Capsular Polysaccharides in *Bacteroides thetaiotaomicron* and Their Role in Mediating Interactions with Host Immunity and Bacteriophage

by

Nathan T. Porter

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in the University of Michigan
2017

Doctoral Committee:

Associate Professor Eric C. Martens, Chair
Professor Matthew R. Chapman
Professor Philip C. Hanna
Professor Patrick D. Schloss

Nathan T. Porter

porternt@umich.edu

ORCID iD: 0000-0001-8523-6909

Acknowledgements

I would like to thank my thesis advisor and mentor Dr. Eric Martens for his guidance over the course of my doctoral training. He has taught me to ask interesting questions and to seek out the best methods to answer those questions. Additionally, Eric has helped me to take on leadership of my own projects and not only improve my skills at the bench, but also develop my own projects—starting from a concept, progressing to design and execution of key experiments, and extending to analysis and publication of the findings. Eric, thank you for allowing me to grow and increase in confidence as a scientist and a leader.

I would also like to thank my committee members: Dr. Matthew Chapman, Dr. Philip Hanna, and Dr. Patrick Schloss, for their invaluable feedback and guidance. Specifically, thank you for encouraging me to focus on specific mechanisms while also generating large collections of strains and testing broad hypotheses.

I also appreciate the priceless assistance and guidance multiple department administrators have given me, as well the many other members of the department who have given me feedback or let me borrow equipment for long periods of time. Particularly, thank you to members of the laboratories of Dr. Harry Mobley, Dr. Michele Swanson, and Dr. Kathryn Eaton. I will eventually return your equipment, and will return it in great condition!

This work was made possible through generous funding from the University of Michigan Cellular Biotechnology Training Program, the Rackham Predoctoral Fellowship, and the National Institutes of Health.

My extended lab “family” also deserves special recognition, with members of the Martens and Koropatkin lab groups providing expert guidance, especially Nicholas Pudlo. Yao Xiao, Ursula Waack, and others also provided immeasurable support in numerous ways throughout graduate school which otherwise would not have been manageable.

Lastly, my family has constantly been supportive of me, through candidacy, late-night and long experiments, and countless hours writing and re-writing papers. Camille, though these are just a few of the many ways you have blessed my life, thank you for discussing and re-discussing scientific topics outside of your normal expertise, and for making graduate school fun and helping me to maintain vision of a greater purpose.

Table of Contents

Acknowledgements ii

List of Tables vii

List of Figures viii

Abstract x

CHAPTER 1 Introduction 1

Abstract 1
Introduction 1
Influence of dietary polysaccharides on the microbiota-host symbiosis 5
Host glycoconjugates shape and are shaped by the microbiota 7
Changes in host intestinal glycosylation in response to symbiont colonization 7
Host- and microbiota-catalyzed changes during disease 9
Biological effects of microbially produced glycans 13
An underexplored repository of polysaccharide diversity and function 13
Synthesis 14
Diversity 15
Contributions to the Diet-Microbiota-Host Axis During Intestinal Colonization 15
Links between diet and MPG expression 16
Interactions with the immune system 17
Exclusion of pathogens 20
Degradation by other microbes 21
Protection from bacteriophages 22
Conclusion and prospectus 24
Chapter outline 26

Notes 27
Literature cited 28

CHAPTER 2 A Subset of Polysaccharide Capsules in the Human Symbiont *Bacteroides thetaiotaomicron* Promote Increased Competitive Fitness in the Mouse Gut 38

Abstract 38
Introduction 38
Results 41
Long-term colonization of mice with B. theta reveals dynamic expression of CPS over time 41
Competition of single CPS-expressing strains reveals differential success in the mouse gut 45
Intestinal anti-CPS IgA levels correlate with increased competition among single CPS-expressing strains 55
Diet alterations do not have major impact on in vivo competition or on in vitro growth rate 59
The ability to express multiple CPS confers an additional advantage in gut competition 61
Extensive CPS biosynthetic locus diversity in sequenced B. theta isolates 66
CPS5 and CPS6 biosynthetic loci are prevalent in human gut metagenomic samples 69
Discussion 70
Methods 73
Strains and culture conditions. 73
Gnotobiotic mouse experiments 78
Resequencing of single CPS-expressing strains. 79
RNA isolation and quantitative PCR (qPCR) for relative cps expression 80
Quantitative PCR (qPCR) 81
Preparation of purified CPS and monosaccharide analysis 81
Single CPS-expressing strain characterization and validation 84
Relative cps expression from previous published datasets 86
Quantification of fecal IgA via ELISA 87
RNA-Seq analysis on bacteria monoassociated in germ-free mice 88
Bacterial growth rate analysis 89
Antibiotic and antimicrobial susceptibility assays 89
Laser capture microdissection and DNA extraction 90
Identification of cps biosynthetic loci in sequenced isolates and in metagenomes 91
Statistical analysis. 95

Data and software availability. 98
Notes 98
References 99

CHAPTER 3 Bacteriophage Target and Modulate Expression of Diverse Capsular Polysaccharides in the Gut Symbiont *Bacteroides thetaiotaomicron* 103

Abstract 103
Introduction 104
Results 107
Discussion 115
Methods 118
Strains and media conditions 118
Host range assays 120
Phage treatment to detect modulation of cps gene expression 121
RNA extraction and quantitative PCR (qPCR) for determination of relative cps gene expression ..122
Data representation and statistical analysis 122
References 124

CHAPTER 4 Discussion 127

Introduction 127
Chapter Summary 128
Specific roles for individual CPS 132
Advantages of encoding and switching between multiple CPS types 135
Future research 138
Determination of the CPS synthesis locus regulatory hierarchy 138
Defining the mechanism(s) of individual phage infection strategies 139
Final conclusions 141
References 143

List of Tables

Table 2.1 Single nucleotide polymorphisms (SNPs) from each of the 10 strains competed in this study 49

Table 2.2 Fold changes in bacterial genes in vivo: single CPS-expressing strains relative to the acapsular strain 50

Table 2.3 Minimum inhibitory concentrations of ciprofloxacin, polymyxin B, and human beta-defensin 3 (μg/ml). 65

Table 2.4 Strains used in this study 75

Table 2.5 Plasmids used in this study 76

Table 2.6 PCR primers used in this study 76

Table 2.7 qPCR primers used in this study 77

Table 2.8 Additional information on relative cps expression from previously published datasets 86

Table 2.9 Strains, genes, and capsular polysaccharide synthesis loci in comparative genomic and metagenomic analyses 92

Table 3.1 Phages used in this study and details on their isolation 107

Table 3.2 Bacterial strains and plasmids used in this study 118

Table 3.3 Primers used in this study 119

List of Figures

Figure 1.1 Representative diversity of polysaccharides derived from various dietary, host, and microbial sources. 4

Figure 1.2 Examples of how polysaccharides mediate interactions between the gut microbiota and host. 12

Figure 1.3 Multiple roles of microbially produced glycans (MPGs) in symbiotic bacterial fitness in the gut. 14

Figure 2.1 B. theta dynamically alters expression of its cps loci over time in the mouse gut. 42

Figure 2.2 Expression of cps loci by wild-type B. theta in vivo. 44

Figure 2.3 Competition of single CPS-expressing strains in germ-free mice reveals distinct advantages conferred by specific CPS. 46

Figure 2.4 Competition of single CPS-expressing strains in vivo, data from individual mice (see also Figure 2.3). 48

Figure 2.5 Consistency in single CPS-expressing strain relative abundance and cps locus relative expression over the length of the intestinal tract, and in vivo competition of single CPS-expressing strains with altered host diet. 56

Figure 2.6 Fecal IgA levels in in vivo competition of single CPS-expressing strains. 58

Figure 2.7 Specific growth rate of wild-type and single CPS-expressing strains in medium containing various monosaccharides or polysaccharides. 60

Figure 2.8 In vivo competition of the single CPS-expressing strains with the acapsular and WT strains. 62
Figure 2.9 Individual mice from in vivo competition with wild-type (WT) B. theta and susceptibility to innate immune factors (see also Figure 2.8). 64
Figure 2.10 Conservation of cps loci from B. theta VPI-5482 in other Bacteroidetes strains..... 67
Figure 2.11 Heterogeneity of cps loci at a single genomic site (Site 1). 68
Figure 2.12 Prevalence of B. theta cps biosynthetic loci in human metagenomic samples. 69
Figure 2.13 Monosaccharide composition of purified CPS. 83
Figure 2.14 Generation and validation of the single CPS-expressing strains utilized in this study. 85
Figure 3.1 Host range of B. thetaiotaomicron-targeting phages on strains expressing different CPS types. 110
Figure 3.2 Titers of Cluster 1 and Cluster 3 phages on strains with specific CPS locus deletions implicate different mechanisms for phage-CPS interactions. 112
Figure 3.3 Treatment of wild-type B. thetaiotaomicron with a low concentration of phage variably modulates CPS locus expression. 114
Figure 3.4 Treatment of wild-type B. thetaiotaomicron with a high concentration of phage modulates CPS locus expression. 116
Figure 4.1 Roles of individual Bacteroides CPS in interactions with the environment and the host immune response. 134
Figure 4.2 Expressing multiple CPS promotes Bacteroides resilience in the gut. 136

Abstract

Members of the *Bacteroides* and *Parabacteroides* genera, a prevalent portion of the gut microbiota, typically encode multiple loci for the synthesis of capsular polysaccharides (CPS). Strains may possess more than a dozen of these loci in a single genome, and locus expression is tightly regulated and may be highly dynamic. While some individual CPS have been characterized and confer advantageous properties on the bacterium (e.g. host immunomodulation), the functions of the majority of CPS are unknown. The model gut commensal *Bacteroides thetaiotaomicron* (*B. theta*) encodes for 8 distinct CPS synthesis loci, and these are coordinately regulated with other cellular functions such as carbohydrate utilization. I created a novel set of isogenic *B. theta* strains that each express one of the bacterium's eight CPS. These strains enable the identification of advantages conferred by individual CPS, as well as the role that encoding and expressing multiple CPS may play in optimal survival within the host organism.

This collection of single CPS-expressing strains was pooled and inoculated into germ-free mice, in which we could alter host genetic and dietary parameters. We found that specific CPS provided an advantage over others when in direct competition, and that adjusting levels of host immune response (especially adaptive immunity and IgA levels) altered the stringency of intra-strain competition. The CPS5-expressing strain outcompeted all other single CPS-expressing strains in the presence of adaptive immunity and even successfully competed with the

wild-type strain until the community was perturbed with antibiotics, which selected for the wild-type strain.

Additionally, I tested the ability of different CPS-expressing strains to be infected by bacteriophages. I determined that a large collection of bacteriophages isolated on this panel of *B. theta* strains are only able to infect a subset of the bacterial strains, directly linking CPS to bacteriophage resistance. Wild-type *B. theta* alters its CPS locus expression when exposed to bacteriophage and is able to quickly recover from phage infection, indicating that another purpose of encoding for the synthesis of multiple CPS is to render subpopulations of the bacterium phage resistant. My work describes roles for specific CPS, identifies advantages of switching between multiple CPS types, and provides a unique resource to further study CPS produced by human-associated commensal bacteria.

CHAPTER 1

Introduction

Abstract

The human intestine harbors a dense microbial ecosystem (microbiota) that is different between individuals, dynamic over time, and critical for aspects of health and disease. Dietary polysaccharides directly shape the microbiota because of a gap in human digestive physiology, which is equipped to assimilate only proteins, lipids, simple sugars, and starch, leaving nonstarch polysaccharides as major nutrients reaching the microbiota. A mutualistic role of gut microbes is to digest dietary complex carbohydrates, liberating host-absorbable energy via fermentation products. Emerging data indicate that polysaccharides play extensive roles in host-gut microbiota symbiosis beyond dietary polysaccharide digestion, including microbial interactions with endogenous host glycans and the importance of microbial polysaccharides. In this review, we consider multiple mechanisms through which polysaccharides mediate aspects of host-microbe symbiosis in the gut, including some affecting health. As host and microbial metabolic pathways are intimately connected with diet, we highlight the potential to manipulate this system for health.

Introduction

Polysaccharides are the most chemically diverse molecules in living systems[1]. This is due partly to the fact that they are composed of more than 20 common substituent sugars when one

counts different monosaccharides plus their variable pyranose (six-atom) and furanose (five-atom) ring forms (Figure 1.1). However, unlike nucleic acids and proteins, which are linked by uniform phosphodiester and peptide bonds, polysaccharides can be extensively diversified by their linkage patterns, which may be either α or ß and can be branched at more than two positions on a single substituent. Finally, carbohydrates can be covalently coupled to other common biological molecules, such as proteins and lipids (i.e., glycoconjugates), adopt secondary structure similar to proteins, and even exist in crystalline form, such as cellulose and some forms of starch.

Despite the astronomical diversity of polysaccharide combinations that may be synthesized from common component sugars, relatively little of this theoretical diversity has been explored by natural biological systems. A substantial portion of the known diversity exists in plants, fungi, and marine algae, including the variable pectins, hemicelluloses, α-mannans, and α-/ß-galactans (e.g., agarose and carrageenan) that are essential for the physiology and structural integrity of these organisms[2–5]. It is likely that even more diversity exists in the *O*- and *N*-linked glycolipid and glycoprotein conjugates found in mammalian tissues and secreted products, such as mucus and shed host cells[6–8].

Finally, microbes may be the most diverse and underexplored repositories of glycobiological content, given their ability to build not just essential structural polysaccharides (peptidoglycan and lipopolysaccharide) but also an extensive array of polysaccharide capsules and exopolysaccharides, many of which contain rare component sugars. All three of the above sources of glycobiological diversity (plant, animal, and microbial) converge in the human distal gut, where a dense and dynamic community of symbiotic microbes (microbiota) exists in homeostasis (from a microbial perspective, perhaps turbulent disequilibrium) with the host. The

gut microbiota plays many important roles in health and disease and is, in turn, influenced by the dynamic compositional and physiological changes that are exerted in this ecosystem by diet. As such, dietary polysaccharides are an obvious focus of rational interventions involving the microbiota. However, other host and microbial carbohydrate-based physiological pathways have also emerged as important factors in health and disease, and some of these are influenced by diet. In this chapter, we consider the many interconnected carbohydrate-based transactions that take place in the gut and involve the microbiota. We focus on the sources and chemical details of the complex carbohydrates that mediate these interactions, as well as emerging frontiers in understanding and manipulating these complex relationships to promote health.

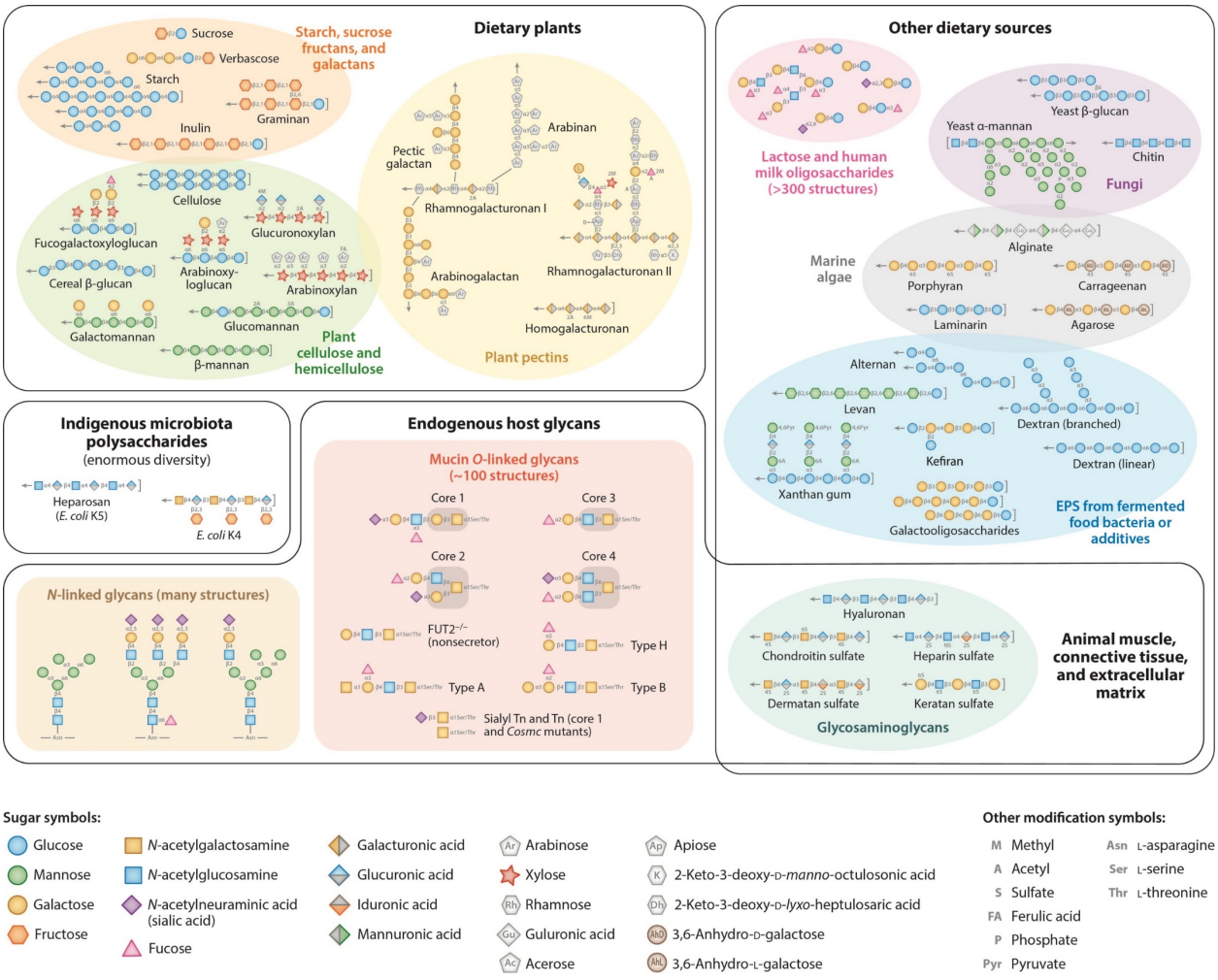


Figure 1.1 Representative diversity of polysaccharides derived from various dietary, host, and microbial sources.

Brackets indicate reducing ends of polysaccharides and arrowheads indicate the possibility of extended polymer length, which may be very long in some cases (e.g., corn starch with up to 10^6 glucose units per chain). Most of the polysaccharides shown are known to be degraded by human gut bacteria. The three major sources of polysaccharides (diet, endogenous host glycans, and microbially produced glycans indigenous to the microbiota) are arranged in order of increasing diversity, beginning in the upper left and proceeding clockwise through indigenous microbiota polysaccharides. For the latter, only two Escherichia coli *capsule (CPS) structures are shown, which overlap with the host glycosaminoglycans heparin and chondroitin sulfate. However, the diversity of microbiota CPS structures is likely astronomical and is still largely unexplored. Linkage types (α or ß) between sugars are indicated, and where the donor sugar is linked via carbon 1 to another monosaccharide, this number is not indicated (e.g., a ß1,4 linkage between two sugars is written as ß4).*

Influence of dietary polysaccharides on the microbiota-host symbiosis

The genomic era, coupled to reinvigorated culturing efforts for human gut microorganisms, has accelerated mechanistic research toward understanding a mutualistic behavior that has long been attributed to human gut microbes: the critical importance of these organisms in digesting dietary fiber polysaccharides[9]. While the human genome only encodes a handful of gastrointestinal enzymes that mostly target starch—thereby releasing glucose for direct absorption in the small intestine—some individual human gut bacteria produce hundreds of individual enzymes with collective catalytic specificities ranging far beyond starch[10]. For example, the Gram-negative symbiont *Bacteroides thetaiotaomicron*, one of the first common gut bacteria after *Escherichia coli* to have its genome sequenced, encodes over 300 enzymes in the glycoside hydrolase, polysaccharide lyase, carbohydrate esterase, and sulfatase families[11,12]. Despite this impressive armament of degradative enzymes, the sequenced type strain of this species is still incapable of utilizing a number of common polysaccharides as nutrient sources, most notably plant cell wall hemicelluloses that are abundant in grains, nuts, fruits, and vegetables[13]. However, systematic investigations of a few additional *Bacteroides* species (*B. ovatus*, *B. cellulosilyticus*, *B. xylanisolvens*), each with even larger polysaccharide-degrading enzyme repertoires than *B. thetaiotaomicron*'s, revealed abilities to degrade hemicellulosic polysaccharides and, like studies of *B. thetaiotaomicron*, identified the genes involved[13–16]. Even more intriguingly, studies focusing on the ability of human gut bacteria to catabolize relatively rare dietary polysaccharides, such as carrageenan, agarose, porphyran, and alginate polymers found in edible seaweed, have revealed the presence of these abilities in geographically and culturally distinct populations[17–19].

An impressive series of follow-up studies have targeted select multiprotein catabolic systems in some of the species noted above for in-depth biochemical and molecular genetic studies of the enzymatic, regulatory, and transport mechanisms involved[20–27]. Although many of these detailed efforts have focused on members of the Gram-negative *Bacteroidetes*, one of the dominant phyla colonizing the gut, similar studies are emerging for members of genera belonging to the two abundant Gram-positive phyla, *Firmicutes* and *Actinobacteria*[28–33]. The known mechanisms through which members of all three of these phyla degrade dietary polysaccharides, ranging from milk oligosaccharides to plant and fungal cell walls, have recently been reviewed in detail[34] and therefore are not covered in depth here. These mechanisms mostly involve either energy-dependent Sus-like transport systems coupled to outer membrane-associated and periplasmic enzymes in the *Bacteroidetes*[35,36] or ABC-transport systems coupled to degradative enzymes in the *Firmicutes* and *Actinobacteria*[29,30,32]. In addition, a third mechanism, cellulosome-like scaffolded enzyme systems, has recently been discovered in some human gut members of *Firmicutes* (namely members of the genus *Ruminococcus*) that so far have been found to use these systems to target cellulose and resistant starch[28,37].

While substantial need for mechanistic investigation remains, especially in Gram-positive gut bacteria, it has become clear that most of the known dietary fiber polysaccharide structures expected to be introduced into the gut (Figure 1.1), even insoluble cellulose, can be utilized by some gut bacteria. Two open questions are, how much variation exists across individuals, and how much redundancy exists among individual bacterial members of a single microbiota? The answer to each of these questions will be important for understanding how individual members of the microbiota metabolize various dietary polysaccharides, if they mostly do so alone or as consortia, and how resilient communities are to loss of species that conduct key digestive

functions. With respect to the latter point, it was recently shown in gnotobiotic mice colonized with a human microbiota that prolonged dietary fiber starvation is capable of catalyzing irreversible extinction of some microbial groups that are more consistently supported by a fiber-rich diet[38]. In this context, it has been hypothesized that the gut microbiota of modern humans—especially those living in industrialized nations where fiber consumption has diminished in recent decades—has lost carbohydrate-metabolizing bacteria over multiple generations[39]. If this is true, we may need to look to human populations that have retained high-fiber, unprocessed diets for repositories of microorganisms that can eventually be reintroduced (i.e., as probiotics) into humans with low digestive diversity.

Host glycoconjugates shape and are shaped by the microbiota

Changes in host intestinal glycosylation in response to symbiont colonization

A second major source of glycans present in the gut is the pool of endogenous molecules attached to cell surfaces, shed epithelial cells, and secreted mucus. The structures of these glycans, typically conjugated to either lipids or proteins (*O*-linked to serine or threonine or *N*-linked to asparagine), are distinct from those of dietary polysaccharides other than milk oligosaccharides, which are themselves produced by host mammals and can overlap substantially with *O*-linked glycans (Figure 1.1). The interactions between gut microbes and host glycans are already known to be manifold and can contribute to beneficial or detrimental outcomes for both the host and gut microbes. Interestingly, the glycobiology of the intestinal surface has been shown to be both temporally and geographically dynamic and responsive to microbiological signals, leading to the idea that changes in intestinal glycosylation occur in part to influence the microbiota. One of the first examples of this phenomenon is α-fucosylation of ileal epithelial

cells during development of the murine gut[40]. In weaned mice with a conventional microbiota, there is heavy fucosylation of the ileal surface. In genetically identical mice raised germfree, this fucosylation is absent but can be restored (even in adult germfree mice), if a single species (*B. thetaiotaomicron*) or a conventional microbiota is introduced. Even more interesting, *B. thetaiotaomicron* mutants that cannot utilize L-fucose as a nutrient source fail to elicit ileal fucosylation in germfree mice, suggesting the presence of a two-way signaling pathway by which microbial foraging of fucose from host glycans triggers increased production by the host[40,41]. This phenomenon of commensal-mediated cell surface fucosylation was subsequently recapitulated in cultured murine ileal organoids injected only with *B. thetaiotaomicron* culture. As observed in vivo[40], inoculation correlates with transcriptional activation of murine *FUT2* (α-1,2-fucosyltranferase), which encodes the enzyme responsible for building this host glycan linkage[42].

More recently, increased small intestinal fucosylation has also been linked to systemic signals associated with bacterial infection[43]. Intraperitoneal injection into mice of Toll-like receptor 2, 4, or 9 ligands that signal the presence of various bacterial molecules (triacylated lipoprotein, lipopolysaccharide, or CpG DNA, respectively) triggers a rapid increase in small intestine fucosylation that is also dependent on Fut2 activity. While the small intestine harbors a relatively low-density microbial community, newly fucosylated glycoproteins are degraded as they transit to the colon in a microbiota-dependent fashion. Thus, it has been hypothesized that such a response exists to provide endogenous sources of glycan-derived nutrients to the microbiota during times of stress due to bacterial infection[43]. Such a response may suppress the activity or proliferation of potential microbiota-resident pathogens, either by selectively feeding nutrients to indigenous mutualists or by altering host glycan structures to make them less

accessible to opportunistic pathogens. In the context of these studies highlighting microbiota-mediated intestinal glycan fucosylation, ~20% of North Americans and Europeans exhibit the nonsecretor blood group phenotype, attributed to loss of *FUT2* (Figure 1.1); the effect of this sequence variation has been shown to directly affect the composition and physiology of the microbiota in a diet-dependent fashion[44,45].

Host- and microbiota-catalyzed changes during disease

The studies described above examined changes in fucosylation of glycans secreted in the small intestine, a site of lower microbial colonization that may pass these nutrients to the distal gut. Glycoconjugates in the colon, a site with high colonization density, are also critical to the diet-microbiota-host axis. One of the most prominent sources of colonic glycans is secreted mucus, a mixture of mucin glycoproteins (Muc2 and Muc5ac are abundant in human and mouse colon) and other molecules, such as immunoglobulins[46]. The most abundant form of glycan nutrients attached to mucins are *O*-linked glycans (Figure 1.1), composing 50-80% of secreted mucin mass[47,48]. A number of taxonomically diverse gut bacteria have been shown to utilize mucus, including *B. thetaiotaomicron*[49,50], although clearly not all species possess this ability[51,52].

Mucus is secreted by goblet cells as a gel-forming layer overlying colonic epithelial cells. It functions, in part, as a physical barrier that protects the host from the dense microbiota (located just microns away in the gut lumen), as well as from invading pathogens[46]. A number of important microbiota-mucus interactions have been revealed in recent years. Near-complete elimination of this barrier through deletion of the murine *Muc2* gene results in the microbiota living immediately adjacent to host tissue, causing inflammation and eventual colorectal

cancer[53,54]. While mutations this severe are not known to naturally exist in humans, more subtle mucin glycosylation defects have been created in mouse models that also lead to increased susceptibility to either spontaneous or chemically induced inflammation resembling inflammatory bowel disease (IBD). Some of these mutations involve elimination of core glycosyltransferase enzymes (Figure 1.1), which like Fut2 mentioned above are involved in synthesizing host glycoconjugates[55,56]. Whereas Fut2 adds terminal fucose to existing chains, core 1 and core 3 glycosyltransferases build broad sets of common base structures that are subsequently extended and diversified. Thus, eliminating these key activities results in broad reductions in *O*-glycan diversity, potentially making it easier for colonizing microbes to degrade these protective structures and erode the function of this barrier. While mutations in the core glycosyltransferases noted above have also not yet been directly associated with human intestinal disorders, mutations in the X chromosome-linked endoplasmic reticulum protein *Cosmc* (a chaperone for core 1 glycosyltransferase) have been associated with human IBD. A recent study found that *Cosmc*-deficient mice fail to synthesize normal mucin glycosylations, instead decorating these important barrier molecules with truncated *O*-linked glycans, such as Tn antigen and sialyl-Tn (Figure 1.1)[57]. This defect in mucus glycosylation, in turn, leads to alterations in the colonic microbiota and spontaneous inflammation, underscoring the conclusion that genetic alterations in colonic glycoconjugates can predispose the host to disease.

While the examples discussed above outline paths by which gut microbes trigger host glycosylation or elicit disease in the absence of normal mucin glycan development, additional recent studies have revealed that the microbiota can be driven to erode the colonic mucous layer in the absence of dietary polysaccharides[58,59]. Because the dietary polysaccharides that collectively compose fiber from plants and other sources (Figure 1.1) are a predominant source

of nutrition for the microbiota, the absence of these nutrients forces some members of the microbiota to rely instead on indigenous host glycans. A shift from utilizing dietary versus host polysaccharide nutrients has been observed in gnotobiotic mice colonized with just one bacterium, *B. thetaiotaomicron*[50], which is otherwise metabolically programmed to preferentially degrade dietary fiber polysaccharides before mucin glycans[60–62] and may otherwise avoid foraging on host carbohydrates in the presence of sufficient dietary input. An important feature of the study noted above (in which a low-fiber diet catalyzed microbiota-mediated erosion of the colonic mucous layer) is that the wild-type mouse host remained healthy despite its diminished mucous barrier. However, when mice fed high-fiber (thick mucus) and low-fiber (thin mucus) diets were challenged with the enteric pathogen *Citrobacter rodentium*, those with a microbiota-eroded mucous layer experienced accelerated disease progression that resulted in lethal colitis in some animals[58].

Collectively, the studies described above underscore that both mutualistic bacteria and pathogens may alter the glycan landscape at different regions in the intestine. Moreover, these studies suggest the existence of a very complex system in which both dietary polysaccharides and members of the gut microbiota influence the host's response (Figure 1.2). Other emerging literature has linked *Bacteroides* commensals to release of sugars (e.g., sialic acid) for other bacteria, including pathogens, and has begun to uncover polysaccharide-digesting food chains consisting of multiple bacterial species[52,63,64]. Even bacteriophages, some of which have the ability to bind to mucous glycans via conserved lectin or immunoglobulin-like domains in their virion proteins, have entered the picture and have been hypothesized to provide a layer of adaptive immunity to the mucous layer by retaining high populations of phages that have recently killed the most successful colonizers[65,66]. Still, an even more complex and only lightly

explored horizon exists in the suite of polysaccharide capsules, exopolysaccharides, and cell walls that are produced by the many different symbiotic bacteria that are indigenous to the gut microbiota or that are present in fermented foods (Figure 1.1). Emerging roles for polysaccharides in this final category are considered in the next section of this chapter.

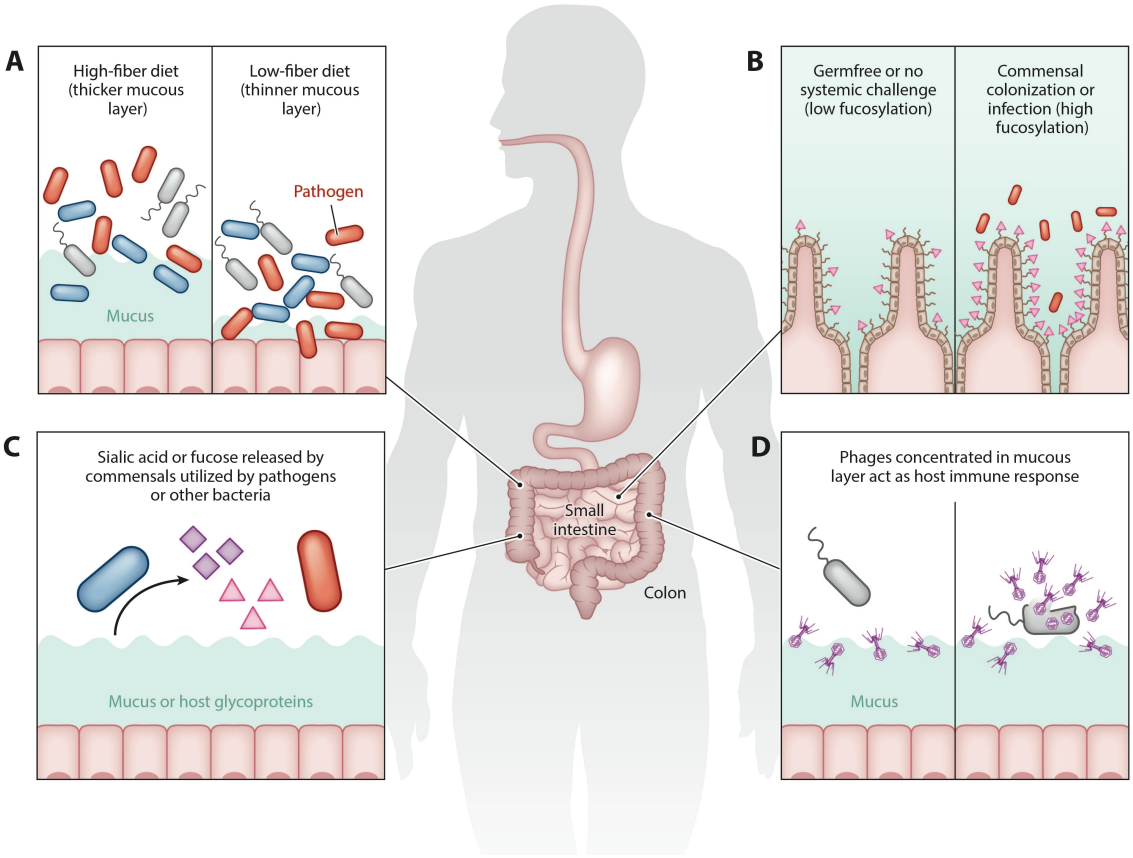


Figure 1.2 Examples of how polysaccharides mediate interactions between the gut microbiota and host.

A) On a high-fiber diet fewer bacteria access the colonic mucous glycoprotein layer for nutrients, but when dietary fiber is low, this barrier can be eroded by mucin-degrading bacteria, facilitating pathogen access to the host epithelium. B) When germfree mice are colonized with a conventional microbiota or a single species such as Bacteroides thetaiotaomicron *or when conventional mice are challenged with a systemic infection signal (bacterial TLR ligands), the small intestine epithelium increases the amount of surface and secreted fucosylated glycans. C) The metabolic activity of some commensal bacteria (including* Bacteroides spp.*) may release free sugars that are accessible by pathogens (*Clostridium difficile *or* Salmonella spp.[63]*) or other members of the microbiota, such as* Escherichia coli[43]*, that lack their own enzymes to liberate these sugars. D) Bacteriophages that have been released from recently killed bacteria can remain bound in the colonic mucous layer, potentially functionalizing it with phage-based adaptive immunity that is specific for bacterial populations that have recently been killed by these viruses.*

Biological effects of microbially produced glycans

An underexplored repository of polysaccharide diversity and function

In addition to facilitating dietary polysaccharide degradation, many gut microbes also possess the ability to synthesize a wide variety of glycan structures. These include exopolysaccharides (EPS), microbial glycans that are loosely associated with the cell surface and may easily slough off into the extracellular environment; and capsular polysaccharides (CPS), glycans that are more firmly attached to the cell surface. However, whether a glycan is attached to the cell surface or not is unclear in many gut microbes; thus, these terms are often used interchangeably. For the purposes of this chapter, we use the term microbially produced glycans (MPGs) to encompass both CPS and EPS. MPGs have been studied in many pathogenic microorganisms or their close symbiotic relatives, revealing extensive complexity, but unfortunately they have been studied far less in nonpathogenic microbes that compose most of the gut microbiota. Much of this research has focused on models like *Escherichia coli*, and inferences to symbiotic capsules are drawn from these systems and from studies on other pathogens. There is enormous diversity of capsule types displayed just by various *E. coli* strains (~80 known K antigens[67]), and emerging data from genomic sequencing efforts have revealed that polysaccharide synthesis genes are both highly abundant and overrepresented in gut bacteria relative to environmental samples[68,69]. Thus, gut bacterial MPGs are likely the most diverse group of polysaccharides relevant to host-microbiota interactions in the gut. Based on studies of just a handful of MPGs discussed below, it is likely that their biological effects are also broad (Figure 1.3), influencing host immunological responses, bacterial fitness, and other important aspects of the host-microbe relationship.

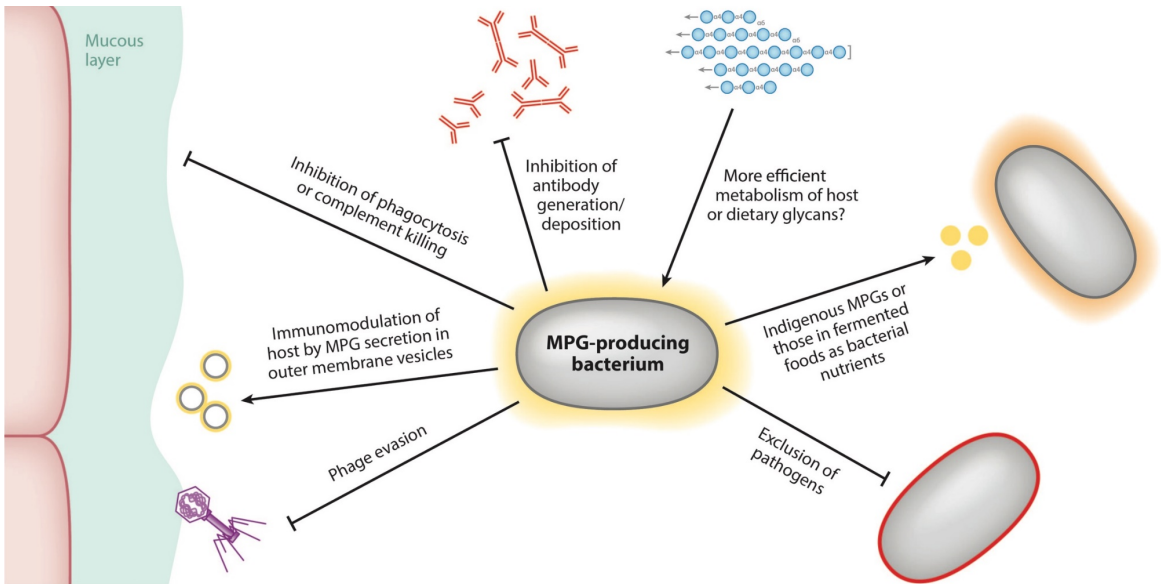


Figure 1.3 Multiple roles of microbially produced glycans (MPGs) in symbiotic bacterial fitness in the gut.

Various mechanisms by which MPGs enhance bacterial fitness are shown. No single model system has demonstrated all of these mechanisms, indicating that the MPG structure and/or bacterial niche dictates the role of each polysaccharide.

Synthesis

Multiple mechanisms exist for MPG synthesis, with the best-studied systems being in *E. coli*[67]. These include the Wzy-dependent mechanism, where assembly of repeat unit oligosaccharides occurs via glycosyltransferases in the cytoplasm; these oligosaccharides are then polymerized into longer chains on the cell surface (Gram-positive bacteria) or the periplasm, with subsequent export to the cell surface (Gram-negative bacteria). This contrasts with other mechanisms such as ABC transporter-dependent synthesis, where the entire glycan chain is polymerized in the cytoplasm. Synthase-dependent and glucansucrase/fructansucrase-dependent pathways produce relatively less complex glycans, with one or two monosaccharides incorporated into the glycan. All four of these pathways have been reviewed extensively elsewhere[67,70,71].

Diversity

Extensive genetic diversity of MPG biosynthetic loci in individual species of several common and abundant Gram-negative and Gram-positive genera of gut bacteria is well known, including *Escherichia*[67], *Lactobacillus*[72–75], *Bifidobacterium*[76], and *Bacteroides*[77,78]. Many gut bacteria encode only a single genomic locus for MPG synthesis, which may be variable among related strains or species. However, lactic acid bacteria often possess both a glucansucrase (or fructansucrase) for synthesizing a homopolysaccharide and a locus encoding a heteropolysaccharide, and *E. coli* strains often have a CPS biosynthetic locus as well as a locus for synthesis of the EPS colanic acid. Alternatively, members of some bacterial groups encode multiple gene clusters that enable them to produce larger numbers of MPGs[70,79]. Notably, some gut-associated *Bacteroidales* (*Bacteroides* and *Parabacteroides* genera) can encode as many as a dozen distinct CPS synthesis loci per genome, with expression of these factors being regulated in some species by a complex network of phase-variable promoters and transcription factors[80–82]. These latter observations imply additional benefits, at least for some species, for producing multiple, diverse CPS types in the gut environment and consistently varying their expression[83].

Contributions to the Diet-Microbiota-Host Axis During Intestinal Colonization

To persist in competitive gut ecosystems, intestinal microbes must either adhere to the mucosal surface or grow faster than the rate at which they are killed or excreted from the gut. While production of MPGs could be viewed as a waste of resources that would otherwise be devoted to microbial growth, based on their presence in widely divergent gut bacteria these structures apparently offer substantial benefits. Many probiotic strains that produce EPS benefit by longer gut persistence times or higher titers in mice, at least when competed against isogenic

EPS-deficient (EPS^-) strains[79,84–86]. However, this is not always the case. One study compared two *Lactobacillus* strains that differed in their ability to persist in the mouse gut and identified an EPS synthesis locus in the persisting strain that was also transcriptionally upregulated in the mouse gut. However, deletion of this locus did not affect gut persistence times[87]. Additionally, strain-to-strain variation exists in the benefits conferred by producing MPGs. Although Sims and colleagues found an EPS-producing (EPS^+) strain of *Lactobacillus reuteri* to be more competitive than an isogenic EPS^- strain[86], this was not the case in a different strain of *L. reuteri*[88]. However, the addition of *L. johnsonii* as another competitor gave an increased advantage to the wildtype, indicating that capsules may provide an advantage against a broad range of microbial competitors beyond close relatives[88]. Similar variability has been seen for *Bacteroides* spp., for which individual strains typically express multiple capsules. Whereas a *Bacteroides fragilis* mutant expressing just one of its eight CPS (PSH) outcompetes an acapsular mutant in vivo, the PSH-expressing strain competes equally with wild-type *B. fragilis* capable of expressing different capsules (individual capsules in this species are alphabetically named PSA through PSH)[89]. However, other mutants expressing just PSA, PSB, or PSC are all outcompeted against wild type in vivo[90]. Differences in competition in these studies likely point to mechanistic differences in capsules as fitness factors and underscore the need to better understand the precise roles these capsules play in symbiotic organisms.

