Bacteria-virus interactions in the vaginal microbiome reduce herpes virus infectivity

Amanda N. D. Adams#, Gin Glick, Desmond Richmond-Buccola, Miranda Gavitt, Smita Gopinath*

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA;

#anadams@hsph.harvard.edu, @AndaAdams, https://www.linkedin.com/in/amanda-n-d-adams/; *https://www.gopinathlab.org/

Abstract

The vaginal microbiome is an important determinant of host health and the first barrier encountered by sexually transmitted pathogens during infection. Among the vaginal microbiome, Lactobacilli are associated with reduced susceptibility to viral infection, but the mechanisms by which various Lactobacilli strains reduce viral infectivity remain poorly understood. Using a collection of human vaginal microbial strains, we show that the prominent vaginal strain, *Lactobacillus crispatus* reduces infectivity of sexually transmitted pathogen Herpes Simplex Virus (HSV). Reduction of HSV infectivity is species specific, with *L. crispatus* reducing infection and disease better than gut-associated *L. reuteri*. Active cell metabolism is not required as UV-killed *L. crispatus* retain the ability to reduce herpes infection. Since one of the most abundant structures on the outside of the *L. crispatus* cell is peptidoglycan, we assessed whether peptidoglycan could reduce HSV infection. We found that commercially available purified peptidoglycan from multiple bacterial sources reduced herpes infection *in vitro* and *in vivo* in a mouse model of genital herpes infection. Mice were susceptible to reinfection, indicating that immunological memory is not activated. Cleavage of the glycosidic linkages in the peptidoglycan chain with lysozyme restored virus infectivity *in vitro* and *in vivo* suggesting that antiviral effects are dependent on longer peptidoglycan chains. Current studies aim to determine how Lactobacilli peptidoglycan contributes to a reduction in HSV infectivity focusing on HSV entry receptors and what species-specific peptidoglycan modifications allow *L. crispatus* to reduce infectivity better than other Lactobacilli. Such results provide a greater understanding of the ways that the vaginal microbiome serves as a physical barrier to infection and why some vaginal communities promote better antiviral protection than others.

The vaginal microbiome as a primary defense against pathogens

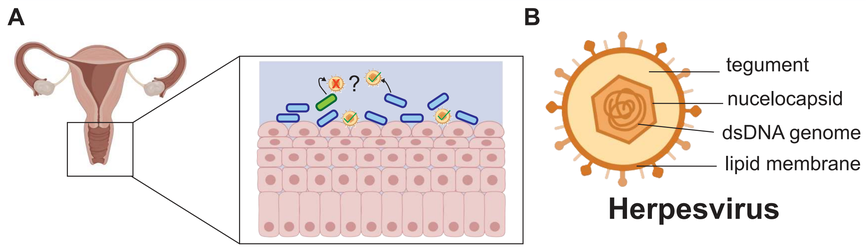


Figure 1. (A) The vaginal mucosa is colonized by an ecosystem of microbes that protect the host from invading pathogens, including viruses. The microbiome, which is largely dominated by Lactobacillus species, is the first barrier encountered by an exogenous pathogen. The vaginal microbiome can secrete molecules that interact with pathogens or the host to influence invasion. Loss of Lactobacilli spp. is linked to increased risk for viral disease, including herpes. (B) Herpes is an enveloped neurotropic dsDNA virus that infects the mucosal epitheia and establishes a lifelong latent infection in the dorsal root ganglia. In the work presented here, we investigate mechanisms by which the vaginal microbimoe influences herpes infection.

What makes some vaginal microbiomes better at protecting against herpes infection than others?

Vaginal *Lactobacilli* reduce HSV-2 infection *in vitro*

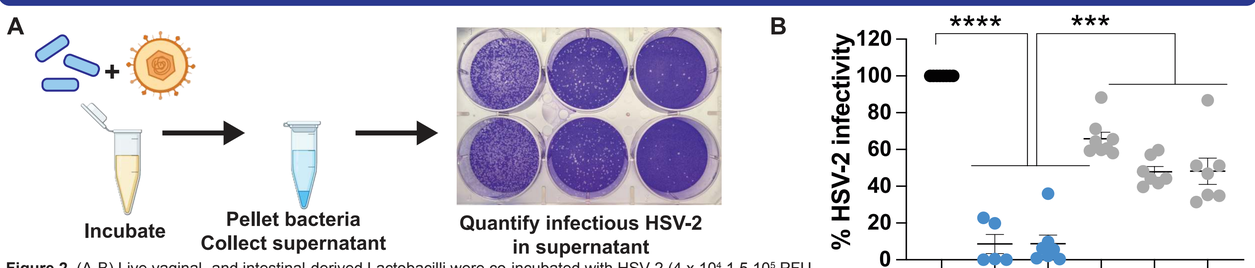
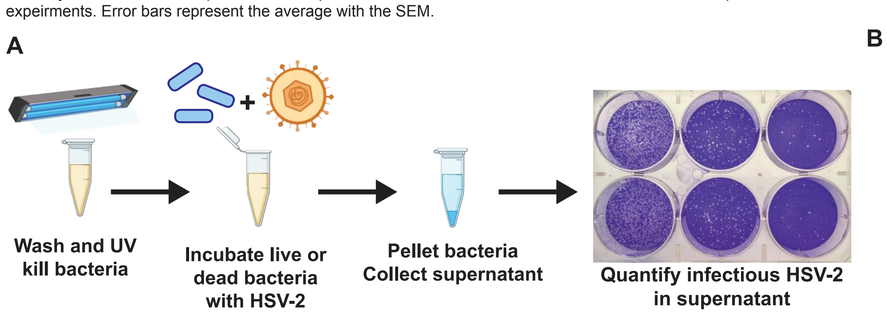


Figure 2. (A-B) Live vaginal- and intestinal-derived Lactobacilli were co-incubated with HSV-2 (4 x 10^4 1.5 10^5 PFU + 10^7 - 10^8 CFU, normalized by OD_{600}) for four hours in 1x PBS at 37C with HSV-2 alone in 1x PBS as a control. Mixtures were then pelleted at 500g to remove bacteria and supernatants plated onto confluent Vero cell monolayers at different dilutions. Cells were then incubated for 48 hours in growth limiting conditions with human IgG. Cells were fixed in crystal violet and plaques counted to determine supernatant infectivity. Counts were normalized to the average of HSV-2 alone. (B) Both vaginal and intestinal Lactobacilli suppressed HSV-2 infectivity, with vaginal Lactobacilli suppressing more than intestinal Lactobacilli. Significance was determined with ordinary one-way anova without multiple corrections, p-value <0.0001 **** and <0.001 ***. N= 5-8 across 3 independent expeirments. Error bars represent the average with the SEM.



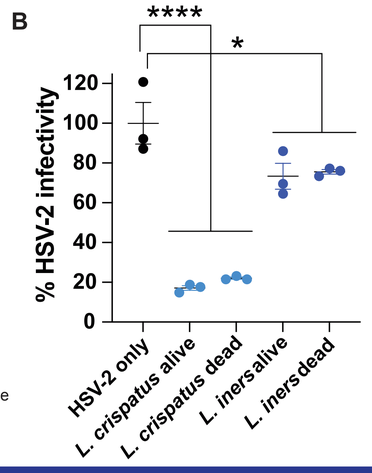
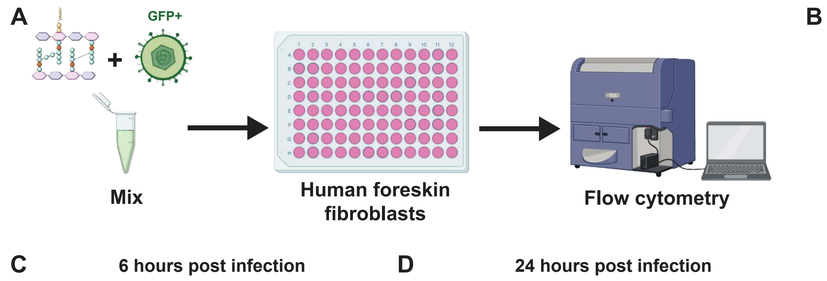


Figure 3. Vaginal Lactobacilli were normalized by OD600, washed 2x in PBS, and then UV killed before supernatants were incubated with HSV-2 for one hour at 37C (40,000 PFU + up to 108 CFU) with HSV-2 alone in PBS as a control. Infectious HSV-2 particles were quantified using a plaque assay on veros. Percent infectivity was quantified and normalized to the average of HSV-2 only. Error bars show the mean and the SEM and one-way anova reveals significance with a p-value of <0.0001 **** or <0.05 *.

Peptidoglycan reduces HSV-1 infectivity *in vitro*



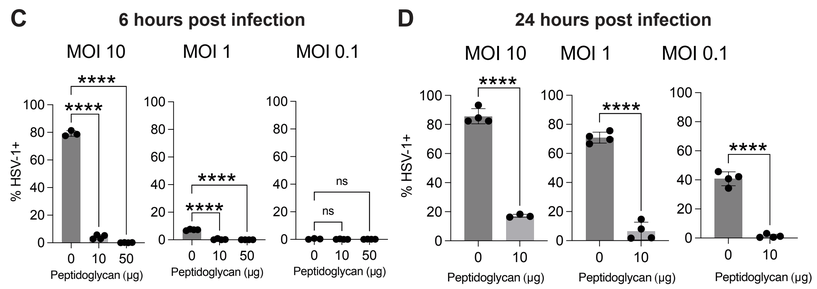
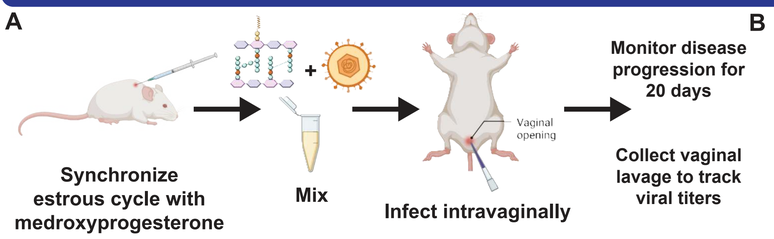
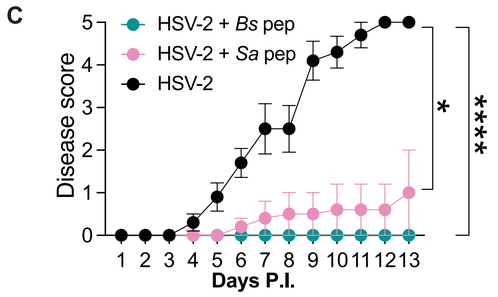


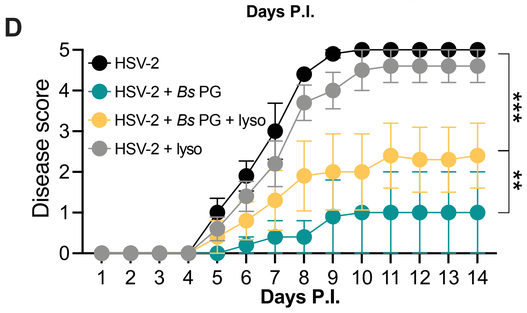
Figure 5. (A-D) To determine if PG could reduce infectivity of the herpes strain HSV-1, HSV-1 expressing a GFP-tagged capsid (GFP-VP26) was combined with [0.2mg/ml, 50 µg) *Bs* PG and added to immunocompetent human foreskin fibroblasts. After infection, GFP+ cells were quantified using flow cytometry (BD FACSymphony). (B-C) At 6 hours post infection (P.I.), in the absence of PG, 80% of cells at multiplicity of infection (MOI) 10 were GFP+. Cells that were incubated with 50µg PG were uninfected. (D) As the infection progressed 24 hours, 20% of live PG-treated cells were GFP+. These results demonstrate that PG can reduce HSV-1 infectivity *in vitro*. Error bars show the mean and the SEM and one-way anova reveals significance with a p-value of <0.0001****.

Peptidoglycan reduces HSV-2 infectivity *in vivo*





Disease score
1. genital erythema
2. genital hair loss
3. ruffled fur, hunched posture
4. hind limb paralysis
5. death



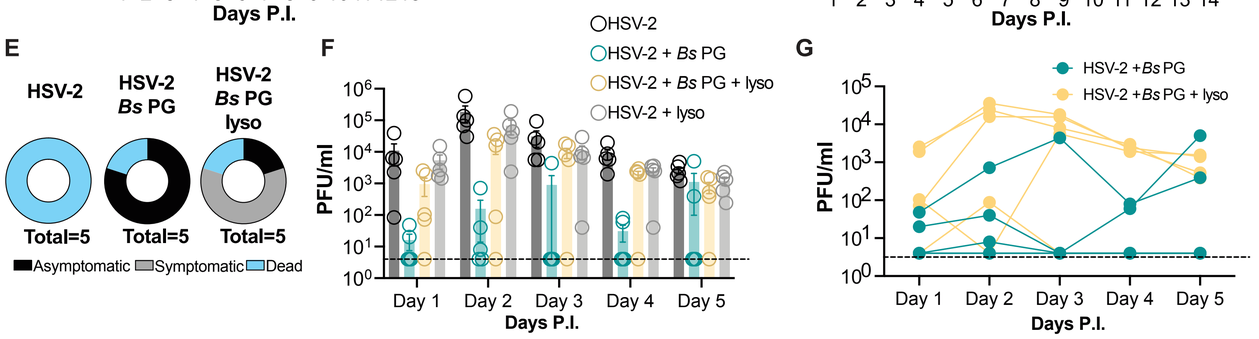
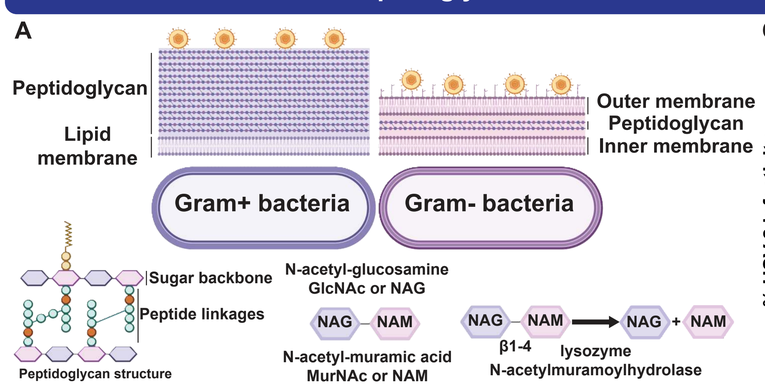
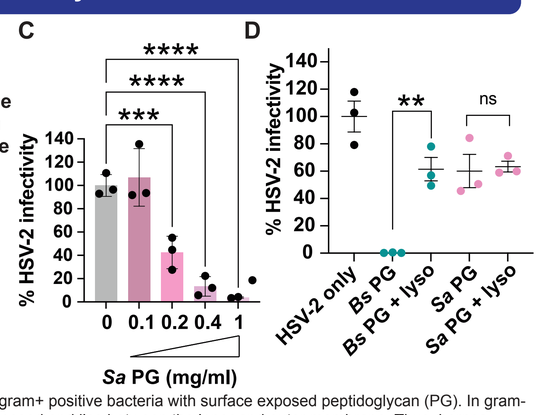


Figure 6. (A-G) To determine if PG could reduce HSV-2 infection *in vivo*, mice were infected intravaginally with a lethal dose of HSV-2 (10,000 PFU) with or without 50ug of B. *Bs* or *Sa* PG. (A) To synchronize disease progression, mouse estrous cycles are synchronized with 2mg of medroxyprogesterone injected subcutaneously 5-7 days before infection. Early in infection, the virus replicates in mucosal epithelial cells and infectious virus titers can be evaluated by collecting vaginal lavage daily and conducting plaque assays on Veros. Four days P.I., the virus infects enervating neurons and travels to the dorsal root ganglion. In humans, the virus enters latency, but in mice, the virus continues to replicate, traveling down neurons to fresh epithelial sites where viral infection results in inflammation and morbidity that can be scored. (B) All mice that received HSV-2 alone died within 14 days P.I. (n=5-10) (B-C) All the mice that received 50ug of *Bs* PG survived and only one *Sa* mouse died. In the experiment shown in (B), no *Bs* treated mice showed signs of disease, whereas some *Sa* treated mice did show symptoms, though these were statistically significantly different from the untreated mice. (D) In a separate experiment, to determine if the NAG-NAM linkage was required for this protection *in vivo*, mice were infected with HSV-2, 50ug *Bs* PG, and 10mM lysozyme (n=5). Mice that were treated with *Bs* PG showed significantly less disease burden than untreated mice. Mice that were treated with Bs PG and lysozyme showed signficantly more disease than mice treated with *Bs* PG alone, suggesting that the NAM-NAM linkage is imporatant for PG protection from HSV-2 infection *in vivo*. (E) Among the *Bs* PG/lysozyme treated mice, most of the mice were symptomatic, which is in contrast to the mice treated with *Bs* mice treated alone which were largely asymptomatic. (F) Viral titer tracking from vaginal lavage in all four treatments within the the 5 day window critical for disease establishment revealed that *Bs* PG treated mice were able to clear the virus by day one P.I. Mice treated with *Bs* PG and lysozyme did not clear the virus within the first few days P.I. This suggests that *Bs* PG blocks early viral establishment. (G) Curves represent the vaginal viral titers for a single mouse, revealing the heterogeneity in progression of virus establishment in individual mice treated with PG. The limit of detection for PFUs is 4 plaques (dotted line). Disease scores were compared using 2-way anova with multiple comparisons with p-value <0.0001 ****, <0.001 ***, <0.01 **, and <0.05 *. Kaplan-Meier survival curves were evaluated with Mantel-Cox test, p-value <0.001 **.

Peptidoglycan reduces HSV-2 infectivity *in vitro*





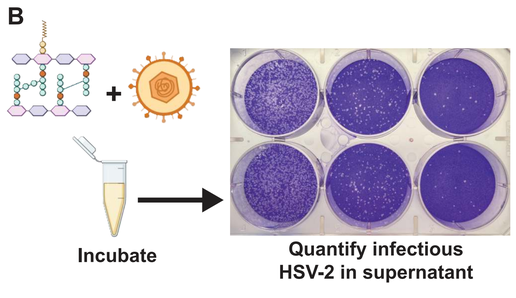


Figure 4. (A) Lactobacilli are gram+ positive bacteria with surface exposed peptidoglycan (PG). In gram- bacteria, PG is not surface exposed and lies between the inner and outer membrane. Thus, herpes virions could potentially interact with the PG of gram+ bacteria and the outer membrane and its surface structures in gram- bacteria. PG is typically made up of repeating units of N-acetyl-muramic acid (NAM) and N-acetyl-glucosamine (NAG). These glycan chains are linked via peptides. Lysozyme is able to cleave peptidoglycan by hydrolyzing the β1-4 linkage between NAG and NAM in susceptible bacteria, including Lactobacilli. Some bacteria, like *Staphylococcus aureus (Sa)*, are not susceptible to lysozyme activity due to protective modifications on the PG. (B-C) To test the effect of PG on HSV-2 infectivity, HSV-2 was incubated with commercially available *Sa* PG and added to vero cells before plaquing to quantify HSV-2 infectivity (C). *Sa* PG reduced HSV-2 infectivity in a dose dependent manner (C). (D) To test the effect of PG on HSV-2 infectivity in the absence of pre-incubation, [0.2mg/ml] PG from *Bs* and *Sa* was mixed with HSV-2 and added directly to veros. *Bs* PG reduced HSV-2 infectivity more than equivalent amounts of *Sa*PG and this effect was reduced for *Bs* when 10mM lysozyme was added on the veros at the same time. *Sa* PG reduction of HSV-2 infectivity was not reduced by lysozyme. Overall, these data suggest that PG is able to reduce HSV-2 infectivity *in vitro*, and that this reduction requires intact PG NAG NAM bonds. Error bars show the mean and the SEM and one-way anova reveals significance with a p-value of <0.0001****, <0.001***, or <0.01 **.

Future directions

Moving forward, we are excited to dissect the interactions between vaginal viruses and the microbial cell surface of Lactobacilli and other clinically important vaginal microbes. We are keenly interested in determining what makes certain vaginal bacteria better or worse at protecting from viral disease by purifying and characterizing the structure of the cell surface using microscopy, chemistry, and glycobiology.

***Lactobacillus reuteri* HM-102** **Lactobacillus iners HM-704** **Lactobacillus crispatus HM-637**

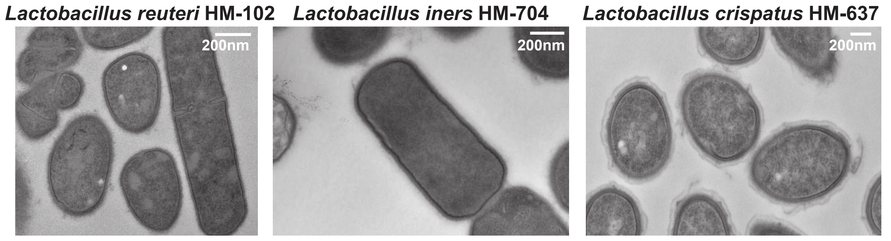


Figure 7. Lactobacilli were washed 1x in PBS and fixed in paraformaldehyde and gluteraldehyde and sectioned into 70nm slices for TEM. Cells were imaged using Tecnai G2 Spirit Bio-Twin TEM with NANOSPRT43 camera, 3000x direct magnification.

Acknowledgments

We thank David Knipe's lab of Harvard Medical School for providing us with GFP-HSV-1. We thank our funders: Harvard T.H. Chan School of Public Health (HSPH), HSPH Dean's Activation Award, HSPH Postdoctoral Association Travel Award, and the Searle Scholars Program. We also thank the members of the labs of Smita Gopinath, Wendy Garrett, and Flaminia Catteruccia for their feedback, assistance, and resources. Figures made with Biorender.

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

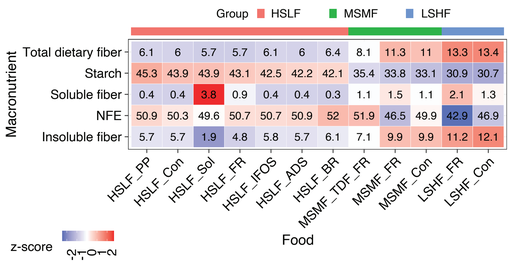
Amrisha Bhosle[1,2,3], Matthew I. Jackson[4], Aaron Walsh[2], Eric A. Franzosa[1,2,3], Curtis Huttenhower[1,2,3,5], Dayakar V. Badri[4]

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

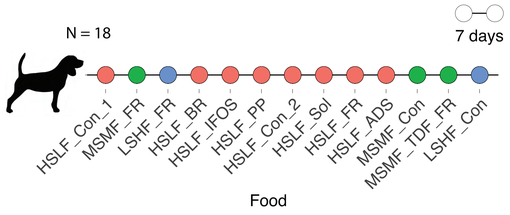
Dietary fiber and the microbiome

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

Design of dietary fiber study

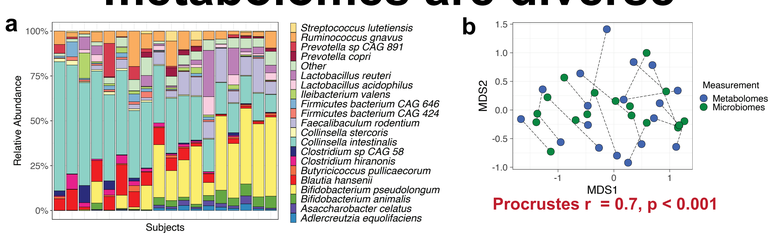


12 foods containing different sources and amounts of fiber were tested. They were classfied into 3 groups - high starch low fiber (HSLF) (n = 7), medium starch medium fiber (MSMF) (n = 3), and low starch high fiber (LSHF) (n = 2).



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

Baseline microbiomes and metabolomes are diverse

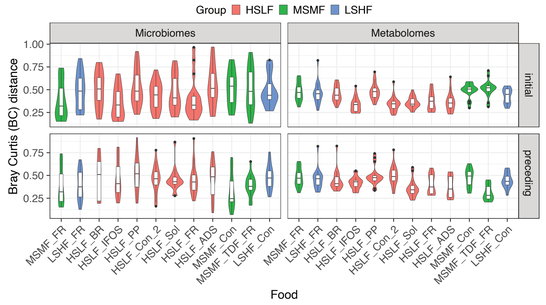


(**a**) *Collinsella* and *Bifidobacterium* sp were the most abundant species in baseline gut microbiomes of dogs followed by Firmicutes(36.98%) and Bacteroidetes(3.23%), Proteobacteria(0.44%), and Fusobacteria (0.005%). (**b**) Abundances of 818 metabolites were assayed from the same samples. Inter-individual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

Fiber affects the metabolome more than microbiome composition

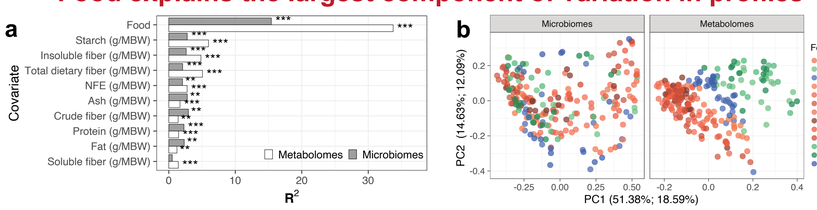
Similar foods give rise to similar metabolomes but not microbiomes

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



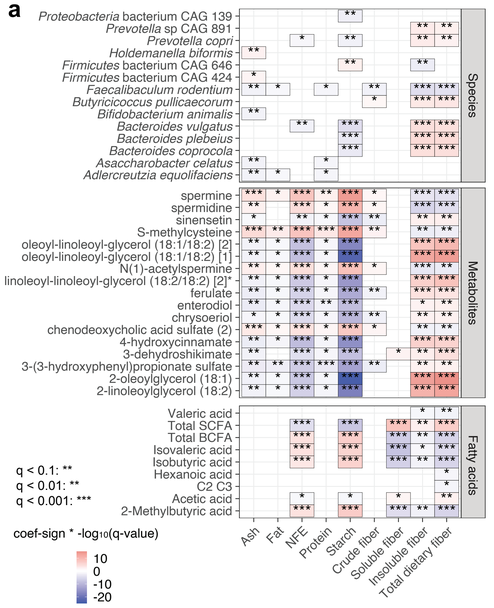
MSMF_FR microbiome was most similar to HSLF_Con1. Microbiomes in response to only 2 of the remaining 6 HSLF foods appeared to be similar to HSLF_Con1. Adjacent similar foods did not result in similar microbiomes. On the other hand, metabolomes following consumption of similar foods were similar.

Food explains the largest component of variation in profiles

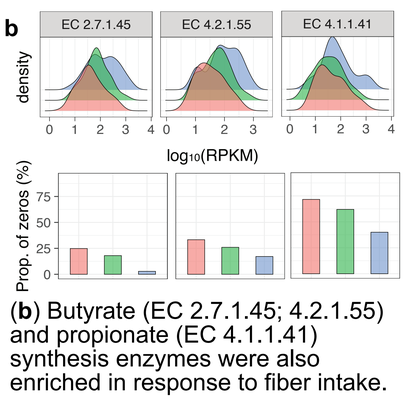


(**a**) Among the macronutrients, starch and insoluble fiber explained most of the variance in metabolomes. (**b**) Bray-Curtis distance based ordination shows similarity only of metabolomes in response to test foods (colored by group).

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



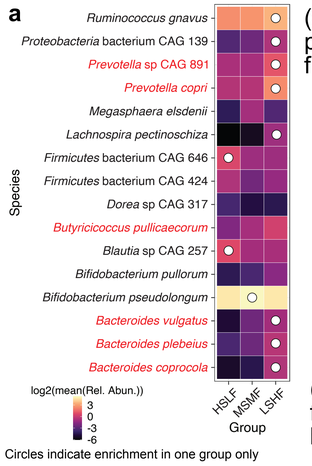
(**a**) Associations were determined using univariate linear models. 19 species, 569 metabolites, and 10 fatty acids were significantly ($q < 0.25$) associated with total dietary fiber. Fiber-enriched features included SCFA producers such as *Butyricicoccus* and *Bacteroidetes* sp, SCFAs, acylglycerols,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



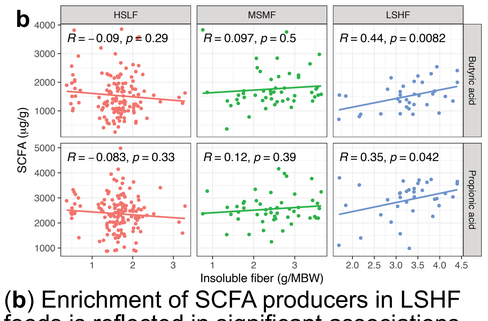
(**b**) Butyrate (EC 2.7.1.45; 4.2.1.55) and propionate (EC 4.1.1.41) synthesis enzymes were also enriched in response to fiber intake.

Strength of associations with fiber varies by food group and subject

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

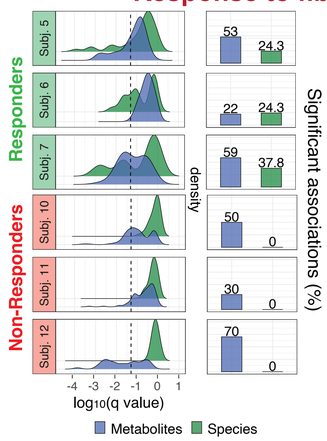


(**a**) One-way ANOVA shows that most SCFA producers that are positively associated with fiber (red) are more abundant in LSHF foods.



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

Response to fiber is highly personalized



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

Conclusions

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

Acknowledgements

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

http://huttenhower.sph.harvard.edu

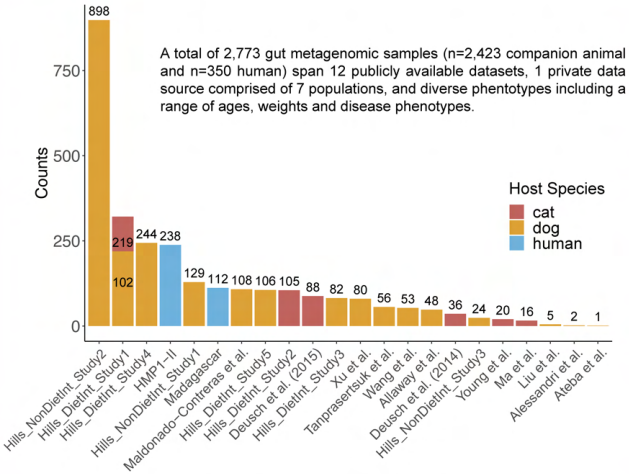
A comprehensive profile of the companion animal gut microbiome integrating reference-based and reference-free methods

Tobyn Branck[1,2,4], Zhiji Hu[1,2], William A. Nickols[1,2], Aaron M. Walsh[1,2,3], Amrisha Bhosle[1,2,3], Meghan I. Short[1,2,3], Jacob Nearing[1,2,3], Artemis Louyakis[4], Dayakar Badri[4], Christoph Brockel[4], Kelsey N. Thompson[1,2,3], Curtis Huttenhower[1,2,3,5]

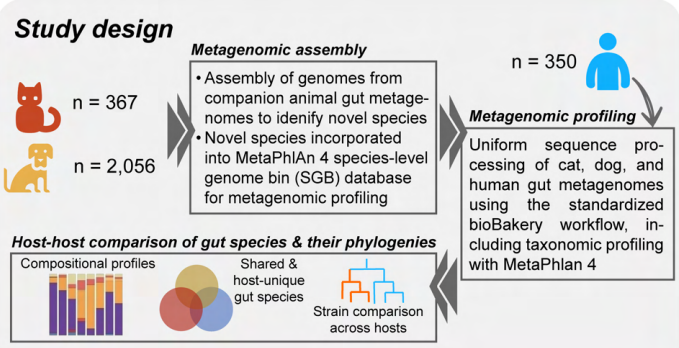
[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [2]Harvard Chan Microbiome in Public Health Center, Harvard T. H. Chan School of Public Health, Boston, MA, USA, [3]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA,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

The gut microbiome of companion animals is relatively underexplored, despite its relevance to animal health, human owner health, and basic microbial community biology. Here, we provide the most comprehensive analysis of the canine and feline gut microbiome to date, incorporating 2,423 stool shotgun metagenomes (2,056 dog and 367 cat) spanning 12 publicly available datasets (n=513) and 7 populations new to this study (n=1,910). These are compared with two human populations of 238 baseline gut metagenomes from the Human Microbiome Project 1-II and 112 gut metagenomes from a human population from Madagascar, both processed in an identical manner to the companion animal metagenomes. All microbiomes were characterized using both reference-based taxonomic and functional profiling, as well as de novo assembly and metagenomic assembled genomes (MAG) clustered into species genome bins (SGBs). We identified SGBs that are shared across hosts, unique to companion animals, and lastly those that are host specific. Among shared SGBs, we observed evidence of niche (host) specialization as well as varying degrees of interhost transfer.

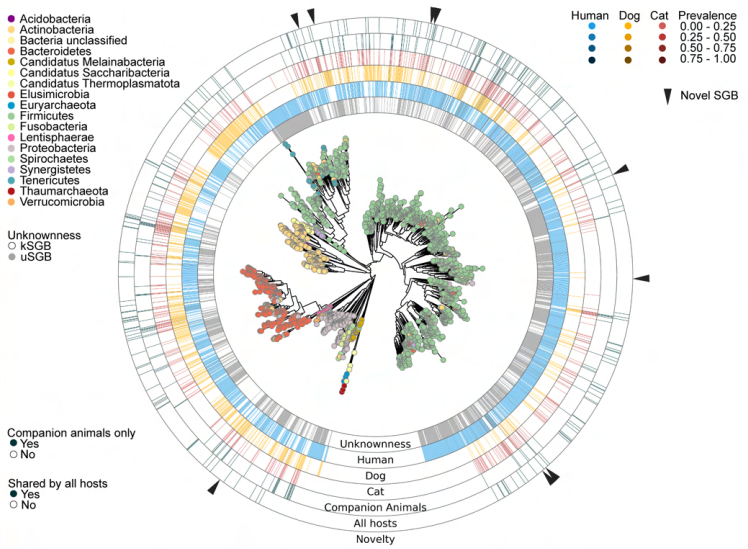
Samples & study design for large-scale companion animal microbiome analysis



Study design

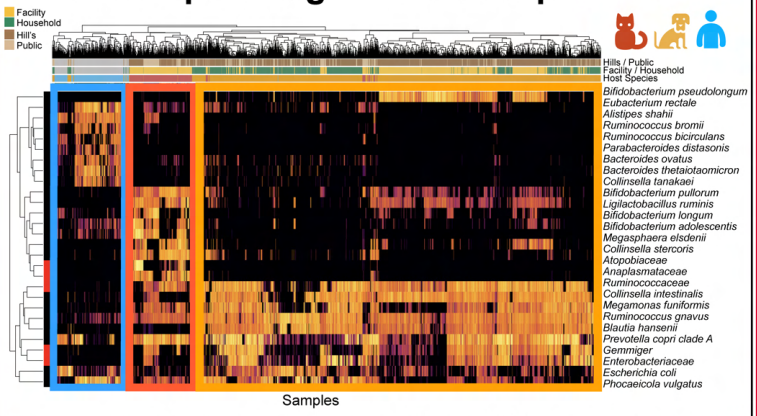


Novel, shared, and unique gut microbes in companion animal and human hosts



Phylogenetic relationship of genomes representing 2,364 SGBs identified in companion animal and human gut microbiomes. Of these, 1,045 are considered to be "known" SGBs (kSGBs) - those that include at least one genome with prior taxonomic assignment - while the remaining 1,310 are considered "unknown" SGBs (SGBs without confident taxonomy). Notably, 9 novel SGBs (distinct from SGBs in the current database) were identified in our dataset through metagenomic assembly.

