2020

Mechanistic Studies to Understand the Role of Human Gut Microbiota in Colonization Resistance and Gut Health

Sudeep Ghimire
South Dakota State University

MECHANISTIC STUDIES TO UNDERSTAND THE ROLE OF HUMAN GUT MICROBIOTA IN COLONIZATION RESISTANCE AND GUT HEALTH

BY

SUDEEP GHIMIRE

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Biological Sciences

Specialization in Veterinary Microbiology

South Dakota State University

2020

DISSERTATION ACCEPTANCE PAGE

Sudeep Ghimire

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

JOY SCARIA
Advisor Date

Jane Hennings
Department Head Date

Dean, Graduate School Date

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Joy Scaria. His guidance throughout the projects helped me to overcome countless hurdles and he has pushed me to grow as a researcher.

I would also like to thank my committee members: Dr. Eric Nelson and Dr. Benoit St. Pierre for their guidance and support. Additionally, I would like to thank Dr. Kincheal Doerner, former Dean of Graduate School at South Dakota State University for his time to put on valuable constructive comments during my research period. Andrew Foley and Mitchell Keena, former undergraduate students in my lab, were instrumental to work with in my initial days in the lab. Also, thank you to my former lab members Dr. Milton Thomas, Dr. Chayan Roy and Dr. Roshan Kumar for their supervision. Furthermore, the effort of current Postdoctoral scholars in the lab, Dr. Abhijit Maji and Dr. Samara Mattiello was invaluable to encourage and push me to work when task seemed impossible.

Supapit Wongkuna; a visiting Ph.D. scholar from Mahidol University, Bangkok, Thailand worked with me for nearly 3 years in our lab and deserves special thanks for her continuous support, understanding and encouragement throughout my pursuit for the degree. I would also like to thank all my current lab members without whom my collaborative projects would go nowhere. Lastly, I would also like to thank Julie for her help with manuscripts, my friends Lokraj Joshi, Dr. Mahesh Shrestha and Tirth Uprety, my well-wishers and family for being the source of ever flowing inspiration.

TABLE OF CONTENTS

ABSTRACT

MECHANISTIC STUDIES TO UNDERSTAND THE ROLE OF HUMAN GUT MICROBIOTA IN COLONIZATION RESISTANCE AND GUT HEALTH

SUDEEP GHIMIRE

2020

Human gut microbiota is comprised of thousands of species which fall into two major categories or "enterotypes" based on dominating bacteria: "*Bacteroides*" and "*Prevotella*". *Bacteroides* enterotype dominates the microbiota of the western population because of a high protein and fat diet, whereas *Prevotella* enterotype is dominate in the gut of the eastern population because of a high carbohydrate diet. While most of the microbiota studies focused on the *Bacteroides* enterotype using both metagenomics and culturomics, *Prevotella* enterotype is understudied, especially in the field of culturomics. The structure of the *Prevotella* dominated gut microbiome is revealed by metagenomic studies, but the individual bacterial species and their characteristics require culturomics and phenotypic characterization studies. To understand how microbes assemble into communities and provide colonization resistance against pathogens, we used the culturomics approach to capture the bacterial diversity of the *Prevotella* dominated community. We cultured 1590 isolates belonging to 102 bacterial species and characterized them for their substrate utilization, short chain fatty acid production and inhibition of *Clostridium difficile*. We found that 66 bacterial species were able to inhibit *C. difficile in vitro*. Furthermore, using a combinatorial community assembly approach, we identified the minimum number of bacteria required to resist *C. difficile in vitro* is twelve. During this culturomics, we also isolated two novel bacterial species from the *Prevotella* dominated microbiome.

Furthermore, to understand how the *Prevotella* dominated community is shaped by certain prebiotic substrates, we combined rice bran and quercetin to analyze their effect on microbiome composition in minibioreactors. The combination of these prebiotics significantly reduces potential pathogenic bacteria by alleviating propionate levels. Also, there are studies suggesting inhibition of *C. difficile* may occur by conversion of bile acids by *C. scindens*. We analyzed the genetic differences of five *C. scindens* strains and tested their ability to inhibit *C. difficile in vitro*. We found a variation in *C. scindens* genomes and their ability to inhibit *C. difficile* in the presence and absence of bile. This research improves our understanding of gut microbiota using culturomics and identifies mechanisms utilized by gut microbes to inhibit *C. difficile in vitro*, which is valuable step in the development of biotherapeutics for gut diseases in future.

Key words: Human gut microbiome, culturomics, colonization resistance, minibioreactor, *Clostridium difficile*

CHAPTER 1: IDENTIFYING *CLOSTRIDIOIDES DIFFICILE*-INHIBITING GUT COMMENSALS USING CULTUROMICS, PHENOTYPING, AND COMBINATORIAL COMMUNITY ASSEMBLY

Sudeep Ghimire [a, b], Chayan Roy [a, b], Supapit Wongkuna[a,b], Linto Antony [a, b], Abhijit Maji [a, b], Mitchel Chan Keena [a], Andrew Foley [a], and Joy Scaria [a, b*]

[a]Department of Veterinary and Biomedical Sciences, South Dakota State University, Brookings, SD, USA.

[b]South Dakota Center for Biologics Research and Commercialization, SD, USA.

Abstract

A major function of the gut microbiota is to provide colonization resistance, wherein pathogens are inhibited or suppressed below infectious level. However, the fraction of gut microbiota required for colonization resistance remains unclear. We used culturomics to isolate a gut microbiota culture collection comprising 1590 isolates belonging to 102 species. Estimated by metagenomic sequencing of fecal samples used for culture, this culture collection represents 34.57% of taxonomic diversity and 70% functional capacity. Using whole genome sequencing we characterized species representatives from this collection, and predicted their phenotypic traits, further characterizing isolates by defining nutrient utilization profile and short chain fatty acid (SCFA) production. When screened using a co-culture assay, 66 species in our culture collection inhibited *C. difficile*. Several phenotypes, particularly, growth rate, production of SCFAs, and the utilization of mannitol, sorbitol or succinate correlated with *C. difficile* inhibition. We used a combinatorial community assembly approach to formulate defined bacterial mixes inhibitory to *C. difficile*. When 256 combinations were tested, we found both species composition and blend size to be important in inhibition. Our results show that the interaction of bacteria with each other in a mix and with other members of gut commensals must be investigated for designing defined bacterial mixes for inhibiting *C. difficile in vivo*.

Importance

Antibiotic treatment causes instability of gut microbiota and the loss of colonization resistance, allowing pathogens such as *C. difficile* to colonize, causing recurrent infection and mortality. Although fecal microbiome transplantation has shown to be an effective treatment for *C. difficile* infection (CDI), a more desirable approach would be the use of a defined mix of inhibitory gut bacteria. *C. difficile*-inhibiting species and bacterial combinations we identify herein improve our understanding of the ecological interactions controlling colonization resistance against *C. difficile* and could aid the design of defined bacteriotherapy as a non-antibiotic alternative against CDI.

Introduction

Normal functioning of the human gut requires a balanced interaction between our mucosal surface, diet, the microbiota, and its metabolic by-products. A major determinant of the gut homeostasis is the presence of a healthy, diverse commensal microbiota which prevents pathogenic bacteria from colonizing the gut or keeping their number below pathogenic levels. This function of the gut microbiome is called colonization resistance (1, 2). Perturbations of the gut microbiome referred to as dysbiosis could result in the loss of colonization resistance (3). Dysbiosis and loss of gut microbiome colonization resistance caused by antibiotics, for example, can predispose people to enteric infections. *Clostridioides difficile* infection (CDI) of the gut following antibiotic treatment is a clear demonstration of this phenomenon. *Clostridioides difficile* is a Gram-positive, spore forming anaerobe and is the leading cause of antibiotic induced diarrhea in hospitalized patients (4).

Antibiotic treatment of CDI often causes recurrence (5). It has been shown that infusion of fecal microbiome from healthy people into CDI patients gut can resolve CDI and prevent recurrence (6, 7). This procedure termed as fecal microbiome transplantation (FMT) has become a common treatment for CDI (8). However, there has been concerns regarding the long-term health consequence of FMT. Recently, there has been reports of weight gain (9) and mortality due to transfer of multi-drug resistant organisms following FMT (10).

The development of defined bacterial mixes from the healthy microbiota, which can resolve CDI, poses an alternative to FMT (11). However, the exact number of species needed in an efficacious defined bacterial mix for CDI remains unknown, and was reportedly in the range of 10–33 species when defined bacteriotherapy was tested in a limited number of patients (12, 13). A mix of spore-forming species tested in phase II clinical trials, despite its initial success, resulted later in recurrence (14). The use of high-throughput anaerobic gut bacterial culturing, coupled with sequencing, improves the cultivability of the gut microbiota (15-17), facilitating the development of culture collections of gut commensals that can be screened for identification of species conferring colonization resistance, or an understanding of the ecological interactions stabilizing or destabilizing colonization resistance.

Here we report the cultivation, using culturomics, of 1590 gut commensals comprising 102 species from healthy human donors. We phenotyped and sequenced genomes of the representative species in this culture collection. We then screened the strains to identify species inhibiting *C. difficile*. A combinatorial community assembly approach in which 256 strain combinations were tested to identify the species interactions that improve or diminish *C. difficile* inhibition. Our results show that both species composition and interactions are both important determinants of *C. difficile* inhibition phenotype. Our approach and culture library—besides advancing the understanding of bacterial community interactions determining colonization resistance—could prove useful in other studies probing the role of gut microbiota in host health.

Materials and Methods

Fecal sample collection, culture conditions, isolate identification, and genome sequencing: We collected fecal samples with the approval of the Institutional Review Board (IRB), South Dakota State University. All procedures were performed following IRB guidelines. We collected fresh fecal samples from six healthy adult donors from Brookings, South Dakota, USA with no antibiotic consumption during the previous year. We transferred fecal samples to an anaerobic chamber within 10 minutes of voiding and diluted them 10-fold with phosphate-buffered saline, and mixed individual fecal samples in equal ratio to make the pooled sample for culturing. We used a modified brain heart infusion (mBHI) broth (ingredients listed in Supplementary Text 1) for culture. We used chloroform and heat treatment to isolate spore-forming species. We plated the 10^4 dilution of the pooled sample on each of the media conditions described (See supplementary text 1) in strictly anaerobic conditions. We selected 1590 isolates from all culture conditions and identified them using MALDI-ToF (MALDI Biotyper, Bruker Inc) against reference spectra (Supplementary Table 1). We identified isolates unidentified at this step using 16S rRNA gene sequencing, for which we prepared total genomic DNA using the OMEGA E.Z.N.A genomic DNA isolation kit (Omega Bio-tek, GA) according to the manufacturer's protocol. We amplified full length bacterial 16s rDNA using the universal forward (27F) and reverse (1492R) primers, respectively, under standard PCR conditions. We sequenced the amplified DNA using the Sanger dideoxy method. We trimmed raw sequences generated for low-quality regions from either end and constructed consensus sequences from multiple primers using Genious (18) with default parameters. Based on full length 16S rRNA gene sequence similarities, we determined the phylogenetic relationships of the

isolates. For bacteria initially identified using MALDI-TOF, we extracted full length 16s RNA gene sequences from the genome using Barrnap (https://github.com/tseemann/barrnap) for phylogenetic tree creation. The sequences were aligned using MUSCLE (19) and evolutionary distances were computed using Jukes-Cantor method. Finally we created the Neighbor Joining tree with a bootstrap of 1000 replicates using MEGA6 (20).

To further characterize the strain library, we selected representative isolates from each species for whole genome sequencing, extracting genomic DNA from overnight culture of the isolates using the OMEGA E.Z.N.A genomic DNA isolation kit (Omega Bio-tek, GA) according to the manufacturer's protocol. We prepared sequencing libraries the Nextera XT kit, sequencing them using Illumina 2x 300 paired-end sequencing chemistry on the Miseq platform (See Supplementary Text 1). We first filtered the raw reads for quality and sequencing adaptors with PRINSEQ (21), and then assembled them *de novo* using Unicycler (22), performing quality check of the assembly results with QUAST (23) and Bandage (24) all in default mode. We performed gene calling using Prokka (25) with a minimum ORF length of 100 bp.

Characterization of donor fecal samples using shotgun metagenome sequencing: We extracted total community DNA from 0.25 g of each donor fecal sample using MoBio Powersoil DNA isolation kit according to the manufacturer's instructions. To enrich microbial DNA, using previously published protocol, we depleted host DNA in the isolated total DNA before sequencing (26). We prepared the sequencing library from 0.3 ng of the

enriched DNA using the Nextera XT library preparation kit (Illumina inc. San Diego, CA), performing sequencing following the protocols used for bacterial genomes described above. After quality correction, we removed human host reads using Bowtie2 v.1.1.2(27) and performed taxonomy assignment using Kaiju (28) in greedy mode. We searched reads against the proGenomes (29) reference database of protein sequences containing a non-redundant set from more than 25,000 genomes from every species cluster recovered by specI (30), using the NCBI non-redundant database for comparative analysis. Thereafter, we calculated the Simpson dominance index (D), Shannon Diversity Index (H) and Shannon Equitability Index (E_H), using an assembly-based approach to characterize the donor fecal metagenomes. For this, we assembled reads *de novo* using default metaSPAdes (SPAdes 3.12.0) (31), specifically designed for assembly of complex metagenomic communities. For initial assembly, we error corrected reads using spades-hammer with default parameters, thereafter checking assembly results with default MetaQUAST v 5.0 (32). We removed contigs of less than 500 bp from the resultant datasets, and performed ORF predictions on the filtered contigs with MetaGeneMark (33) with a minimum length cutoff of 100 bp.

Computation of gene repertoire and the functional analysis of isolate genomes and donor fecal metagenomes: To determine the gene repertoire in the isolate genomes and donor fecal metagenomes, we constructed a non-redundant gene catalog from our total data using cd-hit (34, 35) for comparison against the previously published integrated catalog of reference genes in the human gut microbiome (36). After gene calling, we clustered the concatenated datasets from the culture library using cd-hit at >95% identity with 90%

overlap. We checked these datasets using BLAT to avoid over-representation in the gene catalog. We then compared our gene dataset against the previously mentioned integrated catalog (36). For functional mapping, we mapped amino acid sequences from all the individual datasets and clustered non-redundant sets against the EggNOG database v3.0 (37).

For gaining a better insight into the putative and hypothetical population genomes present within the donor fecal samples, but not isolated using the culturomics approach, population genome bin creation is considered superior to taxonomy assignment of the raw reads. We therefore constructed population genome bins from the metagenomes using MaxBin2with acceptance criteria of 90% completeness and < 5% redundancy, mapping back raw reads on the assembled contigs using Bowtie2 (38) for coverage information. We further analyzed all high-confidence bins with specI (30) for species cluster determination. To determine the abundance of isolated species from the pooled donor fecal samples, we measured the coverage by read mapping with Bowtie2 (38) at 95% identity level. For the functional analysis, we performed KEGG annotation for the ORFs obtained from the pooled donor fecal metagenome and the isolate genomes. We searched data from all comparisons for KO modules using GhostKOALA (39), and performed hierarchical clustering of the datasets to generate heat maps with R (http://www.R-project.org/). We used Traitar (40) under default parameters to predict 67 phenotypes from the whole genomes of all species in the culture collection.

Phenotypic characterization of the isolated strains: To correlate the genomic features with phenotypes, we further characterized the strains for which genomes were sequenced by determining the following phenotypic properties:

Carbon source utilization: We determined the ability to utilize 95 carbon sources using Biolog Biolog AN MicroPlate^{TM}. Briefly, we grew strains on mBHI plates anaerobically for 24 h–48 h at 37^0C. We used a sterile cotton swab to scrape cells from the plates and suspended them in AN inoculating fluid (OD_{650} < 0.02), using 100 µl of this suspension to inoculate AN MicroPlate^{TM} in duplicate, and incubated them at 37^0C anaerobically. We took OD_{650} readings at 0 h and 48 h post-inoculation and normalized the results for growth against water and 0-h OD_{650} values.

Production of SCFA: To analyze the SCFAs produced by isolates, we grew strains in mBHI for 24 hours in anaerobic conditions and added 800 µl of the bacterial culture to 160 µl of freshly prepared 25% (w/v) m- phosphoric acid, and froze them at -80^0C. We thawed samples and centrifuged them (>20,000×g) for 30 min. We used 600 µl supernatant for injection into the TRACE1310 GC system (ThermoScientific, USA) for SCFA analysis.

Identification of C. difficile*-inhibiting strains:* We used a co-culture assay in which pathogen and test strains were cultured together at a ratio of 1:9 to identify *C. difficile*-inhibiting strains in our library, using only those strains reaching OD600 of 1.5 after 24 h of growth in mBHI; 82 species met this criterion. We used *C. difficile* strain R20291 as the reference strain in the first assay. Briefly, we grew all test strains and *C. difficile* R20291

in mBHI medium anaerobically at 37^0C and adjusted the OD_{600} to 0.5. The pathogen and the test were mixed together at a ratio of 9:1 and incubated for 18 hours anaerobically at 37^0C. We then plated 10^{-5} & 10^{-6} dilutions onto *C. difficile* selective agar, using monoculture of *C. difficile* R20291 as a positive control. We compared colony forming units (CFUs) enumerated from co-culture plates against the C. *difficile* R20291 control. In identifying *C. difficile*-inhibitors in healthy and CDI patients, we calculated the frequency of 16 top *C. difficile*-inhibitors in our collection by metagenome read mapping from a previously published dataset (41).

Sequence data Availability: All metagenomic datasets from this project has been deposited in NCBI SRA under the bioproject PRJNA494584. Raw whole genome sequence data for all the isolates has been deposited under the BioProject PRJNA494608.

Results

Single medium based culturomics retrieves high species diversity from donor fecal samples: In this study, we used metagenome sequencing to characterize the fecal microbiome composition of healthy human donors and culturomics to develop a strain library to identify *C. difficile* inhibiting species. As a first step, donor's fecal samples were characterized using shotgun metagenome sequencing. To this end, fecal samples from six donors were sequenced individually and also after pooling in equal proportion. We used high sequencing depth for the metagenome sequencing. Collectively, the datasets from all

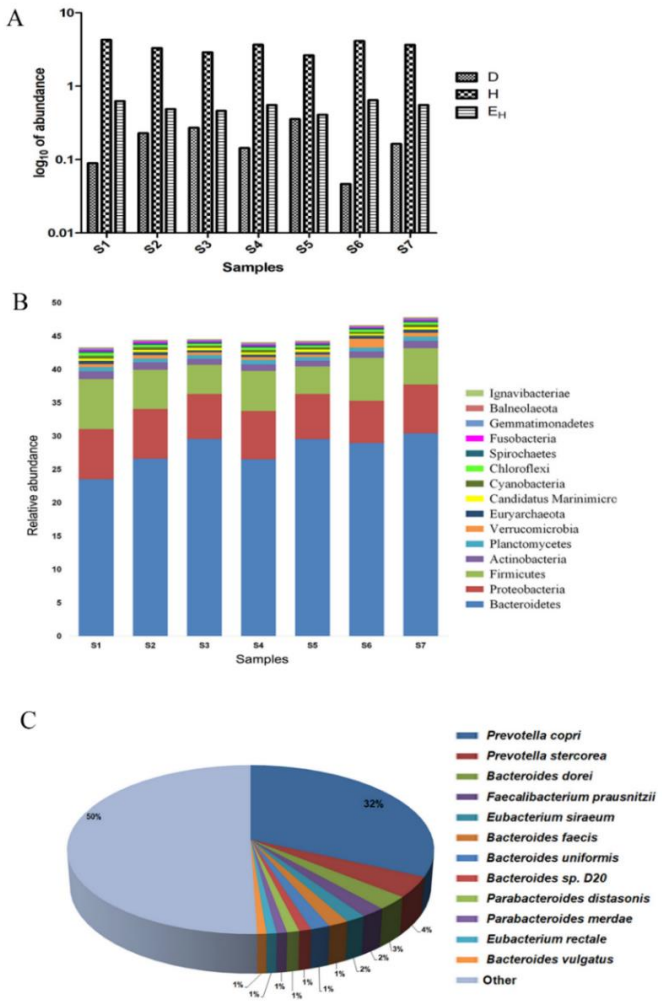


Figure 1: Gut microbiome composition of donor's fecal samples. A. Variations in the Simpson Dominance (D), Shannon Diversity (H) and Shannon Equitability (EH) indices calculated based on the raw reads taxonomic affiliations for individual fecal samples (S1–S6) and pooled fecal sample (S7). B. Phylum level distribution of bacteria for individual fecal samples (S1–S6) and pooled fecal sample (S7). We calculated relative abundance values as compared to the total identifiable reads from any dataset. C. Species-level distribution of the pooled donor fecal sample (S7). We performed taxonomic classification using Kaiju. We referred all species with under 1% abundance to "other" category.

samples constituted 48.9 Gb of data. To determine the diversity of the samples, Simpson dominance index (D), Shannon diversity index (H) and Shannon equitability index (E_H) were calculated for individual samples as well the pooled material. This statistical analysis (Hutcheson t-test, $p > 0.1$) revealed that all donor samples were similar in diversity indices and pooling the samples in equal proportion maintained the overall population structure of the individual samples (Figure 1A). The donors in this study were recent migrants to the United States from Asia and were expected to have a high proportion of *Prevotella* (*Prevotella* enterotype) in the gut microbiome. Consistent with this expectation, taxonomic diversity of the samples in phylum level (Figure 1B) showed the dominance of *Bacteriodetes* while the most abundant genus was *Prevotella* (Figure 1C).

Using culturomics, we developed a strain library from the pooled fecal samples. Previous studies have prevalently used assorted media to isolate gut bacteria (15). For mechanistic studies to understand the microbial community interaction, strains need to be pooled in a single nutrient medium. While the approach of using different media conditions is useful in retrieving high diversity, strains isolated in various media conditions may not be able to grow in a single media. This would prevent the use of a culture library in community assembly studied in a universal medium. To avoid this problem, we used a modified Brain Heart Infusion (mBHI) as the base medium for culturing. We isolated several strains from nine species from the mBHI medium. We reasoned that if the species that were growing fast in the mBHI medium is suppressed, additional species diversity could be isolated using the same medium. We therefore supplemented mBHI with different combinations of antibiotics selected to suppress the formerly dominant strains, also using heat shock and

chloroform treatment to select for spore-forming species. We used 12 conditions for culturing (Supplementary Text 1), selecting 1590 colonies from these conditions. We determined strain species identity using MALDI-ToF, and 16S rRNA sequencing (Supplementary Table 1). We thus isolated 93 more species from the same sample, increasing the total diversity in our strain library to 102 species (Figure 2). In Figure 3A, we present the frequency of each species isolated in each culture condition. We further examined whether our approach could isolate high- and low-abundance species from the sample. We therefore sequenced the genome of one representative isolate from each species in our library (Supplementary Table 2), mapping the metagenome reads (Supplementary Table 3) against the individual species genomes. We mapped 37,669,789 reads obtained from the pooled sample to species whole genomes (Figure 3B), matching 19,109,642 reads—34.57% of the total metagenomic diversity—to the metagenomic reads. Our single medium-based culturomics method was able to isolate about 34.57% of the pooled culture sample diversity. We mapped about 20% reads against the *Prevotella copri* SG-1727 genome isolated from six different conditions—unsurprising, as the fecal donors belonged to the *Prevotella* enterotype. However, we isolated low-abundance species such as *Olsenella umbonata*, for which only 0.5% reads were mapped, in eight different conditions (Figure 3B). We isolated both low- and high abundance species with our technique and used metagenome binning to estimate the number of species missed by our method. Matching metagenome bins against cultured species genomes, we found 50 matching culture isolates and 33 bins matching none, indicating that our method failed to cultivate those metagenome bins (Supplementary Table 4).

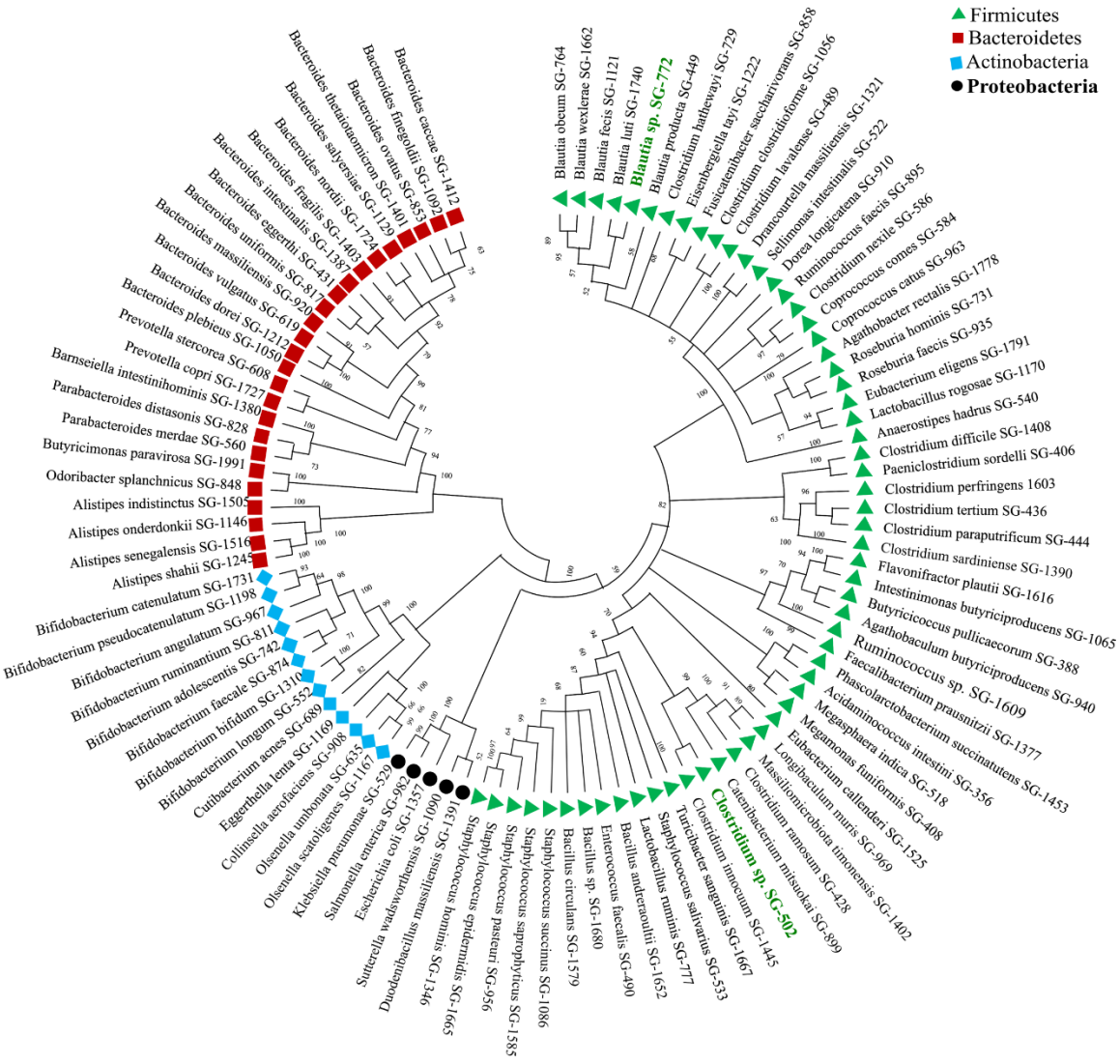


Figure 2: Neighbor-Joining tree of the full length 16S rRNA gene sequences of 102 cultured species isolated from the pooled donor fecal sample. We computed the evolutionary distances using the Jukes-Cantor method, presented in the units of the number of base substitutions per site in MEGA6. Symbols and colors represent four different bacterial phyla, as referred to the legend. We have highlighted putative novel species (n = 2) with "green" text.

