CHARACTERIZING THE HUMAN GUT MICROBIOME THROUGH CULTURE

CHARACTERIZING THE DIVERSITY AND COMPLEXITY OF THE HUMAN GUT MICROBIOME THROUGH THE COMBINATION OF CULTURE AND CULTURE-INDEPENDENT METHODS

By

JENNIFER T. LAU, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2017)
Hamilton, Ontario (Biochemistry and Biomedical Sciences)

TITLE: Characterizing the diversity and complexity of the human gut microbiome through the combination of culture and culture-independent methods

AUTHOR: Jennifer T. Lau, B.Sc (McMaster University)

SUPERVISOR: Dr. Michael G. Surette

NUMBER OF PAGES: xv, 210

Lay Abstract

Bacteria that inhabit the human intestine are important for health, and are involved in several diseases; therefore, it is critical to determine the roles of specific bacteria. I describe a method that results in the growth and recovery of most bacteria in stool, which allows them to be studied in detail. The differences, both in behaviour and in DNA sequences, found within two different bacterial groups were characterized, and extensive variability was observed between closely related bacteria. I studied which bacteria and their functions might be important in Irritable Bowel Syndrome (IBS) by using our method for growing stool bacteria combined with sequencing of all DNA in the stool, but could not find strong support for specific bacteria causing IBS symptoms. This work shows how the ability to grow and isolate bacteria, combined with studying their DNA, allows for better understanding of their functions in the human intestine.

Abstract

The human gut microbiome is the collection of all organisms and their genetic content that inhabit the gastrointestinal tract. An overwhelming number of studies have associated the gut microbiota with health and disease, but with little consensus on which specific bacterial groups are important for causing or maintaining either state. A majority of microbiome studies only identify associations between the gut microbiome and health status, and determining causation requires the isolation and growth of bacterial isolates for further experiments. The goal of this thesis is to demonstrate that the combination of culture-based and culture-independent methods describes greater complexity and diversity in the human gut microbiota than observed by either approach alone. In the first study, a method of culture-enriched molecular profiling could capture the majority of bacterial groups found in fecal samples. Additionally, when compared to culture-independent 16S rRNA gene sequencing, culture detected more bacterial taxa. This method was applied to the targeted culture of the commensal Lachnospiraceae family. The second study explored the diversity in the isolated Lachnospiraceae strains, and compared the genetic diversity of the strains to reference genomes, revealing functional and genetic heterogeneity within the bacterial family. The third study characterized the intra-species phenotypic and genetic diversity in *Escherichia coli. E. coli* diversity was extensive between individuals, but also within-individuals, in both the phenotypes and genetic profiles. Lastly, a method of culture-enriched metagenomics was applied to a murine IBS microbiota transfer model to identify bacterial members of the microbiota and their functional pathways that may be responsible for the development of gastrointestinal

and behavioural IBS phenotypes, although no bacterial groups could be conclusively associated with symptoms. Together, the work described demonstrates that culture and culture-independent methods are complementary, and provides more resolution into the structure and diversity of the human gut microbiome than either approach in isolation.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Mike Surette, for the opportunity to work in your lab for the past six years. You have created the greatest environment to explore my scientific curiosities, to learn and grow as a scientist, and you have put together a group of the most fantastic people that I have had the pleasure to work alongside – many of whom will remain life-long friends. I could not have asked for a more supportive mentor; thank you for your endless encouragement, the excitement and passion you bring to science, your intelligence that I am in constant awe of, for calming me down when I feel anything but, and most of all, for making me a better scientist.

Thank you to my committee members, Dr. Premysl Bercik and Dr. Brian Coombes, for challenging me during committee meetings and giving me a new perspective to look at things, for your interest in my projects and for supporting my career developments. I would also like to thank Lisa Kush, who has always gone above and beyond whenever I needed help.

The experience of grad school would not have been the same without the members of the Surette lab, both current and former, that I have had the pleasure to work with and learn from. Thank you to all for your help and support over the years, for the lunchtime company, for everything you guys have taught me, and for sharing your really amazing science with me. I am so happy to be able to consider many of you my friends. Thank you to Dr. Fiona Whelan, for without you, this PhD would not be what it was. Thanks for always being there when I'm feeling overwhelmed, for sharing the latest lab news, for being someone to bounce ideas off, and for being the voice of encouragement

whenever it was needed. You'll always be my favourite person to finish a crossword with. Also thank you to the many people that I have collaborated with over the past several years, who have provided me with the chance to broaden my horizons; afterall, collaborations are what drive science.

Thank you to my parents, and Justin, for giving me the opportunity to pursue my love of science, even if you didn't always understand it. I would not have been able to do it without your support. Thanks to Sophie for your daily texts, which are always entertaining, and a reminder that there's more to life outside the grad school bubble. Thank you to my friends, both back home and at McMaster (and now also all over!), for being there through all these years, and for making the time fly by. I'm so proud of how far we've come, and so inspired by all of you. And of course, thank you to Ryan Buensuceso. I wouldn't have finished this PhD without you, your support, and your encouragement. Thanks for believing that I can do anything; I sure am glad I sat behind you in safety training that first week.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI-2	autoinducer-2
AIEC	Adherent-invasive *Escherichia coli*
AMP	antimicrobial peptide
ASD	Autism Spectrum Disorder
ASF	Altered Schaedler flora
BDNF	brain-derived neurotrophic factor
BHI	brain heart infusion
CD	Crohn's Disease
CRC	colorectal cancer
DGGE	denaturing gradient gel electrophoresis
DNA	deoxyribonucleic acid
DSS	dextran sodium sulphate
FISH	fluorescent *in-situ* hybridization
FMT	fecal microbiota transplant
GABA	gamma-aminobutyric acid
GF	germ-free
GI	gastrointestinal
h	hour
H_2O_2	hydrogen peroxide
HFD	high fat diet
HMP	Human Microbiome Project
HV	healthy volunteer
IBD	Inflammatory Bowel Disease
IBS	Irritable Bowel Syndrome
Ig	immunoglobulin
IL	interleukin
LI	Lachnospiraceae isolate
LPS	lipopolysaccharide
MDD	Major Depressive Disorder
MIC	minimum inhibitory concentration
mRNA	messenger ribonucleic acid
NOD2	nucleotide-binding oligomerization domain-containing protein 2
OTU	operational taxonomic unit
PCR	polymerase chain reaction
PD	Parkinson's Disease
PI-IBS	Post-infectious Irritable Bowel Syndrome
PMB	polymyxin B
PSA	polysaccharide A
qPCR	quantitative polymerase chain reaction
RDP	Ribosomal Database Project
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid

SCFA	short chain fatty acid
SD	standard deviation
SFB	segmented filamentous bacteria
SNP	single nucleotide polymorphism
SPF	specific-pathogen free
T1D	Type I diabetes
T2D	Type II diabetes
TLR	Toll-like receptor
T_{reg} cells	regulatory T cells
TRFLP	terminal restriction fragment length polymorphism
UC	Ulcerative Colitis
V3	variable region 3
WT	wild-type

Chapter 1

Introduction

The human microbiome

The human microbiome is a collection of all microorganisms living within and on the human body. These microbes are often referred to as commensals, meaning they benefit from being associated with a host without providing an advantage in return. However, most of these relationships are more likely to be mutualistic as it is now known that the microbiome plays a critical role in human health and development. The term microbiome refers to a defined community of all microorganisms, their genes and their surrounding environment[1]. The organisms themselves are referred to as the microbiota, and the metagenome encompasses all the genes and genomes of the microorganisms in the community. In the human microbiome, bacteria are the most commonly studied members, but this community also includes viruses (including bacteriophage), archaea, fungi, and eukaryotes such as protozoa.

An initiative of the US National Institute of Health known as the Human Microbiome Project (HMP) has contributed much of the knowledge on the communities associated with the healthy human body[2]. This program was also crucial to the development of several clinical, experimental and computational tools and protocols for studying the microbiome, in addition to adding 800 reference genomes to a public strain collection[2]. In the first phase of the project, the structure and functions of the microbiome across 15 body sites (plus an additional 3 vaginal sites sampled for female participants) were characterized in 242 healthy volunteers[3]. Additionally, samples were collected from 131 individuals at a second time point to determine the stability of the microbiome. From the results of this survey, extensive variability was found in the microbial community

composition between individuals, and also between different body sites within the same subject[3]. The primary clustering of samples was based on body site, suggesting that the microbiome of a particular body site between individuals was more similar than different body sites within a single individual. Highlighting the extensive inter-individual diversity in the microbiome, there were no taxa shared among all habitats and across all individuals, although shotgun metagenomics revealed that functional pathways were more conserved. Repeated samples from an individual were also more similar to each other compared to samples between individuals, indicating that the healthy microbiome is stable over time. These results were similar to those reported by Costello *et al.* in an earlier survey of 27 body sites from 7-9 individuals, where samples were found to group by body habitat, and individuals had high interpersonal variability but exhibited minimal temporal variability[4]. Describing the characteristics of the healthy human microbiome, albeit in a relatively homogenous population, provides an important foundation for future studies that attempt to link changes in the microbiome to various disease states.

The human gut microbiome

The human gut microbiome is the most described human-associated microbial community, due in part to the efforts of the HMP and its European counterpart, the MetaHIT (Metagenomics of the Human Intestinal Tract) consortium[5]. It comprises all the microorganisms inhabiting the gastrointestinal (GI) tract, from the mouth to the rectum, but is usually sampled via stool specimens or intestinal biopsies. Although the intestine contains 10^{11}-10^{12} bacteria[6], it is a community dominated by only five phyla: mainly

Firmicutes and Bacteroidetes, and to a lesser extent, Actinobacteria, Proteobacteria and Verrucomicrobia[7]. The gut microbiome is involved in many aspects of health and is critical in maintaining intestinal homeostasis.

Functions of the gut microbiome

The gut microbiome is critical for the maintenance of homeostasis and for proper function of the GI tract. This is highlighted in studies with germ-free (GF) mice that are raised in the absence of bacteria, which have demonstrated that intestinal function is altered in comparison to colonized mice. GF mice have increased intestinal transit time and altered intestinal contractility compared to Specific-pathogen-free (SPF) animals, indicating that bacteria are required for normal GI motility[8]. The microbiota also maintains the integrity of the intestinal epithelial barrier, as antibiotic treatment of mice results in increased permeability of the mucosa[9]. Impaired barrier function can lead to an inflammatory response, which may be a mechanism driving the pathophysiology of chronic inflammatory states such as Inflammatory Bowel Disease (IBD)[10]. The intestinal mucus layer also provides an important barrier between the microbiota and intestinal epithelial cells, and studies have shown that there is a change in the abundance of goblet cells, thickness of the mucus layer and composition of mucins after conventionalization of GF mice[11,12]. Further evidence that the gut microbiota influences the development of the intestinal mucus layer comes from a study that showed different microbiota compositions differentially affected the penetrability of the inner mucus layer, which has important implications for diseases associated with altered intestinal barrier function,

such as IBD[13]. Another important function of the gut microbiota is bile acid conversion, where the intestinal bacteria transform primary bile acids (cholic acid and chenodeoxycholic acid) from the gall bladder, to secondary bile acids (e.g. deoxycholate) for reabsorption to the liver and recycling of bile acids[14].

The gut microbiota is also required for complete digestion of dietary substrates, as the human body lacks the necessary enzymes to degrade complex polysaccharides like those found in plants[15]. These complex polysaccharides travel largely intact through the small intestine to the colon, where bacteria have the ability to ferment resistant carbohydrates (such as cellulose, pectin, and resistant starch) into short chain fatty acids (SCFA) like acetate, propionate and butyrate[16]. These SCFAs are integral to bacterial cross-feeding networks that maintain community structure, but are also involved in microbiota-host interactions. Butyrate is used as an energy source by colonocytes[17], and has also been shown to have anti-inflammatory properties[18,19] and improve epithelial barrier function[20]. As such, it is not surprising that butyrate-producing bacteria, such as Ruminococcaceae and Lachnospiraceae, are often reported to be decreased in the gut microbiota of IBD patients[21,22]. Additionally, specific strains of *Bifidobacterium* have been shown to prevent infection from *Escherichia coli* via the production of acetate from carbohydrate utilization[23].

The development of the innate and adaptive immune system is dependent on the gut microbiota. Immunological tolerance to the high load of commensals in the GI tract has likely developed through the co-evolution of the host and microbiome. Tolerance is largely mediated by regulatory T cells (T_{reg} cells) that maintain immune homeostasis and

control inflammation. SCFA produced by the gut microbiota have been shown to be anti-inflammatory through the induction of T_{reg} cells, and their production of the anti-inflammatory cytokine interleukin (IL)-10[24]. Specific members of the gut microbiota have also been reported to have direct effects on T_{reg} cell activity. Polysaccharide A (PSA) from *Bacteroides fragilis* promoted the conversion of $CD4^+$ T cells to $Foxp3^+$ T_{reg} cells via TLR2 signaling of dendritic cells[25,26]. Interestingly, *B. fragilis* PSA was also able to prevent and ameliorate chemically-induced colitis in mice. Atarashi *et al.* demonstrated that spore-forming Clostridia were able to restore the number of T_{reg} cells in GF mice, which are reduced in the large intestine compared to conventionalized mice, and the Clostridia strains were protective in chemically-induced colitis[27]. Th17 cells are also involved in intestinal and immune homeostasis, as they are important for antimicrobial peptide (AMP) secretion, maintenance of the epithelial barrier, and recruitment of neutrophils during infection; however, they may also play a role in driving inflammation in IBD and rheumatoid arthritis[28,29]. In mice, the segmented filamentous bacteria (SFB) are potent inducers of Th17 cells, but this requires adherence to intestinal epithelial cells[28,30]. In a recent study which screened human gut microbiota isolates using monocolonization of GF mice, *Bifidobacterium adolescentis*, as well as several other bacteria, were able to induce Th17 cells to a similar extent as SFB[31]. It was also demonstrated that immune responses to members of the gut microbiota were not predicted by the phylogeny of bacteria, and humans have likely evolved to harbor a gut microbiota with redundant bacterial members that can stimulate similar immune cell interactions, as

demonstrated by the ability of both Gram-negative *B. fragilis* and Gram-positive Clostridia to induce IL-10 production by T_{reg} cells[32].

Production of AMPs as part of the innate immune system is affected by the gut microbiota as colonization of GF mice induces Paneth cell expression of RegIIIγ[33]. The production of AMPs allows for the physical separation of the microbiota from the intestinal epithelial surface, which prevents triggering of immune responses through direct bacterial-host interactions that can result in an unwarranted inflammation[34]. Induction of AMPs may be dependent on specific members of the microbiota as monocolonization with *Bifidobacterium breve* could upregulate RegIIIγ, while monocolonization with *E. coli* JM83 did not[35]. Stimulation of the host immune response, which can result in upregulation of AMPs, can protect against invading pathogenic bacteria in a process known as colonization resistance. Antibiotic treatment is a risk factor for developing *Clostridium difficile* infection due to loss of the antibiotic-susceptible commensals, but colonization by commensal strains could prevent *C. difficile* infection[36].

The human gut microbiota possesses a large functional capacity for chemical transformations, and thus has the ability to impact the efficacy and toxicity of drugs. The gut bacterium *Eggerthella lenta* has been shown to have the ability to inactivate the cardiac drug digoxin, and this inactivation was enhanced by the presence of the gut microbiota[37]. Recently, several studies have shown that the gut microbiota also affects the activity of chemotherapy drugs for cancer treatment. The efficacy of immunotherapy drugs using antibodies to cytotoxic T lymphocyte protein 4 (CTLA4) and programmed cell death protein 1 ligand 1 (PDL1) depended on the microbiome as CTLA4 antibodies

were not effective in GF or antibiotic treated mice, but were successful in SPF mice and in mice monocolonized with *Bacteroides* strains[38,39]. Similarly, anti-PDL1 therapy was shown to be more effective if mice also received a *Bifidobacterium* cocktail[39]. Other studies have used the *Caenorhabditis elegans* model, where bacteria were shown to metabolize the fluoropyrimidines cancer drugs in a strain specific manner[40,41]. *E. coli* and *Comamonas* were able to either increase or decrease the responsiveness of *C. elegans* to different chemotherapies through different mechanisms of bacterial metabolism[40].

More recently, studies have linked the gut microbiota with behaviour via the brain-gut axis. Proposed mechanisms of communication include the production and detection of SCFAs, components of the immune system, production of neurotransmitters by commensals, and vagus nerve stimulation[42]. Yano *et al.* showed that spore-forming members of the gut microbiota increased serotonin production in the colon, and also increased serum levels of serotonin, which could have systemic effects[43]. Further evidence that the gut microbiota can influence behaviour comes from a mouse study where antibiotic treatment resulted in increased exploratory behaviour and higher levels of brain-derived neurotrophic factor (BDNF)[44]. This study also demonstrated that colonization of GF BALB/c mice with the microbiota from NIH Swiss mice increased exploratory behaviour and levels of brain BDNF, while colonization of NIH Swiss mice with BALB/c microbiota reduced exploratory behaviour, showing an intriguing link between the gut microbiota and the development of behaviours.

The gut microbiota in disease

Since the human gut microbiota has been implicated in many aspects of host health and physiology, unsurprisingly, it has also been implicated in several disease states. There have been associations with the gut microbiota and GI diseases such as IBD, Irritable Bowel Syndrome (IBS), colorectal cancer (CRC), as well as diseases outside the GI tract, including metabolic and neurological disorders.

The gut microbiota in gastrointestinal diseases and disorders

Inflammatory Bowel Disease

Inflammatory bowel disease (IBD), which includes ulcerative colitis (UC) and Crohn's disease (CD), is defined as chronic inflammation of the GI tract, and the gut microbiome contributes to the inflammatory response in genetically predisposed individuals[45]. Although studies have not defined a causative role of the gut microbiome in the initiation of IBD, it likely contributes to the severity of intestinal inflammation as fecal diversion has been shown to reduce inflammation in humans, and GF animals usually do not develop inflammation[46]. Additionally, genetic defects in bacterial recognition by the immune system are associated with increased risk of developing IBD. This includes the nucleotide-binding oligomerization domain 2 (NOD2) gene, which is involved in intracellular detection of bacterial muramyl dipeptide, and SNPs in this gene are associated with an increased risk for CD[47,48]. Another CD associated risk gene is *Atg16l1*, which encodes a protein in the autophagosome pathway for processing intracellular bacteria[49], and a T cell selective deletion resulted in a loss of T_{reg} cells and

spontaneous intestinal inflammation in mice[50]. Studies comparing the gut microbiota of IBD patients with healthy controls have shown differences in the community composition, although there is little consensus on the bacterial changes associated with IBD. This is likely due to differences in patient populations including disease subtype (e.g. ileal CD vs UC), treatments, and differences in methodology such as the type of tissue studied, and microbiome analysis techniques utilized. A recent study has also shown that the gut microbiota of IBD patients is less stable overtime compared to controls, which adds another confounding factor for studies attempting to define microbial changes in IBD using only one sampling timepoint[51]. In general, studies often report a decrease in health-associated bacteria such as Lachnospiraceae, Ruminococcaceae (including *Faecalibacterium prausnitzii)*, Bacteroidetes and an increase in Enterobacteriaceae[21,22,52]. These differences were observed in new-onset, treatment naïve pediatric CD patients when compared to healthy controls[53], indicating that these microbiome changes occur early in disease development. *F. prausnitzii* is often reported to be decreased in the gut microbiota of CD patients. It has been shown to induce IL-10 production and to decrease levels of TNF-α and IL-12, resulting in reduced disease severity during chemically-induced colitis in mice[54]. Conversely, increased levels of attaching and invading *E. coli* (AIEC) are associated with CD. AIEC have the ability to invade the ileal mucosa and replicate in macrophages[55,56], and induce a pro-inflammatory response in a chronic infection mouse model of CD[57]. Examining the immunoglobulin A (IgA) coated bacteria from IBD patients provided further evidence that the gut microbiota is involved in the inflammatory response, as colonization of GF mice with the high IgA-

coated fraction of the microbiota induced inflammation during chemically-induced colitis[58].

Since the gut microbiome is implicated in maintaining inflammation in IBD patients, a few studies have attempted to use fecal microbiota transplants (FMT) as a therapeutic treatment (mainly for UC), based on the success of FMT as a treatment for *C. difficile* infections, although with limited success[59]. Although route of administration, FMT dose, and clinical end points varied between studies, remission rates ranged from 24-34% of participants. Rossen *et al.* reported no difference in remission rates between patients receiving FMT or placebo[60], whereas Moayyedi *et al.* found a statistically significant remission rate in a randomized control trial of FMT for UC[61]. A recent study found that one dose of a two-donor FMT resulted in 34% of UC patients achieving clinical response, and similarly to Moayyedi *et al.*, observed that there was increased diversity in patients after FMT[62]. Although the specific bacterial groups and mechanism involved in successful FMTs have yet to be elucidated, this may be a potential treatment for IBD and further clinical trials are required.

Irritable Bowel Syndrome

Irritable bowel syndrome (IBS) is the most prevalent functional GI disorder, characterized by recurring abdominal pain and alterations in the consistency or frequency of bowel movements[63]. The cause of IBS is unknown and likely to be multifactorial[64], however, the gut microbiota can affect GI motility[65], the brain-gut axis[44], low grade inflammation[66,67], visceral sensitivity[68], and intestinal epithelial barrier function[69], which

are involved in the pathophysiology of the disorder. There are no biomarkers for the diagnosis of IBS, but efforts have been made to describe changes in the gut microbiota that may be used as a signature for the disorder. Although several studies have compared the gut microbiota of IBS patients to healthy controls, there has been little consensus on which bacterial groups are associated with the disorder[64,70]. A recent systematic review and meta-analysis comparing IBS microbiota studies that used qPCR to quantify bacterial groups reported a consistent decrease in *Lactobacillus, Bifidobacterium* and *F. prausnitzii* in IBS compared to healthy controls[71]. Decreased abundance of *F. prausnitzii* in IBS patients has also been reported in 16S rRNA gene sequencing studies[72,73]. A larger study with 110 IBS patients and 39 healthy controls used machine learning to determine that 90 bacterial taxonomic groups were predictive of IBS symptom severity, and *Faecalibacterium, Oscillibacter, Blautia* and *Coprococcus* were positively associated with less severe symptoms and health[74].

The strongest clinical evidence for a causative role of the gut microbiota in IBS comes from post-infectious IBS (PI-IBS), where up to 26% of patients have recurrent abdominal pain and altered bowel movements after an acute enteric infection. These IBS symptoms can be long-term, even though the initial infection has been cleared and the causative pathogen is no longer detectable[75,76]. Two recent IBS microbiota transfer studies in mice have also implicated the gut microbiota as a causative agent in IBS. Fecal microbiota transplant from IBS-D (diarrhea) patients to GF mice resulted in altered serum metabolites, increased GI transit and intestinal epithelial barrier permeability, and immune activation in some animals when compared to mice receiving healthy control

FMT[77]. Interestingly, anxiety-like behaviour was also transferred from IBS-D patients with co-morbid anxiety, providing further evidence that IBS is a brain-gut disorder. A separate study colonized mice with fecal microbiota from IBS-C (constipation) patients, and reported slower GI transit and decreased SCFA levels in animals[8]. These studies demonstrate that several symptoms of IBS can be transferred through the gut microbiota, which likely plays a role in the pathophysiology of the disorder.

Colorectal cancer

The gut microbiota has also been implicated in the development of colorectal cancer. Similarly to the microbiota changes reported for IBD, the microbiota of CRC patients are enriched in opportunistic pathogens such as *Enterococcus*, *Escherichia*, *Klebsiella*, and *Streptococcus*, and decreased in butyrate-producing bacteria such as *Roseburia* and Lachnospiraceae, compared to healthy controls[78]. Studies have also reported an increase in *B. fragilis* in CRC patients[78,79], and enterotoxigenic *B. fragilis* can cause diarrhea, inflammation and induce the formation of colonic tumors in Min mice through the activation of Th17 cells[79]. Another example of bacterial toxins in CRC was demonstrated in a spontaneous colitis model, where *E. coli* that harboured a polyketide synthase island, encoding the genotoxin colibactin, was able to induce invasive carcinomas[80]. *Fusobacterium nucleatum* has also been associated with CRC, as it is enriched in carcinomas compared to adjacent healthy colon tissue[81], and could accelerate the development of tumors in genetically susceptible mice[82]. In a mouse model of inflammation-induced CRC, changes in the microbiota was associated with intestinal

tumor formation, including increased *Bacteroides, Odoribacter* and *Akkermansia*, while *Prevotellaceae* and *Porphyromonadaceae* were decreased[83]. Colonizing GF mice with microbiota from mice with tumors increased tumourigenesis, but antibiotic treatment could decrease the number and size of tumors, implicating the gut microbiota in the formation of CRC tumours.

The gut microbiota in metabolic disorders

Obesity

Some of the earliest studies of the gut microbiome investigated its association with obesity. The gut microbiota was shown to have a critical role in obtaining energy from food, as conventionalization of GF mice resulted in increased body fat and insulin resistance in the absence of increased dietary intake[84]. The gut microbiota of obese mice was found to be different from their lean littermates, with decreased Bacteroidetes and increased Firmicutes[85,86]. This altered microbiome could harvest more energy from the diet, and this phenotype was transmissible to GF animals. Mechanistically, the obese microbiome was enriched in enzymes involved in breaking down indigestible polysaccharides, resulting in increased levels of butyrate and acetate. Interestingly, a recent study showed a link between acetate and obesity; increased acetate production by the gut microbiota activated the parasympathetic nervous system to increase insulin secretion, and induced hyperphagia and obesity in rats[87]. Insulin resistant rats on a high fat diet (HFD) had higher levels of acetate, and antibiotic treatment resulted in decreased insulin secretion, which could be reversed by acetate infusion. These effects could be

prevented by vagotomy, which provided a novel link between the brain, gut microbiota and host metabolism. In a study involving twins discordant for obesity, transferring the human gut microbiota to GF mice led to increased body mass and adiposity in the animals receiving the obese twin microbiota compared to the mice receiving the lean microbiota. As observed previously, there were higher levels of Bacteroidales (including *B. uniformis, B. vulgatus,* and *B. cellulosilyticus*) in the lean microbiota, and this microbiome was capable of increased SCFA fermentation and bile acid transformation, while the obese microbiome was enriched for branched chain amino acid metabolism, demonstrating that functional differences in the obese and lean microbiome may be contributing to the disorder. Although these early studies reported differences in the gut microbiota composition between healthy and obese individuals, re-analysis of the combined data from multiple studies suggests that these differences are much more subtle than initially reported[88], and functional differences may be more informative for describing the role of the gut microbiota in obesity.

Type 2 Diabetes

In addition to its role in obesity, the gut microbiota has also been implicated in Type 2 diabetes (T2D), which is often a comorbid disease. Differences in the gut microbiome between T2D patients and healthy controls have been reported in a metagenomics study, although the differences between the two groups were moderate[89]. T2D gut microbiota was decreased in some butyrate-producers, and increased in *Akkermansia, Desulfovibrio* and opportunistic pathogens. Functionally, there were

increased genes related to oxidative stress, and 50 gene markers could be used to classify the T2D microbiome from the healthy microbiome. A randomized, placebo controlled double-blind study found that the mechanism of the T2D drug metformin involves the gut microbiome. Transferring fecal samples from subjects before and after metformin treatment to GF mice on a HFD resulted in improved glucose tolerance in mice with the metformin-treated donor microbiota. Metformin increased fecal propionate and butyrate, as well as increased plasma bile salts, which was accompanied by an increase in bile salt hydrolase genes. This suggests that alterations in the gut microbiome may be involved in T2D, and targeting the gut microbiota may be an effective therapeutic for an increasingly prevalent disorder.

A link between the innate immune system and the gut microbiota may also play a role in T2D since $NOD2^{-/-}$ mice exhibit a different microbiota than WT mice, and $NOD2^{-/-}$ mice are also more resistant to insulin, a phenotype which can be transferred to WT GF mice[90]. $NOD2^{-/-}$ mice also had more gut bacteria translocating to adipose tissue, which led to inflammation and increased insulin resistance. Additionally, antibiotic treatment reduced fasting blood glucose in $NOD2^{-/-}$ mice on a HFD, but not WT mice, providing further evidence for a link between the gut microbiome and T2D.

Type 1 Diabetes

Although the mechanisms of Type 1 diabetes (T1D) are different than T2D, as T1D is an autoimmune disorder, alterations in the gut microbiota may lead to an improper immune response and the development of disease. Wen *et al.* found that SPF non-obese

diabetic (NOD) MyD88$^{-/-}$ mice did not develop T1D, but GF NOD MyD88$^{-/-}$ mice did, and that colonization with a simple, defined bacterial community (Altered Schaedler's flora, ASF) reduced the incidence of T1D[91]. Another study exposed NOD mice to continuous low dose or pulsed antibiotics early in life, to mimic early childhood exposures[92]. T1D incidence was higher and onset of disease was accelerated in the pulsed antibiotics group, whose gut microbiota was decreased in Bacteroidetes and *Bifidobacterium*, and increased in *Enterococcus* and *Akkermansia mucinophila*. The animals with the pulsed antibiotics also had decreased Th17 and T_{reg} cells. These mouse studies show that an interaction between the gut microbiota and the immune system may be implicated in T1D. In human studies, the gut microbiota of children with T1D was different than healthy controls, with increased *Bacteroides*, and decreased *Bifidobacterium*, lactate- and butyrate-producers[93,94]. A metagenomic study comparing a small number of patients with T1D to controls also found decreased levels of butyrate producers, mucin degraders, and increased adhesion and flagella synthesis in the gut microbiome of T1D patients, which could trigger an autoimmune response. Supporting previously observed results, a study of European infants with high rates of autoimmune disease (from Finland and Estonia) had primary exposure to LPS from Bacteroides, while Russian infants with low rates of autoimmune disease had LPS exposure from *E. coli*[95]. *E. coli* LPS could protect mice from developing T1D but *Bacteroides dorei* LPS inhibited immune signaling and tolerance, and was not protective in a mouse model. This suggests that exposure to species-specific bacterial products affects education of the immune

system, which likely plays a role in the development of autoimmune disorders such as T1D.

The gut microbiota in neurological disorders

Anxiety and Depression

GI symptoms are often comorbidities of neurological disorders, and several recent studies have provided evidence of a brain-gut microbiome axis that may be important in neurological disorders. GF mice exhibit less anxiety-like behaviour than SPF mice, and have altered BDNF and 5-hydroxytryptamine (serotonin) receptor expression in the brain[96]. A separate animal study determined that a maternal separation model of early life stress altered the gut microbiota compared to control mice, and interestingly, the development of anxiety-like behaviour was dependent on the presence of the commensal bacteria since behavioural changes were not observed in GF mice[97]. Depression-like behaviour is also decreased in GF mice, and microbiota transfer studies have shown that depression can be transferred from human donors with Major Depressive Disorder (MDD) to mice[98,99]. Human studies comparing the gut microbiota of MDD patients with healthy controls suggests an altered microbiome in depressed patients, with reports of increased Shannon diversity, increased Enterobacteriaceae and *Alistipes*, and decreased *Faecalibacterium* and Ruminococcaceae[100], along with differences in carbohydrate and amino acid metabolism[99]. Jiang *et al.* found a negative correlation between the abundance of *Faecalibacterium* and severity of depressive symptoms[100], but further studies are required to determine if this association is cause or effect. A handful of studies have

provided insight into the role of the gut microbiota in depression, by using pre- or probiotics to reduce depression-like behaviour in animals. Normal mice receiving the probiotic *Lactobacillus rhamnosus* exhibited decreased depression and anxiety-like behaviour, along with decreased corticosterone release during stress tests[101]. Gamma-amino butyric acid (GABA) neurotransmitter receptors were altered in the brain, but severing the vagus nerve negated the effect of *L. rhamnosus* on behaviour and GABA receptors, indicating that the vagus nerve is the mechanism for brain-gut communication in this case. Two other mouse studies that looked at the effect of prebiotics on behaviour found decreased depression and anxiety-like behaviours during stresss exposure, accompanied by changes in corticosterone and SCFA levels[102,103]. An intriguing study recently indicated a potential link between the inflammasome, gut microbiome and anxiety/depression as animals genetically deficient in inflammasome signaling had decreased anxiety and depressive behaviour, along with an altered gut microbiota compared to WT animals[104]. Additionally, treating stressed mice with an inflammasome agonist reduced depressive behaviour, and increased *Akkermansia*, *Blautia* and Lachnospiraceae in the gut microbiota. This proposes a novel link between the gut microbiota and the brain, mediated by the immune system. As a majority of these studies linking the gut microbiota to anxiety and depression have been performed in animal models, it remains to be seen if findings will translate to human subjects.

Autism Spectrum Disorder and Parkinson's Disease

Two other neurological disorders that may be associated with the gut microbiota are Autism Spectrum Disorder (ASD) and Parkinson's disease (PD). GI dysfunction is a common symptom in PD patients, and precedes the onset of motor symptoms[105]. Several studies have compared the gut microbiota of ASD or PD patients with healthy controls and report alterations in community composition; however, as is the case with many microbiome studies, there is little consensus on which bacterial groups are increased or decreased, and results are often conflicting[106–110]. More convincing evidence has come from animal models of these disorders, including two studies using mouse models of ASD, where behavioural abnormalities could be attenuated by altering the gut microbiota. In a maternal immune activation model of ASD, offspring had GI comorbidities similar to human patients, including increased intestinal permeability and alterations in intestinal microbiota[111]. Colonizing ASF mice with *B. fragilis* improved GI permeability and behaviour symptoms, but the mechanism was independent of the immune system and *B. fragilis* PSA production. Metabolomics revealed increased metabolites regulated by the gut microbiota in ASD mice, including serum 4-ethylphenylsulphate (4EPS), indolepyruvate and serotonin, but these were decreased upon *B. fragilis* colonization. Treating mice with 4EPS induced anxiety-like behaviour in animals, demonstrating the impact of gut microbiota-produced metabolites on the development of ASD behaviours. Another study found that mothers on a long-term HFD produced a large proportion of offspring with ASD-like behaviours compared to chow fed mothers, and the gut microbiota was also different between the groups, with a decrease in *Lactobacillus reuteri*

observed as the most significant difference[112]. Adding *L. reuteri* to the drinking water of mice restored deficits in social interactions, but not anxiety and repetitive behaviours. A study using a mouse model of Parkinson's disease, where animals overexpress alpha-synuclein protein, resulting in aggregates that are hallmarks of PD, found that the gut microbiota was required for the development of symptoms since GF mice had fewer aggregates and did not develop motor symptoms[113]. Symptoms in PD mice could also be attenuated by antibiotic treatment, and colonizing GF mice with microbiota from PD patients resulted in motor dysfunction, and also altered levels of SCFAs. Treating GF mice with SCFAs also led to alpha-synuclein aggregation and motor deficits. Taken together, animal studies of ASD and PD indicate that changes in the microbiome, leading to altered metabolites such as neurotransmitters and SCFAs, may be important to the development of these disorders, and altering the microbiota may be a therapeutic avenue.

Studying the human gut microbiome

Culture-independent methods (Table 1)

With the advancements of sequencing technologies in the past several years, high-throughput sequencing methods are now fast and cost-effective enough to be accessible by many research labs, and the development of open-source bioinformatics tools by the community has made analysis of large datasets feasible. This has allowed large-scale population surveys like the HMP and MetaHIT to be feasible, providing valuable reference data for the microbiome of healthy controls for use as comparisons in other studies. These projects employed 16S rRNA gene sequencing and shotgun metagenomics

to explore both structure and function of the human microbiome. Marker gene sequencing, such as the bacterial 16S rRNA gene (part of the 30S small subunit ribosome), allows for determination of the taxonomic groups present in a microbiome sample, and their relative abundances. PCR primers that are complementary to the conserved regions of the 16S rRNA gene are used to amplify the variable regions, where differences in DNA sequences reflect bacterial phylogeny. The variable regions are then sequenced, and similar sequences are binned based on identity thresholds (such as 97% similarity) into operational taxonomic units (OTUs), using de novo or reference based algorithms. Representative sequences from each OTU can then be compared to reference databases for taxonomic assignments; however, this is usually limited to genus-level resolution due to the short length of the variable regions sequenced. Primer bias is an important consideration here as some primer sets are less effective for amplifying certain bacterial groups[114].

Metagenomics allows for the survey of the functional potential of a community, and includes the viral, fungal and archaeal population, in addition to bacterial. DNA is extracted from the sample and sheared to equal sized fragments for sequencing. Reads are either directly compared to reference genomes or assembled first, and then contigs are compared to reference databases and genomes. Metagenomics provides information about community composition, allows for identification of genes and pathways present and their abundances, and also allows for analysis of SNPs and genetic variation within a population. This allows for greater resolution of taxonomy, and bacteria can be identified to the species-level, and sometimes even with strain specificity[115]. The HMP used

metagenomics to show that there were body site-specific functions associated with the microbiome; for example, carbohydrate degradation and hydrogen sulfide production were processes predominantly found in the gut when compared to other body sites[3].

Other methods for surveying bacterial communities that were more commonly employed before the rise in popularity of high-throughput sequencing were also based on targeting the 16S rRNA gene or other gene markers. These methods include fluorescent in situ hybridization (FISH), terminal restriction fragment length polymorphisms (TRFLP), denaturing gradient gel electrophoresis (DGGE), and qPCR using primers for specific bacterial groups.

Other 'omic methods have started to become more common in microbiome studies, although not yet to the extent that 16S rRNA gene or metagenomic sequencing is utilized. DNA based methods can provide information on the composition of a community and the genes that are present within the sample, but do not indicate whether the genes are being actively expressed, and usually cannot measure short term changes in the microbiome (i.e. on the order of minutes). Metatranscriptomics is the sequencing of mRNA from a community to survey gene expression of the active population. RNA is extracted from a sample, and rRNA is depleted before conversion of mRNA to cDNA for high-throughput sequencing. Although metatranscriptomics allows for measurement of rapid changes to a community, RNA is less stable than DNA so it is important that the sample be preserved to maintain RNA integrity. Additionally, in order to be quantitative, metagenomics needs to be performed on the same sample to account for DNA copy number of genes[116]. Metaproteomics and metabolomics is the study of the proteins, and

metabolites and small molecules, respectively, from a microbiome community. Total protein or metabolites are extracted from a sample, separated by liquid chromatography, and analyzed by mass-spectrometry before comparison to a database for identification and quantification of peptides and molecules. These methods allow for the detection of host- and microbial-derived molecules; however, total extraction of all proteins and metabolites from a sample is challenging, and identification of molecules and the sources of metabolites can be difficult as databases are not as complete as genome databases. As microbiome research progresses, the combination of several 'omic methods and systems level analyses will provide a more complete picture of host-associated communities[117]. A study employing metagenomics and metatranscriptomics on stool samples from eight healthy individuals found that 22% of transcripts were differentially regulated by at least 10-fold compared to the abundance of DNA[118]. Underexpressed genes included those involved in sporulation, and upregulated genes included ribosome biogenesis and methanogenesis, indicating that the archaea *Methanobrevibacter smithii* was highly transcriptionally active in the gut. Transcriptional profiles were also found to be variable between individuals, compared to the more conserved functional profiles. This demonstrates the power of integrating multi-omic data; however, analysis of this data will require more sophisticated approaches such as machine learning for the identification of patterns from high-throughput data sets[119].

Culture-dependent methods

Although studies of the human microbiome have increased exponentially in the past decade, most of the work has relied solely on culture-independent methods. As discussed above, there has been much insight gained into the roles of the gut microbiome in human health, and studies have associated changes in the gut microbiome with several diseases. Most of these studies have used culture-independent methods to study the microbiome, but as a limitation of such methods, have resulted in a lack of consensus on specific microbiome alterations. Studies often contradict each other due to differences in methodology and human populations sampled. Additionally, a majority of studies are observational, and there is currently not enough evidence to determine whether the gut microbiome is causative or merely correlated to most disease states. Another limitation of most previous studies is a lack of taxonomic resolution, especially in 16S rRNA gene sequencing studies, and as a result, species and strain-level variability are rarely accounted for. Compared to conventional microbiological culture, culture-independent methods are usually less laborious, so studies with large sample sizes can be undertaken, and longitudinal surveys are more practical to carry out. Additionally, early sequencing studies of the human gut microbiome reported that a majority of the bacteria were not culturable, as much more diversity was revealed through sequencing than previously described by culture[7,120]. This further contributed to the replacement of culture with culture-independent molecular methods as the microbiome field advanced. However, it is now apparent that a bottleneck exists to prove that specific bacteria can contribute to the initiation or maintenance of a disease, or if it has a beneficial, health-promoting role.

Most human microbiome studies can only derive correlative associations and generate hypotheses, but do not demonstrate causality. The hypotheses being proposed from these studies need to be tested empirically, using isolated bacteria or defined bacterial communities to definitively associate with phenotypes via modified Koch's postulates[121], and eventually using genetic knockouts and complementation to confirm mechanisms of action. Unfortunately, bacterial cultivation and isolation has lagged behind in the microbiome field, and there is a wide gap in the bacteria described by culture-independent studies and isolates available in bacterial culture collections, with prevalent and abundant bacterial groups lacking reference strains[122,123].

In addition to allowing for experiments that will define the role of bacteria in health and disease, there are several advantages of bacterial culture compared to culture-independent methods. Culture allows for determination of the viable community in a sample, and can help differentiate the colonizing population of a microbiota community from DNA in the environment. Airway samples can contain contaminating DNA from inhaled air, and oral samples likely contain DNA from ingested food, which would be detected through DNA sequencing methods. However, the ability to grow an organism from a sample means it is much more likely that the bacteria are contributing to the community being studied. The targeted use of selective media can allow for the growth and detection of less abundant bacteria from a microbiota sample. For example, MacConkey agar is extremely successful at enriching Enterobacteriaceae, which usually comprises only a small proportion of a fecal sample[124]. Isolation also makes whole genome sequencing of strains more accessible, since well-assembled genomes can be

difficult to obtain from shotgun metagenomics, except for the most dominant members of a community. Increasing the number of complete genomes in reference databases is particularly important as reference databases are essential for 'omic studies, and genomes from isolates can help relate functions to species[125]. Culture allows for the study of bacterial phenotypes, including microbe-host and microbe-microbe interactions. Importantly, the ability to culture microorganisms will allow access to the products and metabolites, such as antibiotics, that could be exploited for therapeutic use[126,127], and lead to the development of synthetic microbial communities for treating diseases where the pathophysiology involves an altered microbiome.

Early culture-dependent studies of the human gut microbiome

The development of anaerobic culture allowed access to bacteria from the human gut microbiota that could not previously be grown, as the majority of the bacteria inhabiting the GI tract are obligate anaerobes. Early studies characterized the gut microbiota by culturing fecal samples, and described the overall diversity and detailed the bacterial groups present in individuals, including describing changes with diet and disease. Seminal work by Moore *et al.* cultured the fecal bacteria of 20 Japanese-Hawaiian males using anaerobic tubes, and provided quantitative and qualitative descriptions of 113 types of bacteria from 1147 isolates[128]. This included results of biochemical tests, descriptions of metabolic capabilities, gas production, cell and colony morphologies of each species. A later study of 88 individuals with various risks of developing colon cancer described 371 taxa from 5350 isolates, and found 15 bacterial

groups were associated with high risk of colon cancer, including *Bacteroides* and *Bifidoacterium*, while *Lactobacillus* and *Eubacterium aerofaciens* (now *Collinsella aerofaciens*) was associated with low risk[129]. This study was also one of the first to find that individuals had their own distinctive microbial composition.

Studies using both culture and culture-independent methods

More recent studies of the human gut microbiota using traditional bacterial culture have also employed culture-independent molecular methods as a complementary approach for identification of the taxa cultured[130–132]. Contrary to what was previously reported in the earliest culture-independent studies, these studies have shown that cultured microbes represent a substantial fraction of the human gut microbiota, and a recent literature review reported that more than 50% of the bacterial diversity is culturable[133]. Using a high-throughput method of anaerobic culture, Goodman *et al.* created personalized culture collections of thousands of isolates that represented the taxa and functions present in fecal samples[130]. 632 OTUs were observed in the complete fecal sample, and by culturing on GMM medium, 316 OTUs (50%) were detected. Colonization of GF mice with cultured communities resulted in similar biogeographical distribution along the GI tract and similar changes to diet compared to mice colonized with complete fecal communities. By comparing species detected in metagenomic sequencing of fecal samples to colonies grown on YCFA medium, Browne *et al.* demonstrated that 74% of 741 species could be cultured[132]. Another study designed media with the goal of culturing a representative proportion of the human gut microbiota,

combined with high-throughput sequencing to characterize the communities grown on each of the 10 media tested[131]. Most of the media used in this study captured 50-71% of the family-level taxa, but 54% of OTUs were not cultured. Interestingly, they reported that 113 OTUs were cultured but not in the culture-independent sequencing. Several other studies have employed a method of culturomics, where large-scale culture using many conditions is followed by identification of colonies using MALDI-TOF mass spectrometry and genome sequencing of novel isolates. These studies have screened more than 900 000 colonies, resulting in the cultivation of 247 previously undescribed species[134,135]. Culturomics was able to identify 77% of 1525 bacteria identified in the gut. In contrast to other microbiome studies, culturomics utilizes a large number of culture conditions, including enrichment in blood and rumen fluid before subculture onto agar plates, phage depletion of dominant bacteria to allow growth of less abundant organisms, and media with high salt concentrations to enrich for halophiles[135].

Although the studies discussed above used traditional culture on agar plates or broth, a microfluidics workflow was successfully utilized to cultivate target bacteria from clinical samples, where cells are grown into microcolonies in nanoliter wells on a chip[136]. This approach led to the isolation of a new genus related to *Oscillibacter* from the HMP's Most Wanted list. Another non-classical culture approach is the use of continuous culture systems, or chemostats, which allows for controlled experiments of gut microbiota perturbations[137,138]. Although chemostat communities do not reproduce the compositions of the original inoculum, steady state, complex communities can be maintained for

experimental manipulation, and communities reflect the individuality of the gut microbiota[139].

A useful application of bacterial culture is the targeting of bacteria from the gut microbiota based on specific phenotypes. Rettedal *et al.* used combinations of 16 antibiotics to determine the phylogenetic distribution of antibiotic resistance profiles in cultured fecal samples[131]. High throughput sequencing combined with selective culture on media with antibiotics resulted in successful isolation of previously uncultured bacteria. Treating fecal samples with ethanol, followed by culture on media containing the secondary bile acid taurocholate, allowed for phenotypic culture of sporulating bacteria[132]. This landmark paper showed that 60% of the genera found in the gut microbiota are spore-formers, which likely plays an important role in host-to-host transmission of obligate anaerobes that make up a majority of gut commensals. Another study cultured 22 fecal samples on media containing gluten to isolate and characterize bacteria that could metabolize gluten, which were identified by halos of clearing on the gluten media around the colonies[140]. From 144 strains isolated, 97 were found to be involved in gluten metabolism, and thus could potentially play a role in celiac disease. Culturing based on phenotypes is a method that can be applied to identify and isolate strains with characteristics of interest, as opposed to phenotyping strains after isolation, likely reducing the number of colonies that would need to be screened.

The continued need for culturing the gut microbiome

As a result of the combined efforts from the last century, the first 1000 cultured species from the human gastrointestinal microbiota have been described[124]. These cultured species mostly represent the dominant members of the community and are comprised of colonic bacteria, as fecal samples are the most accessible for cultivation[130,133,141]. Increased efforts and novel cultivation approaches may be required to obtain more cultured representatives of low abundant organisms. Additionally, the GI tract can harbor hundreds of bacterial strains, and strains are rarely shared between individuals[115,142] . This is important as strains can have differences in functions, and characterizing the gut microbiota at the strain level is important for describing diversity and population structure. To capture strain differences, several isolates of a species would have to be cultivated for proper representation of the diversity and functional repertoire within taxa. For example, the cardiac glycoside reductase involved in digoxin inactivation by *E. lenta* was only found in some strains, so was not a characteristic universally present across the species and may have been missed if only one representative isolate from the species was studied[37]. Arthur *et al.* found that only *E. coli* strains with a polyketide synthase (*pks)* island could cause DNA damage via production of a genotoxin and development of invasive carcinoma, while *pks*$^{-/-}$ strains did not[80]. Another example of strain-level importance was demonstrated in a FMT study for metabolic syndrome, where no changes were observed at the species-level after FMT, but strain-level analysis showed co-existence of donor strains with recipient strains, and also instances of strain-level

replacement after FMT[143]. This suggests that strain-level differences within species of the gut microbiota may be an important consideration for microbiota therapies.

The combination of culture and next-generation sequencing will be required for advancing the microbiome field. Reference databases contain a large proportion of unannotated genes so it is imperative to isolate more strains to characterize and determine their metabolic capabilities and fill in the missing gaps. To take full advantage of metagenomic sequencing (and other 'omic methods) will necessitate more efforts in bacterial culture so that the full spectrum of phylogenetic diversity and functions in a community is captured[141]. Additionally, cultured strains from the gut microbiota will allow for testing of phenotypes, and for determining the functional diversity among commensals. Some phenotypes, such as minimum inhibitory concentrations (MIC) of antibiotics cannot be determined from genetics, and requires experimental testing of strains. As it is evident that the human gut microbiota plays many functions and may be involved in many disorders, it will be critical to culture microorganisms to determine the exact bacteria involved and their mechanisms of action contributing to gut homeostasis or in driving disease.