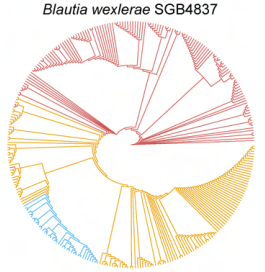
Host-specific gut microbial profiles



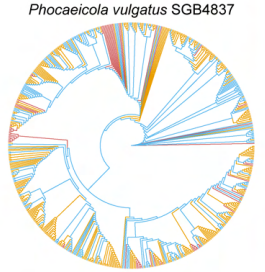
Samples (taxonomic profiles) clustered together with those from the same host species. Certain species (e.g., *Megasphaera elsdensii* and *Collinsella tanakaei)* were present and abundant in companion animals but absent in humans. Other species (e.g., *Ruminococcus gnavus*, *Phocaeicola vulgatus*) were present in all hosts.

Lineage-specific divergence within gut species shared among hosts

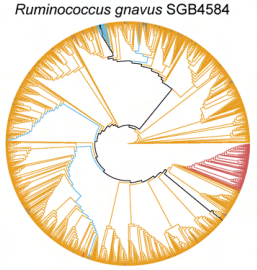
Blautia wexlerae SGB4837



Phocaeicola vulgatus SGB4837

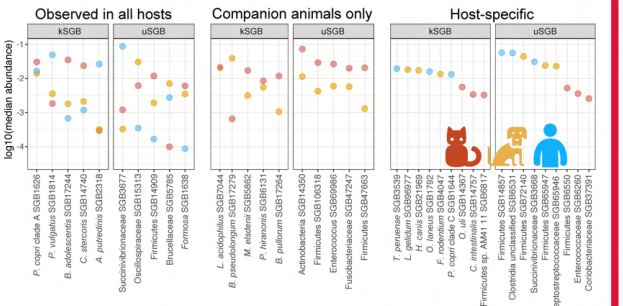


Ruminococcus gnavus SGB4584



We observed patterns of genome similarity between strains from different hosts that suggest evidence of 1) lineage-specific evolution of microbial species and 2) interhost transfer. In the former, distinct clustering of strains belonging to SGBs such as *Blautia wexlerae* SGB4837 and *Ruminococcus gnavus* SGB4584 suggests host-specific adaptation of microbial species. On the other hand, SGBs such as *Phocaeicola vulgatus* SGB4837 had relatively higher genomic similarity amongst strains from different hosts, potentially a result of frequent cross-host transfer.

Variable host-specificity of microbes



The SGBs were either host-unique (#SGBs unique in cats: 46, dogs: 277, humans: 1451), shared only by companion animals (189 SGBs), or shared across companion animals and humans (184 SGBs). When shared, some SGBs were more abundant in certain hosts, such as *Bifidobacterium adolescentis* SGB17244 or *Alistipes putredinis* SGB2318, suggesting differential niche (host) specialization. In some cases, shared SGBs such as *Prevotella copri* clade A SGB1626 had similar abundance across all hosts.

Acknowledgments

We would like to thank Hill's Pet Nutrition and the National Institute of Health for funding this work. We would also like to thank all of the pet partners and those that take care of them.

For questions, please contact Tobyn Branck at tob555@g.harvard.edu
http://huttenhower.sph.harvard.edu

Using machine learning and longitudinal multi-omics microbiome data to predict celiac disease development

Ivan Duran[1,2,3], Maureen M. Leonard[1,2], Alessio Fasano[1,2], Ali R. Zomorrodi[1,2]

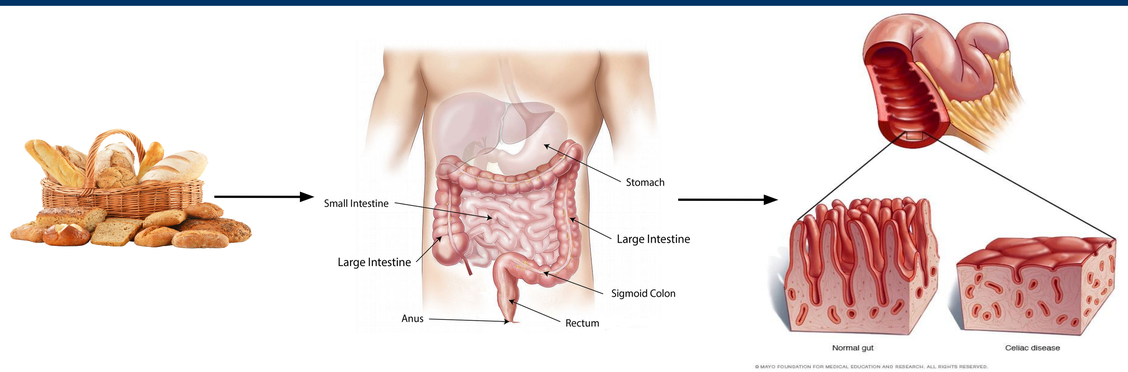
MassGeneral Hospital for Children
Mucosal Immunology and Biology Research Center

[1]Mucosal Immunology and Biology Research Center, Pediatrics Department, Massachusetts General Hospital, [2]Harvard Medical School, [3]Harvard FAS

1. Abstract

The gut microbiome is intrinsically dynamic and studies that collect longitudinal microbiome data to assess the dynamics of the gut microbiota during disease development or progression, or after a therapeutic intervention are increasing in frequency. However, efficient computational tools to harness multi-omics longitudinal microbiome data to predict clinical outcomes are underdeveloped. In this project, we aim to develop new machine learning (ML) tools to predict clinical outcomes by making use of time-series microbiome multi-omics data. As a case study, we used longitudinal metagenomic and metabolomic data from a prospective, longitudinal birth cohort study of children at high risk of Celiac Disease (CD) and sought to predict CD development in these subjects using pre-onset data. To this end, we trained Random Forest classifiers combined with an efficient feature selection scheme using several pieces of clinical metadata along with species, strains, pathways, and metabolites abundance data before disease onset as features (predictors). Our analyses revealed that clinical metadata alone are not accurate predictors of disease development (F1-score = 68.67%, 10-fold C.V.). However, we were able to achieve a high prediction performance of 93% (F1-score, 10-fold C.V.) using the abundance of only one pathway at 9 months of age and 100% (F1-score, 10-fold C.V.) using the abundance of only seven microbial strains at 15 months of age. This pilot study demonstrates the utility of ML for inferring key temporal microbiome signatures that are highly predictive of host clinical status. It also lays the foundation for building early predictive tools that would enable physicians to plan for preventive strategies before the clinical manifestation of disease.

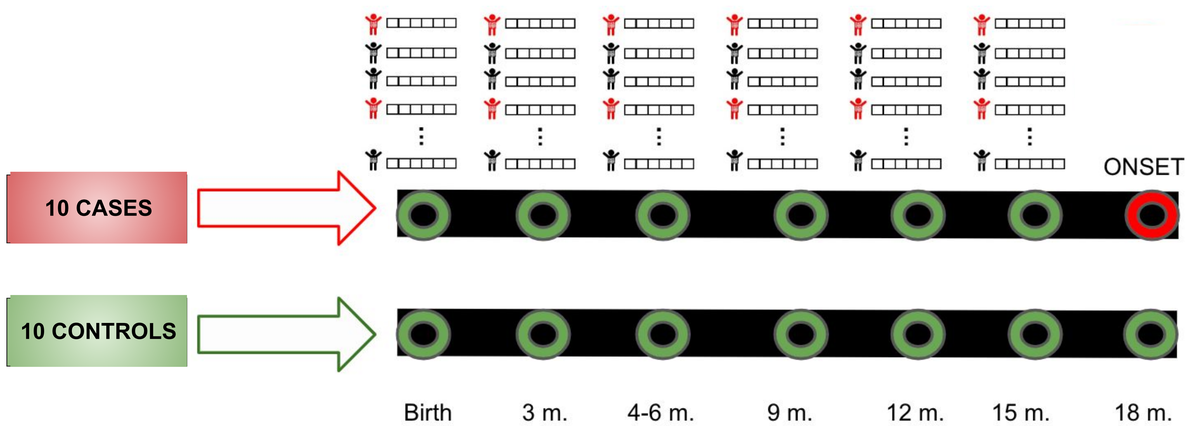
2. Celiac disease (CD) and the gut microbiome



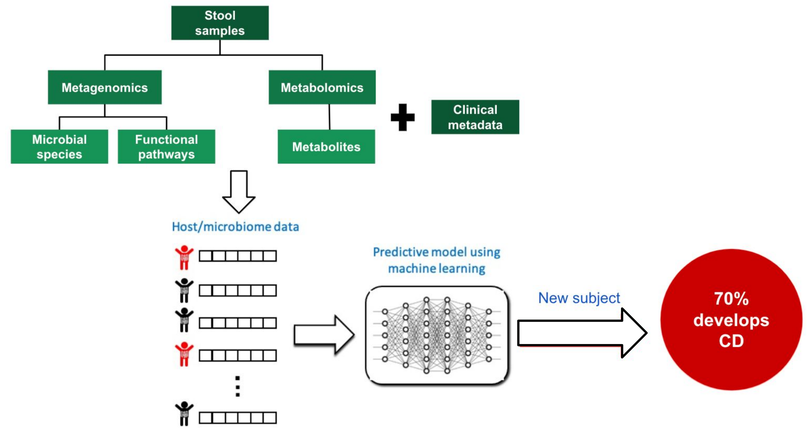
- CD is an autoimmune disorder where immune cells reacting to ingested gluten damage microvilli in the small intestine.
- ~2 million people in the US and 1% of the global population have CD
- Although genetic risk and gluten exposure are necessary, they are not sufficient to trigger the onset.

Recent studies show that the gut microbiome also has role in its pathogenesis

3. Stool samples from the CDGEMM study



4. Multi-omics microbiome data and clinical metadata to predict CD



5. Feature selection to remove features without predictive power

Too many features can worsen predictive power

50 features

20 subjects	Feat. 1	Feat. 2	Feat. #	Feat. 50
Subj. 1	###	###	...	###
Subj. 2	0	0	...	###
Subj. #	0	0	...	###
Subj. 20	###	###	...	###

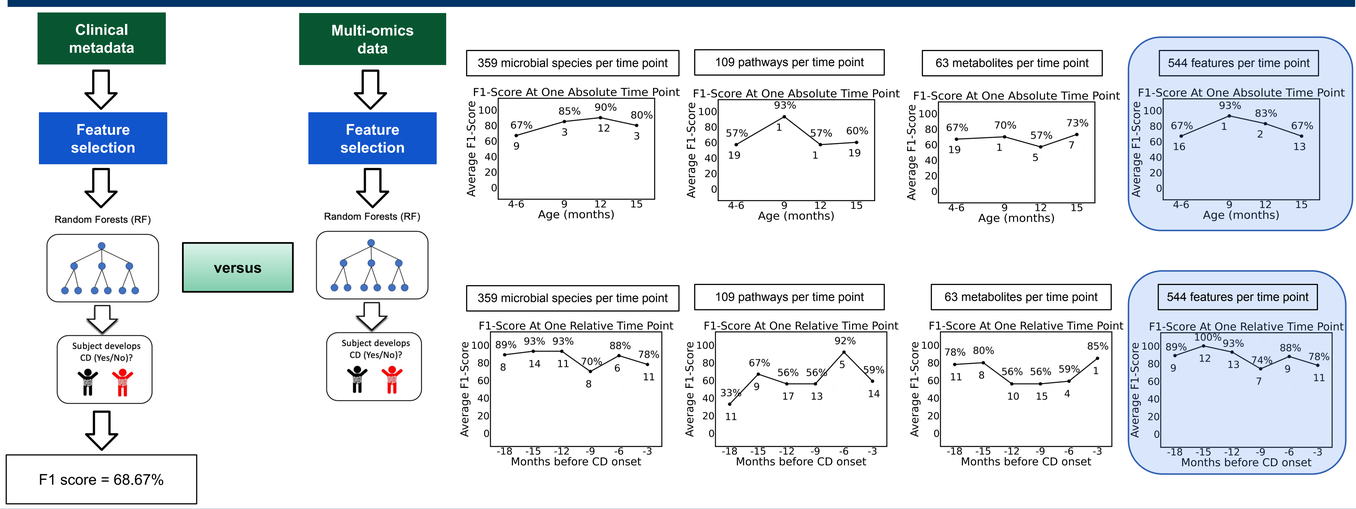
Feature selection

2 features

20 subjects	Feat. 2	Feat. 50
Subj. 1	###	###
Subj. 2	0	###
Subj. 3	0	###
Subj. 4	###	###

Feature selection can remove features and increase predictive power

6. Clinical metadata are not good predictors of CD but multi-omics data are



7. Summary and conclusions

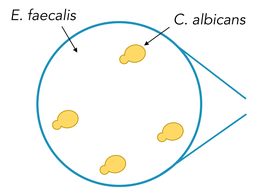
- A machine learning algorithm (Random Forests) was built based on fecal samples to predict who will develop CD
- Clinical metadata alone are not good predictors of CD onset
- We will scale up machine learning analyses and increase our findings using a more comprehensive, complex dataset

Acknowledgments: CDGEMM Team, Zomorrodi Lab Members

Cross-Kingdom Interactions between *Candida* albicans and *Enterococcus faecalis* in the Gut Microbiome

Haley E. Gause, Alexander Johnson
University of California, San Francisco

INTRODUCTION



Adult Gut Microbiome
- >1000 different species
- High bacteria, low fungi (2%)

Infant Gut Microbiome
- Low species diversity
- High fungal abundance
- Two common species:
  - *Candida species*
  - *Enterococcus faecalis (gram (+) bacteria)*

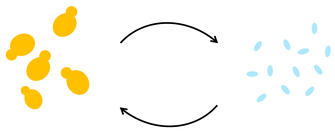
Relationship between *Candida albicans* and *Enterococcus faecalis* well documented

Mouse and Human Gut
After gut disruption, presence of Candida species associated with increased *E. faecalis* colonization (Zhai 2020, Mason 2012)

C. elegans
Colonization with EITHER *C. albicans* or *E. faecalis*
→ invasive growth, worm death
Colonization with BOTH species
→ stable gut colonization
(Cruz et al. 2013)

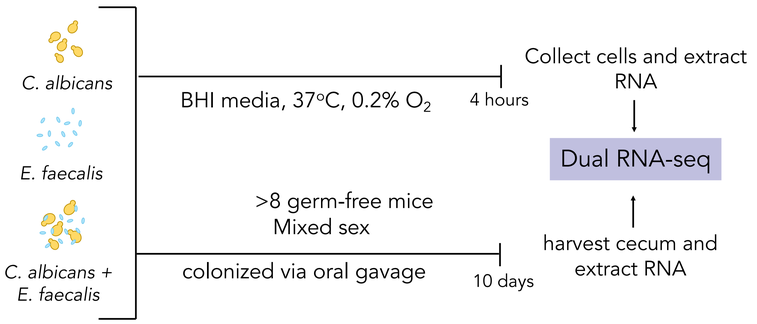
RESEARCH QUESTION

How do *Candida albicans* and *Enterococcus faecalis* interact as members of the Gut Microbiome?



Approach: Identify genes underlying interactions

METHODS - Dual RNA-Seq



RESULTS I – *C. albicans* has Robust Response to *E. faecalis*

Majority of the variance among samples is explained by presence/ absence of *E. faecalis (PC1 = 52%)*

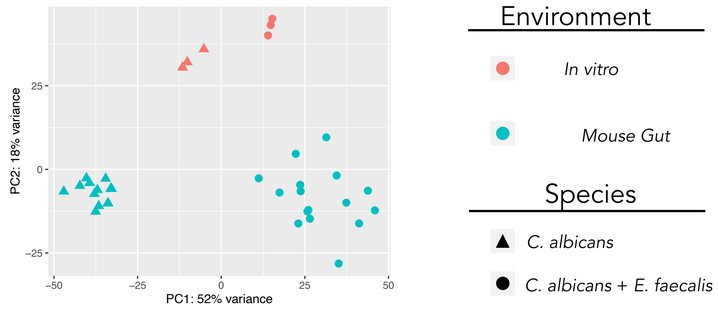


Figure 1: Principle-component analysis (PCA) plot showing that the *Candida* RNA-seq patterns from the four different conditions are very well separated from each other and highly reproducible within each condition.

C. albicans Genes ↑ ≥ 4-fold w/ *E. faecalis*

Significant overlap between *C. albicans* response to *E. faecalis* in Mouse gut and *in vitro* conditions

Figure 2: Overlap of *C. albicans* genes upregulated ≥ 4-fold ($p < 0.01$) in presence of *E. faecalis* in mouse gut and in vitro experiments

RESULTS II – Pheromone Response ↑ w/ *E. faecalis*

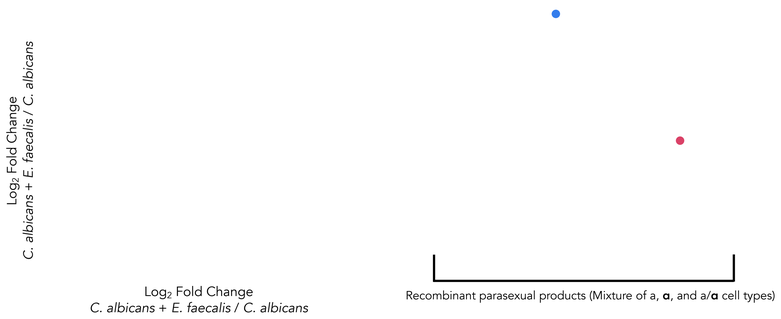


Figure 3: (Left) Expression of Upregulated C. albicans genes in mouse and *in vitro*. Colored dots correspond to steps in parasexual cycle in right panel. (Right) Diagram of the parasexual cycle in *C. albicans* (modified from Bennett (2015)).

E. faecalis induces genes in the parasexual cycle in *C. albicans*
Surprising because (1) mating incompetent a/α cells were used in co-cultures
(2) no precedent for bacteria or other species to induce parasexual cycle

RESULTS II – ↑ Transcription Factors w/ *E. faecalis*

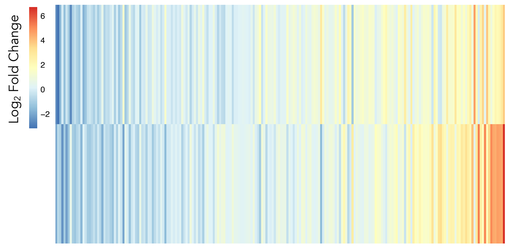
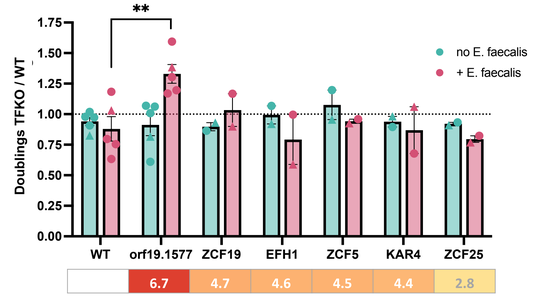


Figure 4: Heatmap of *C. albicans* expression of all 200+ Transcription Factors with *E. faecalis* compared to growth alone

orf19.1577	ZCF25
RON1	ZCF5
EFH1	ZCF26
WOR1	ZCF19
KAR4	

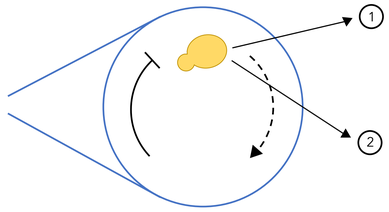
E. faecalis strongly upregulates a small number of transcription factors, many of which are uncharacterized and not expressed without *E. faecalis*. Additionally, in previously published screens, many of these transcription factors have no phenotype in a variety of lab conditions when deleted (Homann et al. 2009)



The most upregulated TF has a growth advantage over wildtype only in the presence of *E. faecalis*

Figure 5: Growth of Transcription Factor Knockouts (TFKOs) with and without *E. faecalis*, shown as normalized to internal WT controls.

CONCLUSIONS AND FUTURE DIRECTIONS



1. ↑ Parasexual Cycle
   → Investigate if canonical signaling pathway activates cycle or atypical activation
2. ↑ Transcription Factors
   → Use library of TF Knockouts to determine role in interactions with *E. faecalis* and genes regulated

ACKNOWLEDGEMENTS & REFERENCES

Thank you to the entire Johnson Lab for their advice and support. Additional thanks to the UCSF Gnotobiotics Core Facility and Jessie Turnbaugh for their expertise and services. Funding was provided by The Microbiology Society, and UCSF TETRAD, MPHD T32 , Discovery Fellowship, ASGD and the Graduate Division. References: (1) West et al. (2021) Microbiome. (2) Zhai et al. (2020) *Nat. Med.* (3) Mason et al. (2012) *Infect Immun.* (4) Cruz et al. (2013) *Infect Immun.* (5) Bennett, RJ. (2015) *Curr Opin Microbiol.* (6) Homann et al. (2009) *Plos Genetics.*

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

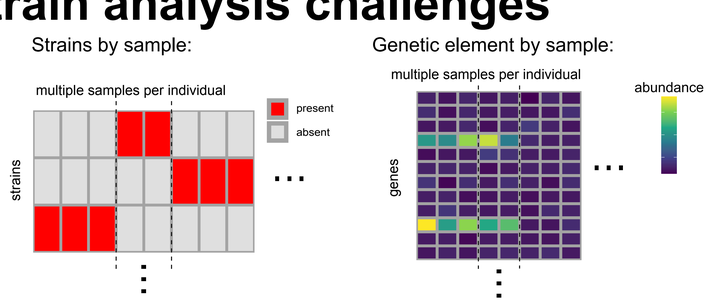
Andrew R. Ghazi[1,2], Yan Yan[1], Eric A. Franzosa[1,2], Curtis Huttenhower[1,2,3]

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

[3]Harvard Chan Microbiome in Public Health Center

Strain variation can strongly influence the impact of microbes on their environments, however methods for quantifying these important differences have been lacking. Sequencing-based microbiome data with strain-level resolution has several features that make traditional statistical methods challenging to use, including high dimensionality, individual-specific strain carriage, and complex phylogenetic relatedness. We present ANPAN, an R package that consolidates methods for strain statistics in three key components. First, adaptive filtering methods specifically designed to assess microbial strain profiles are combined with linear models to facilitate the identification of strain-specific genetic elements associated with host health outcomes. Second, phylogenetic generalized linear mixed models are used to characterize the effect of strain-level community structure. Finally, random effects models are used to account for species abundance when assessing the impact of pathway abundance on outcome status. We validated our methods by simulation, showing that we achieve improved estimation and classification statistics compared to current methodologies. We then applied our methods to a dataset of 1,262 colorectal cancer patients, identifying functionally adaptive genes and strong phylogenetic effects associated with CRC status.

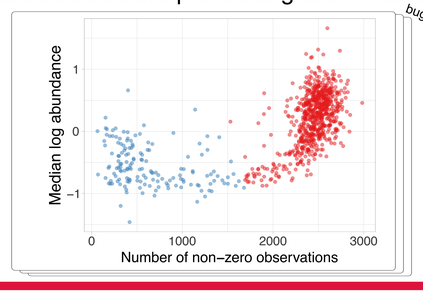
Strain analysis challenges



Microbial strains can be defined using arbitrarily specific nucleotide identity cutoffs. However, as a result of the finely-resolved nature of the data, unique strains rarely recur across individuals. Therefore, strain-level statistical modeling requires methods that can aggregate subspecies structure across samples, either by 1) inspecting genetic elements that recur across samples, 2) quantifying the similarity of strains across samples by phylogeny or 3) identifying differential pathway carriage by group.

Adaptive filtering of gene profiles

The genes of a given species may be poorly covered in a sample if the species is absent or insufficiently abundant. These samples can't provide information on the effects of genes in the species and should be discarded to avoid bias. Applying k-means clustering to summary statistics of each microbe in each sample allows assignment of whether the species' genes are well or poorly covered. Samples where the species is poorly covered are discarded before proceeding to the modeling step.



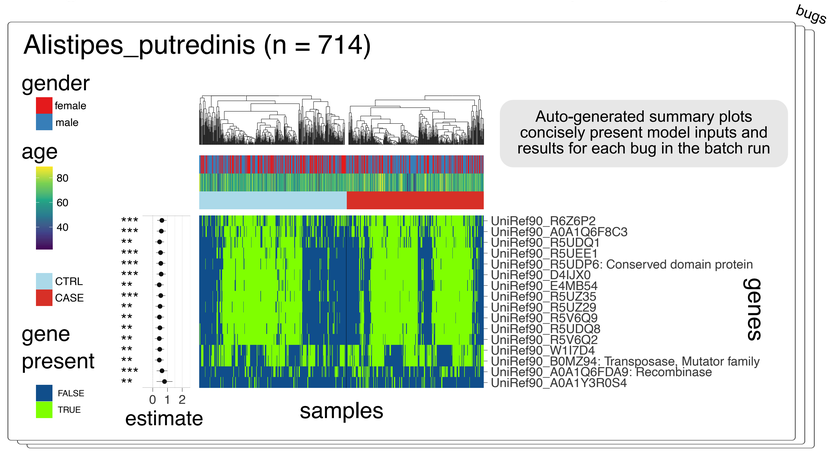
Adaptive filtering identifies species with well covered genes

well covered
poorly covered

Modeling associations of microbial genes with clinical outcomes

To identify genes associated with the outcome (CRC), the filtered data are analyzed alongside relevant covariates using either:

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

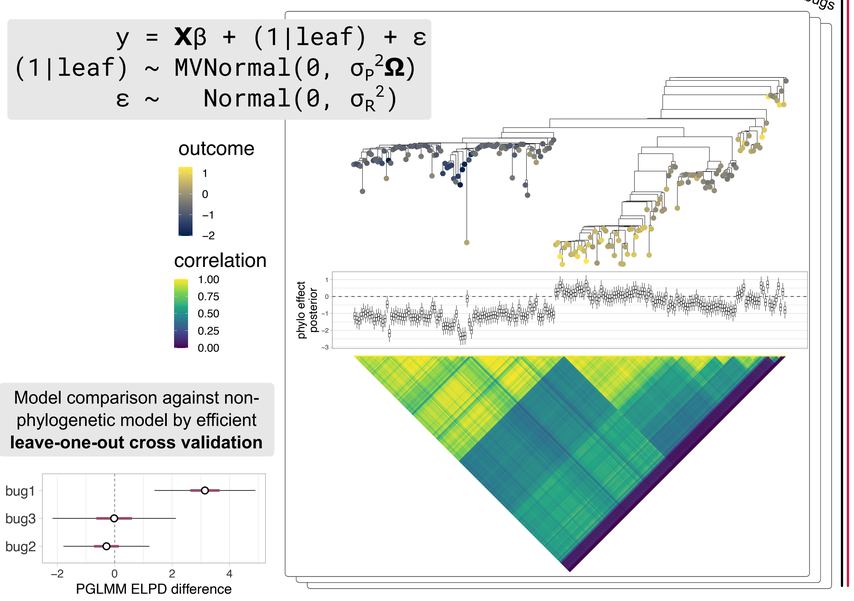


Auto-generated summary plots concisely present model inputs and results for each bug in the batch run

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 **Ω** implied by the inter-leaf distances of the tree.

$$y = \mathbf{X}\beta + (1|leaf) + \varepsilon$$
$$(1|leaf) \sim MVNormal(0, \sigma_P^2\mathbf{\Omega})$$
$$\varepsilon \sim Normal(0, \sigma_R^2)$$

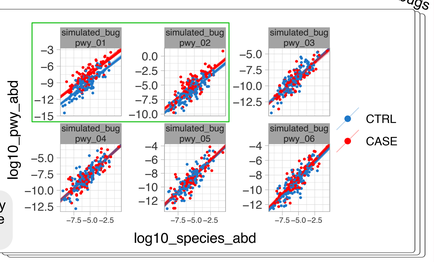


Random effects pathway model

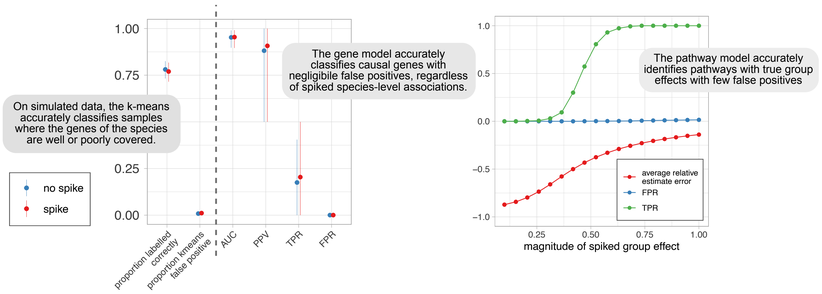
```
log_pwy_abd ~ log_species_abd + (1|pwy) + (0+group|pwy)
```

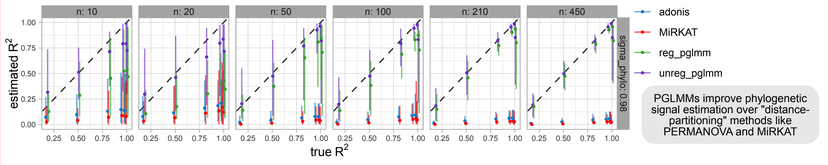
To infer the impact of a gene pathway on an outcome phenotype, a random effects model is used to assess the impact of group membership on pathway abundance while accounting for abundance of the relevant taxa.

Clearly separated posterior distributions by group indicate increased pathway carriage in the CASE group



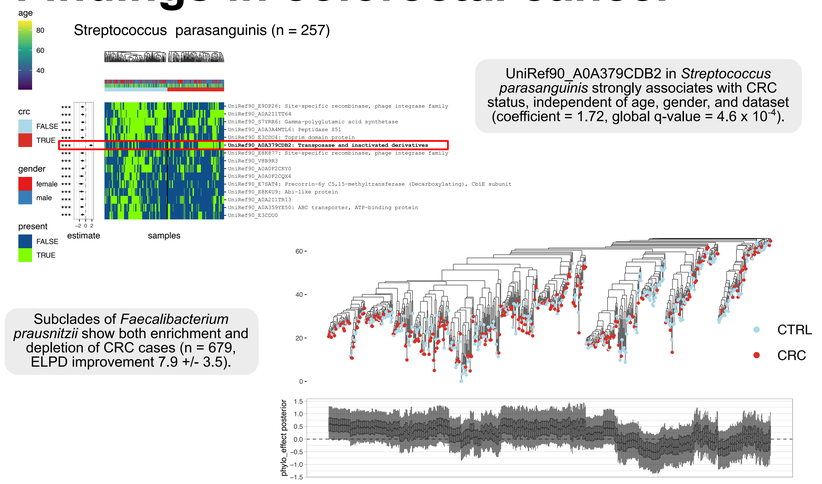
Simulation based evaluations





PGLMMs improve phylogenetic signal estimation over "distance-partitioning" methods like PERMANOVA and MiRKAT

Findings in colorectal cancer



Acknowledgments

We appreciate the help the Stan Developer team and users on the Stan forums who helped the development of this work. This work was supported by NIH NIDDK R24DK110499.

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

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

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

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

MassGeneral Hospital for Children
Mucosal Immunology and Biology Research Center

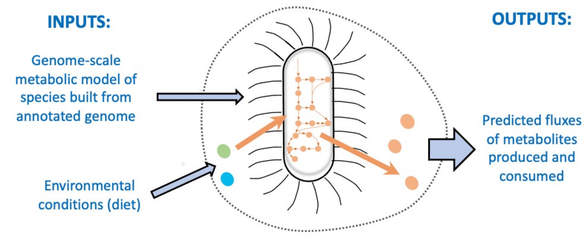
Background

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

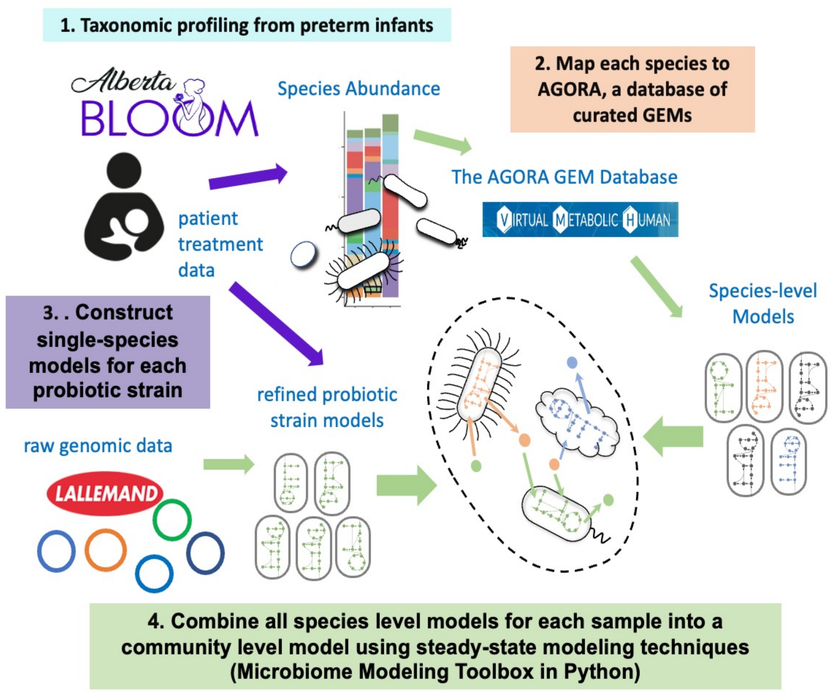
Objective: Use metagenomic and metadata from the BLOOM study to construct metabolic models of the preterm gut community to better understand the probiotic treatment's impact on the maturation of the preterm infant gut microbiota.