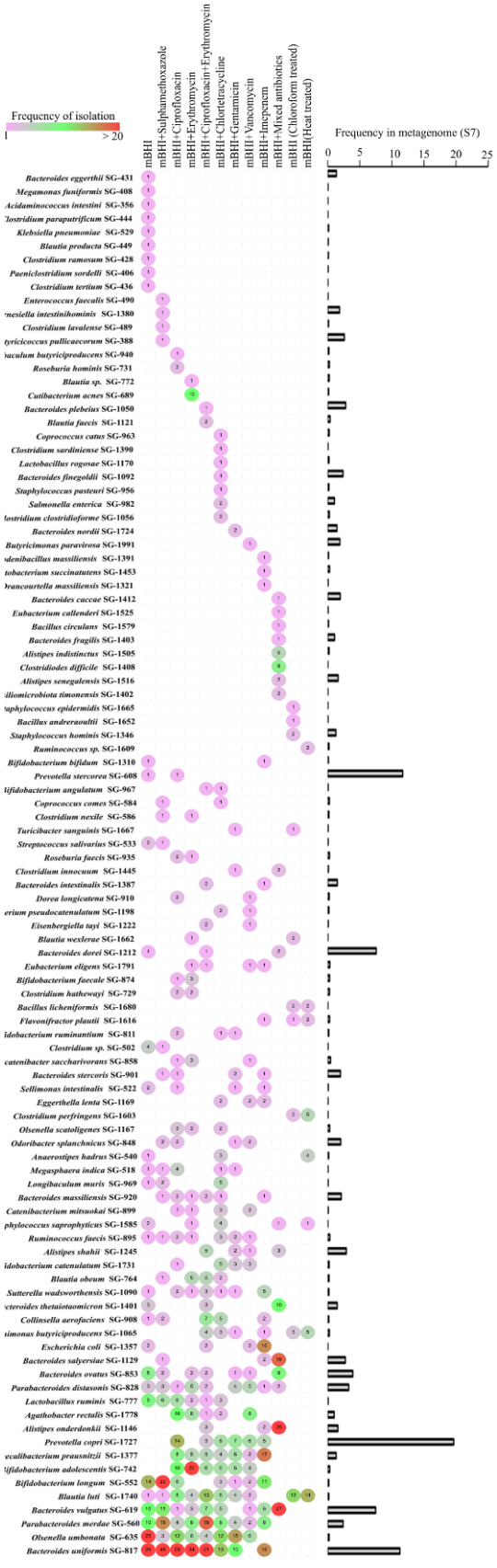
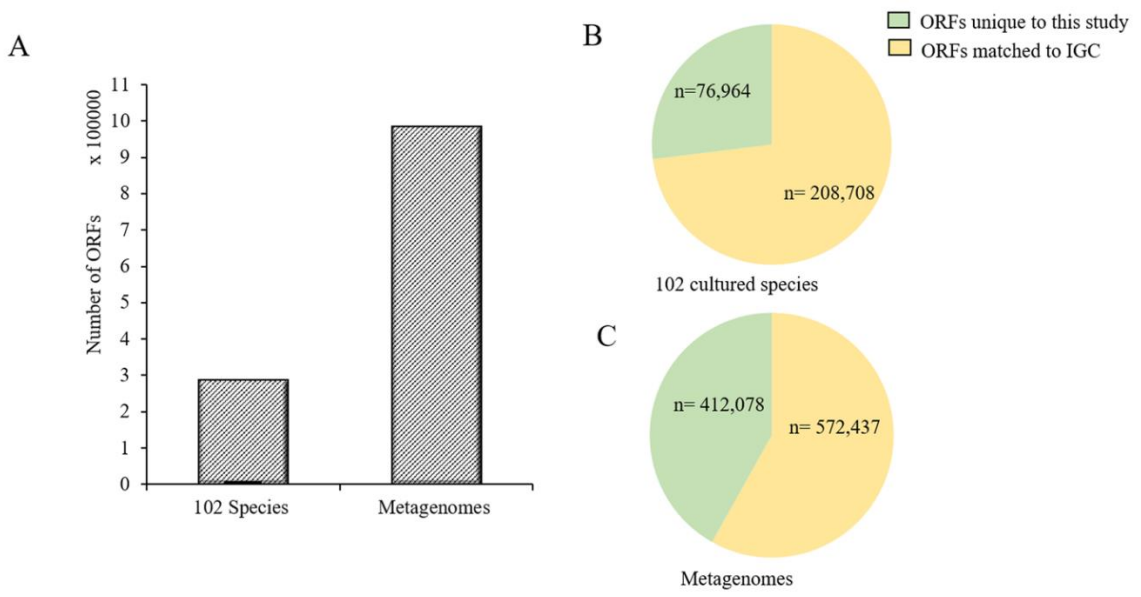


Figure 3: Efficiency of species retrieval in the culture conditions tested. A. Frequency of isolation of 102 species recovered. **B.** Prevalence of individual species as found from read mapping against donor fecal metagenome.

Single medium based culturomics retrieves most frequent gut bacteria in human populations and increases gene repertoire in the integrated gene catalog of the gut microbiome: The availability of high-quality metagenome data from many human samples facilitates the identification of frequently present gut bacterial species in healthy populations. A recent study defined 71 bacteria present across 2144 human fecal metagenomes (42). To quantify the number of these present in our culture library, we matched the whole genomes of 102 bacterial species and 33 metagenome bins against these 71 frequent species (42), finding our culture library contained 65 of these most frequent bacteria (Supplementary Table 4). To determine the relationship between the gene repertoire in our culture library and the integrated gene catalog (IGC) of the human microbiome, we compared the two datasets. For the comparison against IGC, a non-redundant gut microbiome gene set of 9.879 million genes generated using 1,267 human gut metagenomes from Europe, America and China (36), we generated a non-redundant gene set from our cultured genomes and the sample metagenome used for culturing. We created 984,515 open reading frames (ORFs) from sample metagenomes and 285,672 ORFs from cultured species genomes (Supplementary Figure 1A and Supplementary Table 3) and mapped them against the IGC. Of 285,672 ORFs obtained from cultured genomes, we matched 208,708 ORFs (73.05%) to the IGC, while 572,437 ORFs (58.14%) of 984,515 ORFs obtained from the sample metagenome matched the IGC (Supplementary Figure 1B and 1C). In other words, the IGC lacked 26.94% of the ORFs from the cultured isolates and 41.85% of the ORFs from the sample metagenome. This shows the potential for expansion of the IGC if more *Prevotella* enterotype donor fecal samples, such as those used in our study, are sequenced. Our results also show that sequencing cultured species



Supplementary Figure 1: A. Numbers of open reading frames (ORFs) found in this study in genomes and metagenomes and its comparison with IGC. A. Numbers of non-redundant ORFs predicted in 102 cultured species and donor fecal metagenomes. We generated the number of non-redundant ORFs at 95% identity cutoff for donor metagenomes and 102 isolates. **B.** Comparison of the non-redundant ORFs generated form 102 cultured species with the existing integrated human gene catalog (IGC). **C.** Comparison of the non-redundant ORFs generated from donor metagenomes with the existing IGC.

genomes identifies genes otherwise missing in metagenome sequencing because of low depth or assembly issues.

A large number of species in the culture library inhibits C. difficile in vitro: A healthy microbiota suppresses pathogen growth in the gut. To identify *C. difficile*-inhibiting species in our culture library, we screened it against *C. difficile* using a co-culture assay. Slow-growing strains would be outcompeted by *C. difficile*. We therefore used 82 moderately- or fast-growing species in the co-culture assay. When tested, 66 species inhibited *C. difficile* to varying degrees (Figure 4). In this screen, *Bifidobacterium adolescentis* strain SG-742 was the most efficient inhibitor. Furthermore, all *Bifidobacteria* inhibited *C. difficile*, signifying their importance in colonization resistance to this pathogen. The *Lachnospiraceae* family were major inhibitors. In a surprising result, 16 species in our co-culture assay increased the growth of *C. difficile* (Figure 4), a finding of clinical significance, in that a high abundance of these species may confer a higher risk of CDI.

Most inhibitors are acetate or butyrate producers: Gut bacteria metabolize diverse substrates to produce short chain fatty acids (SCFAs) in the gut (43, 44). SCFAs—particularly butyrate—act as gut epithelial immune modulators, energy sources for host intestinal cells, and pathogen-inhibitors (45, 46). To determine the relationship between SCFAs produced and *C. difficile* inhibition, we estimated the SCFAs produced by all species used in the co-culture assay (Figure 5). Our results show that the strains produce mainly acetate (Figure 5A, Supplementary Table 5). Comparing all strains at phylum level

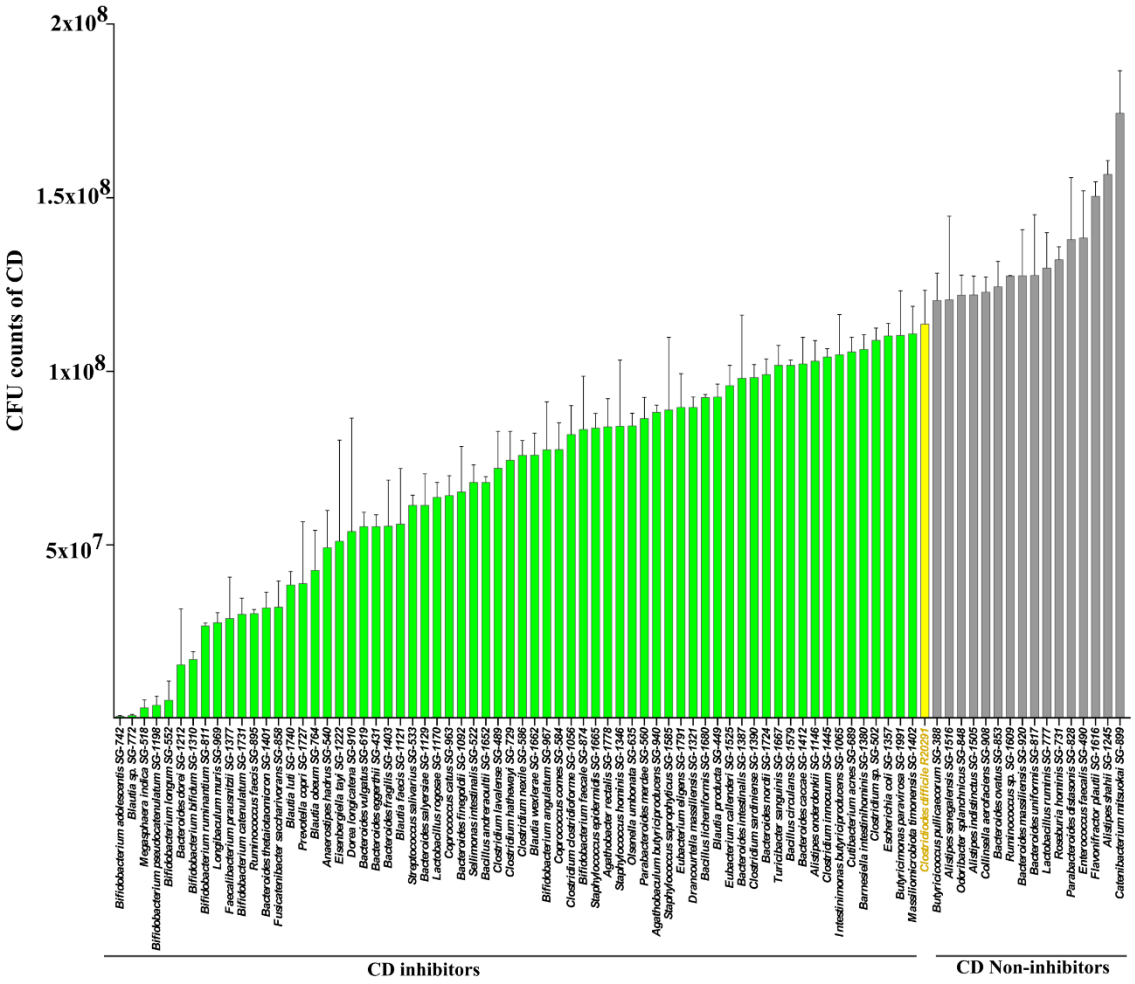


Figure 4: Inhibition of *C. difficile* by individual species *in vitro*. *C. difficile* CFU counts for every individual co-culture assay performed for 82 test species in triplicate. Error bars represent standard deviations of the three independent experiments for each inhibition assay for individual species. “Green,” “yellow” and “gray” bars represent CFU counts of *C. difficile* for inhibitors, control *C. difficile*, and non-inhibitors, respectively in co-culture.

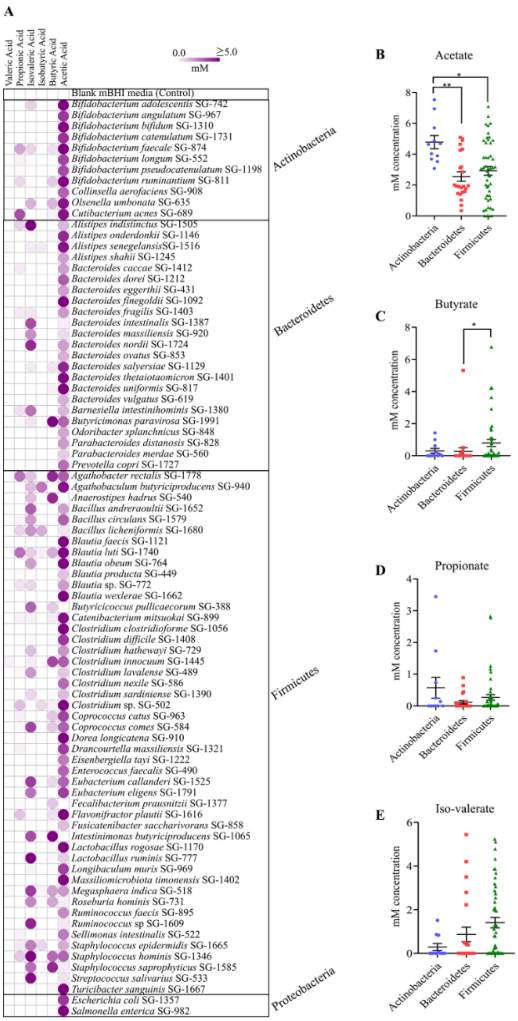


Figure 5: Short chain fatty acid (SCFA) produced by cultured species. A. SCFA production by 82 species used for inhibition assay against *C. difficile*. We measured acetate, propionate, isobutyrate, butyrate, isovalerate and valerate using gas chromatography, expressed in mM concentration. The figure represents mean SCFA measurements from duplicate samples. Color scale bar: "white": no SCFAs production; "dark": ≥5 mM SCFA production. **B.** acetate, **C.** butyrate, **D.** propionate and **E.** isovalerate production in three major phyla from 80 bacteria when oriented according to phyla (Kruskal–Wallis tests, $p<0.05$). SCFAs levels for two members of Proteobacteria phyla are not shown in B, C, D, and E.

using the non-parametric Kruskal–Wallis test, we found Actinobacteria and Bacteroidetes to yield the most acetate (Figure 5B). Firmicutes produced significantly higher butyrate than Bacteroidetes (Figure 5C). Other SCFA production did not differ significantly (Figure 5D & 5E). The majority of high acetate- or butyrate producers were *C. difficile*-inhibitors.

Relationship between nutrient utilization, C. difficile inhibition and species prevalence in patient populations: Commensal species suppress pathogens in the gut chiefly by competing for nutrients (47-49). The Biolog AN MicroPlate™ test panel provides a standardized method to identify the utilization of 95 nutrient sources by anaerobes. Investigating the relationship between nutrient utilization and pathogen inhibition, we determined the nutrient utilization of all species in our library. *C. difficile* is known to exploit mannitol, sorbitol or succinate to invade and produce infection in the human gut (50, 51). Consistent with this observation, we found 27 *C. difficile*-inhibitors that utilized all three nutrients as carbon sources (Figure 6). We hypothesized that if the *C. difficile*-inhibitors we identified were active in *C. difficile* suppression *in vivo*, their abundance would be reduced in the gut during antibiotic treatment and CDI. Hence, we determined the abundance of the top 16 (top 25%) *C. difficile*-inhibitors in our library from a study of healthy and CDI patient gut microbiomes (41). We determined the frequency of the top 16 *C. difficile*-inhibiting isolates in the following groups of human gut metagenomes; (a) CDI patients, (b) antibiotic-exposed but no CDI patients, and (C) no antibiotic exposure and no CDI (healthy) people. Consistent with our hypothesis, the top 16 *C. difficile*-inhibitors in our screen constituted about 20% of total abundance in the heathy microbiome but were in low-abundance in both antibiotic-treated and CDI patient gut samples (Figure 7A).

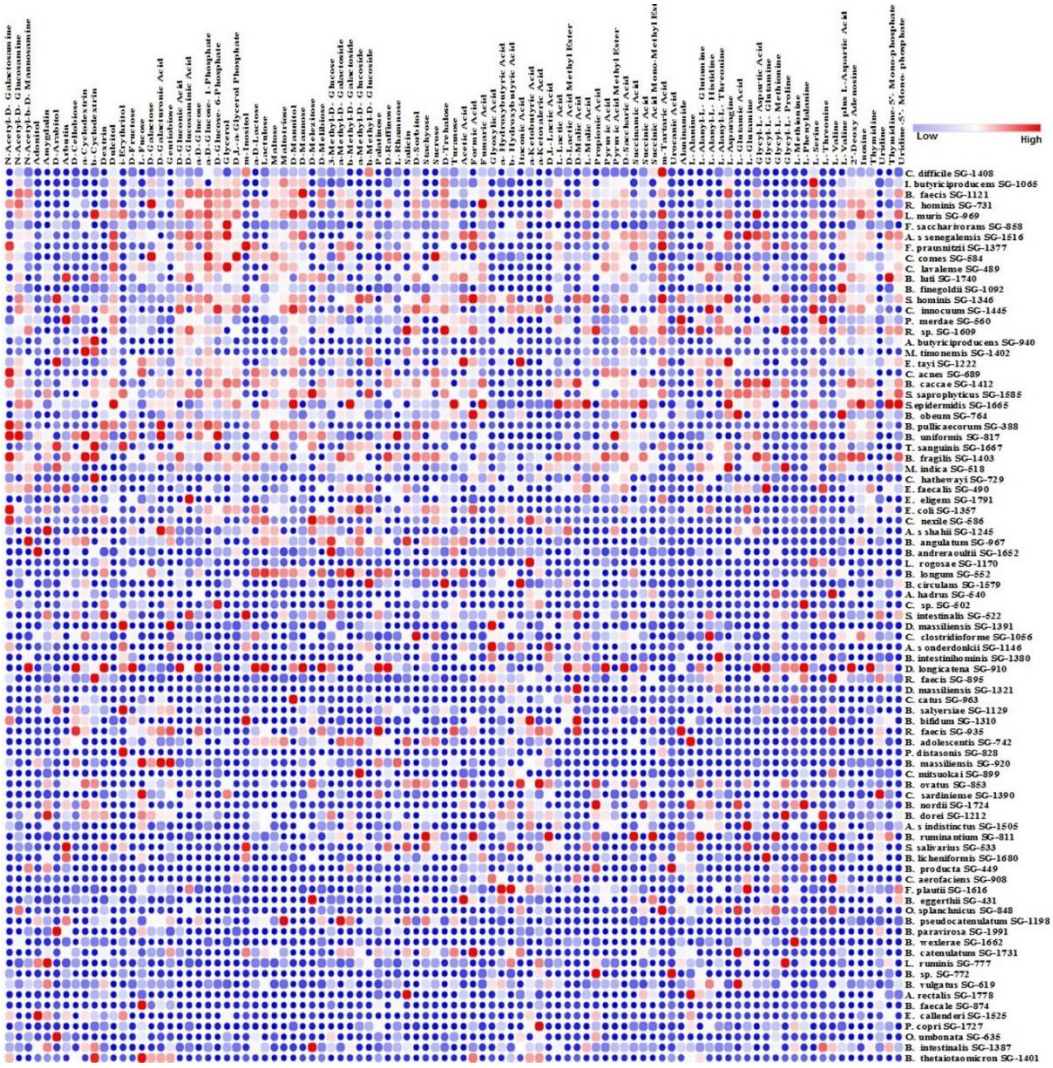


Figure 6: Utilization of 95 nutrient sources by the cultured species. The heatmap represents two independent substrate utilization tests normalized against control. Columns and rows represent nutrients and strains respectively. We considered growth of ≥ 20% in any substrate as compared to the control as positive. “Blue,” “white” and “red” represent low, medium and high utilization of the carbon source, respectively. Top row shows nutrient utilization of *C. difficile.* All other strains are arranged in descending order of nutrient utilization similarity when compared to *C. difficile*. We used nearest neighbor clustering based on the Pearson correlation to identify nutrient utilization similarity of other strains when compared to *C. difficile*.

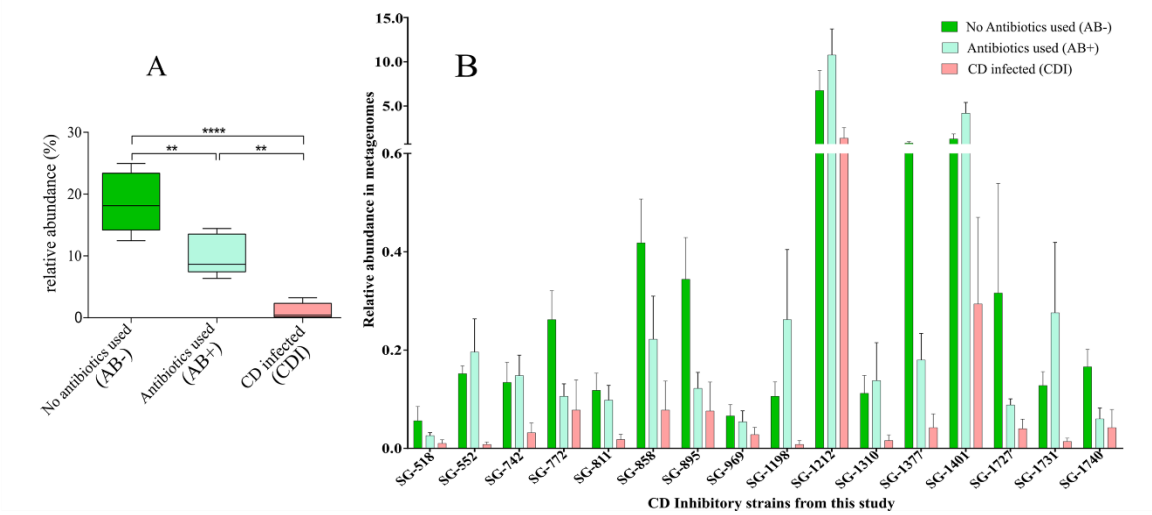
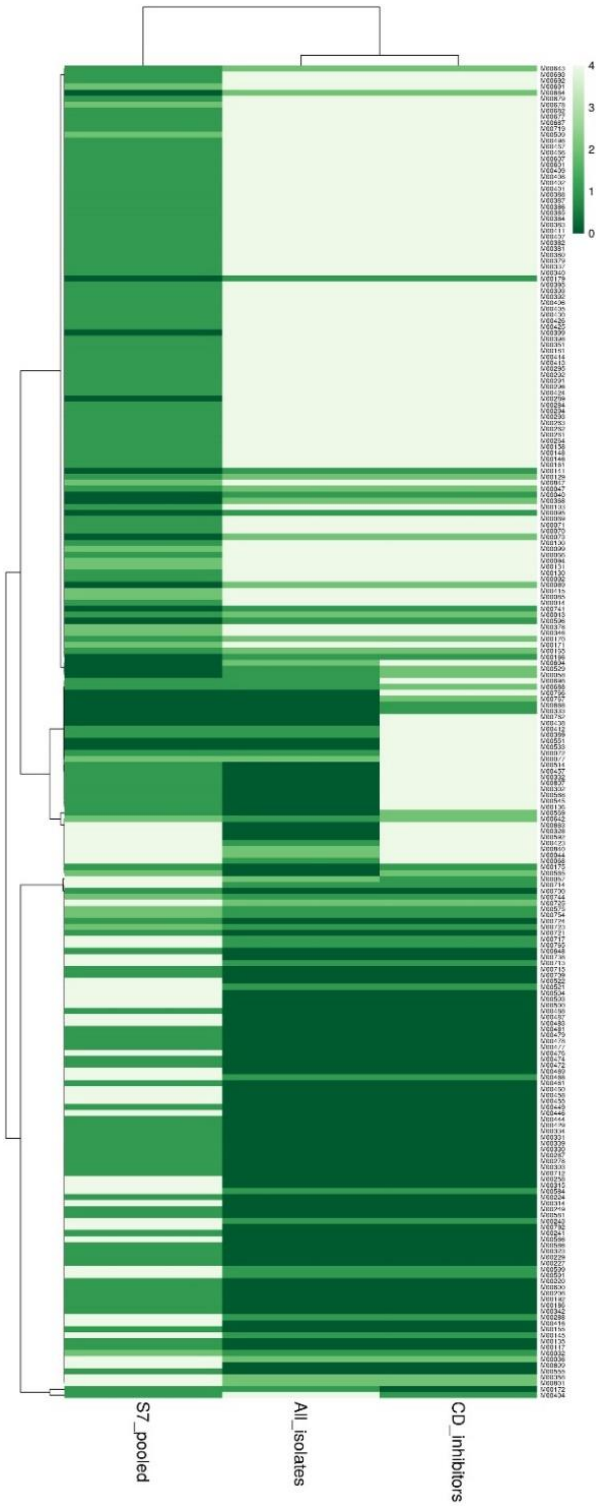


Figure 7: Frequency of the top 25% *C. difficile* inhibitors obtained in this study in the gut microbiome of *C. difficile* infected (CDI) and non-*CDI* patients. AB+ and AB- represent samples from non-*C. difficile* infected patients treated with antibiotics and non-*C. difficile* infected patients without antibiotic treatment, respectively. **A**. The combined abundance of top 25% CD-inhibitors (n = 16) from this study in CDI, AB+, and AB- metagenomes. **B**. Individual abundance of the same 16 strains in CDI, AB+, and AB- metagenomes. We obtained public metagenomes for CDI, AB+, and AB- from Milani et al., 2016.



Supplementary Figure 2: Hierarchical clustering of KEGG modules from fecal sample metagenome, all the 102 species (All_isolates), and subsets that were found to inhibit *C. difficile* R20291 (CD_inhibitors). We annotated predicted ORFs from the pooled donor fecal metagenome and consortium of cultures for KO modules by searching against the KEGG database using Ghostkoala. We indicate the completeness of the KEGG modules by the color gradient where we refer to a complete module by "0" and the absence of a whole module by "4."

Interestingly, *Prevotella copri* (SG-1727)—among the most abundant species in the donor samples—was among the species depleted during antibiotic treatment and CDI (Figure 7B).

Analysis of genotype–phenotype relationships: To identify other phenotypes not covered in our phenotype assays and to link phenotype with genotype, we used Traitar (52) to predict 67 phenotypes from the genomes of all species in our library. The substrate utilization phenotypes based on the Traitar prediction mostly matched the Biolog phenotypes (Figure 8A, Supplemental table 6). The first two clusters (green and red, lower Figure 8A) comprised mostly the pathogen-inhibitors tested here. The defining traits in these clusters related to sugar hydrolysis, mostly matching the Biolog phenotype data. The other two clusters (sky blue and mixed) comprised mostly pathogenic species and slow growers. Notable traits for pathogen clusters were catalase activity, beta hemolysis activity, growth in glycerol and high osmo-tolerance (Figure 8A). To further differentiate these traits, we performed principal component analysis (PCA) on the Traitar data. The PCA plot (Figure 8B) revealed four distinct clusters (*C. difficile*-inhibitors, -non-inhibitors, pathogens, and slow-growing strains) that explained 67.1% total variance in the first two principal components. Slow-growing strains clustered furthest from *C. difficile*-inhibitors—logically for any colonization-resistant bacteria, as a fast growth rate would more likely outcompete the pathogen.

To understand the genomic basis of colonization-resistant and pathogen-inhibiting strain genomes, we used KEGG modules—characteristic gene sets that can be linked to specific metabolic capacities or other phenotypic features of a genome. We identified 515 modules

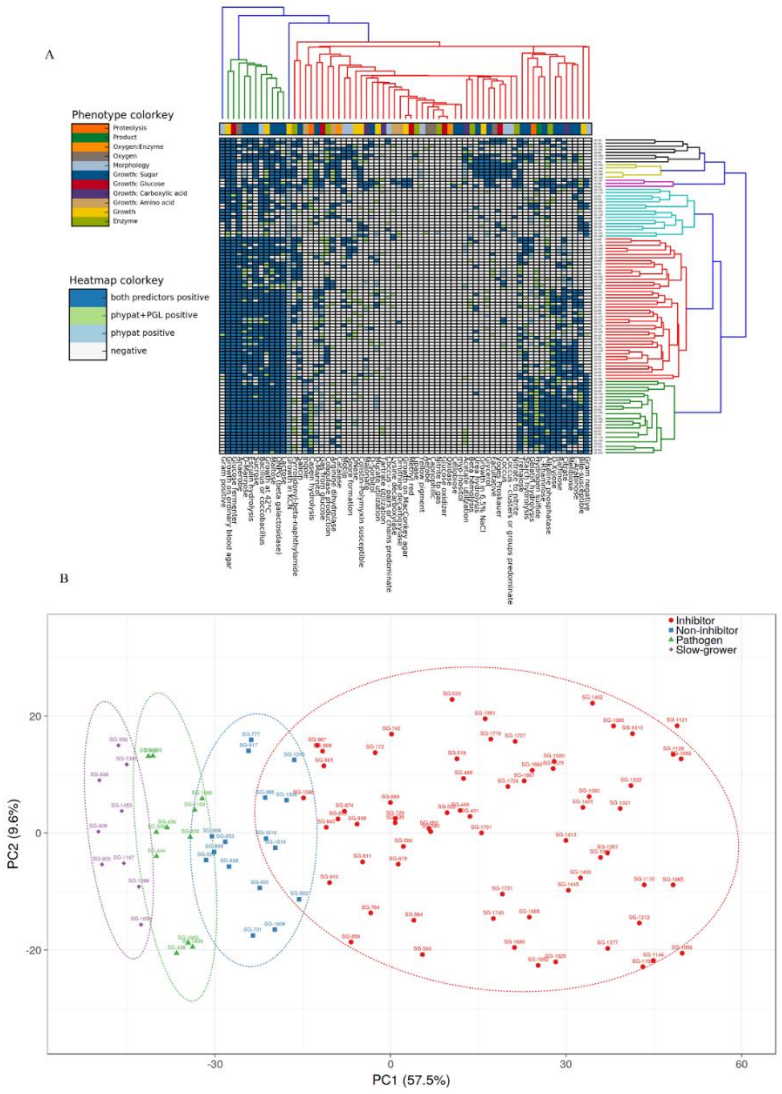


Figure 8: Prediction of phenotypes from the genomes of the 102 species in the culture library. A. Clustering of 102 species based on 67 traits predicted using the Traitar package. Each column represents one of 67 traits whereas rows represent 102 species from this study. The color scheme of the columns further depicts 11 phenotypic properties from proteolysis to enzyme production. **B.** PCA using the combined predicted traits from Pfam annotation. X- and Y-axes show Principal Component 1 & 2, explaining 57.5% and 9.6% of the total variance, respectively. Prediction ellipses are based on 0.95 confidences. Color scheme in the legends represents four different categories of isolates.

in the sample metagenome and 476 modules in the strain genomes. 432 modules were common between our sample metagenome and our strain genomes, demonstrating our ability to retrieve 77.28% of the metabolic functional capacity of the fecal microbiome using our culture method. 82 modules unrecovered in our isolated species comprised pathways related to environmental information processing, several components of cell signaling, DNA replication and repair pathways, lipid metabolism, RNA and protein processing, and the ubiquitin system (Supplementary Figure 2). We then tried to identify differences in KEGG modules associated with pathogen-inhibiting and -non-inhibiting strain genomes, finding 26 modules present in *C. difficile*-inhibitors were absent in non-*C. difficile*-inhibitors. Some important modules absent in non-inhibiting species genomes were M00698 (multidrug resistance efflux pump BpeEF-OprC), M00332 (Type III secretion system), M00438 (Nitrate/nitrite transport system), and M00551 (Benzoate degradation). Further work is necessary to understand how these functional modules relate to colonization resistance.