Links between diet and MPG expression

There are multiple mechanisms by which MPGs may enhance bacterial survival in the gut (Figure 1.3). Since commensal gut microbes provide a critical capacity to digest dietary polysaccharides, capsules with specific structures or properties may provide optimal access to

certain nutrients or facilitate the most efficient use of the bacterial cell's resources. When lactic acid bacteria are cultured on sucrose, for example, they can hydrolyze sucrose into its monosaccharide constituents, glucose and fructose, and utilize a portion of this energy pool to directly synthesize extracellular glucans or fructans (see EPS derived from fermented food bacteria in Figure 1.1)[70]. The amount of EPS produced varies with carbohydrate growth source[75]. Moreover, in *Bacteroides thetaiotaomicron*, changes in the host diet alter expression of this species' eight CPS synthetic loci when it is the sole colonizer in vivo. Specifically, in low dietary fiber conditions in the mouse gut, where *B. thetaiotaomicron* can access only host-derived glycans as a main nutrient source, it shifts capsule expression away from CPS4 (the most preferred on a high-fiber diet) toward CPS5 and CPS6 (in this species capsules are named CPS1 through CPS8)[50,91]. Additionally, molecular genetic studies that either inappropriately activated polysaccharide degrading systems *in vitro* or impaired their function *in vitro* and *in vivo*[92,93], suggest that particular capsules equip this species to proliferate optimally in conditions where different dietary or host-derived nutrients are available. However, direct evidence connecting expression of any particular CPS to optimal use of a specific polysaccharide nutrient is still lacking.

Interactions with the immune system

The capsules produced by pathogenic microbes protect them in various ways from the host immune response[94], including from innate immune mechanisms, such as complement-mediated killing or phagocytosis. Although MPGs in symbiotic microbes do not generally appear to provide resistance to prominent environmental challenges like bile acids (although bile can

increase EPS expression in one strain[95]), they may provide increased resistance to complement and/or antimicrobial peptides in certain bacteria[84,89].

Nonpathogenic microbes can also avoid host recognition via MPG production. EPS^+ variants of one *Bifidobacterium breve* strain with two EPS loci elicit lower levels of inflammatory cytokines and inflammatory cell types than their EPS^- counterparts (which lack both EPS loci)[79]. Higher levels of plasma cells, higher titers of fecal IgA, and increased levels of T cells producing inflammatory cytokines were seen in mice exposed to EPS^- strains. Additionally, the EPS^+ wild-type strain is able to persist longer than EPS^- strains in wild-type mice, but this difference is abolished in B cell-deficient mice, indicating a role for the EPS in evading adaptive immune responses. Serum against the EPS^+ wild-type strain was able to agglutinate wild-type cells but poorly agglutinated EPS^- cells and vice versa, indicating that the presence of EPS protects underlying cell surface antigens from detection[79]. This is supported by work in *L. johnsonii* where antibodies to an EPS^+ strain bound to an EPS^- strain at higher titers than to wild type, suggesting that EPS in part acts to block antibody recognition. A mutant overexpressing EPS had similar levels of bound antibody compared to wild type, indicating that the EPS itself is a poor immunogen[96].

While MPGs may conceal the microbe from the host immune system, some MPGs can directly modulate its effects. The best studied of these are termed zwitterionic polysaccharides (ZPS) and are capable of manipulating both the innate and adaptive immune systems[97,98]. ZPS contain alternating positive and negative charges, which are essential to their modulatory functions. The prototypical ZPS, PSA from *B. fragilis*, was first studied in a rodent model of anaerobic abscess formation. Treatment with sterile cecal contents along with *B. fragilis* or purified PSA (but not a PSA-deficient mutant) elicits disease in this model, whereas CPS from

other pathogenic organisms were not capable of this effect[77,99,100]. Chemically modifying PSA and other ZPS to neutralize their positive or negative charges reduced their ability to induce abscesses. Moreover, modification of a non-abscess-inducing capsule to contain both positively and negatively charged residues enabled it to elicit disease (abscess formation), indicating that it is the zwitterionic nature of the polysaccharides that allows them to induce this process[100].

Interestingly, without sterile cecal contents, prophylactic administration of purified PSA reduced abscess formation instigated by diverse pathogens, apparently working in a nonspecific fashion to suppress disease progression. Injection of naive mice with T cells from PSA-treated mice was sufficient to reduce abscess formation, pointing to a direct effect on the host immune system[101]. PSA was later studied in the context of its role as a mutualistic factor influencing the adaptive immune system, including $CD4^+$ T helper cells. PSA is capable of increasing $CD4^+$ T cell populations in germfree mice and altering the balance of Th1 and Th2 (subtypes of T helper cells that promote different immune responses to diverse pathogens) immune responses to levels that are similar to those in conventionally raised animals[102]. Additionally, PSA stimulates Tregs (T helper cells that act to dampen or halt the immune response) to produce more anti-inflammatory IL-10 and reduce inflammatory Th17 cell levels (T helper cells that instigate a pro-inflammatory immune response). Reducing levels of Th17 cells allows *B. fragilis* to reside in close association with the host epithelium, a niche preferred by this bacterium[103]. Neff and colleagues used the genes necessary for creating PSA's positively charged motif to probe genomes of phylogenetically diverse bacteria for potential ZPS synthesis loci[104]. This motif is found widely throughout divergent bacterial isolates, and ZPS-encoding bacteria induce greater expression of IL-10 and Treg induction in human cells than those without a similar motif. Deletion of one ZPS synthesis locus from *Bacteroides cellulosilyticus* resulted in less IL-10

expression and fewer Tregs[104]. These data support the idea that ZPS from diverse taxa can directly alter the host immune response.

Other MPGs may play similar roles as the ZPS previously described. For example, an EPS^- strain of *Bifidobacterium longum* elicits greater levels of proinflammatory cytokines than an isogenic EPS^+ strain. Addition of purified EPS can at least partially rescue wild-type levels of some of these cytokines, and EPS^+ (but not EPS^-) strains reduce inflammation in a mouse model of colitis[105]. As this EPS does not contain zwitterionic charges, it may act through a different mechanism than the ZPS described above. In another model, intraperitoneal injection of purified *Bacillus subtilis* EPS reduces disease in a *C. rodentium* model of bacterial enteric infection and hyperplasia. Injection of naive mice with peritoneal cells isolated from the mice treated with EPS was sufficient to reduce disease, indicating a direct effect on host cells[106]. Differences in induced immune cell and cytokine profiles exist for isogenic EPS^+ and EPS^- strains of other bacteria. For instance, treatment of macrophages with EPS^+ *Pediococcus parvulus* elicited lower inflammatory cytokine levels (and more anti-inflammatory IL-10) than treatment with an EPS^- strain[107]. Moreover, mice colonized with EPS^+ *L. reuteri* had higher levels of Treg cells in the spleen compared to mice not colonized by *Lactobacillus* or colonized with an EPS^- strain[86]. As with PSA and *B. subtilis* EPS, treatment of cultured cells or animal models with purified MPGs in these systems could isolate the effects of the bacterium from the direct effects of the capsule and establish whether these and other glycans directly affect the host immune response.

Exclusion of pathogens

While there is evidence that bacterial pathogens take advantage of indigenous microbes to establish infection via mechanisms that either directly involve carbohydrate metabolism or

fermentative end products[63,108,109], mutualistic microbes may likewise influence pathogens via MPG-based mechanisms. For example, a recent study showed that treatment of mice with an EPS-producing *B. breve* strain (but not an isogenic EPS-deficient strain) prior to infection with the pathogen *C. rodentium* reduced pathogen levels in stool[79]. One mechanism by which mutualistic bacteria may exclude intestinal pathogens is by inhibiting their attachment to host cells. Individual additions of several different *L. reuteri* EPS (including the starch-like polysaccharide reuteran; Figure 1.1) reduces hemagglutination of enterotoxigenic *E. coli* (ETEC) with red blood cells (a proxy assay for binding to mucus and other glycan-based host receptors)[110]. Moreover, treatment of pig small intestine with reuteran in an ex vivo model reduces attachment of ETEC to host cells[111]. Interestingly, not all EPS influence pathogen toxicity by reducing pathogen abundance and may therefore work via other mechanisms. For instance, in another model system, while treatment with EPS-producing *B. subtilis* did not reduce *C. rodentium* titers, treatment with either the EPS^+ strain or its purified EPS (but not with an isogenic EPS^- strain) did alleviate disease symptoms[106,112]. As discussed above, such mechanisms could involve alterations to the host's immune or inflammatory state.

Degradation by other microbes

Several studies have identified changes in complex gut-derived communities upon treatment with MPGs. These studies typically involve the addition of various purified polysaccharides to a fecal slurry of bacteria followed by identification of changes in community structure, including increases in *Bifidobacteria*, *Bacteroides*, and other taxa[113–116]. However, these studies either suffer from low-throughput identification of individual taxa or rely on characterization of higher-order taxonomic groups via PCR-based techniques. Recent advances

in sequencing technology have enabled higher-throughput assays to determine the microbial taxa most affected by MPG treatment and may point to mechanisms of community changes across individuals. Additionally, although effects on community structure may be due to MPG utilization as a nutrient source in some cases, not all MPGs are degraded by the gut microbiota. Degradation of several purified MPGs (but not many others) was demonstrated in a fecal slurry-based culture from one individual[114]. Additionally, EPS from *Lactococcus lactis* that was fed to rats was recovered undigested from the stool[117]. *B. thetaiotaomicron* illustrates both of these scenarios as it can degrade levan from *L. reuteri* but poorly degrades other glycans from *L. reuteri*[118]. From the viewpoint of the microorganism producing the glycan, providing MPGs as a nutrient for neighboring bacteria may be one mechanism underlying a mutualistic relationship. For instance, cross-feeding of diet-based oligosaccharides has been established between strains of *B. ovatus* and *Bacteroides vulgatus*. Additional unidentified growth-promoting factors exist in this relationship and could be MPGs[119,120]. However, MPG degradation may also be detrimental to the producing organism, and this may be one way in which bacteria undermine competitors to increase their own gut fitness.

Protection from bacteriophages

Bacteriophages are ubiquitous on the planet, and the gut is no exception. While studies focused on the taxonomic assembly of the human gut microbiome (the collection of organisms living in the gut), usually focused on bacteria, now abound thanks to next-generation sequencing, little is known about intestinal viruses, including bacteriophages[121]. Bacteriophages isolated from human stool samples can target bacteria in model gut communities in mice, but few phage receptors on gut bacteria have been identified[122]. MPGs provide one way for bacteria to shield

cell surface receptors from phage recognition and adsorption. For instance, one group compared the formation of spontaneous phage-resistant mutants in isogenic CPS^+ and CPS^- *E. coli* strains[123]. Whereas the wild-type CPS^+ strain produced mucoid (indicating increased capsule production) phage-resistant mutants, abolishing CPS production eliminated generation of phage-resistant mutants, implicating capsule as a key factor in mediating phage resistance[123]. Moreover, experimental production of K1 CPS in a hybrid *E. coli* strain renders the cell resistant to infection by T7 phage[124].

Unfortunately for the bacterium, phages have at least two mechanisms to circumvent this protective layer. First, the MPG itself may provide a receptor for the phage, as has been well described for various *E. coli* capsules[125–127]. Second, phages often encode depolymerases (hydrolases or lyases) capable of degrading the capsule, allowing the phage to tunnel through the thick CPS layer and bind to receptors on the bacterial surface[126,128]. At least some phages are able to infect strains of more than one capsule type—these produce two or more depolymerases[126,129,130].

The great diversity in CPS types in *E. coli* and in other species may be, at least in part, an arms race between phage and bacterium, as bacteria gradually develop new surface capsules and phage adapt to bind to or depolymerize these layers. Interestingly, phages with depolymerases acting on the same CPS substrate may be genomically divergent from each other (and most similar to other phages that do not encode these same depolymerases)[131]. Additionally, adaptor proteins can connect various depolymerases to the outer surface of the phage virion, allowing a more facile acquisition and incorporation of new, diverse degradative enzymes and likely explaining the tropism of coliphages discussed above[131,132]. These data indicate a role for horizontal gene transfer in tailoring a phage to its host capsule type.

While CPS-binding phages are relatively well studied in *E. coli*, little is known about phage-MPG dynamics in other commensal organisms. EPS may play a role in blocking phage infection of *L. lactis*, as an EPS$^-$ strain was less sensitive to phage infection, although the difference was marginal[117]. Additionally, one study noted the inability of a phage to infect encapsulated strains of *L. lactis* while the phage was still able to degrade their EPS[133]. Finally, one group identified unique EPS structures produced by different subtypes of an EPS synthesis locus containing variable glycosyltransferases[72]. Altering a locus to change its subtype (and EPS structure) also yielded a corresponding change in phage specificity from one subtype to the other, providing direct evidence for phage susceptibility and EPS type[72]. Whereas phages for many bacteria would only have to target a single capsule on an individual cell, switching between synthesis of multiple capsule types provides a potential mechanism for *Bacteroides* to evade its viral predators. Previously one group noted an unstable CPS phenotype in a *B. thetaiotaomicron* strain (in hindsight, likely due to capsule switching) that correlated with productive phage infection[134]. Others noted similar phenotypes in other *Bacteroides* strains that correlated with differences in susceptibility to phage infection[135,136]. Experiments controlling for MPG expression are needed to better isolate the role of MPG in blocking/facilitating phage infection in symbiotic organisms other than *E. coli*.

Conclusion and prospectus

It is clear that polysaccharides play many central roles in the human-microbe symbiosis that occurs in the distal gut. The accessibility of dietary polysaccharides to the distal gut microbiota—due to limitations of host enzymatic potential—have been known for a long time. Still, the precise impacts of dietary polysaccharides on microbiota physiology and subsequent

effects on host health are just now being unraveled in molecular mechanistic detail. The landscape of host-derived glycans, such as those attached to mucin glycoproteins and other glycoconjugates, may be even more complex than those of dietary glycans. These molecules change dynamically due to host status (microbial colonization or infection) and vary in composition along the length of the gastrointestinal tract. Moreover, because lack of dietary fiber forces some members of the microbiota to utilize host-derived glycans as a nutrient source, the status of host glycan pools, such as the mucous layer, is directly coupled to diet. Finally, the vast array of different MPGs that are either produced by microbes used in food production/fermentation or indigenous to our gut microbiota play several roles, some of which (e.g., immune modulation by some bacterial MPGs) directly affect human health. While gut bacteria arguably produce most MPGs for their own immediate benefit, such as surviving attacks from the host immune system, phages, or environmental challenges, it is perhaps not surprising that the large number of glycobiological configurations that have been explored by gut bacteria have produced other biological effects.

MPGs could be harnessed to benefit human health in several ways. First, purified immunomodulatory glycans, such as *B. fragilis* PSA, could be used to directly treat inflammatory disease. Such approaches have been successful in the treatment of diseases of the nervous system and colon in mouse models[137,138]. An alternative approach might be to introduce species that naturally produce, or are engineered to produce, biologically active MPGs as probiotics, perhaps even coupled to feeding of dietary polysaccharides that are specifically chosen to support the producing strain. Given the potentially enormous, yet still mostly unexplored, number of connections between dietary, host, and microbial polysaccharides, it is likely that these molecules share more than a chemical lexicon. Indeed, the effects of

polysaccharides from each of these three sources are intrinsically linked, raising the possibility that with better understanding of these manifold interactions we will be able to easily manipulate the effects of these molecules (for example, through changes in diet or microbiota) to improve human health.

Chapter outline

While MPGs such as CPS may potentially carry out various functions in mediating bacterial-environment interactions, few studies have tested the direct role of CPS in these interactions through genetic knockouts or other means, including in members of the prevalent *Bacteroides* genus. This dissertation examines multiple pressures that could affect bacterial fitness in the gut: dietary changes, host immune response targeting, and bacteriophage predation.

In Chapter 2, I describe and validate a novel set of isogenic strains of the model gut symbiont *Bacteroides thetaiotaomicron*. Each of these strains expresses just one of the eight CPS that the wild-type strain expresses (or none at all), allowing us to isolate the role of each individual CPS in various competitive scenarios. Specific CPS confer a fitness advantage on the bacterium when competed against the other CPS-expressing strains in germ-free mice. Increasing host adaptive immunity selects for one particular CPS type, CPS5. On the other hand, dietary changes (both *in vitro* and *in vivo*) appear to exert marginal effects on fitness of the various CPS-expressing strains.

In Chapter 3, I describe a novel collection of bacteriophages isolated on the single CPS-expressing strains. Each phage has a limited host range on the panel of bacterial strains, only infecting a subset of strains expressing different CPS. Isolated phages phenotypically cluster into at least 3 broad groups. The majority of phage isolates in two of these clusters infect the

acapsular strain, indicating that these likely have a cell surface target that is obscured by the CPS layer. Phages in the last cluster generally do not robustly infect the acapsular strain, and these phages likely exploit CPS as a required receptor. Similar phenotypes have emerged from studies in *E. coli* (as described above[124–127]). These and our experiments in *Bacteroides* implicate CPS in facilitating commensal bacterial interactions with both the host and bacteriophage.

Notes

Portions of this chapter were reprinted and modified with permission from Porter, N. T. & Martens, E. C. The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu. Rev. Microbiol.* **71,** 349–369 (2017).

Literature cited

1. Bertozzi, C. R. & Rabuka, D. Structural basis of glycan diversity. in *Essentials of Glycobiology. 2nd edition* (eds. Varki, A., Cummings, R. D. & Esko, J.) (Cold Spring Harbor Press, 2009).
2. Arana, D. M. *et al.* The role of the cell wall in fungal pathogenesis. *Microb Biotechnol* **2,** 308–320 (2009).
3. Caffall, K. H. & Mohnen, D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res* **344,** 1879–1900 (2009).
4. Hehemann, J. H., Boraston, A. B. & Czjzek, M. A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Curr Opin Struct Biol* **28,** 77–86 (2014).
5. Pauly, M. *et al.* Hemicellulose biosynthesis. *Planta* **238,** 627–642 (2013).
6. Holmén Larsson, J. M., Thomsson, K. A., Rodríguez-Piñeiro, A. M., Karlsson, H. & Hansson, G. C. Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. *Am. J. Physiol. Gastrointest. Liver Physiol.* **305,** G357-63 (2013).
7. Stanley, P., Schacter, H. & Taniguchi, N. N-glycans. in *Essentials of Glycobiology* (eds. Varki, A., Cummings, R. D. & Esko, J.) (Cold Spring Harbor, 2009).
8. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front Genet* **6,** 81 (2015).
9. Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3,** 289–306 (2012).
10. El Kaoutari, A. *et al.* The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11,** 497–504 (2013).
11. 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).
12. Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science (80-.).* **299,** 2074–2076 (2003).
13. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9,** e1001221 (2011).
14. Despres, J. *et al.* Unraveling the pectinolytic function of Bacteroides xylanisolvens using a RNA-seq approach and mutagenesis. *BMC Genomics* **17,** 147 (2016).
15. Despres, J. *et al.* Xylan degradation by the human gut Bacteroides xylanisolvens XB1A(T) involves two distinct gene clusters that are linked at the transcriptional level. *BMC Genomics* **17,** 326 (2016).
16. McNulty, N. P. *et al.* Effects of diet on resource utilization by a model human gut

microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome. *PLoS Biol.* **11,** e1001637 (2013).

17. Hehemann, J. H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464,** 908–912 (2010).

18. Hehemann, J.-H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 19786–19791 (2012).

19. Thomas, F. *et al.* Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut Bacteroides. *Env. Microbiol* **14,** 2379–2394 (2012).

20. Bagenholm, V. *et al.* Galactomannan catabolism conferred by a polysaccharide utilization locus of Bacteroides ovatus: enzyme synergy and crystal structure of a beta-mannanase. *J Biol Chem* (2016). doi:10.1074/jbc.M116.746438

21. Cuskin, F. *et al.* How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc. Natl. Acad. Sci. U. S. A.* (2012). doi:10.1073/pnas.1212034109

22. Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506,** 498–502 (2014).

23. Reddy, S. K. *et al.* A beta-mannan utilization locus in Bacteroides ovatus involves a GH36 alpha-galactosidase active on galactomannans. *FEBS Lett* **590,** 2106–2118 (2016).

24. Rogowski, A. *et al.* Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6,** 7481 (2015).

25. Schwalm, N. D., Townsend, G. E. & Groisman, E. A. Prioritization of polysaccharide utilization and control of regulator activation in Bacteroides thetaiotaomicron. *Mol. Microbiol.* **104,** 32–45 (2017).

26. Schwalm, N. D., Townsend, G. E. & Groisman, E. A. Multiple Signals Govern Utilization of a Polysaccharide in the Gut Bacterium Bacteroides thetaiotaomicron. *MBio* **7,** e01342-16 (2016).

27. Raghavan, V., Lowe, E. C., Townsend, G. E., Bolam, D. N. & Groisman, E. A. Tuning transcription of nutrient utilization genes to catabolic rate promotes growth in a gut bacterium. *Mol. Microbiol.* **93,** 1010–25 (2014).

28. Ben David, Y. *et al.* Ruminococcal cellulosome systems from rumen to human. *Env. Microbiol* **17,** 3407–3426 (2015).

29. Cockburn, D. W. *et al.* Molecular details of a starch utilization pathway in the human gut symbiont Eubacterium rectale. *Mol Microbiol* **95,** 209–230 (2015).

30. Ejby, M. *et al.* An ATP binding cassette transporter mediates the uptake of α-(1,6)-linked dietary oligosaccharides in Bifidobacterium and correlates with competitive growth on these substrates. *J. Biol. Chem.* **291,** 20220–31 (2016).

31. Ejby, M. *et al.* Structural basis for arabinoxylo-oligosaccharide capture by the probiotic Bifidobacterium animalis subsp. lactis Bl-04. *Mol Microbiol* **90,** 1100–1112 (2013).

32. Garrido, D., Kim, J. H., German, J. B., Raybould, H. E. & Mills, D. A. Oligosaccharide binding proteins from Bifidobacterium longum subsp. infantis reveal a preference for host glycans. *PLoS One* **6,** e17315 (2011).

33. Morrill, J. *et al.* The GH5 1,4-beta-mannanase from Bifidobacterium animalis subsp. lactis Bl-04 possesses a low-affinity mannan-binding module and highlights the diversity of mannanolytic enzymes. *BMC Biochem* **16,** 26 (2015).

34. Cockburn, D. W. & Koropatkin, N. M. Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. *J. Mol. Biol.* **428,** 3230–52 (2016).

35. Foley, M. H., Cockburn, D. W. & Koropatkin, N. M. The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cell. Mol. Life Sci.* **73,** 2603–17 (2016).

36. 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).

37. Ze, X. *et al.* Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic Firmicutes bacterium Ruminococcus bromii. *MBio* **6,** e01058-15 (2015).

38. Sonnenburg, E. D. *et al.* Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529,** 212–215 (2016).

39. Sonnenburg, E. D. & Sonnenburg, J. L. Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab* **20,** 779–86 (2014).

40. Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. A model of host-microbial interactions in an open mammalian ecosystem. *Science (80-.).* **273,** 1380–1383 (1996).

41. Hooper, L. V., Xu, J., Falk, P. G., Midtvedt, T. & Gordon, J. I. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9833–8 (1999).

42. Engevik, M. A. *et al.* Loss of NHE3 alters gut microbiota composition and influences Bacteroides thetaiotaomicron growth. *Am J Physiol Gastrointest Liver Physiol* **305,** G697-711 (2013).

43. Pickard, J. M. *et al.* Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. *Nature* **514,** 638–641 (2014).

44. Kashyap, P. C. *et al.* Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17059–64 (2013).

45. Kelly, R. J., Rouquier, S., Giorgi, D., Lennon, G. G. & Lowe, J. B. Sequence and expression of a candidate for the human Secretor blood group alpha(1,2)fucosyltransferase

gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J. Biol. Chem.* **270,** 4640–9 (1995).

46. McGuckin, M. A., Linden, S. K., Sutton, P. & Florin, T. H. Mucin dynamics and enteric pathogens. *Nat Rev Microbiol* **9,** 265–278 (2011).

47. Axelsson, M. A., Asker, N. & Hansson, G. C. O-glycosylated MUC2 monomer and dimer from LS 174T cells are water-soluble, whereas larger MUC2 species formed early during biosynthesis are insoluble and contain nonreducible intermolecular bonds. *J Biol Chem* **273,** 18864–18870 (1998).

48. Johansson, M. E., Sjovall, H. & Hansson, G. C. The gastrointestinal mucus system in health and disease. *Nat Rev Gastroenterol Hepatol* **10,** 352–361 (2013).

49. 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,** 447–57 (2008).

50. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307,** 1955–9 (2005).

51. Hoskins, L. C. & Boulding, E. T. Mucin degradation in human colon ecosystems. Evidence for the existence and role of bacterial subpopulations producing glycosidases as extracellular enzymes. *J Clin Invest* **67,** 163–172 (1981).

52. 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).

53. 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 U S A* **105,** 15064–15069 (2008).

54. Van der Sluis, M. *et al.* Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* **131,** 117–29 (2006).

55. Bergstrom, K. *et al.* Core 1- and 3-derived O-glycans collectively maintain the colonic mucus barrier and protect against spontaneous colitis in mice. *Mucosal Immunol.* **10,** 91–103 (2017).

56. Fu, J. *et al.* Loss of intestinal core 1–derived O-glycans causes spontaneous colitis in mice. *J. Clin. Invest.* **121,** 1657–1666 (2011).

57. Kudelka, M. R. *et al.* Cosmc is an X-linked inflammatory bowel disease risk gene that spatially regulates gut microbiota and contributes to sex-specific risk. *Proc Natl Acad Sci U S A* **113,** 14787–14792 (2016).

58. Desai, M. S. *et al.* A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167,** 1339–1353 e21 (2016).

59. Earle, K. A. A. *et al.* Quantitative imaging of gut microbiota spatial organization. *Cell Host Microbe* **18,** 478–88 (2015).

60. Lynch, J. & Sonnenburg, J. Prioritization of a plant polysaccharide over a mucus carbohydrate is enforced by a Bacteroides hybrid two-component system. *Mol. Microbiol.*

85, 478–491 (2012).

61. Pudlo, N. A. *et al.* Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. *MBio* **6,** e01282-15 (2015).

62. Rogers, T. E. *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol. Microbiol.* **88,** 1–15 (2013).

63. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502,** 96–9 (2013).

64. Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* **6,** 1535–1543 (2012).

65. Barr, J. J. *et al.* Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 10771–6 (2013).

66. Barr, J. J., Youle, M. & Rohwer, F. Innate and acquired bacteriophage-mediated immunity. *Bacteriophage* **3,** e25857 (2013).

67. Whitfield, C. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annu. Rev. Biochem.* **75,** 39–68 (2006).

68. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158,** 412–421 (2014).

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

70. Hijum, S. a F. T. Van, Kralj, S., Ozimek, L. K., Dijkhuizen, L. & Geel-schutten, I. G. H. Van. Structure-Function Relationships of Glucansucrase and Fructansucrase Enzymes from Lactic Acid Bacteria Structure-Function Relationships of Glucansucrase and Fructansucrase Enzymes from Lactic Acid Bacteria. *Microbiol. Mol. Biol. Rev.* **70,** 157–176 (2006).

71. Yother, J. Capsules of Streptococcus pneumoniae and other bacteria: paradigms for polysaccharide biosynthesis and regulation. *Annu. Rev. Microbiol.* **65,** 563–81 (2011).

72. Ainsworth, S. *et al.* Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. *MBio* **5,** e00880-14 (2014).

73. Berger, B. *et al.* Similarity and differences in the Lactobacillus acidophilus group identified by polyphasic analysis and comparative genomics. *J. Bacteriol.* **189,** 1311–1321 (2007).

74. Molenaar, D. *et al.* Exploring Lactobacillus plantarum genome diversity by using microarrays. *J. Bacteriol.* **187,** 6119–6127 (2005).

75. Raftis, E. J., Salvetti, E., Torriani, S., Felis, G. E. & O'Toole, P. W. Genomic diversity of Lactobacillus salivarius. *Appl. Environ. Microbiol.* **77,** 954–65 (2011).

76. Ferrario, C. *et al.* Modulation of the eps-ome transcription of bifidobacteria through simulation of human intestinal environment. *FEMS Microbiol. Ecol.* **92,** 1–12 (2016).

77. Coyne, M. J. *et al.* Polysaccharide biosynthesis locus required for virulence of Bacteroides

fragilis. *Infect. Immun.* **69,** 4342–50 (2001).

78. Patrick, S. *et al.* Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of Bacteroides fragilis. *Microbiology* **156,** 3255–69 (2010).

79. Fanning, S. *et al.* Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc. Natl. Acad. Sci.* **109,** 2108–2113 (2012).

80. Chatzidaki-Livanis, M., Coyne, M. J. & Comstock, L. E. A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of Bacteroides fragilis. *J. Bacteriol.* **191,** 7288–95 (2009).

81. Chatzidaki-Livanis, M., Weinacht, K. G. & Comstock, L. E. Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont Bacteroides fragilis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 11976–80 (2010).

82. Coyne, M. J., Weinacht, K. G., Krinos, C. M. & Comstock, L. E. Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 10446–51 (2003).

83. Coyne, M. J. & Comstock, L. E. Niche-specific features of the intestinal Bacteroidales. *J. Bacteriol.* **190,** 736–42 (2008).

84. Lebeer, S., Claes, I. J. J., Verhoeven, T. L. A., Vanderleyden, J. & De Keersmaecker, S. C. J. Exopolysaccharides of Lactobacillus rhamnosus GG form a protective shield against innate immune factors in the intestine. *Microb. Biotechnol.* **4,** 368–74 (2011).

85. Lindström, C., Holst, O., Nilsson, L., Oste, R. & Andersson, K. E. Effects of Pediococcus parvulus 2.6 and its exopolysaccharide on plasma cholesterol levels and inflammatory markers in mice. *AMB Express* **2,** 66 (2012).

86. Sims, I. M. *et al.* Structure and functions of exopolysaccharide produced by gut commensal Lactobacillus reuteri 100-23. *Isme J* **5,** 1115–1124 (2011).

87. Denou, E. *et al.* Identification of genes associated with the long-gut-persistence phenotype of the probiotic Lactobacillus johnsonii strain NCC533 using a combination of genomics and transcriptome analysis. *J. Bacteriol.* **190,** 3161–3168 (2008).

88. Walter, J., Schwab, C., Loach, D. M., Gänzle, M. G. & Tannock, G. W. Glucosyltransferase A (GtfA) and inulosucrase (Inu) of Lactobacillus reuteri TMW1.106 contribute to cell aggregation, in vitro biofilm formation, and colonization of the mouse gastrointestinal tract. *Microbiology* **154,** 72–80 (2008).

89. 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. U. S. A.* **105,** 13099–104 (2008).

90. 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. U. S. A.* **105,** 3951–6 (2008).

91. Bjursell, M. K., Martens, E. C. & Gordon, J. I. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. *J. Biol. Chem.* **281,** 36269–79 (2006).

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

93. Sonnenburg, E. D. *et al.* A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 8834–9 (2006).

94. Roberts, I. S. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu. Rev. Microbiol.* **50,** 285–315 (1996).

95. Ruas-Madiedo, P., Gueimonde, M., Arigoni, F., De Los Reyes-Gavilán, C. G. & Margolles, A. Bile affects the synthesis of exopolysaccharides by Bifidobacterium animalis. *Appl. Environ. Microbiol.* **75,** 1204–1207 (2009).

96. Dertli, E. *et al.* Structure and biosynthesis of two exopolysaccharides produced by Lactobacillus johnsonii FI9785. *J. Biol. Chem.* **288,** 31938–31951 (2013).

97. Avci, F. Y. & Kasper, D. L. How bacterial carbohydrates influence the adaptive immune system. *Annu. Rev. Immunol.* **28,** 107–130 (2010).

98. Mazmanian, S. K. & Kasper, D. L. The love–hate relationship between bacterial polysaccharides and the host immune system. *Nat. Rev. Immunol.* **6,** 849–858 (2006).

99. Onderdonk, A. B., Kasper, D. L., Cisneros, R. L. & Bartlett, J. G. The capsular polysaccharide of Bacteroides fragilis as a virulence factor: comparison of the pathogenic potential of encapsulated and unencapsulated strains. *J. Infect. Dis.* **136,** 82–89 (1977).

100. Tzianabos, A. O., Onderdonk, A. B., Rosner, B., Cisneros, R. L. & Kasper, D. L. Structural features of polysaccharides that induce intra-abdominal abscesses. *Science* **262,** 416–9 (1993).

101. Tzianabos, A. O., Kasper, D. L., Cisneros, R. L., Smith, R. S. & Onderdonk, A. B. Polysaccharide-mediated protection against abscess formation in experimental intra-abdominal sepsis. *J. Clin. Invest.* **96,** 2727–2731 (1995).

102. Mazmanian, S. K. S. S. K. *et al.* An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,** 107–118 (2005).

103. Round, J. L. *et al.* The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332,** 974–7 (2011).

104. Neff, C. P. *et al.* Diverse intestinal bacteria contain putative zwitterionic capsular polysaccharides with anti-inflammatory properties. *Cell Host Microbe* **20,** 1–13 (2016).

105. Schiavi, E. *et al.* The surface-associated exopolysaccharide of Bifidobacterium longum 35624 plays an essential role in dampening host proinflammatory responses and repressing local TH17 responses. *Appl. Environ. Microbiol.* **82,** 7185–7196 (2016).

106. Jones, S. E., Paynich, M. L., Kearns, D. B. & Knight, K. L. Protection from intestinal

inflammation by bacterial exopolysaccharides. *J. Immunol.* **192,** 4813–4820 (2014).

107. De Palencia, P. F. *et al.* Probiotic properties of the 2-substituted (1,3)-β-D-glucan-producing bacterium Pediococcus parvulus 2.6. *Appl. Environ. Microbiol.* **75,** 4887–4891 (2009).

108. Cameron, E. A. & Sperandio, V. Frenemies: Signaling and nutritional integration in pathogen-microbiota-host interactions. *Cell Host Microbe* **18,** 275–284 (2015).

109. Ferreyra, J. A. *et al.* Gut microbiota-produced succinate promotes C. difficile infection after antibiotic treatment or motility disturbance. *Cell Host Microbe* **16,** 770–7 (2014).

110. Wang, Y., Gänzle, M. G. & Schwab, C. Exopolysaccharide synthesized by Lactobacillus reuteri decreases the ability of enterotoxigenic Escherichia coli to bind to porcine erythrocytes. *Appl. Environ. Microbiol.* **76,** 4863–6 (2010).

111. Chen, X. Y., Woodward, A., Zijlstra, R. T. & Gänzlea, M. G. Exopolysaccharides synthesized by Lactobacillus reuteri protect against enterotoxigenic Escherichia coli in piglets. *Appl. Environ. Microbiol.* **80,** 5752–5760 (2014).

112. Jones, S. E. & Knight, K. L. Bacillus subtilis-mediated protection from Citrobacter rodentium-associated enteric disease requires espH and functional flagella. *Infect. Immun.* **80,** 710–719 (2012).

113. Bello, F. D., Walter, J., Hertel, C. & Hammes, W. P. In vitro study of prebiotic properties of levan-type exopolysaccharides from Lactobacilli and non-digestible carbohydrates using denaturing gradient gel electrophoresis. *Syst. Appl. Microbiol.* **24,** 232–7 (2001).

114. Ruijssenaars, H. J., Stingele, F. & Hartmans, S. Biodegradability of food-associated extracellular polysaccharides. *Curr. Microbiol.* **40,** 194–199 (2000).

115. Salazar, N., Gueimonde, M., Hernández-Barranco, A. M., Ruas-Madiedo, P. & De Los Reyes-Gavilán, C. G. Exopolysaccharides produced by intestinal Bifidobacterium strains act as fermentable substrates for human intestinal bacteria. *Appl. Environ. Microbiol.* **74,** 4737–4745 (2008).

116. Salazar, N. *et al.* Exopolysaccharides produced by Bifidobacterium longum IPLA E44 and Bifidobacterium animalis subsp. lactis IPLA R1 modify the composition and metabolic activity of human faecal microbiota in pH-controlled batch cultures. *Int. J. Food Microbiol.* **135,** 260–267 (2009).

117. Looijesteijn, P. J., Trapet, L., De Vries, E., Abee, T. & Hugenholtz, J. Physiological function of exopolysaccharides produced by Lactococcus lactis. *Int. J. Food Microbiol.* **64,** 71–80 (2001).

118. van Bueren, A. L., Saraf, A., Martens, E. C. & Dijkhuizena, L. Differential metabolism of exopolysaccharides from probiotic lactobacilli by the human gut symbiont Bacteroides thetaiotaomicron. *Appl. Environ. Microbiol.* **81,** 3973–3983 (2015).

119. Rakoff-Nahoum, S., Coyne, M. J. & Comstock, L. E. An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr Biol* **24,** 40–49 (2014).

120. Rakoff-Nahoum, S., Foster, K. R. & Comstock, L. E. The evolution of cooperation within

the gut microbiota. *Nature* **533,** 255–259 (2016).

121. Ogilvie, L. A. & Jones, B. V. The human gut virome: A multifaceted majority. *Front. Microbiol.* **6,** 1–12 (2015).

122. Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. & Gordon, J. I. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 20236–41 (2013).

123. Radke, K. L. & Siegel, E. C. Mutation preventing capsular polysaccharide synthesis in Escherichia coli K-12 and its effect on bacteriophage resistance. *J. Bacteriol.* **106,** 432–7 (1971).

124. Scholl, D., Adhya, S. & Merril, C. Escherichia coli K1's capsule is a barrier to bacteriophage T7. *Appl. Environ. Microbiol.* **71,** 4872–4 (2005).

125. Gupta, D. S. *et al.* Coliphage K5, specific for E. coli exhibiting the capsular K5 antigen. *FEMS Micro Lett.* **14,** 75–78 (1982).

126. Nimmich, W., Schmidt, G. & Krallmann-Wenzel, U. Two different Escherichia coli capsular polysaccharide depolymerases each associated with one of the coliphage φK5 and φK20. *FEMS Microbiol. Lett.* **82,** 137–141 (1991).

127. Whitfield, C. & Lam, M. Characterisation of coliphage K30, a bacteriophage specific for Escherichia coli capsular serotype K30. *FEMS Microbiol. Lett.* **37,** 351–355 (1986).

128. Bayer, M. E., Thurow, H. & Bayer, M. H. Penetration of the polysaccharide capsule of Escherichia coli (Bi161/42) by bacteriophage K29. *Virology* **94,** 95–118 (1979).

129. Scholl, D., Rogers, S., Adhya, S. & Merril, C. R. Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of Escherichia coli. *J. Virol.* **75,** 2509–2515 (2001).

130. Schwarzer, D. *et al.* A multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis. *J. Virol.* **86,** 10384–98 (2012).

131. Stummeyer, K. *et al.* Evolution of bacteriophages infecting encapsulated bacteria: lessons from Escherichia coli K1-specific phages. *Mol. Microbiol.* **60,** 1123–35 (2006).

132. Leiman, P. G. *et al.* The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. *J. Mol. Biol.* **371,** 836–849 (2007).

133. Saxelin, M. L., Nurmiaho, E. L., Korhola, M. P. & Sundman, V. Partial characterization of a new C3-type capsule-dissolving phage of Streptococcus cremoris. *Can. J. Microbiol.* **25,** 1182–7 (1979).

134. Burt, S., Meldrum, S., Woods, D. R. & Jones, D. T. Colonial variation, capsule formation, and bacteriophage resistance in Bacteroides thetaiotaomicron. *Appl. Environ. Microbiol.* **35,** 439–443 (1978).

135. Booth, S. J., Van Tassell, R. L., Johnson, J. L. & Wilkins, T. D. Bacteriophages of Bacteroides. *Rev. Infect. Dis.* **1,** 325–336 (1979).

136. Keller, R. & Traub, N. The characterization of Bacteroides fragilis bacteriophage recovered from animal sera: observations on the nature of Bacteroides phage carrier cultures. *J. Gen. Virol.* **24,** 179–89 (1974).

137. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–5 (2008).

138. Ochoa-Repáraz, J. *et al.* A polysaccharide from the human commensal Bacteroides fragilis protects against CNS demyelinating disease. *Mucosal Immunol.* **3,** 487–495 (2010).

CHAPTER 2

A Subset of Polysaccharide Capsules in the Human Symbiont *Bacteroides thetaiotaomicron* Promote Increased Competitive Fitness in the Mouse Gut

Abstract

Capsular polysaccharides (CPS) play multiple roles in protecting bacteria from host and environmental factors, and many commensal bacteria can produce multiple capsule types. To better understand the roles of different CPS in competitive intestinal colonization, we individually expressed the eight different capsules of the human gut symbiont *Bacteroides thetaiotaomicron*. Certain CPS were most advantageous *in vivo,* and increased anti-CPS IgA correlated with increased fitness of a strain expressing one particular capsule CPS5, suggesting that it promotes avoidance of adaptive immunity. A strain with the ability to switch between multiple capsules was more competitive than those expressing any single capsule except CPS5. After antibiotic perturbation, only the wild-type, capsule-switching strain remained in the gut, shifting to prominent expression of CPS5 only in mice with intact adaptive immunity. These data suggest that different capsules equip mutualistic gut bacteria with the ability to thrive in various niches, including those influenced by immune responses and antibiotic perturbations.

Introduction

The microbial community (microbiota) residing in the human intestine encodes many functions that are not represented in the human genome. Some of these functions complement

physiological deficits of humans, such as our inability to digest dietary fiber polysaccharides[1]. Other commonly observed functions serve less defined roles in shaping the membership and physiology of the microbiota. One example is the propensity of many individual human gut bacteria to synthesize a variety of different capsular polysaccharides (CPS) on their surface[2,3]. The ability to cover their cell surfaces and even vary these coatings could protect bacteria from host immune responses[4], other bacteria, bacteriophage or even allow them to alter host physiology[5,6].

Bacteroides thetaiotaomicron (*B. theta*) is a model gut symbiont that degrades a wide variety of dietary, host and microbial glycans[7–9]. The type strain of this species also dedicates a substantial portion of its genome (182 genes contained in 8 distinct loci) to CPS production, a typical phenomenon so far observed among members of the *Bacteroides* and *Parabacteroides* genera[3]. These loci may vary widely between strains of the same species[10], partly due to lateral gene transfer or other events[11,12]. The *Bacteroides* have complex mechanisms for regulating expression of CPS loci, including invertible promoters that are recombinationally flipped between "on" and "off" positions[13–15], NusG-like antitermination factors[16], and *trans* locus inhibitors[17]. It has been hypothesized that this regulation allows for one CPS locus to be expressed by an individual cell[17], while equipping cells in the greater population with the ability to explore expression of different capsules. Such a strategy could pre-adapt subpopulations of *B. theta* and related symbionts to better compete in environmental conditions for which specific CPS are most advantageous.

There are multiple roles that CPS could play in the fitness of intestinal symbiotic microbes (reviewed in ref. 6). First, bacteria may optimize cell energy expenditures by synthesizing CPS similar in composition to sugars that the bacterium is metabolizing.

Alternatively, as the capsule may be hundreds of nanometers thick[18] and acts as a barrier to acquisition of some nutrients[19], conditional expression of some CPS may facilitate passage of various polysaccharides to the cell surface. In support of this hypothesis, *B. theta* CPS loci have previously been found to be coordinately regulated with genes involved in the degradation of some host glycans[18].

In addition to roles in nutrient acquisition, CPS may benefit gut mutualists in similar ways as they do pathogens, including: evasion of the complement system and/or phagocytosis[20–23], shielding of the bacterium from detection through molecular mimicry[24], or directly modulating the immune system[5,25,26]. Indeed, each of these roles has gained support in studies of just a few gut *Bacteroides*: *B. fragilis* CPS provide complement resistance[27]; CPS in *B. fragilis* and other species display molecular mimicry of fucosylated host glycoconjugates[28]; and zwitterionic CPS from *B. fragilis* and others modulates host cytokine levels[29,30]. Finally, production of antibodies against a dominantly expressed *B. theta* CPS (CPS4) results in down-regulation of CPS4 and switching to other CPS[4], suggesting that these bacterial capsules are targeted by adaptive immunity and the ability to dynamically express alternatives is an advantage in evading this response.