Central hypothesis and thesis objectives

As the human gut microbiome field moves from observational studies characterizing the community, to translational studies for developing methods to manipulate the microbiome for disease treatment, access of organisms through culture is becoming increasingly important. Although a few studies have demonstrated that a large proportion of the human

gut microbiota is accessible by culture, further work is required to develop accessible culture methods that use more than one type of media to capture the cultivable gut microbiota. The overarching hypothesis of this thesis is that culture and culture-independent methods are complementary approaches for studying the human gut microbiome, and that the combined approach provides more information on the diversity and complexity of the human gut microbiome than either method alone. I will also apply both methodologies to determine the potential role of the microbiome in Irritable Bowel Syndrome. As the gut microbiome has been associated with gastrointestinal, as well as neurological disorders, IBS is an especially relevant condition for studies of the gut microbiota since it is a disorder that involves the microbiota-gut-brain axis.

Aim 1 Culture-enriched molecular profiling of human fecal samples allows for the investigation of the diversity of the gut microbiota, and will also allow for the determination of the culturable proportion of the gut microbiota. **(Chapter 2)**

Aim 2 Profiling the genetic and phenotypic diversity of Lachnospiraceae and *Escherichia coli* isolates will reveal large diversity within these commensal groups of the human gut microbiota. **(Chapter 3 & 4)**

Aim 3 Applying culture-enriched metagenomics to a murine IBS microbiota transfer model will allow for the determination of microbiota members responsible for gastrointestinal and behavioral IBS phenotypes. **(Chapter 5)**

Table 1. Comparison of culture-independent methods for studying the microbiome

	16S rRNA gene sequencing	Metagenomics	Metatranscriptomics	Metaproteomics	Metabolomics
Molecule	16s rDNA	Total DNA	mRNA	Proteins/peptides	Metabolites
Information	Composition	Composition/ function	Function	Function	Function
Population	Bacteria only	All microbes*	All microbes*	Microbes & Host	Microbes & Host
Taxonomic resolution	Genus-level	Strain-level	Strain-level	Source unknown	Source unknown
Cost	Low	High	High	Intermediate	Intermediate
Reference Databases	Good	Good	Good	Poor	Poor
Advantages	Most available resources	Functional & compositional info	Observe changes in short time scales	Observe changes in short time scales	Small amount of biomass required
Caveats	Primer bias	Computationally intensive	Difficult to get high quality RNA in sufficient amounts, rapid degradation	Large biomass required	Cannot link metabolites to source

* may contain host DNA/transcripts if not removed

Chapter 2

Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling

Preface

This chapter consists of work previously published as:

Lau JT, Whelan FJ, Herath I, Lee CH, Collins SM, Bercik P and Surette MG. Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Med.* **8**, 72 (2016) doi:10.186/s13073-016-0327-7

Contributions:

JTL and IH performed the experiments. Volunteer and patient samples were provided by CHL, SCM, and PB. Data analysis was performed by JTL, FJW, and MGS. JTL, PB, and MGS designed the study and wrote the manuscript.

Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling

Jennifer T. Lau[1], Fiona J. Whelan[1], Isiri Herath[1], Christine H. Lee[2,3], Stephen M. Collins[4], Premysl Bercik[4] and Michael G. Surette[1,4]*

[1]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada, L8S 4K1
[2]Department of Medicine, Division of Infectious Diseases, McMaster University, Hamilton, ON, Canada, L8S 4K1
[3]Hamilton Regional Laboratory Medicine Program, Hamilton, ON, Canada, L8N 4A6
[4]Department of Medicine, Farncombe Family Digestive Health Research Institute, McMaster University, Hamilton, ON, Canada, L8S 4K1

*Corresponding Author: Michael G. Surette; surette@mcmaster.ca; (905) 525-9140 ext. 21964; 1280 Main St W, HSC 3N-9, Hamilton, ON, L8S 4K1; fax: (905) 522-3454

Author Emails:
JTL – lauj8@mcmaster.ca
FJW – whelanfj@mcmaster.ca
IH – isiri91@gmail.com
CHL – clee@stjosham.on.ca
SMC – scollins@mcmaster.ca
PB – bercikp@mcmaster.ca
MGS – surette@mcmaster.ca

Abstract

Background

The human gut microbiota has been implicated in most aspects of health and disease; however, most of the bacteria in this community are considered unculturable, so studies have relied on molecular-based methods. These methods generally do not permit the isolation of organisms, which is required to fully explore the functional roles of bacteria for definitive association to host phenotypes. Using a combination of culture and 16S rRNA gene sequencing, referred to as culture-enriched molecular profiling, we show that the majority of the bacteria identified by 16S sequencing of the human gut microbiota can be cultured.

Methods

Five fresh, anaerobic fecal samples were cultured using 33 media, and incubation of plates anaerobically and aerobically resulted in 66 culture conditions for culture-enriched molecular profiling. The cultivable portion of the fecal microbiota was determined by comparing the Operational Taxonomic Units (OTUs) recovered by 16S sequencing of the culture plates to OTUs from culture-independent sequencing of the fecal sample. Targeted isolation of Lachnospiraceae strains using conditions defined by culture-enriched molecular profiling was carried out on two fresh stool samples.

Results

We show that culture-enriched molecular profiling, utilizing 66 culture conditions combined with 16S rRNA gene sequencing, allowed for the culturing of an average of 95% of the OTUs present at greater than 0.1% abundance in fecal samples. Uncultured

OTUs were low abundance in stool. Importantly, comparing culture-enrichment to culture-independent sequencing revealed that the majority of OTUs were detected only by culture, highlighting the advantage of culture for studying the diversity of the gut microbiota. Applying culture-enriched molecular profiling to target Lachnospiraceae strains resulted in the recovery of 79 isolates, 12 of which are on the Human Microbiome Project's "Most Wanted" list.

Conclusions

We show that through culture-enriched molecular profiling, the majority of the bacteria in the human gut microbiota can be cultured and this method revealed greater bacterial diversity compared to culture-independent sequencing. Additionally, this method could be applied for the targeted recovery of a specific bacterial group. This approach allows for the isolation of bacteria of interest from the gut microbiota, providing new opportunities to explore mechanisms of microbiota-host interactions, and the diversity of the human microbiota.

Background

The gastrointestinal microbiota is a highly diverse community, but the majority of the bacteria are considered unculturable since more recent sequencing-based studies have revealed greater diversity than previously detected by culture[7,120]. Consequently, most studies characterizing the human gut microbiome have relied on culture-independent sequencing methods. These studies have provided insights on the community composition of the gut microbiota in healthy individuals[144], how it changes with environmental perturbation[145,146], and its potential role in a variety of diseases[72,147]. However, there are limitations to the information that can be obtained from molecular approaches alone, and the isolation of organisms is required to define the roles of specific bacteria in causing or maintaining healthy and disease states. Culture also determines the viable population in a community, while most molecular methods do not distinguish between DNA obtained from live or dead cells. Furthermore, culture using selective media allows for the growth and detection of less abundant bacteria that may be missed by insufficient sequencing depth in culture-independent studies[134].

A few recent studies have attempted to characterize the culturable human gut microbiota by combining culture with next-generation sequencing[148], but only a maximum of 50% of the Operational Taxonomic Units (OTUs) detected by 16S rRNA gene sequencing of fecal samples were cultivable[130,131]. Lagier *et al.* developed a culturomics method, which used 212 conditions for the cultivation of the fecal microbiota. However, bacterial colonies were identified with MALDI-TOF mass spectrometry, making it difficult to directly compare to OTUs obtained from

pyrosequencing in order to determine the culturable proportion of the microbiota[134]. 174 novel bacterial species were identified in this study, demonstrating that culture is still a valuable method for exploring the gut microbiome. Another study by Rettedal *et al.* tested several culture conditions to capture a representative proportion and maintain the overall community structure of the gut microbiota[131]. From 10 media, 88% of family level taxonomic groups were recovered, and interestingly, 40% more OTUs were found by culture than by culture-independent methods, demonstrating the potential advantages of culture for capturing microbial diversity in the gut. Another study has shown that personalized culture conditions, although only representing 50% of the phylotypes observed from culture-independent sequencing, could colonize germ-free mice in a similar manner to the complete fecal samples[130].

In this study, a method of culture-enriched molecular profiling[149], which combines extensive bacterial culture with 16S rRNA gene sequencing (Fig. 1A), was used to investigate the proportion of the microbiota that was readily cultured from the gastrointestinal tract. Using 66 culture conditions and applying the same 16S rRNA gene sequencing method to both the cultured community and the fecal samples, we demonstrate that the majority of OTUs could be detected through culture-enriched molecular profiling and culture detected greater diversity than culture-independent methods. Culture-enriched molecular profiling was further applied to the targeted culturing of Lachnospiraceae isolates.

Materials and Methods

Sample collection

This study was approved by the Hamilton Integrated Research Ethics Board and donors provided consent prior to participation. Healthy volunteers (HV1-7) had no gastrointestinal (GI) symptoms and did not use antibiotics within six months of the study. IBS patients (IBS1-4) were diagnosed as diarrhea-predominant or mixed subtype based on Rome III criteria[150], and recruited from the GI Clinical Investigation Unit at McMaster University.

Fresh fecal samples were transferred to sterile specimen containers immediately following defecation and stored in airtight bags containing an anaerobic pouch (Oxoid, UK) and ice-pack until transfer to an anaerobic chamber (5% CO_2, 5% H_2, 90% N_2; Shel Labs, Cornelius, OR), which occurred within 1-5 hours of collection. Afterwards, all work was completed inside an anaerobic chamber. Samples were mechanically mixed with a sterile spatula and cultured as described below.

Culturing of fecal samples

Inside an anaerobic chamber, 0.1 g of fecal sample was diluted in 900 µl of pre-reduced brain heart infusion (BHI) broth (BD, Sparks, MD) with 0.05% L-cysteine hydrochloride hydrate (10^0 dilution). 100 µl of 10^{-3} and 10^{-5} dilutions were plated on pre-reduced 100 mm agar plates (media types listed below). One set of media was incubated at 37^{o}C for 5 days in an anaerobic chamber and another set was incubated at 37^{o}C in 5% CO_2 for 3 days. After incubation, colonies were collected from each plate by adding 1 ml BHI broth

and scraping the surface of plates with a cell scraper. Both dilutions (10^{-3} and 10^{-5}) of each media were combined and 500 µl of the harvested colonies was frozen in 10% skim milk at -80^{o}C as stocks and 500 µl was used for DNA extraction as described below.

Media used in this study included (Table S1): brain heart infusion (BHI) agar (BD), BHI + 0.5 g/L L-cysteine hydrochloride hydrate, 10 mg/L hemin and 1 mg/L Vitamin K (Supplement Set A), BHI + 10 mg/L colistin sulphate and 5 mg/L naladixic acid (Supplement Set B), BHI + Supplement Set A and B, BHI + Supplement Set A and 20 mg/L gentamycin, BHI + Supplement Set A and 1% propionic acid, 0.2X BHI or M9 minimal media (BD) + one of: 1 g/L inulin, 0.5 g/L pectin, 0.5 g/L cellulose, 0.5 g/L mucin or 0.5 g/L starch, Bifidobacterium Selective Media (Fluka, St. Louis, MO), Gut Microbiota Medium[130], phenylethyl alcohol agar (BD) + 5% sheep's blood (Cedarlane, Canada), Bacteroides Bile Esculin agar[151], Actinomyces Isolation agar (BD), Columbia blood agar (BD) + 5% sheep's blood, colistin naladixic agar (BD) + 5% sheep's blood, McKay agar[152], chocolate agar (BD), tryptic soy agar (BD) + 5 g/L yeast extract (BD) + Supplement Set A, fastidious anaerobe agar (Neogen, Lansing, MI), deoxycholate agar[151], MacConkey agar (BD), kanamycin vancomycin laked blood agar[151], mannitol salt agar (BD), de Man Rogosa Sharpe agar (BD) and cooked meat broth (Fluka) + 1.2% agar. All media was prepared following manufacturer's instructions or as previously described unless otherwise mentioned.

For experiments showing that cultured communities were represented by viable organisms (Fig. S4), 100 mg of frozen IBS1 sample was diluted 1:10 with 70% isopropanol and heated to 70^{o}C for 1 min to sterilize (No Growth samples) or DNA

extracted from IBS1 sample was used (DNA samples). The samples were diluted to 10^{-5} with BHI with 0.05% L-cysteine as described above, and cultured anaerobically on phenylethyl alcohol agar (PEA) and M9 + cellulose plates for 1hr at 37°C before plates were scraped and DNA was extracted as described below.

For the comparison of fresh and frozen samples, one aliquot was immediately cultured and another was stored at -80°C for up to one year (IBS1 & HV2 – 4 months, IBS3 – 1 year). To compare anaerobic and aerobic samples, fecal samples were divided during collection; one portion was stored with an anaerobic pouch at 4°C until culturing and the other was stored at 4°C without an anaerobic pouch (aerobic sample).

DNA extraction and 16S rRNA gene sequencing

DNA extraction and purification were performed as previously described[153]. Briefly, 500 µl of harvested colonies or 0.1 g of fecal sample was mechanically homogenized with 0.2 g of 0.1 mm glass beads (MoBio, Carlsbad, CA) (additional 0.2 g of 2.8 mm glass beads added to fecal samples) in 800 µl of 200 mM $NaPO_4$, pH 8 and 100 µl guanidine thiocyanate-EDTA-*N*-lauroyl sarcosine. Enzymatic lysis with 50 µl lysozyme (100 mg/ml), 50 µl mutanolysin (10 U/µl) and 10 µl RNase A (10 mg/ml) for 1h at 37°C was followed by the addition of 25 µl 25% sodium dodecyl sulfate (SDS), 25 µl Proteinase K (20mg/ml,) and 75 µl 5M NaCl, and further incubated for 1h at 65°C. Supernatants were collected and DNA extracted with phenol-chloroform-isoamyl alcohol (25:24:1, Sigma, St. Louis, MO) and further purified using DNA Clean and Concentrator – 25 columns (Zymo, Irvine, CA) as per manufacturer's instructions. Isolated DNA was stored at -20°C.

PCR amplification of the V3 region of the 16S rRNA gene was performed as previously described[154] with the following modifications: a 60 µl reaction containing 1.25 mM $MgCl_2$, 2.5 mM of each dNTP, 100 nM of each barcoded primer and 1.25 U Taq was divided into 3 x 20 µl reactions for amplification. PCR conditions consisted of an initial denaturation at 94°C for 2 min, 30 cycles of 94°C for 30 sec, 50°C for 30 sec, 72°C for 30 sec, followed by a final elongation at 72°C for 10 min. Purified PCR products were sequenced using the Illumina MiSeq platform by the McMaster Genome Facility (Hamilton, ON, Canada).

Sequence processing and analysis

16S rRNA gene sequence processing was completed as previously described[153]. In brief, Illumina sequence reads were trimmed to the forward and reverse primers of the V3 region with Cutadapt[155] and paired-end sequences were aligned with PANDAseq[156]. OTUs were binned at 97% similarity using AbundantOTU[157] and taxonomy was assigned using the Ribosomal Database Project (RDP) classifier[158] against the Greengenes reference database (Feb. 4th, 2011 release)[159] using Quantitative Insights Into Microbial Ecology (QIIME)[160]. Unassigned OTUs and singletons were not included. The total number of reads for the culture-enrichment experiment was 30 581 472. For analysis of the culture-enriched experiments, only OTUs with at least 10 sequence reads were included and considered cultured since the number of reads between replicates were less reproducible below this depth (data not shown).

Relative abundance taxonomic summaries, beta diversity and rarefactions were completed with QIIME. The 16S rRNA gene phylogeny depicting OTUs from culture-independent, culture-enriched or both methods (Fig. 2B) was created by pruning the 4 Feb 2011 release of the Greengenes phylogeny clustered to 97% similarity. This phylogeny was pruned using QIIME and visualized using GraPhlAn v0.9.7[161].

Lachnospiraceae isolation and identification

Freshly collected HV7 and IBS4 fecal samples were cultured anaerobically on BHI + 1g/L inulin (BHI+inu) and cooked meat agar (BEEF). 146 colonies were streaked for purity and single colonies were boiled in 5% Chelex for 15 min to lyse cells. Colony PCR to amplify the 16S rRNA gene was performed using 8f (5'-AGAGTTTGATCCTGGCTCAG-3') and 927r (5'-CCGTCAATTCCTTTRAGTTT-3') primers. PCR conditions consisted of an initial denaturation at 94^{o}C for 2 min, 32 cycles of 94^{o}C for 1 min, 56^{o}C for 1 min, 72^{o}C for 2 min, followed by a final elongation at 72^{o}C for 10 min. PCR products were Sanger sequence by Beckman Coulter Genomics (Danvers, MA) using the 8f primer. Taxonomic assignment of isolates were obtained with the online RDP Classifier[158] using 16S rRNA training set 10 with a confidence threshold of 80%. 16S rRNA gene sequences of the isolates were compared against the Lachnospiraceae consensus sequences from the Human Microbiome Project's (HMP) Most Wanted List (obtained on Nov. 18, 2014) using Megablast in Geneious v5.6.4[162].

Lachnospiraceae phylogeny

The 16S rRNA sequence of 107 Lachnospiraceae Type Isolates obtained from RDP[158] (on Nov. 12, 2014) were aligned using MUSCLE v3.8.31[163] (Edgar, 2004). jModelTest v2.1.1[164] was used to determine that the data most appropriately fit the Generalized Time-Reversible (GTR) model with invariable sites and a discrete gamma distribution. A maximum likelihood molecular phylogeny of these isolates was obtained using raxmlGUI v1.3.1[165] implementing RAxML v8.0.26[166] with 100 bootstrap replicates. pplacer v1.1[167] was used to add the 16S rRNA gene sequences of 79 cultured Lachnospiraceae isolates, obtained as described above. The resulting phylogeny was visualized and midpoint rooted using FigTree v1.4.0[168].

Results

In this study, the cultivable proportion of the human gut microbiota was determined from the fecal samples of 2 healthy donors and 3 patients with Irritable Bowel Syndrome (IBS). IBS samples were included since the culturable bacterial community may be affected by health status, and IBS is a common intestinal disorder in which the microbial composition of the gut is altered and implicated in the pathogenesis of the disorder[169]. Using a method of culture-enriched molecular profiling[149], anaerobic fecal samples were cultured using 33 media, and incubation of plates anaerobically and aerobically resulted in 66 culture conditions (Fig. 1A). The media used were comprised of commercially available components, and without the requirement of complex additions like rumen fluid or fecal extracts used in previous studies[134,170], and were chosen based on

previous culture-enriched molecular profiling studies, with additional media added for specific cultivation of the gut microbiota including Bacteroides Bile Esculin, Bifidobacterium Selective media, Gut Microbiota medium[130], and the addition of prebiotics and resistant carbohydrates[170] (Table S1). The cultivable portion of the fecal microbiota was determined by pooling all colonies from each type of media (per fecal sample) and performing16S rRNA gene sequencing on each plate pool individually, referred to as culture-enriched molecular profiling. A portion of the pooled colonies was reserved for future bacterial isolation. Since the same DNA extraction and sequencing methods were used for both the cultured plate pools and the fecal sample, the OTUs from the culture could be directly compared to the OTUs from culture-independent sequencing of the original fecal sample to determine which OTUs were cultured on each type of media.

Culture-enriched molecular profiling detects the majority of OTUs in fecal samples

An average of 76% of all OTUs observed in culture-independent sequencing of the fecal samples was also detected by culture-enrichment (Fig. 1B). The largest proportion of OTUs cultured was 93% from the Healthy Volunteer 1 (HV1) sample. No obvious differences were observed in the proportions of OTUs cultured between IBS and healthy samples (an average of 79% and 73% of all OTUs were cultured, respectively). Using thresholds of 0.01% and 0.1% (in the culture-independent sequencing), an average of 81% and 95% of OTUs were cultured, respectively (Fig. 1B), indicating that culture-enriched molecular profiling could detect the majority of the abundant members of the

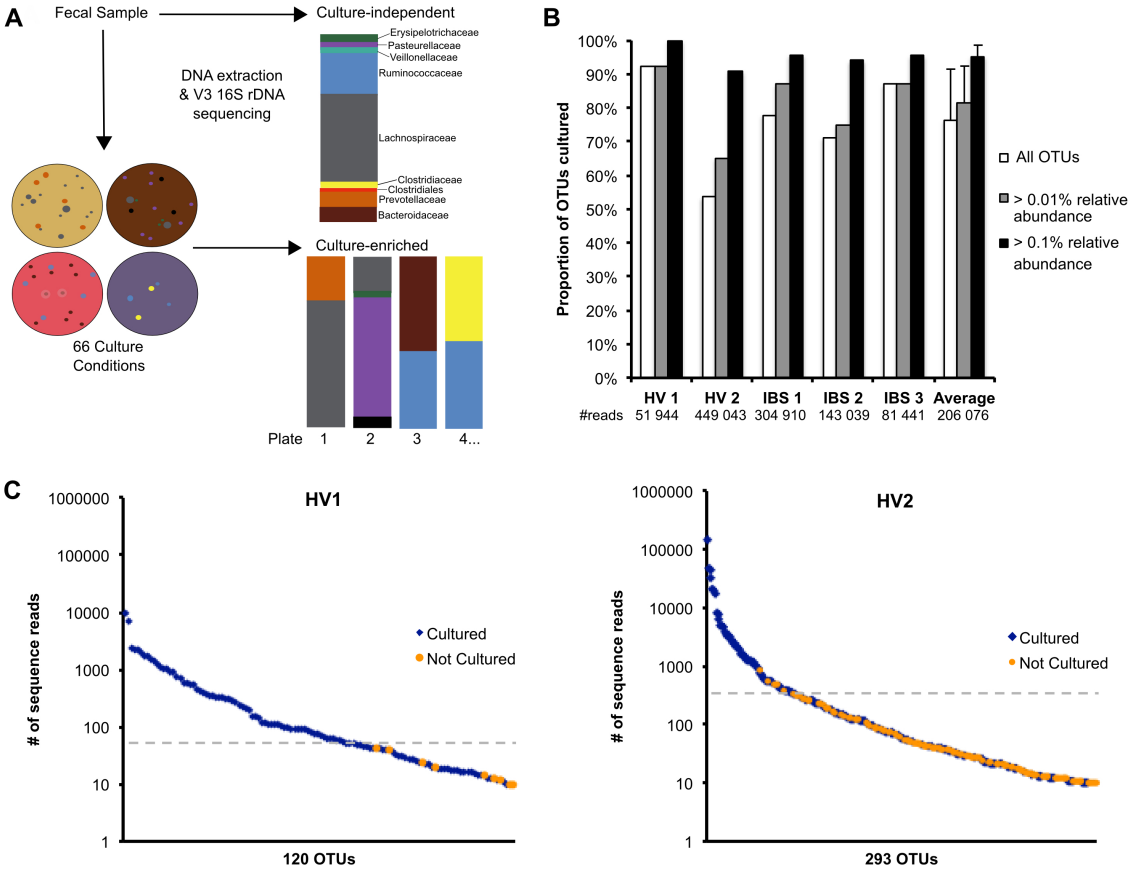


Figure 1. Culture-enriched molecular profiling captures the majority of OTUs. A) Schematic of culture-enriched molecular profiling. Fecal samples were cultured using 66 conditions and colonies were harvested from each plate. 16S rRNA gene sequencing of DNA extracted from colonies was compared to sequencing of the fecal sample. B) OTUs from 16S rRNA gene sequencing of culture-enriched fecal samples were compared to culture-independent sequencing. The number of reads in the culture-independent sequencing of each sample is below the x-axis. The average proportion of OTUs cultured from each sample is shown, error bars represent SD. C) All OTUs from culture-independent sequencing of HV1 and HV2 samples were ranked by abundance and compared to OTUs detected by culture-enriched sequencing. Each point represents one OTU and dotted line indicates 0.1% relative abundance in the culture-independent sequencing.

fecal community. A comparison of several OTU picking methods, including AbundantOTU[157] and UCLUST[171] against different reference databases, resulted in similar proportions cultured for each sample, indicating that these results were robust to

the method of OTU clustering (Fig. S1). However, the proportion of OTUs cultured was affected by culture-independent sequence depth; as the number of culture-independent sequence reads increased, the percent of OTUs cultured decreased (Fig. S2). Shallow sequencing depth (< 5000 reads) in culture-independent sequencing, as used in previous studies[130], would artificially inflate the proportion of the cultured microbiota due to incomplete culture-independent sampling of fecal samples. For each sample, ranking the OTUs from culture-independent sequencing by abundance reveals a few highly abundant OTUs and a long tail distribution of lesser abundant OTUs, as expected in human gut microbiota communities[172] (Fig. 1C, S3). Above 0.1% relative abundance, all OTUs from HV1 were recovered by culture and only 5 OTUs from HV2 were not cultured (Fig. 1C). Uncultured OTUs were of low abundance (< 0.8% relative abundance) in the culture-independent sequencing. Twelve OTUs with relative abundances greater than 0.1% were not cultured from the donor samples, and included Cyanobacteria, Clostridia, Mollicutes and Bacteroidetes. However, 9 of these 12 OTUs were detected by culture in the other fecal samples.

Several bacterial genera were present at less than 1% abundance in culture-independent sequencing but recovered by culture at higher abundances on plates (Fig. S4E, S5), thus demonstrating the advantages of culture for recovering low abundance bacteria from the GI microbiota. Comparison of each plate community to the culture-independent fecal samples showed that no single media accurately represented the composition of the fecal sample, indicating that OTUs detected by culture represent viable organisms and not DNA deposited on plates (Fig. S4A-D). To further demonstrate

this, a sterilized fecal sample from the IBS1 donor and DNA extracted from the same sample was plated on 2 media, representing cultured communities most similar and most different to the composition of the fecal sample (Fig. S4B-D). The composition of the "No growth" and "DNA" controls were very different than either the fecal sample or the pooled plate. Indeed, the negative growth controls from both types of media were most similar to each other, providing evidence that 16S sequencing of plates represent growth of viable cells, and not DNA or dead cells deposited from the fecal sample. Goodman *et al.* have previously shown that sequencing of pooled colonies cultured from a fecal sample diluted 10 000-fold resulted in less than 2% of the sequences being derived from non-viable cells[130].

The minimal conditions for culture-enrichment reflects the inter-individual heterogeneity of the human gut microbiota

The cultured communities on each media differed between samples, reflecting the inter-individual heterogeneity of the microbiota[144]; thus, a minimal set of media that could capture the majority of OTUs in all samples would be difficult to predict, similar to results previously reported for the airway microbiome[149]. By determining the minimum culture conditions needed to recover all culturable OTUs present at greater than 0.1% in the fecal samples, we observed that the set of culture conditions required was different for all 5 fecal samples (Fig. S6). Interestingly, 23 of the 33 anaerobic culture conditions were required across the 5 samples, indicating that a majority of the media used was non-redundant, and additional fecal samples may necessitate the use of the other media. There

was a distinct difference in the bacterial communities recovered from anaerobic incubation compared to aerobic growth (Fig. S5), *Escherichia* was the most abundant OTU on the majority of aerobic media used so future studies could reduce the number of aerobic conditions.

The cultured community is affected by storage conditions of the fecal sample

Since immediate culturing of anaerobic fecal samples may not always be feasible, we determined the affect of freezing at -80^{o}C or exposure to ambient oxygen on the culturable fecal community. Fresh, anaerobic fecal samples were cultured as done previously, and the culture-enriched profiles were compared to those from fecal samples that were frozen or exposed to oxygen (Fig. S7A-E). All conditions tested showed that the cultured microbiome profile was different after freezing, with Bray-Curtis dissimilarity distances greater than 0.3 (Fig. S7A,D), while fecal samples were differentially affected by aerobic exposure, with Bray-Curtis dissimilarity distance ranging from 0.15 to 0.65 (Fig. S7B,C,E). Since the composition of the fecal samples were different from each other, and each media used selected for different bacteria, no abundant bacterial groups were identified as consistently altered after storage. Optimal recovery of bacteria required stool samples to be kept anaerobic immediately after collection and plated without freezing, as exposure to oxygen and storage at -80^{o}C altered the culturable community.

Culture-enriched molecular profiling captures greater bacterial diversity than culture-independent sequencing

Importantly, we observed that culture-enrichment recovered more OTUs than culture-independent sequencing. Taking into account all OTUs detected in the 5 samples by both methods, 1051 OTUs (67% of the total) were detected only by culture-enrichment, 390 OTUs (25%) were found by both methods, and 118 OTUs (8%) were detected only by culture-independent sequencing. For each sample, more OTUs were observed by culture-enriched sequencing than culture-independent sequencing (Fig. 2A). This suggests that even with deep sequencing of a fecal sample (average depth = 206 076 reads), culture detects greater bacterial diversity compared to culture-independent methods. The OTUs that were only detected by culture were mainly distributed across 6 phyla: Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria and Tenericutes. These low abundance organisms may include bacteria from the colonic mucosa and other sites in the gastrointestinal tract where they may be present at higher local abundance but at too low abundance in stool to be detected by the culture-independent sequencing depths used in this study. OTUs that were culture-only or detected by both methods were evenly distributed across the bacterial taxonomy (Fig. 2B), indicating that cultured OTUs represent members of the fecal microbiota and not plate contamination. Uncultured OTUs included a Tenericutes branch and a few Firmicutes clades.

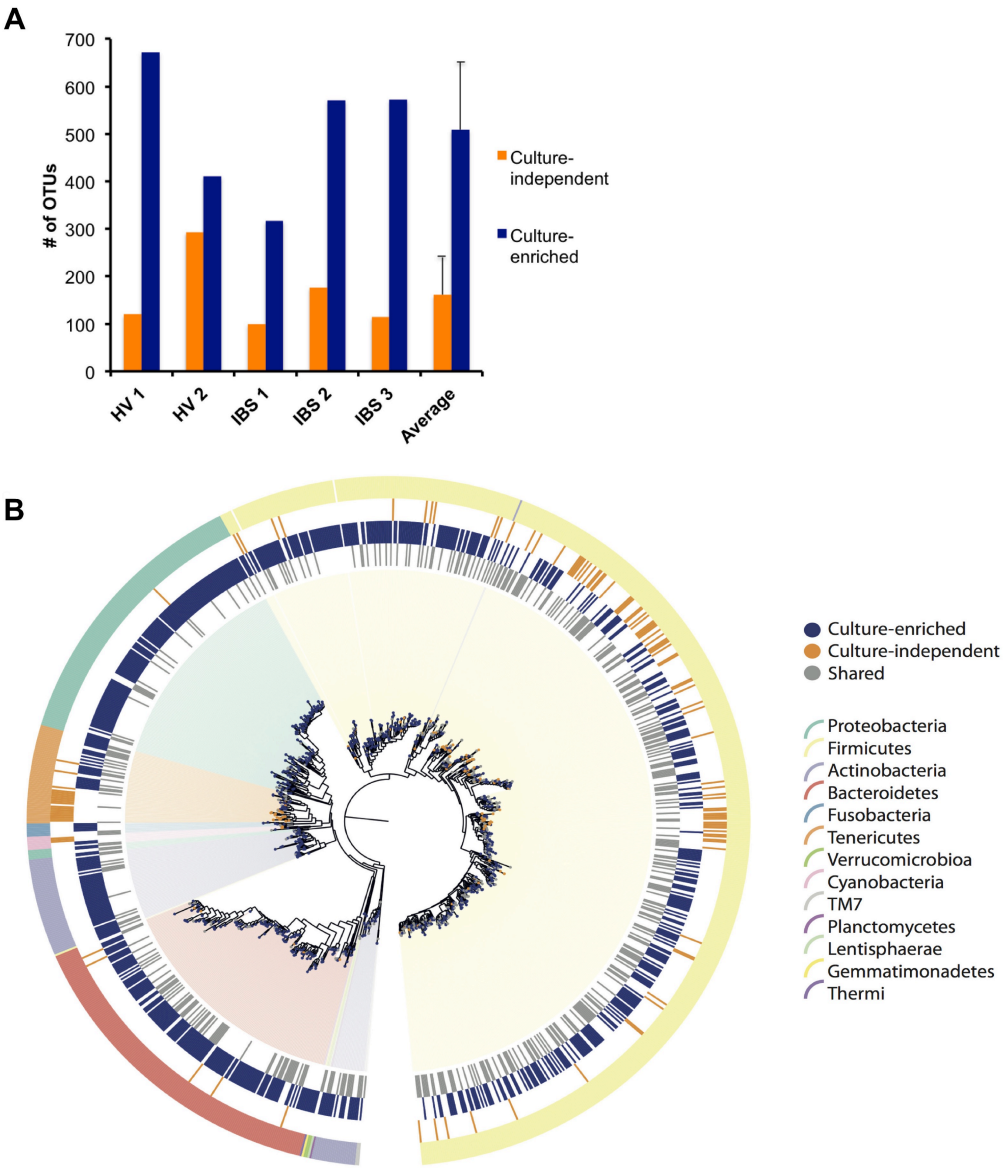


Figure 2. Culture-enriched molecular profiling detects more OTUs than culture-independent sequencing. A) A comparison of the number of OTUs detected by culture-independent and culture-enriched sequencing for each fecal sample. The average number of OTUs detected by each method is shown, error bars represent SD. B) Phylogeny of Greengenes 16S rRNA gene sequences to which OTUs were assigned (clustered to 97% sequence similarity). The branch nodes and 3 innermost rings are labeled based on detection by culture-independent, culture-enriched, or both methods. The outer ring and background are colored by phyla, as indicated in the legend.

Targeted culturing of Lachnospiraceae isolates

Culture-enriched molecular profiling was applied to additional healthy (HV3) and IBS (IBS4) fecal samples for the targeted culturing of Lachnospiraceae isolates to demonstrate that a bacterial group of interest could be isolated from fecal samples when the bacterial composition was not known by using results of culture-enriched profiling of previous samples (Fig. S5). There is extensive Lachnospiraceae diversity in the human gut microbiota, but this prevalent and abundant family is poorly represented by reference genomes[123]. From analysis of the culture-enriched profiling of previous fecal samples, it was determined that communities grown on anaerobic BHI+inu and BEEF agar were dominated by Lachnospiraceae (Fig. S5). Consistent with our culture-enrichment results, inulin has been shown to increase growth of butyrate-producing bacteria, including Lachnospiraceae, both *in vivo*[173] and *in vitro*[174]. HV3 and IBS4 fecal samples were cultured anaerobically on BHI+inu and BEEF, and colonies were isolated (as opposed to plate pooling). Sanger sequencing of the 16S rRNA gene of 146 isolated colonies resulted in the identification of 79 Lachnospiraceae isolates, including representatives of the genera *Blautia*, *Marvinbryantia*, *Ruminococcus*, *Dorea*, *Eubacterium*, *Anaerostipes*, *Clostridium* and *Coprococcus* (Table S2). Additionally, 18 isolates had less than 97% similarity to Ribosomal Database Project (RDP) reference isolates at the genus level[158], as supported by the phylogenetic distance of these isolates to the 107 RDP Lachnospiraceae reference sequences (Fig. 3). These 18 isolates may be novel species not currently represented by the RDP Type Isolates. Comparison of the 16S rRNA gene sequences of the Lachnospiraceae isolates to the Human Microbiome Project's (HMP) "Most Wanted"

taxa resulted in a match to 12 sequences with 100% identity, including 4 organisms of medium priority (Table S3). The "Most Wanted" taxa represent uncultured but abundant

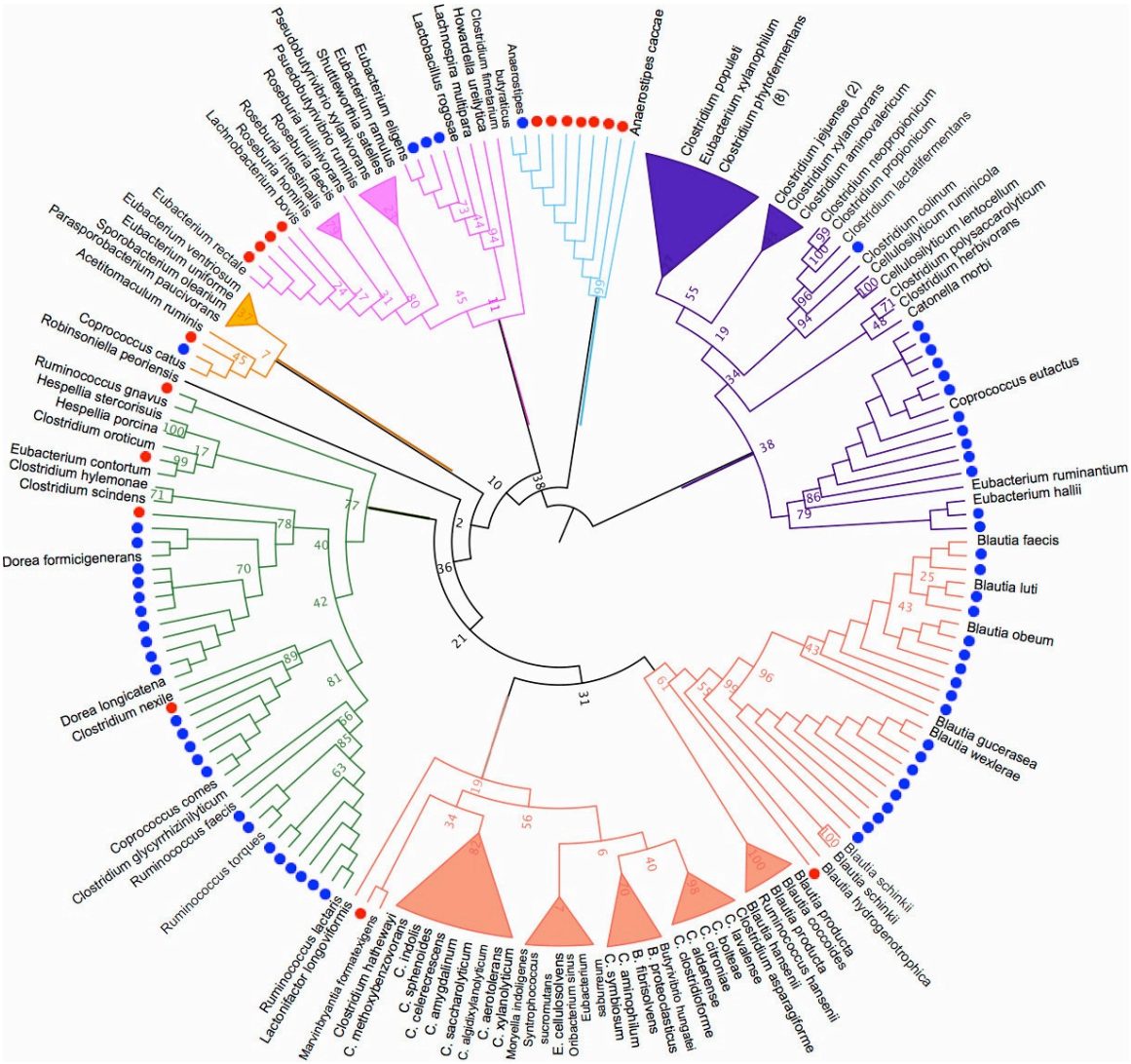


Figure 3. 16S rRNA maximum-likelihood tree of 79 Lachnospiraceae isolates in relation to 107 RDP Lachnospiraceae type isolates. V1-V4 16S rRNA gene sequences of cultured Lachnospiraceae isolates were placed into a phylogeny created with RDP Lachnospiraceae type isolates. Blue represents cultured isolates; red represents cultured isolates with <97% similarity to RDP reference isolates. Branch labels show bootstrap values. Branch colors highlight clades for visual aid. C - *Clostridium*, B – *Butyrivibrio*. E – *Eubacterium.*

and prevalent organisms from the HMP 16S rRNA gene sequencing data set[122]. Isolation of novel representatives of the gut microbiome and whole genome sequencing of these organisms will add critical information to reference genome collections, and lead to more accurate assignments for metagenomics, transcriptomics and proteomic studies.

Discussion

Although other studies have attempted to culture the gut microbiota using complicated methods such as chemostats[136] and microfluidic devices[175], we show that the use of extensive culture conditions and simple agar plates results in the comprehensive cultivation of bacteria from the gut microbiota. Using significantly fewer and less complex culture condtions than Lagier *et al.*[134], we were still able to capture the majority of OTUs seen by culture-independent sequencing. Indeed, we believe that this is the first study showing that > 50% of the OTUs from a fecal sample can be recovered by culture. We also show that culture revealed greater bacterial diversity than detected by culture-independent sequencing. These organisms, absent or underrepresented in culture-independent profiling, may contribute to host phenotypes so methods to detect and isolate this population are critical for future studies. We did not observe a difference in the proportion of OTUs that could be cultured between IBS and healthy fecal samples, however, our sample size was small and further studies should be carried out to confirm this. By applying the data generated from culture-enriched molecular profiling in the first part of this study, we were able to carry out the targeted isolation of Lachnospiraceae. This resulted in the recovery of up to 18 species not represented by a reference database

from only 2 fecal samples and 2 media conditions, demonstrating that culture is valuable for accessing novel members of the gut microbiota.

Although most fecal OTUs were cultured in this study, some OTUs likely contained more than one strain, and some strains may have been missed by culture; metagenomics may sort this out. Some OTUs were detected only by culture-enrichment or culture-independent sequencing, suggesting that both methods are complementary for studying the gut microbiota since the combination is more sensitive than either method alone. The uncultured OTUs may have been missed due to the following reasons: stochastic sampling if the bacteria were rare in the sample, DNA could have originated from dead cells, slower growing organisms may have been out-competed by less fastidious members of the community or the media used in this study may have been missing a nutrient required for growth or prevented growth of a bacterial partner critical for survival. This may also explain why some OTUs were not detected by culturing in one fecal sample but recovered from another donor.

Consistent with previous studies[131,134], we observed many OTUs that were only detected by culture. Certain bacterial groups may be under-represented by primers used in 16S sequencing of complex communities, such as stool, since poor primer alignment would result in these sequences being outcompeted during PCR of mixed community DNA[123]. However, these bacteria can be grown using selective media and subsequently detected by culture-enriched molecular profiling since there may be less competing DNA in 16S rDNA amplification of the plate pools. As previously reported, bacteria present at $<10^6$ cells per gram of feces is likely to be missed by sequencing depths used in most

microbiome studies[134], but the ability to grow less abundant bacteria on media would increase recovery of DNA for sequencing these rarer members of the population. Additionally, DNA from bacterial spores in stool may be underrepresented by 16S sequencing, depending on the method of DNA extraction[175], but germination of spores during culture would allow for increased detection of spore-forming populations.

Culture-enriched molecular profiling for the targeted recovery of specific bacterial groups will be a powerful tool for the study of the gut microbiota since bacteria of interest could be recovered after sequencing has revealed which media supports its growth, as represented in Fig. S5. Furthermore, since this method allows for the development of a bacterial community on a plate, it does not exclude syntrophic interactions that may be required for growth of some bacteria[176]. In addition, subsequent culturing of plate pools using combinations of antibiotics and other selective media components can be used to develop more refined culture conditions for specific bacteria[131]. The application of this method for the recovery of novel representatives of the gut microbiota will add to reference databases and lead to more accurate assignment for future "omic" studies. To advance the microbiome field beyond correlative investigations, and to test hypotheses generated from culture-independent studies, it is critical to isolate bacteria of interest from the gut microbiota, and culture-enriched molecular profiling allows for such isolation.

Conclusions

We demonstrate that the majority of the human gut microbiota can be captured by culture-enriched profiling, and this method can be applied to the recovery of specific bacterial groups. This method also highlights the ability of culture for recovering low abundance bacteria, and for revealing diversity that may be underrepresented by other methods. Isolation of bacteria will allow us to explore the therapeutic potential of bacterial products, which may directly affect the host or microbial community. Access to the cultured human microbiota will offer detailed functional characterization of bacteria and will facilitate the discovery of their biological activities during host-bacteria and inter-bacterial interactions in health and disease.

Abbreviations

OTU – operational taxonomic unit
GI – gastrointestinal
HV – healthy volunteer
IBS – irritable bowel syndrome
SD – standard deviation
RDP – Ribosomal Database Project
HMP – Human Microbiome Project
BHI – brain heart infusion agar
BEEF – cooked meat agar

Ethics approval

All participants signed informed consent before enrollment. This study was approved by the Hamilton Integrated Research Ethics Board (REB #11-445). This study conformed to the Helsinki Declaration.

Consent for publication

Not applicable

Availability of data and materials

The datasets (16S rRNA sequences) supporting the conclusions of this article have been deposited in the Sequence Read Archive (SRA) under the BioProject ID PRJNA321331.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from Crohn's and Colitis Canada awarded to MGS, and Canadian Institutes of Health Research to MGS, PB, and SMC. MGS is supported as a Canada Research Chair in Interdisciplinary Microbiome Research.

Author Contributions: JTL and IH performed the experiments. Volunteer and patient samples were provided by CHL, SCM, and PB. Data analysis was performed by JTL, FJW, and MGS. JTL, PB, and MGS designed the study and wrote the manuscript.

Acknowledgements

We would like to thank MI Pinto-Sanchez for recruiting IBS patients for this study and E Verdu for helpful advice regarding the manuscript.

Additional Data

Additional File 1: Table S1. List of media used for culture-enriched molecular profiling. (XLSX)

Additional File 2: Fig. S1. Comparison of OTU picking method on the proportion of OTUs cultured. **Fig. S2.** Comparison of sequencing depth on the proportion of OTUs cultured. **Fig. S3.** Cultured and uncultured OTUs ranked by abundance for IBS1-3 samples. **Fig. S4.** Comparison of cultured communities to culture-independent sequencing of fecal samples. **Fig. S5.** Heat map of family-level taxa abundances in culture-enriched and culture-independent sequencing. **Fig. S6.** Culture conditions required to capture the most abundant OTUs from each fecal sample **Fig. S7.** Comparison of culturing fresh fecal samples compared to frozen samples, and anaerobic fecal samples compared to aerobic samples. (DOCX)

Additional File 3: Table S2. Classification of Lachnospiraceae isolates cultured from Donor HV7 and IBS4 fecal samples. (XLSX)

Additional File 4: Table S3. HMP Most Wanted matches of Lachnospiraceae isolates from HV7 and IBS4 samples. (XLSX)

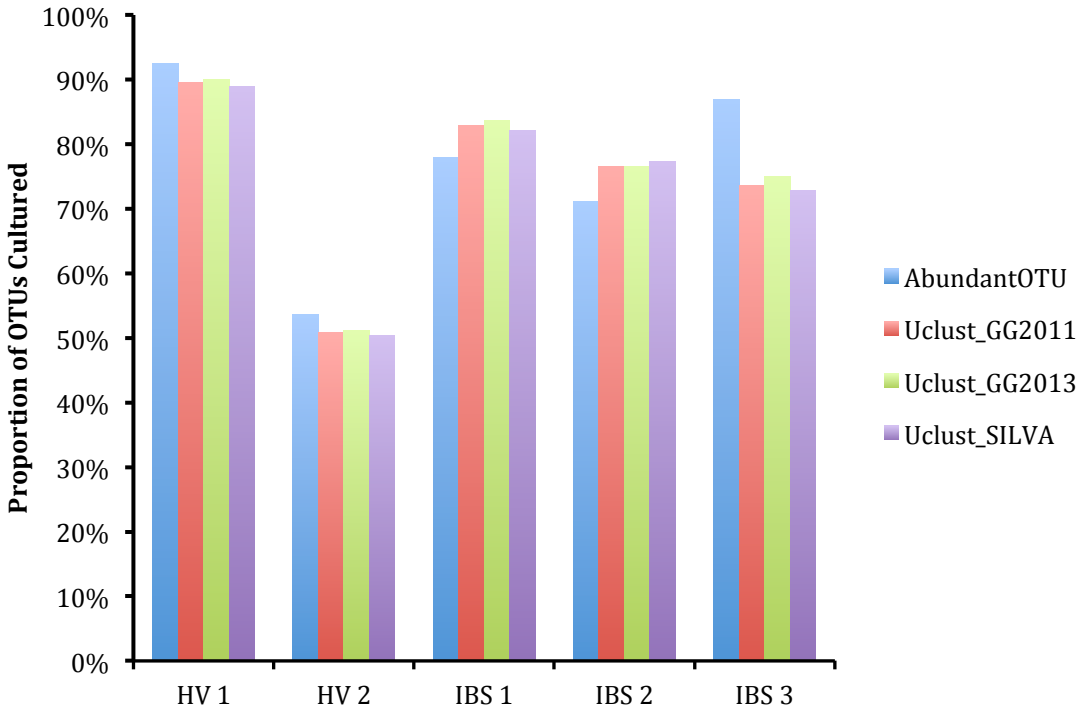


Figure S1. Proportion of OTUs culture is not dependent on OTU picking method. OTUs from 16S rRNA gene sequencing of culture-enriched fecal samples were compared to culture-independent sequencing. OTUs were picked using AbundantOTU, or Uclust using Greengenes 2011, Greengenes 2013 or SILVA as reference databases as seeds.