Design of defined mix of *C. difficile*-inhibitors: While single strain vs pathogen co-culture assay is informative in identifying pathogen inhibiting strains, the inhibition patterns are likely to change when inhibiting species interact as a community. These communities may express emergent properties difficult to predict from the individual members (52). After defining the isolate phenotypes, we used a combinatorial assembly of bacteria from our culture collection to design a tractable mix of *C. difficile*-inhibiting isolates. In the first set of experiments, we mixed 15 inhibiting species in equal proportion and tested them against *C. difficile* using the co-culture assay used for individual strains.

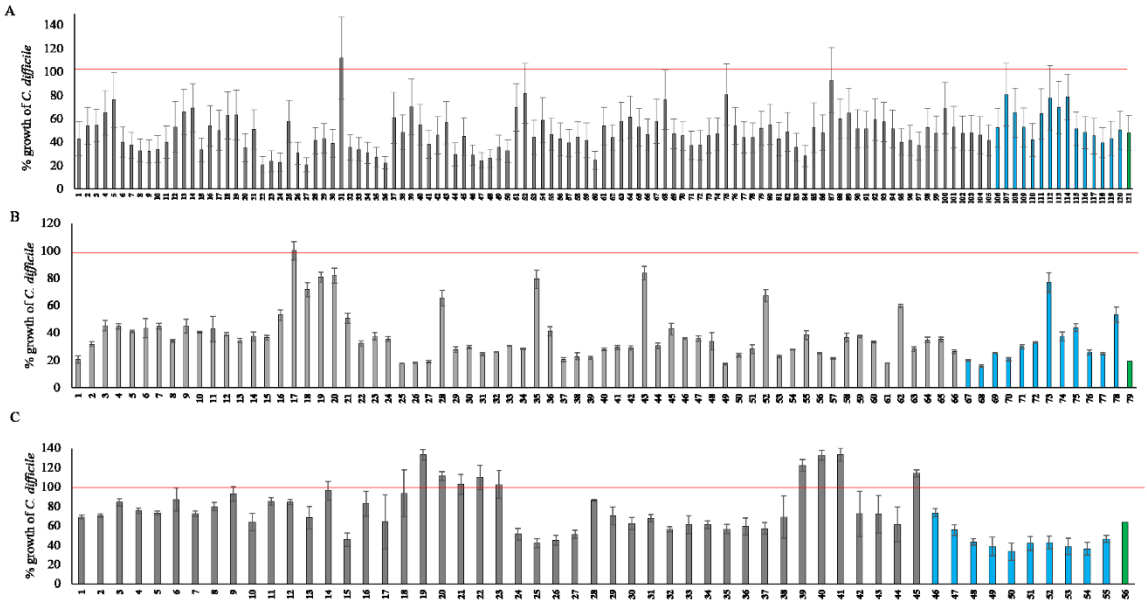


Figure 9: *C. difficile* inhibition efficiency of the consortia A. Combinations starting with 15 species pool **B.** Combinations starting with 12 species pool **C.** Combinations starting with 10 species pool. “Green,” “blue” and “gray” bars represent the parent blend, blend with one bacterial species removed at a time and blend with two bacterial species removed at a time, respectively. “Red” line represents *C. difficile* control growth. We normalized the growth of *C. difficile* in each consortium to control *C. difficile* growth, so as to obtain relative *C. difficile* growth (represented as % *C. difficile* growth) in each consortium condition. We performed each experiment in triplicate, and error bars represent the standard error of the mean.

S

Table 1 describes the overall properties of these isolates, selected based on the criteria of the medium pH not dropping below 5.6 after 24 h growth, and representation of overall taxonomic diversity at the family level. Investigating how changes in the mix composition could affect the inhibition capacity, we removed one or two species at a time from the mix of 15 species to create additional mixes, thus testing 121 mixes listed in Supplementary Table 8 against *C. difficile* in co-culture assay format. As shown in Figure 9, the removal of strains from the 15-species mix had both positive and negative effects. When we removed species, several mixes were less effective than the 15-species mix, and the removal of two species increased the inhibition efficiency in a species dependent manner. Out of 121 mixes tested, mix number 22, comprising the species listed in Supplementary Table 7, most effectively inhibited *C. difficile* growth (by 79.41%). This clearly demonstrates the dependence of inhibition efficiency on the species composition of the defined mix used in the co-culture assay. Seeking the minimum number of species necessary for an effective *C. difficile*-inhibiting mix, we performed another set of experiments in which either one or two species at a time were removed from the parent mix of 12 species. In this round, we tested 79 bacterial mixes comprising species listed in Supplementary Table 8 in the co-culture assay. As shown in Figure 9, removal of species mostly reduced inhibition efficiency when compared with the parent blend. Again, the efficiency was dependent on the species composition of the mix. We performed a third set of experiments to determine mixes comprising under ten species that would impact *C. difficile* inhibition efficiency. Removal of species from the 10-species parent set diminished inhibition efficiency overall; however, some mixes increased the growth of *C. difficile*. Since all the strains we used individually inhibited *C. difficile*, the enhancement

of *C. difficile* growth by these set mixes clearly demonstrates that individual strain phenotype can be overridden by species community interactions. In this case, a set of *C. difficile*-inhibitory species, when mixed in a particular combination, increased *C. difficile* growth. Overall, our results demonstrate that new phenotypes masking the individual strain phenotype could emerge when microbial consortia are formed, and this emergent property needs to be taken into account while designing defined *C. difficile*-inhibiting bacterial mixes.

Discussion

We developed a gut commensal culture collection from healthy human donors and identified *C. difficile*-colonization-resistant strains. As human gut microbiome composition varies across populations, donor selection for culture is an important consideration. Depending on the proportion of *Bacteriodes* and *Prevotella,* the human gut microbiome has been classified into enterotypes (53). The Asian and African populations—two-thirds of the human population fall into the *Prevotella* enterotype. (54). Information about the colonization-conferring species in *Prevotella*-dominant gut microbiomes is limited; we therefore chose recent Asian immigrants in the US as the fecal donors in this study. For culture, we used a pooled sample from six donors before culturing, which could have positive and negative consequences. Pooling can save substantial time and resources. For instance, after pooling we analyzed 1590 colonies; were this to be done individually, the number of analyzed colonies would have been 9540. Pooling, however, may distort the microbiome composition of individual samples, creating artificial population assemblages (55). For gut microbiome ecology studies, a gut commensal culture collection isolated from

fecal donors of similar gut microbiome composition could be more useful than that from many different people. The publicly available gut microbiota culture collection from the human microbiome project is isolated from 265 people (56). Other similar culture collections are also isolated from at least over 100 donors (15, 17). Although such collections are useful as reference strains, since the donors may have different microbiome compositions, the strains isolated may not form stable ecologies if mixed together. In contrast, our collection, from a limited number of donors having similar microbiome composition, may form a stable ecology when mixed, and could be more useful for studies to understand underlying interactions determining colonization resistance and other traits.

Two general approaches have been used previously for developing gut microbiota culture collections: the use of culturing samples in several—often up to 64—different nutrient media conditions, to isolate diversity (15, 57, 58); or the use of a single medium, needing less time and resource but retrieving fewer species (59). In understanding the types of bacterial communities responsible for suppressing pathogen growth in the gut, the assemblage of simple to complex bacterial communities from cultured strain libraries and testing of such consortia against pathogens is commonly performed. Although the use of different nutrient media is highly efficient in isolating the maximum number of species from gut samples, strains so isolated may not grow in a common nutrient medium, diminishing the utility of the strains in community assembly studies. We therefore used a single medium-based approach using mBHI for culturing of the fecal bacteria. As shown in Figures 2 and 3, we isolated 102 species—representing about 34.57% species diversity determined by sequencing the donor fecal sample—by adjusting the mBHI medium

(Supplementary Figure 1)—comparable to other single medium-based approaches (59-61). Furthermore, results in Supplementary Table 3 and Supplementary Figure 2 shows that the 34.57% diversity isolated represents over 70% functional capacity of the donor gut microbiome. According to the insurance hypothesis of microbiota function, more than one species performing the same function is recruited in an ecosystem to allow for functional redundancy (62, 63), possibly explaining the recovery of 70% function from 34.57% species diversity in our library.

Many strains in our culture collection shown in Figures 4 inhibit *C. difficile* at varying levels. Several phenotypes—particularly, growth rate, production of SCFAs, and the utilization of mannitol, sorbitol or succinate—correlated with the *C. difficile*-inhibitor phenotype, consistent with previous reports that restoration of depleted SCFAs in the gut resolved CDI (64, 65) and competition between *C. difficile* and commensals for nutrients and increased availability of mannitol, sorbitol or succinate allowed *C. difficile* invasion of the gut (50, 51). Top inhibiting species in our collection were also depleted in the CDI patient gut, indicating their role in providing colonization resistance against *C. difficile* (Figure 7). The formation of many different defined bacterial mixes using these inhibitory strains may improve the inhibition capacity of the individual strains. Overall, we tested 256 defined mixes using the combinatorial community assembly approach. The combinatorial community assembly method presented in Figure 9 shows two important parameters defining the efficacy of the defined mixes—the number and type of species in the mix. Reducing the number of species in the mix from 15 to 12, did not diminish the overall inhibition capacity, but reducing the number to ten species did so. Furthermore, many of

those mixes increased, rather than inhibiting, *C. difficile* growth. Clearly, adding too many species in a mix does not improve inhibition. The threshold of peak efficacy is 12 species in the conditions we tested. Our results also underline how, when species are pooled in a sub-optimal ecology, undesirable traits could emerge; strains in combination could produce new phenotypes not observed individually. Previous work to identify a defined mix of *C. difficile*-inhibiting bacteria also identified varying mix numbers. For instance, more than 20 years ago, Tvede and Rask-Madsen showed that infusion of a 10-bacterial mix into a patient's colon could resolve CDI (12). Another study in a small patient population found treatment with a 33-bacterial mix could alleviate CDI. In a mouse model, *Clostridium scindens* was a more efficient *C. difficile*-inhibitor when mixed with a defined pool of other commensal bacteria (66). Likewise, a mix of six phylogenetically diverse bacteria alleviated CDI in a mouse model (67). The complexity of the gut microbiota and its variations across populations make the design of defined bacterial *C. difficile*-inhibiting mix not simply a matter of mixing large numbers of diverse species. Since our results show that the *C. difficile*-inhibiting phenotype changes substantially depending on the microbial interaction, design of a defined bacterial mix requires a deeper understanding of how inhibiting species interact with themselves and the members of the total commensal community.

Conclusion

Overall, we demonstrated that a high percentage of the cultivable fraction of gut microbiota from healthy human donors are *C. difficile*-inhibitors *in vitro*. Defined bacterial mixes can enhance the inhibition capacity of individual strains. However, depending on the ecology

of the mix, new phenotypes could emerge. For instance, a mix of bacteria, rather than inhibiting, can increase the growth of *C. difficile*. In designing defined *C. difficile*-inhibiting bacterial mixes *in vivo*, the interaction of bacteria with each other in a mix and with other members of gut commensals demands investigation. The approach of combinatorial testing of strains with well-defined phenotypes used in the present study is a step in that direction.

Tables

For Supplementary Text 1 and Supplementary Table 1-7 please refer:

Ghimire et al., 2020. Identification of *Clostridioides difficile*-inhibiting gut commensals using culturomics, phenotyping, and combinatorial community assembly, msystems, 5(1), DOI: 10.1128/mSystems.00620-19

Literature Cited

1. van der Waaij D, Berghuis-de Vries JM, Lekkerkerk L-v. 1971. Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. J Hyg (Lond) 69:405-11.
2. Rolfe RD. 1984. Role of volatile fatty acids in colonization resistance to Clostridium difficile. Infection and immunity 45:185-91.
3. Adamu BO, Lawley TD. 2013. Bacteriotherapy for the treatment of intestinal dysbiosis caused by Clostridium difficile infection. Curr Opin Microbiol 16:596-601.
4. Kelly CP, Pothoulakis C, LaMont JT. 1994. Clostridium difficile colitis. The New England journal of medicine 330:257-62.
5. Petrof EO, Khoruts A. 2014. From stool transplants to next-generation microbiota therapeutics. Gastroenterology 146:1573-1582.
6. Khoruts A, Dicksved J, Jansson JK, Sadowsky MJ. 2010. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. J Clin Gastroenterol 44:354-60.
7. Khoruts A, Sadowsky MJ, Hamilton MJ. 2015. Development of fecal microbiota transplantation suitable for mainstream medicine. Clin Gastroenterol Hepatol 13:246-50.
8. Bakken JS, Borody T, Brandt LJ, Brill JV, Demarco DC, Franzos MA, Kelly C, Khoruts A, Louie T, Martinelli LP, Moore TA, Russell G, Surawicz C, Fecal Microbiota Transplantation W. 2011. Treating Clostridium difficile infection with fecal microbiota transplantation. Clin Gastroenterol Hepatol 9:1044-9.

9. Alang N, Kelly CR. 2015. Weight gain after fecal microbiota transplantation. Open Forum Infect Dis 2:ofv004.

10. FDA. 2019. FDA In Brief: FDA warns about potential risk of serious infections caused by multi-drug resistant organisms related to the investigational use of Fecal Microbiota for Transplantation. https://www.fda.gov/news-events/fda-brief/fda-brief-fda-warns-about-potential-risk-serious-infections-caused-multi-drug-resistant-organisms. Accessed

11. Collins J, Auchtung JM. 2017. Control of Clostridium difficile Infection by Defined Microbial Communities. Microbiol Spectr 5.

12. Tvede M, Rask-Madsen J. 1989. Bacteriotherapy for chronic relapsing Clostridium difficile diarrhoea in six patients. Lancet 1:1156-60.

13. Petrof EO, Gloor GB, Vanner SJ, Weese SJ, Carter D, Daigneault MC, Brown EM, Schroeter K, Allen-Vercoe E. 2013. Stool substitute transplant therapy for the eradication of Clostridium difficile infection: 'RePOOPulating' the gut. Microbiome 1:3.

14. Khanna S, Pardi DS, Kelly CR, Kraft CS, Dhere T, Henn MR, Lombardo MJ, Vulic M, Ohsumi T, Winkler J, Pindar C, McGovern BH, Pomerantz RJ, Aunins JG, Cook DN, Hohmann EL. 2016. A Novel Microbiome Therapeutic Increases Gut Microbial Diversity and Prevents Recurrent Clostridium difficile Infection. J Infect Dis 214:173-81.

15. Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, Caputo A, Cadoret F, Traore SI, Seck EH, Dubourg G, Durand G, Mourembou G, Guilhot E, Togo A, Bellali S, Bachar D, Cassir N, Bittar F, Delerce J, Mailhe M, Ricaboni D, Bilen M,

Dangui Nieko NP, Dia Badiane NM, Valles C, Mouelhi D, Diop K, Million M, Musso D, Abrahao J, Azhar EI, Bibi F, Yasir M, Diallo A, Sokhna C, Djossou F, Vitton V, Robert C, Rolain JM, La Scola B, Fournier PE, Levasseur A, Raoult D. 2016. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol 1:16203.

16. Forster SC, Kumar N, Anonye BO, Almeida A, Viciani E, Stares MD, Dunn M, Mkandawire TT, Zhu A, Shao Y, Pike LJ, Louie T, Browne HP, Mitchell AL, Neville BA, Finn RD, Lawley TD. 2019. A human gut bacterial genome and culture collection for improved metagenomic analyses. Nat Biotechnol 37:186-192.
17. Zou Y, Xue W, Luo G, Deng Z, Qin P, Guo R, Sun H, Xia Y, Liang S, Dai Y, Wan D, Jiang R, Su L, Feng Q, Jie Z, Guo T, Xia Z, Liu C, Yu J, Lin Y, Tang S, Huo G, Xu X, Hou Y, Liu X, Wang J, Yang H, Kristiansen K, Li J, Jia H, Xiao L. 2019. 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. Nat Biotechnol 37:179-185.
18. Wajid B, Serpedin E. 2016. Do it yourself guide to genome assembly. Brief Funct Genomics 15:1-9.
19. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-7.
20. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-9.
21. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863-4.

22. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

23. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-5.

24. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350-2.

25. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.

26. Thomas M, Webb M, Ghimire S, Blair A, Olson K, Fenske GJ, Fonder AT, Christopher-Hennings J, Brake D, Scaria J. 2017. Metagenomic characterization of the effect of feed additives on the gut microbiome and antibiotic resistome of feedlot cattle. Sci Rep 7:12257.

27. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-9.

28. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7:11257.

29. Mende DR, Letunic I, Huerta-Cepas J, Li SS, Forslund K, Sunagawa S, Bork P. 2017. proGenomes: a resource for consistent functional and taxonomic annotations of prokaryotic genomes. Nucleic Acids Res 45:D529-D534.

30. Mende DR, Sunagawa S, Zeller G, Bork P. 2013. Accurate and universal delineation of prokaryotic species. Nat Methods 10:881-4.

31. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824-834.

32. Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. Bioinformatics 32:1088-90.

33. Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 38:e132.

34. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150-2.

35. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-9.

36. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T, Juncker AS, Manichanh C, Chen B, Zhang W, Levenez F, Wang J, Xu X, Xiao L, Liang S, Zhang D, Zhang Z, Chen W, Zhao H, Al-Aama JY, Edris S, Yang H, Wang J, Hansen T, Nielsen HB, Brunak S, Kristiansen K, Guarner F, Pedersen O, Dore J, Ehrlich SD, Meta HITC, Bork P, Wang J, Meta HITC. 2014. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 32:834-41.

37. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol 34:2115-2122.

38. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

39. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726-731.

40. Weimann A, Mooren K, Frank J, Pope PB, Bremges A, McHardy AC. 2016. From Genomes to Phenotypes: Traitar, the Microbial Trait Analyzer. mSystems 1.

41. Milani C, Ticinesi A, Gerritsen J, Nouvenne A, Lugli GA, Mancabelli L, Turroni F, Duranti S, Mangifesta M, Viappiani A, Ferrario C, Maggio M, Lauretani F, De Vos W, van Sinderen D, Meschi T, Ventura M. 2016. Gut microbiota composition and Clostridium difficile infection in hospitalized elderly individuals: a metagenomic study. Sci Rep 6:25945.

42. Costea PI, Coelho LP, Sunagawa S, Munch R, Huerta-Cepas J, Forslund K, Hildebrand F, Kushugulova A, Zeller G, Bork P. 2017. Subspecies in the global human gut microbiome. Mol Syst Biol 13:960.

43. Flint HJ, Duncan SH, Scott KP, Louis P. 2015. Links between diet, gut microbiota composition and gut metabolism. Proc Nutr Soc 74:13-22.

44. Morrison DJ, Preston T. 2016. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes 7:189-200.

45. Cherrington CA, Hinton M, Pearson GR, Chopra I. 1991. Short-chain organic acids at ph 5.0 kill Escherichia coli and Salmonella spp. without causing membrane perturbation. J Appl Bacteriol 70:161-5.

46. Roe AJ, McLaggan D, Davidson I, O'Byrne C, Booth IR. 1998. Perturbation of anion balance during inhibition of growth of Escherichia coli by weak acids. J Bacteriol 180:767-72.

47. Fabich AJ, Jones SA, Chowdhury FZ, Cernosek A, Anderson A, Smalley D, McHargue JW, Hightower GA, Smith JT, Autieri SM, Leatham MP, Lins JJ, Allen RL, Laux DC, Cohen PS, Conway T. 2008. Comparison of carbon nutrition for pathogenic and commensal Escherichia coli strains in the mouse intestine. Infect Immun 76:1143-52.

48. Maltby R, Leatham-Jensen MP, Gibson T, Cohen PS, Conway T. 2013. Nutritional basis for colonization resistance by human commensal Escherichia coli strains HS and Nissle 1917 against E. coli O157:H7 in the mouse intestine. PLoS One 8:e53957.

49. Deriu E, Liu JZ, Pezeshki M, Edwards RA, Ochoa RJ, Contreras H, Libby SJ, Fang FC, Raffatellu M. 2013. Probiotic bacteria reduce salmonella typhimurium intestinal colonization by competing for iron. Cell Host Microbe 14:26-37.

50. Spiga L, Winter MG, Furtado de Carvalho T, Zhu W, Hughes ER, Gillis CC, Behrendt CL, Kim J, Chessa D, Andrews-Polymenis HL, Beiting DP, Santos RL, Hooper LV, Winter SE. 2017. An Oxidative Central Metabolism Enables Salmonella to Utilize Microbiota-Derived Succinate. Cell Host Microbe 22:291-301 e6.

51. Theriot CM, Koenigsknecht MJ, Carlson PE, Jr., Hatton GE, Nelson AM, Li B, Huffnagle GB, J ZL, Young VB. 2014. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. Nat Commun 5:3114.

52. Newman DK, Banfield JF. 2002. Geomicrobiology: how molecular-scale interactions underpin biogeochemical systems. Science 296:1071-7.

53. Costea PI, Hildebrand F, Arumugam M, Backhed F, Blaser MJ, Bushman FD, de Vos WM, Ehrlich SD, Fraser CM, Hattori M, Huttenhower C, Jeffery IB, Knights D, Lewis JD, Ley RE, Ochman H, O'Toole PW, Quince C, Relman DA, Shanahan F, Sunagawa S, Wang J, Weinstock GM, Wu GD, Zeller G, Zhao L, Raes J, Knight R, Bork P. 2018. Enterotypes in the landscape of gut microbial community composition. Nat Microbiol 3:8-16.

54. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Dore J, Consortium M, Weissenbach J, Ehrlich SD, Bork P, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, et al. 2011. Enterotypes of the human gut microbiome. Nature doi:10.1038/nature09944.

55. Auchtung JM, Robinson CD, Britton RA. 2015. Cultivation of stable, reproducible microbial communities from different fecal donors using minibioreactor arrays (MBRAs). Microbiome 3:42.

56. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, Brady A, Creasy HH, McCracken C, Giglio MG, McDonald D, Franzosa EA, Knight R, White O, Huttenhower C. 2017. Strains, functions and dynamics in the expanded Human Microbiome Project. Nature 550:61-66.

57. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, Trape JF, Koonin EV, La Scola B, Raoult D. 2012. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 18:1185-93.

58. Lagier JC, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, Levasseur A, Rolain JM, Fournier PE, Raoult D. 2018. Culturing the human microbiota and culturomics. Nat Rev Microbiol doi:10.1038/s41579-018-0041-0:540-550.

59. Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G, Gordon JI. 2011. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proc Natl Acad Sci U S A 108:6252-7.

60. Rettedal EA, Gumpert H, Sommer MO. 2014. Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. Nat Commun 5:4714.

61. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD. 2016. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature 533:543-546.

62. Yachi S, Loreau M. 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. Proc Natl Acad Sci U S A 96:1463-8.

63. Relman DA. 2012. The human microbiome: ecosystem resilience and health. Nutr Rev 70 Suppl 1:S2-9.

64. McDonald JAK, Mullish BH, Pechlivanis A, Liu Z, Brignardello J, Kao D, Holmes E, Li JV, Clarke TB, Thursz MR, Marchesi JR. 2018. Inhibiting Growth of

Clostridioides difficile by Restoring Valerate, Produced by the Intestinal Microbiota. Gastroenterology 155:1495-1507 e15.

65. Seekatz AM, Theriot CM, Rao K, Chang YM, Freeman AE, Kao JY, Young VB. 2018. Restoration of short chain fatty acid and bile acid metabolism following fecal microbiota transplantation in patients with recurrent Clostridium difficile infection. Anaerobe 53:64-73.

66. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, Littmann E, van den Brink MR, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint NC, Xavier JB, Pamer EG. 2015. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature 517:205-8.

67. Lawley TD, Clare S, Walker AW, Stares MD, Connor TR, Raisen C, Goulding D, Rad R, Schreiber F, Brandt C, Deakin LJ, Pickard DJ, Duncan SH, Flint HJ, Clark TG, Parkhill J, Dougan G. 2012. Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice. PLoS Pathog 8:e1002995.

CHAPTER:2 GENOME SEQUENCE AND DESCRIPTION OF *BLAUTIA BROOKINGSII* SG772 SP. NOV., A NOVEL BACTERIAL SPECIES ISOLATED FROM THE HUMAN FECES

Sudeep Ghimire, Supapit Wongkuna, Roshan Kumar, Eric A Nelson, Jane Christopher-Hennings, Joy Scaria[1]*

*Corresponding Author

Email: joy.scaria@sdstate.edu

[1]Department of Veterinary & Biomedical Sciences, South Dakota State University, Brookings, SD 57007

Abstract

An anaerobic isolate SG772 belonging to the genus *Blautia* was isolated from healthy human fecal sample. When compared using 16s rRNA sequence identity, SG772 showed only 94.46% similarity with its neighbor species *B. stercoris.* Since strain SG772 showed both phenotypic and genomic differences with other members of the type species within the genus *Blautia*, we propose the designation of SG772 as novel species "*Blautia brookingsii* SG772^{T}".

Keywords: *Blautia brookingsii* SG772^{T}, culturomics, gut microbiota, taxogenomics

Introduction

The human gut microbiome of European and North American population is typically dominated by *Bacteroides* while *Prevotella* dominates Asian and African populations(1). To determine the cultivable microbial diversity of the *Prevotella* dominant human gut microbiota, fecal samples were cultured from six healthy donors using a high-throughput culturomics approach(2). A new species of a bacterium belonging to the genus *Blautia* was isolated during this study. We characterized this strain using the recently proposed taxono-genomics strategy which utilizes the combination of phenotypic and genomic characterization of the strain to define new species. Herein, we describe strain *Blautia brookingsii* SG772^{T} (DSM 107275=CCOS 1888) that was isolated from the healthy human gut microbiota.

Materials and Methods, Results and Discussion

Isolation and growth conditions

The strain was isolated on a Brain Heart Infusion (BHI) agar medium after 48 hours of incubation at 37^{o}C in strict anaerobic condition (85% nitrogen, 10% carbon-dioxide, 5% hydrogen). The strain grew in the pH range of 5.5-7.5 with optimum growth at 6.8. No growth of was observed after 20 min of thermal shock at 80^{o}C suggesting its non-spore forming nature.

Phenotypic characteristics

The colonies grown in agar plates appeared to be round, whitish, convex and smooth after 48 hours of anaerobic incubation on BHI plate. Cells were gram-positive and coccobacillus in shape with no flagellum suggesting non-motile nature. Overnight culture of the strain $SG772^T$ in the broth was centrifuged at 10,000×g. The supernatant was then removed to obtain the pellet which was mixed with 30% glycerol solution and frozen to -80^0C and sent to BGSU Center for Microscopy (Bowling Green, OH) for scanning electron microscopy. The size of the bacterium under scanning electron microscopy was between 0.5-0.8 x 1.8-2.5 µm (Figure. 1).

Using API ZYM strip, enzymatic activities were determined. Positive enzymatic reactions were observed for esterase (C4), esterase lipase (C8), leucine arylamidase, acid phosphatase, α-galactosidase, β-galactosidase, α-glucosidase and N-acetyl-β-glucosaminidase. Carbon source utilization of the strain $SG772^T$ was determined using BIOLOG AN plate. The strain SG772T was grown as lawn in BHI plate for 48 hours and the cells were scooped using sterile cotton swab to mix in BIOLOG AN fluid until the OD_{600} reached 0.01. 100 µl of the mixture was pipetted into 96 wells of the BIOLOG AN plate and incubated for 24 hours. The initial and final OD were measured and compared to control. Those wells with OD growth >20% compared to control were considered as growth. Thus, this test revealed that the strain utilized 21 substrates with maximum preference for propionic acid followed by L-alanyl-L-glutamine (Table 1). Furthermore, it was able to utilize complex carbohydrates like palatinose, rhamnose, arbutin along with

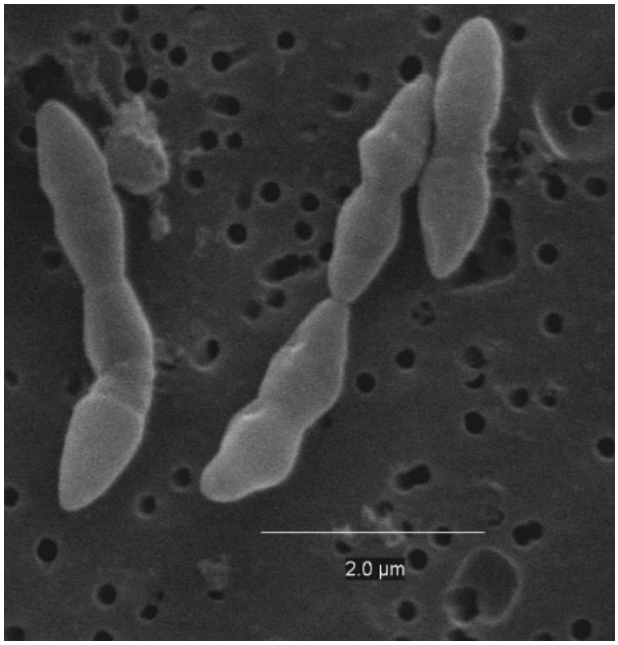


Figure 1: Scanning electron micrograph of *Blautia brookingsii* SG772^{T} sp. nov.