Genome-Scale Metabolic Models

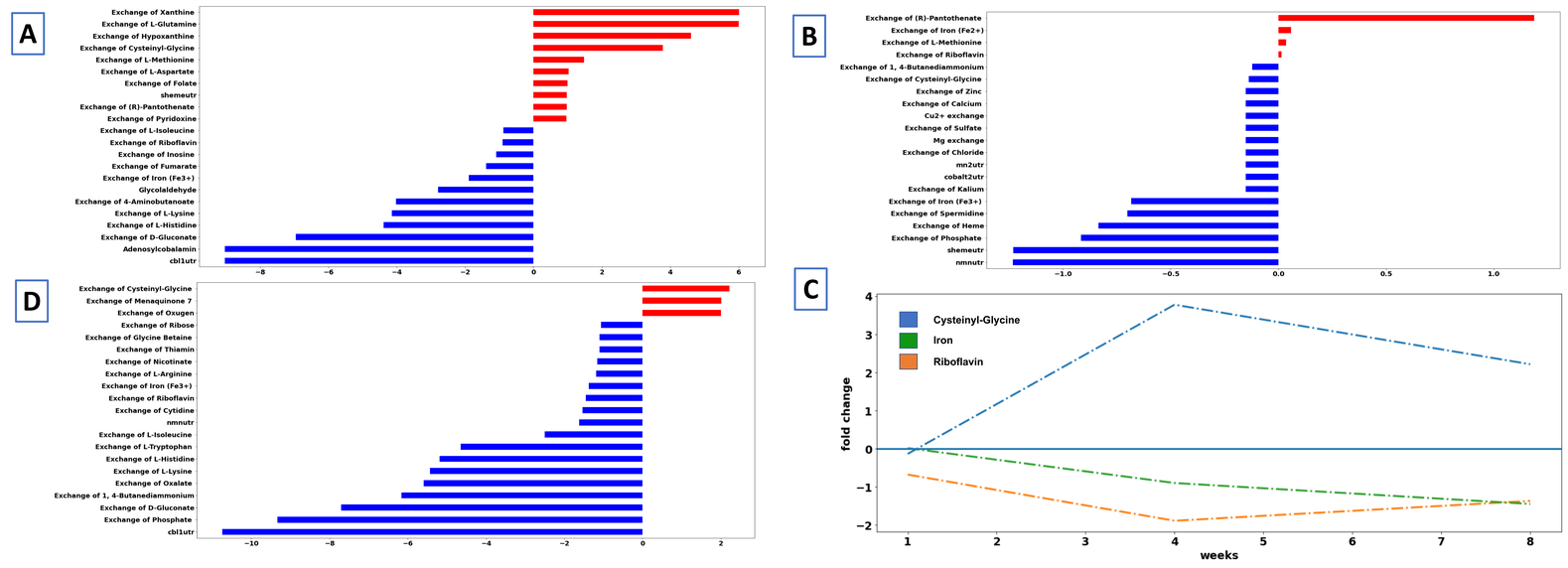
- Assume a cell can be approximated by the network of its metabolic pathways and can be analyzed to trace a metabolite's production back to a specific microbial species in the gut.



Methods

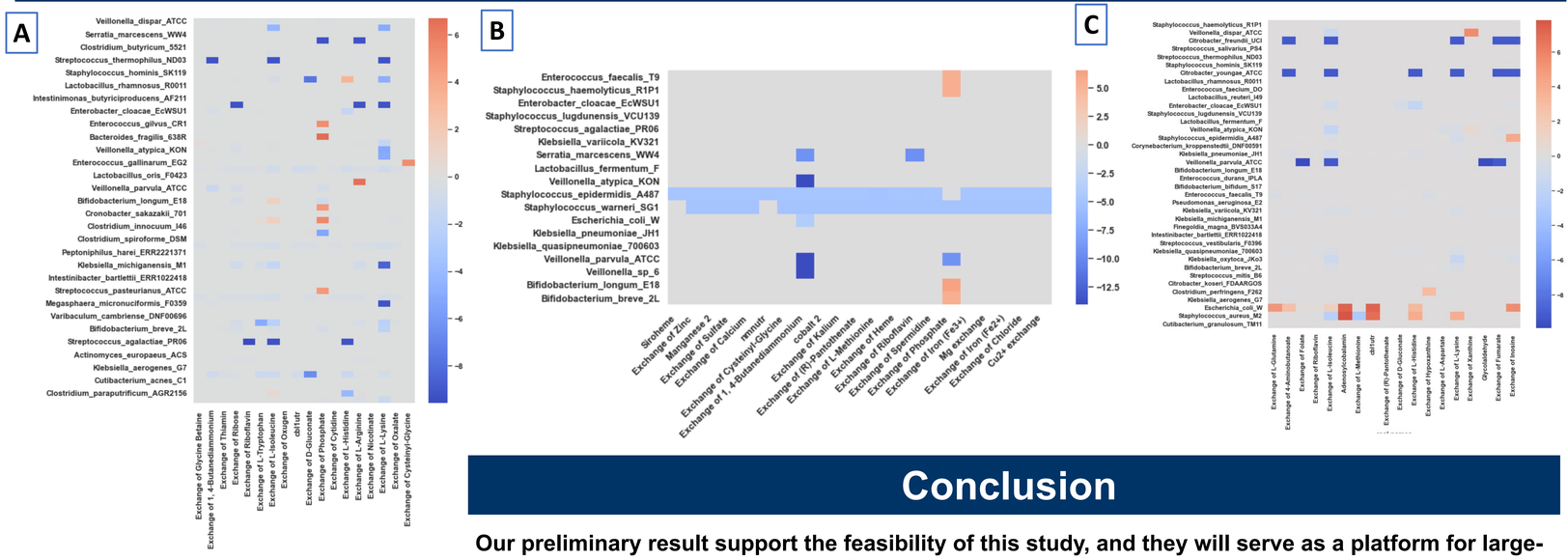


Differential Production of Metabolites on Probiotic Treatment



Community Models of weeks 1, 4, and 8 preterm infant fecal sampling A ,B,C are the log fold change of the top 20 differentially produced metabolites between infants on probiotic treatment and controls. They represent sampling from weeks 1,4, and 8 from birth, respectively. D represents the log fold change of three metabolites differentially produced over the 8 weeks.

Species-Metabolite Linkages of Differentially Produced Metabolites



Species-Metabolite Linkage Heatmaps

A is week 8 community modeling, B is week 1, and C is week 4. The metabolites shown are the top 20 differentially produced between probiotic and controls community models.

Conclusion

Our preliminary result support the feasibility of this study, and they will serve as a platform for large-scale computational studies of the function of probiotics on preterm infant microbiome development

Next Steps:

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

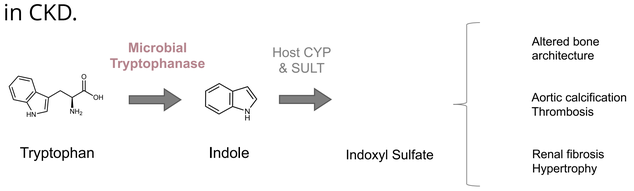
Gut Microbial Tryptophanase and the Uremic Toxins of Chronic Kidney Disease

Amanda L. Graboski[1], Mark E. Kowalewski[2], Joshua B. Simpson[2], Xufeng Cao[3], Jianan Zhang[2], Daniel P. Flaherty[3], Matthew R. Redinbo[2]

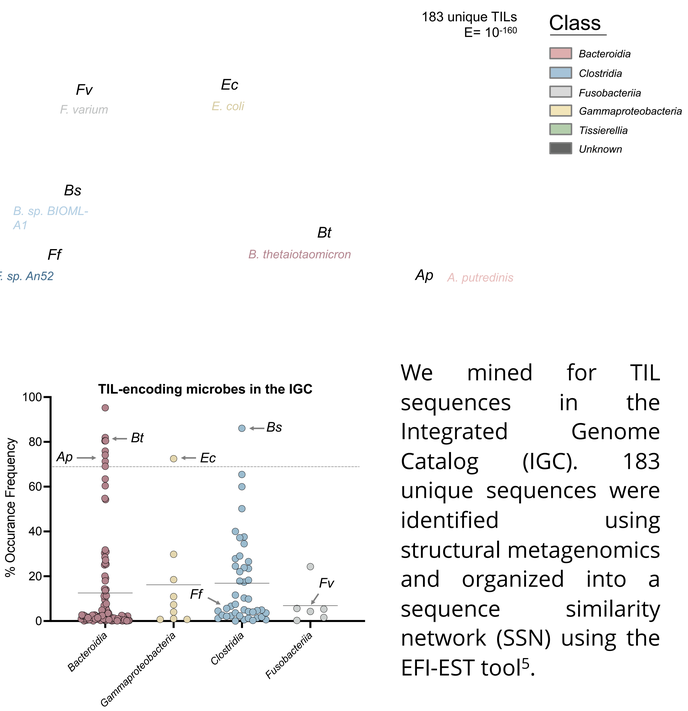
1 Department of Pharmacology, University of North Carolina, Chapel Hill, NC
2 Department of Chemistry, University of North Carolina, Chapel Hill, NC
3 Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN

Introduction

Chronic kidney disease (CKD) afflicts nearly 800 million people worldwide and is one of the fastest growing causes of mortality[1]. A key consequence of a diseased kidney is the serum retention of toxic compounds, known as uremic toxins, that have a broad impact on human physiology. One of the most damaging uremic toxins is indoxyl sulfate (IS), a metabolite produced through the gut microbial metabolism of tryptophan by the enzyme tryptophanase (TIL)[2]. Previous studies reveal that the genetic elimination of TIL in an artificial microbiome of mice resulted in no detectable serum levels of indoxyl sulfate and reduced kidney injury[3,4]. Here we use classical biochemistry, protein crystallography, and medicinal chemistry techniques to reduce the production of IS with application for reducing uremic toxicity in CKD.

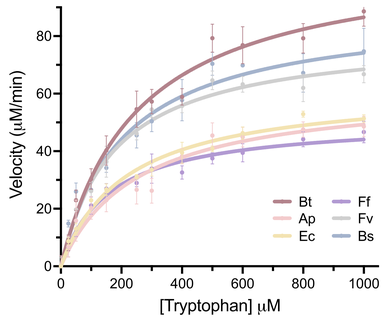


Defining the "Tryptophanase-ome"



We mined for TIL sequences in the Integrated Genome Catalog (IGC). 183 unique sequences were identified using structural metagenomics and organized into a sequence similarity network (SSN) using the EFI-EST tool[5].

Minimal functional and structural differences across diverse TILs

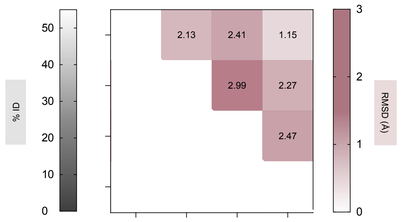


	Bt	Ap	Ec	Fv	Ff	Bs
K_M (μM)	260 ± 13	260 ± 31	215 ± 27	177 ± 16	163 ± 23	206 ± 30
V_{max} (μM/min)	125 ± 8	62 ± 4	62 ± 2	81 ± 4	51 ± 3	90 ± 9
k_{cat} (min^{-1})	54 ± 4	31 ± 2	89 ± 3	81 ± 3	26 ± 2	45 ± 4
k_{cat}/K_M (μM^{-1} min^{-1})	0.21 ± 0.02	0.12 ± 0.01	0.42 ± 0.05	0.46 ± 0.04	0.16 ± 0.02	0.22 ± 0.01

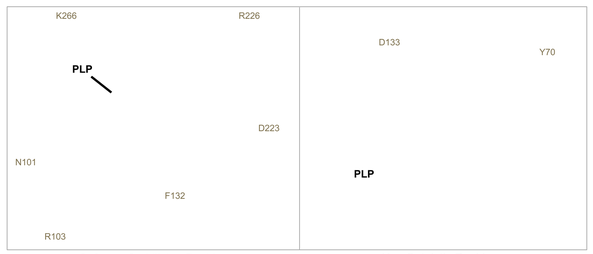
N=3 biological replicates. Ordinary one-way ANOVA with Tukey's multiple comparison.

TILs in the gut microbiome exhibit minor differences in functional activity despite deriving from diverse taxa. K_{cat}/K_M values are maximally 2- to 3-fold different and there is no significant difference in K_M.

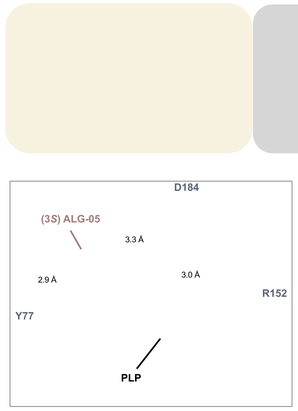
Despite low sequence identity (32-49%) between structures, the active site architecture and carbon backbone alignment (1.15 - 2.99 Å) of TILs are remarkably conserved across the human gut microbiome.

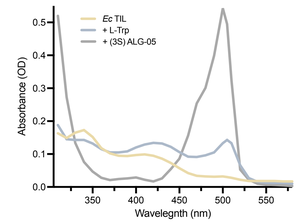


Bt *Bs*



Fv *Ec*

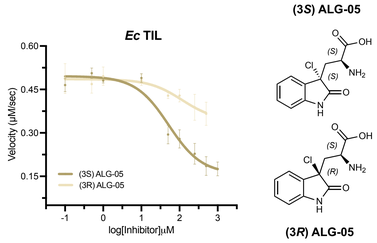




(3S) ALG-05 binding pose reveals introduced halogen bonding interactions with Y77 and R152 that stabilize the quinonoid complex.

Bs TIL co-crystalized with (3S) ALG-05. 2.07 Å resolution, [Fo-Fc] maps contoured at σ=1

Optimization of a pan-acting, transition state analog of TILs



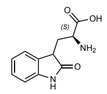
(3S) ALG-05

(3R) ALG-05

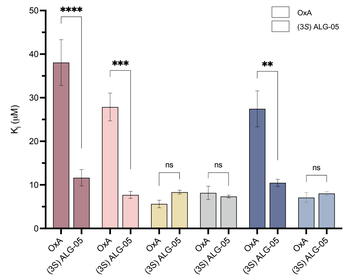
Oxindolyl-L-alanine (OxA), a pre-existing TIL inhibitor, displays variable activity across diverse TILs. OxA was used as the starting scaffold for our medicinal chemistry campaign which resulted in the identification of ALG-05. C3 stereochemistry plays a vital role in potency.

N=3 biological replicates. K_i values were calculated using IC_{50} values and the Cheng-Prusoff equation.

Enzyme	(3S) ALG-05 K_i (μM)	Affinity Fold-Change (L-Trp)
Bt	11 ± 2	24x
Ap	7.7 ± 0.8	34x
Ec	8.4 ± 0.5	26x
Fv	7.3 ± 0.3	24x
Ff	10 ± 1	16x
Bs	8.0 ± 0.5	26x

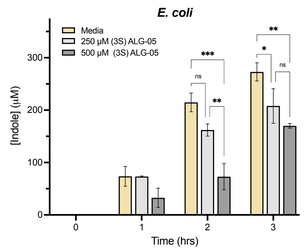


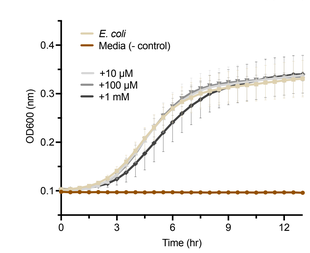
Oxindolyl-L-alanine

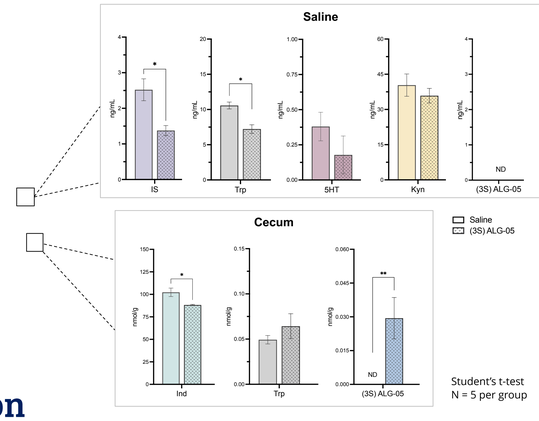


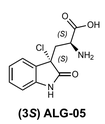
(3*S*) ALG-05 is non-lethal to microbes and lowers IS levels *in vivo*

N=3 replicates, (±SEM)









(3S) ALG-05

Student's t-test
N = 5 per group

Conclusion

Gut microbial TILs display nearly identical structural and functional characteristics despite harboring low sequence identity and deriving from diverse taxa. Here, we leverage this homogeneity to aid in the creation of a pan-acting transition state analog. (3*S*) ALG-05 is non-lethal to microbes at physiologically relevant doses and successfully reduces serum IS levels in mice. Thus, it represents a promising targeted therapeutic to reduce gut-derived uremic toxins in CKD.

References

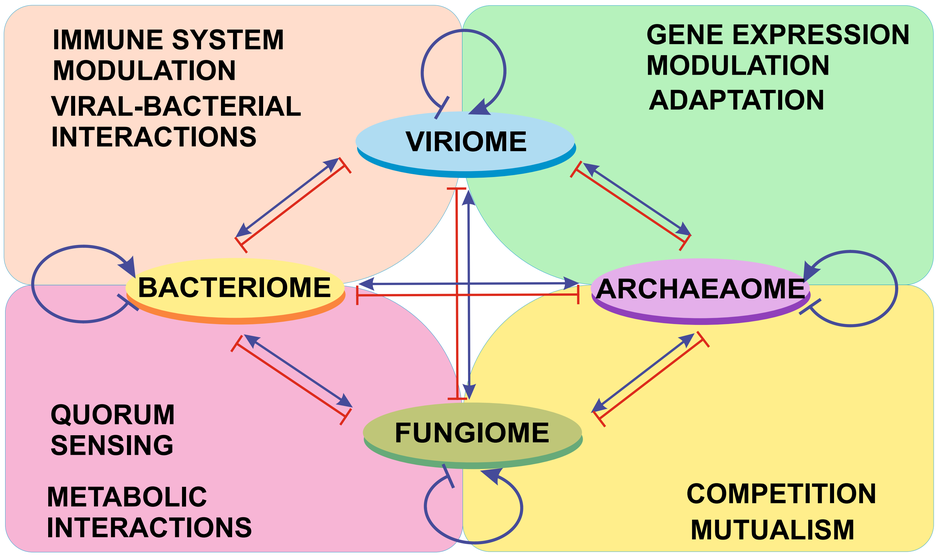
1. Mills et al, *Kidney Int*, 2010.
2. Vanholder et al, *Toxins*, 2018.
3. Devlin et al, *Cell Host Microbe*, 2016.
4. Lobel et al, *Science*, 2020.
5. Zallot et al, Biochemistry, 2019.

WHAT DOES THE RATIO BETWEEN BLOOD BACTERIAL AND FUNGAL MICROBIOME ABUNDANCE TELL US ABOUT FUNGAL-BACTERIAL INTERACTIONS?

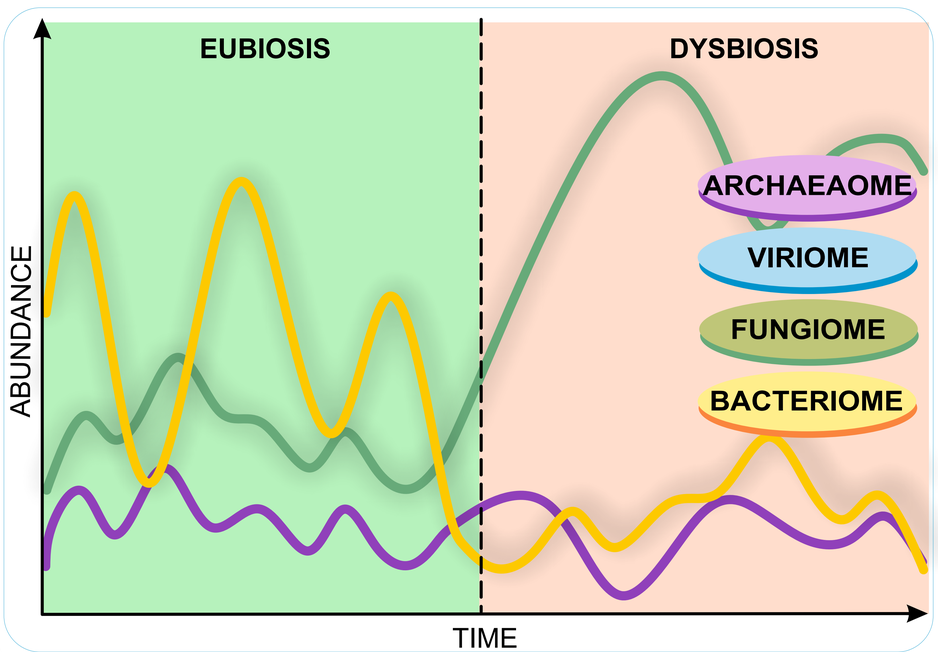
Yordan Hodzhev, Borislava Tsafarova, Vladimir Tolchkov, Reni Kalfin, Stefan Panaiotov
National Center of Infectious and Parasitic Diseases, Bulgaria, e-mail: jordanqvo@gmail.com Institute of Neurobiology, Bulgarian Academy of Sciences, Sofia, Bulgaria

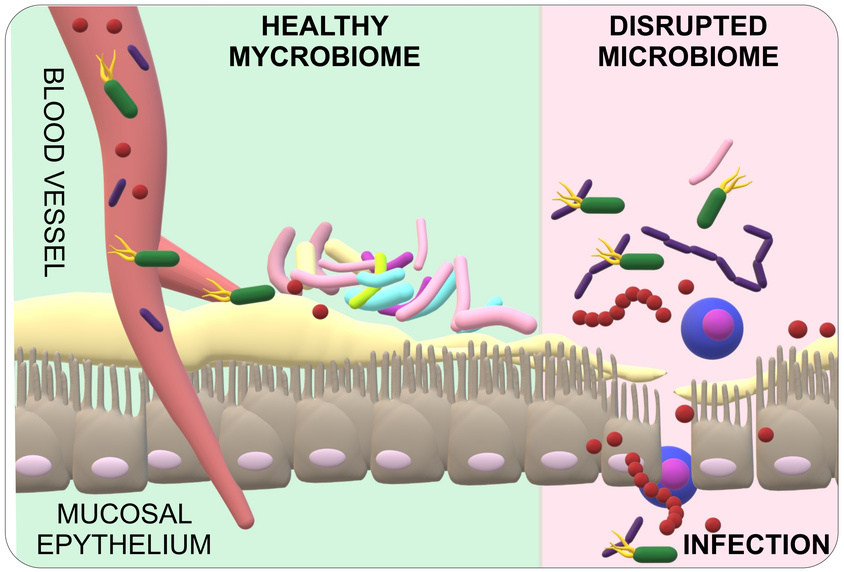
BACKGROUND

The human body is home to a diverse range of microorganisms, including bacteria (**bacteriome**), fungi (**fungiome**), archaea (**archaeaome**), and viruses (**viriome**, including **phageome**). These microorganisms are collectively referred to as the human **microbiome.** They interact in various ways, and these interactions can have significant impacts on human health.



The dynamics of interactions among bacteria, fungi, archaea, and viruses within the human body are complex and multifaceted. They can be influenced by various factors, including the host's genetics, diet, age, environment, and lifestyle.





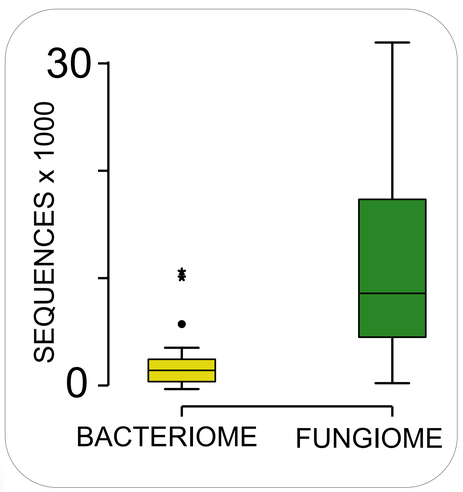
AIM

The aim of the present meta-analysis was to explore possible interactions between microbial and fungal communities. Blood group and gender data were included to assess the findings' biological relevance.

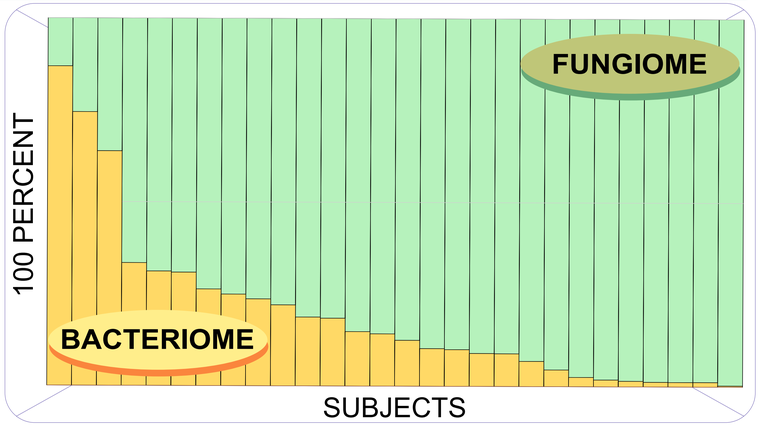
METHODS

3 ml of venous whole blood was collected from 28 subjects (14 females, 7 of each blood group - A, B, AB, O). Blood was lysed in d. water and the human DNA was treated with DNase. Microbial DNA was isolated by applying treatment with 4% SDS for microbial lysis. I solated DNA was divided into two subsamples and 16S and ITS metagenomic analysis was applied for each subject. Microbial total and relative abundance were calculated. Then the bacterial vs. fungal (B/F) reads ratio was analyzed. Data were subjected to nonparametric statistical evaluation (Kruskal-Wallis) of gender and blood group effects.

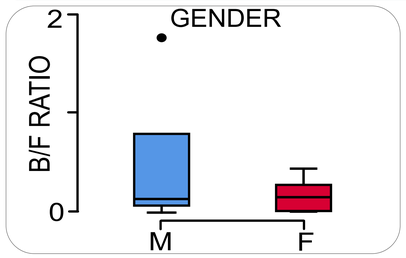
RESULTS



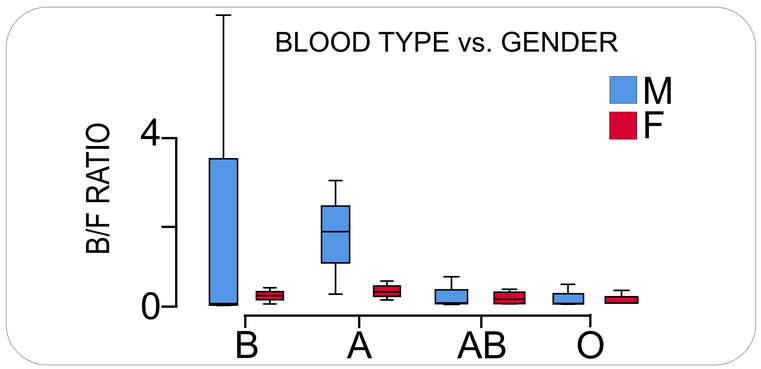
(1) The overall fungal sequence number (median=8579) was higher than the bacterial (median = 1062; Related-Samples Wilcoxon Signed Rank Test, Z =383; P<0.001).



(2) Individually, the B/F ratio varied significantly across subjects spanning from full fungal dominance to an almost complete lack of fungal sequences.



(3) The mean B/F ratio was higher for males (mean B/F=0.95) as compared to females (mean BF = 0.18; P<0.001).



(4) The blood type had an impact on the B/F ratio. For individuals of blood groups, A and B the ratio were around 1 and 0.2 (P<0.05) for individuals of blood groups AB and O.

CONCLUSION

In conclusion, despite the overall fungal dominance the B/F ratio showed high individual variability ranging from almost full fugal dominance to negligible fungal presence. The dependence of the B/F by gender and blood group suggests that it reflects the physiological status of the host. It could be hypothesized that B/F could serve as a health diagnostic index. It is worth testing the therapeutic correction of B/F in clinical practice.

ACKNOWLEDGMENTS - This research was funded by the Bulgarian National Science Fund within National Science Program VIHREN, contract number KP-06-DV/10-21.12.2019.

Alterations in the Fungal Microbiome in Ulcerative Colitis

Katie Hsia, MD[1]; May Fu, DO[1]; Laleh Montaser Kouhsari, MD[1]; Khalid Algarrahi, MD[1]; Hannah Chen, MD, PhD[1,2]; Dominique Michaud, MD[2]; Sushrut Jangi, MD[1,2]

1. Tufts Medical Center, 2. Tufts University School of Medicine

Introduction

Background

- Gut fungi, especially **Candida**, is known to drive immunogenicity in mouse models of IBD
- Gut fungal communities in human IBD (ulcerative colitis) have not been well-characterized
- A prior study showed that **high Candida** in stool can prognosticate favorable responses to fecal microbial transplant

Aim

- To assess the diversity and differential abundance of the fungal microbiome in active and quiescent ulcerative colitis

Methods

Data Source

- The Study of a Prospective Adult Research Cohort with IBD from the Crohn's and Colitis Foundation
- Stool metagenomics, clinical metadata
- Internal Transcribed Spacer based deep sequencing of fungal rDNA

Data Analysis

- Alpha diversity: Observed, Shannon
- Beta diversity: Unifrac and NMDS
- Differential Abundance: DESeq2

Results

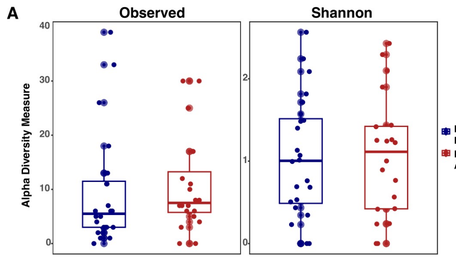
Study Cohort

		Endoscopic Status			Endoscopic Status			Biologic Use		
	Total Cohort	Activity	Remission	p	Activity	Remission	p	Exposed	Naive	p
n	98	25	28		19	23		66	31	
Age, mean (SD)	44.9 (14.3)	48.5 (13.4)	47.7 (15.7)	0.76	46.2 (14.5)	45.1 (12.1)	0.99	48.6 (13.8)	51.6 (14.9)	0.37
Gender, female, frequency	53%	52%	64.3%	0.36	68.4%	60.9%	0.61	50%	61.3%	0.65
Disease duration, years, median (IQR)	2 (0-4)	0 (0-4)	2 (0-4)	0.44	0 (0-1.75)	2 (0-4)	0.27	2.65 (0.25-4)	3 (0-4.25)	0.96
Fecal calprotectin, mean (SD),mg/g	34.9 (84.7)	99.3 (143.3)	8.78 (14.7)	<0.05	89.7 (152.1)	23.6 (49.1)	.09	39.3 (90.3)	40.93 (77.1)	0.26

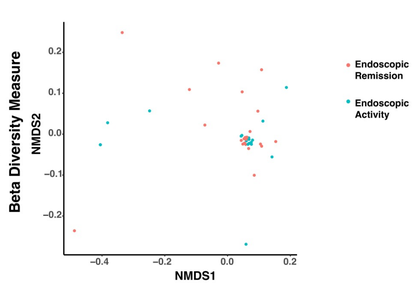
Differential Abundance in Active vs Quiescent UC

log2 fold change	p-adj	Phylum	Genus
4.54	$<1x10^{-4}$	Ascomycota	Saccharomyces
2.56	<0.05	Ascomycota	Candida
-4.94	$< 1 \times 10^{-6}$	Ascomycota	Penicillium

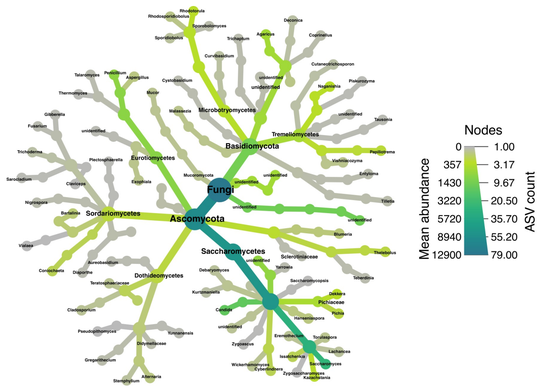
Alpha Diversity



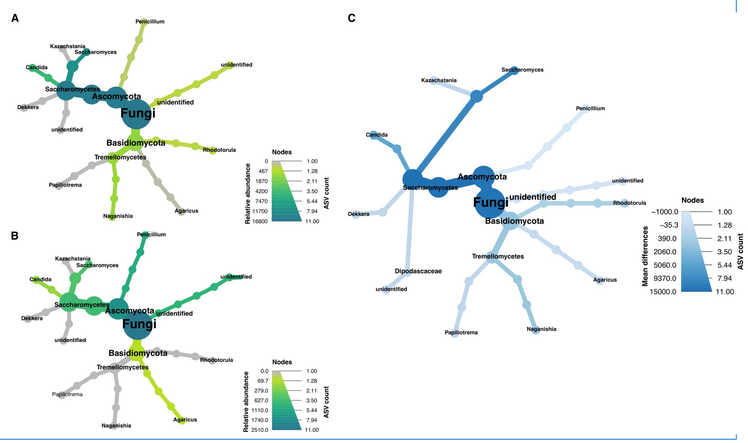
Beta Diversity



Mean Abundances of Fungal Genera in Total Cohort



Relative Abundances in Active vs Quiescent UC



Results

Alpha and Beta Diversity

- No significant differences in alpha or beta diversity of the fungal community were observed in active vs quiescent ulcerative colitis

Differential Abundance

- **Candida** and **Saccharomyces** had a significantly increased relative abundance in patients with active UC vs quiescent UC
- Increased **Candida** was also observed even after adjusting for age, gender, and immunosuppressive exposure

Take Away

- **Candida** and **Saccharomyces** may be linked to active inflammation in ulcerative colitis
- Elevations in **Candida** in active UC appear unrelated to immunosuppressive exposure
- Future studies in other cohorts may strengthen this association, allowing for development of personalized approaches to treating UC, including FMT, in patients with elevated Candida

Acknowledgements

Thank you to Crohn's and Colitis Foundation, the KL2 Career Develpoment Award Program, The Charlton Research Grant, and the Natalie V. Zucker Research Award. Thanks also to the Tufts Clinical and Translational Science Institute and the Tufts High Performance Computing Cluster.

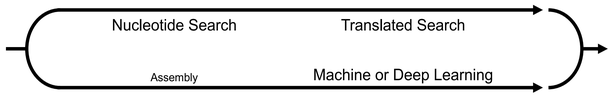
Integrating reference- and assembly-based methods for improved viral identification from microbial community sequencing

Jordan Jensen[1,2], Ya Wang[2,3,4], Moreno Zolfo[5], Philipp C. Münch[3,6], Nicola Segata[5], Eric A. Franzosa[2,3,4], Curtis Huttenhower[1,2,3,4]

[1]Department of Immunology and Infectious Diseases, Harvard University, Boston, MA, USA; [2]Harvard Chan Microbiome in Public Health Center, 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; [6]Department for Computational Biology of Infection Research, Helmholtz Center for Infection Research, Braunschweig, Germany.

