal	189	1,632,935	616,937
Airways	137	2,340,437	348,374
Skin	418	10,286,622	3,203,302
Gut	10,002	1,239,066,644	48,026,249
Vagina	88	478,160	125,190
Animal gut	155	20,154,453	5,082,104
Non Animal Hosts	65	5,239,191	3,488,590
Environmental samples	1,266	231,457,516	55,061,188
Total	14,183	1,676,565,730	117,629,181

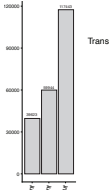
Singletons	77,371,367
Non Singletons	40,257,814

Figure 1. Description of gene catalogue (A) and genetic similarities between ecologies (B)

B



A



B

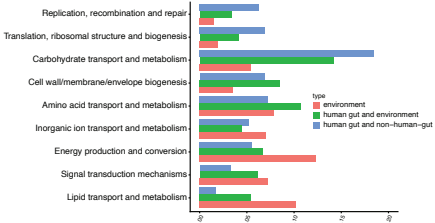


Figure 2. Number of genes shared between ecologies (A) and enriched functions (B)

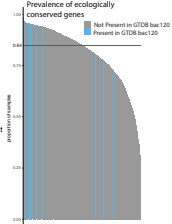


Figure 3. Prevalence of 1,864 globally conserved genes