Taken together, the ability to vary CPS may enable gut bacteria to evade host immune responses. However, since no collection of CPS from symbiotic gut bacteria have been systematically studied in a single species, it is unclear what general roles these prominent cellular features play, especially in the highly abundant gut-dwelling *Bacteroides*. In this study, we measured the contributions of individual CPS towards enabling *B. theta* to compete in the mammalian gut in the face of various selective pressures, including diet, changing immune responses, and antibiotic pressure. We reveal that specific CPS provide advantages in the gut

environment, and that the abundance of strains expressing individual capsules can be shaped by adaptive immunity. Moreover, we determined that the ability of *B. theta* to dynamically express multiple CPS provides an advantage over expression of any single CPS, although antibiotic perturbation is needed to reveal this competitive phenotype. This study shows the importance of capsules within the quotidian functions of a single commensal bacterium and highlights the importance of these massively diversified structures in equipping bacteria for colonization of the mammalian intestinal tract.

Results

Long-term colonization of mice with B. theta reveals dynamic expression of CPS over time

Experiments with the well-studied type strain of *Bacteroides thetaiotaomicron* (*B. theta*) have shown that it varies expression of its 8 different CPS in different *in vivo* conditions (summarized in Figure 2.2G). A limitation of these experiments is that they were short-term colonizations of gnotobiotic mice, providing only a "snapshot" of CPS locus expression at a single time point generally 10-30 days after inoculation or diet change. Longer-term colonization of mice could select for different CPS than those dominantly expressed *in vitro* or may reveal dynamic changes in locus expression. Thus, we inoculated *B. theta* into C57BL/6 (herein, referred to as wild-type) germ-free mice and measured transcriptional changes in CPS expression over a period of approximately 5 months (Figure 2.1; combined data in Figure 2.2A-C). To determine if dietary conditions[31] or adaptive immunity[4] play a role in long-term CPS expression, we also periodically fed one group of wild-type mice a low-fiber diet and colonized $Rag1^{-/-}$ mice (adaptive immune-deficient, fed a high-fiber diet).

Quantitative PCR (qPCR)-based measurements of fecal bacterial community *cps* locus transcription at regular intervals revealed that CPS expression quickly changed from the inoculum, which was enriched for *cps3* transcription (shown at day 0). However, expression

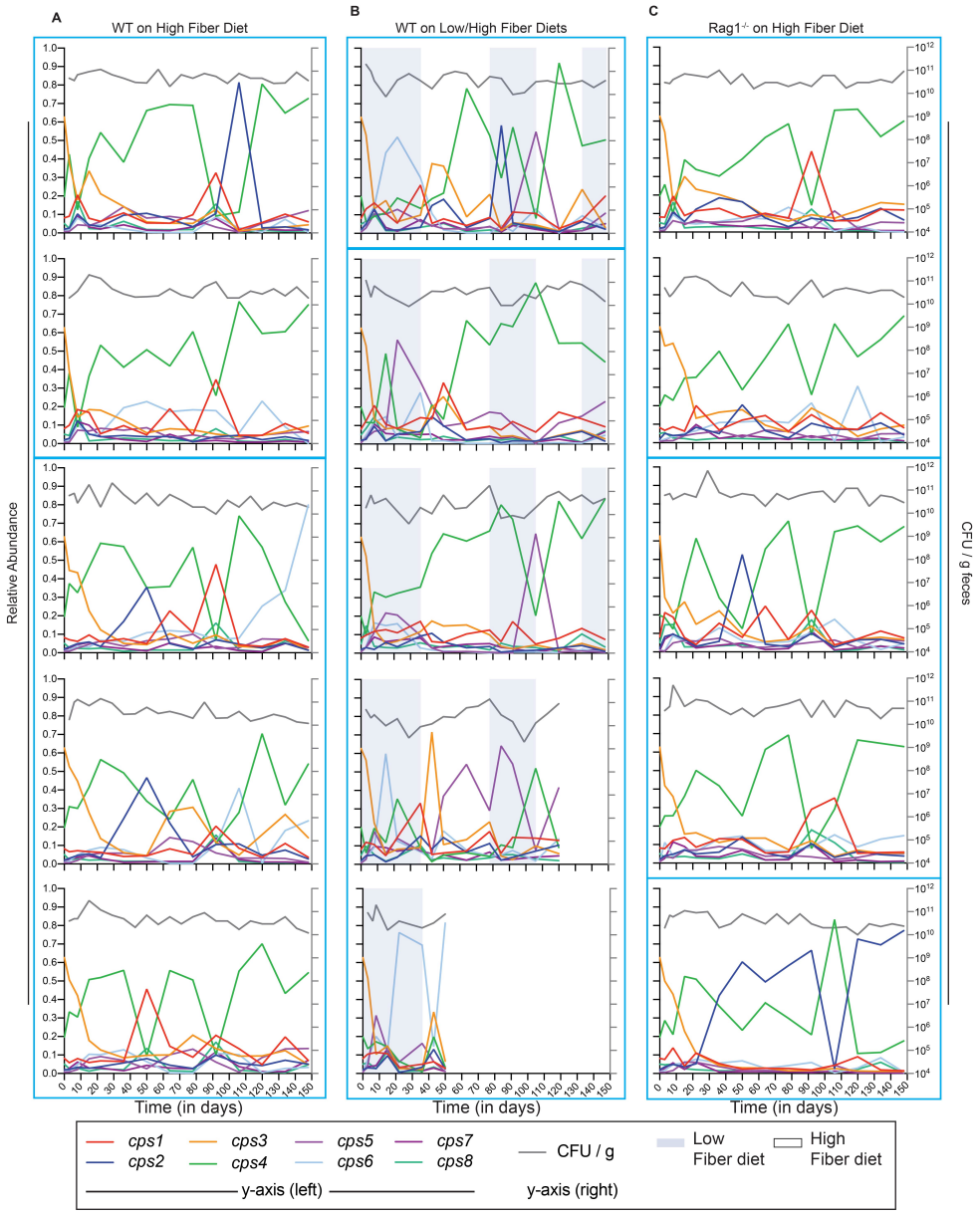


Figure 2.1 B. theta *dynamically alters expression of its* cps *loci over time in the mouse gut.*

Wild-type (WT) B. theta was inoculated into one of 3 groups (n=5 mice/group), each group in its own column and each panel representing data from an individual mouse. A) wild-type mice fed a high-fiber diet. B) wild-type mice switched between low-fiber and high-fiber diets. C) Rag1$^{-/-}$ mice fed a high-fiber diet. Shown are relative cps *locus expression in RNA extracted from stool (left y-axis) and colony forming units per gram of stool (CFU / g) (right y-axis). Blue boxes indicate groups of co-housed mice. Expression of the inoculum is represented as "Day 0". Two mice in panel B died prematurely.*

proceeded to change dynamically in individual mice over time, ultimately exhibiting a unique profile of temporal expression of different CPS for each mouse (Figure 2.1). Dynamic changes in CPS expression occurred on both high- and low-fiber diets as well as in adaptive immune-deficient mice, suggesting that these factors were not exclusively responsible for driving expression changes. As previously observed in shorter-term experiments (Figure 2.2G), expression of CPS4 was most often highest in wild-type and Rag1$^{-/-}$ mice fed a high-fiber diet. However, there were periodic decreases in CPS4 expression (up to ~62% total change) that did not always coincide with changes in total *B. theta* abundance. Additionally, some decreases spanned multiple time points, collectively suggesting that these periodic events truly reflect large-scale changes in community capsule expression. CPS1 and CPS2, and to a far lesser degree CPS6, were the capsules that were expressed antagonistically with CPS4 when mice were fed a high-fiber diet. However, in most cases in which CPS4 was dominantly expressed prior to another capsule, the community resumed expression of CPS4 within 40 or fewer days. Interestingly, bacteria in stool taken from mice on the day of sacrifice could express a different CPS than in the cecum of the same mouse (Figure 2.2D-F).

Compared to wild-type or Rag1$^{-/-}$ mice fed a high-fiber diet, those fed a low-fiber diet exhibited different CPS expression. Consistent with previous studies of *B. theta*-colonized mice consuming low fiber[31,32], 4 of 5 mice showed dominant expression of CPS5 and/or CPS6 at some time during 35 days of low-fiber feeding and this trend was never observed in the same time frame for the high fiber-fed mice. As the low-fiber diet eventually leads to illness in some *B. theta*-monoassociated animals, we switched the low-fiber diet group to the high-fiber diet at day 35. Immediately following the diet switch, CPS3 expression increased transiently in most

animals, but then gave way to a different dominant capsule (CPS4, CPS5 or CPS6). Additional

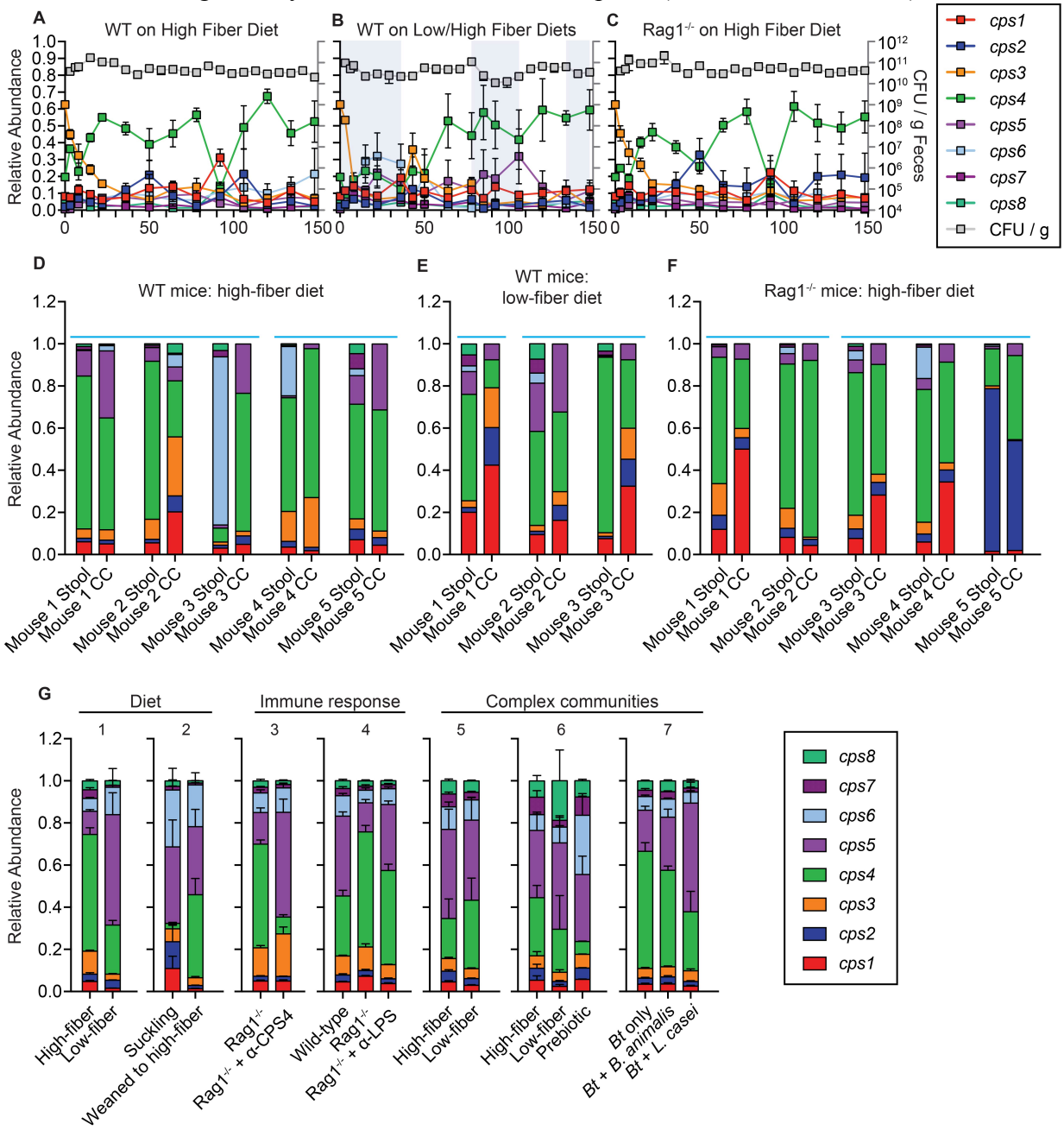


Figure 2.2 Expression of cps *loci by wild-type* B. theta *in vivo.*

A-C) Pooled relative cps *locus expression data for mice shown in Figure 2.1 (n = 5 mice/group). D-F) Relative abundance of* cps *locus expression in stool and cecal contents, both on day of sacrifice. The stool sample for each mouse is the same as the last time point (Day 148) shown in Figure 2.1. Blue lines indicate co-housed mice. A, D) Wild-type mice fed a high-fiber diet. B, E) Wild-type mice oscillated between long-term feeding of a low-fiber and high-fiber diet. Mice had been fed the low-fiber diet for 2 weeks prior to sacrifice. Two of these mice died early, likely due to infighting between cage mates; thus no cecal samples are available for these two mice. C, F) Rag1$^{-/-}$ mice fed a high-fiber diet. G) Expression of* cps *loci by wild-type B. theta in various conditions, extracted from previously published datasets (n = 3-9 mice/group; see Methods). Each panel represents a different study. Data are represented as mean ± SEM. WT: Wild-type (mice); CC: Cecal contents;* Bt*:* B. theta*;* B. animalis*:* Bifidobacterium animalis*;* L. casei*:* Lactobacillus casei.

diet switches may have triggered changes in CPS expression, although the occurrence of these is difficult to attribute to diet change alone given the different responses observed and the dynamic expression in constant feeding of the high-fiber diet. We utilized Dirichlet regression to more quantitatively compare changes in relative cps expression among the 3 experimental groups (see *Methods*). While there is great variation among mice even within experimental groups, possibly masking some transcriptional changes, expression of CPS2 was found to be significantly lower when mice were fed a low-fiber diet ($p = 0.0144$). Significant decreases in CPS4 ($p < 0.0001$) and significant increases in CPS6 ($p = 0.0002$) were also observed in low-fiber fed mice but these changes were only associated with the first period of low-fiber feeding. Taken together, these changes in CPS expression suggest that specific capsules may be advantageous *in vivo* and that the corresponding selective forces are dynamic and influenced by diet.

Competition of single CPS-expressing strains reveals differential success in the mouse gut

To better understand the role of individual CPS in the fitness of gut commensals, without the influence of the complex molecular regulation that is known to govern *Bacteroides cps* gene expression[17], we employed a series of *B. theta* variants that each encodes and expresses one of its 8 CPS biosynthetic loci (*cps* loci)[15] (see *Methods* for details on strain generation and validation). These strains allowed us to isolate the role of individual CPS in subsequent experiments. We first sought to test the hypothesis that expression of distinct CPS would be advantageous in specific immunological environments. We pooled the 8 single CPS-expressing strains, along with an acapsular strain, and inoculated these into individually housed germ-free mice of one of three genotypes: wild-type mice, mice with a deficient adaptive immune response ($Rag1^{-/-}$), and mice

with deficiencies in innate immune signaling (MyD88$^{-/-}$/Trif$^{-/-}$). All mice were fed a standard

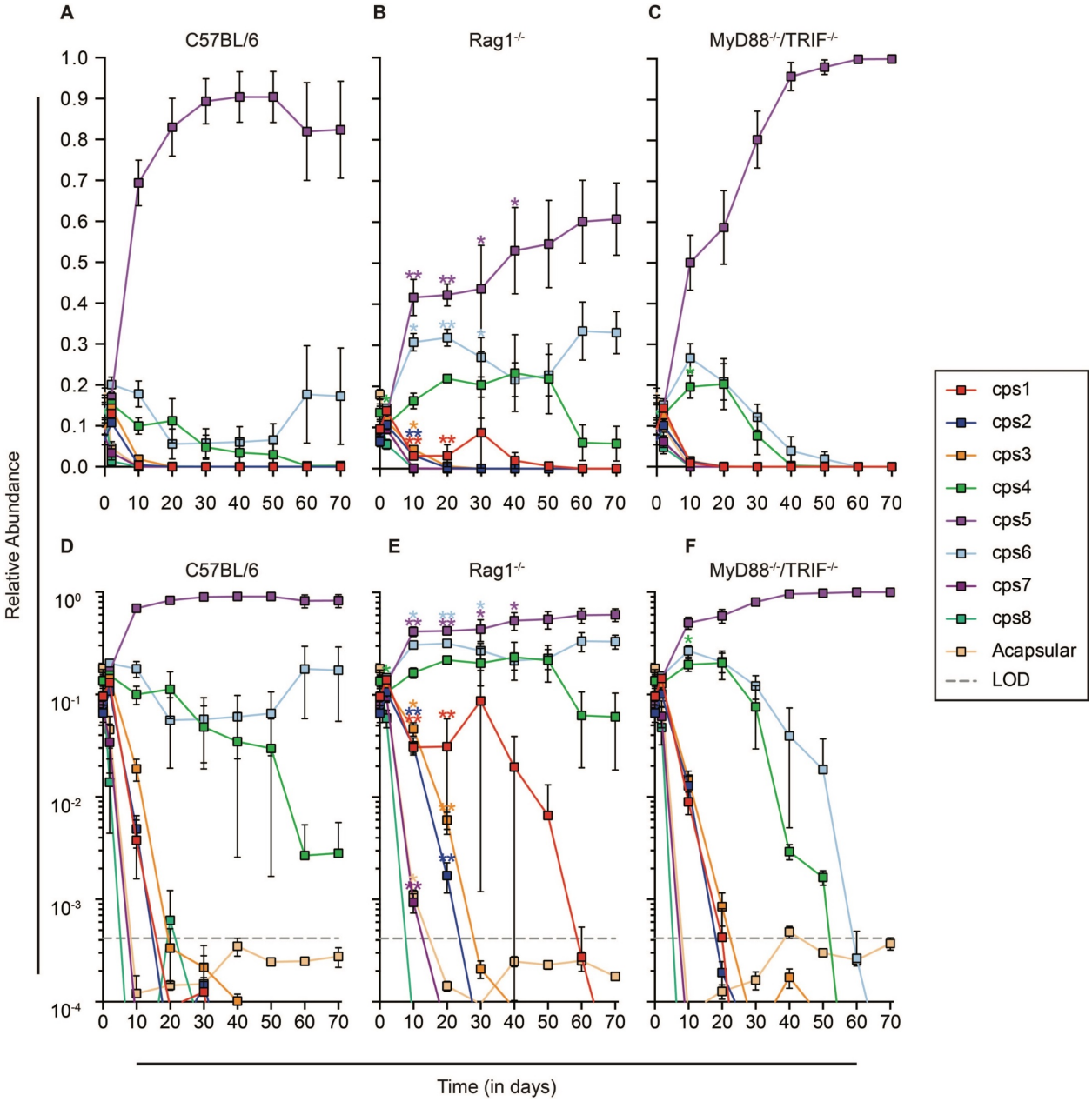


Figure 2.3 Competition of single CPS-expressing strains in germ-free mice reveals distinct advantages conferred by specific CPS.

Eight single CPS-expressing strains and an acapsular strain were pooled and inoculated into germ-free mice of 3 genotypes (n=5 mice/group): A, D) wild-type; B, E) MyD88$^{-/-}$/Trif$^{-/-}$; C, F) Rag1$^{-/-}$. Panels A-C include a linear axis for relative abundance, whereas panels D-F use a logarithmic axis to reveal strains that have fallen below the limit of detection. All mice were fed a high-fiber diet. Relative abundance of each strain in stool was determined at regular intervals via qPCR. The relative abundance in the inoculum is represented as "Day 0". Each mouse was individually caged for this experiment. Data are represented as mean ± SEM. Asterisks (in matching colors to bacterial strain) indicate significant differences ($p < .05$, ** $p < .01$; Kruskal-Wallis test, with Benjamini-Hochberg correction) in relative strain abundance between wild-type mice and the mutant mouse genotype shown in that panel at each time point. LOD: Limit of detection.*

high-fiber diet, and relative levels of each strain in stool over time were determined via qPCR.

Although all strains were present in the inoculum at relatively equal levels, 3 strains outcompeted the others within a few days after inoculation (Figure 2.3 for average strain abundance over time; Figure 2.4 for individual mouse data). The strain expressing CPS5 ("cps5 strain") predominated in the stool of 4 of the 5 wild-type mice, with the cps4 and cps6 strains found at lower levels or outcompeted by the end of the experiment. In one wild-type mouse, the cps5 and cps6 strains appeared to be codominant. However, in adaptive immune deficient ($Rag1^{-/-}$) mice, a more stable competition occurred over time that reflected both higher abundance of the cps6 and cps4 strains at most time points compared to wild-type mice and less propensity for these strains to diminish at later time points, especially cps4 in some individual mice (Figure 2.4). Specifically, while the cps5 strain was most abundant in four of the five $Rag1^{-/-}$ mice by the end of the time course, the cps4 and cps6 strains persisted in all mice and maintained higher abundance over time, with a statistically significant lower abundance of the cps5 strain over the first 40 days. Even though they were still outcompeted, several other strains (cps1, cps2, cps3, and cps7) also maintained significantly higher levels in $Rag1^{-/-}$ mice in the first 20 days. This was especially apparent in one mouse, where the cps1 strain maintained a persistent population for several weeks before dropping below the limit of detection. The less stringent dominance of the cps5 strain in $Rag1^{-/-}$ mice reinforces that adaptive immune responses play a role in CPS-mediated fitness[4] and suggests that CPS5 could be more resilient under these conditions.

Compared to $Rag1^{-/-}$ mice, an opposite trend was observed in $MyD88^{-/-}/Trif^{-/-}$ mice (deficient in innate immunity). In all 5 mice, the cps5 strain became the most dominant strain by day 10. While the cps4 and cps6 strains maintained detectable populations for several weeks in individual mice, by the end of the experiment, the cps5 strain was the only strain remaining, with

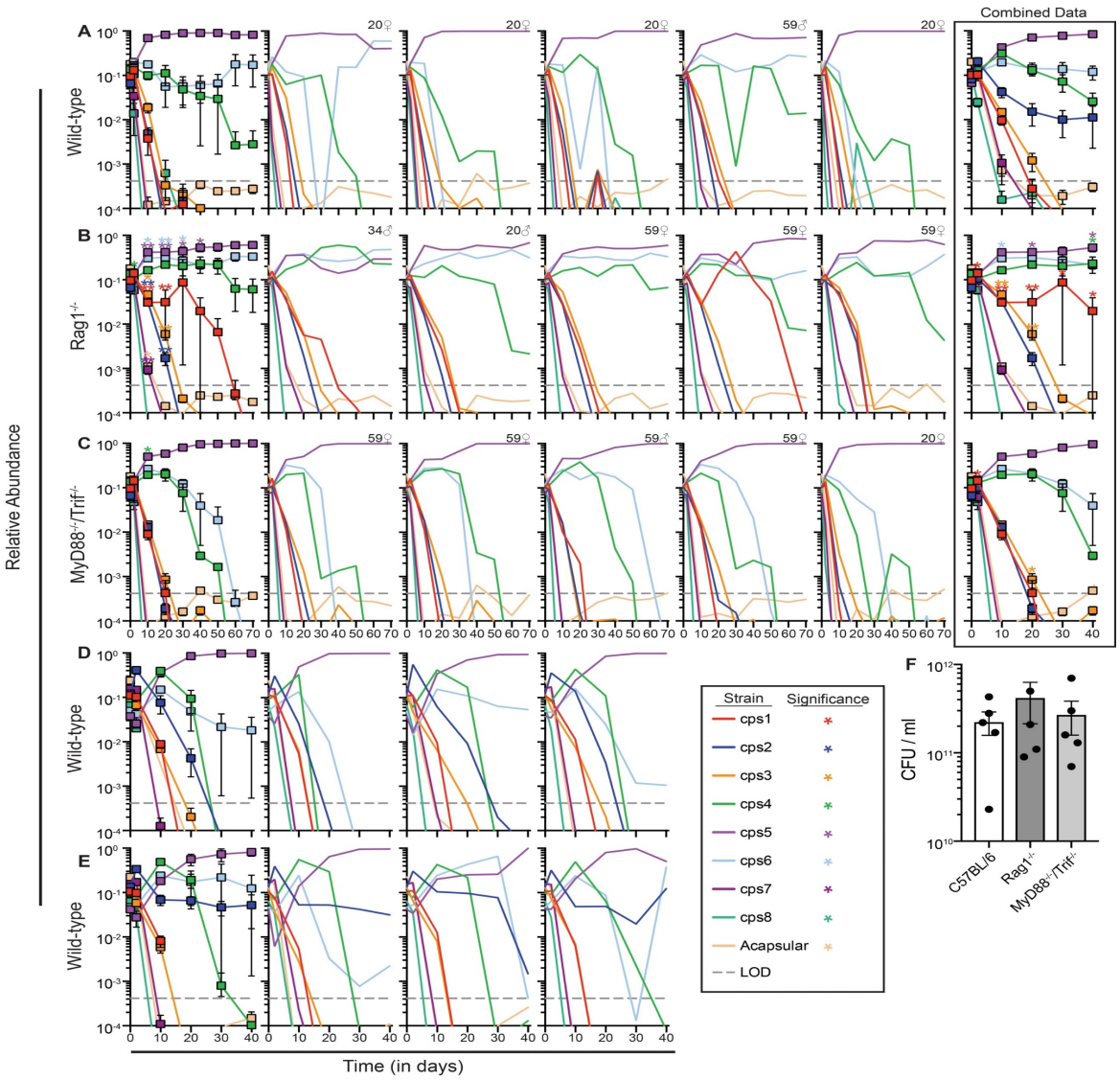


Figure 2.4 Competition of single CPS-expressing strains in vivo, data from individual mice (see also Figure 2.3).

A-C) 9 strains (8 single CPS-expressing strains and an acapsular strain) were pooled and inoculated into germ-free mice of 3 genotypes: A) C57BL/6 (wild-type, WT); B) Rag1$^{-/-}$; C) MyD88$^{-/-}$/Trif$^{-/-}$ (n = 5 mice/group). All mice were fed a high-fiber diet. Relative abundance of each strain in stool was determined at regular intervals. Each mouse in panels A-C was individually housed. Age at inoculation (in days) and gender are indicated at the top right of each panel. D-E) The experiment shown in Figure 2.3 was repeated in wild-type mice (n = 3 mice/group) with the same community of strains as before, including either: D) the original cps5 strain (mice co-housed); or E) a rederived cps5 strain (mice co-housed). The first panel in A-E represents pooled data, with subsequent panels representing individual mice. The last panel in A represents pooled data from all WT mice in Figure 2.3A, Figure 2.4D-E, and Figure 2.5A (n = 14 mice). Asterisks in panels B and C indicate significant differences (p < .05, ** p < .01) in relative strain abundance between WT (panel A) and the indicated group, using a Kruskall-Wallis test with Benjamini-Hochberg correction. F) Colony forming unit (CFU) counts in cecal contents for mice in panels A-C. Data are represented as mean ± SEM as applicable. LOD: Limit of detection.*

all other strains (including the cps6 and cps4 strains) outcompeted below the limit of detection.

Despite the differences in strain abundance, all 3 groups of mice had similar bacterial load at the

end of the experiment (Figure 2.4F). These data further suggest that CPS5 provides an advantage to *B. theta in vivo*, while deficiency in innate immunity provokes a fiercer competition between colonizing strains.

Table 2.1 Single nucleotide polymorphisms (SNPs) from each of the 10 strains competed in this study

Strain	Locus	Reference Position	Reference Base	Called Base	SNP %	Notes
Acapsular	BT1476	1819506	A	C	98.40%	Next to homopolymeric run of 7 Cs
cps6	5' BT1725	2124399	T	-	82.90%	Next to homopolymeric run of 7 Ts
Wild-type	BT2273	2854561	C	T	99.20%	Non-synonymous; V77M
cps1	BT2790	3460467	T	A	100.00%	Synonymous
Acapsular	BT2790	3460467	T	A	100.00%	Synonymous
cps6	BT3029	3834329	GGAT TGTTG	-	96.50%	3-codon deletion in putative Na+/Glucose symporter gene, uncharacterized
cps8	3' BT3570	4624960	A	G	100.00%	3' non-coding region
cps5	TAG	6142712	-	GTCG ACGC	100.00%	Extra restriction site in pNBU2-bla-tetQb barcode vector

Besides our genome sequencing data demonstrating few mutations among the strains used in this study (Table 2.1), we sought to rule out secondary mutations by repeating the competition experiment described above with a newly constructed cps5 strain. Mirroring our previous results, the rederived cps5 strain outcompeted all other strains in two of three mice, and was codominant with the cps6 strain in the last mouse (Figure 2.4E). While some variation existed among the 4 replicates of this competition in wild-type mice (Figure 2.3, Figure 2.4D-E, and Figure 2.5D), including greater persistence of the cps2 strain over the cps4 strain in some cases, combined data from wild-type mice similarly exhibited significant differences in cps5, cps6, and other strain abundances compared to $Rag1^{-/-}$ mice (Figure 2.4). We also sought to determine if there were specific global gene expression profiles shared by the most competitive strains (cps4, cps5, and cps6; "winners") compared to some of the strains that were consistently

outcompeted (cps1, cps2, and acapsular; "losers") by performing RNA-Seq on bacteria from monoassociated mice. When directly comparing all *in vivo* expression profiles for winners vs. losers (n=9 each) there were no significant changes in gene expression under the conditions used (see *Methods*), although a few strain-specific changes in mostly genes of unknown function were observed (Table 2.2). From these data we conclude that CPS4, CPS5, and CPS6 provide a direct advantage to the bacterium *in vivo* that is not reflected in strong changes in global transcription.

Table 2.2 Fold changes in bacterial genes in vivo: single CPS-expressing strains relative to the acapsular strain

Locus Tag	cps1 Fold Change	cps2 Fold Change	cps4 Fold Change	cps5 Fold Change	cps6 Fold Change
BT0004		0.05			
BT0026			39.41		
BT0029					0.14
BT0117					0.01
BT0208		0.11			
BT0222					0.16
BT0223					0.16
BT0229		0.02			
BT0234					0.14
BT0292	0.04	0.02	0.0046	0.04	0.01
BT0293	0.04	0.02	0.01		0.03
BT0294	0.04	0.02	0.01	0.06	0.01
BT0376	65.32		35.33		0.06
BT0377	13012.92				
BT0378	13789.90				
BT0379	16661.96				
BT0380	23552.50				
BT0381	13292.10				
BT0382	4352.74				
BT0383	5167.78				
BT0384	12329.75				
BT0385	5575.96				
BT0386	16167.73				
BT0387	9535.73				
BT0388	13113.55				
BT0389	7886.23				
BT0390	8234.18				
BT0391	10841.37				
BT0392	12031.00				
BT0393	9935.02				
BT0394	6109.75				
BT0395	6780.31				
BT0396	6239.93				
BT0397	7244.56				
BT0398	10684.92				
BT0399	13157.20				
BT0400	33.29		27.96		
BT0401	5.25		19.47		

BT0402			14.19		
BT0437					0.20
BT0438	0.20				0.15
BT0442	0.18				
BT0447					0.17
BT0448					0.15
BT0449	0.18				
BT0451					0.17
BT0452					0.19
BT0453	0.18				0.18
BT0462		32184.03			
BT0463		18185.34			
BT0464		12684.39			
BT0465		12147.69			
BT0466		14747.84			
BT0467		5035.72			
BT0468		5996.19			
BT0469		3498.12			
BT0470		11920.40			
BT0471		10075.51			
BT0472		11796.46			
BT0473		3717.86			
BT0474		6151.24			
BT0475		8270.42			
BT0476		34564.23			
BT0477		9179.07			
BT0478		9123.56			
BT0479		11012.08			
BT0480		8155.00			
BT0481		11770.63			
BT0482		4903.95			
BT0503					0.12
BT0507			6.18		
BT0522					0.12
BT0525					0.17
BT0614	569.71				
BT0782			103.06		
BT0801					0.01
BT0821					0.20
BT0866			5.35		
BT0867			5.15		
BT0924		0.13			0.04
BT0925		0.14			0.06
BT0926					0.14
BT0967				0.02	0.02
BT0998					0.09
BT0999					0.12
BT1028					0.14
BT1029	0.15				0.11
BT1030					0.06
BT1092	7.50		9.17		
BT1268					0.16
BT1338			3490.46		
BT1339			3903.48		
BT1340			2197.16		
BT1341			3851.72		
BT1342		10.86	9350.30		

BT1343			8025.09		
BT1344			9719.10		
BT1345			15386.09		
BT1346			5527.10		
BT1347			10903.34		
BT1348			5795.61		
BT1349			16063.07		
BT1350			2367.81		
BT1351			13093.98		
BT1352			6241.53		
BT1353			16346.82		
BT1354			4117.18		
BT1355			14538.65		
BT1356			5799.82		
BT1357			4039.19		
BT1358			13109.21		
BT1501		0.03	0.09	0.10	0.10
BT1502		0.03	0.10	0.09	0.09
BT1506		0.02			
BT1507		0.04			
BT1617					0.12
BT1640				55.10	
BT1641				17.36	
BT1642				300.54	
BT1643				9434.13	
BT1644				13338.36	
BT1645				4330.22	
BT1646				3064.43	
BT1647				4405.41	
BT1648				13074.56	
BT1649				29673.42	
BT1650				4313.39	
BT1651				13194.04	
BT1652				11856.98	
BT1653				14837.27	
BT1654				5012.69	
BT1655				85153.15	345.46
BT1656				1467.89	
BT1701		0.01			
BT1702					21.32
BT1703					6.40
BT1705					200.86
BT1706					38306.26
BT1707					65646.94
BT1708					15753.60
BT1709					9922.68
BT1710					8122.54
BT1711					10096.06
BT1712					14385.71
BT1713					14844.50
BT1714					5173.06
BT1715					9727.04
BT1716					6746.02
BT1717					9274.28
BT1718					10530.77
BT1719					13865.36
BT1720					14624.04

BT1721					10964.20
BT1722					11386.53
BT1723					15236.22
BT1724					20183.89
BT1725					828.92
BT1738					0.09
BT1786					0.06
BT1792			0.20		0.07
BT1826	0.02	0.01	0.01	0.02	0.01
BT1829					0.20
BT1864					0.17
BT1865					0.20
BT1960					0.02
BT2065					0.18
BT2106					0.19
BT2184					0.15
BT2185					0.13
BT2203					0.19
BT2311					0.19
BT2346					0.16
BT2362	5.62				
BT2409					0.15
BT2481					0.03
BT2485		0.12		0.13	
BT2486		0.07		0.11	
BT2529					0.03
BT2542					0.15
BT2583					0.05
BT2586	33.41				
BT2598			18.54		
BT2600		0.03			
BT2601		0.02	0.02		
BT2613		0.04			
BT2636	0.02				
BT2636					0.02
BT2647					0.07
BT2655			12.01	16.71	
BT2656			19.67		
BT2657			21.15	36.16	
BT2658			25.23	32.41	
BT2659			23.40	32.66	
BT2660			32.35	57.34	
BT2661			0.08	0.00	
BT2671					9.65
BT2675					0.08
BT2676					0.06
BT2677					0.10
BT2790					0.18
BT2953					0.17
BT3122					0.18
BT3129					0.18
BT3278					0.20
BT3345					0.13
BT3346					0.15
BT3347					0.12
BT3358					0.15
BT3477	0.13				0.09

BT3488					0.18
BT3489					0.18
BT3503	0.03				
BT3504					0.02
BT3505					0.02
BT3506					0.02
BT3512					0.12
BT3513	0.17				0.08
BT3525	0.14				
BT3525					0.11
BT3526	0.15				0.09
BT3679		0.18			
BT3830					0.17
BT3866					0.17
BT4013			6.48		
BT4014	8.99		13.21		
BT4016			79.98		
BT4021					0.04
BT4223				0.17	0.12
BT4225		0.12		0.08	
BT4226		0.18		0.12	
BT4227		0.11		0.04	0.06
BT4233				0.18	
BT4234		0.14			0.19
BT4266	8.72		11.74		5.79
BT4267	13.99		14.20		7.28
BT4268	13.14		14.50		8.77
BT4269	12.97		13.17		8.02
BT4270	11.86		11.12		7.88
BT4271	5.26		5.23		
BT4272	5.96				
BT4420					0.08
BT4422	0.12				
BT4423	0.17				0.06
BT4424	0.12				0.06
BT4425	0.14				0.08
BT4426	0.14				0.06
BT4427	0.16				
BT4428	0.11				0.09
BT4429	0.10		0.19		0.09
BT4430	0.18				0.15
BT4435		0.04			
BT4540					0.19
BT4594					0.14
BT4625	23.17				
BT4626	0.14				0.10
BT4735			52.56		
BT4744					0.17

Longitudinal sampling of the stool might not adequately represent geographical differences in population structure in the intestinal lumen vs. the mucus layer or along the intestinal length. Laser capture microdissection of bacteria from longitudinal (ileum, cecum, colon) or lateral (lumen, mucus layer) locations in the Rag1$^{-/-}$ mice from Figure 2.3 generally showed similar patterns among sites and stool (Figure 2.5A-B). Interestingly, the small intestine (ileum) was more permissive to strains other than cps4, cps5, and cps6 strains. A lack of geographical difference in *cps* transcript (Figure 2.5C) was confirmed in whole colon segments from mice monoassociated with wild-type *B. theta*, indicating that gross location within the intestinal tract, at least in the absence of a complex microbiota, is not a major driver of which CPS type is most advantageous.

Intestinal anti-CPS IgA levels correlate with increased competition among single CPS-expressing strains

Each of the 3 groups of mice in our competition in Figure 2.3 exhibited a different competition profile. Whereas the Rag1$^{-/-}$ mice produce no intestinal IgA, the wild-type and MyD88$^{-/-}$/Trif$^{-/-}$ mice produce adaptive immune responses. To determine whether differences in CPS-specific IgA levels between these two groups correlated with competition outcomes, we used ELISA with each purified CPS as antigen to quantify the magnitude of the CPS-specific IgA response in stool samples over time from the wild-type and MyD88$^{-/-}$/Trif$^{-/-}$ mice shown in Figure 2.3. Consistent with development of adaptive immune responses during or after the first week of colonization, levels of anti-CPS IgA in wild-type mice at day 5 were low but significantly increased by day 25 (Figure 2.6A). At this time point, IgA levels were significantly

higher against CPS1 and CPS3 than against CPS5, with significantly higher anti-CPS4 titers

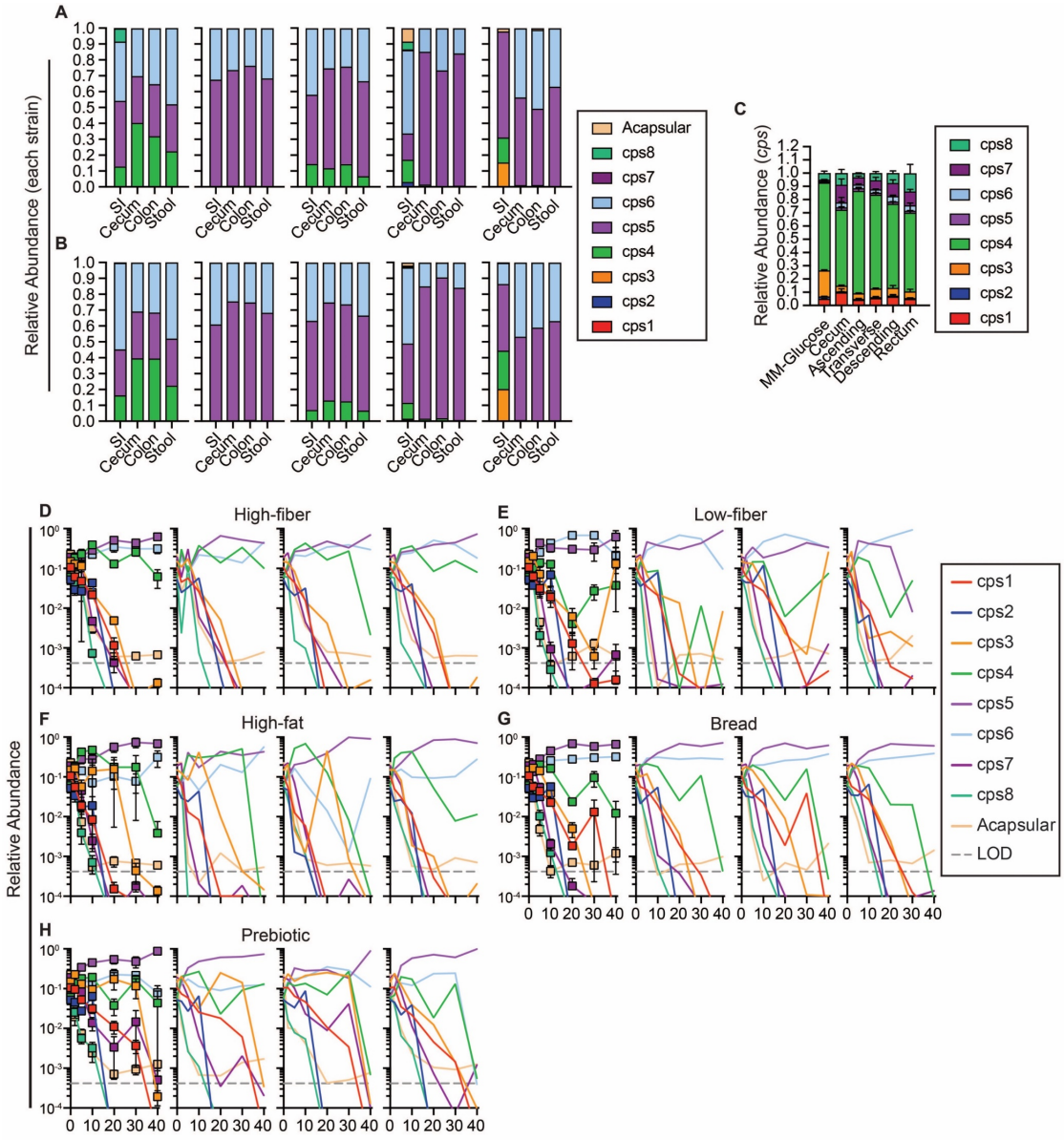


Figure 2.5 Consistency in single CPS-expressing strain relative abundance and cps locus relative expression over the length of the intestinal tract, and in vivo competition of single CPS-expressing strains with altered host diet.