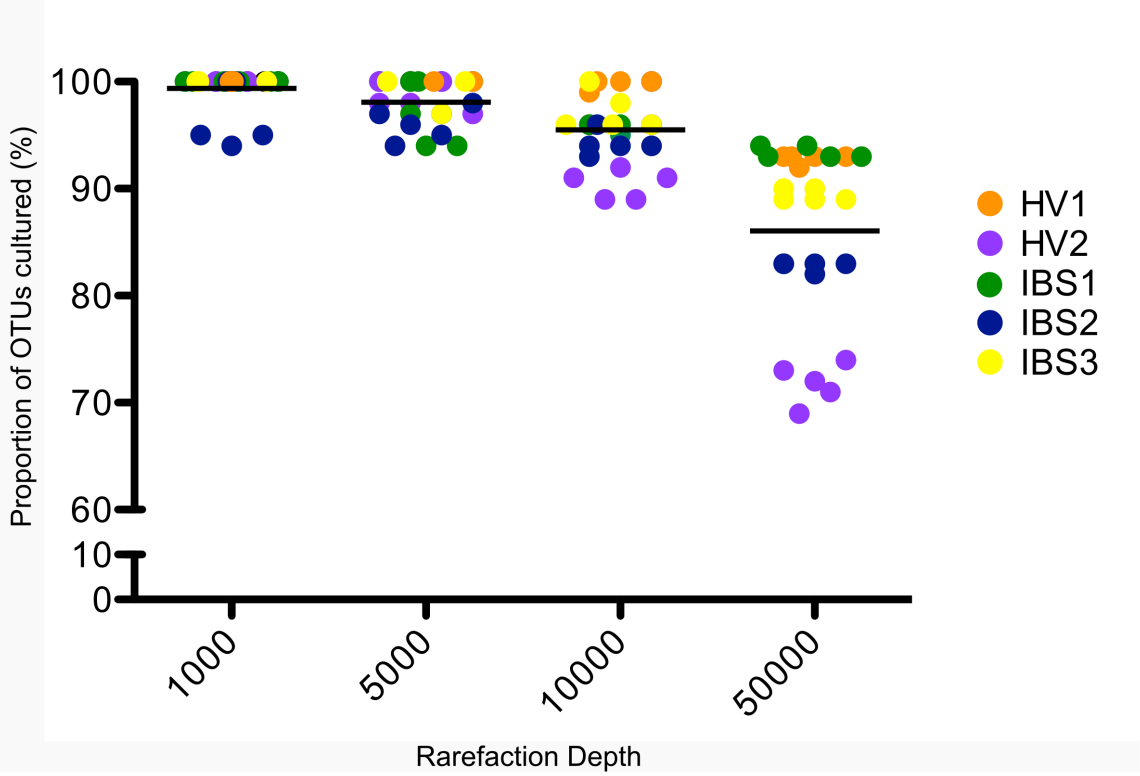


Figure S2. Proportion of OTUs cultured is dependent on culture-independent sequencing depth. Samples were rarified to 1 000, 5 000, 10 000 and 50 000 sequence reads to determine, *in silico,* the proportion of OTUs that would be cultured at each experimental sequencing depth. Each rarefaction was performed five times per sample, each represented as one point. Lines indicate the average proportion of OTUs cultured at each rarefaction depth.

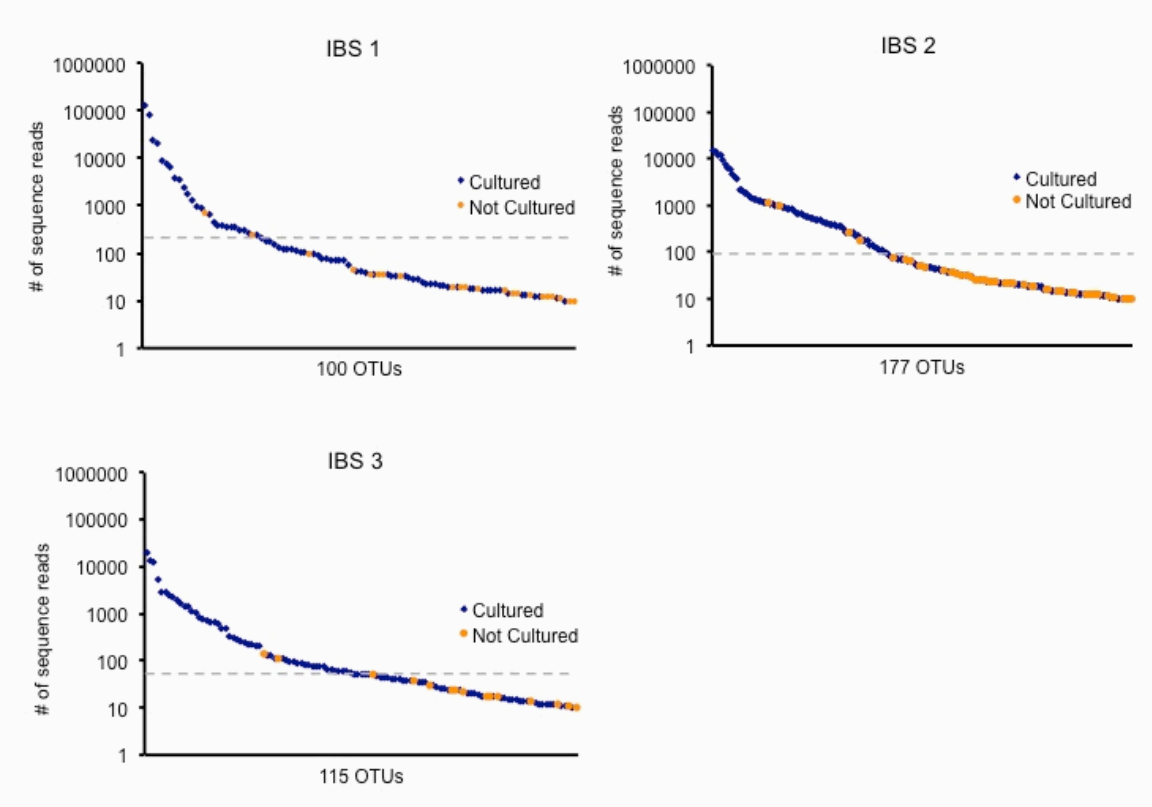


Figure S3. The majority of the uncultured OTUs are low abundance in culture-independent sequencing of the fecal samples. All OTUs from culture-independent sequencing of fecal samples from IBS 1-3 donors were ranked by abundance and compared to OTUs detected by culture-enriched sequencing to determine if each OTU was cultured or not cultured. Each point represents one OTU and the dotted line indicates a cutoff of 0.1% abundance in the culture-independent sequencing. The number of OTUs in each sample is indicated on the x-axis.

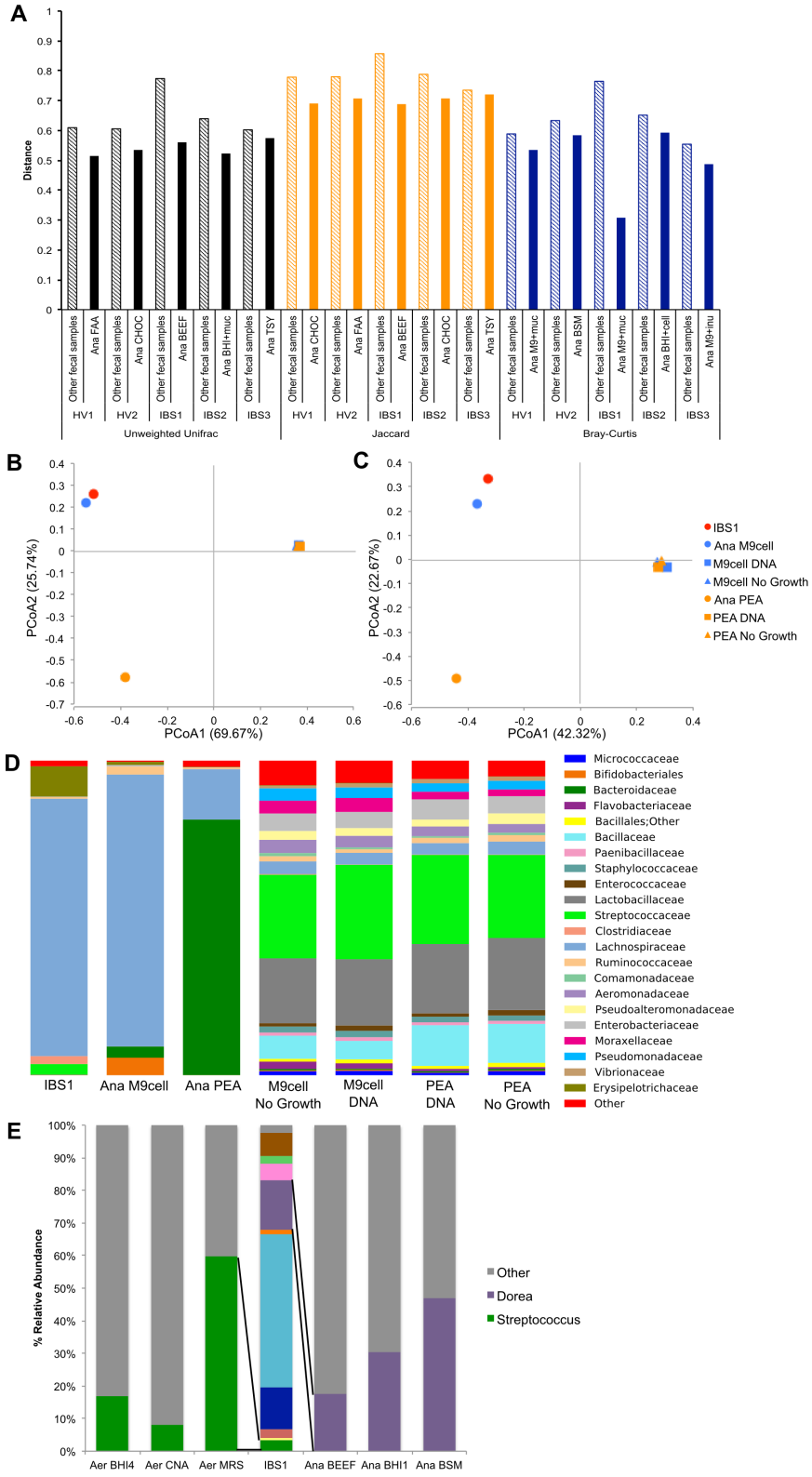

A
Distance
Unweighted Unifrac
Jaccard
Bray-Curtis
B
PCoA2 (25.74%)
PCoA1 (69.67%)
C
PCoA2 (22.67%)
PCoA1 (42.32%)
IBS1
Ana M9cell
M9cell DNA
M9cell No Growth
Ana PEA
PEA DNA
PEA No Growth
D
IBS1
Ana M9cell
Ana PEA
M9cell No Growth
M9cell DNA
PEA DNA
PEA No Growth
Micrococcaceae
Bifidobacteriales
Bacteroidaceae
Flavobacteriaceae
Bacillales;Other
Bacillaceae
Paenibacillaceae
Staphylococcaceae
Enterococcaceae
Lactobacillaceae
Streptococcaceae
Clostridiaceae
Lachnospiraceae
Ruminococcaceae
Comamonadaceae
Aeromonadaceae
Pseudoalteromonadaceae
Enterobacteriaceae
Moraxellaceae
Pseudomonadaceae
Vibrionaceae
Erysipelotrichaceae
Other
E
% Relative Abundance
Aer BHI4
Aer CNA
Aer MRS
IBS1
Ana BEEF
Ana BHI1
Ana BSM
Other
Dorea
Streptococcus

Figure S4. Cultured communities are not similar to fecal samples. A) Unweighted Unifrac, binary Jaccard and Bray-Curtis distances of culture-independent sequencing of 5 samples compared to their most similar cultured community (solid bars). For comparison, the distance between each fecal sample and samples from the other 4 donors is shown (hatched bars). Samples were rarified to 2900 sequences. B) Bray-Curtis PCoA, C) Unweighted Unifrac PCoA, and D) family-level taxonomic summaries of IBS1 culture-independent sample, anaerobic PEA and M9cell cultured communities, sterilized stool and IBS1 DNA plated on PEA or M9cell. Samples were rarified to 18 000 sequences for beta-diversity analyses. E) Taxonomic summaries for IBS1 and the relative abundances of *Dorea* and *Streptococcus* in different cultured communities. PEA – Phenylethyl alcohol agar, cell - cellulose

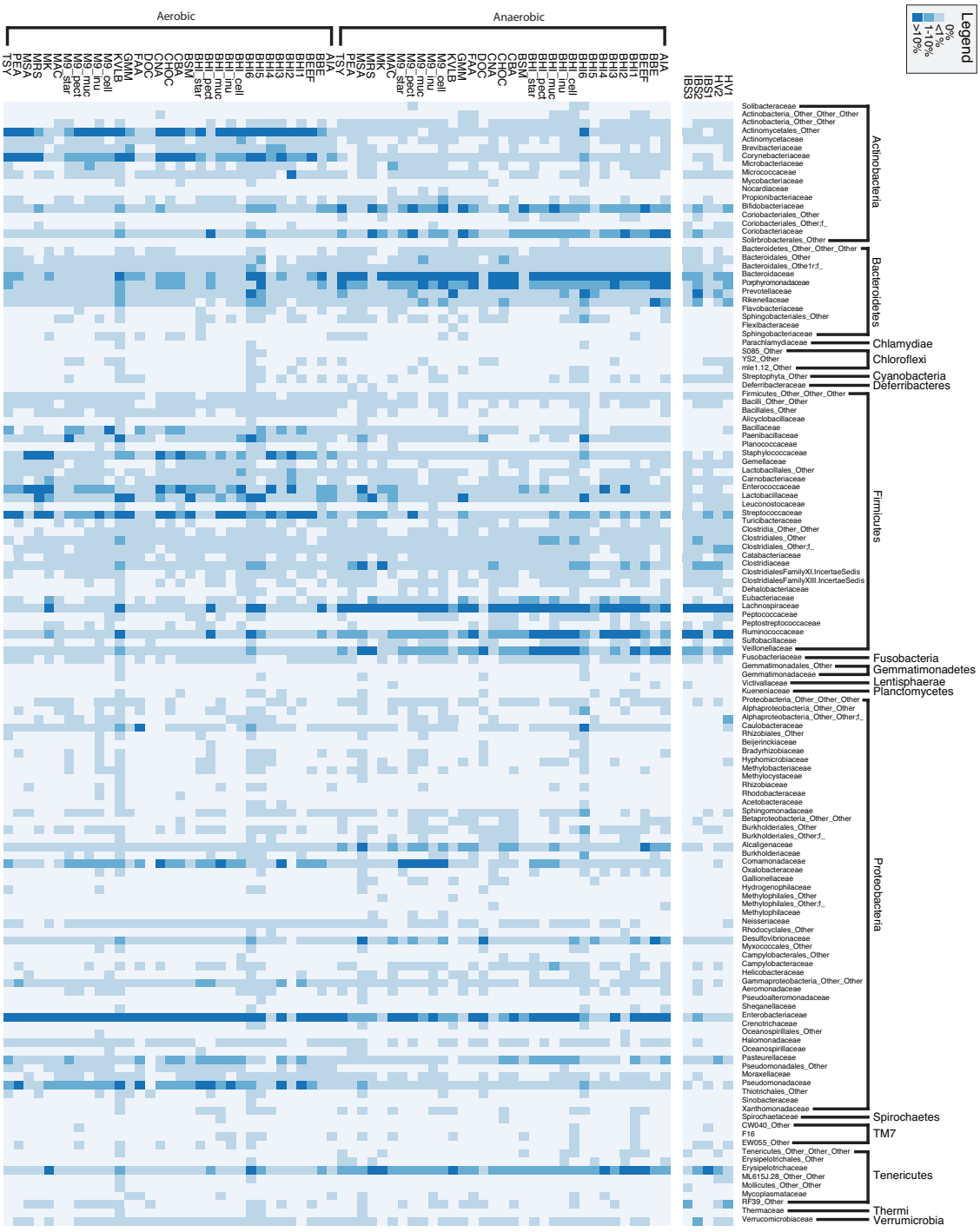


Figure S5. Heat map of family-level taxa abundances in each type of media and in culture-independent sequencing of fecal samples. Each row indicates either a donor fecal sample or a media condition. Each column represents family-level taxa, sorted by

phyla. Increasing grades of blue represent greater relative abundance, from 0% to greater than 10%. Refer to Table S1 for media used in this study.

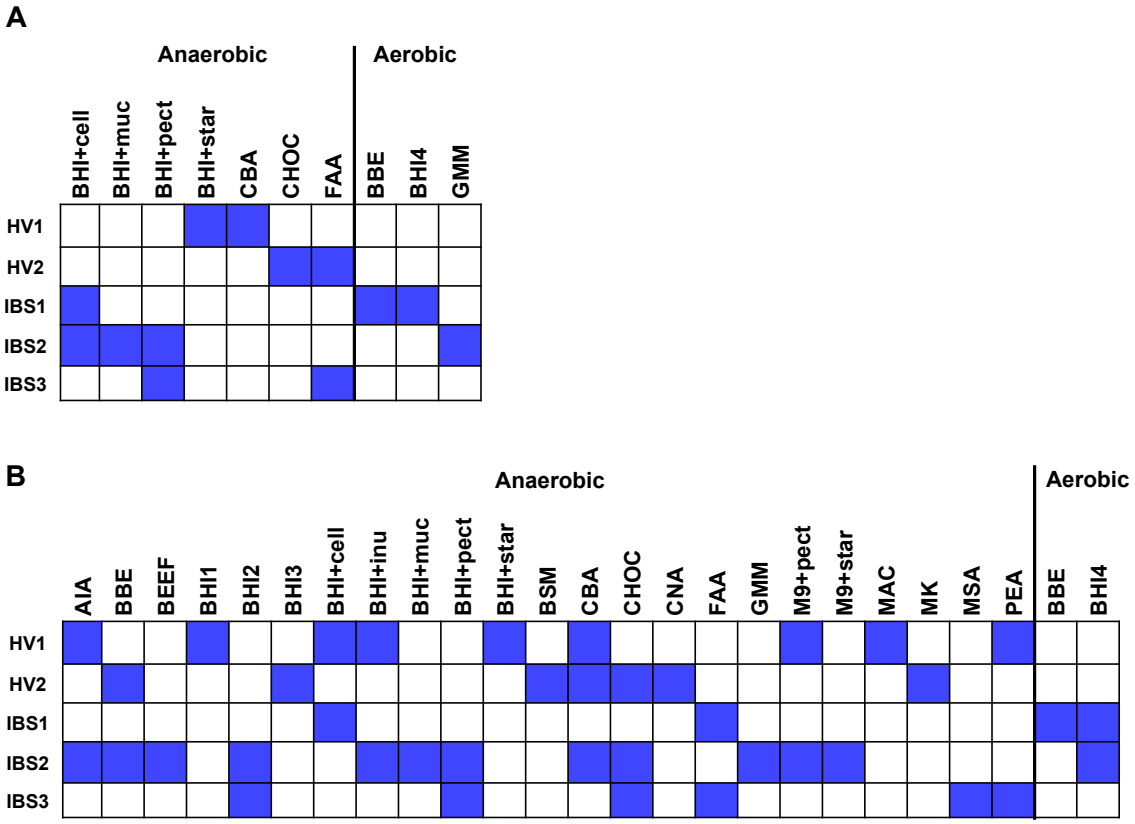
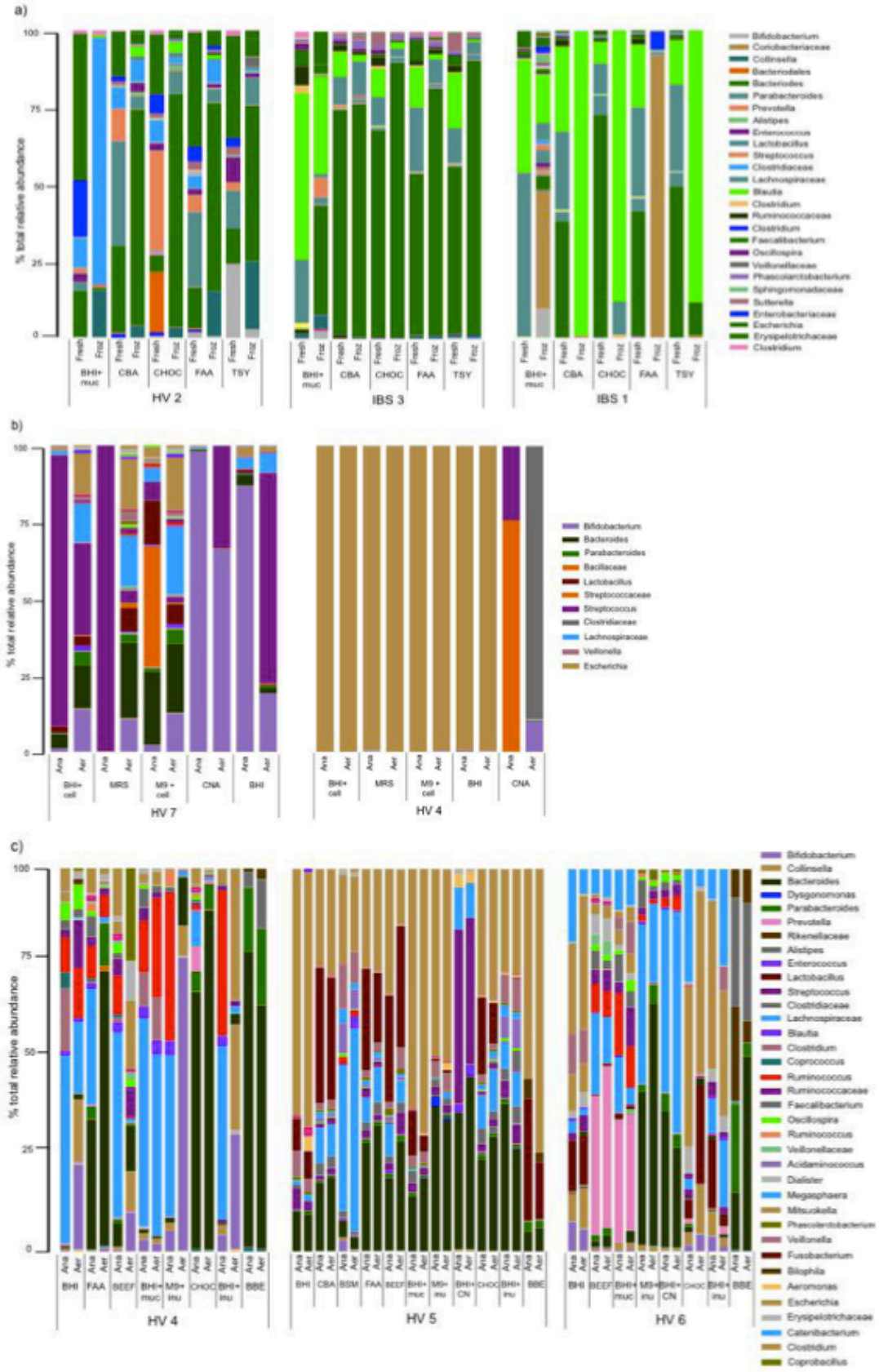


Figure S6. Unique set of culture conditions required to capture the OTUs from each fecal sample. Minimum culture conditions required to recover all culturable OTUs (defined as being at least 0.1% relative abundance in plate pools) present at A) greater than 1%, and B) greater than 0.1% relative abundance in culture-independent sequencing of five fecal samples.


a)
% total relative abundance
HV 2
IBS 3
IBS 1
b)
% total relative abundance
HV 7
HV 4
c)
% total relative abundance
HV 4
HV 5
HV 6

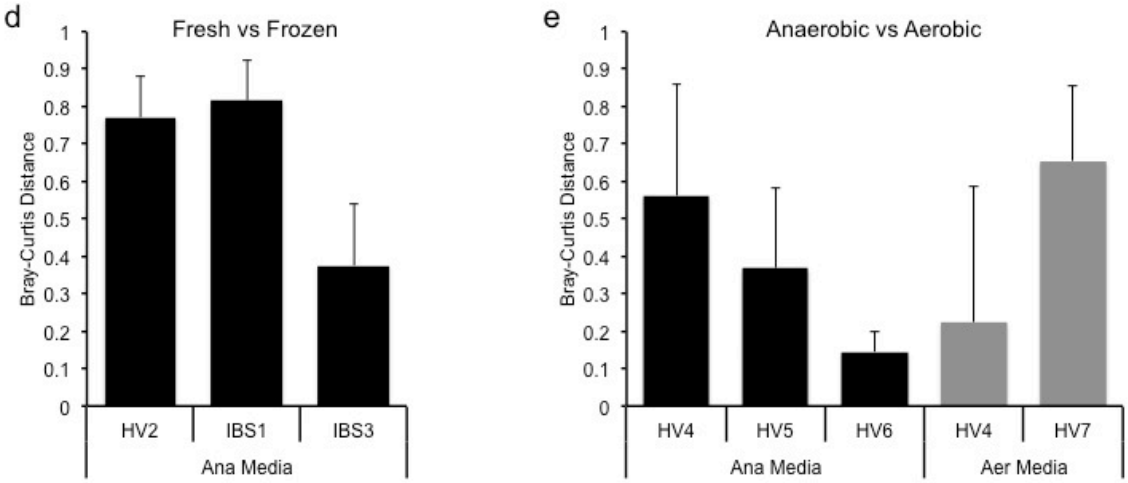


Figure S7. 16S rRNA gene sequencing taxonomic summaries of cultured fresh fecal samples compared to frozen samples, and anaerobic fecal samples compared to aerobic samples. A) Samples from 3 donors were cultured immediately after collection (Fresh) or after long-term storage at -80^{o}C (Frozen). B) Samples from 2 donors were divided into 2 aliquots, one was exposed to oxygen before culturing (Aer) and the other was maintained anaerobic until culturing (Ana) on aerobic media. C) Samples from 3 donors were divided into 2 aliquots, one was exposed to oxygen before culturing (Aer) and the other was maintained anaerobic until culturing (Ana) on anaerobic media. All bacterial groups present at greater than 2% abundance are included in the legend. D) Average Bray-Curtis distances between cultured fecal samples before and after freezing, cultured on 5 types of anaerobic media. 3 samples were cultured immediately after collection or after long-term storage at -80^{o}C. All samples rarified to 3932 reads. E) Average Bray-Curtis distances comparing cultured fecal samples, with (Aerobic) or without (Anaerobic) oxygen exposure. Samples were cultured on 5 types of aerobic media and 7 types of anaerobic media. All sample rarified to 8625 reads. Error bars represent S.D. BHI – brain heart infusion, muc – mucin, CBA – Columbia blood agar, CHOC – chocolate agar, FAA – fastidious anaerobe agar, TSY – tryptic soy yeast agar, cell – cellulose, MRS – de Man Rogosa Sharpe agar, CNA – colistin naladixic acid, BEEF – cooked meat agar, inu – inulin, BBE – bacteroides bile esculin agar, CBA – Columbia blood agar, BSM – bifidobacterium selective agar.

Table S1. Media used for culture-enriched molecular profiling

Media	Additional Supplements
Brain heart infusion agar (BHI1)	Supplement Set A[1] (BHI3) Supplement Set B[2] (BHI2) Supplement Set A + B (BHI4) Supplement Set A + 20 mg/L gentamycin (BHI5) Supplement Set A + 1% propionic acid (BHI6)
M9 minimal media or 0.2X BHI	One of: 1 g/L inulin (inu) 0.5 g/L pectin (pect) 0.5 g/L cellulose (cell) 0.5 g/L mucin (muc) 0.5 g/L starch (star)
Gut Microbiota Medium (GMM)	
Columbia blood agar (CBA) Chocolate agar (CHOC)	5% sheep's blood
Tryptic soy agar (TSY) Fastidious anaerobe agar (FAA) Cooked meat agar (BEEF)	Supplement Set A + 5 g/L yeast extract
Bifidobacterium Selective Media (BSM)	
Phenylethyl alcohol agar (PEA) Actinomyces isolation agar (AIA)	5% sheep's blood
Colistin naladixic acid agar (CNA)	5% sheep's blood
McKay agar (MK) Mannitol salt agar (MSA) de Man Rogosa Sharpe agar (MRS)	
Bacteroides bile esculin agar (BBE) Deoxycholate agar (DOC) MacConkey agar (MAC) Kanamycin vancomycin laked blood agar (KVLB)	

1) Supplement Set A consists of 0.5 g/L L-cysteine hydrochloride hydrate, 10 mg/L hemin and 1 mg/L Vitamin K

2) Supplement Set B consists of 10 mg/L colistin sulphate and 5 mg/L naladixic acid

Abbreviations used for media and supplements shown in parentheses.

Table S2. Classification of Lachnospiraceae isolates cultured from Donor HV7 and IBS4 fecal samples

Isolate	Donor & Media	Sequence Length (bp)	RDP Classification (Root[100%] Bacteria[100%] Firmicutes[100%] Clostridia[100%] Clostridiales[100%] Lachnospiraceae[100%])
Isolate1	IBS4 BEEF	746	Dorea[100%]
Isolate2	HV7 BEEF	703	Blautia[100%]
Isolate3	HV7 BEEF	776	Coprococcus[100%]
Isolate4	HV7 BEEF	728	Coprococcus[100%]
Isolate5	HV7 BEEF	761	Coprococcus[99%]
Isolate6	HV7 BEEF	766	Blautia[100%]
Isolate7	HV7 BEEF	778	Anaerostipes[96%]
Isolate8	HV7 BEEF	704	Dorea[100%]
Isolate9	HV7 BEEF	677	Coprococcus[100%]
Isolate10	HV7 BEEF	748	Ruminococcus2[100%]
Isolate11	HV7 BEEF	748	Ruminococcus2[100%]
Isolate12	HV7 BEEF	760	Coprococcus[99%]
Isolate13	HV7 BEEF	749	Anaerostipes[92%]
Isolate14	HV7 BEEF	778	Anaerostipes[92%]
Isolate15	HV7 BEEF	734	Coprococcus[100%]
Isolate16	HV7 BEEF	703	Blautia[100%]
Isolate17	HV7 BEEF	766	Coprococcus[100%]
Isolate18	HV7 BEEF	746	Coprococcus[100%]
Isolate19	HV7 BEEF	707	Lachnospiracea_incertae_sedis[77%]
Isolate20	HV7 BEEF	707	Coprococcus[100%]
Isolate21	HV7 BEEF	749	Coprococcus[65%]
Isolate22	HV7 BEEF	704	Blautia[100%]
Isolate23	HV7 BEEF	720	Blautia[100%]
Isolate24	IBS4 BHI+inu	711	Dorea[100%]
Isolate25	IBS4 BHI+inu	788	Blautia[100%]
Isolate26	IBS4 BHI+inu	787	Blautia[100%]
Isolate27	HV7 BHI+inu	736	Coprococcus[100%]
Isolate28	HV7 BHI+inu	745	Blautia[100%]
Isolate29	HV7 BHI+inu	763	Blautia[100%]
Isolate30	HV7 BHI+inu	747	Coprococcus[100%]
Isolate31	IBS4 BHI+inu	769	Blautia[100%]
Isolate32	HV7 BHI+inu	736	Coprococcus[100%]
Isolate33	HV7 BHI+inu	705	Coprococcus[100%]

Isolate34	HV7 BHI+inu	707	Roseburia[80%]
Isolate35	IBS4 BEEF	746	Dorea[100%]
Isolate36	IBS4 BEEF	779	Anaerostipes[95%]
Isolate37	IBS4 BEEF	706	Ruminococcus2[100%]
Isolate38	IBS4 BEEF	705	Blautia[100%]
Isolate39	IBS4 BEEF	732	Dorea[99%]
Isolate40	IBS4 BEEF	705	Dorea[100%]
Isolate41	IBS4 BHI+inu	747	Coprococcus[86%]
Isolate42	HV7 BEEF	778	Lachnospiracea_incertae_sedis[99%]
Isolate43	HV7 BEEF	744	Lachnospiracea_incertae_sedis[98%]
Isolate44	HV7 BEEF	707	Lachnospiracea_incertae_sedis[100%]
Isolate45	HV7 BHI+inu	745	Blautia[100%]
Isolate46	HV7 BHI+inu	704	Blautia[100%]
Isolate47	HV7 BEEF	708	Anaerostipes[94%]
Isolate48	HV7 BEEF	708	Anaerostipes[96%]
Isolate49	IBS4 BHI+inu	705	Ruminococcus2[100%]
Isolate50	IBS4 BHI+inu	774	Lachnospiracea_incertae_sedis[78%]
Isolate51	IBS4 BEEF	746	Lachnospiracea_incertae_sedis[67%]
Isolate52	IBS4 BEEF	744	Ruminococcus2[100%]
Isolate53	HV7 BHI+inu	745	Blautia[100%]
Isolate54	HV7 BHI+inu	734	Coprococcus[100%]
Isolate55	HV7 BHI+inu	748	Ruminococcus2[100%]
Isolate56	HV7 BHI+inu	746	Dorea[100%]
Isolate57	HV7 BHI+inu	752	Clostridium XlVb[100%]
Isolate58	HV7 BHI+inu	746	Dorea[100%]
Isolate59	HV7 BHI+inu	734	Coprococcus[100%]
Isolate60	HV7 BHI+inu	746	Coprococcus[99%]
Isolate61	HV7 BHI+inu	776	Anaerostipes[92%]
Isolate62	HV7 BHI+inu	707	Ruminococcus2[98%]
Isolate63	IBS4 BEEF	771	Dorea[100%]
Isolate64	IBS4 BEEF	710	Dorea[100%]
Isolate65	HV7 BHI+inu	785	Lachnospiracea_incertae_sedis[82%]
Isolate66	HV7 BEEF	746	Blautia[100%]
Isolate67	HV7 BEEF	768	Blautia[100%]
Isolate68	IBS4 BHI+inu	766	Blautia[100%]
Isolate69	IBS4 BEEF	702	Lachnospiracea_incertae_sedis[100%]
Isolate70	IBS4 BEEF	712	Clostridium XlVa[74%]
Isolate71	HV7 BEEF	763	Blautia[100%]
Isolate72	IBS4 BHI+inu	800	Clostridium XlVa[86%]
Isolate73	IBS4 BHI+inu	766	Blautia[100%]

Isolate74	IBS4 BHI+inu	748	Ruminococcus2[100%]
Isolate75	IBS4 BHI+inu	739	Clostridium XlVa[94%]
Isolate76	IBS4 BHI+inu	749	Anaerostipes[100%]
Isolate77	HV7 BHI+inu	734	Lachnospiracea_incertae_sedis[73%]
Isolate78	HV7 BHI+inu	744	Lachnospiracea_incertae_sedis[100%]
Isolate79	HV7 BHI+inu	734	Coprococcus[100%]

Table S3. Most Wanted matches of Lachnospiraceae isolates from HV7 and IBS4 samples

Isolate	Donor & Media	MostWanted OTU ID	MostWanted Priority Group	MostWanted RDP Summary	Alignment Length	% Pairwise Identity	E-Value	Consensus Sequence
Isolate1	IBS4 BEEF	otu_321_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Dorea(100)	450	100	0	GCAAGTCGAGCGAAGCACTTAAGTTTGATTCTTCGGATGAAGACTTTTGTGAC TGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGAT AACAGTTAGAAATGACTGCTAATACCGCATAAGACCACGGTACCGCATGGTAC AGTGGTAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGGTAGTTG GTGGGGTAACGGCCTACCAAGCCGACGATCAGTAGCCGACCTGAGAGGGTGA CCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGT GGGGAATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAAGGA TGAAGTATTTCGGTATGTAAACTTCTATCAGCAGGGAAGAAAATGACGGTACCT GACTAAGAAGCCCCGGCTAACTACGTG
Isolate3	HV7 BEEF	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	439	100	0	TGCAAGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGAC GGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAAA CGACTGGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAAC ACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAAC GGCCCACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACAT TGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATT GCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTATTT CGGTATGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAA GCACCGGCTAAATACGTG
Isolate4	HV7 BEEF	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	434	100	0	GTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGACGGGTG AGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAAACGACT GGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAACACTCC GGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAACGGCCC ACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACATTGGGA CTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA ATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTATTTCGGTA TGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAAGCACC GGCTAAATACGTG
Isolate7	HV7 BEEF	otu_308_V3V5	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Anaerostipes(98)	395	100	0	CGGTATGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAA GCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTT ATCCGGAATTACTGGGTGTAAAGGGTGCGTAGGTGGTATGGCAAGTCAGAAGT GAAAACCCAGGGCTTAACTCTGGGACTGCTTTTGAAACTGTCAGACTGGAGTG CAGGAGAGGTAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGG AGGAACATCAGTGGCGAAGGCGGCTTACTGGACTGAAACTGACACTGAGGCA CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA CGATGAATACTAGGTGTCGGGGCCG
Isolate9	HV7 BEEF	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	435	100	0	AGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGACGGGT GAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAAACGAC TGGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAACACTCC GGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAACGGCCC ACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACATTGGGA CTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACA ATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTATTTCGGTA TGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAAGCACC GGCTAAATACGTG
Isolate13	HV7 BEEF	otu_308_V3V5	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Anaerostipes(98)	367	100	0	CGGTATGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAA GCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTT ATCCGGAATTACTGGGTGTAAAGGGTGCGTAGGTGGTATGGCAAGTCAGAAGT GAAAACCCAGGGCTTAACTCTGGGACTGCTTTTGAAACTGTCAGACTGGAGTG CAGGAGAGGTAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGG AGGAACATCAGTGGCGAAGGCGGCTTACTGGACTGAAACTGACACTGAGGCA CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
Isolate14	HV7 BEEF	otu_308_V3V5	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Anaerostipes(98)	396	100	0	CGGTATGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAA GCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTT ATCCGGAATTACTGGGTGTAAAGGGTGCGTAGGTGGTATGGCAAGTCAGAAGT GAAAACCCAGGGCTTAACTCTGGGACTGCTTTTGAAACTGTCAGACTGGAGTG CAGGAGAGGTAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGG AGGAACATCAGTGGCGAAGGCGGCTTACTGGACTGAAACTGACACTGAGGCA CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA CGATGAATACTAGGTGTCGGGGCCGT
Isolate15	HV7 BEEF	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	438	100	0	GCAAGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGACG GGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAAAC GACTGGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAACA CTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAACG GCCCACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACATT GGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTG CACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTATTTC GGTATGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAAG CACCGGCTAAATACGTG

Isolate	Donor & Media	MostWanted OTU ID	MostWanted Priority Group	MostWanted RDP Summary	Alignment Length	% Pairwise Identity	E-Value	Consensus Sequence
Isolate16	HV7 BEEF	otu_556_V3V5	LOW	Clostridia(100);Clostridiales(100); Incertae Sedis XIV(100);Blautia(100)	369	100	0	GGAAACCCTGATGCAGCGACGCCGCGTGAAGGAAGAAGTATCTCGGTATGTAA ACTTCTATCAGCAGGGAAGATAGTGACGGTACCTGACTAAGAAGCCCCGGCTA ACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTT ACTGGGTGTAAAGGGAGCGTAGACGGACTGGCAAGTCTGATGTGAAAGGCGG GGGCTCAACCCCTGGACTGCATTGGAAACTGTTAGTCTTGAGTGCCGGAGAGG TAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACC AGTGGCGAAGGCGGCTTACTGGACGGTAACTGACGTTGAGGCTCGAAAGCGT
Isolate21	HV7 BEEF	otu_903_V3V5	MEDIUM	Clostridia(100);Clostridiales(100); Lachnospiraceae(97);Coprococcus(89)	430	100	0	GGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAAGG AAGAAGTATTTCGGTATGTAAACTTCTATCAGCAGGGAAGAAAATGACGGTAC CTGACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGG GGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGAGCGTAGGCGGTCTGAC AAGTCAGAAGTGAAAGCCCGGGGCTCAACTCCGGGACTGCTTTTGAAACTGC CGGACTAGATTGCAGGAGAGGTAAGTGGAATTCCTAGTGTAGCGGTGAAATGC GTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAAATG ACGCTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGT CCACGCCGT
Isolate22	HV7 BEEF	otu_556_V3V5	LOW	Clostridia(100);Clostridiales(100); Incertae Sedis XIV(100);Blautia(100)	369	100	0	GGAAACCCTGATGCAGCGACGCCGCGTGAAGGAAGAAGTATCTCGGTATGTAA ACTTCTATCAGCAGGGAAGATAGTGACGGTACCTGACTAAGAAGCCCCGGCTA ACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTT ACTGGGTGTAAAGGGAGCGTAGACGGACTGGCAAGTCTGATGTGAAAGGCGG GGGCTCAACCCCTGGACTGCATTGGAAACTGTTAGTCTTGAGTGCCGGAGAGG TAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACC AGTGGCGAAGGCGGCTTACTGGACGGTAACTGACGTTGAGGCTCGAAAGCGT
Isolate27	HV7 BHI+inu	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	440	100	0	ATGCAAGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGA CGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAA ACGACTGGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAA CACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAAC GGCCCACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACAT TGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATT GCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTATTT CGGTATGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAA GCACCGGCTAAATACGTG
Isolate28	HV7 BHI+inu	otu_556_V3V5	LOW	Clostridia(100);Clostridiales(100); Incertae Sedis XIV(100);Blautia(100)	410	100	0	GGAAACCCTGATGCAGCGACGCCGCGTGAAGGAAGAAGTATCTCGGTATGTAA ACTTCTATCAGCAGGGAAGATAGTGACGGTACCTGACTAAGAAGCCCCGGCTA ACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTT ACTGGGTGTAAAGGGAGCGTAGACGGACTGGCAAGTCTGATGTGAAAGGCGG GGGCTCAACCCCTGGACTGCATTGGAAACTGTTAGTCTTGAGTGCCGGAGAGG TAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACC AGTGGCGAAGGCGGCTTACTGGACGGTAACTGACGTTGAGGCTCGAAAGCGT GGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
Isolate30	HV7 BHI+inu	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	441	100	0	CATGCAAGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGG ACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGA AACGACTGGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAA ACACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTA ACGGCCCACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCAC ATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATA TTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTAT TTCGGTATGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGA AGCACCGGCTAAATACGTG
Isolate32	HV7 BHI+inu	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	440	100	0	ATGCAAGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGA CGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAA ACGACTGGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAA CACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAAC GGCCCACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACAT TGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATT GCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTATTT CGGTATGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAA GCACCGGCTAAATACGTG
Isolate33	HV7 BHI+inu	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	438	100	0	GCAAGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGACG GGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAAAC GACTGGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAACA CTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAACG GCCCACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACATT GGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTG CACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTATTTC GGTATGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAAG CACCGGCTAAATACGTG
Isolate35	IBS4 BEEF	otu_321_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Dorea(100)	450	100	0	GCAAGTCGAGCGAAGCACTTAAGTTTGATTCTTCGGATGAAGACTTTTGTGAC TGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGAT AACAGTTAGAAATGACTGCTAATACCGCATAAGACCACGGTACCGCATGGTAC AGTGGTAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGGTAGTTG GTGGGGTAACGGCCTACCAAGCCGACGATCAGTAGCCGACCTGAGAGGGTGA CCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGT GGGGAATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAAGGA TGAAGTATTTCGGTATGTAAACTTCTATCAGCAGGGAAGAAAATGACGGTACCT GACTAAGAAGCCCCGGCTAACTACGTG

Isolate	Donor & Media	MostWanted OTU_ID	MostWanted Priority Group	MostWanted RDP Summary	Alignment Length	% Pairwise Identity	E-Value	Consensus Sequence
								CAAGTCGAACGGAAGCACTTTGCTTAGATTCTTCGGATGAAGAGGATTGTGACT
								GAGCGGCGGACGGGTGAGTAATACGCGTGGGTAACCTGCCTCATACAGGGGGAT
Isolate37 IBS4 BEEF	otu_381_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(97);Coprococcus(47)	449	100	0		AACAGTTAGAAATGACTGCTAATACCGCATAAGACCACAGCATCGCATGGTGC
								GTAAGGTAACGGCTTACCAAGGCGACGATCAGTAGCCGACCTGAGAGGGTGA
								GGGGCATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGCG
								CCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGT
								AAGAATATTCCGATATGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTAC
								CTGACTAAGAAGCACCGGCTAACTACGTG
								AGCAGCTGGAAACGGCTGGTAATACCGCATAAGCGCACAATGTTGTCGATGACAT
								GCAAGTCGAGCGAAGCACTTAAGTTTGATTCTTCGGATGAAGACTTTTGTGAC
								TGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGAT
								AACAGTTAGAAATGACTGCTAATACCGCATAAGACCACGGTACCGCATGGTAC
								AGTGGTTAAAAACTCCGGTGTATGAGATGGACCCGCGTCTGATTAGGTTAGTTG
Isolate40 IBS4 BEEF	otu_321_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Dorea(100)	450	100	0		GTGGGGTAACGGCCTACCAAGCCGACGATCAGTAGCCGACCTGAGAGGGTGA
								CCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGT
								GGGGAATATTGCACAATGGGAGGAAACCTGATGCAGCGACGCCGCGTGAAGGA
								TGAAGTATTCCGTATGTAAACTTCTATCAGCAGGGAAGAAAATGACGGTACCT
								GACTAAGAAGCCCCGGCTAACTACGTG
								AGCAGCTGGAAACGGCTGGTAATACCGCATAAGCGCACAATGTTGTCGATGACAT
								GGTGTGAAAAACTCCGGTGGTATAAGATGGACCCGCGTCTGATTAGCTAGTTG
Isolate42 HV7 BEEF	otu_237_V1V3	MEDIUM	Clostridia(100);Clostridiales(100); Lachnospiraceae(9);Coprococcus(26)	344	100	0		GTGAGATAACAGCCCACCACAGGGCGAATCAGTAGCCGACCTGAGAGGGTGA
								CCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGT
								GGGGAATATTGCACAATGGGAGAAACTCTGATGCAGCGACGCCGCGTGAGTGA
								AGAAGTATTCGTATTGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCT
								GACTAAGAAGCCCCGGCTAACTACGTG
								TGGCGCAGCGGTGAGATACGCGTGGGTAACCTGCCCTGTACAGGGGATAAC
								AGTCGAAATGACTGCTAATACCGCATAAGACCACAGCACCGCATGGTGCAGG
								GGTAAAACTCCGGTGGTACAGGATGGACCCGCGTCTGATTAGCTGGTTGGTG
Isolate47 HV7 BEEF	otu_451_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(72);Coprococcus(4)	393	100	0		AGGTAACGGCTCACCAAGGCGACGATCAGTAGCCGGCTTGAGAGGAGTGAACG
								GCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGG
								GAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAG
								AAGTATCTCGGTATGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCTG
								ACTAAGAAGCCCCGGCTAACTACGTG
								AATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGGCGAAGA
								AGTATTTCGGTATGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCTGAC
								TAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCGA
								AGCGTTATCCGGATTTACTGGGTGTAAAGGGAGCGTAGACGGCAATTCAAGTCT
Isolate51 IBS4 BEEF	otu_561_V3V5	MEDIUM	Clostridia(100);Clostridiales(100); Lachnospiraceae(62);Coprococcus(29)	426	100	0		TGATGTGAAAGGCCCGGGCTCAACCCCGGGACTGCATTGGAAACTGTCTCAATCT
								AGAGTACCGGAGGGGTAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAT
								ATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACGGTAACTGACGTT
								GAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACG
								CCGT
								GGAAACCCTGATGCAGCGACGCCGTGAAGGAAGAAGTATCTCGGTATGTAA
								ACTTCTATCAGCAGGGAAGAATGTGACGGTACCTGACTAAGAAGCCCCGGCTA
								ACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTT
Isolate53 HV7 BHI+inu	otu_556_V3V5	LOW	Clostridia(100);Clostridiales(100); Incertae Sedis XIV(100);Blautia(100)	410	100	0		ACTGGGTGTAAAGGGAGCGTAGACGGCACTTGGCAAGTCTGATGTGAAAGGCGG
								TAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACC
								AGTGGCGAAGGCGGCTTACTGGACGGTAACTGACGTTGAGGCTCGAAAGCGT
								GGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
								GCAAGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGACG
								GGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAAAC
								GACTGGTAATACCGCATAAGCGCACAGTGTGCGCATGCACACAGTGTGAAAACA
								CTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAACG
Isolate54 HV7 BHI+inu	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	438	100	0		GCCCACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACATT
								GGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTG
								CACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGGTTTC
								GGTATTGTAAAGCTCTATCAGCAGGGAAGAATAATGACGGTATCCTGACTAAGAAG
								CACCGGCTAACTACGTG
								GCAAGTCGAACGAAGCACTTAAGTTTGATTCTTCGGATGAAGACTTTTGTGAC
								TGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGAT
								AACAGTTAGAAATGACTGCTAATACCGCATAAGACCACGGTACCGCATGGTAC
								AGTTGGTTAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGGTTAGTTG
Isolate56 HV7 BHI+inu	otu_321_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Dorea(100)	450	100	0		GTGGGGTAACGGCCTACCAAGCCGACGATCAGTAGCCGACCTGAGAGGGTGA
								CCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGT
								GGGGAATATTGCACAATGGGAGGAAACCTGATGCAGCGACGCCGCGTGAGGGA
								TGAAGTATTCCGTATGTAAACTTCTATCAGCAGGGAAGAAAATGACGGTACCT
								GACTAAGAAGCCCCGGCTAACTACGTG
								TGATGCAGCAACGCCGCGTGAAGGAAGAAGGTTTTCGGATCGTAAACTTCTAT
								CAACAGGGACGAAGAAAGTGACGGTACCTGAATAAGAAGCCCCGGCTAACTA
Isolate57 HV7 BHI+inu	otu_514_V3V5	MEDIUM	Clostridia(96);Clostridiales(96); Lachnospiraceae(69);Acetitomaculum(38)	406	100	0		CGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGAATTACTG
								GGTGTAAAGGGAGCGTAGGCGGCACGCCAAGTCAGATGTGAAAGCCCGGGCTCAACTA
								TTAACCTCGCGGATTGCATTTGGAACTGGCGGAGCTAGAGTACGGGAGAGGAAA
								GCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAAGAACACCAGT
								GGCGAAGGCGGCTTTCTGGACTGAACTGACGCTGAGGCTCGAAAGCGTGGG
								GAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT

Isolate	Donor & Media	MostWanted OTU ID	MostWanted Priority Group	MostWanted RDP Summary	Alignment Length	% Pairwise Identity	E-Value	Consensus Sequence
Isolate58	HV7 BHI+inu	otu_321_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Dorea(100)	450	100	0	GCAAGTCGAGCGAAGCACTTAAGTTTGATTCTTCGGATGAAGACTTTTGTGAC TGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGAT AACAGTTAGAAATGACTGCTAATACCGCATAAGACCACGGTACCGCATGGTAC AGTGGTAAAAACTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGGTAGTTG GTGGGGTAACGGCCTACCAAGCCGACGATCAGTAGCCGACCTGAGAGGGTGA CCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGT GGGGAATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAAGGA TGAAGTATTTCGGTATGTAAACTTCTATCAGCAGGGAAGAAAATGACGGTACCT GACTAAGAAGCCCCGGCTAACTACGTG
Isolate59	HV7 BHI+inu	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	438	100	0	GCAAGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGACG GGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAAAC GACTGGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAACA CTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAACG GCCCACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACATT GGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTG CACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTATTTC GGTATGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAAG CACCGGCTAAATACGTG
Isolate61	HV7 BHI+inu	otu_451_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(72);Coprococcus(4)	393	100	0	TGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCCTGTACAGGGGGATAAC AGTCAGAAATGACTGCTAATACCGCATAAGACCACAGCACCGCATGGTGCAGG GGTAAAAACTCCGGTGGTACAGGATGGACCCGCGTCTGATTAGCTGGTTGGTG AGGTAACGGCTCACCAAGGCGACGATCAGTAGCCGGCTTGAGAGAGTGAACG GCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGG GAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAG AAGTATCTCGGTATGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCTG ACTAAGAAGCCCCGGCTAACTACGTG
Isolate64	IBS4 BEEF	otu_395_V3V5	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Dorea(100)	363	100	0	TGCAGCGACGCCGCGTGAAGGATGAAGTATTTCGGTATGTAAACTTCTATCAGC AGGGAAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGCCA GCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAA AGGGAGCGTAGACGGCTGTGCAAGTCTGAAGTGAAAGGCATGGGCTCAACCT GTGGACTGCTTTGGAAACTGTGCAGCTAGAGTGTCGGAGAGGTAAGTGGAATT CCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGG CGGCTTACTGGACGATGACTGACGTTGAGGCTCGAAAGCGTGGGGA
Isolate66	HV7 BEEF	otu_556_V3V5	LOW	Clostridia(100);Clostridiales(100); Incertae Sedis XIV(100);Blautia(100)	410	100	0	GGAAACCCTGATGCAGCGACGCCGCGTGAAGGAAGAAGTATCTCGGTATGTAA ACTTCTATCAGCAGGGAAGATAGTGACGGTACCTGACTAAGAAGCCCCGGCTA ACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTT ACTGGGTGTAAAGGGAGCGTAGACGGACTGGCAAGTCTGATGTGAAAGGCGG GGGCTCAACCCCTGGACTGCATTGGAAACTGTTAGTCTTGAGTGCCGGAGAGG TAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACC AGTGGCGAAGGCGGCTTACTGGACGGTAACTGACGTTGAGGCTCGAAAGCGT GGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
Isolate67	HV7 BEEF	otu_556_V3V5	LOW	Clostridia(100);Clostridiales(100); Incertae Sedis XIV(100);Blautia(100)	433	100	0	GGAAACCCTGATGCAGCGACGCCGCGTGAAGGAAGAAGTATCTCGGTATGTAA ACTTCTATCAGCAGGGAAGATAGTGACGGTACCTGACTAAGAAGCCCCGGCTA ACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTT ACTGGGTGTAAAGGGAGCGTAGACGGACTGGCAAGTCTGATGTGAAAGGCGG GGGCTCAACCCCTGGACTGCATTGGAAACTGTTAGTCTTGAGTGCCGGAGAGG TAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACC AGTGGCGAAGGCGGCTTACTGGACGGTAACTGACGTTGAGGCTCGAAAGCGT GGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATA CTAGGTGTTGG
Isolate78	HV7 BHI+inu	otu_398_V3V5	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(98);Coprococcus(53)	354	100	0	TCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAAGCTCCGGCTAAAT ACGTGCCAGCAGCCGCGGTAATACGTATGGAGCAAGCGTTATCCGGATTTACTG GGTGTAAAGGGTGCGTAGGTGGCAGTGCAAGTCAGATGTGAAAGGCCGGGGC TCAACCCCGGAGCTGCATTTGAAACTGCTCGGCTAGAGTACAGGAGAGGCAG GCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGT GGCGAAGGCGGCCTGCTGGACTGTTACTGACACTGAGGCACGAAAGCGTGGG GAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
Isolate79	HV7 BHI+inu	otu_437_V1V3	LOW	Clostridia(100);Clostridiales(100); Lachnospiraceae(100);Coprococcus(100)	438	100	0	GCAAGTCGAACGATTGAAGCTGGAGCTTGCTCCGGCCGATTTAGTGGCGGACG GGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAGCAGTTGGAAAC GACTGGTAAAACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAACA CTCCGGTGGTATGAGATGGACCCGCGTCTGATTAGCTAGTTGGTGAGGTAACG GCCCACCAAGGCGACGATCAGTAGCCGGCCTGAGAGGGTGAACGGCCACATT GGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTG CACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGATGAAGTATTTC GGTATGTAAAGCTCTATCAGCAGGGAAGATAATGACGGTACCTGACTAAGAAG CACCGGCTAAATACGTG

Chapter 3

Characterization of Lachnospiraceae In The Human Gut Microbiota

Preface

Research presented as part of this chapter has been prepared for publication.

