COLONIZATION OF THE INTESTINAL SURFACE BY INDIGENOUS MICROBIOTA

Thesis by

Gregory Paul Donaldson

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2018

(Defended April 26, 2018)

© 2018

Gregory Paul Donaldson

All rights reserved

ORCID: 0000-0002-8551-374X

ACKNOWLEDGEMENTS

I would like to foremost thank my advisor, Sarkis Mazmanian, for support, guidance, and maintaining confidence in me through highs and lows of research. My thesis committee, Dianne Newman, Victoria Orphan, and Raymond Deshaies, provided astute criticism as well as encouragement, at appropriate times. The opportunity to come to Caltech was enabled by the mentorship of Vincent Lee, who taught me how to work with bacteria when I was an undergraduate at the University of Maryland.

I have benefited from proximity to an incredible group of scientists in the Mazmanian lab, many of whom contributed ideas, criticism, and support. Melanie Lee, who discovered the scientific basis of this thesis, taught me nearly all of the technical skills involved. As projects did not pan out the way I expected, Elaine Hsiao showed me how to be dauntless in research.

And I am grateful to my parents and sisters, for the lifelong encouragement that has led me to make movies, to dance, and to discover something new.

ABSTRACT

The mammalian gut evolved to foster the development and maintenance of a community of specific bacterial symbionts that persist for years. *Bacteroides fragilis* is one of a number of species that are able to colonize the mucus of the large intestine in mice and humans. This thesis explores the mechanisms and functions of mucosal colonization, most notably by using reductionist approaches with gnotobiotic mice. Harnessing genetics on both the host and microbial side allowed the dissection of a pathway by which immunoglobulin A enhances mucosal colonization by *B. fragilis*. Novel colonization assays were developed to explore the importance of mucosal colonization to bacterial fitness. Finally, an enrichment method for host-associated bacterial transcriptomics was used to define the behavior of this symbiont within the mucus layer.

PUBLISHED CONTENT AND CONTRIBUTIONS

Chapter 1:

Donaldson G.P., Lee S.M., and Mazmanian S.K. (2016) "Gut biogeography of the bacterial microbiota." ***Nature Reviews Microbiology*** 14 (1): 20-32. DOI: 10.1038/nmicro3552

G.P.D. conceived the topics, wrote the bulk of the review, and drew the figures.

Chapter 2:

Donaldson G.P., Ladinsky M.S., Yu K.B., Sanders J.G., Yoo B.B., Chou W.C., Conner M.E., Earl A.M., Knight R., Bjorkman P.J., and Mazmanian S.K. (2018) "Gut microbiota utilize immunoglobulin A for mucosal colonization." ***Science.*** DOI: 10.1126/science.aaq0926

G.P.D. conceived the project, led the design and interpretation of all experiments, performed most of the experiments, and wrote the paper.

Chapter 3:

Donaldson G.P., Chou W.C., Manson A.L., Abeel T., Ciulla D.M., Rogov P., Bochicchio J., Giannoukos G., Melnikov A., Earnst P.B., Chu H., Earl A.M., and Mazmanian S.K. "Hybrid selection RNA-seq reveals mucosal colonization factors in *Bacteroides fragilis*." To be submitted for publication.

G.P.D. conceived the project, performed all experiments except hybrid selection RNA-seq, led the interpretation of results, and wrote the paper.

TABLE OF CONTENTS

C h a p t e r 1

GUT BIOGEOGRAPHY OF THE BACTERIAL MICROBIOTA

Donaldson G.P., Lee S.M., and Mazmanian S.K.

This chapter was published in 2016 in *Nature Reviews Microbiology* 14 (1): 20-32. DOI: 10.1038/nmicro3552

PREFACE

Animals assemble and maintain a diverse yet host-specific gut microbial community. In addition to characteristic microbial compositions along the longitudinal axis of the intestines, discrete bacterial communities form in microhabitats, such as the gut lumen, colon mucus layers, and colon crypts. In this review, we examine how spatial distribution of symbiotic bacteria among physical niches in the gut impacts the development and maintenance of a resilient microbial ecosystem. We consider novel hypotheses for how nutrient selection, immune activation, and other mechanisms control the biogeography of bacteria in the gut and discuss the relevance of this spatial heterogeneity to health and disease.

INTRODUCTION

Humans and other mammals harbor a complex gastrointestinal microbiota, which includes all three domains of life (Archaea, Bacteria and Eukaryota). This extraordinary symbiosis, formed via a series of exposures to environmental factors, is initiated upon contact with the vaginal microbiota during birth[1]. Abrupt changes during the first year of life follow a pattern that corresponds to gestational age in both mice[2] and humans[3], which suggests that strong deterministic processes shape the composition of the microbiota during development. These population shifts may be explained by influences from diet, the developing immune system, and chemical exposures, as well as potential founder effects of initial colonizers. Founder effects are not well understood in the mammalian gut, but the profound changes in host gene expression that occur in response to microorganisms, and the great potential for syntrophic interactions between bacteria suggest that early colonizers may have long-term effects on the establishment of the microbiota. The immune system imposes selective pressure on the microbiota through both innate and adaptive mechanisms such as antimicrobial peptides[4], secreted immunoglobulin A (IgA)[5], and other contributing factors[6] (see below). However, current research suggests that diet may have the greatest impact on microbiota assembly.

Prior to weaning, breast milk plays a crucial part in shaping the microbial community composition via transmission of the milk microbiota to the infant gut[7], protection from harmful species by secreted maternal antibodies[8], and selection for certain species by milk oligosaccharides, which can be used by microorganisms as carbon sources[9]. For example, in *in vitro* competitive growth experiments, *Bifidobacterium longum* benefits from its ability to use fucosylated oligosaccharides that are present in human milk to outgrow other bacteria that are usually present in the gut microbiota, such as *Escherichia coli* and *Clostridium*

perfringens[10]. Several species of *Bacteroides* can also utilize fucosylated oligosaccharides as carbon sources[11], suggesting that their colonization may be aided by prebiotic properties of milk. Accordingly, children of mothers with nonfunctional fucosyltransferase 2, an enzyme required for fucosylation of milk oligosaccharides, display lower levels of fecal Bifidobacteria and *Bacteroides* species[12]. The importance of diet in determining the composition of the microbial community in the gut is also highlighted by the observation that transition to solid foods coincides with establishment of an adult-like microbiota.

The adult intestinal microbiota consists of hundreds to thousands of species, dominated by the Bacteroidetes and Firmicutes phyla[13]. This ecosystem is distinct from that of any other microbial habitats that have been surveyed[14], and includes many species that exist nowhere else in nature, indicating that coevolution of the host with its gut microbial symbionts (including commensals and mutualists) has generated powerful selective mechanisms. A recent study of how different microbial communities colonize gnotobiotic animals showed that deterministic mechanisms (presumably host-microorganism interactions) led to reproducible shaping of the microbiota regardless of the source of the input community[15].

The adult intestinal microbiota is also partially stable, as a core of ~40 bacterial species (accounting for 75% of the gut microbiota in terms of abundance) persists for at least a year in individuals[16]. A more extensive longitudinal study found that 60% of all bacterial strains within an individual persisted for five years[17]. During severe perturbations such as antibiotic treatment, the fecal community is depleted to a low-diversity consortium, but after a recovery period membership and relative abundance largely resemble the pretreatment state[18]. Some

species that are depleted to undetectable levels in stool are later recovered[18], suggesting that there may be reservoirs of bacterial cells that can re-seed the intestinal lumen.

The mucus layer, crypts of the colon, and appendix are examples of privileged anatomical sites, protected from the fecal stream and accessible only to certain microorganisms. In this review, we highlight relevant features of spatial heterogeneity of bacterial species and communities in the gut microbiota, and discuss the impact of microbial localization in engendering specific and stable colonization with profound implications for health and disease.

MICROBIAL COMPOSITION OF THE GUT

The mammalian lower gastrointestinal tract contains a variety of distinct microbial habitats along the small intestine, cecum, and large intestine (colon). Physiological variation along the lengths of the small intestine and colon include chemical and nutrient gradients, as well as compartmentalized host immune activity, which are known to influence bacterial community composition. For example, the small intestine is more acidic, and has higher levels of oxygen and antimicrobials than the colon (Figure 1A). Therefore, the small intestine microbial community is dominated by fast-growing facultative anaerobes that tolerate the combined effects of bile acids and antimicrobials, while still effectively competing for simple carbohydrates that are available in this region of the gastrointestinal tract. Bile acids, secreted through the bile duct at the proximal end of the small intestine, are bactericidal to certain species due to their surfactant properties and are known to broadly shape the composition of the microbiota, especially in the small intestine. For example, feeding mice excess bile acids generally stimulates the growth of Firmicutes and inhibits Bacteroidetes[19]. Additionally, the shorter transit time in the small intestine compared to colon (an order of magnitude shorter, despite the increased length of the small intestine) is thought to make bacterial adherence to tissue or mucus an important factor for persistent colonization of the small intestine.

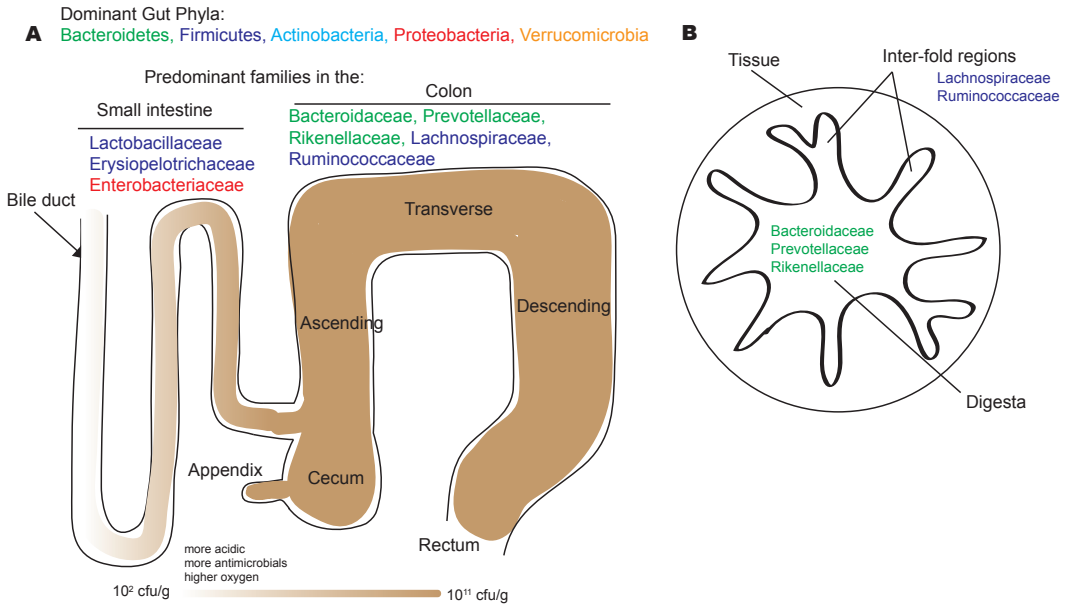


Figure 1: Microbial habitats in the human lower gastrointestinal tract. The dominant bacterial phyla in the gut are the Bacteroidetes, Firmicutes, Actinobacteria, Proteobacteria, and Verrucomicrobia. A) The dominant bacterial families of the small intestine and colon reflect physiological differences along the length of the gut. For example, a gradient of oxygen, antimicrobial peptides (including bile acids, secreted by the bile duct), and pH limits the bacterial density in the small intestinal community, whereas the colon carries high bacterial loads. In the small intestine, Lactobacillaceae and Enterobacteriaceae dominate, whereas the colon is characterized by the presence of Bacteroidaceae, Prevotellaceae, Rikenellaceae, Lachnospiraceae, and Ruminococcaceae. B) A cross-section of the colon shows the digesta – which is dominated by Bacteroidaceae, Prevotellaceae and Rikenellaceae – and the inter-fold regions of the lumen – which are dominated by Lachnospiraceae and Ruminococcaceae.

In ileostomy samples from humans, the small intestine was found to exhibit lower bacterial diversity than the colon, and was highly enriched in certain Proteobacteria and *Clostridium* species[20]. Furthermore, a metatranscriptomic analysis revealed that the expression of genes involved in central metabolism and in pathways responsible for import of simple sugars by facultative anaerobes was greatly enriched in ileal samples, compared to fecal samples[20]. In mice, Lactobacillaceae and Proteobacteria (especially Enterobacteriaceae) are enriched in the small intestine[21] (Figure 1A). Although bacteria in the small intestine are potentially competing with the host for nutrients, host-derived bile acids and antimicrobial peptides limit bacterial growth to low densities in proximal regions. Only at the distal end of the small intestine (in the terminal ileum) do bacterial densities reach saturating levels similar to those found in the large intestine (Figure 1A).

The cecum and colon cultivate the most dense and diverse communities of all body habitats. Mice, like most herbivorous mammals, have a large cecum between the small and large intestine where plant fibers are slowly digested by the microbiota. Humans have a small pouch-like cecum with an attached appendix, a thin tube-like extension (Figure 1A). In the cecum and colon, microorganisms are responsible for the breakdown of otherwise 'resistant' polysaccharides that are not metabolized during transit through the small intestine. Lower concentrations of antimicrobials, slower transit time, and a lack of available simple carbon sources facilitate the growth of fermentative polysaccharide-degrading anaerobes, notably those of the high-abundance families Bacteroidaceae and Clostridiaceae. In the mouse, the cecum is enriched in Ruminococcaceae and Lachnospiraceae, while the colon is enriched in Bacteroidaceae and Prevotellaceae[21]. Rikenellaceae are prominent in both the cecum and

colon[21]. Various host factors drive community differences over the cross-sectional axis of the gut. The entire wall of the colon folds over itself, creating compartments between folds (inter-fold regions) that are distinct from the central lumenal compartment (Figure 1B). In mouse studies that used laser capture microdissection to profile the composition of the microbial communities in discrete regions, significant differences were observed between the central lumen compartment and the inter-fold region[22,23]. Specifically, the Firmicutes families Lachnospiraceae and Ruminococcaceae were enriched between folds while the Bacteroidetes families Prevotellaceae, Bacteroidaceae, and Rikenellaceae were enriched in the digesta[22]. Relative to the digesta, the inter-fold regions are likely to contain greater amounts of mucus, which can serve as a nutrient source for certain bacteria.

Gut microhabitats: mucus and colon crypts

Throughout the human small intestine and colon, specialized epithelial cells called goblet cells secrete a mucus layer of varying thickness that partially or fully covers the epithelium depending on the region, creating a boundary between the gut lumen and host tissue (Figure 2A and 2B). The small intestine harbors a single, tightly-attached mucus layer (Figure 2A)**,** whereas in the colon, mucus is organized into two distinct layers: an outer, loose layer, and an inner, denser layer that is firmly attached to the epithelium (Figure 2B). As mentioned above, bacterial densities are much higher in the colon, compared to the small intestine, and examination of the colon by fluorescence *in situ* hybridization (FISH) has shown that the inner mucus layer appears essentially sterile next to the densely populated outer layer[24]. In addition to mucus density itself serving as a physical obstacle for microorganisms, antimicrobial molecules and oxygen secreted from the epithelium accumulate higher local

concentrations within the mucosa, especially in the small intestine, greatly restricting potential microbial inhabitants.

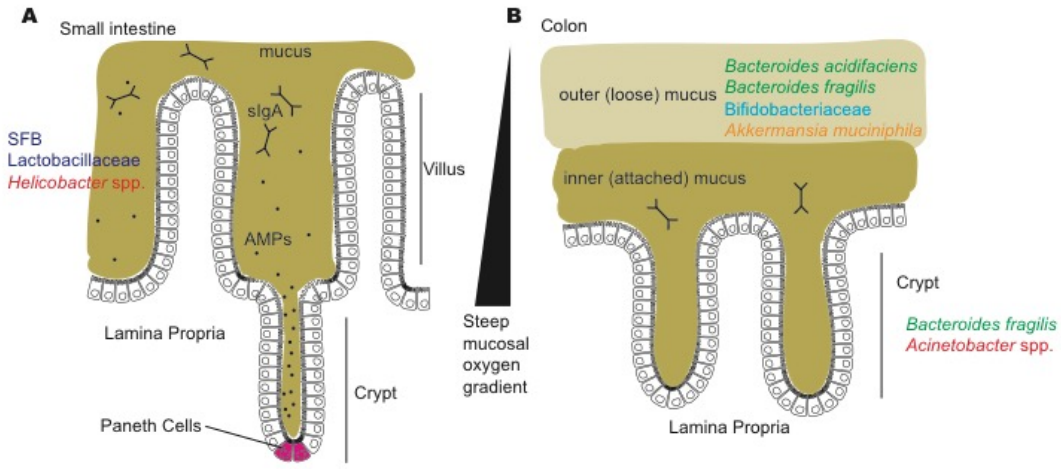


Figure 2: The mucus layers of the small intestine and colon. Several factors limit the ability of gut bacteria to access host cells, including the mucus layers in the small intestine and the colon; antimicrobial peptides in the small intestine, including those produced by Paneth cells at the base of the crypts; secreted immunoglobulin A (sIgA) in both the small intestine and colon; and a steep oxygen gradient that influences which bacteria are capable of surviving close to the epithelial surface. A) The surface of the small intestine is shaped into villi and crypts and is colonized by certain adherent species, including segmented filamentous bacteria (SFB), Lactobacillaceae and *Helicobacter* spp. B) The colon has two distinct mucus structures: the outer layer is colonized by mucin-degrading bacteria and is characterized by the presence of *Bacteroides acidifaciens*, *Bacteroides fragilis,* Bifidobacteriaceae, and *Akkermansia muciniphila* and the inner layer and crypts are penetrated at low density by a more restricted community that includes *Bacteroides fragilis* and *Acinetobacter* spp.

Mucus is continuously secreted and the outer layers are sloughed off, generating 'islands' of mucus that are carried into the fecal stream[25]. In mice, a viscosity gradient of the gel-forming mucus increases from the proximal colon (which includes the cecum and the ascending and transverse colon) to distal colonic sites (which includes the descending colon and the sigmoid colon that connects to the rectum). Accordingly, there are more mucus-associated bacteria in the proximal region[26]. Mucosal biofilm formation in the proximal colon is conserved from mammals to amphibians[27], suggesting an ancient, evolutionarily conserved origin of this region for interactions with bacteria. Therefore, the mucus layers of the gastrointestinal tract create environments that are distinct, protected habitats for specific bacterial ecosystems that thrive in proximity to host tissue.

Divergence between the mucosal and lumenal (digesta-associated) colonic communities has been observed in several mammals including humans[28], macaques[29], mice[30], cows[31], and flying squirrels[32]. More specifically, human colon biopsy and swab samples have revealed a distinct mucosal community enriched in Actinobacteria and Proteobacteria compared to the lumen community[33]. Certain species are highly enriched in colon mucus, such as the mucin-degraders *Bacteroides acidifaciens* in mice[34], *Bacteroides fragilis* in macaques[29], and *Akkermansia muciniphila* in mice and humans[34,35] (Figure 2B). Human mucosal communities in biopsy[36-38] and lavage[39] samples of the colon contain significant variability between sample locations less than one centimeter apart, suggestive of the existence of mucosal microbial populations in patches. Interestingly, an imaging study using approaches that carefully preserve the structure of feces also identified discrete patches; individual groups of bacteria were found to spatially vary in abundance from undetectable to saturating

levels[25]. This spatial niche partitioning in feces may be reflective of aggregates of interacting microorganisms, heterogeneity of nutrient availability in plant fibers, or microenvironments in mucosally-associated communities that imprint the digesta as it transits through the gut. Therefore, microbial profiling of fecal samples, which is the most common strategy employed in microbiome studies, represents an incomplete and skewed view of even the colon, which has distinct mucosal communities and spatial heterogeneity that is lost upon sample homogenization.

Some bacteria completely penetrate the mucus and are able to associate directly with the epithelium, within the crypts of the colon. Crypt-associated microorganisms were first described using electron microscopy[40,41]. Many subsequent imaging studies likely failed to observe or underestimated the number of tissue-associated bacteria because common washing and fixing methods can remove mucosal biofilms[42]. This led to the hypothesis that the mucosal surface is largely devoid of microbial colonization in healthy individuals. However, imaging studies using Carnoy's fixative, which is known to preserve the mucosal layer, found that there are bacteria in a significant fraction of colonic crypts in healthy mice[43] and humans[44]. More recent work using laser microdissection and sequencing to profile mouse crypt-associated communities revealed that the community is especially dominated by *Acinetobacter* spp. and is generally enriched for Proteobacteria capable of aerobic metabolism[23] (Figure 2B). Evasion of immune responses and particular metabolic activities are likely required for crypt occupancy by microorganisms specialized to reside in close proximity to the host. A well-characterized example of this adaptation is the ability of the human symbiont *B. fragilis* to enter crypts of the proximal colon of mice via a process requiring both modulation of the immune system[45] and utilization of specific host-derived

nutrients[46] (see below). While dogma has emerged that microorganisms contact mucosal surfaces exclusively in disease states, it appears that life-long physical associations between specific members of the microbiota and their hosts represent symbioses forged over millennia of co-evolution.

MECHANISMS RESPONSIBLE FOR GUT BIOGEOGRAPHY

Several factors influence the biogeography of bacteria within the gut, including diet, antimicrobials, mucus and adherence, and the host immune system.

Diet and nutrients

Bacterial metabolism in the gut likely contributes to the localization of particular groups of microorganisms. Because fatty acids and simple carbohydrates from food are absorbed and depleted during transit through the small intestine, sustainability of the colonic bacterial ecosystem requires growth by fermentation of complex polysaccharides, the principal carbon sources that reach the colon. Best studied in this regard are *Bacteroides* species, which are able to catabolize polysaccharides derived from the diet and from the host[47]. Compared to other gut bacteria, *Bacteroides* have the largest number and diversity of genes involved in polysaccharide degradation[48]. This extensive array of polysaccharide utilization systems is dominated by those resembling the starch utilization system (Sus), originally described in *Bacteroides thetaiotaomicron* [49]. Sus systems consist of lipid-anchored enzymes either secreted or displayed on the bacterial cell surface that can catabolize particular complex glycans into smaller oligosaccharides, which are then imported through a dedicated outer membrane transporter (Figure 3A). In the gut, *Bacteroides* species use Sus-like systems to break down dietary polysaccharides and host-derived mucin glycans[50]. The genome of *B. thetaiotaomicron* encodes 88 Sus-like systems presumably with different glycan specificities, providing remarkable metabolic flexibility[51]. Based on these findings, *Bacteroides* species, and *B. thetaiotaomicron* in particular, are sometimes referred to as "generalists," capable of occupying a variety of metabolic niches depending on the

availability of diverse polysaccharide nutrients.

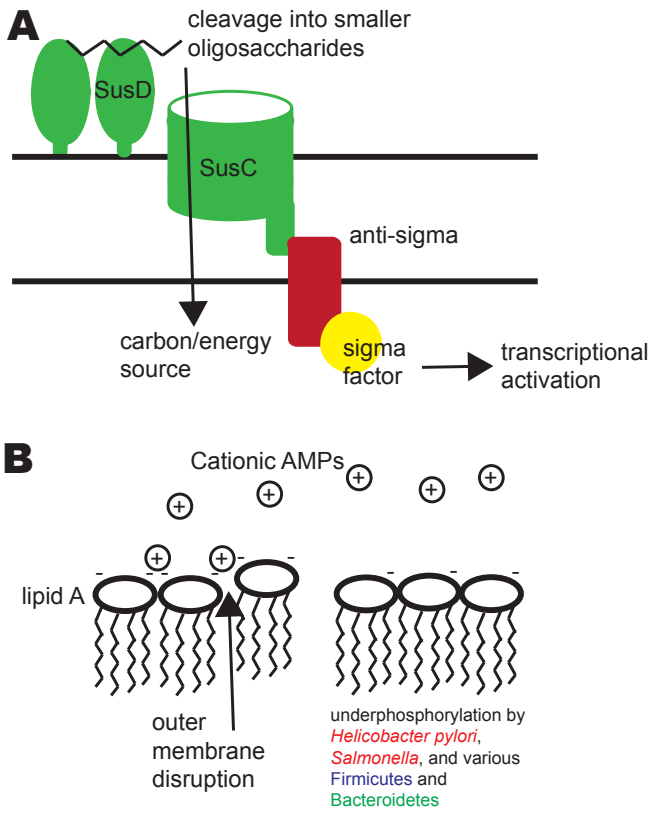


Figure 3: Bacterial colonization determinants. Several factors affect the localization of bacteria within the gastrointestinal tract, including the ability to utilize different glycans and to resist antimicrobial peptides (AMPs). A) Sus-like systems in *Bacteroides* species allow the utilization of complex polysaccharides from the diet or the host. The figure illustrates a generalized schematic of a Sus-like system. Homologues of SusD and other outer membrane lipid-anchored enzymes bind and cleave the glycans (such as starch) into smaller oligosaccharides that are then imported by the SusC-like outer membrane transporter. Interaction with the cognate glycan often leads to transmembrane signaling to activate gene regulatory mechanisms, such as a two-component system or a transmembrane anti-sigma factor which releases and activates a sigma factor. Downstream transcriptional regulation

allows *Bacteroides* species to respond to local availability of glycans. B) Cationic AMPs in the small intestine, which also pass into the colon via the fecal stream, disrupt bacterial outer membranes by interacting with negative charges on their surface. By removing phosphate groups from lipid A of lipopolysaccharide (LPS), pathogens and commensals alike–such as *Helicobacter pylori*, *Salmonella* spp., and various Firmicutes and Bacteroidetes–reduce the negative charge on their membranes and evade attack by cationic AMPs.

Diet-derived polysaccharides control microbial community composition in the lumen of the colon. Unsurprisingly, the influence of diet is readily apparent in studies that profile the fecal community. A study of humans that completely switched between plant and animal-based diets showed that the microbiome abruptly shifts with diet[52]. Over small time scales this effect is reversible, suggesting that these changes represent transient ecosystem adaptations via blooms of particular species in the lumen while the mucosal reservoir remains unchanged. Many studies of *Bacteroides* glycan metabolism have shown that restricting the polysaccharide content of the mouse diet allows selection for species (or strains) that are capable of metabolizing the complex glycans present, such as fructans[53], human milk oligosaccharides[11], fucosylated mucin glycans[54], and mannan[55]. Presumably, the variety of Sus-like systems present in the genomes of *Bacteroides* provides the metabolic plasticity to persist in the gut despite short and long-term changes in nutrient availability. However, even in terms of monosaccharide and disaccharide utilization, there is a hierarchy of bacteria that are more efficient consumers, which helps explain how diet can dramatically and rapidly change the composition of the fecal community. Importantly, the nutrient environment of the gut lumen may be in a dynamic state of flux due to potential meal-to-meal variability, especially in omnivorous mammals.

In contrast to the variable conditions in the gut lumen, mammals likely maintain a more consistent nutrient balance in the mucosa, which serves as a stable positive selection factor for certain species of bacteria. Mucus degradation and metabolism by gut microorganisms provides access to privileged spatial niches and therefore a competitive advantage over other species, both indigenous and invasive. For example, several studies have shown that the

ability to grow in an *in vitro* mucus culture is generally predictive of the ability of a bacterial species to colonize the mouse gut[56,57]. MUC2 alone is coated with over 100 different O-linked glycan structures in humans[58]. These glycans differ between mice and humans[59], and differences in complex glycan "preference" by various bacterial species are a suggested mechanism of host-specific selection of a characteristic microbiome profile. In agreement, computational models have shown that positive selection at the epithelium via the ability to metabolize specific nutrients can be a more powerful mechanism for shaping host-associated microbial communities than negative selection driven by antimicrobials[60].

A. muciniphila, a prominent symbiont in many mammals, is one of the most effective mucin degraders *in vitro*[35] and is consistently found at high abundance in the mucus layer in humans[35] and mice[34]. Consumption of mucus glycans as a carbon and energy source allows *A. muciniphila* and other mucin-degraders to colonize the gut independently of the animal's diet, providing a clear advantage to the bacteria during conditions of nutrient deprivation. Accordingly, levels of *A. muciniphila* increase in fasting Syrian hamsters[61] and hibernating ground squirrels[62]. Similarly, during intestinal inflammation in mice, the community metatranscriptome indicates increased mucin utilization with a corresponding increase in abundance of the mucin-degrading *B. acidifaciens*[63]. In gnotobiotic mice, restriction of complex polysaccharides in the diet causes the generalist *B. thetaiotaomicron* to shift its metabolism to utilize mucin glycans[50]. Further work has revealed that mutations in Sus-like systems involved in mucin glycan utilization in *B. thetaiotaomicron* cause a defect in competitive colonization and in vertical transmission of bacteria from mother to pup[64]. Therefore, the ability to utilize mucus as a carbon and energy source contributes to the ability

of some microorganisms to stably colonize the host and transfer to offspring across generations. Not surprisingly, genetic manipulation of enteric mucus production in mice changes microbial community composition[54,65]. In turn, gut bacteria affect transcription of mucin-encoding genes in mice[66]. Overall, development of a healthy mucosa is a collaborative, bi-directional event between the host and the gut microbiota, creating an environment that allows the specific members to establish persistent colonization via utilization of host-derived glycans.

In some cases, the ability of a bacterium to colonize the gut may be determined by its ability to utilize a specific, yet limiting, nutrient. Bacterial species-specific carbohydrate utilization systems termed commensal colonization factors (CCFs) have been identified in *B. fragilis* and *Bacteroides vulgatus*, and allow these bacteria to colonize saturable nutrient niches[46]. This discovery was made based on the observation that gnotobiotic mice colonized with a specific *Bacteroides* species are resistant to colonization by the same species, but not colonization by closely related species. A genetic screen revealed that a set of genes encoding the CCF system was required for this intra-species colonization resistance phenotype (Box 1), suggesting that CCFs are responsible for defining the species-specific niche. Accordingly, when the *ccf* genes from *B. fragilis* were expressed in *B. vulgatus*, the resulting hybrid strain gained the ability to colonize an alternate niche. The CCF system was also required for penetration of *B. fragilis* into the crypts of the colon and long-term resilience to intestinal perturbations such as antibiotic treatment and gastroenteritis. Collectively, these data suggest that while metabolic flexibility allows bacterial adaptation in the lumen environment, the

occupation of a narrowly-defined, tissue-associated niche is likely very important for stable colonization by some bacteria.

Box 1: Colonization resistance.

One of the benefits afforded by the microbiota to the host is colonization resistance to pathogens. Invasive species of bacteria are inhibited from colonizing the gut because they are unable to displace indigenous species that have gained a strong foothold. After years of studying colonization resistance against pathogens in gnotobiotic animals in the 1960's and 70's, Rolf Freter theorized that the ability of a bacterial species to colonize the gut is determined by its ability to utilize a specific, limiting nutrient[135]. This notion has been well supported by studies showing that colonization resistance to pathogens is mediated by the availability of nutrient niches in the cases of *Escherichia coli*[136] and *Clostridium difficile*[137]. But Freter's hypothesis reached even further, suggesting that the relative amounts of limiting nutrients could dictate the abundance of each species in the indigenous community. Correspondingly, the variety of host-derived growth substrates could explain the stable diversity of the gut microbiota if individual species have evolved to specialize in the uptake and metabolism of specific, limiting nutrients, such as in the case of *Bacteroides fragilis*[46]. The concept of spatial niche partitioning being governed by host production of specific and scarce nutrient resources is attractive, and may help explain both long-term persistence and resilience of the microbiome, as well as colonization resistance to pathogens.

Antimicrobials

Specialized epithelial immune cells called Paneth cells reside at the base of the crypts of the small intestine, secreting an array of antimicrobials that restrict the growth of bacteria that are found near the mucosal surface[4]. Many of these molecules are cationic antimicrobial peptides that interact with and disrupt negatively charged bacterial membranes (Figure 3B). Modifications to lipid A, a major component of the outer membrane of gram-negative bacteria, are known to confer resistance to cationic antimicrobial peptides in several pathogens[67]. Interestingly, underphosphorylation of this lipid portion of LPS, a modification shared with the pathobiont *Helicobacter pylori*, was found to be important for resilient colonization by *B. thetaiotaomicron* during inflammation[68] (Figure 3B).

The concentration of a variety of antimicrobials is higher toward the proximal end of the small intestine, creating a gradient that leads to a higher abundance and diversity of bacteria in distal locations (Figure 1A). For example, the lectin RegIIIγ is bactericidal to gram-positive bacteria that dominate the small intestine because it binds to and disrupts their exposed peptidoglycan layer. RegIIIγ is required to prevent massive infiltration of the mucosa and microbial invasion of the tissue[69]. In addition to RegIIIγ, the innate immune system deploys many other antimicrobials (such as alpha-defensins from Paneth cells and beta-defensins from neutrophils) with differing specificities to limit access to the epithelium[70], and resistance to these host-derived antimicrobial peptides is a general feature of many indigenous gut species of Firmicutes and Bacteroidetes[68].

In addition to these antimicrobials, gut bacteria, which are largely anaerobic, must contend

with reactive oxygen species produced by aerobic host metabolism. Rapid dilution and consumption of oxygen secreted from the host tissue generates a gradient of oxygen that decreases in concentration from tissue to lumen (Figure 2). Accordingly, the mucosal community is enriched in genes required for resistance to reactive oxygen species[33]. Notably, although all *Bacteroides* species are classified as obligate anaerobes, *B. fragilis* can use oxygen as a terminal electron acceptor at nanomolar concentrations[71]. *B. fragilis* and tissue-associated, microaerophilic Lactobacillaceae express catalase, superoxide dismutase, and other enzymes to inactivate reactive oxygen species[72]. Altogether, these mechanisms restrict access to the epithelium to a subset of bacterial species that not only can utilize nutrients found only at the tissue boundary, but can survive host antimicrobial strategies as well.

Mucus and adhesion

To access the epithelium, pathogens and commensals alike must contend with the mucus barrier and the immune system (Figure 4). Secreted MUC2 forms peptide crosslinks to create a viscous gel-like substance[73], serving as a barrier and host defense mechanism[74]. In mice lacking MUC2, the crypts of the colon are filled with bacteria and the tissue is covered in biofilms[24], indicating that the gel-forming mucus is the primary barrier to tissue association by the microbiota at large. However, certain bacteria are able to penetrate the mucus by swimming or eating their way through.

In the gut, bacterial motility is generally restricted due to the immunogenicity of flagellin, which is a ligand for Toll-like Receptor 5 (TLR5)[75], and the viscosity of mucus limits the effectiveness of swimming (Figure 4). Still, the enteric pathogen *Salmonella enterica*

subspecies *enterica serovar* Typhimurium depends on flagella and chemotaxis to penetrate the mucus layer and to reach host tissue[76]. *E. coli* and close relative *Shigella flexneri* opt for an alternative strategy of secreting a mucin-binding serine protease, Pic, which rapidly digests mucus (Figure 4). Interestingly, Pic also causes hypersecretion of mucus, which may interfere with the ability of indigenous bacteria to compete with the pathogen[77]. Similarly, another family of mucus-degrading proteins, M60-like peptidases, are conserved in pathogens and commensal mucosal bacteria from the Proteobacteria, Firmicutes, Bacteroidetes, and other phyla[78]. In enterotoxigenic *E. coli*, an M60-like peptidase was required for association with villi in the mouse small intestine[79].