BAQLaVa Methodology

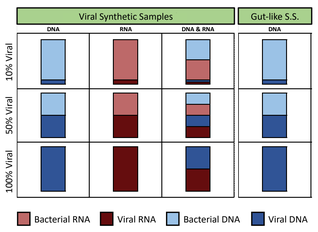
Capturing an accurate representation of the viral members of a microbial community presents significant experimental and computational challenges. To address these limitations, we developed **BAQLaVa** (**B**ioinformatic **A**pplication for **Q**uantification and **La**beling of **V**iral t**a**xonomy), which integrates both reference- and assembly-based methods to generate viral profiles from shotgun DNA or RNA sequencing. Here, we have evaluated BAQLaVa with 1) *in silico* simulated data representing virus across all viral realms, 2) synthetic gut viromes, and 3) human gut metagenomes and metatranscriptomes.



BAQLaVa employs a tiered reference-based search, first to a nucleotide database, and subsequently to a protein database. In parallel, reads are assembled into contigs and classified with a neural net trained to predict viral taxonomy at the genus level.

Evaluation of BAQLaVa with complex meta'omes

Right A set of viral synthetic meta'omes were created by clustering all RefSeq & GenBank viral genomes deposited after Jan 1, 2021. Gut-specific synthetic viromes were made from predicted viral genomes identified from a set of MAGs assembled from human gut metagenomes (Benler et al. 2021).



TPR, FPR, and F1 scores for synthetic meta'ome mapping to BAQLaVa databases

Left We observed high true positive and low false positive rates across nucleic acid types, compositions, and simulated sources from reads mapped to the BAQLaVa nucleotide and protein databases.

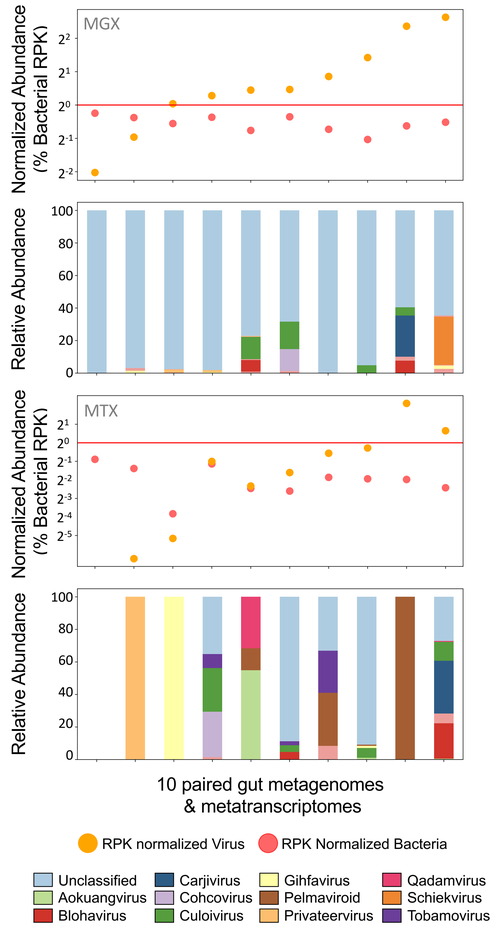
After this high-performance mapping, downstream filters are applied by BAQLaVa to prevent false-positive species observations.

The gut virome is abundant but underrepresented in databases

BAQLaVa identifies characterized viruses from RefSeq (nucleotide) and ICTV (translated) databases, as well as uncharacterized viruses by mapping to viral MAG databases (nucleotide).

We used BAQLaVa to profile ten metagenome (MGX, top) and metatranscriptome (MTX, bottom) samples from paired human gut samples (ibdmdb.org). Bacterial abundances were obtained for the same samples via upstream analysis with MetaPhlAn and HUMAnN.

Metagenome & metatranscriptome viral mapping through BAQLaVa



Our results show we can often capture a 1:1 and potentially higher virus:bacteria ratio with BAQLaVa.

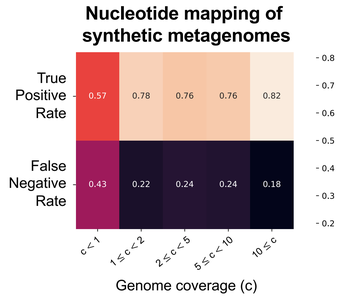
A large fraction of virus identified by BAQLaVa originates from viral MAG databases, indicating that an abundance of virus in the gut has not yet been well-studied. Use of novel databases can boost sensitivity and overcome this limitation that would otherwise severely restrict viral profiling.

CrAssphage are abundant in the gut: Among the shared genera observed highly present in both MGX & MTX samples were *Culoivirus*, *Blohavirus*, *Cohcovirus*, and *Carjivirus*, all members of the novel *Crassvirales* order.

Normalized Abundance: Bacterial and viral abundances in RPK were normalized to the total potential bacterial coverage (red line), which was calculated based on a model bacterial genome length of 4.6 Mbp (*E. coli*). **Relative Abundance:** Virome community profiles are shown at the genus level.

BAQLaVa combines approaches for improved performance

Nucleotide mapping of synthetic metagenomes

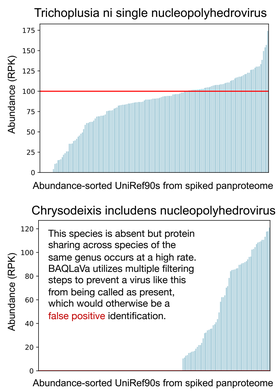


Left Nucleotide search as a function of genome coverage reveals that even at low coverages, BAQLaVa is able to report viral assignments for a majority of viral reads.

Multi-step translated approach prevents false positive calls

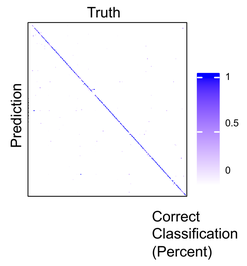
We defined an ICTV species' panproteome as all UniRef90 protein families that map to an ICTV genome with ≥ 80% protein coverage and ≥ 90% identity. Requiring 50% of a species' proteins to be detected avoids false positive detections from homology with minimal impact to sensitivity.

Right Protein set abundances from two species of the same genus, one present in a simple synthetic metagenome (top) and the other absent (bottom).



Deep learning complements limitations of reference-based viral search

Confusion Matrix for Genus-level Classification



Results from direct reference observation are complemented by taxonomic predictions generated via deep learning. This achieves an expanded viral profile.

Left A neural net, deepG, was trained to predict genus-level taxonomy for 474 ICTV genera (all with at least 10 unique species). Predicting on assembled contigs of ≥ 2kb produced the highest balanced accuracy across genera.

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 Astellas Pharma Inc.

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

Dissecting the role of the human microbiome in COVID-19 via metagenome-assembled genomes

Shanlin Ke[1]; Scott T. Weiss[1]; Yang-Yu Liu[1]

[1]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02115, USA.

Abstract

Coronavirus disease 2019 (COVID-19) is often accompanied by gastrointestinal symptoms. However, little is known about the relation between the human microbiome and COVID-19. Here we used whole-metagenome shotgun sequencing data together with assembly and binning strategies to reconstruct metagenome-assembled genomes (MAGs) from 514 COVID-19 related nasopharyngeal and fecal samples in six independent cohorts. We reconstructed a total of 11,584 medium-and high-quality microbial MAGs and obtained 5403 non-redundant MAGs (nrMAGs) with strain-level resolution. We found that there is a significant reduction of strain richness for many species in the gut microbiome of COVID-19 patients. The gut microbiome signatures can accurately distinguish COVID-19 cases from healthy controls and predict the progression of COVID-19. Moreover, we identified a set of nrMAGs with a putative causal role in the clinical manifestations of COVID-19 and revealed their functional pathways that potentially interact with SARS-CoV-2 infection. Finally, we demonstrated that the main findings of our study can be largely validated in three independent cohorts.

Introduction

To better understand the relationship between the human microbiome and COVID-19, we applied state-of-the-art metagenome assembly and binning strategies to reconstruct microbial population genomes directly from microbiome samples of COVID-19 patients and controls. Our major goals were to construct a COVID-19 related metagenomic genome catalog to identify novel taxa and strain-level differences that are likely related to the clinical manifestations of SARS-COV-2 infection.

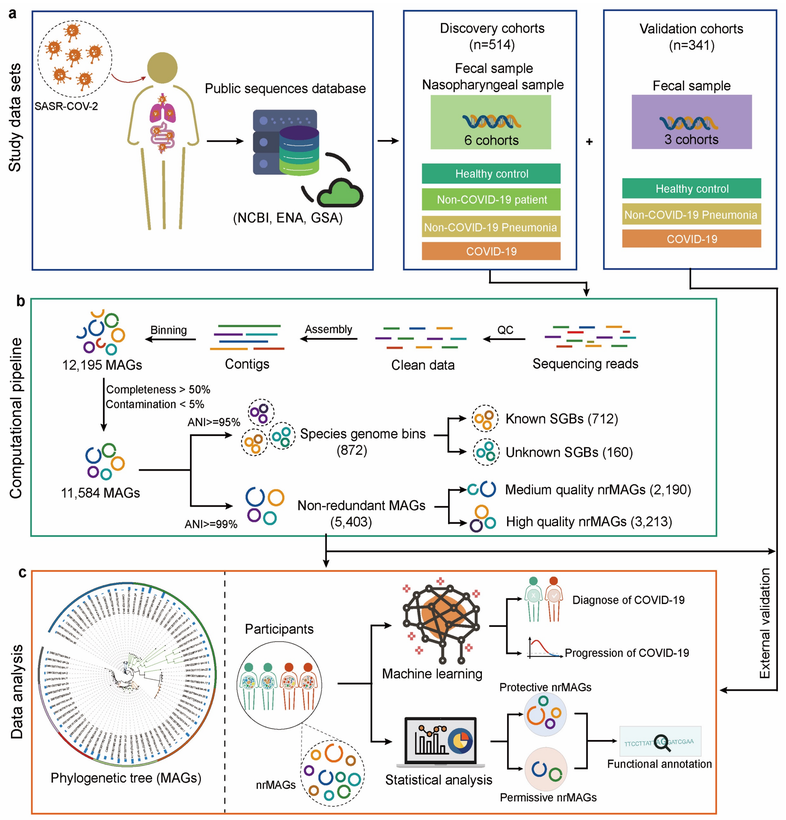


Figure 1. Conceptual framework of study.

Methods and Materials

We collected the raw WMS sequencing data of 514 microbiome samples (359 individuals) and 341 microbiome samples (278 individuals) from 6 and 3 publicly available datasets with different technical settings, respectively (**Fig.1**).

We applied state-of-the-art metagenome assembly and binning strategies to reconstruct microbial population genomes directly from microbiome samples of COVID-19 patients and controls in the discovery cohorts.

Results

- After quality control, we performed metagenomic assembly and binning on those microbiome samples from the discovery cohorts and recovered 11,584 MAGs, 872 SGBs, and 5403 non-redundant MAGs (nrMAGs) (**Fig.2**).

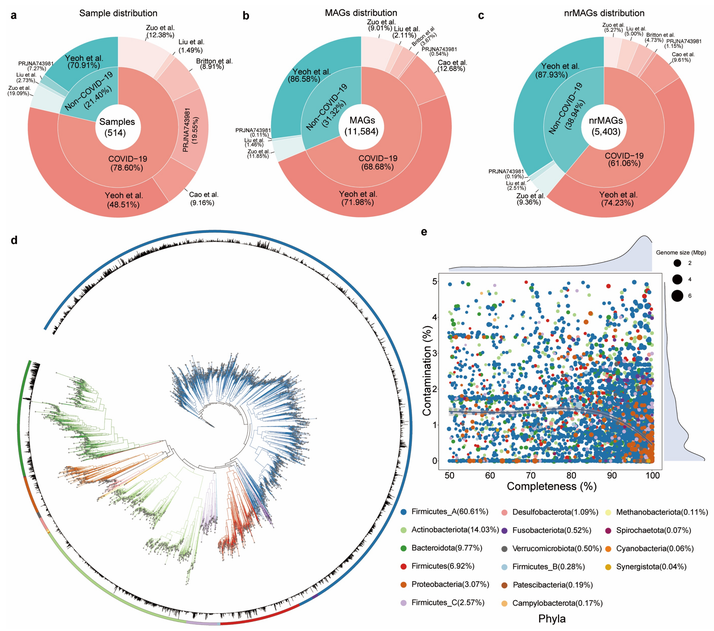


Figure 2. Reconstruction of MAGs from the discovery cohorts.

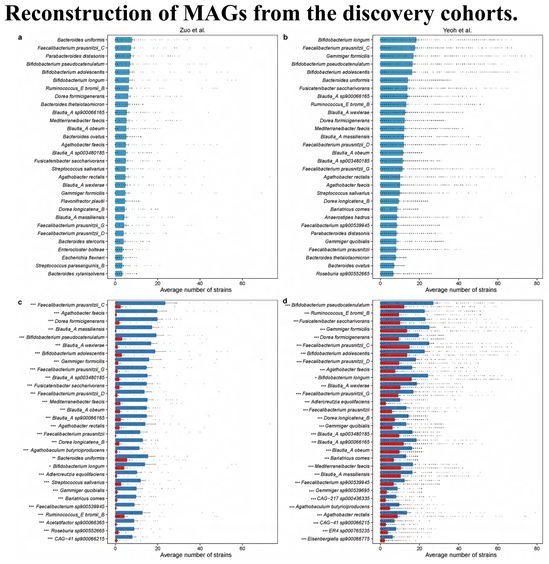


Figure 3. COVID-19 related changes in strain richness of microbial species.

- COVID-19 patients lost many strains of multiple microbial species (**Fig.3**).

- nrMAGs accurately predict the progression of COVID-19 (**Fig.4**). And we observed some opportunistic pathogens were associated with the progression of COVID-19, including nrMAGs from *Klebsiella quasivariicola*, *Klebsiella pneumoniae*, and *Escherichia coli.*

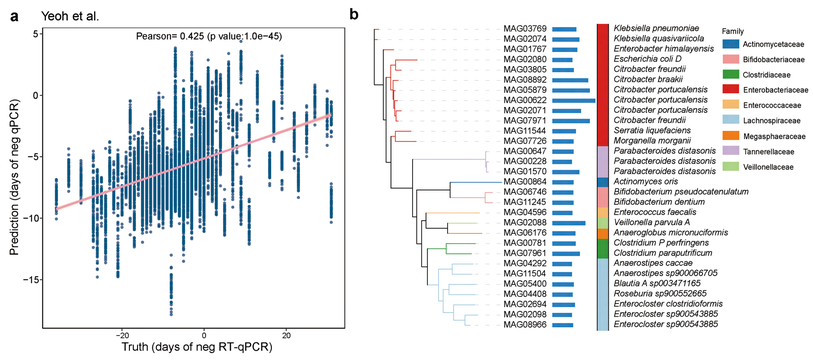


Figure 4. Machine learning model predicts the progression of COVID-19.

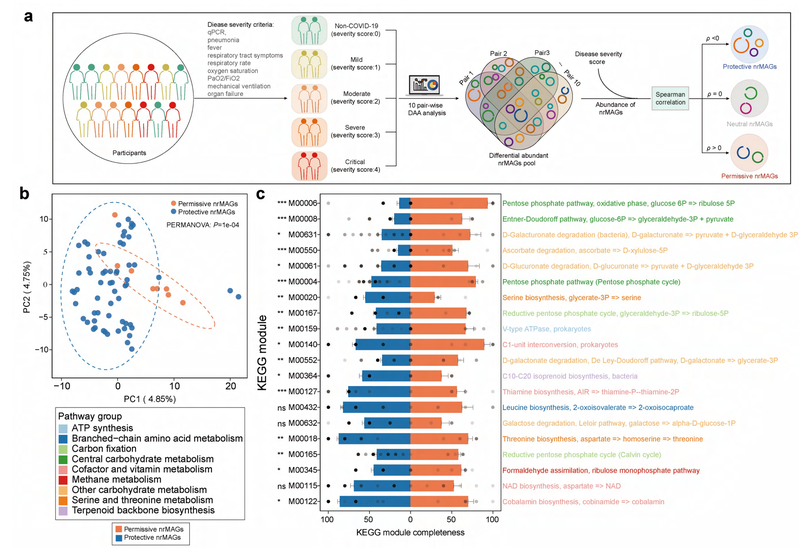


Figure 5. Genome annotation of permissive and protective nrMAGs.

- We identified a set of nrMAGs with a putative causal role in the clinical manifestations of COVID-19 using GMPT pipeline (**Fig.5a**) and revealed their functional pathway (i.e., pentose phosphate pathway) that potentially interact with SARS-CoV-2 infection (**Fig.5b-c**).
- Reactions of the pentose phosphate pathway plays an important role in the production of aromatic amino acids (providing the RNA backbone precursors ribose 5-phosphate and erythrose 4-phosphate).
- The aromatic amino acids in the juxtamembrane domain of the SARS-CoV S glycoprotein play critical roles in receptor-dependent virus-cell and cell-cell fusion.
- A previous study reported that the UK mutation (N501Y) producing aromatic-aromatic interactions that provide for stronger binding between receptor and spike.

Together, these results suggest that specific microbes (permissive nrMAGs may play a role in mediating SRAS-CoV-2 entry into host cells through pentose phosphate pathway and aromatic amino acids.

Conclusions

The presented results highlight the importance of incorporating the human gut microbiome in our understanding of COVID-19.

Reference

Ke, S., Weiss, S.T. & Liu, YY. Dissecting the role of the human microbiome in COVID-19 via metagenome-assembled genomes. *Nature Communications* 13, 5235 (2022). https://doi.org/10.1038/s41467-022-32991-w

Acknowledgements

We would like to thank Dr. Yun Kit Yeoh and Dr. Siew C Ng for sharing the phenotypic data with us. We thank Xu-Wen Wang, Zheng Sun, Tong Wang, Darius Schaub, Yunyan Zhou, and Xiaochang Huang for valuable discussions. Yang-Yu Liu acknowledges the funding support from the National Institutes of Health (R01AI141529, R01HD093761, RF1AG067744, UH3OD023268, U19AI095219, and U01HL089856).

Functional diversification of plant small molecules by the gut microbiome tunes intestinal homeostasis

Gavin A. Kuziel[1], Gabriel L. Lozano[1], Corina Simian[2], Emmanuel Stephen-Victor[1], Talal A. Chatila[1], Jing-Ke Weng[2], Seth Rakoff-Nahoum[1]

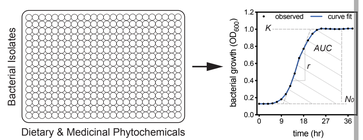
Department of Pediatrics, Harvard Medical School, Boston MA [1]
Department of Biology, Massachusetts Institute of Technology, Cambridge MA [2]

Introduction

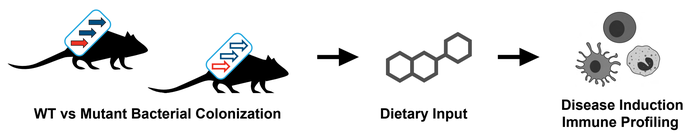
Diet is instrumental in driving the composition and dynamics of the gut microbiome and in the development and prevention of human disease. Unlike our understanding of carbohydrate-microbe interactions, there is a dearth of knowledge as to plant small molecule (phytochemical)-microbe interactions, whether these molecules are metabolized by gut bacteria and how products of phytochemical catabolism affect microbiome composition or host physiology. Here we show that diverse gut symbionts leverage distinct genetic systems to bioactivate dietary phytochemicals to immunomodulatory metabolites. Our findings provide new insight into the role of the microbiome in the activation of abundant dietary phytochemicals and the effects of these metabolic transformations on the maintenance of intestinal homeostasis and protection from enteric disease.

Approach

- Determine the scope of phytochemical metabolism across prominent human gut bacteria utilizing techniques within microbiology and culturomics
- Identify and characterize the genetic and enzymatic basis for phytochemical metabolism, leveraging techniques within microbial genetics and molecular biochemistry

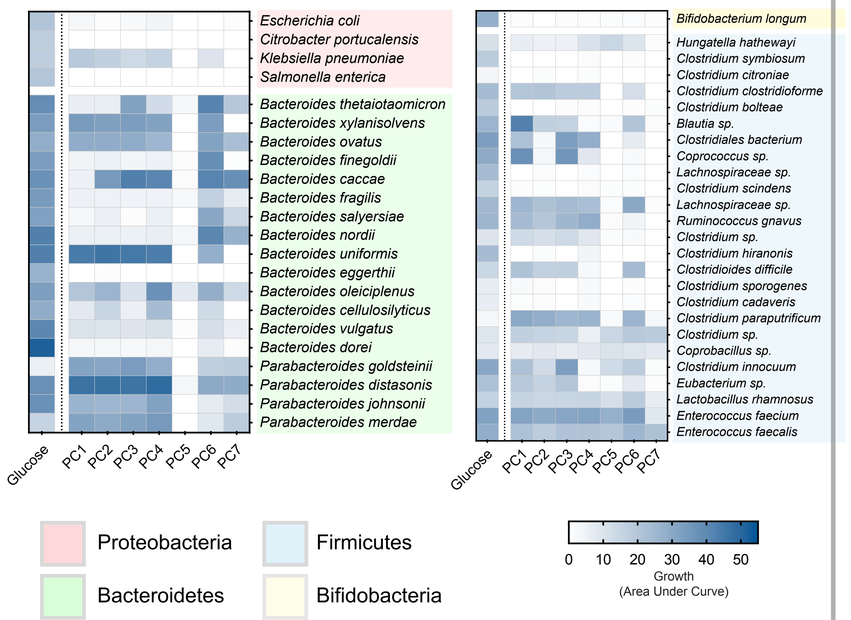


- Assess the pro- or anti-homeostatic effects of phytochemical metabolism on host physiology using coupled in vitro and in vivo models of intestinal disease such as colitis or colorectal cancer

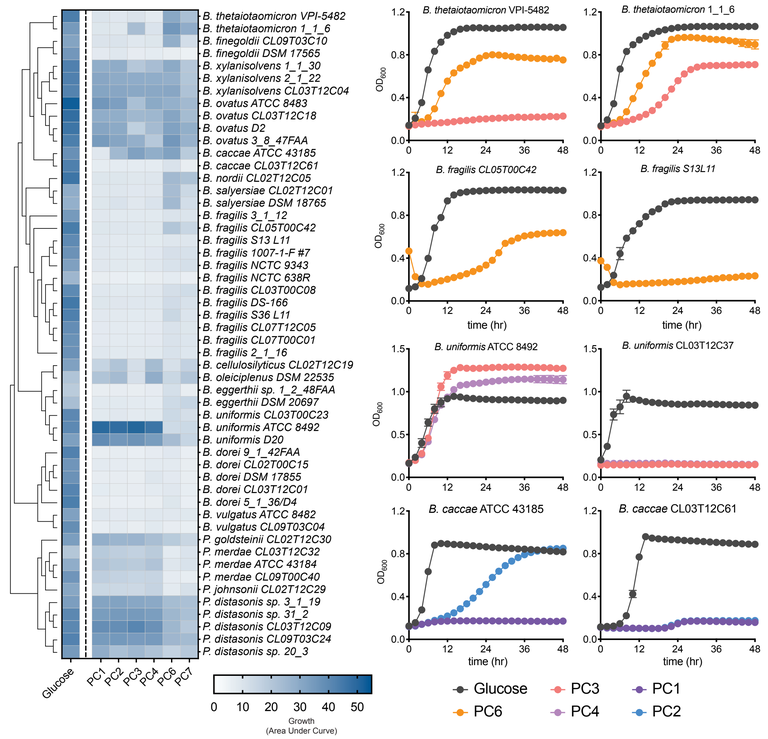


Results

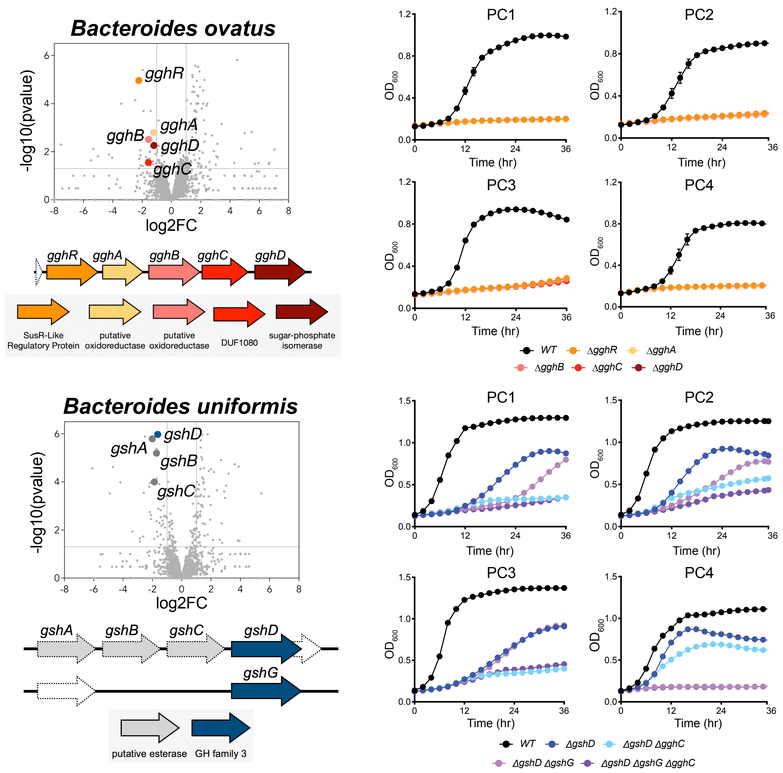
1 Diverse human gut bacteria metabolize dietary phytochemicals



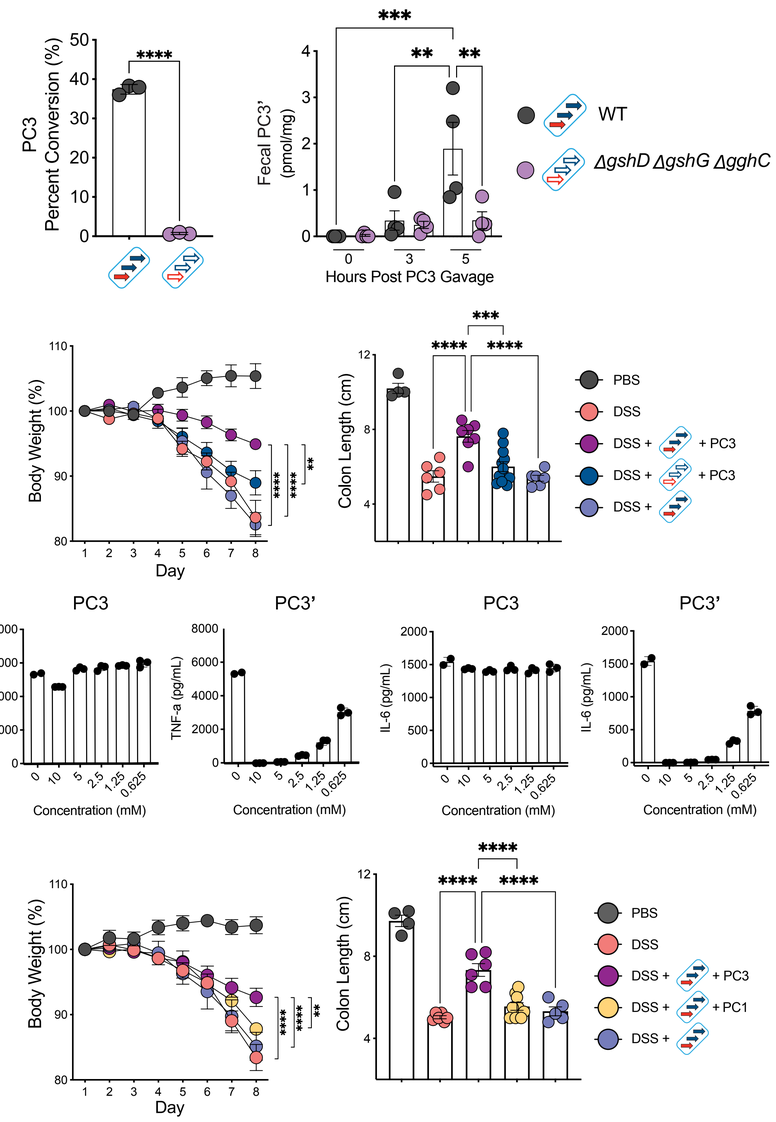
2 Species-level and chemical-level variation is extensive within *Bacteroides* phytochemical metabolism



3 *Bacteroides* species leverage divergent genetic mechanisms for phytochemical metabolism



4 Phytochemical bioactivation by a *Bacteroides* metabolic specialist differentially tunes intestinal homeostasis



Future Directions

- Identify the cellular and molecular circuitry underlying **PC3'**-mediated protection from DSS-induced experimental colitis
- Determine whether microbial bioactivation of phytochemicals protects against other diseases such as infection, cancer, or food allergy

gkuziel@g.harvard.edu
linkedin.com/in/gavinkuziel

BROAD INSTITUTE
NATIONAL INSTITUTES OF HEALTH
PEDIATRIC INFECTIOUS DISEASES SOCIETY
Food Allergy Science Initiative
BURROUGHS WELLCOME FUND
THE PEW CHARITABLE TRUSTS

Genetic diversity of commensal *Blastocystis* gut protists reveals strain-specific changes in host-interfacing pathways

Abigail Lind[1], Ami Bhatt[2], and Katie Pollard[1]

[1]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, [2]Department of Genetics, Stanford University, Stanford, CA

Eukaryotes in the human gut microbiome

Microbial eukaryotes (protists, fungi) and microscopic animals are found alongside bacteria and archaea in natural microbial systems, including host-associated microbiomes.

Eukaryotes in human gut microbiomes are incredibly diverse and span a wide range of the eukaryotic tree of life (Figure 1).

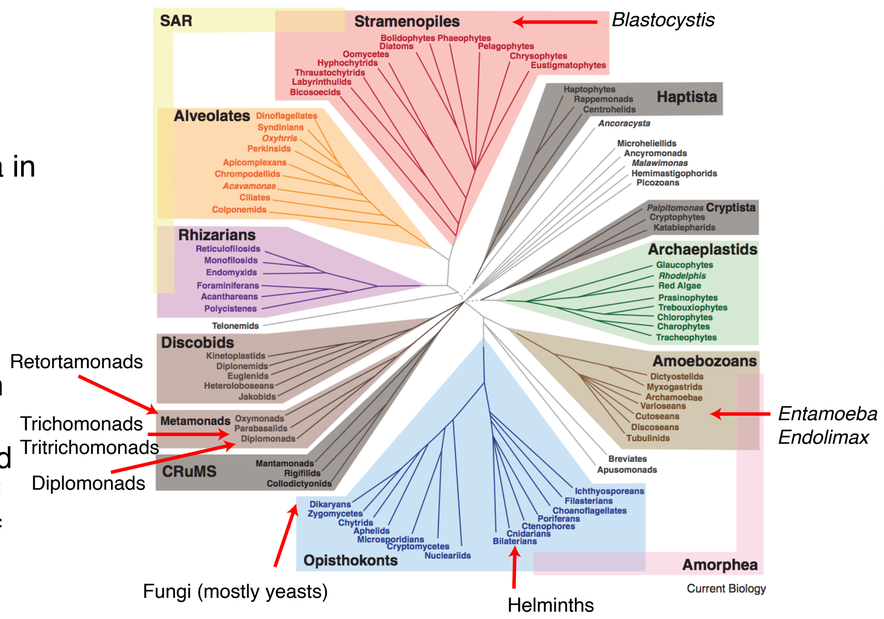


Figure 1. The eukaryotic tree of life[1]. Species belonging to taxonomic groups marked with red arrows are found in human g microbiomes.

Blastocystis is the most common gut eukaryote, correlates with health & differences in microbiota

The stramenopile protist *Blastocystis* is the most prevalent commensal eukaryotic gut colonizer (Figure 2). *Blastocystis* is more common in individuals without gut inflammation and correlates with lowered markers of gut inflammation and metabolic syndrome[3]. Certain microbiota co-occur and co-exclude with *Blastocystis* (Figure 3).

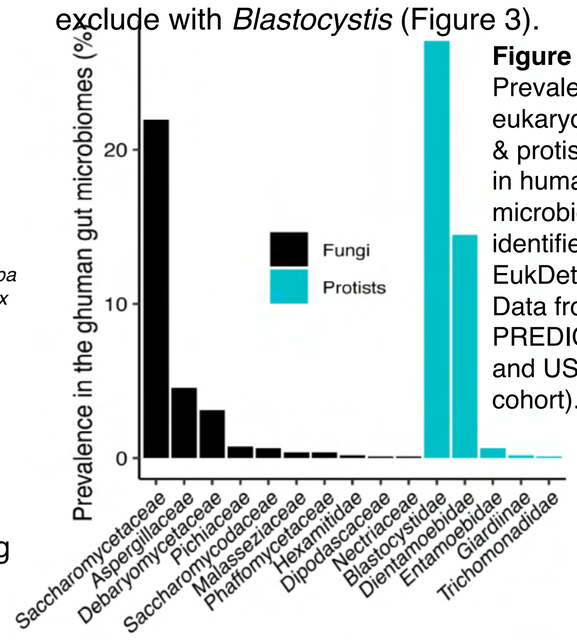


Figure 2. Prevalence of eukaryotic fungal & protist species in human gut microbiomes, as identified with EukDetect[2]. Data from PREDICT[3] (UK and US healthy cohort).

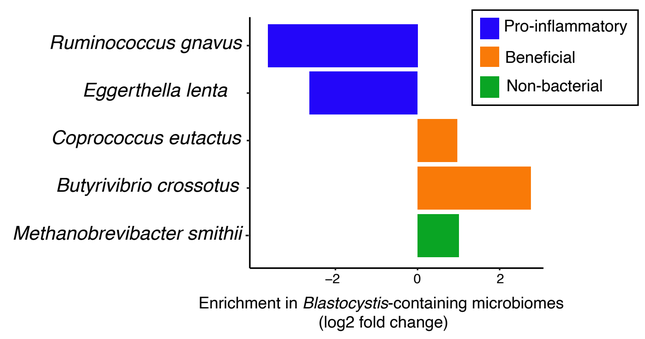


Figure 3. Bacteria that co-occur or co-exclude with *Blastocystis* in the human microbiome. Numbers indicate relative abundance. Data from PREDICT[3].

Blastocystis is genetically diverse with a derived morphology

The *Blastocystis* genus is comprised of over 20 known subtypes that colonize the gastrointestinal tract of animals.