Lau JT and Surette MG. Characterization of Lachnospiraceae diversity in the human gut microbiota. 2017.

Contributions:

JTL performed the experiments. JTL and MGS designed the experiments, analyzed the data, and wrote the manuscript.

Abstract

The Lachnospiraceae family is one of the most abundant members of the human gut microbiota, and one of the most phylogenetically diverse bacterial groups. The abundance of Lachnospiraceae is reported to be altered in several disease states, and this family may be important in the microbiota's role for maintaining health. However, few reference isolates and genomes are publically available for this family, and the extent of diversity is potentially underrepresented by the small number of strains. In this study, we characterized the phenotypic diversity of 42 Lachnospiraceae isolates from the human gut microbiota for *in vitro* phenotypes that may be important for survival and interactions in the gastrointestinal tract. The genome sequences of 14 isolates were compared to reference genomes to determine the potential heterogeneity within this family. The Lachnospiraceae isolates exhibited diversity in the majority of phenotypes assessed, although the prevalence of resistance to several antibiotics was low. Intra-species variability was observed between isolates collected from the same donor. Genomic comparisons reveal differences between strains and their closest RefSeq genomes, including differences in prophage insertions, predicted secondary metabolites and presence of antibiotic resistance genes. This study revealed insight into the significant functional and genetic heterogeneity within this family, including intra-species diversity, which will be important to consider in understanding the roles of this family in the human gut microbiome.

Introduction

Lachnospiraceae, also known as Clostridium cluster XIVa or *Clostridium coccoides* group, is the most abundant and diverse bacterial family in the human gastrointestinal (GI) tract, making up to 50% of the total bacterial community[124]. It is part of the Firmicutes phyla and Clostridia class, and contains 24 genera, making it one of the most phylogenetically diverse bacterial groups of the human microbiome[125]. Lachnospiraceae are mostly found to inhabit the mammalian GI tract[177] but are also common members of the oral cavity although at low abundance[178,179]. All members are obligate anaerobes and have been shown to be extremely sensitive to oxygen, with some strains unable to grow after two minutes of air exposure[180]. Although originally described as non-spore forming bacteria, recent studies have shown experimentally[132] and genetically[179] that most Lachnospiraceae species can form endospores, which is likely important for host-to-host transmission of these obligate anaerobes. This family is important for the breakdown of complex polysaccharides from the diet, and is a major producer of the short chain fatty acid butyrate, although not all genera, such as *Dorea,* are capable of making butyrate[177].

One mechanism by which Lachnospiraceae contribute to the gut microbiota's role in maintaining intestinal health is through the production of butyrate, which is important for regulation of gut motility and the immune system[181]. Members of this family have also been shown to provide colonization resistance against *Clostridium difficile* infection[36]. Additional evidence implicating Lachnospiraceae to health include studies that have reported decreased Lachnospiraceae in patients with colorectal cancer[78],

diabetes[89,94], Inflammatory Bower Disease (IBD)[21,22] and Irritable Bowel Syndrome (IBS)[68,74]. However, other studies, have found that *Dorea*, a major gas producer, is increased in IBS patients and may play a role in symptoms of bloating[72,182], and some Lachnospiraceae may be involved in the initiation of intestinal inflammation in DSS colitis[183]. Despite this potential impact on host physiology, as well as the prevalence and abundance of Lachnospiraceae in the gut microbiota, many genera are not well-characterized and there are relatively few cultured representatives[123], likely due to difficulties in working with strains which are highly sensitive to oxygen exposure.

Although Lachnospiraceae OTUs from different species are often grouped together due to lack of resolution in 16S rRNA gene sequencing studies, we hypothesize that isolates from this family express diverse phenotypes, and genetic diversity is likely not fully represented by reference strains. Indeed, novel Lachnospiraceae species are continuously being isolated from human fecal samples[132,184] and a strain-level metagenomics study found that *Roseburia*, a species of the Lachnospiraceae family, had highly variable genes[115]. Using strains isolated from a previous study, we investigated the phenotypic variability in 42 isolates and compared the genome sequences of 14 strains to their closest reference genomes to determine the potential heterogeneity in this family. Lachnospiraceae phenotypes characterized include those that may be important for colonization of the GI tract, including production of exoenzymes, motility and antibiotic resistance. As a clear role for Lachnospiraceae in GI health and disease has not been definitively described, but microbiome studies frequently report changes in this family when comparing healthy to disease samples, it is important to isolate, characterize and

sequence genomes of Lachnospiraceae strains to further understand their ecological role in the human gut.

Materials and methods

Lachnospiraceae isolation and identification

Lachnospiraceae isolates were obtained from a healthy donor (Donor 1) and an IBS donor (Donor 2) as previously described[184]. In brief, freshly collected fecal samples were cultured anaerobically on BHI + 1g/L inulin and cooked meat agar. Strains were streaked to purify and 16S rRNA gene was amplified performed using 8f (5'-AGAGTTTGATCCTGGCTCAG-3') and 927r (5'-CCGTCAATTCCTTTRAGTTT-3') primers. PCR conditions consisted of an initial denaturation at 94^{o}C for 2 min, 32 cycles of 94^{o}C for 1 min, 56^{o}C for 1 min, 72^{o}C for 2 min, followed by a final elongation at 72^{o}C for 10 min. PCR products were Sanger sequenced by Beckman Coulter Genomics (Danvers, MA) using the 8f primer. Sequences were identified by BLAST of the 16S rRNA gene regions against NCBI's 16S ribosomal RNA sequences (Bacteria and Archaea) database.

Lachnospiraceae phenotyping

Frozen skim milk stocks were streaked onto cooked meat broth (Fluka) + 1.2% agar and grown for 48h, at 37^{o}C in an anaerobic chamber (5% CO_2, 5% H_2, 90% N_2, Shel Labs). All media used was pre-reduced overnight in the anaerobic chamber prior to use. Colonies were resuspended in cooked meat broth and isolates were phenotyped by pin-

replication onto cooked meat agar base plates unless otherwise specified. Resuspensions were replicated onto cooked meat agar as a positive control for growth. Hemolysis was determined on Columbia blood agar with 5% sheep's blood, beta-galactosidase activity was determined by spreading 50 mg/ml X-gal on cooked meat agar, swimming was detected on 0.25% (w/v) agar and swarming on 0.6% agar + 0.5% glucose plates. DNase, protease and H_2O_2 production was determined as previously described[185]. Beta-galactosidase and H_2O_2 production was assessed after removing the plates from the anaerobe chamber for 1h.

Antibiotic resistance was measured on cooked meat agar plates containing ampicillin (2 and 8 μg/ml), vancomycin (2 and 8 μg/ml), clindamycin (1 and 4 μg/ml), metronidazole (1 and 4 μg/ml), chloramphenicol (8 and 32 μg/ml) or polymyxin B (16, 64 and 256 μg/ml). All plates were assessed after 48h of anaerobic growth at 37^{o}C.

Production of AI-2 was measured from supernatants of Lachnospiraceae cultures grown anaerobically in cooked meat broth for 24h, at 37^{o}C. The reporter strain *Vibrio harveyi* MM32 was grow in Autoinducer Bioassay (AB) medium overnight at 30^{o}C, and then diluted 1/5000 in AB medium, as previously described[186]. Luminescence and OD_{600nm} was measured in 96-well black plates with clear bottoms using a BioTek Synergy plate reader at 30^{o}C for 20h, with reads every 30min. Each well contained 10% supernatant, and was overlayed with mineral oil to prevent volume loss. Positive controls were supernatants from 2 commensal *Escherichia coli* strains subcultured for 4h in LB+0.5% glucose as previously described[186].

Genome assembly and comparative genomics

DNA was extracted from 14 Lachnospiraceae strains from colonies on cooked meat agar after 48h growth, using the Promega Wizard Genomic DNA Purification kit. Strains were sequenced on the Illumina HiSeq by the McMaster Genomics Facility (Hamilton, ON) and assembled with A5-miseq[187] (Supplementary Table 1). The closest reference genome for each isolate was identified by BLAST of the 16S rRNA regions against NCBI's RefSeq Representative Genomes database. Genomes were compared using BLAST Ring Image Generator (BRIG), following the tutorials[188]. Assemblies were analysed with PHASTER for prophages (phaster.ca), specifying metagenomic contigs[189], PRISM3[190] for prediction of secondary metabolites (magarveylab.ca/prism/), and the Comprehensive Antibiotic Resistance Database (CARD) Resistance Gene Identifier (card.mcmaster.ca/analyze/rgi), with only perfect and strict hits reported[191].

Results

Phenotypic diversity of Lachnospiraceae strains demonstrates inter- and intra-species variability

In this study, 42 Lachnospiraceae strains isolated from two donors (Supplementary Figure 1) were phenotyped using agar plate-based assays, and genomes from 14 strains were sequenced and compared to publically available reference strains. The phenotypes assessed in this study included those that have a potential role in colonizing the GI tract, and interactions with host cells and other microbes. H_2O_2 production was observed to some extent in all *Coprococcus comes* strains from Donor 1,

but expression was most apparent in the *Anaerostipes hadrus* strains, and the *Clostridium lactifermentans* and *Eubacterium rectale* isolates. A majority of the Lachnospiraceae isolates also had beta-galactosidase and protease activity on X-gal and skim milk plates, respectively. Additionally, 25 strains could degrade extracellular DNA. No hemolytic activity was observed on sheep's or human blood plates (data not shown). Both swimming and swarming motility were characterized but only swimming was observed, and only in a few isolates. Both types of movement require flagella, but swarming is a multi-cellular phenomenon that was not observed under the conditions tested.

The Lachnospiraceae isolates were separated into four clusters based on similarities in phenotype profiles for 9 different assays (Figure 1). In general, isolates of the same species grouped together, although there were several discrepancies observed. For example, the six *Blautia wexlerae* strains were found across three phenotype clusters, and while four *B. wexlerae* strains from Donor 1 grouped together, a fifth strain that was also from the same donor, had a divergent phenotype profile. Also of note was that *A. hadrus* strains from both donors grouped together although the phenotype profiles were slightly variable between the strains, e.g. DNase production and swimming motility were different between the Donor 1 isolates. However, some species had strains that were more similar to other within-donor strains than to strains from a different donor, as is the case for Donor 1 *C. comes* isolates. Interestingly, with the exception of *Coprococcus eutactus* strains from Donor 1, which all had identical phenotypes, no other strains within a species had identical phenotype profiles. This demonstrates that intra-species phenotypic diversity is apparent even between strains derived from the same individual.

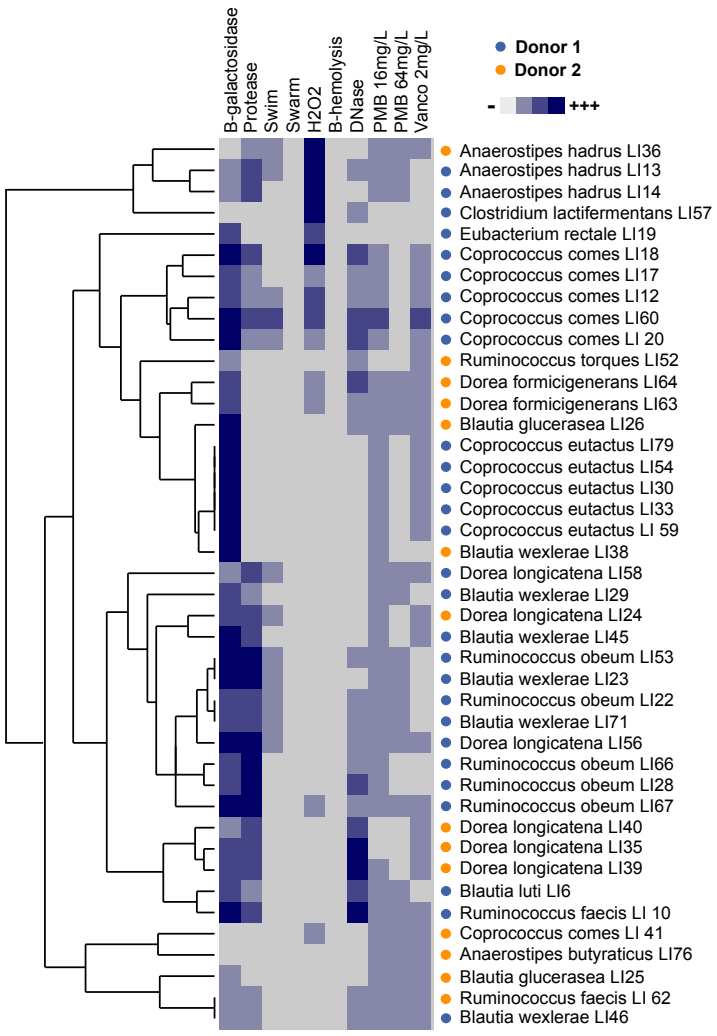


Figure 1. Extensive phenotypic diversity observed from Lachnospiraceae isolates of two donors. Hierarchical clustering of isolates based on production of β-galactosidase, protease activity, H_2O_2 production, DNase activity, β-hemolysis on sheep's blood, swimming/swarming motility, resistance to polymyxin B (16 & 64 μg/ml) and vancomycin (2 μg/ml). Darker blue indicates higher expression of phenotype, Donor indicated by legend. PMB – polymyxin B, vanco – vancomycin.

Resistance of Lachnospiraceae isolates to six antibiotics of different classes were characterized to further look at phenotypic variability within this family. There was no growth of any strains detected at the lowest concentrations of ampicillin, clindamycin,

metronidazole and chloramphenicol tested (Figure 2). Lachnospiraceae isolates were only resistant to PMB and vancomycin, although this was also variable between isolates. 16/42 isolates were resistant to 16 μg/ml PMB, 20/42 strains were resistant to 64 μg/ml PMB, and no isolates could grow on 256 μg/ml (data not shown). Six isolates could not grow at the lowest concentration of PMB used (16 μg/ml), even though PMB is not reported to be effective against Gram-positives. 13 isolates were sensitive to 2 μg/ml of vancomycin, and 29 isolates were resistant. No isolates could grow at 8 μg/ml vancomycin. Intra-species diversity was also observed for antibiotic resistance profiles. One *Ruminococcus obeum* strain from Donor 1 was resistant to 2 μg/ml vancomycin, while the other four *R. obeum* strains from Donor 1 were not. Additionally, three *R. obeum* strains were resistant to 64 μg/ml PMB, while the other two strains could only grow at 16 μg/ml PMB. Although all four *Dorea longicatena* strains were resistant to 2 μg/ml vancomycin, they differed in resistance to 64 μg/ml PMB. *B. wexlerae* strains from Donor 1 also demonstrated variability in resistance profiles, as similarly observed for the other phenotypes characterized (Figure 1).

The ability of Lachnospiraceae to produce the inter-species quorum sensing signal, AI-2, was measured using the luminescence reporter *Vibrio harveyi* MM32 (Figure 3). Unlike the previously characterized phenotypes, AI-2 production was fairly consistent between isolates, and very similar levels were measured within a species. The

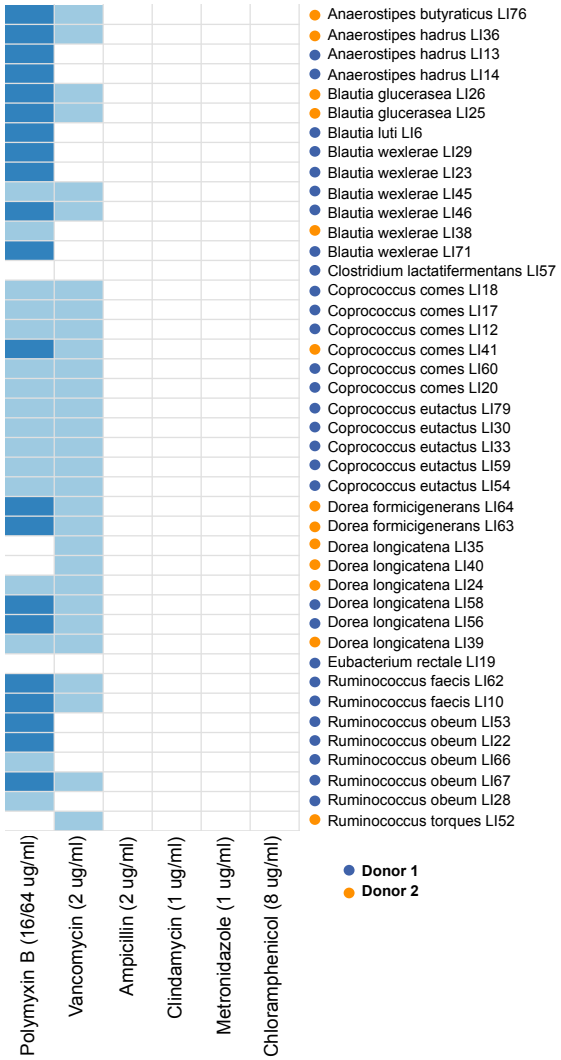


Figure 2. Antibiotic resistance of Lachnospiraceae isolates from two donors show low resistance profiles. Concentrations of antibiotics tested are at or below the EUCAST clinical breakpoint for Gram-positive anaerobes (Ver. 6, 2016), except for polymyxin B (PMB), which did not have a breakpoint concentration. Isolates resistant to 16 µg/ml PMB are light blue, isolates resistant to 64 µg/ml PMB are dark blue, sensitive isolates are white.

average production of luminescence by Lachnospiraeae isolates was 145-fold over

background, which was much higher than the positive *E. coli* control (average 50-fold

over background). *E. rectale* produced slightly lower levels of AI-2 (100-fold) and *R. torques* was the only isolate tested that did not appear to make AI-2.

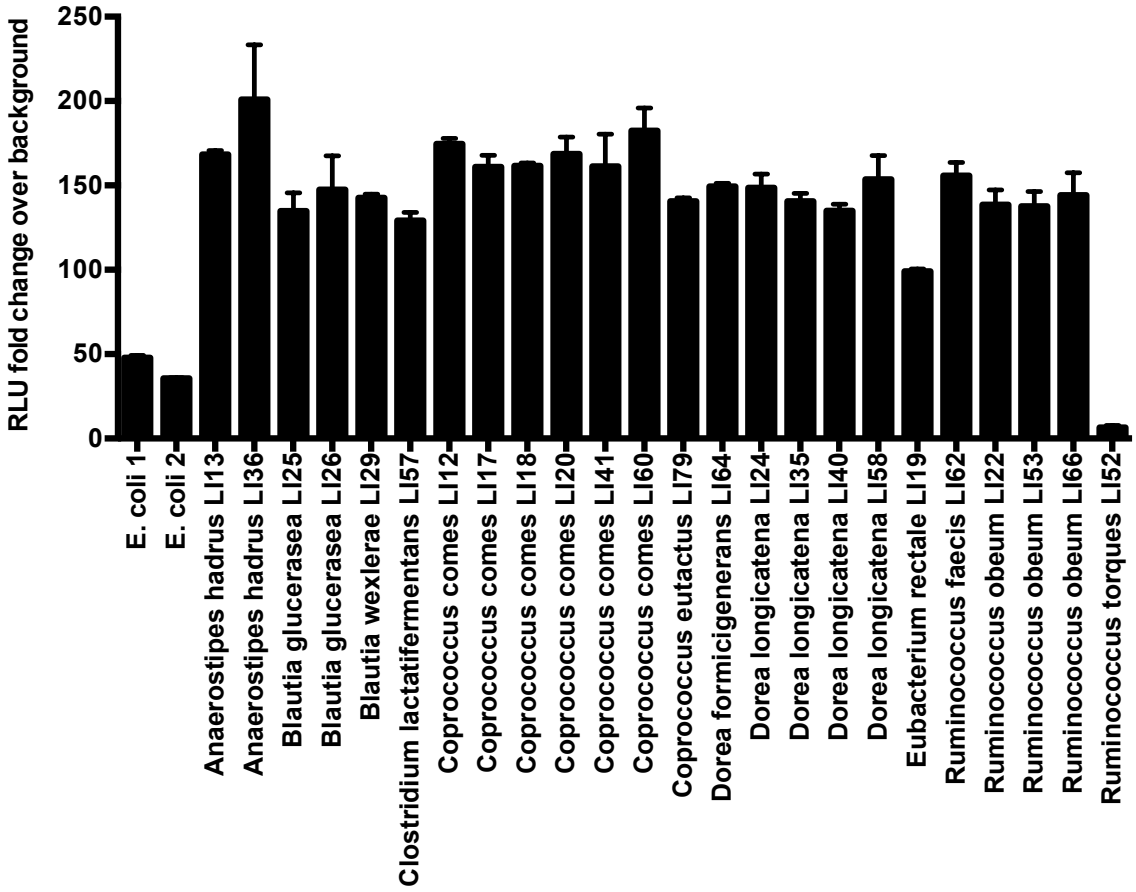
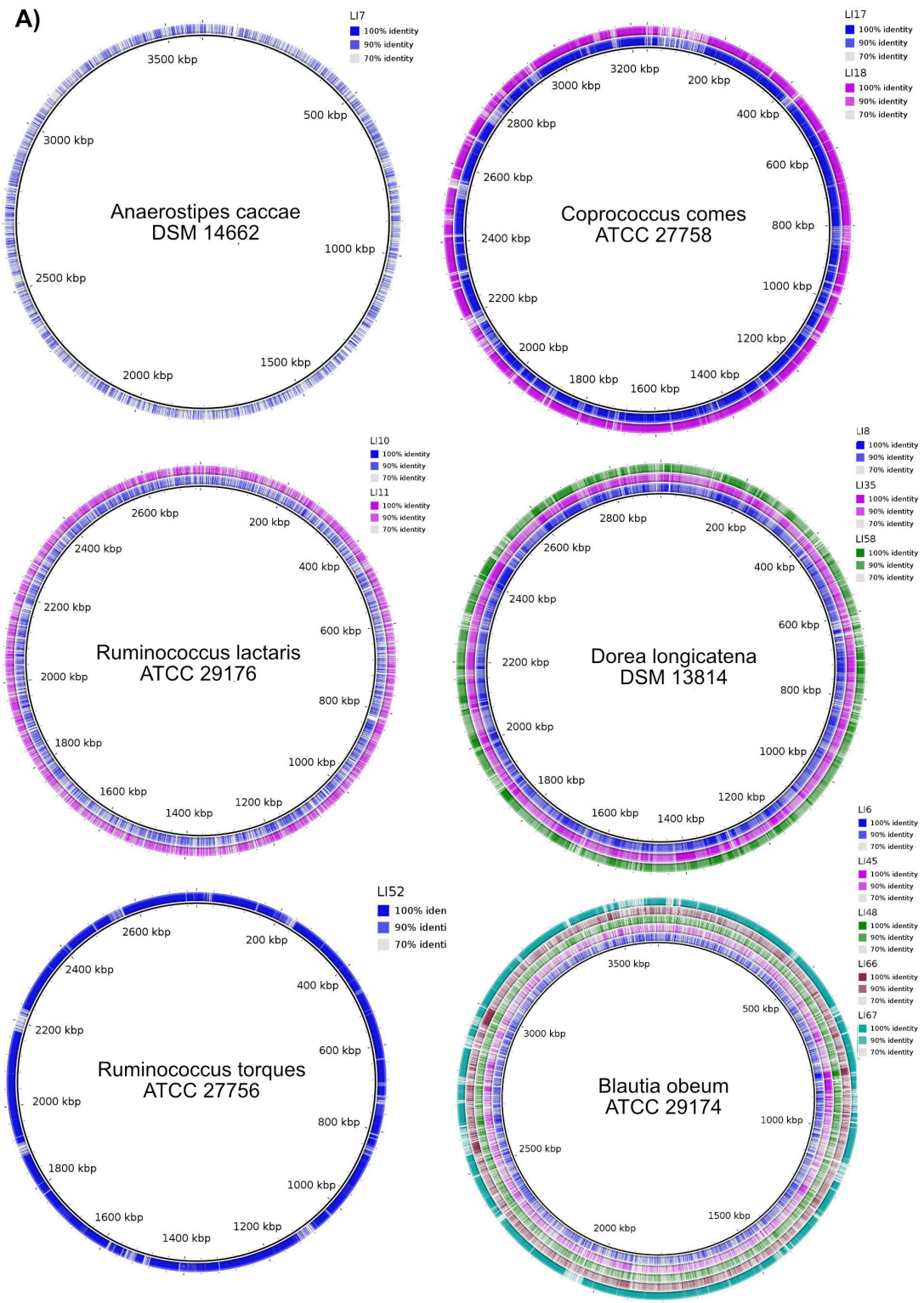


Figure 3. Lachnospiraceae produce high levels of the interspecies signalling molecule, autoinducer-2 (AI-2). Production of AI-2 measured using *Vibrio harveyi* MM32 as a reporter. Supernatant of 24h Lachnospiraceae cultures were incubated with *V. harveyi* MM32 for 20h, and luminescence measured every 30 min. Log-phase *E. coli* grown in LB+glucose was a positive control for AI-2 production. Bars represent max RLU over background luminescence (*V. harveyi* MM32 only), error bars represents standard deviation of triplicates.

Genome comparisons of Lachnospiraceae strains with reference sequences demonstrates genetic diversity

The genomes of 14 Lachnospiraceae strains were sequenced, contigs were assembled and compared with the closest reference genomes of each isolate to determine genetic diversity. Each Lachnospiraceae strain was compared to the closest reference genome and visualized by BRIG[188] (Figure 4A). Reference genomes ranged from 2.731 to 3.626 Mbp. No isolates had identical genomes to reference sequences, as indicated by low identity or uncovered regions (gaps) in the comparisons to references. For example, LI7 had low identity when compared to the *Anaerostipes caccae* DSM 14662 genome, as indicated by the light coloured ring. LI17 and LI18 strains appeared to be very similar to each other, as did LI10 & LI11. *C. comes* and *Ruminococcus torques* comparisons had the highest identity to the Lachnospiraceae strains for regions with coverage, as indicated by the darker coloured rings. However, there were still gaps/breaks in comparison to the references, which were areas of the references that were not found in the sequences of the Lachnospiraceae isolates. In the *B. obeum*-group, LI67 was most similar to the reference genome but compared to the other 4 Lachnospiraceae strains in this group, it contained unique areas of low identity in comparison to the reference. Indeed, all five isolate sequences for which *B. obeum* was the closest reference were different from each other, suggesting potential intra-species genetic diversity. Next, each Lachnospiraceae strain was used as the BRIG reference for the closest RefSeq genome (Figure 4B) to identify any regions of the isolate genomes not present in the references. The Lachnospiraceae


A)
LI7
100% identity
90% identity
70% identity
Anaerostipes caccae
DSM 14662
LI17
LI18
Coprococcus comes
ATCC 27758
LI10
LI11
Ruminococcus lactaris
ATCC 29176
LI8
LI35
LI58
Dorea longicatena
DSM 13814
LI52
Ruminococcus torques
ATCC 27756
LI6
LI45
LI48
LI66
LI67
Blautia obeum
ATCC 29174

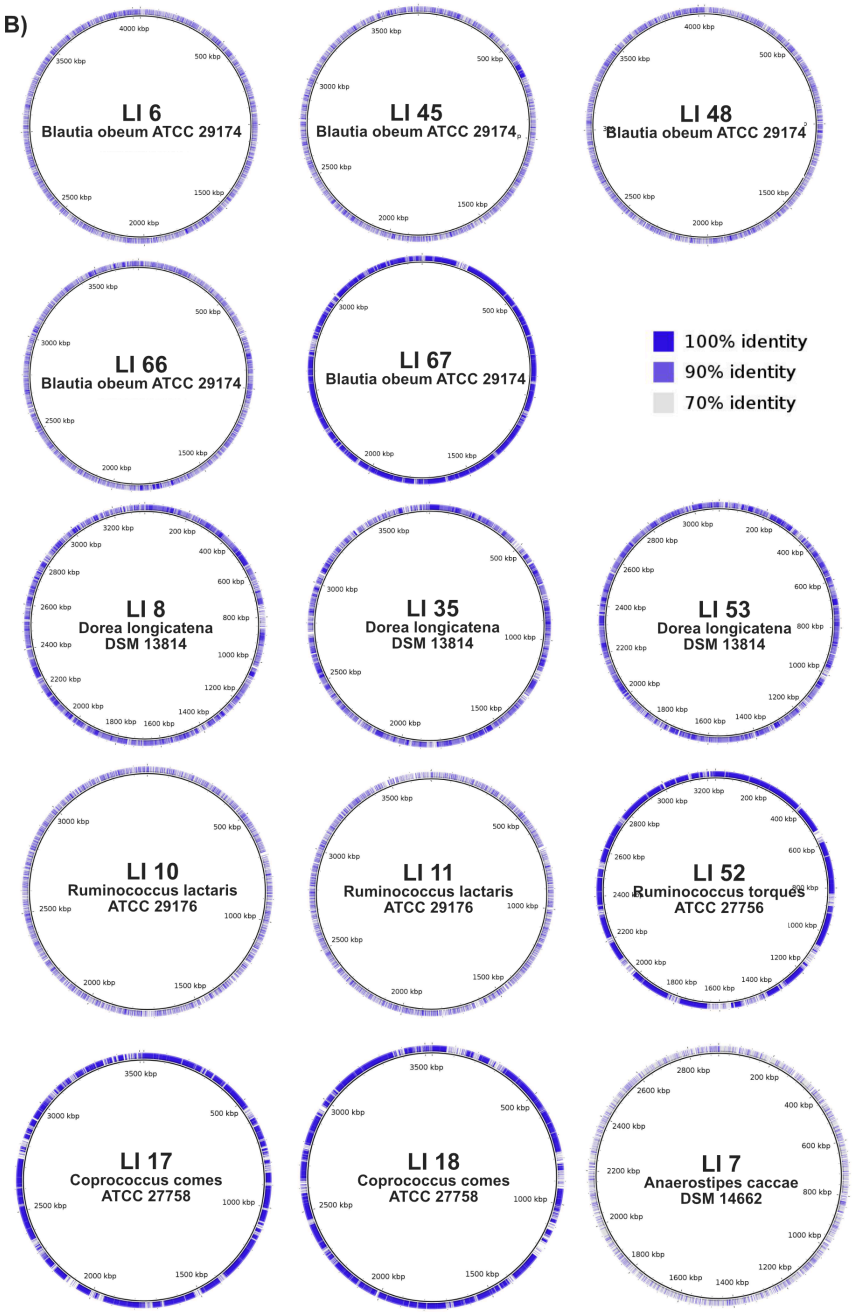


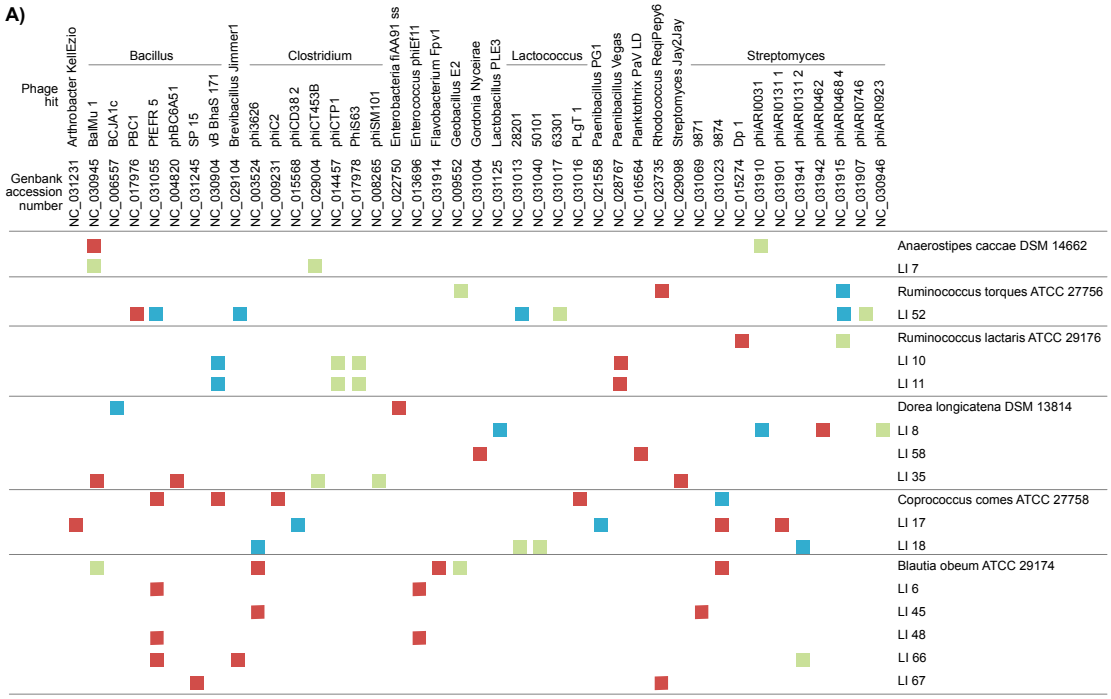
Figure 4. BRIG visualization of Lachnospiraceae genome comparisons demonstrate genetic diversity between isolates and reference sequences. A) Comparison of Lachnospiraceae genomes using RefSeq representative genomes as reference sequence. Innermost black ring represents the reference genomes, coloured rings represent the most closely related Lachnospiraceae strains, as indicated in the legends. B) Comparison of RefSeq representative genomes using Lachnospiraceae strains as reference sequence.

Name of RefSeq genomes indicated under the size of each assembly. BLAST identities are indicated as a sliding colour scale in legends.

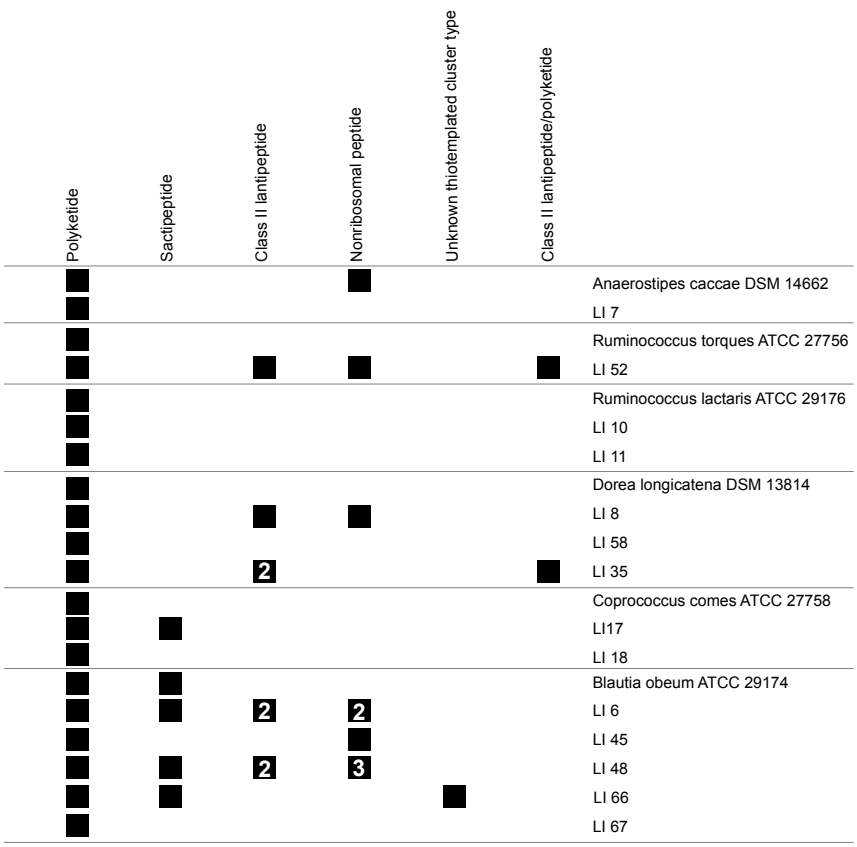
genome assemblies ranged from 2.909-4.042 Mbp. As previously observed, in the *B. obeum* group, LI67 had the highest identity with the *B. obeum* ATCC 29174 RefSeq genome (i.e. the blue ring is the darkest). LI52, LI17 and LI18 also had high identity with their respective reference genomes, whereas the other comparisons showed less similarity across the genomes (i.e. rings are lighter blue). It is interesting to note that even when Lachnospiraceae genomes were similar to the reference, there are still several regions without coverage, where segments of the isolate genome were not found in the reference sequences.

Genome sequences were also analyzed using freely available, web-based tools to assess genomic features. The insertion of prophages were detected by PHASTER[189] and 41 prophages were found across the 20 genomes (14 isolates and 6 references) (Figure 5A). 19 prophages were detected in the six reference genomes, but only four of these were also found in the genomes of the closest isolates. This indicates that there is little overlap in prophage sequences between the Lachnospiraceae strains and the reference strains. 25 prophages were identified in the Lachnospiraceae strains but not found in any of the six references. LI10 and LI11 had identical prophage profiles, as did LI6 and LI48, although the 16S rRNA gene sequence identified these strains as different species (LI6 classified as *Blautia luti* and LI48 classified as *A. hadrus*). LI10 and LI11 genomes also contained polyketide sequences as predicted by PRISM3[190], as did the *Ruminococcus lactaris* ATCC 29176 genome (Figure 5B). However, several isolates had more

secondary metabolite pathways encoded in the genome than the references, including the *B. obeum*-group, *D. longicatena*-group and the *R. torques*-group. For example, LI48 genome contained seven secondary metabolites pathways and LI6 had six, while their closest reference, *B. obeum,* was only predicted to produce two natural products (Supplementary Figure 2). Only two isolates had fewer secondary metabolite pathways than identified in their closest reference (LI7 and LI67). Identification of antibiotic resistance genes by CARD[191] revealed that *mfd*, *ileS* and *LlmA* genes, implicated in resistance to fluoroquinolones, mupirocin and antibiotic target modification, respectively, were found in all isolates and references (Figure 5C). This suggests the presence of a core resistome within the Lachnospiraceae family. However, no isolate was found to have an



B)


Polyketide
Sactipeptide
Class II lantipeptide
Nonribosomal peptide
Unknown thiotemplated cluster type
Class II lantipeptide/polyketide
Anaerostipes caccae DSM 14662
LI 7
Ruminococcus torques ATCC 27756
LI 52
Ruminococcus lactaris ATCC 29176
LI 10
LI 11
Dorea longicatena DSM 13814
LI 8
LI 58
LI 35
Coprococcus comes ATCC 27758
LI17
LI 18
Blautia obeum ATCC 29174
LI 6
LI 45
LI 48
LI 66
LI 67

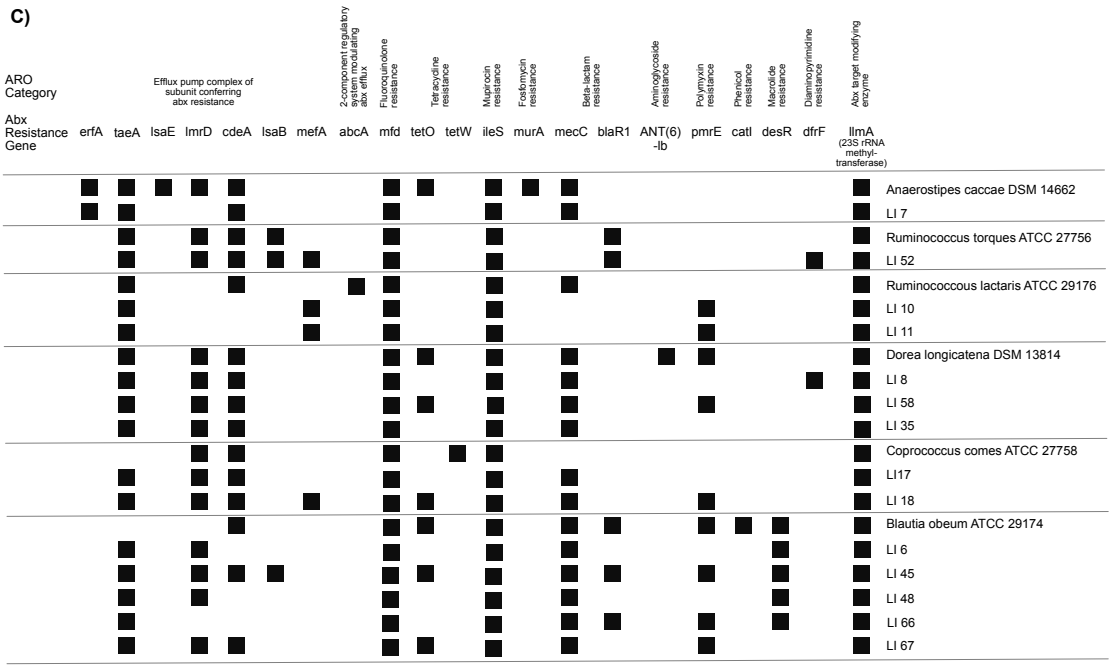


Figure 5. Comparison of 14 Lachnospiraceae genomes with closest reference genomes for identification of A) prophages, B) secondary metabolite predictions, and C) antibiotic resistance genes show genetic diversity. A5-miseq assembled contigs of Lachnospiraceae isolates and NCBI reference genomes were analyzed with A) PHASTER, colours indicated completeness of prophage regions: green – intact, blue – questionable, red - incomplete B) PRISM3 and C) Comprehensive Antibiotic Resistance Database (CARD) Resistance Gene Identifier, with only strict and perfect hits reported. Numbers in boxes represent number of pathways identified for each secondary metabolite class.

identical set of resistance genes compared to their reference, demonstrating that reference isolates likely do not represent the antibiotic resistance profiles of all strains within a species. Similar to the identification of prophages and secondary metabolites, LI10 and LI11 did have the same antibiotic resistance gene profiles, and are likely clonal strains, but did have a different profile than *R. lactaris* ATCC 29176. Within-donor intra-species diversity was also apparent in the antibiotic resistance genes, as LI18 possessed three more genes than LI17, and LI66 & 67 also had unique profiles. Within the *B. obeum*-

group, the *taeA* gene was found in all isolates except for the reference genome, but none of the isolates in this group had an identical CARD profile.