In addition to the ability to penetrate the mucus layer, bacterial adhesion to the epithelium also influences the microbial composition of the gut, especially in the small intestine (Figure 2A). Species of *Helicobacter* adhere to and colonize the the stomach and small intestine tissue via adherence to epithelial surface glycans[80]. Further downstream in the small intestine, segmented filamentous bacteria (SFB) adhere intimately to the epithelial surface, as first described in imaging studies of mice[81]. Host-specific strains of SFB appear to be present in many mammals, including humans[82]. These bacteria were only recently cultured *in vitro* using tissue-cultured enterocytes as a platform to support their growth, reinforcing the idea that they are obligate symbionts with the mammalian gut tissue[83]. Their mechanism of attachment is still a mystery, though the attachment site is marked by accumulation of actin and leaves a visible indentation on the surface of the epithelial cell following removal of the filaments[83]. By virtue of intimate host association, SFB shape the host immune response[84] and impact autoimmune disease in mouse models[85,86].

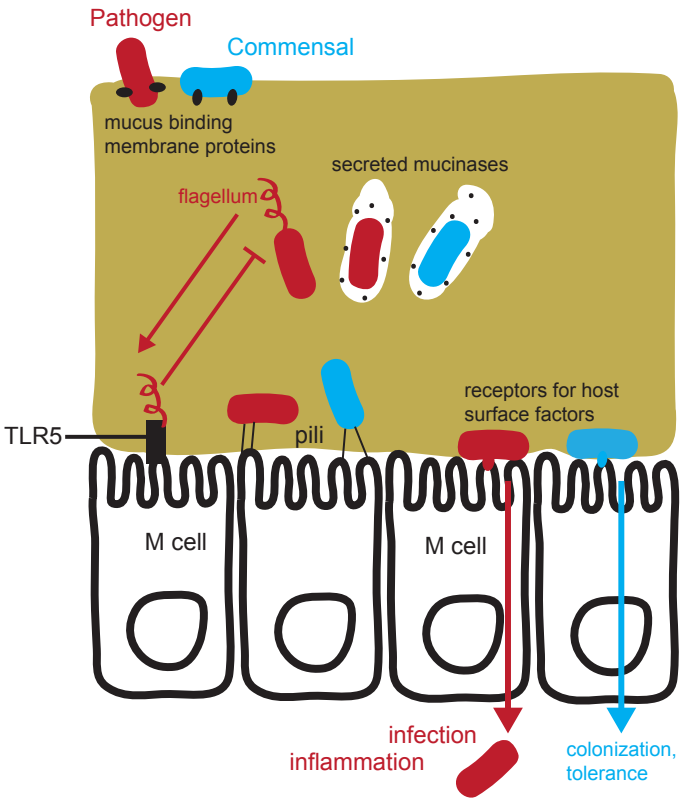


Figure 4: Bacterial access to the epithelium. Both bacterial pathogens (red) and commensals (or mutualists; blue) have the ability to cross the mucus layer and access the gut epithelium. Lectins and other mucus-binding proteins facilitate initial interactions with the mucus layer. Mucinases and proteases are used to degrade mucus for bacteria to "eat" their way through, while some pathogens such as *Salmonella* spp. use flagella to swim through the viscous mucus. TLR5 sensing of flagellin effectively leads to inhibition of flagellar biosynthesis for most bacteria in the gut. Adherence to the tissue is achieved by both commensals and pathogens through pili, lectins, and other outer-membrane proteins that target ligands on the epithelial cell surface. Adherence facilitates gut colonization for both commensals and pathogens, and also allows tissue invasion by pathogenic bacteria. Microfold cells (M cells) are specialized immune sentinel epithelial cells that detect gut bacteria and are also exploited by many pathogens as a means of translocation across the epithelium.

The molecular mechanisms underlying how microorganisms attach to host tissue have been well-studied in pathogens[87]. Although all of these features were initially discovered and described in pathogens, they are found in many commensal bacteria. Bacteria adhere to mucus and epithelial surfaces by deploying outer membrane proteins, capsules, lectins, adhesins, and fimbriae (attachment pili) (Figure 4). For example, the non-invasive pathogen *Vibrio cholerae* forms a layer of adhered cells on the wall of the small intestine using toxin-coregulated pili (TCP) [88]. *V. cholerae* also binds mucins using an outer membrane N-acetyl-D-glucosamine (GlcNAc)-binding protein, which may also facilitate penetration of the mucus and access to the epithelium[89]. Without attachment, these naturally plankton-associated marine bacteria are unable to colonize the gut, and thus are avirulent. *E. coli* possesses a great number of lectins with diverse sugar specificities allowing it to bind mucins as well as other glycoproteins and extracellular matrix components of epithelial cells[90]. Invasive pathogens also depend on adherence factors as a preceding step to penetration and infection of the tissue. *Listeria monocytogenes* expresses a surface protein, internalin A, which binds epithelial E-cadherin (a host cell adhesion protein) as a first step before exploiting actin to induce phagocytosis[91]. Studies of *S.* Typhimurium also reveal a critical role of apical surface attachment in inducing neutrophil-mediated inflammation, which appears to paradoxically promote infection[92] by providing a competitive advantage for the pathogen over the resident microbiota[93].

Beneficial microorganisms also adhere to particular regions of the epithelium and can serve to exclude adherent pathogens by occupying limited binding sites, although little is known about the underlying mechanisms or functions of this process (Box 1). Early imaging studies

revealed that *Lactobacillus* spp. that form adherent layers on the epithelium in the rat stomach prevent yeast[94] and staphylococcal[95] adherence to the epithelium. Members of the family Lactobacillaceae (such as *Lactobacillus* and *Lactoccocus*) that colonize the small intestine and stomach have become model systems for studying adhesion by commensals, with exopolysaccharides, pili, and cell wall-anchored proteins found to be involved in interacting with mucus, extracellular matrix proteins, and other molecular targets on the epithelial cell surface[96]. Notably, cell wall-anchored mucus-binding proteins (MUBs) unique to lactobacilli are known to be involved in both adherence and aggregation[97]. Strain-specific diversity in adherence and aggregation factors underlies the host specificity of *Lactobacillis reuteri*, indicating that tissue-associated biofilm formation is fundamental to colonization by this species[98]. Other means of attachment involve mechanisms conserved with pathogens, such as adhesive pili in *Lactobacillus rhamnosus* that bind mucus[99]. Analogous mechanisms can be found in unrelated species such as *Bifidobacterium bifidum*, which uses pili to bind extracellular matrix proteins, contributing to bacterial aggregation[100].

Collectively, these studies suggest that interactions with mucus and adherence to intestinal epithelial cells appear to be adaptations used by pathogens during infection, as well as strategies employed by commensals during persistent colonization (Figure 4).

Immunomodulation

In order to persist in the gut, non-pathogenic bacteria that intimately associate with host tissue must be tolerated by the immune system. The mucosa is inundated with large amounts of secreted immunoglobulin A (sIgA) to monitor the microbiota. Many bacteria in the gut are

coated in sIgA, and this subpopulation broadly resembles the mucosal population[101]. Certain adherent species such as *Helicobacter* spp. and SFB are especially highly coated in sIgA[102]. Binding of sIgA to bacteria may contribute to mucosal biofilm formation, which serves as a barrier to pathogen adherence[103]. Gnotobiotic studies with *Rag1* knockout mice (which effectively have no adaptive immune system) showed that experimental coating of *B. thetaiotaomicron* with sIgA reduces microbial fitness but also leads to reduced inflammatory signaling and changes to bacterial gene expression[5,104]. Through these mechanisms, sIgA mediates homeostasis between the host and the microbiota, as well as potential pathogens at mucosal surfaces. Furthermore, natural antibodies have evolved to recognize bacterial capsular polysaccharides; while largely studied in the context of infectious agents, such antibodies may also represent an evolutionarily conserved strategy used by the host to sense indigenous bacterial species. However, examples on how the immune system can dependably distinguish between harmful and beneficial microorganisms remain limited.

An alternate view is that the immune system is not "hard-wired" to discriminate between various classes of microorganisms, but rather that specific species have adapted to promote their own immunologic tolerance. A few examples of active, species-specific immunomodulation by beneficial microorganisms suggest that some bacteria display signals that ensure their own tolerance by the immune system (Figure 5). *B. fragilis* is one of the best-understood gut bacteria in terms of immunomodulation. A component of its capsule, polysaccharide A (PSA), signals through an antigen-presenting cell intermediary to stimulate production of IL-10 by an anti-inflammatory subset of immune cells, regulatory T cells[105], contributing to the ability of *B. fragilis* to enter the mucus layer of the colon[45] (Figure 5).

Surface fucosylation of the bacterial capsule also contributes to *B. fragilis* fitness in the gut, perhaps by mimicking the host cell surface to elicit a tolerogenic immune response[106]. Through these specific molecular signals, *B. fragilis* induces an anti-inflammatory immune profile that facilitates its own colonization. Similarly, exopolysaccharides of *Bifidobacterium breve* promote immune tolerance by decreasing the production of pro-inflammatory cytokines and preventing a B-cell response[107] (Figure 5). Through a less well-defined mechanism, *B. breve* also induces IL-10 production by regulatory T cells[108]. Notably, both *B. fragilis* and species of *Bifidobacterium* are known to closely associate with the host, which may necessitate immunomodulation to prevent an inflammatory reaction against these bacteria. Similarly, adherent SFB stimulate the development of a subset of T helper cells, Th17 cells, which are required for normal SFB colonization and also confer resistance to the pathogen *Citrobacter rodentium*[84] (Figure 5; Box 1). Clostridia are able to induce regulatory T cells, but a population of many species is much more effective than single isolates or combinations of a few, suggesting this is a combined effect of production of different metabolites, such as short chain fatty acids, by different species (see below)[109] (Figure 5). Similarly, a defined community of eight mouse gut bacterial species (including several members of families Clostridiaceae and Lactobacillaceae) referred to as the altered Schaedler's flora[110], was also shown to modulate immune responses mediated by regulatory T cells. Therefore, it is likely that many other beneficial microorganisms have co-evolved with the immune system to facilitate stable long-term colonization.

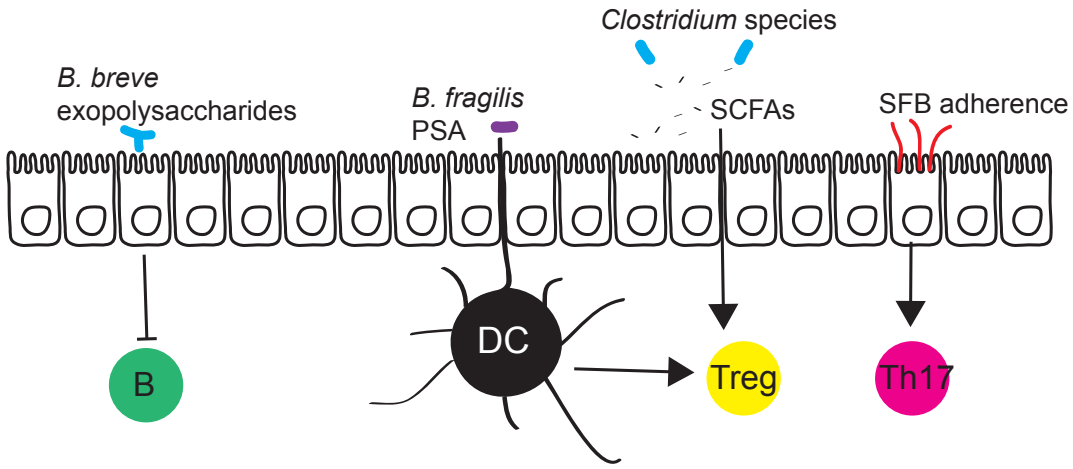


Figure 5: Immunomodulation by commensal gut bacteria. Commensal gut bacteria induce immunomodulation via interaction with epithelial cells, antigen presenting cells (such as dendritic cells (DCs)), and via production of signaling metabolites. The exopolysaccharides of adherent *Bifidobacterium breve* reduce the production of inflammatory cytokines to dampen B cell responses. The capsular polysaccharide PSA of *Bacteroides fragilis* and short-chain fatty acids (SCFAs) produced by many species of *Clostridia* (and other genera) stimulate the production of the anti-inflammatory interleukin-10 (IL-10) by regulatory T cells. Segmented filamentous bacteria (SFB) intercalate between microvilli of epithelial cells and stimulate the development of Th17 cells, which are important for mucosal immunity to extracellular pathogens.

Several non-specific signals in the gut also promote tolerance towards beneficial microorganisms. Short chain fatty acids such as butyrate, propionate, and acetate are the end-products of anaerobic fermentation of sugars, which is the dominant metabolism in the colon. The development of regulatory T cells is stimulated by these molecules[111,112], which could be a more general way for the immune system to recognize beneficial bacteria or to assess the total fermentative productivity of the community. Mucus is another non-specific anti-inflammatory signal. When MUC2 is taken up by dendritic cells in mice, it inhibits the expression of pro-inflammatory signals[113], raising the possibility that indigenous mucin-degraders may induce host tolerance by being co-presented with mucus. Pathogens also have an arsenal of anti-inflammatory mechanisms to suppress the immune system to promote infection[114]. Particularly perplexing is the fact that features traditionally regarded as virulence factors in pathogens, such as capsular polysaccharides and pili, are also colonization factors in beneficial bacteria. Our notion of the defining characteristics of pathogens is likely clouded by a historical under-appreciation of similar colonization strategies used by beneficial species (Figure 4). It is not surprising that similar mechanisms of host association (mucus penetration, adherence, immune modulation) are used by pathogenic and commensal bacteria alike; however, a key distinguishing feature is that commensals have either not evolved traits resembling traditional virulence factors, or have evolved additional features or modifications to offset the host-response to such factors. This perspective suggests that commensal bacteria may have reached an immunologic and metabolic 'truce' with their host, enabling persistent establishment of defined microbial habitats and elaborate microbial biogeographies.

MICRO-BIOGEOGRAPHY IN HEALTH AND DISEASE

Microhabitats in the gut are likely to contribute to the development and stability of microbial communities because spatially stratified niches facilitate greater diversity. In mouse pups, the fecal microbiota is initially dominated by Proteobacteria, a signature of the small intestine, but switches following weaning to Clostridia and *Bacteroides*, which are characteristic of the adult colon[2]. The sequential development of the microbiota thus may occur from proximal to distal compartments, which makes sense as dispersal in the gut is largely unidirectional along the fecal stream. Because of this restriction on dispersal, depletion of beneficial species, especially in the colon, could be catastrophic without a mechanism to replenish the community. Therefore, protected regions that are less susceptible to variable conditions in the gut may serve as reservoirs of bacterial cells that can seed growth in the lumen, possibly after an environmental insult (Figure 6). In the case of *B. fragilis*, mutants that are unable to colonize the crypts of the colon are less resilient to intestinal perturbations such as antibiotic treatment and enteric infection[46]. This is also a proposed function of the human appendix, which has a mucus and bacteria-filled lumen contiguous with the cecum[115]. The appendix is protected from the fecal stream, yet harbors a diverse microbial community and a contingent of specialized immune cells. The appendix is also phylogenetically widespread and evolved independently at least twice, providing strong evidence that this is not a vestigial structure as once believed[27]. In the rabbit appendix, indigenous bacteria coordinate the education of B and T-lymphocytes, suggesting that these tissue-associated niches are venues for immunomodulation[116]. Microhabitats such as crypts, mucus, and the appendix may be crucial to facilitate immune homeostasis, protect microbial inhabitants from competitors, and re-populate the gut following catastrophic perturbations

that alter bacterial community structure or deplete certain species from the lumen.

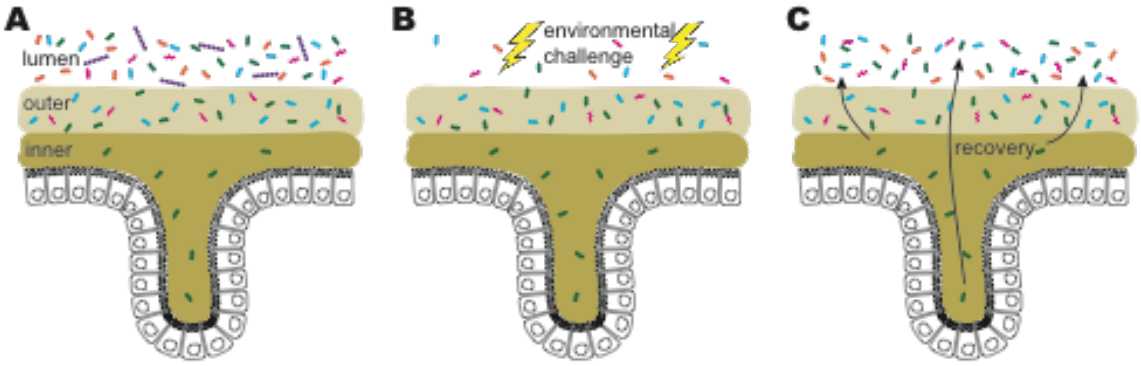


Figure 6: Gut microhabitats as reservoirs of bacterial diversity. Specific niches such as crypts, the inner mucus, and the appendix may be crucial to facilitate immune homeostasis, protect microbial inhabitants from competitors, and re-populate the gut following perturbations that alter bacterial community structure or deplete certain species from the lumen. A) A subset of species is able to penetrate the inner mucus layer and enter crypt spaces. B) Environmental challenges such as diet perturbations, abnormalities in gastrointestinal motility, and antibiotic consumption massively alter the lumen community. However, the more stable mucosal environment and crypts protect important bacterial species. C) The crypts and mucosa serve as reservoirs to repopulate the lumen.

Micro-biogeography alterations during disease

The adverse effects of altered composition of the healthy microbiota, known as dysbiosis, on host health have long been appreciated. Increasing clinical evidence links dysbiosis with various immune, metabolic, and neurological disorders in both intestinal and extra-intestinal sites. For example, inflammatory bowel disease (IBD) is associated with changes in the gut microbiota, characterized by decreased abundance of *Clostridia*[117-119] and overall reduction in bacterial diversity[118-120]. Childhood asthma is correlated with low intestinal microbial diversity during the first month of life[121]. The obesity-associated microbiota is characterized by reduced microbial diversity and, in some studies, an increased *Firmicutes:Bacteroidetes* ratio[122]. In recent years, the role of gut dysbiosis in the pathogenesis of chronic liver diseases [123,124], colorectal cancer (CRC) [125,126], and even neuropsychiatric dysfunctions[127] has been explored in animal models and humans. For clinical applications, profiling of the fecal microbiota has been widely used as a surrogate for the gastrointestinal bacterial community due to non-invasive and straightforward sample collection; however, fecal populations may be less informative than mucosal biopsies in defining disease-associated dysbiosis[128], a notion that requires additional experimental support. Below, we detail two examples that illustrate the importance of micro-biogeography alterations of the gut microbiota during disease: IBD and hepatic encephalopathy.

IBD is characterized by inflammation of the gastrointestinal tract resulting in pain, vomiting, diarrhea, and other complications including severe weight loss and behavioral changes. Generally, IBD is categorized into two syndromes, Crohn's disease, which may involve inflammation throughout the gastrointestinal tract (mouth to anus), and ulcerative colitis,

where pathology is restricted to the large intestine. For over a decade, studies have attempted to define a pattern of dysbiosis associated with IBD and yielded inconsistent and sometimes contradicting results[129]. Studies that focused on fecal microbiota reported wide inter-individual differences in composition with overall microbial diversity reduced[118] in Crohn's disease patients compared to healthy controls. However, a study in which human gut microbiota were assessed for the ability to drive colitis pathology in mice found that bacteria contributing to the disease are highly coated in sIgA[102], suggesting that the mucosal or tissue-associated population is most relevant. Human studies based on biopsy samples elucidated several consistent features in line with this hypothesis: patients had increased concentration of bacteria on the mucosal surface[130]; decreased microbial diversity; [119,120]; decreased abundance of *Clostridium* species [117]; and increased number of *Enterobacteriaceae* (especially adherent, invasive *E. coli*) in ileal mucosa[131]. Most recently, both the lumen- and mucosa-associated microbiota were profiled in a large cohort of new-onset, treatment naïve pediatric patients with Crohn's disease and non-IBD controls. Analysis of the mucosal microbiota revealed a significant drop in species richness, an increase in *Enterobacteriaceae,* a decrease in *Clostridiales*, and significant changes in several other previously unidentified taxa. Importantly, these dysbiotic signatures were lost when stool samples were examined[128]. Intriguingly, a laser-capture microdissection study of colon crypt mucus in patients with ulcerative colitis found that they had lower levels of crypt-associated bacteria[132]. Overall, these studies highlight that distinguishing between fecal and mucosal microbial communities is particularly important for finding a reproducible microbial signature of IBD. Moving from correlations to a potential causal etiology of the microbiota for IBD and other disorders will require further study of mucosal communities, focusing on the interactions between the host

and microbiota.

Biogeographical changes in the gut microbiota may also influence liver function. Hepatic encephalopathy is a neuropsychiatric complication of cirrhosis and direct sequelae of gut dysbiosis. As a result of impaired liver function and the presence of porto-systemic shunts (bypass of the liver by the circulatory system), toxic metabolites produced by the gut microbiota evade liver catabolism and cross the blood-brain barrier, leading to cerebral toxicity[123]. Interestingly, a comparison of the fecal microbiota of cirrhotic patients with and without hepatic encephalopathy showed minimal differences, whereas analyzing the microbiota composition of the colonic mucosa revealed significant changes, including lower *Roseburia* and higher *Enterococcus*, *Veillonella*, *Megasphaera*, *Burkholderia*, and *Bifidobacterium* in cirrhotic patients with hepatic encephalopathy[133]. The bacterial genera over-represented in the mucosa of patients with hepatic encephalopathy (*Megaspheara*, *Veillonella*, *Burkholderia*, and *Bifidobacterium*) were also correlated with poor cognition, higher inflammation, and higher clinical severity score. In summary, mucosal dysbiosis in the gut, but not in the composition of the fecal community, significantly correlates with the severity of chronic liver disease phenotypes, including hepatic encephalopathy.

CONCLUSION

We have highlighted evidence that the microbiota is biogeographically stratified within the gut on different spatial scales and axes. Progress towards a functional understanding of the microbiota necessitates increased attention to microhabitats within the gut ecosystem, and to spatial relationships between microorganisms and between microorganisms and the host.

Fecal community profiling enabled by next-generation sequencing provides a valuable picture of the diversity, specificity, stability, and developmental dynamics of the gut microbiota, but focusing on measurements of abundance in feces neglects the importance of mucus and tissue-associated organisms and cannot account for spatial distributions. Similarly, studies in gnotobiotic animals allow a reductionist approach to studying host-microorganism interactions akin to methods traditionally employed by microbiologists studying pathogens, but this simplified methodology is likely to miss important contributions from interspecies interactions. The functional study of gut microbial ecology using "meta-omics" techniques enables one to account for the behaviors of the community as a whole, but attributing functions to particular microbial members remains a challenge in community-level ecology. Therefore, testing unifying hypotheses using both reductionist and ecological approaches will be essential to our understanding of the microbiota and its biological functions.

More than half a century ago, in "Microorganisms Indigenous to Man," the microbiologist Theodor Rosebury lamented on the lack of a general theory for influences that control composition of the microbiota, the roles of individual members, and functions that affect the host[134]. With the true complexity of the problem revealed recently by sequencing advances, research is only now in a position to fulfill Rosebury's call for a general theory. Rolf Freter's nutrient niche hypothesis[135], which states that limiting nutrients control the population level of species that are particularly adept at utilizing them, provides a metabolic foundation to explain some of the nascent observations in the field. But when Freter proposed his ideas, we were unaware of the role of immunomodulation by non-pathogens, which requires access

to the tissue. Based on evidence outlined in this review, we propose that the host presents limiting nutrients as well as attachment sites in privileged locations. Furthermore, the immune system has an active role in allowing only beneficial species to access these locations during homeostasis. Selection for particular species close to the epithelium creates protected, stable reservoirs for microorganisms to persist in the face of rapidly changing conditions in the gut lumen. Through localized, immune-facilitated, and adherence-dependent nutrient selection, the host maintains stability of a diverse community of microbial symbionts.

ACKNOWLEDGEMENTS

Thanks to Elaine Hsiao, Brittany Needham and Timothy Sampson for critical comments on the manuscript. G.P.D. is supported by an NSF Graduate Research Fellowship (DGE-1144469). Work in the Mazmanian laboratory is supported by funding from the National Institutes of Health (GM099535, DK078938, MH100556), the Emerald Foundation, and the Simons Foundation.

REFERENCES

1. Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA* **107,** 11971–11975 (2010).
2. Hasegawa, M. *et al.* Transitions in oral and intestinal microflora composition and innate immune receptor-dependent stimulation during mouse development. *Infect. Immun.* **78,** 639–650 (2010).
3. La Rosa, P. S. *et al.* Patterned progression of bacterial populations in the premature infant gut. *Proc Natl Acad Sci USA* **111,** 12522–12527 (2014).
4. Bevins, C. L. & Salzman, N. H. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nat. Rev. Microbiol.* **9,** 356–368 (2011).
5. Peterson, D. A., McNulty, N. P., Guruge, J. L. & Gordon, J. I. IgA response to symbiotic bacteria as a mediator of gut homeostasis. *Cell Host Microbe* **2,** 328–339 (2007).
6. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9,** 313–323 (2009).
7. Fernández, L. *et al.* The human milk microbiota: origin and potential roles in health and disease. *Pharmacol. Res.* **69,** 1–10 (2013).
8. Rogier, E. W. *et al.* Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression. *Proc Natl Acad Sci USA* **111,** 3074–3079 (2014).
9. Yu, Z.-T., Chen, C. & Newburg, D. S. Utilization of major fucosylated and sialylated human milk oligosaccharides by isolated human gut microbes.

Glycobiology (2013). doi:10.1093/glycob/cwt065

10. Yu, Z.-T. *et al.* The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota. *Glycobiology* **23,** 169–177 (2013).

11. Marcobal, A. *et al.* Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **10,** 507–514 (2011).

12. Lewis, Z. T. *et al.* Maternal fucosyltransferase 2 status affects the gut bifidobacterial communities of breastfed infants. *Microbiome* **3,** 425 (2015).

13. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).

14. Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* **6,** 776–788 (2008).

15. Seedorf, H. *et al.* Bacteria from diverse habitats colonize and compete in the mouse gut. *Cell* **159,** 253–266 (2014).

16. Martínez, I., Muller, C. E. & Walter, J. Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. *PLoS ONE* **8,** e69621 (2013).

17. Faith, J. J. *et al.* The long-term stability of the human gut microbiota. *Science* **341,** 1237439 (2013).

18. 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 Suppl 1,** 4554–4561 (2011).

19. Islam, K. B. M. S. *et al.* Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology* **141,** 1773–1781 (2011).

20. Zoetendal, E. G. *et al.* The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J* **6,** 1415–1426 (2012).

21. Gu, S. *et al.* Bacterial Community Mapping of the Mouse Gastrointestinal Tract. *PLoS ONE* **8,** e74957 (2013).

22. Nava, G. M., Friedrichsen, H. J. & Stappenbeck, T. S. Spatial organization of intestinal microbiota in the mouse ascending colon. *ISME J* **5,** 627–638 (2011).

23. Pédron, T. *et al.* A crypt-specific core microbiota resides in the mouse colon. *MBio* **3,** (2012).

24. Johansson, M. E. V. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc Natl Acad Sci USA* **105,** 15064–15069 (2008).

25. Swidsinski, A., Loening-Baucke, V., Verstraelen, H., Osowska, S. & Doerffel, Y. Biostructure of fecal microbiota in healthy subjects and patients with chronic idiopathic diarrhea. *Gastroenterology* **135,** 568–579 (2008).

26. Swidsinski, A. *et al.* Viscosity gradient within the mucus layer determines the mucosal barrier function and the spatial organization of the intestinal microbiota. *Inflamm. Bowel Dis.* **13,** 963–970 (2007).

27. SMITH, H. F. *et al.* Comparative anatomy and phylogenetic distribution of the mammalian cecal appendix. *J. Evol. Biol.* **22,** 1984–1999 (2009).

28. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308,** 1635–1638 (2005).

29. Yasuda, K. *et al.* Biogeography of the intestinal mucosal and lumenal microbiome in the rhesus macaque. *Cell Host Microbe* **17,** 385–391 (2015).

30. Wang, Y. *et al.* Regional mucosa-associated microbiota determine physiological expression of TLR2 and TLR4 in murine colon. *PLoS ONE* **5,** e13607 (2010).

31. Malmuthuge, N., Griebel, P. J. & Guan, L. L. Taxonomic identification of commensal bacteria associated with the mucosa and digesta throughout the gastrointestinal tracts of preweaned calves. *Applied and Environmental Microbiology* **80,** 2021–2028 (2014).

32. Lu, H.-P. *et al.* Spatial heterogeneity of gut microbiota reveals multiple bacterial communities with distinct characteristics. *Sci Rep* **4,** 6185 (2014).

33. Albenberg, L. *et al.* Correlation Between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota in Humans and Mice. *Gastroenterology* (2014). doi:10.1053/j.gastro.2014.07.020

34. Berry, D. *et al.* Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. *Proc Natl Acad Sci USA* (2013). doi:10.1073/pnas.1219247110

35. Png, C. W. *et al.* Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am. J. Gastroenterol.* **105,** 2420–2428 (2010).

36. Hong, P.-Y., Croix, J. A., Greenberg, E., Gaskins, H. R. & Mackie, R. I. Pyrosequencing-based analysis of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-heterogeneity. *PLoS ONE* **6,**

e25042 (2011).

37. Zhang, Z. *et al.* Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J* (2013). doi:10.1038/ismej.2013.185

38. Nava, G. M., Carbonero, F., Croix, J. A., Greenberg, E. & Gaskins, H. R. Abundance and diversity of mucosa-associated hydrogenotrophic microbes in the healthy human colon. *ISME J* **6,** 57–70 (2012).

39. Tong, M. *et al.* A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. *PLoS ONE* **8,** e80702 (2013).

40. Davis, C. P., Mulcahy, D., Takeuchi, A. & Savage, D. C. Location and description of spiral-shaped microorganisms in the normal rat cecum. *Infect. Immun.* **6,** 184–192 (1972).

41. Savage, D. C. & Blumershine, R. V. Surface-surface associations in microbial communities populating epithelial habitats in the murine gastrointestinal ecosystem: scanning electron microscopy. *Infect. Immun.* **10,** 240–250 (1974).

42. Palestrant, D. *et al.* Microbial biofilms in the gut: visualization by electron microscopy and by acridine orange staining. *Ultrastruct Pathol* **28,** 23–27 (2004).

43. Swidsinski, A., Loening-Baucke, V., Lochs, H. & Hale, L.-P. Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence in situ hybridization study in mice. *World J. Gastroenterol.* **11,** 1131–1140 (2005).

44. Swidsinski, A., Weber, J., Loening-Baucke, V., Hale, L.-P. & Lochs, H. Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *J. Clin. Microbiol.* **43,** 3380–3389 (2005).

45. Round, J. L. *et al.* The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332,** 974–977 (2011).

46. Lee, S. M. *et al.* Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501,** 426–429 (2013).

47. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10,** 323–335 (2012).

48. Kaoutari, A. E., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11,** 497–504 (2013).

49. Reeves, A. R., Wang, G. R. & Salyers, A. A. Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. *Journal of Bacteriology* **179,** 643–649 (1997).

50. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307,** 1955–1959 (2005).

51. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284,** 24673–24677 (2009).

52. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505,** 559–563 (2013).

53. Sonnenburg, E. D. *et al.* Specificity of polysaccharide use in intestinal

bacteroides species determines diet-induced microbiota alterations. *Cell* **141,** 1241–1252 (2010).

54. Kashyap, P. C. *et al.* Genetically dictated change in host mucuscarbohydrate landscape exerts a diet-dependenteffect on the gut microbiota. *Proc Natl Acad Sci USA* (2013). doi:10.1073/pnas.1306070110

55. Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517,** 165–169 (2015).

56. Wadolkowski, E. A., Laux, D. C. & Cohen, P. S. Colonization of the streptomycin-treated mouse large intestine by a human fecal Escherichia coli strain: role of growth in mucus. *Infect. Immun.* **56,** 1030–1035 (1988).

57. Gries, D. M., Pultz, N. J. & Donskey, C. J. Growth in cecal mucus facilitates colonization of the mouse intestinal tract by methicillin-resistant Staphylococcus aureus. *J. Infect. Dis.* **192,** 1621–1627 (2005).

58. Larsson, J. M. H., Karlsson, H., Sjövall, H. & Hansson, G. C. A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19,** 756–766 (2009).

59. Thomsson, K. A. *et al.* Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2. *Glycobiology* **22,** 1128–1139 (2012).

60. Schluter, J. & Foster, K. R. The evolution of mutualism in gut microbiota via host epithelial selection. *PLoS Biol.* **10,** e1001424 (2012).

61. Sonoyama, K. *et al.* Response of gut microbiota to fasting and hibernation in Syrian hamsters. *Applied and Environmental Microbiology* **75,** 6451–6456

(2009).

62. Carey, H. V., Walters, W. A. & Knight, R. Seasonal restructuring of the ground squirrel gut microbiota over the annual hibernation cycle. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **304,** R33–42 (2013).

63. Schwab, C. *et al.* Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *ISME J* **8,** 1101–1114 (2014).

64. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4,** 447–457 (2008).

65. Sommer, F. *et al.* Altered mucus glycosylation in core 1 o-glycan-deficient mice affects microbiota composition and intestinal architecture. *PLoS ONE* **9,** e85254 (2014).

66. Bergström, A. *et al.* Nature of bacterial colonization influences transcription of mucin genes in mice during the first week of life. *BMC Res Notes* **5,** 402 (2012).

67. Needham, B. D. & Trent, M. S. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nat. Rev. Microbiol.* **11,** 467–481 (2013).

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

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

70. Gallo, R. L. & Hooper, L. V. Epithelial antimicrobial defence of the skin and intestine. *Nat. Rev. Immunol.* **12,** 503–516 (2012).

71. Baughn, A. D. & Malamy, M. H. The strict anaerobe Bacteroides fragilis grows in and benefits from nanomolar concentrations of oxygen. *Nature* **427,** 441–444 (2004).

72. Miyoshi, A. *et al.* Oxidative stress in Lactococcus lactis. *Genet. Mol. Res.* **2,** 348–359 (2003).

73. Johansson, M. E. V., Larsson, J. M. H. & Hansson, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci USA* **108 Suppl 1,** 4659–4665 (2011).

74. Pelaseyed, T. *et al.* The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol. Rev.* **260,** 8–20 (2014).

75. 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).

76. Stecher, B. *et al.* Flagella and chemotaxis are required for efficient induction of Salmonella enterica serovar Typhimurium colitis in streptomycin-pretreated mice. *Infect. Immun.* **72,** 4138–4150 (2004).

77. Navarro-Garcia, F. *et al.* Pic, an autotransporter protein secreted by different pathogens in the Enterobacteriaceae family, is a potent mucus secretagogue. *Infect. Immun.* **78,** 4101–4109 (2010).

78. Nakjang, S., Ndeh, D. A., Wipat, A., Bolam, D. N. & Hirt, R. P. A novel

extracellular metallopeptidase domain shared by animal host-associated mutualistic and pathogenic microbes. *PLoS ONE* **7,** e30287 (2012).

79. Luo, Q. *et al.* Enterotoxigenic Escherichia coli secretes a highly conserved mucin-degrading metalloprotease to effectively engage intestinal epithelial cells. *Infect. Immun.* **82,** 509–521 (2014).

80. Mahdavi, J. *et al.* Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. *Science* **297,** 573–578 (2002).

81. Davis, C. P. & Savage, D. C. Habitat, succession, attachment, and morphology of segmented, filamentous microbes indigenous to the murine gastrointestinal tract. *Infect. Immun.* **10,** 948–956 (1974).