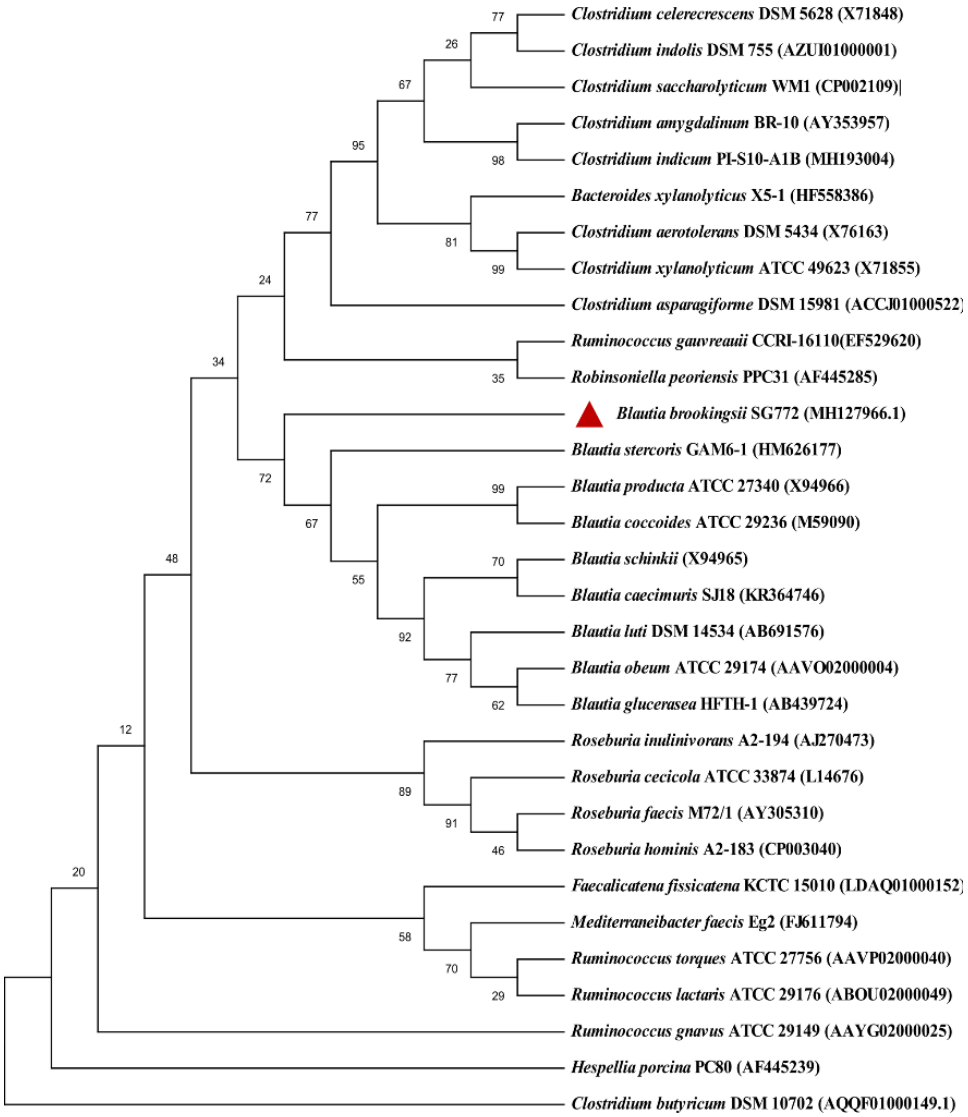


Figure 2: Phylogenetic tree highlighting the position of strain SG772^{T} with regard to other closely related species. The consensus phylogenetic tree showing the was constructed using 16S rRNA gene sequences from strain SG772^{T} and its top 30 neighbors obtained from EzTaxon. The evolutionary history was inferred using Neighbor-Joining method using Kimura 2-paramter method. The numbers at the nodes indicate bootstrap values expressed as percentages of 1000 replications in MEGA X. *Clostridium butyricum* DSM10702 was taken as an outgroup. GenBank accession numbers are shown in parentheses.

organic acids such as α-ketovaleric acid, urocanic acid, glycoxylic acid, fumaric acid, galacturonic acid and α-hydroxybutyric acid. Comparison of the substrate utilization of strain SG772^T with other neighbor *Blautia* strains are given in Table 1. Additionally, cellular morphology and optimum pH for growth varied among the different species (Table 1). Antibiotic sensitivity test using disk diffusion technique displayed that the strain SG772^T was resistant to tetracycline and streptomycin. However, it was susceptible to chloramphenicol, ampicillin, erythromycin, and novobiocin.

Strain Identification

After initial isolation of the strain SG772^T, MALDI-TOF was performed resulting no identification (score < 1.7). Therefore, DNA of the strain was isolated and 16S rRNA gene was amplified and sequenced using Applied Biosystems 3500xL, Genetic analyser (Applied Biosystems, MA, USA). The 16S rRNA sequences were trimmed and assembled using Genious 10.2.3 (NJ, US). A continuous stretch of 1440 bp of the 16S rRNA gene of strain SG772^T was obtained and searched against EzTaxon- e-server (3) for identification. This comparison showed that the closest neighbor for strain SG772^T is *B. stercoris* GAM6-1 sharing only 94.46% sequence similarity (Figure 2).

Genome sequencing and comparison

The genomic DNA was isolated using E.Z.N.A DNA isolation kit (Omega, Biotek) following manufacture's protocol. The sequencing was performed using llumina MiSeq

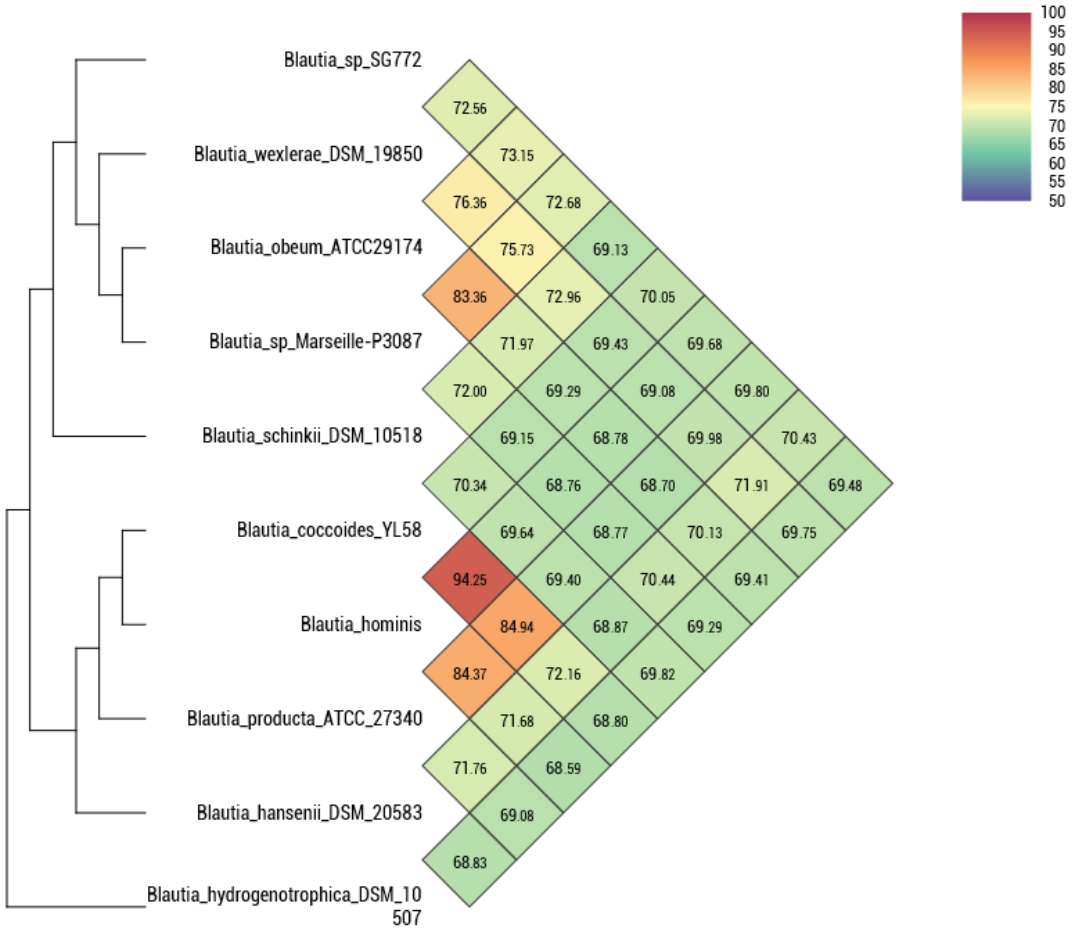


Figure 3: Average Nucleotide Identity (*ANI*) comparison of *Blautia brookingsii* strain SG772^{T} sp. nov., with other closely related species with standing nomenclature. Heatmap was generated with OrthoANI values calculated using the OAT software.

(Illumina Inc, CA), 2×250 paired end chemistry. The reads were assembled using SPAdes 3.9.0 (4) and the contigs were validated using Quast (5).

Assembly of the strain SG772^{T} genome produced 52 contigs with the genome size of 3,411,519 bps (N50 = 156,020) and 44.07% GC content. The largest contig was 401,555 bps while the smallest was of 610 bps. The genome contains 57 tRNA genes, 1 rRNA gene and 3061 coding sequences. Additionally, for SG772^{T} genome, the open reading frames (ORF) were predicted using Prodigal 2.6 (6) in Prokka software package (7). Transfer RNA genes were predicted using Aragorn 1.2 (8) and rRNA genes were annotated using Barrnap 0.7 (https://github.com/tseemann/barrnap). The comparison of SG772^{T} genome with its close species revealed several differences in genome size, G+C content and RNA copies. The closely related species of the strain SG772^{T} for which the genomes were available NCBI were obtained and the degree of similarity was estimated using OrthoANI software (9). For the closest neighbors (Fig 2), OrthoANI values ranged from 69.13 to 73.15%. The type strain SG772^{T} was observed to be close to *Blautia wexlerae* DSM19850 with 73.15% genomic identity while *B. schinkii* DSM10518 was the farthest (Figure. 3).

Conclusion

As the sequence identity of the phylogenetically closest validated species is < 98.7 %: the threshold recommended to define a species as per nomenclature (10), we propose the strain SG772^{T} as a new species *Blautia brookingsii* sp.nov. strain SG772^{T} (*broo.king.sii*, L.,

neut., adj., *brookingsii*, based on the acronym of the Brookings city where the type strain was first isolated). The 16S rRNA gene sequence of SG772^{T} is deposited in GenBank under accession number MH127966.1. Whole genome sequences are deposited in GenBank under BioProject PRJNA436995. Strain SG772^{T} was deposited in the DSMZ-German Collection of Microorganisms and Cell Cultures GmbH under DSM 107275 and Culture Collection of Switzerland (CCOS) under CCOS 1888.

Table

Table 1: Physiological and substrate utilization characteristics of strain SG772^{T} compared with its phylogenetic neighbors.

(Refer to **Ghimire et al., 2020.** Genome sequence and description of *Blautia brookingsii* SG772 sp. nov., a novel bacterial species isolated from human faeces, New Microbes New Infections, Vol 34, 100648 DOI: https://doi.org/10.1016/j.nmni.2019.100648)

Literature Cited

1. Costea PI, Hildebrand F, Arumugam M, Backhed F, Blaser MJ, Bushman FD, et al. Enterotypes in the landscape of gut microbial community composition. Nat Microbiol. 2018;3(1):8-16.
2. Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol. 2016;1:16203.
3. Chun J, Lee JH, Jung Y, Kim M, Kim S, Kim BK, et al. EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. Int J Syst Evol Microbiol. 2007;57(Pt 10):2259-61.
4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77.
5. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-5.
6. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
7. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
8. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32(1):11-6.

9. Lee I, Ouk Kim Y, Park SC, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016;66(2):1100-3.

10. Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol. 2014;64(Pt 2):384-91.

11. Park SK, Kim MS, Roh SW, Bae JW. Blautia stercoris sp. nov., isolated from human faeces. Int J Syst Evol Microbiol. 2012;62(Pt 4):776-9.

12. Ezaki T, Li N, Hashimoto Y, Miura H, Yamamoto H. 16S ribosomal DNA sequences of anaerobic cocci and proposal of Ruminococcus hansenii comb. nov. and Ruminococcus productus comb. nov. Int J Syst Bacteriol. 1994;44(1):130-6.

13. Liu C, Finegold SM, Song Y, Lawson PA. Reclassification of Clostridium coccoides, Ruminococcus hansenii, Ruminococcus hydrogenotrophicus, Ruminococcus luti, Ruminococcus productus and Ruminococcus schinkii as Blautia coccoides gen. nov., comb. nov., Blautia hansenii comb. nov., Blautia hydrogenotrophica comb. nov., Blautia luti comb. nov., Blautia producta comb. nov., Blautia schinkii comb. nov. and description of Blautia wexlerae sp. nov., isolated from human faeces. Int J Syst Evol Microbiol. 2008;58(Pt 8):1896-902.

14. Park SK, Kim MS, Bae JW. Blautia faecis sp. nov., isolated from human faeces. Int J Syst Evol Microbiol. 2013;63(Pt 2):599-603.

15. Rieu-Lesme F, Morvan B, Collins MD, Fonty G, Willems A. A new H2/CO2-using acetogenic bacterium from the rumen: description of Ruminococcus schinkii sp. nov. FEMS Microbiol Lett. 1996;140(2-3):281-6.

16. Simmering R, Taras D, Schwiertz A, Le Blay G, Gruhl B, Lawson PA, et al. Ruminococcus luti sp. nov., isolated from a human faecal sample. Syst Appl Microbiol. 2002;25(2):189-93.

17. Lawson PA, Finegold SM. Reclassification of Ruminococcus obeum as Blautia obeum comb. nov. Int J Syst Evol Microbiol. 2015;65(Pt 3):789-93.

CHAPTER 3: RICE BRAN AND QUERCETIN COMBINED PRODUCE A POSITIVE SYNERGISTIC EFFECT ON HUMAN GUT MICROBIOTA, REDUCE THE POPULATION OF *ENTEROBACTERIACEAE* FAMILY, AND ELEVATE THE LEVEL OF PROPIONATE WHEN DETERMINED USING A BIOREACTOR MODEL

[a, b]Sudeep Ghimire, [a, b]Supapit Wongkuna, [c]Ranjini Sankaranarayanan, [d]Elizabeth P. Ryan, [c]G. Jayarama Bhat, and [a, b]Joy Scaria

[a]Department of Veterinary and Biomedical Sciences, South Dakota State University, Brookings, SD, USA.

[b]South Dakota Center for Biologics Research and Commercialization, SD, USA.

[c]Department of Pharmaceutical Sciences, South Dakota State University, Brookings, SD, USA.

[d]Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO, USA

Running title: Impact of rice bran and quercetin on human gut microbiota

Abstract

Diet is one of the prominent determinants of gut microbiota composition significantly impacting human health. Recent studies with dietary supplements such as rice bran and quercetin have been shown to provide a beneficial impact on the host by positively influencing the gut microbiota. However, the specific bacterial species impacted when rice bran or quercetin is present in the diet is not well understood. Therefore, in this study, we used a minibioreactor array system as a model to determine the effect of quercetin and rice bran individually, as well as in combination, on gut microbiota without the confounding host factors. We found that rice bran exerts higher shift in gut microbiome composition when compared to quercetin. At the species level, *Acidaminococcus intestini* was the only significantly enriched taxa when quercetin was supplemented, while 15 species were enriched in rice bran supplementation and 13 were enriched when quercetin and rice bran were supplemented in combination. When comparing the short chain fatty acid production, quercetin supplementation significantly enriched isobutyrate production while propionate dominated the quercetin and rice bran combined group. Higher levels of propionate were highly correlated to the lower abundance of the potentially pathogenic *Enterobacteriaceae* family. These findings suggest that the combination of rice bran and quercetin serve to enrich beneficial bacteria and reduce potential opportunistic pathogens. However, further *in vivo* studies are necessary to determine the synergistic effect of rice bran and quercetin on host health and immunity.

Importance

Rice bran and quercetin are dietary components that shape host health by interacting with the gut microbiome. Both these substrates have been reported to provide nutritional and immunological benefits individually. However, considering the complexity of the human diet, it is useful to determine how the combination of food ingredients such as rice bran and quercetin influences the human gut microbiota. Our study provides insights into how these ingredients influence microbiome composition alone and in combination *in vitro*. This will allow us to identify which species in the gut microbiome are responsible for biotransformation of these dietary ingredients. Such information is helpful for the development of synbiotics to improve gut health and immunity.

Introduction

Gut microbiome influences health and disease (1-3) and is highly affected by diet (4-6). Some dietary components are known to enhance host health by promoting the growth of beneficial bacteria in the gut, termed as "prebiotics" (7, 8). Quercetin and rice bran have been used as prebiotics in mice and pig experiments, showing significant improvements to the gut microbiota (9-11). Rice bran, a byproduct of the rice milling process, is available, affordable, sustainable, and a globally produced source of prebiotics. It is known to supplement nutrients (12), increase beneficial bacteria growth, enhance gut mucosal immunity (10), and prevent diseases (9, 13-15). Similarly, quercetin (3,3′,4′,5,7-pentahydroxyflavone) represents an important subgroup of flavonoids found in fruits and leafy vegetables (16). Quercetin has received substantial attention in the past few years from the scientific community for exerting anti-inflammatory effects (17, 18) and its potential health-promoting properties in the treatment or prevention of cardiovascular diseases (19), lung and colorectal cancers (20, 21), and colitis (11, 22). Furthermore, the mammalian body can endure high levels of quercetin without any significant adverse health effects (23).

Despite multiple studies corroborating the beneficial effect of quercetin and rice bran on the host, their effect on microbiota varies significantly from study to study and has lower taxonomic resolution (22, 24-26). Inter-individual variation in *in vivo* studies are often due to multiple host-related factors which makes it difficult to interpret microbiome results (27). In contrast, studying the effect of dietary ingredients on the microbiome in the absence of confounding host factors can help to better understand the complex microbial interactions. Bioreactors have been used as model systems to study how dietary ingredients

shape microbiomes (28). The use of bioreactors allows precise control of the environmental conditions that affect microbiome composition which provide increased reproducibility and reveal microbial interactions in a more defined way. Minibioreactor array is an *in vitro* anaerobic model system that simulates hindgut conditions for growth of complex, stable microbiota without interference of host factors (29). Furthermore, this system helps to identify microbial biotransformations and allows for measuring metabolites produced (30). We used minibioreactor array systems to gain deeper understanding of how the microbiota responds to quercetin and/or rice bran without host interference, and hypothesized that the combination of quercetin and rice bran will have a synergistic positive effect on the gut microbiota and microbiota metabolism. We observed that the quercetin and rice bran combination in the minibioreactors were effective to significantly reduce *Enterobacteriaceae* family members and resulted in higher propionate production. This study provides novel insights to species-level shifts in the human gut microbiome in the presence of quercetin and/or rice bran supplementation through an *in vitro* model.

Materials and Methods

Donor samples, mini-bioreactor array preparation and sample collection: We obtained fresh fecal samples from six healthy donors with no prior history of antibiotic consumption in the past year. The pooled fecal sample from six individuals was used as the inoculum as our prior study showed that pooled sample represented the individual microbial composition (31).

The modified BHI medium (31) was used as a control medium. Heat-stabilized rice bran (RBT 300) was purchased from Rice Bran Technologies (Sacramento, CA, USA), and the extract was prepared as described previously (14) and then added to the control medium (final concentration: 2 mg/mL): designated as RB. Quercetin was added to the media at a final concentration of 75 mg/L and is referred to as QC. The final experimental group consisted of modified BHI medium with the additions of quercetin (75 mg/L) and rice bran extract (2 mg/mL) and is identified as QC+RB. Mini-bioreactors (MBRAs) were sterilized, assembled and the experiment was performed as described previously (32) with minor modifications. Briefly, the input and output on Watson Marlow pumps were set at 1 rpm and 2 rpm respectively. The rotating magnetic stirrer was set at 130 rpm. The media (Control, QC, RB, and QC+RB) (Figure 1A) were allowed to flow continuously for 24 hours each in triplicate. Three hundred microliters of the inoculum was introduced into all wells with a retention time of 16 hours. The continuous flow model was operated up to 21 days post-inoculation (Figure 1B). Five hundred microliters of the media was collected for sequencing at day 0 (inoculum), days 4, 7, 14 and 21 and directly frozen to -80^{o}C. Also,

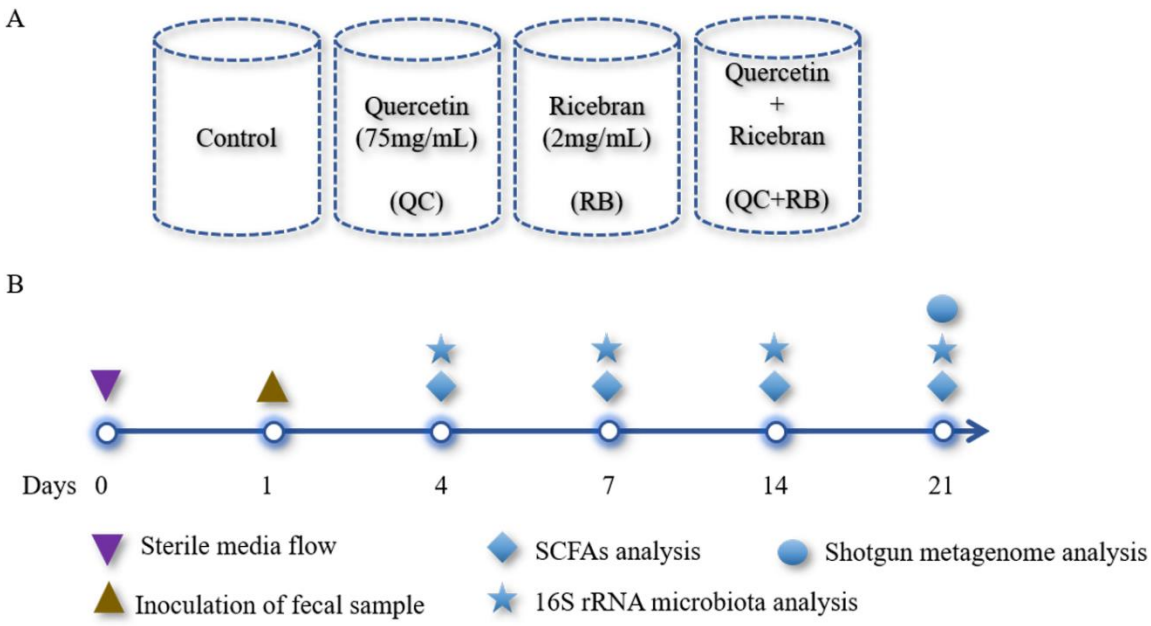


Figure 1: **Overview of the study design**: **A)** A schematic diagram shown the design of four different media conditions used in this study. Single substrates (quercetin or rice bran) or the mixture of quercetin and rice bran were used at the final concentration of 75 mg/L and 2 mg/mL as shown above in the base medium BHI (see Methods section). Final pH of the media was adjusted to 6.8±0.2 for all conditions. Each condition was run in triplicate and were inoculated with the same fecal inoculum (see Methods section). **B)** Outline of the bioreactor experiment showing time points for fecal inoculation and sample collection for SCFAs and microbial community analysis.

samples for short chain fatty acid (SCFAs) determination were collected at days 4, 7, 14 and 21 post-inoculation with dietary treatments and controls (Figure 1B).

Microbial DNA extraction and sequencing: DNA isolation was performed on 50 samples including duplicate inoculum samples. The DNA was extracted from 500 µl of the sample using a Powersoil DNA isolation kit (MoBio Laboratories Inc, CA) following the manufacturer's instructions. After extraction, the quality of DNA was measured using NanoDrop^{TM} one (Thermo Fisher Scientific, DE) and quantified using Qubit Fluorometer 3.0 (Invitrogen, CA). The DNA samples were stored at -20^{o}C until further use. To analyze the variation of the microbial composition over time, all samples were amplicon sequenced using an Illumina MiSeq platform with paired-end V3 chemistry. The library was prepared using an Illumina Nextera XT library preparation kit (Illumina Inc, CA) targeting V3-V4 regions of the 16S rRNA. The libraries were bead normalized and multiplexed before loading into the sequencer. We also performed shotgun metagenome sequencing on twelve samples obtained from day 21 to identify species level taxonomical differences between the experimental groups. We used the Nextera XT kit (Illumina, San Diego, CA) for the preparation of the shotgun metagenome sequencing library. The library was then sequenced using paired end 300 base sequencing chemistry using a MiSeq platform.

***Data analysis*:** The time-series changes in the microbial communities were analyzed using 16S rRNA community analysis in Quantitative Insights into Microbial Ecology framework (QIIME, Version 2.0) (33). Briefly, the demultiplexed reads obtained were quality filtered

using q2-demux plugin and denoised applying DADA2 (34). All amplicon sequence variants were aligned with *mafft* (35) to construct a phylogeny with fasttree2 (36). The outputs rooted-tree.qza, table.qza, taxonomy.qza were then imported into R (37) for analysis using *phyloseq* (38). Shannon diversity and bray curtis dissimilarity indices were calculated as alpha and beta diversity metrices. Kruskal-Wallis test at p=0.05 was performed to compare the species richness between the groups. The reads were normalized by rarefying to 30,000 and the taxonomy was assigned to amplicon sequence variants using the q2-feature-classifier (39) using Greengenes as the reference (40). Rarefying the reads to 30,000 were enough to estimate total diversity and taxonomy (Supplementary Figure 1A). Initially, a total of 947 amplicon variants were identified from 50 samples. The average number of non-chimeric reads per sample obtained in Qiime2 pipeline was 92,512±27,631 (mean ± SD). The amplicon sequence variants obtained were filtered to select those which are present in at least 20% of samples with a count of 10 each or amplicon sequence variants with > 0.001% of total median count reducing the total number to 581.

For shotgun metagenomes, raw fastq sequences were quality controlled using Fastqc (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and host reads were removed using metaWRAP pipeline (41). Filtered reads were analyzed for taxonomy using Kaiju (42) against the proGenomes database (43) using default parameters. The percentage abundance of each taxon was plotted using Explicit v2.10.5 (44). The samples from day 21 yielded a total of 39,332,649 reads of which 70. 86% were classified to 161 different species, whereas 185,188 (5.55%) and 9,278,647 (23.59%) reads remained as unclassified

bacteria or unassigned to non-viral species, respectively. In addition, the raw counts obtained for each taxon were used for calculation of alpha and beta diversity using the *phyloseq* package in R. For identification of differentially abundant taxa in QC, RB and QC+RB compared to control medium, the raw counts of abundance of each taxon was then Log10(x+1) transformed and fed to DESEq2 (45) package in R. Bacterial taxa which significantly altered from the control were further filtered out with the criteria of at least log2foldchange (Log2FC) of ≥ 2 and p*adj* value > 0.05. The enriched taxa were selectively analyzed for their co-relation to SCFAs phenotypic data using Spearman correlation method in R.

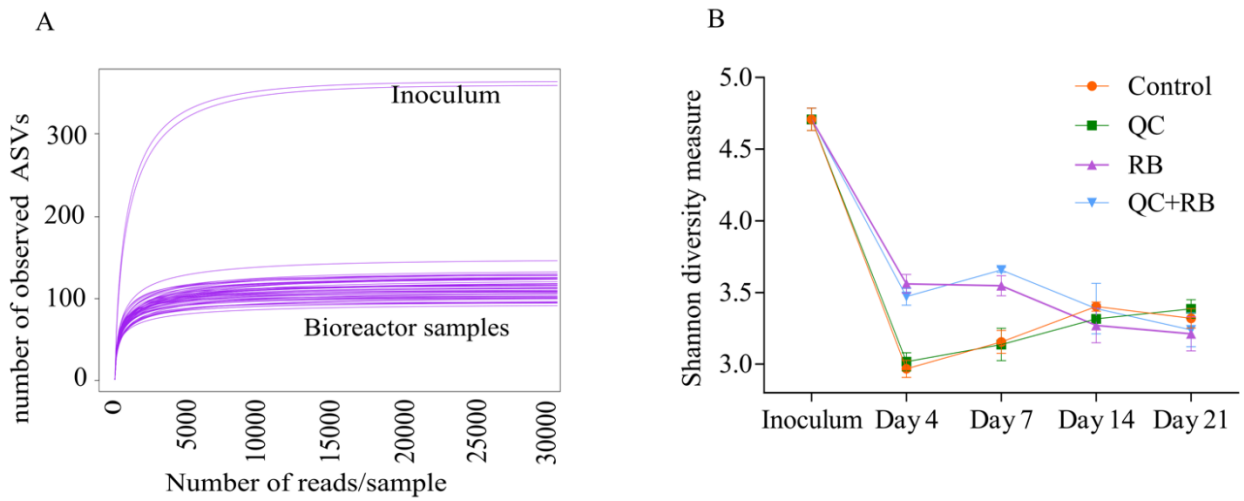
Estimation of Short Chain Fatty Acids (SCFAs): For the estimation of SCFAs, 800 µl of samples were collected from each mini-bioreactor, mixed with 160 µl of 25% m-phosphoric acid and frozen at -80°C until further analysis. Later, the frozen samples were thawed and centrifuged (>15,000 × g) for 20 min. Five hundred microliters of supernatant was collected in the tubes before loading into the gas chromatography- mass spectrometry (GC-MS) (Agilent Technologies, USA) for analysis (31). The SCFA concentrations were compared between the groups using the Kruskal-Wallis test followed by the Dunn test with Benjamini-Hochberg correction in R and visualized using GraphPad Prism 6.0.