Discussion

In this study, the diversity within the Lachnospiraceae strains from two donors were characterized phenotypically and genetically. The phenotypes of 42 Lachnospiraceae strains exhibited extensive diversity, and although diversity was expected between different species, diversity was also observed between strains of the same species, even when strains were isolated from one individual. Genetic diversity was also observed although only a limited number of strains were sequenced. Genetic variability was detected between strains, and when strains were compared to reference genomes. It was particularly striking that, except for *C. comes*, no strains within a species had the same phenotype profiles. *C. comes* may have less functional diversity compared to other species, and based on genomic comparisons, the two strains sequenced (LI17 and LI18) were also similar to each other. However, the genomes still possessed differences in the predicted prophages, secondary metabolites and antibiotic resistance genes, indicating that these were likely not clonal strains. As the production of exoenzymes such as proteases and DNases are usually associated with pathogens, it was interesting to note that several Lachnospiraceae isolates expressed activity during these assays, however it is not known if these would be produced within the host. Potential production of hydrogen peroxide was also unexpected as it is produced from aerobic respiration, which is unlikely to be the case for the anaerobic Lachnospiraceae. This may

be false-positives from the Prussian blue assay and should be confirmed with hydrogen peroxide probes[192].

Lachnospiraceae isolates demonstrated a lack of resistance to antibiotics as there was no growth in 4 of the 6 antibiotics tested. As this family makes up a large proportion of gut microbiota, the antibiotic sensitivity of Lachnospiraceae likely contributes to the perturbations observed in the microbiota after antibiotic treatment[146,193]. Importantly, strains within a species had different resistant profiles to PMB and vancomycin, in addition to unique sets of resistant genes, which indicates that antibiotic resistance profiles cannot be generalized. This could have implications for microbiome studies involving antibiotics. Although there were several antibiotic resistance genes identified by CARD in a majority of the isolates, further studies should determine if resistance to the antibiotics predicted is observed experimentally. The sensitivity of isolates to PMB, even at 16 μg/ml, was unexpected as Cullen *et al*. had shown by screening 7 Firmicutes from the gut microbiome, that the lowest PMB MIC observed was 53 μg/ml, and 4 strains had MICs >1024 μg/ml[194], but we did not detect growth of Lachnospiraceae isolates at 256 μg/ml. This may be due to differences in antibiotic testing methods as E-test strips were used in that study, but may also reflect the phenotypic heterogeneity among isolates.

The phenotypes in this study were characterized for each strain in isolation, but expression of phenotypes could be affected by the host or by the presence of other community members, especially if Lachnopsiraceae are capable to responding to AI-2, which was not explored in this study. It has not been experimentally shown if Lachnospiraceae are able to respond to AI-2, or are only able to produce the signal,

although preliminary analysis has identified homologues of *lsrB* and *lsrD* genes, which are part of an AI-2 regulated operson[195]. Additionally, a l*uxS* homologue was found in the *R. torques* LI52 strain that produced low AI-2, suggesting the possibility that AI-2 may accumulate later in growth, as observed in other organisms[196]. Further studies are required to determine this, as well as the concentrations required to impact gene expression, and what processes, if any, are controlled by this quorum sensing. In other bacteria, AI-2 controls the expression of virulence factors and biofilm formation[195], and in *Clostridium perfringens,* AI-2 signaling is involved in toxin production[197]. Two recent studies have shown interesting roles for AI-2 signaling in the gut. In the first study, *Ruminococcus obeum* was able to restrict *in vivo* infection by *Vibrio cholerae* in a AI-2 dependent mechanism[198]. Thompson *et al.* demonstrated that high levels of AI-2 favoured the expansion of Firmicutes, including a Lachnospiraceae OTU, after antibiotic treatment. These studies indicate the potential impact on gastrointestinal health through interspecies signaling by the Lachnospiraceae.

Through comparative genomics of the Lachnspiraceae strains and the reference genomes, it was observed that genetic differences existed. It is possible that this family has a large pangenome, which could explain the gaps in the BRIG comparisons for the reference genomes and the strains in this study, as well as the heterogeneity observed in prophages, secondary metabolites and antibiotic resistance genes. The low identity coverage of a majority of the 14 strain genomes when compared to the references indicated that these RefSeq genomes were not good representatives of the genomic content of the commensal strains. As none of the RefSeq genomes were complete

assemblies, and the number of references sequences available for each species ranged from one to seven (in comparison to *E. coli*, which has 6783 NCBI assemblies, 358 of which are complete), it is clear that more representatives from this family need to be isolated and sequenced.

By comparing commensal strains with the closest RefSeq genomes, it was observed that sections of the isolate genomes were not in the reference strains, and further annotation is required to determine if any genes are encoded in those segments. As reference genomes are used to map reads in metagenomic studies, it is crucial that the functional potential of a population is represented. However, many genes are predicted to encode hypothetical proteins, and further experimental studies and bioinformatics analyses would be required to characterize the Lachnospiraceae genomes. Although novel methods have been developed to genetically manipulate the Gram-negative *Bacteroides* from the gut microbiota[199,200], these methods are lacking for anaerobic Gram-positive bacteria but will be required to experimentally test functions of unknown genes. From PHASTER predictions, it was observed that all genomes contained potential phage regions, but the presence of functional phages needs to be tested experimentally by inducing the lytic cycle as prophage regions could be defective or cryptic[189]. The presence of functional prophages is important as prophages can impact community dynamics if the lytic cycles is induced, and prophages can contribute to intra-species bacterial variation in traits including antibiotic resistance. The predicted production of secondary metabolites by Lachnospiraceae isolates, especially those not present in reference strains, is also of interest as natural products form the basis of many pharmaceutical drugs, including

antimicrobials, so discovering potential functionally important products of these strains should be pursued.

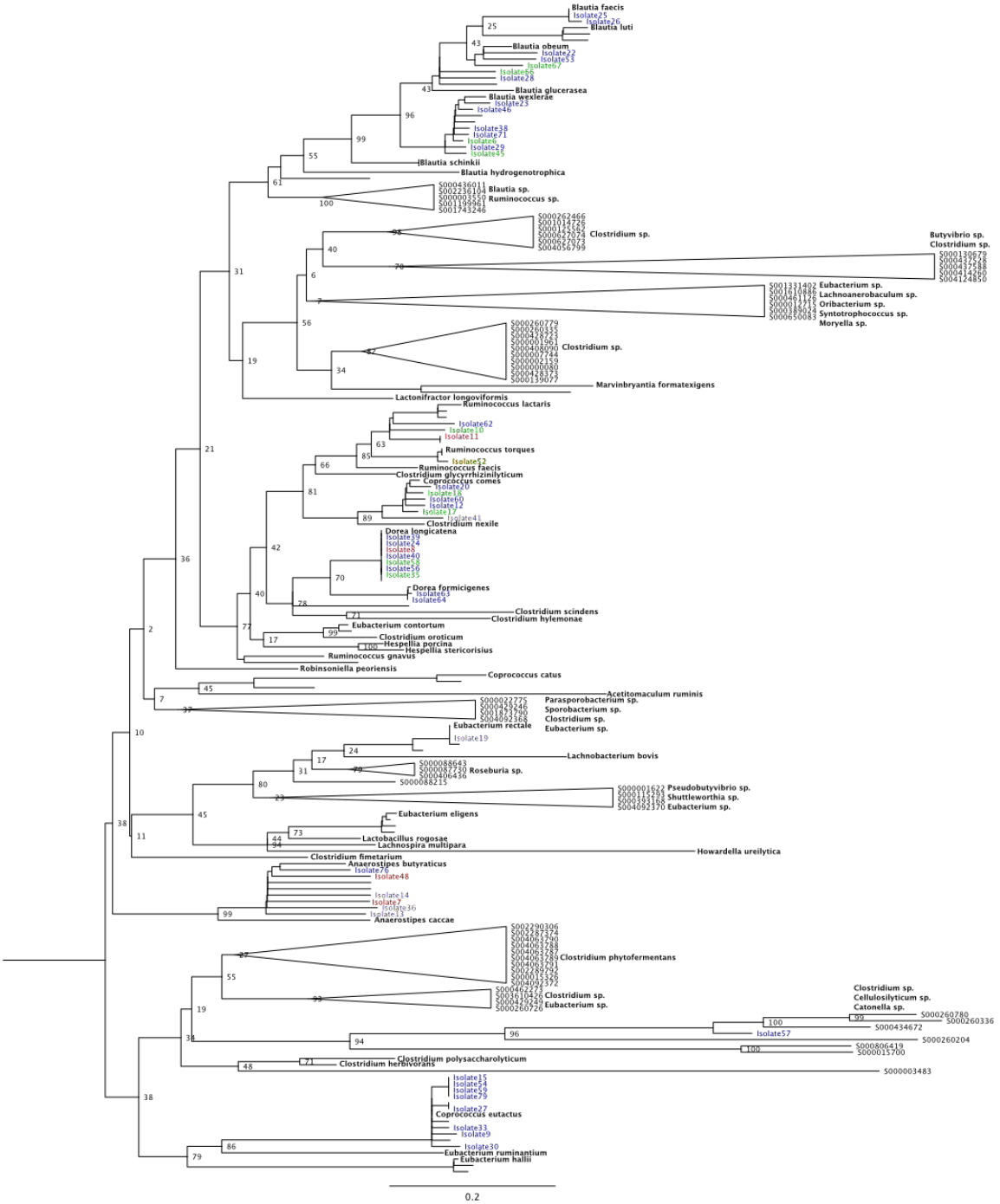
The Lachnospiraceae isolates in this study were obtained from the stool samples of only two donors that were cultured on two media, and from 140 colonies sequenced, 72 were Lachnospiraceae. This demonstrates that several species from this family are readily isolated and amenable to further studies. There are 24 genera in the family, but only seven were represented in this study, so there is likely much more functional diversity in the Lachnospiraceae, and phenotypic and genetic representation of more isolates is required to represent the population. As 16S rRNA gene sequencing studies usually only identify Lachnospiraceae at the family or genus-level, it is important to recognize the species and intra-species diversity that is being overlooked, and that reference strains may not represent the functional potential of the whole species. This may have important implications for understanding the role of Lachnospiraceae in maintaining health.

Acknowledgements

We thank JC Szamosi for the helpful discussion and advice on genome comparisons.

Table 1. Assembly statistics of Lachnospiraceae isolates and RefSeq reference genomes. Lachnospiraceae assembly information obtained from output of A5-Miseq.

Genome	Total length	Number of scaffolds	N50
Blautia obeum ATCC 29174	3 625 508	54	147 259
LI6	4 042 288	153	105 854
LI45	3 696 807	64	480 071
LI48	4 034 684	143	102 241
LI66	3 727 485	80	163 438
LI67	3 473 882	52	184 195
Dorea longicatena DSM 13814	2 915 433	36	177295
LI8	3 343 305	104	73 481
LI35	3 713 661	144	59 504
LI58	3 107 685	69	101 749
Ruminococcus torques ATCC 27756	2 741 706	41	209 206
LI52	3 236 280	73	106 766
Coprococcus comes ATCC 27758	3 242 215	13	847 455
LI17	3 518 088	113	101 550
LI18	3 530 653	62	150 219
Ruminococcus lactaris ATCC 29176	2 731 235	47	91 712
LI10	3 461 809	128	74 099
LI11	3 706 193	482	71 026
Anaerostipes caccae DSM 14662	3 606 936	26	261 254
LI7	2 909 446	74	80 667



Supplementary Figure 1. 16S rRNA gene (V1-V4) tree of Lachnospiraceae isolates in relation to RDP Lachnospiraceae type isolates. Isolates in red were sequenced only, isolate in blue were phenotyped only, isolates in green were sequenced and phenotyped. Adapted from Lau *et al.* Genome Med. 2016

Supplementary Figure 2. Visualization of predicted secondary metabolite clusters from PRISM3 for LI48 and LI6 isolate genomes, and *Blautia obeum* ATCC 29174.

Chapter 4

Characterization of genetic and phenotypic diversity of *Escherichia coli* isolates from the human gut microbiota

Preface

Research presented as part of this chapter has been prepared for publication.

Lau JT, Dhanjal S and Surette MG. Characterization of genetic and phenotypic diversity of *Escherichia coli* isolates from the human gut microbiota. 2017.

Contributions:

JTL and SD performed the phenotyping experiments and analyzed the data, SD isolated the strains and did the RAPD PCR profiling, JTL carried out the cytotoxicity assays and pulse-field gel electrophoresis. JTL and MGS designed the experiments and wrote the manuscript.

Abstract

There have been few studies characterizing the extent of within-species diversity of commensal bacteria of the gastrointestinal tract. Although extensive diversity exists in pathogenic *Escherichia coli*, the strain variability of commensals contributes to the overall diversity of the human gut microbiome and should be described. In this study, both phenotypic and genomic approaches were applied to characterize the diversity of *E. coli* from four individuals. 78-94 isolates from each donor were characterized for several *in vitro* phenotypes, and genetic diversity in the overall population was assessed. Additionally, one individual was sampled twice to determine changes in diversity over time. Phenotypic diversity was observed among isolates from the same individual, and between individuals, the extent of variability differed in each *E. coli* population. Resistance profiles against nine antibiotics differed between isolates from the same sample, and some isolates were also cytotoxic to host cells, while other within-donor strains were not. Additionally, genetic diversity was observed in strains isolated from an individual, and this did not always agree with the amount of phenotypic diversity observed. Changes to the dominant strain were also observed in an individual over time, with a replacement in the most prevelant genotype and phenotype between two sampling time points. The combination of phenotypic and genetic approaches in this study allowed for greater resolution of within-species diversity than either method alone. The diversity characterized in *E. coli* may extend to other commensal groups, and could play an important role in the functional capacity of the human gut microbiome.

Introduction

Escherichia coli is a common but low abundant commensal of the human gut microbiota, and one of the earliest colonizers of the newborn gastrointestinal tract[3,201,202]. However, pathogenic *E. coli* strains can be responsible for causing intestinal and extra-intestinal infections through several different mechanisms[203]. Through the use of next-generation sequencing, microbiome studies have described the composition and functional capacity of the overall community, but less is known about the within-species diversity among commensal bacteria, including *E. coli*. This can be attributed to the lack of species-level resolution in most 16S rRNA gene sequencing studies. Strain-level differences are only attainable through shotgun metagenomics if sequencing depth is sufficient to provide good coverage, but even this can be difficult for lesser abundant bacteria. *E. coli* species diversity has been studied through the isolation of strains, with a focus on pathogenic pathovars, and studies on commensal strains usually survey the genetic variability in only a small number of isolates per individual, which may not adequately capture the complexity of the entire population[204–207]. It is important to describe the overall strain diversity and population structure of a community since this contributes to the total functional repertoire of the microbiome.

Studies of pathogens, like *Pseudomonas aeruginosa* during chronic infection of cystic fibrosis airways[208], and *E. coli* from extra-intestinal infections[209,210], show that bacteria adaptively evolve during the course of an infection, resulting in genetic and phenotypic diversification. Mouse studies have also shown that *E. coli* adapts to the host during colonization, whereby mutations that allow for adaptation to the gut environment

confer increased fitness and lead to increased diversity[211–213]. The resulting within-species diversity is a direct result of host factors, including exposures to the immune system[213], nutrient availability[211], rate of transit through the gastrointestinal tract[214], and also inter- and intra-species competition with other members of the microbiota[215].

In this study, we applied both phenotypic and genetic approaches to explore the diversity within an extensively sampled *E. coli* population from four individuals. Isolates from each individual was assessed for phenotypes including hemolysis, motility, biofilm production, antibiotic resistance and cytotoxic effect. Genetic variability of *E. coli* within an individual was determined through a combination of random amplified polymorphic DNA (RAPD) PCR and pulse-field gel electrophoresis (PFGE). Other studies have shown that greater bacterial diversity can be revealed through phenotypes, such as antibiotic resistance profiles, than through genetic approaches, such as ribotyping and PFGE[205,208]. Additionally, one individual was profiled at two time points to determine changes in the diversity of *E. coli* over time. This combination of approaches will allow for greater resolution of heterogeneity within commensal *E. coli* populations, and provides insight into the within-species diversity present in the human gut microbiota.

Materials and methods

E. coli isolation and identification from fecal samples

Fresh fecal samples were collected from two healthy donors (Donor 1 & 2) and two Irritable Bowel Syndrome patients (Donor 3 & 4) and immediately stored at -80°C. Donors provided consent prior to participation. 0.1 g of sample was resuspended in 1 mL

of Luria-Bertani (LB) broth (Difco), serial dilutions were plated onto MacConkey agar and incubated at 37°C overnight. 96 colonies were isolated from each sample, streaked for purity, and frozen in 1-ml deep 96-well plates in 25% glycerol at -80°C.

The isolates were confirmed as *E. coli* with primers targeting the universal stress protein (*uspA*), as previously described[216]. For each strain, a colony was boiled in 50 µL 5% Chelex for 15 minutes, and supernatant was used as template for PCR. Reactions contained 5 µL of DNA, 0.2 pmol of forward primer (5'- CCG ATA CGC TGC CAA TCA GT-3'), 0.2 pmol of reverse primer (5'-ACG CAG ACC GTA GGC CAG AT-3'), 1 U of Taq, 0.2 mM of each dNTP (New England Biolabs), 5 µL of 10X PCR buffer (Invitrogen), and 1.5 mM $MgCl_2$. PCR cycling conditions was as follows: 95°C for 5 minutes, 30 cycles of 94°C 2 min, 64°C 1 min, 72°C 1 min, final extension 72°C for 10 minutes.

Phenotypic Assays

Phenotypic assays were performed by pin-replicating from 96-well 200 µl LB overnight cultures. All incubations were at 37^{o}C unless otherwise specified. The isolates were pin-replicated onto 150 mm MacConkey agar plates and incubated overnight for description of colony morphologies. Hemolysis was determined on Columbia blood agar with 5% sheep blood, and incubated for 48 hours. Swimming was measured on LB with 0.25% (w/v) agar, and incubated for 4 hours, then plates were scanned and zones of motility were measured by Fiji[217]. Swarming was measured on LB with 0.6% (w/v) agar and 0.5% glucose, and overnight incubation. Biofilm production was performed in 96-well plates in

LB using the crystal violet staining, as previously described[218]. For each isolate, biofilm production was normalized against cell growth (OD_{600nm}) and classified as positive if normalized OD_{595nm} was greater than 0.1. Auxotrophy was investigated by first pin-replicating into M9 minimal medium to dilute residual LB, then pin-replicated again to grow the isolates in M9 minimal medium with 0.5% glucose, with and without 0.1% casamino acids (Sigma-Aldrich), shaking. Isolates were classified as auxotrophic if growth in supplemented media was at least two-times higher than in non-supplemented media. High RNA polymerase sigma S (RpoS) producers were identified by iodine staining of glycogen as previously described, since glycogen production is regulated by RpoS[209]. Red dry and rough (RADR) colony morphology, a phenotype also dependent on RpoS[219], was determined by growth on LB without NaCl and supplemented with 40 μg/ml Congo red and 20 μg/ml Coomassie blue, and plates were incubated for 48h at 28°C.

Resistance to ampicillin, cephalexin and ciprofloxacin was used for biotyping, and determined by pin-replicating from 96-well 200 μl LB overnight cultures onto LB agar plates containing one of three antibiotic concentrations: clinical breakpoint concentration for Enterobacteriaceae according to EUCAST 2015[220], four-times below the clinical breakpoint concentration to identify highly sensitive isolates, and four-times above the clinical breakpoint concentration to identify highly resistant isolates. Growth was assessed after overnight incubation. Resistance to polymyxin B, colistin, aztreonam, ceftazidime, chloramphenicol, gentamicin and trimethoprim was determined by pin-

replicating 96-well 200 µl LB overnight cultures into 200 µl LB containing antibiotics at the clinical breakpoint concentration. Growth after overnight incubation with shaking was measured by $OD_{600nm,}$ and resistance was classified as OD_{600nm} greater than 0.2. Heatmap of antibiotic resistance profiles was generated in R using gplots.

Cytotoxicity of HT-29 cells

Human colonic epithelial HT-29 cells were seeded in 96-well plates at a density of $1x10^5$ cells per well for 24 hours in McCoy's 5S Modified Medium (Gibco) with 10% fetal bovine serum (FBS; Gibco). One hour before the addition of *E. coli*, medium was replaced with 50 µl of fresh medium without FBS. Log-phase *E. coli* cultures were diluted to OD_{600nm} 0.5 in cell culture medium and 100 µl was added to HT-29 cells, and incubated for 4 hours at 37^oC, 5% CO_2. Lactate dehydrogenase release was measured with the Promega CytoTox 96 Non-radioactive Cytotoxicity Assay according to manufacturer's instructions.

Random Amplified Polymorphic DNA (RAPD) PCR

To determine the number of genetic strains within an individual, 20 µl of an overnight culture in LB was mixed with 50 µl 5% Chelex, boiled for 15 min and 1 µl of supernatant was used as DNA template. PCR was performed as previously described[221] with RAPD-1254 primer (5′-CCG CAG CCA A-3′). PCR products were visualized with a 2% agarose gel, separated at 60V for 2 hours.

Pulse field gel electrophoresis (PFGE)

PFGE was used to determine strain relatedness between individuals, and 12 isolates from each donor were profiled. PFGE was performed according to the Standard Operating Procedure for PulseNet PFGE of *E. coli* O157:H7, *E. coli* non-O157, *Salmonella* serotypes, *Shigella sonnei* and *Shigella flexneri* (https://www.cdc.gov/pulsenet/pdf/ecoli-shigella-salmonella-pfge-protocol-508c.pdf*)*. Colonies were streaked on MacConkey for overnight growth, and chromosomal DNA was digested with *Xba*I. Electrophoresis of digested fragments was performed using BioRad CHEF-DR III in 1% agarose gels. Running conditions were as follows: 0.5X TBE buffer, 14^{o}C, initial switch time 6.76 s, final switch time: 35.58 s, voltage: 6V, included angle: 120^{o}, run time: 18 hours. Gel images were inverted and DNA patterns were analyzed with GelCompar II using the Dice similarity coefficient with 1% band tolerance. Patterns were clustered by Unweighted Pair Group Method with Arithmetic Mean (UPGMA).

Results

E. coli isolates exhibit within- and between-individual phenotypic diversity

The *E. coli* population from the gut microbiota of four individuals was extensively sampled, with a minimum of 78 isolates characterized from each donor. The phenotypes of each donor are summarized in Table 1. Between the four individuals, the proportion of isolates that were positive for each phenotype tested was different, demonstrating inter-individual variability in the expression of *E. coli* phenotypes. Aside from swarming motility, which was exhibited by greater than 95% of all isolates from each individual,

and formation of RDAR colony morphotypes, which was expressed by fewer than 3% of isolates, there were differences in the expression of all other phenotypes across the four populations.

To compare the overall within-sample phenotypic diversity of each individual, each unique combination of the 11 phenotypes observed in Table 1 was designated as a unique biotype. The phenotypes included in the biotype classification were: colony morphology on MacConkey agar, β-hemolysis on sheep's blood, swimming, swarming, biofilm production, amino acid auxotrophy, resistance to ampicillin, cefalexin, and ciprofloxacin, glycogen staining and RDAR morphotypes. A larger number of biotypes represents greater phenotypic diversity within a sample, thus, the *E. coli* isolates from Donor 3 and 4 (with 29 and 38 biotypes, respectively) were more phenotypically diverse than Donor 1 & 2 isolates, which comprised 14 and 6 biotypes, respectively (Figure 1). Interestingly, this correlated with colony morphologies, as Donor 1 and 2 only had a few different colony types, whereas 16 and 17 different colonies were described from Donor 3 and 4. The distribution of isolates into biotypes was also different between the donors; Donor 3 isolates mainly belonged to a single dominant biotype, whereas the other three samples had two or more common biotypes. However, all four populations had several biotypes that were unique (i.e. no other isolate tested from the same individual expressed the identical combination of phenoypes).

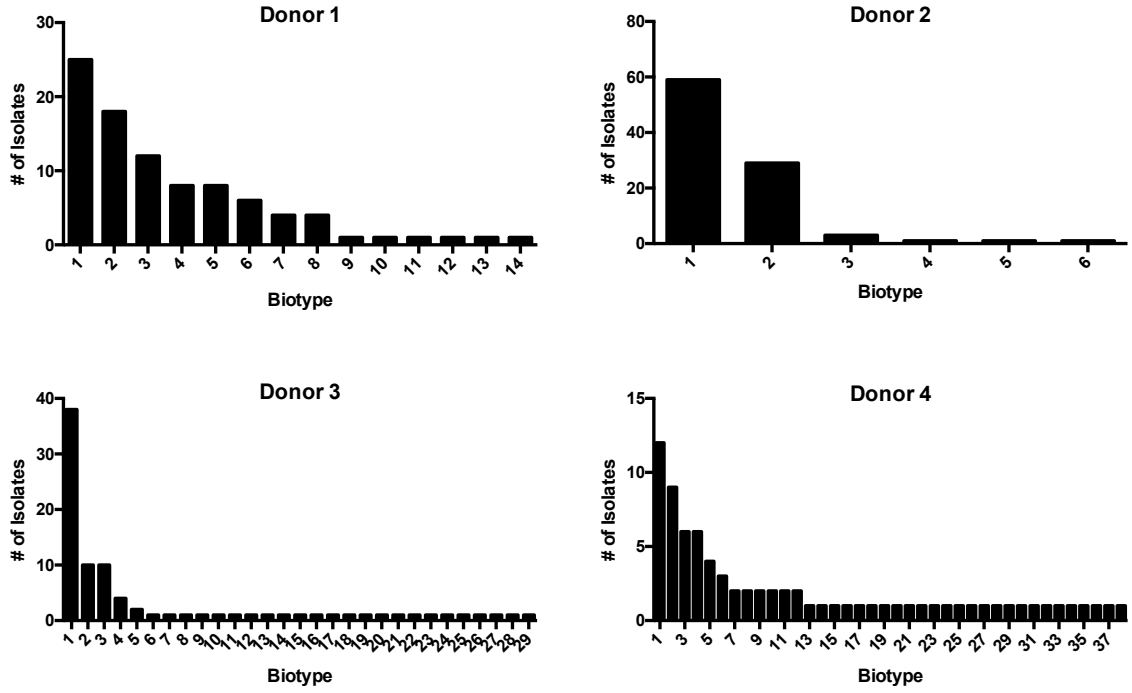


Figure 1. Distribution of biotypes (unique combination of phenotypes) for *E. coli* isolates from 4 individual samples. Phenotypes were only compared within-sample, and included colony morphology on MacConkey agar, beta-hemolysis on sheep's blood, swimming, swarming, biofilm production, amino acid auxotrophy, and resistance to ampicillin, cephalexin, ciprofloxacin, glycogen staining and RDAR morphotypes.

Resistance to nine antibiotics were tested, and resistance profiles of the *E. coli* isolates was observed to be variable between individuals (Figure 2). Donor 2 had the highest amount of antibiotic resistance, with growth of all isolates observed in ampicillin, cefalexin, gentamicin, ciprofloxacin and trimethoprom at clinical breakpoint concentrations. Donor 1 had the least resistant profile. Comparing all donors, no antibiotic had the same level of resistance across all individuals, except for chloramphenicol, which did not allow the growth of any isolate. Within-individual diversity in antibiotic resistance was also observed among isolates, as not every isolate

expressed the same antibiotic resistance profile. For example, although all 91 isolates from Donor 1 were ampicillin resistant, only 35 of those isolates were also resistant to trimethoprim.

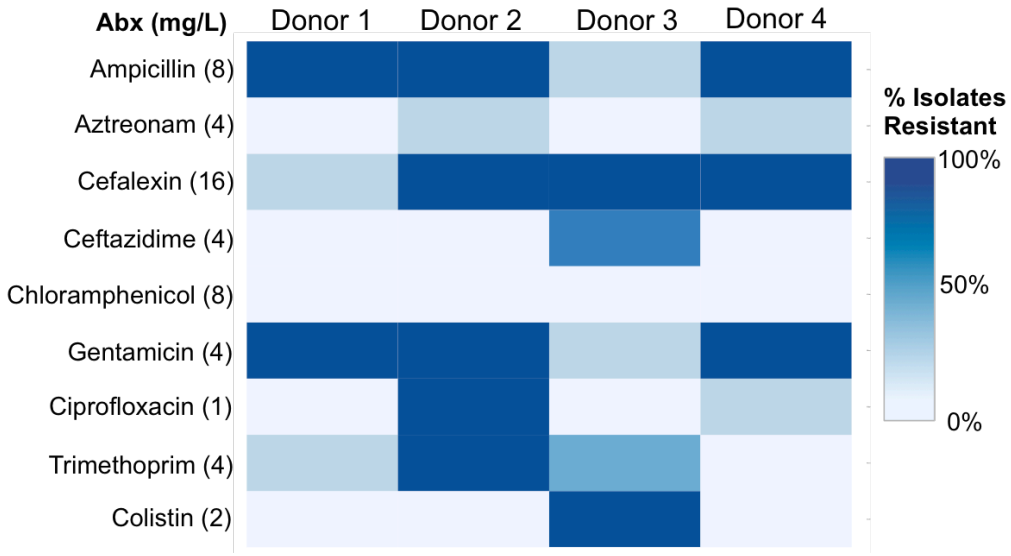


Figure 2. Antibiotic resistance profiles of *E. coli* isolates from four individuals demonstrates extensive variability. Heatmap shows proportion of isolates resistant to 9 antibiotics at the EUCAST clinical breakpoints for Enterobacteriaceae (Ver 5.0, 2015). Breakpoint concentrations (mg/L) are in parentheses next to antibiotic name.

E. coli strains isolated from the same individual can differentially affect host cells

To determine if *E. coli* strains from the same individual could vary in their interactions with host cells, five isolates from each of the four samples was co-incubated with human colonic epithelial HT-29 cells, and the cytotoxic effect of the bacteria was determined by measuring release of lactate dehydrogenase (Figure 3). Interestingly, 4 out of 5 isolates from Donors 1-3 were able to induce host cell death at greater than 50% of the Max LDH control, and one isolate from each donor had similar levels of LDH as the

untreated controls. All isolates from Donor 4 were comparatively less cytotoxic, with the greatest LDH release of 24%.

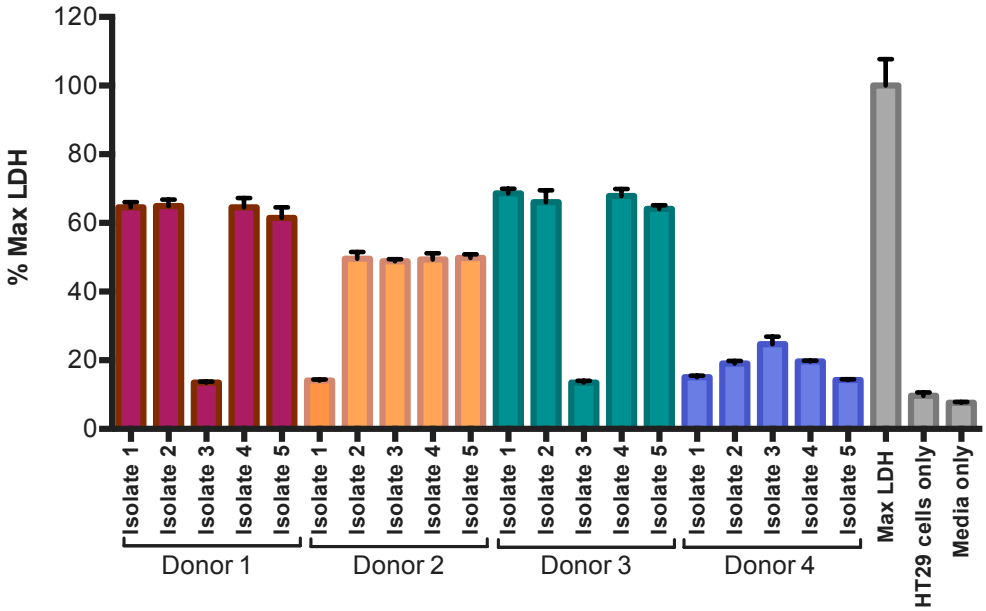


Figure 3. Cytotoxicity of *E. coli* isolates is variable within-individual. Five *E.* coli isolates from each of 4 donors were co-incubated with HT-29 cells for 4 hours. Lactate dehydrogenase (LDH) release from the supernatant was normalized to the Max LDH control (complete lysis of untreated HT-29 cells). Error bars represent SD, n=3.

Genetic diversity is observed among *E. coli* isolates and may not be indicative of phenotypic variability

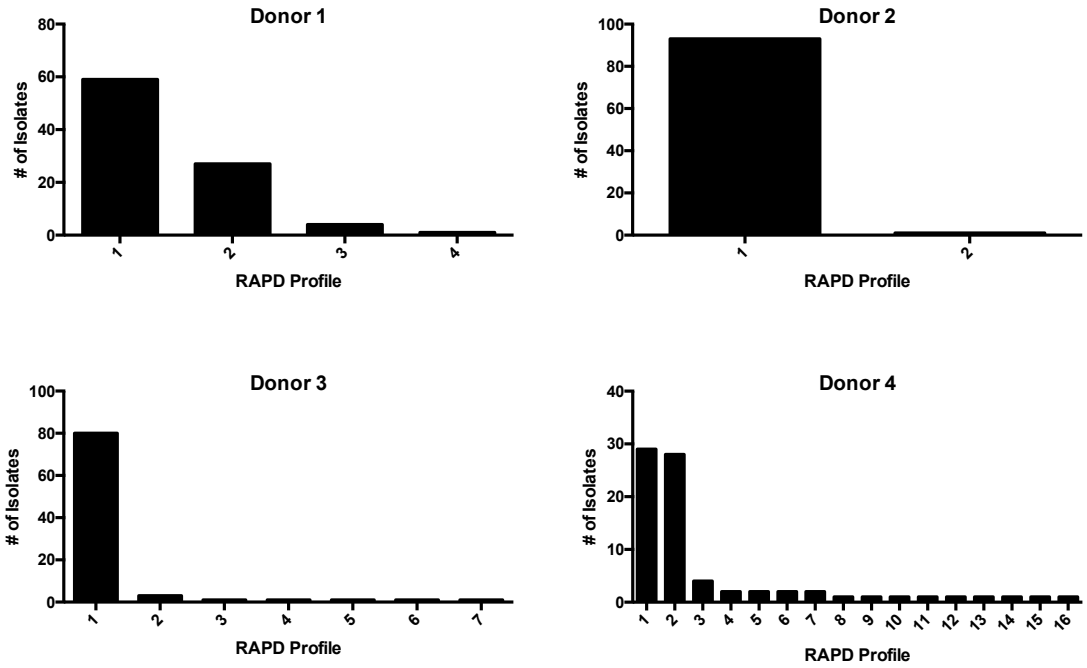
To screen the overall genetic diversity of *E. coli* isolates within an individual, RAPD PCR profiles were compared within each sample (Figure 4A). Donor 2 had the least genetic diversity, with only two RAPD patterns observed, while Donor 4 had the most diversity with 16 RAPD profiles. Donor 1 and 3 isolates were distributed across four and seven RAPD profiles, respectively. Donor 2 and 3 were dominated by one RAPD pattern, suggesting the presence of one dominant genetic clone, whereas Donor 1 and 4

isolates reflected two dominant RAPD patterns. Interestingly, for all individuals, there were fewer RAPD profiles than biotypes, indicating that commensal *E. coli* have less genetic diversity than phenotypic diversity. Genotyping an additional 356 isolates from Donor 1 did not increase the number of RAPD profiles observed, suggesting that genetic diversity in this sample was represented by the initial screening of 91 isolates (Suppl. Fig. 1).

To increase the resolution of genetic diversity and to compare genotypes of *E. coli* isolates between individuals, 12 isolates from each donor were profiled by PFGE, including isolates with different biotypes and RAPD profiles (Figure 4B). For Donor 1, 2, and 4 (with the exception of one Donor 4 isolate), all isolates grouped by donor. Isolates were more genetically similar within an individual than between individuals, and shared strains were not observed, although Donor 3 isolates were divided into two clades, and one was closely related to Donor 1 isolates. It was also observed that RAPD profiles did not always agree with the results of PFGE, as genetic diversity was underestimated by RAPD profiles from Donor 1 and 2. Donor 2 isolates were seperated into two distinct PFGE clusters, but all isolates had identical RAPD patterns.

Interestingly, genetic diversity did not correspond to the phenotypic diversity in all four samples. For example, the first three isolates from Donor 4 had identical PFGE patterns, so were clonal strains, but exhibited variability in cefalexin and ciprofloxacin resistance, biofilm production and auxotrophy. Conversely, for the 12 Donor 2 isolates, all phenotype profiles were identical except for auxotrophy, but the isolates are divided into two distinct PFGE clusters, which was not reflected in the phenotyping. Donor 3

A)


Donor 1
# of Isolates
RAPD Profile
Donor 2
# of Isolates
RAPD Profile
Donor 3
# of Isolates
RAPD Profile
Donor 4
# of Isolates
RAPD Profile

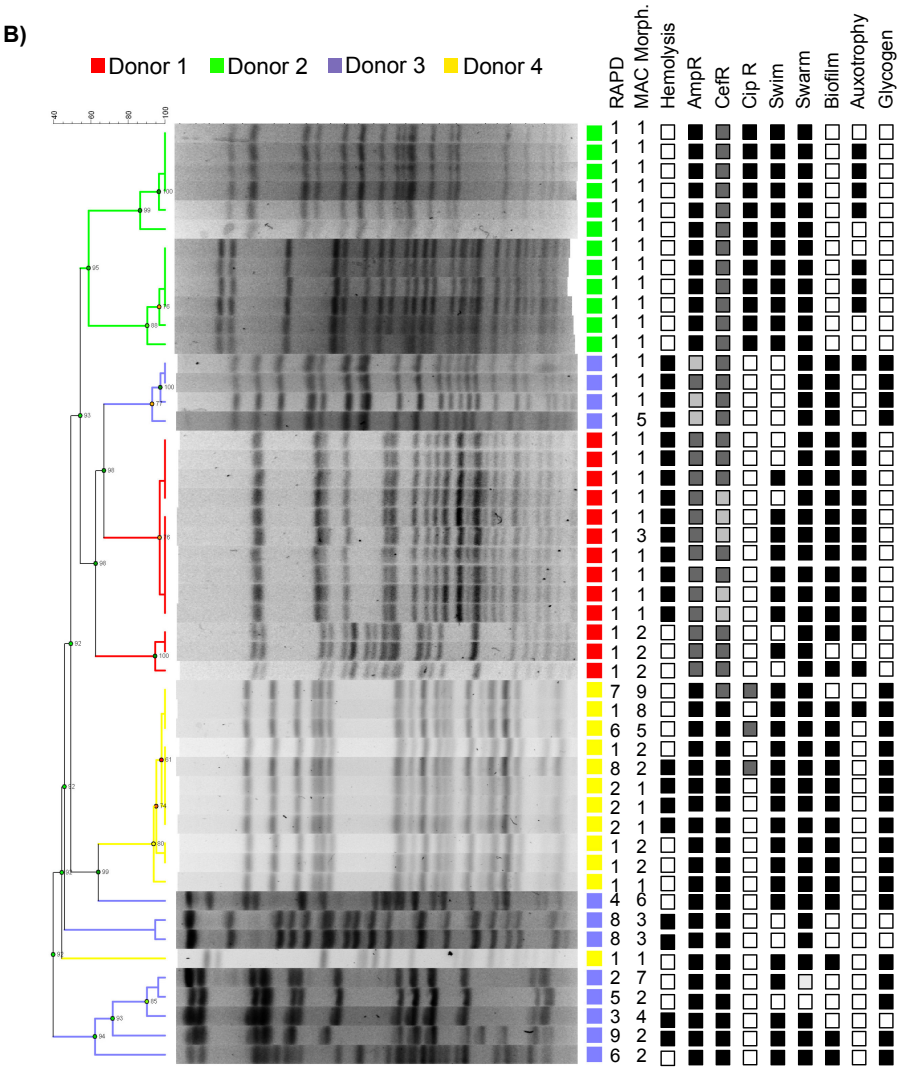


Figure 4. Genetic diversity of *E. coli* varies between four individuals A) Distribution of random amplified polymorphic DNA (RAPD) PCR profiles for *E. coli* isolates, compared within-sample. B) Pulse-field gel electrophoresis of *Xba*I fragments from 12 *E. coli* isolates from each donor, compared with Dice similarity coefficient and clustered by UPGMA. RAPD PCR pattern and colony morphology type on MacConkey (MAC) compared within-individual only. Expression of hemolysis, ampicillin (AmpR), cephalexin (CefR) and ciprofloxacin (CipR) resistance, swim, swarm, biofilm production, amino acid auxotrophy and glycogen production for each isolate indicated by heat map. White – negative, black – positive, antibiotic resistance – darker shades indicate growth in increasing concentrations of antibiotics.

strains had the most genetic diversity when compared to isolates from the other three individuals and this corresponded to the highly variable phenotypic patterns in the Donor 3 isolates.

To determine if the number of *E. coli* characterised did not fully represent the diversity in the population due to undersampling, 356 additional isolates from the Donor 1 sample were characterized, and compared to the biotypes observed from the first 91 isolates screened (Suppl. Fig. 1). The 91 original isolates made up 14 biotypes, and the additional 356 isolates comprised 21 biotypes. However, the additional biotypes were almost all unique to one isolate, so the first 91 isolates sampled a representative distribution of the biotypes present in this population, but undersampled the rare biotypes. Since Donor 3 and 4 had many more unique biotypes than Donor 1, it is likely that continued sampling of *E. coli* isolates would uncover more phenotypic diversity in those individuals.

E. coli phenotypic variability changes over time in an individual

To determine if the heterogeneity of the *E. coli* population changes over time within an individual, strains were isolated from two samples from Donor 2, taken 9 weeks apart. There appeared to be little change in the overall diversity of the *E. coli* population since the number of biotypes and RAPD profiles remained similar (Figure 5). However, the dominant RAPD profile from the first sample was not found in the 94 isolates from the second sample, suggesting strain replacement had occurred within the donor. Although the overall number of biotypes in the population was consistent, at the first time

point, the majority of isolates belonged to 2 biotypes, but at the second time point, there was only one dominant phenotype profile. Examining the individual phenotypes tested between the two samples, it was observed that there was a striking change in the proportion of *E. coli* strains that were beta-hemolytic, increasing from no hemolytic isolates to 89% of isolates expressing hemolysis in the second sample (Figure 5E). The proportion of isolates that were auxotrophic and isolates resistant to cephalexin and ciprofloxacin decreased from 100% to 6 and 0%, respectively (Figure 5E). This change of phenotypes resulted in no overlap of biotypes between the two time points, further indicating that strain replacement had occurred during the sampling period.

Discussion

In this study, a minimum of 78 *E. coli* colonies per sample was isolated from four individuals, and phenotypes were compared across isolates. Additionally, the genetic diversity of the *E. coli* populations were assessed with RAPD PCR and PFGE. We observed that the amount of genetic and phenotypic diversity varied across the individuals, with some donors possessing more strains with distinct phenotypes and genetic profiles than others. There was a correlation between the biotypes observed and the number of genetic strains present as determined by RAPD PCR. Additionally, we observed extensive within-species diversity in the donor samples, as *E. coli* isolates from the same individual could express up to 38 unique phenotype profiles (i.e. biotypes). PFGE revealed that isolates from an individual were generally more similar to each other

than isolates from another individual. This individuality of the human gut microbiome has been previously characterized by studies, including the Human Microbiome Project[3],

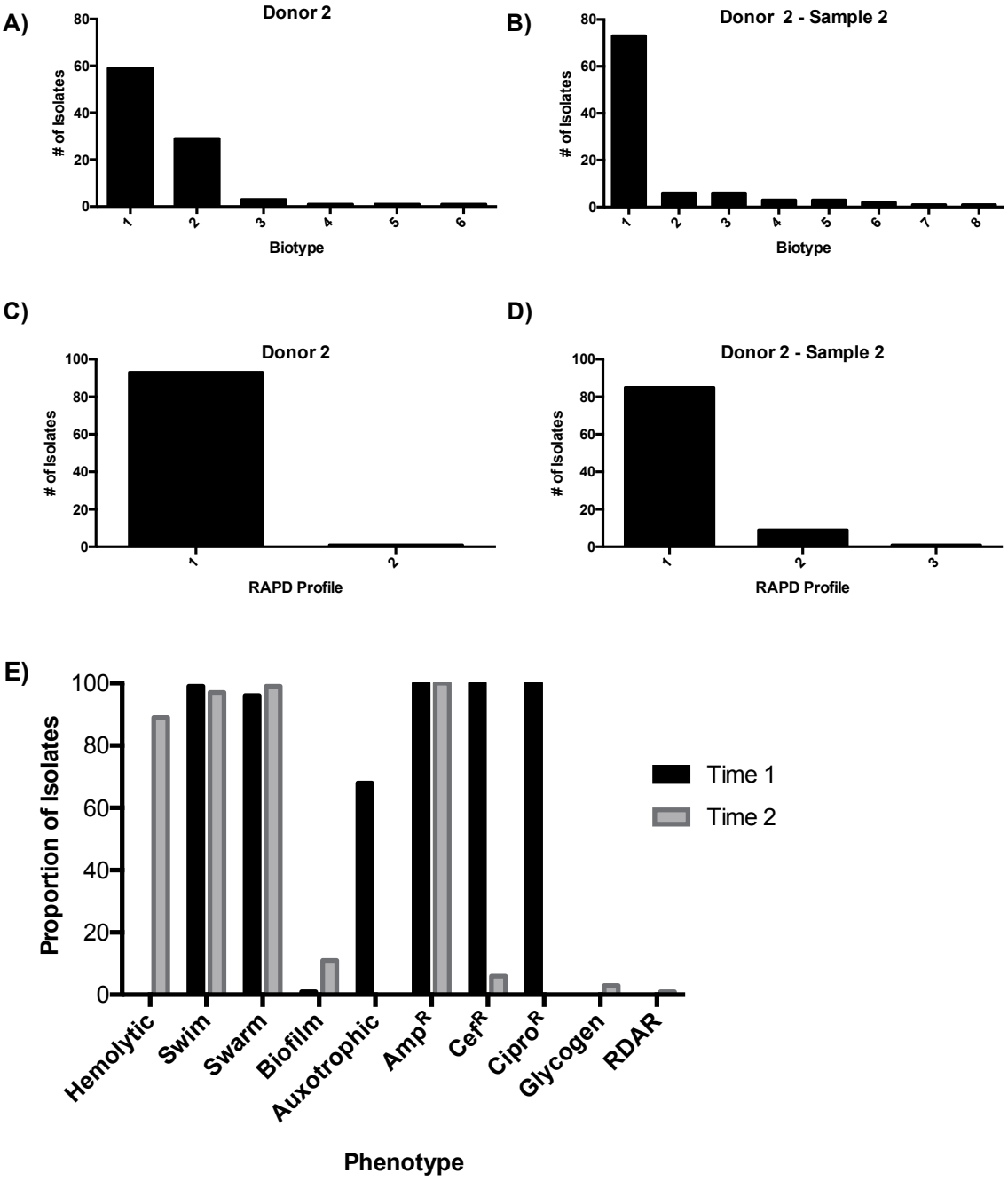


Figure 5. *E. coli* genetic and phenotypic profiles from a healthy donor is variable over time. A-D) Distribution of *E. coli* biotypes and RAPD PCR profiles from a healthy fecal sample during sampling time point 1 (A & C) or from a sample taken 9 weeks later (B & D). E) Proportion of *E. coli* isolates positive for several phenotypes from two Donor 2 fecal samples, 9 weeks apart.

and the within-species variability described in this study further contributes to the inter-individual variability of the human gut microbiome.

Several interesting results were observed in the different phenotype assays in this study. Among the most interesting was the diversity and prevelance of antibiotic resistance among the *E. coli* from the four individuals, as donors were not, to our knowledge, exposed to antibiotics at the time of, or in the three months preceding, sampling. This meant that the change in antibiotic resistance profiles over the two time points for Donor 2 occurred without pressure from antibiotic use, and may reflect the natural dynamics of the population over time. Resistance to ampicillin, chloramphenicol, ceftazidime, aztreonam, gentamicin and trimethoprim in commensal *E. coli* has previously been reported[222,223] but colistin resistance in Donor 3 was not expected. We did not look for plasmids in this study, which can carry antibiotic resistance such as the *mcr*-1 gene conferring colistin resistance[224]. The number of strains observed that were capable of forming rdar colonies was much lower than previously reported, where screens of fecal *E. coli* isolates found 21-36% of strains were positive for this morphotype[225,226]. This discrepency may be a result of host variabiliy, but could also be a result of false-negatives in our study. This could be confirmed with Calcofluor staining for cellulose production, which is required for rdar morphology. The number of isolates that expressed swimming motility was variable between donors, and this phenotype could potentially be

selected for by the host immune system, as flagellin specific immunoglobulins have been shown to prevent bacterial motility and result in down-regulation of flagellin expression[227].