82. Yin, Y. *et al.* Comparative analysis of the distribution of segmented filamentous bacteria in humans, mice and chickens. *ISME J* **7,** 615–621 (2013).

83. Schnupf, P. *et al.* Growth and host interaction of mouse segmented filamentous bacteria in vitro. *Nature* (2015). doi:10.1038/nature14027

84. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139,** 485–498 (2009).

85. Lee, Y. K., Menezes, J. S., Umesaki, Y. & Mazmanian, S. K. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc Natl Acad Sci USA* **108 Suppl 1,** 4615–4622 (2011).

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

87. Sansonetti, P. J. War and peace at mucosal surfaces. *Nat. Rev. Immunol.* **4,** 953–964 (2004).

88. Taylor, R. K., Miller, V. L., Furlong, D. B. & Mekalanos, J. J. Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc Natl Acad Sci USA* **84,** 2833–2837 (1987).

89. Bhowmick, R. *et al.* Intestinal adherence of Vibrio cholerae involves a coordinated interaction between colonization factor GbpA and mucin. *Infect. Immun.* **76,** 4968–4977 (2008).

90. Mouricout, M. Interactions between the enteric pathogen and the host. An assortment of bacterial lectins and a set of glycoconjugate receptors. *Adv. Exp. Med. Biol.* **412,** 109–123 (1997).

91. Lecuit, M. *et al.* A transgenic model for listeriosis: role of internalin in crossing the intestinal barrier. *Science* **292,** 1722–1725 (2001).

92. McCormick, B. A., Colgan, S. P., Delp-Archer, C., Miller, S. I. & Madara, J. L. Salmonella typhimurium attachment to human intestinal epithelial monolayers: transcellular signalling to subepithelial neutrophils. *J. Cell Biol.* **123,** 895–907 (1993).

93. Winter, S. E. *et al.* Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature* **467,** 426–429 (2010).

94. Savage, D. C. Microbial interference between indigenous yeast and lactobacilli in the rodent stomach. *Journal of Bacteriology* **98,** 1278–1283 (1969).

95. Morotomi, M., Watanabe, T., Suegara, N., Kawai, Y. & Mutai, M. Distribution of indigenous bacteria in the digestive tract of conventional and gnotobiotic rats. *Infect. Immun.* **11,** 962–968 (1975).

96. Sengupta, R. *et al.* The role of cell surface architecture of lactobacilli in host-

microbe interactions in the gastrointestinal tract. *Mediators Inflamm.* **2013,** 237921–16 (2013).

97. Mackenzie, D. A. *et al.* Strain-specific diversity of mucus-binding proteins in the adhesion and aggregation properties of Lactobacillus reuteri. *Microbiology (Reading, Engl.)* **156,** 3368–3378 (2010).

98. Frese, S. A. *et al.* Molecular characterization of host-specific biofilm formation in a vertebrate gut symbiont. *PLoS Genet.* **9,** e1004057 (2013).

99. Ossowski, von, I. *et al.* Mucosal adhesion properties of the probiotic Lactobacillus rhamnosus GG SpaCBA and SpaFED pilin subunits. *Applied and Environmental Microbiology* **76,** 2049–2057 (2010).

100. Turroni, F. *et al.* Role of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in modulating bacterium-host interactions. *Proc Natl Acad Sci USA* (2013). doi:10.1073/pnas.1303897110

101. Kubinak, J. L. *et al.* MyD88 signaling in T cells directs IgA-mediated control of the microbiota to promote health. *Cell Host Microbe* **17,** 153–163 (2015).

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

103. Mathias, A. & Corthésy, B. N-Glycans on secretory component: mediators of the interaction between secretory IgA and gram-positive commensals sustaining intestinal homeostasis. *Gut Microbes* **2,** 287–293 (2011).

104. Peterson, D. A. *et al.* Characterizing the Interactions Between a Naturally-primed Immunoglobulin A and its Conserved Bacteroides thetaiotaomicron Species-specific Epitope in Gnotobiotic Mice. *J Biol Chem* (2015).

doi:10.1074/jbc.M114.633800

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

106. Coyne, M. J., Reinap, B., Lee, M. M. & Comstock, L. E. Human symbionts use a host-like pathway for surface fucosylation. *Science* **307,** 1778–1781 (2005).

107. Fanning, S. *et al.* Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc Natl Acad Sci USA* (2012). doi:10.1073/pnas.1115621109

108. Jeon, S. G. *et al.* Probiotic Bifidobacterium breve induces IL-10-producing Tr1 cells in the colon. *PLoS Pathog.* **8,** e1002714 (2012).

109. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous Clostridium species. *Science* **331,** 337–341 (2011).

110. Geuking, M. B. *et al.* Intestinal bacterial colonization induces mutualistic regulatory T cell responses. *Immunology* **34,** 794–806 (2011).

111. Arpaia, N. *et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504,** 451–455 (2013).

112. Smith, P. M. *et al.* The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science* (2013). doi:10.1126/science.1241165

113. Shan, M. *et al.* Mucus enhances gut homeostasis and oral tolerance by delivering immunoregulatory signals. *Science* **342,** 447–453 (2013).

114. Monack, D. M., Mueller, A. & Falkow, S. Persistent bacterial infections: the

interface of the pathogen and the host immune system. *Nat. Rev. Microbiol.* **2,** 747–765 (2004).

115. Randal Bollinger, R., Barbas, A. S., Bush, E. L., Lin, S. S. & Parker, W. Biofilms in the large bowel suggest an apparent function of the human vermiform appendix. *J. Theor. Biol.* **249,** 826–831 (2007).

116. Hanson, N. B. & Lanning, D. K. Microbial induction of B and T cell areas in rabbit appendix. *Dev. Comp. Immunol.* **32,** 980–991 (2008).

117. Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W. F. & Veldhuyzen van Zanten, S. J. O. Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J. Clin. Microbiol.* **44,** 4136–4141 (2006).

118. Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55,** 205–211 (2006).

119. Walker, A. W. *et al.* High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol.* **11,** 7 (2011).

120. Ott, S. J. *et al.* Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* **53,** 685–693 (2004).

121. Abrahamsson, T. R. *et al.* Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin. Exp. Allergy* **44,** 842–850 (2014).

122. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human

gut microbes associated with obesity. *Nature* **444,** 1022–1023 (2006).

123. Garcovich, M., Zocco, M. A., Roccarina, D., Ponziani, F. R. & Gasbarrini, A. Prevention and treatment of hepatic encephalopathy: focusing on gut microbiota. *World J. Gastroenterol.* **18,** 6693–6700 (2012).

124. Henao-Mejia, J. *et al.* Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* (2012). doi:10.1038/nature10809

125. Zhu, Q., Gao, R., Wu, W. & Qin, H. The role of gut microbiota in the pathogenesis of colorectal cancer. *Tumour Biol.* **34,** 1285–1300 (2013).

126. Wu, N. *et al.* Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb Ecol* **66,** 462–470 (2013).

127. Collins, S. M., Surette, M. & Bercik, P. The interplay between the intestinal microbiota and the brain. *Nat. Rev. Microbiol.* **10,** 735–742 (2012).

128. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15,** 382–392 (2014).

129. Petersen, C. & Round, J. L. Defining dysbiosis and its influence on host immunity and disease. *Cell. Microbiol.* **16,** 1024–1033 (2014).

130. Swidsinski, A. *et al.* Mucosal flora in inflammatory bowel disease. *Gastroenterology* **122,** 44–54 (2002).

131. Baumgart, M. *et al.* Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *ISME J* **1,** 403–418 (2007).

132. Rowan, F. *et al.* Bacterial colonization of colonic crypt mucous gel and disease

activity in ulcerative colitis. *Ann. Surg.* **252,** 869–875 (2010).

133. Bajaj, J. S. *et al.* Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. *Am. J. Physiol. Gastrointest. Liver Physiol.* **303,** G675–85 (2012).

134. Rosebury, T. *Microorganisms Indigenous to Man*. (McGraw-Hill, 1962).

135. Freter, R., Brickner, H., Botney, M., Cleven, D. & Aranki, A. Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. *Infect. Immun.* **39,** 676–685 (1983).

136. Maltby, R., Leatham-Jensen, M. P., Gibson, T., Cohen, P. S. & Conway, T. Nutritional Basis for Colonization Resistance by Human Commensal Escherichia coli Strains HS and Nissle 1917 against E. coli O157:H7 in the Mouse Intestine. *PLoS ONE* **8,** e53957 (2013).

137. Wilson, K. H. & Perini, F. Role of competition for nutrients in suppression of Clostridium difficile by the colonic microflora. *Infect. Immun.* **56,** 2610–2614 (1988).

C h a p t e r 2

GUT MICROBIOTA UTILIZE IMMUNOGLOBULIN A FOR MUCOSAL COLONIZATION

Donaldson G.P., Ladinsky M.S., Yu K.B., Sanders J.G., Yoo B.B., Chou W.C., Conner M.E., Earl A.M., Knight R., Bjorkman P.J., and Mazmanian S.K.

This chapter was published in 2018 in *Science*. DOI: 10.1126/science.aaq0926

Abstract

The immune system responds vigorously to microbial infection, while permitting life-long colonization by the microbiome. Mechanisms that facilitate the establishment and stability of the gut microbiota remain poorly described. We discovered that a sensor/regulatory system in the prominent human commensal *Bacteroides fragilis* modulates its surface architecture to invite binding of immunoglobulin A (IgA). Specific immune recognition facilitated bacterial adherence to cultured intestinal epithelial cells and intimate association with the gut mucosal surface *in vivo*. The IgA response was required for *B. fragilis*, and other commensal species, to occupy a defined mucosal niche that mediated stable colonization of the gut through exclusion of exogenous competitors. Therefore, in addition to its role in pathogen clearance, we propose that IgA responses can be co-opted by the microbiome to engender robust host-microbial symbiosis.

Main Text

At birth, ecological and evolutionary processes commence to assemble a complex microbial consortium in the animal gut. Community composition of the adult human gut microbiome is remarkably stable during health, despite day-to-day variability in diet and diverse environmental exposures. Instability, or dysbiosis, may be involved in the etiology of a variety of immune, metabolic, and neurologic diseases (*1*, *2*). Longitudinal sequencing studies indicate a majority of bacterial strains persist within an individual for years (*3*), and for most species there is a single, persistently dominant strain (*4*) (termed “single-strain stability"). Mucus and components of the innate and adaptive immune systems are thought to influence microbiome stability, independently of diet. For example, immunoglobulin A (IgA), the main antibody isotype secreted in the gut, shapes the composition of the intestinal microbiome via currently unknown mechanisms (*5*-*8*). IgA deficiency in mice increases inter-individual variability in the microbiome (*9*) and decreases diversity (*10*, *11*). The direct effects of IgA on bacteria have largely been studied in the context of enteric infection by pathogens (*12*). However, early studies of IgA in the healthy gut found that the majority of live bacterial cells in feces are bound by IgA (*13*), reflecting a steady-state IgA response to persistent indigenous microbes (*14*). Studies show that IgA promotes adherence of commensal bacteria to tissue-cultured intestinal epithelial cells (*15*, *16*), though the *in vivo* implications of this observation are unclear. Furthermore, lack of IgA, the most common human immunodeficiency, does not affect lifespan and only modestly increases susceptibility to respiratory and gastrointestinal infections (*17*), raising the question of why the immune system evolved to invest the considerable energy to produce several grams of IgA daily (*18*).

*Bacteroides fragili*s is an important member of the human gut microbiome, with beneficial properties that ameliorate inflammatory and behavioral symptoms in preclinical animal models (*19*-*22*). This commensal exhibits remarkable single-strain stability (*23*, *24*) and enriched colonization of the gut mucosal surface (*25*). To explore physical features of *B. fragilis* interaction with the host epithelium, we used transmission electron microscopy (TEM) to visualize colonic tissues of mono-colonized mice. *B. fragilis* commonly formed discrete aggregates of tightly-packed cells on the apical epithelial surface (Fig. 1A) and penetrated the glycocalyx layer of transmembrane mucins, nearly contacting the microvilli (Fig. 1B and fig. S1A and B). Intact *B. fragilis* cells were also found nestled in the ducts of the crypts of Lieberkühn (Fig. 1C and S1C). We previously identified a genetic locus in *B. fragilis*, named the commensal colonization factors (*ccfABCDE*), which is necessary for colonization of colonic crypts (*26*). To assess how these genes affect bacterial localization to the mucosal surface, we mono-colonized mice with a *ccfCDE* (Δ*ccf*) mutant. By TEM, *B. fragilis* Δ*ccf* was only found as sparse, individual cells within the epithelial mucosa, excluded from contact with the glycocalyx (Fig. 1D and E), and never observed in aggregates as for wild-type bacteria (Fig. 1F). *B. fragilis* burden in the colon lumen was identical between strains (fig. S2A), suggesting that the CCF system is required specifically for bacterial aggregation within mucus.

High-resolution tomograms of bacterial cells *in vivo* revealed the presence of a thick, fuzzy capsule layer covering wild-type *B. fragilis* (Fig. 1G), which was significantly reduced in *B. fragilis* Δ*ccf* (Fig. 1H and 1I). We sought to investigate the bacterial physiology underlying this ultrastructural change, and potential corresponding effects on colonization.

The *ccf* locus is highly induced during gut colonization (*26*) and bacterial growth in mucin O-glycans (*27*), indicating the CCF system may sense a specific host-derived glycan. The *ccf* genes are homologous to polysaccharide utilization systems in which a sigma factor (*ccfA*) is activated by extracellular glycan sensing (*28*), thus we hypothesized that *ccfA* may activate genes involved in mucosal colonization. We overexpressed *ccfA* in *B. fragilis* and assessed global gene expression by RNAseq during *in vitro* growth (without overexpression *ccf* is poorly expressed in culture (*26*)). Of the non-*ccf* genes regulated by *ccfA*, 24 out of 25 genes mapped to the biosynthesis loci for capsular polysaccharides A and C (PSA and PSC) (Fig. 2A, 2B and Table S1). Correspondingly, *ccf* mutation decreased expression of PSC and increased expression of PSA *in vivo* (Fig. 2C). While phase variation of capsular polysaccharides is known to influence general *in vivo* fitness of *B. fragilis* (*29*, *30*), these studies identify a pathway for transcriptional regulation of specific polysaccharides in the context of mucosal colonization.

We modeled single-strain stability using a horizontal transmission assay, wherein co-housing animals respectively harboring isogenic strains of wild-type *B. fragilis* resulted in minimal strain transmission from one animal to another (Fig. 2D, S2A). This intra-species colonization resistance is provided through bacterial occupation of a species-specific nutrient or spatial niche (*26*). However, as previously reported (*26*), if mice are colonized initially with *B. fragilis* Δ*ccf*, animals were permissive to co-colonization by wild-type *B. fragilis* after co-housing (Fig. 2E, S2B), indicating a CCF-dependent defect in niche saturation. Mice harboring a mutant in the biosynthesis genes for PSC (ΔPSC) showed highly variable co-colonization by wild-type bacteria (Fig. 2F, S2C). Interestingly, we

observed an unexpected increase in expression of the PSB biosynthesis genes in this mutant (Fig. 2H), which may compensate for the loss of PSC. We generated a strain defective in synthesizing both PSB and PSC (ΔPSB/C), and mice mono-associated with the double mutant were consistently unable to maintain colonization resistance (Fig. 2G, S2D-F), though the strain retained *ccf* expression (fig. S2G). Despite lack of competition in a mono-colonized setting and equal levels of colonization in the colon lumen (fig. S2H), the *B. fragilis* Δ*ccf* and ΔPSB/C strains were defective in colonization of the ascending colon mucus (Fig. 2I), reflecting impaired saturation of the mucosal niche. Accordingly, when we imaged the ΔPSB/C strain *in vivo* employing TEM, though the capsule was not as thin as in *B. fragilis* Δ*ccf* (fig. S2I and J), the hallmark epithelial aggregation phenotype was abrogated compared to wild-type bacteria (fig. S2K and L). Therefore, we conclude that the CCF system regulates capsule expression to mediate *B. fragilis* mucosal colonization and single strain stability.

To investigate host responses contributing to mucosal colonization, we defined the transcriptome of the ascending colon during colonization with wild-type *B. fragilis* or *B. fragilis* Δ*ccf*. Remarkably, 7 of the 14 differentially expressed genes encode immunoglobulin variable chains (Fig. 3A and table S2). We did not observe any elevation of immune responses in Δ*ccf*–colonized mice (fig. S3A), indicating that changes in mucosal association are not caused by inflammation. Accordingly, we tested whether capsular polysaccharide regulation by *ccf* affects IgA recognition of bacteria (*31-33*). In fecal samples from mono-colonized animals, wild-type *B. fragilis* was highly coated with IgA, which was significantly diminished in Δ*ccf* and ΔPSB/C strains (Fig. 3B, 3C, and

S4A). We observed no difference between these strains in the induction of total fecal IgA (Fig. 3D), reflecting equivalent stimulation of nonspecific IgA production (*10*, *34*, *35*). To test bacteria-specific responses, IgA extracted from feces of mice mono-colonized with *B. fragilis* was evaluated for binding to bacteria recovered from mono-colonized *Rag1*$^{-/-}$ mice (*in vivo*-adapted, yet IgA-free bacteria). Western blots of bacterial lysates showed that strong IgA reactivity to capsular polysaccharides was abrogated in the Δ*ccf* and ΔPSB/C strains (Fig. 3E and F). Although IgA can be polyreactive (*10*, *34*, *35*), binding to lysates of *Bacteroides* was species-specific (fig. S4B) and required induction of IgA following bacterial colonization (fig. S4C and D). Accordingly, in a whole bacteria binding assay, IgA induced by wild-type bacteria maximally coated wild-type *B. fragilis* compared with the Δ*ccf* and ΔPSB/C strains (Fig. 3G). IgA induced by *B. fragilis* Δ*ccf* exhibited reduced binding to wild-type bacteria (Fig. 3G). The addition of IgA to *in vivo*-adapted, IgA-free bacteria increased adherence of *B. fragilis* to intestinal epithelial cells in tissue culture (Fig. 3H), yet had no effect on bacterial viability (fig. S4E). Cell lines known to produce more mucus (*36*) exhibited a greater capacity for IgA-enhanced *B. fragilis* adherence (fig. S4F), consistent with prior work showing that IgA binds mucus (*36*-*38*). Importantly, IgA-enhanced adherence was decreased if targeted bacteria lack *ccf* or PSB/C, or if the IgA tested was induced by a *ccf* mutant or *Bacteroides thetaiotaomicron* (Fig. 3H and S4G). While pathogenic bacteria elaborate capsular polysaccharides for immune evasion, these results suggest *B. fragilis* deploys specific capsules for immune attraction, potentially enabling stable mucosal colonization.

We determined whether IgA coating promotes *B. fragilis* colonization in mice. Using the horizontal transmission paradigm, *Rag1*$^{-/-}$ mice colonized with wild-type *B. fragilis* were readily co-colonized by an isogenic strain from a co-housed animal (fig. S5A and B), showing loss of colonization resistance in the absence of adaptive immunity. We next treated wild-type mice with an anti-CD20 antibody (fig. S5C) (*39*) to deplete B cells (fig. S5D-F), thus reducing total fecal IgA levels (fig. S5G) and eliminating IgA coating of wild-type *B. fragilis* during mono-colonization (Fig. 4A). IgA recovered from isotype control treated mice, also mono-colonized with *B. fragilis*, promoted adherence of wild-type bacteria to epithelial cells *in vitro*, while IgA from anti-CD20 treated mice had no effect despite being exposed to *B. fragilis* antigens (Fig. 4B). In the horizontal transmission assay, B cell depleted mice mono-colonized with *B. fragilis* were readily invaded by wild-type bacteria, while isotype control-injected animals retained colonization resistance (Fig. 4C and S5H). Therefore, active B cell responses to *B. fragilis* colonization enhance single-strain stability.

As B cell depletion eliminates all antibody isotypes, germ-free IgA$^{-/-}$ mice (*40*) were generated and mono-colonized with *B. fragilis*. We did not observe compensatory coating by IgM (fig. S6A). In a horizontal transmission assay with wild-type (BALB/c) and IgA$^{-/-}$ mice, lack of IgA allowed co-colonization by challenge strains (Fig. 4D, S6B-D), indicating that IgA specifically contributes to single-strain stability. This feature was reproduced in mice with a full microbial community "spiked" with genetically marked *B. fragilis* strains (fig. S6E and F), revealing that single-strain stability of an individual bacterial species occurs in the context of a complex community. Mono-colonized IgA$^{-/-}$

mice harbored reduced levels of live bacteria in the colon mucus compared to wild-type mice (Fig. 4E), though they had greater numbers of bacteria in the colon lumen (fig. S6G). TEM images of ascending colon tissues reveal that in $IgA^{-/-}$ animals, wild-type *B. fragilis* failed to aggregate on the epithelial surface (Fig. 4F and 4G), similar to the *ccf* and PSB/C mutants in wild-type animals. *B. fragilis* cells also formed aggregates in feces in the presence of IgA (fig. S7), indicating that enhanced mucosal colonization may be due to increased aggregation or growth (*41*) within mucus. These findings converge to support a model whereby *ccf* regulates expression of specific capsular polysaccharides to attract IgA binding, allowing for robust mucosal colonization and single-strain stability.

Beyond *B. fragilis*, we tested whether IgA shapes a complex microbiome following controlled introduction of mouse microbiota to germ-free BALB/c or $IgA^{-/-}$ mice. One month following colonization, despite similar microbiome profiles in feces of both mouse genotypes (fig. S8A), we observed differences for specific taxa (Table S3). We also identified a defect in community stratification between the colonic mucus and lumen of $IgA^{-/-}$ mice (Fig. 4H and S8B), revealing that IgA is required to individualize microbiome profiles between these two anatomic locations. Remarkably, a highly mucus-enriched exact sequence variant (ESV), mapping uniquely to *B. fragilis*, was significantly decreased in the mucus of $IgA^{-/-}$ mice compared to BALB/c mice (Fig. 4I and S9A), naturally supporting our observations from mono-colonized mice. To extend this analysis to other microbial species, we identified Rikanellaceae, *Blautia sp.*, and segmented filamentous bacteria (SFB) as being highly IgA-coated (fig. S9B) (*35*), and assessed the abundance of these taxa in the colonic or ileal mucus. *Blautia sp*. and segmented filamentous bacteria (SFB)

displayed increased mucosal association in the absence of IgA (Fig. 4I) (*42*), demonstrating that IgA can protect the intestinal barrier. However, similar to *B. fragilis*, Rikanellaceae were highly abundant in colon mucus and significantly depleted in $IgA^{-/-}$ mice (Fig. 4I). We conclude that IgA-enhanced mucosal colonization occurs within complex communities for multiple strains of *B. fragilis* and other species of the gut microbiome.

Classically viewed, the immune system evolved to prevent microbial colonization. However, not only do animals tolerate a complex microbiome, in the case of *B. fragilis* provoking an immune response paradoxically enables intimate association with its mammalian host. Related commensal bacteria may also benefit from actively engaging IgA during symbiosis, as $Rag2^{-/-}$ mice devoid of adaptive immunity harbor fewer *Bacteroides* (*43*), and both B cell deficient and $IgA^{-/-}$ animals display decreased colonization by the Bacteroidaceae family (*44*). IgA has been previously shown to increase adherence of *Escherichia coli* (*15*), *Bifidobacterium lactis,* and *Lactobacillus ramnosus* (*16*) to tissue-cultured epithelial cells, suggesting that these microorganisms may also benefit from IgA to establish a mucosal bacterial community. Mucosal microbiome instability or loss of immunomodulatory species may underlie the link between IgA deficiency and autoimmune diseases in humans (*45*). Interestingly, while IgA-coated bacteria from individuals with IBD (*46*) or nutritional deficiencies (*47*) exacerbate respective pathologies in mice, IgA-coated bacteria from healthy humans protect mice from disease (*47*). We propose that during health, IgA fosters mucosal colonization of microbiota with beneficial properties (*9*), while disease states may induce (or be caused by) IgA responses to pathogens or pathobionts that disrupt healthy microbiome equilibria. Indeed, computational models indicate that IgA

can both maintain indigenous mucosal populations and clear invasive pathogens (*48*). In addition to serving as a defense system, we discover that adaptive immunity evolved to engender intimate association with members of the gut microbiome.

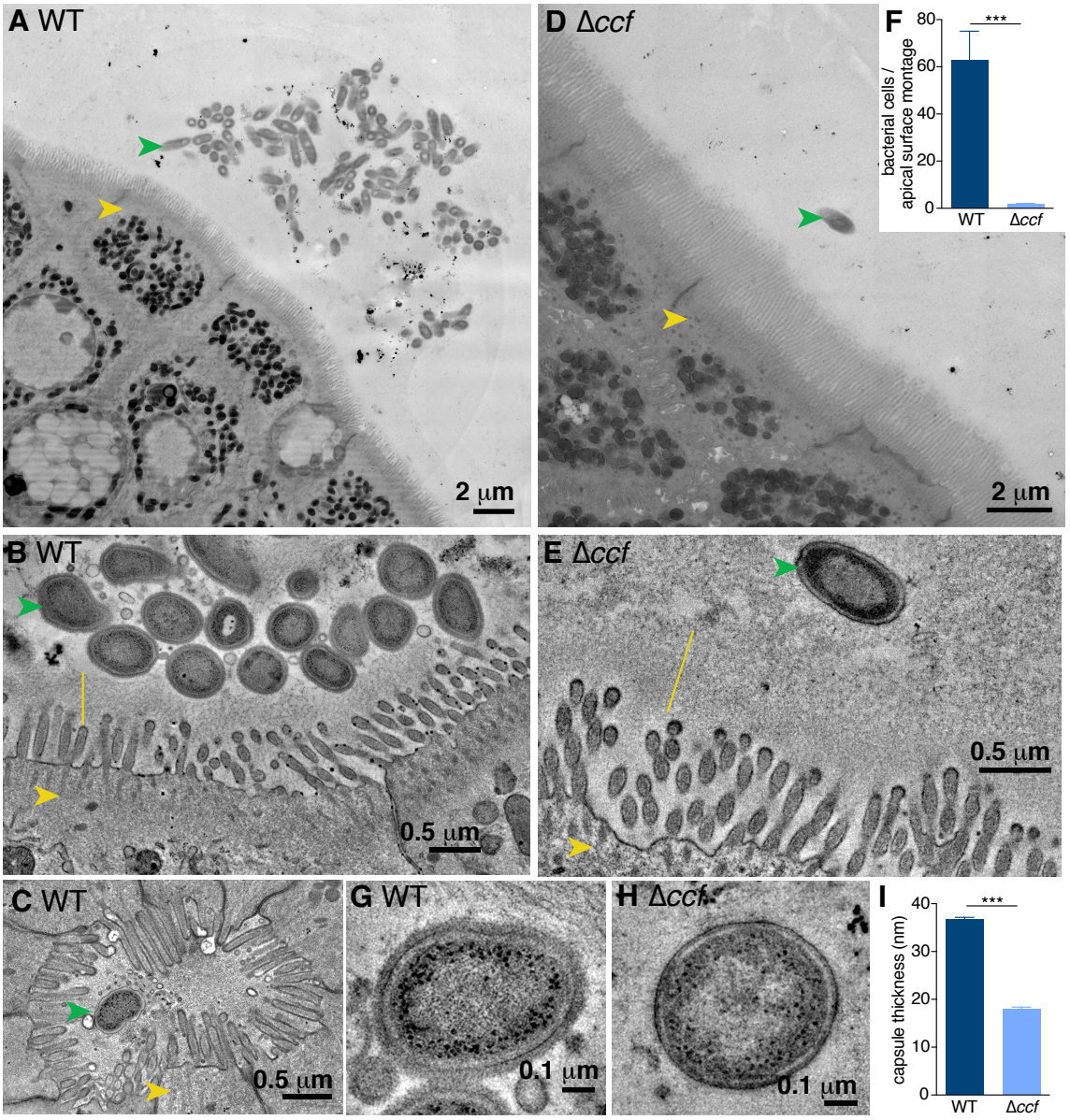


Fig. 1. ***Bacteroides fragilis*** **resides as aggregates on the colon epithelium in a CCF-dependent manner**. (**A**) Representative transmission electron microscopy (TEM) projection and (**B**) high-resolution tomogram of epithelial-associated wild-type *B. fragilis* in mono-colonized mice. Ascending colons of mice harbored aggregates of *B. fragilis* (green arrow) under non-pathogenic conditions that made tight associations with the glycocalyx (yellow line) overlying intestinal epithelial cells (IECs, yellow arrow). (**C**)

Tomogram of wild-type *B. fragilis* penetrating deep into the duct of a crypt of Lieberkühn. (**D**) Representative TEM projection image and (**E**) tomogram of epithelial-associated *B. fragilis* Δ*ccf*. The absence of the CCF system abrogated formation of bacterial aggregates and prevented intimate association with the glycocalyx. (n = 3 mice per group, about 1 mm epithelium scanned per mouse). (**F**) Quantification of bacterial cells per projection montage (A and D) of epithelial-associated bacteria (unpaired t test, n = 7, 8 images from 4 mice per group). (**G** and **H**) Tomogram of the bacterial surface of wild-type *B. fragilis* (G) in comparison to *B. fragilis* Δ*ccf* (H) revealed a thick fuzzy capsule for wild-type bacteria residing in the colons of mice. (**I**) Measurement of capsule thickness (unpaired t test, n = 10 cells from 3 mice per group) (*** $p < 0.001$).

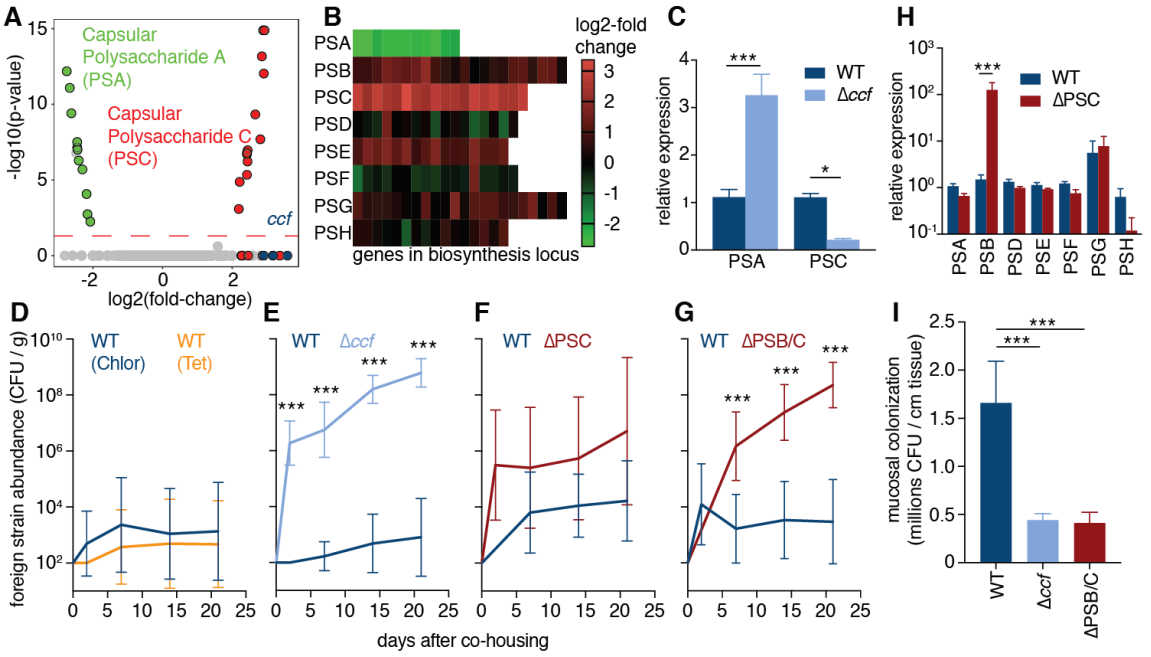


Fig. 2. Specific capsular polysaccharides, regulated by *ccf*, are necessary for single-strain stability. (**A**) RNAseq gene expression analysis of *B. fragilis* overexpressing *ccfA* during laboratory culture growth, relative to empty vector control (n = 3). Green symbols represent PSA genes; red symbols represent PSC genes; blue symbols represent *ccf* genes. (**B**) Heat map of expression levels for all capsular polysaccharide loci in *B. fragilis* following *ccfA* overexpression during growth in culture. (**C**) Relative expression using qRT-PCR (ΔΔCt normalized to gyrase) of RNA from colon lumen contents of mice mono-colonized with *B. fragilis* or *B. fragilis* Δ*ccf* (Sidak 2-way ANOVA, n = 4). (**D-G**) Abundance of foreign strains exchanged between pairs of co-housed mice each mono-colonized with the indicated strains, in colony forming units (CFU) per gram of feces (Sidak repeated measure 2-way ANOVA on log-transformed data, geometric mean and 95% CI, n = 9-12 pairs per plot). (**H**) Relative expression levels of capsular polysaccharides analyzed by qRT-PCR (ΔΔCt normalized to gyrase) of RNA from colon lumen contents of

mice mono-colonized with *B. fragilis* or *B. fragilis* ΔPSC (Sidak 2-way ANOVA, n = 3, 4). (**I**) Plating of CFU from ascending colon mucus of mice mono-colonized with *B. fragilis* strains (Tukey ANOVA, n = 8) (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