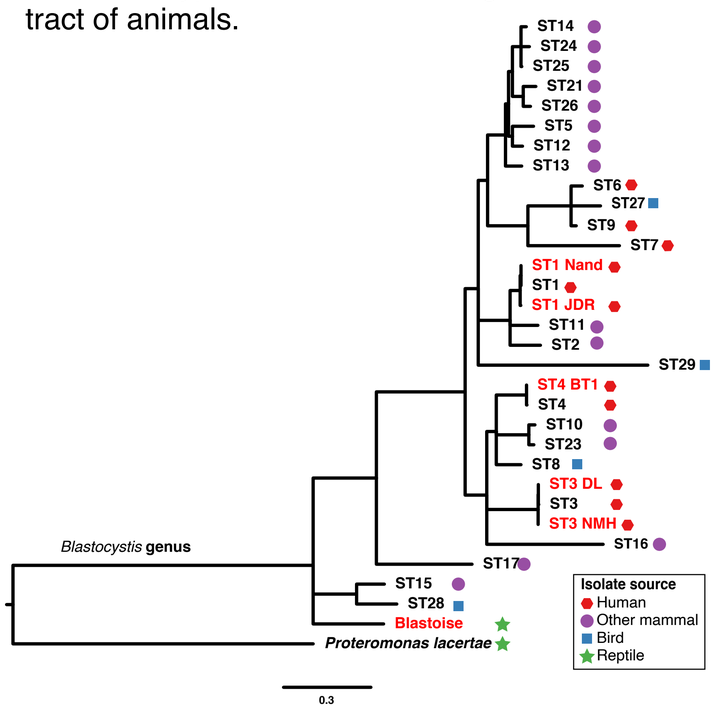


Figure 4. 18S rRNA phylogeny of all described *Blastocystis* subtypes. *Proteromonas lacertae* is used as an outgroup. Branches with bootstrap values less than 80 are collapsed. Sequences in red were sequenced in this study.

Blastocystis is a stramenopile, but does not have characteristic stramenopile morphology. It lacks flagella, and in culture appears in multiple different cell forms (large central vacuole, amoeboid-like protrusions) (Figure 5).

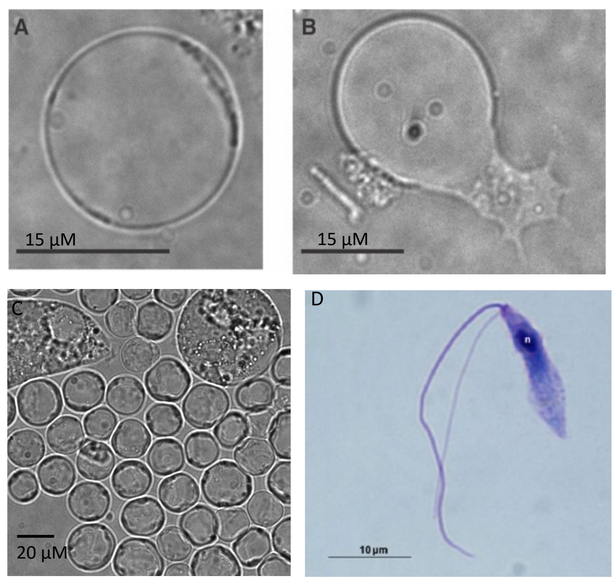


Figure 5. *Blastocystis* morphology. (A) Vacuolar ST1 (JDR) (B) Amoeboid ST3 (DL) in co-culture with bacteria, (C) Vacuolar and granular Blastoise, (D) *Proteromonas lacertae* (credit V. Perez-Brocal)

High-quality genomes spanning the diversity of the *Blastocystis* genus

We cultured 6 *Blastocystis* strains and used long-read Nanopore sequencing, Illumina DNA and RNA sequencing, and for one strain (BT1) Phase Genomics Hi-C scaffolding to generate highly continguous, annotated genomes (Figure X).

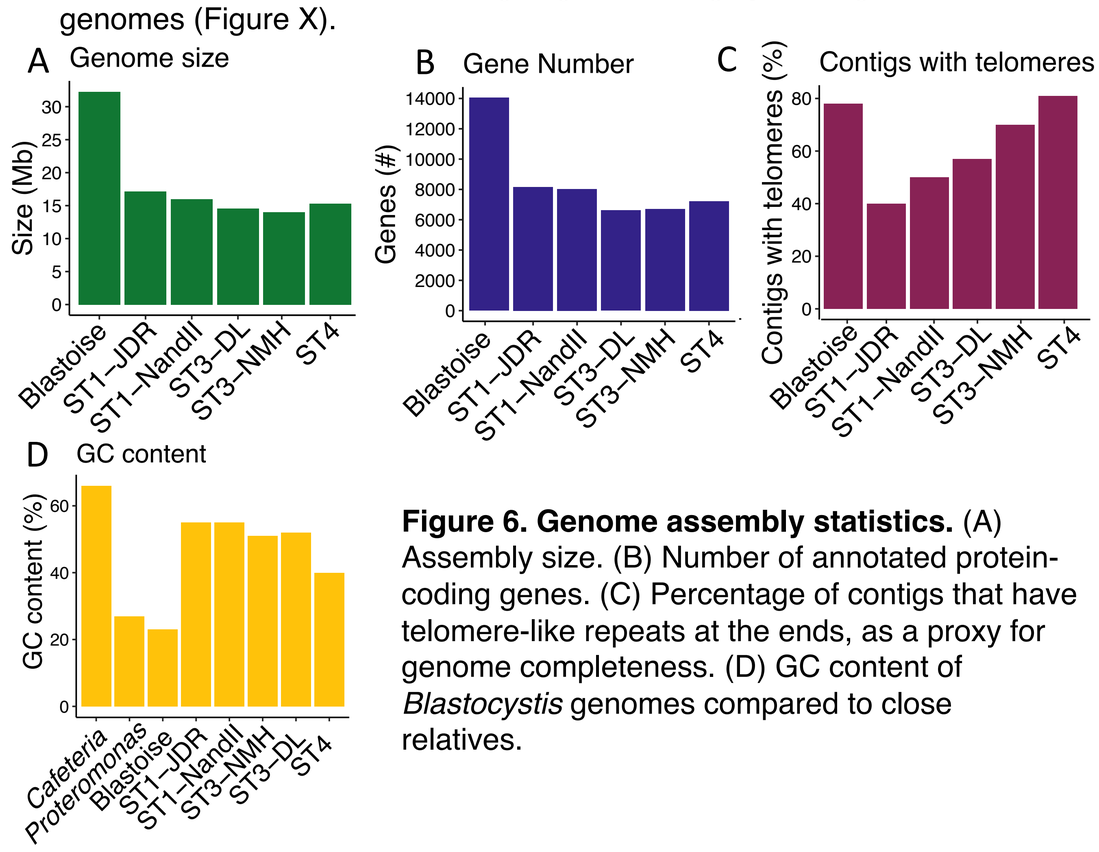


Figure 6. Genome assembly statistics. (A) Assembly size. (B) Number of annotated protein-coding genes. (C) Percentage of contigs that have telomere-like repeats at the ends, as a proxy for genome completeness. (D) GC content of *Blastocystis* genomes compared to close relatives.

Blastocystis has lost morphological genes related to cell body shape and flagella

All *Blastocystis* subtypes have lost most flagellal genes, reduced their molecular motors, and lost ion channels that protists use to move flagellal hairs. These gene losses underly the change in cell morphology seen in *Blastocystis* relative to other stramenopile protists.

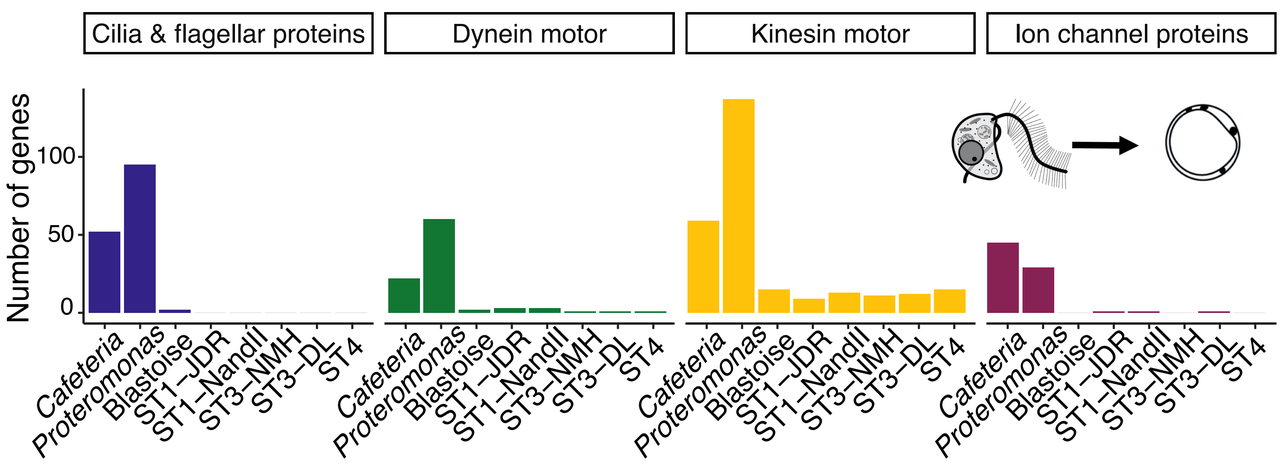


Figure 7. Morphological genes lost in *Blastocystis*.

Expanded diversity of glycosyltransferases underlying antigenic diversity in *Blastocystis*

Human subtypes of *Blastocystis* have undergone large gene family expansions of glycosyltransferase gene families. *Blastocystis* isolates are antigenically diverse, and this diversity in glycosyltransferase may generate diversity in cell surface sugars.

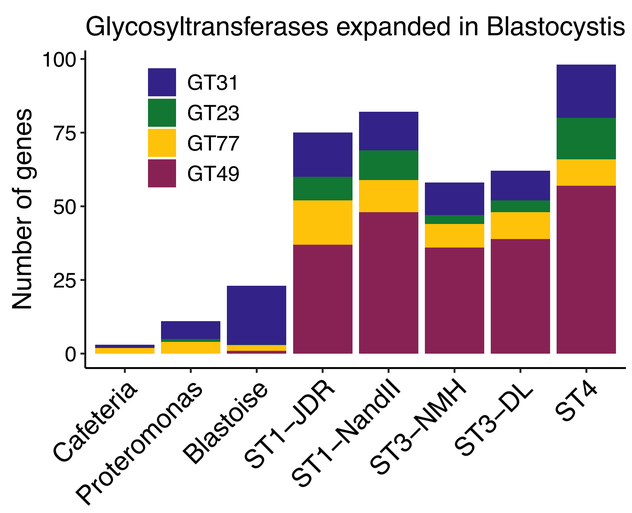


Figure 8. Glycosyltransferase families expanded in *Blastocysis*.

Subtype-specific expansion of horizontally transferred adhesins

Blastocystis and *Proteromonas* species contain a family of adhesins with homologs in bacteria but without homologs in other eukaryotes. This gene family has expanded in some subtypes.

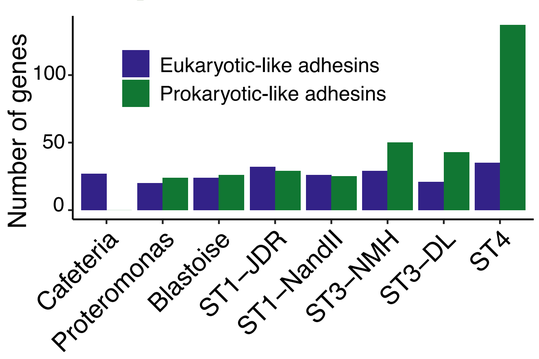
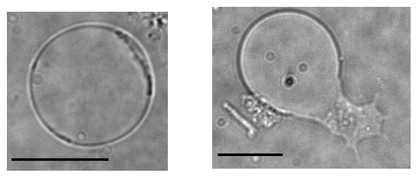


Figure 9. Expanded adhesins in *Blastocystis*.

Blastocystis forms pseudopodia-like extrusions in co-culture with bacteria, and these adhesins may aid in forming these extrusions.



Vacuolar form; axenic culture

Pseudopodia-like extrusion; bacterial co-culture

Figure 10. Pseudopodia-like extrusions from *Blastocystis* in bacterial co-culture.

Summary

- *Blastocystis* is most common gut eukaryote
- High quality genomes of 6 *Blastocystis* strains
- *Blastocystis* has lost morphological genes
- Subtype-specific expansions of host-facing genes including:
  - Horizontally transferred membrane-anchored cadherins
  - Glycosyltransferases underlying antigenic diversity

References

1. Keeling and Burki. *Current Biology*, 2019.
2. Lind and Pollard. *Microbiome*, 2021.
3. Asnicar et al. *Nature Medicine*, 2021.
4. Gentekaki et al. *PLoS Biology*, 2017.

Contact: abigail.lind@gladstone.ucsf.edu allind @AbigailLLind

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

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

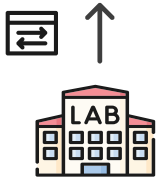
[1]Department of Bistatistitcs, Harvard TH. Chan School of Public Health,
[2]Department of Immunology and Infecteous Diseases, Harvard T.H. Chan School of Public Health

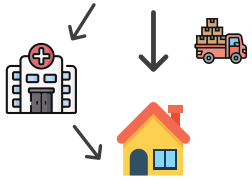
The Microbiome Collection Core at the Harvard T.H. Chan School of Public Health (HCMCC) was established in response to a strong demand among the research community for validated microbiome sample collection kit configurations and easy usability for in-home sampling. Under the umbrella of the Harvard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

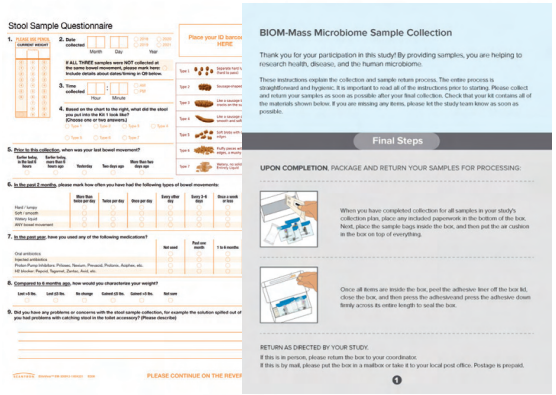
Pre-paid return shipment

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**

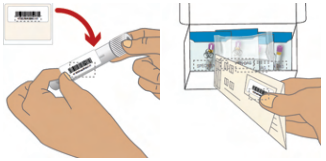
A scalable gut and oral microbiome sample collection platform

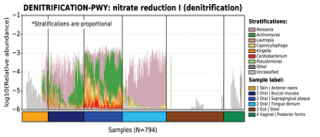
This customizable microbiome sample collection kit avoids the need for expensive, bulky, and inconvenient ice packs by providing several different room temperature storage media that are also compatible with multiple different molecular assays including **any combination of amplicon (16S), metagenomic, 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 acilitate 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[(i)]
Microbiome Collection Core Manager: Steven Medina
Microbiome Analysis Core Director: Xochitl Morgan
HCMPH Co-Directors: Wendy Garret, Curtis Hunntenhower

Follow us on Twitter @hutlab[(ii)]

Follow us on Mastodon @hutlab.mstdn.science[(iii)]

Incorporating metabolic activity, taxonomy and community structure to improve microbiome-based predictive models for host phenotype prediction

Mahsa Monshizadeh and Yuzhen Ye

Computer Science Department, Luddy School of Informatics, Computing and Engineering, Indiana University Bloomington

Abstract

The human gut microbiome play key roles in human health and diseases. We developed MicroKPNN, a prior-knowledge guided interpretable neural network for microbiome-based human host phenotype prediction. The prior-knowledge used in MicroKPNN includes the metabolic activities of different bacterial species, phylogenetic relationships, and bacterial community structure. Application of MicroKPNN to seven gut microbiome datasets (involving five different human diseases including inflammatory bowel disease, type 2 diabetes, liver cirrhosis, colorectal cancer, and obesity) shows that incorporation of the prior knowledge helped improve the microbiome-based host phenotype prediction. MicroKPNN outperformed fully-connected neural network based approaches in all seven cases, with the most improvement of accuracy in the prediction of type 2 diabetes. MicroKPNN outperformed a recently developed deep-learning based approach DeepMicro, which selects the best combination of autoencoder and machine learning approach to make predictions, in six out of the seven cases. More importantly, we showed that MicroKPNN provides a way for interpretation of the predictive models. Our results suggested that the metabolic potential of the bacterial species contributed more than the two other sources of prior knowledge. MicroKPNN is publicly available at https://github.com/mgtools/MicroKPNN.

Methods

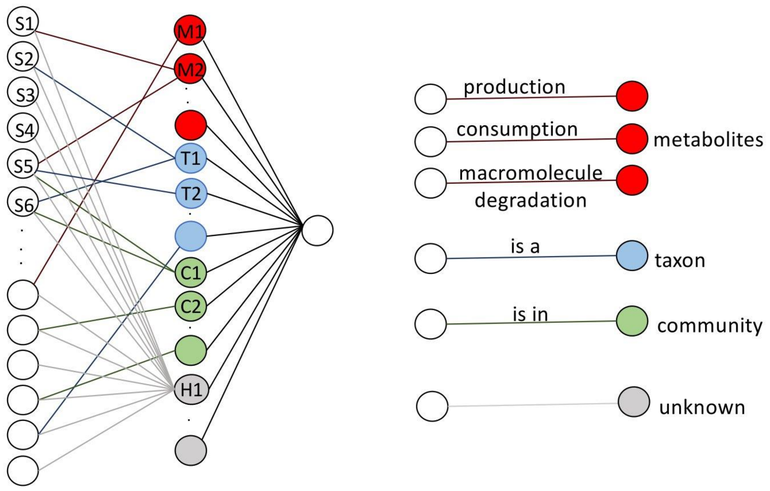


Figure 1. The neural network structure used in MicroKPNN. It is composed of three layers (shown on the left). In the input layer, each node is a species, and the hidden layer includes nodes of four different groups: metabolites (red), taxa (blue), communities (green), and unknown hidden nodes (gray). The links between the input nodes and the nodes in the hidden layer represent different biological meanings (shown on the right).

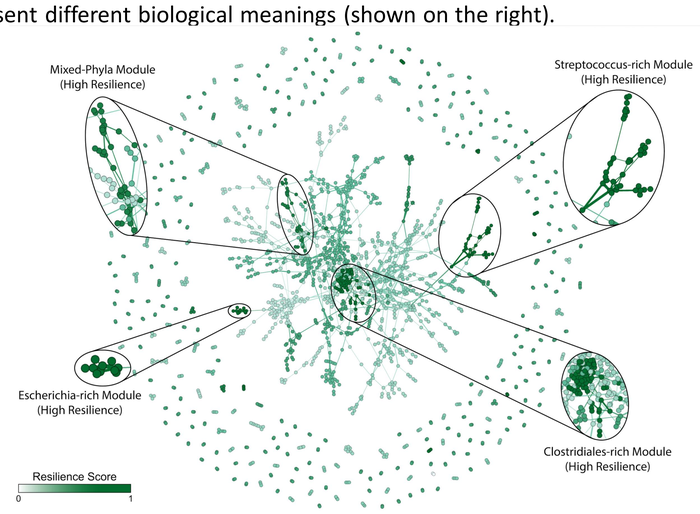


Figure 2. MicroKPNN uses bacterial communities that were inferred from a microbiome association network (see Lam and Ye, 2022). In this network, nodes (species) are colored by module resilience, a metric we proposed to quantify the tendence of different bacterial species forming bacterial communities.

Optimization of MicroKPNN and comparison with other methods

Table 1. Summary of best performing neural network architecture for each dataset and their average AUC.

dataset	no. of nodes in different groups in the hidden layer: all	taxon (rank)	metabolite	community	unknown	avg. AUC
IBD	519	176 (genus)	234	29	80	0.953
EW-T2D	309	34 (order)	234	31	10	0.820
C-T2D	365	27 (class)	240	38	60	0.753
Cirrhosis	354	40 (order)	239	35	40	0.969
Colorectal	403	65 (family)	234	34	70	0.846
Obesity	479	184 (genus)	233	32	30	0.699
Obesity-multi	822	190 (order)	276	336	20	0.875

Table 2. Comparison of MicroKPNN with different methods including NNs that are fully connected (fc-NN) in averaged AUC and standard deviation (in parenthesis).

Dataset	MicroKPNN	fc-NN (keras)	fc-NN (KPNN)	DeepMicro[a]	EPCNN[c]
IBD	**0.953** (0.019)	0.865 (0.031)	0.678 (0.089)	0.873 (0.067)	NA
EW-T2D	0.820 (0.047)	0.595 (0.065)	0.580 (0.062)	**0.829** (0.087)	0.789 (0.056)
C-T2D	0.753 (0.013)	0.675 (0.033)	0.723 (0.019)	0.725 (0.056)	**0.813** (0.024)
Cirrhosis	**0.969** (0.009)	0.823 (0.022)	0.947 (0.021)	0.888 (0.025)	0.953 (0.007)
Colorectal	0.846 (0.020)	0.624 (0.038)	0.764 (0.053)	0.809 (0.103)	**0.906** (0.013)
Obesity	**0.699** (0.045)	0.539 (0.023)	0.608 (0.022)	0.674 (0.076)	NA
Obesity-multi	**0.875** (0.017)	0.820 (0.014)	0.826 (0.045)	0.763 (0.042)[b]	NA

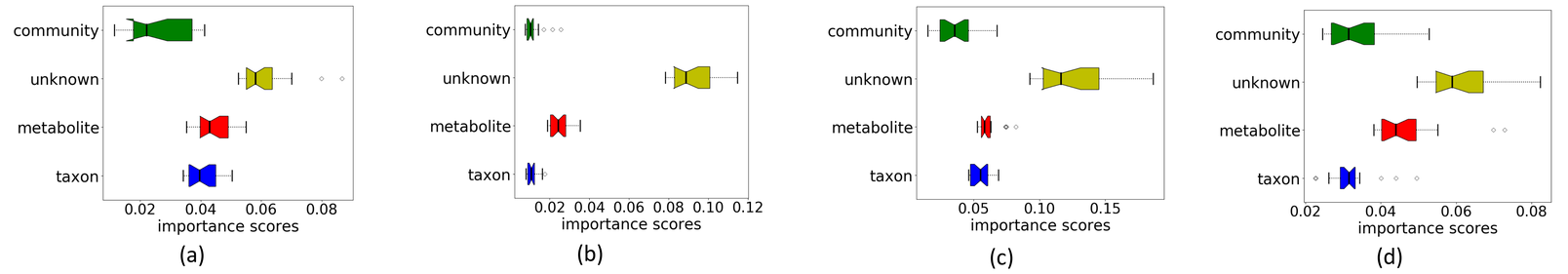


Figure 3. Contributions of the different groups of hidden nodes to the prediction as measured by importance scores. (a) IBD (b) C_T2D (c) Obesity (d) Cirrhosis.

A case study: What can MicroKPNN tell about liver cirrhosis prediction?

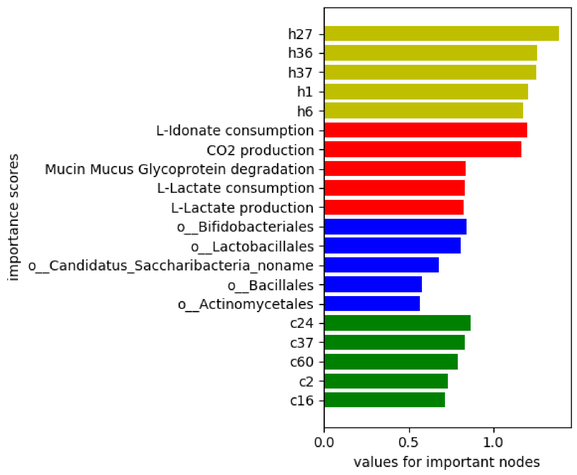


Figure 4. Importance scores of the hidden nodes for microbiome-based liver cirrhosis prediction. The top five most important nodes of each group in the hidden layer for prediction of cirrhosis are shown in the plot. The bars are highlighted in different colors: yellow for unknown nodes, red for metabolites, blue for taxa, and green for communities.

- L-Idonate, CO2, and mucin glycoprotein were the top three most important metabolite nodes that contributed to the prediction.
- Among the bacterial species that are involved in mucin glycoprotein degradation (i.e., mucin consumers), we observed that Ruminococcus gnavus was highly elevated in cirrhosis patients. Increase of R. gnavus was found to be implicated in the degradation of elements from the mucus layer providing an explanation for the impaired intestinal barrier function and systematic inflammation in LC patients.
- Lactate consumption and production were also picked up as important nodes by MicroKPNN, suggesting the importance of the bacterial species that produce and/or digest these metabolites.
- It is well known that bacteria produce intermediate fermentation products including lactate, but these are normally detected at low levels in feces from healthy individuals due to extensive utilization of them by other bacteria.
- Among the taxon nodes, Bifidobacteria had the highest importance score; previous studies have shown that patients with chronic liver disease have varying degrees of intestinal microflora imbalance with a decrease of total Bifidobacterial counts.

Discussion and Future Work

MicroKPNN uses a simple architecture, but by leveraging on prior knowledge of microbial species, it provides promising predictions of host phenotype based on microbiome composition as shown on all seven datasets. Comparison of the importance scores of different prior knowledge showed that the metabolic activities had the largest impact on the performance of predictions. The difference between the relative importance scores of the hidden nodes with that of the unknown nodes indicates the knowledge gap between the microbial species and their interaction with human hosts. The predictive models we built in this work are based on species abundance. It has been shown (including our own work) that using bacterial genes typically (not always) results in better predictive models. A future direction of our work is to expand MicroKPNN so that it can take gene abundance as the input for microbiome-based prediction.

References

- Monshizadeh, M and Ye, Y. Incorporating metabolic activity, taxonomy and community structure to improve microbiome-based predictive models for host phenotype prediction. doi: https://doi.org/10.1101/2023.01.20.524948.
- Lam, T. J. & Ye, Y. Meta-analysis of microbiome association networks reveal patterns of dysbiosis in diseased microbiomes. Sci Rep 12, 17482 (2022).
- Fortelny, N. & Bock, C. Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. Genome biology 21, 1–36 (2020)
- Oh, M. & Zhang, L. DeepMicro: deep representation learning for disease prediction based on microbiome data. Scientific reports 10, 1–9 (2020).
- Chen, X. et al. Human disease prediction from microbiome data by multiple feature fusion and deep learning. Iscience 25, 104081 (2022).

Acknowledgement

This work was supported by NSF 2025451 and NIH 1R01AI143254.

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

Xochitl C. Morgan[1], Lauren J. McIver[1], Thomas Kuntz[1], Curtis Huttenhower[1,2]
[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health [2]Broad Institute of MIT and Harvard

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

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

Fully-collaborative support for all stages of funded investigations

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

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

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

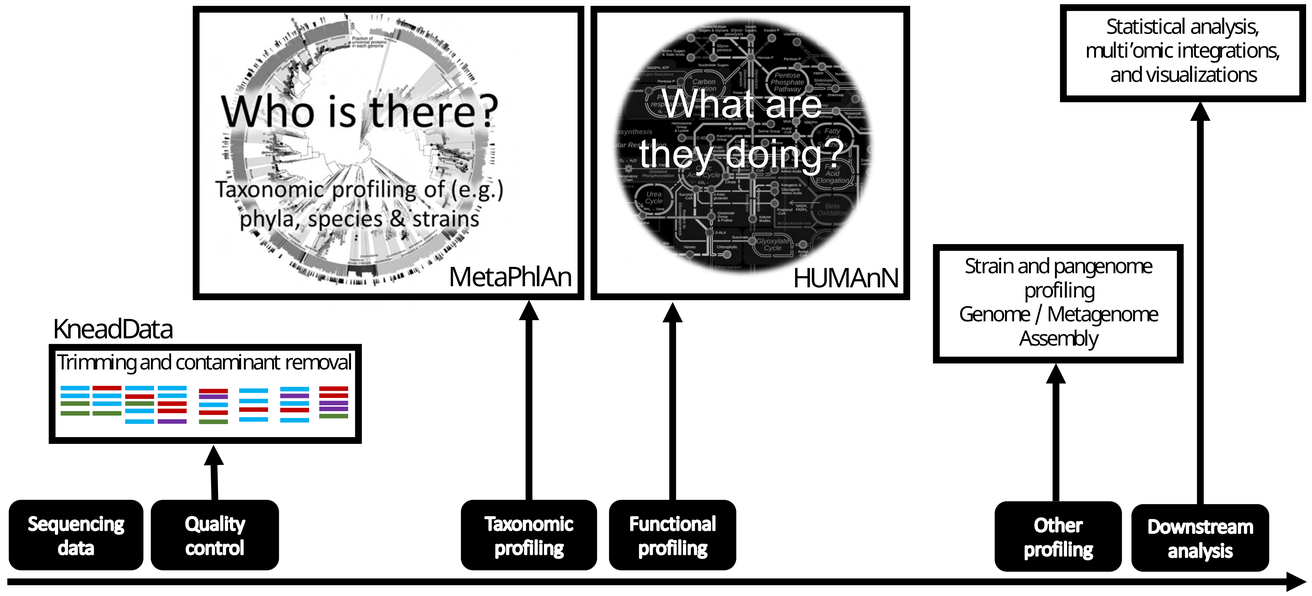
Interpretation
- Results
- Discussion
- Manuscript writing/editing
- Response to reviewers

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

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

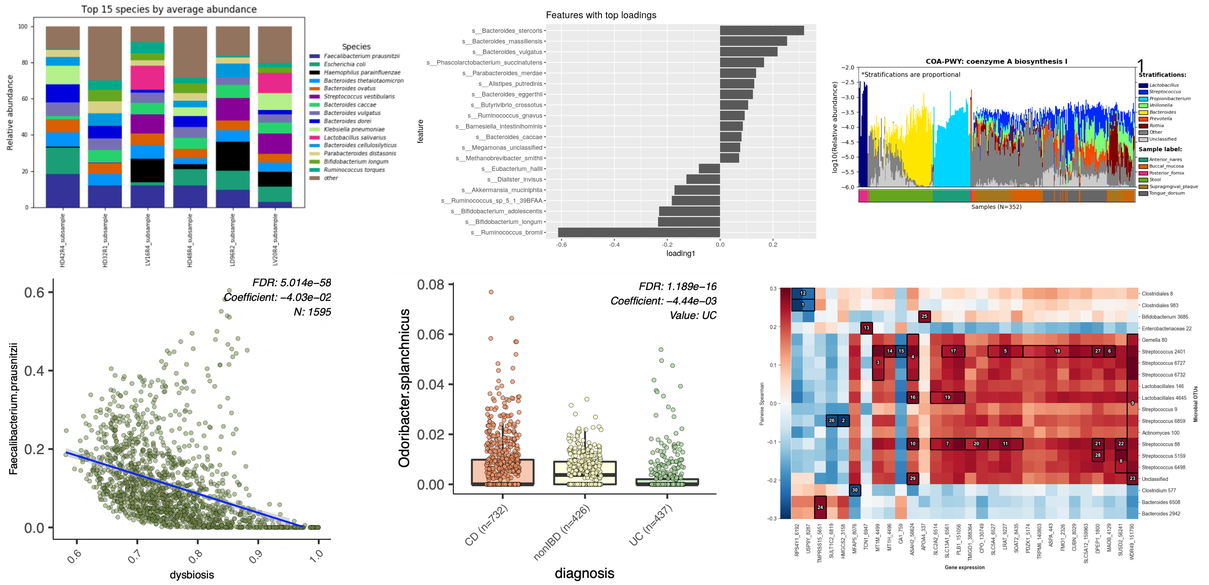
Contact us at: hcmac@hsph.harvard.edu

Microbial community profiling



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

Downstream analysis and statistics



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

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

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

Gut microbiome-metabolome interactions during ketogenic diets of varied composition

Jacob T. Nearing[1,2,3], Kelsey N Thompson[1,2,3], Amrisha Bhosle[1,2,3], Veronica Perdomo[2], William A. Nickols[2], Tobyn Branck[2,4], Dayakar Badri[4], Curtis Huttenhower[1,2,3], Matthew Jackson[4]

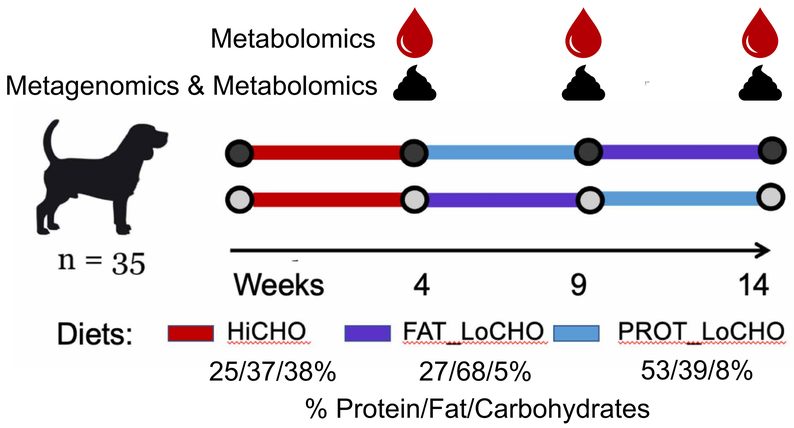
[1]Broad Institute of Harvard and MIT - [2]Harvard T.H. Chan School of Public Health - [3]Harvard Chan Microbiome in Public Health Center -[4]Hill's Pet Nutrition

Low carb diets and the microbiome

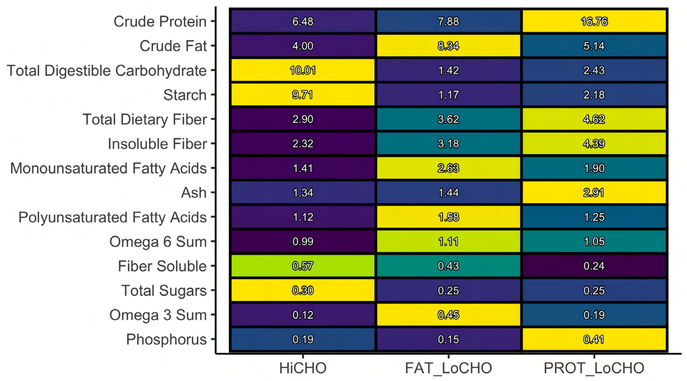
Diets that come close to eliminating all dietary carbohydrates (low carb) force the body to switch from utilizing carbohydrates to fat or protein as its main energy source (Swink et al., 1997). When energy is primarily driven by the usage of fat this state known as ketosis and is characterized by the production of ketone bodies such as 3-hydroxybutyrate (BHB). This state has been associated with a number of health benefits. For example, recent evidence has suggested that alterations in the gut microbiome by ketogenic diets may play a role in improving colitis (Kong et al., 2021) and epileptic symptoms (Dahlin & Prast- Nielsen, 2019). However, the degree to which the microbiome is impacted by replacing dietary carbohydrates with protein or fat remains unclear.