Results

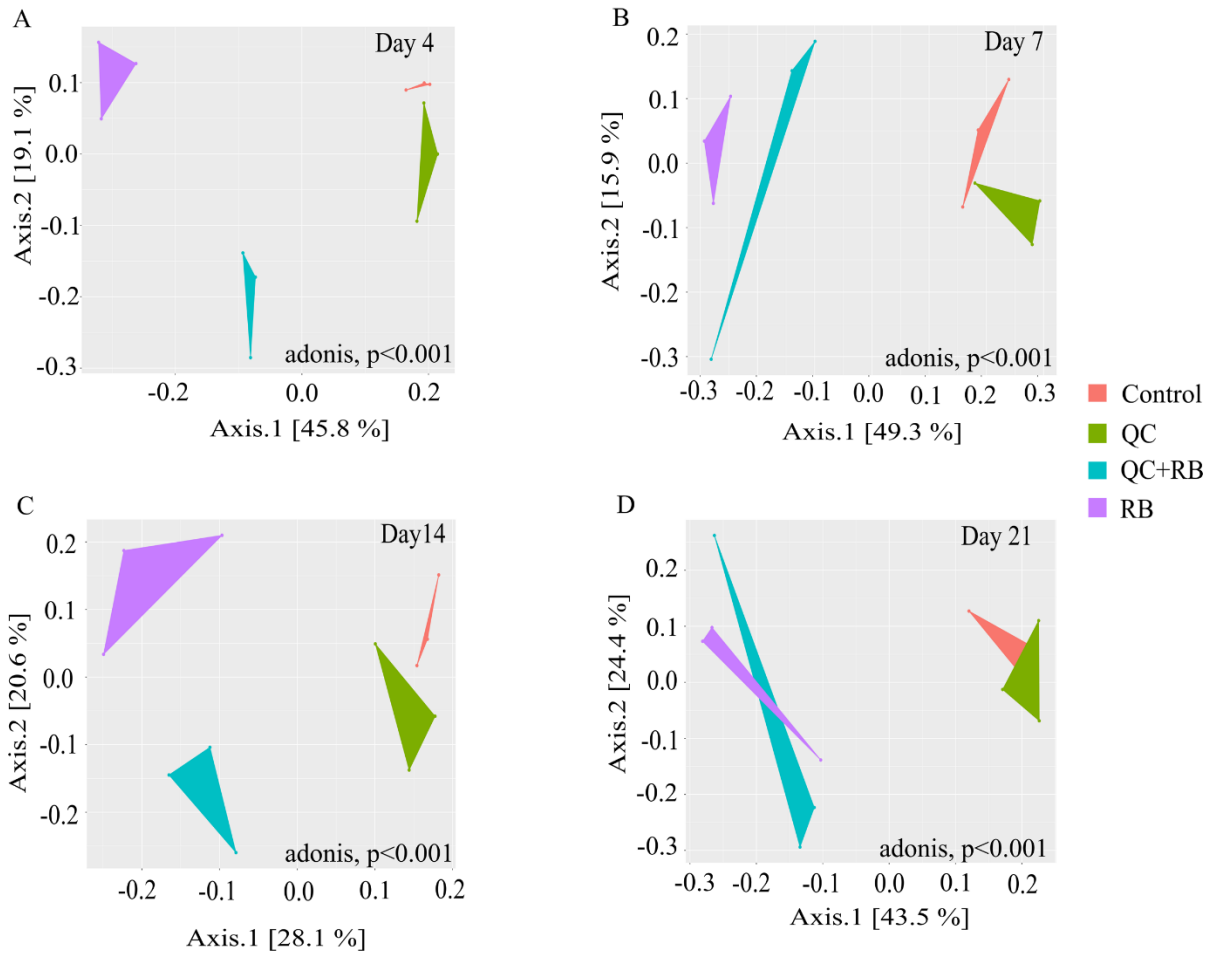
Rice bran produced larger shift in microbiota composition when compared to quercetin

The microbial community richness indicated by the Shannon diversity index showed significant differences in the early days (days 4 and 7) due to addition of rice bran in the media whether alone or in combination with quercetin (Supplementary Figure 1B). No significant changes in the richness was observed among the groups on day 14 and day 21, suggesting the stabilization of the microbial communities by day 14. Also, on day 21, shotgun metagenome analysis showed no differences in the richness between the four conditions (Figure 2B). In contrast, marked differences between the communities were evident as early as day 4 by beta diversity analysis (Supplementary Figure 2A). Rice bran supplementation showed a greater shift in the bacterial community as shown on days 7, 14 and 21 (Supplementary Figure 2B, 2C, 2D). Based on Bray-Curtis distance, RB and QC+RB treatments were similar to one another and clustered separately from control and QC. This supports that there was a limited shift in the community following addition of quercetin when compared to rice bran (Supplementary Figure 2). Shotgun metagenome sequencing at day 21 also indicated significant differences in the community profile, primarily influenced by the addition of rice bran extract in the medium (Figure 2C).

When examined taxonomically, the baseline inoculum composition was dominated by *Prevotellaceae* (29.9±0.87%), followed by *Ruminococcaceae* (18.82±0.81%) and *Lachnospiraceae* (12.66±0.4%). Also, *Enterobacteriaceae* was found to be very low in



Supplementary Figure 1: Exploratory analysis of predicted taxa and alpha diversity of microbiota in quercetin and rice bran: control, QC, RB and QC+RB medium and inoculum obtained using 16S rRNA community analysis A) Rarefaction curves measuring the bacterial diversity in fecal communities. The curves are based on V3-V4 16S rRNA gene sequences obtained from a total of 48 samples (12 each from control, QC, RB, and QC+RB) from day 4, 7, 14 and 21 and duplicate inoculums **B)** Alpha diversity of four groups (control, QC, RB and QC+RB) at day 4, 7, 14 and 21.



Supplementary Figure 2: Beta diversity assessment of the microbiota communities formed at days 4, 7, 14 and 21 among four groups of media condition (Control, QC, RB and QC+RB) determined using 16S rRNA sequence analysis. The communities were ordinated using the "MDS" method followed by Bray-Curtis distance calculation to visualize the difference between the groups. Statistically, PERMANOVA using adonis was calculated on the beta diversity with p=0.004. The upper panel represents the beta diversity in early stages (day 4 and day 7) and the lower panel represents in later stages (day 14 and day 21).

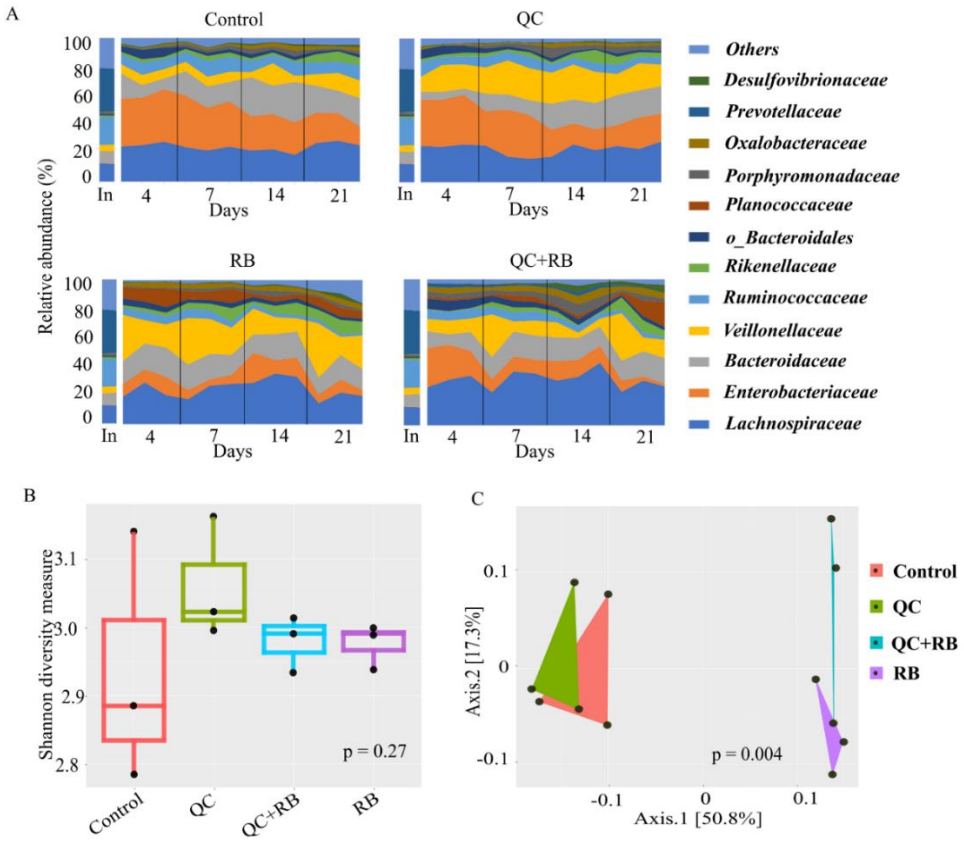


Figure 2: Gut microbiota compositional changes following quercetin and rice bran supplementation A) Temporal family-level composition of the most abundant 12 bacterial taxa at day 4, 7, 14 and 21 and inoculum determined using 16S rRNA community profiling for Control, QC, RB and QC+RB conditions. **B)** Alpha diversity measure (Shannon diversity) between Control, supplemented with QC, supplemented with RB and supplemented with QC+RB determined using shotgun metagenome sequence analysis at day 21. Kruskal-Wallis test was performed for the Shannon diversity indices obtained for the groups (p=0.27). **C)** Beta diversity among the control, QC, RB and QC+RB groups were obtained using shotgun metagenome sequence analysis on day 21. "MDS" ordination followed by Bray-Curtis distance calculation was used to visualize the differences between the groups (adonis, p= 0.004).

inoculum (0.33±0.097%). However, the abundance of *Prevotellaceae* was reduced to ~ 0.0 % after 21 days in all treatment groups suggesting that not all taxa in the inoculum were supported for growth. Over time, there was a major shift in the composition of the microbiota between the groups starting as early as day 4 (Figure 2A). The control and quercetin medium were dominated by *Enterobacteriaceae,* followed by *Lachnospiraceae* at day 4. However, by day 21, *Lachnospiraceae* dominated the control and QC conditions followed by *Bacteroidaceae*. The abundance of *Enterobacteriaceae* was reduced from 34.16±1.92 % and 33.26±1.11% on day 4 to 17.62±4.2% and 18.53±3.88% in control and QC medium respectively by day 21 (Figure 2A, 2B). Even though ~50% reduction of abundance of *Enterobacteriaceae* was observed for both control and QC medium at day 21 compared to day 4, the population of *Enterobacteriaceae* tend to stabilize. However, *Enterobacteriaceae* were highly reduced in both RB and QC+RB conditions. In RB medium, *Enterobacteriaceae* was only reached to 10.212±2.14% at day 4 which further reduced to 6.33±2.59% by day 21. At day 21, *Veillonellaceae* and *Lachnospiraceae* dominanted the RB medium. Similarly, the dominant family were *Lachnospiraceae, Veillonellaceae* and *Bacteroidaceae* in the quercetin and rice bran combination (QC+RB) on day 21. Strikingly, on day 21, the abundance of *Enterobacteriaceae* was reduced by ~18.22% to 4.20±3.16% compared to day 4 on QC+RB condition. At day 21, when mean abundance of *Enterobacteriaceae* was compared between control, QC, RB and QC+RB, significant reduction was observed in RB and QC+RB (Kruskal-Wallis, p=0.0064).

On day 21, we performed shotgun metagenome analysis to identify the species level differences among the four groups (Figure 3). Based on DeSEQ2 analysis, 53 species were

significantly altered in QC+RB medium. The combination (QC+RB) enriched 13 additional species and reduced 13 others when the cut off value of Log2FC≥ 2 and p*adj* value > 0.05 was applied (Supplementary Table 1). Specifically, members of *Enterobacteriaceae, Escherichia* sp. TW09308, *E. coli, E. albertii*, *Citrobacter koseri*, and *C. youngae* were reduced. The *Citrobacter* species, including *C. rodentium,* were also reduced in RB medium; however, *Escherichia* species were not significantly reduced in this medium.

Further, 13 species were altered significantly in the quercetin group (Supplementary Figure 3A), and 70 species in the rice bran group compared to control (Supplementary Figure 3B). With a cut off Log2FC≥ 2 and p*adj* value > 0.05, *Acidaminococcus intestine* was the only enriched taxa in media supplemented with quercetin while *Coprococcus catus, Dorea longicatena, Anaerostipes caccae, Eubacterium limosum* and *Enterococcus faecalis* were highly reduced. Interestingly, the population of *Flavonifractor plautii*, a flavonoid metabolizing bacterium (46, 47) was significantly reduced in quercetin supplementation. Rice bran supplementation alone enriched 15 species, including six different species from *Veillonella* (Supplementary Table 1).

Quercetin and rice bran combination yield higher propionate levels and reduces members of *Enterobacteriaceae* family

Along with the changes in the taxonomy of the microbial community, diet alters the metabolic profile of a community (48, 49). To understand how these dietary substrates have

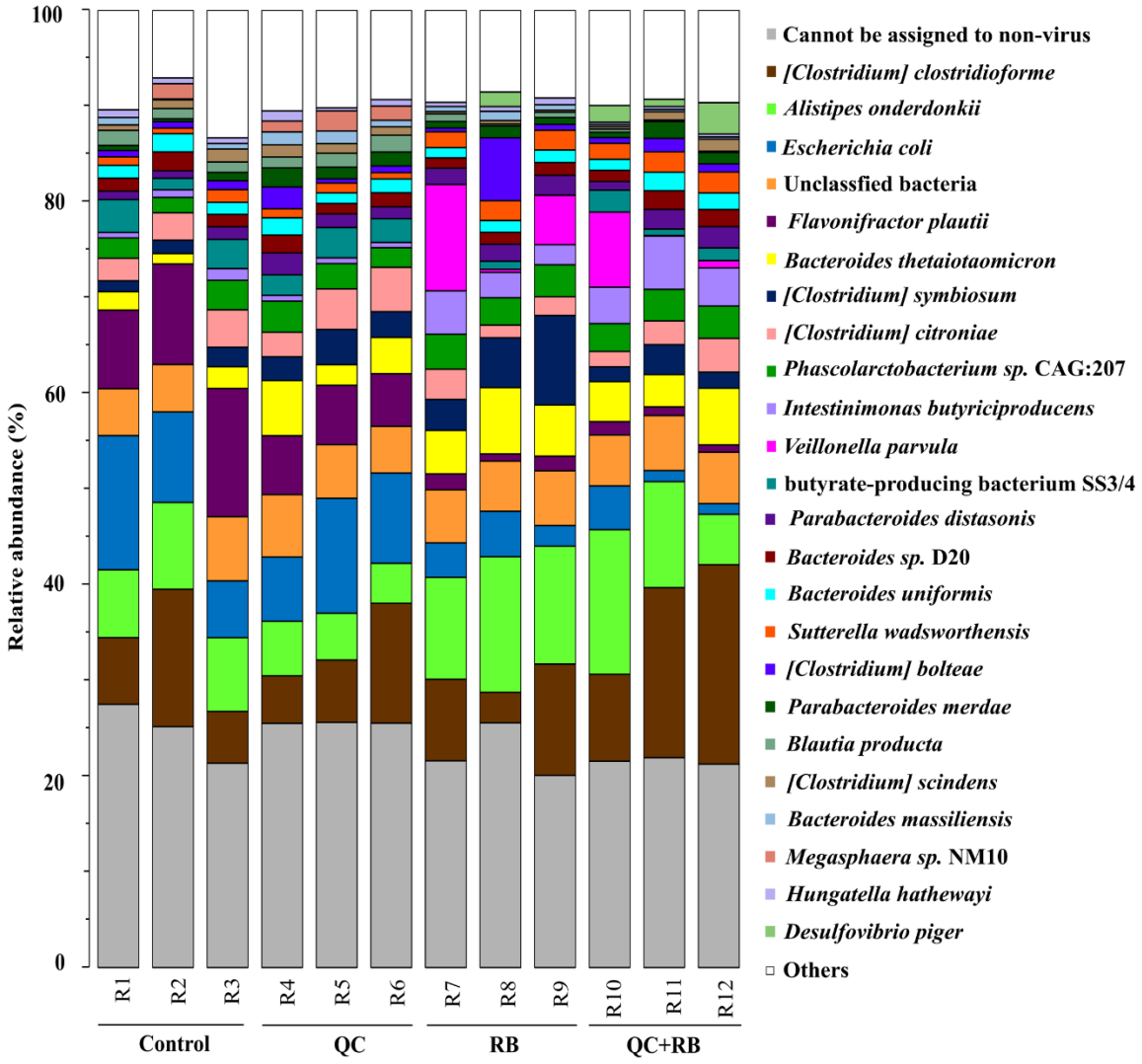
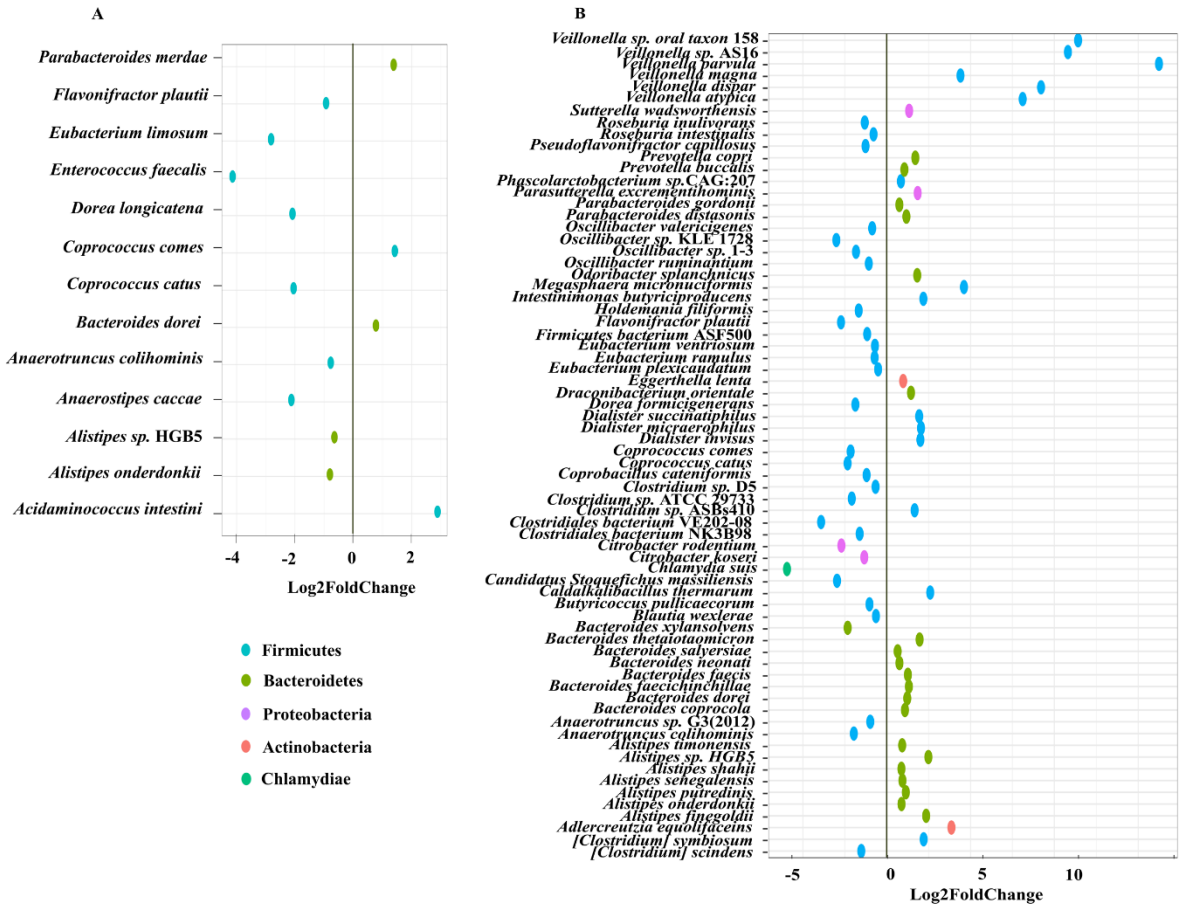


Figure 3: Impact of quercetin (QC) and rice bran (RB) supplementation on bacterial taxonomy. Species-wise variation of the most abundant (top 25) of the bacterial taxa determined by shotgun metagenome sequencing at day 21 (endpoint) for control, QC, RB and QC+RB conditions in triplicate.

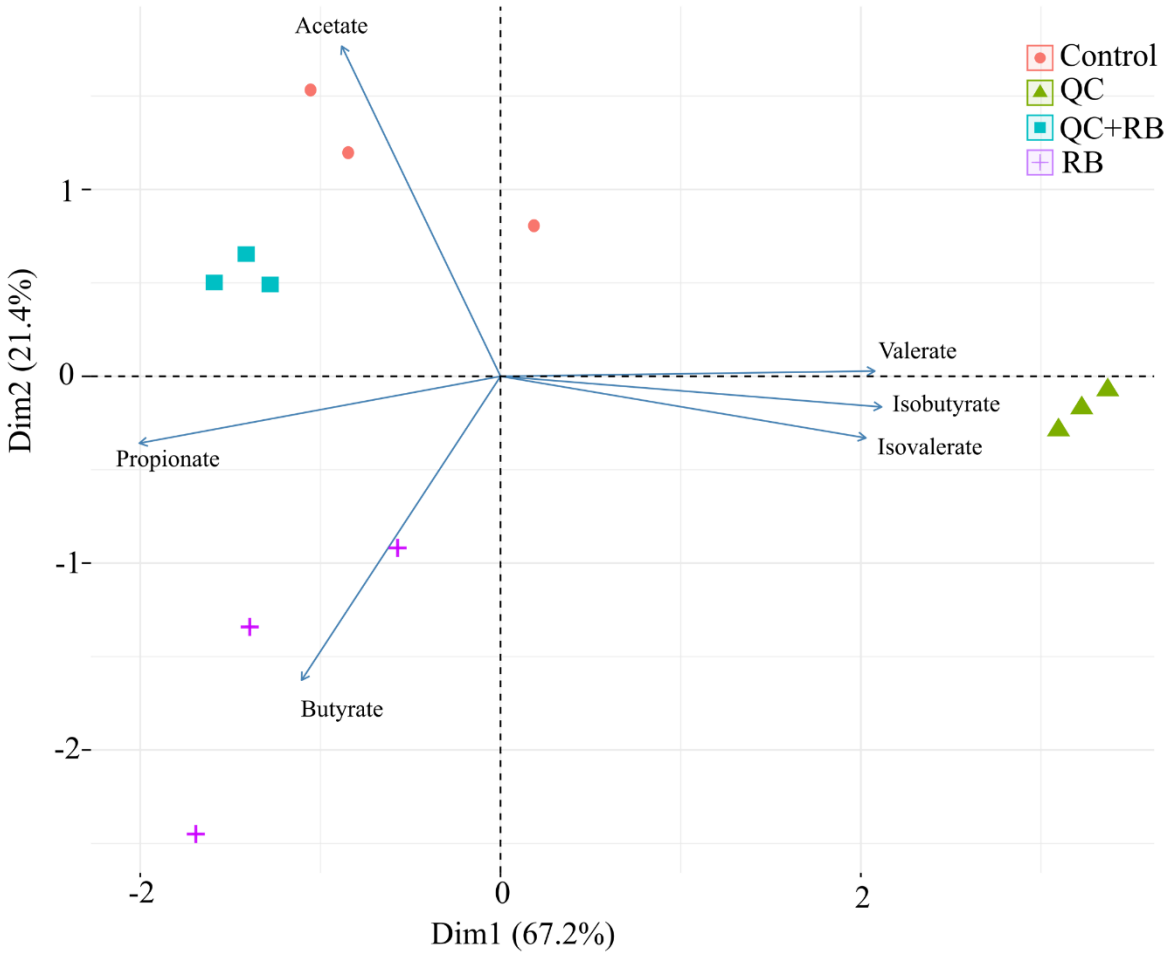
Figure 4: Impact of quercetin and rice bran (QC+RB) supplementation on bacterial species that are significantly altered in QC+RB when compared to the control medium. The differential significance of the species was calculated by using DESEq2 (See Supplementary Table 1) from shotgun metagenome sequence analysis at day 21. The left column represents bacterial species-level taxa while the colors indicate their corresponding phyla ("Blue" =*Firmicutes*, "Brown" =*Bacteroidetes*, "Pink" =*Proteobacteria*, "Orange" =*Actinobacteria*, "Green" =*Chlamydiae*).



Supplementary Figure 3: Impact of quercetin and rice bran (QC, RB) supplementation on bacterial species that are significantly altered in A) QC and B) RB when compared to the control medium. The endpoint (day21) shotgun metagenome sequence analysis was performed and the differential significance of the species was calculated by using DESEq2 (See Supplementary Table 1). The left column represents bacterial species-level taxa while the colors indicate their corresponding phyla ("Blue" =*Firmicutes*, "Brown" =*Bacteroidetes*, "Pink" =*Proteobacteria*, "Orange" =*Actinobacteria*, "Green" =*Chlamydiae*).

altered the fermentation potential of the microbiota, we measured SCFAs from each minibioreactor on days 4, 7, 14 and 21. We observed that the amount of each SCFA produced was stable by day 14 and at the endpoint, with marked differences between the groups (Figure 5A, 5B, 5C, 5D, 5E). At day 21, communities formed in the control and RB weighted strongly towards acetate and butyrate production respectively (Supplementary Figure 4). Statistically at day 21, acetate production in QC+RB group (48.01±3.3 mM) was significantly different from the RB group (33.33±2.44 mM) (Figure 5F). The butyrate production in the RB medium was higher but not significantly different compared to the other groups (Figures 5H). Similarly, the medium supplemented with quercetin and rice bran (QC+RB) weighted towards the production of propionate and butyrate (Supplementary Figure 4). Rice bran alone and combined quercetin and rice bran supplementation significantly raised propionate levels by at least 3 folds compared to the QC group alone (Figure 5G). In contrast, QC supplementation in the medium, weighted towards the production of minor SCFAs, i.e. valerate, isovalerate and isobutyrate. However, only isobutyrate production was significantly higher in QC compared to QC+RB group (Figure 5I).

SCFAs production by microbiota has been associated with pathogen inhibition to benefit the host (50). Physiological levels of SCFAs are reported to reduce *Enterobacteriaceae* members by pH mediated action (51). Specifically, propionate production by a propionate producing consortium has been shown to reduce antibiotic induced dysbiosis (52). In this study, as higher production of propionate weighted towards QC+RB medium at day 21, we estimated the correlation of abundances of significantly altering members of



Supplementary Figure 4: Biplot showing ordination of short-chain fatty acid production and bacterial communities formed at day 21 in control, QC, RB and QC+RB conditions. Abundances of the bacterial communities at day 21 determined by shotgun metagenomics were used with SCFAs profiles of day 21 to generate the plot.

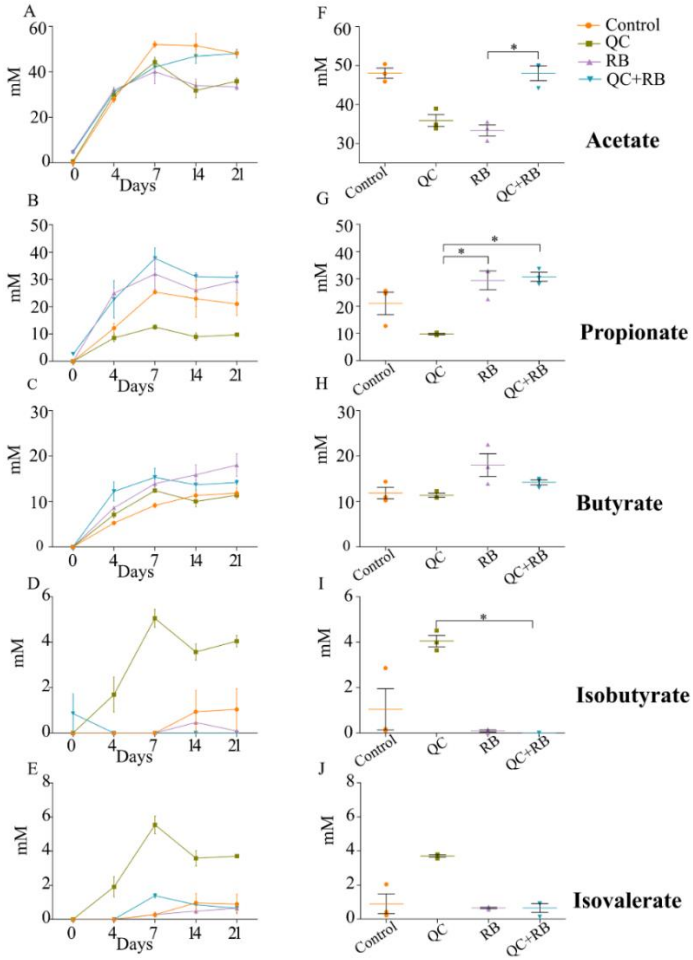


Figure 5: Effect of quercetin (QC) and rice bran (RB) on short Chain Fatty Acids (SCFAs) production in QC, RB, and QC+RB compared to control medium in minibioreactors. A-E represent the periodic variation of acetate, propionate, butyrate, isobutyrate and isovalerate at day 0, 4, 7, 14 and 21 respectively. F-J represent the comparison of concentrations of acetate, propionate, butyrate, isobutyrate and isovalerate respectively from control, QC, RB and QC+RB medium at day 21 (endpoint). Kruskal-Wallis test was performed between the groups and *posthoc* (Dunn test) analysis was performed to identify the different significant groups. "*" represents significance at 0.05. Error bars represent standard error of the mean of data obtained from three different bioreactors.