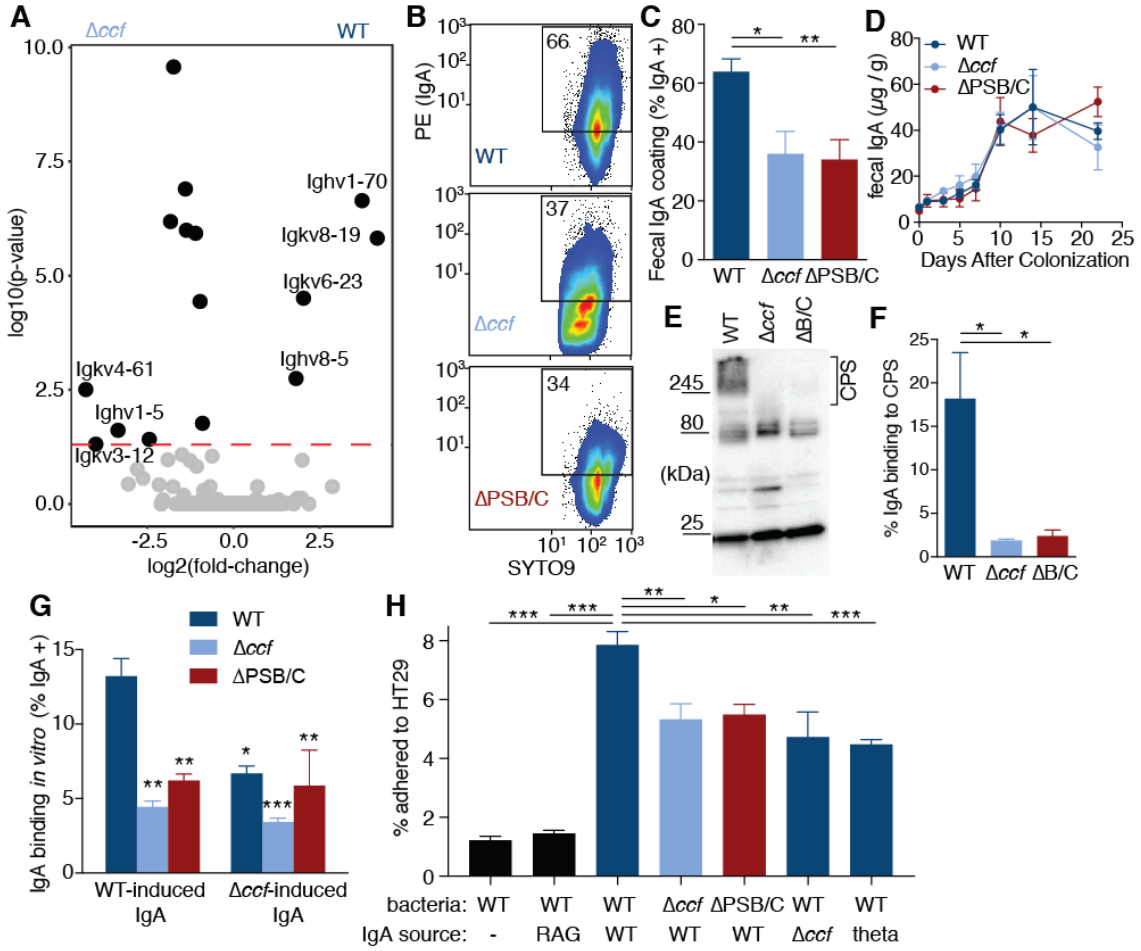


Fig. 3. *B. fragilis* induces a specific IgA response, dependent on *ccf* regulation of surface capsular polysaccharides, which enhances epithelial adherence. (**A**) RNAseq gene expression analysis of RNA recovered from whole ascending colon tissue of mice mono-colonized with *B. fragilis* or *B. fragilis* Δ*ccf* (n = 3). (**B**) Flow cytometry plots and (**C**) quantification of IgA coating of *B. fragilis* from feces of mice mono-colonized with various strains (Tukey ANOVA, n = 11-12). (**D**) ELISA for total fecal IgA in mono-colonized mice (Sidak repeated measure 2-way ANOVA, not significant, n = 4). (**E**) Bacterial lysates from feces of mono-colonized *Rag1*$^{-/-}$ mice probed in Western blots with fecal IgA from *B. fragilis* mono-colonized mice and (**F**) quantification of the proportional

signal from IgA binding to capsular polysaccharides (CPS) (over 245 kDa) (Tukey ANOVA, n = 3 mice). (**G**) Binding of fecal IgA extracted from mono-colonized mice to various strains of *B. fragilis*. Source of IgA is mice colonized with either WT *B. fragilis* or *B. fragilis* Δ*ccf*. Because *ccf* is expressed *in vivo*, IgA-free bacteria from feces of mono-colonized *Rag1*$^{-/-}$ mice were used as the target for IgA binding (Tukey 2-way ANOVA, *significantly different from WT bacteria with WT IgA, n = 3). (**H**) *In vitro* epithelial cell adherence assay using IgA extracted from Swiss Webster mice (or *Rag1*$^{-/-}$, second column) mono-colonized with *B. fragilis* or *B. thetaiotaomicron* (theta; last column). IgA-free but *in vivo*-adapted bacteria were isolated from mono-colonized *Rag1*$^{-/-}$ mice (Tukey ANOVA, n = 4 mice as the source of bacteria) (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

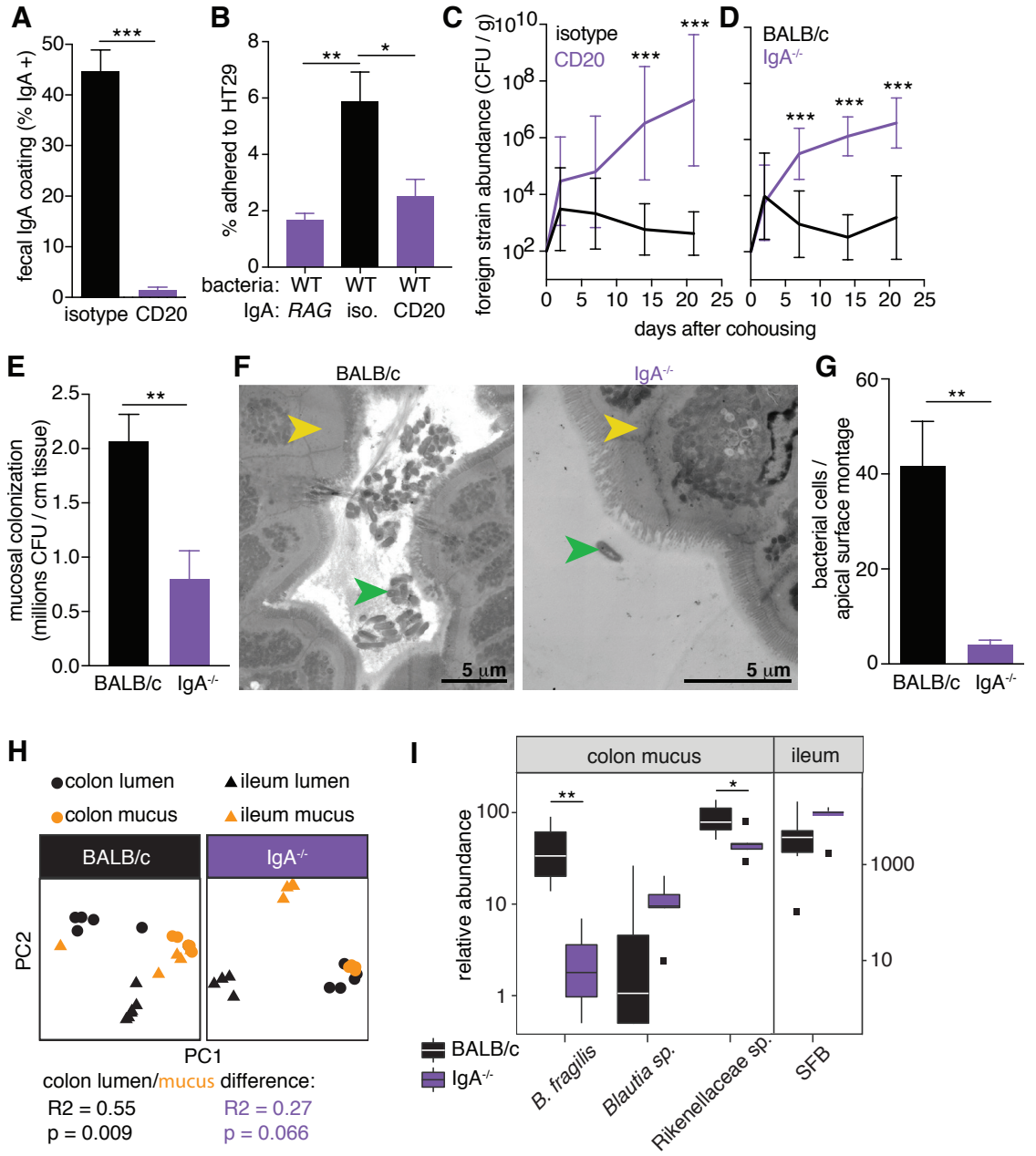


Fig. 4. IgA production *in vivo* is necessary for single-strain stability, mucosal colonization, and epithelial aggregation. (A) IgA coating of wild-type *B. fragilis* in feces following injection of anti-CD20 or isotype control antibody (unpaired t test, n = 8). (B) Epithelial cell adherence assay of wild-type *B. fragilis* incubated with IgA extracted from indicated mono-colonized mice (Tukey ANOVA, n = 4 mice as the source of bacteria). (C)

Abundance of foreign strains exchanged between pairs of wild-type *B. fragilis* mono-colonized mice treated with anti-CD20 or an isotype control (Sidak repeated measure 2-way ANOVA on log-transformed data, n = 10). (D) Foreign strains exchanged between pairs of BALB/c and BALB/c $IgA^{-/-}$ mice mono-colonized with wild-type *B. fragilis* (Sidak repeated measure 2-way ANOVA on log-transformed data, n = 9). (E) CFU plating of ascending colon mucus of wild-type and $IgA^{-/-}$ mice mono-colonized with wild-type *B. fragilis* (unpaired t test, n = 9). (F) Representative TEM projections of ascending colon (yellow arrow: epithelial cell) from mice mono-colonized with wild-type *B. fragilis* (green arrow) (n = 3 mice per group, about 1 mm epithelium scanned per mouse) and (G) quantification of bacterial cells per projection montage (unpaired t test, n = 7, 6 images from 3 mice per group) (H) Principle coordinate analyses of weighed UniFrac distances of 16S community profiles of ex-germ-free BALB/c and BALB/c $IgA^{-/-}$ mice transplanted with a complex mouse microbiota (Adonis test within colon for lumen/mucus difference). (I) Relative abundance of *B. fragilis* and highly IgA-coated ESVs in ex-germ-free mice (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

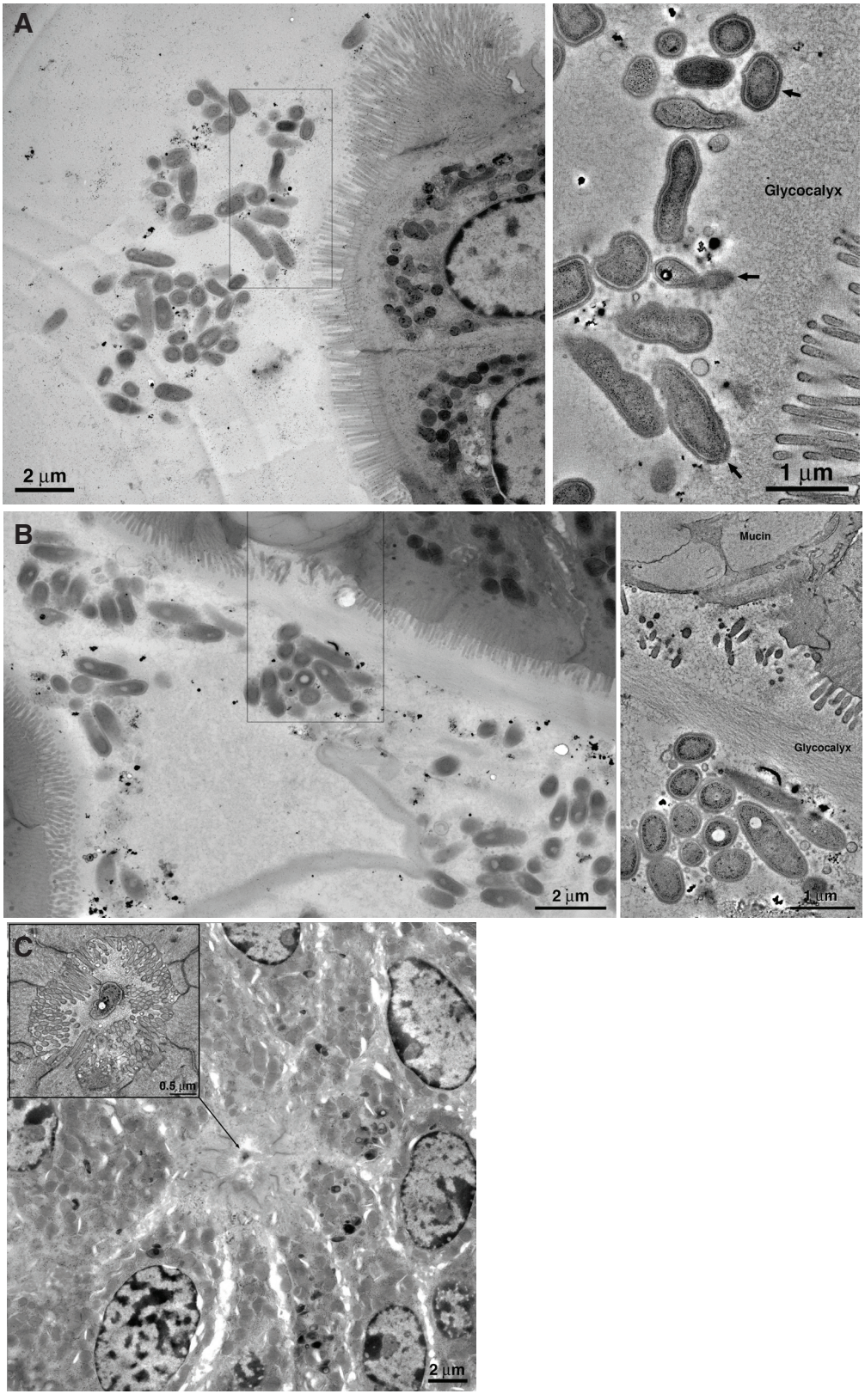

A
2 μm
Glycocalyx
1 μm
B
2 μm
Mucin
Glycocalyx
1 μm
C
0.5 μm
2 μm

Fig. S1. Additional EM images from mice mono-colonized with wild-type *B. fragilis*. (**A** and **B**) Example aggregates of *B. fragilis* on the epithelial surface in projection images (left) with high resolution tomograms of the marked region (right). Arrows indicate examples of bacteria penetrating the glycocalyx. (**C**) Projection and inset tomogram of an additional example of *B. fragilis* in a crypt of Lieberkühn.

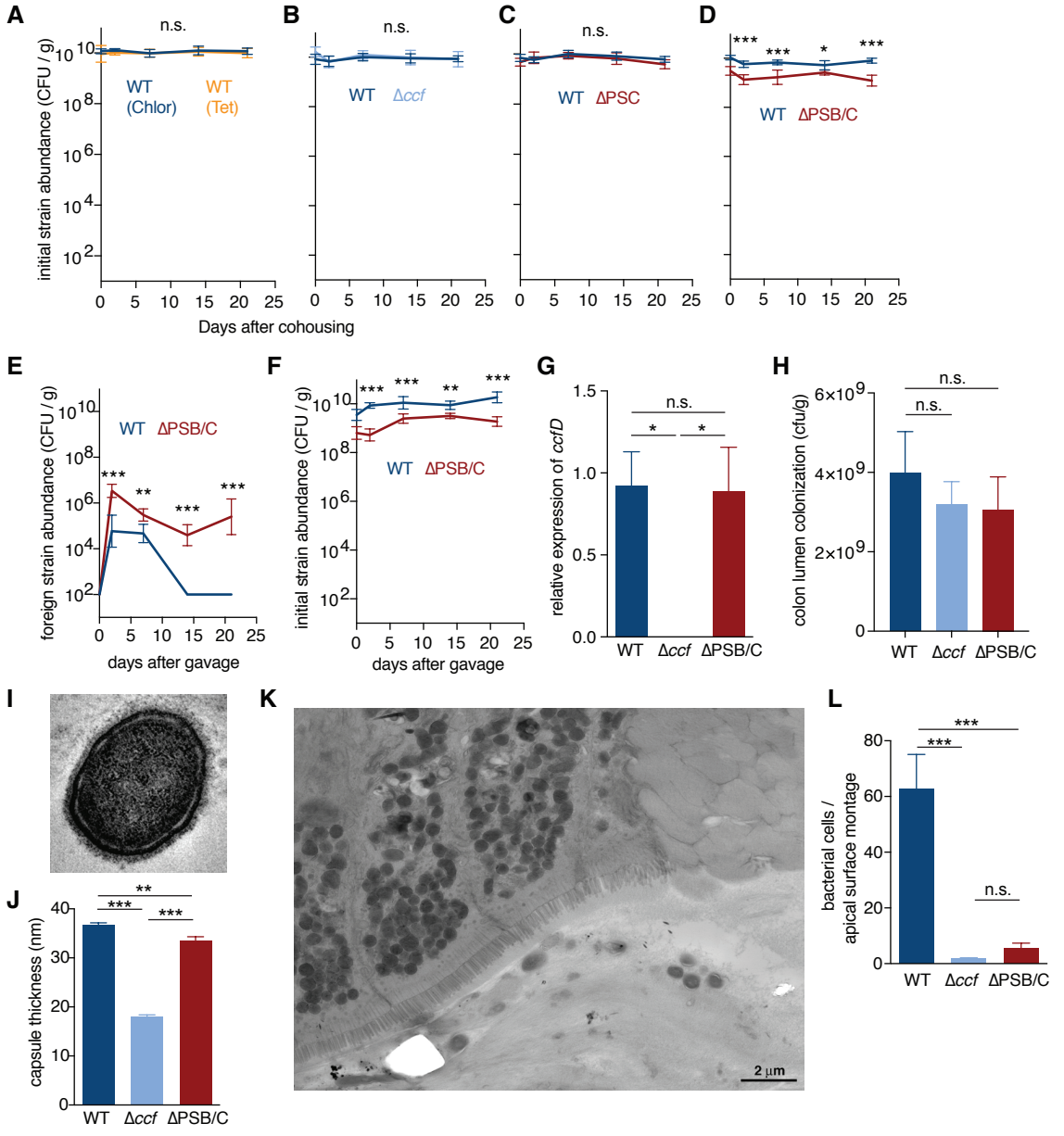


Fig. S2. Additional *in vivo* phenotypes for various *B. fragilis* strains. (**A**) Abundance of the initial strains corresponding to Fig. 2D, (**B**) Fig. 2E, (**C**) Fig. 2F, and (**D**) 2G (n = 9-12). (**E**) Abundance of foreign strains in mice originally mono-colonized with indicated strain (n = 6) and challenged by gavage. (**F**) Abundance of the original strains corresponding to E (geometric mean and 95% confidence interval plotted for all line graphs with Sidak repeated measure 2-way ANOVA on log-transformed data). (**G**) qRT-PCR

analysis for *ccf* expression of RNA extracted from colon lumen contents of mono-colonized mice, indicating that PSB/C mutation does not affect *ccf* expression (Tukey ANOVA, n = 3). (**H**) Quantification of bacteria in the colon lumen of mono-colonized mice, corresponding to Fig. 2I (Tukey ANOVA, n = 8). (**I**) High resolution tomogram (image width = 600 nm) of *B. fragilis* ΔPSB/C in the colon mucosa of a mono-colonized mouse and (**J**) quantification of the capsule thickness including data from Fig. 1I for wild-type and Δ*ccf* (Tukey ANOVA, n = 10). (**K**) Example projection montage of the apical surface of the epithelium of mice mono-colonized with *B. fragilis* ΔPSB/C shows aggregates of only small numbers of bacteria. (**L**) Quantification of number of epithelial associated bacteria including data from Fig. 1F for wild-type and Δ*ccf* (Tukey ANOVA, n = 7, 8, 8 images from 4 mice per group) (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

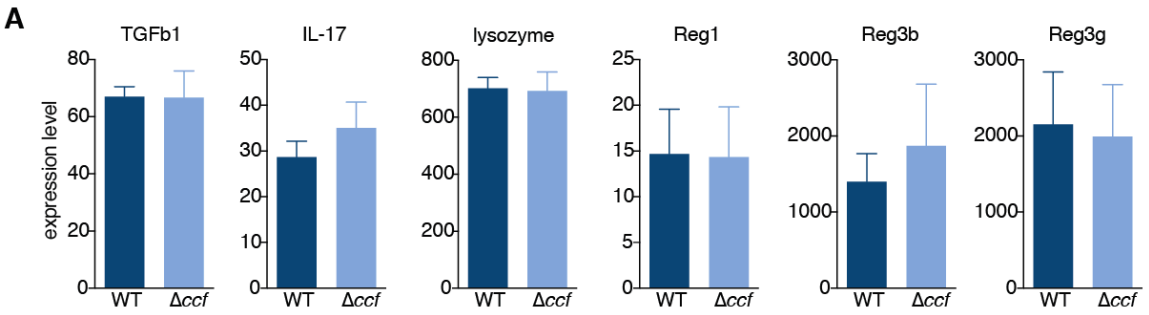


Fig. S3. Host inflammatory profiles of *B. fragilis* and *B. fragilis* Δ*ccf* show no differences. (A) Expression levels of inflammatory cytokines, lysozyme, and antimicrobial peptides from ascending colon RNAseq of mice mono-colonized with *B. fragilis* or *B. fragilis* Δ*ccf* reveal an indistinguishable inflammatory response (n = 3). Beta-defensins, TNF alpha, IL-1, and IL-6 were not expressed at detectable levels.

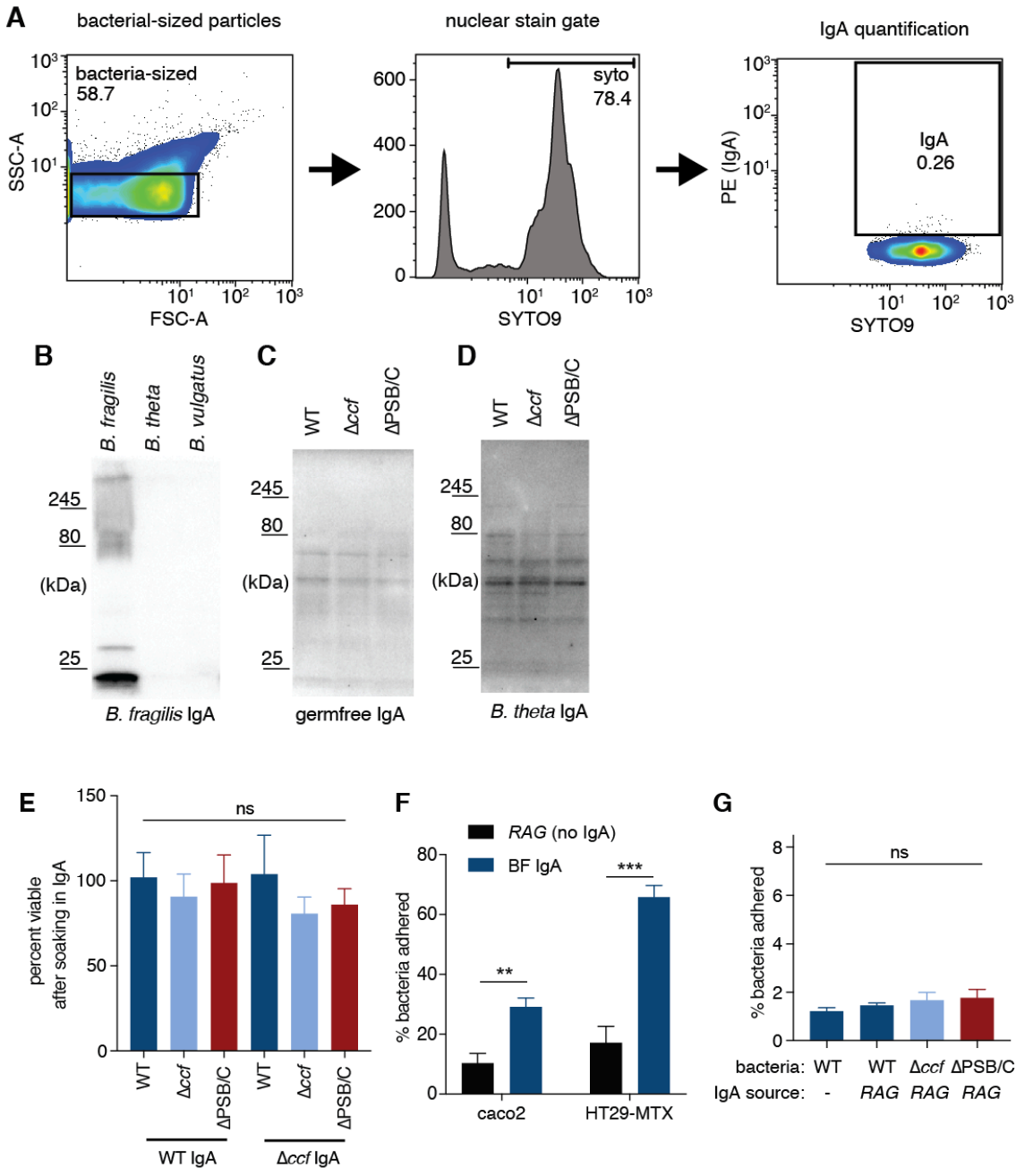


Fig. S4. Interactions between IgA and *B. fragilis*. (A) Gating strategy for assessing IgA coating of bacteria. Gates were defined using single-stained samples. (B) Lysates of culture-grown *Bacteroides* species (10^8 CFU loaded per well for all blots) probed in western blots with fecal IgA from *B. fragilis* mono-colonized mice, indicating minimal cross-reactivity to other species. (C) Lysates of *B. fragilis* strains from the feces of mono-colonized *Rag1*$^{-/-}$ mice

probed in western blots with fecal IgA from germ-free and (D) *B. thetaiotaomicron* mono-colonized mice, with images overexposed to show minimal binding (blots B-D repeated at least 3 times). (E) Plating CFU before and after coating with IgA *in vitro* suggests no impact of IgA on the viability of *B. fragilis* (Tukey ANOVA, n = 8). (F) Tissue-cultured epithelial cell adherence assay with various cell lines (in addition to those used in the main text), using wild-type *B. fragilis* from feces of mono-colonized *Rag1*$^{-/-}$ mice and IgA from mice mono-colonized with *B. fragilis* (Sidak 2-way ANOVA, n = 5 mice as the source of bacteria) (G) Additional controls for Fig. 3H (first two bars are the same data as in Fig. 3H) indicate no baseline difference in epithelial adherence between strains of *B. fragilis* from feces of mono-colonized *Rag1*$^{-/-}$ mice, without the addition of IgA (Tukey ANOVA, n = 4) (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

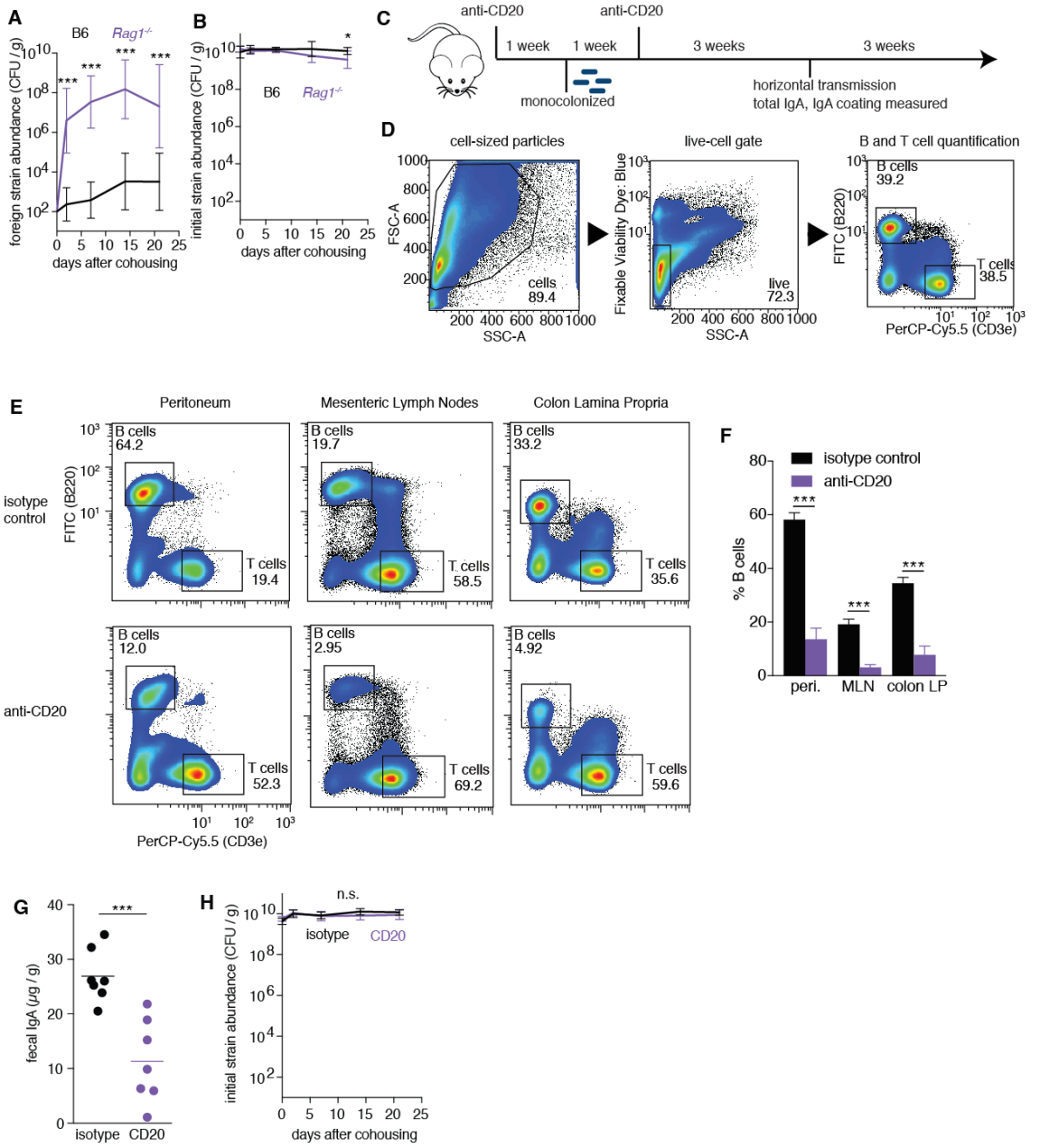


Fig. S5. *B. fragilis* colonization phenotypes in *RAG1*$^{-/-}$ and B cell depleted mice. (**A**) Abundance of foreign strains exchanged between pairs of wild-type *B. fragilis* mono-colonized mice, either C57BL/6 (B6) and *Rag1*$^{-/-}$, in colony forming units (CFU) per gram of feces (geometric mean and 95% confidence interval plotted for all line graphs with Sidak repeated measure 2-way ANOVA on log-transformed data, n = 9). (**B**) Abundance of initial

strains corresponding to **A**. (**C**) Anti-CD20 experimental timeline with germ-free Swiss Webster mice. (**D**) Gating strategy for quantification of B cells to confirm depletion by anti-CD20 antibody and (**E**) example plots. (**F**) Quantification of B cells as a proportion of cell populations in the peritoneum, mesenteric lymph nodes, and lymphocyte fraction of the colon lamina propria (LP) following injection of anti-CD20 or isotype control antibody (Sidak 2-way ANOVA, n = 4, 5). (**G**) IgA concentration in feces as assessed by ELISA 3 weeks after second injection of anti-CD20 or isotype control antibody (unpaired t test, n = 7). (**H**) Abundance of initial strains for Fig. 4C (n = 10) (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

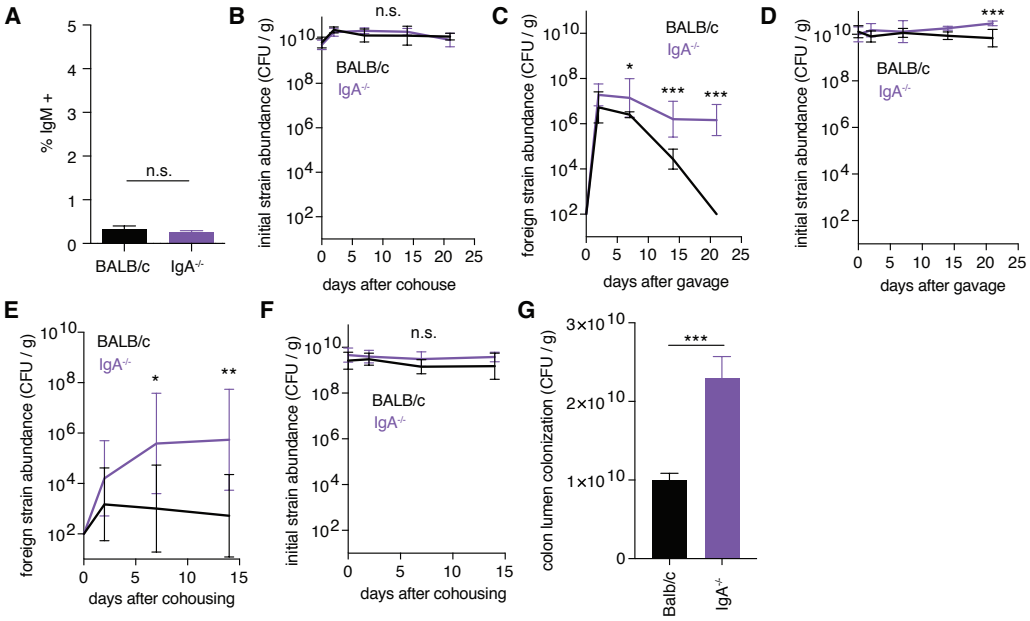


Fig. S6. *In vivo* phenotypes for *B. fragilis* in BALB/C and BALB/c IgA$^{-/-}$ mice. (**A**) Percent of *B. fragilis* coated in IgM in mono-colonized BALB/c and BALB/c IgA$^{-/-}$ mice (unpaired t test, n = 4). (**B**) Abundance of initial strains corresponding to Fig 4D (geometric mean and 95% confidence interval plotted for all line graphs with Sidak repeated measure 2-way ANOVA on log-transformed data, n = 9). (**C**) Abundance of foreign strains in mice mono-colonized with *B. fragilis* and challenged by gavage with *B. fragilis* (n = 5). (**D**) Abundance of initial strains corresponding to **C**. (**E**) Abundance of foreign strains of *B. fragilis* in SPF mice colonized with *B. fragilis* and co-housed (n = 10). (**F**) Abundance of initial strains corresponding to **E**. (**G**) Quantification of colon lumen colonization corresponding to Fig 4E (unpaired t test, n = 9). (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

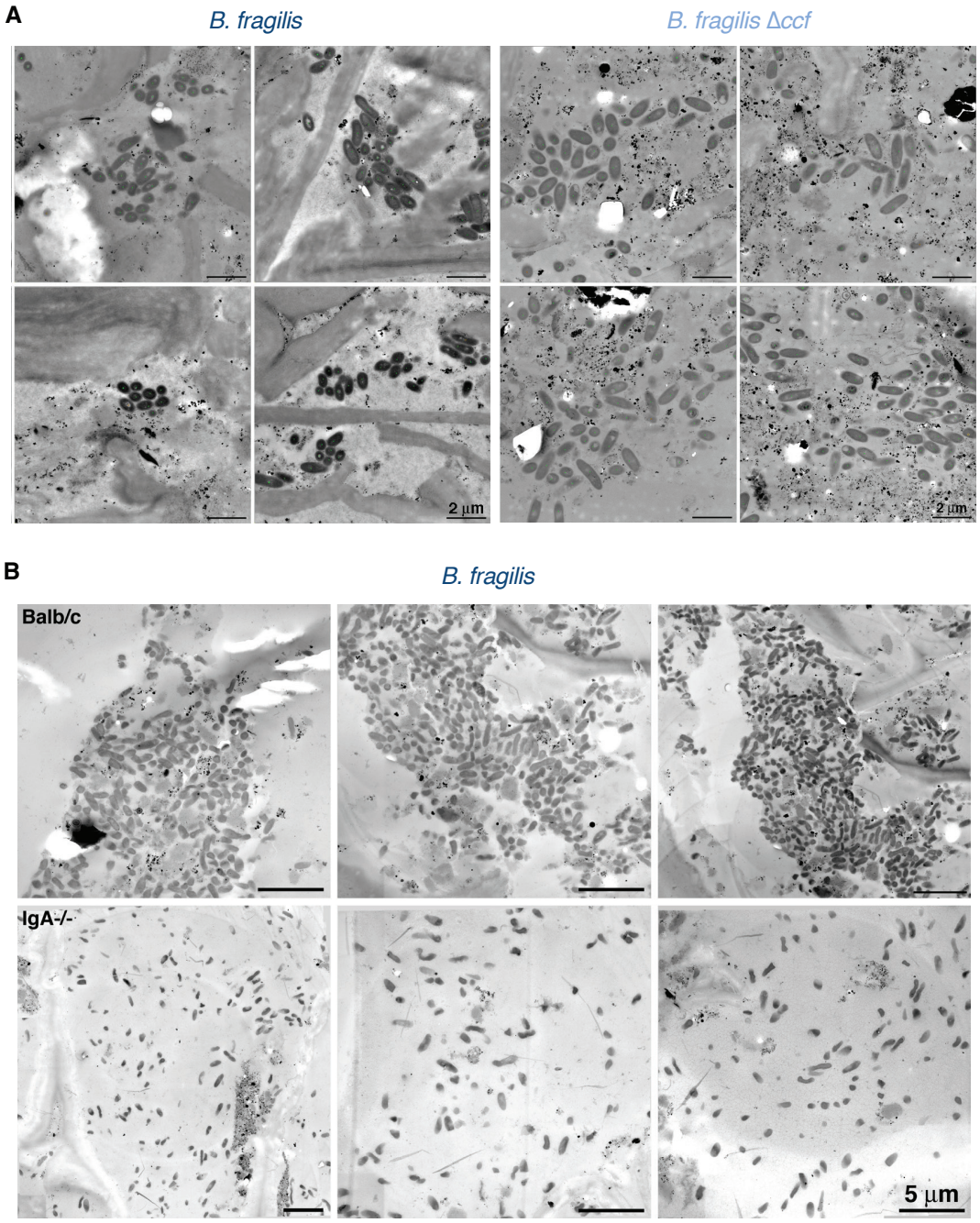


Fig. S7. Aggregation of *B. fragilis in vivo* is affected by IgA and *ccf*. (**A**) Transmission electron micrographs of colon lumen contents of Swiss Webster mice mono-colonized with *B. fragilis* or *B. fragilis* Δ*ccf* (4 representative images for each). (**B**) Transmission electron

micrographs of feces of BALB/c and BALB/c $IgA^{-/-}$ mice mono-colonized with *B. fragilis* (3 representative images for each).