Experimental design

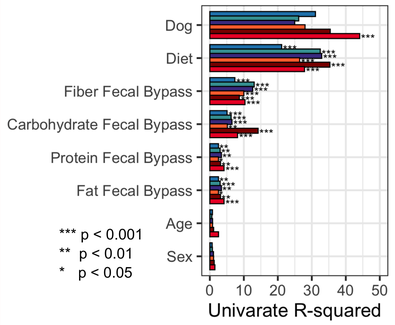
- Beagles were randomly assigned two different food orders
- All dogs were initally fed a standard "high" carb food
- All samples were collected in a fasted state



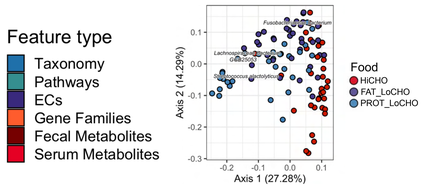
Median dietary intake by metabolic body weight (grams/(kg body weight$^{0.75}$))



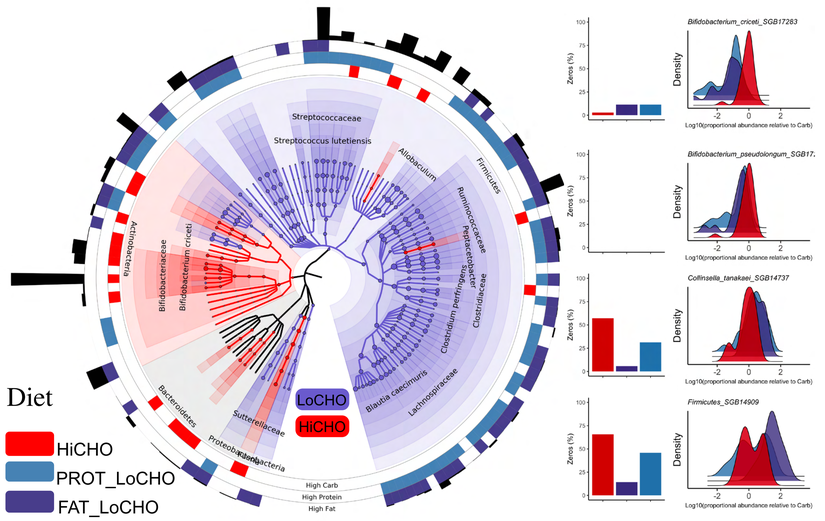
Food change is associated with overall microbiome and metabolome profiles



We evaluated the assocation between dietary change and overall microbiome (Bray-Curtis) and blood and stool metabolome profiles (Manhattan distance) using PERMANOVA tests.

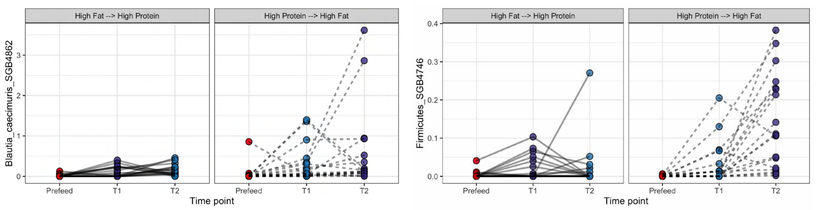


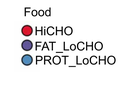
Low carb foods are associated with both broad and specific taxonomic changes



- MaAsLin2 linear mixed models: ~ age + gender + (1|animal) + diet
- Inner circle: assocaited with carbohydate levels
- Outer circle: associated with specific food group

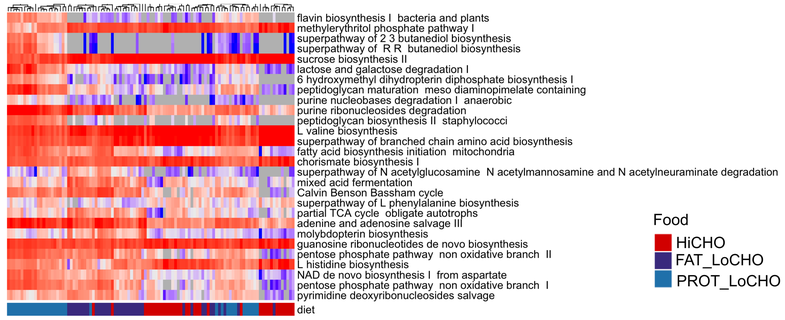
Order of food impacts microbial abundance





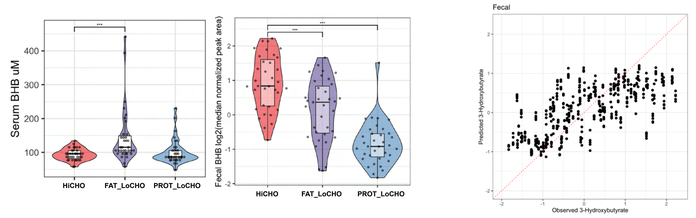
MaAsLin2 linear mixed models: ~ age + gender + (1|animal) + diet order

A broad range of microbiome functional pathways differ with food change



MaAsLin2 linear models on functional relative abundance were run using diet, age, and gender as fixed effects and animal as a random effect.

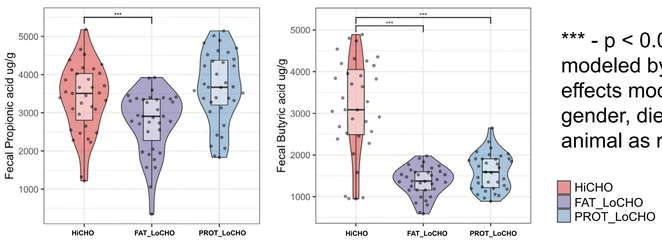
Low carb foods result in differing serum and fecal 3-hydroxybutyrate (BHB) levels



*** - p < 0.001 as modeled by a linear mixed effects model with age, gender, diet as fixed effects and animal as a random effect

Random forest modelling predicts fecal BHB levels using taxonomic composition. Models were validated using ten repeat five fold cross validation

Not all low carb foods are equal in their impact on microbial fermentation



*** - p < 0.001 as modeled by a linear mixed effects model with age, gender, diet as fixed and animal as random effects

HiCHO
FAT_LoCHO
PROT_LoCHO

Acknowledgments

We would like to thank Hill's Pet Nutrition and the National Institutes of Health for funding this work. We would additionally like to thank all of the animals and those that care for them. All software used for this project can be found at the below link through the bioBakery.

Http://huttenhower.sph.harvard.edu

email: nearing@broadinstitute.org

The landscape of novel lateral gene transfer events in the human microbiome

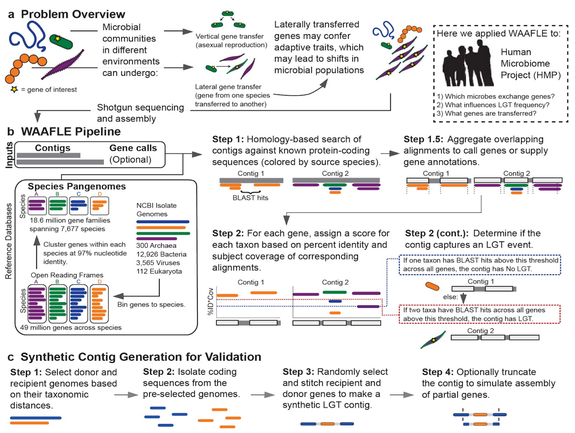
Etienne Nzabarushimana[1,2*], Tiffany Y. Hsu[2*], Dennis Wong[4], Chengwei Luo[3], Robert G. Beiko[4], Morgan Langille[5], Curtis Huttenhower[2,3], Long H. Nguyen[1,2†], Eric A. Franzosa[2,3†]
*co-lead
†co-supervised

[1]Massachusetts General Hospital and Harvard Medical School. Boston, MA, USA, [2]Harvard T.H. Chan School of Public Health. Boston, MA, USA, [3]The Broad Institute of MIT and Harvard, Cambridge, MA, USA,
[4]Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada, [5]Department of Pharmacology, Dalhousie University, Halifax, Nova Scotia, Canada

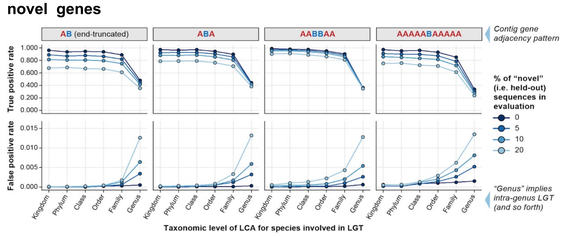
ABSTRACT

Lateral gene transfer (LGT) is an important mechanism for genomic diversification in microbial populations and communities[1], including the human microbiome[2]. While previous work has surveyed ancient LGT events in human-associated microbial isolate genomes[3], the scope, and dynamics of novel LGT events in human microbiomes are not well understood. We addressed this by developing and validating a computational method (**W**orkflow to **A**nnotate **A**ssemblies and **F**ind **L**GT **E**vents or WAAFLE) to profile novel LGT events from assembled metagenomes. We assessed WAAFLE on synthetic contigs containing spiked LGTs and identified intergenus LGTs with >91% sensitivity and >99.9% specificity. For more challenging intragenus LGT (due to congeneric overlap), we report a still-respectable 51% sensitivity. Applying WAAFLE to >2K human metagenomes from diverse body sites, we identified >100K high-confidence putative, novel LGT events. These events were enriched for mobile elements (as expected), as well as restriction-modification and transport functions, both being particularly intriguing areas for further study given their putative role in viral/phage-mediated LGT defense. LGT frequency was quantifiably influenced by biogeography, the phylogenetic similarity of the involved taxa, and the ecological abundance of the involved taxa. Our findings suggest that LGT is an active process in the human microbiome, occurring far more frequently than previously suspected

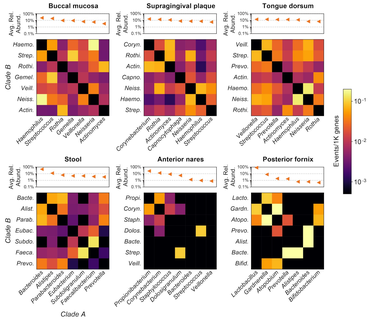
METHOD



WAAFLE finds most expected LGTs with few false positives and is robust to novel genes

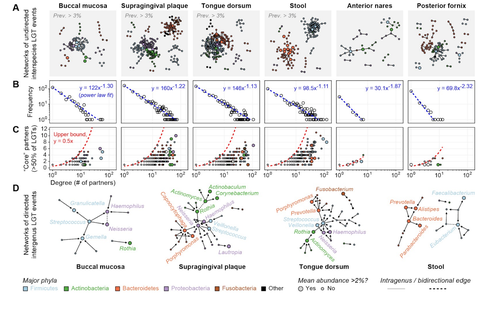


RESULTS



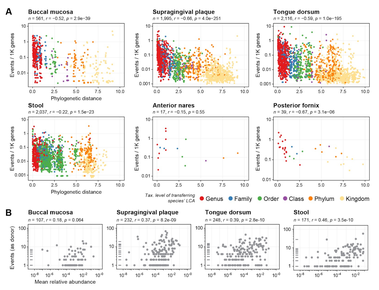
Rates of undirected LGT among major genera of the human microbiome

Six body sites with metagenome sequencing from at least 20 individuals in the HMP1-II are shown; the three body sites in the top row are all from the oral cavity. Heatmap values indicate the density (rate) of undirected LGT between major genera from the body site, with "major genera" defined based on ranked average relative abundance. Rates are computed over first-visit samples from HMP subjects.



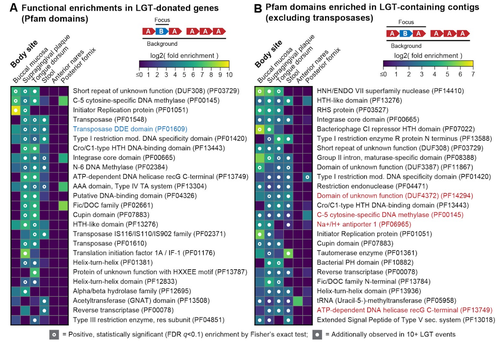
Preferential attachment in the human microbiome LGT network

(**A**) LGTs shown as undirected edges between species (nodes) across six major microbiome sites. Edges from the oral and gut site are filtered for 3% population prevalence, while all edges are shown for the sparser nares and fornix sites (complete predictions in Table S2). Nodes are colored according to major phyla (top 5 by mean abundance) and sized according to species relative abundance. (**B**) Node degrees in the (unfiltered) networks from 'a' follow power-law distributions, with many low-degree species and a long tail of high-degree (hub) species. (**C**) LGT events involving hub species are often dominated by a small number of LGT partners. (**D**) Directed edges are drawn from donor to recipient genera from the oral and gut sites. Edges are filtered for 3% population prevalence, with directionality requiring at least a two-fold preference for the donor role (bidirectional edges are dashed).



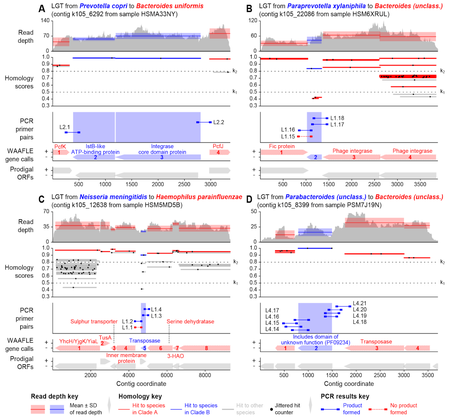
Phylogenetic distance and donor abundance as determinants of LGT rate

(**A**) Negative relationships between density of undirected LGT for species pairs (normalized by their combined total assembly size) and phylogenetic distance at six major HMP body sites. Pairs are colored according to "remoteness" of the LGT (i.e., the taxonomic level of the LCA of the two transferring species). (**B**) Positive relationships between species' frequencies as an LGT donor (inferred from directed LGT events) and species' mean body site abundances at four major HMP body sites. The nares and fornix sites were not sufficiently well represented among directed LGTs and were excluded from this analysis. Horizontal marks in the y-axis margin represent species that occurred as donors, but which were never detected by MetaPhlAn2 (i.e., having zero mean abundance). In both "a" and "b," only species (or species pairs) contributing at least 100 genes across assembled metagenomes were considered. Correlation ("r") values are Spearman's rank correlation; *p*-values are two-tailed.



Novel LGTs are enriched for mobile elements and transport functions

(**A**) Fold enrichments for Pfam[4]domains among transferred genes from inter-genus LGT events relative to all genus-resolved genes in first-visit HMP metagenomes. Dots indicate statistically significant positive enrichments based on Fisher's exact test, with nominal p-values subjected to FDR control (target FDR=0.1). Only domains seen in 10+ LGT events from at least one body site were considered. The top-25 such domains by mean log-scaled fold enrichment are shown. (**B**) Fold enrichments for Pfam domains among inter-genus LGT contigs relative to single-genus contigs. The top-25 domains were selected and plotted as in panel 'A', with seven transposase domains excluded to highlight other functions. (**C**) Taxonomic composition of LGT-enriched Pfam domains at oral and gut sites. The first example (blue title) is based on counts from panel 'A'; all other examples (red titles) are based on counts from panel 'B'.



Experimental support for novel LGT events in human stool

A-D :Each panel describes a single contig from an HMP2[5] non-IBD control metagenome that contains a PCR-supported LGT event. "Read depth" shows variation in the depth of sample reads matching sites along the contig, as computed during WAAFLE's quality control procedures. Bands show the mean and standard deviation of read depth for each WAAFLE-identified ORF. "Homology scores" show matches to this contig from WAAFLE's protein-coding reference sequences, along with the k_1 and k_2 homology thresholds used by the algorithm. Each alignment is represented by a thin gray line (indicating coverage), at a particular height (homology score), with a black dot placed randomly within the line (to facilitate counting alignments in densely aligning regions). "PCR primer pairs" shows the locations of the designed primers (line endpoints) and amplified products; some LGT events had more than one primer pair per endpoint.. "WAAFLE gene calls" and "Prodigal ORFs" show the location and direction of WAAFLE- and Prodigal-suggested ORFs along the contig and are largely in agreement. Features shown in red correspond to the putative recipient clade, while those shown in blue correspond to the putative donor clade. Functional annotations are taken from either 1) the UniProt-assigned name of the best homolog at each locus, or 2) the UniProt-predicted domain composition of that homolog.

CONCLUSIONS

- Novel methodology for culture-independent LGT detection and profiling in complex microbial communities
  - WAAFLE's focus on raw (unbinned) short-read metagenomic contigs improves sensitivity and avoids the daunting technical challenge of assembling complete microbial genomes from metagenomes
- New insights into the landscape of LGT events in the human microbiome:
  - LGT is an active process in the human microbiome, occurring far more frequently than previously suspected
  - LGT frequency is quantifiably influenced by biogeography, the phylogenetic similarity of the involved taxa, and the ecological abundance of the involved taxa
  - LGTs are enriched for mobile elements, as well as restriction-modification and transport functions typically associated with the destruction of foreign DNA (and a theoretical impedance to LGT) and for which their relative overrepresentation may suggest a selective advantage that ironically promoted their lateral dissemination and fixation

REFERENCES

1. Brito, I. L. (2021). Examining horizontal gene transfer in microbial communities. Nature Reviews. Microbiology, 19(7), 442–453

2. Vatanen, T., Jabbar, K. S., Ruohtula, T., Honkanen, J., Avila-Pacheco, J., Siljander, H., Stražar, M., Oikarinen, S., Hyöty, H., Ilonen, J., Mitchell, C. M., Yassour, M., Virtanen, S. M., Clish, C. B., Plichta, D. R., Vlamakis, H., Knip, M., & Xavier, R. J. (2022). Mobile genetic elements from the maternal microbiome shape infant gut microbial assembly and metabolism. Cell, 185(26), 4921–4936.e15

3. Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., & Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. Nature, 480(7376), 241–244

4. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., & Punta, M. (2014). Pfam: the protein families database. Nucleic Acids Research, 42(Database issue), D222–D230

5. Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., Andrews, E., Ajami, N. J., Bonham, K. S., Brislawn, C. J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T. G., Hall, A. B., Lake, K., Landers, C. J., Mallick, H., Plichta, D. R., … Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature, 569(7758), 655–662

ACKNOWLEDGEMENTS

- Dr. Tiffany Y. Hsu
- Dr. Long Nguyen and the Nguyen Lab
- Dr. Eric Franzosa
- Dr. Curtis Hutenhower and the Hutenhower lab
- Collaborators and co-authors
- Funding sources

CONTACT INFORMATION

Etienne Nzabarushimana, PhD, MPH:

enzabarushimana@mgh.harvard.edu

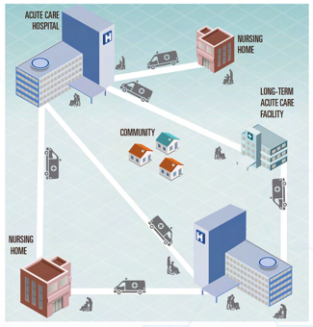
Candida auris and the great ESKAPE: the skin as a reservoir for antibiotic resistance and transmission in American nursing homes

Diana M. Proctor[1], Sean P. Conlan[1], Sarah E. Samson[2], Mary K. Hayden[2], Julia A. Segre[1]

[1]National Institutes of Health, Bethesda, MD, USA
[2]Rush University Medical Center, Chicago, IL, USA.

Introduction

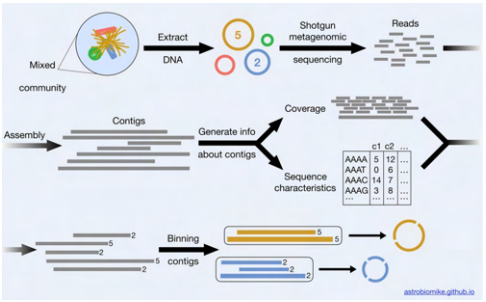
Breaches of infection control practices in nursing homes impacts the larger healthcare ecosystem in America.



CDC: Antibiotic Resistance Threats in the United States, 2019

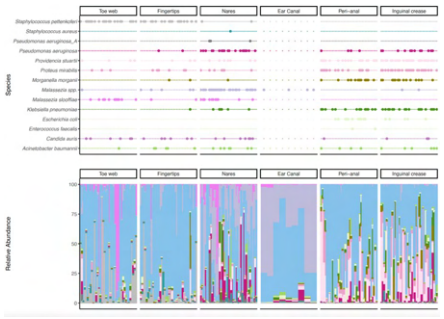
Methodology

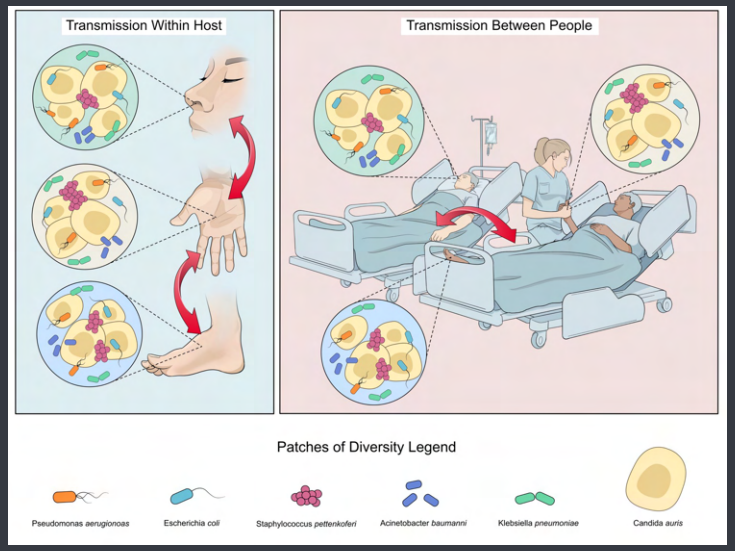
- 36 nursing home patients
- Shotgun metagenomics (N=210)
- Whole genome sequencing (N=75)

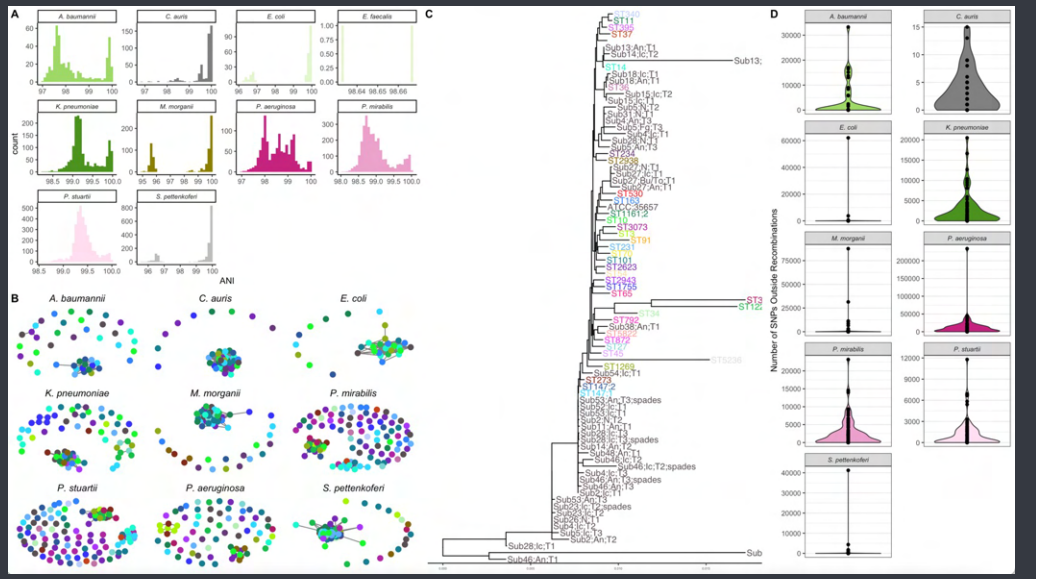


Results

Genome resolved metagenomics yields >300 near complete genomes for Candida auris and the full ESKAPE

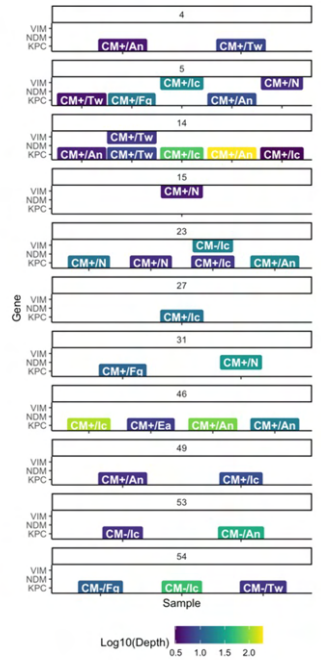




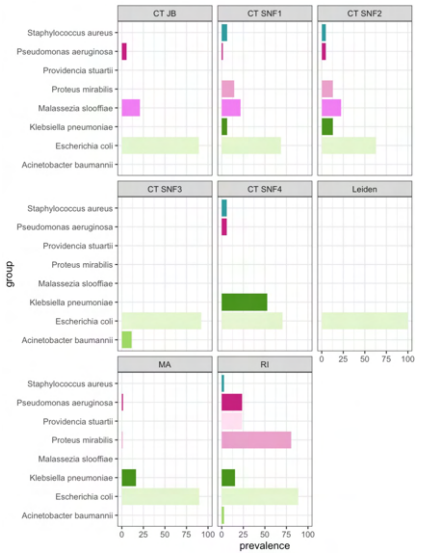


Results

Many individuals harbor resistance genes on skin sites that had been identified by a clinical microbiology lab in rectal or blood samples over 310 days prior, on average.



Similar results seen in 6 other American nursing homes



THE ROLE OF THE GUT MICROBIOME IN THE ASSOCIATION BETWEEN CITRUS FRUIT AND RISK OF DEPRESSION

Chatpol Samuthpongtorn,[1] Allison Chan,[1] Wenjie Ma,[1,2] Fenglei Wang,[3] Long H. Nguyen,[1,2]
Dong D. Wang,[3,6] Olivia I. Okereke,[4,5] Curtis Huttenhower,[6,7,8] Andrew T. Chan,[1,2,5,7,8], Raaj S. Mehta[1,2,8]

[1]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School
[2]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School
[3]Department of Nutrition, Harvard T.H. Chan School of Public Health
[4]Department of Psychiatry, Brigham and Women's Hospital and Harvard Medical School
[5]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital
[6]Department of Biostatistics, Harvard T.H. Chan School of Public Health
[7]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health
[8]Broad Institute of MIT and Harvard

INTRODUCTION

Diet is known to alter the risk of depression. Increasing data also demonstrate a causal role of the gut microbiome in mental illness, via the gut-brain axis. However, it remains unclear how diet and the microbiome mechanistically influence depression risk in humans.

OBJECTIVES

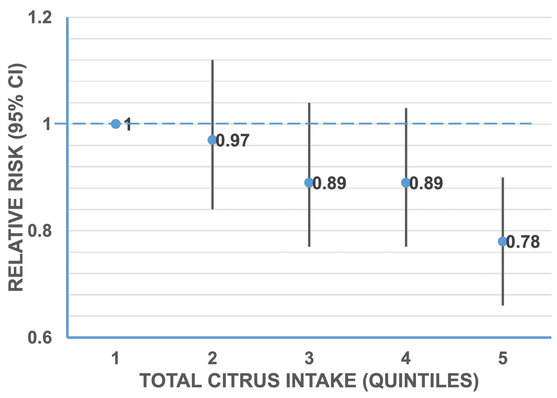
We assessed how gut microbial species and features mediate the association between depression and citrus, a food group that possibly protects against risk of depression.

METHODS

We conducted a prospective study in the Nurses' Health Study II (NHSII) between 2003 and 2017 among 32,427 middle-aged women free of depression at baseline. Citrus intake was determined using validated food frequency questionnaires collected every 4 years. Depression was defined according to physician-diagnosis and antidepressant use. Between 2013-2014, 207 NHSII participants enrolled in a nested substudy, providing up to 4 stool samples (profiled by shotgun metagenomics) and a blood sample (profiled by LC-MS-based metabolomics). Cox proportional hazard models were used to relate citrus intake with depression risk. Linear mixed effects models were used to relate diet with gut microbial features, and microbial features with depression. We also associated microbial features with a depression-risk score, derived according to levels of circulating serotonin and GABA. All models were adjusted for multiple dietary, medication and lifestyle variables including age, BMI, calorie/alcohol intake, and diet quality. We validated our findings in the Men's lifestyle Validation Study (MLVS), a subcohort of 307 men in the Health Professionals Follow-up Study (HPFS). Finally, we used a linear mixed-effects model to examine the role of gut microbial RNA with host transcriptomic gene expression from colon biopsies of 132 Human Microbiome Project 2 (HMP2) participants.

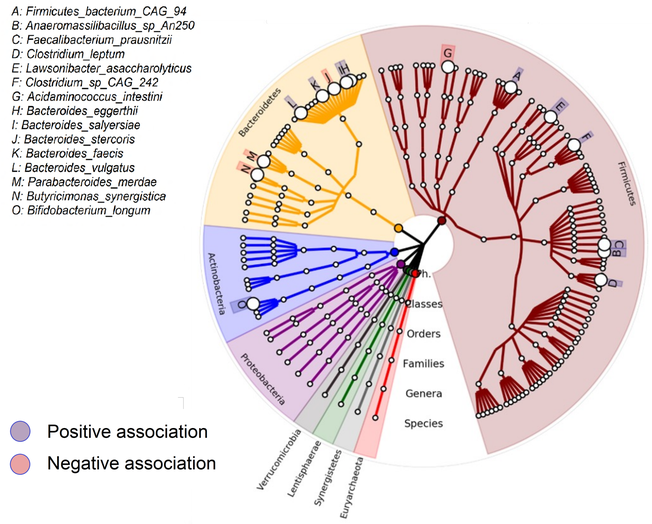
RESULTS

Nurses' Health Study II (NHSII)



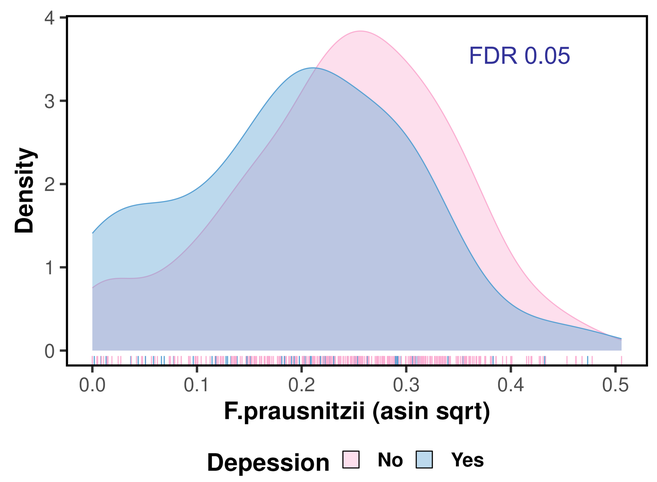
Total citrus intake was associated with a lower risk of incident depression (p_{trend} 0.001), with a multivariable relative risk of 0.78 (95% CI, 0.66-0.90), comparing extreme quintiles.

Mind-Body Study (MBS)



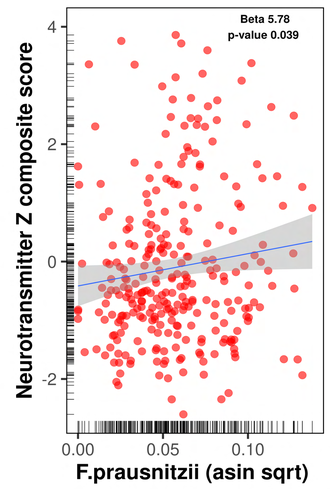
We found 15 species out of 144 whose abundance was significantly associated with total citrus intake using linear mixed effects models (FDR = 0.25)

Mind-Body Study (MBS)



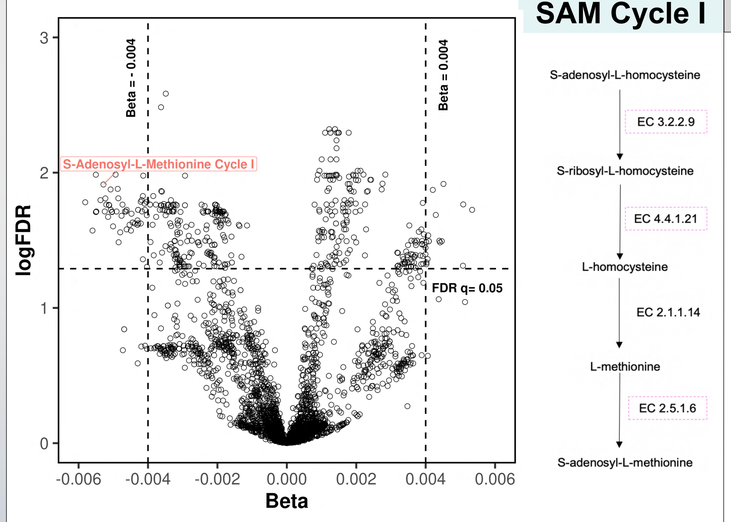
Among these 15 microbial species, *F. prausnitzii* were higher in non-depressed individuals compared to depressed participants

Men's lifestyle Validation Study (MLVS)



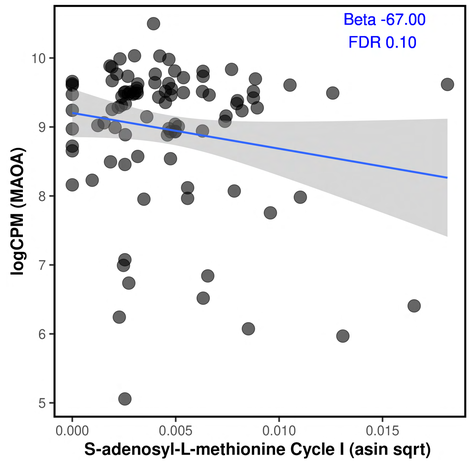
Greater abundance of *F. prausnitzii* was also associated with our metabolomics-based depression-risk score in the MBS (p 0.027), and in the MLVS (p 0.039).

SAM Cycle I



In an exploratory analysis of gut microbial pathways, S-Adenosyl-L-Methionine (SAM) cycle I, encoded by *F. prausnitzii*, was reduced in depressed participants.

Human Microbiome Project 2 (HMP2)



Greater abundance of the SAM cycle I pathway was associated with decreased monoamine oxidase A (MAOA) gene expression in colon

CONCLUSION

Citrus → *F. prausnitzii* and its pathway(SAM cycle I) → ↓MAOA gene → ↓Depression

Greater citrus intake was prospectively associated with lower risk of depression, and with greater abundance of *F. prausnitzii.* Genes encoded by *F. prausnitzii* that produce SAM (a compound known to have antidepressant properties) may help explain these findings, via modulation of intestinal neurotransmitter production. These data offer a potential mechanism by which diet influences the gut microbiome to reduce risk of depression.

ACKNOWLEDGEMENTS

We would like to thank the participants and staff of the Nurses' Health Study II for their valuable contribution.