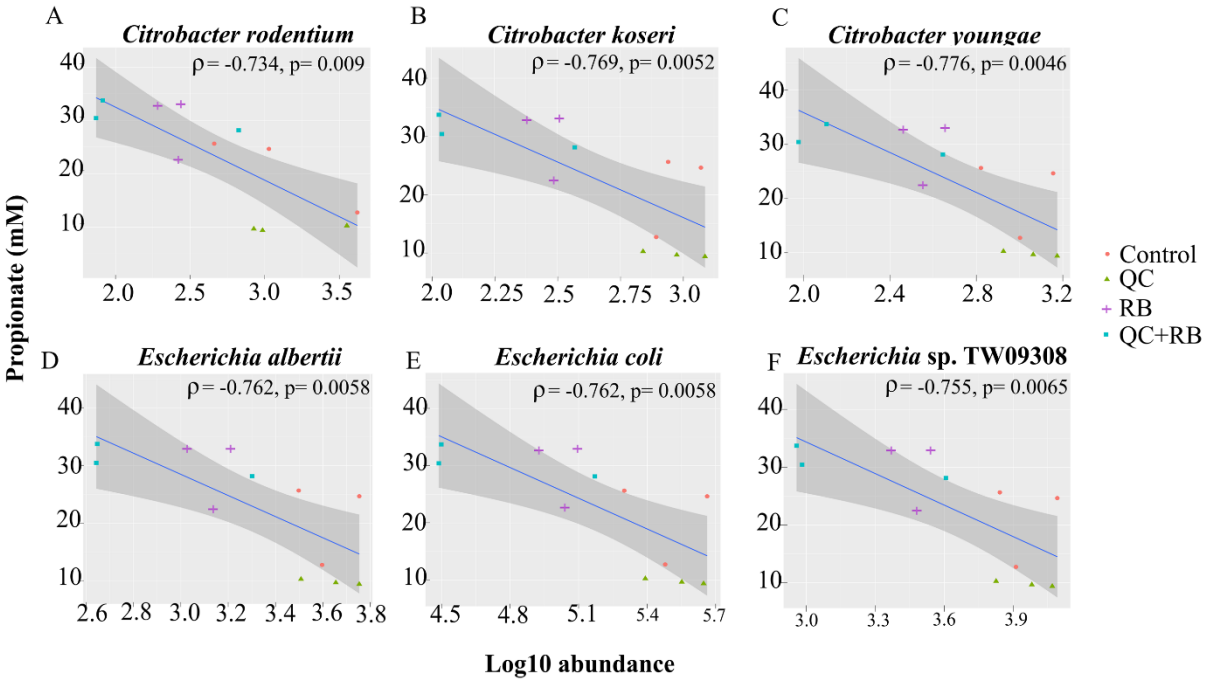


Figure 6: Spearman correlation of *Enterobacteriaceae* family members (*Citrobacter* and *Escherichia* species) log10 abundances at day 21 determined by shotgun metagenomics sequence analysis to levels of propionate in the medium. A negative correlation is expressed by negative values of correlation coefficient "rho" (ρ) with corresponding p-values.

Enterobacteriaceae to propionate levels. Significant high negative correlations were observed between the abundance of the members of *Enterobacteriaceae* family and propionate with media change (Figure 6). *Citrobacter rodentium* (ρ= -0.734, p= 0.009), *C. koseri* (ρ= -0.769, p=0.0052), *C. youngae* (ρ= -0.776, p= 0.0046), *Escherichia albertii* (ρ= -0.762, p= 0.0058), *E. coli* (ρ= -0.762, p= 0.00587) and *Escherichia* sp. TW09308 (ρ= -0.755, p= 0.0065) (Figure 6A, B, C, D, E, F) were greatly reduced in QC+RB medium, whereas propionate levels were at least 3 fold and 1.5 fold higher compared to QC and control medium .

Discussion

In vivo studies have shown changes in microbial communities due to quercetin or rice bran and describe the improvement of colonization resistance and reduction of colon cancer (11, 14, 53). However, these studies are marked by high variations of the microbiota composition likely due to host factors. Additionally, very little is known about the combined effect of quercetin and quercetin on gut microbiota. Thus, we focused on understanding the impact of quercetin and rice bran separately and in combination using a minibioreactor array model (32). Also, to gain insights into taxonomical composition, we performed 16S rRNA sequencing over time and shotgun metagenome sequencing at the endpoint.

Using modified BHI medium, we were able to capture 30.93 % of the total amplicon sequence variants from the minibioreactors compared to the inoculum. This finding is

slightly higher than previous results where the MBRA cultivated only 15-25% of the initial fecal community (29). The reason for the higher number of cultivated bacteria may be due to the modified BHI medium which has been shown to cultivate a wide range of gut bacteria (31). On the other hand, the highly dominant *Prevotellaceae* was lost in all four conditions in this study suggesting that fermenter adapted communities can lose much of their initial diversity (54). The alpha diversity indicated that the richness of the communities in bioreactors stabilized by day 14 (Supplementary Figure 1B). However, the taxonomical differences were evident as early as day four following inoculation. The changes in the relative abundances of taxa after day 14 were lower in control and quercetin supplemented medium whereas rice bran and combined quercetin with rice bran conditions had a more homogenized microbial composition after day 4 (Figure 2A). This indicates that the time required for the stabilization of communities may vary with the substrate used in minibioreactors in contrast to the previously determined timeframe of one week following fecal inoculation (29). Such changes in the time required for stabilization and diversity between the groups can be attributed ecologically to the stochastic rearrangement of the microbial species and their interactions at early stages because of selection pressure by substrate (55).

Rice bran is a nutrient-dense food with a unique profile and ratio of bioactive phytochemicals such as gamma oryzanol, tocotrienols, ferulic acid, vitamin B, beta-sitosterol and many others (12). It has been reported to be effective in preventing *Salmonella typhimurium* (9, 14, 56), rotavirus (57, 58) and norovirus (15) infections by priming intestinal immune cells. However, very few *in vivo* studies have highlighted its

effect on gut microbiota and the reports have striking differences. In a clinical trial, Zambrana et al., (2019) showed significant enrichment of *Veillonella*, *Megasphaera* and *Dialister* species at the genus level from gut samples of children from either Nicaragua or Mali at 12 months' time (24). Another study by Sheflin et al., (2015) showed a significant increase in the abundance of *Methanobrevibacter smithii, Paraprevotella clara, Ruminococcus flavefaciens, Dialister succinatiphilus, Bifidobacterium* sp., *Clostridium glycolicum, Barnesiella intestinihominis, Anaerostipes caccae* and *Ruminococcus bromii* OTUs after heat stabilized rice bran was fed to people (3 g/day) (25). The differences in the enriched taxa could be because of the unique inoculum used in this study. However, both above studies had lower resolution and report enrichment of different taxa which could be attributed to a different variety of rice bran being used and individualized host factors (59-61). The compounding hosts' factors along with variation in age and geography (62), diet pattern (63), lifestyle (64-66), etc. play a crucial role in determining the gut microbiota composition, thus masking the actual effect of the substrate alone upon the diverse community. In contrast, this study supplies species-level resolution eliminating host interference and shows *Veillonella Prevotella, Dialister, Bacteroides* and *Alistipes* species are significantly enriched while *Oscillibacter, Eubacterium* and *Citrobacter* species are significantly reduced (Supplementary Figure 3B).

Similar to rice bran, previous studies analyzing microbial composition after quercetin supplementation yielded variable results among different hosts. *Enterobacteriaceae* and *Fusobacteriaceae* were reported to be positively related to quercetin supplementation whereas *Suttrellaceae* and *Oscillospiraceae* were found to be negatively correlated (26).

However, Lei et al., (2019) reported an increase in abundances of *Bifidobacterium, Bacteroides, Lactobacillus* and *Clostridium*, with a reduction of *Fusobacterium* and *Enterococcus* in mice fed with quercetin (11). With higher resolution and removal of host factors, our results contrast with both studies and show enrichment of *Acidaminococcus intestini* and decreased abundances of *E. limosum, E. faecalis, A. caccae, D. longicatena* and *C. catus*. The domination of *Enterobacteriaceae* in medium supplemented with quercetin (Figure 2A, Supplementary Figure 3C) might have resulted in lower enrichment of the bacterial taxa as *Enterobacteriaceae* has been reported to affect quercetin metabolism by directly or indirectly inhibiting quercetin degrading bacteria (26).

When the combined effect of quercetin and rice bran were analyzed, we find that majority of the microbial shift is due to the supplementation of rice bran (Figure 2, 3, Supplementary Figure 2). Most of the taxa enriched or decreased in the combination were very similar to those affected by rice bran supplementation alone. Compared to quercetin and rice bran supplementation separately, the combination was observed to significantly reduce members of the *Enterobacteriaceae* family (*Escherichia* and *Citrobacter* sp.) (Supplementary Table 1, Figure 6). *Enterobacteriaceae* consists of class of pathogens that are low in abundance but have potential to grow and dominate during dysbiotic conditions (67-69). The reduction of *Enterobacteriaceae* by combined quercetin and rice bran suggests a possibly beneficial effect on the host.

The reduction of *Enterobacteriaceae* members was highly correlated with greater propionate levels in quercetin and rice bran combined medium (Figure 5G, Figure 6).

Propionate, along with other SCFAs produced by the gut microbiota, has been previously implicated in regulating the intestinal morphology and functions (70). Interestingly, unlike butyrate that is used as an energy source for the colonocytes, the health benefits of propionate are not restricted to the colon. Propionate has been shown to decrease liver lipogenesis, and hepatic and plasma cholesterol levels in rats. It has also been implicated in reducing obesity, by stimulating satiety in mice models. Importantly, propionate has also been demonstrated as a potential anti-inflammatory and anti-cancer agent. The reported anti-inflammatory abilities of propionate are suggested to occur through the inhibition of Nuclear Factor-Kappa B and suppression of IL-6 mRNA and other immune-related gene expression, while its anti-cancer effects are thought to occur through inhibition of Histone Deacetylases (HDACs) and regulation of the AP-1 pathway (71-73). Therefore, our observations of increased propionate levels, along with previous reports on the health benefits of propionate, demonstrates the importance of both flavonoids and fiber in the diet. This study, for the first time, reports the combined effect of quercetin and rice bran on the gut microbial composition in the absence of interfering host factors. Here, we report that the gut microbial composition was altered favorably by quercetin and rice bran to result in a significant reduction of opportunistic pathogens that could potentially provide additional health benefits to the host.

Conclusion

Overall, the combined effect of quercetin and rice bran to shift gut microbiota to reduce *Enterobacteriaceae* was correlated with higher levels of propionate production *in vitro*. Even though the impact of these dietary ingredients separately is described to be beneficial to human health, their combined effect was not known. Our results show that a combination of substrates such as quercetin and rice bran will be beneficial in excluding enteric pathogens in the gut. However, it is crucial to culture, isolate and characterize the bacteria enriched in quercetin and rice bran supplemented medium to develop potential synbiotic formulations for gut health and immunity. Further, *in vivo* validation of such a combination would be necessary to understand its implication on the host.

Raw sequence data from 16S rRNA amplicon sequencing and shotgun metagenome sequencing associated with this study has been deposited in NCBI Sequence Read Archive under accession number PRJNA606575.

Literature Cited

1. Maji A, Misra R, Dhakan DB, Gupta V, Mahato NK, Saxena R, Mittal P, Thukral N, Sharma E, Singh A, Virmani R, Gaur M, Singh H, Hasija Y, Arora G, Agrawal A, Chaudhry A, Khurana JP, Sharma VK, Lal R, Singh Y. 2018. Gut microbiome contributes to impairment of immunity in pulmonary tuberculosis patients by alteration of butyrate and propionate producers. Environ Microbiol 20:402-419.
2. Pulikkan J, Maji A, Dhakan DB, Saxena R, Mohan B, Anto MM, Agarwal N, Grace T, Sharma VK. 2018. Gut Microbial Dysbiosis in Indian Children with Autism Spectrum Disorders. Microb Ecol 76:1102-1114.
3. Gentile CL, Weir TL. 2018. The gut microbiota at the intersection of diet and human health. Science 362:776-780.
4. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559-63.
5. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. 2009. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med 1:6ra14.
6. Dhakan DB, Maji A, Sharma AK, Saxena R, Pulikkan J, Grace T, Gomez A, Scaria J, Amato KR, Sharma VK. 2019. The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. Gigascience 8.

7. Slavin J. 2013. Fiber and prebiotics: mechanisms and health benefits. Nutrients 5:1417-35.

8. Gibson GR, Roberfroid MB. 1995. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. J Nutr 125:1401-12.

9. Goodyear A, Kumar A, Ehrhart EJ, Swanson KS, Grusak MA, Leach JE, Dow SW, McClung A, Ryan EP. 2015. Dietary rice bran supplementation prevents Salmonella colonization differentially across varieties and by priming intestinal immunity. Journal of Functional Foods 18:653-664.

10. Henderson AJ, Kumar A, Barnett B, Dow SW, Ryan EP. 2012. Consumption of rice bran increases mucosal immunoglobulin A concentrations and numbers of intestinal Lactobacillus spp. J Med Food 15:469-75.

11. Lin R, Piao M, Song Y. 2019. Dietary Quercetin Increases Colonic Microbial Diversity and Attenuates Colitis Severity in Citrobacter rodentium-Infected Mice. Front Microbiol 10:1092.

12. Zarei I, Brown DG, Nealon NJ, Ryan EP. 2017. Rice Bran Metabolome Contains Amino Acids, Vitamins & Cofactors, and Phytochemicals with Medicinal and Nutritional Properties. Rice (N Y) 10:24.

13. Islam J, Koseki T, Watanabe K, Ardiansyah, Budijanto S, Oikawa A, Alauddin M, Goto T, Aso H, Komai M, Shirakawa H. 2017. Dietary Supplementation of Fermented Rice Bran Effectively Alleviates Dextran Sodium Sulfate-Induced Colitis in Mice. Nutrients 9.

14. Kumar A, Henderson A, Forster GM, Goodyear AW, Weir TL, Leach JE, Dow SW, Ryan EP. 2012. Dietary rice bran promotes resistance to Salmonella enterica serovar Typhimurium colonization in mice. BMC Microbiol 12:71.

15. Lei S, Ramesh A, Twitchell E, Wen K, Bui T, Weiss M, Yang X, Kocher J, Li G, Giri-Rachman E, Trang NV, Jiang X, Ryan EP, Yuan L. 2016. High Protective Efficacy of Probiotics and Rice Bran against Human Norovirus Infection and Diarrhea in Gnotobiotic Pigs. Front Microbiol 7:1699.

16. D'Archivio M, Filesi C, Di Benedetto R, Gargiulo R, Giovannini C, Masella R. 2007. Polyphenols, dietary sources and bioavailability. Ann Ist Super Sanita 43:348-61.

17. Stewart LK, Soileau JL, Ribnicky D, Wang ZQ, Raskin I, Poulev A, Majewski M, Cefalu WT, Gettys TW. 2008. Quercetin transiently increases energy expenditure but persistently decreases circulating markers of inflammation in C57BL/6J mice fed a high-fat diet. Metabolism 57:S39-46.

18. Comalada M, Camuesco D, Sierra S, Ballester I, Xaus J, Galvez J, Zarzuelo A. 2005. In vivo quercitrin anti-inflammatory effect involves release of quercetin, which inhibits inflammation through down-regulation of the NF-kappaB pathway. Eur J Immunol 35:584-92.

19. Egert S, Bosy-Westphal A, Seiberl J, Kurbitz C, Settler U, Plachta-Danielzik S, Wagner AE, Frank J, Schrezenmeir J, Rimbach G, Wolffram S, Muller MJ. 2009. Quercetin reduces systolic blood pressure and plasma oxidised low-density lipoprotein concentrations in overweight subjects with a high-cardiovascular

disease risk phenotype: a double-blinded, placebo-controlled cross-over study. Br J Nutr 102:1065-74.

20. Darband SG, Kaviani M, Yousefi B, Sadighparvar S, Pakdel FG, Attari JA, Mohebbi I, Naderi S, Majidinia M. 2018. Quercetin: A functional dietary flavonoid with potential chemo-preventive properties in colorectal cancer. J Cell Physiol 233:6544-6560.

21. Linsalata M, Orlando A, Messa C, Refolo MG, Russo F. 2010. Quercetin inhibits human DLD-1 colon cancer cell growth and polyamine biosynthesis. Anticancer Res 30:3501-7.

22. Hong Z, Piao M. 2018. Effect of Quercetin Monoglycosides on Oxidative Stress and Gut Microbiota Diversity in Mice with Dextran Sodium Sulphate-Induced Colitis. Biomed Res Int 2018:8343052.

23. Harwood M, Danielewska-Nikiel B, Borzelleca JF, Flamm GW, Williams GM, Lines TC. 2007. A critical review of the data related to the safety of quercetin and lack of evidence of in vivo toxicity, including lack of genotoxic/carcinogenic properties. Food Chem Toxicol 45:2179-205.

24. Zambrana LE, McKeen S, Ibrahim H, Zarei I, Borresen EC, Doumbia L, Boré A, Cissoko A, Douyon S, Koné K, Perez J, Perez C, Hess A, Abdo Z, Sangaré L, Maiga A, Becker-Dreps S, Yuan L, Koita O, Vilchez S, Ryan EP. 2019. Rice bran supplementation modulates growth, microbiota and metabolome in weaning infants: a clinical trial in Nicaragua and Mali. Scientific Reports 9:13919.

25. Sheflin AM, Borresen EC, Wdowik MJ, Rao S, Brown RJ, Heuberger AL, Broeckling CD, Weir TL, Ryan EP. 2015. Pilot dietary intervention with heat-

stabilized rice bran modulates stool microbiota and metabolites in healthy adults. Nutrients 7:1282-300.

26. Tamura M, Hoshi C, Kobori M, Takahashi S, Tomita J, Nishimura M, Nishihira J. 2017. Quercetin metabolism by fecal microbiota from healthy elderly human subjects. PLoS One 12:e0188271.

27. Zoetendal EG, Akkermans AD, De Vos WM. 1998. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. Appl Environ Microbiol 64:3854-9.

28. Chung WSF, Walker AW, Vermeiren J, Sheridan PO, Bosscher D, Garcia-Campayo V, Parkhill J, Flint HJ, Duncan SH. 2019. Impact of carbohydrate substrate complexity on the diversity of the human colonic microbiota. FEMS Microbiol Ecol 95.

29. Auchtung JM, Robinson CD, Britton RA. 2015. Cultivation of stable, reproducible microbial communities from different fecal donors using minibioreactor arrays (MBRAs). Microbiome 3:42.

30. PNAS. 2017. Environmental Chemicals, the Human Microbiome, and Health Risk: A Research Strategy. *In* PNAS (ed), Environmental Chemicals, the Human Microbiome, and Health Risk: A Research Strategy doi:10.17226/24960, Washington (DC).

31. Ghimire S, Roy C, Wongkuna S, Antony L, Maji A, Keena MC, Foley A, Scaria J. 2020. Identification of Clostridioides difficile-Inhibiting Gut Commensals Using Culturomics, Phenotyping, and Combinatorial Community Assembly. mSystems 5.

32. Auchtung JM, Robinson CD, Farrell K, Britton RA. 2016. MiniBioReactor Arrays (MBRAs) as a Tool for Studying C. difficile Physiology in the Presence of a Complex Community. Methods Mol Biol 1476:235-58.

33. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37:852-857.

34. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581-3.

35. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-66.

36. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.

37. Team RC. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria

38. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217.

39. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90.

40. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6:610-8.

41. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158.

42. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7:11257.

43. Mende DR, Letunic I, Huerta-Cepas J, Li SS, Forslund K, Sunagawa S, Bork P. 2017. proGenomes: a resource for consistent functional and taxonomic annotations of prokaryotic genomes. Nucleic Acids Res 45:D529-D534.

44. Robertson CE, Harris JK, Wagner BD, Granger D, Browne K, Tatem B, Feazel LM, Park K, Pace NR, Frank DN. 2013. Explicet: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. Bioinformatics 29:3100-1.

45. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

46. Gupta A, Dhakan DB, Maji A, Saxena R, P.K. VP, Mahajan S, Pulikkan J, Kurian J, Gomez AM, Scaria J, Amato KR, Sharma AK, Sharma VK. 2019. Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. mSystems 4:e00438-19.

47. Braune A, Blaut M. 2016. Bacterial species involved in the conversion of dietary flavonoids in the human gut. Gut Microbes 7:216-34.

48. Sonnenburg JL, Backhed F. 2016. Diet-microbiota interactions as moderators of human metabolism. Nature 535:56-64.

49. Sonnenburg ED, Sonnenburg JL. 2014. Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. Cell Metab 20:779-786.

50. Hung CC, Garner CD, Slauch JM, Dwyer ZW, Lawhon SD, Frye JG, McClelland M, Ahmer BM, Altier C. 2013. The intestinal fatty acid propionate inhibits Salmonella invasion through the post-translational control of HilD. Mol Microbiol 87:1045-60.

51. Sorbara MT, Dubin K, Littmann ER, Moody TU, Fontana E, Seok R, Leiner IM, Taur Y, Peled JU, van den Brink MRM, Litvak Y, Baumler AJ, Chaubard JL, Pickard AJ, Cross JR, Pamer EG. 2019. Inhibiting antibiotic-resistant Enterobacteriaceae by microbiota-mediated intracellular acidification. J Exp Med 216:84-98.

52. El Hage R, Hernandez-Sanabria E, Calatayud Arroyo M, Props R, Van de Wiele T. 2019. Propionate-Producing Consortium Restores Antibiotic-Induced Dysbiosis in

a Dynamic in vitro Model of the Human Intestinal Microbial Ecosystem. Front Microbiol 10:1206.

53. Zhang XA, Zhang S, Yin Q, Zhang J. 2015. Quercetin induces human colon cancer cells apoptosis by inhibiting the nuclear factor-kappa B Pathway. Pharmacogn Mag 11:404-9.

54. McDonald JA, Schroeter K, Fuentes S, Heikamp-Dejong I, Khursigara CM, de Vos WM, Allen-Vercoe E. 2013. Evaluation of microbial community reproducibility, stability and composition in a human distal gut chemostat model. J Microbiol Methods 95:167-74.

55. Vellend M. 2010. Conceptual synthesis in community ecology. Q Rev Biol 85:183-206.

56. Rubinelli PM, Kim SA, Park SH, Roto SM, Nealon NJ, Ryan EP, Ricke SC. 2017. Differential effects of rice bran cultivars to limit Salmonella Typhimurium in chicken cecal in vitro incubations and impact on the cecal microbiome and metabolome. PLoS One 12:e0185002.

57. Yang X, Twitchell E, Li G, Wen K, Weiss M, Kocher J, Lei S, Ramesh A, Ryan EP, Yuan L. 2015. High protective efficacy of rice bran against human rotavirus diarrhea via enhancing probiotic growth, gut barrier function, and innate immunity. Sci Rep 5:15004.

58. Nealon NJ, Yuan L, Yang X, Ryan EP. 2017. Rice Bran and Probiotics Alter the Porcine Large Intestine and Serum Metabolomes for Protection against Human Rotavirus Diarrhea. Front Microbiol 8:653.

59. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich CF, Funai K, Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters WA, Knight R, Newgard CB, Heath AC, Gordon JI. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science 341:1241214.

60. Stewart JA, Chadwick VS, Murray A. 2005. Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. J Med Microbiol 54:1239-42.

61. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. 2014. Human genetics shape the gut microbiome. Cell 159:789-99.

62. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. 2012. Human gut microbiome viewed across age and geography. Nature 486:222-7.

63. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science 334:105-8.

64. Cook MD, Allen JM, Pence BD, Wallig MA, Gaskins HR, White BA, Woods JA. 2016. Exercise and gut immune function: evidence of alterations in colon immune

cell homeostasis and microbiome characteristics with exercise training. Immunol Cell Biol 94:158-63.

65. Benedict C, Vogel H, Jonas W, Woting A, Blaut M, Schurmann A, Cedernaes J. 2016. Gut microbiota and glucometabolic alterations in response to recurrent partial sleep deprivation in normal-weight young individuals. Mol Metab 5:1175-1186.

66. Karl JP, Margolis LM, Madslien EH, Murphy NE, Castellani JW, Gundersen Y, Hoke AV, Levangie MW, Kumar R, Chakraborty N, Gautam A, Hammamieh R, Martini S, Montain SJ, Pasiakos SM. 2017. Changes in intestinal microbiota composition and metabolism coincide with increased intestinal permeability in young adults under prolonged physiological stress. Am J Physiol Gastrointest Liver Physiol 312:G559-G571.

67. Seksik P, Rigottier-Gois L, Gramet G, Sutren M, Pochart P, Marteau P, Jian R, Dore J. 2003. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. Gut 52:237-42.

68. Walker AW, Sanderson JD, Churcher C, Parkes GC, Hudspith BN, Rayment N, Brostoff J, Parkhill J, Dougan G, Petrovska L. 2011. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. BMC Microbiol 11:7.

69. Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne A, Lee YJ, Dubin KA, Socci ND, Viale A, Perales MA, Jenq RR, van den Brink MR, Pamer EG. 2012. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clin Infect Dis 55:905-14.

70. Scheppach W. 1994. Effects of short chain fatty acids on gut morphology and function. Gut 35:S35-8.

71. Hosseini E, Grootaert C, Verstraete W, Van de Wiele T. 2011. Propionate as a health-promoting microbial metabolite in the human gut. Nutr Rev 69:245-58.

72. Tedelind S, Westberg F, Kjerrulf M, Vidal A. 2007. Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel disease. World J Gastroenterol 13:2826-32.

73. Nepelska M, Cultrone A, Beguet-Crespel F, Le Roux K, Dore J, Arulampalam V, Blottiere HM. 2012. Butyrate produced by commensal bacteria potentiates phorbol esters induced AP-1 response in human intestinal epithelial cells. PLoS One 7:e52869.

CHAPTER 4: *CLOSTRIDIUM SCINDENS* INHIBITS *CLOSTRIDIOIDES DIFFICILE* BY BILE ACID DEPENDENT AND INDEPENDENT MECHANISMS: GENETIC AND PHENOTYPIC ANALYSIS

Sudeep Ghimire [1,2], Kinchel C. Doerner [3], Jennifer M. Auchtung [4], Robert A. Britton [5], Joy Scaria [1,2]*

[1]Department of Veterinary and Biomedical Sciences, South Dakota State University, Brookings, SD, USA.

[2]South Dakota Center for Biologics Research and Commercialization, SD, USA.

[3]Department of Biology and Microbiology, South Dakota State University, Brookings, SD, USA.

[4]Department of Food Science and Technology, University of Nebraska, Lincoln, NE, USA.

[5]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA

#Address correspondence to Joy Scaria, joyscaria@gmail.com

Abstract

Clostridioides (*Clostridium*) *difficile* is a gram-positive, endospore-forming, anaerobic rod that causes severe gastrointestinal disease resulting in thousands of deaths every year worldwide. It is the most common cause of hospital acquired diarrhea with a 15-30% relapsing rate. *C. difficile* growth is inhibited by the secondary bile acids deoxycholic acid and lithocholic acid. *Clostridium scindens*, an intestinal bacterium capable of 7-α-dehyroxylation of primary bile acids (i.e., cholic and chenodeoxycholic acids) to secondary bile acids, has demonstrated promise to treat *C. difficile* infections. This study compares genetic diversity among five *C. scindens* strains and uses batch and continuous flow mini-bioreactor culture systems to determine the ability of these strains to inhibit the growth of *C. difficile* strain R20291. All five strains inhibited *C. difficile* R20291, with two distinct mechanisms observed. Two *C. scindens* strains directly inhibited *C. difficile* R20291 in the absence of cholic acid but also inhibited *C. difficile* R20291 in a dose dependent manner when cholic acid was included. Three *C. scindens* strains directly inhibited *C. difficile* R20291 but did not further inhibit when cholic acid was added. Filter-sterilized culture supernatant from each of the five *C. scindens* strains inhibited growth of *C. difficile* R20291 when added to the growth media of *C. difficile* R20291, but inhibition was significantly reduced when heat treated culture supernatant was used indicating the presence of heat-labile inhibitors in the culture supernatant. Results indicate that *C. scindens* inhibits *C. difficile* in both a bile acid-dependent and independent manner.

Importance

Clostridium scindens is normally found in the gastrointestinal tracts of humans and modifies bile acids produced by the host. These modified bile acids are thought to inhibit the growth of *Clotridioides* (*Clostridium difficile*), a potent intestinal pathogen. Here, we study the genomes of five strains of *C. scindens*, perform a genome wide analysis for each strain, and determine if *C. scindens* inhibits *C. difficile*. Results indicate that all *C. scindens* strains tested inhibit *C. difficile* growth by production of secretory extracellular alkaloids. The presence of bile acids enhances *C. difficile* inhibition by two *C. scindens* strains. Substantial differences in the nature and the degree of inhibition of *C. scindens* against *C. difficile* are observed. Such metabolic and genetic variations in capabilities of *C. scindens* strains to inhibit *C. difficile* is an important consideration during *C. difficile* infection treatments. Utilizing *C. scindens* strains to inhibit *C. difficile* through bile acid-independent mechanisms could lead to improved bacteriotherapy against *C. difficile* infection.

Key words: *Clostridium scindens*, *Clostridiodes difficile*, bioreactor, dual inhibitory mechanism

Introduction

Clostridioides (*Clostridium*) *difficile* is a Gram-positive, endospore forming bacterium producing mild to severe diarrhea leading to pseudomembranous colitis and death (1). In the United States, *C. difficile* causes 450,000 infections and 29,000 deaths annually (2). *C. difficile* infection can be effectively treated by vancomycin and fidaxomicin (3), however, use of these and other antibiotics perturb the intestinal microbial community which may increase the severity of the *C. difficile* infection (4). Furthermore, following antibiotic treatment, 15-30% of *C. difficile* infections relapse into an active infection (3, 5). Fecal-microbiota transplantation has been recommended as an alternative therapy against such relapsing infections with ~90% success rate (6-8). Use of fecal microbiota transplantation is debated among scientists because of the variable bacterial community of the fecal transplant, difference in the recipient genotype (9), potential for transfer of pathogens (10), differences in dosage and administration methods. A better understanding of the colonic microbiota and how fecal microbiota transplantation therapy is effective would be advantageous. Recently, the microbial metabolism of bile acids has been recognized as being an important component of *C. difficile* control in the intestine.

Bile acids are synthesized in the liver from cholesterol and secreted into the proximal small intestine as a component of bile. The primary bile acids in humans are cholic acid (CA; 3α,7α,12α-trihydroxy-5β-cholanic acid) and chenodeoxycholic acid (CDCA; 3α,7α-dihydroxy-5β-cholanic acid) each of which are conjugated to glycine or taurine in the liver at the 3 position. While in the small intestine, bile acids aid in the absorption of lipids and lipid soluble vitamins with approximately 95% reabsorbed by the distal small intestine and

returned to the liver via enterohepatic circulation. Approximately 5% of bile acids escape reabsorption and enter the colon and are exposed to the colonic microflora (11). Many colonic bacterial species are known to deconjugate both the glycine and taurine moieties resulting in free CA and CDCA (12). These free bile acids, in turn, undergo a variety of oxidation-reduction reactions but the most quantitatively important and irreversible reaction is 7α-dehydroxylation of primary bile acids to secondary bile acids (11). This reaction generates (DCA; 3α,12α-dihydroxy-5β-cholanic acid) from CA and lithocholic acid (LCA; 3α-hydroxy-5β-cholanic acid) from CDCA.