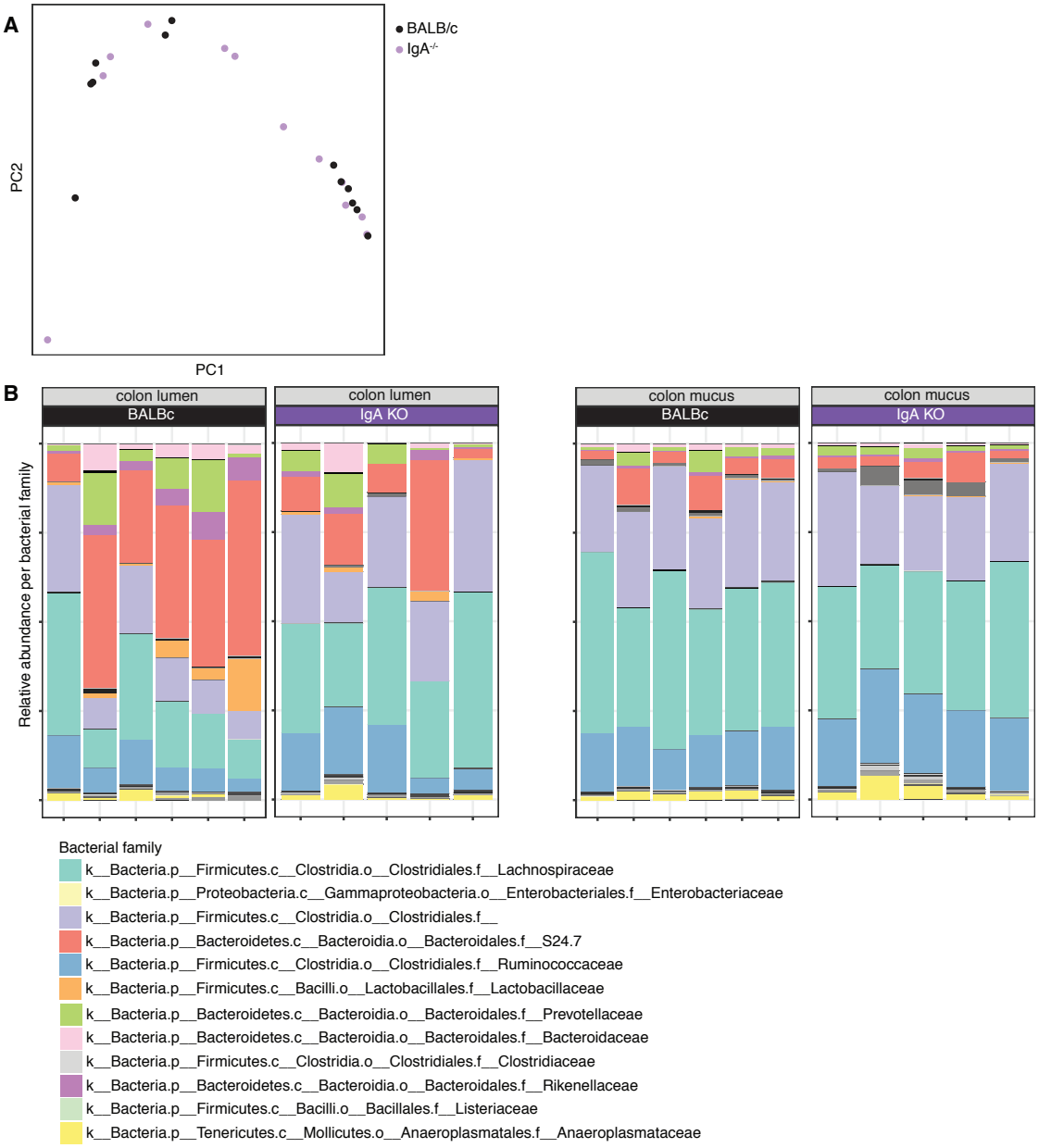


Fig. S8. 16S microbiome profiling of ex-germ-free mice. (**A**) Principle coordinate analysis of weighted UniFrac distances of fecal microbiomes between BALB/c and BALB/c IgA$^{-/-}$ mice. (**B**) Relative abundance of bacterial families in colon and ileal lumen and mucus in ex-germ-free BALB/c and IgA$^{-/-}$ mice. Most abundant 12 families assigned colors; remaining families assigned random shades of gray.

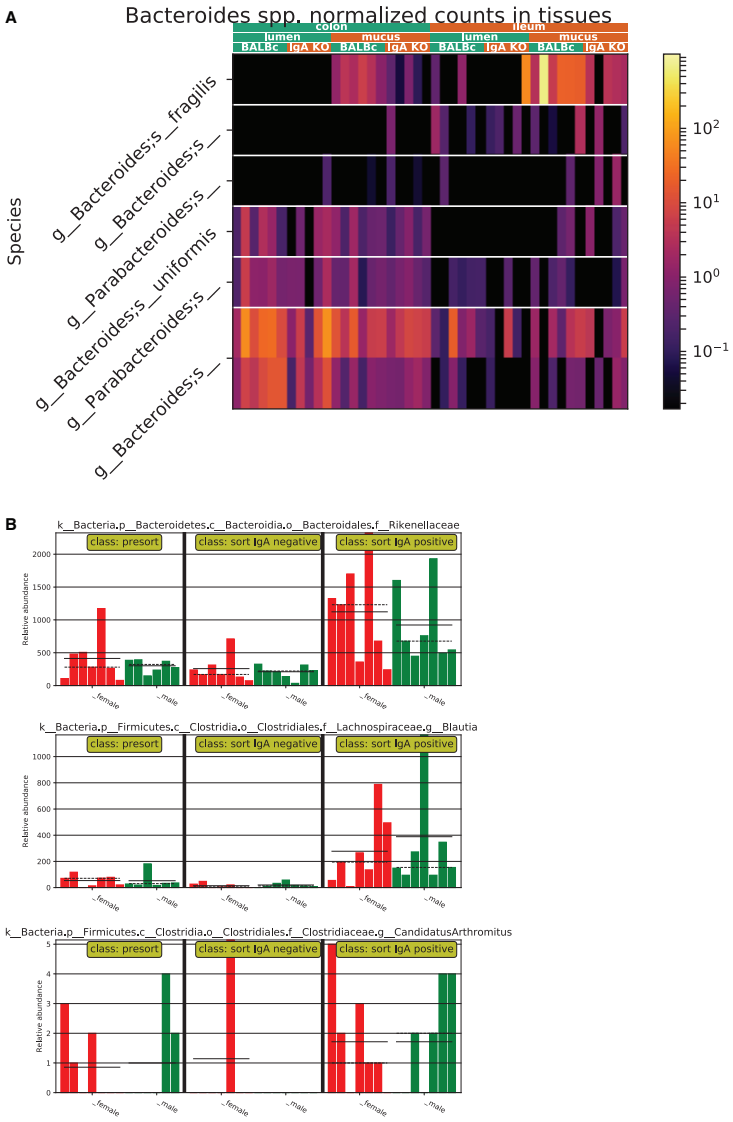


Fig. S9. Analyses of exact sequence variants (ESVs) of interest in ex-germ-free mice. (A) Heatmap of relative abundance of ESVs assigned to *Bacteroides* or *Parabacteroides* in the mucus and lumen of the colon and ileum of BALB/c and BALB/c IgA$^{-/-}$ mice. (B) Relative abundance of highly IgA-bound taxa (LEfSe LDA effect size > 2) before and after sorting (ESVs in Figure 4I are the most abundant representatives of these taxonomic groups; "Candidatus Arthromitus" = SFB).

Table S1.

Differentially expressed (adjusted p-value <0.05) genes in RNAseq of *B. fragilis* grown in culture with and without overexpression of *ccfA*. Genes are ordered by the fold-change.

Locus ID	Gene	Log2(fold-change) (ccfA overexpression / vector)	Adjusted P-value
BF3583	ccfA	4.08	2.139E-106
BF1021	wcfG (PSC)	2.93	1.29145E-15
BF1009	upcY (PSC)	2.91	9.25035E-13
BF1011	rmlA2 (PSC)	2.89	1.29145E-15
BF1023	wcfI (PSC)	2.85	6.7139E-14
BF1017	wcfD (PSC)	2.81	2.02687E-08
BF1022	wcfH (PSC)	2.66	4.63823E-10
BF1024	wcfJ (PSC)	2.44	1.07951E-07
BF1019	wcfE (PSC)	2.44	5.99448E-07
BF1012	rmlC1 (PSC)	2.42	4.56061E-06
BF1015	wcfB (PSC)	2.42	1.98807E-07
BF1025	wcfK (PSC)	2.42	1.65081E-07
BF1010	upcZ (PSC)	2.21	1.33346E-05
BF1020	wcfF (PSC)	2.19	0.000822972
BF1377	wcfS (PSA)	-2.07	0.005651271
BF1376	wcfR (PSA)	-2.15	0.001828779
BF1369	wzx3 (PSA)	-2.18	8.44926E-05
BF1373	wcfO (PSA)	-2.30	2.01292E-06
BF1374	wcfP (PSA)	-2.41	5.18014E-07
BF1372	wzy3 (PSA)	-2.44	9.8695E-08
BF1370	wcfM (PSA)	-2.45	7.08576E-08
BF1371	wcfN (PSA)	-2.45	3.10176E-08
BF0466	Putative protein	-2.48	1.65081E-07
BF1375	wcfQ (PSA)	-2.62	3.87573E-10
BF1368	upaZ (PSA)	-2.67	8.15223E-12
BF1367	upaY (PSA)	-2.75	6.50178E-13

Table S2.

Differentially expressed (adjusted p-value < 0.05) genes in mice mono-colonized with *B. fragilis* or *B. fragilis* Δ*ccf*. Genes are ordered by the absolute value of the fold-change.

Gene	**Log2(fold-change) (WT/Δ*ccf*)**	**Adjusted P-value**
Igkv4-61	-4.29	0.003125755
Igkv8-19	4.18	1.51E-06
Igkv3-12	-3.99	0.049083696
Ighv1-70	3.74	2.28E-07
Ighv1-5	-3.35	0.024434919
Gpr63	-2.44	0.038541457
Igkv6-23	2.03	3.13E-05
Cyr61	-1.83	6.56E-07
Ighv8-5	1.82	0.001789717
Nr4a1	-1.74	2.69E-10
Dusp1	-1.39	1.27E-07
Fos	-1.36	1.02E-06
Egr1	-1.10	1.19E-06
Sik1	-0.97	3.69E-05
Klf2	-0.90	0.017155274

Table S3.

DEseq analysis of exact sequence variants (ESVs) differing in abundance according to mouse genotype in either feces or the lumen and mucus of the ileum and colon of ex-germ-free mice.

Enriched in:	**Sample type**	**log2(fold-change)**	**Adjusted p-value**	**Best taxonomic assignment**
BALB/c	mucus and lumen	-8.634412444	0.000312846	Mollicutes RF39
BALB/c	mucus and lumen	-3.905449189	0.000551256	*Bacteroides fragilis*
BALB/c	mucus and lumen	-3.519511063	0.001725039	Lachnospiraceae
IgA$^{-/-}$	mucus and lumen	2.821434619	0.001983777	*Ruminococcus*
IgA$^{-/-}$	mucus and lumen	5.7119259	0.001221637	Clostridiales
IgA$^{-/-}$	mucus and lumen	5.039308072	2.76E-05	Lachnospiraceae
IgA$^{-/-}$	mucus and lumen	3.467202384	2.76E-05	*Ruminococcus gnavus*
IgA$^{-/-}$	mucus and lumen	7.097420438	4.84E-08	Alphaproteobacteria BD7-3
BALB/c	feces	-24.05671228	3.37E-23	Mollicutes RF39
BALB/c	feces	-7.354148271	1.01E-06	Lachnospiraceae
BALB/c	feces	-5.70390907	0.00543052	*Coprococcus*
IgA$^{-/-}$	feces	4.450515436	0.008991039	Clostridiales
IgA$^{-/-}$	feces	2.283274756	0.001376256	*Adlercreutzia*
IgA$^{-/-}$	feces	8.12403428	2.37E-08	Alphaproteobacteria BD7-3
IgA$^{-/-}$	feces	20.67955327	3.84E-20	Clostridiales

Table S4.

Bacterial strains and plasmids used in this study.

Strain / Plasmid	**Description**	**Source**
Bacteroides fragilis NCTC9343	Type strain, parent for all mutants in this study	ATCC
Bacteroides thetaiotaomicron ATCC 29148	Type strain	ATCC
Bacteroides vulgatus ATCC8482	Type strain	ATCC
B. fragilis Δ*ccf*	An in-frame deletion within the operon containing *ccfC*, *ccfD*, and *ccfE* (BF3581-79)	(*26*)
B. fragilis ΔPSC	An in-frame deletion of non-regulatory genes in the PSC locus (BF1011-26)	This study
B. fragilis ΔPSB	An in-frame deletion of non-regulatory genes in the PSB locus (BF1895-1914)	(*67*)
B. fragilis ΔPSB/C	ΔPSC mutation made in a ΔPSB background	This study
pFD340	*Escherichia coli* (Carbenicillin) / *Bacteroides* (Erythromycin) shuttle vector	(*68*)
pFD340-*ccfA*	Overexpression of the sigma factor *ccfA* using the IS4351 promoter on pFD340	This study
pFD340-chlor	Marker plasmid for mouse colonization experiments (Erythromycin, Chloramphenicol)	(*26*)
pFD340-tet	Marker plasmid for mouse colonization experiments (Erythromycin, Tetracycline)	(*26*)
pNJR6	Suicide plasmid for allelic exchange (Kanamycin for *E. coli*, Erythromycin for *B. fragilis*)	(*69*)

Table S5.

Primers used for cloning, generating mutants, and qPCR.

Primer	Sequence	Purpose	Source
ΔPSC primer-1	**GTGGATCC**AAATGCGTTGCTTTTGCTTT	Left flank PSC 5'	This study
ΔPSC primer-2	**AAACCATGG**TTCGAAATCGTTTTGCTTC A	Left flank PSC 3'	This study
ΔPSC primer-3	**GATTTCGAA**CCATGGTTTATGCTGGCTT T	Right flank PSC 5'	This study
ΔPSC primer-4	**TTGGATCC**AACACTACGCCTACCCGATG	Right flank PSC 3'	This study
PSC WT 1	GGAGGATGTTTGAATTGGTGG	PSC WT check 5'	This study
PSC WT 2	CCCGCTTAATGCCCTAAAAT	PSC WT check 3'	This study
PSC KO 1	GGAGGATGTTTGAATTGGTGG	PSC KO check 5'	This study
PSC KO 2	TATCCTGATGTTCTGCTTTTCCG	PSC KO check 3'	This study
ccfA 1	**AAGGATCC**TGCGCAACTGATATTGTTAG AA	*ccfA* (BF3583) cloning 5'	This study
ccfA 2	**AAGAGCTCC**GAAATCTACTCAGTGTAAA TGGA	*ccfA* (BF3583) cloning 3'	This study
q. gyrB 1	GTGAATGAGGACGGCAGTTT	qPCR gyrase	This study
q. gyrB 2	CTCGATGGGGATGTTTTGTT	qPCR gyrase	This study
q. PSA 1	TTGTATCCGCAAGGGAGAGA	qPCR PSA	This study
q. PSA 2	CGCTCCATACTGCCCATATT	qPCR PSA	This study
q. PSB 1	GCTTTTGGCTTAATGCTTGTTGG	qPCR PSB	This study
q. PSB 2	GCCTAGAAGTACAATTAGCCCGA	qPCR PSB	This study
q. PSC 1	TGTTTGGTGGCTGCTACTTG	qPCR PSC	This study
q. PSC 2	AGGTGAAGTTTGAAGCCAAGG	qPCR PSC	This study
q. PSD 1	CAATTTGGGAGGTGCGTTGT	qPCR PSD	This study
q. PSD 2	ACGACCAATCCAAAACCCCA	qPCR PSD	This study
q. PSE 1	TGCCTCCCTGTTGGTGAAAA	qPCR PSE	This study
q. PSE 2	AGCGTTAGCCAAACTCCGTA	qPCR PSE	This study
q. PSF 1	TTCTATCGTTCAGCGTGCGA	qPCR PSF	This study
q. PSF 2	TGCCCATACGCCAAATCCTT	qPCR PSF	This study
q. PSG 1	CAAGTACACCTGTCAGTAGTTTGC	qPCR PSG	This study
q. PSG 2	GCAACTTCCAATTCCTAACAAAAGA	qPCR PSG	This study
q. PSH 1	GGAAAACAGTCGGAATGGCTC	qPCR PSH	This study
q. PSH 2	TTCCACACACGCAGACACAA	qPCR PSH	This study
q. *ccfA* 1	GGAATTTGCATGACACTTAT	qPCR *ccfA*	(*26*)
q. *ccfA* 2	CTGAGAGGTTTCATCTTCTG	qPCR *ccfA*	(*26*)
q. *ccfB* 1	AGTGTCCCCACTTCATCGTC	qPCR *ccfB*	(*26*)
q. *ccfB* 2	TGAAACTTTTGCCGGAGAAT	qPCR *ccfB*	(*26*)
q. *ccfC* 1	GATGAACTGATAGCCCATTA	qPCR *ccfC*	(*26*)

q. *ccfC* 2	TAGCGATGACTAAAGGTGTT	qPCR *ccfC*	(*26*)
q. *ccfD* 1	CGGTTATATGCTTTTCAAAC	qPCR *ccfD*	(*26*)
q. *ccfD* 2	CAAATAGAAATCTGCCAAAC	qPCR *ccfD*	(*26*)

Materials and Methods

Bacterial strains, media, and plasmids

Bacteroides fragilis NCTC9343 was grown in Brain Heart Infusion (BD) supplemented with 5 μg/ml hemin (Frontier Scientific) and 5 μg/ml vitamin K1 (Sigma) in an anaerobic atmosphere of 80% nitrogen, 10% carbon dioxide, and 10% hydrogen. For selection, 200 μg/ml gentamicin, 10 μg/ml erythromycin, 10 μg/ml chloramphenicol, or 2 μg/ml tetracycline was added. For *in vivo* studies, *B. fragilis* carried a marker plasmid pFD340-Chlor or pFD340-Tet which confer resistance to chloramphenicol and tetracycline, respectively, allowing identification of multiple strains within an animal without affecting their fitness (*26*). PSC was deleted by allelic exchange with a suicide plasmid (pNJR6) as previously described (*26*). Details for all strains and plasmids are listed in Table S4 and primers for cloning in Table S5.

Mice

All mouse experiments were performed in accordance with the NIH Guide for the Care and Use of Laboratory Animals using protocols approved by the Institutional Animal Care and Use Committee at the California Institute of Technology. Swiss Webster, C57BL/6, *Rag1*$^{-/-}$, and BALB/c mice from Taconic Farms and BALB/c IgA$^{-/-}$ mice from the Baylor College of Medicine were C-section rederived germ-free and bred in flexible film isolators. Unless specifically stated, Swiss Webster mice were used for all experiments due to ease of breeding in gnotobiotic conditions. Eight-week-old germ-free mice were transferred to autoclaved microisolator cages supplied with autoclaved chow (LabDiet 5010) and autoclaved water containing 10μg/ml erythromycin (to select for marker plasmids) and 100μg/ml gentamicin (to which *Bacteroides* are naturally resistant). Mice were mono-

colonized by a single oral gavage of 10^8 CFU in 100µl of HBSS with 1.5% sodium bicarbonate and maintained for at least 4 weeks prior to any imaging, mucosal plating, gene expression, or colonization experiments. Colonization was confirmed for all mice and monitored over time using freshly collected fecal samples. Feces were weighed, mashed and vortexed in 1 ml BHI, spun at 400 g for 1 minute to pellet debris, and then diluted for CFU plating.

For depletion of B cells, 8-week-old germ-free Swiss Webster mice were injected intraperitoneally with 250 µg of monoclonal anti-mouse CD20 antibody (clone 5D2 from Genentech) or 250 µg of an isotype control (IgG2a from Bio X Cell). One week later, mice were mono-colonized. One week after that, the mice were given a second identical antibody injection. Four weeks after mono-colonization, the mice were used for horizontal transmission experiments.

Horizontal transmission and sequential gavage colonization assays

For horizontal transmission experiments, pairs of mono-colonized female mice were cohoused in a new autoclaved cage for 4 hours during the day (16 hours overnight for BALB/c mice, which exhibited a higher threshold for invasion), then individually housed in new autoclaved cages. For each experiment, the antibiotic resistance markers (Tet and Chlor) were swapped for half the animals (for example: WT-Tet vs. Δ*ccf*-Chlor, and then WT-Chlor vs. Δ*ccf*-Tet) to ensure that the difference in horizontal transmission was due to the bacterial or mouse genotype and not due to the different antibiotic resistance markers. The same assay was performed in specific pathogen free (SPF) BALBc mice by first treating with ciprofloxacin and metronidazole as described previously (*26*) to allow colonization with *B. fragilis*. For sequential gavage experiments, mice mono-colonized for

three weeks were gavaged with 10^8 CFU of the challenge strain in HBSS with 1.5% sodium bicarbonate and subsequently single-housed. The community-level comparison of BALB/c mice (Fig. 4H and I) was made by gavaging 8-10 week-old mice with 100 µl cecal contents (donors: 8 week-old Swiss Webster mice delivered from Taconic, 2 male and 2 female cecums dissected and pooled) homogenized to saturation in HBSS with 1.5% sodium bicarbonate. Mucus and lumen samples were collected 5 weeks after colonization.

Sample preparation for electron microscopy

A 1 cm portion of ascending colon from mono-colonized mice was excised and immediately fixed with an ice-cold solution of 3% glutaraldehyde, 1% paraformaldehyde, 5% sucrose in 0.1 M sodium cacodylate trihydrate, without flushing of the intestinal contents. Tissues were prefixed for 1 hour at 4°C, then transferred to a petri dish containing 5% sucrose in 0.1 M cacodylate buffer. Tissues were cut into ~1-2 mm^3 blocks with a #11 scalpel and placed into brass planchettes (Type A; Ted Pella, Inc., Redding, CA), prefilled with cacodylate buffer containing 10% Ficoll (70 kD, Sigma-Aldrich) which serves as an extracellular cryoprotectant. Excess buffer was removed with Whatman filter paper and the sample covered with a Type B brass planchette. Samples were ultra-rapidly frozen with a HPM-010 high-pressure freezing machine (Leica Microsystems, Vienna Austria), then transferred immediately under liquid nitrogen to cryotubes (Nunc) containing a frozen solution of 2.5% osmium tetroxide, 0.05% uranyl acetate in acetone. The tubes were loaded into a AFS-2 freeze-substitution machine (Leica Microsystems) precooled to -100°C. Samples were processed at -90°C for 72 hours, warmed over 12 hours to -20°C, held at that temperature for 6-10 hours, and then warmed to 4°C for 1 hour. The fixative was removed and the samples rinsed four times with cold acetone, infiltrated into Epon-

Araldite resin (Electron Microscopy Sciences, Port Washington PA), and flat-embedded between two Teflon-coated glass slides. Resin was polymerized at 60°C for 48 hours.

Electron microscopy and dual-axis tomography

Flat-embedded colon samples were observed with a stereo dissecting microscope to ascertain preservation quality and select appropriate regions for EM study. These were extracted with a microsurgical scalpel and glued to the tips of plastic sectioning stubs. Semi-thick (300 or 400 nm) serial sections were cut with a UC6 ultramicrotome (Leica Microsystems) using a diamond knife (Diatome, Ltd. Switzerland). Sections were placed onto Formvar-coated copper-rhodium 1 mm slot grids (Electron Microscopy Sciences) and stained with 3% uranyl acetate and lead citrate. Gold beads (10 nm) were placed on both surfaces of the grid to serve as fiducial markers for subsequent image alignment. Grids were placed in a dual-axis tomography holder (Model 2040, E.A. Fischione Instruments, Export PA) and imaged with a Tecnai TF30ST-FEG transmission electron microscope (300 KeV; FEI Company/ThermoFisher Scientific, Hillsboro OR) equipped with an XP1000 2k x 2k CCD camera (Gatan, Inc. Pleasanton CA). Tomographic tilt-series and large-area montaged overviews were acquired automatically using the SerialEM software package (*49*). For tomography, samples were tilted +/- 64° and images collected at 1° intervals. The grid was then rotated 90° and a similar series taken about the orthogonal axis. Tomographic data was calculated, analyzed, and modeled using the IMOD software package (*50*, *51*). For quantification of epithelial associated bacteria (Fig. 1F and 4J), sections were surveyed until a bacterium on the apical epithelial surface was identified, and then a ~20 x 20 μm montaged projection image was acquired. Bacterial cells were identified by morphology and counted. About 1 mm length of epithelium was scanned per

mouse, including an area of around 20 µm from the apical epithelial surface into the mucus. For measurement of the capsule thickness (Fig. 1I), the distance between the outer membrane and edge of the electron-dense capsule layer was measured in tomograms. Ten different measurement sites on the same cell were averaged for each cell.

Mucosal scraping for CFU quantification or sequencing analysis

The ascending colon was dissected and cut open longitudinally. Lumen content was removed, and the flat tissue was washed with HBSS until no feces could be observed. A 1 cm length of the tissue was cut, making an approximately 1 cm x 1 cm square of flat tissue. Mucus was removed from the epithelial surface by holding one corner of the tissue and using light pressure with a sterile plastic 1.8 cm cell scraper (BD Falcon). For sequencing analysis, mucus was frozen at this point. For CFU quantification, the mucus was moved into a 2 ml screw cap tube with 1.4 mm ceramic beads (Lysing Matrix D, MP Biomedicals) and 1 ml of HBSS and bead beat on medium for 1 minute to homogenize the mucus. This homogenate was diluted for CFU plating.

RNA isolation

Pelleted mid-log bacterial cultures, freshly harvested colon lumen contents, or whole colon tissue were immediately lysed by bead-beating in a mixture of 500 µl buffer (0.2 M NaCl and 20 mM EDTA), 210 µl 20% SDS, and 500 µl phenol, chloroform, and isoamyl alcohol (Ambion AM9720). The aqueous phase was separated by centrifugation and moved to a new tube for a second extraction with phenol, chloroform, and isoamyl alcohol. Then, 50 µl of 3 M sodium acetate and 500 µl of cold ethanol were mixed in to the aqueous fraction and placed on ice for 20 minutes. RNA was pelleted and washed once with cold 70% ethanol, and then resuspended in 100 µl water. RNA was further purified using the Qiagen

RNeasy mini kit according to the manufacturer instructions. DNA was removed using Turbo DNase (Ambion AM2238) for one hour at 37 C before applying to a second Qiagen RNeasy column, including an on-column Qiagen RNase-free DNase digest.

qRT-PCR

First-strand cDNA synthesis from total RNA was achieved using the manufacturer's instructions for the iScript cDNA synthesis kit (Bio-Rad). Reactions were run on an ABI PRISM 7900HT Fast Real-Time PCR System (Applied Biosystems) with Power SYBR Green PCR Master Mix (Applied Biosystems). Relative quantification was performed using the ΔΔCt method with *gyrB* (DNA gyrase) as the housekeeping control gene. The mean Ct value from 3 technical replicates was used for each biological replicate. Primers are listed in Table S5.

RNAseq

For bacterial RNAseq (Fig. 2A), 10 ml liquid cultures were harvest at mid-log phase and total RNA was isolated as described above (RNA isolation). Libraries were prepared using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB E7530) according to the manufacturer's instructions and sequenced on an Illumina HiSeq2500 in single read rapid run mode with single-end 50 bp reads. For each sample, at least 2.5 million reads mapped to protein-coding genes. The bacterial RNAseq analysis package Rockhopper (*52*) was used for quality filtering, mapping, and differential expression analysis.

For mouse RNAseq (Fig. 3A), a 1 cm length of opened and washed ascending colon tissue was prepped for total RNA as described above (RNA isolation). Ribosomal RNA was depleted using a Ribo-Zero Magnetic Gold Epidemiology Kit (Epicentre/Illumina).

Sequencing was done on an Illumina HiSeq2500 to produce 101-bp paired-end reads. Reads were trimmed with a Phred quality score cut-off of 20 using fastq_quality_trimmer from the FASTX toolkit, version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). Reads shorter than 20 bp after adaptor and poly(A)-trimming were discarded. Trimmed reads were aligned to the mouse genome (build GRCm38.p4) by STAR (*53*), read counts were quantified using HTseq (*54*), and differentially expressed genes were identified using edgeR (*55*).

Bacterial flow cytometry

Fresh fecal pellets were mashed and vortexed in 1 ml of HBSS per 100 mg feces. Large debris was removed by centrifugation at 400 x g for 5 minutes. Supernatants, containing fecal bacteria, were diluted to 10^6 CFU per well in 96-well v-bottom plates, washed with HBSS, and stained with a monoclonal rat anti-mouse IgA conjugated to PE (eBioscience clone mA-6E1) or monoclonal rat anti-mouse IgM conjugated to PE (Biolegend clone RMM-1) at 1:250 and SYTO 9 nuclear stain (Molecular Probes) at 1:1000. After washing twice with HBSS, stained bacteria were analyzed on a MACSQuant VYB (Miltenyi). Data were analyzed using FlowJo.

For IgA affinity studies, IgA was extracted from fresh fecal samples by mashing and vortexing in 1 ml HBSS per 100 mg feces, then removing the insoluble fraction by centrifuging at 16,000 x g for 10 minutes at 4 C, and collecting the supernatant containing soluble IgA. The pellet was resuspended in the same volume for a second extraction. The combined supernatant was then sterile filtered, heat inactivated at 56 C for 30 minutes, and then stored at 4 C for up to 1 month (degradation was not observed by ELISA in this timeframe). IgA concentrations were measured using a Ready-SET-Go mouse IgA ELISA

kit (eBioscience) according to the manufacturers protocol. Binding assays were performed on fecal-derived bacteria from mono-colonized *Rag1*$^{-/-}$ mice so that the bacteria were *in vivo*-adapted (in which *ccf* is induced and the thick capsule is formed), but IgA-free (because *Rag1*$^{-/-}$ mice do not make antibody). 10^6 CFU per well in 96-well plates was incubated on a shaker with 200 µl IgA at a concentration of 1 µg / ml for 30 minutes and then stained for flow cytometry.

Western blot

Fresh fecal pellets from *Rag1*$^{-/-}$ mice were mashed and vortexed in 1 ml of BHI per 100 mg feces. Large debris was removed by centrifugation at 400 x g for 5 minutes. This supernatant or late-log cultures of bacteria were pelleted at 8,000 g and resuspended at 10X concentration in BHI. The concentrate was mixed 1:1 with 2X Laemmli sample buffer and boiled for 10 minutes to lyse bacteria. 20 µl of lysate (10^8 CFU) was loaded per well in tris-glycine gels with a 4-20% polyacrylamide gradient (Novex WedgeWell), ran in a tris-glycine buffer with 0.1% SDS (Bio-Rad). Gels were transferred at 4 C overnight to a PVDF membrane in the same buffer supplemented with 20% methanol. PVDF blots were blocked for 1 hour in PBS-T (PBS with 0.1% Tween20) with 5% powdered milk, washed with PBS-T, and then stained with IgA, extracted from feces as described above, at a final concentration of 0.1 µg / ml in PBS-T with 2.5% powdered milk for 2 hours. Blots were washed four times with PBS-T, stained with an HRP goat anti-mouse IgA (Southern Biotech 1040-05) diluted 1:1000 in PBS-T with 2.5% powdered milk for 1 hour, washed four times again, and then developed with a Clarity western ECL substrate (Bio-Rad).

Tissue-cultured epithelial cell adherence assay

Intestinal epithelial cell lines HT29, HT29-MTX, and Caco-2 were maintained in high glucose DMEM with 4 mM L-glutamine, 4.5 g / L glucose (HyClone), supplemented with 10% fetal bovine serum (FBS) (Gibco), and penicillin-streptomycin solution (Corning). Cells were seeded in flat 96-well plates at a density of 25,000 cells per well and then grown for seven days past confluence to allow production of mucus. IgA was extracted and bound to fecal-derived bacteria from mono-colonized *Rag1*$^{-/-}$ mice so that the bacteria were *in vivo*-adapted but IgA-free as described above (Bacterial flow cytometry for IgA coating and affinity). IgA-coated bacteria were pelleted and resuspended in high-glucose DMEM without antibiotics or FBS. Epithelial cells were washed once with HBSS and then incubated with 10^6 coated bacteria for 2 hours at 37 C in an anaerobic atmosphere of 80% nitrogen, 10% carbon dioxide, and 10% hydrogen to allow adherence of bacteria. Bacterial growth was not observed within this timeframe either on epithelial cells or in media alone. Wells were washed twice with 200 µl HBSS to remove unbound bacteria, and then trypsinized with 50 µl of 0.05% trypsin in HBSS (Corning) for 20 minutes at 37 C to disassociate epithelial cells. Then 50 µl of BHI was added and cells were vigorously resuspended before dilution plating for CFU. Fraction bound was calculated as the output CFU / input CFU. For all epithelial cell adherence assays, the average of 4 technical replicates (4 separate binding reactions with the same IgA and bacteria on 4 wells of cells) is reported for each biological replicate.

B cell flow cytometry

Cells were isolated from the peritoneum, mesenteric lymph nodes, and colon tissues. The peritoneal space was injected with PBS and massaged for thirty seconds to suspend cells. Mesenteric lymph nodes were dissected and single cell suspensions were obtained by

grinding tissues through a 100 μm cell strainer. Colonic lamina propria lymphocytes were isolated by first flushing lumen contents away with PBS, and removing the longitudinal and circular muscle layers of the colon by micro-dissection. The remaining tissue was incubated in 1 mg / mL Collagenase II (Dibco) for 1 hour and then cells were filtered through a 100 μm cell strainer. Cells were spun down, resuspended and separated by a 40%/80% (v/v) Percoll (GE Healthcare) density gradient. All cells were washed in complete RPMI buffer, allowed to rest at 37 C for at least 1 hour in a 5 % CO_2 tissue culture incubator, and incubated in 5% mouse serum for 15 minutes before proceeding to staining for flow cytometry. Cells were stained with a LIVE/DEAD fixable Violet dye (Life Technologies), FITC-conjugated B220 antibody (eBioscience clone RA3-6B2), and PerCP-Cy5.5-conjugated CD3e antibody (eBioscience clone 145-2C11) for 20 minutes at 4 C. Cells were washed in HBSS with 1% BSA and fixed in 2% PFA prior to acquisition on a MACSQuant Analyzer (Miltenyi). Data were analyzed using FlowJo.