Material from Rag1$^{-/-}$ mice used in Figure 3 was captured via laser capture microdissection from: A) the middle of the lumen, or B) the mucus layer. DNA isolated from this material was subjected to a non-specific amplification of the barcode vector region, followed by qPCR to the specific barcode sequence in each strain and calculation of relative abundance of each strain. Each panel represents an individual mouse, and upper and lower panels correspond to the same mouse. C) Mice (n = 3 mice) were monoassociated with B. theta for 10 days and RNA was extracted from whole intestinal segments as used in a previous study (Hickey et al., 2015). Relative cps locus expression was determined from each intestinal segment (small intestine and cecum, as well as regions of the distal colon). D-H) 9 strains (8 single CPS-expressing strains and an acapsular strain) were pooled and inoculated into wild-type germ-free mice fed one of 5 diets: D) high-fiber diet, E) low-fiber diet, F) high-fat diet, G) bread diet, or H) prebiotic diet. At regular intervals, DNA was extracted from stool for determining relative abundance of each strain via qPCR. The first panel in D-H represents pooled data, with each subsequent panel representing individual, co-housed mice (n = 3 mice/group). Pooled data are represented as mean ± SEM. SI, small intestine; MM-Glucose: minimal medium containing glucose; LOD: limit of detection.

(compared to anti-CPS5 titers) emerging by day 45. This is noteworthy given that the intestinal community consisted mainly of the cps5, cps4, and cps6 strains, illustrating that the most competitive strain in these mice (the cps5 strain) is not also the most targeted by IgA. Interestingly, over the entire experiment, levels of anti-CPS IgA were generally higher in the innate immune-deficient (MyD88$^{-/-}$/Trif$^{-/-}$) mice (Figure 2.6B), which correlated with greater competition among strains and eventual elimination of the cps4 and cps6 strains (Figure 2.3). To measure the significance of this change in anti-CPS IgA production, we used a univariate mixed model with mouse genotype as a predictor, determining that levels of IgA in the MyD88$^{-/-}$/Trif$^{-/-}$ mice were significantly higher ($p < 0.001$) than those of wild-type mice. A multivariate model including time showed no statistical difference in anti-CPS IgA levels between mouse genotypes at the first two time points, indicating that differences between mouse genotypes at days 45 ($p < 0.001$) and 65 ($p < 0.001$) drive the significance between groups—correlating with the outcompetition of both the cps4 and cps6 strains from the MyD88$^{-/-}$/Trif$^{-/-}$ mice. We also directly correlated CPS-specific IgA titers with relative strain abundance data (Figure 2.6C-D). All capsules other than CPS5 were associated with a negative or non-significant correlation coefficient, whereas there was a strong positive relationship between anti-CPS5 IgA titers and cps5 strain abundance. While this positive relationship did not reach our level of significance ($\alpha = 0.05$) in wild-type mice ($r = 0.4372$, $p = 0.0612$), it was significant in the MyD88$^{-/-}$/Trif$^{-/-}$ mice ($r = 0.6207$, $p = 0.006$). The increased IgA response in MyD88$^{-/-}$/Trif$^{-/-}$ mice was not limited to *B. theta* CPS, as ELISA using whole acapsular *B. theta* (Figure 2.6E) and quantification of total IgA

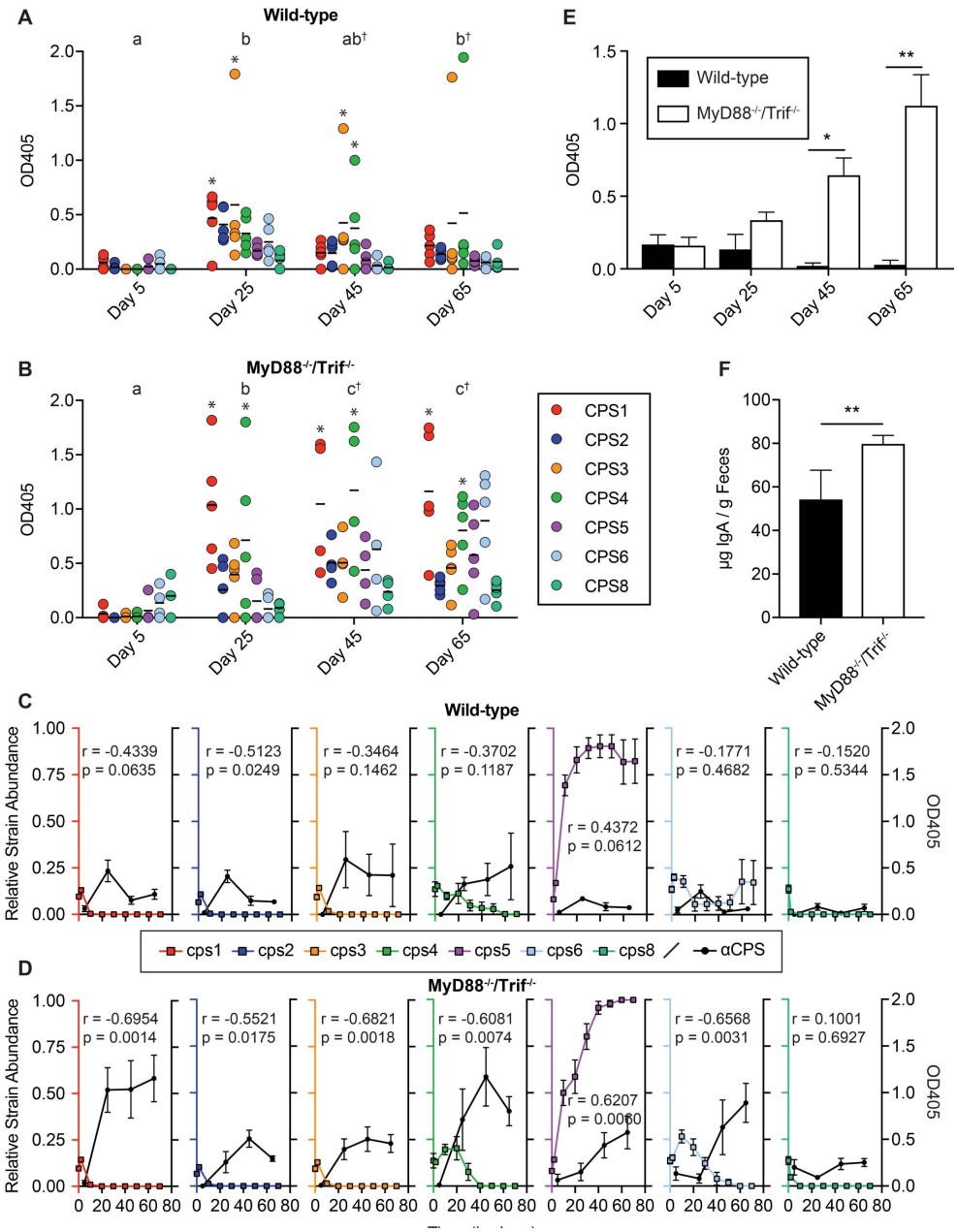


Figure 2.6 Fecal IgA levels in in vivo competition of single CPS-expressing strains.

IgA was isolated from stool from A) wild-type and B) MyD88$^{-/-}$/Trif$^{-/-}$ mice displayed in Figure 2.3 (n=5 mice/group). IgA was quantified via an ELISA using purified CPS as an antigen. Black bars indicate the mean optical density at 405 nm (OD405) against each purified CPS. Significance was determined using a series of linear mixed models. Distinct lower case letters denote significantly different time points within a host genotype ($p < .05$). Obelisks (†) denote significant differences between host genotypes at the same time point ($p < .05$). Asterisks (*) indicate samples within a time point with significantly higher OD405 values than the corresponding value for CPS5 ($p < .05$). C-D) Relative strain abundance (left y axis) and corresponding CPS-specific IgA titers (right y axis) are plotted over time for C) wild-type mice; D) MyD88$^{-/-}$/Trif$^{-/-}$ mice. Pearson correlation coefficients (r) and p values for each correlation (p) are also indicated. E) Fecal IgA to the acapsular strain from wild-type and MyD88$^{-/-}$/Trif$^{-/-}$ mice was quantified via ELISA. F) Total IgA from wild-type and MyD88$^{-/-}$/Trif$^{-/-}$ mice at day 65 was quantified via ELISA. Significant differences between mouse genotypes in panels E-F were determined at each time point via Mann-Whitney test ($p < .05$, ** $p < .01$). For panels C-F, data are represented as mean ± SEM.*

(Figure 2.6F) both showed a correspondingly stronger response in MyD88$^{-/-}$/Trif$^{-/-}$ mice. These

data implicate the adaptive immune response (particularly, IgA levels) in driving changes in relative abundance seen in the three genotypes of mice tested.

Diet alterations do not have major impact on in vivo competition or on in vitro growth rate

Previously, we determined that activation of individual glycan degradation systems induced changes in CPS locus expression[18] and others have seen changes in CPS locus expression *in vivo* upon switching mouse diet[31,32]. To investigate whether specific CPS conferred advantages in different dietary nutrient environments we used two approaches. First, we inoculated the same mixture of strains (8 single CPS-expressing strains and the acapsular strain) into wild-type mice fed one of 5 different diets (Figure 2.5D-H). While somewhat less stringent competition occurred compared to our previous experiment (Figure 2.3), the same strains (cps5, cps6, and cps4) predominated in high fiber-fed mice. While the cps6 strain dominated for most of the time in low fiber-fed mice (similar to results in Figure 2.1), on the whole we observed little difference among groups of mice fed different diets. Second, we excluded the immune response as a factor by growing each single CPS-expressing strain in minimal medium containing one of 28 different monosaccharides or polysaccharides as the sole carbon source. Overall, we observed few significant differences in growth rate, which were mostly slower growth of some single CPS-expressing strains on monosaccharides (Figure 2.7). Taken together, our *in vitro* and *in vivo* data indicate that competition for nutrients, which must permeate each surface capsule to be utilized for growth[19], is insufficient to explain the dominance of cps4, cps5 and cps6 strains in most conditions.

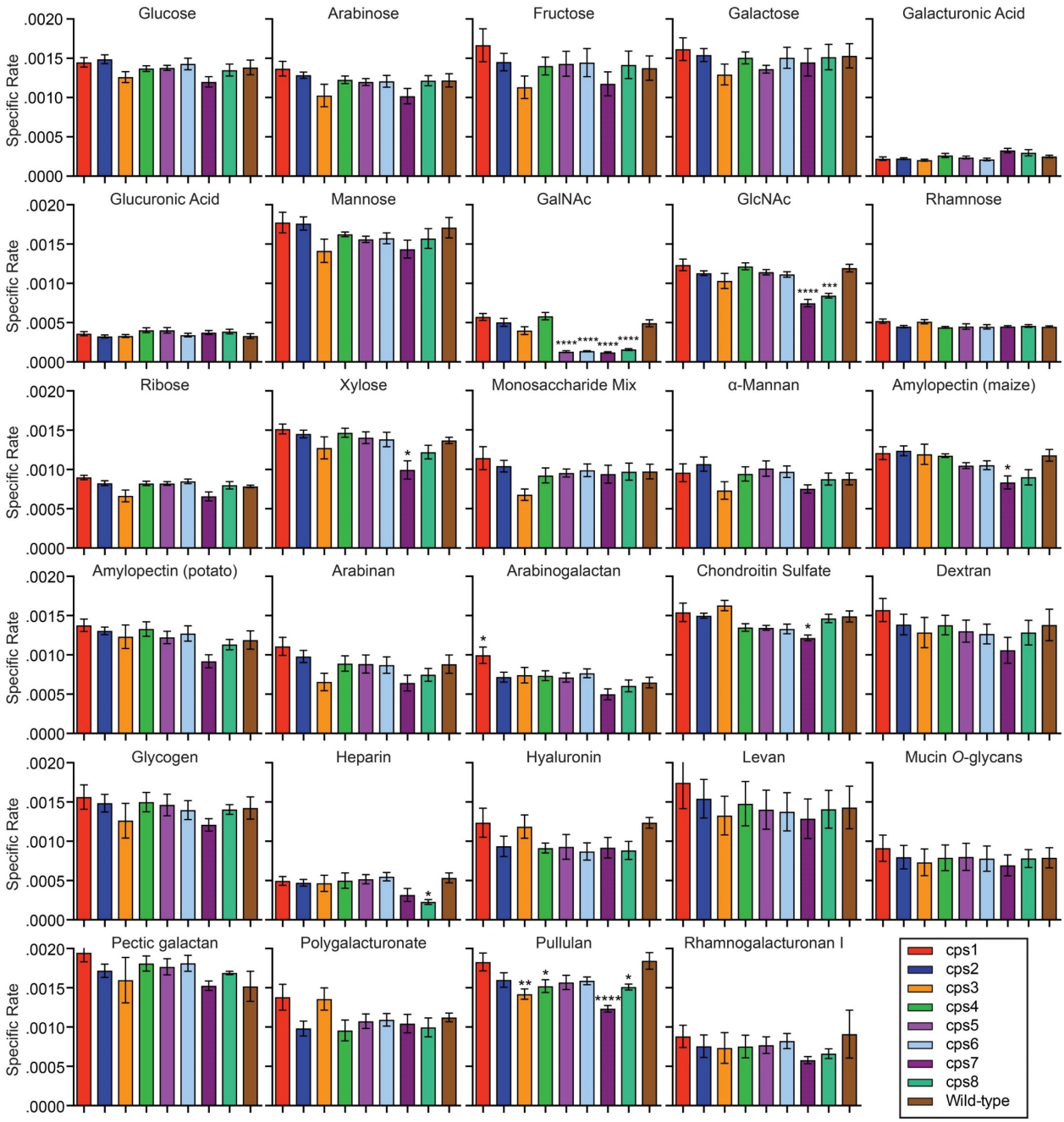


Figure 2.7 Specific growth rate of wild-type and single CPS-expressing strains in medium containing various monosaccharides or polysaccharides.

*Strains were inoculated into fresh minimal medium containing 0.5% of the indicated carbohydrate (except 1.0% for O-glycans and rhamnogalacturonan I). The monosaccharide mix contains equal concentrations of each of the individual monosaccharides at a final concentration of 0.5%. Data are represented as mean ± SEM (n = 8 independent replicates for glucose, and 3-5 independent replicates for other sugars). Significant differences in growth rate were calculated via one-way ANOVA followed by Dunnett's multiple comparisons tests to the wild-type strain with Holm-Sidak correction. * ($p < 0.05$); ** ($p < 0.01$); **** ($p < 0.0001$).*

The ability to express multiple CPS confers an additional advantage in gut competition

B. theta and its relatives have the potential to switch expression of their CPS loci to adapt to changing conditions[14], including the development of an immune response[4]. While some single CPS-expressing strains of *B. fragilis* have been competed against an isogenic wild-type strain *in vivo*, studies have yielded conflicting results. In one, a strain expressing only one of its capsules (PSH) competed equally with the wild-type for 15 days[27]. However, in another report, 3 other single CPS-expressing strains (PSA, PSB, and PSC) were outcompeted by the wild-type strain over a 7-day period[33]. Thus, it is unclear as to whether phase-variable expression of CPS loci provides an advantage over any individual CPS. Our collection allows us to test all 8 single CPS-expressing *B. theta* strains, which may compete differently with wild-type. Thus, we next tested the hypothesis that wild-type *B. theta* (WT) would exhibit increased fitness compared to the isogenic single CPS-expressing strains because it possesses the ability to dynamically adapt its CPS to thrive in potentially more diverse but unknown selective conditions in the gut. We repeated our previous *in vivo* competition by inoculating germ-free mice with the single CPS-expressing strain community, plus the acapsular strain, along with WT *B. theta* (Figure 2.8, Figure 2.9).

Similar to previous experiments, the CPS5 capsule was the most advantageous for the single CPS-expressing strains. Interestingly, the cps5 strain persisted at substantial levels alongside WT, although in wild-type mice (with adaptive immunity) WT *B. theta* competed slightly better at most later time points with the cps5 strain periodically dropping to low levels in some mice before regaining abundance. In contrast, in Rag1$^{-/-}$ mice, the cps5 strain competed better overall, with WT *B. theta* persisting at lower, dynamic levels in all but one mouse. As previously observed in Rag1$^{-/-}$ mice, the cps4 and cps6 strains were maintained at low levels, and

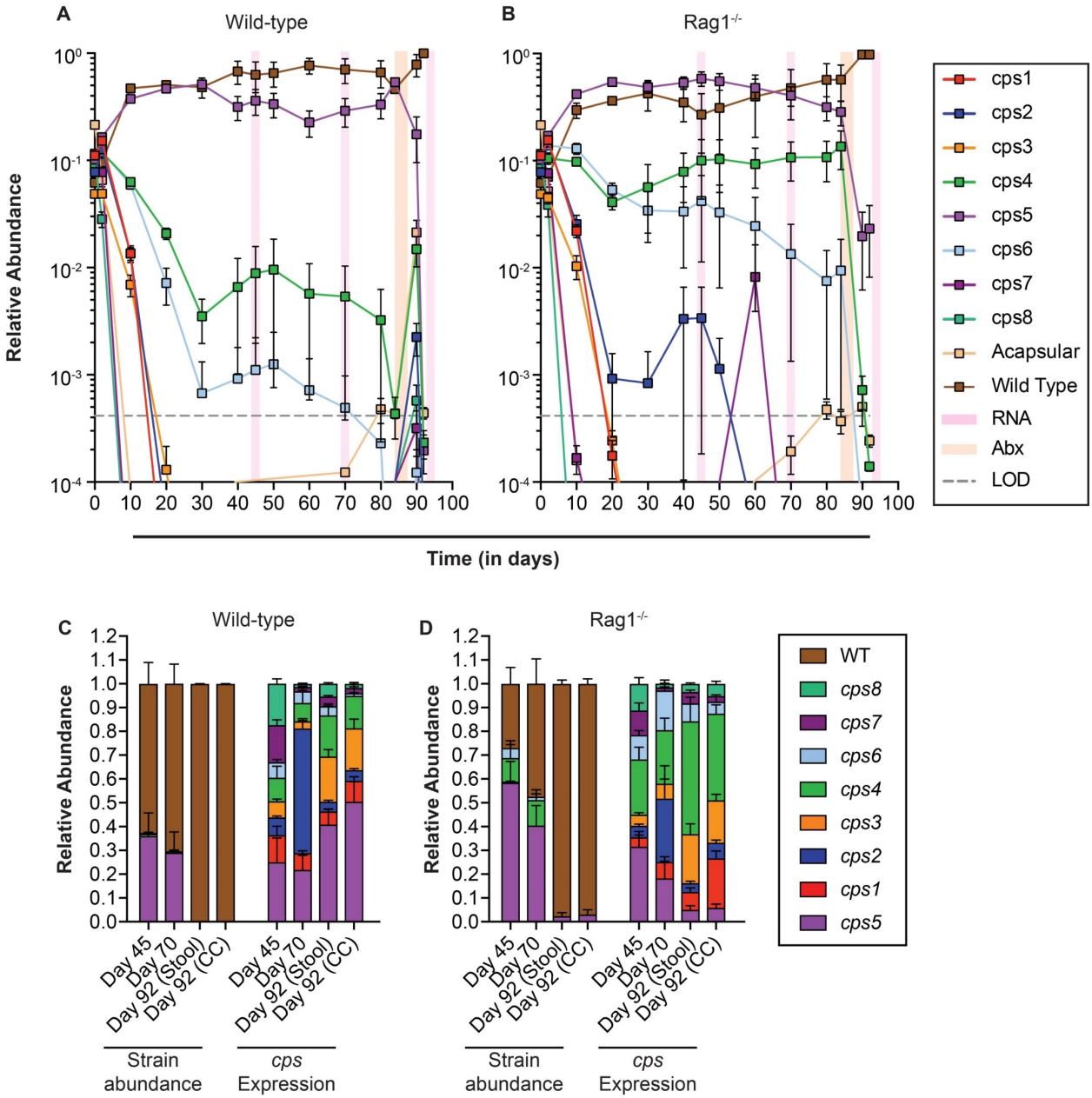


Figure 2.8 In vivo competition of the single CPS-expressing strains with the acapsular and WT strains.

10 strains (8 single CPS-expressing strains, the acapsular strain, and the WT strain) were pooled and inoculated into germ-free mice of 2 genotypes (n=5 mice/group): A) wild-type (all co-housed); B) Rag1$^{-/-}$ (in two cages). All mice were fed a high-fiber diet. Relative abundance of each strain in stool was determined via qPCR. Ciprofloxacin was administered in the drinking water of all mice from days 84-88. C-D) Relative abundance of bacterial strains and relative expression of CPS loci in C) wild-type and D) Rag1$^{-/-}$ mouse stool at days 45, 70, and 92, and from cecal contents at mouse sacrifice (day 92). Data are represented as mean ± SEM. LOD: Limit of detection.

other strains emerged at low levels in some individuals, corroborating our previous experiments (Figure 2.9B). Since WT and cps5 strains persisted together, we hypothesized that WT would

also mainly express its *cps5* locus, which generally provided the most advantage in our previous competitions. We sampled stool at two different time points to determine relative *cps* locus expression in fecal RNA. Opposite our expectations, we observed a variety of capsules expressed in the mixed community of fecal *B. theta*. As the cps5 strain can only express its *cps5* locus, expression of CPS5 was higher than that of other loci at days 45 and 70 and similar to the abundance of the cps5 strain (Figure 2.8C). At day 45, each of the remaining seven capsules was expressed above ~5% total locus transcript, presumably by WT *B. theta*. At day 70, this relatively equal proportion of capsule expression was skewed in favor of CPS2 expression, one of the CPS favored in our long-term experiment (Figure 2.1). Strikingly, this same pattern was observed in the separate group of colonized Rag1$^{-/-}$ mice (Figure 2.8D), albeit with increased presence of CPS4 and CPS6 locus transcripts, presumably due to the higher abundance of the respective single CPS strains.

Given the ability of WT *B. theta* to persist alongside the cps5 strain (and others in Rag1$^{-/-}$ mice) by largely expressing a different capsule repertoire, we hypothesized that creating a perturbation to the community would reveal a differential advantage for one of these two populations. To test this, we treated the mice with ciprofloxacin, an antibiotic that kills *Bacteroides* but does not sterilize the gut[34] and to which the cps5 and WT *B. theta* strains are equally susceptible (Table 2.3). Administering antibiotics in water for 4 days (days 84-88) to both *B. theta*-colonized wild-type and Rag1$^{-/-}$ mice reduced the level of fecal bacteria by at least 7 orders of magnitude. We then allowed the respective populations to regrow in the presence or absence of a developed adaptive immune response. In support of our hypothesis, and despite 4/10 mice having higher cps5 populations prior to antibiotic treatment, the *B. theta* populations in all mice recovered in favor of the WT *B. theta* strain. This post-antibiotic dominance of the

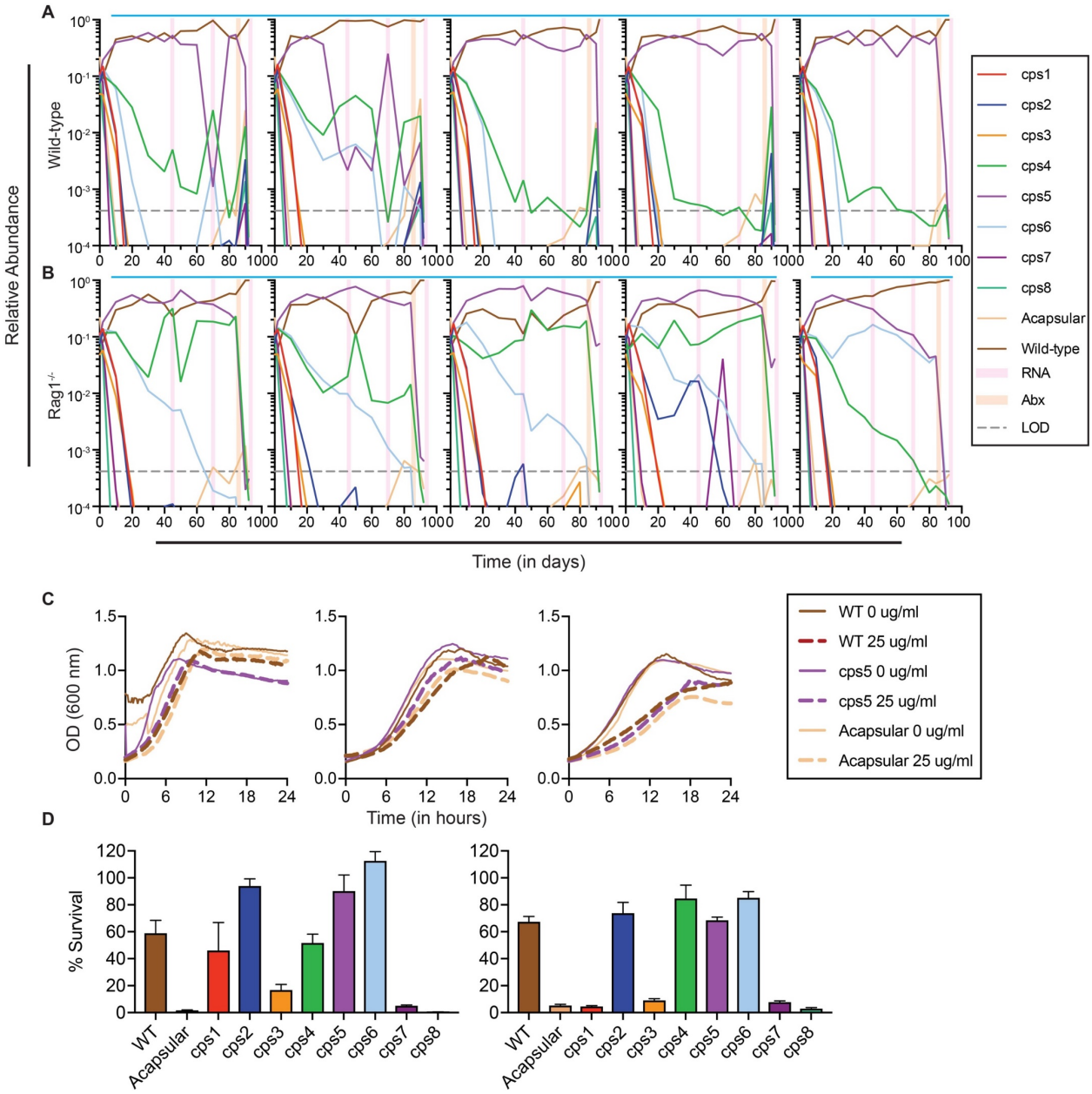


Figure 2.9 Individual mice from in vivo competition with wild-type (WT) B. theta and susceptibility to innate immune factors (see also Figure 2.8).

10 strains (WT, the 8 single CPS-expressing strains, and the acapsular strain) were pooled and inoculated into germ-free mice fed a high-fiber diet. Mice were either A) wild-type, or B) Rag1$^{-/-}$. DNA was extracted from stool for determining relative abundance of each strain via qPCR. Ciprofloxacin was administered via drinking water for the indicated time, shown by the pink bars ("Abx"). Orange bars ("RNA") indicate times at which relative cps *locus expression was also determined. Each panel represents an individual mouse in each group. Blue lines indicate co-housed mice. C) WT, cps5, and acapsular strains were grown in TYG medium in the presence (25 µg/ml) or absence of human beta-defensin 3. Each panel represents a separate experiment. D) 3 replicates of each strain (WT, acapsular, and the 8 single CPS-expressing strains) were incubated with normal human serum or freshly inactivated serum for 1 hour before plating on solid media. The percentage of serum-killed cells [(serum-treated - inactivated serum-treated) / inactivated serum-treated] was calculated. Two independent experiments are shown. Data for all experiments are represented as mean ± SEM, as applicable. LOD: Limit of detection.*

WT strain occurred in all wild-type mice, and in 3 of 5 Rag1$^{-/-}$ mice (2 mice harbored cps5 at 5-

10%), suggesting that adaptive immunity explains a small part of this phenotype. The dominance of the WT strain was unlikely to be explained by differential growth on carbohydrates (Figure 2.7) or to killing mechanisms such as antimicrobial peptides (Table 2.3, Figure 2.9C) or complement (Figure 2.9D).

Table 2.3 Minimum inhibitory concentrations of ciprofloxacin, polymyxin B, and human beta-defensin 3 (μg/ml).
Mean ± SEM displayed for each strain incubated with each antimicrobial compound. Kruskal-Wallis tests for each antimicrobial compound for each strain versus wild-type were not significant at p = 0.05. ND = Not determined.

Strain	Ciprofloxacin	Polymyxin B	Human Beta-Defensin 3
Wild-type	16 ± 0	341.3 ± 85.3	> 25
Acapsular	10.67 ± 2.67	384.0 ± 128.0	> 25
cps1	13.33 ± 2.67	298.7 ± 112.9	ND
cps2	13.33 ± 2.67	256.0 ± 0	ND
cps3	10.67 ± 2.67	384.0 ± 128	ND
cps4	10.67 ± 2.67	298.7 ± 112.9	ND
cps5	10.67 ± 2.67	341.3 ± 85.3	> 25
cps6	16 ± 0	192.0 ± 64.0	ND
cps7	13.33 ± 2.67	181.3 ± 74.7	ND
cps8	8 ± 0	192.0 ± 64.0	ND

Drastic reduction of the cps5 and other strains after antibiotics provided the opportunity to measure which CPS were expressed by WT *B. theta* after it expanded to dominate colonization. Surprisingly, in wild-type mice (with 100% WT *B. theta* in all 5 mice), WT *B. theta* displayed a new pattern of *cps* expression that varied from its pre-antibiotic profile by exhibiting far less *cps2* expression and prominent expression of *cps5*, followed by expression of *cps3* and *cps1*, plus others (Figure 2.8C). This same pattern was observed in post-antibiotic-treated Rag1$^{-/-}$ mice (Figure 2.8D), with the notable exception that far less *cps5* was expressed by the population (with most *cps5* expression found in the two mice with low populations of the cps5 strain), which instead favored corresponding increases in *cps1*, *cps4* and *cps6*. While it is possible that high *cps2* expression at Day 70 is caused by the same dynamic changes in locus expression observed in Figure 2.1, it is unlikely that the observed changes in *cps5* expression,

exclusively found in all wild-type mice, is due to a similar phenomenon, as we never observed a large (>20%) expression change of *cps5* in wild-type or Rag1$^{-/-}$ mice fed this diet. Collectively, our data indicate that the ability to shift expression of multiple CPS provides an advantage to *B. theta* that is realized most strongly after antibiotic perturbation and that CPS5 confers some advantage(s) in the face of the adaptive immune response.

Extensive CPS biosynthetic locus diversity in sequenced B. theta isolates

As *Bacteroides fragilis* strains harbor broad diversity in CPS biosynthetic loci[10], we wondered whether the specific CPS that we determined to be advantageous in gnotobiotic mice would be prevalent in other *B. theta* strains and in humans. We probed the diversity of *B. theta* CPS loci in 14 sequenced members of this species (all with >98% 16S rDNA identity to the type strain). Interestingly, only 2 of the 8 CPS loci from the type strain (*cps2* and *cps6*) are found in half or more of the strains (Figure 2.10A), raising the question of whether other strains encode different CPS loci and if they reside at the same genomic positions. Based on variations in gene content and homology, we identified a total of 49 unique *cps* loci among these 14 strains (Figure 2.10B). Strikingly, only one of these loci (*cps8* from the VPI-5482 type strain) appeared in other non-*B. theta* strains (~45,000 genomes probed), suggesting that, while diverse, *Bacteroides* capsules are largely shared among closely related lineages. The *cps8* locus is on a putative mobile element, which may explain its presence in more divergent bacteria (Figure 2.10C).

Breaks in the draft genome assemblies around *cps* loci hindered the determination of the genomic site for several *cps* loci; however, when location could be determined, homologous loci occupied the same genomic site. Identical *cps* were not limited to isolates from a single host

species, as examples from humans were also found in isolates from goats, cows, pigs and/or mice.

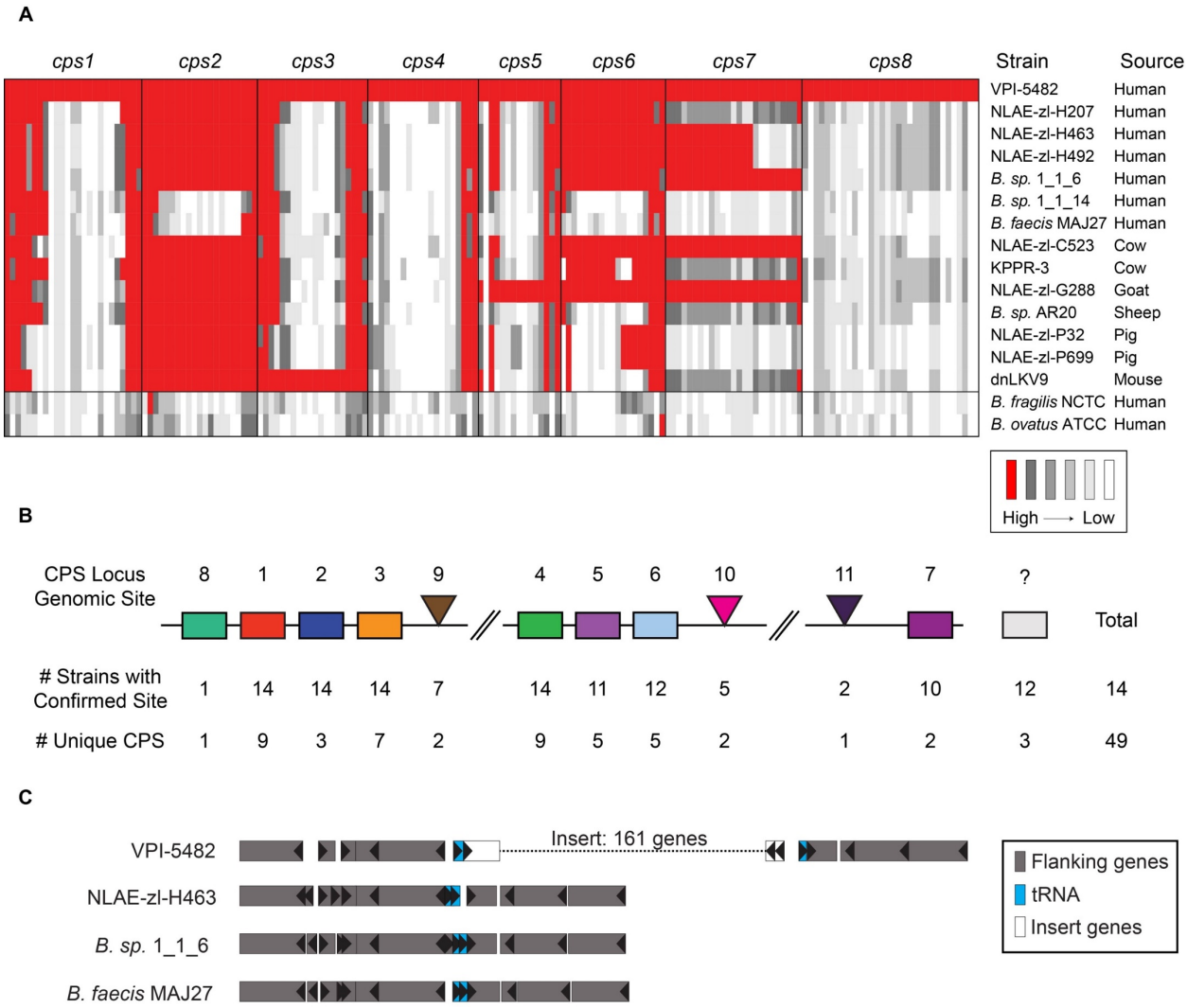


Figure 2.10 Conservation of cps *loci from B. theta VPI-5482 in other Bacteroidetes strains.*

A) The genomes of 14 B. theta *strains (>98% full-length 16S identity to type strain VPI-5482) or closely-related strains were probed with genes from the 8* cps *loci in the* B. theta *type strain.* B. fragilis *NCTC 9343 and* B. ovatus *ATCC 8483,* Bacteroides *species containing distantly related* cps *loci* [3]*, are included as controls. Each gene corresponds to a column of boxes underneath its cognate cps locus (e.g.* cps7*). Red boxes indicate homology between the gene from the type strain and the genome of the probed strain. B) A simple schematic of the* B. theta *pangenome illustrates the relative location of the 49 unique* cps *loci identified in the 14 strains. Rectangular boxes indicate genomic sites utilized by the type strain, whereas triangular wedges indicate sites found only in other strains. Because of the frequency of genomic breaks at sites of* cps *loci, a "confirmed site" indicates the presence of any number of genes involved in CPS synthesis. C) The* cps8 *biosynthetic locus is encoded on a mobile genetic element. tRNA genes flanking the* cps8 *locus in VPI-5482 were identified, and 4-5 flanking genes on each side were used to identify the same genomic location in other closely related strains. These flanking genes were homologous in all loci. The 161-gene insert in VPI-5482 includes the* cps8 *locus and putative mobilization functions.*

A more detailed inspection of a single genomic site, Site 1 (*cps1* locus in the type strain) revealed that genes flanking the *cps* locus were well conserved, as were genes for regulation (*upxY*/*upxZ*-like) and export (*wza*-like/*wzc*-like), at the 5' and 3' ends (Figure 2.11). While a few strains had gene clusters at this site that shared homologous genes and even regions of local synteny, we identified a total of 9 unique loci in which most of the genes involved in polymerization or modification of the growing glycan (e.g., glycosyltransferases, acetyltransferases, epimerases) were not homologous. Additionally, glycosyltransferase (GT) content between these loci was different, in both combination and number, suggesting that each locus encodes the ability to produce a distinct glycan.

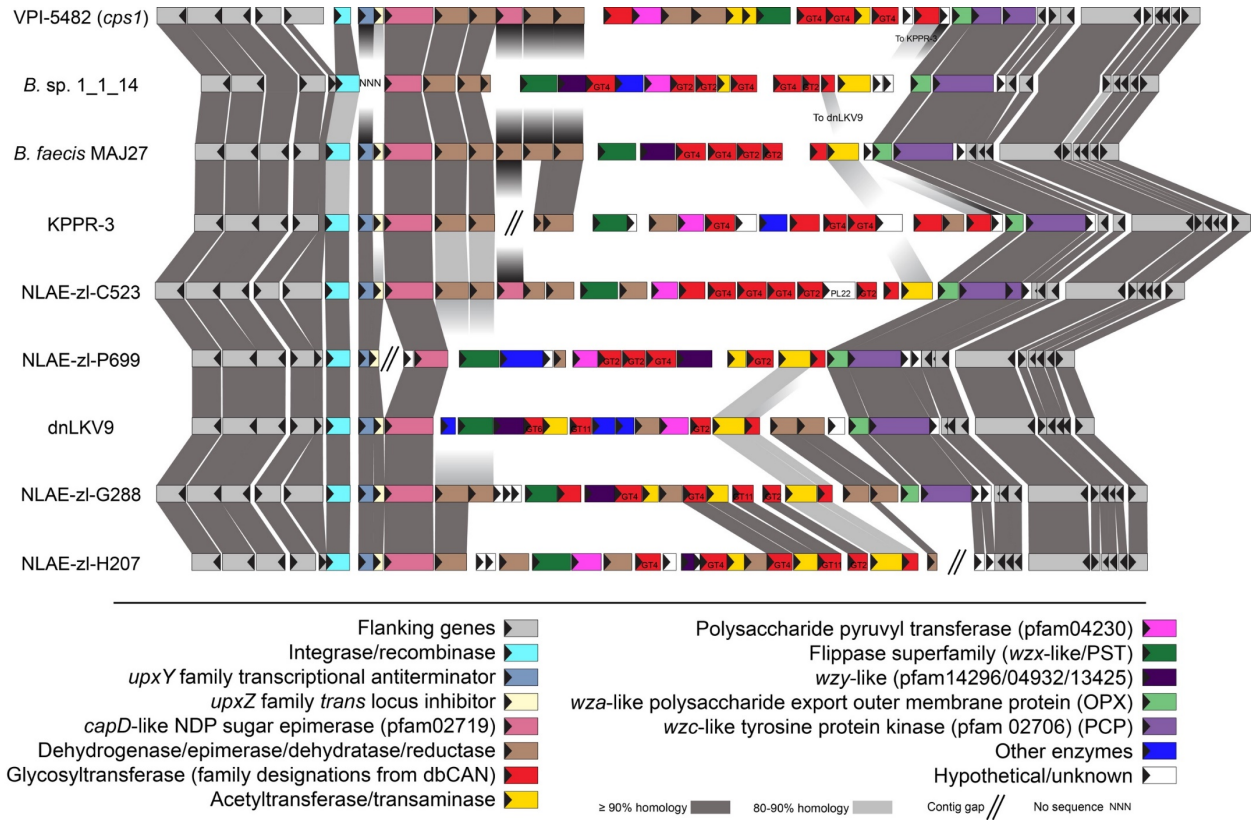


Figure 2.11 Heterogeneity of cps *loci at a single genomic site (Site 1).*

All of the genes present in 9 different cps *locus variants found at the same site as the* cps1 *locus in* B. theta *VPI-5482 were analyzed for pairwise homology to each other. Gray boxes connecting genes indicate homology (>90% identity in dark grey, 80-90% identity in light grey). Genes are color-coded according to information found in the profile for each gene on the Integrated Microbial Genomes website. Annotations from the CAZy database are displayed within each gene as applicable.*

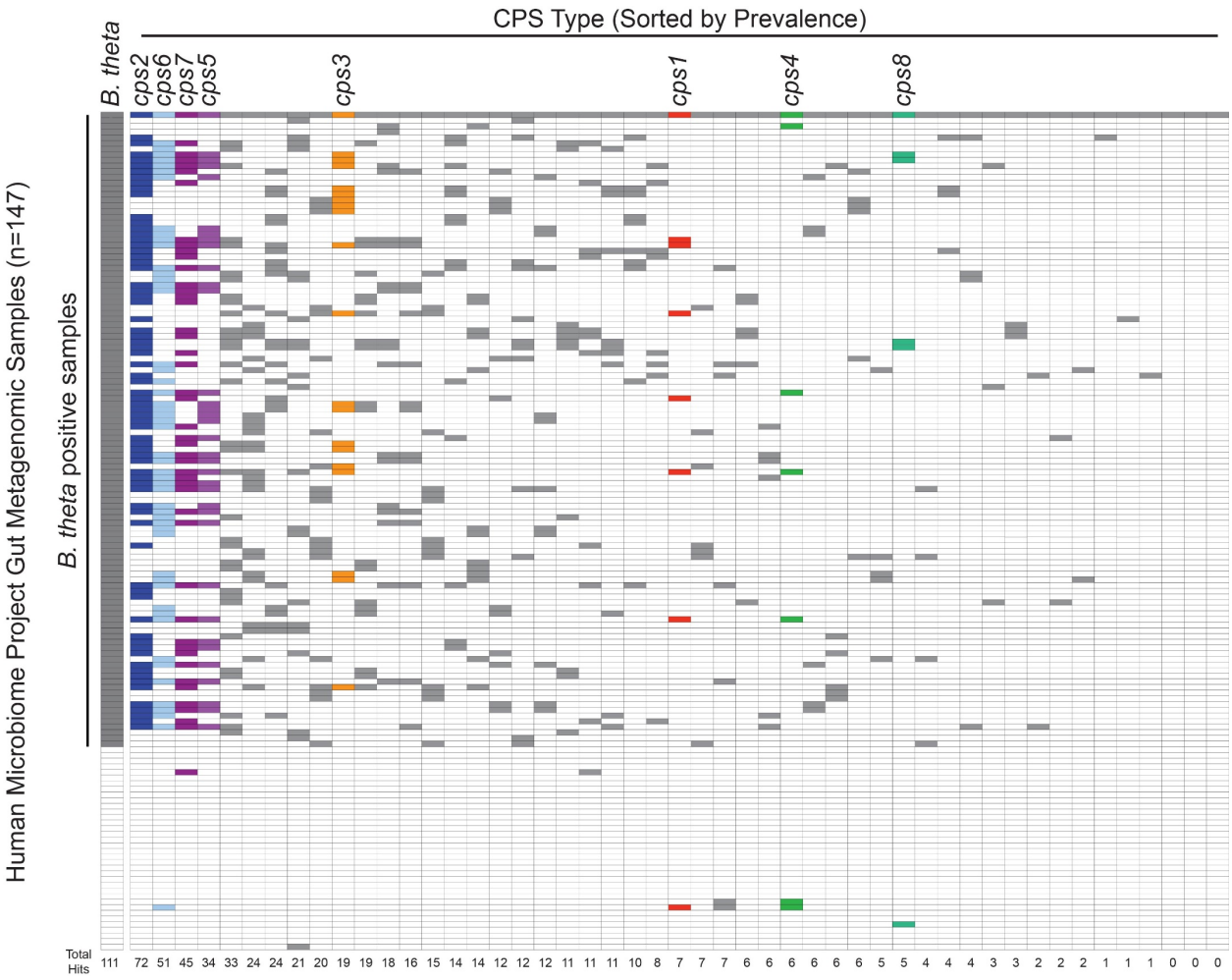


Figure 2.12 Prevalence of B. theta cps *biosynthetic loci in human metagenomic samples.*

The 147 gut metagenomic samples from the Human Microbiome Project were probed with genes from each of the 49 identified B. theta cps *loci, as well as a set of genes specific to* B. theta *sequenced isolates. Each row represents a single metagenome, whereas each column represents a unique* cps *locus. Colored boxes indicate metagenomic samples with positive hits for the corresponding locus.* cps *loci are sorted by prevalence in metagenomic samples.*

CPS5 and CPS6 biosynthetic loci are prevalent in human gut metagenomic samples

To more broadly survey the presence of *B. theta* CPS diversity we extended our analysis beyond cultured bacterial isolates to human fecal community samples. We probed metagenomic data from Human Microbiome Project stool samples[35] to determine the prevalence, in healthy adults, of the various capsule variants that we identified. Of the 48 *B. theta*-specific loci (excluding *cps8*), 45 were detected within at least one metagenomic sample (Figure 2.12). These

ranged in prevalence from approximately 1-65% of all *B. theta*-positive samples. The 3 most prevalent loci in sequenced isolates (*cps2*, *cps6* and *cps7*) were also the 3 loci most frequently identified in metagenomic samples (65, 46, and 41% prevalence in *B. theta*-positive samples, respectively), indicating that these loci may confer a competitive advantage in the human gut. Interestingly, while only found in 2 of the 14 sequenced *B. theta* isolates, the *cps5* locus was the fourth most frequently identified *cps* locus in the metagenomic samples (31% of *B. theta*-positive samples). Taken together with our experimental data, CPS5 and other CPS undoubtedly provide unique, niche-specific advantages to the bacterium during competitive colonization of the gut. At least for CPS5, this may be due in part to its ability to facilitate evasion of the host immune response in mice and, based on the prevalence of the *cps5* locus in the human metagenome, this advantage may extend to the human gut as well.