With four individuals, the sample size of this study was small and may not be representative of the diversity observed across a larger population; however the results show that within an individual, we could observe diversity within one species of gut microbiota by using both phenotypic and genetic approaches. *E. coli* isolates were obtained from frozen stool samples which may have selected for more stable phenotypes; however, a comparison of *P. aeruginosa* from fresh and frozen sputum samples demonstrated that no specific loss of phenotypes was observed[228]. Another study had shown that greater *E. coli* genetic diversity was found in biopsy samples than stool[207], therefore the isolation of bacteria from fecal samples in this study may be underrepresenting the diversity present in each donor. Another limitation of this study was the characterization of phenotypes *in vitro*, which may not be similarly expressed *in vivo*, but the results of this study does reflect the potential heterogeneity of the *E. coli* population in the human gut.

Strain diversity within an individual may be a result of colonization with more than one strain (as indicated by isolates with distinct PFGE patterns) or adaptive radiation may be occuring, as some isolates with identical PFGE profiles (i.e. are likely to be clonal isolates) possessed differences in phenotype expression. Additionally, strain replacement may occur over time and contribute to the diversity of the community, as observed for Donor 2. This could potentially be the results of phage dynamics or the successful

colonization of a new strain from the environment. The presence of hypermutator strains[215,229] and phase variation within strains can also contribute to phenotypic variability[230]. The gastrointestinal tract is a complex and dynamic ecosystem, and host factors like the immune system, GI motility and the presence of distinct niches may select for strains or result in adaptation to increase fitness. The ileum is a very different environment from the distal colon in regards to pH, oxygen exposure, presence of immune cells and antimicrobial peptides and bacterial load, and these factors may all be important drivers of bacterial diversity. Additionally, the gut microbiome is a heavily populated environment and bacterial competition for space and nutrients, and the ability to exploit available niches results in increased phenotype variability[215]. In an experiment with mono-colonization of mice with *E. coli* K12 to study the adaptation of a strain to the gut, mutations resulting in increased fitness were observed after 24 days and phenotypic replacement of strains occurred within 2 days[212]. Although mono-colonization studies over simplify the complexity of the human gut, this study indicates that adaptation to the host results in intra-species phenotypic and genetic diversity.

E. coli has been reported to be increased in GI disorders such as Inflammatory Bowel Disease[22] and IBS[231], but it is not yet clear if this is important for the initiation or maintainence of disease. Additionally, studies have not determined if there is only an increase in the amount of *E. coli*, or also increased strain diversity. Although greater phenotypic and genetic diversity was present in Donor 3 & 4, who had IBS, compared to healthy Donors 1 & 2, the sample size was too small to make conclusions about *E. coli* diversity in IBS, and further studies are warranted. The isolates from IBS samples were

not able to induce more host cell cytoxicity, and Donor 4 isolates were actually the least cytotoxic, but *E. coli* strains with a pathogenic phenotype have been implicated in other diseases. Attaching and invading *E. coli* (AIEC) are increased in Crohn's disease[55,56] and *E. coli* strains harbouring a *pks* island encoding colibactin are enriched in IBD and colorectal cancer patients, and these strains promoted invasive carcinoma in a mouse model[80].

In most microbiome studies, the strain-level diversity reported in this study is not captured at the species or OTU level, and furthermore, phenotypic diversity may not be represented by molecular methods. For example, genome sequences will not provide information on the extent of antibiotic resistance in a strain, or variations in motility and biofilm production, but these phenotypes may impact the ability of strains to colonize the GI tract. The pangenome of *E. coli* is large[206,232], so more diversity in this species may be observed than in other species of the gut microbiota, and this has been suggested by a metagenomic study on strain-level variation in the gut[115]. However, this requires confirmation from future studies that include *in vitro* experiments, especially for less characterized bacterial groups. Inter-individual and within-individual diversity of *Bifidobacterium* has been reported[233], suggesting that the results in this study may be reflected in other commensal species. Within-species diversity has important implications for microbiome studies as the use of reference strains may not be representative of the entire species, and this diversity will contribute to the individuality and functional potential of the human gut microbiome.

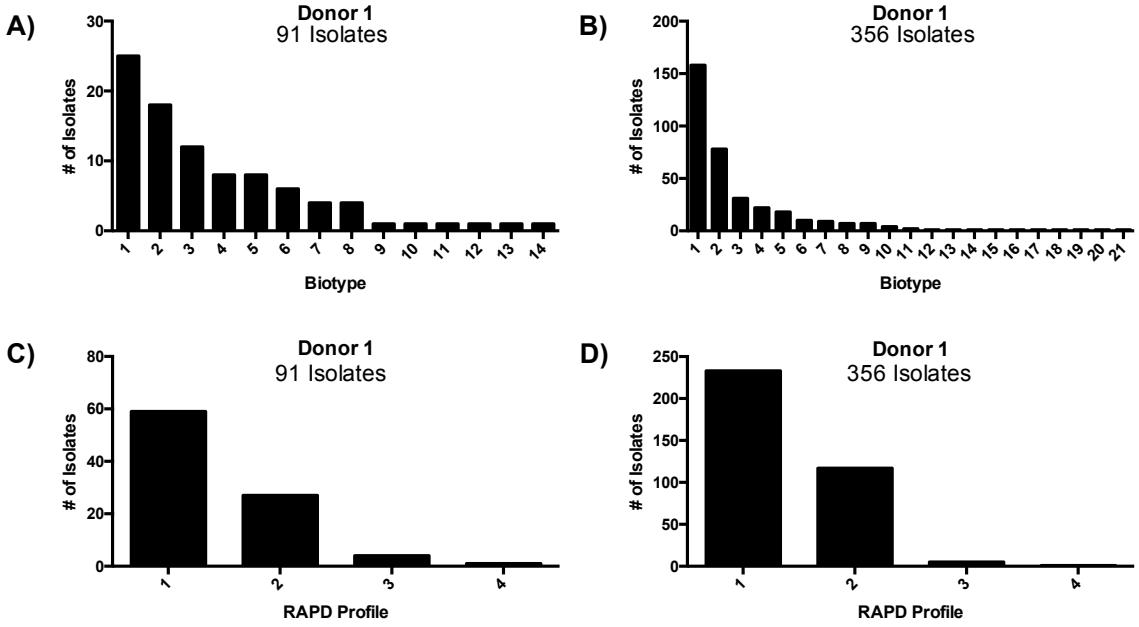
Acknowledgements

We would like to thank SP Bernier and ML Workentine for insightful discussion, and the Verdu lab for the kind gift of the HT-29 cells.

Table 1. Summary of *E. coli* phenotypes from four donors.

	Colony Morph	Hemolytic (%)	Swim (%)	Swarm (%)	Biofilm (%)	Auxotrophic (%)	Amp^R (%)	Cef^R (%)	$Cipro^R$ (%)	Glycogen (%)	RDAR (%)	# Biotypes
Donor 1 (91 isolates)	1	71	69	100	76	80	100	36	0	0	1	14
Donor 2 (94 isolates)	2	0	99	96	1	68	100	100	100	0	0	6
Donor 3 (88 isolates)	16	83	6	98	87	15	40	100	0	20	0	29
Donor 4 (78 isolates)	17	59	100	96	94	21	99	99	22	90	3	38

Number of isolates screened indicated under sample name. % indicates proportion of isolates positive for each phenotype.
Amp^R: ampicillin resistance, Cef^R: cefalexin resistance, $Cipro^R$: ciprofloxacin resistance, RDAR: red dry and rough morphotype



Supplementary Figure 1. Distribution of *E. coli* biotypes (A & B) and RAPD PCR profiles (C & D) from a healthy donor fecal sample where 91 isolates were screened (A & C) or 356 isolates were screened (B & D).

Chapter 5

Culture-enriched metagenomics in an Irritable Bowel Syndrome human-mouse microbiota transfer model

Preface

Research presented as part of this chapter has been prepared for submission.

Lau JT, Lu J, Pigrau M, Rossi L, Shah M, Verdu EF, Bercik P and Surette MG. Culture-enriched metagenomics in an Irritable Bowel Syndrome human-mouse microbiota transfer model. 2017.

Contributions:

JTL, LR and MS performed the culturing, DNA extractions and sample preparations for sequencing. JL and MP performed the mouse experiments and measurement of GI motility and behaviours. JTL and MGS conducted all sequencing analyses. JTL, EFV, PB and MGS designed the experiments. JTL, PB and MGS wrote the manuscript.

Abstract

Irritable bowel syndrome (IBS) is a prevalent functional gastrointestinal (GI) disorder characterized by abdominal pain and altered bowel movements, and is often accompanied by comorbid anxiety and depression. Pathophysiology of IBS likely involves the brain-gut axis, and studies have shown that there are differences in the gut microbiome between patients and healthy controls. Additionally, microbiota transfer studies have demonstrated that IBS-like symptoms can be transferred to germ-free animals via colonization with human fecal samples, suggesting the microbiota can initiate symptoms. In this study, three groups of germ-free mice were colonized with IBS microbiota, and the microbiota of the mice with the most IBS-like phenotypes were compared to the mice with the most health-like phenotypes within each group, to identify bacteria that may play a role in the generation of GI and behaviour phenotypes. The microbiota was profiled by 16S rRNA gene sequencing, and in addition, the most phenotypically divergent animals were sequenced by shotgun metagenomics and culture-enriched metagenomics. In culture-enriched metagenomics, cecal samples were cultured and plates capturing the majority of the bacterial diversity were subjected to metagenomic sequencing. We hypothesized that the combination of these approaches would allow for the association of bacteria to IBS symptoms, and improve metagenomic sequencing assemblies for identification of bacterial pathways involved. Although many bacterial taxa and functional pathways were found to be different between high and low IBS-like mice, bacterial groups were not consistently associated across experiments. Culture-enriched metagenomics did result in longer average contig assemblies than metagenomics of the mouse samples alone. The

culture-enriched metagenomics method described here can be applied to future studies to provide functional insight into the bacterial groups involved in the development of IBS symptoms.

Introduction

Irritable bowel syndrome (IBS) is a chronic functional bowel disorder, involving the brain-gut axis, with a high prevalence both world-wide and in Canada[234,235]. IBS is defined by Rome IV criteria as recurrent abdominal pain that is associated with defecation, altered frequency or consistency of stool, with symptoms persisting for at least one day a week in the previous three months, for a duration of at least six months[234]. Patients are grouped into four subtypes, IBS with diarrhea (IBS-D), with constipation (IBS-C), mixed subtype and unsubtyped. In addition to gastrointestinal (GI) symptoms, patients also report high rates of psychiatric comorbidities, including anxiety and depression[236]. The cause of IBS is likely to be multifactorial as stress, prior GI infection and genetics[76,234,237] have been implicated in the onset of the disorder, and a significant body of work has now provided evidence for the involvement of the gut microbiota in IBS.

Several studies have compared the gut microbiome of IBS patients to healthy controls, and although most studies report alterations in the abundance of bacterial groups, results are sometimes contradictory, and no taxonomic groups have been convincingly associated with IBS. A recent review of IBS microbiome studies found that an increase in the Firmicutes to Bacteroidetes ratio, decreased lactobacilli, bifidobacteria

and alpha diversity, and increased strepotococci and *Ruminococcus* has been reported by more than one study[64]. A meta-analysis by Liu *et al.* found that combining the results of 13 studies showed a significant decrease in *Lactobacillus*, *Bifidobacterium* and *Faecalibacterium prausnitzii* in IBS compared to controls[71]. Studies have also found associations of specific bacterial groups with IBS symptoms. One study reported that Gamma-proteobacteria was positively correlated with IBS symptoms, and several Clostridia species were associated with flatulence and distension, while *Bifidobacterium* was negatively associated with pain[72]. Using machine learning on a cohort of 110 IBS patient and 39 controls, 90 operational taxonomic units (OTUs) were associated with severe IBS symptoms but no specific taxonomic groups were identified[74]. However, another study found that IBS symptom severity was positively associated with *Ruminococcus torques*[238]. Applying supervised learning to a pediatric cohort found that *Alistipes* was correlated with greater pain frequency[182]. In IBS-C patients, an altered microbiome profile with decreased hydrogen consumers, including methanogens, was associated with bloating and abdominal pain[239]. Jeffrey *et al.* reported microbiota signatures for both GI symptoms and psychological symptoms in IBS, with higher levels of Cyanobacteria and Proteobacteria associated with increased bloating, increased mental component and pain, while Actinomycetales was inversely associated with depression[240]. From these studies, it is apparent that IBS symptoms are not consistently associated with any particular bacterial groups. Extensive heterogeneity in the human gut microbiome makes cross-sectional comparisons between patients and controls difficult, and limits the ability to detect significant correlations between phenotypes and the microbiota[3]. Most

studies include limited sample time points so instability in the microbiome is a further confounding factor, which may be particularly important in IBS as symptoms are recurrent.

A limited number of animal studies that have transferred fecal samples from IBS patients to germ-free mice demonstrate that the gut microbiota is capable of initiating IBS symptoms. A study by De Palma *et al.* successfully transferred both gut dysfunction and co-morbid anxiety from IBS fecal samples to germ-free (GF) mice[77]. Mice receiving IBS-D microbiota exhibited increased GI transit times, intestinal barrier dysfunction, and in a subset of mice there was also immune activation when compared to mice colonized with control microbiota. Most intriguingly, the transfer of anxiety-like behaviour only occurred in mice with immune activation. Mice receiving the same donor samples had similar microbiomes, and the individuality of donors was reflected in groups of mice. Bacterial groups that were associated with IBS were Lachnospiraceae and Bacteroidaceae. A separate study involving IBS-C patients demonstrated that fecal samples could transfer slower GI transit phenotypes to colonized animals[8]. The mice had reduced levels of short chain fatty acids (SCFAs), as a result of an altered gut microbiome with increased *Bacteroides uniformis, B. vulgatus,* and *Parabacteroides distasonis*. These microbiome transfer studies show that the gut microbiota can initiate several symptoms of IBS; however, it is unlikely that the entire community is required for the transfer of phenotype, but the specific bacterial groups and functional pathways involved have yet to be determined.

The goal of this study was to determine if specific bacterial taxa and functions are associated with IBS-like symptoms in mice receiving the microbiota of IBS patients. Colonization stochastically results in some microbiome heterogeneity within a cohort of mice receiving the same donor sample, and this is accompanied by heterogeneity of IBS-like phenotypes within a group[77]. We hypothesized that since the major bacterial groups within a cohort would be similar, it would be possible to identify the low abundant bacteria involved in the development of phenotypes since "microbiome noise", i.e. the inter-individual heterogeneity that makes microbiome associations difficult, would be reduced. For each human donor and mouse, 16S rRNA gene sequencing was performed, and within each of the 3 cohorts of animals, the microbiomes of the 3 mice with the highest IBS-like phenotypes and the 3 mice with the lowest IBS-like phenotypes were also characterized by culture-enrichment and shotgun metagenomics (Figure 1). Identification of the bacterial groups involved in the pathophysiology of IBS may lead to novel therapeutic targets for the treatment of this prevalent disorder.

Materials and methods

IBS microbiota colonization of germ-free mice

Three independent IBS microbiota transfer experiments were performed. Frozen IBS fecal samples were diluted anaerobically in sterile saline (1:10) and germ-free mice were colonized through intragastric gavage of 200µl of sample. Mice were housed for three weeks in vented racks, with 12h light/dark cycle and had ad libitum access to sterile food

and water. IBS-like phenotypes were assessed after three weeks of colonization. Experiments were approved by McMaster's Animal Research Ethics Board.

Experiment 1: To optimize the diversity of phenotypes after colonization, two stool samples from a single donor, taken 21 weeks apart and during high and low IBS symptom periods (Donor 1.1 and Donor 1.2 samples, respectively), were used to gavage 15 GF NIH Swiss mice (8 and 7 animals per sample, respectively). Based on the phenotype data however, this approach did not appear to increase the phenotypic heterogeneity between mice.

Experiment 2: To evaluate the impact of host genetics on the transfer of IBS patient microbiota and IBS phenotypes, Donor 1.2 sample was used to colonize 15 GF C57BL/6 mice.

Experiment 3: 15 GF NIH Swiss mice were colonized with IBS Donor 2 sample.

Measurement of IBS-like phenotypes in mice

Anxiety-like behaviour was assessed using a light/dark box preference test as previously described[97] using an automated system (Med Associates, St. Albans, Vermont). Each mouse was placed in the center of an illuminated box connected to a smaller dark box and movement was recorded for 10 minutes. Total time spent in the illuminated area, distance traveled and average velocity were measured. Decreased time in the light box is indicative of anxiety-like behaviour. A tail suspension test was also performed as a measure of depression phenotype, where longer time immobile is indicative of depression-like behaviour.

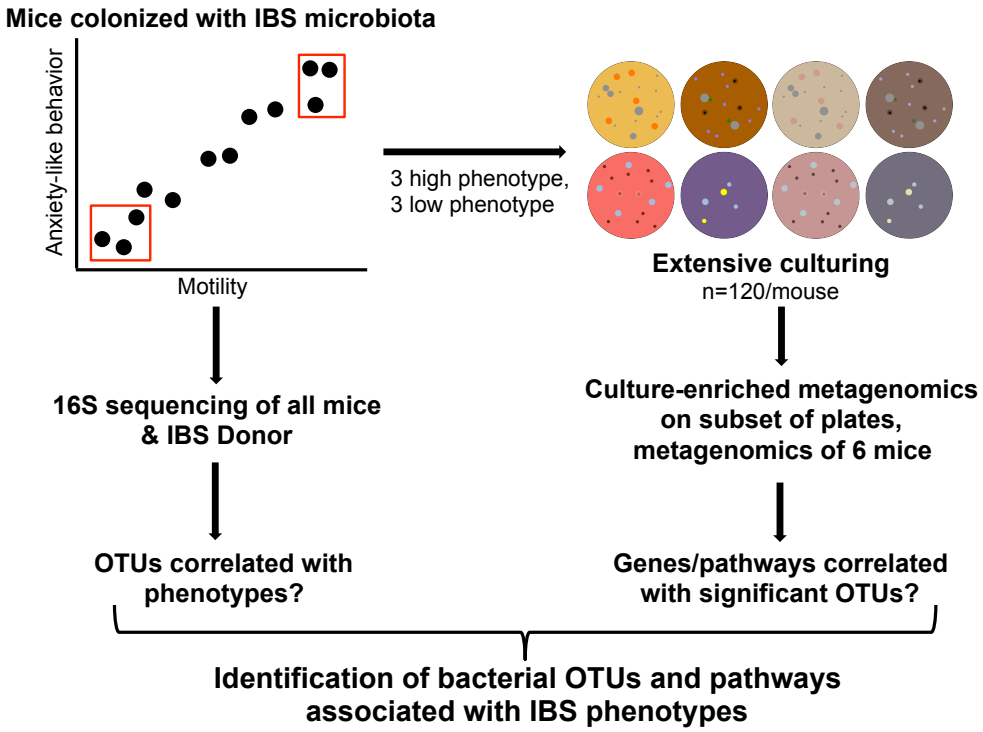


Figure 1. Overall experimental design. Outline of IBS microbiota transfer experiments to determine bacterial groups and pathways associated with IBS phenotypes.

Gastrointestinal motility was assessed by videofluoroscopy as described previously[241]. Five small steel beads (diameter=0.79mm, Bal-tec, California, USA) plus barium (0.1 mL, 40% dilution) were gavaged intragastrically to each mouse. A second gavage of barium (0.2 mL, 40% dilution) was performed 170 minutes later. Ten minutes later, the mouse was immobilized in a Plexiglass restrainer and fluoroscopic video recording (duration 10-20 seconds) was then obtained using digital videofluoroscope. Video images were digitized and analyzed by using NIH Image 1.62 software (http://rsb.info.nih.gov/nih-image). A simple scoring system was used to measure transit score. An individual bead was given a score that was dependent upon its location within

the GI tract (stomach 0, proximal small bowel 1, distal small bowel 2, cecum 3, colon 4, expelled 5). The scores of each bead within a single animal were then summed together to give a total transit score.

16S rRNA gene sequencing and analysis

Bacterial profiling was carried out by amplification of the V3 region of the 16S rRNA gene, sequencing was performed on the Illumina MiSeq and the data processed through an in-house bioinformatic pipeline as previously described[153,184]. The final sequence statistics after QC and filtering is summarized in Supplemental Table 1. Samples were rarified to 28 000 reads for beta-diversity analyses. To determine taxa associated with specific phenotypes, Linear Discriminant Analysis effect size (LEfSe)[242] was used. Spearman correlation of OTUs and phenotypes were analyzed with Graphpad Prism.

Shotgun metagenomic sequencing and analysis

DNA was extracted from human fecal samples, mouse cecal samples and plate pools as previously described[184] and DNA was sheared to 500 bp fragments by sonication. Shotgun metagenomic libraries were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs) according to manufacturer's instructions and samples were pair-end sequenced on the Illumina HiSeq at the McMaster Genome Facility. Reads were quality trimmed, and adaptors and primers removed using Trimmomatic (default parameters)[243]. Paired, trimmed reads were interleaved using custom scripts, and subsequently run through the HUMAnN2 (Human Microbiome

Project Unified Metabolic Analysis Network) pipeline for profiling microbial pathways[244]. Data was filtered for pathways with coverage of at least 0.3 and significant pathways were determined by LEfSe implemented in Galaxy, by stratifying the 6 mice samples by the 3 highest and 3 lowest animals for each phenotypic test.

Culture-enriched metagenomics

Bacterial culture of cecal microbiota was carried out on 6 mice from each experiment selected based on phenotype. Each sample was transferred to an anaerobic environment immediately after collection. Culturing was performed as previously described using 60 culture conditions[184]. For each cultured community, 16S rRNA gene sequencing was performed and, using a custom OTU coverage algorithm, the plates that captured all OTUs greater than 0.1% in mouse samples were identified and subsequently shotgun metagenomic sequenced. The taxonomic assignments of metagenomic samples was profiled by MetaPhlAn2[245]. Metagonomic reads were assembled using MEGAHIT[246] and assemblies were assessed by MetaQUAST[247].

Results

16S rRNA gene sequencing of IBS microbiota colonized mice

In this study, three sets of germ-free mice were colonized with fecal samples from IBS patients, and for each experiment, the light/dark box, GI motility and tail suspension tests were used to measure transfer of neurological and gastrointestinal IBS-like phenotypes. Each test showed a distribution of measurements for the animals within each

experiment, even though the mice were of identical genetic backgrounds and the same donor sample was used for colonization (Figure 2). It was also observed that the different IBS-like phenotypes were sometimes discordant for a mouse, so not all mice chosen for culturing and metagenomic sequencing were consistently high (or low) for all three tests.

16S rRNA gene sequencing of the human donor samples and the recipient mice cecal samples from the three experiments showed that the mice grouped based on their donors (Figure 3A). Animals from Experiment 1 and 2 had the same donor so more overlap in the microbiomes of these groups was observed, and mice colonized with IBS2 samples clustered separately. This was especially apparent on the PC1 vs PC3 plot, where the mouse samples from Experiment 1 and 2 were also separated. As previously reported, the human donor fecal samples were separated from the mice in in the PCoA plot, but the two samples from IBS1 were more similar than the IBS2 sample. The unweighted pair group method arithmetic mean (UPGMA) tree further showed that the mice colonized with the two IBS1 samples formed distinct groups (Figure 3B). Despite being similar in composition, it is interesting that the subtle differences in the two samples resulted in separation of the colonized mice. This was also of note because Experiment 1 and 2 mice were different strains, indicating that the effect of the donor had a stronger impact on the microbiome than the genetic background. Within each cohort, genus-level taxa summaries show the microbiome variability between the mice (Figure 3C). For example, *Akkermansia* was not visibly apparent in 5 mice from Experiment 1, but was in the other 10. As observed in the PCoA, the two samples from IBS1 had very similar compositions. Additionally, the mice in Experiment 1 and 2 appeared more similar to the donor sample

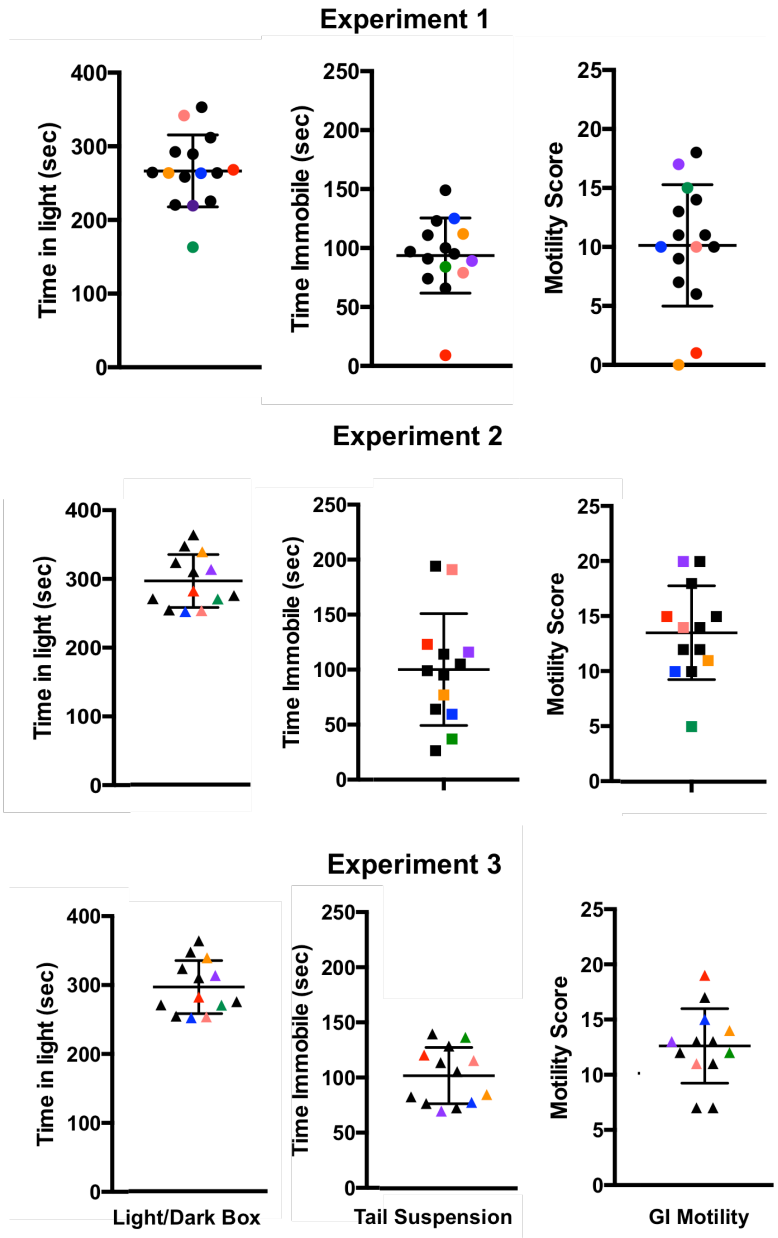


Figure 2. Mice exhibit phenotypic variability after IBS microbiota transfer. Results of light/dark box test, GI motility assessment and tail suspension test for each experiment show distribution for each phenotype. Coloured symbols represents mice with cecal contents cultured and shotgun metagenomic sequenced; each colour represents the same mouse across the 3 phenotypes.

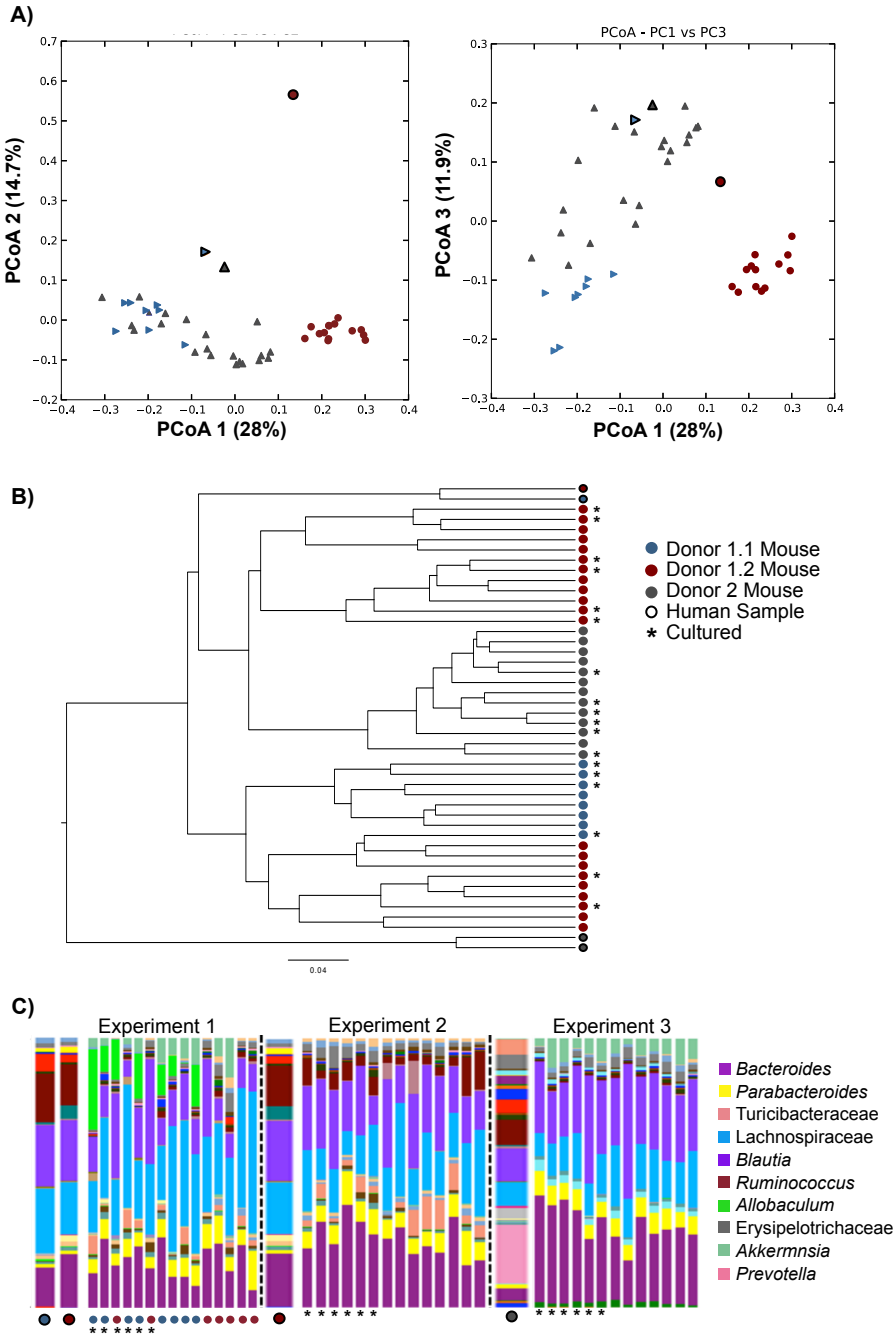


Figure 3. IBS mouse microbiomes group based on donor. Results of 16S rRNA gene sequencing of IBS donor samples and mouse cecal contents from 3 microbiota transfer experiments. A) Bray-curtis (BC) PCoA plot based on samples rarified to 28 000 reads, B) Unweighted pair group method with arithmetic mean (UPGMA) clustering of BC distances C) Genus-level taxonomic summaries. Colours indicate human IBS donor, * indicates mice that were cultured and metagenomic sequenced.

compared to the mice in Experiment 3. However, it was obvious that some bacterial groups were expanded in the mice and lower abundance in the donors, including *Allobaculum* and *Akkermansia* in Experiment 1, and decreased *Ruminococcus* in Experiment 1 and 3.

As mentioned, the two groups of mice in Experiment 1 and 2 that were both colonized with Donor 1.2 clustered separately by Bray-Curtis beta-diversity, with the potential implication that OTUs from the donor sample may be differentially transferred to mice of different genetic backgrounds. 18 OTUs were transferred to only C57BL/6 or NIH Swiss mice, 10 and 8 OTU, respectively (Figure 4). Differentially transferred OTUs were defined as OTUs that had <10 reads in all mice of one background, and >10 reads in at least 2 mice in the other background. The most abundant OTUs detected in only the NIH Swiss mice was *Akkermansia* OTU11, and the most abundant OTU only in the C57BL/6 mice was Lachnospiraceae OTU 40. It should be noted that these OTUs were all low abundance in the Donor 1.2 samples used to colonize both strains of mice, with the highest abundance of 0.15% in the 16S rRNA gene sequencing of the donor sample, and 10 of the differentially transferred OTUs were not detected at all.

For each experiment, OTUs that were significantly different between mice with the highest and lowest phenotype for each test was determined by LEfSe, and from these OTUs, those that had significant correlations with IBS-like phenotypes are listed in Table 2. All correlated OTUs were part of the Firmicutes or Bacteroidetes phyla. For GI motility, 3 OTUs were positively correlated with faster GI transit times, consisting of 1 *Bacteroides* & 2 *Clostridium* OTUs. Interestingly, *Clostridium* OTU51 was positively

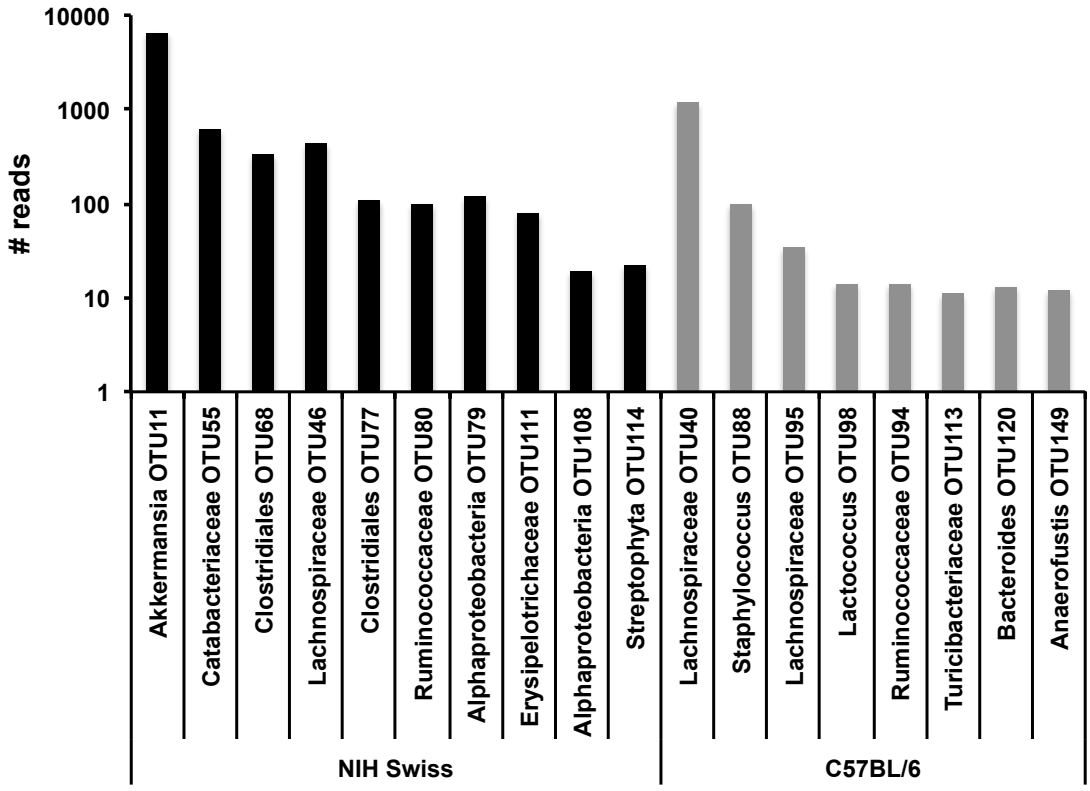


Figure 4. Several OTUs were only found in one genetic background in mice colonized with the same donor. OTUs that were found either in NIH Swiss or C57BL/6 mice colonized with Donor 1.2 sample. OTUs were included only if they were <10 reads in all mice of one background, and >10 reads in at least 2 mice in the other background.

correlated with faster GI transit times in both Experiment 1 and 2, but this OTU was not present in the donor used for Experiment 3, and no OTUs were correlated with GI motility in this set of animals. In the tail suspension tests, each experiment had one significantly correlated OTU but they were all different from each other. The results of the light/dark box test had the most correlated OTUs, with some negatively correlated with time spent in the light and others positively correlated, where decreased time in the light is interpreted as anxiety-like behavior. Erysipelotrichaceae OTU86 positively correlated with increased time in the light in Experiment 3 but negatively correlated with

Experiment 2. However, due to a lack of taxonomic resolution in 16S rRNA gene sequencing, more than 1 species could be in this OTU, which may explain this discrepancy between experiments, especially as Experiment 2 and 3 animals were colonized with different IBS donors.

Shotgun metagenomic sequencing of IBS microbiota colonized mice

To overcome the taxonomic limitations and lack of functional information from 16S rRNA gene sequencing, shotgun metagenomics was performed on the three mice within each experiment exhibiting the most and least IBS-like phenotypes. Mouse metagenomic sequencing was analyzed by the HUMAnN2 (Human Microbiome Project Unified Metabolic Analysis Network) pipeline[244] to profile microbial pathways in each sample. The animals in Experiment 2 showed the most consistent IBS-like behaviour across all 3 phenotypes measured i.e. mice with fast GI transit times also displayed the highest anxiety-like and depression-like behaviours. Within this cohort, 7 pathways were significantly higher in the least IBS-like group, and all the pathways belonged to members of the Bacteroidetes, specifically, *Alistipes shahii* & *Bacteroides faecis* (Figure 5). 16 pathways were at a higher abundance in the most IBS-like group, and pathways were from *Parabacteroides merdae, Clostridium asparagiforme,* unclassified Clostridiales*,* and *Eggerthella lenta*. All significant pathways with coverage greater than 30% for each phenotype in the three transfer experiments are reported in Tables 3-5, and dozens of pathways belonging to several groups of phylogenetically diverse taxa were significantly associated with each phenotype. 23 pathways were associated with the tail

suspension test, 31 with results of the light/dark box test, and 36 with rates of GI motility (Table 3-5).

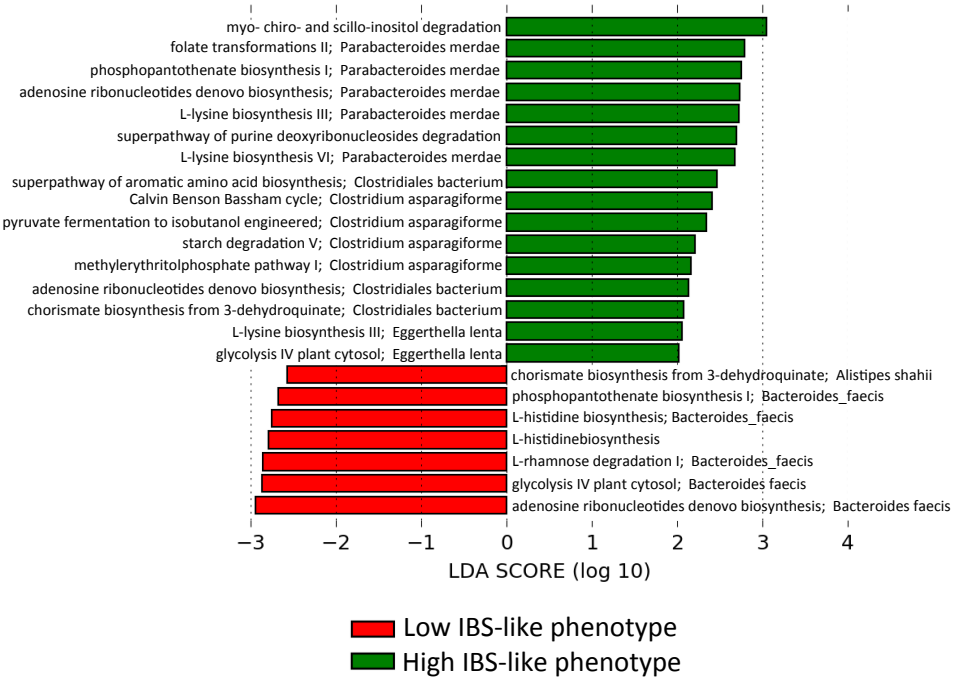


Figure 5. Several pathways were significantly associated with the most IBS-like phenotype and least IBS-like phenotype groups from Experiment 2. Pathways were assigned by Humann2 analysis of metagenomic sequencing of mouse cecal samples, and significance determined by LEfSe.

Additionally, several significant pathways were represented in multiple bacteria due to the high functional redundancy in gut microbiome[3]. For example, 2 pathways of L-lysine biosynthesis (PWY 2942 & PWY 5097) were significantly associated with GI motility, and the pathways were assigned to *Eggerthella lenta, Parabacteroides merdae, P. distasonis, Akkermansia muciniphila, Paraprevotella clara*, and an unclassified taxa.

Some significant pathways were assigned to taxa that had also been determined to be associated with IBS-like phenotypes in 16S rRNA gene profiling, demonstrating correlations between 16S rRNA and metagenomics data. For example, Experiment 1 pathways associated with GI motility were found in *Bacteroides* and *Clostridium*, and OTUs of these genera were also associated with motility in the 16S rRNA gene analysis (Table 2). Specifically, 10 *Clostridium* pathways were positively associated with a faster GI motility phenotype and *Clostridium* OTU51 also correlated with motility in Experiment 1 and 2, but due to short length of V3 16S sequencing, it is currently difficult to confidently match *Clostridium* sp. OTU51 to a metagenomic taxa.

Culture-enriched metagenomic sequencing of IBS microbiota colonized mice

Although correlations were observed between bacterial groups and IBS-like phenotypes from 16S and metagenomic sequencing, culture-enriched metagenomic sequencing should provide even greater information about the functionality of the community. Culture-enriched metagenomics would produce greater depth of coverage since cultured communities are less complex than culture-independent sequencing of mouse samples, and may allow for better assemblies and more complete genomic information. Six mouse samples from each experiment were cultured using 30 media, and plates were incubated anaerobically and aerobically, and cultured communities from each media were 16S sequenced. By comparison of OTUs from the mouse samples to the cultured OTUs, we found that the cultured OTUs accounted for greater than 98% of the total abundance of the cecal microbiome in all three experiments (Figure 6). Additionally,

an average of 81-89% of the total OTUs in each group of mice were represented in culture. Since most OTUs and the majority of the community abundance was cultured, it was feasible to apply culture-enriched metagenomics to characterize the gut microbiome of the IBS colonized mice. Shotgun metagenomics of the cultured plate pools should represent the majority of taxa detected in the culture-independent sequencing of the mouse cecal samples in each experiment. The subset of plates for each experiment that was required to capture all culturable OTUs present at greater than 0.1% relative abundance in the mouse samples was determined, and these plates were shotgun metagenomic sequenced. To cover the bacterial diversity represented in the mouse samples required metagenomic sequencing of 6, 7, and 9 plates for Experiment 1, 2, and 3, respectively.

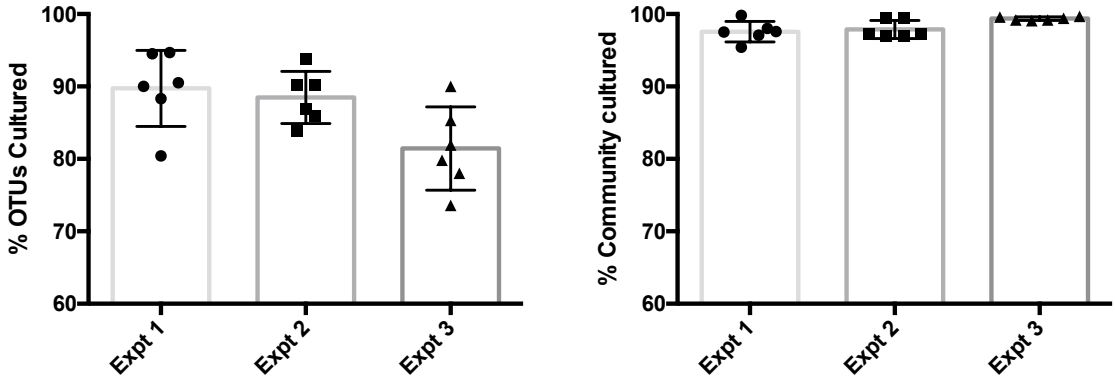


Figure 6. The majority of the mouse cecal community was cultured. Percentage of total OTUs and the total community abundance (based on abundance of each OTU) cultured from 6 mice in each IBS microbiota transfer experiment, based on 16S rRNA gene sequencing.

From the results of culture-enriched metagenomics, several phylogenetically diverse taxa, including Actinobacteria, Firmicutes, Bacteroidetes and Proteobacteria,

made up less than 1% of the culture-independent metagenomic sequencing of the 6 mice in each experiment but were greater than 1% of a culture-enriched community. This was the case for nine taxa in Experiment 1, seven taxa in Experiment 2 and 16 taxa in Experiment 3 (Table 6). *Flavonifractor plautii, Clostridium ramosum*, unclassified *Eggerthella*, and *Escherichia coli* were found in higher abundance by culture, in more than one experiment. Aside from increasing the relative abundance of several taxa, we also wanted to determine if culture-enrichment resulted in better metagenomic assemblies than using reads from mouse samples alone. Shotgun metagenomic sequences were assembled using Megahit[246], which provides high accuracy and fast assemblies, and the quality of assemblies was assessed using MetaQUAST[247]. The mouse-only assemblies resulted in a greater number of contigs of at least 500bp in length than the assemblies that included the culture-enriched reads (Figure 7). The largest contig size was similar between assemblies in Experiment 1, higher in the mouse-only assemblies in Experiment 2, and in Experiment 3, the largest contig in the combined mouse and culture-enriched assembly was 119 660 bp longer than largest contig in the mouse-only assembly. N50 is the minimum contig length that contains 50% of the bases in an assembly (i.e. the minimum contig length to cover 50% of the genome), and is commonly used to assess the quality of assemblies. The N50 length was always higher in the assemblies containing the culture-enriched reads, indicating that these assemblies had longer contigs on average than the mouse-only assemblies.

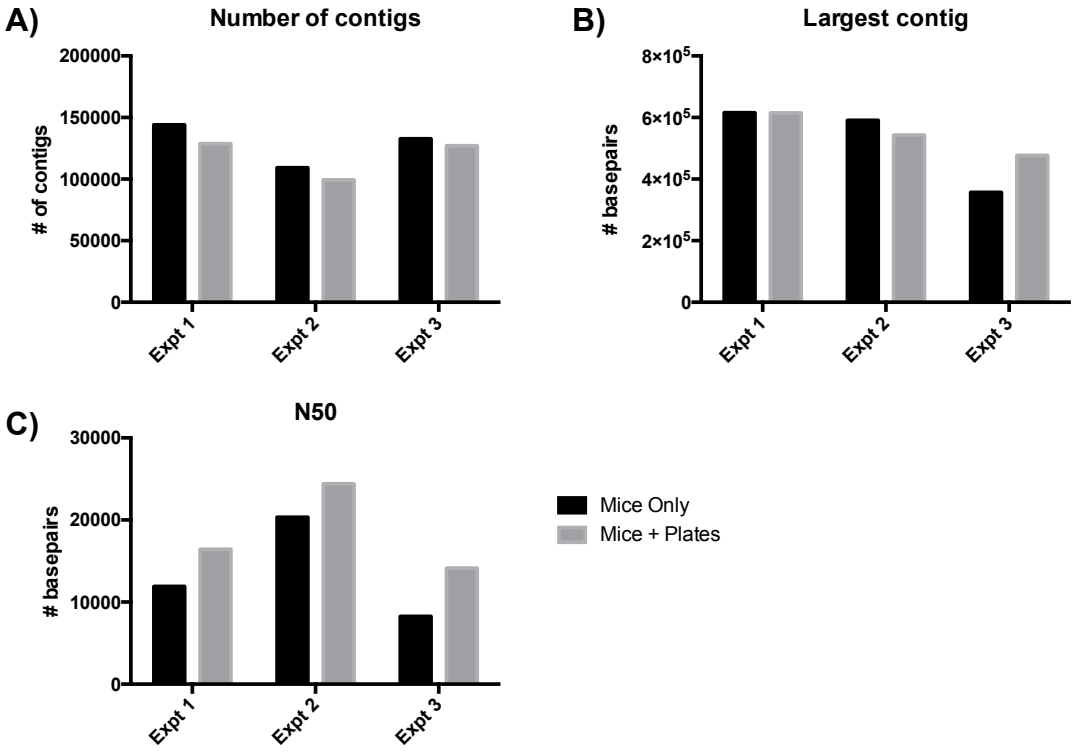


Figure 7. Assemblies from culture-enrichment resulted in fewer contigs but higher N50s compared to assemblies from mouse samples alone. MetaQUAST statistics of Megahit metagenomic assemblies for mouse samples only, or mouse samples and cultured plates. A) Total number of contigs greater than 500bp B) Largest assembled contig C) N50 length.