Both primary and secondary bile acids are reported to affect *C. difficile* germination and outgrowth. Microscopic examination indicates taurocholic acid (TCA), CA, and DCA facilitate germination of *C. difficile* endospores. CA (12 mM) and DCA (1.9 mM) each inhibit growth of *C. difficile* vegetative cells, but CA at 1.2 mM does not affect cell growth on agar plates (13, 14). These observations have led to improved endospore recovery and germination on selective agar plates for clinical testing (13).

Secondary bile acids have been reported to inhibit *C. difficile*. Studer et al. (15) demonstrated that a combination of 12 murine intestinal isolates, when given to germ free mice, are not protective against a *C. difficile* challenge but addition of *C. scindens,* an intestinal species known to 7α-dehydroxylate bile acids (16), to the microbial mix was protective during the early stages of infection. Using a comprehensive analysis of murine and human intestinal microbiota, Buffie et al. (17) correlated *C. scindens* with inhibition

of *C. difficile* growth. Additionally, these workers used an *in vivo* murine model to show *C. scindens* was protective against a *C. difficile* infection and DCA was also correlated with *C. scindens* and inhibition of *C. difficile*. In contrast to the bile acid theory, Kang et al. (18) observed that *C. difficile* inhibited *C. scindens* growth in culture by secretion of proline-base cyclic peptide antimicrobials and, surprisingly, that at least two strains of bile acid 7α-dehydroxylating bacteria, *C. scindens* ATCC 35704 and *C. sordellii* ATCC 9714, secrete tryptophan-derived antimicrobial peptides which inhibit *C. difficile*. Although, *C. scindens* associated production of DCA in the colon has been shown to provide protection against *C. difficile* (15, 17), high accumulation of DCA is toxic to colonocytes (19). Colonocytes undergo major changes in their DNA and metabolism properties that affects cell division to produce colon cancer (20-25). High levels of DCA can predispose to colon cancer, the third most common cause of cancer death in the US (26, 27).

This study compares the genomes of 5 strains of *C. scindens* and tests their ability to inhibit *C. difficile* R20291, a hypervirulent ribotype 027 strain, in both batch and continuous culture. *C. scindens* ATCC 35704 was originally isolated from feces from a health human subject and shown to convert 17-hydroxylated corticoids to androstans (C19 corticoids) and also exhibits bile acid 7α-dehydroxylation (15). *C. scindens* TH82, *C. scindens* M18, *C. scindens* I10, and *C. scindens* Y1113 were also isolated from feces from a healthy human and possess bile acid 7α-dehydroxylation activity (28, 29). All strains exhibit high levels of bile acid 7α-dehydroxylation activity (30) but otherwise exhibit unique phenotypic or metabolic characteristics (28, 29). Our results show that among these *C. scindens* strains, both bile acid-dependent and bile acid-independent inhibition of *C. difficile* are present.

Materials and Methods

Strains and culture conditions: *C. scindens* TH82, *C. scindens* M18, *C. scindens* Y1113, *C. scindens* I10 were originally obtained from the P.B. Hylemon (Virgina Commonwealth University, Richmond, VA USA). *C. scindens* ATCC 35704 was purchased from American Type Culture Collection (Manassas, VA, USA). *C. difficile* R20291 was obtained from the laboratory Dr. Trevor Lawley (Sanger Institute, UK). Brain Heart Infusion broth modified (BHIM) was prepared as indicated: 37 g/L brain heart extract, 5.0g/L yeast extract, 10.0g/L pectin, 10.0g/L inulin , 0.3 g/L L-cysteine, 0.25 mg/L resazurin, and 1.0 mg/L menadione was added to 900 ml water, pre-reduced and sterilized by autoclaving at 121^{o}C and 15 psi for 30 min. Following autoclaving, 100 ml of filter sterilized 4-morpholinoethanesulfonic acid hydrate (1 M, MES hydrate) was added as the buffering agent and pH was adjusted to 6.8 using 10 M sodium hydroxide. Finally, 1.7ml of acetate (30 mM), 2ml of propionate (8 mM), 2ml butyrate (4mM), 100µl isovalerate (1mM), 1ml hemin (0.5 mg/ml), 10 ml of ATCC vitamin and ATCC mineral mixtures respectively were added per liter of media. All cultures were manipulated using an anaerobic chamber (Coy Lab Products, Grass Lake, MI) containing 5% CO_2, 10% H_2, and 85% N_2 maintained at 37^{o}C.

Genome sequencing assembly, annotation, comparative analysis and secondary metabolite prediction: To generate high quality genomes, strains were sequenced using Illumina short read (Illumina Inc, CA) and the Oxford Nanopore MinION long read (SQL-LSK108, Oxford Nanopore technologies, UK) sequencing technologies. For Illumina sequencing, genomic DNA was isolated from 0.5 ml of overnight culture using E.Z.N.A.®

Bacterial DNA Kit (Omega Biotek, GA). Sequencing library was then prepared using Nextera XT Kit (Illumina Inc, CA). Libraries were sequenced using 250 base paired end chemistry on an Illumina MiSeq platform. FASTQ files were generated using Casava v1.8.2 pipeline (Illumina, Inc, CA).

High molecular weight DNA required for MinION sequencing was extracted following modified phenol: chloroform method as described previously (41). High molecular weight DNA was quantified using Qubit dsDNA HS assay kit (Invitrogen, OR) kit and Qubit® 3.0 Flurometer (Thermo Fisher Scientific Inc., MA). For Oxford Nanopore sequencing, 1.5-2.0 µg of high molecular weight DNA was fragmented by centrifugation using G-tube (Covaris Inc., MA), and the ends were repaired using NEBNext End Repair/dA-Tailing Module followed by ligation of adapters. MinION long read sequencing was performed using the 1-directional (1D) sequencing kit according to manufacturer's protocol (SQL-LSK108, Oxford Nanopore technologies, UK).

To compare the genetic makeup of the *C. scindens* strains, hybrid genomes were generated using short and long reads. Initially, long reads were assembled using miniasm (42). The assembled long reads were then used to scaffold assembly of the short reads using Unicycler v 0.4.0 (43) to produce hybrid assembly. Genomic features were evaluated and validated from hybrid sequences using Quast (44). Total number of tRNAs were predicted using ARAGORN (45). Also, the assembled genome sequences were annotated using Rapid Annotation using Subsystem Technology (46). Amino acid FASTA sequences

obtained from RAST were used to calculate identities between the sequences using two way Amino Acid Identity (AAI) calculator with minimum identity of 20% (47). To determine the diversity and variations of bile acid inducible (*bai*) genes which are responsible for bile acid 7α-dehydroxylation, the hybrid genomes of five *C. scindens* strains were separately examined for the presence of *bai* genes using CLC genomics workbench 9.5.3 (Qiagen Bioinformatics, CA) at 80% identity level. *C. scindens* VPI 12708 *bai* operon sequences were used as reference sequences for *bai* genes comparison [GenBank: U57489.2]. Additionally, these 5 hybrid assemblies of the *C. scindens* strains were annotated using Prokka 1.13 (48). The obtained protein sequences were visualized using Snapgene Viewer 4.2.6 (GSL Biotech LLC) to locate *bai* operon. Genes *bai*A1, *bai*F, *bai*E and *bai*B were annotated by Prokka and remaining genes were manually located using blastp in nrNCBI with default parameters. Amino acid sequences of *bai*A1/3, *bai*A2, *bai*B, *bai*CD, *bai*E, *bai*F, *bai*G, *bai*H, *bai*I and *bai*JKL were separately aligned employing the ClustalO 1.2.4 web server. To identify the presence of *mcbB* gene encoding the production of 1-acetyl-β-carboline in *C. scindens*, know sequence of mcbB gene from other species was used for sequence search against *C. scindens* genomes. mcbB gene from *C. innocuum* 2959 in NCBI database (accession number NZ_KB850948.1) and ten other *Streptococcus* species (*McbB* gene accessions: EFM29959.1, VEI59856.1, VTQ24966.1, CIZ00836.1, EMP71847.1, RGR60169.1, RKV90253.1, RFU54189.1, AXQ78541.1 and RFU54189.1) in UniprotKB database were used for homology search using CLC genomics workbench 9.5.3 (Qiagen Bioinformatics, CA) at 30% sequence identity and 50% coverage.

Batch co-culture of C. scindens strains with C. difficile: To analyze the effect of genetic differences in phenotypic expression related to resistance against *C. difficile,* in vitro batch co-culture was performed. *C. scindens* strains and *C. difficile* R20291 were maintained inside anaerobic chamber in BHIM medium. To examine the effect of bile acids on *C. difficile* inhibition by *C. scindens*, strains were co-cultured in various concentrations of CA and DCA. Concentrations were similar to health human bile acid fecal water concentrations (49). Cultures were grown overnight in BHIM and OD_{600} was adjusted to 0.5 using BHIM. Each *C. scindens* strain was inoculated with *C. difficile* in the ratio 9:1 (vol/vol) in 1 ml of BHIM in triplicate. The mixture was incubated at 37^0C for 24 hours anaerobically then serially diluted and spread on *C. difficile* selective agar to enumerate the *C. difficile* as described before (SQL-LSK108, Oxford Nanopore technologies, UK).

Co-culture of C. scindens strains with C. difficile in continuous flow mini-bioreactors: C. *scindens* strains were co-cultured in BHIM with *C. difficile* in continuous flow model using minibioreactors arrays (MBRAs) as described previously with 8 hour retention time (50). Similar to batch co-culture assays, *C. scindens* and *C. difficile* were mixed at a 9:1 (vol/vol) ratio in duplicate. Co-culture of *C. scindens* and *C. difficile* were incubated CA (50 µg/ml; 116 µM) and BHIM with DCA (50 µg/ml; 121 µM). Samples were collected prior inoculation and 1, 2- and 3-days post inoculation. Samples were enumerated for *C. difficile*. For short chain fatty acid analysis, eight hundred microliters of the bacterial sample from MBRAs were added to 160 µl of freshly prepared 25% (w/v) m-phosphoric acid and frozen at -80^0C. For analysis, samples were thawed and centrifuged (>20,000×g)

for 30 min and 600 µl supernatant was subjected to gas chromatography (Agilent Technologies, USA).

Testing for the production of antimicrobials from C. scindens against C. difficile R20291: *C. scindens* strains were grown for 48 hours in BHIM medium then centrifuged at 10,000 g for 5 min. The supernatant was filtered using a 0.22 µm filter and moved into the anaerobic chamber and diluted in ratio 1:1 with the 1X BHIM medium. pH was adjusted to 6.8. An overnight culture of *C. difficile* R20291 was adjusted to OD_{600} = 0.5. Twenty microliters of OD_{600} adjusted suspension was added to 1 ml of the 1:1 diluted cell free supernatant (heat treated and untreated) and incubated for 24 hours in triplicate. *C. difficile* R20291 was grown in 50% BHIM diluted with anaerobic PBS as a positive control. After 24 hours, the cultures were serial diluted with anaerobic PBS and plated on *Clostridium difficile* selective agar and incubated anaerobically for 24 hours for enumeration. To access the activity of 1- acetyl-β-carboline against *C. difficile* R20291, we grew the *C. difficile* in BHIM supplemented with 25 µg/mL of commercially available 1-acetyl-β-carboline (Aobious Inc, MA) under both heat-treated (65^0C for 20 minutes) and untreated conditions in triplicate. OD_{630} was measured at 0, 24 and 48 hours and analyzed to obtain inhibition of *C. difficile* which was expressed as percentage inhibition compared to the control.

Statistical analysis: Oneway ANOVA was performed in Graphpad Prism v6.0.1. Dunnett test was used to compare mean values of treatments with control ($\alpha = 0.05$).

Results

Genomic analysis of C. scindens strains: To gain insights into the genomic diversity that might impact the *C. difficile* inhibition capacity of *C. scindens*, genomes of five strains known to exhibit bile acid 7α-dehydroxylation activity, were sequenced. Each genome was sequenced using Oxford Nanopore long read and Illumina short read platforms. Hybrid genome assembly produced complete genomes for *C. scindens* I10, *C. scindens* M18, while near complete genomes were produced for *C. scindens* ATCC 35704, *C. scindens* TH82 and *C. scindens* Y1113. (Table 1). The *bai* genes, responsible for bile acid 7α-dehydroxylation, were distributed in two different chromosomal regions of the *C. scindens* genomes, the first being a larger complete *bai* operon consisting of *bai*B, *bai*CD, *bai*E, *bai*A2, *bai*F, *bai*G, *bai*H and *bai*I, while second operon consisted of *bai*A1/3, *bai*J, *bai*K and *bai*L. Both larger and smaller operons were present in *C. scindens* M18 and *C. scindens* Y1113 while only *bai*A1/3 was present for *C. scindens* ATCC 35704, *C. scindens* I10 and *C. scindens* TH82 (Figure 1A). The *bai*JKL operon was not present in *C. scindens* ATCC 35704, *C. scindens* I10 or *C. scindens* TH82. In addition, a second copy of *bai*A2 was present in *C. scindens* ATCC 35704, *C. scindens* TH82 and *C. scindens* Y1113 but absent in *C. scindens* M18 and *C. scindens* I10. The *bai*A1/3 and *bai*JKL operons were separated in *C. scindens* M18 and *C. scindens* Y1113 by two genes, crtK-2-TspO (tryptophan rich sensory protein) and *hcaR* (LysR family transcriptional regulator) (Figure 1B). Unlike *C. scindens* VPI 12708, *C. scindens* M18 and *C. scindens* Y1113 do not possess a LysR type transcription regulator between the *bai*JKL operon and the *bai*A1 gene. Furthermore, as upstream and downstream genes could potentially regulate gene expression, we checked for differences in upstream and downstream genes in both operons. Among the 5 strains,

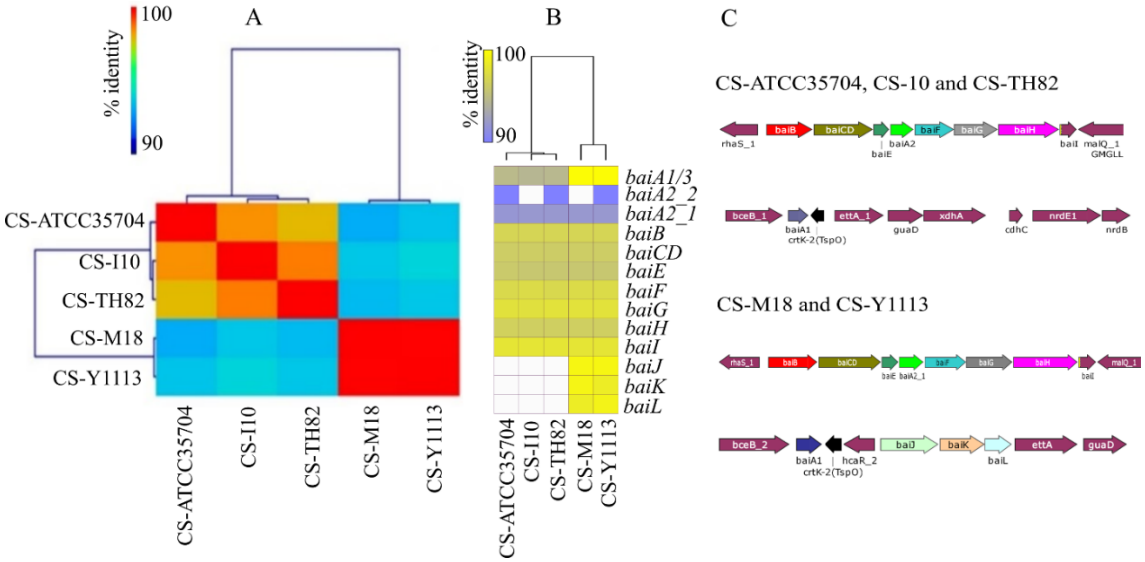


Figure 1. Genomic differences among the *Clostridium scindens* (CS) strains. A) Average amino acid identity of CS strains. Average amino acid identity (AAI) matrix was obtained using amino acid identity calculator and visualized using MeV 4.9.0. B) Presence of bile acid inducible (bai) genes in bai operon in different *Clostridium scindens* (CS) strains. *Clostridium scindens* VPI 12708 *bai* genes were used for comparison and % identity of each gene was represented as a heat map. C) Two groups of *Clostridium scindens* were formed based on the presence of *bai*JKL genes. The presence of *bai*JKL and *hca*R genes in *Clostridium scindens* -M18 and *Clostridium scindens* Y1113 are shown compared to *Clostridium scindens* ATCC 35704, CS-I10 and CS-TH82. Two separate rows of genes represent separate chromosomal regions in each individual genome.

no differences in the large operon were observed up to 8 upstream and downstream genes. For the second operon, genes in upstream of the *bai*A1 differed from -5 position whereas an additional hypothetical protein (winged helix DNA binding domain protein) was detected in downstream 6th position for only *C. scindens* M18 and *C. scindens* Y1113.

Doerner et al. (30) has demonstrated the differences in the dehydroxylation of CA to DCA. Among these five strains, *C. scindens* Y1113 was observed to be the highest dehydroxylating bacteria followed by *C. scindens* I10, *C. scindens* M18, *C. scindens* ATCC 34704 and *C. scindens* TH82. Such differences could be attributed to the presence of the *bai* genes in the *bai* operons. Even though all strains exhibited the presence of larger operon *bai*BCDEA2FGHI in this study, the difference was observed for *bai*JKL which might contribute to the differences in bile acid conversion. On other hand, the *bai*A1/3 was also present in all the strains but variation in the amino acid composition was observed. The *bai*A1/3 gene of *C. scindens* ATCC 35704, *C. scindens* I10 and *C. scindens* TH82 was determined to have 9 amino acid differences compared to *C. scindens* M18 and *C. scindens* Y1113 (Supp. Figure 1). The differences were in positions 25, 99, 111,118, 133, 204, 209, 210 and 235. Position 111 represented a change in histidine (H), present in *C. scindens* ATCC 35704, *C. scindens* I10 and *C. scindens* TH82, to asparagine (N) in *C. scindens* M18 and *C. scindens* Y1113. At position 111, the positively charged side chain of histidine, present in *C. scindens* ATCC 35704, *C. scindens* I10 and *C. scindens* TH82, was altered to a polar uncharged side chain of asparagine in *C. scindens* M18 and *C. scindens* Y1113. Similarly, at position 118, isoleucine, present in *C. scindens* ATCC 35704, *C. scindens* I10

```
CS-ATCC       MKLVQDKITIITGGTRGIGFAAAKIFIENGAKVSIFGETQEEVDTALAQLKELYPEEEVL
       60
CS-I10        MKLVQDKITIITGGTRGIGFAAAKIFIENGAKVSIFGETQEEVDTALAQLKELYPEEEVL
       60
CS-TH82       MKLVQDKITIITGGTRGIGFAAAKIFIENGAKVSIFGETQEEVDTALAQLKELYPEEEVL
       60
CS-M18        MKLVQDKITIITGGTRGIGFAAAKLFIENGAKVSIFGETQEEVDTALAQLKELYPEEEVL
       60
CS-Y1113      MKLVQDKITIITGGTRGIGFAAAKLFIENGAKVSIFGETQEEVDTALAQLKELYPEEEVL
       60
              ************************:***********************************

CS-CSATCC     GFAPDLTSRDAVMAAVGTVAQKYGRLDVMINNAGITMNTVFSRVSEEDFKHIMDINVIGV
      120
CS-I10        GFAPDLTSRDAVMAAVGTVAQKYGRLDVMINNAGITMNTVFSRVSEEDFKHIMDINVIGV
      120
CS-TH82       GFAPDLTSRDAVMAAVGTVAQKYGRLDVMINNAGITMNTVFSRVSEEDFKHIMDINVIGV
      120
CS-M18        GFAPDLTSRDAVMAAVGTVAQKYGRLDVMINNAGITMNSVFSRVSEEDFKNIMDINVNGV
      120
CS-Y1113      GFAPDLTSRDAVMAAVGTVAQKYGRLDVMINNAGITMNSVFSRVSEEDFKNIMDINVNGV
      120
              ************************************:***********:****** **

CS-ATCC       FNGAWSAYQCMKEAKQGVIINTASVTGIYGSLSGIGYPTSKAGVIGLTHGLGREIIRKNI
      180
CS-I10        FNGAWSAYQCMKEAKQGVIINTASVTGIYGSLSGIGYPTSKAGVIGLTHGLGREIIRKNI
      180
CS-TH82       FNGAWSAYQCMKEAKQGVIINTASVTGIYGSLSGIGYPTSKAGVIGLTHGLGREIIRKNI
      180
CS-M18        FNGAWSAYQCMKDAKQGVIINTASVTGIYGSLSGIGYPTSKAGVIGLTHGLGREIIRKNI
      180
CS-Y1113      FNGAWSAYQCMKDAKQGVIINTASVTGIYGSLSGIGYPTSKAGVIGLTHGLGREIIRKNI
      180
              ************:***********************************************

CS-ATCC       RVVGVAPGVVDTDMTKGLPPEILDDYLKSFPMKRMLKPEEIANVYLFLASDLANGITATT
      240
CS-I10        RVVGVAPGVVDTDMTKGLPPEILDDYLKSFPMKRMLKPEEIANVYLFLASDLANGITATT
      240
CS-TH82       RVVGVAPGVVDTDMTKGLPPEILDDYLKSFPMKRMLKPEEIANVYLFLASDLANGITATT
      240
CS-M18        RVVGVAPGVVDTDMTKGLPPEILEDYLKTLPMKRMLKPEEIANVYLFLASDLASGITATT
      240
CS-Y1113      RVVGVAPGVVDTDMTKGLPPEILEDYLKTLPMKRMLKPEEIANVYLFLASDLASGITATT
      240
              ***********************:****::***********************.******

CS-ATCC       ISVDGAYRP*   249
CS-I10        ISVDGAYRP*   249
CS-TH82       ISVDGAYRP*   249
CS-M18        ISVDGAYRP*   249
CS-Y1113      ISVDGAYRP*   249
```

Supplementary Figure 1: Sequence alignment of amino acids of *bai*A1 gene for five *C. scindens* strains using ClustalO. The red fonts show the amino acids and their positions variation.

Table 1. Genomic features of *Clostridium scindens* strains compared in this study.

Features	ATCC 35704	I10	M18	TH82	Y1113
Total contigs	3	1	1	141	10
Largest contig (bp)	3,536,561	3,435,579	4,020,163	298,262	1,165,028
Total length (bp)	3,635,563	3,435,579	4,020,163	3,321,200	3,937,901
GC (%)	46.23	46.36	47.38	46.66	47.42
N50	3,536,561	3,435,579	4,020,163	61,248	897,417
L50	1	1	1	16	2
Orthologous protein clusters	3179	3165	3690	3077	3667
Coding sequence	3822	3548	3819	3372	3721
RNAs	65	64	64	52	60
tRNAs	58	57	56	51	55

and *C. scindens* TH82, was substantially different than the hydrophobic side chain of asparagine present in *C. scindens* M18 and *C. scindens* Y1113 (Supp. Figure 1).

Kang et al. (18) recently reported that *C. scindens* ATCC 35704 produce 1-acetyl-β-carboline, which has an inhibitory effect against *C. difficile*. However, this study did not identify the gene responsible for the production of 1-acetyl-β-carboline in *C. scindens*. In other organisms, *mcbB* mediates the Pictet Spengler reaction of L-tryptophan and oxaloacetaldehyde to produce carbolines (31). To determine whether the *mcbB* homolog is present in *C. scindens*, we searched the genomes of all strains sequenced in this study using all known *mcbB* sequences from *C. innocuum* and *Streptococcus* species. This search did not find any *mcbB* homolog in *C. scindens* when 30% sequence identity and 50% coverage was used as search cut off.

***Phenotypic analysis of C. scindens strains*:** All five *C. scindens* strains were tested for inhibition of growth of *C. difficile* R20291 using batch co-cultures with varying concentrations of bile acids (Figure 2). When grown in the absence of bile acid, all *C. scindens* strains significantly reduced growth of *C. difficile* R20291. Reduction of growth ranged from 8.3±0.02 $\log_{10}$ CFU of *C. difficile* R20291 grown in monoculture (control) to 7.8±0.08 $\log_{10}$ when grown with *C. scindens* TH82 or 5.8±0.06 $\log_{10}$ when grown with *C. scindens* M18. Additionally, when co-cultured with *C. scindens* TH82 or *C. scindens* Y1113, CFUs of *C. difficile* R20291 were reduced in a dose-dependent manner when concentrations of 12.5 – 50 μg/ml of CA were included in the medium. Addition of CA did

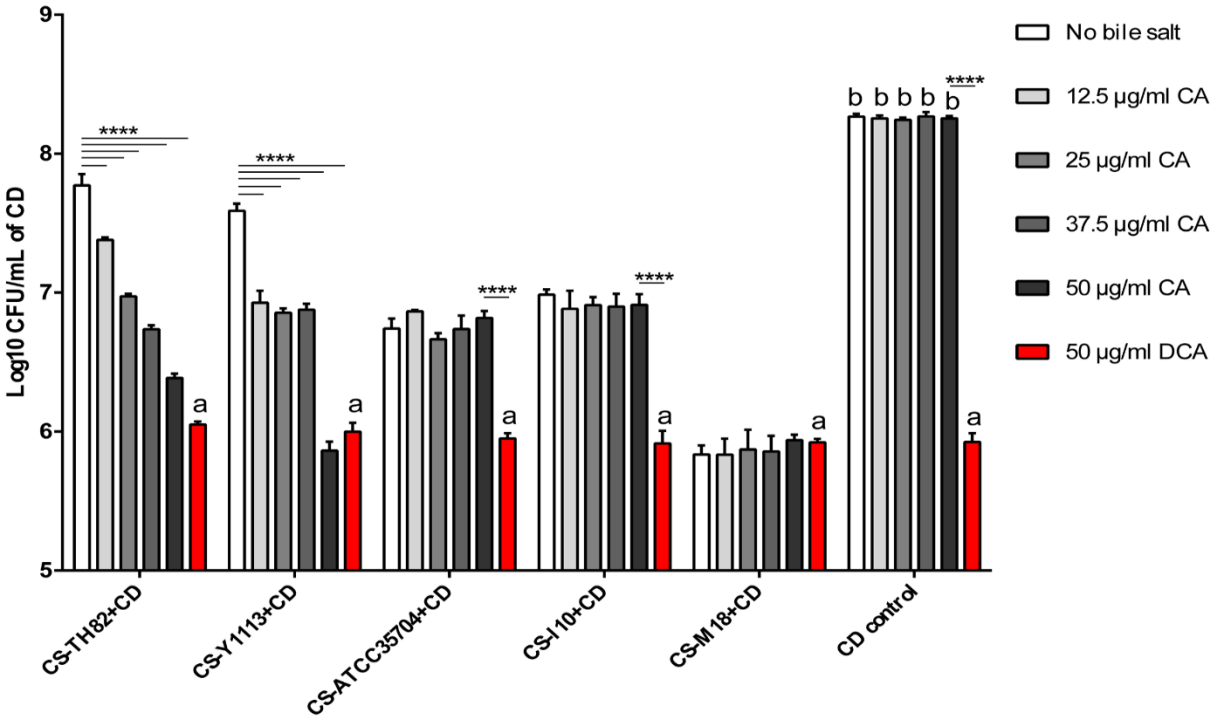


Figure 2. Growth of *Clostridioides* (*Clostridium*) *difficile* (CD) when tested using batch co-culture with *Clostridium scindens* (CS) strains in the presence and absence of cholic acid (CA) and deoxycholic acid (DCA). *Clostridium difficile* monoculture and all CS cocultures in the presence of DCA are shown in red; 'a' indicates statistical significance ($p<0.05$) when compared to the corresponding control culture without DCA. *C. difficile* in monoculture, with and without CA, is indicated by 'b'; no significant difference of *C. difficile* CFUs were observed. '****' represents statistical significance at $p < 0.001$.

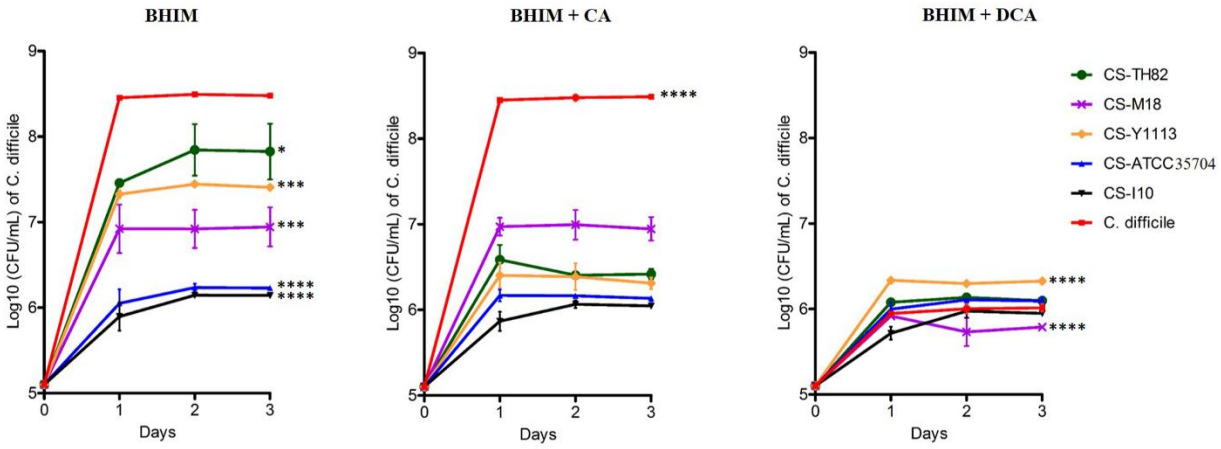


Figure 3. Levels of *C. difficile* R20291 when cocultured in continuous flow minibioreactor with *C. scindens* (CS) strains in the presence of cholic acid (CA) and deoxycholic acid (DCA). *C. difficile* was determined by counting colony forming units (CFU) every 24 h for 3 days. A) BHIM medium. B) BHIM containing 50 µg/ml of CA. C) BHIM medium containing 50 µg/ml of DCA. Statistical analysis was performed by comparing *C. difficile* CFUs in pure culture to *C. difficile* CFUs in coculture. Error bars represent standard error of mean of two experiment performed in duplicate. One way ANOVA and tukey test was performed at the 3rd day between the groups where '*', '**', '***', and '****' represent statistical significance at $p = 0.05$, $p = 0.01$, $p = 0.001$ and $p < 0.001$ respectively.

not increase the growth inhibition of *C. difficile* R20291 by *C. scindens* M18, *C. scindens* ATCC 35704 or *C. scindens* I10. Addition of DCA (50 μg/ml) reduced the growth of *C. difficile* R20291 from 8.2±0.02 $\log_{10}$ to 5.98 ±0.06 $\log_{10}$. None of the *C. scindens* tested further reduced growth of *C. difficile* R20291 cultured in the presence of DCA when compared to *C. difficile* R20291 grown in monoculture in the presence of DCA (Figure 2).