Discussion

Little has been uncovered about the specific roles that CPS play in bacterial survival in the gut and why members of this phylum encode so many variations of these surface coatings. With its full structure known and much of its immunomodulatory activity described[5,36,37], by far the best-studied *Bacteroides* CPS is the zwitterionic polysaccharide A (PSA) from *B. fragilis*, one of 8 CPS produced by the type strain of this species. Deletion of only PSA abrogated these immunomodulatory changes and protective effects, indicating that the other 7 CPS in this strain likely play different roles.

Our single CPS-expressing *B. theta* strains provide a means to test the functional roles of CPS in nutrition and other functions. Our current work fails to reveal broad support for the hypothesis that CPS expression is tightly tuned to specific polysaccharides. Instead, our findings,

combined with others' previous results, support a critical role for Bacteroidetes CPS in coping with aspects of the adaptive immune response and recovery from antibiotic challenge. Specifically, while CPS4, CPS5, and CPS6 are advantageous *in vivo* regardless of host immune proficiency, CPS4 and CPS6 are clearly less advantageous than CPS5 in the face of an adaptive immune response, especially one that is strengthened by removing MyD88/Trif (Figure 2.3, Figure 2.8A-B). However, removing host adaptive immunity neither eliminates dynamic expression of capsules by WT *B. theta in vivo* (Figure 2.1) nor equilibrates levels of competing single CPS-expressing strains, suggesting that other selective pressures are also important.

In confirmation of a previous study in *B. fragilis*[27], a strain expressing a single CPS (CPS5) co-colonizes to high levels with WT *B. theta* especially under stable conditions (e.g. consistent diet, lack of adaptive immunity). It is intriguing that, during coexistence with the cps5 strain, WT *B. theta* produces multiple CPS, which appear to mostly exclude CPS5. At least two scenarios may account for this phenomenon. First, the complex *cps* locus regulatory hierarchy employed in *B. fragilis*[14,17] likely applies to *B. theta* as well. As CPS3 and CPS4 are the most dominantly expressed in WT *B. theta in vitro* (see inoculum at "Day 0" in Figure 2.1), regulatory factors encoded in these CPS synthesis loci may actively downregulate the expression of other loci (including CPS5) in the absence of strong selective pressures. Second, having the facultative ability to shift production to different CPS may provide the bacteria with an advantage to fill some niches better than the cps5 strain. In light of these possibilities, it is even more remarkable that, although the WT strain is equally susceptible to antibiotic inhibition *in vitro*, the WT strain prevails as the dominant colonizer after antibiotic challenge in both adaptive immune proficient and deficient mice, but resorts to dominant expression of CPS5 only in wild-type mice with intact adaptive immunity. This observation supports a model in which CPS5 provides *B. theta*

with the greatest advantage to occupy a particular niche in the gut that is influenced by adaptive immunity. The "size" of this hypothetical niche must be relatively large based on the >50% abundance of the CPS5 strain in single strain competitions. After antibiotic treatment, WT *B. theta* prevails, perhaps because a capsule other than CPS5 promotes better outgrowth or access to an antibiotic-resistant niche. But after outgrowth in wild-type mice, a subpopulation of WT *B. theta* adjusts its CPS expression and fills the CPS5 niche, which may either not exist or be smaller in adaptive immune-deficient mice. It is unclear whether WT *B. theta* would continue to maintain high levels of *cps5* expression in adaptive immune-proficient mice, or whether expression would eventually decrease to levels similar to those in Rag1$^{-/-}$ mice. Further studies investigating community perturbations could provide more evidence for specific niches favoring CPS5 or other capsule types.

Such a conditionally advantageous niche exists for *B. fragilis*, which possesses a gene cluster (*ccfA-E*, for crypt colonization factor) that allows it to colonize the lumen of colonic crypts. Wild-type *B. fragilis* is more resistant to antibiotic perturbation than a *ccf* mutant unable to colonize deep within crypts[34]. Several homologous loci exist in *B. theta*, which was recently shown to also colonize colonic crypts[38], and one of these loci (BT0865-BT0867) is upregulated approximately 5-fold in the cps4 strain in mice (Table 2.2). However, the cps4 strain is also outcompeted during antibiotic perturbation, suggesting that a more complex interplay between expression of specific CPS loci and niche-specific colonization factors may be involved. Additionally, we determined that relative abundance of the single CPS-expressing strains in our previous experiment (Figure 2.3, Figure 2.5) did not vary along the length of the colon or between the mucus layer and lumen of the gut. Thus, the niches implied in our study may not refer to grossly geographical habitats but rather niches that could be based on any number or

combination of other undefined factors that are not necessarily exclusive of each other. A complex, constantly-shifting environment will likely dynamically change the size of many of these niches, necessitating phase-variable expression of fitness factors, such as capsules, to pre-adapt some part of the population to thrive in the most important niches. In experiments involving single CPS-expressing strains, CPS5 is most often dominant but this is not an exclusive phenomenon. The presence of other abundant strains that may be even more dominant in individual mice, indicates that capsules such as CPS4 and CPS6 also provide substantial advantages either by promoting access to other competing niches or by promoting good access to the same niche as CPS5.

Finally, we identify numerous, genetically diversified CPS loci in sequenced *B. theta* isolates and human gut samples. Similar broad diversity has been found in other species of gut Bacteroidetes[3,10] as well as a more phylogenetically diverse set of genomes from the human gut microbiome[2], underscoring the importance of CPS diversity to the fitness of bacteria that inhabit this niche. Given the extensive diversification of gut bacterial CPS and the currently limited knowledge of individual CPS function, we believe that future work into the roles of these CPS will provide a much greater understanding of the ways in which symbiotic bacteria interact with both their hosts and the complex and competitive microbiota in which they live.

Methods

Strains and culture conditions. Bacterial strains used in this study are listed in Table 2.4. Strains were routinely grown in TYG medium[39] (10 g/L tryptone, 5 g/L yeast extract, 4 g/L glucose, 100 mM KH_2PO_4 (pH 7.2), 15 mM NaCl, 8.5 mM $(NH_4)_2SO_4$, 4 mM L-cysteine, 200 μM L-histidine, 100 μM $MgCl_2$, 50 μM $CaCl_2$, 1.9 μM hematin, 1.4 μM $FeSO_4 \cdot 7H_2O$, 1 μg/mL vitamin K_3, and 5

ng/mL vitamin B_{12}), minimal medium[7] (100 mM KH_2PO_4 (pH 7.2), 15 mM NaCl, 8.5 mM $(NH_4)_2SO_4$, 4 mM L-cysteine, 200 μM L-histidine, 100 μM $MgCl_2$, 50 μM $CaCl_2$, 1.9 μM hematin, 1.4 μM $FeSO_4 \bullet 7H_2O$, 1 μg/mL vitamin K_3, 5 ng/mL vitamin B_{12}, and 0.5% glucose or other carbohydrate as indicated), or on brain heart infusion agar plates containing 10% horse blood (Quad Five, Rygate, Montana) (BHI-blood plates). Antibiotics were added as needed (gentamicin: 200 μg/mL; erythromycin: 25 μg/mL; 5-fluoro-2'-deoxyuridine: 200 μg/mL). Bacteria were routinely grown at 37 °C in an anaerobic chamber (Coy Laboratory Products Inc., Grass Lake, Michigan) or in borosilicate tubes using the $NaHCO_3$/pyrogallol method for anaerobiosis[39] (after placing a sterile cotton ball in the top of a borosilicate tube, burning the cotton ball and pushing it midway into the tube, 200 μL of saturated $NaHCO_3$ solution and 200 μL of 35% pyrogallol solution were pipetted onto the cotton ball, a rubber stopper was used to seal the tubes, and the tubes were incubated at 37 °C.).

Locus deletions and promoter locking were performed in the *B. thetaiotaomicron* VPI-5482 *tdk*$^-$ background as described previously[40]. *E. coli* S17-1 λpir cells (carrying the pExchange plasmid, containing genomic flanks of the locus of interest that have been ligated together) were mated on BHI-blood plates aerobically at 37 °C with an approximately equal number of *B. thetaiotaomicron* recipient cells. After one day, selection for merodiploid *B. thetaiotaomicron* cells was carried out by plating on BHI-blood plates containing gentamicin and erythromycin and incubating for 2 days anaerobically. Resistant colonies were restreaked and regrown under the same conditions for 2 days, and then 10 resistant colonies were grown anaerobically overnight in TYG medium. Counterselection was carried out by pooling the overnight cultures and plating on BHI-blood plates containing 5-fluoro-2'-deoxyuridine. After 2 days of anaerobic

growth, colonies were restreaked and regrown under the same conditions. Individual colonies were then picked and screened for loss of the locus or sequence of interest. Deletion constructs for *cps* loci were previously described[41] (see also Table 2.5), and primers for promoter locking are listed in Table 2.6. *cps3* promoter locking data in Figure 2.14C was previously shown in Hickey et al., 2015.

Table 2.4 Strains used in this study

Strain	Genotype	Features	Reference
Bacteroides thetaiotaomicron (*B. theta*) *tdk*$^-$ +Tag1	ATCC 29148 *tdk*$^-$ ::pNBU2-bla-tetQb Tag1	Parent strain of single *cps*-expressing strains with unique barcode inserted ("wild type")	Martens, Chiang, & Gordon, 2008
B. theta cps1	*tdk*$^-$ Δ*cps*2-*8 cps1*-lock ::pNBU2-bla-tetQb Tag3	*cps1*-expressing strain with invertible promoter locked in "on" orientation and unique barcode inserted	Hickey et al., 2015
B. theta cps2	*tdk*$^-$ Δ*cps1* Δ*cps*3-*8* ::pNBU2-bla-tetQb Tag11	*cps2*-expressing strain with unique barcode inserted	Hickey et al., 2015
B. theta cps3	*tdk*$^-$ Δ*cps1-2* Δ*cps4-8 cps3*-lock ::pNBU2-bla-tetQb Tag4	*cps3*-expressing strain with invertible promoter locked into "on" orientation and unique barcode inserted	Hickey et al., 2015
B. theta cps4	*tdk*$^-$ Δ*cps1-3* Δ*cps5-8* ::pNBU2-bla-tetQb Tag14	*cps4*-expressing strain with unique barcode inserted	Hickey et al., 2015
B. theta cps5	*tdk*$^-$ Δ*cps1-4* Δ*cps6-8 cps5*-lock ::pNBU2-bla-tetQb Tag6	*cps5*-expressing strain with invertible promoter locked into "on" orientation and unique barcode inserted	Hickey et al., 2015
B. theta cps6	*tdk*$^-$ Δ*cps1-5* Δ*cps7-8 cps6*-lock ::pNBU2-bla-tetQb Tag8	*cps6*-expressing strain with invertible promoter locked into "on" orientation and unique barcode inserted	Hickey et al., 2015
B. theta cps7	*tdk*$^-$ Δ*cps1-6* Δ*cps8* ::pNBU2-bla-tetQb Tag16	*cps7*-expressing strain with unique barcode inserted	Hickey et al., 2015
B. theta cps8	*tdk*$^-$ Δ*cps1-7 cps8*-lock ::pNBU2-bla-tetQb Tag9	*cps8*-expressing strain with invertible promoter locked into "on" orientation and unique barcode inserted	Hickey et al., 2015
B. theta acapsular	*tdk*$^-$ Δ*cps1-8* ::pNBU2-bla-tetQb Tag18	Strain lacking all *cps* loci and with unique barcode inserted	Hickey et al., 2015
B. theta rederived cps5	*tdk*$^-$ Δ*cps1-4* Δ*cps6-8 cps5*-lock ::pNBU2-bla-tetQb Tag21	*cps5*-expressing strain with invertible promoter locked into "on" orientation and unique barcode inserted; rederived from a *tdk*$^-$ Δ*cps1-4* strain	This study
B. theta tdk$^-$	ATCC 29148 *tdk*$^-$	Parent strain of all deletion strains; used in initial *cps* expression assay	Koropatkin et al. 2008
B. theta cps1-lock	*tdk*$^-$ Δ*cps*2-*8 cps1*-lock	*cps1*-expressing strain with invertible promoter locked in "on" orientation; used in initial *cps* expression assay	This study
B. theta cps2-only	*tdk*$^-$ Δ*cps1* Δ*cps*3-*8*	*cps2*-expressing strain; used in initial *cps* expression assay	This study
B. theta cps3-lock	*tdk*$^-$ Δ*cps1-2* Δ*cps4-8 cps3*-lock	*cps3*-expressing strain with invertible promoter locked into "on" orientation; used in initial *cps* expression assay	This study
B. theta cps4-only	*tdk*$^-$ Δ*cps1-3* Δ*cps5-8*	*cps4*-expressing strain; used in initial *cps* expression assay	This study

B. theta cps5-lock	*tdk*⁻ Δ*cps1-4* Δ*cps6-8 cps5*-lock	*cps5*-expressing strain with invertible promoter locked into "on" orientation; used in initial *cps* expression assay	This study
B. theta cps6-lock	*tdk*⁻ Δ*cps1-5* Δ*cps7-8 cps6*-lock	*cps6*-expressing strain with invertible promoter locked into "on" orientation; used in initial *cps* expression assay	This study
B. theta cps7-only	*tdk*⁻ Δ*cps1-6* Δ*cps8*	*cps7*-expressing strain; used in initial *cps* expression assay	This study
B. theta cps8-lock	*tdk*⁻ Δ*cps1-7 cps8*-lock	*cps8*-expressing strain with invertible promoter locked into "on" orientation; used in initial *cps* expression assay	This study
B. theta Δ*cps*-All	*tdk*⁻ Δ*cps1-8*	Strain lacking all *cps* loci; used in initial *cps* expression assay	Rogers et al. 2013

Table 2.5 Plasmids used in this study

Plasmid	**Features**	**Reference**
pNBU2-bla-tetQb Tag21	pNBU2-bla-tetQb vector with unique barcode sequence inserted (ATGTTCGATCATCAGTTCAGTAGCCATGGC)	Patrick Degnan, U. of Illinois and Andy Goodman, Yale University
pExchange Δ*cps6*	For deletion of the *cps6* locus to generate the rederived cps5 strain	Rogers et al. 2013
pExchange Δ*cps7*	For deletion of the *cps7* locus to generate the rederived cps5 strain	Rogers et al. 2013
pExchange Δ*cps8*	For deletion of the *cps8* locus to generate the rederived cps5 strain	Rogers et al. 2013

Table 2.6 PCR primers used in this study

Primer	**Sequence**	**Use**
cps1 lock left F	*GCGGTCGAC*GAAGCCGAATACGGTATCAACG	*cps1* promoter locking
cps1 lock left R	AGAAACAGATCCATACTGACTATC	*cps1* promoter locking
cps1 lock right F	GATAGTCAGTATGGATCTGTTTCTTTATGACC-AGCCGCAAATGAAAAGGTATAAAGCATAAA-CGAACAATGAA	*cps1* promoter locking
cps1 lock right R	*GCGTCTAGA*GCAACTAGTTCTCCAACCAGTC	*cps1* promoter locking
cps3 lock left F	*GCGGTCGAC*CGGTATCAACGTATATGCGTATG	*cps3* promoter locking
cps3 lock left R	ACACCAGCATCAGTGCTGATTAT	*cps3* promoter locking
cps3 lock right F	ATAATCAGCACTGATGCTGGTGTCTCCATAA-AGGCGCATACCGACTAAAAGTACACTTCATA-ATAAATCGGG	*cps3* promoter locking
cps3 lock right R	*GCGTCTAGA*CATCTACATTGACAAGCTCACCC	*cps3* promoter locking
cps5 lock left F	*GCGGTCGAC*TGCTTCGGAGTATCTATAATCGG	*cps5* promoter locking
cps5 lock left R	CACACCGATAAACACTGTATATCA	*cps5* promoter locking
cps5 lock right F	TGATATACAGTGTTTATCGGTGTGGATTAC-GGCGTGATAGATTGGTGTAACAAACATAG-GGAAAGCTAAACA	*cps5* promoter locking
cps5 lock right R	*GCGTCTAGA*CATTAACCAGTTCGCCAACAAG	*cps5* promoter locking
cps6 lock left F	*GCGGTCGAC*GATTTCCACGTATATGCGTATGC	*cps6* promoter locking
cps6 lock left R	CACACCGATAAACACTGTATATCA	*cps6* promoter locking
cps6 lock right F	TGATATACAGTGTTTATCGGTGTGCTTCCTG-CCCGGCTAAAAACACGAAACAAGCATAGGG-CAAGCAGAAA	*cps6* promoter locking
cps6 lock right R	*GCGTCTAGA*GCCAAAGGAGAACTGTTCATGC	*cps6* promoter locking
cps8 lock left F	*GCGGTCGAC*GCGTTTCCCTTTCAAACATGTC	*cps8* promoter locking

cps8 lock left R	CGACAATTTTTACTCAACATTATTG	*cps8* promoter locking
cps8 lock right F	CAATAATGTTGAGTAAAAATTGTCGTCAAAT-CCGGGGACTGGGCTTAGAGTGATACATGTGA-ATAAAACAATCT	*cps8* promoter locking
cps8 lock right R	*GCGTCTAGA*TCCGGTACTGCAATCCTTCAC	*cps8* promoter locking
Universal outer F	GTGGACAACAAGCCAGGGATG	Non-specific amplification of barcode vector
Universal outer R	CCTACATTGACATAGCGGATGA	Non-specific amplification of barcode vector
cps1 internal F	CCTTATATACCTAATTTCCATTGG	Determination of orientation of *cps1* locus promoter
cps1 promoter OFF	GGCGATCACTCCATTGATGG	Determination of orientation of *cps1* locus promoter
cps1 promoter ON	TTGAGATGATCAGTGAGTCG	Determination of orientation of *cps1* locus promoter
cps3 internal F	TTCATGCTTAGGAATTTATATGCC	Determination of orientation of *cps3* locus promoter
cps3 promoter OFF	CCCCACCTTGTTCAAATACTCG	Determination of orientation of *cps3* locus promoter
cps3 promoter ON	GCCAAGTTCCGTGGAGTTCC	Determination of orientation of *cps3* locus promoter
cps5 external F	GAGTGCTATCAAAGGGTTGTTG	Determination of orientation of *cps5* locus promoter
cps5 promoter ON	CAAATGCAGGCATAGCTGATGA	Determination of orientation of *cps5* locus promoter
cps5 promoter OFF	TCATCAGCTATGCCTGCATTTG	Determination of orientation of *cps5* locus promoter
cps6 external F	TTTAAAAGAAAGCAATAGCGAACG	Determination of orientation of *cps6* locus promoter
cps6 promoter ON	GGAAAATGACAATATCCGTGTTAT	Determination of orientation of *cps6* locus promoter
cps6 promoter OFF	ATAACACGGATATTGTCATTTTCC	Determination of orientation of *cps6* locus promoter
cps8 internal F	TCCTTGAAACTGCCAAACAGG	Determination of orientation of *cps8* locus promoter
cps8 promoter OFF	TTGCCCCCCATGGCGTTAGG	Determination of orientation of *cps8* locus promoter
cps8 promoter ON	CAAGTTGATTTTGAAATCCTTGC	Determination of orientation of *cps8* locus promoter
NBU2 att1 F	CCTTTGCACCGCTTTCAACG	Determination of pNBU2 insertion site
NBU2 att1 R	TCAACTAAACATGAGATACTAGC	Determination of pNBU2 insertion site
NBU2 att2 F	TATCCTATTCTTTAGAGCGCAC	Determination of pNBU2 insertion site
NBU2 att2 R	GGTGTACCTGGCATTGAAGG	Determination of pNBU2 insertion site

Table 2.7 qPCR primers used in this study

Primer	**Sequence**	**Use**
BT0396 F	GCGTGGTGGTGGCATACTCTTCT	*cps1* locus expression
BT0396 R	ACTTACGCCTGCCACCAATGTTAG	*cps1* locus expression
BT0463 F	CTATTTCGGTATTGATGTTGGCTGGTA	*cps2* locus expression
BT0463 R	TCTCCGATAATAAATGCTTGGGCTAAT	*cps2* locus expression
BT0612 F	CCATAAAGGGGCTGAGCATTGTTC	*cps3* locus expression

BT0612 R	AGTCATCATAGGTAGCGGCTTGTAGTG	*cps3* locus expression
BT1339 F	CGGTTAGGAGGAGTTTCGTTTTTC	*cps4* locus expression
BT1339 R	CAAGATCATCCGCAATACCTGTTA	*cps4* locus expression
BT1646 F	TGGGAGCTATGGGCAAAAGTAT	*cps5* locus expression
BT1646 R	TTCCGCCAAAGTAACAGACCTC	*cps5* locus expression
BT1714 F	AATTCTTTTGGGCGTACTTATGGTGA	*cps6* locus expression
BT1714 R	CAACTTTTAGCCTATCCGGTGTGAAC	*cps6* locus expression
BT2862 F	GGGCAGCCGATATTCTTGA	*cps7* locus expression
BT2862 R	ACGCCGTTATTCGCTCTTTT	*cps7* locus expression
BT0040 F	GTCGGCGGAATGCTGATGATGAACT	*cps8* locus expression
BT0040 R	CGCCTGACGGAAAGACGAACCA	*cps8* locus expression
Tag1	ATGTCGCCAATTGTCACTTTCTCA	Quantification of barcoded wild-type strain
Tag3	TTATGACCAGCCGCAAATGAAAAG	Quantification of barcoded cps1 strain
Tag11	ATGCCGCGGATTTATTGGAAGAAG	Quantification of barcoded cps2 strain
Tag4	CTCCATAAAGGCGCATACCGACTA	Quantification of barcoded cps3 strain
Tag14	GGCACGCCATTCTTCATCTAACTG	Quantification of barcoded cps4 strain
Tag6	GATTACGGCGTGATAGATTGGTGT	Quantification of barcoded cps5 strain
Tag8	CTTCCTGCCCGGCTAAAAACACGA	Quantification of barcoded cps6 strain
Tag16	GTTTATCCGTAACTAGCAGTAACC	Quantification of barcoded cps7 strain
Tag9	TCAAATCCGGGGACTGGGCTTAGA	Quantification of barcoded cps8 strain
Tag18	TTGAATTTTACCTTCCCAGTGTCC	Quantification of barcoded acapsular strain
Tag21	ATGTTCGATCATCAGTTCAGTAGC	Quantification of barcoded rederived cps5 strain
Universal Tag R	CACAATATGAGCAACAAGGAATCC	Reverse primer for all barcode qPCR primers

Gnotobiotic mouse experiments. All animal experiments were performed with approval from the University Committee on Use and Care of Animals at the University of Michigan. Mice were randomly assigned into groups by a technician unfamiliar with the project. Animal numbers for each experiment were chosen based on the minimum numbers required in previous studies to observe significant changes in the readouts measured in each experiment (gene expression, community member changes, antibody titers, etc.). Germ-free mice of a C57BL/6 background (wild-type, MyD88$^{-/-}$/TRIF$^{-/-}$, or Rag1$^{-/-}$, as indicated in each experiment) of varying ages and genders were used, dependent on availability, and none of these were involved in any previous

experiments. Mixed gender groups with a wide age range (20-59 days old at inoculation, see Figure 2.4A-C) were used in the experiment for Figure 2.3, whereas all other experiments involved a single gender common to all groups within that experiment. Groups within an experiment were age matched to the greatest extent possible. For Figure 2.1, male mice were inoculated at 50-84 days of age. For Figure 2.4D-E, female mice age 35-47 days were used. For Figure 2.7, female mice age 64-80 days were used. For Figure 2.8, male mice 27-36 days of age were used. For the RNA-Seq experiment performed for Table 2.2, female mice 45-56 days of age were used.

Mice were inoculated with 100 µL of bacterial culture (individual strain for monocolonization experiments; or for competition experiments, strains pooled in equal volumes at approximately OD_{600} 1.0, with the acapsular strain volume doubled). Mice were routinely fed a standard, autoclaved high-fiber diet (LabDiet 5013, LabDiet). For experiments involving other diets [low-fiber diet TD.130343 (Harlan Laboratories, Madison, WI)[42], prebiotic diet (TD.130342, Harlan)[43], bread-based diet (TD.130682, Harlan)[9], or high-fat diet (TD.96132, Harlan)[44]], mice were pre-fed the corresponding diet for 7 days prior to inoculation with *Bacteroides thetaiotaomicron* (*B. theta*) strains. To reduce bacterial load *in vivo* in Figure 2.8, mice were switched to water containing 0.625 mg/mL ciprofloxacin for 4 days, prior to switching back to regular water. Reduction of bacterial load was confirmed via spot plating of stool samples.

Resequencing of single CPS-expressing strains. For strain genomic sequencing, pellets from 35 mL overnight cultures were resuspended in 15 mL TE and incubated with 0.5 mL 20% SDS and 200 µL of 20 mg/mL proteinase K at approximately 50 °C for 2 hours. Samples were extracted

twice with 1:1 phenol:chloroform, followed by two extractions with pure chloroform. DNA was precipitated using sodium acetate and isopropanol and centrifuged to form a pellet. After washing in 70% ethanol and drying, the pellet was resuspended in 5 mL TE overnight at 4 °C. Samples were treated with RNase prior to extractions with phenol, 1:1 phenol:chloroform, and two extractions with chloroform. DNA was precipitated with sodium acetate and ethanol, washed in 70% ethanol, dried, and resuspended in TE. 700 µL of 100 ng/µL DNA was fragmented in a Sonics Vibra-cell GE505 ultrasonicator at 25% amplitude for 2 minutes (30 seconds on, 30 seconds off). DNA libraries were prepared using the TruSeq DNA LT Sample Prep Kit (Illumina Inc., San Diego, California). Libraries were pooled and run in 1 lane of 50 base pair single read sequencing on an Illumina HiSeq 2500 at Michigan State University. Genomes were assembled in SeqMan NGen (DNASTAR Inc., Madison, Wisconsin) and SNPs against the sequenced wild type strain (*tdk*$^-$ + Tag1) were identified in SeqMan Pro (DNASTAR Inc.) (see Table 2.1). CPS locus deletions were also validated from sequencing assemblies.

RNA isolation and quantitative PCR (qPCR) for relative cps expression. For *in vitro* samples, strains were streaked onto agar media and grown anaerobically for 2 days. 3 individual colonies were selected for each strain and inoculated into TYG medium overnight, then were subcultured 1:50 into minimal medium containing 0.5% glucose and grown using the $NaHCO_3$/pyrogallol method. Upon reaching OD_{600} 0.6-0.8, the supernatant was removed, and 1.5 mL of each culture was resuspended in 500 µL RNAprotect (Qiagen) and then handled according to the manufacturer's instructions. RNA was extracted via the RNeasy Mini Kit, treated with the TURBO DNA-free Kit (Ambion) and purified again with the RNeasy Mini Kit.

RNA was extracted from flash frozen mouse stool or cecal contents via bead beating[45] for

5 minutes in 500 µL buffer (200 mM NaCl, 200 mM Tris, 20 mM EDTA at pH 8) with 210 µL of 20% SDS and 500 µL phenol:chloroform:isoamyl alcohol (125:24:1, pH 4.3, Fisher Scientific), followed by ethanol precipitation, purification using the RNeasy Mini Kit, treatment with TURBO DNA-free Kit, and repurification with the RNeasy Mini Kit. For measuring relative *cps* locus expression from inocula, the inoculum was passaged into minimal medium and handled as for other *in vitro* samples above. cDNA was prepared with up to 1 µg RNA using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions.

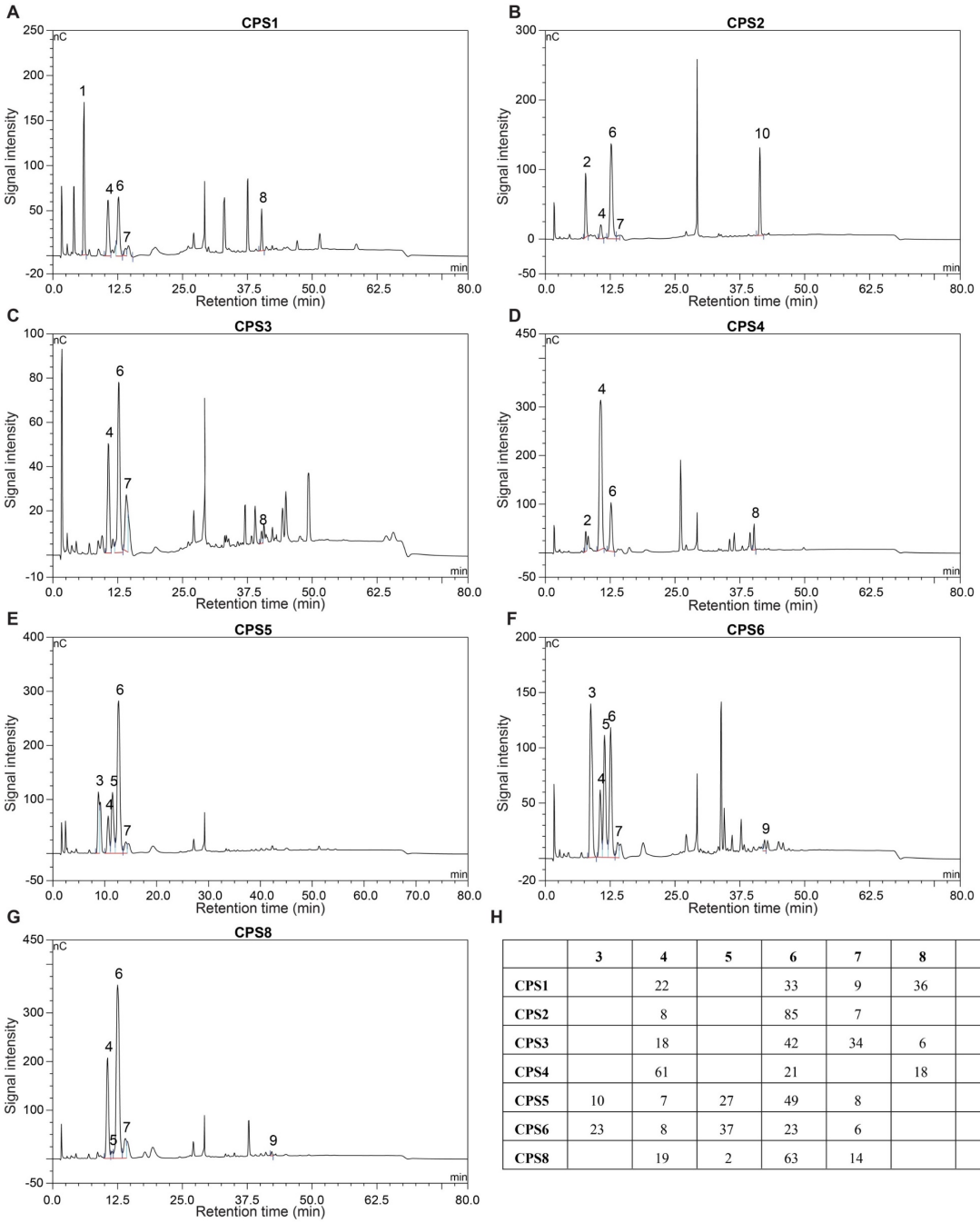
Quantitative PCR (qPCR). qPCR analyses were performed using a Mastercycler ep realplex instrument (Eppendorf) and KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems, Wilmington, MA) with 500 nM primers and 10 ng DNA/cDNA as appropriate. Primers are listed in Table 2.7. For quantification of barcoded strains, a touchdown protocol was used: 40 cycles of 95 °C for 3 seconds, annealing at a variable temperature for 20 seconds, and 72 °C for 8 seconds. For the first 6 cycles annealing temperature started at 68 °C and then dropped one °C each for the subsequent 5 cycles. The annealing temperature was 62 °C for the remaining 34 cycles. Samples were normalized to a DNA standard curve for each strain. For *cps* locus gene expression, qPCR was performed for 40 cycles of 95 °C for 3 seconds, 55 °C for 20 seconds, and 72 °C for 8 seconds. Samples were normalized to a DNA standard curve of DNA from the wild-type strain. Both protocols were followed by a melting curve analysis for determination of amplicon purity. Relative abundance of strains/*cps* locus genes was then calculated.

Preparation of purified CPS and monosaccharide analysis. Purified CPS were prepared via the hot water phenol method[18]. Briefly, 10 mL overnight cultures of each strain grown in TYG medium were inoculated into 4 L of minimal medium containing 0.5% glucose as the sole

carbohydrate source for 2 days. Stationary phase cultures were centrifuged, flash frozen in liquid nitrogen, and stored at -80 °C until use. Cells were resuspended in 100 mL hot water prior to the addition of an equal volume of phenol. The mixture was stirred for 1 hour at 65 °C and allowed to cool at 4 °C overnight. Preparations were then centrifuged and the aqueous phase was dialyzed exhaustively against deionized distilled water (12-14 kDa cutoff). Preps were lyophilized to dryness and resuspended in buffer containing 20 mM Tris HCl (pH 7.4), 2.5 mM $MgCl_2$, and 0.5 mM $CaCl_2$. Both RNase A and DNase I were added and samples were rotated gently overnight at 37 °C. Proteinase K was then added and samples were incubated overnight at 65 °C. Phenol was added to the sample prior to vortexing and centrifuging for phase separation, and the aqueous phase was removed. Samples were again dialyzed exhaustively, prior to lyophilization to dryness. The cps7 strain did not produce sufficient product for downstream analysis, likely because of lower *cps7* transcription and thus lower CPS production. Glycogen contamination in the samples was assayed using a quantitative iodine assay[46]: 50 μL of 5 mg/mL sample was added to a 96-well microtiter plate, and 50 μL of iodine reagent (5 mM I_2 + 5 mM KI) was added. Color development was measured at 580 nm and compared to a standard curve of glycogen to determine glycogen levels in each sample. Samples displayed variable amounts of bacterial glycogen contamination (Figure 2.13H), which is also evident in relative abundance of glucose in purified CPS preparations (Figure 2.13A-G).

For monosaccharide composition analysis, purified CPS were analyzed via high pH anion exchange chromatography using pulsed amperometric detection (HPAEC-PAD) following acid hydrolysis at the UCSD Glycotechnology Core, using a uronic acid profile and authentic sugar standards (glucose, galactose, mannose, fucose, N-acetyl-glucuronic acid, N-acetyl-galacturonic

acid, glucuronic acid, galacturonic acid, iduronic acid, guluronic acid, and mannuronic acid).



H

	3	4	5	6	7	8	9	Gly
CPS1		22		33	9	36		5
CPS2		8		85	7			0
CPS3		18		42	34	6		0
CPS4		61		21		18		2
CPS5	10	7	27	49	8			24
CPS6	23	8	37	23	6		3	7
CPS8		19	2	63	14		1	21

Figure 2.13 Monosaccharide composition of purified CPS.

HPAEC-PAD traces for purified preparations of A) CPS1, B) CPS2, C) CPS3, D) CPS4, E) CPS5, F) CPS6, and G) CPS8. Due to low yield, CPS7 was not analyzed. H) Summary of relative percentage of quantified monosaccharides, and percentage of glycogen ("Gly") contamination in each sample. Numbers on peaks (panels A-G) and numbers on column headings (panel H) correspond with the following detected monosaccharides: 1) N-acetyl-fucosamine, 2) Rhamnose, 3) N-acetyl-galactosamine, 4) N-acetyl-glucosamine, 5) Galactose, 6) Glucose, 7) Mannose, 8) Galacturonic acid, 9) Glucuronic acid, 10) unknown.

Single CPS-expressing strains were confirmed to have different monosaccharide profiles (see Figure 2.13).

Single CPS-expressing strain characterization and validation. A set of *B. theta* strains that each encodes one of the type strain's 8 CPS biosynthetic loci were previously generated[15]. These were produced by sequential deletion of all but one of the type strain's 8 *cps* loci in parallel mutant lines (Figure 2.14A). Invertible promoters that flip between "on" and "off" orientations[13] precede 5 of the 8 *cps* loci. To eliminate this as a confounding variable, the "outside" inverted terminal repeat region (i.e., lying upstream of the expressed promoter in its "on" state) of these 5 promoters was deleted to lock them in the expressed orientation (Figure 2.14B-C). To enable identification of each strain in a mixture, a unique 24 bp barcode sequence was inserted into each of these 10 strains at one of the two NBU2 ($tRNA^{ser}$) sites[40]. As the panel of CPS-expressing strains is crucial for many of the experiments in this work, each strain was extensively validated. Genomic sequencing of each strain identified few mutations compared to the other strains (Table 2.1), and expected deletions, promoter locking, and barcode vector integration were confirmed in genome assemblies. Furthermore, for experiments in Figure 2.4E, the cps5 strain was rederived from an intermediate strain harboring just 4 *cps* loci (*cps5*, *cps6*, *cps7*, and *cps8*), with the reasoning that any secondary mutations found in this ancestral strain would be found in multiple strains, including other strains that competed both dominantly and poorly *in vivo*. Expression of individual *cps* loci in the 8 single CPS-expressing strains was confirmed by qPCR. As expected, each of the 8 single *cps*-encoding strains expressed only its respective *cps*, with all but *cps7* being transcribed to similarly high levels. In contrast, a population of the wild-type (tdk^-) parent strain was able to express all of the loci (Figure 2.14D). Unlike other strains possessing CPS

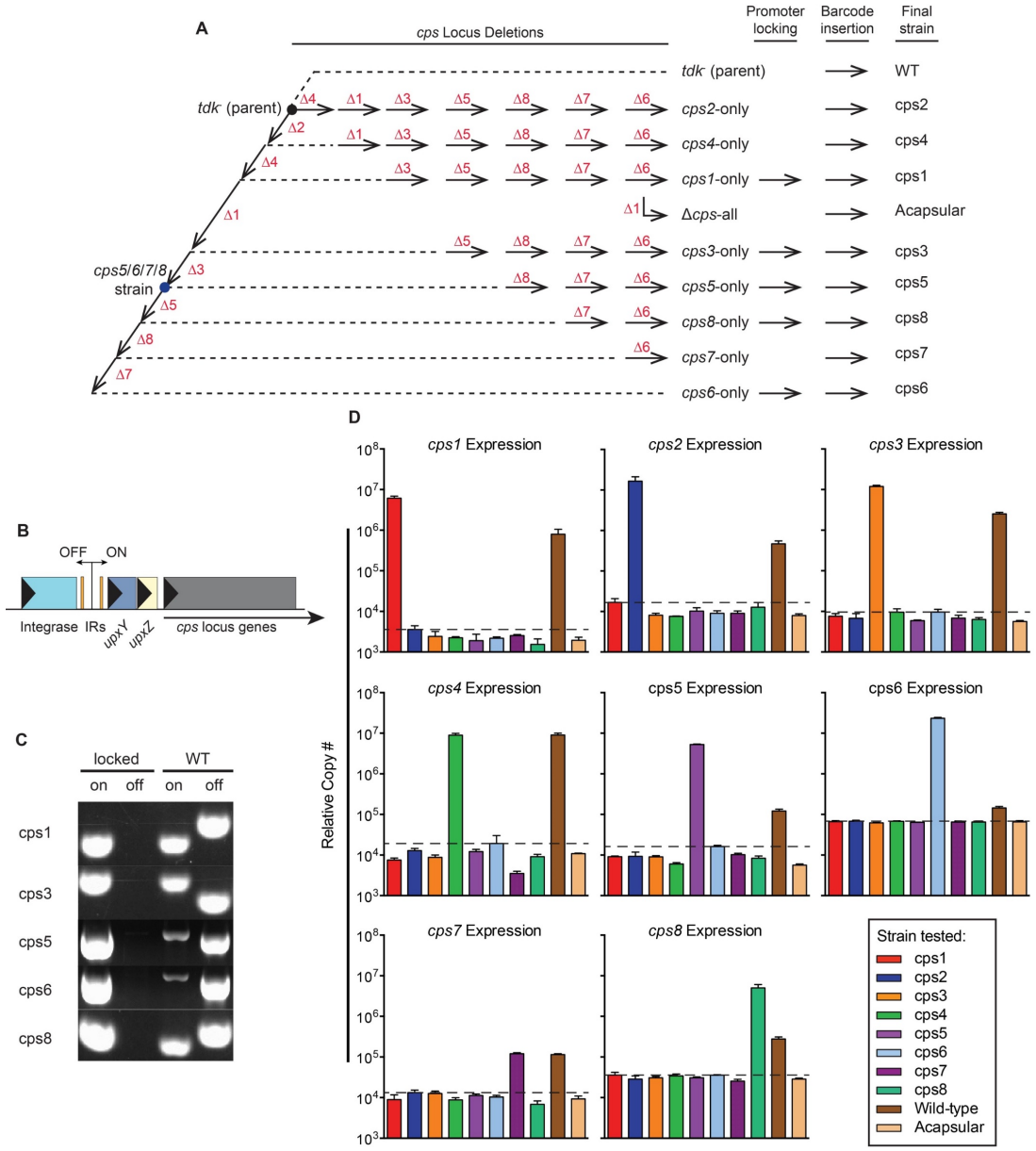


Figure 2.14 Generation and validation of the single CPS-expressing strains utilized in this study.