Discussion

A combination of 16S rRNA gene sequencing, metagenomics, and bacterial culture methods was applied to an IBS microbiota transfer model to determine if bacterial groups and their pathways could be associated with the IBS-like phenotypes. We rationalized that comparison of mice across a group, with the most and least IBS-like phenotypes, would result in more robust microbiome associations since the phenotypes should be driven by the donor microbiome, and the similar microbiome compositions

within a group would allow for the correlation of low abundant taxa to IBS behaviours. However, a significant caveat of this study was that IBS-like phenotypes may not be stable, and the distribution of behaviours measured could change overtime e.g. a mouse with high anxiety-like behaviour for one measurement may exhibit discordant behaviours if another test was used, or if the phenotype was measured longitudinally. This would be complicated to evaluate since repeated phenotype measurements can impact results due to induced stress and learned behaviours, but could be an explanation for why it was difficult to identify mice with consistently high IBS-like phenotypes in each group.

Interestingly, from the 16S rRNA gene sequencing, we observed OTUs that were potentially differentially transferred between two different genetic backgrounds of mice that were colonized with a single donor sample. Although these OTUs were low abundance in the 16S profile of the donor, they were detected in the mice, demonstrating the potential enrichment of OTUs during colonization of mice. We have previously reported the detection of OTUs by culture that were not seen in the culture-independent sequencing of fecal samples, potentially as a result of sequencing depth[184]. Colonization of germ-free mice can also be a potential selective environment for low abundant OTUs in fecal samples. An example of this was *Akkermansia* OTU11, which was only found in the NIH Swiss mice and was less than 10 reads in the Donor 1.2 sample. *Akkermansia*, which currently has one described species, *A. muciniphila*, is a mucosally-adherent bacterium[248], and this could by a possible explanation of why it was not detected in the fecal sample (where it was likely below the level of detection), but was found in the cecal contents of the microbiota humanized mice, since cecal contents contain adherent

organisms due to collection methods. These observations have implications for microbiota-transfer studies, as OTUs not detected in human fecal samples may be enriched in animals, and the genetic background of mice will impact the OTUs that are transferred. This may lead to studies where transfer of phenotypes from humans to animals does not occur because the bacteria groups of importance could not successfully colonize.

For each experiment, we were able to identify bacterial groups and pathways that were associated with the three phenotypes measured – GI transit rates, anxiety-like behaviour, and depression-like behaviour. *Bacteroides* and Firmicutes OTUs from the 16S rRNA gene sequencing were significantly correlated with IBS phenotypes but because of the lack of taxonomic resolution by this method, species-level resolution was not obtained and it is not known if an OTU contains more than one bacterial species. Shotgun metagenomics resulted in the identification of bacterial species and specific pathways that were significantly associated with IBS-like phenotypes in mice. A recent study looking at human stool consistency (which is correlated with GI transit times) and its association with the gut microbiota composition, showed that *Bacteroides* was increased in loose stools (associated with faster transit), and *Akkermansia* was associated with slower transit and harder stools[249]. Interestingly, from the metagenomic analysis, we observed that *Akkermansia* was increased in the mice with slower GI transit in Experiment 3 and all significant *Akkermansia* pathways in Experimentt 3 were associated with slow GI transit. Also similar to what has been previously reported, *Bacteroides* (and *Bacteroides* pathways) were associated with faster GI transit in Experiment 1, but in

Experiment 2 *Bacteroides faecis* pathways increased in slower GI transit. These discrepancies may be due to species differences as *Bacteroides* are a functionally diverse genera.

We also observed that several pathways associated with L-lysine biosynthesis were significant by biomarker discovery. L-lysine is a serotonin receptor 4 (5-HT4) antagonist, and was shown to block contractions of guinea pig ileum *in vitro* and inhibited diarrhea, as well as inhibiting serotonin-mediated anxiety in rats *in vivo*[250]. 10 pathways of L-lysine biosynthesis were associated with GI motility, and were assigned to *Eggerthella lenta, Parabacteroides merdae, P. distasonis, Akkermansia muciniphila, Paraprevotella clara*, and unclassified taxa. It was also interesting that several pathways for chorismate biosynthesis were associated with the light/dark box test since chorismate is a precursor for indole and indole derivatives, and the aromatic amino acids including tryptophan which is required for synthesis of the neurotransmitters serotonin and melatonin[251]. Pathways of chorismate biosynthesis were also associated with GI motility, which is intriguing as changes in levels of serotonin can influence motility of the GI tract[252].

Culture-enriched metagenomics of the IBS microbiota transfer model demonstrated the ability to culture taxa that were low abundance in mice or human samples, at a higher abundance than detected in culture-independent metagenomics. This shows the potential of culture-enriched metagenomics for accessing the functional potential of the gut microbiota. Contigs from potentially important strains could be assembled from the culture-enriched metagenomic reads, which would not be possible

from the culture-independent metagenomics. For example, several bacterial pathways associated with GI motility were from *Bacteroides* and *Clostridium*, and several species from this group made up less than 1% of the community in the mouse samples, but were found at a greater abundance in the cultured communities. Evaluation of assemblies from the two methods found that the culture-enriched assemblies had larger N50s than culture-independent assemblies, however, further analysis is required to determine the accuracy of assemblies. Communities with highly similar organisms, such as multiple strains of a species, may result in misassembled contigs. Contigs (or unassembled reads) could be binned to improve assemblies and strain-level differences could be analyzed; however, several tools are available for each of these processes, which require benchmarking to determine the best method. This is computationally intensive, but other studies have shown discrepancies in results including the number of strains identified and their taxonomic assignments depending on the bioinformatics tools chosen[253].

From this study, several bacterial groups and functional pathways that may play a role in the transfer of IBS-like symptoms were identified. However, due to the potential lack of robustness of phenotypes, the lack of consistent results between experiments, and the small number of IBS donors used, these potential bacterial signatures require further studies to confirm if they are important to the pathophysiology of IBS. We describe a novel approach of combining 16S rRNA gene sequencing, shotgun metagenomic sequencing and bacterial culture that could be applied to future studies to provide insight into the involvement of specific members the gut microbiome in the development of IBS.

Acknowledgements

We would like to thank FJ Whelan for helpful discussion regarding metagenomic analysis and Nestle Research Center for funding.

Table 1. Summary of patient samples and mouse experiments

Expt	Donor	IBS Type	Birmingham Score	HAD-A Score	β-defensin	Mouse Strain	# of mice
1	1.1	IBS-D	14	2	100	NIH Swiss	7
	1.2	IBS-D	5	0	9	NIH Swiss	8
2	1.2	IBS-D	5	0	9	C57Bl/6	15
3	2	IBS-mix	25	12	74	NIH Swiss	15

HAD-A, Hospital anxiety and depression Scale for anxiety

Table 2. OTUs from 16S rRNA gene sequencing significantly correlated with each IBS-like phenotype. Comparisons between mice with high and low phenotype was completed for each experiment independently.

GI motility

OTU	Expt	Spearman rho	p-value
168 *Bacteroides*	1	0.682	0.005
417 *Clostridium*	1	0.723	0.002
51 *Clostridium*	1	0.596	0.02
	2	0.591	0.03

Tail Suspension Test

OTU	Expt	Spearman rho	p-value
127 Lachnospiraceae	1	-0.636	0.01
771 *Clostridium*	2	0.592	0.03
133 *Bacteroides*	3	0.577	0.04

Light/Dark Box Test

OTU	Expt	Spearman rho	p-value
249 Rikenellaceae	1	-0.604	0.02
85 Clostridiales	2	-0.582	0.03
86 Erysipelotrichaceae	2	-0.555	0.04
	3	0.748	0.003
37 *Clostridium*	2	0.662	0.01
312 *Parabacteroides*	2	-0.762	0.002
133 *Bacteroides*	3	-0.802	0.001
154 Ruminococcaceae	3	-0.577	0.004

Table 3. Significant metagenomic pathways for Light/Dark box

Pathway [1]	Expt	Avg rel. abundance in High Light time	Avg rel. abundance in Low Light time	Pathway coverage
ARO_PWY_chorismate biosynthesis I.g__Parasutterella_s__Parasutterella_excrementihominis	1	**8.9033E-06**	5.8967E-06	0.60
ARO_PWY_chorismate biosynthesis I.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	1.15E-06	**3.55E-06**	0.30
ARO_PWY_chorismate biosynthesis I.g__Clostridium_s__Clostridium_scindens	3	0	**1.4333E-06**	0.64
ARO_PWY_chorismate biosynthesis I.g__Lachnospiraceae_noname_s__Lachnospiraceae_bacterium_5_1_57FAA	3	3.4667E-07	**1.4367E-06**	0.62
COA_PWY_1_coenzyme A biosynthesis II_mammalian_.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47 FAA	1	5.000E-07	**2.877E-06**	0.36
COA_PWY_1_coenzymeA biosynthesis II_mammalian_.g__Akkermansia_s__Akkermansia_muciniphila	3	1.208E-04	**2.152E-04**	0.95
PANTO_PWY_phosphopantothenate biosynthesis I.g__Bacteroides_s__Bacteroides_faecis	1	1.145E-05	**2.533E-05**	0.75
PANTO_PWY_phosphopantothenate biosynthesis I.g__Bacteroides_s__Bacteroides_faecis	2	**2.350E-05**	9.880E-06	0.65
PANTO_PWY_phosphopantothenate biosynthesis I.g__Parabacteroides_s__Parabacteroides_merdae	2	1.6313E-05	**3.7100E-05**	0.59
PANTO_PWY_phosphopantothenate biosynthesis I.g__Akkermansia_s__Akkermansia_muciniphila	3	1.091E-04	**1.905E-04**	0.55
NONMEVIPP_PWY_methylerythritol phosphate pathway I.g__Clostridium_s__Clostridium_asparagiforme	2	9.4167E-07	**3.6033E-06**	0.45
NONMEVIPP_PWY_methylerythritol phosphate pathway I.g__Akkermansia_s__Akkermansia_muciniphila	3	9.969E-05	**1.780E-04**	0.50
NONMEVIPP_PWY_methylerythritol phosphate pathway I.g__Bacteroides_s__Bacteroides_xylanisolvens	3	**3.5033E-05**	1.6867E-05	0.44
NONMEVIPP_PWY_methylerythritolphosphate pathway I.g__Erysipelotrichaceae_noname_s__Clostridium_ramosum	3	5.473E-06	**9.910E-06**	0.41
PWY_1042_glycolysis IV_plant cytosol_.g__Bacteroides_s__Bacteroides_faecis	1	1.197E-05	**3.147E-05**	0.52
PWY_1042_glycolysis IV_plant cytosol_.g__Bacteroides_s__Bacteroides_faecis	2	**3.1867E-05**	1.0517E-05	0.52
PWY_1042_glycolysis IV_plant cytosol_.g__Eggerthella_s__Eggerthella_lenta	2	0	**2.4567E-06**	0.60
PWY_2942_L_lysine biosynthesis III.g__Bacteroides_s__Bacteroides_faecis	1	1.0560E-05	**2.9767E-05**	0.59
PWY_2942_L_lysine biosynthesis III.g__Eggerthella_s__Eggerthella_lenta	2	0	**2.2967E-06**	0.47
PWY_2942_L_lysine biosynthesis III.g__Parabacteroides_s__Parabacteroides_merdae	2	2.170E-05	**4.4333E-05**	0.77
PWY_2942_L_lysine biosynthesis III.g__Akkermansia_s__Akkermansia_muciniphila	3	1.113E-04	**1.996E-04**	0.37
PWY_2942_L_lysine biosynthesis III.g__Parabacteroides_s__Parabacteroides_distasonis	3	**5.250E-05**	3.443E-05	0.84
PWY_2942_L_lysine biosynthesis III.g__Paraprevotella_s__Paraprevotella_clara	3	**1.8833E-05**	1.0643E-05	0.52
PWY_2942_L_lysine biosynthesis III.g__Paraprevotella_s__Paraprevotella_xylaniphila	3	**1.4373E-06**	6.0667E-07	0.34
PWY_2942_L_lysine biosynthesis III.unclassified	3	**7.7567E-05**	5.5400E-05	0.99
PWY_5097_L_lysine biosynthesis VI.g__Bacteroides_s__Bacteroides_faecis	1	9.940E-06	**2.743E-05**	0.54
PWY_5097_L_lysine biosynthesis VI.g__Parabacteroides_s__Parabacteroides_merdae	2	2.0867E-05	**4.1167E-05**	0.71

PWY_5097_L_lysine biosynthesis VI.g__Parabacteroides_s__Parabacteroides_distasonis	3	**4.9067E-05**	3.2167E-05	0.82
PWY_5097_L_lysine biosynthesis VI.g__Paraprevotella_s__Paraprevotella_clara	3	**1.9267E-05**	1.1187E-05	0.55
PWY_5097_L_lysine biosynthesis VI.unclassified	3	**8.4815E-05**	6.0300E-05	1.00
PWY_3841_folate transformations II.g__Parabacteroides_s__Parabacteroides_merdae	2	1.7433E-05	**4.1567E-05**	0.65
PWY_3841_folate transformations II.g__Akkermansia_s__Akkermansia_muciniphila	3	9.7464E-05	**1.7145E-04**	0.50
PWY_3841_folate transformations II.g__Bacteroides_s__Bacteroides_ovatus	3	**1.166E-04**	6.117E-05	0.77
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Parasutterella_s__Parasutterella_excrementihominis	1	**8.8567E-06**	5.6667E-06	0.60
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Alistipes_s__Alistipes_shahii	2	**0.00000255**	0	0.58
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	2	0	**2.7167E-06**	0.39
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	2.137E-06	**3.730E-06**	0.54
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Clostridium_s__Clostridium_scindens	3	0	**1.3467E-06**	0.59
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.unclassified	3	**8.528E-05**	6.0033E-05	1.00
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Bacteroides_s__Bacteroides_faecis	1	1.0013E-05	**3.1533E-05**	0.51
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Bacteroides_s__Bacteroides_faecis	2	**3.510E-05**	1.0217E-05	0.84
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	1	2.520E-06	**4.440E-06**	0.61
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	2	8.4333E-07	**4.9467E-06**	0.61
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	1.7767E-06	**5.8033E-06**	0.63
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Parabacteroides_s__Parabacteroides_merdae	2	2.290E-05	**4.640E-05**	0.80
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Parabacteroides_s__Parabacteroides_merdae	3	**3.13E-05**	2.24E-05	0.53
PWY_7219_adenosine ribonucleotides denovo biosynthesis	3	**1.052E-03**	1.016E-03	1.00
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Akkermansia_s__Akkermansia_muciniphila	3	1.280E-04	**2.218E-04**	0.53
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Alistipes_s__Alistipes_onderdonkii	3	0	**2.6837E-06**	0.59
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Erysipelotrichaceae_noname_s__Eubacterium_biforme	3	**1.7633E-05**	9.1433E-06	0.59
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Parabacteroides_s__Parabacteroides_distasonis	3	**5.3567E-05**	3.4433E-05	0.76
PWY_7229_superpathway of adenosine nucleotides denovo biosynthesis I.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	1	1.9833E-06	**4.070E-06**	0.62
PWY_7229_superpathway of adenosine nucleotides denovo biosynthesis I.unclassified	3	**8.7946E-05**	6.8133E-05	0.98
RHAMCAT_PWY_L-rhamnose degradation I.g__Bacteroides_s__Bacteroides_faecis	1	1.3033E-05	**3.0367E-05**	0.85
RHAMCAT_PWY_L-rhamnose degradation I.g__Bacteroides_s__Bacteroides_faecis	2	**3.1033E-05**	9.7667E-06	0.91
PWY_6737_starchdegradationV.g__Clostridium_s__Clostridium_asparagiforme	2	4.6233E-06	**9.420E-06**	0.98
PWY_6737_starch degradation V.g__Akkermansia_s__Akkermansia_muciniphila	3	9.9275E-05	**1.789E-04**	0.55

PWY_6737_starch degradation V.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	3.4333E-07	**2.5233E-06**	0.48
PWY_6737_starch degradation V.g__Lachnospiraceae_noname_s__Lachnospiraceae_bacterium_3_1_57FAA_CT1	3	0	**9.2233E-07**	0.81
PWY_7111_pyruvate fermentation to isobutanol_engineered_.g__Clostridium_s__Clostridium_asparagiforme	2	2.330E-06	**6.9133E-06**	0.78
PWY_7111_pyruvate fermentation to isobutanol_engineered_.g__Clostridium_s__Clostridium_bolteae	3	6.0167E-06	**1.0753E-05**	0.53
CALVIN_PWY_Calvin_Benson_Bassham cycle.g__Clostridium_s__Clostridium_asparagiforme	2	1.806E-06	**8.4667E-06**	0.87
COMPLETE_ARO_PWY_superpathway of aromatic aminoacid biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	2	0	**2.7467E-06**	0.31
HISTSYN_PWY_L-histidine biosynthesis.g__Bacteroides_s__Bacteroides_faecis	2	**2.340E-05**	6.630E-06	0.30
PWY_7237_myo- chiro- and scillo-inositol degradation.unclassified	2	3.0867E-05	**7.216E-05**	0.64
BRANCHED_CHAIN_AA_SYN_PWY_super pathway of branched aminoacid biosynthesis.unclassified	3	**8.0472E-05**	5.7467E-05	0.88
COA_PWY_coenzymeA biosynthesis I.g__Akkermansia_s__Akkermansia_muciniphila	3	1.188E-04	**2.050E-04**	0.76
COA_PWY_coenzymeA biosynthesis I.g__Bacteroides_s__Bacteroides_ovatus	3	**1.138E-04**	6.1367E-05	0.54
DTDPRHAMSYN_PWY_dTDP_L-rhamnose biosynthesis I.g__Bacteroides_s__Bacteroides_ovatus	3	**1.843E-04**	1.040E-04	1.00
ILEUSYN_PWY_L-isoleucine biosynthesis I_from threonine_.unclassified	3	**8.773E-05**	6.210E-05	0.90
NONOXIPENT_PWY_pentosephosphate pathway_non-oxidative branch_	3	3.975E-04	**4.486E-04**	1.00
PANTOSYN_PWY_pantothenate and coenzymeA biosynthesis I.g__Akkermansia_s__Akkermansia_muciniphila	3	1.170E-04	**2.048E-04**	0.66
PWY_5100_pyruvate fermentation to acetate and lactate II.g__Peptostreptococcaceae_noname_s__Clostridium_difficile	3	**1.9167E-06**	8.5133E-07	0.84
PWY_5686_UMP biosynthesis.g__Alistipes_s__Alistipes_onderdonkii	3	4.400E-07	**2.940E-06**	0.72
PWY_5695_urate biosynthesis_inosine 5-phosphate degradation.unclassified	3	**2.9933E-05**	2.3233E-05	0.75
PWY_6121_5_aminoimidazole ribonucleotide biosynthesis I	3	5.158E-04	**6.017E-04**	0.99
PWY_6122_5_aminoimidazole ribonucleotide biosynthesis II	3	4.700E-04	**5.569E-04**	0.98
PWY_6122_5_aminoimidazole ribonucleotide biosynthesis II.g__Holdemania_s__Holdemania_filiformis	3	0	**6.36E-07**	0.66
PWY_6122_5_aminoimidazole ribonucleotide biosynthesis II.unclassified	3	**7.865E-05**	6.120E-05	0.94
PWY_6277_superpathway of 5-aminoimidazole ribonucleotide biosynthesis	3	4.700E-04	**5.569E-04**	0.98
PWY_6277_superpathway of 5-aminoimidazole ribonucleotide biosynthesis.g__Holdemania_s__Holdemania_filiformis	3	0	**6.36E-07**	0.66
PWY_6277_superpathway of 5-aminoimidazole ribonucleotide biosynthesis.unclassified	3	**7.865E-05**	6.120E-05	0.94
PWY_7221_guanosine ribonucleotides denovo biosynthesis.g__Akkermansia_s__Akkermansia_muciniphila	3	1.173E-04	**2.050E-04**	0.87
PWY_7221_guanosine ribonucleotides denovo biosynthesis.g__Desulfovibrio_s__Desulfovibrio_piger	3	0	**0.00000219**	0.45
PWY0_1586_peptidoglycan maturation_meso-diaminopimelate containing_.g__Akkermansia_s__Akkermansia_muciniphila	3	1.497E-04	**2.475E-04**	0.98
TRNA_CHARGING_PWY_tRNA charging.unclassified	3	**7.6401E-05**	5.9267E-05	1.00
VALSYN_PWY_L-valine biosynthesis.unclassified	3	**8.773E-05**	6.210E-05	0.90

Table 4. Significant metagenomic pathways for Tail Suspension Test

Pathway [1]	Expt	Avg rel. abundance in high immobile time	Avg rel. abundance in low immobile time	Max pathway coverage
CALVIN_PWY_Calvin_Benson_Bassham cycle.g__Clostridium_s__Clostridium_asparagiforme	2	**8.4667E-06**	1.806E-06	0.87
CALVIN_PWY_Calvin_Benson_Bassham cycle	3	4.621E-04	**4.329E-04**	1.00
NONMEVIPP_PWY_methylerythritol phosphate pathway I.g__Clostridium_s__Clostridium_asparagiforme	2	**3.6033E-06**	9.4167E-07	0.45
NONMEVIPP_PWY_methylerythritol phosphate pathway I	3	7.320E-04	**7.339E-04**	1.00
NONMEVIPP_PWY_methylerythritol phosphate pathway I.g__Paraprevotella_s__Paraprevotella_clara	3	1.3297E-05	**2.5267E-05**	0.61
PANTO_PWY_phosphopantothenatebiosynthesis I.g__Parabacteroides_s__Parabacteroides_merdae	1	**1.370E-05**	1.9117E-06	0.59
PANTO_PWY_phosphopantothenatebiosynthesis I.g__Parabacteroides_s__Parabacteroides_merdae	2	**3.710E-05**	1.6313E-05	0.59
PANTO_PWY_phosphopantothenate biosynthesis I.g__Bacteroides_s__Bacteroides_faecis	2	9.880E-06	**2.350E-05**	0.65
PANTO_PWY_phosphopantothenate biosynthesis I	3	6.561E-04	**6.641E-04**	1.00
PANTO_PWY_phosphopantothenate biosynthesis I.g__Bacteroides_s__Bacteroides_xylanisolvens	3	1.3867E-05	**2.4033E-05**	0.50
PWY_1042_glycolysis IV_plant cytosol_.g__Bacteroides_s__Bacteroides_faecis	2	1.0517E-05	**3.1867E-05**	0.52
PWY_1042_glycolysis IV_plant cytosol_.g__Eggerthella_s__Eggerthella_lenta	2	**2.4567E-06**	0	0.60
PWY_1042_glycolysis IV_plant cytosol_.g__Desulfovibrio_s__Desulfovibrio_piger	3	1.8133E-06	**3.640E-06**	0.48
PWY_2942_L-lysine biosynthesis III.g__Eggerthella_s__Eggerthella_lenta	2	**2.2967E-06**	0	0.47
PWY_2942_L-lysine biosynthesisIII.g__Parabacteroides_s__Parabacteroides_merdae	2	**4.4333E-05**	2.170E-05	0.77
PWY_2942_L_lysinebiosynthesis III	3	7.9474E-04	**7.7630E-04**	1.00
PWY_2942_L-lysine biosynthesis III.g__Bacteroides_s__Bacteroides_uniformis	3	4.5033E-05	**5.6467E-05**	0.53
PWY_3841_folate transformations II.g__Parabacteroides_s__Parabacteroides_distasonis	1	1.720E-05	**2.8267E-05**	0.65
PWY_3841_folate transformations II.g__Parabacteroides_s__Parabacteroides_distasonis	3	1.5427E-05	**4.3633E-05**	0.46
PWY_3841_folate transformations II.g__Parabacteroides_s__Parabacteroides_merdae	2	**4.1567E-05**	1.7433E-05	0.65
PWY_3841_folate transformations II	3	6.7629E-04	**7.0181E-04**	1.00
PWY_5097_L-lysine biosynthesis VI.g__Parabacteroides_s__Parabacteroides_merdae	2	**4.1167E-05**	2.0867E-05	0.71
PWY_5097_L-lysine biosynthesisVI	3	7.330E-04	**7.248E-04**	1.00
PWY_5097_L-lysine biosynthesis VI.g__Bacteroides_s__Bacteroides_uniformis	3	4.3167E-05	**5.500E-05**	0.35
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Alistipes_s__Alistipes_shahii	2	0	**2.550E-06**	0.58
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Alistipes_s__Alistipes_shahii	3	4.4333E-06	**6.6333E-06**	0.81
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	2	**2.7167E-06**	0	0.39
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Clostridium_s__Clostridium_sp_KLE_1755	3	0	**1.3233E-05**	0.41

PWY_6527_stachyose degradation.unclassified	1	3.870E-05	**9.965E-05**	0.84
PWY_6527_stachyose degradation.g__Clostridium_s__Clostridium_sp_KLE_1755	3	0	**1.7333E-05**	0.82
PWY_6737_starchdegradationV.g__Clostridium_s__Clostridium_asparagiforme	2	**9.420E-06**	4.6233E-06	0.98
PWY_6737_starchdegradationV.g__Clostridium_s__Clostridium_sp_KLE_1755	3	0	**1.1793E-05**	0.66
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Anaerotruncus_s__Anaerotruncus_colihominis	1	4.2667E-07	**6.5467E-06**	0.76
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Lachnospiraceae_noname_s__Lachnospiraceae_bacterium_7_1_58FAA	1	**2.640E-06**	0	0.61
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Parabacteroides_s__Parabacteroides_merdae	1	**1.580E-05**	5.7933E-06	0.81
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Parabacteroides_s__Parabacteroides_merdae	2	**4.640E-05**	2.290E-05	0.80
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Bacteroides_s__Bacteroides_faecis	2	1.0217E-05	**3.51E-05**	0.84
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	2	**4.9467E-06**	8.4333E-07	0.61
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Bacteroides_s__Bacteroides_uniformis	3	4.9733E-05	**6.190E-05**	0.61
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Bacteroides_s__Bacteroides_xylanisolvens	3	2.8267E-05	**3.8733E-05**	0.99
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Clostridium_s__Clostridium_sp_KLE_1755	3	0	**1.6333E-05**	0.59
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Desulfovibrio_s__Desulfovibrio_piger	3	2.3233E-06	**4.4067E-06**	0.93
COMPLETE_ARO_PWY_superpathway of aromatic aminoacid biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	2	**2.7467E-06**	0	0.31
HISTSYN_PWY_L-histidine biosynthesis.g__Bacteroides_s__Bacteroides_faecis	2	6.630E-06	**2.340E-05**	0.30
PWY_7111_pyruvate fermentation to isobutanol_engineered_.g__Clostridium_s__Clostridium_asparagiforme	2	**6.9133E-06**	2.330E-06	0.78
PWY_7237_myo-chiro- and scillo-inositol degradation.unclassified	2	**7.216E-05**	3.0867E-05	0.64
RHAMCAT_PWY_L-rhamnose degradationI.g__Bacteroides_s__Bacteroides_faecis	2	9.7667E-06	**3.1033E-05**	0.91
NONOXIPENT_PWY_pentosephosphate pathway_non-oxidative branch_.g__Bacteroides_s__Bacteroides_xylanisolvens	3	1.690E-05	**2.8933E-05**	0.83
NONOXIPENT_PWY_pentosephosphate pathway_non-oxidative branch_.g__Clostridium_s__Clostridium_bolteae	3	8.040E-06	**1.0943E-05**	0.79
PWY_6121_5-aminoimidazole ribonucleotide biosynthesis I.g__Desulfovibrio_s__Desulfovibrio_piger	3	2.8367E-06	**3.1867E-06**	0.56
PWY_6122_5-aminoimidazole ribonucleotide biosynthesis II.g__Desulfovibrio_s__Desulfovibrio_piger	3	2.740E-06	**3.0033E-06**	0.54
PWY_6277_superpathway of 5-aminoimidazole ribonucleotide biosynthesis.g__Desulfovibrio_s__Desulfovibrio_piger	3	2.740E-06	**3.0033E-06**	0.54
PWY_6317_galactose degradation I_Leloir pathway_.g__Clostridium_s__Clostridium_sp_KLE_1755	3	0	**1.3497E-05**	0.69
PWY_6317_galactose degradation I_Leloir pathway_.g__Lachnospiraceae_noname_s__Lachnospiraceae_bacterium_5_1_57FAA	3	**1.243E-06**	3.98E-07	0.46
PWY_7229_superpathway of adenosine nucleotides denovo biosynthesis I.g__Clostridium_s__Clostridium_asparagiforme	3	**6.3733E-06**	2.9767E-06	0.68
PWY66_422_D-galactose degradation V_Leloir pathway_.g__Clostridium_s__Clostridium_sp_KLE_1755	3	0	**1.3497E-05**	0.69
PWY66_422_D-galactose degradation V_Leloir pathway_.g__Lachnospiraceae_noname_s__Lachnospiraceae_bacterium_5_1_57FAA	3	**1.243E-06**	3.98E-07	0.46

PYRIDNUCSYN_PWY_NAD biosynthesis I_from aspartate_.g__Alistipes_s__Alistipes_shahii	3	**5.1167E-06**	7.9333E-06	0.72

Table 5. Significant metagenomic pathways for GI motility

Pathway [1]	Expt	Avg rel. abundance in Fast motility	Avg rel. abundance in Slow motility	Pathway coverage
CALVIN_PWY_Calvin_Benson_Bassham cycle.g__Clostridium_s__Clostridium_asparagiforme	1	**0.00001349**	6.0933E-06	0.64
CALVIN_PWY_Calvin_Benson_Bassham cycle.g__Clostridium_s__Clostridium_asparagiforme	2	**8.4667E-06**	1.806E-06	0.87
COMPLETE_ARO_PWY_superpathway of aromatic amino acid biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	2	**2.7467E-06**	0	0.31
COMPLETE_ARO_PWY_superpathway of aromatic amino acid biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	1.7400E-06	**3.2267E-06**	0.33
NONMEVIPP_PWY_methylerythritol phosphate pathway Ig__Clostridium_s__Clostridium_asparagiforme	1	**0.00014205**	0.00008976	1.00
NONMEVIPP_PWY_methylerythritol phosphate pathway I.g__Clostridium_s__Clostridium_asparagiforme	2	**3.6033E-06**	9.4167E-07	0.45
NONMEVIPP_PWY_methylerythritol phosphate pathway I.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00013151	**0.00014617**	0.50
NONMEVIPP_PWY_methylerythritol phosphate pathway I.g__Bacteroides_s__Bacteroides_xylanisolvens	3	**2.8833E-05**	2.3067E-05	0.44
NONMEVIPP_PWY_methylerythritol phosphate pathway I.g__Erysipelotrichaceae_noname_s__Clostridium_ramosum	3	8.0933E-06	**7.2900E-06**	0.41
PANTO_PWY_phosphopantothenate biosynthesis I.g__Bacteroides_s__Bacteroides_faecis	2	9.8800E-06	**2.3500E-05**	0.65
PANTO_PWY_phosphopantothenate biosynthesis I.g__Parabacteroides_s__Parabacteroides_merdae	2	**3.7100E-05**	1.6313E-05	0.59
PANTO_PWY_phosphopantothenate biosynthesis I.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00013886	**0.00016071**	0.55
PANTO_PWY_phosphopantothenate biosynthesis I.g__Bacteroides_s__Bacteroides_finegoldii	3	3.7767E-06	**3.4267E-06**	0.39
PWY_1042_glycolysis IV_plant cytosol_.g__Bacteroides_s__Bacteroides_faecis	2	1.0517E-05	**3.1867E-05**	0.52
PWY_1042_glycolysis IV_plant cytosol_.g__Eggerthella_s__Eggerthella_lenta	2	**2.4567E-06**	0	0.60
PWY_1042_glycolysis IV_plant cytosol_.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00019294	**0.00022055**	0.76
PWY_1042_glycolysis IV_plant cytosol_.g__Bacteroides_s__Bacteroides_finegoldii	3	5.0600E-06	**4.5233E-06**	0.74
PWY_1042_glycolysis IV_plant cytosol_.g__Barnesiella_s__Barnesiella_intestinihominis	3	**1.2753E-05**	6.7267E-06	0.68
PWY_2942_L_lysine biosynthesis III.g__Eggerthella_s__Eggerthella_lenta	2	**2.2967E-06**	0	0.47
PWY_2942_L_lysine biosynthesis III.g__Parabacteroides_s__Parabacteroides_merdae	2	**4.4333E-05**	2.1700E-05	0.77
PWY_2942_L_lysine biosynthesis III.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00014536	**0.00016555**	0.37
PWY_2942_L_lysine biosynthesis III.g__Parabacteroides_s__Parabacteroides_distasonis	3	**4.3400E-05**	4.3533E-05	0.84
PWY_2942_L_lysine biosynthesis III.g__Paraprevotella_s__Paraprevotella_clara	3	**1.5700E-05**	1.3777E-05	0.52
PWY_2942_L_lysine biosynthesis III.unclassified	3	**7.0334E-05**	6.2633E-05	0.99
PWY_3841_folate transformations II.g__Parabacteroides_s__Parabacteroides_merdae	2	**4.1567E-05**	1.7433E-05	0.65
PWY_3841_folate transformations II.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00012564	**0.00014328**	0.50
PWY_3841_folate transformations II.g__Bacteroides_s__Bacteroides_ovatus	3	**9.7344E-05**	8.0409E-05	0.77
PWY_5097_L_lysine biosynthesis VI.g__Parabacteroides_s__Parabacteroides_merdae	2	**4.1167E-05**	2.0867E-05	0.71

PWY_5097_L_lysine biosynthesis VI.g__Parabacteroides_s__Parabacteroides_distasonis	3	**4.1633E-05**	3.9600E-05	0.82
PWY_5097_L_lysine biosynthesis VI.g__Paraprevotella_s__Paraprevotella_clara	3	**1.6133E-05**	1.4320E-05	0.55
PWY_5097_L_lysine biosynthesis VI.unclassified	3	**7.8082E-05**	6.7033E-05	1.00
PWY_5695_urate biosynthesis_inosine 5'-phosphate degradation.unclassified	1	**0.00012392**	0.0000421	0.85
PWY_5695_urate biosynthesis_inosine 5'-phosphate degradation.unclassified	3	0.0000282	2.4967E-05	0.75
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Bilophila_s__Bilophila_wadsworthia	1	**1.8157E-05**	7.0533E-06	0.74
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Clostridium_s__Clostridium_asparagiforme	1	**9.8667E-06**	4.3267E-06	0.46
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Alistipes_s__Alistipes_shahii	2	0	**0.00000255**	0.58
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	2	**2.7167E-06**	0	0.39
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	2.6233E-06	**3.2433E-06**	0.54
PWY_6163_chorismate biosynthesis from 3-dehydroquinate.unclassified	3	**8.3313E-05**	6.2000E-05	1.00
PWY_6737_starch degradation V.g__Blautia_s__Ruminococcus_gnavus	1	0	**0.00000514**	0.76
PWY_6737_starch degradation V.g__Lachnospiraceae_noname_s__Lachnospiraceae_bacterium_2_1_58FAA	1	0	**5.5333E-06**	0.90
PWY_6737_starch degradation V.g__Clostridium_s__Clostridium_citroniae	1	**0.00000585**	1.4693E-06	0.84
PWY_6737_starch degradation V.g__Clostridium_s__Clostridium_hylemonae	1	**1.9283E-05**	3.9600E-06	1.00
PWY_6737_starch degradation V.g__Clostridium_s__Clostridium_asparagiforme	1	**0.0000189**	0.00000726	0.99
PWY_6737_starch degradation V.g__Clostridium_s__Clostridium_asparagiforme	2	**9.4200E-06**	4.6233E-06	0.98
PWY_6737_starch degradation V.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00012899	**0.00014921**	0.55
PWY_6737_starch degradation V.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	9.3000E-07	**1.9367E-06**	0.48
PWY_7111_pyruvate fermentation to isobutanol_engineered_.g__Clostridium_s__Clostridium_hathewayi	1	**2.2667E-05**	6.3433E-06	0.51
PWY_7111_pyruvate fermentation to isobutanol_engineered_.g__Clostridium_s__Clostridium_asparagiforme	1	**1.4967E-05**	8.2633E-06	0.86
PWY_7111_pyruvate fermentation to isobutanol_engineered_.g__Clostridium_s__Clostridium_asparagiforme	2	**6.9133E-06**	2.3300E-06	0.78
PWY_7111_pyruvate fermentation to isobutanol_engineered_.g__Clostridium_s__Clostridium_bolteae	3	9.2600E-06	**7.5100E-06**	0.53
PWY_7111_pyruvate fermentation to isobutanol_engineered_.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	3.3400E-06	**4.5333E-06**	0.64
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Bacteroides_s__Bacteroides_vulgatus	1	**9.3312E-05**	4.9033E-05	0.92
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Bacteroides_s__Bacteroides_xylanisolvens	1	**3.3833E-05**	1.4657E-05	0.59
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Clostridium_s__Clostridium_asparagiforme	1	**1.6733E-05**	5.3900E-06	0.92
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Clostridium_s__Clostridium_hylemonae	1	**9.0133E-06**	1.6353E-06	0.54
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Bacteroides_s__Bacteroides_faecis	2	1.0217E-05	**3.5100E-05**	0.84
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	2	**4.9467E-06**	8.4333E-07	0.61
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	2.6067E-06	**4.9733E-06**	0.63

PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Parabacteroides_s__Parabacteroides_merdae	2	**4.6400E-05**	2.2900E-05	0.80
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Parabacteroides_s__Parabacteroides_merdae	3	**2.8567E-05**	2.5133E-05	0.53
PWY_7219_adenosine ribonucleotides denovo biosynthesis	3	**0.00105927**	0.00100928	1.00
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00016588	**0.00018388**	0.53
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Alistipes_s__Alistipes_onderdonkii	3	2.0533E-06	**6.3033E-07**	0.59
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Erysipelotrichaceae_noname_s__Eubacterium_biforme	3	**1.7633E-05**	9.1433E-06	0.59
PWY_7219_adenosine ribonucleotides denovo biosynthesis.g__Parabacteroides_s__Parabacteroides_distasonis	3	**4.5033E-05**	4.2967E-05	0.76
PWY_7221_guanosine ribonucleotides denovo biosynthesis.g__Bacteroides_s__Bacteroides_vulgatus	1	**7.1533E-05**	3.5967E-05	0.65
PWY_7221_guanosine ribonucleotides denovo biosynthesis.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00014504	**0.00017728**	0.87
PWY_7229_super pathway of adenosine nucleotides denovo biosynthesis I.g__Clostridium_s__Clostridium_asparagiforme	1	**1.3017E-05**	4.7433E-06	0.54
PWY_7229_super pathway of adenosine nucleotides denovo biosynthesis I.unclassified	1	**0.00015538**	0.00010696	0.95
PWY_7229_super pathway of adenosine nucleotides denovo biosynthesis I.unclassified	3	**8.2380E-05**	7.3700E-05	0.98
ANAGLYCOLYSIS_PWY_glycolysis III_from glucose_.g__Bacteroides_s__Bacteroides_xylanisolvens	1	**0.000028**	0.00001322	0.41
PWY_5659_GDP_mannose biosynthesis.g__Bacteroides_s__Bacteroides_xylanisolvens	1	**2.4633E-05**	1.2807E-05	0.44
PWY_6317_galactose degradation I_Leloir pathway_.g__Clostridium_s__Clostridium_hylemonae	1	**1.1143E-05**	1.2033E-06	0.83
PWY0_1296_purine ribonucleosides degradation.g__Blautia_s__Ruminococcus_gnavus	1	3.1367E-07	**3.0433E-06**	0.46
PWY0_1296_purine ribonucleosides degradation.g__Lachnospiraceae_noname_s__Lachnospiraceae_bacterium_2_1_58FAA	1	0	**4.1433E-06**	0.63
PWY66_422_D_galactose degradation V_Leloirpathway_.g__Clostridium_s__Clostridium_hylemonae	1	**1.1143E-05**	1.2033E-06	0.83
HISTSYN_PWY_L-histidine biosynthesis.g__Bacteroides_s__Bacteroides_faecis	2	6.6300E-06	**2.3400E-05**	0.30
PWY_7237_myo__chiro_and scillo_inositol degradation.unclassified	2	**7.216E-05**	3.0867E-05	0.64
RHAMCAT_PWY_L_rhamnose degradation I.g__Bacteroides_s__Bacteroides_faecis	2	9.7667E-06	**3.1033E-05**	0.91
ARO_PWY_chorismate biosynthesis I.g__Clostridiales_noname_s__Clostridiales_bacterium_1_7_47FAA	3	1.6633E-06	**3.0367E-06**	0.30
BRANCHED_CHAIN_AA_SYN_PWY_superpathway of branched amino acid biosynthesis.unclassified	3	**7.7105E-05**	6.0833E-05	0.88
COA_PWY_1_coenzymeA biosynthesis II_mammalian_.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00015759	**0.00017845**	0.95
COA_PWY_coenzymeA biosynthesis I.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00015223	**0.00017154**	0.76
COA_PWY_coenzymeA biosynthesis I.g__Bacteroides_s__Bacteroides_ovatus	3	**9.70E-05**	7.8164E-05	0.54
DTDPRHAMSYN_PWY_dTDP_L-rhamnose biosynthesis I.g__Bacteroides_s__Bacteroides_ovatus	3	**0.00015822**	0.00013005	1.00
ILEUSYN_PWY_L-isoleucine biosynthesis I from threonine_.unclassified	3	**8.4032E-05**	6.5800E-05	0.90
NONOXIPENT_PWY_pentosephosphate pathway non-oxidative branch	3	0.00042414	**0.00042192**	1.00
PANTOSYN_PWY_pantothenate and coenzymeA biosynthesis I.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00014907	**0.00017278**	0.66
PWY_5686_UMP biosynthesis.g__Alistipes_s__Alistipes_onderdonkii	3	2.2500E-06	**1.1300E-06**	0.72
PWY_6121_5_aminoimidazole ribonucleotide biosynthesis I	3	0.0005766	**0.00054093**	0.99

PWY_6122_5_aminoimidazole ribonucleotide biosynthesis II	3	0.00053008	**0.00049679**	0.98
PWY_6122_5_aminoimidazole ribonucleotide biosynthesis II.unclassified	3	**7.5016E-05**	6.4833E-05	0.94
PWY_6277_superpathway of 5-aminoimidazole ribonucleotide biosynthesis	3	0.00053008	**0.00049679**	0.98
PWY_6277_superpathway of 5-aminoimidazole ribonucleotide biosynthesis.unclassified	3	**7.5016E-05**	6.4833E-05	0.94
PWY0_1586_peptidoglycan maturation_meso-di-aminopimelate containing_.g__Akkermansia_s__Akkermansia_muciniphila	3	0.00018716	**0.00021**	0.98
TRNA_CHARGING_PWY_tRNA charging.unclassified	3	**7.2134E-05**	6.3533E-05	1.00
VALSYN_PWY_L_valinebiosynthesis.unclassified	3	**8.4032E-05**	6.5800E-05	0.90

Table 6. Taxa from culture-enriched metagenomics that were not in metagenomic sequencing of mouse samples. For each mouse cohort, taxa that were maximally detected at <1% abundance in mice but >1% of culture plates are listed. * indicates cultured taxa found in more than one experiment

Expt 1	**Culture Abundance**	**Mice Abundance**
Coriobacteriaceae	4.469	0.855
Eggerthella, unclassified	3.414	0.122
Bacteroides dorei	1.098	0.208
Flavonifractor plautii *	5.420	0.183
Dorea, unclassified	1.229	0.453
Lachnospiraceae bacterium 7_1_58FAA	1.949	0.113
Clostridium bartletti	2.847	0.133
*Clostridium ramosum**	4.235	0.037
Coprobacillus, unclassified	36.474	0.554
Expt 2		
Eggerthella, unclassified *	3.860	0.489
Staphylococcus epidermidis	17.801	0
Enterococcus faecalis	7.562	0.125
Flavonifractor plautii *	2.284	0.148
Anaerotruncus	1.103	0.314
Escherichia coli *	6.737	0.024
Escherichia, unclassified	1.766	0.083
Expt 3		
Bacteroides finegoldii	2.844	0
Odoribacter splanchnicus	1.741	0.207
Parabacteroides, unclassified	1.052	0.223
Alistipes indistinctus	1.006	0.130
Alistipes onderdonkii	9.946	0.782
Alistipes shahii	2.291	0.818
Clostridium asparagiforme	2.276	0.761
Clostridium hylemonae	1.175	0.225
Clostridium bacterium 1_7_47FAA	1.178	0.230
Lachnospiraceae bacterium 5_1_63FAA	1.388	0.012
Clostridium ramosum *	1.487	0.316
Erysipelotrichaceae bacterium 3_1_53	8.448	0.165
Sutterellaceae, unclassified	1.157	0.757
Bilophila wadsworthia	2.293	0.056
Desulfovibrio piger	18.335	0.128
Escherichia coli *	1.390	0.055

Supplementary Table 1. 16S rRNA sequencing and metagenomic read depths

16S	Expt	Min. read depth	Max. read depth
	1	28,651	109,761
	2	35,156	60,993
	3	41,263	68,990

Metagenomics	Expt		Min. read depth	Max. read depth
	1	Donor 1.1	16,252,584	
		Donor 1.2	8,915,116	
		Mice	7,287,073	14,250,969
		Plates	2,517,445	12,642,712
	2	Donor 1.2	8,915,116	
		Mice	9,587,261	22,350,177
		Plates	2,985,428	7,688,310
	3	Donor 2	2,871,585	
		Mice	9,051,600	12,984,785
		Plates	4,460,237	8,420,315

Supplementary Table 2. Statistically different OTUs from 16S rRNA gene for each phenotype.

Tail Suspension Test

Expt	OTU	High/Low immobile time	LDA Score	p-value	High Mean	Low Mean
1	106 Clostridiaceae	Low	2.671	0.032	0.0000	0.0007
	127 Lachnospiraceae	Low	2.616	0.028	0.0000	0.0077
	318 Alphaproteobacteria	Low	2.115	0.032	0.0000	0.0001
2	41 Turicibacteraceae	Low	4.162	0.034	0.0202	0.0619
	48 Alcaligenaceae	Low	3.164	0.032	0.0015	0.0017
	75 *Alistipes*	Low	2.853	0.034	0.0007	0.0013
	141 *Clostridium*	Low	2.553	0.033	0.0003	0.0007
	2 *Enterococcus*	Low	2.393	0.034	0.0002	0.0004
	83 *Adlercreutzia*	Low	2.358	0.034	0.0003	0.0006
	133 *Bacteroides*	Low	2.008	0.034	0.0001	0.0001
	97 Lachnospiraceae	High	2.609	0.034	0.0006	0.0001
	228 *Bacteroides*	High	2.669	0.019	0.0001	0.0000
3	48 Alcaligenaceae	Low	2.934	0.033	0.0001	0.0019

Light/Dark Box

Expt	OTU	High/Low Light time	LDA Score	p-value	High Mean	Low Mean
1	19 *Coprobacillus*	Low	3.916	0.034	0.0046	0.0170
	1 *Escherichia*	Low	3.589	0.034	0.0004	0.0001
	13 *Bacteroides*	High	3.254	0.034	0.0008	0.0015
	37 *Clostridium*	High	3.462	0.032	0.0004	0.0014
	71 Ruminococcaceae	High	3.52	0.034	0.0007	0.0006
	133 *Bacteroides*	High	4.206	0.034	0.0001	0.0001
	11 *Bacteroides*	High	4.329	0.034	0.0504	0.0301
	39 *Allobaculum*	High	4.646	0.034	0.0645	0.0240
2	10 *Bacteroides*	Low	4.589	0.034	0.0106	0.0134
	32 Lachnospiraceae	Low	3.069	0.034	0.0187	0.0041
	371 Clostridiales	Low	2.686	0.034	0.0000	0.0001
	55 *Ruminococcus*	Low	2.266	0.034	0.0042	0.0006
	85 Clostridiales	Low	2.254	0.034	0.0002	0.0000
	249 Rikenellaceae	High	2.545	0.034	0.0000	0.0002
	154 Ruminococcaceae	High	2.759	0.034	0.0005	0.0007
	39 *Allobaculum*	High	2.822	0.034	0.0645	0.0240
	31 *Oscillospira*	High	2.875	0.034	0.0024	0.0021

	22 *Parabacteroides*	High	3.427	0.034	0.0080	0.0116
	12 *Parabacteroides*	High	3.763	0.034	0.0163	0.0232
3	None					

GI Motility

Expt	OTU	Fast/Slow Transit Time	LDA Score	p-value	High Mean	Low Mean
1	26 *Akkermansia*	Low	4.584	0.034	0.0422	0.0606
	371 Clostridiales	Low	2.69	0.034	0.0000	0.0001
	34 Ruminococcaceae	Low	2.612	0.034	0.0003	0.0006
	46 Ruminococcaceae	High	2.496	0.034	0.0015	0.0009
2	51 *Clostridium*	High	2.795	0.043	0.0025	0.0013
3	65 Ruminococcaceae	High	2.578	0.049	0.0006	0.0001

Chapter 6

Conclusions and Future Directions

Contributions to the field

Within the work of this thesis, I show that applying both culture-based and culture-independent methods allows for greater insight into the complexity and diversity of the human gut microbiota, than either method alone. In the first study, the use of culture-enriched molecular profiling of human stool samples demonstrated that the majority of the bacterial groups could be captured through culturing on agar plates, and that culture was able to detect more microbial diversity than culture-independent 16S rRNA gene sequencing. I also show that culture-enriched molecular profiling could be used for the targeted culturing of a specific bacterial group, the Lachnospiraceae. From the isolated strains, I explored the genetic and phenotypic diversity within species of this family. To more comprehensively characterize the phenotypic and genetic strain diversity within the human gut microbiota, the *E. coli* population in four individuals was studied. Since we determined that culture-enrichment could capture most of the bacterial groups of the gut microbiota, it was applied to a metagenomic study with the goal of identifying bacteria and specific pathways involved in transferring IBS-like phenotypes to germ-free mice from human IBS fecal samples.