Sorting IgA-bound bacteria for sequencing

This method was carried out as previously described (*35*). Two fecal pellets were mashed and resuspended in 1 ml PBS with 0.5% BSA (this buffer used throughout the protocol). The homogenate was centrifuged at 400 g for 5 minutes to remove debris, and the supernatant was passed through a 30 μm filter. Bacteria were pelleted by centrifugation at 8000 g for 5 minutes and resuspended in 1 ml buffer plus 35 μl of anti-IgA PE (eBioscience clone mA-6E1) and stained for 20 minutes at 4 C. Stained bacteria were pelleted, washed once, then resuspended in 1 ml buffer plus 100 μl ultrapure anti-PE magnetic beads (Miltenyi) and incubated for 15 minutes at 4 C. Bead-bound, stained bacteria were pelleted and resuspended in 500 μl buffer and then run over an MS column (Miltenyi) on an

OctoMACS separator (Miltenyi). Bound bacteria were eluted and re-loaded on the column three times to further purify.

16S amplicon sequencing

DNA was extracted from tissue scrapings, fecal samples, and IgA-sorted bacterial pellets using the Qiagen PowerMag Soil extraction kit and a ThermoFisher KingFisher magnetic bead purification robot. Extracted DNA was amplified and sequenced as in Caporaso et al (*56*). Briefly, the V4 region of the 16S ribosomal RNA gene was amplified in triplicate using the 515F/805R primers from the Earth Microbiome Project (*57*) and amplicons pooled and sequenced on an Illumina MiSeq instrument. Sequences were uploaded to the Qiita analysis platform for demultiplexing and primary data processing (https://qiita.ucsd.edu/). Exact sequence variants (ESVs) were identified and chimeric sequences removed using the Deblur plugin in Qiita (*58*). For phylogenetic analyses, ESVs were inserted into the GreenGenes reference phylogeny (release 13_8, (*59*))using SEPP (*60*). Taxonomy was assigned per ESV according to its placement in the GreenGenes phylogeny using the fragment-insertion pluging in Qiime 2 v2017-12 (https://qiime2.org). Samples yielding less than 1000 Deblurred sequences were excluded from subsequent analysis.

16S diversity analyses

We used the Galaxy implementation LEfSe (*61*) to identify taxonomic groups that were significantly enriched (LDA effect size ≥ 2) in bacterial pellets purified using anti-IgA antibodies, coding pellet sort status as a class and mouse sex as subclass and normalizing to 10^6 counts per sample. To identify specific ESVs differentially represented in tissues or feces of BALB/c and BALB/c $IgA^{-/-}$ mice, we used the DESeq2 (*62*) in R (*63*), adding

a pseudocount of 1 to all ESVs and fitting models using a negative binomial distribution. Analyses of microbial community beta-diversity were performed using the Weighted UniFrac metric (*64*) calculated from count tables rarefied to 1000 sequences in Qiime2. Principle Coordinate Analyses and Adonis tests of group differences in central tendency (*65*) were performed in the Vegan package in R (*66*).

Statistical methods

All bar graphs show mean values with standard error of measurement. Log-scale line graphs for horizontal transmission and sequential gavage assays show the geometric mean and 95% confidence interval. Post-hoc corrections for multiple comparisons were the Tukey method when comparing all groups with each other and Sidak for other types of comparisons.

Acknowledgments

We thank Drs. Elaine Hsiao, June Round, Hiutung Chu, and members of the Mazmanian laboratory for critical review of this manuscript. The anti-CD20 antibody was provided under an MTA from Genentech. We appreciate technical support from Taren Thron, the Caltech Office of Laboratory Animal Resources, Caltech Genomics Laboratory, and Caltech Flow Cytometry Facility. G.P.D. was supported by an NIH training grant (5T32 GM07616) and NSF Graduate Research Fellowship (DGE-1144469). Support for this research was provided by grants from the National Institutes of Health to the Broad Institute (U19AI110818), the NIH (P50 GM082545 and AI04123) to P.B.J., the NIH (GM099535 and DK083633) to S.K.M., and the Heritage Medical Research Institute to S.K.M. All data and code to understand and assess the conclusions of this research are available in the main text, supplementary materials, and via the following repositories: EMBL-EBI accession ERP107727 and NCBI Bioproject accessions PRJNA445716 and PRJNA438372.

References

1. A. B. Hall, A. C. Tolonen, R. J. Xavier, Human genetic variation and the gut microbiome in disease. *Nat. Rev. Genet.* (2017), doi:10.1038/nrg.2017.63.

2. T. C. Fung, C. A. Olson, E. Y. Hsiao, Interactions between the microbiota, immune and nervous systems in health and disease. *Nat. Neurosci.* **20**, 145–155 (2017).

3. J. J. Faith *et al.*, The long-term stability of the human gut microbiota. *Science*. **341**, 1237439 (2013).

4. D. T. Truong, A. Tett, E. Pasolli, C. Huttenhower, N. Segata, Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.*, gr.216242.116 (2017).

5. S. Fagarasan *et al.*, Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora. *Science*. **298**, 1424–1427 (2002).

6. S. Kawamoto *et al.*, The inhibitory receptor PD-1 regulates IgA selection and bacterial composition in the gut. *Science*. **336**, 485–489 (2012).

7. A. J. Macpherson, Y. Köller, K. D. McCoy, The bilateral responsiveness between intestinal microbes and IgA. *Trends Immunol.* **36**, 460–470 (2015).

8. J. L. Kubinak, J. L. Round, Do antibodies select a healthy microbiota? *Nat. Rev. Immunol.* **16**, 767–774 (2016).

9. J. L. Kubinak *et al.*, MyD88 signaling in T cells directs IgA-mediated control of the microbiota to promote health. *Cell Host Microbe*. **17**, 153–163 (2015).

10. F. Fransen *et al.*, BALB/c and C57BL/6 Mice Differ in Polyreactive IgA Abundance, which Impacts the Generation of Antigen-Specific IgA and

Microbiota Diversity. *Immunology*. **43**, 527–540 (2015).

11. S. Kawamoto *et al.*, Foxp3(+) T cells regulate immunoglobulin a selection and facilitate diversification of bacterial species responsible for immune homeostasis. *Immunology*. **41**, 152–165 (2014).

12. N. J. Mantis, S. J. Forbes, Secretory IgA: arresting microbial pathogens at epithelial borders. *Immunol. Invest.* **39**, 383–406 (2010).

13. L. A. van der Waaij, P. C. Limburg, G. Mesander, D. van der Waaij, In vivo IgA coating of anaerobic bacteria in human faeces. *Gut*. **38**, 348–354 (1996).

14. K. E. Shroff, K. Meslin, J. J. Cebra, Commensal enteric bacteria engender a self-limiting humoral mucosal immune response while permanently colonizing the gut. *Infect. Immun.* **63**, 3904–3913 (1995).

15. R. R. Bollinger *et al.*, Human secretory immunoglobulin A may contribute to biofilm formation in the gut. *Immunology*. **109**, 580–587 (2003).

16. A. Mathias *et al.*, Potentiation of polarized intestinal Caco-2 cell responsiveness to probiotics complexed with secretory IgA. *J Biol Chem*. **285**, 33906–33913 (2010).

17. L. Yel, Selective IgA deficiency. *Journal of clinical immunology* (2010), doi:10.1007/s10875-009-9357-x.

18. M. E. Conley, D. D. A. O. I. medicine, 1987, Intravascular and mucosal immunoglobulin A: two separate but related systems of immune defense? *Am Coll Physicians*, doi:10.7326/0003-4819-106-6-892.

19. S. K. Mazmanian, J. L. Round, D. L. Kasper, A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature*. **453**, 620–625 (2008).

20. J. Ochoa-Repáraz *et al.*, Central nervous system demyelinating disease protection by the human commensal Bacteroides fragilis depends on polysaccharide A expression. *J. Immunol.* **185**, 4101–4108 (2010).

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

22. H. Chu *et al.*, Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science*. **352**, 1116–1120 (2016).

23. M. Scholz *et al.*, Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat. Methods*. **13**, 435–438 (2016).

24. M. Yassour *et al.*, Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Science Translational Medicine*. **8**, 343ra81–343ra81 (2016).

25. K. Yasuda *et al.*, Biogeography of the intestinal mucosal and lumenal microbiome in the rhesus macaque. *Cell Host Microbe*. **17**, 385–391 (2015).

26. S. M. Lee *et al.*, Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature*. **501**, 426–429 (2013).

27. N. A. Pudlo *et al.*, Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans. *MBio*. **6** (2015), doi:10.1128/mBio.01282-15.

28. E. C. Martens, R. Roth, J. E. Heuser, J. I. Gordon, Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J Biol Chem*. **284**, 18445–18457 (2009).

29. M. J. Coyne, M. Chatzidaki-Livanis, L. C. Paoletti, L. E. Comstock, Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont Bacteroides fragilis. *Proc Natl Acad Sci USA*. **105**, 13099–13104 (2008).

30. C. H. Liu, S. M. Lee, J. M. Vanlare, D. L. Kasper, S. K. Mazmanian, Regulation of surface architecture by symbiotic bacteria mediates host colonization. *Proc Natl Acad Sci USA*. **105**, 3951–3956 (2008).

31. D. A. Peterson, N. P. McNulty, J. L. Guruge, J. I. Gordon, IgA response to symbiotic bacteria as a mediator of gut homeostasis. *Cell Host Microbe*. **2**, 328–339 (2007).

32. N. L. Zitomersky, M. J. Coyne, L. E. Comstock, Longitudinal analysis of the prevalence, maintenance, and IgA response to species of the order Bacteroidales in the human gut. *Infect. Immun.* **79**, 2012–2020 (2011).

33. K. Moor *et al.*, Analysis of bacterial-surface-specific antibodies in body fluids using bacterial flow cytometry. *Nat Protoc*. **11**, 1531–1553 (2016).

34. M. Shimoda, Y. Inoue, N. Azuma, C. Kanno, Natural polyreactive immunoglobulin A antibodies produced in mouse Peyer's patches. *Immunology*. **97**, 9–17 (1999).

35. J. J. Bunker *et al.*, Natural polyreactive IgA antibodies coat the intestinal microbiota. *Science*. **358**, eaan6619 (2017).

36. H. L. Gibbins, G. B. Proctor, G. E. Yakubov, S. Wilson, G. H. Carpenter, SIgA binding to mucosal surfaces is mediated by mucin-mucin interactions. *PLoS ONE*. **10**, e0119677 (2015).

37. A. R. Biesbrock, M. S. Reddy, M. J. Levine, Interaction of a salivary mucin-

secretory immunoglobulin A complex with mucosal pathogens. *Infect. Immun.* **59**, 3492–3497 (1991).

38. A. Phalipon *et al.*, Secretory component: a new role in secretory IgA-mediated immune exclusion in vivo. *Immunology*. **17**, 107–115 (2002).

39. G. Sarikonda *et al.*, Transient B-cell depletion with anti-CD20 in combination with proinsulin DNA vaccine or oral insulin: immunologic effects and efficacy in NOD mice. *PLoS ONE*. **8**, e54712 (2013).

40. S. E. Blutt, A. D. Miller, S. L. Salmon, D. W. Metzger, M. E. Conner, IgA is important for clearance and critical for protection from rotavirus infection. *Mucosal Immunol*. **5**, 712–719 (2012).

41. K. Moor *et al.*, High-avidity IgA protects the intestine by enchaining growing bacteria. *Nature*. **103**, 3–19 (2017).

42. K. Suzuki *et al.*, Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc Natl Acad Sci USA*. **101**, 1981–1986 (2004).

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

44. J. Mirpuri *et al.*, Proteobacteria-specific IgA regulates maturation of the intestinal microbiota. *Gut Microbes*. **5**, 28–39 (2014).

45. K. Singh, C. Chang, M. E. Gershwin, IgA deficiency and autoimmunity. *Autoimmun Rev*. **13**, 163–177 (2014).

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

47. A. L. Kau *et al.*, Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. *Science Translational Medicine*. **7**, 276ra24–276ra24 (2015).

48. K. McLoughlin, J. Schluter, S. Rakoff-Nahoum, A. L. Smith, K. R. Foster, Host Selection of Microbiota via Differential Adhesion. *Cell Host Microbe*. **19**, 550–559 (2016).

49. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

50. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

51. D. N. Mastronarde, Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the IMOD package. *J Microsc*. **230**, 212–217 (2008).

52. R. McClure *et al.*, Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res*. **41**, e140 (2013).

53. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

54. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).

55. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2010).

56. J. G. Caporaso *et al.*, Ultra-high-throughput microbial community analysis on

the Illumina HiSeq and MiSeq platforms. *ISME J*. **6**, 1621–1624 (2012).

57. L. R. Thompson *et al.*, A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. **551**, 457–463 (2017).

58. A. Amir *et al.*, Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*. **2** (2017), doi:10.1128/mSystems.00191-16.

59. T. Z. DeSantis *et al.*, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*. **72**, 5069–5072 (2006).

60. S. Mirarab, N. Nguyen, T. Warnow, SEPP: SATé-enabled phylogenetic placement. *Pac Symp Biocomput*, 247–258 (2012).

61. J. N. Paulson, O. C. Stine, H. C. Bravo, M. Pop, Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods*. **10**, 1200–1202 (2013).

62. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

63. R. C. Team, R: A language and environment for statistical computing (2013).

64. C. Lozupone, R. Knight, UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*. **71**, 8228–8235 (2005).

65. M. J. Anderson, A new method for non-parametric multivariate analysis of variance. *Austral Ecology*. **26**, 32–46 (2001).

66. J. Oksanen, R. Kindt, P. Legendre, B. O'Hara, *The Vegan Package—Community Ecology Package. R package version 2.0-9* (2007).

67. M. J. Coyne, W. Kalka-Moll, A. O. Tzianabos, D. L. Kasper, L. E. Comstock,

Bacteroides fragilis NCTC9343 produces at least three distinct capsular polysaccharides: cloning, characterization, and reassignment of polysaccharide B and C biosynthesis loci. *Infect. Immun.* **68**, 6176–6181 (2000).

68. C. J. Smith, M. B. Rogers, M. L. McKee, Heterologous gene expression in Bacteroides fragilis. *Plasmid*. **27**, 141–154 (1992).

69. A. M. Stevens, N. B. Shoemaker, A. A. Salyers, The region of a Bacteroides conjugal chromosomal tetracycline resistance element which is responsible for production of plasmidlike forms from unlinked chromosomal DNA might also be involved in transfer of the element. *Journal of Bacteriology*. **172**, 4271–4279 (1990).

C h a p t e r 3

HYBRID SELECTION RNA-SEQ REVEALS MUCOSAL COLONIZATION FACTORS IN *BACTEROIDES FRAGILIS*

Donaldson G.P., Chou W.C., Manson A.L., Abeel T., Ciulla D.M., Rogov P., Bochicchio J., Giannoukos G., Melnikov A., Earnst P.B., Chu H., Earl A.M., and Mazmanian S.K.

To be submitted for publication.

ABSTRACT

Gut microbiota inhabit a variety of microenvironments within the gastrointestinal tract, and a subset of this community specializes in colonizing the epithelial surface. Transcriptomic analysis of bacteria within host-associated niches is challenging due to the overabundance of host RNA. We applied hybrid selection RNA-seq to compare the transcriptome of *Bacteroides fragilis* in the colon lumen, mucus and tissue of mono-colonized mice. Hybrid selection increased reads mapping to the *B. fragilis* genome by 48 and 154-fold in mucus and tissue, allowing for high fidelity comparisons across sample sites. In the mucus and tissue, *B. fragilis* up-regulated many genes involved in protein synthesis, indicating these bacteria are active and thriving in the mucosal niche. Rather than broad changes in metabolic pathways, a specific sulfatase and glycanase were highly induced in mucus and tissue. Mutants in these individual genes exhibited a mucus-specific defect in colonization of the mouse colon, which was disadvantageous during competitive colonization experiments. Furthermore, the mucosal glycanase was required for *B. fragilis*-mediated immunomodulation and protection from colitis. Therefore, colonization of the colon mucus is mutually beneficial for this bacterial symbiont and its mammalian host.

INTRODUCTION

The mammalian gastrointestinal tract hosts a microscopic ecosystem of bacteria, protists, fungi, and viruses. For the host, this community is critical to digestion, nutrient extraction, and inhibition of pathogen colonization. Though most microbiome profiling relies on homogenized fecal samples, the gut community is spatially structured in important ways[1]. Notably, mucus secreted from intestinal goblet cells provides both a physical barrier that limits microbial colonization of the intestinal surface[2] and a carbon/energy source for the indigenous microbiome[3-5]. Accordingly, the mucus and lumen of the intestines exhibit distinct community compositions[6-9]. Some bacteria also directly associate with the intestinal surface, such as segmented filamentous bacteria[10,11], adherent *Lactobacillus*[12], and a community of crypt-resident bacteria[13-15]. Generally, studying the relevance of spatial structure is a challenge because of the dynamic nature of the gut and complexity of the community. Imaging studies can provide valuable snapshots of gut organization[16,17], but the functional relevance of localization is difficult to infer. Mechanisms and implications of mucosal colonization by indigenous gut bacterial species remain relatively poorly defined.

Attempts to investigate bacterial transcriptomics in host tissues or mucus are complicated by the miniscule quantity of bacterial RNA present compared to host RNA. Separating bacterial cells from host cells can be done by size at the cellular level, prior to RNA extraction, but this risks bacterial stress responses during the procedure. As typical bacterial mRNAs have *in vivo* half-lives of just a few minutes[18], rapid dissection and RNA stabilization is critical to high fidelity transcriptomics. This favors an alternative strategy: to separate bacterial transcripts at the molecular level, following nucleic acid extraction. Here we used hybrid

selection to enrich bacterial cDNA out of total cDNA preparations using biotinylated probes complementary to the whole bacterial genome. Host cDNA can then be washed away, allowing captured bacterial cDNA to be eluted and amplified for sequencing (Fig. 1a). This method was originally developed for the purification of re-sequencing targets in the human genome[19] and was then adapted to enrich pathogen DNA in clinical samples dominated by human genetic material[20,21]. We used a unified protocol to isolate total RNA from proximal colon lumen, mucus, and tissue samples from gnotobiotic mice mono-colonized with *Bacteroides fragilis* (Fig. 1a). *B. fragilis* is a mammalian gut symbiont that binds mucin[22], penetrates colonic crypts of Lieberkuhn in mice[23], and is enriched in the mucosal population in primates[8].

RESULTS

Hybrid selection enriches bacterial RNA is mucus and tissue samples

To test this enrichment method on murine gut samples, we performed RNA sequencing (RNA-seq) with and without hybrid selection (Supplementary Table 1). Without hybrid selection, bacterial RNA represented 50% of the total RNA in the lumen, but only 0.6% and 0.1% of total RNA in mucus and tissue samples, respectively (Fig. 1b). After hybrid selection, we observed a reduction in unaligned reads and reads mapping to the mouse genome, with a corresponding increase in reads mapping to the *B. fragilis* genome (Supplementary Table 1). The percentage of total reads mapping to the *B. fragilis* genome increased 48 and 154-fold in the mucus and tissue, respectively (Figure 1b). This striking enrichment resulted in three times more coverage over *B. fragilis* genes in the mucus sample, and almost ten times more gene coverage in tissue samples (Figure 1c). Gene expression

levels between the hybrid selected and non-hybrid selected samples were highly correlated within sample sites (Supplementary Fig. 1), indicating that the procedure did not globally skew the transcriptome.

Hybrid selection also improved our ability to perform comparisons across microenvironments. Limited bacterial transcript coverage in the mucus and tissue without hybrid selection led to a large number of genes with near-zero expression levels (555 and 1034, respectively, Figure 1d). For most of these, we were able to measure a higher expression value in the hybrid-selected samples (159 and 299 remained with near-zero expression values in mucus and tissue, respectively, after hybrid selection). Assuming that the majority of transcripts should be unchanged across sample sites, correlation of the more reliable lumen transcriptome with the host-associated samples can be used as a measure of fidelity. Hybrid selection substantially improved the correlation of gene expression levels between lumen and mucus as well as lumen and tissue (Figure 1d). As a result, hybrid selection greatly increased the number of genes identified in a differential expression analysis between different microenvironments (Figure 1e). This method greatly enriched bacterial transcripts in host-associated sample types without skewing the transcriptome, facilitating spatial comparisons within the gut.

Differentially expressed genes between gut micro-environments

We proceeded to analyze differentially expressed genes between lumen, mucus, and tissue samples with hybrid selection (Fig. 2, Supplementary Fig. 2). As compared to the lumen, in the mucus 26 genes were significantly up-regulated and 42 were down-regulated (Figure 2a,

Supplementary Table 2). Also compared to the lumen, in the tissue 52 genes were up-regulated and 47 were down-regulated (Figure 2a, Supplementary Table 3). No genes had significantly different expression values between mucus and tissue. Of the differentially expressed genes, 37 changed in the same direction for both mucus and tissue, while 31 genes significantly changed only in mucus and 62 changed significantly only in tissue. We did not observe broad shifts in metabolic or other pathways, though the Pfam[24] for glycosyl transferase group 1 was slightly overrepresented in the lumen (Supplementary Table 4).

We examined genes down-regulated in the mucus and tissue to assess the relative behavior of *B. fragilis* in the fecal stream. A universal stress gene, *uspA* (BF2495), was more highly expressed in the lumen than mucus (Supplementary Table 2b). Interestingly, an adenine-specific DNA methyltransferase (BF1252) was the most highly expressed in lumen compared to tissue (Fig. 2b). DNA methylation in bacteria is widespread and important for genome protection[25]. Furthermore, two bacterial histone-like proteins (BF3379, BF4220) were also more highly expressed in the lumen (Supplemental Tables 2b and 3b), which may also be involved in genome protection[26]. Compared to the mucus and tissue, the behavior of *B. fragilis* in the lumen points toward persistence rather than growth.

Another one of the most down-regulated genes in mucus and tissue was the flippase transporter (BF0737) for Polysaccharide G (PSG) (Fig. 2b). Four other genes in the PSG biosynthesis locus were also significantly down-regulated during host association (Supplementary Fig. 2a, Supplementary Tables 2b and 3b), and the biosynthesis locus for PSG as a whole trended toward lumen enrichment (Supplementary Fig. 2c). A characteristic

feature of *Bacteroides* species is the presence of multiple biosynthesis loci for capsular polysaccharides. *B. fragilis* has 8 such loci, and its ability to switch between multiple capsules contributes to fitness in the gut[27,28]. But the other 7 capsular polysaccharide biosynthesis loci (PSA-PSF and PSH) were not differentially expressed. This suggests that PSG has a unique lumen-specific role in the physiology of *B. fragilis*. Taken together with the stress response and genome protection, perhaps this prepares the bacteria for survival outside of the host.

The genes up-regulated in the mucus and tissue support the theory that *B. fragilis* is adapted to living on the intestinal surface. Though anaerobic, *B. fragilis* is actually well-equipped[29] to contend with reactive oxygen species emanating from the host[9]. Indeed, both subunits of alkyl hydroperoxide reductase (a reactive oxygen species resistance enzyme), *ahpC* (BF1210) and *ahpF* (BF1209)[30], were induced in mucus and tissue (Fig. 2c, Supplementary Tables 2a and 3a). Interestingly, 14 tRNAs were up-regulated in the mucus and 26 in the tissue whereas no tRNAs were more highly expressed in the lumen (Supplementary Fig. 2c, Supplementary Tables 2a and 3a). Ribonuclease P (RNase P), the ribozyme that cleaves the precursor RNA on tRNAs to form mature tRNAs, was also up-regulated in both the mucus and tissue (Supplementary Fig. 2d). A number of ribosome-related genes were up-regulated in the tissue, including 30S and 50S subunits (Supplementary Fig. 2b, Supplementary Table 3a). Collectively, this indicates that *B. fragilis* expands its capacity for protein synthesis in the mucus and tissue.

Rather than observing global metabolic shifts between microenvironments, we observed a small number of individual metabolic genes with spatially differentiated expression patterns. The two most up-regulated genes in host-associated samples (mucus and tissue) were BF3134, a glycanase (glycosyl hydrolase family 13), and BF3086, one of 17 sulfatases in the genome (Fig. 2b, Supplementary Fig. 2e). The induction of these enzymes is likely for the metabolism of mucosal glycans, which have unique structures and are often sulfated[31]. We would not have been able to confidently detect up-regulation of these two genes in the tissue samples without hybrid selection, as the average number of reads per sample without hybrid selection was only 14.

We examined the conservation of these two genes across a set of 92 *Bacteroides* and *Parabacteroides* genomes (see Materials and Methods). The average pairwise nucleotide identity within the 13 *B. fragilis* BF3134 and BF3086 genes was 99.7% and 99.5%, respectively, while the average pairwise identity between the non-*B. fragilis* genomes in our dataset was 66% and 69%, respectively, indicating significant conservation within *B. fragilis*, but significant diversity outside of *B. fragilis* (Supplementary Fig. 3a). Interestingly, we found a matching 36 bp gapped motif (Supplementary Fig. 3b) upstream of both BF3086 and BF3134 which was conserved in 13 closely-related *B. fragilis* genomes upstream of orthologous genes (Supplementary Fig. 3a). Three "*B. fragilis*" genomes that did not have the motif upstream of orthologues for these two genes appeared to actually be different species (see "Motif scanning" methods). The conservation of these host-association genes across strains of *B. fragilis* motivated us to investigate their importance by making in-frame deletions of both genes.

BF3086 and BF3134 are mucosal colonization factors

To assess the function of these enzymes, we compared growth of the BF3086 and BF3134 mutants with wild-type *B. fragilis* in minimal media with defined carbon sources. All three strains showed similar growth profiles in several dietary polysaccharides (Supplementary Fig. 4a-c). Their growth was also similar in purified pig mucin (Supplementary Fig. 4d). We sought to more accurately model mucosal growth *in vitro* using bulk mucus from germ-free mice (see Methods). *B. fragilis* grew rapidly, but the BF3134 mutant saturated at a lower CFU / ml (Fig. 3a). Furthermore, while wild-type bacteria had a stable stationary phase, the BF3134 mutant entered a rapid death phase (Fig. 3a). The BF3086 mutant also exhibited a significantly faster death phase (Fig. 3a). This lack of persistence may reflect the mutants' inability to use less accessible or lower abundance mucosal glycans that remain after log-phase growth. Because growth in mucus is involved in association with the epithelial surface, we tested the ability of these mutants to adhere to mucus-producing tissue-cultured epithelial cells. While BF3086 exhibited a mild defect in adherence, BF3134 showed a dramatic defect (Fig. 3b).

To test how these mucosal growth, persistence, and adherence phenotypes affect colonization in mice, we first mono-colonized mice with wild-type or mutant *B. fragilis* strains. All strains colonized mice stably and reached the same CFU level in feces (Fig. 3c), so we sacrificed the animals to assess spatial colonization differences. Although there were also equal amounts of bacteria in the colon lumen (Fig. 3d), both BF3086 and BF3134 mutants were defective in colonizing the colon mucus (Fig. 3e). We tested whether this difference in

mucosal colonization provides advantages to *B. fragilis* using a competitive colonization assay against wild-type bacteria. The BF3134 mutant had a dramatic defect, being eventually cleared from the animal (Fig. 4a). Consistent with the subtle *in vitro* phenotypes, the BF3086 mutant did not show a competitive disadvantage (Fig. 4b).

We have previously shown that *B. fragilis* mutants exhibiting defects in mucosal colonization are also unable to exclude competitors of the same species [23] (Chapter 2 of this thesis). In horizontal transmission assays, while mice colonized with wild-type *B. fragilis* (WT initial) remained essentially mono-colonized, mice initially colonized with either the BF3086 or BF3134 mutant were substantially colonized by the wild-type bacteria (Fig. 4c and 4d). This is consistent with a model whereby saturation of the mucosal niche prevents invasion by a foreign strain. The implication that animals would maintain long-term colonization by only a single strain of *B. fragilis* has been observed in longitudinal metagenomics studies in humans [32,33].

BF3134 is required for *B. fragilis* protection from colitis

While mucosal colonization has clear benefits for the bacteria, how this might affect the host is less obvious. *B. fragilis* protects mice from pathology in multiple models of colitis through the induction of anti-inflammatory IL-10 production in regulatory T cells (Tregs)[34,35]. Because this requires delivery of *B. fragilis* polysaccharide A to dendritic cells[35], we hypothesized that mucosal colonization may be critical for its immunomodulatory effects. We tested this in mice mono-colonized with *B. fragilis* strains subjected to the DNBS model of experimentally-induced colitis.

Control mice were treated with ethanol (the carrier for DNBS) to assess any baseline differences between groups. The weight-loss associated with this control treatment was consistent whether mice were colonized with wild-type, BF3086 mutant, or BF3134 mutant strains (Supplementary Fig. 5a). The length of the colon was also consistent across these control groups (Supplementary Fig. 5b). When treated with DNBS, mice mono-colonized with wild-type *B. fragilis* begin to recover by day 3 (Fig. 5a). However, mice mono-colonized with the BF3134 mutant exhibited continued decline, having lost significantly more weight by day 3 (Fig. 5a). Accordingly, this group of mice had significantly shorter colon lengths (Fig. 5b), a hallmark of worsened disease. The BF3086 mutant, which had less dramatic phenotypes *in vitro* and *in vivo*, was more similar to wild-type in having a protective effect in the mice.

We hypothesized that defective mucosal colonization led to a loss of *B. fragilis* anti-inflammatory effects on Tregs[34-36]. To test this, we isolated T cells from mesenteric lymph nodes 3 days after inducing colitis. Mice mono-colonized with the BF3134 mutant had similar amounts of pro-inflammatory IL-17 producing Tregs (Fig. 5c) but significantly less anti-inflammatory IL-10 producing Tregs (Fig. 5d). This indicates *B. fragilis* glycanase BF3134-dependent mucosal colonization is required for the protective anti-inflammatory effect of the symbiont.

DISCUSSION

Comparisons of bacterial behavior in micro-environments allows for a more sophisticated understanding of host-microbe interactions. Hybrid selection greatly improved our ability to assess transcriptomics of host-associated bacteria. Unlike *Bacteroides thetaiotaomicron* [6], *B. fragilis* did not broadly alter metabolic pathways along the radial axis of the gut. We found instead that *B. fragilis* deployed a specific set of mucosal colonization factors during host association, including a sulfatase (BF3086) and a critical mucosal glycanase (BF3134). Sulfatases were previously shown to be important for mucosal glycan foraging by *B. thetaiotaomicron*[37]. Though association with the intestinal surface may be treacherous for many bacteria, *B. fragilis* can tolerate oxygen[29], is resistant to many antimicrobial peptides[38], and benefits from immunoglobulin A binding (Chapter 2 of this thesis). During colitis, bacterial penetration of mucus is generally associated with pathology[39-41], but here we found that mucosal colonization by *B. fragilis* was protective. We conclude that the localization of *B. fragilis* in the mucus layer is fundamental to its ecology as well as its immunomodulatory effects.

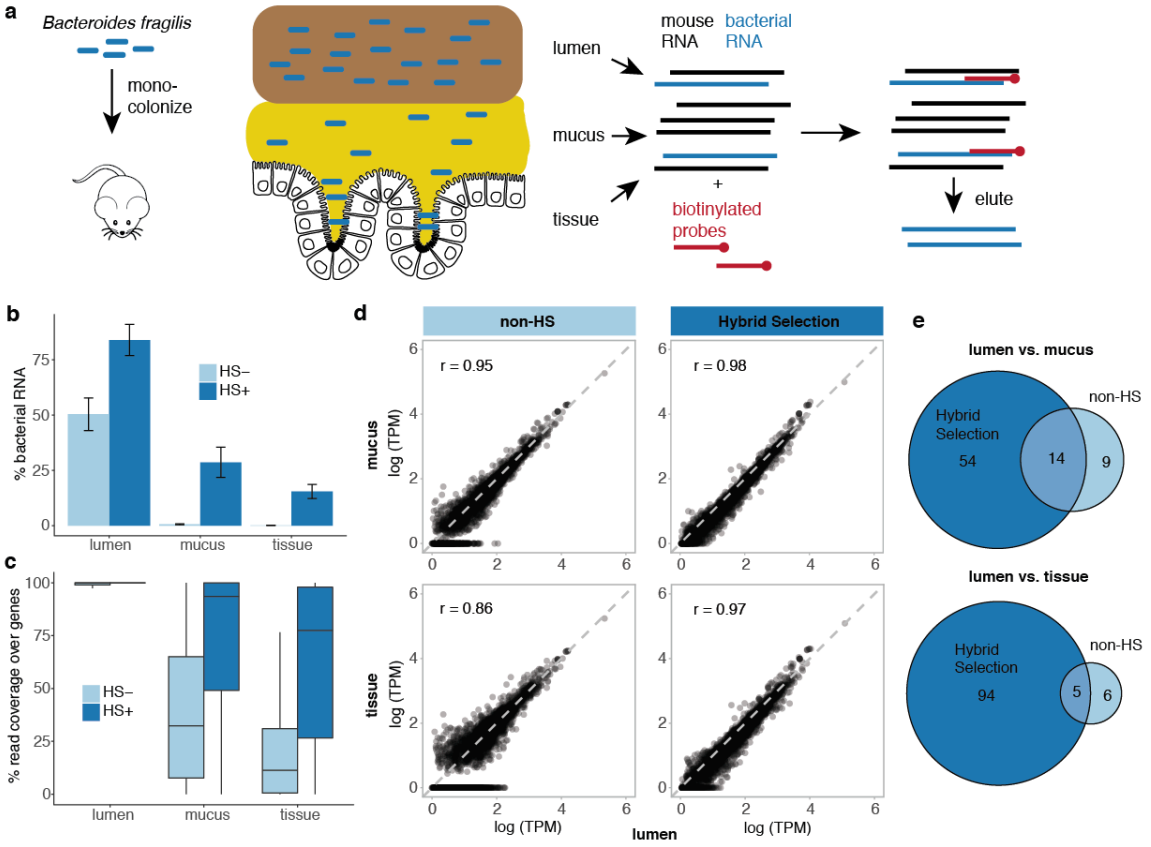


Fig. 1 | Hybrid selection enriches for bacterial RNA in host-associated gut samples. a, Colon lumen, mucus, and tissue samples were collected from mice mono-colonized with *B. fragilis*. Whole genome baits were used to select *B. fragilis* RNA from total RNA preparations. **b**, Average RNA-seq reads mapping to the *B. fragilis* genome with and without hybrid selection. **c**, Average read coverage of *B. fragilis* genes. **d**, Correlation in gene expression between different sample sites was improved with hybrid selection. All genes are plotted. **e**, Hybrid selection increases the number of genes identified in differential expression analyses between sample sites.