A) Schematic of the genetic manipulation required to generate each of the 10 strains used in this study. For each of the single CPS-expressing strains, 7 of the 8 cps *biosynthetic loci were sequentially deleted, and all 8 loci were deleted to yield the acapsular strain. Invertible promoters (5/8 loci) were locked as needed. A barcode vector was inserted into each strain for identification in a mixed population. The CPS5-expressing strain was rederived from the cps5/6/7/8-encoding strain (blue dot) and this rederived strain was tested in Figure 2.4E only. B) General schematic of the promoter region of phase-variable* cps *loci (*cps1, cps3, cps5, cps6, *and* cps8*). Inverted terminal repeats (IRs) flanking the promoter region facilitate promoter inversion between "on" and "off" states. C) Phase-variable promoters are locked into the "on" orientation. Genomic DNA from the 5 single CPS-expressing strains containing invertible promoters was used in a PCR to determine promoter orientation after locking the promoter as described above, with the parent strain (tdk-, WT) employed as a promoter-inverting control. The forward primer for each strain is found within the phase variable region and a reverse primer is found downstream and within the locus ("on"), or upstream of the locus ("off").* cps3 *locus data was shown previously*[15]*. D) Each of the barcoded single CPS-expressing strains expresses a single* cps *locus. Each strain was grown in triplicate to mid-log, and qRT-PCR for a single gene in each locus was performed. Data are represented as mean ± SEM.*

biosynthetic genes, the CPS7-expressing strain ("cps7 strain") that exhibited low expression sank

to the bottom of the tube when grown in static liquid culture, similar to the acapsular strain of *B. theta.* This phenomenon has also been reported for acapsular *B. fragilis*[27] and, together with low transcription, suggests that CPS7 is not fully functional. To measure differences in the chemical composition of the capsules, the CPS was extracted from each single CPS-expressing strain and quantitative sugar analysis was performed as described above. When grown in a defined medium containing glucose as the sole carbohydrate source, each strain produced CPS composed of different sugars or different stoichiometric abundances (Figure 2.13, note that CPS7 was not assayed due to low abundance of extractable CPS). These data indicate that each of these single CPS-expressing strains only produces a single CPS, allowing isolation of the role of individual CPS in experiments in this work.

Relative cps *expression from previous published datasets.* Microarray or RNA-Seq datasets measuring gene expression in samples containing *B. thetaiotaomicron* VPI-5482 were selected from the Gene Expression Omnibus website (http://www.ncbi.nlm.nih.gov/geo/; see Table 2.8 for information on each sample set). In order to control for inter-experimental or inter-group variation, expression data for each of the genes in a *cps* biosynthetic locus was averaged within each individual sample, and the relative abundance of expression of each of the *cps* loci was calculated. Trends highlighted in some of the original studies (such as increased expression of the *cps5* locus[31]) were still evident in the current analysis.

Table 2.8 Additional information on relative cps *expression from previously published datasets*
dpi: days post inoculation

Panel number	Strain of *B. thetaiotaomicron*	Diet	Mouse genotype	Other community members	Time of sampling (dpi)	# of animals / group (n)	GEO/SRA Accession Number	Reference
1	VPI-5482	High- vs. low-fiber	NMRI	No	10	9 (high-fiber), 3 (low-	GSE2231	31

						fiber)		
2	VPI-5482	Suckling or weaned pups	NMRI	No	17/30	6	GSE5279	32
3	VPI-5482	High-fiber	C57BL/6, Rag1-/-	No	10	4	GSE9019	4
4	VPI-5482	High-fiber	C57BL/6, Rag1-/-	No	10	4	GSE65455	47
5	VPI-5482	High- vs. low-fiber	C57BL/6	Yes (12 species)	42	7	GSE48532	48
6	VPI-5482	High- vs. low-fiber, or prebiotic	Swiss Webster	Yes (13 species)	54	3	SRP092461, SRP092458, SRP092453	43
7	VPI-5482	High-fiber	C57BL/6	Yes/No	20	3	GSE5867	49

Quantification of fecal IgA via ELISA. IgA was isolated from mouse stool by preparing a 10% solution in PBS-BSA (1% BSA), vortexing for 20 minutes, centrifuging for 20 minutes at 14,000 rpm (18,400 rcf), and collecting the supernatant. For quantification of anti-CPS or anti-acapsular strain IgA responses, plates were coated with 50 µL of a 0.2 mg/mL solution of purified CPS in bicarbonate buffer (15 mM sodium carbonate, 35 mM sodium bicarbonate, 3 mM sodium azide, pH 8.5) and incubated for 2 hours at room temperature. Plates were then washed 3 times in PBS-T (0.05% Tween 20), then blocked with PBS-BSA (30 min at room temperature), and again washed with PBS-T. The fecal supernatant containing IgA was diluted 1:10 in PBS-BSA and added to the plate (50 µL per well) before an overnight incubation at 4 °C. Wells were washed 3 times with PBS-T prior to the addition of 50 µL of horseradish peroxidase (HRP)- conjugated goat anti-mouse IgA (Southern Biotech; diluted 1:1000 in PBS-BSA). After a 2 hour incubation, samples were washed three times with PBS-T, and 50 µL of biotinyl tyramide solution (diluted 1:500 in 0.1 M sodium borate buffer, pH 8, supplemented with 0.03% H_2O_2) was added to each well. Samples were incubated for 15 minutes at room temperature and washed 3 times in PBS-T; 50 µL of Neutralite Avidin-HRP (Southern Biotech; diluted 1:2000 in PBS-BSA) was added and incubated at room temperature for 2 hours. Samples were then washed 3 times in PBS-T and

revealed using 100 µL of 1 mM 2,2'-azino-di-[3-ethylbenzthiazoline]-sulfonate solution (ABTS, Roche) in citrate buffer (100 mM citric acid, 50 mM sodium phosphate, pH 4.2), containing 0.03% H_2O_2. Samples were incubated for 15 minutes at room temperature prior to measuring absorbance at 405 nanometers. As possible (depending on amount of IgA isolated), samples were run in duplicate. For quantification of total IgA from stool, plates were coated with 50 µL of a 1:1000 dilution of goat anti-mouse IgA (Southern Biotech) in bicarbonate buffer and incubated for 2 hours at room temperature. Plates were then washed 3 times in PBS-T, then blocked with PBS-BSA (30 minutes at room temperature). Serial dilutions of the fecal supernatant were added to the plate (50 µL per well), starting with a 1:10 dilution, followed by serial 1:4 dilutions in BSA-PSA. Plates were incubated at room temperature for 2 hours. Wells were washed 3 times with PBS-T prior to the addition of 50 µL of horseradish peroxidase (HRP)-conjugated goat anti-mouse IgA (diluted 1:1000 in PBS-BSA). After a 2 hour incubation, samples were washed three times with PBS-T, and 100 µL of 1 mM ABTS solution in citrate buffer supplemented with 0.03% H_2O_2 was added to each well, prior to measuring absorbance at 405 nanometers. A standard curve of mouse purified IgA (Southern Biotech) was used to calculate IgA titers from each sample.

RNA-Seq analysis on bacteria monoassociated in germ-free mice. Groups of 3 C57BL/6 wild-type mice were colonized with a single strain of B. thetaiotaomicron VPI-5482 (the cps1, cps2, cps4, cps5, cps6, or acapsular strains) for 4 weeks. Mice were fed a standard high-fiber diet (LabDiet 5013, LabDiet). RNA was extracted as described above and then treated with Ribo-Zero rRNA Removal Kit (Illumina Inc.) and concentrated using RNA Clean and Concentrator -5 kit (Zymo Research Corp, Irvine, CA). Sequencing libraries were prepared using TruSeq

barcoding adaptors (Illumina Inc.), and 24 samples were multiplexed and sequenced with 50 base pair single end reads in one lane of an Illumina HiSeq instrument at the University of Michigan Sequencing Core. Demultiplexed samples were analyzed via Arraystar software (DNASTAR, Inc.) using RPKM normalization and the default parameters. Gene expression for each of the 5 CPS-expressing strains was normalized to the acapsular strain, and gene expression significantly upregulated or downregulated (5-fold cutoff, and significant by a moderated t test with Benjamini-Hochberg correction) was included in Table 2.2.

Bacterial growth rate analysis. Growth rate analysis was essentially as described previously[8]. Briefly, strains were grown overnight in TYG medium, and were subcultured and grown overnight in minimal medium containing 0.5% glucose. Cells were washed in double strength minimal medium containing no carbohydrate, prior to resuspension in the same medium. 100 µL aliquots of the cell suspensions were then inoculated into 96-well plates containing an equal volume of 1.0% solution of the desired carbohydrate (2.0% for *O*-glycans or rhamnogalacturonan I) and grown for 4 days anaerobically. Absorbance values were measured using an automated plate reading device (BioTek Instruments). Specific rate [(max OD_{600} - min OD_{600}) / time] was calculated using Microsoft Excel.

Antibiotic and antimicrobial susceptibility assays. Inhibition of bacterial growth by ciprofloxacin, polymyxin B, and human beta-defensin 3 were measured in a broth-based assay as described for antimicrobial peptide resistance[50]. Cultures were diluted to approximately OD_{600} 0.05 in TYG medium, and 100 µL diluted culture was added to serially diluted concentrations of antibiotic/antimicrobial peptide. All 3 assays were set up anaerobically. For the assay containing human beta-defensin 3, plates were then incubated aerobically for 90 minutes before being

placed back in an anaerobic environment. Growth was continuously monitored in an automated plate reading device and was compared to wells in which no compound was added; if wells achieved less than half the absorbance of the control at late log (~OD_{600} 0.8), growth was considered to be inhibited at that concentration. For the complement-mediated killing assay[27], 1 mL overnight culture of bacteria was diluted to OD_{600} 0.2, washed, and resuspended in 1 mL PBS containing 5 mM $MgCl_2$. 10 µL of normal human complement serum or freshly inactivated serum (incubated in a 56 °C water bath for 30 minutes) was added to 3 wells containing 90 µL of the bacterial resuspension. Wells were incubated anaerobically for 1 hour prior to serial dilution and plating on solid media. Each of the 3 replicates were normalized to counts for the heat inactivated complement [(inactive serum-treated counts – serum-treated counts) / inactive serum-treated counts] and then averaged and expressed as percent survival.

Laser capture microdissection and DNA extraction. Methanol Carnoy's fixed intestinal segments were paraffin embedded prior to microscope slide generation. Slides were deparaffinized in xylene, twice incubated in 100% ethanol, and retreated with xylene prior to drying briefly and placing in a dessicator for approximately 1 hour. Mucus layer (~30 µm immediately adjacent to epithelium) or lumen-associated (middle of lumen) samples were taken for each intestinal segment (ileum, cecum, and distal colon) of the $Rag1^{-/-}$ mice using an Arcturus Veritas microdissection instrument (Applied Biosystems). DNA was extracted from each sample with the PicoPure DNA extraction kit (Applied Biosystems). Prior to qPCR for quantifying relative abundance of each strain, a PCR for non-specific amplification of the barcoding vector was employed (primers are listed in Table 2.6 and Table 2.7). 10 ng of extracted DNA were amplified in duplicate for each sample with Platinum Pfx (Invitrogen) with the following

protocol: 95 °C for 5 minutes; 20 cycles of: 95 °C for 30 seconds, 58 °C for 30 seconds, and 72 °C for 30 seconds; final extension at 72 °C for 10 minutes. After PCR, duplicate samples were pooled and purified using Qiagen MinElute PCR Purification Kit and diluted (ileum diluted 1:10, colon segments diluted 1:100) prior to use in qPCR.

Identification of cps biosynthetic loci in sequenced isolates and in metagenomes. Strains of the *B. theta* genomic clique in the Integrated Microbial Genomes (IMG) database[51], as well as *Bacteroides faecis* MAJ27 (each with a >98% identity hit to the VPI-5482 16S rDNA) were probed with genes from the 8 *cps* loci from the VPI-5482 strain using the IMG "Genome Gene Best Homologs" tool with an initial percent identity cutoff of 20%. Strains with an identified homolog meeting identity cutoffs (90%, 80%, 60%, 40%, 20%, 0%) were shaded according to the color scale indicated in Figure 2.10A (from "high"/90% to "low"/0% homology). Type strains of *B. ovatus* (ATCC 8483) and *B. fragilis* (NCTC 9343) were included as controls. To identify other *cps* loci within each genome, multiple strategies were undertaken. First, homologs of flanking genes on both sides of known *cps* loci were investigated for the presence of neighboring, novel loci in other strains. Then, a blastp search with an E-value cutoff of 1e-5 was performed for conserved regulatory genes (*upxY* and *upxZ* family genes from the *cps5* locus, BT1656 and BT1655) and for conserved export machinery genes (BT0398 from *cps1* and BT1654 from *cps5*). To identify novel *cps* loci, each of the 14 strains was used as a reference strain in the "Genome Gene Best Homologs" tool and compared against the other 13 strains. A sufficient length of contiguous genome assembly was required to either identify the genomic location and/or compare gene content and homology to other loci. Although, gaps in assembled genomes prohibited full locus comparison in several cases. Genes in homologous loci were

typically >98% similar to one another at the amino acid level. 3 genes within each locus were identified via this method that appeared to be highly conserved only among those strains that encoded the entire locus. A blastn search (E-value cutoff 1e-5) of all isolate genomes in IMG was performed, and each gene was confirmed to be specific to its cognate locus and not found within other genomes. Only the *cps8* locus, which is found on a mobile genetic element, was found within other *Bacteroides* species (these were confirmed to contain a locus homologous to *cps8*). Genes from each locus were not found to be present outside of the *Bacteroides* genus. 49 unique *cps* biosynthetic loci were thus identified.

The 3 unique genes from each locus were also used to probe for each of the loci in the 147 Human Microbiome Project gut metagenomic samples found on the IMG website. Each sample was probed with the 3 genes from each of the 49 unique loci via blastn with more stringent criteria (E-value cutoff 1e-50, >100 bp hit, 95% identity). A sample with at least 2 out of 3 genes present was considered positive for the corresponding locus. 3 genes conserved within all 14 isolate genomes (but not found within other isolate genomes on IMG) were used to determine the presence of *B. theta* within each metagenomic sample. Genes used are listed in Table 2.9.

Table 2.9 Strains, genes, and capsular polysaccharide synthesis loci in comparative genomic and metagenomic analyses

Strain	**Unique Locus Type**	**Genes in Locus**	**Gene 1**	**Gene 2**	**Gene 3**
VPI-5482	0	(Conserved *B. theta* genes)	BT0271	BT3954	BT4297
VPI-5482	1A	BT0376-BT0400	BT0384	BT0390	BT0392
NLAE-zl-G288	1B	G288DRAFT_04050-G288DRAFT_04075	G288DRAFT_04064	G288DRAFT_04066	G288DRAFT_04067
NLAE-zl-P32	1C	P32DRAFT_00714-P32DRAFT_00716;	P32DRAFT_04748	P32DRAFT_04754	P32DRAFT_04756

		P32DRAFT_04747-P32DRAFT_04764			
dnLKV9	1D	C799_01601-C799_01622	C799_01610	C799_01613	C799_01617
KPPR-3	1E	IE52DRAFT_03468-IE52DRAFT_03472; IE52DRAFT_00849-IE52DRAFT_00868	IE52DRAFT_00856	IE52DRAFT_00860	IE52DRAFT_00862
NLAE-zl-C523	1F	C523DRAFT_03082-C523DRAFT_03105	C523DRAFT_03089	C523DRAFT_03092	C523DRAFT_03094
NLAE-zl-H207	1G	H207DRAFT_04956-H207DRAFT_04979; H207DRAFT_00557-H207DRAFT_00558	H207DRAFT_04964	H207DRAFT_04966	H207DRAFT_04970
sp. 1_1_14	1H	HMPREF9007_01300-HMPREF9007_01322	HMPREF9007_01305	HMPREF9007_01313	HMPREF9007_01314
B. faecis MAJ27	1I	KCYDRAFT_02823-KCYDRAFT_02842	KCYDRAFT_02832	KCYDRAFT_02833	KCYDRAFT_02835
VPI-5482	2A	BT0462-BT0481	BT0469	BT0471	BT0472
sp. 1_1_14	2B	HMPREF9007_01400-HMPREF9007_01413;	HMPREF9007_01405	HMPREF9007_01408	HMPREF9007_01409
B. faecis MAJ27	2C	KCYDRAFT_02888-KCYDRAFT_02904	KCYDRAFT_02892	KCYDRAFT_02896	KCYDRAFT_02897
VPI-5482	3A	BT0596-BT0615	BT0605	BT0607	BT0610
NLAE-zl-H463	3B	H463DRAFT_01871-H463DRAFT_01872; H463DRAFT_03984-H463DRAFT_04009	H463DRAFT_03995	H463DRAFT_03999	H463DRAFT_04001
B. faecis MAJ27	3C	; KCYDRAFT_03614-KCYDRAFT_03650	KCYDRAFT_03635	KCYDRAFT_03640	KCYDRAFT_03641
NLAE-zl-C523	3D	C523DRAFT_04750-C523DRAFT_04787	C523DRAFT_04761	C523DRAFT_04770	C523DRAFT_04773
NLAE-zl-G288	3E	G288DRAFT_04483-G288DRAFT_04516	G288DRAFT_04496	G288DRAFT_04500	G288DRAFT_04503
NLAE-zl-H207	3F	H207DRAFT_01156-H207DRAFT_01178; H207DRAFT_04252-H207DRAFT_04253	H207DRAFT_01162	H207DRAFT_01165	H207DRAFT_01170
sp. 1_1_14	3G	HMPREF9007_04174-HMPREF9007_04188; HMPREF9007_01983-HMPREF9007_01992	HMPREF9007_04180	HMPREF9007_04184	HMPREF9007_04186
VPI-5482	4A	BT1338-BT1358	BT1343	BT1346	BT1352
NLAE-zl-G288	4B	G288DRAFT_03286-G288DRAFT_03317	G288DRAFT_03290	G288DRAFT_03296	G288DRAFT_03305
KPPR-3	4C	IE52DRAFT_04698-IE52DRAFT_04710; IE52DRAFT_02183-IE52DRAFT_02187	IE52DRAFT_04700	IE52DRAFT_04701	IE52DRAFT_04703

NLAE-zl-P32	4D	P32DRAFT_04638-P32DRAFT_04648; P32DRAFT_03863-P32DRAFT_03865	P32DRAFT_04640	P32DRAFT_04643	P32DRAFT_04644
dnLKV9	4E	C799_00539-C799_00552; C799_00537-C799_00538	C799_00545	C799_00547	C799_00550
NLAE-zl-C523	4F	C523DRAFT_04191-C523DRAFT_04210	C523DRAFT_04199	C523DRAFT_04201	C523DRAFT_04202
NLAE-zl-H207	4G	H207DRAFT_00491-H207DRAFT_00517; H207DRAFT_02827-H207DRAFT_02828	H207DRAFT_00493	H207DRAFT_00503	H207DRAFT_00506
sp. AR20	4H	IE59DRAFT_04535-IE59DRAFT_04549; IE59DRAFT_03033-IE59DRAFT_03035	IE59DRAFT_04537	IE59DRAFT_04539	IE59DRAFT_04541
B. faecis MAJ27	4I	KCYDRAFT_03889-KCYDRAFT_03897; KCYDRAFT_03898-KCYDRAFT_03904	KCYDRAFT_03891	KCYDRAFT_03895	KCYDRAFT_03896
VPI-5482	5A	BT1642-BT1656	BT1647	BT1650	BT1651
sp. 1_1_6	5B	BSIG_05102-BSIG_05113; BSIG_04226	BSIG_05106	BSIG_05109	BSIG_05110
dnLKV9	5C	C799_00203-C799_00217	C799_00208	C799_00214	C799_00216
NLAE-zl-C523	5D	C523DRAFT_03947-C523DRAFT_03958	C523DRAFT_03952	C523DRAFT_03953	C523DRAFT_03955
sp. 1_1_14	5E	HMPREF9007_02960-HMPREF9007_02973; HMPREF9007_04401-HMPREF9007_04406	HMPREF9007_02967	HMPREF9007_02970	HMPREF9007_02971
VPI-5482	6A	BT1707-BT1725	BT1710	BT1715	BT1716
dnLKV9	6B	C799_00133-C799_00148;	C799_00138	C799_00139	C799_00141
sp. 1_1_14	6C	HMPREF9007_04325-HMPREF9007_04348	HMPREF9007_04332	HMPREF9007_04333	HMPREF9007_04341
sp. AR20	6D	; IE59DRAFT_03884-IE59DRAFT_03906	IE59DRAFT_03891	IE59DRAFT_03892	IE59DRAFT_03893
B. faecis MAJ27	6E	KCYDRAFT_04301-KCYDRAFT_04316	KCYDRAFT_04309	KCYDRAFT_04313	KCYDRAFT_04314
VPI-5482	7A	BT2862-BT2886	BT2880	BT2884	BT2872
NLAE-zl-H207	7B	H207DRAFT_03140-H207DRAFT_03168	H207DRAFT_03149	H207DRAFT_03151	H207DRAFT_03153
VPI-5482	8A	BT0037-BT0068	BT0044	BT0045	BT0052

NLAE-zl-H207	9A	H207DRAFT_04437-H207DRAFT_04458; H207DRAFT_00656-H207DRAFT_00660	H207DRAFT_04444	H207DRAFT_04448	H207DRAFT_04449
dnLKV9	9B	C799_00973-C799_01005	C799_00982	C799_00992	C799_01001
NLAE-zl-G288	10A	G288DRAFT_03748-G288DRAFT_03774	G288DRAFT_03759	G288DRAFT_03760	G288DRAFT_03762
B. faecis MAJ27	10B	KCYDRAFT_04608-KCYDRAFT_04634;	KCYDRAFT_04620	KCYDRAFT_04622	KCYDRAFT_04627
NLAE-zl-H207	11A	H207DRAFT_00406-H207DRAFT_00430	H207DRAFT_00409	H207DRAFT_00411	H207DRAFT_00416
sp. AR20	0A	; IE59DRAFT_01957-IE59DRAFT_01978	IE59DRAFT_01963	IE59DRAFT_01965	IE59DRAFT_01967
sp. AR20	0B	IE59DRAFT_03239-IE59DRAFT_03267	IE59DRAFT_03249	IE59DRAFT_03254	IE59DRAFT_03258
sp. AR20	0C	IE59DRAFT_04693-IE59DRAFT_04705 (either site 5 or 6 based on adjacent amidase gene)	IE59DRAFT_04698	IE59DRAFT_04699	IE59DRAFT_04701

For *cps* locus synteny analysis (Figure 2.11), representative loci from each of the 9 unique *cps* locus types at genomic Site 1 (location of *cps1*) were compared using the IMG "Genome Gene Best Homologs" tool. Especially for some conserved gene families (e.g. *upxY*/*upxZ*) the closest homolog for a gene in one strain was found in a *cps* locus at a different genomic site. For these genes, Blastp (E-value cutoff 1e-2) was performed. Cutoffs of 80% or 90% identity were used to label genes as homologous.

Statistical analysis. None of the investigators were blinded to sample identity during any of the described experiments and no samples or measurements were excluded from analysis. For each experiment, the total number of biological replicates "n" is indicated in the corresponding legend along with a description of what is considered a replicate (mice, replicate cultures, antibody titers, etc.). Statistical analyses (other than Dirichlet regression and linear mixed models) were performed in Prism software (GraphPad Software, Inc., La Jolla, CA). As data did not generally

follow a normal distribution, nonparametric tests (Mann-Whitney/Kruskal-Wallis) were used to determine statistical significance. Unless otherwise indicated, statistical significance is indicated as follows: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; and **** $p < 0.0001$. Number of animals used (n) and statistical tests used are indicated in each figure legend. For comparing relative abundance of *cps* locus expression (Figure 2.1) between pairs of experimental groups, Dirichlet regression was performed in R using the alternate parameterization in the "DirichletReg" package[52] (version 0.6-3). The parameters in this representation of the Dirichlet distribution are the relative proportions of *cps* locus expression and the total expression, which is a measure of precision. The parameterization models the logarithm of the ratio of the relative expression of locus cps*c* (referring to CPS loci *cps1*, *cps2*, *cps3*, *cps4*, *cps5*, *cps6*, or *cps8*) to the relative proportion of the reference *cps7* locus as a linear combination of the given variables of interest. The variables of interest were diet and low-fiber diet interval for comparison of the wild-type, high-fiber fed mice with the wild-type mice whose diets oscillated between intervals of low- and high-fiber diets. The variable of interest for comparison of the wild-type, high-fiber fed mice with the $Rag1^{-/-}$, high-fiber fed mice, on the other hand, was mouse genotype. The precision for a given model was allowed to vary by group in cases where a likelihood ratio test indicated at significance level $p < 0.05$ that this model was superior to a corresponding model with constant precision.

For the IgA titer assay in Figure 2.6, it was necessary to use the entire fecal sample to provide enough IgA for the assay. Thus, to correlate specific anti-CPS IgA titers with strain abundance, a good estimation of the relative strain abundance at the time of the IgA titer data was generated by averaging the two closest relative strain abundances to the IgA time point (e.g.

for a time point on day 25, the Day 20 and Day 30 abundances were averaged). A Pearson correlation was then performed.

Linear mixed models for Figure 2.6 were created and run in R using the package "lme4"[53] (version 1.1-12). First, to examine the effect of mouse genotype on anti-CPS fecal IgA levels, a univariate model dependent only on mouse genotype as a predictor was employed: $IgA_{ijk} = A_0 + A_1 Genotype_i + R_i + e_{ijk}$ where *i* denotes an individual mouse, *j* denotes time, *k* denotes CPS type, R_i is the random effect of the mouse and e_{ijk} is the error term. The variable IgA_{ijk} refers to the normalized OD_{405} values shown in Figure 2.6 for subject *i* at time *j* for CPS type *k* and $Genotype_i$ is an indicator variable of the mouse genotype. This means that wild-type has value 0 and MyD88$^{-/-}$/Trif$^{-/-}$ has value 1. Next, the contributions of both mouse genotype and day were examined in a multivariate model: $IgA_{ijk} = A_0 + A_1 Genotype_i + \sum_{d=25,45,65} B_d Day_{dj} + \sum_{d=25,45,65} C_d Genotype_i * Day_{dj} + R_i + e_{ijk}$ with notation as described above. $Day_{d,j}$ are indicator variables for days 25, 45, and 65. Furthermore, comparisons of anti-CPS5 titers to other anti-CPS titers at each time point were carried out using another model incorporating identity of each targeted CPS: $IgA_{ijk} = A_0 + A_1 Genotype_i + \sum_{d=25,45,65} B_d Day_{dj} + \sum_{t=1,2,3,4,6,8} C_t Type_{tk} + \sum_{d=25,45,65} D_d Genotype_i * Day_{dj} + \sum_{t=1,2,3,4,6,8} E_t Genotype_i * Type_{tk} + \sum_{\substack{d=25,45,65 \\ t=1,2,3,4,6,8}} F_{dt} Day_{dj} * Type_{tk} + R_i + e_{ijk}$ where $Type_{tk}$ are indicator variables for CPS types CPS1, CPS2, CPS3, CPS4, CPS6, and CPS8, with other notation as described above. Linear combinations of coefficients in the mixed models represent differences between groups and between time points for each group. Significance was determined for each difference by testing the null hypothesis that the linear combination of

coefficients representing the given difference is 0.

Data and software availability. Genome sequencing and RNA-Seq data from this study have been deposited in the NCBI Short-Read Archive (SRA) under the BioProject ID PRJNA395217 (BioSample IDs for genomes: SAMN07373223-SAMN07373232; BioSample IDs for RNA-Seq data: SAMN07374632- SAMN07374649).

Notes

This work has been reprinted and modified with permission from Porter, N. T., Canales, P., Peterson, D. A. & Martens, E. C. A Subset of Polysaccharide Capsules in the Human Symbiont *Bacteroides thetaiotaomicron* Promote Increased Competitive Fitness in the Mouse Gut. *Cell Host Microbe* **22,** 494–506 (2017).

References

1. El Kaoutari, A. *et al.* The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11,** 497–504 (2013).
2. Donia, M. S. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158,** 1402–1414 (2014).
3. Coyne, M. J. & Comstock, L. E. Niche-specific features of the intestinal Bacteroidales. *J. Bacteriol.* **190,** 736–42 (2008).
4. 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).
5. Mazmanian, S. K. S. S. K. *et al.* An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,** 107–118 (2005).
6. Porter, N. T. & Martens, E. C. The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu. Rev. Microbiol.* **71,** 349–369 (2017).
7. 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,** 447–57 (2008).
8. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9,** e1001221 (2011).
9. Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517,** 165–169 (2015).
10. Patrick, S. *et al.* Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of Bacteroides fragilis. *Microbiology* **156,** 3255–69 (2010).
11. Xu, J. *et al.* Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol.* **5,** e156 (2007).
12. Martens, E. C., Kelly, A. G., Tauzin, A. S. & Brumer, H. The devil lies in the details: How variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *J. Mol. Biol.* **426,** 3851–3865 (2014).
13. Krinos, C. M. *et al.* Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414,** 555–558 (2001).
14. Coyne, M. J., Weinacht, K. G., Krinos, C. M. & Comstock, L. E. Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 10446–51 (2003).
15. Hickey, C. A. *et al.* Colitogenic Bacteroides thetaiotaomicron antigens access host immune cells in a sulfatase-dependent manner via outer membrane vesicles. *Cell Host Microbe* **17,** 672–680 (2015).

16. Chatzidaki-Livanis, M., Coyne, M. J. & Comstock, L. E. A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of Bacteroides fragilis. *J. Bacteriol.* **191,** 7288–95 (2009).

17. Chatzidaki-Livanis, M., Weinacht, K. G. & Comstock, L. E. Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont Bacteroides fragilis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 11976–80 (2010).

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

19. Cameron, E. A. *et al.* Multifunctional nutrient-binding proteins adapt human symbiotic bacteria for glycan competition in the gut by separately promoting enhanced sensing and catalysis. *MBio* **5,** e01441-14 (2014).

20. Avery, O. T. & Dubos, R. The protective action of a specific enzyme against Type III Pneumococcus infection in mice. *J. Exp. Med.* **54,** 73–89 (1931).

21. Thurlow, L. R., Thomas, V. C., Fleming, S. D. & Hancock, L. E. Enterococcus faecalis capsular polysaccharide serotypes C and D and their contributions to host innate immune evasion. *Infect. Immun.* **77,** 5551–7 (2009).

22. Horwitz, M. A. & Silverstein, S. C. Influence of the Escherichia coli capsule on complement fixation and on phagocytosis and killing by human phagocytes. *J. Clin. Invest.* **65,** 82–94 (1980).

23. Cunnion, K. M., Zhang, H.-M. & Frank, M. M. Availability of complement bound to Staphylococcus aureus to interact with membrane complement receptors influences efficiency of phagocytosis. *Infect. Immun.* **71,** 656–62 (2003).

24. Cress, B. F. *et al.* Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiol. Rev.* **38,** 660–97 (2014).

25. Vecchiarelli, A. *et al.* Purified capsular polysaccharide of Cryptococcus neoformans induces interleukin-10 secretion by human monocytes. *Infect. Immun.* **64,** 2846–9 (1996).

26. Jelacic, T. M. *et al.* Exposure to Bacillus anthracis capsule results in suppression of human monocyte-derived dendritic cells. *Infect. Immun.* **82,** 3405–16 (2014).

27. 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. U. S. A.* **105,** 13099–104 (2008).

28. Coyne, M. J., Reinap, B., Lee, M. M. & Comstock, L. E. Human symbionts use a host-like pathway for surface fucosylation. *Science* **307,** 1778–81 (2005).

29. Round, J. L. *et al.* The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332,** 974–7 (2011).

30. Neff, C. P. *et al.* Diverse intestinal bacteria contain putative zwitterionic capsular polysaccharides with anti-inflammatory properties. *Cell Host Microbe* **20,** 1–13 (2016).

31. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial

symbiont. *Science* **307,** 1955–9 (2005).

32. Bjursell, M. K., Martens, E. C. & Gordon, J. I. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. *J. Biol. Chem.* **281,** 36269–79 (2006).

33. 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. U. S. A.* **105,** 3951–6 (2008).

34. Lee, S. M. *et al.* Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501,** 3–8 (2013).

35. Markowitz, V. M. *et al.* IMG/M-HMP: a metagenome comparative analysis system for the Human Microbiome Project. *PLoS One* **7,** e40151 (2012).

36. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–5 (2008).

37. Shen, Y. *et al.* Outer membrane vesicles of a human commensal mediate immune regulation and disease protection. *Cell Host Microbe* **12,** 509–20 (2012).

38. Whitaker, W. R., Shepherd, E. S. & Sonnenburg, J. L. Tunable Expression Tools Enable Single-Cell Strain Distinction in the Gut Microbiome. *Cell* **169,** 538–546.e12 (2017).

39. Holdeman, L. V. E. *Anaerobe Laboratory Manual*. (Anaerobe Laboratory Virginia Polytechnic Institute & State University, 1977).

40. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16,** 1105–15 (2008).

41. Rogers, T. E. *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol. Microbiol.* **88,** 1–15 (2013).

42. Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506,** 498–502 (2014).

43. Desai, M. S. *et al.* A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167,** 1339–1353 e21 (2016).

44. Reed, M. J. *et al.* A new rat model of type 2 diabetes: the fat-fed, streptozotocin-treated rat. *Metabolism.* **49,** 1390–4 (2000).

45. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 11070–5 (2005).

46. Xiao, Z., Storms, R. & Tsang, A. A quantitative starch–iodine method for measuring alpha-amylase and glucoamylase activities. *Anal. Biochem.* **351,** 146–148 (2006).

47. 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.* **290,** jbc.M114.633800 (2015).

48. McNulty, N. P. *et al.* Effects of diet on resource utilization by a model human gut

microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome. *PLoS Biol.* **11,** e1001637 (2013).

49. Sonnenburg, J. L., Chen, C. T. L. & Gordon, J. I. Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. *PLoS Biol.* **4,** 2213–2226 (2006).

50. Cullen, T. W. *et al.* Gut microbiota. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. *Science* **347,** 170–5 (2015).

51. Markowitz, V. M. *et al.* IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* **42,** D560–D567 (2014).

52. Maier, M. J. DirichletReg : Dirichlet Regression for Compositional Data in R. 13 (2014).

53. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67,** 51 (2015).

CHAPTER 3

Bacteriophage Target and Modulate Expression of Diverse Capsular Polysaccharides in the Gut Symbiont *Bacteroides thetaiotaomicron*

Abstract

Individual members of the prominent gut *Bacteroides* genus typically synthesize multiple, distinct surface capsular polysaccharides (CPS). Little is known about the roles that the majority of these CPS play. We used a collection of 10 isogenic strains of the model human gut symbiont *Bacteroides thetaiotaomicron* that each either expresses a different CPS, multiple CPS (wild-type), or no CPS (acapsular) to isolate a diverse collection of 52 bacteriophages. Evaluating the infectivity of these phages on the entire panel of bacterial strains revealed that each phage is only able to infect a subset of the strains. Moreover, *in vitro* infections of wild-type *B. thetaiotaomicron*, which is naturally equipped to shift between expression of different CPS, revealed that individual phages are capable of modulating the repertoire of CPS synthetic genes that are expressed, presumably by selecting against strains that express the CPS they target. The independently isolated phages can be clustered into phenotypically similar groups based on their infection profiles of strains expressing different CPS, suggesting that phage with similar profiles are either related or have evolved to recognize similar CPS motifs. Taken together, our results reveal distinct roles for *Bacteroides* CPS in modulating phage susceptibility, either by serving as receptors themselves or blocking access to cell surface receptors that phage recognize. Given the diversity of CPS encoded by members of the same species, our results provide important insight

into the evolutionary adaptations of gut bacteria to avoid phage predation and also hold important implications for the potential use of phage to therapeutically target these diverse bacterial symbionts.

Introduction

The human gut microbiota (the collection of microorganisms inhabiting the human intestinal tract) includes a diverse population of bacteria, with hundreds of species present in the communities of individual people[1]. The gut microbiota varies greatly among individuals and over time, with many factors contributing to these differences, including exposure throughout life, diet and the host immune response[2–4]. Members of the prominent *Bacteroides* genus encode many degradative enzymes for the utilization of dietary fiber polysaccharides[5]. These functions, often grouped into co-regulated gene clusters (polysaccharide utilization loci, PULs) provide a competitive advantage to the bacterium in the competitive gut environment by allowing them to access different polysaccharide nutrients[6–8]. An additional, but comparatively understudied adaptation of the *Bacteroides,* is that they typically encode several capsular polysaccharide (CPS) synthesis loci within a single strain. While a few of these CPS take part in immune evasion or immunomodulation[9–13] the sheer diversity and number of CPS synthesis loci in the *Bacteroides* suggests that they could also fill other orthogonal roles as well[13–16].

Bacterial viruses or bacteriophage, like the bacteria on which they prey, vary greatly among individuals and are even responsive to host dietary changes (likely due to corresponding shifts in bacterial community composition)[17,18]. However, much less is understood about the human gut virome than its bacterial counterpart. Additionally, while *Bacteroides*-specific phage are species- and often even strain-specific[19–21], little is known about the receptors required for

viral attachment and entry. Notably, many fewer viral genomes exist for phages infecting the prominent Bacteroidetes phylum than for those targeting other gut phyla, complicating the identification of potential receptors via bioinformatic means[22]. Using *in vitro* binding assays, Puig and colleagues identified two possible outer membrane protein receptors for one particular *B. fragilis*-targeting bacteriophage[23], but these targets have not been confirmed by genetics or other methods.

While many phages target protein receptors, some have evolved to utilize polysaccharides, including CPS, present on the cell surface as receptors[24]. This has been particularly evident in *Escherichia coli*, where inactivation of particular CPS synthesis genes or enzymatic removal of the capsule renders the host cell resistant to infection by its cognate phage[25–27]. Additionally, other gut bacteria, such as *Campylobacter jejuni* and *Lactococcus lactis*, are resistant to specific phage isolates when capsule synthesis genes are inactivated[28,29]. Notably, Ainsworth and colleagues replaced EPS synthesis genes in *Lactococcus lactis* encoding one capsule type with those encoding another. This rendered the bacterium resistant to phages targeting the first EPS type but susceptible to those targeting strains that produced the second EPS type, demonstrating that capsule is required for productive phage infection[29].

CPS can in some cases also block access to proteinaceous or other phage receptors located on the cell surface, abolishing or partially blocking infection[30,31]. Many phages encode "depolymerase" enzymes (i.e., glycoside hydrolases and polysaccharide lyases) to degrade the capsule layer and provide access to the cell surface, also indicating the effectiveness of capsule in blocking phage in the absence of these enzymes[26,32–34]. Notably, a single virion may encode multiple depolymerases to broaden its bacterial host range to include more diverse CPS-expressing strains[35,36].

While some phage-receptor interactions have been relatively well characterized, little is known about the role and identity of phage receptors in the *Bacteroides*. However, some clues in three classic articles on *Bacteroides* phage[19–21] point to capsules playing a critical role in *Bacteroides* phage infection. Collectively, these studies observed that: 1) Phages are species- and strain-specific within the *Bacteroides* genus, indicating that the cellular determinants of infection are largely confined to, but heterogeneous within, a species. 2) *Bacteroides* exhibit "pseudolysogeny", wherein only a proportion of the cells in a population are actively infected, and a proportion of uninfected cells can be isolated and readily infected. Additionally, plaque assays may exhibit great variability among experiments with wild-type strains, at least for some phages. These observations suggest a phase-variable factor is required for infection. 3) Via microscopy, presence/absence (or thickness) of capsule has been qualitatively correlated with productive phage infection.

Given the observations that *Bacteroides* CPS are highly variable even within a single species[13,15] and employ a complex regulatory hierarchy that diversifies expression within a population and over time[13,37], CPS are ideal candidates for phage receptors in the *Bacteroides*. A lack of studies describing phage receptors in the *Bacteroides* may be due to the complexity of isolating a single phage-capsule interaction, as individual strains typically encode for multiple, phase-variable CPS synthesis loci[14,37,38]. Employing a panel of strains of the model symbiont *B. thetaiotaomicron* that each expresses a different CPS, we isolated 52 bacteriophages and characterized their host range on strains expressing various CPS types. We show that these phages infect a limited set of the single CPS-expressing strains, and provide evidence that some *B. thetaiotaomicron*-specific bacteriophage use capsule as a receptor whereas others are blocked by the presence of specific CPS.

Results

We first set out to generate a large collection of bacteriophages that might target different bacterial CPS. To this end, we used the wild-type strain (which can express 8 distinct CPS types), an acapsular strain, and an isogenic panel of 8 single CPS-expressing strains (designated as "cps1" to "cps8" strains) as host strains for isolating phages from a local wastewater treatment plant (Ann Arbor, Michigan) at two locations within the plant. We also varied isolation conditions and sampled at multiple time points, with the goal of generating a diverse collection of phages exhibiting a wide range of phenotypes (see further details on this strain collection in *Methods* and Table 3.1). Plaque morphologies varied greatly among the strains, ranging in size from < 0.5 mm to 3 mm or greater and ranging from very cloudy to clear. All phages were purified at least 3 times by picking isolated plaques and replating, and high titer phage lysates were generated for each of the 52 phages.