Contrary to what has been widely reported in microbiome studies, we have shown that the majority of the bacteria comprising the human gut microbiota can be cultured[184], and this has been supported by recent work from others[130–132,135]. Most importantly, we observed that more bacterial OTUs were found by culture of fecal samples than culture-independent 16S rRNA gene sequencing. The potential implications of this is that culture-independent studies, especially those utilizing low sequencing depth, may be missing or

underrepresenting bacterial groups that could be important for the function or disease being studied. These 'culture-only' OTUs are most likely to be low abundance; however, low abundant bacterial groups could still have functions that impact the community and host. Furthermore, if these bacteria are cultivable, they must be viable and are likely contributing to the overall community.

It was important to develop an accessible method that would allow for most microbiology labs to culture the majority of bacterial groups from the fecal microbiota. Although we relied on chemically undefined media, they are commercially available or contain readily available ingredients, and aside from requiring an anaerobic environment, culturing on agar plates does not utilize specialized equipment. Most gut microbiota studies that have been previously published have used only a few types of culture media, and often only one, but this would require many more replicate plates in order to isolate low abundant bacteria, assuming that all gut bacteria could be supported by one medium[130,132]. In our study, we take advantage of several selective media, and the addition of antibiotics to nonselective media, to more readily capture these less abundant members of the microbiota. The success of this approach is reflected in the increased diversity of bacterial OTUs we observed by culture compared to culture-independent methods.

A major challenge of microbiome studies is determining whether differences in the microbiome are causative or correlated in health and disease. Many culture-independent microbial studies on human diseases have implicated members of the gut microbiota by comparing the microbial communities between patients and healthy

individuals, or diseased and healthy sites (e.g. in colorectal cancer). However, these changes in the microbiome may merely be the consequences of disease pathology, such as inflammation, or they could be important in the etiology or maintenance of disease. Bacterial culture would complement these studies, as bacteria predicted to be important for health or disease should be isolated and tested experimentally to confirm their roles and mechanisms of action. It will be crucial to determine cause and effect, as the gut microbiota could be a potential therapeutic target in diseases where the bacterial community plays an active role in pathophysiology.

As discussed, there have been studies implicating the gut microbiota with several diseases and disorders, and the development of therapies that target the microbiome is an active area of research. Due in part to the efficacy of fecal microbiota transfers (FMT) for *Clostridium difficile* infections and chronic disorders like ulcerative colitis[60,61,254], some therapies that are intended to manipulate gut microbiome are based on live bacteria[255,256]. Due to regulatory restrictions, whole stool samples have limited use, and there is a push to replace donor samples with defined communities. This relies on the ability to culture potentially therapeutic strains to study their efficacy and safety for specific disease indications. As different isolates/strains within a species are unlikely to be equally effective, and indeed, part of this may be host-dependent, each potential live therapeutic needs to be fully characterized, both phenotypically and genetically. Culture allows for phenotyping of isolates and more informative genetic sequencing to characterize strain diversity in the gut microbiota, which could have implications for health and disease.

Through *in vitro* phenotyping and genetic analysis of *E. coli* and Lachnospiraceae isolates, the diversity of these taxa was characterized, and intra-species diversity was found to be prevalent. Although genetic studies have revealed intra-species diversity in the gut, we show that there is also extensive phenotypic variability. Phenotypic variability can be difficult to associate with genetic differences, especially for complex phenotypes such as biofilm formation and swarming motility, although we have not included high-resolution genomic analysis like gene coverage, SNPs and indel identifications (this work is currently on-going). Additionally, the full spectrum of phenotypic heterogeneity, such as levels of antibiotic resistance, would not be apparent with genetic data alone. This diversity adds to the complexity of the bacterial composition in the gut, and also contributes to the inter-individual heterogeneity of the microbiome. It is particularly important to understand the functional role of Lachnospiraceae species, and the potential variability within species, to elucidate their contributions to the community and the host, as they are abundant members of the gut microbiota in many healthy individuals and have been implicated in maintenance of health. Some strains may be more beneficial than others – for the Lachnospiraceae, there may be variability in the production of butyrate or in the ability to provide colonization resistance against *C. difficile* infection. Additionally, the Lachnospiraceae strains characterized included differences in gene content, including pathways for secondary metabolite production, which could encode bioactive products.

The observed co-existence of strains within the human gut raises some intriguing considerations. First, functionally similar strains would likely compete for resources if they occupied the same physical space, but can different strains of a species occupy

different niches in the gut? As the GI tract is not a homogenous environment, with differences in pH, GI motility rates, oxygen content, and bacterial density along the GI tract, and also cross-sectionally through the lumen, biogeographical differences could create distinct niches that allow for intra-species diversity to develop. Secondly, how are strains distributed among a population of individuals; are there differences in geographic locations and ethnicities? One study reported that the phylogenetic composition of *E. coli* in subjects from 10 populations across 3 continents differed by geographic location, although the extent that diet and other lifestyle factors contribute to this is not known[257]. Additionally, how are strains acquired, whether successively or simultaneously, and how stable are strains? Our work, and others, suggests that the *E. coli* population may be dynamic over the course of a few months[205], but could other bacterial groups have more stable functional and genetic profiles? Most importantly, what does intra-species diversity mean for the host? Does colonization with several strains provide increased resilience to perturbation? Although these questions are beyond the scope of this thesis, they are important to consider given our observations of phenotypic and genetic diversity.

An important implication of the intra-species diversity in the human gut microbiota is that the use of references strains is unlikely to be representative of the functional and genetic diversity of a species, especially groups with extensive strain-level variation like *E. coli* and Lachnospiraceae[115]. Reference isolate libraries will require more cultured strains to represent this heterogeneity, and should be accompanied by improvements in reference genomes databases. Metagenomic studies are becoming more common as the cost of sequencing decreases and bioinformatics tools improve to handle

large datasets, but reference datasets have not kept pace. Commensal strains are not widely represented compared to pathogenic bacteria, and most reference genomes are not complete assemblies, and as a result, significant proportions of metagenomic data cannot be assigned functions.

From our studies, we do not know if the functional diversity described is relevant *in vivo* and if this heterogeneity impacts the host. However, insight into this has been provided from a recent study where germ-free mice were mono-colonized with 53 different species to determine their immunomodulatory effects. Some mice were mono-colonized with different strains of the same *Bacteroides* species, and intra-species differences in the lymphocyte populations after colonization were observed[32]. A study of FMT in metabolic syndrome found coexistence of donor and recipient strains, and donor strains were more likely to colonize if a similar strain was already present in the recipient[143]. This suggests that the host has a role in selecting intra-species strain diversity, but whether this is done actively through the immune system, or as a result of environmental factors like the presence of other bacteria, remains to be determined. Combined, these studies indicate the potential of strain diversity to impact the host, and for the host to select for intra-species diversity. However, the FMT study was performed using only shotgun metagenomics and in the monocolonization study, only two strains of each *Bacteroides* species included were tested, so it is important to confirm these results with *in vivo* experiments, and with increased number of strains from phylogenetically diverse bacterial groups, respectively. Although we show that intra-species diversity is extensive in the gut microbiota, in order to determine if this is important in human health

and disease, strains need to be detected from metagenomics. Indeed, a metagenomic study has shown preliminary evidence that strain-level diversity is different between healthy individuals and those with IBD or obesity[115]. However, most of the genes that differentiated strain-level resolution were not functionally annotated. Culture-enriched metagenomics is an approach that could potentially result in improved strain-level sequence data.

I could not determine if bacterial groups and specific functions were associated with the development of IBS-like symptoms in mice colonized with human IBS microbiota. Although both 16S rDNA and metagenomic sequencing found several significantly different bacteria and pathways between IBS or health-like phenotypes in mice, the inconsistencies of transferred phenotypes and the lack of consistent microbiome differences between experiments made it difficult to form conclusions. However, the methods and approach used in this study can be applied to future studies to gain further insight into the contributions of gut bacteria in the development of IBS symptoms. Instead of comparing mice within groups, a study where groups of mice colonized with IBS samples are compared with mice colonized with healthy control samples, an approach that had previously found a potential IBS microbiota signature[77], would be a more robust method to defining bacterial mechanisms in IBS. Using a larger number of mice would also overcome the variability in measuring phenotypes, especially in behaviours. We did find that culture-enriched metagenomics resulted in larger assembled contigs, and improvements in bioinformatics, either assembly or binning tools, should provide better coverage of genomes from cultured communities. We hypothesize that culture-enriched

metagenomics will result in improved assembly of contigs from bacterial communities, and this will result in more mapped reads for metagenomic studies looking at differences in microbial genes between IBS and healthy individuals. Other work from our lab has demonstrated this for respiratory samples[258]. The use of this approach can be applied to future IBS microbiome studies to obtain more conclusive results.

Future Directions

Recent studies show that culture captures mostly the more abundant members of the gut microbiota, and more effort is required to isolate and characterize the rare bacterial groups, as they can be involved in important bacterial and host interactions. Insights into the growth requirements for fastidious bacteria may be elucidated from metagenomics, as the genome of a bacterium can lead to the development of a media for targeted culture of "uncultivable" species[259]. This approach will likely be augmented with the development of third generation sequencing technologies; longer reads can be obtained than from current shotgun metagenomic strategies, which would significantly reduce the complexity of assembling genomes from sequenced communities. Additionally, the focus of this thesis, and most microbiome studies, has been on bacterial members of the gut microbiota, but archaea and fungi are also important functional members of this community[260]. For example, methanogenic archaea are critical for efficient bacterial fermentation by removing hydrogen from the environment. Culture collections of archaea and fungi would be important for studies of inter-kingdom interactions that are part of the GI ecology.

Our studies were limited to only a few human samples, obtained from healthy controls and IBS patients. It would be important to expand these studies to a larger sample size to determine if our observations are consistent. Additionally, we have only cultured and sequenced fecal samples. As the mucosally-adherent microbiota differs from the fecal microbiota population[261], the cultivability of biopsies and the strain diversity from these samples should be characterized and compared to the fecal community. Species that are adherent to the host are likely functionally distinct from luminal bacteria, and have increased host interactions that could be important for disease pathophysiology.

Comprehensive libraries of the gut microbiota, including bacteria, fungi, archaea, and both fecal and mucosal isolates, are critical for further experiments to test hypotheses generated from microbiome studies, and for development of potential therapeutics. Additionally, novel commensal bacteria may produce pharmaceutically relevant products like antibiotics, so strains should continually be isolated. However, a bottleneck exists for high-throughput identification of cultured isolates. Current next-generation sequencing technologies do not currently provide long enough read lengths to reliably assign species from 16S rRNA genes, Sanger sequencing is not economically feasible for a large number of isolates, and identification by MALDI-TOF mass spectrometry, as used by clinical labs, is highly dependent on custom databases that are variable between labs. Isolates should be rigorously described and characterized phylogenetically, genetically and phenotypically to ensure that previously identified bacteria are not being renamed as new species, as this will lead databases to becoming inflated with unreliable and unmatched 16S rRNA gene sequences and genomes.

Overall Significance

An impressive amount of information about the gut microbiota, its functions and composition, has been determined by studies using either a culture-independent or culture-dependent approach. However, moving forward, studies should apply a complementary perspective to combine these two methods, since indispensible information can be gathered by both approaches. The studies in this thesis have combined both culture and molecular profiling, and the use of both methods provides much more information and has higher sensitivity than either method alone. Comparison of bacteria detected by culture-enrichment and culture-independent molecular profiling (both 16S rRNA gene sequencing and shotgun metagenomics) resulted in groups that were only identified by one method, and would be missed if one approach were chosen over the other. Both phenotyping and genotyping bacterial diversity revealed intra-species heterogeneity in two distinct bacterial groups. However, it would not be possible to determine the extent of antibiotic diversity through genetics, and the phylogenetic relatedness of strains would not be conclusive if determined solely from phenotypes, demonstrating the complementarity of culture-based and genetic characterization. As the microbiome field progresses, the ability to determine the role of specific microbial groups in disease and the development of novel therapies will depend on the detection of all bacteria present in a community, and the isolation and characterization of bacterial strains. This will require the combination of culture and molecular profiling. The ideas presented here may be applied to investigations of microbiomes of other body sites that have also been implicated in health and disease (e.g. lung, oral cavity, skin, and vagina),

and the methods proposed can be used for characterization of these communities. In conclusion, the studies presented in this thesis demonstrate that bacterial culture can be applied to the gut microbiota to describe the diversity of the community, and in combination with culture-independent molecular methods, provides insight in to the gut microbiome in health and disease.

References

1. Marchesi, J. R. & Ravel, J. The vocabulary of microbiome research: a proposal. *Microbiome* **3,** 31 (2015).

2. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486,** 215–221 (2012).

3. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–14 (2012).

4. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326,** 1694–7 (2009).

5. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).

6. Sender, R., Fuchs, S. & Milo, R. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* **14,** 1–14 (2016).

7. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308,** 1635–8 (2005).

8. Touw, K. *et al.* Mutual reinforcement of pathophysiological host-microbe interactions in intestinal stasis models. *Physiol. Rep.* **5,** e13182 (2017).

9. Vaishnava, S., Behrendt, C. L., Ismail, A. S., Eckmann, L. & Hooper, L. V. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc. Natl. Acad. Sci.* **105,** 20858–20863 (2008).

10. McGuckin, M. A., Eri, R., Simms, L. A., Florin, T. H. J. & Radford-Smith, G. Intestinal barrier dysfunction in inflammatory bowel diseases. *Inflamm. Bowel Dis.* **15,** 100–113 (2009).

11. Kandori, H., Hirayama, K., Takeda, M. & Doi, K. Histochemical, lectin-histochemical and morphometrical characteristics of intestinal goblet cells of germfree and conventional mice. *Experimental animals / Japanese Association for Laboratory Animal Science* **45,** 155–160 (1996).

12. Fukushima, K. *et al.* Colonization of microflora in mice: Mucosal defense against luminal bacteria. *J. Gastroenterol.* **34,** 54–60 (1999).

13. Jakobsson, H. E. *et al.* The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Rep.* **16,** 164–177 (2015).

14. Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D. J. & Hylemon, P. B. Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* **7,** 22–39 (2016).

15. Cockburn, D. W. & Koropatkin, N. M. Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. *J. Mol. Biol.* **428,** 3230–3252 (2016).

16. Wong, J. M., de Souza, R., Kendall, C. W., Emam, A. & Jenkins, D. J. Colonic health: fermentation and short chain fatty acids. *J Clin Gastroenterol* **40,** 235–243 (2006).

17. Roediger, W. Utilization of nutrients by isolated epithelial cells of the rat colon. *Gastroenterology* **83,** 424–9 (1982).

18. Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461,** 1282–1286 (2009).

19. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **506,** 254–254 (2014).

20. Wang, H. B., Wang, P. Y., Wang, X., Wan, Y. L. & Liu, Y. C. Butyrate enhances intestinal epithelial barrier function via up-regulation of tight junction protein claudin-1 transcription. *Dig. Dis. Sci.* **57,** 3126–3135 (2012).

21. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104,** 13780–5 (2007).

22. Morgan, X. C. *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* **13,** R79 (2012).

23. Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469,** 543–547 (2011).

24. Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* **341,** 569–573 (2013).

25. Shen, Y. *et al.* Outer membrane vesicles of a human commensal mediate immune regulation and disease protection. *Cell Host Microbe* **12,** 509–520 (2012).

26. Round, J. L. & Mazmanian, S. K. Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12204–12209 (2010).

27. Atarashi, K. *et al.* Induction of colonic regulatory T cells. *Science* **331,** 337–342 (2011).

28. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139,** 485–98 (2009).

29. Wu, H. J. *et al.* Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* **32,** 815–827 (2010).

30. Atarashi, K. *et al.* Th17 cell induction by adhesion of microbes to intestinal epithelial cells. *Cell* **163,** 1–14 (2015).

31. Tan, T. G. *et al.* Identifying species of symbiont bacteria from the human gut that, alone, can induce intestinal Th17 cells in mice. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E8141–E8150 (2016).

32. Geva-Zatorsky, N. *et al.* Mining the human gut microbiota for immunomodulatory organisms. *Cell* **168,** 928–943 (2017).

33. Cash, H. L., Whitham, C. V, Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313,** 1126–30 (2006).

34. Vaishnava, S. *et al.* The antibacterial lectin RegIIIg promotes the spatial segregation of microbiota and host in the intestine. *Science* **334,** 255–258 (2011).

35. Natividad, J. M. M. *et al.* Differential induction of antimicrobial REGIII by the intestinal microbiota and *Bifidobacterium breve* NCC2950. *Appl. Environ. Microbiol.* **79,** 7745–7754 (2013).

36. Reeves, A. E., Koenigsknecht, M. J., Bergin, I. L. & Young, V. B. Suppression of *Clostridium difficile* in the gastrointestinal tracts of germfree mice inoculated with a murine isolate from the family Lachnospiraceae. *Infect. Immun.* **80,** 3786–3794 (2012).

37. Haiser, H. J. *et al.* Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* **341,** 295–298 (2013).

38. Vétizou, M. *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* **350,** 1079–1084 (2015).

39. Sivan, A. *et al.* Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* **350,** 1084–1089 (2015).

40. García-González, A. P., Ritter, A. D., Shrestha, S. & Erik, C. Bacterial metabolism affects the *C. elegans* response to cancer chemotherapeutics. *Cell* **169,** 431–441 (2017).

41. Scott, T. A. *et al.* Host-microbe co-metabolism dictates cancer drug efficacy in *C. elegans*. *Cell* **169,** 1–15 (2017).

42. Collins, S. M., Surette, M. & Bercik, P. The interplay between the intestinal microbiota and the brain. *Nat. Rev. Microbiol.* **10,** 735–42 (2012).

43. Yano, J. M. *et al.* Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* **161,** 264–276 (2015).

44. Bercik, P. *et al.* The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice. *Gastroenterology* **141,** 599–609, 609.e1–3 (2011).

45. Hanauer, S. B. Inflammatory Bowel Disease : Epidemiology, Pathogenesis, and Therapeutic Opportunities. **12,** 3–9 (2006).

46. Winslet, M. C., Allan, A., Poxon, V., Youngs, D. & Keighley, M. R. Faecal diversion for Crohn's colitis: a model to study the role of the faecal stream in the inflammatory process. *Gut* **35,** 236–242 (1994).

47. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411,** 603–606 (2001).

48. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411,** 599–603 (2001).

49. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.* **39,** 207–211 (2007).

50. Kabat, A. M. *et al.* The autophagy gene Atg16l1 differentially regulates Treg and TH2 cells to control intestinal inflammation. *Elife* **5,** 1–27 (2016).

51. Halfvarson, J. *et al.* Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* **2,** 17004 (2017).

52. Joossens, M. *et al.* Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* **60,** 631–637 (2011).

53. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15,** 382–92 (2014).

54. Sokol, H. *et al. Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105,** 16731–6 (2008).

55. Darfeuille-Michaud, A. *et al.* High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. *Gastroenterology* **127,** 412–421 (2004).

56. Martinez-Medina, M. *et al.* Molecular diversity of *Escherichia coli* in the human gut: New ecological evidence supporting the role of adherent-invasive *E. coli* (AIEC) in Crohn's disease. *Inflamm. Bowel Dis.* **15,** 872–882 (2009).

57. Small, C.-L. N., Reid-Yu, S. A., McPhee, J. B. & Coombes, B. K. Persistent infection with Crohn's disease-associated adherent-invasive *Escherichia coli* leads to chronic inflammation and intestinal fibrosis. *Nat. Commun.* **4,** 1–12 (2013).

58. Palm, N. W. *et al.* Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158,** 1000–1010 (2014).

59. Lopez, J. & Grinspan, A. Fecal microbiota transplantation for Inflammatory Bowel Disease. *Gastroenterol. Hepatol. (N. Y).* **12,** 374–9 (2016).

60. Rossen, N. G. *et al.* Findings from a randomized controlled trial of fecal transplantation for patients with Ulcerative Colitis. *Gastroenterology* **149,** 110–118 (2015).

61. Moayyedi, P. *et al.* Fecal microbiota transplantation induces remission in patients with active Ulcerative Colitis in a randomized controlled trial. *Gastroenterology* **149,** 102–109 (2015).

62. Jacob, V. *et al.* Single delivery of high-diversity fecal microbiota preparation by colonoscopy is safe and effective in increasing microbial diversity in active Ulcerative Colitis. *Inflamm. Bowel Dis.* **23,** 903–911 (2017).

63. Longstreth, G. F. *et al.* Functional bowel disorders. *Gastroenterology* **130,** 1480–91 (2006).

64. Bhattarai, Y., Muniz Pedrogo, D. A. & Kashyap, P. C. Irritable bowel syndrome: a gut microbiota-related disorder? *Am. J. Physiol. - Gastrointest. Liver Physiol.* **312,** G52–G62 (2017).

65. Kashyap, P. C. *et al.* Complex interactions among diet, gastrointestinal transit, and gut microbiota in humanized mice. *Gastroenterology* **144,** 967–77 (2013).

66. Belmonte, L. *et al.* Role of toll like receptors in irritable bowel syndrome: differential mucosal immune activation according to the disease subtype. *PLoS One* **7,** e42777 (2012).

67. Chadwick, V. S. *et al.* Activation of the mucosal immune system in irritable bowel syndrome. *Gastroenterology* **122,** 1778–1783 (2002).

68. Crouzet, L. *et al.* The hypersensitivity to colonic distension of IBS patients can be transferred to rats through their fecal microbiota. *Neurogastroenterol. Motil.* **25,** e272–82 (2013).

69. Camilleri, M. & Gorman, H. Intestinal permeability and irritable bowel syndrome. *Neurogastroenterol. Motil.* **19,** 545–552 (2007).

70. Simrén, M. *et al.* Intestinal microbiota in functional bowel disorders: a Rome foundation report. *Gut* (2012).

71. Liu, H. N. *et al.* Altered molecular signature of intestinal microbiota in irritable bowel syndrome patients compared with healthy controls: A systematic review and meta-analysis. *Dig. Liver Dis.* 1–7 (2017).

72. Rajilić-Stojanović, M. *et al.* Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology* **141,** 1792–801 (2011).

73. Carroll, I. M., Ringel-Kulka, T., Siddle, J. P. & Ringel, Y. Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome. *Neurogastroenterol. Motil.* **24,** 521–30, e248 (2012).

74. Tap, J. *et al.* Identification of an intestinal microbiota signature associated with severity of Irritable Bowel Syndrome. *Gastroenterology* **152,** 111–123.e8 (2017).

75. Marshall, J. K. *et al.* Incidence and epidemiology of irritable bowel syndrome after a large waterborne outbreak of bacterial dysentery. *Gastroenterology* **131,** 445–50; (2006).

76. Thabane, M., Kottachchi, D. T. & Marshall, J. K. Systematic review and meta-analysis: The incidence and prognosis of post-infectious irritable bowel syndrome. *Aliment. Pharmacol. Ther.* **26,** 535–544 (2007).

77. De Palma, G. *et al.* Transplantation of fecal microbiota from patients with irritable bowel syndrome alters gut function and behavior in recipient mice. *Sci. Transl. Med.* **9,** 1–15 (2017).

78. Wang, T. *et al.* Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J.* **6,** 320–329 (2012).

79. Wu, S. *et al.* A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat. Med.* **15,** 1016–1022 (2009).

80. Arthur, J. C. *et al.* Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota. *Science* **338,** 120–124 (2012).

81. Kostic, A. D. *et al.* Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. 292–298 (2012).

82. Kostic, A. D. *et al. Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14,** 207–215 (2013).

83. Zackular, J. P. *et al.* The gut microbiome modulates colon tumorigenesis. *MBio* **4,** 1–9 (2013).

84. Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101,** 15718–23 (2004).

85. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci.* **102,** 11070–11075 (2005).

86. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–31 (2006).

87. Perry, R. J. *et al.* Acetate mediates a microbiome–brain–β-cell axis to promote metabolic syndrome. *Nature* **534,** 213–217 (2016).

88. Sze, M. A. & Schloss, P. D. Looking for a signal in the noise: Revisting obesity and the Microbiome. *MBio* **7,** 1–9 (2016).

89. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490,** 55–60 (2012).

90. Denou, E. *et al.* Defective NOD2 peptidoglycan sensing promotes diet-induced inflammation, dysbiosis, and insulin resistance. *EMBO Mol. Med.* **7,** 259–274 (2015).

91. Wen, L. *et al.* Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455,** 1109–1113 (2008).

92. Livanos, A. E. *et al.* Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. *Nat. Microbiol.* **1,** 16140 (2016).

93. De Goffau, M. C. *et al.* Fecal microbiota composition differs between children with b-Cell autoimmunity and those without. *Diabetes* **62,** 1238–1244 (2013).

94. Murri, M. *et al.* Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. *BMC Med.* **11,** 46 (2013).

95. Vatanen, T. *et al.* Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* 842–853 (2015).

96. Neufeld, K. M., Kang, N., Bienenstock, J. & Foster, J. a. Reduced anxiety-like behavior and central neurochemical change in germ-free mice. *Neurogastroenterol. Motil.* **23,** 255–265 (2011).

97. De Palma, G. *et al.* Microbiota and host determinants of behavioural phenotype in maternally separated mice. *Nat. Commun.* **6,** 7735 (2015).

98. Kelly, J. R. *et al.* Transferring the blues: Depression-associated gut microbiota induces neurobehavioural changes in the rat. *J. Psychiatr. Res.* **82,** 109–118 (2016).

99. Zheng, P. *et al.* Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. *Mol. Psychiatry* **21,** 786–796 (2016).

100. Jiang, H. *et al.* Altered fecal microbiota composition in patients with major depressive disorder. *Brain. Behav. Immun.* **48,** 186–194 (2015).

101. Bravo, J. A. *et al.* Ingestion of *Lactobacillus* strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proc. Natl. Acad. Sci.* **108,** 16050–16055 (2011).

102. Burokas, A. *et al.* Targeting the microbiota-gut-brain axis: prebiotics have anxiolytic and antidepressant-like effects and reverse the impact of chronic stress in mice. *Biol. Psychiatry* 1–16 (2016).

103. Tarr, A. J. *et al.* The prebiotics 3'Sialyllactose and 6'Sialyllactose diminish stressor-induced anxiety-like behavior and colonic microbiota alterations: Evidence for effects on the gut-brain axis. *Brain. Behav. Immun.* **50,** 166–177 (2015).

104. Wong, M. L. *et al.* Inflammasome signaling affects anxiety- and depressive-like behavior and gut microbiome composition. *Mol. Psychiatry* **21,** 797–805 (2016).

105. Savica, R. *et al.* Medical records documentation of constipation preceding Parkinson disease: A case-control study. *Neurology* **73,** 1752–1758 (2009).

106. Kang, D. W. *et al.* Reduced incidence of *Prevotella* and other fermenters in intestinal microflora of autistic children. *PLoS One* **8,** (2013).

107. De Angelis, M. *et al.* Fecal microbiota and metabolome of children with autism and pervasive developmental disorder not otherwise specified. *PLoS One* **8,** 1–18 (2013).

108. Keshavarzian, A. *et al.* Colonic bacterial composition in Parkinson's disease. *Mov. Disord.* **30,** 1351–1360 (2015).

109. Scheperjans, F. *et al.* Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov. Disord.* **30,** 350–358 (2015).

110. Hill-Burns, E. M. *et al.* Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome. *Mov. Disord.* **32,** 739–749 (2017).

111. Hsiao, E. Y. *et al.* Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155,** 1451–1463 (2013).

112. Buffington, S. a. *et al.* Microbial reconstitution reverses maternal diet-induced social and synaptic deficits in offspring. *Cell* **165,** 1762–1775 (2016).

113. Sampson, T. R. *et al.* Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's Disease. *Cell* **167,** 1469–1480.e12 (2016).

114. Kuczynski, J. *et al.* Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* **13,** 47–58 (2011).

115. Greenblum, S., Carr, R. & Borenstein, E. Extensive strain-level copy-number variation across human gut microbiome species. *Cell* **160,** 583–594 (2015).

116. Morgan, X. C. & Huttenhower, C. Meta'omic analytic techniques for studying the intestinal microbiome. *Gastroenterology* **146,** 1437–1448.e1 (2014).

117. Jansson, J. K. & Baker, E. S. A multi-omic future for microbiome studies. *Nat. Microbiol.* **1,** 16049 (2016).

118. Franzosa, E. A. *et al.* Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci.* **111,** E2329–E2338 (2014).

119. Franzosa, E. A. *et al.* Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat. Rev. Microbiol.* **13,** 360–372 (2015).

120. Hayashi, H., Sakamoto, M. & Benno, Y. Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiol. Immunol.* **46,** 535–48 (2002).

121. Byrd, B. A. L., Segre, J. a & Koch, R. Adapting Koch's postulates. *Science* **351,** 224–6 (2016).

122. Fodor, A. a *et al.* The 'most wanted' taxa from the human microbiome for whole genome sequencing. *PLoS One* **7,** e41294 (2012).

123. Jumpstart, C. H. M. P. D. G. W. G. Evaluation of 16S rDNA-Based Community Profiling for Human Microbiome Research. *PLoS One* **7,** e39315 (2012).

124. Rajilić-Stojanović, M. & de Vos, W. M. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol. Rev.* **38,** 996–1047 (2014).

125. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489,** 220–230 (2012).

126. Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158,** 1402–1414 (2014).

127. Nakatsuji, T. *et al.* Antimicrobials from human skin commensal bacteria protect against *Staphylococcus aureus* and are deficient in atopic dermatitis. *Sci. Transl. Med.* **9,** 1–12 (2017).

128. Moore, W. E. C. & Holdeman, L. V. Human Fecal Flora : The Normal Flora of 20 Japanese-Hawaiians. *Appl. Microbiol.* **27,** 961–979 (1974).

129. Moore, W. E. C. & Moore, L. H. Intestinal floras of populations that have a high risk of colon cancer. *Appl. Environ. Microbiol.* **61,** 3202–3207 (1995).

130. Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **108,** 6252–6257 (2011).

131. Rettedal, E. A, Gumpert, H. & Sommer, M. O. A. Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. *Nat. Commun.* **5,** 4714 (2014).

132. Browne, H. P. *et al.* Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533,** 543–436 (2016).

133. Lagkouvardos, I., Overmann, J. & Clavel, T. Cultured microbes represent a substantial fraction of the human and mouse gut microbiota. *Gut Microbes* **0976,** 00–00 (2017).

134. Lagier, J. C. *et al.* Microbial Culturomics: Paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* **18,** 1185–1193 (2012).

135. Lagier, J. C. *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat. Microbiol.* **1,** 16203 (2016).

136. Ma, L. *et al.* Gene-targeted microfluidic cultivation validated by isolation of a gut bacterium listed in Human Microbiome Project's Most Wanted taxa. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 9768–73 (2014).

137. Gibson, G. R., Cummings, J. H. & Macfarlane, G. T. Use of a three-stage continuous culture system to study the effect of mucin on dissimilatory sulfate reduction and methanogenesis by mixed populations of human gut bacteria. *Appl. Environ. Microbiol.* **54,** 2750–2755 (1988).

138. Gibson, G. R. & Wang, X. Enrichment of bifidobacteria from human gut contents by oligofructose using continuous culture. *FEMS Microbiol. Lett.* **118,** 121–7 (1994).

139. McDonald, J. A. K. *et al.* Evaluation of microbial community reproducibility, stability and composition in a human distal gut chemostat model. *J. Microbiol. Methods* **95,** 167–74 (2013).

140. Caminero, A. *et al.* Diversity of the cultivable human gut microbiome involved in gluten metabolism: Isolation of microorganisms with potential interest for coeliac disease. *FEMS Microbiol. Ecol.* **88,** 309–319 (2014).

141. Walker, A. W., Duncan, S. H., Louis, P. & Flint, H. J. Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends Microbiol.* 1–8 (2014).

142. Zhu, A., Sunagawa, S., Mende, D. R. & Bork, P. Inter-individual differences in the gene content of human gut bacterial species. *Genome Biol.* **16,** 82 (2015).

143. Li, S. S. *et al.* Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science* **352,** 586–590 (2016).

144. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–14 (2012).

145. David, L. a *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505,** 559–63 (2014).

146. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl Acad. Sci. USA* **108,** 4554–61 (2011).

147. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15,** 382–92 (2014).

148. Sommer, M. O. Advancing gut microbiome research using cultivation. *Curr. Opin. Microbiol.* **27,** 127–132 (2015).

149. Sibley, C. D. *et al.* Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PLoS One* **6,** e22702 (2011).

150. Drossman, D. a. The functional gastrointestinal disorders and the Rome III process. *Gastroenterology* **130,** 1377–90 (2006).

151. Atlas, R. *Handbook of microbiological media*. (CRC Press, 2004).

152. Sibley, C. D. *et al.* McKay agar enables routine quantification of the 'Streptococcus milleri' group in cystic fibrosis patients. *J. Med. Microbiol.* **59,** 534–40 (2010).

153. Whelan, F. *et al.* The loss of topography in the microbial communities of the upper respiratory tract in the elderly. *Ann Am Thorac Soc* **11,** 513–21 (2014).

154. Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl. Environ. Microbiol.* **77,** 3846–52 (2011).

155. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17,** 10–2 (2011).

156. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13,** 31 (2012).

157. Ye, Y. Identification and quantification of abundant species from pyrosequences of 16S rRNA by consensus alignment. *Proc IEEE Int. Conf. Bioinform Biomed* **2010,** 153–7 (2011).

158. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–7 (2007).

159. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–72 (2006).

160. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7,** 335–6 (2010).

161. Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3,** e1029 (2015).

162. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).

163. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).

164. Posada, D. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **25,** 1253–1256 (2008).

165. Silvestro, D. & Michalak, I. raxmlGUI: a graphical front-end for RAxML. *Org Divers Evol* **12,** 335–7 (2012).

166. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).

167. Matsen, F. A, Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11,** 538 (2010).

168. Rambaut, A. & Drummond, A. Figtree v1.3.1:tree figure drawing tool. (2009).

169. Collins, S. M. A role for the gut microbiota in IBS. *Nat. Rev. Gastroenterol. Hepatol.* **11,** 497–505 (2014).

170. Kim, B. S., Kim, J. N. & Cerniglia, C. E. In vitro culture conditions for maintaining a complex population of human gastrointestinal tract microbiota. *J. Biomed. Biotechnol.* **2011,** 838040 (2011).

171. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).

172. Zhou, Y. *et al.* Biogeography of the ecosystems of the healthy human body. *Genome Biol.* **14,** R1 (2013).

173. Van den Abbeele, P. *et al.* Arabinoxylans and inulin differentially modulate the mucosal and luminal gut microbiota and mucin-degradation in humanized rats. *Environ. Microbiol.* **13,** 2667–2680 (2011).

174. Scott, K. P., Martin, J. C., Duncan, S. H. & Flint, H. J. Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro. *FEMS Microbiol. Ecol.* **87,** 30–40 (2014).

175. Kuske, C. *et al.* Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. *Appl. Environ. Microbiol.* **64,** 2463–72 (1998).

176. Kaeberlein, T., Lewis, K. & Epstein, S. S. Isolating 'Uncultivable' microorganisms in pure culture in a simulated natural environment. *Science* **296,** 1127–1129 (2002).

177. Rainey, F. A. *Lachnospiraceae fam. nov. Bergey's Manual of Systematics of Archaea and Bacteria. 1-2*. (Bergey's Manual Trust, 2009).

178. Dewhirst, F. E. *et al.* The human oral microbiome. *J. Bacteriol.* **192,** 5002–5017 (2010).

179. Meehan, C. J. & Beiko, R. G. A phylogenomic view of ecological specialization in the lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol. Evol.* **6,** 703–713 (2014).

180. Duncan, S. H., Louis, P. & Flint, H. J. Cultivable bacterial diversity from the human colon. *Lett. Appl. Microbiol.* **44,** 343–50 (2007).

181. Puertollano, E., Kolida, S. & Yaqoob, P. Biological significance of short-chain fatty acid metabolism by the intestinal microbiome. *Curr. Opin. Clin. Nutr. Metab. Care* **17,** 139–44 (2014).

182. Saulnier, D. M. *et al.* Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome. *Gastroenterology* **141,** 1782–91 (2011).

183. Nakanishi, Y., Sato, T. & Ohteki, T. Commensal Gram-positive bacteria initiates colitis by inducing monocyte/macrophage mobilization. *Mucosal Immunol.* 1–9 (2014).

184. Lau, J. T. *et al.* Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Med.* **8,** 72 (2016).

185. Grinwis, M. E. *et al.* Characterization of Streptococcus milleri group isolates from expectorated sputum of adult patients with cystic fibrosis. *J. Clin. Microbiol.* **48,** 395–401 (2010).

186. Kim, W. & Surette, M. G. Coordinated regulation of two independent cell-cell signaling systems and swarmer differentiation in *Salmonella enterica* serovar typhimurium. *J. Bacteriol.* **188,** 431–440 (2006).

187. Coil, D., Jospin, G. & Darling, A. E. A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* **31,** 587–589 (2015).

188. Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12,** 402 (2011).

189. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44,** 1–6 (2016).

190. Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* **11,** 119 (2017).

191. Jia, B. *et al.* CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45,** D566–D573 (2017).

192. Rhee, S., Chang, T. S., Jeong, W. & Kang, D. Methods for detection and measurement of hydrogen peroxide inside and outside of cells. *Mol. Cells* **29,** 539–549 (2010).

193. Theriot, C. M., Bowman, A. a & Young, V. B. Antibiotic-induced alterations of the gut microbiota alter secondary bile acid production and allow for *Clostridium difficile* spore germination and outgrowth in the large intestine. *mSphere* **1,** e00045–15 (2015).

194. Cullen, T. *et al.* Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. *Science* **347,** 170–175 (2015).

195. Federle, M. J. & Bassler, B. L. Interspecies communication in bacteria. *J. Clin. Invest.* **112,** 1291–1299 (2003).

196. Surette, M. G. & Bassler, B. L. Quorum sensing in *Escherichia coli* and *Salmonella typhimurium. Proc. Natl. Acad. Sci. U. S. A.* **95,** 7046–50 (1998).

197. Ohtani, K., Hayashi, H. & Shimizu, T. The luxS gene is involved in cell – cell signalling for toxin production in *Clostridium perfringens*. *Mol. Microbiol.* **44,** 171–179 (2002).

198. Hsiao, A. *et al.* Members of the human gut microbiota involved in recovery from *Vibrio cholerae* infection. *Nature* **515,** 423–6 (2014).

199. Lim, B., Zimmermann, M., Barry, N. a. & Goodman, A. L. Engineered regulatory systems modulate gene expression of human commensals in the gut. *Cell* **169,** 547–558.e15 (2017).

200. Whitaker, W. R., Shepherd, E. S. & Sonnenburg, J. L. Tunable expression tools enable single-cell strain distinction in the gut microbiome. *Cell* **169,** 538–546.e12 (2017).

201. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17,** 690–703 (2015).

202. Rajilić-Stojanović, M. & de Vos, W. M. The First 1,000 Cultured Species of the Human Gastrointestinal Microbiota. *FEMS Microbiol.* (2014).

203. Croxen, M. A & Finlay, B. B. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat. Rev. Microbiol.* **8,** 26–38 (2010).

204. Bailey, J. K., Pinyon, J. L., Anantham, S. & Hall, R. M. Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants. *J. Med. Microbiol.* **59,** 1331–9 (2010).

205. Anderson, M. A, Whitlock, J. E. & Harwood, V. J. Diversity and distribution of *Escherichia coli* genotypes and antibiotic resistance phenotypes in feces of humans, cattle, and horses. *Appl. Environ. Microbiol.* **72,** 6914–22 (2006).

206. Rasko, D. a *et al.* The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* **190,** 6881–93 (2008).

207. Gordon, D. M., O'Brien, C. L. & Pavli, P. *Escherichia coli* diversity in the lower intestinal tract of humans. *Environ. Microbiol. Rep.* **7,** 642–648 (2015).

208. Workentine, M. L. *et al.* Phenotypic heterogeneity of *Pseudomonas aeruginosa* populations in a Cystic Fibrosis patient. *PLoS One* **8,** 1–10 (2013).

209. Levert, M. *et al.* Molecular and evolutionary bases of within-patient genotypic and phenotypic diversity in *Escherichia coli* extraintestinal infections. *PLoS Pathog.* **6,** e1001125 (2010).

210. Chen, S. L. *et al.* Genomic diversity and fitness of *E. coli* strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. *Sci. Transl. Med.* **5,** 184ra60 (2013).

211. De Paepe, M. *et al.* Trade-off between bile resistance and nutritional competence drives *Escherichia coli* diversification in the mouse gut. *PLoS Genet.* **7,** e1002107 (2011).

212. Barroso-Batista, J. *et al.* The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps. *PLoS Genet.* **10,** (2014).

213. Barroso-Batista, J., Demengeot, J. & Gordo, I. Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut bacteria. *Nat. Commun.* **6,** 8945 (2015).

214. O'Brien, C. L. & Gordon, D. M. Effect of diet and gut dynamics on the establishment and persistence of *Escherichia coli. Microbiology* **157,** 1375–84 (2011).

215. Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W. & Wilson, D. J. Within-host evolution of bacterial pathogens. *Nat. Rev. Microbiol.* **14,** 150–62 (2016).

216. Chen, J. & Griffiths, M. W. PCR differentiation of *Escherichia coli* from other gram-negative bacteria using primers derived from the nucleotide sequences flanking the gene encoding the universal stress protein. *Lett. Appl. Microbiol.* **27,** 369–371 (1998).

217. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–82 (2012).

218. O'Toole, G. A. Microtiter Dish Biofilm Formation Assay. *J. Vis. Exp.* **47,** (2011).

219. Dong, T., Chiang, S. M., Joyce, C., Yu, R. & Schellhorn, H. E. Polymorphism and selection of *rpoS* in pathogenic *Escherichia coli*. *BMC Microbiol.* **9,** 118 (2009).

220. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 5.0, 2015. http://www.eucast.org. (2015). at <http://www.eucast.org/>

221. Pacheco, A. B. F., Guth, B. E. C., De Almeida, D. F. & Ferreira, L. C. S. Characterization of enterotoxigenic *Escherichia coli* by random amplification of polymorphic DNA. *Res. Microbiol.* **147,** 175–182 (1996).

222. Briñas, L. *et al.* beta-Lactamases in ampicillin-resistant *Escherichia coli* isolates from foods, humans, and healthy animals. *Antimicrob. Agents Chemother.* **46,** 3156–3163 (2002).

223. Saenz, Y. *et al.* Mechanisms of resistance in multiple-antibiotic-resistant *Escherichia coli* strains of human, animal, and food origins. *Antimicrob. Agents Chemother.* **48,** 3996–4001 (2004).

224. Haenni, M. *et al.* Co-occurrence of extended spectrum β lactamase and MCR-1 encoding genes on plasmids. *Lancet Infect. Dis.* **16,** 281–282 (2016).

225. Bokranz, W., Wang, X., Tschäpe, H. & Römling, U. Expression of cellulose and curli fimbriae by *Escherichia coli* isolated from the gastrointestinal tract. *J. Med. Microbiol.* **54,** 1171–1182 (2005).

226. White, A. P. *et al.* Intergenic sequence comparison of *Escherichia coli* isolates reveals lifestyle adaptations but not host specificity. *Appl. Environ. Microbiol.* **77,** 7620–7632 (2011).

227. Cullender, T. C. *et al.* Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut. *Cell Host Microbe* **14,** 571–581 (2013).

228. Poonja, A. *et al.* Effect of freezing sputum on *Pseudomonas aeruginosa* population heterogeneity. *J. Cyst. Fibros.* **16,** 353–357 (2017).

229. Matic, I. *et al.* Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* **277,** 1833–1834 (1997).

230. Davidson, C. J. & Surette, M. G. Individuality in bacteria. *Annu. Rev. Genet.* **42,** 253–68 (2008).

231. Duboc, H. *et al.* Increase in fecal primary bile acids and dysbiosis in patients with diarrhea-predominant irritable bowel syndrome. *Neurogastroenterol. Motil.* **24,** 513–20, e246–7 (2012).

232. Kaas, R. S. *et al.* Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* **13,** 577 (2012).

233. Turroni, F. *et al.* Exploring the diversity of the Bifidobacterial population in the human intestinal tract. *Appl. Environ. Microbiol.* **75,** 1534–1545 (2009).

234. Ford, A. C., Lacy, B. E. & Talley, N. J. Irritable Bowel Syndrome. *N. Engl. J. Med.* **376,** 2566–78 (2017).

235. Fedorak, R. N., Vanner, S. J., Paterson, W. G. & Bridges, R. J. Canadian Digestive Health Foundation Public Impact Series 3: Irritable bowel syndrome in Canada. Incidence, prevalence, and direct and indirect economic impact. *Can. J. Gastroenterol.* **26,** 252–6 (2012).

236. Lacy, B. E. *et al.* Bowel disorders. *Gastroenterology* **150,** 1393–1407e5 (2016).

237. Ek, W. E. *et al.* Exploring the genetics of irritable bowel syndrome: a GWA study in the general population and replication in multinational case-control cohorts. *Gut* **0,** 1–9 (2014).

238. Malinen, E. *et al.* Association of symptoms with gastrointestinal microbiota in irritable bowel syndrome. *World J. Gastroenterol.* **16,** 4532–4540 (2010).

239. Chassard, C. *et al.* Functional dysbiosis within the gut microbiota of patients with constipated-irritable bowel syndrome. *Aliment. Pharmacol. Ther.* **35,** 828–38 (2012).

240. Jeffery, I. B. *et al.* An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut* **61,** 997–1006 (2012).

241. Reed, D. E. *et al.* Bead study: a novel method to measure gastrointestinal transit in mice. *Neurogastroenterol. Motil.* **26,** 1663–8 (2014).

242. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12,** R60 (2011).

243. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

244. Abubucker, S. *et al.* Metabolic Reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* **8,** e1002358 (2012).

245. Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12,** 902–903 (2015).

246. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31,** 1674–1676 (2014).

247. Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics* **32,** 1088–1090 (2016).

248. Belzer, C. & de Vos, W. M. Microbes inside--from diversity to function: the case of Akkermansia. *ISME J.* **6,** 1449–58 (2012).

249. Vandeputte, D. *et al.* Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* 1–6 (2015).

250. Smriga, M. & Torii, K. L-Lysine acts like a partial serotonin receptor 4 antagonist and inhibits serotonin-mediated intestinal pathologies and anxiety in rats. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 15370–5 (2003).

251. Zhang, L. S. & Davies, S. S. Microbial metabolism of dietary components to bioactive metabolites: opportunities for new therapeutic interventions. *Genome Med.* **8,** 46 (2016).

252. Manocha, M. & Khan, W. I. Serotonin and GI Disorders: An Update on Clinical and Experimental Studies. *Clin. Transl. Gastroenterol.* **3,** e13 (2012).

253. Sczyrba, A. *et al.* Critical Assessment of Metagenome Interpretation − a benchmark of computational metagenomics software. *bioRxiv* 099127 (2017).

254. Bakken, J. S. *et al.* Treating clostridium difficile infection with fecal microbiota transplantation. *Clin. Gastroenterol. Hepatol.* **9,** 1044–1049 (2011).

255. Petrof, E. O. *et al.* Stool substitute transplant therapy for the eradication of Clostridium difficile infection: 'RePOOPulating' the gut. *Microbiome* **1,** 3 (2013).

256. Khanna, S. *et al.* A novel microbiome therapeutic increases gut microbial diversity and prevents recurrent *Clostridium difficile* infection. *J. Infect. Dis.* **214,** 173–181 (2016).

257. Escobar-Páramo, P. *et al.* Large-scale population structure of human commensal *Escherichia coli* isolates. 1–4 (2004). doi:10.1128/AEM.70.9.5698

258. Whelan, F. Investigations of the microbial communities of the respiratory tract in the elderly and in Cystic Fibrosis via culture-dependent and -independent approaches. (McMaster University, 2017).

259. Bomar, L., Maltz, M., Colston, S. & Graf, J. Directed culturing of microorganisms using metatranscriptomics. *MBio* **2,** 1–8 (2011).

260. Hoffmann, C. *et al.* Archaea and Fungi of the human gut microbiome: Correlations with diet and bacterial residents. *PLoS One* **8,** e66019 (2013).

261. Zoetendal, E. G. *et al.* Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl. Environ. Microbiol.* **68,** 3401 (2002).