C. scindens M18, *C. scindens* ATCC 35704, *C. scindens* Y1113, *C. scindens* TH82 and *C. scindens* I10 were also tested for inhibition of *C. difficile* R20291 grown in continuous flow culture using a mini-bioreactor array. Similar to batch co-culture, all 5 *C. scindens* strain inhibited *C. difficile* R20291 (Figure 3A) with statistical significance ($p<0.05$). *C. scindens* ATCC 35704 and *C. scindens* I10 were the most inhibitory, each decreasing *C. difficile* R20291 growth by 2.3 $\log_{10}$. Addition of CA (50 μg/ml) substantially increased *C. difficile* R20291 inhibition by *C. scindens* TH82 and *C. scindens* Y1113 but did not affect the inhibition of the other three *C. scindens* strains (Figure 3B). DCA (50 μg/ml) effectively inhibited growth of *C. difficile* R20291 in continuous culture reducing control levels of 8.5±0.01 $\log_{10}$ CFU to 6.0 ±0.04 $\log_{10}$ CFU. Furthermore, addition of *C. scindens* strains to *C. difficile* R20291 cultured in DCA did not significantly reduce *C. difficile* R20291 levels with the exception of *C. scindens* M18 which reduced growth from 6.0 ±0.04 $\log_{10}$ to 5.8±0.04 $\log_{10}$ (Figure 3C).

Steady-state production of the short chain fatty acids, acetate, propionate, and butyrate was determined in continuous culture of the five *C. scindens* strains with *C. difficile* R20291

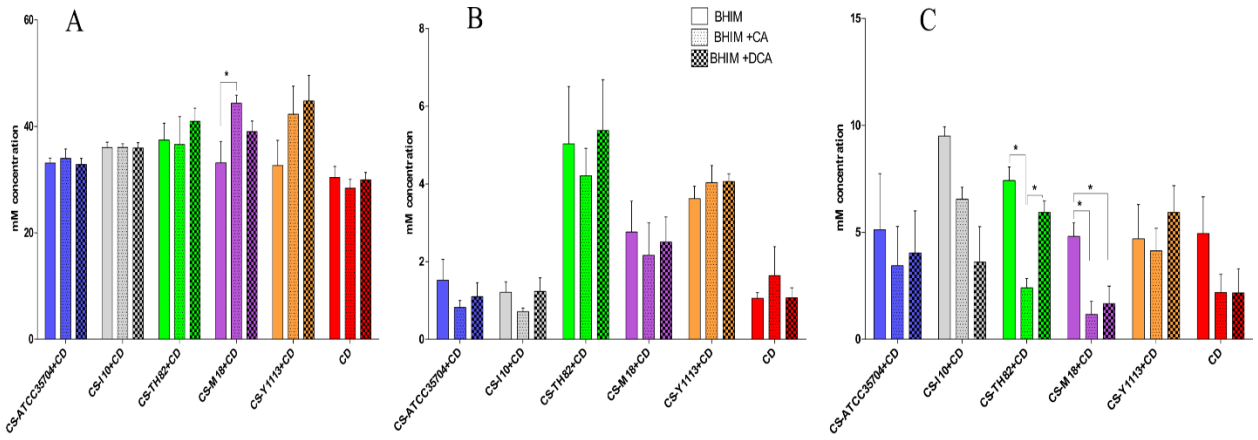


Figure 4. Concentrations of short chain fatty acids (SCFAs) obtained from coculture of *C. scindens* (CS) strains with *C. difficile* R20291 in continuous flow minibioreactor in BHIM medium with cholic acid (CA, 50 µg/ml) and deoxycholic acid (DCA, 50 µg/ml) in duplicate in two experiments. A, B and C represent concentrations of acetate, propionate and butyrate on day 3, respectively. Statistical comparison using Oneway ANOVA. '*' represents statistical significance at p=0.05.

(Figure 4). Significant differences were not observed when CA or DCA was added to the culture medium for most strains. Acetate values ranged from 28 to 45 mM. Values ranged from 0.7 to 5.4 mM for propionate and 1.2 to 9.5 mM for butyrate. However, the *C. scindens* M18 and *C. difficile* R20291 coculture provided two notable expectations. First, this coculture displayed a significant increase in acetate from 33.2 to 44.4 mM upon the addition of CA and, second, displayed a reduction of butyrate from 4.8 mM (control) to 1.2 and 1.7 mM with addition CA and DCA, respectively. Also, cocultures of *C. scindens* TH82 and *C. difficile* R20291 indicated a decrease of butyrate production of 7.4 mM (control) to 2.4 mM with addition of CA but addition of DCA did not significantly inhibit.

Each *C. scindens* strain was tested for extracellular material which could inhibit growth of *C. difficile* R20291 (Figure 5A). Culture supernatant was collected from each strain and diluted 1X with fresh medium then used as growth medium for *C. difficile* R20291. Data indicate that each strain was inhibitory; *C. scindens* ATCC 35704 was the most inhibitory and reduced *C. difficile* R20291 growth from 7.7±0.2 $\log_{10}$ to 6.9±0.04 $\log_{10}$ while *C. scindens* TH82 was the least inhibitory and reduced growth to 7.2±0.01 $\log_{10}$. As shown in Figure 5A, the inhibition was significantly reduced, but not abolished when culture supernatant was heat treated. To determine whether heat treatment affects the inhibitory capacity of 1-acetyl-β-carboline on *C. difficile*, heat treated and untreated commercially available 1-acetyl-β-carboline was tested against *C. difficile* R20291 in identical conditions. Data in Figure 5B Shows that heat treatment did not affect the inhibitory capacity of 1-acetyl-β-carboline. Taken together, our results indicate that factors other than

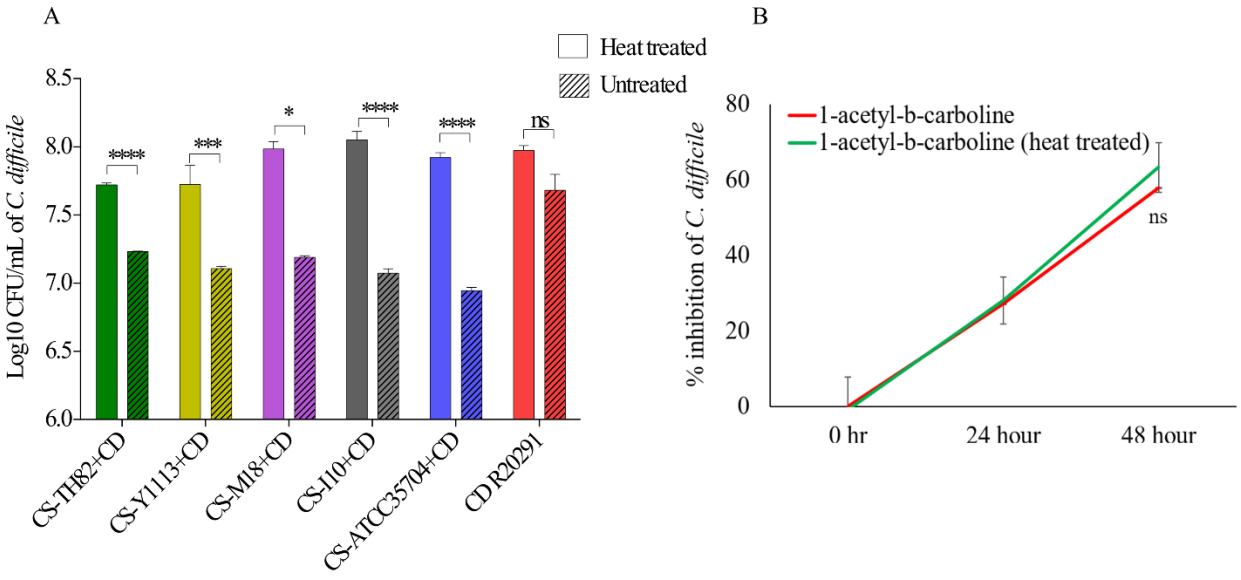


Figure 5. (A) Inhibition of *C. difficile* R20291 (CD) by the cell free culture supernatant obtained from *C. scindens* (CS) strains cultured in BHIM medium. (B) Effect of heating on the *C. difficile* R20291 inhibition capacity of 1- acetyl-β-carboline. One-way ANOVA was used for statistical analysis and each mean value was compared to CD. "***" and "****" represents p-values at 0.001 and < 0.0001.

1-acetyl-β-carboline are responsible for the bile acid independent inhibition of *C. difficile* R20291 by the *C. scindens* strains used in this study.

Discussion

Data presented here indicate that all five strains of *C. scindens* inhibit the growth of *C. difficile* R20291 in batch and continuous-flow culture using two different mechanisms of inhibition. First, *C. scindens* M18, *C. scindens* ATCC 35704, and *C. scindens* I10 all inhibit *C. difficile* R20291 in the absence of CA; the presence of CA does not increase inhibition. *C. scindens* M18 was most effective at inhibiting *C. difficile* growth in batch culture, causing up to 2.4 $\log_{10}$ reduction (270-fold) of *C. difficile* R20291 growth. In continuous-flow culture, *C. scindens* I10 and *C. scindens* ATCC 35704 were more effective, inhibiting *C. difficile* R20291growth approximately 2.3 $\log_{10}$. Thus, the first mechanism of inhibition by *C. scindens* is bile acid independent. Second, *C. scindens* TH82 and *C. scindens* Y1113 inhibited *C. difficile* R20291 in the absence of CA but showed a dose-dependent increase of inhibition with addition of CA in batch culture and a CA-dependent increase in inhibition in continuous-flow culture. In total, these results indicate that *C. scindens* TH82 and *C. scindens* Y1113 inhibit *C. difficile* R20291 by both a bile acid independent and bile acid dependent mechanisms and *C. scindens* ATCC 35704, *C. scindens* I10, and *C. scindens* M18 inhibit *C. difficile* R20291 only by a bile acid independent mechanism. The bile acid independent mechanism of *C. scindens* TH82 and *C. scindens* Y1113 is less effective at inhibiting *C. difficile* R20291 than the bile acid independent mechanisms of the other three strains.

These two strains, in the absence of CA, were less effective than *C. scindens* M18, *C. scindens* ATCC 35704, or *C. scindens* I10 but achieved similar inhibitory effects with the addition of CA or DCA at 50 ug/ml. The inhibitory action is dependent upon the bile acid CA and likely not due to CA directly, but due to the conversion of CA to DCA which is a more hydrophobic than CA and acts as detergent to disrupt cellular membranes and inhibit cell growth.

Kang et al. (18) have proposed a model for both bile acid dependent and bile acid independent inhibition of *C. difficile*. These workers isolated three compounds from *C. scindens*; turbomycin A and 1,1,1-tris(3-indolyl)-methane from *C. scindens* ATCC 9714 and 1-acetyl-β-carboline from *C. scindens* ATCC 35704. Further experiments indicated that turbomycin A and 1-acetyl-β-carboline effectively inhibited *C. difficile* ATCC R9689 and the addition of DCA or LCA, but not CA, significantly increased this inhibition. The authors propose the combination of bacteriostatic compounds produced by bile acid 7α-dehydroxylating bacteria coupled with the detergent-like properties of hydrophobic secondary bile acids effectively inhibit *C. difficile in vivo.*

Data presented here are generally consistent with this model, but inconsistencies remain. Kang et al. (18) show that *C. scindens* ATCC 35704 does not inhibit *C. difficile* ATCC R9689, in spite of the fact *C. scindens* ATCC 35704 is producing 1-acetyl-β-carboline. Addition of CA to the co-culture also did not inhibit *C. difficile* ATCC R9689. Addition

of *C. scindens* VPI 12708 to the co-culture of *C. scindens* ATCC 35704 and *C. difficile* ATCC R9689 did inhibit the growth of *C. difficile* ATCC R9689. However, it is unclear what factors *C. scindens* VPI 12708 contributed, as control co-culture experiments with *C. scindens* VPI 12708 and *C. difficile* ATCC R9689 were not reported. *C. scindens* VPI 12708 may have acted alone or perhaps in synergy with *C. scindens* ATCC 35704. Our data indicate that a growing *C. scindens* ATCC 35704 culture or culture supernatant can effectively inhibit *C. difficile* which is consistent with this strain producing bacteriostatic compounds. However, the heat treatment of culture supernatant reduced the inhibitory effect while heat treated 1-acetyl-β-carboline did not show any reduction in inhibition. These data suggest that a protein mediated mechanism is present in addition to 1-acetyl-β-carboline mediated inhibition of *C. difficile* R20291 by *C. scindens* ATCC 35704 in our experiments. Homology based searches using known mcbB genes responsible for the production of 1-acetyl-β-carboline in other bacterial species did not identify any homologs in all *C. scindens* strains used in this study. Clostridial genomes in general exhibit high sequence variation (32). This high sequence divergence may be the reason why homology-based searches failed to identify 1-acetyl-β-carboline encoding gene in our searches.

In a recent study, freshly isolated, clinical *C. difficile* strains were administered to conventional mice then disease severity was measured and correlated to virulence factors and genomic features (33). The results indicated that pathogenic *C. difficile* strains produce a range of disease severities which are not necessarily correlated with known virulence factors. For example, *C. difficile* intestinal burden, toxin production, sporulation rate, and germination efficiency were not associated with disease severity but displayed

wide variations across the strains tested. Tolerance to the secondary bile acid lithocholic acid was correlated with disease severity but also displayed wide variation. The *C. difficile* RT027 genotype, which represented 6 of the 21 strains tested, was correlated with the greatest disease severity and mortality, except for one RT027 strain which was the least pathogenic of all 21 strains. The workers went on to show that no fewer than nine groups of genes, other than the previously recognized pathogenicity loci, were found to correlate with disease severity. Burns et al. (34) have shown that significant variation exists in the rate of sporulation of *C. difficile* strains and Heeg et al. (35) have shown that significant variation exists in germination of *C. difficile* endospores in response to bile acids and other factors. These data show that virulence factors and genetic variation among *C. difficile* strains is substantial which suggests sensitivity of *C. difficile* strains toward bacteriostatic factors may vary. While in general agreement, our data differ from other studies (18) which could be due to differences between the *C. difficile* strains employed. This study used *C. difficile* R20291, a 'hypervirulent' RT027 strain isolated from a recent disease outbreak in the United Kingdom (36), whereas Kang et al. (18) used the 'historic' *C. difficile* R9689 (37). In batch culture experiments, Kang et al. (18) observed *C. difficile* R9689 was not sensitive to *C. scindens* ATCC 35704 in the presence of CA indicating *C. difficile* R9689 may not be sensitive to 1-acetyl-β-carboline at the levels produced by *C. scindens* ATCC 35704. In this study, *C. scindens* ATCC 35704 effectively inhibited *C. difficile* R20291 (Figure 2) suggesting *C. difficile* R20291 is sensitive to the bacteriostatic compound, 1-acetyl-β-carboline, at the levels produced by *C. scindens* ATCC 35704 (18). Our data suggest *C. difficile* R20291 has a varied response to CA. We show that *C. difficile* R20291 cultured with *C. scindens* TH82 or Y1113 display a dose-dependent response to

CA however when *C. difficile* R20291 is cultured with *C. scindens* ATCC 35704, *C. scindens* I10, or *C. scindens* M18, CA does not have an effect. This suggests that *C. difficile* R20291 may be differentially inhibited by CA in the presence various *C. scindens* strains possibly due to the bacteriocins produced by the *C. scindens* strains. In sum, these observations indicate that phenotypic and genetic variation of *C. difficile* strains used to study the inhibitory effects of *C. scindens* may significantly influence the results.

Conclusion

Hence, the data presented here indicate that *C. scindens* effectively inhibits *C. difficile* in both batch and continuous culture. Growth inhibition of *C. difficile* by some *C. scindens* strains proceed by a bile acid dependent mechanism, as evidenced by the dose dependent inhibition by CA in the presence of some *C. scindens* strains. All *C. scindens* tested also inhibit *C. difficile* in a bile acid independent manner, as indicated by the observation spent*C. scindens* culture supernatant inhibited *C. difficile.* Since antibiotic treatment often causes recurrence of *C. difficile* infection, identifying resident members of the commensal gut microbiota that inhibit *C. difficile* growth and provide colonization resistance has been proposed as a means of developing defined bacteriotherapy against *C. difficile* infection. *C. scindens* is a member of gut commensal that has shown high effectiveness in inhibiting *C. difficile*. However, focusing on using secondary bile acid production of *C. scindens* alone as a means of developing bacteriotherapy may have long-term health complications for patients. For example, high activity bile acid dehydroxylating strains can induce colorectal cancer by producing DCA in colon which induce colonal stem cells in colonic epithelium to develop colon cancer (38-40). Therefore, utilizing *C. scindens* strains which

inhibit *C. difficile* through bile-independent mechanisms could be an effective means of developing defined bacteriotherapy against *C. difficile* without the long-term risk of colon cancer causation.

Literature Cited

1. Khan FY, Elzouki AN. 2014. Clostridium difficile infection: a review of the literature. Asian Pac J Trop Med 7S1:S6-S13.
2. Lessa FC, Winston LG, McDonald LC. 2015. Burden of Clostridium difficile infection in the United States. N Engl J Med 372:2369-70.
3. Sirbu BD, Soriano MM, Manzo C, Lum J, Gerding DN, Johnson S. 2017. Vancomycin Taper and Pulse Regimen With Careful Follow-up for Patients With Recurrent Clostridium difficile Infection. Clin Infect Dis 65:1396-1399.
4. Bakken JS, Borody T, Brandt LJ, Brill JV, Demarco DC, Franzos MA, Kelly C, Khoruts A, Louie T, Martinelli LP, Moore TA, Russell G, Surawicz C, Fecal Microbiota Transplantation W. 2011. Treating Clostridium difficile infection with fecal microbiota transplantation. Clin Gastroenterol Hepatol 9:1044-9.
5. Johnson S. 2009. Recurrent Clostridium difficile infection: a review of risk factors, treatments, and outcomes. J Infect 58:403-10.
6. Gough E, Shaikh H, Manges AR. 2011. Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent Clostridium difficile infection. Clin Infect Dis 53:994-1002.
7. Bakken JS. 2009. Fecal bacteriotherapy for recurrent Clostridium difficile infection. Anaerobe 15:285-9.
8. Kassam Z, Lee CH, Yuan Y, Hunt RH. 2013. Fecal microbiota transplantation for Clostridium difficile infection: systematic review and meta-analysis. Am J Gastroenterol 108:500-8.

9. Pamer EG. 2014. Fecal microbiota transplantation: effectiveness, complexities, and lingering concerns. Mucosal Immunol 7:210-4.

10. FDA. 2019. FDA In Brief: FDA warns about potential risk of serious infections caused by multi-drug resistant organisms related to the investigational use of Fecal Microbiota for Transplantation. https://www.fda.gov/news-events/fda-brief/fda-brief-fda-warns-about-potential-risk-serious-infections-caused-multi-drug-resistant-organisms. Accessed

11. Vlahcevic ZRD, Heuman DM, Hylemon PB. 1996. Physiology and pathophysiology of enterhepatic circulation of bile acids. , p 376-417. *In* Zakim D, Boyer T (ed), Hepatology: a textbook of liver disease, 3 ed, vol 1. W.B. Saunders, Philadelphia.

12. Hayakawa S. 1973. Microbiological transformation of bile acids. Adv Lipid Res 11:143-92.

13. Wilson KH. 1983. Efficiency of various bile salt preparations for stimulation of Clostridium difficile spore germination. J Clin Microbiol 18:1017-9.

14. Wilson KH, Kennedy MJ, Fekety FR. 1982. Use of sodium taurocholate to enhance spore recovery on a medium selective for Clostridium difficile. J Clin Microbiol 15:443-6.

15. Studer N, Desharnais L, Beutler M, Brugiroux S, Terrazos MA, Menin L, Schurch CM, McCoy KD, Kuehne SA, Minton NP, Stecher B, Bernier-Latmani R, Hapfelmeier S. 2016. Functional Intestinal Bile Acid 7alpha-Dehydroxylation by Clostridium scindens Associated with Protection from Clostridium difficile Infection in a Gnotobiotic Mouse Model. Front Cell Infect Microbiol 6:191.

16. Morris GN, Winter J, Cato EP, Ritchie AE, Bokkenheuser VD. 1985. Clostridium scindens sp. nov., a Human Intestinal Bacterium with Desmolytic Activity on Corticoids. International Journal of Systematic and Evolutionary Microbiology 35:478-481.

17. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, Littmann E, van den Brink MR, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint NC, Xavier JB, Pamer EG. 2015. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature 517:205-8.

18. Kang JD, Myers CJ, Harris SC, Kakiyama G, Lee IK, Yun BS, Matsuzaki K, Furukawa M, Min HK, Bajaj JS, Zhou H, Hylemon PB. 2019. Bile Acid 7alpha-Dehydroxylating Gut Bacteria Secrete Antibiotics that Inhibit Clostridium difficile: Role of Secondary Bile Acids. Cell Chem Biol 26:27-34 e4.

19. Bernstein H, Bernstein C, Payne CM, Dvorak K. 2009. Bile acids as endogenous etiologic agents in gastrointestinal cancer. World J Gastroenterol 15:3329-40.

20. Ha YH, Park DG. 2010. Effects of DCA on Cell Cycle Proteins in Colonocytes. J Korean Soc Coloproctol 26:254-9.

21. Narisawa T, Magadia NE, Weisburger JH, Wynder EL. 1974. Promoting effect of bile acids on colon carcinogenesis after intrarectal instillation of N-methyl-N'-nitro-N-nitrosoguanidine in rats. J Natl Cancer Inst 53:1093-7.

22. Hill MJ. 1990. Bile flow and colon cancer. Mutat Res 238:313-20.

23. Zeng H, Botnen JH, Briske-Anderson M. 2010. Deoxycholic acid and selenium metabolite methylselenol exert common and distinct effects on cell cycle,

apoptosis, and MAP kinase pathway in HCT116 human colon cancer cells. Nutr Cancer 62:85-92.

24. Farhana L, Nangia-Makker P, Arbit E, Shango K, Sarkar S, Mahmud H, Hadden T, Yu Y, Majumdar AP. 2016. Bile acid: a potential inducer of colon cancer stem cells. Stem Cell Res Ther 7:181.

25. Bernstein C, Holubec H, Bhattacharyya AK, Nguyen H, Payne CM, Zaitlin B, Bernstein H. 2011. Carcinogenicity of deoxycholate, a secondary bile acid. Arch Toxicol 85:863-71.

26. Siegel RL, Miller KD, Jemal A. 2016. Cancer statistics, 2016. CA Cancer J Clin 66:7-30.

27. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, Stein KD, Alteri R, Jemal A. 2016. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin 66:271-89.

28. Kitahara M, Takamine F, Imamura T, Benno Y. 2000. Assignment of Eubacterium sp. VPI 12708 and related strains with high bile acid 7alpha-dehydroxylating activity to Clostridium scindens and proposal of Clostridium hylemonae sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 50 Pt 3:971-8.

29. Takamine F, Imamura T. 1995. Isolation and characterization of bile acid 7-dehydroxylating bacteria from human feces. Microbiol Immunol 39:11-8.

30. Doerner KC, Takamine F, LaVoie CP, Mallonee DH, Hylemon PB. 1997. Assessment of fecal bacteria with bile acid 7 alpha-dehydroxylating activity for the presence of bai-like genes. Appl Environ Microbiol 63:1185-8.

31. Mori T, Hoshino S, Sahashi S, Wakimoto T, Matsui T, Morita H, Abe I. 2015. Structural Basis for beta-Carboline Alkaloid Production by the Microbial Homodimeric Enzyme McbB. Chem Biol 22:898-906.

32. Paredes-Sabja D, Shen A, Sorg JA. 2014. Clostridium difficile spore biology: sporulation, germination, and spore structural proteins. Trends Microbiol 22:406-16.

33. Lewis BB, Carter RA, Ling L, Leiner I, Taur Y, Kamboj M, Dubberke ER, Xavier J, Pamer EG. 2017. Pathogenicity Locus, Core Genome, and Accessory Gene Contributions to Clostridium difficile Virulence. MBio 8.

34. Burns DA, Heeg D, Cartman ST, Minton NP. 2011. Reconsidering the sporulation characteristics of hypervirulent Clostridium difficile BI/NAP1/027. PloS one 6:e24894-e24894.

35. Heeg D, Burns DA, Cartman ST, Minton NP. 2012. Spores of Clostridium difficile clinical isolates display a diverse germination response to bile salts. PloS one 7:e32381.

36. Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, Lawley TD, Sebaihia M, Quail MA, Rose G, Gerding DN, Gibert M, Popoff MR, Parkhill J, Dougan G, Wren BW. 2009. Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol 10:R102.

37. Rolfe RD, Finegold SM. 1979. Purification and characterization of Clostridium difficile toxin. Infect Immun 25:191-201.

38. Ajouz H, Mukherji D, Shamseddine A. 2014. Secondary bile acids: an underrecognized cause of colon cancer. World J Surg Oncol 12:164.

39. Cheng K, Raufman JP. 2005. Bile acid-induced proliferation of a human colon cancer cell line is mediated by transactivation of epidermal growth factor receptors. Biochem Pharmacol 70:1035-47.

40. Pai R, Tarnawski AS, Tran T. 2004. Deoxycholic acid activates beta-catenin signaling pathway and increases colon cell cancer growth and invasiveness. Mol Biol Cell 15:2156-63.

41. Green MR, Sambrook J. 2017. Isolation of High-Molecular-Weight DNA Using Organic Solvents. Cold Spring Harb Protoc 2017:pdb prot093450.

42. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103-10.

43. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

44. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-5.

45. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11-6.

46. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich

C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC genomics 9:75.

47. Rodriguex-R LM, Konstantinidis, K.T. . 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints doi:4:e1900v1 doi:10.7287/peerj.preprints.1900v1.

48. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.

49. Ditscheid B, Keller S, Jahreis G. 2009. Faecal steroid excretion in humans is affected by calcium supplementation and shows gender-specific differences. Eur J Nutr 48:22-30.

50. Auchtung JM, Robinson CD, Farrell K, Britton RA. 2016. MiniBioReactor Arrays (MBRAs) as a Tool for Studying C. difficile Physiology in the Presence of a Complex Community. Methods Mol Biol 1476:235-58.

OVERALL CONCLUSION AND FUTURE DIRECTION

This study examined the *Prevotella* dominated human gut microbiota utilizing culturomics and addressed questions regarding individual bacteria colonization resistance against *C. difficile in vitro*. According to results from this study, the majority of the *Prevotella* dominated gut microbiota confer colonization resistance against *C. difficile*. However, a detailed mechanism of resistance against the pathogen by each species needs further investigation. Additionally, using the combinatorial assembly technique, we found that the minimum set of bacteria required as a community to provide protection against *C. difficile* is twelve. With detailed phenotypic characterization of the individual species in terms of substrate utilization, short chain fatty acid production, genetic traits and inhibition against *C. difficile* presented in this study, it is now possible to undertake the ecological community assembly, resilience and colonization resistance studies using a *Prevotella* dominated microbial community. These studies will help us develop the microbial therapeutics against multidrug resistance gut pathogens such as *C. difficile* and understand chronic inflammatory gut diseases such as Crohn's disease and inflammatory bowel disease.

The ability to isolate two novel species from a *Prevotella* dominated microbiota on a single medium in this culturomics study suggests this microbiota enterotype is not well defined. Also, the effect of prebiotic substrates such as rice bran and quercetin on this microbiota includes the potential benefits of reduced *Enterobacteriaceae* by enhancing propionate production. However, further culturing and isolation of the microbes enriched in the combined rice bran and quercetin substrate is warranted due to their potential as synbiotics that are useful in preventing gut dysfunction.

The hypothesis that *C. difficile* is inhibited by bile dependent mechanisms of *C. scindens* is supported by this study. However, strain wise variations are apparent when genetically and phenotypically analyzing for inhibition against *C. difficile*. Additionally, some strains of *C. scindens* were also able to inhibit *C. difficile* even in the absence of bile. Thus, use of non-bile converting, but *C. difficile* inhibiting, *C. scindens* strains would not only reduce the negative impact of microbe mediated high deoxycholate production in the gut but also provide colonization resistance against *C. difficile*.

With interpersonal variation, the key question of how the gut microbiota assembles and how it affects colonization resistance against gut pathogens remains poorly understood. Even though our study provides insights into species that inhibit *C. difficile in vitro*, further *in vivo* studies are required to explore their full potential as a community. This knowledge will help develop biotherapeutics against gut diseases.