Table 3.1 Phages used in this study and details on their isolation

Designation	Date isolated	Original isolation strain	Bile in original isolation?	Direct Plating/Enrichment?	Site (at Ann Arbor Wastewater Treatment Plant)
ARB007	8/12/15	WT	No	Enrichment	East Plant
ARB009	10/8/15	Acapsular	No	Enrichment	East Plant
ARB010	10/8/15	cps1	No	Enrichment	East Plant
ARB014	10/8/15	cps5	No	Enrichment	East Plant
ARB016	10/8/15	cps7	No	Enrichment	East Plant
ARB017	10/8/15	cps8	No	Enrichment	East Plant
ARB019	10/8/15	Acapsular	No	Enrichment	East Plant
ARB021	2/8/16	cps1	Yes	Enrichment	West Plant
ARB022	2/8/16	cps5	Yes	Enrichment	West Plant
ARB024	2/8/16	Acapsular	Yes	Enrichment	West Plant
ARB025	2/8/16	WT	Yes	Enrichment	West Plant
ARB026	2/8/16	cps1	No	Enrichment	West Plant
ARB028	2/8/16	cps5	No	Enrichment	West Plant
ARB029	2/8/16	cps7	No	Enrichment	West Plant
ARB030	2/8/16	cps7	No	Enrichment	West Plant
ARB031	2/8/16	cps8	No	Enrichment	West Plant
ARB032	2/8/16	Acapsular	No	Enrichment	West Plant
ARB033	2/8/16	WT	No	Enrichment	West Plant
ARB034	2/10/16	cps1	Yes	Plate	West Plant
ARB037	2/10/16	cps5	Yes	Plate	West Plant

ARB038	2/10/16	cps1	No	Plate	West Plant
ARB041	2/10/16	Acapsular	No	Plate	West Plant
ARB043	2/10/16	cps5	No	Plate	West Plant
ARB044	2/10/16	cps5	No	Plate	West Plant
ARB045	2/10/16	cps5	No	Plate	West Plant
ARB047	2/10/16	cps8	No	Plate	West Plant
ARB050	2/18/16	Acapsular	No	Plate	West Plant
ARB0178	2/18/16	Acapsular	No	Plate	West Plant
ARB057	3/10/16	Acapsular	No	Plate	West Plant
ARB059	3/10/16	Acapsular	No	Plate	West Plant
ARB069	4/14/16	cps1	No	Plate	West Plant
ARB070	4/14/16	cps1	No	Plate	West Plant
ARB072	4/14/16	cps1	No	Plate	West Plant
ARB074	4/14/16	cps1	No	Plate	West Plant
ARB077	4/14/16	cps1	No	Plate	West Plant
ARB078	4/14/16	cps1	No	Plate	West Plant
ARB079	4/14/16	cps1	No	Plate	West Plant
ARB080	4/14/16	cps1	No	Plate	West Plant
ARB081	4/14/16	cps1	No	Plate	West Plant
ARB082	4/14/16	cps1	No	Plate	West Plant
ARB101	4/14/16	cps3	No	Plate	West Plant
ARB104	4/14/16	cps3	No	Plate	West Plant
ARB105	4/14/16	cps3	No	Plate	West Plant
ARB133	4/14/16	cps7	No	Plate	West Plant
ARB135	4/14/16	cps7	No	Plate	West Plant
ARB144	4/14/16	cps7	No	Plate	West Plant
ARB148	4/14/16	cps7	No	Plate	West Plant
ARB152	4/14/16	cps8	No	Plate	West Plant
ARB154	4/14/16	cps8	No	Plate	West Plant
ARB156	4/14/16	cps8	No	Plate	West Plant
ARB163	4/14/16	cps8	No	Plate	West Plant
ARB166	4/14/16	Acapsular	No	Plate	West Plant

We next tested whether each phage was able to infect each of the 10 isolation strains, or whether phages exhibited more limited host ranges. To effectively investigate the host range for dozens of strains would be challenging using traditional agar overlay assays; thus we employed a spot titer assay for semi-quantitative comparisons of infectivity on each bacterial strain (Figure 3.1). 1 µl of serial dilutions of phage lysates (from approximately 10^3 to 10^6 plaque forming units (PFU)/ml) were plated on each of the 10 bacterial strains (n = 3 replicates), and phage titers were normalized to the bacterial strain that typically exhibited the highest phage titer (the "preferred host strain"). More than half of phages plaqued most efficiently on their isolation strain; however, a substantial number (21 of 52 phages) were not isolated on their preferred host strain

(e.g. ARB7 was isolated on the wild type strain but produced the most plaques on the acapsular strain), or exhibited variable preferences in their host range (e.g. while ARB78 was isolated on the CPS1-expressing "cps1" strain, it variably produced the most plaques on the cps1, cps2, and wild-type strains). Interestingly, while great variation existed in relative host titers within the phage collection, grouping the phages based on their host ranges revealed 3 broad clusters (Figure 3.1). The topmost cluster ("Cluster 1") predominantly infected the cps7, cps8, and acapsular strains. The phages in the second cluster ("Cluster 2") generally exhibited robust infection of the acapsular strain, as well as limited infection of some combination of the cps1, cps5, cps7, and/or cps8 strains (with some strains also infecting the wild-type strain). Lastly, those in the bottom cluster ("Cluster 3") prioritized infection of the cps1, cps2, or cps3 strains, and also tended to have the highest infection rates of the wild-type strain (though plaques on the wild-type host were small and cloudy, data not shown). In summary, bacterial strains expressing different (or no) capsules display variable susceptibility to individual phages, implicating capsule in either facilitating or blocking phage infection.

The different phenotypes seen among the 3 clusters could point to distinct mechanisms of entry for phages in each cluster. Specifically, many of the phages in Clusters 1 and 2 robustly infect the acapsular strain, indicating that a cell surface receptor other than capsule exists. However, many of these phages also infect unique subsets of the single CPS-expressing strains. In these cases, specific CPS may either passively facilitate access to underlying phage receptors (for example by not blocking phage entry while others do) or may be targeted as co-receptors (either case resulting in higher plaquing efficiency). For phages that do not efficiently infect the acapsular strain (including many in Cluster 3), one or more specific CPS may either serve as a direct, exclusive receptor for phage, or binding to specific CPS may be a necessary co-receptor.

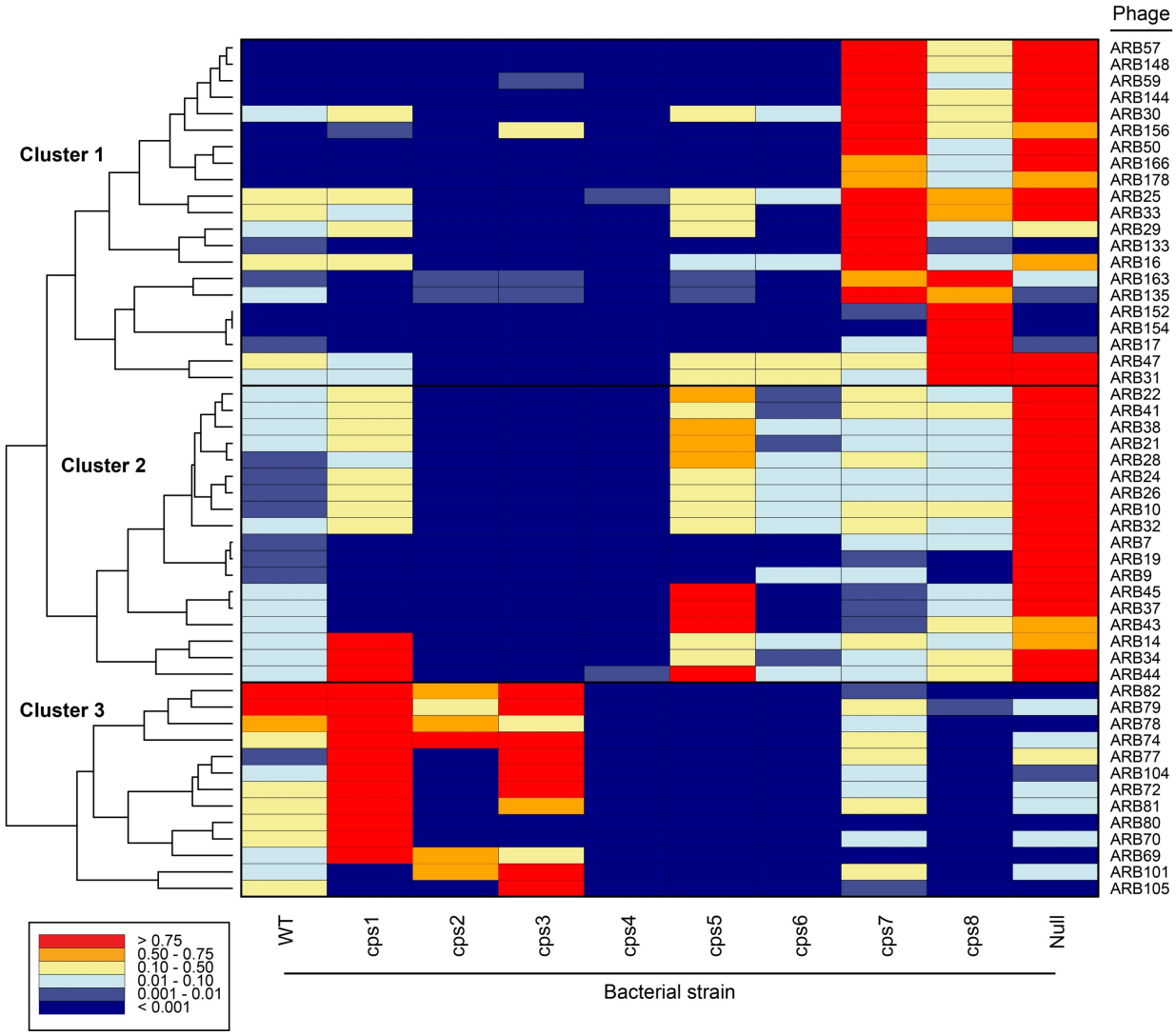


Figure 3.1 Host range of B. thetaiotaomicron*-targeting phages on strains expressing different CPS types.*

52 bacteriophages isolated and purified on the wild-type, acapsular, or the 8 single CPS-expressing strains were tested in a spot titer assay to determine phage host range. 10-fold serial dilutions of each phage ranging from approximately 1E6 to 1E3 plaque-forming units (PFU) / ml were spotted onto top agar plates containing the 10 bacterial strains, plates were grown overnight, and phage titers were then calculated. A heatmap showing titers normalized to the preferred host strain was then generated, and phage host ranges were clustered by phenotype. Each row in the heatmap corresponds to an individual phage, whereas each column corresponds to one of the 10 host strains.

To begin to differentiate between capsule as either a phage receptor or a "shield" protecting an underlying receptor, we further investigated a subset of 7 phages from Clusters 1 and 3 that were able to infect the wild-type strain, which is able to express all 8 of its CPS synthesis loci. We hypothesized that deleting a small number of CPS loci might render the cells immune to phage infection, especially in cases where the CPS is required as a receptor or co-

receptor. To this end, we titered phages (n = 6 replicates for each) on a set of bacterial strains with specific CPS locus deletions tailored to each phage's previously determined host range. We compared these to titers on the wild-type and the preferred host strains (Figure 3.2). Of the Cluster 1 phages, ARB25 produced very similar titers on the wild-type and acapsular strains, as well as on all other strains tested, including a CPS locus mutant lacking the 4 *cps* loci that were correlated with higher phage infection in our initial assay. While a strain lacking the *cps5* locus had significantly lower titers than the wild-type strain, deletion of the *cps5* locus in combination with other loci did not result in significantly lower titers. For the other Cluster 1 phage (ARB47), deletion of specific CPS synthesis loci likewise had no significant effect on phage titer. However, both the acapsular and CPS8-expressing ("cps8") strains produced higher titers (as expected based on our initial assay), but plaques on these strains were visibly clearer than on all other strains tested, indicating more replication on these two strains than on any of the individual CPS locus deletion strains. These data suggest that another surface receptor likely exists, and that (except in the potential case of ARB47/CPS8) CPS does not act as a co-receptor for these phages.

We also tested 5 phages from Cluster 3 that exhibited different host ranges (ARB72, ARB78, ARB82, ARB101, and ARB105). Each of these phages formed plaques best on either the cps1 or cps3 strains, with clearer plaques on these strains than on the wild-type strain (Figure 3.2). The ability of these phages to infect the acapsular strain was consistent with our previous host range assay, though low numbers of plaques, close to the limit of detection, were counted in a small number of replicates for ARB78 and ARB82 (1/6 and 2/6 replicates, respectively; see *Methods* on "Host range assays"). Contrary to the data from the Cluster 1 phages, deletion of single CPS loci did reduce (but did not eliminate) phage infection by Cluster 3 phages.

Strikingly, deletion of two or more CPS loci resulted in complete loss of titer in most cases, with

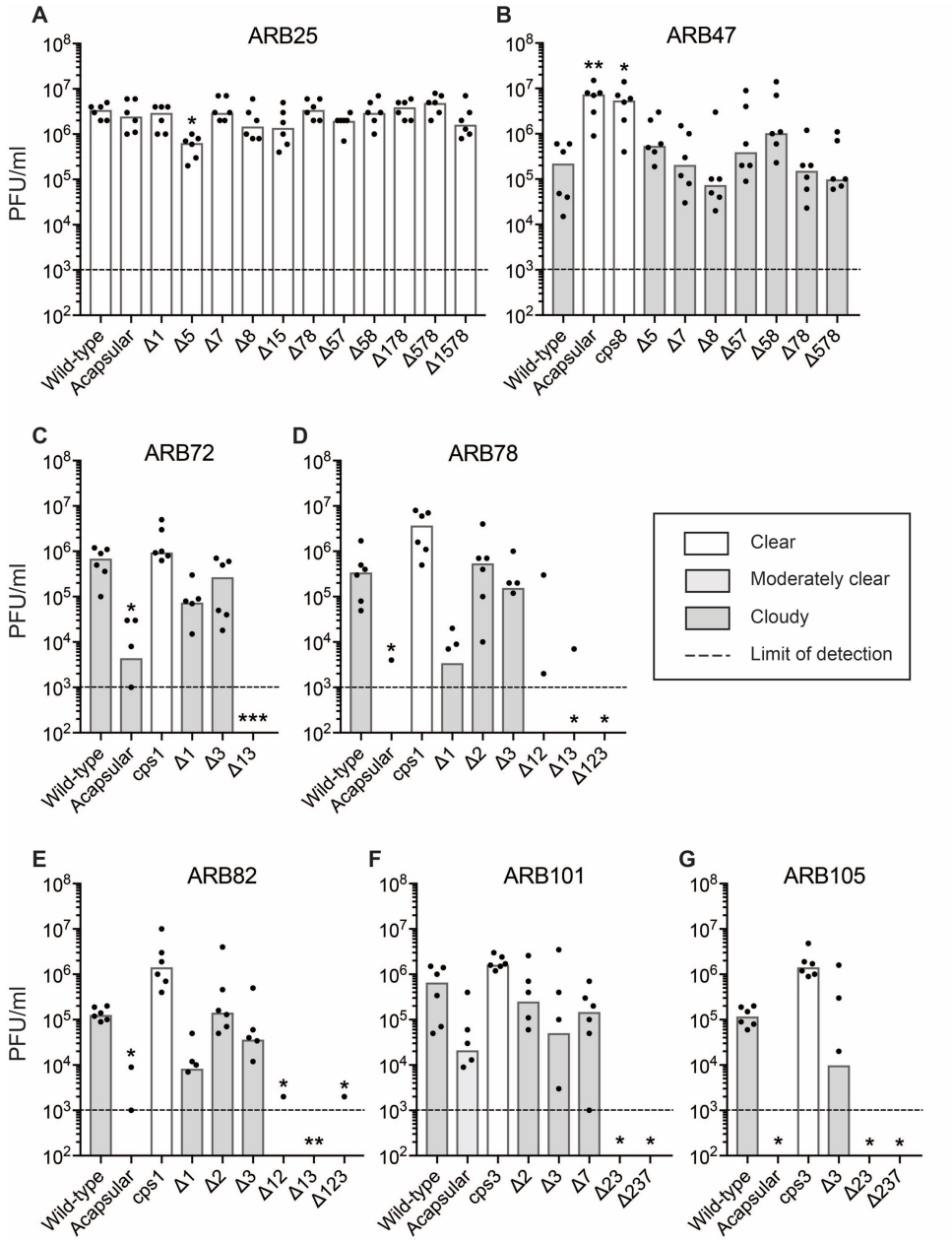


Figure 3.2 Titers of Cluster 1 and Cluster 3 phages on strains with specific CPS locus deletions implicate different mechanisms for phage-CPS interactions.

*7 phages were titered on the wild-type strain, the acapsular strain, their respective preferred host strain, and a set of bacterial strains harboring limited CPS locus deletions that correspond to their host range in Figure 3.1 (n = 6 replicates/phage). "Δ" followed by a number(s) refers to specific CPS synthesis locus deletions (e.g. "Δ15" indicates the cps1 and cps5 synthesis loci were deleted). Phages infecting the wild-type strain were selected from Cluster 1: A) ARB25, B) ARB47; and from Cluster 3: C) ARB72, D) ARB78, E) ARB82, F) ARB101, and G) ARB105. The cloudiness of plaques for each phage was qualitatively assessed and compared to that of the preferred host strain. Bars for strains generating plaques as clear as those on the preferred host strain were colored white, whereas bars for strains generating more cloudy plaques were colored gray as indicated. Significant differences in phage titers compared to titers on the wild-type strain were calculated via Kruskal Wallis test followed by Dunn's multiple comparisons test. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. PFU/ml: Plaque forming units/ml lysate.*

most other replicate titers close to the limit of detection (and potentially false positive counts). These data support the hypothesis that specific capsules (CPS1, CPS2, CPS3 and/or CPS7) are necessary or at least optimal for efficient infection by these Cluster 3 phages.

Given that CPS type is correlated with resistance to phage attack (e.g. ARB25 fails to infect strains with CPS2 or CPS3, Figure 3.1), we wondered whether the wild-type bacterial strain (which can express 8 different CPS types) would be relatively resistant to phage attack. We treated wild-type *B. thetaiotaomicron* with one of four phages (ARB25, ARB47, ARB72, and ARB105) at either a low (Figure 3.3) or high (Figure 3.4) multiplicity of infection (MOI) and monitored bacterial growth by measuring optical density at 600 nm (OD_{600}). Though they initially grew similarly to control cultures treated with heat-killed phage, OD_{600} for cells treated with a low MOI of Cluster 1 phages (ARB25 or ARB47) declined within 3-5 hours after infection (Figure 3.3). Interestingly, OD_{600} began to increase again within 1-hour post decline, indicating a large proportion of the population had remained resistant to phage infection. Interestingly, bacterial cultures originating from different clones displayed somewhat variable resistance to these phages with the growth of one clone barely delayed by live phage treatment. This clone reached our growth endpoint (> 0.6 OD_{600}) at least 1.5 hours before the other two clones. We questioned whether this reduction in cell number resulted in a relative change in *cps* gene expression. Similar to wild-type cells grown in BPRM, the heat-killed phage treatment groups exhibited predominantly *cps3* and *cps4* locus expression. However, treatment with either live ARB25 or ARB47 resulted in a dramatic loss of *cps1* and *cps4* expression, with the phage-treated cells expressing mainly *cps3*. Similar growth and expression phenotypes occurred in the cells treated with a high MOI of ARB25 or ARB47 (Figure 3.4), with the only difference being a quicker loss of OD_{600} when a high MOI was used. Dirichlet regression (see *Methods*) confirmed

that *cps* expression changes for *cps1*, *cps3*, and *cps4* loci were significant ($p < 0.01$ in all cases) for both phages at both low and high MOI. Notably, the most resistant of the bacterial clones in each of the two replicates (low and high MOI) exhibited similar *cps* expression to the other

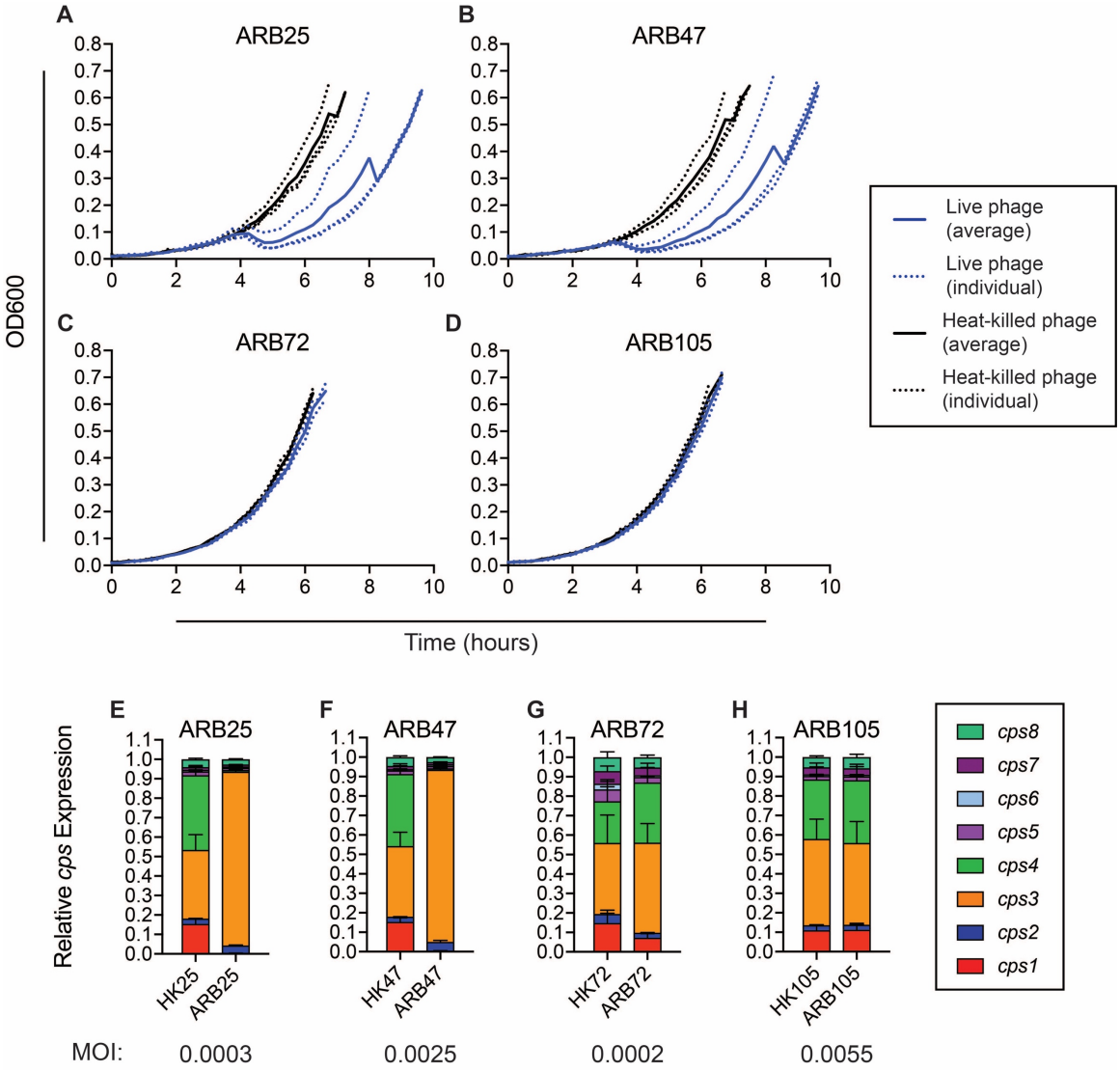


Figure 3.3 Treatment of wild-type B. thetaiotaomicron *with a low concentration of phage variably modulates CPS locus expression.*

5 ml cultures of wild-type B. thetaiotaomicron *(n = 3 clones/condition) at approximately optical density at 600 nanometers (OD600) of 0.01 were treated with either live or heat-killed phage at low multiplicity of infection (MOI). Phages assayed were from Cluster 1: A/E) ARB25; B/F) ARB47; or from Cluster 3: C/G) ARB72; D/H) ARB105. A-D) OD600 was monitored over time for each culture until the culture reached OD600 greater than 0.6. Solid lines indicate the average of 3 replicates for each condition, whereas dotted lines depict growth of individual replicates. E-H) RNA was isolated from cultures at the end of the growth experiment, and quantitative PCR (qPCR) to cDNA was used to measure relative* cps *gene expression in samples treated with live or heat-killed phage. MOI for each live phage was calculated (# phage / # bacteria in each tube) and is indicated below each panel. HK: Heat-killed phage.*

clones after treatment with live phage but expressed lower levels of *cps1* than the other clones in heat-killed phage treatment groups. This lower expression of one CPS allowing phage infection in our host range assays (Figure 3.1) may be the reason for these clones' marginally increased resistance.

In contrast to the Cluster 1 phages, phages from Cluster 3 (ARB72 and ARB105) did not demonstrate a detectable difference in growth compared to heat-killed controls at either low or high MOI conditions. This could in part be due to the poor ability of these phages to infect even their preferred host strain—yielding tiny, cloudy plaques compared to the Cluster 1 phages examined. These phages likely are only able to infect a proportion of the cells in the population even when expressing a suitable CPS type and thus may only be able to exert an influence on a small proportion of the population, especially at a low MOI. Consistent with this explanation, no significant difference in *cps* expression could be detected in the low or high MOI conditions. However, treatment with a high MOI qualitatively reduced *cps1* expression in the ARB72 treatment (from ~13% to 6% relative expression) and expression of *cps3* in the ARB105 treatment (from ~38% to 26% relative expression), consistent with the phages' demonstrated host range phenotypes (Figure 3.1). Taken together, these data indicate that *Bacteroides*-targeting phage can modulate host *cps* expression even while only infecting a subset of the population.

Discussion

Work from our lab and from others has implicated specific Bacteroides CPS in protecting the bacterium from the host immune response or even in modulating the host immune response[9,13,39]. Still, other biological and possibly orthogonal roles of the majority of *Bacteroides*

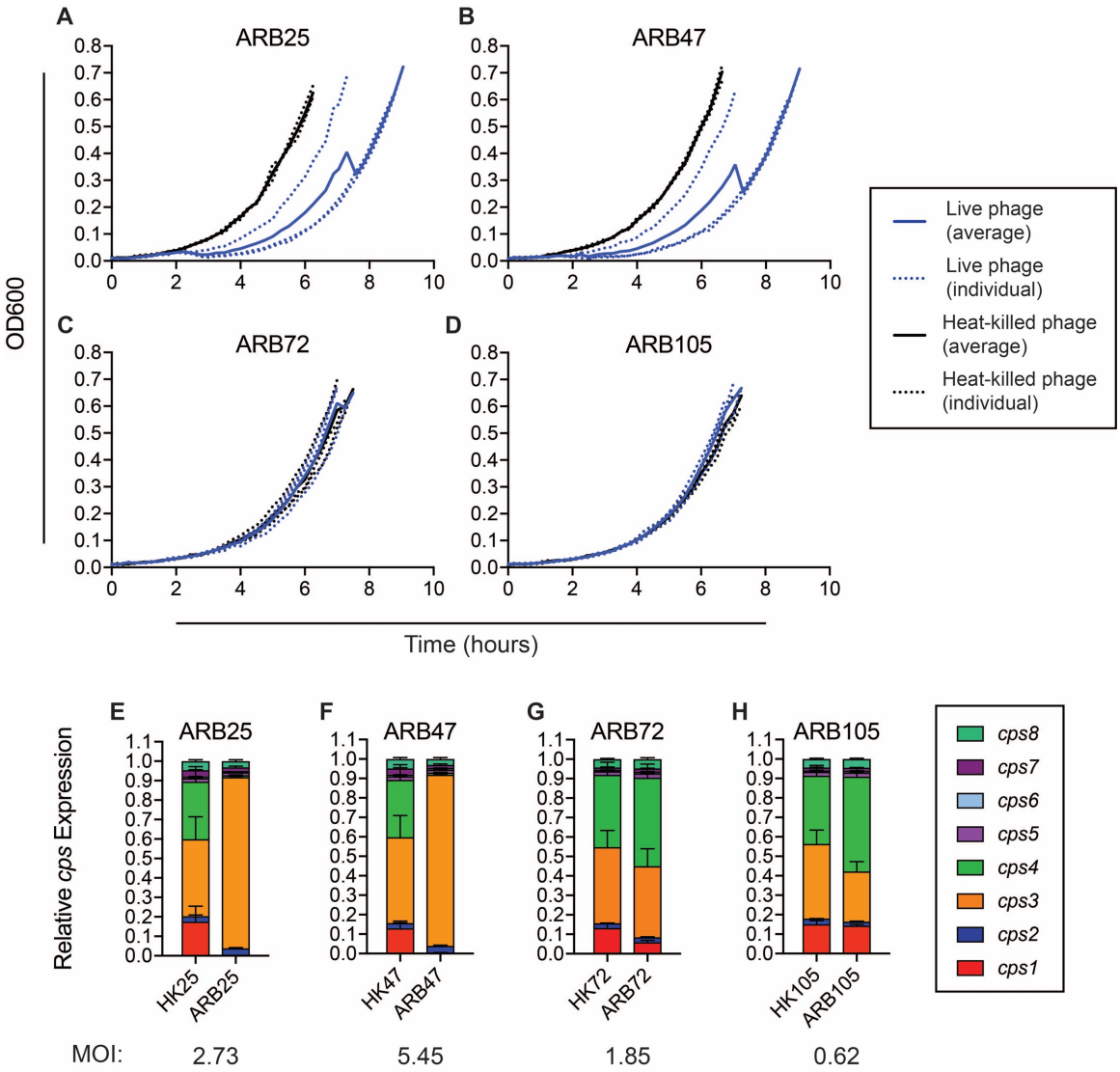


Figure 3.4 Treatment of wild-type B. thetaiotaomicron *with a high concentration of phage modulates CPS locus expression.*

5 ml cultures of wild-type B. thetaiotaomicron *(n = 3 clones/condition) at approximately OD600 of 0.01 were treated with either live or heat-killed phage at high MOI. Phages assayed were from Cluster 1: A/E) ARB25; B/F) ARB47; or from Cluster 3: C/G) ARB72; D/H) ARB105. A-D) OD600 was monitored over time for each culture until the culture reached OD600 greater than 0.6. Solid lines indicate the average of 3 replicates for each condition, whereas dotted lines depict growth of individual replicates. E-H) RNA was isolated from cultures at the end of the growth experiment, and qPCR to cDNA was used to measure relative* cps *gene expression in samples treated with live or heat-killed phage. MOI for each live phage was calculated (# phage / # bacteria in each tube) and is indicated below each panel. HK: Heat-killed phage.*

CPS are unexplored. Our collection of *B. thetaiotaomicron* strains expressing individual CPS allows us to test a previously inaccessible hypothesis: that *Bacteroides*-targeting phage populations are both blocked and/or aided by the many capsules expressed by their host. Our

data indicate that expression of specific CPS is linked to phage susceptibility. There may be multiple mechanisms by which CPS could promote or prevent phage infection. First, CPS may mask other surface receptors from phage binding. Under this model, only certain CPS would block a particular phage; phage may bind to other CPS as a co-receptor, or may specifically degrade these CPS with depolymerases. This may be the case for many of the Cluster 1 phages, as both ARB25 and ARB47 infect the acapsular strain and a limited subset of single CPS-expressing strains. Second, CPS itself may act as a phage receptor (which may be the case for Cluster 3). These mechanisms need not be exclusive of each other, and different phages likely use different mechanisms. Future experiments will investigate the ability of specific phages to bind to or degrade pure capsule. While we have failed to detect degradation of purified CPS *in vitro* using a limited panel of phages (data not shown), employing more sensitive assays and testing a wider set of phages may identify some phages with the capacity to degrade CPS structures.

To date, very few *Bacteroides* phage have been sequenced[22]. If these enzymes exist, genome sequencing of a large collection of phages with known CPS-influenced host ranges will allow us to identify the genes targeting these CPS and determine whether phage encode for carbohydrate-binding proteins or depolymerases. Comparative genomics of these phage genes may also enable us to identify the motifs necessary for binding to specific capsules, whether phages modify the same proteins to bind diverse CPS structures, or whether convergent evolution has led to multiple strategies for circumvention of the CPS layer to infect the cell. Our new collection of over 50 phages targeting the genetically tractable type strain of *B. thetaiotaomicron* will also allow us to further determine the strategies by which both bacterium and phage thwart the success of the other. A greater knowledge of these bacterium-virus

interactions will increase our understanding of how mutualistic bacteria survive in the gut, and potentially how we can modulate them to increase or maintain human health.

Methods

Strains and media conditions. The bacterial strains used in this study are listed in Table 3.2. Bacteria were routinely cultured in *Bacteroides* Phage Recovery Medium (BPRM)[40]: per 1 liter of broth, 10 g meat peptone, 10 g casein peptone, 2 g yeast extract, 5 g NaCl, 0.5 g L-cysteine monohydrate, 1.8 g glucose, 0.12 g $MgSO_4$ heptahydrate, and 1 ml of a 0.05 g/ml $CaCl_2$ solution were added; after autoclaving and cooling to approximately 55 °C, 10 ml hemin solution (0.1% w/v in 0.02% NaOH) and 25 ml 1 M Na_2CO_3 solution were added. Bacteria and phage were routinely cultured in an anaerobic chamber or in anaerobic jars at 37 °C.

Table 3.2 Bacterial strains and plasmids used in this study

Strain	Genotype	Features	Reference
Strains used in Figures 3.1, 3.3, 3.4			
Wild-type	*tdk*$^-$ ::pNBU2-bla-tetQb Tag1	Parent strain of all other derivatives with Tag1 barcode vector	Martens, Chiang, & Gordon, 2008
cps1	*tdk*$^-$ Δ*cps*2-*8 cps1*-lock ::pNBU2-bla-tetQb Tag3	CPS1-expressing strain with Tag3 barcode vector	Hickey et al., 2015
cps2	*tdk*$^-$ Δ*cps1* Δ*cps*3-*8* ::pNBU2-bla-tetQb Tag11	CPS2-expressing strain with Tag11 barcode vector	Hickey et al., 2015
cps3	*tdk*$^-$ Δ*cps1-2* Δ*cps4-8 cps3*-lock ::pNBU2-bla-tetQb Tag4	CPS3-expressing strain with Tag4 barcode vector	Hickey et al., 2015
cps4	*tdk*$^-$ Δ*cps1-3* Δ*cps5-8* ::pNBU2-bla-tetQb Tag14	CPS4-expressing strain with Tag14 barcode vector	Hickey et al., 2015
cps5	*tdk*$^-$ Δ*cps1-4* Δ*cps6-8 cps5*-lock ::pNBU2-bla-tetQb Tag6	CPS5-expressing strain with Tag6 barcode vector	Hickey et al., 2015
cps6	*tdk*$^-$ Δ*cps1-5* Δ*cps7-8 cps6*-lock ::pNBU2-bla-tetQb Tag8	CPS6-expressing strain with Tag8 barcode vector	Hickey et al., 2015
cps7	*tdk*$^-$ Δ*cps1-6* Δ*cps8* ::pNBU2-bla-tetQb Tag16	CPS7-expressing strain with Tag16 barcode vector	Hickey et al., 2015
cps8	*tdk*$^-$ Δ*cps1-7 cps8*-lock ::pNBU2-bla-tetQb Tag9	CPS8-expressing strain with Tag9 barcode vector	Hickey et al., 2015
Acapsular	*tdk*$^-$ Δ*cps1-8* ::pNBU2-bla-tetQb Tag18	Strain lacking all CPS synthesis loci with Tag18 barcode vector	Hickey et al., 2015
Strains used in Figure 3.2		**Used in Figure 3.2**	
Wild-type	*tdk*$^-$	Parent strain of all other derivatives	Koropatkin et al. 2008

cps1	*tdk*⁻ Δ*cps2-8*	Strain expressing CPS1 locus	Porter et al. 2017
cps3	*tdk*⁻ Δ*cps1*-2, Δ*cps4-8*	Strain expressing CPS3 locus	Porter et al. 2017
cps8	*tdk*⁻ Δ*cps1-7*	Strain expressing CPS8 locus	Porter et al. 2017
Acapsular	*tdk*⁻ Δ*cps1-8*	Strain lacking all CPS synthesis loci	Rogers et al. 2013
Δcps1	*tdk*⁻ Δ*cps1*	Strain lacking only *cps1* locus	This study
Δcps2	*tdk*⁻ Δ*cps2*	Strain lacking only *cps2* locus	This study
Δcps3	*tdk*⁻ Δ*cps3*	Strain lacking only *cps3* locus	This study
Δcps5	*tdk*⁻ Δ*cps5*	Strain lacking only *cps5* locus	This study
Δcps7	*tdk*⁻ Δ*cps7*	Strain lacking only *cps7* locus	This study
Δcps8	*tdk*⁻ Δ*cps8*	Strain lacking only *cps8* locus	This study
Δcps1/5	*tdk*⁻ Δ*cps1*, Δ*cps5*	Strain lacking *cps1* and *cps5* loci	This study
Δcps7/8	*tdk*⁻ Δ*cps7-8*	Strain lacking *cps7* and *cps8* loci	This study
Δcps5/7	*tdk*⁻ Δ*cps5*, Δ*cps7*	Strain lacking *cps5* and *cps7* loci	This study
Δcps5/8	*tdk*⁻ Δ*cps5*, Δ*cps8*	Strain lacking *cps5* and *cps8* loci	This study
Δcps1/7/8	*tdk*⁻ Δ*cps1*, Δ*cps7-8*	Strain lacking *cps1*, *cps7*, and *cps8* loci	This study
Δcps5/7/8	*tdk*⁻ Δ*cps5*, Δ*cps7-8*	Strain lacking *cps5*, *cps7*, and *cps8* loci	This study
Δcps1/5/7/8	*tdk*⁻ Δ*cps1*, Δ*cps5*, Δ*cps7-8*	Strain lacking *cps1*, *cps5*, *cps7*, and *cps8* loci	This study
Δcps1/2	*tdk*⁻ Δ*cps1-2*	Strain lacking *cps1* and *cps2* loci	This study
Δcps1/3	*tdk*⁻ Δ*cps1*, Δ*cps3*	Strain lacking *cps1* and *cps3* loci	This study
Δcps2/3	*tdk*⁻ Δ*cps2-3*	Strain lacking *cps2* and *cps3* loci	This study
Δcps1/2/3	*tdk*⁻ Δ*cps1-3*	Strain lacking *cps1*, *cps2,* and *cps3* loci	This study
Δcps2/3/7	*tdk*⁻ Δ*cps2-3*, Δ*cps7*	Strain lacking *cps2*, *cps3*, and *cps7* loci	This study
Plasmids used	**In *Escherichia coli* S17-1 λpir**	**Used for generation of bacterial knockouts**	
pExchange Δ*cps1*		Deletion of *cps1* synthesis locus	Rogers et al. 2013
pExchange Δ*cps2*		Deletion of *cps2* synthesis locus	Rogers et al. 2013
pExchange Δ*cps3*		Deletion of *cps3* synthesis locus	Rogers et al. 2013
pExchange Δ*cps4*		Deletion of *cps4* synthesis locus	Rogers et al. 2013
pExchange Δ*cps5*		Deletion of *cps5* synthesis locus	Rogers et al. 2013
pExchange Δ*cps6*		Deletion of *cps6* synthesis locus	Rogers et al. 2013
pExchange Δ*cps7*		Deletion of *cps7* synthesis locus	Rogers et al. 2013
pExchange Δ*cps8*		Deletion of *cps8* synthesis locus	Rogers et al. 2013

Table 3.3 Primers used in this study

Primer	Sequence	Use
BT0396 F	GCGTGGTGGTGGCATACTCTTCT	*cps1* locus expression
BT0396 R	ACTTACGCCTGCCACCAATGTTAG	*cps1* locus expression
BT0463 F	CTATTTCGGTATTGATGTTGGCTGGTA	*cps2* locus expression
BT0463 R	TCTCCGATAATAAATGCTTGGGCTAAT	*cps2* locus expression
BT0612 F	CCATAAAGGGGCTGAGCATTGTTC	*cps3* locus expression
BT0612 R	AGTCATCATAGGTAGCGGCTTGTAGTG	*cps3* locus expression
BT1339 F	CGGTTAGGAGGAGTTTCGTTTTTC	*cps4* locus expression
BT1339 R	CAAGATCATCCGCAATACCTGTTA	*cps4* locus expression
BT1651 F	GTAGCATTAGGAACCAGCCCACTTTA	*cps5* locus expression
BT1651 R	ACGACGTTTTATATTCCCGACAA	*cps5* locus expression
BT1714 F	AATTCTTTTGGGCGTACTTATGGTGA	*cps6* locus expression
BT1714 R	CAACTTTTAGCCTATCCGGTGTGAAC	*cps6* locus expression
BT2872 F	CCAACTACCGGCCTGCTGAA	*cps7* locus expression
BT2872 R	TCGGGCTAATCGTAACACCATCTT	*cps7* locus expression
BT0066 F	TGCCTCCGCCAACCTCGTCACCTTC	*cps8* locus expression
BT0066 R	GTTCGTTGCCCTTGCTGCTACCCGTTCC	*cps8* locus expression

Bacteriophage strains were isolated using the double agar layer plaque assay[40] under multiple conditions from two locations at the Ann Arbor, Michigan Wastewater Treatment Plant (approximate GPS coordinates 42.270387, -83.663741) (Table 3.1): either with or without an enrichment step prior to plating; and with or without bile in the top agar. Primary effluent was filtered through a 0.22 μm PVDF filter. Up to 1 ml filtered wastewater or wastewater concentrated via a 100 kDa size exclusion column was added to each 0.5 ml overnight culture and 4.5 ml BPRM top agar (0.35% agar), and the mixture was poured on top of a BPRM agar plate. Plates were incubated anaerobically overnight. Single, isolated plaques were picked at least 3 times for each purified bacteriophage strain. To promote a diverse collection of phages, no more than 5 plaques from the same plate were included if they exhibited similar host range phenotypes (see below). When using individual enrichment cultures, only a single plaque was included in the collection. High titer phage stocks were generated by adding approximately enough phage to the culture/top agar mixture to generate a "lacey" pattern of bacteria on the plate. Following overnight incubation of each plate, 5 ml of sterile phage buffer (5 ml of 1 M Tris pH 7.5, 5 ml of 1 M MgSO4, 2 g NaCl, and filled to a total of 500 ml with ddH_2O) was added to the plate to resuspend the phage. The lysate was spun at 5,500 rcf for 10 minutes to clear debris and then filter sterilized through a 0.22 μm PVDF filter. Where indicated, phage were heat killed by heating to 95 °C for 30 minutes, and heat-killed phage had no detectable plaque forming units (PFU)/ml with a limit of detection of 100 PFU/ml.

Host range assays. For host range assays in Figures 3.1 and 3.2, bacterial strains indicated in each figure were streaked onto agar plates of Brain Heart Infusion supplemented with 10% horse blood (Quad Five, Rygate, Montana) and grown anaerobically for 2-3 days. A single colony was

picked for each bacterial strain, inoculated into 5 ml BPRM, and grown anaerobically overnight. 0.5 ml of bacterial cultures were added to 4.5 ml BPRM top agar, poured on top of a BPRM agar plate, and allowed to solidify. 1 µl of 1:10 serial dilutions of phage stocks (approximately 1E6 PFU/ml to 1E3 PFU/ml) were spotted onto strains indicated in each figure. After spots dried, plates were incubated anaerobically for 15-24 hours prior to counting plaques. Large, clear plaques (e.g. those generated by ARB25) were typically easier to count than for phage generating very small (<1 mm) plaques (e.g. ARB72) even on their preferred host strain (with some phages also generating very cloudy plaques on some strains). In order to not inappropriately deem a host strain "resistant" to phage infection, we attempted to be uniformly liberal in counting plaques for these phages. Thus, low phage titers near the limit of detection may in some cases actually reflect a complete resistance of the host to the corresponding phage.

Phage treatment to detect modulation of cps *gene expression.* For each experiment in Figure 3.3 and Figure 3.4, 3 individual colonies of wild-type *B. theta* were picked from agar plates and grown overnight in BPRM. Each culture was then diluted to an optical density at 600 nm (OD600) of approximately 0.01. 200 µl of heat-killed phage or live phage were added to the diluted cultures, and bacterial growth was monitored by measuring OD600 every 15-30 minutes using a GENESYS 20 spectrophotometer (Thermo Scientific). For determination of relative *cps* gene expression, once a culture surpassed OD600 0.6, it was centrifuged at 7700 rcf for 2.5 minutes, the supernatant was decanted, and the pellet was immediately resuspended in 1 ml RNA-Protect (Qiagen) and then treated according to the manufacturer's instructions. All cultures were taken between OD600 0.6-0.8. RNA-stabilized cell pellets were stored at -80 °C until use.