CONTACT INFORMATION

Chatpol Samuthpongtorn, M.D.
Research fellow at Clinical and Translational Epidemiology Unit (CTEU)
Email: csamuthpongtorn@mgh.harvard.edu

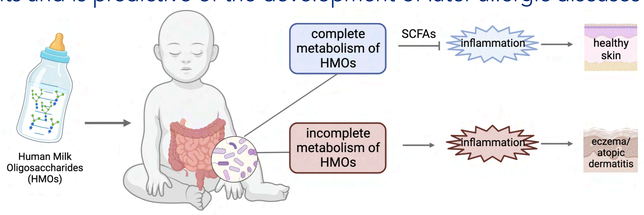
Phylogenetic analysis of bacteria associated with HMO metabolism from gut metagenomes of US infants with eczema

Prioty F Sarwar, Deniz Uzun, Kevin S Bonham, Vanja Klepac-Ceraj

Department of Biological Sciences, Wellesley College, Wellesley, MA 02481

Background

- Human milk plays a key role in the development of the innate and adaptive immune system and gut barrier integrity
- Human milk oligosaccharides (HMOs) are indigestible by infants but acts as a prebiotic to shape the infant gut microbiota
- Eczema is an inflammatory skin condition that affects up to 20% of US infants and is predictive of the development of later allergic diseases

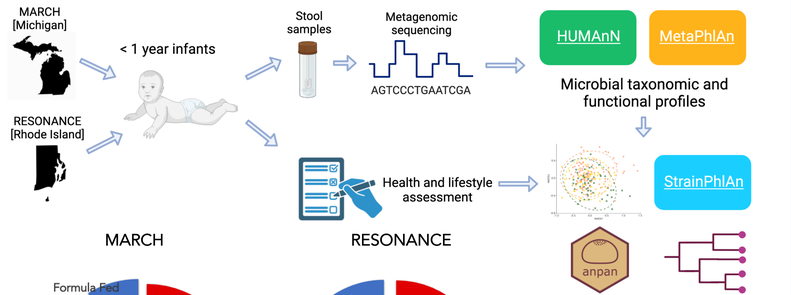


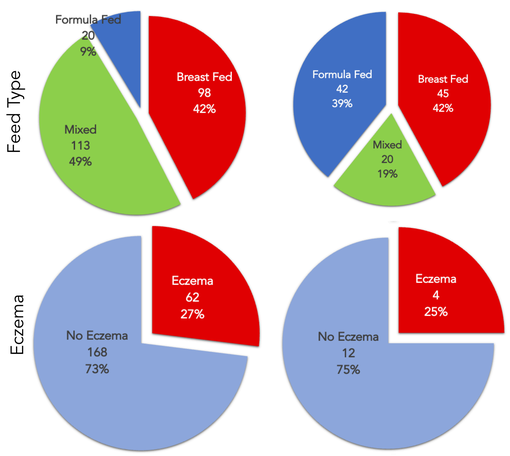
Hypothesis: HMO metabolism (partially) drives the protective effect of human breast milk from eczema

Research Aim

Aim: Identify HMO metabolizing bacteria and the specific genes that correlate with the development of AD/eczema in infants

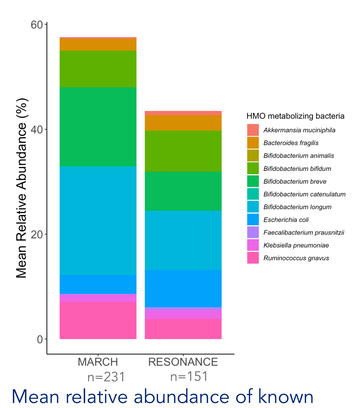
Study Design & Characteristics



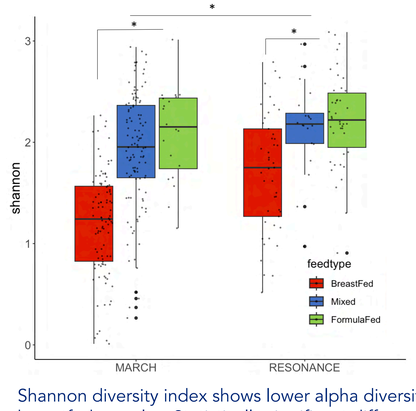


		MARCH	RESONANCE
No. of Samples		231	151
Sex	Male	114	74
	Female	115	62
Mean Age		4.1 months	6.9 months

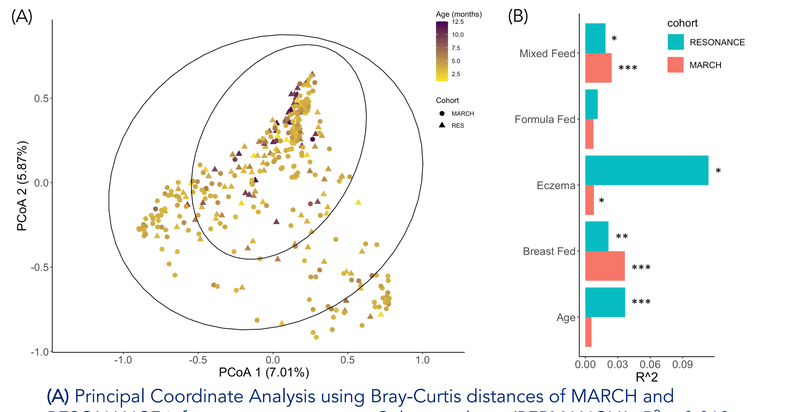
Results: Community Level Metrics



Mean relative abundance of known HMO metabolizers in the MARCH and RESONANCE cohorts.



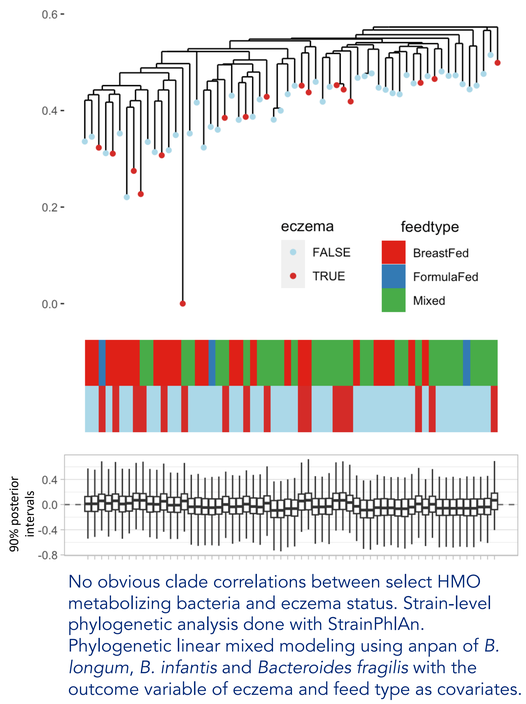
Shannon diversity index shows lower alpha diversity for breastfed samples. Statistically significant differences in alpha diversity observed by cohort and feed type (two-way ANOVA, $p < 0.0001$).



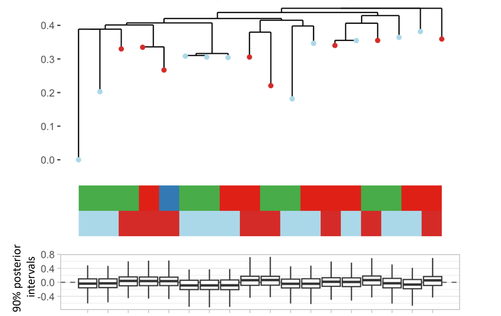
(A) Principal Coordinate Analysis using Bray-Curtis distances of MARCH and RESONANCE infant gut metagenomes Cohort and age (PERMANOVA; $R^2 = 0.018$ p-value < 0.0001) have significant effects on sample beta diversity. (B) Taxonomic profile PERMANOVA by cohort (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Results: Strain level analysis of select HMO metabolizing bacteria from MARCH cohort

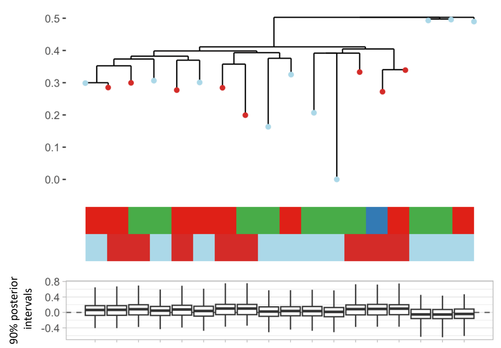
Bifidobacterium longum subsp. *infantis* (n = 60)



Bifidobacterium longum subsp. *longum* (n = 18)



Bacteroides fragilis (n = 18)



No obvious clade correlations between select HMO metabolizing bacteria and eczema status. Strain-level phylogenetic analysis done with StrainPhlAn. Phylogenetic linear mixed modeling using anpan of *B. longum*, *B. infantis* and *Bacteroides fragilis* with the outcome variable of eczema and feed type as covariates.

Acknowledgements

We would like to thank our collaborators Sarah S Comstock and the Huttenhower lab, and the grants from NIH and Wellcome that funded this project. Full acknowledgements and citations: tinyurl.com/bdfzafnr.

Conclusions

- The MARCH and RESONANCE are distinct cohorts that span similar age groups of infancy with RESONANCE skewed towards older infants.
- The two cohorts show differences in taxonomic alpha and beta diversity. RESONANCE cohort has a lower mean relative abundance of HMO metabolizers than MARCH.
- *B. infantis*, *B. longum* and *B. fragilis* assembled from the infant gut metagenomes do not show any clear associations with eczema.

Future Directions

- Complete strain-level analysis of HMO metabolizers in the MARCH and RESONANCE cohorts.
- Phylogenetics Analysis of specific HMO metabolizing gene sets from the infant gut metagenomes of the two cohorts.
- Functional gene set enrichment analysis of our infant gut metagenomes using the CAZY database.

A multi-kingdom genome catalog of the early-life human skin microbiome

Zeyang Shen[1], VITALITY team[2], Kirsten P. Perrett[2], Pamela A. Frischmeyer-Guerrerio[3], Julia A. Segre[1]

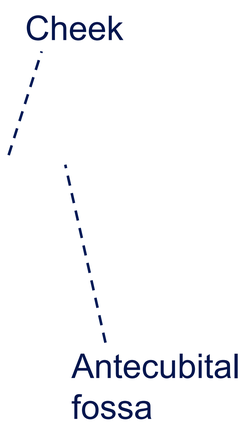
[1]National Human Genome Research Institute, Bethesda, MD, [2]Murdoch Children's Research Institute, Parkville, Australia, [3]National Institutes of Allergy and Infectious Diseases, Bethesda, MD

Introduction

- We recently published the Skin Microbial Genome Collection (SMGC), which greatly expanded the reference genomes for human skin microbiome.
- However, the current reference genomes are largely based on samples from adults in North America and **lack representation from infants** and individuals from other continents.
- Infant skin provides a different cutaneous environment for microbes and a unique habitat to study skin microbiome.

Methods

- We used **ultra-deep shotgun metagenomic sequencing** to profile the skin microbiota of the cheek and antecubital fossa of **212 infants** at age 2-3 months and 12 months who were part of the VITALITY trial in Australia.
- Each sample yielded a median of 28.6 million non-human reads.
- We built **metagenome-assembled genomes (MAGs)** from each sample by using MEGAHIT for assembly, and MetaBAT, MaxBin, and CONCOCT for binning.



The Early-Life Skin Genomes (ELSG) catalog is comprised of 9,194 bacterial genomes from 1,029 species, 206 fungal genomes from 13 species, and 39 eukaryotic viral sequences

- Among 1,029 bacterial species, 699 are newly found on human skin, expanding the total phylogenetic diversity by 56%.

Phylogenetic diversity by phylum

- Fungal specificity of early-life skin.
- Viral diversity of early-life skin.

Methods (cont'd)

- Bacterial MAGs were checked for quality with CheckM and GUNC, fungal MAGs with EukCC, and viral sequences with CheckV. All the MAGs included in the ELSG catalog have **>50% completeness and <10% contamination**.
- Taxonomy was assigned by GTDB-Tk for bacterial MAGs; >95% ANI compared to GenBank genomes for fungal MAGs; BLASTn to the nt database for viral sequences.
- The ELSG catalog improved the classification rate of metagenomic reads by 25% based on Kraken.

Discussion

- We present the largest multi-kingdom genome collection for early-life skin microbiome, which is also the first skin microbial genome collection based on samples from Oceania.
- This resource will be useful to initiate studies of childhood cutaneous disorders, such as atopic dermatitis that typically has an age of onset in infancy.

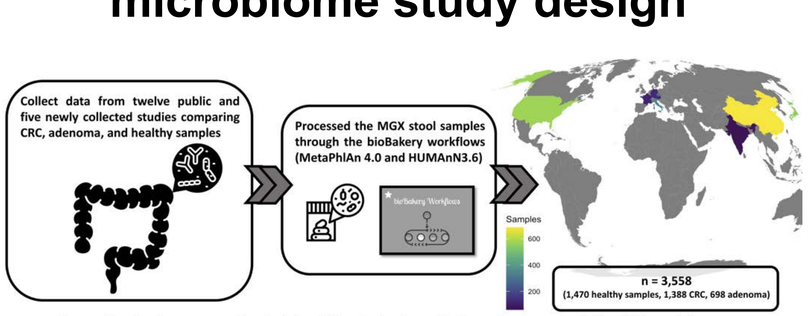
Identifying strain-specific associations in colorectal cancer

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

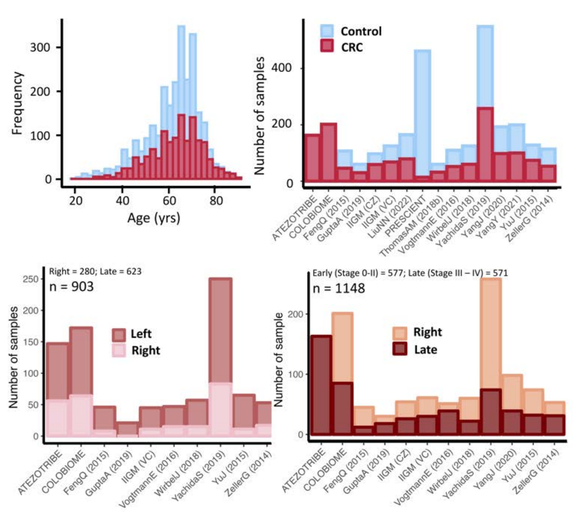
[1]Department of Biostatistics, T.H. Chan School of Public Health, Harvard University; [2]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard; [3]Harvard Chan Microbiome in Public Health Center, Harvard T. H. Chan School of Public; [4]Department CIBIO, University of Trento; 5Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School

In colorectal cancer development, the progression from healthy intestinal cells to benign tumors (adenomas) and then to more malignant forms has profound impacts on the composition of the intestinal microbiota. Additionally, the factors influencing this progression are idiopathic but likely involve a combination of genetics, local tumor environment, and extrinsic factors such as diet. Here, we focus on further elucidating the role of the gut microbiome, a large component of the tumor microenvironment, in cancer initiation and progression by considerably expanding on the current largest meta-analysis to include a total of 3,558 samples from 17 public and private studies. Through expanded sample size, increased resolution of the computational tools, and bioinformatic advances, we have improved the understanding of the gut ecosystem in CRC. While biomarkers of CRC have been well studied in the past, these analyses have focused on species who have previously been isolated. Leveraging new tools, we expand that analysis here to all known microbial taxa and provided novel strain-level insights into the role of the gut microbiome in CRC.

Expanded meta-analysis of CRC-microbiome study design



Several cohorts were collected for this study (n = 5) from Europe and the US and from publicly available studies (n =12) but only when all required metadata was jointly available. Once all raw samples were collected, they were uniformly processed through the bioBakery workflows using KneadData for QC, MetaPhlAn 4 for taxonomic profiling, HUMAnN 3.6 for functional profiling, and StrainPhlAn 4 for sub-species level comparisons.

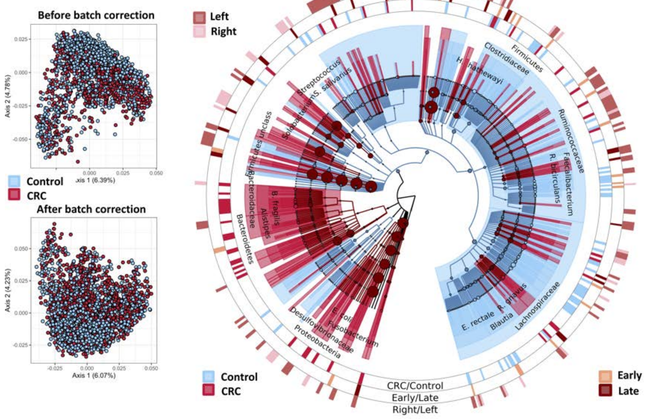


Samples spanned early onset and typical onset CRC samples and covered large areas of the globe in terms of country of origin.

For a subset of the CRC samples, we curated additional information about CRC Stage and sidedness.

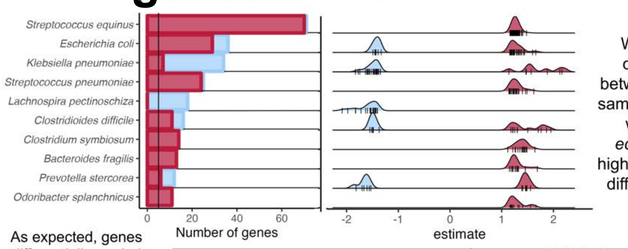
Robust biomarkers of CRC across country, colon location, and stage

Batch correction for study-wise difference reduced the effect of study from 8.0% to 3.6% (PERMANOVA Bray-Curtis on study ID).



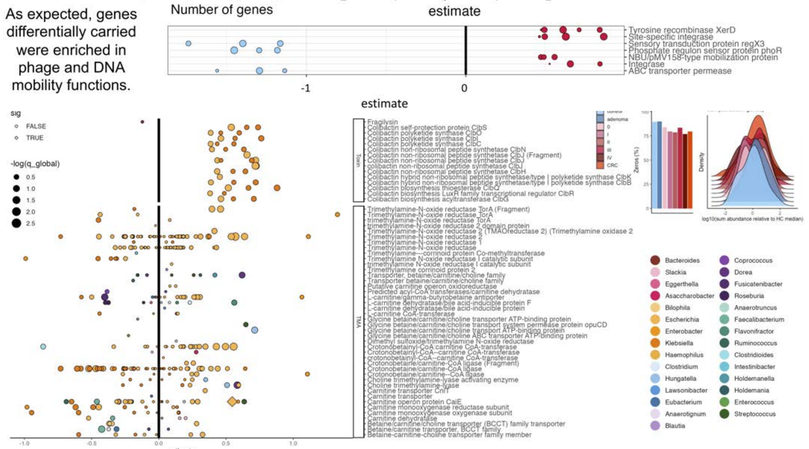
In agreement with previous studies, many species were found to associate with CRC/Control, Early/Late stage, and Right/Left tumor location. This included well known species such as *Fusobacterium nucleatum* and *Bacteroidetes fragilis*. Several new species were also found to robustly associated with CRC including Solobacterium spp. and *Hungatella hathewayi*.

Many species differentially carry genes in health and CRC



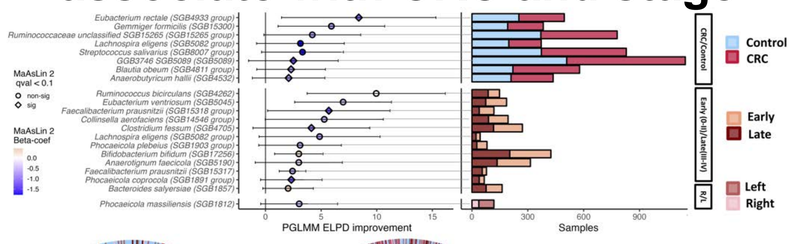
Within species gene coverage difference between CRC and control samples varied by species with *Streptococcus equinus* exhibiting the highest rate of significantly different genes between phenotypes.

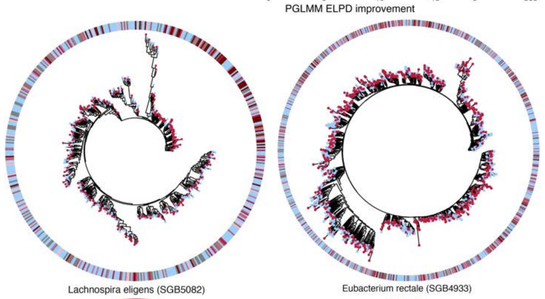
As expected, genes differentially carried were enriched in phage and DNA mobility functions.



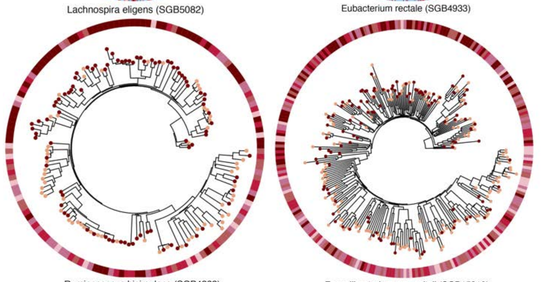
Genes previously quantified to be associated with CRC were found to be differentially carried within species. Colibactin exhibited a trend towards being more likely to be carried in CRC by *E. coli* and *Klebsiella spp*, however increased carriage was not statistically significant. While cutC gene involved in TMAO production did not exhibit the same within species carriage patterns.

Sub-species clades strongly associate with CRC and stage



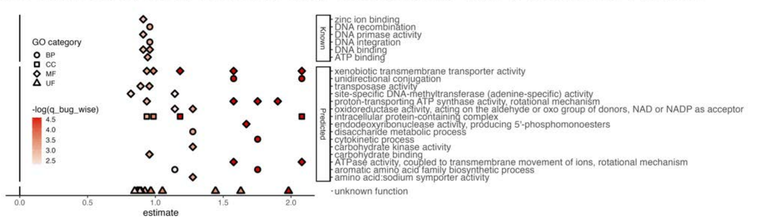


Within species phylogenetic signal was calculated using PGLMMs. (**Top**) Improvement in the model explaining CRC when using the phylogeny in the model. Within species associations were largely independent of species-level associations with CRC, indicating that sub-species clades can be associated with CRC without gain or loss of the species in the gut ecosystem.



(**Bottom**) Broadly sub-clade associations were observed in several species highlighted by *Lachnospira eligens* (previously *Eubacterium*) and *Eubacterium rectale*. Stronger sub-clade behavior was qualitatively observed in the Early/Late comparisons as seen in *Ruminococcus bicirculans* and *F. prausnitzii*.

Ruminococcus bicirculans was identified to be carrying several genes (n = 52) more in late stage (III-IV) CRC than in early stage (0-II) CRC. This suggests either a role for these genes in continued survival of this species under increasing environmental stressors, medication or host homeostasis, or role for the late stage associated sub-clade to become more virulent.



Acknowledgments

We appreciate the valuable scientific contributions of many people on this project including Segata and Huttenhower laboratory members and Prescient Metabiomics. This work was supported by a CRUK Grand Challenge grant, team OPTIMISTICC.

http://huttenhower.sph.harvard.edu

Phage predation of the dominant skin microbiome commensal

A. Delphine Tripp[1,2], Jacob S. Baker[2], Evan B. Qu[2], Tami D. Lieberman[2]

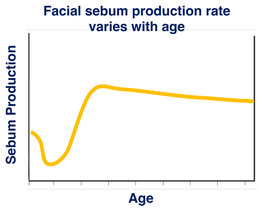
[1]Department of Systems Biology, Harvard University

[2]Institute for Medical Engineering and Science, Massachusetts Institute of Technology

HARVARD MEDICAL SCHOOL | SYSTEMS, SYNTHETIC, AND QUANTITATIVE BIOLOGY

Massachusetts Institute of Technology

Cutibacterium acnes and its phage are a tractable model system to study predator-prey dynamics in the human skin microbiome

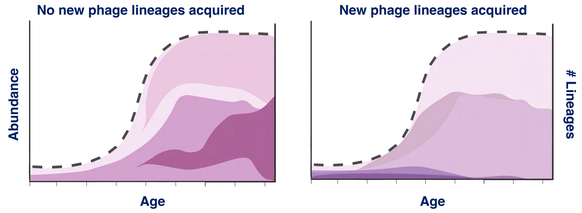


Bacteriophages (predators) promote diversity in microbial communities by generating bacterial (prey) species- and strain-level population fluctuations. The contribution of phage predation on the skin microbiome has not been extensively studied, where it is a likely determinant of on-person colonization. Here, we combine culture-based whole genome sequencing and shotgun metagenomic approaches to examine on-person predator-prey dynamics of the highly abundant and ubiquitous skin commensal *C. acnes* and its phage on sebaceous skin. During early life and adolescence, sebaceous skin is particularly conducive to colonization, in contrast adult skin maintains a stable composition. It is hypothesized that throughout adolescence vertical and horizontal acquisition from parental and non-parental sources creates unique and stable microbiomes across adult individuals. The mechanisms behind this age-dependent selective skin colonization are unclear. This motivates our investigation into this period of ecological disturbance as a unique window to understand how bacterial and phage communities assemble in a human ecosystem.

Questions:

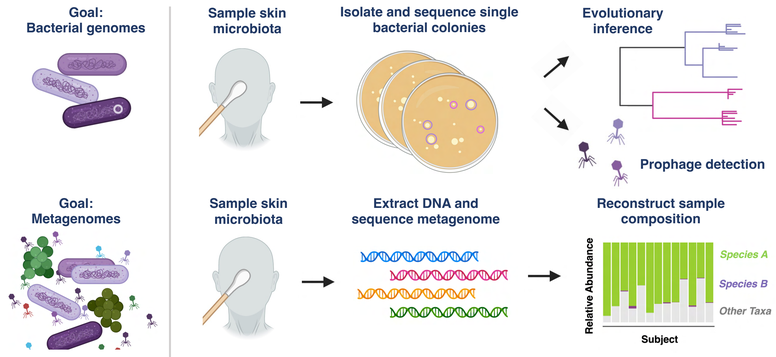
What is the prevalence of *C. acnes* phage populations on human skin?

When are *C. acnes* phage acquired?



A combination of culture-based genomics and metagenomics facilitates mechanistic understanding of on-person ecology and evolution

To study *C. acnes* and its phage, we collected longitudinal facial swabs from K-8 classmates and their family members. We cultured *C. acnes* bacterial isolates and sequenced metagenomes from the faces of 56 individuals across 24 families, including timepoints sampled every six months up to two years apart. We supplemented this with *C. acnes* genomes and skin metagenomes from public datasets.



Findings:

Novel ssDNA lysogen discovered in *C. acnes* genomes

C. acnes prophage carriage is rare and sparse across phylogroups

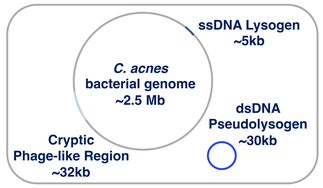
Individuals' *C. acnes* bacteria are dominated by a single prophage type

Although prevalent, *C. acnes* phage populations are not ubiquitous across people

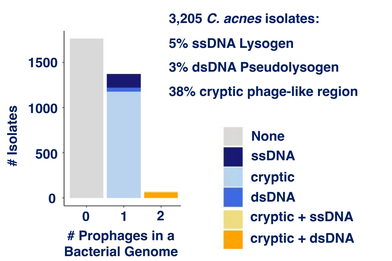
Individual's cutotype shapes on-person *C. acnes* phage population structure

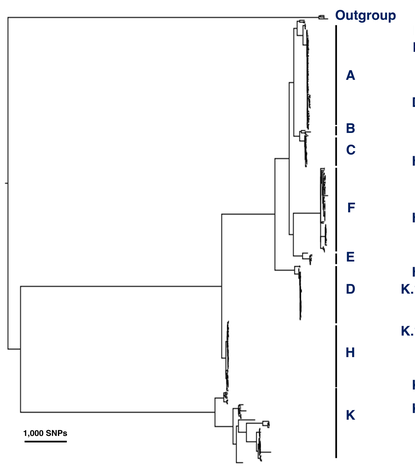
A special thanks to members of the Lieberman lab and Fatima Hussain for advice and support. Thank you to our funding support and study participants.

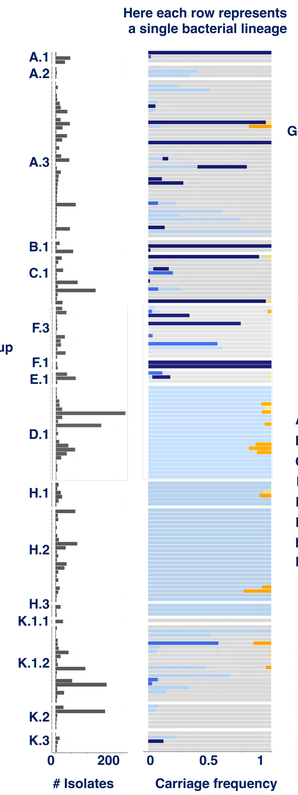
C. acnes prophage carriage is rare and sparse across phylogroups

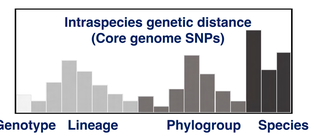


Prophage-like elements found in the genomes of *C. acnes* bacteria include a novel single-stranded DNA (ssDNA) phage, and a previously documented double-stranded DNA tailed (dsDNA) pseudolysogen and cryptic phage-like region.

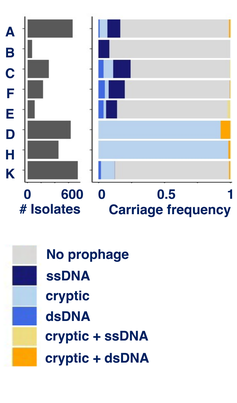






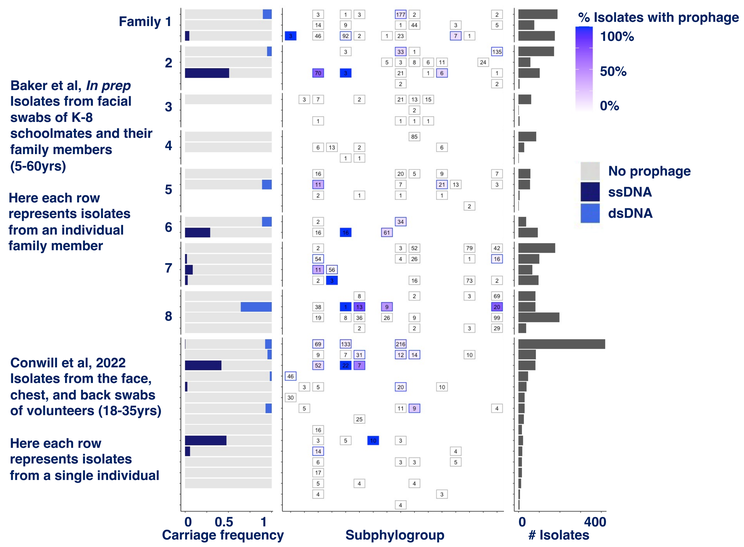


The ssDNA and dsDNA prophage are found in a minority of isolates with high variability across and within lineages suggesting rapid gain/loss. The cryptic phage-like region is found in all phylogroup type D/H isolates yet carried sparsely across other phylogroups suggesting possible mobilization.

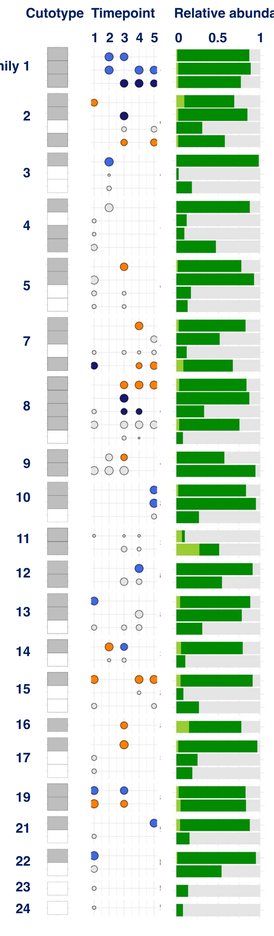


The human host shapes the assembly and structure of on-person *C. acnes* phage populations

Individuals' skin microbiomes harbor multiple co-existing subphylogroups of *C. acnes* bacteria that are dominated by a single type of prophage. Phylogenetic reconstruction of prophage sequences is required to determine if on-person phage populations originate from a single or multiple colonization events.



Individuals' skin microbiomes can be assigned a cutotype based on the dominance of *C. acnes* bacteria in their metagenomes. Younger individuals' skin microbiome begin with a low relative abundance of *C. acnes* and with maturation transition to a *C. acnes* dominated community. Although we never simultaneously observed both phage as lysogens in any individuals' bacterial isolates, we do observe both co-existing stably in metagenomes where both free phage and prophage are collectively detected. On some individuals' we detected only a single phage type, and across younger subjects with non *C. acnes* dominant communities we did not detect either phage. In public datasets we also observe metagenomes lacking either phage independent of *C. acnes* bacteria relative abundance. Together these findings motivate further investigation into the role of the human host environment in shaping on-person phage populations.

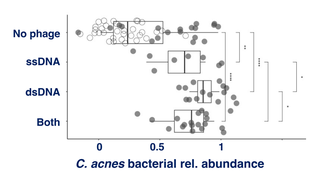


Baker et al, *In prep*
Metagenomics from longitudinal (six-month intervals) of facial swabs from K-8 schoolmates and their family members (5-60yrs)

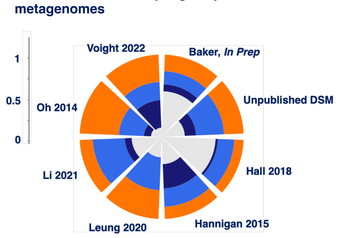
Here each row represents metagenomes from an individual family member

C. acnes dominant
Not *C. acnes* dominant
Both
dsDNA
ssDNA
No phage
Other *Cutibacterium*
C. acnes
Other taxa

Young individuals in the Baker dataset with non *C. acnes* dominant microbiomes lack phage



Detection of *C. acnes* phage in public skin metagenomes



Scalable virome enrichment methods for microbial community detection and quantification

Ya Wang[1,2,3]; Jordan Jensen[1,3]; Eric Franzosa[1,2,3]; Kelsey Thompson[1,2,3]
Krista Cortez[4], Seth Rakoff-Nahoum[4], Curtis Huttenhower[1,2,3]

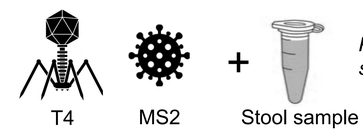
[1]Harvard T.H. Chan School of Public Health [2]Broad Institute of MIT and Harvard [3]Harvard Chan Microbiome in Public Health Center [4]Boston Children's Hospital

Viruses are important but often overlooked members of most microbial communities, including the human gut, where many remain uncharacterized. This is due to a combination of both computational and experimental limitations: viral nucleotides are difficult to enrich and extract, and once sequenced, their uniqueness and divergence make them difficult to classify. The limitations of high-throughput sequencing approaches to address this have been noted previously, but few study has evaluated the efficiency of specific protocols for retaining viral nucleotides from a community while depleting non-viral members. Here, we present our work benchmarking varied experimental protocols to isolate virus-like particles (VLP) from gut microbial communities. Different experimental parameters were evaluated to develop an optimized protocol, which was further validated in mock communities (viruses representing common gut viral families) and in spiked stool samples. The optimized protocol efficiently reduced bacterial signals below the detection limit in mock viral communities. In spiked stool samples, the protocol depleted bacterial signals by approximately 100-fold - although, notably, this still left non-viral nucleotides in the majority in many cases. Different viral clades were also differentially affected by changes in experimental parameters, leading to bias relative to the ground truth. We thus provide a standardized and optimized protocol for gut VLP isolation, with known limits of detection and differential extraction efficiency among potential viral targets.