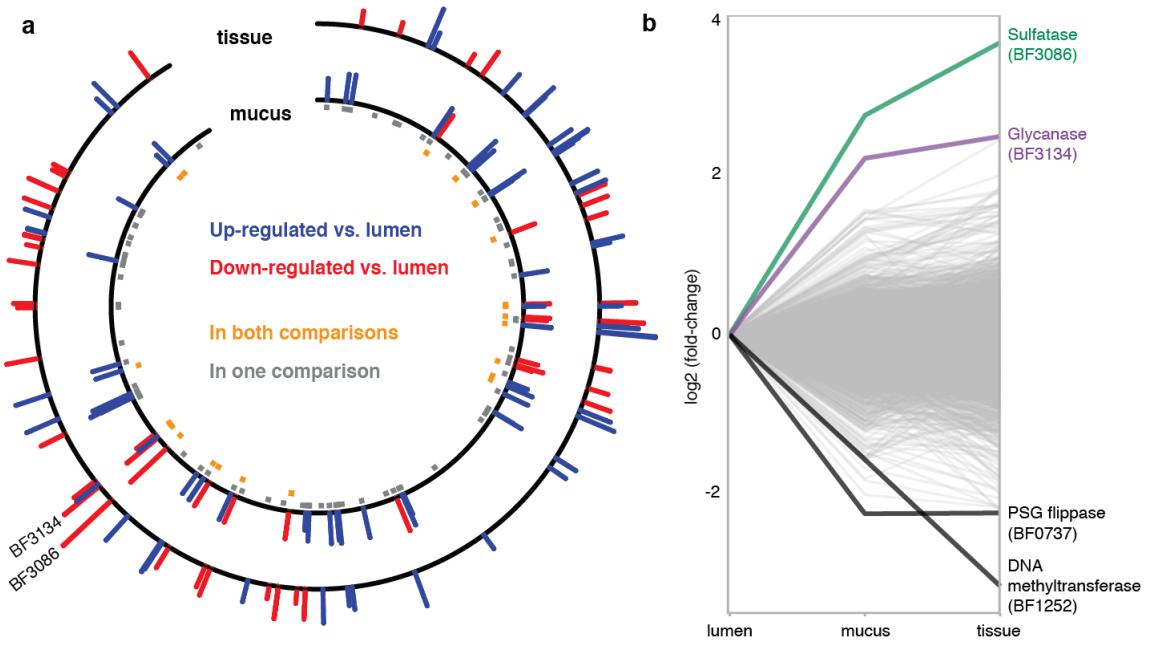


Fig. 2 | *B. fragilis* gene expression across microenvironments in the colon a, Genes differentially expressed between the lumen and mucus (inner circle) and lumen and tissue (outer circle). Squares in the innermost ring indicate genes differentially expressed in both (orange) or only one of the two comparisons (grey). **b**, Fold change in expression for individual genes in mucus and tissue, with respect to the lumen.

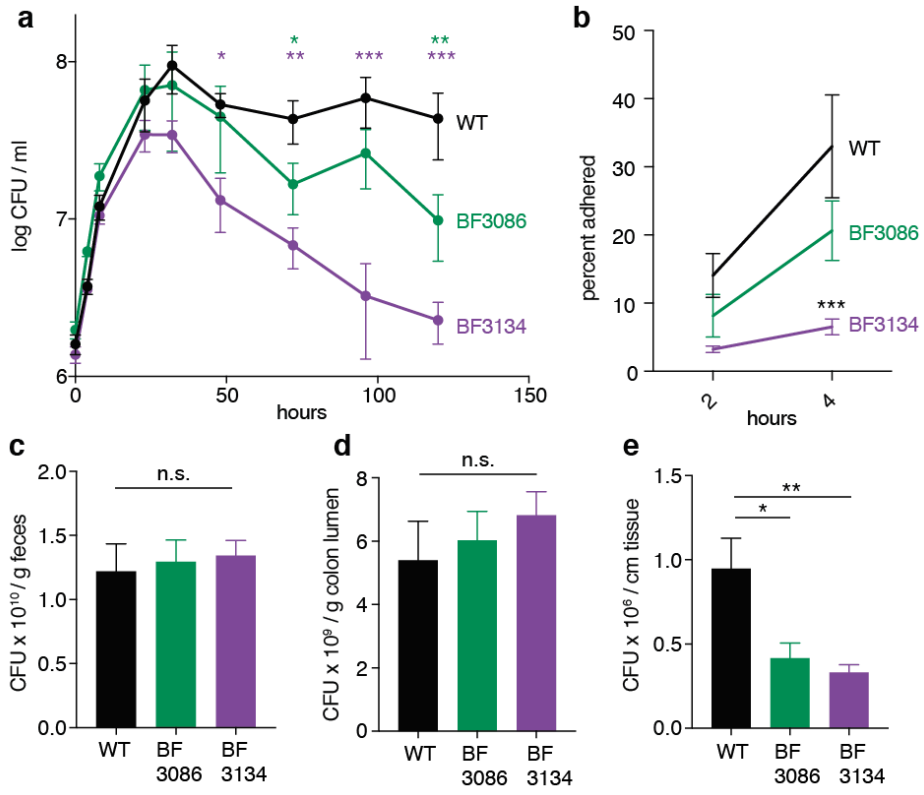


Fig. 3 | BF3086 and BF3134 are mucosal colonization factors. a, Growth of *B. fragilis* and mutants in BF3086 and BF3134 in defined minimal media with mouse mucus from germ-free mice as the sole carbon source (Tukey 2-way ANOVA on log-transformed data, n = 4 where each n is a different preparation of mucus). **b**, Adherence of *B. fragilis* strains to mucus-producing, tissue-cultured Caco-2 epithelial cells (Tukey 2-way ANOVA, n = 4). **c**, Quantification of bacteria in feces one month after mono-colonization with indicated strains of *B. fragilis* (Tukey ANOVA, n = 6). These mice were sacrificed to quantify **d**, colon lumen and **e**, colon mucus colonization levels (Tukey ANOVA, n = 6) (all panels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

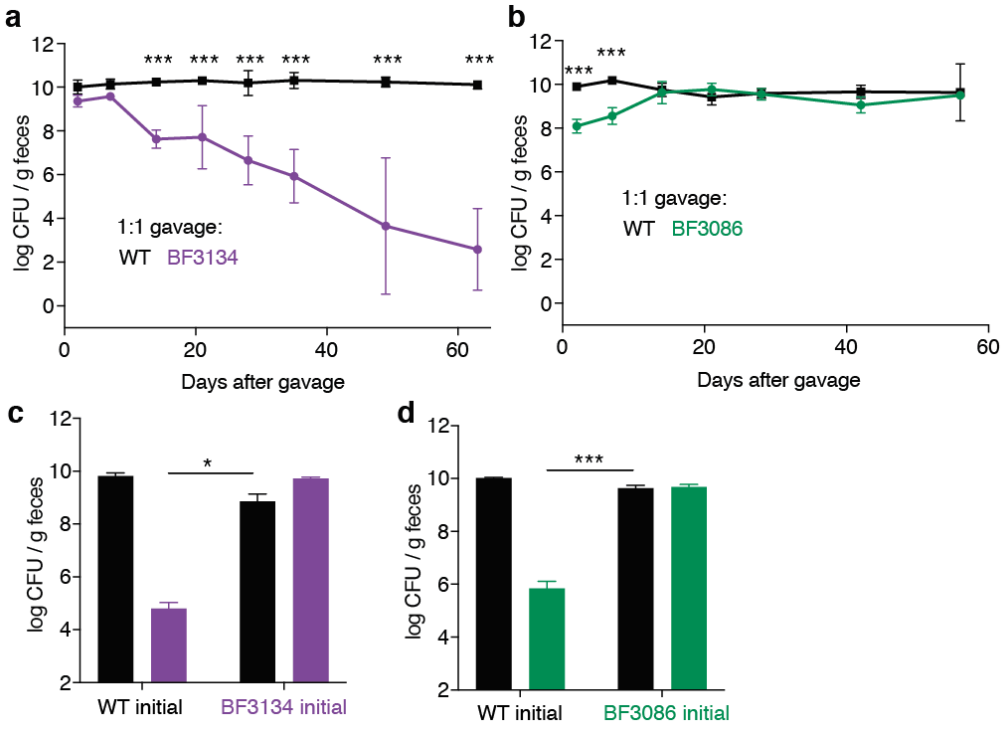


Fig. 4 | Competitive colonization phenotypes of BF3086 and BF3134. a, Germ-free mice were gavaged with a 1:1 mixture of 10^8 CFU each of wildtype and ΔBF3134 *B. fragilis*. Colonization of each strain was tracked using antibiotic resistance markers (Sidak 2-way ANOVA of log-transformed data, n = 4) **b**, The same 1:1 gavage experiment with wildtype and ΔBF3086 *B. fragilis* (Sidak 2-way ANOVA of log-transformed data, n = 4). **c**, Horizontal transmission between pairs of mice mono-colonized with either wild-type *B. fragilis* (WT initial) or ΔBF3134 (BF3134 initial). Two weeks after separating the mice, levels of initial and foreign strains are graphed (Tukey 2-way ANOVA of log-transformed data, n = 6). **d**, Horizontal transmission with wild-type and ΔBF3086-colonized mice (Tukey 2-way ANOVA of log-transformed data, n = 6) (all panels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

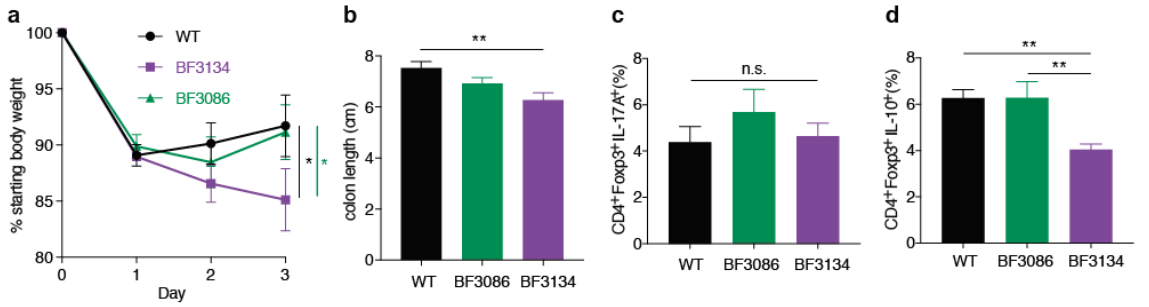
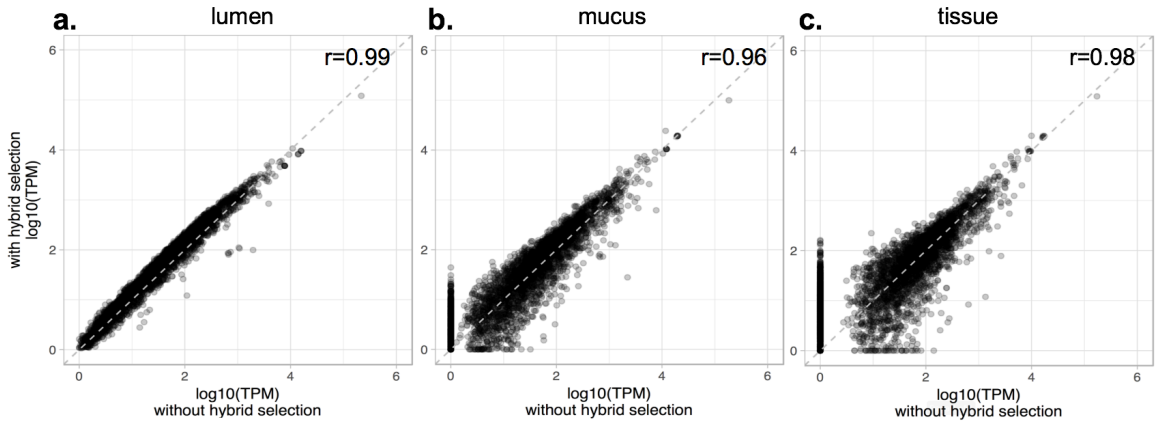
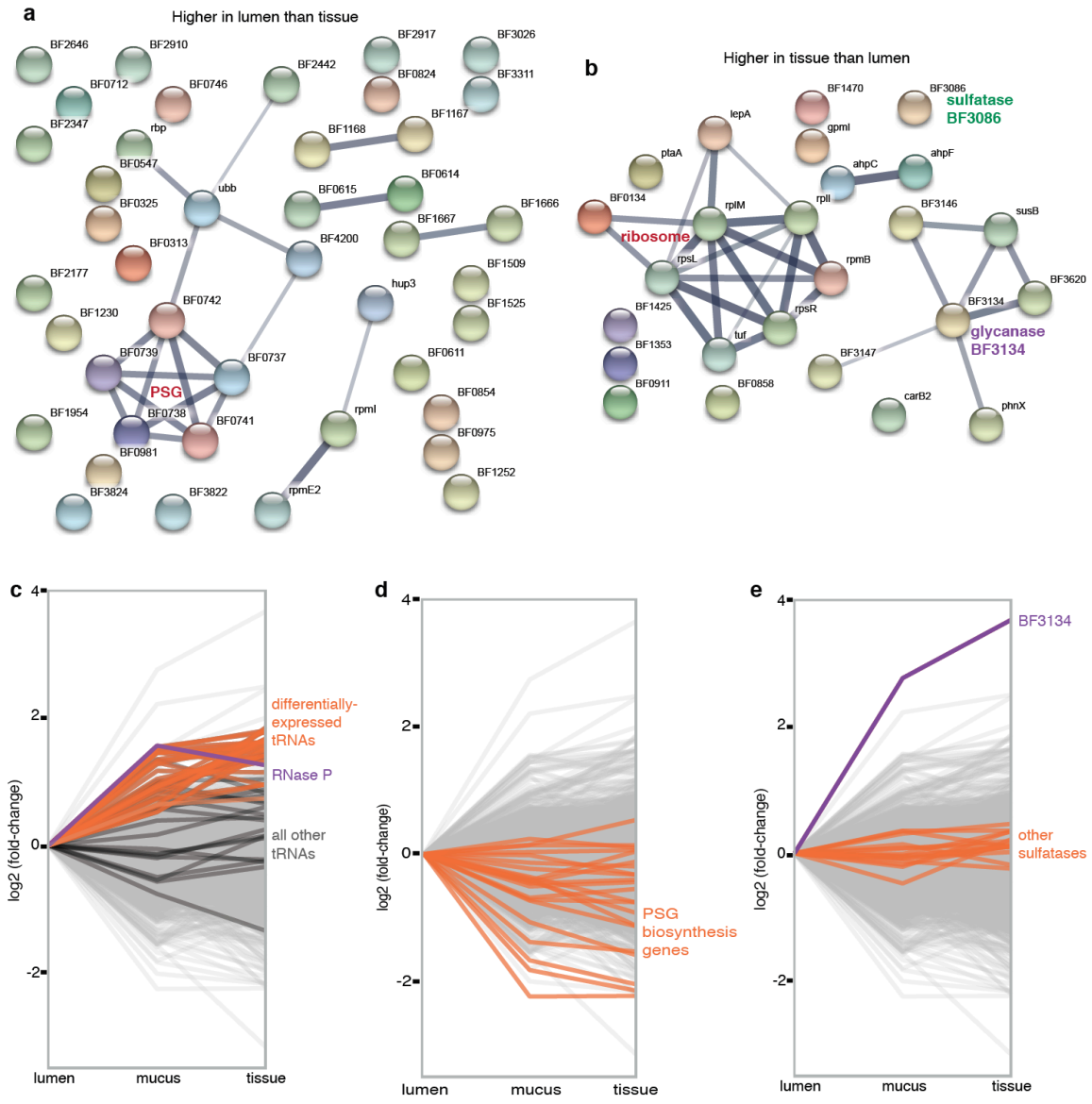


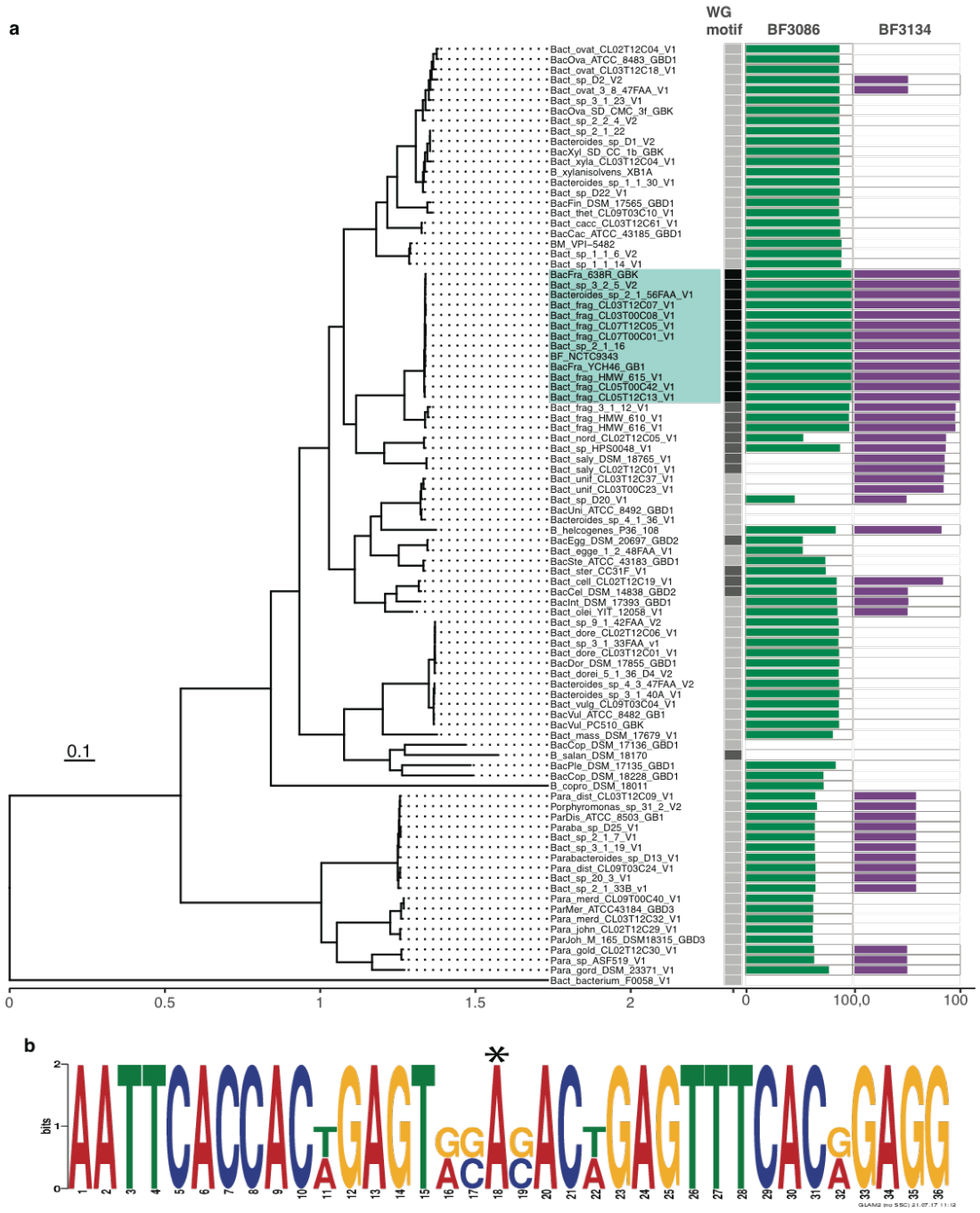
Fig. 5 | Effect of mucosal colonization on colitis. a, Mice were mono-colonized with *B. fragilis* strains for one month before DNBS colitis induction. Body weight of the mice was measured every 24 hours and graphed as a percentage of their starting weight on day 0 (Tukey ANOVA). **b**, 72 hours after induction mice were sacrificed and the length of the colon from rectum to the cecal junction was measured (Tukey 2-way ANOVA). **c**, Images. **d**, Scores. **e**, Lymphocytes isolated from mesenteric lymph nodes of mono-colonized, DNBS-induced mice were analyzed using flow cytometry. IL-17A-producing T cells quantified as a percent of total Foxp3+ regulatory T cells (Tukey ANOVA). **f**, IL-10-producint T cells quantified as a percent of total Foxp3+ regulatory T cells (Tukey ANOVA) (all panels: n = 10, 9, 9, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).



Supplementary Figure 1 | Comparison of gene expression with and without hybrid selection. Each gene is represented by a single dot. **a**, The correlation coefficient between lumen samples is 0.99. **b**, The correlation coefficient between mucus samples is 0.96. **c**, The correlation coefficient between tissue samples is 0.98.

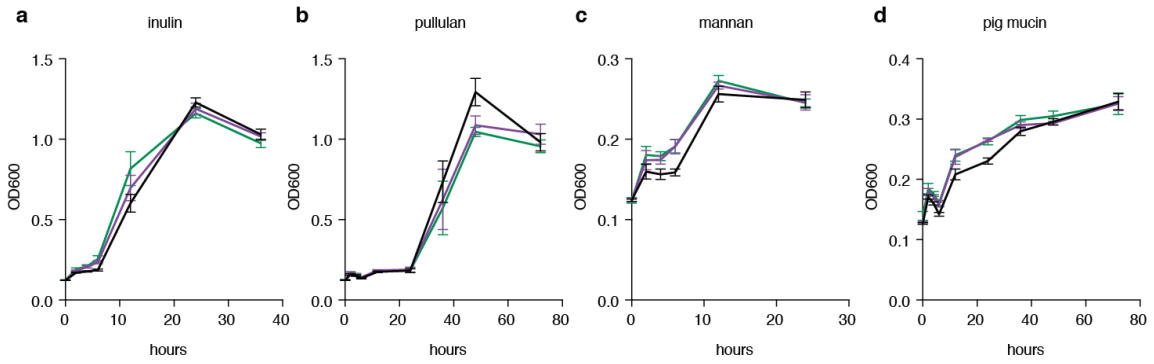


Supplementary Figure 2 | Differentially expressed genes in micro-environments. a, STRING[54] analysis of genes more highly expressed in the lumen than tissue. The thickness of connecting lines indicates confidence in relationships between genes. **b**, Same analysis of genes more highly expressed in the tissue than lumen (tRNAs are not included). **c-e**, Fold change in expression for individual genes in mucus and tissue, with respect to the lumen with indicated genes highlighted.

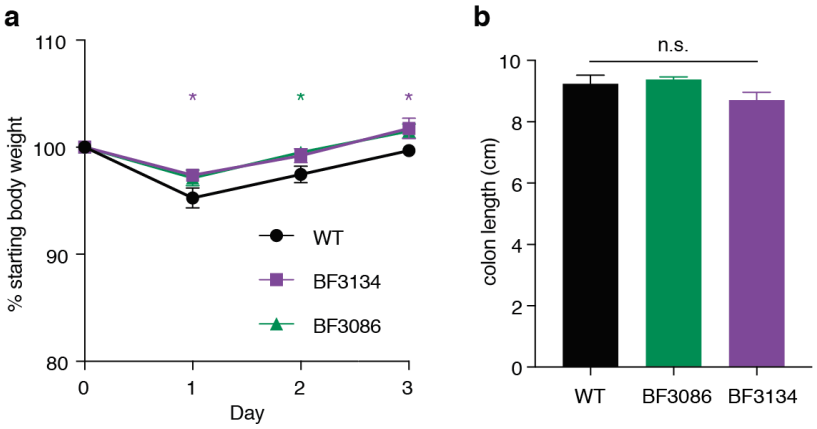


Supplementary Figure 3 | Conservation of BF3086, BF3134, and a potential regulatory motif. a, Phylogeny of 92 *Bacteroides* and *Parabacteroides* strains[50] showing conservation of BF3086, BF3134, and the upstream motif in *B. fragilis* strains (teal box). The black

squares indicate the presence of the conserved upstream motif, using the GALM2Scan algorithm[51]. Black corresponds to a motif "match" (0-2 mismatches). Green and purple bar graphs indicate the percent protein sequence identity of orthologues to BF3086 and BF3134. **b**, Sequence of the conserved motif upstream of both genes.



Supplementary Figure 4 | Growth of *B. fragilis* in minimal defined media with various carbon sources. a, inulin. **b**, pullulan. **c**, mannan. **d**, pig mucin (n = 8 for all).



Supplementary Figure 5 | Vehicle control (ethanol)-treated mice for DNBS colitis experiment. a, Mice mono-colonized with indicated strains of *B. fragilis* for one month were treated with 50% ethanol, the vehicle control for DNBS colitis induction. Mice were weighed every 24 hours, graphed as a percentage of their weight at day 0 (Tukey 2-way ANOVA, n = 5, 4, 4). **b**, 72 hours after treatment the mice were sacrificed and the length of the colon was measured from rectum to the cecal junction (Tukey 2-way ANOVA, n = 5, 4, 4).

Supplementary Table 1 | RNA-seq read counts and mapping statistics. Average values of three replicates and standard deviations are indicated.

	Sample	Total RNA-seq reads	% Reads mapped to *B. fragilis* genome	% Reads mapped to mouse genome	% Reads unaligned	Median of % read coverage over *B. fragilis* genes
Without hybrid selection	**Lumen**	30,037,711 ± 4,838,289	50.4 ± 7.3	19.5 ± 8.2	30 ± 2.42	100 ± 0.0
	Mucus	30,851,659 ± 380,774	0.6 ± 0.2	66.9 ± 2.1	32.5 ± 2.1	31.9 ± 8.5
	Tissue	30,284,943 ± 3,142,595	0.1 ± 0.0	67.1 ± 0.6	32.8 ± 0.6	8.3 ± 4.0
With hybrid selection	**Lumen**	33,719,927 ± 5,268,780	84 ± 7.0	11.9 ± 6.3	4.2 ± 0.7	100 ± 0.0
	Mucus	30,509,921 ± 2,638,387	28.6 ± 6.8	60.9 ± 6.0	10.5 ± 0.9	95 ± 4.3
	Tissue	31,786,088 ± 5,102,116	15.4 ± 3.2	71.2 ± 2.5	13.3 ± 0.7	79.1 ± 4.3

Supplementary Table 2a | Fold change and adjusted p-value (FDR) of genes most up-regulated in mucus compared to lumen.

old locus tag	new locus tag	log2(fold change)	FDR	Product Name
BF3086	**BF9343_RS14795**	**2.75**	**9.2E-17**	**sulfatase**
BF3134	**BF9343_RS15035**	**2.22**	**1.8E-13**	**glycosyl hydrolase family 13**
BF2146	BF9343_RS10075	1.99	1.1E-02	5-amino-6-(5-phosphoribosylamino)uracil reductase
BF1418	BF9343_RS06550	1.59	1.8E-02	α-galactosidase
-	BF9343_RS21640	1.56	3.9E-06	RNase P RNA component class A
BF1209	BF9343_RS05605	1.46	1.3E-02	NADH dehydrogenase
BF1158	BF9343_RS05365	1.44	1.3E-04	α-glucosidase
BF1210	BF9343_RS05610	1.31	1.2E-03	alkyl hydroperoxide reductase
BF3147	BF9343_RS15100	1.29	3.0E-04	LacI family transcriptional regulator
BF1394	BF9343_RS06445	1.20	1.5E-04	RNA polymerase sigma factor
BF1425	BF9343_RS06580	0.97	8.0E-03	membrane protein
BF3146	BF9343_RS15095	0.85	2.7E-02	SusC/RagA family TonB-linked outer membrane protein
tRNAs				
BFt38	BF9343_RS12180	1.55	8.6E-08	tRNA-Gly
BFt41	BF9343_RS13845	1.53	1.6E-07	tRNA-Asp
BFt07	BF9343_RS02255	1.53	6.9E-06	tRNA-Phe
BFt37	BF9343_RS12175	1.52	1.4E-07	tRNA-Leu
BFt42	BF9343_RS13850	1.46	3.3E-06	tRNA-Asp
BFt13	BF9343_RS04120	1.45	5.1E-07	tRNA-Tyr
BFt06	BF9343_RS02250	1.45	1.8E-05	tRNA-Pro
BFt33	BF9343_RS12155	1.35	6.2E-06	tRNA-Gly
BFt12	BF9343_RS04115	1.32	6.2E-06	tRNA-Gly
BFt15	BF9343_RS05655	1.30	1.8E-02	tRNA-Ser
BFt36	BF9343_RS12170	1.27	4.0E-05	tRNA-Leu

BFt23	BF9343_RS09900	1.17	1.1E-02	tRNA-Asp
BFt34	BF9343_RS12160	1.04	7.0E-04	tRNA-Leu
BFt35	BF9343_RS12165	0.99	3.3E-03	tRNA-Gly

Supplementary Table 2b | Fold change and adjusted p-value (FDR) of genes most down-regulated in mucus compared to lumen.

old locus tag	new locus tag	log2(fold change)	FDR	Product Name
BF3278	BF9343_RS15800	-2.54	4.0E-06	hypothetical protein
BF3296	BF9343_RS15895	-2.29	6.3E-05	hypothetical protein
BF0737	BF9343_RS03450	-2.25	5.9E-10	Polysaccharide G (PSG) flippase
BF0611	BF9343_RS02855	-2.01	1.1E-07	amino-terminal protease
BF3276	BF9343_RS15795	-1.90	1.7E-03	recombinase
BF3294	BF9343_RS15885	-1.88	6.5E-03	hypothetical protein
BF0739	BF9343_RS03460	-1.83	9.3E-05	hypothetical protein
BF2112	BF9343_RS09910	-1.82	3.2E-03	RNA-directed DNA polymerase
BF0466	BF9343_RS21570	-1.81	4.9E-03	hypothetical protein
BF2365	BF9343_RS11175	-1.74	4.0E-03	hypothetical protein
BF1641	BF9343_RS07560	-1.73	2.8E-02	hypothetical protein
BF2397	BF9343_RS11345	-1.71	1.3E-02	hypothetical protein
BF0738	BF9343_RS03455	-1.68	1.8E-04	glycosyltransferase
BF2266	BF9343_RS10670	-1.68	6.8E-03	hypothetical protein
BF0632	BF9343_RS02940	-1.60	1.3E-02	biotin-[acetyl-CoA-carboxylase] synthetase
BF1252	BF9343_RS05815	-1.57	1.5E-03	DNA methyltransferase
BF2487	BF9343_RS11795	-1.56	4.7E-02	chloramphenicol acetyltransferase
BF2371	BF9343_RS11205	-1.55	1.6E-02	hypothetical protein
BF2785	BF9343_RS13330	-1.55	4.5E-02	N-acetylmuramoyl-L-alanine amidase
BF3797	BF9343_RS18370	-1.55	1.9E-02	hypothetical protein
BF0144	BF9343_RS00645	-1.54	1.3E-02	hypothetical protein
BF0117	BF9343_RS00530	-1.53	8.0E-03	hypothetical protein
BF3410	BF9343_RS16455	-1.53	1.3E-02	DNA-directed RNA polymerase sigma-70 factor
BF3379	BF9343_RS16290	-1.49	6.5E-07	histone H1

BF1525	BF9343_RS07025	-1.48	4.0E-03	hypothetical protein
BF1037	BF9343_RS04885	-1.46	2.7E-02	hypothetical protein
BF1568	BF9343_RS07220	-1.44	4.6E-04	DNA-binding protein
BF2917	BF9343_RS13960	-1.41	7.4E-03	hypothetical protein
BF0742	BF9343_RS03475	-1.39	2.0E-03	capsular polysaccharide-like protein
BF2943	BF9343_RS14090	-1.33	1.3E-02	anti-sigma factor
BF4220	BF9343_RS20550	-1.33	1.2E-03	DNA-binding protein HU-beta
BF3140	BF9343_RS15065	-1.29	3.6E-05	50S ribosomal protein L31 type B
BF2104	BF9343_RS09865	-1.20	1.3E-02	membrane protein
BF0033	BF9343_RS00165	-1.20	2.9E-02	hypothetical protein
BF1640	BF9343_RS07555	-1.18	1.3E-02	hypothetical protein
BF0615	BF9343_RS02870	-1.15	1.0E-03	hypothetical protein
BF1168	BF9343_RS05415	-1.13	3.0E-04	hypothetical protein
BF4025	BF9343_RS19545	-1.10	1.4E-02	30S ribosomal protein S21
BF1509	BF9343_RS06955	-1.08	4.5E-02	hypothetical protein
BF2495	BF9343_RS11850	-0.96	4.9E-03	universal stress protein
BF1167	BF9343_RS05410	-0.88	8.0E-03	preprotein translocase subunit YajC
BF4200	BF9343_RS20455	-0.79	2.8E-02	glycosyl transferase family 1

Supplementary Table 3a | Fold change and adjusted p-value (FDR) of the most up-regulated genes in tissue compared to lumen.

old locus tag	new locus tag	log2(fold change)	FDR	Product Name
BF3086	**BF9343_RS14795**	**3.67**	**1.9E-38**	**sulfatase**
BF3134	**BF9343_RS15035**	**2.49**	**8.2E-11**	**glycosyl hydrolase family 13**
BF1209	BF9343_RS05605	2.45	5.2E-14	NADH dehydrogenase
BF1158	BF9343_RS05365	2.00	2.2E-04	α-glucosidase
BF3488	BF9343_RS16825	1.87	8.2E-03	phosphonoacetaldehyde hydrolase
BF0858	BF9343_RS04030	1.68	4.2E-02	MFS transporter
BF1210	BF9343_RS05610	1.61	2.0E-06	alkyl hydroperoxide reductase
BF1470	BF9343_RS06775	1.58	4.0E-02	DNA primase
BF3147	BF9343_RS15100	1.49	4.8E-04	LacI family transcriptional regulator
-	BF9343_RS21640	1.26	8.5E-05	hypothetical protein
BF3631	BF9343_RS17525	1.19	1.4E-04	30S ribosomal protein S18
BF0911	BF9343_RS04290	1.18	2.3E-02	cystathionine beta-lyase
BF1425	BF9343_RS06580	1.17	1.9E-03	membrane protein
BF3146	BF9343_RS15095	1.16	6.3E-04	SusC/RagA family TonB-linked outer membrane protein
BF3808	BF9343_RS18420	1.07	2.8E-03	carbamoyl phosphate synthase large subunit
BF3620	BF9343_RS17465	0.97	1.0E-02	4-α-glucanotransferase
BF1353	BF9343_RS06265	0.96	1.2E-02	iron transporter FeoB
BF3630	BF9343_RS17520	0.91	9.8E-03	50S ribosomal protein L9
BF0134	BF9343_RS00595	0.90	2.0E-02	DEAD/DEAH box helicase
BF4007	BF9343_RS19435	0.87	1.8E-02	30S ribosomal protein S12
BF3787	BF9343_RS18310	0.83	2.3E-02	50S ribosomal protein L13
BF2510	BF9343_RS11925	0.83	1.6E-02	50S ribosomal protein L28
BF0423	BF9343_RS01970	0.79	3.7E-02	phosphate acetyltransferase