RNA extraction and quantitative PCR (qPCR) for determination of relative cps gene expression. Relative *cps* synthesis locus expression was assayed similar to our previous work[13]. Briefly, crude RNA was isolated using the RNeasy Mini Kit (Qiagen) then treated with the TURBO DNA-free Kit (Ambion) followed by an additional isolation using the RNeasy Mini Kit. cDNA was then synthesized using SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions using random oligonucleotide primers (Invitrogen). qPCR analyses for *cps* locus expression were performed on a Mastercycler ep realplex machine (Eppendorf). Expression of each of the 8 *cps* synthesis loci was quantified using primers to a single gene in each locus (primers are listed in Table 3.3) and normalized to a standard curve of DNA from wild-type *B. theta*. Relative abundance of expression for each locus was then calculated. A custom-made SYBR-based master mix was used for qPCR: 20 µl reactions were made with ThermoPol buffer (New England Biolabs), and containing 2.5 mM $MgSO_4$, 0.125 mM dNTPs, 0.25 µM each primer, 0.1 µl of a 100 X stock of SYBR Green I (Lonza), and 500 U Hot Start *Taq* DNA Polymerase (New England Biolabs). 10 ng of cDNA was used for each sample, and samples were run in duplicate. A touchdown protocol with the following cycling conditions was used for all assays: 95 °C for 3 minutes, followed by 40 cycles of 3 seconds at 95 °C, 20 seconds of annealing at a variable temperature, and 20 seconds at 68 °C. The annealing temperature for the first cycle was 58 °C, then dropped one degree each cycle for the subsequent 5 cycles. The annealing temperature for the last 34 cycles was 52 °C. These cycling conditions were followed by a melting curve analysis to determine amplicon purity.

Data representation and statistical analysis. The heatmap and dendrogram for Figure 3.1 were generated in R using the "heatmap" function. Other graphs were created in Prism software

(GraphPad Software, Inc., La Jolla, CA). Statistical significance in this work is denoted as follows unless otherwise indicated: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Statistical analyses other than Dirichlet regression were performed in Prism. Dirichlet regression was performed in R using the package "DirichletReg" (version 0.6-3), employing the alternative parameterization as used previously[13,41]. Briefly, the parameters in this distribution are the proportions of relative *cps* gene expression and the total *cps* expression, with *cps7* expression used as a reference. The variable of interest used is phage viability (live versus heat-killed phage). Precision was allowed to vary by group given this model was superior to a model with constant precision, as determined by a likelihood ratio test at significance level $p < 0.05$.

References

1. Eckburg, P. B. *et al.* Diversity of the Human Intestinal Microbial Flora. *Science* **308,** 1635–8 (2005).
2. Cockburn, D. W. & Koropatkin, N. M. Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. *J. Mol. Biol.* **428,** 3230–52 (2016).
3. Hooper, L. V, Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336,** 1268–73 (2012).
4. Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the human body. *Nature* **509,** 357–360 (2014).
5. El Kaoutari, A. *et al.* The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11,** 497–504 (2013).
6. Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517,** 165–169 (2015).
7. Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506,** 498–502 (2014).
8. 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,** 447–57 (2008).
9. 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).
10. Round, J. L. *et al.* The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332,** 974–7 (2011).
11. Shen, Y. *et al.* Outer membrane vesicles of a human commensal mediate immune regulation and disease protection. *Cell Host Microbe* **12,** 509–20 (2012).
12. Neff, C. P. *et al.* Diverse intestinal bacteria contain putative zwitterionic capsular polysaccharides with anti-inflammatory properties. *Cell Host Microbe* **20,** 1–13 (2016).
13. Porter, N. T., Canales, P., Peterson, D. A. & Martens, E. C. A Subset of Polysaccharide Capsules in the Human Symbiont Bacteroides thetaiotaomicron Promote Increased Competitive Fitness in the Mouse Gut. *Cell Host Microbe* **22,** 494–506 (2017).
14. Coyne, M. J. & Comstock, L. E. Niche-specific features of the intestinal Bacteroidales. *J. Bacteriol.* **190,** 736–42 (2008).
15. Patrick, S. *et al.* Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of Bacteroides fragilis. *Microbiology* **156,** 3255–69 (2010).
16. Porter, N. T. & Martens, E. C. The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu. Rev. Microbiol.* **71,** 349–369 (2017).

17. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466,** 334–8 (2010).

18. Minot, S. *et al.* The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* **21,** 1616–25 (2011).

19. Booth, S. J., Van Tassell, R. L., Johnson, J. L. & Wilkins, T. D. Bacteriophages of Bacteroides. *Rev. Infect. Dis.* **1,** 325–336 (1979).

20. Cooper, S. W., Szymczak, E. G., Jacobus, N. V. & Tally, F. P. Differentiation of Bacteroides ovatus and Bacteroides thetaiotaomicron by means of bacteriophage. *J. Clin. Microbiol.* **20,** 1122–1125 (1984).

21. Keller, R. & Traub, N. The characterization of Bacteroides fragilis bacteriophage recovered from animal sera: observations on the nature of Bacteroides phage carrier cultures. *J. Gen. Virol.* **24,** 179–89 (1974).

22. Ogilvie, L. A. & Jones, B. V. The human gut virome: A multifaceted majority. *Front. Microbiol.* **6,** 1–12 (2015).

23. Puig, A., Araujo, R., Jofre, J. & Frias-Lopez, J. Identification of cell wall proteins of Bacteroides fragilis to which bacteriophage B40-8 binds specifically. *Microbiology* **147,** 281–8 (2001).

24. Rakhuba, D. V., Kolomiets, E. I., Szwajcer Dey, E. & Novik, G. I. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Polish J. Microbiol.* **59,** 145–155 (2010).

25. Gupta, D. S. *et al.* Coliphage K5, specific for E. coli exhibiting the capsular K5 antigen. *FEMS Micro Lett.* **14,** 75–78 (1982).

26. Nimmich, W., Schmidt, G. & Krallmann-Wenzel, U. Two different Escherichia coli capsular polysaccharide depolymerases each associated with one of the coliphage φK5 and φK20. *FEMS Microbiol. Lett.* **82,** 137–141 (1991).

27. Whitfield, C. & Lam, M. Characterisation of coliphage K30, a bacteriophage specific for Escherichia coli capsular serotype K30. *FEMS Microbiol. Lett.* **37,** 351–355 (1986).

28. Sørensen, M. C. H. *et al.* Phase variable expression of capsular polysaccharide modifications allows Campylobacter jejuni to avoid bacteriophage infection in chickens. *Front. Cell. Infect. Microbiol.* **2,** 11 (2012).

29. Ainsworth, S. *et al.* Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. *MBio* **5,** e00880-14 (2014).

30. Scholl, D., Adhya, S. & Merril, C. Escherichia coli K1's capsule is a barrier to bacteriophage T7. *Appl. Environ. Microbiol.* **71,** 4872–4 (2005).

31. Sørensen, M. C. H. *et al.* Primary isolation strain determines both phage type and receptors recognised by Campylobacter jejuni bacteriophages. *PLoS One* **10,** e0116287 (2015).

32. Bayer, M. E., Thurow, H. & Bayer, M. H. Penetration of the polysaccharide capsule of Escherichia coli (Bi161/42) by bacteriophage K29. *Virology* **94,** 95–118 (1979).

33. O'Leary, T. R., Xu, Y. & Liu, J. Investigation of the substrate specificity of K5 lyase A from K5A bacteriophage. *Glycobiology* **23,** 132–141 (2013).

34. Clarke, B. R., Esumeh, F. & Roberts, I. S. Cloning, expression, and purification of the K5 capsular polysaccharide lyase (KflA) from coliphage K5A: Evidence for two distinct K5 lyase enzymes. *J. Bacteriol.* **182,** 3761–3766 (2000).

35. Scholl, D., Rogers, S., Adhya, S. & Merril, C. R. Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of Escherichia coli. *J. Virol.* **75,** 2509–2515 (2001).

36. Leiman, P. G. *et al.* The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. *J. Mol. Biol.* **371,** 836–849 (2007).

37. Chatzidaki-Livanis, M., Weinacht, K. G. & Comstock, L. E. Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont Bacteroides fragilis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 11976–80 (2010).

38. Coyne, M. J., Weinacht, K. G., Krinos, C. M. & Comstock, L. E. Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 10446–51 (2003).

39. Mazmanian, S. K. S. S. K. *et al.* An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,** 107–118 (2005).

40. Araujo, R. *et al.* Optimisation and standardisation of a method for detecting and enumerating bacteriophages infecting Bacteroides fragilis. *J. Virol. Methods* **93,** 127–136 (2001).

41. Maier, M. J. DirichletReg : Dirichlet Regression for Compositional Data in R. 13 (2014).

CHAPTER 4

Discussion

Introduction

Members of the *Bacteroides* compose one of the most prevalent genera in the human gut microbiome. Individual strains or species in this group dedicate a substantial portion of their genomic potential to carbohydrate-active enzymes[1]. The model organism *Bacteroides thetaiotaomicron* (*B. theta*) dedicates approximately 18% of its genome to loci enabling the breakdown of a diverse array of dietary and host-derived carbohydrates[2], an indication of the importance of glycan degradation to the fitness of this species. *B. theta* invests an additional 3% of its genome in encoding eight loci for the synthesis of distinct capsular polysaccharides (CPS). The importance of these loci is exemplified by both the sheer number of CPS synthetic loci in individual strains of this single species[3] and in other members of the genus[4], as well as the apparent diversity in CPS types within and among various species of this genus[3,5]. However, besides a few examples related to interaction with the host immune response[6–8], little is known about the roles of individual CPS structures in these symbiotic organisms. My work in this dissertation has both expanded the knowledge of the function of some of these CPS structures, and also provides key tools for future studies into the role of CPS in both immune evasion and bacteriophage resistance.

Chapter Summary

Much of the work described in this dissertation focuses on a key set of isogenic mutants generated from the type strain of *B. theta* VPI-5482. The wild-type *B. theta* strain encodes and expresses 8 different CPS. Multiple layers of regulation control the expression of individual CPS synthesis loci. First, invertible promoter regions alternate between "on" (allows expression of CPS synthesis genes) and "off" (prohibits CPS gene expression) orientations in five of the eight CPS loci (Figure 2.14). Second, expression of transcriptional anti-termination genes is required in *Bacteroides fragilis* for proper transcriptional read-through of each CPS synthesis locus[9], and homologous genes are found in each of the eight loci in *B. theta*. Third, many of the CPS loci in both *B. fragilis* (8/8) and *B. theta* (5/8) encode *trans*-locus repressors that inhibit transcriptional read-through of a subset of other CPS synthesis loci in the genome, but not the cognate locus[10]. Additional regulation via an unknown mechanism(s) may occur, such as during the sensing of specific host glycans, as mutational activation of some glycan degradation systems in *B. theta* was correlated with altered CPS gene expression[11].

The complex regulation involved in governing the expression of multiple CPS loci is intriguing and may very well serve the evolutionary advantage of pre-adapting some individuals within every population with potentially advantageous survival traits. However, it confounds the contributions of individual CPS to facilitate bacterial interactions with the environment, such as glycan acquisition, host immune interactions, and bacteriophage resistance. Thus, we created a panel of isogenic *B. theta* strains that each encodes for and expresses just one of the eight CPS from the wild-type strain[12]; additionally, we created an acapsular mutant lacking all CPS synthesis loci[13]. We extensively characterized these strains to discount the possibility that secondary mutations could confound our analyses. Genome sequencing of each strain revealed

very few single nucleotide polymorphisms among the strains other than CPS locus deletions (Table 2.1). Moreover, whereas the wild-type strain expresses various amounts of each CPS synthesis locus, each single CPS-expressing strain only expressed its associated locus (Figure 2.14). While the structures of each of these CPS are not known, strains exhibited distinct profiles of monosaccharides when grown on minimal medium with only glucose as the carbon source (Figure 2.13). The acapsular strain could not express any of the CPS loci and routinely settled to the bottom of the culture during overnight growth, similar to what has previously been seen for acapsular derivatives of *Bacteroides* strains[10,14,15]. This confirmatory data allowed us to confidently attribute changes in phenotype to expression of specific CPS in our subsequent assays.

With our validated strains in hand (the wild-type, acapsular, and eight single CPS-expressing strains; each barcoded with a unique sequence tag to enable identification in mixed populations), we were prepared to evaluate their respective contributions to bacterial fitness *in vitro* and *in vivo*. In Chapter 2 we tested the contribution of CPS on gut fitness by pooling together the eight single CPS-expressing strains with the acapsular strain and competing them against each other in germ-free mice. A subset of the strains (those expressing CPS4, CPS5, and CPS6) typically outcompeted all other strains, with the strain expressing CPS5 typically being most dominant (see Figure 2.4). Changes in host diet had little effect on the relative abundance of these three dominant strains, with a low-fiber diet providing a marginal fitness advantage to the CPS6-expressing strain (Figure 2.5). In contrast, altering host immunity significantly altered the relative abundance of the CPS-expressing strains. Specifically, increasing levels of host fecal IgA were correlated with more stringent competition among the strains. Whereas many CPS-expressing strains persisted for significantly longer in the gut in adaptive immune deficient Rag1^{-}

$^{-/-}$ mice (which produce no IgA), the CPS5-expressing strain outcompeted all others in MyD88$^{-/-}$/Trif$^{-/-}$ mice, which were confirmed to produce higher levels of IgA than wild-type mice (Figure 2.3, Figure 2.6). Both total IgA levels, as well as IgA specifically targeting CPS, were increased in MyD88$^{-/-}$/Trif$^{-/-}$ mice, and CPS5 was the only capsule positively associated with increasing IgA levels (Figure 2.6). Additionally, development of anti-CPS5 IgA responses was significantly slower than responses against several other, less abundant CPS types (Figure 2.6). These data indicate that CPS5 allows the bacterium to evade host adaptive immune responses better than other CPS.

Interestingly, the wild-type strain (which can express eight different CPS), expresses mainly its *cps3* and *cps4* genes when grown *in vitro* and expresses mainly *cps4* synthesis genes *in vivo* on standard plant-rich diets (Figure 2.1). As CPS5 is often only expressed at a low level yet is so advantageous when CPS expression is forced (i.e. the CPS5-expressing strain is only able to produce CPS5), we tested the hypothesis that the ability to express multiple CPS types provides an additional advantage. Competition experiments revealed that wild-type *B. theta* competes well with the CPS5-expressing strain, with a slight advantage in the presence of adaptive immunity (Figure 2.8). Surprisingly, wild-type expressed all of its CPS, and its competitive colonization was characterized by dynamic changes in CPS expression over time. Wild-type exhibited an additional advantage after antibiotic reduction of the community. Following antibiotic exposure, only the wild-type strain was detectable in most mice. Under these conditions, the wild-type strain strongly increased its relative expression of the *cps5* genes in an adaptive immune-dependent fashion. While these data clearly indicate an advantage conferred by CPS5 in the presence of adaptive immunity, it also points to increased fitness conferred by the ability to express a variety of CPS types.

The experiments in Chapter 2 provide evidence for a single CPS (CPS5) specifically aiding in host immune evasion. Still, seven other CPS types exist in this strain with no specific ascribed function. We also probed 13 additional *B. theta* genomes and found another 41 distinct CPS synthesis loci, with different combinations and types of glycosyltransferases, acetyltransferases, and epimerases (Figures 2.10, Figure 2.11). The enormous diversity in CPS within and among strains of the same species implies there are additional selective pressures other than the host immune response that select for diverse cell surface coatings for these bacteria.

To this end, in Chapter 3, we isolated a collection of bacteriophage using the 10 isogenic strains (the wild-type strain, the acapsular strain, and the eight single CPS-expressing strains). Evaluating the infectivity of each of these phages on the whole panel of bacterial strains revealed diverse host ranges among these phages (Figure 3.1). Very few phages targeted a single host. Rather, the majority of the phages infected a subset of the single CPS-expressing strains, as well as the wild-type and/or acapsular strains. Clustering of phage isolates based on host range phenotype revealed distinct infection patterns. Phages in two of these clusters, termed "Cluster 1" and "Cluster 2", typically exhibited robust infection of the acapsular strain. Thus, phages in these clusters likely target various cell surface receptors other than CPS, such as outer membrane proteins or lipopolysaccharide (LPS). Phages in the last cluster ("Cluster 3") typically exhibited poor to no infection of the acapsular strain, implicating CPS as a likely target. Subsequent assays studying strains with a limited set of CPS locus deletions revealed that phage infection by several Cluster 3 phages (but not Cluster 1 phages) was dependent on CPS (Figure 3.2). These data support distinct mechanisms of attachment and infection for members of each phage cluster.

The ability to encode multiple CPS provided an advantage in *in vivo* competition (Chapter 2). Thus, we tested the hypothesis that the wild-type strain of *B. theta* was also resistant to phage infection due to its ability to vary these surface structures. We treated wild-type with live or heat-killed phage, using phages from both Cluster 1 and Cluster 3 (Figure 3.3, Figure 3.4). Interestingly, cell growth quickly and consistently recovered from phage infection by Cluster 1 phages, with culture turbidity beginning to increase within an hour of its decline (Figure 3.3). Cells treated with Cluster 3 phages did not demonstrate a noticeable drop in cell density. Phages from both clusters appeared to induce changes in *cps* gene expression in wild-type *B. theta*, though changes were greater when treating with Cluster 1 phages. Collectively, the experiments in this dissertation demonstrate the utility of bacteria encoding and expressing multiple CPS types, including in the face of host immunity and bacteriophage challenge.

Specific roles for individual CPS

CPS in pathogens play various important roles in evading the host immune system[16–18]. While CPS from beneficial gut bacteria have been less studied, support for a similar role in immune evasion is mounting[3,7,14,19–21]. For instance, Polysaccharide A (PSA) from *B. fragilis* has been well characterized for its ability to directly act on multiple host immune cell types to suppress inflammatory responses[7,8,22–24]. Additionally, Neff et al. screened a broad array of gut bacteria for the ability to modulate anti-inflammatory responses in human cells[21]. Though only studying the presence of a single, positively charged motif, they demonstrated that the presence of this motif (which is present in several *Bacteroides* species and strains) was strongly correlated with increased levels of the anti-inflammatory cytokine IL-10. While we have not identified CPS3 as promoting bacterial fitness *in vivo* (at least in competition with other CPS-expressing

strains), it is predicted to have the same positively charged motif. Thus, in addition to CPS5, CPS3 may also play a role in modulating the host immune response.

Other CPS may also have specific, yet undetermined functions. Growth of each CPS-expressing strain on a panel of individual monosaccharides and polysaccharides did not reveal widespread changes in growth rate among the strains. However, while changing host diet did not greatly alter strain competition *in vivo* (Figure 2.5), wild-type expression of CPS6 increased when host diets were limited in fiber intake (Figure 2.1). This finding agrees with results from previous studies that observed differences in CPS expression when mice were fed diets low in plant fiber (either fiber-deficient, or fed mother's milk)[25,26]. *B. theta* preferentially uses dietary glycans before host glycans[27]. Under low-fiber conditions, *B. theta* switches to degradation of host-derived glycans[25,26]. Mutationally activating certain host glycan-degrading loci *in vitro* also resulted in distinct shifts in CPS production[11]. While we have determined that individual CPS-expressing strains do not differ in their ability to degrade host mucus-derived *O*-linked glycans (Figure 2.7), certain CPS could facilitate the acquisition of other, yet untested host-derived glycans.

Alternatively, it is possible that *B. theta* switches its CPS production upon sensing host glycans for other reasons. As *B. theta* only upregulates expression of its host glycan utilization loci when concentrations of dietary fibers are low, this could act as a signal that the bacterium is entering a new niche or is nearing the host epithelium. In this case, the bacterium would switch its CPS to become more resistant to certain host derived antimicrobial factors. We[3] and others[14] have determined that CPS type (or no CPS) affects cell susceptibility to complement (Figure 2.9). Higher host IgA concentrations in the mucus layer could also select against specific CPS. We also investigated the effect of CPS type on antimicrobial peptide resistance. While we found

no significant difference in resistance to polymyxin B or human beta-defensin 3 (Figure 2.9,

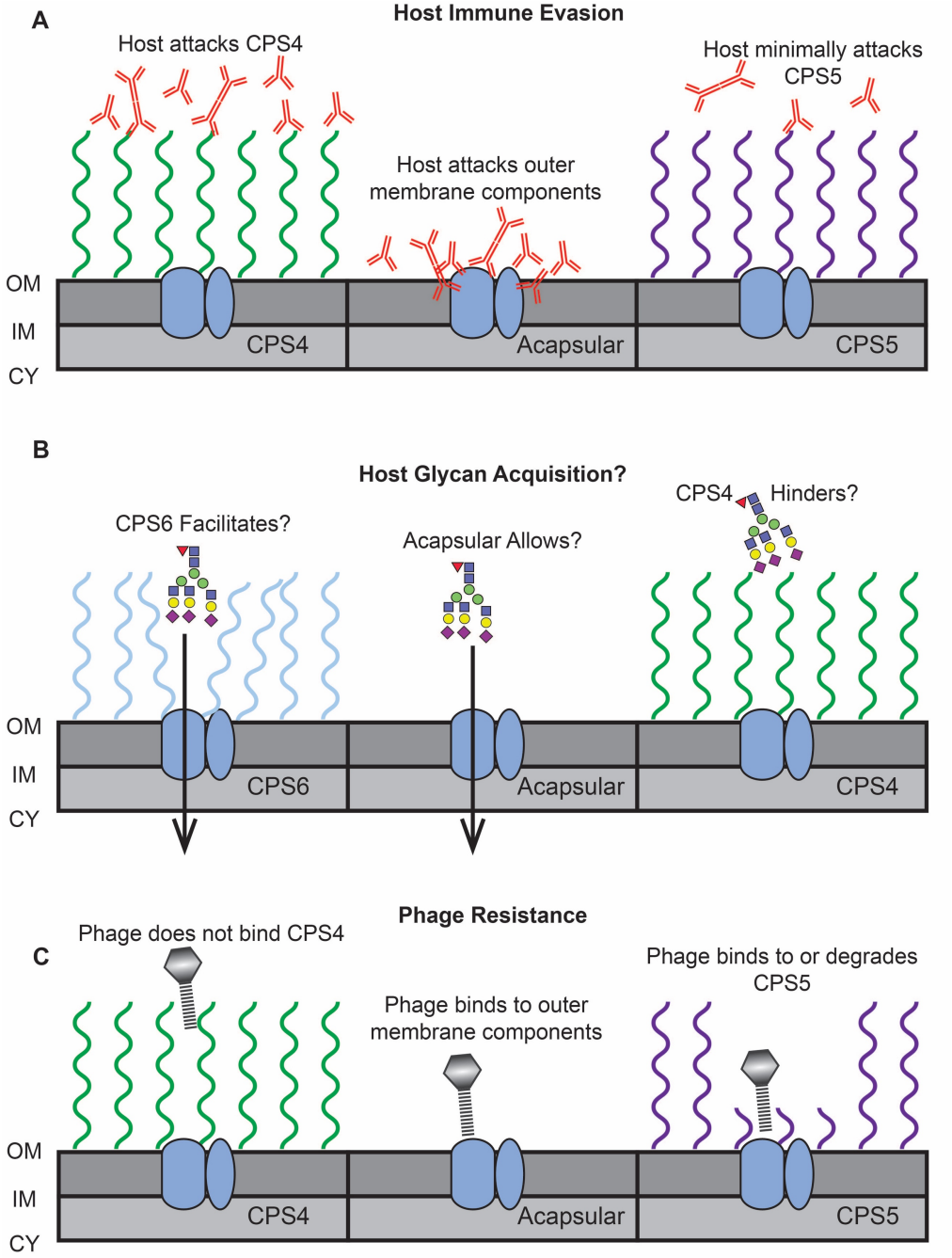


Figure 4.1 Roles of individual Bacteroides *CPS in interactions with the environment and the host immune response.*

A) Host IgA responses against CPS4 (left) develop more quickly than against CPS5 (right) (see Figure 2.3). The CPS likely protects other cell surface antigens (e.g. nutrient acquisition systems) from immune targeting and neutralization. B) Certain CPS (potentially CPS6, left) could facilitate host glycan acquisition better than other CPS types (e.g. CPS4, right). This is supported by CPS locus expression changes in low-fiber conditions (see Figure 2.1) under which B. theta *upregulates its host-derived glycan utilization genes. C) Some phages bind to cell surface receptors (potentially nutrient acquisition systems) in the absence of CPS. Only certain CPS (such as CPS4, left) are able to block phage infection. Others (e.g. CPS5, right) may passively allow phage through the CPS layer, act themselves as a receptor/co-receptor, or may be actively degraded by phage polymerases. OM: Outer membrane; IM: Inner membrane; CY: Cytoplasm.*

Table 2.3), other antimicrobial peptides could differentially target cells based on CPS type. These dietary or immune factors (or other, yet unidentified factors) could drive the changes seen in mice fed low-fiber diets. Potential roles of *Bacteroides* CPS are outlined in Figure 4.1.

Advantages of encoding and switching between multiple CPS types

The experiments in this thesis show that wild-type *B. theta* dynamically changes expression of its multiple CPS synthesis loci over time when monoassociated in germ-free mice (Figure 2.1). This occurs even in the absence of the adaptive immune response or dietary changes, indicating that these dynamic changes may be "hardwired" into the regulatory machinery for each CPS locus. While it is yet unproven, it has been hypothesized that the complex regulatory hierarchy of CPS synthesis only allows a single CPS to be produced in a single cell at a given time[10]. If this is the case, switching CPS production would likely be a very costly endeavor, requiring the cell to synthesize many new enzymes as well as a new CPS with each switch. Such a cost to the cell should require the benefits to also be high.

Such benefits may come to the cell in the form of resilience, which, similar to Sommer and colleagues[28], we define as the ability for the population to recover or rebound following perturbation (an outside selective force on the population). *B. theta* maintains a population with phase-variable CPS production. Under "normal" (no perturbation) conditions, the relative proportions of CPS expression in a population of cells remain comparatively stable. However, as a selective pressure is applied to the population that differentially affects cells with distinct CPS types, the relative expression of CPS changes. Once the perturbation has ended and "normal" conditions return, relative CPS expression may gradually shift back to be similar to the original population (Figure 4.1). The various features of the regulatory hierarchy—invertible promoters,

trans-locus inhibitors, and possibly other mechanisms—allow for changes in CPS expression but provide a "set point" for return to the original expression profile.

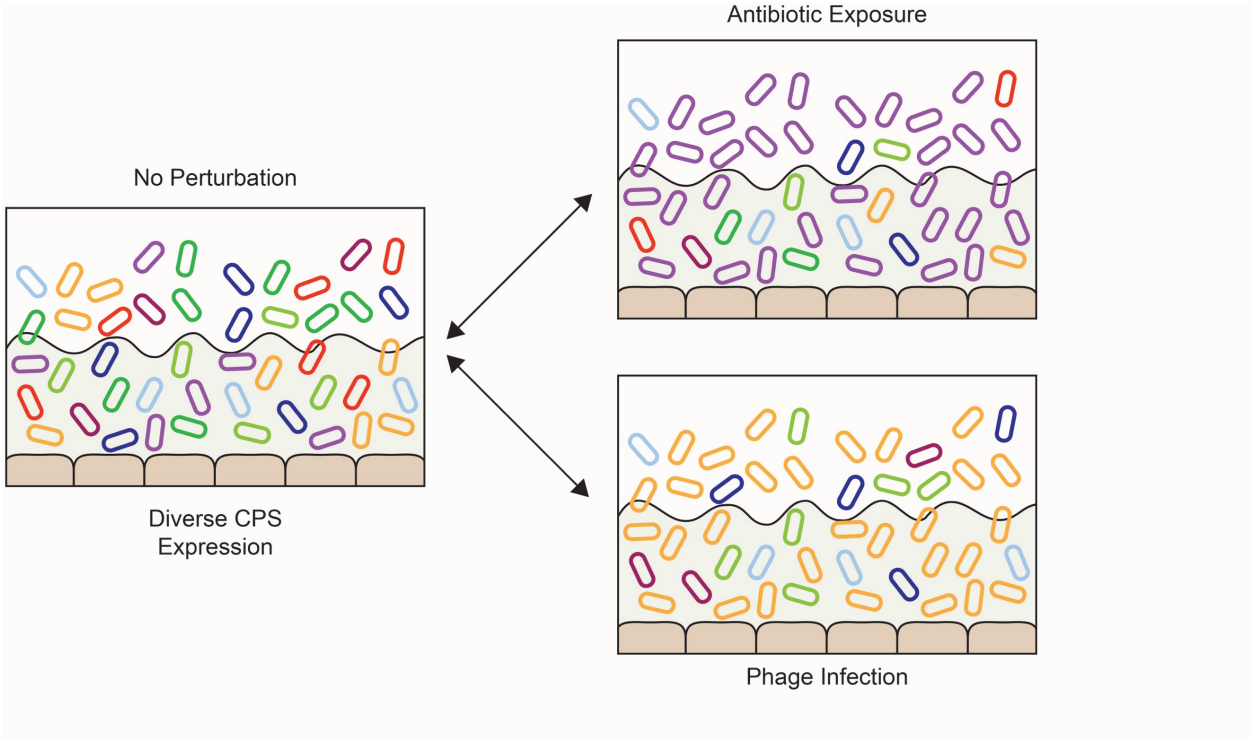


Figure 4.2 Expressing multiple CPS promotes Bacteroides *resilience in the gut.*

Under "normal" conditions (in the absence of strong selective pressures), the B. theta *population expresses various amounts of each CPS type. After exposure to strong selective pressures (e.g. antibiotic exposure, phage infection) a proportion of the population that is resistant to the perturbation experiences outgrowth. Potentially, upon waning of the selective pressure, CPS expression returns to a state allowing expression of multiple CPS types.*

The long-term colonization of mice with wild-type *B. theta* demonstrates this principle of homeostasis in CPS expression (Figure 2.1). CPS locus expression in either wild-type or Rag1$^{-/-}$ mice is mainly characterized by *cps4* expression, with low level expression of other loci at most time points. Occasionally, *cps4* expression decreases dramatically and is temporarily replaced by expression of a different locus until *cps4* again becomes the most abundantly expressed locus. Events such as diet changes (such as the low- to high-fiber dietary oscillations in Figure 2.1) select for the expression of other loci (e.g. *cps5* or *cps6*) but dominant *cps4* eventually resumes in some of these mice. Increased *cps5* gene expression in wild-type *B. theta* after antibiotic

exposure may be another example (Figure 2.8). As the experiment only continued for a few days after antibiotic perturbation, we may have failed to see the wild-type population return to an expression profile that again favored *cps4*.

Bacteriophage infection of the wild-type strain (Figure 3.3) provides another example of the advantages of a population expressing diverse CPS structures. We showed that a population of cells that would have otherwise expressed mainly the *cps3*, *cps4*, and *cps1* loci abolished both *cps1* and *cps4* expression after infection with either of two bacteriophages (Figure 3.3, Figure 3.4). Others have shown that individual clones of a *Bacteroides fragilis* strain infected with phage can be divided into phage resistant and phage susceptible populations[29]. Bacterial colonies that no longer contained phage produced cultures that were again susceptible to phage infection (i.e. they had reverted to a susceptible state after removal of the selective pressure), whereas colonies that still contained phage particles (i.e. with selective pressure maintained) remained resistant. This supports our model that being able to dynamically express and switch expression of CPS allows a proportion of any given population to withstand a range of selective pressures.

In these cases of resilience, individual CPS might not have a specific, defined role to perform. Rather, each CPS may provide an advantage solely by being structurally distinct from the other CPS. Thus, even if a phage is able to target several distinct structural motifs found in various CPS, it is unlikely that a single phage would be able to infect the entire population. Moreover, the enormous diversity in CPS structures among strains provides additional "insurance" by decreasing the odds that a phage released from one strain would also target a neighboring strain.

It is unclear whether encoding for multiple, phase variable capsules occurs frequently outside of the *Bacteroides* and *Parabacteroides* genera. However, one notable example is a

strain of the Gram positive bacterium *Bifidobacterium breve* that encodes biosynthetic loci for two exopolysaccharide (EPS) types. Alternating expression between these two loci is governed by an invertible promoter[30], resembling *Bacteroides* phase-variable CPS expression in some respects. Loci for polysaccharide or oligosaccharide synthesis are very common in sequenced bacterial genomes, comprising approximately 40% of all biosynthetic gene clusters[31]. As encoding genes for synthesizing multiple capsule types appears to provide a great advantage in dealing with many different environmental pressures, we expect that this phenomenon may be more widespread than previously detected.

Future research

Determination of the CPS synthesis locus regulatory hierarchy

Much of the research in this dissertation focused on a panel of *B. theta* strains that are forced to express particular CPS. These strains enabled me to purify individual CPS, isolate a bacteriophage collection (including many that do not infect the wild-type strain), and to isolate the roles that specific CPS may play in both immune evasion and bacteriophage resistance. However, studies with these strains purposely overlook the function of CPS phase variation. Experiments involving the wild-type strain have revealed unexpected findings. For instance, when competing with the strain expressing only CPS5, the wild-type strain expresses a variety of CPS types (Figure 2.8). This raises the question of whether or not the wild-type strain is more fit by adapting to other niches that are more accessible through expression of other CPS types. Another, non-exclusive explanation would be that transcription of *cps5* is repressed by multiple other CPS synthesis loci. In this case, a strong selective pressure would be needed to increase *cps5* expression in the wild-type strain.

Similarly, treatment of wild-type *B. theta* with the phages ARB25 or ARB47 induces drastically decreased expression of *cps4* genes (Figure 3.3, Figure 3.4), even though these phages exhibit little to no replication on the CPS4-expressing strain. This may be due to changes in *trans*-locus repression due to the lack of *cps1* gene expression (and thus no expression of the CPS1 locus *trans*-locus repressor) in the population. While the mechanisms for hierarchical expression of CPS have been elucidated in *B. fragilis*, no work on the CPS locus hierarchy in *B. theta* has been conducted. Our single CPS-expressing strains would facilitate testing of pairwise combinations of individual CPS with individual *trans*-locus inhibitors. Repression of CPS locus transcription in strains containing a single locus should lead to the strain sinking in static culture, similar to the acapsular strain. This sinking phenotype can be determined both qualitatively (through visual examination) and quantitatively (through spectrophotometric readings) and compared to controls expressing no *trans*-locus inhibitors. Follow-up methods directly measuring *cps* gene expression can confirm repression locus transcription. This new information would be extremely informative in identifying the mechanisms underlying the two scenarios outlined above, and future studies would also benefit in understanding this hierarchy in advance.

Defining the mechanism(s) of individual phage infection strategies

The work in this dissertation has demonstrated that bacteriophage targeting *B. theta* must do so in the context of the CPS layer. Individual phages are only able to infect a subset of the strains expressing different CPS (Figure 3.1). Whereas some phages infect only a subset of the bacterial strains producing a CPS, other phages are also able to infect the acapsular strain, implying that other cell surface receptors are involved (e.g. an outer membrane protein, or LPS).

Identifying and confirming cell surface receptors, whether they are protein or glycan, will greatly increase our understanding of the infection strategies of these diverse phages.

Multiple strategies may be undertaken to determine the identity of receptors other than CPS. For example, we and others have successfully used transposon mutant libraries to screen thousands of clones for a phenotype of interest[32], and screening mutants for phage resistance can identify potential receptors. Second, binding assays of phage to bacterial protein have identified putative phage receptors in other bacteria, including in a *B. fragilis* strain[33]. Third, screening outer membrane proteomes of phage-infected vs uninfected samples may also yield insightful hits. For each of these strategies, follow-up assays using gene knockouts or purified protein would be needed to verify potential receptors.

The type strain of *B. thetaiotaomicron* encodes 88 polysaccharide utilization loci (PULs) for degrading and importing distinct glycan substrates[2]. *B. theta* typically upregulates these systems by 100-fold or more in the present of their cognate glycans, and expression of these PULs may also be phase-variable[13]. Interestingly, the first phage receptors identified in *E. coli* include transporters for various nutrients, including iron, vitamin B12, and maltose[34]. It is intriguing to consider that many nutrient transport systems in *B. thetaiotaomicron*, including PULs, may be targeted by phage. Thus, an alternative strategy for identifying phage receptors is to use bacteria grown on a specific polysaccharide (or other nutrient) of interest, which should flood the cell surface with the associated PUL (or other complex for nutrient import). This should enrich for phage isolates that target the degradation machinery. Evaluating the infectivity of the phages in the presence or absence of the polysaccharide of interest (to modulate expression levels of the intended target) will help confirm phages specific for a PUL of interest, and deletions of specific genes in the pathway of interest can subsequently be made and tested.

Finally, few genomes of phage infecting the *Bacteroides* have been sequenced[35]. Sequencing of phenotypically diverse *Bacteroides*-specific phage will provide molecular insight into the diversity of these phages and may identify mechanisms of host adaptation. Besides preventing the use of comparative genomics approaches in studying phage biology, a limited number of genome sequences also hampers the identification of *Bacteroides*-specific phage in metagenomic studies. Sequencing the genomes of the *B. theta*-targeting phages isolated here, as well as sequencing a more diverse collection of Bacteroidetes-associated phages with associated host range data, will greatly increase the usefulness of bioinformatics tools in characterizing phages and their genomes in future studies.

Final conclusions

While mechanistic studies on members of the gut microbiota (including *B. theta*) have increased dramatically in recent years, little is still known about the adaptations that allow commensal microbes to successfully compete and thrive in the gut. My thesis scrutinized the role of CPS in increasing bacterial survival and resilience against diverse threats and changing environmental conditions. Using a collection of uniquely constructed CPS-expressing strains, we were able to identify a subset of these CPS that provided an advantage in the mouse gut, and one CPS (CPS5) that provided a substantial fitness advantage in the presence of adaptive immunity. Perhaps even more important than this work is the potential these strains provide to study the roles of commensal bacterial CPS in multiple settings. We are not aware of any similar collection of isogenic commensal strains. Thus far, we have employed these strains to characterize the monosaccharide composition of each CPS (Figure 2.13), to develop new immune assays (Figure 2.6), and to characterize the role of CPS in many others scenarios related to host diet, host

immune response, and phage predation. We envision that in the future these strains or their derivatives will likewise facilitate many additional studies. An increased understanding of how the *Bacteroides* respond to such challenges will empower future strategies for maintaining specific members of this genus in the gut or even modulating their function. As they target specific bacterial receptors and modulate bacterial gene expression, bacteriophage may provide one strategy to accomplish this goal. Such approaches to modulate key members of the gut microbiome may lead to new methods to improve human health.

References

1. El Kaoutari, A. *et al.* The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11,** 497–504 (2013).
2. 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,** 447–57 (2008).
3. Porter, N. T., Canales, P., Peterson, D. A. & Martens, E. C. A Subset of Polysaccharide Capsules in the Human Symbiont Bacteroides thetaiotaomicron Promote Increased Competitive Fitness in the Mouse Gut. *Cell Host Microbe* **22,** 494–506 (2017).
4. Coyne, M. J. & Comstock, L. E. Niche-specific features of the intestinal Bacteroidales. *J. Bacteriol.* **190,** 736–42 (2008).
5. Patrick, S. *et al.* Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of Bacteroides fragilis. *Microbiology* **156,** 3255–69 (2010).
6. 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.* **290,** jbc.M114.633800 (2015).
7. Round, J. L. *et al.* The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332,** 974–7 (2011).
8. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–5 (2008).
9. Chatzidaki-Livanis, M., Coyne, M. J. & Comstock, L. E. A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of Bacteroides fragilis. *J. Bacteriol.* **191,** 7288–95 (2009).
10. Chatzidaki-Livanis, M., Weinacht, K. G. & Comstock, L. E. Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont Bacteroides fragilis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 11976–80 (2010).
11. Martens, E. C., Roth, R., Heuser, J. E. & Gordon, J. I. Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J. Biol. Chem.* **284,** 18445–18457 (2009).
12. Hickey, C. A. *et al.* Colitogenic Bacteroides thetaiotaomicron antigens access host immune cells in a sulfatase-dependent manner via outer membrane vesicles. *Cell Host Microbe* **17,** 672–680 (2015).
13. Rogers, T. E. *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol. Microbiol.* **88,** 1–15 (2013).
14. 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. U. S. A.* **105,** 13099–104 (2008).

15. 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. U. S. A.* **105,** 3951–6 (2008).

16. Cunnion, K. M., Zhang, H.-M. & Frank, M. M. Availability of complement bound to Staphylococcus aureus to interact with membrane complement receptors influences efficiency of phagocytosis. *Infect. Immun.* **71,** 656–62 (2003).

17. Cress, B. F. *et al.* Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiol. Rev.* **38,** 660–97 (2014).

18. Roberts, I. S. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu. Rev. Microbiol.* **50,** 285–315 (1996).

19. Coyne, M. J., Reinap, B., Lee, M. M. & Comstock, L. E. Human symbionts use a host-like pathway for surface fucosylation. *Science* **307,** 1778–81 (2005).

20. 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).

21. Neff, C. P. *et al.* Diverse intestinal bacteria contain putative zwitterionic capsular polysaccharides with anti-inflammatory properties. *Cell Host Microbe* **20,** 1–13 (2016).

22. Mazmanian, S. K. & Kasper, D. L. The love–hate relationship between bacterial polysaccharides and the host immune system. *Nat. Rev. Immunol.* **6,** 849–858 (2006).

23. Round, J. L. & Mazmanian, S. K. Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12204–9 (2010).

24. Shen, Y. *et al.* Outer membrane vesicles of a human commensal mediate immune regulation and disease protection. *Cell Host Microbe* **12,** 509–20 (2012).

25. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307,** 1955–9 (2005).

26. Bjursell, M. K., Martens, E. C. & Gordon, J. I. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. *J. Biol. Chem.* **281,** 36269–79 (2006).

27. Pudlo, N. A. *et al.* Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. *MBio* **6,** e01282-15 (2015).

28. Sommer, F., Moltzau Anderson, J., Bharti, R., Raes, J. & Rosenstiel, P. The resilience of the intestinal microbiota influences health and disease. *Nat. Publ. Gr.* (2017). doi:10.1038/nrmicro.2017.58

29. Keller, R. & Traub, N. The characterization of Bacteroides fragilis bacteriophage recovered from animal sera: observations on the nature of Bacteroides phage carrier cultures. *J. Gen. Virol.* **24,** 179–89 (1974).

30. Fanning, S. *et al.* Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc. Natl. Acad. Sci.* **109,** 2108–2113 (2012).

31. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158,** 412–421 (2014).

32. Goodman, A. L. *et al.* Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6,** 279–89 (2009).

33. Puig, A., Araujo, R., Jofre, J. & Frias-Lopez, J. Identification of cell wall proteins of Bacteroides fragilis to which bacteriophage B40-8 binds specifically. *Microbiology* **147,** 281–8 (2001).

34. Braun, V. FhuA (TonA), the career of a protein. *J. Bacteriol.* **191,** 3431–3436 (2009).

35. Ogilvie, L. A. & Jones, B. V. The human gut virome: A multifaceted majority. *Front. Microbiol.* **6,** 1–12 (2015).