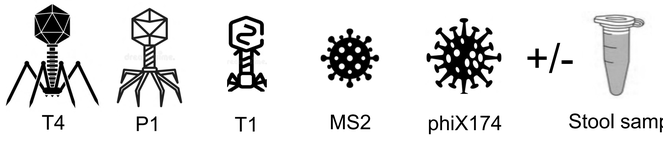
Study design

Experimental parameters known to affect viral extraction from previous works: storage buffer, filtration combinations, methods for nucleic acid concentration after filtration, and various enzymatic treatments.

Combinations of the parameters tested by qPCR in stool samples spiked with simple synthetic viral communities comprising equal amount of an sRNA phage (MS2) and a dsDNA phage (T4).



Protocol optimization in stool samples with synthetic viral cultures of equal T4 and MS2

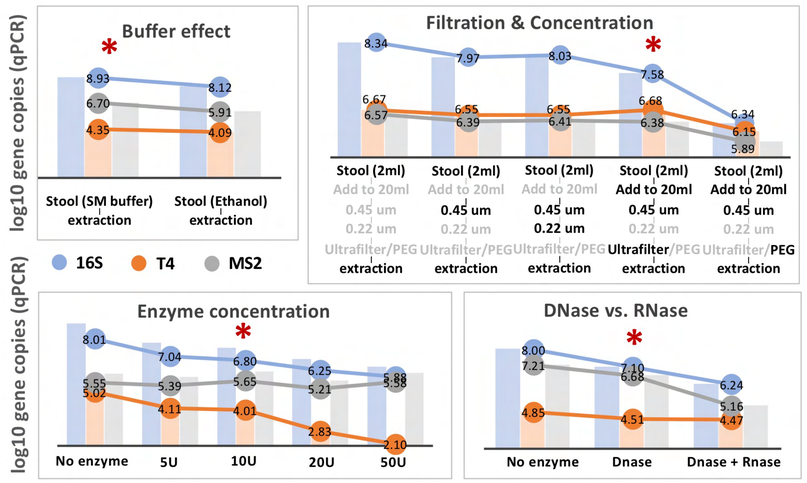


VLP protocol evaluation in mock viral communities of five equally mixed viruses and in spiked stool samples

VLP protocol application in preemie stool samples

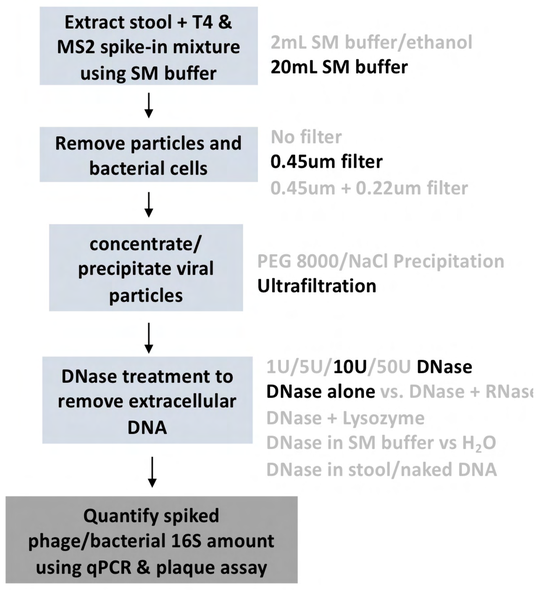
Optimized protocol evaluated in mock communities (equally mixed viruses representing common gut viral families) and in spiked stool samples; further applied in stool samples from preemie babies.

Evaluating different experimental parameters for gut VLP isolation



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.

Purification efficiency of a gut VLP isolation protocol was found affected by various experimental parameters:

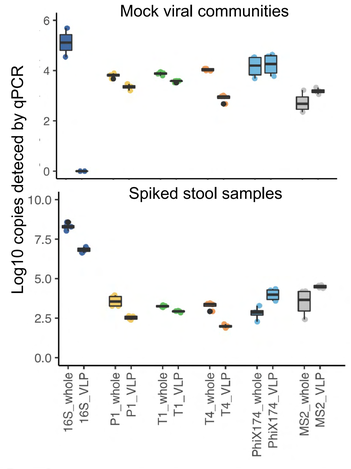


- Ethanol did not differ greatly from SM buffer
- 0.45 μm filter, 100kDa centrifugal filter concentration depletd the most bacteria while preserving the most viruses;
- 10 Unit DNase treatment being most effective
- RNase caused negative effect to RNA viruses.

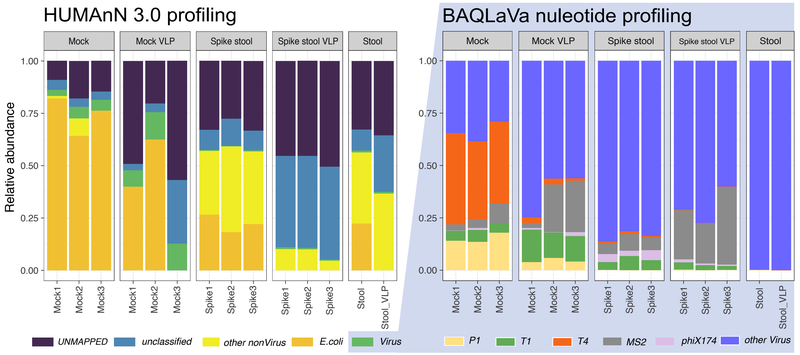
Finalized VLP isolation protocol includes parameters shown in bold.

Optimized VLP protocol depleted bacterial nucleotides from spiked stool

- The optimized VLP protocol reduced bacterial signals by 10^5 copies/ml in mock viral communities, and about 100-fold in spiked stool samples
- P1, T1 and T4 phages were slightly depleted in samples treated with VLP protocol, while phiX174 and MS2 phages enriched.
- Non-viral nucleotides still left in the majority, and different viral clades were differentially affected.



Virome profiling of VLP-enriched and whole community metatranscriptomes



Translation search using HUMAnN identified both bacteria and viruses. Abundant *E. coli* reads likely indicate contaminations in original phage cultures; Viral communities further profiled using a newly-developed integrative computational method, BAQLaVa (Bioinformatic Application for Quantification and Labeling of Viral taxonomy).

Ongoing works

We are currently carrying out analysis of metagenomic and metatranscriptomic sequencing from VLP-treated preemie stool samples to evaluate the protocol on real-world samples at scale. We are also continuing to improve BAQLaVa for virome profiling. Together, we hope these tools will improve experimental and bioinformatic capabilities for gut virome profiling.

Acknowledgments

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

http://huttenhower.sph.harvard.edu

CRISPR spacer acquisition is a rare event in human gut microbiome

Anni Zhang*, Jeffry Gaston, Eric Alm&
anniz44@mit.edu

Alm Lab

Abstract

Host-parasite interactions are vital for all living organisms, including humans versus SARS-COV2. Here, we investigated the host-parasite interactions in the human gut microbiome using temporal whole-genome sequencing (WGS) datasets and metagenomes from healthy individuals. We found that spacer acquisition by CRISPR systems, which defend bacteria against phages, is rare in the human gut microbiome, occurring at an average rate of one spacer per 2,000-5,000 cell divisions, over a period of 6-17 years.

Bifidobacterium longum acquires spacers significantly faster than other species. We identified six highly prevalent consecutive spacers in the same order in *B. longum* from 14 human subjects in the United States and Europe, located on different parts of the *B. longum* genome, but within a highly similar neighborhood (50k-135k bp). This indicates that horizontal gene transfer is the primary contributor to spacer acquisition in *B. longum*.

We developed a model to investigate factors impacting phage infection and CRISPR spacer acquisition. Our model found that low bacterial abundance and frequent dilution events decrease phage infection and selection pressure, resulting in a reduced spacer acquisition rate. Longitudinal metagenome analysis revealed a significant correlation (spearman rho=0.75, p = 9E-9) between bacterial species abundance and spacer acquisition rate.

These findings suggest that CRISPR may not be the primary risk for effective phage therapy for the majority of human microbiome, which may inform future efforts involving phage therapy and pandemic defense.

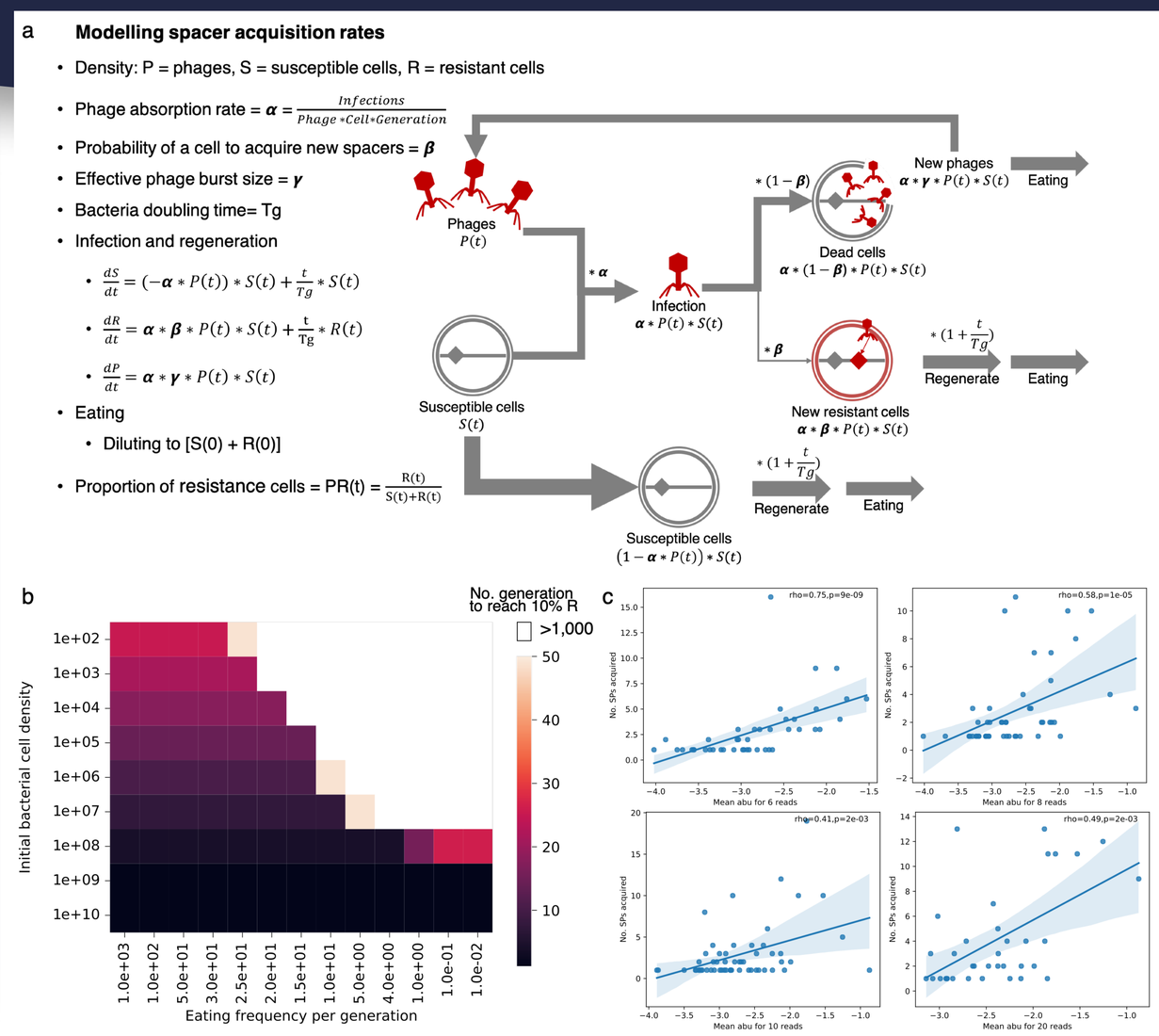


Figure 3. Modelling phage infection and CRISPR spacer acquisition.

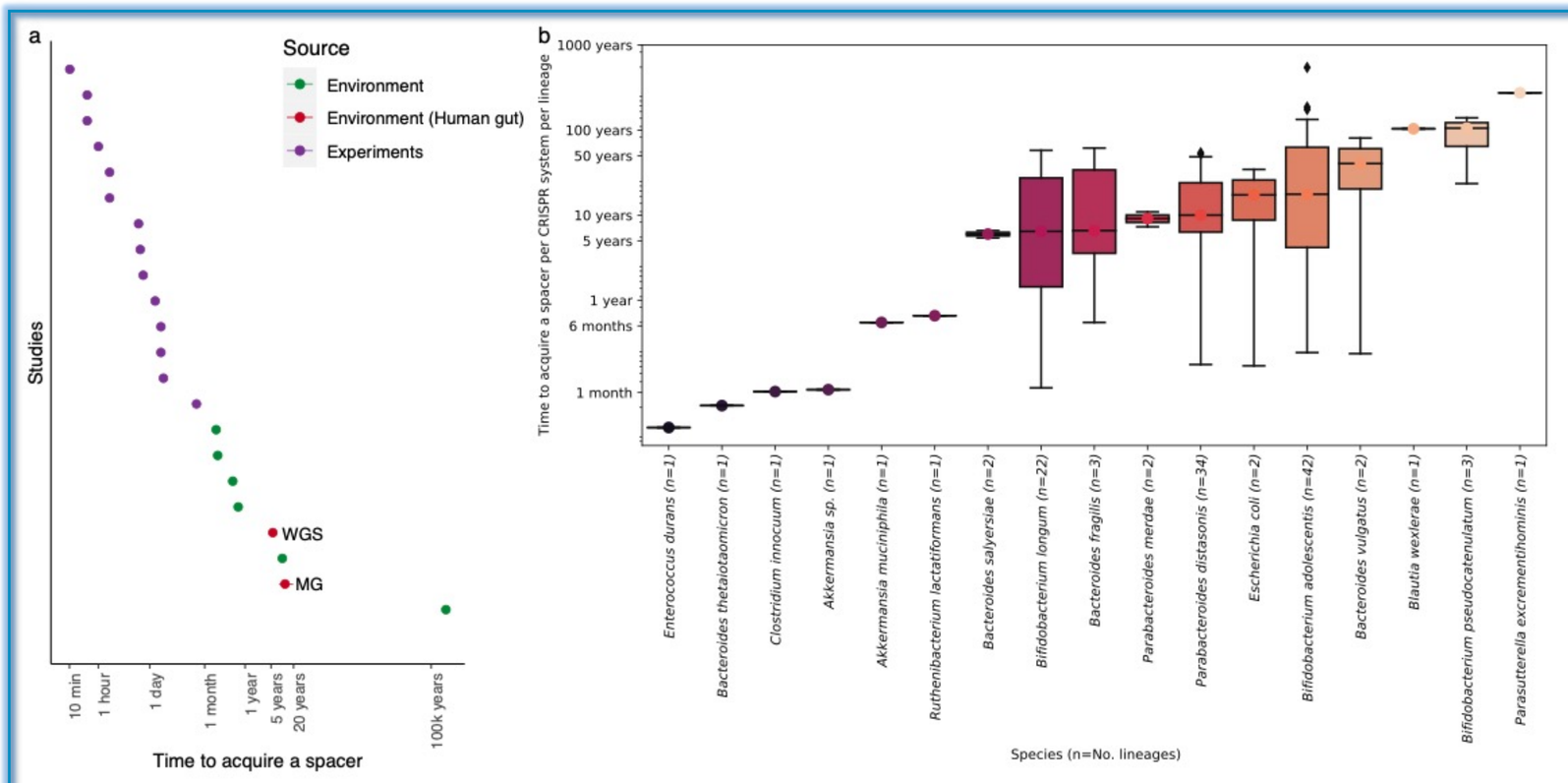


Figure 2. Spacer acquisition rate in human gut microbiome compared to that in other environments.

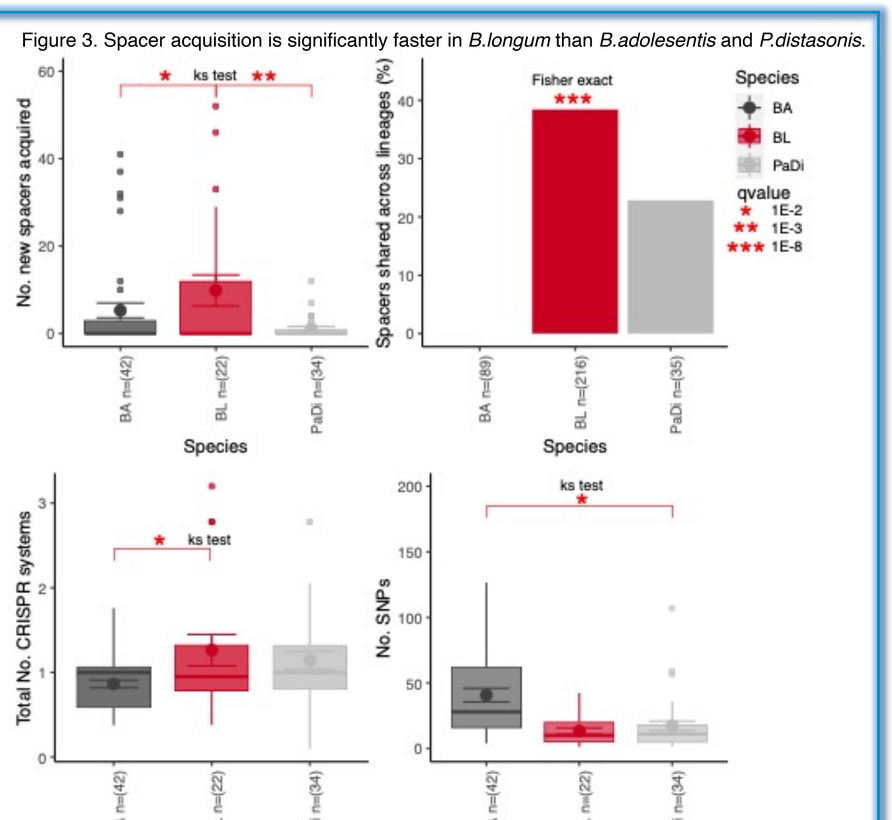


Figure 3. Spacer acquisition is significantly faster in *B.longum* than *B.adolesentis* and *P.distasonis*.

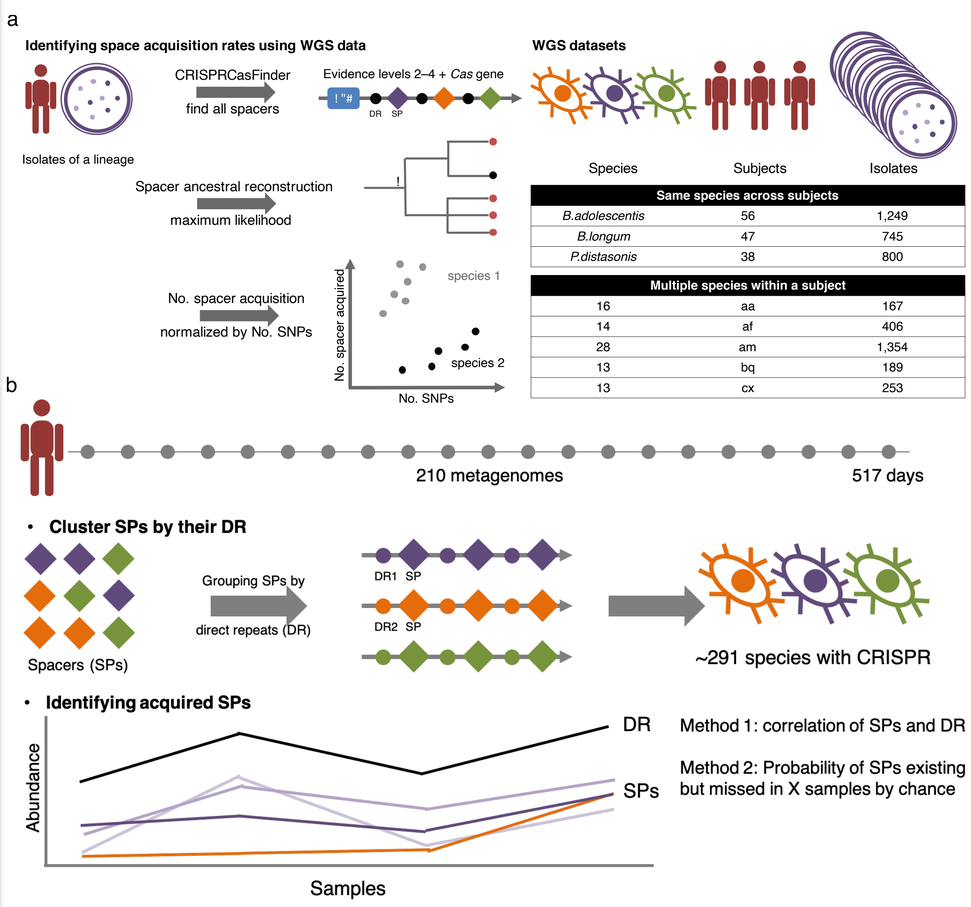


Figure 1. Work flow of spacer acquisition identification using population genetic analysis.

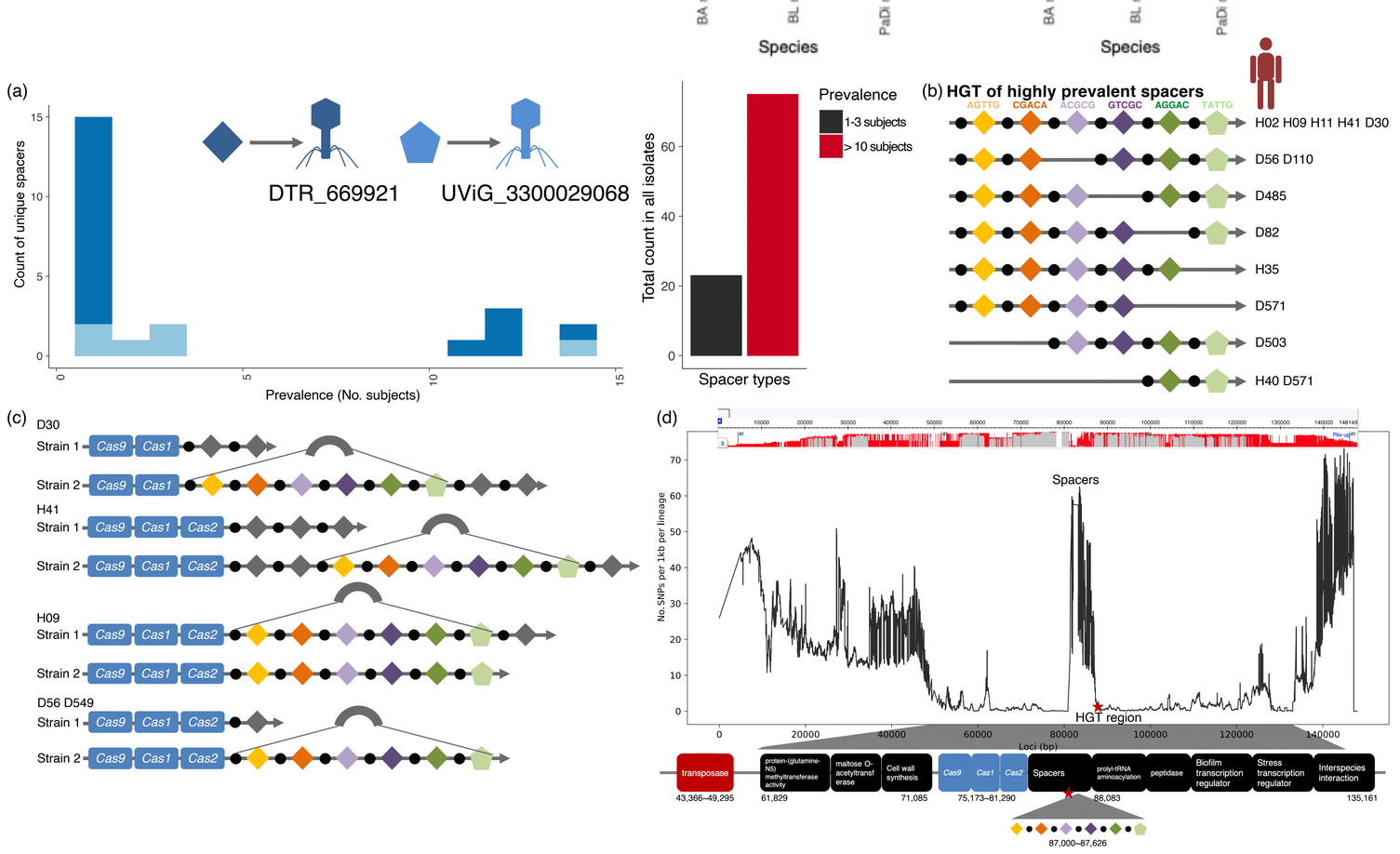


Figure 4. Spacers acquired in B.longum lineages by horizontal gene transfer (HGT)

Conclusion

1. Spacer acquisition is a rare event in human gut microbiome, which agrees with previous literature
   - 1 spacer per 2,000-5,000 cell divisions (whole genome sequencing + metagenomes)

2. Spacer acquisition rate varies among species
   - Acquired spacers in *B.longum* were spread through horizontal gene transfer (HGT)

3. Spacer acquisition rate correlates positive with bacterial abundance

Predicting functions of uncharacterized gene products from microbial communities

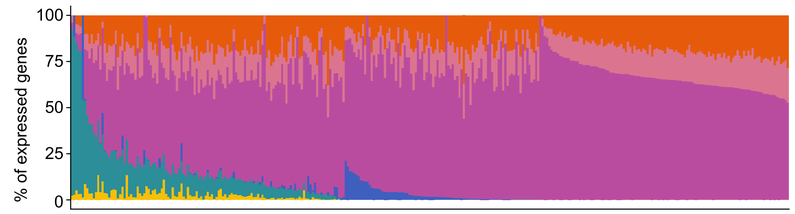
Yancong Zhang[1,2,3], Amrisha Bhosle[1,2,3], Sena Bae[2,3], Kelly Eckenrode[2,3], Xueying Huang[2,3], Jingjing Tang[2,3], Danylo Lavrentovich[4], Lana Awad[2], Ji Hua[2], Xochitl C. Morgan[2,3], Andy Krueger[5], Wendy S. Garrett[1,2,3], Eric A. Franzosa[1,2,3], Curtis Huttenhower[1,2,3]

[1]Broad Institute, [2]Harvard T. H. Chan School of Public Health, [3]Harvard Chan Microbiome in Public Health Center, [4]Department of Systems, Synthetic, and Quantitative Biology, Harvard Medical School, [5]Takeda Pharmaceutical Company Limited

THE HARVARD CHAN CENTER FOR THE MICROBIOME IN PUBLIC HEALTH

Metagenomes are enriched for genes of unknown function

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 (FUGAsseM) to systematically predict functions for uncharacterized microbial proteins by integrating high-dimensional meta-omics data and applied our method to the Integrative Human Microbiome Project (HMP2).



Protein classes (and abbreviations)

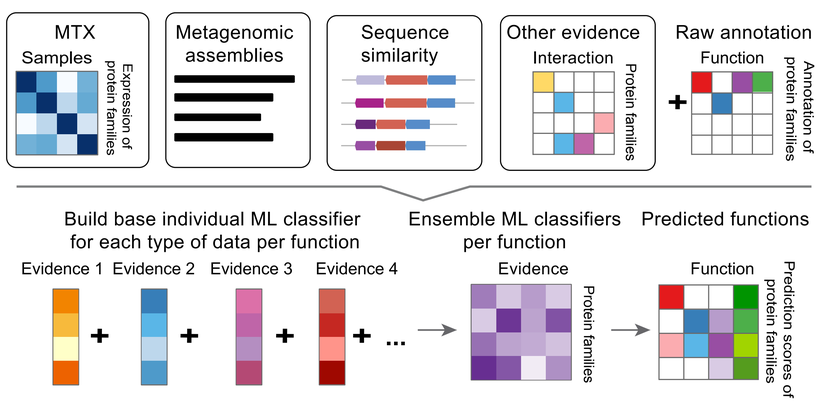
- SC = Strong homology to UniProtKB proteins with informative BP terms
- SC_nonInfo = Strong homology to UniProtKB proteins with non-informative BP terms
- SU = **S**trong homologs to **U**ncharacterized UniProtKB proteins
- UPI = Strong homologs to uncharacterized **UniP**arc proteins
- RH = **R**emote **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" Gene Ontology (GO) biological process (BP) terms, i.e. each BP term contains >1% of annotated genes without any child term passing the criteria.

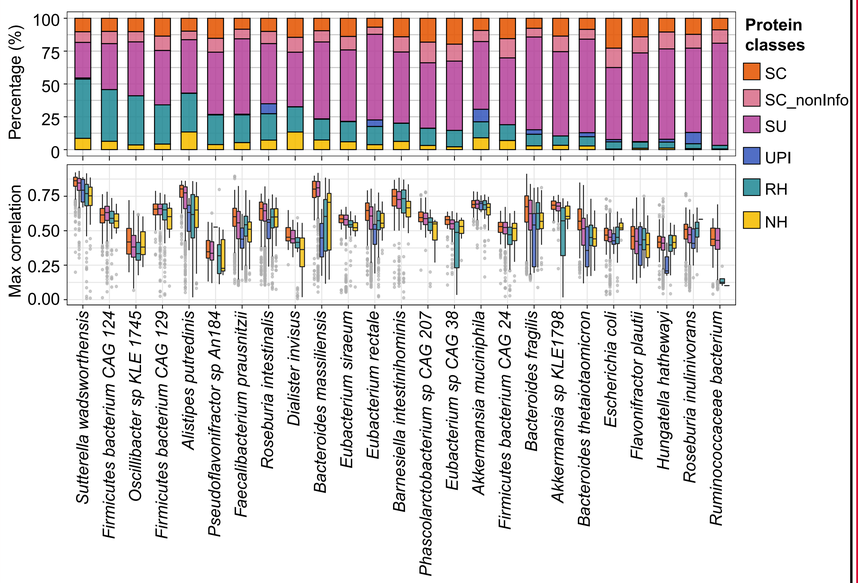
FUGAsseM for function prediction from microbiome

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

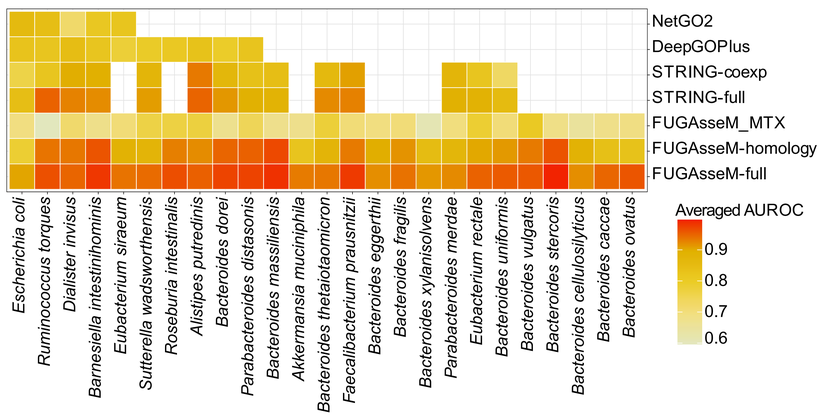


Microbial community meta-omics data improve gene function prediction

Among the 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 transfers of functional annotation under "guilt-by-association" logic. Among the subset of these species with isolate data, proteins linked in STRING networks tend to have higher correlation among MTX networks.



FUGAsseM accurately predicts previously unseen functional annotations

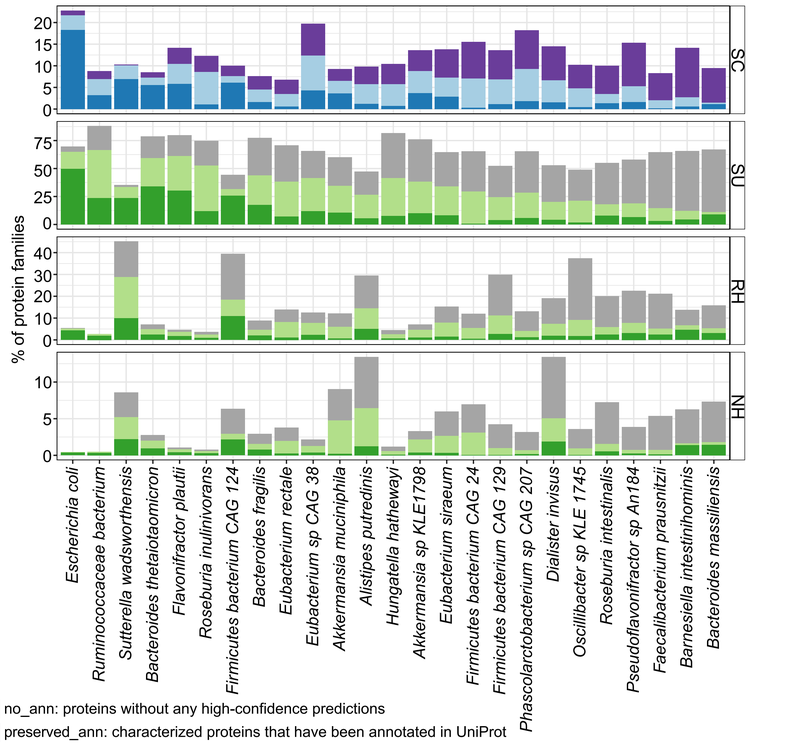


We evaluated FUGAsseM by comparing with other methods for function prediction:

- FUGAsseM's accuracy is improved by aggregating other community-wide data;
- FUGAsseM shows comparable predictions to state-of-art single-organism tools where they overlap;
- FUGAsseM applies to many more species from communities.

Novel characterizations for 1,000s* of microbial genes in the human gut

We applied FUGAsseM to to the 1,595 human gut metagenomes and 800 metatranscriptomes from HMP2, predicting functions of proteins from stratified species in the community.



Here, we summarize the high-confidence BP annotations newly assigned to the 25 species containing the largest numbers of novel (uncharacterized) proteins:

- Species showed different levels of functional characterization;
- Both characterized proteins and uncharacterized proteins were better functional annotated.

Conclusions

- MTX-based coexpression patterns are informative for gene function prediction in microbial communities;
- FUGAsseM predicts functions with high accuracy;
- FUGAsseM refines the functional landscape of microbiomes.

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