BF4022	BF9343_RS19510	0.77	3.8E-02	hypothetical protein
BF0241	BF9343_RS01105	0.76	4.0E-02	phosphoglycerate mutase
BF2581	BF9343_RS12325	0.76	3.0E-02	elongation factor 4
tRNA				
BFt59	BF9343_RS19130	1.85	4.8E-04	tRNA-Ala
BFt60	BF9343_RS19135	1.83	8.6E-04	tRNA-Ile
BFt37	BF9343_RS12175	1.80	2.1E-09	tRNA-Leu
BFt72	BF9343_RS21010	1.79	1.8E-04	tRNA-Ala
BFt38	BF9343_RS12180	1.77	9.7E-09	tRNA-Gly
BFt30	BF9343_RS11815	1.67	9.1E-03	tRNA-Arg
BFt56	BF9343_RS18890	1.67	7.2E-04	tRNA-Ala
BFt33	BF9343_RS12155	1.63	2.8E-07	tRNA-Gly
BFt36	BF9343_RS12170	1.63	2.5E-08	tRNA-Leu
BFt06	BF9343_RS02250	1.63	2.0E-06	tRNA-Pro
BFt07	BF9343_RS02255	1.62	1.2E-05	tRNA-Phe
BFt51	BF9343_RS18075	1.62	8.3E-04	tRNA-Ala
BFt39	BF9343_RS13155	1.62	1.4E-03	tRNA-Ala
BFt73	BF9343_RS21015	1.54	2.3E-02	tRNA-Ile
BFt13	BF9343_RS04120	1.54	1.6E-06	tRNA-Tyr
BFt29	BF9343_RS11810	1.52	2.4E-02	tRNA-Arg
BFt45	BF9343_RS15740	1.50	3.2E-03	tRNA-Ala
BFt12	BF9343_RS04115	1.50	3.2E-06	tRNA-Gly
BFt31	BF9343_RS11820	1.45	3.4E-02	tRNA-Arg
BFt34	BF9343_RS12160	1.41	6.1E-07	tRNA-Leu
BFt46	BF9343_RS15745	1.41	4.8E-02	tRNA-Ile
BFt41	BF9343_RS13845	1.35	2.3E-04	tRNA-Asp
BFt35	BF9343_RS12165	1.32	1.5E-05	tRNA-Gly
BFt42	BF9343_RS13850	1.28	1.0E-03	tRNA-Asp

BFt63	BF9343_RS19520	0.96	1.7E-02	tRNA-Gly
BFt64	BF9343_RS19525	0.95	1.2E-02	tRNA-Tyr

Supplementary Figure 3b | Fold change and adjusted p-value (FDR) of genes most down-regulated in tissue compared to lumen.

old locus tag	new locus tag	log2 (fold change)	FDR	Product Name
BF1252	BF9343_RS05815	-3.16	8.2E-11	DNA methyltransferase
BF0737	BF9343_RS03450	-2.24	3.1E-04	Polysaccharide G (PSG) flippase
BF0313	BF9343_RS01440	-2.22	2.5E-04	hypothetical protein
BF1230	BF9343_RS05705	-2.19	1.2E-04	hypothetical protein
BF1525	BF9343_RS07025	-2.18	2.4E-07	hypothetical protein
BF0611	BF9343_RS02855	-2.17	6.2E-06	amino-terminal protease
BF0739	BF9343_RS03460	-2.16	4.8E-04	hypothetical protein
BF0738	BF9343_RS03455	-2.06	2.3E-04	glycosyltransferase
BF2177	BF9343_RS10220	-2.00	1.5E-03	hypothetical protein
BF2917	BF9343_RS13960	-1.96	1.7E-04	hypothetical protein
BF3379	BF9343_RS16290	-1.93	9.7E-09	histone H1
BF3311	BF9343_RS15970	-1.91	2.3E-02	cupin
BF2442	BF9343_RS11570	-1.88	6.7E-05	hypothetical protein
BF4220	BF9343_RS20550	-1.83	2.4E-06	DNA-binding protein HU-beta
BF1509	BF9343_RS06955	-1.81	6.3E-04	hypothetical protein
BF3026	BF9343_RS14490	-1.80	1.4E-02	phosphohydrolase
pBF 9343.09	BF9343_RS21360	-1.79	9.1E-03	hypothetical protein
BF0854	BF9343_RS04010	-1.78	2.0E-03	hypothetical protein
BF2910	BF9343_RS13935	-1.73	1.4E-02	DNA-directed RNA polymerase sigma-70 factor
BF0975	BF9343_RS04600	-1.71	1.6E-02	DNA-directed RNA polymerase sigma-70 factor
-	BF9343_RS05700	-1.65	2.0E-02	conjugal transfer protein
BF0741	BF9343_RS03470	-1.59	1.6E-02	hypothetical protein
BF0824	BF9343_RS03860	-1.59	2.3E-02	hypothetical protein

pBF 9343.20c	BF9343_RS21415	-1.55	8.2E-03	hypothetical protein
BF0742	BF9343_RS03475	-1.54	4.4E-03	capsular polysaccharide-like protein
BF3140	BF9343_RS15065	-1.51	6.2E-06	50S ribosomal protein L31 type B
BF0615	BF9343_RS02870	-1.48	4.8E-04	hypothetical protein
BF0712	BF9343_RS03330	-1.48	2.4E-02	hypothetical protein
BF3883	BF9343_RS18770	-1.41	4.3E-02	ubiquitin
BF0547	BF9343_RS02550	-1.39	1.5E-02	hypothetical protein
BF1666	BF9343_RS07680	-1.31	1.8E-04	membrane protein
BF2347	BF9343_RS11095	-1.25	2.3E-02	hypothetical protein
BF2646	BF9343_RS12645	-1.21	2.3E-02	heat-shock protein
pBF 9343.16	BF9343_RS21395	-1.21	2.3E-02	hypothetical protein
BF1167	BF9343_RS05410	-1.20	6.5E-04	preprotein translocase subunit YajC
BF0746	BF9343_RS03495	-1.13	5.8E-03	transposase
BF4200	BF9343_RS20455	-1.12	9.0E-03	glycosyl transferase family 1
BF1696	BF9343_RS07830	-1.09	2.3E-02	50S ribosomal protein L35
BF1168	BF9343_RS05415	-1.02	5.1E-03	hypothetical protein
BF2361	BF9343_RS11155	-1.00	1.8E-02	RNA-binding protein
BF3822	BF9343_RS18490	-0.97	2.3E-02	tight junction protein ZO-3
BF3824	BF9343_RS18500	-0.97	1.2E-02	hypothetical protein
BF0981	BF9343_RS04630	-0.95	2.0E-02	hypothetical protein
BF1954	BF9343_RS09125	-0.95	4.0E-03	RNA polymerase sigma factor
BF0325	BF9343_RS01505	-0.94	4.5E-02	hypothetical protein
BF1667	BF9343_RS07685	-0.88	3.3E-02	hypothetical protein
BF0614	BF9343_RS02865	-0.84	3.8E-02	glycosyl transferase family A

Supplementary Table 4a: Significantly enriched functional domains in the 68 genes differentially expressed in mucus compared to lumen.

Pfam functional domain	adjusted p-value (FDR)	# differentially expressed genes containing the functional domain	# total genes in the genome containing the functional domain
Glycosyl transferases group 1	0.004	3	28
Bacterial DNA-binding protein	0.0059	2	16
GTB	0.0242	3	51

Supplementary Table 4b: Significantly enriched functional domains in the 99 genes differentially expressed in tissue compared to lumen.

Pfam functional domain	adjusted p-value (FDR)	# differentially expressed genes with the functional domain	# total genes in the genome with the functional domain
Elongation factor Tu domain 2	0.0028	2	7
50S ribosome-binding GTPase	0.0036	3	16
Elongation factor Tu GTP binding domain	0.0069	2	10
Glycosyl transferases group 1	0.0185	3	28
Type I phosphodiesterase / nucleotide pyrophosphatase	0.0324	2	19

Supplementary Table 4c: Significantly enriched functional domains in the 130 genes differentially expressed in both mucus and tissue, as compared to lumen (FDR < 0.05).

Pfam functional domain	adjusted p-value (FDR)	# differentially expressed genes with the functional domain	# total genes in the genome with the functional domain
Elongation factor Tu domain 2	0.0045	2	7
50S ribosome-binding GTPase	0.0066	3	16
Sigma-70 region 2	0.0106	5	43
Elongation factor Tu GTP binding domain	0.0106	2	10
Sigma-70, region 4	0.0115	5	44

ECF sigma factor	0.0128	2	11
Glycosyl transferases group 1	0.0287	3	28
Bacterial DNA-binding protein	0.0294	2	16
Bacterial regulatory proteins, luxR family	0.0396	2	18
Type I phosphodiesterase / nucleotide pyrophosphatase	0.0453	2	19

METHODS

Bacteria strains and media

B. fragilis NCTC9343 was cultured in Brain Heart Infusion (BD) with 5 μg/ml hemin (Frontier Scientific) and 5 μg/ml vitamin K1 (Sigma) or a defined minimal media[42,43] in an 80% nitrogen, 10% carbon dioxide, and 10% hydrogen atmosphere. For growth in mouse mucus, crude mucus was isolated as described below (Separation of colon lumen, mucus, and tissue) from the entire colon of germ-free mice into defined minimal media (mucus from one whole colon homogenized in 5 ml of final media). Where appropriate, 200 μg/ml gentamicin, 10 μg/ml erythromycin, 2 μg/ml tetracycline, or 10 μg/ml chloramphenicol were used in selective media. For competitive colonization and horizontal transmission, marker plasmids pFD340-Chlor (providing resistance to erythromycin and chloramphenicol) or pFD340-Tet (providing resistance to erythromycin and tetracycline) were used to distinguish two strains, as before[23]. In-frame deletions in BF3086 (1305 bp deleted) and BF3134 (1686 bp deleted) were made using allelic exchange with the pNJR6 suicide vector as previously described[23].

Mice and colonization experiments

All animal experiments were performed in accordance with guidelines in the NIH Guide for the Care and Use of Animals and protocols approved by the Caltech Institutional Animal Care and Use Committee. Swiss Webster mice from Taconic Farms were c-section re-derived germ-free and bred in flexible film isolators. For most experiments, germ-free mice were transferred at 6-8 weeks of age to sterile micro-isolator cages with autoclaved food (LabDiet 5010) and water. Mice were mono-colonized by a single gavage with 10^8 CFU of

B. fragilis in 100 µl of HBSS with 1.5% sodium bicarbonate (or a 1:1 mix of 10^8 CFU each of two strains, for competitive colonization). Mice were maintained mono-colonized for 4 weeks prior to subsequent experimentation. Colonization was monitored in fresh fecal samples which were weighed, mashed, and vortexed in 1 ml BHI and diluted for plating CFU. For mucosal CFU plating, mucus was isolated as described below (Separation of colon lumen, mucus, and tissue). For competitive colonization and horizontal transmission assays, water was supplemented with 100 µg/ml gentamicin (*Bacteroides* are naturally resistant) and 10 µg/ml erythromycin (to select for marker plasmids). Horizontal transmission was assayed as previously[23], by co-housing mice for 4 hours in a fresh sterile cage and then separating into single-housing in fresh sterile cages.

Separation of colon lumen, mucus, and tissue

Animals were sacrificed one at a time and samples were quickly processed through the bead-beating lysis step (< 10 minutes from sacrifice to lysis). Lysed samples were kept at 4 C until all samples were collected. First, the proximal colon was dissected. A 1 cm segment of the proximal (ascending) colon, starting at the cecal junction, was taken for sampling. The segment was opened longitudinally and ~100 mg of lumen content was collected for the "lumen" sample. The rest of the lumen content was removed with forceps. The tissue was washed by vigorously shaking with forceps in a dish of HBSS for 20 seconds (changing forceps grip 3 times during the process). The washed tissue was carefully observed to ensure no lumen content remained. Tissue was dabbed in a dry sterile petri dish to remove excess buffer. Mucus was scraped from the surface of the tissue using a sterile plastic 1.8 cm cell

scraper (BD Falcon), collected as the “mucus” sample. The remaining tissue was collected as the “tissue” sample.

RNA purification

All samples were subjected to the same RNA preparation protocol, which was the only method we found to reproducibly provide high-quality and high-yield RNA from lumen, mucus, and tissue. Samples were immediately lysed by bead-beating for 1 minute in 2 ml lysing matrix B tubes (MP Biomedicals) with 500 µl buffer (0.2 M NaCl and 20 mM EDTA), 210 µl 20% SDS (Ambion AM9820), and 500 µl phenol, chloroform, isoamyl alcohol mixture (Ambion AM9720). After centrifugation at 13,000 rpm for 3 minutes at 4 C, the aqueous phase was added to a new microcentrifuge tube with 500 µl of the phenol, chloroform, isoamyl alcohol mixture, and mixed by inversion. The centrifugation and isolation of aqueous phase was repeated, yielding about 300 µl. Next, 30 µl of 3 M sodium acetate (Ambion AM9740) and 300 µl of -20 C 100% ethanol was mixed in by inversion. Samples were left on ice for 20 minutes, then centrifuged at 13,000 rpm for 20 minutes at 4 C. The supernatant was decanted and 500 µl of -20 C 70% ethanol was added and vortexed before centrifuging at 13,000 rpm for 5 minutes at 4 C. The supernatant was decanted and the tubes were inverted and air-dried for 5 minutes after wiping the lips of tubes dry on Kimwipes. 100 µl of water was added to the dried pellets, which were then frozen at -20 C. The next day, samples were thawed and 350 µl RLT buffer (Qiagen) with 1% 2-mercaptoethanol was added. Tubes were vortexed for 20 minutes, at which point the pellets were completely dissolved. The samples were then loaded on Qiagen RNeasy mini columns and purified according to the manufacturer’s instructions. Nucleic acids were eluted in 50 µl

of water and quantified using a NanoDrop. Up to 10 µg of this was taken into a 60 µl DNase reaction with 4 µl Turbo DNase (Ambion AM2238) at 37 C for an hour. 40 µl of water and 350 µl buffer RLT were added and the samples were loaded onto a second Qiagen RNeasy mini column. The second column purification was performed according to the manufacturer's instructions including the Qiagen on-column DNase digest. The final total RNA was eluted in 50 µl water.

Preparation of whole-transcriptome fragment libraries (pond) for hybrid selection

The isolated RNA was first quantified and qualified by Qubit and Agilent Bioanalyzer. The RNA was then fragmented by FastAP (Thermo Scientific) and linked to barcoded adapters. The fragmented and barcoded RNA was pooled to perform ribosomal RNA depletion using Ribo-Zero Magnetic Gold Epidemiology Kit (Epicentre/Illumina). The cDNA was generated from the RNA through template-switching RT-PCR. After exonuclease I treatment and PCR enrichment, the RNA was used for hybrid selection.

Hybrid selection probes (bait) construction

Whole genome bait (WGB) was generated at the Broad Institute. For input, 3 µg of *B. fragilis* NCTC 9343 DNA was sheared for 4 minutes on a Covaris E210 instrument set to duty cycle 5, intensity 5 and 200 cycles per burst. The mode of the resulting fragment size distribution was 250 bp. End repair, addition of a 3'-A, adaptor ligation and reaction clean-up followed the Illumina's genomic DNA sample preparation kit protocol, except that the adapter consisted of oligonucleotides 5'-TGTAACATCACAGCATCACCGCCATCAGTCxT-3' ('x' refers to an exonuclease I-resistant phosphorothioate linkage) and 5'-

[PHOS]GACTGATGGCGCACTACGACACTACAATGT-3'. The ligation products were cleaned up (Qiagen), amplified by 8 to 12 cycles of PCR on an ABI GeneAmp 9700 thermocycler in Phusion High-Fidelity PCR master mix with HF buffer (NEB Ipswich, Massachusetts, United States) using PCR forward primer 5'-CGCTCAGCGGCCGCAGCATCACCGCCATCAGT-3' and reverse primer 5'-CGCTCAGCGGCCGCGTCGTAGTGCGCCATCAGT-3' (ABI Carlsbad, California, United States). Initial denaturation was 30 s at 98°C. Each cycle was 10 s at 98°C, 30 s at 50°C and 30 s at 68°C. PCR products were size-selected on a 4% NuSieve 3:1 agarose gel followed by QIAquick gel extraction. To add a T7 promoter, size-selected PCR products were re-amplified as above using the forward primer 5'-GGATTCTAATACGACTCACTATACGCTCAGCGGCCGCAGCATCACCGCCATCA GT-3'. Qiagen-purified PCR product was used as template for whole genome biotinylated RNA bait preparation with the MEGAshortscript T7 kit (Ambion)[19,20].

Hybrid selection

Using the designed baits, hybridization was conducted at 65°C for 66 h with 2 μg of 'pond' libraries carrying standard or indexed Illumina paired-end adapter sequences and 500 ng of bait in a volume of 30 μl. After hybridization, captured DNA was pulled down using streptavidin Dynabeads (Invitrogen Carlsbad, California, United States). Beads were washed once at room temperature for 15 minutes with 0.5 ml 1 × SSC/0.1% SDS, followed by three 10-minute washes at 65°C with 0.5 ml pre-warmed 0.1 × SSC/0.1% SDS, re-suspending the beads once at each washing step. Hybrid-selected DNA was eluted with 50 μl 0.1 M NaOH. After 10 minutes at room temperature, the beads were pulled down, the supernatant

transferred to a tube containing 70 µl of 1 M Tris-HCl, pH 7.5, and the neutralized DNA desalted and concentrated on a QIAquick MinElute column and eluted in 20 µl.

Sequencing

Each sample was sequenced on Illumina HiSeq2500 at the Broad Institute to produce 101-bp paired-end reads. Sequence data have been deposited in the NCBI Short Read Archive under a project accession number, PRJNA438372.

RNA-seq read processing and mapping

Identifiers for all RNA-seq experiments are listed in Supplementary Table 1. The RNA-seq reads were trimmed with a Phred quality score cut-off of 20 by the program fastq_quality_trimmer from the FASTX toolkit, version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). Reads shorter than 20 bp after adaptor- and poly(A)-trimming were discarded before mapping. Trimmed RNA-seq reads were aligned to the *B. fragilis* NCTC 9343 genome (NC_003228.3) and the mouse genome (genome build GRCm38.p4) in parallel. RNA-seq read mapping to the bacterial and mouse genome were performed by Bowtie2[44] and STAR[45], respectively. The mapping results were used to calculate read counts over each gene of the bacterial and mouse genomes by bedtools[46] and HTseq[47], respectively. We also calculated read coverage and transcripts per million (TPM) for each bacterial gene for examining the distribution of bacterial gene expression.

Differential gene expression analysis

Read counts for each bacterial gene were used to analyze differential gene expression using the edgeR package[48]. Bacterial genes with more than 10 uniquely mapped reads in each of three replicates were considered to be detected and were retained for the differential gene expression analysis. Genes with an adjusted p-value < 0.05 in the edgeR analysis were considered differentially expressed.

Functional enrichment and other statistical analysis

We annotated the bacterial genes with Pfam[24] using the Broad Institute's prokaryotic annotation pipeline[49]. To assess functional enrichment of differentially expressed genes, we calculated statistical significance using the hypergeometric function with adjustment for multiple hypothesis testing. Adjusted p-values < 0.05 were considered enriched. We calculated Pearson correlation coefficients to determine if the expression of genes (in TPM) were comparable between two samples. We used the Wilcoxon signed-rank test to compare the expression of a gene family in two different samples. p-values < 0.05 were considered significant.

Comparative genomics and motif analysis of 92 *Bacterioides* and *Parabacteroides* genomes

Using a comparative analysis of 92 diverse genome sequences related to *B. fragilis* [50], ,which included 23 Bacteroides and 5 Parabacteroides species, we identified 43 BF3134 orthologs in 43 strains and 117 BF3086 orthologs in 83 strains. We constructed multiple alignments of the nucleotide sequences of these groups of orthologs for BF3086 and BF3134, which we used to calculate pairwise sequence identities to measure conservation levels. We searched

the upstream regions of BF3086 and BF3134 in *B. fragilis* NCTC 9343 for conserved motifs, or potential binding sites for transcription factors, using GLAM2[51] from the MEME suite[52]. We further examined the presence and conservation of this potential regulatory motif using GLAM2Scan[51] on our set of 92 diverse Bacteroides and Parabacteroides genome sequences (Supplementary Figure S3)[50]. Presence of the motif was defined by having 2 or less mismatches. Of the 16 *B. fragilis* genomes, we used multiple sequence alignments to confirm that the 3 divergent *B. fragilis* are missing the predicted motif (Supplementary Figure S4 and S5). Excluding three same-patient samples, the *B. fragilis* genomes containing the motif had a pairwise average nucleotide identity (ANI) value of 98%, indicating that these strains are members of the same species, but not clonally related. In contrast, pairwise ANI values between *B. fragilis* strains with and without the motif averaged 86%, below the threshold commonly used to describe species[49,53].

Epithelial cell adherence assay

The stable intestinal epithelial cell line Caco-2 was maintained in high glucose DMEM with 4 mM L-glutamine, 4.5 g / L glucose (HyClone), supplemented with 10% fetal bovine serum (FBS) (Gibco), and penicillin-streptomycin solution (Corning). 25,000 cells were seeded in flat 96-well plates and grown for ten days past confluence (~13 days after seeding) to allow production of mucus. One day prior to the adherence assay, the media was changed to high-glucose DMEM without antibiotics or FBS. Bacteria from fresh cultures in BHI were pelleted and re-suspended in high-glucose DMEM without antibiotics or FBS. Epithelial cells were washed once with warm HBSS and then incubated with 10^6 CFU of bacteria for 2 or 4 hours at 37 C in an anaerobic atmosphere of 80% nitrogen, 10% carbon dioxide, and

10% hydrogen. Wells were washed twice with 200 µl warm HBSS to remove unbound bacteria, and then trypsinized with 50 µl of 0.05% trypsin in HBSS (Corning) for 20 minutes at 37 C to disassociate epithelial cells. Cells were vigorously resuspended with 50 µl BHI added to the trypsin before dilution plating for CFU. Fraction bound was calculated as the output CFU / input CFU.

DNBS colitis

Mice were mono-colonized at 3 weeks of age and colitis was induced 4 weeks later. Mice were anesthetized using isofluorane and 5% DNBS in 50% ethanol (or 50% ethanol only) was administered rectally through a 3.5F catheter (Instech Solomon) inserted 4 cm into the colon. Mice were subsequently kept upside-down for 1 minute to prevent leakage. Mice were weighed every 24 hours and sacrificed at 72 hours post-induction. The whole colon was dissected, fixed in 10% buffered formalin, paraffin embedded, sectioned, and then stained with hematoxylin and eosin (HE).

Isolation of mesenteric lymph node lymphocytes and flow cytometry

Mesenteric lymph nodes (MLN) were isolated and processed by dissociating tissues through a 70 µm cell strainer (BD Falcon) to generate single cell suspensions. Cells were washed in ice cold PBS. For flow cytometry analysis, cells were labelled with the LIVE/DEAD fixable violet dead cell stain kit (Life Technologies), with empirically titrated concentration of PE-Cy7-conjugated anti-mouse CD4 (RM4-5, eBioscience). For intracellular staining, cells were fixed and permeabilized with the Foxp3/Transcription factor buffer kit (eBioscience), followed by staining with the following antibodies: FITC-conjugated anti-mouse IFNγ

(XMG1.2, eBioscience), PE-conjugated anti-mouse IL-10 (JES5-16E3, eBioscience), PerCP-Cy5.5 anti-mouse IL-17A (eBio17B7, eBioscience), and APC-conjugated anti-mouse Foxp3 (FJK16s, eBioscience). Cell acquisition was performed on a Miltenyi MACSQuant (Miltenyi), and data was analyzed using FlowJo software suite (TreeStar).

REFERENCES

1. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14,** 20–32 (2015).
2. Johansson, M. E. V. & Hansson, G. C. Immunological aspects of intestinal mucus and mucins. *Nat. Rev. Immunol.* (2016). doi:10.1038/nri.2016.88
3. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307,** 1955–1959 (2005).
4. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4,** 447–457 (2008).
5. Berry, D. *et al.* Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. *Proc Natl Acad Sci USA* (2013). doi:10.1073/pnas.1219247110
6. Li, H. *et al.* The outer mucus layer hosts a distinct intestinal microbial niche. *Nat Commun* **6,** 8292 (2015).
7. Wang, Y. *et al.* Regional mucosa-associated microbiota determine physiological expression of TLR2 and TLR4 in murine colon. *PLoS ONE* **5,** e13607 (2010).

8. Yasuda, K. *et al.* Biogeography of the intestinal mucosal and lumenal microbiome in the rhesus macaque. *Cell Host Microbe* **17,** 385–391 (2015).

9. Albenberg, L. *et al.* Correlation Between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota in Humans and Mice. *Gastroenterology* (2014). doi:10.1053/j.gastro.2014.07.020

10. Davis, C. P. & Savage, D. C. Habitat, succession, attachment, and morphology of segmented, filamentous microbes indigenous to the murine gastrointestinal tract. *Infect. Immun.* **10,** 948–956 (1974).

11. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139,** 485–498 (2009).

12. Vélez, M. P., De Keersmaecker, S. C. J. & Vanderleyden, J. Adherence factors of Lactobacillus in the human gastrointestinal tract. *FEMS Microbiol. Lett.* **276,** 140–148 (2007).

13. Savage, D. C. Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* **31,** 107–133 (1977).

14. Rowan, F. *et al.* Bacterial colonization of colonic crypt mucous gel and disease activity in ulcerative colitis. *Ann. Surg.* **252,** 869–875 (2010).

15. Pédron, T. *et al.* A crypt-specific core microbiota resides in the mouse colon. *MBio* **3,** (2012).

16. Earle, K. A. *et al.* Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell Host Microbe* **18,** 478–488 (2015).

17. Mark Welch, J. L., Hasegawa, Y., McNulty, N. P., Gordon, J. I. & Borisy, G. G. Spatial organization of a model 15-member human gut microbiota

established in gnotobiotic mice. *Proc Natl Acad Sci USA* **114,** E9105–E9114 (2017).

18. Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S. & Cohen, S. N. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* **99,** 9697–9702 (2002).

19. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27,** 182–189 (2009).

20. Melnikov, A. *et al.* Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biol.* **12,** R73 (2011).

21. Matranga, C. B. *et al.* Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biol.* **15,** 519 (2014).

22. Huang, J. Y., Lee, S. M. & Mazmanian, S. K. The human commensal Bacteroides fragilis binds intestinal mucin. *Anaerobe* **17,** 137–141 (2011).

23. Lee, S. M. *et al.* Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501,** 426–429 (2013).

24. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44,** D279–85 (2016).

25. Blow, M. J. *et al.* The Epigenomic Landscape of Prokaryotes. *PLoS Genet.* **12,** e1005854 (2016).

26. Burroughs, A. M., Kaur, G., Zhang, D. & Aravind, L. Novel clades of the HU/IHF superfamily point to unexpected roles in the eukaryotic centrosome,

chromosome partitioning, and biologic conflicts. *Cell Cycle* **16,** 1093–1103 (2017).

27. Liu, C. H., Lee, S. M., Vanlare, J. M., Kasper, D. L. & Mazmanian, S. K. Regulation of surface architecture by symbiotic bacteria mediates host colonization. *Proc Natl Acad Sci USA* **105,** 3951–3956 (2008).

28. Coyne, M. J., Chatzidaki-Livanis, M., Paoletti, L. C. & Comstock, L. E. Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont Bacteroides fragilis. *Proc Natl Acad Sci USA* **105,** 13099–13104 (2008).

29. Baughn, A. D. & Malamy, M. H. The strict anaerobe Bacteroides fragilis grows in and benefits from nanomolar concentrations of oxygen. *Nature* **427,** 441–444 (2004).

30. Rocha, E. R. & Smith, C. J. Role of the alkyl hydroperoxide reductase (ahpCF) gene in oxidative stress defense of the obligate Anaerobe bacteroides fragilis. *Journal of Bacteriology* **181,** 5701–5710 (1999).

31. Thomsson, K. A. *et al.* Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2. *Glycobiology* **22,** 1128–1139 (2012).

32. Scholz, M. *et al.* Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat. Methods* **13,** 435–438 (2016).

33. Yassour, M. *et al.* Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Science Translational Medicine* **8,** 343ra81–343ra81 (2016).

34. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–625 (2008).

35. Chu, H. *et al.* Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science* **352,** 1116–1120 (2016).

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

37. Benjdia, A., Martens, E. C., Gordon, J. I. & Berteau, O. Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont, Bacteroides thetaiotaomicron. *J Biol Chem* **286,** 25973–25982 (2011).

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

39. Schultsz, C., Van Den Berg, F. M., Kate, Ten, F. W., Tytgat, G. N. & Dankert, J. The intestinal mucus layer from patients with inflammatory bowel disease harbors high numbers of bacteria compared with controls. *Gastroenterology* **117,** 1089–1097 (1999).

40. Swidsinski, A. *et al.* Comparative study of the intestinal mucus barrier in normal and inflamed colon. *Gut* **56,** 343–350 (2007).

41. Johansson, M. E. V. *et al.* Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model. *PLoS ONE* **5,** e12238 (2010).

42. Varel, V. H. & Bryant, M. P. Nutritional features of Bacteroides fragilis subsp. fragilis. *Appl Microbiol* **28,** 251–257 (1974).

43. Kotarski, S. F. & Salyers, A. A. Isolation and characterization of outer membranes of Bacteroides thetaiotaomicron grown on different carbohydrates. *Journal of Bacteriology* **158,** 102–109 (1984).

44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

45. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

46. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

47. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).

48. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

49. Lebreton, F. *et al.* Emergence of epidemic multidrug-resistant Enterococcus faecium from animal and commensal strains. *MBio* **4,** (2013).

50. Sefik, E. *et al.* Individual intestinal symbionts induce a distinct population of RORγ$^+$ regulatory T cells. *Science* **349,** 993–997 (2015).

51. Frith, M. C., Saunders, N. F. W., Kobe, B. & Bailey, T. L. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput. Biol.* **4,** e1000071 (2008).

52. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37,** W202–8 (2009).

53. Lebreton, F. *et al.* Tracing the Enterococci from Paleozoic Origins to the Hospital. *Cell* **169,** 849–861.e13 (2017).

54. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* **45,** D362–D368 (2017).

ACKNOWLEDGMENTS

We would like to thank Eric Martens, Dirk Gevers, Chris Desjardins, Brian Haas, and Jonathan Livny for helpful discussions. G.P.D. was supported by an NIH training grant (5T32 GM07616) and NSF Graduate Research Fellowship (DGE-1144469). This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Grant Number U19AI110818 to the Broad Institute.

Chapter 4

THESIS CONCLUSION

"The plagues fade, but microbes remain; and vestiges of the memory of the formidable ones cling to those that live innocently, even beneficently, on man."

-Theodor Rosebury, 1969, *Life on Man*

Almost a century before Darwin's "Origin of Species," Antonie van Leeuwenhoek turned his microscope toward his own body and discovered that we do not live alone. His letters describe a diverse oral and gut microbiome during health, which appeared to change during disease. Despite these early indicators of its importance, the indigenous microbiome of animals was largely an afterthought through the 20th century. In the context of association with animals, bacteria were studied as pathogens.

The development of anaerobic culture methods and gnotobiotics facilitated a revolution in microbiome research in the 1960's. René Dubos was by then an eminent bacteriologist at the Rockefeller University, known for being the first to *deliberately* discover an antibiotic. Inspired by his observations about the importance of nutrition in immunity, he turned his laboratory to a reductionist study of the mouse intestinal microbiome. In a remarkable series of papers, Russell Schaedler and Dubos described the sequential development of the gut microbial community in pups[1], its dependence on diet[2], spatial differences along the length of the gut[2], and a specific mucosal community composition[2]. Furthermore, they established the experimental paradigm of colonizing germ-free mice with individual species or defined communities[3].

Dubos and colleagues understood that studying feces was a poor approximation of what was happening inside the gut. Dwayne Savage, another member of Dubos's group, applied imaging techniques to the study of the mucosal microbiome[4], later discovering segmented filamentous bacteria[5] and crypt-associated bacteria[6]. Savage championed the idea that the intestinal surface was heavily colonized during health, providing protection from pathogen colonization[7]. Decades later, segmented filamentous bacteria would become an important model for interactions with immunoglobulin A (IgA)[8] and effects of the microbiome on T cell development[9]. Despite their age and limited impact at the time, the studies from Dubos, Schaedler, and Savage provide a still-relevant foundation for molecular symbiosis research.

Armed with genetics and molecular biology in addition to traditional techniques, this thesis explores the same question that inspired Dubos: how do bacteria colonize the gut during health? The chapters contain several examples of mechanisms of mucosal colonization and the importance of gut biogeography to the form and function of the microbiome. They provide support for the hypothesis that the mucus serves as a privileged niche and reservoir for bacteria. This microbiome reservoir hypothesis was proposed as a function of the human appendix in 2007 by Randall Bollinger and William Parker at Duke University[10]. Four years earlier, the same group hypothesized that antibodies (IgA) could enhance mucosal colonization[11]. The reductionist approach described in this thesis, using colonization experiments in gnotobiotic mice, provides mechanistic support for these high-level concepts.

The study of mucosal colonization by indigenous bacteria necessitates an immunological awareness. The science of immunology has been historically framed by pathogenesis, though

the immune system itself evolved in the context of symbiosis. Obligate microbial symbionts and the animal gut are bound together by co-evolution. The mutualistic relationship between *Bacteroides fragilis* and the mammalian immune system should therefore not come as a surprise. But the manner in which this symbiosis manifests is surprising in that there are many parallels to pathogenesis. Bacterial capsules have been well-described as virulence factors, allowing pathogens to cloak themselves to avoid immune attacks, such as antibody recognition. This paradigm is reversed for *B. fragilis*, which uses a capsule to attract antibody binding. Here, antibodies don't act as weapons, but as anchors for mucosal symbiosis. The insults that most bacteria would struggle to endure in the mucus do not appear to perturb *B. fragilis*, which lives and thrives on the intestinal surface. The effect of mucosal colonization on the host is also unintuitive from a pathogenesis-framed immunological perspective. Mucosal *B. fragilis* provides a protective anti-inflammatory signal.

To view the immune system in its entirety as a barrier preventing microbial colonization is to miss half of the story. The immune system must exclude threatening microbes, but it allows long-term colonization of the surface of tissues all over the body. It might be tempting to interpret the tolerance of the gut microbiome as the immune system recognizing these bacteria as "self." But the data here indicates otherwise: the immune system recognizes indigenous mucosal microbiota as allies.

REFERENCES

1. Schaedler, R. W., Dubos, R. & Costello, R. The development of the bacterial flora in the gastrointestinal tract of mice. *J. Exp. Med.* **122,** 59–66 (1965).
2. Dubos, R., Schaedler, R. W., Costello, R. & Hoet, P. Indigenous, normal, and autochthonous flora of the gastrointestinal tract. *J. Exp. Med.* **122,** 67–76 (1965).
3. Schaedler, R. W., Dubos, R. & Costello, R. Association of germfree mice with bacteria isolated from normal mice. *J. Exp. Med.* **122,** 77–82 (1965).
4. Savage, D. C., Dubos, R. & Schaedler, R. W. The gastrointestinal epithelium and its autochthonous bacterial flora. *J. Exp. Med.* **127,** 67–76 (1968).
5. Davis, C. P. & Savage, D. C. Habitat, succession, attachment, and morphology of segmented, filamentous microbes indigenous to the murine gastrointestinal tract. *Infect. Immun.* **10,** 948–956 (1974).
6. Savage, D. C. & Blumershine, R. V. Surface-surface associations in microbial communities populating epithelial habitats in the murine gastrointestinal ecosystem: scanning electron microscopy. *Infect. Immun.* **10,** 240–250 (1974).
7. Savage, D. C. Microbial interference between indigenous yeast and lactobacilli in the rodent stomach. *Journal of Bacteriology* **98,** 1278–1283 (1969).
8. Suzuki, K. *et al.* Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc Natl Acad Sci USA* **101,** 1981–1986 (2004).
9. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139,** 485–498 (2009).
10. Randal Bollinger, R., Barbas, A. S., Bush, E. L., Lin, S. S. & Parker, W. Biofilms in the large bowel suggest an apparent function of the human vermiform appendix. *J.*

Theor. Biol. **249,** 826–831 (2007).

11. Bollinger, R. R. *et al.* Human secretory immunoglobulin A may contribute to biofilm formation in the gut. *Immunology* **109,** 580–587 (2003).